

RICONOSCIMENTO E TRACCIAMENTO DELLE PERSONE

Riconoscimento Full-Body

Le videocamere sono ampiamente utilizzate per la sorveglianza basata sulla visione. Sistemi completamente automatici dovrebbero eseguire (1) il rilevamento di una persona, (2) il tracciamento di una persona e (3) il riconoscimento di una persona.

Per quanto riguarda il riconoscimento di una persona (scenario non invasivo), un approccio consiste nel considerare l'aspetto esteriore complessivo della persona da riconoscere. La grande varietà di colori e textures degli abiti e la possibilità di combinare giacche, camicie, pantaloni, scarpe e accessori come sciarpe e cappelli rappresentano una soluzione per l'identificazione. Gli stili differenti e le tipiche combinazioni di colori possono essere associati a un'unica persona e quindi possono essere utilizzati per l'identificazione (si veda la figura riportata nella slide successiva).

Per una descrizione opportuna dell'aspetto di una persona, possono essere utilizzati algoritmi di estrazione del colore e descrittori delle textures.

RICONOSCIMENTO E TRACCIAMENTO DELLE PERSONE

Riconoscimento Full-Body



a



b



c



d



e

Persone differenti il cui aspetto differisce a causa delle loro dimensioni e a causa del colore e delle textures del loro abbigliamento (a-d). Tuttavia, un abbigliamento simile potrebbe non permettere una distinzione appropriata degli individui (d-e).

RICONOSCIMENTO E TRACCIAMENTO DELLE PERSONE

Riconoscimento Full-Body

Color Features

Per quanto riguarda le features associate al colore si possono considerare gli istogrammi relativi al colore e i *Color Structure Descriptor (CSD)*.

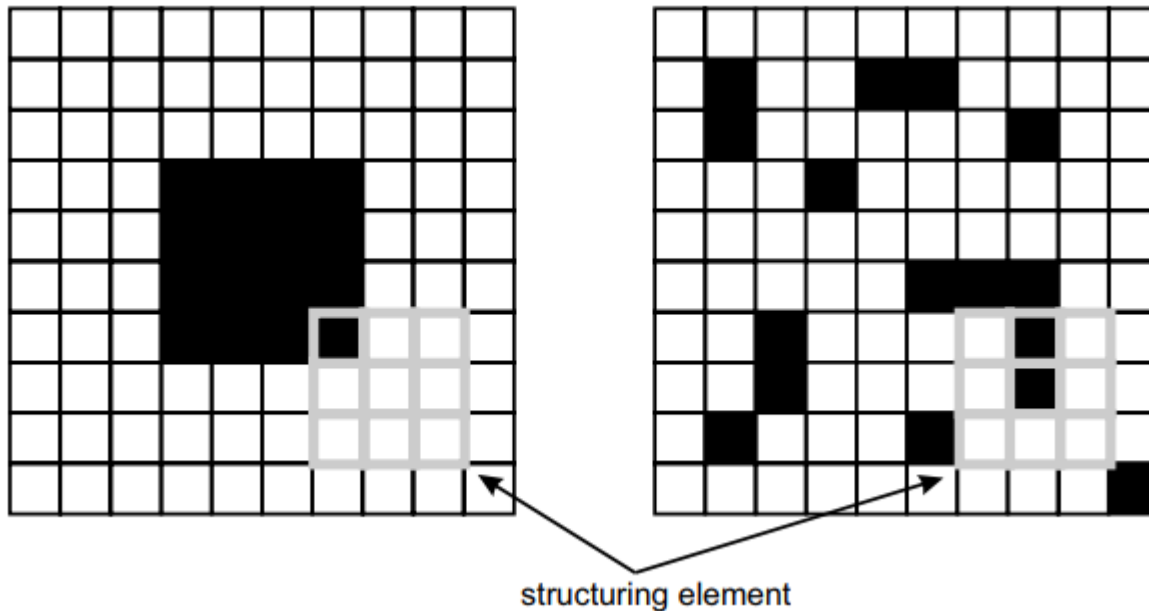
Gli istogrammi sono comunemente utilizzati per descrivere l'aspetto degli oggetti contando il numero di pixel di ciascun colore presente. Ogni bin dell'istogramma rappresenta un singolo colore o un intervallo (range) di colori. Con l'aumentare del numero di bin, crescono le dimensioni dell'istogramma e quindi aumenta la richiesta di risorse di memoria. Se si rappresenta un range di colori con un bin, aumenta l'invarianza al rumore: piccole variazioni di colore non hanno effetto poiché viene incrementato lo stesso bin nella costruzione dell'istogramma. Quindi occorre trovare un compromesso tra l'aspetto computazionale, l'invarianza al rumore (piccolo numero di bin) e la capacità di differenziazione (grande numero di bin). La dimensione di un istogramma è determinata dal numero di canali che vengono considerati; ad esempio, un istogramma costruito sullo spazio RGB è tre volte più grande di un istogramma associato al valore di intensità.

RICONOSCIMENTO E TRACCIAMENTO DELLE PERSONE

Riconoscimento Full-Body

Color Features

Il CSD può essere definito come un'estensione dell'istogramma associato al colore: con il *CSD* non viene rappresentata solo la distribuzione del colore ma si considera anche la distribuzione spaziale dei colori. Utilizzando il *CSD*, nella costruzione dell'istogramma non si considera solo il singolo pixel ma un suo intorno; ciò viene ottenuto mediante lo *structuring element*.

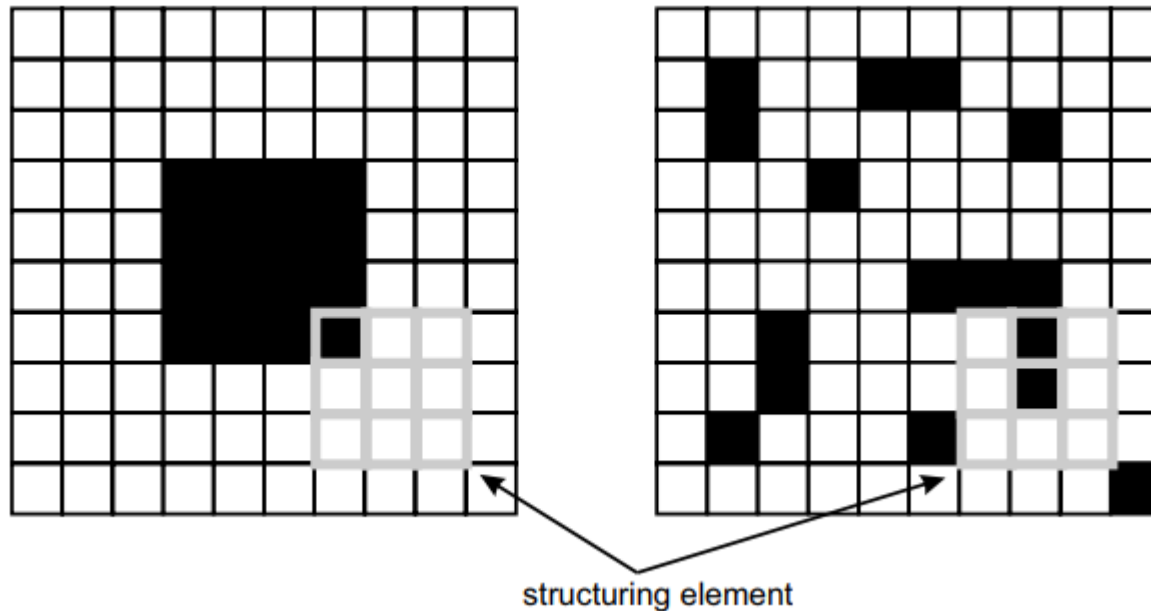


Structuring elements 3x3 posizionati in due immagini per considerare l'intorno di uno stesso pixel. In entrambi i casi, i bins dell'istogramma riferiti ai colori bianco e nero sono incrementati di 1.

RICONOSCIMENTO E TRACCIAMENTO DELLE PERSONE

Riconoscimento Full-Body

La figura mostra l'utilizzo dello *structuring element* (in questo esempio esso ha una dimensione 3x3 ed è evidenziato in grigio) in due immagini caratterizzate da due colori (bianco e nero). Entrambe le immagini hanno una dimensione di 10x10 pixels (16 pixels sono neri). Per costruire un istogramma *CSD*, lo *structuring element* viene traslato per riga sull'intera immagine. Un bin dell'istogramma viene incrementato di 1 se il colore corrispondente può essere trovato nello *structuring element* anche se

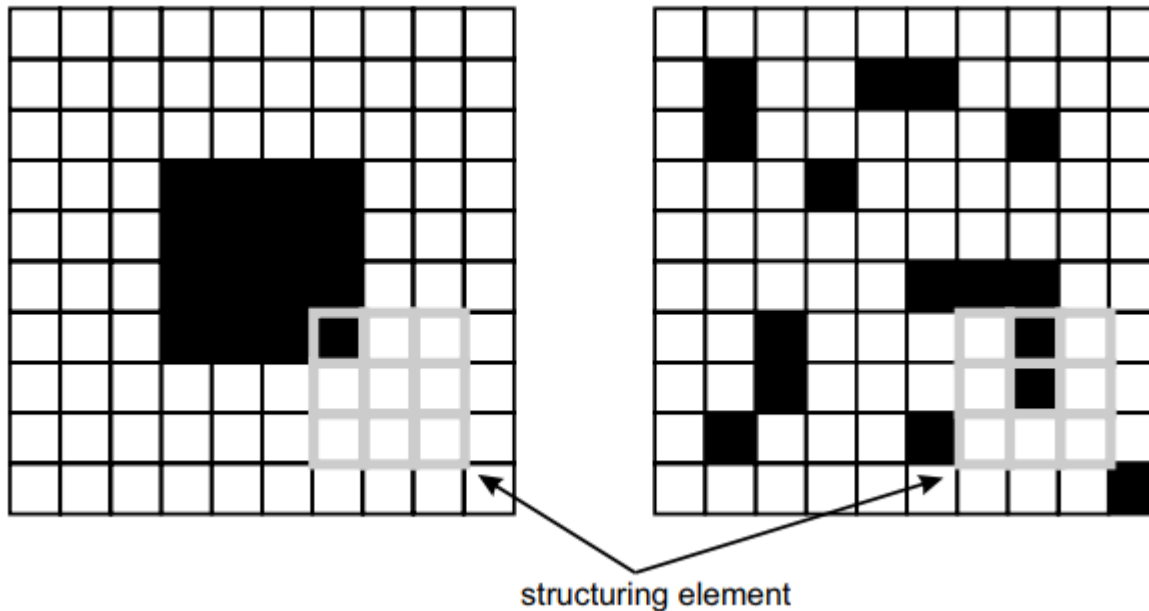


Structuring elements 3x3 posizionati in due immagini per considerare l'intorno di uno stesso pixel. In entrambi i casi, i bins dell'istogramma riferiti ai colori bianco e nero sono incrementati di 1.

RICONOSCIMENTO E TRACCIAMENTO DELLE PERSONE

Riconoscimento Full-Body

più di un pixel ha questo colore. Nell'esempio riportato in figura, nell'immagine di sinistra vengono trovati 1 pixel nero e 8 pixel bianchi. I valori del bin riferito ai colori bianco e nero vengono incrementati di 1. Ciò accade anche nella figura di destra, anche se in questo caso nello *structuring element* si hanno 2 pixel neri e 7 pixel bianchi. Applicando lo *structuring element*, i bin dell'istogramma avranno i valori 36 (nero) e 60 (bianco) per l'immagine a sinistra, 62 (nero) e 64 (bianco) per

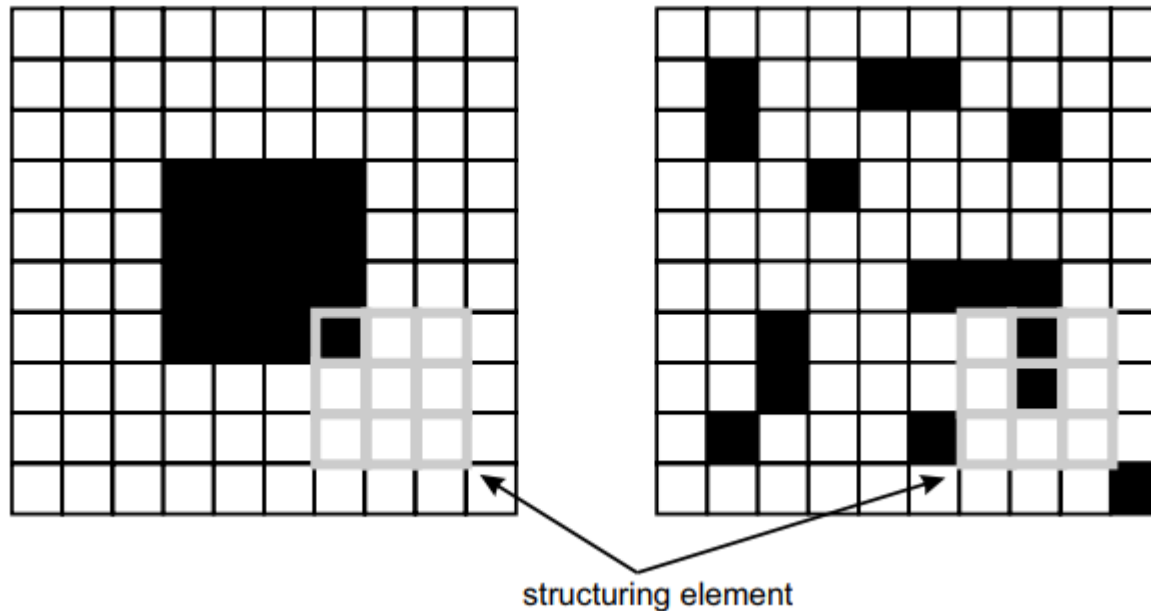


Structuring elements 3x3 posizionati in due immagini per considerare l'intorno di uno stesso pixel. In entrambi i casi, i bins dell'istogramma riferiti ai colori bianco e nero sono incrementati di 1.

RICONOSCIMENTO E TRACCIAMENTO DELLE PERSONE

Riconoscimento Full-Body

l'immagine a destra. Come può essere notato per i pixel di colore nero, un colore maggiormente distribuito è caratterizzato da un valore più grande del bin nell'istogramma. Un istogramma classico costruito per queste due immagini ha valore dei bin pari a 16 (nero) e 84 (bianco), quindi le immagini non possono essere differenziate tramite un istogramma classico. Un istogramma *CSD* mostra invece la differenza. Il *CSD* è stato definito nello standard MPEG-7.



Structuring elements 3x3 posizionati in due immagini per considerare l'intorno di uno stesso pixel. In entrambi i casi, i bins dell'istogramma riferiti ai colori bianco e nero sono incrementati di 1.

RICONOSCIMENTO E TRACCIAMENTO DELLE PERSONE

Riconoscimento Full-Body

Texture Features

L'abbigliamento non è caratterizzato solo dal colore, ma anche dalla texture. Anche la texture può contenere informazioni importanti sull'aspetto di una persona.

Per quanto riguarda le features associate alla texture, si possono utilizzare le features basate sul concetto di Oriented Gaussian Derivatives (OGD) e sul concetto di Homogeneous Texture Descriptor (HTD). Le features risultanti garantiscono buoni risultati, soprattutto se combinate con altre features.

RICONOSCIMENTO E TRACCIAMENTO DELLE PERSONE

Tracciamento camera-based

Numerosi task richiedono la percezione e la comprensione del mondo reale. Come gli occhi umani, le immagini delle camere sono input adeguati per fornire le informazioni richieste e quindi la visione artificiale è uno strumento che può essere utilizzato in tale ambito. Alcune abilità fondamentali per la comprensione delle scene sono la percezione e l'analisi delle persone nel loro ambiente naturale (compito «Looking at People» della visione artificiale). Tali abilità sono un elemento di analisi comune in molte applicazioni, come ad esempio robot di servizio autonomi, videosorveglianza, riconoscimento di persone, interfacce uomo-macchina.

Il tracciamento delle persone (*people tracking*) si riferisce al rilevamento degli umani in una sequenza di immagini e al tenere traccia dei loro movimenti nella scena. Spesso il tracciamento consente di ricavare informazioni utili per analisi successive, come ad esempio l'analisi della postura del corpo, della posizione delle persone nella scena tridimensionale, del comportamento o dell'identità.

Le situazioni del mondo reale sono molto diverse e molto articolate per permettere un'interpretazione senza conoscenza a priori, quindi il tracciamento è un problema complesso (si veda la figura riportata nella slide seguente).

RICONOSCIMENTO E TRACCIAMENTO DELLE PERSONE

Tracciamento camera-based

Inoltre, le persone hanno un'ampia varietà di forme e di aspetto. Un osservatore umano risolve il compito di analisi dell'immagine attraverso la comprensione tridimensionale della situazione nella sua interezza, utilizzando una vasta conoscenza di base su ogni singolo elemento della scena, sulle leggi fisiche e sul comportamento umano.



(a)



(b)



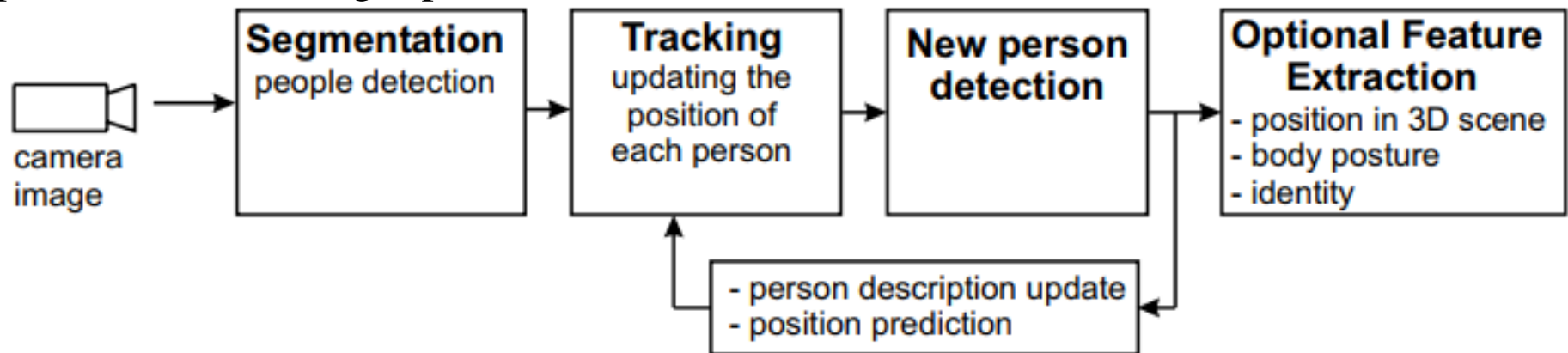
(c)

Situazioni diverse richiedono approcci più o meno complessi in base al numero di persone presenti nella scena, alla prospettiva della camera e alla distanza della camera.

RICONOSCIMENTO E TRACCIAMENTO DELLE PERSONE

Tracciamento camera-based

Come mostrato nella figura, un sistema di tracciamento delle persone è caratterizzato da diverse parti. Viene eseguita una segmentazione, nella quale avviene una ricerca delle regioni dell'immagine caratterizzate da un'alta probabilità di rappresentare persone. Tale fase di segmentazione precede la fase di tracking; la fase di tracking utilizza la conoscenza e le osservazioni delle immagini precedenti per associare ogni persona monitorata alla regione corrispondente e per rilevare nuove persone che entrano nella scena. Comunque, esistono anche approcci che combinano le fasi di segmentazione e di tracking. La soluzione del problema di tracciamento delle persone richiede quindi di svolgere diverse fasi, come ad esempio il rilevamento iniziale e la segmentazione delle persone nell'immagine, il tracking e, se richiesto, il calcolo della posizione reale di ogni persona.



Schema della struttura generale di un sistema di tracciamento delle persone.

RICONOSCIMENTO E TRACCIAMENTO DELLE PERSONE

Tracciamento camera-based

Segmentazione

La segmentazione di un'immagine indica il processo di raggruppamento di pixel con alcune proprietà comuni e può essere quindi definita come un problema di classificazione. Nel caso di tracciamento delle persone, l'obiettivo di tale fase è quello di trovare i pixel dell'immagine che sono associati a persone.

Il risultato della segmentazione è rappresentato da una matrice di valori booleani $\mathbf{M}(x, y)$. Tale matrice etichetta ogni pixel dell'immagine $\mathbf{I}(x, y)$ come membro di una delle due classi «background scene» o «probably person» (si veda la parte b della figura):

$$\mathbf{M}(x, y) = \begin{cases} 0 & \text{if background scene} \\ 1 & \text{if moving foreground object (person)} \end{cases}$$



(a)



(b)



(c)

Segmentazione dell'immagine: a) immagine acquisita dalla camera; b) foreground segmentation; c) separate person segmentation.

RICONOSCIMENTO E TRACCIAMENTO DELLE PERSONE

Tracciamento camera-based

Segmentazione

La segmentazione a due classi spiegata nella slide precedente è sufficiente in numerose applicazioni. Invece, il tracking individuale di persone durante la sovrapposizione tra persone richiede un'ulteriore segmentazione delle regioni in primo piano (si veda la parte c della figura); ogni pixel deve essere etichettato come una delle classi: scena di sfondo (background scene), n -esima persona, primo piano (foreground) sconosciuto (quindi è probabile che ci sia una nuova persona). Tale risultato di segmentazione può essere rappresentato da singole maschere del primo piano (individual foreground masks) $\mathbf{M}_n(x, y)$ per ogni persona tracciata $n \in \{1, \dots, N\}$ e da una maschera $\mathbf{M}_0(x, y)$ per rappresentare i pixel del primo piano (foreground) rimanenti.



(a)



(b)



(c)

Segmentazione dell'immagine: a) immagine acquisita dalla camera; b) foreground segmentation; c) separate person segmentation.

RICONOSCIMENTO E TRACCIAMENTO DELLE PERSONE

Tracciamento camera-based

Foreground segmentation

Si ricordi come è stata descritta la parte b dell'immagine riportata nella slide precedente: essa rappresenta la *foreground segmentation*. La maggior parte delle applicazioni di tracciamento delle persone utilizza una o più camere fisse. Il vantaggio di questo tipo di configurazione è che ogni pixel dell'immagine di sfondo (background image) $\mathbf{B}(x, y)$, cioè l'immagine che rappresenta la scena vuota e statica senza persone (si veda la parte a della figura), ha un valore costante finché non ci sono variazioni di illuminazione.



a) *background model*



b) *camera image*



c) *background subtraction*



d) *shadow reduction*



e) *morphological filtering*

Fasi della foreground segmentation.

A REALTA' VIRTUALE

RICONOSCIMENTO E TRACCIAMENTO DELLE PERSONE

Tracciamento camera-based

Foreground segmentation

Un approccio per rilevare oggetti in movimento consiste nell'etichettare come primo piano (foreground) tutti i pixel dell'immagine iniziale $\mathbf{I}(x, y, t)$ (al tempo t) che differiscono rispetto a $\mathbf{B}(x, y)$ di una quantità superiore a una soglia predefinita θ :

$$\mathbf{M}(x, y, t) = \begin{cases} 1 & \text{if } \|\mathbf{B}(x, y) - \mathbf{I}(x, y, t)\| > \theta \\ 0 & \text{otherwise} \end{cases}$$



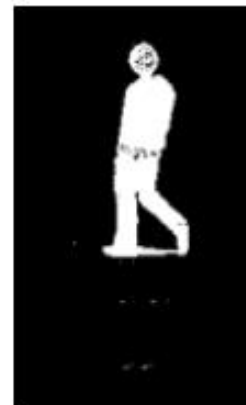
a) background model



b) camera image



c) background subtraction



d) shadow reduction



e) morphological filtering

Fasi della foreground segmentation.

A REALTA' VIRTUALE

RICONOSCIMENTO E TRACCIAMENTO DELLE PERSONE

Tracciamento camera-based

Tracking

La maschera associata al primo piano $\mathbf{M}(x, y)$ (definita nelle slide precedenti) etichetta ogni pixel dell'immagine come «potential foreground object» o come «image background». Tuttavia, fino a questo momento, non è stato assegnato alcun significato alle regioni in primo piano: questo è il compito della fase di tracking.

Il tracciamento (tracking) indica il processo di associazione dell'identità di un oggetto in movimento (persona) allo stesso oggetto dei precedenti frame di una sequenza di immagini. Viene quindi seguita la traiettoria dell'oggetto nel piano dell'immagine o rispetto alle coordinate del mondo reale. Per raggiungere tale obiettivo, vengono utilizzate alcune features dell'oggetto tracciato, come ad esempio la predizione della posizione o le descrizioni dell'aspetto. Queste features vengono sfruttate per classificare ed etichettare le regioni associate al primo piano (cioè le regioni che non rappresentano lo sfondo).

RICONOSCIMENTO E TRACCIAMENTO DELLE PERSONE

Tracciamento camera-based

Tracking

Esistono due tipologie di algoritmi di tracciamento:

- *tracking following detection*: tutti i candidati ad essere classificati come umani nell'immagine vengono rilevati prima che venga effettuato il tracking. In tale algoritmo, il tracking consiste principalmente in un problema di classificazione; tale problema utilizza varie *features* per assegnare ogni candidato a una delle descrizioni interne associate a tutte le persone osservate.
- *combined tracking and detection*: le fasi di rilevamento e di tracking sono combinate. La predizione dell'area dell'immagine di ogni persona tracciata viene utilizzata come inizializzazione per cercare la nuova posizione. Tale ricerca può essere effettuata utilizzando la *segmented foreground mask* o l'immagine iniziale. Se la ricerca viene effettuata utilizzando l'immagine iniziale, il processo include spesso anche la fase di segmentazione, utilizzando le descrizioni del colore associate a tutte le persone in modo da migliorare la qualità della segmentazione o anche per separare le persone quando si verifica l'occlusione.

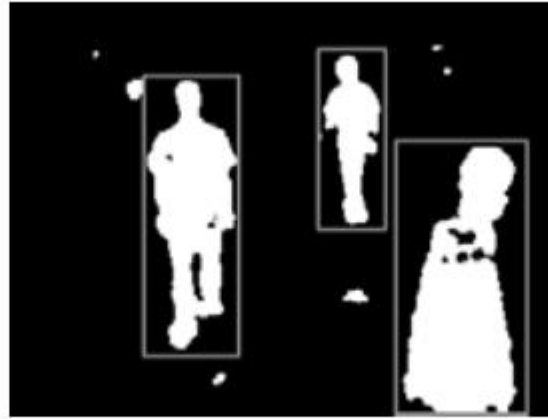
RICONOSCIMENTO E TRACCIAMENTO DELLE PERSONE

Tracciamento camera-based

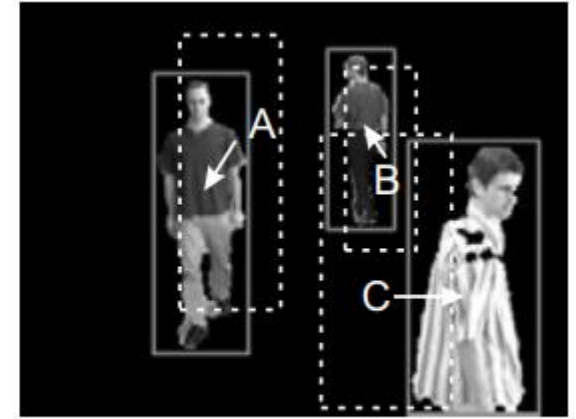
Tracking



(a)



(b)



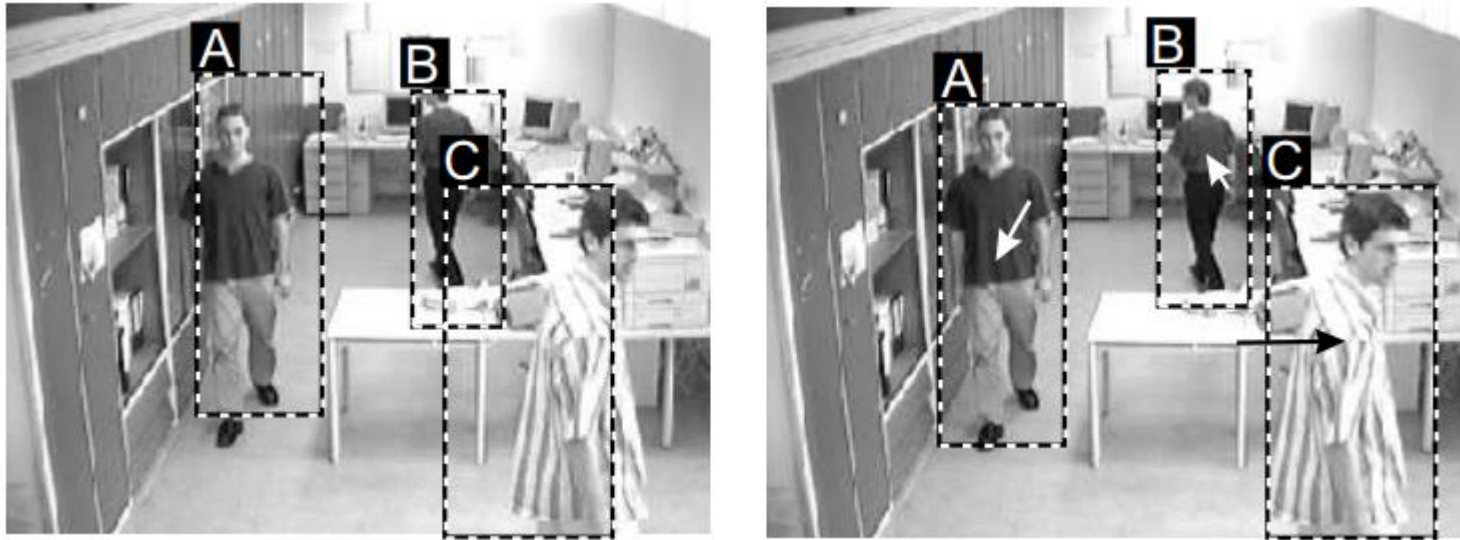
(c)

Tracking following detection: a) immagine acquisita dalla camera; b) foreground segmentation e person candidate detection; c) people assignment in base alla distanza e all'aspetto.

RICONOSCIMENTO E TRACCIAMENTO DELLE PERSONE

Tracciamento camera-based

Tracking



Combined tracking and detection con adattamento iterativo del modello.

RICONOSCIMENTO E TRACCIAMENTO DELLE PERSONE

Prestazioni di un sistema di riconoscimento

Si ipotizzi di utilizzare un sistema di riconoscimento per il controllo degli accessi. Il sistema deve fornire un risultato positivo (accesso consentito) o negativo (accesso negato). Le prestazioni del sistema possono essere valutate utilizzando una matrice di confusione (si veda la figura) e calcolando alcune metriche.

		Classe predetta	
		positivo	negativo
Classe reale	positivo	true positive (TP)	false negative (FN)
	negativo	false positive (FP)	true negative (TN)

Matrice di confusione per una classificazione binaria.

RICONOSCIMENTO E TRACCIAMENTO DELLE PERSONE

Prestazioni di un sistema di riconoscimento

True Positive (TP): il sistema consente l'accesso; tale risultato è corretto.

False Positive (FP): il sistema consente l'accesso; tale risultato è errato.

False Negative (FN): il sistema nega l'accesso; tale risultato è errato.

True Negative (TN): il sistema nega l'accesso; tale risultato è corretto.

		Classe predetta	
		positivo	negativo
Classe reale	positivo	true positive (TP)	false negative (FN)
	negativo	false positive (FP)	true negative (TN)

Matrice di confusione per una classificazione binaria.

RICONOSCIMENTO E TRACCIAMENTO DELLE PERSONE

Prestazioni di un sistema di riconoscimento

Esempi di metriche che possono essere calcolate sono:

$$Accuratezza = \frac{TN + TP}{TN + TP + FN + FP}$$

$$Precisione = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F_1 \text{ score} = \frac{2 \cdot Precisione \cdot Recall}{Precisione + Recall}$$



chiamato anche
Sensitivity

		Classe predetta	
		positivo	negativo
Classe reale	positivo	true positive (TP)	false negative (FN)
	negativo	false positive (FP)	true negative (TN)

Matrice di confusione per una classificazione binaria.

RICONOSCIMENTO E TRACCIAMENTO DELLE PERSONE

Prestazioni di un sistema di riconoscimento

Esempi di metriche che possono essere calcolate sono:

$$\text{Specificity} = \frac{TN}{FP + TN}$$

		Classe predetta	
		positivo	negativo
Classe reale	positivo	true positive (TP)	false negative (FN)
	negativo	false positive (FP)	true negative (TN)

Matrice di confusione per una classificazione binaria.

Riferimenti Bibliografici

- [1] Kraiss, K. -F. (2006). Advanced Man-Machine Interaction: Fundamentals and Implementation. Springer-Verlag Berlin Heidelberg. ISBN-10: 3-540-30618-8
- [2] Paramartha Dutta, Asit Barman (2020). Human Emotion Recognition from Face Images. Springer Singapore. ISBN: 978-981-15-3883-4