

Proponente: Ettore Stella

Fruitore: Nicola Mosca

Supervisione: Hamid Aghajan

#### Metodi per la ricostruzione multi-vista del movimento umano

Il progetto è stato iniziato nell'ambito del programma CNR Short Term Mobility 2008, presso la Stanford University. La short term mobility è avvenuta dal 5 al 25 Novembre 2008, presso il Wireless Systems Network Laboratory - Department of Electrical Engineering, sotto la supervisione del prof. Hamid Aghajan.

Il progetto è stato ideato al fine di affrontare il problema del riconoscimento di eventi in scene affollate. Quest'ultimo è diventato un argomento molto importante negli ultimi anni, a causa della crescente attenzione ed interesse in differenti contesti che spaziano dalla video-sorveglianza alla fruizione di eventi sportivi, dalla sicurezza all'intrattenimento. Nell'ambito del riconoscimento delle attività umane, uno dei modi per facilitare tale processo è quello di controllare la scena attraverso un approccio multi-camera, e laddove possibile, multi-sensoriale.

Il ricorso all'uso di più punti di vista consente di svincolarsi da alcuni problemi di elaborazione delle immagini, tipici dell'approccio ad una singola vista. In questo modo è possibile catturare ed identificare nello spazio tridimensionale gli attori presenti nella scena e attraverso l'analisi del loro movimento nella zona di interesse, cercare di classificarne il comportamento.

Il periodo di Short Term Mobility è stato principalmente sfruttato per confrontare le attività dell'ISSIA CNR, ed in particolare del gruppo di visione, con le attività del laboratorio Wireless Systems Network Laboratory della Stanford University e per impostare un progetto di ricerca comune nell'ambito della ricostruzione del movimento umano tramite un approccio multi-vista. Il confronto delle attività e del "modus operandi" dei due gruppi di ricerca ha

permesso di identificare affinità e punti di divergenza, con relativi vantaggi e svantaggi, dei differenti approcci utilizzati per poi delineare una strategia di base con cui affrontare la ricostruzione multi-vista. In particolare, è risultato evidente il comune interesse nei contesti di monitoraggio delle attività umane con approccio multi-vista, seppure con differenti finalità (l'ISSIA CNR è più focalizzato sui contesti di videosorveglianza, WSNL mostra maggiore interesse nei contesti di assisted living).

Nell'ambito del progetto, sono stati identificati differenti tasks, che, partendo dall'analisi di più punti di vista della stessa scena, possono portare alla ricostruzione, ed eventualmente alla interpretazione, di quanto osservato.

Al fine di limitare la complessità iniziale affrontata nella risoluzione del problema, ci si è limitati all'osservazione di contesti dinamici attraverso un set di telecamere fisse. Questo vincolo permette di identificare gli oggetti in movimento della scena facendo ricorso a tecniche, abbastanza collaudate, di background subtraction.

Il sistema ideato consiste di tre moduli in cascata, il primo dei quali operante con le immagini 2d ottenute dalle sequenze filmate. La skeletonizzazione 2d, assieme allo sfruttamento di vincoli esistenti nel contesto delle posture umane (ogni articolazione ha particolari gradi di libertà), permette ad un secondo modulo, operante in 3d, di effettuare una fusione spazio-temporale delle informazioni, per poi passare, in una ultima fase, ad interpretare la scena osservata.

Durante il periodo relativo alla short term mobility sono stati principalmente esplorati i passi relativi alla elaborazione delle immagini 2d, con particolare attenzione alla skeletonizzazione delle sagome, essenziale per ricostruire in 3d la postura del soggetto e stabilire delle corrispondenze per la stima della postura model based.

Al fine di identificare l'approccio di skeletonizzazione più adatto, si è lavorato inizialmente con sagome create ad-hoc, in modo da mostrare i migliori risultati ottenibili con i vari ap-

procci, per poi successivamente valutare la degradazione delle performance delle tecniche man mano che le sagome venivano compromesse da rumore, in condizioni simili a quanto ottenibile nella realtà.

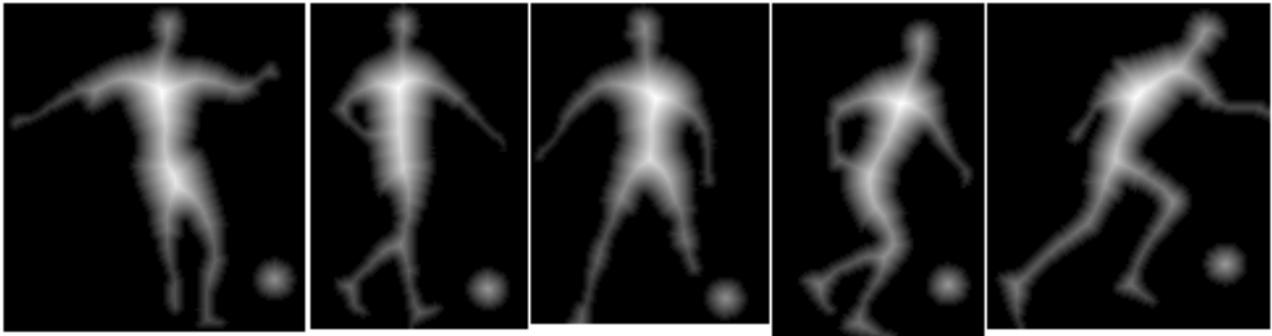
I due approcci di skeletonizzazione messi a confronto sono stati l'operatore morfologico di thinning e le distance transform (di cui la medial axis transform, MAT, identifica lo scheletro interno). Mentre il primo approccio offre sostanzialmente benefici prestazionali, non riesce a garantire dei risultati accurati e risulta eccessivamente sensibile al rumore di fondo. Il metodo MAT, basato sulle distance transform ha sostanzialmente dimostrato una minore sensibilità al rumore e risultati decisamente più accurati. Il risultato delle distance transform su una sagoma, è sostanzialmente una immagine a scala di grigio che mostra come ogni pixel all'interno della sagoma è distante dai bordi della sagoma stessa.



*Sagome usate per la valutazione dei metodi di skeletonizzazione*

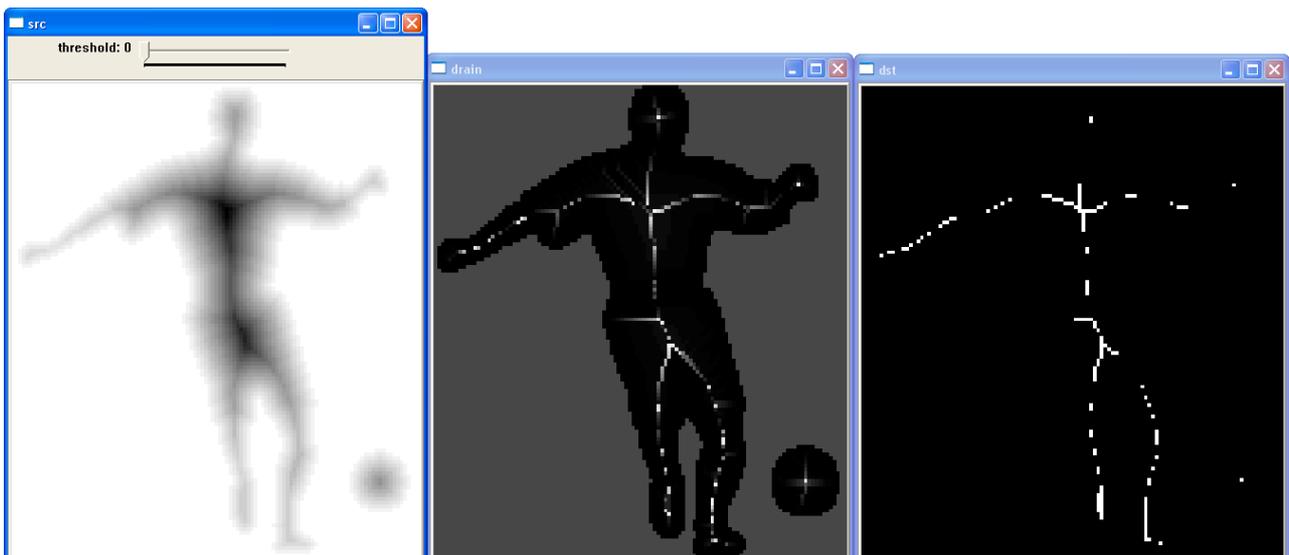
E' successivamente necessario estrarre le caratteristiche dello scheletro 2d ottenuto. Mentre questo compito si rivela sostanzialmente istantaneo visivamente, la computazione delle features 2d al calcolatore richiede un ulteriore passaggio che corrisponde all'identificazione di un set limitato di pixels che possano descrivere lo scheletro più succintamente.

E' stato dunque messo a punto un procedimento basato sul gradient descent. Il meccanismo è intuitivamente semplice: si interpreta l'immagine a scala di grigi come se fosse una mappa altimetrica (i pixels della sagoma corrispondono ai punti più alti, i punti più interni appartengono al fondo della "vallata"). Successivamente si fanno "piovere" delle gocce



*Immagini a scala di grigio risultanti dalla distance transform*

d'acqua sulla mappa altimetrica. Laddove l'acqua si accumulerà, identificherà le zone (i punti) da considerare significativi dello scheletro. Un successivo passaggio di elaborazione trasformerà questa mappa di punti sparsi, in un insieme di segmenti connessi tra di loro.



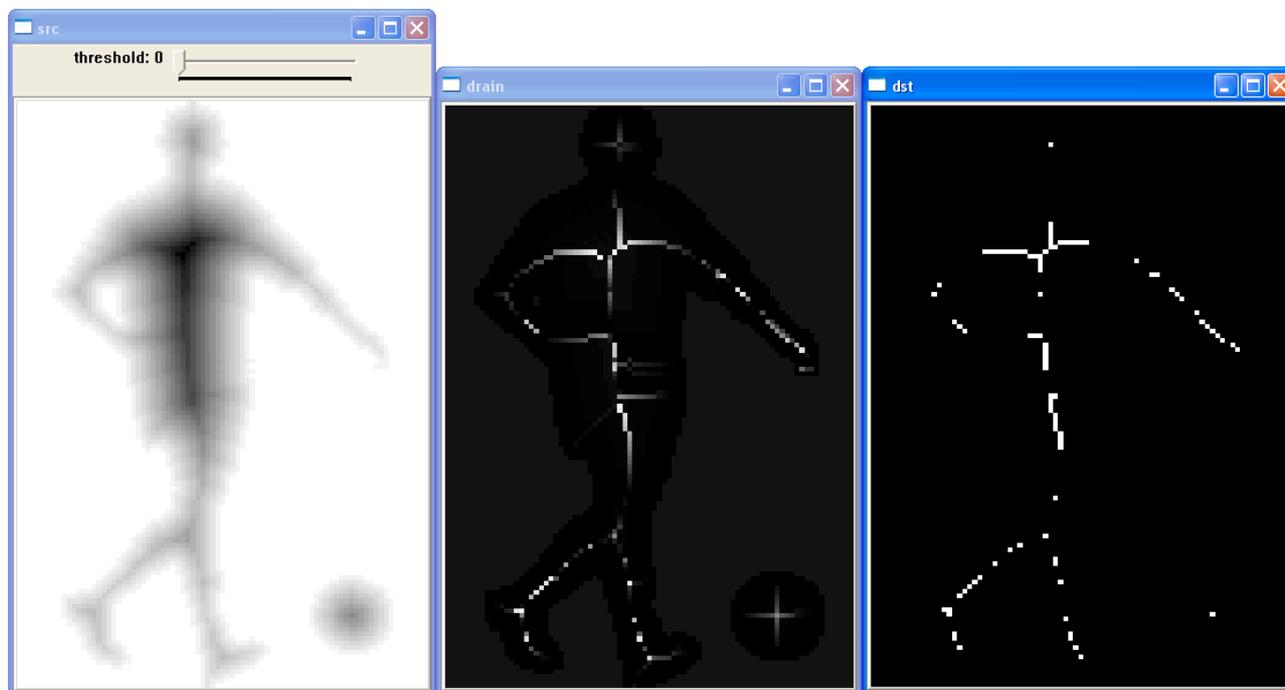
*Esempio di individuazione dei pixel appartenenti allo scheletro della sagoma tramite sistemi partecellari e gradient descent*

Le successive fasi prevederanno:

2. Fusione spazio-temporale di tipo ibrido
  1. Individuazione di un modello umano 3d di riferimento
  2. Stima della postura model based
  3. Stima della postura template based
  4. Confronto e fusione dei risultati

## 3. Interpretazione della scena

1. Analisi posture tramite Hidden Markov Model
2. Interazione tra persone



*Un ulteriore esempio dei risultati ottenibili attraverso la scheletronizzazione 2d proposta*

Il lavoro è attualmente concentrato sulla identificazione delle features significative per il mapping tra più viste bidimensionali per l'identificazione della postura umana in 3d. Per fare questo, si sta ponendo l'attenzione sulla stima della postura model-based cercando di sfruttare il più possibile le conoscenze relative allo scheletro umano (come i gradi di libertà degli arti) in situazioni in cui la scena possa presentare ambiguità interpretative.

Bari, 3 Febbraio 2009

Il fruitore

Dott. Nicola Mosca

Il proponente

Dott. Ettore Stella