

Le applicazioni di video-sorveglianza avevano già iniziato a fare capolino negli anni '90, ma è solo dopo gli attentati di New York nel 2001 e Londra nel 2005 che c'è stata una diffusione significativa. Nel solo Regno Unito, per esempio, nel 2009 si contavano già oltre 4 milioni di dispositivi di video-sorveglianza installati. Secondo studi recenti, il mercato ha generato un volume di affari su scala mondiale compreso tra 7,25 e 14 miliardi di dollari e dovrebbe crescere fino a raggiungere quota 46 miliardi entro il 2013. I principali ambiti d'impiego, secondo una ricerca di Multimedia Intelligence, riguardano il settore bancario, dei trasporti, l'industria, la vendita al minuto, il settore pubblico, la scuola e l'università. Nel 2008 la principale attività di video-sorveglianza ha riguardato il controllo del traffico. Tradizionalmente, i sistemi di video-sorveglianza attiva si basano sulla presenza di un operatore che controlla le immagini di una o più telecamere analogiche, su uno o più schermi, per individuare eventi di varia natura. Diverse ricerche hanno tuttavia dimostrato che tale approccio ha grosse limitazioni per quanto concerne l'affidabilità. In particolare si è riscontrato che l'attenzione dell'operatore diminuisce drasticamente, al disotto di livelli accettabili, già dopo 20 minuti di osservazione continua; più o meno allo stesso modo, l'operatore non è in grado di monitorare

# LA VIDEO-SORVEGLIANZA SI FA INTELLIGENTE

IL CONNUBIO TRA COMPUTER VISION E SICUREZZA NEL SETTORE EMBEDDED SANCISCE LA NASCITA DEI SISTEMI DI VIDEO-SORVEGLIANZA INTELLIGENTE. SIA CHE SI PARLI DI SISTEMI ANTIFURTI CHE DI MONITORAGGIO DELLE FOLLE, DI CONTROLLO DEL TRAFFICO O DI DIAGNOSTICA REMOTA, DALL'ELETTRONICA ARRIVA UN AIUTO IN PIÙ PER LA NOSTRA SICUREZZA

di Mariano Severi

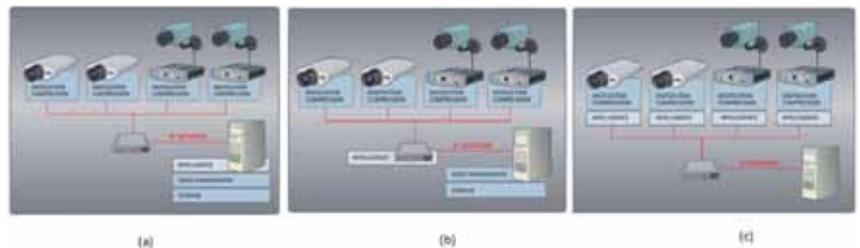


Fig. 1 - Architettura delle reti di video-sorveglianza intelligente: a) centralizzata; b) ibrida; c) distribuita

più di 12 camere simultaneamente per più di 15 minuti e comunque la probabilità di reagire istantaneamente a un evento è

soltanto di 1 su 1.000. I tradizionali sistemi di video-sorveglianza, come molti degli impianti televisivi a circuito chiuso (Closed Circuit Television - CCTV) cui siamo abituati, hanno così tipicamente solo funzione di dissuasione oppure servono come strumento di indagine successivo.

I continui progressi tecnologici nel settore dei sistemi di visione hanno determinato negli ultimi quindici anni una lenta transizione verso l'impiego di camere digitali dotate di connettività di rete (camere Internet Protocol). La tendenza è stata accelerata nell'ultimo periodo dalla crescente disponibilità, a costi contenuti, di sensori a elevata risoluzione, dalla diffusione di protocolli di connessione wireless high-speed (come il WiMax o l'LTE), dalla definizione di nuovi algoritmi di compressione dati efficienti (come l'H.264). Se ancora nel 2008 si stimava che addirittura il 95% delle oltre 40 milioni di camere installate nel mondo fosse di tipo analogico, già un anno dopo, indagini di mercato riportavano che ogni 4 nuove camere instal-

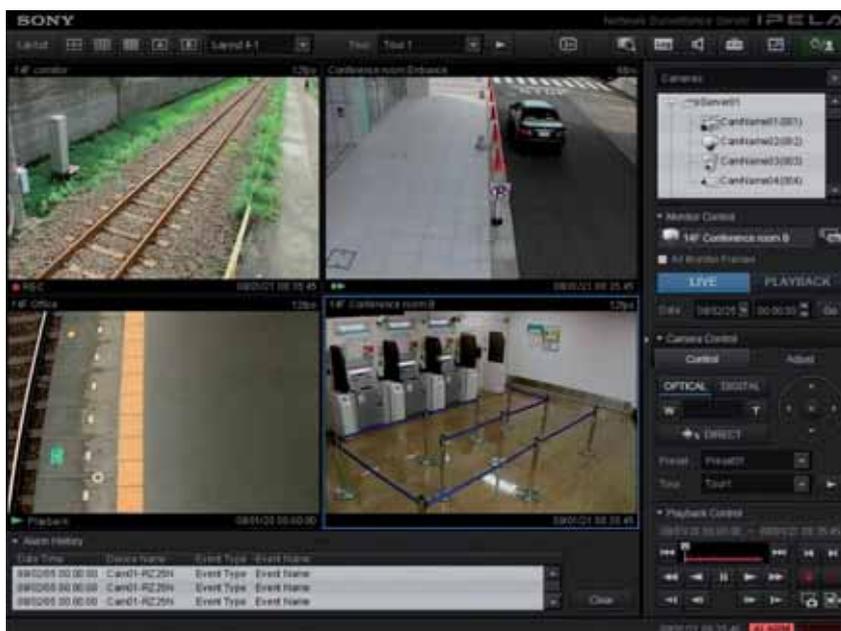


Fig. 2 - La soluzione RealShot Manager di Sony per applicazione di video sorveglianza intelligente

late 1 era di tipo digitale. Nel 2009 sono stati emessi i principali due standard per quanto concerne le camere IP, Onvif e Psia; il primo è supportato da Axis e Sony che da sole detengono il 45% del mercato mentre il secondo è promosso Cisco e da una serie di produttori minori come Panasonic e Honeywell.

Nel settore della sicurezza, la transizione verso camere IP a elevata risoluzione ha determinato la nascita di nuove applicazioni di video-sorveglianza intelligente (Intelligent Video Surveillance o Smart Surveillance). Con tale termine s'intendono tutte quelle applicazioni in grado di estrarre automaticamente dalle immagini acquisite informazioni di vario tipo, sotto forma di meta-data, che elaborate in tempo reale consentono di individuare eventi di interesse. L'elaborazione automatica dell'immagine consente di superare le limitazioni di un sistema tradizionale basato su un operatore, assicura operatività continua e risposta 'immediata' all'evento, riduce la quantità di informazioni che è necessario memorizzare, supporta l'integrazione delle informazioni da più camere o la fusione di dati di diversa natura per la discriminazione dell'evento. È stato, per esempio, recentemente dimostrato un sistema di video-sorveglianza intelligente che utilizza le immagini provenienti da due camere per una ricostruzione stereo-scopica della scena al fine di determinare l'esatta posizione di un oggetto; in risposta a un evento, la posizione determinata è utilizzata per controllare il puntamento e la messa a fuoco di una terza camera PTZ (Pan-Tilt-Zoom) impiegata per avere una visione di dettaglio della zona di interesse. Un'applicazione analoga ma a basso costo può essere realizzata mediante una rete di sensori wireless i cui dati sono usati per il controllo di una camera IP.

### **Architetture centralizzate, distribuite e ibride**

Come in tutte le altre applicazioni, i sistemi di video-sorveglianza intelligenti possono avere architettura centralizzata o distribuita. Nel primo caso (figura 1a) le immagini acquisite sono compresse e inviate a un server remoto per l'elaborazione dei dati. Tale soluzione semplifica il progetto delle camere e facilita la combinazione dei dati ma richiede server con elevata capacità di calcolo e comporta elevata occupazione di banda sulla rete; risulta così tipicamente poco scalabile e soprattutto meno affidabile. Nelle solu-

zioni distribuite (figura 1c) gli algoritmi di elaborazione delle immagini sono invece eseguiti sulla camera stessa; vengono quindi inviate al server centrale solo le informazioni utili, in seguito al riconoscimento di un evento. Nelle architetture più complesse, la video-camera stessa dispone anche di capacità di memorizzazione locale dei dati in formato non compresso o a elevata qualità. Per gli utenti che intendano ridurre gli investimenti riutilizzando parti delle infrastrutture pre-esistenti, sono pure possibili



**Fig. 3 - Camere Sony con tecnologia Depa per applicazioni di video-sorveglianza intelligente**

soluzioni ibride (figura 1b) in cui l'elaborazione dei dati avviene nei router o negli switch di rete. I vantaggi delle reti distribuite (e in minor misura, di quelle ibride) sono ovviamente un minore uso della banda per la trasmissione dei dati, la riduzione dei costi di mantenimento dei server centrali e dei costi operativi, una migliore scalabilità e manutenibilità dell'architettura. La realizzazione di camere IP intelligenti è resa possibile oggi a costi contenuti grazie ai continui progressi tecnologici nel settore, soprattutto, dei processori embedded; la piattaforma privilegiata è spesso l'architettura ARM. L'implementazione degli algoritmi di analisi video richiede tipicamente il 15% della potenza di calcolo dei DSP oggi disponibili rendendo possibile la creazione di reti distribuite con 200 nodi. Mpeg-4/H.264 e Mjpeg sono le soluzioni tipicamente adottate per la trasmissione delle immagini compresse, ma nuovi schemi di compressione per ridurre ulteriormente il bit-rate sono in fase di studio. La diffusione, in particolare, dello standard H.264 potrebbe essere nel prossimo futuro frenata qua-

lora fossero approvate le richieste di brevetti attualmente in fase di discussione, il che comporterebbe la necessità per gli utilizzatori di pagare diritti per l'adozione del codec. Tra le iniziative più interessanti vi è certamente la definizione del nuovo standard open-source VP8, supportato da Google. Mpeg-4 raggiunge maggiori fattori di compressione rispetto a Mjpeg, risultando così preferibile per applicazioni di streaming in tempo reale, ma è caratterizzato da maggiore complessità computazionale, latenza più elevata e minore

robustezza. La connessione delle camere può essere di tipo wired o wireless. In questo secondo caso, in particolare, oltre il 3G, la tanto attesa diffusione delle reti WiMax (una volta rimossi i vincoli imposti dalle risoluzioni adottate a garanzia di una maggiore sicurezza ma che, di fatto, hanno frenato l'evoluzione del mercato negli ultimi anni) dovrebbe in particolare consentire l'impiego di video camere intelligenti per applicazioni di video-sorveglianza anche a bordo di stazioni mobili, come auto della polizia, autobus o ambulanze, seppur a bassa velocità. In caso contrario, infatti, pro-

blemi di fading multipath del segnale indurrebbero elevata frequenza di errore, riducendo la banda di trasmissione utile. Per stazioni fisse su piccola scala sono state realizzate reti di video-sorveglianza che adottano un protocollo 802.11. Nella maggior parte delle applicazioni, le reti sono strutturate con architettura mesh. Schemi WPA, WPA2, Tkip e AES sono tipicamente adottati per assicurare la protezione e la sicurezza dei dati. È interessante citare che vi è pure un crescente interesse nella realizzazione di applicazioni sotto Android per telefoni cellulari e smart phone che permettono di monitorare da remoto camere di video-sorveglianza. Nell'ambito delle connessioni cablate invece, l'infrastruttura privilegiata è spesso Ethernet, grazie anche a una crescente diffusione della tecnologia PoE (Power over Ethernet), che consente la distribuzione dell'alimentazione alla camera attraverso lo stesso cavo di rete. Dal punto di vista dei protocolli, sono correntemente implementate soluzioni Http, TCP/IP, Dhcp, Ddns, RTP/Rtsp; più di recente sono stati propo-

sti pure degli schemi che adottano lo standard Rtcp e si basano su server video cooperativi per migliorare la qualità del servizio nelle architetture su larga scala. In questo stesso ambito, in letteratura è stato proposto l'utilizzo della linea Adsl per la connessione delle camere tra loro e ai server remoti.

## Un esempio di soluzione distribuita

Diversi sono i produttori di soluzioni di video-sorveglianza intelligenti. RealShot Manager (figura 2), per esempio, è la soluzione software lanciata da Sony per la gestione da remoto delle proprie camere basate su tecnologia Depa (Distributed Enhanced Processing Architecture). Le caratteristiche principali sono il monitoraggio e la ricerca intelligenti grazie alle

codifica simultanea dello stream video in due diversi formati per la memorizzazione locale e la trasmissione remota, funzionalità giorno/notte per assicurare visibilità eccellente anche in condizioni di luminosità minima. Offre capacità integrate di Intelligent Motion Detection (utilizzando gli ultimi 15 frame acquisiti per consentire di ridurre al minimo i falsi allarmi) e Intelligent Object Detection (consentendo, per esempio, di rilevare la presenza di un oggetto rimasto nella scena per oltre 40 secondi o rimosso dalla stessa).

## Video-sorveglianza intelligente e video content analysis

Gli algoritmi di elaborazione delle immagini adottati nei sistemi di video-sorveglianza intelligente vanno genericamente

gnition) per il controllo di accesso. La maggior parte degli algoritmi si basa su schemi di natura statistica che permettono l'identificazione dello sfondo dell'immagine e la segmentazione delle regioni in primo piano. Regioni persistenti nel tempo, chiamate blob, sono quindi classificate e tracciate per il riconoscimento di oggetti o persone. I sistemi più avanzati supportano poi l'identificazione di eventi complessi mediante algoritmi basati su reti neurali, reti Bayesiane o modelli di Markov nascosti e adottano tecniche di ricostruzione 3D per ridurre significativamente la probabilità di falsi allarmi. In questo modo, l'analisi video non si riduce alla semplice localizzazione di eventi ma promette l'interpretazione degli stessi.

Uno dei progetti più interessanti in questo senso è, per esempio, il programma Samurai (Suspicious and Abnormal behaviour Monitoring Using a network of cAmeras for situation awareness enhancement), un'attività finanziata nell'ambito del Settimo Programma Quadro Europeo che ha come obiettivo la realizzazione di un innovativo sistema di sorveglianza intelligente per il monitoraggio di persone e veicoli all'interno e nelle aree circostanti le infrastrutture pubbliche critiche. Cubea, invece, è stato un progetto finanziato dal programma Eurostars di Eureka e coordinato da Search-Lab, conclusosi nell'ottobre 2010 e che ha visto la realizzazione di un sistema di video-sorveglianza intelligente per il monitoraggio di persone e beni in centri di vendita al dettaglio. Sulla base delle informazioni acquisite, il sistema è in grado di prevedere il comportamento dei clienti consentendo di migliorare la disposizione sia delle merci negli scaffali sia delle locandine pubblicitarie.

Un ambito di interesse particolare nel settore della video-sorveglianza intelligente, in effetti, in relazione anche agli aspetti di sicurezza civile, è proprio quello del monitoraggio delle folle, sebbene esperimenti recenti abbiamo dimostrato che la tecnologia per tali applicazioni non è ancora del tutto matura. I sistemi di video-sorveglianza intelligenti infatti tendono a funzionare bene prevalentemente in condizioni di luminosità costante o in assenza di continue occlusioni del campo di vista, nel qual caso hanno difficoltà a isolare gli oggetti in primo piano dallo sfondo della scena.



Fig. 4 - Esempi di object detection e tracking

funzionalità Video Motion Filter Alarm (allarme sui filtri di movimento) e Video Motion Filter Search (ricerca sui filtri di movimento). Il software supporta la riproduzione simultanea di immagini da più videocamere anche durante la registrazione, offre la disponibilità di funzioni per la maschera di alcune zone della scena (a tutela di eventuale privacy), e per la registrazione audio. Tra le camere supportate da RealShot Manager vi sono per esempio la SNC-RZ50P (figura 3), una video camera PTZ (Pan-Tilt-Zoom) dotata di zoom ottico 26x con velocità di rotazione di 300° al secondo e possibilità di definire fino a 16 posizioni di ritorno con una precisione di 0,064°. La camera implementa integrazione di frame dinamico per assicurare elevata qualità delle immagini in movimento, compressione Jpeg, Mpeg-4 e H.264 con, eventualmente, supporto per

sotto il nome di Video Analytics o Video Content Analysis (VCA). Algoritmi di questo tipo sono usati pure in altri settori, come quello della domotica, automobilistico, trasporti o intrattenimento. Secondo recenti stime di Frost & Sullivan, si tratta di un mercato che valeva 138 milioni di dollari nel 2008 e che dovrebbe crescere esponenzialmente nei prossimi anni. Tra le funzionalità principali che rientrano in questo ambito vi sono, per esempio, le tecniche di identificazione di un oggetto all'interno di una scena; è così possibile riconoscerne e tracciarne il movimento (Motion Detection and Tracking) oppure notarne la rimozione, utile nel caso dei sistemi antifurto. Vi sono pure algoritmi di riconoscimento facciale (il più famoso è forse quello di Viola-Jones) o riconoscimento targhe basati, per esempio, su funzionalità OCR (Optical Character Reco-