ULISSE2000

Alitalia

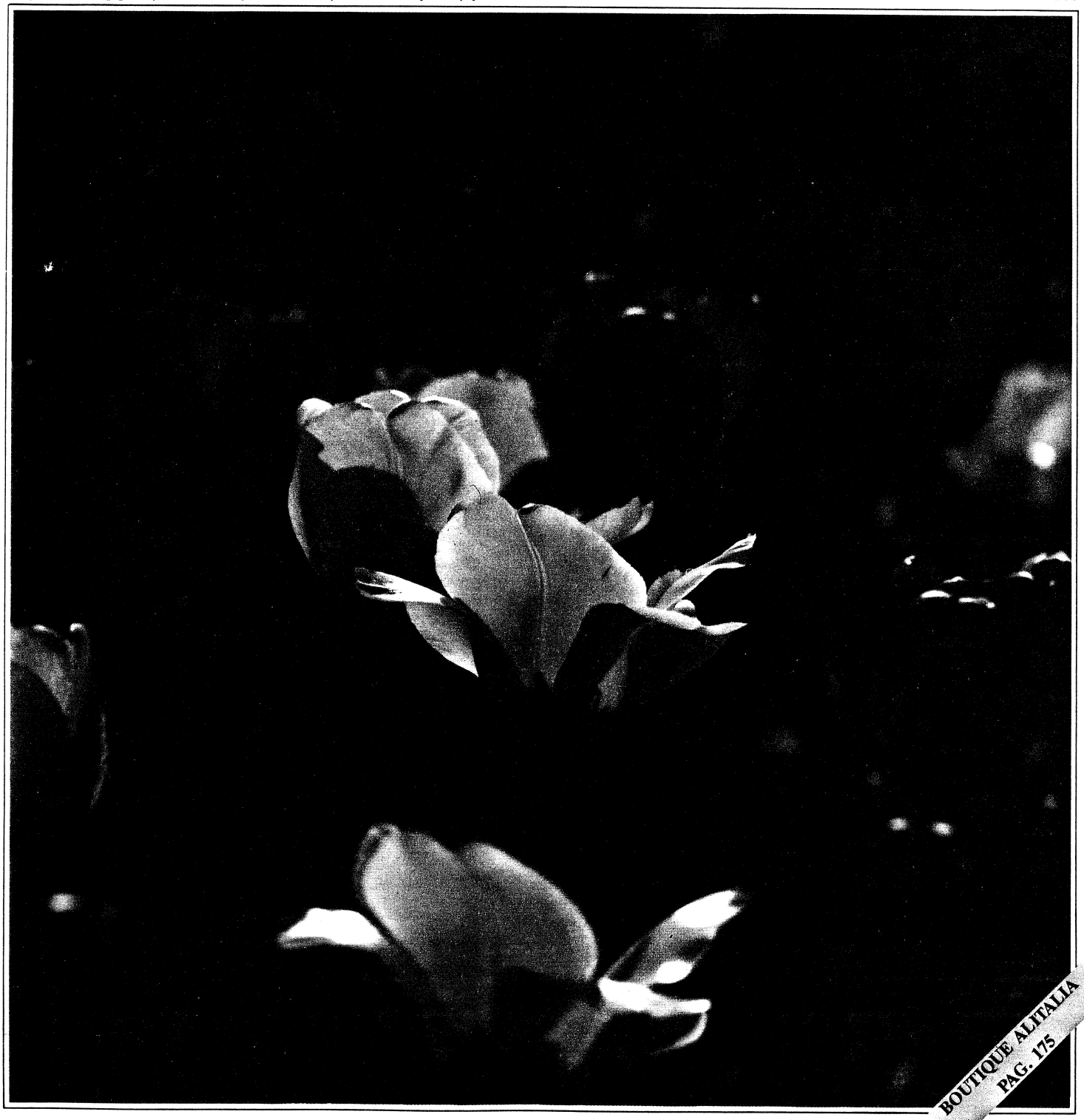
ON BOARD:

Giulietta Masina Tomaso Poggio Mike J. Murphy
Maurizio Beretta Omar Sharif Alfredo Pieroni
Gino Gullace Stephen King Roger Moore Shirley Temple

63

copia omaggio per voi your complimentary copy

GIUGNO 1989



BOUTIQUE ALITALIA
PAG. 175

SPEDIZ. IN ABB. POSTALE - GRUPPO III - 70% - BERGAMO

Oltre l'immagine

di Tomaso Poggio

Computer capaci di «vedere» imitano i meccanismi del cervello umano. Le ricerche sulla visione artificiale hanno fatto passi da gigante.

Gli esseri umani sono animali con un senso della vista particolarmente sviluppato. Circa il 40% della neocorteccia del cervello umano — la parte che nel corso dell'evoluzione si è sviluppata per ultima e che caratterizza dunque i primati — è consacrata alla visione. La nostra comprensione del mondo dipende a tal punto dalla visione, che in diverse lingue «vedere» significa non soltanto decodificare segnali luminosi, ma anche comprendere, intuire. La visione è intelligenza. Immaginate una persona che guarda la televisione: a partire dalla luce che guizza sullo schermo bidimensionale, lo spettatore ricrea un mondo tridimensionale fatto di persone, luoghi e oggetti. I processi mentali che, sulla base del modello luminoso che si forma sulla retina, portano a una rappresentazione interiore del mondo non sono meno intelligenti dell'analisi e interpretazione dei sintomi che un medico effettua per giungere alla diagnosi. Tuttavia, quando ci stupiamo delle facoltà del cervello umano, pensiamo piuttosto alle capacità deduttive del logico che non all'abilità dell'uomo comune di riconoscere un volto.

I progressi compiuti nel campo della visione artificiale nel corso degli ultimi vent'anni hanno messo in luce un paradosso che invita all'umiltà. La visione, infatti, si è rivelata non solo intelligente, ma anche più difficile da comprendere e riprodurre dei più sofisticati ragionamenti matematici. Attualmente esistono computer che possono imitare il lavoro di medici, avvocati e consulenti finanziari, ma nessun robot è in grado di rimpiazzare, ad esempio, cuochi e giardinieri. Se si considera la percezione in termini di risoluzione dei problemi connessi all'elaborazione di informazioni, la percezione risulta probabilmente più profonda e complessa del pensiero logico-razionale.

L'obiettivo delle ricerche sulla visione artificiale è di costruire macchine capaci di vedere e, al tempo stesso, di comprendere la visione umana. Dall'epoca della sua nascita, circa vent'anni fa, la visione artificiale ha fatto passi da gigante. A poco a poco

BEYOND THE IMAGE

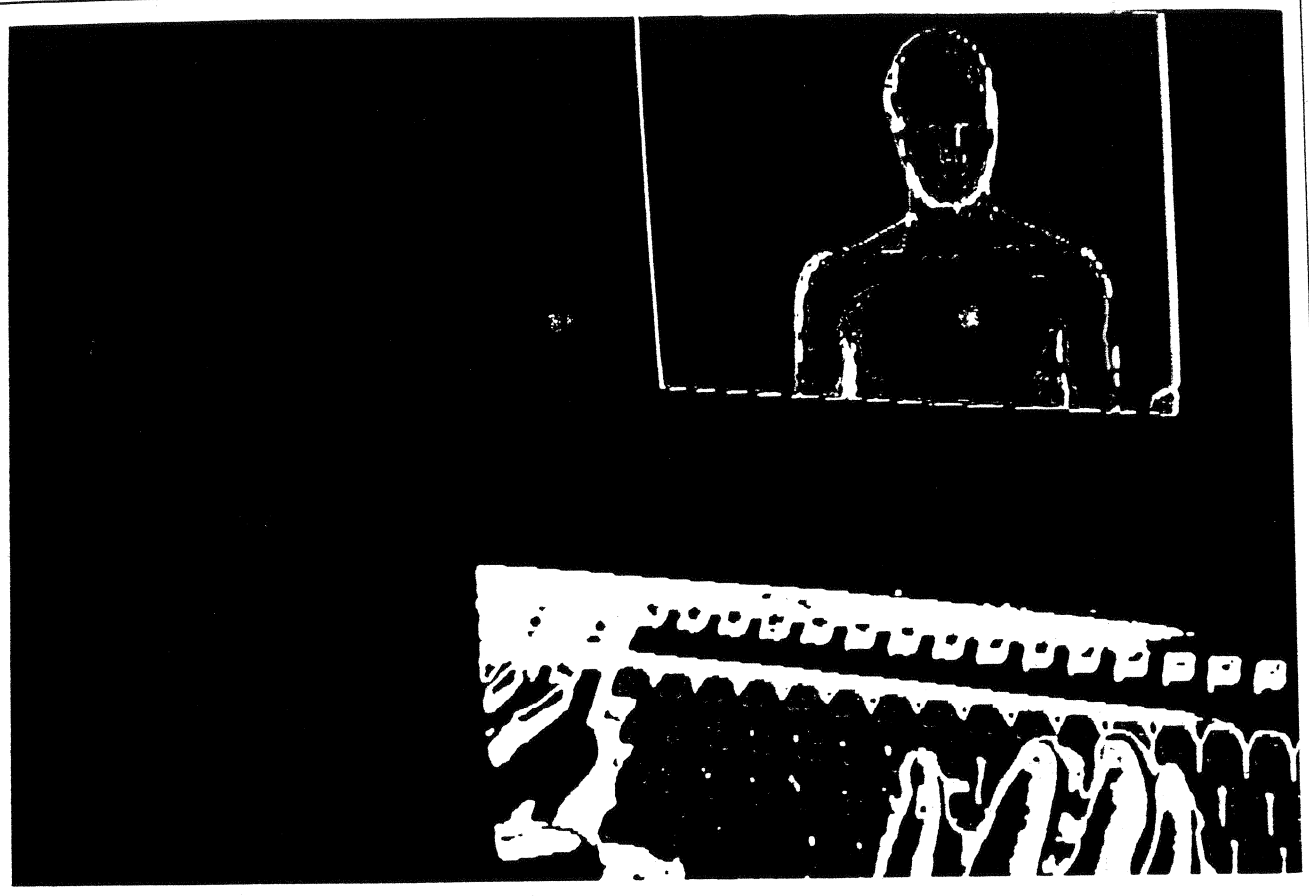
by Tomaso Poggio

Computers that can "read" imitate the mechanisms of the human brain: research on artificial vision has made incredible progress.

We humans are highly visual animals. Almost 40% of our neocortex — the part of the brain that arrived latest in evolution, and hence is most characteristic of primates — is dedicated to vision. Vision is so integral to our underestanding of the world that "to see" means not only to decode light signals, but also to comprehend the thrust of a verbal argument. Vision is intelligence. Imagine watching television; from the flickering light on a two-dimensional screen, you create a three-dimensional world of people, places, and things. The mental processes that lead from the pattern of light on your retina to an internal picture of the world are as intelligent as the analysis and interpretation that lead a doctor from symptoms to diagnosis. Yet when we marvel at the human brain, we are more apt to prize the deductive powers of the logician than the skill of the average person in recognizing a face.

Progress in machine vision over the past twenty years has revealed a humbling irony. Vision is not only intelligent, but also more difficult to understand and recreate than the most sophisticated mathematical reasoning. In fact, computers now exist that imitate the work of physicians, lawyers, and financial advisors, but today's robots could not possibly replace gardners or cooks. Looked at as information processing problems, perception is probably more deep and complex than logic and rational thought.

The goal of machine vision research is to build machines that can see and, at the same time, to understand human vision. Machine vision has made much progress since its birth about twenty years ago. It is becoming a solid science, based on sophisticated theories. It has important



L'obiettivo delle ricerche sulla visione artificiale è di costruire macchine capaci di vedere.

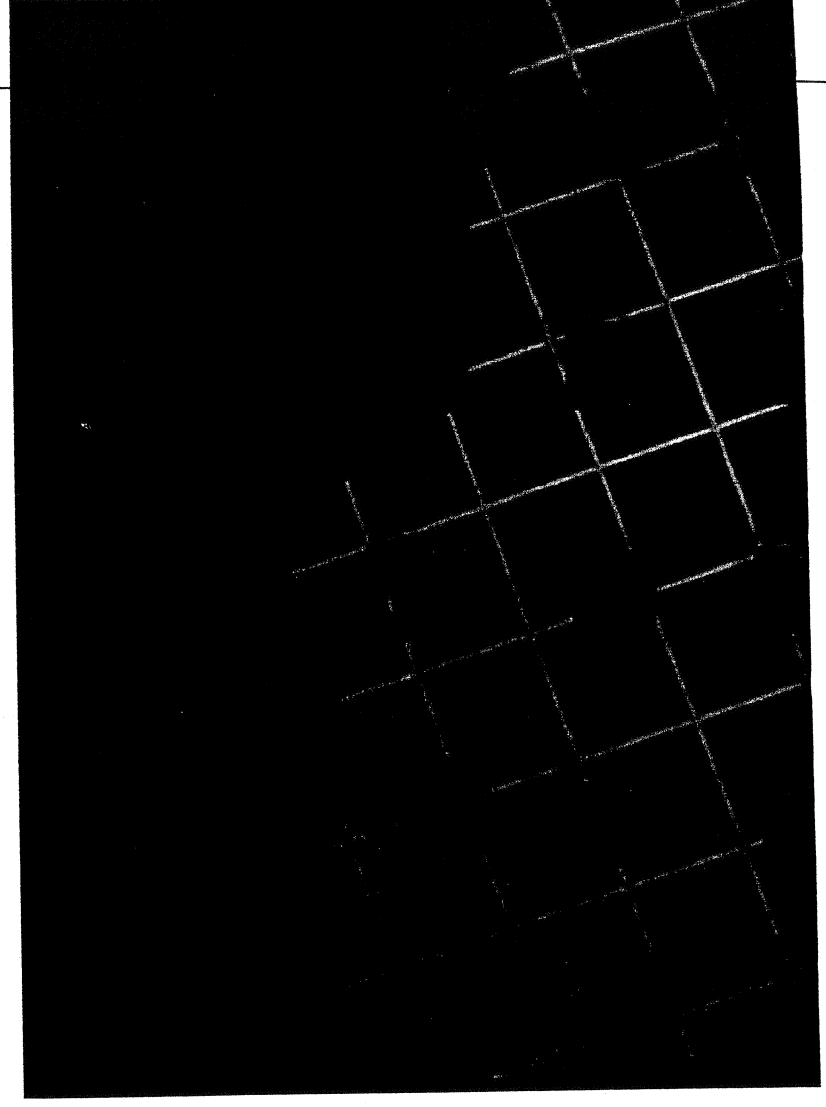
The aim of research into artificial vision is to produce machines that can see.

si è trasformata in una disciplina scientifica, basata su teorie sofisticate e ricca di potenziali applicazioni, ad esempio nel campo dei collaudi e della robotica industriale. Tuttavia, essa si trova ad affrontare un problema, o meglio un insieme di problemi, di ardua risoluzione. In cosa consiste la complessità della visione? Consideriamo l'immagine digitale con cui ha inizio il processo della visione. In una telecamera digitale, tale immagine consiste in una grande e ordinata disposizione di numeri che traduce la luminosità dei vari *pixel* (elementi di rappresentazione), così come è misurata dalla telecamera — che non è altro che una disposizione ordinata di cellule fotoelettriche. L'equivalente biologico delle cellule fotoelettriche sono i fotorecettori della retina, minuscole cellule che misurano la quantità di luce per ciascun punto dell'occhio e la traducono in un segnale elettrico che viene poi inviato alle altre cellule della retina, e da lì al cervello. È così che ha inizio il processo della visione; per ricavare informazioni sugli oggetti tridimensionali che hanno dato origine all'immagine, sulla suddetta disposizione di numeri o di potenziali vanno poi effettuate diverse operazioni complesse. Il risultato di questa multiforme elaborazione di informazioni è ciò che noi percepiamo, qualcosa, cioè, di assai diverso da ciò che gli occhi misurano: il frutto di operazioni complesse che hanno luogo nel cervello, al di fuori del campo della coscienza.

Una delle difficoltà presenti nell'elaborazione delle immagine necessaria alla visione è la quantità stessa di informazioni implicate nel processo. Un'im-

applications, for example in industrial inspection and robotics. But it is also a problem, or rather a set of problems, that is extremely difficult. Why is vision so hard? Consider the digital image with which the process of vision begins. It is a large array of numbers which represent the brightness of the various pixels (picture elements) as measured by a digital camera, which is nothing other than an array of photo cells. The photoreceptors in our retina are the biological equivalent of the photo cells. These are tiny cells that measure the amount of light at each position in the eye, and transduce it into an electrical signal that is then sent to other cells in the retina, and from there to the brain. This is the beginning of the process of vision; complex operations must be performed on this array of numbers or voltages in order to extract information about the three-dimensional objects that gave rise to the image. The result of this complex processing is what you perceive, and it is not what the eyes measure, but rather the result of complex operations that go on in the brain, hidden from conscious awareness.

One of the difficulties in the image processing required is simply the raw amount of information. A typical image in machine vision might be composed of one million pixels. Each pixel holds an eight-bit number. The total amount of information, although far less than a human retinal image (the retina has more than one hundred million photoreceptors), is still a staggering eight million bits. Multiply that number by the number of images per second that the camera must deliver in order to mimic the human eye, and



magine media, nella visione artificiale, può essere composta da un milione di *pixel*. Ciascun *pixel* è espresso da un numero ad otto bit. La quantità totale di informazione, sebbene di gran lunga inferiore a quella di un'immagine retinica umana (la retina ha più di cento milioni di fotorecettori), ammonta comunque a ben otto milioni di bit. Se si moltiplica questo numero per il numero di immagini al secondo che la telecamera deve fornire per imitare l'occhio umano, il tasso di trasmissione delle informazioni si alza enormemente, fino a raggiungere la cifra astronomica di diverse centinaia di bit al secondo. Questo significa che persino le più semplici operazioni matematiche effettuate dalla macchina sul flusso di immagini richiedono miliardi di moltiplicazioni e addizioni al secondo. Anche impiegando diverse migliaia di personal computer che operino congiuntamente, l'impresa risulterebbe pressoché irrealizzabile.

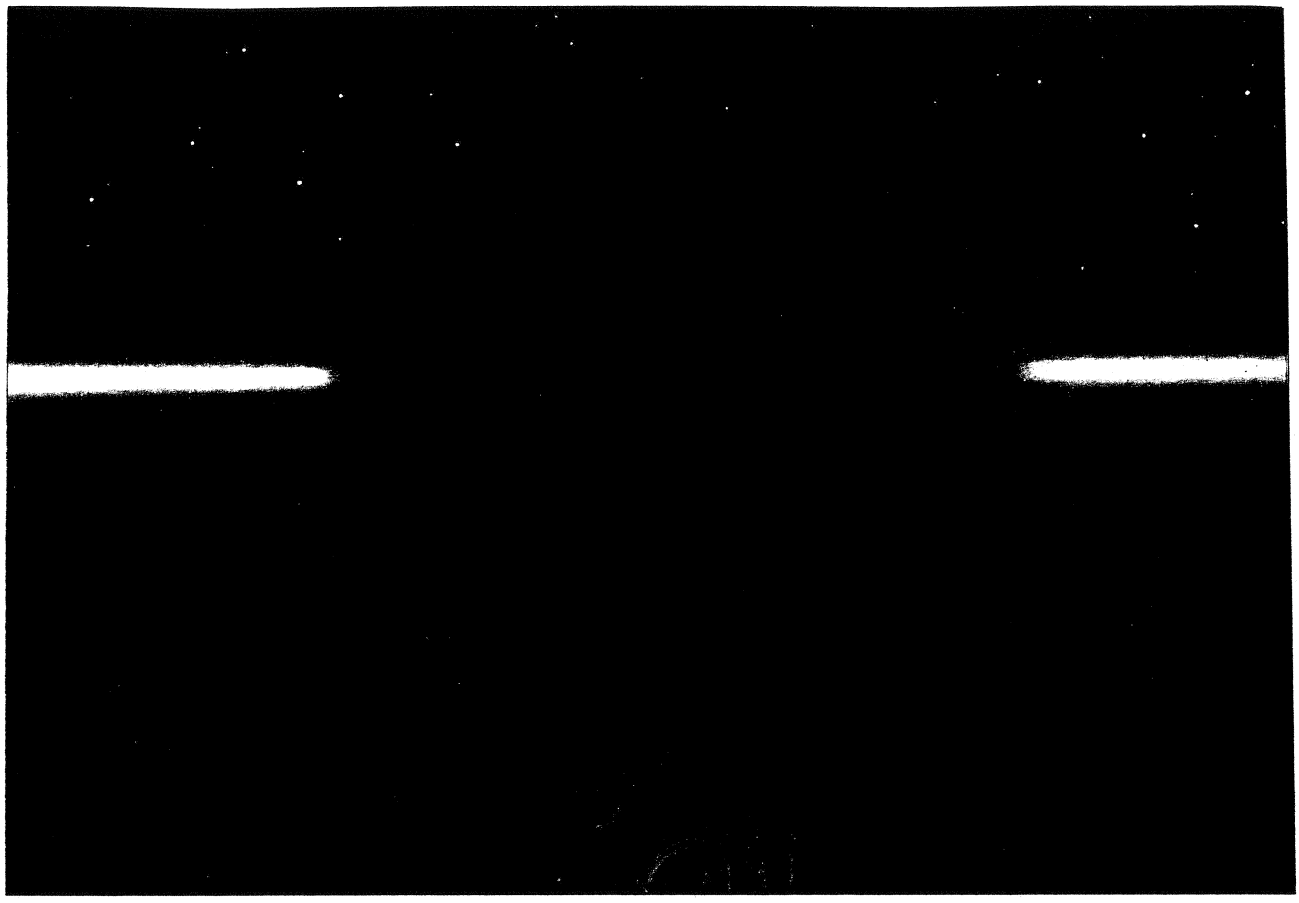
Paradossalmente, il vero problema della visione è che tutte le informazioni contenute in un'immagine non sono mai sufficienti. Un'immagine non è altro che la proiezione bidimensionale di una superficie tridimensionale. Nel processo di formazione dell'immagine, molte informazioni vanno perdute nella proiezione delle superfici su due dimensioni. Il problema di come computare le immagini, tenendo

«La visione è intelligenza - dice Tomaso Poggio - ed è più difficile da riprodurre dei più sofisticati ragionamenti matematici».

"Vision is intelligence," says Tomaso Poggio, "and is more difficult to reproduce than the most sofisticated mathematical reasoning."

the information transmission rate climbs to at least several hundred million bits per second. Thus, even the simplest mathematical operations that the machine performs on the flow of images requires billions of multiplications and additions per second. Many thousands of personal computers working together could barely do the job.

Ironically, the real problem in vision is that all the information in an image is never enough. An image is just the two-dimensional projection of a three-dimensional surface. In the imaging process, much information is lost as the surfaces are projected onto two dimensions. The problem of how to compute the images given the surfaces of the objects that originate them is already solved — this is the problem of classical optics and computer graphics; it is also the problem of the artist who attempts to reproduce on canvas a three-dimensional environment. The problem of vision, both biological and artificial, is the inverse problem: given a two-dimensional array of intensity values, find the three-dimensional arrangement of objects, surfaces, and surface properties that produced it. Extracting the features is hard because information is missing; this missing information must be replaced by hypotheses based on typical properties of the visual world such as the rigidity of objects, which are likely to be respected. As stated, the problem appears intractable. The features that must be recovered — the colors, textures, and spatial



presenti le superfici degli oggetti da cui esse hanno origine, è stato risolto: è lo stesso problema dell'ottica classica e dei computer grafici e, prima ancora, è il problema dell'artista che cerca di riprodurre sulla tela un ambiente tridimensionale. Il problema della visione, sia biologica che artificiale, è esattamente l'opposto: data una disposizione bidimensionale di valori di intensità luminosa, la visione deve risalire alla disposizione tridimensionale di oggetti, superfici e proprietà delle superfici che l'hanno generata. Estrapolare i tratti fondamentali dell'immagine non è facile, perché parte delle informazioni manca e deve essere sostituita da ipotesi basate sulle proprietà tipiche dell'universo visivo, quale ad esempio la rigidità degli oggetti, che con ogni probabilità sono valide in tutte le situazioni percettive. Formulato in questi termini, il problema appare ir risolvibile. I tratti che devono essere dedotti — colore, grana e rapporti spaziali tra gli oggetti — sono inestricabilmente mescolati al resto della matrice numerica. Buona parte degli sforzi compiuti nel campo della visione artificiale durante l'ultimo decennio tendeva a individuare un certo numero di proprietà generali dell'universo visivo che fosse possibile sfruttare per risolvere il problema della visione (per individuarne, cioè, i cosiddetti «limiti naturali»). I risultati ottenuti sono molto incoraggianti. La visione artificiale ha sviluppato diverse tecniche che, utilizzando queste ipotesi a priori, elaborano degli algoritmi in grado di computare con successo parecchie tracce

La visione artificiale è ancora ben lontana dall'eguagliare il sistema visivo umano.

Artificial vision has a long way to go to reach the perfection of the human eye.

relationships of objects, the position and color of the light source — are hopelessly entangled in the matrix of numbers. Much of the effort in machine vision in the past ten years has gone toward identifying *a priori* properties of the visual world that can be exploited to solve the inverse problem of vision (known as "natural constraints"). This has been remarkably successful. Machine vision has developed techniques to exploit these *a priori* assumptions in algorithms that successfully compute several independent elementary visual cues such as motion, color, texture, and depth. The tactics of computer vision can be characterized as three general activities. The first is to segregate visual tasks from one another, that is, to define exactly the distinct visual cues, such as color, texture, motion, depth, and shading that must be extracted from the image. The second is to identify the natural constraints that cover each task. The final step is to fit these natural constraints into an algorithm that works. This strategy has resulted in the development of isolated modules, each of which computes an important property from the image, such as motion, texture, and color. For example, to compute the distance of an object from the viewer (an important task for obstacle avoiding systems) the machine vision approach starts with two images. Each image is taken from a slightly different viewpoint, as if from two eyes separated by the bridge of a nose. Because of this, the pixels that correspond to an object in view will be slightly displaced in the left image compared to the right. The further away the object, the greater the displacement. Thus by measuring the

visive elementari indipendenti le une dalle altre, quali il movimento, il colore, la grana e la profondità.
C'è un'importante analogia fra i sistemi di visione artificiale e il cervello. I nostri neuroni sembrano, in linea di massima, suddividere i vari compiti in maniera simile. Un neurone che discrimina il colore, ad esempio, in generale trascura le informazioni concernenti la disparità delle immagini destra e sinistra. Nella corteccia esistono differenti aree preposte alla visione e ciascuna di esse è specializzata, almeno in parte, nell'elaborazione di tracce diverse. Ad esempio, Gian Poggio, un noto fisiologo della John Hopkins Medical School, ha scoperto nella corteccia visiva delle scimmie alcuni neuroni che sono coinvolti nella valutazione della profondità sulla base di immagini binoculari. Per quanto concerne la visione artificiale, sono stati costruiti computer che, a partire da algoritmi che estrapolano tracce visive quali il movimento, il colore, la grana e la profondità, sono in grado di trattare in maniera discretamente efficace le immagini naturali. Si tratta di un risultato eccellente, se confrontato con quanto era possibile fare solo cinque o dieci anni fa.
La visione artificiale, tuttavia, è ancora ben lontana dall'eguagliare le prestazioni del sistema visivo umano. In parte, la maggiore elasticità ed efficienza della visione umana — la sua capacità di funzionare adeguatamente in condizioni di luce e in ambienti estremamente diversificati — è dovuta al fatto che essa consiste in operazioni ben più complesse della semplice suddivisione della percezione visiva in varie attività separate. Sebbene nelle fasi iniziali il sistema visivo analizzi l'immagine in termini di tracce visive multiple, in maniera simile alla visione artificiale, nelle fasi successive tali tracce vengono sottoposte a un processo di sintesi. Per questo motivo, attualmente le ricerche sulla visione artificiale condotte presso l'Artificial Intelligence Laboratory del MIT si concentrano sul problema dell'integrazione delle varie tracce visive in un sistema unitario che riesca a vedere in tempo reale.
La Macchina per la Visione è un sistema multimodulare consistente di di-

Una delle difficoltà presenti nell'elaborazione dell'immagine è la quantità stessa di informazioni implicate nel processo.

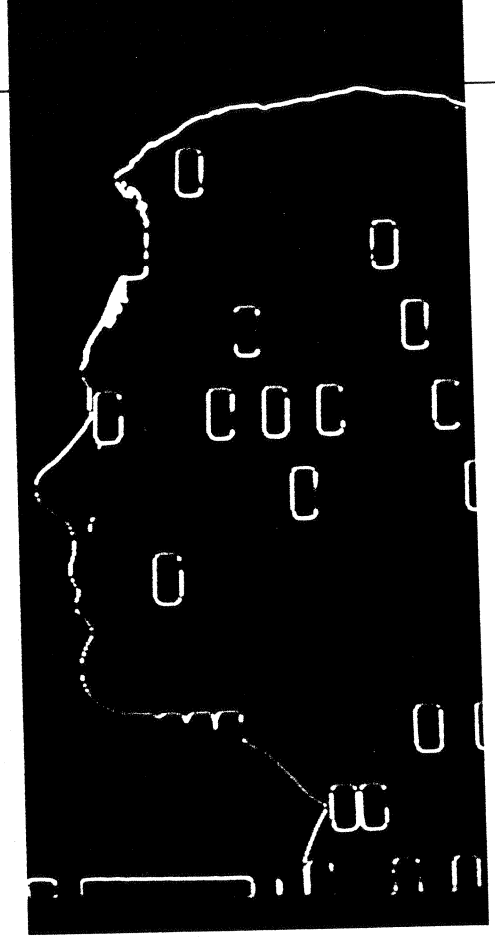
One of the problems in processing images is the amount of data involved.

displacement, one can calculate the depth of an object. But to do so, one first needs to match the regions in the left and right image that correspond to the same object. This matching problem is at the heart of most vision algorithms for computing depth. Natural constraints are called on to solve it. By assuming that most surfaces are smooth, that is, that they do not consist of many jagged leaps in depth, one can develop stereo algorithms that succesfully compute distance of surface from pairs of images as our brain does.
There is an important analogy between artificial vision systems and brains. Our own neurons seem, in an approximate way, to divide duties similarly. A color selective neuron, for example, generally cares little about disparity. There are different visual areas in the cortex, each one specialized, at least to some extent, for different cues. For instance, Gian Poggio, a well-known physiologist at Johns Hopkins Medical School, has found neurons in the visual cortex of the monkey that are involved in the computation of depth from stereo images. On the computer side of vision, algorithms that extract elementary visual cues such as motion, color, texture, and depth are today reasonably successful on natural images. This represents a great achievement compared to what was possible only five or ten years ago.
Computer vision, however, is still far from achieving the performance of the human visual system. Part of the reason for the superior flexibility and robustness of human vision — the fact that it works under so many different illumination conditions and in so many different environments — is because biological vision does more than simply divide vision into separate tasks. Although in the early stages the visual system analyzes the image in terms of multiple visual cues, as machine vision does, in later stages the cues are combined. For these reasons, one of the main preoccupations of current work in machine vision at the Artificial Intelligence Laboratory at MIT is the problem of integrating many different visual cues into a system that can see in real time.
The Vision Machine is a multi-modular system consisting of several modules which attempts to solve the general problem of vision from images to object recognition. The ultimate goal of the system is to

versi elementi che cerca di risolvere il problema generale della visione, dalla percezione delle immagini al riconoscimento degli oggetti. L'obiettivo ultimo del sistema è di riconoscere oggetti e persone semplicemente esplorando visivamente l'ambiente circostante. Nell'attuale configurazione della Macchina per la Visione, numerosi moduli diversi valutano tracce quali la stereopsi, il movimento, la grana, il colore e la luminosità dei contorni, ricavate dalla sequenza di immagini fornite da due telecamere. Tutti i moduli della Macchina per la Visione sono collegati a un supercomputer che opera parallelamente, detto Computer di Collegamento, contenente migliaia di semplici processori che lavorano simultaneamente cooperando tra di loro. La Macchina per la Visione è quindi un apparecchio che, in un certo senso, prende a modello la retina e il cervello (organi prototipici che operano in sinergia). Oltre al Computer di Collegamento, il sistema comprende un dispositivo di ricezione degli input, costituito da un occhio mobile che ruota e cattura immagini digitali con le sue due telecamere. Questo apparecchio possiede una mobilità paragonabile a quella dei nostri occhi e della nostra testa, che possono ruotare, muoversi verso l'alto o verso il basso, e guardare in differenti direzioni verso diversi punti dell'ambiente circostante.

Cosa abbiamo imparato a proposito della visione biologica lavorando sulla visione artificiale? Uno dei risultati della nostra ricerca è la nozione di moduli separati che elaborano diverse tracce visive contemporaneamente. Nel cervello, in maniera non dissimile dalla Macchina per la Visione, sembrano esserci aree distinte preposte al trattamento di diverse tracce visive: un'area specializzata nella valutazione del movimento, un'altra che estrapola le informazioni relative al colore, una terza che si occupa di misurare la profondità sulla base della disparità binoculare, e così via. Queste, tuttavia, sono semplificazioni grossolane dei dati, che dovrebbero essere prese con cautela, ma che sembrano comunque indicare una certa modularità del processo della visione (simulata, per l'appunto, dalla Macchina per la Visione), in parte confermata peraltro dai più recenti risultati degli studi sulla percezione umana. In conclusione, diversi campi di ricerca scientifica sembrano convergere verso un'identica descrizione dell'organizzazione complessiva della visione.

Tomaso POGGIO, docente presso l'Artificial Intelligence Laboratory del Massachusetts Institute of Technology (MIT) di Cambridge (USA). È autore di numerose pubblicazioni in diversi campi scientifici.



La Macchina per la Visione è un apparecchio che prende a modello la retina e il cervello.

The Vision Machine is a device modeled on the retina and the brain.

recognize objects and people just by looking around the room. In the present configuration of the Vision Machine, several different modules compute cues such as stereopsis, motion, texture, color, and brightness contours from the sequence of images provided by the two cameras. In the Vision Machine, the various cues are later integrated to obtain information about the attributes of three-dimensional objects such as their shape and the properties of their surface. One of the results of this processing is a kind of line drawing of the main contours in the scene. This line drawing is later used in the recognition stage. All the modules in the Vision Machine are implemented on a parallel supercomputer called the Connection Machine, with thousands of simple processors working simultaneously toward a solution in a cooperative manner. The Vision Machine is therefore a parallel device which in a sense uses the retina and the brain (prototypical parallel organs) as models. In addition to the Connection Machine, the system consists of an input device, an eye-head system that looks around and captures digital images with its two cameras. The eye-head system has several degrees of freedom, similar to the degrees of freedom of our eyes and our head: they can rotate up and down, and look in different directions toward different points in the environment.

What does this work on artificial vision tell us about biological vision? One of the results of our research is the notion of separate modules that process different visual cues at the same time. In the brain, as in the Vision Machine, there seem to be separate visual areas dedicated to different visual cues: an area specialized in computing motion, another specialized in extracting information about color, and still another specialized in extracting depth information from binocular disparity.

These are great oversimplifications of the data, and should be regarded with caution. They suggest, however, a certain modularization of the process of vision (which the Vision Machine emulates), similar to what the study of human perception suggests.

Thus, different areas of science seem to be converging on the same description of the overall organization of vision.

Tomaso POGGIO, professor at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology, Cambridge, is the author of numerous scientific papers.