Dal tracking alla classificazione

VISIONE ARTIFICIALE

dott. Alessandro Ferrari

Social Q&A

@vs_AR
#askVisionary
www.vision-ary.net

Object Tracking: possibili approcci

1. Tracking Frame by Frame: il tracking è emulato. Si applica continuamente (senza coerenza spazio-temporale) un localizzatore di oggetti 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 «innovativa».

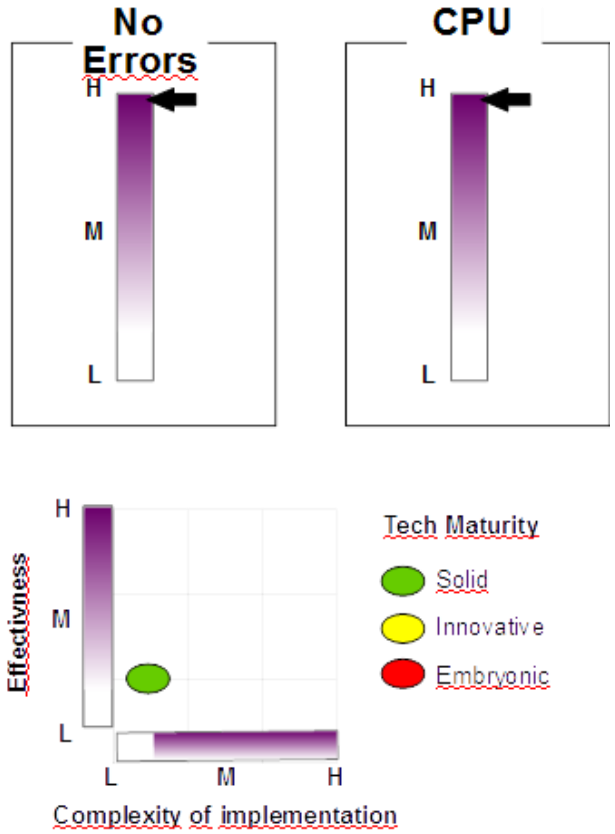
Queste definizioni non sono formalmente corrette ma rendono l'idea sui possibili approcci che si possono utilizzare per tracciare un oggetto. Nella fattispecie, il tracking by detection è riconducibile al caso del tracking by likelihood ma «storicamente» ha introdotto un approccio innovativo in letteratura e per questo si è preferito presentarlo separatamente.

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.

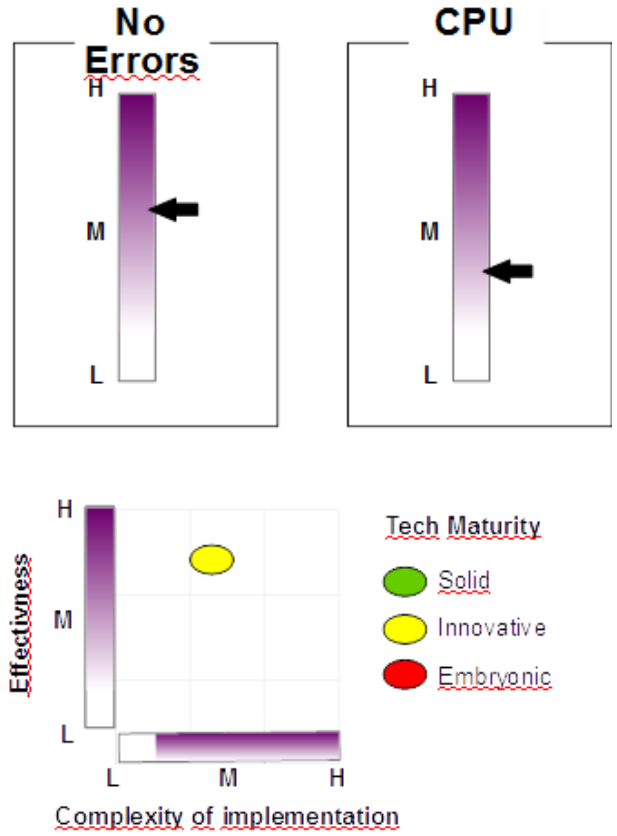


Tracking by likelihood

- 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 framework probabilistici la conferma della previsione avviene tramite la 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

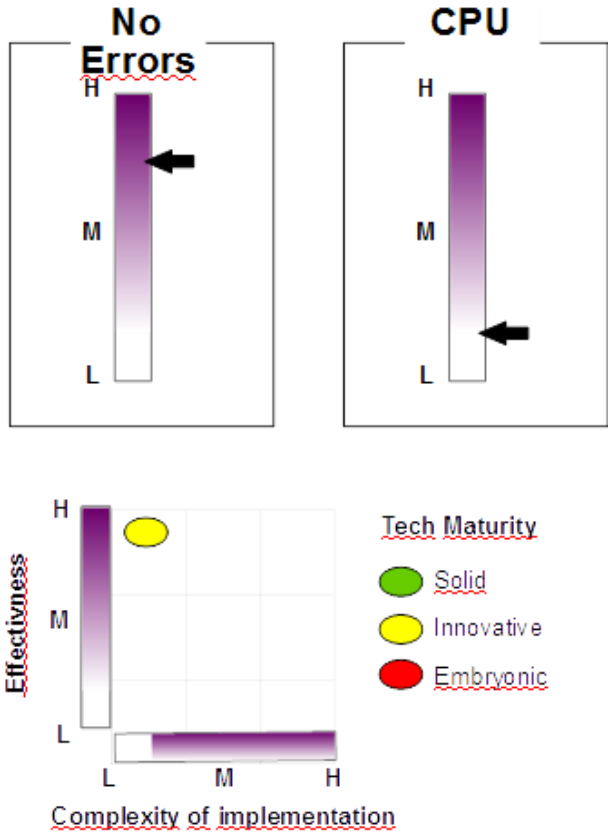
https://www.youtube.com/watch?v=W2qR60hrD2w

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.
  - Se il localizzatore di oggetti non fornisce una stima probabilistica (ma solo un responso T/F) è necessario trovare una misura adeguata.
  - Totale dipendenza dal tipo di oggetto tracciato.
  - Efficienza esasperata, moltiplicatore 50-100x.



Tracking di oggetti: stima probabilistica

Un tipico problema di interesse nella computer vision è quello di stimare lo stato di un oggetto che evolve nel tempo attraverso una serie di misure che si effettuano su di esso. L'oggetto tracciato viene descritto (modellato) attraverso un vettore xt che contiene tutte le caratteristiche che si ritengono rilevanti alla descrizione dell'oggetto (bordi, colore, punti di controllo, ecc…). L'osservazione dell'oggetto viene strutturata in un secondo vettore chiamato zt che contiene il responso di misurazione raccolto in fase di osservazione. Infine, Zt contiene la storia delle osservazioni zt effettuate nel tempo.

Obiettivo: stimare lo stato (non solo la posizione) di un oggetto che evolve nel tempo attraverso una serie di misure su di esso.

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 con un approccio di tipo probabilistico e pertanto si prestano in maniera diretta ad essere inclusi in un procedimento probabilistico ricorsivo, 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: predizione e aggiornamento

La ricorsione 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 un efficiente meccanismo di aggiornamento della conoscenza alla luce di una nuova osservazione non appena essa si rende disponibile.

Il modello evolutivo: predizione

- **Spostamento deterministico (deterministic drifting)**. La densità di probabilità di x trasla secondo una legge deterministica all'interno dello spazio degli stati che rappresentano l'oggetto. Alla base di questo fenomeno vi è la dinamica evolutiva dell'oggetto tracciato che assumiamo di conoscere a priori e che, si ipotizza, preveda fedelmente il comportamento dell'oggetto stesso. Ad esempio se volessimo tracciare una sfera puntiforme lungo una retta e conoscessimo la velocità dell'oggetto ad un certo istante t, potremmo prevedere in maniera precisa la posizione dell'oggetto all'istante successivo t + 1, secondo le classiche leggi orarie che descrivono il moto di oggetti puntiformi.

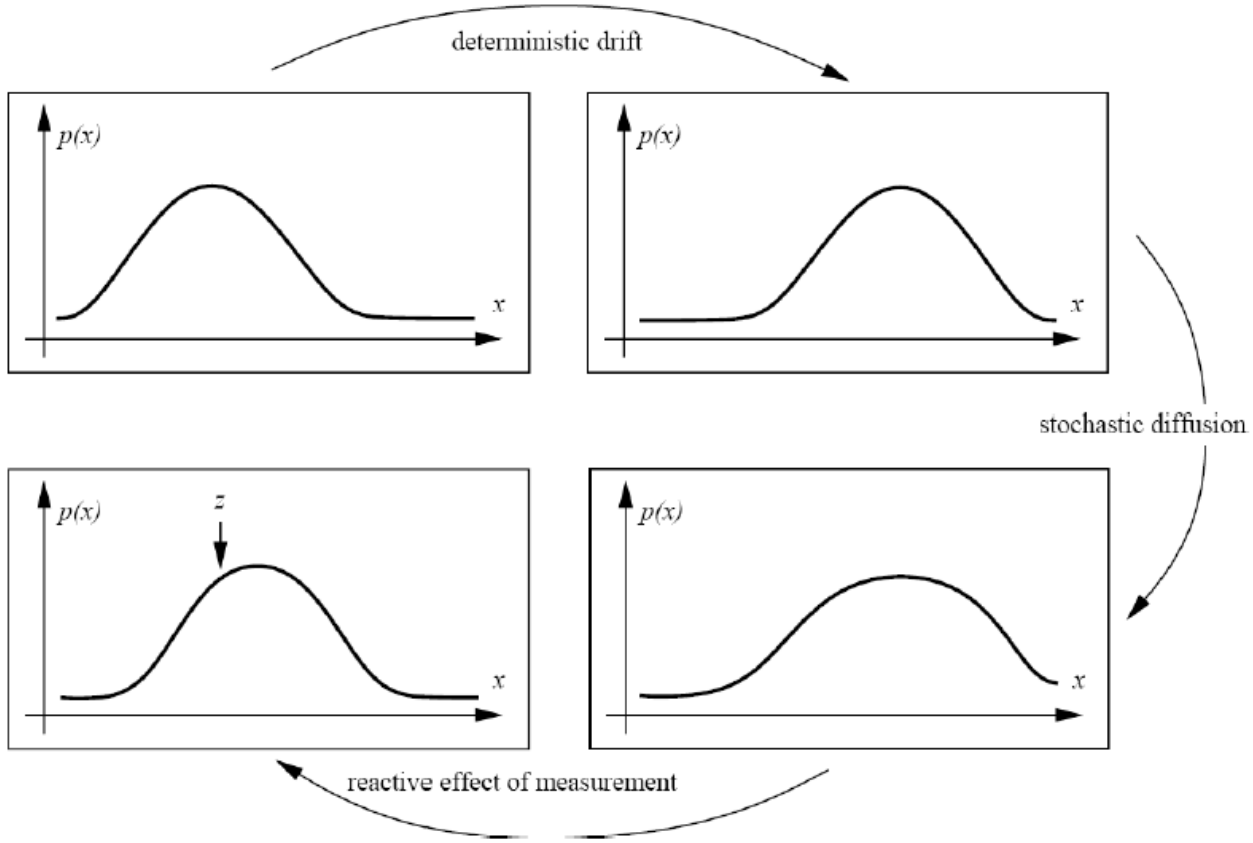
- **Diffusione stocastica (stochastic difusion).** Siccome la dinamica deterministica dell'oggetto in genere non è nota, o comunque non definita in maniera esatta, nella seconda fase, la varianza della distribuzione traslata viene modicata secondo un modello stocastico che mira ad allargare la distribuzione nello spazio degli stati; in genere il modello è rappresentato da rumore bianco introdotto ad hoc. Il rumore genera una perturbazione dello spazio delle ipotesi che permette di «muoversi» in un intorno della soluzione ottimale stimata al passo precedente evitando «il collassamento» delle ipotesi all'interno di «minimi locali».

Il modello di osservazione

- In conclusione del ciclo, la distribuzione traslata e diffusa nei due precedenti step viene aggiornata secondo la verosimiglianza dell'osservazione effettuata. La moda della distribuzione corrisponde (ad esempio) alla configurazione che ha ottenuto un riscontro (in termini di verosimiglianza) maggiore degli altri. Quest'ultima fase rappresenta l'aggiornamento della conoscenza sul sistema.

- Le situazioni reali sono tuttavia più complesse: in primo luogo la presenza di occlusioni e falsi rilevamenti (false detection) dà vita ad una sorta di competizione tra le osservazioni che in un certo modo forza la distribuzione ad assumere conformazioni multi-modali, in cui le mode possono muoversi in direzioni anche totalmente opposte tra loro; inoltre è naturale estendere il caso base (di tracciamento di un solo oggetto) al caso in cui gli oggetti da tracciare contemporaneamente siano diversi. Per questi motivi è necessario utilizzare un approccio multi-modale (quindi non gaussiano) per la definizione del sistema di tracciamento

Il modello evolutivo: rappresentazione grafica 1D monomodale



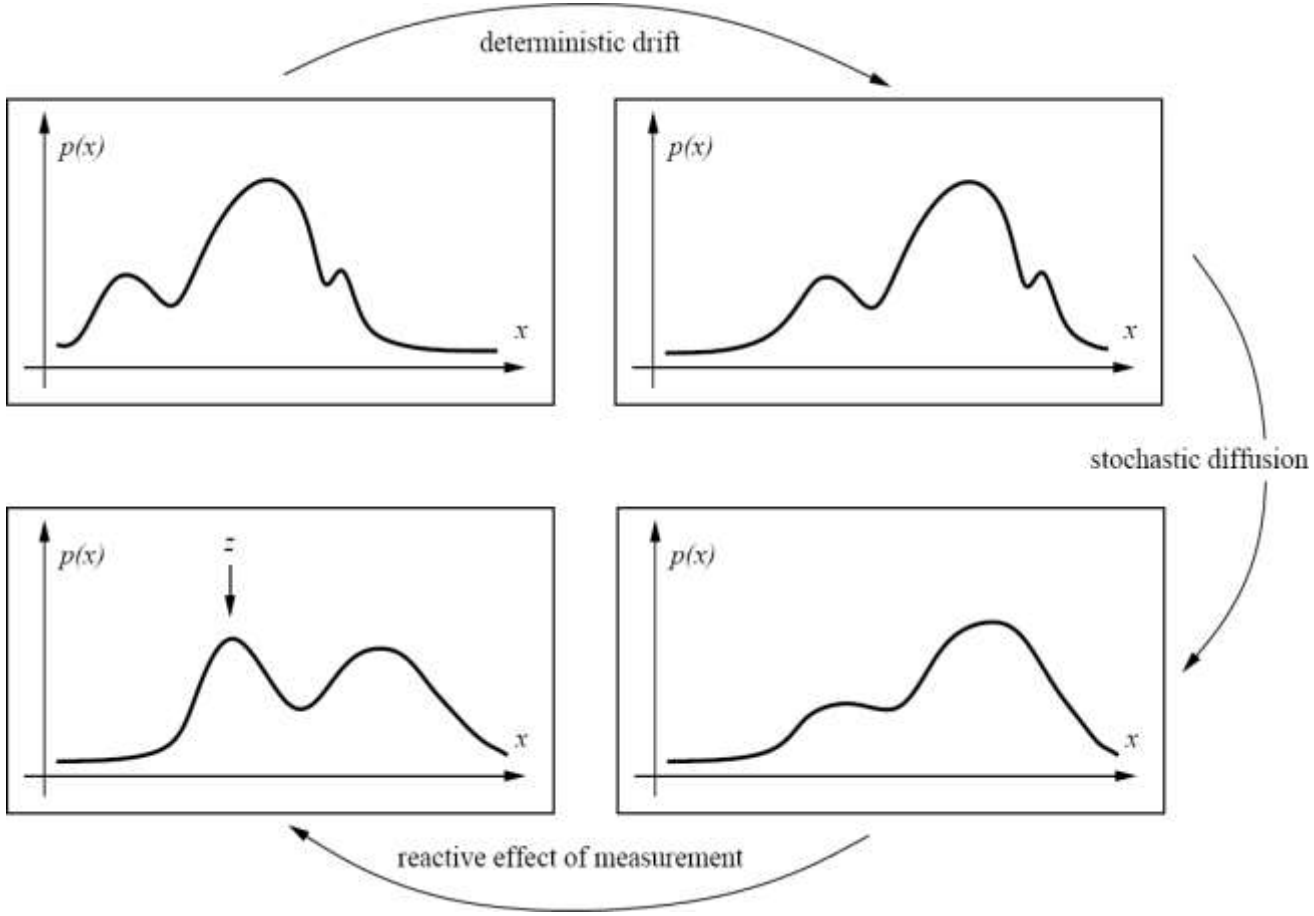
Dal mono-modale al multi-modale: ipotesi di Gaussianità

I processi monomodali sono tipici delle situazioni mono-tracking, ad esempio una persona di fronte alla webcam. In condizioni non controllate (ma non solo), la «gaussianità» del processo è solo teorica. Infatti il framework di tracking può assumere aspetto multi-modale a causa di:

- distribuzione a priori non gaussiana.
- La densità dell'evoluzione può assumere forma non gaussiana data la natura del processo o data la presenza di una dinamica non lineare del sistema.
- Processo di osservazione non gaussiano. In ultima analisi, a causa di occlusioni del moto, il processo di osservazione potrebbe non risultare gaussiano.

Un approccio per ovviare alle problematiche di cui sopra è quello di utilizzare una mistura additiva di N gaussiane (additive gaussian mixture model) per gestire la multi-modalità delle distribuzioni in gioco. Questa strategia complica la trattazione formale del problema nonché l'implementazione vera e propria a livello algoritmico.

Il modello evolutivo: rappresentazione grafica 1D multimodale



Particle Filter: dal continuo al discreto

L'idea chiave alla base dell'approccio particle filter è di rappresentare la distribuzione a posteriori p(xt|Zt) attraverso un set di campioni S.

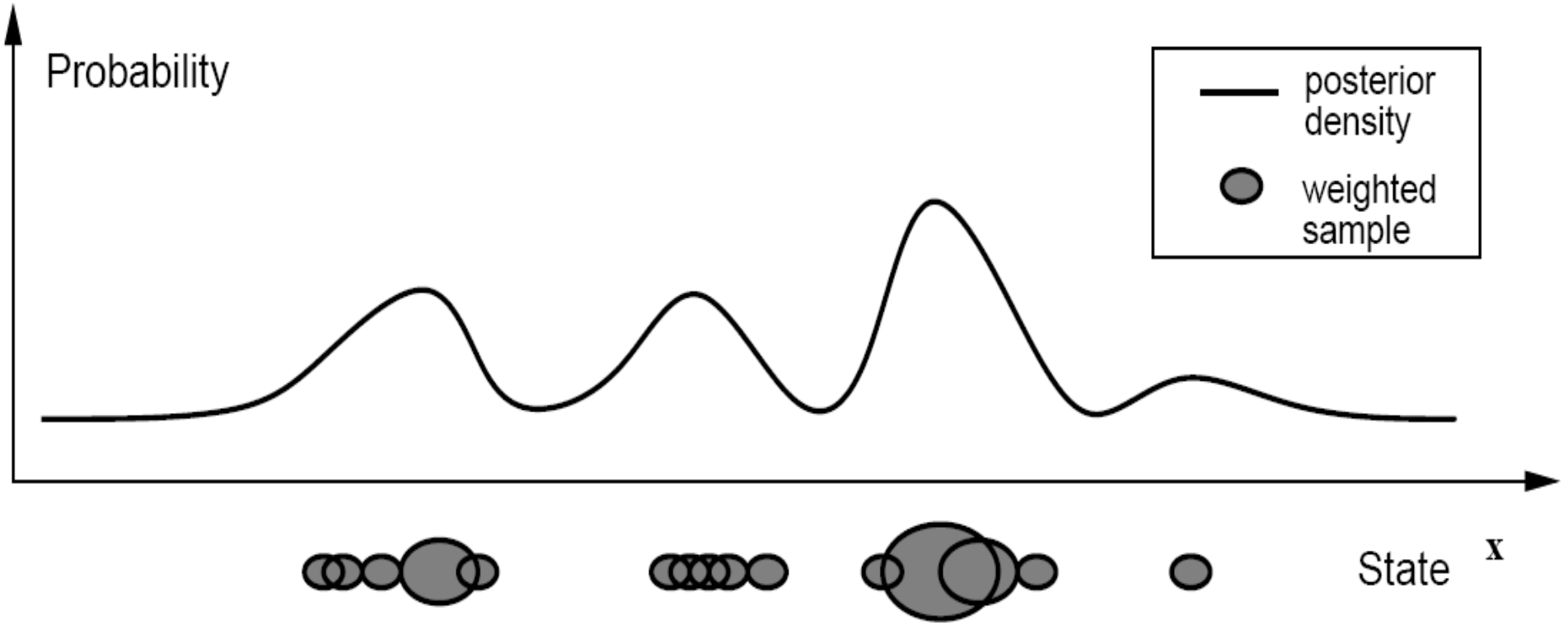
$$\mathcal{S} = \left\{ \mathrm{s}_t^{(i)}, \pi_t^{(i)} \right\};\ i = 1, \ldots, N$$

dove s rappresenta una possibile configurazione dell'oggetto e π il peso (normalizzato) che ne identifica l'importanza assunta durante il processo di analisi. La densità a posteriori sarà pertanto stimata come:

$$p(\mathrm{x}_t = \mathrm{s}_t | Z_t) \approx \sum_{i=1}^{N} \pi_t^{(i)} \delta(\mathrm{s}_t - \mathrm{s}_t^{(i)})$$

Spesso ci si riferirà ai campioni s anche col termine ipotesi poiché ogni campione rappresenta un possibile stato che il nostro oggetto può assumere. L'affidabilità è proporzionale al numero dei campioni N utilizzato; si dimostra infatti che per N che tende a infinito la rappresentazione approssimata converge alla rappresentazione funzionale della distribuzione a posteriori.

Modello a particelle: particle filtering

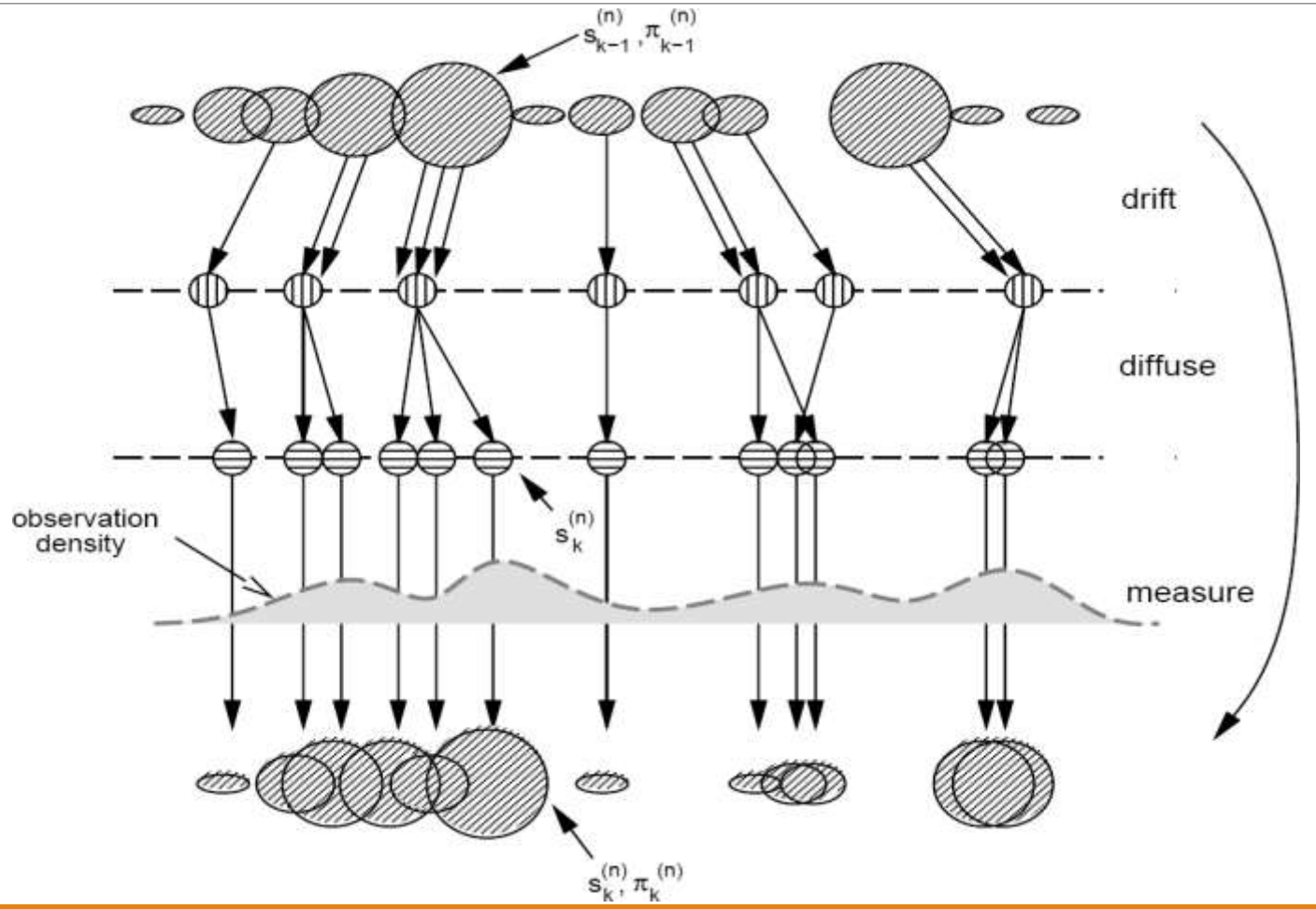


Particle Filtering: vantaggi

Utilizzando un'approssimazione discreta tramite N campioni non è necessario definire una rappresentazione funzionale della distribuzione a posteriori (che generalmente non è nota). Inoltre L'efficienza computazionale è sotto controllo e dipende da N, scelto a priori da noi o tarato in maniera iterativa secondo delle logiche di retroazione.

L'algoritmo originale di CONDENSATION è un ciclo continuo di predizione e aggiornamento di un set S di campioni di numerosità fissata N che ha come obiettivo la stima iterativa e approssimata della probabilità a posteriori di una variabile casuale x che rappresenta l'oggetto in analisi. Ad ogni frame, l'algoritmo genera dal precedente set di campioni un nuovo set di campioni secondo tre fasi ben distinte.

Tracking by CONDENSATION



Ciclo di CONDENSATION

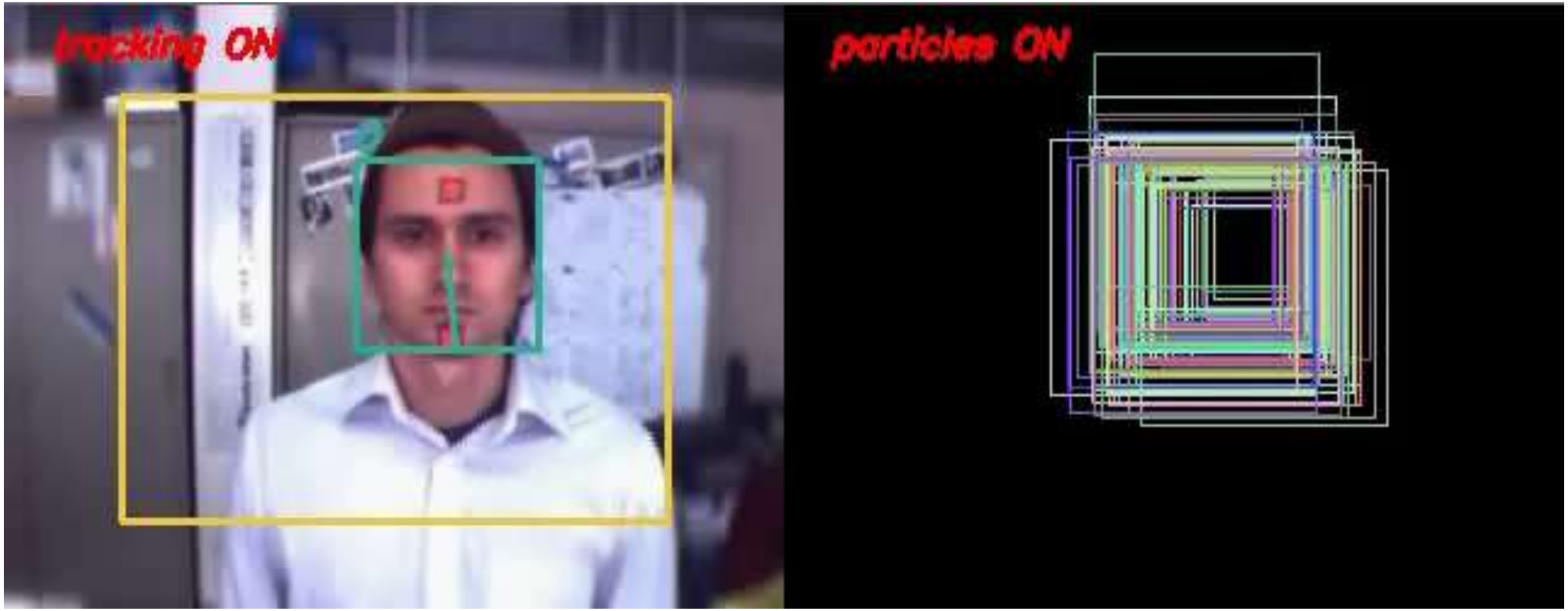
1. Campionamento (Sampling): un numero fissato di campioni N viene estratto dal set di campioni disponibile all'istante precedente; le particelle con peso elevato verranno scelte con frequenza maggiore rispetto alle analoghe con pesi bassi. La ciclicità del campionamento permette la selezione preponderante delle particelle che hanno un peso rilevante, evitando così situazioni di degenerazione della capacità descrittiva dell'insieme S.

2. Previsione (Drift and Diffusion): l'insieme campionato viene a questo punto sottoposto a un passo di predizione secondo le modalità evidenziate in precedenza, ovvero si calcola l'evoluzione delle particelle campionate. I campioni dell'insieme generato non sono ancora associati a nessun peso che ne identichi l'importanza.

3. Misura (Measure): i campioni che descrivono lo stato dell'oggetto tracciato vengono valutati secondo un modello di osservazione che assegna a ciascuno di essi un peso proporzionale alla verosimiglianza dell'osservazione. L'insieme di campioni S così ottenuto rappresenta un'approssimazione della distribuzione a posteriori del sistema.

CONDENSATION: prova su campo

CONDENSATION: natura generica delle particelle (pose estimation).



CONDENSATION: gestione occlusioni

Dal Tracking alla Classificazione

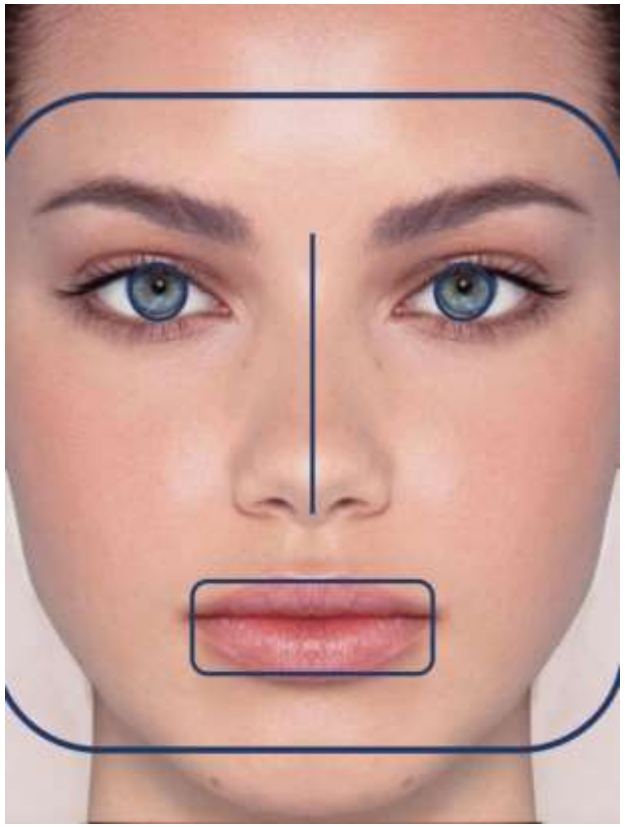
- Abbiamo ottenuto un tracking affidabile, tuttavia la bounding box (per come è costruita la procedura di addestramento) non offre una precisa localizzazione delle caratteristiche dell'oggetto.

- Questa imprecisione in genere si traduce in una difficoltà ulteriore che si somma alla complessità del problema di classificazione. Sostanzialmente se non si rettifica in maniera "sensata" l'immagine potrebbe essere non possibile assolvere al task di classificazione.

- Nel caso dei volti, la variazione di posa è gestita (entro certi limiti) dal detector in maniera trasparente al «programmatore», in ottica di classificazione questa variabilità va invece gestita con la necessaria cura.

- Per i volti si usano generalmente due approcci:
  - Sub part detection
  - ASM/AAM

Sub Part detection



- L’obiettivo è riportare la faccia localizzata in una “faccia” media che renda la classificazione più agevole. Bisogna definire una dimensione “target” e definire almeno due (meglio tre) punti di controllo sui quali proiettare la faccia localizzata.
- Il metodo più efficiente prevede la localizzazione di alcune feature facciali da utilizzare come “ancore” per la normalizzazione.
- La localizzazione di occhi e bocca è un procedimento standardizzato che può essere effettuato in maniera analoga a quanto fatto per le facce (boosted cascade).
- Normalizzare sugli occhi è facile, sulla bocca più difficile.

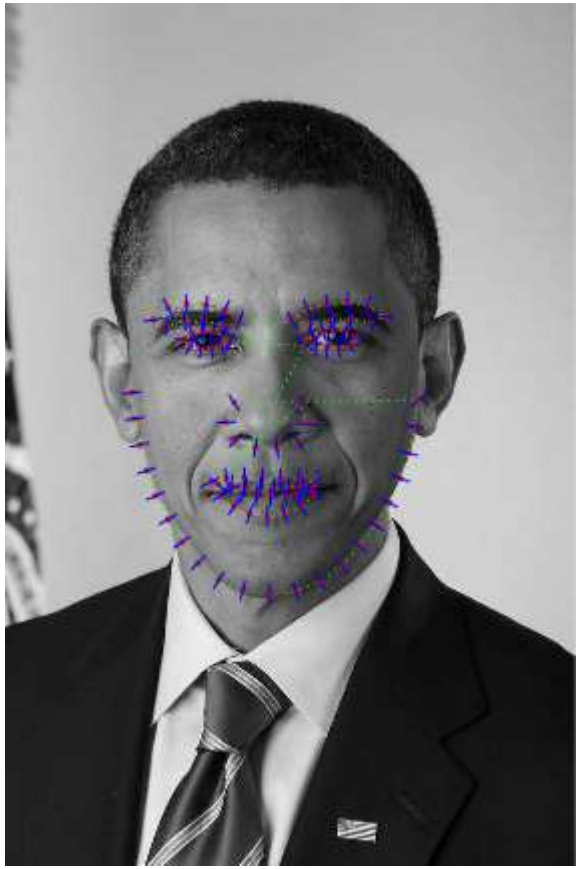
ASM/AAM normalization

La normalizzazione basata sulle sub-part potrebbe non essere sufficientemente precisa per diversi motivi, tra cui:

1. Instabilità dei punti di controllo: se localizzo erroneamente uno dei punti la proiezione è completamente sbagliata.
2. Instabilità del metodo di ricerca: come per le facce, localizzare un “oggetto” con una boosted cascade potrebbe non offrire la sufficiente precisione.

Esistono approcci più robusti seppur computazionalmente più costosi. I più noti sono quelli basati su ASM o AAM. Sostanzialmente si crea a priori un modello dell’oggetto basato su una serie di “landmark” che identificano bordi e corner dell’oggetto “facilmente” ritrovabili nelle immagini. La topologia determinata da questi punti è detta “shape” dell’oggetto. Nel caso delle AAM viene data grande importanza alla tessitura dell’oggetto e il modello generato comprende anche informazione sulla tessitura.

ASM fitting pre-classificazione



Social Q&A

@vs_AR
#askVisionary
www.vision-ary.net