

# Indicizzazione e Recupero delle Immagini

- Introduzione
- Differenti approcci di Indicizzazione e Recupero
- Tecniche Text-Based
- Tecniche Color-Based
- Miglioramenti delle tecniche di base
- Retrieving basato sulla forma

# Introduzione

- L'indicizzazione ed il recupero delle immagini costituisce un settore di ricerca in cui si sono raggiunti risultati molto importanti e avanzati.
- I risultati ottenuti in questa di ricerca sono spesso sfruttati nei database relazionali di tipo commerciale.

Nell'indicizzazione e ricerca si distinguono 4 approcci diversi basati su:

1. **Memorizzazione di attributi strutturati**
2. **Riconoscimento degli oggetti**
3. **Annotazioni libere**
4. **Estrazione di Feature di basso livello**

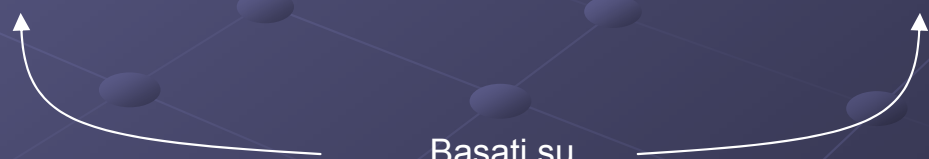
1. e 3. rientrano nei metodi tradizionali

2. e 4. fanno parte del “content-based image retrieval”

# Schema degli approcci principali

Attributi di Struttura	Riconoscimento degli oggetti	Testo	Caratteristiche di basso livello
Nome_file Categoria Data creazione Soggetto Autore ....  DBMS	Computer vision OCR .....	Annotazioni      Incompletezza e soggettività	Colori Texture .....

Basati su contenuti



# Memorizzazione di attributi strutturati

Questa strategia è fondata sulla memorizzazione di informazioni per ognuna delle immagini del D.B.M.S.:

- Nome del file,
- Data di creazione,
- Autore,
- Categoria dell'immagine
- Soggetto
- .....

L'indicizzazione e la ricerca possono quindi essere effettuate con tecniche standard dei DBMS

I limiti più evidenti di questa metodologia sono:

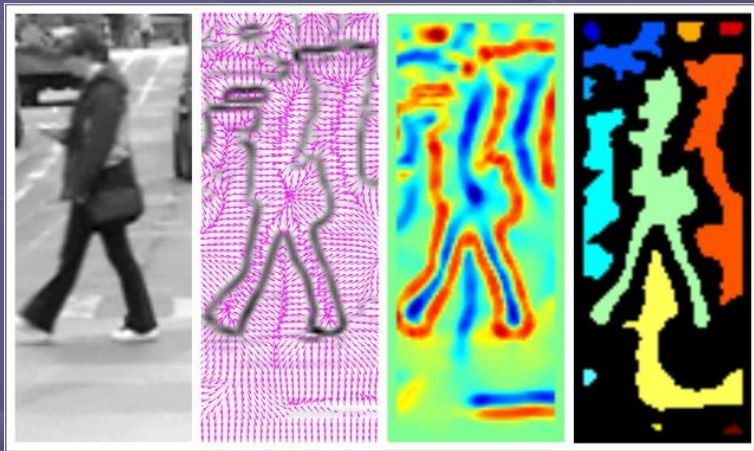
- Gli attributi possono non descrivere in maniera completa le immagini
- Le query sono limitate ai soli attributi memorizzati

# Riconoscimento degli oggetti

Questa strategia è fondata su tecniche sofisticate di estrazione di feature e di algoritmi di riconoscimento degli oggetti presenti nella scena

Caratteristiche principali:

- Sono computazionalmente molto pesanti
- Richiedono algoritmi estremamente sofisticati
- Godono di una buona prestazione solo per domini applicativi ristretti
- Sono argomento di ricerca avanzata (**Computer Vision**)





# Annotazioni libere (Text-based image retrieval)

- Le immagini sono descritte con **testo libero** (non controllato).
- Le query sono nella forma di “**parole chiave**” o testo libero (con eventuale utilizzo di operatori booleani)
- La ricerca usa gli **algoritmi convenzionali di IR** (Information retrieval) basati sulla ricerca di similarità tra query e testo descrittivo delle immagini
- Essendo una tecnica “**manuale**”, occorre notare che:
  - Il testo deve essere il più possibile completo e consistente
  - E' necessario l'utilizzo di un dizionario per ricercare anche su sinonimi (uomo/donna/bambino = persona)
  - Il testo introdotto potrebbe essere soggettivo → occorre utilizzare tecniche di *relevance feedback*

## **Vantaggi:**

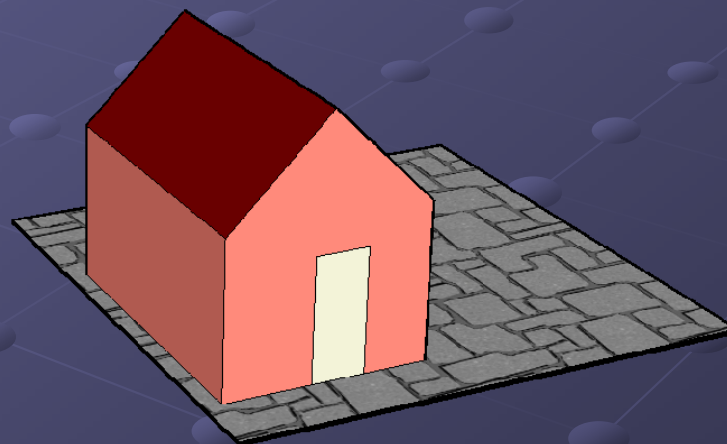
Si possono “catturare” anche concetti di alto livello (cioè astratti) presenti nell'immagine (per es. una persona che sorride o che è triste)

# Estrazione di feature di basso livello

L' estrazione di feature di basso livello è definita su tecniche di indicizzazione e ricerca basate sul contenuto (**content-based**)

Possono prendere in considerazione una o più delle seguenti caratteristiche:

- **COLORE**
- **FORMA**
- **TEXTURE**



Molti sistemi reali sono basati su tali metodi

# Algoritmi basati sull'analisi del colore

Ogni immagine è memorizzata assegnando ai pixel tre valori numerici ad ognuno dei canali di colore (ad esempio RGB)

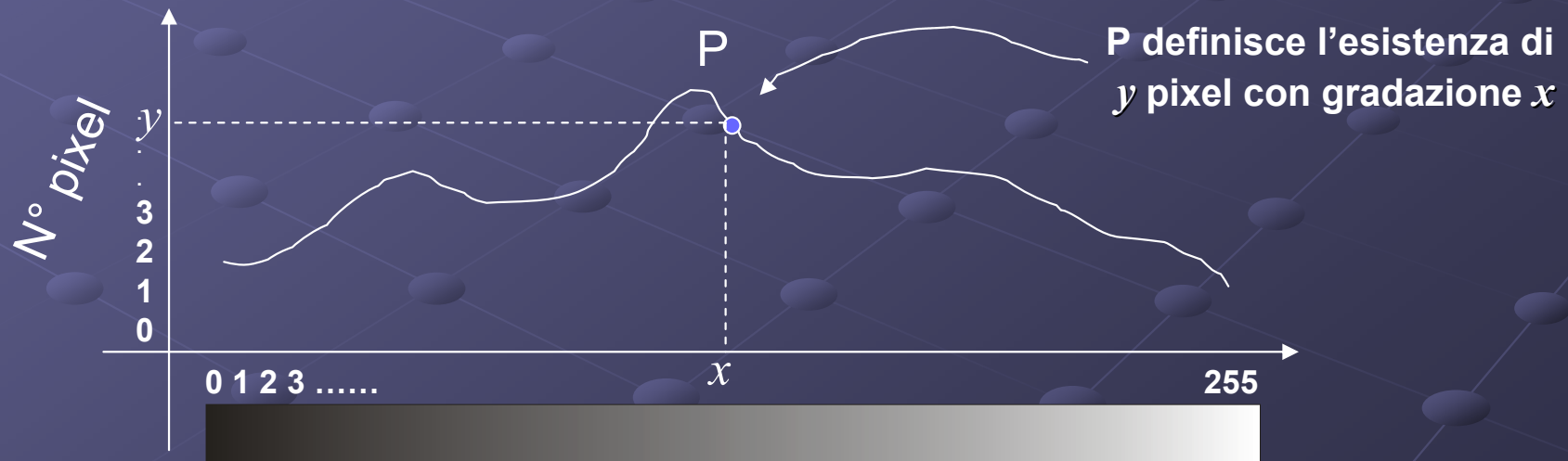
Ogni canale è discretizzato in  $m$  intervalli (**quantizzazione dei colori**): il numero totale di combinazioni diverse (bins) è  $m^3$

(es: 16 intervalli di colore  $\rightarrow$  4096 bins)

Si definisce **ISTOGRAMMA di COLORE**:

il vettore  $H(M) = (h_1, h_2, \dots, h_j, \dots, h_n)$

in cui  $h_j$  rappresenta il numero di pixel dell'immagine  $M$  che ricadono nel bin  $j$ .





# Confronto degli Istogrammi di immagini

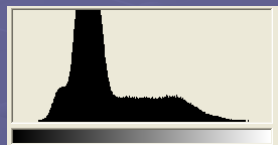
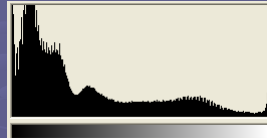
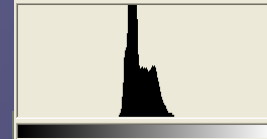


Foto ORIGINALE



ALTO CONTRASTO



BASSO CONTRASTO



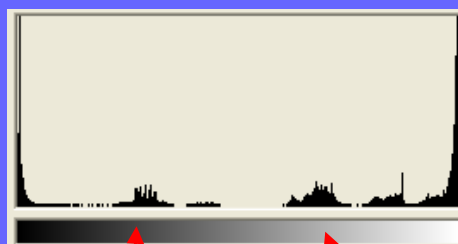
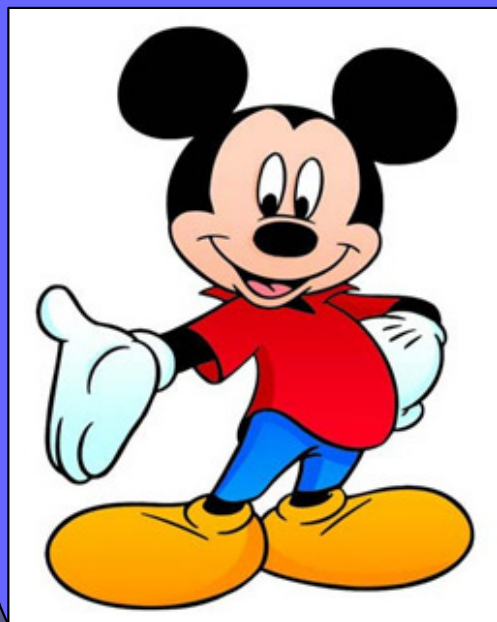
BASSA LUMINOSITA'



ALTA LUMINOSITA'

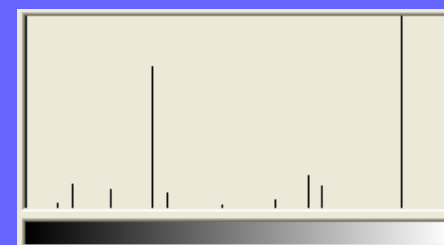
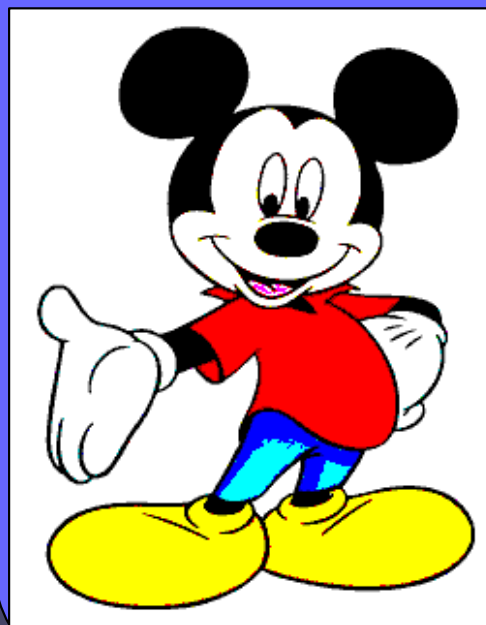
# Confronto degli Istogrammi di immagini (2)

**Disegno con 8 bit/pixel (max colori =  $2^8=256$ )**



Un numero adeguato di colori garantisce “buone” sfumature (passaggi dolci di gradazione): nell’istogramma sono presenti “raggruppamenti”

**Disegno con 4 bit/pixel (max colori =  $2^4=16$ )**



Un numero minore di colori implica un maggiore “schiacciamento” dell’immagine: l’istogramma contiene quindi “segmenti” ben identificabili

# Indicizzazione su istogramma di colore

- Per ogni immagine si calcola l'istogramma di colore  **$H(M)$**  che verrà poi **utilizzato come indice dell'immagine  $M$** .
- Per la ricerca delle immagini nel DB serve definire una **misura di distanza** tra l'istogramma dell'immagine query e quelli delle immagini contenute nel database.
- Date 2 immagini  $A$  e  $B$ , la misura di distanza più semplice è data da:

$$d(A, B) = \sum_{i=1}^n |a_i - b_i|$$

in cui  $a_i$  e  $b_i$  sono il numero pixel delle immagini  $A$  e  $B$  che ricadono nel bin  $i$ -esimo



# Esempio di calcolo della distanza

Si supponga di avere 3 immagini 8x8 pixels aventi i seguenti istogrammi:

$$H1 = (8, 8, 8, 8, 8, 8, 8, 8)$$

$$H2 = (7, 7, 7, 7, 9, 9, 9, 9)$$

$$H3 = (2, 2, 10, 10, 10, 10, 10, 10)$$

Quindi, le distanze tra le immagini sono rispettivamente:

$$d(H1, H2) = 1+1+1+1+1+1+1+1=8$$

$$d(H1, H3) = 6+6+2+2+2+2+2+2=24$$

$$d(H2, H3) = 5+5+3+3+3+1+1+1=23$$

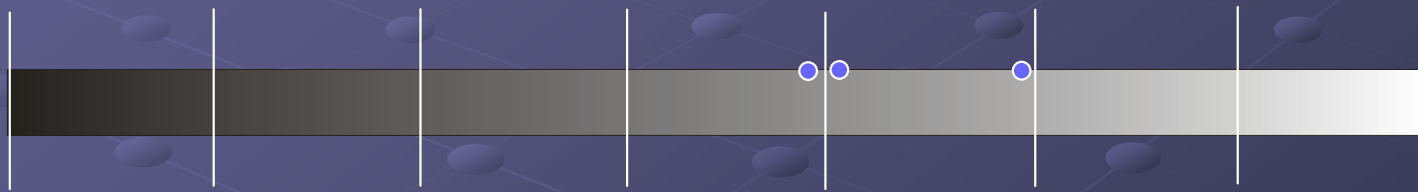
Pertanto le immagini 1 e 2 sono più simili che 1 e 3 oppure 2 e 3

Tale metodo di base è soggetto ad una serie di problemi e può fornire risultati fortemente incorretti se non si applicano delle varianti necessarie per trattare in maniera migliore l'informazione sul "colore".

# Problemi relativi all'indicizzazione tramite istogramma del colore

La discretizzazione dello spazio dei colori in classi (bins) non tiene conto della similarità dei colori:

- Due bins adiacenti vengono considerati totalmente diversi
- Il posizionamento della linea di demarcazione tra i bins influisce fortemente sull'istogramma e sul calcolo della distanza



La rappresentazione dei colori non e' univoca:

- Dipendenza dal sistema di rappresentazione
- Dipendenza dal device



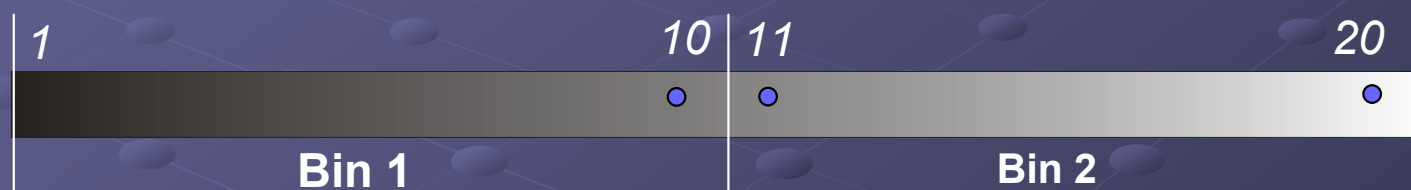
# Istogrammi del colore e similarità

- L' esempio precedente non considera la similarità tra i colori dei bins:
- due immagini con colori percepiti come simili ma numericamente diversi risultano molto distanti tra di loro.

## *Esempio:*

Si supponga di avere due bin che rappresentano gli insiemi di colori 1-10 e 11-20 (in cui i colori con numeri vicini sono simili tra di loro). Quindi il colore 10 viene assegnato al bin 1, ed i colori 11 e 20 vengono assegnati al bin 2

Per paradosso il colore 11 viene considerato uguale al 20 ma diverso dal 10



- Sono stati proposti vari metodi per superare tale problema:
  - Distanza tra i bins
  - Istogramma cumulativo
  - Istogramma pesato percettivamente

# Distanza tra i bins

La distanza tra bins si definisce come misura di similarità (inverso della distanza) calcolata bin-per-bin. E' quindi possibile definire l'istogramma  $Z$  di similarità tra due istogrammi  $X$  ed  $Y$  nel seguente modo:

$$\|Z\| = Z^T A Z$$

in cui:

$A$  è la matrice simmetrica di similarità dei colori in cui  $a(i,j) = a(j,i) = 1 - d(c_i, c_j)/d_{max}$

$d(c_i, c_j)$  è la distanza tra i colori in uno spazio di uniforme

$d_{max}$  è la distanza massima tra due colori

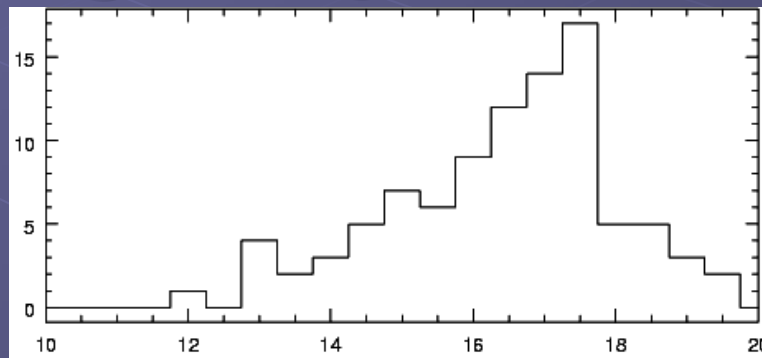
Se due colori sono molto diversi  $d(c_i, c_j)$  e' prossima a  $d_{max}$  e  $a(i, j)$  è vicino allo 0 contribuendo al calcolo della similarità

# Istogramma cumulativo

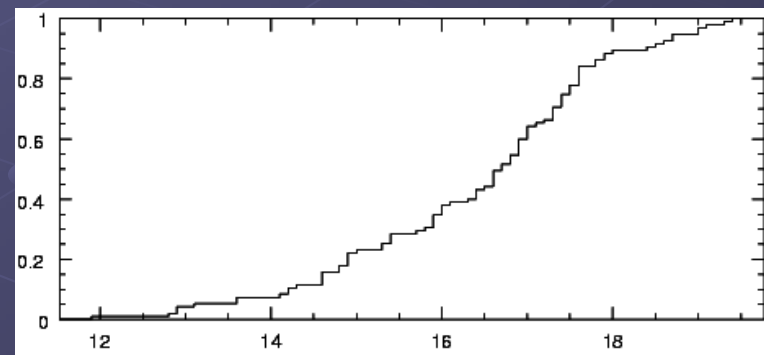
L'istogramma cumulativo non considera la distanza tra i bin ma risolve parzialmente il problema creando delle classi cumulative.

Un istogramma cumulativo (Cumulative Histogram) per un'immagine  $M$  è definito da:

$$\mathbf{CH}(M) = (ch_1, ch_2, \dots, ch_n) \quad \text{in cui} \quad ch_i = \sum_{j \leq i} h_j$$



istogramma



istogramma cumulativo

Per il calcolo della distanza si usa la normale [distanza euclidea](#).

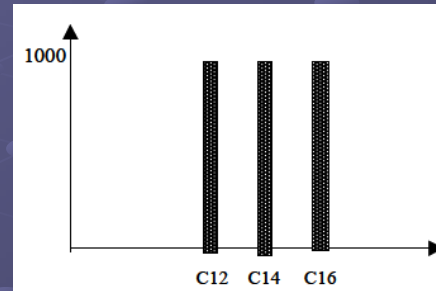
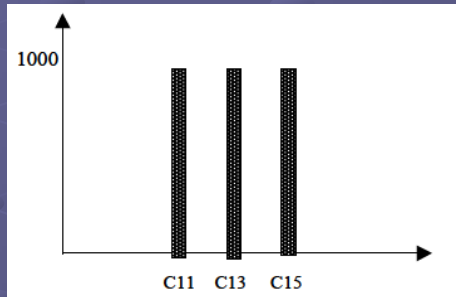
Prestazione: buona per valori bassi di  $i$ . Per valori alti non riflette la similarità tra i colori

# Istogramma pesato percettivamente (PWH - **P**ercentually **W**eighted **H**istogram)

A causa del processo di quantizzazione dei colori e di “mappatura” sui bin corrispondenti, **un singolo colore originario potrebbe essere simile a colori appartenenti a più di un unico bin**  $\Rightarrow$  immagini costituite da colori simili potrebbero avere istogrammi la cui misurazione di distanza genera un valore elevato  
Se si assume che i bin siano ordinati da C1 a Cn, l'esempio di seguito illustrato mostra gli istogrammi di due immagini A e B che non avendo alcun colore comune avranno una distanza pari a 6000

RGB

H(A)



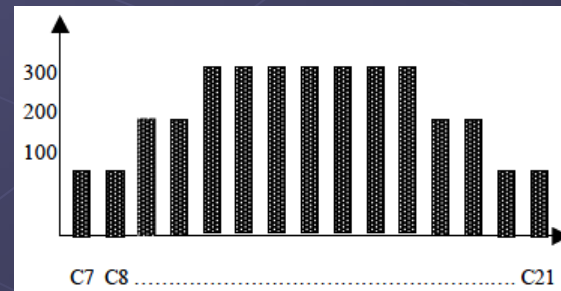
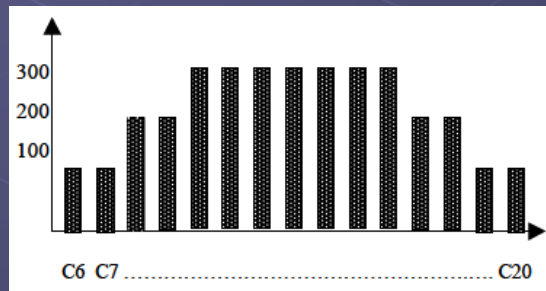
H(B)

Una soluzione consiste nel **cambiare lo spazio di rappresentazione dei colori**, da **RGB** a **CIE Luv**  
Il cambiamento di rappresentazione dei colori comporta che ogni singolo colore in RGB può avere 10 diverse **corrispondenze “pesate”** in CIE Luv (si assume che per ogni pixel in RGB, la somma dei pesi relativi alle corrispondenze in CIE Luv sia uguale a 1).

Infine si costruisce l'istogramma dell'immagine rappresentata nel nuovo dominio CIE Luv

CIE Luv

H(A)



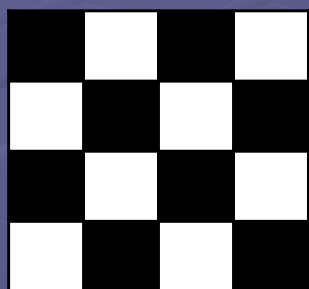
H(B)

Ora la nuova “distanza” è 600 (cioè molto inferiore alla precedente cioè 6000)  $\Rightarrow$  maggiore similitudine

# Limiti dell'approccio Color-Based: relazioni spaziali

Una limitazione dei sistemi di indicizzazione di immagini basate sui colori consiste nel fatto che generalmente **vengono ignorate le relazioni spaziali tra i pixel**.

Ad esempio le due immagini seguenti verrebbero sempre considerate uguali:

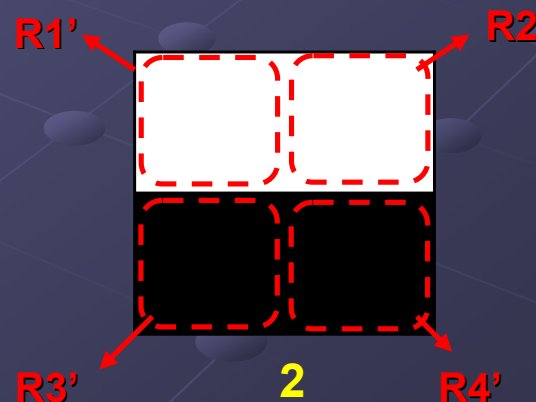
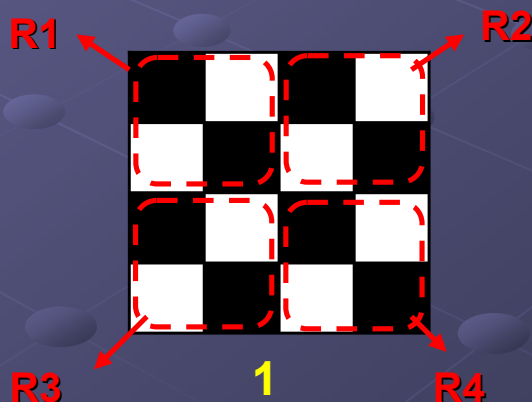


1



2

Per risolvere il problema si possono utilizzare tecniche che prevedono la **suddivisione delle immagini in regioni più piccole** e applicare la tecnica basandosi sugli istogrammi delle sotto-regioni corrispondenti delle immagini





# Limiti dell'approccio Color-Based: background

## Mascheramento ad opera dello sfondo

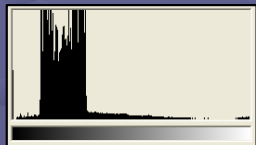
Gli istogrammi sono fortemente condizionati dalla presenza di grandi blocchi di colore omogeneo e risulta che le immagini che contengono un oggetto in primo piano (**foreground**) siano trattate in maniera errata in quanto nell'istogramma i pixel dello sfondo (**background**) sono maggiormente rappresentati e “mascherano” l'oggetto in primo piano.

### *Esempio:*

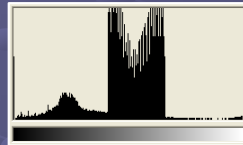
Immagine 1: macchina rossa su sfondo nero

Immagine 2: macchina rossa su sfondo marrone

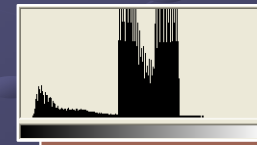
Immagine 3: macchina blu su sfondo marrone



1



2



3

2 e 3 sono considerate più vicine rispetto a 1 e 2

La possibile soluzione può consistere nell'utilizzare **tecniche di segmentazione** (automatiche o semiautomatiche) per suddividere le immagini in soggetto e sfondo e, quindi, calcolare separatamente i due istogrammi usandoli singolarmente

## Indicizzazione e Ricerca basate sulla forma

- Si basa su algoritmi di segmentazione delle immagini che sono in grado di suddividere una immagine in singoli oggetti attraverso metodi automatici o semiautomatici.
- I sistemi efficienti in tale ambito devono possedere le seguenti proprietà:
  - Ogni forma deve avere una rappresentazione unica, invariante rispetto a **traslazione, rotazione o scala**



- Forme simili devono avere una rappresentazione simile in modo tale che la ricerca possa essere basata sulla distanza tra la rappresentazione delle forme
- Il processo di ricerca è spesso realizzato in modalità “query by example” e il sistema ritorna come risposta la forma che meglio approssima quella disegnata dall’utente

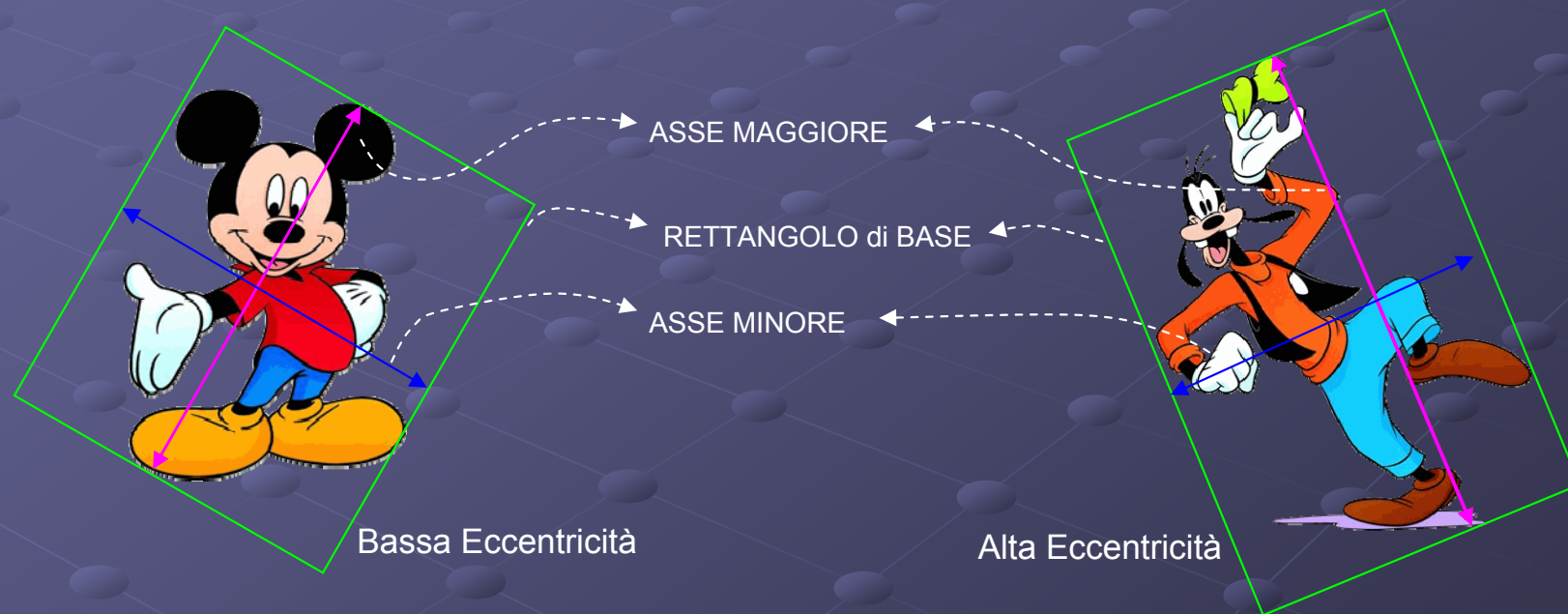
# Definizioni

**ASSE MAGGIORE:** segmento che congiunge i due punti della forma che sono più distanti fra di loro

**ASSE MINORE:** segmento perpendicolare all'asse maggiore e tale che il rettangolo il cui lati sono paralleli ai due assi racchiuda completamente l'intera forma

**RETTANGOLO DI BASE:** il rettangolo descritto nella precedente (esso coincide con il più piccolo rettangolo che contiene l'intera figura)

**ECCENTRICITA':** rapporto tra la lunghezza dell'asse maggiore e la lunghezza dell'asse minore



Le caratteristiche elencate possono essere considerate basilari in quanto forniscono una prima caratterizzazione delle forme e sono utilizzati durante l'indicizzazione e la ricerca

# Descrizioni numeriche delle forme

Per la descrizione numerica delle forme sono state adottate molteplici strategie tra cui:

- Metodi basati su **formulazioni matematiche** che descrivono un'immagine come funzione  $f(x,y)$  dove  $x$  e  $y$  sono le coordinate di un pixel:
  - Momenti invarianti
  - Momenti centrali normalizzati
- Metodi che cercano di descrivere le forme mediante **le coordinate dei punti** che le definiscono
  - Metodo dei descrittori di Fourier
  - Istogramma dei lati significativi
  - Lista ordinata dei punti interessanti
  - Elastic template matching

Una limitazione “classica” delle precedenti metodologie è dovuta al fatto che trattano con difficoltà l'invarianza rispetto alle traslazioni, rotazioni e cambiamenti di scala necessari per ottenere un buon sistema di indicizzazione e ricerca basato sulle forme.

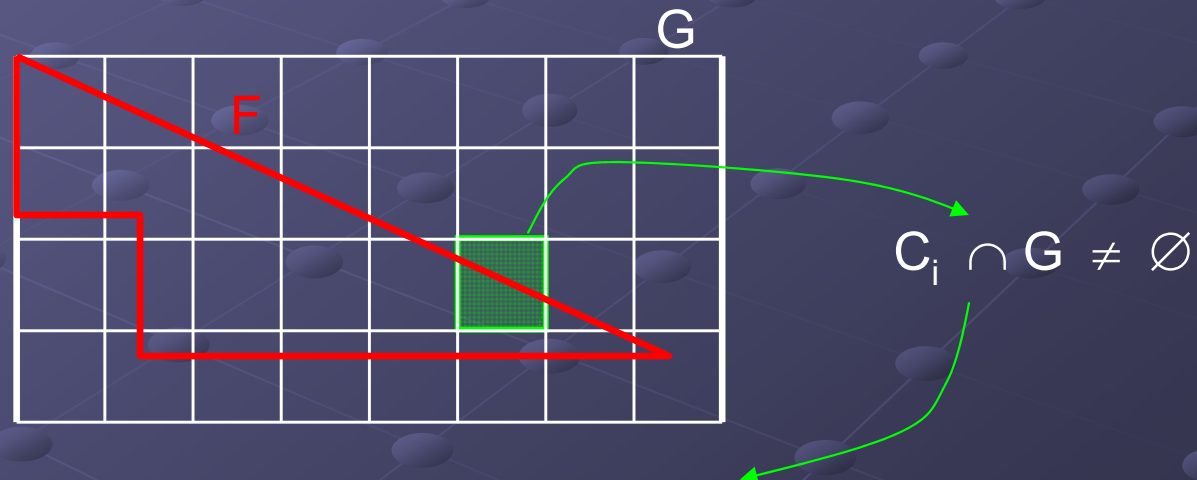


# Rappresentazione delle forme basata su REGIONI

La rappresentazione delle forme basata su regioni è una metodica semplice i cui risultati sono molto promettenti. Data una forma  $F$ , la metodologia adottata è così descritta:

- Si sovrappone una griglia  $G$  che contenga completamente  $F$ .
- Per ogni cella  $C_i$  di  $G$ , se  $C_i \cap F = \emptyset$  allora  $C_i = 0$  altrimenti  $C_i = 1$
- Si ottiene una stringa binaria che rappresenta la forma leggendo gli 1 o 0 da sinistra a destra e dall'alto al basso.

*Esempio:*



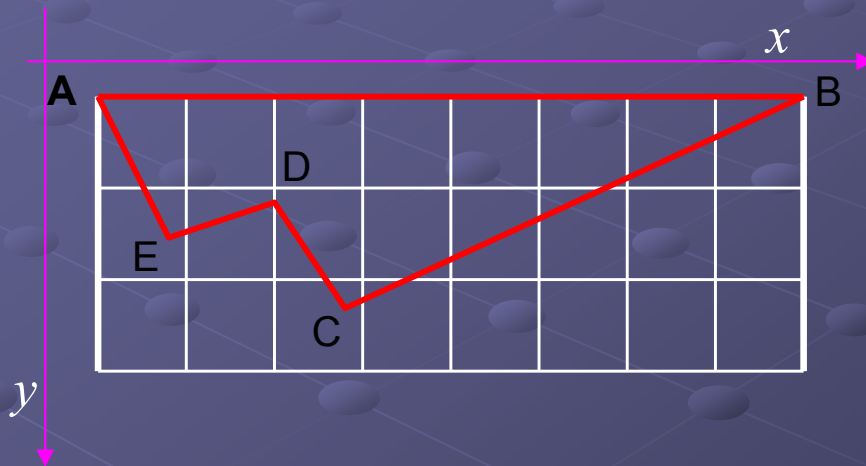
*Rappresentazione:* 11100000111110000111111001111111



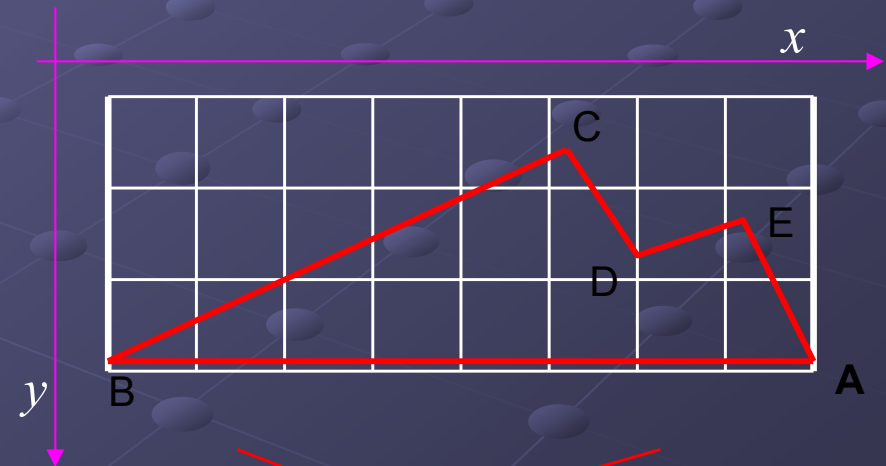
# Normalizzazione rispetto alle rotazioni

Per ottenere l'invarianza rispetto alla rotazione, la forma viene ruotata in modo tale che il suo asse maggiore sia allineato con l'asse  $x$  (orizzontale): Il suddetto allineamento è ugualmente possibile ruotando tutto di  $180^\circ$  ed andrebbero quindi esaminati 2 casi.

Per evitare di raddoppiare le rappresentazioni di stringhe binarie, generalmente si preferisce, invece, di "raddoppiare" la query e cioè, in altre parole, si effettua una trasformazione sulla query stessa che viene riproposta quindi per la seconda volta.



Rappresentazione: Riga1,Riga2,Riga3



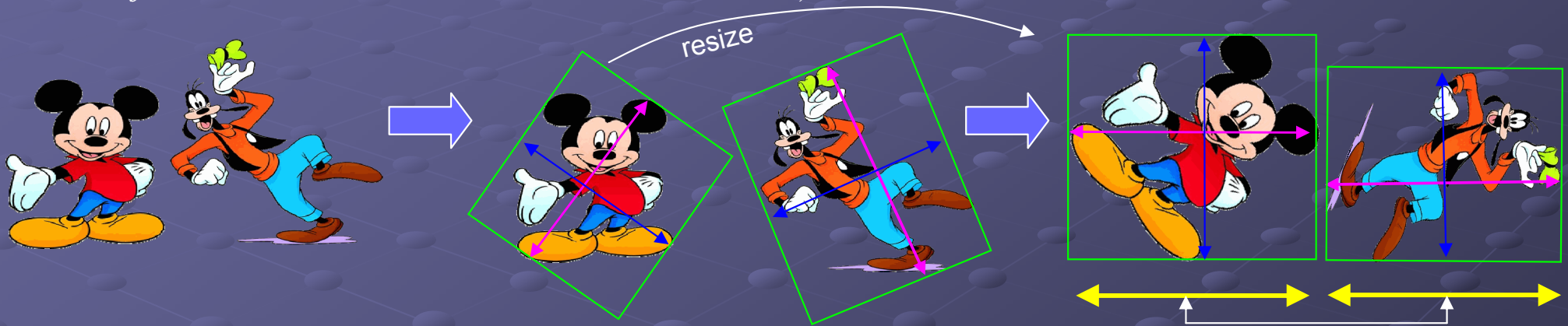
~~Rappresentazione: Riga3<sub>rev</sub>, Riga2<sub>rev</sub>, Riga1<sub>rev</sub>~~

Query1 = Riga1, Riga2, Riga3

Query2 = Riga3<sub>rev</sub>, Riga2<sub>rev</sub>, Riga1<sub>rev</sub>

# Normalizzazione rispetto ai cambiamenti di scala

E' possibile effettuare una normalizzazione di scala ridimensionando tutte le forme presenti nel DataBase in modo tale che abbiano tutte lo stesso asse maggiore (di numero di pixel). In altre parole, fissata la larghezza  $x$  di una forma è possibile che possa variare solo la sua altezza  $y$  (finché non se ne conosce anche l'eccentricità)



Il criterio di **univocità di rappresentazione di una forma** è pienamente determinato se:

- Viene fissata la CellSize (per es. 10x10 pixel)
- Si presume che l'asse maggiore sia univoco
- Si effettuano le operazioni di normalizzazione di Rotazione
- Si effettuano le operazioni di normalizzazione di Scala
- Per ogni forma si conosce la sua eccentricità

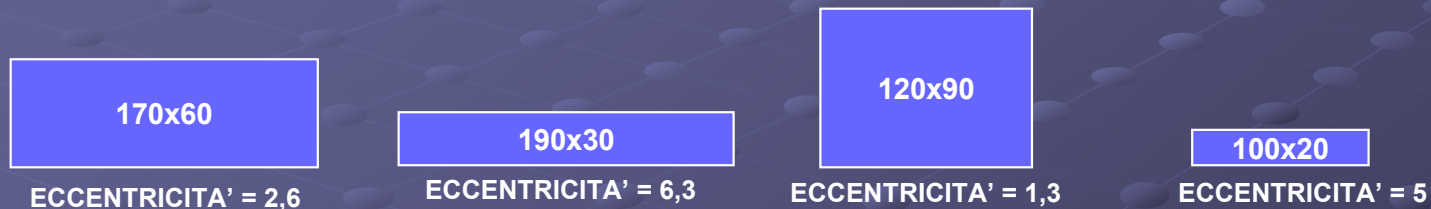
La coppia (**Stringa\_binaria**; **Eccentricità**) determina univocamente la forma

# Calcolo della similarità

Il calcolo della similarità viene generalmente compiuto in 2 fasi:

## FASE 1: **Eccentricità molto diverse**

Se due forme hanno eccentricità molto diversa tra di loro, allora non c'è alcun bisogno di calcolare la similarità in quanto possiamo assumere che, di conseguenza, le figure siano molto diverse.



## FASE 2: **Eccentricità uguale o simile**

- Se due forme hanno **uguale** eccentricità (e quindi la loro stringa binaria ha uguale lunghezza), la distanza tra di loro si calcola come numero di posizioni delle stringhe che hanno valore diverso

*Esempio:*

tra **1111111111100000** e **1111111111111100** la distanza è **3**

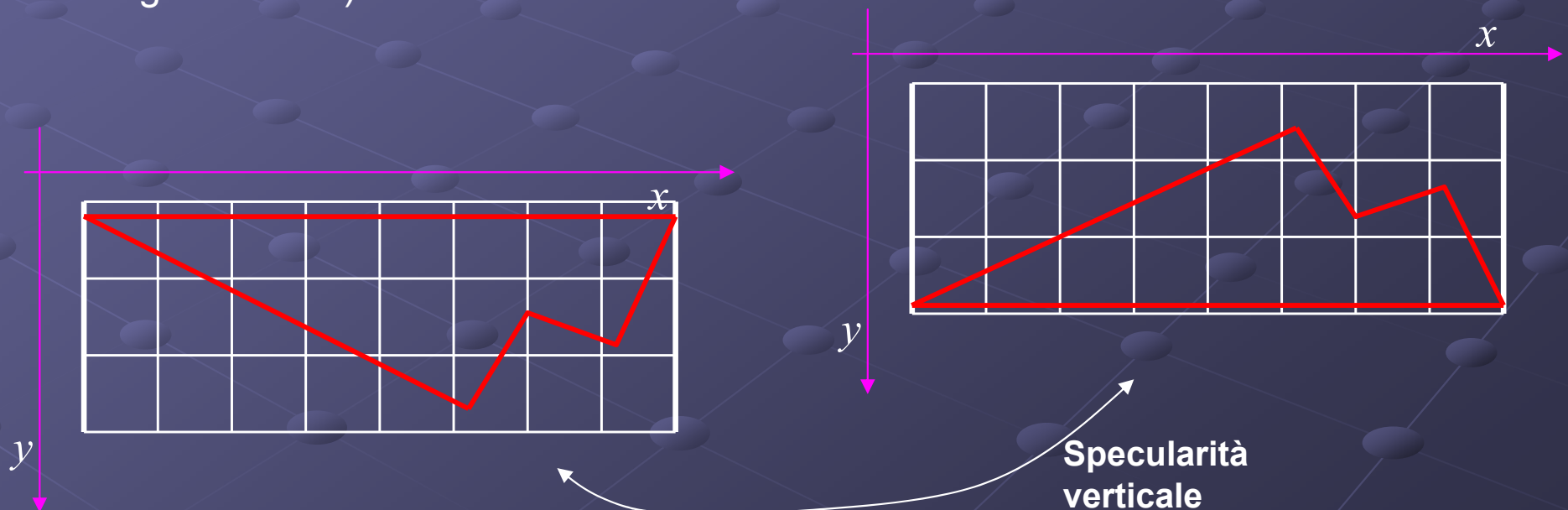


- Se due forme hanno **eccentricità diversa di poco**, si possono aggiungere zeri alla stringa che descrive la forma più "bassa" e quindi applicare la distanza come sopra

# Forme speculari

Un tipo di normalizzazione “estesa” considera anche l’ulteriori trasformazioni che potrebbero riguardare una figura: una buona strategia di riconoscimento dovrebbe essere resistente anche rispetto alle **trasformazioni di specularità orizzontale e verticale**:

Anche in questo caso si preferisce non appesantire il database memorizzando 2 sequenze come indice della forma ma, piuttosto, è preferibile effettuare un maggior numero di ricerche **applicando tali trasformazioni alla forma da ricercare** (considerando anche la rotazione di  $180^\circ$ , ogni ricerca dovrà essere realizzata con 4 stringhe diverse)





# Riepilogo

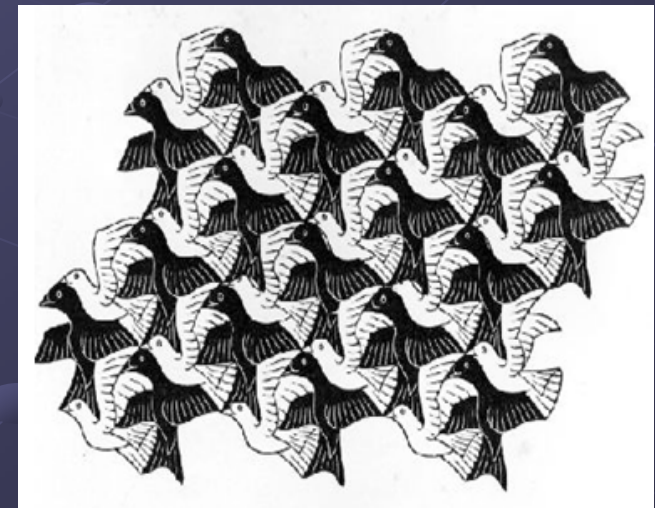
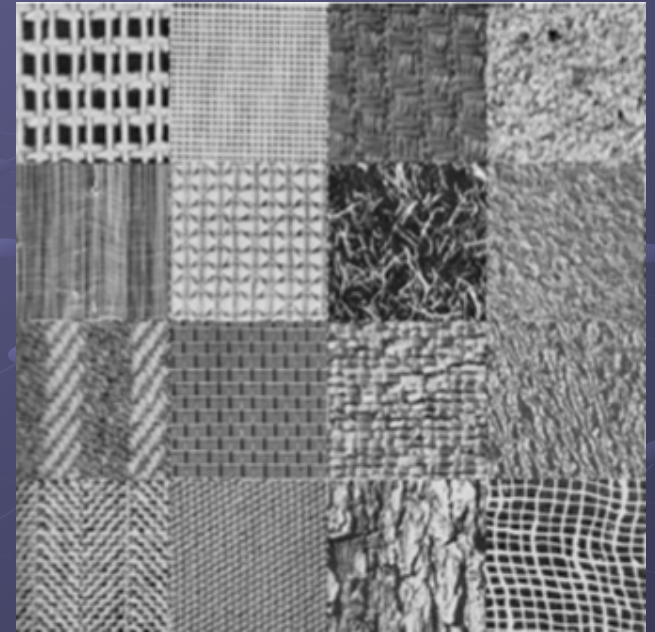
- Un sistema di indicizzazione e ricerca di forme basato su Regioni può essere realizzato in modo che sia invariante a rotazioni, traslazioni, scalatura e trasformazioni speculari, nel seguente modo:
  1. Si determinano gli **assi maggiore e minore** e l'**eccentricità** della forma
  2. Si **ruota** la forma in modo che l'asse maggiore coincida con l'asse X
  3. Si **scala** la forma in modo che l'asse maggiore abbia una lunghezza prefissata
  4. Si sovrappone una **griglia regolare** alla forma
  5. Si assegnano 1 e 0 alle celle e si calcola la stringa binaria scorrendo le celle del grigliato da sx a dx e dall'alto in basso
  6. L'indice della forma è costituito dalla stringa binaria e dal numero di celle che la forma occupa in direzione Y (le due informazioni sono memorizzate nel database come rappresentazione della forma stessa)
- Durante la ricerca la forma da trovare è elaborata in modo analogo ottenendo 4 stringhe diverse (rotazione di  $180^\circ$  e 2 mirroring) e per ogni forma presente nel database che ha una eccentricità simile **si calcola la distanza** (si considera la distanza minima tra le quattro rappresentazioni della forma)



# Metodi basati sulle texture

## Texture Analysis

- La Texture (detta anche tessellatura) descrive una “percezione” dell’immagine che è difficilmente descrivibile e riguarda aree caratterizzate da comuni caratteristiche di intensità e struttura
  - Granularità (fine o grossa)
  - Contrasto
  - Direzionalità (direzine dominante della texture)
  - **Regolarità**
  - ecc...
- Ci sono esempi di sistemi reali che utilizzano una misurazione matematica (più o meno sofisticata) di tali concetti e la utilizzano per indicizzare e ricercare immagini.



# Altre metodologie

La compressione di immagine sfrutta già **caratteristiche “intime” dell’immagine** stessa. Proprio questi parametri di “sintesi” possono essere sfruttati per indicizzare l’immagine

*esempio:*

i coefficienti DCT nella compressione JPEG

i coefficienti Wavelet nella compressione Wavelet

Memorizzazione basata su un **modello descrittivo dell’immagine**

*esempio:*

utilizzo delle tecniche di compressione FRATTALE

Memorizzazione delle **relazioni tra gli oggetti che sono presenti in una immagine**. E’ simile all’utilizzo della informazione topologica nei GIS: tale relazione consente di inviare al sistema richieste quali:

“trova tutte le immagini contenenti un lago vicino ad una montagna“

*oppure*

“trova le immagini contenenti una autostrada sulla sinistra di una foresta”