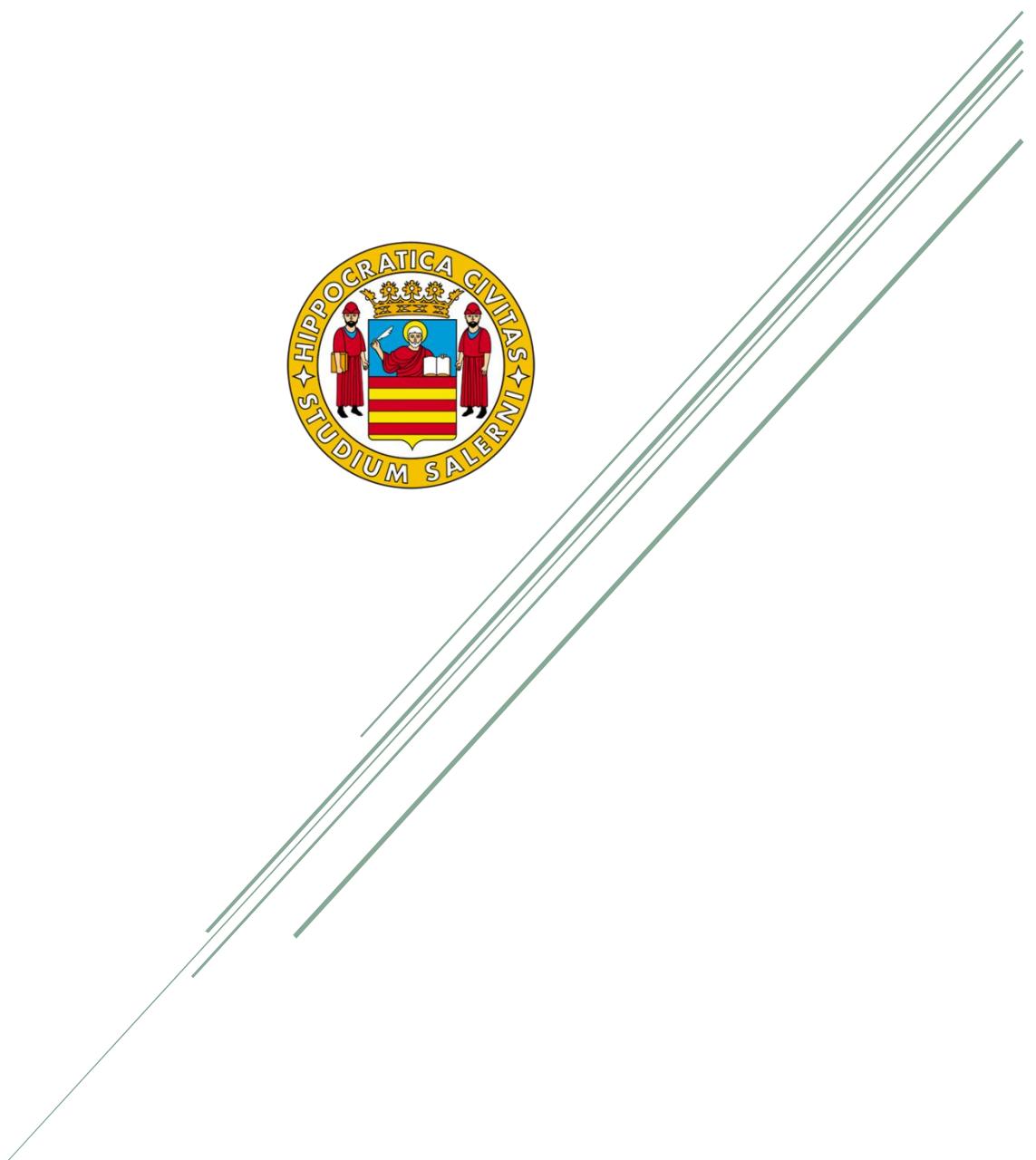


APPUNTI FONDAMENTI DI VISIONE ARTIFICIALE E BIOMETRIA

Corso di Laurea Magistrale – Sicurezza Informatica



Sommario

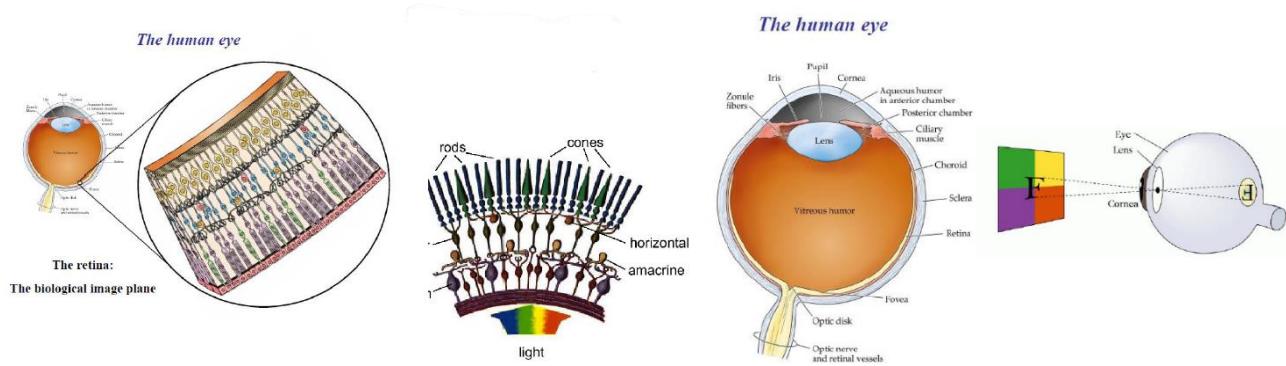
HVS (Human Visual System) and GESTALT Theory: Premessa al face Recognition	1
Definizioni di base Campionamento, Istogramma	11
Introduzione alla biometria	23
Image Filtering	37
Trasformata di Fourier	62
Content Based Image Retrieval (CBIR).....	80
Mercato della Biometria: Panoramica sulle Principali Biometriche	89
Modelli di Colore	102
Face Detection	109
Face Recognition.....	116
L'iride	124
Multibiometric Systems	130

HVS (Human Visual System) and GESTALT Theory: Premessa al face Recognition

L'occhio è una sfera avente un diametro di 20 millimetri e possiede diverse membrane concentriche quali:

- **Cornea** (trasparente).
- **Sclera** (opaca).
- **Coroide** (rete di vasi sanguigni)
 - Corpo Ciliare (agisce sul cristallino).
 - Iride (diaframma con un foro: pupilla che va dai 2 a 8 millimetri).
- **Retina**
 - *Fovea*:
 - Coni (circa 6 milioni) sensibili ai colori: Photopic Vision.
 - Bastoncelli (75/150 milioni): Scotopic Vision.

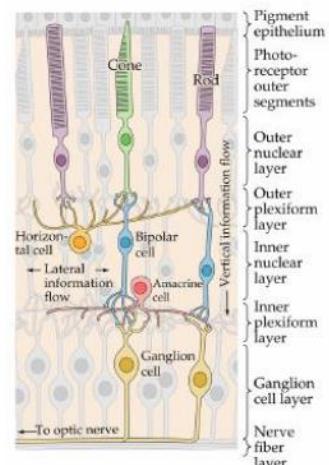
Ed infine il **cristallino**, composto per il 60/70% da acqua, 6% da grasso e proteine. Esso assorbe l'8% dello spettro visibile (infrarossi e ultravioletti).



La **retina** risulta essere una varietà di celle:

- Fotorecettori: circa 100 milioni (95% bastoncelli e 5% coni).
- Cellule Bipolari
- Cellule Gangliari
- Cellule Amacrine
- Cellule orizzontali

Nervo ottico, fatto di assoni di neuroni secondari (Gangliari). 1 milione di assoni.



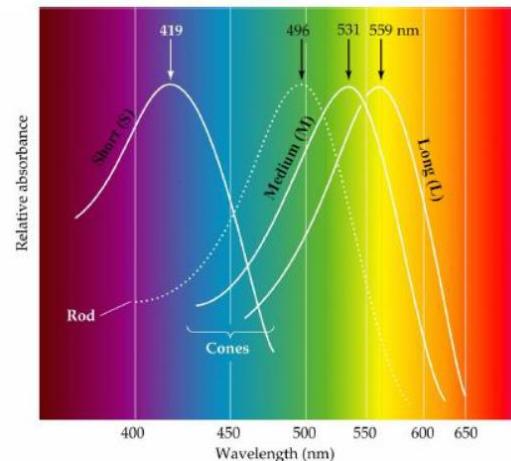
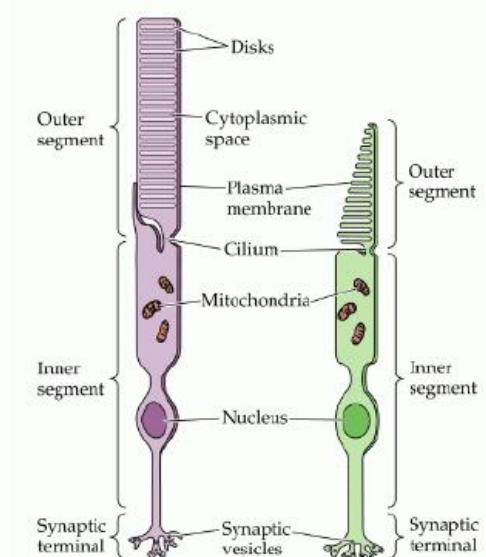
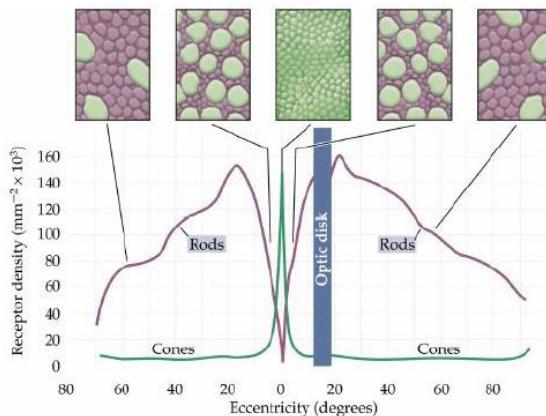
I bastoncelli sono:

- Estremamente sensibili alla luce,
- Emettono una risposta a singolo fotone,
- Bassa risoluzione spaziale,
- Profilo a risposta singola,
- B&W visione notturna (scotopica).

I coni:

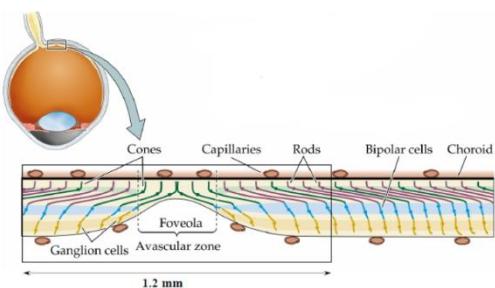
- Relativamente insensibile alla luce,
- 100 fotoni per una risposta paragonabile ai bastoncelli,
- Alta risoluzione spaziale,
- Profilo di risposta diverso,
- Visione a colori alla luce del giorno (fotopica).

Di seguito la distribuzione attraverso la retina di coni e bastoncelli.



La fovea è:

- Densità estremamente elevata dei recettori dei coni.
- Pochi bastoncelli. Nessuno nella foveola.
- Strati neurali spostati.
- Nessun vaso sanguigno.
- Relazione 1:1 con cellule bipolari e gangliari.



The Gestalt Theory of Visual Perception

I personaggi principali della psicologia della Gestalt erano Max Wertheimer, Kurt Koffka e Wolfgang Kohlers. Sebbene in seguito gli psicologi siano stati fortemente influenzati dalla teoria della Gestalt, specialmente nella psicologia sociale (ad esempio, Solomon Asch), non vi è mai stata alcuna seconda generazione di teorici della Gestalt formati dagli originali.

L'origine della parola è tedesca che significa forma o forma. Sviluppato dallo psicologo tedesco Max Wertheimer nel 1910. Dopo alcuni esperimenti con uno stroboscopio (più comunemente noto come libro a fogli mobili), concluse che l'occhio prende semplicemente tutti gli stimoli visivi e che il cervello organizza le sensazioni in un'immagine coerente.

Gli psicologi della Gestalt hanno ulteriormente perfezionato il lavoro di Wertheimer per concludere che la percezione visiva era il risultato dell'organizzazione di elementi o forme sensuali in vari gruppi. Gli elementi discreti all'interno di una scena sono combinati e compresi dal cervello attraverso una serie di quattro leggi di raggruppamento.

"Il tutto è diverso dalla somma delle sue parti."

L'idea di base della psicologia della Gestalt è che le caratteristiche sensoriali (linee, bordi, angoli, colori, ecc.) sono combinate dal cervello per formare un nuovo modello o configurazione, per produrre qualcosa che non esiste nell'insieme delle caratteristiche prese uno per uno.

L'obiettivo principale di questa idea era la psicologia atomistica di Wundt e Von Helmholtz.

Sensazione e Percezione

Sensazione: Il processo attraverso il quale i sensi raccolgono stimoli visivi, uditivi e altri stimoli e li trasmettono al cervello; informazioni sensoriali registrate nel cervello ma che non sono state interpretate.

Percezione: Il processo mediante il quale le informazioni sensoriali sono organizzate e interpretate attivamente dal cervello.

Se si fa un elenco di tutte le singole sensazioni che si provano quando la figura la si vede come un vaso, e si fa anche un elenco delle sensazioni quando lo si vede come due facce, le due liste sarebbero identiche. Ma i concetti sono diversi, quindi nelle liste deve mancare ciò che li rende diversi.



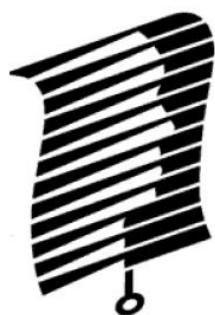
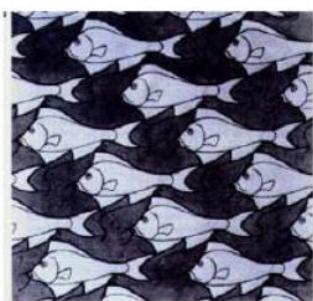
I principi della psicologia della Gestalt

- **Figure-Ground:** l'organizzazione dipende da ciò che vediamo come figura (oggetto) e da ciò che percepiamo di fondo (contesto).
- **(Legge di) prossimità:** gli oggetti si avvicinano nello spazio o nel tempo percepiti come appartenenti ad un solo insieme.
- **(Legge della) somiglianza:** gli oggetti che hanno caratteristiche simili sono percepiti come unità.
- **(Legge della) continuità:** tendiamo a percepire figure o oggetti come appartenenti ad un insieme se sembrano formare un modello continuo.
- **(Legge della) chiusura:** rileviamo che le figure con delle lacune siano complete.

Figure-Ground

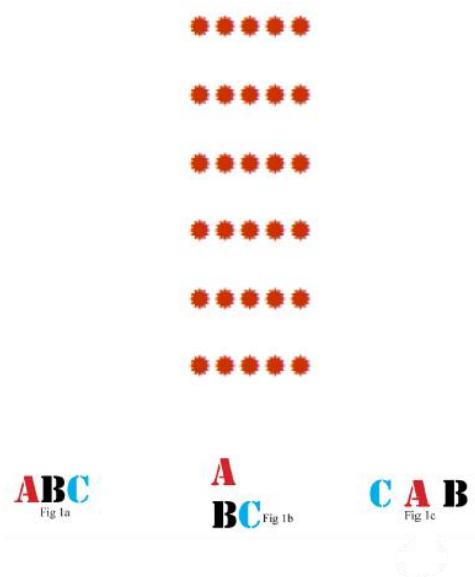
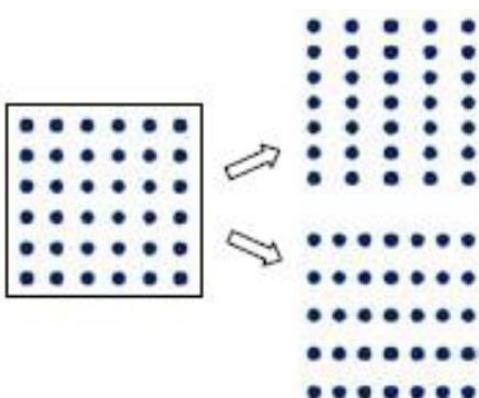
Nel 1915, lo psicologo danese della Gestalt, Edgar Rubin sperimentò modelli reversibili di figure (noti nella pittura e nella fotografia come spazio positivo e negativo). Ha progettato questa immagine per mostrare che l'oggetto poteva essere interpretato come due facce o un vaso. Tuttavia, il cervello non può vedere entrambe le immagini contemporaneamente: si deve prendere una decisione consapevole se vedere, nel disegno, una faccia o un vaso.

Quando un'immagine contiene due o più regioni distinte, si dividerà l'immagine in *figura* e *sfondo*.



Gestalt: Legge di Prossimità

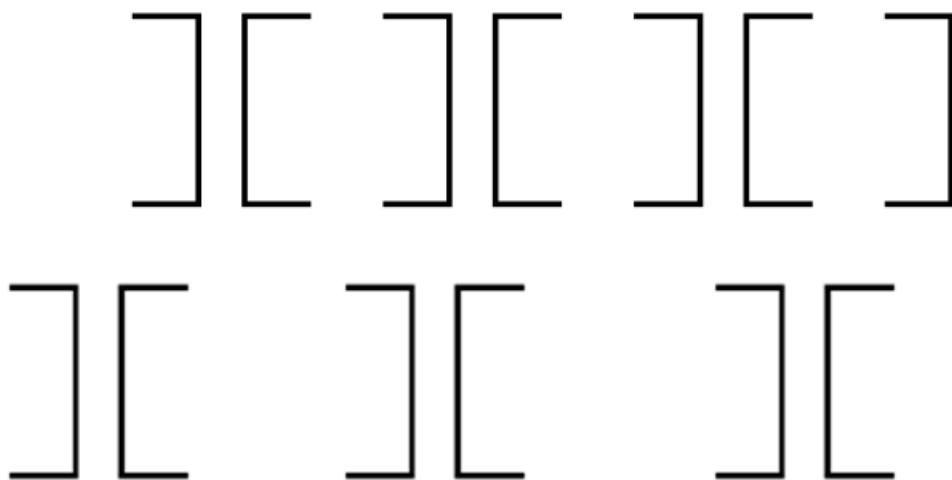
(La regola di prossimità) Questa regola si applica al modo in cui gli elementi di un modulo sono raggruppati in base alla loro prossimità o vicinanza. Questi singoli pezzi possono essere diversi o uguali.



Il cervello associa più facilmente oggetti vicini gli uni agli altri rispetto a quelli più distanti. Due amici in piedi uno accanto all'altro saranno considerati più strettamente correlati di qualcun altro a venti metri di distanza.



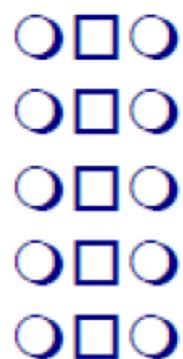
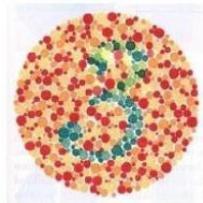
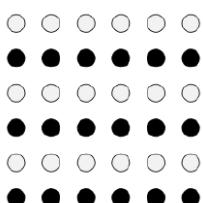
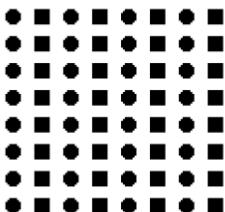
Gli elementi raggruppati vicini saranno percepiti come appartenenti allo stesso gruppo.



Gestalt: Legge di Similarità

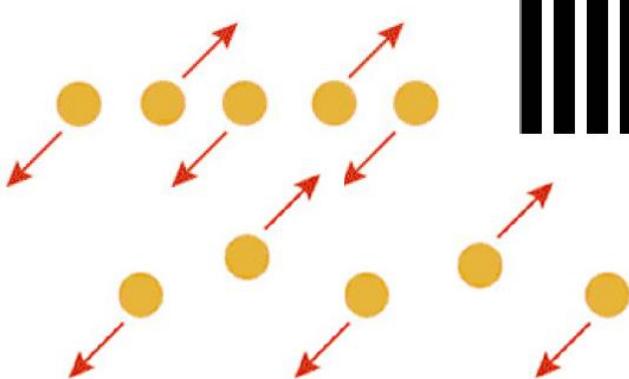
Similarità: vediamo questa forma come colonne anziché come righe a causa della somiglianza.

Gli elementi che condividono le qualità saranno percepiti come parte della stessa forma. Tali elementi vengono uniti con tanta maggior coesione quanto maggiore è la loro somiglianza.



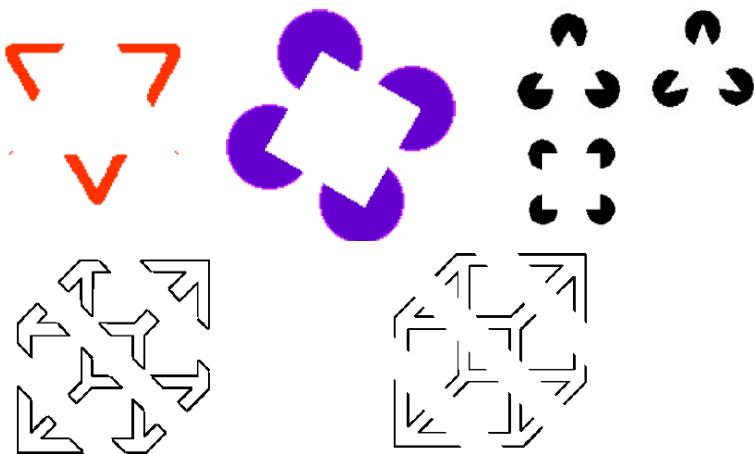
Gestalt: Legge di Continuità (o destino comune)

Continuità: ciò che vediamo è determinato dal flusso di linee. La legge del destino comune afferma che quando gli oggetti si muovono nella stessa direzione, tendiamo a vederli come un'unità.



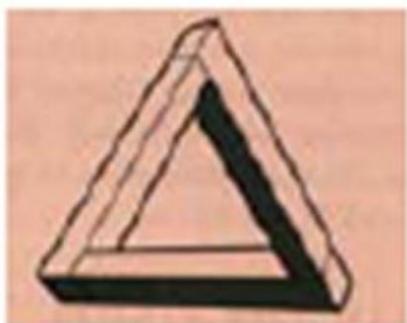
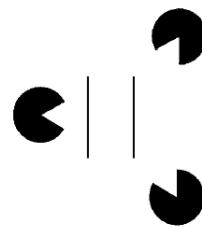
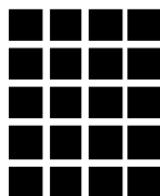
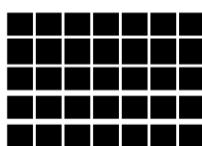
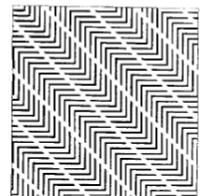
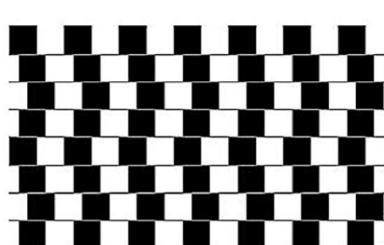
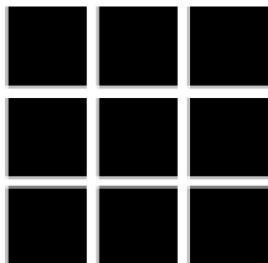
Gestalt: Legge di Continuazione (Chiusura)

Il cervello non preferisce cambiamenti improvvisi o insoliti nel movimento di una linea. La linea può essere una linea continua nel senso tradizionale o può essere una serie di oggetti messi insieme per formare una linea. Gli oggetti non in quella linea saranno separati mentalmente.



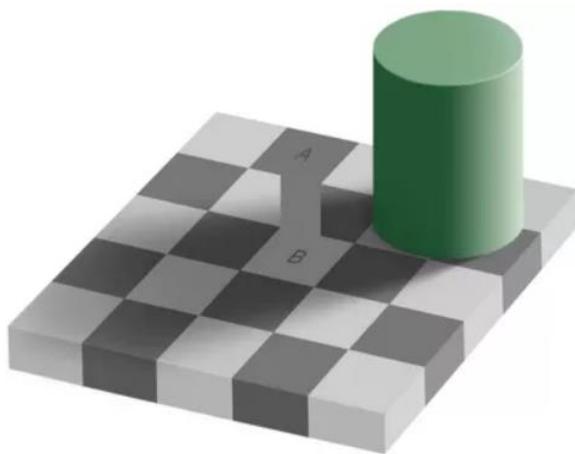
Illusione

Un'illusione è una distorsione di una percezione sensoriale o cognitiva, causata dal modo in cui il cervello è solito organizzare ed interpretare le informazioni che riceve. Le illusioni possono coinvolgere tutti i sensi, ma quelle ottiche sono le più emblematiche e conosciute, dal momento che la vista spesso prevarica gli altri sensi.



L'illusione di Adelson

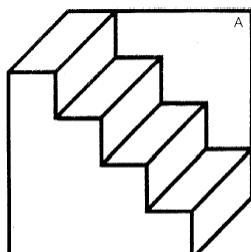
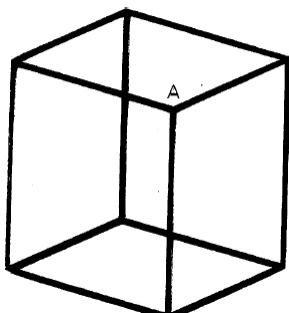
Adelson ha costruito una scacchiera dove i due quadrati segnati con le lettere A e B, per quanto possa risultare strano, hanno esattamente la stessa tonalità di grigio. La prova si può ottenere copiando e incollando in un qualunque programma di editing di immagini i due quadrati. Il nostro sistema visivo cerca di determinare il grado di grigio dei quadrati della scacchiera. Una superficie bianca in ombra riflette meno luce di una superficie nera in piena luce. Per decidere dunque dove sono le ombre e come compensare la loro presenza per determinare la tonalità di grigio della superficie, il nostro sistema visivo deve ricorrere a qualche stratagemma. Il primo è basato sul contrasto locale. Sia in condizione di luce che di ombra un quadrato più luminoso dei quadrati vicini è probabilmente più luminoso della media e viceversa. Nella nostra illusione il quadrato chiaro in ombra (B) è



circondato da quadrati neri. Questo fa sì che, nonostante il quadrato sia fisicamente scuro, esso appaia più luminoso se confrontato con quelli vicini. Al contrario i quadrati neri, esterni all'ombra proiettata dal cilindro verde, (come il quadrato A) sono circondati da quadrati più luminosi, che appaiono nel confronto relativo, come più scuri. Un secondo stratagemma del nostro sistema visivo è basato sul concetto che le ombreggiature spesso presentano dei contorni morbidi mentre per le superfici con un motivo i contorni sono netti. L'apparato visivo tende ad ignorare i cambiamenti graduali di intensità della luce al fine di determinare il colore di una superficie senza essere fuorviato dalle ombre. Nell'esempio specifico l'ombra disegnata sulla scacchiera è calibrata in modo da scurire i quadrati chiari rendendoli uguali a quelli scuri. Il cervello interpreta il contesto dell'immagine e si convince che si tratti di un'ombra, in quanto presenta contorni morbidi e perché risulta visibile l'oggetto che proietta l'ombra. La luminosità dei due quadrati viene inoltre influenzata dalla connessione a forma di X dei 4 quadrati adiacenti. I video in basso effettuano una verifica sul colore dei quadranti e vi faranno capire, più facilmente, l'illusione.



INDEX
CAROUSEL
BHD EYELI



Curiosità – Illusione della Thatcher

L'Effetto Thatcher o Thatcher illusion è un fenomeno caratterizzato dalla difficoltà nel riconoscere cambiamenti nei tratti somatici di un volto mostrato a rovescio, i quali apparirebbero invece ovvi in un volto orientato normalmente.

L'effetto è illustrato da due immagini di volti apparentemente identiche, una ruotata di 180°. La seconda immagine è alterata - mostrando occhi e bocca specchiati in senso verticale - tuttavia non è possibile rendersi conto delle modifiche fino a che non la si orienta normalmente.

Ciò si pensa sia dovuto ad uno specifico modulo cognitivo psicologico legato alla percezione dei volti, particolarmente marcato nei volti orientati verticalmente. I volti tendono ad apparire unici nonostante siano molto simili. È stato ipotizzato che l'uomo abbia sviluppato specifici processi nel differenziare i volti che fanno affidamento tanto sulla loro configurazione (ovvero la relazione strutturale dei tratti somatici) quanto agli specifici dettagli di un viso, come occhi, naso e bocca.

Inoltre, gli occhi e la bocca sono gli elementi che convogliano le informazioni più significative per il nostro cervello, addestrato a trarre da questi elementi informazioni sulle intenzioni e le emozioni degli altri. Alcune evidenze scientifiche suggeriscono che il cervello abbia una regione specializzata nel fare esclusivamente questo: riconoscere facce. Si tratta dell'area facciale fusiforme, che negli studi di neuroimaging si attiva particolarmente se un soggetto osserva volti, ed è più silente quando osserva altri tipi di stimoli. Inoltre, esiste una condizione patologica chiamata prosopoagnosia che impedisce a chi ne soffre di riconoscere volti noti, lasciando intatta praticamente ogni altra facoltà cognitiva.

Le capacità di riconoscimento di un sistema biometrico sono invarianti rispetto all'orientamento dell'immagine. L'occhio umano non lo è!



Il volto di Marte

Che esista o meno un'area specificatamente coinvolta in questo processo, è indubbio che gli esseri umani abbiano sviluppato una particolare abilità a riconoscere volti. Abilità utilissima per farci vivere assieme ai nostri simili e talvolta persino sovrautilizzata. Si pensi alla faccia su Marte.

La **pareidolia** è la tendenza del cervello umano a ricondurre a forme note oggetti o profili naturali o artificiali dalla forma casuale. È una tendenza istintiva che ricrea immagini significative da materiale non strutturato, mischiando percezione sensoriale con immaginazione.

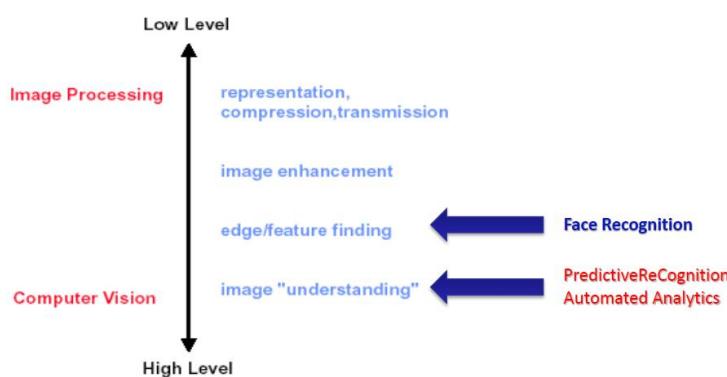
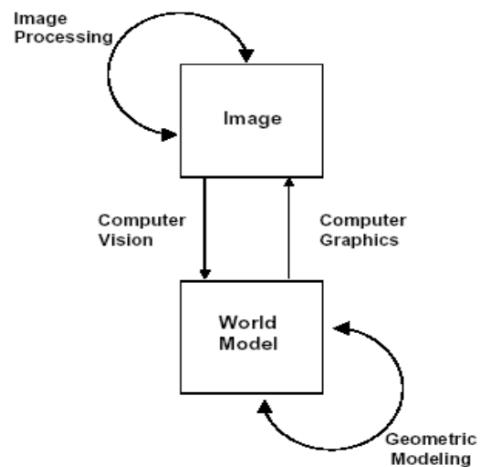


Definizioni di base Campionamento, Istogramma

Come si evince dal grafico sulla destra, si nota come nel quadrato dove viene rappresentata una immagine del mondo reale, applicando una tecnica di **Image Processing**, si ricava una ulteriore immagine che differisce dalla precedente per una qualche motivazione, provando ad estrarre degli aspetti qualificanti (si pensi alla radiografia in campo medico per evidenziare una anomalia), infatti la freccia presente in figura parte e arriva sulla stessa immagine.

Quando invece si effettua una operazione di **Computer Vision**, si collega un qualche aspetto dell'immagine del mondo reale, a qualcosa di più astratto quello che risiede nel mondo della conoscenza; automatizzando determinati passaggi. La Computer Vision risulta essere una evoluzione rispetto all'Image Processing, dove quest'ultima non fa nient'altro che trasformare i pixel senza interpretarli. Invece la Computer Vision assegna a questi pixel un significato sulla base di una informazione fornita da un esperto, interpretando il contenuto semantico dell'immagine sulla base di un dominio di competenza.

La modellazione geometrica rappresenta oggetti astratti e la **Computer Graphics** li rende visibili sottoforma di immagine (si pensi alla progettazione di una nuova auto che passa dall'idea alla realizzazione grafica).



tentano di capire cosa è presente dentro l'immagine, considerata un'attività di alto livello (interpretazione e predizione). Vi sono svariate applicazioni dell'Image Processing e Computer Vision:

- **Robotics**
 - Object Recognition (assembly line)
 - Autonomous Vehicles Obstacle
 - Avoidance
- **Earth Remote Sensing**
 - Pollution Control (River, City)
 - Landsat Image Analysis, Synthesis and Coding
- **Arial Photography**
 - Image Enhancement (espionage)
 - Missile Guidance
 - Geological Mapping
- **Astronomy**
 - Astronomical Image Enhancement
 - Chemical/Spectral Analysis
- **Medicine**
 - Computerized Scanners (MRI, CT, etc)
 - Radiological Organ Segmentation
- **Chemistry**
 - Chemical Composition
 - Molecular Imaging
 - Electron Microscope
- **Graphics**
 - Image Warping
 - Animation
 - Texture/Image Mapping
- **Security and Forensic**
 - Biometric
 - Videosurveillance

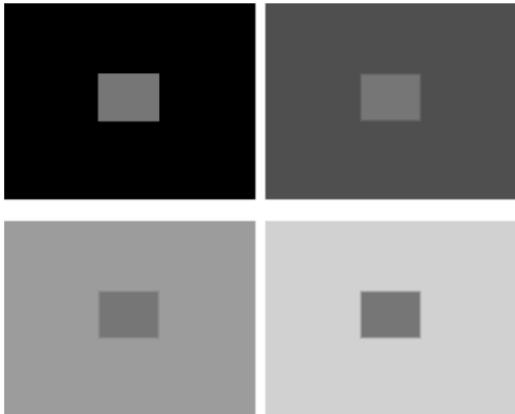
Da ciò che si è spiegato in precedenza si può capire che l'Image Processing risulta essere un'attività di basso livello, siccome essa non si addentra nella semantica dei pixel (non si entra nell'interpretazione di ciò che è presente nell'immagine) come le operazioni di compressione, rappresentazione e trasmissione.

Man mano che ci si muove sulla linea, si va verso la Computer Vision ovvero quelle tecniche di elaborazione che

Intensità e Luminosità

Intensità dell'immagine: è l'energia luminosa emessa da un singolo punto dell'immagine e dipende dal dispositivo. Ovvero un dispositivo in grado, in un singolo punto, di emettere una energia luminosa diversa; maggiore è l'intensità prodotta e più il device sarà sofisticato.

Luminosità dell'immagine: il dispositivo emette una certa intensità luminosa che viene percepita dal soggetto sotto forma di luminosità. Essa è ciò che viene percepito della luminosità emessa da una singola unità (singolo pixel). Dipende dal contesto (ed è anche soggettiva), che la rendono variabile.

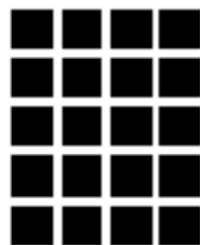


Come si vede nell'esempio (a sinistra) tutti i quadratini hanno esattamente la stessa intensità ma appaiono progressivamente più scuri man mano che lo sfondo diventa più chiaro.

Un altro esempio è quello di considerare queste sei strip (in basso), ognuna di esse possiede una intensità costante all'interno della strip ma variabile da strip a strip. Ognuna di esse confina a destra con una strip di intensità più chiara e a sinistra con una strip con intensità più scura. Il fenomeno della luminosità indotta è che viene percepita è progressivamente più scura muovendosi da sinistra verso destra, man mano che ci si muove verso una strip più chiara, la precedente risulta essere più scura, mantenendo un profilo più scuro verso la strip più chiara e un profilo più chiaro verso la strip più scura pur essendo di intensità costante.



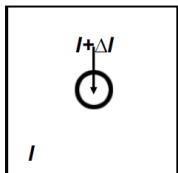
Un altro esempio ancora è il seguente dove vengono percepiti ad intermittenza dei puntini di colore grigio. Pur essendo l'intensità costante lungo le intersezioni, la presenza della varianza chiaro/scuro tende a far comparire questi puntini grigi all'interno delle intersezioni. È ancora una dimostrazione che l'intensità è costante ma la luminosità viene contestualizzata all'immagine che si sta osservando.



Un altro esempio relativo alla luminosità, ovvero alla capacità umana di percepire variazioni di luminosità è legata alla Legge di Weber. Esso posizionava un utente di fronte a uno schermo, il quale nella porta posteriore (non visibile all'utente) aveva un cannone luminoso che sparava una certa intensità luminosa. Succedeva che lo schermo veniva illuminato con una certa intensità L e il centro dello schermo veniva illuminato con $L + \delta(L)$. Inizialmente il centro dello schermo aveva la stessa intensità dello sfondo senza esser percepita nessuna differenza. Facendo variare le due intensità, si chiedeva all'utente fin quando fosse stato possibile individuare queste variazioni di intensità. Weber ha dedotto che l'essere umano possiede una buona capacità di

Good Brightness Discrimination:
 $\Delta I/I = \text{small value}$

Poor Brightness Discrimination:
 $\Delta I/I = \text{large value}$



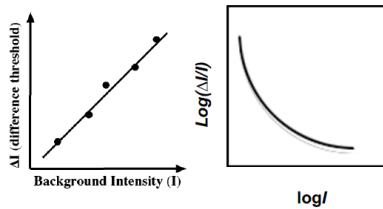
In general, ΔI needed for just noticeable difference (JND) over background I was found to satisfy:

$$\frac{\Delta I}{I} = \text{constant}$$

(I is intensity, ΔI is change in intensity)

Weber's Law:

$$\text{Perceived Brightness} = \log(I)$$

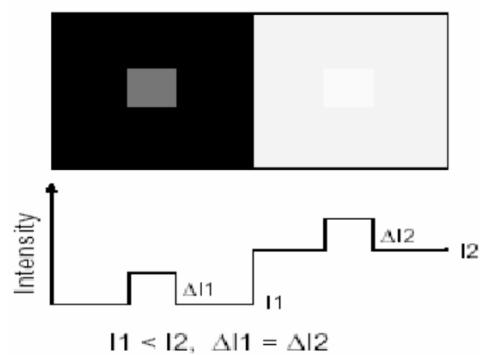


l'intensità dello sfondo, direttamente aumenta la variazione che dovrà essere prodotta per essere visibile.

Image Contrast

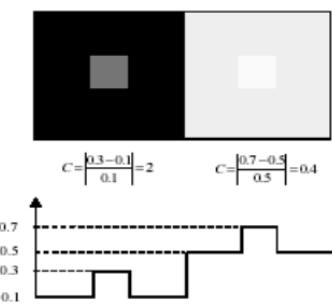
Un altro concetto importante che spesso si confonde nel definire una immagine è il **contrasto di una immagine**. Esso dovrebbe fornire una idea di quanto gli elementi presenti in una immagine sono distinti uno dall'altro. Queste due immagini hanno un contrasto diverso da un punto di vista percettivo anche se dal

punto di vista numerico no. L'immagine di sinistra presenta due rettangoli uno con uno sfondo uniforme ed un altro. All'interno vi è un altro rettangolo più chiaro. Seguendo l'intensità che rende l'idea di quale è la variazione tra lo sfondo e l'oggetto, lo sfondo si mantiene costante fin quando non si intercetta il rettangolo grigio, poi vi è un incremento ΔI_1 che mantiene costante l'intensità per un certo periodo per poi decrescere e ritornare di nuovo sullo sfondo. Quindi la variazione indotta nel passare dallo sfondo all'oggetto è un ΔI_1 , quindi una variazione di luminosità quantificabile in ΔI_1 . Il secondo rettangolo è simile al primo ma con intensità diverse, partendo con una intensità



più chiara, mantenendosi costante fin quando non viene intercettato il rettangolo più chiaro. Quindi anche in questo caso vi è una variazione di luminosità ΔI_2 costante per poi scendere nuovamente e mantenersi costante fino alla fine. La variazione indotta nel passaggio dallo sfondo all'oggetto è ΔI_2 . Le due variazioni sono equivalenti $\Delta I_1 = \Delta I_2$, ma l'immagine di sinistra è meglio contrastata, in quanto l'oggetto centrale si nota in maniera più distinta rispetto allo sfondo, benché le variazioni siano equivalenti. Quindi se si volesse misurare il contrasto di una immagine non si potrebbe semplicemente calcolare la variazione tra un oggetto e un altro (sfondo e oggetto centrale), ma va considerato un ulteriore elemento, ovvero una sorta di normalizzazione. Per ottenere il valore numerico che evince quale immagine è meglio contrastata rispetto ad un'altra, si può

$$C = \left| \frac{I_p - I_n}{I_n} \right|$$



normalizzare rispetto allo sfondo, come si vede dalla figura dove viene assegnato un valore alla luminosità per ogni rettangolo e relativo sfondo e poi vengono effettuate le operazioni di normalizzazione dividendo per il valore dello sfondo ottenendo due valori, risultati dalla formula in rosso, che permettono di individuare il contrasto di una immagine, in questo caso siccome $2 > 0.4$, si può dire che la prima immagine presenta un contrasto migliore all'immagine precedente.

Ora si presentano varie immagini che hanno ognuna un contrasto leggermente differente:



A sinistra una immagine ideale, dove il livello del contrasto è sufficientemente buono in quanto tutti i particolari presenti nell'immagine sono facilmente individuabili.



Low contrast



High contrast

Invece in queste altre due foto si può notare come in quella con basso contrasto non tutti i particolari sono evidenti. L'immagine ad alto contrasto porta a confondere vari particolari (capelli e occhi vengono confusi).



Non contrast Abdominal Imaging



Contrast-enhanced Abdominal Imaging

Intervenendo sull'immagine, prima che essa venga elaborata e/o osservata, è possibile esaltare il contrasto con determinate operazioni sui pixel affinché alcuni oggetti all'interno dell'immagine vengano enfatizzati e visibili in modo migliore. Di conseguenza migliorare anche

il risultato dell'elaborazione che si prevede di effettuare.



Non contrast MR Cardiac Imaging



Contrast-enhanced MR Cardiac Imaging

Image Formation

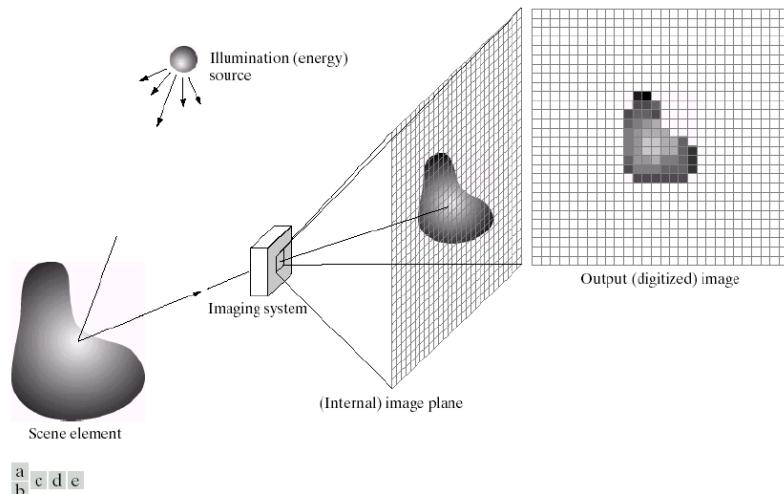
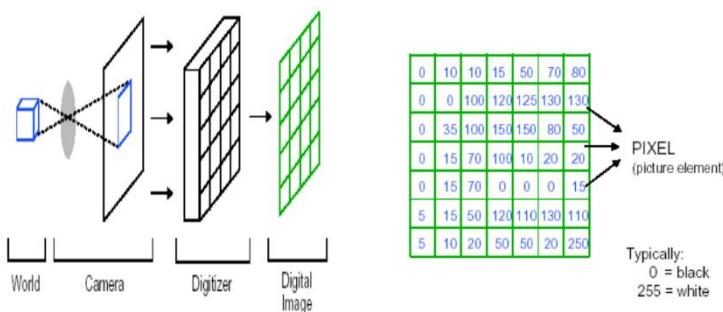


FIGURE 2.15 An example of the digital image acquisition process. (a) Energy (“illumination”) source. (b) An element of a scene. (c) Imaging system. (d) Projection of the scene onto the image plane. (e) Digitized image.

$$f(x,y) = \text{reflectance}(x,y) * \text{illumination}(x,y)$$

Reflectance in [0,1], illumination in [0,inf]

Fondamentalmente, si può effettuare la digitalizzazione perché si ha una sorgente luminosa che illumina l'oggetto per effetto della riflessione dei raggi luminosi sull'oggetto, il sensore acquisisce le informazioni ottenute, di rimbalzo sull'oggetto e produce una rappresentazione digitalizzata dell'immagine.



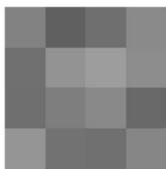
pixel (acronimo di picture element).

Il processo di **digitalizzazione di una immagine** risulta essere molto importante in quanto digitalizzando una immagine o un video, si perde inevitabilmente una parte dell'informazione. Quando si acquisisce una immagine con una singola sorgente (dispositivo) si passa da una rappresentazione tridimensionale ad una rappresentazione bidimensionale, e questo passaggio da 3D a 2D comporta irrimediabilmente una perdita di informazioni. Se tale processo non viene svolto nella maniera adeguata il processo di digitalizzazione viene aggravato da un ulteriore perdita di informazione.

Come si vede dalla figura, vi è l'oggetto tridimensionale, che viene acquisito dal dispositivo il quale all'interno presenta una rete la quale, in ogni posizione prevede un sensore incaricato ad acquisire una porzione del mondo reale e trasformare tale acquisizione in termini numerici, calcolando la quantità di luce che viene riflessa in quella determinata posizione convertendola in numero, composta da

Sampling (Campionamento)

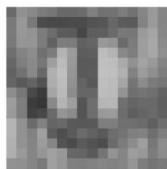
Si vuol costruire l'immagine attraverso un'operazione di **sampling** (campionamento progressivo). Il campionamento è l'operazione attraverso la quale si decide quanti sensori utilizzare nella fase di digitalizzazione. Quindi la fase di digitalizzazione è l'intero processo che viene diviso in due fasi, una di esse è il campionamento. Come si evince dalla figura verranno utilizzati in modo progressivo sempre più sensori



$4 \times 4 = 16$



$8 \times 8 = 64$

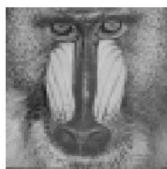


$16 \times 16 = 256$

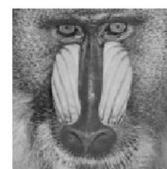
risultati della moltiplicazione tra numero di righe e numero di colonne. Man mano che aumentano i sensori e la porzione di acquisizione di ognuna di essa diventa sempre più piccola, l'immagine assume una qualità migliore.



$32 \times 32 = 1024$



$64 \times 64 = 4096$



$128 \times 128 = 16384$



Risoluzione spaziale: da sinistra a destra 256x256, 128x128, 64x64 e 32x32 pixel

In questo caso vi è un deterioramento dell'immagine passando da 256x256 ad una progressiva blocchetizzazione 32x32 che è sintomo di basso campionamento.

Quantizzazione

L'altra fase del processo di digitalizzazione di una immagine, oltre al campionamento, è la quantizzazione. Mentre il campionamento concerne il numero di sensori (pixel) impegnati nella fase di acquisizione, **la quantizzazione** è legata al numero di livelli di intensità luminosa (colori) che ogni sensore può produrre.



1 bit per pixel



2 bit per pixel



3 bit per pixel



4 bit per pixel



8 bit per pixel

Nelle figure a sinistra, viene assegnato per ogni sensore un determinato numero di bit. Ogni sensore converte la quantità di luce proveniente dal mondo reale in base al numero di bit assegnati. Man mano che aumentano i numeri di bit, di conseguenza aumentano il numero di livelli di intensità luminosa migliorando sensibilmente la qualità dell'immagine. Tutte le immagini hanno conservato un campionamento costante (non sono variati il numero di sensori) con quantizzazione variabile.



Risoluzione spaziale: da sinistra a destra 256x256, 128x128, 64x64 e 32x32 pixel



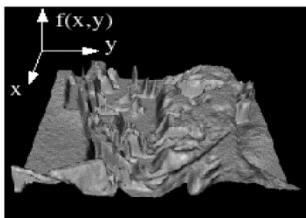
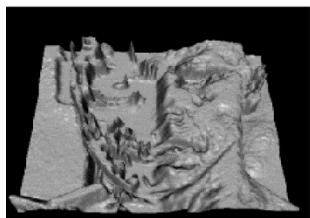
Risoluzione sulle ampiezze: da sinistra a destra 16, 8, 4 e 2 livelli

Come nell'esempio precedente, nelle 4 immagini in alto viene mantenuto costante la quantizzazione facendo variare il campionamento. Le 4 immagini in basso hanno mantenuto costante il campionamento e fatto variare la quantizzazione.

Poiché la fase di digitalizzazione è costituita da campionamento e quantizzazione, se si sbaglia una delle due fasi si compromette il risultato finale, che in situazioni critiche non è accettabile. Quindi quantizzazione e campionamento vanno calibrate in maniera distinta e rigorosa, anche in base al contesto in cui si opera.

What is an image?

Fondamentalmente l'immagine è un insieme di informazioni luminose che vengono catturate da questi sensori. Tali informazioni luminose possono essere convertite in una informazione tridimensionale.



Come si vede dall'immagine, ad ogni posizione (x, y) nella rappresentazione, corrisponde un valore $f(x, y)$ che risulta essere il livello di luminosità catturato dal sensore. La rappresentazione tridimensionale di una immagine in termini di $f(x, y)$ può essere quella dove i colori **più bassi** sono quelli tendenzialmente **più scuri**, man mano che si va verso una zona **più chiara** vi sono dei **rilievi**. Quindi il valore $f(x, y)$ tende a crescere dando l'effetto di tridimensionalità.

Si può quindi immaginare l'immagine come un insieme di valori $f(x, y)$ dove ad ogni pixel (x, y) si

assegna un valore di luminosità, che produce la rappresentazione tridimensionale.

$i \downarrow$	$j \longrightarrow$	62	79	23	119	120	105	4	0
10		10	9	62	12	78	34	0	
10		58	197	46	46	0	0	48	
176		135	5	188	191	68	0	49	
2		1	1	29	26	37	0	77	
0		89	144	147	187	102	62	208	
255		252	0	166	123	62	0	31	
166		63	127	17	1	0	99	30	

Questa è la rappresentazione matriciale, presente in memoria, dove ogni valore corrisponde ad un livello di luminosità variabile a seconda se il sensore assegnato ad ogni pixel ha catturato un livello di intensità scuro o chiaro.



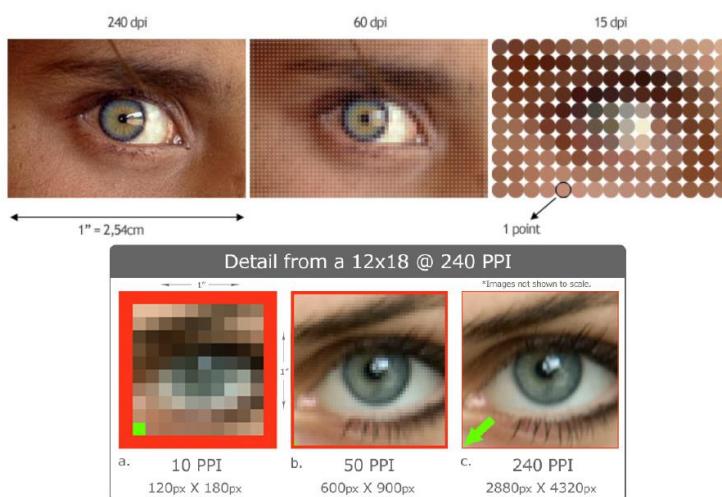
x =	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	
y =	41	210	209	204	202	197	247	143	71	64	80	84	54	54	57	58
42	206	190	203	197	195	210	207	56	63	58	53	53	61	62	51	
43	207	192	201	198	213	156	69	65	57	55	52	53	60	50		
44	216	206	211	193	202	207	208	57	69	60	55	77	49	62	61	
45	221	206	211	194	196	197	220	56	63	60	55	46	97	58	57	
46	209	214	224	199	194	193	204	173	64	60	59	51	62	58	48	
47	204	212	213	208	191	190	191	214	60	62	66	76	51	49	55	
48	214	215	215	207	208	180	172	188	69	72	55	49	56	52	56	
49	209	205	214	205	204	196	187	196	86	62	66	87	57	60	48	
50	208	209	205	205	203	202	186	174	185	149	71	63	55	55	45	56
51	207	210	211	199	217	194	183	177	209	90	62	64	52	93	52	
52	208	205	209	209	197	194	183	187	187	239	58	68	61	51	56	
53	204	206	203	209	195	203	188	185	183	221	75	61	58	60	60	
54	200	203	199	236	188	197	183	190	183	196	122	63	58	64	66	
55	205	210	202	203	199	197	196	181	173	186	105	62	57	64	63	

Risoluzione

La risoluzione lega il campionamento e la dimensione fisica. Va divisa in due elementi: **Temporale e Spaziale**. **Temporale** (per i video): definisce il numero di immagini al secondo da usare (FPS). Per la visione umana si usano da 15 a 30 immagini al secondo. Al di sotto di tale valore vi è una bassa risoluzione temporale. **Spaziale** (singole immagini): definisce il numero di pixel da usare (in alternativa DPI, dot per inch: il numero di pixel al centimetro).



di blocchetizzazione è quasi totalmente scomparso. Quindi pur avendo due immagini di uguale dimensione, ma con un numero di pixel maggiore l'una rispetto all'altra, si evince come quest'ultima sia qualitativamente migliore, legata al numero di DPI.



Queste due immagini hanno la stessa dimensione fisica ma sono profondamente diverse. A destra l'effetto di blocchetizzazione è molto più evidente rispetto all'immagine di sinistra. Tale effetto è molto evidente dove vi sono i dettagli più importanti; infatti l'immagine è stata acquisita con 72 DPI. L'immagine di sinistra acquisita con 150 DPI, risulta essere sensibilmente migliore e l'effetto

Un altro esempio dove vengono mostrate tre immagini, con la stessa dimensione fisica, ma con DPI differenti e di conseguenza qualità differenti. (PPI, punti per pollice).

Image Processing

Come si è detto in precedenza una immagine è assimilabile a una funzione tridimensionale $f(x, y)$. Quindi quando si effettua una operazione di Image Processing che definisce una nuova immagine g in termini di una immagine esistente f , effettuando una **trasformazione** t . Si trasforma un pixel, non interpretando il contenuto dell'immagine, ma viene effettuata una trasformazione t su tale pixel per trasformarlo in un altro pixel.

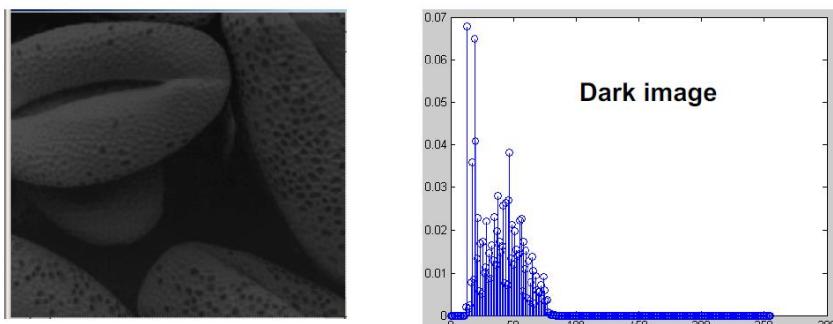
$$g(x, y) = t(f(x, y))$$

Fondamentalmente si possono effettuare varie tipologie di trasformazioni sull'immagine; sulla base anche delle immagini che si hanno a disposizioni, come:

- **Immagini Binarie**, ovvero in bianco e nero, con un bit per pixel.
- **Immagini in Scala di Grigio**, dove oltre al bianco e il nero vi sono varie sfumature di grigio, e la quantità di sfumature in termini di livelli di grigio, dipendono dalla quantizzazione. Una quantizzazione a 8 bit, si avranno complessivamente 256 colori, dove 0 è nero, 255 bianco e le restanti sono sfumature di grigio.
- **Immagini a colori**, dove si compone ogni singolo pixel attraverso una serie di componenti che dipendono dal modello di colore che si utilizza.

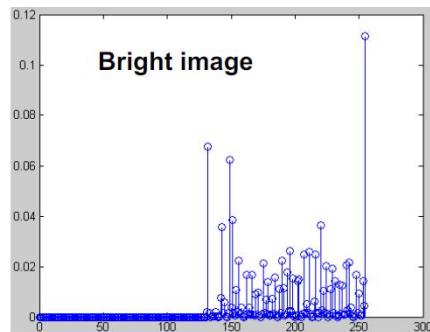
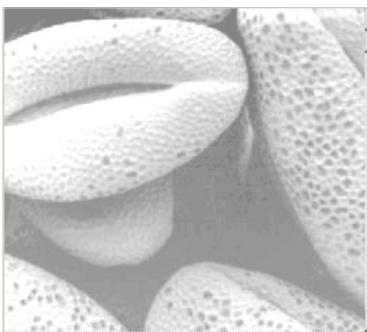
Image Enhancement: Histogram Based Methods

Una misura importante che può essere assegnata all'immagine è quella dell'**istogramma**. Non è nient'altro che una rappresentazione statistica dell'immagine, attraverso il quale è possibile avere una distribuzione di quelli che sono i livelli di grigio. Si ha il numero di occorrenze per ogni tipologia di colore presente nell'immagine. Ad esempio, di seguito una immagine con il relativo istogramma ottenuto: sull'asse delle ascisse viene rappresentato il risultato della quantizzazione, ovvero il numero di colori rappresentabili. Poiché questa immagine è stata digitalizzata usando 8 bit per pixel, sull'asse delle scisse ci saranno 256 colori.

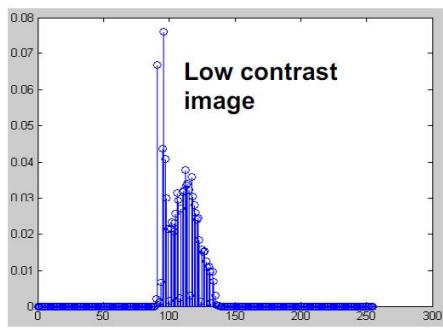
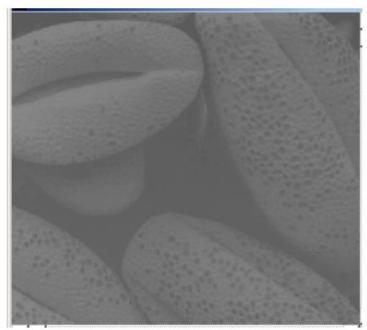


nell'intervallo di colori che varia da 0 a 75, e poche occorrenze di colori chiari (fortemente addensato a sinistra).

L'asse delle ordinate riporta il numero di occorrenze per ognuno dei colori. Ciò significa che ognuno dei picchi rappresenta quante volte il colore corrispondente è rappresentato nell'immagine. Dato che l'immagine presenta una prevalenza di colori scuri, di conseguenza nell'istogramma vi sarà una concentrazione di picchi

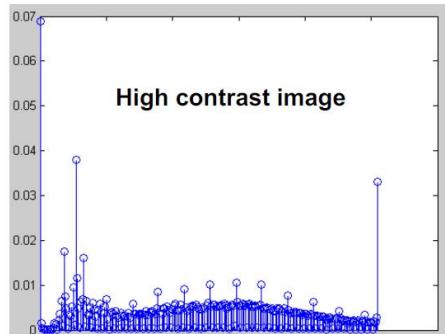
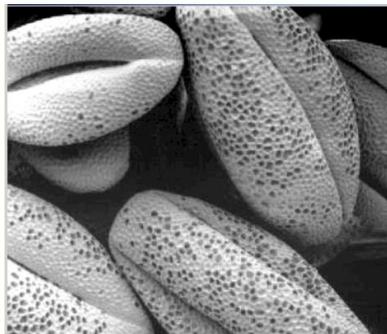


Viceversa, in questa immagine dove si ha una prevalenza di colori chiari, molto luminosa, vi sono dei picchi nella parte destra, ovvero nella parte chiara.

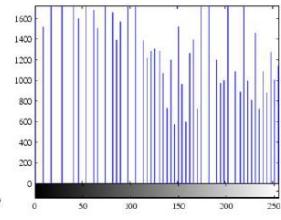
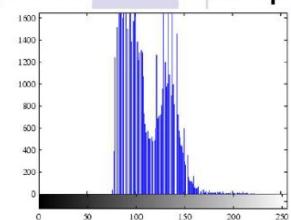


In precedenza, quando si è parlato del contrasto, si è detto che in presenza di una immagine poco contrastata è possibile intervenire per poterla trasformare. Per capire tale concetto, ovvero che una foto è poco contrastata, l'istogramma assumerà la forma in figura, con una campana molto

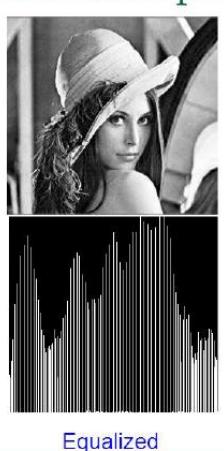
stretta concentrata in pochi colori. Quando un istogramma si presenta in questa modalità, si può effettuare una operazione di **Equalizzazione dell'istogramma**: andando a stiracchiare l'istogramma per ottenere una redistribuzione dei colori, avendo in output la seguente immagine:

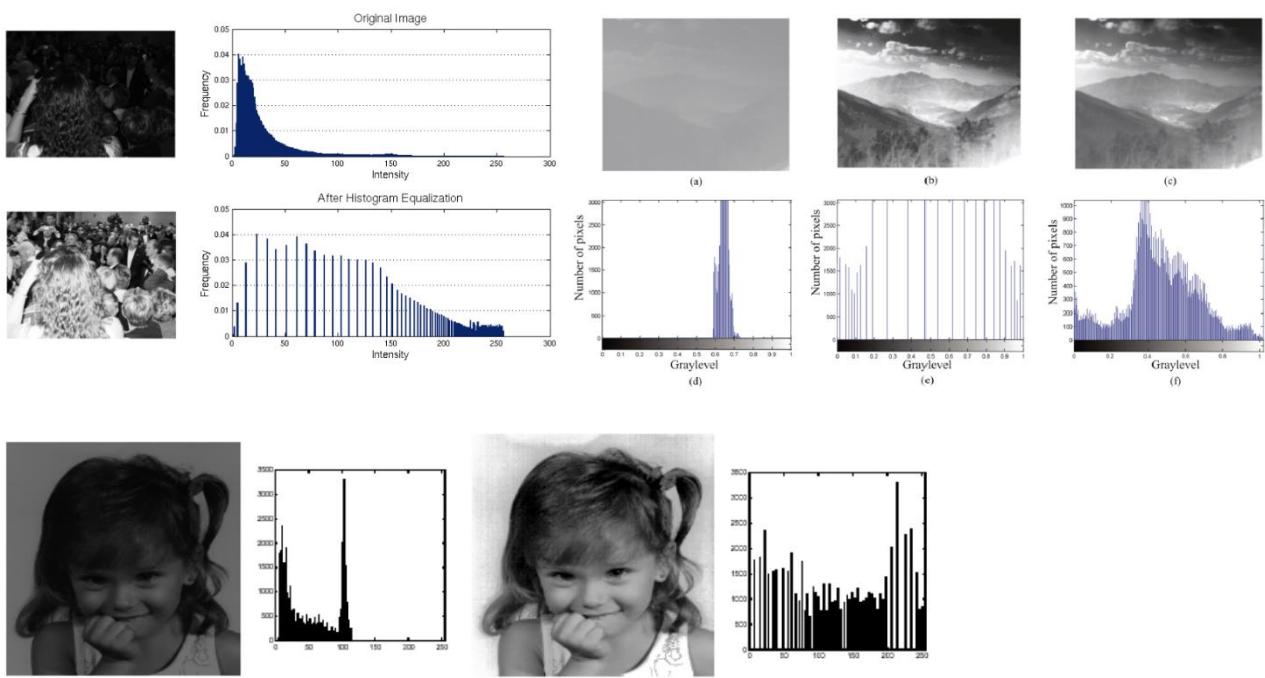


Histogram Equalization Example



Histogram equalization: Example





Tutti gli esempi mostrati, con una campana stretta su pochi colori hanno beneficiato dell'operazione di equalizzazione avendo come output una immagine di qualità superiore rispetto a prima.

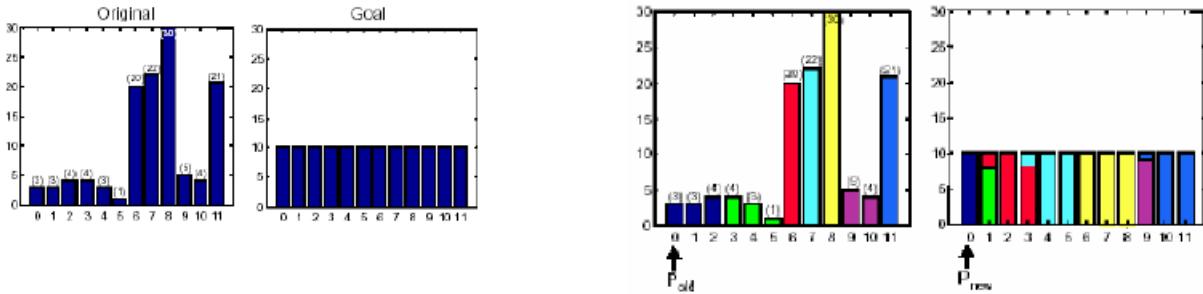
Equalizzazione Istogramma

L'equalizzazione, come detto, non è nient'altro che una rappresentazione statistica, per conoscere la ricorrenza in termini di rappresentazione di ogni colore. Poiché ogni colore rappresenta un certo numero di pixel, la somma delle occorrenze di tutti i colori darà come risultato il numero totale di pixel. Si tratta di rendere l'istogramma il più piatto possibile. Questo potrebbe rendere l'immagine più leggibile, anche se la cosa è opinabile (in alcuni casi). Supponiamo che l'immagine di partenza sia di dimensioni 512x512, ovvero che sia composta da 2^{18} punti (numero dei campioni). Questo significa che, detto $H(i)$ l'istogramma (somma dei valori sull'asse delle ordinate), deve valere che:

$$\sum_{i=0}^{255} H(i) = 2^{18}$$

Se si vuol rendere l'istogramma piatto significa che si vuole vuole

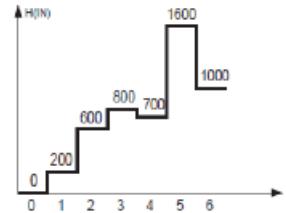
$$H(i) = 2^{18}/256 = 2^{10} = 1024, \forall i.$$



Naturalmente non si potrà ottenere un'esatta uguaglianza, visto i limiti dell'elaborazione, ma si tenta di avvicinarsi il più possibile. Nel primo livello di uscita si vorrebbe ottenere 1024: se si sommano i primi tre livelli (0, 1 e 2) dell'ingresso otteniamo 800, se si sommano i primi quattro 1600. Poiché 800 è più vicino a 1024 di 1600 scegliamo di sommare i primi tre.

Passando al secondo livello dell'uscita: anche lui dovrebbe essere riempito con 1024 livelli ma, per evitare accumuli dell'errore, bisogna considerare che nei primi due livelli voglio $1024 + 1024 = 2048$ punti. Poiché nel primo ne sono già stati inseriti 800 si dovrà cercare di approssimare, per il secondo, il valore di 1248. Quindi si sceglie di inserire i livelli 3 e

IN	OUT
0	0
1	0
2	0
3	1
4	1
5	3
...	...



4 dell'ingresso, visto che $800 + 700 = 1500$ è più vicino a 1248 di 800. Nel terzo livello di uscita ci va ancora 1024 e la somma dei primi tre deve essere 3072. Ne sono stati già sistemati $800 + 1500 = 2300$ e quindi ne restano 772. Poiché il primo livello da sistemare dell'ingresso (il 5) vale 1600 e $1600 - 772 = 828$ si lascia vuoto il terzo livello dell'uscita. Nel quarto livello di uscita devo approssimare $4096 - 2300 = 1996$, se viene inserito il livello di ingresso 5 ne mancano 396, si inserisce 5+6 ne avanzano $2600 - 1996 = 604$ quindi scelgo 5.

Si analizzano ora le impostazioni generali per effettuare una equalizzazione. Si è detto che tale operazione trasforma l'istogramma iniziale G in un altro istogramma \hat{G} (cappelletto).

$$\hat{G} = T(G)$$

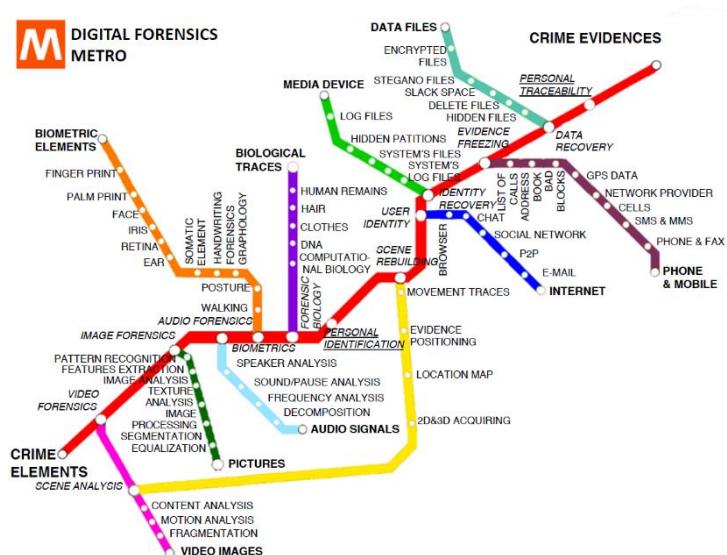
Una volta acquisita l'immagine e aver preso coscienza della campana relativamente stretta, si applica la trasformazione T all'istogramma per produrre il nuovo istogramma \hat{G} (cappelletto). Questa trasformazione T deve seguire tre regole tassonomiche (ordine di importanza):

1. L'istogramma deve essere il **più piatto possibile**, senza grossa variazione tra le occorrenze.
2. L'**ordine dei livelli di grigi** deve essere mantenuto, con piccole oscillazioni.
3. L'istogramma deve essere **non frammentato**, non vi devono essere dei "salti".

Introduzione alla biometria

In questa rappresentazione vi è una linea principale (la linea rossa) che rappresenta la dorsale principale del contesto forense. Gli elementi del crimine e l'evidenza del crimine. Ovvero due modi di approccio: **analizzare** ciò che successo dopo che il crimine è reso evidente, oppure **prevedere** gli eventi prima che il crimine possa essere realizzato.

Il problema della biometria, di una certa rilevanza, corrisponde al verificare se l'identità dichiarata corrisponde all'identità reale.



Autenticazione di un soggetto

I processi di autenticazione di un soggetto possono essere molto diversificati. Ovvero un processo di associazione dell'identità ad un soggetto attraverso un processo di **verifica o riconoscimento**:

- Una persona riconosce una persona con sistemi e supporti di identificazione come carta d'identità, codice fiscale. Oppure una persona ne riconosce un'altra per conoscenza diretta (ambito familiare).
- Un sistema informatico riconosce una persona utilizzando sistemi e supporti di identificazione come smart card, biometriche, ecc...

Queste metodologie presentano dei limiti come ad esempio, in presenza di un PIN (login e password), è possibile dimenticarlo oppure essere carpito da un impostore. Anche nel caso delle chiavi fisiche (smart card) che si possono rompere o perdere.

Questi sistemi non consentono la distinzione tra il reale possessore e un impostore.

Per superare i limiti dei sistemi di autenticazione attualmente in uso:

- Password (qualcosa che si possiede), carte magnetiche o Smart, chiavi, passaporto, ecc...
- Conoscenza (qualcosa che si conosce), password, PIN, ecc...
- **Biometria** (ciò che si è), una caratteristica fisica o comportamentale non associabile a nessun'altra persona.

Si deve superare il meccanismo di autenticazione tradizionale basato sul concetto che l'utente possiede, spostando l'attenzione su qualcosa che "l'utente è effettivamente".

Cosa è la Biometria?

Dal greco bios (vita) e metros (misura), la biometria è basata sul riconoscimento automatico di una persona sulla base di caratteristiche discriminanti.

La biometria è la disciplina che si occupa di misurare tratti fisiologici e comportamentali degli esseri viventi, offrendone una rappresentazione quantitativa attraverso modelli matematici e statistici per costruirne un template ed effettuare eventuali confronti.

Le tecnologie biometriche costituiscono metodi automatizzati di verifica o riconoscimento dell'identità di una persona, basati sul rilevamento di una o più **caratteristiche fisiologiche o comportamentali** dell'individuo, oppure il confronto con una immagine precedentemente acquisita.

I principi portanti della biometria sono:

- Ogni persona è **unica**, dove alcune biometrie sono più specifiche di altre.
- Individuazione delle **caratteristiche somatiche** che rendono unico un individuo.
- Metodologie per la **misurazione e quantificazione** di tali caratteristiche.
- **Classificazione** degli individui sulla base delle misure effettuate.

Livelli di autenticazione

Something you have:

- Token
 - Key
 - Card or badge

Something you know:

- Password
- PIN
- A memory "unique" to you

Something you are:

- Biometric
 - Physiological
 - Behavioral



I livelli di autenticazione normalmente utilizzati, dipendono dal livello di sicurezza che si intende raggiungere. Vi sono dei livelli di sicurezza in cui il livello di autenticazione deve essere molto forte.

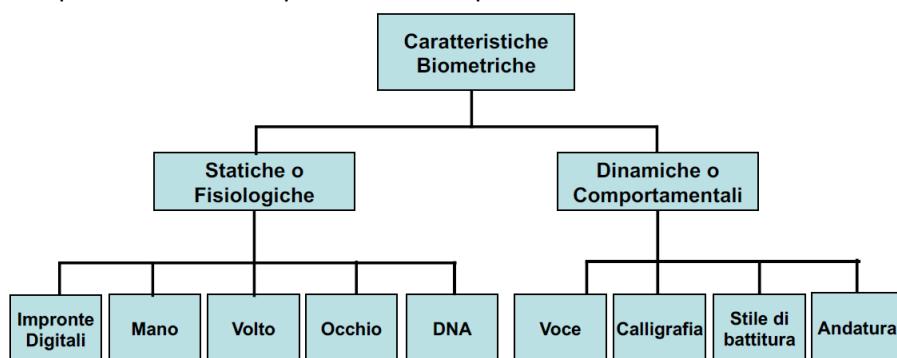
Partendo dal basso, si individuano i livelli di sicurezza meno forti, come ad esempio quelli legati a qualcosa che si conosce come una password, PIN ecc... Un livello superiore è il

cosiddetto Token, basato su qualcosa che si deve avere come una chiave, un badge ecc, quasi sempre legate a qualcosa che si conosce per aumentarne il livello di sicurezza... Infine, vi è un'informazione biometrica, fisiologica o comportamentale, quindi qualcosa che realmente "si è" che possa essere eventualmente abbinata alle informazioni appena descritte. L'incrocio di questi tre livelli di sicurezza produce un livello di sicurezza molto elevato. Le altre combinazioni tra di esse rendono comunque alto il livello di sicurezza.

Caratteristiche Biometriche

Parlando di caratteristiche biometriche è possibile effettuare una classificazione: come si vede in figura ci sono le caratteristiche biometriche **statiche o fisiologiche** e le caratteristiche **dinamiche o comportamentali**. Le prime, tendenzialmente, hanno una loro stabilità nel tempo ma soprattutto non dipendono dal comportamento della persona e non possono essere variate sulla base di aspetti emotivi. Viceversa, le

caratteristiche dinamiche dipendono fortemente dagli aspetti emotivi della persona, e possono essere anche variabili, avendo un impatto sulla base di aspetti emotivi o legati alla salute o stile di vita.



Tecniche riconosciute dallo Standard X9.84 - 2003

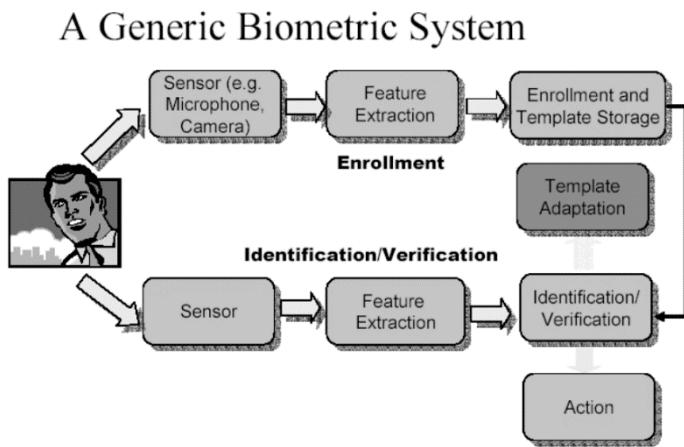
Per poter utilizzare una biometrica nel contesto forense, è necessario che esse rientrino in una normativa, affinché esse siano legalmente valide. Di seguito si elencano varie biometriche utilizzate attraverso un riconoscimento automatico in un contrasto forense:

- Biometria delle impronte – riconoscimento delle **impronte**
- Biometria dell'occhio – riconoscimento dell'**iride** e della **retina**
- Biometria del volto – riconoscimento **facciale** (foto, infrarosso)
- Biometria della **mano** – geometria delle dita
- Biometria della firma – riconoscimento della **firma**
- Biometria della **voce** – riconoscimento vocale
- Biometria auricolare – riconoscimento dell'**orecchio**
- Digitazione di tasti
- DNA



Architettura di un Sistema Biometrico

Un sistema biometrico in grado di operare sulle componenti appena descritte, che possano essere trattate in maniera automatica. Di seguito uno schema di un sistema biometrico generico:



Un sensore che sia in grado di trasformare l'informazione biometrica in un segnale digitale. Una volta digitalizzato tale dato, vengono estratte delle caratteristiche, che dipendono dal tipo di biometrica utilizzata. Tali caratteristiche saranno la base della creazione del modello matematico (**template**). Questo modello diviene il template rispetto al quale, successivamente si faranno operazioni di verifica e riconoscimento. Con la creazione del template, si è in grado di poter riconoscere un determinato individuo, riacquisendolo in maniera analoga a come appena descritto, e una volta creato il template temporaneo verificare se esso è presente nell'archivio per produrre un risultato.

Registrazione di un soggetto (Enrollment)

Si associa un insieme di caratteristiche all'identità di un soggetto, raccogliendo dei dati ed estraendo delle caratteristiche. Il modello estratto (template) è memorizzato in un database o su un supporto portatile (smart card). Questo processo può essere effettuato singolarmente o in blocco (batch Enrollment).

Verifica e Riconoscimento

Non sempre una stessa biometrica può essere utilizzata sia verifica che per fare riconoscimento. Dipende dalle condizioni, una delle quali è proprio scegliere se svolgere Verifica o Riconoscimento. Nel caso delle Verifica si parla di confronto **Uno a Uno**, dove un individuo fornisce la propria identità (mediante carta d'identità, codice utente ecc, in un qualsiasi modo, e il sistema una volta che è stata acquisita l'identità,

effettua un match 1 a 1, acquisendo il template della persona in real time, confrontandolo con quello memorizzato, e confermando o negando l'identità dichiarata da un individuo.

Viceversa, nella fase di riconoscimento dove si verifica un confronto **Uno a Molti**, ovvero quando l'utente non fornisce la propria identità (esempio telecamera al gate aeroportuale). In questo caso si acquisisce il dato biometrico e lo si confronta con quelli memorizzati in vari storage. Quindi si stabilisce l'identità di un soggetto a partire da un insieme di persone registrate.

Formazione delle Biometriche

Alcune biometriche risultano essere più discriminanti di altre. Durante la fase di **angiogenesi**, nel grembo materno impatta maggiormente sulla formazione di due delle biometriche, più discriminanti e stabili nel corso della vita di un utente, ovvero **impronta digitale** e **iride**. La fase di angiogenesi è il processo fenotipico di sviluppo di nuovi vasi sanguigni a partire da altri già esistenti.

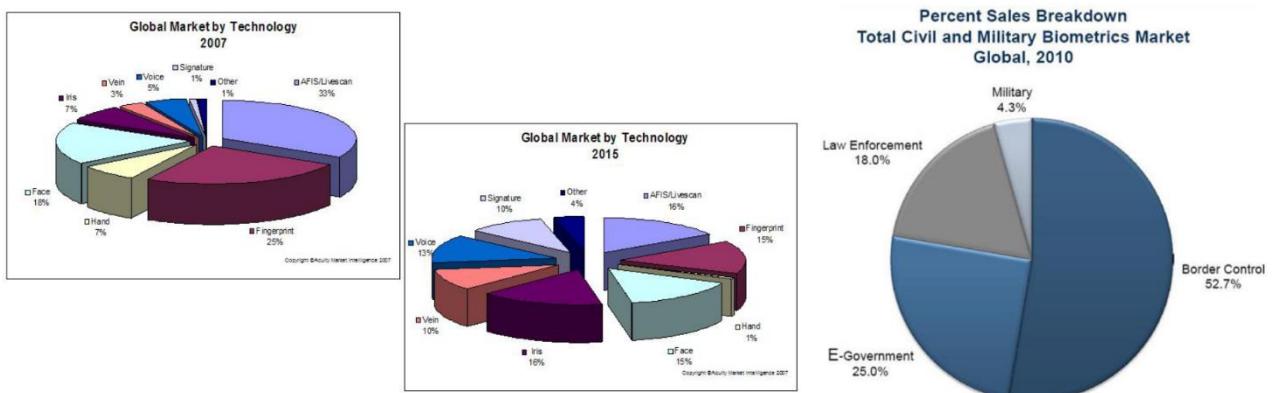
<i>Biometric characteristic</i>	<i>genotypic*</i>	<i>randotypic*</i>	<i>behavioral**</i>	<i>Biometric Trait</i>	<i>Permanence over time</i>
Fingerprint (only minutia)	o	ooo	o	Fingerprint (Minutia)	oooooooooooo
Signature (dynamic)	oo	o	ooo	Signature(dynamic)	oooo
Facial geometry	ooo	o	o	Facial structure	oooooo
Iris pattern	o	ooo	o	Iris pattern	oooooooooooo
Retina (Vein structure)	o	ooo	o	Retina	oooooooooooo
Hand geometry	ooo	o	o	Hand geometry	oooooooooooo
Finger geometry	ooo	o	o	Finger geometry	oooooooooooo
Vein structure of the hand	o	ooo	o	Vein structure of the back of the hand	oooooo
Ear form	ooo	o	o	Ear form	oooooo
Voice (Tone)	ooo	o	oo	Voice (Tone)	ooo
DNA	ooo	o	o	DNA	oooooooooooo
Odor	ooo	o	o	Odor	ooooooo ?
Keyboard Strokes	o	o	ooo	Keyboard strokes	oooo

Nella prima tabella in alto (dove più cerchi indicano una corrispondenza forte), vengono analizzate alcune biometriche, e per ognuna di esse vengono individuati tre possibili fattori di caratterizzazione, **genotipico**, **randotipico** (o fenotipico) e **comportamentale**. Concentrandosi sui primi due si può evincere che le componenti genotipiche vengono trasmesse direttamente attraverso il genoma dei genitori. La componente randotipica non è condivisa con nessuno ed è strettamente legata all'utente. Quanto più è forte quest'ultima componente tanto più si riesce a differenziare una persona, anche rispetto ai parenti. L'altro elemento che si analizza è legato alla durata nel tempo di una biometrica, nella seconda tabella. Se una biometrica risulta essere molto variabile, non è pratico utilizzarla per lunghi periodi di tempo. La tabella indica la robustezza delle biometrie, dove il numero di cerchi è relazionato alla robustezza di tale biometrica, ovvero più cerchi più robustezza. Quindi una biometrica se è stata fortemente condizionata da elementi randotipici risulta essere più discriminante.

Nella tabella in seguito vengono mostrati i sensori con cui vengono catturate le caratteristiche biometriche.

Biometric characteristic	Sensor
Fingerprint (Minutia)	capacitive, optic, thermal, acoustic, pressure sensitive
Signature (dynamic)	Tablet
Facial Structure	Camera
Iris pattern	Camera
Retina	Camera
Hand geometry	Camera
Finger geometry	Camera
Vein structure of the hand	Camera (infrared)
Ear form	Camera
Voice (Timbre)	Microphone
DNA	Chemical Lab
Odor	Chemical sensors
Keyboard Strokes	Keyboard
Comparison: Password	Keyboard

E un po' di mercato...



Antico Egitto e Biometria

2570 AC: La Grande Piramide di Giza (Cheope) è rimasta l'edificio più alto del mondo per circa 3800 anni. La Piramide di Cheope a Giza, anche detta Grande piramide, è la più grande delle tre piramidi (Chefren e Micerino). L'unica delle sette meraviglie del mondo antico ancora esistente. Dimensioni:

- Altezza: 146,6 metri (oggi 138).
- La base della piramide copre oltre 5 ettari di superficie, formando un quadrato di circa 230,34 metri per lato.
- Sono state scelte pietre di calcare, pesanti ognuna dagli 800 kg alle 4 tonnellate e rappresentano circa il 97% del materiale usato.
- Per le camere interne sono stati usati monoliti di granito, pesanti dalle 20 alle 80 tonnellate. Il peso totale si aggira intorno ai 7 milioni di tonnellate. Il volume totale è di circa 2.600 .000 m³.
- Numero di operai impiegati: 10.000 - 100.000.
- 1% di avvicendamento settimanale a causa di incidenti.
- Tempi di realizzazione: 20-40 anni.
- Operai dell'edilizia operanti in Italia (1992-2012) ≈ 40.000.

La domanda è come si distinguevano al momento del pagamento?

Principalmente fu utilizzato il concetto che l'ossatura umana non si modifica più a partire dai 20 anni, quindi ogni scheletro è diverso. Quindi le misure antropometriche utilizzate furono:

- Lunghezza braccio e avambraccio.
- Lunghezza coscia e gamba.
- Altezza.
- Proporzioni tra le precedenti.

Storia Antica, Moderna

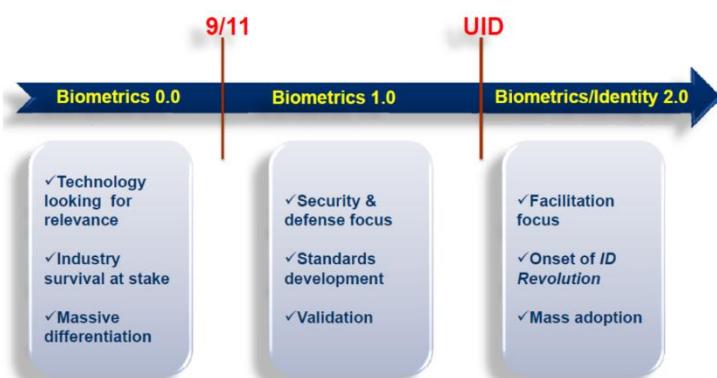
[2600 AC --- 1892 DC] = 4492 anni

- 2600 ac – In Egitto si misura la lunghezza delle braccia degli operai per distinguergli
- 500 ac – Gli Assiro-Babilonesi usano le impronte digitali sulle tavolette d'argilla per validare i contratti
- 1300 dc – in Cina si prendono le impronte delle mani e dei piedi dei bambini per distinguergli (Joao de Barros)
- 1686 dc – A Bologna, Marcello Malpighi scopre le caratteristiche uniche delle impronte digitali
- 1880 dc – Adolphe Bertillon propugna l'uso dell'antropometria. Il metodo detto Bertillonage sarà in uso fino a primi anni del secolo XX
- 1892 – Galton sviluppa il primo sistema di classificazione per impronte
- 1892 – Primo caso criminale risolto sulla base delle impronte digitali

[1900 DC --- 2011 DC] = 112 anni

- 1901 – Scotland Yard attiva il Metropolitan Police Fingerprint Bureau
- 1903-1905 – USA introducono fingerprint (Sistema Penitenziario/FBI)
- 1908 – ARMY\NAVY\Marines introducono fingerprint
- 1970-1980 – ARMY\NAVY\Marines introducono fingerprint\geometria della mano
- 1993 – INPASS
- 2001 – MIT Technology Review indica la biometria come una delle 10 tecnologie che cambieranno il mondo
- 2002 – BAT (Biometric Automated Toolset)
- 2004 – VISIT (Visitors and Immigrant Status Indicator Technology)
- 2006 – HIIDE (Handled Interagency Intelligent Detection Equipment)
- 2006 – BTF (Biometric Task Force)
- 2008 – FBI e NSA adottano un sistema rapido per IR
- 2011 – Individuazione di una fingerprint latente (9/11/2001)

.... e Contemporanea



L'utilizzo della biometria in un contesto ampio comporta una serie di problemi: utilizzato come strumento di riconoscimento e tracciamento di una persona, può intaccare alcuni aspetti etici e della privacy. Quindi quando si parla di biometria bisogna coniugare più aspetti:



Storia della Biometria

A. Bertillon

Bertillon, Louis Adolphe

[Paris, 1821 – Neuilly 1883]

French statistician and demographer; doctor; and cofounder of the school of Anthropology of Paris.

Alphonse Bertillon

[Paris, 1853 – id., 1914]

French anthropologist and criminologist

(Chief of criminal identification for the Paris Police) who devised the Bertillon system (1880)

A system formerly used for identifying persons by means of a detailed record of body measurements, physical description, and photographs (called metric frames and stereo metric photography).



[Also, A. Bertillon, as a forensic expert, and using a geometrical process and probability calculations, wrongly concluded that a handwritten document had been produced by A. Dreyfus...Rennes Trial Feb. 1899
<http://www.law-forensic.com/forerunners.htm>]

Bertillon fu il primo ad esasperare il concetto di classificazione già fondato dagli egiziani in precedenza.

Gli egiziani usavano l'altezza di una persona per identificarla durante il pagamento del salario. Il sistema Bertillon (1853-1914) creò un nuovo sistema per identificare i criminali

Principi

- L'ossatura umana non si modifica più partire dai 20 anni.
- Ogni scheletro è diverso.

Misure antropometriche

- Lunghezza braccio e dita.
- Altezza e larghezza testa.
- Lunghezza piedi.

Descrizione generale del corpo (proporzioni, armonia dei movimenti, etc.), caratteristiche mentali e morali

Peculiarità della pelle

- Nei, Porri, Cicatrici, Tataggi.

Bertillon creò una scheda per classificare i pregiudicati, dove essi venivano classificati sulla base di alcune caratteristiche biometriche descritte poc'anzi. Questa scheda riportava dati che avevano una certa affidabilità, benché nascessero da misurazioni piuttosto variabili dati gli strumenti utilizzati.

By dividing each of the measurements into small, medium and large groupings,

Bertillon could place the dimensions of any single person into one of the 243 distinct categories.

Further subdivisions by eye (7) and hair color provide 1,701 separate groupings.

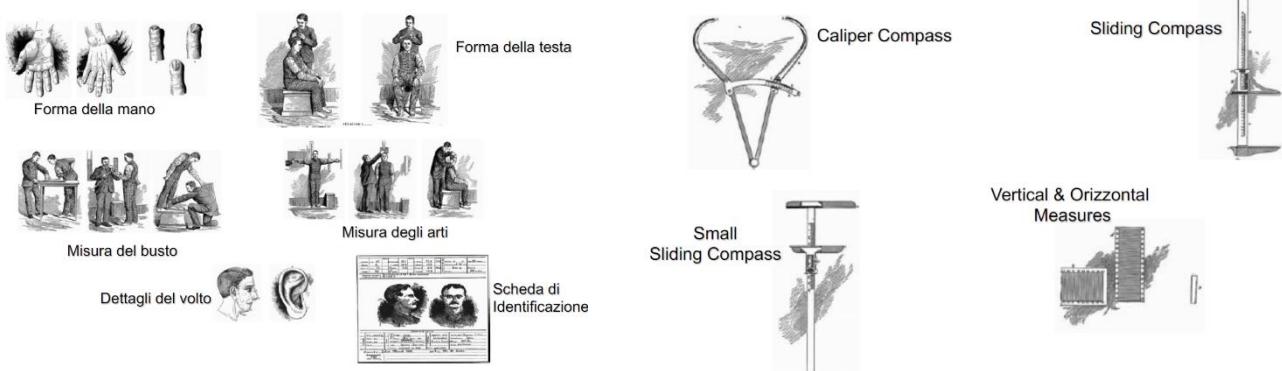
For example, in a file a 5,000 records, each of the primary categories would hold about 20 cards.

Therefore, it was not difficult to compare the new record to each of the other cards in the same category.



The first criminal identification card filed by the New York State Bertillon Bureau

Tale scheda era basata sulle biometriche seguenti:



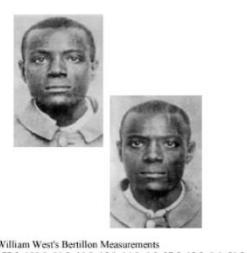
Il **Bertillionage** (sistema di classificazione) composto da queste schede che andavano a comporre un archivio, fu adottato in varie parti del mondo. Esso durò per qualche anno fin quando non si verificò un caso di omonimia, che portò questo sistema di classificazione al fallimento data la comprovata poca affidabilità. Il sistema fallì in quanto nel 1901 e nel 1903 furono registrati due detenuti William West e Will West con misure antropometriche simili.



The case of Will West

In 1903,

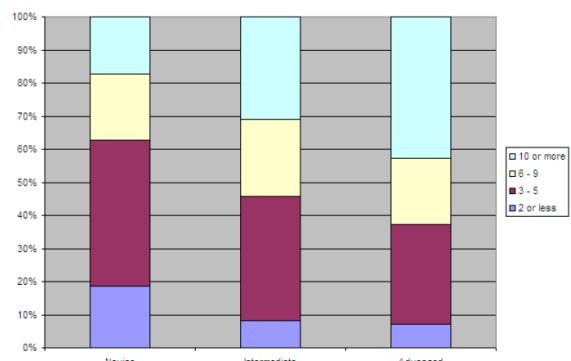
- Will West as a new prisoner;
- Will West's Bertillon measurements and photography done;
- Will West denied ever being incarcerated in the facility;
- Using Bertillon measurements, the system retrieves the Bertillon card for a William West;
- Will West continued to deny that the William West card was his;
- Subsequent investigation disclosed that William West was already incarcerated in the facility in 1901, and still a prisoner...



Il sistema di Bertillon fu sostituito, anche perché in quel periodo fu messo a punto un sistema di classificazione delle persone molto più affidabile, basato su caratteristiche biometriche delle impronte. Vennero scoperte nella struttura tridimensionale delle impronte che esistevano delle caratteristiche molto piccole non sempre visibili (le **minuzie**) che rendevano la possibilità di confusione tra due persone molto bassa.

Quante password hai?

La problematica che si affronta con le applicazioni biometriche è legato alla possibilità di sostituire o integrare i sistemi tradizionali di riconoscimento, basati su password. 6 persone su 10 hanno subito almeno un furto d'identità (clonazione carta di credito/bancomat, falsificazione CI, accesso ad aree riservate, accesso a informazioni riservate, etc.).



Proprietà di una caratteristica biometrica

- **Universalità:** quando si sceglie una biometria, ogni persona deve possedere tale biometria;
- **Distintività (Unicità):** due persone devono essere sufficientemente distinguibili in base a tale biometria;
- **Permanenza (Stabilità):** la biometria deve rimanere invariante (rispetto ai termini di confronto) nel tempo;
- **Misurabilità:** essa deve essere misurabile quantitativamente.
- **Performance (Efficacia/Efficienza):** l'acquisizione, l'estrazione delle caratteristiche e il confronto delle entità non devono essere troppo costose in termini di tempo, memoria ed efficienza del sistema;
- **Accettabilità:** la procedura di acquisizione deve essere tollerata da una vasta porzione della popolazione (utenti a cui viene sottoposta la biometrica);
- **Eludibilità:** la probabilità che il sistema venga raggiunto o indotto in errore deve essere minimizzata.

Nella tabella seguente vengono comparati secondo tre livelli le varie biometrie in relazione alle proprietà appena presentate.

Biometrics	Universality	Uniqueness	Permanence	Collectability	Performance	Acceptability	Circumvention
Face	H	L	M	H	L	H	L
Fingerprint	M	H	H	M	H	M	H
Hand Geometry	M	M	M	H	M	M	M
Keystroke Dynamics	L	L	L	M	L	M	M
Hand vein	M	M	M	M	M	M	H
Iris	H	H	H	M	H	L	H
Retina	H	H	M	L	H	L	H
Signature	L	L	L	H	L	H	L
Voice	M	L	L	M	L	H	L
Facial Thermogram	H	H	L	H	M	H	H
DNA	H	H	H	L	H	L	L

H=High, M=Medium, L=Low

Tipologia di Utente

Nella definizione della biometria più adatta ad un determinato contesto operativo, bisogna considerare se l'utente è:

- **Cooperativo:** è interesse dell'utente che il sistema lo riconosca (un impostore cerca di farsi riconoscere).
- **Non-Cooperativo:** l'utente è indifferente al processo di riconoscimento (un impostore cerca di non farsi riconoscere).

A seconda della tipologia di utente, cambia la tipologia di progettazione di un sistema biometrico.

Un altro aspetto che riguarda l'utente è:

- **Abituato/non abituato:** frequenza con cui gli utenti del sistema si prestano al riconoscimento biometrico (più volte al giorno, settimanalmente, mensilmente, ecc.).
- **Consapevole/inconsapevole:** se l'utente è o no al corrente del processo di riconoscimento in atto.

Tipologia di Applicazione

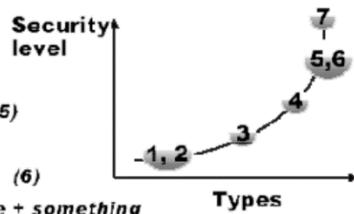
Altra possibile considerazione riguarda l'effettuare il riconoscimento con una procedura:

- **On-line:** l'utente è in attesa finché il sistema non risponde (tempi brevi di risposta).
- **Off-line:** l'utente viene acquisito ma non attende una risposta immediata.

Metodi di Autenticazione

Seven types of authentication:

- ✓ **Something you know** (1)
 - ✓ e.g. PIN code, mother's maiden name, birthday
- ✓ **Something you have** (2)
 - ✓ e.g. Card, key
- ✓ **Something you know + something you have** (3)
 - ✓ e.g. ATM card + PIN
- ✓ **Something you are – Biometrics** (4)
 - ✓ no PIN to remember, no PIN to forget
- ✓ **Something you have + something you are** (5)
 - ✓ Smart Card
- ✓ **Something you know + something you are** (6)
- ✓ **Something you know + something you have + something you are** (7)



Misura delle Prestazioni

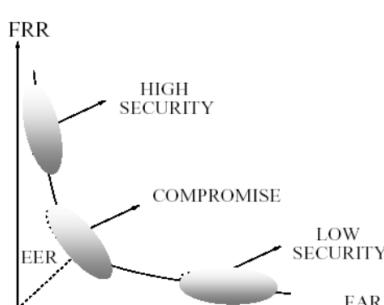
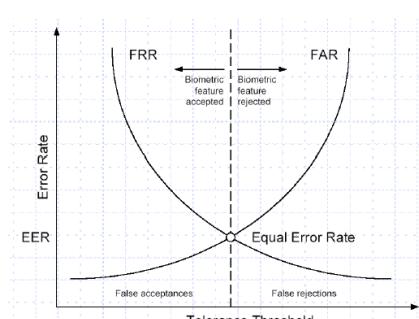
Le prestazioni di un sistema biometrico possono essere misurate con diverse modalità, innanzitutto comprendendo il comportamento di tale sistema biometrico. Tendenzialmente ci possono essere quattro possibili casi:

- L'identità **corrisponde** ed il soggetto è **accettato**.
- L'identità **corrisponde** ma il soggetto è **respinto** (errore).
- Un **impostore** è **accettato** (errore).
- Un **impostore** è **rifiutato**.

Non è possibile riconoscere un individuo senza ammettere una **tolleranza d'errore**. La soglia di tolleranza è cruciale e dipende fortemente dall'applicazione. Una soglia troppo bassa causa molti errori di **Tipo I** (rifiuto di soggetti registrati – **FRR** (False Recognition Rate)). Una soglia troppo alta causa molti errori di **Tipo II** (accettazione di impostori) – **FAR** (False Acceptance Rate). $FAR = FRR$.

La scelta più utilizzata è **Equal Error Rate (EER)**.

Si pensi ad un sistema biometrico dove bisogna stabilire la giusta soglia di tolleranza. In figura sono costruiti dei grafici, dove sull'asse delle ascisse viene posta la soglia di tolleranza, e sull'asse delle ordinate viene posto l'errore commesso (essi dipendono dalla soglia di tolleranza). In figura viene mostrato FAR che, quando la soglia di tolleranza risulta essere piccola, l'errore commesso è molto basso, ma man mano che la soglia di tolleranza aumenta, ammettendo un errore maggiore, la probabilità che un impostore possiede nel farsi accettare cresce in maniera esponenziale. L'FRR si comporta in maniera opposta e cioè quando la soglia di tolleranza è molto bassa, allora la probabilità che l'utente reale venga riconosciuto, è più alta. Tale probabilità scende man mano che la soglia di tolleranza tende ad aumentare.



Quando si progetta un sistema biometrico, bisogna individuare la soglia ottimale (EER) che incrocia le due curve FRR e FAR. In questa logica si fa lavorare per un certo periodo il sistema biometrico facendo variare la soglia di tolleranza e gli errori mappando le due curve, ed infine si raggiunge all'intersezione delle due, che rappresenta un giusto compromesso. Spesso nella progettazione di un sistema biometrico si tende a privilegiare un sistema biometrico che commetta un numero alto di errori FRR a patto di minimizzare gli errori FAR. Si preferisce avere una soglia di tolleranza molto bassa che porterà ad avere un certo numero di soggetti reali, rifiutati. Ma che limita molto l'accesso del numero di impostori.

Variabilità delle Biometrie

I dati biometrici a differenza delle tradizionali metodologie di autenticazione non possono essere persi, trasferiti ad altri o dimenticati. L'acquisizione (digitalizzazione dei dati) può determinare una variabilità delle biometrie:

- **Intra_Classe**, cioè la variabilità delle biometrie per un individuo.
- **Inter_Classe**, cioè la similarità delle biometrie tra individui diversi.

Entrambe le variabilità rappresentano elementi di disturbo per il riconoscimento dell'individuo attraverso tecniche biometriche.

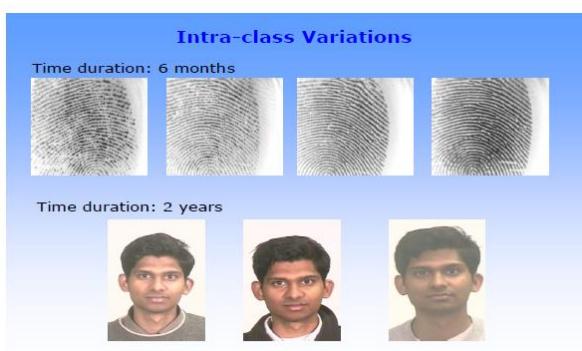
La variabilità Intra_Classe può essere resa variabile da:

- Il medio/breve termine. Alcune biometriche sono condizionate dal trascorrere del tempo.
- Il lungo termine: l'età.
- La posa.
- Le parziali occlusioni.
- L'espressione.
- L'illuminazione.

La variabilità Inter_Classe, invece, è legata alla presenza di:

- Gemelli.
- Sosia.
- Ecc...

E quindi la similitudine che ci può essere tra soggetti diversi.



Un esempio di un soggetto, tendenzialmente giovane, con le proprie impronte digitali. Il volto subisce delle trasformazioni piuttosto limitate in un arco di tempo di due anni. Le impronte digitali invece rimangono intatte.



La variabilità Intra_Classe, legata anche alla posa, dove un soggetto che cambia la posa può rendere più difficile la propria autenticazione.



La variazione di espressione può determinare trasformazioni del volto che superano la tolleranza imposta su un sistema di riconoscimento biometrico.



Espressione, Posa e Parziali Occlusioni (occhiali).



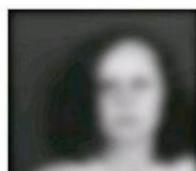


Con il variare dell'illuminazione il processo di identificazione potrebbe diventare molto ostico, pur essendo il soggetto in una posa corretta.

La similarità Inter_Classe è legata ai rapporti di parentela



Un piccolo esempio di riconoscimento in situazioni particolarmente stressanti.



Count of Q2	Q2											Total
Gender	A 1	A 2	A 3	A 4	A 5	A 6	A 7	A 8	A 9	A 10	d	Total
Female	8%	3%	4%	4%	16%	8%	0%	4%	2%	51%	100%	
Male	9%	5%	3%	6%	6%	5%	1%	2%	1%	63%	100%	
Grand Total	9%	4%	3%	5%	9%	6%	1%	3%	1%	59%	100%	

This is a smoothed picture (i.e. no detail information).

Try to find this person among the pictures below.

Choose one of the checkboxes under the pictures



1.

2.

3.

4.

5.



6.

7.

8.

9.

10.

	Match to Name? _____	<input type="checkbox"/> Very well _____ years ago <input type="checkbox"/> Somewhat _____ years ago <input type="checkbox"/> Not at all _____ years ago or <input type="checkbox"/> never		<input type="checkbox"/> Recognize?		<input type="checkbox"/> Recognize?
	Match to Name? _____	<input type="checkbox"/> Very well _____ years ago <input type="checkbox"/> Somewhat _____ years ago <input type="checkbox"/> Not at all _____ years ago or <input type="checkbox"/> never		<input type="checkbox"/> Recognize?		<input type="checkbox"/> Recognize?
	Match to Name? _____	<input type="checkbox"/> Very well _____ years ago <input type="checkbox"/> Somewhat _____ years ago <input type="checkbox"/> Not at all _____ years ago or <input type="checkbox"/> never		<input type="checkbox"/> Recognize?		<input type="checkbox"/> Recognize?
	Match to Name? _____	<input type="checkbox"/> Very well _____ years ago <input type="checkbox"/> Somewhat _____ years ago <input type="checkbox"/> Not at all _____ years ago or <input type="checkbox"/> never		<input type="checkbox"/> Recognize?		<input type="checkbox"/> Recognize?
	Match to Name? _____	<input type="checkbox"/> Very well _____ years ago <input type="checkbox"/> Somewhat _____ years ago <input type="checkbox"/> Not at all _____ years ago or <input type="checkbox"/> never		<input type="checkbox"/> Recognize?		<input type="checkbox"/> Recognize?

In questo esempio vi sono 5 persone fotografate a distanza di diversi anni, mischiate con altre 5 persone. L'effetto dell'età è fondamentale sulle modifiche del volto, la quale risulta poco stabile nel tempo.

Image Filtering

Si mostreranno in seguito le prime trasformazioni che si possono eseguire su di una immagine. In generale una trasformazione dell'immagine è caratterizzata nel seguente modo:

$$g(x, y) = T[f(x, y)]$$

dove:

- f è l'immagine input
- g è l'immagine elaborata
- T è l'operatore di trasformazione applicata ai singoli pixel dell'immagine.

L'operatore che si applica all'immagine può avere diverse caratteristiche matematiche. In genere si possono definire degli operatori lineari se godono di determinate proprietà come:

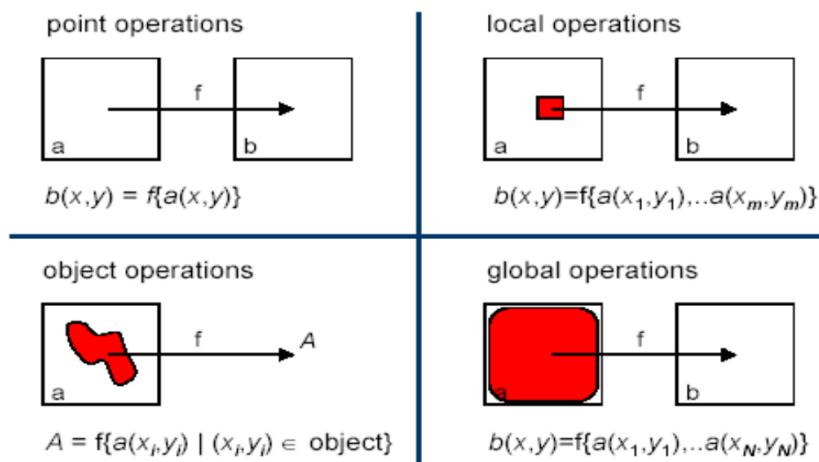
- **Omogeneo** se $T\{af[n]\} = aT\{f[n]\}$
- **Associativo** se $T\{f_1[n] + f_2[n]\} = T\{f_1[n]\} + T\{f_2[n]\}$

Il fatto che un operatore possa essere lineare o meno non ha un impatto sulla tipologia di trasformazione che si va ad applicare.

Innanzitutto, va fatta una classificazione delle trasformazioni T che possono essere applicate ad una determinata immagine/video:

- Tecniche operanti nel **dominio spaziale**, ovvero dove vi è una manipolazione diretta dei pixel.
- Tecniche operanti nel **dominio delle frequenze** (trasformata di Fourier), ovvero vi è una trasformazione dell'immagine nel dominio delle frequenze e conseguente manipolazione di esse. Tali frequenze non agiscono sui pixel ma agiscono su una trasformazione dei pixel su un concetto diverso. Come si può intuire, tali tecniche implicano un costo computazionale aggiuntivo compensato da una qualità della trasformazione dell'immagine migliore.
- **Tecniche ibride**, ovvero una combinazione di varie tecniche.

Classificazione delle Tecniche



L'altro elemento che si può aggiungere nell'ottica della classificazione delle tecniche, è quello che permette di dividere le tecniche di trasformazione in 4 famiglie.

- **Point Operations:** le tecniche che rientrano in questa classe risultano essere molto semplici, ovvero trasformazioni dove il pixel di destinazione, nella trasformazione, dipende solo dal pixel di partenza.
- **Local Operations:** hanno una struttura più aggregata, dove il pixel di destinazione non dipende solo da un pixel di partenza, ma anche da un suo intorno quadrato.
- **Object Operations:** il pixel di destinazione dipende dal pixel di partenza e da un oggetto che ha un suo significato semantico, non di forma regolare come nel caso della Local Operations.
- **Global Operations:** ogni pixel dell'immagine di destinazione, dipende da tutti i pixel dell'immagine di partenza.

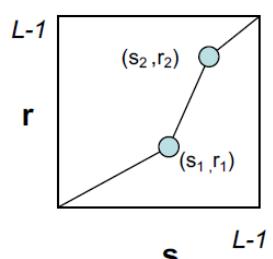
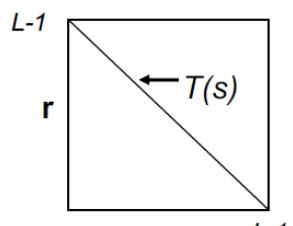
Tecniche operanti nel dominio spaziale

Si analizza ora qualche semplice trasformazione: partendo con il Point Processing, dove ogni pixel di destinazione dipende solo dal corrispettivo pixel di partenza. In questa rappresentazione grafica, sull'asse delle ascisse si hanno gli input gray level (livelli di grigio dell'immagine di partenza), r invece rappresenta il valore output e T effettua la trasformazione. Per ogni pixel situato su s , T ne ottiene una trasformazione che lo porta ad assumere un nuovo valore lungo la retta $T(s)$.

Questa trasformazione che si sta effettuando, $T(s)$, è una trasformazione monotona decrescente, che prende il nome di **Negativo**.

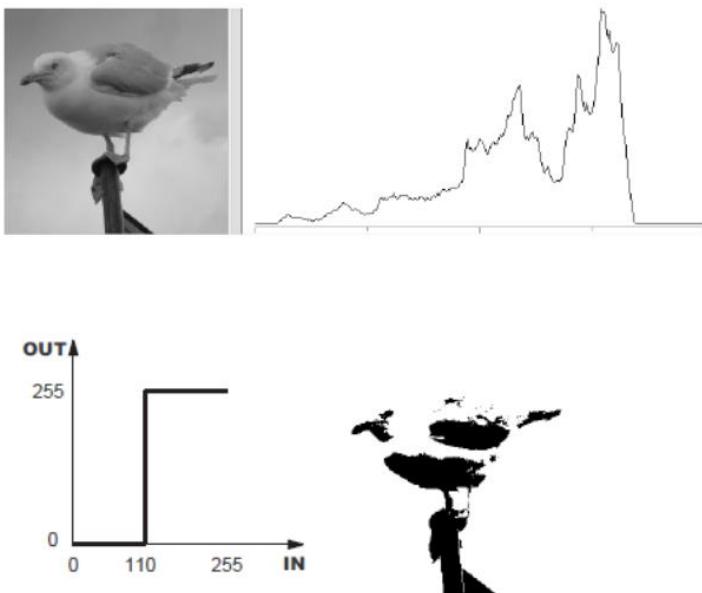
È possibile appore anche altre operazioni come la **Binarizzazione** (T è monotona crescente) di una immagine. In essa i valori ottenuti in fase di digitalizzazione, mappati sull'asse delle ascisse, sono valori che devono essere distribuiti soltanto su due colori, ovvero il bianco e il nero. Ottenendo un'immagine di output, appunto, bianco e nera.

Oltre alla binarizzazione, agendo sui due cerchi (a destra) è possibile anche effettuare il **Contrast Stretching** dell'immagine, operando direttamente sull'istogramma per migliorare il contrasto dell'immagine in questione.



Binarizzazione

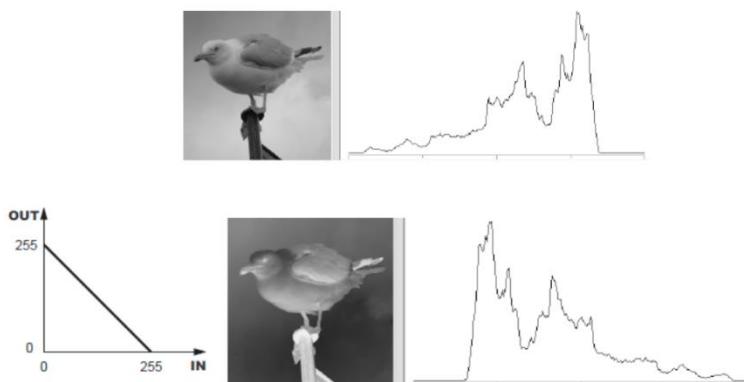
Questa trasformazione individua una soglia in 110 (variabile), dove tutti i colori da 0 fino a 110 diventeranno nero, e tutti i colori che partono da 110 fino a 255 diventeranno bianco. Ottenendo la stessa immagine di partenza, ma binarizzata.



Negativo

Vi è un'immagine con il proprio istogramma. Quest'ultimo, si ricordi, è rappresentato da un'asse delle ascisse

dove vi sono tutti i livelli di grigio frutto della quantizzazione, e sull'asse delle ordinate vi è il numero di occorrenze per ognuno dei livelli di grigio. Produrre il negativo di tale immagine, significa associare ad ogni pixel il livello di grigio complementare. Infatti, tra le due immagini intercorre una differenza specifica: i pixel chiari sono diventati più scuri e viceversa. La trasformazione quindi in presenza di colori scuri si occupa di tramutarli in colori chiari.



Progressivamente, sull'asse delle ascisse, i colori diventando più chiari, e con la progressione della funzione T , essa tende a renderli più scuri ottenendo il negativo dell'immagine. L'istogramma risultante è ad immagine speculare di quello di partenza (ribaltato rispetto al valore mediano).

Andando avanti ci si concentra sulle Local Operations, dove la trasformazione T è definita in un intorno predefinito di (x, y) .

Con i filtri spaziali (mask processing) è possibile effettuare varie operazioni. Queste trasformazioni sono definite dei filtri spaziali. Si possono individuare tre categorie di filtri che si andranno ad applicare alle immagini:

- **Lowpass** (passa basso): attenua o elimina le alte frequenze, ovvero contorni e dettagli.
- **Highpass** (passa alto): attenua o elimina le basse frequenze, come contrasto e intensità.
- **Bandpass** (passa banda): attenua frequenze in una banda predefinita.

Filtri Spaziali

Questi filtri rientrano nella categoria dei Local Operations dove ogni pixel di destinazione dipende dal pixel di partenza più il suo intorno. Per essere costruiti, in genere vengono create delle mascherine quadrate,

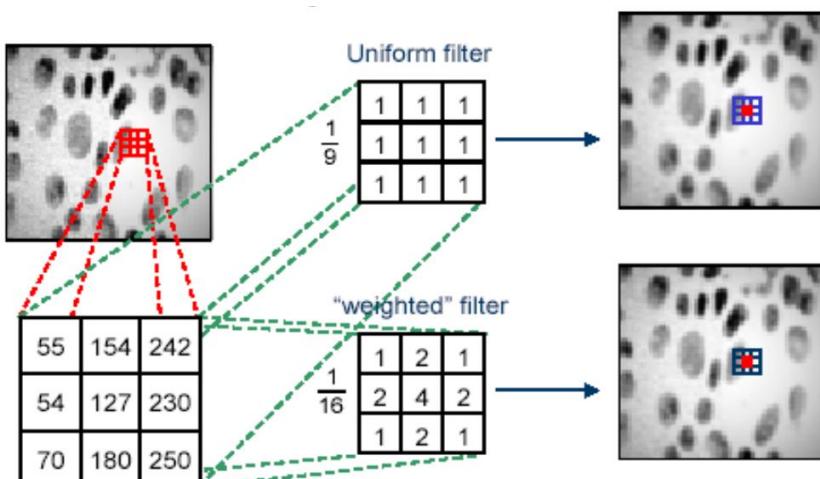
W_1	W_2	W_3
W_4	W_5	W_6
W_7	W_8	W_9

generalmente di dimensioni dispari 3×3 , 5×5 ecc... e dentro tali mascherine verranno posti dei pesi, che chiaramente andranno a determinare la funzionalità del filtro. Sulla base dei valori dei valori $W_1 \dots W_9$, distribuiti all'interno della mascherina si otterrà un risultato diverso. Una volta costruita, essa si sovrapporrà all'immagine centrandola nel pixel (x, y) che avrà come corrispondente il peso W_5 e si effettuerà la moltiplicazione tra ciascuno di questi pesi con il corrispondente valore di luminosità presente nell'immagine:

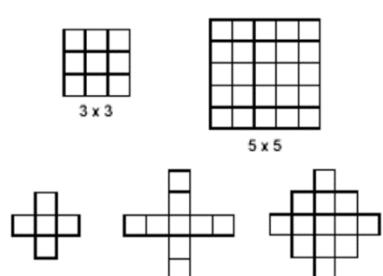
$$R = W_1 Z_1 + W_2 Z_2 + W_3 Z_3 + W_4 Z_4 + W_5 Z_5 + W_6 Z_6 + W_7 Z_7 + W_8 Z_8 + W_9 Z_9$$

dove z_1, \dots, z_9 rappresentano i pixel dell'immagine ai quali è sovrapposto il filtro.

Il valore R sarà quello proiettato nell'immagine di destinazione nella posizione (x, y) .



Da come si può notare nella figura a sinistra, nell'immagine viene presa una zona specifica che rappresenta vari pixel. Si applica una maschera composta da tutti 1 ed un fattore di normalizzazione di $1/9$. La maschera con tutti 1 sarà sovrapposta a quella dell'immagine iniziale e dopo i vari calcoli, si otterrà la media della zona in questione. Il pixel in questione sarebbe la media della porzione dell'immagine considerata.



Questa prima classe di filtri verrà utilizzata per costruire i cosiddetti **Smoothing Filter**, cioè i filtri utilizzati nelle operazioni di preprocessing. Infatti, può capitare che l'immagine contenga del “rumore” oppure sia troppo ricca di dettagli. Con tali filtri è possibile intervenire, anche se essi possono produrre un effetto di **Blurring** se l'immagine è troppo ricca di dettagli, e quindi se ne vogliono rimuovere alcuni di essi per rendere l'immagine più funzionale per l'estrazione di oggetti. Oppure l'immagine presenta del rumore che può esser stato introdotto o nella fase di digitalizzazione oppure durante la fase di trasmissione. Questi filtri ricadono nella categoria dei filtri Lowpass (bassa passo) ovvero i filtri che lasciano transitare solo le informazioni che risiedono nelle basse frequenze come contrasto e intensità, i filtri Lowpass si dividono in **lineari** e **non lineari**. Il primo filtro che verrà affrontato è il Filtro **Media** (aritmetica) il quale fa parte dei filtri lineari. I filtri non lineari invece, sono il Filtro **Mediana**, Filtro Max e Filtro Min.

Filtri Spaziali: La Media

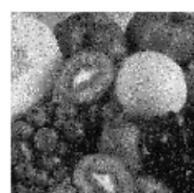
Tale filtro non differisce dal concetto di media aritmetica. Quindi ogni pixel dell'immagine destinazione viene sostituito con la **media dei pixel** appartenenti ad un'area ben definita dell'immagine. Piuttosto che prendere soltanto il pixel (x, y) nell'immagine input, viene preso la media di tutti i pixel connessi con il pixel (x, y) . Effettuando la media (con la divisione per 9) ottenendo il filtro di destinazione.

- S= insieme di pixel connessi con (x,y)
- M= numero di pixel coinvolti ($|S|+1$)

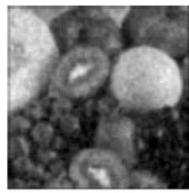
$$g(x,y) = \frac{1}{M} \sum_{f(n,m) \in S} f(n,m)$$

$$\frac{1}{9} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

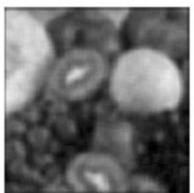
L'applicazione di un filtro media ad un'immagine/video che presenta rumore, si veda l'esempio seguente, viene effettuato proprio per eliminare tale rumore prodotto da un determinato motivo, per ricondurre



Salt & Pepper
Noise



3 X 3 Average



5 X 5 Average



7 X 7 Average

l'immagine ad una qualità migliore. Applicando un filtro media 3×3 si inizia col notare che parte del rumore e dei dettagli sono scomparsi, sfumando l'immagine. Utilizzando un filtro più di dimensioni maggiori, quando si esegue la media di un intorno più grande, in output vi sarà una eliminazione del rumore e dei dettagli. Il progressivo aumento della dimensione del filtro, fa sì che il rumore venisse diminuito, ma ha comportato anche un effetto di eliminazione delle informazioni più significative presenti nell'immagine come contorni e dettagli. Più cresce il filtro più aumenta l'impatto delle

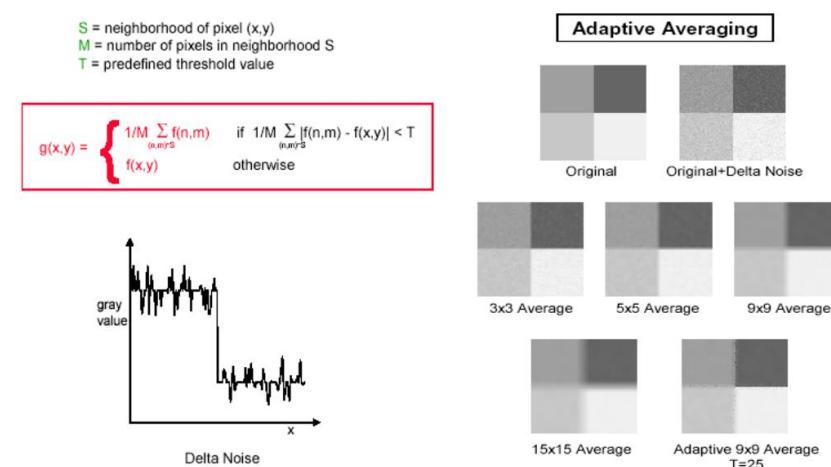
informazioni eliminate.

Il **contorno** di un oggetto è una **transizione**, tra un oggetto e un altro, come una sorta di interruzione.



Filtri Spaziali: La Media Adattiva

È possibile modificare il filtro media per ottenere un risultato qualitativamente migliore. La media adattiva va a risolvere varie problematiche della media, come la confusione dei contorni. Ogni qual volta essa viene utilizzata effettua un confronto tra il pixel (x, y) e tutti i quanti i pixel che sono nel suo intorno. In presenza di piccole variazioni si applica la media, ma in presenza di grandi variazioni si lasciano invariati i pixel di riferimento. Essa **lavora sui bordi**.



Applicando un filtro media sempre più grande, l'immagine tende a sfocarsi e perdere rumore ma anche dettagli. Quindi la media adattiva può essere una soluzione che consente di preservare in parte dettagli e i contorni degli oggetti avendo un buon effetto sulla rimozione del rumore.

Filtri Spaziali: La Mediana

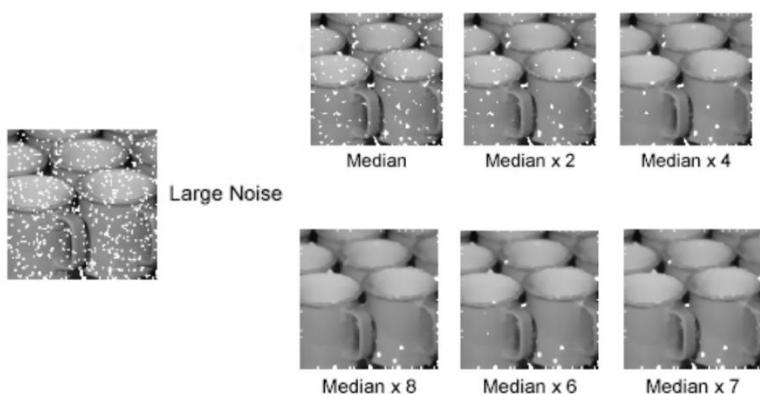
$$g(x,y) = \text{mediana}\{f(x,y)\}_{f(x,y) \in S}$$

30	10	20
10	250	25
20	25	30

→ 10, 10, 20, 20, 25, 25, 30, 30, 250
median

Un'altra possibile soluzione che può essere utilizzata per rimuovere il rumore è il Filtro Mediana. Esso corrisponde esattamente al concetto di mediana statistica per il quale si considerano i pixel di una regione, li si ordina in maniera crescente e poi ci si posiziona sul valore centrale. La mediana è proprio tale valore centrale. Rispetto alla media, la mediana **non viene influenzata dalle code**, ovvero il valore

centrale non viene condizionato particolarmente dalla presenza di significativi cambiamenti agli estremi, cosa che avviene nella media. Quando viene applicata la mediana ad un'immagine con rumore, si ottengono i seguenti risultati.

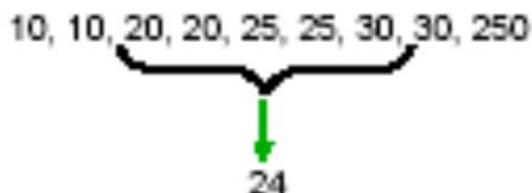


Come si può notare, invece che aumentare la dimensione del filtro mediana, si reitera lo stesso filtro. Si parte sempre da filtri 3x3. Quindi si è prese tale filtro in cui si è calcolata la mediana e lo si è applicato all'immagine. Successivamente lo si applica più volte, così facendo si determina un filtraggio progressivo ma utilizzando sempre lo stesso filtro, mantenendo la stessa dimensione. Si può notare che

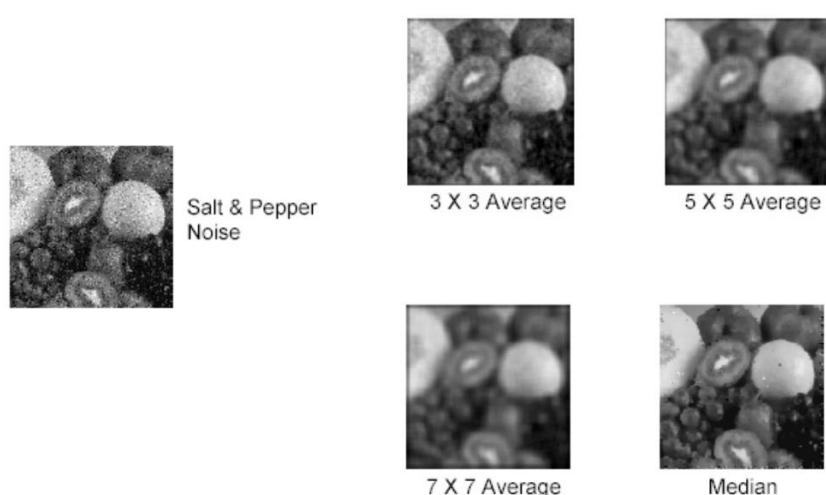
all'immagine, pur essendoci alterazione sui bordi, l'effetto Blurring è molto meno evidente rispetto alla media tradizionale.

Alle volte è possibile combinare media e mediana, ottenendo dei **filtri ibridi**.

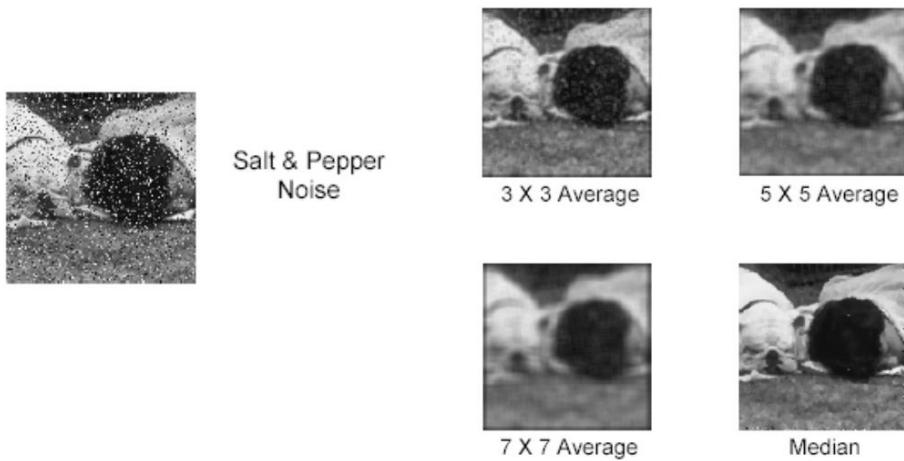
Median + Average: average the k central values.



Media vs Mediana



Si parte sempre da una immagine con rumore "sale e pepe". E poi si presenta la stessa immagine con i filtri media (normale) 3x3, 5x5 e 7x7, e infine una mediana 3x3. In generale nessuno dei due filtri ha lavorato in modo ottimale, dato anche il molto rumore distribuito nell'immagine di partenza ma la mediana risulta dare risultati più soddisfacenti.

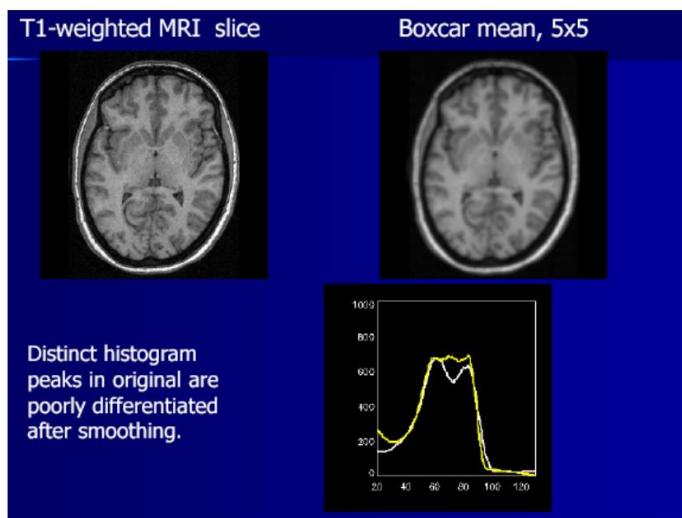


Anche in questo caso il filtro mediana sembra produrre dei risultati migliori rispetto al filtro media.

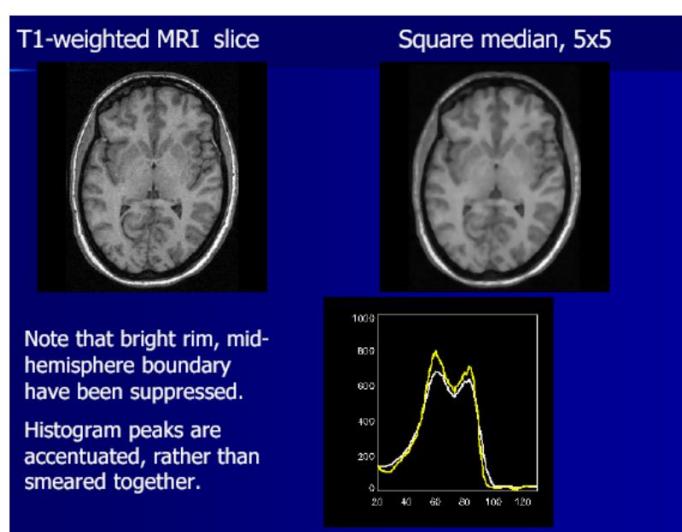
Riassuntivamente:

La Media presenta una **eccessiva velatura (Blurring)** al crescere della maschera e delle iterazioni con conseguente perdita dei dettagli. Essa però **preserva** con la versione adattive le **forme** (scalini).

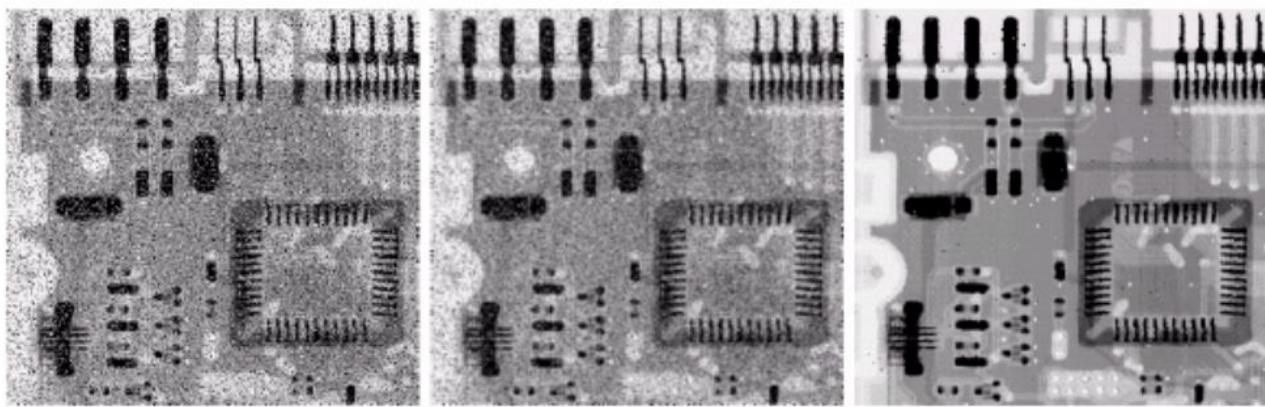
La Mediana ha una **riduzione del rumore (noise reduction)** senza eccessiva perdita di dettagli (**Blurring limitato**). Non preserva le forme (scalini).



Un uso particolare di questi filtri lo si evince nell'ambito medico, precisamente nelle risonanze magnetiche. Il rumore indotto nell'esame finale (movimenti del paziente) tende ad essere eliminato con dei filtri di preprocessing per renderla migliore dopo l'acquisizione. In figura l'istogramma delle due immagini, dove essi sono molto simili; ciò fa capire che non ci sono state grosse trasformazioni all'interno dell'immagine.



In questo caso viene applicato un Filtro Mediana. Si può notare che la mediana risulta essere leggermente migliore rispetto alla media precedente. Anche in questo caso i due istogrammi sono molto simili, ciò significa che non sono state introdotte grosse trasformazioni.



a b c

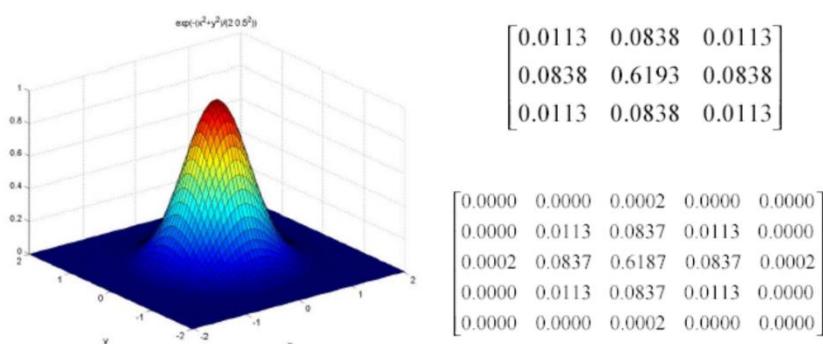
FIGURE 3.37 (a) X-ray image of circuit board corrupted by salt-and-pepper noise. (b) Noise reduction with a 3×3 averaging mask. (c) Noise reduction with a 3×3 median filter. (Original image courtesy of Mr. Joseph E. Pascente, Lixi, Inc.)

In questo caso la mediana riesce ad ottenere ottimi risultati rispetto alla media tradizionale.

Smoothing Gaussiano

Il funzionamento del Filtro Gaussiano è simile a quello del filtro mediana con la differenza che il contributo di ciascun **pixel limitrofo ha un peso diverso**, definito dalla distribuzione spaziale della gaussiana.

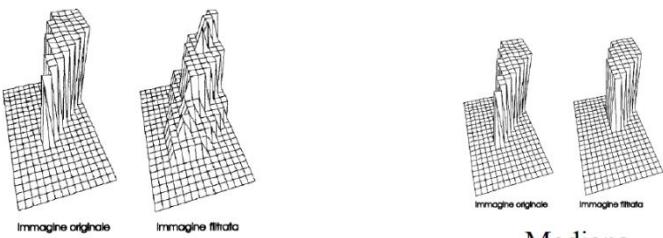
Nel campo dell'elaborazione delle immagini la gaussiana viene troncata limitandone l'estensione ad una zona di dimensioni $N * N$ (i valori più frequenti per N sono 3, 4 e 5).



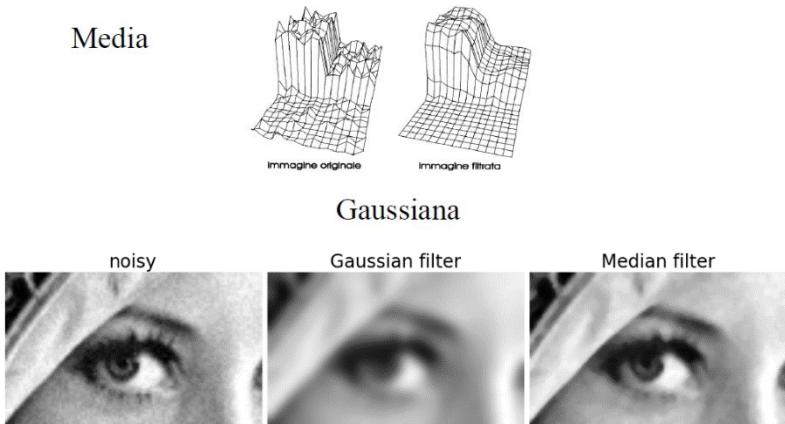
progressivamente diminuire il peso dei pixel man mano che ci si allontana dal centro. La somma di tutti i pesi sarà **sempre uguale a uno**. In prossimità del pixel di riferimento, si assegnerà un valore maggiore.

Quindi possono essere costruiti nuovi filtri che, fondamentalmente, sono dei Filtri Media pesati. Mentre prima il peso assegnato dalla media era uguale per tutti quanti, ora si adopera una selezione: più ci si è vicini alla posizione (x, y) maggiore sarà il peso in quanto maggiore sarà la probabilità che quel pixel appartenga alla regione di interesse; più ci si allontana dal pixel (x, y) maggiore sarà la probabilità di non appartenere alla regione di interesse e quindi si cerca di contaminare il meno possibile la trasformazione.

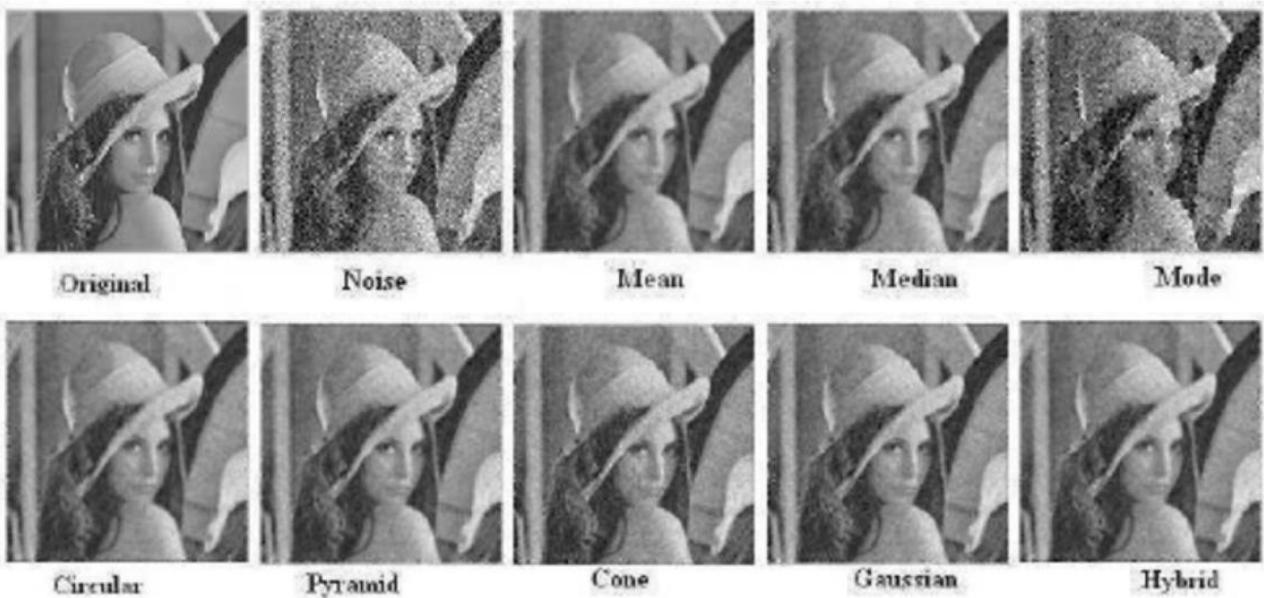
Si parte dal presupposto che i pixel in prossimità della zona centrale, sono quelli più importanti. Quindi se si vuol posizionare la maschera su un pixel (x, y) , non si deve assegnare poi lo stesso peso a tutti gli altri pixel ma si cerca di concentrare la maggior parte del peso in prossimità del pixel (x, y) , e



Con un Filtro gaussiano si riesce ad eliminare le zone di rumore e a preservare gli scalini, proprio perché in prossimità di essi i pixel otterranno una distribuzione di pesi differenziata.

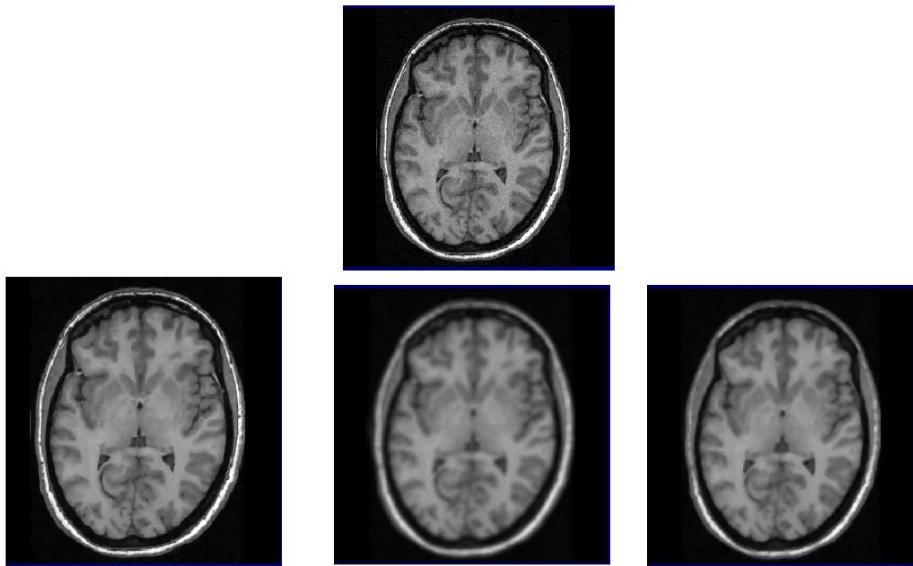


Operando con maschere 5x5, nel filtro gaussiano l'effetto Blurring è molto evidente. Nel filtro mediana l'effetto Blurring è meno evidente, ma il rumore risulta essere più consistente.



Un'immagine e i relativi output di vari filtri. L'immagine di partenza risulta essere già compromessa, e con l'aggiunta di rumore, produce un risultato non proprio positivo sull'applicazione di filtri di questo tipo.

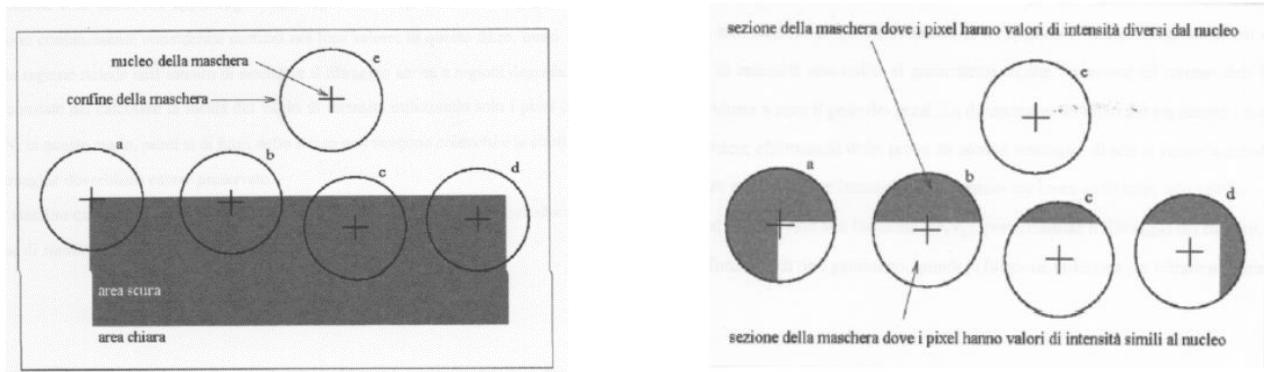
Originale, Gaussiano,Media,Mediana



Come si nota con un Filtro Gaussiano si riscontrano i migliori risultati rispetto agli altri filtri Media e Mediana.

SUSAN: Smoothing Segment Univalue Assimilating Values

Tale filtro esaspera il concetto Gaussiano. Infatti, permane l'idea di Filtro Gaussiano dove però i pesi non vengono stabiliti a priori come fatto in precedenza, ma in questo caso i pesi vengono calcolati **run-time**. Quindi viene applicata la maschera all'immagine e si va a controllare quali pixel ricadono nella stessa area del pixel centrale. Si assegneranno pesi alti a tutti i pixel che appartengono all'aerea del pixel centrale e si tenderà a dare peso minore ai pixel che si trovano più in lontananza rispetto al centro.



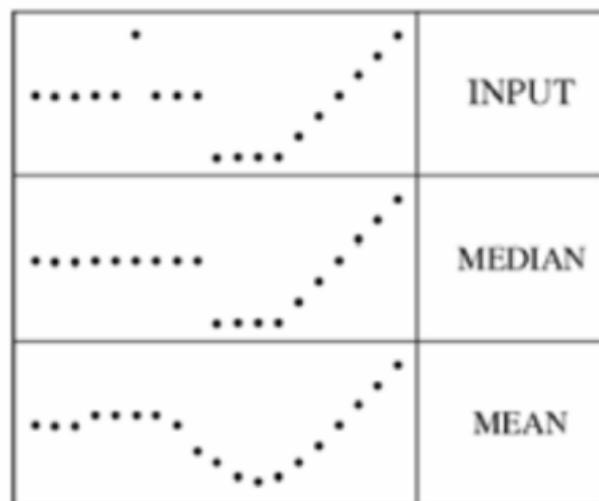
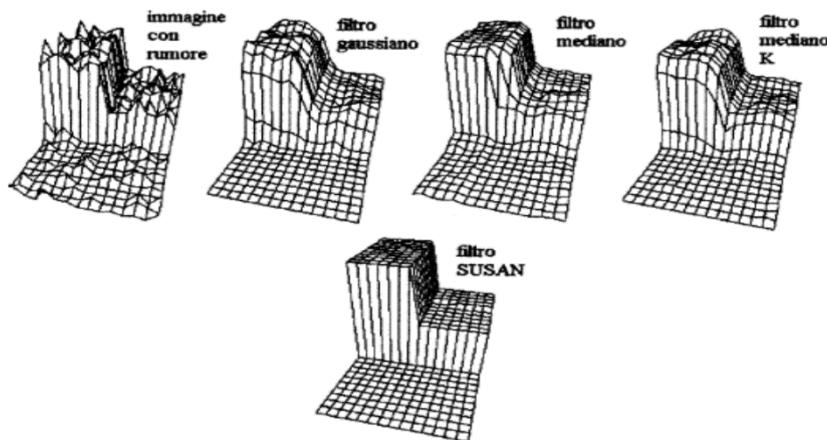
Questo calcolo lo si effettua per mezzo della formula seguente:

$$c(r, r_0) = e^{-\left(\frac{I(r) - I(r_0)}{t}\right)^2}$$

in cui $I(r)$ è il valore di intensità dei pixel all'interno della maschera. $I(r_0)$ quello del nucleo, t è un valore di soglia sull'intensità e $c(r, r_0)$ rappresenta il peso che il pixel in esame ha nel calcolo della media.

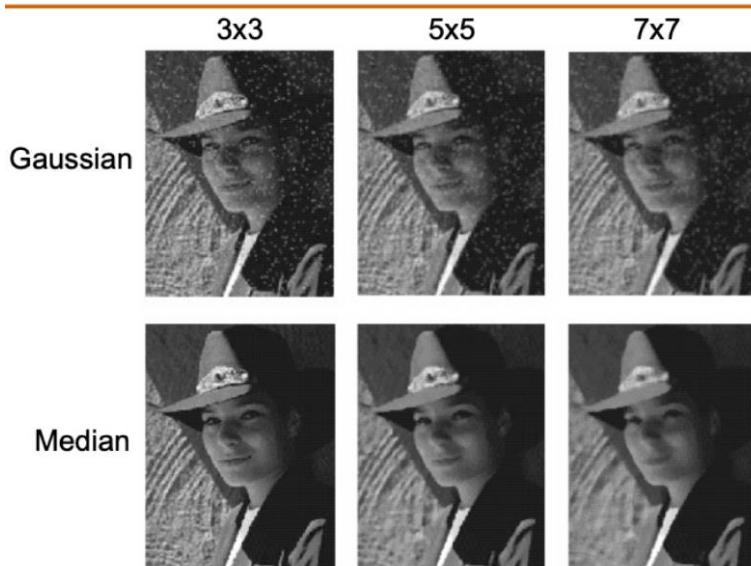
Oltre al peso dovuto alla funzione $c(r, r_0)$, per effettuare il filtraggio del rumore, si utilizza una funzione di tipo gaussiano. Se l'area dell'USAN è **nulla** il valore dell'intensità del pixel in esame è calcolato applicando il filtro mediano descritto precedentemente. Quindi si avrà ancora un filtro gaussiano, ma sarà una maschera in cui i pesi saranno calcolati di volta in volta sulla base dell'immagine. Il **vantaggio** di questo filtro è che i pesi dipenderanno dall'immagine e quindi non sarà un filtro statico costruito in partenza. Il **disavvantaggio** di questo filtro è la grande onerosità in termini computazionali. Nell'immagine seguente viene vista una rappresentazione

tridimensionale dei vari filtri applicati a delle immagini. Il filtro SUSAN **preserva gli scalini**, risultando essere un filtro ottimale fornendo output migliori.



Anche il filtro mediano garantisce tali aspetti, soprattutto in presenza di scalini. Uno dei problemi fondamentali è quella di preservare la distanza tra due oggetti, cosa che il filtro mediano riesce ad ottenere a differenza del filtro media. Quest'ultima non presenta un distacco tra un oggetto ed un altro, causando l'effetto di velatura molto evidente all'interno delle immagini e che induce una perdita dei contrasti.

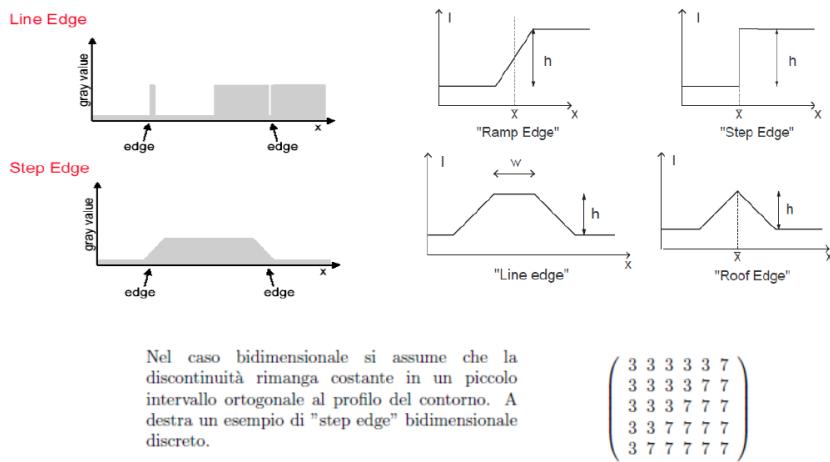
Median vs. Gaussian filtering



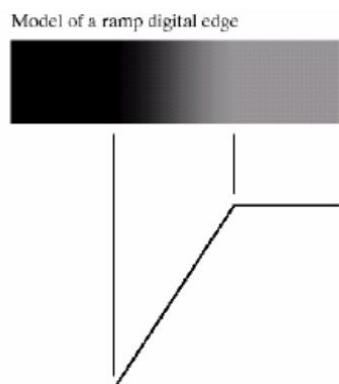
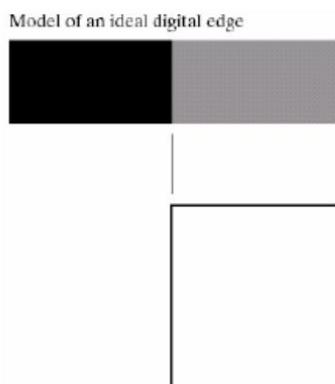
Edge Detection

Si introduce un'altra classe di filtri che hanno come obiettivo quello di trattenere le informazioni che risiedono nelle alte frequenze, lavorando in modo esattamente speculare a quelli precedenti. Tali filtri quindi trattengono i punti di controllo, ovvero i **dettagli**. Gli **Edge Detector** sono filtri che hanno la capacità di individuare le aree di transizione tra regioni omogenee:

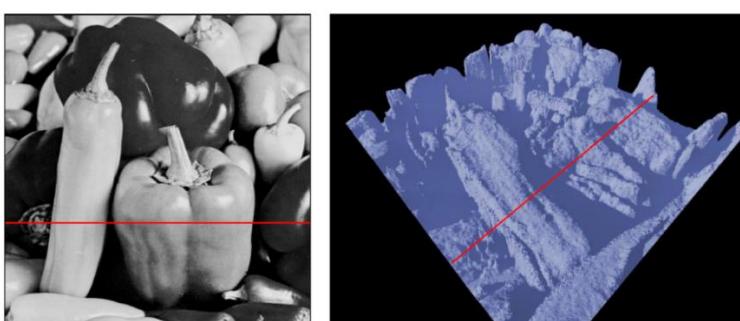
- **Punti isolati.**
- **Linee** (orizzontali, verticali, oblique).
- **Contorni** (insieme di punti con inizio e fine coincidenti).



Come detto, gli edge possono avere forme differenziate; tendenzialmente si identificano come una **variazione** di una serie di pixel o anche un singolo pixel, che ha una caratteristica ben definita: ovvero si passa da un'area a **transizione costante** (vicina alla 0) ad un'area in cui c'è una **variazione** (o perturbazione). Questa variazione può avere varie forme: rampa, scalino, ecc...

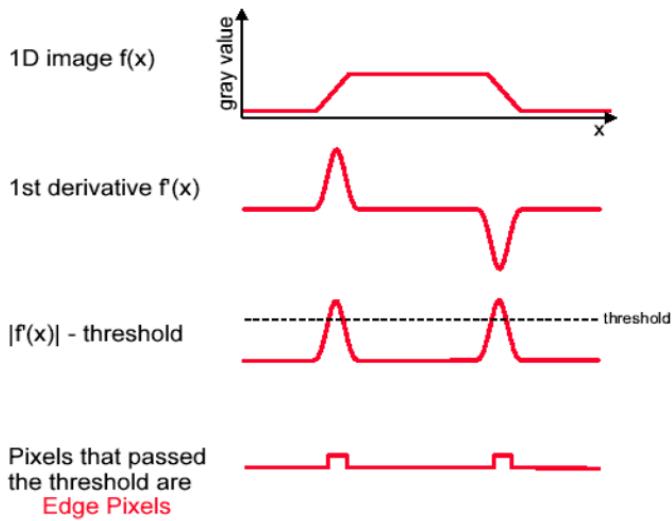


Tra le possibili situazioni di transizione tipiche, vi sono quelle a **rampa** e **scalino**. Il contorno a scalino presenta una transizione brusca tra due zone uniformi, invece nella transizione a rampa il passaggio tra le due regioni risulta essere più attenuato.



Nella figura vi è una immagine e la propria rappresentazione tridimensionale sotto forma di livelli di intensità. L'obiettivo è quello di individuare una sorta di detector che vada a mettere in evidenza le cime più alte delle variazioni di luminosità. La linea in rosso ha il compito di individuare quelle transizioni che hanno più alta probabilità di essere dei punti di contorno, non delle semplici variazioni all'interno dell'oggetto ma delle **variazioni significative** che indicano una transizione da un oggetto ad un altro.

contorno, non delle semplici variazioni all'interno dell'oggetto ma delle **variazioni significative** che indicano una transizione da un oggetto ad un altro.



Si supponga di avere una rappresentazione di un'immagine con una certa variabilità di toni di grigio. Sull'asse dell'ascisse vi sono i pixel e sull'asse delle ordinate i valori di grigio. Se viene calcolato la derivata prima della funzione in questione, si avrà come risultato delle informazioni sulla presenza di variazioni; quindi la derivata prima risulta essere **uguale a zero** quando **non vi sono valori significativi** e risulta avere una **variazione** laddove **esiste una variazione dei livelli**. Ritorna ad esser costante in presenza della zona uniforme, per poi trovare un'altra variazione e terminare in modo uniforme.

Quello che si vuol fare è capire se questi punti di massimo e minimo individuati dalla derivata prima, possono essere significativi. Quindi si vuol individuare una soglia che può dare informazioni circa i punti individuati al fine di comprendere se possono essere variazioni utili ad individuare un passaggio da un'area uniforme ad un'altra. Questi pixel presi singolarmente o in modo aggregato, sono quelli che possono essere potenzialmente etichettati come pixel di transizione ovvero **Edge Pixels**.

La difficoltà fondamentale è capire la soglia utile, e se essa può essere resa universale oppure deve essere adattata a classi di immagini, e se viene individuata per classi di immagini si può esser sicuri che questa soglia vada bene per tutta quanta l'immagine. Ovviamente **NO** in quanto la soglia è molto difficile da individuare e non sempre quella individuata per classi di immagini vada bene. Spesso tale soglia all'interno della stessa immagine permette sia di individuare pixel significativi ma individua potenziali pixel di contorno che non risultano esserlo. Questo pone un limite dell'applicazione di tali filtri alle immagini digitali.

Edge Detection: Punti Isolati

I punti isolati sono le discontinuità più semplici da individuare, in quanto non contribuiscono a definire la **morfologia** di un oggetto. Il punto corrispondente alla posizione centrale della maschera può essere considerato un punto isolato quando, per esempio con riferimento alla maschera mostrata si ha: $R > T$, essendo T un opportuno valore di soglia (non negativo). Anche in questo caso, quando si individua un punto isolato affinché tale punto possa essere etichettato come punto di discontinuità, deve esserci una soglia T che viene individuata. Tale soglia T deve far sì che una volta applicata la maschera ad una determinata area ed una volta calcolato il valore R , si è in grado di etichettare il pixel a cui corrisponde R come pixel di edge, soltanto se il valore è abbastanza significativo. Quindi il valore di grigio di quel punto isolato deve essere

abbastanza differente da quello dei suoi vicini. Se non vi è tale differenza evidente, non si può etichettare questo come punto isolato. In figura un esempio, in cui vi è una prima possibile rappresentazione di un filtro di Edge Detector basato sulla derivata prima. La maschera 3x3 presenta **valori negativi** e un **valore positivo centrale**, e la somma di tutti valori corrisponde a 0. Quindi in linea col concetto di derivata prima, questo filtro deve garantire che quando viene applicato all'immagine si dovrà avere

$$\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix} * \begin{bmatrix} 5 & 5 & 5 & 5 & 5 \\ 5 & 5 & 5 & 100 & 5 \\ 5 & 5 & 5 & 5 & 5 \end{bmatrix} = \begin{bmatrix} -95 & -95 & -95 \\ 0 & 760 & -95 \\ 0 & -95 & -95 \end{bmatrix}$$

Large Positive values = **light** point on dark surround
Large Negative values = **dark** point on light surround

che nelle zone uniformi, il risultato deve essere zero, quindi il pixel kernel non è un pixel candidato ad essere

un pixel di contorno, quando invece si trovano delle variazioni, esse devono essere evidenziate. Il principio portante di tali maschere, come detto, è che la **somma dei pesi deve essere uguale a zero**, per assicurare il concetto di rapporto incrementale. Quest'ultimo sarà nullo nelle aree in cui non c'è variazione, e ci sarà un rapporto incrementale evidente positivo o negativo quando esiste una variazione.

Edge Detection: Linee

È possibile costruire dei filtri che sono capaci di individuare non soltanto i punti, ma anche le **linee**, quindi privilegiare quei pixel che sono presenti lungo linee uniformi di variazione.

$$\begin{bmatrix} -1 & -1 & -1 \\ 2 & 2 & 2 \\ -1 & -1 & -1 \end{bmatrix} \quad \begin{bmatrix} -1 & -1 & 2 \\ -1 & 2 & -1 \\ 2 & -1 & -1 \end{bmatrix} \quad \begin{bmatrix} -1 & 2 & -1 \\ -1 & 2 & -1 \\ -1 & 2 & -1 \end{bmatrix} \quad \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}$$

— / | \

Come il filtro precedente, hanno la caratteristica di avere pesi negativi e positivi, e la somma di tutti i pesi è uguale a zero. Ognuno dei filtri riesce ad individuare una particolare caratteristica di linea.

Il primo individua punti di contorno che hanno uno sviluppo orizzontale, il secondo

invece punti di contorno che hanno uno sviluppo prevalentemente obliqui, il terzo punti di contorno che hanno uno sviluppo prevalentemente verticale, ed infine il quarto individua punti di contorno che hanno uno sviluppo obliqui ma con un senso rispetto al secondo. La prima maschera risponde più intensamente alle linee orizzontali (spesse un pixel): con uno sfondo costante, il massimo di R si ha infatti quando la linea coincide con la riga centrale della maschera. Analogamente, la seconda maschera risponde meglio a linee orientate a 45°, la terza a linee verticali e la quarta a linee orientate a 45°.

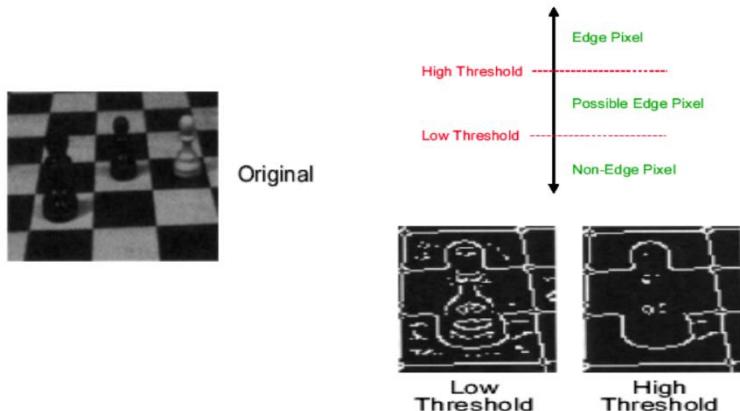
Si supponga di passare tutte e quattro le maschere su una immagine, e che in un certo punto si avranno R_1 , R_2 , R_3 , R_4 risposte, diverse, delle quattro maschere. Per capire se il pixel appartiene potenzialmente ad una maschera bisogna verificare quale valore R_i risulta essere più grande rispetto agli altri; una volta individuato tale valore, allora il punto di edge che si sta etichettando è associato alla maschera R_i e si trova lungo una linea specifica di quella determinata maschera.

Edge Detection: Contorni

L'estrazione dei contorni (edge) è sicuramente uno degli argomenti che hanno ricevuto più attenzione nella letteratura sull'image processing. Il contorno di un oggetto rappresenta infatti la **separazione** tra l'oggetto e lo sfondo o tra l'oggetto ed altri oggetti, per cui la sua estrazione è molto spesso il primo passo verso l'individuazione dell'oggetto. Un edge si presenta in una immagine come il confine tra due regioni caratterizzate da proprietà dei livelli di grigio in qualche modo distinguibili. Nel seguito si ipotizzerà che le regioni in questione siano sufficientemente omogenee, di modo che la determinazione della transizione tra le due regioni sia possibile sulla sola base della discontinuità dei valori di grigio. Le prime tecniche di edge detection che verranno analizzate sono basate sull'applicazione di un operatore locale di derivata.

Per rendere le regioni sufficientemente omogenee tra loro è possibile applicare uno Smoothing in modo che vengano individuate soltanto le variazioni importanti.

Per costruire dei filtri di edge detector basati sul principio di derivata è molto semplice, rispettando poche regole. Infatti, si prende una mascherina di dimensione dispari 3x3, 5x5, ecc... si assegnano i pesi in modo tale che essi siano tra di loro bilanciati con valori positivi e negativi e assicurando che la somma sia uguale a zero.



Uno dei problemi che si presenta è quello di stabilire se è necessario introdurre una **soglia**. L'introduzione di una soglia è una delle problematiche più importanti nelle fasi di costruzione degli edge detector. Nella figura vi è una immagine qualitativamente scadente, ed effettuare un'operazione di edge detector su questa immagine è piuttosto difficile estraendo il profilo degli oggetti che sono presenti. La forma del pedone bianco è facilmente

ottenibile siccome essa riesce a distinguersi all'interno della scacchiera. Il pedone nero invece, si confonde con il tassello nero alle sue spalle. Quindi la determina della soglia che permette di tirar fuori il profilo dell'oggetto è abbastanza complicata. I progettisti di edge detector hanno a disposizione questa soglia e possono decidere se utilizzare una soglia alta o bassa e questo produce degli effetti differenti nella qualità dell'immagine. Un edge detector trasforma un'immagine in una **immagine bianco e nera** quindi con solo due livelli di grigio, in quanto un edge detector deve solo stabilire se ogni pixel di partenza è un pixel di contorno oppure no. Se un pixel **non è di contorno** viene etichettato come **nero** (valore 0), se un pixel **è di contorno** viene etichettato come **bianco** (valore 255).

La scelta della **Threshold** (soglia) è fondamentale per stabilire quali sono i pixel tendenzialmente candidabili a diventare pixel di contorno; quindi una volta stabilita la soglia, tutti i pixel che si trovano al di sopra della soglia sono pixel di contorno, viceversa se si trovano sotto tale soglia non sono pixel di contorno. Infine, c'è una fascia grigia, che potrebbe oscillare tra una soglia troppo bassa o troppo alta.

Con **Low Threshold** si individuano potenziali pixel di contorno anche pixel che hanno variazioni piuttosto piccole. Tali variazioni possono introdurre come potenziali pixel di controllo anche pixel che non lo sono. Questo tipo di soglia permette di individuare il profilo del pedone nero.

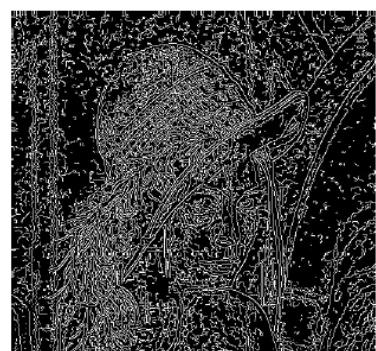
Con **High Threshold** si corre il rischio di ridurre troppo il numero di pixel che vengono etichettati come pixel di contorno, dando come risultato finale una quasi scomparsa totale della morfologia dell'oggetto. Il pedone ha perso tutta la parte centrale.



High threshold

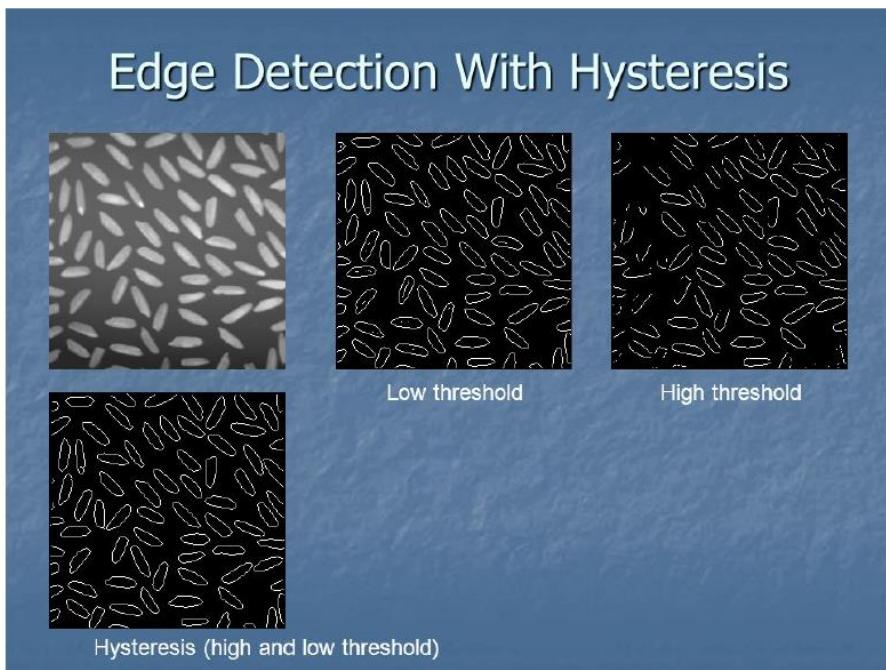


Good threshold



Low threshold

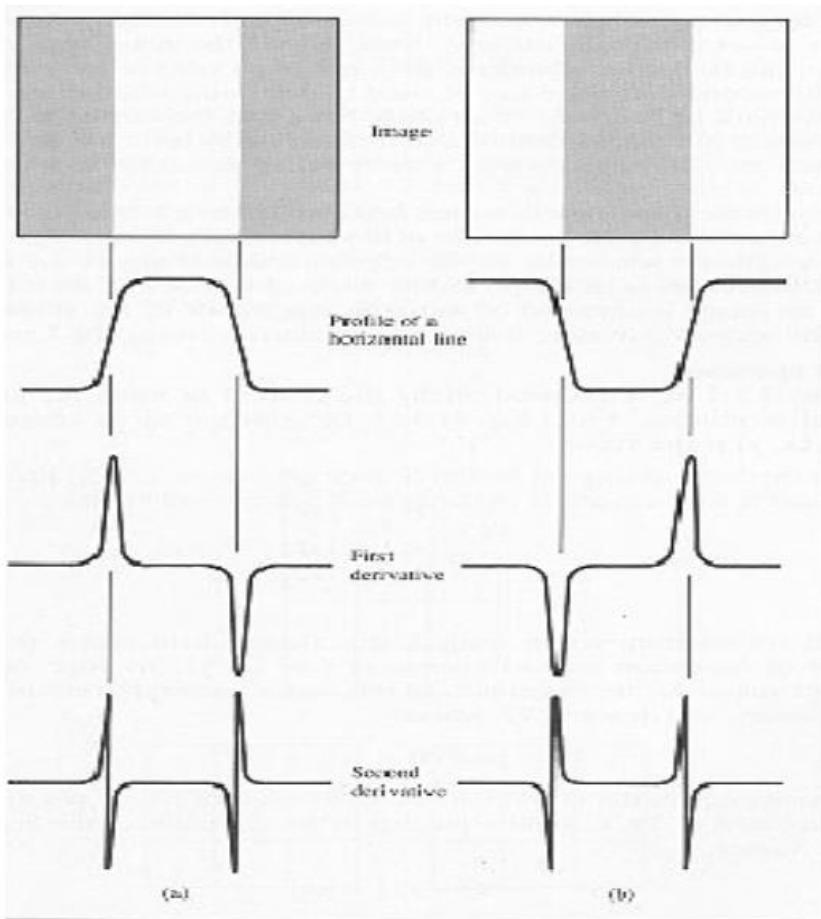
Edge Detection With Hysteresis



Hysteresis, ovvero una soglia intermedia dove vengono presi tutti i valori compresi tra una soglia alta e una soglia bassa.



In questo esempio all'immagine iniziale viene applicato uno Smoothing, affinché vengano tirati fuori dei contorni più puliti rispetto agli innumerevoli dettagli che presenta la foto originale. Infatti, il risultato mostra dei contorni abbastanza limpidi che definiscono la morfologia degli oggetti. In seguito, anche diverse tipologie di soglie. Grazie allo Smoothing è stato possibile non tirar fuori le piccole variazioni, producendo un buon risultato. L'ultima immagine rappresenta la sovrapposizione tra i contorni e l'immagine originale.



Il fatto che la derivata prima e la derivata seconda del profilo siano significativamente diverse da 0 soltanto in corrispondenza alle transizioni costituisce la motivazione dell'uso di operatori derivativi per l'estrazione dei contorni.

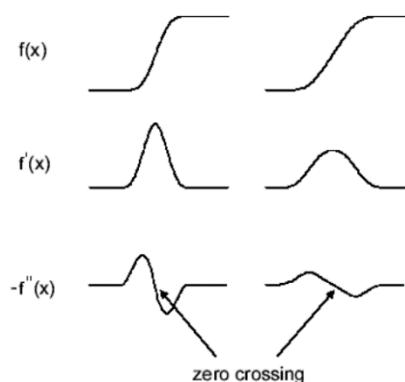
La **derivata prima** del profilo è **positiva** in corrispondenza di una transizione scuro-chiaro, **negativa** in corrispondenza di una transizione chiaro-scuro, **nulla** nelle zone a livello di grigio costante.

La **derivata seconda** è **positiva** in prossimità di un contorno, dalla parte scura del contorno stesso, **negativa** dalla parte chiara del contorno, nulla nelle zone a livello di grigio costante, ed esibisce un passaggio per lo **zero** o **zero crossing** esattamente in corrispondenza delle transizioni.

Ovvero, la **derivata prima** indica la

presenza di un massimo, la **derivata seconda** fornisce l'**esatta posizione del massimo**, nella posizione in cui il rapporto incrementale smette di crescere per iniziare la decrescita.

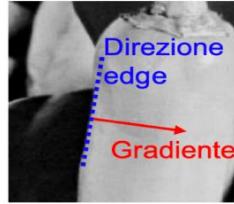
Riassumendo, il valore della derivata prima può essere utilizzato per determinare la presenza di contorni in una immagine. Gli zero crossing della derivata seconda ne possono consentire la precisa localizzazione. Il segno della derivata seconda permette di stabilire l'appartenenza di un pixel al versante scuro o al versante chiaro di un contorno. L'applicazione dei concetti precedentemente illustrati necessita tuttavia di alcune cautele, essenzialmente legate alla natura digitale delle immagini.



Edge Detection: Il Gradiente

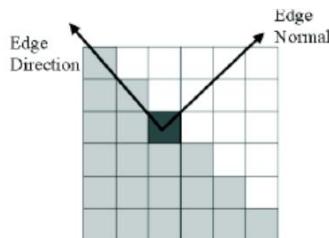
Il gradiente rappresenta uno degli strumenti attraverso il quale si può cercare di individuare le **zone di transizione**. Il gradiente di una funzione a due variabili viene calcolato come in figura. Del gradiente è possibile calcolare due informazioni: **l'entità del gradiente** (la grandezza della variazione) e la **direzione del gradiente**, ovvero in che direzione si muove il contorno. La direzione dell'edge è quella perpendicolare rispetto al gradiente. Il gradiente è uno strumento sofisticato che permette di calcolare due informazioni: se un pixel è etichettato come pixel di contorno sulla base della

Gradient	$\nabla f(x,y) = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix}$
----------	--



Gradient Magnitude $\sqrt{(\frac{\partial f}{\partial x})^2 + (\frac{\partial f}{\partial y})^2}$

Gradient Direction $\tan^{-1}(\frac{\partial f}{\partial y} / \frac{\partial f}{\partial x})$



dimensione del gradiente, infatti più è grande la variazione, più il pixel ha probabilità elevate di essere etichettato come pixel di contorno. Il secondo aspetto è legato alla direzione del contorno che viene calcolata come perpendicolare rispetto alla direzione del gradiente.

La rappresentazione del gradiente può essere effettuata attraverso vari operatori, le rappresentazioni 2x2 sono troppo sensibili al rumore, quindi si usano le 3x3. Tutte quelle che si vedranno con linee orizzontali o verticali con valori zero sono rappresentazioni discrete del gradiente.

Edge Detection: Roberts e Prewitt

- Molto Sensibili al Rumore

Roberts Edge Detector

$$F_A = f(x,y) - f(x-1,y-1)$$

$$F_B = f(x-1,y) - f(x,y-1)$$

$$A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \quad B = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

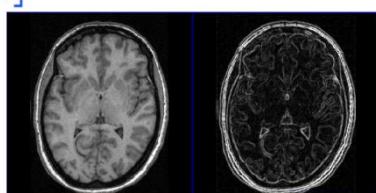
Prewitt Edge Detector

$$A = \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix} \quad B = \begin{bmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}$$

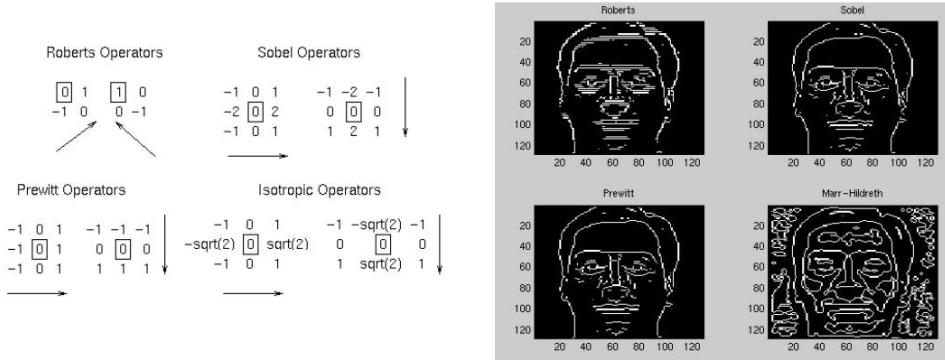
Edge Detection: Sobel

Sobel Edge Detector

$$A = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \quad B = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}$$



Edge Detection: Roberts, Sobel, etc.



Edge Detection: Laplaciano

Il Laplaciano è una rappresentazione della derivata seconda, che a differenza del gradiente, è caratterizzato dall'avere un pixel centrale positivo, le direzioni principali con valori negativi, e gli altri posti uguali a zero. Si prendono i coefficienti della derivata seconda della funzione (x, y) rispetto a x , a due variabili. E questi poi costituiscono poi i coefficienti che si andranno a riprodurre all'interno della matrice. Questa rappresentazione del Laplaciano è la **rappresentazione discreta di una derivata seconda**.

Approximation of second derivative (horizontal):

$$\frac{\partial^2 f(x,y)}{\partial x^2} = f''(x,y) = f(x+1,y) - f(x,y) = \\ = [f(x+1,y) - f(x,y)] - [f(x,y) - f(x-1,y)] \\ = f(x+1,y) - 2f(x,y) + f(x-1,y)$$

convolution with: $[1 \ -2 \ 1]$

Approximation of second derivative (vertical):

convolution with:

1
-2
1

Laplacian Operator

$$\nabla^2 = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right)$$

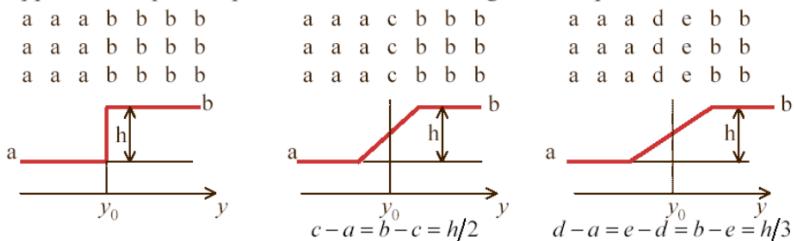
convolution with:

0	-1	0
-1	4	-1
0	-1	0

- ✓ Una versione (normalizzata) del laplaciano riferito agli 8-vicini, quindi con le differenze delle derivate prime mediate su tre righe e tre colonne, è:

$$\mathbf{H} = \frac{1}{8} \begin{bmatrix} -2 & 1 & -2 \\ 1 & 4 & 1 \\ -2 & 1 & -2 \end{bmatrix}$$

- ✓ Applichiamo questo operatore ai modelli di edge visti in precedenza:



Operatori basati sul laplaciano

- ✓ Si ottiene, con riferimento alla riga centrale dei tre modelli:

$$\begin{array}{ccccccc} \text{grigi} & a & a & a & b & b & b \\ \text{edge} & 0 & 0 & -\frac{3}{8}h & \frac{3}{8}h & 0 & 0 \end{array}$$

gradino

$$\begin{array}{ccccccc} a & a & a & c & b & b & b \\ 0 & 0 & -\frac{3}{16}h & 0 & \frac{3}{16}h & 0 & 0 \end{array}$$

rampa più ripida $c - a = b - c = h/2$

$$\begin{array}{ccccccc} \text{grigi} & a & a & a & d & e & b & b \\ \text{edge} & 0 & 0 & -\frac{h}{8} & 0 & 0 & \frac{h}{8} & 0 \end{array}$$

rampa meno ripida
 $d - a = e - d = b - e = h/3$

Per l'edge a rampa più ripida, il laplaciano rileva correttamente il contorno in corrispondenza allo ZC

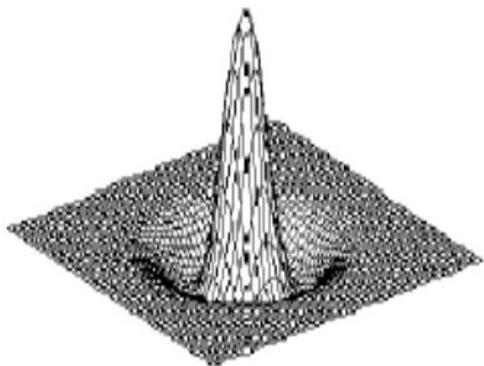
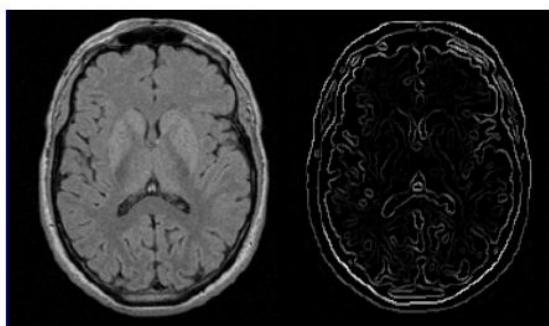
Per l'edge a gradino, lo ZC si trova fra i due pixel a sinistra e a destra del contorno, e per convenzione l'edge è collocato sul pixel a destra dello ZC

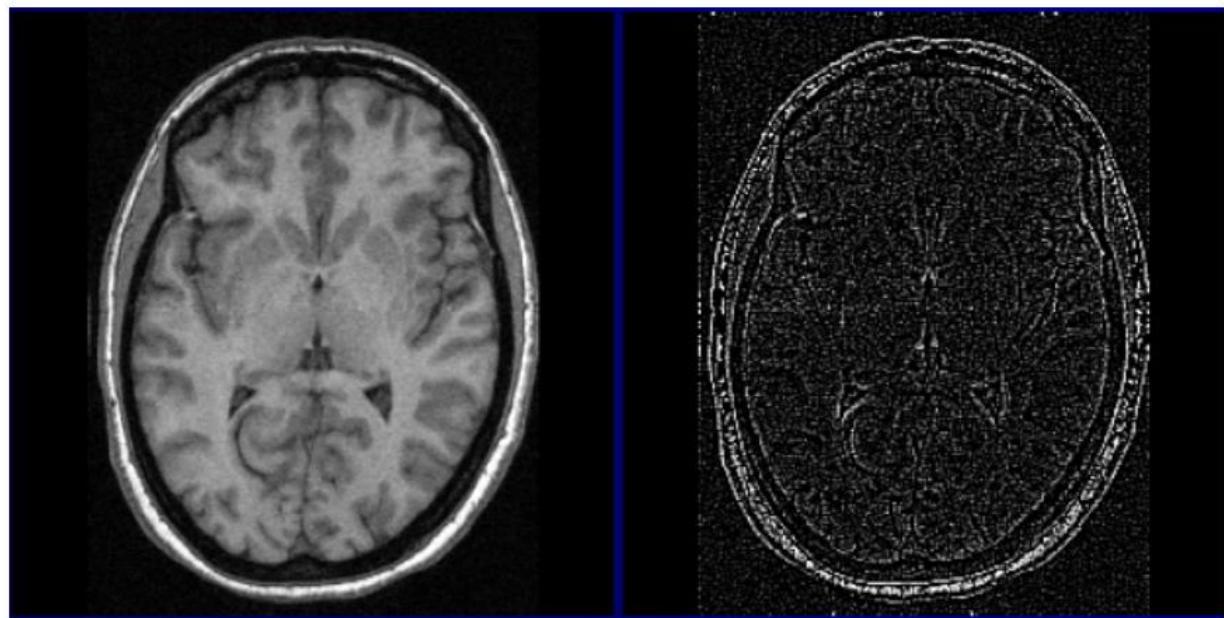
Neanche nell'ultimo caso lo ZC è localizzato esattamente su un pixel, anzi si ha un edge *doppio*

Variations on Laplace Operators:

All are approximations of:

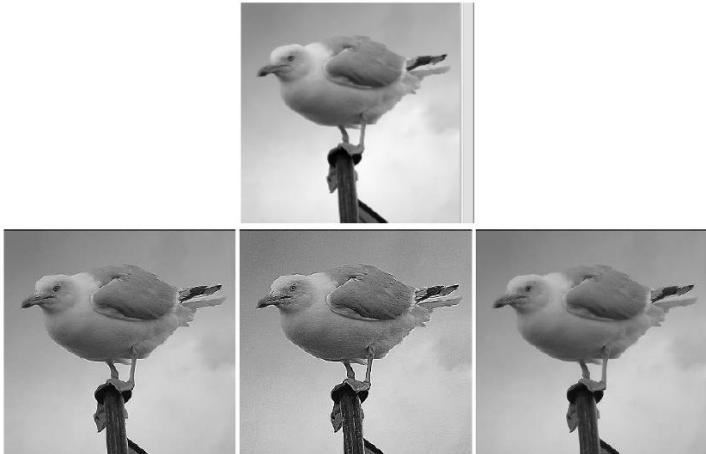
$$\begin{bmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{bmatrix} \quad \begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix} \quad \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$$





Sharpening

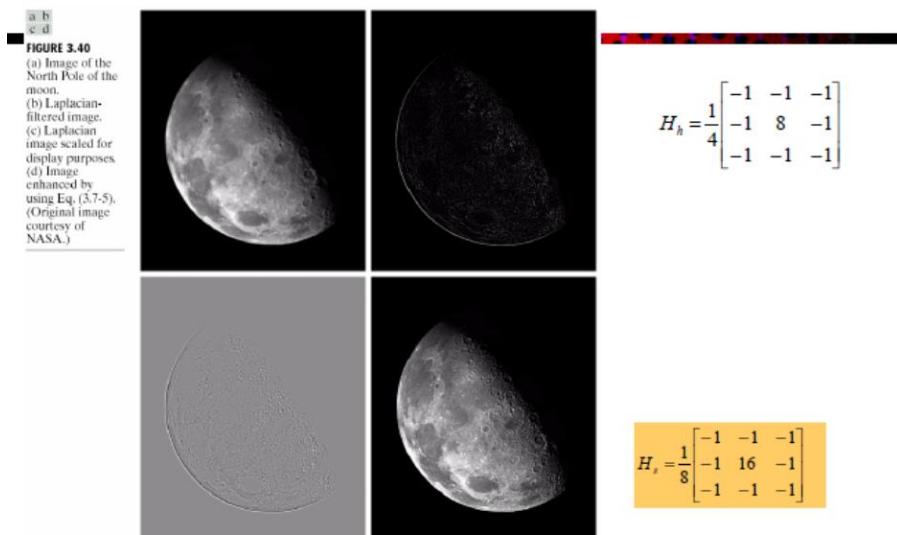
Un'operazione di Sharpening tende ad **evidenziare i dettagli fini**, attraverso un'operazione che prevede prima l'estrazione dei contorni e poi la sovrapposizione dei contorni con l'immagine originale. Quindi non si ottiene una immagine in bianco e nero ma, con le operazioni di Sharpening, dopo aver estratto i contorni essi vengono sovrapposti all'immagine, per renderli ancor più evidenti.



$$M_1 = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix} \quad M_2 = \begin{bmatrix} -1 & -1 & -1 \\ -1 & 9 & -1 \\ -1 & -1 & -1 \end{bmatrix} \quad M_3 = \begin{bmatrix} 1 & -2 & 1 \\ -2 & 5 & -2 \\ 1 & -2 & 1 \end{bmatrix}$$

Da come si evince in figura, si individua una immagine di partenza e successivamente si applica lo Sharpening utilizzando filtri differenziati con l'obiettivo però di non dare in output la semplice immagine dei contorni ma di sovrapporre i contorni ottenuti all'immagine originale.

Example of Sharpening



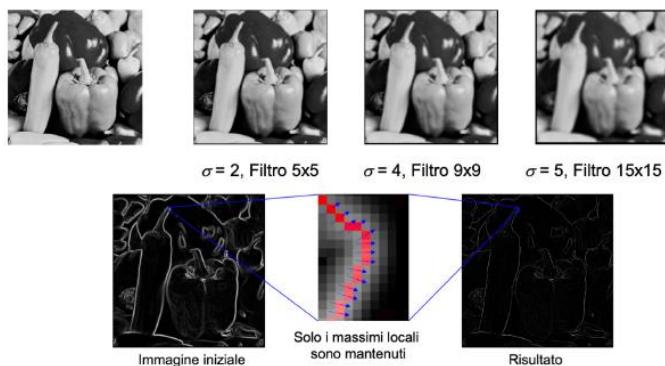
Canny Edge Detector

Uno degli operatori maggiormente utilizzati per estrarre i contorni è l'operatore edge detector Canny, il quale si è dimostrato molto efficace. Questo operatore mette insieme tutto ciò che è stato detto finora, e sperimentalmente si è visto che è in grado di produrre risultati molto significativi. Questo perché il metodo di Canny ha un'elevata possibilità di produrre edge connessi, cioè **contorni chiusi** o piuttosto **continuativi** che sono uno degli obiettivi che si ricercano all'interno delle operazioni di edge detection. Quando si vuole effettuare l'estrazione di contorni e preservare la chiusura di essi, Canny fornisce questo tipo di prestazioni. L'algoritmo di Canny prevede:

1. Smoothing gaussiano dell'immagine, per eliminare le piccole transazioni.
2. Calcolo del gradiente.
3. Soppressione dei non-massimi in direzione ortogonale all'edge.
4. Selezione degli edge significativi mediante isteresi.

La qualità del risultato dell'operatore Canny dipende da:

- Ampiezza della gaussiana nella prima fase.
- Dimensione del filtro nella prima fase.
- T1 e T2, soglie per l'isteresi nell'ultima fase.



SOGLIATURA CON ISTERESI (Doppia Soglia)

Due soglie, una bassa ed una alta, che vengono confrontate con il gradiente in ciascun punto.

- Valore del gradiente:

- **Inferiore** alla soglia bassa, il punto è scartato;
- **Superiore** alla soglia alta, il punto è accettato come parte di un contorno;
- **Compreso** fra le due soglie, il punto è accettato solamente se *contiguo* ad un punto già precedentemente accettato.

Output: Immagine binaria dove ciascun pixel è marcato come contorno o no

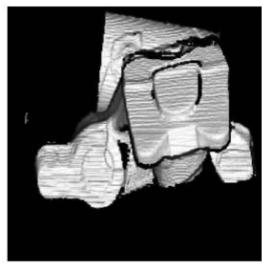
Canny edge detector – Esempi



1) Immagine originale – 2) Smoothing gaussiano – 3) Modulo del gradiente
4) Soppressione non-massimi – 5) Selezione edge – 6) Edge sovrapposti a immagine originale



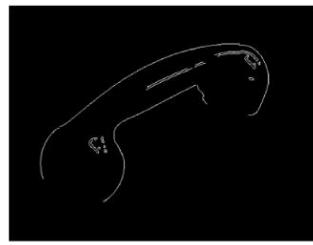
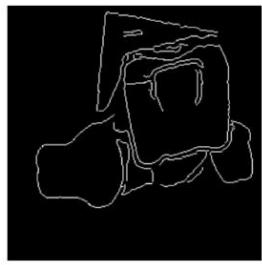
$\sigma=2$ F:5x5 T₁=0.3 T₂=0.15



$\sigma=2$ F:5x5 T₁=0.4 T₂=0.2



$\sigma=1$ F:3x3 T₁=0.3 T₂=0.15



Trasformata di Fourier

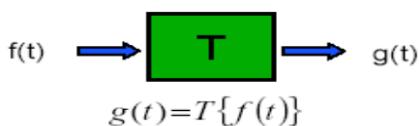
Viene fatta una premessa iniziale: quando si vuol operare e quindi applicare dei filtri per rimuovere delle informazioni per rendere più evidenti altre informazioni, tale operazione è possibile effettuarla in due modalità: operando direttamente sui **pixel** oppure trasformare l'immagine in un altro dominio e trovare vantaggio nell'evidenziare informazioni significative. L'obiettivo è costruire dei filtri che non agiscono direttamente sui pixel, ma agiscono su altre informazioni correlate con i pixel ma aventi natura diversa. È chiaro che quando si effettua una trasformazione di un'immagine da un dominio all'altro, si deve essere in grado di portare l'immagine da un dominio all'altro attraverso una trasformazione, ma fare anche l'**operazione inversa**, riportando l'immagine dal dominio nuovo al dominio dei pixel. Questa operazione può esser fatta con la **trasformata di Fourier**. Intuitivamente si capisce che questo tipo di operazione comporta un **aggravio computazionale**, per cui l'obiettivo è quello di capire se questo aggravio computazionale produce dei risultati qualitativamente migliori.

Sistema Lineare

Le trasformazioni lineare è una trasformazione che agisce su un segnale in ingresso $f(t)$, che può essere

- A **linear system T** gets an **input $f(t)$** and produces an **output $g(t)$** :

associato ad un'immagine. Tale segnale entra in questo box verde e produce un risultato $g(t)$.



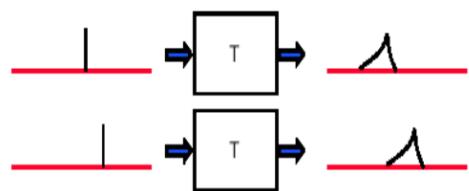
- In the discrete case:

- input: $f[n]$, $n = 0, 1, 2, \dots$
- output: $g[n]$, $n = 0, 1, 2, \dots$

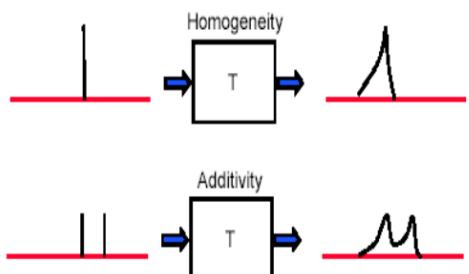
$$g[n] = T\{f[n]\}$$

Per essere un sistema lineare, esso deve godere di alcune proprietà:

- **Homogeneity:** $T\{af[n]\} = aT\{f[n]\}$



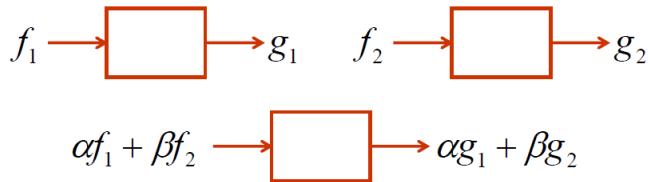
- **Additivity:** $T\{f_1[n] + f_2[n]\} = T\{f_1[n]\} + T\{f_2[n]\}$



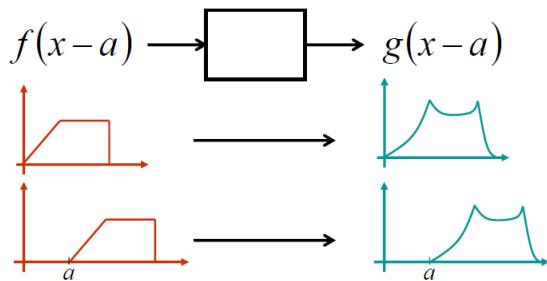
Linear Shift Invariant Systems (LSIS)

Un aspetto che si vuol introdurre nei sistemi lineari è la cosiddetta **proprietà dello Shift Invariant**. Cioè dei sistemi che sono invarianti al tempo. Un sistema è shift invariant quando il suo comportamento nel tempo **non cambia** e l'unica variazione che si ottiene è la **traslazione** del risultato. Si può dire che un sistema gode dell'invarianza allo shift se ha un comportamento come nel seguito:

Linearity:

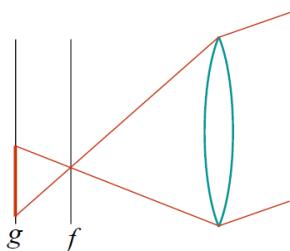


Shift invariance:



Se all'istante t il sistema riceve un segnale di questo tipo e produce un risultato simile, allora il sistema è Shift Invariant se all'istante $t = a$, con una traslazione temporale sull'asse delle ascisse, il sistema riceve un segnale analogo al precedente in input ma traslato temporalmente, allora produrrà come risultato lo stesso risultato precedente ma con una traslazione temporale. Se ciò avviene allora il sistema gode dell'invarianza allo shift. Quindi uguale comportamento ma traslato nel tempo.

Example of LSIS



Defocused image (g) is a processed version of the focused image (f)

Ideal lens is a LSIS $f(x) \rightarrow \boxed{\text{LSIS}} \rightarrow g(x)$

Linearity: Brightness variation

Shift invariance: Scene movement

(not valid for lenses with non-linear distortions)

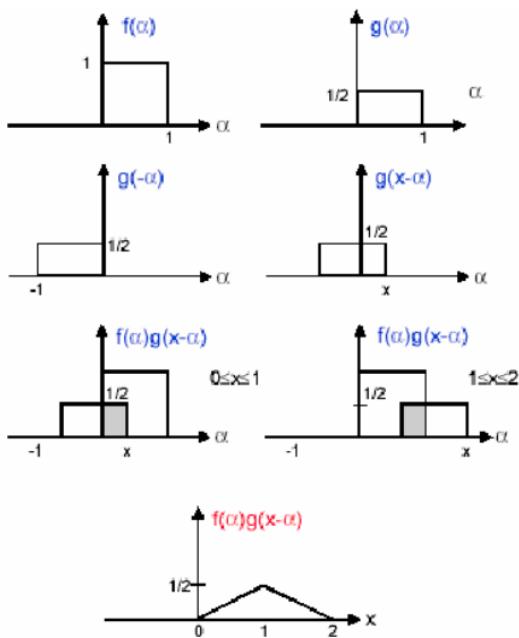
Convolution in Continuous Case

La **convoluzione** tra due funzioni, nel caso continuo, si scrive nel seguente modo:

$$(f * g)(x) = \int_{-\infty}^{\infty} f(\alpha)g(x-\alpha) d\alpha$$

la presenza dell'integrale indica che il risultato di questa operazione, sarà un'area che si andrà a ritagliare.

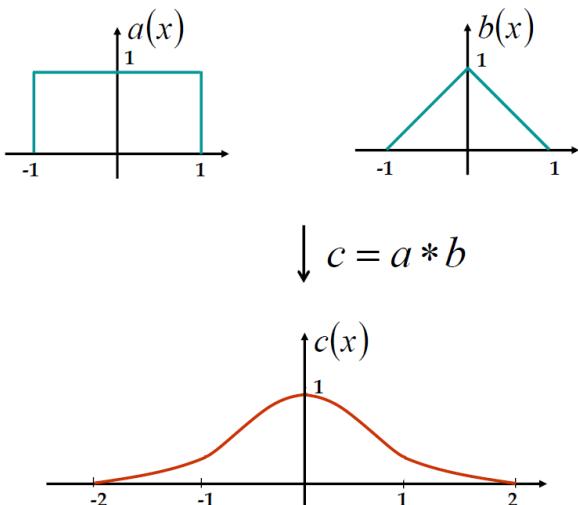
Si mostrano degli esempi:



La convoluzione fra f e g produce l'area che viene intersecata dalle due funzioni. Nell'ultimo grafico vi è la rappresentazione geometrica della convoluzione fra le due funzioni.

La convoluzione è un'operazione tra due funzioni di una variabile che consiste nell'integrare il prodotto tra la prima e la seconda traslata di un certo valore

Convolution - Example



Convolution Kernel – Impulse Response

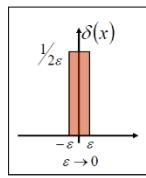
Il Delta di Dirac è un impulso unitario, è la più semplice delle funzioni che permette di rappresentare ogni

$$f \rightarrow \boxed{h} \rightarrow g \quad g = f * h$$

funzione in relazione al Delta di Dirac. Esso è una funzione:

- What h will give us $g = f$?

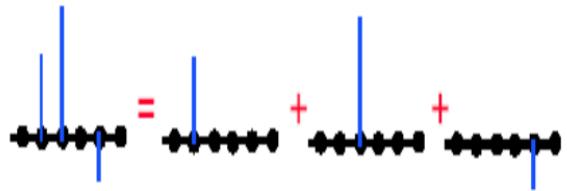
Dirac Delta Function (Unit Impulse)



$$\delta[n-k] = \begin{cases} 0 & \text{where } n \neq k \\ 1 & \text{where } n = k \end{cases}$$

Se viene considerata una funzione di questo tipo, si può rappresentare una qualsiasi funzione:

$$f[n] = \sum_{j=-\infty}^{\infty} f[j] \delta[n-j]$$



La funzione può essere decomposta in somma lineare di impulsi scalati e shiftati, ciascuno dei quali viene intercettato dal Delta di Dirac.

Shift-Invariant Linear System and Convolution

Si analizza ora cosa è possibile effettuare con i sistemi shift Invariant sfruttando anche la convoluzione.

Proof:

Shift-Invariant Linear System is a Convolution

- $f[n]$ input sequence
- $g[n]$ output sequence
- $h[n]$ the system impulse response:

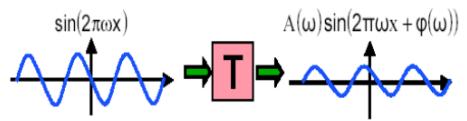
$$h[n] = T\{\delta[n]\}$$

$$\begin{aligned} g[n] &= T\{f[n]\} = T\left\{ \sum_{j=-\infty}^{\infty} f[j] \delta[n-j] \right\} \\ &= \sum_{j=-\infty}^{\infty} f[j] T\{\delta[n-j]\} \quad (\text{from linearity}) \\ &= \sum_{j=-\infty}^{\infty} f[j] h[n-j] \quad (\text{from shift-invariance}) \\ &= f * h \end{aligned}$$

The output is a sum of scaled and shifted copies of impulse responses.

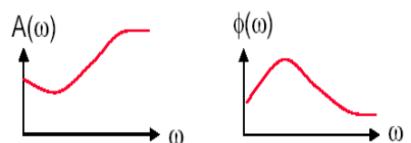
L'intento è costruire una trasformazione che permetta di ottenere dei risultati. Lavorando sui pixel, T è stato costruito con le matrici e i pesi relativi, ottenendo T diversi tra loro. In questo caso si vuol costruire T affinché agisca sulle funzioni trigonometriche. Quindi si vuol costruire una trasformazione lineare e shift invariante, ed applicata ad una funzione sinusoidale può agire su due aspetti: non modificare la frequenza della funzione ma modificarne **l'ampiezza** e la **fase**. E ci si concentrerà sulla modifica dell'ampiezza.

Quindi la risposta di un sistema lineare shift invariante ad una funzione seno è una copia scalata e shiftata della funzione seno con la stessa frequenza.



$$e^{i2\pi\omega x} \quad A(\omega)e^{i\phi(\omega)}e^{i2\pi\omega x}$$

- The **frequency response** (or Transfer Function) of a system:



$$\delta(x) \rightarrow \boxed{T} \rightarrow h(x)$$

Impulse response

Nel dominio spaziale si fa entrare un impulso e il sistema risponde con una funzione trasformata.

Nel dominio delle frequenze si fa entrare un'onda sinusoidale e il sistema è in grado di modificare la sua ampiezza.

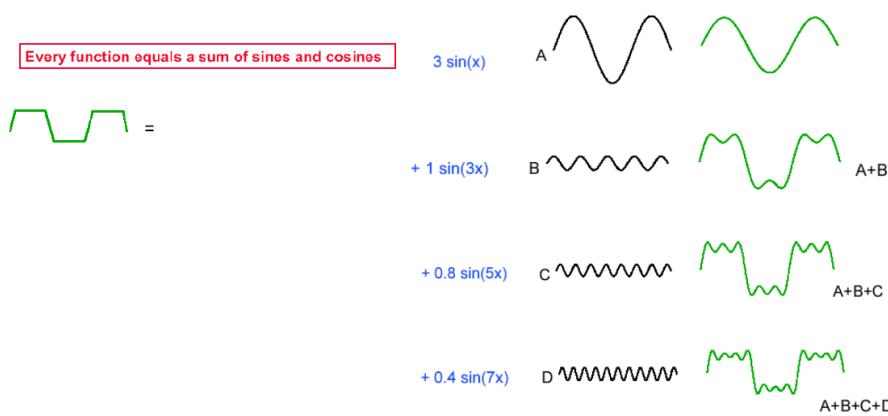
$$A(\omega)e^{i\phi(\omega)} = H(\omega)$$

$$e^{i2\pi\omega x} \rightarrow \boxed{T} \rightarrow H(\omega)e^{i2\pi\omega x}$$

Frequency response

La Trasformata di Fourier

Fourier affermò che una qualsiasi funzione periodica, può essere scritta sempre sotto forma di funzione seno e coseno pesate, e con frequenze differenti.



Come si può notare, le varie funzioni sommate tra loro riescono ad approssimare la funzione di partenza. Con infinite frequenze, aumenta maggiormente la frequenza che si avvicina alla funzione di partenza. Quindi, infiniti punti nel dominio spazio-tempo corrispondono a infinite frequenze nel dominio delle frequenze.

Numero finito di pixel nel dominio delle frequenze, finite frequenze nel dominio delle frequenze.

Analisi delle frequenze

Si formalizza ora la Trasformata di Fourier. Nella rappresentazione formale:

The Fourier Transform:

It is possible to express any signal as a sum of shifted and scaled sinusoids at different frequencies.

$$f(x) = \sum_{\omega} F(\omega) e^{i2\pi\omega x}$$

Or

$$f(x) = \int F(\omega) e^{i2\pi\omega x} d\omega$$

La funzione $f(x)$ è la funzione nello spazio-tempo (nel contesto dei pixel è dominio di pixel). Quindi tale funzione è rappresentabile come la sommatoria di ω , dove essa rappresenta le frequenze di $F(\omega)$ che è l'ampiezza (ovvero il peso) moltiplicato per $e^{i2\pi\omega x}$. Quindi per ogni frequenza ω , si assegna un peso stabilito. Questa trasformata di Fourier permette di prendere un segnale e rappresentarlo in termini di funzioni sinusoidali a frequenza variabile ed ampiezza variabile, cioè pesate. Ciò può essere scritto sotto forma di sommatoria oppure di integrale.

Per essere completi dal punto di vista della trasformata di Fourier si necessita di una soluzione che dia la possibilità di passare dal dominio spazio-tempo al dominio delle frequenze e viceversa. Il primo passaggio da effettuare è quello di costruire le frequenze partendo dalla funzione. $F(\omega)$ rappresenta l'insieme delle ampiezze che si assegnano alle varie frequenze. Questa trasformazione porta dal dominio spazio-tempo al dominio delle frequenze. Quando si vuol fare l'operazione inversa si applica la **funzione inversa**. Se vi è un ritorno dal dominio delle frequenze al dominio spazio-tempo, caratterizzato dall'assenza di modifiche è assolutamente indolore ovvero non si perdono operazioni. Se l'informazione nel dominio spazio-tempo è composta da n pixel, si avranno n frequenze nel dominio delle frequenze e viceversa.

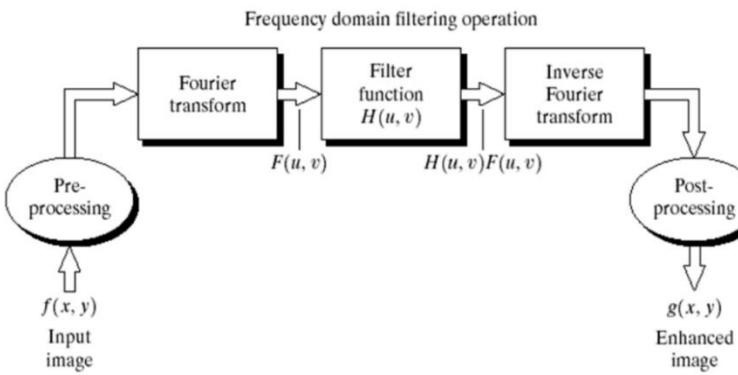
The **inverse Fourier Transform** composes a signal $f(x)$ given $F(\omega)$:

$$f(x) = \int_{\omega} F(\omega) e^{i2\pi\omega x} d\omega$$

The **Fourier Transform** finds the $F(\omega)$ given the signal $f(x)$:

$$F(\omega) = \int_x f(x) e^{-i2\pi\omega x} dx$$

Si sintetizza graficamente ciò che è stato fatto.



Si parte da una immagine nel dominio dei pixel $f(x, y)$, dove x e y corrispondono alle coordinate della posizione del pixel ed f corrisponde al valore di luminosità associato a questa posizione. L'immagine successivamente viene processata mediante una trasformata di Fourier e quindi si vanno a rappresentare i pixel in termini di frequenze, come nell'enunciato precedente. Ora l'idea è costruire un filtro nel dominio delle frequenze che può essere applicato ad F attraverso una operazione di moltiplicazione esattamente come fatto nei filtri spaziali. Dopo fatto ciò si applica la trasformata inversa per ottenere $g(x, y)$. Quindi invece che applicare la trasformazione ad f , si applica la trasformazione ad una rappresentazione nel dominio delle frequenze di f .

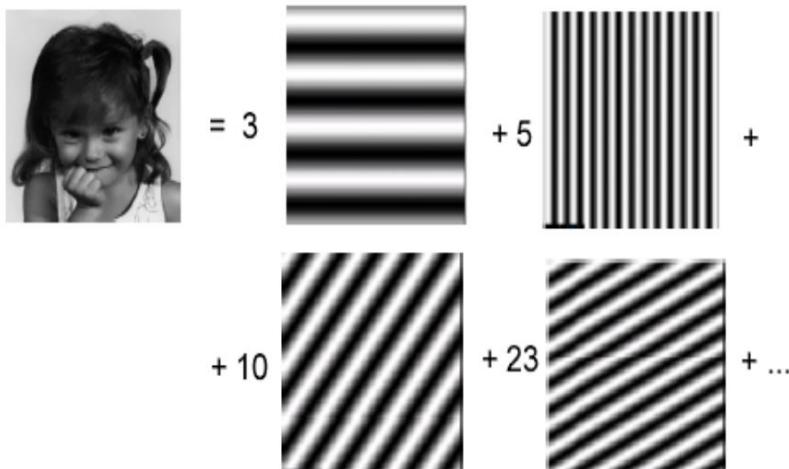
Trasformata di Fourier discreta

The Discrete Fourier Transform (DFT) is defined as:

$$F(u) = \frac{1}{N} \sum_{x=0}^{N-1} f(x) e^{-\frac{2\pi i u x}{N}} \quad u = 0, 1, 2, \dots, N-1$$

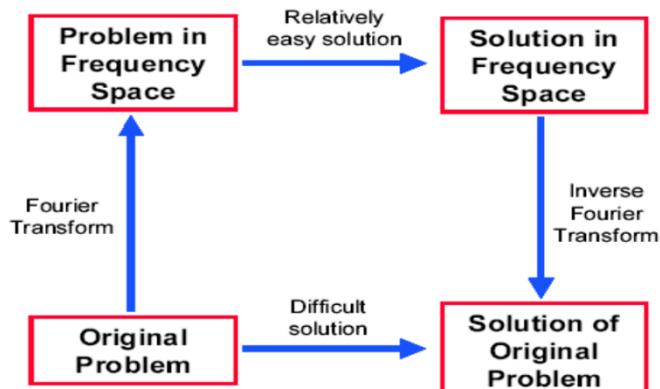
The Inverse Discrete Fourier Transform (IDFT) is defined as:

$$f(x) = \sum_{u=0}^{N-1} F(u) e^{\frac{2\pi i u x}{N}} \quad x = 0, 1, 2, \dots, N-1$$



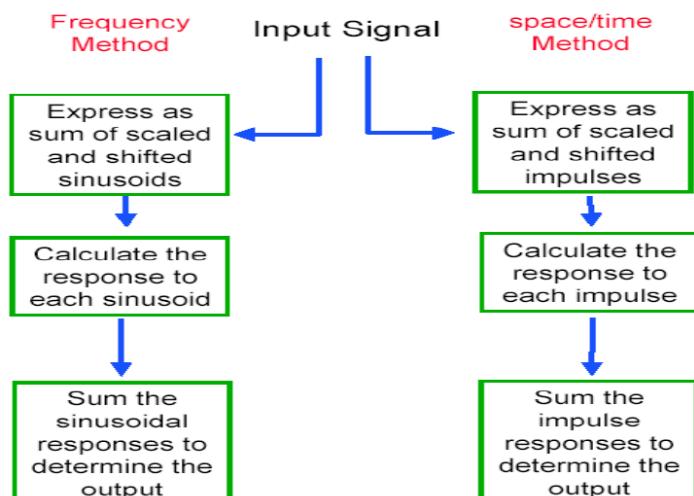
Tendenzialmente, quando si applica la trasformata di Fourier si cerca un certo numero di funzioni seno e coseno. Come si vede in figura, dalla combinazione tra le varie funzioni si ottiene la funzione di partenza.

Spatial vs Frequency Domain



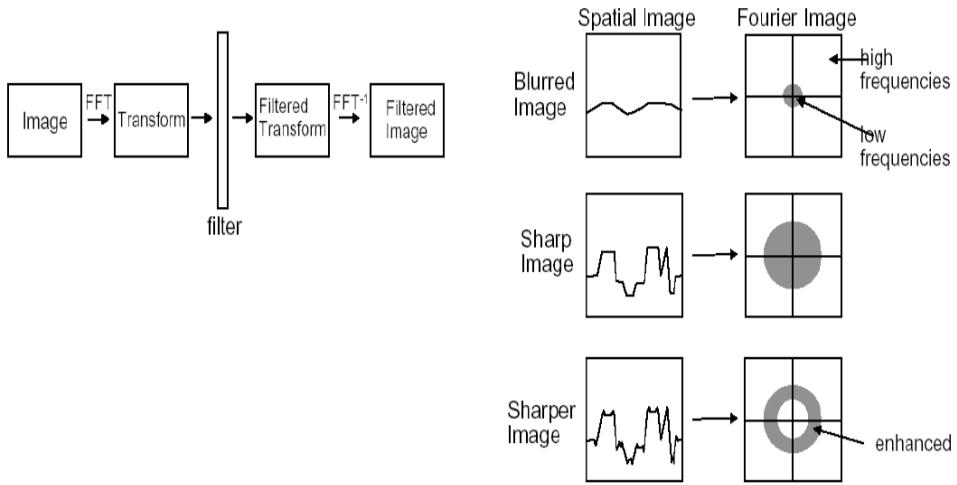
Il problema iniziale è filtrare l'immagine, quindi in qualche modo bisogna risolvere un problema nel dominio-spazio-tempo; in tale dominio **vi sono delle difficoltà** perché i pixel hanno una natura molto diversificata, quindi ciò può comportare dei problemi. Ed è per questo che si fa ricorso all'utilizzo di soglie che non sempre possono essere precise. Potrebbe essere più funzionale applicare la trasformata di Fourier per tentare di risolvere il problema operando in un dominio

differente ovvero il **dominio delle frequenze**, dove dovrebbe essere più semplice rilevare qualche soluzione per il problema in questione. Si dovrebbe verificare che filtrare una immagine attraverso la propria rappresentazione nel dominio delle frequenze, risulta essere più semplice e questa soluzione, attraverso la trasformata di Fourier inversa, che riporta al dominio spazio-tempo comporterebbe avere la soluzione del problema originale. La trasformata di Fourier, applicata ad un'immagine $N \times N$, computazionalmente viene a costare $O(N^2)$, quindi un aumento di complessità rispetto alla linearità che si verificava nelle trasformate lineari. La **Fast Fourier Transform** costa $O(n \log n)$.

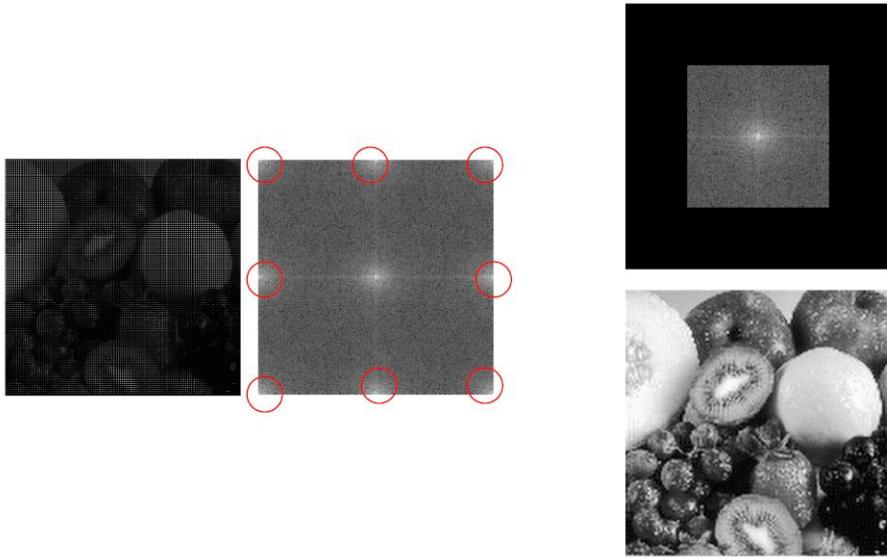


$$G(\omega) = F(\omega)H(\omega)$$

$$g(x) = f(x) * h(x)$$



Filtering Example



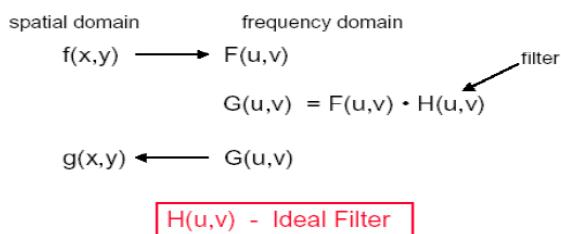
In figura, vi è una immagine con rumore. È molto difficile rimuovere il rumore con i filtri enunciati in precedenza, di fianco vi è la rappresentazione della trasformata di Fourier, dove viene rappresentato lo **spettro di Fourier**, ovvero la rappresentazione delle frequenze o meglio delle ampiezze, che si sono ottenute nella trasformata. Le ampiezze nello spettro di Fourier si possono

interpretare come: tendono ad essere **più alte** in prossimità delle **basse frequenze**, e tendono progressivamente ad **azzerarsi** man mano che le frequenze **aumentano**. I bagliori cerchiati in rosso indicano delle ampiezze significative nelle alte frequenze, il che è atipico per le immagini e ciò significa che verosimilmente l'immagine è corrotta dalla presenza di rumore, e quindi tale aree vanno eliminate. Per fare tale operazione vanno azzerate le ampiezze che riguardano le alte frequenze, ovvero una **moltiplicazione per zero** per tutti i coefficienti delle alte frequenze, e una **moltiplicazione per uno** per le frequenze che si trovano nella parte bassa e media. Ottenendo uno spettro di Fourier come in alto, e il risultato dell'immagine, operando l'anti trasformata di Fourier è una immagine senza rumore e qualitativamente migliore. Con un filtro spaziale non si sarebbe mai ottenuto tale risultato.

La trasformata di Fourier rappresenta attraverso il suo spettro, la **distribuzione delle ampiezze** delle varie funzioni trigonometriche utilizzate, quindi una immagine senza rumore o con pochi dettagli, avrà un bagliore localizzato vicino allo zero (centro), infatti vi sono poche frequenze basse e quasi nulla di alte frequenze. Se l'immagine presenta molti contorni e molti dettagli, lo spettro di Fourier avrà un bagliore molto più ampio, che sta ad indicare che ci sono molte frequenze con ampiezze più significative.

Low Pass Filter

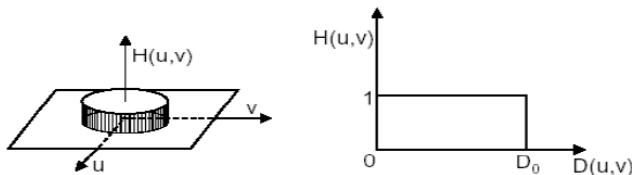
Quando si costruisce un filtro, bisogna decidere quali frequenze conservare e quali no. La prima cosa fattibile è la costruzione di un filtro passa basso, ovvero un filtro che lascia **transitare** solamente **basse frequenze**. Tale filtro viene costruito sulla base dell'individuazione delle frequenze che si vogliono conservare e per



$$H(u,v) = \begin{cases} 1 & D(u,v) \leq D_0 \\ 0 & D(u,v) > D_0 \end{cases}$$

$$D(u,v) = \sqrt{u^2 + v^2}$$

D_0 = cut off frequency



queste frequenze si pensa ad una moltiplicazione per uno, e per le frequenze che si vogliono tagliare quindi una moltiplicazione per zero. Il filtro diventa quindi un moltiplicatore, attuato sulle frequenze. D_0 rappresenta la frequenza di taglio, quella rispetto alla quale si stabilisce una soglia.

Sharpening (High Pass) Filter

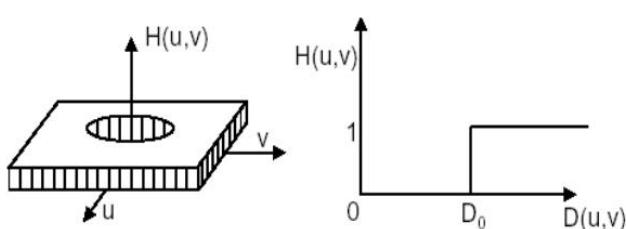
Il filtro passa alto sarà opposto al precedente, esso mette in evidenza i dettagli.

H(u,v) - Ideal Filter

$$H(u,v) = \begin{cases} 0 & D(u,v) \leq D_0 \\ 1 & D(u,v) > D_0 \end{cases}$$

$$D(u,v) = \sqrt{u^2 + v^2}$$

D_0 = cut off frequency



Band Pass Filter

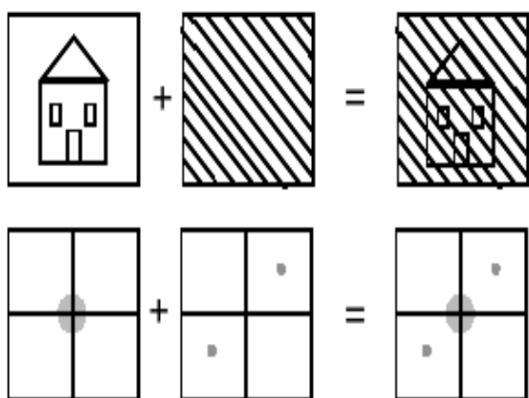
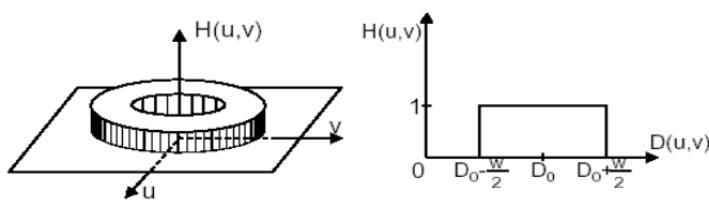
Prende una parte delle basse frequenze e una parte delle alte frequenze.

$$H(u,v) = \begin{cases} 0 & D(u,v) \leq D_0 - \frac{w}{2} \\ 1 & D_0 - \frac{w}{2} \leq D(u,v) \leq D_0 + \frac{w}{2} \\ 0 & D(u,v) > D_0 + \frac{w}{2} \end{cases}$$

$$D(u,v) = \sqrt{u^2 + v^2}$$

D_0 = cut off frequency

w = band width



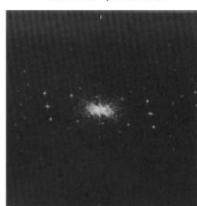
Come si vede in figura, vi è una immagine senza rumore, il cui spettro di Fourier risulta essere del tipo sotto disegnato. Quando si aggiunge del rumore, esso si annida nella alte frequenze, ma mai nelle basse frequenze. Quando vi è la somma delle due immagini vi è uno spettro di Fourier come mostrato. Se si moltiplica per zero le frequenze del rumore, esso scomparirà senza intaccare la qualità dell'immagine.

Filtering Example

Original Noisy image



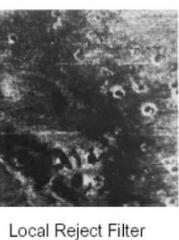
Fourier Spectrum



Original Noisy image



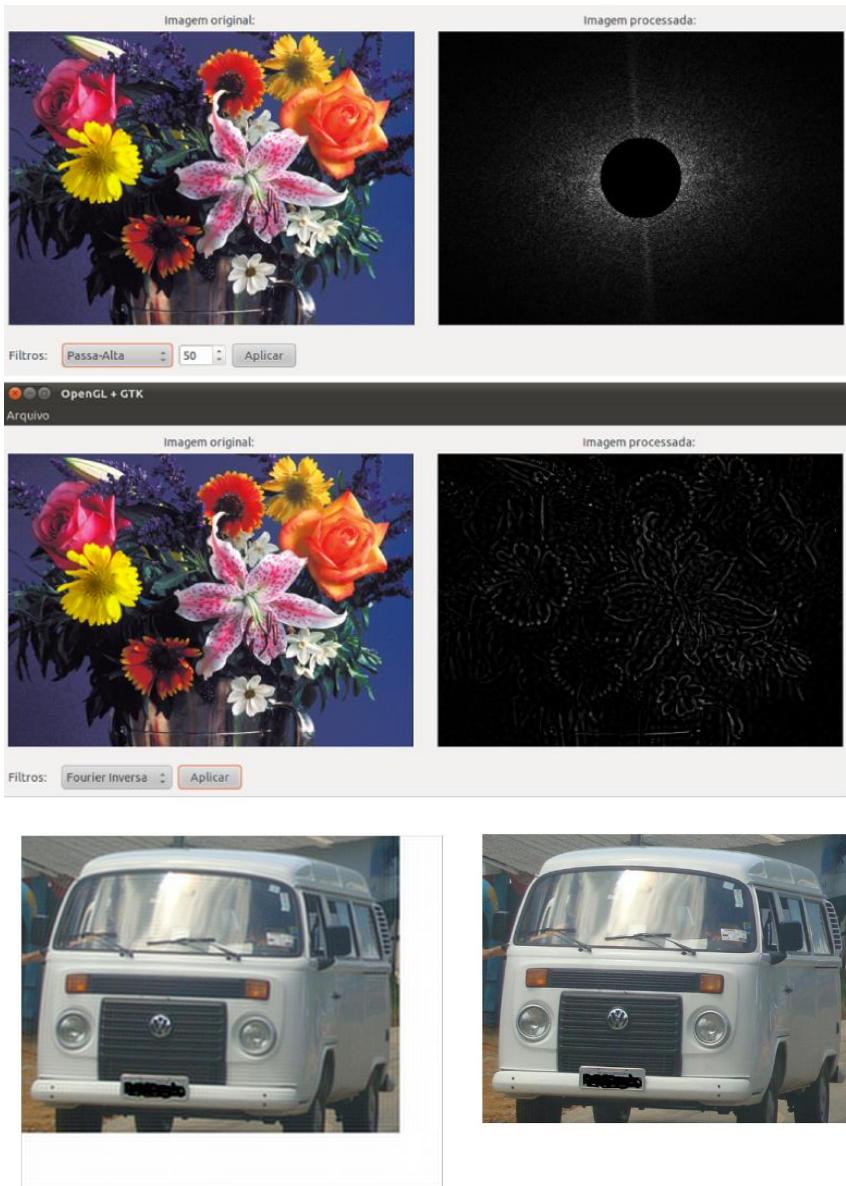
Fourier Spectrum



Local Reject Filter



Band Reject Filter

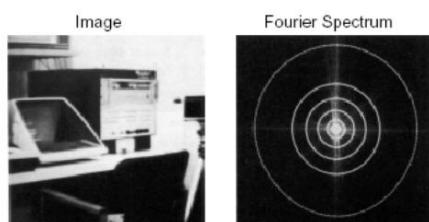


In questo caso vi è una immagine ricca di dettagli. Osservando lo spettro di Fourier, si può notare come c'è stata l'applicazione di un filtro passa alto, facendo transitare soltanto le alte frequenze, ovvero solo i contorni. Il bagliore risulta essere molto ampio dato il numero elevato di dettagli, ed essi vengono evidenziati tagliando le basse frequenze (azzerandole). Ottenendo come risultato i contorni dell'immagine iniziale.

Frequency Bands

È possibile calcolare la percentuale di **energia** dell'immagine **che si conserva dell'immagine** se si va a costruire un filtro progressivo. Più ci si allontana dalle basse frequenze, minore è il contributo informativo che quelle frequenze danno all'occhio umano.

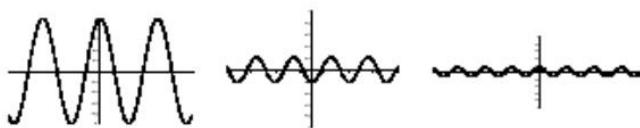
- Percentage of image power enclosed in circles (small to large):
90, 95, 98, 99, 99.5, 99.9



Fourier Transform

- Computing Time
 - $O(n^2)$
 - Fast Fourier Transform (FFT)
 - $O(n \log n)$

What is FT ?



FT decomposes a function into a weighted sum of sinusoidal functions
=> We can reconstruct the original function:

$$p(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} Fp(\omega) e^{i\omega t} d\omega$$

FT maps a function to its frequencies

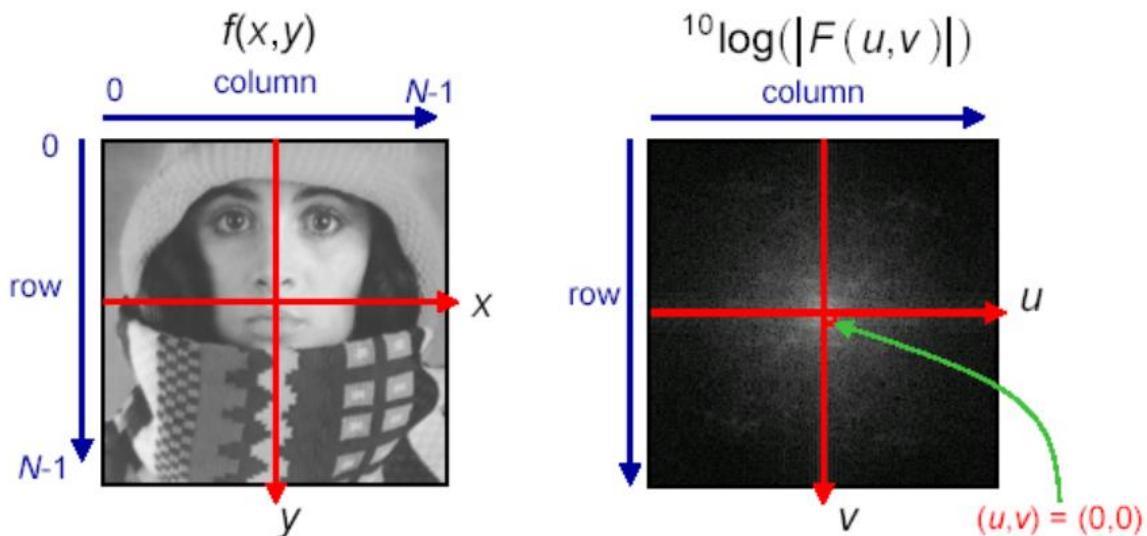
$$\text{Fourier Transform of } p \quad i^2 = -1$$

$$Fp(\omega) = \int_{-\infty}^{\infty} p(t) e^{-i\omega t} dt$$

Angular frequency Continuous function

$$e^{-i\omega t} = \cos(\omega t) - i \sin(\omega t)$$

- $F(u,v)$ is the complex amplitude of the eigenfunction $\exp(j(2\pi/N)(ux+vy))$
Note that $\exp(j(2\pi/N)(ux+vy)) = \cos((2\pi/N)(ux+vy)) + j \sin((2\pi/N)(ux+vy))$
 - Standard display is the logarithm of the magnitude: $\log(|F(u,v)|)$



Fourier spectrum

Linear Systems and Responses

	Spatial Domain	Frequency Domain
Input	f	F
Output	g	G
Impulse Response	h	
Freq. Response		H
Relationship	$g = f * h$	$G = FH$

Si analizzano le analogie tra il dominio delle frequenze e il dominio spaziale (pixel). Il segnale iniziale corrisponde all'immagine e per essa vi è la corrispondente trasformata di Fourier. L'output prodotto nel dominio delle frequenze è un segnale G che opportunamente trasformato con l'anti trasformata di Fourier diventa g . Per costruire il filtro si individuano quali frequenze tagliare e quali conservare;

sulla base di tale scelta viene effettuata la moltiplicazione $G = FH$ e si ottiene G che con l'opportuna anti trasformata si tramuta in g . L'analogia tra i due domini indica che si potrebbe effettuare la stessa operazione nel dominio spaziale, ma non effettuando una moltiplicazione ma bensì una convoluzione tra f ed h dove h dovrebbe essere ottenuto come anti trasformata di Fourier di H . Però mentre H è facile da costruire nel dominio delle frequenze (vettore di 1 e 0 per conservare o tagliare), nel dominio spaziale la costruzione di h è molto complessa, infatti non si ha possibilità di costruirlo, se non facendo l'anti trasformata di H ; in più la relazione tra i due domini non implica la stessa operazione, infatti da un lato c'è la **convoluzione** e dall'altro una **moltiplicazione**.

The Convolution Theorem

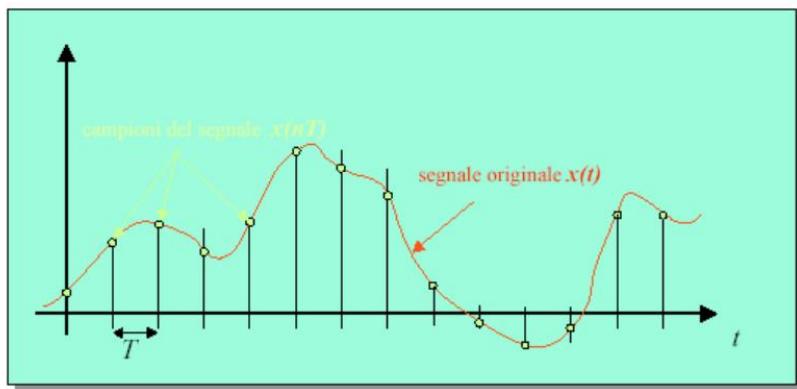
$$\begin{array}{c|c} g = f * h & g = f h \\ \text{implies} & \text{implies} \\ G = F H & G = F * H \end{array}$$

Convolution in one domain is multiplication in the other and vice versa

Campionamento

È il processo attraverso il quale si trasforma un segnale continuo, quindi analogico, in un segnale discreto. Normalmente avendo un segnale continuo nel dominio spazio-tempo, si può individuare un campionamento a passo t , che tende ad individuare un **sottoinsieme di punti** (un insieme discreto e finito di punti) che dovrebbero approssimare il segnale di partenza. Nel dominio spazio-tempo determinare con certezza la dimensione del passo di campionamento risulta essere piuttosto complicato, sapendo che la qualità del segnale discreto ricostruito è condizionata dal passo di campionamento. In genere più è ampio il passo di campionamento minore è la qualità del segnale discreto.

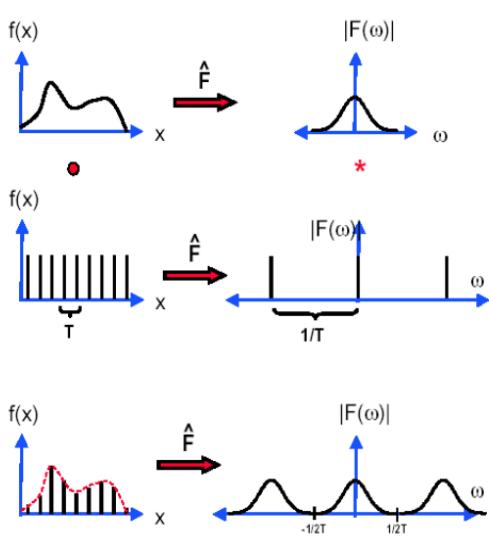
Campionare i segnali (discretizzare nel tempo)



- T è detto **periodo (o passo) di campionamento**
- $f_c = 1/T$ è detta **frequenza di campionamento**

Il **teorema del campionamento** permette di determinare un passo di campionamento attraverso il quale si può quantificare la quantità di informazione che si andrà a **perdere**; stabilendo anche la **qualità** dell'informazione che verrà persa. Per fare questo si ha bisogno di effettuare un passaggio per un dominio differenziato rispetto a quello spazio-tempo (che nel caso delle immagini corrisponde a quello dei pixel). Si dovrà quindi passare alla rappresentazione in termini di trasformata di Fourier.

Campionamento immagine

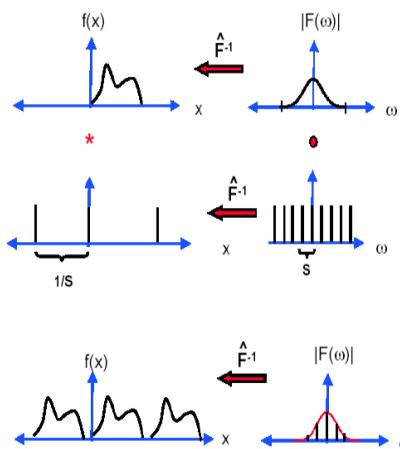


In generale, si immagini di avere un segnale nel dominio spazio-tempo, tutti i grafici a sinistra sono tutti grafici del dominio spazio-tempo; a destra il dominio delle frequenze. Si sa che un segnale qualsiasi nel dominio spazio-tempo può essere trasformato, attraverso Fourier, in una sua rappresentazione ovvero lo spettro di Fourier; esso è la rappresentazione in valore assoluto dei coefficienti assegnati alle armoniche. Se nel dominio spazio-tempo si volesse effettuare un campionamento, si dovrebbero produrre una serie di impulsi unitari che dovrebbero intercettare il segnale al passo t . Più fitta sarà questa serie di impulsi maggiore sarà la probabilità di ricostruire in maniera fedele il segnale. Se si effettua la trasformata di Fourier su questa serie di impulsi a passo t si otterrà nel dominio delle frequenze, quindi come spettro di Fourier,

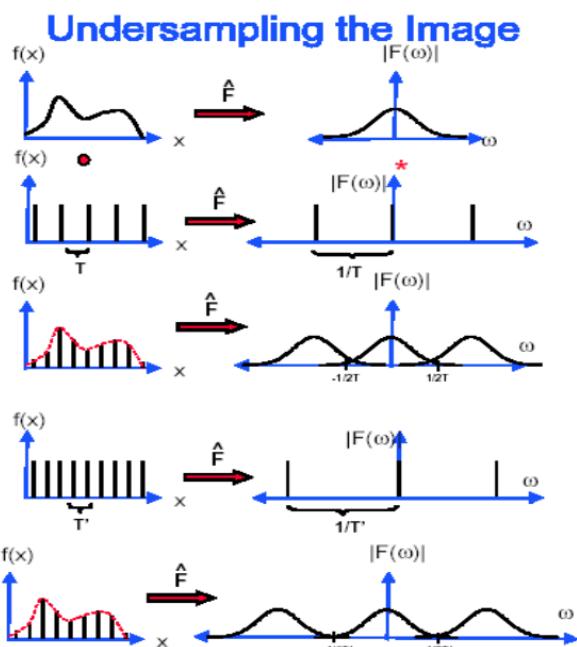
una serie di impulsi a passo $1/t$. Più sarà **piccolo il passo di campionamento nel dominio spazio-tempo, più sarà grande l'intervallo nel dominio delle frequenze**.

Sapendo che tra i due domini, per il teorema di convoluzione, esiste una corrispondenza specifica, effettuare un'operazione di moltiplicazione tra questa serie di impulsi e il segnale nel dominio spazio-tempo, equivale ad effettuare la convoluzione tra il segnale in valore assoluto e la frequenza di campionamento. La convoluzione produce **repliche del segnale originario**, avendo una serie di gaussiane replicate lungo il dominio di riferimento; questo corrisponderà fondamentalmente a quello ottenuto, in termini di moltiplicazione nel dominio spazio-tempo. Le operazioni effettuate verificano la corrispondenza tra l'effettuare un campionamento nel dominio spazio-tempo e mostrare i risultati per le operazioni effettuate, in tale dominio.

- Sampling a function $F(\omega)$ with impulse train of cycle S produces replicas in the image domain with cycle $1/S$



Si mostra ora cosa succederebbe se si volesse effettuare un campionamento nel dominio spazio-tempo o nel dominio delle frequenze.



Nel dominio spazio-tempo dovrebbe essere trovato un passo di campionamento t relativamente piccolo affinché esso garantisca la conservazione del maggior numero di punti sul segnale originario. Più t aumenta, più il numero di punti presi in considerazione diminuisce. Se si effettua la stessa considerazione nel dominio delle frequenze: poiché la trasformata di Fourier del segnale t , produce una serie di campioni a passo $1/t$, nel dominio delle frequenze, più il passo di campionamento nel dominio spazio-tempo risulta essere piccolo, più i segnali unitari risultano essere distanziati. Questo distanziamento, in termini di convoluzione, ha effetto su quanto le armoniche risultano essere intersecate. Più t è piccolo, più le armoniche saranno **disgiunte**, più t è grande più le armoniche tenderanno a **sovraporsi**. Quando le armoniche sono molte sovrapposte, una parte delle

informazioni vanno tra di loro a **confondersi**, ovvero le informazioni che risiedono nelle alte frequenze come contorni e dettagli, risultano essere ambigue (un punto corrisponde a due punti di due armoniche), quindi l'ideale sarebbe poter ottenere delle curve che non vanno ad intersecarsi ma che siano **contigue**. Per stabilire il giusto passo di campionamento, bisogna stabilire un passo di campionamento:

$$\frac{1}{T} > 2\omega_{\max}$$

Il **teorema di Nyquist** permette di ottenere delle curve che siano contigue senza sovrapposizioni, riducendo anche il problema dell'aliasing, ovvero la non perdita delle informazioni che risiedono nelle alte frequenze e quindi la non perdita di dettagli e contorni.

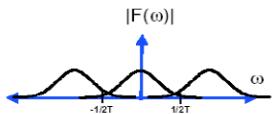
Nyquist Theorem (Sampling Theorem)

- If the maximal frequency of $f(x)$ is ω_{\max} , it is clear from the above replicas that ω_{\max} should be smaller than $1/2T$.

- Alternatively:

$$\frac{1}{T} > 2\omega_{\max}$$

- Nyquist Theorem:** If the maximal frequency of $f(x)$ is ω_{\max} the sampling rate should be larger than $2\omega_{\max}$ in order to fully reconstruct $f(x)$ from its samples.
- If the sampling rate is smaller than $2\omega_{\max}$ overlapping replicas produce **aliasing**.



Dalle proprieta' della trasformata di Fourier (moltiplicazione per esponenziali complessi, oppure convoluzione delle trasformate) e' immediato verificare che il segnale campionato ha come trasformata di Fourier la ripetizione periodica della trasformata $X(f)$ del segnale continuo $x(t)$, con periodo pari alla frequenza di campionamento $f_c=1/T$, moltiplicata per $f_c=1/T$

$$X_c(f) = \frac{1}{T} \sum_{k=-\infty}^{\infty} X(f - k/T)$$

Se le infinite repliche traslate in frequenza di $X(f)$ non si sovrappongono e' facile estrarre mediante filtraggio $X(f)$ da $X_c(f)$, cioe' riottenere $x(t)$ da $x_c(t)$. Ovviamente cio' richiede una conoscenza a priori della banda B occupata da $x(t)$. Tale valore e' disponibile, o facilmente misurabile, nella grande maggioranza dei casi di interesse pratico.

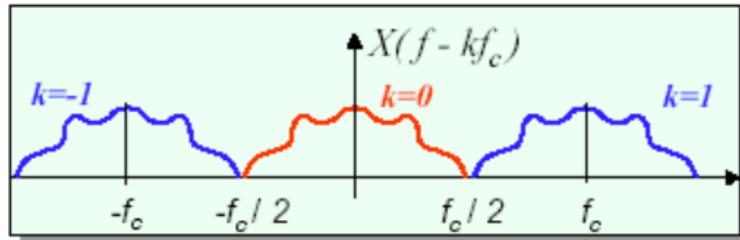
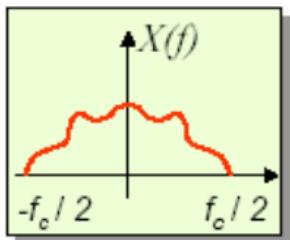
Teorema del campionamento

Se **e' noto a priori** che il segnale tempo continuo $x(t)$ non contiene frequenze maggiori di $f_c/2$ e inferiori a $-f_c/2$, esiste un legame univoco tra il segnale continuo nel tempo e i suoi campioni $x(nT)$.

Se un segnale $x(t)$ e' campionato con frequenza di campionamento f_c almeno doppia della massima frequenza contenuta e' perfettamente ricostruibile (*le repliche in frequenza sono disgiunte*). Altrimenti le repliche sono sovrapposte e vi sono frequenze alle quali non e' possibile distinguere tra repliche diverse.

$$x(t) \xrightarrow{\text{TF}} X(f)$$

$$x(t) \exp(j2\pi f_c t) \xrightarrow{\text{TF}} X(f - kf_c)$$



Come mostrato in figura, questa è la situazione ideale dove nel dominio delle frequenze si dovrebbe ottenere **repliche della rappresentazione** dello spettro di Fourier che siano contigue, scegliendo come frequenza di campionamento quella che consente la **non sovrapposizione** nelle alte frequenze. Il teorema di Nyquist permette di individuare un passo di campionamento che evita la perdita di informazioni, stabilendo quale valore di campionamento utilizzare nel dominio delle frequenze, tale da ottenere un numero discreto di punti che conservino tutte le informazioni originarie del segnale. Questo è vero solo se il segnale possiede banda limitata ovvero se esiste un ω_{\max} . Se quest'ultimo non esiste, ovvero le gaussiane non convergono verso zero ma si muovono all'infinito senza mai toccare l'asse delle ascisse, allora è chiaro che non si potrebbe avere una ricostruzione perfetta del segnale nel dominio discreto, e quindi si andrà a dimensionare quali informazioni devono essere tagliate (in genere quelle nelle altissime frequenze). Il processo di campionamento nel dominio delle frequenze permette di stabilire che se il segnale ha banda limitata, si può campionare il segnale senza perdita di informazioni, se viceversa il segnale non è a banda limitata, quindi le frequenze tendono all'infinito, allora si possono quantificare le informazioni che andranno perse. Questo permette di stabilire quante informazioni perdere e la qualità di esse.

Content Based Image Retrieval (CBIR)

Sono sistemi che si sono sviluppati con l'aumento della mole di immagini e video oltre il quale una **indicizzazione** non è più fattibile. Si tratta di sistemi dove il recupero delle immagini viene effettuato sulla base del contenuto. Quindi, un CBIR, è sistema il cui obiettivo è quello di consentire il recupero di immagini da un database che siano **rilevanti** rispetto ad una query; il sistema è in grado di fornire come risposta un insieme di immagini che possiedono delle caratteristiche di **similarità** con le immagini che si è utilizzate come query o con la query stessa. Il concetto di similarità può avere diversi significati quindi bisogna capire se si fa riferimento all'intera immagine o ad una parte dell'immagine stessa. Trovare una immagine che sia simile ad una query è concetto astratto che non sempre si riesce a codificare precisamente.



Le immagini nel quadrato in alto sono ipotetiche query, e lungo l'asse orizzontale vi sono una serie di immagini che presentano delle caratteristiche simili all'immagine di partenza.

Le analogie di similarità sono molto forti tra loro, ma bisogna generalizzare all'ambito applicativo nel quale viene effettuata la query.

I CBIR realizzano i **match inesatti**, dove si intende che la risposta è individuabile in un insieme di immagini ordinate per similarità dove l'utente effettua una successiva scelta. Quindi a fare la differenza sul concetto di similarità, spesso oltre ad essere l'utente è l'ambito applicativo in cui si opera.

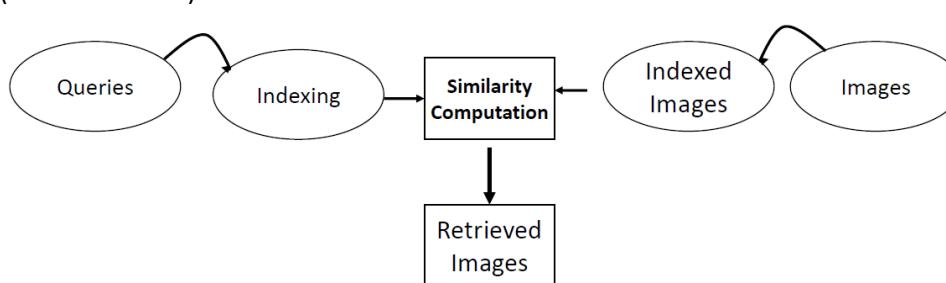
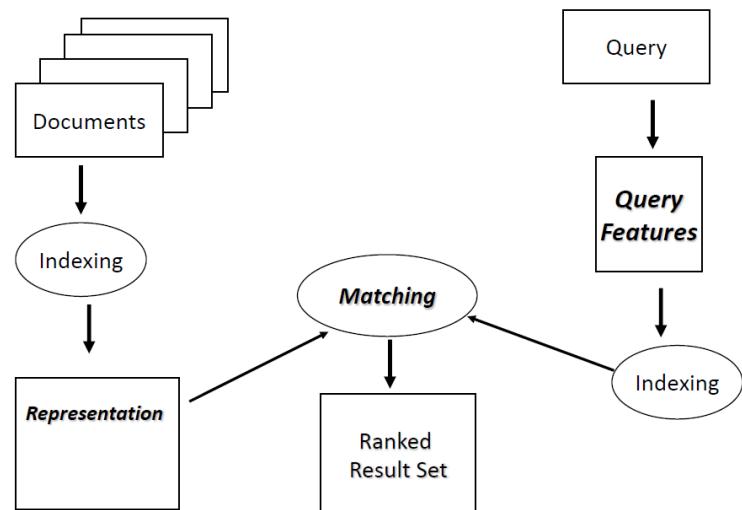
In alcuni casi il concetto di similarità potrebbe essere legato, non solo all'intera immagine, ma ad una parte di essa, per esempio la distribuzione dei colori in una parte dell'immagine, e quindi verificare se vi sono delle immagini che possiedono la stessa distribuzione di colori rispetto ad un'area specifica dell'immagine della query.



Viene analizzata ora la differenza dell'indicizzazione, l'interrogazione e il recupero in un CBIR rispetto a sistemi tradizionali in cui viene utilizzato il **text retrieval** (il sistema recupera solo il file con quel dato nome). Ovviamente, il processo di indicizzazione basato su contenuto è molto più complesso, in quanto va specificata la realizzazione di un query, successivamente come indicizzare e rappresentare un oggetto multimediale e infine come costruire l'architettura di un CBIR.

Il classico sistema di information retrieval è un sistema abbastanza schematico, dove ci sono un insieme di documenti dove ciascuno di essi viene opportunamente indicizzato e rappresentato sotto forma di metadati all'interno di un database. L'indice viene associato al documento e ne rappresenta il suo metadato rispetto al quale vengono effettuate le ricerche.

Durante la fase di query, viene posta appunto la query: in questo caso l'oggetto è stato indicizzato in forma testuale (la query è testuale) l'indice rappresentato attraverso le tecniche di matching viene individuato nella sua rappresentazione e viene fornita poi la risposta che può essere composta di una singola immagine (match esatto) o una serie di immagini (match inesatto).



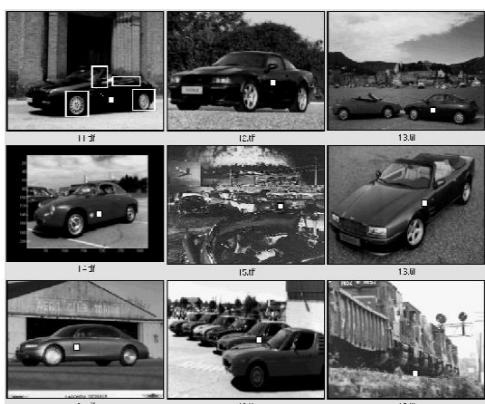
Il concetto di CBIR è analogo al precedente, l'unica cosa che varia nel è il concetto di match inesatto rispetto al precedente, dove vi è un match esatto. Infatti, non si vuol trovare un'unica immagine, ma si cerca un concetto di similitudine, ovvero un insieme di risposte nel quale poter navigare e cercare poi le immagini che sono più pertinenti rispetto al contesto.

Nel contesto dei Text retrieval si sa che indicizzare una foto/video, il testo è fondamentalmente oggettivo, associando alla parte testuale un significato semantico.

Il concetto di immagine è più complesso, innanzitutto perché i pixel non hanno un significato semantico in quanto ad un livello di luminosità non è possibile associare un significato semantico, in quanto esso dipende dal contesto dell'immagine, e non dal livello di luminosità. Nell'**Image Retrieval** è possibile essere interessati a sottoinsieme dell'immagine (parti di essa), quindi effettuare una segmentazione di tali oggetti per poterli descrivere. La divisione di immagini in oggetti semanticamente significativi è un'operazione abbastanza complessa. Infine, la descrizione di un oggetto visuale dipende da molti fattori come il punto di vista, l'illuminazione ecc... un oggetto può anche cambiare dal punto di vista visuale a seconda del contesto in cui è individuato.

Cos'è la Similarità

La prima cosa che deve essere definita quando si utilizza un CBIR è il concetto di similarità, in quanto esso permette di concentrare l'attenzione sulle caratteristiche da estrarre da un'immagine.



Nel caso delle auto il concetto di similarità potrebbe essere il modello, il colore dell'auto, ecc ... oppure si cerca una caratteristica **globale** o **locale**, quindi si è interessati alle ruote, i finestrini, sostanzialmente ad un particolare aspetto del design oppure all'intera immagine. Tutto ciò ha un'importanza in quanto la definizione di similarità è legata alla definizione delle caratteristiche che si vogliono estrarre, o al concetto di **metrica** per misurare se due immagini sono distanti o simili, e all'individuazione delle **non-metric distance**, quindi misure di similarità che non si basano su metriche tradizionali.

Il primo elemento considerato quando si introduce il concetto di similarità è che le misure di similarità introducono due elementi di disturbo:

- **False Alarms**: immagini non significative inserite nell'insieme risposta (answer set).
- **False Dismissals**: immagini significative non inserite nell'insieme risposta, ritenuto anche il *più grave*.

Quindi la progettazione di un CBIR deve essere fatta in modo tale da limitare i due problemi appena elencati ma in particolare i False Dismissals, perché se un'immagine non viene recuperata, essa può tramutarsi in un problema serio. Nel caso di un False Alarms si può scartare l'immagine non pertinente al contesto.

Image Features

L'immagine è un concetto piuttosto complesso il quale può essere interpretato a vari livelli e ciò dipende da vari fattori. Nelle immagini è possibile immaginare tre livelli di caratteristiche:

- **Caratteristiche primitive** (basso livello): colori, forme, relazioni spaziali, ecc...
- **Caratteristiche derivate o logiche**, dove si cerca un certo oggetto all'interno dell'immagine come un volto, o un oggetto di certa forma ecc...
- **Caratteristiche astratte** che sono in grado di descrivere aspetti astratti come eventi, significati emotivi ecc...

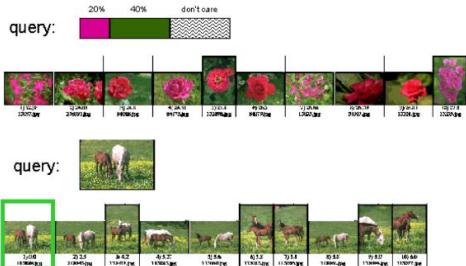
Ovviamente più si passa dalle caratteristiche primitive a quelle astratte, più è complesso il processo di indicizzazione e quindi estrazione delle caratteristiche.

Image Features: Primitive

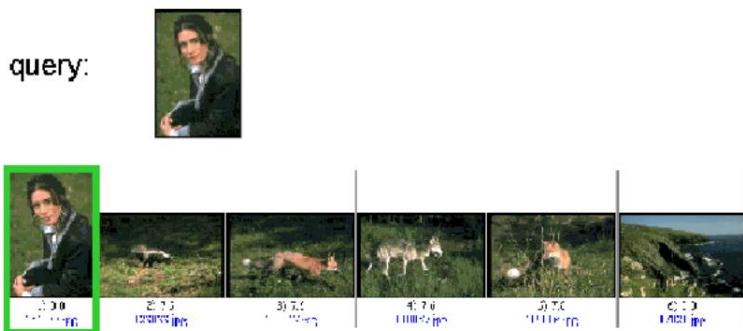
Il livello 1 è il più semplice riuscendo a descrivere in maniera abbastanza efficiente le immagine sottoforma di **caratteristiche primitive**. Si possono immaginare **query elementari** basate sulla ricerca per colore e posizione di un determinato oggetto, oppure tutte le immagini che contengono un elemento all'interno di un altro. O più comunemente si utilizzano delle query dove viene richiesta la ricerca di immagini simili all'immagine query fornita; quindi si parte da un'immagine e si chiede al sistema di recuperare tutte le immagini simili rispetto all'immagine iniziale.

1. Find all images that have 20% of purple and 40% of green
2. Find all images that have 60% of green

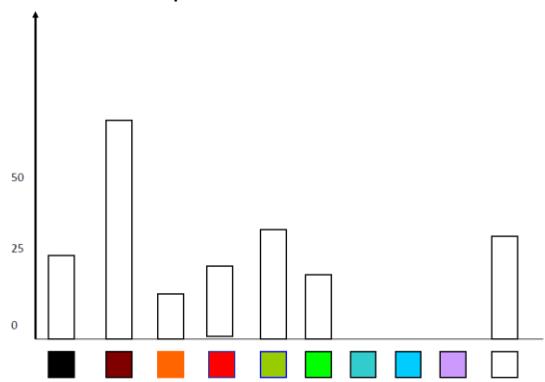
Examples: Color-based Retrieval



Color Retrieval - Bad Example



Pur avendo compreso che i colori o la distribuzione di essi, può essere utilizzata come caratteristiche primitive, in alcuni casi possono produrre dei risultati significativi e in altre occasioni no. Quindi le caratteristiche primitive, in particolare i colori, possono essere utilizzati per discriminare immagini molto diverse tra loro; avendo un database di immagini eterogenee provenienti da vari ambiti applicativi o contrasti sociali, le caratteristiche primitive come il colore possono essere utilizzate per classificare immagini simili, con la forte probabilità di introdurre molti False Alarms. Si analizza l'utilizzo dei colori per effettuare un'indicizzazione: il modo più semplice è utilizzare l'istogramma dei colori, quindi avendo per ogni immagine la propria distribuzione in termini di colori, dove sull'asse delle ascisse vi sono tutti i possibili colori e sull'asse delle ordinate le occorrenze di ciascuno di essi, ottenendo un vettore *n-dimensional* che rappresenterà la chiave di indicizzazione dell'immagine; tale vettore presenta per ogni elemento un colore e il numero associato a quella posizione risulta essere la percentuale con cui quel colore viene rappresentato all'interno dell'immagine.



un'indicizzazione: il modo più semplice è utilizzare l'istogramma dei colori, quindi avendo per ogni immagine la propria distribuzione in termini di colori, dove sull'asse delle ascisse vi sono tutti i possibili colori e sull'asse delle ordinate le occorrenze di ciascuno di essi, ottenendo un vettore *n-dimensional* che rappresenterà la chiave di indicizzazione dell'immagine; tale vettore presenta per ogni elemento un colore e il numero associato a quella posizione risulta essere la percentuale con cui quel colore viene rappresentato all'interno dell'immagine.

Una volta individuate le caratteristiche e la loro rappresentazione, bisogna individuare una modalità per effettuare il match. È possibile utilizzare una metrica euclidea dove viene calcolata una distanza tra le immagini, rappresentate come vettori, nello spazio *n-dimensionale*.

- Euclidean Metric??

- *n-dimensional space*

- Image 1: ($C_{11}, C_{12}, C_{13}, \dots, C_{1n}$)
- Image 2: ($C_{21}, C_{22}, C_{23}, \dots, C_{2n}$)

- Huge Computing Time

- DFT + Euclidean Metrics

- Cut off frequencies (keeping first k coefficients)

- $k < n$

taglio) molto più piccolo di n . Quindi piuttosto che operare su tutti gli elementi, si trasforma il vettore nello spazio delle frequenze ottenendo n frequenze e di esse si portano avanti soltanto le prime k . Con un'operazione simile calcolando una differenza su k elementi, la loro distanza nello spazio delle frequenze sicuramente non aumenterà.

Il **Teorema del Parseval** enuncia che se X è la trasformata di Fourier di una sequenza discreta x (come ad esempio un istogramma) si avrà che:

$$\sum_{i=0}^{n-1} x_i^2 = \sum_{u=0}^{n-1} X_u^2$$

$$D\left(\vec{x}, \vec{y}\right) = D\left(\vec{X}, \vec{Y}\right)$$

where \vec{X}, \vec{Y} are Fourier transform of \vec{x} and \vec{y} respectively

$$D_k\left(\vec{X}, \vec{Y}\right) \leq D_n\left(\vec{X}, \vec{Y}\right)$$

where \vec{X}, \vec{Y} are Fourier transform of \vec{x} and \vec{y} respectively, and $k \leq n$

distanza sulle prime k frequenze, essa è sicuramente minore o uguale della distanza euclidea su n frequenze, e si avrà come risultato che $D_k \leq D_n$. Quindi il teorema di Parseval permette di stabilire fondamentalmente che, introducendo questo passaggio, non si avranno False Dismissals. Ciò significa che se si fosse stato in presenza di immagini simili utilizzando tutte le componenti dell'istogramma, utilizzandone soltanto k la distanza non aumenta e quindi se l'immagine è pertinente (x simile ad y) lo sarà ancora quando si utilizzano solo k componenti. Scegliendo un k molto piccolo, o comunque più piccolo di n , non vengono introdotti False Dismissals rispetto alla ricerca effettuata considerando tutte le n componenti. Il concetto di False Dismissals è legato all'utilizzo di **tutte le componenti**.

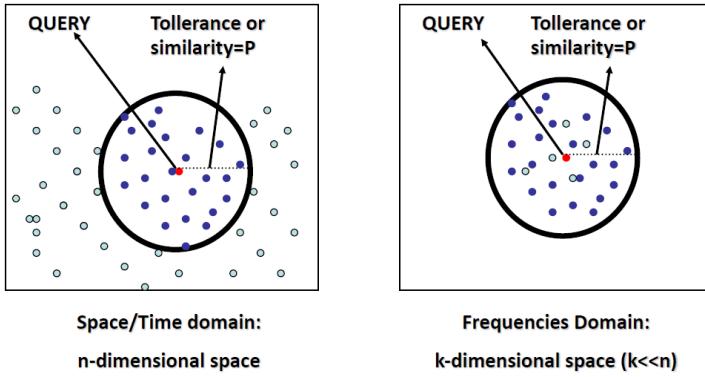
la sommatoria degli n elementi nel dominio spazio-tempo è **uguale** alla sommatoria degli n elementi nel dominio delle frequenze, avendo una **conservazione dell'energia**. Quindi, se si sommano gli n campioni nel dominio spazio-tempo, tale somma sarà uguale alla sommatoria dei coefficienti delle armoniche nel dominio delle frequenze.

Considerando due vettori (due istogrammi) nel dominio spazio-tempo dove D è una distanza euclidea o una qualsiasi metrica, si è a conoscenza che la distanza tra x e y è uguale alla distanza delle rispettive trasformate di Fourier X e Y , con D qualsiasi metrica.

Come detto in precedenza, nel dominio delle frequenze si tenta di tagliare alcune frequenze, conservando solo le frequenze più basse; considerando una metrica euclidea calcolando la

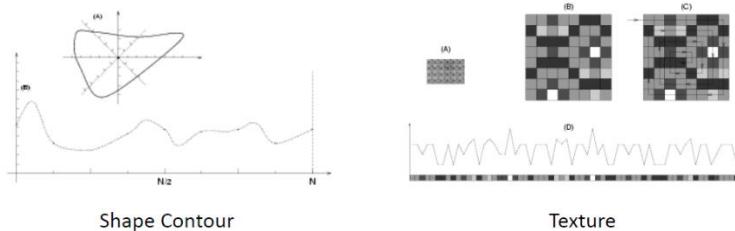
Se il sistema, utilizzando l'istogramma dei colori introduceva dei False Dismissals, lo introdurrà anche utilizzandone k , ma l'idea è di porsi nel caso in cui si considerano tutte le n componenti, si cerca una strategia per non prendere tutte le componenti ma soltanto k e dimostrare che pur prendendone k si ottiene lo stesso risultato che si otteneva prendendo tutte le n . Quindi due immagini simili con n , continuano ad essere simili utilizzando solo k componenti.

- Query
- Included in answer set [Distance from query $\leq P$]
- Not included in answer set (n -dimensional) or false alarms (k -dimensional)



cerchio alcune immagini che prima erano al di fuori di esso. Infatti, considerando una dimensionalità più bassa k , alcuni di quelli che su n sarebbero stati esclusi, adesso con k più piccolo di n potrebbero rientrare, rappresentando i False Alarms. Applicando il teorema di Parseval, passando per Fourier e riducendo la dimensione non si avranno dei False Dismissals ma si potrebbero introdurre dei False Alarms.

Attraverso la trasformata di Fourier si ottiene un ulteriore elemento di applicazione per ridurre la dimensionalità di spazi particolarmente ampi, senza compromettere il risultato finale che si sta cercando.



segna monodimensionale. La forma del contorno attraverso la rappresentazione monodimensionale è composta da un gran numero di punti. Su questo segnale è possibile applicare la trasformata di Fourier e prendere in considerazione soltanto k componenti con k molto piccolo, avendo una rappresentazione del contorno senza introdurre False Dismissals. Un'altra possibilità è quella di poter rappresentare una tessitura con analogo concetto.

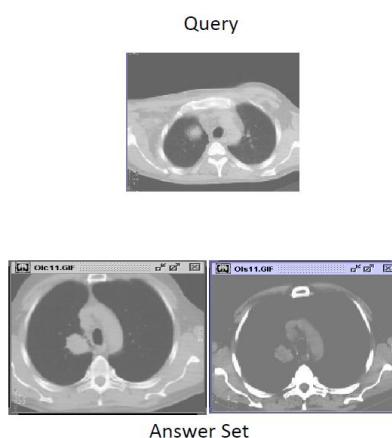
Facendo questa operazione è possibile introdurre dei False Alarms, in figura una simulazione nello spazio bidimensionale. Quando viene diminuita la dimensionalità dello spazio, le immagini che in precedenza non erano significative che si trovavano all'esterno del cerchio rappresentato, ora diventano significative, e tutte le similarità che erano presenti in precedenza continuano ad esserci, ma abbassando la dimensionalità potranno entrare nel

Altre applicazioni del teorema di Parseval combinato con la trasformata di Fourier, possono essere trovate nel caso in cui si voglia rappresentare la forma di un oggetto. Essa può essere rappresentata considerando il baricentro dell'oggetto e considerando le distanze di ciascun punto del contorno dal baricentro, ottenendo un

Image Features: Logical

Le caratteristiche logiche risultano esse più complesse da estrarre da un'immagine, infatti esse richiedono tecniche di computer vision, spesso realizzate attraverso una conoscenza specifica del dominio di riferimento. I CBIR utilizzano poco questo tipo di caratteristiche in quanto risultano essere complesse e computazionalmente onerose da implementare.

- *Find all CT images that have a symmetric circular nodule in left lung, vertically included between spine and aorta.*

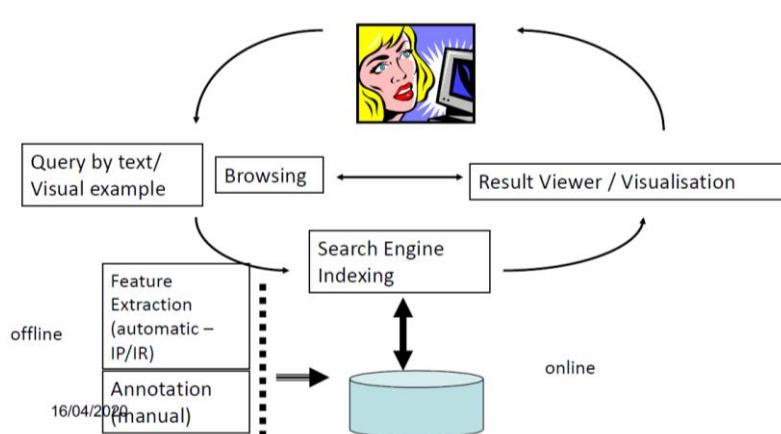


Un esempio di caratteristiche logiche potrebbe essere quello di un CBIR che permette di effettuare query su patologie simili.

Image Features: Abstract

Ancor più complessa risulta essere la fase in cui le immagini vengono indicizzate con caratteristiche astratte. Si passa quindi ad attributi visuali che descrivono situazioni particolari come ad esempio contesti emotivi, culturali ecc.... e richiedono un processo di elaborazione dell'immagine piuttosto complesso.

Architecture of Visual Information retrieval system – New generation



In figura viene mostrata l'architettura di un sistema CBIR di nuova generazione. Concettualmente non è diversa da un sistema tradizionale di archiviazione, con l'unica differenza che la query può essere posta o sotto forma di esempio o sotto forma di immagine. Fornendo al sistema una immagine, ed esso sarà in grado di estrarre determinate caratteristiche dall'immagine e di ricercare all'interno del database immagini le cui caratteristiche sono

simili a quelle poste dall'immagine iniziale. Quindi il processo di annotazione visuale risulta essere automatico e non supervisionato dall'utente. Molti sistemi combinano i due aspetti sia automatico che manuale.

Query Paradigms

I paradigmi che vengono utilizzati sui CBIR possono essere due:

- **Query by Example (QbE):** l'utente fornisce un'immagine come query.
- **Query by Sketch (QbS):** l'utente costruisce lo sketch dell'immagine.

Measures for CBIR

L'efficienza di un CBIR viene misurata in vari modi, dove ogni misura dipende spesso dal dominio di applicazione:

- **Efficienza**: si misura con il tempo di risposta medio per effettuare un recupero.
- **Efficacia**: riduzione della quantità di false Alarms e false Dismissals.
- **Minimizzare False Alarms**.
- **Minimizzare False Dismissals**.

Per quanto riguarda i False Alarms e i False Dismissals esistono due misure che sono:

- **Recall**: capacità del sistema nel recuperare tutte le immagini significative.
- **Precision**: la capacità del sistema nel recuperare solo le immagini significative.

Queste due misure vengono calcolate nel seguente modo: si supponga di avere una collezione di immagini (database) con TOT elementi, e per ogni possibile query (supervisione di un esperto) si associa a ciascuna delle immagini un Rango che va da 1 a REL , dove REL rappresenta, per ogni possibile query che può esser effettuata, il **grado di similarità**. Infatti, se rispetto ad una query si associa ad una immagine un rango 1, essa è più simile alla query rispetto ad un'immagine con rango 2 ecc... più **cresce il rango** più l'immagine risulta essere **difforme** in termini di similitudine dall'immagine query.

Ideal Rank (IR)

$$IR = \sum_{r=1}^{REL} \frac{r}{REL}$$

Per ogni query si avranno un certo numero di immagini rilevanti. **IR** dovrebbe essere un valore molto piccolo.

Average Rank (AR)

$$AR = \sum_{r=1}^{REL} \frac{Rank_r}{REL}$$

- **Effectiveness = $(AR - IR)$**

– $[0; (TOT-REL)]$

- 0 ($AR = IR$) Perfect Retrieval
- $(TOT - REL)$ Worst Case

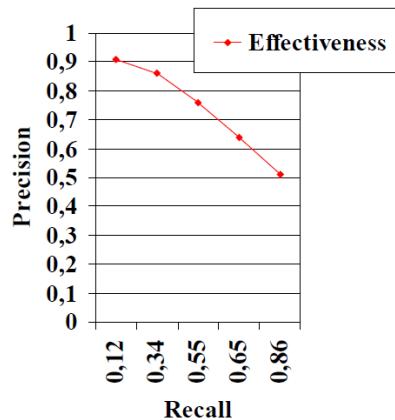
La **Normalized Recall** è una misura efficace per misurare il comportamento di un CBIR. È possibile misurare CBIR differenti e controllare mediamente quali sono i valori di NR prodotti.

- **Normalized Recall (NR)**

– $[0;1]$

$$NR = 1 - \frac{(AR - IR)}{(TOT - REL)}$$

- Recall vs Precision



numero di immagini significative che si inseriscono. Aumentando l'errore, inoltre, succede che viene inserita anche informazione inutile. Ciò fa in modo che la precisione tende a diminuire. Quindi con valori di Recall piccoli, la precisione è molto alta, infatti si introducono pochi False Alarms. Ma quando la Recall tende ad aumentare, siccome si aumenta la tolleranza, la Recall cresce in quanto più abile nel recuperare tutte le immagini ma la Precision ne risente e diminuisce. Questo rapporto permette di stabilire l'efficacia di un sistema CBIR. Un sistema CBIR ideale dovrebbe mantenere una Precision piuttosto alta, man mano che la curva tende a scendere verranno introdotti molti False Alarms.

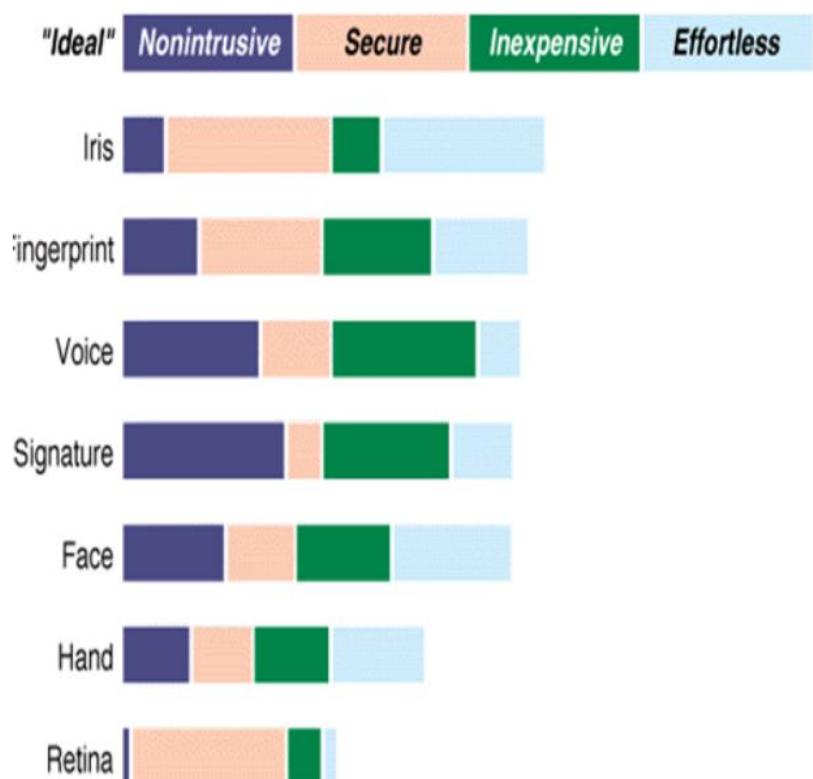
Quando si interroga un CBIR bisogna introdurre una **tolleranza**, ovvero un errore che si vuol commettere.

Per capire se il sistema è in grado di recuperare tutte le immagini significative, bisogna aumentare progressivamente l'errore. Aumentando l'errore, la capacità del sistema nel recuperare tutte le immagini significative, ovviamente aumenta, e aumentano anche il

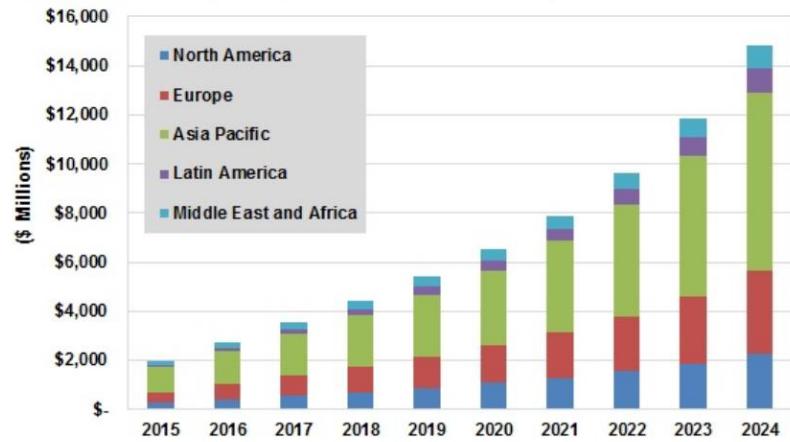
Mercato della Biometria: Panoramica sulle Principali Biometriche

Comparazione di Tecnologie Biometriche

	Universality	Distinctiveness	Permanence	Collectability	Performance	Acceptability	Unspoofability
Fingerprint	M	H	H	M	H	M	M
Iris	H	H	H	M	H	L	H
Retina	H	H	M	L	H	L	H
Hand Geometry	M	M	M	H	M	M	M
Palmprint	M	H	H	M	H	M	M
Hand Vein	M	M	M	M	M	M	H
Voice	M	L	L	M	L	H	L
Face	H	L	M	H	L	H	L
Face Therm.	H	H	L	H	M	H	H
DNA	H	H	H	L	H	L	L

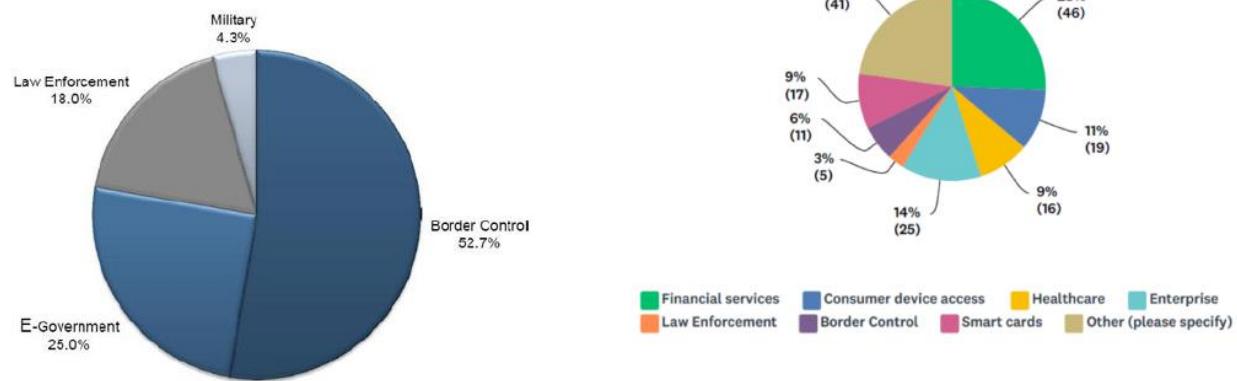


Annual Biometrics Revenue by Region, World Markets: 2015-2024



Source: Tractica

Percent Sales Breakdown
Total Civil and Military Biometrics Market
Global, 2010



Biometrics	Vendors	Market Share	Applications
Fingerprint	54%	34%	Law enforcement; civil government; enterprise security; medical and financial transactions
Hand Geometry	5%	26%	Time and attendance systems, physical access
Face Recognition	23%	15%	Transaction authentication; picture ID duplication prevention; surveillance
Voice Authentication	10%	11%	Security, V-commerce
Iris Recognition	3%	9%	Banking, access control

Fingerprint

Le caratteristiche globali non sono sufficienti per il riconoscimento, infatti tali caratteristiche vengono utilizzate solo per **classificare** (clustering) le impronte (divisione in classi).

Gli archi sono più rari, mentre sono più frequenti i loop e le spirali. Queste caratteristiche, come in figura, sono macro-caratteristiche quindi possono essere condivise da più di una persona e come tali non consentono la discriminazione della persona stessa.



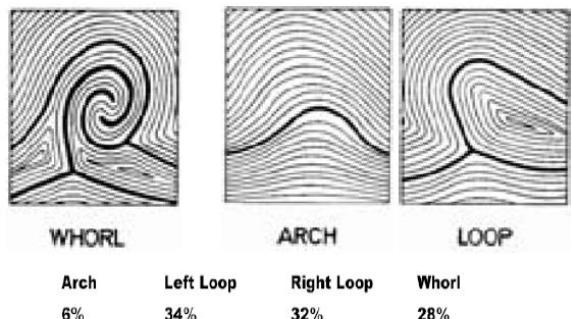
LOOP
In a loop pattern, the ridges enter from either side, re-curve and pass out or tend to pass out the same side they entered.



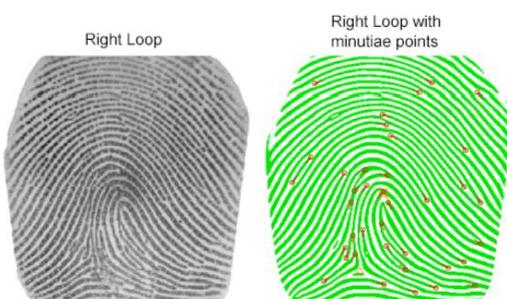
ARCH
In an arch pattern the ridges enter from one side, make a rise in the center and exit generally on the opposite side.



WHORL
In a whorl pattern, the ridges are usually circular.



bifurcation ending bridge lake island



Right Loop
Right Loop with minutiae points
In generale su ogni impronta digitale si possono rilevare fino a 150 minutie. Per dimostrare che l'impronta digitale appartiene ad una determinata persona, delle 150 minutie potenzialmente presenti sull'impronta digitale, in un contesto legale, ne bastano poco più di una decina.

Examples of Ridge Characteristics



Vantaggi:

- Universalità
- Unicità
- Permanenza (Stabilità)
- Misurabilità
- Accettabilità
- Efficacia
- Acquisizione
 - Attiva e Passiva
 - Basso costo dei sensori

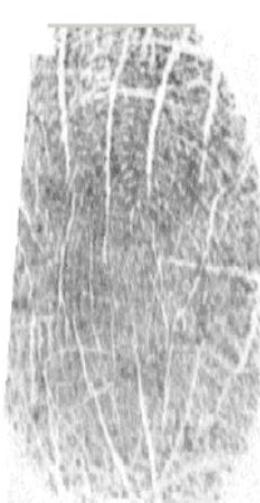
Svantaggi:

- **Efficienza**
 - memoria della chiave di ricerca: 256 bytes – 1.2k
 - Risoluzione: 500 dpi
 - Tempo di ricerca
- **Stabilità**
 - Abrasioni, Rughe...
 - Sporco, Sudore...
 - Disidratazione (over 65)
- **Insidia**
 - Relativamente facile da duplicare (Physical spoofing)

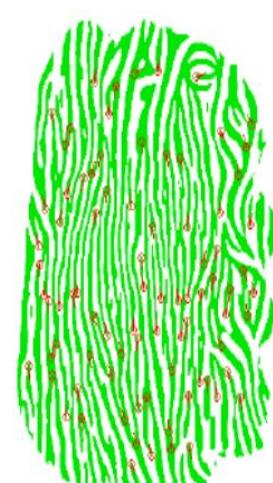
Fingerprint: Stabilità?



Example image of a 22 year old subject



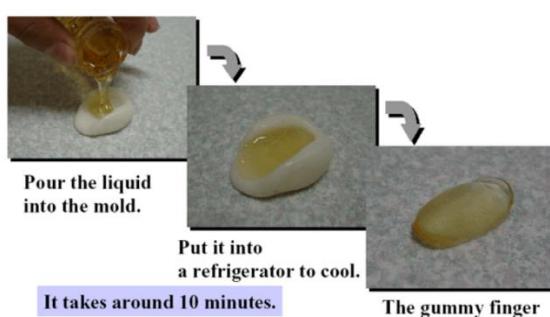
Example image of an 81 year old subject



Physical spoofing: Metodo di Matsumoto



- Only a few dollars' worth of materials



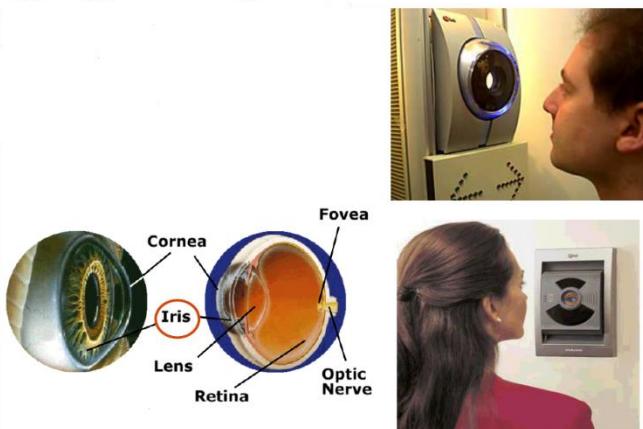
You can place the “gummy finger” over your real finger. Observers aren't likely to detect it when you use it on a fingerprint reader.

Iride

L'iride ha caratteristiche dal punto di vista fisico, molto simili a quelle delle impronte digitali, come delle minuzie con una distribuzione randomica sulla sua superficie. Essa è la zona colorata intorno alla pupilla e la struttura complessa ne costituisce la chiave di autenticazione.

La parte colorata dell'occhio umano, con le sue circa 300 caratteristiche misurabili, è forse l'unica caratteristica fisica maggiormente peculiare di un individuo:

- Non è possibile di cambiamenti nel corso del tempo.
- Non può essere modificata artificialmente.
- La probabilità di trovare sulla Terra due iridi uguali è praticamente nulla (1 su 10 seguito da 78 zeri).
- Ogni iride umano ha infatti una struttura unica, al punto che persino l'iride destra e sinistra della stessa persona sono differenti.
- Impossibile da rigenerare artificialmente, ma può essere asportata chirurgicamente e trapiantata.
- Lenti a contatto e traumi non hanno effetto sulla sua morfologia.



- È situato in una posizione più esterna alla retina, per cui i metodi per la scansione di essa risultano essere più semplici e meno invasivi.

Metodi di rilevazione dell'iride

- Un sensore, collocato a circa 40 centimetri di distanza dalla persona esaminata, inizia a fotografare i margini dell'occhio; quindi, attraverso un certo numero di scansioni successive, si individuano i contorni dell'iride come una corona circolare.
- Poi si seleziona un quadratino alla volta di quest'area che rappresenta l'area da decodificare. Il disegno di questa regione è convertito in un codice di 512 byte, il cosiddetto Iris Code che, confrontato con quelli archiviati nel database, identifica o meno la persona.

Vantaggi:

- Universalità
- Unicità
- Permanenza (Stabilità)
- Misurabilità
- Efficacia
- Insidia
 - Physical spoofing: mediante chirurgia

Svantaggi:

- **Efficienza**
 - Memoria della chiave di ricerca: 256–512 bytes
 - Risoluzione: 200/500 dpi
 - Tempo di ricerca
- **Accettabilità**
- **Acquisizione**
 - Attiva
 - Elevato costo dei sensori

Retina

La retina si trova nella parte interna dell'occhio, e la sua scansione avviene attraverso la mappatura dei capillari sul fondo oculare.

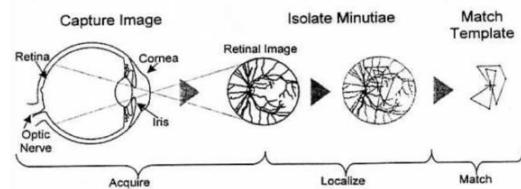
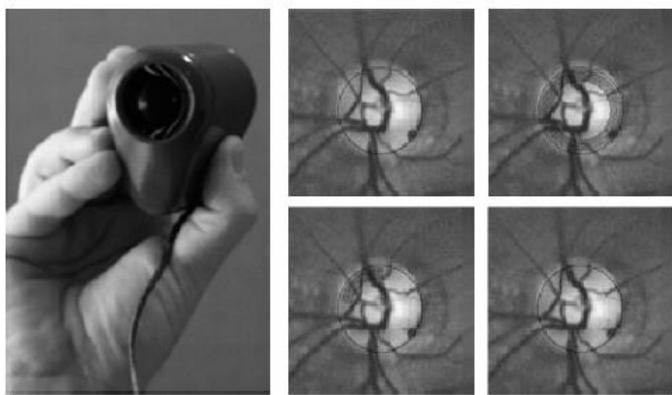
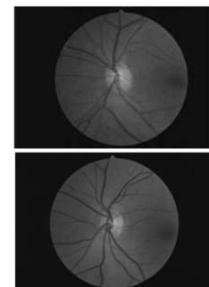


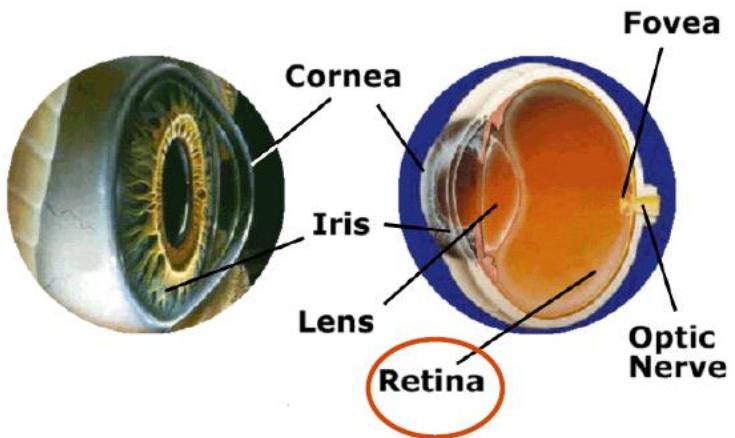
Exhibit 11-6. Retinal recognition process.



È basato sull'acquisizione e la verifica della mappa vascolare della retina dell'occhio umano: ci sono circa 180 caratteristiche misurabili.

Simon e Goldstein nel 1935 mostrarono che la forma delle vene nella retina sono uniche per ogni individuo. La rete vascolare della retina è una delle caratteristiche che presenta un indice di variabilità nel tempo moderatamente basso. È situata in una posizione più interna rispetto all'iride, per cui i metodi per la scansione sono più invasivi. Impossibile da rigenerare artificialmente, ma può essere asportata chirurgicamente e trapiantata. L'acquisizione delle proprietà della retina utilizza un raggio di luce che effettua la scansione della retina. I raggi riflessi producono un'immagine in intensità della struttura rilevata. Ci sono due metodi differenti di acquisizione della retina:

- Metodo attivo: l'utente deve effettuare dei movimenti con la testa (20-40 cm) affinché il dispositivo di scansione possa individuare la posizione corretta;
- Metodo passivo: l'utente resta fermo e una serie di dispositivi di scansione in cooperazione individuano la posizione corretta.



Vantaggi:

- Universalità
- Unicità
- Permanenza (Stabilità)
- Misurabilità
- Efficacia
- Insidia
 - Physical spoofing: mediante chirurgia
- Efficienza
 - Memoria della chiave di ricerca: 96-256 bytes
 - Risoluzione: 100/200 dpi
 - Tempo di ricerca

Svantaggi:

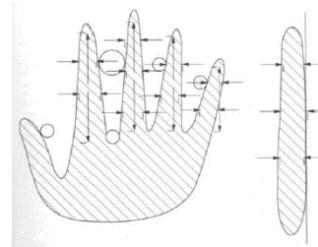
- Efficienza
 - Memoria della chiave di ricerca: 96-256 bytes
 - Risoluzione: 100/200 dpi
 - Tempo di ricerca
- Accettabilità
 - Maggior cooperazione rispetto all'iride
- Stabilità
 - Alcune patologie possono alterare la mappa
- Acquisizione
 - Attiva
 - Elevato costo dei sensori

Geometria della mano

Consiste nel riconoscimento delle persone mediante la verifica delle misure e della conformazione della mano e consente un discreto coefficiente di univocità.

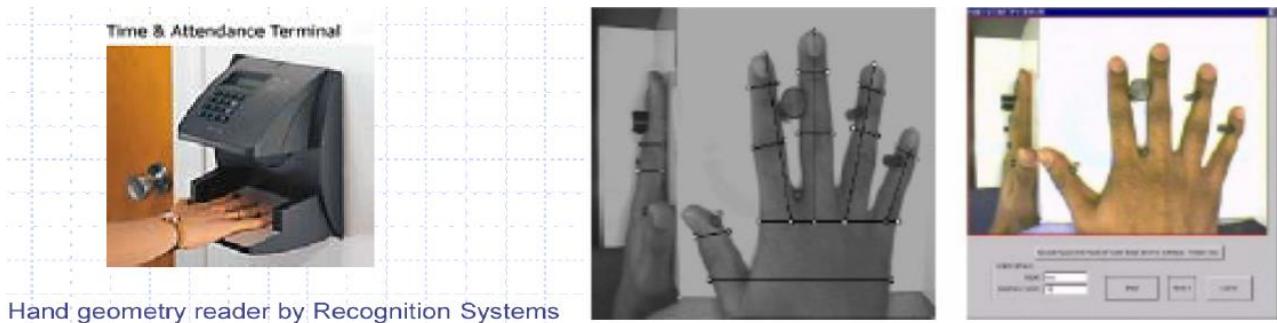
Caratteristiche: forma, locazione e dimensione di mano, dita e nocche. Misura le caratteristiche fisiche della mano:

- Lunghezza delle dita.
- Larghezza della mano.
- Spessore delle dita.



Le misure fisiche della mano sono catturate da un DISPOSITIVO CCD e la sagoma della mano viene memorizzata tridimensionalmente. La mano viene posta su un piano dove vi sono 5 pioli per posizionarla nel modo giusto. Due immagini acquisite:

- Dall'alto.
- Di lato.



Vantaggi:

- Universalità
- Misurabilità
- Accettabilità
- Efficienza
 - Memoria della chiave di ricerca: 10 bytes
 - Risoluzione: 72/100 dpi
 - Tempo di ricerca

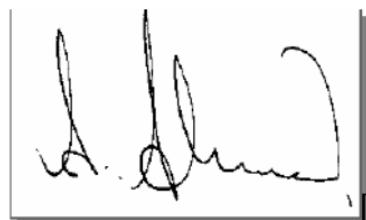
Svantaggi:

- Unicità
- Insidia
 - Relativamente facile da duplicare (Physical spoofing)
- Efficacia
- Stabilità
 - Alcune patologie possono alterare la morfologia
- Acquisizione
 - Attiva

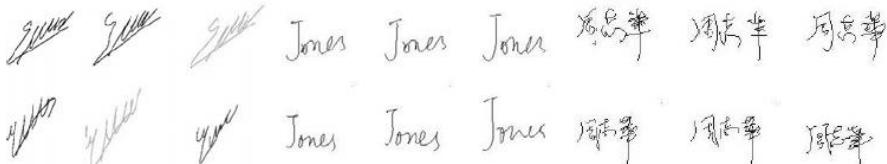
Firma

La firma è una biometria prevalentemente comportamentale, rispetto alle precedenti che risultano essere fisiologiche. È una tecnica riconosciuta ai fini dell'autenticazione, ma è soggetta ad un riflesso condizionato non facilmente riproducibile, soprattutto in real time. La firma è utilizzata anche in processi di riconoscimento. Si definiscono schemi di caratterizzazione per la dinamica dei movimenti durante la firma:

- Angolazione della penna.
- Pressione esercitata con la penna.
- Tempo di esecuzione.
- Velocità e accelerazioni.
- Proprietà geometriche.

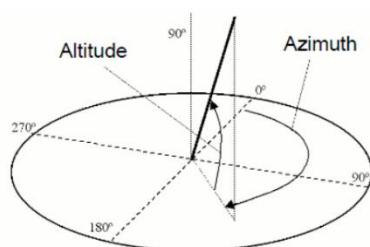
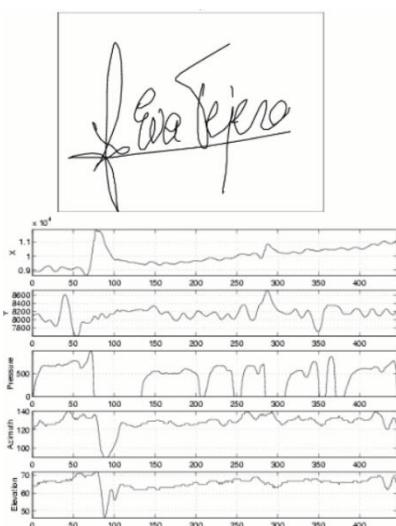


Eccessiva variabilità intra-classe. Può essere dimenticata (a differenza di altre biometriche). Nella modalità di acquisizione offline contano solo le caratteristiche morfologiche (forma geometrica) che sono facili da duplicare. Nella modalità online (più difficile da realizzare mediante hardware speciale) si considerano altre caratteristiche (velocità, accelerazione, inclinazione delle penne...) più difficili da replicare.



- First row = authentics
- Second row = impostors

Alcuni esempi di come possono essere estratte le caratteristiche relative alla firma



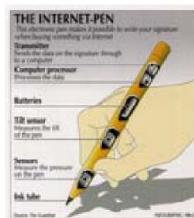
Features:

1. coordinate X
2. coordinate Y
3. pressure
4. pen azimuth ($0^\circ - 359^\circ$)
5. pen altitude ($0^\circ - 90^\circ$)

Alcuni esempi di dispositivi utilizzati per estrarre le informazioni dette in precedenza.



Electronic pen [LCI-SmartPen]



The Internet-Pen



Digitising tablet by Wacom Technologies



Digitising tablet [Hesu Signature Pad] by BS Biometric Systems GmbH

Vantaggi:

- **Universalità**
- **Misurabilità**
- **Accettabilità**
- **Insidia**
 - Difficile da duplicare (modalità on line)
- **Efficienza**
 - Risoluzione: 72/100 dpi
- **Acquisizione**
 - Attiva e Passiva (on line e off line)

Svantaggi:

- **Unicità**
- **Efficienza**
 - Memoria della chiave di ricerca: 500 bytes – 1000 bytes
 - Tempo di ricerca
- **Insidia**
 - Relativamente facile da duplicare (modalità off line)
- **Efficacia**
- **Stabilità**
 - Variabilità intraclasse
- **Acquisizione**
 - Costo dei dispositivi on line (acquisizione attiva)

Riconoscimento Vocale

Determinazione di un modello acustico dipendente dal linguaggio e dal soggetto. Il modello acustico riflette:

- Anatomia (dimensione e conformazione della bocca e della gola),
- Comportamento (timbro di voce, modo di parlare)

Richiede complesse elaborazioni del segnale acustico (analisi di spettro, periodicità, etc....). Le caratteristiche della voce di un parlante sono dovute sia a differenze fisiologiche, sia al particolare stato d'animo in cui si trova. Dipendenti dal Testo: esiste un insieme prefissato di frasi per Enrollment e identificazione. Con digitazione del Testo: esiste un insieme prefissato di parole, ma l'utente ne digita alcune e poi le pronuncia. Indipendenti dal Testo: il soggetto viene riconosciuto a prescindere dalla frase pronunciata.

Vantaggi:

- Universalità
- Misurabilità
- Accettabilità
- Acquisizione
 - Attiva e Passiva
 - Costo dei dispositivi

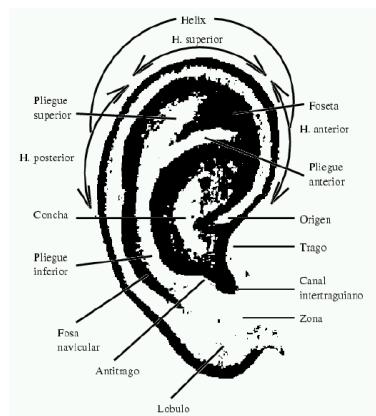
Svantaggi:

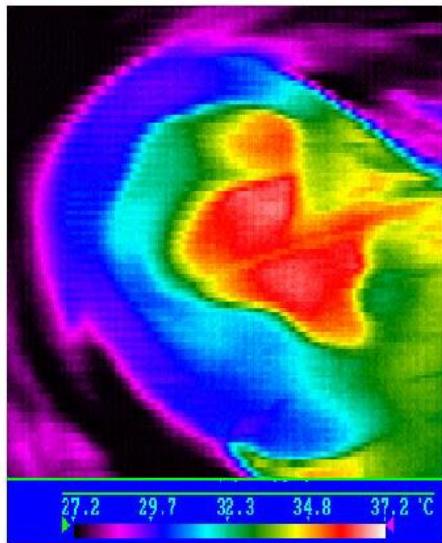
- Unicità
- Efficienza
 - Memoria della chiave di ricerca: 3K-5K
 - Tempo di ricerca
- Insidia
 - Relativamente facile da duplicare (Physical Spoofing)
- Efficacia
- Stabilità
 - Variabilità intraclasse (condizioni fisiologiche, emotività, rumori di fondo etc.)
- Acquisizione
 - Difficile la modalità passiva

Orecchio

L'acquisizione avviene attraverso una serie di caratteristiche uniche che caratterizzano la morfologia di ciascun orecchio. Essa non muta nel tempo. Quindi:

- Mediante foto – il più studiato.
- Mediante punti di interesse – usato per l'autenticazione di criminali.
- Termografia – in fase di sperimentazione.





Termografia



Sistema Iannarelli

Vantaggi:

- Universalità
- Unicità
- Efficacia
- Stabilità
 - Bassa variabilità
- Misurabilità
- Acquisizione
 - Costo dei dispositivi

Svantaggi:

- Efficienza
 - Memoria della chiave di ricerca: 500-1000 byte
 - Tempo di ricerca
- Insidia
 - Media difficoltà nella duplicazione (Physical Spoofing)
- Acquisizione
 - Attiva
 - Parziali occlusioni (cappelli, orecchini...)
- Accettabilità
 - Forte cooperazione dell'utente

Volto

Il volto viene utilizzato sia per la verifica sia per il riconoscimento dell'identità. Acquisizione dell'input:

- Immagine fissa o video.
- Colori o Toni di grigio.
- Tecniche 2D/3D (Verifica):
 - Stereo.
 - Con illuminazione controllata.
- Face Detection (C'è un volto??).
- Face Segmentation (Dov'è il volto (still image)???).
- Face Tracking ((Dov'è il volto (video image)??)).

Face detection & segmentation



Vantaggi:

- Universalità
- Efficienza
 - Memoria della chiave di ricerca: 84 bytes – 2k
 - Tempo di ricerca
- Efficacia
 - Verifica
- Misurabilità
- Accettabilità
- Acquisizione
 - Costo dei dispositivi
 - Attiva e Passiva

Svantaggi:

- Unicità
 - Sosia
- Efficacia
 - Riconoscimento
- Stabilità
 - Variabilità intraclasse
- Insidia
 - Facile da raggirare (Physical spoofing)

Modelli di Colore

L'occhio umano possiede uno spettro visibile, per avere un'idea della gamma di colori che l'occhio umano possa percepire, si fa riferimento alla **luce bianca che attraversa un prisma ottico**.

La luce visibile in realtà non è nient'altro che una serie di onde elettromagnetiche nell'intervallo [380nm, 780nm].

- Ultravioletto <380 nm (<10nm Raggi X, Raggi Gamma, Raggi Cosmici)
- Infrarosso >780 nm (>1500nm onde radio)

La percezione del colore avviene attraverso la combinazione di tre stimoli primari (**tricromia**): a livello della retina viene filtrato il segnale e diviso in tre colori: blu, rosso e verde (verde-giallo).



Il colore di un oggetto è percepito come luce riflessa dall'oggetto stesso. Un oggetto che **riflette** luce in modo omogeneo sull'intero spettro visibile è percepito di colore **bianco**. Un oggetto che **assorbe** luce in modo omogeneo sull'intero spettro visibile è percepito di colore **nero**. Un oggetto verde riflette luce la cui lunghezza d'onda ricade nell'intervallo [500 nm, 570 nm] e assorbe quasi totalmente le altre lunghezze d'onda [380 nm, 499 nm]U[571 nm, 780]. Il nostro occhio, a differenza dell'orecchio, non è in grado di effettuare un'analisi spettrale, ma riporta una sensazione risultante dalla combinazione di tutte le lunghezze d'onda visibili. È inoltre impossibile comunicare la sensazione corrispondente ad un certo stimolo. Quello che possiamo comunicare è che due stimoli diversi producono la stessa sensazione.

La teoria di Young-Helmholtz (1801) ipotizza che per riprodurre la sensazione di un colore si possano combinare opportunamente tre radiazioni primarie (sistemi tricromici). La prima conferma sperimentale ottenuta da esperimenti sull'occhio si ebbe negli esperimenti di Wright & Guild (1928-1931).

Percezione e Colori del CIE (Commission Internationale de l'Eclairage)

Questa organizzazione si preoccupò di classificare i colori, come colori primari:

- Red (700 nm).
- Green (546.1 nm).
- Blue (435.8 nm).

Ne fu classificato anche un quarto:

- $Y = 0.299\text{Red} + 0.587\text{Green} + 0.114\text{Blue}$ (luminanza dello stimolo).

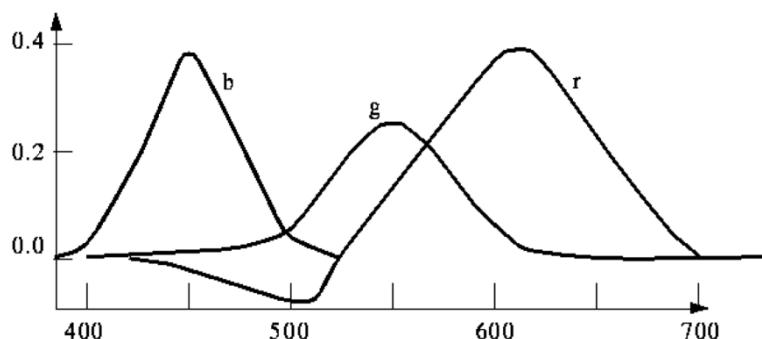
I coni (visione fotopica) sono divisi in tre gruppi:

- L (long) sensibili al rosso.
- M (middle) sensibili al verde.
- S (short) sensibili al blu.

I bastoncelli sono sensibili nello stesso modo (invarianti) alle varie frequenze dello spettro luminoso: producono una visione (scotopica) basata solo sulle differenze di luminosità.

Percentuale (Y) di intensità dei tre primari (RGB) per comporre tutti i colori dello spettro visibile (X) [Wright & Guild (1928-1931)].

I coefficienti negativi indicano che alcuni colori non possono essere rappresentati dalla somma dei tre primari.



Modelli di Colore

Classi di Immagini a Colori:

1. **True colors** (Colori Veri) è ottenuta mediante composizione (sottrattiva o additiva di tre componenti –HSB, RGB, CYM, YIQ–) Ogni componente è quantizzata con un numero definito di bit.
2. **Pseudo-Colors** (Colori Falsi) è ottenuta assegnando ad ogni intervallo di colori veri un colore medio.

Il modello di colore **RGB** è utilizzato per realizzare dispositivi di proiezione quali monitors, TV e nell'elaborazione di immagini. Viene utilizzato anche per immagini satellitari.

- **Additivo:** Si addiziona luce (RGB) al nero

Il modello di colore **CYM** è utilizzato per realizzare dispositivi di stampa.

- **Sottrattivo:** Si sottrae luce (CYM) al bianco

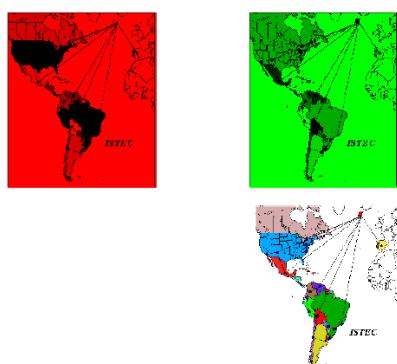
Il modello di colore **HSB** (HSV) è utilizzato nell'Elaborazione di Immagini

- **Combinazione** di Hue (Tonalità), Saturazione (Saturation), Luminosità (Brightness)

Il modello di colore **YIQ** (YUV) è utilizzato nelle trasmissioni TV e nell'elaborazione di immagini

- Sfrutta la maggiore sensibilità dell'occhio umano alla **luminanza** (immagini a livelli di grigio)

Modelli di Colore: RGB



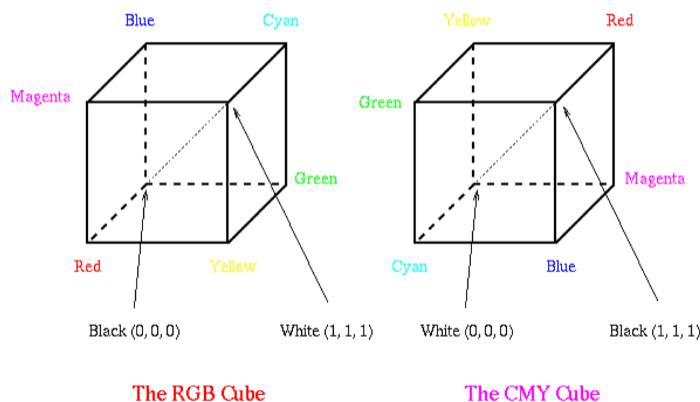
La tabella dei colori che utilizza il modello RGB è:

R	G	B	colore
255	255	255	bianco
255	0	0	rosso
0	255	0	verde
0	0	255	blu
0	0	0	nero

Modelli di Colore: CMY



Relazioni tra i Cubi di Colore

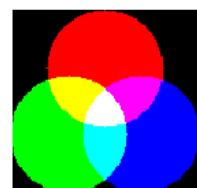


Modelli di Colore

- magenta = rosso + blu.
- ciano = verde + blu.
- giallo = rosso + verde.

Altre possibili combinazioni:

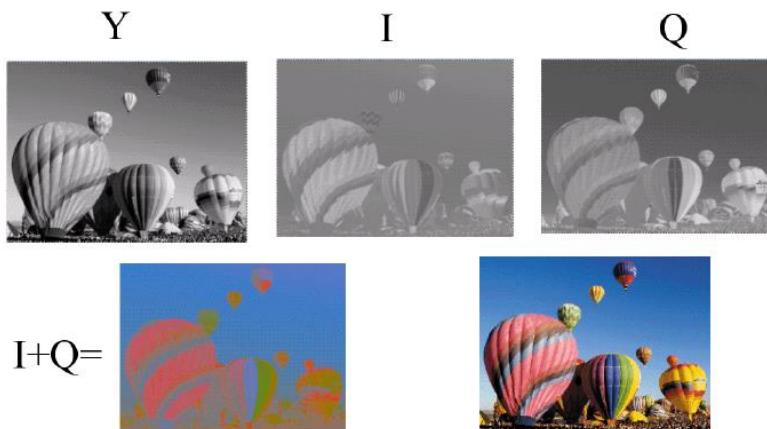
- bianco = blu (primario) + giallo (secondario).
- bianco = verde (primario) + magenta (secondario).
- bianco = rosso (primario) + ciano (secondario).



Modelli di Colore: YIQ

- Y = Luminanza (proporzionale alla quantità di luce percepita dall'occhio)
- I = Inphase
- Q = Quadrature ($I + Q$ = informazioni sul colore)

L'occhio umano è più sensibili alla variazione di luminosità che alle variazioni di tono e saturazione.



Conversione tra i Modelli di Colore

$$\begin{bmatrix} C \\ Y \\ M \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

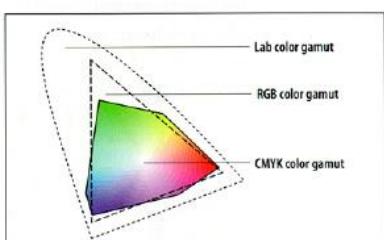
$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.275 & -0.321 \\ 0.212 & -0.528 & 0.311 \end{bmatrix} \times \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

Relazione tra Modelli di Colori

Il **gamut** è l'insieme dei colori che possono essere realizzati dalla combinazione di tre primari. Il modello **LAB** (L = Luminosità, a = asse verde-rosso, b = asse blu-giallo) copre tutti i colori nello spettro visibile.

Il gamut **RGB** è minore del **LAB**, quindi alcuni colori (giallo puro, ciano puro) non possono essere visualizzati sul monitor.

Il gamut **CMYK** è il più piccolo (ma non è un semplice sottoinsieme del gamut RGB).



Modelli di Colore

Quantizzazione:

- Immagini Binarie (1 bpp).
- Immagini in scala di grigio (8 bpp, 16 bpp).
- Immagini a colori (16 bpp, 24 bpp, 32 bpp, 48 bpp, 64 bpp).
- Immagini a pseudo colori o falsi colori (8 bpp).

Modelli di Colore: Scala di Grigio

8 bpp



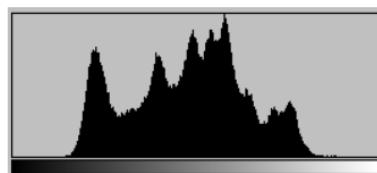
6 bpp



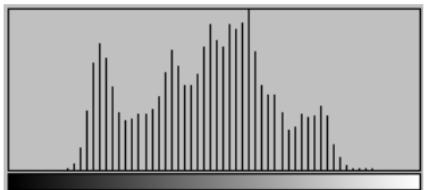
4 bpp



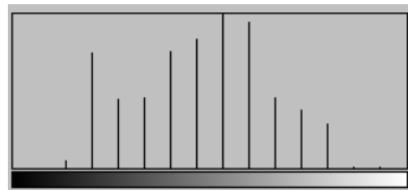
8 bpp



6 bpp



4 bpp



8 bpp



6 bpp

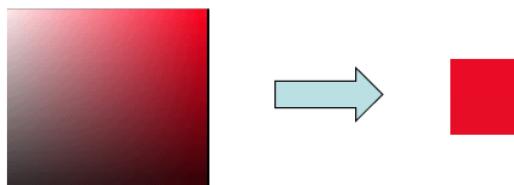


4 bpp



Modelli di Colore: Pseudocolori

I Pseudo-Colors (Colori Falsi) è ottenuta assegnando ad ogni intervallo di colori veri un colore medio.



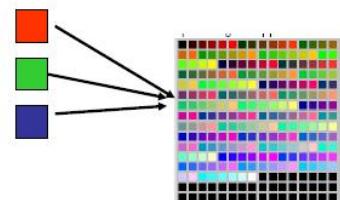
Quantizzazione cromatica

Come la risoluzione spaziale anche la risoluzione cromatica è un compromesso costo prestazioni. Non sempre è necessario avere a disposizione tutti i colori delle immagini truecolor (24 bit per pixel, 8 per ogni canale), specie in applicazioni tipiche della grafica. È quindi possibile risparmiare spazio di memorizzazione e ridurre il carico computazionale, riducendo il numero di colori. Per ridurre il numero di livelli di colore:

- Si usano meno bit per banda (es. 5-5-5 o 5-6-5 nel caso dell'icolor a 15 e 16 bit rispettivamente).
- Si usa una look-up table (mappa di colori): si sceglie un numero finito di colori (es. 256) memorizzati su una tabella ed il valore del pixel è un puntatore a tale tabella che contiene terne RGB.

Selezione della palette

L'elenco dei colori disponibili è chiamato **palette** (più comune dell'italiano paletta, comunque corretto). “Anticamente” la scheda CGA aveva una paletta a 4 colori, l’EGA a 16 colori e la “rivoluzione” si ebbe con la scheda VGA che consentiva una paletta a 256 colori!



look-up table

Palette ottimizzata

È possibile ottimizzare la riduzione dei colori scegliendo una paletta che rappresenti al meglio lo spazio dei colori dell’immagine. Una possibile tecnica per ottenere una paletta ottimizzata è quella proposta da Paul Heckbert nel 1980, nota come Median Cut Algorithm:

1. Si calcola l’istogramma tridimensionale dei colori contenuti nell’immagine;
2. Si trova il minimo parallelepipedo (box) orientato come gli assi che contenga tutti i colori.
3. Si sceglie la diagonale del box e si taglia nel punto che lascia lo stesso numero di elementi nelle due metà (taglio mediano).
4. Si ripete il taglio del box con più elementi (maggior numero di colori) fino a quando non si sono ottenuti tanti box quanti i colori richiesti. I colori rappresentativi saranno dati dalla media dei colori inclusi nel box.

Confronto tra Palette



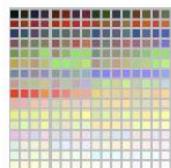
◆ Immagine originale a 24 bit



◆ Immagine a 8 bit con palette standard



◆ Immagine a 8 bit con palette ottimizzata



Modelli di Colore: Pseudocolori

Ris=481x321, Depth=24 bpp, Spazio=463 K



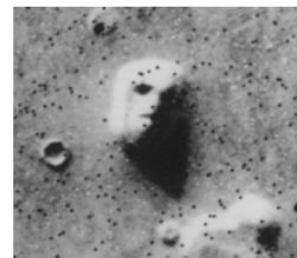
Ris=481x321, Depth=8 bpp, Spazio=154 K



Face Detection

Il volto è una biometria al limite, in quanto essa è soggetta a **cambiamenti temporali**, ma allo stesso tempo è la biometrica con la quale si riconoscono le persone. La pareidolia è l'illusione subcosciente che tende a ricondurre a forme note, oggetti o profili (naturali o artificiali) dalla forma casuale.

La prosopagnosia è un deficit percettivo acquisito o congenito del sistema nervoso centrale che impedisce ai soggetti che ne vengono colpiti di riconoscere correttamente i volti.

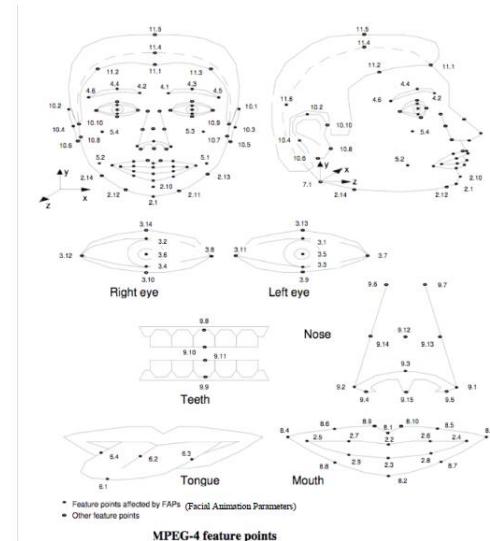


Un celebre caso di pareidolia: il [Volto su Marte](#), una formazione rocciosa marziana ripresa dalla sonda [Viking](#) 1, che appare come un volto in particolari condizioni e angolazioni di luce

Il volto è largamente utilizzato come biometria, infatti i due più importanti fattori che sanciscono il successo di una biometria sono **affidabilità** e **accettabilità**. Il riconoscimento dell'iride è il sistema più affidabile, ma è anche quello più intrusivo. Le impronte digitali sono più facilmente accettate, ma non applicabili a soggetti non consenzienti. Il volto possiede un'accettabilità molto elevata, mentre l'affidabilità deve essere ancora migliorata.

Ulteriori vantaggi possono essere: È naturale riconoscere una persona dal viso e lo è di solito per le persone da fotografare (ad es. impronte digitali possono essere associate al cattivo sentimento di essere sospettato di essere un criminale), alto tasso di riconoscimento in condizioni controllate, i dispositivi di acquisizione sono facili da implementare nell'ambiente operativo, è possibile integrarlo nelle applicazioni di controllo remoto, è possibile integrarlo nell'accesso logico (login).

Ma il volto è una struttura tridimensionale piuttosto complessa che quando viene acquisita, soprattutto se viene acquisita da un solo punto di vista, diventa bidimensionale perdendo una parte delle informazioni. Questa sua complessità è il limite che si ottiene quando si effettua l'acquisizione da un solo punto di vista, essa determina una certa inaffidabilità legata ad un altro limite sostanziale del volto e cioè la sua **eccessiva variabilità intra-classe**, infatti essa introdotta all'interno di un sistema di riconoscimento bisogna immaginare un aggiornamento periodico del template.



Applicazioni della Face Recognition



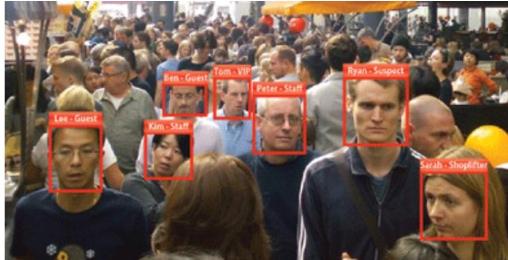
Forensics



iOS face recognition app –



Border control



In a crowd!

Problemi

I problemi, come accennato in precedenza, legati alla biometria del volto sono diversi, come ad esempio l'eccessiva variabilità tra le persone e la similarità interpersonale (sosia o chirurgia plastica).

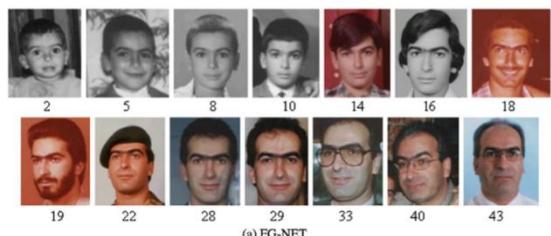
Un altro aspetto fortemente perturbativo dei sistemi di Face Recognition sono le perturbazioni che possono essere indotte dall'acquisizione; il volto è una delle poche biometriche che può essere variata dall'individuo a differenza di altre, infatti si possono produrre delle variazioni volontarie sul volto (espressioni o smorfie). Quando a queste variazioni si aggiungono ulteriori fattori perturbativi come la variazione di posa e illuminazione, si creano le **PIE variations** (**Pose, Illumination, Expression**). Esse sono in realtà le variazioni che rendono più difficile il processo di acquisizione. Questi elementi perturbativi sono elementi che intervengono quando l'utente non è collaborativo.





L'altro fattore che non è indotto dalla persona e dalle condizioni ambientali è il cosiddetto **Ageing** (A-PIE, PIE variations + Ageing), ovvero la variazione dell'età.

Esso è un aspetto fortemente perturbatore nel processo di Face Recognition. Le caratteristiche somatiche tendono a modificarsi in intervalli di tempo piuttosto piccoli, e tendono progressivamente a determinare effetti che col tempo possono rendere totalmente inefficaci i processi di riconoscimento.



29



40

(b) MORPH

Quando si sviluppa una tecnica di face Recognition, essa la si testa su database comuni per comprendere in modo comparativo, se l'idea algoritmica prodotta risulta essere migliore rispetto ad altre. Negli anni sono stati prodotti diversi database e quando la comunità scientifica produce un nuovo algoritmo, tende ad utilizzare uno più database per avere una misura comparativa con quello che è stato prodotto in precedenza.

Verifica e Riconoscimento

Esse sono le due tipologie di applicazione in cui il volto può essere utilizzato.

Verifica: confronto Uno a Uno. Conferma l'identità di un individuo dichiarata mediante carta d'identità, codice utente, ecc...

Riconoscimento: confronto Uno a Molti. Stabilisce l'identità di un soggetto a partire da un insieme di persone registrate.

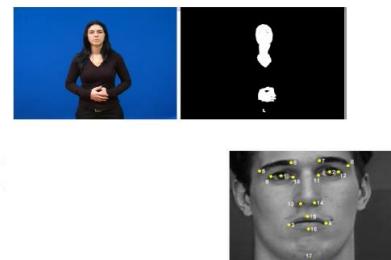
Struttura di un Riconoscitore Facciale

Il primo passo è legato ad individuare all'interno di un video/immagine la presenza di un volto. Una volta che il volto è stato individuato con qualche tecnica, allora è possibile darlo in input al sistema di classificazione che deve estrarre le caratteristiche principali. Il primo problema da risolvere, è quello di costruire un **face detector efficace**, ovvero come capire se in un'immagine vi è un volto.

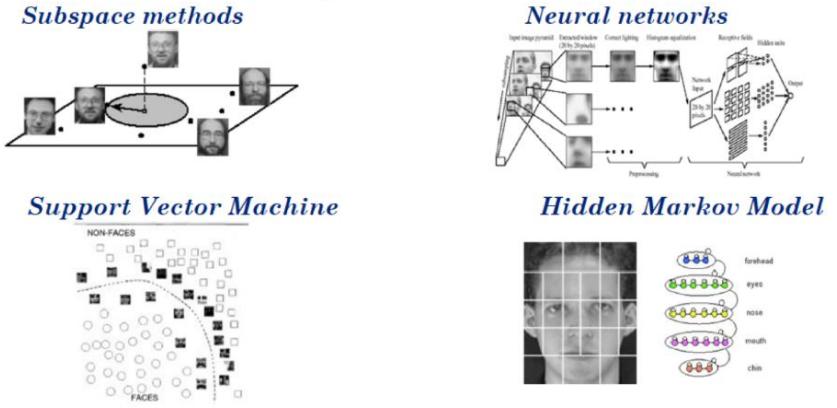


Face localization: some approaches

Le tecniche per effettuare operazioni di Face Detection sono diverse. Le più elementari ma comunque estremamente performanti sono quelle basate su caratteristiche di basso livello che possono riguardare alcune caratteristiche dei pixel (luminosità del volto o geometria del volto). Infatti, tutti i punti che individuano occhi, narici, sopracciglia ecc... creano una costellazione di informazioni che possono essere legate da una serie di relazioni geometriche. Oppure si può immaginare di avere un template di un volto rappresentabile con un modello matematico. Tutte queste sono caratteristiche di basso livello nelle quali si possono individuare le tecniche di face detection **features-based**; esse sono tecniche che sfruttano alcune proprietà di basso livello del volto per individuarlo all'intero di una immagine.

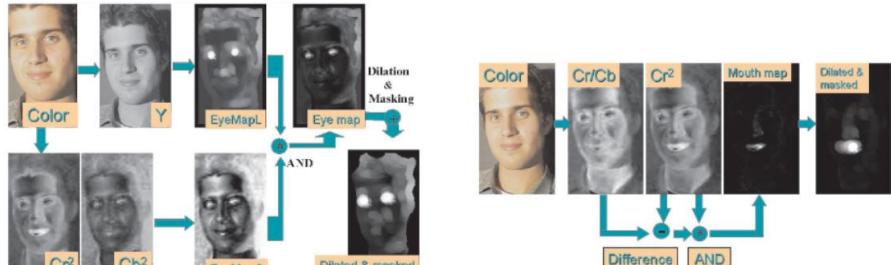


Poi ci sono degli approcci che piuttosto che basarsi sulle caratteristiche elementari del volto sfruttano le immagini nella loro totalità, le cosiddette tecniche **image-based**. Queste tecniche si basano sull'addestramento di modelli computazionali, in grado di riconoscere determinati oggetti sulla base di un pre-addestramento effettuato. Si costruisce un dataset di addestramento, dove all'interno ci sono sia immagini contenenti volti e sia immagini non contenenti volti, e si allena il sistema a riconoscere soltanto i volti.



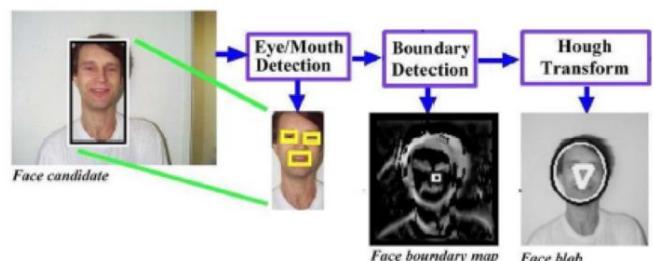
Eye/Mouth localization

Una delle tecniche basate sulle caratteristiche di basso livello sfrutta la relazione tra il modello RGB e YCC. Quest'ultimo è quello più utilizzato nell'image processing; è un modello nel quale è possibile separare un componente dell'immagine, rispetto a delle altre, concentrando in questa componente (Y) la maggior parte dell'informazione significativa. La transizione tra i due modelli avviene attraverso una trasformazione lineare che mappa il modello RGB nel modello YCC. Nel caso del modello YCC l'informazione che interessa non è contenuta nella componente di luminanza Y, ma piuttosto nelle altre due componenti C C (crominanza). Nella componente di crominanza sono particolarmente evidente alcune caratteristiche del volto, in particolare occhi e bocca; presa questa componente di crominanza si applicano una serie di operatori (equalizzazione e AND) che tendono a massimizzare gli aspetti che interessano occhi e bocca. La loro individuazione permette di avere una certa probabilità che l'immagine in considerazione contenga un volto. Tale tecnica produce molti falsi positivi (volti dove non ci sono).



Face Contour

Un approccio che rientra più nell'ambito geometrico prevede l'individuazione della bocca e degli occhi ma vincola essi ad una certa relazione geometrica. Ovvero deve crearsi un triangolo che gode di alcune proprietà specifiche. Tale triangolo viene anche legato ad esser circoscritto da una circonferenza che rappresenta il volto vero e proprio. Si passa ad un approccio di natura geometrica in quanto ci si basa non solo sulla presenza di occhi e bocca ma anche sulle caratteristiche geometriche che vengono calcolate sulla base di una serie di operatori.



Face Detection: Viola-Jones

Paul Viola e Michael Jones hanno proposto uno degli approcci di maggior successo (finora) alla localizzazione degli oggetti (incluso in OpenCV).

L'algoritmo è **image-based** e può essere applicato al rilevamento del volto (ma anche al rilevamento di occhi e bocca in una strategia gerarchica). Esso richiede di **creare un classificatore** inizialmente addestrato utilizzando più istanze della classe da identificare (esempi positivi), e diverse istanze di immagini che non contengono alcun oggetto della classe ma che possono causare un errore (esempi negativi). Il Training è progettato per **estrarre diverse funzionalità** dagli esempi e per selezionare quelle più discriminanti. Il modello statistico che viene costruito in modo incrementale contiene tali informazioni.

Misses (un oggetto presente non viene rilevato) o **False Alarms** (un oggetto viene rilevato ma non è presente) possono essere ridotti riqualificando l'aggiunta di nuovi esempi adatti (positivi o negativi).

L'algoritmo di Viola-Jones è un algoritmo **composito** in cui ci sono alcuni aspetti che di fatto mettono insieme un po' tutto quello che è stato analizzato in precedenza nell'ambito dell'elaborazione delle immagini.

È un algoritmo molto veloce in quanto vengono posti in modo preliminare al processo di riconoscimento vero e proprio, utilizzando dei classificatori, dei filtri e attraverso la combinazione di essi, mette a punto una strategia gerarchica affinché riesca ad individuare se in un'immagine è presente un volto.

L'algoritmo di Viola-Jones sfrutta le cosiddette **caratteristiche di Haar**, cioè delle caratteristiche tipiche del volto, condivise da tutti i volti umani. Queste caratteristiche presenti nel volto vengono rappresentate da dei **filtri** che vengono **sovraposti** all'immagine. Tali filtri cercano di verificare la presenza di alcuni aspetti tipici del volto umano, come ad esempio la regione perioculare (occhi) è generalmente più scura della zona sottostante (zigomi) così come il ponte del naso è normalmente più luminoso degli occhi.

Tendenzialmente si possono utilizzare filtri molti simili a questi per cercare di capire se all'interno di una

immagine c'è un'alternanza di regione scura/chiara, perché tale alternanza potrebbe essere sintomo di presenza di occhi e zigomi. Così come un'alternanza nero/bianco/nero potrebbe essere funzionale all'avere individuato occhio/ponte del naso/occhio.

È chiaro che queste sono condizioni necessarie ma non sufficienti per individuare un volto, ed è chiaro **che di questi filtri se ne dovranno considerare diversi** da far passare sul volto al fine di poter avere una certezza che quel che viene isolato è effettivamente un volto.

Una volta individuato il filtro da far scorrere sull'immagine lo si sovrappone su di essa e si verifica la variazione fra zona chiara e zona scura: si moltiplicano per 1 i pixel dell'area bianca e per 0 i pixel dell'area nera e si effettua una differenza, che poi viene controllata se essa **superà una certa soglia**; se si supera una certa soglia allora vuol dire che quel filtro ha una forte probabilità di aver localizzato una regione dell'immagine.

Il valore di una determinata funzione è sempre semplicemente la somma dei pixel all'interno di rettangoli chiari sottratti dalla somma dei pixel all'interno rettangoli ombreggiati. Caratteristiche rettangolo:

Valore = Σ (pixel nell'area nera) - Σ (pixel nell'area bianca).

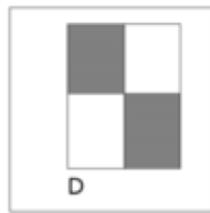
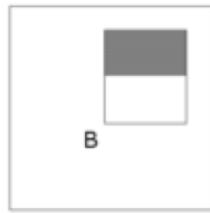
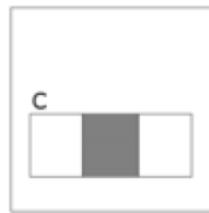
Tre tipi: due, tre, quattro rettangoli, Viola e Jones hanno utilizzato le funzioni a due rettangoli. Ad esempio: la differenza di luminosità tra i rettangoli bianchi e neri su un'area specifica. Ogni funzione è correlata a una speciale posizione nella finestra secondaria.



Haar Feature that looks similar to the bridge of the nose is applied onto the face

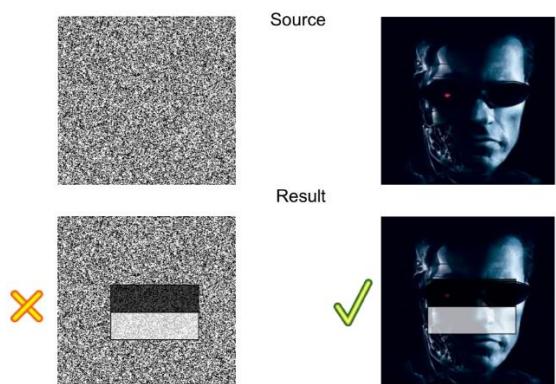


Haar Feature that looks similar to the eye region which is darker than the upper cheeks is applied onto a face



Example use of rectangular features

Example



Una serie di filtri vengono applicati sull'immagine: diversi filtri dove ognuno dovrebbe individuare una parte significativa dell'immagine. Quando viene individuata un'area potenzialmente assimilabile ad un'immagine il quadro si blocca in quanto si è in presenza del primo frame utilizzabile per classificare l'immagine come volto. I filtri possono avere diverse dimensioni e le finestre possono essere di diversa grandezza.



An example of an early stage in the Haar cascade. Each black and white patch represents a feature that the algorithm hunts for in the image.

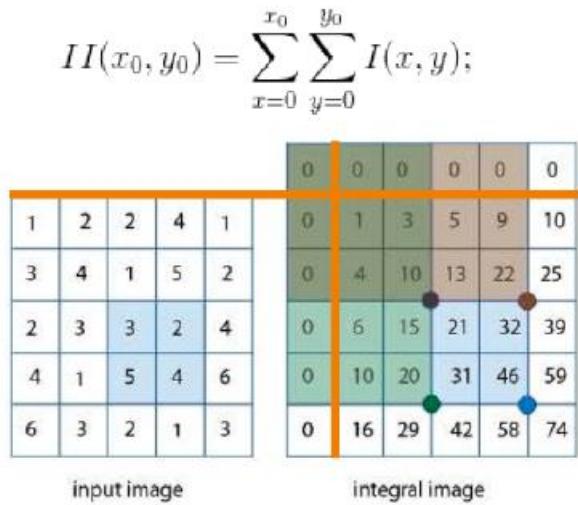
Come si vede in figura, vengono mostrate alcune delle caratteristiche che possono essere utilizzate da Viola-Jones. Uno dei problemi di tale algoritmo è che dovrebbe unire efficacia ad efficienza, quindi bravo nell'evidenziare i potenziali volti in un tempo piuttosto rapido. Il problema è che ogni volta viene applicata una mascherina come in figura sull'immagine, devono essere effettuate delle semplici **operazioni** (moltiplicazioni e sottrazioni) **ripetute un gran numero di volte** su tutta quanta l'immagine, in quanto tali operazioni vengono effettuate per ognuno dei filtri con dimensione anche variabile dei filtri stessi. Per ridurre il costo computazionale l'immagine viene trasformata nella cosiddetta **Integral Image**.

numero di volte su tutta quanta l'immagine, in quanto tali operazioni vengono effettuate per ognuno dei filtri con dimensione anche variabile dei filtri stessi. Per ridurre il costo computazionale l'immagine viene trasformata nella cosiddetta **Integral Image**.

L'algoritmo

L'algoritmo utilizza delle Haar features in combinazione con una nuova rappresentazione dell'immagine detta Integral Image. Le features hanno basso costo computazionale e la nuova struttura dati permette di effettuare l'analisi in tempo costante indipendentemente dalla dimensione delle regioni analizzate. Viene introdotto un metodo di selezione di feature di Haar attraverso l'algoritmo AdaBoost di Freud Shapire (1995). Questa strategia permette di eliminare in addestramento la maggior parte delle feature di scarsa capacità discriminante e selezionare solo quelle più efficaci per il problema.

Integral Images

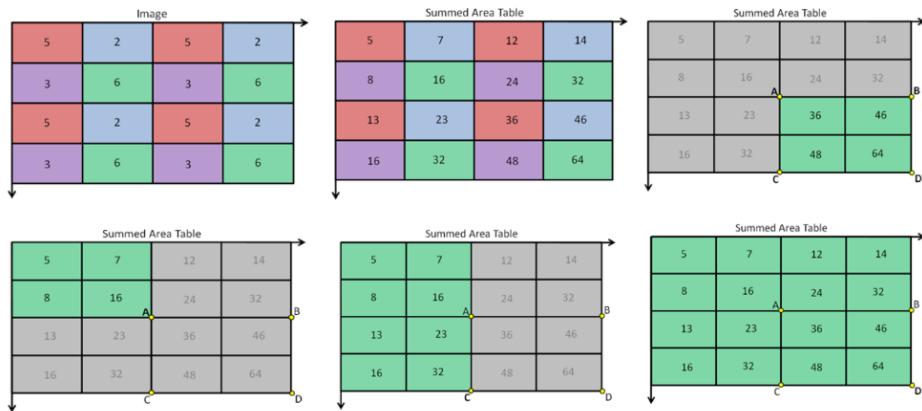


L'immagine integrale non è nient'altro che una rappresentazione/trasformazione più efficiente dell'immagine per realizzare queste moltiplicazioni e sottrazioni. Come in figura, l'immagine integrale viene costruita sommando i pixel in maniera progressiva. Quindi ogni pixel rappresenta la somma dei pixel precedenti in maniera crescente partendo **da sinistra verso destra e dall'alto verso il basso**. La rappresentazione dell'immagine in termini di immagine integrale permette di realizzare la somma dei pixel in una **qualsiasi area** (rettangolo o quadrato di riferimento) attraverso un numero più efficiente di operazioni, ovvero **solamente tre operazioni**.

Indipendentemente dalla dimensione del filtro si potrà sempre realizzare la somma dei pixel all'interno di un'immagine in termini di: A - B - C + D facendo fondamentalmente solo tre operazioni elementari; questo risulta essere un grande vantaggio in quanto applicando un filtro di HAAR sull'immagine, si può avere in tempo costante (le sole tre operazioni) indipendentemente dal numero di pixel che sono coperti, la somma dei pixel. Poiché fondamentalmente nei filtri di HAAR si deve effettuare la somma dei pixel per poi fare la differenza, la riuscita di queste operazioni velocemente è un grande vantaggio.

Once the integral image is obtained, the sum of pixels in any rectangular region can be obtained in constant time ($O(1)$ time complexity) by the following expression:

$$\text{Sum} = \text{Bottom right} + \text{top left} - \text{top right} - \text{bottom left}$$

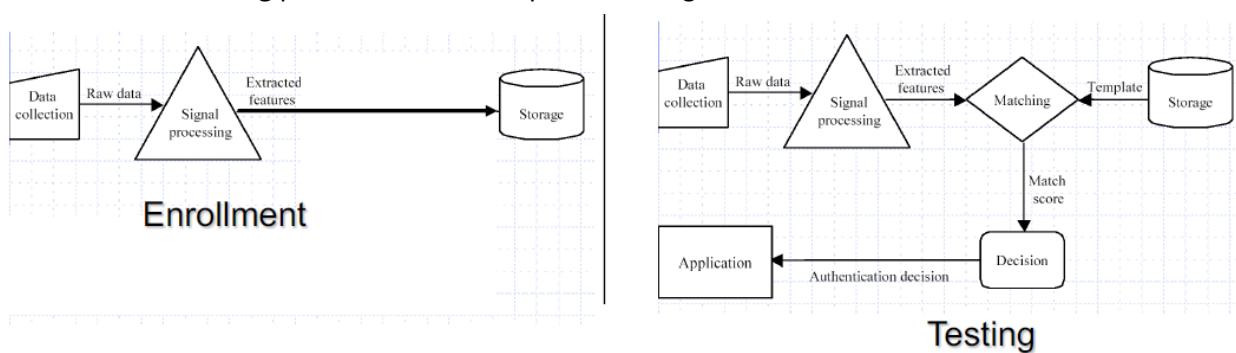


Boosting

La scelta dei filtri e il peso assegnato viene fatto con dei classificatori Ada Boosting.

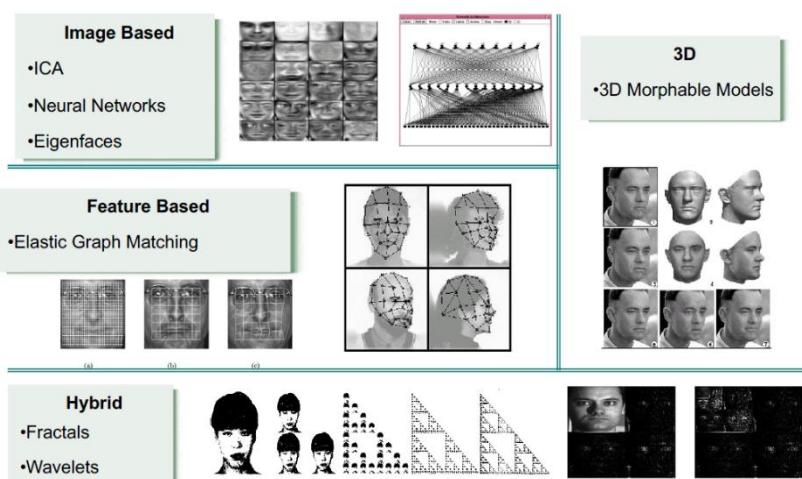
Face Recognition

Una volta effettuata l'operazione di detection, ovvero una volta individuato il volto all'interno di un'immagine, successivamente si procede all'estrazione di questo volto dall'immagine, precisamente le caratteristiche del volto che permettono di effettuarne il riconoscimento. Per poter fare il **riconoscimento** del volto bisogna avere un **template**, ovvero un modello corrispondente affinché possa essere confrontato e verificato. In precedenza, si è accennato che un sistema di riconoscimento biometrico prevede due fasi: **Enrollment** in cui si crea il template, dove il dato viene acquisito attraverso un sensore e le informazioni estratte con l'image processing vengono catalogate sotto forma di template in un database. Questo processo può essere effettuato singolarmente o in blocco (batch Enrollment). L'altra fase, quella di **Testing** prevede l'acquisizione nuovamente dei dati attraverso la stessa tipologia di sensore, l'estrazione delle caratteristiche e il recupero del template dal database per il confronto con i dati estratti. Attraverso una soglia di tolleranza si verifica se il **matching** produce un risultato positivo o negativo.



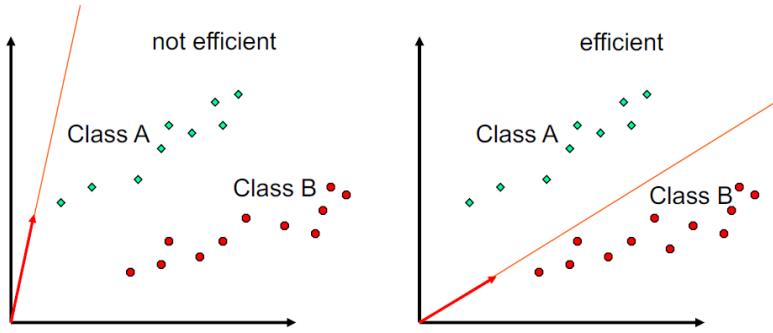
Sistemi esistenti in Letteratura

In letteratura esistono svariati approcci di tecniche di face Recognition.



ICA e PCA

Tra le tecniche più utilizzate soprattutto in passato per il Face Recognition c'è la **PCA (Principal Component Analysis)** che rientra nella categoria delle Image Based. Tale tecnica sotto alcuni aspetti ricorda la trasformata di Fourier, ma con un approccio totalmente diverso. Piuttosto che considerare i pixel, essi vengono associati a delle variabili geometriche. Vettori (le direzioni) in cui è massima la variazione (varianza) tra i vettori stessi. L'**obiettivo** principale della PCA è di individuare dei vettori che siano efficienti per rappresentare determinati dati. Si veda in figura, ad esempio si supponga di avere due classi di dati A e B; nella prima rappresentazione



non è possibile affermare che il vettore possiede una direzione rappresentativa della classe A o della classe B, infatti la rappresentazione non indica il comportamento delle due classi. Esse pur avendo dati distinti, possiedono delle caratteristiche in comune, ovvero una distribuzione delle informazioni rispetto al piano cartesiano

piuttosto simile. Nella seconda rappresentazione invece, il vettore è rappresentativo delle due classi; infatti il vettore e le due classi condividono la direzione e quindi il coefficiente angolare della retta associata.

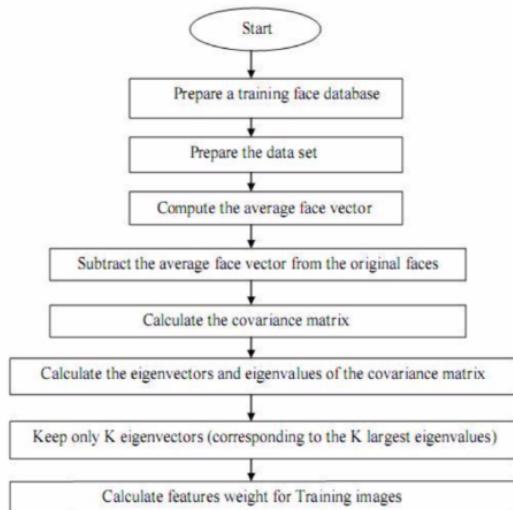
Si assuma di avere delle **immagini quadrate di dimensione N** , in cui il numero di pixel risulta essere $N \times N$. Con M il **numero di immagini del database** e P il **numero di persone coinvolte all'interno del database**; quindi con M diverso da P (per ogni persona si possono avere più immagini nel database).

L'algoritmo della PCA non fa nient'altro che cercare di individuare all'interno di un insieme di immagini, le **immagini più significative**; quindi tentare di trovare un numero di immagini molto più piccolo di M tale che queste immagini siano **rappresentative** di tutto il dataset. Quindi trovare una strategia che in presenza di un dataset molto corposo, possa dare in output un **insieme molto più piccolo** di immagini significative. Tali immagini significative possono rappresentare tutto il dataset; riassumendo si tenta di ridurre il problema di rappresentazione ad un problema più piccolo trasformando la dimensionalità del problema.

L'idea è quella di rappresentare **ogni volto** attraverso la **combinazione lineare** dei volti più significativi, i cosiddetti **Eigenfaces**. Quindi ogni volto significativo contribuirà alla rappresentazione di un volto all'interno dell'immagine.

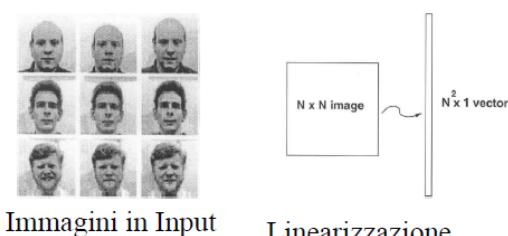
PCA: Il Metodo

1. «Training set» costituito da un insieme di immagini delle stesse dimensioni e normalizzate in maniera da evidenziare caratteristiche del volto, quali occhi e bocca.
2. Proiezione in un sottospazio, nel quale vengono messe in risalto le caratteristiche salienti di un volto. Tale riduzione dimensionale, infatti, permette di escludere l'informazione che non viene considerata rilevante e precisamente decomponne la struttura di un volto in una combinazione di componenti ortogonali, scorrelate tra loro, dette eigenface.
3. Ogni immagine di volto, può successivamente essere rappresentata come una somma pesata (vettore delle feature) di queste eigenfaces, raccolte in un vettore monodimensionale. Il confronto di un'immagine di volto con le altre presenti nel training set viene effettuato semplicemente valutando la distanza tra questi vettori di caratteristiche locali.
4. La Matrice di Covarianza: Matrice Simmetrica che rappresenta la variazione di ogni variabile rispetto alle altre (inclusa se stessa).
 - Se valori grandi" (rispetto il valore atteso) di X si accostino a valori grandi di Y, oppure "valori piccoli" (rispetto il valore atteso) di X con valori piccoli di Y, la covarianza è maggiore o uguale a 0. In questo caso si dice che X e Y sono **correlate positivamente**.
 - Se invece valori piccoli di X tendono ad accoppiarsi con valori grandi di Y o viceversa, la covarianza sarà minore o uguale a 0 e le variabili aleatorie in questione si diranno **correlate negativamente**.



Eigenfaces – Sistemi Lineari

Si prendono in input le immagini e per ognuna di esse si effettua una **linearizzazione**, ovvero ogni immagine viene rappresentata come vettore, si calcola la **media** su tutti i quanti i vettori e tale media viene **sottratta** da ciascun vettore. Quest'ultima operazione fa in modo che venga rimossa la parte in **comune** tra tutti i volti, ottenendo quindi le immagini che andranno a costituire la base per la matrice di covarianza; si scelgono le immagini **meno correlate** tra di loro (quelle più vicine ad essere linearmente indipendenti). Una volta individuato l'insieme dei vetri (immagini) meno correlate, è possibile rappresentare tutta la base di dati attraverso una **combinazione lineare** degli Eigenfaces. Quindi la scelta delle immagini più indipendenti e più significative permette la generazione di tutta quanta la base di dati.



$$\Psi = \frac{1}{M} \sum_{i=1}^M \Gamma_i \quad \Phi_i = \Gamma_i - \Psi$$

Media su tutte i volti A ciascun vettore togliamo la media



Risolvendo un particolare sistema lineare
estraiemo un insieme di volti di riferimento.

$$\text{Face} = 0.9571 * \text{Face 1} - 0.1945 * \text{Face 2} + 0.0461 * \text{Face 3} + 0.0586 * \text{Face 4}$$

Ciascun volto viene rappresentato come una combinazione dei volti di riferimento. I coefficienti utilizzati per la rappresentazione costituiscono il vettore di caratteristiche

Eigenfaces, the algorithm

- The database

$$\begin{array}{l} \text{Face 1} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_{N^2} \end{pmatrix} & \text{Face 2} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_{N^2} \end{pmatrix} & \text{Face 3} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_{N^2} \end{pmatrix} & \text{Face 4} = \begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_{N^2} \end{pmatrix} \\ \text{Face 5} = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_{N^2} \end{pmatrix} & \text{Face 6} = \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_{N^2} \end{pmatrix} & \text{Face 7} = \begin{pmatrix} g_1 \\ g_2 \\ \vdots \\ g_{N^2} \end{pmatrix} & \text{Face 8} = \begin{pmatrix} h_1 \\ h_2 \\ \vdots \\ h_{N^2} \end{pmatrix} \end{array}$$

Eigenfaces, the algorithm

- We compute the average face

$$\vec{m} = \frac{1}{M} \begin{pmatrix} a_1 + b_1 + \dots + h_1 \\ a_2 + b_2 + \dots + h_2 \\ \vdots \\ a_{N^2} + b_{N^2} + \dots + h_{N^2} \end{pmatrix}, \quad \text{where } M = 8$$



Eigenfaces, the algorithm

- Now we build the matrix which is N^2 by M

$$A = [\vec{a}_m \vec{b}_m \vec{c}_m \vec{d}_m \vec{e}_m \vec{f}_m \vec{g}_m \vec{h}_m]$$

- The covariance matrix which is N^2 by N^2

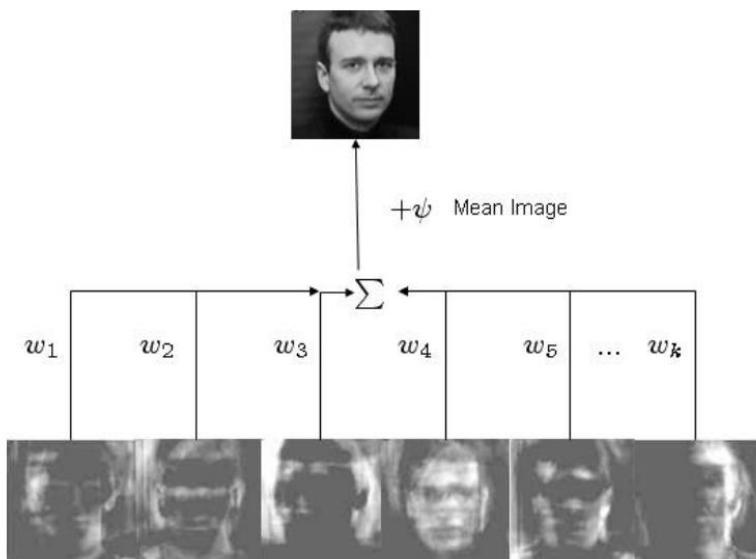
$$Cov = AA^T$$

Covarianza positiva: indica che due variabili tendono a muoversi nella stessa direzione.

Covarianza negativa: rivela che due variabili tendono a muoversi in direzioni inverse.

PCA ed Eigenfaces

La rappresentazione di ogni volto all'interno del database avviene attraverso la combinazione lineare dei vettori linearmente indipendenti o che quelli che si sono classificati come potenzialmente linearmente indipendenti.



Example backprojection



Example training set

Mean image from the training set



Eigenfaces from the training set

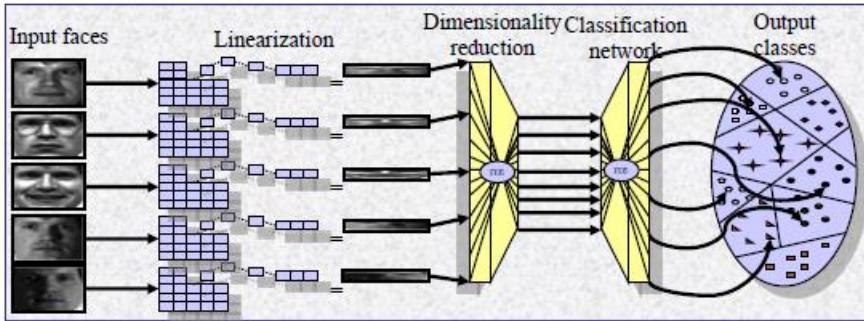
Riassumendo, il principio alla base della PCA è la ricerca all'interno della base di dati delle immagini meno correlate aspettandosi che esse siano dal punto di vista numerico molto inferiore alla dimensione dell'intero database. Quindi, se il database possiede M immagini, si cercano k immagini con k molto piccolo, tale da poter rappresentare tutte le immagini del database attraverso la combinazione lineare delle k scelte. Esse vengono scelte grazie alla matrice di covarianza che permette di stabilire quali immagini risultano essere meno correlate tra esse; meno le immagini sono correlate più sono tra di loro indipendenti. Le immagini vengono assimilate da dei vettori *n-dimensional*. La matrice di covarianza permette di stabilire quante sono le immagini. La scelta delle immagini dipende anche **dall'accuratezza** che si vuol ottenere. Più è **piccola** la dimensione del set di Eigenfaces scelto, **meno** è l'accuratezza ottenuta nella rappresentazione delle immagini (maggior errore).

Vantaggi e Svantaggi

- La fase di identificazione è veloce.
- Se si conservano gli autovettori è possibile ricostruire l'informazione originaria.
- La fase di training è lenta.
- Se si aggiunge un numero consistente di nuovi soggetti è necessario il retraining del sistema.
- Elevata sensibilità a variazioni di illuminazione, posa, occlusioni, ...

Reti Neurali

Una rete neurale simula il funzionamento dei neuroni nel cervello. Ciascun neurone è rappresentato da una funzione matematica basata sul calcolo delle probabilità. Per il riconoscimento dei volti, l'ottimo sarebbe utilizzare un neurone per ciascun pixel. Questo approccio richiede troppi neuroni. La soluzione consiste nell'utilizzare una rete di neuroni per « riassumere » l'immagine in un vettore più piccolo. Mentre una seconda rete effettua il riconoscimento vero e proprio.

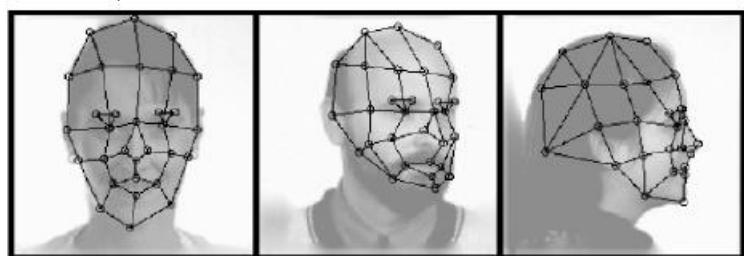


Vantaggi e Svantaggi

- Riducono l'ambiguità fra soggetti appartenenti a classi simili.
- Con opportuni accorgimenti sono robuste alle occlusioni
- Richiedono più di una immagine per il training
- Alcune reti sono soggette a problemi di:
 - **Overfitting:** la rete ha la stessa dimensione dell'input.
 - **Overtraining:** se le immagini sono grandi il metodo converge lentamente
 - **Database size:** quando il numero di soggetti cresce diventano inefficienti.

Sistemi basati su Grafi

Attraverso filtri e funzioni di localizzazione vengono localizzati sul volto un insieme di punti di riferimento. Questi punti vengono collegati da archi pesati e si ottiene un grafo. Ad ogni volto è associato un grafo, per cui confrontare due volti significa confrontare due grafi.

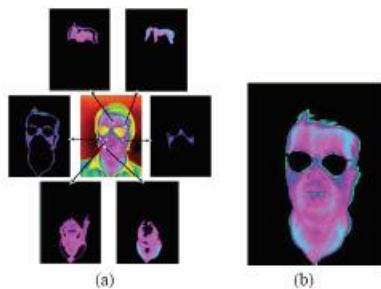


Vantaggi e Svantaggi

- Sono robusti rispetto alle variazioni di posa
- Sono robusti rispetto alle variazioni di illuminazione
- Non richiedono il retraining del sistema
- Il processo di testing è molto lento perché richiede il confronto fra grafi (NP-Hard).

Termogramma

L'immagine del volto viene acquisita mediante un sensore termico. Il sensore rileva le variazioni di temperatura dell'epidermide del volto. L'immagine viene segmentata e indicizzata.

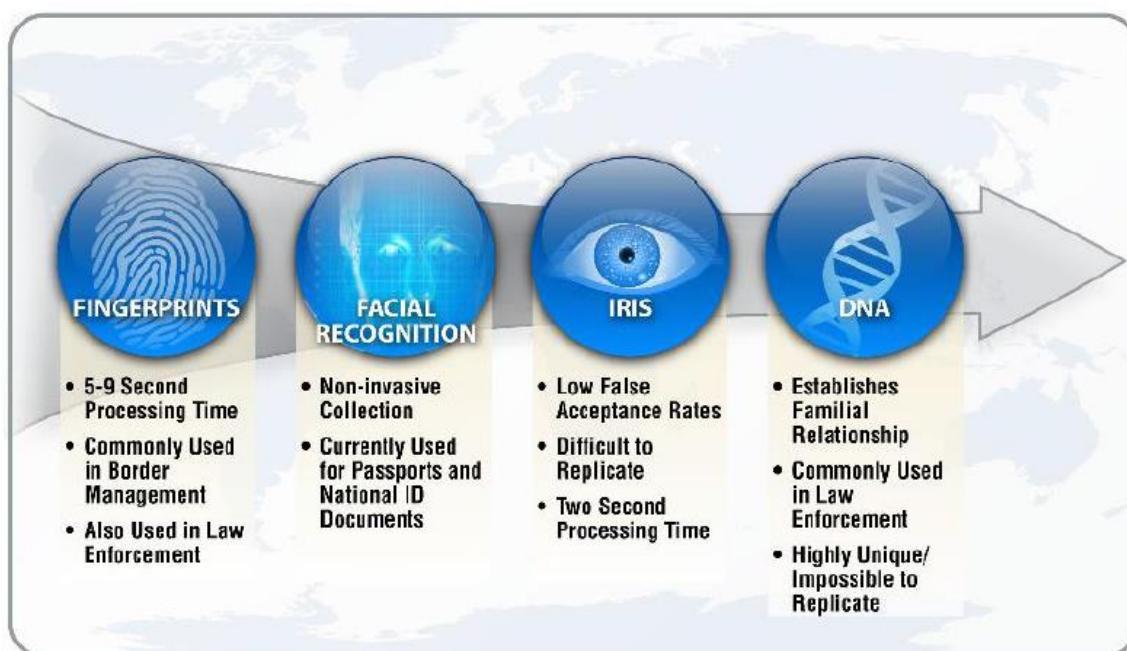


Vantaggi e Svantaggi

- Sono robusti rispetto alle variazioni di illuminazione
- Sono robusti rispetto alle variazioni di tempo
- Sono efficienti anche in caso di ambienti esterni
- Richiedono dispositivi di acquisizione costosi
- I dispositivi di acquisizione sono troppo sensibili ai movimenti del soggetto e offrono limitate risoluzioni
- Dipendono dallo stato emotivo del soggetto
- La presenza di un vetro fra il soggetto e il dispositivo rende inefficace l'acquisizione.

L'iride

Some Desirable Properties



red arrow pointing to the DNA row

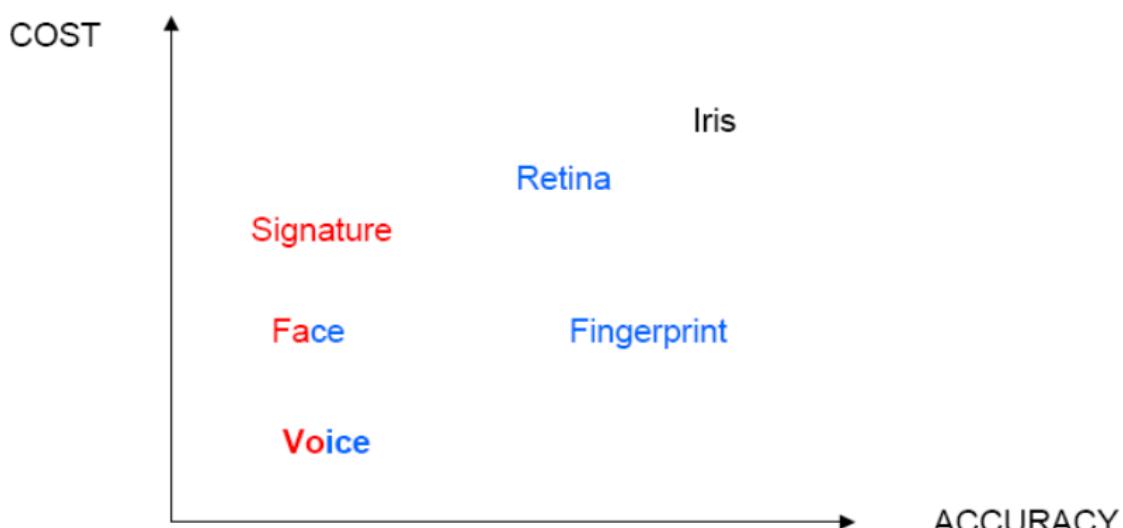
Biometric Trait	Comfort	Accuracy	Availability	Costs
Finger Print	oooooooo	ooooooo	oooo	ooo
Signature (dynamic)	ooo	oooo	oooooo	oooo
Facial geometry	oooooooooooo	oooo	oooooooo	oooooo
Iris	oooooooooooo	oooooooooooo	oooooooooooo	oooooooooooo
Retina	ooooooo	oooooooo	ooooo	oooooooo
Hand geometry	oooooo	oooo	ooooooo	oooooo
Finger geometry	ooooooo	ooo	oooooooo	oooo
Vein Structure of the back of the hand	oooooo	oooooo	ooooooo	oooooo
Ear form	ooooo	oooo	oooooooo	oooooo
Voice	oooo	oo	ooo	oo
DNA	o	oooooooo	oooooooooooo	oooooooooooo
Odor	?	oo	ooooooo	?
Keyboard strokes	oooo	o	oo	o
Comparison: Password	oooooo	oo	oooooooo	o

green = best red = worst

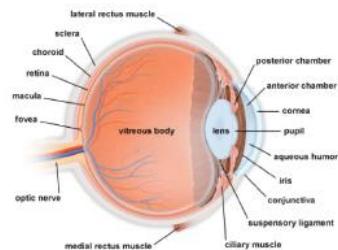


<i>Biometric characteristic</i>	<i>genotypic*</i>	<i>randotypic*</i>	<i>behavioral**</i>	<i>Biometric Trait</i>	<i>Permanence over time</i>
Fingerprint (only minutia)	o	ooo	o	Fingerprint (Minutia)	oooooooooooo
Signature (dynamic)	oo	o	ooo	Signature(dynamic)	oooo
Facial geometry	ooo	o	o	Facial structure	oooooooooooo
Iris pattern	o	ooo	o	Iris pattern	oooooooooooo
Retina (Vein structure)	o	ooo	o	Retina	oooooooooooo
Hand geometry	ooo	o	o	Hand geometry	oooooooooooo
Finger geometry	ooo	o	o	Finger geometry	oooooooooooo
Vein structure of the hand	o	ooo	o	Vein structure of the back of the hand	oooooooooooo
Ear form	ooo	o	o	Ear form	oooooooooooo
Voice (Tone)	ooo	o	oo	Voice (Tone)	ooo
DNA	ooo	o	o	DNA	oooooooooooo
Odor	ooo	o	o	Odor	oooooooooooo?
Keyboard Strokes	o	o	ooo	Keyboard strokes	oooo

Comparison of biometric techniques



Dal punto di vista anatomico, l’iride è un muscolo avente una funzione cromatica, che agisce sul foro presente al centro dell’occhio, ovvero la pupilla. Quest’ultima è il foro attraverso il quale entra la luce del sole, che va a impattarsi sulla retina. I muscoli che agiscono attraverso movimenti “involontari” hanno il compito di comprimere e far dilatare la pupilla per consentire l’entrata di una certa quantità di luce. Inoltre, dal punto di vista scientifico le **striature presenti sull’iride** (pattern irregolari) caratterizzano la cromaticità dell’iride. È possibile individuare tra due consanguinei anche una stessa cromaticità dell’iride, ma ciò che viene diversificato sono appunto tali striature casuali (solchi e creste) che vanno a caratterizzare la parte cromatica. Quindi nell’iride si va a rilevare la **distribuzione** di tali pattern irregolari, piuttosto che la cromaticità del colore, infatti tali pattern rendono unica l’iride dal punto di vista dell’identificazione. Addirittura, l’iride destra è diversa da quella sinistra. L’iride si trova tra la cornea e il cristallino.

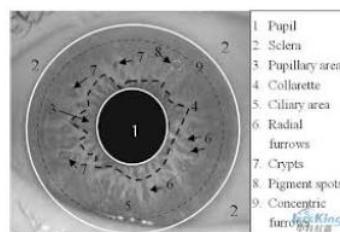
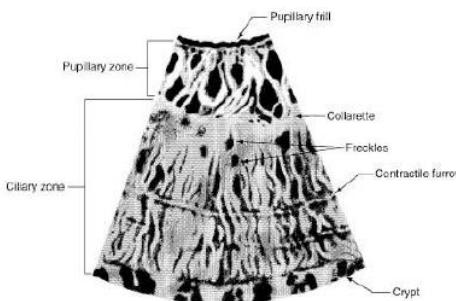


Vantaggi: l'iride è ben visibile benché protetta dalle palpebre. Non varia nel tempo ed è un tratto biometrico estremamente discriminante rispetto altri individui, ma anche all'interno dello stesso individuo. L'immagine può essere acquisita senza avere un contatto diretto (**contactless**). L'acquisizione dell'iride può avvenire in due modalità differenti: **infrared** e **nel visibile**.

Svantaggi: la superficie sulla quale bisogna poter effettuare l'estrazione delle caratteristiche è molto piccola, nella migliore delle ipotesi si parla di poco più di 3 cm^2 . Per poter acquisire l'iride, si dovrebbe disporre di dispositivi ad elevata risoluzione, che ovviamente dovrebbero adattarsi ad altre situazioni di contorno, ad esempio l'utente non dovrebbe indossare degli occhiali scuri ecc... per altre biometriche una parziale occlusione non ostacola un tentativo di riconoscimento.

L'iride come accennato può essere acquisita sia nel visibile che nell'infrarosso. La differenza sostanziale è legata alla quantità di **melanina**. Ovvero nel caso della luce visibile, la melanina assorbe la luce visibile, evidenziando gli strati che compongono l'iride sono visibili. L'immagine ottenuta con la luce visibile non permette di estrarre tutte le informazioni tridimensionali per una serie di problematiche.

Nel caso della luce infrarossa, essa viene utilizzata nelle operazioni di verifica (utente altamente collaborativo), in questo caso la melanina riflette la maggior parte della luce infrarossa e assorbe la maggior parte della luce visibile, rendendo limitata la riflessione. La tessitura è più visibile e questo consente un elevato livello di sicurezza soprattutto nel contesto della verifica, appunto infatti si rende più adatta in sistemi biometrici basati sul riconoscimento dell'iride.



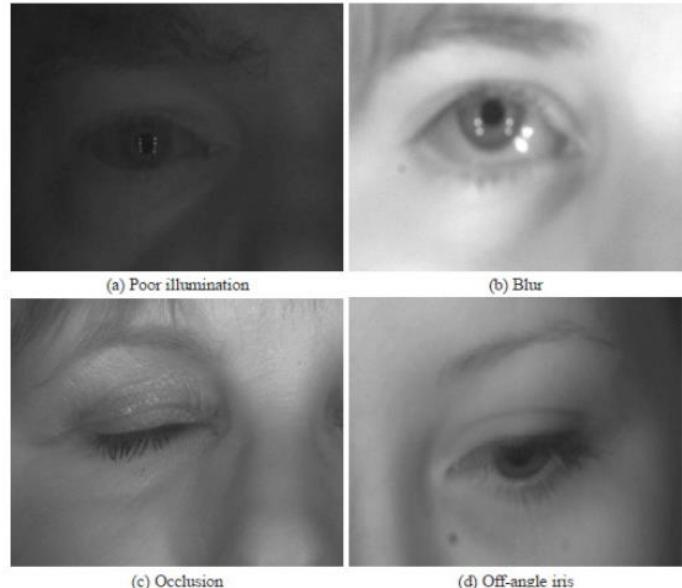
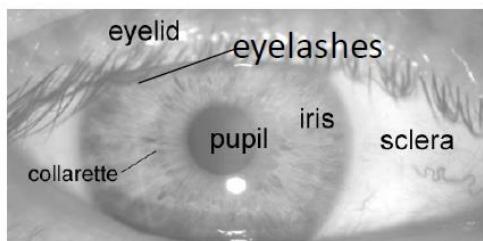
Near Infrared
(NIR: 700–900 nm)



Visible
Wavelengths

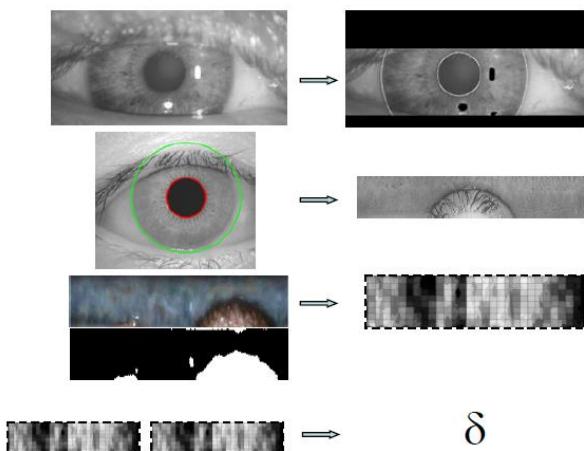


Uno dei problemi fondamentali dell'iride è la presenza di una serie di elementi che da un lato possono essere considerati **perturbatori**, e dall'altro elementi **caratteristici** della zona perioculare. In figura si possono evidenziare tutti gli elementi caratteristici della zona perioculare. Tutti questi elementi possono essere considerati parte di disturbo che possono essere combinate con un'altra serie di tipiche situazioni rumorose per l'acquisizione dell'iride come una scarsa illuminazione, o un'immagine con Blurring, un'occlusione oppure l'angolazione da cui viene effettuato lo scatto può essere sfavorevole per l'acquisizione. Tutte queste caratteristiche tendono a rendere difficile sia il processo di detection e sia il processo di estrazione delle caratteristiche.



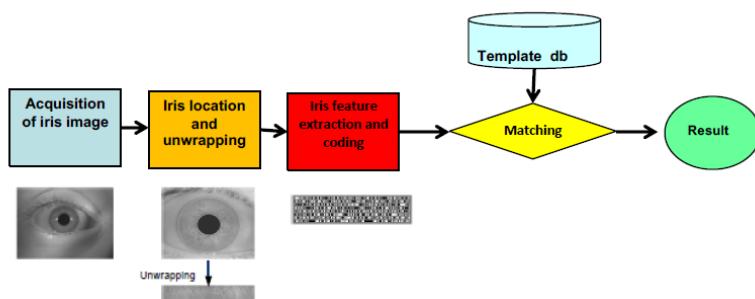
Normalmente un algoritmo per estrarre le caratteristiche dell'iride è caratterizzato da una serie di passi:

1. **Segmentazione** (o detection), quindi individuare che all'interno di un'immagine è presente un iride.
2. **Normalizzazione**.
3. **Coding**, che permette di estrarre le caratteristiche sulla base di un modello diversificato da caso a caso.
4. **Matching**, dove si confronta il codice prodotto nella fasi precedenti.



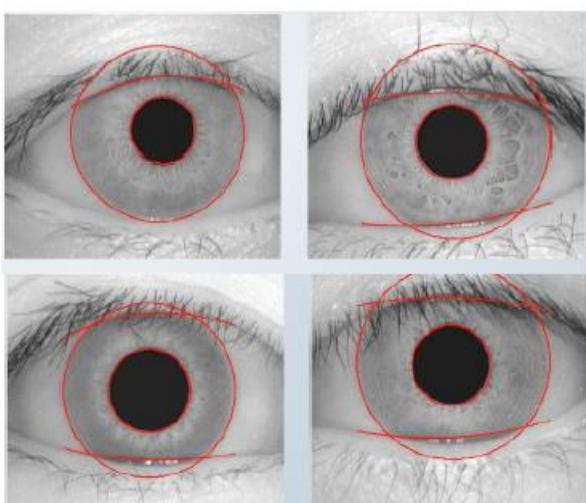
The first and most famous: Daugman

Il primo algoritmo, completamente automatico, utilizzato per l'analisi dell'iride fu prodotto da Daugman. Esso propose la prima metodologia automatica basata su un algoritmo piuttosto efficiente di **Iris Location** (detection). Perché il problema fondamentale nell'iride è legato alla necessità di **individuare** un'iride all'interno di una immagine. La detection dell'iride risulta essere fondamentale.



Il processo di estrazione delle feature è banale, una volta individuato l'iride l'estrazione delle caratteristiche è un problema secondario. Il problema principale come detto è riuscire ad

individuare l'iride. Se non si inizia tale processo o lo si sbaglia, tutto ciò che viene dopo risulta essere fortemente compromesso. Daugman in un primo momento associò l'individuazione di un'iride all'individuazione di **circonferenze** che fossero in grado di determinare sia il **cerchio della pupilla** sia il **cerchio più grande dell'iride**. Quindi due cerchi **concentrici** tra di loro con lo stesso centro e raggio variabile. Esso immaginò di avere a disposizione un filtro gaussiano, fondamentale per la determina di tali cerchi, per evitare che le variazioni presenti all'interno dell'iride possano determinare dei False Alarms. Immaginò quindi un Edge Detector che producesse delle circonferenze a raggio variabile che andavano ad effettuare una convoluzione con un filtro gaussiano, quindi l'immagine veniva lisciata e veniva prodotta poi una circonferenza (edge detector circolare) che individuasse le **transizioni più significative**. Tale algoritmo si fermava quando si raggiungeva il cerchio a raggio massimo. Partendo dalla zona centrale (x_0, y_0) il primo cerchio che veniva prodotto a raggio massimo era quello che determina la transizione **tra la pupilla e l'iride**, e il secondo cerchio a raggio massimo che veniva prodotto era quello che determina la transizione **tra l'iride e la sclera**. Quindi effettuando la convoluzione tra il filtro gaussiano che effettua uno Smoothing dell'immagine ed evita la produzione di false Alarms e la generazione di cerchi a dimensione sempre maggiore, si determina questo operatore integro-differenziale che infine determina i candidati ideali a identificare l'iride.

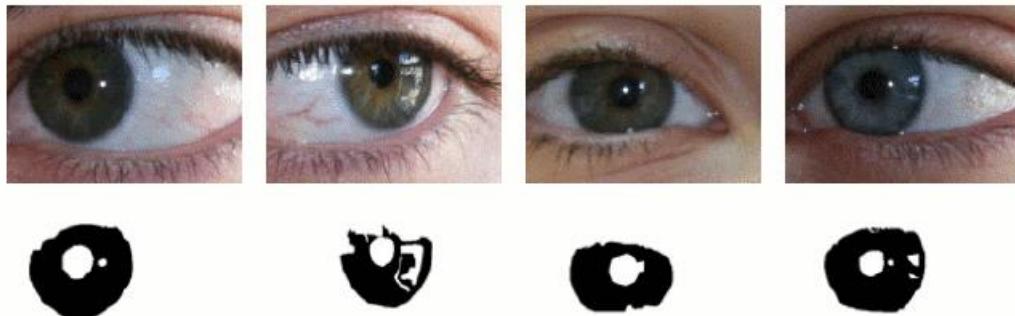


In figura si può osservare come opera l'algoritmo di Daugman. In maniera sequenziale a partire da un punto di un ipotetica iride inizia col generare una serie di cerchi che vanno a determinare le transizioni più elevate. Ogni volta questi cerchi vengono combinati con un algoritmo gaussiano che serve ad evitare variazioni non significative. I due cerchi con raggio massimo vengono poi etichettati per creare la **corona circolare** che determina l'area utile sulla quale andare ad effettuare l'estrazione delle caratteristiche. Ovviamente quest'area utile è solo parziale (in figura vi è un'immagine nel dominio infrared).

Riassumendo si tratta di un banale filtro circolare come visto in precedenza, solo che esso lavora in simultanea con un filtro gaussiano onde evitare falsi allarmi nella zone dell'iride, tentando di arrivare fino alla transizione con la sclera in quanto essa rappresenta la variazione massima di transizione tra un'area e l'altra.

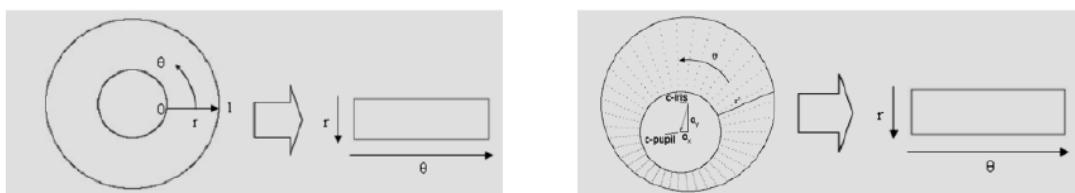
General Iris Segmentation

Una volta generato tale profilo, si etichettano i **punti neri** come punti appartenenti all'iride e i **punti bianchi** come punti non appartenenti all'iride. Si crea fondamentalmente una maschera che determina quali sono le aree sensibili. All'interno dell'iride ci possono essere dei puntini che vengono esclusi in quanto rappresentano zone di riflesso riducendo l'area dalla quale poter estrarre le caratteristiche. L'output della segmentazione, cioè dell'Iris Detection, è in realtà una **maschera in bianco e nero**.



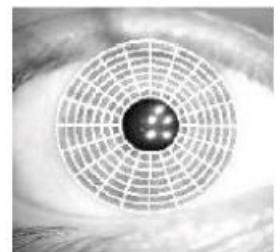
Daugman: iris unwrapping

Daugman nell'algoritmo aveva previsto una fase di **normalizzazione**, in quanto la dimensione dell'area significativa di un'iride era molto variabile. Essendo molto variabile poteva accadere di dover confrontare due iridi con dimensioni variabili dello spazio utile da cui estrarre le caratteristiche. Poiché l'iride, a differenza dell'impronta digitale, non presenta delle caratteristiche puntuali che vengono estratte ma che viene considerata nella sua intera superficie, vengono considerati tutti i pixel appartenenti alla superficie dell'iride. Si rese necessario che, dato l'intervento di elementi perturbatori, l'iride fosse normalizzato, proprio perché le aree utili erano diverse. Daugman applicò quindi vari algoritmi combinati tra loro che attraverso la combinazione tra coordinate polari e altri elementi di normalizzazione facessero in modo che tutte le superfici delle iridi estratte potessero avere sempre la **stessa dimensione**; creando così un vettore a dimensione fissa indipendentemente dall'area utile dell'iride estratta, proprio per non creare problemi nella fase di matching.



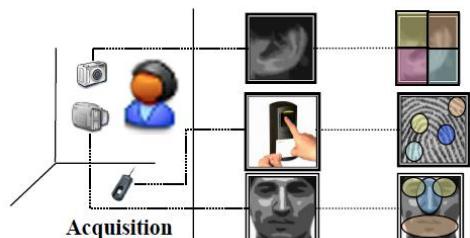
Daugman: feature extraction

Tra gli algoritmi di feature extraction più utilizzati, c'è una variante della trasformata di Fourier basata sui **filtri di Gabor**: si utilizzano dei filtri per catturare alcune informazioni significative dell'iride. Una volta individuata l'area di pertinenza, si tenta di estrarre da essa alcune informazioni aggregate. I filtri di Gabor sono una variante della trasformata di Fourier perché piuttosto che utilizzare la trasformata di Fourier del segnale, utilizzano la trasformata di Fourier combinata con la trasformata di Fourier di una gaussiana. In realtà il filtro era caratterizzato da una **convoluzione tra la trasformata di Fourier e la trasformata di Fourier di una gaussiana**. Quindi l'immagine viene proiettata nel dominio delle frequenze e non dei pixel. Infine, si avrà un vettore delle caratteristiche che banalmente viene identificato con i coefficienti di Gabor e su di essi viene effettuato un match attraverso la distanza di Hamming.

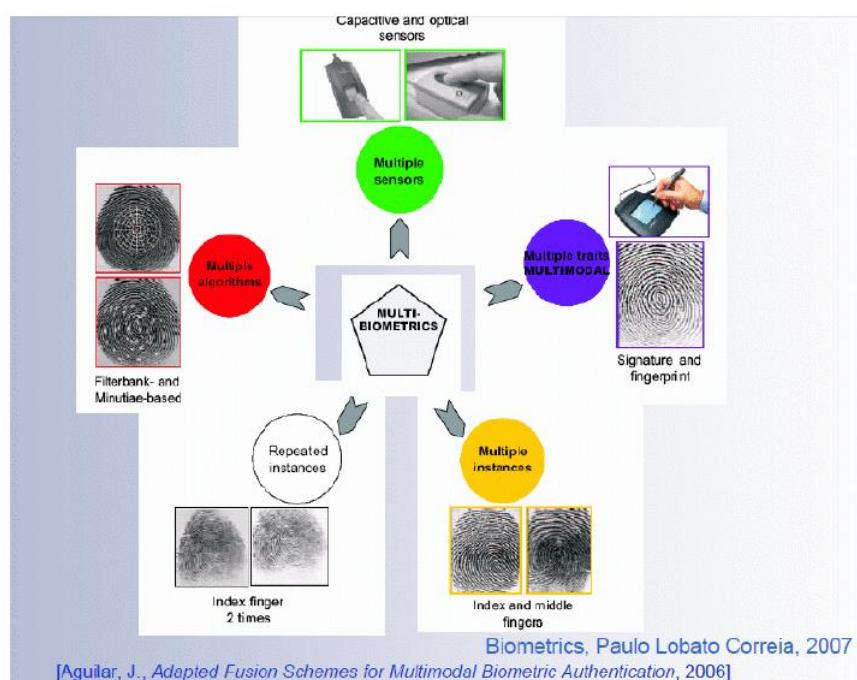


Multibiometric Systems

L'idea di sistema multi-biometrico nasce dalla necessità di aumentare la sicurezza nel contesto di verifica e riconoscimento di una persona. La maggior parte dei sistemi presenti si basa su una **singola biometria**. Ciò li rende vulnerabili a possibili attacchi e scarsamente robusti per una serie di problemi. Un sistema multimodale fornisce una soluzione efficace, poiché gli svantaggi dei singoli sistemi possono essere controbilanciati grazie alla disponibilità di più biometrie.



Si possono avere vari tipi di sistemi multi-biometrici:



Le **modalità** sono le seguenti:

- **Multimodal:**



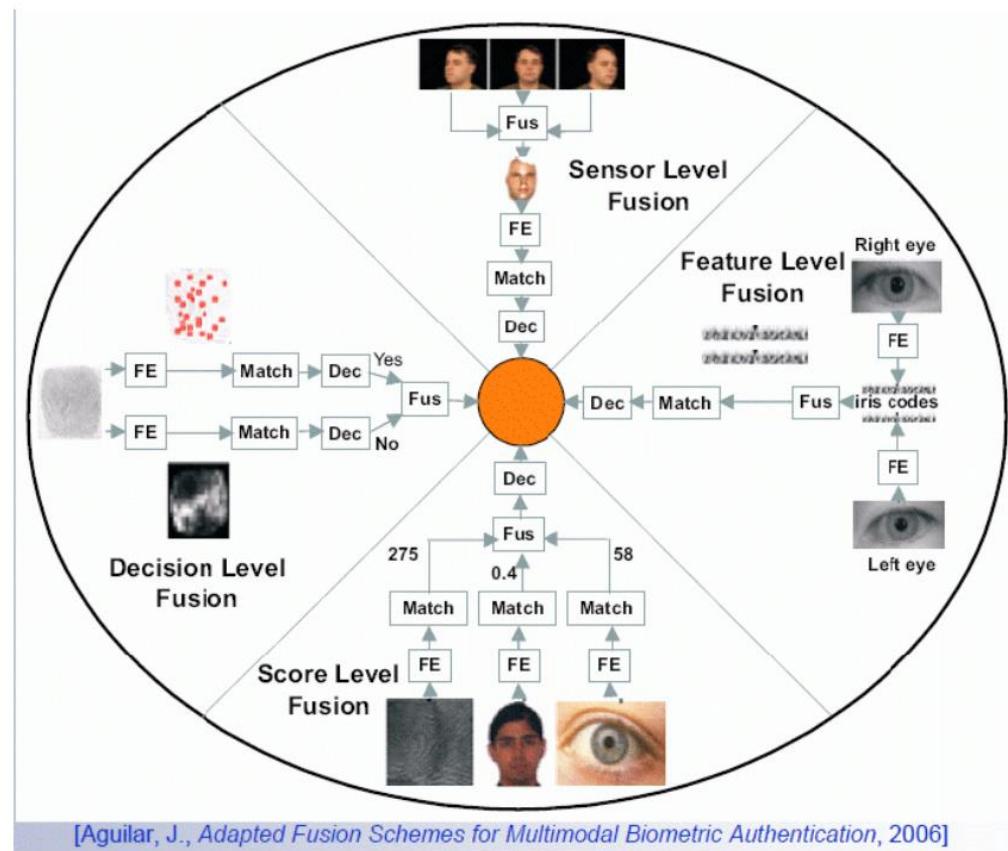
- **Multibiometric:**



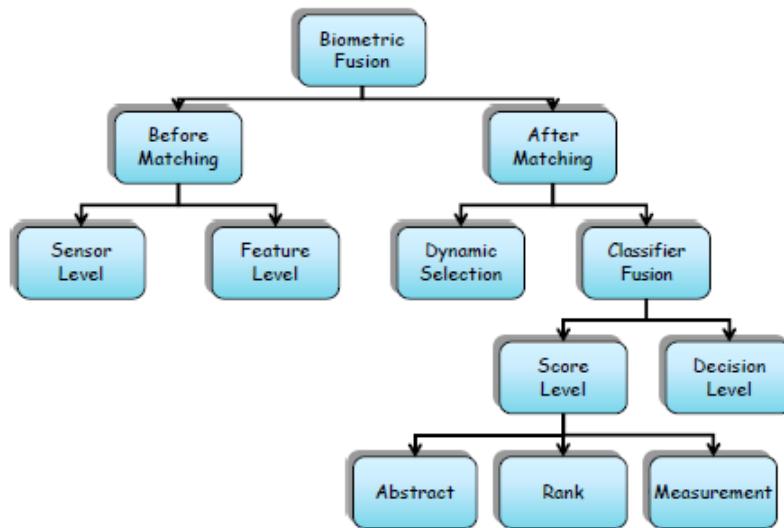
- **Multiexpert:**



Una volta decisa quale tipo di soluzione multi-biometrica adottare, resta da stabilire come combinare le informazioni differenziate.

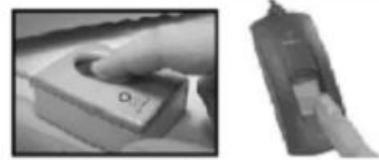


La **combinazione** delle diverse biometrie può essere eseguita in ciascuno dei quattro moduli di sistema.



Scenari di fusione

Unica biometria e Multiplo sensore: sono combinate le informazioni dello stesso tratto biometrico acquisito però con due tipi di sensori diversi.



Multipla biometria: sono combinate le caratteristiche di due biometrie diverse, un volto ed un'impronta.



Unica biometria e Multipla unità: si combinano due elementi diversi appartenenti alla stessa classe biometrica (l'impronta di due dita diverse).



Unica biometria e Multipla acquisizione: si combinano più sessioni di acquisizione dello stesso tratto biometrico (più impressioni della stessa impronta).



Sistemi multimodali: scelte progettuali

Nello sviluppo di un sistema multimodale bisogna considerare diversi aspetti progettuali:

- La scelta delle biometrie.
- Il livello del sistema a cui fare l'integrazione.
- La metodologia utilizzata per l'integrazione.
- Requisiti funzionali.

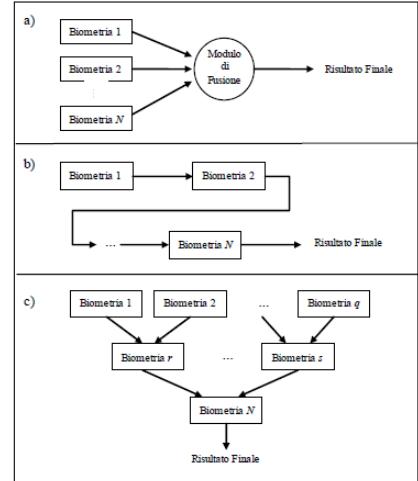
Insieme di biometrie

La scelta del numero e del tipo di biometria da valutare dipende soprattutto dalla natura del sistema. Un sistema governativo ha esigenze di sicurezza e di autenticità dell'utente superiori rispetto ad un sistema commerciale; la scelta di una biometria più affidabile quale l'iride o l'impronta digitale è favorita. Un sistema embedded su telefonini o computer palmari privilegia l'uso di biometrie rilevabili con l'hardware in dotazione al dispositivo, quindi voce, volto o firma.

Architettura del sistema

Un sistema multimodale può essere progettato secondo tre distinte architetture:

- **In parallelo.**
- **In serie.**
- **A livello gerarchico.**



Progettazione in parallelo

Le informazioni estratte dalle molteplici biometrie sono utilizzate simultaneamente per realizzare il riconoscimento. Le acquisizioni sono svolte contemporaneamente e le valutazioni delle singole biometrie sono realizzate indipendentemente e poi combinate mediante opportune tecniche di fusione. Questa tecnica può essere utilizzata per entrambe le modalità di riconoscimento, verifica e identificazione. Offre vantaggi maggiori nella prima, in quanto garantisce un accertamento dell'individuo più forte basando la decisione su più parametri di valutazione.

Progettazione in serie

Le valutazioni delle singole biometrie, generalmente due o tre al massimo, sono effettuate in cascata. Questa tecnica può essere utilizzata in entrambe le modalità di riconoscimento. Essa è più vantaggiosa nella modalità di identificazione, poiché può essere utilizzata come strumento di filtraggio dei possibili candidati ad ogni processo di confronto. Solitamente si utilizzano diversi tipi di biometrie, una prima di rapido confronto ma di limitata attendibilità per eseguire una prima selezione di una rosa di candidati, sui quali applicare, quindi, una seconda biometria più lenta ma più affidabile per la decisione finale.

Progettazione a livello gerarchico

Si usano metodologie di classificazione individuali per le varie biometrie, quindi i risultati sono memorizzati in una struttura ad albero che facilita la decisione finale. Questa modalità è particolarmente utile nel caso di numerose biometrie da integrare.

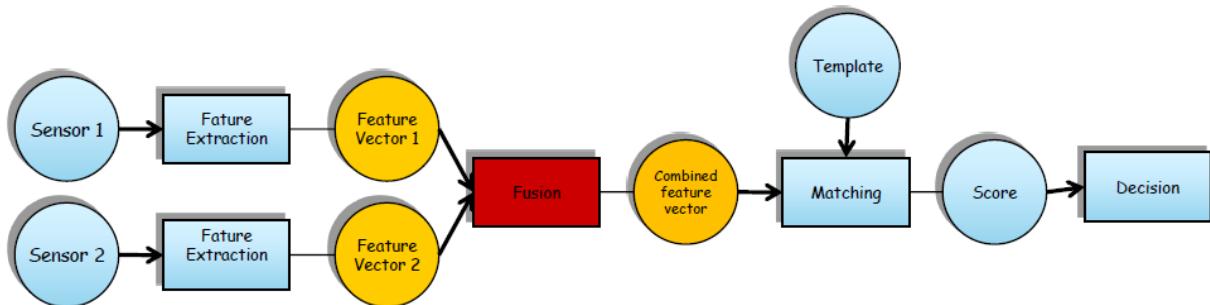
Livelli e strategie di fusione

In un sistema biometrico le fasi di Verifica o Identificazioni sono suddivisibili essenzialmente in quattro moduli:

- Il modulo di acquisizione della specifica biometria,
- Il modulo di estrazione delle caratteristiche e creazione del template relativo alla biometria rilevata,
- Il modulo di confronto,
- Il modulo di decisione.

Feature level fusion

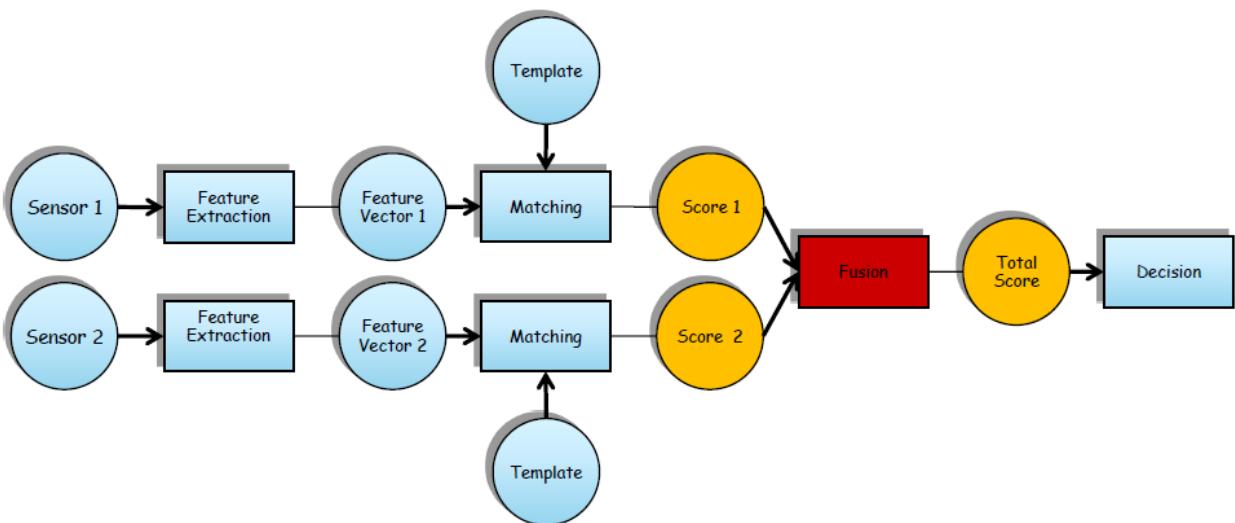
Le funzioni che sono state estratte con tecniche possibilmente diverse possono essere fuse per creare un nuovo vettore di caratteristiche per rappresentare l'individuo.



1. Sono previsti risultati migliori, poiché sono ancora presenti molte più informazioni.
2. Possibili problemi:
3. Feature set incompatibili.
4. La combinazione di vettori di caratteristiche può causare “differenza della dimensionalità”.
5. Potrebbe essere necessario un match più complesso.
6. I vettori combinati possono includere dati rumorosi e / o ridondanti.

Score level fusion

Diversi algoritmi di matching restituiscono un insieme di punteggi che sono fusi per generare un singolo punteggio finale.

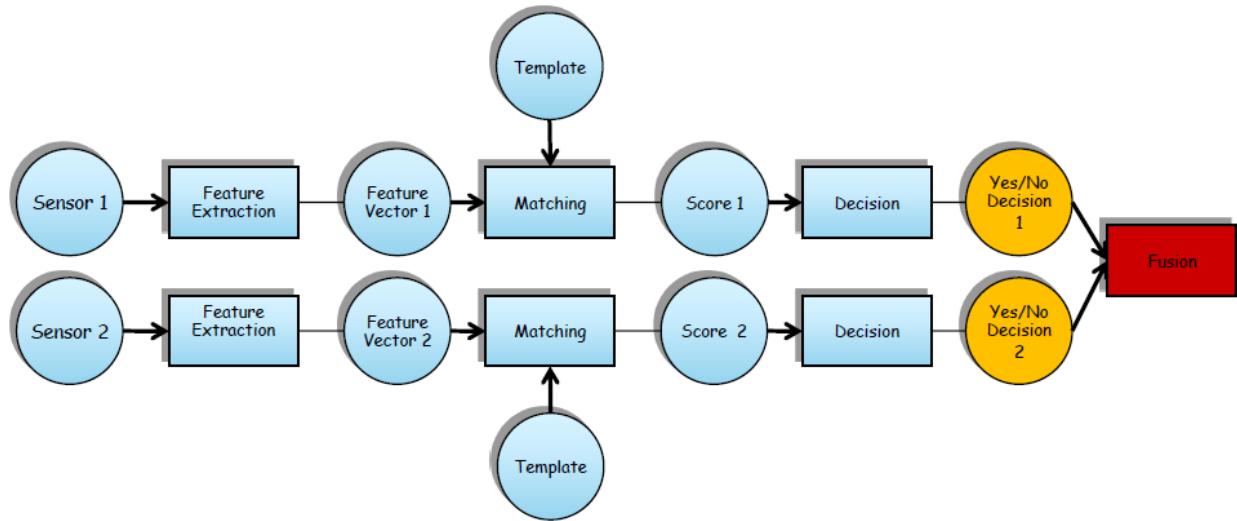


Transformation-based: i punteggi di diversi match vengono prima normalizzati (trasformati) in un dominio comune e poi combinati usando le regole di fusione.

Classifier-based: i punteggi di diversi classificatori sono considerati funzioni e sono inclusi in un vettore di caratteristiche. Un classificatore binario viene addestrato per discriminare tra vettori di punteggi autentici e impostori (NN-Neural Networks, SVM - Support Vector Machine).

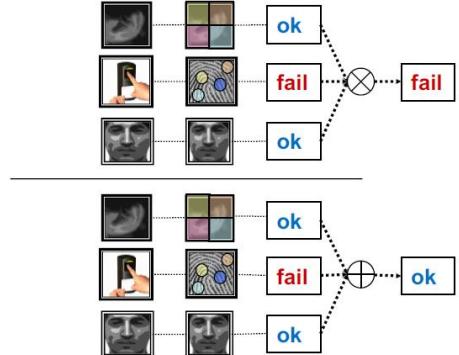
Decision level fusion

Ogni classificatore emette la propria decisione (accetta / rifiuta per verifica o per identificazione). La decisione finale viene presa combinando le singole decisioni secondo una regola di fusione.



Sono possibili diverse strategie di combinazione. I più semplici implicano una semplice combinazione logica:

- **Combinazione seriale di AND.**
 - Richiedete l'autenticazione globale.
 - Tutte le decisioni positive.
 - Questo migliora il FAR.
- **Combinazione parallela di OR.**
 - L'utente può essere autenticato.
 - Anche per una singola modalità biometrica.
 - Questo migliora FRR.
- Un'ulteriore importante regola di fusione a livello decisionale è **Majority Voting**.



Politiche di fusione

Ritardare la fusione risulta sconveniente, perché:

- Il modulo decisionale riceve solo una minima parte delle informazioni acquisite e conosce solo l'esito del confronto delle singole biometrie
- Comporta l'implementazione e quindi l'elaborazione di due sottosistemi ognuno dedicato alla specifica biometria.

Conviene fare uso di una strategia di fusione il prima possibile poiché si conserva una ricchezza di informazioni maggiore fornita dai dati acquisiti.

Il template derivante dalla fusione può risultare di dimensioni considerevoli, diventando pesante da gestire dal sistema incidendo così sulle performance. Questa strategia non sempre è attuabile in quanto spesso si lavora con sistemi chiusi nei quali non sono note le relazioni tra i dati acquisiti e i template generati oppure può essere difficile fondere caratteristiche provenienti da biometrie diverse. La scelta più vantaggiosa resta una fusione al modulo di confronto.

Strategia di fusione

Al livello decisionale è possibile combinare con una strategia **AND/OR**:

- Richiedere che siano superate tutte le prove biometriche (AND) oppure solo alcune (OR).
- Esempio (Volto-Impronta-Mano): Impronta AND (Volto OR Mano).

Al livello di confronto è possibile utilizzare una strategia di combinazione **pesata**:

- Si associa un peso a ciascuna biometria e quindi anche il risultato dei singoli confronti avranno un peso diverso in fase decisionale
- Altre strategie utilizzano funzioni di combinazione matematiche come la somma, il prodotto, la media, o il massimo o minimo dei risultati dei confronti, queste tecniche necessitano di un ulteriore processo di normalizzazione dei tali risultati