

Videosorveglianza come supporto interattivo

*Franco Dischi**

Riassunto

La videosorveglianza non è e non può più essere intesa come un sistema di acquisizione delle immagini fine a se stesso. Il "prodotto" audiovisivo acquisito, in aggiunta alla sorveglianza e alla sicurezza, rappresenta un'utile fonte di informazioni in caso di conservazione e analisi automatica nella pianificazione urbana per l'ottimizzazione delle risorse del territorio e dei mezzi di supporto, per esempio il monitoraggio ambientale a tutela dell'habitat, del territorio e dell'ecosistema.

Si tratta di modelli comportamentali precognitivi di analisi video per un contesto percettivo della situazione di pericolo.

Résumé

La vidéosurveillance n'est pas et ne peut plus être considérée comme un simple système d'acquisition d'images. Le "produit" audiovisuel acquis, en plus de la surveillance et de la sécurité, constitue une source d'informations très utile en cas de stockage et d'analyse automatique de données dans le domaine de l'aménagement urbain afin d'optimiser les ressources du territoire et les dispositifs auxiliaires, par exemple la surveillance de l'environnement pour protéger l'habitat, le territoire et les écosystèmes.

Il s'agit de modèles de comportement précognitifs d'analyse vidéo, dans un contexte perceptif de la situation de danger.

Abstract

Video surveillance is not and cannot be considered a system of image acquisitions "end in itself".

The acquired audio-visual "product", in addition to surveillance and security, provides a useful source of information in case of storage and automatic analysis of data in urban planning to optimise land resources and means of support, for example environmental monitoring to protect habitat, land and ecosystem.

These are behavioural precognitive models of video analysis, for a perceptive context of the situation of danger.

L'inizio dello studio per lo sviluppo di strumenti di identificazione da utilizzarsi su vasta scala e basati su apparati di video ripresa fu una conseguenza dell'epoca del terrorismo.

Tra gli addetti ai lavori si era sviluppata una particolare sensibilità nel dotarsi di strumenti d'aiuto per l'identificazione della persona in modo da scoraggiare gli eventi delittuosi, gettando così le basi per uno più specifico e inarrestabile ciclo evolutivo, ora definita videosorveglianza interattiva.

Le prime applicazioni vendute su larga scala inerenti la visualizzazione di immagini con successiva registrazione di tipo "proprietario"

(da qui l'acronimo TVCC, Televisione a Circuito Chiuso) sono avvenute proprio ambito bancario.

Lo sviluppo economico del paese ha poi esteso l'utilizzo di apparati di video ripresa, impiegandoli congiuntamente a sistemi: antintrusione, antirapina, antitaccheggio, controllo del territorio, sorveglianza di aree ad alto rischio, vigilanza ambientale; lo stesso corpo di Protezione Civile impiega apparati mobili installandoli in zone a rischio in arco temporale specifico proprio a scopo preventivo. Di fronte a richieste di messa in sicurezza e quindi a progetti di sistemi di protezione per aree sensibili tramite l'installazione, ad esempio, di videocamere in

* Notifier Italia.

ambasciate, tribunali, siti militari, aeroporti, carceri – luoghi questi in cui diventa di primaria importanza la questione sicurezza - la TVCC risulta essere di gran lunga apprezzata rispetto ai sistemi elettronici tradizionali, se poi abbiniamo le immagini all'audio, i metadati raccolti e archiviati diventano estremamente utili come tecnologia investigativa e d'*intelligence*.

E' innegabile come il tema "sicurezza" sia uno degli argomenti rappresentanti le ultime campagne elettorali proprio per far fronte alla sempre maggiore richiesta di protezione in ambito territoriale e sociale. I cittadini, consapevoli e coscienti riguardo alla loro tutela personale e diritto alla sicurezza, fanno esplicita richiesta alle istituzioni, le quali hanno risposto emanando il decreto legge 23 febbraio 2009, n. 11, "recante misure urgenti in materia di sicurezza pubblica e di contrasto alla violenza sessuale, nonché in tema di atti persecutori", convertito in legge 23 aprile 2009, n. 38, che ha modificato il tempo massimo di conservazione delle immagini raccolte mediante l'uso della videosorveglianza, previsto dalla legge sulla privacy, aumentandolo dalle originarie 24 ore a sette giorni, fatte salve speciali esigenze di ulteriore conservazione.

Non possiamo esimerci dal riconoscere come l'evoluzione elettronico-tecnologica ci meravigli quotidianamente con apparati la cui sofisticazione rasenta l'incredibile pur adattandoci (molte volte nostro malgrado) all'evoluzione della specie. Non da meno lo sono stati gli apparati di video ripresa le cui prestazioni sono notevolmente progredite, crescendo in maniera inversamente proporzionale al loro prezzo. Più economici e più

performanti, sono diventati papabili anche in aree d'utilizzo, impensabili fino ad un decennio fa incrementandone, di fatto, la diffusione.

Il comparto "videosorveglianza" ha visto aumentare la domanda in modo esponenziale e questo ha procurato denaro fresco alle aziende produttrici che hanno potuto così investire nello sviluppo di nuovi e sempre più performanti apparati dedicati.

La TVCC abbinata alla video registrazione locale è stato il mezzo più innovativo utilizzato in questi ultimi anni non tanto per abbattere il rischio ed accrescere l'effetto deterrente, ma al fine del riconoscimento (a posteriori) dell'evento delittuoso o di pericolo e poterne analizzare il frutto.

L'esperienza, gli studi e l'applicazione delle moderne tecnologie ci consentono di utilizzare una soluzione filosofica e sistemistica atta a modificare la condizione di prevenzione e protezione non solo antirapina, com'era stata pensata, ma applicabile a contesti molto vasti. L'evoluzione tecnologica ha consentito di modificarne l'uso grazie proprio al miglioramento inarrestabile in questo settore, passando in pochissimi anni dagli ormai vetusti VCR (Video Cassette Recorder), ai NVR IP (Network Video Recorder) veri e propri Server di rete Ethernet per video digitale in qualità DVD o MegaPixel fullHD le cui immagini possono essere viste e gestite da qualsiasi parte nel mondo ci si trovi, tramite una connessione di rete sia essa filare o wireless.

La legge sulla privacy, poc'anzi richiamata, in questi anni ha sollevato non pochi dubbi sull'utilizzo massivo di videocamere collocate in

ogni dove. Come far conciliare la richiesta di sicurezza col diritto alla privacy?

Provocatoriamente, se una persona è moralmente integra ed onesta perché mai dovrebbe soffrire della sindrome del Grande fratello?

E' altresì vero che l'opera di George Orwell ha avuto molto successo ed è stata tradotta e letta dall'intero cosmo allo slogan "*Il Grande Fratello vi guarda*" dove si narra appunto l'utilizzo di telecamere come strumento di repressione al fine di annullare il libero arbitrio dei cittadini dello stato di Oceania.

Per non parlare dell'idiosincrasia che ha scatenato l'introduzione del passaporto e carta d'identità elettronica dove tra i dati contenuti sono presenti le impronte digitali. Con un minimo di raziocinio è facile intuire come sia semplice lasciare tracce attraverso i nostri movimenti sia fisici e sia pecuniari (la normale vita sociale) movimenti bancari, carta di credito, bancomat, telepass, carta cliente (ipermercati), codice fiscale, tessera sanitaria, TV digitale, telefono cellulare, Internet, biglietti aerei o ferroviari, tutti facilmente tracciabili e senza far uso di microspie, microfoni, localizzatori GPS o apparati più professionali in dotazione alle forze d'investigazione.

Ecco che, sgonfiata la percezione d'essere spiati dalle telecamere, si accetta di buon grado l'uso di sistemi di video sorveglianza poiché s'intuisce come l'impiego, più o meno dissimulato sul territorio, possa fornire al normale cittadino un senso di protezione, riscontrabile nelle cronache quotidiane le quali riportano la cattura di malfattori grazie proprio alle registrazioni video.

1. Non solo telecamere.

L'efficienza funzionale è stata migliorata sotto tutti gli aspetti, la convergenza verso l'uso

pressoché assoluto delle tecnologie digitali si traducono in maggiore definizione delle immagini, maggiore numero di pixel a disposizione per l'elaborazione, maggiore dettaglio e precisione d'analisi permettendo ai software IVS (Intelligent Video Software) di riuscire a discriminare all'interno del singolo fotogramma ogni minima variazione attivando procedure e processi automatici di riconoscimento o scansione che sicuramente sfuggirebbero all'occhio umano.

La video analisi ha subito un'escalation progettuale, grazie all'immediatezza del recupero degli investimenti fatti, per la messa a punto di sistemi automatici di riconoscimento targhe.

Il progetto finanziato dalle sanzioni ha contribuito a fondare le basi per la rielaborazione in 3D del contenuto della scena e arrivare a livelli di sofisticazione tali da consentire il riconoscimento automatico dell'asportazione o il lascito di oggetti, il conteggio di persone, individuazione pre-cognitiva di una situazione di pericolo come il formarsi di un assembramento, una manifestazione. E' altresì facile intuire come sia d'aiuto un sistema automatico, che allerta l'operatore nel centro di controllo di polizia o della società autostrade dove convergono centinaia di telecamere, se un'auto che percorre in senso inverso alla normale marcia, se occupa per tot tempo la corsia d'emergenza, segnala il ribaltamento o il fermo di un veicolo in galleria, ecc..

A prescindere dalle telecamere e dal sistema d'acquisizione, è il software d'analisi che riconosce la scena non classificandola "normale" e crea l'evento d'allarme.

La potenza di calcolo necessaria ad elaborare scene complesse è uno dei limiti dello sviluppo di applicativi che sono tendenzialmente più maturi rispetto all'hardware su cui dovranno girare; più precisamente, un evoluto software d'analisi ha la necessità d'estrazione di database strutturati e distribuiti formati da grandi collezioni multimediali registrati su reti di telecamere e microfoni dislocati in siti reali.

Il prodotto flusso audiovisivo, in aggiunta alla sorveglianza di sicurezza, rappresenta un'utile fonte d'informazioni in caso di memorizzazione per la futura analisi e corretta pianificazione/ottimizzazione delle risorse "ambiente" in cui calare e configurare il "prototipo tipo".

L'analisi si basa su dati di tipo sensoriale, fonte da cui creare il modello per conto di due tipi di conoscenza. Da un lato abbiamo il contenuto della conoscenza, caratterizzata da un livello di eventi primitivi (ad esempio, classificazione degli oggetti e loro traiettorie), dall'altro, un secondo strato di eventi semantici in grado di fornire conoscenze utili per una visione più complessa di episodi, circostanze e relazioni con cui interagiscono.

L'acquisizione di dati su lunghi periodi consente di affinare i modelli comportamentali estratti proprio da grandi collezioni al fine di plasmarli attraverso uno specifico linguaggio ontologico: soggetto – relazione – oggetto, dove il significato dello studio dell'essere o dell'esistere sono concetti fondamentali di categoria e di relazione basilari per creare dei modelli matematici su cui si costruisce l'intelligenza artificiale.

Si evince come la correlazione di una struttura dati gerarchica, applicata ad un dominio di server cui adottare relazioni, assiomi, regole per la

condivisione della conoscenza, porta a diverse classificazioni condivisibili e applicabili a regole di ragionamento deduttivo che sfociano nelle diverse tecniche di *problem solving*. Il sistema è, e sarà, così d'aiuto all'operatore, soprattutto trattasi di Centro di Controllo Unico territoriale, dove le immagini ed i dati da elaborare fanno capo a centinaia o migliaia di telecamere distribuite.

2. La sensazione di pericolo.

Le pubbliche amministrazioni, gli enti locali, le forze di polizia col coordinamento del governo centrale si stanno attivando affinché si possano coordinare gli interventi ottimizzando le risorse e riducendo i tempi ed i costi. Le cariche istituzionali capitoline, in testa a tutti, hanno subito pressioni mediatiche dovute principalmente a stupri, susseguiti in un brevissimo arco temporale.

Tali situazioni destano preoccupazione e allarme sociale con gravi ripercussioni sul livello di sicurezza percepito dai cittadini.

E' in un simile contesto che le amministrazioni pubbliche richiedono, al fine di innalzare i livelli di sicurezza, contributi alle aziende private e, pertanto, la sicurezza sul territorio non poteva non prevedere, tra le varie iniziative, anche l'installazione di sistemi di videosorveglianza, così detta, *mobile*. Da qualche tempo è prassi installare telecamere sul territorio, nelle stazioni, nelle grandi vie di comunicazioni, nelle metropolitane, dove lo spostamento di grandi masse di persone ha necessariamente necessità di un maggiore controllo.

Ora si sta facendo di più. RFI (Rete Ferroviaria Italiana), assieme alle associazioni di categoria comprendenti tutti i mezzi mobili di trasporto urbano ed extraurbano, ha un piano per la

sicurezza della circolazione rotabile con piani di sviluppo d'impianti di sicurezza di terra e di bordo, tanto da dotare le carrozze ed i bus di sistemi automatici interattivi di sicurezza attiva compreso la video sorveglianza attraverso apparati che rispondono all'acronimo MVR (Mobile Video Recorder) per identificare quel dispositivo espressamente realizzato ai fini della video sorveglianza sui mezzi di trasporto.

L'MVR, realizzato espressamente per essere installato su mezzi mobili quali treni, navi, traghetti, metropolitana, pullman, bus, mezzi di soccorso, è in grado di sfruttare tutte le caratteristiche intrinseche di un DVR/NVR atto a sopportare condizioni gravose d'ambiente, in primis le vibrazioni, registrando in locale i filmati e l'audio collegandosi quando necessita, con uno o più centri di controllo, territorialmente distribuiti.

Integrabile con un GIS (Geographic Information System) attraverso una connessione satellitare con una semplice antennina GPS, il mezzo su cui è installato l'MVR diventa così perfettamente rintracciabile sulle mappe elettroniche (ad esempio di Google Map con licenza gratuita) sempre aggiornate e visibili in formato iconografico.

La mappatura satellitare interattiva al centro di controllo fa intervenire un altro aspetto oltre la video sorveglianza che è quello gestionale: il mezzo è individuabile immediatamente grazie alla localizzazione GPS oltre alla ricezione dell'eventuale allarme scatenato da un evento di pericolo qualunque esso sia; ciò darà modo all'operatore di vedere e colloquiare con il personale a bordo gestendo puntualmente le forze

d'intervento a seconda della tipicità dell'allarme ricevuto.

Un sistema automatico e configurabile coordina gli eventi d'allarme ricevuti, ad esempio se il mezzo è al di fuori del percorso prestabilito, se al di sotto di una velocità minima impostata, se rimane fermo per un tempo superiore al consentito, controlla il percorso pre-determinato attraverso tracking multi mezzo, oltre a registrare in piena autonomia sia l'audio che il video.

Va da sé che un sistema così configurato può essere calato in diverse realtà come ad esempio: mezzi pubblici o di pubblica sicurezza, Protezione civile, Vigili del fuoco, porta valori, mezzi pesanti, tram, bus, treni, vigilanza, ecc.

Oltre a fungere da deterrente, le immagini acquisite possono essere inoltrate a diversi centri di controllo distribuiti territorialmente contenendo così la spesa nella sua globalità innalzando di molto il livello di tranquillità del cittadino.

L'identificazione di volti per il riconoscimento automatico tra la folla di persone ricercate si palesa nella maggior parte dei recenti film polizieschi. L'applicazione analitica sulla geometria del volto per il riconoscimento biometrico di un viso, anche celato da baffi posticci, barba o parrucca, è allo studio da decenni con risultati più o meno soddisfacenti. L'avvento delle telecamere MegaPixel ed i nuovi software d'analisi sono in grado di dettagliare maggiormente l'immagine riducendo il grado d'errore grazie all'alta definizione; di contro non si riesce ad avere un sistema stabile perché il software di elaborazione deve continuamente essere rivisto adattandolo ai nuovi formati d'acquisizione.

3. Prevenzione e videosorveglianza attiva.

Nel campo della video sorveglianza attiva si sta lavorando a tale scopo cioè anticipare il fatto, prevenendone con buon anticipo l'accadimento, preavvisando l'operatore che potrà allarmare e coordinare le forze d'intervento prima che succeda. Il Content Analysis confronta con dei dati probabilistici acquisiti in reale ed elaborati su milioni d'informazioni multimediali la scena nel suo contesto, non solo pochi frame dell'accaduto, quindi, ma precognizione su quello che sta accadendo.

Mettere a punto una tecnologia in grado di estrarre ed analizzare il contenuto di un'enorme mole di dati la cui elaborazione sia in grado di fornire modelli di condotta ai fruitori delle informazioni, consente di affinare lo studio degli atteggiamenti e modi di agire dei soggetti attraverso la scienza dell'analisi comportamentale.

Potendo disporre di grandi collezioni multimediali registrate su reti di telecamere e microfoni dislocati in ambienti ad alta frequentazione, i metadati elaborati ed affinati sul lungo periodo consentono di sviluppare e valutare i contenuti secondo classi d'appartenenza per ottimizzare dei modelli probabilistici d'accadimento.

Più persone o gruppi di studio potranno partecipare allo sviluppo di una piattaforma unica e coordinata atta a raccogliere nel tempo metadati significativi e più il sistema crescerà affinandosi.

Lo sforzo congiunto tendenzialmente serve a formulare un'elaborazione per l'estrazione dei componenti meta-dati dai sub-sistemi in un contesto di gestione automatizzata della situazione come sensibilizzazione e supporto alle decisioni.

E' facile portare ad esempio il comportamento di un soggetto umano o lo studio dello stesso nella folla. Sistemi di tracking ed analisi del

comportamento difficilmente sbagliano a riconoscere un ladro d'auto mentre si aggira tra le vetture di un parcheggio, due persone mentre si scambiano una busta, l'aumento repentino dal passo alla corsa per un borseggio o lo svenimento per un malore.

Proviamo a calare l'applicazione anche in contesti diversi da quello criminologico, in mano alla protezione civile, ai vigili del fuoco, agli studiosi del CNR e meteorologi, il "prodotto" audio-visivo acquisito, in aggiunta alla sorveglianza e la sicurezza, rappresenta un'utile fonte d'informazioni in caso di conservazione e analisi automatica nella pianificazione urbana per l'ottimizzazione delle risorse del territorio e dei mezzi di supporto, per esempio il monitoraggio ambientale a tutela dell'habitat, del territorio o dell'ecosistema o una forma d'aiuto e intervento per persone disabili, anziani o comunque in difficoltà; si potrebbero anticipare le tempeste o i cicloni, lo smottamento del terreno e le frane, gli incendi boschivi, slavine e valanghe, i campi d'applicazione sono veramente vasti.

Tornando all'origine del progetto, creare un modello di riferimento è fondamentale per i due tipi di apprendimento: da un lato, la sicurezza, aiuto agli operatori e *decision makers* fornendo, oltre le immagini e i dati sensoriali, anche una descrizione contestuale. D'altra parte, il contenuto della conoscenza, caratterizzata da un livello base di eventi primitivi (ad esempio, oggetti e traiettorie), seguita da uno strato superiore di eventi di tipo semantico, in grado di fornire conoscenze utili per una visione più complessa di eventi e le relazioni che tra queste interagiscono.

4. Autoapprendimento video audio.

I DVR più evoluti sono dotati di Object Detection ovvero la possibilità di riconoscimento del lascito di un oggetto nella scena inquadrata (borsa con un ordigno alla stazione) o l'asportazione di un oggetto dalla scena (un vaso in un museo, una collana dal gioielliere), quello che manca è il riconoscimento dell'oggetto e la sua classificazione.

Le altre informazioni da catalogare devono tenere conto della non staticità, quindi la distanza e velocità di spostamento, oltre la traiettoria ed il dimensionamento rispetto al background. Le registrazioni acquisite da telecamere sono scomposte e segmentate separando l'individuazione di un percorso anomalo rispetto a quanto pre-classificato come normale comportamento.

La differenza tra un "rivelatore" di sagoma umana utilizzabile per contare quanti entrano o escono è sostituito da un auto apprendimento del sistema che impara dall'esempio precedente, costruendo un modello 3D (tri-dimensionale), autocostruendosi un modello percettivo evoluto.

Qualsiasi oggetto che vibra in un mezzo elastico (ad esempio, aria, acqua, materiale composito, ecc.) può produrre un suono udibile. Il suono è in realtà un disturbo fisico che si propaga attraverso qualsiasi mezzo elastico. Queste variazioni di pressione o compressioni delle particelle d'aria nelle vicinanze della sorgente sonora causano coesione delle forze che tendono a ripristinare il normale stato e la pressione ne propaga le oscillazioni.

L'analisi della "scena uditiva" si riferisce al compito cognitivo di coinvolgere l'orecchio nell'individuazione e nella costruzione della scena e alla presenza di oggetti da miscugli di suoni

presenti in un complesso acustico ambientale più vasto.

Provate a chiudere gli occhi e sperimentare come il sistema uditivo sia responsabile di individuare quanti suoni sono presenti nell'ambiente (ad esempio, il numero di fonti), da dove provengono (ubicazione delle sorgenti) e che cosa significa (vale a dire, la natura e l'interpretazione delle fonti).

E' comunemente accettato che l'analisi uditiva della scena si svolge in due fasi. In primo luogo, il segnale è scomposto in elementi discreti sensoriali su un piano temporale, appunto tempo-frequenza prodotto dalla fonte sonora, quindi si suddividono i "flussi" in un processo percettivo denominato raggruppamento.

Ascoltare si traduce nell'isolamento e nel riconoscimento delle sorgenti sonore in un complesso di suoni nell'ambiente. I flussi audio sono il concetto centrale per l'analisi della scena uditiva rappresentando le diverse sorgenti, consentendone l'individuazione.

Il nostro cervello è in grado di creare forma mentis acquisite da sensazioni o comunque elementi sensoriali attraverso un complesso meccanismo che governa l'elaborazione e ne raggruppa gli elementi. Il dominio visivo così come quello uditivo sono in grado di formulare catalogazioni sensoriali sui singoli oggetti nell'ambiente, quindi ugualmente applicabili ed associabili per l'elaborazione di una scena più complessa.

Va da sé che uno strumento diventa interattivo quando si ha la possibilità di programmarlo, per calarlo nell'applicazione ontologica specifica. Perché sia effettivamente efficace ed efficiente,

serve una corretta e puntuale configurazione al fine di renderlo usabile per l'utente finale in modo produca i risultati attesi.

L'acronimo TVCC è rimasto, ma il prodotto ha subito una tale e rilevante metamorfosi che poco ha a che fare con il vecchio concetto di televisione a circuito chiuso, le applicazioni vanno ben oltre la video sorveglianza. Il monitoraggio industriale sui cicli di produzione automatizzata o il controllo dei robot, la prevenzione delle rotture strutturali attraverso telecamere termiche hanno portato la tecnologia a far beneficiare l'evoluzione anche nel campo della sicurezza.

Chiudendo con un esempio applicato ai bagagli che imbarchiamo sui voli intercontinentali, questi passando sotto uno scanner a raggi X collegato ad un ingresso di un DVR/NVR, che ne acquisisce e cataloga il contenuto inviando i filmati allo scalo d'arrivo o a quelli intermedi dove un ulteriore ed analogo sistema, oltre a "visionarne" l'interno ne confronta il contenuto rispetto allo scanning di partenza al fine di verificare non tanto l'asportazione di oggetti, bensì l'inserimento di armi, ordigni, droga durante la fase di trasbordo.