RITA CUCCHIARA
ImageLab - Dipartimento di Ingegneria dell'Informazione Università di Modena e Reggio Emilia

VIDEO SORVEGLIANZA PER L'INDIVIDUAZIONE DI PERSONE E L'ANALISI COMPORTAMENTALE

In questo articolo si parla delle nuove frontiere di visione artificiale nella videosorveglianza di persone in ambienti pubblici e privati ed in particolare di analisi comportamentale. Sono poi presentate alcuni progetti in corso presso l'ImageLab di Modena

1. Video sorveglianza oggi

Sono trascorsi più di 40 anni da quando, nel 1969, il primo sistema di video sorveglianza fu istallato al Muncipal Building del City Hall di New York: in quegli anni venivano impiegate le prime telecamere analogiche a CCD, collegate in TV a circuito chiuso a monitor e sistemi VCR di videoregistrazione a cassetta. Queste tecnologie, impiegate per molti anni, hanno sostanzialmente modificato il mondo della vigilanza e hanno permesso la realizzazione di grandi e piccoli centri di controllo. Miliardi di ore di video sono stati trasmessi e memorizzati in tutto il mondo, prima attraverso immagini a bassa risoluzione e a livello di grigio, poi con una visibilità sempre migliore, con l'uso del colore e della sensibilità all'infrarosso per i periodi notturni. La tecnologia analogica è stata via via sostituita da quella digitale sia nell'acquisizione sia nella registrazione delle immagini e dei video.
I sistemi di videosorveglianza oggi, come in Figura 1 consentono l'impiego di telecamere digitali di diversa natura, fisse e brandeggiabili, di DVR (digital video recorder) o di sistemi basati su PC per salvare i dati visuali, di collegamenti wireless o wired con grandi centri di controllo e di possibilità di accesso remoto anche con qualsiasi dispositivo. I concetti di Universal Multimedia Access, ora ben consolidati nel mondo consumer si applicano direttamente anche nella sicurezza e sorveglianza.
La videosorveglianza non è però solo hardware; è sempre più necessario l'impiego di software e di tools complessi per elaborare, gestire e memorizzare le informazioni multimediali connessi alla sorve-

glianza stessa. Da una parte sono indispensabili soluzioni di DBMS e sistemi informativi per la gestione e la comunicazione delle informazioni, dall'altra stanno sempre più emergendo sistemi che impiegano tecniche di Visione Artificiale per l'elaborazione dei dati visuali. In questo caso, non si parla semplicemente di Imaging, ossia delle tecniche per il migliorare il contenuto visivo di immagini e video in ausilio all'operatore umano, ma di Computer Vision (in italiano Visione Artificiale).

La Visione Artificiale è una disciplina informatica nata agli inizi degli anni '80 come campo applicativo dell'Intelligenza Artificiale che si occupa di fornire metodologie e strumenti per l'interpretazione automatica e la comprensione di immagini e video. E' una disciplina complessa, altamente multidisciplinare, che coniuga tecniche ed algoritmi propri dell'analisi del segnale digitale, soluzioni di apprendimento automatico e modelli di ragionamento automatico, algoritmi di pattern recognition statistico, modelli di geometria 3D, competenze di colorimetria e tools di trattamento di dati multimediali, per fornire tecnologie abilitanti a una vastissima gamma di applicazioni, sia web-based che stand-alone, ormai computazionalmente accessibili anche sui moderni pc multimediali o su interfacce palmari e smart-phone.

Le applicazioni "storiche" di visione artificiale riguardavano principalmente il mondo militare e della robotica industriale, i sistemi per elaborazioni di immagini da satellite o astrofisiche, il medical imaging . Negli ultimi decenni invece si stanno sviluppando nuovi campi applicativi, che anche grazie alla disponibilità di librerie open-source e al diffondersi delle competenze nel settore possono diventare appannaggio anche di piccole imprese sia ad uso interno, sia per la creazione di prodotti innovativi facilmente proponibili nel mercato dell'ICT. Tra esse si vogliono citare solo alcuni ambiti applicativi quali l'human machine interaction, i l'augmented reality, il recupero di dati visuali da web e da archivi multimediali e da ultimo certamente la video sorveglianza. Anzi in questo settore, ancora trainante nell'economia mondiale, la ricerca militare ed accademica ha sempre dialogato in modo molto

I sistemi di videosorveglianza di nuova generazione stanno sempre più adottando tecniche e soluzioni di visione artificiale per l'elaborazione in modo automatico di dati visuali, l'interpretazione della scena e l'ausilio al personale addetto alla sicurezza.

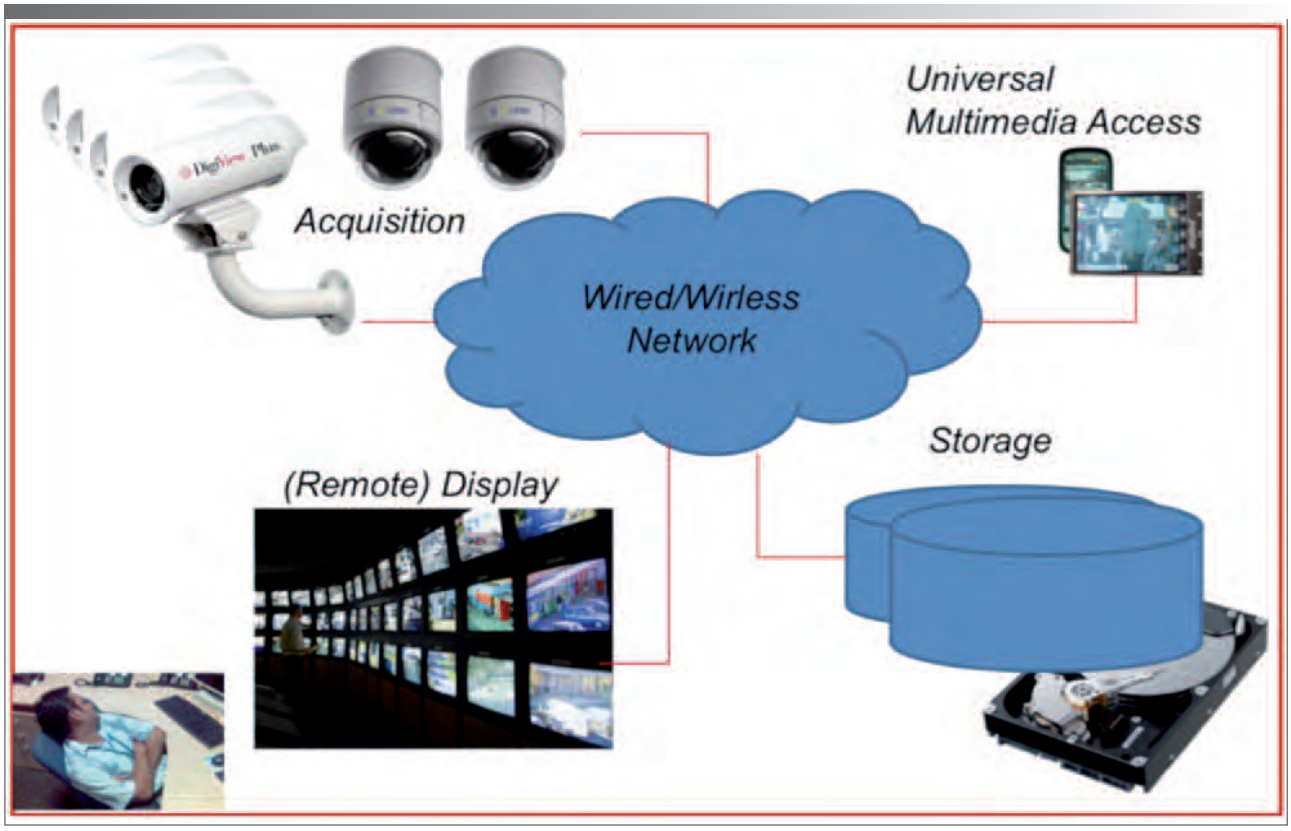

Figura 1 - Sistemi di videosorveglianza

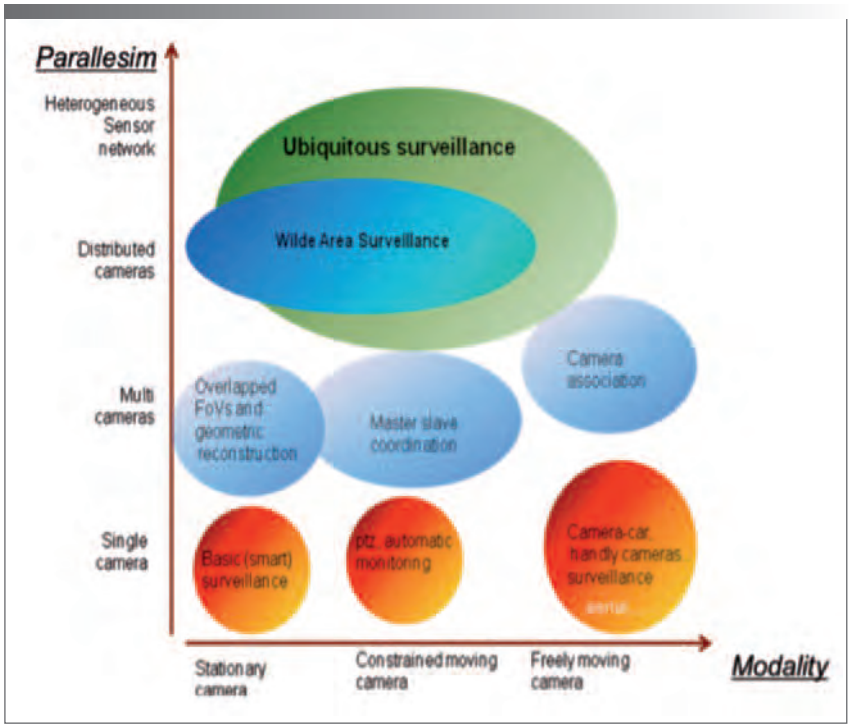

Figura 2 - Evoluzioni dei sistemi di video sorveglianza

stretto con il mondo industriale.

Ora la visione artificiale può offrire molti strumenti per la sicurezza, per l'estrazione in tempo reale di dati dai video acquisiti dalle migliaia di telecamere istallate per la sorveglianza sia di enti pubblici , sia di soggetti privati. Servizi come la lettura targhe, l'estrazione di dati visuali biometrici (impronte digitali, riconoscimento del volto, body scanner), il controllo degli accessi, il monitoraggio perimetrale e la sicurezza ambientale (ad esempio per riconoscimento di fumo ed incendio) sono ora realizzabili con investimenti ragionevolmente limitati.

Le applicazioni più studiate negli anni '90 si orintavano principalmente nel settore dei trasporti, per il controllo del traffico urbano ed extrairbano e nel controllo degli accessi per la sorveglianza perimetrale. Dall'inizio del nuovo millennio sia i progetti di ricerca sia i più innovativi sistemi commerciali stanno focalizzando la loro attenzione sul controllo automatico delle persone in ambienti chiusi ed aperti.

Gli orizzonti della videosorveglianza si stanno allargando verso quella che nel prossimo futuro sarà chiamata "live surveillance" con sensori mobili, distribuiti ed eterogenei. Lo schema di figura 2 vuole mostrare questi nuovi orizzonti nei termini di multi modalità nelle acquisizioni video e di parallelismo. Nel vertice si vedono i sistemi tradizionali basati su singole telecamere fisse, che però possono essere accompagnate da telecamere brandeggiabili (PTZ), ossia con movimento vincolato e programmabile, fino a telecamere totalmente in movimento come quelle in dotazione su autoveicoli, o indossabili da agenti della sicurezza.

Questa multi modalità non crea problemi alla visualizzazione dato che le tecniche di compressione e stabilizzazione del movimento (MPEG, H264..) sono ormai ben consolidate, ma propongono difficoltà sempre crescenti per gli algoritmi di visione artificiale per la segmentazione ed il tracciamento di oggetti mobili e la separazione del moto dell'oggetto dal moto relativo del sensore.

Sono però in atto molti progetti di ricerca in questi campi dell'analisi di movimento con risultati assai promettenti tanto da prevederne un uso commerciale in tempi molto brevi..

Salendo verso il parallelismo si incontrano sempre più sistemi di sorveglianza dove più telecamere sono impiegate con FOVs (Field of Views, campi di vista) sia parzialmente sovrapposti sia totalmente disgiunti. Nel primo caso (multi-camera) come nei sistemi a multiprocessore, i sistemi di visione devono elaborare su una "memoria condivisa" ossia condividendo informazioni visuali. In questo caso impiegando paradigmi di geometria 3D si può anche ricostruire la scena osservata e mantenere una ridondanza delle informazioni.

Quando invece i FOVs sono disgiunti e le telecamere possono coprire anche vaste aree, si parla più spesso di WAN (Wide Area Surveillance) o di ubiquitous surveillance. In questo contesto ogni sistema lavora in modo disaccoppiato e distribuito con scambi di messaggio per trasferire solo informazioni di alto livello dopo l'elaborazione dei video. Un problema in questo momento assai dibattuto, nel mondo della ricerca, è la risoluzione del problema della "re-identification" automatica, che permette di rispondere a domande quali "questa persona che appare è già stata vista da un'altra telecamera e dove?" Il problema non è banale perché l'apparenza visuale cambia non solo prospetticamente ma anche dal punto di vista colorimetrico: le telecamere non sono mai uniformemente tarate e le condizioni visuali ed atmosferiche sono sempre diverse. Si parla perciò di "soft biometry" per indicare l'uso di caratteristiche visuali biometriche non certe (come la tessiture dei vestiti, l'andatura o l'altezza stimata) che pur non avendo singolarmente caratteri di unicità, correttamente interpretate possono portare a questa soluzione utilissima sia nel mondo della sorveglianza in real-time che nel mondo dell'analisi forense.

Per finire la sorveglianza sta sempre più dirigendosi nel settore della multi modalità e multimedialità impiegando non solo telecamere ma reti di sensori distribuiti, come microfoni, GPS, RFID e smart card per gestire in modo congruente dati di natura diversa ma sempre inerenti al campo della sicurezza. Le ricerche nella Data Fusion e nei multi-classificatori diventeranno nel prossimo futuro sempre più vincenti. Un esempio di risultato di una calibrazione multipla tra sensori mobili con telecamere e RFID, prodotto alla Facoltà di Ingegneria di Modena e' mostrato nella figura 5 b.

■ 2. Analisi comportamentale

Ora che i sistemi di visione artificiale assieme a reti di sensori possono individuare con sufficiente precisione la presenza di persone nello spazio è anche possibile iniziare a collezionare informazioni sul movimento delle persone.
Lo scopo finale è quello di ottenere sistemi stabili capaci di inferire il comportamento dei singoli individui, mediante l'analisi delle loro azioni e delle interazioni tra loro e con l'ambiente.
Nuove generazioni di sistemi di video sorveglianza possono ottenere stabili informazioni sulle singole persone, in ambienti non affollati o dove normalmente le persone sono naturalmente o forzatamente "serializzate" come porte, scale mobili, tornelli, stazionamenti davanti a sistemi bancari di pagamento automatico etc. In questi casi le attività che ora vengono analizzate in modo automatico mediante tecniche di Visione Artificiale sono molteplici:

- La postura delle persone e il susseguirsi di pose diverse (posture erette, chinate, accovacciate etc);
- Il modo di muoversi ("gate analysis"), di correre o di camminare;
- Le gestualità ("gesture analysis") soprattutto nelle interazioni con interfacce uomo macchina quali quelle con distributori automatici, benzinai o ATM;
- La traiettoria ossia il modo di procedere nello spazio soprattutto non condizionato da percorsi obbligati.

Attualmente queste analisi sono oggetto di ricerca e di sviluppo di tools che sfruttano metodi di apprendimento automatico e di "Pattern Recognition" soprattutto statistico. Tramite la raccolta di grandi quantità di dati catalogati con comportamenti esemplari, il sistema può diventare in grado di sintetizzare il miglior classificatore di attività e conseguentemente di azioni protratte nel tempo. Azioni ripetute dagli stessi individui o da gruppi di individui nel tempo costituiscono poi pattern di normalità che sono impiegati per inferire comportamenti tipici o anche comportamenti anomali o in generale significativi.

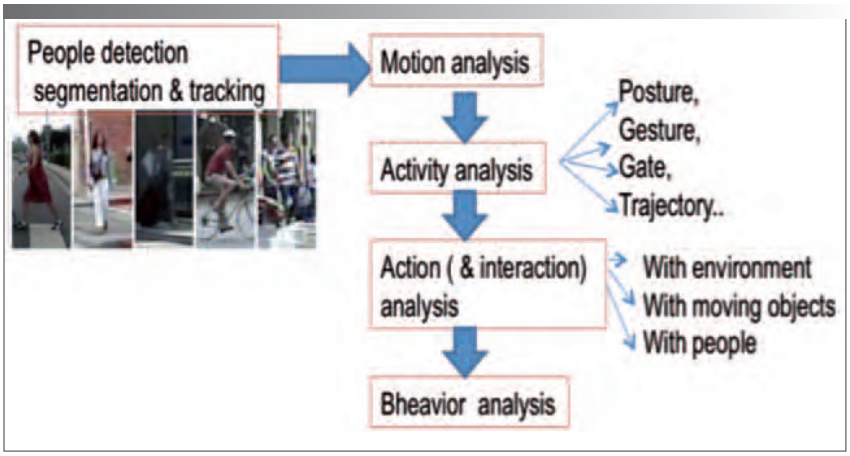

Figura 3 - I passi dell'analisi comportamentale

Le differenti attività che possono essere monitorate dipendono fortemente dalla applicazione di riferimento. Se lo scopo è osservare il comportamento di persone davanti ad uno sportello bancomat può interessare soprattutto la postura e la gestualità se non l'espressione del viso, argomento di interesse di una branca emergente della pattern recognition chiamata "affective computing".
Se invece si vuole osservare il comportamento di persone in spazi aperti come centri commerciali o aeroporti, la traiettoria e il modo di muoversi può diventare determinante.
Diverso è il caso in cui si voglia analizzare il comportamento di persone in zone affollate. In questo caso la ricerca scientifica non ha ancora dato risultati stabili per l'analisi dei singoli individui in code e affollamenti (cosa difficile anche ad un operatore umano), ma è in atto una grande attività di ricerca prototipale per lo studio di parametri di movimento e di flussi di persone per il riconoscimento di situazioni anomale quali creazione di code, deflusso di gente tramite gate, persone in direzioni anomale rispetto alle masse per ottenere in generale una analisi comportamentale della folla e non del singolo individuo. In questo caso sono stati prodotti molti studi sia in ambienti chiusi come metropolitane e stazioni sia in ambienti aperti: in particolare vengono spesso analizzati video di maratone e di altre manifestazioni sportive.
Ci sono molti progetti in atto per entrambi i tipi di analisi comportamentale: tra essi, il progetto europeo THIS (Transportation Hub with Video Intelligent Systems) finanziato dal 2010 al 2012 dalla comunità Europea (nella Divisione JLS Justice Legalitee Securite). Il consorzi di THIS vede la presenza di due aziende italiane, Bridge129 srl e Vitrociset spa e di 4 centri di ricerca europei: oltre al centro interdipartimentale per la sicurezza, CRIS dell'Università di Modena e Reggio Emilia, che coordina il progetto, lavoreranno insieme centri di ricerca di Barcellona, Budapest e Atene.

■ 3. Analisi di traiettorie

L'analisi del comportamento può iniziare dalla elaborazione di dati altmente significativi quali le traiettorie e i pattern di movimento. Nella figura 4 è mostrato il risultato di una sperimentazione in corso presso l'ImageLab dell'Università di Modena e Reggo Emilia.
Quest'attività nata inizialmente nell'ambito del progetto BESAFE finanziato dalla NATO (progetti "Science for Peace"), ora continuerà nel citato progetto THIS. Lo scopo è di analizzare in modo automatico le traiettorie delle persone che attraversano uno spazio aperto al fine di verificarne consuetudini ed anomalie.
Come detto, le potenzialità applicative di queste soluzioni informatiche sono estremamente interessanti. Innanzitutto nell'ambito della sicurezza possono essere impiegate come sup-

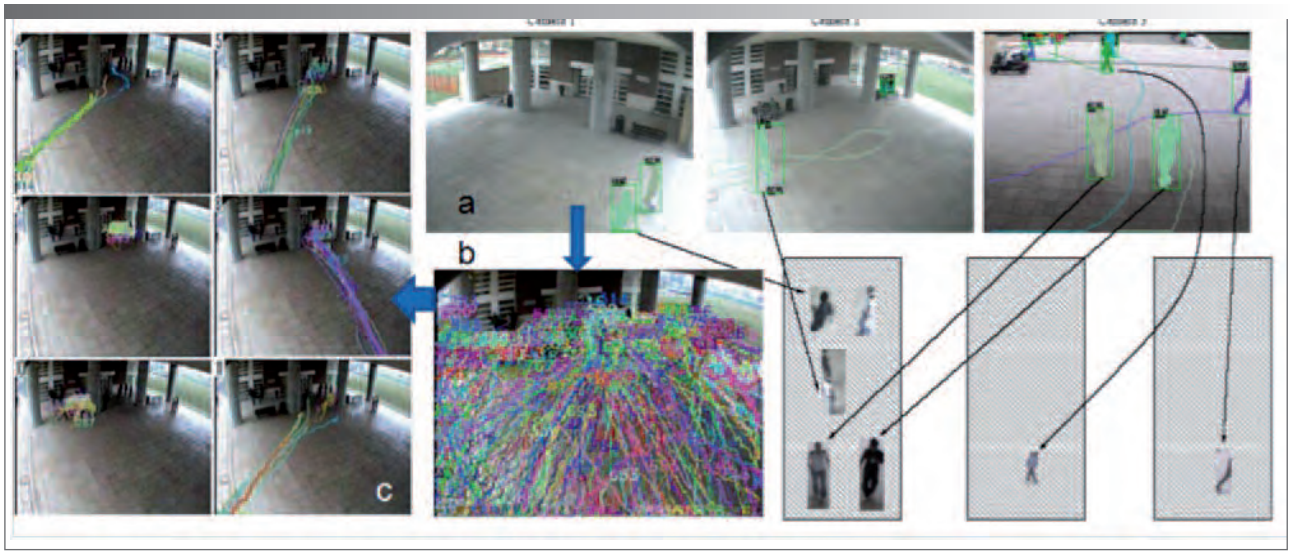

Figura 4 - Un esempio di analisi di traiettorie: a) segmentazione e tracking di persone da più telecamere con FOV sovrapposti; b) un insieme di 1000 traiettorie tracciate nel tempo; c) un esempio di classificazione delle traiettorie in pattern di similarità per analizzare i comportamenti frequenti e quelli anomali.

porto all' analisi forense per studiare comportamenti anomali (nel senso di non frequenti), interazioni tra persone e possibili connessioni tra individui. Si tratta di nuovi strumenti molto sofisticati di Data Mining su dati visuali. Se impiegate in sistemi in tempo reale, collegate ai sistemi di video sorveglianza permetteranno di individuare l'incorrere di situazioni di potenziale pericolo o interesse per le forze dell'ordine. Queste tecniche possono essere impiegate anche per applicazioni "civili" non legate alla sicurezza quanto all' analisi statistica del comportamento di persone, come lo studio dei percorsi tipici di utenti in musei o in centri commerciali, o la fruizione di servizi pubblici in parchi e in ambienti aperti.

Questi studi rappresentano lo stato dell'arte della ricerca scientifica nell'ambito della sorveglianza di persone. Nel caso specifico, le persone vengono identificate mediante metodi di "moving object detection" basati sulla analisi delle differenze tra l'immagine corrente presa da telecamera fissa e una immagine di riferimento (Background) ottenuta statisticamente e aggiornata nel tempo per tener conto dei cambi di luminosità, delle condizioni atmosferiche etc. Le parti diverse dallo sfondo, vengono poi classificate tra oggetti apparenti, ombre, artefatti vari (come fumo, riflessi etc) e veri oggetti in movimento che a seconda del contesto possono poi essere validati come persone veicoli o altri oggetti.

Queste tecniche, note con il nome di "background suppression", sono ormai consolidate , ne esistono diverse soluzioni alcune delle quali anche disponibili su librerie open-souce come le OpenCV, e sono implementate su molti sistemi commerciali. Esse sono efficaci quando il grado di affollamento non è particolarmente elevato per poter discriminare singoli o gruppi di individui. Per potere eseguire il calcolo della traiettorie è necessario compiere un processo di "tracking" ossia di inseguimento nel tempo per associare lo stesso identificatore sempre alla stessa persona anche in presenza di più persone che si occludono tra loro e di parziali perdite visuali di un singolo individuo nel tempo. Ciò può accadere in molti casi, sia per un errore nel sistema di segmentazione sia nel caso di reali occlusioni come quando, ad esempio, una persona passa dietro ad una colonna.

Poi è necessario un'opera di ricostruzione del mondo 3D a partire dal piano immagine per potersi riferire al "sistema mondo" ed estrarre l'informazione corretta della posizione della persona. Spesso non è necessaria una vera e propria ricostruzione 3D ma è sufficiente riportare le immagini ad alcuni punti di riferimento stabili, come ad esempio sul piano del pavimento ottenuto tramite una trasformazione omografica.

Questa attività diventa più complicata se si vuole costruire la traiettoria di una persona che attraversa i campi di vista (FOVs da Field of Views) di più telecamere come nelle immagini in figura 4 in cui tre telecamere osservano parti della stessa scena. Calibrando le telecamere nel 3D o almeno eseguendone una corretta omografia e' possibile mantenere la traccia delle persone e anche memorizzarne l'aspetto ripreso da più telecamere. Questo potrà servire per migliorare un possibile sistema di riconoscimento automatico.

In figura 5 si vede un esempio di ricostruzione omografica dello spazio coperto da due telecamere dove il piano immagine di una telecamera viene mappato (distorto) nello spazio immagine della seconda. La persona si vede rappresentata due volte, come viene ripresa effettivamente dalle due telecamere.

Grazie a questi sistemi si potrà mantenere in un database, aggiornato frame dopo frame, le informazioni non solo di tutta la scena ma anche dei singoli protagonisti: il loro aspetto fi-

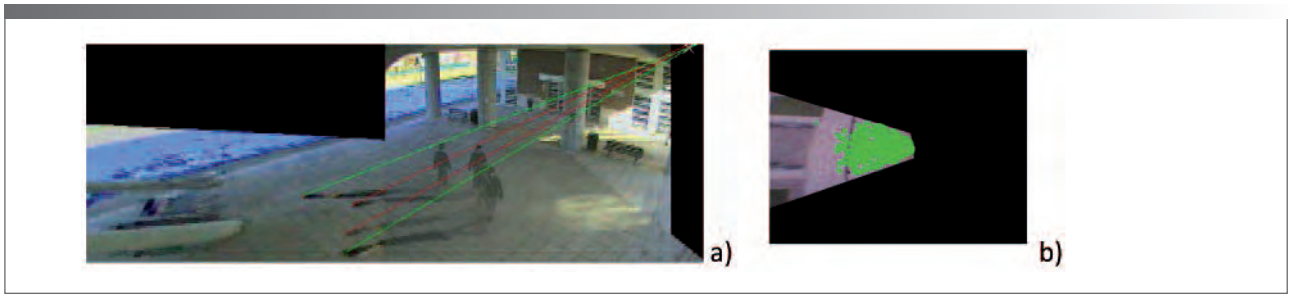

Figura 5 - Esempio di a) omografia e ricostruzione epipolare da due telecamere con FOVs parzialmente sovrapposte e b) di calibrazione nello spazio omografico tra telecamere e RFID

sico, la loro altezza o altre caratteristiche biometriche (si veda ad esempio figura 4), il volto se il sistema e' dotato di specifiche telecamere PTZ coordinate, la traiettoria sul pavimento, la velocità, la accelerazione e molte altre informazioni visuali.
Questi dati, che nell'ambito della Pattern Recognition vengono denominate "Features", sono poi soggetti a processi di classificazione supervisionata o non supervisionata (clustering) per inferire conoscenza sulle azioni e attività dei singoli individui. Tali azioni se protratte e ripetute nel tempo e correlate alle situazioni specifici (ad esempio a quando vengono svolte in concomitanza di quali altri eventi) possono poi portare ad una interpretazione dei comportamenti frequenti ed anomali.
Sempre nell'esempio di figura 4 l'immagine centrale (b) mostra visualmente la raccolta di un migliaio di traiettorie. Ogni traiettoria e' descritta come una sequenza di punti nel tempo, a cui può essere associata la velocità o altre informazioni scalari. Le traiettorie sono poi confrontate tra loro per ottenere gruppi, o cluster, di similarità. Le tecniche qui impiegate sono molto vicine a quelle nate per sistemi di similarità di dati visuali nelle ricerche sul Web, o in altri sistemi di content-based information retrival. Un aspetto complicato è dato dal fatto che le traiettorie non sono mai della stessa lunghezza sia per errori, sia per il rumore insito nel sistema e per la impredicibilità del movimento delle persone. Per questo in questi prototipi si usano tecniche di allineamento dei dati e di "inexact matching" che si usano nell'analisi di stringhe e recentemente in bioinformatica per l'analisi di sequenze di genoma.

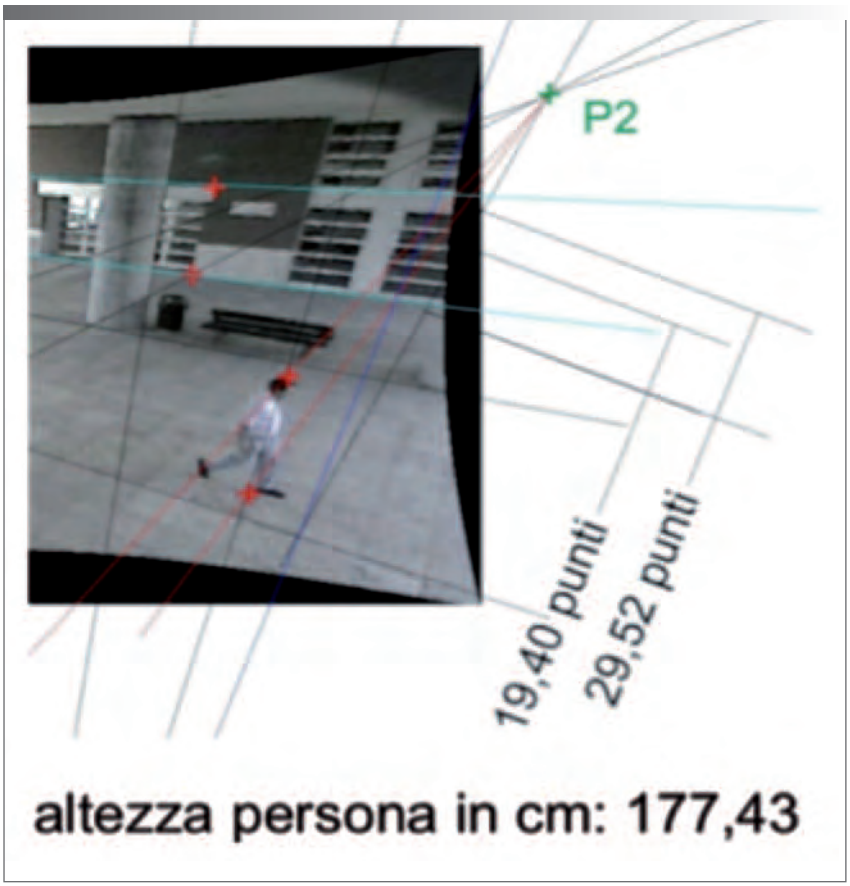

Figura 6 - Attraverso la correzione della distorsione delle lenti, e alla calibrazione si può rimuovere la distorsione e avere informazioni metriche su oggetti nella scena.

Si può comunque andare ben oltre: anzichè osservare la posizione delle traiettorie, possiamo osservare le deviazioni angolari e cioè la forma della traiettoria nello spazio e nel tempo (con la velocità) per trovare comportamenti simili anche se non protratti esattamente nello stesso punto dello spazio. Ad esempio due persone che girano intorno ad una panchina o a una colonna fumando una sigaretta hanno una forma di traiettoria simile indipendentemente dalla posizione della panchina o della colonna stessa.
Naturalmente ciò che è stato descritto fino ad ora è il risultato di un'attività di ricerca scientifica, che porta a sistemi prototipali da ingegnerizzare nel tempo. Ma come ingegnere posso affermare che tra la teoria e la realizzazione di un sistema commerciale sufficientemente generalizzato per l'analisi di azioni di persone non passerà un tempo troppo lungo. ■