

**Simone Arcagni**

*Computer Vision e Imaging come “visionica”.*  
*Verso la Visual Intelligence \**

Niente è natura, tutto artificio.  
(Lucrezio)

## 1. Visivo digitale. Premessa: i video

Ogni report e indicatore attesta che oggi, nel mondo dei dispositivi di comunicazione la testualità più prodotta, fruita e condivisa è quella visiva, e nello specifico l'audiovisivo. Quel mondo, cioè, di video di diversa natura, produzione, genere, estetica e finalità che circolano in rete sui diversi device in varie piattaforme e siti<sup>1</sup>, e che definisce un visivo in continua espansione, anzi, per essere ancora più specifici, in una espansione “esponenziale”<sup>2</sup>.

## 2. Hyper-visual

Nell'ultimo dei report annuali che rilascia IBM a proposito degli sviluppi tecnologici più interessanti e con maggiore capacità di impatto nella nostra società chiamato *5 in 5*, la componente visiva rappresenta lo snodo fondamentale, tanto che il titolo del documento è *The Invisible Made Visible*<sup>3</sup>. Si parla di sistemi di *Imaging* (MRI e EEG<sup>4</sup>) in

\* Estratto del saggio: S. Arcagni, “Computer Vision e Imaging come ‘visionica’. Verso la Visual Intelligence”, in A. Rabbito (a cura di), *La cultura visuale del XXI secolo. Cinema, teatro e new media*, Meltemi, Milano, di prossima pubblicazione.

<sup>1</sup> Per una mappatura di questo universo di forme, modi, pratiche e tecnologie rimando al mio *Visioni digitali. Web, video e nuove tecnologie*, Einaudi, Torino, 2016.

<sup>2</sup> Il rimando è alla natura “esponenziale” delle nuove tecnologie. In questa direzione si muovono, in particolare, gli studi della Singularity University sulla base della Legge di Moore sugli sviluppi esponenziali delle tecnologie digitali. Si veda la voce Treccani a proposito della legge di Moore: [http://www.treccani.it/enciclopedia/legge-di-moore\\_\(Enciclopedia-della-Scienza-e-della-Tecnica\)/](http://www.treccani.it/enciclopedia/legge-di-moore_(Enciclopedia-della-Scienza-e-della-Tecnica)/), e il volume del fondatore della Singularity University, Ray Kurzweil, *La Singolarità è vicina*, Apogeo, Milano, 2008.

<sup>3</sup> <http://research.ibm.com/5-in-5/>

<sup>4</sup> MRI sta per Magnetic Resonance Imaging. Si tratta di una tecnica di generazione di immagini usata in medicina a fini diagnostici e basata sul principio fisico della risonanza magnetica nucleare. L'EEG è invece l'elettroencefalografia che, tramite elettrodi, registra l'attività elettrica dell'encefalo.

grado di supportare la misurazioni di varianti nel linguaggio e nella scrittura dei pazienti (per esempio di Parkinson o Alzheimer) e capaci di creare complessi sistemi visivi a supporto di medici e pazienti. Viene citato l'Hyperimaging, quel ramo degli studi della Computer Vision che permette alla macchina di vedere e predire il visibile oltre lo spettro della luce combinando dispositivi di analisi di immagini a onde millimetriche, videocamere a varia visibilità e sensori. Mentre il cosiddetto Macroscopio è l'utilizzo di algoritmi e software di Machine Learning per analizzare, catalogare e combinare un vasto numero di informazioni riguardanti il mondo fisico e renderle alla portata della nostra vista e della nostra comprensione.

Ma basti pensare a quanto si appresta a realizzare Facebook nel 2018 creando, da una parte, una piattaforma mainstream per la Realtà Aumentata e costruendo, quindi, il mondo Facebook intorno alle immagini e ai video e formalizzando il rapporto tra uomo e ambiente e informazioni sulla base delle camere. Fatto che implica uno spostamento fondamentale dalla tastiera alla camera e che sta già avendo ripercussioni a livello sociale, culturale e cognitivo. E sempre Facebook ha appena lanciato – in fase Beta – *Spaces*, una “zona” di Facebook in cui le interazioni avvengono nei mondi della Realtà Virtuale accessibili agli avatar degli utenti registrati. Un universo di scambi tridimensionale, a 360° e immersivo.

Ma l'elenco degli investimenti in ricerca sui dispositivi visivi digitali e le loro applicazioni potrebbe essere lunga, basti accennare agli algoritmi che regolano le auto a guida senza conducente, ai famosi algoritmi di ricerca di Netflix, allo slancio della robotica di nuova generazione e all'impatto rivoluzionario che le tecniche di Medical Imaging stanno apportando in ambito medico, chirurgico e farmaceutico.

### 3. La questione dispositivo. Una questione di contesto

«Il dispositivo è ciò che regola il rapporto tra lo spettatore e le sue immagini *in un determinato contesto simbolico*<sup>5</sup>.»

Questo il pensiero di Jacques Aumont, e nel nostro caso specifico il *determinato contesto simbolico* è il contesto culturale, sociale ed economico determinato dalle nuove tecnologie informatiche. Il dispositivo visivo di oggi nasce a stretto contatto, non tanto delle tecnologie digitali, quanto della logica culturale che le tecnologie digitali hanno e stanno determinando in una infinita negoziazione all'interno della società. Una

<sup>5</sup> J. Aumont, *L'immagine*, Lindau, Torino, 2007, p. 199.

logica culturale talmente nuova da poter essere definita “rivoluzionaria”<sup>6</sup>, in quanto in grado di modificare i paradigmi e di sovvertire modelli consolidati nel tempo. E, infatti, Aumont continua: «Inoltre, lo studio del dispositivo è esso stesso necessariamente uno studio storico: non esiste dispositivo al di fuori della storia<sup>7</sup>».

#### 4. “Visionica”

Ciò implica che lo studio dell’immagine e del visivo contemporaneo non può prescindere dallo studio dello specifico dispositivo visivo digitale. Se ne era già accorto Paul Virilio che dopo aver studiato la “logistica della percezione” nel suo volume *Guerra e cinema*<sup>8</sup>, aveva dato seguito alle sue riflessioni sulla logica dei dispositivi visivi nel volume *La macchina che vede*. Siamo agli inizi degli anni Ottanta e Virilio prevede una sorta di materia di studi futuristica o futuribile:

Non si parla forse della prossima produzione di una «macchina che vede» capace non solo di riconoscere i contorni delle forme, ma anche di interpretare completamente il campo visivo, la messinscena, vicina o lontana, di un ambiente circostante complesso? Non si parla forse anche di una nuova disciplina tecnica, la “visionica”, ossia la possibilità di ottenere una *visione senza sguardo* in cui la videocamera dipenderebbe da un computer ed eserciterebbe per la macchina, invece che per un qualsiasi telespettatore, la capacità d’analisi dell’ambiente circostante, l’interpretazione automatica del senso degli avvenimenti, e ciò nel campo della produzione industriale, della gestione di stock, o anche in quello della robotica militare<sup>9</sup>?

La “visionica” non esiste, ma esiste una disciplina, tutta interna al mondo dell’informatica e della cibernetica, che in qualche modo si è assunta proprio questa finalità. Si tratta della Computer Vision (CV da ora in avanti). E prima di affrontare lo specifico di questa disciplina vorrei concentrare l’attenzione sul fatto che è lo studio della CV a permetterci di capire il dispositivo digitale che definisce il visivo contemporaneo. La CV è il luogo dove, non solo vengono realizzati algoritmi, software, hardware e dispositivi visivi, ma anche - e soprattutto - il luogo in cui viene definito il campo del visivo, della visività, esplorando potenzialità e orizzonti. Non è il luogo

<sup>6</sup> Cfr. L. Floridi, *The Fourth Revolution. How the Infosphere is Reshaping Human Reality*, Oxford University Press, Oxford, 2014.

<sup>7</sup> J. Aumont, op. cit., p. 199.

<sup>8</sup> P., *Guerra e cinema. Logistica della percezione*, Lindau, Torino, 2002.

<sup>9</sup> P. Virilio, *La macchina che vede. L’automazione della percezione*, Sugarco, Milano, 1989, p. 123.

dell'immagine, ma il luogo in cui si crea ciò che produce l'immagine e ciò che permette all'immagine di prendere e assumere forme diverse per scopi e finalità differenti. Risulta quindi fondamentale lo studio della CV per poter definire, nello specifico, l'oggetto "immagine contemporanea", fornendo così agli studi sul visivo un oggetto concreto, sia nelle sue componenti tecnologiche che, soprattutto, nelle sue potenzialità visive, nella sua natura visiva, in una parola nella sua logica culturale.

## 5. Tra mappe e filosofia

Per Eric Rodenbeck:

La visualizzazione delle informazioni sta diventando qualcosa di più che un insieme di strumenti, tecnologie e tecniche per comprendere il significato di insiemi enormi di dati. Sta emergendo come un medium a sé stante<sup>10</sup>.

Io non penso si possa parlare di medium, quanto di un processo visivo con una logica e una natura tecnologica specifica che deve essere sondata. E questa operazione deve essere fatta secondo una prassi ben delineata, per esempio, da Marshall McLuhan in una nota intervista realizzata da "Playboy":

A volte me lo chiedo. Sto esplorando. Non so dove ciò mi porterà. Il mio lavoro è rivolto allo scopo pratico di cercare di comprendere il nostro ambiente tecnologico e le sue conseguenze psichiche e sociali. Ma i miei libri fanno parte del processo, piuttosto che essere il prodotto finale della scoperta; il mio scopo è quello di utilizzare i fatti come strumenti d'indagine, come mezzi di comprensione e di ricerca di modelli, invece che di impiegarli nel senso sterile e tradizionale di classificazione di dati, di categorie e contenitori. Voglio creare mappe per nuovi territori, piuttosto che segnalare dove si trovano le vecchie postazioni<sup>11</sup>.

Oppure declinando l'affermazione di Deleuze e Guattari:

La scienza non ha nessun bisogno della filosofia per i suoi compiti. In compenso, quando un oggetto è scientificamente costruito con funzioni, per esempio uno spazio geometrico, resta da cercarne il concetto filosofico che non è assolutamente dato nella funzione<sup>12</sup>.

<sup>10</sup> E. Rodenbeck, citato da Salvatore Iaconesi, Oriana Persico, *La cura*, Codice, Torino, 2016, p. 99.

<sup>11</sup> Intervista a Marshall McLuhan, "Playboy", 1969, <https://sites.google.com/site/uominiculture/people/marshall-mcluhan/intervista-mcluhan-1969>

<sup>12</sup> G. Deleuze, F. Guattari, *Che cos'è la filosofia?*, Einaudi, Torino, 1996, p. 111.

Potremmo così dire che il concetto filosofico della CV è quello che abbiamo definito come la sua logica culturale.

## 6. I “nuovi territori” della CV. La macchina che vede

«Computer Vision is the science and technology of making *machine that see*<sup>13</sup>».

Ecco allora che ci imbattiamo ne “la macchina che vede” citata da Virilio, proprio negli studi di CV. Ma andiamo con ordine e proviamo a tracciare innanzitutto una sintetica storia di questo campo di ricerca<sup>14</sup>. Il primo personaggio chiave della CV è il famoso scienziato Marvin Minsky che fonda negli anni Sessanta il Laboratorio di Intelligenza Artificiale al MIT di Boston. Nel 1966 propone al suo studente Gerald Jay Sussman di passare l'estate provando a connettere una camera a un computer e chiedere alla macchina di descrivere ciò che “vede”. Minsky sa che il processo di apprendimento che prefigura la costruzione di una qualche forma di intelligenza, a partire dall'immagazzinamento di dati e informazioni, fino alla loro trasformazione e elaborazione, avviene nell'uomo fondamentalmente attraverso la vista. L'intuizione è quella di dotare il computer di questa facoltà. Ma il compito si preannuncia come piuttosto complesso e per affrontarlo invita al suo laboratorio lo studioso David Marr, scienziato con alle spalle studi in matematica, psicologia e neuroscienze. Definito il fondatore della neuroscienza computazionale, studia la visione come un processo in grado di definire un'immagine tridimensionale a partire da una bidimensionale. La rappresentazione tridimensionale diviene quindi la base da cui una macchina può apprendere, e quindi provare a prendere decisioni e realizzare azioni autonome.

La CV può così essere definita il campo di ricerca che studia come abilitare i computer alla comprensione e alla interpretazione delle informazioni visuali presenti in immagini o video.

Nel 1982 (quindi quasi in contemporanea con la “visionica” di Virilio) esce, postumo, *Vision*<sup>15</sup> di David Marr che definisce i caratteri della CV. Con il volume di Marr, non solo nasce la disciplina “Computer Vision”, ma se ne delinea il tratto essenziale: costruire dispositivi computazionali in grado di vedere ma anche, a partire da questo senso acquisito, di

<sup>13</sup> R. Cipolla, S. Battiato, G. M. Farinella (eds. by), *Computer Vision. Detection, Recognition and Reconstruction*, Springer, Berlin Heidelberg, 2010, p. VII (corsivo mio).

<sup>14</sup> Una sinetica quanto esaustiva storia (e bibliografia) della CV si può trovare in id.

<sup>15</sup> D. Marr, *Vision. A Computational Investigation into the Human Representation and Processing of Visual Information*, The MIT Press, Cambridge (Mass.), 2010.

comprendere e ragionare. La CV si caratterizza per l'approccio multidisciplinare accogliendo gli studi delle neuroscienze, del cognitivismo, quelli psicologici, quelli fisiologici e persino le teorie sul visivo in ambito filosofico come, per esempio, la Fenomenologia.

Il libro di David Marr, e i successivi saggi del suo più stretto collaboratore Tomaso Poggio, hanno il merito di fondare una disciplina che ha come fine quello di permettere alla macchina di vedere. L'idea iniziale è semplicemente quella di studiare il fenomeno della vista sotto ogni aspetto per poi replicarlo artificialmente. Quella di Marr è fondamentalmente uno studio che si basa sulla analisi dell'occhio e della funzione visiva e sulla replicabilità computazionale della visione. Marr analizza e replica, attraverso algoritmi, la percezione, lo sguardo, i soggetti, le forme, i contorni, i colori.

Se l'*Imaging* si occupa di processare le immagini digitali, la CV entra direttamente nella funzione visiva, nella produzione delle immagini. I due ambiti sono evidentemente separati anche se vengono affrontati all'interno dello stesso campo di ricerca, anche perché sempre più si percepisce l'esigenza di processare immagini, manipolarle, costruirle e definirle nel processo anche di creazione. Gli ambiti sono sì definiti nell'applicativo ma sempre più sfumati invece nelle pratiche di ricerca.

All'interno di una disciplina che, almeno inizialmente, mira all'imitazione dell'occhio umano, Marr, con un'intuizione tanto semplice quanto rivoluzionaria, rende esplicito un elemento fondamentale: la visione computazionale non è la visione umana. Per quanto possa sembrare banale, questa affermazione, ha il merito di gettare le basi per una futura, sempre più totale, indipendenza della CV dalla mera mimesi dell'apparato visivo umano.

## Bibliografia

- S. Arcagni, *Visioni digitali. Web, video e nuove tecnologie*, Einaudi, Torino, 2016.  
 R. Kurzweil, *La Singolarità è vicina*, Apogeo, Milano, 2008.  
 J. Aumont, *L'immagine*, Lindau, Torino, 2007.  
 L. Floridi, *The Fourth Revolution. How the Infosphere is Reshaping Human Reality*, Oxford University Press, Oxford, 2014.  
 P. Virilio, *Guerra e cinema. Logistica della percezione*, Lindau, Torino, 2002.  
 P. Virilio, *La macchina che vede. L'automazione della percezione*, Sugarco, Milano, 1989.  
 S. Iaconesi, O. Persico, *La cura*, Codice, Torino, 2016.  
 Intervista a Marshall McLuhan, "Playboy", 1969.  
 G. Deleuze, F. Guattari, *Che cos'è la filosofia?*, Einaudi, Torino, 1996.  
 R. Cipolla, S. Battiato, G. M. Farinella (eds. by), *Computer Vision. Detection, Recognition and Reconstruction*, Springer, Berlin Heidelberg, 2010.  
 D. Marr, *Vision. A Computational Investigation into the Human Representation and Processing of Visual Information*, The MIT Press, Cambridge (Mass.), 2010.

## **Computer Vision e Imaging come “visionica”. Verso la Visual Intelligence**

Il pervasivo e ubiquo visivo contemporaneo è prodotto da un particolare occhio che è quello della macchina computazionale. A questa “macchina che guarda” già agli inizi degli anni '80 Paul Virilio assegnava funzioni e caratteri talmente specifici da determinare un cambiamento di paradigma. Ora che i laboratori di *computer vision* – la disciplina nata all'interno dell'informatica per abilitare la macchina a vedere – sono diventati il fondamento del visivo contemporaneo abbiamo l'obbligo di studiare lo specifico tecnologico e gli impatti che esso produce.

PAROLE CHIAVE: computer vision, artificial intelligence, imaging, Virilio, visivo.

## **Computer Vision and Imaging as “visionic”. Towards the Visual Intelligence**

The contemporary pervasive and ubiquitous visual is a product by a particular eye. The Eye of the computational machine. At this “machine that looks” in the very beginning of the 80s Paul Virilio assigned functions and characters so specific to determine a paradigm shift. “Computer vision” is the discipline born within the Informatics to enable the machine to see. And because it represents the foundation of contemporary visual, we have the obligation to study the technological specific and the impacts it produces.

KEYWORDS: computer vision, artificial intelligence, imaging, Virilio, visual.