Università degli Studi di Salerno

Dottorato di Ricerca in Informatica e Ingegneria dell'Informazione
Ciclo 30 – a.a. 2016/2017

TESI DI DOTTORATO / PH.D. THESIS

Real-time face analysis for gender recognition on video sequences

ANTONIO GRECO

SUPERVISOR: **PROF. MARIO VENTO**

PHD PROGRAM DIRECTOR: **PROF. PASQUALE CHIACCHIO**

Dipartimento di Ingegneria dell'Informazione ed Elettrica
e Matematica Applicata
Dipartimento di Informatica

L’attività di ricerca è nata con l’obiettivo di identificare un metodo per riconoscere il genere di una persona analizzando in tempo reale immagini estratte da sequenze video registrate con telecamere di sorveglianza classiche. Tale compito può sembrare semplice per un essere umano, ma non lo è altrettanto per un algoritmo di computer vision. Anche su immagini di alta qualità gli algoritmi di riconoscimento automatico del genere devono essere progettati per essere in grado di effettuare una corretta classificazione di volti di età e razza diversa, con pose e dimensioni differenti, in presenza occlusioni e così via.
Quando l’elaborazione avviene in tempo reale su immagini acquisite in ambienti reali, le difficoltà aumentano in manera considerevole. Infatti, in scenari realistici le persone sono inconsapevoli della presenza della telecamera e possono effettuare movimenti bruschi che, insieme alla bassa qualità delle immagini di sorveglianza, rendono i volti ancora più rumorosi e caratterizzati da motion blur, variazioni di orientamento e dimensioni differenti. Inoltre, la necessità di associare una singola decisione ad ogni persona (e non di effettuare soltanto una classificazione per ogni volto) in tempo reale, impone di progettare un rapido algoritmo per il riconoscimento del genere, capace di identificare una persona in frame differenti e di effettuare la classificazione rapidamente.
Il vincolo sul tempo reale diventa ancora più stringente se si considera che uno degli obiettivi di questa attività di ricerca è la progettazione di un algoritmo adatto per un’architettura di embedded vision.
Infine, il compito è reso ancor più complicato dalla mancanza di benchmark e protocolli standard per la valutazione degli algoritmi di riconoscimento del genere.
Nella tesi, l’attenzione si è concentrata in primo luogo sull’analisi di immagini di alta qualità, non estratte da sequenze video, in modo da identificare le feature più adatte per il riconoscimento del genere. A tale scopo, un algoritmo di allineamento è stato applicato per normalizzare la posa delle facce e per ottimizzare le prestazioni delle fasi successive di elaborazione. Su tali volti allineati sono stati applicati due multi-esperti per il riconoscimento del genere.
Il primo multi-esperto combina le decisioni di tre classificatori addestrati con valori dei pixel, istogrammi LBP e feature di HOG. Tali classificatori sono in grado di prendere la propria decisione analizzando rispettivamente il colore, la trama e la forma del volto. Le decisioni dei singoli esperti sono state combinate con una votazione a maggioranza pesata, che tiene conto delle performance dei singoli esperti sulla classe in esame. Tale regola ha dimostrato, nell’analisi sperimentale, di essere la più adatta per fondere le decisioni dei tre classificatori.
Il secondo metodo combina invece un classificatore basato sui filtri COSFIRE, ovvero feature di forma addestrabili fornendo in ingresso parti del volto, e un altro esperto che prende la sua decisione estraendo i descrittori SURF in particolari punti del volto, detti facial landmarks. La complementarità dei due tipi di features è stata dimostrata con un’analisi sperimentale e le decisioni dei due classificatori sono state combinate con un secondo livello di classificazione in cascata.
L’analisi sperimentale su immagini è stata effettuata sui dataset GENDER-FERET e LFW con un protocollo standard, così da rendere possibile un onesto confronto delle prestazioni. Tale valutazione dimostra che il classificatore basato su COSFIRE e SURF ottiene le migliori performance su entrambi i dataset (94.7% su GENDER-FERET e 99.4% su LFW), anche rispetto ad altri metodi allo stato dell’arte. Inoltre, anche le prestazioni ottenute dal multi-esperto basato su raw, LBP e HOG sono risultate molto elevate (93.0% su GENDER-FERET e 98.4% su LFW).
In seguito all’analisi preliminare svolta sulle immagini, l’attenzione è stata spostata sui video. A tale scopo, un nuovo dataset, chiamato UNISA-dataset, è stato acquisito in diversi ambienti reali (università e supermercati) con telecamere di sorveglianza classiche. Una parte di queste immagini èstata resa disponibile pubblicamente. In queste sequenze video le persone sono inconsapevoli di essere inquadrate, dunque le immagini sono significativamente più complicate di quelle disponibili nei dataset standard.
Tale dataset è stato utilizzato innanzitutto per un’approfondita analisi del tempo di elaborazione richiesto dagli algoritmi sopra descritti. L’attività di profiling dimostra che l’algoritmo di allineamento dei volti è molto oneroso e non è adatto per l’elaborazione in tempo reale, così come il

calcolo dei descrittori basati su SURF e COSFIRE. Considerando che i pixel dell'immagine non sono affidabili se i volti non sono allineati, l'analisi ha permesso di concludere che il miglior classificatore per il ricoscimento del genere in tempo reale è quello basato su HOG, che ha dimostrato maggiore efficienza ed efficacia rispetto a quello basato su LBP. Infine la valutazione dei tempi ha dimostrato che, nonostante il classificatore basato su feature di HOG sia capace di elaborare immagini in tempo reale sulle classiche architetture server, non è in grado di fare lo stesso su sistemi di embedded vision a basso costo.

Alla luce di queste osservazioni, l'unica soluzione che permette di raggiungere l'obiettivo è un'architettura multi-sensore. Essa è costituita da una telecamera zenitale dedicata al conteggio delle persone in tempo reale, da un'altra telecamera installata in posizione frontale per inquadrare i volti delle persone che si avvicinano e da un dispositivo embedded a basso conto per effettuare in tempo reale il riconoscimento del genere. In tale architettura, la telecamera che effettua conteggio persone invia una notifica all'altra telecamera quando identifica il passaggio di almeno una persona, fornendo informazioni sulla posizione in cui potrebbero essere rilevati i volti. In tal modo, l'algoritmo di riconoscimento del genere può essere applicato soltanto ad una sottoregione dell'immagine, riducendo in maniera significativa il carico computazionale e i falsi positivi.

Tale architettura consente di elaborare più immagini ad alta risoluzione, raggiungendo contemporaneamente l'obiettivo di massimizzare l'accuratezza e riconoscere il genere in tempo reale su dispositivi embedded a basso costo. Inoltre, l'algoritmo è in grado di effettuare il tracking delle persone, in modo da associare la stessa identità alle classificazioni dei volti della stessa persona in frame diversi, ottenendo un'accuratezza superiore al 90%.

Un'approfondita nalisi sperimentale sull'UNISA-dataset dimostra l'efficacia del motodo proposto e la sua adeguatezza per il riconoscimento del genere in tempo reale.