

UNIVERSITA' DEGLI STUDI DI FERRARA

FACOLTA' DI LETTERE E FILOSOFIA

C.A.R.I.D.

Centro di Ateneo per la Ricerca, l'Innovazione Didattica e l'Istruzione a Distanza

CORSO DI LAUREA

in

**TECNOLOGO DELLA COMUNICAZIONE
AUDIOVISIVA E MULTIMEDIALE**

Tesi di Laurea

Titolo

“Computer Vision”, un passo verso la vita artificiale?

Relatore:

prof. Andrea Pizzirani

Correlatore:

prof. Paolo Frignani

Laureando:

Gianni Gazzetto

Anno Accademico 2006/2007

Indice generale

Introduzione.....	5
Il processo visivo: cosa significa “vedere”?	13
Il fondamento: la luce.....	17
L'apparato ottico nell'uomo e negli altri esseri viventi.....	24
L'occhio umano.....	27
Funzione e Struttura	28
Posizionamento e visione stereoscopica.....	28
Formazione dell'immagini.....	30
La cornea e l'umore acqueo.....	35
La pupilla e l'iride.....	36
Il cristallino.....	37
Umore vitreo.....	38
La retina.....	38
Il “processo di riconoscimento” delle immagini.....	49
Il cervello.....	49
Dall'occhio al cervello.....	50
La corteccia cerebrale.....	53
La percezione integrata.....	62
Le cellule neurali, codifica ed elaborazione dei segnali.....	65
Approcci alla visione.....	73
Strutturalismo.....	75
Gestalt.....	77
La teoria ecologica di Gibson.....	81
Costruttivismo.....	83
La Teoria dell'elaborazione dell'informazione.....	86
La metafora del computer.....	87
Il linguaggio visivo.....	95
Cos'è la Computer Vision?.....	99
Definizioni.....	99
Legami tra tecnologia e arte.....	103
Discipline correlate.....	104
Le immagini digitali.....	108
Cos'è un'immagine?.....	108
Cosa sono le immagini digitali?.....	109
Alcune nozioni per l'acquisizione.....	111
Immagini a modulazione d'intensità/luminanza.....	111
Campionamento.....	112
Quantizzazione.....	114
Il rumore.....	115
Immagini spaziali o di profondità.....	116
Strumenti di acquisizione.....	117
Scanner.....	117
Fotocamera Digitale.....	118

Telecamere.....	119
Elaborazione automatica dell'immagine.....	120
Come funziona? - Tecnologie implementate.....	126
Informazioni dall'immagine.....	126
L'istogramma dei livelli di grigio.....	127
Spettro di frequenza spaziale.....	128
Operazioni puntuali.....	130
L'operazione di soglia.....	131
Operazioni locali.....	132
Operazioni aritmetiche.....	133
Operazioni logiche.....	134
Filtri.....	134
Filtri passa-basso – riduzione del rumore.....	135
Filtri passa-alto – aumento del dettaglio.....	137
Rilevazione dei bordi.....	137
Rilevazione delle linee.....	141
Estrazione di contorni.....	141
Texture.....	143
Operazioni globali.....	145
Operazioni con l'istogramma d'intensità.....	145
La trasformata di Fourier.....	149
Estrazione di superfici dalle ombre.....	151
Stereo-Fotometrica.....	151
La profondità.....	152
Profondità dalle immagini d'intensità.....	152
La profondità nelle immagini spaziali.....	155
Sistemi di Visione Attiva.....	156
Operazioni a livello di oggetti.....	157
Superfici.....	158
Il movimento.....	159
Riconoscimento degli oggetti.....	162
Classificazione degli oggetti - Pattern recognition.....	165
Relazioni con l'intelligenza artificiale.....	169
Possibili applicazioni.....	173
Quali saranno le conseguenze?.....	183
Conseguenze economiche.....	187
Conseguenze sociali.....	192
Conseguenze psicologiche.....	199
Conclusioni.....	211
Bibliografia.....	215
Sito-grafia.....	219
Filmografia.....	225

Introduzione

Le motivazioni che spingono gli uomini a compiere determinate azioni non sono sempre chiare agli stessi, la consapevolezza spesso è una dura conquista. A volte le scelte sono casuali; in altri casi, si segue semplicemente la strada già tracciata; più raramente si riesce a valutare e programmare il prosieguo delle azioni che incideranno sulla propria esistenza.

Una premessa così filosofica sembra avere poco a che fare con l'argomento che mi accingo ad affrontare: un nuovo ambito della ricerca applicata denominata “computer vision”¹ e le sue possibili relazioni con altri campi tra cui la vita artificiale. Il termine computer richiama una visione tecnica, matematica, automatica; parlando di vita artificiale il pensiero di molti va ai più recenti film di fantascienza o alle ricerche più avanzate della scienza.

Eppure alcune loro applicazioni sono già entrate nelle nostre case, a volte senza che ce ne sia resi conto. È sufficiente ricordare, per la prima, un semplice programma per il riconoscimento, tramite scansione elettronica, di testi cartacei e, per la seconda il Tamagochi, quel gioco elettronico, che, simulando un pulcino lasciato alle cure dei bambini, è stato causa delle preoccupazioni di più di qualche genitore.

Diffondere la conoscenza della tecnologia che influenza la nostra vita: questo è il motivo che mi ha spinto e mi spinge a riflettere sull'argomento, a trattare l'ampio tema² del rapporto tra la tecnologia, la scienza e l'uomo.

L'obiettivo di questo lavoro vorrebbe essere di diffondere la conoscenza di alcuni aspetti tecnologici della computer vision, finora riservati a professionalità specialistiche quali ingegneri o programmatori informatici tra un pubblico più ampio, con l'auspicio di suscitare interesse e riflessioni sui possibili sviluppi di questa tecnica e prospettare i possibili vantaggi³.

1 Si usa anche “Machine vision”, ma il termine ha una accezione riguardante le applicazioni industriali, e viene tradotto in italiano con “visione artificiale”, la preferenza da me accordata a “computer vision” è in ragione della sua diffusione e alla generalità della materia cui si riferisce.

2 Per la distinzione fra tema e argomento si veda Max Giovagnoli, *Come si fa una tesi di laurea con il computer e internet*, Milano, Tecniche Nuove, 2003 pp. 18-19

3 Questo è quanto avviene normalmente all'interno di un gruppo di lavoro, ossia, uno scambio di conoscenze e competenze tra i membri che lo compongono al fine di giungere a un risultato migliore. La comunicazione viene quindi ad assumere un ruolo essenziale, tanto che spesso per

L'opportunità di valutare in anticipo le possibili evoluzioni di una ricerca non è sempre stata a disposizione dell'uomo, anzi, in passato, alcune scoperte sono giunte casualmente (es: il fuoco, la ruota, la polvere da sparo, il nastro adesivo). Tuttavia prevedere il possibile evolversi di alcune ricerche sta divenendo necessario per un efficiente ed efficace, utilizzo delle risorse disponibili e quindi assicurare un futuro migliore per l'uomo e il suo ambiente. Infatti, come dice Antonio Zichichi: “Spazio, Tempo, Massa, Energia e Cariche sono le quantità fondamentali della nostra esistenza nell'Immanente. E sono tutte finite.”⁴. In tal senso la computer vision può essere sia oggetto di osservazione, sia uno strumento di osservazione per altri fenomeni.

Il tema potrebbe assumere un carattere più generale riferendosi non solo a “come”, e cioè con quali artefatti, si possano tecnicamente raggiungere certi risultati/prodotti che soddisfano i nostri bisogni, ed eventualmente quali siano questi bisogni, ma soprattutto il “perché”, i motivi che ci spingono verso queste mete, e la valutazione di queste mete/bisogni.

E' chiara, fin da queste prime righe, la mia ispirazione a uno dei più conosciuti teorici dei media, Marshall McLuhan, e alle sue riflessioni: “In una cultura come la nostra, abituata da tempo a frazionare e dividere ogni cosa al fine di controllarla, è forse sconcertante sentirsi ricordare che, per quanto riguarda le sue conseguenze pratiche, il *medium* è il messaggio. Che in altre parole le conseguenze individuali e sociali di ogni medium, cioè di ogni estensione di noi stessi, derivano dalle nuove proporzioni introdotte nelle nostre questioni personali da ognuna di tali estensioni o da ogni nuova tecnologia.”⁵ McLuhan, non sa o non vuole pronunciarsi sulla bontà del processo di estensione della comunicazione⁶ e lo considera un problema aperto a “...un'ampia gamma di soluzioni. Ma è praticamente impossibile dare una risposta a qualsiasi domanda su tali estensioni dell'uomo senza prenderle tutte in esame nel loro insieme”⁷.

È questo il tipo di approccio necessario per la comprensione di ogni medium, un

facilitarla sono inserite figure specifiche come il comunicatore. Cfr. G.P. Quaglino, S. Casagrande, A. Castellano, *Gruppo di Lavoro Lavoro di Gruppo*, Milano, Raffaello Cortina Editore, 1992, p. 109.

4 Intrage, *L'universo tra scienza e fede*, in rete alla URL <http://interviste.intrage.it/zichichi/> 03 luglio06

5 Marshall McLuhan, *Gli strumenti del comunicare*, trad. it. di E. Capriolo, Milano, Net, 2002, p. 15. (ed. orig. *Understanding Media*, 1964).

6 Ibidem p. 9

7 Ibidem

approccio sistemico, cioè il considerare le interazioni dell'elemento preso in esame con gli altri elementi del sistema. Per comprendere il processo visivo è necessario esaminarlo in relazione agli altri sensi, alla funzione svolta per l'uomo: la funzione percettiva, di relazione con l'ambiente, il mondo, il reale.

Significativa, al riguardo, è una famosa metafora di David Marr, uno dei primi e principali ricercatori a proporre un approccio computazionale nell'ambito delle scienze visive: “Come è impossibile capire il volo di un uccello esaminando solo le penne delle ali, ma è necessario analizzare i vincoli e le limitazioni della gravità al volo e come l'uccello riesca a superarli, così per capire la visione non è sufficiente occuparsi solo della neurofisiologia della visione, ma è necessario individuare una teoria che renda conto delle limitazioni, dei vincoli che si oppongono alla visione e di come è possibile neutralizzarli. Non è sufficiente avere una conoscenza di un solo livello, ma si deve essere capaci di descrivere le risposte dei neuroni, di predire i risultati degli esperimenti di psicologia e di scrivere un programma al calcolatore che analizzi e interpreti gli input visivi nel modo desiderato”⁸. Tale affermazione si apre a varie interpretazioni come afferma Tomás Maldonado, esperto e critico d'informatica, “una metafora, talvolta, nasconde qualcosa di più che un sapere insufficiente sulle questioni discusse”. In effetti, la visione è sempre stata un argomento che dà luogo a molte controversie; probabilmente, nei suoi pensieri, Marr voleva chiedere collaborazione e convincere i più dubbiosi del fatto che riprodurre la visione è possibile⁹.

McLuhan propone una prospettiva storica, identificando un'era meccanicistica di partenza, caratterizzata dalla frammentazione delle attività umane, per arrivare all'era elettrica, dove vige l'automazione e si ha una ricomposizione dell'uomo sia nei ruoli sia nelle capacità richieste per ricoprirli. Egli individua soprattutto le relazioni tra medium-società-individui, le peculiarità sociali che hanno portato alla realizzazione di un determinato medium, come questo influenzi la collettività e i singoli, le spinte evolutive e le possibili dinamiche. Cercherò in questo mio elaborato di fare un percorso simile, partendo dall'uomo, e dal suo principale senso per l'orientamento e la costruzione della

8 Riccardo Valle, *Introduzione*, in *Mente umana, mente artificiale*, a cura di Riccardo Valle, Milano, Giangiacomo Feltrinelli, 1989 pp. XI-XII

9 Cfr Tomás Maldonado, *Critica della ragione informatica*, Milano, Giangiacomo Feltrinelli, 2006 p. 170

propria immagine del mondo¹⁰, la vista, al fine di comprendere come le più recenti ricerche tentino di riprodurre automaticamente il processo. Concluderò ipotizzando quali possono essere gli effetti di una tale tecnologia a livello economico, sociale, psicologico, ecc. nella consapevolezza che, tale conclusione, è più un punto di partenza che di arrivo.

È mia intenzione dimostrare, nel più ampio dibattito che riguarda il rapporto uomo-tecnologia, come quest'ultima abbia una valenza positiva (presa di posizione che McLuhan non si è sentito di fare), e, in particolare, come la computer vision, sia necessaria alla nostra stessa sopravvivenza. Cercherò, di convincere, come Marr, anche i più scettici (un esempio di come vi sia timore verso le nuove tecnologie può essere dato da alcune opere cinematografiche anche recenti come Terminator 3).

Per quanto riguarda il metodo, penso di essermi comportato come in tutti i casi in cui si affronta un argomento abbastanza nuovo e poco conosciuto: partendo da una base, ho espanso le mie cognizioni inizialmente a macchia d'olio, successivamente secondo interesse e pertinenza. Il primo spunto è venuto dal testo “Le immagini digitali” di Roberto Marangoni e Marco Geddo¹¹, dove gli autori dedicano, senza particolari approfondimenti, un capitolo alla computer vision. Intuendo le potenzialità di questa tecnologia ho cercato documenti e informazioni, prima in internet, per semplicità di accesso e reperibilità, poi in biblioteca, per monografie e miscellanee più complete, ritornando a internet per i documenti riguardanti i progetti più recenti. Dal punto di vista comunicativo è interessante notare il fatto che alcuni testi sono in lingua inglese, anche se gli autori sono italiani. Questo prova l'oramai avvenuta affermazione dell'inglese come “lingua ufficiale” in ambito scientifico.

Nel mio procedere ho sempre tenuto presente tre concetti che, penso, sia utile riportare:

- a “L'aver pensato rettamente non è un merito: statisticamente è quasi inevitabile che tra le molte idee sballate, confuse o banali che gli si presentano alla mente, qualcuna che ve ne sia di perspicua o addirittura geniale; e com'è venuta a lui, può esser certo che sarà venuta pure a qualcun altro.”¹²questo è quanto afferma Italo Calvino, attraverso il suo personaggio, in Palomar. Di

10 Cf. P. Frignani, P. Rizzati, *Didattica della comunicazione*, Ferrara, Tecom Projet, 2003 pp. 25-26

11 Roberto Marangoni, Marco Geddo, *Le Immagini digitali*, Milano, Hoepli Editore, 2000 pp. 74-77

12 Italo Calvino, *Palomar*, Milano, Arnoldo Mondadori editore, 2003 p. 103

conseguenza ho cercato di documentare il più possibile le mie ipotesi, seguendo le tracce di chi ha già considerato certi argomenti e confrontando le conclusioni con altri autori. Inoltre, dove possibile, ho tentato di comprendere l'influenza esercitata dalla cultura del tempo sulle tesi esposte¹³.

b Il “Discorso sul metodo” e sulla morale provvisoria di Cartesio¹⁴. Per questi il sapere assume la caratteristica del dubbio: per ogni concetto è necessario dimostrare con evidenza che è assolutamente chiaro, privo di qualsiasi confusione con altre nozioni oscure e incerte, quindi inattaccabile da ogni immaginabile dubbio¹⁵. Come lui e la maggior parte di scienziati e ricercatori, ho incontrato notevoli difficoltà ad applicare quest'atteggiamento: il dubbio e la ricerca di chiarezza richiedono tempi lunghi, forse infiniti. Nel frattempo, per non rimanere immobili, sono necessarie delle regole provvisorie, definite da Cartesio “morale provvisoria”:

1. “La prima era di obbedire alle leggi e ai costumi del mio paese, serbando fede alla religione nella quale Dio mi ha fatto la grazia di essere educato sin dall'infanzia, e regolandomi nel resto secondo opinioni più moderate, lontane da ogni eccesso, e comunemente seguite dalle persone più assennate con le quali dovevo vivere.”¹⁶
2. “La seconda massima era di esser fermo e risoluto, per quanto potevo, nelle mie azioni, e di seguire anche le opinioni più dubbie, una volta che avessi deciso di accettarle, con la stessa costanza come se fossero le più sicure: imitando in ciò i viaggiatori, i quali, se si trovano smarriti in una foresta, non debbono aggirarsi or di qua e ora di là, e tanto meno fermarsi, ma

13 Riprendendo con ciò la tesi di Paul K. Feyerabend: i fatti non esistono “nudi”, ovvero al di fuori delle teorie, ma solo nell'ambito di determinati “quadri mentali”, in quanto lo scienziato vede solo ciò che questi lo inducono a vedere.

Cfr. P. Frignani, P. Rizzati, *Didattica della comunicazione*, Op. cit. p. 293

14 Salvatore Guglielmino, Herman Grosser, *Il sistema letterario – Guida alla storia letteraria e all'analisi testuale – Cinquecento e Seicento Settecento*, Milano, G. Principato, 1988 pp. 250-254

15 Cfr. Salvatore Guglielmino, Herman Grosser, *Il sistema letterario – Guida alla storia letteraria e all'analisi testuale – Cinquecento e Seicento Settecento*, op. cit., p. 250

16 Ibidem

camminare sempre nella stessa direzione, e non mutarla per deboli ragioni, ancorché, l'abbiano scelta a caso, perché, così, anche se non vanno proprio dove desiderano, arriveranno per lo meno, alla fine, in qualche luogo dove probabilmente si troveranno meglio che nel fitto della boscaglia.”¹⁷

3. “La mia terza massima fu di vincere sempre piuttosto me stesso che la fortuna, e di voler modificare piuttosto i miei desideri che l'ordine delle cose nel mondo; e in generale di assuefarmi a credere che nulla, all'infuori dei nostri pensieri, è interamente in nostro potere, in modo che, quando abbiam fatto del nostro meglio riguardo alle cose che sono fuori di noi, se qualcosa non ci riesce, vuol dire ch'essa non dipende assolutamente da noi.”¹⁸

Quindi, nell'impossibilità di una verifica puntuale e completa, terrò per valide le teorie pertinenti più diffuse ed affermate, eventualmente evidenziando le problematiche e le possibili alternative, cercando di evitare contraddizioni e, in fine, che ulteriori approfondimenti potrebbero cambiare quanto esposto.

- c “Nessuno scienziato pensa in formule” è con questa idea che Einstein e Infeld scrissero “L'evoluzione della fisica. Sviluppo delle idee dai concetti iniziali alla relatività e ai quanti.”, chiaramente con l'obiettivo di diffondere il più possibile la conoscenza della materia. Presumo, inoltre, che essi stessi fossero convinti che, la conoscenza del mondo, avvenisse per ipotesi che debbono necessariamente essere verificate, ponendo così una congiunzione tra il mondo percepito dai nostri sensi e il mondo fisico misurabile con gli strumenti di volta in volta adeguati.

Come loro, è mia intenzione utilizzare un linguaggio accessibile, evitando, la formalizzazione matematica, specificando il significato delle sigle e/o acronimi, e servendomi di termini di uso comune, come consigliato da Tullio De Mauro¹⁹. Questo non significa, tuttavia, che non impiegherò la necessaria

17 Ibidem

18 Ibidem

19 <http://www.tulliodemauro.it/default.htm> 9 gennaio 2007

terminologia tecnica, nel tentativo di semplificare il messaggio a tutti i costi. Significa che, utilizzerò gli strumenti disponibili (immagini, diagrammi, disposizione visiva del testo, ecc), al fine di rendere comprensibile l'esposizione²⁰, nonché la necessaria ridondanza per una semplice memorizzazione dei concetti e una scorrevole lettura .

20 Cfr. Riccardo Fedriga, *Il Cittadino Lettore*, Milano, Sylvestre Bonnard, 2005 p. 29

Il processo visivo: cosa significa “vedere”?

“In principio Dio creò il cielo e la terra. Ora la terra era informe e deserta e le tenebre ricoprivano l'abisso e lo spirito di Dio aleggiava sulle acque.

Dio disse: «Sia la luce!». E la luce fu.”

Gen 1,1-3²¹

Senza voler parafrasare Isaac Asimov che, nel suo libro “In Principio”, interpreta parola per parola il capitolo della Genesi dal punto di vista storico e scientifico, questa citazione mi permette d'introdurre alcune riflessioni, prima sulla natura del problema e poi sulla sua storia e le possibili relazioni con altri campi del sapere²².

Vedere non è il risultato di soli processi biologici, ma anche psicologici, e interessa vari aspetti dell'esistenza. Per tale ragione, sull'argomento sono intervenuti fisici, fisiologi, esponenti religiosi, filosofi, artisti, neuro-psicologi, linguisti e, da buoni ultimi, chi si occupa di scienze matematiche e d'informatica²³. Dal punto di vista storico, tutte le civiltà che si sono succedute nel corso dei secoli, oltre a interrogarsi sull'origine della vita e dell'universo (cosmogonia), si sono chieste quale fosse il rapporto tra l'uomo e il mondo, e in specie si sono interessate alla percezione visiva quale canale di relazione. I filosofi greci, ad esempio, pensavano che il processo visivo fosse dovuto a particelle che uscivano dagli occhi per toccare gli oggetti, come dita che tastano²⁴.

Vedere può apparire intuitivamente semplice, ma una “visione²⁵” più attenta si presenta

21 AAVV, *La bibbia di Gerusalemme*, Bologna, Centro Editoriale Dehoniano, 2000 p. 35

22 Cfr Arnaldo Bagnasco, Marzio Barbagli, Alessandro Cavalli, *Corso di sociologia*, Bologna, il Mulino, 1997 p.174

23 Cfr. Stephen E. Palmer, *Vision Science – photons to phenomenology*, USA, MIT Press, 1999 p. 4

24 Richard L. Gregory, *Occhio e cervello – La psicologia del vedere*, trad. it. Alberta Rebaglia, Milano, Raffaello Cortina, 1998 p. 1, (ed. originale *Eye and Brain*, 1998)

25 Sembra un gioco di parole, ma mi riferisco alla differenza evidenziata da Denis Kaufman (Dziga Vertov, teorico e regista cinematografico russo) tra guardare, fermarsi superficialmente sugli aspetti effimeri, e vedere, ossia, partendo dall'osservazione degli stessi aspetti, trarre delle conclusioni non effimere, sostanziali.

Cfr. Carlo Alberto Pinelli, *L'ABC del documentario*, Roma, Dino Audino Editore, 2001 cap. 4

come fenomeno complesso, la cui funzione è certamente quella di “acquisire una conoscenza del nostro mondo”²⁶, mettendoci in relazione con ciò che ci circonda al fine ultimo di sopravvivere²⁷, reagendo agli eventi che stanno accadendo. È chiaro che per riuscire in questo vi è una stretta interdipendenza con gli altri sensi, in via primaria con il tatto²⁸, che permette di dare certezza alla percezione visiva.

Parlando di sopravvivenza, è d'obbligo richiamare esplicitamente la teoria dell'evoluzione di Charles Darwin, esposta nell'opera “L'origine delle specie” e alle sue conseguenze²⁹. Queste riguardano vari aspetti, che si possono ricondurre, da un lato al mondo scientifico, dall'altro alla sfera sociale. Pur non volendo entrare nel dibattito mai del tutto assopito³⁰ tra favorevoli e contrari, dico subito che nel prosieguo dell'elaborato ritengo tale tesi probabile, e con poche remore³¹. Darwin stesso, comunque, ammetteva alcuni scogli, come è evidente nel capitolo “Difficoltà della Teoria”³². È singolare che uno di questi riguardasse la formazione di organi perfetti come gli occhi. Le loro capacità, la loro struttura, pongono serie difficoltà a credere che l'evoluzione, attraverso la selezione naturale, abbia potuto raggiungere tale risultato. Ulteriore perplessità era dovuta alle modalità con cui si possono acquisire e modificare gli istinti, quindi non tanto la struttura/forma fisica/biologica dell'animale uomo, ma i suoi comportamenti. Ancora oggi tali problematiche sono, per certi versi, irrisolte e che si ripresentano nel riprodurre artificialmente organi complessi come gli occhi o processi comportamentali-sociali³³.

Vedere quindi è essenzialmente un “processo” per l'acquisizione d'informazioni-

26 Semir Zeki, *La visione dall'interno – Arte e cervello*, trad. it. Paolo Pagli e Giovanna de Vivo, Torino, Bollati Boringhieri, 2003 p. 21, (ed. originale *Inner Vision. An Exploration of Art and the Brain*, 1999)

27 Cfr. Richard L. Gregory, *Occhio e cervello – La psicologia del vedere*, op. cit., p. 3

28 Ibidem p. 9

29 <http://it.wikipedia.org/wiki/Evoluzione> 23 agosto 2006

30 Cfr. <http://mmmgroup.altervista.org/i-darwin.html> 23 agosto 2006

31 Sono interessanti, sull'argomento, le interviste fatte a Jérôme Lejeune (1926-1994), genetista di fama internazionale, che dà la propria opinione in merito al rapporto tra scienza e fede, reperibili ai seguenti indirizzi:

<http://www.disf.org/ScienziatiCredenti/Lejeune.asp> 23 agosto 2006

<http://lgxserver.uniba.it/lei/rassegna/021006b.htm> 23 agosto 2006

32 http://pages.britishlibrary.net/charles.darwin/texts/origin_6th/origin6th_06.html 23/08/2006

33 Richard L. Gregory, *Occhio e cervello – La psicologia del vedere*, op. cit., pp. 35-36

conoscenza sull'ambiente-mondo che ci circonda (e nel quale siamo immersi) e implica il coinvolgimento di organi biologici e di fattori cognitivi. Nel momento in cui cerchiamo di simulare la vista biologica, da considerarsi come punto di riferimento³⁴ (anche se il parallelismo tra le due presenta delle difficoltà), dovremo valutare non solo le caratteristiche dell'hardware ma soprattutto i requisiti del software, che deve fare riferimento a una conoscenza acquisita³⁵, nonché possedere la capacità di modificarla, tipici temi trattati dall'intelligenza artificiale³⁶.

Per comprendere è quindi necessario conoscere “l'hardware” fornitoci dall'evoluzione biologica; questo potrebbe anche non essere sufficiente, come dichiara David H. Hubel: “All'inizio potremmo pensare che una conoscenza dettagliata delle connessioni dell'occhio con il cervello e dei circuiti all'interno del cervello sia sufficiente a permettere di dedurre i principi generali di funzionamento. Purtroppo questo si verifica solo in minima parte.”³⁷. Ma “Quando abbiamo a che fare con una invenzione dell'uomo, non abbiamo alcun dubbio che il macchinario e le sue parti abbiano funzioni comprensibili. Per capirle è sufficiente leggere le istruzioni relative. Analogamente in biologia, abbiamo sviluppato la fede nel valore funzionale, e in definitiva nella possibilità di comprensione, di strutture che non abbiamo inventato, ma che si sono sviluppate e perfezionate nel corso dei milioni di anni di evoluzione. Il problema del neurobiologo (non il solo, per la verità) è quello di cogliere le correlazioni tra l'ordine e la complessità di una struttura e la sua funzione.”³⁸. Se “l'hardware” non è di semplice comprensione, quando passiamo al “software” il problema si complica ulteriormente, in quanto l'oggetto d'indagine siamo noi stessi e le nostre sensazioni, quindi un campo altamente soggettivo³⁹. Infatti numerose sono le teorie psicologiche che propongono un quadro interpretativo, tra le altre quelle recenti della psicofisica che si legano a quelle

34 Cfr. Martin D. Levine, *Vision in Man and Machine*, USA, McGraw-Hill, 1985 p. XIV

35 Ramesh Jain, Rangachar Kasturi, Brian G. Schunck, *Machine Vision*, U.S.A., McGraw-Hill, 1995 pp. 5-6

36 Marco Pedroni, Giorgio Poletti, *U.D. Fondamenti di Informatica*, S. Bartolomeo in Bosco (Fe), Tecomproject, 1999 pp. 142-143

37 David H. Hubel, *Occhio, Cervello e Visione*, trad. it. Enrica Maria Fava, Bologna, Nicola Zanichelli Editore, 1989 p. 14, (ed. originale *Eye, Brain, and Vision*, 1988)

38 Ibidem p. 15

39 Cfr Wolfgang Köhler, *La psicologia della Gestalt*, tra. it. Giannantonio De Toni, Milano, Giangiacomo Feltrinelli, 1984 pp. 13-30, (ed. originale *Gestalt Psychology*, 1947)

della neuropsicologia e della psicobiologia⁴⁰.

Avendo quindi un'idea di quante e quali possono essere le possibili vie d'approfondimento, iniziamo tenendo presente che la visione umana è fondata sulla percezione della luce, come sottolineano i primi versetti del capitolo biblico della genesi, anche se esistono sistemi alternativi per la costruzione di immagini che si basano su segnali diversi, come l'emissione di raggi laser o onde⁴¹ (ad esempio i pipistrelli, o il sonar)... forse gli antichi greci non erano poi così in errore.

40 Cfr. Martin D. Levine, *Vision in Man and Machine*, op. cit., pp. 6-9

41 Emanuele Trucco, Alessandro Verri, *Introductory Techniques for 3-D Computer Vision*, New Jersey, Prentice Hall, 1998 p. 16

Il fondamento: la luce

La luce è fondamentale non solo per vedere ma anche per vivere. Essa è essenziale per la fotosintesi⁴² degli organismi vegetali che sono, a loro volta, il primo anello della catena alimentare (non a caso si ritiene i primi organismi viventi furono le alghe grigio-azzurre⁴³). Come potremmo vivere senza la luce? Cerchiamo di comprenderla più da vicino.

Per Newton, è necessario distinguere il mondo fisico dal mondo percepito. Nel primo trattiamo grandezze fisiche che sono oggettivamente misurabili, nel secondo tutto è soggettivo e non misurabile⁴⁴. La distinzione, fatta per motivi d'analisi, potrebbe generare qualche confusione⁴⁵ in quanto i due mondi sono collegati: tutti abbiamo esperienza diretta della luce ma definirla a livello fisico non è semplice.

In fisica, la natura della luce è stata oggetto di molte ricerche, tanto che ne costituisce una parte specifica denominata ottica⁴⁶. Inizialmente, come riporta Richard L. Gregory, si sono affermate due teorie, tra loro rivali: “Isaac Newton riteneva fosse costituita da un flusso di particelle; mentre il fisico olandese Christiaan Huygens sosteneva che gli impulsi luminosi si propagavano attraverso un mezzo impercettibile e omni-pervasivo – l'etere -, concepito sotto forma di sfere elastiche in contatto fra loro: qualunque perturbazione verrebbe trasmessa in tutte le direzioni dalle sfere, accostate le une alle altre, sotto forma di un'onda; onda che costituisce la luce”⁴⁷.

Possiamo definire la prima teoria corpuscolare e la seconda ondulatoria, ma entrambe non spiegavano completamente il fenomeno. Si è così passati, con James Clerk Maxwell⁴⁸, alla teoria elettromagnetica, utilizzata oggi nella maggioranza delle

42 http://it.encarta.msn.com/encyclopedia_761572911_2/Fotosintesi.html 27 agosto 2006

43 http://www.funsci.com/fun3_it/protisti/mostra.htm 27 agosto 2006

44 <http://www.boscarol.com/pages/cs/000-colore.html> 26 agosto 2006

45 Cfr Wolfgang Köhler, *La psicologia della Gestalt*, Op. cit., pp. 13-30

46 <http://it.wikipedia.org/wiki/Ottica> 28 settembre 2006

47 Richard L. Gregory, *Occhio e cervello – La psicologia del vedere*, Op. cit., p.22

48 http://it.wikipedia.org/wiki/James_Clerk_Maxwell 27 agosto 2006

http://en.wikipedia.org/wiki/James_Clerk_Maxwell 27 agosto 2006

applicazioni, e poi alla teoria quantistica di Max Planck⁴⁹ (ripresa anche da Albert Einstein per la teoria della relatività).

La teoria elettromagnetica migliora quella ondulatoria, interrogandosi sulla natura delle onde⁵⁰. Lo spazio è ivi definito come campo o spettro elettromagnetico (uno spazio in ogni punto del quale esistono sia azioni elettriche sia azioni magnetiche), e la luce come insieme di onde elettromagnetiche.

Per risolvere alcuni problemi sulla trattazione del corpo nero⁵¹, Max Planck propose che lo scambio di energia fra il campo elettromagnetico e la materia potesse avvenire solo tramite pacchetti discreti di energia (quanti) chiamati fotoni. Successivamente Albert Einstein mostrò, studiando l'effetto fotoelettrico, che i fotoni non erano solo un artificio matematico, ma oggetti reali.

Si arriva quindi al concetto dualistico della luce come onda-particella, i quanti “possiedono sia le proprietà caratteristiche dei corpuscoli sia quelle delle onde, tant'è che in molti casi il fotone viene descritto come un “pacchetto” d'onde, composto da un numero tanto maggiore di onde quanto minore è la loro lunghezza.”⁵²

Le onde luminose (formate da un'alternanza di cavi e creste) sono caratterizzate da quattro grandezze fondamentali: l'ampiezza (il massimo spostamento dell'onda in altezza), la lunghezza d'onda (la distanza tra due creste successive), la frequenza (il numero di oscillazioni complete compiute in un secondo) e il periodo (l'intervallo di tempo in cui l'onda compie un'oscillazione completa).

49 http://it.wikipedia.org/wiki/Max_Planck 27 agosto 2006

http://en.wikipedia.org/wiki/Max_Planck 27 agosto 2006

50 Diffusa era la metafora tra luce ed acqua, la prima si propaga nello spazio come le vibrazioni della seconda sulla superficie dello stagno; ma se per l'onda nello stagno è chiaro che è l'acqua a vibrare, nel caso delle onde luminose, cos'è che vibra?

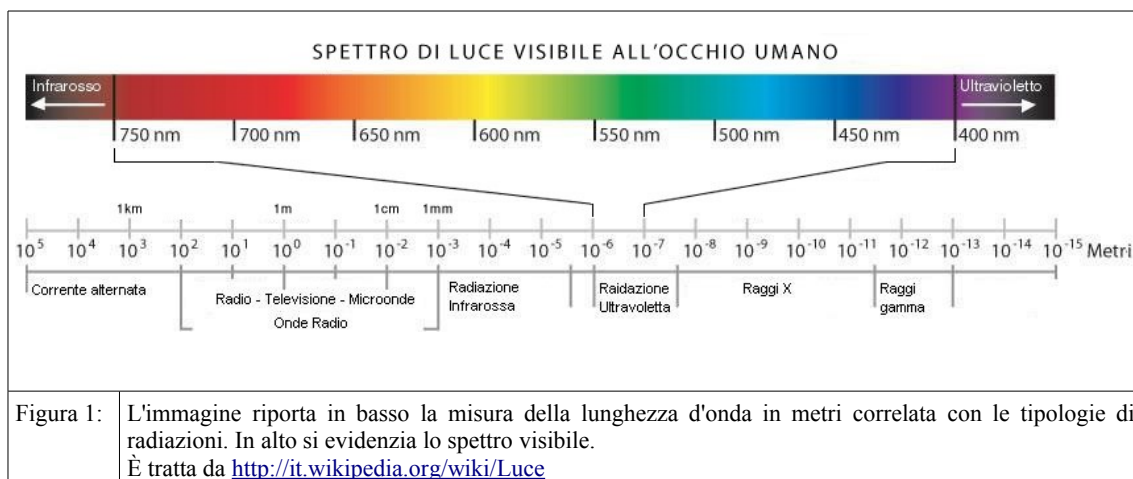
http://www.geocities.com/codadilupo_2000/lucesuono.htm 27 agosto 2006

51 Un oggetto che assorbe tutta la radiazione elettromagnetica incidente (e quindi non ne riflette). Nonostante il nome, il corpo nero irradia comunque, e deve il suo nome solo all'assenza di riflessione. Lo spettro (intensità della radiazione emessa ad ogni lunghezza d'onda) di un corpo nero è caratteristico, e dipende unicamente dalla sua temperatura.

http://it.wikipedia.org/wiki/Corpo_nero 27 agosto 2006

52 Richard L. Gregory, *Occhio e cervello – La psicologia del vedere*, Op. cit., p. 32

La luce è solo una delle onde elettromagnetiche esistenti. Più precisamente: “Le lunghezze d'onda della luce sono comprese fra 0,38 micrometri⁵³ (3800 Å⁵⁴, luce blu) e 0,76 micrometri (7600 Å, luce rossa). Per un musicista, il rapporto (1:2) di queste lunghezze d'onda limite (o frequenze) è di un'ottava. L'occhio umano percepisce quindi un'ottava delle oscillazioni elettromagnetiche. La radiazione di 0,5 micrometri corrisponde al colore verde. Una serie di duemila onde di luce verde è lunga un millimetro. Onde più corte di 3800 Å e più lunghe di 900 Å sono radiazioni ultraviolette (UV), mentre le onde più lunghe di 7600 Å (fino a 0,1 mm) sono radiazioni infrarosse (IR)⁵⁵. In base alla lunghezza d'onda⁵⁶ possiamo quindi distinguere sia il colore⁵⁷ sia il tipo di radiazioni: raggi ultravioletti, raggi X, raggi gamma, luce visibile, raggi infrarossi, microonde, radioonde⁵⁸. Possiamo anche renderci conto di quanto sia limitata la percezione visiva rispetto alla totalità dello spettro.



Vi è sempre stato il dubbio se la luce viaggiasse a velocità finita o se raggiugesse

- 53 Il micron, abbreviazione del meno usato microméto, è un'unità di misura della lunghezza corrispondente a un milionesimo di metro.
http://it.wikipedia.org/wiki/Micrometro_%28unit%C3%A0_di_misura%29 28 novembre 2006
- 54 L'angstrom (Å), scritto anche ångström o angström, è un'unità di lunghezza non appartenente al SI e pari a 10⁻¹⁰ metri, 0.1 nanometri o 100 picometri. L'angstrom prende nome dal fisico svedese Anders Jonas Ångström, uno dei padri della spettroscopia.
<http://it.wikipedia.org/wiki/Angstrom> 28 novembre 2006
- 55 <http://www.pd.astro.it/othersites/venere/ESO/g1.htm> 27 agosto 2006
- 56 O alla frequenza: infatti sono tra loro in rapporto costante, secondo la formula $v \cdot \lambda = 3 \cdot 10^8$ m/s, dove v è la frequenza, λ la lunghezza d'onda e $3 \cdot 10^8$ m/s la velocità della luce.
- 57 Si ricordi il famoso esperimento di Newton che riuscì a scomporre e ricomporre lo spettro di un raggio di luce mediante prismi ottenendo sette colori: rosso, arancione, giallo, verde, azzurro, indaco e violetto.
- 58 Cfr. Roberto Marangoni, Marco Geddo, *Le Immagini digitali*, Op. cit., p. 92

istantaneamente qualsiasi regione dello spazio. Nel 1676 Ole Romer⁵⁹, in base ad alcune osservazioni sull'irregolarità delle eclissi dei satelliti di Giove, concluse che la luce viaggia alla velocità finita di circa 300.000 km/s nel vuoto assoluto (in base alla teoria della relatività di Einstein, è la più alta velocità raggiungibile da un corpo dotato di massa).

Questo significa che, se aggiungiamo al tempo impiegato dai fotoni per raggiungerci i nostri tempi di reazione, vediamo sempre al passato e per poter agire necessitiamo di funzionalità predittive.

La velocità diminuisce nel momento in cui la luce si propaga in un corpo trasparente⁶⁰ in relazione all'indice di rifrazione⁶¹ del corpo stesso (variabile in base alla lunghezza d'onda e alla densità del materiale), permettendo ai prismi di deviare la luce e alle lenti di formare le immagini⁶². Infatti: "Un oggetto non è visibile se non emette radiazione luminosa. Tranne il caso di oggetti che generano direttamente la luce⁶³, in generale il colore di una superficie dipende dalle caratteristiche della superficie stessa e dalla intensità e direzione della luce che la colpisce."⁶⁴ Questo dipende dal fatto che i fotoni viaggiano con traiettorie perfettamente lineari, ma solo finché colpiscono la superficie di un oggetto. In quasi ogni caso la superficie produce un cambiamento radicale nel comportamento del fotone che la colpisce. Sono questi cambiamenti, indotti dalla superficie nel comportamento dei fotoni, che forniscono alla visione le informazioni riguardanti le stesse superfici presenti nell'ambiente. Le sole superfici che non cambiano il comportamento dei fotoni sono quelle completamente trasparenti, quindi letteralmente invisibili (se esistono). Tutte le superfici interagiscono con la luce in modo sufficiente,

59 http://it.wikipedia.org/wiki/Olaus_Roemer 27 agosto 2006

60 Alcuni corpi si lasciano attraversare dalla luce: sono cioè *trasparenti*, come il vetro, certe materie plastiche, il cristallo, l'acqua stessa, purché pulita e non troppo profonda; altri, come il vetro smerigliato, la stoffa, la carta, non permettono di distinguere con precisione attraverso di essi gli oggetti, pur lasciando passare la luce: si tratta allora di corpi *traslucidi*. Molti altri corpi, infine, non fanno assolutamente passare la radiazione luminosa: un pezzo di legno o di metallo, un sasso, un blocco di carbone sono corpi *opachi*.

61 <http://it.wikipedia.org/wiki/Rifrazione> 27 agosto 2006

http://it.wikipedia.org/wiki/Indice_di_rifrazione 27 agosto 2006

62 Richard L. Gregory, *Occhio e cervello – La psicologia del vedere*, Op. cit., p. 25

63 Sorgenti luminose come: il sole, una lampadina, un fuoco, ecc

64 Cfr. Roberto Marangoni, Marco Geddo, *Le Immagini digitali*, Op. cit., p. 122

nella maggior parte delle condizioni ambientali, a essere percepite da un attento osservatore.

Il fotone che incide su una superficie può passarvi attraverso (materiali trasparenti), essere assorbito (corpi oscuri) o essere riflesso (materiali opachi). La luce può passare direttamente attraverso la superficie o essere curvata (rifratta). Di queste interazioni tra fotoni e superfici, la riflessione è la più importante per la visione. Infatti, la luce riflessa è stata cambiata dalla sua interazione con la superficie e contiene, quindi, delle informazioni sulla stessa. Inoltre, la luce riflessa è in seguito disponibile per incidere/colpire la superficie ricettiva dell'occhio di un osservatore, che la può trasmettere così al sistema visivo. Anche la luce non riflessa è importante in quanto permette di comprendere alcuni aspetti della percezione visiva: la luce assorbita per la visione dei colori, la luce passante/trasmessa per la percezione della trasparenza.

Riflettendo un fotone, una superficie ne altera la traiettoria: il fotone rimbalza sulla superficie in una direzione che dipende sia dalla sua provenienza sia dalla struttura microscopica della superficie. Questo significa che le superfici agiscono come fonti di luce secondaria, ossia che la luce non arriva solo direttamente, come nel caso di illuminazione solare senza nuvole o di una lampadina, ma anche in modo riflesso da ogni possibile superficie presente nell'ambiente⁶⁵.

I fenomeni di riflessione, dispersione e diffrazione della luce sono alla base della nostra percezione di distanze e colori. L'uomo non si limita ad acquisire passivamente i segnali luminosi, ma opera sugli stessi successive valutazioni ed elaborazioni: vedere è un processo attivo.

Prima di analizzare, per quanto possibile, questo processo, propongo alcune riflessioni sulle caratteristiche della luce rispetto ad altri segnali quali il suono, gli odori ecc., nonché sulle interdipendenze tra i vari sensi. La luce si distingue dagli altri segnali per la velocità e la capacità di variazione. La prima ci dà informazioni su oggetti,

⁶⁵ Cfr. Stephen E. Palmer, *Vision Science – photons to phenomenology*, USA, MIT Press, 1999 pp. 16-17

luoghi/ambienti o avvenimenti lontani da noi, informazioni utili sia per decidere in che direzione muoversi sia per “prepararsi”, decidere “cosa fare”⁶⁶. La seconda, ci informa sia sulla probabile distanza⁶⁷ sia, attraverso i colori, sulle probabili qualità degli oggetti stessi (es.: la mela rossa, il cielo scuro, ecc).

Le onde sonore non sono altrettanto veloci⁶⁸ (es.: vediamo prima il lampo e poi sentiamo il tuono) e non hanno la stessa propagazione. Il segnale sonoro è però facilmente riproducibile (è sufficiente far vibrare una corda, mentre la luce richiede una reazione elettrochimica-magnetica) e semplice da controllare. Pensiamo al linguaggio verbale, al sonar o ai pipistrelli che grazie al fenomeno dell'eco possono localizzare gli oggetti/ostacoli presenti nell'ambiente (eco-localizzazione). Al limite c'è anche una differenza d'approccio: con la luce noi ci limitiamo a recepire delle onde/particelle che comunque andrebbero a colpire l'oggetto in quanto emesse da una sorgente per noi generalmente incontrollabile (es.: il solare), comunque di grandezza infinitamente piccola, quindi non modifichiamo l'oggetto (l'eccezione potrebbe essere rappresentata da organismi fotosensibili, ma anche i sali d'argento usati in fotografia), insomma non siamo invasivi e potremmo essere anonimi. Poi, il suono comporta il movimento di materia, di atomi se non di molecole (es.: aria, acqua), quindi anche delle sollecitazioni fisiche sugli oggetti (ad esempio mi è capitato di sentir vibrare i vetri delle finestre in occasione di lancio di fuochi d'artificio per feste paesane) perciò se utilizziamo il suono siamo più invasivi, la fonte del suono è quasi sempre identificabile, se siamo noi, probabilmente non siamo anonimi.

Anche l'olfatto ci permette di avere informazioni a distanza, limitate però in portata e precisione (metri, decine di metri, chilometri?), in quanto l'odore non si propaga in modo uniforme. L'olfatto è d'altro canto molto più attendibile per quanto riguarda la composizione chimica delle sostanze (es.: gpl, metano, carne avariata, pane, biscotti appena sfornati, ecc).

66 Ibidem p. 6

67 Cfr. Lamberto Maffei, Adriana Fiorentini, *Arte e Cervello*, Bologna, Zanichelli, 1995 pp. 75-76

68 Nell'aria, la velocità del suono è di 331,5 m/s a 0°C e di 343 m/s a 20°C cfr.

http://it.wikipedia.org/wiki/Velocit%C3%A0_del_suono 30 agosto 2006

Con il tatto acquisiamo certezze: possiamo stabilire ciò che è solido, ciò è ruvido, ecc. Ma se abbiamo certezze significa anche che siamo coinvolti da ciò che ci circonda, nel bene, se questo ci fornisce delle possibilità (nutrirci, ecc), e nel male, (come toccare un metallo incandescente).

Infine il gusto, che come l'odorato ci dà informazioni chimiche sulla sostanza che abbiamo già assunto al nostro interno, dandoci un'ultima possibilità di espellerla prima che provochi danni.

Penso sia chiaro come la nostra sopravvivenza sia strettamente legata alla conoscenza del mondo/ambiente/universo che ci circonda, quindi alla nostra capacità di reperire informazioni il più possibile affidabili, ossia avere una percezione veridica. Interessante quanto afferma Stephen Palmer: “Questo è quasi sempre il caso della visione, ed è per questo, probabilmente, che la prendiamo quasi sempre per sicura. Sembra quasi una finestra perfettamente trasparente sulla realtà. Ma è veramente così? [...] Io argomenterò che la visione non è una finestra trasparente sulla realtà, ma una costruzione attiva, un significativo modello dell'ambiente che permette al percettore di predire che cosa succederà nel futuro così da poter effettuare azioni appropriate, aumentando le possibilità di sopravvivere.”⁶⁹.

69 Stephen E. Palmer, *Vision Science – photons to phenomenology*, Op. cit., pp. 6-7

L'apparato ottico nell'uomo e negli altri esseri viventi

Se la luce è il mezzo, il canale, per la diffusione dell'informazione visiva, è necessario disporre anche di un apparato recettore in grado di captarne i segnali. È utile a tal fine un'analisi che confronti i vari sistemi sia nel corso dell'evoluzione, a livello diacronico, sia al momento attuale nei vari organismi, a livello sincronico.

Come afferma Richard L. Gregory è necessario, in aggiunta, considerare i rapporti tra l'apparato recettore e quello che si occuperà di elaborare l'informazione: “A cosa potrebbe mai servire un cristallino fatto a metà, o anche un cristallino in grado di focalizzare un'immagine, se non vi fosse poi un cervello capace di servirsene? E come potrebbe d'altronde essersi formato un cervello con tali capacità prima che vi fosse l'occhio in grado di fornirgli informazioni visive? In ambito evolutivo non può esservi alcuna programmazione, alcun tipo di previsione, alcun controllo sperimentale per eliminare quanto può essere inefficace. Occhio e cervello si sono evoluti attraverso un processo lento e casuale di tentativi ed errori. Per ripercorrere le tappe dello stesso dobbiamo cercare di rintracciare i possibili vantaggi incrementali di ciascuno stadio, seppure ammettendo che qualcosa di vantaggioso per un verso possa poi rivelarsi utile in un campo totalmente differente.”⁷⁰.

Non solo alcuni minerali sono sensibili alla luce: anche a livello biochimico, quindi nei vegetali e negli animali, avvengono delle reazioni basate sulla luce⁷¹. Ad esempio è facile osservare come alcune piante dotate di mobilità attiva seguano lo spostamento della luce.

L'analisi della probabile evoluzione ce la fornisce ancora Gregory con alcuni esempi: “La reazione alla luce si verifica perfino in esseri unicellulari, e nelle forme animali più evolute troviamo cellule specializzate che funzionano come fotorecettori sensibili al

70 Richard L. Gregory, *Occhio e cervello – La psicologia del vedere*, Op. cit., pp. 35-36

71 Un approfondimento delle relazione fra regni minerale, vegetale e animale, nonché di chimica organica si può trovare in:

<http://www.accademiaxl.it/biblioteca/virtuale/percorsi/testi/percorsi.asp?page=17> 17 settembre 2006

movimento. Queste cellule possono essere sparse su tutta la pelle (come nel lombrico) oppure raggruppate, spesso lungo una piega cutanea o in una depressione della superficie corporea, secondo un adattamento che prelude al vero e proprio occhio capace di formare le immagini.

Sembra probabile che gli elementi fotorecettori si siano collocati in questi recessi per trovare protezione dalla luce abbagliante, che riduceva la loro capacità di individuare ombre in movimento, segnali dell'avvicinarsi di un pericolo. In queste loro sedi primitive i recettori erano sottoposti al rischio di venire occlusi da polveri e piccoli corpi estranei, che sovrapponendosi a essi potevano precludere la visione della luce. Per ovviare a tale inconveniente si formò allora, per effetto di mutamenti fortuiti, una membrana di protezione; la quale, in seguito, divenne più spessa al centro e si trasformò in una vera e propria lente (il cristallino). In un primo tempo il cristallino serviva soltanto per potenziare la luminosità, ma in seguito venne a formare immagini effettive. Un esempio di occhio primitivo si può ancora osservare in un mollusco marino, la patella⁷²; e un altro mollusco, il *Nautilus*⁷³, è fornito di un occhio ancor più primitivo, privo di cristallino ma provvisto di un piccolo foro che serve a formare le immagini. L'interno dell'occhio del *Nautilus* è lavato dal mare in cui esso vive, mentre gli occhi più evoluti, simili al nostro, sono pieni di un liquido appositamente prodotto (l'umor acqueo) che sostituisce l'acqua di mare; e le lacrime umane, con il loro sapore salato, sono una sorta di riproduzione degli oceani primordiali che bagnavano i primi occhi.⁷⁴

Gregory prosegue analizzando il funzionamento dell'occhio degli invertebrati⁷⁵ che ha

72 Mollusco marino commestibile, frequente lungo i litorali del Mediterraneo, dove vive attaccato a rocce in parte sommerse; ha una conchiglia di colore verdastro e di forma conica alquanto appiattita.

http://www.sapere.it/gr/DictionarySearchServlet?DS_action=ItalySearch&DS_resType=14&DS_use_rInput=patella 13 settembre 2006

73 Considerato estinto in seguito ai ritrovamenti fossili risalenti al Paleozoico, il *Nautilus*, un mollusco cefalopode (con tentacoli inseriti nella parte anteriore della testa che servono da piedi) tetrabranchiato (possiede cioè due paia di branchie, anziché un solo paio come gli altri cefalopodi), è stato osservato per la prima volta in vita solamente nel 1829, pertanto è classificato come fossile vivente.

[http://it.wikipedia.org/wiki/Nautilus_\(mollusco\)](http://it.wikipedia.org/wiki/Nautilus_(mollusco)) 17 settembre 2006

74 Richard L. Gregory, *Occhio e cervello – La psicologia del vedere*, Op. cit., p. 37

75 È il gruppo di animali più vasto: comprende tutti gli organismi che non hanno uno scheletro interno. Ne fanno parte sia quegli esseri primitivi formati da poche decine di cellule, sia gli insetti più evoluti

una struttura più complessa dell'occhio umano poiché è formato da molte lenti con un piccolo gruppo di recettori, circa uno per ogni singolo cristallino. Esso ha delle prestazioni migliori di quello dell'uomo specie nella rilevazione del movimento: a riprova di ciò si pensi alla libellula che cattura le prede in volo, o alla capacità di recepire i raggi ultravioletti, come nel caso delle api. Interessanti, negli occhi composti, sono i meccanismi che permettono l'adattamento alla luce e alla penombra, o il funzionamento della lente cilindrica, che per le sue caratteristiche richiama i moderni strumenti utilizzati in endoscopia medica o le fibre ottiche⁷⁶.

Gregory, passando attraverso l'analisi del sistema visivo della *Copilia* (un microorganismo acquatico⁷⁷), arriva ad ipotizzare “che l'elaborazione visiva sia derivata dall'elaborazione della sensazione tattile, caratterizzata da una utilità immediata ed essenzialmente più semplice. Può essere suggestivo pensare che vi siano due tipi di tatto: una ricezione tattile attiva ottenuta muovendo le ciglia, o le dita, e una ricezione passiva legata alla percezione delle forme. In effetti, si può scorrere lentamente con le dita sulla superficie degli oggetti, oppure sperimentare direttamente la forma di piccoli oggetti, nella loro unitarietà, portandoli a contatto con un'area estesa della pelle.”⁷⁸. Così facendo egli supera l'*impasse* dovuta alle possibili differenze evolutive tra occhio e cervello.

Non solo, l'originale meccanismo visivo a scansione utilizzato dalla *Copilia* potrebbe fungere da esempio per strumenti dediti a particolari applicazioni, e permettere di introdurre una riflessione sull'evoluzione epistemologica della scienza. Infatti, quando nel 1891, Selig Exner osservò che il sistema ricettivo della *Copilia* e la lente cilindrica ad esso unita compivano "dei continui movimenti attivi", egli non riuscì a darne una spiegazione. Fu con l'avvento della televisione che si ebbero le nozioni necessarie alla comprensione. La scienza, e noi con essa, ha quindi bisogno di “modelli di riferimento”

(come le formiche e le api). Ma la stragrande maggioranza degli invertebrati è composta da animali vermiformi, di ogni tipo e dimensione, che hanno colonizzato tutti gli ecosistemi terrestri. Cfr: <http://www.sapere.it/tca/minisite/scienza/tuttozoologia/id100003.html> 17 settembre 2006

76 Cfr Richard L. Gregory, *Occhio e cervello – La psicologia del vedere*, Op. cit., p. 42

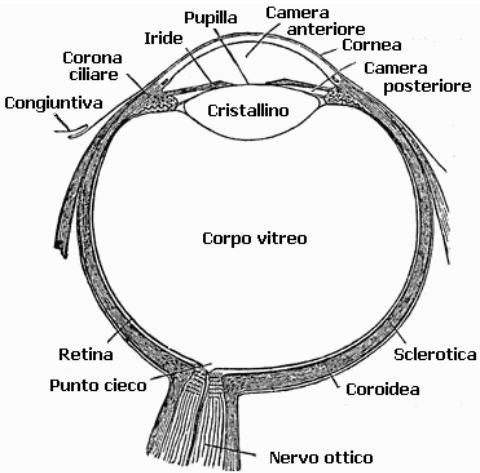
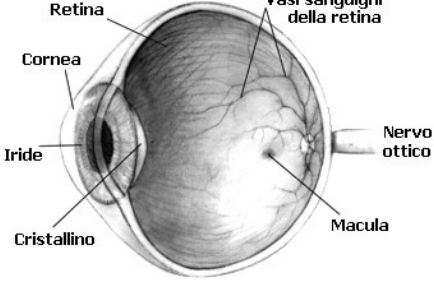
77 <http://www.richardgregory.org/papers/copilia/curious-eye-copilia.pdf#search=%22copilia%22> 20 settembre 2006

78 Cfr Richard L. Gregory, *Occhio e cervello – La psicologia del vedere*, Op. cit., pp 43, 46,47

per comprendere, avere una conoscenza più completa o avvicinarsi alla verità⁷⁹.

L'occhio umano⁸⁰

La brevissima panoramica sulla possibile evoluzione dei sistemi ottici e l'ipotesi che la vista discenda dal tatto dovrebbero trasmettere sia l'idea della complessità dell'argomento, sia il fatto che la natura ha risolto con “dispositivi” diversi il problema della percezione visiva. Nel riprodurla dovremo quindi valutare qual è l'obbiettivo e quali le possibili criticità. Con questo fine, cerchiamo di comprendere le peculiarità dell'occhio umano, quale modello di riferimento, analizzandone le funzionalità e la struttura.

	
<p>Figura 2 Sezione orizzontale obliqua dell'occhio http://it.wikipedia.org/wiki/Immagine:Occhio_umano_spaccato.png 13 settembre 2006</p>	<p>Figura 3 http://it.wikipedia.org/wiki/Immagine:Occhio_umano.png 13 settembre 2006</p>

79 Il riferimento è alla contrapposizione tra Karl Popper e Thomas S. Kuhn, se il primo sostiene che la confutabilità di una teoria consente la proposizione di successive ipotesi che, pur destinate ad essere falsificate, consentono un progresso verso la verità, il secondo afferma che nella storia c'è progresso non perché ci si avvicina a qualche meta assoluta (la verità), ma perché ci si allontana sempre più da stadi primitivi di ricerca. In altri termini nella scienza non c'è progresso “verso” qualcosa, ma “a partire da” qualcosa. Cfr Paolo Frignani, Paola Rizzati, *Didattica della comunicazione*, San Bartolomeo in Bosco (Fe), 2003 p. 293

80 Si veda, come possibile approfondimento, gli appunti di neurofisiologia della visione con cenni di anatomia e embriologia realizzati da P.P. Battaglini per gli scopi divulgativi del Centro Interdipartimentale per le Neuroscienze B.R.A.I.N. Dell'Università di Trieste all'indirizzo <http://fc.units.it/ppb/visione/> 12 ottobre 2006

Funzione e Struttura

Il sistema ottico può essere considerato come un meccanismo che converte/trasforma gli stimoli/energia luminosa che si imprime nella retina in un flusso in uscita simile a un segnale elettrico. Come si può osservare dalla fig. 2, la luce entra attraverso la cornea, passa in sequenza la camera anteriore, la pupilla (apertura nell'iride), il cristallino, e il corpo vitreo (o humor vitreo), prima di imprimersi sullo strato di fotorecettori che costituiscono il retro della retina. Questa è responsabile della reale trasformazione da energia luminosa in segnali nervosi nella forma di una sequenza d'impulsi a frequenza modulata⁸¹.

Nonostante la sua forma globulare avente un diametro di circa 25 mm, l'occhio umano è funzionalmente simile a una telecamera con un campo visivo che copre un'area di circa 160° in ampiezza per 135° d'altezza. Come ogni altro sistema ottico, soffre di vari tipi di distorsioni geometriche e cromatiche⁸². Si tenga presente infatti che solo il 50 per cento della luce che attraversa la cornea arriva alla retina. Il sistema ottico quindi proietta un'immagine riconoscibile ma molto imperfetta sulle cellule recettrici⁸³, tuttavia in quantità e qualità sufficiente a fornire al sistema visivo informazioni eccellenti riguardanti l'ambiente⁸⁴.

Posizionamento e visione stereoscopica

Oltre alla struttura interna è opportuno valutare il posizionamento degli occhi e il loro numero. Essi sono collocati sulla linea immaginaria che divide orizzontalmente a metà la testa, all'interno di due fori quasi emisferici del cranio, denominati orbite oculari, che li tengono saldamente in posizione e permettono, attraverso la rotazione, di dirigerli come si desidera o necessita. Ogni occhio si muove grazie a sei piccoli ma potenti muscoli denominati extraoculari, controllati da specifiche aree del cervello. Il loro

81 Cfr. Martin D. Levine, *Vision in Man and Machine*, Op. cit., pp. 60-61

82 Cfr David A. Forsyth, Jean Ponce, *Computer Vision: A Modern Approach*, Prentice Hall, 2002 pp. 13-15

83 Cfr. Martin D. Levine, *Vision in Man and Machine*, Op. cit., p. 62

84 Cfr. Stephen E. Palmer, *Vision Science – photons to phenomenology*, Op. cit., p. 24

numero è in ragione del loro funzionamento: sono organizzati in tre coppie, e i muscoli di ogni coppia sono antagonisti, così da determinare movimenti in uno dei tre piani ortogonali (perpendicolari) dello spazio. I movimenti sono necessari per sondare diverse regioni del campo visivo senza dover muovere l'intero capo e per la messa a fuoco di oggetti posti a distanze diverse. Per i due occhi, il compito di seguire un oggetto deve essere svolto con una precisione dell'ordine dei minuti di arco⁸⁵, altrimenti vedremmo doppio⁸⁶. Palpebre e ciglia proteggono l'occhio, le lacrime lo tengono umido e pulito.

Alcuni di questi aspetti sono veri anche per occhi non umani. Tuttavia alcune specie, come i piccioni e i gufi, non possono muovere gli occhi nelle orbite e sono costretti a spostare l'intero capo. Questo limita la velocità con cui essi percepiscono gli eventi nell'ambiente, semplicemente per il fatto che la testa è più pesante e difficile da spostare degli occhi. In aggiunta la posizione differisce in modo notevole nelle varie specie. Nell'uomo gli occhi sono entrambi posizionati nella parte frontale del capo, mentre in molti animali si trovano più vicini ai lati. Il posizionamento frontale fornisce un campo visivo con un'ampia area di sovrapposizione dei due occhi, limitandone però la copertura totale. Il beneficio della sovrapposizione del campo visivo (detta anche visione binoculare o stereovisione) consiste nel permettere la percezione della profondità. La capacità di stimare la distanza in modo attendibile è un vantaggio evolutivo per i predatori, che hanno bisogno dell'informazione per attaccare: è per questo che gli occhi dei cacciatori tendono a essere posizionati frontalmente con sovrapposizione dei campi. All'opposto, campi visivi panoramici sono vantaggiosi per le possibili prede per monitorare in modo più esteso possibile l'ambiente che le circonda al fine di individuare possibili pericoli, quindi gli occhi sono posizionati lateralmente. Altri animali hanno gli occhi posizionati frontalmente o lateralmente per ragioni diverse, ma tutte riflettono lo stesso bilanciamento tra un'accurata percezione della profondità e copertura dell'ambiente⁸⁷.

85 Il minuto d'arco è una misura degli angoli e corrisponde a 1/60 di grado. Viene così denominato per non essere confuso con le misure del tempo. http://it.wikipedia.org/wiki/Primo_%28geometria%29
22 febbraio 2007

86 David H. Hubel, *Occhio, Cervello e Visione*, Op. cit., p. 43

87 Cfr. Stephen E. Palmer, *Vision Science – photons to phenomenology*, Op. cit., p. 26

Formazione dell'immagini

A questo punto, date alcune informazioni sulla composizione della luce e sulla struttura dell'occhio, è opportuno chiedersi, prima di vedere più in dettaglio alcuni componenti dello stesso e quanto (relativamente poco) si conosce del cervello, come si formano le immagini ottiche.

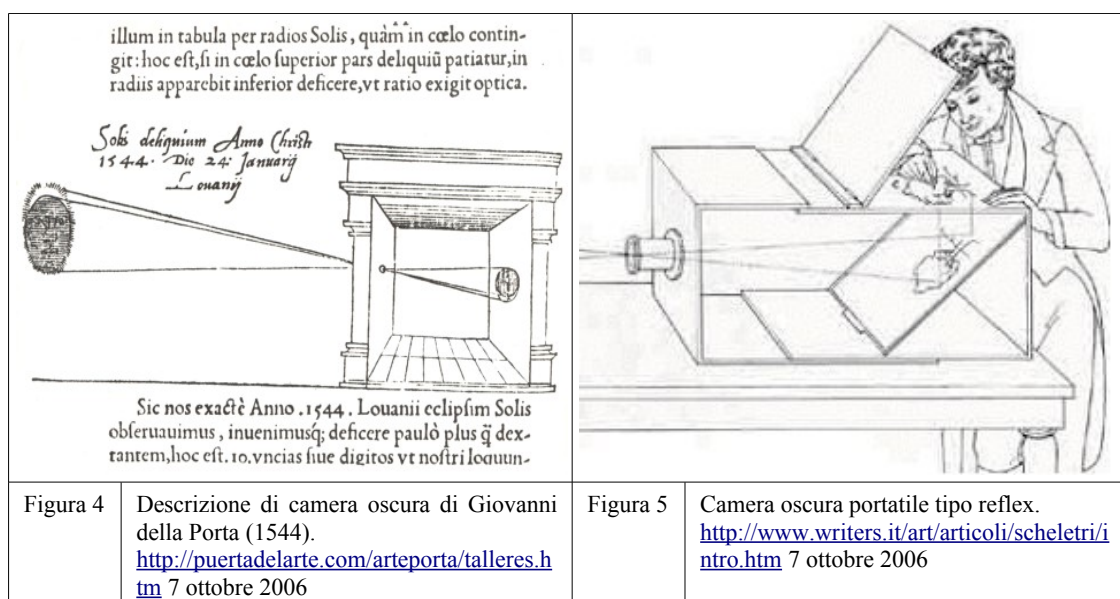
Platone (427-347 a.C.), filosofo greco, credeva che un “fuoco interno” facesse nascere dei raggi che partivano dagli occhi verso gli oggetti percepiti. Questa teoria, denominata emanazione o emissiva, venne ripresa da Euclide (365-275 a.C.), probabilmente discepolo di Platone, e inserita nella sua opera “*Ottica*”. Epicuro (341-270 a.C.) la rigettava credendo, invece, che piccole repliche degli oggetti fossero, in qualche modo, rapidamente trasmesse nella mente attraverso gli occhi. Galeno (130-200 d.C.) elaborò l'idea con dettagli fisiologici, proponendo che dopo l'emanazione dei raggi questi interagissero con gli oggetti e ritornassero agli occhi. Egli credeva che nel cristallino i raggi, interagissero poi con uno “spirito visivo” che va e ritorna dal cervello agli occhi portando con sé le repliche degli oggetti percepiti⁸⁸. Fu Alhazen (965-1040 d.C.), filosofo arabo, che intuì come l'occhio funzionasse in modo simile a una camera con un foro in una parete⁸⁹ che permetteva la formazione, nella parete opposta a quella forata, di immagini interne, anche se capovolte, delle figure di oggetti esterni alla camera. Egli progettò così la camera oscura che fu successivamente perfezionata da Giovan Battista Della Porta (1535-1615 d.C.), che inserì nel foro una lente focalizzante, per ottenere immagini più brillanti e riuscire a individuare forme e colori⁹⁰. Fu infine Johannes Kepler (Keplero 1571-1630 d.C.) che formulò la moderna teoria fisiologica dell'ottica con un'approssimazione accettabile. Teoria successivamente sperimentata da padre Cristoph Scheiner (1573-1650 d.C.) nel 1625 e adottata da René Descartes (Cartesio 1595-1650 d.C.)⁹¹.

88 Ibidem p.24

89 Interessante la denominazione inglese “pinhole camera”, che tradotta letteralmente sarebbe camera con foro di spillo.

90 Cfr Richard L. Gregory, *Occhio e cervello – La psicologia del vedere*, Op. cit., p. 48

91 Cfr. Lamberto Maffei, Adriana Fiorentini, *Arte e Cervello*, Op. cit., p. 3



La formazione delle immagini ottiche è quindi la trasformazione di un mondo che ha tre dimensioni spaziali in rappresentazioni bidimensionali. Gli oggetti sono immersi in spazi tridimensionali illuminati dalla luce, che viene riflessa dalle loro superfici nell'occhio dell'osservatore lungo linee rette. I fotoni entrano nell'occhio formando un'immagine bidimensionale capovolta sul fondo. L'oggetto esterno è spesso riferito come stimolo distale (significa lontano dall'osservatore), e la sua proiezione sul retro della retina come stimolo prossimo (vicino all'osservatore). La grandezza dell'immagine dell'oggetto nell'occhio è solitamente definita dal suo angolo visivo: il numero di gradi sottesi dall'immagine dai suoi estremi fino al punto focale dell'occhio. Quest'angolo misura la dimensione spaziale dello stimolo prossimo e non quello distale. Lo stesso oggetto esterno sottenderà un angolo inferiore quando è lontano dall'osservatore e uno maggiore quando gli sarà vicino. La relazione tra dimensione e distanza dell'oggetto e dimensione dell'immagine è importante per comprendere come percepiamo la grandezza e il posizionamento degli oggetti⁹².

Forse il fatto più importante, per quanto riguarda il processo di formazione delle immagini e la comprensione del processo visivo, è che l'immagine sul fondo dell'occhio ha solo due dimensioni spaziali al posto di tre, cioè è simile alla proiezione su una superficie curva. Questo significa che vitali informazioni riguardanti lo spazio si

92 Cfr. Stephen E. Palmer, *Vision Science – photons to phenomenology*, Op. cit., p. 20

perdono passando dal mondo reale tridimensionale (in seguito 3-D) alla sua rappresentazione oculare bidimensionale (in seguito 2-D). La grandezza persa è la profondità: la distanza che va dal punto focale d'osservazione alla superficie ambientale che riflette la luce. Per percepire il mondo in 3-D la profondità deve essere in qualche modo recuperata dalle informazioni fornite dall'immagine ottica bidimensionale⁹³.

Il processo di trasformazione da 3-D a 2-D ubbidisce a determinate leggi fisiche e può quindi essere analizzato matematicamente. Nel caso specifico si usano le **proiezioni geometriche**: gli studi di come uno spazio a dimensioni superiori è mappato in uno a dimensioni inferiori. La proiezione geometrica può determinare esattamente, per una data scena 3-D, dove ogni punto di questa sarà proiettato in un piano 2-D creando la relativa immagine, e quali saranno le proprietà invariabili delle possibili immagini in varie proiezioni. Nella visione dinamica, la proiezione, della scena considerata, parte dalle quattro dimensioni strutturali dello spazio-tempo, passa per lo spazio tridimensionale del flusso ottico che si stende nel tempo e arriva alle superfici bidimensionali nel fondo dell'occhio.

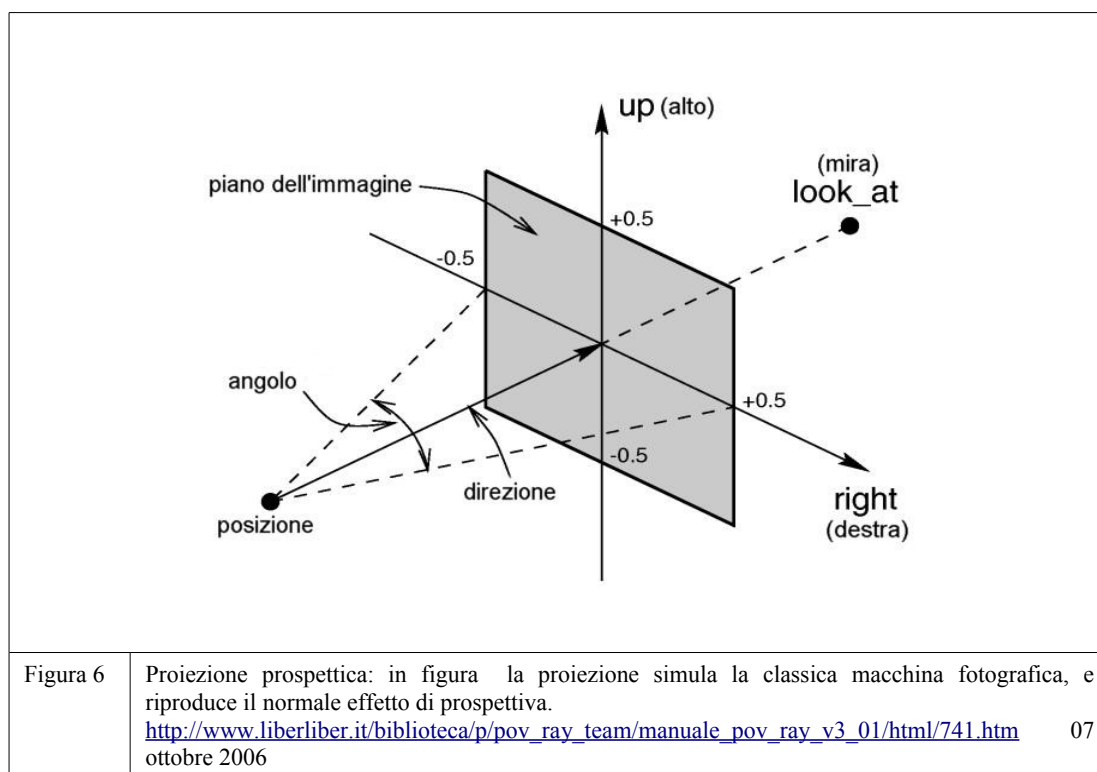
La proiezione geometrica sembra perciò essere lo strumento matematico ideale per comprendere la formazione delle immagini. Il problema è che da sola non può riprodurre la complessità dei fenomeni ottici in quanto non contiene le strutture appropriate per la modellazione della riflessione della luce, dell'assorbimento o della rifrazione. In un mondo saturo di superfici opache, ad esempio, solo la luce riflessa da superfici vicine raggiunge l'occhio. I fotoni provenienti da punti lontani saranno o assorbiti o riflessi dalle superfici più vicine, senza arrivare, perciò, agli occhi.

Per superare questi problemi si ritorna al modello della camera oscura, quindi alle osservazioni fatte da Alhazen. Poiché i fotoni viaggiano in linee rette, la luce che cade su ogni punto del piano dell'immagine della camera oscura vi giunge riflessa (o emessa) da un punto specifico dell'ambiente. Questo punto sta sul raggio che parte dal punto sul piano dell'immagine esterno e che attraversa il foro della camera. Il tutto è facilmente

93 Ibidem

comprensibile nelle figure 4 e 5, che sono prime rudimentali riproduzioni di fenomeni osservabili, ma soprattutto in figura 6, che rappresenta lo schema di funzionamento di una macchina fotografica e/o dell'occhio⁹⁴. Il punto “posizione” è all'interno della camera oscura, quindi è il piano su cui cade/incide il raggio di luce, il punto “look_at” o “mira” è un punto nello spazio ambientale, il “piano dell'immagine” è la parete con il foro o l'eventuale lente. Per questi punti passa la linea retta percorsa dai fotoni.

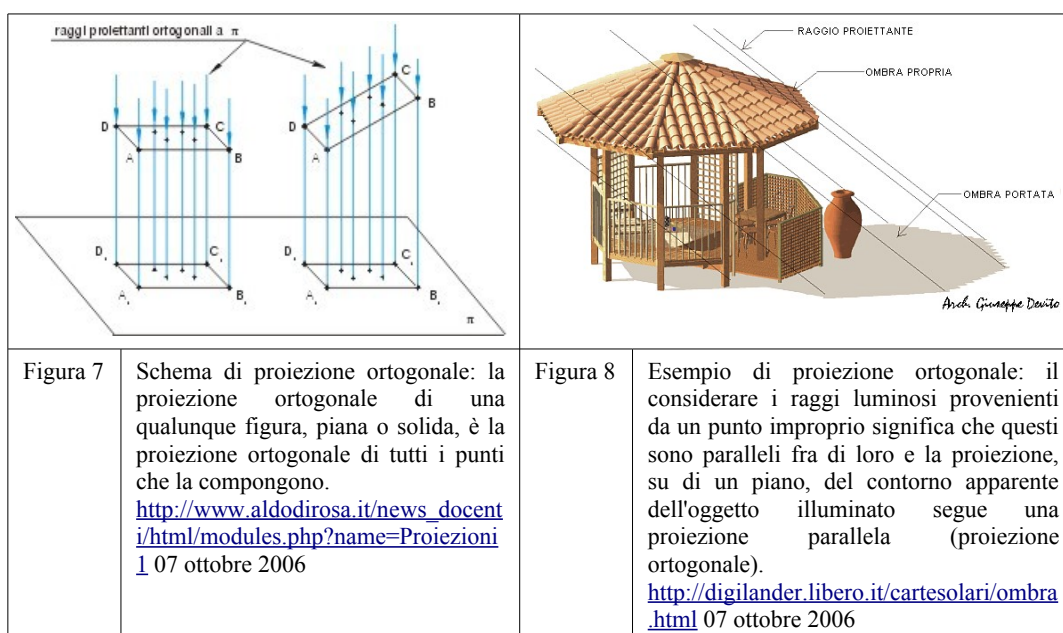
Questa situazione dà la proiezione prospettica (o polare) della geometria di base: il processo di formazione dell'immagine dove la luce converge in un unico punto focale (o polo).



Buone foto realizzate con la camera oscura non sono così semplici da ottenere come la presente descrizione può far sembrare. Per avere un'immagine chiara e vivace, è necessario che il foro sia molto piccolo; circa 0,4 mm di diametro sono l'ideale. A causa di questa ridotta apertura, solo pochissima luce incide sulla parete interna. Se il foro viene allargato, facendo passare più luce, l'immagine diventa confusa poiché tutta la

94 Ibidem p. 21

luce non passa più per un singolo punto ma attraverso più punti. Questo problema viene superato ponendo una lente convessa trasparente (un obiettivo) sul foro, così da curvare i raggi di luce entrate in un punto (detto punto focale) dietro alla stessa lente ma leggermente distanziato. Perciò l'obiettivo fornisce un “foro virtuale” con un suo punto focale che rende l'immagine proiettata sul fondo nuovamente definita e chiara, con il vantaggio, rispetto alla camera oscura di partenza, di essere più brillante in quanto entra più luce. Nell'occhio umano sia cornea che cristallino cooperano per la messa a fuoco.



Come si può ben intuire le elaborazioni matematiche relative alla proiezione prospettica sono complesse. Al fine di semplificarle spesso usata una “prospettiva debole” denominata proiezione ortografica o parallela per fornire un modello geometrico della formazione dell'immagine. In questo caso si considera l'immagine come formata da raggi di luce che viaggiano paralleli gli uni agli altri e perpendicolari al piano dell'immagine, piuttosto che raggi che convergono sull'obiettivo o sul foro della camera oscura.

La proiezione ortogonale permette di semplificare l'elaborazione matematica in quanto ignora la dimensione di profondità, le distanze dall'immagine agli oggetti, del mondo, mentre tutte le informazioni spaziali sul piano perpendicolare alla direzione visiva sono preservate senza cambiamenti. Questo significa che, quando la distanza che va

dall'immagine all'oggetto è sufficientemente grande, rispetto alla profondità dello stesso, la proiezione ortografica è una buona approssimazione della proiezione prospettica. Da vicino, invece, le differenze sono significative. Quindi, la proiezione ortografica può essere concepita come un caso limite della proiezione prospettica, dove la distanza tra l'oggetto e il punto focale è infinita. La differenza è che nella seconda a una distanza infinita l'oggetto è un punto, nella prima è un'immagine estesa nello spazio⁹⁵.

La visione è quindi il problema inverso alla formazione delle immagini: come riottenere da immagini ottiche di una scena la conoscenza degli oggetti che la formano. Sebbene sembri facile invertire il processo, esso risulta alquanto difficile, se non impossibile. Se da uno spazio tridimensionale si può ricavare una sola immagine bidimensionale, non è vero il contrario, ossia, ogni punto dell'immagine può essere mappato in un infinito numero di punti nell'ambiente. Si sa che è possibile solo in quanto il sistema visivo umano vi riesce con accuratezza in molte circostanze. Sul come vi riesca, vi sono numerose teorie, la più nota delle quali afferma che tale sistema elabori numerose assunzioni altamente probabili circa la natura dell'ambiente e le condizioni in cui è visto. I vincoli posti da queste assunzioni permettono di risolvere il “problema inverso” nella maggior parte dei casi. La visione è per questo un processo euristico, dove si inferiscono le condizioni più probabili dell'ambiente che ha prodotto l'immagine. Il processo è euristico in quanto usa regole empiriche di inferenza – basate su assunzioni addizionali – che non sono sempre valide e quindi, qualche volta, possono portare a conclusioni errate. Poiché sono elaborate nel cervello le vedremo in seguito, ora esaminiamo nel dettaglio i componenti dell'occhio seguendo il percorso fatto dai fotoni.

La cornea e l'umore acqueo

La cornea (che è coperta da una pellicola di lacrime) è la superficie frontale dell'occhio che curva la luce per formare l'immagine. È trasparente ed è sorretta da uno strato opaco di membrana fibrosa denominata “sclera”⁹⁶. Essa si caratterizza per la mancanza di vasi sanguigni e per questo ricava il suo nutrimento dall'umore acqueo, questo la rende

⁹⁵ Ibidem p. 23

⁹⁶ Cfr. Martin D. Levine, *Vision in Man and Machine*, Op. cit., p. 67

praticamente indipendente dal resto dell'organismo (per questa ragione i trapianti di cornea sono sicuri: gli anticorpi non possono raggiungerla e distruggerla). L'umore acqueo è completamente rinnovato ogni quattro ore circa, in quanto viene continuamente secreto e assorbito; alcune volte, in questo liquido, sono presenti delle impurità che compaiono davanti agli occhi e, poiché proiettano la loro ombra sulla retina, vengono viste come macchie vaganti nello spazio⁹⁷. Il percorso aria-cornea-umore acqueo è responsabile di approssimativamente due-terzi della capacità ottica dell'occhio (42 diottrie contro le possibili da 57 a 62 totali).

La pupilla e l'iride

Dopo aver attraversato la cornea e l'umore acqueo la luce passa attraverso la pupilla, un diaframma o apertura al centro dell'iride, che è, appunto, un muscolo a forma anulare. Il pigmento⁹⁸ presente nell'iride assume una vasta gamma di colori. Non a caso, iride significa arcobaleno, in greco.

La funzione dell'iride è di contrarsi, restringendo il foro pupillare, sia quando la luce è intensa, per ridurre l'apertura della lente, sia quando gli occhi convergono, per vedere oggetti vicini. Ciò aumenta la profondità di campo che consente di mettere a fuoco distintamente, il processo è analogo a quello della riduzione del diaframma in una macchina fotografica⁹⁹. Le variazioni del diametro del diaframma sono ottenute con la contrazione di due fasci di muscoli lisci, lo sfintere e il dilatatore pupillare. Lo sfintere è responsabile della costrizione e ha un moto circolare parallelo all'iride; il dilatatore, con fibre radiali disposte come i raggi di una ruota, apre la pupilla. In cooperazione questi muscoli possono posizionarsi in 16 diverse aperture variando il diametro da 2mm a 8mm. Per massimizzare la messa a fuoco il meccanismo di controllo tende a tenere l'apertura più stretta possibile. L'operazione richiede circa 10s per l'intero diametro. L'attivazione del controllo è estremamente complessa e dovuta a numerosi fattori, ma

97 Cfr Richard L. Gregory, *Occhio e cervello – La psicologia del vedere*, Op. cit., pp. 50-51

98 Il pigmento è una sostanza colorata che assorbe una limitata fascia di lunghezza d'onda della luce incidente, riflettendo il resto. Il pigmento può essere organico o inorganico, biologico o prodotto chimicamente.

99 Cfr Richard L. Gregory, *Occhio e cervello – La psicologia del vedere*, Op. cit., p. 53

principalmente alla quantità di luce incidente e allo stato emotivo dell'osservatore. In verità, una recente indagine ha fornito una lista di 23 diverse cause di variazione, tra cui le preferenze sessuali, le attitudini politiche, la fatica, e la lunghezza d'onda di colori e suoni, nonché alcuni disordini neurologici.

Riassumendo, variare la dimensione della pupilla ha tre obiettivi. Il primo, come sopra esposto, consiste nel controllare la quantità di luce che s'imprime sulla retina. Il secondo, conosciuto come risposta alla vicinanza, tende a mettere a fuoco gli oggetti vicini, quindi adeguarsi alla profondità di campo. Il terzo, importante in condizioni di luce brillante, è la riduzione dell'apertura per evitare distorsioni dell'immagine. Sembra che solo i primi due siano quantitativamente controllabili da stimoli esterni¹⁰⁰.

Il cristallino

Oltre la pupilla, la luce passa attraverso il cristallino, che fornisce il terzo rimanente del potere diottrico. Il suo compito principale è di compiere gli aggiustamenti necessari per mettere a fuoco oggetti a distanze diverse. Nell'uomo la lente cristallina è formata da un insieme di strati sottili, tra loro sovrapposti, ed è fissata da una membrana, la zonula, che ne regola, anche, lo stato di tensione¹⁰¹. Il funzionamento è simile a quello della messa a fuoco di una macchina fotografica: si varia la distanza tra la lente e la pellicola; nell'uomo, invece, cambia la forma del cristallino, mettendo in tensione o rilasciando i tendini che lo fissano alla periferia, così può diventare più sferico per vedere gli oggetti vicini e più appiattito per vedere quelli lontani. Queste variazioni di forma sono determinate da un insieme di muscoli radiali chiamati ciliari. Il processo avviene in tempi rapidi, circa 0,4s (dopo i quarantacinque anni, il cristallino si irrigidisce e viene meno la capacità di messa a fuoco: per superare questo problema Benjamin Franklin inventò gli occhiali bifocali). Lo stimolo necessario a far contrarre i muscoli ciliari, e modificare la forma del cristallino, è dato dall'input visivo, questo, a sua volta, è collegato al riflesso che controlla la convergenza degli occhi¹⁰².

100 Cfr. Martin D. Levine, *Vision in Man and Machine*, Op. cit., p. 68

101 Cfr Richard L. Gregory, *Occhio e cervello – La psicologia del vedere*, Op. cit., pp. 54-55

102 Cfr David H. Hubel, *Occhio, Cervello e Visione*, Op. cit., p. 44

Umore vitreo

L'umore vitreo è un liquido gelatinoso responsabile del mantenimento della forma del globo oculare. La luce passa attraverso il liquido alla retina, che a sua volta si occupa della conversione elettro-ottica delle forme di luce che incidono sui fotorecettori della retina ¹⁰³.

La retina

Il termine retina etimologicamente discende da "rete", o "tunica a forma di ragnatela", ed è dovuto al fatto della presenza di un fitto intreccio di vasi sanguigni nella membrana retinica ¹⁰⁴.

Prima di esaminarne la struttura ricordo, come accennato nel paragrafo “Formazione delle immagini”, che fu Keplero, nel 1604, a rendersi conto che la retina è il primo stadio della percezione visiva: essa infatti costituisce lo schermo su cui vengono proiettate le immagini provenienti dalla lente cristallina. L'ipotesi è stata in seguito controllata sperimentalmente nel 1625 da Scheiner che, prima asportò la parte posteriore dell'occhio di un bue delle membrane più esterne (la sclerotica e la corioide), quindi, osservò una piccola immagine capovolta sulla retina lasciata allo scoperto. La retina appariva simile a una pellicola diafana (trasparente, esile). Successivamente, Cartesio ripeté l'esperimento, descrivendolo nella sua *Diottrica* del 1637, e vi aggiunse delle chiare schematizzazioni. È sorprendente che la scienza sia giunta a questa scoperta relativamente tardi nella sua storia, e ancor più strano è che non sia stato apprezzato fin da subito l'ampio spettro delle sue implicazioni. Oggi vi è ancora chi tenta di conservare l'idea secondo cui la percezione è direttamente in contatto con il mondo degli oggetti, ignorando la presenza dello schermo ottico della retina e la complessità fisiologica del processo di elaborazione dell'immagine necessario a creare le percezioni. Il sapersi isolati dal mondo esterno può, probabilmente, essere fonte di timore e, portare ad

103 Cfr. Martin D. Levine, *Vision in Man and Machine*, Op. cit., p. 68

104 <http://www.etimo.it/?term=retina&find=Cerca> 25 febbraio 2007

illusioni ottiche¹⁰⁵.

La retina è una parte del cervello, ma si è separata da questo nel corso dell'evoluzione; tuttavia vi è ancora un fascio di fibre che funge da connessione tra le parti: il nervo ottico. La retina ha la forma di una lamina, come altre strutture del sistema nervoso centrale. Il suo spessore è di circa un quarto di millimetro. È costituita da cinque strati: tre strati di neuroni, separati da due strati che contengono le sinapsi fatte dagli assoni e dai dendriti¹⁰⁶, come visibile in figura 10.

<p>Figura 9</p>	<p>Schema di trasmissione di segnali in neuroni e del neurone, di proprietà di P. Forster. http://it.wikipedia.org/wiki/Neurone 9 ottobre 2006.</p>	<p>Figura 10</p>	<p>Schema dei livelli neurali nella retina, tratto da Henry Gray (1825–1861). <i>Anatomy of the Human Body</i>. 1918. http://www.bartleby.com/107/illus882.html 10 ottobre 2006</p>

Prima di esaminarli è opportuno comprendere come sia fatto e come funzioni l'elemento base del cervello e della retina: il neurone. Esso è una cellula specializzata che, grazie ad una complessa serie di reazioni biochimiche interne, unisce l'attività di altri neuroni (come input) ad essa collegati e ne diffonde/propaga (in output) il risultato ad altri in

105 Cfr Richard L. Gregory, *Occhio e cervello – La psicologia del vedere*, Op. cit., pp. 73-74

106 Cfr David H. Hubel, *Occhio, Cervello e Visione*, Op. cit., p. 46

successione¹⁰⁷. Il neurone può essere suddiviso in cinque parti:

1. **Dendriti:** sono delle ramificazioni che partono dal corpo della cellula nervosa. Sono specializzati nel rispondere ai segnali provenienti da altri neuroni o dall'ambiente esterno. La loro forma ramificata, rendendo disponibile un'ampia superficie, li rende adatti alla ricezione dei segnali. I dendriti dei neuroni sensoriali sono in grado di reagire a stimoli ambientali specifici come la pressione, gli odori, la luce o il calore, e questo grazie ad alcuni speciali adattamenti della loro membrana. I dendriti presenti nei neuroni del cervello e del midollo spinale, reagiscono ai neurotrasmettitori chimici rilasciati da altri neuroni. La loro membrana è dotata di recettori proteici che si legano a neurotrasmettitori specifici e trasmettono, come risultato di questa unione, segnali elettrici¹⁰⁸.
2. **Il soma (corpo cellulare):** integra i segnali elettrici provenienti dai dendriti e assicura le funzioni vitali del neurone. I segnali confluiscono al corpo cellulare del neurone viaggiando lungo i dendriti. Questo, si comporta come un centro di integrazione, "interpreta" i segnali e "decide" se produrre un potenziale d'azione: il segnale elettrico di uscita (output) del neurone. È provvisto di un'ampia varietà di organuli, come qualsiasi altra cellula organica, quindi è in grado di sintetizzare le proteine, i lipidi e i carboidrati. Funge anche da coordinatore per le attività metaboliche della cellula¹⁰⁹.
3. **L'assone:** è la via che porta a destinazione i segnali elettrici generati dal corpo cellulare. Normalmente, l'assone, è una fibra lunga e sottile, che si protende dal corpo cellulare. Per questo motivo il neurone è la cellula più lunga del corpo umano. Singoli assoni, ad esempio, partono dal midollo spinale e arrivano alle dita dei piedi, misurando più di un metro. Gli assoni sono le linee di distribuzione, attraverso le quali si propagano i potenziali d'azione in direzione centrifuga verso le estremità del neurone. Gli assoni sono per lo più avvolti in un fascio di nervi, come i fili di un cavo elettrico. A differenza di questo, che

107 Cfr. Stephen E. Palmer, *Vision Science – photons to phenomenology*, Op. cit., p. 28

108 http://www.benessere.com/salute/atlante/sist_nervoso.htm 12 ottobre 2006

109 Ibidem

dissipa energia nel tragitto tra la centrale e l'utente, la membrana plasmatica degli assoni fa pervenire, alle estremità del neurone, potenziali d'azione di intensità immutata, grazie, per buona parte, alla guaina mielinica che offre un ottimo isolamento¹¹⁰. Il segnale trasmesso è codificato in via principale attraverso la frequenza di trasmissione¹¹¹.

4. **I terminali sinaptici:** essi comunicano con altri neuroni, muscoli e ghiandole. Sono come dei rigonfiamenti delle estremità ramificate degli assoni. La trasmissione dei segnali alle altre cellule avviene a livello dei terminali sinaptici. La maggior parte di essi contiene una sostanza chimica specifica, detta neurotrasmettitore, che viene rilasciata come reazione a un potenziale d'azione che percorre l'assone. I terminali sinaptici di un neurone possono essere connessi ad una ghiandola, ad un muscolo con dendriti o ad un corpo cellulare di un secondo neurone, in questo modo il segnale in uscita (output) della prima cellula può diventare segnale in entrata (input) per la seconda¹¹².
5. **La sinapsi:** è il piccolo spazio che esiste fra i terminali di un neurone e i dendriti di un altro. Il neurotrasmettitore che viene rilasciato nella sinapsi, stimola i dendriti del neurone più vicino posizionandosi in zone specifiche della sua membrana.

Vedremo in seguito, parlando del cervello, come i neuroni abbiano un'enorme importanza per la codifica e il processo biochimico dei segnali, per ora, sapere come sono formati è sufficiente per procedere nella comprensione delle varie parti della retina.

La prima fila di cellule, nella parte posteriore della retina, è costituita dai fotorecettori: i bastoncelli e i coni, così denominati in base all'aspetto/forma che presentano se osservati al microscopio (la prima volta vennero descritti, in maniera piuttosto approssimativa, nel 1835 da G. R. Treviranus). I bastoncelli, sono circa 120 milioni, molto più numerosi dei coniche sono circa 8 milioni. Sono responsabili della visione in

110 Ibidem

111 Cfr. Stephen E. Palmer, *Vision Science – photons to phenomenology*, Op. cit., p. 29

112 http://www.benessere.com/salute/atlante/sist_nervoso.htm 12 ottobre 2006

condizione di ridotta illuminazione, in specie notturna, e sono fuori combattimento in piena luce. I coni non rispondono alle basse intensità luminose, ma sono responsabili della capacità di distinguere i dettagli fini, e della visione dei colori. La visione alla luce diurna, di pertinenza dei coni, viene denominata fotopica. Mentre la visione scotopica, fornita dai bastoncelli, è relativa alla luce crepuscolare. La visione mesopica, è riferita al caso intermedio, dato dalla luce lunare, che si colloca tra la brillantezza della luce solare e la debole luminosità delle stelle. Essa può risultare ingannevole¹¹³.

Se si considerano le dimensioni dei recettori retinici ci si rende conto di quanto possano essere sensibili: il più piccolo di essi misura 1 pm (picometro¹¹⁴), circa il doppio della lunghezza d'onda della luce rossa, dimensioni inferiori sono difficilmente ottenibili. La sensibilità, definita come la più piccola separazione rilevabile tra due linee adiacenti, date queste dimensioni, risulta molto inferiore del diametro della singola cellula a cono. Il cervello ricostruisce i gradienti di intensità in base alla quantità di segnali che riceve da un certo numero di recettori, quindi raggiunge una maggiore acuità visiva nella rilevazione delle linee che non dei punti¹¹⁵.

Le quantità relative di coni e bastoncelli variano in modo evidente nelle diverse zone della retina. Al centro, dove la visione del dettaglio è migliore, vi sono solo coni. Quest'area priva di bastoncelli, detta fovea, ha un diametro di circa mezzo millimetro. I coni sono presenti su tutta la retina, ma, nella fovea, sono in maggior numero e densamente stipati¹¹⁶.

Soltanto i coni provvedono alla visione cromatica. Tale conclusione è frutto di numerose ricerche effettuate sugli occhi di diversi animali, analizzando la loro struttura retinica e confrontando la capacità di distinguere i colori (rilevata attraverso una serie di osservazioni sul comportamento). Infatti si è constatando che non vi sono molti coni

113 Cfr Richard L. Gregory, *Occhio e cervello – La psicologia del vedere*, Op. cit., p. 78

114 Il picometro (simbolo: pm) è un'unità di misura della lunghezza nel SI (Sistema Internazionale di unità di misura), uguale a 10^{-12} metri. Altre unità di misura, con le relative corrispondenze, sono: 1 pm = 1000 femtometri; 100 pm = 1 ångström; 1000 pm = 1 nanometro.

115 Cfr Richard L. Gregory, *Occhio e cervello – La psicologia del vedere*, Op. cit., p. 80

116 Cfr David H. Hubel, *Occhio, Cervello e Visione*, Op. cit., p. 46

nella periferia della retina umana, da dove non si ottiene la visione a colori. Si è osservato inoltre che, la fovea, fornisce prestazioni migliori, a livello di dettagli visivi e cromatici, ma essa, risulta meno sensibile di altre regioni della retina ricche di bastoncelli (perciò gli astronomi "escludono" la fovea quando desiderano osservare stelle molto deboli. Essi cercano di far cadere l'immagine in una regione retinica ricca di bastoncelli più sensibili. Questo, però non consente di rilevare i colori delle stelle). Queste zone ricche di bastoncelli, secondo Gregory si sono formate in stadi evolutivi remoti, per questo egli le definisce "primitive"¹¹⁷.

Una domanda interessante è come i fotorecettori, coni e bastoncelli, siano in grado di cambiare l'energia elettromagnetica dei protoni in attività neurale. È un processo complesso e veramente ingegnoso conosciuto in modo abbastanza completo. Sia coni che bastoncelli sono formati da due parti: il segmento interno, che contiene il nucleo e l'apparato produttivo cellulare, e il segmento esterno che contiene bilioni di pigmenti molecolari fotosensibili. Questi pigmenti molecolari sono incastonati nelle membrane di migliaia di dischi impilati come "frittelle" perpendicolari al lungo asse del segmento esterno. Il pigmento nei bastoncelli è detto rodopsina¹¹⁸, e si conosce come questa molecola fotosensibile converta la luce in energia elettrochimica. Quando un fotone colpisce una molecola di rodopsina e ne è assorbito, questa cambia la sua forma in modo tale da alterare il flusso della corrente elettrica all'interno e vicino al pigmento molecolare. La parte interessante è che questa complessa reazione biochimica produce dei cambiamenti nella membrana esterna del recettore. Questi cambiamenti sono poi veicolati lungo la membrana esterna fino alla regione sinaptica del recettore, dove trasmettitori chimici influenzano il neurone successivo. Le variazioni elettriche che risultano da ogni fotone assorbito dallo stesso recettore sono unite nella risposta della sua membrana esterna. La variazione complessiva del potenziale elettrico dalla parte interna a quella esterna della cellula è graduale, continuo, non come in molte altre parti

117 Gregory, seguendo la teoria evolucionista darwiniana ritiene che si siano formati prima i bastoncelli e successivamente i coni.

Cfr Richard L. Gregory, *Occhio e cervello – La psicologia del vedere*, Op. cit., p. 79

118 Detta anche "porpora visiva", è conosciuta da più di un secolo ed è stata studiata in modo esteso, in quanto può essere velocemente estratta e studiata in vitro.

Cfr. Martin D. Levine, *Vision in Man and Machine*, Op. cit., p. 76

del sistema nervoso. In termini quantitativi, la risposta graduale di un fotorecettore è una funzione logaritmica del numero dei fotoni assorbiti. Questo significa che la stessa crescita complessiva in uscita richiede un'aggiunta di pochi fotoni a bassi livelli ma in quantità maggiori a livelli superiori. Questa complessa catena di reazioni nel segmento esterno è denominata sbiancamento del pigmento poiché le variazioni nella forma molecolare acquisite attraverso la luce ne causano anche la variazione di colore. Prima che una molecola di rodopsina sia sbiancata dalla luce appare violacea, dopo è quasi trasparente. Poiché è trasparente, è difficile che ne assorba un altro fotone. Per riacquistare il suo stato iniziale, la rodopsina, deve attendere l'intervento di un enzima contenuto nell'epitelio dei pigmenti dietro la retina¹¹⁹. Diversamente dai bastoncelli, i coni hanno tre differenti tipi di foto-pigmenti, ognuno dei quali risponde a una diversa lunghezza d'onda. Sono difficili da estrarre e vengono studiati in loco con la riflessione e il densimetro. Questi tre tipi differenti di coni, con diverse caratteristiche di banda passante, sono responsabili di tre diversi canali informativi che caratterizzano la visione a colori, come accennato. Ci si è chiesti se i coni interagiscano fra loro. Nella retina della tartaruga questo è evidente. Nel caso dei vertebrati sembra che questa interattività, che porta ad associazione elettrica, tenda a ridurre il disturbo dei fotorecettori a bassi livelli di luce. È anche interessante notare che l'acuità nella visione spaziale non ne è degenerata: anzi, la comunicazione fra fotorecettori può migliorare la percezione dell'immagine ad alti livelli di luce¹²⁰.

In una posizione particolare della retina, detta “macula lutea”, ci sono alcuni processi spettrali addizionali. Questa è un pigmento giallo schermato, che si trova solo nei primati, che funge da filtro per la parte blu-violetta dello spettro della luce. Questo effetto di filtraggio, aggiunto a quello effettuato dalle lenti del sistema ottico sulle onde corte, e all'assenza nella regione centrale della fovea di coni sensibili al blu, attenua, in modo significativo, la capacità di vedere le alte frequenze (ossia: radiazioni ultraviolette, raggi x e gamma)¹²¹.

119 Cfr. Stephen E. Palmer, *Vision Science – photons to phenomenology*, Op. cit., pp. 31-32

120 Cfr. Martin D. Levine, *Vision in Man and Machine*, Op. cit., p. 77

121 Ibidem p. 78

Dato che coni e bastoncelli sono nella parte posteriore della retina, la luce che arriva deve attraversare gli altri due strati per stimolarli. Non è ben compreso perché la retina si sviluppi in questo modo, a rovescio. Alcuni studiosi si rifanno allo sviluppo embriologico della retina (nei vertebrati avviene a partire dalla corteccia cerebrale)¹²². Un'altra ragione possibile, di tipo funzionale, è, l'esistenza, dietro i fotorecettori, di una fila di cellule che contengono un pigmento nero, la melanina (che si trova anche nella pelle). La melanina assorbe la luce che ha attraversato la retina, impedendo che sia riflessa all'indietro e si diffonda all'interno dell'occhio, si comporta come la vernice nera all'interno di una macchina fotografica. Questo tipo di cellule sono anche di supporto per la ri-sintesi chimica del pigmento visivo dei recettori, precedentemente sbiancato dalla luce. È necessario, per svolgere queste due funzioni, che il pigmento melanico sia posizionato vicino ai recettori. Quindi se i recettori si trovassero nella parte anteriore della retina, le cellule pigmentate si dovrebbero collocare tra essi e lo strato successivo di cellule, in una regione dove già vi sono assoni, dendriti e sinapsi. Così, invece, gli strati anteriori ai recettori sono trasparenti e, probabilmente, non offuscano molto l'immagine. Tuttavia nel millimetro centrale, la già menzionata fovea dove maggiore è l'acuità visiva, le conseguenze di un eventuale lieve offuscamento sarebbero veramente disastrose. Questa soluzione dev'essere il risultato ottimale del processo evolutivo, che, spostando lateralmente gli altri strati ha formato un anello più spesso, ed esposto i coni centrali direttamente alla luce. La fovea risulta quindi infossata, depressa¹²³.

Interessante è anche osservare la disposizione dei recettori retinici, che risulta a forma di matrice esagonale di coni con la presenza, negli spazi intermedi, di bastoncelli dalle dimensioni inferiori. Questo tipo di suddivisione e posizionamento, chiamato tassellamento, è ottimale, nel senso che ogni elemento ha un numero massimo di vicini equidistanti. Tale schema risulta impraticabile in un computer a scopi generici, nei quali il piano dell'immagine è suddiviso in rettangoli. Perciò, l'elemento base dell'immagine digitale nei computer, il pixel, risulta di forma quadrata¹²⁴.

122 Cfr Richard L. Gregory, *Occhio e cervello – La psicologia del vedere*, Op. cit., p. 76

123 Cfr David H. Hubel, *Occhio, Cervello e Visione*, Op. cit., p. 46

124 Cfr. Martin D. Levine, *Vision in Man and Machine*, Op. cit., pp. 75-76

Il secondo strato nel nostro ordine, risalendo verso la parte anteriore dell'occhio, è posto tra coni-bastoncelli e le cellule gangliari, si tratta di una posizione intermedia della retina. Esso contiene tre tipi di cellule nervose: le cellule bipolari, le cellule orizzontali e le cellule amacrine. Le cellule bipolari ricevono l'input dai recettori, come dimostra il disegno della Figura 10. Molte di esse si connettono direttamente con le cellule gangliari. Le cellule orizzontali connettono i recettori e le cellule bipolari. Sono connessioni (spesso del tipo di molti a uno) relativamente lunghe che scorrono parallele agli strati della retina; in modo simile, le cellule amacrine connettono le cellule bipolari e alle cellule gangliari (ancora connessioni di tipo molti a uno)¹²⁵.

Sulla superficie anteriore della retina c'è il terzo strato composto dalle cellule gangliari. Gli assoni di queste attraversano la retina, si raccolgono in un fascio a livello del disco ottico, quindi lasciano l'occhio andando a costituire il nervo ottico. I punti in cui i nervi ottici lasciano l'occhio sono denominati scotomi e sono zone cieche. Queste zone non costituiscono un grave problema in quanto vi è una cooperazione tra la visione bioculare e il cervello per costruire la parte mancante¹²⁶. Ciascun occhio contiene circa 128 milioni di fotorecettori, suddivisi in coni e bastoncelli, ma solo un milione di cellule gangliari, quindi lo spessore del nervo ottico risultante è inferiore. Le ridotte dimensioni del nervo ottico ne aumentano la flessibilità, che, a sua volta permette un miglioramento dei movimenti degli occhi. L'enorme differenza tra i fotorecettori e le cellule gangliari pone il problema di come venga mantenuto il dettaglio nell'informazione visiva. La risoluzione sta nell'analisi delle connessioni all'interno della retina. L'informazione che attraversa la retina può seguire due strade¹²⁷. La prima è una via diretta dai fotorecettori alle cellule bipolari e gangliari¹²⁸. La seconda è indiretta: possono essere interposte, tra recettori e cellule bipolari delle cellule orizzontali, oppure, delle cellule amacrine tra cellule bipolari e cellule gangliari¹²⁹. Il primo a descrivere queste connessioni dettagliatamente fu Ramon y Cajal intorno al 1900¹³⁰. Hubel precisa che “La via diretta

125 Cfr David H. Hubel, *Occhio, Cervello e Visione*, Op. cit., p. 47

126 Cfr Richard L. Gregory, *Occhio e cervello – La psicologia del vedere*, Op. cit., p. 84

127 Cfr David H. Hubel, *Occhio, Cervello e Visione*, Op. cit., p. 47

128 Ibidem

129 Ibidem

130 Ibidem

è altamente specifica, o privata, nel senso che uno solo, o pochi, recettori si connettono con una cellula bipolare, e solo una, o poche, cellule bipolari si connettono con una cellula gangliare. La via indiretta è più diffusa, o estesa, possedendo più ampie connessioni laterali. L'area totale dello strato posteriore occupata dai recettori che si connettono con una sola cellula gangliare dello strato anteriore, direttamente e indirettamente, è solo di circa un millimetro. Quest'area costituisce il campo recettivo¹³¹ della cellula gangliare, ossia la regione di retina la cui stimolazione luminosa modifica l'attività della cellula gangliare.¹³²

Questa struttura è presente in generale in tutta la retina. Tuttavia nel dettaglio le connessioni si diversificano molto passando dalla fovea (dove è massima la nostra capacità di discriminare i fini dettagli) alla regione più esterna, la periferia¹³³ (dove la visione diventa relativamente grossolana): cambiano le caratteristiche del collegamento tra il recettore e la cellula gangliare. Nella fovea e nelle sue vicinanze normalmente un singolo cono è connesso a una singola cellula bipolare, e, quest'ultima a una cellula gangliare. Allontanandosi dal centro, numerosi recettori confluiscono sulle cellule bipolari e queste convergono sulle cellule gangliari. Questo sistema di congiunzioni presente in gran parte della retina, assieme alla via privata della fovea e dintorni, fornisce una spiegazione di come la nostra visione risulti dettagliata anche in presenza di un rapporto di 128 a 1 tra i recettori e le fibre del nervo ottico¹³⁴.

In sintesi si può affermare che parte del processo percettivo avviene direttamente nell'occhio, che risulta quindi parte integrante del cervello, poiché i fotorecettori sono interconnessi e danno luogo a "campi recettivi". Questi risultano più ampi nel caso di luce debole, in quanto è sacrificata l'acuità necessaria a distinguere i dettagli minuti per

131 In senso proprio, il termine campo recettivo si riferisce semplicemente ai recettori specifici che fanno capo a una data cellula del sistema nervoso, con l'interposizione di una o più sinapsi. Attualmente il termine tende a includere la partizione del campo o, meglio, tende a dare una indicazione di come bisogna stimolare una regione per far rispondere la cellula. Cfr David H. Hubel, *Occhio, Cervello e Visione*, Op. cit., pp. 52-53

132 Ibidem p. 48

133 Si possono distinguere la fovea vera e propria di circa 0,3 mm, la "parafovea" di circa 2,5 mm, la "perifovea" con l'anello interno di circa 2,5 mm ed esterno di 5,5 mm e infine la "retina periferica" che costituisce il 97,25 per cento della superficie concava della retina. Cfr. Martin D. Levine, *Vision in Man and Machine*, Op. cit., pp. 73-75

134 Cfr David H. Hubel, *Occhio, Cervello e Visione*, Op. cit., pp. 47-49

favorire una maggiore sensibilità. Questo procedimento, gestito nell'occhio a livello neurale, è assimilabile alla possibilità di optare per l'utilizzo nella macchina fotografica di una pellicola più veloce e più granulosa per le riprese effettuate con scarsa luminosità¹³⁵.

Si aggiunga la presenza dell'inibizione laterale, la si potrebbe assimilare a un meccanismo neurale per rendere più definite le immagini. Funziona come segue: i fotorecettori più interessati dallo stimolo luminoso inibiscono quelli vicini meno coinvolti. Sembra, che questo meccanismo venga impiegato per filtrare il numero di segnali inviati al cervello, in quanto evidenzia i contorni, che risultano, spesso, molto significativi¹³⁶.

In generale, è possibile affermare che l'occhio umano ha carattere polifunzionale e, quindi gli mancano molte delle caratteristiche specifiche di altri tipi di occhio. Ne risulta che l'informazione visiva è poco selettiva e non vincolata a speciali esigenze. Il compito di operare le scelte opportune entro un ampio ventaglio di dati relativamente neutri viene lasciato, di volta in volta, al cervello¹³⁷.

Pur non essendo entrati nello specifico, se non per quanto servirà in seguito, e tralasciando alcuni aspetti di cui si parlerà più avanti, come la percezione delle distanze, dovrebbe essere fin d'ora chiaro che il processo visivo negli organismi viventi è costituito da una serie di analisi in ordine gerarchico lungo il percorso che va dagli occhi al cervello.

135 Cfr Richard L. Gregory, *Occhio e cervello – La psicologia del vedere*, Op. cit., p. 78

136 Ibidem

137 Ibidem p. 80

Il “processo di riconoscimento” delle immagini

Se l'occhio è l'organo che si occupa di captare i segnali luminosi e le immagini, è del cervello il compito di elaborarli e, poi utilizzarli.

Il cervello

Passando dagli occhi al cervello, la conoscenza scientifica di strutture (anatomiche, neurobiologiche, ecc) e funzioni scema visibilmente. Questo si verifica in parte per la mancanza di adeguati strumenti d'osservazione¹³⁸, e in parte si deve alla materia osservata. Per esaminare un minerale lo si può scomporre, frantumare, ecc senza alterarne la natura, ma ciò non è possibile in biologia.

Per sezionare un organo appartenente a un essere vivente bisogna innanzitutto reperirlo e, quindi, causarne la “morte”. Questo comporta problemi di natura umana, etica e sociale¹³⁹. Si aggiunga che lo studio di un tessuto vivo non è paragonabile allo studio di uno morto, tra l'altro soggetto a degenerazione.

Per questi motivi, si ricorre agli animali che presentano organi simili a quelli umani, come gatti e scimmie, in specie se vi è necessità di sezionarli¹⁴⁰. Purtroppo non si possono avere da questi delle risposte verbali. Anche se fosse possibile fare esperimenti sull'uomo vi sarebbe un grado di soggettività tale da intaccarne l'attendibilità.

138 Ad esempio il già citato microscopio, pur essendo presente in forme rudimentali da tempi immemori, è stato perfezionato nel 1590, quindi relativamente di recente, da due olandesi, Zaccharias Janssen e suo figlio. Essi osservarono che aumentando il numero delle lenti in un tubo e regolandone la distanza e la grandezza, l'oggetto osservato diventava incomparabilmente più grande. Era l'antesignano del microscopio composto e del cannocchiale sviluppato poi da Galileo in Italia.
<http://www3.iperbole.bologna.it/bologna/malpighi/mostra/microsco.htm> 17 ottobre 2006

139 In passato, per studi di anatomia, venivano usati i corpi dei defunti, con preferenza per donne e delinquenti in quanto si riteneva non avessero l'anima. Inoltre chi esercitava questo tipo di studi veniva spesso perseguitato; anche Leonardo da Vinci ebbe di questi problemi.
http://it.wikipedia.org/wiki/Anatomia_umana#Storia_dell.27anatomia 17 ottobre 2006
http://it.wikipedia.org/wiki/Leonardo_da_Vinci 17 ottobre 2006
http://www.sapere.it/tca/MainApp?srvc=dcmnt&url=/tc/scuola/percorsi/DP_anatomia/anatomia1.jsp 17 ottobre 2006

140 Contro la procedura di vivisezione esistono numerose associazioni e si è giunti ad alcuni progetti di legge, si veda:
<http://www.novivisezione.org/> 17 ottobre 2006

Per questi motivi, la ricerca è progredita lentamente. Come per altre scienze, molto ci si aspetta dall'utilizzo del computer, in particolare dalla computer grafica, sia come strumento per visualizzare e migliorare le immagini ottenute con sonde o altri meccanismi¹⁴¹, sia per le sue possibilità di simulazione di eventi e comportamenti che si verificano nel corpo umano¹⁴².

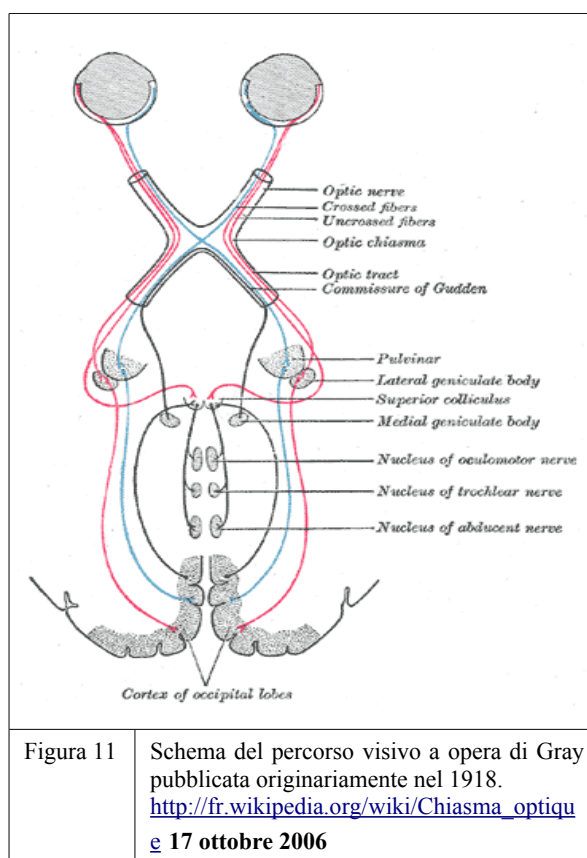
Per comprendere le funzioni del cervello è necessario conoscerne, almeno in parte, la struttura fisica. Hubel afferma, probabilmente a ragione, che il modo migliore è seguire il percorso che va dall'occhio al cervello.

Dall'occhio al cervello

Dallo strato della retina di coni e bastoncelli si passa a quello intermedio di cellule neurali bipolari, orizzontali e amacrine, per arrivare a quello delle cellule gangliari i cui assoni lasciano l'occhio formando il nervo ottico, un fascio dello spessore di una matita che contiene circa un milione di fibre.

141 Ad esempio le recenti tecniche di scansione PET (Tomografia ad emissione di positroni), NMR C (Risonanza Magnetica Nucleare) e CT (Tomografia computerizzata). Cfr Richard L. Gregory, *Occhio e cervello – La psicologia del vedere*, Op. cit., p. 98

142 Cfr Richard L. Gregory, *Occhio e cervello – La psicologia del vedere*, Op. cit., p. XII



I due nervi ottici, come si può vedere da figura 11, s'incontrano prima di raggiungere il cervello nel chiasma ottico¹⁴³. Qui le fibre che provengono dalla parte della retina vicina al naso si incrociano e passano dalla parte opposta del cervello, mentre le altre, provenienti dalla parte della retina vicina alle tempie, rimangono dove si trovano¹⁴⁴.

Di conseguenza, anche le informazioni s'incrociano: quelle che provengono dalla parte sinistra del campo visivo vanno all'emisfero destro e viceversa. Dato che il campo visivo dei due emisferi si sovrappone, quando si guarda frontalmente, ognuno dei due emisferi ne possiede i relativi dati. Questi vengono tra loro elaborati e integrati in modo da permettere la percezione della profondità.

Successivamente, gran parte delle fibre si dirige al nucleo del genicolato laterale del talamo e poi alla corteccia occipitale (o corteccia visiva primaria). La parte minore, invece, va al collicolo superiore, che sembra occuparsi di informazioni primarie

143 La denominazione deriva dal fatto che assomiglia alla lettera greca x, pronunciata "chi".

144 Cfr. Martin D. Levine, *Vision in Man and Machine*, Op. cit., p. 80

riguardanti la posizione spaziale degli oggetti e il controllo dei movimenti oculari¹⁴⁵. Da qui molte si dirigono al nucleo del pulvinar del talamo per poi finire nella corteccia occipitale come le prime.

Il nucleo del pulvinar media i riflessi pupillari fornendo alcuni segnali di feedback dai vari stadi del processo visivo. Il talamo, che contiene anche il corpo del genicolato laterale, è una parte del cervello dove convergono tutti i segnali sensoriali ad eccezione dell'olfatto. Si suppone che in esso vi sia qualche interazione di basso livello che integri i dati raccolti dal corpo sull'ambiente.

Il corpo genicolato laterale del talamo è una tappa principale nel percorso visivo umano, in quanto una frazione significativa delle fibre ottiche si uniscono qui, come fanno altre provenienti da parti diverse del sistema nervoso centrale. È una struttura laminale (striata) che contiene sei distinti livelli di cellule. Non vi sono motivi per ritenere che le principali analisi visive si svolgano qui; nondimeno è evidente che queste cellule sono coinvolte nella visione a colori.

Vi sono inoltre delle altre fibre, che provengono dalla fovea e che attraversano il chiasma, che non passano né per il nucleo genicolato laterale, né per il collicolo superiore, ma si dirigono a una formazione reticolare, le cui funzioni non sono ben chiare, ricongiungendosi poi ad altre parti della corteccia.

Fu Ramon Y Cajal a pubblicare, nel 1893, il primo studio completo sull'anatomia del cervello umano. Circa 30 anni prima, Golgi ne aveva stabilito l'organizzazione, usando una tecnica che permetteva la colorazione selettiva di alcuni gruppi di cellule. Più tardi, verso la fine del diciannovesimo secolo, si comprese che la corteccia è divisa in aree differenti ognuna con proprie funzioni.

Come già detto, alla fine del percorso sensorio c'è la corteccia cerebrale. Vediamola più in dettaglio.¹⁴⁶

145 Cfr. Stephen E. Palmer, *Vision Science – photons to phenomenology*, Op. cit., p. 35

146 Cfr. Martin D. Levine, *Vision in Man and Machine*, Op. cit., p. 84

La corteccia cerebrale

Il cervello non è una struttura uniforme. Esso è anzi diviso primariamente in tre parti: una mediana detta diencefalo (la cui struttura principale è il talamo) e due parti simmetriche e pari, gli emisferi destro e sinistro, che costituiscono il telencefalo. Essi sono connessi da una lamina di fibre nervose, chiamata corpo calloso ¹⁴⁷.

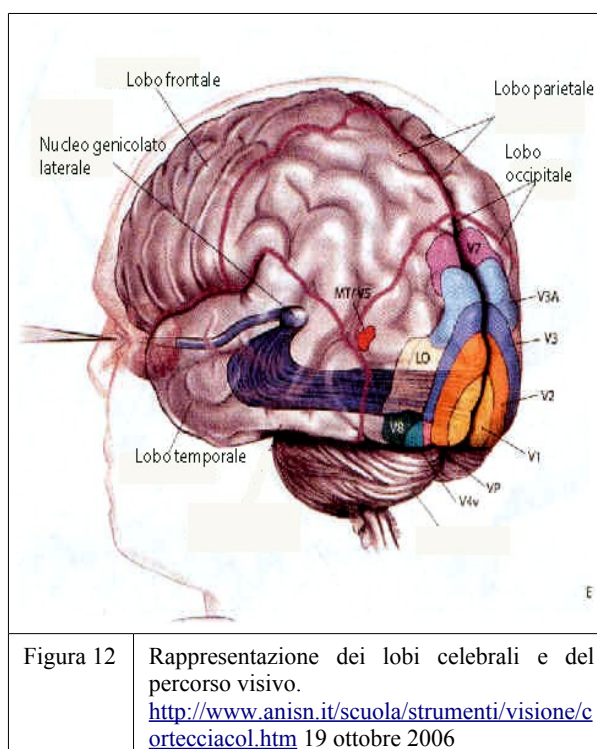


Figura 12 Rappresentazione dei lobi cerebrali e del percorso visivo.
<http://www.anisn.it/scuola/strumenti/visione/cortecciacol.htm> 19 ottobre 2006

I due emisferi cerebrali si suddividono in sotto-parti, come si vede nelle figure 12-13, denominate lobi: frontale, occipitale, temporale e parietale. Si presume che ciascuno di questi abbia una funzione differente: il lobo frontale viene associato alla pianificazione, alla strategia e all'azione; il lobo parietale è implicato nella sensibilità, in particolare del tatto; il lobo occipitale controlla la visione; quello temporale si occupa sia delle emozioni, sia di alcuni aspetti della memoria.

147 http://www.sapere.it/tca/MainApp?srcv=vr&url=/5/307_1 03 novembre 2006

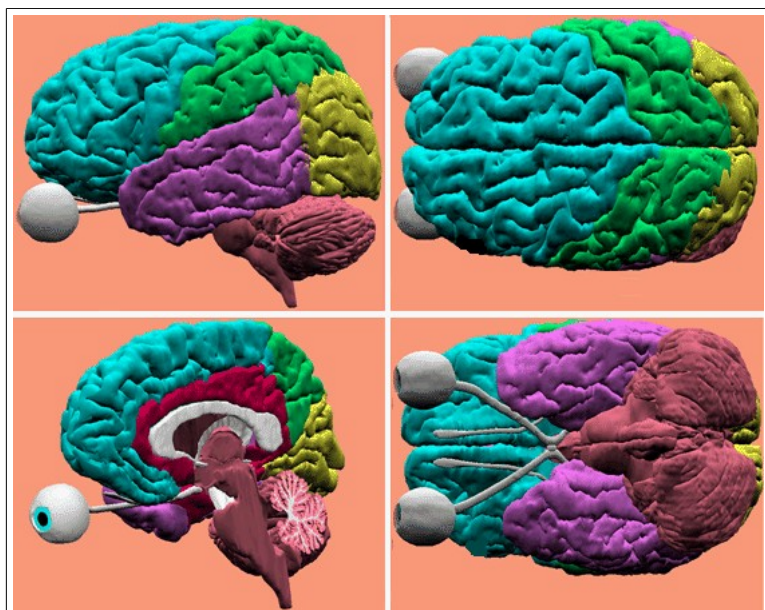


Figura 13: Ricostruzione in 3D dei lobi cerebrali da posizione laterale e da sopra. Si distinguono in celeste il lobo frontale, in verde il lobo parietale, in viola il lobo temporale e in giallo-oro il lobo occipitale. Negli spaccati sono evidenziati il nervo e il chiasma ottico.
<http://www.biocfarm.unibo.it/aunsnc/telenf.html> 7 novembre 2006

Le pareti degli emisferi sono costituite dalla sostanza grigia¹⁴⁸, la corteccia cerebrale¹⁴⁹, che contiene circa il 70 per cento dei neuroni presenti nel sistema nervoso centrale umano, percentuale che ne attesta l'importanza. Per quanto riguarda le sue caratteristiche fisiche, si tratta di una lamina ripiegata che va dai 2 ai 4 mm di spessore, perfettamente inserita nel cranio. Se dispiegata, la corteccia cerebrale occuperebbe circa $1/7 \text{ m}^2$. I neuroni sono circa 10^5 per millimetro quadrato di superficie, così che l'intera corteccia può essere considerata un network di 10^{10} di neuroni. Queste celle non sono posizionate a caso, ma sono organizzate in livelli, che alternano regioni densamente popolate con altre pressoché deserte. In una particolare regione si può notare un alto grado di uniformità delle celle¹⁵⁰.

148 Così denominata in quanto è il colore che assume nei cervelli non vivi conservati.

149 Nell'uso corrente, tutta la massa nervosa contenuta nel cranio, che invece è detta più propriamente encefalo, di cui il cervello costituisce la parte anteriore e superiore. Oltre al cervello, infatti, l'encefalo comprende il tronco cerebrale (bulbo, ponte di Varolio e mesencefalo) e il cervelletto.
http://www.sapere.it/tca/MainApp?srvc=vr&url=/5/307_1 03 novembre 2006

150 Cfr. Martin D. Levine, *Vision in Man and Machine*, Op. cit., p. 84

Si può spiegare questa struttura osservando che le fibre del nervo ottico si connettono alle cellule del corpo genicolato laterale. Gli assoni di queste ultime vanno poi alla corteccia visiva primaria. Questo insieme di connessioni ha un'organizzazione topografica: esiste una relazione sistematica (punto a punto) tra la mappa di una struttura e quella della successiva, in altre parole i diversi punti della retina, i punti corrispondenti del corpo genicolato laterale e quelli della corteccia possono essere congiunti da una linea continua. I processi nella corteccia, considerando il percorso seguito dalle fibre e i vari punti di elaborazione, risultano quindi localizzati in direzione laterale¹⁵¹.

È l'osservazione clinica che prova, da tempo e in modo significativo, l'organizzazione topografica delle vie visive. Quando si verifica una lesione anche di una piccola area della corteccia visiva primaria, si sviluppa una cecità localizzata, lo stesso effetto che si avrebbe se fosse lesa la parte corrispondente della retina¹⁵².

Queste scoperte sono una conquista relativamente recente, come dichiara Hubel: “Esiste dunque una mappa sistematica dell'universo visivo nel corpo genicolato laterale e nella corteccia. Negli anni Cinquanta non era del tutto chiaro il significato della mappa. A quel tempo non era accettata l'idea che il cervello eseguisse delle operazioni sulle informazioni che riceve, trasformandole in modo da renderle meglio utilizzabili. Si aveva la sensazione che la scena visiva fosse in qualche modo contenuta nel cervello: il problema era il significato che il cervello le attribuiva; forse non era un problema del cervello ma della mente.”¹⁵³. Oggi, grazie anche allo stesso Hubel e a Torsten Wiesel, si è appreso che la corteccia visiva primaria trasforma profondamente l'informazione che riceve¹⁵⁴.

È interessante scoprire che queste operazioni/trasformazioni/elaborazioni sono localizzate in punti distinti della corteccia. Il primo a formulare questa ipotesi fu

151 Cfr David H. Hubel, *Occhio, Cervello e Visione*, Op. cit., p. 70

152 Ibidem

153 Ibidem

154 Ibidem

Salomon Eberhard Henschen (1847-1930), seguito, successivamente da altri ricercatori. Essi mostrarono che la retina è connessa ad una zona specifica della corteccia cerebrale e non con l'intera corteccia. Questa zona del cervello fu detta inizialmente retina corticale, successivamente corteccia sensorio-visiva, e, di recente, corteccia visiva primaria o, con indicazione clinica abbreviata, area V1. In altre parole, esiste una zona specifica della corteccia cerebrale implicata nel peculiare processo della visione. Agli inizi del XX secolo Henschen faticò molto per sostenere la propria posizione, infatti molti credevano che le vie ottiche fossero connesse ad un'ampia area del cervello e che la visione non potesse essere ricondotta ad una sua parte specifica. La sua teoria era considerata un'assurda «mania di localizzazione» (*localisation à outrance*)¹⁵⁵.

Studi successivi hanno comunque dimostrato che esistono più aree interessate alla visione che circondano l'area V1. Queste aree sono state denominate in modi diversi: nel contesto di questo elaborato si userà una terminologia semplificata, e cioè V2, V3, V4, V5 o MT (Medial Temporal cortex, corteccia temporale mediana) e così via (fig. 12). I numeri crescenti non indicano una gerarchia, le varie zone cooperano in parallelo alla percezione visiva. V1 e V2 hanno il compito di selezionare e distribuire i segnali alle altre aree visive, mentre le altre zone sono specializzate nell'elaborazione e nella percezione di aspetti peculiari della scena visiva. A questa conclusione si è giunti solo negli ultimi 30 anni, perciò non è presente nelle prime teorie sulle funzioni e sul meccanismo visivo del cervello¹⁵⁶.

I primi studi che considerarono la localizzazione, l'organizzazione topografica e le osservazioni cliniche delle lesioni all'area V1 portarono alla conclusione che quanto era visto dall'occhio veniva interpretato dalle aree corticali circostanti (V2, V3, ecc, che hanno invece funzioni visive specifiche), definite “corteccia associativa”. La visione sarebbe quindi stata un processo essenzialmente passivo, distinto dalla comprensione¹⁵⁷.

Questa teoria, ebbe comunque degli effetti positivi. Infatti la «retina corticale», o V1,

155 Cfr. Semir Zeki, *La visione dall'interno – Arte e cervello*, Op. cit. pp. 31

156 Ibidem p.32

157 Ibidem

divenne in breve l'area più studiata del cervello visivo, e forse di tutto il cervello: si è così scoperto che alla fovea (la zona centrale della retina), corrisponde una quantità enorme di corteccia in rapporto alla parte periferica della retina che risulta invece sottodimensionata in considerazione della sua estensione sulla retina stessa¹⁵⁸. Quindi la «mappa retinica» nell'area corticale VI, se rapportata ad una comune lastra fotografica, risulta essere una trasposizione non conforme e affetta da deformazioni. Questa mappa evidenzia una parte specifica del campo visivo, così questa trasformazione viene denominata fattore di magnificazione corticale.

D'altro canto, tale teoria ha ritardato la comprensione della vera natura attiva del processo di visione, che scaturisce dall'interazione delle varie aree, le quali selezionano i dati disponibili, confrontano l'informazione selezionata con i ricordi immagazzinati e generano l'immagine, similmente al processo messo in atto da un artista¹⁵⁹.

L'idea che le immagini del mondo visibile s'imprimano sulla retina e, una volta trasmesse, vengano ricevute e focalizzate da una parte del cervello visivo, l'area VI, per essere successivamente interpretate da un'altra area corticale, è oramai superata¹⁶⁰.

Gran parte dei processi ha luogo in parallelo in sub-aree diverse, localizzate in zone topograficamente distinte. Ogni regione proietta delle fibre a numerose altre, anche se non a tutte. Le connessioni sono generalmente bidirezionali, cioè se l'area X è connessa con Y, Y a sua volta si connette con X. I due collegamenti non sono però completamente simmetrici in quanto si originano e terminano in livelli differenti della corteccia¹⁶¹. Si suppone, quindi, che la visione sia organizzata secondo un sistema modulare in parallelo, come visibile in figura 14. Curiosamente, Semir Zeki afferma che anche l'esperienza estetica è riconducibile allo stesso principio, confermando, in un certo senso, queste ipotesi¹⁶².

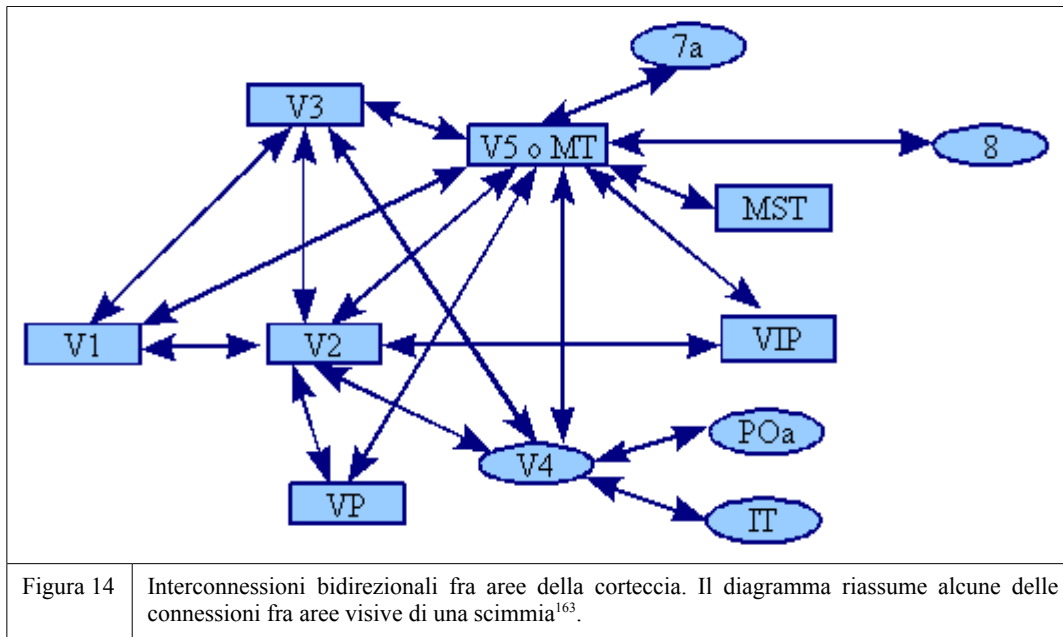
158 Ibidem p.34

159 Ibidem p. 38

160 Ibidem pp. 79-80

161 Cfr. Stephen E. Palmer, *Vision Science – photons to phenomenology*, Op. cit., p. 40

162 Cfr. Semir Zeki, *La visione dall'interno – Arte e cervello*, Op. cit. pp. 80



I segnali che raggiungono l'area V1 possono riferirsi ad aspetti diversi come: al colore, alla luminosità, al movimento, alla forma, alla profondità, e così via. Le cellule che ricevono i segnali delle varie caratteristiche visive sono raggruppate in diversi livelli dell'area V1, ragion per cui si dice abbia una struttura laminare o striata. Tra questi livelli sono importanti, ai fini della visione dei colori, i cosiddetti blobs. Sono delle piccole isole, abbastanza numerose, che si caratterizzano per un elevato metabolismo, e in cui sono stipate le cellule che sono sensibili alle diverse lunghezze d'onda della luce. le varie ripartizioni specializzate di V1, successivamente, inviano i loro segnali ad altre aree visive, sia direttamente sia utilizzando una zona intermedia che circonda V1, conosciuta come V2 (si veda ancora fig. 14)¹⁶⁴.

Il compito dell'area V1 è dunque, principalmente di smistare i segnali visivi, come un ufficio postale centrale, inviandoli alle diverse aree visive della corteccia che la circonda. Questa operazione comporta, a sua volta, una diversa specializzazione di ogni gruppo di aree, che dipende dal tipo di segnali che esse ricevono. Perciò il “cervello visivo” è un insieme eterogeneo di molte zone, la più importante delle quali è V1. Zeki lo definisce “sistema di elaborazione specializzato”: un sistema globale dedicato ad aspetti specifici della scena visiva, che comprende sia le cellule specializzate di V1 che

163 Stephen E. Palmer, *Vision Science – photons to phenomenology*, Op. cit., p. 42

164 Cfr. Semir Zeki, *La visione dall'interno – Arte e cervello*, Op. cit. pp. 80

le aree visive altrettanto specializzate destinatarie, sia direttamente che indirettamente dei segnali inviati dalle prime¹⁶⁵.

Poiché le singole cellule che costituiscono il cervello visivo presentano un'alta selettività al tipo di segnale o agli stimoli che ricevono ne consegue un'alta specializzazione funzionale dello stesso. Per inciso, una cellula può essere selettiva al colore e reagire al rosso, per esempio, ma non agli altri colori laddove altre cellule dello stesso tipo reagiranno in modo diverso ad altri colori. Queste cellule non risentono della direzione in cui lo stimolo si muove, è sufficiente che questo sia del colore adatto. In modo simile sono indifferenti alla forma, ossia reagiscono allo stimolo del colore adatto, indipendentemente dalla sua forma: una barra verticale o orizzontale, un rettangolo, un cerchio, un quadrato. Ancora, una cellula può selezionare un'altra qualità della scena visiva, ad esempio le linee con un'orientazione specifica, o i movimenti in una direzione stabilita, ecc. Nuovamente la selezione di un attributo specifico è associata all'indifferenza verso gli altri. Una cellula selettiva al movimento in una particolare direzione di un oggetto (cellula direzionalmente selettiva) non s'interessa né al suo colore né, di solito, della sua forma; praticamente le cellule più altamente selettive alla direzione, reagiscono in via ottimale a punti in movimento, ignorando le forme grandi e precise. Cellule che colgono lo stimolo di linee con un orientamento particolare saranno sensibili a questo, tralasciando il colore dello stimolo o dello sfondo su cui esso si staglia¹⁶⁶.

La specializzazione è, probabilmente, fondamentale per il complesso meccanismo cerebrale diretto a cogliere l'essenza degli attributi. La questione riguarda ora in che modo questo obiettivo venga raggiunto nei singoli sistemi specializzati, a tal proposito sono state avanzate diverse ipotesi sui meccanismi neurologici alla base della costanza dell'oggetto o del colore.

La specializzazione funzionale costituisce quindi, una prima soluzione escogitata dal cervello per risolvere il problema di conoscere il mondo, ossia le sue componenti

165 Ibidem pag. 82

166 Ibidem pag. 84

invarianti. La tipologia d' informazioni che il cervello non deve considerare per cogliere le caratteristiche principali di un oggetto, come il colore, differisce notevolmente dal tipo d'informazione che deve tralasciare per cogliere l'essenza di un altro attributo, come la dimensione. Nel primo caso è importante cogliere l'esatta lunghezza d'onda della luce proveniente solo da una superficie, nell'altro la distanza dall'osservatore. Probabilmente, a livello operativo, è più efficiente per il cervello valutare i diversi tipi di segnali in aree diverse, costruendole con un'anatomia e una fisiologia adatte a cogliere le caratteristiche principali di alcuni attributi distintivi. In sintesi, è probabile che esso abbia optato per un'elaborazione simultanea e parallela di diverse caratteristiche della scena visiva¹⁶⁷.

Una delle prove che attestano la specializzazione delle funzioni presente nel cervello umano avviene con l'utilizzo di sistemi che individuano le variazioni del flusso sanguigno in determinate zone del cervello stesso. Nel momento in cui le cellule della corteccia reagiscono ad uno stimolo si verifica un incremento della loro attività, nel dettaglio aumenta la frequenza d'impulsi la loro emessa rispetto allo stato di riposo. L'accelerazione di questa attività provoca, a sua volta, un aumento dell'attività metabolica che, a cascata, richiede un maggiore apporto di sangue ossigenato. Alcune tecniche sofisticate permettono di convertire in immagini questo processo, quindi di misurare l'aumento del flusso sanguigno, specifico di un'area. Così si può ottenere, con relativa precisione, la posizione dell'area considerata nel cervello. Questo metodo permette inoltre, di osservare che, se un individuo normale vede una scena policroma (come una configurazione astratta in cui non sono riconoscibili gli oggetti, ad esempio le realizzazioni di Mondrian in fig. 15) le variazioni del flusso sanguigno, nelle regioni cerebrali, si limita all'area V1, in quanto è il punto in cui convergono tutti i segnali dalla retina, e a una zona esterna, il complesso delle aree V4. Invece, se la stessa persona osserva una composizione di quadretti neri e bianchi che si muovono in direzioni diverse, il cambiamento del flusso sanguigno si verifica ancora in V1 – in quanto è il centro di raccolta principale di tutti i segnali visivi – ma non in V4 (si consideri per completezza che esso si verifica in un'area esterna a V1 e distinta da V4, l'area V5).

167 Ibidem pag. 84-85

Ulteriori esperimenti hanno mostrato che altre caratteristiche della scena visiva, come il riconoscimento di volti noti, vengono elaborate in aree diverse¹⁶⁸.

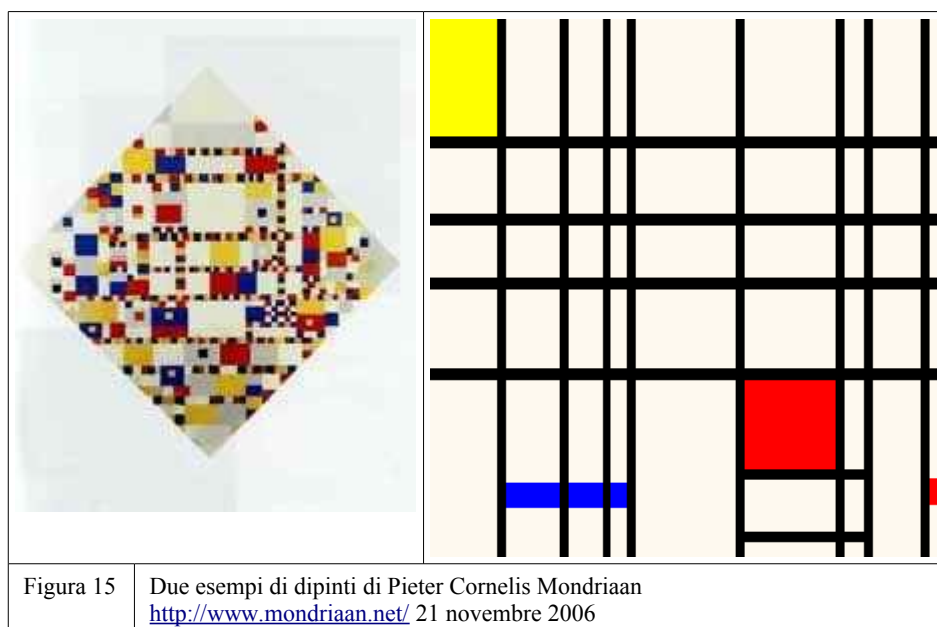


Figura 15 Due esempi di dipinti di Pieter Cornelis Mondriaan
<http://www.mondriaan.net/> 21 novembre 2006

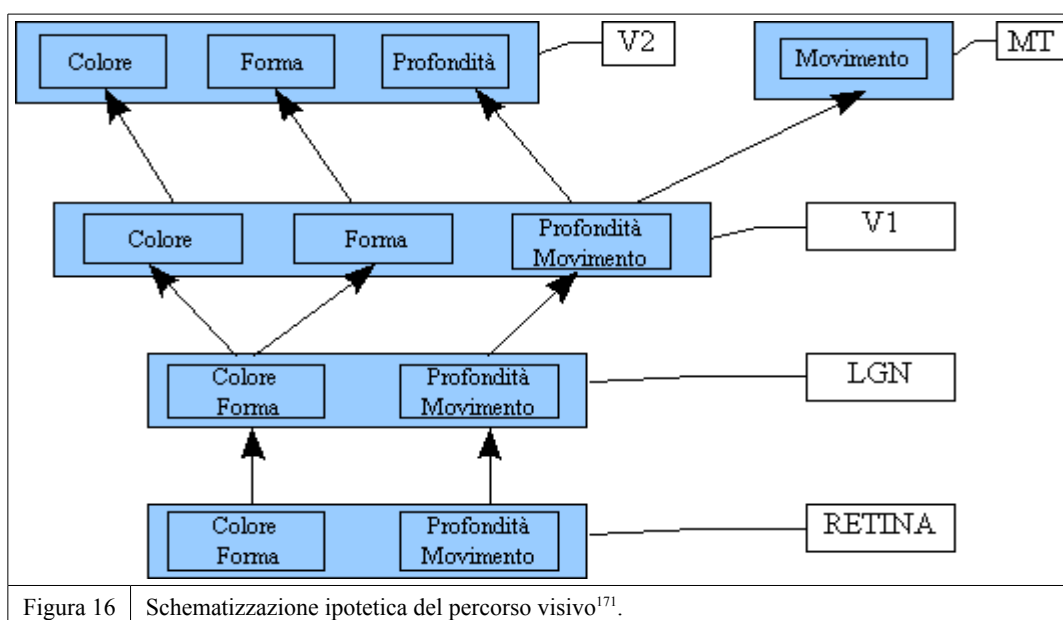
Sono stati Livingstone e Hubel a proporre che le informazioni relative a colore, forma, profondità e movimento fossero elaborate in differenti percorsi neurali a partire dalla retina, evidenziando relazioni simili a quelle riportate in figura 14. Essi tracciarono il percorso che va dalle cellule gangliari retiniche (differenziato in due canali, uno per colore e forma e l'altro per profondità e movimento), passa per il nucleo genicolato laterale e arriva a V1 e V2. Da queste aree partono i collegamenti verso livelli più alti della corteccia, dove vengono elaborate informazioni specifiche: il movimento e la visione stereoscopica della profondità nell'area V5, il colore in V4, la forma in numerosi e diversi centri intermedi dell'area IT (Infero Temporal cortex, corteccia temporale inferiore). Da qui, il percorso per forma e colore porterebbe alla parte inferiore del sistema, denominata “what”, incaricata d'identificare gli oggetti, mentre il percorso relativo a profondità e movimento si dirige verso la parte dorsale/superiore del sistema, chiamata “where”, incaricata della localizzazione degli oggetti¹⁶⁹.

La natura dei processi visivi nelle aree superiori della corteccia è meno chiara e

168 Ibidem pag. 87

169 Stephen E. Palmer, *Vision Science – photons to phenomenology*, Op. cit., p. 43

conosciuta che non quella dell'area V1. Ad esempio, si è scoperto che alcune cellule dell'area IT si attivano alla vista delle mani di una scimmia, mentre altre dalla visione del volto. La natura dei processi spaziali che avvengono tra V1 e IT rimane comunque misteriosa. L'analisi del movimento che avviene in area MT fornisce dati all'area MST (Media Superior Temporal cortex, corteccia temporale mediana superiore) e a numerose altre aree parietali. Ma anche qui si conosce poco degli specifici processi che avvengono in questi punti terminali. Se la ricerca per quanto riguarda la corteccia delle scimmie ha fornito risultati vaghi e a volte controversi, la comprensione del funzionamento e della struttura delle aree corticali umane è anche meno chiara¹⁷⁰.



La percezione integrata¹⁷²

Se si è certi che le caratteristiche della scena visiva sono elaborate separatamente, non si ha prova certa che siano anche percepite separatamente. Si ritiene che nel cervello si abbia un qualche tipo di integrazione, che permette di riunire assieme i risultati delle operazioni eseguite dai distinti sistemi di elaborazione visiva, ricavandone un'immagine unificata del mondo che vediamo, un'immagine in cui tutti gli attributi sono percepiti simultaneamente. Sono state effettuate varie ipotesi su come ciò avvenga ; la via

170 Ibidem

171 Ibidem

172 Cfr. Semir Zeki, *La visione dall'interno – Arte e cervello*, Op. cit. pp. 88-91

migliore per affrontare il problema in termini scientifici consiste probabilmente nel porsi la questione dell'esistenza di una “[...] registrazione temporale precisa dei risultati dell'attività compiuta dai diversi sistemi di elaborazione.”¹⁷³.

Quindi, il quesito è: tutte le caratteristiche della scena visiva elaborate dalle differenti aree, vengono immesse in una registrazione sincronizzata? Se si considera un lasso di tempo relativamente lungo, superiore al secondo, vengono percepiti tutti gli attributi in sincronia. È uno stimolo per comprendere come si forma l'immagine visiva integrata. Se si considera che un secondo è costituito da 1000 millisecondi, si comprende che risulta un tempo molto lungo in termini neurologici; infatti è sufficiente un impulso inferiore a un millisecondo per attraversare una barriera sinaptica (punto di contatto tra cellule nervose) e di circa 35 millisecondi perché i primi segnali visivi arrivino alla corteccia, anche se la maggioranza arriva più tardi, dopo 70-80 millisecondi. Si può presumere che se li si esamina in un intervallo temporale molto breve, si potrebbe trovare il processo d'integrazione ipotizzato?

In alcuni esperimenti si sono misurati i tempi necessari a percepire il colore, la forma e il movimento, dimostrando che vengono percepiti separatamente: prima viene il colore poi la forma, quindi il movimento. Tra la percezione del colore e quella del movimento passano circa 60-80 millisecondi. Si può inferire che anche i sistemi percettivi siano in possesso di una specializzazione funzionale, e che nella visione vi sia una gerarchia temporale, sovrapposta ai sistemi di elaborazione parallela che sono diversamente localizzati. Ne consegue che se due attributi di un medesimo oggetto - per esempio il colore e la direzione del suo movimento- variano in un brevissimo lasso di tempo, il cervello coglie prima il cambiamento di colore e dopo quello di direzione, dato che percepisce il colore prima del movimento. In generale, sembra che il cervello non abbia la capacità, se si considerano intervalli temporali brevissimi, di collegare ciò che accade in tempo reale; esso, al massimo, riesce ad unire i risultati dei suoi sistemi di elaborazione. Questo implica il verificarsi di errori di collegamento in tempo reale, o quantomeno delle approssimazioni.

173 Ibidem p. 89

I risultati forniti da questi esperimenti danno importanti indicazioni sul funzionamento del cervello visivo. Dimostrano come differenti sistemi di elaborazione percepiscano l'elemento di propria competenza in tempi differenti. Questo suggerisce quindi che le funzioni di elaborazione e percezione siano svolte dagli stessi sistemi, ciò, a sua volta, permette d'ipotizzare vi sia una serie di sistemi paralleli per la percezione e l'elaborazione. I risultati delle operazioni compiute da sistemi di elaborazione separati sono le differenti percezioni. Si può quindi considerare l'esistenza di una rete di sistemi percettivi e di elaborazione distribuiti spazialmente. Inoltre, se, per definizione, la percezione è un evento conscio, si deduce che si percepisce ciò di cui si ha coscienza mentre, ciò di cui si è inconsapevoli non viene percepito. Il fatto che percepiamo due attributi, quali il colore e il movimento, in momenti separati, porta alla deduzione esistano anche consapevolezze separate, ognuna corrispondente all'attività di uno specifico sistema di elaborazione-percezione, e, che, queste distinte consapevolezze non siano sincronizzate tra loro. Quindi, per avere una percezione cosciente della scena, è necessario unire, non tanto le differenti attività dei diversi sistemi percettivo-elaborativi, ma le microcoscienze ottenute dai diversi sistemi percettivo-elaborativi.

Gli esperimenti menzionati indicano anche un'altra conclusione: i vari sistemi percettivo-elaborativi,, possiedono un elevato grado di autonomia pur interagendo tra loro. Questo è confermato anche dalla letteratura patologica in materia già accennata.

La specializzazione funzionale e la successiva integrazione dei processi suggerisce che il cervello non è un semplice cronista che limita la sua attività alla registrazione passiva della realtà fisica del mondo esterno, ma è un soggetto attivo che partecipa alla creazione dell'immagine visiva, con proprie regole e programmi.

Queste regole e programmi sono stati oggetto d'analisi in diverse scuole di pensiero ed hanno portato alla formulazione di varie teorie, di cui si farà cenno in seguito, ma prima è necessario concludere l'accento fatto in precedenza sulla funzione e struttura delle cellule neurali.

Le cellule neurali, codifica ed elaborazione dei segnali

La struttura generale delle cellule neurali è già stata trattata nel paragrafo relativo alla retina; ora è il momento di approfondire il loro funzionamento, da una parte per quel che riguarda la codifica e l'elaborazione dei segnali, dall'altra per la loro stessa diversificazione/tipologia. Questo al fine di completare quanto esposto a riguardo della corteccia cerebrale e di fornire delle indicazioni sulle reti neurali, un settore in via di sviluppo per quanto riguarda l'intelligenza artificiale.

L'importanza di comprendere il funzionamento dei neuroni sta nel fatto che questi sono l'unità d'analisi più appropriata per il sistema visivo, nel senso che si ritiene di poter spiegare la visione in termini di modelli di trasmissione che si notano dall'interazione di molti singoli neuroni. L'esistenza di un metodo per lo studio dei singoli neuroni consente agli scienziati di tracciare uno schema delle interconnessioni dell'intero sistema visivo, specificando quale operazione è compiuta da ogni singolo neurone¹⁷⁴.

I segnali nervosi sono trasmessi sotto forma di sequenze di impulsi elettrici (detti potenziali d'azione) a frequenza modulata, i quali costituiscono la sola forma di comunicazione in entrata e uscita del cervello con il resto del corpo. È risaputo che la modulazione di frequenza offre un alto grado di stabilità e di rigetto del rumore rispetto alla modulazione d'ampiezza, il che spiega perché i neuroni funzionano in questo modo¹⁷⁵.

Cerchiamo di meglio comprendere come avviene la generazione dei segnali e dei potenziali d'azione: i segnali dipendono dal movimento di ioni attraverso la membrana della cellula (in altre parole, dalla sua permeabilità ionica). Normalmente all'interno della cellula vi è un eccesso di ioni di potassio, mentre all'esterno vi è un eccesso di ioni di sodio. La cellula in qualche modo trattiene all'interno gli ioni di potassio e all'esterno quelli di sodio, in modo che la concentrazione di ciascuno ione ai due lati della membrana cellulare rimanga diversa. Nella cellula esiste inoltre una pompa sodica che

174 Stephen E. Palmer, *Vision Science – photons to phenomenology*, Op. cit., p. 64

175 Cfr. Martin D. Levine, *Vision in Man and Machine*, Op. cit., p. 85

continua a espellere ioni di sodio man mano che questi vi penetrano. In ogni caso, esiste una differenza di potenziale elettrico di circa 80 mV (microvolt) tra i due lati della membrana cellulare, e l'interno è sempre elettricamente negativo rispetto all'esterno¹⁷⁶. Nello specifico, se si considera la parte interna di un assone a riposo, esso ha carica elettrica negativa rispetto alla superficie. Al comparire di uno stimolo - ad esempio quando un fotorecettore retinico è colpito dalla luce - la parte centrale della fibra diviene positiva, originando a un flusso di corrente che si diffonde come un'onda lungo il nervo. Quest'onda si muove con una velocità molto inferiore se rapportata a quella di un segnale elettrico lungo un conduttore: nelle fibre più ampie e spesse viaggia a una velocità di circa cento metri al secondo (360 km l'ora), mentre in quelle più sottili difficilmente raggiunge il metro al secondo (valori così bassi crearono stupore nei fisiologi del XIX secolo che ritenevano dovesse essere superiore alla velocità della luce). Le fibre più spesse, normalmente hanno un rivestimento di sostanze grasse - la guaina mielinica - che permette d'isolare ciascuna fibra da quelle vicine, oltre ad aumentare la velocità di conduzione dei potenziali d'azione¹⁷⁷.

Il tempo di reazione agli stimoli - di circa un ventesimo di secondo - caratteristico della specie umana è dovuto in parte al tempo di propagazione lungo nervi periferici del segnale elettrico, e in parte ad un ritardo necessario alla commutazione delle sinapsi cerebrali, che si occupano di elaborare e dirigere i segnali sensoriali e motori. Come visto, questi ritardi permettono di ottenere informazioni sui differenti processi cerebrali (elaborazione di contorni, colore, forme, movimento, ecc)¹⁷⁸.

I tempi di commutazione delle sinapsi, e quindi la velocità dell'azione potenziale, dipendono dalle proprietà della codifica di frequenza dello stimolo: la latenza e il periodo refrattario (o di refrattarietà). La latenza, o tempo di risposta effettivo, è definito come il tempo che trascorre dall'applicazione dello stimolo alla registrazione della variazione. Essa decresce all'aumentare dello stimolo. Il suo comportamento è simile

176 Cfr. Isaac Asimov, *Il libro di Biologia*, trad. it. Alessandra Fois, Milano, Arnoldo Mondadori, 1987 p. 351 (tit. originale *Asimov's guide to Science* 1984)

177 Cfr Richard L. Gregory, *Occhio e cervello – La psicologia del vedere*, Op. cit., p. 103-104

178 Ibidem

per assoni di cellule diverse, anche se i tempi possono differire da cellula a cellula. Il periodo refrattario è il tempo minimo che passa tra due stimoli successivi che provocano due azioni potenziali consecutive. Infatti, una reazione della cellula è provocata solo da quegli stimoli che superano un valore di soglia. Questo valore dipende anche dal periodo refrattario. C'è un periodo morto, denominato periodo refrattario assoluto, prima del quale è impossibile avere una reazione. Passato questo periodo, il valore di soglia del secondo impulso decresce esponenzialmente all'aumentare del periodo refrattario, raggiungendo il punto in cui il neurone sarà nuovamente in grado di reagire/eccitarsi. Come per la latenza, questo periodo si comporta similmente per diversi assoni, variando solo d'ampiezza. Ne consegue che anche in caso di uno stimolo costante, gli impulsi che ne derivano saranno comunque controllati da questi due parametri, e avranno un proprio andamento. Non solo, si ricava che il modo migliore per ottenere una reazione è variare continuamente lo stimolo, e che questo si propaga solo in avanti, dato che la porzione che si è appena eccitata non può eccitarsi nuovamente finché non è trascorso il periodo di riposo¹⁷⁹.

È importante sottolineare che non tutti gli stimoli provocano una reazione della cellula ma solo quelli che superano un valore di soglia. Infatti questo tipo di reazione può essere implementata/riprodotta artificialmente, sia dai programmi software, sia dai circuiti elettronici, come quelli denominati unità logica a soglia (ULS). Questa unità calcola una somma pesata dei suoi ingressi, confronta questa somma con un valore di soglia e produce un 1 se la soglia viene superata, altrimenti 0. In applicazioni in cui vi sono solo due azioni possibili, una singola ULS può a volte, essere in grado di calcolare l'azione corretta a partire da una appropriata codifica delle informazioni in entrata (vettore delle caratteristiche). Per problemi più complessi è necessaria una rete di tali elementi. È per questo che le reti neurali vengono così denominate, in quanto le ULS sono considerate modelli semplificati di neuroni biologici, che si attivano o meno a seconda dell'intensità e dell'attivazione delle loro connessioni sinaptiche¹⁸⁰.

179 Cfr. Martin D. Levine, *Vision in Man and Machine*, Op. cit., p. 89-90

180 Cfr Nils J. Nilsson, *Intelligenza artificiale*, trad. it. Salvatore Gaglio, Apogeo, 2002, p. 47-48 (ed. originale *Artificial Intelligence: a New Synthesis*, 1998)

Dunque, questo è il processo che trasforma il potenziale elettrico¹⁸¹ presente nell'assone in una serie d'impulsi a frequenza modulata. Proseguendo è interessante vedere come questi si propagano e permettono l'elaborazione del segnale. L'assone si connette ai dendriti delle altre cellule attraverso i terminali sinaptici. Queste sinapsi sono delle connessioni chimiche che impiegano delle sostanze trasmettitori per convogliare informazioni oltre il proprio limite. Gli impulsi di potenziale azione condotti lungo l'assone sono convertiti dalla sinapsi in potenziale elettrico nei dendriti, solitamente definito potenziale post-sinaptico (PostSynaptic Potential, PSP). Il PSP è direttamente proporzionale alla quantità di sostanza rilasciata e si satura in presenza di quantità elevate. Questo accade facilmente in quanto il tempo necessario affinché la sostanza attraversi lo spazio sinaptico è superiore al tempo intercorrente tra un impulso e il successivo. Quindi nuovo potenziale è sommato a quello rimanente del precedente impulso, facendo decadere il PSP (a seguito di latenza e periodo refrattario), portando ciò che si dice un basso potenziale. Se ne ricava che l'ampiezza di questa depolarizzazione¹⁸² dendritica è proporzionale alla frequenza media alla quale gli impulsi arrivano alla sinapsi, una forma di codifica a potenziale di frequenza.

Il tutto inizia da circa 10 neuroni diversi: sembra che ogni neurone debba proiettare una quantità elevata di sinapsi ai neuroni a sé connessi (per ogni neurone vi sono dalle mille alle centomila sinapsi di questo tipo). In ragione del gran numero di vie di stimolazione, una media di 104 fino a 105 PSP al secondo si propagano attraverso le sinapsi. Le giunzioni sinaptiche si trovano normalmente tra assone e dendriti, tuttavia vi possono essere anche tra assone e assone, tra dendriti e dendriti, tra assone e soma¹⁸³.

Un aspetto interessante dal punto di vista computazionale è che possono esistere due tipi di congiunzioni sinaptiche: eccitatorie e inibitorie, che non sono però sinonime di incrementi positivi o negativi. In aggiunta, la quantità di depolarizzazione o iper-

181 Il potenziale elettrico, noto anche con il nome di potenziale scalare, viene indicato dalla lettera V, o a volte anche dalla greca ϕ . Esso si ricava a partire dal lavoro del campo elettrico.

http://it.wikipedia.org/wiki/Potenziale_elettrico 12 dicembre 2006

182 La depolarizzazione è una variazione/escursione positiva rispetto allo stato di riposo del neurone situato a circa -60 mV, quindi normalmente negativo.

183 Cfr. Martin D. Levine, *Vision in Man and Machine*, Op. cit., p. 91-92

polarizzazione¹⁸⁴ è pesata da ogni singola sinapsi, comportando una moltiplicazione parametrica. L'ampiezza del parametro può dipendere da determinate caratteristiche anatomiche come la grandezza della sinapsi e la distanza di questa dalla soma, ossia la lunghezza del dendrite. Ne risulta che l'adattamento/controllo è possibile anche prima del gap sinaptico, se l'assone si dirama nel punto terminale e vi sono due connessioni giustapposte. Questo tipo di sinapsi è più raro, e utilizzato per trasmettere un potenziale relativamente basso (rispetto alle frequenze degli impulsi in arrivo alla congiunzione), il quale controlla l'incremento della quantità di sostanza rilasciata dal terminale. Quindi aggiunte incrementalmente vengono utilizzate per modulare l'ampiezza del segnale in uscita dall'assone (potenziale pre-sinaptico) e il successivo "effetto domino". Questo meccanismo viene definito "inibizione pre-sinapsi", in quanto sono coinvolte solo le sinapsi dell'assone che portano una depolarizzazione¹⁸⁵.

È curioso notare come questo meccanismo eccitatorio/inibitorio delle singole cellule sia riprodotto a livello generale del sistema nervoso. Le fibre nervose si raggruppano infatti per formare due sistemi nervosi chiamati ortosimpatico e parasimpatico. Questi agiscono sull'intero organismo con effetti opposti. Le fibre ortosimpatiche accelerano il battito cardiaco, mentre quelle parasimpatiche lo rallentano; le fibre ortosimpatiche diminuiscono la secrezione dei succhi gastrici, mentre quelle parasimpatiche la stimolano, e via dicendo¹⁸⁶.

A questo punto è corretto chiedersi cosa accade alla moltitudine di dendriti uniti al soma. Un processo additivo, nel corpo cellulare, porta a una media pesata di tutti gli stimoli (inibitori e non) in entrata, che vengono convertiti nella frequenza in uscita. Quanto esposto è una semplificazione del processo, che non è ancora del tutto chiaro.

È bene riconsiderare come il sistema nervoso comunica con l'ambiente visivo esterno. L'intensità della luce è trasformata dalle cellule specifiche (gangliari, ecc), in basso potenziale elettrico. La domanda che sorge è: quali sono le caratteristiche di queste e

184 Escursione negativa, si veda la nota precedente relativa alla depolarizzazione.

185 Cfr. Martin D. Levine, *Vision in Man and Machine*, Op. cit., p. 91-92

186 Cfr. Isaac Asimov, *Il libro di Biologia*, Op. cit. p. 347

delle altre cellule coinvolte nella visione?. La risposta viene dalle ricerche di Hubel e Torstin Wiesel, che registrarono l'attività delle singole cellule dell'area striata del cervello di un gatto quando gli ponevano dinanzi agli occhi semplici forme visive. Erano, in genere, barre luminose, proiettate su uno schermo posto di fronte al gatto. Hubel e Wiesel notarono che alcune cellule corticali del primo stadio di elaborazione (V1) davano segni di attività solo se la barra veniva posta in una determinata direzione; per inclinazioni diverse non registravano alcuna attività. Altre cellule tuttavia, rispondevano a inclinazioni diverse; alcune cellule cerebrali più interne erano sensibili a caratteristiche più generali, e si dimostravano attive indipendentemente da dove veniva stimolata la retina; altre ancora erano sensibili solo al movimento, e alcune di queste solo al movimento verso una specifica direzione. Questi esperimenti mostrano come nel cervello vi siano dei meccanismi specifici per la selezione di alcune caratteristiche degli oggetti. La percezione si realizza unendo queste caratteristiche¹⁸⁷.

Per comprendere e avere un approccio più conciso è necessario riprendere il concetto di campo recettivo, già accennato trattando la retina. Il primo a darne una descrizione formale fu Edgar Douglas Adrian, seguito dalle ricerche di Hartline, Barlow e Stephen Kuffler¹⁸⁸: si tratta, semplicemente, di un modello di cellule ricettrici del segnale luminoso che, a loro volta, suscitano la risposta di singole cellule sparse lungo il percorso visivo. Kuffler in particolare ne individuò due tipi: centro-on e centro-off. Egli sperimentò, ma non fu il solo, che illuminando la retina con un raggio di luce di una certa ampiezza si aveva una certa frequenza di risposta. Aumentando o diminuendo la superficie colpita si aveva sempre una diminuzione della risposta, per cui intuì che vi era una sorta di antagonismo tra una zona circolare interna (centro) e un anello circostante. Le cellule centro-on rispondono quando si illumina il centro del campo recettivo e la risposta diminuisce quando si illumina anche la parte circostante, che ha quindi una funzione inibitoria. Le cellule centro-off non rispondono quando si illumina il centro (che ha funzione inibitoria), ma quando si colpisce l'anello esterno¹⁸⁹.

187 Cfr Richard L. Gregory, *Occhio e cervello – La psicologia del vedere*, Op. cit., p. 110

188 Cfr. Martin D. Levine, *Vision in Man and Machine*, Op. cit., p. 102

189 Stephen E. Palmer, *Vision Science - Photons to Phenomenology*, Op. cit., p. 147

La tecnica della mappatura dei campi recettivi spiega la prima parte dell'esperimento di Hubel e Wiesel, ossia come si possono registrare delle variazioni/risposte (in gergo viene spesso usato “firing”, sparare) nella corteccia stimolando la retina.

Per quanto riguarda il resto dell'esperimento, si può dire che la scoperta, come spesso avviene, fu casuale: infatti, i due ricercatori si aspettavano che vi fossero delle reazioni a un singolo punto e non a determinate caratteristiche dell'immagine. Comunque, ulteriori approfondimenti rivelarono che vi erano tipi diversificati di cellule corticali con campi recettivi dalle caratteristiche differenti. Esse furono poi classificate in tre tipi: semplici, complesse e iper-complesse¹⁹⁰.

Le cellule semplici furono così definite in quanto la loro risposta a uno stimolo complesso poteva essere prevista in base alla loro risposta a singoli punti di luce. Quindi il loro campo recettivo può essere individuato rilevando la risposta a singoli punti di luce che colpiscono vari punti della retina. Poiché alcune di queste cellule rispondono con più veemenza a un margine luminoso con un corretto orientamento e una corretta polarità (da chiaro a scuro o da scuro a chiaro), esse sono state spesso denominate *rilevatrici di bordi*. Altre hanno una lunga regione centrale che è sia eccitante sia inibitoria, con campi antagonisti da entrambi i lati, per questo rispondono bene sia a linee luminose che scure, e vengono denominate *rilevatori di linee o barre*. Hubel e Wiesel supposero che questi tipi di rilevatori di bordi e di linee potessero essere formati da una serie di unità centro-periferia nel nucleo genicolato laterale¹⁹¹.

Le cellule complesse sono le più comuni nella corteccia striata (circa 75% della corteccia striata). Sebbene anch'esse abbiano lunghi campi recettivi, differiscono dalle cellule semplici per alcuni aspetti:

1. **Non linearità.** Esse sono altamente non lineari, nel senso che rispondono poco o nulla a piccoli punti stazionari. Ne risulta che il preciso orientamento dei loro campi recettivi non può essere desunto con tecniche che misurino la loro risposta

190 Ibidem p. 151

191 Ibidem

a singoli punti proiettati sulla retina.

2. ***Sensibili al movimento***. Tendono ad essere altamente reattive a linee o bordi in movimento posizionati nel loro campo recettivo. Spesso sono anche specializzate per direzione (sinistra-destra, alto-basso, in obliquo, ecc).
3. ***Insensibili alla posizione***. Non riconoscono la posizione di alcuni stimoli. Lievi differenze nel posizionamento di una barra o di un reticolo, ad esempio, non influiscono molto sulla loro risposta.
4. ***Estensione spaziale***. Le cellule complesse tendono ad avere dei campi recettivi mediamente più ampi delle cellule semplici.

La prima cellula corticale con cui Hubel e Wiesel “parlarono” attraverso il bordo della loro diapositiva era, infatti, una cellula complessa. Le caratteristiche appena esposte spiegano perché i tentativi fatti con luce diffusa, punti fissi, ecc, fossero andati a vuoto.

Hubel e Wiesel proposero che queste cellule fossero costituite dall'integrazione delle risposte di molte cellule semplici. Generalmente sembra questo il caso, tuttavia alcune ricevono direttamente gli stimoli dal nucleo genicolato laterale¹⁹².

Il terzo tipo di cellule, le iper-complesse, è stato così definito dai due ricercatori in quanto i loro campi ricettivi sono ancora più selettivi. La loro caratteristica più sorprendente consiste nel fatto che se si allunga una linea o un bordo oltre una certa lunghezza la loro risposta diminuisce, mentre risulta essere più vigorosa nel caso di linee o margini più corti. Per questo motivo vengono dette cellule end-stopped (di fine corsa). Le ricerche più recenti postulano che queste cellule siano cellule semplici o complesse con aggiunta questa caratteristica, piuttosto che una categoria a sé. In aggiunta si ritiene che il grado di “end-stopping” sia un continuo piuttosto che un fenomeno del tipo “o tutto o niente”¹⁹³.

A questo punto sono state delineate per sommi capi le strutture anatomiche di occhio e

192 Ibidem p. 153

193 Ibidem pp. 153-154

cervello. E' il momento di dare uno sguardo ad alcune teorie relative alla visione.

Approcci alla visione¹⁹⁴

Nonostante si sia già accennato a differenti teorie è bene, prima di riportare quali siano le principali riguardanti la visione, operare una riflessione sul concetto di teoria. Quando si cerca di comprendere un fenomeno complesso, considerare semplicemente tutti i fatti rilevanti spesso non è sufficiente. È necessaria una teoria. Una teoria è un insieme integrato di dichiarazioni (definite ipotesi) riguardanti i meccanismi sottesi o i principi che non solo organizzano e spiegano i fatti conosciuti, ma che permettono anche di prevedere quelli che stanno per accadere.

Esse hanno tutte in comune il fatto di essere inadeguate in quanto per un dato dominio ci può essere solo una “corretta” teoria. Ma allora perché studiarle, analizzarle, ecc? La ragione sta nel fatto che comprenderne gli errori e tentare di migliorarle è un componente essenziale dell'attività scientifica, come affermavano Popper e Kuhn. Ovviamente le ipotesi di una teoria devono essere internamente consistenti, cioè non contraddirsi fra loro, pena il rigetto della teoria stesa. Non solo, esse devono essere anche esternamente consistenti, ossia non contraddire alcuni (idealmente tutti) fatti ormai associati del dominio considerato. Perciò la teoria è considerata seriamente solo se risulta logicamente ed empiricamente consistente.

Tuttavia, può accadere che vi siano più teorie logicamente ed empiricamente consistenti che giustifichino un insieme di fenomeni (ad esempio, per la luce, la contrapposizione tra teoria ondulatoria e corpuscolare, confluite poi in una sola). Come si decide qual è la migliore? Gli scienziati credono nel principio del rasoio di Ockham. Questi dichiara che la teoria migliore è quella più parsimoniosa, cioè quella che può contare su risultati empirici con il minor numero di assunzioni.

Avendo così introdotto l'argomento, passiamo a considerare le teorie della visione, e come spesso succede, partiamo da una domanda: ***“Perché le cose appaiono come***

¹⁹⁴ Il paragrafo che segue è quasi tutto tratto dal cap. 2 di Stephen E. Palmer, *Vision Science - Photons to Phenomenology*, Op. cit. a cui si rimanda.

sono?” Tale quesito è stato posto da Kurt Koffka, illustre psicologo della Gestalt.

Vi sono molte possibili risposte, ma nell'arco del tempo sono emerse le seguenti tre argomentazioni:

1. La contrapposizione Ambiente – Organismo:

Con la risposta “*Perché il mondo è fatto così*” si prende la posizione di chi ritiene che l'analisi degli stimoli esterni sia il modo corretto per comprendere la percezione. In particolare si suggerisce che si dovrebbe esaminare il tipo d'informazione presente nello stimolo prossimo (nella retina) con lo stimolo distale (esterno, nell'ambiente)¹⁹⁵.

Sulla stessa lunghezza d'onda le risposte “*Poiché siamo come siamo*” o “*Poiché il nostro sistema visivo è fatto così com'è*”. Tutte e tre le risposte, e le possibili posizioni intermedie, difendendo un approccio interno alla struttura percettiva, enfatizzando la natura dell'organismo piuttosto che la natura del mondo circostante.

2. La contrapposizione Empirico – Innato:

La risposta “*Perché abbiamo imparato a vederle così*” dichiara esplicitamente che la visione dipende dall'esperienza accumulata attraverso l'interazione con l'ambiente/mondo.

L'alternativa “*Perché siamo nati per vedere in questo modo*” afferma che vediamo in quanto il vedere è una nostra capacità innata, assicurata dai nostri meccanismi neurali. L'apprendimento sarebbe avvenuto durante l'evoluzione a livello di specie. Il fatto che queste abilità abbiano bisogno di un periodo di maturazione non incide su questa visione, infatti essa ritiene che la

195 Si veda il paragrafo sulla retina.

programmazione genetica sia predominante.

3. Atomismo contro olismo:

“Dipende dal modo in cui appare ogni singola unità del campo visivo”. Con questa risposta gli atomisti sostengono che sia possibile ricostruire l'intero campo visivo mettendo insieme l'esperienza visiva di ogni singola regione.

“Dipende da come è organizzato l'intero campo visivo”. Con questa risposta “olisitca”, che si contrappone alla precedente, si considera insufficiente aggiungere le singole unità, in quanto la loro valenza è data dall'insieme e nell'insieme. In questo modo la percezione di una parte sarà influenzata dalle altre. Questo approccio sottende che gli stimoli siano organizzati in modo che le proprietà di interi oggetti o scene prendano il posto di quelle delle varie regioni locali.

A questi approcci s'aggiunge quello comportamentale, che differisce dagli altri dal punto di vista metodologico. Esso, infatti, non risponde alla domanda così come formulata da Koffka il quale, ritenendo la visione un evento conscio, aveva un atteggiamento introspettivo. La domanda sottesa di stampo comportamentista è: *“Che cosa ci permette di fare la visione?”*. La risposta è quanto mai complessa e sarà l'oggetto delle parti successive della tesi, almeno per quanto riguarda la computer vision.

Altre teorie possono essere considerate nei vari approcci alla percezione visiva. Eccone una breve spiegazione.

Strutturalismo

In termini temporali, è una delle prime teorie riguardanti la percezione visiva. Lo Strutturalismo rientra nella scuola filosofica denominata “Empirismo inglese”, con

riferimenti specifici negli scritti di Locke, Berkeley e Hume. Il padre fondatore è ritenuto Wilhelm Wundt, che la diffuse in Germania; successivamente la teoria arrivò negli Stati Uniti, grazie al suo discepolo Edward Titchener.

La teoria strutturalista propone la visione come risultato di un processo che parte da unità sensitive atomiche – definite primitive, ossia elementi indivisibili d'esperienza in una data modalità sensitiva – e arriva ad evocare ricordi di altre unità atomiche sensitive, che sono già state associate in memoria. Queste associazioni sono predisposte per avvenire ogni volta che un'esperienza sensoriale è sufficientemente vicina, nello spazio e nel tempo, ad un numero sufficiente di apparizioni. Nel caso della visione, le unità sensibili sono predisposte all'esperienza visiva del colore in ogni minuta regione localizzata nel campo visivo, che si presume provenga dai singoli fotorecettori della retina. Si ritiene che queste singole sensazioni siano unite in percezioni attraverso la semplice concatenazione: ossia, messe insieme come quando si crea un'immagine con la sovrapposizione di diverse diapositive.

L'esperienza visiva non richiama solo la memoria visiva, ma anche quella uditiva, tattile, olfattiva, e gustativa (ad esempio la mela rossa). La percezione è quindi programmata per avvenire a mezzo di processi associativi rapidi e inconsci, che vanno oltre la memoria acquisita attraverso le precedenti esperienze. Di conseguenza, come un osservatore diviene sempre più esperto del proprio ambiente/mondo attraverso meccanismi associativi, la sua percezione diventa sempre più ricca, più accurata e complessa.

Dai termini utilizzati è possibile notare l'analogia tra la teoria strutturalista e la chimica: la relazione tra semplici sensazioni e percezioni complesse riprende quella tra atomi e molecole. Ciò che unisce il tutto è l'associazione.

Tale descrizione mostra come al centro della teoria strutturalista vi sia un approccio empirista e atomista. Dal punto di vista metodologico i sostenitori dello Strutturalismo utilizzarono l'introspezione guidata. Questa consiste nel cercare di “guardarsi dentro” e

osservare le proprie esperienze. Questa capacità non è comune, tanto che per acquisirla gli strutturalisti propongono un periodo di formazione guidato da un esperto.

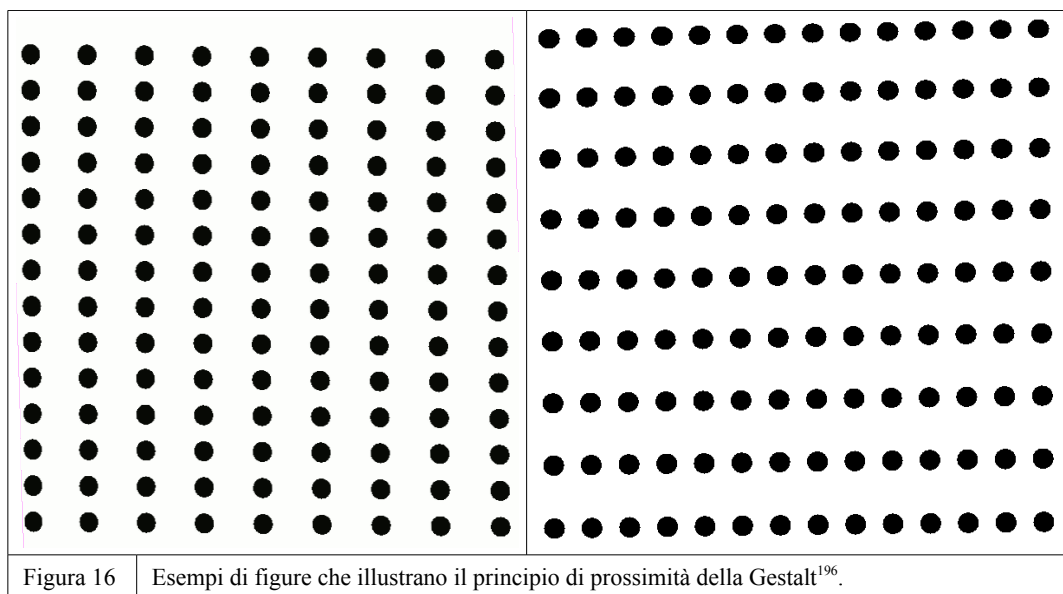
Le grandi difficoltà incontrate dalla teoria sono strettamente legate al metodo utilizzato. Esso risulta troppo soggettivo, mentre il periodo di formazione più che migliorare i risultati li può influenzare. Per questi motivi la teoria strutturalista ha lasciato spazio alle teorie che si basano sull'analisi comportamentale.

Gestalt

Storicamente il movimento della Gestalt sorse come reazione allo Strutturalismo. Il termine, di origine tedesca, significa semplicemente “forma intera” o “configurazione”.

Questa teoria, come descritta originariamente nelle opere di Koffka, Wertheimer e Köhler, afferma che la percezione opera attraverso schemi innati di cui è possibile studiare le proprietà e le leggi di organizzazione. Il motto “L'intero è maggiore della somma delle parti” rende bene la posizione dei gestaltisti: la percezione è il risultato dell'organizzazione delle sensazioni, più che la loro associazione. L'intero possiede una propria struttura non rilevabile dalle singole parti che lo compongono.

Sono stati inoltre postulati dei principi di organizzazione percettiva per giustificare la probabilità con cui alcune percezioni si verificano più di altre. Alcuni di questi riguardano il modo in cui si raggruppano gli elementi delle figure, altri la separazione della figura dallo sfondo.



Per quanto riguarda il raggruppamento abbiamo¹⁹⁷:

1. Il **principio di prossimità**:

Ad esempio, in figura 16, i puntini neri nella parte sinistra si organizzano percettivamente in righe verticali in quanto sono più vicini tra loro lungo l'asse verticale che non lungo l'orizzontale; viceversa nella parte destra i puntini neri formano righe orizzontali per la maggiore prossimità lungo una direzione orizzontale.

2. L'effetto cambierebbe se i pallini fossero neri e bianchi in quanto entrerebbe in gioco il **principio di somiglianza** che prevale su quello della prossimità.
3. Vi è poi il **principio di continuazione** secondo il quale la continuità delle linee viene preferita ai cambiamenti bruschi.

4. Infine il **principio di chiusura**, secondo cui le forme chiuse sono preferite a

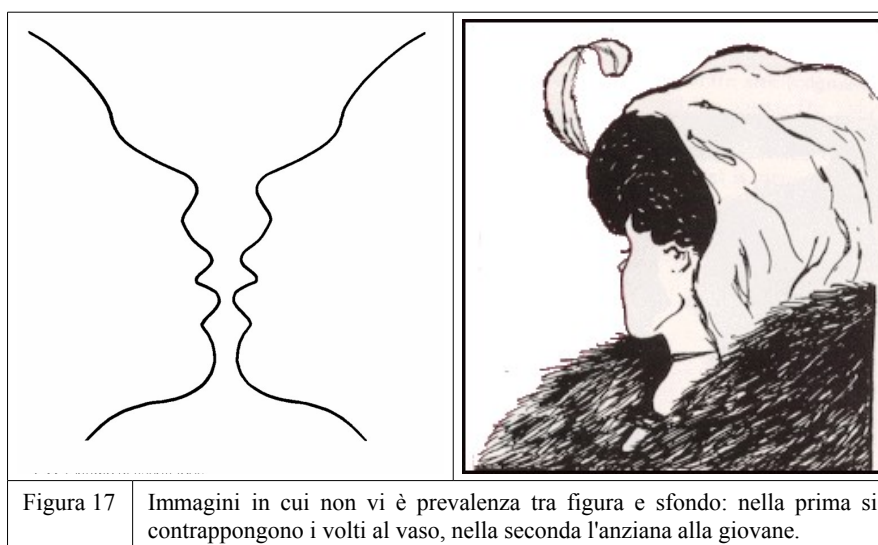
196 Ariana Fiorentini, Lamberto Maffei, *Arte e cervello*, Zanichelli, Bologna, 1995 p. 9

197 Ibidem

quelle aperte.

Con questi principi i punti, che di base hanno solo le caratteristiche di colore, grandezza e posizione, vengono organizzati in oggetti più complessi come linee, che hanno ulteriori caratteristiche quali lunghezza, orientamento e curvatura, confermando così il motto olistico¹⁹⁸.

La Gestalt ha formulato degli altri assunti per la separazione di una figura dallo sfondo. Ad esempio, a parità di altre condizioni, una figura verrà percepita come simmetrica agli assi orizzontali e verticali, e preferibilmente vista come superficie più piccola all'interno di una superficie più grande, identificata come sfondo. Il tutto è sintetizzato dal cosiddetto **principio della pregnanza**, secondo cui tra le varie organizzazioni geometricamente possibili prevale quella che possiede la forma "migliore, più semplice e più stabile"¹⁹⁹.



Talvolta si possono tuttavia costruire delle immagini in cui non vi è una chiara prevalenza figura-sfondo (fig. 17). In questi casi si alternano due percezioni a seconda di cosa si identifica come figura e di cosa si riconosce come sfondo²⁰⁰.

198 Stephen E. Palmer, *Vision Science - Photons to Phenomenology*, Op. cit. p. 51

199 Ariana Fiorentini, Lamberto Maffei, *Arte e cervello*, Zanichelli, Bologna, 1995 p. 10

200 Ibidem

Per la Gestalt la percezione è cogliere le caratteristiche essenziali dell'oggetto che forniscono informazioni sulla sua struttura globale, piuttosto che la sua completezza o esattezza. Percepire non è un'operazione d'inferenza probabilistica o un'ipotesi, come asserisce il costruttivismo, in quanto si fonda su leggi predeterminate che strutturano l'informazione sensoriale²⁰¹.

Per questo movimento l'empirismo non è alla base della percezione: ad esempio la capacità di organizzare gli oggetti non dipende dall'apprendimento per esperienza, ma sorge dall'interazione della struttura cerebrale con la struttura dello stimolo.

Se la contrapposizione tra empirico-innato è solo un argomento marginale, l'interesse dei gestaltisti per la relazione tra mente e cervello era così forte che postularono la teoria dell'**isomorfismo psicofisiologico**: le esperienze percettive (psicologiche) sono strutturalmente le stesse (isomorfiche) dei sottostanti eventi cerebrali (fisiologici).

Tuttavia l'isomorfismo psicofisico non è sufficiente per definire una teoria neurologica del funzionamento del cervello. Wolfgang Köhler approfondì l'argomento proponendo una prima spiegazione denominata **fisica della Gestalt**: un sistema fisico dinamico che converge verso uno stato di equilibrio di minima energia (portando la metafora delle bolle di sapone, indipendentemente dalla forma iniziale, evolvono sempre verso quella di una perfetta sfera). La seconda proposta, più specifica, fu che i meccanismi responsabili della percezione fossero dei campi elettromagnetici generati da eventi elettrici in milioni di neuroni. Questa spiegazione venne ritenuta plausibile ai tempi di Köhler, in quanto rispettava tutte le proprietà della fisica della Gestalt. In seguito, numerosi esperimenti dimostrarono che la distruzione dei campi elettrici non inibiva le capacità percettive. Questo portò ad un affievolimento del favore riconosciuto al movimento gestaltico, che comunque conosce un certo ritorno d'interesse nella veste delle teorie connettiviste dinamiche²⁰².

201 Ibidem

202 Stephen E. Palmer, *Vision Science - Photons to Phenomenology*, Op. cit. p. 51-53

La teoria ecologica di Gibson

Un'altra teoria classica della percezione visiva è quella elaborata da James J. Gibson dell'università di Cornell a New York.

Gibson accolse nella propria teoria l'enfasi sulle caratteristiche olistiche della percezione, tipiche del movimento della Gestalt. Egli riteneva, infatti, che la percezione degli oggetti non dovesse mai essere considerata separatamente dalla struttura dello sfondo su cui essi si proiettano. Per comprendere meglio la percezione di un oggetto, propose di analizzare la struttura dell'ambiente circostante l'oggetto stesso, che definì la “sua ecologia”²⁰³.

In effetti la teoria dell'ottica ecologica si basa più sulle informazioni percettive presenti nell'ambiente che sui meccanismi interni al cervello. Così facendo Gibson precede la moderna teoria computazionale.

Lo scopo della teoria ecologica era di specificare come le strutture del mondo (dovute alla riflessione della luce sulle superfici degli oggetti) illuminavano quello che egli definì matrice ottica ambientale (tutti i raggi che convergono in un determinato punto d'osservazione), permettendo così la percezione dell'ambiente attraverso il campionamento delle informazioni. In termini leggermente diversi, Gibson propose di individuare quali caratteristiche dello stimolo prossimo fornissero informazioni circa lo stimolo distale²⁰⁴.

Egli osservò che uno sfondo strutturato (una texture), come ad esempio un terreno o un prato, genera un gradiente di densità degli elementi che lo compongono (dimensione dei ciottoli di un selciato, grana di fili d'erba ecc.): gli elementi più vicini sono più distanziati di quelli più lontani. E' questo gradiente di densità che dà luogo alla percezione di una superficie inclinata che recede dall'osservatore. Un oggetto che si staglia contro uno sfondo di questo tipo non viene percepito isolatamente, come se fosse

203 Ibidem

204 Ibidem

sospeso nell'aria. Anzi, è proprio la parte di sfondo che l'oggetto nasconde che dà luogo alla percezione nella sua complessità, che dà informazioni sulle dimensioni dell'oggetto, sulla sua posizione ecc. Ad esempio, un albero verticale in un prato è visto come tale perché nasconde una parte del gradiente di densità del fondo, mentre la microstruttura dell'albero stesso (foglie, grana della corteccia dei rami ecc.) è costante²⁰⁵.

Si osservi che gli aspetti della scena visiva su cui Gibson concentrò la sua attenzione sono quelli che si mantengono invariati con il cambiare del punto di osservazione, come la presenza di un gradiente di densità nello sfondo, la verticalità dell'albero ecc. Per la teoria di Gibson è quindi cruciale considerare come si modifica la scena visiva durante il movimento dell'osservatore, e quali proprietà di questa scena rimangono invariate.

Le modificazioni complessive dell'intera scena visiva durante il movimento dell'osservatore consentono di ricavare, oltre agli elementi fissi/essenziali, le informazioni che definiscono la direzione del movimento dell'osservatore stesso²⁰⁶.

Gibson credeva quindi che nello stimolo retinico vi fossero sufficienti informazioni da permettere ad un organismo di percepire l'ambiente in modo non ambiguo e diretto, contrapponendosi anch'egli all'ipotesi dell'inferenza inconscia del costruttivismo (vedasi prossimo capitolo).

Sfortunatamente anche l'analisi di Gibson riesce difficilmente a ricostruire un mondo 3D da un'immagine 2D, anche con l'aggiunta della dimensione temporale dovuta al movimento. Il punto è che anche il tempo è una dimensione ambientale, e averla aggiunta al modello è ancora insufficiente per ricostruire un mondo che così possiede quattro dimensioni. Il problema della ricostruzione dell'ambiente presenta ancora troppe soluzioni possibili.

205 Ariana Fiorentini, Lamberto Maffei, *Arte e cervello*, Zanichelli, Bologna, 1995 p. 11

206 Ibidem

Costruttivismo

Il costruttivismo è un movimento che fonda le sue radici in vari rami del pensiero scientifico a noi contemporaneo: cibernetico, psicologico, operazionista²⁰⁷. È quindi, per natura, multidisciplinare. Esso incorpora in parte le teorie sulla visione appena esposte, e probabilmente per questo è oggi l'approccio più seguito.

I principali assunti del costruttivismo sono²⁰⁸:

- l'individuo partecipa attivamente alla costruzione della conoscenza;
- in ogni soggetto esiste una struttura cognitiva di base che dà una determinata forma all'esperienza;
- l'uomo viene visto come un sistema auto-organizzantesi che protegge e mantiene la propria integrità (autopoiesi).

Concetto fondamentale del costruttivismo è che la conoscenza umana, l'esperienza, l'adattamento, sono caratterizzati da una partecipazione attiva dell'individuo. Il reale non è considerato come qualcosa di oggettivo, indipendente dal soggetto che ne fa esperienza, ma è il soggetto stesso che lo crea, agendo attivamente per la sua costruzione. In sintesi: *“L'ambiente, così come noi lo percepiamo, è una nostra invenzione.”*²⁰⁹.

La possibilità di una conoscenza oggettiva è in dubbio. L'osservazione diretta dei fenomeni non è più considerata fonte privilegiata di conoscenza obiettiva. Quindi: *“Tutto ciò che è detto è detto da qualcuno.”*. Per inciso: non esistono fatti “nudi”, al di

207 <http://www.oikos.org/voncostrutt.htm> 20 dicembre 2006 di Ernst von Glasersfeld

208 http://it.wikipedia.org/wiki/Costruttivismo_%28psicologia%29 20 dicembre 2006

209 Ibidem

fuori delle teorie; ogni osservazione è ritenuta possibile solo alla luce di teorie, e nessuna conoscenza è data dall'ambiente, ma è sempre lo sviluppo di una conoscenza precedente²¹⁰.

Per comprendere il fenomeno della visione i costruttivisti partono dal fatto che il nostro sistema visivo riesce regolarmente a ricostruire l'ambiente 3D da un'immagine 2D in modo accettabile. Dato che i dati informativi, come considerati da Gibson, sono insufficienti, la domanda è: “Com'è possibile?”. L'ipotesi dei costruttivisti è che vi sia una sorgente informativa ulteriore rispetto l'immagine retinica, che aggiunge informazioni a quelle reperite con il sistema visivo, così da individuare la più probabile soluzione al problema di ricostruzione dell'ambiente 3D fra le infinite logicamente possibili. In un certo senso il sistema visivo ricrea l'immagine²¹¹.

Le teorie costruttiviste sulla visione si basano sull'estrazione di informazioni ambientali dai modelli di stimolo retinico che Gibson descrisse nella sua teoria ecologica. Viene anche riconosciuta l'importanza delle proprietà emergenti di linee, bordi, angoli, e delle figure intere, come proposto dai teorici della Gestalt.

La teoria costruttivista della visione può essere ricondotta agli studi di Hermann von Helmholtz²¹², un brillante fisico tedesco, nonché matematico e fisiologo, contemporaneo di Wundt e del strutturalismo. Molte sue idee di base sono state pubblicate originariamente nel testo “Handbuch der Physiologischen Optik” (1856-1867) (Manuale di fisiologia ottica), divenuto un punto di riferimento fondamentale per gli studi sulla visione dei colori.

Una delle sue idee che perdurano nel tempo e che risulta centrale nella teoria costruttivista consiste nell'assunto che la visione dipenda da un processo di *inferenza inconscia*. Al contrario di Gibson, Helmholtz riconobbe il salto logico che esiste tra le informazioni ottiche direttamente disponibili dalla retina e la conoscenza percettiva che

210 Ibidem

211 Stephen E. Palmer, *Vision Science - Photons to Phenomenology*, Op. cit. p. 56

212 <http://www.anisn.it/scuola/strumenti/visione/helmholtz.htm> 20 dicembre 2006

da esse deriva. Da qui propose, come appena accennato, che questo poteva essere superato attraverso delle “assunzioni” nascoste che venivano aggiunte alle immagini retiniche per ottenere delle “conclusioni” percettive relative all'ambiente. Queste assunzioni dovevano essere dei processi d'inferenza/deduttivi tali da trasformare gli insufficienti dati bidimensionali in una interpretazione percettiva corretta dell'ambiente 3D. Il processo deve essere inconscio in quanto, diversamente dai normali processi d'inferenza coinvolti nel pensiero e nella risoluzione di problemi, le persone non hanno consapevolezza di come o quando o perché o addirittura che stanno attuando inferenze visive.

È interessante chiedersi su quali basi si compiano queste inferenze. Helmholtz propose il **principio di verosimiglianza**: la visione arriva all'interpretazione che è più vicina allo stato dei fatti nel mondo esterno che, a sua volta, potrebbe aver causato lo stimolo retinico.

Il principio di verosimiglianza è spesso considerato antitetico al principio di pregnanza (spesso detto principio minimo) della Gestalt. Tuttavia non è semplice distinguere queste due teorie in quanto, ad esempio, somiglianza e semplicità sono spesso collegate tra loro.

Il processo che porta ad unire gli stimoli retinici, ma più in generale i dati sensori, con le inferenze e le deduzioni inconse, risulta avere natura **interpretativa euristica**: il sistema visivo perpetua delle inferenze circa le probabili condizioni ambientali che possono aver prodotto l'immagine. Il processo è euristico in quanto le regole probabilistiche usate sono di natura empirica, quindi approssimative, e perciò spesso, ma non sempre, vere. Se false, portano ad errate conclusioni che si manifestano come illusioni ottiche.

Le assunzioni nascoste messe in atto dal sistema visivo sono molte e articolate. Un esempio può essere che alcuni bordi ad ampia gradazione come soffitti, pavimenti, muri e simili siano allineati con la gravitazione verticale o orizzontale ad essi. Normalmente

quest'assunzione è esatta, e può servire per percepire correttamente gli oggetti che ci circondano. Se falsa, come in una stanza inclinata di un parco giochi, la nostra percezione è sorprendentemente storta, bieca. Questo succede perché tendiamo a ricondurre la situazione all'assunzione che in condizioni normali è corretta.

Un'interpretazione stretta o letterale delle inferenze inconsce porta a dire che il meccanismo della percezione si compie applicando in sequenza le regole della logica simbolica o risolvendo equazioni matematiche. Sembra che questi siano i tipi di processo supposti da Helmholtz a suo tempo. Oggi vi sono delle spiegazioni più plausibili provenienti dall'approccio computazionale. Ad esempio le reti connettive possono raggiungere conclusioni percettive basandosi in parte sui dati sensori entranti, in parte sulle assunzioni incorporate nel modello di interconnessioni tra le proprie unità neurali. Queste reti sono in grado d'effettuare inferenze sulla base di assunzioni euristiche, senza usare la logica simbolica o equazioni matematiche.

La Teoria dell'elaborazione dell'informazione²¹³

Il paradigma dell'elaborazione dell'informazione è un concetto che assimila il cervello umano a un algoritmo di calcolo. Tale modello è stato applicato con successo non solo nella percezione visiva, ma anche a svariati fenomeni cognitivi quali la percezione uditiva, la memoria, il linguaggio, il giudizio, il pensiero e la risoluzione di problemi. L'approccio dell'elaborazione dell'informazione domina così questi temi, tanto che alcuni suggeriscono che vada a formare un “paradigma di Kuhn” per la cognizione.

Per “paradigma di Kuhn” s'intende un dato insieme di assunzioni di lavoro che una certa comunità scientifica condivide nell'operare ricerche per un determinato argomento. Le assunzioni di un paradigma comprendono normalmente presunzioni teoriche o meta-teoriche sul modo di concepire gli argomenti principali e le metodi d'approccio.

Gli sviluppi delle teorie visive hanno sempre seguito l'evoluzione tecnologica. In questo

213 Stephen E. Palmer, *Vision Science - Photons to Phenomenology*, Op. cit. pp. 70-85

caso la spinta significativa è stata l'invenzione del computer elettronico.

La metafora del computer

L'invenzione del computer ha influenzato la ricerca teorica in due modi distinti, ma correlati tra loro:

1. Da una parte è divenuto lo strumento per testare le nuove supposizioni sui processi visivi e avere un riscontro su immagini reali. Questo ha fatto sorgere la computer vision come campo specifico di ricerca all'interno dell'informatica, o meglio scienza dei calcolatori. Il suo scopo è di programmare il calcolatore in modo che possa comprendere il mondo che lo circonda come fa l'uomo.
2. La seconda influenza è anche più rilevante: ha fornito l'analogia teorica tra processi mentali ed elaborazione dell'informazione. Per inciso: la mente è considerata come un programma che gira su una macchina chiamata cervello; la mente è il “software” e il cervello l’“hardware”.

Questo ha portato realmente a realizzare dei programmi basati su teorie percettive o, comunque, a descrivere queste teorie come dei processi di elaborazione.

Alcuni teorici hanno creduto che i fattori in comune ai binomi mente/cervello e programma/computer siano così estesi che un calcolatore adeguatamente programmato per vedere può realmente avere esperienze visive consce. Prospettiva denominata “AI forte” e contrapposta alla “AI debole”, in cui si ritiene che questo tipo di macchina stia solo simulando “processi” mentali, più o meno consci. Chiaramente è in corso un serio dibattito, con agguerrite e competenti fazioni.

Rimanendo nell'analogia tra mente/cervello e programma/computer, può risultare utile

esaminare quanto si conosce dell'elaborazione delle informazioni su strumenti come i computer al fine di comprendere come funziona la mente e rendere così più fruttuose le relative ricerche. A tal fine vengono utili le riflessioni di David Marr (1945-1980), che propose, nel suo libro, "Vision: A Computational Investigation into the Human Representation and Processing of Visual Information" pubblicato postumo nel 1982, tre livelli descrittivi per la comprensione di sistemi di elaborazione complessi: computazionale, algoritmico, e d'implementazione. Così facendo egli ha fornito un'analisi metateoretica (una teoria che indaga sulla natura delle teorie relative ad un argomento) al paradigma dell'elaborazione dell'informazione.

I tre livelli di Marr sono distinti e tutti necessari per comprendere il fenomeno:

1. Livello computazionale:

È il livello più astratto, definito come l'insieme di vincoli informativi disponibili per mappare/tracciare il percorso che parte con le informazioni in entrata e finisce con i risultati in uscita. Specifica, quindi, quali calcoli devono essere eseguiti e su quali informazioni, ma senza specificare come questi debbano essere realizzati. Un esempio, molto semplice, può essere un termostato per regolare la temperatura dell'abitazione collegato alla caldaia. Sopra o sotto determinati punti di soglia esso si spegne o accende.

2. Il livello algoritmico.

Esso è più specifico del precedente in quanto descrive come i calcoli sono effettuati in termini di operazioni del processo di elaborazione delle informazioni. Chiaramente vi possono essere modi diversi per ottenere lo stesso risultato, in questo senso il livello algoritmico corrisponde in modo più stretto al concetto di programma in campo informatico.

Per costruire l'algoritmo è necessario indicare una rappresentazione delle

informazioni in entrata e in uscita, e specificare l'insieme di processi da applicare sulle stesse.

3. Il livello implementativo.

Questo livello descrive come realmente un algoritmo è incorporato in un processo fisico all'interno di un sistema. Così come un programma può essere eseguito su computer diversi per costruzione, così lo stesso algoritmo può essere incorporato in strumenti fisicamente differenti. Per “strumenti fisicamente differenti” sono da intendersi sia vari strumenti fisici come computer, servo-meccanismi (es. il termostato già citato), ecc, ma anche il cervello (con particolare riferimento alla neuroanatomia e neurofisiologia dei meccanismi di percezione, cognizione e azione)²¹⁴.

Interessante è vedere come questi livelli descrittivi siano stati ripresi, con una prospettiva più psicologica, da Palmer e Kimchi che hanno cercato di individuare e analizzare le assunzioni implicite sottostanti alle teorie dell'elaborazione dell'informazione nella psicologia cognitiva. I livelli sono i seguenti:

1. Livello informativo.

La descrizione del funzionamento dei processi mentali può essere fatta in termini di eventi informativi, ognuno dei quali consiste di tre parti: l'informazione in entrata, l'operazione su questa e il risultato (l'informazione elaborata). La differenza con l'approccio di Marr sta nel fatto che qui ci si riferisce esplicitamente alla mente.

2. Decomposizione ricorsiva.

Questa assunzione implicita va a definire le teorie che possono rientrare nella macro area dell'elaborazione dell'informazione. In tale area non può infatti

214 <http://kybele.psych.cornell.edu/~edelman/marr/marr.html> 23 dicembre 2006

rientrare, ad esempio, la teoria di Gibson, che, pur descrivendo il funzionamento dei processi informativi, rifiuta esplicitamente di analizzare la loro rappresentazione ed esecuzione.

La decomposizione ricorsiva è essa stessa un processo che permette di scindere/specificare un evento informativo di livello superiore in un certo numero di eventi informativi che lo compongono stabilendo, inoltre, l'ordine temporale delle relazioni tra gli stessi. Il fatto che sia ricorsiva significa che può essere eseguita sui risultati fino ad ottenere degli elementi inscindibili, definiti primitive. Il problema verte sulla natura di queste; la proposta è di distinguere tra primitive “software” e “hardware”. Le prime sono un insieme di operazioni computazionali sufficienti ad eseguire un compito (livello logico). Le seconde sono quelle che riguardano operazioni che devono essere svolte da componenti fisici del sistema come i neuroni (livello fisico). È importante comprendere che per quanto si scomponga, non si arriva mai realmente al sistema fisico. Si rimane comunque a livello informativo. Questo è dovuto al fatto che l'informazione è per natura diversa dalla materia fisica o dall'energia, anche se necessita di queste per sussistere.

La principale differenza con il secondo livello di Marr è che quest'ultimo era inteso come una unità singola, a se stante, mentre qui è considerata come composta da molti livelli vicini tra loro e ordinati in via gerarchica.

L'efficacia di questo metodo sta nel fatto che permette di comprendere a poco a poco la complessità del sistema dell'elaborazione dell'informazione. Ogni successiva decomposizione rimuove la complessità implicitamente presente al livello precedente rendendo espliciti i suoi componenti.

Chiaramente questo funziona se il sistema ha una struttura gerarchica. Herbert Simon sostiene che questo sia vero per il sistema informativo umano, almeno per approssimazione. Egli parla di sistema “quasi scomponibile”. Per comprendere occorre distinguere tra le interazioni interne ai singoli componenti del sistema e

quelle che avvengono tra gli stessi. Un sistema scomponibile è caratterizzato da interazioni tra componenti trascurabili se comparate con quelle interne agli stessi. In un sistema quasi scomponibile le interazioni sono deboli ma non trascurabili. L'opinione che il sistema umano sia quasi scomponibile è evidentemente contrastata dal movimento della Gestalt, ma anche altri sollevano dubbi.

3. Incorporazione fisica (implementazione).

In un sistema fisico l'informazione è rappresentata da stati diversi del sistema, mentre i processi (le operazioni) che utilizzano queste informazioni sono posti in essere da variazioni di stato.

È il caso di riflettere sulle modalità d'implementazione o incorporazione fisica. I componenti fondamentali di un sistema computazionale sono la rappresentazione e i processi.

La rappresentazione si riferisce ad uno stato/posizione del sistema visivo che sostituisce/significa una proprietà dell'ambiente, un oggetto, un evento; è quindi un modello di ciò che rappresenta. Da questo punto di vista, la rappresentazione è solo una occorrenza di un più vasto sistema di rappresentazioni che include due mondi legati e distinti: il *mondo rappresentato* esterno al sistema e il *mondo rappresentante* interno al sistema (denominato rappresentazione interna o semplicemente rappresentazione)²¹⁵. La relazione tra i due mondi è data da una struttura simile. Spesso si tratta di una relazione di *isomorfismo*, una mappatura di un dominio nell'altro, così che le relazioni tra gli oggetti del primo siano riprodotte nel secondo. Più esplicitamente, l'isomorfismo sottintende una variazione causale fra i due mondi: una variazione nel mondo fisico (es: temperatura dell'aria) comporta una variazione nella rappresentazione (es: valori del termostato). Questo è importante sia per avere una rappresentazione sempre aggiornata, sia affinché questa sia attendibile e non accidentale.

215 Stephen E. Palmer, *Vision Science - Photons to Phenomenology*, Op. cit. p. 77

Oltre ad esservi una relazione causa-effetto, per avere una rappresentazione è necessario vi sia un processo che la utilizzi. Pena l'inutilità, non esistenza, della stessa.

I processi sono i componenti attivi di un sistema computazionale che trasformano o operano sulle informazioni cambiando una informazione nella successiva. Senza rappresentazioni non vi sarebbe nulla su cui lavorare, senza processi nulla sarebbe compiuto.

Alcuni dei principali processi riguardano l'esplicitazione delle informazioni presenti nelle rappresentazioni. Ossia, i processi hanno bisogno di una rappresentazione del mondo, che tuttavia di per sé non fornisce le necessarie informazioni sull'ambiente. Ciò che serve, nel caso di un'immagine retinica, sono i limiti tra le varie regioni retiniche proiettati da superfici diverse, dove queste superfici sono localizzate in uno spazio 3D; come sono sistemate per formare oggetti significativi, e come questi potrebbero essere disposti. Tutte informazioni non evidenti nella risposta retinica allo stimolo ambientale e che devono essere estratte dalla mente/cervello.

Come suggerito da Helmholtz l'estrazione delle informazioni avviene tramite un processo d'inferenza sul risultato dell'associazione delle informazioni esterne con quelle interne.

Il processo d'inferenza parte da delle premesse per arrivare a delle conclusioni. Tralasciando le differenze che vi possono essere tra l'inferenza logica e quella percettiva, consideriamo quale sia l'oggetto di quest'ultima e come possa avvenire.

Vi sono due tipi d'inferenza: deduttiva e induttiva. La prima si caratterizza per portare a delle conclusioni certe se sono vere le premesse (es: se “Tutti gli uomini sono mortali” e “Gianni è un uomo” abbiamo “Gianni è mortale”). Se le premesse sono false, le conclusioni non possono essere provate.

L'inferenza induttiva è intrinsecamente incerta e probabilistica anche se le premesse

sono vere. Un classico esempio di questo secondo tipo d'inferenza è la conclusione che le persone siano mortali. Normalmente non la si pensa una conclusione, in quanto la si usa come premessa a un sillogismo, ma su quali basi la si ritiene fondata? Non può essere dedotta dal fatto che molte persone sono morte, perché ci potrebbero essere delle persone viventi o che devono ancora nascere che non moriranno. La generalizzazione delle esperienze passate non assicura l'affidabile astrazione che tutte le persone siano mortali. L'induzione ha questo carattere d'incertezza, poiché si basa su fatti evidenti ma incompleti o probabilistici.

Molte inferenze del sistema visivo sono induttive, tuttavia possono essere ricondotte a deduttive adottando delle assunzioni nascoste. Alcuni ricercatori, come Rock e Gregory, suggeriscono che le assunzioni che portano ad una percezione veridica siano acquisite con l'apprendimento. Esse possono riguardare, ad esempio, l'effetto di convergenza di linee parallele all'orizzonte, alcuni vincoli ambientali come la gravità, ecc.

Normalmente più assunzioni sono coinvolte nella percezione della singola scena. Nella maggior parte dei casi esse concordano e portano a una conclusione comune; tuttavia vi sono casi in cui questo non accade. Si pone allora il problema d'individuare quelle corrette. Una soluzione è considerarle come *vincoli deboli*: informazioni restrittive che possono essere considerate ma anche sostituite da altre. La loro forza può variare, tuttavia anche le più conflittuali possono essere considerate per eventuali inferenze visive. Una possibile alternativa è utilizzare una *logica fuzzy*, introdotta da Lotfi Zadeh, piuttosto che una logica standard per modellare le inferenze. La logica fuzzy permette proposizioni che hanno diversi gradi di verità al posto dei normali valori “vero” e “falso”. Ulteriore possibilità è di trattare il problema come *inferenza probabilistica*, utilizzando il teorema di Baies. Il vantaggio di lavorare con questa struttura è il suo essere intrinsecamente probabile, e la possibilità che offre d'integrare in una singola struttura matematica diverse prove. Esso è inoltre compatibile con la logica fuzzy cosicché i due approcci possono essere utilizzati insieme.

Al di là della o delle procedure adottate, appare evidente la natura euristica di questi

processi, che possono portare, ma non sempre, alla corretta ricostruzione dell'ambiente. Inoltre, l'aggiunta di vincoli di supporto (assunzioni nascoste) sembra essere l'unico modo in cui la percezione può compiersi. Le informazioni aggiunte non devono però essere dichiarate esplicitamente come regole logiche, in forma di proposizioni, come nei sillogismi. Possono anche essere incorporate in un modello di connessioni in una rete neurale.

L'ulteriore aspetto da considerare è se i processi seguano un orientamento del tipo “bottom-up” o “top-down”. Considerando la retina come il livello base e le successive tappe del percorso visivo come livelli superiori ci si può chiedere quale sia l'incidenza di ognuno sul risultato finale. Se consideriamo più influenti i primi stadi avremo una elaborazione “data-driven” ossia guidata dai dati. Al contrario, se sono i livelli superiori a correggere, in base alle assunzioni nascoste, le informazioni retiniche, l'elaborazione sarà del tipo “hypothesis-driven” o “expectation driven processing”. Intuitivamente si ritiene che la visione sia di tipo bottom-up, ma ciò si è rivelato vero solo per i primi stadi, e non per l'intero processo. Come affermato in precedenza, sia la struttura anatomica del cervello (che prevede collegamenti del tipo bi-direzionale), sia il fatto che la percezione del presente porti a delle aspettative future, al fine di poter agire in tempo, implicano una componente del sistema di tipo top-down. Ad esempio si è constatato che la rilevazione delle lettere è strettamente legata al fatto che queste appartengano a parole conosciute o insignificanti. Tuttavia non è ancora chiaro dove inizi l'influenza tra i due processi.

Dopo questa esposizione teorica dei vari approcci alla visione è opportuno porre, prima di analizzare più nel dettaglio l'aspetto computazionale, una riflessione sul linguaggio visivo.

Il linguaggio visivo²¹⁶

Questo capitolo ha un carattere interlocutorio: da una parte è utile per richiamare e sintetizzare alcuni concetti presentati, dall'altra per porre alcune riflessioni e introdurre il capitolo centrale riguardante le tecniche utilizzate per la computer vision.

La prima riflessione riguarda la possibilità di considerare il sistema visivo come uno strumento per “intendere” un linguaggio, normalmente definito visivo e, prima ancora, se questo linguaggio esiste come tale, sia cioè un sistema di comunicazione con una propria struttura o grammatica²¹⁷.

A tal fine è utile considerare il concetto sopraesposto di rappresentazione e vedere come quest'ultima sia presente fin dalla comparsa dell'uomo. Infatti i cacciatori del paleolitico²¹⁸ ci hanno tramandato le loro incisioni sulle pareti delle caverne²¹⁹ rappresentanti spesso gli animali da loro visti o uccisi.

Possono sorgere alcuni interrogativi: come fecero i loro compagni a comprendere che quelle linee rappresentavano un bisonte piuttosto che un altro animale? Come fece il disegnatore a spiegare, in assenza del linguaggio, a cosa si riferisse il disegno? E, più in generale, come possono delle linee richiamare alla memoria visiva dell'osservatore delle figure complesse? L'immagine percepita è molto diversa dal profilo che si può tracciare di un animale, o di un albero, ecc. In aggiunta le immagini dell'animale e quelle dei disegni attivano cellule retiniche diverse, e certamente stimoleranno zone cerebrali diverse. Questi stimoli, sebbene diversi, devono rievocare lo stesso “oggetto” dalla memoria. Lo stesso risultato si può raggiungere richiamando alla mente rumori e suoni tipici di un animale, oppure riproducendone la forma o le azioni con movimenti di mani e corpo. Oggi, con l'uso del linguaggio verbale e/o della scrittura, in forza delle loro

216 Cfr. Ariana Fiorentini, Lamberto Maffei, *Arte e cervello*, Zanichelli, Bologna, 1995 pp. 46-48

217 Cfr. Giorgio Graffi, Segio Scalise, *Le lingue e il linguaggio - Introduzione alla linguistica*, Bologna, Il Mulino, 2003 p. 16

218 Letteralmente età della pietra antica che va da circa 2,5 milioni di anni (in Africa) a 10 mila anni fa. <http://es.wikipedia.org/wiki/Paleol%C3%ADtico> 28 dicembre 2006

219 Tra le più conosciute quelle di Altamira in Spagna. <http://www.artepreistorica.it/links/altamira.asp> 28 dicembre 2006

proprietà simboliche e di astrazione, è sufficiente una parola, come «bisonte», per richiamare alla memoria l'oggetto-immagine e le relative caratteristiche distintive.

Il segno, o meglio il contorno, è quindi uno stimolo particolarmente efficace per il cervello. Pensare che le immagini degli oggetti del mondo esterno siano fisicamente ridisegnate e dipinte con le loro forme e colori nella nostra corteccia appare, oggi, assurdo. Esse, piuttosto, sono ridotte ad invarianti, a modelli visivi, e memorizzate come simboli dall'attività di alcune zone neurali.

Come accennato sono stati David Hubel e Torsten Wiesel a fornire la dimostrazione scientifica che i neuroni della corteccia visiva reagiscono, in modo preferenziale, al contorno degli oggetti del mondo esterno. Per questo lavoro fu loro dato nel 1981 il premio Nobel. I loro esperimenti mostrarono che le cellule neurali della corteccia sono reattivi solo a stimoli visivi rappresentati da linee o bordi con particolare orientamento e dimensioni. È molto probabilmente, a livello della corteccia, le immagini che incidono sulla retina, sono ridotte ai propri contorni o a segmenti di essi e, nell'ottica del costruttivismo, sono ricostruite da una rielaborazione di questa informazione. Si può supporre che siano queste informazioni a costituire la grammatica del linguaggio visivo.

Nel passaggio dalla retina alla corteccia il linguaggio dei neuroni evolve. Se nella retina i neuroni «vedono» l'immagine come un insieme di punti, nella corteccia i neuroni «vedono» i contorni dell'immagine.

Attualmente l'esempio più chiaro di come il cervello può simbolizzare l'informazione proveniente dai sensi è dato dalle risposte della corteccia visiva ai contorni degli oggetti, più che ad altre loro caratteristiche o attributi.

Non a caso, Ariana Fiorentini e Lamberto Maffei affermano: “Sembra quindi che i segni siano un linguaggio primitivo proprio del nostro sistema nervoso, una caratteristica che origina dalle proprietà fisiche del cervello: dalle sue connessioni, dalle sue intrinseche caratteristiche anatomiche e funzionali. Queste proprietà sono, in gran parte, innate,

anche se possono essere modificate o perfezionate con l'esperienza. Sono proprietà comuni a tutti gli uomini, e costituiscono quindi, come le parole della lingua di un popolo, una base/patrimonio per comunicare.²²⁰

I segni tracciati dal disegnatore paleolitico esprimevano quindi il vocabolario grafico del cervello, comune agli altri esseri umani del suo tempo. Le immagini erano state da loro acquisite e tradotte attraverso lo stesso linguaggio, basato sugli stessi simboli, ossia, essenzialmente dei contorni. La memoria riconduce questi ricordi alla percezione, essi si presentano del tutto simili a quelli percepiti con l'esperienza sensoriale e, quindi, facilmente riconoscibili.

Forse le origini del linguaggio segnico sono antiche quanto quelle del linguaggio della parola, collegabili alla capacità dell'uomo di emettere suoni e rumori. Secondo Noam Chomsky e altri linguisti, la struttura del linguaggio deriva dalla struttura del cervello umano. Chomsky sostiene che solo in parte il linguaggio viene appreso con l'esperienza, in quanto tutte le lingue hanno una "grammatica mentale" comune che è propria del cervello dell'uomo e che viene geneticamente ereditata. Chomsky porta a sostegno della sua teoria di come sia possibile capire frasi mai sentite prima o che si riferiscono ad eventi completamente estranei a chi ascolta. È sufficiente che la frase rispetti una determinata struttura grammaticale necessaria per legare tra loro nomi, aggettivi e verbi.

La comprensione del linguaggio avviene in quanto vi sono delle regole comuni a tutte le lingue per collegare le parole. Regole che Chomsky attribuisce a proprietà della struttura cerebrale.

Per il linguaggio parlato l'esistenza di questa proprietà è ipotetica; mentre per la visione sono disponibili prove sperimentali a supporto dell'esistenza di un linguaggio visivo neurale. Il fatto che i neuroni della corteccia visiva preferiscano certe forme piuttosto che altre potrebbe far supporre l'esistenza di un equivalente della grammatica mentale postulata da Chomsky per il linguaggio.

220 Ariana Fiorentini, Lamberto Maffei, *Arte e cervello*, Op. Cit. p. 48

Con la teoria computazionale, si è visto come questa struttura possa essere ricondotta, attraverso i passaggi proposti prima da Marr e poi da Palmer e Kimchi, a proposizioni logiche implementabili in un computer, anche se vi sono ancora notevoli difficoltà dovute sia a limiti “hardware”, sia di know-how. Ciò che segue è quindi il tentativo di presentare quanto si conosce finora e di prevedere le possibili evoluzioni.

Cos'è la Computer Vision?

Definizioni.

Rispondere a questa domanda è meno semplice di quanto si possa pensare. Dare una definizione a una materia in continua evoluzione, che presenta aspetti variegati e permette approcci diversificati, è a dir poco arduo. Perciò è meglio riportarne più d'una, nella speranza di trasmettere non solo delle spiegazioni formali, ma un'intuizione. In effetti, comunque, una definizione traccia un limite di separazione tra ciò che è compreso e ciò che viene escluso ma ciò, specie durante la fase di ricerca, può risultare più uno svantaggio che un vantaggio.

È utile fare una premessa di carattere terminologico: i termini “computer vision” e “machine vision” vengono usati spesso con lo stesso senso, anche se il secondo si riferisce più ad applicazioni industriali; il termine viene tradotto in italiano con “visione artificiale”. La preferenza da me accordata a computer vision è in ragione della sua diffusione in ambito della comunità scientifica internazionale e, in secondo luogo, poiché si riferisce alla generalità della materia²²¹.

David Forsyth e Jean Ponce, affermano: “Noi vediamo la computer vision – o semplicemente 'vision', scusandoci con chi studia la visione umana o animale – come un'impresa che usa metodi statistici per districarsi nell'immensità di dati utilizzando dei modelli costruiti con l'aiuto della geometria, della fisica e delle teorie dell'apprendimento. Perciò, dal nostro punto di vista, la visione si basa sulla comprensione dei principi della camera oscura²²² e dei processi fisici della formazione dell'immagine, per ottenere semplici inferenze dai valori di singoli pixel, combinare l'informazione disponibile da molte immagini in un'unità coerente, imporre ordine su gruppi di pixel per separarli gli uni dagli altri o inferire informazioni sulla forma, e

221 Cfr Linda G. Shapiro, George C. Stockman, *Computer Vision*, New Jersey, Prentice Hall, 2001 p. 1 si veda anche http://it.wikipedia.org/wiki/Computer_vision 20 gennaio 2007

222 O, comunque, di tutti gli strumenti in grado di acquisire delle immagini: macchine fotografiche, cineprese, telecamere, scanner, ecc (n.d.t.)

riconoscere oggetti usando informazioni geometriche o tecniche probabilistiche.”²²³.

Linda G. Shapiro e George C. Stockman la definiscono come “il campo dove informazioni significative devono essere ricavate/estratte automaticamente dalle immagini²²⁴” con l'obiettivo di “prendere decisioni utili riguardanti oggetti fisici reali e scene basate sulle immagini percepite”²²⁵.

Emanuele Trucco e Alessandro Verri preferiscono porsi delle domande più precise: “Che tipo di problemi stiamo affrontando? E come programmiamo di risolverli?”. Il problema dovrebbe essere “ricavare/elaborare le proprietà di un mondo 3D da una o più immagini digitali”, basandosi su proprietà geometriche (es.: forma e posizione di oggetti solidi) e dinamiche (es.: la velocità degli oggetti). Per loro, essenzialmente, il computer dovrebbe interpretare le immagini²²⁶.

Ramesh Jain, Rangachar Kasturi e Brian G. Schunck riprendono il concetto “*A machine vision system recovers useful information about a scene from its two-dimensional projections*”²²⁷ ma poi aggiungono che “Decidere richiede sempre la conoscenza dell'applicazione o dello scopo. Come vedremo, a ogni stadio in *machine vision* le decisioni devono essere prese dal sistema. L'enfasi sui sistemi di *machine vision* si è massimizzato sull'automazione delle operazioni di ogni livello, e questi sistemi dovrebbero utilizzare la conoscenza per riuscirci. Questa conoscenza include modelli delle caratteristiche, della formazione dell'immagine, modelli di oggetti, e le relazioni fra oggetti.”²²⁸.

Avvincente è quanto riportato da Microsoft sulla sezione del sito dedicato alla ricerca sulle tecnologie visive: “Lo scopo della ricerca sulla computer vision è di dotare il computer dell'abilità di comprendere le immagini ferme e in movimento. Anche se noi,

223 Cfr David A. Forsyth, Jean Ponce, *Computer Vision: A Modern Approach*, Op. Cit., p. xvii

224 Cfr Linda G. Shapiro, George C. Stockman, *Computer Vision*, Op. Cit., p. xvii

225 Ibidem p.1

226 Emanuele Trucco, Alessandro Verri, *Introductory Techniques for 3-D Computer Vision*, Op. Cit, p. 1

227 Ramesh Jain, Rangachar Kasturi, Brian G. Schunck, *Machine Vision*, Op. Cit., p. 1

228 Ibidem pp. 5-6

come esseri umani, possiamo dare un senso alle fotografie e ai video, per un computer esse sono solo una matrice di numeri rappresentante la luminanza e il colore di un pixel. Come possiamo ottenere da questa matrice di numeri la comprensione che c'è una ragazza che sta giocando a pallone davanti all'edificio? Quanto alto sta gettando la palla? Qual'è la struttura di fondo della casa? Questi sono gli argomenti che ci interessano!”²²⁹.

Roberto Marangoni e Marco Geddo, da cui ho preso spunto, affermano: “Possiamo ora pensare alla nostra immagine non più come a una collezione di pixel tra loro scorrelati, ma come a un insieme di *oggetti* ciascuno con proprietà *misurabili* quali per esempio la lunghezza, l'area, il livello medio di luminosità, ecc.

Questo schema di rappresentazione costituisce il primo passo verso la “Computer Vision” (visione artificiale)[...].

L'immagine deve quindi esser pensata come un messaggio scritto in un linguaggio nel quale gli oggetti identificati sono le parole e le relazioni sono i connettivi.”²³⁰.

Ognuna di queste definizioni ha degli aspetti in comune con le altre, e ognuna focalizza un aspetto particolare della materia. Esse sono comunque, in qualche modo incomplete, in quanto si limitano a descrivere un procedimento (ricavare/estrarre informazioni dalle immagini), o uno scopo (prendere decisioni utili riguardanti oggetti fisici reali e scene basate sulle immagini percepite). Possono solo descrivere in quanto si basano su un linguaggio che possiamo definire, sulle orme di Francesco Antinucci, logico-simbolico. A questo linguaggio se ne può contrapporre uno di tipo percettivo-motorio, ossia non basato sull'utilizzo di simboli, astrazioni e di una logica sottostante, ma sul coinvolgimento dei sensi percettivi e sul movimento.

Utilizzare il linguaggio percettivo-motorio per definire la computer vision appare difficile, eppure, forse, il cinema di fantascienza l'ha già fatto. Ad esempio, nel film

229 http://research.microsoft.com/vision/#research_groups 30 dicembre 2006

230 Roberto Marangoni, Marco Geddo, *Le Immagini digitali*, Op. cit., pp. 74-75

“Terminator” di James Cameron²³¹, divenuto un “cult” per gli appassionati del genere. Ciò che qui interessa non sono l'ambientazione o gli artifici usati dal regista per creare suspense e paura o reazioni emotive, ma quelli utilizzati per rendere credibile l'interpretazione di Schwarzenegger-Terminator. Dopo ben 35 minuti e 42 secondi, durante i quali si vede Schwarzenegger coinvolto in fenomeni quantomeno anomali (l'arrivo in questo tempo con lampi d'energia, essere colpito a morte e rialzarsi, ecc) e compiere azioni violente (l'uccisione a sangue freddo di più persone), Cameron suggerisce la vera natura del personaggio: un robot. L'aspetto interessante e però costituito dal “come” lo fa, e cioè con una soggettiva che porta lo spettatore ad identificarsi con il Terminator, a vedere come il Terminator.

Il regista utilizza la soggettiva per cinque volte nell'arco del film. Esaminando nel dettaglio le relative scene, si possono cogliere i messaggi impliciti utili all'analisi:

- la prima soggettiva s'incontra nella scena in cui il Terminator insegue i due personaggi principali, appena incontrati nella discoteca. Con essa Cameron fa capire che il Terminator è un robot, ponendo così le basi per l'attendibilità delle dichiarazioni di Michael Biehn, l'interprete di Kyle Reese (l'altro essere venuto dal futuro);
- nella scena girata alla stazione di polizia, la soggettiva riafferma la natura inumana della macchina e, a seguire, la sua presunta superiorità nei confronti dell'uomo. Da notare che il robot elimina la luce prodotta dall'impianto elettrico in quanto a lui non necessaria;
- nella stanza affittata all'hotel, la soggettiva mostra uno schermo dove appare un menù di scelta tra risposte alternative usate per rispondere al proprietario;
- nel residence dove gli altri due personaggi hanno trascorso la notte l'uso

231 “*The Terminator*”, di James Cameron, USA, 1984.

della soggettiva non appare del tutto motivato, se non per indicare l'indifferenza della macchina nei confronti degli esseri viventi (si pensi al cane e all'altro inquilino);

- infine, lo spettatore si immedesima nel Terminator quando questi sale sul camion per innestare la marcia. In tale occasione, il regista voleva probabilmente sottolineare la differenza di questa macchina rispetto alle altre di cui si è soliti servirsi. Questo robot è in grado di apprendere il funzionamento di altre macchine, quindi di utilizzarle.

La soggettiva “inumana” si percepisce dal fatto che la visione è resa in rosso e nero, quasi fosse a infrarossi. A questa sono aggiunte, di volta in volta, delle video-grafie riprodotte simboli di puntamento, o coordinate incomprensibili (almeno per gli umani; in effetti si tratta di un linguaggio numerico, tipico delle macchine), o la riproduzione del meccanismo del cambio... insomma degli elementi che “svelano” come la macchina “pensa”.

Purtroppo in questo elaborato non si può trascendere dal linguaggio logico-simbolico, ma si spera che la memoria sia di aiuto per richiamare le scene sopra menzionate.

Come si può notare, l'opera di Cameron richiama più o meno esplicitamente alcune delle definizioni esposte dando, nel contempo, sia la misura di come la visione possa essere ritenuta una funzione semplice, ma si riveli un fenomeno complesso. Il film mostra quelle che sono le fasi del processo visivo, ossia l'acquisizione di un'immagine e la sua interpretazione/elaborazione al fine di agire.

Legami tra tecnologia e arte.

Certamente vi sono altre opere, più o meno conosciute, che potrebbero essere portate ad esempio per definire/trasmettere l'essenza della computer vision. Questa considerazione permette di riprendere l'argomento della supposta reciproca influenza tra arte e tecnologia. In via esplicita corre l'obbligo di citare, ancora una volta, Asimov: “Non si

deve, tuttavia, sottovalutare interamente l'influenza dei racconti di fantascienza. All'inizio degli anni '50, uno studente della Columbia University, Joseph F. Engelberger, lesse *Io, Robot* e come risultato, fu contagiato da una passione, che durò tutta la vita, per il lavoro sui robot.

Nel 1956 Engelberger incontrò George C. Devol Jr., che, due anni prima, aveva ottenuto il primo brevetto per un robot industriale. Egli chiamò il suo sistema di controllo e di memoria tramite calcolatore *universal automation* (automazione universale) o, abbreviato, *unimation*²³².

Engelberger e Devol fondarono insieme la Unimation Inc. e, in seguito, Devol ottenne da trenta a quaranta brevetti. [...]

Fu solo con la comparsa dei <<microchip>> che i robot progettati dalla Unimation acquistarono un interesse commerciale. Ben presto la Unimation divenne la più importante e la più redditizia fabbrica di robot nel mondo.

Iniziò così l'era del *robot industriale*. [...]²³³.

Discipline correlate.

È opportuno, dopo aver considerato il rapporto tra arte e tecnologia e, in precedenza, le relazioni tra biologia, psicologia e computer vision, esaminare gli stretti collegamenti tra quest'ultima e altre discipline informatiche che, comunque, si differenziano in alcuni aspetti. Ecco le principali:

5. Elaborazione dell'immagine.

Lo scopo principale di questa materia è la trasformazione di immagini in altre immagini; il recupero delle informazioni è lasciato all'uomo. Quest'area

232 Per una cronologia degli avvenimenti più importanti si veda http://trueforce.com/Articles/Robot_History.htm 2 gennaio 2006.

233 Isaac Asimov, *Il libro di Biologia*, Op. cit. pp. 386-387

s'interessa del miglioramento dell'immagini, degli algoritmi di compressione, della correzione delle immagini sfuocate. Da sottolineare, invece, che la computer vision si occupa del recupero delle informazioni in modo automatico, o con un minimo intervento umano: quindi essenzialmente della "comprensione" dell'immagine. Gli algoritmi utilizzati per l'elaborazione delle immagini vengono utilizzati nella computer vision per evidenziare particolari informazioni ed eliminare il rumore²³⁴.

6. Computer grafica.

Si occupa della creazione di immagini partendo da primitive geometriche quali linee, cerchi o superfici a forma libera. Queste tecniche giocano un ruolo significativo nella visualizzazione e nella realtà virtuale. La computer vision ha come scopo il problema inverso: misurare/stimare le primitive geometriche e altre caratteristiche dall'immagine. La computer grafica è quindi la sintesi delle immagini, mentre la computer vision ne rappresenta l'analisi. In passato queste due aree non erano molto collegate, ma ultimamente lo sono sempre più. La computer vision utilizza la rappresentazione di curve e superfici, e diverse altre tecniche della computer grafica. Viceversa, quest'ultima utilizza diverse tecniche della computer vision per fornire al computer i modelli necessari alla creazione di immagini realistiche. I due campi sono resi contigui dalla visualizzazione e dalla realtà virtuale²³⁵.

7. Pattern recognition (riconoscimento di modelli).

Questa disciplina si occupa della classificazione di dati numerici e simbolici. Molte tecniche statistiche e sintattiche sono state sviluppate per la classificazione dei modelli. Esse sono importanti nella computer vision per il riconoscimento degli oggetti. Infatti, molte applicazioni industriali si basano sulla *pattern recognition*. Ma la computer vision normalmente richiede ulteriori

234 Cfr. Ramesh Jain, Rangachar Kasturi, Brian G. Schunck, *Machine Vision*, Op. Cit., p. 4

235 Ibidem

tecniche²³⁶. Infatti molti metodi sviluppati in passato erano adatti per oggetti 2D o 3D posizionati secondo alcuni vincoli, ma del tutto insoddisfacenti per un ambiente 3D generico²³⁷.

8. Intelligenza artificiale.

Si occupa della progettazione di sistemi intelligenti e dello studio degli aspetti computazionali dell'intelligenza. L'intelligenza artificiale è utilizzata, nella fase successiva all'estrazione delle caratteristiche dall'immagine, per l'analisi delle scene da seguirsi attraverso l'elaborazione di rappresentazioni simboliche dei contenuti delle stesse. Si possono considerare tre livelli del processo di visione posto in essere dall'intelligenza artificiale: percezione, cognizione e azione. La percezione traduce i segnali esterni in simboli, la cognizione manipola i simboli, l'azione traduce i simboli in segnali che portano dei cambiamenti all'esterno (nel mondo). Molte tecniche dell'intelligenza artificiale svolgono un ruolo importante in vari aspetti della computer vision: infatti quest'ultima viene spesso considerata un sottoinsieme della prima.

La progettazione e l'analisi di reti neurali sono campi in continuo fermento. L'impiego di reti neurali per risolvere problemi relativi alla visione artificiale cresce costantemente. Non vi sono però tecniche consolidate applicabili alla computer vision, ragion per cui se ne farà solo limitato cenno nel proseguo della tesi²³⁸.

9. Fotogrammetria.

Si occupa di ricavare misurazioni affidabili e accurate da immagini remote. Questa disciplina ha un collegamento più flebile con la computer vision rispetto

236 Ibidem p. 4-5

237 Cfr. Emanuele Trucco, Alessandro Verri, *Introductory Techniques for 3-D Computer Vision*, Op. Cit., p. 3

238 Cfr. Ramesh Jain, Rangachar Kasturi, Brian G. Schunck, *Machine Vision*, Op. Cit., p. 5

alle precedenti. La fotogrammetria si differenzia dalla computer vision per gli alti livelli di accuratezza ricercati. Inoltre, non tutta la computer vision si occupa di misurazione.

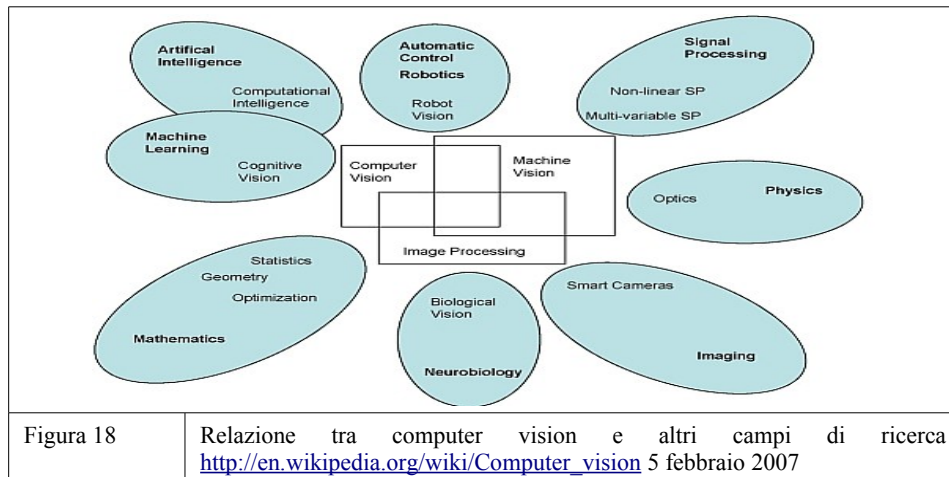


Figura 18

Relazione tra computer vision e altri campi di ricerca
http://en.wikipedia.org/wiki/Computer_vision 5 febbraio 2007

Le immagini digitali

Questo paragrafo ha lo stesso titolo del testo di Marangoni e Geddo, ma non ne vuole essere un omaggio. Esso è invece la base d'obbligo della computer vision e di tutte le operazioni sulle immagini che si possono eseguire con il calcolatore.

Sono necessarie alcune considerazioni, già in precedenza esposte, ma che è utile richiamare: le immagini rivestono un ruolo centrale in gran parte delle attività umane. La comunicazione delle informazioni e la trasmissione di concetti attraverso le immagini sono universalmente presenti nell'umano agire. Ecco le due ragioni essenziali:

- l'uomo usa la vista per osservare e analizzare l'ambiente che lo circonda: tende quindi a rappresentare per immagini ciò che vede;
- l'immagine sintetizza una grande quantità di informazioni.

Si pensi alla cartina di una città (o una mappa qualsiasi): è un'immagine che mostra in modo semplice, completo e sintetico le informazioni relative alla rete stradale: le parole richiederebbero più tempo e spazio²³⁹.

E' possibile introdurre una distinzione tra immagini reali e artificiali/sintetiche. Nelle prime rientrano quelle ottenute attraverso strumenti come macchine fotografiche, telecamere, scanner..... insomma strumenti di acquisizione di una scena reale. Le seconde sono quelle costruite dall'uomo, principalmente attraverso il calcolatore, che non necessariamente hanno un corrispettivo reale.

Cos'è un'immagine?

Nel capitolo dedicato alla luce e nel paragrafo "Formazione delle immagini" sono stati esposti alcuni processi fisici e geometrici coinvolti con la sua formazione. Ma, essenzialmente, cos'è un'immagine? In generale, è possibile definire un'immagine come

239 Cfr. Roberto Marangoni, Marco Geddo, *Le Immagini digitali*, Op. cit., pp. 1-2

“un'area con una determinata distribuzione di colori”²⁴⁰. In tale definizione è possibile riconoscere sia le immagini retiniche (o in generale quelle attribuibili a esseri viventi), sia le immagini ottenute con strumenti analogici (es: quadri, le classiche foto, i film su pellicola, ecc) nonché le immagini digitali (es: immagini digitali sintetiche, fotografie digitali, ecc).

Cosa sono le immagini digitali?

Che cosa distingue un'immagine digitale dalle altre? Tautologicamente si potrebbe dire: è digitale! Ma cosa significa?

Con digitale si vuole esprimere il concetto di un qualcosa che può essere compreso da un calcolatore, cioè una rappresentazione numerica²⁴¹. Digitale deriva da digit che in inglese significa cifra; a sua volta digit deriva dal latino digitus che significa dito. In definitiva, è digitale ciò che è rappresentato con i numeri che si contano, appunto, con le dita²⁴².

Per immagine digitale s'intende l'immagine che è stata **digitalizzata**, ossia rappresentata in numeri. Come avviene questo processo? Il primo passo è dividere il soggetto in unità distinte. L'immagine è divisa in una griglia di elementi della figura, detti anche pixel (contrazione della locuzione inglese picture element); ne si fa, tecnicamente, un *campionamento spaziale*. Il dettaglio raggiungibile e la complessità della griglia (ossia la sua *risoluzione*) variano a seconda di quanto è sofisticato il sistema di acquisizione.

Successivamente avviene la *quantizzazione cromatica*, ossia viene assegnato a ciascun pixel un valore numerico relativo che ne rappresenta il colore medio²⁴³.

Queste matrici numeriche bidimensionali possono a loro volta dare luogo a due tipi

240 Ibidem p. 4

241 Lev Manovic, *Il linguaggio dei nuovi media*, trad. it. Roberto Merlini, Milano, Fres, 2004 p. 46 e seguenti, (ed. originale *The Language of New Media*, 2001)

242 [http://it.wikipedia.org/wiki/Digitale_\(informatica\)](http://it.wikipedia.org/wiki/Digitale_(informatica)) 4 gennaio 2007

243 Cfr. Roberto Marangoni, Marco Geddo, *Le Immagini digitali*, Op. cit., p. 5 e Ben Long, *Fotografia Digitale. Il Manuale. 3A ed.*, trad. it. Riccardo Mori, Milano, Apogeo, 2005, pp. 6-7, (ed. originale *Complete Digital Photography*, 2002)

d'immagine:

4. Immagini a **modulazione d'intensità o luminanza**: le normali fotografie che codificano l'intensità della luce, acquisite tramite i normali sensori alla luce (misurano la quantità di luce che si imprime sul sensore).
5. Immagini **spaziali**: codificano la forma e la distanza (stimano direttamente le strutture 3D della scena osservata attraverso varie tecniche). Sono ottenute con l'utilizzo di sonar o scanner laser.

A seconda della natura dell'immagine i numeri possono quindi rappresentare valori diversi, quali l'intensità della luce, le distanze o altre quantità fisiche. La prima considerazione che si può trarre è che la relazione tra l'immagine e il mondo rappresentato dipende dal processo di acquisizione e quindi dal sensore utilizzato. La seconda considerazione riguarda il fatto che ogni informazione contenuta nell'immagine deve essere ricavata da una matrice numerica²⁴⁴.

Quest'ultima affermazione mi consente d'introdurre un'ulteriore distinzione tra le immagini digitali, che potremmo ricondurre a differenze logiche o, meglio, di memorizzazione. Possiamo infatti distinguere tra immagini **raster/bitmap** e immagini **vettoriali**. Le prime, definite da una griglia (mappa di bit,) sono quelle più utilizzate. Le seconde sono definite quasi per contrapposizione alle prime, ossia dall'assenza della mappa e si caratterizzano per essere formate da curve o tracciati definiti da entità matematiche dette vettori²⁴⁵, insomma da *primitive geometriche* che rappresentano l'immagine²⁴⁶. Se è semplice passare da un'immagine vettoriale ad una bitmap attraverso un processo di “rasterizzazione” (d'obbligo in fase di stampa e per la visualizzazione a schermo), il processo inverso è così arduo da essere ritenuto “praticamente impossibile”²⁴⁷. Questo tipo di problematica è simile a quella già discussa nel paragrafo

244 Cfr. Emanuele Trucco, Alessandro Verri, *Introductory Techniques for 3-D Computer Vision*, Op. Cit, p. 16

245 Cfr. Pizzirani A., *Teoria e tecnica di elaborazione delle immagini*, pp. 14-17

246 Cfr. <http://www.dmi.unict.it/~gdibiasi/didattica/ixg/lezione1.pdf> p. 104 7 gennaio 2007

247 Cfr. Marco Pedroni, Giorgio Poletti, *U.D. Fondamenti di Informatica*, Op. Cit, p. 137

“Formazione delle immagini”, ossia la ricostruzione da un'immagine 2D presente nella retina all'ambiente 3D che l'ha generata: questa può essere un'applicazione della computer vision²⁴⁸.

Sempre per quanto riguarda la memorizzazione/archiviazione delle immagini è da tener presente che nel corso del tempo si sono proposti e affermati vari formati (gif, jpg, png, ecc), diversificati per meccanismi di compressione e tipologia di informazioni considerate.

Alcune nozioni per l'acquisizione

Immagini a modulazione d'intensità/luminanza

È necessario richiamare alcuni concetti relativi alla luce e alla formazione dell'immagine approfondendo, brevemente, l'aspetto computazionale.

Considerando l'immagine come un'area con una determinata distribuzione di colori, è interessante comprendere come questi vengano disposti al fine di fornire un significato²⁴⁹. Riprendendo il concetto di “formazione delle immagini ottiche” come la trasformazione “di un mondo che ha tre dimensioni spaziali in rappresentazioni bidimensionali”, si hanno sia le regole per la disposizione dei punti nell'area, sia le informazioni necessarie alla sua interpretazione. Le regole di disposizione sono principalmente quelle della geometria e dell'ottica²⁵⁰. Per l'interpretazione, è l'intuito che suggerisce si tratti di immagini reali, scaturite dalla riflessione della luce sulle superfici degli oggetti.

Per la luce è sufficiente ricordare che, normalmente, essa si diffonde nell'etere in linea retta, cambiando direzione quando i quanti rimbalzano sulle superfici o le attraversano.

248 Cfr. <http://www.dmi.unict.it/~gdibiasi/didattica/ixg/lezione1.pdf> p. 1112 7 gennaio 2007

249 Cfr. Marvin Minsky, *Significato e definizione*, in *Mente umana, mente artificiale*, a cura di Riccardo Valle, Op. Cit. pp. 253-259

250 Cfr. Ramesh Jain, Rangachar Kasturi, Brian G. Schunck, *Machine Vision*, Op. Cit., p. 6-10

I meccanismi simili alla camera oscura utilizzano questa proprietà della luce per ottenere immagini bidimensionali da proiezioni di ambienti tridimensionali. I punti dell'ambiente si posizionano sul piano dell'immagine (la parete interna alla camera opposta a quella forata) in base alla proiezione prospettica o, con determinati vincoli, alla prospettiva debole. Le “macchine fotografiche” (in senso lato), successive alla camera oscura utilizzano un obiettivo per focalizzare la luce sul piano dell'immagine. Qui, nelle macchine digitali, viene posizionato il CCD (*Charge Coupled Device*, dispositivo ad accoppiamento di carica) che si occupa di *campionare* la luce e di convertirla in segnali elettrici. Sono segnali molto deboli, che devono essere amplificati prima di essere inviati a un convertitore analogico-digitale (*frame-grabber*) che li trasforma in numeri. Questi vengono poi elaborati da un processore e archiviati²⁵¹.

Il CCD è un chip di silicio ricoperto da una serie di piccoli elettrodi chiamati *photo-site* (foto-elementi). Sistemati in una griglia, esiste un photo-site per ogni pixel di un'immagine. È il numero di photosite che determina la risoluzione di un CCD²⁵².

Campionamento

La risoluzione di un'immagine indica il suo grado di qualità. Generalmente il termine è riferito alle immagini digitali, ma anche una fotografia analogica ha una sua risoluzione. La risoluzione è la misura della densità dei pixel, cioè dei puntini elementari che formano l'immagine. Tale densità è riferita all'unità di lunghezza, di solito al pollice (ppi, pixel per inch o dpi, dot per inch). Per alcuni dispositivi, la densità dei pixel differisce nelle due dimensioni (per esempio negli scanner d'immagini); in tali casi è necessario indicare sia la risoluzione orizzontale sia quella verticale. Uno schermo di computer ha valori di risoluzione di 72 dpi per il Macintosh, e 96 per i pc. Le attuali stampanti casalinghe permettono di stampare immagini con risoluzioni di alcune centinaia di dpi. La risoluzione equivalente di una normale pellicola fotografica è di 3-4.000 dpi²⁵³.

251 Cfr. Ben Long, *Fotografia Digitale. Il Manuale. 3A ed.*, Op. Cit., pp. 19-20

252 Ibidem p. 23

253 <http://it.wikipedia.org/wiki/Risoluzione> 11 gennaio 2007

Le dimensioni assolute dell'immagine bitmap dipendono dal numero di pixel che la compongono lungo l'altezza e la larghezza.

Dimensione e risoluzione dei pixel sono entità inversamente proporzionali: tenendo costante la dimensione (larghezza e altezza), un'immagine ad alta risoluzione avrà più pixel, ma più piccoli, rispetto ad un'immagine con risoluzione inferiore²⁵⁴.

Tornando al processo di acquisizione, la superficie del CCD deve essere, per captare la luce, caricata di elettroni. Quando viene colpita dalla luce, gli elettroni si agglomerano sopra la griglia di photo-site. Maggiore è la luce, maggiore sarà il numero di elettroni agglomerati sul photo-site. Successivamente all'esposizione, la macchina deve semplicemente misurare la quantità di carica a ogni photo-site, stabilendo così quanta luce ha inciso in ogni punto determinato. Questa matrice d'incidenza è poi passata al convertitore analogico-digitale²⁵⁵, che la trasformerà in numeri.

Il termine “dispositivo di accoppiamento di carica” deriva da come le cariche dei singoli photo-site vengono interpretate dalla macchina: successivamente all'esposizione, le cariche della prima fila di photo-site sono trasferite al dispositivo/registo di uscita (read out register), dove vengono amplificate per poi essere inviate al convertitore analogico-digitale. Ogni fila di cariche è elettricamente unita alla fila successiva in modo che, dopo che una fila è stata processata (letta e cancellata), le successive si muovono posizionandosi sullo spazio appena lasciato libero. Una volta processate (e cancellate) tutte le cariche, il CCD si ricarica ed è pronto a ripetere il processo²⁵⁶.

I photo-site reagiscono solo alla quantità di luce ricevuta, rimanendo indifferenti al relativo colore. Per ottenere una percezione al colore è necessario operare un filtraggio Red Green Blue (RGB)²⁵⁷. Con il metodo introdotto da Maxwell ogni photo-site viene colorato da un filtro, rosso, verde o blu. Questa combinazione è detta *allineamento*

254 Cfr. Pizzirani A., *Teoria e tecnica di elaborazione delle immagini*, p. 14

255 Cfr. Ben Long, *Fotografia Digitale. Il Manuale. 3A ed.*, Op. Cit., p. 23

256 Ibidem p. 24

257 *Red, Green, Blue*, i tre colori in cui viene scomposta la luce nello spazio colore additivo.

(array) di filtri a colori e segue uno schema abbastanza standard per la maggior parte dei CCD. Con questi filtri il CCD può produrre immagini distinte per ogni colore, ma incomplete in quanto mancanti dei pixel coperti dagli altri filtri; per ottenere l'immagine completa, le immagini dei vari colori devono essere interpolate fra loro. Questo è quanto avviene su macchine a singolo CCD, tuttavia si trovano macchine che ne hanno uno per ogni colore²⁵⁸.

Dal CCD si ottiene un segnale elettrico continuo, il segnale video, che viene elaborato dal frame grabber o, meglio, digitalizzato. Il risultato è una matrice bidimensionale rettangolare di righe e colonne, contenente valori interi, archiviata in memoria²⁵⁹.

Come detto, l'operazione di digitalizzazione avviene in due passaggi: campionamento e quantizzazione. La prima, appena trattata, inizia nella matrice del CCD e suddivide la scena 3D nella matrice bidimensionale; da questa fase dipende la risoluzione dell'immagine e, di conseguenza, la sua qualità.

Quantizzazione

L'operazione di quantizzazione cromatica consiste nell'assegnare a ciascun pixel uno o più valori numerici che ne definiscano il colore. L'operazione non è semplice in quanto, come già visto, dipende non solo dalla luce, ma anche dalla superficie dell'oggetto e dall'osservatore. Esso è infatti un campo di ricerca aperto della psicofisica. Tuttavia, come accennato poc'anzi, Maxwell risolse, nel 1869, il problema. Egli agì per approssimazione, cercando di ridurre le infinite variazioni dello spettro a tre colori primari: il rosso, il verde e il blu²⁶⁰.

Questa scomposizione è ancor oggi utilizzata per i tubi catodici delle televisioni e dei monitor, mentre per la stampa si usa la rappresentazione CMY (ciano, magenta e giallo) cui si aggiunge il nero per risolvere alcuni problemi d'assorbimento della carta e

258 Cfr. Ben Long, *Fotografia Digitale. Il Manuale. 3A ed.*, Op. Cit., p. 26

259 Cfr. Emanuele Trucco, Alessandro Verri, *Introductory Techniques for 3-D Computer Vision*, Op. Cit, p. 29

260 Cfr. Ben Long, *Fotografia Digitale. Il Manuale. 3A ed.*, Op. Cit., p. 20

per migliorare la resa cromatica.

La scala d'intensità di ciascuno dei colori primari non può essere casuale; l'intervallo, tra il valore minimo ed il valore massimo dell'intensità di ogni colore è ripartito in un certo numero di livelli. Più sono i livelli, maggiore è la precisione nel rendere il colore; normalmente si utilizzano 256 livelli per ogni colore primario. Il numero di livelli è, di solito, espresso in bit, mentre, il termine *banda* indica che si tratta di informazioni relative al colore. Un'immagine a 256 livelli di colore in RGB viene spesso definita come a 24 bit per pixel, 8 per ogni colore primario²⁶¹.

Le immagini in bianco e nero sono un tipo particolare di quantizzazione, in quanto a ogni pixel corrisponde un valore numerico che fornisce informazioni solamente sulla sua luminosità. Queste immagini si dicono a *livelli di grigio* e sono particolarmente utili al riconoscimento delle forme geometriche: per questo sono spesso prese come base per la computer vision²⁶².

Le *immagini binarie* sono un caso particolare delle immagini in bianco e nero, in quanto i valori d'intensità associati a ciascun pixel sono solo due. Le sfumature sono ottenute con una tecnica particolare definita *dithering*, che sfrutta un'illusione ottica²⁶³.

Il rumore

Quanto finora scritto è basato sull'assunto che l'immagine digitale sia "perfetta", ossia non presenti difetti quali riflessi, tonalità di colore errate, pixel contigui con colori contrastanti, ecc. Generalizzando, si hanno dei pixel con valori nell'immagine che non sono quelli attesi, poiché si sono corrotti durante la fase di acquisizione, quindi non sono utili allo scopo dell'elaborazione. Come conseguenza, i valori dei pixel di due immagini della stessa scena, ottenuti con la stessa macchina e con le stesse condizioni di luce non sono mai esattamente gli stessi. Queste fluttuazioni producono degli errori nei calcoli basati sui valori dei pixel; è quindi necessario considerare la quantità di rumore

261 Cfr. Roberto Marangoni, Marco Geddo, *Le Immagini digitali*, Op. cit., p. 8

262 Ibidem p. 10-11

263 Ibidem

presente nell'immagine in modo da non inferire conclusioni errate (un po' come accade per le illusioni nella visione umana).

Immagini spaziali o di profondità²⁶⁴

In molte applicazioni si utilizza la visione per stimare le distanze; ad esempio, per tenere i veicoli lontani dagli ostacoli, controllare la produzione di oggetti ,e più in generale rilevare la forma dalle superfici.

Un'immagine che misura l'intensità della luce si dimostra carente per rilevare le distanze, in quanto i valori dei pixel sono collegati alle superfici geometriche solo indirettamente (ossia, dipendono sia dalle proprietà geometriche e fisiche di queste, sia dalle condizioni d'illuminazione).

Tuttavia è possibile rilevare direttamente la forma degli oggetti attraverso dei sensori che misurano direttamente le distanze spaziali, ottenendo delle immagini spaziali, ossia delle matrici i cui pixel esprimono la distanza tra un quadro di riferimento e un punto visibile della scena. Viene così riprodotta direttamente la struttura 3D della scena.

Per quanto riguarda la rappresentazione, le immagini spaziali possono essere rese in tre modi distinti: come un'immagine d'intensità, nella forma di una “nuvola di punti” o in forma r_{ij} . La prima forma corrisponde alle immagini finora considerate. La seconda è una lista di coordinate 3D con un dato riferimento (es: un piano ,ecc): essa non richiede un ordine preciso e suggerisce direttamente l'idea della forma. L'ultima è una matrice che riporta i valori relativi alla profondità dei singoli punti lungo le direzioni degli assi x y dell'immagine, rendendo esplicite le informazioni spaziali.

Per quanto riguarda i sensori, essi possono misurare la profondità di un singolo punto, la distanza e la forma dei profili delle superfici o superfici complete. Si possono distinguere in attivi e passivi. I primi proiettano energia (es.: un tipo di luce, impulsi sonar) sulla scena e ne rilevano la posizione per ottenere la misura; o utilizzano l'effetto

²⁶⁴ Cfr. Emanuele Trucco, Alessandro Verri, *Introductory Techniques for 3-D Computer Vision*, Op. Cit, pp. 40-47

di cambiamenti controllati di alcuni parametri del sensore (es.: la messa a fuoco). I secondi si affidano solamente alla misurazione intensità luminosa dell'immagine per rilevare la profondità.

I sensori utilizzano vari principi fisici di funzionamento come i radar e i sonar, interferometria Moiré, la messa a fuoco/sfocatura attiva e la triangolazione²⁶⁵.

Quest'ultima è particolarmente interessante in quanto utilizza le macchine a rilevazione d'intensità (fotografiche, telecamere, ecc) partendo, quindi, da una conoscenza già acquisita. Questi sensori forniscono mappe di coordinate 3D accurate e dense, sono semplici da comprendere e costruire, nonché molto diffusi. La rilevazione dell'immagine scaturisce dall'interazione di un proiettore con una telecamera. Il primo fornisce sia la luce (il mezzo) che il piano di luce (coordinate geometriche) di riferimento, mentre la seconda effettua le rilevazioni²⁶⁶.

Strumenti di acquisizione

Dopo aver esaminato la complessità e la bellezza dell'occhio umano è quasi triste constatare i limiti degli strumenti di acquisizione oggi disponibili. Probabilmente per questo motivo Paul Sajda, della Columbia University, cerca di interfacciare le capacità di acquisizione delle immagini del sistema visivo umano con alcuni algoritmi della computer vision²⁶⁷. Di seguito non si tratteranno queste sperimentazioni d'avanguardia, ma ci si limiterà a riportare alcuni strumenti di uso comune specificandone, brevemente, le caratteristiche.

Scanner

Sono stati, probabilmente, fra i primi strumenti utilizzati per la digitalizzazione delle immagini: il loro processo di acquisizione parte infatti da immagini già stampate su carta o da negativi.

265 Ibidem p.42

266 Cfr Linda G. Shapiro, George C. Stockman, *Computer Vision*, Op. Cit., pp. 47-49

267 http://newton.bme.columbia.edu/mainTemplate.htm?liinc_news.htm 13 gennaio 2007 <http://punto-informatico.it/p.aspx?id=1568530&r=PI> 13 gennaio 2007

<http://www.wired.com/news/technology/medtech/0,71364-0.html> 13 gennaio 2007

Vi sono due tecnologie di scansione, quella a CCD in parte già vista, e quella PMT (*Photo Multiplier Tubes*). Entrambe convertono diversi livelli di luminosità in segnali elettrici a variazione continue (analogici), i quali passano al convertitore A/D (analogico/digitale) che effettua i processi di campionatura (suddivisione del segnale in quantità discrete) e quantizzazione (ossia l'attribuzione di un'etichetta numerica alle quantità discrete ottenute)..

La tecnologia PMT permette di riprodurre una gamma cromatica più vasta; tuttavia essa è più costosa, richiede un più ampio livello professionale e un elevato controllo, e accetta solo originali flessibili²⁶⁸.

Fotocamera Digitale

È il dispositivo più comune e flessibile per l'acquisizione di immagini. Utilizza normalmente la tecnologia CCD. È molto simile alla normale macchina fotografica; la differenza sta nel fatto che la pellicola viene sostituita con un sottile strato di celle allo stato solido, necessarie alla conversione dell'energia luminosa in cariche elettriche. Dalla matrice del CCD dipende direttamente la risoluzione e, in parte, la dimensione dell'immagine.

I sensori possono essere anche di tipo CMOS (*Complimentary Metal Oxide Semiconductors*), ovvero semiconduttori complementari a ossido di metallo. Questi sono normalmente più economici dei sensori CCD in quanto costruiti con tecnologie comuni ai diffusi chip per calcolatori. La differenza si nota anche nei consumi: i sensori CMOS sembrano infatti essere meno esigenti, assicurando una vita più lunga alle batterie della macchina e meno problemi di surriscaldamento. Permettono anche di integrare più funzioni nello stesso chip agevolando, quindi, la costruzione di macchine più compatte²⁶⁹.

Le fotocamere di ultima generazione possono essere collegate direttamente a un computer (normalmente via USB o standard IEEE 1394) sia per l'archiviazione delle

268 Cfr. Pizzirani A., *Teoria e tecnica di elaborazione delle immagini*, p. 5

269 Cfr. Ben Long, *Fotografia Digitale. Il Manuale. 3A ed.*, Op. Cit., pp. 30

immagini, sia per il controllo dello scatto o delle impostazioni della macchina.

Alcuni apparecchi offrono anche caratteristiche avanzate quali l'autofocus, pre-impostazioni per situazioni o effetti particolari, ecc.

Telecamere

Le telecamere, aventi come obiettivo la creazione di immagini visionabili dall'uomo, registrano sequenze d'immagini con una frequenza di 30 fotogrammi al secondo, permettendo la rappresentazione del movimento degli oggetti. Si ha quindi una dimensione aggiunta alle coordinate spaziali contenute nella singola immagine. Per permettere una percezione armoniosa dell'immagine si utilizzano 60 mezzi fotogrammi al secondo; se ogni singola immagine può essere considerata come una successione di righe, questi mezzi fotogrammi sono dati dagli insiemi delle righe dispari e pari che scorrono in successione alternata.

Le telecamere che creano immagini per un utilizzo automatico possono registrare immagini a qualsiasi velocità e non necessitano di utilizzare la tecnica dei mezzi fotogrammi.

La tecnologia a CCD utilizzata nelle telecamere per la computer vision ha qualche volta sofferto l'utilizzo di standard video ottimizzati per l'uomo. Un primo problema è dato dalla complessità dovuta al fatto che righe pari e dispari dell'immagine sono intrecciate, processo non necessario per la macchina. Secondo, molte matrici di CCD hanno un rapporto 4:3 tra ampiezza e altezza, come la maggioranza dei video per l'uomo. Pixel quadrati e un rapporto a 1 tra le dimensioni è più favorevole alla computer vision in quanto facilitano l'elaborazione. Il mercato ha guidato i costruttori verso standard umani: gli sviluppatori di sistemi visivi automatici hanno perciò dovuto adattarsi o sostenere costi superiori per strumenti prodotti su misura in quantità limitate²⁷⁰.

270 Cfr Linda G. Shapiro, George C. Stockman, *Computer Vision*, Op. Cit., p. 26

Elaborazione automatica dell'immagine

Dopo aver cercato di definire in cosa consiste la ricerca/tecnologia della computer vision e l'immagine digitale, nonché alcuni strumenti per la relativa acquisizione, si può parlare di automatizzazione del processo di elaborazione. A tal fine è importante comprendere la relazione tra la geometria della formazione dell'immagine e la rappresentazione delle immagini nel computer. Deve essere chiaro il collegamento tra la notazione matematica utilizzata nello sviluppo degli algoritmi della computer vision e quella algoritmica utilizzata nei programmi²⁷¹.

Come detto un pixel è un campione dell'intensità dell'immagine quantizzato a un valore intero e l'immagine è una matrice bidimensionale di pixel. Gli indici $[i,j]$ di un pixel sono valori interi che specificano le righe e le colonne della matrice. Il pixel $[0,0]$ è posizionato nell'angolo in alto a sinistra dell'immagine. I valori dell'indice i si susseguono dall'alto verso il basso, mentre quelli di j si dirigono da sinistra verso destra. Questo tipo di notazione corrisponde strettamente alla sintassi della matrice utilizzata nei programmi. La posizione dei punti sul piano dell'immagine ha come coordinate x e y . La coordinata y (ordinate) corrisponde alla direzione verticale, la x (ascisse) a quella orizzontale. L'asse delle y va dal basso verso l'alto, quella delle x *sinistra* verso destra. Quindi i valori riportati dalla matrice degli indici i e j sono in ordine rovescio rispetto ai valori riportati dalla matrice delle coordinate relative alle posizioni x e y . È necessario quindi eseguire degli algoritmi per passare da un sistema di coordinate all'altro²⁷².

In un sistema per la formazione dell'immagine, ogni pixel occupa un'area definita sul piano dell'immagine. Le posizioni sul piano dell'immagine possono essere quindi rappresentate da frazioni di pixel. La matrice dei pixel del software corrisponde alla griglia di posizioni sul piano dell'immagine da cui è ottenuta²⁷³.

271 Cfr. Ramesh Jain, Rangachar Kasturi, Brian G. Schunck, *Machine Vision*, Op. Cit., pp. 12-13

272 Ibidem

273 Ibidem

Questa premessa ci permette di comprendere quali elaborazioni possono essere compiute sull'immagine, o meglio sulla matrice di valori che la rappresenta. Tuttavia è necessario considerare il livello, o meglio la posizione, su cui si opera. Quindi occorre considerare ogni algoritmo in base alle trasformazioni che pone in essere, a quali sono i dati richiesti e i risultati forniti. Certamente l'elaborazione si svolge sull'immagine, ma come risultato si hanno dei simboli rappresentanti, per esempio, l'identità e la posizione di un oggetto. La quantità di dati elaborati da un sistema visivo è enorme, quindi è richiesto hardware dalle capacità adeguate. Negli ultimi anni si sono sperimentate anche architetture particolari, come le accennate reti neurali. Analizzare quali siano le caratteristiche delle operazioni significa, implicitamente, valutare le possibili richieste computazionali²⁷⁴.

A seconda di quali dati vengono trattati possiamo distinguere i seguenti livelli²⁷⁵:

5. A livello dei punti (puntuale)

Alcune operazioni si basano solamente su un punto dell'immagine. Un esempio è l'operazione di soglia.

Questa operazione è efficientemente implementata con una *look-up table*.

6. A livello locale

L'intensità dei punti nell'immagine risultato dipende non solo da un singolo punto dell'immagine di partenza, ma anche da quelli che gli sono adiacenti/vicini. La ridistribuzione dei punti (*smoothing*) e la rilevazione dei contorni sono operazioni locali. Per questo motivo sono adatti a queste operazioni sistemi in grado di effettuare calcoli matriciali o quelli *Single Instruction, Multiple Data* (SIMD). In pratica, sono operazioni facilmente implementate su elaboratori paralleli ed eseguite in tempo reale.

²⁷⁴ Ibidem

²⁷⁵ Ibidem pp. 14-17

7. A livello globale

Il risultato dipende dall'intera immagine, e può essere una nuova immagine o un'interpretazione simbolica della prima. Un istogramma dei valori d'intensità o una trasformata di Fourier sono esempi di operazioni globali.

La complessità dei processi globali rallenta l'elaborazione nei sistemi visivi; purtroppo molte operazioni sono nella loro natura globali.

8. A livello oggetto

Questo livello è il più specifico della computer vision, dato che i precedenti sono la base anche per altre materie come l'elaborazione dell'immagine. Le dimensioni, l'intensità media, la forma, e altre caratteristiche dell'oggetto devono essere valutate perché il sistema le riconosca. Al fine di determinare queste proprietà vengono effettuate delle operazioni solamente sui pixel appartenenti all'oggetto. Ma il punto centrale è: cos'è un oggetto? Come si può rilevare?

Gli oggetti sono normalmente definiti dal loro particolare contesto. In effetti, molte operazioni in computer vision sono svolte per trovare la posizione di un oggetto nell'immagine. Tuttavia, definire cos'è un oggetto è come trovarsi nella condizione del “gatto che si morde la coda”. Per valutare le caratteristiche dell'oggetto abbiamo bisogno di sapere quali punti appartengono a tale oggetto, ma per identificarli è necessario sapere da quali caratteristiche sono contraddistinti. Sono stati fatti molti sforzi per distinguere le figure dallo sfondo, o raggruppare i punti in oggetti.

Questi livelli richiamano le teorie psicologiche della visione e la loro evoluzione. La teoria strutturalista è collegabile al primo livello, ossia la visione come percezione di punti. La Gestalt e la teoria ecologica di Gibson ai tre successivi. Per quanto riguarda il costruttivismo, esso suppone anche una “comprensione” dell'immagine che non è strettamente collegabile a questi livelli, se non come fase successiva, quella definita da

Marr come “The Category-Based stage”.

Marr sosteneva che si può decomporre il livello algoritmico in quattro stadi di elaborazione/percezione dell'immagine retinica, di complessità crescente: *image-based*, *surface-based*, *object-based* e *category-based*.

Questi livelli possono essere utilizzati sia per l'analisi del sistema visivo umano che di uno artificiale in quanto l'immagine retinica può essere considerata alla stregua di un'immagine digitale: i singoli recettori possono essere infatti considerati, anche se in via molto approssimativa, come singoli pixel.

Ogni stadio è definito in base al tipo di rappresentazione che fornisce e dai processi necessari ad elaborarla partendo dalla precedente rappresentazione. Nel dettaglio si ha²⁷⁶:

1. Image-based.

È lo stadio immediatamente successivo alla percezione ottica dell'immagine. Si occupa di rilevare i bordi e le linee, collegandoli tra loro in modo più globale, unendo le immagini stereoscopiche, definendo regioni bidimensionali nell'immagine e rilevando altre caratteristiche di base, quali i termini delle linee e le cosiddette “bolle”. Le caratteristiche bidimensionali delle immagini descrivono la loro struttura e organizzazione prima di essere interpretate come proprietà di scene tridimensionali.

2. Surface-based.

Il secondo stadio riguarda le modalità di recupero delle proprietà intrinseche delle superfici visibili nell'ambiente esterno che possono aver prodotto le caratteristiche riscontrate nello stadio precedente. La fondamentale differenza tra questi sta nel

²⁷⁶ Stephen E. Palmer, *Vision Science - Photons to Phenomenology*, Op. cit. pp. 87-92

fatto che il secondo rappresenta le informazioni sull'ambiente in termini di disposizioni nello spazio tridimensionale delle superfici visibili, mentre il primo si riferisce alle caratteristiche dell'immagine in termini di modelli ipotetici bidimensionali.

Il concetto di una rappresentazione esplicita basata sulle superfici, introdotto inizialmente da Gibson, divenne popolare quando fu quantitativamente formulato dai teorici della computer vision e implementato in simulazioni funzionanti sull'elaboratore. Marr, Barrow e Tennenbaum fecero delle rappresentazioni basate sulle superfici quasi nello stesso periodo (1978), e descrissero degli algoritmi che potevano ricavarle realmente da immagini in scala di grigio.

Marr chiamò la sua rappresentazione “*2.5-D sketch*” per enfatizzare il fatto che si colloca tra la struttura 2-D basata sull'immagine e quella 3-D ottenuta dalla rappresentazione basata sugli oggetti.

3. Object-based.

È in questo stadio che le rappresentazioni includono realmente informazioni tridimensionali. Affinché il sistema visivo sia in grado di gestirle, è necessario si facciano ulteriori assunzioni implicite, riguardanti la natura del mondo visivo, poiché le inferenze attuate si basano anche su superfici non visibili o parti di superfici. Questo stadio è chiamato *Object-based* in quanto il considerare superfici non viste comporta una rappresentazione esplicita degli oggetti presenti nell'ambiente. Questo può essere ottenuto in due modi:

- a. Con la semplice estensione della rappresentazione basata sulle superfici fino a includere quelle non viste all'interno di uno spazio tridimensionale.
- b. Concependo gli oggetti come entità intrinsecamente tridimensionali rappresentati, a loro volta, da disposizioni di alcuni insiemi di forme primitive

tridimensionali. Questo approccio può essere definito volumetrico, in quanto gli oggetti sono rappresentati come volumi di particolare forma.

4. *Category-based.*

Se il fine della percezione è fornire all'organismo accurate informazioni sull'ambiente per aiutarlo a sopravvivere e riprodursi, l'ultimo stadio della percezione deve riguardare la riacquisizione delle proprietà funzionali degli oggetti: cosa apportano all'organismo (solievo, benessere, nonché desideri, scopi e motivazioni).

Questo processo è denominato *category-based stage* in quanto è convinzione ampiamente condivisa che le proprietà funzionali degli oggetti siano accessibili tramite un processo di categorizzazione.

La categorizzazione (o "*pattern recognition*" riconoscimento di modelli) è un metodo per desumere le funzioni evolutive rilevanti che propone il coinvolgimento di due operazioni:

- a. Il sistema visivo classifica un oggetto come appartenente a un insieme di categorie conosciute basandosi sulle sue proprietà visibili: forma, dimensioni, colore e posizione.
- b. L'identificazione così ottenuta permette di accedere alle conoscenze acquisite riguardanti la tipologia dell'oggetto, tra cui le sue funzioni e le aspettative sul suo comportamento. Ad esempio, una tazza è considerata utile per contenere liquidi e, quindi, per bere.

I vantaggi di un simile approccio dovrebbero essere evidenti. Tuttavia esiste un approccio alternativo. Il sistema visivo è in grado di percepire le funzioni dell'oggetto più o meno direttamente, registrandole/acquisendole dalle sue

caratteristiche visibili senza una preventiva categorizzazione. I primi a sostenere questa ipotesi furono i teorici della Gestalt, parlando di *caratteri fisiognomici*. Gibson concordò con questa impostazione, espandendo la sua teoria della percezione diretta all'inclusione delle funzioni.

È possibile, e anche molto probabile, che entrambi i processi siano utilizzati nella percezione delle funzioni degli oggetti. Alcuni come sedie e tazze hanno delle caratteristiche funzionali così intrinsecamente connesse al loro aspetto che non necessitano del processo di categorizzazione per comprenderne l'utilizzo. Altri, come calcolatori e telefoni, hanno delle funzioni così slegate dalla loro forma che non possono fare a meno del procedimento di categorizzazione.

Questi quattro stadi del processo visivo rappresentano l'ipotesi più plausibile sulla struttura della percezione visiva. Essi sono elencati nell'ordine in cui dovrebbero essere logicamente svolti, tanto che ognuno di essi deve essere completato prima che il successivo abbia luogo. Come accennato, le connessioni nel cervello suggeriscono invece l'opposto, ossia dei meccanismi di continui feedback.

Nel prossimo capitolo si analizzeranno nel dettaglio le operazioni basilari della computer vision, associate ai vari livelli.

Come funziona? - Tecnologie implementate

Al fine di offrire le basi su cui poter elaborare analisi e considerazioni si riportano di seguito alcune delle tecniche alla base della computer vision e dell'elaborazione dell'immagine, rilevandone le analogie con la visione umana.

Informazioni dall'immagine

L'immagine è ricca d'informazioni: per questo essa dev'essere interpretata e le relative informazioni estratte. Uno dei fini della computer vision (non il solo) è eseguire questo compito automaticamente.

Il primo passaggio, eseguito dai sensori, è trasformare l'immagine ottica in una rappresentazione numerica. Questo “semplice” passaggio permette già di avere una conoscenza più approfondita dell'immagine stessa, ossia i valori di luminosità di ogni pixel. Non solo, permette anche di conoscere come i valori dei pixel si ripetono nello spazio.

Nel proseguo dell'elaborato si considereranno le sole immagini binarie (in bianco e nero) o a livelli di grigio, le prime utilizzate dai ricercatori²⁷⁷. La procedura di trasformazione delle immagine in bianco e nero può essere comunque facilmente estesa alle immagini a colori.

L'istogramma dei livelli di grigio

È una delle modalità più semplici per rappresentare il contenuto informativo di un'immagine. Per costruirlo si supponga di avere un'immagine a 256 livelli di grigi: in ascissa vengono riportati i valori possibili di ogni pixel, da 0 (corrispondente al nero) a 255 (corrispondente al bianco); in ordinata la quantità (il numero) di pixel che corrisponde a ogni singolo valore di grigio²⁷⁸.

L'istogramma ottenuto riassume la distribuzione dell'informazione cromatica presente nell'immagine, divenendo un valido strumento per determinate operazioni sulla stessa, quali l'appartenenza di un pixel all'oggetto, l'eliminazione di alcuni disturbi, l'esaltazione di particolari, ecc²⁷⁹.

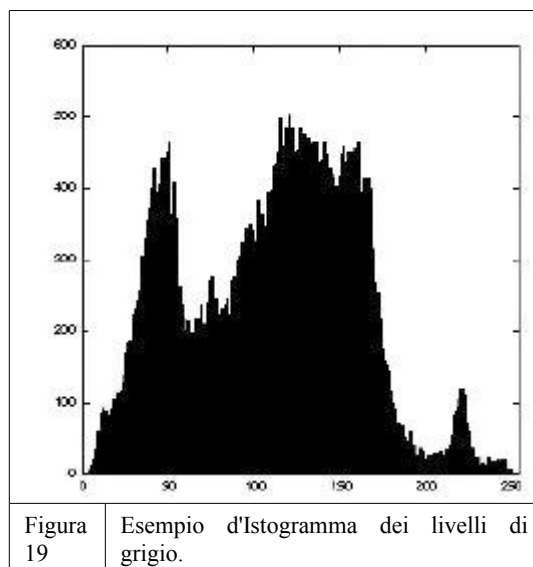
Per inciso, l'istogramma fornisce informazioni *quantitative* sulla distribuzione dei livelli di grigio nell'immagine tralasciando, tuttavia, ogni informazione sulla distribuzione spaziale²⁸⁰.

277 Cfr. Stephen E. Palmer, *Vision Science - Photons to Phenomenology*, Op. cit. p. 60 e Ramesh Jain, Rangachar Kasturi, Brian G. Schunck, *Machine Vision*, Op. Cit., p. 25

278 Cfr. Roberto Marangoni, Marco Geddo, *Le Immagini digitali*, Op. cit., p. 43

279 Ibidem

280 Ibidem p. 45



Spettro di frequenza spaziale

Il concetto di frequenza è già stato introdotto trattando la luce come grandezza fisica, in particolare definendola una caratteristica dell'onda elettromagnetica. Si tratta di una grandezza unidimensionale (il numero di oscillazioni compiute in un dato periodo di tempo), mentre le immagini, e di conseguenza le matrici o *reticoli* che le rappresentano, sono bidimensionali, hanno cioè righe e colonne. Sorgono quindi alcuni quesiti:

3. *Com'è possibile utilizzare la frequenza per ottenere le immagini/reticoli?*

Se si considera una riga di pixel, ognuno con un dato valore, e la si replica per le successive, si ottiene un'immagine: abbiamo un segnale bidimensionale come replica di un segnale unidimensionale. Questa è la forma bidimensionale più elementare che si possa avere²⁸¹.

4. *Ma cosa misura la frequenza?*

Si supponga che un primo reticolo sia costruito con una riga per metà da pixel bianchi e per il resto neri, mentre il secondo reticolo da $\frac{1}{4}$ di bianchi, $\frac{1}{4}$ da neri, $\frac{1}{4}$ bianchi e $\frac{1}{4}$ neri, e i successivi con quantità via via dimezzate.

²⁸¹ Cfr. Roberto Marangoni, Marco Geddo, *Le Immagini digitali*, Op. cit., p. 45

Possiamo considerare il bianco e il nero come le due opposte creste dell'onda e la frequenza come la misura delle variazioni che avvengono in ciascun reticolo. Per i suddetti reticoli, avremo quindi una frequenza di 2, 4, 8, ecc.

I reticoli con frequenze così regolari sono composti da linee verticali; per averne di diversi è necessario considerare tutti i loro parametri²⁸²:

4. **Ampiezza** (o contrasto): si riferisce alla differenza di luminosità tra la parte più chiara e quella più scura, che corrisponde alla differenza d'altezza dell'onda tra il picco e la valle sotto il profilo della luminanza.
5. **Orientamento**: si riferisce all'inclinazione delle barre bianche o nere; è espressa in gradi partendo dalla posizione verticale e procedendo in senso orario.
6. **Fase**: si riferisce alla posizione della curva in riferimento a un punto determinato.
7. **Frequenza spaziale**: si riferisce all'ampiezza delle righe (i reticoli a bassa frequenza hanno linee spesse, quelli ad alta frequenza linee fini).

Modificando “a piacere” questi parametri si possono ottenere i reticoli più vari, in pratica corrispondenti a tutte le immagini acquisibili dall'occhio umano.

Per visualizzare il contenuto in frequenze di un reticolo qualsiasi è possibile utilizzare un istogramma simile a quello usato per la quantizzazione dei livelli di grigio: nelle ascisse vi sono i valori delle frequenze spaziali presenti nel reticolo, nelle ordinate, è indicata la percentuale di ciascun pixel dell'immagine in corrispondenza a ciascun valore di frequenza. Questo istogramma può essere considerato, in via euristica, lo spettro di frequenza del reticolo²⁸³.

Poter manipolare lo spettro di frequenza permette di intervenire sulle proprietà spaziali

282 Cfr. Stephen E. Palmer, *Vision Science - Photons to Phenomenology*, Op. cit. p. 160

283 Cfr. Roberto Marangoni, Marco Geddo, *Le Immagini digitali*, Op. cit., p. 47

dei pixel, ossia di aumentare o diminuire lo spessore delle linee che costituiscono una delle primitive dell'immagine. Come questo sia possibile si vedrà analizzando le operazioni globali sull'immagine; ora, invece si considereranno le operazioni sui singoli punti²⁸⁴.

Operazioni puntuali

Una delle più semplici elaborazioni eseguibili su un'immagine è la sostituzione dei pixel corrispondenti a un determinato valore con pixel di altro valore. Un esempio è la sostituzione di un valore di grigio dell'immagine di partenza con un altro nell'immagine derivata. La variazione sarà tanto più evidente quanto più alto sarà il numero di pixel coinvolti o la differenza tra il valore di partenza e quello di sostituzione²⁸⁵.

Un'elaborazione puntuale è detta omogenea se il risultato dipende solo dal valore del pixel cui è applicata; per lo stesso motivo vengono anche dette manipolazioni della scala dei grigi o dei colori. Se il risultato dell'elaborazione dipende anche dalla posizione del pixel nell'immagine, si parla di elaborazioni puntuali non omogenee.

Ecco alcune tipiche elaborazioni puntuali:

4. Aggiunta o sottrazione di una costante a tutti i pixel (per compensare sotto o sovraesposizioni).
5. Clipping (ritagliare determinate aree per renderle visibili all'osservatore o stabilirne l'appartenenza a una regione)²⁸⁶.
6. Espansione del contrasto.
7. Inversione della scala dei grigi(negativo).

284 Ibidem p. 48

285 Ibidem p. 39

286 <http://it.wikipedia.org/wiki/Clipping> e <http://en.wikipedia.org/wiki/Clipping> 20 gennaio 2007

8. Modifica (equalizzazione o specifica) dell'istogramma.

9. Presentazione in colore falso.

10. Soglia.

Quest'ultima elaborazione puntuale, di seguito spiegata, è tra quelle più spesso utilizzate.

L'operazione di soglia

L'operazione di soglia è comune in psicofisica e nell'analisi dei segnali. Consiste essenzialmente nello stabilire un valore di riferimento per procedere, in successione, al confronto con altri valori al fine di eseguire una determinata azione. Nel nostro caso, la soglia viene confrontata con i valori dei pixel componenti l'immagine e l'azione consiste nella loro sostituzione con grandezze diverse a seconda siano inferiori o superiori al valore di riferimento²⁸⁷.

$$P(m, n) = \begin{cases} 255 & \text{se } p(m, n) \geq k \\ 0 & \text{se } p(m, n) < k \end{cases}$$

In questo caso, un esempio in formula può forse essere più chiaro.

Se con k si indica un valore di soglia nei livelli di grigio dell'immagine, i valori del pixel dell'immagine elaborata $[P(m,n)]$ saranno zero se il valore di partenza $[p(m,n)]$ è inferiore a k , 255 se superiore. Come si può dedurre, questa operazione è utile per trasformare un'immagine a colori o a scala di grigi in un'immagine binaria. È un primo metodo per poter distinguere i valori d'illuminazione dei pixel che appartengono ad un oggetto rispetto a quelli appartenenti allo sfondo, ossia per eseguire una segmentazione/suddivisione dell'immagine.

287 Cfr. Roberto Marangoni, Marco Geddo, *Le Immagini digitali*, Op. cit., p. 39

La segmentazione è una procedura per suddividere l'immagine in sotto-immagini, denominate regioni che, a loro volta, rappresentano oggetti o parti di oggetti. L'operazione di soglia e la segmentazione sono, per le immagini binarie, sinonimi²⁸⁸.

L'immagine binaria ottenuta permette di elaborare le proprietà geometriche e topologiche degli oggetti per poi utilizzarli: il valore di soglia deve essere perciò scelto in relazione all'illuminazione e all'indice di riflessione degli oggetti. Questo implica una conoscenza del dominio sotto analisi, in quanto lo stesso valore di soglia applicato ad un'altra immagine potrebbe non funzionare. La scelta della soglia ottimale, se fatta dall'uomo è spesso frutto dell'esperienza. In alcuni casi all'avvio del sistema si fanno alcuni tentativi per determinarne interattivamente il valore adatto²⁸⁹.

La determinazione automatica della soglia ottimale è il primo passo per l'analisi dell'immagine con i sistemi di computer vision. Molte tecniche sono state sviluppate per utilizzare la distribuzione della luminosità nell'immagine e la conoscenza degli oggetti d'interesse per selezionare automaticamente l'appropriato valore di soglia.

Si può notare il parallelismo tra quest'operazione e la funzione svolta dai bastoncelli nell'occhio umano: infatti questi si occupano di rilevare la luminosità, individuando solo movimenti e ombre. Sono considerati da Gregory come la forma più primitiva di visione²⁹⁰, è forse per questo che non abbiamo difficoltà a comprendere le immagini binarie, come le linee tratteggiate o *silhouettes*.

Operazioni locali

Si parla di elaborazioni locali quando le trasformazioni dipendono non solo da un singolo punto dell'immagine di partenza, ma anche da quelli ad esso adiacenti/vicini, ossia da quello che viene definito l'"intorno" del pixel. Avremo un intorno di 3'3 quando il pixel in posizione centrale, identificato con $p(m,n)$, è circondato da un quadrato di lato 3 pixel; un intorno di 5'5 quando il lato è di 5 pixel, e così via. Vengono

288 Cfr. Ramesh Jain, Rangachar Kasturi, Brian G. Schunck, *Machine Vision*, Op. Cit., p. 29

289 Ibidem p. 27

290 Cfr Richard L. Gregory, *Occhio e cervello – La psicologia del vedere*, Op. cit., p. 79

scelti intorno di lato dispari per fare in modo che il pixel considerato sia esattamente al centro dell'intorno. Vi possono essere eccezioni per alcune operazioni particolari. Il lato è scelto in base al tipo di filtraggio da compiere, con preferenza di intorni piccoli per ridurre la quantità di calcoli da effettuare²⁹¹.

$p(m-1,n-1)$	$p(m,n-1)$	$p(m+1,n-1)$
$p(m-1,n)$	$p(m,n)$	$p(m+1,n)$
$p(m-1,n+1)$	$p(m,n+1)$	$p(m+1,n+1)$
Figura 20	Esemplificazione di una maschera/filtro, in giallo il pixel centrale, in blu l'intorno; nello specifico si tratta di un filtro di media, che assegna, cioè, al pixel centrale la media calcolata sui valori dell'intorno.	

Le operazioni locali possono essere di tipo aritmetico o logico²⁹².

Operazioni aritmetiche

Sono tra le più utilizzate nell'elaborazione delle immagini a basso livello (riduzione del rumore, miglioramento di qualità, estrazione di contorni, ecc...).

Il risultato di un'operazione locale è sempre un'immagine, che ha perso però delle informazioni, dato che il livello di grigio iniziale del pixel, sottoposto a elaborazione locale, non può più essere recuperato.

Queste operazioni sono eseguite a livello locale perché applicare un filtro a tutti i pixel di una immagine è una operazione che richiede lunghi tempi di elaborazione: si pensi che, per applicare una maschera 3x3 ad una immagine da 512x512 richiede 9 moltiplicazioni e 8 addizioni in ogni posizione, per un totale di 2.359.296 moltiplicazioni e 2.097.152 addizioni. Negli anni passati, per superare questi problemi è

291 Cfr. Roberto Marangoni, Marco Geddo, *Le Immagini digitali*, Op. cit., pp. 40-41

292 <http://jada1.unime.it/visilab/AppuntiCV/operazioni.htm> 20 gennaio 2007

stato sviluppato hardware dedicato capace di effettuare operazioni logico-aritmetiche in parallelo, a velocità di video-frame (consentendo quindi l'elaborazione in tempo reale di filmati). Oggigiorno, anche elaboratori di fascia medio-bassa, visto il generale aumento delle performance computazionali, possono essere usati allo scopo.

Operazioni logiche

Le operazioni logiche usate nell'elaborazione delle immagini sono quelle booleane: *AND*, *OR*, *NOT*.

Ogni altra operazione logica può essere ricavata tramite una loro combinazione: è per questo che esse sono definite complete.

Si possono applicare esclusivamente alle immagini binarie e tra pixel omologhi (cioè aventi le stesse coordinate) di due immagini.

Sono utili per le operazioni di estrazione di caratteristiche, analisi della forma, applicazione di maschere, ecc...

Filtri

L'operazione di soglia è un esempio di filtro che può essere applicato alle immagini. I filtri sono operazioni che possono essere eseguite sia punto per punto, sia locali²⁹³.

Il termine filtro prende origine dall'analisi dei segnali, dove le operazioni in generale vengono definite *filtraggi*, mentre *filtro* indica l'operazione specifica messa in atto.

Essi possono essere utilizzati per migliorare l'immagine, sia sopprimendo il rumore, sia mettendo in risalto alcune caratteristiche. Per riuscire in tale compito essi agiscono sulla parte alta dello spettro di frequenza dell'immagine, dove si posizionano i punti isolati o le piccole macchie.

293 Cfr. Roberto Marangoni, Marco Geddo, *Le Immagini digitali*, Op. cit., p. 40

Per eliminare il rumore si utilizzano dei filtri passa-basso o il motion blur (nel caso in cui l'operatore si sia mosso durante lo scatto) o del rumore statico per sequenze d'immagini in movimento.

Filtri passa-basso – riduzione del rumore

Essi eliminano le alte frequenze lasciando intatte le basse. Poiché i punti di alta frequenza spaziale coincidono anche con i contorni degli oggetti, se le alte frequenze vengono eliminate “indiscriminatamente” questi risultano sfocati, e l'immagine risulta poco nitida.

I filtri passa-basso rimuovono alcuni tipi di rumore comuni, come²⁹⁴:

- *Sale e pepe*: si tratta di punti variamente dispersi nell'immagine con luminosità bianca e nera.
- Rumore da *impulsi*: si tratta di punti bianchi sparsi per l'immagine.
- Rumore *gaussiano*: diversamente dai precedenti il rumore è dovuto a una distribuzione normale o gaussiana delle variazioni di luminosità.

I filtri passa-basso più utilizzati sono:

1. Filtro di media locale

Si esegue assegnando la media dell'intorno di un pixel nell'immagine di origine al corrispondente nell'immagine derivata.

Si tratta di un filtro passa basso in quanto prendendo i valori medi dei livelli di

294 Cfr. Ramesh Jain, Rangachar Kasturi, Brian G. Schunck, *Machine Vision*, Op. Cit., p. 118

grigio, automaticamente si riducono le brusche variazioni di luminosità; tuttavia, se l'intorno è troppo ampio, vengono sfuocati i contorni.

Il contributo al valore del pixel derivato può venire in modo uguale da tutti i pixel dell'intorno oppure, se ve n'è necessità, in modo differenziato/pesato. L'importante, se non si vogliono alterare i valori globali dell'immagine, è che la somma dei valori dell'intorno sia pari a 1.

$1/9$	$1/9$	$1/9$	$1/10$	$1/10$	$1/10$	$1/16$	$1/8$	$1/16$
$1/9$	$1/9$	$1/9$	$1/10$	$2/10$	$1/10$	$1/8$	$1/4$	$1/8$
$1/9$	$1/9$	$1/9$	$1/10$	$1/10$	$1/10$	$1/16$	$1/8$	$1/16$

Figura 21 Esempi di filtri locali passa-basso, a sinistra il filtro di media non pesato; al centro un filtro di media con maggiore importanza al pixel centrale; a destra il filtro attribuisce via via meno valore al pixel centrale, a quelli che sono in verticale e orizzontale per finire su quelli delle diagonali.

2. Filtro di mediana.

È un modo per ottenere l'eliminazione del rumore ad altissime frequenze lasciando quasi inalterata l'immagine.

Il filtro considera l'intorno del pixel disponendoli in ordine crescente o decrescente. Il valore centrale della lista (la mediana) sarà il valore del pixel nell'immagine di destinazione. Questo filtro non sostituisce il valore medio ma, più semplicemente, “sopprime” i valori troppo scostanti dalla media dell'intorno.

Il filtro di mediana è più accurato del precedente filtro di media locale. Può essere ancora migliorato, applicandolo solo se la differenza tra il valore della mediana dell'intorno del pixel e quello di partenza è inferiore ad una determinata soglia. In questo modo il filtro è più selettivo, elimina solo il rumore e non incide su altre caratteristiche dell'immagine²⁹⁵.

295 Cfr. Roberto Marangoni, Marco Geddo, *Le Immagini digitali*, Op. cit., pp. 56-57

3. Filtri Gaussiani – *Smoothing*

Sono una classe di filtri per la distribuzione lineare i cui pesi sono scelti in base alla funzione di Gauss. Essi sono ottimi filtri per la rimozione del normale rumore, assimilabili ai filtri di media, ma con alcune proprietà particolari come il fatto che il peso decresce verso zero all'allontanarsi dall'origine (ciò significa che i valori dell'immagine vicini alla posizione centrale sono più importanti dei valori più remoti)²⁹⁶.

Filtri passa-alto – aumento del dettaglio

I filtri possono essere utilizzati anche per evidenziare i contorni degli oggetti, perciò aumentare il contenuto delle alte frequenze dello spettro e diminuire le basse.

L'uso di filtri passa-alto per evidenziare i particolari fini dell'immagine è detto *sharpening* (rafforzamento), in quanto aumenta il contrasto medio dell'immagine.

-1	-1	-1	0	-1	0	1	-2	1
-1	+9	-1	-1	+5	-1	-2	+5	-2
-1	-1	-1	0	-1	0	1	-2	1

Figura 22 Esempi di filtri passa alto per l'esaltazione di contorni

Rilevazione dei bordi

I primi stadi del processo visivo identificano le caratteristiche delle immagini che sono rilevanti per stimare la struttura e le proprietà degli oggetti della scena. I bordi, una di queste caratteristiche, sono variazioni locali significative nell'immagine e quindi estremamente utili per la sua analisi. Normalmente si trovano tra i margini/limiti di due regioni diverse. Riuscire ad individuarli è il primo passo per poter estrarre ulteriori informazioni dall'immagine, e associarle ad oggetti effettivamente presenti nella scena²⁹⁷.

²⁹⁶ Cfr Linda G. Shapiro, George C. Stockman, *Computer Vision*, Op. Cit., p. 151

²⁹⁷ Cfr. Ramesh Jain, Rangachar Kasturi, Brian G. Schunck, *Machine Vision*, Op. Cit., p. 140

Può essere una base di partenza il considerare che i punti che definiscono un bordo sono quelli posizionati nella zona delle alte frequenze spaziali e che un filtro passa-alto è in grado di evidenziarli. Ma per estrarre un bordo è necessario individuare solo i pixel che appartengono allo stesso, eliminando gli altri: costruire cioè un'immagine binaria con tutti i pixel appartenenti al contorno aventi valore 255 e gli altri 0. Operando solo sulle frequenze sarebbe sufficiente eliminare quelle basse e mantenere quelle alte. Spesso, però, si preferisce utilizzare dei filtri direttamente sull'immagine, o meglio sul dominio spaziale²⁹⁸.

Gli algoritmi che si occupano della rilevazione dei bordi sviluppano solitamente le seguenti fasi²⁹⁹:

1. **Filtraggio:** dato che il calcolo dei gradienti si basa sui valori d'illuminazione, solo due punti sono soggetti a rumore o altri accidenti nei calcoli discreti; i filtri sono quindi utilizzati per migliorare l'esecuzione del rilevatore di contorni con riferimento al rumore. Tuttavia è necessario valutare la rilevanza dei bordi e la quantità di rumore. Un filtro troppo forte riduce spesso anche i margini.
2. **Miglioramento:** per facilitare la rilevazione dei bordi è essenziale determinare i cambiamenti d'intensità vicino al punto. È necessario evidenziare i pixel dove vi sono significativi cambi nei valore locali d'intensità, e questo normalmente avviene valutando l'ampiezza del gradiente.
3. **Rilevamento:** dato che molti punti dell'immagine hanno valori diversi da zero, e non tutti sono bordi, è necessario individuare un metodo per individuare quali punti sono contorni. A tal fine si usano, solitamente, i criteri di soglia.
4. **Localizzazione:** è un algoritmo incluso solo in alcuni rilevatori. Si tratta di valutare la posizione del bordo basandosi sulla risoluzione, richiesta da alcune applicazioni, dei sub-pixel. Può essere rilevato l'orientamento del bordo.

298 Cfr. Roberto Marangoni, Marco Geddo, *Le Immagini digitali*, Op. cit., p. 64

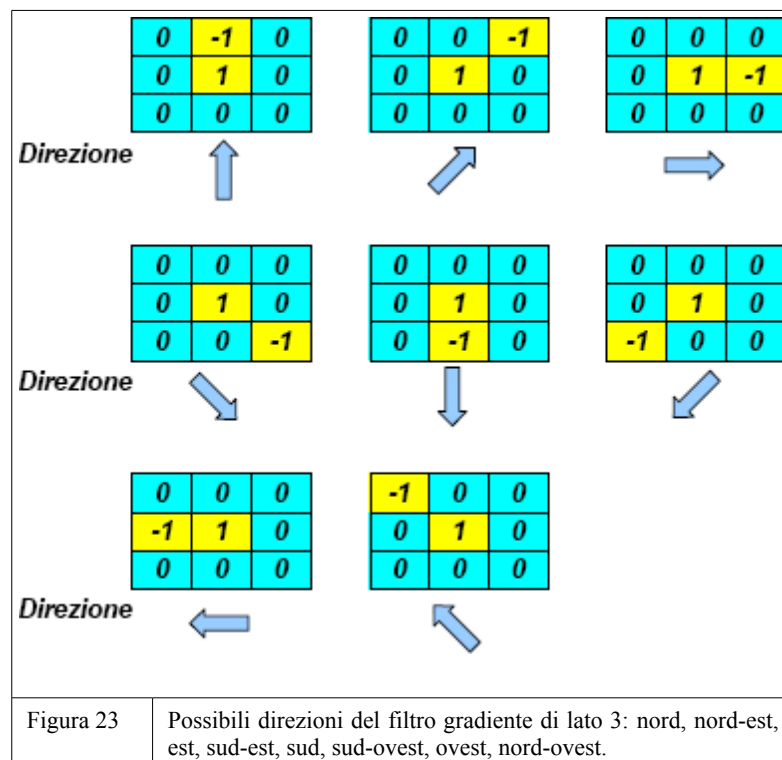
299 Cfr. Ramesh Jain, Rangachar Kasturi, Brian G. Schunck, *Machine Vision*, Op. Cit., pp. 145-146

I filtri normalmente utilizzati per rilevare i bordi sono i seguenti³⁰⁰:

1. Filtri gradiente

Prendono il nome dall'operatore gradiente che, nelle funzioni continue, esprime per ogni punto, la variazione della funzione. Poiché le immagini digitali sono discrete, e non continue, è necessario applicare gli operatori discreti analoghi ad esso, ovvero il rapporto incrementale.

Basandosi sul fatto che i bordi risultano spesso in brusche variazioni della luminosità, essi cercano di massimizzare le differenze esistenti tra pixel consecutivi, in modo da assegnare al pixel dell'immagine trasformata un valore tanto maggiore, quanto più ampia è la differenza tra i pixel considerati nell'immagine d'origine. La *direzione* del filtro dipende dai pixel considerati, come è possibile vedere in figura 23.



300 Cfr. Roberto Marangoni, Marco Geddo, *Le Immagini digitali*, Op. cit., pp. 64-68

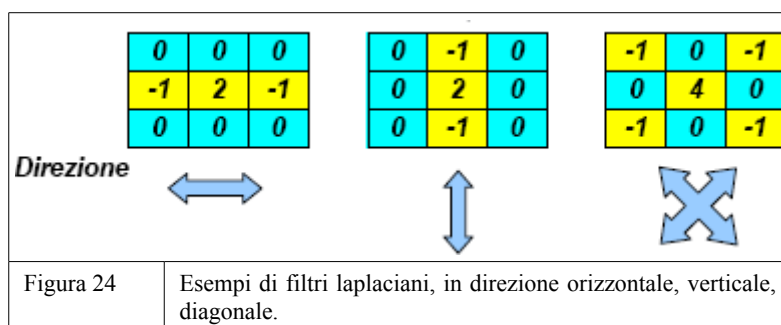
Da notare che la somma delle maschere dei filtri gradiente o, comunque, dei filtri per l'estrazione di contorni, è pari a zero e non a uno, in quanto si vuole esaltare i contorni. Per questo l'immagine derivata è tutta buia, tranne che per i pixel dei contorni che sono chiari.

Questo tipo di filtri presenta comunque alcuni problemi: essi tendono ad amplificare qualunque variazione brusca della luminosità (quindi di rumore e falsi contorni inclusi) e, peggio, non sempre individuano tutti i contorni, in quanto ignorano spesso le transizioni oggetto/sfondo (o oggetto/oggetto) cui corrispondono limitate variazioni della funzione di luminosità.

2. Filtro laplaciano.

È stato studiato per superare alcuni dei problemi accennati dei filtri gradiente.

Il filtro laplaciano si basa sull'omonimo operatore. Dal punto di vista delle funzioni continue, quest'ultimo è una derivata seconda, ossia misura la variazione della variazione (prima derivata) della funzione di luminosità. Poiché questo filtro si basa sulla rapidità con cui si passa dall'oggetto allo sfondo, esso risulta meno sensibile al rumore e più efficace nell'identificare i contorni. Come il precedente può essere orientato in diverse direzioni, riportate in figura.



3. Filtro di Prewitt

Il filtro di Prewitt adotta un approccio di tipo gradiente, ma usa maschere in cui

sono contrapposte tra loro delle zone di pixel, anche orientate.

4. Filtro di Sobel

Il filtro di Sobel, di tipo non-lineare, fornisce ottimi risultati. Data la sua complessità, ne si omette la spiegazione dettagliata.

Rilevazione delle linee

Una volta rilevati i bordi è importante riuscire a ricondurli a delle linee. Una linea può essere modellata come due bordi vicini tra loro ma con opposta polarità, separati da una distanza irrilevante. Sebbene sia un'operazione molto semplice, al lato pratico presenta difficoltà non indifferenti, tanto che una soluzione univoca non è stata ancora trovata. L'implementazione in un singolo sistema della rilevazione di bordi e linee sembra ancora un problema insormontabile³⁰¹.

Estrazione di contorni

I bordi devono essere in un secondo tempo collegati per rappresentare i confini di una regione. Questa rappresentazione è detta contorno. I contorni possono essere aperti o chiusi. Quelli chiusi corrispondono ai confini di una regione, e i pixel appartenenti alla regione possono essere individuati da un apposito algoritmo. Un contorno aperto può essere parte del confine di una regione. Quest'ultima può contenere dei salti (ossia vi sono dei punti vuoti nella regione) in quanto il contrasto tra regioni può non essere rilevato a causa della debolezza della sua intensità. Questo in quanto la soglia del rilevatore di bordi può essere stata impostata ad un valore troppo alto oppure il contrasto lungo determinate posizioni di confine può essere troppo debole rispetto a regioni confinanti dell'immagine cosicché le singole soglie non risultino funzionali per tutta l'immagine. Contorni aperti sono presenti anche quando frammenti di linee sono collegati assieme, ad esempio in un dipinto al tratto o in un manoscritto³⁰².

Un contorno può essere rappresentato sia da una lista ordinata di bordi o da una curva.

301 Cfr. Ramesh Jain, Rangachar Kasturi, Brian G. Schunck, *Machine Vision*, Op. Cit., pp. 180-181

302 Ibidem pp. 186-187

Una curva è un modello matematico di un contorno. Esempi di curve includono segmenti di linee o funzioni cubiche spline³⁰³. Ci sono alcuni criteri per una buona rappresentazione dei contorni:

1. Efficienza: il contorno dovrebbe avere una rappresentazione semplice e compatta.
2. Accuratezza: il contorno dovrebbe adattarsi in modo accurato alle caratteristiche dell'immagine.
3. Efficace: il contorno dovrebbe essere appropriato per le operazioni che saranno successivamente effettuate.

L'accuratezza della rappresentazione è determinata dalla forma della curva usata per modellarla, dalla performance dell'algoritmo usato per produrla e dall'accuratezza del rilevatore di bordi. La più semplice rappresentazione di un contorno è una lista ordinata dei suoi bordi. Questa è tanto più accurata quanto più è la stima della posizione dei bordi, ma risulta essere così la forma meno compatta e può essere inutilizzabile per la successiva analisi dell'immagine. Trovare il modello di curva che meglio si adatta ai bordi ne aumenta l'accuratezza, poiché gli errori nella posizione dei bordi sono ridotti attraverso la media, e l'efficienza ne è accresciuta attraverso una più appropriata e più compatta rappresentazione per le successive operazioni. Ad esempio, un insieme di bordi che sono vicini ad una linea possono essere rappresentati in modo più efficiente adattando questa ai bordi. Così facendo si semplificano i calcoli successivi quali: determinare l'orientamento o la lunghezza della linea, accrescerne l'accuratezza (l'errore quadratico medio tra la linea stimata e la linea "vera" sarà minore dell'errore tra la linea "vera" e ognuno dei bordi)³⁰⁴.

Vi sono anche altri tipi di approcci all'estrazione di contorni, suggeriti da alcune scoperte della psicofisica moderna, che si basano sulle trasformate di Fourier. Il procedimento è più complesso dell'individuazione delle differenze tra i livelli di grigio.

303 http://it.wikipedia.org/wiki/Interpolazione_spline 28 gennaio 2007

304 Cfr. Ramesh Jain, Rangachar Kasturi, Brian G. Schunck, *Machine Vision*, Op. Cit., pp. 186-187

Dato che questo metodo valuta l'intera immagine e solo successivamente individua dei punti localizzati, lo si esaminerà trattando le operazioni globali.

Texture

Le textures (“tessiture”) giocano un ruolo importante in molti sistemi di computer vision, come l'ispezione di superfici, la classificazione di scene, la determinazione dell'orientamento e della forma delle superfici³⁰⁵.

La texture è caratterizzata dalla distribuzione spaziale di livelli di grigio in una zona. Esse non possono essere perciò definite in un punto. La risoluzione alla quale un'immagine è osservata determina come la texture è percepita. Ad esempio, osservando un'immagine di un pavimento piastrellato da grande distanza si possono notare le strutture formate dalle disposizioni delle piastrelle, mentre sfuggono i disegni, formati dalle singole unità. Quando la stessa scena è osservata a distanza ravvicinata, così che solo alcune piastrelle rientrino nel campo visivo, si percepiscono gli intrecci formati dalla disposizione dei disegni dettagliati che compongono ogni piastrella³⁰⁶.

Si può quindi definire una texture come una ripetizione di un modello locale di variazione d'intensità, che è troppo debole per essere distinto, alla risoluzione da cui lo si osserva, come oggetto separato. Un insieme di pixel connessi che soddisfa una data proprietà di grigio che si ripete in una data regione dell'immagine costituisce quindi una zona di texture. Un testo su foglio bianco può essere considerato una texture. Ogni carattere è dato da un insieme di pixel connessi e che hanno un determinato valore di grigio o colore. Disponendo i caratteri in linea e le linee in sequenza, come elementi della pagina, si ha una texture ordinata³⁰⁷.

Nell'analisi delle textures si devono affrontare principalmente tre problemi³⁰⁸:

305 Ibidem p. 234

306 Ibidem

307 Ibidem p. 235

308 Ibidem p. 235-236

1. Classificazione

Il problema riguarda l'identificazione della texture in esame da un insieme dato di classi. L'algoritmo d'analisi ricava le caratteristiche distintive da ogni area per favorire la classificazione del modello. Ad esempio, da delle foto aeree si possono estrarre i modelli per i terreni agricoli, le foreste, le zone urbane, ecc. Per semplicità spesso si assume che i margini tra le varie regioni siano già stati determinati. Nella classificazione delle strutture, i metodi statistici sono usati in modo estensivo.

2. Segmentazione

È il problema opposto al precedente. Se prima si rilevavano le caratteristiche di una regione per classificarla, qui si segmenta/suddivide automaticamente l'immagine determinando, sempre automaticamente, i margini che dividono le varie regioni di texture.

3. Rilevazione della forma dalle textures.

Grazie ad alcune variazioni sul piano dell'immagine riguardanti le proprietà delle textures, come densità, grandezza, e orientamento, è possibile ricavare informazioni sulla forma e sull'orientamento delle superfici.

Riassumendo, una texture si può definire o come una serie di elementi (*texel* = TEXTure Element) uniti tra loro in modo regolare e ripetuto o come la misurazione della disposizione quantitativa dell'intensità in una data regione.

Nel primo caso, considerato come un approccio strutturale, per costruire e analizzare la texture è sufficiente individuare la forma degli elementi e le relazioni. È il metodo utilizzato nella maggior parte delle immagini artificiali.

Il secondo caso, denominato approccio statistico, risulta più interessante, in quanto

applicabile alle immagini reali. Infatti, quantità numeriche o statistiche che descrivono una texture possono essere calcolate partendo da un'immagine a scala di grigi (o a colori) autonomamente. Questo approccio, forse meno intuitivo, è efficiente in termini di calcolo, e può andar bene sia per la segmentazione sia per la classificazione delle textures³⁰⁹.

Operazioni globali

Si è già accennato come, in generale, sia possibile acquisire delle informazioni sull'immagine considerata attraverso l'istogramma dei livelli di grigio (o comunque, dei colori) e lo spettro di frequenze spaziali. Questi strumenti possono essere utilizzati per modificare l'immagine, come spiegato nei paragrafi seguenti.

Operazioni con l'istogramma d'intensità

L'istogramma d'intensità indica la distribuzione quantitativa dei livelli di grigio presenti nell'immagine. Queste informazioni possono essere utilizzate per determinare automaticamente il valore di soglia da applicare a una determinata immagine, e quindi alla segmentazione della stessa³¹⁰.

Come già accennato, per riuscire in questo compito l'algoritmo di segmentazione deve avere conoscenza degli oggetti presenti nella scena, del loro possibile utilizzo e dell'ambiente circostante. Questa conoscenza in particolare può riguardare:

1. L'indice di rifrazione (luminosità caratteristica) degli oggetti.
2. Le dimensioni degli oggetti.
3. La frazione dell'immagine occupata dagli oggetti.
4. Il numero di tipi differenti d'oggetti appartenenti all'immagine.

309 Cfr Linda G. Shapiro, George C. Stockman, *Computer Vision*, Op. Cit., p. 213-214

310 Cfr. Ramesh Jain, Rangachar Kasturi, Brian G. Schunck, *Machine Vision*, Op. Cit., p. 77

Un algoritmo che utilizzi questo tipo di conoscenza e arrivi a determinare autonomamente il corretto valore di soglia può essere definito automatico³¹¹.

Ecco alcuni algoritmi applicati ad oggetti che, per semplicità, si suppongono scuri perché illuminati da una luce di fondo. In questo modo si possono attribuire agli oggetti i valori di grigio sotto una certa soglia, allo sfondo quelli sopra la soglia stessa³¹²:

1. Metodo *P-Tile* (percentuale di copertura).

È forse il metodo più semplice, in quanto parte dalla conoscenza dell'area o delle dimensioni degli oggetti per stabilire la soglia da applicare all'immagine. Si suppone che in una data applicazione gli oggetti occupino una determinata percentuale dell'immagine. In questo modo si può suddividere l'istogramma dell'immagine di partenza e scegliere una o più soglie per assegnare determinate percentuali di pixel agli oggetti.

Questo metodo ha un uso chiaramente limitato. Solo poche applicazioni, come i lettori di pagine, permettono di applicarlo alla generalità dei casi.

2. Metodo della moda.

Se gli oggetti dell'immagine hanno gli stessi valori di grigi, diversi dallo sfondo, e non vi è rumore, si può assumere che l'istogramma presenti due picchi separati, quindi la soglia può essere individuata in un punto qualsiasi tra questi due valori. Nella pratica questo non succede, in quanto, per vari fattori sono presenti molti valori intermedi di grigio, con la conseguenza che non si avranno picchi separati ma delle creste d'onda, è possibile, tuttavia collocare la soglia nella valle tra due creste dell'istogramma.

Il problema d'individuare picchi e valli non è insignificante e molti sono i tentativi

311 Ibidem

312 Ibidem pp. 78-85

di risolverlo. Uno è di ignorare i picchi locali, considerando solo quelli che si trovano a certe distanze. La rilevazione dei picchi si basa sull'altezza degli stessi e sulla profondità delle valli, mentre la distanza tra valli e picchi è ignorata.

Questo approccio può essere generalizzato alle immagini aventi molti oggetti con diversi valori medi di grigio.

3. Selezione interattiva della soglia

Si stabilisce una soglia iniziale che viene successivamente affinata. Ci si aspetta che dalla soglia applicata alla prima immagine se ne ricavi una che possa essere usata per individuare una soglia migliore. Dall'algoritmo di modifica dipende il successo di questo approccio.

4. Soglia adattativa.

Se l'illuminazione della scena è ineguale, lo schema precedente può non essere adatto. La mancata uniformità può essere dovuta a delle ombre o alla direzione dell'illuminazione. In questi casi gli stessi valori di soglia non possono essere usati per tutta l'immagine.

L'approccio adattativo analizza l'istogramma dell'intera immagine, per poi suddividerla in sotto immagini. Di ognuna di queste si calcola poi la soglia, basandosi sui relativi istogrammi. La segmentazione finale dell'immagine scaturirà dalla riunione del tutto.

5. Soglie variabili.

Sempre nel caso di illuminazione ineguale, si possono approssimare i valori d'intensità dell'immagine attraverso semplici funzioni piane o biquadratiche. La funzione adatta è in gran parte determinata dai valori di grigio dello sfondo.

L'istogramma e la soglia possono essere valutati in relazione al livello base dato dalla funzione appropriata.

6. Doppia soglia.

In molte applicazioni, i valori di grigi appartenenti agli oggetti sono conosciuti. Vi possono infatti essere valori di grigio aggiunti appartenenti sia agli oggetti che allo sfondo. In questi casi si può utilizzare una prima soglia per ricavare gli oggetti principali e un secondo metodo per migliorarli. Spesso, quest'ultimo è una seconda soglia. Con essa si dovrebbero individuare i pixel che hanno un vicino sicuramente appartenente all'oggetto oppure, usando le caratteristiche di luminosità rilevate con l'istogramma, individuare i punti da includere nella regione dell'oggetto.

Questo algoritmo implementa i principi di somiglianza e prossimità spaziale. I pixel dei margini hanno valori vicini ai pixel “*core*” già appartenenti all'oggetto, dato che i due insiemi di pixel sono adiacenti nell'istogramma, e sono anche vicini nello spazio in quanto confinanti.

La limitazione più pesante incontrata dal metodo di segmentazione basato sull'istogramma è la perdita delle informazioni relative alla posizione dei valori d'intensità dell'immagine. Molte immagini con diversa distribuzione spaziale possono avere istogrammi d'intensità simili. La natura globale dell'istogramma limita la sua applicabilità a scene complesse. Esso non evidenzia il fatto importante che punti dello stesso oggetto solitamente sono vicini in quanto appartenenti alla stessa superficie³¹³. Per questo è necessario operare sullo spettro di frequenza spaziale.

Un ulteriore utilizzo dell'istogramma è la redistribuzione dei valori di grigio in quelle immagini che hanno i valori d'intensità posizionati in un raggio ridotto, ad esempio quelle a basso contrasto. Questa operazione è definita equalizzazione dell'istogramma e permette di aumentare il contrasto e spesso, di conseguenza, la qualità dell'immagine³¹⁴.

313 Ibidem p. 86

314 Ibidem p. 112

La trasformata di Fourier

La trasformata di Fourier permette di modificare le immagini agendo sullo spettro di frequenza. Questo è possibile in quanto l'immagine è riconducibile ad un segnale.

Un segnale è definito come una successione di valori discreti o continui derivati dalla misurazione della variazione di una qualche grandezza nel tempo e nello spazio. Ricordando come si è ottenuto il reticolo da una singola linea, si può dire che un'immagine è un segnale bidimensionale discreto derivato dal campionamento spaziale della funzione di luminosità³¹⁵.

Al fine di analizzare il contenuto in frequenze spaziali dell'immagine è necessario utilizzare la trasformata di Fourier discreta, ossia generare un segnale bidimensionale discreto derivato. Il segnale così ottenuto sarà composto da numeri complessi, dove nella parte reale sono contenute le informazioni relative alla frequenza, mentre nella parte immaginaria quelle relative alla fase³¹⁶.

Il modulo (valore assoluto) di questa funzione bidimensionale discreta è, quindi, una funzione bidimensionale reale, che rappresenta il contenuto delle frequenze spaziali dell'immagine di partenza. E' possibile utilizzare questa definizione di spettro di frequenza e sostituirla alla precedente, ottenuta con un metodo euristico³¹⁷. Dato che il modulo è un segnale, esso può a sua volta essere rappresentato da un'immagine e, quindi, manipolato con un qualunque programma di elaborazione delle immagini, agendo direttamente sui valori dei pixel³¹⁸.

Grazie alla trasformata di Fourier si possono eseguire importanti operazioni

315 Cfr. Roberto Marangoni, Marco Geddo, *Le Immagini digitali*, Op. cit., p. 48

316 Ibidem

317 L'euristica (dal greco εὐρίσκω, heurisko, letteralmente "scopro" o "trovo") è una parte dell'epistemologia e del metodo scientifico.

È quella parte della ricerca il cui compito è di favorire l'accesso a nuovi sviluppi teorici o scoperte empiriche. Si definisce infatti procedimento euristico un metodo di approccio alla soluzione dei problemi che non segue un chiaro percorso, ma si affida all'intuito e allo stato temporaneo delle circostanze, al fine di generare nuova conoscenza. È opposto al procedimento algoritmico.

<http://it.wikipedia.org/wiki/Euristica> 25 gennaio 2007

318 Cfr. Roberto Marangoni, Marco Geddo, *Le Immagini digitali*, Op. cit., p. 49

sull'immagine quali:

1. Rimuovere il rumore ad alte frequenze dall'immagine.
2. Estrarre le caratteristiche strutturali (*texture*) che possono essere utilizzate per individuare la tipologia di oggetti presenti in una regione dell'immagine.
3. Compressione delle immagini.

Il primo punto è già stato in parte trattato nel paragrafo relativo ai filtri.

Il secondo è un approccio alternativo all'estrazione dei contorni e di altre caratteristiche dell'immagine suggerito dalle teorie della psicofisica. Consiste, molto semplicemente, nello scomporre l'immagine in serie di Fourier (l'immagine viene rappresentata attraverso la somma di funzioni periodiche di seno e coseno) e nel costruire la funzione rilevando i punti in cui le fasi delle armoniche concordano. Secondo queste ricerche, i contorni dovrebbero coincidere con i punti in cui la concordanza di fase ha un massimo locale. Questo approccio, nonostante la complessità, sembra essere quello più vicino alla visione umana, dato che produce illusioni simili.

Proprio per questo, e si è già al terzo punto, la trasformata di Fourier è utile per comprimere le immagini. In effetti, lo standard JPEG si basa su di essa, permettendo di considerare solo i dettagli dell'immagine importanti per la visione umana e diminuendo, parallelamente, lo spazio occupato in memoria³¹⁹.

Tra le operazioni che possono essere compiute in quello che possiamo definire lo spazio di frequenza (l'immagine rappresentante la funzione bidimensionale reale) vi è la convoluzione³²⁰. Questa operazione consiste nel far scorrere una maschera (un'altra matrice contenente dei valori definiti *pesi*) sull'immagine, centrandola, in sequenza, su ciascun pixel; calcolare il prodotto tra i valori dei pixel e i pesi relativi alle varie

319 Ibidem p. 71

320 Ibidem p. 49

posizioni e, infine, sommare tutti i pixel considerati³²¹. Normalmente tale operazione è impiegata in modo analogo ai filtri, con la differenza che si considera l'intera immagine. Essa è importante in quanto, per una delle proprietà della trasformata di Fourier, se la si applica al dominio spaziale è come si eseguisse il prodotto delle trasformate nel dominio di frequenza. L'equivalenza delle due operazioni è garantita dal fatto che la trasformata di Fourier è invertibile (*antitrasformata* di Fourier). Calcolare un prodotto è molto più efficiente a livello computazionale che calcolare una convoluzione, utilizzando la trasformata di Fourier per passare da un dominio all'altro si possono diminuire i tempi di elaborazione.

Oltre alle operazioni che modificano l'intera immagine, ve ne sono alcune che operano a livello globale, piuttosto che locale o puntuale, per estrarre informazioni riguardanti le superfici, le dimensioni, ecc

Estrazione di superfici dalle ombre

E' possibile costruire una mappa dell'indice di riflessione dove venga registrata la luminosità dei pixel in funzione dell'orientamento della superficie su cui giacciono i punti nella scena reale. In tal modo, avendo illuminazioni fisse durante la formazione dell'immagine e conoscendo l'indice di riflessione della superficie, si possono tradurre le variazioni dell'orientamento in variazioni d'illuminazione dell'immagine stessa³²².

Ponendo alcuni vincoli, ad esempio che si tratti di superfici lisce, è successivamente possibile risalire alle loro forme. Questo costituisce comunque uno svantaggio, in quanto le superfici reali non sono sempre lisce³²³.

Stereo-Fotometrica

Il metodo della stereo-fotometrica registra immagini multiple (almeno due, di norma

321 Ibidem p. 42

322 Cfr. Ramesh Jain, Rangachar Kasturi, Brian G. Schunck, *Machine Vision*, Op. Cit., p. 269

323 Ibidem

tre) di un oggetto illuminato in sequenza, da differenti sorgenti di luce³²⁴. Questo al fine di ricostruire più parti di superfici e misurarne le dimensioni delle forme-oggetti inferendone la profondità. Per riuscire in quest'operazione si stabiliscono dei vincoli relativi all'illuminazione della scena³²⁵ o alla fissità degli oggetti.

Uno dei vantaggi di questo metodo consiste nel fatto che i punti delle immagini sono perfettamente registrati l'uno con l'altro: non hanno cioè problemi di corrispondenza, in quanto sia la fotocamera sia la scena sono fissi. D'altra parte, questo costituisce anche uno svantaggio, in quanto condizioni così rigide sono difficili da realizzare e/o controllare. Il metodo rileva inoltre la profondità solo in modo indiretto.

La profondità

E' risaputo che nella rappresentazione 2D della scena 3D si perde la dimensione della profondità e che il processo inverso costituisce un problema non immediatamente risolvibile. Calcolare la distanza relativa di vari punti della scena dal punto di acquisizione è uno dei compiti più importanti per un sistema di computer vision. Uno dei metodi più utilizzati per riuscirvi è acquisire una copia di immagini utilizzando due macchine fotografiche di cui si conosca la distanza che le divide, in pratica simulando una stereo-visione. In alternativa si può utilizzare una telecamera, spostandola secondo necessità³²⁶.

Esistono anche le immagini spaziali (*range image*), in cui la profondità è rilevata direttamente grazie a sensori che si basano sui principi del radar e della triangolazione.

Profondità dalle immagini d'intensità

Nel caso si voglia simulare la stereo-visione (ossia acquisendo due immagini, una a destra, l'altra a sinistra), è sufficiente, per comprendere come i punti della scena 3D siano localizzati nello spazio, utilizzare la geometria e l'algebra. La misurazione avviene attraverso la tecnica della triangolazione, misurando la parallasse tra due immagini

324 Cfr Linda G. Shapiro, George C. Stockman, *Computer Vision*, Op. Cit., pp. 471-472

325 Cfr David A. Forsyth, Jean Ponce, *Computer Vision: A Modern Approach*, Op. Cit., pp. 81-82

326 Cfr. Ramesh Jain, Rangachar Kasturi, Brian G. Schunck, *Machine Vision*, Op. Cit., p. 289

ottenute riprendendo l'oggetto da due posizioni diverse, separate da una distanza sufficiente³²⁷.

Il metodo della parallasse è descritto molto semplicemente da Asimov nel suo libro di Fisica: “Un metodo per calcolare le distanze cosmiche è quello basato sulla *parallasse*. Il significato di questo termine è facile da spiegarsi: mettete un dito alla distanza di una decina di centimetri dagli occhi e guardatelo prima con l'occhio sinistro, poi con il destro; vedrete il vostro dito spostarsi rispetto allo sfondo, perché avete cambiato il vostro punto di vista. Se ora ripetete lo stesso procedimento tenendo il dito più lontano, per esempio alla distanza del braccio teso, esso si sposterà ancora rispetto allo sfondo, ma meno di prima; l'entità dello spostamento può quindi servire a determinare la distanza del dito dai vostri occhi.

Naturalmente, se un oggetto dista una ventina di metri il cambiamento di posizione quando lo si guarda con l'uno o l'altro occhio comincia a essere troppo piccolo per essere misurato; si deve avere una <<linea di base>> maggiore della distanza tra i due occhi. Per ottenere uno spostamento maggiore del punto di vista basterà guardare l'oggetto prescelto da una determinata posizione, poi spostarsi, per esempio di qualche metro a destra, e guardare di nuovo: ora la parallasse è sufficiente per essere facilmente misurata e si può determinare la distanza. È proprio a questo metodo che si ricorre per determinare l'ampiezza di un fiume o di un burrone.”³²⁸.

La profondità di vari punti della scena può essere quindi recuperata conoscendo la distanza (disparità) tra punti corrispondenti nelle diverse immagini³²⁹.

È da notare che a causa della natura discreta delle immagini digitali, i valori della disparità sono numeri interi, a meno che particolari algoritmi siano usati per aumentarne l'accuratezza. Solitamente, però, si preferisce aumentare la distanza della <<linea di

327 Cfr Linda G. Shapiro, George C. Stockman, *Computer Vision*, Op. Cit., p. 397

328 Isaac Asimov, *Il libro di Fisica*, trad. it. di Carla Sborgi, Milano, Arnoldo Mondadori, 1986 pp. 24-25, (ed. originale *Asimov's New Guide to Science*, 1984)

329 Cfr. Ramesh Jain, Rangachar Kasturi, Brian G. Schunck, *Machine Vision*, Op. Cit., p. 291

base>> così da aumentare anche la disparità. Questa soluzione introduce, a sua volta, altre problematiche relative, ad esempio, all'ampiezza del campo visibile o alle distorsioni introdotte dalla prospettiva³³⁰.

Questa tecnica dà per implicito che si possa identificare le coppie di punti congiunti nelle immagini stereo. Tuttavia questo non è così semplice, tanto da costituire il famoso problema di *corrispondenza*: per ogni punto dell'immagine di sinistra, si trovi il corrispondente in quella di destra. Le soluzioni finora trovate si fondano sulla rilevazione di alcune caratteristiche per identificare i singoli punti, come l'appartenenza a bordi o regioni già individuate³³¹.

Oltre a quello della parallasse ci sono anche altri metodi basati su tecniche alternative che possono fornire alcune informazioni/suggerimenti relativi alla profondità. L'interposizione è, forse, una delle informazioni nascoste più importanti: gli oggetti che sono più vicini occludono parti degli oggetti che sono lontani; il riconoscimento delle occlusioni fornisce una profondità relativa (es.: una persona davanti a un muro è più vicina al sensore del muro; un uomo dietro un'auto è più lontano dell'auto)³³².

Anche le dimensioni relative degli oggetti sono importanti: un'auto lontana apparirà più piccola e lenta di un'auto vicina.

Il punto di vista da cui si osserva è rilevante per la profondità: si pensi ad una porta aperta che proietta nella retina la forma di un trapezio e non di un rettangolo; il bordo più lontano appare più corto rispetto al più vicino, a causa dell'effetto scorcio (*foreshortening*).

Le textures delle superfici, come detto, cambiano in relazione sia alla distanza dall'osservatore sia al loro orientamento.

330 Ibidem

331 Ibidem p. 293

332 Cfr Linda G. Shapiro, George C. Stockman, *Computer Vision*, Op. Cit., p. 40

Sono stati implementati, negli anni '80, vari metodi per usufruire di queste informazioni/suggerimenti. Tali metodi, tutti indiretti, acquisiscono informazioni sia sulla forma sia sulla profondità: si parla infatti, in generale, di *shape from X*, e nello specifico, di *shape from shading* (forma dalle ombre), *shape from texture*, *shape from focus* (forme dalla messa a fuoco), *shape from motion* (forma dal movimento), nonché di *photometric stereo* (stereo-fotometrica).

Si approfondirà ora l'estrazione di profondità dalla messa a fuoco, rimandando l'estrazione di forme dal movimento all'operazione sugli oggetti. Gli altri metodi sono stati già considerati.

Grazie al fatto che i sistemi ottici hanno una profondità di campo finita, solo gli oggetti a una distanza appropriata appaiono a fuoco nell'immagine, mentre altri risultano confusi/indistinti. Alcuni algoritmi sono stati realizzati per sfruttare quest'effetto. L'immagine è modellata come una convoluzione di immagini a fuoco con una funzione di diffusione del punto³³³ determinata in ragione dei parametri della fotocamera e della distanza degli oggetti dalla stessa. La profondità è recuperata dalla stima della confusione nell'immagine e utilizzando una conosciuta o stimata funzione di diffusione lineare. Questa ricostruzione presenta dei problemi di calcolo; tuttavia è utile nelle applicazioni che richiedono informazioni dettagliate sulla profondità³³⁴.

La profondità nelle immagini spaziali

Come trattato in precedenza, le immagini spaziali (o mappe di profondità) sono ottenute da apparecchi che misurano la distanza di ogni punto della scena all'interno dell'angolo visivo e li registrano come una funzione bidimensionale.

Due metodi molto usati per la formazione delle immagini spaziali sono la triangolazione e il radar (*Radio Detection And Ranging*)³³⁵.

333 http://en.wikipedia.org/wiki/Point_spread_function 1 febbraio 2007

334 Cfr. Ramesh Jain, Rangachar Kasturi, Brian G. Schunck, *Machine Vision*, Op. Cit., p. 300

335 <http://it.wikipedia.org/wiki/Radar> 1 febbraio 2007

I sistemi ad illuminazione strutturata, usati in modo estensivo in computer vision, si basano sulla triangolazione per calcolare la profondità.

I sistemi radar per la formazione delle immagini usano rilevatori (*finder*) acustici e laser (*Light Amplification by the Stimulated Emission of Radiation*, amplificazione della luce attraverso l'emissione di radiazioni³³⁶) per ottenere le mappe di profondità³³⁷.

Sistemi di Visione Attiva³³⁸

Considerati le due precedenti tipologie d'immagini e i relativi sistemi d'acquisizione, si può introdurre una riflessione sulle loro caratteristiche.

Molti sistemi hanno delle caratteristiche fissate, che includono sia sensori passivi come fotocamere e telecamere, sia attivi come rilevatori laser di profondità.

Tuttavia si sta sempre più diffondendo l'idea che sistemi di visione attiva, di natura contrapposta ai precedenti, dove i parametri e le caratteristiche di acquisizione sono dinamicamente controllati da sistemi d'interpretazione della scena, siano cruciali per la percezione della stessa. Questo è ciò che accade negli esseri viventi, dove i dati sono acquisiti in modo attivo.

I sistemi di visione attiva possono impiegare sensori attivi o passivi. Tuttavia, in questo caso, gli stadi parametrici dei sensori, come messa a fuoco, apertura, vergenza³³⁹ e illuminazione sono controllati per acquisire i dati utili a semplificare il compito d'interpretazione della scena stessa.

La visione attiva è essenzialmente un processo d'acquisizione intelligente dei dati, controllato dai parametri misurati e stimati, nonché dai possibili errori provenienti dalla

336 Per una comprensione del funzionamento del laser si veda Isaac Asimov, *Il libro di Fisica*, Op. Cit., pp. 492-498

337 Cfr. Ramesh Jain, Rangachar Kasturi, Brian G. Schunck, *Machine Vision*, Op. Cit., p. 300

338 Ibidem p. 305

339 Il movimento simultaneo di entrambi gli occhi nei sensi opposti (convergenza e divergenza) per ottenere o effettuare la singola visione binoculare. <http://en.wikipedia.org/wiki/Vergence> 1 febbraio 2007

scena.

Una precisa definizione di questi parametri dipendenti dalla scena e dal contesto richiede una completa comprensione non solo delle proprietà della formazione delle immagini e dei sistemi di elaborazione, ma anche delle loro interdipendenze.

Operazioni a livello di oggetti

Operare a livello di oggetti significa anzitutto identificarli. Le operazioni precedenti (puntuali, locali, globali), reiterabili, dovrebbero aver contribuito a migliorare l'immagine, ad evidenziarne alcune caratteristiche e a conoscerne alcune informazioni. Tuttavia, prima di riuscire a riconoscere gli oggetti contenuti in una scena è necessario “interpretare” queste informazioni. Ad esempio, selezionando e unendo i bordi al fine di ottenere i contorni, ed eseguendo una segmentazione più accurata dell'immagine. Proprio la segmentazione è strettamente collegata al riconoscimento degli oggetti: senza almeno un parziale riconoscimento degli oggetti, essa non può essere eseguita, e senza di essa gli oggetti non possono essere riconosciuti. **Segmentazione**

Si è già accennato a questo processo esaminando l'operazione di soglia, l'estrazione di bordi, la rilevazione delle textures e le operazioni con l'istogramma d'intensità; è comunque opportuno richiamarne alcuni aspetti.

Con segmentazione si indica la procedura che consente di suddividere, in base a criteri predefiniti, un'immagine in aree considerate omogenee. Per certi versi questa procedura funziona in modo opposto all'estrazione dei contorni: quest'ultima ricerca delle discontinuità nella funzione di luminosità, mentre le tecniche di segmentazione ricercano pixel con valori d'intensità simili. La maggior parte delle tecniche utilizzate, sebbene presentino differenze anche notevoli, si basano su una filosofia comune detta *region growing*³⁴⁰.

L'operazione di *region growing* procede come segue: dato un pixel di partenza, lo si

340 Cfr. Roberto Marangoni, Marco Geddo, *Le Immagini digitali*, Op. cit., p. 70

etichetta (ossia si procede al *labeling*: al pixel viene assegnato un codice interno completamente slegato dalla funzione di luminosità) e si copia tale etichetta su tutti i pixel dell'intorno, sempre che non differiscano oltre una determinata soglia dal valore del pixel di riferimento. Questa procedura è reiterata per tutti i pixel considerati e s'interrompe quando le differenze di soglia non consentano di proseguire in nessuna direzione. Regioni contrassegnate con la stessa etichetta hanno una buona probabilità di appartenere al medesimo oggetto, o quantomeno a parti di un medesimo oggetto, fornendone quindi una descrizione³⁴¹. Errori di segmentazione possono portare ad una non perfetta corrispondenza tra regioni e oggetti, ragion per cui per una corretta interpretazione dell'immagine è necessaria una conoscenza specifica degli oggetti stessi.

Da quanto detto si desume che i più importanti principi della segmentazione sono due: la somiglianza dei valori e la vicinanza spaziale. Due pixel possono appartenere alla stessa regione se hanno caratteristiche di luminosità simili o se sono vicini tra loro³⁴².

Superfici³⁴³

Le operazioni sulle superfici, come quelle sul movimento, sono a cavallo tra quelle globali e quelle sugli oggetti. Per poter risalire agli oggetti è infatti necessario considerare l'intera scena. Per questa ragione i due problemi principali relativi alle superfici nella computer vision riguardano la loro identificazione e la successiva segmentazione. Le superfici devono essere ricostruite dalle misurazioni della profondità, tenendo conto che vi possono essere oggetti a sé stanti. Poi esse vengono segmentate in vari tipi per permetterne il riconoscimento degli oggetti e per una loro migliore considerazione.

Molti oggetti 3D, specialmente manufatti, possono essere facilmente descritti in termini di forma e posizione delle superfici di cui sono costituiti. La descrizione delle superfici è utilizzata per la classificazione degli oggetti, la stima della posizione, l'ingegneria inversa, ed è onnipresente in computer graphics.

341 Ibidem p. 71

342 Cfr. Ramesh Jain, Rangachar Kasturi, Brian G. Schunck, *Machine Vision*, Op. Cit., p. 74

343 Ibidem p. 365

Il movimento

Finora si sono considerate le elaborazioni su singole immagini, o al limite su due immagini acquisite simultaneamente. Introducendo il movimento si amplia la prospettiva in quanto l'elaborazione dell'immagine abbraccia la dimensione temporale. Più precisamente, si può disporre delle informazioni visive che possono essere estratte da variazioni spaziali e temporali presenti in una sequenza d'immagini³⁴⁴.

La dimensione temporale nel processo visivo è importante principalmente per due ragioni. In primo luogo, l'apparente movimento degli oggetti sul piano dell'immagine dà una forte indicazione per comprendere struttura e movimento tridimensionale. Secondo, i sistemi visivi biologici utilizzano il movimento per inferire le proprietà dell'ambiente 3D con una bassa conoscenza *a priori* dello stesso³⁴⁵.

Il dato di partenza per un sistema di analisi di una scena in movimento è una sequenza di fotogrammi presa da un mondo in continua evoluzione. Anche la telecamera che riprende può essere in movimento. Ogni fotogramma rappresenta un'immagine della scena in un particolare istante. I cambiamenti, della scena possono essere dovuti al movimento della telecamera, allo spostamento degli oggetti, a variazioni d'illuminazione, o a quelli di struttura, dimensioni o forma degli oggetti³⁴⁶.

Normalmente si presume che le variazioni della scena siano dovute a spostamenti della telecamera o degli oggetti, e che queste siano rigide o quasi-rigide; altre variazioni non sono ammesse³⁴⁷.

Il sistema deve rilevare i cambiamenti, determinare le caratteristiche del movimento dell'osservatore e degli oggetti, caratterizzare il movimento attraverso astrazioni di alto livello, ricostruire la struttura degli oggetti e riconoscere gli oggetti in movimento³⁴⁸.

344 Cfr. Emanuele Trucco, Alessandro Verri, *Introductory Techniques for 3-D Computer Vision*, Op. Cit., p. 178

345 Ibidem

346 Cfr. Ramesh Jain, Rangachar Kasturi, Brian G. Schunck, *Machine Vision*, Op. Cit., p. 406

347 Ibidem

348 Ibidem

In applicazioni quali *video editing* e *video database* potrebbe essere richiesto di rilevare delle *macro* variazioni nella sequenza. Queste variazioni suddivideranno il segmento in parti tra loro collegate in quanto simili per tipo di movimenti della telecamera e tipo di scene nella sequenza³⁴⁹.

Per quanto riguarda le relazioni tra telecamera (o punto d'osservazione), oggetti e ambiente, si possono presentare le seguenti situazioni, ognuna delle quali richiede una diversa tecnica d'analisi³⁵⁰:

1. Camera fissa, un singolo oggetto fisso, sfondo fisso.

Inserita per completezza, è una semplice scena statica, praticamente una foto su cui applicare le tecniche/algoritmi già esposti.

2. Camera fissa, un singolo oggetto in movimento, sfondo fisso.

L'oggetto in movimento sullo sfondo comporta dei movimenti dei pixel nell'immagine associati all'oggetto. La rilevazione di questi pixel può svelare la forma dell'oggetto così come la sua velocità e percorso. Questo tipo di sensori è normalmente utilizzato per la sicurezza e la sorveglianza³⁵¹.

3. Camera fissa, più oggetti in movimento, sfondo fisso.

Il movimento di uno o più oggetti può essere tracciato per ottenere una traiettoria o un percorso dai quali sarà possibile trarre indicazioni sul comportamento dell'oggetto. È il caso di una telecamera usata per analizzare il comportamento di alcune persone che entrano in un edificio per affari o altro lavoro. Diverse telecamere possono essere utilizzate per ottenere diversi punti di vista dello stesso oggetto, permettendo quindi di elaborare percorsi tridimensionali. Possibili

349 Ibidem

350 Cfr Linda G. Shapiro, George C. Stockman, *Computer Vision*, Op. Cit., p. 252

351 Ibidem

applicazioni sono l'analisi del movimento di atleti o di pazienti in riabilitazione. Vi è anche un sistema in via di sviluppo che traccia i movimenti, durante un incontro di tennis, della palla e dei giocatori fornendo l'analisi degli elementi del gioco³⁵².

4. Camera in movimento, scena relativamente costante.

Una telecamera in movimento provoca dei cambiamenti nelle immagini dovuti al suo stesso movimento, anche se l'ambiente non cambia. Si può utilizzare questo movimento in modi diversi. Ad esempio si può ottenere una più ampia visione dell'ambiente rispetto all'osservazione da un singolo punto fisso: è il caso di un movimento panoramico di macchina. Il movimento della camera può anche fornire informazioni sulla profondità relativa degli oggetti, in quanto le immagini di quelli vicini cambiano più velocemente di quelle relative ai lontani. In terzo luogo, esso può dare la percezione o la misurazione della forma di oggetti 3D vicini: i molteplici punti di vista permettono infatti di effettuare calcoli trigonometrici simili alla visione stereo. Elaborando o analizzando il contenuto di film o video, è spesso importante rilevare la posizione e l'istante in cui la telecamera ha effettuato una panoramica o uno zoom: possiamo così ottenere delle informazioni su come, e in che modo, la scena è vista dal sistema³⁵³.

5. Camera in movimento, diversi oggetti in movimento.

Questa situazione presenta i problemi relativi al movimento più difficili da risolvere e probabilmente più importanti in quanto riguardano situazioni in cui sono in movimento i sensori ma anche una grande quantità di oggetti nella scena osservata. È il caso di un veicolo che si muove nel traffico di punta, o di alcune telecamere che devono seguire automaticamente degli oggetti in movimento³⁵⁴.

Da tener presente che una sequenza di fotogrammi offre molte più informazioni per

352 Ibidem

353 Ibidem

354 Ibidem

comprendere una scena, ma aumenta in proporzione anche la quantità di dati da elaborare. Applicare quindi tecniche per l'analisi di una scena statica ad ogni fotogramma di una sequenza occorrono elevate capacità di calcolo. Questa esigenza va ad aggiungersi alle difficoltà finora rilevate. Tuttavia, le ricerche finora condotte per l'analisi di scene dinamiche hanno fornito soluzioni che ne facilitano l'estrazione d'informazioni rispetto a quelle statiche³⁵⁵.

L'analisi di scene dinamiche avviene in tre fasi³⁵⁶:

1. Periferica.

Riguarda l'estrazione d'informazioni approssimative circa l'attività presente in una scena; tali informazioni saranno utilizzate, nelle fasi successive, per decidere quale parte della scena richiede una maggiore attenzione.

2. Attenzione.

Si focalizza sulle parti individuate dalla fase precedente per estrarre informazioni che possono essere utilizzate per riconoscere gli oggetti, analizzare il loro movimento, stendere una lista cronologica degli eventi presenti nella scena, e così via.

3. Cognitiva.

Applica alla scena la conoscenza pregressa relativa agli oggetti e ai movimenti con lo scopo di comprendere esattamente quali sono e cosa sta accadendo.

Riconoscimento degli oggetti

Un sistema per il riconoscimento degli oggetti è progettato per ricercare oggetti del mondo reale, basandosi su un'immagine che lo rappresenta e utilizzando dei modelli

355 Cfr. Ramesh Jain, Rangachar Kasturi, Brian G. Schunck, *Machine Vision*, Op. Cit., p. 407

356 Ibidem p. 408

conosciuti. Per definizione, infatti, il riconoscimento implica che le descrizioni degli oggetti, o modelli, siano già disponibili; non si può riconoscere ciò che non si conosce. Questo compito, semplice e istantaneo per l'uomo, è molto difficile per la macchina³⁵⁷.

Il problema di riconoscere gli oggetti può essere quindi ricondotto ad un problema di etichettatura/denominazione basato su modelli. Se si ha un'immagine contenente uno o più oggetti, e un insieme di etichette corrispondenti a un insieme di modelli conosciuti dal sistema, questo dovrebbe assegnare ad ogni regione, o gruppo di regioni dell'immagine, l'appropriata etichetta. È evidente l'importanza di suddividere/segmentare correttamente l'immagine al fine di denominare e riconoscere gli oggetti³⁵⁸.

Il sistema per il riconoscimento dovrebbe contenere i seguenti elementi:

1. Il database dei modelli (detto anche *modelbase*)

Contenente tutti i modelli conosciuti dal sistema. Questi dipendono dall'approccio utilizzato per il riconoscimento e possono variare da una loro descrizione qualitativa o funzionale ad una precisa informativa geometrica delle superfici. Una caratteristica è un attributo dell'oggetto rilevante per la descrizione e il riconoscimento dell'oggetto in relazione con gli altri. Dimensioni, colori e forme sono caratteristiche normalmente utilizzate.

Il database è organizzato in modo da eliminare, nella fase d'ipotesi, le informazioni relative ad oggetti che possono generare confusione³⁵⁹.

2. Il rilevatore di caratteristiche

Il rilevatore di caratteristiche applica degli operatori all'immagine e identifica la

357 Cfr. Ramesh Jain, Rangachar Kasturi, Brian G. Schunck, *Machine Vision*, Op. Cit., p. 459

358 Ibidem

359 Ibidem p. 460-461

posizione di caratteristiche che possono aiutare a formare alcune ipotesi sugli oggetti. Il sistema utilizza le caratteristiche che sono collegate agli oggetti da analizzare e ai modelli presenti nel database³⁶⁰.

3. Un generatore d'ipotesi

Questo componente utilizza le caratteristiche rilevate per inferire quali oggetti hanno buone probabilità di essere presenti nella scena. Questo permette anche d'individuare in quali parti dell'immagine questi sono posizionati e di conseguenza ridurre le aree d'analizzare³⁶¹.

4. Un verificatore di ipotesi

Utilizza i modelli degli oggetti per verificare le ipotesi, quindi elimina le forme/informazioni che possono portare in errore o confondere. Solo le forme, superfici, ecc che con tutta probabilità corrispondono realmente ad oggetti vengono selezionate e considerate come entità uniche³⁶².

Tutti i sistemi per il riconoscimento degli oggetti utilizzano dei modelli che presentano caratteristiche sia esplicite che implicite, utilizzate dai rilevatori in fase d'analisi. I componenti per la formulazione e la verifica delle ipotesi hanno un'importanza variabile a seconda dell'approccio utilizzato per il riconoscimento. Alcuni utilizzano il generatore d'ipotesi e selezionano gli oggetti solo in base alle loro probabilità di essere presenti nella scena. Un esempio sono i metodi che si basano sulla classificazione dei modelli (*pattern classification*). Dall'altra parte troviamo molti sistemi d'intelligenza artificiale che si concentrano sulla fase di verifica delle corrispondenze tra gli oggetti rilevati e le relative ipotesi.

360 Ibidem

361 Ibidem

362 Ibidem

Classificazione degli oggetti - Pattern recognition

A questo punto, è bene riassumere brevemente il percorso finora fatto. Si è visto cos'è un'immagine digitale, come acquisirla, quali operazioni eseguire per migliorarla ed estrarne informazioni, e come queste, aggiunte ad altre già conosciute presenti in un database, ci portino ad identificare degli oggetti.

Effettuato il riconoscimento, è possibile ottenere ulteriori informazioni sugli oggetti, quali le dimensioni, l'area, localizzarne il centro, ecc³⁶³; in aggiunta è possibile operare su di essi, ad esempio raggruppandoli per ottenere oggetti compositi (tecnica utilizzata per comprimere il segnale video da trasmettere a distanza³⁶⁴), ma anche per individuare dei volti umani in un'immagine³⁶⁵.

Tuttavia, aver riconosciuto degli oggetti e possedere alcune informazioni su di essi non è equiparabile alla funzione svolta dalla percezione visiva per l'uomo. Vedere ci pone in relazione con il nostro ambiente, permettendoci non solo di riconoscere, ma anche e soprattutto di utilizzare o, comunque, attribuire una funzionalità a ciò che ci circonda³⁶⁶. Semplicemente guardando, individuiamo gli oggetti utili ai nostri scopi. Se così non fosse, se la visione non permettesse di utilizzare ciò che ci circonda, se fosse limitata alla sola percezione di forme, posizioni, orientamenti, colori e ogni altra proprietà fisica, non si potrebbe far altro che navigare in un mondo tridimensionale, evitando di urtare gli oggetti di cui è composto. Nella migliore delle ipotesi si potrà riprodurre qualche oggetto, sempre che si conosca e si disponga del materiale di cui è costituito³⁶⁷.

Tutti i processi finora esaminati hanno lo scopo ultimo di percepire le funzionalità degli oggetti. È questo il vantaggio evolutivo offerto dalla visione. Per l'uomo, la capacità di utilizzare degli oggetti costituisce un tema complesso³⁶⁸, soggetto a continue disamine,

363 Cfr. Roberto Marangoni, Marco Geddo, *Le Immagini digitali*, Op. cit., p. 72

364 Ibidem p. 86

365 Cfr. David A. Forsyth, Jean Ponce, *Computer Vision: A Modern Approach*, Op. Cit., p. 537

366 Cfr. Stephen E. Palmer, *Vision Science - Photons to Phenomenology*, Op. cit. p. 409

367 Ibidem

368 Tanto da essere stato individuato come un anello evolutivo, si parla infatti di *Homo faber*

riguardante anche la struttura culturale e interpersonale della società³⁶⁹. Trattasi comunque di un argomento che va oltre gli obiettivi iniziali di questo elaborato, e che quindi non si espone per lasciare invece spazio al modo in cui è possibile percepire le funzioni. Esistono essenzialmente due tipi di approccio, già accennati: una è la *percezione diretta* postulata prima dalla *Gestalt*, ed estesa alle funzioni da James J. Gibson (che introduce il concetto di *affordances*)³⁷⁰, l'altro è quello della categorizzazione.

Probabilmente l'uomo utilizza entrambi questi metodi per percepire le funzionalità degli oggetti, tanto che alcuni ricercatori presumono vi siano, a tal fine, parti distinte del cervello³⁷¹; nella computer vision prevale, al momento, il secondo.

Per la percezione diretta è sufficiente ricordare che essa si basa sull'assunto che le funzioni di un oggetto siano ricavabili dalla sua forma o struttura, senza il bisogno di averne una precedente esperienza o memoria³⁷².

La categorizzazione è un processo più complesso, in quanto richiede una prima percezione delle proprietà intrinseche di un oggetto al fine di determinarne l'appartenenza a una certa classe e, successivamente, il richiamo dalla memoria delle funzioni ad essa collegate. Da notare che non ci sono limiti teorici alla memorizzazione di funzioni legate ad ogni oggetto; tuttavia, probabilmente ciò non è fisicamente conveniente, ed è forse per questo che gli esseri umani tendono a dimenticare ciò che non usano³⁷³.

All'interno della macro categoria della classificazione è possibile avere diversi approcci, fra questi la *Pattern recognition* e le reti neurali³⁷⁴.

369 Cfr. Stephen E. Palmer, *Vision Science - Photons to Phenomenology*, Op. cit. p. 409

370 Il termine è difficilmente traducibile in quanto coniato da Gibson stesso, sta ad indicare ciò che un oggetto permette, quindi, con una forzatura si potrebbe rendere con "permissioni" o "possibilità/funzionalità"

371 Cfr. Stephen E. Palmer, *Vision Science - Photons to Phenomenology*, Op. cit. p. 413 e 412

372 Ibidem p. 410

373 Ibidem p. 413

374 Cfr. Ramesh Jain, Rangachar Kasturi, Brian G. Schunck, *Machine Vision*, Op. Cit., p. 474

Il termine pattern recognition può creare confusione in quanto si riferisce a una materia quasi autonoma, che si occupa di immagini 2D; esso comprende varie tecniche d'analisi, che possono essere utilizzate non solo per il riconoscimento e la classificazione degli oggetti, ma anche per la successiva verifica di corrispondenza³⁷⁵. La verifica differisce dal riconoscimento e classificazione in quanto si occupa del confronto di un singolo oggetto rilevato con uno o al massimo due modelli dati da immagini o classificazioni³⁷⁶.

I problemi più rilevanti per la pattern recognition sono, da un lato, la costruzione delle classi di modelli e, dall'altro, i metodi per collegarvi i potenziali oggetti (tratti dall'insieme di caratteristiche rilevate).

Il primo problema è stato finora risolto con la creazione di database che abbracciano conoscenze specifiche. L'esempio classico è un programma di riconoscimento testi, la cui conoscenza/database/memoria riguarda le caratteristiche dei caratteri dell'alfabeto³⁷⁷. Il problema di espandere le classi di modelli a più materie è più complesso di quanto possa sembrare, tanto che alcuni ritengono possa essere superato solo riproducendo la struttura fisica del cervello umano, ossia con le reti neurali, permettendo alle macchine di apprendere autonomamente. Nella *machine learning*, le operazioni avvengono in modo *unsupervised*, ossia senza l'intervento umano, e il sistema determina autonomamente sia il numero sia la struttura delle classi³⁷⁸.

Il collegamento tra oggetti e modelli presenta delle difficoltà legate alla somiglianza e/o complessità degli elementi, nonché alla possibilità di sovrapposizioni. Per questo sono stati implementati degli algoritmi complessi che per brevità si citano solamente: classificatori che utilizzano i vicini più prossimi, classificatori di Bayes, classificatori che utilizzano la classe media più vicina, tecniche strutturali, classificatori che utilizzano una struttura decisionale ad albero o dati multidimensionali.

375 Cfr Linda G. Shapiro, George C. Stockman, *Computer Vision*, Op. Cit., pp. 92-93

376 Cfr. Ramesh Jain, Rangachar Kasturi, Brian G. Schunck, *Machine Vision*, Op. Cit., p. 481

377 Cfr. Roberto Marangoni, Marco Geddo, *Le Immagini digitali*, Op. cit., p. 76 e Cfr Linda G. Shapiro, George C. Stockman, *Computer Vision*, Op. Cit., p. 98-100

378 Cfr Linda G. Shapiro, George C. Stockman, *Computer Vision*, Op. Cit., pp. 119-126

Come anticipato, anche le reti neurali possono essere usate per implementare una classificazione di modelli e oggetti. La loro convenienza risiede nel fatto che possono utilizzare anche classi di contorni non lineari³⁷⁹ per ripartire l'insieme di caratteristiche rilevate. I contorni sono ottenuti sottoponendo la rete ad apprendimento/training. Durante questa fase, vengono mostrati al sistema molti esempi di oggetti da riconoscere. Se gli esempi dell'insieme di prova sono selezionati accuratamente, al fine di prevedere tutti i possibili oggetti che si troveranno nella fase di riconoscimento, la rete potrà apprendere la classificazione dei contorni nelle sue caratteristiche spaziali. Nella fase di riconoscimento, essa si comporterà come gli altri algoritmi di classificazione. I vantaggi, rispetto a questi, sono la capacità di considerare classi di bordi non lineari, la possibilità di apprendimento e l'auto-apprendimento della rete. Tuttavia, vi sono dei limiti che riguardano l'impossibilità d'inserire/trasferire la conoscenza già acquisita relativa ad un determinato dominio, operazione semplice nei normali computer, e le difficoltà di controllo/correzione (*debugging*) delle performance del sistema³⁸⁰.

379 In matematica un sistema è non lineare (si usa anche *nonlineare*) quando il suo comportamento non è riconducibile alla semplice somma delle parti che lo descrivono o di loro multipli. Questo significa che in un sistema lineare è possibile fare delle assunzioni e delle approssimazioni, mentre, per un sistema non lineare, questo non è possibile. <http://en.wikipedia.org/wiki/Nonlinear> 11 febbraio 2007

380 Cfr. Ramesh Jain, Rangachar Kasturi, Brian G. Schunck, *Machine Vision*, Op. Cit., p. 479

Relazioni con l'intelligenza artificiale

L'interrogativo di fondo posto all'inizio è se la computer vision, o meglio la riproduzione della percezione in generale, debba essere un passaggio obbligato per replicare/costruire la vita. Dovrebbe essere oramai chiaro che per ora non esiste una risposta che possa essere certa, soddisfacente e condivisa. Inoltre, il quesito stesso è per molti versi troppo sintetico e necessita di essere interpretato. È ora necessario precisare il significato di vita artificiale.

La vita artificiale (detta anche *Alife o alife*) è sia una materia di ricerca sia una forma d'arte che esamina i sistemi collegati alla vita, ai suoi processi ed evoluzioni. Questo avviene attraverso simulazioni che utilizzano modelli computazionali, robotici e biochimici (che vengono, rispettivamente denominati, approcci “*soft*”, “*hard*”, e “*wet*”). Dato che la simulazione al computer è per il momento prevalente, parlando di vita artificiale ci si riferisce spesso a questa³⁸¹. Il termine è stato coniato da Christopher Langton verso la fine degli anni '80, quando ha tenuto la prima "Conferenza Internazionale sulla Sintesi e Simulazione dei Sistemi Viventi" (anche nota come Artificial Life I) presso il Laboratorio Nazionale di Los Alamos, nel 1987³⁸².

La vita artificiale, quindi, non si pone semplicemente lo scopo di ricreare la vita biologica (esperimento riuscito a Stanley Lloyd Miller nel 1952³⁸³), ma lo studio, la riproduzione, la sintesi di interi sistemi viventi³⁸⁴.

Riuscire a riprodurre un sistema “vivente”, sebbene possa fornire interessanti

381 http://en.wikipedia.org/wiki/Artificial_life#_ref-1 11 febbraio 2007

382 http://it.wikipedia.org/wiki/Vita_artificiale 11 febbraio 2007

383 Cfr. Isaac Asimov, *Civiltà extraterrestri*, trad. it. Paola Cusumano, Massimo Parizzi, Milano, Arnoldo Mondadori, 1986 p. 166, (ed. originale *Extraterrestrial Civilizations*, 1979), oppure dello stesso autore *Il libro di Biologia*, Op. cit. p. 154 e in rete http://www.cosediscienza.it/bio/07_vita.htm 11 febbraio 2007

384 È questa l'approccio forte alla vita artificiale, sostenuto tra gli altri da Tom Ray, che ritiene la vita sia un'astrazione indipendente dal substrato materiale. Tom Ray dimostra questi suoi convincimenti con un programma di simulazione denominato Tierra.
http://en.wikipedia.org/wiki/Artificial_life#_ref-1 11 febbraio 2007
<http://www.his.atr.jp/~ray/> 11 febbraio 2007
<http://www.his.atr.jp/~ray/tierra/index.html> 11 febbraio 2007

informazioni sull'evoluzione umana, non dimostra però come si sia potuta sviluppare l'intelligenza. Questa è uno dei fattori, ma non l'unico, che ha permesso all'uomo di "dominare" l'ambiente-mondo che lo circonda³⁸⁵. La presenza di un essere intelligente influenza, notevolmente, la configurazione complessiva di un sistema/ambiente vivente.

È bene precisare che l'intelligenza non porta solo vantaggi. Essendo legata al volume e alla struttura del cervello, e questo a quella del corpo, se ne trae che un organismo dev'essere relativamente grande. Questo comporta che la sua presenza nell'ambiente non possa superare un certo numero. Inoltre, per avere un effettivo vantaggio sulle altre specie, tale specie deve vivere abbastanza a lungo (altrimenti non riuscirebbe ad apprendere abbastanza), e riprodursi quindi con una certa lentezza³⁸⁶. Se ne può dedurre che la realizzazione di sistemi per l'intelligenza artificiale, capaci di memorizzare ingenti quantità di dati per tempi incalcolabili, può apportare dei vantaggi non indifferenti.

Definire esattamente cosa è l'intelligenza non è semplice: la si potrebbe descrivere come l'insieme dato dalle funzioni conoscitive, adattative e immaginative, in possesso dell'uomo e di alcuni animali, grazie ai loro cervelli. L'intelligenza è quindi riconducibile alle capacità di ragionare, apprendere, risolvere problemi, comprendere le idee e il linguaggio³⁸⁷. Per quanto riguarda le macchine, sapere se saranno in grado di compiere operazioni simili (in breve, di pensare) è un quesito cui molti hanno cercato, invano, di dare una risposta. Allan Turing ha bypassato il problema proponendo un test: si tratta di un gioco che consiste, in versione semplificata, nel tentativo di una macchina, di convincere un intervistatore umano che essa è umana. Il test è importante, non per il risultato, ma in quanto propone la possibilità di etichettare come intelligente una macchina in base alle capacità ad essa richieste. Può essere una proposta opinabile, tuttavia essa fornisce un riferimento per valutare il grado d'intelligenza raggiunto dalle macchine³⁸⁸.

385 Cfr. Isaac Asimov, *Civiltà extraterrestri*, Op. Cit., p. 1191-194

386 Ibidem p. 192

387 Cfr Nils J. Nilsson, *Intelligenza artificiale*, Op. Cit., p. 21 si veda anche [http://it.wikipedia.org/wiki/Intelligenza_\(psicologia\)](http://it.wikipedia.org/wiki/Intelligenza_(psicologia)) 11 febbraio 2007

388 Ibidem p. 24-25

A questo punto è possibile riformulare la domanda iniziale: la computer vision è un passo obbligato per l'intelligenza artificiale? In questo caso, la risposta è certamente affermativa. Sono molti a sostenere questa ipotesi, a volte considerandola scontata. Percepire significa infatti apprendere con la mente, ed è quindi indispensabile per conoscere. È per questo che lo studio della percezione è divenuto un importante ambito di ricerca, pur differendo da quello classico della risoluzione di problemi (*problem solving*), del “ragionamento”. Quest'ultimo tipo di ricerca richiede sia fornita, per procedere alla formulazione di una risposta, un'ampia conoscenza della materia: il problema deve essere ben posto³⁸⁹, in quanto la risposta dev'essere unica e ben definita. La percezione, invece, è caratterizzata da problemi mal posti dove le informazioni per decidere sono insufficienti, e si può giungere a soluzioni diverse e non ben definite (es.: illusioni)³⁹⁰.

Se la computer vision è importante per la ricerca sull'intelligenza artificiale, è vero anche l'inverso, ossia vi è un rapporto di interdipendenza. La percezione permette la conoscenza e questa, a sua volta, consente la percezione. Infatti, l'utilizzo di una conoscenza *a priori* permette di restringere la classe di soluzioni possibili di un problema mal posto (cercando di ricondurlo ad un problema ben posto)³⁹¹. In modo più formale, è possibile dire che l'approccio non è solo di tipo bottom-up, come avviene per la vita artificiale, ma anche di tipo top-down, in quanto devono essere utilizzate delle conoscenze già acquisite (assunzioni nascoste) e vi devono essere dei meccanismi di continuo feed-back³⁹². È grazie all'utilizzo dell'intelligenza artificiale che diviene possibile, automaticamente, trasformare una rappresentazione numerica dell'immagine in una rappresentazione simbolica, attraverso i passaggi intermedi d'identificazione e descrizione degli oggetti³⁹³. Questo è particolarmente importante se si vogliono implementare sistemi di visione attiva.

389 La distinzione tra problemi ben e mal posti è quella definita da Hadamard, Cfr. Tomaso Poggio, *Visione: l'altra faccia dell'Intelligenza Artificiale*, in *Mente umana, mente artificiale*, a cura di Riccardo Valle, Op. Cit. p. 283

390 Cfr. Tomaso Poggio, *Visione: l'altra faccia dell'Intelligenza Artificiale*, in *Mente umana, mente artificiale*, a cura di Riccardo Valle, Op. Cit. pp. 278-279

391 Ibidem pp. 283-184

392 Ibidem p. 282

393 Ibidem p. 292

È necessario precisare che le assunzioni utilizzate per implementare l'algoritmo possono risultare, in casi specifici, false e provocare quindi delle *illusioni ottiche*³⁹⁴.

Prima di passare alle possibili applicazioni, è importante sottolineare che ai livelli più alti dell'elaborazione delle immagini, dove le operazioni utilizzano simboli, vi è un avvicinamento sostanziale all'altro ramo dell'intelligenza artificiale, il *problem solving*: in entrambi i casi la ricerca è focalizzata sulla riduzione delle possibili soluzioni di problemi utilizzando la conoscenza delle probabili cause di un determinato evento/fenomeno. Questo non significa, comunque, che i metodi di risoluzione e il tipo di conoscenza siano gli stessi, né che una soluzione matematica possa sempre esistere³⁹⁵.

394 Ibidem p. 290

395 Ibidem pp. 293-294

Possibili applicazioni

Questa tecnologia è applicabile a un così ampio numero di settori che approfondirne solo alcuni equivarrebbe a fornirne un'idea distorta. Per questo motivo si riporta un elenco, reperito in rete³⁹⁶, dove le imprese che hanno già sviluppato e commercializzato alcuni prodotti sono raggruppate per categorie. Nelle tabelle che seguono compare il nome dell'azienda e la relativa localizzazione, quindi il sito web di riferimento e in fine una descrizione del prodotto.

MobilEye (Jerusalem, Israel)

www.mobileye.com

Realizza sistemi di visione per avvertire gli automobilisti di pericoli, permettendo di controllare la velocità, fornendo, quindi assistenza alla guida.

Smart Eye (Göteborg, Sweden).

www.smarteye.se

Sistemi per monitorare gli occhi del guidatore, rilevando sonnolenza o disattenzione. Assistenza/Gestione del traffico

Appian Technology (Bourne End, Buckinghamshire, UK).

www.appian-tech.com

Sistemi per leggere le targhe dei veicoli.

AutoVu (Montreal, Canada).

www.autovu.com/website/indexEng.html

Sistemi per leggere le targhe dei veicoli.

Image Sensing Systems (St. Paul, Minnesota).

www.imagesensing.com

Hanno creato il sistema Autoscope che, utilizzando delle telecamere ai lati della carreggiata, permette il controllo/la gestione in tempo reale del traffico.

Cinema e televisione

2D3 (Oxford, UK).

www.2d3.com

Sistema per il tracciamento di oggetti in film o video e il rilevamento dei movimenti al fine di permetterne l'elaborazione con la computer graphics.

Hawkeye (Winchester, UK).

www.hawkeyeinnovations.co.uk

Utilizza più telecamere per tracciare i movimenti delle palle da tennis o da cricket, al

396 <http://www.cs.ubc.ca/spider/lowe/vision.html> 13 febbraio 2007

fine di permettere l'arbitraggio o gli eventuali commenti.

Image Metrics (Manchester, England).

www.image-metrics.com

Un sistema di tracciamento per i volti umani, che può essere utilizzato per la mappatura dei movimenti e delle espressioni facciali, allo scopo di riprodurli artificialmente.

Imagineer Systems (Guildford, UK).

www.imagineersoftware.com

Realizza software per l'industria cinematografica basati sulla computer vision.

Mova (San Francisco, California).

www.mova.com

Fornisce misurazioni tridimensionali e il tracciamento di migliaia di punti sui volti o altre superfici per ottenerne l'animazione di personaggi.

Orad (Kfar Saba, Israel).

www.orad.co.il

Sistemi per creare set virtuali, l'analisi di sport, e altre applicazioni di *augmented-reality* in tempo reale.

PVI (Lawrenceville, New Jersey).

www.pvi.tv

Utilizza le tecniche di computer vision per tracciare i movimenti delle telecamere (panoramiche, inclinazioni, zoom) che riprendono scene reali, e inserirvi, in tempo reale, eventuali spot pubblicitari.

QuesTec (Deer Park, New York).

www.questec.com

Sistemi per riprendere azioni sportive e trasmetterle ad una qualità migliore.

REALVIZ (Sophia Antipolis, France).

www.realviz.com

Sistemi software per la cattura del movimento, il tracciamento delle telecamere, la ricostruzione di panorami, e la costruzione di modelli 3D.

Sport Universal (Nice, France).

www.sport-universal.com

Sistema per il tracciamento dei giocatori e della palla durante le competizioni sportive in tempo reale. Prevede l'assistenza umana.

Sportvision (New York, NY).

www.sportvision.com

Sistemi di visione per migliorare le immagini trasmesse di avvenimenti sportivi.

Sistemi visivi a scopi generici

Cognex (Natick, Massachusetts)

www.cognex.com

È una delle più grandi società che si occupa di computer vision. Sviluppano sistemi con compiti d'ispezione e localizzazione, il conteggio delle persone, ecc

Evolution Robotics (Pasadena, California)

www.evolution.com

Sistemi di visione per il riconoscimento di oggetti e navigazione. Le applicazioni includono: robot mobili, rivendite di spezie, e il riconoscimento utilizzando le telecamere dei cellulari.

Neptec (Ottawa, Canada).

www.neptec.com

Sistemi visivi 3D al laser, per essere utilizzati sugli shuttle spaziali, ma utili anche per altri scopi.

Newton Research Labs (Renton, Washington).

www.newtonlabs.com

Sistemi visivi per il tracciamento ad alta velocità e robot mobili.

Point Grey Research (Vancouver, Canada).

www.ptgrey.com

Sistemi di stereo-visione in tempo reale, di visione sferica; hardware per l'acquisizione delle immagini.

Sarnoff (Princeton, New Jersey).

www.sarnoff.com

Sistemi visivi per il tracciamento, la registrazione, la navigazione, la biometrica, ecc

Seeing Machines (Canberra, Australia).

www.seeingmachines.com

Sistemi di tracciamento dei movimenti della testa e del puntamento dello sguardo (direzione fissa).

SpikeNet (Toulouse, France).

www.spikenet-technology.com

Sistemi di visione capaci di apprendere per essere in grado di riconoscere.

TYZX (Menlo Park, California).

www.tyzx.com

Sistemi di stereo-visione in tempo reale, che utilizzano dei chip su misura per unire velocemente le immagini stereo.

Ricerca delle immagini

LTU Technologies (Paris, France).

www.ltutech.com

Recupero delle immagini basate sul contenuto.

Ojos Inc. (Redwood City, California).

www.ojos-inc.com

Ha sviluppato il sistema Riya, per la ricerca e l'etichettatura delle immagini in un database utilizzando il riconoscimento di testi e volti.

Polar Rose (Malmo, Sweden).

www.polarrose.com

Recupero delle immagini basato sul riconoscimento di volti.

Automazione e ispezione industriale: l'industria dell'automobile

BrainTech (Vancouver, Canada).

www.bnti.com

Sistemi per la visione e la guida di robot nell'industria dell'auto.

CogniTens (Ramat-Hasharon, Israel).

www.cognitens.com

Ha sviluppato un sistema accurato di rilevamento di oggetti 3D principalmente per l'industria automobilistica, ma utilizzabile anche in altri settori. Il sistema utilizza 4 telecamere e un proiettore che emana un fascio di luce al fine di evidenziare e riconoscere le textures presenti nella scena.

Perceptron (Plymouth, Michigan).

www.perceptron.com

Produce sistemi tridimensionali di esplorazione al laser.

Automazione e ispezione industriale: l'industria elettronica

ICOS Vision Systems (Heverlee, Belgium).

www.icos.be

Sistemi per l'ispezione elettronica e per l'assemblaggio di componenti e la fabbricazione di semiconduttori.

KLA-Tencor (San Jose, California).

www.kla-tencor.com

Sistemi per l'ispezione e il controllo di processo nella fabbricazione di semiconduttori.

Orbotech (Yavne, Israel).

www.orbotech.com

Sistemi per l'ispezione automatica di schede a circuiti stampati e di video piatti.

RVSI Inspection (Hauppauge, New York).

www.rvsi.com/rvsi/index.html

Sistemi visivi per l'ispezione elettronica e l'assemblaggio.

Automazione e ispezione industriale: l'industria alimentare e l'agricoltura

Dipix Technologies (Ottawa, Canada).

www.dipix.com

Sistemi visivi per l'industria dei cibi cotti. Il sistema controlla la cottura: i colori, la

forma, le dimensioni di pane, dolci, tortillas, ecc

Ellips (Eindhoven, The Netherlands).

www.ellips.nl

Sistemi visivi per l'ispezione e la classificazione di frutta e vegetali.

Automazione e ispezione industriale: la stampa e il tessile

Advanced Vision Technology (Hod Hasharon, Israel).

avt-inc.com

Sistemi per il controllo delle stampe ad alta velocità.

Elbit Vision Systems Ltd. (Yoqneam, Israel).

www.evs.co.il

Sistemi di visione per il controllo dei tessuti e altro.

Mneumonics (Mt. Laurel, New Jersey).

mnemonicsinc.com

Sistemi visivi per il controllo della stampa.

Xiris Automation (Burlington, Ontario, Canada).

www.xiris.com

Sistemi d'ispezione per la stampa e l'imballaggio.

Automazione e ispezione industriale: altri casi

Adept (Livermore, California).

www.adept.com

Robot industriali dotati di visione per il posizionamento di componenti e loro ispezione.

Avalon Vision Solutions (Lithia Springs, Georgia).

www.avalonvisionsolutions.com

Sistemi visivi per l'industria della plastica.

Basler (Ahrensburg, Germany).

www.baslerweb.com

Sistemi di controllo per strumenti ottici, sigillanti, schermi, ecc

Hermery Opto Electronics (Coquitlam, BC, Canada).

www.hermeryopto.com

Sviluppa scanner 3D per segherie e altre applicazioni.

JLI vision (Soborg, Denmark).

www.jli.dk

Sistemi visivi per il controllo industriale nei settori alimentare, della lavorazione del vetro, degli strumenti medici e della lavorazione del ferro.

LMI Technologies (Vancouver, Canada).

www.lmint.com

Sviluppano sistemi di visione 3D al laser per il controllo della produzione di legno, strade, veicoli, ecc

MVTec (Munich, Germany).

www.mvtec.com

Sistemi di controllo e altre applicazioni.

NeuroCheck GmbH (Remseck, Germany).

www.neurocheck.com

Sistemi di controllo per la qualità.

PPT Vision (Eden Prairie, Minnesota).

www.pptvision.com

Sistemi visivi per l'industria farmaceutica, dell'auto, dell'elettronica, ecc

SICK IVP (Linköping, Sweden).

www.ivp.se

Piccole telecamere che utilizzano processori dedicati per applicazioni industriali ad alta velocità.

SIGHTech (San Jose, California).

www.sightech.com

Sistemi visivi in grado di apprendere per il controllo e l'automazione.

Virtek Vision International (Waterloo, Ontario, Canada).

www.virtek.ca

Sistemi di modellazione e controllo basati sul laser.

Wintriss Engineering (San Diego, California).

www.weco.com

Sistemi visivi per il controllo di applicazioni web.

Medicina e biomedicina

Claron Technology (Toronto, Canada).

www.clarontech.com

Utilizza sistemi di stereo-visione in tempo reale per rilevare e tracciare la posizione di marcatori per applicazioni chirurgiche.

CTI Mirada Solutions (Siemens) (Oxford, UK).

www.ctimi.com/portals/ctimi/content/about_mirada.html

Sistemi per l'analisi quantitativa delle immagini mediche, compresa la diagnosi del cancro al seno.

Cynovad (Montreal, Canada).

www.cynovad.com

Sistema per far combaciare il colore protesico dei denti con il colore naturale dei denti

del paziente.

Noesis (St. Laurent, Quebec, Canada).

www.noesisvision.com/index_en.htm

Software per uso biomedico e l'analisi scientifica delle immagini.

TriPath Imaging (Burlington, North Carolina).

www.tripathimaging.com

Sistemi visivi per la rilevazione di macchie sul capezzolo che potrebbero indicare la presenza di cellule anormali.

Sistemi per tracciare i pedoni

Reveal (Auckland, New Zealand).

www.reveal.co.nz

Sistema utilizzato per conteggiare e tracciare i pedoni. Utilizza una telecamera montata sul capo.

Monitoraggio sanitario

Vision IQ (Boulogne-Billancourt, France).

www.vision-iq.com

Il sistema Poseidon monitora le piscine per avvisare di eventuali incidenti e possibili annegamenti.

Sicurezza e biometrica

A4Vision (Sunnyvale, California).

www.a4vision.com

Sistema per l'identificazione di volti che utilizza la luce per la ricostruzione 3D. La compagnia ha creato anche il software per il tracciamento dei volti utilizzato nelle web-cam Logitech.

Activeye (Briarcliff Manor, New York).

www.activeye.com

Sistemi visivi per la sorveglianza, che includono il tracciamento, il monitoraggio degli oggetti e l'analisi dei comportamenti.

Aimetis (Waterloo, Ontario, Canada).

www.aimetis.com

Sistema di sorveglianza intelligente.

Aurora (Northampton, UK).

www.facerec.com

Sistema biometrico per il riconoscimento dei volti.

AuthenTec (Melbourne, Florida).

www.authentec.com

Sistema per il riconoscimento delle impronte digitali basato su nuovo sensore.

Digital Persona (Redwood City, California).

www.digitalpersona.com

Sistemi di riconoscimento delle impronte digitali.

EVITECH (Paris, France).

www.evitech.com

Sistemi di video sorveglianza di dimensioni ridotte.

Equinox (New York, NY).

www.equinoxsensors.com

Sistema di sicurezza che utilizza nuovi sensori, come quelli si basano sulla registrazione degli infrarossi o che sfruttano la luce polarizzata.

Geometrix (San Jose, California).

www.geometrix.com

Riconoscimento di volti che utilizza dati tridimensionali provenienti da immagini stereo.

L-1 Identity Solutions (Stamford, Connecticut).

www.l1id.com

Sistemi di riconoscimento delle impronte, dell'iride e dei volti.

ObjectVideo (Reston, Virginia).

www.objectvideo.com

Prodotti per la video-sorveglianza che consentono il tracciamento, il riconoscimento e la classificazione delle attività.

Vidient (Sunnyvale, California).

www.vidient.com

Sistemi di video sorveglianza con riconoscimento del comportamento.

Modellazione tridimensionale

Creative Dimension Software (Guildford, UK).

www.3dsom.com

Creano modelli 3D da un insieme di immagini.

Eos Systems (Vancouver, Canada).

www.photomodeler.com

PhotoModeler software permette la creazione di modelli tridimensionali che mappano le textures da un ridotto numero di foto. È richiesto un inserimento manuale.

Eyetrionics (Leuven, Belgium).

www.eyetrionics.com

Produce scanner tridimensionali per il corpo umano utilizzando luce strutturata.

InSpeck (Quebec City, Canada).

www.inspeck.com

Usa la proiezione della luce per creare un modello tridimensionale di textures del volto o del corpo umano in frazioni di secondi.

Videogiochi

GestureTek (Toronto, Canada).

www.gesturetek.com

Traccia i gesti per giocare, o interagire con il computer.

Reactrix (Redwood City, California).

www.reactrix.com

Pubblicità interattiva per proiettori che tracciano i gesti umani.

Sony EyeToy

www.eyetoy.com

Utilizza la computer vision per tracciare i movimenti delle mani e del corpo dei giocatori per controllare la Playstation. Le vendite nel 2004 hanno superato i 4 milioni di unità.

Quali saranno le conseguenze?

Da quanto finora esposto, spero si sia acquisita un'idea del funzionamento della visione e della possibilità di riprodurla, e delle attuali implementazioni della computer vision. È normale, credo, chiedersi come questa tecnologia cambierà il mondo, quali saranno gli effetti, i probabili sviluppi. Dato l'argomento, sembra fantascienza; tuttavia, quello che si propone non è una profezia, che per definizione si basa sulle capacità di chiaroveggenza di chi le fa, ma una previsione, che ha come base dei processi logici³⁹⁷. E se proprio la si vuole considerare fantascienza, è utile ricordare l'introduzione alla raccolta "Sogni di Robot", di Asimov: "La fantascienza può dare soddisfazioni del tutto particolari. Per esempio si può andare molto vicini al bersaglio quando si tenta di costruire un quadro attendibile della tecnologia futura. Se si scrive un racconto, e si vive abbastanza a lungo, si può avere la soddisfazione di scoprire che le proprie predizioni erano ragionevolmente accurate e può anche accadere di vedersi considerare come una specie di profeta moderno. A me è successo con le storie sui robot³⁹⁸ e in particolare con il racconto *Luciscultura*, incluso in questa antologia."³⁹⁹.

È anche per queste dichiarazioni che Asimov è considerato un futurologo⁴⁰⁰. La futurologia, pur non essendo considerata scienza⁴⁰¹, riveste un ruolo importante in quanto costituisce un insieme di ricerche volte a prevedere, basandosi anche su analisi e proiezioni scientifiche, le caratteristiche e le possibili tendenze del mondo futuro, con

397 <http://it.wikipedia.org/wiki/Profezia> 14 febbraio 2007

398 Il termine robot deriva dal termine ceco robota, che significa "lavoro pesante" o "lavoro forzato". L'introduzione del termine è dovuta allo scrittore ceco Karel Čapek, che lo usò per la prima volta nel 1920 nel dramma teatrale "I robot universali di Rossum" in breve R.U.R.; il termine "robotica", invece, è stato usato per la prima volta (su carta stampata) da Asimov nel racconto intitolato Circolo vizioso (Runaround, 1942), inserito nella raccolta Io, Robot. Cfr. <http://it.wikipedia.org/wiki/Robot> 15 febbraio 2007

399 Isaac Asimov, *Sogni di Robot*, trad. it. di Mauro Gaffo, Milano, Interno Giallo, 1990 p. 7 (ed. originale, *Robot dreams*, 1986)

400 <http://it.wikipedia.org/wiki/Futurologo> e <http://www.fabiofemiofantascienza.org/FUTURES/FUTUROLOGY2.html> 14 febbraio 2007

401 Si veda l'intervento di Stanisław Lem, scrittore di *science fiction* e filosofo, su quali sono gli ostacoli e le opportunità offerte da questo tipo di ricerche.

<http://www.intercom.publinet.it/ic11/metafut.htm> 14 febbraio 2007

particolare riguardo all'esistenza umana⁴⁰². A tal proposito è interessante considerare un intervento di Francesco d'Arcais, direttore del periodico “Civiltà delle Macchine” dal 1960 al 1979 (anno dell'ultima pubblicazione⁴⁰³): “Il futuro come scienza, dunque? La risposta è sì e no, in uguale misura (da un punto di vista qualitativo, naturalmente), perché esiste una esigenza di prevedere e quindi di programmare, ma ci dev'essere anche la percezione più attenta delle situazioni inattese - non sempre clamorosamente avvertibili - che può modificare tutta una progettazione già in atto.

La futurologia come scienza è nata e si sta sviluppando in un ambito che è soprattutto economico, anche se è richiesto il contributo di uomini appartenenti ad altri settori scientifici e culturali; la cosa si spiega con il fatto che l'economia sta diventando sempre più scienza, nell'accezione comune della parola, sta cioè sempre più matematizzandosi. Al tempo stesso l'economia va assumendo dimensioni e significati un tempo imprevedibili. Siamo dunque in piena fase di trapasso, anche qui (la storia del resto è un fatto unitario, e le ripercussioni sono sempre più immediate e più forti) e si deve perciò stare in guardia da ogni eccesso. Certamente come la scienza che lavora su ipotesi ma è sempre pronta ad abbandonarle quando non sono più valide, così anche la futurologia, che deve partire da ipotesi (ciò che oggi è una prospettiva di estrapolazione) avrà la forza di abbandonarle quando si dimostrassero infeconde.”⁴⁰⁴.

Nelle pagine che seguono, si cercherà di costruire degli scenari futuri, basandosi su alcuni testi di economia, sociologia e psicologia, ma anche sugli scritti di Asimov, evitando, per quanto possibile, di cadere nell'errore di scendere troppo nei dettagli. Tale errore è stato evidenziato, in modo sublime, dall'umorismo di Italo Calvino nella raccolta “Tutte le cosmicomiche”. Più precisamente, nel racconto “Quanto scommettiamo”, il protagonista, *Qfwfq*, si spinge a prevedere e scommettere su una serie interminabile di avvenimenti, compresi quelli più marginali e aleatori, per giungere a realizzare, a sue spese, che ciò non è possibile né vantaggioso⁴⁰⁵. Scendere

402 <http://www.sapere.it/SearchWeb/results.jsp?q=futurologia> 14 febbraio 2007

403 http://www.infocity.go.it/vedi_comunicato.php?id=271 15 febbraio 2007

404 <http://diea.ing.unibo.it/cdm/articolo.php?id=1442&numero=5&anno=1968> 15 febbraio 2007

405 Cfr. Italo Calvino, *Tutte le cosmicomiche*, a cura di Claudio Milanini, Milano, Arnoldo Mondadori, 2002 pp. 82-91

troppo nei dettagli è controproducente in quanto non si tiene conto, come afferma Stanisław Lem, dell'imponderabile, trasmettendo la sensazione che tutto sia determinabile, con le opportune conoscenze, *in primis* tecnologiche. Lem afferma: "Quando guidiamo la civiltà controllando la tecnologia, siamo padroni degli effetti immediati delle nostre azioni. Allo stesso tempo, ignoriamo ogni mezzo di guida non tecnologico - e quindi non strumentale. Dapprima lo ignoriamo perché è più semplice farlo; più in là, come conseguenza di questo processo di selezione, la struttura delle connessioni all'interno della civiltà si dimostra sempre più dipendente da interventi puramente strumentali."⁴⁰⁶ In generale, sussiste il rischio che previsioni dettagliate, tali da poter essere ritenute oggettive, incidano di per sé sul futuro, realizzando quello che viene chiamato *self-fulfilling prophecies*⁴⁰⁷.

In questo elaborato l'intento è esattamente l'opposto. Sulle orme di Umberto Eco, che ha espresso il concetto di opera aperta⁴⁰⁸, si vuole fornire solo alcune indicazioni sulle possibili evoluzioni, lasciando che ognuno possa inferirne e aggiungerne altre. Il motivo è semplice: il futuro non è solo conseguenza del presente, ma dipende dal contributo di ognuno.

Prima di cominciare è necessario chiarire che la suddivisione tra le materie è operata solo ai fini di semplificazione: la realtà/il mondo è unica/o, anche se può essere interpretata/o da angolazioni diverse. L'ordine in cui tali materie appaiono non riflette filosofie di tipo materialista, che si ricollegano al pensiero di Marx, o immaterialista, da ricollegarsi a Hegel. Secondo il primo, sono le condizioni in cui si vive, specie lavorative, che determinano l'autorealizzazione. Per Hegel, invece, la felicità dipende dalle idee, dalle interpretazioni, dalla coscienza⁴⁰⁹. E' plausibile pensare che i vari approcci siano tra loro interdipendenti: le condizioni economiche influenzano la formazione degli individui che costituiscono la società; il pensiero di singoli, chiamati a decidere per molti, può determinare non solo delle prese di posizione collettive, ma

406 <http://www.intercom.publignet.it/ic11/metafut.htm> 15 febbraio 2007

407 Cfr Gianni Losito, *Il potere del pubblico*, Roma, Carrocci, 2002 p. 95

408 Eco attesta che uno stesso testo possa essere interpretato in modi diversi da diversi lettori, Cfr Gianni Losito, *Il potere del pubblico*, Op. cit., p. 101 e 106

409 Paolo Jedlowski, *Il mondo in questione*, Roma, Carrocci, 1998 pp. 41-42.

186

anche le risorse disponibili in un prossimo futuro.

Conseguenze economiche

L'importanza dell'avanzamento/progresso tecnologico per le imprese è una certezza da tempo immemore, specie se considerata in ottica di lungo periodo. I vantaggi che se ne possono trarre in termini di riduzione di costi fissi, di miglioramento della qualità del prodotto e aumento delle quantità prodotte sono essenziali per qualsiasi azienda, al di là del settore d'appartenenza⁴¹⁰. La tecnologia utilizzata incide quasi direttamente sulle dimensioni, sull'organizzazione, sulle strategie, sulla logistica e sul posizionamento delle imprese, specie per quelle che hanno carattere multinazionale⁴¹¹.

Si può anche affermare che la tecnologia costituisce il contesto in cui opera l'impresa. Si possono quindi distinguere, a seconda della tecnologia usata, diversi contesti come, ad esempio: tecnologie per la produzione di soia, della fusione nucleare, ecc. L'impresa, nella sua essenza, costituisce l'insieme di conoscenze e di beni che vincolano le capacità decisionali dell'imprenditore-produttore. Questa è un'estrema semplificazione di un processo, quello decisionale, che risulta molto complesso e coinvolge diversi soggetti⁴¹².

È bene introdurre qualche altro concetto per facilitare l'analisi. Per esempio, il progresso tecnico è riferito o all'innovazione del prodotto o a quella del processo. Questa ripartizione mal si adatta, in generale, alle innovazioni legate all'elettronica e in particolare alla computer vision. È più significativo distinguere, al limite, tra cambiamenti che partono dall'interno di un'azienda o di uno stesso settore, e quelli che si realizzano all'esterno di questi. Così si possono infatti comprendere come i mutamenti del tecnologia incidono sulle relazioni economiche e produttive⁴¹³. Tali mutamenti o, meglio, il grado d'incidenza/ampiezza del loro raggio d'azione sul sistema economico, costituisce un'ulteriore prospettiva, e la si può suddividere in sotto-categorie⁴¹⁴.

410 Hal R. Varian, *Microeconomia*, trad. it di Giulio Codognato, Stefano Chinellato, Venezia, Cafoscarina, 1998 pp. 291-301 (ed. originale *Intermediate Microeconomics. A Modern Approach*, 1987)

411 Francesco Grassivaro, *Le Imprese Multinazionali*, Padova, Cedam, 1991, p. 136

412 Arrigo Opocher, *Lezioni di economia politica*, Padova, Cedam, 1995 pp. 129-130

413 Francesco Grassivaro, *Economia Politica – Teorie e modelli*, Padova, Cleup, 1992 p. 113

414 Ibidem pp. 114-115

1. Innovazioni incrementali.

Sono i miglioramenti di prodotto che avvengono nel breve periodo, e che non ne alterano sensibilmente le caratteristiche. Sono effettuati con lo scopo di migliorarne l'attrattiva nei confronti dei potenziali clienti e, di conseguenza aumentare la profittabilità dell'impresa.

2. Innovazioni che portano alla formazione di un nuovo prodotto.

3. Innovazioni che conducono alla creazione di gruppi di nuovi prodotti.

L'esempio dell'introduzione delle materie plastiche nel sistema economico-produttivo può chiarire il concetto. Essa ha permesso sia la nascita di una miriade di nuovi prodotti sia l'evoluzione della "funzione di produzione", ossia ha cambiato le modalità produttive in diversi settori.

4. Innovazioni che mutano lo scenario economico.

Ciò avviene quando una o più innovazioni determinano un mutamento complessivo dei "modi di produrre", portando alla creazione di nuove merci e alla modifica delle preferenze dei consumatori.

Gli esempi relativi a quest'ultima categoria partono dall'introduzione della macchina a vapore alla scoperta dell'energia elettrica, dalla realizzazione del motore a scoppio all'uso dei combustibili liquidi, fino ad arrivare all'introduzione della microelettronica. C'è chi prevede, come possibili tappe a venire, l'utilizzo di biotecnologie o un più ampio impiego dell'energia solare⁴¹⁵.

Fatte queste premesse, e tenendo presente l'elenco delle possibili applicazioni, si può cercare di trarre alcune conclusioni, necessarie per il prossimo scenario.

415 Ibidem

La computer vision è collegabile alla microelettronica: senza di questa non vi sarebbero probabilmente dei computer sufficientemente potenti e compatti da poter implementare la computer vision stessa; probabilmente non vi sarebbe neppure una teoria computazionale. È lecito chiedersi se la computer vision possa essere compresa tra quelle innovazioni che hanno un più forte impatto sul sistema economico. Ora come ora, la sua condizione è simile a quella dell'energia atomica che non ha ancora inciso profondamente sulla struttura economica, sì da poter identificare un nuovo paradigma tecnologico⁴¹⁶. La computer vision, pur avendo in via ipotetica la possibilità di compiere questo mutamento, rientra per ora tra le innovazioni che portano alla creazione di nuovi gruppi di prodotti. Come descritto nei primi capitoli di questo libro, è possibile infatti avere un aiuto alla guida dei veicoli⁴¹⁷, nella gestione del traffico, nella realizzazione di prodotti audio-visivi; essa permette inoltre la realizzazione di robot mobili, di migliorare l'automazione e il controllo industriale; è applicabile in medicina, può essere utilizzata per ricercare o filtrare informazioni, nonché a scopi ludici. Ne consegue la formazione di nuovi mercati settoriali, con nuove imprese che entrano in competizione, il che si pone, a sua volta, come ulteriore fattore di sviluppo tecnologico⁴¹⁸.

Essa è soprattutto un'innovazione di processo: la sua influenza tende quindi ad espandersi ben oltre il proprio settore. In considerazione della sua natura, eminentemente di controllo e ispezione, si può supporre che essa porterà ad un migliore utilizzo delle risorse disponibili, con una riduzione del materiale di scarto e un aumento delle economie di scala. Il tutto si basa su una sua applicazione all'automazione industriale e in particolare alla robotica. Interessante è a tal proposito quanto riportava Piero Angela, nel suo "Quark Economia", qualche anno fa (1986): "Così come il braccio umano (che è in fondo uno strumento semplice) è in grado di fare cose straordinarie e diversissime, grazie agli impulsi che gli arrivano dal cervello (per

416 Ibidem

417 Si veda ad esempio il sistema di parcheggio assistito installato sulla Touran della Volkswagen http://www.omniauto.it/magazine/articolo/1784/la_touran_con_l%E2%80%98assistente_di_parcheggio.html 16 febbraio 2007

418 Si veda l'ipotesi di David Bordwell e Janet Staiger sulle cause dell'affermazione dell'arte cinematografica (per alcuni versi simili al fenomeno qui considerato) in Lev Manovic, *Il linguaggio dei nuovi media*, trad. it. di Roberto Merlini, Milano, Fres, 2004 p. 237 (edizione originale *The Language of New Media*, 2001)

esempio verniciare, riordinare uno scaffale, o scrivere a macchina), analogamente i bracci dei robot potranno compiere lavori sempre più complessi man mano che aumenteranno le capacità dei computer che li guidano (e anche la capacità umana di guidare i computer).

Uno dei prossimi sviluppi è, per esempio la crescita delle capacità visive, che permetterà ai robot di riconoscere gli oggetti, di scegliere, di adattarsi a situazioni diverse.

Con i futuri calcolatori, detti della quinta generazione, queste capacità dovrebbero compiere un nuovo balzo in avanti, permettendo di interpretare le immagini e persino di leggere testi. Sarà il momento in cui si avranno forse macchine per scrivere capaci di riconoscere la voce umana e di scrivere direttamente sotto dettatura senza più bisogno di dattilografi.⁴¹⁹ Oggi sappiamo che queste previsioni si sono avverate: si pensi, ad esempio, agli ultimi ritrovati della telefonia mobile che permettono d'interagire con i cellulari attraverso la voce.

Asimov, con i suoi racconti di robotica, si spinge ben oltre, probabilmente dipingendo scenari che non si avvereranno mai; tuttavia, non gli si può dar torto quando afferma: “Ma i robot sono fattibili? E, nel caso lo fossero, varrebbe la pena costruirli? Con meccanismi a base di molle e ingranaggi si possono costruire senza difficoltà macchine di forma umana, capaci di imitare le azioni umane, ma l'essenza di un vero robot è la sua capacità di pensare: e la capacità di pensare talmente bene da poter svolgere lavori utili, senza bisogno di un continuo controllo.”⁴²⁰ Asimov, quindi, suggerisce l'idea di robot come strumenti avanzati a disposizione dell'uomo, capaci di comprenderne i desideri e soddisfarli. I robot potrebbero compiere per noi “i lavori semplici, noiosi o ripetitivi, fisici o mentali, che possono essere svolti facilmente – e meglio – da macchine non più complesse di quelle di cui già disponiamo ora.”⁴²¹ L'avanzamento

419 Piero Angela, *QUARK Economia – per capire un mondo che cambia*, Milano, Garzanti, 1987 pp. 47-48

420 Isaac Asimov, *Domani! 71 sguardi su un futuro migliore*, trad. it. di Riccardo Valla, Trezzano sul Naviglio, Euroclub, 1990 p. 90 (ed. originale *Change!*, 1982)

421 *Ibidem* p. 87

tecnologico aumenterebbe quindi le possibilità di scelta e, in sintesi, la nostra libertà. Qualcuno (si pensi al movimento luddista, nato con la prima rivoluzione industriale in Inghilterra nel XIX secolo⁴²²), potrebbe essere spaventato dalla prospettiva che le macchine lo sostituiscano in questi compiti. Ciò è probabilmente dovuto all'attuale struttura socio-economica destinata, come già avvenuto in passato, a cambiare in futuro: il punto è se tale cambiamento sarà “progressivo” o conflittuale⁴²³.

Come si può desumere dall'elenco presentato, le applicazioni della computer vision nel campo della mecatronica⁴²⁴ non sono le uniche possibili. Molte riguardano il trattamento dell'informazione visiva, come ricercare, modificare e archiviare immagini. Per ora il processo ha ancora bisogno, in alcune fasi, dell'intervento umano, ma si arriverà ad un trattamento completamente automatico grazie anche a tecniche di intelligenza artificiale. Se questo accadrà, si assisterà, sempre in via ipotetica, a un'evoluzione del processo decisionale il quale diverrà sempre più automatizzato. Interessante è quanto scrive Lev Manovich, riprendendo il pensiero di Paul Virilio: “Il regime della grande ottica porta inevitabilmente alla politica gestita in tempo reale, una politica che comporta reazioni istantanee a eventi pubblicizzati alla velocità della luce e che, alla fine, potrà essere gestita efficientemente solo da computer interconnessi.”⁴²⁵. Computer in grado di percepire l'ambiente, di apprendere/inferire dalle loro percezioni e capaci di trasformare i vincoli del contesto in opportunità? Possono permettere all'uomo (alle società umane) di liberarsi dalle necessità quotidiane e incombenti, per passare a concentrarsi su emergenze planetarie⁴²⁶, o ad esprimere nuovi bisogni?

422 <http://it.wikipedia.org/wiki/Luddismo> 17 febbraio 2007

423 Isaac Asimov, *Domani! 71 sguardi su un futuro migliore*, Op. cit. pp. 88-89

424 Acronimo di meccanica ed elettronica, è la materia che risulta dalla combinazione sinergica degli studi di ingegneria meccanica, elettronica, informatica e del controllo
<http://es.wikipedia.org/wiki/Mecatr%C3%B3nica> 17 febbraio 2007.

425 Lev Manovic, *Il linguaggio dei nuovi media*, Op. Cit., p. 220

426 Cfr. Antonio Zichichi, *Scienza ed emergenze planetarie*, Milano, R.C.S. Libri e Grandi Opere, 2003 pp. 11-12

Conseguenze sociali

La sociologia è legata alla tecnologia fin dalle sue origini. Queste, infatti, sono riconducibili al periodo della prima rivoluzione industriale inglese (metà del 1700) e alla rivoluzione politica francese (fine 1700)⁴²⁷. In verità, il termine è stato utilizzato per la prima volta da Auguste Comte (1798-1857) in tempi di poco successivi⁴²⁸. Tuttavia, dato che la ricerca sociologica si è affermata solo quando ci si è resi conto che una serie di “regole”, consuetudini, ecc, condivise da più individui, erano cambiate, è a questi periodi dalle evidenti trasformazioni che sono ricollegate le sue radici⁴²⁹.

Il campo d'indagine della sociologia è rappresentato dalle relazioni e dalle istituzioni umane. Essa s'interroga se queste siano rimaste immutate nel tempo e nello spazio, quale sia stato il loro principio e quale la loro evoluzione⁴³⁰.

Comte, sulle orme di Henri de Saint-Simon (1760-1825), postulava che il progresso tecnologico avrebbe portato ad una totale riorganizzazione della società⁴³¹. È nello scritto “Sistema di politica positiva” del 1854 che Comte propone il positivismo come un'idea politica; egli credeva fermamente che si stesse realizzando una società in cui gli uomini avrebbero adottato leggi conformi alla propria natura, leggi individuate da scienziati e tecnici⁴³². È interessante il fatto che Comte, negli ultimi anni della sua vita, recuperò una visione religiosa e si propose come un sacerdote del culto dell'umanità⁴³³. Vi sono pareri discordi su questo suo atteggiamento. Tuttavia, il sociologo Paolo Jedlowski sostiene che Comte debba essersi reso conto che la scienza non può, di per sé, portare all'integrazione sociale, che sono necessari dei valori assoluti in cui gli uomini possano credere⁴³⁴.

427 Cfr. Paolo Jedlowski, *Il mondo in questione*, Op. cit., p. 17

428 Ibidem p. 22

429 Ibidem p. 17

430 Ibidem pp. 13-14

431 Ibidem pp. 13-14

432 Ibidem p. 33

433 Ibidem

434 Ibidem

È necessario introdurre la distinzione tra scienza e tecnica. Il loro rapporto è cresciuto a tal punto che è facile confonderle. La ricerca scientifica o fondamentale si occupa, in via primaria, della conoscenza, ossia della descrizione e spiegazione degli eventi, singoli o ricorrenti, sia naturali sia umani e sociali. La tecnica o ricerca applicata si occupa, invece, di problemi pratici, basandosi sia su conoscenze empiriche sia su conoscenze scientifiche⁴³⁵. Si può inoltre pensare che vi sia un passaggio diretto da una scoperta scientifica alla relativa invenzione tecnologica e da questa all'eventuale sfruttamento economico. Tuttavia, spesso non è così: le esigenze che dirigono la ricerca verso determinate scoperte sono di tipo economico (o militare), solo successivamente l'interesse per la materia accresce e assume valore scientifico⁴³⁶. Un esempio eclatante è il progetto Manhattan che portò alla costruzione delle prime bombe atomiche⁴³⁷. L'innovazione tecnologica, per svilupparsi e diffondersi, deve quindi attirare interessi di natura economica, politica o militare. È inevitabile, anche in questo caso, una competizione per le risorse. È questo il motivo per cui, normalmente, la ricerca fondamentale è sostenuta dalle università e da istituzioni pubbliche, mentre quella applicata è appannaggio delle imprese private. Da rilevare che i finanziamenti militari vanno in entrambe le direzioni, portando innovazioni che possono essere utilizzate anche in ambito civile (emblematico il caso di internet) portando al fenomeno detto *technological fall-out* ("ricaduta tecnologica")⁴³⁸. L'innovazione tecnologia non ha però lo stesso peso per tutti i settori economici. Il settore delle telecomunicazioni, a cui si può ricondurre la computer vision, è tra quelli che ne risentono maggiormente, e che più direttamente influenza l'opinione pubblica⁴³⁹.

Per quanto riguarda il rapporto tra società e scienza è interessante l'opinione del filosofo, logico e matematico Bertrand Russell (1872-1970)⁴⁴⁰. Come afferma Marcello Pera, nella sua prefazione al testo "La visione scientifica del mondo" dello stesso

435 Cfr Arnaldo Bagnasco, Marzio Barbagli, Alessandro Cavalli, *Corso di sociologia*, Op. Cit., p. 234 si veda inoltre Antonio Zichichi, *Scienza ed emergenze planetarie*, Op. Cit. p. 19

436 Cfr Arnaldo Bagnasco, Marzio Barbagli, Alessandro Cavalli, *Corso di sociologia*, Op. Cit., p. 250-251

437 Cfr. Antonio Zichichi, *Scienza ed emergenze planetarie*, Op. Cit. pp. 39-41

438 Cfr Arnaldo Bagnasco, Marzio Barbagli, Alessandro Cavalli, *Corso di sociologia*, Op. Cit., pp. 251-252

439 Ibidem

440 http://it.wikipedia.org/wiki/Bertrand_Russell 18 febbraio 2007

Russell, egli fornì “un'immagine di società scientifica” che “anticipò gran parte della letteratura utopistica negativa di questo secolo”. E ancora: “Anticipò l'occhio del Grande Fratello e l'atmosfera opprimente di *1984* di George Orwell; anticipò lo Stato Mondiale, con i centri d'incubazione e le Sale di condizionatura neo-pavloviana, di Fecondazione e di Predestinazione sociale del *Brave New World* di Aldous Huxley. Quando, nel 1932, anno successivo all'uscita del suo libro, comparve questo romanzo, Russell si lamentò di plagio in una lettera di Stanley Unwin notando, non senza qualche ragione, che l'opera di Huxley era <<soltanto una estensione dei due penultimi capitoli di *Scientific Outlook*>>.”⁴⁴¹ Russell suddivise il testo in tre parti: nella prima e seconda parte disaminò la conoscenza e la tecnica scientifica, mentre nella terza parte elaborò e propose degli scenari di società scientifiche. La prima parte ricollega la nascita della scienza a Galileo Galilei e all'introduzione del metodo scientifico, ne esamina gli eventuali limiti e quindi passa a valutazioni metafisiche sulle relazioni tra scienza e religione. Nella seconda parte si analizza l'applicazione della tecnica a vari ambiti: alla natura inanimata, alla biologia, alla fisiologia, alla psicologia e infine all'ambito sociale. La terza parte è probabilmente la più significativa. Qui Russell evidenziò, volutamente, quali avrebbero potuto essere gli effetti disastrosi di innovazioni tecnologiche, di per sé buone, che fossero diventate l'unico fattore di sviluppo sociale⁴⁴². Egli concepì la società scientifica come “quella che impiega la migliore tecnica scientifica nella produzione, nell'educazione e nella propaganda”⁴⁴³, e “Nessuna società può dirsi pienamente scientifica, se non le è stata data deliberatamente una certa struttura per espletare certe mansioni.”⁴⁴⁴. Arrivò anche a pensare a società artificiali come naturale conseguenza del perfezionamento della scienza sociale, che potrà intervenire sull'uomo sia a livello biologico che psicologico⁴⁴⁵, aggiungendo che i primi tentativi di realizzare queste società sono stati oggetto di fallimento in quanto gli imprevisti sono stati più determinanti di quanto pianificato. Portò ad esempio i casi della Germania nazista e del Giappone nel periodo che va dal 1867 alla seconda guerra mondiale, nonché della

441 Bertrand Russell, *La visione scientifica del mondo*, trad. it. di Emilio A. G. Loliva, Roma-Bari, Gius. Laterza & Figli, 2004 p. X (ed. originante *The Scientific Outlook*, prima 1931, seconda 1949)

442 Cfr. *Ibidem* p. 179

443 *Ibidem* p. 141

444 *Ibidem*

445 *Ibidem* p.142

Russia sovietica (Russell, quando scrisse il suo libro non sapeva, per quest'ultima, quali sarebbero stati i risultati). Tuttavia, egli riteneva che ogni uomo, dotato di sufficiente intelligenza ed energia, si sarebbe cimentato nell'impresa di costruire una società artificiale finché vi sarebbe stata una tecnica che lo avesse permesso, a ciò spinto dallo stimolo della soddisfazione che la costruzione pianificata gli avrebbe procurato⁴⁴⁶.

Sono evidenti le connessioni della computer vision con quanto affermato da Russell. Anzitutto, per le sue origini la computer vision è uno strumento elettro-meccanico e informatico che si adopera per riprodurre un processo biologico e psicologico. Secondo, per le sue applicazioni, orientate essenzialmente all'ispezione, al controllo e, una volta in possesso delle necessarie informazioni, alla decisione/scelta.

Un esempio potrebbe essere la video sorveglianza: oggi le telecamere a circuito chiuso si limitano a registrare delle immagini che devono essere successivamente analizzate dall'uomo; si pensi invece ad una telecamera di sorveglianza che riprende un tentativo di furto inviando le relative immagini a un computer e che quest'ultimo fosse capace di interpretare e comprendere la situazione, quindi avvisare la locale stazione di polizia con un collegamento telematico o, addirittura, azionare dei meccanismi per l'autodifesa. Una simile prospettiva scoraggerebbe sicuramente molti ladri. Ampliare questo tipo di video-sorveglianza a livello planetario significherebbe però realizzare l'occhio del Grande Fratello descritto da George Orwell in *1984*, prospettiva alquanto inquietante. Tuttavia vi è chi, come Asimov, non lo ritiene il peggiore dei futuri possibili ma, sarebbe anche disposto, in via ipotetica, a cedere il testimone: "Ma se i computer diventeranno più intelligenti degli esseri umani, non finiranno per sostituirsi a noi? Be', perché non dovrebbero farlo? Può darsi che oltre ad essere intelligenti siano anche di buon cuore, e che ci lascino consumare per usura. [...]"

Ma consideriamo anche tutto ciò che stiamo facendo in questo stesso momento: a noi, agli altri organismi viventi e al pianeta dove abitiamo. Forse è ora di essere sostituiti. Forse il vero rischio è che i computer non riescano ad arrivare abbastanza in fretta a un

446 Ibidem

livello di sviluppo sufficientemente elevato per sostituirci.

Pensiamoci!”⁴⁴⁷.

Quelle di Russell e Asimov sono due posizioni estreme, assunte, più o meno esplicitamente, per provocare l'interesse nell'opinione pubblica. Un simile intento, sebbene con dichiarazioni meno eclatanti, è anche quello di altri autori, come Zichichi e Stephen W. Hawking, quest'ultimo è tra i ricercatori e diffusori scientifici più conosciuti nell'ambito della cosmologia. In particolare, egli scrive: “Se accettiamo l'idea che non si può impedire alla scienza e alla tecnologia di modificare il nostro mondo, si può almeno tentare di assicurare che i mutamenti siano nelle direzioni giuste. In una società democratica, ciò significa che l'opinione pubblica ha bisogno di avere una comprensione di base della scienza, per poter prendere decisioni con cognizione di causa e non lasciare le decisioni ad esperti. Attualmente il pubblico ha un atteggiamento piuttosto ambivalente verso la scienza. Ha imparato ad attendersi il costante aumento del livello di vita che i nuovi sviluppi nella scienza e nella tecnologia apportano, ma è anche diffidente verso la scienza, perché non la capisce. Questa diffidenza è evidente nella figura dello scienziato folle dei fumetti, che lavora nel suo laboratorio a produrre un Frankenstein, ed è importante elemento di sostegno per i movimenti ecologisti.”⁴⁴⁸.

Di quanto affermato da Hawking, due concetti meritano particolare attenzione: comprensione e direzione. Comprensione che tecnologia e scienza sono comunque artefatti umani, che l'uomo ha voluto per migliorare la propria esistenza. Direzione, non solo come capacità di programmare e pianificare le tappe successive del processo, ma anche di prevederne i probabili sviluppi. Visto in quest'ottica, un sistema globale per rilevare i comportamenti umani, e non semplicemente dedicato ad applicazioni industriali, potrebbe incutere meno sgomento della prospettiva orwelliana. Comprendere che una seppur minima restrizione della propria libertà personale è il fondamento di qualsiasi convivenza/società, potrebbe portare ad accettarne una più restrittiva che, però,

447 Isaac Asimov, *Domani! 71 sguardi su un futuro migliore*, Op. cit. p. 95

448 Stephen W. Hawking, *Buchi Neri e Universi Neonati*, trad. it. di Libero Sosio, Milano, R.C.S. Rizzoli Libri, 1993 p. 50 (ed. originale *Black holes and baby universes and other essays* 1993)

consenta un aumento della qualità di vita e una maggiore autorealizzazione. In aggiunta, da un monitoraggio continuo e approfondito del comportamento umano si potrebbero inferire delle spiegazioni di alcuni fenomeni sociali come il suicidio, già oggetto di studio di Emile Durkheim (1858-1917) nel 1897⁴⁴⁹.

Sistemi globali di sorveglianza sono, per alcuni versi, già presenti. È il caso del controllo, tramite satellite, delle condizioni atmosferiche terrestri. Ma può far riflettere anche il monitoraggio fatto dalla polizia sulla quantità impressionante di video, acquisiti tramite “telefonini” e poi inseriti in rete. Vi sono i sistemi di visione per la lettura automatica delle targhe degli autoveicoli⁴⁵⁰ o, ancora, il moderno sistema usato dall'Unità Analisi Crimini Violenti (UACV) per memorizzare le scene dei delitti e quindi procedere alla loro analisi⁴⁵¹.

I vantaggi offerti dall'automazione in generale sono riportati da Marshall McLuhan nell'ultimo capitolo del suo libro “Gli strumenti del comunicare”, intitolato “Automazione – Imparare un modo di vivere”. Egli la considerava un processo che permetterà di superare la meccanizzazione, ossia la suddivisione di lavoro e delle relative competenze, così come descritte da Adam Smith⁴⁵² (1723-1790) e applicate da Frederick Winslow Taylor⁴⁵³ (1856-1915)⁴⁵⁴.

L'affermazione di McLuhan che “Il lavoro futuro sarà quello di imparare a vivere nell'era dell'automazione” è una previsione che si sta avverando. Le aziende che non si aggiornano tecnologicamente tendono a ridurre i profitti e uscire dal mercato; chi è in possesso di una formazione migliore o semplicemente riesce ad ottenere l'accesso a una maggiore conoscenza ha la possibilità ottenere posizioni sociali “migliori” o semplicemente più redditizie. La maggior parte degli oggetti che ci circonda richiedono, per essere utilizzati, una conoscenza pregressa, come la televisione o il

449 Cfr. Paolo Jedlowski, *Il mondo in questione*, Op. cit., p. 65

450 <http://www.pklab.net/kbase/targhe/> 18 febbraio 2007

451 <http://www.poliziadistato.it/pds/chisiamo/territorio/reparti/scientifica/uacv.htm> 18 febbraio 2007

452 http://it.wikipedia.org/wiki/Adam_Smith 18 febbraio 2007

453 http://en.wikipedia.org/wiki/Frederick_Winslow_Taylor 18 febbraio 2007

454 Marshall. McLuhan, *Gli strumenti del comunicare*, Op. Cit. p. 369

videoregistratore. Questa conoscenza riguarda sia il loro riconoscimento sia il loro funzionamento: non è sufficiente sapere che si tratta di una televisione, è necessario sapere che funziona tramite l'energia elettrica, che riceve le onde elettromagnetiche le quali, a loro volta, trasportano informazioni audio-video utili ad aumentare la nostra conoscenza del mondo o, più semplicemente, a farci sorridere. È necessario inoltre conoscere più processi, spesso in ambiti molto diversi tra loro. Non a caso, sempre McLuhan afferma: “Ogni materia affrontata in profondità suggerisce un rapporto immediato con altre”⁴⁵⁵. Sembra il ritorno ad un'epoca pre-industriale, dove l'artigiano doveva conoscere tutte le fasi del processo. Ma può essere anche interpretata come la constatazione che la nostra esistenza, la nostra fisicità è ancora unica. Se gli sviluppi della computer vision porteranno alla realizzazione di sistemi industriali e di robot in grado di sostituirci nella maggioranza dei compiti più ripetitivi e, sostanzialmente, poco creativi, il futuro lavorativo degli uomini consisterà nel comprendere e indirizzare il sistema. Questo in quanto gli esseri umani, più che le macchine, saranno in grado di individuare nuovi bisogni e nuove problematiche.

La riflessione sulla società è molto complessa e potrebbe essere espansa considerando come le strutture sociali evolveranno, come muterà la disposizione sul territorio, con possibili accentramenti o decentramenti. Tuttavia è preferibile per comprendere le possibili evoluzioni sociali esaminare anche le unità di cui sono composte, ossia considerare il pensiero delle singole persone, come ognuno concepisce o, meglio, è consapevole di sé e del suo ambiente⁴⁵⁶.

455 Ibidem p. 370

456 <http://it.wikipedia.org/wiki/Psicologia> 18 febbraio 2007

Conseguenze psicologiche

Nel 1986 Margaret A. Boden⁴⁵⁷, autorità internazionale nel campo dell'intelligenza artificiale e delle scienze cognitive⁴⁵⁸, affermò: “La maggior parte delle persone è profondamente scettica sulla possibilità che i calcolatori possano simulare la mente umana; per molti già la semplice idea è chiaramente assurda. Per di più, il loro scetticismo e il loro scherno sono di solito accompagnati dal timore. Oggi le teorie che assimilano la mente a un meccanismo appaiono anche più minacciose di quanto apparissero secoli fa simili riflessioni: e questo proprio perché sono ora disponibili risultanze più convincenti”⁴⁵⁹.

Al giorno d'oggi la situazione non è cambiata molto. Le conferme sono nelle dichiarazioni di Hawking, come quella riportata, e di Zichichi sui tempi medi necessari alla diffusione delle scoperte scientifiche: si perdono dai dieci ai vent'anni per passare dal fronte della ricerca avanzata all'insegnamento universitario nelle facoltà scientifiche; altrettanto tempo per arrivare all'insegnamento nella facoltà tecnologiche, e un'ulteriore decade per arrivare alla produzione industriale⁴⁶⁰. In Italia, tra l'altro, manca la propensione alla lettura, specie di materie scientifiche, ritenute dai più troppo complesse e perciò riservate ad esperti⁴⁶¹. La cultura è sempre più cinematografica e televisiva. Però, opere come Terminator (e i seguiti)⁴⁶², Nemico pubblico⁴⁶³, o Universal Soldier⁴⁶⁴, non ispirano certo fiducia visto che mostrano, in successione, la creazione di robot assassini, un controllo planetario dei mezzi di comunicazione e, infine, l'impianto di un microchip in un corpo umano, migliorato con la nanotecnologia, al fine di ottenere un

457 <http://www.sussex.ac.uk/informatics/profile276.html> 19 febbraio 2007

458 http://www.sussex.ac.uk/press_office/media/media190.shtml 20 febbraio 2007

459 Cfr. Margaret A. Boden, *La simulazione della mente al calcolatore è socialmente dannosa*, in *Mente umana, mente artificiale*, a cura di Riccardo Valle, Op. Cit. p. 4

460 Cfr. Antonio Zichichi, *Scienza ed emergenze planetarie*, Op. Cit. p. 40

461 Interessanti sono gli articoli reperibili in rete che riguardano l'atteggiamento degli italiani nei confronti della lettura: http://www.metronews.it/det.php?id=58762&fid=adnk_id e <http://www.aie.it/mercati/visualizza.asp?ID=41> 20 febbraio 2007

462 “*Terminator 2 – Il giorno del giudizio*”, di James Cameron, USA, 1991 (titolo originale, *Terminator 2: judgement day*) e “*Terminator 3 - Le macchine ribelli*”, di Jonathan Mostow, USA, 2003 (titolo originale, *Terminator 3 - Rise of the machines*)

463 “*Nemico Pubblico*”, di Tony Scott, USA, 1998 (titolo originale, *Enemy Of The State*)

464 “*Universal Soldier*”, di Mic Rodgers, USA, 1999 (titolo originale, *Universal Soldier: The Return*)

cyborg⁴⁶⁵ che si auto-dichiara poi successore dell'uomo.

Si può obiettare che questi lavori utilizzino la paura per far presa su un ampio pubblico, a fini economici. Ma anche le ricerche di Sigmund Freud⁴⁶⁶ (1856-1939) e Carl Gustav Jung⁴⁶⁷ (1875-1961) ritenevano che la paura e i desideri inconsci, individuali per il primo, più legati alla storia collettiva per il secondo, fossero elementi importanti per la comprensione delle azioni coscienti.

Non rimane che tentare di comprendere quali siano i timori legati alla riproduzione dei processi cognitivi nelle macchine. Considerando che i vari tipi di percezione sono fondamentali per lo sviluppo delle facoltà intellettuali (le aree del cervello che non sono soggette a stimoli non si sviluppano o si atrofizzano), la computer vision può veramente portare un incremento nello sviluppo dell'intelligenza artificiale, quindi alla costruzione di macchine con capacità umane. Macchine che saranno in grado, forse, di sostituirci. È solo questa la nostra angoscia? Forse c'è di più. Al di là della paura di morire, c'è soprattutto la paura dell'ignoto. Queste angosce sono come delle prigioni psichiche, che impediscono di progredire nella conoscenza del mondo⁴⁶⁸.

Margaret Boden cerca di comprendere queste paure e la disinformazione su cui si fondano: “Ciò che conta è che l'uomo della strada, nella grande maggioranza dei casi, crede che l'Intelligenza Artificiale sia la negazione dei tratti specifici e delle peculiarità essenziali dell'essere umano, e ciò a causa delle conseguenze teoriche di tale disciplina a proposito della natura della mente umana. In altre parole, la maggioranza teme che, se permettiamo che l'immagine dell'uomo sia forgiata a somiglianza del calcolatore, i valori umani debbano passare in secondo piano o essere del tutto negati.”⁴⁶⁹; e ancora, per quanto riguarda la conoscenza di noi stessi: “Le teorie sociali influenzano il modo in cui concepiamo noi stessi e il nostro potenziale umano: cambiando l'immagine di noi

465 <http://it.wikipedia.org/wiki/Cyborg> 20 febbraio 2007

466 http://it.wikipedia.org/wiki/Sigmund_Freud 20 febbraio 2007

467 http://it.wikipedia.org/wiki/Carl_Gustav_Jung 20 febbraio 2007

468 Gareth Morgan, *IMAGE – Le metafore dell'organizzazione*, trad. it. di Massimo Balducci, Milano, Franco Angeli, 2002 p. 277 (ed. originale, *Images of Organization*, 1997)

469 Margaret A. Boden, *La simulazione della mente al calcolatore è socialmente dannosa*, in *Mente umana, mente artificiale*, a cura di Riccardo Valle, Op. Cit. p. 5

stessi, essi ci cambiano. Le ipotesi, spesso inconfessate e largamente inconsece, riguardo a quel che gli esseri umani sono e possono essere, contribuiscono a determinare la nostra vita quotidiana per l'intima connessione che le lega alla nostra morale e al nostro morale.”⁴⁷⁰

Insomma, sembra che se risultasse vera l'equazione che equipara l'uomo a un insieme di processi riproducibili a loro volta su delle macchine, se ne potrebbe trarre che alcune caratteristiche distintive dell'uomo, come pensare, provare sentimenti e scegliere, non esistono in realtà o, comunque, non gli appartengono in via esclusiva, privandolo così della sua dignità in quanto “non siamo altro che macchine”⁴⁷¹.

Ma al momento le differenze tra il cervello umano e i calcolatori sono ancora evidenti: il primo funziona in modo analogico, parallelo, specializzato; i secondi sono dispositivi digitali, seriali, generali⁴⁷².

Il fatto che nessuno sia riuscito, per ora, a produrre un computer “pensante”, lascia spazio alla possibilità che la mente non sia divisibile dal cervello⁴⁷³.

Tuttavia, le ricerche che si pongono come obiettivo la riproduzione della percezione visiva consentono di supporre che l'impresa sia possibile. Le reti neurali degli anni '40 riproducevano la struttura del cervello, ma non raggiungevano la capacità di vedere, in quanto riuscivano a distinguere alcune configurazioni (bordi, contorni, superfici, ecc), ma non ad interpretarle. Quando fu chiaro che la visione non consiste nel semplice riconoscimento di forme, ma richiede un'interpretazione dell'immagine come rappresentazione del mondo reale, si passò a tecniche di computer vision⁴⁷⁴. Come esposto in precedenza, per interpretare un'immagine questa utilizza sia la geometria proiettiva sia delle conoscenze relative all'ambiente considerato. Di recente si è verificato un ritorno alle reti neurali, in quanto dotate della capacità di auto-

470 Ibidem p. 6

471 Cfr. Ibidem p. 7

472 Ibidem p. 8

473 Ibidem p. 9

474 Ibidem p. 10

apprendimento. Questi ultimi sistemi, definiti “connessionistici”, si differenziano dai precedenti anche per le capacità di fornire interpretazioni tridimensionali degli oggetti e la sostanziale indipendenza delle singole unità, per la progettazione delle quali si tiene conto delle teorie dell'interpretazione delle immagini sorte nel frattempo⁴⁷⁵. Tuttavia, come in una rete sinaptica, tali unità possono influenzare il volume delle informazioni (stimoli) scambiate, attraverso l'implementazione di meccanismi di feedback.

La successione di questi tentativi di riproduzione della visione è stata utile anche alla psicologia, che ha potuto verificare alcune delle proprie ipotesi implementandole sui calcolatori. Realizzare un modello sul calcolatore permette infatti di comprendere come un dato sistema deve assolvere un determinato compito e in che modo può farlo. Permette di analizzarne i processi e le funzioni computazionali in modo indipendente dalla struttura fisica sottostante. Questo è, probabilmente, il vero campo d'indagine della psicologia, la caratteristica che la distingue dalle scienze naturali, come la fisica o la neurofisiologia. Queste ultime non riescono a chiarire come due persone possano vedere il mondo in modo diverso, o avere opinioni divergenti, pur avendo delle strutture cerebrali simili. Anzi, per le scienze naturali queste problematiche sono spesso irrilevanti⁴⁷⁶.

Comprendere in modo più approfondito come funziona l'intelligenza artificiale e scoprirne i potenziali vantaggi è d'aiuto per superare alcuni timori, ma è necessario riprendere e approfondire anche altri aspetti: i computer possono realmente “pensare”? È credibile un'ipotesi di *strong AI*⁴⁷⁷? O è più probabile quella di una *weak AI*? Quali sono le differenze tra l'uomo e la macchina, specie nel primo caso?

La risposta al primo quesito non è scontata, tanto d'aver formato due fazioni contrapposte. Da una parte chi, come Daniel C. Dennett⁴⁷⁸, noto filosofo americano, ritiene che ciò sia possibile: è la prospettiva forte dell'intelligenza artificiale. Dall'altra

475 Ibidem p. 11

476 Ibidem p. 14

477 http://en.wikipedia.org/wiki/Strong_ai 21 febbraio 2007

478 http://en.wikipedia.org/wiki/Daniel_Dennett 21 febbraio 2007

c'è la corrente, campeggiata da John Searle⁴⁷⁹, della prospettiva debole.

È curioso che sia stato Searle a coniare il termine *strong AI* definendolo, in un suo articolo del 1980 intitolato “Minds, Brains, and Programs” come: “[...] according to strong AI, the computer is not merely a tool in the study of the mind; rather, the appropriately programmed computer really is a mind, in the sense that computers given the right programs can be literally said to understand and have other cognitive states. In strong AI, because the programmed computer has cognitive states, the programs are not mere tools that enable us to test psychological explanations; rather, the programs are themselves the explanations.”⁴⁸⁰, che si riassume essenzialmente in “... secondo l'intelligenza artificiale forte il computer non sarebbe soltanto uno strumento nello studio della mente; piuttosto, un computer opportunamente programmato è letteralmente una mente”.

Accettando questa definizione e date le attuali tecnologie, sembra difficile che si possa, nel breve termine, giungere alla reale realizzazione di sistemi di *strong AI*. Sembrano invece fattibili sistemi di intelligenza artificiale debole, in quanto è facile constatare come un computer elabori una serie di simboli che non comprende, eseguendo i suoi compiti meccanicamente.

La contrapposizione per il momento è quindi più teorica che pratica: è difficile considerare tutte le informazioni e le argomentazioni e trarne una conclusione certa. È preferibile, al limite, indicare alcune caratteristiche che il sistema di *strong AI* deve possedere⁴⁸¹ per essere definito tale.

Secondo Daniel C. Dennett una di queste caratteristiche potrebbe essere la velocità del sistema. Egli infatti afferma che se si paragonano due sistemi che eseguono lo stesso programma si avrà che “[...] essi hanno le stesse capacità (entrambi “computano la

479 <http://socrates.berkeley.edu/~jsearle/> e http://en.wikipedia.org/wiki/John_Searle 21 febbraio 2007

480 <http://www.ptproject.ilstu.edu/STRONGAI.HTM> 27 febbraio 2007

481 Cfr. Daniel C. Dennett, *Il mito dell'intenzionalità originaria*, in *Mente umana, mente artificiale*, a cura di Riccardo Valle, Op. Cit. p. 24

stessa funzione”), ma grazie alla sua maggiore velocità uno di essi avrà “poteri causali” di cui l'altro manca, e cioè i poteri di controllo necessari, ad esempio, per guidare un corpo in movimento nel mondo reale. È per questo che si può sostenere che il sistema veloce è “letteralmente una mente”, mentre lo stesso non si può sostenere del suo gemello più lento⁴⁸²; e ancora: “Un calcolatore adeguatamente programmato e sufficientemente veloce da interagire con il mondo esterno in tempo reale ha effettivamente una mente, nel senso letterale dell'espressione, indipendentemente dalla sua costituzione materiale, organica e inorganica”⁴⁸³.

La posizione di Dennett, focalizzata su un aspetto fisico, risulta, per certi versi, convincente. Tuttavia, se consideriamo da una parte l'aumento di prestazioni dei calcolatori verificatosi negli ultimi decenni, senza che vi sia stato l'apparire di un sistema dotato di “poteri causali” e, dall'altra, il fatto che, nonostante vi siano molti animali dotati di cervelli simili all'uomo, solo quest'ultimo possiede alcune capacità particolari, si può ritenere vi debbano essere delle altre caratteristiche.

Una di queste dovrebbe essere il “senso dell'io”⁴⁸⁴, come lo definisce Domenico Parisi⁴⁸⁵, noto ricercatore nell'ambito dell'intelligenza artificiale. La coscienza⁴⁸⁶ di sé o auto-consapevolezza è addirittura, seguendo il pensiero di Cartesio, prova della nostra stessa esistenza: non a caso egli coniò la locuzione: “dubito ergo cogito, cogito ergo sum”. Partendo dalla considerazione che si debba considerare vero solo ciò che è evidente, egli iniziò a dubitare di tutto: delle proprie percezioni, della propria realtà corporea, delle scienze matematiche, della nostra origine, del mondo che ci circonda. Di questo processo/metodo vi è, secondo Cartesio, una sola certezza: il fatto che finché dubito è evidente che penso e quindi esisto come sostanza pensante. La conclusione è criticabile, se è vero che non ho nessuna certezza, come posso essere certo che sto pensando? Cartesio stesso comunque rispose affermando che la sua conclusione non era

482 Ibidem p. 25

483 Ibidem p. 26

484 Cfr. Domenico Parisi, *Senso dell'io*, in *Mente umana, mente artificiale*, a cura di Riccardo Valle, Op. Cit. pp. 327-334

485 <http://www.domenicoparisi.it/index.htm> 21 febbraio 2007

486 <http://www.cicap.com/enciclop/at100078.htm> e <http://home.att.net/~mstaique/Italian/consciob.htm> 21 febbraio 2007

frutto di un ragionamento discorsivo, ma un'intuizione immediata, che chiunque sta pensando ha, con evidenza, di sé⁴⁸⁷.

Com'è possibile che una macchina acquisti auto-consapevolezza? Parisi constata che l'uomo è in grado di produrre frasi che contengono la parola io, quindi lo pone come un possibile parametro di misurazione per il “senso dell'io”. Per arrivare a tale obiettivo si possono implementare nella macchina varie strategie, in ordine crescente:

1. Memorizzando le frasi nel calcolatore e predisponendolo ad utilizzarle quando gli vengono poste delle domande che lo riguardano, come: “dove sei nato?”. Questa soluzione non è soddisfacente in quanto assimilabile all'estrazione di dati da un database, tuttavia è un primo passo. Parisi, ipotizzando che la conoscenza-memoria abbia una struttura a grafo, dichiara che si è nella situazione in cui “La conoscenza è già attaccata al nodo”⁴⁸⁸.
2. Un miglioramento si ha se il calcolatore può inferire la risposta dalle sue conoscenze pregresse, senza che questa sia già disponibile. Parisi: “La conoscenza non è attaccata al nodo, ma la macchina la genera internamente inferendola dalle conoscenze già esistenti.”⁴⁸⁹.
3. “La conoscenza non è attaccata al nodo, ma la macchina la acquisisce attraverso la percezione esterna”⁴⁹⁰.

La differenza, rispetto ai punti precedenti è notevole. Qui si presuppone che la macchina non sia dotata solo di circuiti per “far girare” il software, ma anche di sensori, o meglio di un corpo, che permette di percepire l'ambiente e sé stesso in quanto “è tale corpo che deve essere visto, sentito, toccato dal sistema per acquisire

487 <http://www.filosofico.net/desc4.htm>, http://it.wikipedia.org/wiki/Cogito_ergo_sum e <http://www.forma-mentis.net/Filosofia/Cartesio.html#3> 21 febbraio 2007

488 Cfr. Domenico Parisi, *Senso dell'io*, in *Mente umana, mente artificiale*, a cura di Riccardo Valle, Op. Cit. pp. 330

489 Ibidem

490 Ibidem

le conoscenze da usare nelle risposte⁴⁹¹.

4. “La conoscenza non è attaccata al nodo, ma la macchina la acquisisce attraverso la percezione interna.”⁴⁹²

La percezione esterna può essere accessibile a sistemi diversi: la posizione di un oggetto, la forma di una mano, un suono, ecc possono essere rilevate da più uomini, sistemi di acquisizione, ecc

Non così la percezione interna, riservata in via prioritaria, se non esclusiva⁴⁹³, al singolo sistema. Forse in questo caso è possibile riconoscere al sistema un inizio di consapevolezza, in quanto “[...] per il sistema il suo corpo – un nodo tra gli altri nel suo deposito di conoscenze – è almeno in parte ciò su cui acquisisce conoscenze mediante la percezione interna, cioè la percezione riservata a lui”⁴⁹⁴.

Da sottolineare, come sostiene Parisi, che è importante non solo possedere una percezione interna, ma anche riuscire ad esprimerla con il linguaggio formulando frasi contenenti la parola “io” che si basano su questa. Questo, infatti, funge alla dimostrazione della propria coscienza.

5. “La conoscenza non è attaccata al nodo, ma la macchina la acquisisce esaminando il proprio software”⁴⁹⁵.

Quest'ultima possibilità si ricollega alla concezione dualistica che si ha in prima istanza dell'uomo, che viene suddiviso in anima e corpo, in cervello e mente, ecc, e successivamente dei calcolatori, suddivisi in hardware e software. Questo concetto, dato per assodato, è pressoché indiscutibile: ne è prova la sua presenza in quasi

491 Ibidem p. 331

492 Ibidem

493 Le eccezioni all'esclusività dell'accesso potrebbero essere, per l'uomo, da un lato la psicanalisi, dall'altro l'uso di sonde introspettive.

494 Cfr. Domenico Parisi, *Senso dell'io*, in *Mente umana, mente artificiale*, a cura di Riccardo Valle, Op. Cit. pp. 331

495 Ibidem

tutte le culture, per dare spiegazione a fatti e fenomeni che vengono osservati⁴⁹⁶. Tuttavia, vi è stato anche chi ha sollevato obiezioni, come San Tommaso D'Aquino⁴⁹⁷ (1224/1225 – 1274), facendone emergere alcuni difetti sostanziali. In effetti, da quanto finora detto, sembra difficile ma possibile dividere il cervello dalla mente, mentre è più semplice la distinzione tra software e hardware.

Per l'ultimo caso è utile considerare che oggi, grazie a tecnologie come Java, abbiamo del software abbastanza indipendente dalla piattaforma sottostante, e questo facilita nel concepire due categorie diverse. Ma non è sempre stato così, e lo sa bene chi si occupa di *porting* tra sistemi diversi. Per programmare in Assembler, uno dei linguaggi più vicini alla macchina, è necessario sapere con esattezza le caratteristiche del processore su cui il software dovrà “girare”. Il sistema operativo di un palmare è essenzialmente diverso da quello di un pc e, in successione, da quello di un Mac. Se si risale alle origini dell'informatica, sono numerosi gli esempi di operazioni che erano implementate a livello hardware e, quindi, indivisibili da questo (l'esempio classico è il regolo calcolatore).

Ad ogni modo, se si decide di accettare la distinzione tra hardware e software, si comprende come un sistema in grado di esaminare il proprio software⁴⁹⁸ e ricavarne conoscenze nuove, quali i suoi scopi, capacità, limiti, stato presente, ecc, possa essere considerato in possesso del senso dell'io⁴⁹⁹.

In sintesi, Parisi ritiene che, per attribuire un senso dell'io ad un sistema, questo non debba solo avere delle conoscenze su di sé, ma essere in grado di acquisirne di nuove relative sia al proprio “corpo” sia alla propria “mente”. Le conoscenze così ottenute devono poi essere espresse tramite il linguaggio, in quanto “Saper usare il linguaggio (umano) richiede che ciò che il sistema sa sia rappresentato a un livello che viene

496 http://it.wikipedia.org/wiki/Unit%C3%A0_psicofisica 22 febbraio 2007

497 http://it.wikipedia.org/wiki/Tommaso_d%27Aquino 22 febbraio 2007

498 Da non confondere con il caso precedente dove si inferiva da dati memorizzati nel database, qui è tutto il software della macchina ad essere esaminato, l'operazione è assimilabile a una riflessione su sé stessi.

499 Cfr. Domenico Parisi, *Senso dell'io*, in *Mente umana, mente artificiale*, a cura di Riccardo Valle, Op. Cit. pp. 332-333

usualmente chiamato “concettuale”.⁵⁰⁰ Così, attraverso la formulazione di frasi che lo riguardano, il sistema dimostra di aver collegato a sé queste conoscenze, in essenza, si riconosce⁵⁰¹.

La conclusione di Parisi evidenzia la stretta interconnessione presente tra corpo e conoscenza. Il corpo è necessario per la memorizzazione/contenimento della conoscenza, ma ancor più per il suo sviluppo. L'apprendimento, l'inferenza di nuova conoscenza, avviene attraverso la percezione, sia esterna sia interna, prima del “corpo” quindi della “mente”. Se ne può trarre che lo studio di tecnologie che riproducono la percezione e il funzionamento del corpo sono essenziali per lo sviluppo della conoscenza. Percepire/apprendere è la condizione necessaria per operare/esistere nella complessità del modo reale.

Pietro Moraso e Vincenzo Tagliasco, rispettivamente esperti di robotica e bioingegneria, affermano, riferendosi alla rappresentazione della conoscenza: “[...] l'estrema articolazione e ricchezza delle prestazioni umane non può essere completamente capita se non prendiamo in considerazione la stretta interazione tra l'operatore umano e l'ambiente che lo circonda”⁵⁰²; e ancora “[...] sembra ragionevole ipotizzare che il classico principio ingegneristico – secondo il quale il “miglior” modo di capire un sistema intelligente consiste nel costruire una struttura atta a funzionare come il sistema originale – potrebbe costituire un'efficace tecnica di simulazione del complesso meccanismo che sta alla base del funzionamento del cervello. Secondo questo approccio, uno dei modi per formulare precise teorie potrebbe essere quello di mettere a punto modelli computazionali.”⁵⁰³. L'esempio a supporto di queste loro affermazioni è il riconoscimento visivo di oggetti tridimensionali, che richiede un numero di operazioni così elevato che solo la sua implementazione su computer può validare o meno il modello teorico proposto, e quindi portare a successive considerazioni/teorie⁵⁰⁴. Si tratta

500 Ibidem p. 334

501 Ibidem p. 333

502 Pietro Moraso e Vincenzo Tagliasco, *Conoscenza visiva e motoria*, in *Mente umana, mente artificiale*, a cura di Riccardo Valle, Op. Cit. p. 336

503 Ibidem pp. 335-336

504 Ibidem

essenzialmente di tecniche di computer vision.

Queste affermazioni richiamano quelle di Margaret A. Boden: esse, infatti, riportano al vantaggio offerto dall'intelligenza artificiale e, in generale, dalle nuove tecnologie. Rimangono da specificare le differenze tra uomo e macchina nel caso fosse realizzabile l'ipotesi di *strong AI*. La questione non è molto diversa da quella che si pone Zichichi: "L'uomo è una delle numerose forme di materia vivente che esistono su questo pianeta. Ed ecco la prima domanda fondamentale: in che cosa, questa forma di materia vivente, si distingue dalle altre?"⁵⁰⁵.

La prima risposta, che egli propone, è di tipo trascendentale: "[...] l'uomo è l'unica forma di materia vivente che riesca a concepire Dio. Questa risposta non può essere accettata da un ateo, per il fatto semplicissimo che lui, in Dio, non crede. Se non fosse per la Scienza, il discorso tra credenti e non credenti si fermerebbe qui. Lo scoglio logico risulta infatti insormontabile."⁵⁰⁶. La seconda risposta è molto simile: "[...] se vincoliamo lo studio della nostra esistenza alla sfera immanentistica, siamo costretti a concludere che c'è poco da sperare: non è possibile distinguere, in modo netto, l'uomo dagli animali. Anche essi mangiano, dormono, lavorano, cacciano, soffrono, gioiscono, apprezzano la musica e amano la pittura e forse sanno anche contare"⁵⁰⁷.

Zichichi evidenzia così i due approcci alla problematica: quello materialista⁵⁰⁸ e quello immaterialista⁵⁰⁹. È innegabile, come sosteneva Paul Feyerabend (1924-1994), che le opinioni e le credenze influenzino le decisioni e ciò che ognuno è disposto ad accettare o meno. Feyerabend portava ad esempio il caso di Galileo Galilei, fondatore del metodo scientifico.

Se affrontata con un approccio materialista, la questione è quindi rilevante. In questo

505 Antonio Zichichi, *Perché io credo in colui che ha fatto il mondo*, Milano, Il Saggiatore, 2006 p. 67

506 Ibidem

507 Ibidem p. 68

508 <http://it.wikipedia.org/wiki/Materialismo> 22 febbraio 2007

509 http://www.filosofico.net/Antologia_file/AntologiaB/BERKELEY_%20I%20VANTAGGI%20DEL_L%20IMMATE.htm 22 febbraio 2007

caso, la differenza tra le macchine e uomo (e, in generale, gli animali), anche nel caso di *strong AI*, si trova nel substrato materiale. Le macchine si basano su strutture in silicio e ferro, gli esseri viventi su strutture di carbonio. È una distinzione importante, non solo ai fini della possibilità di pensare, come ritiene Searle⁵¹⁰, ma anche per quanto riguarda le origini della vita⁵¹¹. Non a caso, nel suo esperimento di Stanley Lloyd Miller, nel 1952, utilizzò come base una miscela d'acqua, ammoniacca, metano e idrogeno, e vi fece scorrere una scarica elettrica per simulare l'energia solare. Ne ottenne, dopo alcuni giorni, molecole più complesse, tra cui glicina e alanina, due tra gli amminoacidi più semplici presenti nelle proteine⁵¹². Sembra, almeno per ora, che con silicio e ferro, non si ottengano gli stessi risultati. In secondo luogo, seguendo l'ipotesi di Darwin, l'uomo è il risultato di un processo evolutivo durato milioni, se non miliardi, di anni; le macchine sono invece una conquista umana molto recente. Secondo Vincenzo Tagliasco, per acquisire capacità simili all'uomo, è necessario che le macchine siano dotate di processi di apprendimento e adattamento all'ambiente⁵¹³.

In conclusione la computer vision e l'intelligenza artificiale sono e probabilmente rimarranno strumenti utili all'uomo, ma difficilmente saranno in grado di sostituirlo completamente, in forza di una diversa natura. Se vi è una dipendenza, questa è quella delle macchine nei confronti dell'uomo, che dovrà prevedere a che fini costruirle. Questi strumenti permetteranno all'uomo di comprendere meglio sé stesso, di avere a disposizione un'ampia gamma di scelte, di migliorare sé stesso e il mondo in cui vive. L'auspicio per il futuro è che non vi sia una contrapposizione uomo-macchina, ma una sempre maggiore cooperazione con sistemi realmente intelligenti per un miglioramento sia della condizione umana che dell'ambiente-contesto planetario.

510 Cfr Nils J. Nilsson, *Intelligenza artificiale*, Op. Cit., p. 23

511 Cfr. Isaac Asimov, *Civiltà extraterrestri*, Op. Cit., p. 158-167

512 Ibidem p. 166

513 http://it.wikipedia.org/wiki/Intelligenza_artificiale 22 febbraio 2007

Conclusioni

La trattazione di aspetti biologici, psicologici, tecnici e computazionali è stata utile per comprendere l'ampio campo di ricerca della computer vision e la tecnologia che ne deriva. Comprensione che dovrebbe evidenziare le connessioni con l'intelligenza artificiale e la vita artificiale.

Nel passato recente, queste materie sono state ambito esclusivo di addetti ed esperti, ma ora stanno progredendo, e le loro applicazioni incideranno su molti aspetti della vita di ogni essere umano; si è quindi tentato di delinearne gli scenari economici, sociali e individuali.

A livello economico sono già presenti sui mercati diversi prodotti basati sulla computer vision che si posizionano in settori abbastanza diversificati. Oltre a nuovi strumenti di visualizzazione e acquisizione dedicati allo sviluppo della tecnologia stessa, sono commercializzati apparecchi per il controllo dei singoli autoveicoli, e del traffico in generale; nell'industria cinematografica e televisiva sono stati introdotti nuovi sistemi per il tracciamento degli oggetti o dei volti umani utili per le trasmissioni di eventi sportivi o per animare i personaggi; in medicina vi sono sistemi per l'analisi quantitativa di immagini utili, ad esempio, per la diagnosi del cancro al seno, o per far combaciare il colore protesico dei denti con il colore naturale dei denti del paziente. Queste sono solo alcune applicazioni di un lungo elenco che risulta tuttavia limitante in quanto la vera innovazione si sta verificando con il rinnovo dei processi produttivi. La natura di questa tecnologia, votata essenzialmente al controllo, permette e renderà sempre più semplice l'automatizzazione degli stessi. L'intervento umano sarà probabilmente sempre più ridotto, si avranno degli automi in grado di svolgere compiti diversi, addirittura in grado di decidere quanto, come e cosa produrre. Sistemi tra loro integrati, capaci di percepire l'ambiente che li circonda saranno in grado di valutare e decidere autonomamente quali sono le necessità umane da soddisfare in via prioritaria e quali risorse utilizzare. I processi decisionali finora spettanti all'imprenditore e al suo staff saranno condivisi o

delegati ai sistemi intelligenti. All'uomo, liberato da compiti operativi e decisionali di breve termine, spetteranno principalmente funzioni di controllo generale, d'indirizzo e ricerca per nuovi campi del sapere.

Le conseguenze sociali sono strettamente legate a quelle economiche. Se migliorano le tecnologie impiegate in agricoltura, industria e servizi, aumentano proporzionalmente le risorse disponibili in termini di quantità fisiche che di tempo utilizzabile. Quindi, essenzialmente, aumenteranno per ognuno le possibilità di scelta. Questo, probabilmente, significherà più tempo disponibile per apprendere, per acquisire consapevolezza sul proprio essere, o per curare il proprio corpo. Si avrà anche una più equa ripartizione sul territorio in quanto alcune zone ora ostili (es: deserti, ma anche zone periferiche dove scarseggiano i servizi) diverranno più confortevoli, in parte per un migliore utilizzo delle risorse naturali, in parte per la possibilità di avere degli automi con competenze mediche o di sicurezza/salvaguardia, o per l'accesso di queste conoscenze da parte di tutti. Il maggior tempo a disposizione e la possibilità di reperire facilmente informazioni, specie visive, permetterà anche una maggiore sensibilità ai problemi sociali e una più attiva partecipazione per la loro risoluzione. In sintesi, queste tecnologie, affiancate ad altre, come la rete internet, consentiranno dei cambiamenti sociali radicali, di cui si scorgono solo gli inizi.

Se a livello economico e sociale si possono intravedere più vantaggi che svantaggi, a livello psicologico si presentano le maggiori difficoltà. È probabilmente un meccanismo di autodifesa che crea nell'uomo diffidenza verso ciò che non conosce. Queste paure si possono rilevare sia dall'osservazione del comportamento di singoli individui, sia di alcune opere culturali umane come i film di fantascienza, che mostrano la lotta dell'uomo contro la tecnologia. In verità, le paure inconsce nascono dalla mancata conoscenza dell'origine umana. Le ricerche sull'origine dell'universo e della vita sono lontane dal raggiungere un risultato. È forse per questo motivo che, la possibilità che una macchina replichi le capacità umane, incute paura a molti. Tuttavia, una visione più razionale non può che dimostrare come la computer vision e l'intelligenza artificiale siano utili non solo per le loro applicazioni ma anche per giungere a una maggiore

conoscenza dell'uomo. Infatti, sembra che il modo più semplice per comprendere il funzionamento di un sistema complesso è costruirne uno che funzioni in modo simile. La costruzione di sistemi intelligenti potrà portare a delle conoscenze sull'uomo e sull'ambiente maggiori di quelle attuali, contribuendo al contempo a dipanare timori e paure. È necessario andare oltre la contrapposizione uomo-macchina per poter trarre vantaggio dalla costruzione di sistemi realmente intelligenti.

Da quanto esposto credo d'aver mantenuto i propositi iniziali di fornire un primo approccio all'argomento, complesso e in continua evoluzione. In particolare d'aver evidenziato e chiarito le relazioni di questa tecnologia con altri campi del sapere dimostrando, al contempo, che ogni innovazione è frutto delle conoscenze e della cultura presenti fino a quel momento, e costituisce la base per ulteriori sviluppi. In questo senso, si può ben accogliere l'ipotesi evolutiva di Darwin.

Questo tipo di riflessioni sono strettamente collegate alla figura del tecnologo, che per definizione è esperto e studioso di tecnologia. Una persona, quindi, che non solo apprende e utilizza la tecnologia presente, ma che si dovrebbe impegnare per prevedere, conoscere e diffondere quella a venire. Questa è un'esigenza costante nell'era dell'automazione in cui si vive al giorno d'oggi.

La computer vision, come tecnologia, può essere identificata come un processo di controllo; tuttavia, essa può essere collocata nell'ambito più ampio della comunicazione tra strumenti, intesa sia all'interno di una singola unità, con gli strumenti di acquisizione lontani dal centro di elaborazione, sia fra più unità o anche in una struttura a rete, permettendo quindi operazioni diversificate e condivise (elaborazione, archiviazione, ricerca) sulle immagini.

Queste ultime sono alla base della computer vision, per il semplice motivo che, principalmente attraverso di esse l'uomo si relaziona con il proprio ambiente. Esse sono probabilmente fondamentali per come egli concepisce sé stesso e la propria realtà.

Bibliografia

- Antonio Marazzi**, *Antropologia della visione*, Roma, Carocci, 2003
- Antonio Zichichi**, *Perchè io credo in colui che ha fatto il mondo*, Milano, Il Saggiatore, 2006
- Antonio Zichichi**, *Scienza ed emergenze planetarie*, Milano, R.C.S. Libri e Grandi Opere, 1993
- Ariana Fiorentini, Lamberto Maffei**, *Arte e cervello*, Bologna, Zanichelli, 1995
- Arnaldo Bagnasco, Marzio Barbagli, Alessandro Cavalli**, *Corso di sociologia*, Bologna, Il Mulino, 1997
- Arrigo Opocher**, *Lezioni di economia politica*, Padova, Cedam, 1995
- Ben Long**, *Fotografia Digitale. Il Manuale. 3A ed.*, traduzione italiana di Riccardo Mori, Milano, Apogeo, 2005 (edizione originale *Complete Digital Photography*, 2002)
- Bertrand Russell**, *La visione scientifica del mondo*, traduzione italiana di Emilio A. G. Loliva, Roma-Bari, Gius. Laterza & Figli, 2004 (edizione originale *The Scientific Outlook*, 1931)
- David A. Forsyth, Jean Ponce**, *Computer Vision: A Modern Approach*, Prentice Hall, 2002
- David H. Hubel**, *Occhio, cervello e visione*, traduzione italiana di Enrica Maria Fava, Bologna, Nicola Zanichelli, 1989 (edizione originale *Eye, Brain and Vision*, 1988)
- Emanuele Trucco, Alessandro Verri**, *Introductory Techniques for 3-D Computer Vision*, New Jersey, Prentice Hall, 1998
- Emilio Garroni**, *Immagine, linguaggio, figura*, Roma-Bari, Gius. Laterza & Figli, 2005
- Francesco Grassivaro**, *Le Imprese Multinazionali*, Padova, Cedam, 1991
- Francesco Grassivaro**, *Economia Politica – Teorie e modelli*, Padova, Cleup, 1992
- Gareth Morgan**, *IMAGE – Le metafore dell'organizzazione*, traduzione italiana di Massimo Balducci, Milano, Franco Angeli, 2002 (edizione originale *Images of Organization*, 1997)

George Berkeley, *Saggio su una nuova teoria della visione – Trattato sui principi della conoscenza umana*, traduzione italiana di Daniele Bertini, Milano, Bompiani – R.C.S. Libri, 2004 (edizione originale *An Essay Towards A New Theory of Vision – A Treatise Concerning the Principles of Human Knowledge*, 1709-1710)

Gianni Losito, *Il potere del pubblico*, Roma, Carocci, 2002

Hal R. Varian, *Microeconomia*, traduzione italiana di Giulio Codognato, Stefano Chinellato, Venezia, Cafoscarina, 1998 (edizione originale *Intermediate Microeconomics. A Modern Approach*, 1987)

Isaac Asimov, *Il libro di Biologia*, traduzione italiana di Alessandra Fois, Milano, Arnoldo Mondadori, 1987 (edizione originale *Asimov's New Guide to Science*, 1984)

Isaac Asimov, *Il libro di Fisica*, traduzione italiana di Carla Sborgi, Milano, Arnoldo Mondadori, 1986 (edizione originale *Asimov's New Guide to Science*, 1984)

Isaac Asimov, *Civiltà extraterrestri*, traduzione italiana di Paola Cusumano, Massimo Parizzi, Milano, Arnoldo Mondadori, 1986 (edizione originale *Extraterrestrial Civilizations*, 1979)

Isaac Asimov, *Domani! 71 sguardi su un futuro migliore*, traduzione italiana di Riccardo Valla, Trezzano sul Naviglio, Euroclub, 1990 (edizione originale *Change!*, 1982)

Isaac Asimov, *Sogni di Robot*, traduzione italiana di Mauro Gaffo, Milano, Interno Giallo, 1990 (edizione originale *Robot dreams*, 1986)

Italo Calvino, *Tutte le cosmicomiche*, traduzione italiana di Claudio Milanini, Milano, Arnoldo Mondadori, 2002

Lev Manovic, *Il linguaggio dei nuovi media*, traduzione italiana di Roberto Merlini, Milano, Fres, 2004 (edizione originale *The Language of New Media*, 2001)

Linda G. Shapiro, George C. Stockman, *Computer Vision*, New Jersey, Prentice Hall, 2001

Marshall McLuhan, *Gli strumenti del comunicare*, traduzione italiana di Ettore Capriolo, Milano, Net, 2002 (edizione originale *Understanding Media*, 1964)

Martin D. Levine, *Vision in Man and Machine*, USA, McGraw-Hill, 1985

Nils J. Nilsson, *Intelligenza artificiale*, traduzione italiana di Salvatore Gaglio, Apogeo, 2002 (edizione originale *Artificial Intelligence: a New Synthesis*,)

Paolo Frignani, Paola Rizzati, *Didattica della comunicazione*, San Bartolomeo in

Bosco – Ferrara, Tecom Project, 2003

Paolo Jedlowski, *Il mondo in questione*, Roma, Carocci, 1998

Piergiorgio Odifreddi, *Matematico – Strumenti razionali per la comprensione del mondo*, Bologna, Zanichelli, 2005

Piero Angela, *QUARK Economia – per capire un mondo che cambia*, Milano, Garzanti, 1987

Ramesh Jain, Rangachar Kasturi, Brian G. Schunck, *Machine Vision*, , McGraw-Hill, 1995

Richard L. Gregory, *Occhio e cervello. La psicologia del vedere.*, traduzione italiana di Alberta Rebaglia, Milano, Cortina Raffaello, 1998 (edizione originale *Eye and Brain*, 1998)

Roberto Marangoni, Marco Geddo, *Le immagini digitali*, Milano, Ulrico Hoepli, 2000

Semir Zeki, *La visione dall'interno. Arte e cervello.*, traduzione italiana di Giovanna De Vivo, Paolo Pagli, Torino, Bollati Boringhieri, 2003 (edizione originale *Inner Vision. An Exploration of Art and the Brain*, 1999)

Stephen E. Palmer, *Vision Science – photons to phenomenology*, USA, MIT Press, 1999

Stephen W. Hawking, *Buchi Neri e Universi Neonati*, traduzione italiana di Libero Sosio, Milano, R.C.S. Rizzoli Libri, 1993 (edizione originale *Black holes and baby universes and other essays*, 1993)

Tomàs Maldonado, *Critica della ragione informatica*, Milano, Giangiacomo Feltrinelli, 2006

Tomàs Maldonado, *Reale e virtuale*, Milano, Giangiacomo Feltrinelli, 2005

Vincenzo Caglioti, Giuseppina Gini, *Robotica*, Bologna, Zanichelli, 2003

Wolfgang Koler, *La psicologia della Gestalt*, traduzione italiana di Giannantonio De Toni, Milano, Giangiacomo Feltrinelli, 1984 (edizione originale *Gestalt Psychology*, 1947)

Sito-grafia

Anatomia Occhio-Cervello

Associazione degli Insegnanti di Scienze Naturali – ANISN – pagina dedicata alla visione

<http://www.anisn.it/scuola/strumenti/visione/index.htm>

Atlante Anatomico del Sistema Nervoso Centrale

<http://www.biocfarm.unibo.it/aunsnc/Default.htm>

Applicazioni

Elenco di aziende raggruppate per categoria elaborato da David Lowe

<http://www.cs.ubc.ca/spider/lowe/vision.html>

Integrated Machine Vision
Elenco di aziende che producono sistemi visivi

<http://www.spt.fi/eutist/>
<http://www.vision1.com/library/commercial.php>

Autori

Alessandro Verri

<http://www.disi.unige.it/person/VerriA/>

Christiaan Huygens

http://it.wikipedia.org/wiki/Christiaan_Huygens

David Marr

<http://kybele.psych.cornell.edu/~edelman/marr/marr.html>

Fechner Gustav Theodor

<http://www.provincia.venezia.it/lartis/log/musica/fechner1.htm>

<http://www.itapi.eu/itapi-persona-fechner.html>

Isaac Newton

http://en.wikipedia.org/wiki/Isaac_Newton

James Clerk Maxwell

http://it.wikipedia.org/wiki/James_Clerk_Maxwell

Max Planck

http://it.wikipedia.org/wiki/Max_Planck

Oscar Bettelli

<http://www.starrylink.com/editrice/catalogosag.html>

Paul Feyerabend

http://it.wikipedia.org/wiki/Paul_Feyerabend

Richard Gregory

<http://www.richardgregory.org/index.htm>

Sir Isaac Newton (1643 – 1727)

Stephen E. Palmer

Thomas Hobbes

http://www.bbc.co.uk/history/historic_figures/newton_isaac.shtml

<http://socrates.berkeley.edu/~plab/>

http://it.wikipedia.org/wiki/Thomas_Hobbes

Economia

Analisi di borsa con tecniche di
intelligenza artificiale

<http://www.borsanalisi.com/intart.shtml>

Evoluzione

ANISN – pagina dedicata
all'evoluzione

<http://www.anisn.it/>

Antidarwinisti e creazionisti

<http://ulisse.sissa.it/Answer.jsp?questionCod=59441345>

Dubbi sulla teoria evolutiva

<http://mmmgroup.altervista.org/i-darwin.html>

Gli scritti di Darwin

<http://darwin-online.org.uk/>

L'intervento di Jérôme Lejeune

<http://www.disf.org/ScienziatiCredenti/Lejeune.asp>

<http://lgxserver.uniba.it/lei/rassegna/021006b.htm>

Futurologia

Intervista a Barton Kunstler

<http://www.mediamente.rai.it/HOME/BIBLIOTE/intervis/k/kunstler.htm>

L'intervento di Francesco d'Arcais

<http://diea.ing.unibo.it/cdm/articolo.php?id=1442vmero=5&anno=1968>

L'intervento di Stanley L. Jaki

http://www.arianaeditrice.it/articolo.php?id_articolo=6401

Metafuturologia

<http://www.intercom.publignet.it/ic11/metafut.htm>

Intelligenza Artificiale

Artificial intelligence definizione
Wikipedia

http://en.wikipedia.org/wiki/Artificial_intelligence

Associazione Italiana per

<http://aixia.deis->

l'Intelligenza Artificiale - AIIA

ce.unibo.it:8080/aixia/aixia.jsp

Cos'è la “Strong AI”, prospettiva di
un'intelligenza artificiale forte

<http://www.ptproject.ilstu.edu/STRONGAI.HTM>

Definizione d'intelligenza artificiale -

<http://it.wikipedia.org/wiki/Intelligenza>

Wikipedia Definizione di Intelligenza per la psicologia - Wikipedia	a_artificiale http://it.wikipedia.org/wiki/Intelligenza_a_artificiale
Introduzione alle reti neurali artificiali	http://it.wikipedia.org/wiki/Intelligenza_a_(psicologia) http://www.sci.unich.it/~aroli/dida/iasc/articoli/Gori_p.4-20.pdf#search="reti neurali"
Mediamente Learning - introduzione all'intelligenza artificiale	http://www.mediamente.rai.it/learning/corso_02/index.htm
Strong AI vs. Weak AI - Wikipedia, the free encyclopedia	http://en.wikipedia.org/wiki/Strong_ai
Un approfondimento curato da alcuni istituti scolastici	http://www.racine.ra.it/curba/set/privato/index.html#Intro

Introduzione

Computer vision - Wikipedia	http://en.wikipedia.org/wiki/Computer_vision
CVonline - Compendium of Computer Vision	http://it.wikipedia.org/wiki/Computer_vision
Il laboratorio dell'università del Southern California	http://homepages.inf.ed.ac.uk/rbf/CVonline/
Keith Price – una bibliografia on line tra le più aggiornate.	http://iris.usc.edu/USC-Computer-Vision.html
L'intervento di Oscar Bettelli	http://iris.usc.edu/Vision-Notes/bibliography/contents.html
The Computer Vision Homepage	http://newk.alma.unibo.it/oscar/ai22.htm http://www.cs.cmu.edu/~cil/vision.html

Laboratori di Ricerca

CNR - Ricerca sulla segmentazione delle immagini	http://www.cnr.it/commesse/Scheda_Commissa_Descrizione.html?co=1372
Laboratory for Intelligent Imaging and Neural Computing	http://newton.bme.columbia.edu/main/Template.htm?liinc_projects.htm
Università di Messina - laboratorio	http://visilab.unime.it/new/
Vision Technology - Microsoft	http://research.microsoft.com/vision/

Luce

L'Osservatorio astronomico di Padova- "Che cos'è la luce?"

<http://www.pd.astro.it/othersites/venere/ESO/g1.htm>

Luce - Wikipedia
Luce e Suono

<http://it.wikipedia.org/wiki/Luce>
http://www.geocities.com/codadilupo_2000/lucesuono.htm

Luce virtuale

<http://www.lucevirtuale.net/biografie/maxwell.html>

Object Recognition

Object Recognition - di David Marshall 1994-1997
Object Recognition di Kimberly Kirkpatrick

http://www.cs.cf.ac.uk/Dave/Vision_lecture/node50.html
<http://www.pigeon.psy.tufts.edu/avc/kirkpatrick/default.htm>

Operazioni sulle immagini

Introduzione all'elaborazione delle immagini di Nico Lanconelli

http://www.bo.infn.it/~lanconel/my_file/presentations/Dig_image_proc_2.htm

Operazioni tra pixel

<http://jada1.unime.it/visilab/AppuntiCV/operazioni.htm>

Psicologia

Coscienza

<http://home.att.net/~mstaique/Italian/consciob.htm>
<http://www.cicap.com/enciclop/at100078.htm>

Il Costruttivismo e le sue Radici di Ernst von Glasersfeld
Materialismo - Wikipedia

<http://www.oikos.org/voncostrutt.htm>

Unità psicofisica - Wikipedia

<http://it.wikipedia.org/wiki/Materialismo>
http://it.wikipedia.org/wiki/Unità_psicofisica

Società

Il sito ufficiale della Polizia di Stato

<http://www.poliziadistato.it/pds/chisiamo/territorio/reparti/scientifica/uacv.htm>

Sistema di visione per lettura automatica di targhe di autoveicoli

<http://www.pklab.net/kbase/targhe/>

Suono

Numero di Mach - Wikipedia

[http://it.wikipedia.org/wiki/Numero_d
i_Mach](http://it.wikipedia.org/wiki/Numero_d
i_Mach)

Velocità del suono - Wikipedia

[http://it.wikipedia.org/wiki/Velocità_
del_suono](http://it.wikipedia.org/wiki/Velocità_
del_suono)

Visione come Linguaggio

Intervista a Semir Zeki

[http://www.nuovoutile.it/index.php?
story=149&cat=3](http://www.nuovoutile.it/index.php?
story=149&cat=3)

Maria Antonia Ferrante - La Funzione
simbolica nell'uomo preistorico

[http://www.psychomedia.it/neuro-
amp/97-98-sem/ferrante.htm](http://www.psychomedia.it/neuro-
amp/97-98-sem/ferrante.htm)

Vita Artificiale

Artificial life - Wikipedia, the free
encyclopedia

[http://en.wikipedia.org/wiki/Artificial
_life](http://en.wikipedia.org/wiki/Artificial
_life)

Artificial life links

[http://www.alcyone.com/max/links/al
ife.html#Artificial_life](http://www.alcyone.com/max/links/al
ife.html#Artificial_life)

Robot - Wikipedia

<http://it.wikipedia.org/wiki/Robot>

Vita artificiale - Wikipedia

[http://it.wikipedia.org/wiki/Vita_artifi
ciale](http://it.wikipedia.org/wiki/Vita_artifi
ciale)

Filmografia

Chris Columbus, *L'uomo bicentenario*, USA, 1999 (edizione originale *Bicentennial man*)

Basato sull'omonimo racconto di Asimov, il film tratta la questione della differenza tra uomo e macchina

James Cameron, *Terminator*, USA, 1984 (edizione originale *The Terminator*)

Il primo film della serie, che rappresenta il possibile obbiettivo della computer vision e dell'intelligenza artificiale

James Cameron, *Terminator 2 – Il giorno del giudizio*, USA, 1991 (edizione originale *Terminator 2: judgement day*)

John Badham, *Corto circuito*, USA, 1986 (edizione originale *Short Circuit*)

Il protagonista del film è un robot che ha acquisito casualmente la vita.

Jon Amiel, *The core*, Usa/Gran Bretagna, 2003

Sono interessanti le scene che mostrano la realizzazione di sensori visivi

Jonathan Mostow, *Terminator 3 - Le macchine ribelli*, USA, 2003 (edizione originale *Terminator 3 - Rise of the machines*)

La scena relativa alla presa di autocoscienza delle macchine si riferisce chiaramente ad alcune teorie dell'intelligenza artificiale

Mic Rodgers, *Universal Soldier*, USA, 1999 (edizione originale *Universal Soldier: The Return*)

Tony Scott, *Nemico Pubblico*, USA, 1998 (edizione originale *Enemy Of The State*)

Steven Liesberger, *Tron*, USA, 1982

Gabriele Salvatores, *Nirvana*, 1997