Dalla Computer Vision alle interfacce naturali

METODI E MODELLI DI INTERAZIONE

Social Q&A

@vs_AR
#askVisionary
www.vision-ary.net

Obiettivi del seminario

1. Fornire gli elementi di base per comprendere cosa succede «dietro» un sistema interattivo
2. Rudimenti di Computer Vision partendo da esempi concreti
3. Panoramica su casi reali in ambiti diversi: dalla video-sorveglianza all'arte digitale interattiva
4. Fornire i rudimenti di OpenCV per un sistema di localizzazione di oggetti

NON-obiettivi del seminario

1. Non verranno forniti rudimenti di algebra
2. Non verranno forniti rudimenti di statistica e probabilità
3. Non verranno forniti rudimenti di programmazione
4. Non verranno forniti rudimenti di computer graphics
5. Se possibile, eviteremo di citare Minority Report.

Interazione Naturale vs. Computer Vision

1. Con il volto | Face Tracking
2. Con il corpo | Body Tracking
3. Con le mani | Hand Tracking
4. Con gli occhi | Gaze Tracking
5. Con la voce (non è Computer Vision ma le tecniche sottostanti sono spesso le medesime)

Realtà aumentata

Obiettivo: «aumentare» la realtà apponendo un super-strato di informazione digitale che fornisce una esperienza interattiva personalizzata.

Situazione di riferimento: videocamera posizionata in un punto di interesse, ad esempio nello schermo di un laptop nel caso di una webcam.

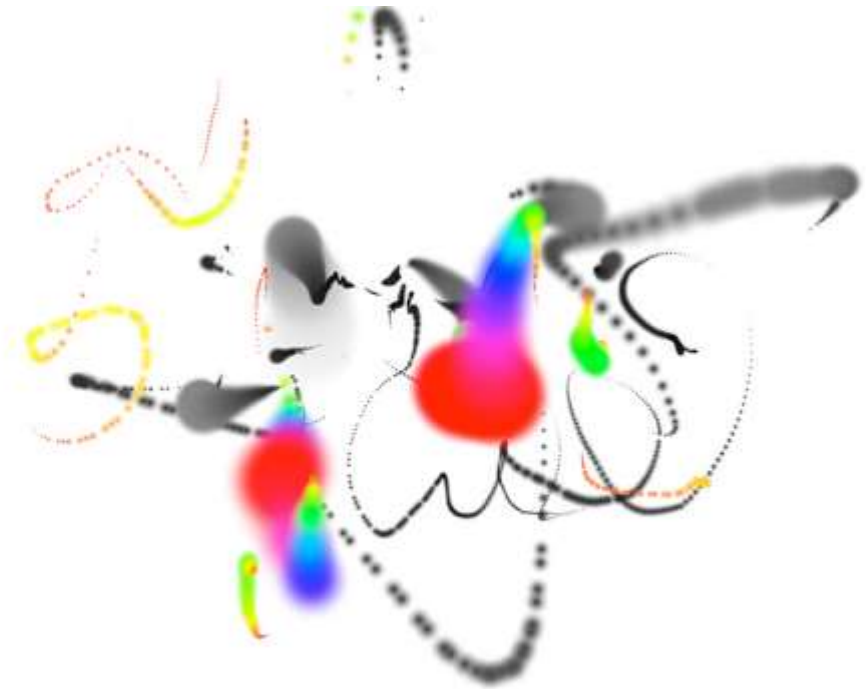
Risultato: l'interazione tra l'utente e il dispositivo di acquisizione produce una sovrapposizione digitale tra le informazioni e il protagonista, creando una esperienza interattiva e personale.

Ambiti di applicazione: Gaming (Kinect, Oculus, ecc..) | Marketing & Advertising (occhiali da sole interattivi, digital signage interattivo, promozione di prodotti) | Medicale (operazioni «aumentate») | Elettronica di consumo (webcam con applicativi ludici) | Arte interattiva

Realtà aumentata col volto

Realtà aumentata con le mani

Arte Interattiva con le mani



@PaoloScoppola

www.paoloscoppola.com

Arte Interattiva con il corpo

Arte Interattiva con il cervello

Detection and Tracking

Obiettivo: localizzare la presenza del volto come indicatore di presenza umana, «inseguirlo» nel tempo mantenendo una coerenza spazio-temporale tra i frame elaborati. Generalmente viene apposta una «bouding box» sul viso localizzato.

Situazione di riferimento: videocamera posizionata in un punto di interesse, ad esempio nello schermo di un laptop nel caso di una webcam.

Risultato: Il sistema di tracciamento (face tracking) è il risultato dell'integrazione di diversi moduli che devono lavorare in tempo reale nonostante la complessità dell'analisi in atto.

Ambiti di applicazione: Video-Sorveglianza (face recognition) | Gaming (Kinect) | Marketing & Advertising (occhiali da sole interattivi, analisi clientela, analisi pubblicità, digital signage) | Medicale (gaze tracking) | Elettronica di consumo (cellulari, macchine fotografiche) | Arte interattiva

Face detection and tracking: perché è interessante?

1. La localizzazione del volto è un chiaro esempio di «back-end» di una interfaccia interattiva moderna.

2. Il volto è un indicatore certo della prossimità umana:
   - stimatore di attenzione visiva;
   - è un modulo di controllo dell'interfaccia molto potente.

3. L'addestramento di un localizzatore di volti è rappresentativo del problema dell'addestramento (seconda parte del seminario).

Object Tracking: possibili approcci

1. Tracking Frame by Frame: il tracking è emulato. Si applica continuamente (senza coerenza spazio-temporale) un localizzatore di volti e si aggregano i risultati nel tempo.

2. Tracking by likelihood: tramite una misura di verosimiglianza si rafforzano le ipotesi di tracciamento più consistenti per mantenere una traccia coerente nello spazio-tempo.

3. Tracking by detection: la misura di verosimiglianza è data dallo stesso localizzatore di volti utilizzato in maniera «furba».

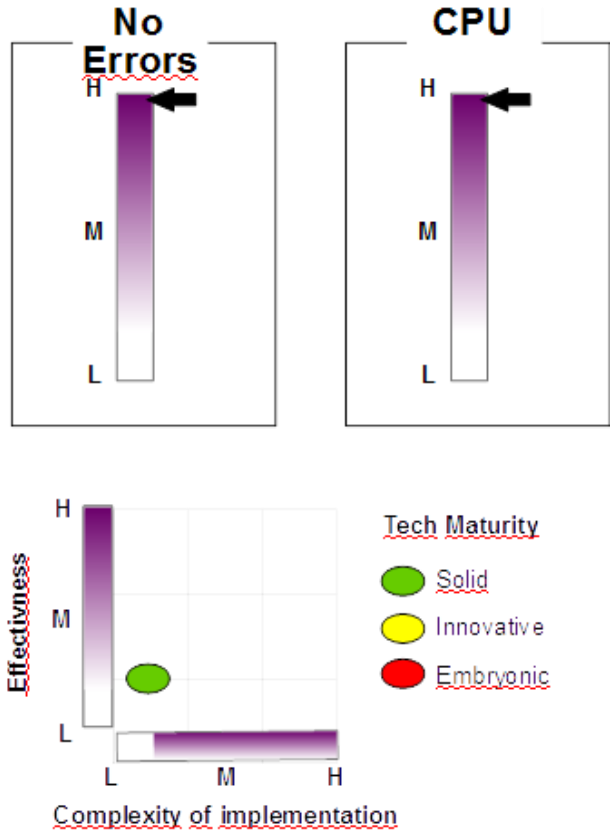
Queste definizioni non sono formalmente corrette ma rendono l'idea sui possibili approcci che si possono utilizzare in computer vision per tracciare un oggetto.

Tracking frame by frame

- Questo approccio si basa su una ricerca frame a frame dei volti presenti nelle immagini che compongono la sequenza video non tenendo conto della correlazione spazio-temporale che è naturalmente insita in una sequenza di immagini.

- Si noti inoltre che tale approccio non permette di realizzare un tracciamento in senso stretto poiché, isolando l'analisi all'interno di ogni singolo frame senza inferire la posizione dell'oggetto di interesse al frame successivo, viene meno la capacità di analizzare la coerenza spazio-temporale del movimento dell'oggetto.

- Le tecniche basate su questo approccio spesso presentano un output simile a quello di un sistema di tracking, poiché i rilevamenti effettuati frame a frame «emulano» il comportamento di un sistema di tracciamento, ma di fatto non lo realizzano realmente. E' necessario aggiungere un ulteriore livello di intelligenza.

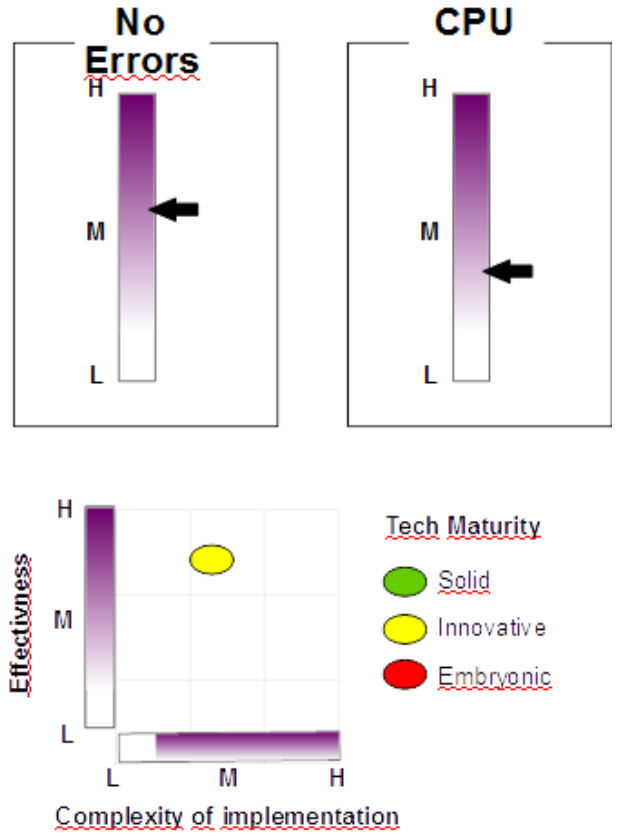


Tracking by likelihood

- Questo approccio si basa su una ricerca dell'oggetto nel frame solo al «tempo zero». Dal frame successivo si utilizza un meccanismo di «propagazione» della posizione dell'oggetto. Nel caso di un framework di stima bayesiana la conferma della previsione avviene tramite verosimiglianza dell'osservazione.

- A differenza del precedente approccio (che utilizza solo informazione all'interno di un singolo frame), questa modalità sfrutta la correlazione della sequenza di immagini introducendo il concetto "logico'' di traccia, come evoluzione della posizione di un oggetto nello spazio e nel tempo. La correlazione spazio-temporale è sfruttata sia per mantenere una logica di tracciamento che per minimizzare il costo computazionale .

- Si noti che:
  - se la misura di verosimiglianza è «smart» si ha flessibilità nel tracking.
  - Generalmente si ha indipendenza dal tipo di oggetto tracciato.
  - Spesso è difficile trovare una misura «leggera».



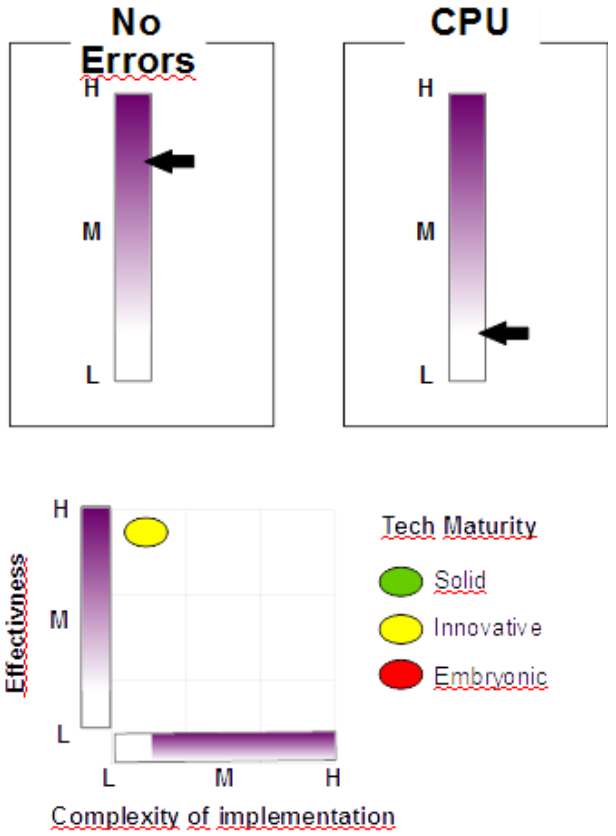
Tracking by likelihood: the TLD

Tracking by detection

- E’ un caso particolare del tracking by likelihood.

- Questo approccio si basa su una ricerca dell’oggetto nel frame solo al «tempo zero». Dal frame successivo si utilizza un meccanismo di «propagazione» della posizione dell’oggetto. La verifica della verosimiglianza avviene tramite lo stesso meccanismo di inizializzazione del tracking, ovvero il face detector.

- Si noti che:
  - Il tracking è meno flessibile poiché solo oggetti «riconoscibili» dal detector possono essere anche tracciati (NO posa e rotazione in piano).
  - La verosimiglianza non richiede altre implementazioni addizionali, va solo «estratta» dal sistema di detection.
  - Totale dipendenza dal tipo di oggetto tracciato.
  - Efficienza esasperata, moltiplicatore 50-100x.



Object Tracking: da dove si parte?

Indipendentemente dall'approccio utilizzato è necessario localizzare il volto almeno al «tempo zero» per l'inizializzazione del sistema. L'inizializzazione può avvenire:

- Manualmente (app per cellulari, siti internet, ecc...). Viene chiesto all'utente, ad esempio tramite il mouse, di inserire una bounding box contenente la faccia o la posizione di occhi e bocca.

- Automaticamente, attraverso un sistema di localizzazione volti vero e proprio che in maniera trasparente all'utente localizza il volto e inizializza il sistema di tracking.

Il primo caso è ovviamente di scarso interesse. Per quanto riguarda il secondo, lo stato dell'arte è rappresentato dall'algoritmo di Viola/Jones.

Il problema dell'addestramento

Quando si affronta un problema di apprendimento sono tre gli elementi in gioco:

1. Dataset (Train Set vs. Test Set)
2. Feature (Haar, LBP, Hog, SIFT, SURF, ecc…)
3. Algoritmo di apprendimento (AdaBoost, SVM, ecc…)

Questi tre elementi ci sono sempre, ma quali dataset, features e algoritmi di apprendimento usare dipende sempre dal problema! Si noti che lo stesso problema può essere affrontato con tecniche diverse.

Localizzatore di volti: Viola & Jones

L'algoritmo di Viola e Jones (2001) è lo standard de facto per la localizzazione del volto. Rispetto ai suoi (falliti) predecessori introduce tre elementi di novità:

1. Utilizzo 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.

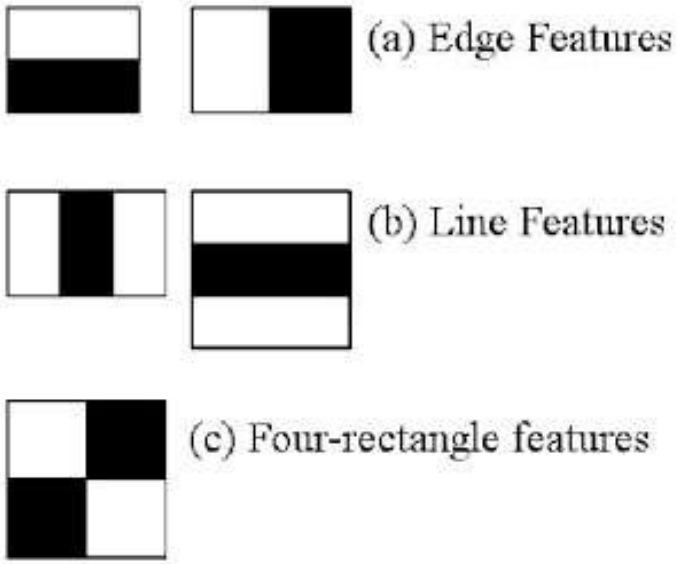
2. Viene introdotto un metodo di selezione di feature di Haar attraverso l'algoritmo **AdaBoost** di Freud e Shapire (1995). Questa strategia permette di eliminare in addestramento la maggior parte delle feature di scarsa capacità discriminante e selezionare solo quelle più efficaci per il problema.

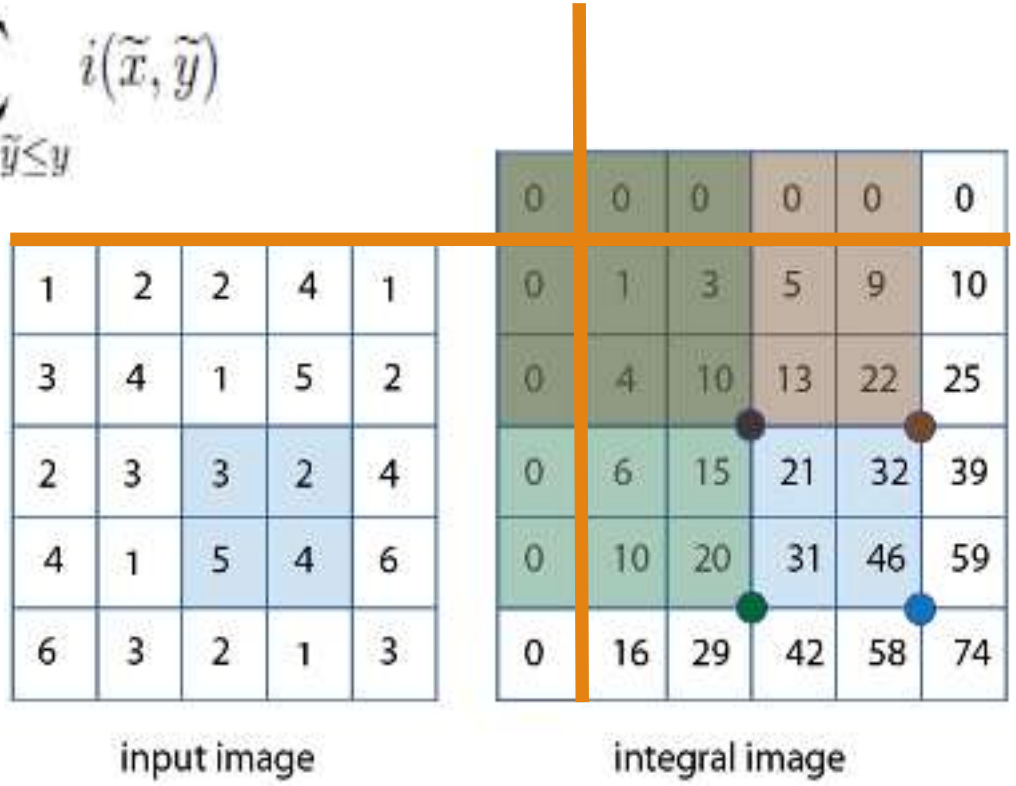
3. Viene introdotta una nuova strategia di analisi dell'immagine basata su **struttura a cascata** dove ogni livello della cascata è un classificatore creato con AdaBoost. La complessità dei livelli cresce man mano che si procede verso la fine della cascata. Le regioni «facili» vengono scartate velocemente ai primi stadi, quelle più «difficili» sono sottoposte a più livelli di verifica. Qualora una regione superi tutti gli stadi viene etichettata come regione contenente una faccia.

Haar features e Integral Image

L'immagine integrale, che ha le stesse dimensioni dell'immagine originale, nella posizione di indice (x,y) contiene la somma dei livelli di grigio di tutti i pixel superiori e a sinistra del punto.

$$ii(x,y) = \sum_{\tilde{x}\leq x\ ,\ \tilde{y}\leq y} i(\tilde{x},\tilde{y})$$





Localizzatore di volti in azione

La fase di training: AdaBoost

- All'interno di una finestra di ricerca di 24 × 24 pixel si possono definire 50.000/100.000 Haar features a fronte di solo 576 pixel. Sebbene la singola feature sia efficiente, non è pensabile applicare tutto l'intero di set in maniera esaustiva (su campo) poiché richiederebbe tempi di calcolo proibitivi.

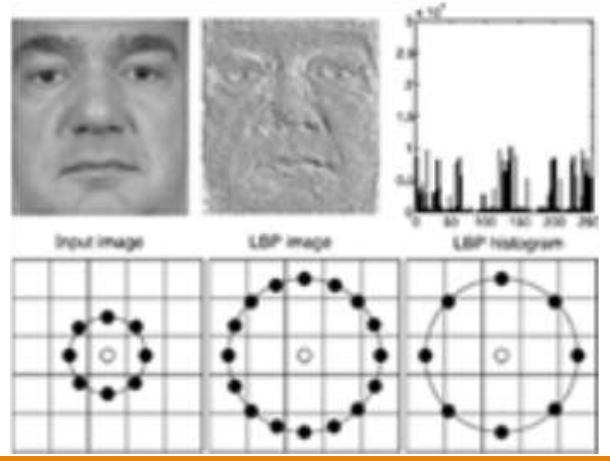
- Viola e Jones hanno dimostrato che è sufficiente applicare un numero limitato di feature discriminanti per localizzare il volto. La chiave dell'algoritmo è individuare le feature più discriminanti e combinarle opportunamente in un classificatore.

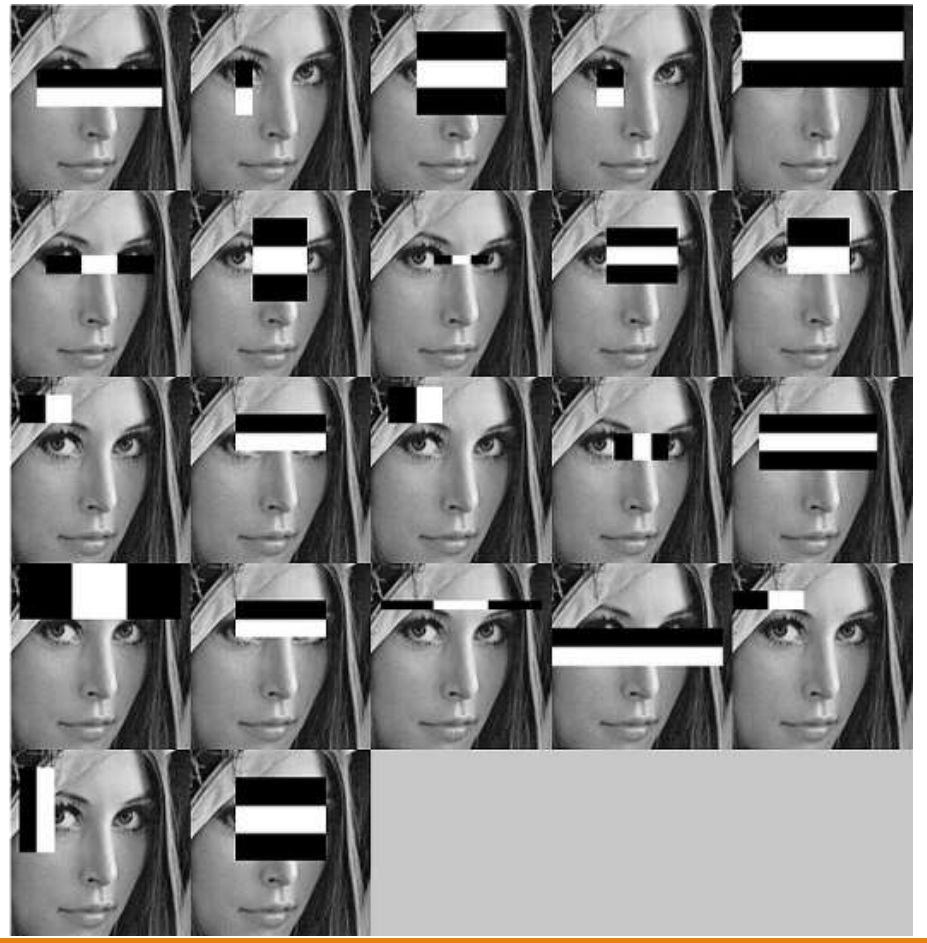
- L'algoritmo di Boosting addestra il classificatore presentandogli, più volte e secondo uno schema iterativo, una sequenza di esempi etichettati (faccia o non faccia). Ad ogni iterazione viene modificata l'importanza di ciascun esempio in modo da enfatizzare gli esempi “difficili” (esempi classificati erroneamente) così da migliorare le performance del classificatore nella successiva iterazione. Al termine delle iterazioni, il classificatore è una combinazione lineare pesata delle feature più discriminanti.

L’importanza del dataset

L’importanza del dataset

L’importanza delle Feature





Art vs Feature: CVDazzle by Adam Harvey

@adamhrv

www.cvdazzle.com

Tracking di oggetti: stima probabilistica

Obiettivo: stimare lo stato (non solo la posizione) di un oggetto che evolve nel tempo attraverso una serie di misure che si effettuano su di esso. L'oggetto tracciato viene modellato attraverso un insieme di caratteristiche che lo descrivono (bordi, colore, punti di controllo, ecc…) .

E’ necessario definire almeno due modelli:

1. **Modello evolutivo** (o dinamica dell'oggetto), descrive l'evoluzione dell'oggetto nel tempo.

2. **Modello di osservazione** determina come valutare lo stato dell'oggetto tracciato nel tempo.

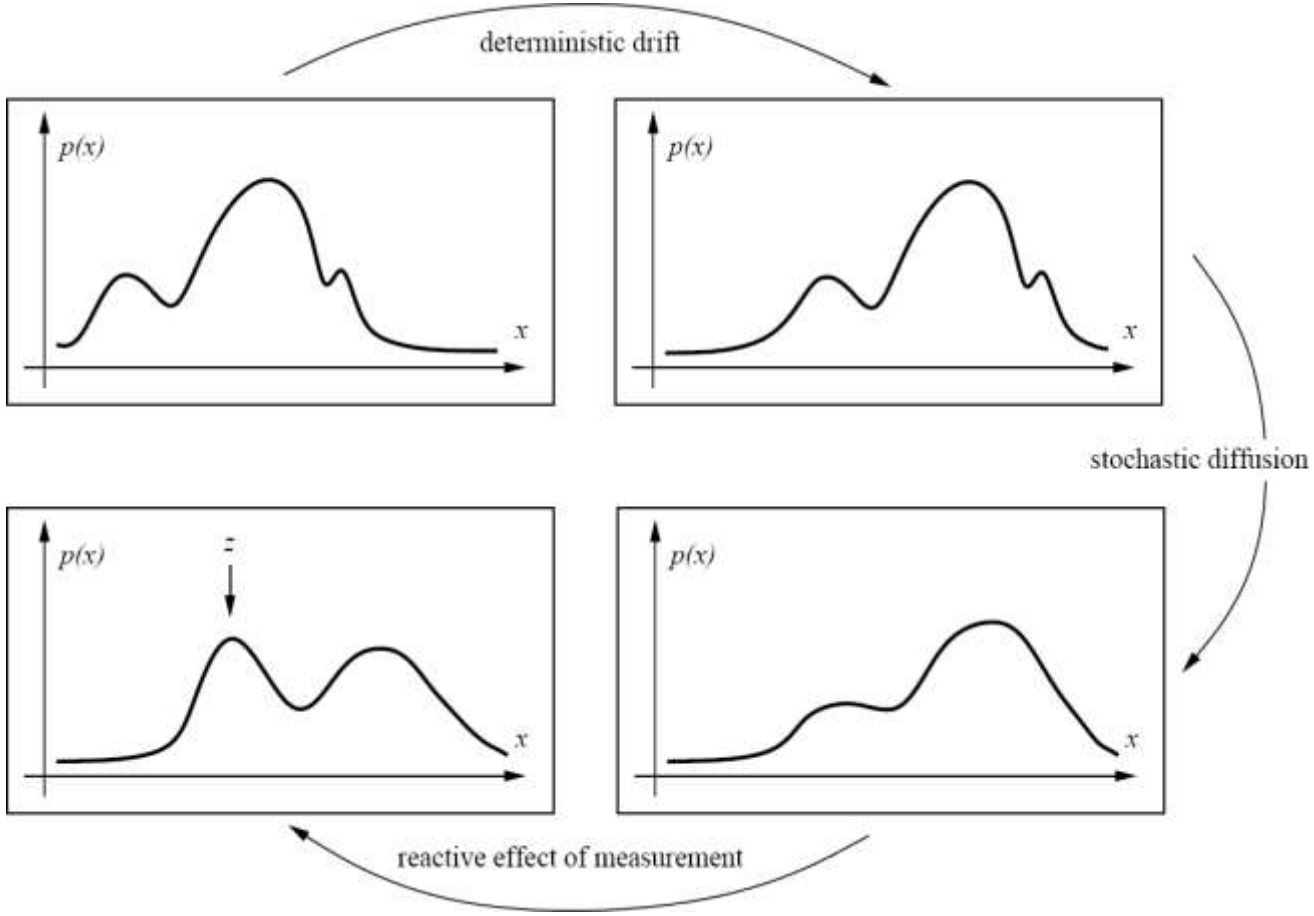
Questi due modelli vengono definiti secondo un approccio di tipo probabilistico e pertanto si prestano in maniera diretta ad essere inclusi in un procedimento ricorsivo di tipo bayesiano, che fornisce un paradigma di stima dell'evoluzione temporale di un oggetto. Nei fatti si vuole stimare la distribuzione a posteriori dell'oggetto a partire dal modello evolutivo e da una serie di osservazioni effettuate su di esso.

Tracking di oggetti: predizione e aggiornamento

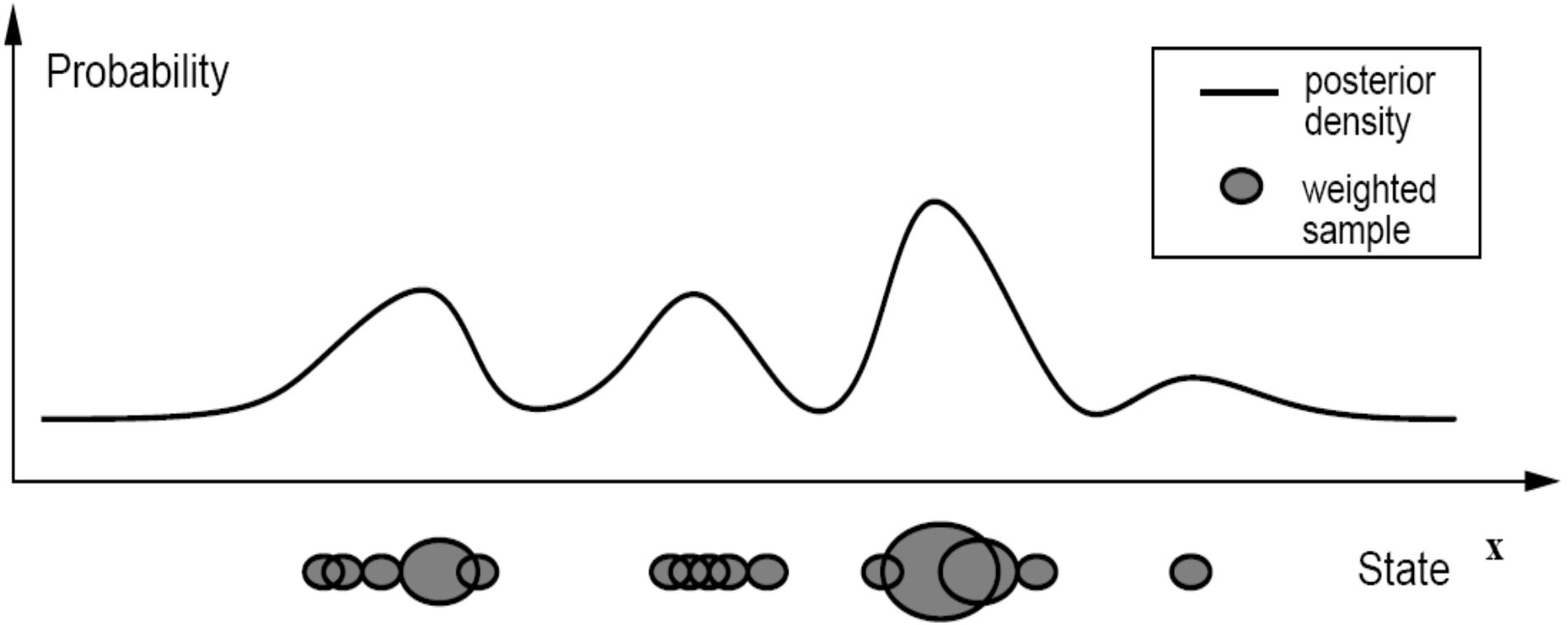
La ricorsione bayesiana si articola in due passi di elaborazione ciclica:

1. **Passo di predizione**: utilizzando il modello evolutivo, inferisce la configurazione che il nuovo stato dell'oggetto assumerà al tempo successivo “deformando'', secondo delle regole che introdurremo a breve, la densità dell'oggetto calcolata al tempo precedente;
2. **Passo di aggiornamento**: attraverso il modello di osservazione, corregge la previsione effettuata nel precedente passo rendendola conforme alla situazione osservata correntemente. Questo risultato è ottenuto tramite il teorema di Bayes che fornisce un efficiente meccanismo di aggiornamento della conoscenza alla luce di una nuova osservazione non appena essa si rende disponibile.

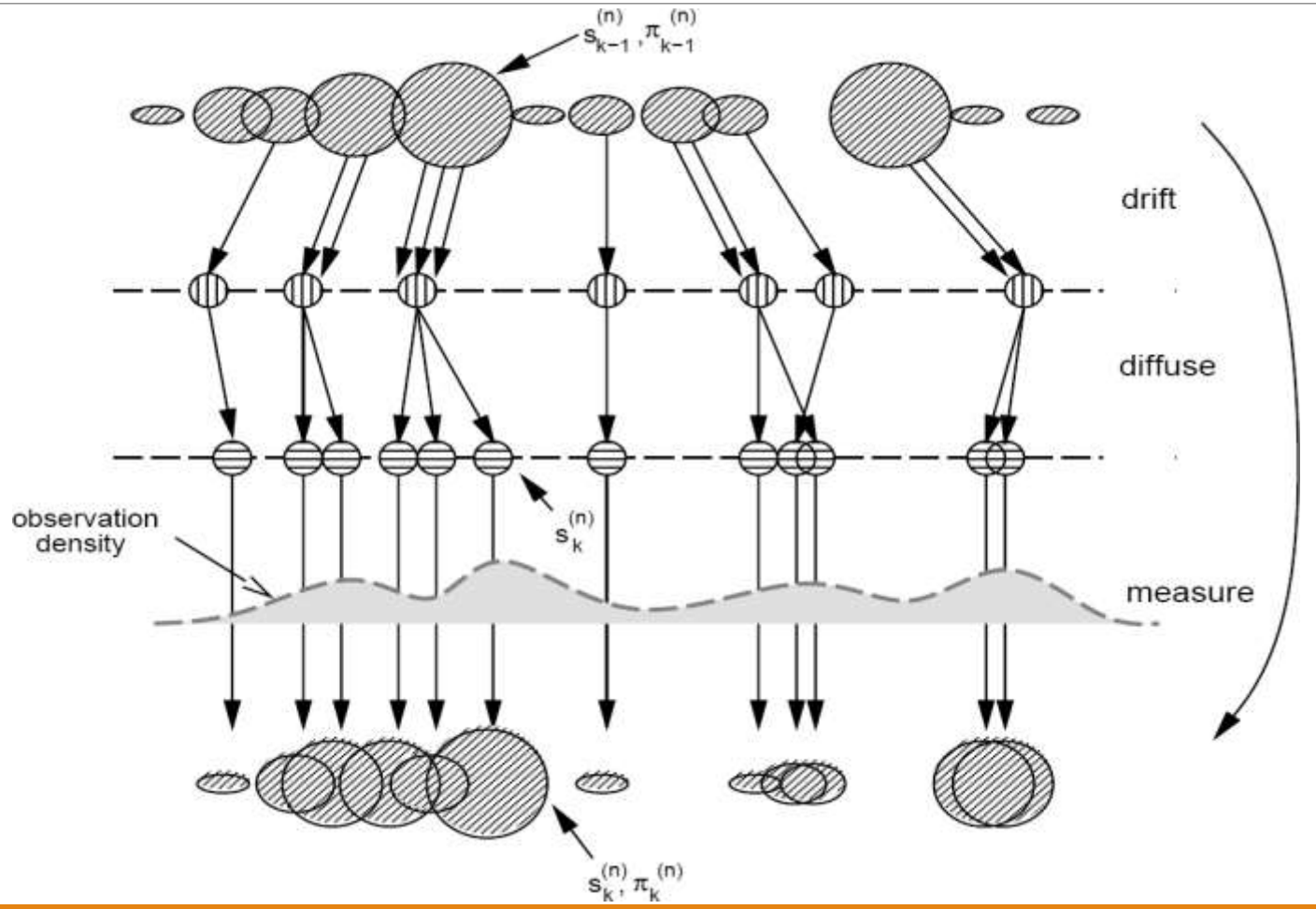
Il modello evolutivo



Modello a particelle: particle filtering



Tracking by CONDENSATION



Approfondimenti

Viola & Jones https://www.cs.cmu.edu/~efros/courses/LBMV07/Papers/viola-cvpr-01.pdf

Lienhart http://www.lienhart.de/Prof._Dr._Rainer_Lienhart/Source_Code_files/ICIP2002.pdf

CONDENSATION http://link.springer.com/article/10.1023%2FA%3A1008078328650

Next Step: OpenCV

1. Nella prossima lezione vedremo come addestrare un sistema di localizzazione oggetti con OpenCV

2. Cosa serve:
   1. Ambiente Windows 7 + Webcam
   2. Visual Studio 10 oppure 11
   3. OpenCV 2.4.9
   4. CMake GUI
   5. XnView

Esercizio per Giovedì: compilare OpenCV

Social Q&A

@vs_AR
#askVisionary
www.vision-ary.net