Riconoscimento immagini

Raimondo Schettini
DISCo - Universita' di Milano Bicocca
Raimondo.schettini@unimib.it

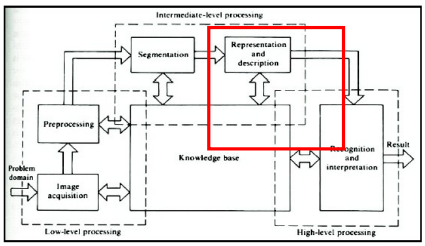
1

I docenti per lezioni ed esercitazioni si avvalgono di slide. Le slide superano abbondantemente il migliaio. Sono state fatte, rifatte, perfezionate negli anni, ma per quanto possano essere ben fatte non saranno saranno mai, da sole, un esaustivo supporto per lo studio. Per comprendere gli argomenti si suggerisce caldamente di seguire attivamente il corso e di prendere appunti. Per lo studio a casa si suggerisce di usare le slide e gli appunti come indice agli argomenti da studiare sul libro, o sui libri a disposizione.
Da quest'anno le slide verranno rese disponibili PRIMA delle lezioni.

Le slide sono rese disponibili in formato elettronico e sono per uso personale.

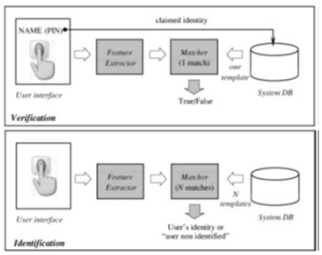
2

Elaborazione delle immagini



3

Riconoscimento vs. classificazione



Riconoscimento o verifica

Classificazione o identificazione

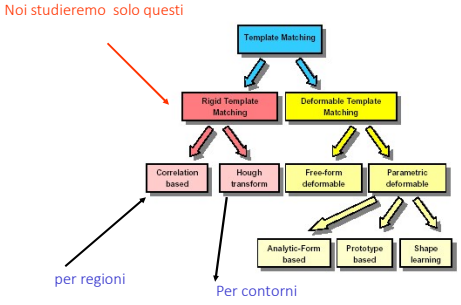
4

Riconoscimento mediante template matching

5

Template matching

Noi studieremo solo questi



6

Template matching

- Il termine **rigid template matching** è molto generico nell'ambito del Pattern recognition, ma normalmente fa riferimento allo "ricerca" di un template **T** all'interno di un'immagine **I** con l'obiettivo di determinare se **I** contiene l'oggetto (*match*) e in quale posizione **T** appare nell'immagine.
- **Global template matching:** tutto l'oggetto è ricercato nell'immagine
- **Local template matching** se si cerca una caratteristica visuale, un particolare come un corner.

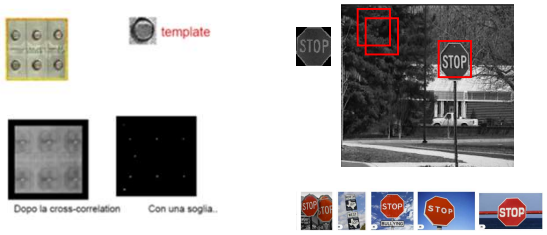
7

Template matching

- Il template **T** è costituito da un oggetto **rigido** (normalmente una immagine in formato raster).
- **T** viene sovrapposto a **I** in tutte le possibili posizioni (**rispetto agli assi X e Y**), ma a seconda dell'applicazione, può essere anche necessario **ruotarlo** e **scalarlo**.
  - Nel seguito denominiamo **T**i le istanze di **T** ottenute dalle trasformazioni precedenti (spostamento in X e Y, rotazione, scala).
- Per ogni istanza **T**i il grado di similarità viene solitamente calcolato minimizzando la distanza o massimizzando la **correlazione** con la porzione di immagine **I** "coperta" da **T**i (che ha la stessa dimensione di **T**i).
- Se l'oggetto è parzialmente occluso, o distorto avremo una matching errato.
- Se ci sono cambiamenti nella scena (illuminazione, ad esempio) il matching potrebbe essere compromesso.

8

Template matching



9

Template matching

Guardiamo al problema della localizzazione di una data immagine di riferimento (template) R all'interno di un'immagine di dimensioni maggiori **I**, che chiamiamo l'immagine di ricerca. Il compito è quello di trovare quelle posizioni in cui il contenuto dell'immagine di riferimento R e la corrispondente sotto-immagine di I sono uguali o più simili.

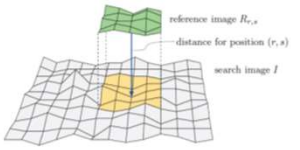
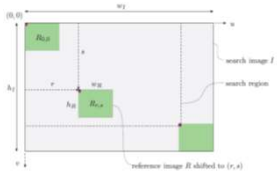


Figure 11.3 Measuring the distance between two-dimensional image functions. The reference image R is positioned at offset (r, s) on top of the search image I.

10

Template matching



L'immagine di riferimento R viene spostata attraverso l'immagine di ricerca I da un offset (r, s) utilizzando le origini delle due immagini come punti di riferimento. Le dimensioni dell'immagine di ricerca ($w_I \times h_I$) e dell'immagine di riferimento ($w_R \times h_R$) determinano la regione di ricerca massima per questo confronto.

11

Template matching

Sono state proposte diverse misure di distanza per le immagini di intensità bidimensionale, comprese le seguenti tre definizioni di base:.

Sum of absolute differences:.

$$d_A(r,s) = \sum_{(i,j)\in R} |I(r+i, s+j) - R(i,j)|;$$

Maximum difference:.

$$d_M(r,s) = \max_{(i,j)\in R} |I(r+i, s+j) - R(i,j)|;$$

Sum of squared differences:.

$$d_E(r,s) = \left[\sum_{(i,j)\in R} (I(r+i, s+j) - R(i,j))^2 \right]^{1/2}.$$

12

Template matching

Per trovare la migliore corrispondenza tra l'immagine di riferimento R e l'immagine di ricerca I, è sufficiente ridurre al minimo il quadrato di dE (che è sempre positivo), che può essere espanso a

$$d_E^2(r,s) = \sum_{(i,j)\in R} \left(I(r+i,s+j) - R(i,j)\right)^2 \qquad (11.5)$$
$$= \underbrace{\sum_{(i,j)\in R} I^2(r+i,s+j)}_{A(r,s)} + \underbrace{\sum_{(i,j)\in R} R^2(i,j)}_{B} - 2\cdot \underbrace{\sum_{(i,j)\in R} I(r+i,s+j)\cdot R(i,j)}_{C(r,s)}.$$

13

Template matching

$$d_E^2(r,s) = \sum_{(i,j)\in R} \left(I(r+i,s+j) - R(i,j)\right)^2 \qquad (11.5)$$
$$= \underbrace{\sum_{(i,j)\in R} I^2(r+i,s+j)}_{A(r,s)} + \underbrace{\sum_{(i,j)\in R} R^2(i,j)}_{B} - 2\cdot \underbrace{\sum_{(i,j)\in R} I(r+i,s+j)\cdot R(i,j)}_{C(r,s)}.$$

Il termine B in Eqn. (11.5) è la somma dei valori al quadrato dei pixel dell'immagine di riferimento R, un valore costante (indipendente da r, s) che può quindi essere ignorato. Se supponiamo che A(r,s) - l' "energia del segnale" - in Eqn. (11.5) sia costante in tutta l'immagine I, allora anche A(r, s) può essere ignorato e la posizione della massima **Correlazione Trasversale – Cross Correlation -** C(r, s) coincide con la migliore corrispondenza tra R e I massimizzando

$$\underbrace{\cdot \sum_{(i,j)\in R} I(r+i,s+j)\cdot R(i,j)}_{C(r,s)}.$$

14

Template matching

L'ipotesi fatta sopra che A(r, s) sia costante non vale per la maggior parte delle immagini, e quindi il risultato della correlazione incrociata varia fortemente con i cambiamenti di intensità nell'immagine I. La **correlazione incrociata normalizzata** -**Normalized Cross correleation -** compensa questa dipendenza tenendo conto dell'energia nell'immagine di riferimento e nella sotto-immagine corrente:

$$C_N(r,s) = \frac{C(r,s)}{\sqrt{A(r,s)\cdot B}} = \frac{C(r,s)}{\sqrt{A(r,s)}\cdot\sqrt{B}}$$
$$= \frac{\sum_{(i,j)\in R} I(r+i,s+j)\cdot R(i,j)}{\left[\sum_{(i,j)\in R} I^2(r+i,s+j)\right]^{1/2}\cdot\left[\sum_{(i,j)\in R} R^2(i,j)\right]^{1/2}}.$$

$C_N(r, s)$ è sempre nell'intervallo [0, 1], indipendentemente dai rimanenti contenuti in I e R.

15

Template matching

a b

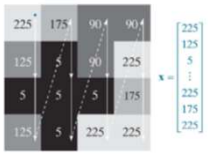


FIGURE 13.1
Using linear indexing to vectorize a grayscale image.

I template e le immagini possono essere rappresentate anche come dei vettori

16

Template matching

Diversa notazione - stesso significato

Correlation based: Data un'immagine I e un'istanza Ti, una misura intuitiva di diversità tra I e Ti è la Sum of Squared Difference (SSD):

$$SSD(\mathbf{I}_{xy}, \mathbf{T}_i) = \left\|\mathbf{I}_{xy} - \mathbf{T}_i\right\|^2 = \left(\mathbf{I}_{xy} - \mathbf{T}_i\right)^T\left(\mathbf{I}_{xy} - \mathbf{T}_i\right) = \left\|\mathbf{I}_{xy}\right\|^2 + \left\|\mathbf{T}_i\right\|^2 - 2\mathbf{T}_i^T\mathbf{I}_{xy}$$

Correlation based: quando i termini $||\mathbf{I}||^2$ e $||\mathbf{T}_i||^2$ sono costanti, minimizzare il primo termine corrispondere a massimizzare la Cross Correlation (CC) tra **I** e $\mathbf{T}_i$:

$$CC(\mathbf{I}_{xy}, \mathbf{T}_i) = \mathbf{T}_i^T\,\mathbf{I}_{xy} = \sum_k \mathbf{T}_i[k]\cdot\mathbf{I}_{xy}[k]$$

17

Template matching

Misure di correlazione normalizzate sono necessarie quando I e Ti non sono costanti

Normalized Sum of Squared Difference (NSSD): la normalizzazione rende NSSD indipendente dal contrasto (range dinamico di valori) di immagine e template.

$$NSSD(\mathbf{I}_{xy}, \mathbf{T}_i) = \frac{\left\|\mathbf{I}_{xy} - \mathbf{T}_i\right\|^2}{\left\|\mathbf{I}_{xy}\right\|\cdot\left\|\mathbf{T}_i\right\|}$$

Normalized Cross-Correlation (NCC): Simile a NSSD ma computazionalmente meno costosa

$$NCC(\mathbf{I}_{xy}, \mathbf{T}_i) = \frac{\mathbf{I}_{xy}^T\mathbf{T}_i}{\left\|\mathbf{I}_{xy}\right\|\cdot\left\|\mathbf{T}_i\right\|}$$

18

Template matching

- **Zero mean Normalized Sum of Squared Differences** (***ZNSDD***) e **Zero mean Normalized Cross-Correlation** (***ZNCC***) rispetto a *NSSD* e *NCC* sono invarianti anche per pattern che, a parità di contrasto (stesso range dinamico), presentano *luminosità medie* diverse

$$ZNSSD(\mathbf{I},\mathbf{T}_i)=\frac{\left\|(\mathbf{I}-\bar{\mathbf{I}})-(\mathbf{T}_i-\bar{\mathbf{T}}_i)\right\|^2}{\left\|(\mathbf{I}-\bar{\mathbf{I}})\right\|\cdot\left\|(\mathbf{T}_i-\bar{\mathbf{T}}_i)\right\|} \qquad ZNCC(\mathbf{I},\mathbf{T}_i)=\frac{(\mathbf{I}-\bar{\mathbf{I}})^T(\mathbf{T}_i-\bar{\mathbf{T}}_i)}{\left\|(\mathbf{I}-\bar{\mathbf{I}})\right\|\cdot\left\|(\mathbf{T}_i-\bar{\mathbf{T}}_i)\right\|}$$

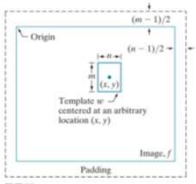
19

Template matching

$$(w\star f)(x,y)=\sum_s\sum_t w(s,t)f(x+s,y+t)$$

$$\gamma(x,y)=\frac{\sum_s\sum_t[w(s,t)-\bar{w}][f(x+s,y+t)-\bar{f}_{xy}]}{\left\{\sum_s\sum_t[w(s,t)-\bar{w}]^2\sum_s\sum_t[f(x+s,y+t)-\bar{f}_{xy}]^2\right\}^{\frac{1}{2}}}$$

Correlation, usando la notazione del Gongalez



20

Template matching

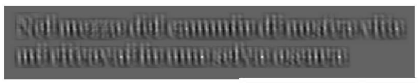
Nel mezzo del cammin di nostra vita
mi ritrovai in una selva oscura

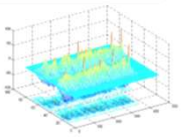
Bisogna riconoscere le istanze della lettera 'a', di cui si ha a disposizione un template

21

Template matching

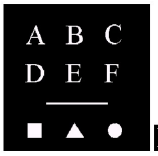
Nel mezzo del cammin di nostra vita
mi ritrovai in una selva oscura

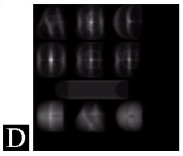




22

Template matching



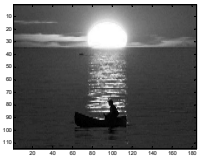


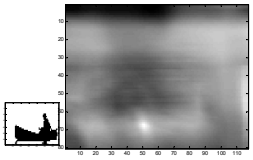
a b c
FIGURE 12.9 (a) Image. (b) Subimage. (c) Correlation coefficient of (a) and (b). Note that the highest (brighter) point in (c) occurs when subimage (b) is coincident with the letter "D" in (a).

1. Come faccio a decidere se la lettera D è presente nell'immagine ?
2. Come faccio a definire dove è presente la lettera D ?

23

Template matching



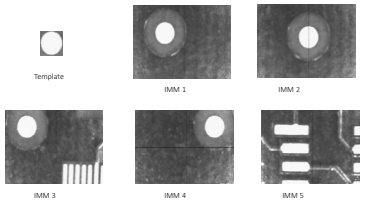


1. Come faccio a decidere se la barca e' presente nell'immagine ?
2. Come faccio a definire dove e' presente la barca ?

24

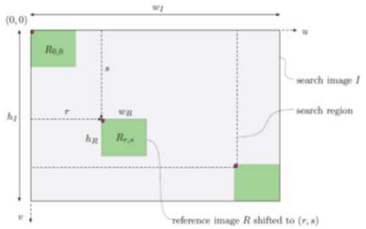
Template matching

Esempio



25

Template matching

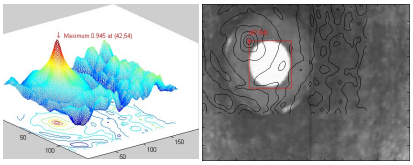


Per ogni pixel dell'immagine da ispezionare valuto la correzione fra il template la porzione di immagine a cui si sovrappone. Ottengo quindi una nuova immagine.

26

Template matching

IMM 1



Mappa di correlazione

Immagine originale, Rettangolo trovato, e mappa di correlazione

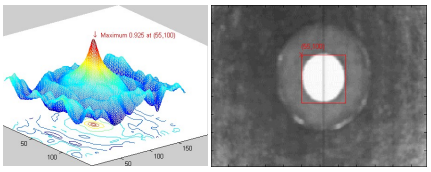
Template

Per indicare la posizione, sovrappongo la bounding box dell'oggetto cercato all'immagine, bentrata sul pixel a massima correlazione

27

Template matching

IMM 2



Mappa di correlazione

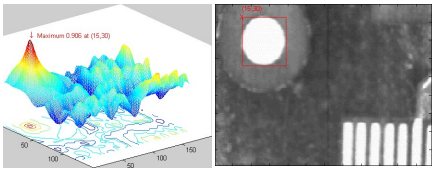
Immagine originale, Rettangolo trovato.

Template

28

Template matching

IMM 3



Mappa di correlazione

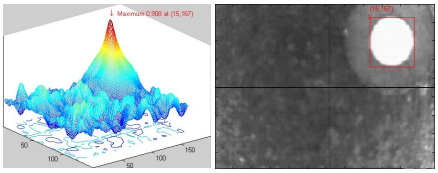
Immagine originale, Rettangolo trovato.

Template

29

Template matching

IMM 4



Mappa di correlazione

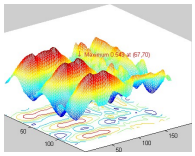
Immagine originale, Rettangolo trovato.

Template

30

Template matching

IMM 5



Mappa di correlazione

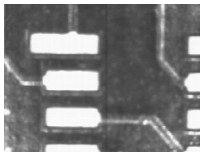


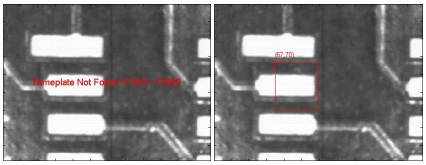
Immagine originale.

Template

31

Template matching

IMM 5



Soglia impostata a 0.800

Soglia impostata a 0.200

Il probelma e' quindi come faccio a trovare il giusto valore di soglia?

32

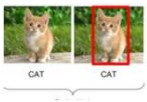
Insiemi di addestramento e valutazione

Come faccio a trovare il giusto valore di soglia?

Si parte sempre da una Raccolta dei Dati

Dati che sono usati per training (per allenare il classificatore / riconoscitore) o di test (per testare il classificatore / riconoscitore)

- Non basta che avere le immagini di training e test, le immagini devono essere annotate, label oggetto (gatto) e posizione (bounding box)



E dalla definizione di una funzione oggettiva di errore.

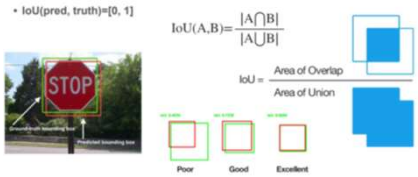
Quando posso dire che ho effettivamente trovato l'oggetto?

33

Funzione di valutazione

- Il coefficiente di Jaccard misura la similarità tra insiemi campionari, ed è definito come la dimensione dell'intersezione divisa per la dimensione dell'unione degli insiemi campionari:

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|}.$$



34

Insiemi di addestramento e valutazione

Il riconoscimento (si/no) e' un problema di classificazione binario, dove le due classi corrispondono a esempi Positivi (vera classe) e Negativi (classe resto del mondo).

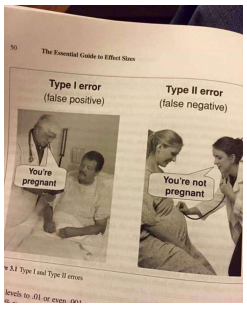
L'apprendimento di un riconoscitore/classificatore supervisionato è tipicamente effettuato a partire da un insieme di addestramento (training set)

l'insieme di esempi utilizzati per valutare le prestazioni di un sistema prende il nome di test set

Training set e test set sono disgiunti

35

Valutazione dei risultati



36

Valutazione dei risultati

In un sistema di ***riconoscimento/localizzazione di oggetti***, abbiamo due tipi di errori:

- **False *riconoscimento/localizzazione*** : percentuale di casi in cui il sistema localizza un oggetto non corretto (questi errori sono detti anche false).
- **Mancate riconoscimento**: percentuale di casi in cui il sistema non riconosce / localizza nessun oggetto, sebbene l'oggetto sia presente (questi errori sono detti anche drop o miss).

37

Valutazione dei risultati

- se il risultato della predizione è positivo *p* e il valore vero è anche positivo *p*, viene chiamato *vero positivo* (true positive - TP);
- se invece il valore vero è negativo, il risultato viene chiamato *falso positivo* (false positive - FP);
- al contrario, si ha un *vero negativo* (true negative - TN) quando entrambi, il risultato e il valore vero, sono negativi;
- un *falso negativo* (false negative - FN) invece si ha quando il risultato è negativo e il valore vero è positivo

38

Valutazione dei risultati

Dato un insieme di immagini, possiamo definire

- *True Positive Rate* (TPR, frazione di veri positivi) e *False Positive Rate* (FPR, frazione di falsi positivi).

$$\bullet TPR = TP/P = TP/(TP + FN)$$
$$\bullet FPR = FP/N = FP/(FP + TN)$$
$$\bullet \text{accuratezza } ACC = (TP + TN)/(P + N)$$

Se fosse un sistema di riconoscimento volti. Criteri di valutazione:

- Frazione di *clienti* (che dichiarano l'identità corretta) che vengono respinti dal sistema (*False Rejection Rate FRR*)
- Frazione di *impostori* che vengono accettati dal sistema (*False Acceptance Rate FAR*)

39

Valutazione dei risultati

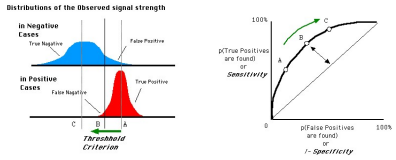
- False e mancate localizzazioni sono spesso ***legate tra loro*** ed entrambe funzione di alcuni parametri di tolleranza del sistema:
  - se si rende il sistema meno tollerante ai falsi in modo da evitare che vengano localizzati oggetti non esistenti, aumenta la probabilità di perdere anche qualche oggetto genuino.
  - Viceversa, se si rende il sistema più tollerante in modo da localizzare tutti gli oggetti presenti (anche quelli "difficili") allora aumenta la probabilità di localizzare anche qualche falso oggetto.

I sistemi possono in genere essere regolati per operare a diversi livelli di tolleranza.

40

Valutazione dei risultati

- Il ROC è anche noto come curva Receiver Operating Characteristic, poiché è un confronto tra due caratteristiche operative (TPR e FPR) al cambiare del criterio.
- Nel nostro caso, la curva ROC viene creata tracciando il valore del True Positive Rate (TPR, frazione di veri positivi) rispetto al False Positive Rate (FPR, frazione di falsi positivi) a varie impostazioni di soglia.



Variando la soglia si può abbassare un indice facendo crescere l'altro

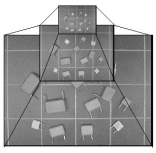
41

Template matching

- Il numero di operazioni richieste cresce linearmente con il numero di istanze e con il numero di pixel di I e T (e quindi quadraticamente rispetto al lato di I e di T).
- In pratica, per alcune applicazioni real-time, l'approccio di base è raramente applicabile.

Per ridurre il costo computazionale, gestire cambiamenti di scala, orientamento, prospettiva, piccole variazioni negli oggetti, rumore, … si può

1. adottare un approccio multi-risoluzione; la strategia di matching diventa però molto più complessa.
2. Aumentare il livello di astrazione, ovvero… (vedi slide successive)



42

Multi-scale template matching

Immagine 512 × 256

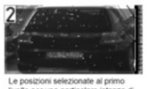
Template 10 digit e 26 caratteri 

Per ogni digit e per ogni carattere consideriamo 3 i stanze dovute a variazioni di scala (diverse distanze dalla telecamera). La scala intermedia ha risoluzione 14 × 20 pixel.
Ogni istanza (108 istanze = (10 + 26) × 3) deve essere sovrapposta all'immagine in tutte le possibili posizioni e genera quindi ulteriori 512×256 istanze (se si trascurano i bordi).
Pertanto occorrere stimare circa 14.155.776 = 108 × 512 × 256 correlazioni ciascuna richiedente almeno 14×20 moltiplicazioni e altrettante somme (nel caso di semplice *CC*).
In totale circa 4×10^9 moltiplicazioni (intere) e altrettante somme.
Quanto tempo occorre per processare un'immagine ? 40 secondi su una macchina capace di eseguire 200×106 operazioni intere al sec.

43

Multi-scale template matching

- Si segue la ricerca su una gerarchia crescente di risoluzioni
- Viene creata una "piramide" di risoluzioni sia per I che per T (ad esempio dimezzando la risoluzione ad ogni livello)
- La ricerca viene eseguita inizialmente sulla risoluzione più bassa, e ai livelli successivi vengono analizzate solo le istanze promettenti (la cui correlazione al livello inferiore eccedeva una data soglia)
- Consente di eseguire una "scrematura" ai livelli iniziali e di perfezionare la localizzazione e filtrare "false somiglianze" ai livelli successivi, per esempio:
  - A metà risoluzione le operazioni si riducono di 16 volte.
  - A 1/4 quarto di risoluzione di 256 volte.
  - A $1/2^n$ di risoluzione di 2^{4n} volte, ma tipicamente oltre a 3, 4 livelli non è possibile operare per mancanza di dettagli

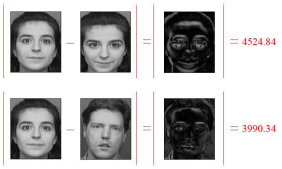


Le posizioni selezionate al primo livello per una particolare istanza di un template

44

Template matching

- Perché non sempre si può usare:
  - Traslazione, rotazione, scala e prospettiva.
  - Deformazione e variabilità dei pattern.
  - Occlusioni
  - Cambiamenti di illuminazione.
  - Rumore e diverse tecniche di acquisizione.



45

Matching / Riconoscimento

Schema algoritmi di riconoscimento basato su features

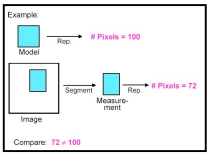
1) rappresentazione/descrizione del modello

2) identificazione di possibili candidati (ROI) nell'immagine (e.g. mediante segmentazione o sliding windows)

3) rappresentazione/descrizione dei candidati (ad esempio colore, texture, numero di pixel ad alto gradiente…)

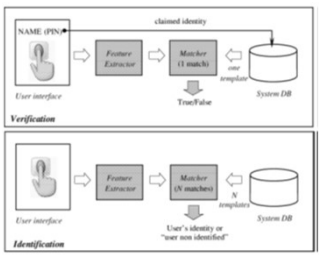
4) confronto delle rappresentazioni/descrizioni dei modelli e dei candidati (valori di soglia trovati su un opportuno training set)

5) decisione



46

Riconoscimento vs. classificazione



Riconoscimento o verifica

Classificazione o identificazione

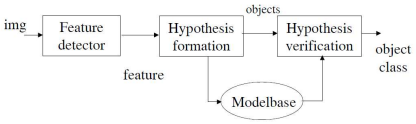
47

Matching / Riconoscimento vs classificazione

Riconoscimento : constatare se una regione candidata corrisponde o no ad un modello di riferimento.

Classificazione : determinare a quale di N classi di prestabilite è da attribuirsi la regione in esame.

Si può definire un classificatore usando N riconoscitori ed una ulteriore regola di decisione a posteriori.



48

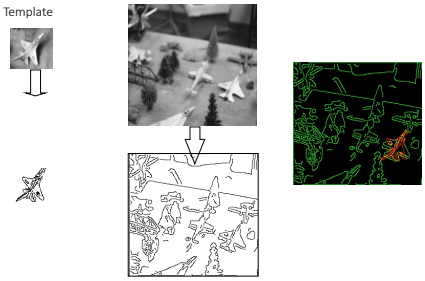
Riconoscimento immagini

Raimondo Schettini
DISCo - Universita' di Milano Bicocca
Raimondo.schettini@unimib.it

49

Matching fra contorni

Template

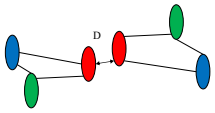


50

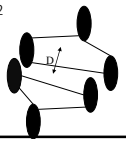
Matching fra contorni

Oss. Il concetto classico di distanza non tiene conto della forma degli oggetti

Esempio 1



Esempio 2



51

Hausdorff Distance

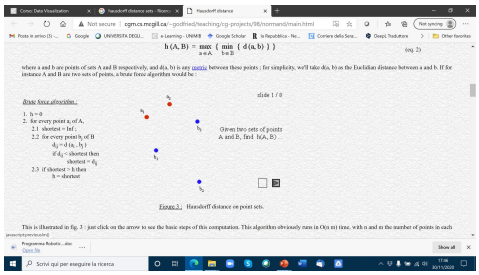
$$h(A,B)=\max_{a \text{ appartiene ad } A} (\min_{b \text{ appartiene } B} (\text{Dist } (a,b))$$

- La distanza di Hausdorff è la distanza massima di un insieme di punti rispetto ad un altro insieme.
- La distanza di Hausdorff è orientata ovvero h(A,B)≠h(B,A).
- Quindi, siano h(A,B) è la distanza da A a B (*forward*) E h(B,A) distanza da B ad A (*backward*), la distanza di Hausdorff generalizzata è definita come:

$$H(A,B)=\max (h(A,B), h(B,A))$$

52

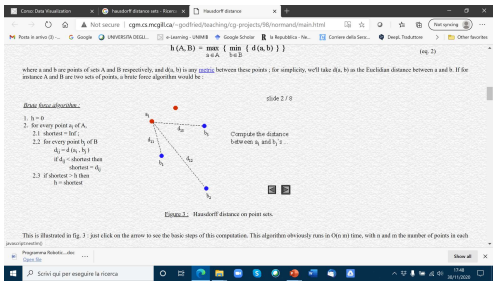
Hausdorff Distance



Hausdorff distance (mcgill.ca)

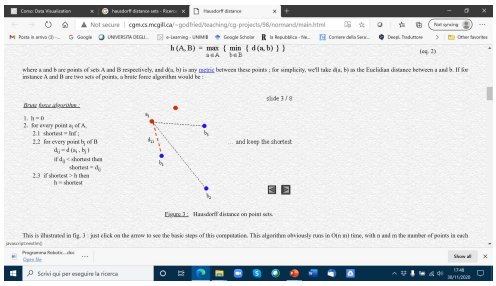
53

Hausdorff Distance



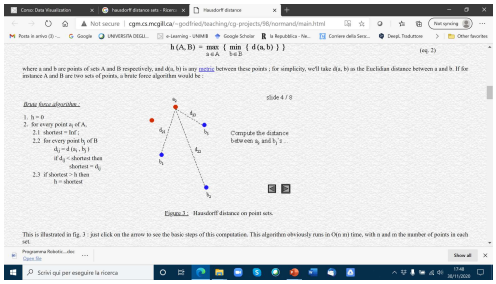
54

Hausdorff Distance



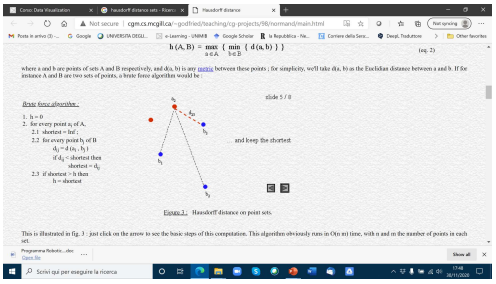
55

Hausdorff Distance



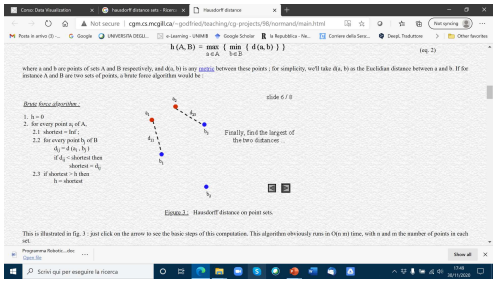
56

Hausdorff Distance



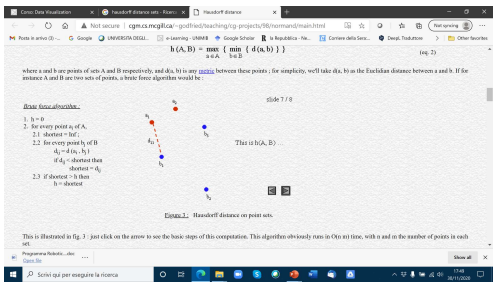
57

Hausdorff Distance



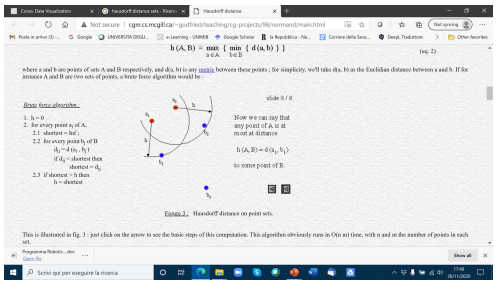
58

Hausdorff Distance



59

Hausdorff Distance



60

Matching fra contorni

- Se un punto X è alla distanza D da un poligono P intendiamo dire che X è a distanza D dal punto più vicino di P.
- Lo stesso dicasi per due poligoni. Se A e B sono due poligoni, la distanza minima è quella più corta tra tutti i punti di A e quelli di B.

- Quindi se D è la funzione distanza, Per ogni punto a della curva A trova il punto della curva di B a distanza minima. Quindi trova tra tutti i punti di A quello che ha minima distanza

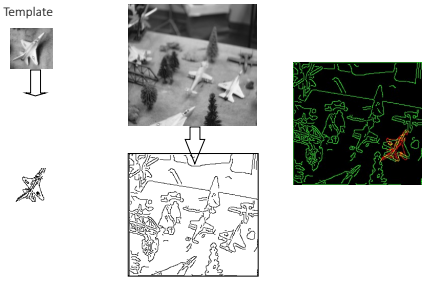
$$D(A,B)=\min_{\text{a appartiene ad A}} (\min_{\text{b appartiene B}} (\text{Dist}(a,b))$$

Si può considerare "*Dist*(·, ·)" come la distanza Euclidea

61

Matching fra contorni

Template



62