

# Implementazione di un sistema di riconoscimento facciale basato sul metodo di Viola e Jones

Lorenzo Cioni  
Apprendimento Automatico  
lore.cioni@gmail.com

## Abstract

*Il riconoscimento facciale gioca un ruolo sempre più determinante in molti settori applicativi, come ad esempio sistemi di sorveglianza, software di editing o applicazioni per l'acquisizione di immagini digitali nei vari dispositivi. In molti casi poi, a causa di alcune specifiche, potrebbe essere necessario dover effettuare dei riconoscimenti in tempo reale, come ad esempio in un video. L'obiettivo di questo elaborato di implementare un sistema di riconoscimento facciale basato sul metodo proposto da P. Viola e M.J. Jones.*

## Future Distribution Permission

The author(s) of this report give permission for this document to be distributed to Unifi-affiliated students taking future courses.

## 1. Introduzione

Il riconoscimento facciale è stato oggetto di ricerca dai primi anni novanta come parte del ramo del *pattern recognition* e della *object detection*. Il riconoscimento di oggetti e di facce presentano circa gli stessi problemi, ma in particolare, nel secondo caso, i problemi più determinanti sono dati dalla differenza di illuminazione, la scala e la rotazione del volto.

A questi problemi classici del riconoscimento di oggetti se ne aggiungono altri come la gran varietà di espressioni del volto umano o alla presenza di elementi che ne potrebbero compromettere il riconoscimento, come ad esempio una particolare capigliatura o semplicemente un paio di occhiali.

L'obiettivo di questo elaborato è implementare il metodo di riconoscimento facciale proposto da Viola e Jones [1] e valutarne le criticità e perfor-

mance.

## 2. Il metodo di Viola e Jones

Il metodo di riconoscimento facciale proposto da Viola e Jones si basa su dei semplici classificatori lineari che operano su delle feature specifiche dell'immagine. Le feature utilizzate sono di tipo binario, molto buone per caratterizzare parti dell'immagine in cui si evidenzia un maggior contrasto: nel caso dei volti queste zone sono tipicamente il naso, gli occhi, etc.

Per consentire di calcolare rapidamente questo tipo di features, Viola e Jones hanno introdotto una nuova rappresentazione dell'immagine chiamata *immagine integrale*. Una volta calcolate queste si procede alla selezione di un loro sottoinsieme utilizzando una versione modificata dell'algoritmo AdaBoost.

Infine, per la classificazione, viene utilizzata una *cascata di classificatori*. I classificatori sono suddivisi in più livelli rendendo possibile il riconoscimento in tempo reale.

## 3. Features

Uno dei componenti essenziali del metodo di Viola e Jones è l'utilizzo di un sistema di feature molto semplice ed un modo altrettanto veloce per calcolarle.

Viene inizialmente definita una *detection window* di dimensione prefissata (generalmente 24 x 24) e si fa scorrere quest'ultima lungo tutta l'immagine da analizzare alla ricerca di facce. In ciascuna posizione vengono analizzate un'insieme di feature locali, in questo caso le

Haar-like.

### 3.1. Le features Haar-like

Le features Haar-like consistono in 2 o pi rettangoli bianchi o neri adiacenti all'interno della *detection window*. A partire da una configurazione di rettangoli viene calcolata la differenza delle somme dei pixel sottostanti una regione bianca e quelli sottostanti ad una regione nera. In questo modo possibile identificare zone in cui vi un contrasto nella direzione indicata dalla posizione dei rettangoli. Una zona omogenea infatti dar valori circa prossimi a 0.

Sia  $X$  un'immagine. Chiamiamo  $W$  l'insieme dei pixel sottostanti una regione *bianca* e  $B$  l'insieme dei pixel sottostanti una regione *nera*. Il valore della feature ottenuto tramite

$$f = \sum_{i \in W} x_i - \sum_{j \in B} x_j$$

Le features utilizzate nel metodo di Viola e Jones sono di 5 tipi diversi, come mostrato in Figura 1.

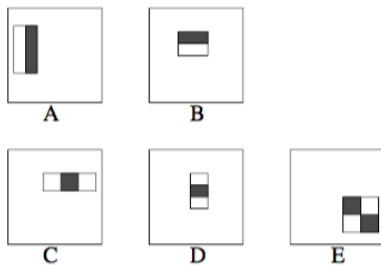


Figure 1. I 5 tipi diversi di Haar-like features utilizzati

Il numero totale di features di tutti e 5 i tipi scattati e traslati all'interno di una *detection window*  $24 \times 24$  di 162,336. Come nell'implementazione di OpenCV sono state per considerate solo le features con un area superiore a 32 pixels: questo per evitare di andare a considerare delle regioni troppo piccole e irrilevanti ai fini del riconoscimento. Con queste particolari restrizioni il numero totale di features si riduce a 108,490.

### 3.2. Immagini integrali

Al fine di poter riconoscere i volti in tempo reale necessario poter valutare queste features in poco tempo. Per fare questo Viola e Jones hanno introdotto una nuova rappresentazione dell'immagine chiamata *immagine integrale*.

L'immagine integrale costruita utilizzando una matrice della stessa dimensione dell'immagine originale con associata a ciascun elemento la somma dei valori dei pixel che si trovano sopra e a sinistra del pixel corrispondente.

Sia  $i$  l'immagine originale e  $I$  la sua rappresentazione integrale. Un generico elemento della matrice integrale si trova con:

$$I(x, y) = \sum_{x' \leq x, y' \leq y} i(x', y')$$

E' possibile giungere alla costruzione di questa matrice iterativamente utilizzando le seguenti due formule:

$$s(x, y) = s(x, y-1) + i(x, y)$$

$$I(x, y) = I(x-1, y) + s(x, y)$$

A questo punto, dato un

## 4. Boosting

### 4.1. Selezione delle features

## 5. Classificazione

### 5.1. La cascata di classificatori

### 5.2. Post-processing

## 6. Risultati

### 6.1. Creazione del dataset

### 6.2. Esperimenti

## 7. Conclusioni

## References

- [1] P. Viola and M. J. Jones. Robust real-time face detection. *Int. J. Comput. Vision*, 57(2):137–154, May 2004.