



UNIVERSITÀ DEGLI STUDI DI BARI ALDO MORO

DIPARTIMENTO DI INFORMATICA
CORSO DI LAUREA IN INFORMATICA

TESI DI LAUREA IN
SISTEMI AD AGENTI

RICONOSCIMENTO BIOMETRICO DI GENERE ED ETNIA ATTRaverso l'ANALISI DEL VOLTO

RELATORE:

Prof.ssa Berardina De Carolis

CO-RELATORE:

Dott. Giuseppe Palestro

LAUREANDO:
Andrea Giove

ANNO ACCADEMICO 2014/2015

Indice

1	Introduzione	5
1.1	Computer Vision	7
1.2	Riconoscimento delle biometriche	8
1.3	Lavori correlati	9
2	Analisi del volto	12
2.1	Metodi per la gestione della dimensionalità	12
2.1.1	Principal Component Analysis	12
2.1.2	Linear Discriminant Analysis	15
2.2	Descrittori	18
2.2.1	Local Binary Patterns	18
2.3	Trattamento del colore	21
2.3.1	RGB	21
2.3.2	YCbCr	22
2.3.3	HSV	23
3	Classificatori per il riconoscimento facciale	25
3.1	Haar Cascade	26
3.2	k-Nearest Neighbors	29

3.3	Support Vector Machines	30
4	OpenCV	33
4.1	Caffe	34
4.1.1	Deep learning	34
5	Riconoscimento biometrico di genere ed etnia attraverso l'analisi del volto	38
5.1	Requisiti e configurazione hardware	38
5.1.1	Configurazione della macchina fisica	39
5.1.2	Configurazione della macchina virtuale	39
5.2	Dati di addestramento	40
5.2.1	Fonti	40
5.2.2	Operazioni di pre-processamento	42
5.3	Architettura del sistema	45
5.3.1	Descrizione della struttura	45
5.3.2	Dinamica di funzionamento	46
5.4	Riconoscimento del genere	50
5.4.1	Esempi di addestramento	50
5.4.2	Scelte implementative	51
5.5	Riconoscimento dell'etnia	54
5.5.1	Esempi di addestramento	54
5.5.2	Scelte implementative	55
5.6	Applicazione Web	58
5.6.1	Interfaccia utente	58
6	Risultati sperimentali	61

Indice

6.1	Riconoscimento del genere	61
6.1.1	Test sugli esempi	61
6.1.2	Test in tempo reale	70
6.2	Riconoscimento dell'etnia	73
6.2.1	Test sugli esempi	73
6.2.2	Test sulle immagini	78
7	Conclusioni e sviluppi futuri	82

1 Introduzione

Nell'ultimo decennio, il campo delle biometriche ha suscitato grande interesse in virtù delle numerose possibilità applicative in ambito accademico e industriale[1]. La videosorveglianza, l'analisi dei contenuti multimediali, le pratiche forensi più innovative sono solo alcuni fra i settori che stimolano lo sviluppo di tecniche avanguardistiche per il riconoscimento automatico dei parametri biofisici per i propri specifici scopi.

In questo contesto, il ramo della biometria noto come "*Soft biometrics*" ha visto la nascita di numerosi approcci alla classificazione di caratteri demografici come il genere, l'etnia e l'età, a partire dall'analisi di immagini e sequenze video. In particolare, le ricerche si sono concentrate sullo studio del volto umano e della sua rappresentazione, al fine di individuare i metodi più efficaci per l'identificazione dei soggetti considerati.

Il presente lavoro di tesi propone un framework moderno e flessibile per il riconoscimento dei parametri biometrici. Particolare attenzione è stata rivolta alla realizzazione di un sistema scalabile, capace di offrire risultati accurati sia in ambienti ad alte prestazioni sia in contesti di esecuzione a risorse più limitate. Nella prospettiva di utilizzare il framework anche sulle macchine più performanti (come ad esempio un server), sono stati esaminati e testati alcuni modelli costruiti impiegando strategie di deep learning.

1 Introduzione

Per sviluppare il sistema è stata utilizzata OpenCV, attraverso la sua interfaccia C++: questa è una vastissima libreria open source legata al mondo della computer vision e offre innumerevoli funzioni e utilità per lavorare analiticamente sulle immagini e i prodotti multimediali in genere. La libreria, giunta alla versione 3.1, mette a disposizione diverse implementazioni degli algoritmi più diffusi nell'ambito del riconoscimento facciale e rappresenta un punto di riferimento per l'intera area di ricerca.

A compimento del lavoro di sperimentazione, è stato realizzato un sito web con l'intento di rendere fruibili le funzionalità sviluppate a chiunque avesse voglia di provare il sistema con il proprio materiale.

L'esposizione degli argomenti è così strutturata: nel *capitolo 1* è illustrata una panoramica generale della computer vision e del campo delle biometriche, discutendo le ricerche precedenti rivolte al riconoscimento automatico dei parametri biometrici. Nel *capitolo 2* sono presentati diversi approcci per l'analisi del volto, dai metodi di riduzione della dimensionalità, all'uso di descrittori, al trattamento di un aspetto potenzialmente chiave come quello del colore. Tali metodi condividono il medesimo fine di estrapolare caratteristiche di alto livello a partire dalla rappresentazione in pixel di un'immagine. Una volta individuate, tali caratteristiche vengono adoperate per addestrare un opportuno classificatore: il *capitolo 3* tratta i principali classificatori presenti in letteratura per assolvere il compito di riconoscimento facciale. Nel *capitolo 4* sono approfonditi la libreria OpenCV e il framework Caffe, usato per addestrare i modelli di deep learning integrati. Nel *capitolo 5* vengono descritti i passi effettuati per sviluppare il sistema, dalla selezione degli esempi di addestramento, passando per la definizione dell'architettura, alle scelte fatte, di metodo e implementazione, per riconoscere il genere e l'etnia. A

questo segue la fase sperimentale, puntualmente riportata nel *capitolo 6*, e comprendente i test per verificare l'accuratezza dei metodi proposti. Infine, nel *capitolo 7*, si tirano le somme dell'intero lavoro svolto, esaminandone criticamente pregi e difetti; sono quindi indagati auspicabili sviluppi futuri per il sistema.

1.1 Computer Vision

La *computer vision* è un campo dell'informatica che studia metodi per l'acquisizione, l'elaborazione, l'analisi e la comprensione delle immagini e, in generale, di dati ad elevata dimensionalità provenienti dal mondo reale, al fine di produrre informazione numerica o simbolica. Per "comprensione", in questo contesto, si intende la trasformazione delle immagini in descrizioni del mondo che possano essere utili per esercitare alcune forme di ragionamento - e di azione - sugli oggetti osservati.

In quanto disciplina scientifica, la computer vision indaga la teoria alla base dei sistemi artificiali che estraggono informazioni a partire dai dati visuali, i quali possono assumere svariate forme: foto, semplici sequenze video fino ad arrivare ai complessi output forniti da uno scanner per uso medico.

Alcune importanti aree della computer vision riguardano il tracciamento video, il riconoscimento di oggetti, l'apprendimento automatico, la ricostruzione di scene e il riconoscimento di eventi.

1.2 Riconoscimento delle biometriche

Le *biometriche* si riferiscono alle metriche relative a caratteristiche proprie dell'essere umano. Gli identificatori biometrici sono attributi distintivi e misurabili usati per etichettare e descrivere gli individui [2]. Sono spesso classificati in identificatori fisiologici e identificatori comportamentali.

- **Identificatori fisiologici:** si riferiscono ad aspetti relativi alla forma del corpo come le impronte digitali, il riconoscimento facciale, il DNA ecc.;
- **Identificatori comportamentali:** riguardano alcuni schemi di comportamento individuali come l'andatura o la voce;

La scelta delle opportune biometriche dipende fortemente dal tipo di sistema applicativo che si vuole costruire: alcune biometriche saranno migliori di altre in base ai livelli di sicurezza e comodità richiesta. Nessuna singola misura è capace di soddisfare tutti i requisiti di una qualunque possibile applicazione.

L'obiettivo principale di un sistema biometrico è quello di identificare univocamente un individuo.

Le cosiddette *soft biometrics* fanno invece riferimento alle caratteristiche fisiche e comportamentali derivate dal modo in cui l'essere umano distingue i suoi simili. Sono generalmente usate in modo complementare ai sistemi biometrici per il riconoscimento dell'identità al fine di migliorarne le prestazioni grazie all'analisi e la scoperta di attributi come il genere, l'etnia, l'età, il colore degli occhi ecc. Jain ha ultimamente riscoperto i tratti soft biometrics come insieme di caratteristiche che trasportano informazioni personali riguardo l'individuo in considerazione, sebbene questi tratti non siano adeguati a

una piena identificazione in quanto mancano di potere discriminatorio e, per alcuni, permanenza.[3]

I tratti soft biometrics ereditano una parte dei vantaggi dei tratti biometrici e in alcuni casi ne estendono le potenzialità. Alcuni dei vantaggi includono la non invadenza (in generale il soggetto non deve compiere azioni specifiche per la raccolta dei dati). Altri vantaggi sono l'efficienza computazionale e l'adattamento al linguaggio naturale umano. Inoltre, non richiedono la registrazione, né il consenso né la cooperazione del soggetto osservato.[4]

In ultima analisi, la loro identificazione si presta facilmente ad arricchire numerosi contesti commerciali, pubblicitari e di indagine statistica.

1.3 Lavori correlati

L'approccio più intuitivo al riconoscimento facciale è probabilmente quello che passa attraverso l'analisi delle caratteristiche geometriche del volto. Uno dei primi sistemi automatici per il riconoscimento facciale è descritto in [5]: alcuni punti marcatori (la posizione degli occhi, delle orecchie, del naso ecc...) vengono usati per costruire un vettore di caratteristiche del singolo volto (calcolando ad esempio la distanza e l'angolo fra i punti individuati, ...); il riconoscimento effettivo è poi svolto sulla base del calcolo della distanza euclidea fra il vettore caratteristico individuato e i vettori di riferimento per ciascuna classe. Tale metodo risulta naturalmente robusto rispetto alle variazioni di luce, ma presenta l'importante problematica che l'accurata identificazione dei punti marcatori è complessa, persino utilizzando gli algoritmi allo stato dell'arte. Uno degli ultimi esperimenti che sfrutta l'approccio geometrico è presentato in [6]: un vettore di caratteristiche in

1 Introduzione

uno spazio 22-dimensionale è stato testato su grandi dataset, mostrando come i connotati geometrici del volto non procurano sufficiente informazione per il conseguimento ottimale del compito di riconoscimento.

Sono state dunque intraprese altre strade e indagati nuovi approcci. Uno di questi è quello definito *olistico*: l'immagine di un volto, normalmente rappresentata come una matrice di pixel di grandi dimensioni, viene proiettata in uno spazio a bassa dimensionalità, all'interno del quale il compito di classificazione diventa più semplice. In [7] vengono confrontate le prestazioni di diversi metodi olistici: viene dimostrato sperimentalmente come l'algoritmo Fisherfaces risulti il migliore, esibendo anch'esso una buona robustezza rispetto alle variazioni di luminosità.

Più recentemente, sono emersi metodi per l'estrapolazione di caratteristiche che operano localmente sulle immagini. Per aggirare il problema dell'elevata dimensionalità dei dati in ingresso, vengono descritte solo piccole regioni di un'immagine, e da queste estratte caratteristiche con la proprietà di conservarsi rispetto a fenomeni come l'occlusione parziale del volto, l'illuminazione e la dimensione ridotta delle immagini. Algoritmi usati per questo scopo sono le Gabor Wavelets [8], Discrete Cosinus Transform [9] e Local Binary Patterns [10].

In [11] è presentata un'indagine approfondita circa diversi metodi per la classificazione di genere, etnia ed età: sono esaminate diverse strategie di pre-processamento delle immagini, verificati i vantaggi e gli svantaggi nell'uso di dataset bilanciati e sbilanciati, sperimentati alcuni fra i più noti algoritmi e descrittori presenti in letteratura.

In [12] viene provata l'indipendenza fra i caratteri di genere ed etnia, comparando le prestazioni di modelli generici con modelli etnici specifici

1 Introduzione

addestrati in funzione del genere e viceversa, mostrando come essi siano sostanzialmente paragonabili.

Infine, le ricerche hanno evidenziato come il colore possa supportare alcuni aspetti chiave del riconoscimento facciale. In particolare, in [13] è studiato lo spazio di colore RGB: gli autori espongono un metodo per rintracciare i volti in un'immagine sulla base del colore della pelle, attraverso l'analisi e la classificazione dei singoli pixel. In [14] vengono presi in considerazione anche gli spazi di colore YCbCr e HSV/HSI dimostrandone la superiorità nel raggiungere il medesimo scopo, grazie alla loro intrinseca peculiarità di separare l'informazione sulla luce dalla tonalità di riferimento.

2 Analisi del volto

2.1 Metodi per la gestione della dimensionalità

Uno dei principali problemi nella rappresentazione delle immagini riguarda l'alta dimensionalità. Si pensi, ad esempio, ad una immagine bidimensionale, in scala di grigi, avente come lunghezza p ed altezza q : essa risiede in uno spazio m -dimensionale dove $m = pxq$, che per una piccola immagine di 100x100 pixel si traduce in una rappresentazione vettoriale in uno spazio 10000-dimensionale.

Intuitivamente, tuttavia, non tutte queste dimensioni sono egualmente utili: siamo dunque alla ricerca di metodi per estrapolare le componenti più dense di informazione.

2.1.1 Principal Component Analysis

La *Principal component Analysis* (PCA) è stata indipendentemente proposta da Karl Pearson (1901) e Harold Hotelling (1933) per trasformare un insieme di potenziali variabili correlate in un insieme più piccolo di variabili indipendenti. L'idea alla base del metodo è che un insieme di dati ad alta dimensionalità è spesso descritto da numerose variabili correlate e che

2 Analisi del volto

dunque, l'informazione rilevante risiede solo in poche dimensioni significative. Il metodo PCA trova le direzioni caratterizzate dalla più alta varianza nei dati, nominandole componenti principali.

Si illustra di seguito una descrizione sintetica del metodo¹: sia $X = x_1, x_2, \dots, x_n$ un vettore di osservazioni $x_i \in R^d$.

1. Si calcola la media μ :

$$\mu = \frac{1}{n} \sum_{i=1}^n x_i$$

2. Si calcola la matrice di covarianza S :

$$S = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)(x_i - \mu)^T$$

3. Si calcolano gli autovalori λ_i e gli autovettori v_i di S

$$Sv_i = \lambda_i v_i, i = 1, 2, \dots, n$$

4. Si Ordinano gli autovettori in senso discendente sulla base dei relativi autovalori. Le k componenti principali sono gli autovettori corrispondenti ai k autovalori più grandi.

Le k componenti principali del vettore di osservazioni x sono date da:

$$y = W^T(x - \mu)$$

dove $W = (v_1, v_2, \dots, v_k)$

La ricostruzione attraverso il metodo PCA è data invece da:

$$x = Wy + \mu$$

¹I passaggi sono presi dalla documentazione ufficiale di OpenCV.

2 Analisi del volto

Esiste ancora un problema irrisolto: si immagini di avere a disposizione 400 immagini di 100x100 pixel. Il metodo PCA risolve la matrice di covarianza $S = XX^T$ dove la dimensione di X risulterebbe pari a 10000x400. Di conseguenza, La dimensione di S sarebbe 10000x10000, andando a occupare approssimativamente 0,8 GB di memoria.

Tuttavia, sapendo che una matrice mxn con $m > n$ può avere solo $n-1$ autovettori, è possibile considerare la decomposizione di $S = X^T X$ di dimensione nxn e quindi:

$$X^T X v_i = \lambda_i v_i$$

e ottenere gli autovettori originali di $S = XX^T$ con la seguente moltiplicazione:

$$XX^T(Xv_i) = \lambda_i(Xv_i)$$

Gli autovettori risultanti sono ortogonali; affinché costituiscano anche una base ortonormale per lo spazio vettoriale occorre trasformarli in vettori di norma unitaria.

Il metodo PCA cerca una combinazione lineare di caratteristiche che massimizzano la varianza totale dei dati, senza prendere in considerazione alcuna informazione circa eventuali classi o etichette ad essi assegnate; pertanto, esiste il rischio che molta della informazione discriminante per una particolare classe venga persa nell'eseguire la proiezione. Infatti, le variazioni tra immagini dello stesso soggetto dovute all'illuminazione o all'inclinazione sono quasi sempre maggiori di quelle dovute ad un cambiamento nell'identità del soggetto.[15]

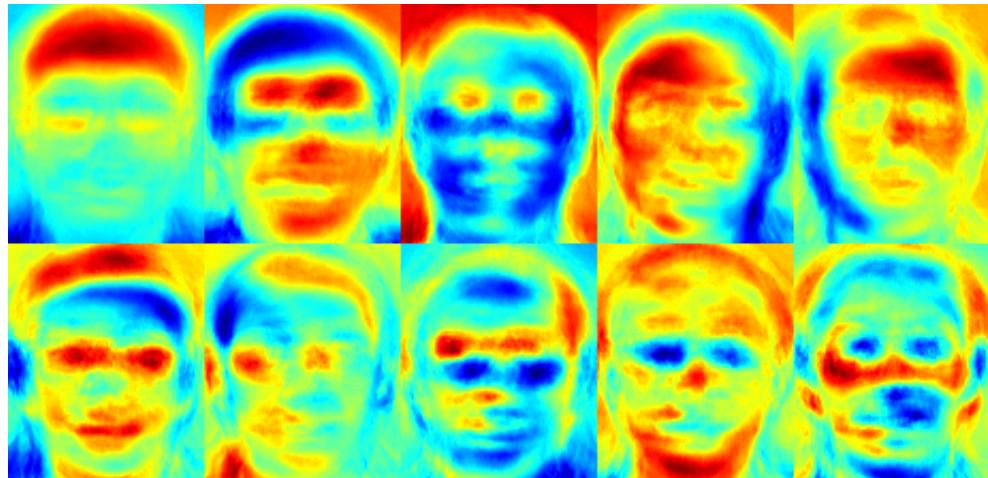


Figura 2.1: Si può notare come le caratteristiche della faccia individuate dal metodo PCA includano anche l'informazione luminosa.

2.1.2 Linear Discriminant Analysis

La *Linear Discriminant Analysis* (LDA), introdotta dallo statista R. A. Fisher, opera una riduzione della dimensionalità dipendente dalle classi di appartenenza. Allo scopo di trovare la combinazione di caratteristiche che separa più adeguatamente le classi, il metodo LDA massimizza il rapporto tra la dispersione inter-classe e quella intra-classe. L'idea è che, all'interno del sottospazio, elementi appartenenti alla medesima classe dovrebbero essere intimamente raggruppati, mentre elementi di classi differenti posti il più lontano possibile gli uni dagli altri.

Si fornisce una descrizione sintetica del metodo²:

1. Sia X un vettore di esempi appartenenti a c classi:

$$X = X_1, X_2, \dots, X_c$$

$$X_i = x_1, x_2, \dots, x_n$$

²I passaggi sono presi dalla documentazione ufficiale di OpenCV.

2 Analisi del volto

2. Le matrici di dispersione sono calcolate nel modo seguente:

$$S_B = \sum_{i=1}^c N_i (\mu_i - \mu)(\mu_i - \mu)^T$$

$$S_W = \sum_{i=1}^c \sum_{x_j \in X_i} N_i (\mu_i - \mu)(\mu_i - \mu)^T$$

3. Si considera la media totale:

$$\mu = \frac{1}{N} \sum_{i=1}^N x_i$$

4. Si considera la media delle singole classi $i \in 1, \dots, c$:

$$\mu_i = \frac{1}{|X_i|} \sum_{x_j \in X_i} x_j$$

5. L'algoritmo di Fisher cerca una proiezione W che massimizzi il criterio di separabilità delle classi:

$$W_{opt} = \operatorname{argmax}_W \frac{|W^T S_B W|}{|W^T S_W W|}$$

Una soluzione per questo problema di ottimizzazione è data risolvendo il problema generale della ricerca dell'autovalore:

$$S_B v_i = \lambda_i S_W v_i$$

$$S_W^{-1} S_B v_i = \lambda_i v_i$$

Esiste un ultimo problema da risolvere: il rango di S_W è al più $N - c$, dove N è il numero di esempi e c il numero di classi. Generalmente, il numero di esempi N è sempre più piccolo delle dimensioni dei dati in input (il numero di pixel), e dunque la matrice di dispersione S_W diventa singolare.

2 Analisi del volto

La questione può essere risolta utilizzando la Principal Component Analysis sui dati, proiettando gli esempi in uno spazio $N - c$ -dimensionale. La Linear Discriminant Analysis è poi impiegata nel sottospazio ridotto, poiché S_W non è più singolare.

Il problema di ottimizzazione può quindi essere riscritto come:

$$W_{pca} = \operatorname{argmax}_W |W^T S_T W|$$
$$W_{fld} = \operatorname{argmax}_W \frac{|W^T W_{pca}^T S_B W_{pca} W|}{|W^T W_{pca}^T S_W W_{pca} W|}$$

La matrice di trasformazione W , che proietta gli esempi nello spazio $(c - 1)$ -dimensionale è data da:

$$W = W_{fld}^T W_{pca}^T$$

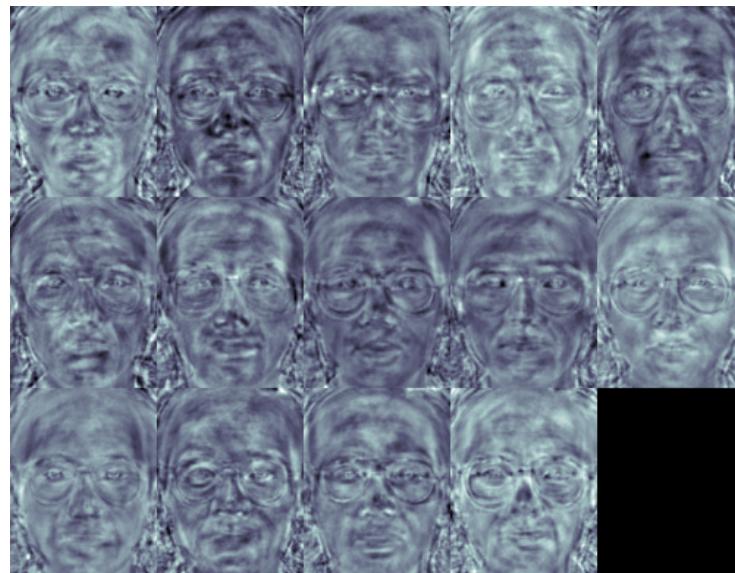


Figura 2.2: Un esempio di applicazione del metodo Fisherfaces.

2.2 Descrittori

In computer vision si parla di *descrittori visuali* per indicare sia le descrizioni di specifiche caratteristiche proprie dei contenuti visivi (forma, colore, texture, movimento, oggetti, persone ecc.) sia la classe di algoritmi o di applicazioni che si occupano di produrre tali descrizioni.

L'uso di tali descrittori rappresenta il primo passo per scoprire le connessioni tra i pixel dell'informazione digitale al fine di cogliere dati ad un più alto livello di astrazione, avvicinandosi alle associazioni e alle interpretazioni visive tipiche del cervello umano.

I descrittori visuali si suddividono in due gruppi principali: [16]

1. **Descrittori generali (General information descriptors)**: comprende descrittori di basso livello il cui compito è dare descrizioni su aspetti visivi come il colore, la forma, texture e movimento;
2. **Descrittori specifici (Specific domain information descriptors)**: prende descrittori che operano ad un livello più astratto, sforzandosi di descrivere oggetti ed eventi all'interno di una scena.

2.2.1 Local Binary Patterns

Local Binary Patterns (abbreviato *LBP*) è un tipo di descrittore visuale usato per la classificazione in computer vision. È noto in letteratura per essere particolarmente efficace nella classificazione di strutture visuali (*textures*).

Un vettore di feature LBP può essere costruito, nella sua forma più semplice, effettuando le operazioni descritte di seguito.

- Si divide l'immagine esaminata in celle (ad esempio da 16x16 pixel).

2 Analisi del volto

- Si confronta ogni pixel in una cella con ciascuno dei suoi 8 pixel vicini (quello in alto a sinistra, in alto al centro e in alto a destra, a sinistra, a destra, in basso a sinistra, in basso al centro e in basso a destra). Si scorrono i pixel in senso circolare, orario o antiorario.
- Quando il valore del pixel centrale è più grande di quello del vicino, si memorizza 0, quando è più piccolo 1. La concatenazione delle cifre così raccolte dà origine a un numero binario di 8 cifre.
- Dopo aver ottenuto questo numero per ciascun pixel in una cella, si calcola l'istogramma della frequenza associata a ciascun valore binario. Opzionalmente si effettua una normalizzazione dell'istogramma.
- Per concludere si concatenano gli istogrammi (eventualmente normalizzati) di tutte le celle in cui è stata divisa l'immagine, ottenendo un vettore di feature che descrive l'intera figura.

Il vettore di feature può a questo punto essere elaborato mediante un qualsiasi algoritmo di machine learning adatto alla classificazione di immagini, come Support Vector Machine.

2 Analisi del volto

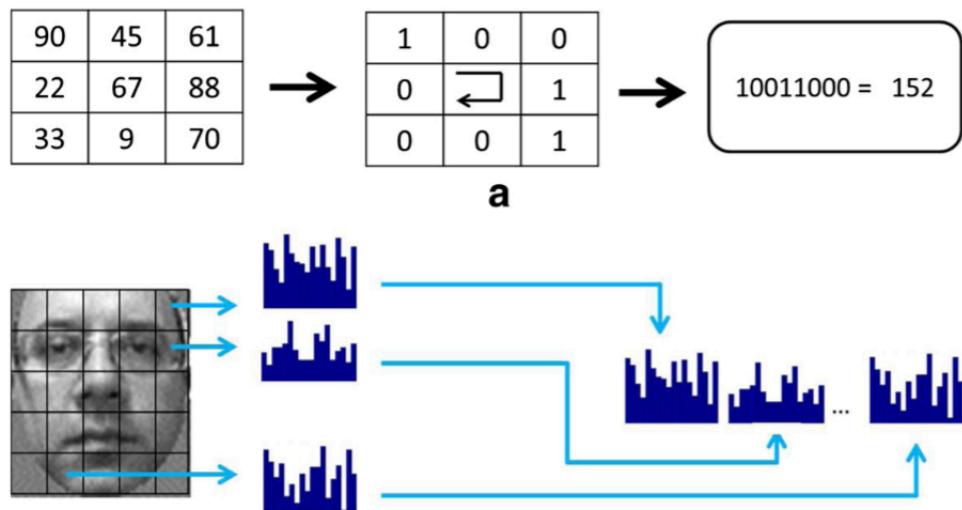


Figura 2.3: Passi per la costruzione di un descrittore LBP.

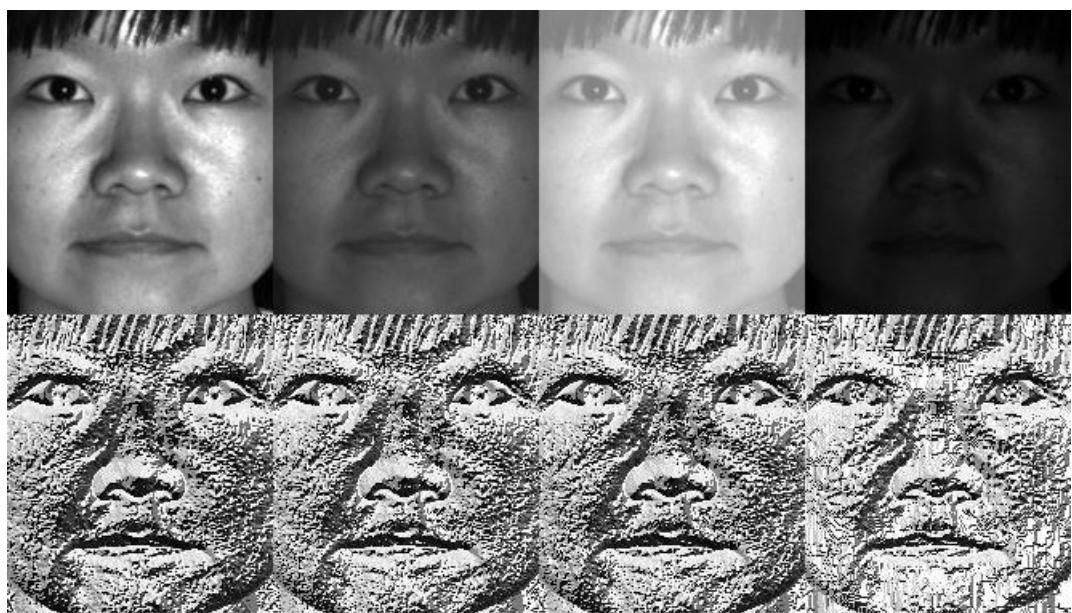


Figura 2.4: Un esempio di applicazione del metodo LBP.

2.3 Trattamento del colore

Uno *spazio dei colori* è la combinazione di un modello di colore e di una appropriata funzione di mappatura di questo modello.

Un modello di colore è un modello matematico astratto che descrive un modo per rappresentare i colori come combinazioni di numeri, - tipicamente tre o quattro valori - detti componenti colore. La funzione di mappatura definisce un insieme di regole specifiche rispetto all'utilizzo del modello. Ad esempio, Adobe RGB e sRGB sono spazi di colore diversi, pur basandosi entrambi su RGB.

2.3.1 RGB

RGB è un modello di colori le cui specifiche sono state descritte nel 1931 dalla CIE (Commission internationale de l'éclairage). Diversamente dalle immagini in scala di grigio e CMYK, tale modello è di tipo additivo e si basa sui tre colori primari rosso (Red), verde (Green) e blu (Blue), da cui il nome deriva.

Più specificamente i tre colori principali corrispondono a forme d'onda di periodo fissato, quali:

- *Rosso*, con una lunghezza d'onda di 700,47 nm;
- *Verde*, con una lunghezza d'onda di 546,09 nm;
- *Blu*, con una lunghezza d'onda di 455,79 nm;

L'uso dei tre primari non è comunque sufficiente a riprodurre tutti i colori dello spettro luminoso. Sono rappresentati solo quelli entro il triangolo definito dalla cromaticità dei primari che possono essere riprodotti tramite sintesi additiva di quantità non negative di tali colori[17].

2.3.2 YCbCr

Il termine YCbCr o Y'CbCr indica una famiglia di spazi colore usata nei sistemi video a componenti e di fotografia digitale. Y' è il componente di luma mentre Cb e Cr indicano i componenti di crominanza.

Lo Y'CbCr non è uno spazio colore assoluto, ma è un modo di codificare l'informazione RGB, e i colori realmente visualizzati dipendono dai coloranti usati dal mezzo di visualizzazione.

I segnali YCbCr sono chiamati YPbPr, e sono creati dalle corrispondenti sorgenti RGB usando due costanti K_B e K_R come segue:

$$YI = K_R R I + (1 - K_R - J_B) G I + K_B B I$$

$$P_B = \frac{1}{2} \frac{B I - Y I}{1 - K_B}$$

$$P_R = \frac{1}{2} \frac{R I - Y I}{1 - K_R}$$

dove K_B e K_R sono normalmente derivata dalla definizione del corrispondente spazio RGB.



Figura 2.5: Da sinistra verso destra: 1) L'immagine originale; 2) La componente Y; 3) La componente Cb; 4) La componente Cr.

2.3.3 HSV

Hue Saturation Value (abbreviato HSV), dall'inglese "tonalità, saturazione e valore", indica sia un metodo additivo di composizione dei colori, sia un modo per rappresentarli in un sistema digitale. Viene anche chiamato HSB da *Hue Saturation Brightness* (tonalità, saturazione e luminosità) o HSI da *Hue Saturation Intensity* (tonalità, saturazione ed intensità). È una variante del modello Hue Saturation Lightness (HSL).

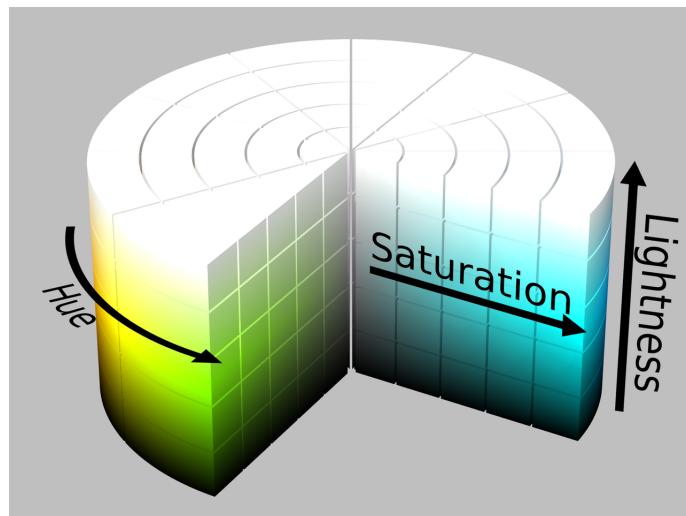


Figura 2.6: Una rappresentazione del modello HSL/HSV

La *tonalità* (H) viene misurata da un angolo intorno all'asse verticale, con il rosso a 0°, il verde a 120° e il blu a 240°, tornando al rosso a 360°. L'altezza del modello rappresenta la *luminosità* (L), ovvero la brillantezza del colore, con lo zero che rappresenta il nero e l'uno il bianco. La *saturazione* (S) ne

2 Analisi del volto

rappresenta la purezza e viene misurata da zero, sull'asse del modello, a una sua superficie.

Il modello HSL, per via della sua rappresentazione, è particolarmente orientato alla prospettiva e ai meccanismi percettivi umani.

3 Classificatori per il riconoscimento facciale

In machine learning *il problema della classificazione* consiste, dato un insieme di categorie, nell'individuare a quale categoria un oggetto osservato appartenga, sulla base di un insieme di dati di addestramento (training-set) contenente altri oggetti la cui categoria di appartenenza è nota.

Il task di classificazione fa capo alla branca del machine learning nota come *supervised learning* (apprendimento supervisionato), la cui caratteristica peculiare sta nella disponibilità di un training-set di oggetti correttamente identificati (etichettati). Esso è contrapposto al *clustering*, sua controparte non supervisionata che esercita il suo compito in assenza di esempi di addestramento.

Ad oggi si è sperimentata una gran molitudine di strategie per la risoluzione automatica del problema di classificazione, e la letteratura documenta l'utilizzo di un gran numero di algoritmi *classificatori*.

Nell'ambito dell'analisi di contenuti multimediali, e in particolare di contenuti video, quello della classificazione automatica è un problema centrale. Parecchi sono gli ambiti sperimentali in cui si è rivelata lo strumento più efficace, se non l'unico possibile, per consentire al calcolatore un discernimento

3 Classificatori per il riconoscimento facciale

autonomo sul contenuto di immagini, video e documenti acustici. Ciò si deve certamente alla natura complessa della codifica del materiale multimediale, e alle difficoltà che questa comporta nell'ideazione di altre soluzioni, basate ad esempio sull'analisi semantica dei contenuti.

La classificazione è l'attività alla base del riconoscimento dei parametri biometrici di cui si occupa il presente lavoro di tesi: ciascun task di riconoscimento è implementato utilizzando uno o più classificatori. Dall'analisi della letteratura è emerso come quelli presentati, siano *algoritmi di classificazione* più utilizzati per il riconoscimento facciale.

3.1 Haar Cascade

Un classificatore *Haar Cascade* (per esteso "cascade of boosted classifiers working with haar-like-features") è un metodo efficace ed efficiente per l'individuazione di oggetti all'interno di un'immagine, proposto da Paul Viola e Michael Jones in [18].

Il metodo adoperato riesce a raggiungere risultati paragonabili allo stato dell'arte, distinguendosi dagli approcci precedenti per l'abilità di identificare i volti molto rapidamente: dove altri sistemi usufruiscono di informazione ausiliaria (il colore, differenze fra le immagini nelle sequenze video ecc.) per ottenere alte prestazioni, l'algoritmo lavora solo con immagini in scala di grigio, lasciando la libertà di integrare sorgenti di informazione extra per migliorare ulteriormente i risultati.

Sono di seguito sintetizzati gli aspetti principali del metodo:

1. Si utilizza una rappresentazione intermedia delle immagini chiamata *immagine integrale*, calcolata dall'immagine di partenza eseguendo poche

3 Classificatori per il riconoscimento facciale

operazioni per ciascun pixel. Una volta ottenuta, permette di valutare le caratteristiche di un'immagine, a prescindere dalla dimensione o dalla zona esaminata, in tempo costante.

2. Si costruisce un classificatore prelevando un ristretto numero di caratteristiche fra quelle più rilevanti, mediante una variazione di AdaBoost [19]. Per assicurare una classificazione veloce, il processo di apprendimento deve escludere la maggior parte delle caratteristiche disponibili e concentrarsi su di un piccolo insieme di attributi critici.
3. Si combinano in successione classificatori sempre più complessi a definire una struttura a cascata che aumenta sensibilmente la velocità di individuazione, concentrando l'attenzione sulle regioni più promettenti dell'immagine. Infatti, una regione non rigettata dal classificatore iniziale viaggia lungo la sequenza, affrontando elaborazioni sempre più complesse, e potendo essere scartata in un qualsiasi passo (stadio) del procedimento.

Più specificamente, sono calcolate tre tipologie di caratteristiche. Il valore di una *two-rectangle feature* si esprime come la differenza tra la somma dei pixel all'interno delle due regioni rettangolari. Le regioni hanno stessa forma e dimensione e sono adiacenti o verticalmente o orizzontalmente. Una *three-rectangle feature* è calcolata a partire dalla somma dei due rettangoli esterni a cui viene sottratta la somma dei pixel del rettangolo centrale. Infine, una *four-rectangle feature* valuta la differenza tra coppie di rettangoli posti in diagonale.

3 Classificatori per il riconoscimento facciale

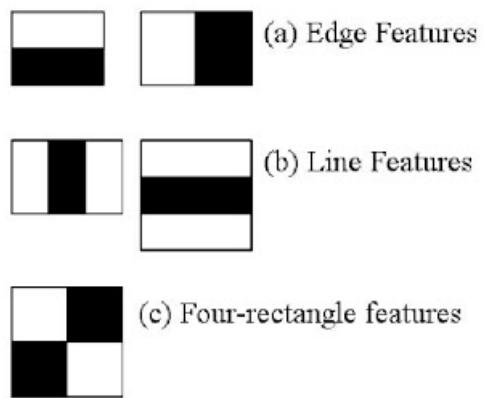


Figura 3.1: Le tre tipologie di caratteristiche elaborate dal metodo Haar Cascade.

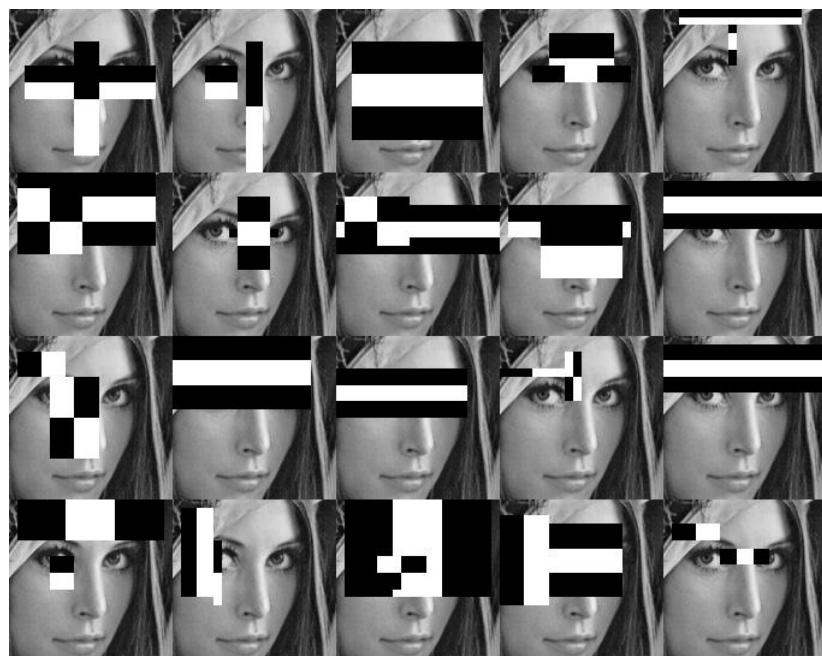


Figura 3.2: Processo di elaborazione a cascata di un classificatore Haar Cascade.

3.2 k-Nearest Neighbors

L'algoritmo *k-Nearest Neighbors* (abbreviato in *k-NN*) è un metodo non parametrico usato per risolvere problemi di classificazione e regressione. Si tratta di uno dei più semplici algoritmi di machine learning.

L'*input* dell'algoritmo è costituito dall'oggetto da classificare e dai k esempi di training a esso più vicini nello spazio delle feature. Per spazio delle feature si intende lo spazio dei vettori n -dimensionali di feature¹ numeriche che rappresentano gli esempi di training e gli oggetti da classificare. L'*output* è una classe di appartenenza: un oggetto viene classificato secondo la classe più comune tra i suoi k vicini². In alcuni casi può essere utile assegnare un peso ai vicini, così da considerare più importante il contributo dei più prossimi.

I vicini sono prelevati da un insieme di oggetti di classe nota, che si può considerare il training-set dell'algoritmo, sebbene quest'ultimo *non* richieda una fase di training esplicita. Gli esempi di training sono vettori in uno spazio di feature multidimensionale e a ciascuno di essi è associata un'etichetta di classe. La fase di addestramento consiste quindi nella sola memorizzazione dei vettori di feature e delle etichette di classe degli esempi di addestramento.

¹In ambito di machine learning e pattern recognition, una feature è una proprietà individualmente misurabile di un fenomeno osservato.

² k è un intero positivo, di solito piccolo.

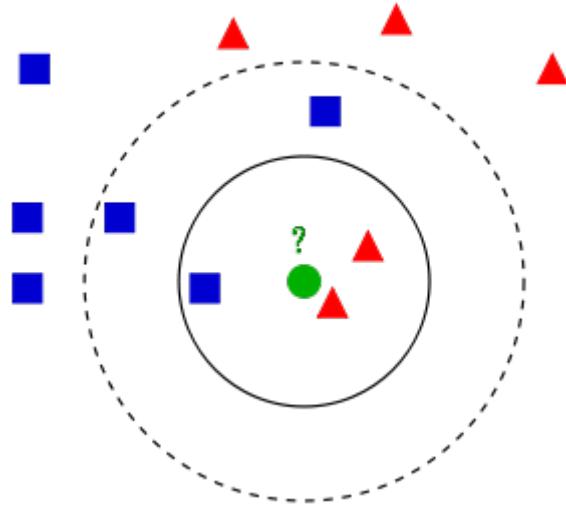


Figura 3.3: L'oggetto da classificare (cerchio verde) potrebbe essere assegnato alla prima classe (quadrati blu), o nella seconda (triangoli rossi). Se $k=3$ (cerchio con linea continua), esso è assegnato alla seconda classe perché ci sono 2 triangoli e soltanto un quadrato all'interno del cerchio. Se $k=5$ (cerchio con linea tratteggiata), l'oggetto è invece assegnato alla prima classe: ci sono 3 quadrati e soltanto due triangoli all'intero del cerchio più esterno.

3.3 Support Vector Machines

Le *macchine a vettori di supporto* (SVM, dall'inglese Support Vector Machines) sono un insieme di metodi di apprendimento supervisionato per la regressione e la classificazione di pattern, sviluppati negli anni '90 da Vladimir Vapnik ed il suo team presso i laboratori Bell AT&T.

Dato un insieme di esempi, ognuno opportunamente marcato come appartenente ad una fra due possibili categorie, un algoritmo di addestramento basato su SVM costruisce un modello attraverso un approccio lineare non

3 Classificatori per il riconoscimento facciale

probabilistico, rappresentando gli esempi come punti in uno spazio, in modo che punti appartenenti a due diverse categorie risultino sensibilmente divisi.

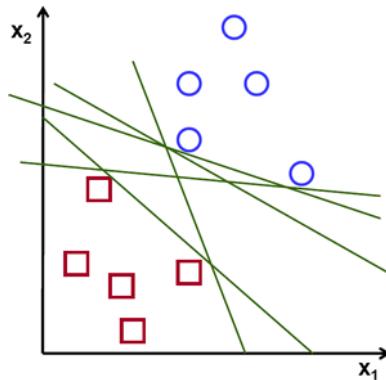


Figura 3.4: In uno spazio bidimensionale un separatore è una linea retta

Più formalmente, una macchina a vettori di supporto costruisce un iperpiano (o un insieme di iperpiani) in uno spazio ad alta dimensionalità. Intuitivamente, un buona separazione è ottenuta dall'iperpiano che abbia distanza massima rispetto ai punti più vicini di classi diverse (il cosiddetto margine geometrico), dal momento in cui maggiore risulta il valore del margine, minore sarà in generale l'errore empirico del classificatore.

I nuovi esempi sono proiettati all'interno dello stesso spazio e la predizione viene calcolata in base alla posizione, rispetto all'iperpiano, in cui il punto ricade.

Le macchine a vettori di supporto possono essere pensate come una tecnica alternativa per l'apprendimento di classificatori polinomiali, contrapposta alle classiche tecniche di addestramento delle reti neurali.

3 Classificatori per il riconoscimento facciale

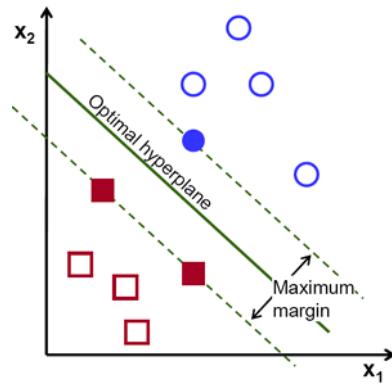


Figura 3.5: Il miglior iperpiano è quello che massimizza il margine geometrico

Alcune applicazioni per cui le SVM sono state utilizzate con successo sono:

- identificazione di facce all'interno delle immagini;
- identificazione di pedoni;
- elaborazione del linguaggio naturale (es. riconoscimento delle cifre);

4 OpenCV

OpenCV (*Open Source Computer Vision*) è una libreria di funzioni multiplatforma, orientata alla Computer Vision in tempo reale. È alla base di tutta l'elaborazione su immagini e video che il framework esegue a runtime.

Ne è stata compilata e usata la versione 3.1.0, per mezzo della sua interfaccia in C++. Per poter avere accesso alle funzionalità contenute nel modulo `face.h`, cui fanno capo gli algoritmi per il riconoscimento facciale, e a quelle presenti in `dnn.h`, dedicato al deep learning, sono stati installati separatamente i moduli extra di cui la libreria è provvista (*OpenCV extra modules*).

Sono di seguito brevemente descritti i moduli della libreria utilizzati nel sistema:

`core` comprende le funzionalità principali.

`imgproc` contiene una serie di utilità per l'elaborazione delle immagini (trasformazioni geometriche, istogrammi, funzioni di disegno ecc. . .).

`highgui` consente la gestione delle interfacce grafiche di alto livello.

`objdetect` espone insieme di classi e metodi per l'individuazione di oggetti (CascadeClassifier, HogDescriptor ecc. . .)

ml	modulo contentente gli algoritmi per l'apprendimento automatico (SVM, kNN, ecc. . .).
dnn	include le implementazioni per diverse strategie di deep learning (dnn, cnn ecc. . .).
face	ospita i descrittori e gli algoritmi per la riduzione della dimensionalità (Eigenfaces, Fisherfaces, LBPH, ecc. . .).

4.1 Caffe

Caffe è un framework sviluppato dal Berkley Vision and Learning Center (BVLC) con il supporto di un'ampia ed attiva comunità di collaboratori [20]. I vantaggi principali della piattaforma sono da ricercarsi nell'architettura, comoda, efficace ed aperta alle estensioni e nella velocità di elaborazione. Secondo gli autori, Caffe rappresenta la più veloce implementazione di *convolutional neural networks* disponibile.

Grazie al modulo extra `dnn.h`, disponibile dalla versione 3.0 di OpenCV, è possibile integrare molto semplicemente un modello addestrato con Caffe e sfruttarlo per effettuare delle previsioni con OpenCV.

Segue una descrizione sommaria della teoria alla base delle tecniche di deep learning.

4.1.1 Deep learning

Il *deep learning* (anche noto come *deep structured learning*, *hierarchical learning* o *deep machine learning*) è una tipologia di apprendimento automatico

che si prefigge l'obiettivo di modellare astrazioni di alto livello dei dati passando per molteplici livelli di elaborazione. Esso appartiene a una famiglia più grande di metodi di machine learning basati sull'apprendimento di *forme di rappresentazione* dei dati.

La ricerca in questo ambito è volta a costruire rappresentazioni efficaci e a creare modelli che apprendono queste rappresentazioni a partire da dati non etichettati su larga scala. Una delle promesse del deep learning è quella di proporre una valida alternativa all'estrazione manuale delle feature per la costruzione degli esempi di training, attraverso algoritmi per l'apprendimento non supervisionato o semi-supervisionato.

Concetti fondamentali

Il deep learning identifica una classe di algoritmi di machine learning che adottano una cascata di livelli di unità di elaborazione non-lineari per l'estrazione di feature e la trasformazione dei dati. Il risultato è l'apprendimento automatico di molteplici livelli di feature (o rappresentazioni dei dati). Le feature di più alto livello sono derivate da quelle di livello più basso, a formare una rappresentazione gerarchica, una gerarchia di concetti: ciascun livello di elaborazione successivo ha per input l'output del livello precedente.

Gli algoritmi in uso in ciascun livello possono essere supervisionati o non supervisionati, e la composizione di un particolare livello di unità di elaborazione non-lineari, in un algoritmo di deep learning, dipende strettamente dal problema da risolvere.

Ciò che distingue gli algoritmi di deep learning (apprendimento approfondito) dai più comuni algoritmi di shallow learning (apprendimento superficiale) è il fatto di trasformare i dati in ingresso in più stadi di elaborazione. Ad

ogni livello il “segnale” (lo stream dei dati) viene trasformato da un’unità di elaborazione – che si comporta come un neurone artificiale – i cui parametri sono stati “appresi” in seguito a un addestramento specifico. Una catena di trasformazioni dall’input all’output rappresenta un *credit assignment path* (CAP). I CAP descrivono possibili connessioni causali tra dati di ingresso e di uscita, e potrebbero variare in lunghezza.

Gli algoritmi di deep learning sono fondati sul concetto di *rappresentazione distribuita*. L’assunzione di fondo per le rappresentazioni distribuite è che i dati osservati sono generati dalle interazioni di fattori organizzati su più livelli. Il deep learning aggiunge l’assunzione che questi livelli di fattori corrispondano a livelli di astrazione (o composizione). Variando il numero dei livelli o della dimensione di ciascun livello si ottengono astrazioni di diversa entità.

L’apprendimento approfondito è impiegato in 1. *attività di apprendimento supervisionato*, laddove la necessità di ingegnerizzare le feature viene aggirata traducendo i dati in rappresentazioni intermedie di dimensioni ridotte e derivando da esse strutture a livelli prive di ridondanze di rappresentazione; 2. *attività di apprendimento non supervisionato*, con importanti benefici visto il gran numero di dati non etichettati in genere disponibili.

Il deep learning ha visto proliferare numerose architetture e modelli differenti. Per gli scopi del presente lavoro di tesi, verrà illustrata solo quella delle reti neurali a convoluzione.

Convolutional Neural Networks

Le Convolutional Neural Networks (d’ora in poi CNN) hanno conquistato sempre più importanza per l’elaborazione visuale e di dati bidimensionali in genere. Una CNN è composta da uno o più strati di “convoluzione”, avente

4 OpenCV

in cima quelli completamente connessi fra loro. Questo modello permette alle CNN di avvantaggiarsi della struttura a due dimensioni dei dati in ingresso. Rispetto ad altre architetture di deep learning, le CNN hanno mostrato risultati superiori sia nelle applicazioni basate sulle immagini che su quelle fondate sull'elaborazione del linguaggio parlato. Inoltre, sono più semplici da addestrare e si contraddistinguono per avere pochi parametri da dover valutare, rendendone l'utilizzo particolarmente attraente. [21]

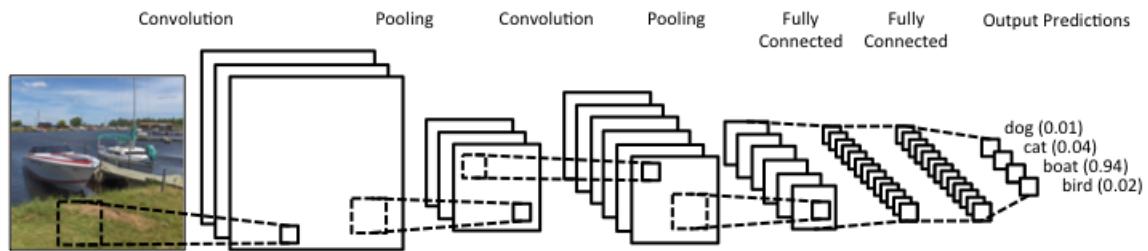


Figura 4.1: Un'immagine esemplificativa del processo di elaborazione di una CNN.

5 Riconoscimento biometrico di genere ed etnia attraverso l'analisi del volto

5.1 Requisiti e configurazione hardware

Lo sviluppo del progetto è stato condotto su una macchina virtuale con sistema operativo Ubuntu 14.04 LTS. Questa scelta si deve al proposito iniziale di integrare nel framework alcuni moduli già funzionanti, progettati per OpenCV e ottimizzati per l'esecuzione in ambiente Linux. Tale integrazione si è in seguito rivelata inconciliabile con l'ultima versione di OpenCV, importante prerogativa del progetto. I moduli in questione, infatti, avrebbero richiesto una retrocessione alla versione 2.4 della libreria, ormai da tempo aggiornata alla 3.1. Così, continuare a lavorare in un ambiente virtualizzato a ridotte prestazioni è sembrato un buon modo per tenere sotto controllo i requisiti hardware del sistema. Questa ulteriore attenzione è coerente con uno degli obiettivi principali del progetto, che mira a rendere possibile l'installazione del framework anche su sistemi embedded, caratterizzati dalla disponibilità di risorse limitate.

5.1.1 Configurazione della macchina fisica

Componente	Specifiche
Modello	MacBook Pro (13 pollici, metà 2012)
Sistema operativo	OS X El Capitan 10.11.4
Processore	Intel Core i7 dual-core a 2,9 GHz (Turbo Boost fino a 3,6 GHz) con 4 MB di cache L3
Memoria	8 GB di memoria DDR3 a 1600MHz
Archiviazione	Unità a stato solido da 256GB
Scheda grafica	Intel HD Graphics 4000
Videocamera	Videocamera FaceTime HD a 720p

5.1.2 Configurazione della macchina virtuale

Componente	Specifiche
Macchina virtuale	VirtualBox 5.0.14 r105127
Sistema operativo	Ubuntu 14.04 LTS
Numero Core	2
Memoria	4 GB
Memoria video	128 MB
Accelerazione 3D	Abilitata
Archiviazione	20 GB

5.2 Dati di addestramento

5.2.1 Fonti

Sono stati sperimentati e utilizzati diversi dataset disponibili gratuitamente su internet a scopo di ricerca.

Le prime prove hanno riguardato *The ORL Database of Faces*, contenente un insieme di foto scattate fra il 1992 e il 1994 presso i laboratori AT&T, a Cambridge. Il dataset comprende 10 immagini differenti per ciascuno dei 40 soggetti; per alcuni di loro, le immagini sono state scattate in momenti diversi, variando la luce, l'espressione facciale (occhi aperti/chiusi, sorridendo/non sorridendo) e alcuni dettagli come la presenza o meno di occhiali. Tutte le foto presentano i soggetti in posizione frontale su uno sfondo omogeneamente scuro. Ogni immagine è 92x112 pixel, in formato PGM.



Figura 5.1: Immagini prese da *The ORL Database of Faces*.

Successivamente, è stato sfruttato per l'addestramento lo *Extended Yale Database B*, il quale comprende 16128 immagini per 28 soggetti in 9 pose e 64 condizioni di luce diverse.

5 Riconoscimento biometrico di genere ed etnia attraverso l'analisi del volto

Entrambi i dataset si sono rivelati insufficienti per il riconoscimento dei parametri biometrici per due ragioni fondamentali: la prima è da attribuirsi alla presenza di un basso numero di soggetti diversi rispetto al quantitativo totale di immagini contenute; la seconda è che ambedue i dataset sono sbilanciati, in favore della componente maschile rispetto alla controparte femminile, risultando, inoltre, completamente inefficaci per il compito di classificazione etnica.

I pessimi risultati iniziali hanno guidato la ricerca in direzione di un dataset che privilegiasse la numerosità di soggetti diversi e che offrisse la possibilità di bilanciare in maniera ottimale gli esempi appartenenti alle diverse categorie.

E' stato dunque preso in considerazione il *IMDB-WIKI dataset* (di cui è stata esaminata solo la parte relativa a Wikipedia), caratterizzato da 62328 foto di soggetti distinti, a varie risoluzioni, appartenenti a ceppi etnici differenti, da cui sono state prelevate manualmente le immagini migliori, al fine di costruire un insieme di esempi di addestramento efficace allo scopo.



Figura 5.2: Immagini prese dal *IMDB-WIKI dataset*.

Infine, per integrare le fonti a disposizione si è fatto uso di *Google Immagini*,

5 Riconoscimento biometrico di genere ed etnia attraverso l'analisi del volto

con la dovuta accortezza di selezionare materiale che fosse contrassegnato per il riutilizzo non commerciale.

5.2.2 Operazioni di pre-processamento

Tutte le immagini sono state sottoposte ad un processo composto dalle seguenti fasi:

1. *Individuazione degli occhi*: sono stati manualmente presi i punti corrispondenti alla posizione dell'occhio sinistro e destro.
2. *Rotazione*: ciascuna immagine è stata raddrizzata e centrata in base della posizione degli occhi al fine di uniformare l'inclinazione del volto;
3. *Ritaglio*: è stata ritagliata la porzione dell'immagine contenente il volto in posizione frontale o con una inclinazione approssimativamente non superiore ai 30 gradi. Rispetto alla posizione degli occhi è stato considerato un margine sinistro e destro pari a 1.3;
4. *Ridimensionamento*: sono state sperimentate diverse dimensioni per l'addestramento dei classificatori: 64x64 pixel, 88x88 pixel, 176x176 pixel.

A seconda del metodo di classificazione adottato, le immagini sono state ulteriormente processate con i seguenti passi:

- *Passaggio in scala di grigi*: in una immagine in scala di grigi ogni pixel è rappresentato con un valore di intensità compreso fra 0 (nero) e 255 (bianco). Esempi di classificatore che richiedono tale rappresentazione in ingresso sono Haar Cascade, Fisherfaces e Local Binary Patterns Histograms;

5 Riconoscimento biometrico di genere ed etnia attraverso l'analisi del volto

- *Equalizzazione:* ogni immagine in scala di grigi può essere graficamente rappresentata attraverso histogrammi che indicano la distribuzione dei valori di intensità dei pixel, permettendo di quantificare facilmente il numero di pixel per ciascun valore di intensità. Il processo di equalizzazione mira a migliorare il contrasto generale dell'immagine in modo da distendere l'intervallo dei valori di intensità.

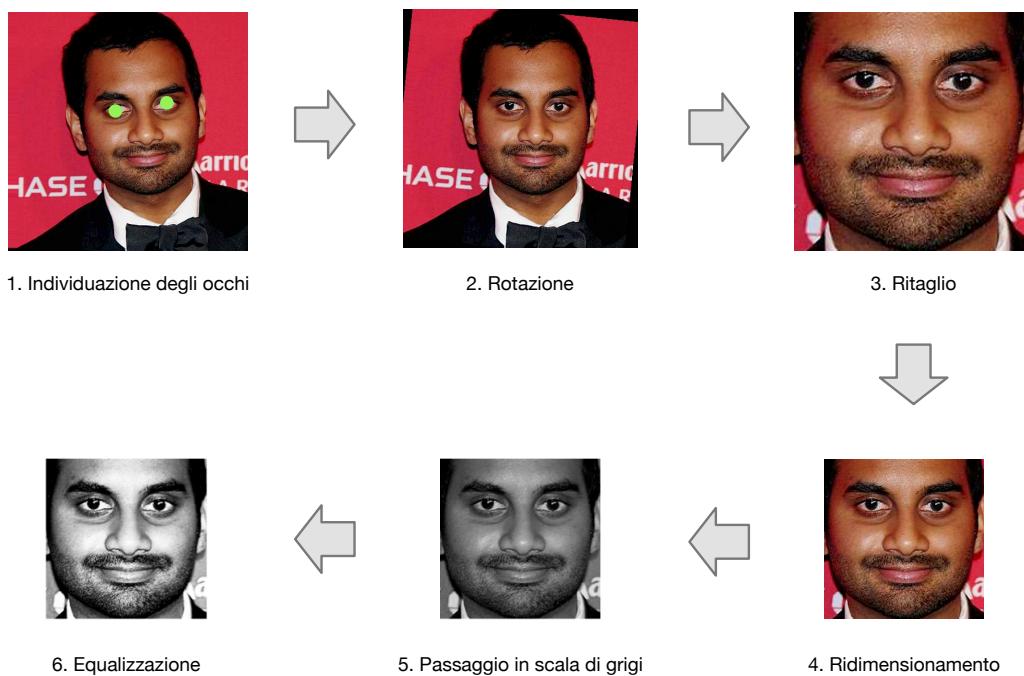


Figura 5.3: Operazioni di pre-processamento

5 Riconoscimento biometrico di genere ed etnia attraverso l'analisi del volto

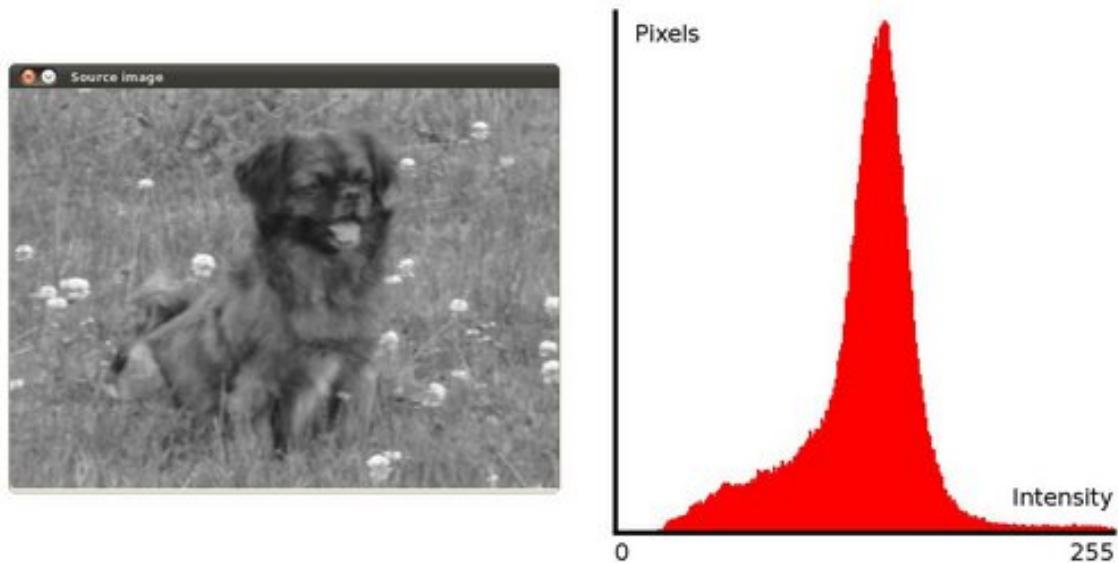


Figura 5.4: La rappresentazione di un'immagine mediante istogrammi.

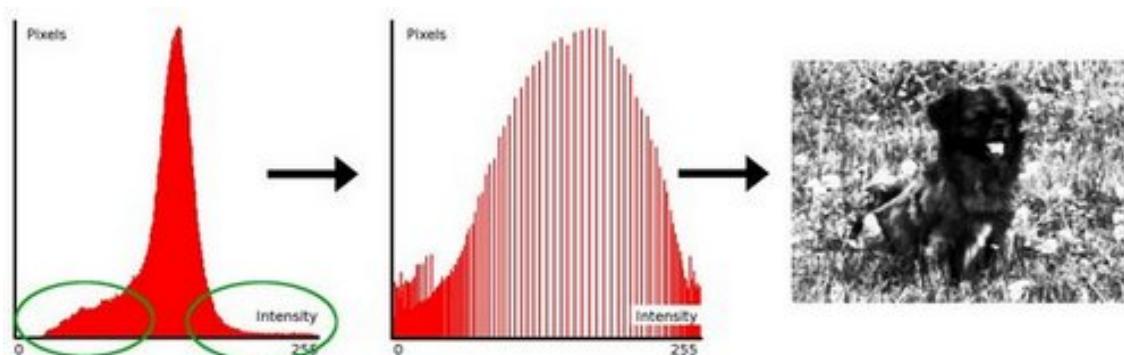


Figura 5.5: Esempio di equalizzazione dell'immagine

5.3 Architettura del sistema

Uno degli obiettivi del presente lavoro di tesi è stato quello di costruire un framework intuitivo e solido, che offrisse la possibilità di essere riusato facilmente ed evoluto a seconda delle proprie esigenze di sviluppo.

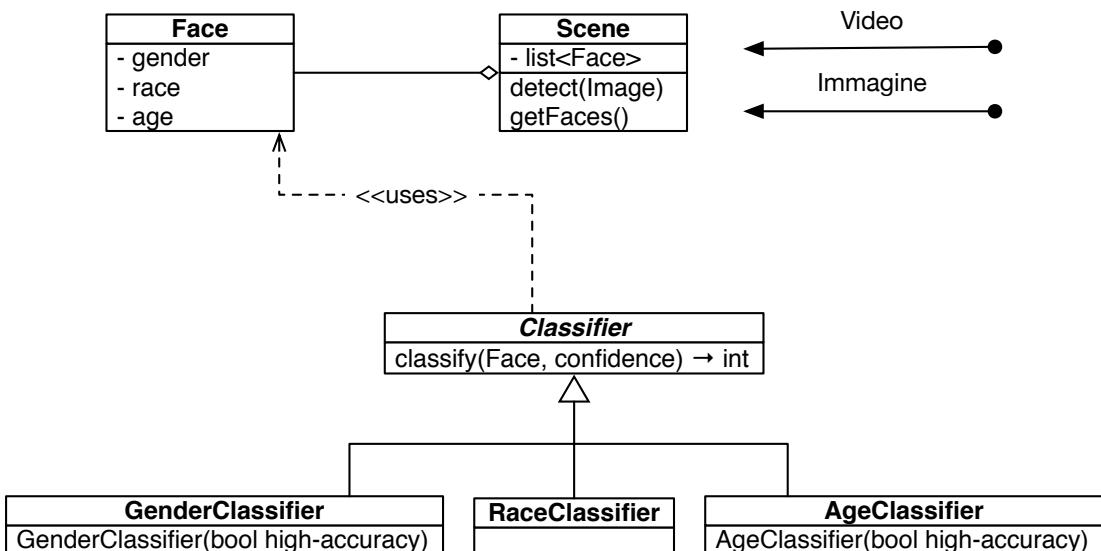


Figura 5.6: Una visione esemplificativa dell'architettura di sistema.

5.3.1 Descrizione della struttura

Sono di seguito illustrate le classi principali dell'architettura software:

- **Scene**: si occupa di analizzare una singola immagine attraverso il metodo `detect`, le cui funzionalità consentono l'individuazione dei volti e

5 Riconoscimento biometrico di genere ed etnia attraverso l'analisi del volto

della posizione degli occhi. Incapsula al suo interno una lista di oggetti di tipo Face.

- **Face:** contiene tutti gli attributi principali per la rappresentazione del volto: la matrice di pixel, la posizione nell'immagine, i punti degli occhi, il genere, l'etnia e l'età stimati. Dispone dei necessari metodi per l'accesso e la modifica dei parametri.
- **Classifier:** rappresenta una interfaccia essenziale per i classificatori biometrici, ridotta all'unico operatore virtuale `classify`, il quale restituisce un intero positivo da mappare con un'opportuna etichetta a seconda della specifica implementazione e lo scopo di classificazione.
- **GenderClassifier:** è specializzazione di Classifier; ingloba gli algoritmi per la previsione del genere sulla base dell'immagine in ingresso.
- **RaceClassifier:** è specializzazione di Classifier; ingloba gli algoritmi per la previsione dell'etnia sulla base dell'immagine in ingresso.

5.3.2 Dinamica di funzionamento

L'esplorazione comincia attraverso la creazione di un oggetto Scene. A seconda che l'elaborazione sia richiesta su di un insieme di immagini differenti o su di una sequenza video, quest'ultimo dovrà essere usato in maniera diversa: nel primo caso dovrà essere creato un oggetto di tipo Scene per ciascuna delle immagini in ingresso e invocato il rispettivo operatore `detect()`; nel secondo caso è altamente consigliato l'utilizzo dello stesso oggetto Scene, premunendosi di invocare ripetutamente il metodo `detect()` per ciascuna frazione di video. Infatti, sono stati sviluppati meccanismi per riconoscere uno stesso

5 Riconoscimento biometrico di genere ed etnia attraverso l'analisi del volto

volto in frame consecutivi, calcolando le differenze di posizionamento tra i volti correnti rispetto a quelli individuati nel frame precedente. Nello specifico, sono considerati i due volti aventi distanza minima di posizionamento e su di essi applicata una soglia empirica, per cui se la distanza calcolata risulta minore o al più uguale a 70 i volti vengono associati e identificati come lo stesso.

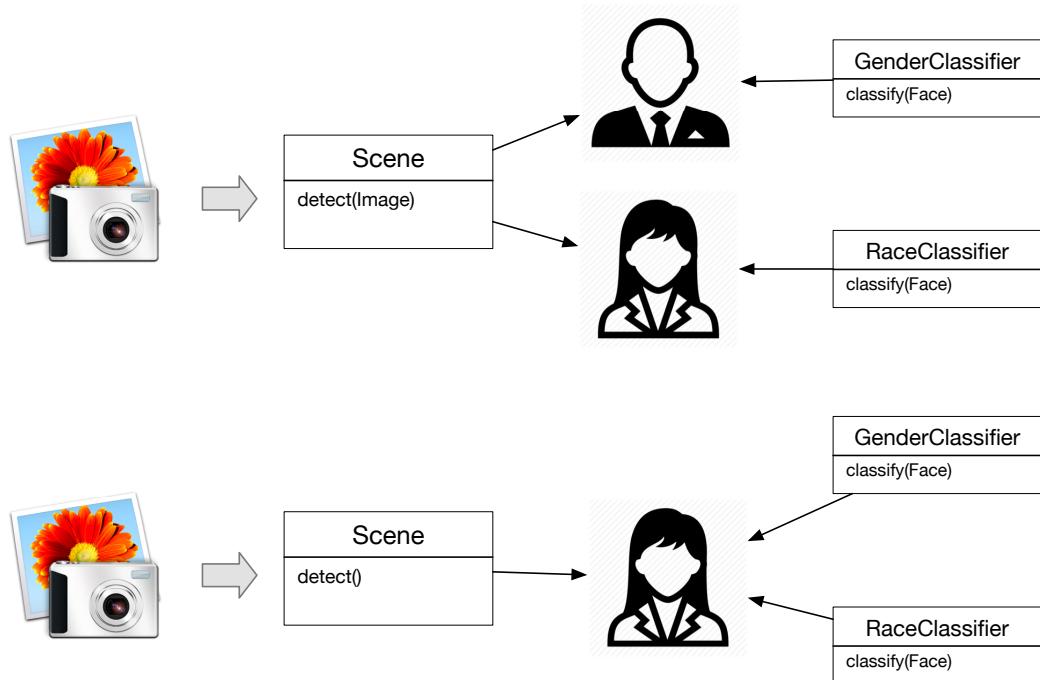


Figura 5.7: Funzionamento del sistema nel caso di immagini differenti in ingresso.

Queste operazioni si sono rilevate necessarie per gestire in maniera ottimale l'analisi di video in tempo reale: senza di esse non sarebbe stato possibile

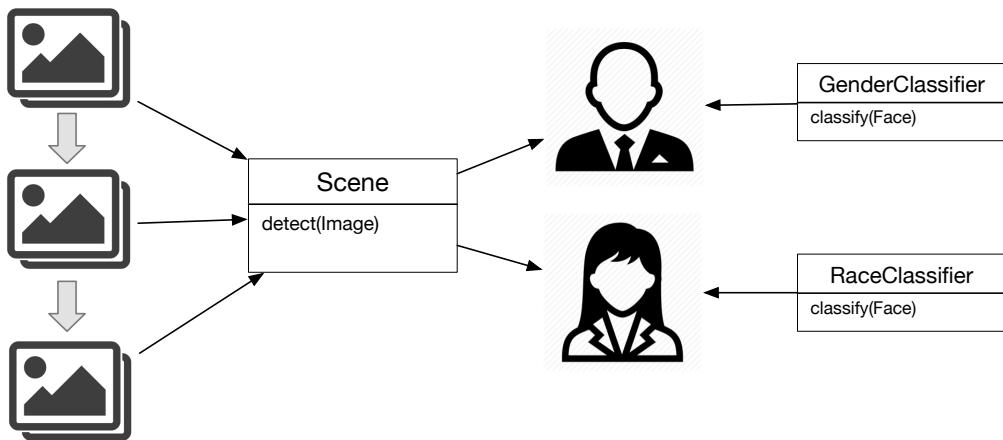


Figura 5.8: Funzionamento del sistema nel caso di una sequenza video in ingresso.

assegnare un'etichetta stabile ad un determinato volto; a questo proposito, si è deciso di conservare i primi 20 risultati per ciascuna tipologia di classificazione e calcolare la previsione più frequente come risposta definitiva. Inoltre, questo approccio incrementa l'efficienza generale del sistema in presenza di soggetti multipli all'interno di una scena. Alternativamente, il sistema avrebbe dovuto elaborare le classificazioni richieste per tutti i volti, per ciascun frammento video. Grazie all'algoritmo proposto, il sistema non prenderà più in considerazione i volti per cui sono state già effettuate il numero di classificazioni richieste per esprimere un giudizio definitivo, permettendo di concentrare le risorse di calcolo sui nuovi volti o quelli ancora parzialmente valutati.

Il processo di individuazione si attua per mezzo di un modello già disponibile con l'installazione di OpenCV 3.1 (`haarcascade-frontalface-alt.xml`) che viene caricato in un oggetto *CascadeClassifier* tramite il metodo `load()`.

5 Riconoscimento biometrico di genere ed etnia attraverso l'analisi del volto

Ad ogni nuova invocazione di `detect()`, viene costruita o aggiornata la lista dei volti (oggetti `Face`) presenti nella scena. Dall'esterno è possibile ottenere un riferimento a questi oggetti attraverso l'operatore `getFaces()`.

Dipendentemente dagli obiettivi di classificazione si potranno istanziare le opportune specializzazioni della classe virtuale `Classifier` e invocare l'operatore `classify()` sugli oggetti `Face` individuati al passo precedente.

5.4 Riconoscimento del genere

Il genere è stata una delle prime biometriche indagate in letteratura. Alcuni studi hanno dimostrato che un essere umano è capace di distinguere con facilità l'appartenenza al sesso maschile e a quello femminile (oltre il 95% di accuratezza dai volti)[22].

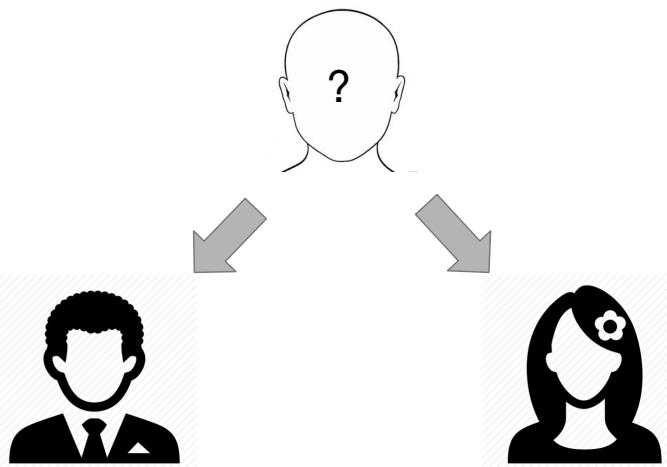


Figura 5.9: Problema della classificazione del genere

5.4.1 Esempi di addestramento

L'insieme di addestramento finale, per ciascuna classe, è il seguente:

- 100 esempi di uomini;
- 103 esempi di donne;

Le immagini sono state selezionate cercando di diversificare i soggetti rispetto all'etnia di appartenenza e alla fascia d'età.

5.4.2 Scelte implementative

Sono stati impiegati due approcci:

- approccio orientato alle *prestazioni*
- approccio orientato all'*accuratezza*

Nel primo caso, è stata posta grande attenzione ai dispositivi a risorse limitate, enfatizzando la leggerezza e l'efficienza del metodo, senza penalizzare troppo l'*accuratezza*. Nel secondo caso, l'elaborazione del risultato è decisamente più costosa in termini di risorse di calcolo, ma garantisce livelli di precisione più alti in sede di classificazione.

Approccio orientato alle prestazioni. In una fase iniziale, sono stati inizialmente confrontati gli algoritmi Fisherfaces e Local Binary Patterns Histograms: sulla base dei risultati ottenuti, il primo si è subito rivelato essere il più promettente e i successivi sforzi sono stati spesi per migliorarne l'*efficacia*.

A questo proposito, i tentativi hanno riguardato la variazione dei parametri relativi agli esempi di addestramento, come la dimensione delle immagini e il numero degli esempi per ciascuna classe.

Per quanto concerne la dimensione delle immagini, sono stati eseguiti test su esempi di: 64x64 pixel, 88x88 pixel, 176x176 pixel. La dimensione fissata in partenza è stata di 88x88 pixel. Si è subito provato a raddoppiare la grandezza delle immagini, verificando un sensibile degrado delle prestazioni, dovuto verosimilmente alla eccessiva dimensionalità nella rappresentazione degli esempi. Si è allora provato a ridurre la dimensione portandola a 64x64 pixel: i test condotti non hanno mostrato netti incrementi o cali significativi nelle prestazioni generali. Pertanto la scelta finale è ricaduta su immagini di

5 Riconoscimento biometrico di genere ed etnia attraverso l'analisi del volto

64x64 pixel, sia in quanto esse risultano più semplici e veloci da elaborare in tempo reale, sia perché permettono la generazione di modelli più leggeri da caricare in memoria durante l'esecuzione del sistema.

I primi test sono stati condotti con 60 immagini per classe. Stabilita l'adozione dell'algoritmo Fisherfaces e la dimensione degli esempi di addestramento ottimale (64x64 pixel), si è provato ad incrementarne il numero prima a 100 per classe, poi a 200. Il primo incremento ha portato a un sensibile miglioramento nei risultati dei test; il passaggio ulteriore a 200 immagini - rispetto a 100 - ha comportato invece un lieve degrado delle prestazioni. La soluzione adottata è stata in definitiva la seconda.

Da ultimo si è tentato di introdurre un passo di equalizzazione delle immagini, da cui è conseguito un miglioramento in termini di accuratezza media riscontrata.

La ridotta dimensione delle immagini e il numero esiguo di esempi di addestramento costituiscono gli elementi che consentono all'algoritmo di classificazione di essere eseguito con un dispendio di risorse molto limitato, in linea con gli obiettivi che hanno guidato l'intero sviluppo del sistema.

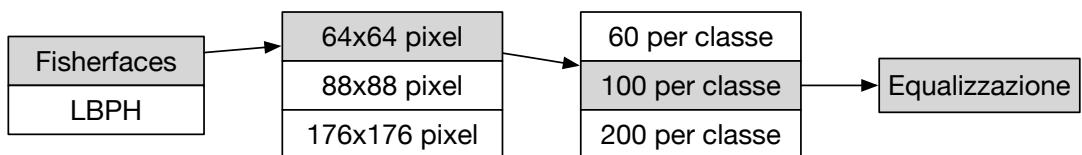


Figura 5.10: Scelte effettuate per il riconoscimento del genere nell'approccio orientato alle prestazioni

Approccio orientato all'accuratezza. A beneficio di chi avesse a disposizione una macchina performante, è stato aggiunto al sistema un modello addestrato mediante strategie di deep learning. Tale modello è stato liberamente fornito da Rothe, Timofte e Van Gool [23] e addestrato mediante il framework Caffe a partire da oltre 500000 immagini provenienti dai dataset di *Wikipedia* e *IMDB*.

L'integrazione ha permesso di saggiare con mano le enormi potenzialità dei metodi di deep learning. A causa della proprietà esterna del modello, non sono stati condotti test formali per valutarne l'accuratezza; tuttavia, le prove effettuate in maniera informale hanno evidenziato risultati sbalorditivi, seppure al prezzo di un oneroso sforzo computazionale, che ha messo a dura prova le prestazioni in tempo reale della macchina a disposizione.

5.5 Riconoscimento dell'etnia

A livello terminologico, è possibile definire il concetto di razza come un “*raggruppamento di individui che presentano un insieme di caratteri fisici ereditari comuni. Nel caso dell'uomo, tali caratteri si riferiscono a caratteristiche somatiche (colore della pelle, tipo di capelli, forma del viso, del naso, degli occhi ecc.)*”. Per etnia, invece, “*si identifica una comunità caratterizzata da omogeneità di lingua, cultura, tradizioni e memorie storiche, stanziate tradizionalmente su un determinato territorio.*” Nel corso degli anni questi concetti sono stati soggetti ad un lungo dibattito fra gli antropologi sulle loro implicazioni e sul loro corretto significato. Senza entrare nel merito di una specifica posizione sull'argomento, è doveroso premettere che, in questa sede, i termini etnia e razza saranno usati in modo interscambiabile, ad indicare esclusivamente le caratteristiche somatiche distintive di un individuo, prescindendo da ogni riferimento culturale.

Ad oggi, la maggior parte degli antropologi distingue dalle 3 alle 5 razze principali: una sintetica tassonomia considera la razza caucasica (Caucasoid), la razza afroamericana (Negroid) e la razza asiatica (Mongoloid), a cui si aggiungono, per una classificazione più completa, la razza dei nativi americani e la razza australoide (Australoid), costituita dagli individui originari delle isole del Pacifico.

Allo stato attuale, il sistema si limita a classificare soggetti appartenenti alle prime tre etnie succitate.

5.5.1 Esempi di addestramento

L'insieme di addestramento finale, per ciascuna classe, è il seguente:

- 100 esempi per l'etnia afro-americana
- 100 esempi per l'etnia asiatica
- 100 esempi per l'etnia caucasica

5.5.2 Scelte implementative

Il presente problema di classificazione ternario è stato interpretato e affrontato attraverso la composizione di due classificatori binari. Il primo tenta di identificare l'appartenenza all'etnia afro-americana: in tal caso il processo si ferma e viene prontamente consegnato il risultato; altrimenti, si procede con un nuovo passo di classificazione che mira a distinguere le persone di etnia caucasica rispetto a quelle di etnia asiatica.

Le motivazioni di questa scelta risiedono nelle modalità con cui intuitivamente una persona affronta il compito di riconoscimento etnico nel caso di soggetti afro-americani, per cui la prima e più evidente informazione utile è data dal colore della pelle a cui si sommano le considerazioni sui lineamenti del viso.

Per assecondare questo modo di ragionare sono stati addestrati e costruiti due modelli, rispettivamente per esaminare il colore della pelle e i tratti somatici. Allo scopo di interpretare correttamente l'informazione sul colore, anche al variare delle condizioni di luce, è stato scelto lo spazio di colore HSV che isola in un singolo campo l'indicazione sulla luminosità. Si sono allora ipotizzate diverse strategie per l'estrapolazione del colore a partire dall'immagine del volto. Un primo problema ha riguardato il *dove* concentrare l'attenzione al fine di scartare informazione potenzialmente inquinante: sono state prese in considerazione la zona del mento, la zona della fronte e la zona

5 Riconoscimento biometrico di genere ed etnia attraverso l'analisi del volto

nasale. Le prime due sono apparse scelte pericolose a causa della possibile presenza di barba o di capelli rispettivamente. La zona nasale, d'altro canto, risulta spesso una delle zone più illuminate del viso, a tal punto da alterare, talvolta, l'informazione sul colore. Fra le possibilità esaminate, quest'ultima si è rivelata sicuramente la migliore, ma si è deciso di includere dei margini che comprendessero anche una parte delle guance per colmare i difetti del metodo. Un secondo problema ha invece riguardato il *come* estrapolare effettivamente il colore. A questo proposito sono state vagliate due possibilità, il colore più frequente e il colore medio. Dalle prove effettuate, il colore medio è risultato il migliore e il più sicuro, anche a causa dei risultati imprevedibili riscontrati con il primo metodo. Si è quindi proceduto ad addestrare un classificatore SVM con i valori di tonalità e saturazione dei colori estratti, sulla base delle immagini di esempio: in questo senso è stata selezionata solo una piccola parte dell'insieme di addestramento, prelevando esclusivamente i soggetti afro-americani con una carnagione molto scura e i soggetti caucasici dalla carnagione molto chiara, scartando gli esempi al limite fra le due classi. Questo lavoro è stato fatto per assicurare che i valori di confidenza della previsione per questo tipo di immagini fosse piuttosto basso, nella prospettiva di combinare la valutazione del classificatore basato sui lineamenti del viso, il cui ruolo in questi casi deve risultare predominante.

In veste di classificatore per i lineamenti del viso, esattamente come per il riconoscimento del genere, anche in questo caso è stato sperimentato l'uso di Fisherfaces e Local Binary Patterns Histograms. Ancora una volta, il primo ha manifestato la sua superiorità rispetto al secondo grazie a valori di accuratezza superiori.

Poiché la previsione dell'algoritmo Fisherfaces si fonda sul classificatore

5 Riconoscimento biometrico di genere ed etnia attraverso l'analisi del volto

kNN, i valori di confidenza calcolati rappresentano le distanze algebriche rispetto al vicinato di esempi più simili, in cui 0 è il valore di massima similarità e valori più alti, nell'ordine anche delle centinaia, celano una progressiva incertezza nel risultato di classificazione. Questa interpretazione di confidenza differisce da quella di SVM, il quale colloca i valori in un intervallo in cui lo 0 rappresenta il valore di similarità minima e valori vicino a 1 o superiori descrivono una maggiore certezza nella previsione. Non essendoci delle soglie prefissate per i valori di confidenza, rispettivamente di minimo e di massimo, sono stati condotti una serie di test mirati ad individuare dei parametri accettabili, al fine di rendere confrontabili le due misure. Nel caso di kNN è stato deciso di rigettare e approssimare al minimo tutti i valori oltre 180; nel caso di SVM sono stati approssimati a 1.15 e tutti i valori ad esso superiori. Una volta espressi entrambi i valori nell'intervallo [0, 1], l'algoritmo esegue un semplice confronto, prendendo in considerazione il risultato di classificazione con la confidenza maggiore.

Per risolvere il problema di distinguere le persone di etnia caucasica da quelle di etnia asiatica è stato ancora una volta sfruttato l'algoritmo Fisherfaces. Una delle idee chiave che ha accompagnato l'intero sviluppo del compito di classificazione etnica è stata quella di puntare l'attenzione sulle caratteristiche fisiche più distintive nel valutare le diverse etnie. In maniera analoga rispetto a quanto fatto con l'etnia afro-americana, si è pensato di sfruttare la principale particolarità dell'etnia asiatica, ovvero gli occhi a mandorla. E' stata dunque ulteriormente ritagliata l'area degli occhi dalle immagini di esempio e addestrato l'algoritmo Fisherfaces e LBPH. Sfortunatamente, le prestazioni sono state al di sotto delle aspettative per entrambi i metodi, a cui è stato quindi preferito l'approccio basato sull'analisi globale del volto.

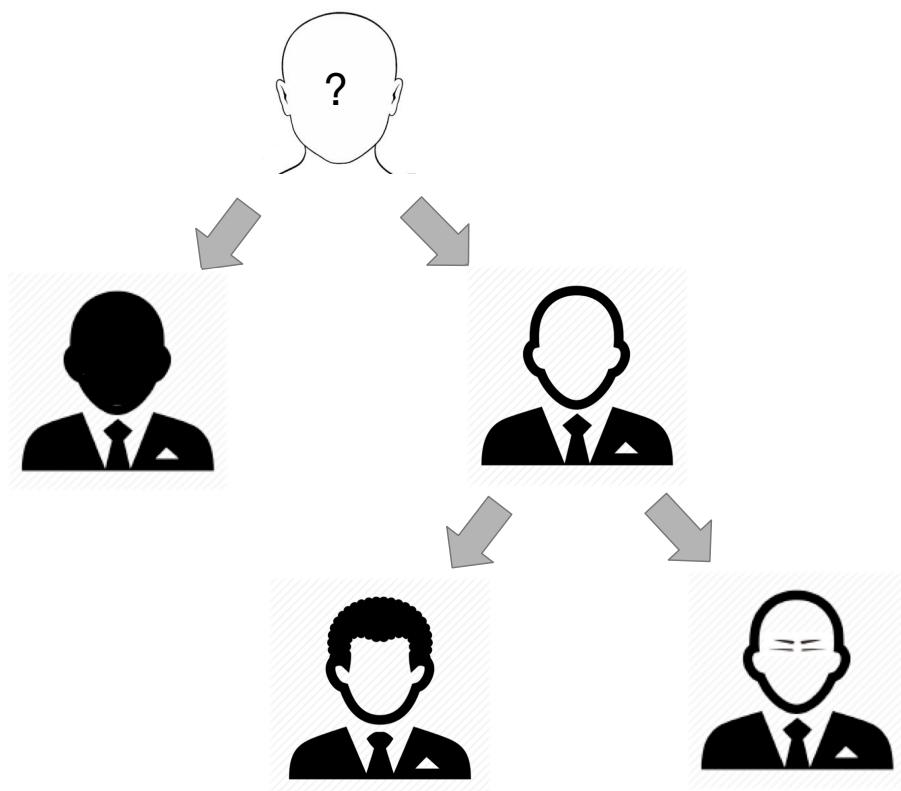


Figura 5.11: Approccio al problema della classificazione dell'etnia.

5.6 Applicazione Web

Si è scelto di condividere pubblicamente una parte delle funzionalità del framework nella forma di un sito web.

5.6.1 Interfaccia utente

Il sito consente di caricare un'immagine dal disco locale e di ottenere, previa consultazione del server su cui il framework è installato, informazioni sul numero dei volti presenti nella foto e sulle caratteristiche biometriche attribuitegli dal sistema.

5 Riconoscimento biometrico di genere ed etnia attraverso l'analisi del volto

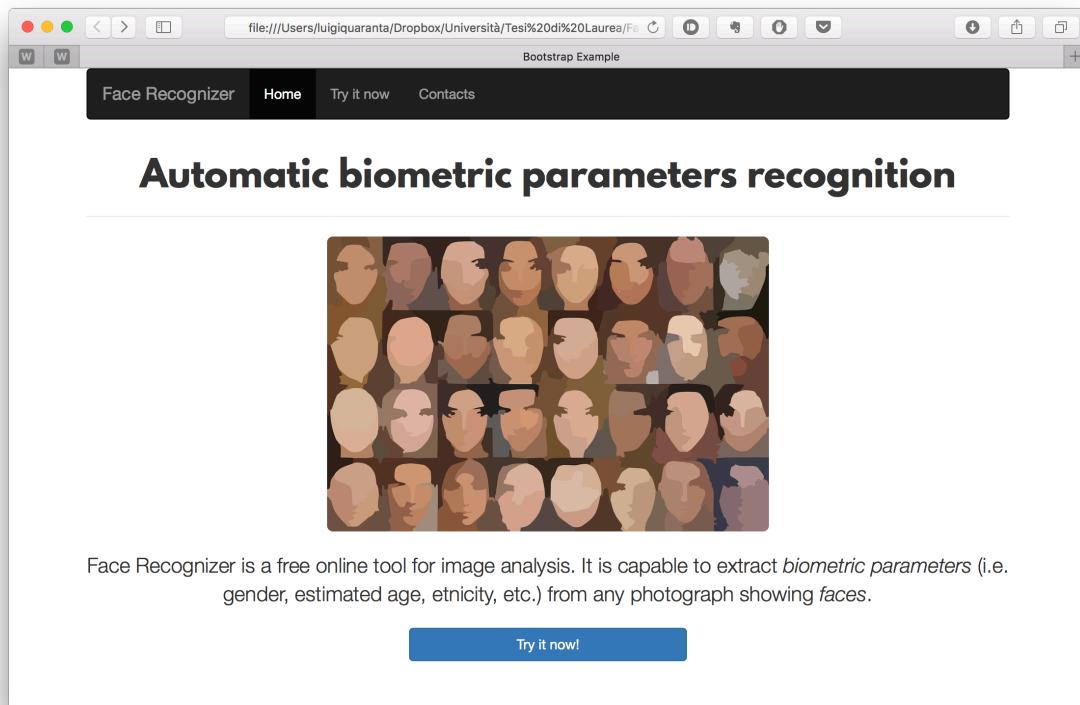


Figura 5.12: Home page del sito.

In seguito al caricamento dell'immagine, l'utente può selezionare i caratteri biometrici su cui desidera che il sistema si esprima. In particolare, per quel che riguarda il riconoscimento del genere e dell'età, gli è anche dato di specificare la tipologia di approccio: quello orientato alle prestazioni o quello orientato all'accuratezza.

La libreria è stata installata su una macchina server apache e viene richiamata da uno script php collegato alla pagina html tramite una procedura ajax. Lo script ha il solo compito di richiedere al sistema operativo del server l'esecuzione del framework, passandogli come parametri la foto da analizzare

5 Riconoscimento biometrico di genere ed etnia attraverso l'analisi del volto

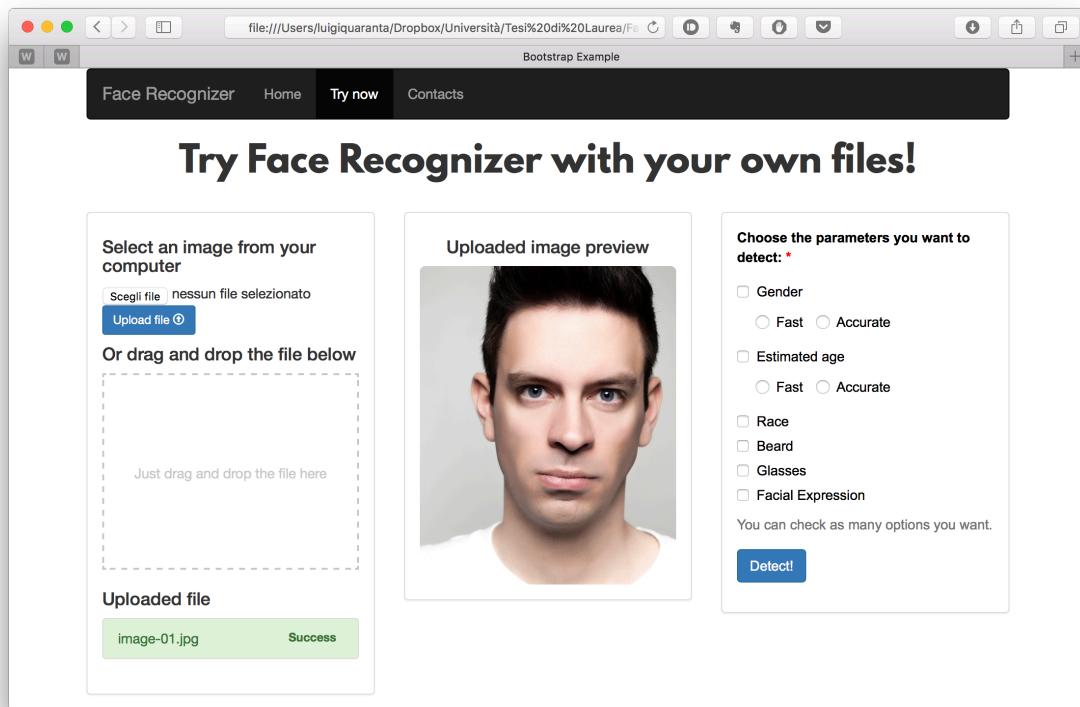


Figura 5.13: Interfaccia di caricamento dell'immagine e visualizzazione dei risultati.

e i moduli da attivare con le relative preferenze.

6 Risultati sperimentali

Il protocollo sperimentale adottato per la valutazione delle prestazioni dei classificatori è la *k-fold cross validation*, con $k = 10$. Sono stati condotti diversi test sia per il genere che per l'etnia, volti a trovare la configurazione migliore al variare dei descrittori e delle caratteristiche dei dati di addestramento. Per i test considerati sono state riportate le matrici di confusione per ciascun fold e calcolate la precisione, il richiamo e l'accuratezza medi.

6.1 Riconoscimento del genere

6.1.1 Test sugli esempi

Caso 1. Genere con algoritmo Fisherfaces, immagini non equalizzate da 88x88 pixel, 100 esempi di addestramento per gli uomini e 103 per le donne .

Fold n. 1		
	Maschio _P	Femmina _P
Maschio _R	10	2
Femmina _R	0	9

Fold n. 2		
	Maschio _P	Femmina _P
Maschio _R	8	2
Femmina _R	2	9

6 Risultati sperimentali

<i>Fold n. 3</i>		
	Maschio _P	Femmina _P
Maschio _R	8	2
Femmina _R	2	9

<i>Fold n. 4</i>		
	Maschio _P	Femmina _P
Maschio _R	7	2
Femmina _R	3	8

<i>Fold n. 5</i>		
	Maschio _P	Femmina _P
Maschio _R	9	2
Femmina _R	1	8

<i>Fold n. 6</i>		
	Maschio _P	Femmina _P
Maschio _R	8	0
Femmina _R	2	10

<i>Fold n. 7</i>		
	Maschio _P	Femmina _P
Maschio _R	7	1
Femmina _R	3	9

<i>Fold n. 8</i>		
	Maschio _P	Femmina _P
Maschio _R	7	1
Femmina _R	3	9

<i>Fold n. 9</i>		
	Maschio _P	Femmina _P
Maschio _R	6	1
Femmina _R	4	9

<i>Fold n. 10</i>		
	Maschio _P	Femmina _P
Maschio _R	7	1
Femmina _R	3	9

6 Risultati sperimentali

- Precisione media: 0.827
- Richiamo medio: 0.818
- F_1 media: 0.822
- Accuratezza media: 0.817

Caso 2. Genere con algoritmo LBPH, immagini non equalizzate da 88x88 pixel, 100 esempi di per gli uomini e 103 per le donne .

<i>Fold n. 1</i>		
	Maschio _P	Femmina _P
Maschio _R	7	1
Femmina _R	3	10

<i>Fold n. 2</i>		
	Maschio _P	Femmina _P
Maschio _R	6	2
Femmina _R	4	9

<i>Fold n. 3</i>		
	Maschio _P	Femmina _P
Maschio _R	6	2
Femmina _R	4	9

<i>Fold n. 4</i>		
	Maschio _P	Femmina _P
Maschio _R	7	3
Femmina _R	3	7

<i>Fold n. 5</i>		
	Maschio _P	Femmina _P
Maschio _R	7	4
Femmina _R	3	6

<i>Fold n. 6</i>		
	Maschio _P	Femmina _P
Maschio _R	10	2
Femmina _R	0	8

6 Risultati sperimentali

Fold n. 7		
	Maschio _P	Femmina _P
Maschio _R	9	4
Femmina _R	1	6

Fold n. 8		
	Maschio _P	Femmina _P
Maschio _R	10	2
Femmina _R	0	8

Fold n. 9		
	Maschio _P	Femmina _P
Maschio _R	8	4
Femmina _R	2	6

Fold n. 10		
	Maschio _P	Femmina _P
Maschio _R	9	2
Femmina _R	1	8

- Precisione media: 0.779
- Richiamo medio: 0.767
- F_1 media: 0.773
- Accuratezza media: 0.769

Caso 3. Genere con algoritmo Fisherfaces, immagini non equalizzate da 64x64 pixel, 100 esempi di addestramento per gli uomini e 103 per le donne .

Fold n. 1		
	Maschio _P	Femmina _P
Maschio _R	9	2
Femmina _R	1	9

Fold n. 2		
	Maschio _P	Femmina _P
Maschio _R	8	2
Femmina _R	2	9

6 Risultati sperimentali

<i>Fold n. 3</i>		
	Maschio _P	Femmina _P
Maschio _R	8	2
Femmina _R	2	9

<i>Fold n. 4</i>		
	Maschio _P	Femmina _P
Maschio _R	8	0
Femmina _R	2	10

<i>Fold n. 5</i>		
	Maschio _P	Femmina _P
Maschio _R	8	2
Femmina _R	2	8

<i>Fold n. 6</i>		
	Maschio _P	Femmina _P
Maschio _R	5	0
Femmina _R	5	10

<i>Fold n. 7</i>		
	Maschio _P	Femmina _P
Maschio _R	7	4
Femmina _R	3	6

<i>Fold n. 8</i>		
	Maschio _P	Femmina _P
Maschio _R	7	1
Femmina _R	3	9

<i>Fold n. 9</i>		
	Maschio _P	Femmina _P
Maschio _R	9	3
Femmina _R	1	7

<i>Fold n. 10</i>		
	Maschio _P	Femmina _P
Maschio _R	7	0
Femmina _R	3	10

6 Risultati sperimentali

- Precisione media: 0.819
- Richiamo medio: 0.803
- F_1 media: 0.811
- Accuratezza media: 0.803

Caso 4. Genere con algoritmo Fisherfaces, immagini equalizzate da 64x64 pixel, 100 esempi per gli uomini e 103 per le donne .

Fold n. 1		
	Maschio _P	Femmina _P
Maschio _R	9	3
Femmina _R	1	8

Fold n. 2		
	Maschio _P	Femmina _P
Maschio _R	10	1
Femmina _R	0	10

Fold n. 3		
	Maschio _P	Femmina _P
Maschio _R	8	2
Femmina _R	2	9

Fold n. 4		
	Maschio _P	Femmina _P
Maschio _R	10	2
Femmina _R	0	8

Fold n. 5		
	Maschio _P	Femmina _P
Maschio _R	10	1
Femmina _R	0	9

Fold n. 6		
	Maschio _P	Femmina _P
Maschio _R	8	0
Femmina _R	2	10

6 Risultati sperimentali

Fold n. 7		
	Maschio _P	Femmina _P
Maschio _R	8	2
Femmina _R	2	8

Fold n. 8		
	Maschio _P	Femmina _P
Maschio _R	7	0
Femmina _R	3	10

Fold n. 9		
	Maschio _P	Femmina _P
Maschio _R	7	1
Femmina _R	3	9

Fold n. 10		
	Maschio _P	Femmina _P
Maschio _R	9	1
Femmina _R	1	9

- Precisione media: 0.877
- Richiamo medio: 0.868
- F_1 media: 0.872
- Accuratezza media: 0.867

Caso 5. Genere con algoritmo Fisherfaces, immagini equalizzate da 64x64 pixel, 202 esempi di addestramento per gli uomini e 201 per le donne .

Fold n. 1		
	Maschio _P	Femmina _P
Maschio _R	21	3
Femmina _R	0	17

Fold n. 2		
	Maschio _P	Femmina _P
Maschio _R	16	2
Femmina _R	5	18

6 Risultati sperimentali

Fold n. 3

	Maschio _P	Femmina _P
Maschio _R	17	4
Femmina _R	3	17

Fold n. 4

	Maschio _P	Femmina _P
Maschio _R	17	4
Femmina _R	3	16

Fold n. 5

	Maschio _P	Femmina _P
Maschio _R	18	0
Femmina _R	2	20

Fold n. 6

	Maschio _P	Femmina _P
Maschio _R	16	6
Femmina _R	4	14

Fold n. 7

	Maschio _P	Femmina _P
Maschio _R	18	6
Femmina _R	2	14

Fold n. 8

	Maschio _P	Femmina _P
Maschio _R	18	3
Femmina _R	2	17

Fold n. 9

	Maschio _P	Femmina _P
Maschio _R	17	3
Femmina _R	3	17

Fold n. 10

	Maschio _P	Femmina _P
Maschio _R	16	3
Femmina _R	4	17

- Precisione media: 0.850
- Richiamo medio: 0.846
- F_1 media: 0.848
- Accuratezza media: 0.846

Di seguito, una tabella riassuntiva con i risultati per i vari casi di test.

Tabella 6.1: Risultati dei test sul riconoscimento del genere.

Test	\bar{P}	\bar{R}	\bar{F}_1	\bar{A}
Caso 1	0.827	0.818	0.822	0.817
Caso 2	0.779	0.767	0.773	0.769
Caso 3	0.819	0.803	0.811	0.803
Caso 4	0.877	0.868	0.872	0.867
Caso 5	0.850	0.846	0.848	0.846

Già a partire dai casi 1 e 2 è possibile scorgere la superiorità del metodo Fisherfaces rispetto a Local Binary Patterns Histograms, motivo per il quale il secondo è stato scartato nelle prove successive.

Come si può notare dalla tabella, il caso 4 - ovvero quello che usa l'algoritmo Fisherfaces su 100 immagini equalizzate circa per classe di 64x64 pixel - risulta il migliore. L'aggiunta di nuovi esempi di addestramento non ha condotto a valori più alti nelle metriche considerate, anzi, ha comportato un lieve peggioramento (caso 5).

6.1.2 Test in tempo reale

È stato condotto un test in tempo reale sul riconoscimento del genere. La prova ha visto la partecipazione di 20 soggetti di sesso maschile e 20 di sesso femminile, tutti di nazionalità italiana. La maggior parte dei partecipanti selezionati aveva un'età compresa tra i 18 e i 30 anni.

Il test si è svolto in un laboratorio dove, per sopperire alle scarse condizioni di luce, è stata posta una lampada LED di fronte ai soggetti esaminati.

Si riportano di seguito la matrice di confusione per l'esperimento. Le etichette della matrice contrassegnate con la lettera “R” si riferiscono alle classi reali, quelle contrassegnate dalla lettera “P” alle classi predette.

Tabella 6.2: Matrice di confusione per il test sul genere in tempo reale.

	Maschio _P	Femmina _P
Maschio _R	18	2
Femmina _R	4	16
% Successo	90%	80%

I risultati ottenuti sono solo discreti, se confrontati con lo stato dell'arte (che si attesta su valori anche oltre il 95% sul dataset FERET), tuttavia appaiono incoraggianti: le percentuali di veri positivi per entrambe le classi si stabilizzano su valori accettabili considerando il contesto dinamico in cui sono state ottenute, decisamente più complesso rispetto alle immagini statiche.

6 Risultati sperimentali

Alla pagine seguente è possibile consultare la tabella con i dati grezzi dell'esperimento.

In occasione della stessa sessione di test è stato provato, in via del tutto informale, il funzionamento del sistema in presenza di più di un soggetto. Il comportamento riscontrato è sembrato robusto e soddisfacente.

6 Risultati sperimentali

Tabella 6.3: Risultati dei test sul riconoscimento del genere in tempo reale.

ID Soggetto	Sesso	Previsione	Risultato
Soggetto 1	maschio	maschio	C
Soggetto 2	maschio	maschio	C
Soggetto 3	maschio	maschio	C
Soggetto 4	maschio	maschio	C
Soggetto 5	maschio	maschio	C
Soggetto 6	maschio	maschio	C
Soggetto 7	maschio	maschio	C
Soggetto 8	maschio	maschio	C
Soggetto 9	maschio	maschio	C
Soggetto 10	maschio	maschio	C
Soggetto 11	maschio	maschio	C
Soggetto 12	maschio	maschio	C
Soggetto 13	maschio	femmina	NC
Soggetto 14	maschio	maschio	C
Soggetto 15	maschio	maschio	C
Soggetto 16	maschio	maschio	C
Soggetto 17	maschio	femmina	NC
Soggetto 18	maschio	maschio	C
Soggetto 19	maschio	maschio	C
Soggetto 20	maschio	maschio	C
Soggetto 21	femmina	femmina	C
Soggetto 22	femmina	femmina	C
Soggetto 23	femmina	maschio	NC
Soggetto 24	femmina	femmina	C
Soggetto 25	femmina	maschio	NC
Soggetto 26	femmina	femmina	C
Soggetto 27	femmina	femmina	C
Soggetto 28	femmina	maschio	NC
Soggetto 29	femmina	femmina	C
Soggetto 30	femmina	maschio	NC
Soggetto 31	femmina	femmina	C
Soggetto 32	femmina	femmina	C
Soggetto 33	femmina	femmina	C
Soggetto 34	femmina	femmina	C
Soggetto 35	femmina	femmina	C
Soggetto 36	femmina	femmina	C
Soggetto 37	femmina	femmina	C
Soggetto 38	femmina	femmina	C
Soggetto 39	femmina	femmina	C
Soggetto 40	femmina	femmina	C

6.2 Riconoscimento dell'etnia

6.2.1 Test sugli esempi

Si riportano di seguito i test principali.

Caso 1. Etnia con Fisherfaces, immagini di 88x88 pixel, 111 esempi per l'etnia afro-americana, 112 per le altre (56 caucasici e 56 asiatici)

<i>Fold n. 1</i>		<i>Fold n. 2</i>			
	Afroameric. _P	Altro _P			
Afroameric. _R	12	2	Afroameric. _R	10	1
Altro _R	0	9	Altro _R	1	11

<i>Fold n. 3</i>		<i>Fold n. 4</i>			
	Afroameric. _P	Altro _P			
Afroameric. _R	9	2	Afroameric. _R	9	1
Altro _R	2	10	Altro _R	2	11

<i>Fold n. 5</i>		<i>Fold n. 6</i>			
	Afroameric. _P	Altro _P			
Afroameric. _R	9	0	Afroameric. _R	10	0
Altro _R	2	11	Altro _R	1	11

6 Risultati sperimentali

<i>Fold n. 7</i>			<i>Fold n. 8</i>		
	Afroameric. _P	Altro _P		Afroameric. _P	Altro _P
Afroameric. _R	9	0	Afroameric. _R	8	1
Altro _R	2	11	Altro _R	3	10

<i>Fold n. 9</i>			<i>Fold n. 10</i>		
	Afroameric. _P	Altro _P		Afroameric. _P	Altro _P
Afroameric. _R	10	3	Afroameric. _R	9	3
Altro _R	1	8	Altro _R	2	8

- Precisione media: 0.878
- Richiamo medio: 0.870
- F_1 media: 0.874
- Accuratezza media: 0.870

Caso 2. Etnia con algoritmo Fisherfaces, immagini di 64x64 pixel, 126 esempi di addestramento per l'etnia asiatica, 122 per l'etnia caucasica

6 Risultati sperimentali

<i>Fold n. 1</i>		
	Asiatico _P	Caucasico _P
Asiatico _R	12	1
Caucasico _R	1	11

<i>Fold n. 2</i>		
	Asiatico _P	Caucasico _P
Asiatico _R	12	2
Caucasico _R	1	10

<i>Fold n. 3</i>		
	Asiatico _P	Caucasico _P
Asiatico _R	11	2
Caucasico _R	2	10

<i>Fold n. 4</i>		
	Asiatico _P	Caucasico _P
Asiatico _R	11	4
Caucasico _R	2	8

<i>Fold n. 5</i>		
	Asiatico _P	Caucasico _P
Asiatico _R	12	3
Caucasico _R	1	9

<i>Fold n. 6</i>		
	Asiatico _P	Caucasico _P
Asiatico _R	13	4
Caucasico _R	0	8

<i>Fold n. 7</i>		
	Asiatico _P	Caucasico _P
Asiatico _R	12	0
Caucasico _R	0	13

<i>Fold n. 8</i>		
	Asiatico _P	Caucasico _P
Asiatico _R	11	6
Caucasico _R	1	7

6 Risultati sperimentali

Fold n. 9			Fold n. 10		
	Asiatico _P	Caucasico _P		Asiatico _P	Caucasico _P
Asiatico _R	11	3	Asiatico _R	10	2
Caucasico _R	1	10	Caucasico _R	2	10

- Precisione media: 0.858
- Richiamo medio: 0.847
- F_1 media: 0.852
- Accuratezza media: 0.847

Caso 3. Etnia con algoritmo Fisherfaces, immagini equalizzate di 64x64 pixel, 126 esempi di addestramento per l'etnia asiatica, 122 per l'etnia caucasica .

Fold n. 1			Fold n. 2		
	Asiatico _P	Caucasico _P		Asiatico _P	Caucasico _P
Asiatico _R	12	0	Asiatico _R	13	2
Caucasico _R	1	12	Caucasico _R	0	10

Fold n. 3			Fold n. 4		
	Asiatico _P	Caucasico _P		Asiatico _P	Caucasico _P
Asiatico _R	11	2	Asiatico _R	12	2
Caucasico _R	2	10	Caucasico _R	1	10

6 Risultati sperimentali

<i>Fold n. 5</i>			<i>Fold n. 6</i>		
	Asiatico _P	Caucasico _P		Asiatico _P	Caucasico _P
Asiatico _R	12	2		11	2
Caucasico _R	1	10		2	10

<i>Fold n. 7</i>			<i>Fold n. 8</i>		
	Asiatico _P	Caucasico _P		Asiatico _P	Caucasico _P
Asiatico _R	11	1		11	0
Caucasico _R	1	12		1	13

<i>Fold n. 9</i>			<i>Fold n. 10</i>		
	Asiatico _P	Caucasico _P		Asiatico _P	Caucasico _P
Asiatico _R	12	2		8	1
Caucasico _R	0	11		4	11

- Precisione media: 0.896
- Richiamo medio: 0.890
- F_1 media: 0.893
- Accuratezza media: 0.891

Segue la tabella riassuntiva dei vari casi:

Tabella 6.4: Risultati dei test sul riconoscimento dell'etnia.

Test	\bar{P}	\bar{R}	\bar{F}_1	\bar{A}
Caso 1	0.878	0.870	0.874	0.870
Caso 2	0.858	0.847	0.852	0.847
Caso 3	0.896	0.890	0.893	0.891

Il Caso 1 mostra già buoni risultati per la classificazione dell'etnia afroamericana, considerando, in previsione, di combinare i suoi risultati con il trattamento del colore. I successivi sforzi hanno mirato a migliorare la distinzione fra l'etnia caucasica e quella asiatica: come si può osservare dalla tabella, l'equalizzazione delle immagini ha portato ad un netto incremento in tutti i campi (Caso 2 e Caso 3).

6.2.2 Test sulle immagini

A causa della difficoltà di reperire soggetti di etnie diverse per una prova in tempo reale, è stato svolto un test su 51 immagini per valutare le prestazioni complessive dell'algoritmo di riconoscimento dell'etnia. L'insieme è così composto: 19 immagini per l'etnia afroamericana, divisi in 9 uomini e 10 donne; 17 immagini per l'etnia asiatica divisi in 9 uomini e 8 donne; 15 immagini per l'etnia caucasica divisi in 7 uomini 8 donne.

6 Risultati sperimentali

Tabella 6.5: Matrice di confusione per il test sull'etnia su di un insieme di immagini.

	Afroamericano _P	Asiatico _P	Caucasico _P
Afroamericano _R	16	1	2
Asiatico _R	1	12	4
Caucasico _R	1	0	14
% Successo	84,2%	70,6%	93,3%

I risultati si mostrano abbastanza in linea con quelli ricavati mediante la k-fold cross validation. Le prestazioni del sistema rispetto all'etnia afroamericana appaiono accettabili ma migliorabili, mentre l'unica reale sorpresa è rappresentata dal riconoscimento dei soggetti asiatici: verosimilmente, l'esiguo numero di esempi di addestramento, unito alla difficoltà dell'algoritmo Fisherfaces di individuarne le caratteristiche peculiari e le differenze rispetto ai caucasici - più sottili che negli altri casi sperimentati - hanno costituito il tallone d'Achille del metodo. Ottime sono invece le percentuali di veri positivi ottenute per l'etnia caucasica.

Segue la tabella con i dati grezzi dell'esperimento.

6 Risultati sperimentali

Tabella 6.6: Risultati dei test sul riconoscimento dell'etnia su un insieme di immagini.

ID Soggetto	Etnia	Previsione	Risultato
Soggetto 1	afroamericano	afroamericano	C
Soggetto 2	afroamericano	afroamericano	C
Soggetto 3	afroamericano	afroamericano	C
Soggetto 4	afroamericano	afroamericano	C
Soggetto 5	afroamericano	afroamericano	C
Soggetto 6	afroamericano	afroamericano	C
Soggetto 7	afroamericano	afroamericano	C
Soggetto 8	afroamericano	caucasico	NC
Soggetto 9	afroamericano	afroamericano	C
Soggetto 10	afroamericano	afroamericano	C
Soggetto 11	afroamericano	caucasico	NC
Soggetto 12	afroamericano	asiatico	NC
Soggetto 13	afroamericano	afroamericano	C
Soggetto 14	afroamericano	afroamericano	C
Soggetto 15	afroamericano	afroamericano	C
Soggetto 16	afroamericano	afroamericano	C
Soggetto 17	afroamericano	afroamericano	C
Soggetto 18	afroamericano	afroamericano	C
Soggetto 19	afroamericano	afroamericano	C
Soggetto 20	asiatico	caucasico	NC
Soggetto 21	asiatico	asiatico	C
Soggetto 22	asiatico	asiatico	C
Soggetto 23	asiatico	asiatico	C
Soggetto 24	asiatico	asiatico	C
Soggetto 25	asiatico	asiatico	C
Soggetto 26	asiatico	asiatico	C
Soggetto 27	asiatico	asiatico	C
Soggetto 28	asiatico	caucasico	NC
Soggetto 29	asiatico	asiatico	C
Soggetto 30	asiatico	asiatico	C

6 Risultati sperimentali

ID Soggetto	Etnia	Previsione	Risultato
Soggetto 31	asiatico	asiatico	C
Soggetto 32	asiatico	asiatico	C
Soggetto 33	asiatico	afroamericano	NC
Soggetto 34	asiatico	caucasico	NC
Soggetto 35	asiatico	asiatico	C
Soggetto 36	asiatico	caucasico	NC
Soggetto 37	caucasico	caucasico	C
Soggetto 38	caucasico	caucasico	C
Soggetto 39	caucasico	caucasico	C
Soggetto 40	caucasico	afroamericano	NC
Soggetto 41	caucasico	caucasico	C
Soggetto 42	caucasico	caucasico	C
Soggetto 43	caucasico	caucasico	C
Soggetto 44	caucasico	caucasico	C
Soggetto 45	caucasico	caucasico	C
Soggetto 46	caucasico	caucasico	C
Soggetto 47	caucasico	caucasico	C
Soggetto 48	caucasico	caucasico	C
Soggetto 49	caucasico	caucasico	C
Soggetto 50	caucasico	caucasico	C
Soggetto 51	caucasico	caucasico	C

7 Conclusioni e sviluppi futuri

Il campo della computer vision è uno spazio indefinitamente aperto alla ricerca e alle innovazioni: la capacità di estrapolare automaticamente il contenuto delle immagini - e del materiale multimediale in genere - rappresenta una frontiera entusiasmante per l'informatica. Il riconoscimento dei parametri biometrici, che si insinua in questo contesto, ne rappresenta solo un piccolissimo, seppur importante aspetto.

Il presente lavoro di tesi non pretende di stravolgere i metodi già presenti in letteratura per il conseguimento del compito di riconoscimento, né riesce nell'intento di superare i risultati allo stato dell'arte, ma rappresenta un solido punto di partenza, adoperando le implementazioni più aggiornate dell'ultima versione di OpenCV, in prospettiva di future applicazioni che si affaccino al panorama sempre più dilagante dell'Internet of Things (IoT) e del mondo dei sistemi embedded. Allo stato attuale dello sviluppo, infatti, il framework descritto si presta a un gran numero di aggiunte e miglioramenti.

Innanzitutto, si potrebbe provare a migliorare l'accuratezza degli algoritmi - e altre eventuali metriche di valutazione - secondo nuove modalità, soprattutto per le componenti più deboli del sistema: primo fra tutti il classificatore dell'etnia che si occupa di distinguere gli individui di razza caucasica da quelli di razza asiatica. Inoltre, sempre nello stesso ambito, si potrebbe estendere

7 Conclusioni e sviluppi futuri

il numero delle classi contemplate, così da includere gli altri principali gruppi etnici e consentire al sistema di erogare risposte più precise e più complete. Migliorie nel riconoscimento del genere, in particolare nella sua versione orientata alle prestazioni, sarebbero altrettanto auspicabili, con l'obiettivo di aumentare l'accuratezza a patto di mantenere gestibili i costi computazionali.

In secondo luogo, il riconoscimento di nuovi parametri biometrici e la loro successiva integrazione delineano un concreto prossimo passo di sviluppo, grazie alla predisposizione naturale del sistema di accogliere nuovi moduli e nuovi classificatori. La stima dell'età, il colore degli occhi e dei capelli, la presenza o meno di barba, sono esempi di caratteri la cui analisi andrebbe approfondita, sia presi indipendentemente, sia per via dei potenziali legami con la predizione del genere e dell'etnia, in vista di possibili collaborazioni al fine di aumentare l'accuratezza dei metodi illustrati. In aggiunta, l'individuazione delle espressioni facciali e la presenza di occhiali, renderebbero l'esame delle caratteristiche del volto più soddisfacente e darebbero vita a numerosi altri sbocchi applicativi.

A questo proposito, si consideri, a titolo di esempio, l'idea di un cartellone pubblicitario elettronico, dotato di uno schermo e di una telecamera, che analizza e classifica il volto dei passanti e in risposta personalizza lo spot, nel tentativo di mostrare una pubblicità sempre mirata e potenzialmente interessante. Si aggiunga il fatto che, lo stesso dispositivo, potrebbe raccogliere e inviare ai centri di elaborazione dati, statistiche sul numero e la tipologia dei passanti che si sono soffermati sul cartellone, e su una certa pubblicità.

Da ultimo, si potrebbe considerare quello della sicurezza uno degli ambiti più fertili per lo sviluppo di questa tecnologia. Un esempio di uso pratico in questo settore potrebbe essere il riconoscimento di potenziali terroristi

7 Conclusioni e sviluppi futuri

in luoghi pubblici, con individuazione e tracking dei soggetti potenzialmente pericolosi, a partire dalle immagini di telecamere di videosorveglianza. Infine, sempre nello stesso campo, la classificazione dell'età di un individuo potrebbe trovare il suo impiego nella certificazione via webcam della maggiore età, da effettuarsi preliminarmente alla distribuzione di materiale o all'erogazione di servizi per adulti su Internet.

Concludendo, sono stati evidenziati diversi stimolanti scenari di sviluppo futuro. Con questi in mente, forti sono le intenzioni di arricchire il sistema di nuove funzionalità e raffinare quelle già presenti, al fine di allinearla alle più recenti scoperte teoriche e alle più efficaci implementazioni divulgata dalla comunità scientifica.

Ringraziamenti

Si ringrazia la professoressa De Carolis per la sua piena disponibilità, la sua fervente creatività, il sano e palpabile ottimismo.

Si ringrazia il dottor Giuseppe Palestra per la sua pragmaticità, il rigore organizzativo e il caldo sostegno morale nei momenti più avvillenti.

Un ringraziamento speciale al mio caro amico e collega Luigi Quaranta, che ha contribuito insieme a me far nascere e crescere il progetto di tesi, plasmandolo con impegno e determinazione, fedele compagno di studio e di risate.

Un ringraziamento speciale alla mia ragazza, detta *Pupa*, per l'amore e l'affetto dimostrati anche nei periodi più difficili, di più forte irascibilità e frustrazione, di questo intenso percorso accademico.

Una affettuoso ringraziamento alla mia famiglia che mi è sempre stata accanto, e a mio fratello, per i suoi preziosi consigli e la sua esperienza.

Una menzione speciale ai miei amici più vicini di vecchia data, Marcello con cui condivido la passione per la musica e Dario con cui condivido quella per i videogiochi;

Una menzione speciale anche per Edoardo Vignola, la cui bonarietà e infinita pazienza, insegnano come la vita vada vissuta con filosofia. Almeno finché qualcuno non porta il Jack Daniels.

7 Conclusioni e sviluppi futuri

Un ringraziamento ai miei colleghi di facoltà Roberto, Danilo e Francesco con cui ho vissuto l’esperienza universitaria con maggiore leggerezza ed entusiasmo.

Un ringraziamento, infine, a tutti coloro che hanno creduto in me e nelle mie capacità, coloro che mi hanno aiutato, con le parole o con le azioni, a continuare a combattere ogni giorno per raggiungere i miei obiettivi, concretizzare i miei sogni e seguire le mie aspirazioni. Un sentito *Grazie*.

Elenco delle figure

2.1	Si può notare come le caratteristiche della faccia individuate dal metodo PCA includano anche l'informazione luminosa.	15
2.2	Un esempio di applicazione del metodo Fisherfaces.	17
2.3	Passi per la costruzione di un descrittore LBP.	20
2.4	Un esempio di applicazione del metodo LBP.	20
2.5	Da sinistra verso destra: 1) L'immagine originale; 2) La componente Y; 3) La componente Cb; 4) La componente Cr.	22
2.6	Una rappresentazione del modello HSL/HSV	23
3.1	Le tre tipologie di caratteristiche elaborate dal metodo Haar Cascade.	28
3.2	Processo di elaborazione a cascata di un classificatore Haar Cascade.	28

Elenco delle figure

3.3 L'oggetto da classificare (cerchio verde) potrebbe essere assegnato alla prima classe (quadrati blu), o nella seconda (triangoli rossi). Se k=3 (cerchio con linea continua), esso è assegnato alla seconda classe perché ci sono 2 triangoli e soltanto un quadrato all'interno del cerchio. Se k=5 (cerchio con linea tratteggiata), l'oggetto è invece assegnato alla prima classe: ci sono 3 quadrati e soltanto due triangoli all'intero del cerchio più esterno.	30
3.4 In uno spazio bidimensionale un separatore è una linea retta	31
3.5 Il miglior iperpiano è quello che massimizza il margine geometrico	32
4.1 Un'immagine esemplificativa del processo di elaborazione di una CNN.	37
5.1 Immagini prese da <i>The ORL Database of Faces</i>	40
5.2 Immagini prese dal <i>MDB-WIKI dataset</i>	41
5.3 Operazioni di pre-processamento	43
5.4 La rappresentazione di un'immagine mediante istogrammi. .	44
5.5 Esempio di equalizzazione dell'immagine	44
5.6 Una visione esemplificativa dell'architettura di sistema. . .	45
5.7 Funzionamento del sistema nel caso di immagini differenti in ingresso.	47
5.8 Funzionamento del sistema nel caso di una sequenza video in ingresso.	48
5.9 Problema della classificazione del genere	50

Elenco delle figure

5.10 Scelte effettuate per il riconoscimento del genere nell'approccio orientato alle prestazioni	52
5.11 Approccio al problema della classificazione dell'etnia.	58
5.12 Home page del sito.	59
5.13 Interfaccia di caricamento dell'immagine e visualizzazione dei risultati.	60

Bibliografia

- [1] T D'Orazio A Distante T Martiriggiano, M Leo. in innovations in applied artificial intelligence. lecture notes in computer science, ed. by m ali, f esposito. face recognition by kernel independent component analysis, vol. 3533 (springer, berlin heidelberg, 2005), pp. 55–58.
- [2] L. Jain, A.; Hong and S. Pankanti. Biometric identification, communications of the acm, 43(2), p. 91–98.
- [3] Sarat C.; Nandakumar Karthik Anil K. Jain, Dass. Soft biometric traits for personal recognition systems. *International Conference on Biometric Authentication*, 2004.
- [4] Soft biometrics, wikipedia, 2014.
- [5] T. Kanade. Picture processing system by computer complex and recognition of human faces. phd thesis, kyoto university. November 1973.
- [6] Poggio T. Brunelli, R. Face recognition through geometrical features. *European Conference on Computer Vision (ECCV)*, 1992, S. 792–800.
- [7] Joao P. Hespanha Peter N. Belhumeur and David J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection.

Bibliografia

IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, 19(7):711–720, July 1997.

- [8] Fellous J. Krüger N. Malsburg C. Wiskott, L. Face recognition by elastic bunch graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19, pages 775–779, 1997.
- [9] K. et al. Messer. Performance characterisation of face recognition algorithms and their sensitivity to severe illumination changes. *icb*. pages 1–11, 2006.
- [10] Hadid A. Ahonen, T. and M. Pietikainen. Face recognition with local binary patterns. *computer vision. ECCV 2004*, pages 469–481, 2004.
- [11] Pierluigi Carcagnì, Marco Del Coco, Dario Cazzato, Marco Leo, and Cosimo Distante. A study on different experimental configurations for age, race, and gender estimation problems. *EURASIP Journal on Image and Video Processing*, 2015(1):1–22, 2015.
- [12] Giovanna Farinella and Jean-Luc Dugelay. Demographic classification: Do gender and ethnicity affect each other? In *Informatics, Electronics & Vision (ICIEV), 2012 International Conference on*, pages 383–390. IEEE, 2012.
- [13] Rabah Attia Anis Youssef Marwa Jmal, Wided Soidene Mseddi. Classification of human skin color and its application to face recognition. the sixth international conferences on advances in multimedia. 2014.

Bibliografia

- [14] Mayank Vatsa Richa Singh Sanjay Kr. Singh, D. S. Chauhan. A robust skin color based face detection algorithm. *Tamkang Journal of Science and Engineering*, 6(4):227–234, 2003.
- [15] Y. Adini Y. Moses and S. Ullman. Face recognition: The problem of compensating for changes in illumination direction. In *European Conf. Computer Vision*, pages 286–296, 1994.
- [16] Visual descriptor, wikipedia.
- [17] R. W. G. Hunt. *The Reproduction of Colour*. Chichester UK, Wiley – IS&T Series in Imaging Science and Technology, 6th edition, 2004.
- [18] Michael Jones Paul Viola. Rapid object detection using a boosted cascade of simple features. *Accepted Conference on Computer Vision and Pattern Recognition*, 2001.
- [19] Yoav Freund and Robert E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Computational Learning Theory: Eurocolt '95*, pages 23–27, Springer-Verlag, 1995.
- [20] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. Caffe: Convolutional architecture for fast feature embedding. *arXiv preprint arXiv:1408.5093*, 2014.
- [21] Deep learning, wikipedia, 2016.
- [22] V. Bruce et al. Sex discrimination: how do we tell the difference between male and female faces?". *Perception*, 22(2):131–152, 1993.

Bibliografia

- [23] Rasmus Rothe, Radu Timofte, and Luc Van Gool. Dex: Deep expectation of apparent age from a single image. In *ICCV, ChaLearn Looking at People workshop*, December 2015.