Alessia Saggese(*)

Dal video all'audio:
audio sorveglianza intelligente

Una violenza in un bagno pubblico; una telecamera di sorveglianza installata all'interno del bagno; un algoritmo di analisi video che elabora la sequenza di immagini e rileva in modo automatico l'evento di interesse (la violenza, nel caso specifico); una sirena che suona e le forze dell'ordine addette alla sicurezza che intervengono prontamente. Scienza o Fantascienza? Molto probabilmente fantascienza, se non altro perché per motivi di privacy non è consentito di installare all'interno di un bagno alcuna telecamera di sorveglianza. Ma non solo. Perché, pur potendo, significherebbe installare una camera in ciascun bagno. E potrebbe essere tutt'altro che banale realizzare un algoritmo di analisi video capace di elaborare in modo automatico le immagini (magari indipendentemente dal posizionamento della telecamera) per il rilevamento di un evento di questo tipo. In questo contesto, quale altra soluzione tecnologica ci viene incontro? Scopriamolo insieme.

(*) Account manager @ A.I. Tech **www.aitech.vision**

Per anni quando si è parlato di analisi audio si è pensato a problemi quali lo speech recognition o lo speaker identification. Ma forse in meno sono a conoscenza del fatto che l'audio analisi può significare molto altro. Può infatti significare analizzare in modo automatico il flusso audio acquisito da un microfono, identificando eventi di interesse, ad esempio legati al mondo della sorveglianza: spari, vetri rotti, o ancora urla.
Eventi di questo tipo, evidentemente, sono tutt'altro che semplici da rilevare da un algoritmo che deve analizzare le immagini acquisite da una telecamera. Sono meno difficili, invece, da rilevare analizzando il flusso audio. Analisi audio che non deve quindi essere considerata solo in contrapposizione con l'analisi video, ossia che non deve essere considerata solo una alternativa all'analisi video, per quegli ambienti in cui la camera magari non può essere installata. Questa infatti può essere considerata una sorta di add-on alla più tradizionale analisi video. In questa direzione si sono spinti alcuni camera manifacturer, che hanno deciso di investire in questa direzione aggiungendo alle più tradizionali funzionalità di analisi video di base anche l'analisi audio.

ANALISI AUDIO

Vi ricordate la differenza tra motion detection (che molti ancora oggi confondono con analisi video) e analisi video intelligente, di cui abbiamo discusso in altri numeri di a&s Italy[1] ? Bene, possiamo dire che oggi lo stesso accade nell'analisi audio. Anche in questo caso, infatti, l'elaborazione automatica del segnale può avvenire a differenti "livelli", più o meno "pregiati". Il più semplice, ma anche meno efficace (quello che potremmo in qualche modo considerare l'equivalente del motion detection nel settore video), è una semplice soglia sul "volume". Se il volume supera un determinato valore soglia (scelto dall'operatore umano durante una preliminare fase di configurazione), allora il sistema genera un allarme. Un po' come dire: se ci sono troppi pixel "in movimento", la cui differenza rispetto al fotogramma precedente è superiore ad una certa soglia (espressa in valore assoluto o in percentuale rispetto all'occupazione di una area), allora il sistema genera un allarme. Perché questo meccanismo non funziona (o quantomeno funziona male)?

PROBLEMATICHE

Per due motivi principali. Il primo: non tutti gli eventi audio "sopra-soglia" corrispondono necessariamente a eventi di interesse. Si pensi, ad esempio, alla possibilità di effettuare una analisi audio di questo tipo all'interno di una stazione. Ciascun treno in transito o in frenata finirebbe per generare un allarme. O meglio, almeno un allarme. Ovviamente si tratterebbe di allarmi non di interesse ai fini della sicurezza. Esattamente ciò che succede ai sistemi di motion detection quando si verifica una improvvisa variazione di illuminazione.
Secondo motivo: il volume non può essere una proprietà "distintiva". Questo dipende dal fatto che il volume dipende dalla distanza a cui ci posizioniamo rispetto alla sorgente (ossia dall'oggetto che genera il suono). Il volume che un essere umano percepisce, infatti, è tanto maggiore quanto minore è la distanza dalla sorgente. Uno sparo a un metro di distanza ci sembra molto più "forte" che lo stesso sparo (quindi con la stessa intensità) a cento metri di distanza. A questo si aggiunge il fatto che all'evento di interesse possono "sommarsi" altri eventi di non interesse, il cosiddetto rumore di fondo: un evento che parte dalla sorgente (lo sparo di una pistola, ad esempio) può combinarsi con gli altri suoni presenti nell'ambiente, fino ad arrivare al dispositivo di acquisizione (il nostro orecchio, o magari un microfono). E un sistema automatico deve essere in grado di distinguere questi differenti contributi, anche se si sono aggiunti a distanze differenti e in istanti temporali differenti. Un po' come rilevare la presenza di un intruso che entra nella scena e si nasconde (anche se parzialmente) dietro un albero.

PROPRIETÀ DISTINTIVE DEL SUONO

Pertanto, al fine di identificare eventi audio di interesse, gli algoritmi di analisi audio si basano su proprietà distintive del suono, alcune caratteristiche salienti tali da distinguere (anche a differenti distanze) le varie tipologie di classi di interesse. Così come accade per l'analisi delle immagini (l'avevamo scoperto con l'analisi dei volti in a&s Italy n. 45/2017), tali caratteristiche sono utilizzate da un software intelligente di Machine Learning, in grado di apprendere automaticamente attraverso una serie di campioni di esempio (ossia una serie di suoni di urla, spari, vetri rotti ad esempio). Software *intelligente* poiché non è l'esperto a codificare le regole (e non è neanche l'operatore umano a scegliere le soglie sul volume), bensì è il sistema ad apprenderle automaticamente grazie a dei campioni di esempio di audio di interesse.
Un altro esempio di come le tecnologie di Machine Learning possono supportare l'operatore nel compito delicato (e a tratti noioso) di monitorare ambienti sensibili.

[1] Vedi a&s Italy n. 30/2014