

ACQUISIZIONE NON INVASIVA DELL'AZIONE UMANA

Si vogliono introdurre alcuni concetti di base riguardanti l'acquisizione non invasiva dell'azione umana mediante tecniche di visione.

ACQUISIZIONE NON INVASIVA DELL'AZIONE UMANA

Espressioni facciali – Introduzione

L'espressione facciale è, oltre al parlato e ai gesti, un canale primario di scambio di informazioni durante la comunicazione naturale. Con la mimica facciale, le emozioni possono essere espresse in modo più rapido, preciso ed efficace di quanto non si possa fare mediante l'uso delle parole.

Le espressioni facciali sono utilizzate attivamente da un relatore per supportare la comunicazione, ma anche intuitivamente per mimare il feedback di un interlocutore. In medicina, i sintomi e i progressi nella terapia possono essere letti dal volto. I linguaggi dei segni utilizzano le espressioni facciali come canale di comunicazione aggiuntivo rispetto alla sintassi e alla semantica di una conversazione. I bambini si concentrano sui volti già dopo pochi minuti dalla nascita e distinguono le persone.

Il battito delle palpebre è applicato, ad esempio, nell'industria automobilistica per stimare la stanchezza del conducente, il contorno delle labbra e il movimento contribuiscono alla robustezza dei riconoscitori vocali e il riconoscimento facciale è utilizzato come metodo biometrico.

ACQUISIZIONE NON INVASIVA DELL'AZIONE UMANA

Espressioni facciali – Introduzione

Nella figura sono riportati alcuni dispositivi hardware per l'acquisizione di espressioni facciali.

Sinistra - Camera montata sulla testa e markers sulla faccia

Destra - Camera montata sulla testa e dispositivo meccanico



ACQUISIZIONE NON INVASIVA DELL'AZIONE UMANA

Espressioni facciali – Introduzione

L'uso di tecniche non invasive per l'acquisizione e il riconoscimento di espressioni facciali prevede di disporre di immagini digitali come inputs. Il fatto che l'aspetto dell'utente possa variare rappresenta una notevole difficoltà per tali obiettivi.

Le variazioni dell'aspetto dell'utente possono avere cause intrinseche ed estrinseche. Le variazioni intrinseche derivano ad esempio dalla forma del viso e dai movimenti effettuati durante la conversazione, mentre gli effetti estrinseci sono dovuti ad esempio alla posizione della camera, all'illuminazione e a possibili occlusioni.

I volti umani sono oggetti non rigidi con un alto grado di variabilità in termini di dimensioni, forma, colore e superficie.

Database contenenti volti sono molto considerati per la sperimentazione dei diversi algoritmi di riconoscimento delle espressioni facciali.

Qualsiasi sistema automatico per il riconoscimento del volto e dei movimenti del volto ha un immenso potenziale nell'identificazione di criminali, nella sorveglianza, nella medicina, nelle interfacce uomo-macchina, ...

ACQUISIZIONE NON INVASIVA DELL'AZIONE UMANA

Espressioni facciali – Introduzione

Ci sono due ragioni per considerare l'umore dell'uomo.

In primo luogo, cosa sono le emozioni? Uno stato mentale che sorge spontaneamente piuttosto che attraverso uno sforzo cosciente. Tale stato mentale è spesso accompagnato da cambiamenti fisiologici e questi cambiamenti fisiologici sono riconosciuti dal mondo esterno. A seconda dei cambiamenti di umore, si può facilmente capire lo stato d'animo.

In secondo luogo, le principali emozioni umane sono soggette a cambiamenti nella vita quotidiana: gioia (happiness), disgusto (disgust), rabbia (anger), tristezza (sadness), sorpresa (surprise) e paura (fear). Queste vengono anche definite come le emozioni universali.

ACQUISIZIONE NON INVASIVA DELL'AZIONE UMANA

Espressioni facciali – Introduzione

Gioia – La gioia è caratterizzata da un'espressione facciale che fa alzare gli angoli della bocca verso l'alto.



Esempio di gioia.

ACQUISIZIONE NON INVASIVA DELL'AZIONE UMANA

Espressioni facciali – Introduzione

Disgusto – Il disgusto è caratterizzato da un'espressione facciale che fa sollevare il labbro superiore, corrugare il ponte nasale e alzare le guance. Sorprendentemente, però, il disgusto è una delle emozioni che fanno diminuire la frequenza cardiaca.



Esempio di disgusto.

ACQUISIZIONE NON INVASIVA DELL'AZIONE UMANA

Espressioni facciali – Introduzione

Paura – Quando si parla di paura, si dice che questa provochi l'innalzamento delle sopracciglia, l'apertura della bocca e uno spalancamento degli occhi maggiore rispetto al normale.



Esempio di paura.

ACQUISIZIONE NON INVASIVA DELL'AZIONE UMANA

Espressioni facciali – Introduzione

Rabbia – La rabbia è caratterizzata da un'espressione facciale che fa abbassare le sopracciglia, fa stringere con forza le labbra e fa sporgere gli occhi.



Esempio di rabbia.

ACQUISIZIONE NON INVASIVA DELL'AZIONE UMANA

Espressioni facciali – Introduzione

Sorpresa – La sorpresa è caratterizzata da un'espressione facciale che inarca le sopracciglia, apre gli occhi e fa abbassare la mascella.



Esempio di sorpresa.

ACQUISIZIONE NON INVASIVA DELL'AZIONE UMANA

Espressioni facciali – Introduzione

Tristezza – La tristezza è caratterizzata da un'espressione facciale che fa abbassare gli angoli della bocca e fa alzare la parte interna delle sopracciglia.



Esempio di tristezza.

ACQUISIZIONE NON INVASIVA DELL'AZIONE UMANA

Espressioni facciali – Database

Le immagini utilizzate per il riconoscimento delle espressioni facciali sono immagini statiche o sequenze di immagini. Una sequenza di immagini contiene potenzialmente più informazioni di un'immagine statica, in quanto una sequenza di immagini raffigura anche le informazioni temporali.

In letteratura, esistono diversi database di immagini (espressioni facciali) che possono essere utilizzati ai fini della ricerca. Un esempio di database è il database Cohn-Kanade (CK+). Alcuni esperimenti su questo database si sono basati sui dati descritti in seguito. I soggetti presenti erano circa 100 studenti universitari iscritti a corsi introduttivi di psicologia. L'età variava dai 18 ai 30 anni. Il 65% era di sesso femminile, il 15% era afroamericano e il 3% era asiatico o latinoamericano. La sala di acquisizione era dotata di una sedia per il soggetto e di due camere Panasonic WV3230, ognuna delle quali era connessa a un registratore video Panasonic S-VHS AG-7500. Una delle telecamere era situata direttamente di fronte al soggetto e l'altra era posizionata a 30° sulla destra del soggetto. Le immagini utilizzate per gli esperimenti fanno riferimento alla camera frontale.

ACQUISIZIONE NON INVASIVA DELL'AZIONE UMANA

Espressioni facciali – Database



Esempio di dataset del database CK+.

ACQUISIZIONE NON INVASIVA DELL'AZIONE UMANA

Espressioni facciali – Database

I soggetti sono stati istruiti a eseguire una serie di 23 espressioni facciali che includevano singole *action units* (AUs) o unità d'azione (ad esempio, AU 12, o angoli delle labbra tirati obliquamente) e combinazioni di unità d'azione (ad esempio, AU 1 + 2, o sopracciglia interne ed esterne sollevate).

Ogni emozione iniziava con un'espressione neutra e terminava con l'espressione facciale selezionata. Prima di effettuare ognuna delle espressioni facciali selezionate, l'espressione facciale è stata descritta e modellizzata.

Sono disponibili espressioni facciali associate alle sei emozioni universali. Le sequenze di immagini dall'espressione neutra all'espressione facciale desiderata sono state digitalizzate in immagini caratterizzate da 640 x 480 o 640 x 490 pixels (immagini in scala di grigi con precisione di 8 bit, formato .png e .jpg).

Le sei emozioni e l'apertura della bocca in assenza di altre *action units* sono state annotate (*labeled*) con il sistema di codifica FACS (Facial Action Coding System), il quale descrive l'espressione del soggetto in termini di *action units*. Inoltre, sono presenti anche etichette (*labels*) associate alle emozioni. Quindi, le sequenze possono essere analizzate sia con le *action units* sia con i prototipi delle emozioni.

ACQUISIZIONE NON INVASIVA DELL'AZIONE UMANA

Espressioni facciali – Database

Il FACS è stato introdotto da Ekman e Friesen (1978) ed esso è caratterizzato da 64 AUs. Assumendo che tutti i muscoli facciali siano rilassati durante un'espressione facciale neutra, ogni AU modella la contrazione di un certo insieme di essi, portando a deformazioni sulla faccia. Le features facciali predominanti sono gli occhi, le sopracciglia, il naso e la bocca. La loro forma e il loro aspetto contengono una grande quantità di informazione.

ACQUISIZIONE NON INVASIVA DELL'AZIONE UMANA

Espressioni facciali – Distance Signature

Un possibile approccio per il riconoscimento delle espressioni facciali si basa sul concetto di *distance signature*. Tale approccio utilizza un modello *Active Appearance Model* (AAM) al fine di rilevare dei punti di riferimento (*landmarks*) sulle immagini associate alle espressioni facciali. I punti di riferimento principali vengono identificati all'interno e nell'intorno delle regioni associate alle sopracciglia, agli occhi, alla bocca e al naso di una tipica immagine che rappresenta la faccia umana.

Tra questi punti di riferimento, vengono considerati tre punti per ogni sopracciglio, quattro punti per ogni occhio, tre punti per il naso e quattro punti associati alla regione della bocca.

Sulla base di questi punti di riferimento, viene formata una griglia (grid) sull'immagine del volto. Viene quindi misurata la distanza Euclidea all'interno della griglia. Le distanze calcolate vengono quindi normalizzate, ottenendo la *normalized distance signature*.

In seguito, viene calcolato un indice di stabilità (*stability index*) di questa *normalized distance signature*.

ACQUISIZIONE NON INVASIVA DELL'AZIONE UMANA

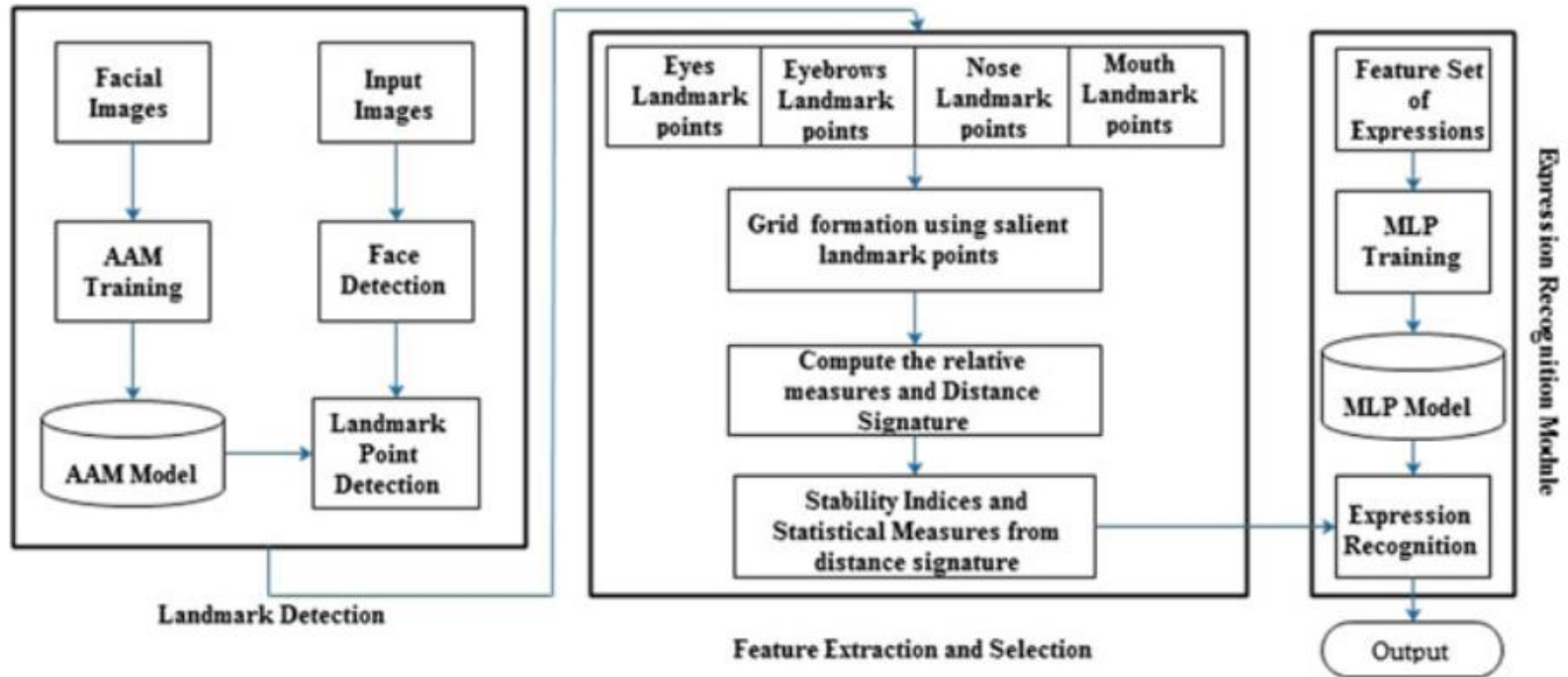
Espressioni facciali – Distance Signature

Vengono inoltre calcolate alcune misure statistiche a partire dalla *normalized distance signature*: range, momento, asimmetria (skewness), curtosi ed entropia.

Tutte le features calcolate vengono sfruttate al fine di riconoscere le differenti categorie di emozioni. Queste features vengono quindi utilizzate come input per l'addestramento (training) di una rete neurale artificiale e poi per la classificazione delle sei espressioni facciali di base.

ACQUISIZIONE NON INVASIVA DELL'AZIONE UMANA

Espressioni facciali – Distance Signature

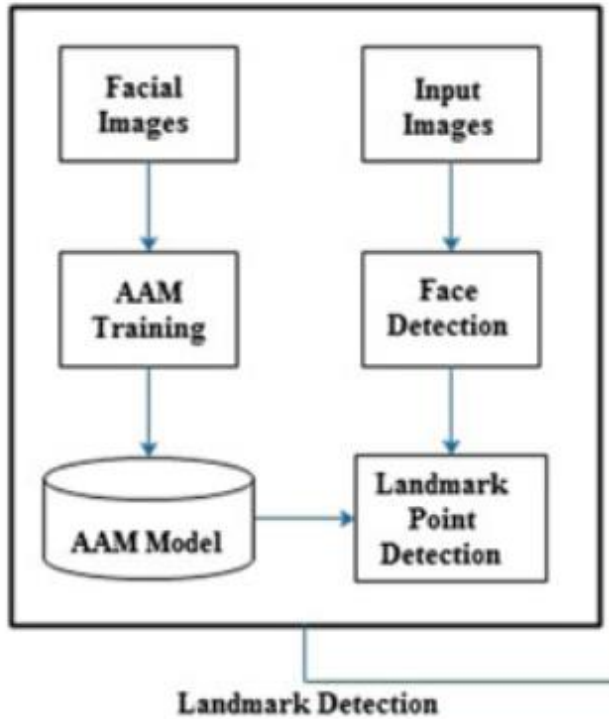


Schema a blocchi dell'approccio basato sulla distance signature.

ACQUISIZIONE NON INVASIVA DELL'AZIONE UMANA

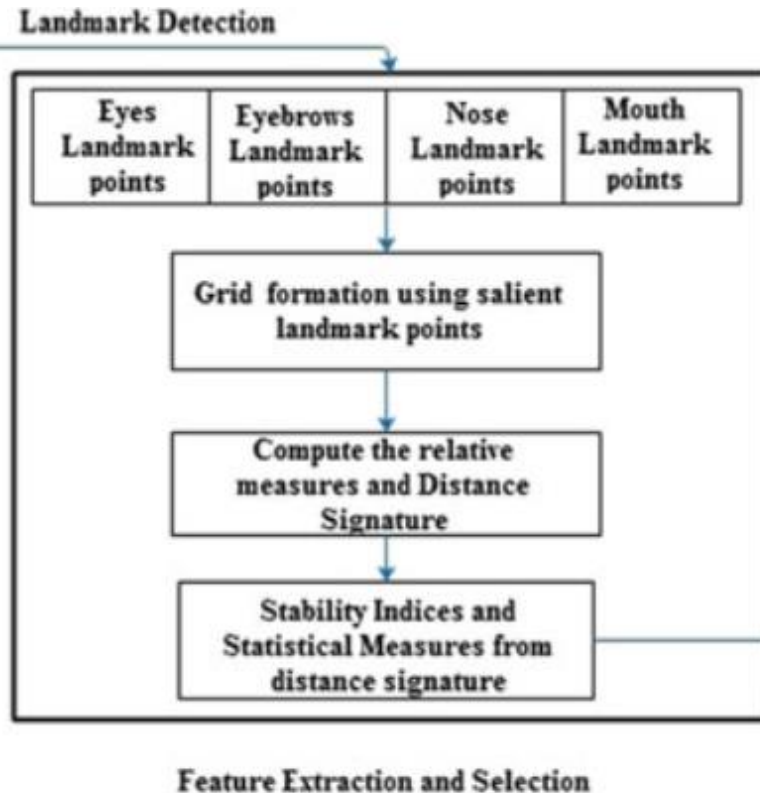
Espressioni facciali – Distance Signature

La tecnica basata su *Active Appearance Model* (AAM) viene utilizzata per rilevare i punti di riferimento dalle immagini del volto.



ACQUISIZIONE NON INVASIVA DELL'AZIONE UMANA

Espressioni facciali – Distance Signature



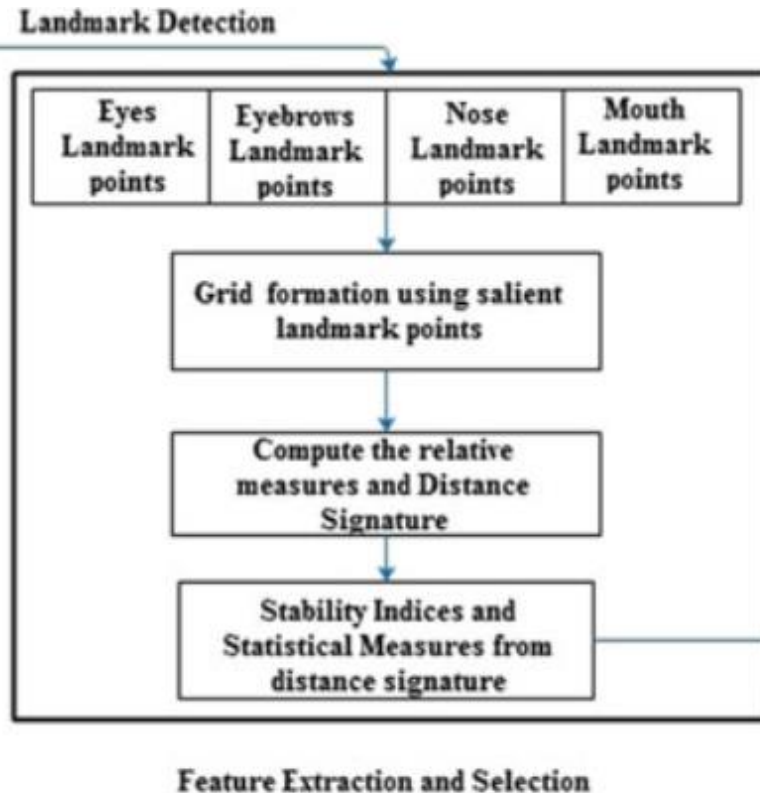
Alcuni punti di riferimento principali vengono estratti da sopracciglia, occhi, bocca e naso per la formazione di una griglia. Tale griglia viene costruita sull'immagine del volto con i punti di riferimento estratti (21).

I 21 punti di riferimento estratti vengono collegati tra loro e riportati in una matrice per la formazione di una griglia.

Viene calcolata la distanza Euclidea per ogni coppia di punti di riferimento della griglia. Le distanze calcolate vengono normalizzate, ottenendo la *normalized distance signature*.

ACQUISIZIONE NON INVASIVA DELL'AZIONE UMANA

Espressioni facciali – Distance Signature



Le proprietà associate al concetto di *distance signature* possono essere applicate per il riconoscimento dell'espressione facciale.

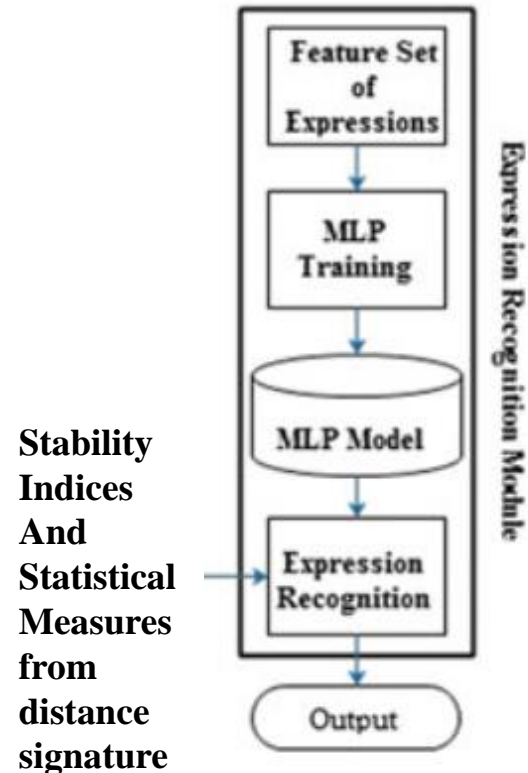
Un indice di stabilità viene derivato dalla *normalized distance signature*.

La *normalized distance signature* viene utilizzata anche per calcolare alcune features statistiche come ad esempio il range, il momento, la curtosi, l'asimmetria e l'entropia in modo da arricchire il significato dell'insieme delle features.

ACQUISIZIONE NON INVASIVA DELL'AZIONE UMANA

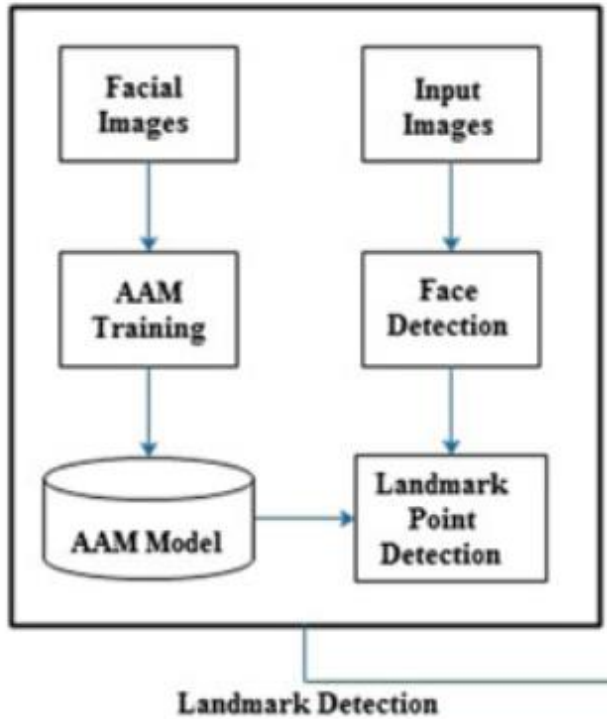
Espressioni facciali – Distance Signature

Le features vengono utilizzate come input per una rete neurale artificiale di tipo MLP (Multilayer Perceptron) per il riconoscimento delle emozioni a partire da espressioni facciali.



ACQUISIZIONE NON INVASIVA DELL'AZIONE UMANA

Espressioni facciali – Distance Signature



Una efficace fase di estrazione delle features è fondamentale per il riconoscimento delle emozioni. Per tale motivo, la corretta identificazione dei punti di riferimento nelle immagini del volto è una sfida importante.

I punti di riferimento sono affidabili per l'estrazione di un insieme di features significativo per le deformazioni e gli spostamenti dei componenti facciali durante la variazione delle emozioni.

Alcuni punti di riferimento sono caratterizzati da proprietà comuni nel cambiamento di espressioni facciali e alcuni di essi hanno caratteristiche importanti per categorizzare le emozioni da una espressione facciale all'altra.

ACQUISIZIONE NON INVASIVA DELL'AZIONE UMANA

Espressioni facciali – Distance Signature



Original Image

Using
AAM
Model

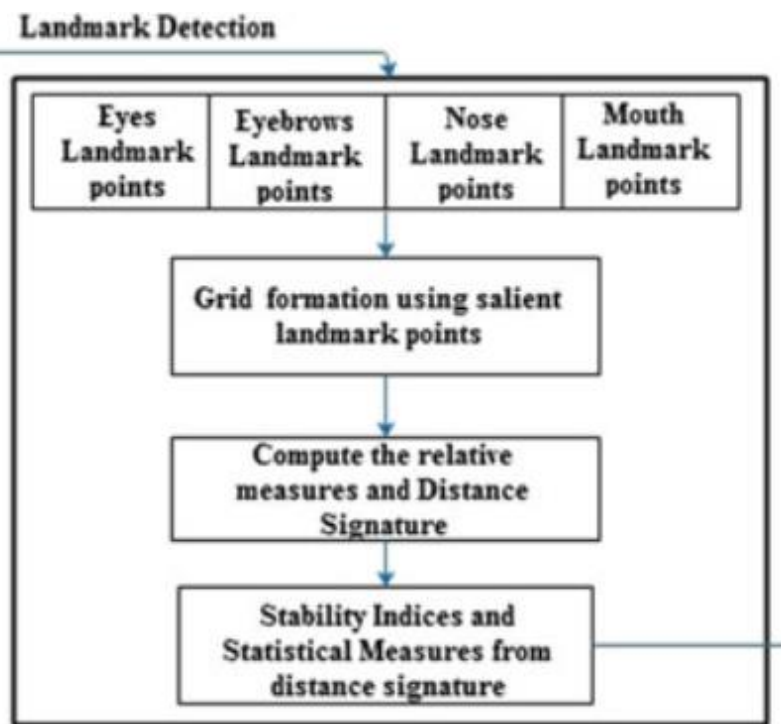


Detected Landmark Points

Rilevamento dei punti di riferimento dell'espressione facciale associata alla gioia mediante AAM.

ACQUISIZIONE NON INVASIVA DELL'AZIONE UMANA

Espressioni facciali – Distance Signature



Feature Extraction and Selection

Gli AAM sono utilizzati per ottenere un insieme di punti di riferimento sull'immagine. Tali punti di riferimento vengono quindi utilizzati per identificare i punti di riferimento principali (21) su bocca, naso, sopracciglia e occhi grazie alla loro sensitività a cambiamenti dell'espressione facciale. Questi punti di riferimento principali sono marcati sulle sopracciglia (uno in posizione centrale tra i due punti di corner e altri due sono punti di corner), sugli occhi (due punti sono in posizione centrale tra palpebra superiore e palpebra inferiore, due sono punti di

corner), sulla bocca (due punti in posizione centrale del labbro superiore e del labbro inferiore, due sono punti di corner delle labbra) e sul naso (due punti di corner della narice e il rimanente nella posizione superiore del naso).

ACQUISIZIONE NON INVASIVA DELL'AZIONE UMANA

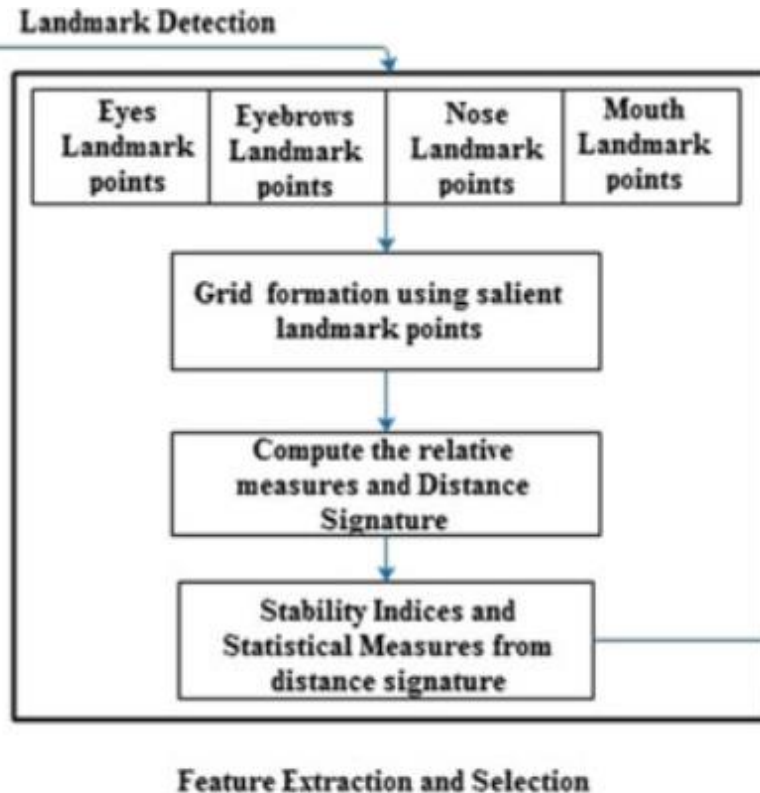
Espressioni facciali – Distance Signature



Rilevamento dei punti di riferimento principali dell'espressione facciale associata alla gioia mediante AAM.

ACQUISIZIONE NON INVASIVA DELL'AZIONE UMANA

Espressioni facciali – Distance Signature



A questo punto, i punti di riferimento principali identificati vengono connessi gli uni agli altri, formando una griglia.

Se sono stati identificati 21 punti di riferimento principali, si hanno 210 possibili distanze ($\frac{21 \cdot (21-1)}{2}$) nella griglia. I 21 punti di riferimento vengono connessi tra loro in modo da formare una griglia.

Si ha quindi che il processo di formazione della griglia si basa sulle seguenti fasi:

- 1) Identificazione dei punti di riferimento utilizzando gli AAM.
- 2) Rilevamento dei punti di riferimento principali nelle regioni di bocca, naso, sopracciglia e occhi.
- 3) I punti di riferimento principali vengono connessi tra loro formando una griglia.

ACQUISIZIONE NON INVASIVA DELL'AZIONE UMANA

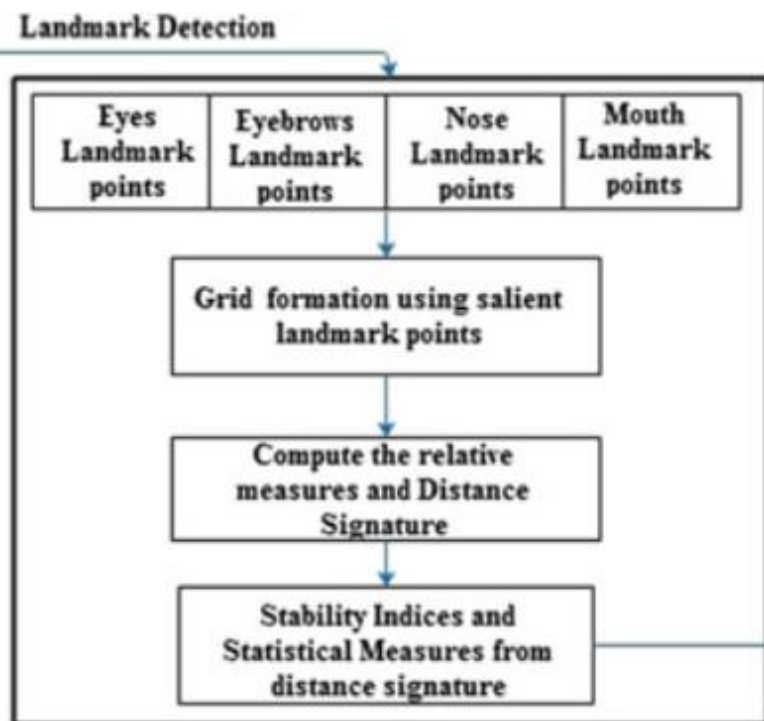
Espressioni facciali – Distance Signature



Formazione della griglia utilizzando i punti di riferimento principali dell'espressione facciale associata alla gioia.

ACQUISIZIONE NON INVASIVA DELL'AZIONE UMANA

Espressioni facciali – Distance Signature



Feature Extraction and Selection

Il riconoscimento di espressioni facciali si basa in larga misura su un'appropriata selezione delle features facciali dalle immagini del volto. Tali features devono avere la capacità di permettere distinzioni precise nel riconoscimento di espressioni facciali. In un sistema di riconoscimento facciale, possono essere quindi considerate diverse features.

Consideriamo le regioni delle sopracciglia, degli occhi, della bocca e del naso per estrarre i punti di riferimento appropriati sulle immagini del volto. Questi punti di riferimento principali

vengono utilizzati per formare una griglia sull'immagine del volto. All'interno della griglia sono disponibili 210 distanze distinte. Si può utilizzare ad esempio la distanza Euclidea (d) tra coppie di punti di riferimento della griglia:

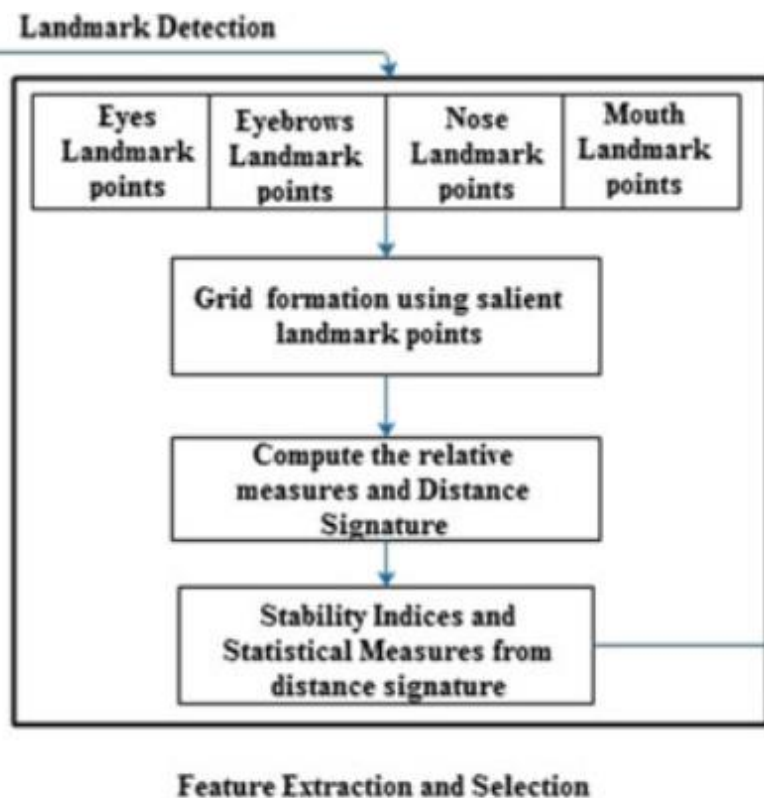
$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

(x_1, y_1) : primo punto

(x_2, y_2) : secondo punto

ACQUISIZIONE NON INVASIVA DELL'AZIONE UMANA

Espressioni facciali – Distance Signature



Tali distanze d vengono normalizzate ottenendo le *normalized distance signatures*:

$$\beta_i^f = \frac{d_i^f}{\sum_{i=1}^n d_i^f} \quad f = 1, 2, \dots, m \text{ and } i = 1, 2, \dots, n$$

dove m è il numero totale di immagini del volto e n è il numero totale di distanze.

Un insieme di valori β_i^f ordinati costituisce la *distance signature*.

I vari termini d_i^f sono le distanze (calcolate come d nella slide precedente). β_i^f sono distanze normalizzate e vengono considerate come una *distance signature*, la quale rappresenta una feature.

Al fine di comprendere meglio il concetto di *distance signature*, consideriamo l'esempio riportato nelle figure delle due slide seguenti.

ACQUISIZIONE NON INVASIVA DELL'AZIONE UMANA

Espressioni facciali – Distance Signature

$m=6$
 $n=210$



a) Surprise



b) Happiness



c) Sadness



d) Fear



e) Disgust



f) Anger

Sei emozioni vengono utilizzate per calcolare la distance signature relativa al database CK+.

ACQUISIZIONE NON INVASIVA DELL'AZIONE UMANA

Espressioni facciali – Distance Signature

$m=6$
 $n=210$

	a)	b)	c)	d)	e)	f)
	1	2	3	4	5	6
1	0.0015	0.0017	0.0017	0.0017	0.0018	0.0017
2	0.0015	0.0017	0.0017	0.0017	0.0018	0.0017
3	0.0025	0.0028	0.0027	0.0028	0.0027	0.0029
4	0.0015	0.0017	0.0018	0.0017	0.0018	0.0016
5	0.0014	0.0016	0.0017	0.0015	0.0017	0.0015
6	0.0074	0.0082	0.0082	0.0080	0.0083	0.0081
7	0.0055	0.0059	0.0059	0.0056	0.0056	0.0058
8	0.0061	0.0067	0.0067	0.0067	0.0068	0.0067
9	0.0066	0.0072	0.0071	0.0072	0.0074	0.0074

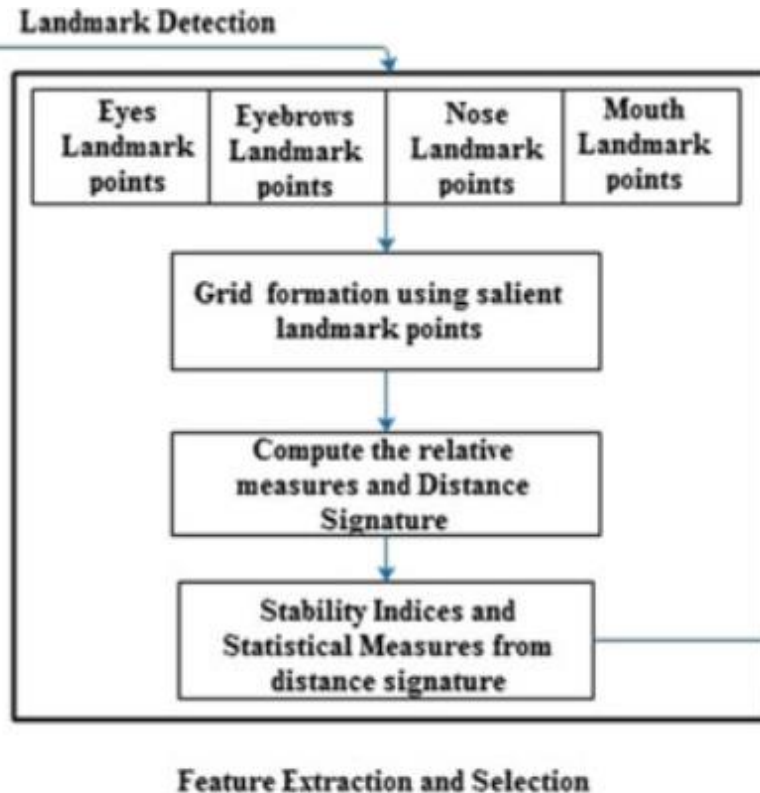
199	0.0077	0.0078	0.0078	0.0077	0.0077	0.0077
200	0.0093	0.0072	0.0067	0.0077	0.0070	0.0068
201	0.0032	0.0034	0.0034	0.0032	0.0031	0.0033
202	0.0045	0.0050	0.0048	0.0049	0.0050	0.0050
203	0.0033	0.0035	0.0035	0.0035	0.0031	0.0035
204	0.0067	0.0074	0.0074	0.0073	0.0076	0.0074
205	0.0063	0.0068	0.0068	0.0067	0.0069	0.0068
206	0.0036	0.0040	0.0039	0.0038	0.0039	0.0039
207	0.0079	0.0076	0.0079	0.0076	0.0075	0.0079
208	0.0067	0.0072	0.0073	0.0071	0.0072	0.0073
209	0.0075	0.0065	0.0070	0.0066	0.0071	0.0070
210	0.0091	0.0071	0.0066	0.0075	0.0069	0.0066

Distance signature di sei emozioni del database CK+.

TUALE

ACQUISIZIONE NON INVASIVA DELL'AZIONE UMANA

Espressioni facciali – Distance Signature



Un'altra feature molto importante per il riconoscimento di espressioni facciali è rappresentata dall'indice di stabilità (stability index) delle immagini del volto. L'indice di stabilità è sensibile per la classificazione di espressioni facciali.

Consideriamo la seguente equazione:

$$(\mu_i^r)^f = \frac{(d_i^r)^f}{(\sum_{i=1}^n d_i^r)^f} \quad f = 1, 2, \dots, m \text{ and } i = 1, 2, \dots, n$$

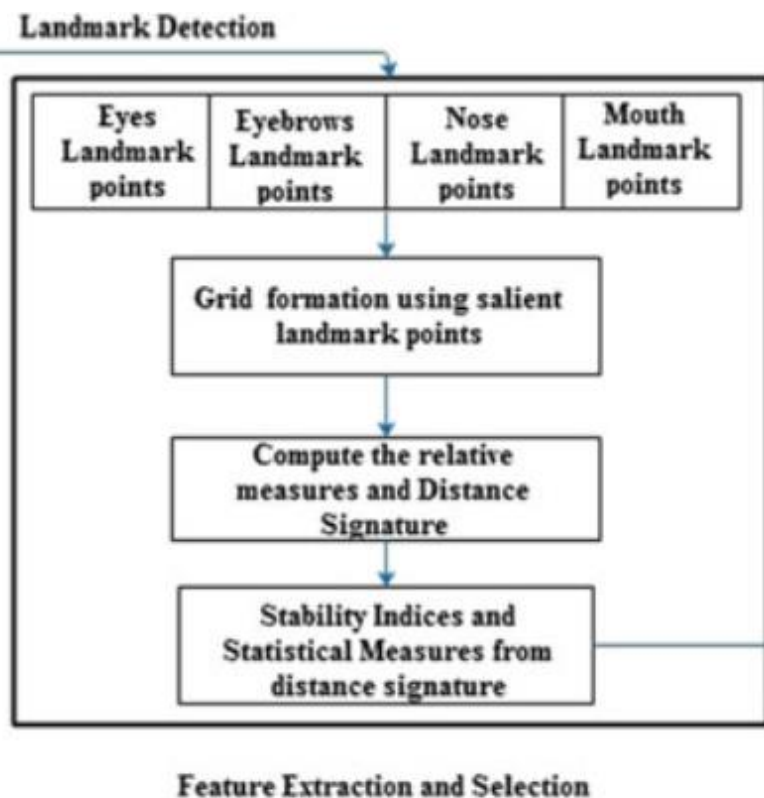
Tale equazione permette di calcolare la *normalized distance signature* di ordine r per ogni immagine del volto.

Calcoliamo quindi l'incremento di ordine r come

$$\Delta(d_i^r)^f = (\mu_i^r)^f - (\mu_i^{r-1})^f$$

ACQUISIZIONE NON INVASIVA DELL'AZIONE UMANA

Espressioni facciali – Distance Signature



$$\Delta(d_i^r)^f = (\mu_i^r)^f - (\mu_i^{r-1})^f$$

Definiamo una soglia λ al fine di analizzare tali incrementi. Data una (piccola) soglia positiva (predefinita) λ , troviamo l'ordine $r1$ più piccolo per il quale si verifica per la prima volta:

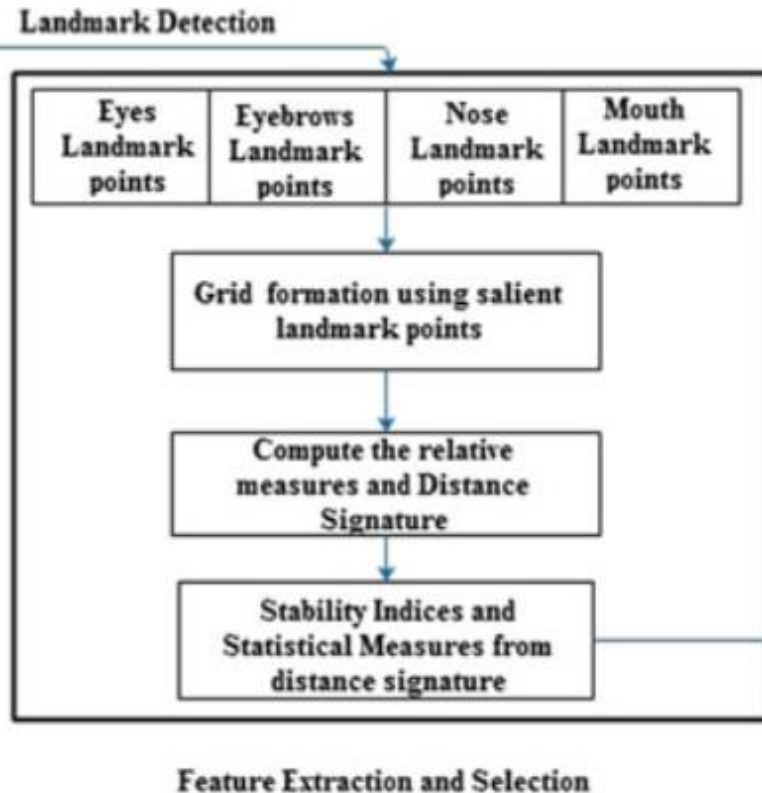
$$\Delta(d_i^{r1})^f = (\mu_i^{r1})^f - (\mu_i^{r1-1})^f \leq \lambda$$

Il valore $r1$ trovato può essere considerato come un indice di stabilità per la *distance signature*.

L'indice di stabilità è associato a $(\mu_i^{r1})^f$.

ACQUISIZIONE NON INVASIVA DELL'AZIONE UMANA

Espressioni facciali – Distance Signature

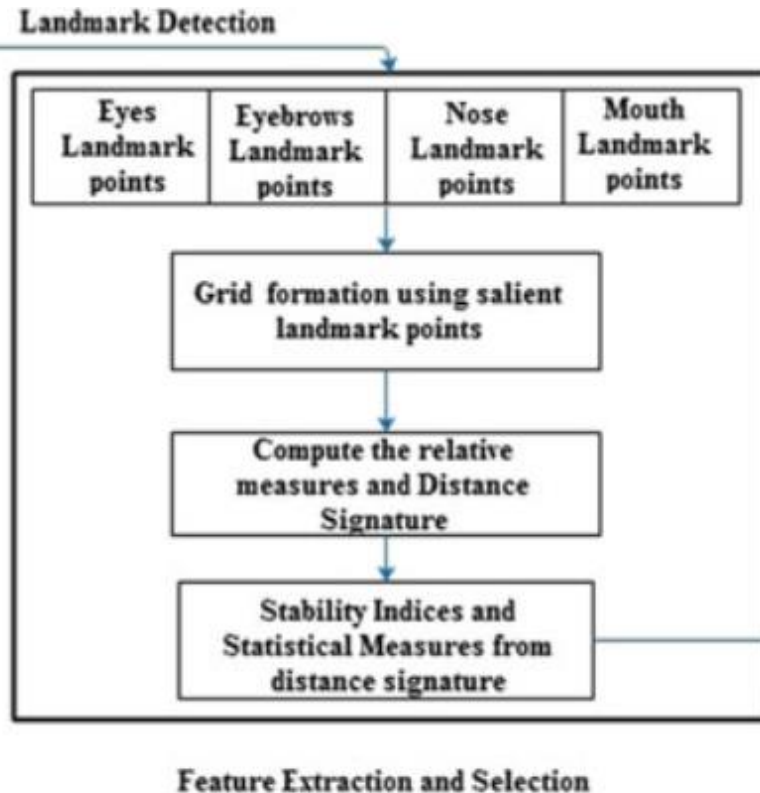


Il calcolo dell'indice di stabilità è quindi caratterizzato dalle seguenti fasi:

- 1) Calcolo delle *normalized distance signatures* a partire dalla griglia delle immagini del volto.
- 2) Calcolo delle *normalized distance signatures* di ordine superiore.
- 3) Selezione di una soglia mediante valutazione degli incrementi di ordine superiore.
- 4) L'ordine superiore associato all'incremento che per la prima volta è minore o uguale alla soglia prestabilita viene considerato come indice di stabilità dell'immagine del volto analizzata.

ACQUISIZIONE NON INVASIVA DELL'AZIONE UMANA

Espressioni facciali – Distance Signature



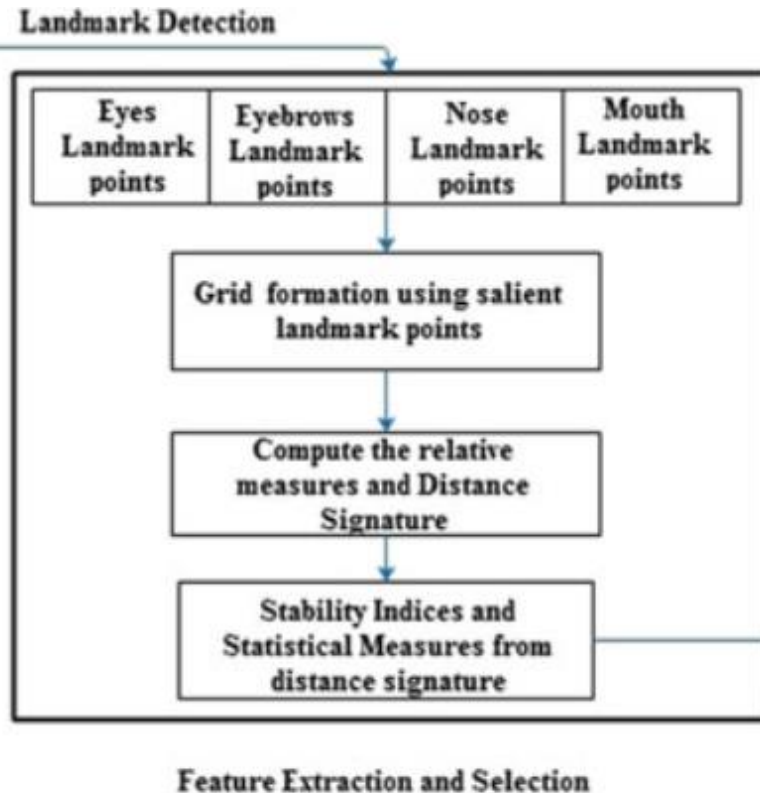
A questo punto, possiamo calcolare alcuni indicatori (misure) statistici associati alla *normalized distance signature*. Le misure statistiche tipicamente considerate sono: range, momento (ordine 1, 2, 3, 4), asimmetria, curtosi, entropia. Tali misure rivelano come sono distribuiti i dati di *distance signature*.

Quindi il *feature vector* calcolato dal blocco di estrazione e selezione delle features (Feature Extraction and Selection) sarà composto da 9 elementi per ogni immagine di ogni classe da riconoscere:

- $(\mu_i^{r1})^f$
- Range
- Momento (ordine 1, 2, 3, 4)
- Asimmetria
- Curtosi
- Entropia

ACQUISIZIONE NON INVASIVA DELL'AZIONE UMANA

Espressioni facciali – Distance Signature



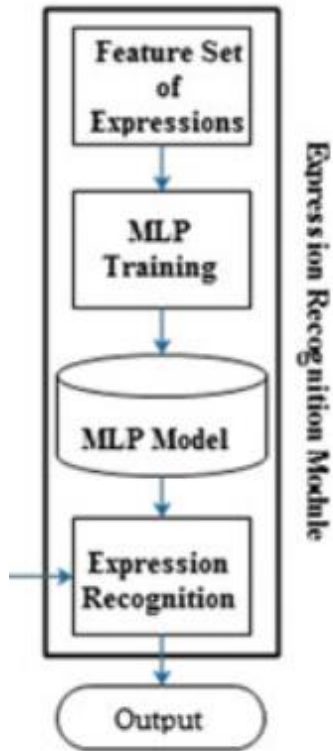
Range: in statistica, il range è definito come la differenza tra il massimo e il minimo valore delle osservazioni. Il range dovrebbe suggerire quanto siano diversificati i valori e, calcolando la differenza tra il valore massimo e minimo, si ottiene una stima della diversificazione dei dati. Ad esempio, supponiamo che un esperimento preveda di scoprire il peso di alcuni oggetti della stessa tipologia e che i valori in grammi siano 320, 367, 423, 471 e 480. In questo caso, il range è semplicemente calcolato come $(480 - 320) = 160$ grammi.

ACQUISIZIONE NON INVASIVA DELL'AZIONE UMANA

Espressioni facciali – Distance Signature

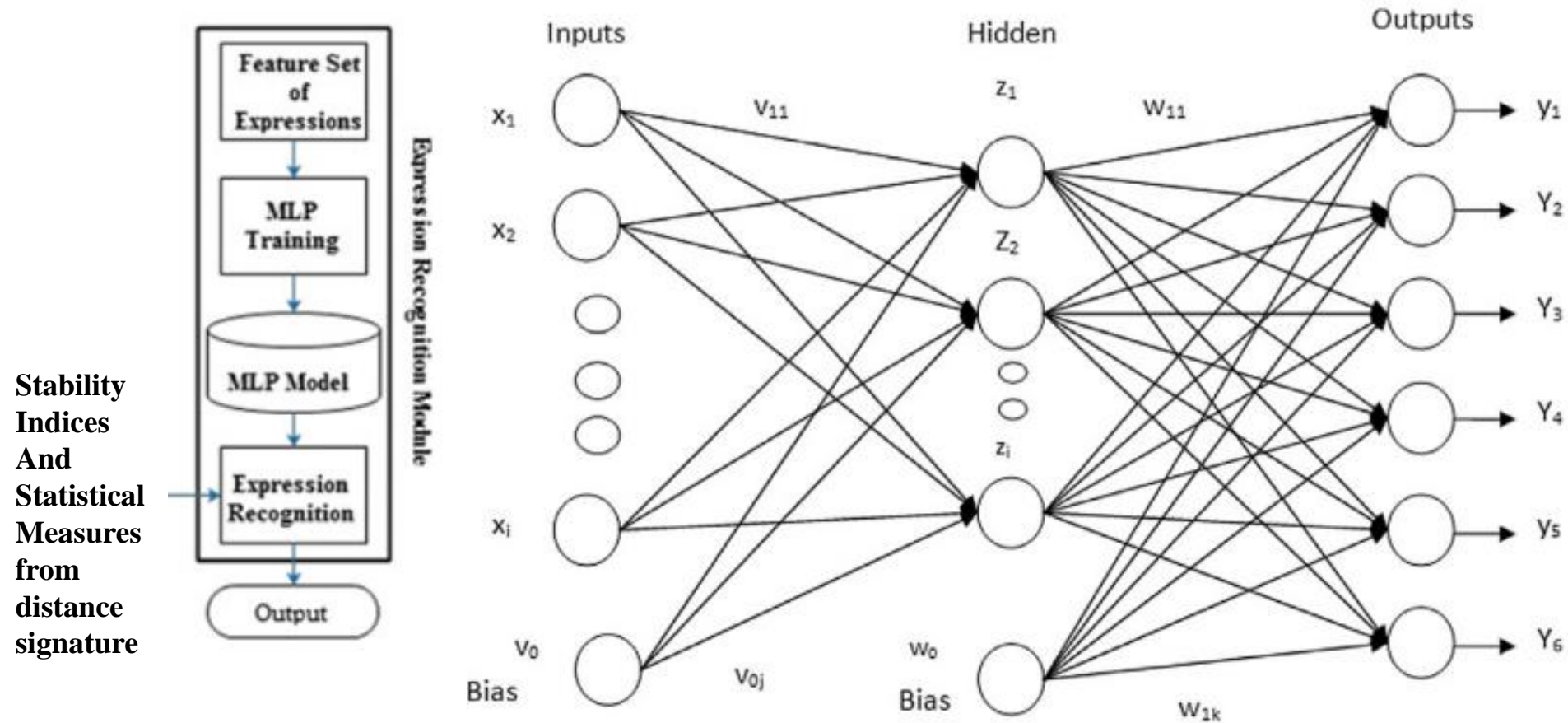
Una rete neurale artificiale di tipo MLP (Multilayer Perceptron) può essere utilizzata per il riconoscimento delle emozioni a partire da espressioni facciali.

La rete neurale MLP viene addestrata (fase di training) su determinati insiemi di features associati a espressioni facciali (nella fase di training le espressioni facciali delle immagini utilizzate sono note). Siamo quindi nel caso di classificazione supervisionata. Grazie alla fase di training, si avrà a disposizione una rete neurale MLP addestrata (MLP Model nella figura). Tale rete neurale addestrata è utilizzata per il riconoscimento delle espressioni facciali (Expression Recognition nella figura). Essa riceve in ingresso il *feature vector* calcolato in precedenza e restituisce come output l'espressione facciale che ha riconosciuto (non è detto ovviamente che il risultato della classificazione sia corretto).



ACQUISIZIONE NON INVASIVA DELL'AZIONE UMANA

Espressioni facciali – Distance Signature



Esempio di architettura della rete neurale artificiale MLP.

ACQUISIZIONE NON INVASIVA DELL'AZIONE UMANA

Espressioni facciali – Distance Signature

Il metodo mostrato è stato sperimentato considerando immagini appartenenti a diversi database, come ad esempio il database Cohn-Kanade (CK+). Sono state selezionate 849 immagini (566 per la fase di training e 283 per la fase di test), ottenendo un tasso di riconoscimento medio uguale al 98.6% nella fase di training. Il tasso di riconoscimento medio può essere calcolato come la media aritmetica tra i tassi di riconoscimento associati alle singole emozioni che si vogliono classificare. Il tasso di riconoscimento è definito come il rapporto tra il numero di classificazioni corrette e il numero totale di classificazioni eseguite.

Per quanto riguarda la fase di test, si riporta nella slide seguente la matrice di confusione.

ACQUISIZIONE NON INVASIVA DELL'AZIONE UMANA

Espressioni facciali – Distance Signature

Classe predetta

Classe reale

	Anger	Disgust	Fear	Happiness	Sadness	Surprise
Anger	49	0	0	0	0	2
Disgust	1	45	0	1	0	0
Fear	0	0	23	1	0	0
Happiness	1	0	0	63	0	0
Sadness	0	0	0	0	47	2
Surprise	0	1	0	1	0	46

Matrice di confusione (CK+ database, fase di test, distance signature).

L'emozione rabbia è stata classificata correttamente in 49 immagini, ma 2 immagini sono state classificate in modo errato con l'emozione sorpresa. L'emozione disgusto è stata classificata correttamente in 45 immagini e 2 immagini sono state confuse con le emozioni rabbia e gioia. L'emozione paura è stata classificata correttamente in 23 immagini e 1 immagine è stata confusa con l'emozione gioia. L'emozione gioia è stata classificata correttamente in 63 immagini, mentre in 1 immagine è stata erroneamente classificata con l'emozione rabbia. L'emozione tristezza è stata classificata correttamente in 47 immagini e 2 immagini sono state confuse con l'emozione sorpresa. L'emozione sorpresa è stata classificata correttamente in 46 immagini e 2 immagini sono state confuse con le emozioni disgusto e gioia.

ACQUISIZIONE NON INVASIVA DELL'AZIONE UMANA

Espressioni facciali – Distance Signature

Classe predetta

Classe reale

	Anger	Disgust	Fear	Happiness	Sadness	Surprise
Anger	49	0	0	0	0	2
Disgust	1	45	0	1	0	0
Fear	0	0	23	1	0	0
Happiness	1	0	0	63	0	0
Sadness	0	0	0	0	47	2
Surprise	0	1	0	1	0	46

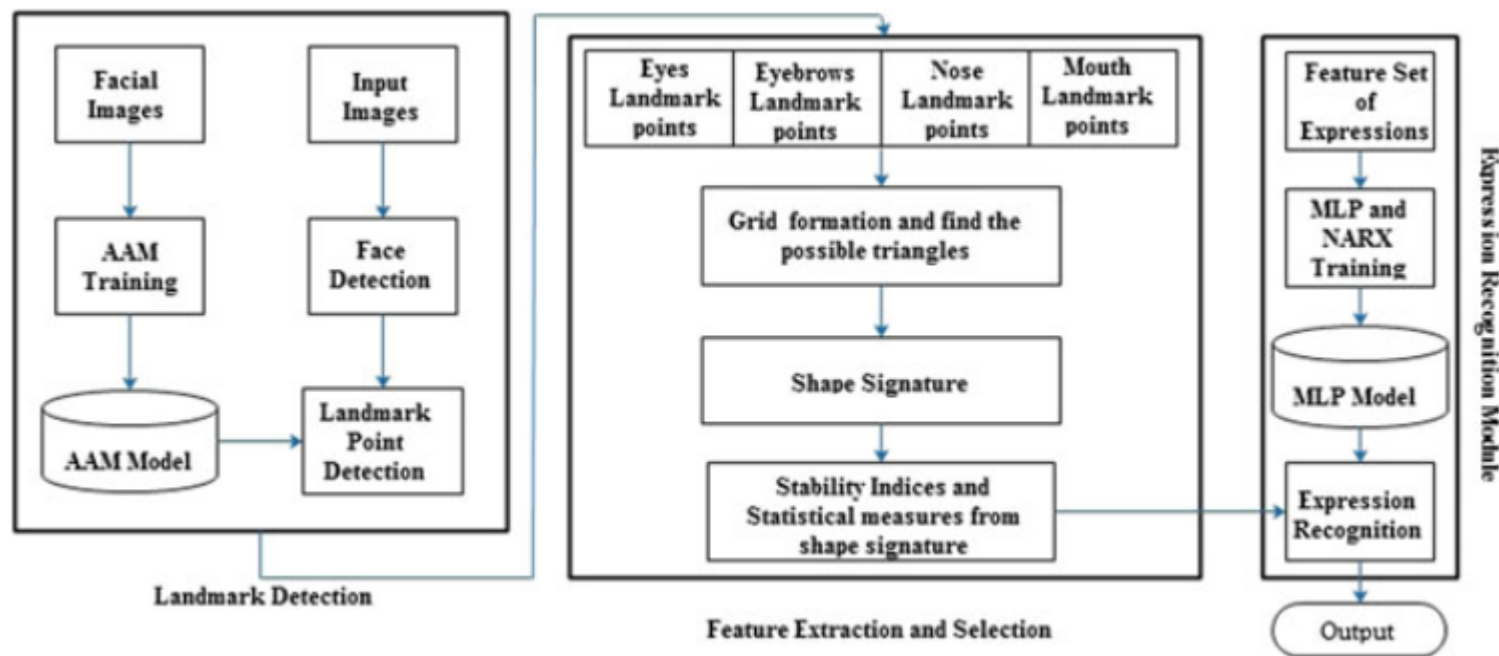
Matrice di confusione (CK+ database, fase di test, distance signature).

	Database	An	Di	Fe	Ha	Sa	Su	Avg.
Distance Signature	CK+	96.1	95.7	95.8	98.4	95.9	95.8	96.3

Tasso di riconoscimento (CK+ database, fase di test, distance signature).

ACQUISIZIONE NON INVASIVA DELL'AZIONE UMANA

Espressioni facciali – Shape Signature



Schema a blocchi dell'approccio basato sulla shape signature.

Ci sono alcune differenze tra tale approccio e l'approccio basato sul concetto di *distance signature*.

I due approcci possono essere utilizzati insieme.

ACQUISIZIONE NON INVASIVA DELL'AZIONE UMANA

Espressioni facciali – Shape Signature

Nell'approccio basato sul concetto di *shape signature*, i punti di riferimento principali, identificati nelle regioni associate alle sopracciglia, agli occhi, alla bocca e al naso, vengono utilizzati per la creazione di triangoli nella griglia di punti ottenuta.



Triangoli ottenuti su tre immagini del database CK+.

ACQUISIZIONE NON INVASIVA DELL'AZIONE UMANA

Espressioni facciali – Shape Signature

Si ricorda che per un generico triangolo i cui lati sono x, y, z :

$$p = x + y + z$$

Perimetro

$$h = p/2$$

Semiperimetro

$$a = \sqrt{h(h-x)(h-y)(h-z)}$$

Area

La *shape signature* viene definita come

$$s_i = \frac{a_i}{p_i^2}$$

dove i rappresenta il triangolo considerato nella griglia ottenuta. Si calcolano quindi le *normalized shape signatures*

$$\mu_i^f = \frac{s_i^f}{\sum_{i=1}^n s_i^f} \quad f = 1, 2, \dots, m \text{ and } i = 1, 2, \dots, n$$

Diagram illustrating the calculation of normalized shape signatures:

- μ_i^f is labeled as *shape signatures*.
- f is labeled as *numero totale di immagini* (total number of images).
- n is labeled as *numero totale di forme* (total number of shapes).

Riferimenti Bibliografici

- [1] Kraiss, K. -F. (2006). Advanced Man-Machine Interaction: Fundamentals and Implementation. Springer-Verlag Berlin Heidelberg. ISBN-10: 3-540-30618-8
- [2] Paramartha Dutta, Asit Barman (2020). Human Emotion Recognition from Face Images. Springer Singapore. ISBN: 978-981-15-3883-4