



# LA VISIONE ARTIFICIALE PER LA VIDEO SORVEGLIANZA

Rita Cucchiara

Gli studi di Visione Artificiale e Pattern Recognition propongono soluzioni sempre più sofisticate per essere applicate alla Video Sorveglianza, un settore emergente nell'ICT. I sistemi di Video Sorveglianza, infatti, non richiedono solo un'efficiente supporto hardware per l'acquisizione e la gestione dei video, ma anche tool software con algoritmi efficaci per l'analisi dei dati in tempo reale. In questo articolo si presentano gli aspetti della Video Sorveglianza e delle tecniche di Visione Artificiale necessarie per diverse applicazioni. Infine vengono illustrati alcuni progetti di ricerca italiani.

3-7

## 1. VIDEO SORVEGLIANZA

La *Video Sorveglianza* è quel settore dell'ICT che si occupa dell'acquisizione, elaborazione e gestione in tempo reale di video provenienti da telecamere installate in luoghi pubblici e privati per molteplici scopi, comunque correlati al concetto di *sicurezza*. Nei Paesi anglosassoni si parla di "*security & safety*", coinvolgendo quindi sia la sfera pubblica, per la lotta alla criminalità e la prevenzione di reati, sia la sfera privata per garantire l'incolumità e la tranquillità personale. Rientrano nella grande categoria di sistemi di Video Sorveglianza tutti gli apparati installati nel contesto urbano (piazze, strade e monumenti), nei luoghi affollati (stazioni, metropolitane, aeroporti, stadi), nelle banche, nelle abitazioni private e nei siti industriali, nonché i sistemi per il monitoraggio del traffico (veicolare, aereo, navale) e quelli per il controllo ambientale (incendi, alluvioni, frane). È indubbio che la Video Sorveglianza sia per il mondo ICT una delle sfide applicative più interessanti degli ultimi anni. La sua progettazione

coinvolge l'*hardware* per i sensori e i sistemi di acquisizione, l'architettura dei sistemi di calcolo, l'infrastruttura di comunicazione *wired* e *wireless*, il *middleware* e soprattutto la sfera algoritmica e cognitiva di competenza della *Visione Artificiale* e della *Pattern Recognition*.

Sebbene le soluzioni architetture e hardware siano ormai ben consolidate, moltissimi problemi rimangono aperti nell'analisi automatica dei video, dove la ricerca scientifica sta investendo al massimo.

La **Visione Artificiale** (*Computer Vision*) fino a pochi anni fa era bollata come un vezzo accademico, una disciplina di nicchia al più assimilata a una sotto-disciplina dell'*Intelligenza Artificiale* (IA). Con l'IA, la Visione Artificiale condivide molti dei meccanismi inferenziali e strumenti per la rappresentazione della conoscenza. Tuttavia, a differenza dell'IA, la computazione oltre che del livello simbolico si occupa del basso livello, spesso chiamato livello "iconico". In altre parole, i processi di *Computer Vision* non possono prescindere dall'essenza dell'input che è fatta di sequenze di

pixel, le cui informazioni visuali sono per natura parziali e corrotte da rumore sintattico (a livello di segnale) e semantico (i cosiddetti "distrattori" nella scena).

La Visione Artificiale si avvale fortemente della **Pattern Recognition**, la disciplina che studia tecniche, spesso statistiche, di classificazione e riconoscimento di forme a partire da dati misurati o calcolati, siano essi visuali, audio, testuali, multimediali in genere o numerici. Ora la Visione Artificiale e la Pattern Recognition sono ben affermate nel mondo dell'ICT e sono materie di studio in molti corsi di laurea di ingegneria informatica e di informatica. Due ottimi testi su tali discipline sono [1, 2].

Questo articolo vuole introdurre i fondamenti della Video Sorveglianza basata sulla Visione Artificiale, mostrando le esigenze hardware e software per i sistemi della prossima generazione. Tali sistemi sfrutteranno a pieno tecniche e algoritmi per la percezione, l'apprendimento automatico (*machine learning*), modelli di attenzione e di ragionamento su im-

magini (*cognitive vision*), tecniche di *clustering* e classificazione (*pattern classification*) e tecniche inferenziali proprie dell'IA, per fornire allarmi in modo automatico, comprendere la scena e il comportamento di persone e oggetti di interesse, e annotare i dati visuali.

**Video Analytics** è il termine che sta emergendo per denotare l'insieme di queste tecniche, sia per i sistemi in tempo reale, sia per i sistemi *off-line* di analisi di dati multimediali legati alla sicurezza: in questo caso la Video Sorveglianza si estende a discipline limitrofe quali il *content-based video retrieval* per il recupero per similarità di dati visuali o il *video mining* per la ricerca di associazioni e correlazioni e l'estrazione di conoscenza dal video.

Si apre la strada ad applicazioni innovative per il supporto all'indagine forense o all'intelligence, e per mercati emergenti nel campo dell'ICT quali l'identificazione di comportamenti in ambienti sociali, pubblici e privati, supermercati o musei. Alcuni esempi di possibili applicazioni sono riportati nella figura 1.



**FIGURA 1**

Applicazioni di video sorveglianza. Dall'alto a sinistra: tracking di persone davanti a vetrine (video di PETS'2002); individuazione di veicoli fermi; rilevazione di pacchi abbandonati (PETS'2006); riconoscimento della testa e del volto da telecamera PTZ. Per informazioni si veda <http://imagelab.ing.unimore.it>

## 2. VIDEO SORVEGLIANZA E L'OPINIONE PUBBLICA

Tutti ora parlano di Video Sorveglianza, sui giornali, TV e Web, spesso intendendo in forma molto riduttiva solo le centinaia di telecamere installate nelle nostre città, nelle strade e negli uffici collegati a centri di controllo dove le immagini dovrebbero essere osservate con continuità da occhi attenti di addetti alla sicurezza. Ovviamente ciò non accade e questi sistemi installati, le cui immagini non sono controllate né da operatori umani né dal calcolatore, hanno un effetto preventivo e reattivo di fatto assai limitato. Le telecamere hanno comunque una funzione deterrente e possono fornire un validissimo aiuto come supporto alle indagini forensi a posteriori.

Ciò nonostante, l'interesse e l'aspettativa dell'opinione pubblica in Italia è decisamente altalenante: da un forte contrasto si è passati al plauso incondizionato e viceversa. Dieci anni fa, mentre negli Stati Uniti tali sistemi si stavano capillarmente distribuendo, in Italia spesso si assolveva la carenza di infrastrutture e denaro pubblico con una legittimazione del garantismo ad oltranza supportato da un dibattito serrato sui potenziali problemi di privacy. Basta ricordare l'abuso di citazioni al grande fratello del "1984" di Orwell e a bui scenari di occulto controllo dei cittadini che presupporrebbero una rete di intelligence e una tecnologia di distribuzione in tempo reale dell'informazione fino a pochi anni fa impensabile.

Negli ultimi anni, invece, e soprattutto negli ultimi mesi, la diffusa sensazione di insicurezza nei luoghi pubblici e nei centri urbani sembra aver fatto dimenticare il problema della privacy e si richiedono a gran voce nuove e capillari installazioni di telecamere come placebo ad ogni illegalità. Sui giornali locali non manca ogni giorno una richiesta di nuove telecamere per la sicurezza dei cittadini, a cui si associa la richiesta dalle forze dell'ordine delle telecamere per il controllo del traffico e delle infrazioni stradali (semafori, velocità ZTL).

Le telecamere installate in Italia solo dagli enti pubblici sono migliaia, centinaia a Roma e a Milano e numerose anche nelle città di provincia se si pensa che nel 2008 il record sembra essere di Reggio Emilia con una telecamera ogni 350 abitanti.

Si diceva però opinione altalenante: infatti

proprio recentemente è stata enfatizzata sui giornali italiani una notizia sull'inutilità dei sistemi di video sorveglianza.

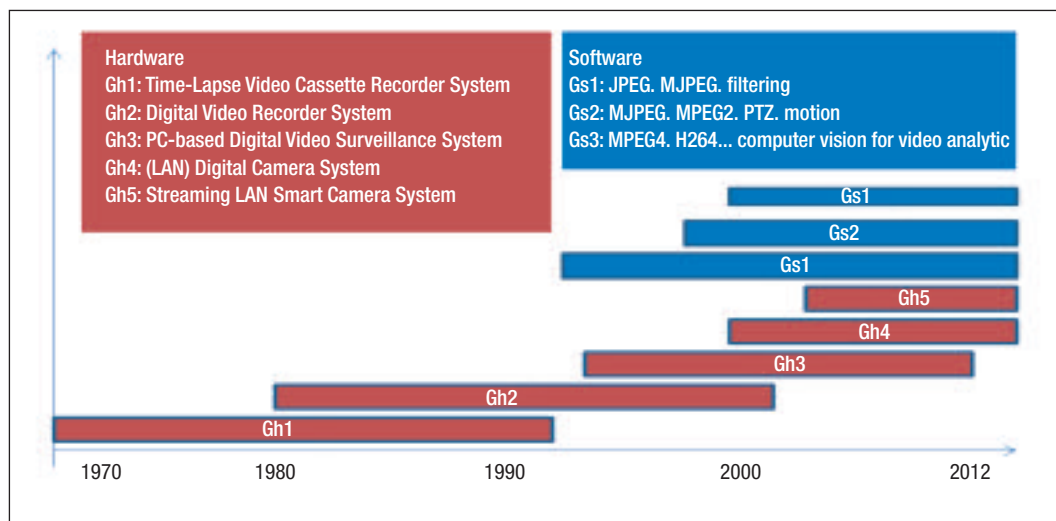
In un'intervista al *The Guardian* del maggio 2008 Mick Neville, ispettore di Scotland Yard, ha bollato la Video Sorveglianza come un "utter fiasco", affermando che l'Inghilterra a fronte dell'enorme diffusione degli apparati (4,2 milioni di telecamere con un rapporto 1 a 14 tra telecamere ed abitanti) e dei miliardi di sterline impiegati, ha sistemi decisamente inefficaci; ha stimato solo in un 3% il miglioramento nella lotta alla criminalità grazie alle telecamere.

Il problema è quello prima richiamato: le forze dell'ordine non possono materialmente osservare in tempo reale centinaia di migliaia di video e l'efficacia di tali sistemi nel momento in cui il fatto si verifica è di fatto nulla. L'ispettore al termine dell'intervista però assolve i sistemi del futuro e ipotizza come unica soluzione plausibile l'affidarsi al software di analisi di immagini per elaborare in tempo reale i video e fornire allarmi.

Da anni la comunità scientifica internazionale sta lavorando sulla Visione Artificiale per trovare soluzioni all'interpretazione della scena in modo automatico. I primi progetti di ampio respiro provengono dagli Stati Uniti e sono stati finanziati dal DARPA alla metà degli anni '90, come i progetti VSAM per l'inseguimento di persone e MugSpot per il riconoscimento del volto. Piano piano anche nel mondo commerciale la proposta di tools di "Intelligent Surveillance" o "smart Surveillance" o "Surveillance with Video Analytics" diventa migliore. Con l'avvento di queste tecnologie l'ente inglese di analisi di sicurezza IMS Research ipotizza una crescita del 65% annuo dal 2004 ad oggi e predice un valore del mercato di 215 milioni di sterline nel 2009 solo per la parte software di *video analytics*.

## 3. ARCHITETTURA DI UN SISTEMA DI VIDEO SORVEGLIANZA

I sistemi di Video Sorveglianza non possono più essere considerati solo il supporto elettronico per l'acquisizione e memorizzazione di video. Essi invece hanno un'architettura complessa, sia hardware che software.



**FIGURA 2**  
Evoluzione delle tecnologie hardware e software per sistemi di video sorveglianza

Per quel che riguarda la parte hardware, si è assistito ad una grande evoluzione negli ultimi cinquant'anni. Il primo sistema publicizzato risale al 1969, installato al Municipal Building del City Hall di New York: in quel tempo i sistemi di prima generazione (Gh1 nella Figura 2) erano costituiti da telecamere CCTV analogiche collegate a VCR (*Video Cassette Recorder*) e monitor. Questa tecnologia è rimasta in auge per più di vent'anni ed è ancora in uso, per la falsa credenza che le video cassette prelevate fisicamente solo da personale autorizzato siano più sicure di altri supporti digitali. La seconda generazione (Gh2 nella Figura 2) ha visto l'impiego di tecnologia digitale nella registrazione: le telecamere analogiche CCTV, solitamente a livello di grigio, venivano collegate a video switch o multiplexer e trasferite a DVR (*Digital Video Recorder*) per salvare qualche immagine selezionata, solitamente in JPEG. Il video switch permetteva quindi sia una visualizzazione analogica sia una registrazione digitale. Questi sono tuttora i sistemi a basso costo in uso in sedi dove non è richiesta un'interazione con il servizio di sorveglianza (per esempio, banche). Una grande trasformazione si è verificata quando le tecnologie informatiche hanno soppiantato quelle puramente elettroniche, impiegando il computer (Gh3 nella Figura 2), sostituendo al multiplexer e al DVR, un PC con video capture card capace di comprimere immagini e video su hard disk. L'aver un PC anche industriale permetteva inoltre un possibile collegamento remoto nonché una so-

stituzione dei monitor CCTV con monitor digitali. Sistemi di questo tipo furono impiegati massivamente: un primo esempio risale all'installazione sempre a New York attorno all'area del World Trade Center dopo il primo attacco di febbraio del 1993. Purtroppo inutilmente. I sistemi di questo tipo, tutt'oggi assai venduti hanno tipicamente 1-16 telecamere collegate, compressione in modo *intra-frame only* (in MJPEG) e con risoluzione e *frame-rate* bassi dipendenti dal numero di telecamere collegate e dalle schede di *frame-grabbing*. Tipicamente questi sistemi visualizzano una sola o poche telecamere per volta a risoluzione completa e a colori mentre registrano al 25% della risoluzione (esempio, formato CIF 320 x 240) a 25 o 30 frame al secondo complessivamente per tutte le telecamere. Questi sistemi sono ottimali per la visualizzazione di piccole installazioni private (negozi, uffici); spesso però sono impiegati anche in grandi centri come nei centri di polizia, ma hanno una qualità ben poco accettabile per l'analisi forense, dando a disposizione pochi frame (esempio, 3 al secondo con 10 telecamere) e a una risoluzione così bassa che spesso rende impossibile l'identificazione degli individui o delle targhe. I sistemi più recenti (Gh4 nella Figura 2) adottano telecamere digitali, collegabili anche alla LAN con tecnologia IP e con processori RISC o con FPGA (*Field Programmable Gate Array*) a bordo che permettono un forte risparmio nell'infrastruttura. Queste telecamere hanno spesso la possibilità di essere controllate in re-

moto e brandeggiate in pan, tilt o zoom (PTZ) e sono spesso impiegate per larghe distanze come ad esempio negli stadi. Le telecamere sono collegate a uno *switch ethernet* che le collega a centri di controllo e di registrazione remoti. In larghe installazioni urbane, “foreste di telecamere” sono collegate in fibra ottica a centri di acquisizione con una topologia a stella da cui i dati sono trasferiti a centri di controllo.

Infine le ultime soluzioni hardware (generazione Gh5) impiegano *smart cameras* con RISC e/o DSP a bordo per permettere una miglior compressione video anche in streaming (si parla di *streaming lan cameras*) con tecnologie MPEG2, MPEG4 o H264; esse riassumono tutte le capacità precedenti, come il controllo PTZ, necessitano di PC per la decompressione video e visualizzazione o per la memorizzazione. L'uso di IP e di streaming naturalmente implica tutti i problemi di sicurezza legati ad internet e quindi non sempre vengono installate come dovrebbero, se non in intranet o in reti sicure.

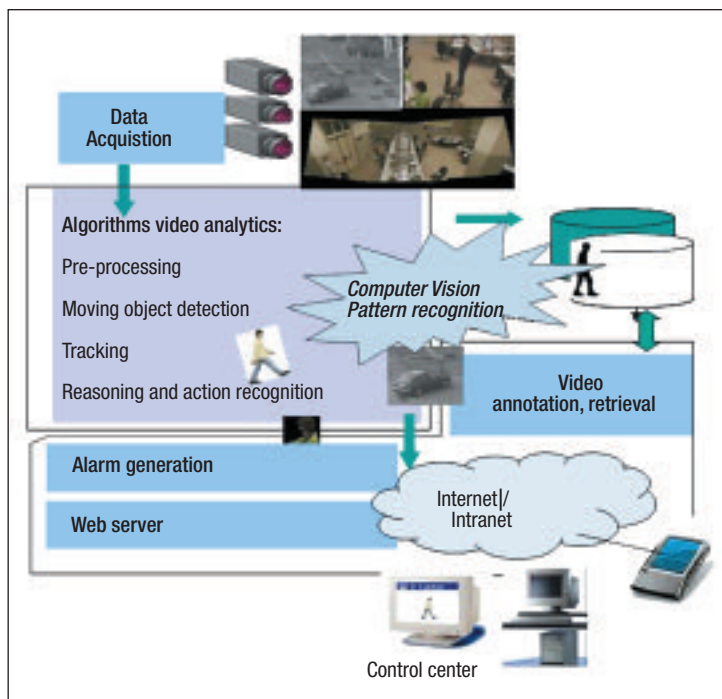
L'impiego di processori, sia a bordo delle telecamere, sia nel sistema di raccolta dei flussi video ha aggiunto la possibilità di un'elaborazione software delle immagini *on-the-fly* lasciando campo aperto all'uso di tecniche più o meno sofisticate di Visione Artificiale.

Anche per il software si può parlare di diverse generazioni commerciali.

Nella prima generazione (Gs1), con tecnologie hardware dalla terza generazione in poi, l'impiego di software era assai limitato, solo a qualche filtraggio digitale per il rumore e alla compressione.

Nella seconda generazione (Gs2) che comprende la quasi totalità di sistemi sul mercato, vengono adottati alcuni semplici algoritmi di elaborazioni di immagini per una buona compressione video e per il “*motion detection*”, di cui se ne parlerà nella prossima sezione.

La nuova generazione (Gs3) che si sta affacciando ora sul mercato ma che probabilmente lo saturerà per il prossimo quinquennio, adotterà in modo massivo la Visione Artificiale, impiegando *tools* per il tracciamento di oggetti in movimento, il riconoscimento di azioni di interesse, la registrazione e la trasmissione selettiva anche su *mobile*, e nuove interfacce multimediali per l'annotazione e il *retrieval*. Nella figura 3 è riportato uno schema completo dove la parte centrale è basata su algoritmi e tecniche di Visione Artificiale e Pattern Recognition. I dati così elaborati generano allarmi automatici anche via web e sono raccolti in un DBMS accessibili anche remotamente e a posteriori. Ora molti colossi aziendali offrono soluzioni di questo tipo come IBM con il sistema S3 [3] o Siemens [4], ma anche molte piccole aziende si stanno muovendo: tra esse citiamo la Wavelet Technology Italia<sup>1</sup> per il sistema realizzato a Reggio Emilia. Stanno inoltre entrando sul mercato *software house* che si occupano solo della parte software, impiegando hardware general-purpose. Molte di queste aziende sono nate come spin off universitari, come l'Object Video<sup>2</sup> americana ora leader mondiale nel settore.



**FIGURA 3**  
Architettura di un sistema di videosorveglianza

#### 4. VISIONE ARTIFICIALE E PATTERN RECOGNITION PER LA VIDEO SORVEGLIANZA

La Visione Artificiale è quella disciplina che si occupa di dare al calcolatore capacità percettive e visive simili a quelle biologiche, almeno in applicazioni specifiche. Si avvale forte-

<sup>1</sup> <http://www.wti.it>.

<sup>2</sup> <http://www.objectvideo.com>.

mente di competenze di geometria, di ottica e di calcolo numerico per definire modelli computazionali al fine di riconoscere il mondo 3D a partire da immagini 2D. È strettamente legata alla *Pattern Recognition*, che si occupa di tecniche di *data analysis*, soprattutto statistiche ma anche sintattiche e strutturali (o *graph-based*) per ragionare sui dati estratti dalle immagini, avendo come fine soprattutto la classificazione supervisionata o non supervisionata. Entrambe le discipline hanno specifiche tecniche e modelli per la Video Sorveglianza, per elaborare i video e riconoscere la scena in tempo reale, e, con vincoli computazionali più rilassati, per analizzare *off-line* i dati relativi alla sicurezza.

Non è possibile fare una trattazione completa di tutte le tecniche impiegate, ma in questo contesto si vogliono piuttosto ricordare i passi fondamentali che un sistema di Video Sorveglianza di nuova generazione deve prevedere.

#### 4.1. Pre-processing

In una fase iniziale tecniche di *image processing* sono sempre impiegate per il miglioramento delle immagini digitali. La teoria dell'*image processing* è ben consolidata. Per un testo di riferimento si veda per esempio, [5]. Anche in base ai sistemi di acquisizione, sono necessarie operazioni di filtraggio, di *noise reduction*, di *image restoration* e di *quality enhancement*. Si usano tecniche di *camera motion compensation* per eliminare le piccole fluttuazioni delle telecamere. Con telecamere PTZ si applicano algoritmi di registrazione per la creazione di mosaici (*mosaicing*), come riferimento dello sfondo (Figura 4).

Nel caso di sistemi con più telecamere e con campi di vista (Field of View, FoV) sovrapposti, si opera la calibrazione manuale o automatica per ottenere la ricostruzione 3D. Si può anche sfruttare una calibrazione parziale per costruire l'omografia e premettere a tutte le telecamere di sovrapporre la loro visuale. Nella figura 5, per esempio, si vede l'output di un sistema con due telecamere con FoV sovrapposte e con la trasformazione omografica del piano immagine di una telecamera nel piano dell'altra. La corrispondenza è ulteriormente garantita dalle linee epipolari che passano per il punto più alto della testa [5].

#### 4.2. Moving object detection

Il passo iniziale consiste nell'estrazione degli oggetti di interesse, che di solito sono, o sono stati in frame passati, in movimento. Sono perciò necessarie tecniche di analisi del moto: o viene stimato il moto di ogni punto dell'immagine, poi raggruppato in cluster omogenei per estrarne forme con la stessa direzione di moto, o viene direttamente effettuata la segmentazione di gruppi di pixel adiacenti caratterizzati da movimento. Nel primo caso si utilizzano tecniche di stima del flusso ottico (*optical flow*) su ogni punto o su punti di interesse quali angoli o speciali *feature* come i SIFT (*Scale Invariant Feature Tracking*) [6]. Questi metodi più complessi si impiegano solitamente in caso di telecamere mobili dove altre tecniche non sono possibili. In caso invece di telecamere fisse, si usano metodi differenziali di *change detection*, che impiegando differenze tra frame o solitamente differenze con frame di riferimento (*background subtraction*) estraggono tutti i punti che sono diversi dallo sfondo e quindi in movimento. In questo contesto c'è una fervente attività di ricerca internazionale per studiare le tecniche più efficienti e reattive per modelli di *background* che tengano conto non solo degli oggetti in movimento ma anche del fat-

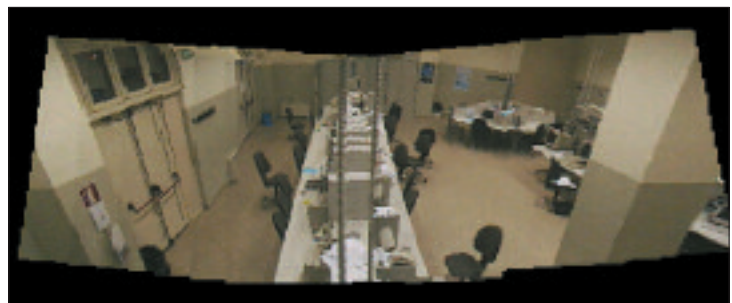


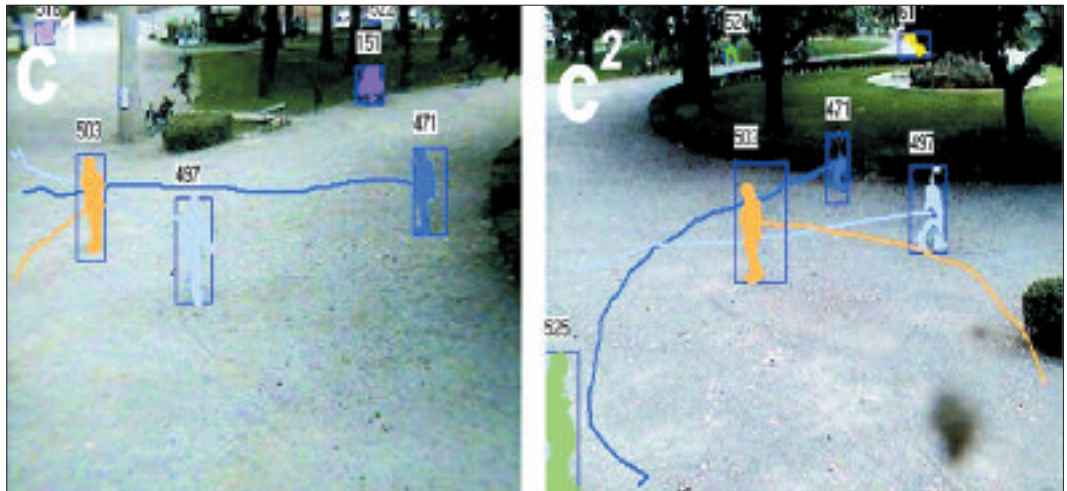
FIGURA 4

Mosaico creato da PTZ indoor



FIGURA 5

Omografia con due telecamere e rette epipolari



**FIGURA 6**  
 Esempio di tracking  
 di persone in un  
 parco (Progetto  
 Piano Telematico  
 ER LAICA- Comune  
 di Reggio Emilia)

to che lo sfondo cambia dinamicamente (ore del giorno, condizioni atmosferiche ecc.). L'approccio più famoso anche se computazionalmente pesante modella ogni pixel con una *Mixture of Gaussians* [7]. I sistemi semplici che propongono solo la "motion detection" si fermano qui e rilevano se punti possono dichiararsi in modo avendo cambiato la loro intensità frame dopo frame. Se il numero di tali punti è alto allora il sistema suppone la presenza di movimento e inizia la registrazione. Invece questo non è che il primo passo a cui devono eseguire altri processi di *labeling* per raggruppare punti adiacenti che potenzialmente fanno parte dell'oggetto in movimento (i blob disegnati nella Figura 1), tecniche di rimozione di artefatti come ombre [8], riflessi e rumore e tecniche di classificazione degli oggetti estratti, in base all'applicazione.

### 4.3. Tracking

Con la segmentazione si ottiene la coerenza spaziale tra i dati di ogni frame: si estraggono così gli oggetti di interesse, che devono essere inseguiti nel tempo (*tracking*) per mantenere anche la coerenza temporale.

Si studiano tecniche sofisticate per il tracking di più oggetti in movimento interagenti, che si occludono gli uni con gli altri o che possono temporaneamente sparire nella scena. I semplici *Kalman Filters* presuppongono un modello di moto lineare che può essere adottato solo in alcuni contesti come il controllo del traffico veicolare, ma che non sono adeguato per oggetti deformabili e con moto variabile quali

le persone. In questo caso si usano modelli statistici discriminativi, come quelli *appearance-based* e generativi, come il famoso algoritmo di *particle filtering* in cui si tengono in vita più stime (particelle) dello stato dell'oggetto poi messe in corrispondenza con l'osservazione ad ogni frame.

Un'interessante *survey sull'object tracking* è quello di Shah *et al.* [10]. L'output del tracking è l'informazione completa dell'oggetto, della sua forma e della traiettoria frame per frame; in caso di più telecamere viene fatto anche il *consistent labeling* per mantenere la stessa identità di ogni individuo se pur acquisito da telecamere diverse. Questo è assai importante nei nuovi sistemi dove si vuole tenere traccia degli oggetti osservati che si muovono in ampie scene coperte da più telecamere. Nella figura 6 si vede un esempio di tracking di persone nel Parco del Popolo di Reggio Emilia, uno dei prototipi del progetto LAICA (piano telematico Emilia Romagna) sviluppato dall'Image-Lab di Modena [11].

### 4.4. Reasoning and Action Recognition

Avendo a disposizione i dati della scena e degli oggetti in movimento, possono iniziare diversi passi di *reasoning* e di riconoscimento che dipendono naturalmente dall'applicazione. Iniziano processi di riconoscimento di eventi di interesse e di interazioni tra oggetti, quali il rilevamento di oggetti abbandonati. In questo caso si impiegano spesso metodi spesso basati su grafi spaziali e temporali che cercano oggetti che cambiano il loro sta-

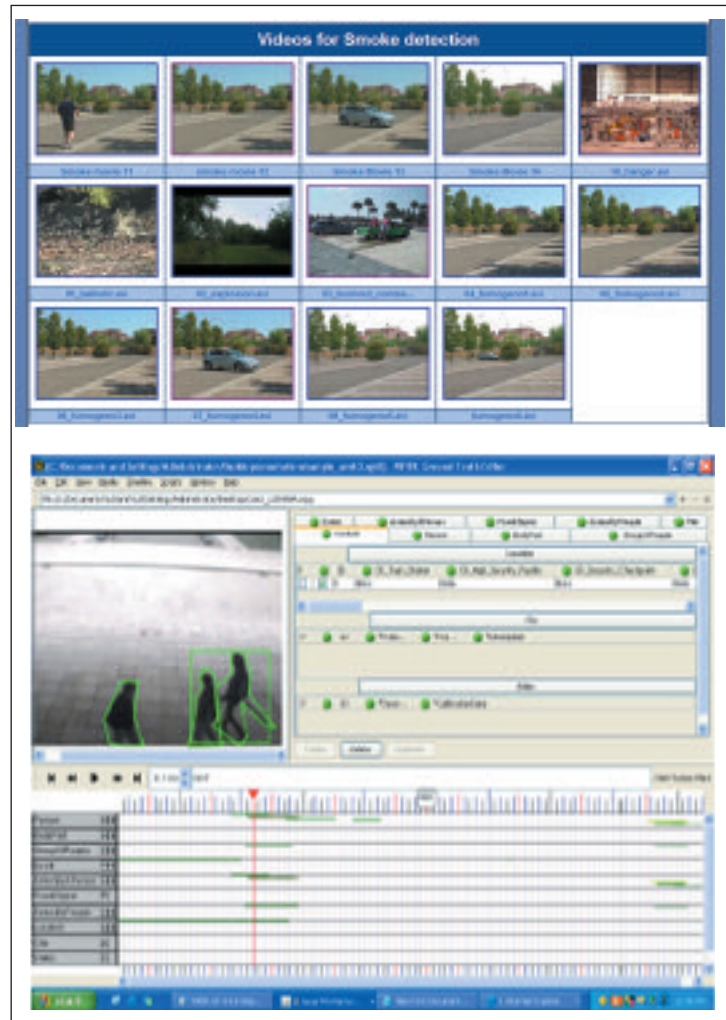
to da moto a fermo e che sono correlati con altri oggetti (persone) in movimento. L'attività di ricerca ora più attiva nel settore è quella dell'*action analysis*, che studia le azioni delle persone, le traiettorie e le modifiche di forma, per monitorarne attività o riconoscere il comportamento (*behavior analysis*). Anche in questo ambito si usano spesso grafi probabilistici quali gli HMM (*Hidden Markov Model*), tecniche di ragionamento con regole e *parser* sintattici. Un buon *survey sull'human motion analysis* è [12].

La ricerca in questo settore sta producendo diversi prototipi interessanti e si stanno diffondendo anche i primi esempi commerciali. Questo è certo uno dei settori dove si stanno concentrando maggiormente gli investimenti pubblici e privati. La ricerca si diffonde anche per il trasmettersi del sapere sull'argomento: molte librerie software sono di disposizione anche per chi si affaccia per la prima volta alla materia come i toolbox di Matlab e le OpenCV<sup>3</sup>, librerie supportate da Intel ma totalmente open source che si arricchiscono di giorno in giorno.

Nella ricerca è però fondamentale il confronto, e questo è uno degli aspetti più critici nel settore. Mentre in altri ambiti i *benchmark* di riferimento sono ormai ben diffusi (esempio, FERET per l'analisi del volto, TRECVID per il video retrieval) nella Video Sorveglianza ci sono tanti progetti poco coordinati. Tra essi si citano i *benchmark* PETS<sup>4</sup>, CAVIAR<sup>5</sup> e il recente VISOR<sup>6</sup> che è un *open repository* di video di sorveglianza, con relativa ontologia, annotazione in XML e strumenti per la visualizzazione e l'annotazione manuale basati sui tools VIPER dell'Univ. of Maryland (Figura 7).

## 5. LA RICERCA IN ITALIA

La Visione Artificiale per la Video Sorveglianza è un campo di ricerca dove l'Italia ha un ruolo di grande rilievo nella scena mondiale. Ne sono prova i tanti convegni internazionali orga-



**FIGURA 7**

In alto: VISOR - repository di video con annotazioni XML e MPEG7. In basso: esempio di annotazione con il tool VIPER

nizzati da ricercatori italiani negli ultimi anni, i tanti progetti di ricerca nazionali ed internazionali. Alcune di queste attività di ricerca sono state presentate a VISIT2008<sup>7</sup>, organizzato dal GIRPR<sup>8</sup> che ha creato uno specifico TC in Video Sorveglianza.

Ricercatori del GIRPR attualmente sono coinvolti in due PRIN (*Progetti di Rilevante Interesse Nazionale*): il primo coordinato dall'U-

<sup>7</sup> VISIT2008: <http://imabelab.ing.unimore.it/visit2008>

<sup>8</sup> Il Gruppo Italiano di Ricercatori in Pattern Recognition (GIRPR) è il capitolo italiano della International Association in Pattern Recognition (IAPR). Raccoglie ricercatori accademici ed industriali che si occupano di pattern recognition, computer vision e trattamento di dati multimediali in Italia. Per conoscere conferenze, eventi, attività e iscriversi al GIRPR, consultare: GIRPR: <http://www-girpr.dis.unina.it/> IAPR: <http://www.iapr.org/>

<sup>3</sup> OPENCV: <http://opencvlibrary.sourceforge.net/>

<sup>4</sup> PETS: <http://www.cvg.cs.rdg.ac.uk/slides/pets.html>

<sup>5</sup> CAVIAR: <http://homepages.inf.ed.ac.uk/rbf/CAVIAR/>

<sup>6</sup> VISOR: <http://imabelab.ing.unimore.it/visor>



niversità Udine, assieme a Roma e Pavia<sup>9</sup> si occupa di Ambient Intelligence ed in particolare di gestione, riconfigurazione ed analisi di prestazioni di reti di sensori e telecamere. Il secondo<sup>10</sup> coordinato da Modena con le Università di Firenze e Palermo e con la Fondazione Kessler di Trento e' invece focalizzato su algoritmi innovativi di visione per la Video Sorveglianza; si vogliono proporre nuovi paradigmi per sistemi di sicurezza compatibili con i vincoli di privacy. Da una parte si vuole garantire la privacy degli individui, ad esempio eliminando in tempo reale dai video i caratteri identificativi biometrici quali i volti, dall'altra si vogliono generare allarmi automatici per eventi di interesse a partire da piattaforme di nuove generazione con telecamere fisse, mobili, PTZ e reti di sensori.

I gruppi di ricerca coinvolti in questi progetti PRIN sono attivi da anni nel settore. Per quel che riguarda l'Università di Udine ricordiamo inoltre il progetto Europeo HamLET (*Hazardous Material Location and Tracking*) e il progetto NATO AMSS-CT (*Advanced Multisensory Surveillance Systems for Combating Terrorism*). Anche l'Università di Modena e Reggio Emilia coordina un progetto NATO (programma *Science for Peace*) chiamato BESAFE (*Behavioral Learning in Surveilled Areas with Feature Extraction*) in collaborazione con la Hebrew University di Israele e collabora con l'Università di Firenze nel progetto Europeo VIDI-VIDEO per il citato portale VISOR. Esistono molti altri gruppi attivi in Italia: presso il DIEE di Cagliari si lavora sull'integrazione di sistemi di video sorveglianza e di autenticazione biometrica (fingerprint) anche in collaborazione con i Ra.C.I.S. dell'arma dei Carabinieri. Presso i laboratori IMM e ISSIA del CNR di Lecce e Bari rispettivamente si studia l'impiego di nuove telecamere (*range sensor*) per l'analisi delle immagini di profondità. Il gruppo TeV della fondazione Kessler di Trento da anni sviluppa sistemi di Video Sorveglianza: si ricorda il progetto pilota finanziato dal Ministero dei Trasporti SCOCA per l'analisi del traffico veicolare in incroci cittadini.

<sup>9</sup> "Ambient intelligence: event analysis, sensor re-configuration and multimodal interfaces".

<sup>10</sup> "FREE SURF Free surveillance in a privacy respectful way".

IL VIPLAB di Verona sta lavorando nel progetto Europeo SAMURAI, per la coordinazione uomo-macchina nella gestione di reti di telecamere. Il Laboratorio LABVIS di Bologna infine da anni svolge attività di ricerca per la video sorveglianza e l'augmented reality nell'esplorazione di luoghi pubblici e musei.


Il Gruppo di Telecomunicazioni di Genova è attivo da anni nella videosorveglianza con risultati di eccellenza internazionali. Tra i diversi progetti si ricorda il VICOM (*Virtual Immersive Communication*), un grande progetto triennale finanziato dal MIUR che oltre alle università italiane ha coinvolto il CNIT, il CNR e l'ISCTI.

Pur ben lungi dall'essere esaustivo, quest'elenco ha voluto mostrare il ricco e dinamico panorama di ricerca italiano nel settore, coordinato dal GIRPR e in stretto legame con la ricerca internazionale nell'ambito dell'IA-PR e nell'IEEE.

La ricerca in Video Sorveglianza è estremamente fruttuosa in termini accademici, sia per progetti, sia per le rilevanti pubblicazioni internazionali e sta guidando la ricerca applicata in stretta collaborazione con aziende italiane e internazionali. Certi che quest'attività sarà trainante nel settore ICT nel prossimo futuro, non possiamo far altro che augurarci un ulteriore coinvolgimento nazionale degli enti di ricerca e governativi, per rafforzare un ambito scientifico che davvero può dirsi un fiore all'occhiello della ricerca italiana ed internazionale.

## Bibliografia

- [1] Forsyth D.A., Ponce J.: *Computer Vision: A Modern Approach*. Prentice Hall 2002.
- [2] Bishop C.: *Pattern Recognition and Machine Learning*. Springer 2006.
- [3] Shu C., Hampapur A., et al.: *IBM smart surveillance system (S3): a open and extensible framework for event based surveillance*. IEEE Conf. AVSS, 15-16 Sept. 2005, p. 318-323.
- [4] Ramesh V.: *Real-time vision at Siemens Corporate Research*. IEEE Conf. AVSS 2005, 15-16 Sept. 2005, p. 300-305.
- [5] Gonzalez R., Woods R.: *Digital Image Processing*. Prentice Hall, 2002.
- [6] Calderara S., Cucchiara R., Prati A.: *Bayesian-competitive Consistent Labeling for People Surveillance*. In: IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 30, n. 2, 2008, p. 354-360.

- 
- [7] Lowe D.G.: Distinctive image features from scale-invariant keypoints. *Int. Journal of Computer Vision*, Vol. 60, n. 2, 2004, p. 91-110.
- [8] Stauffer C., Grimson W.: Learning patterns of activity using real-time tracking. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 22, n. 8, 2000, p.747-757.
- [9] Prati A., Mikic I., Trivedi M.M., Cucchiara R.: Detecting Moving Shadows: Algorithms and Evaluation. In: *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 25, n. 7, July, 2003, p. 918-923.
- [10] Yilmaz A., Javed O., Shah M.: Object tracking: A survey. *ACM Computing Survey*, Vol. 38, n. 4, Dec. 2006.
- [11] Cucchiara R., Prati A., Vezzani R.: *Ambient Intelligence for Security in Public Parks: the LAICA Project*. In: Proc. of IEE Intern. I Symp. on Imaging for Crime Detection and Prevention, London, UK, 2005.
- [13] Moeslund T.B., Hilton A., Krüger V.: A survey of advances In vision-based human motion capture and analysis. In: *Computer Vision and Image Understanding*, Vol. 104, n. 2, 2006, p 90-126.
- [14] Cuntoor N.P., Yegnanarayana B., Chellappa R.: Activity Modeling Using Event Probability Sequences. *IEEE Trans.on Image Processing*, Vol. 17, n. 4, 2008, p.594-607.

RITA CUCCHIARA è Professore Ordinario di Sistemi di Elaborazione presso la Facoltà di Ingegneria di Modena; è docente di Calcolatori Elettronici e di Visione Artificiale. È Vice Preside della Facoltà e coordinatore del corso di "Computer Engineering and Science" della Phd School in ICT. Coordina il Laboratorio ImageLab che svolge attività di ricerca in Visione Artificiale, Pattern Recognition e Multimedia, per la video sorveglianza, la sicurezza, la visione robotica e per sistemi multimediali per Digital library di immagini e video in diversi progetti nazionali ed internazionali. È autore di più di 150 lavori su riviste e congressi internazionali. Per riferimenti alla ricerca e alla pubblicazioni si veda: <http://imagelab.ing.unimore.it>  
E-mail: [rita.cucchiara@unimore.it](mailto:rita.cucchiara@unimore.it)