

# RICONOSCIMENTO DEL LINGUAGGIO DEI SEGNI

## *Riconoscimento di segni mediante features non associate alle mani*

Poiché i linguaggi dei segni sono linguaggi multimodali, differenti canali vengono utilizzati contemporaneamente per trasferire informazione. Una distinzione di base è quella tra i canali associati alle mani e ai gesti e i canali non associati alle mani/associati alla faccia (con i rispettivi parametri).

I parametri non associati alle mani sono indispensabili nel linguaggio dei segni. Essi codificano alcuni aggettivi e contribuiscono alla grammatica. Alcuni segni sono identici riguardo ai gesti e possono essere distinti solo facendo riferimento a parametri non associati alle mani. Questo è ad esempio il caso dei segni «not» e «to» nel German Sign Language (GSL); tali segni possono essere distinti solo facendo riferimento al movimento della testa. Un esempio simile nel British Sign Language (BSL) è rappresentato dai segni «now» e «offer»: essi richiedono di osservare il profilo delle labbra per essere distinti (si vedano le figure delle slide seguenti).

# RICONOSCIMENTO DEL LINGUAGGIO DEI SEGNI

## *Riconoscimento di segni mediante features non associate alle mani*

**A**



**B**



*Nel German Sign Language i segni «not» (A) e «to» (B) sono identici riguardo ai gesti delle mani ma sono diversi riguardo al movimento della testa.*

# RICONOSCIMENTO DEL LINGUAGGIO DEI SEGNI

## *Riconoscimento di segni mediante features non associate alle mani*

**A**



**B**



*Nel British Sign Language i segni «now» (A) e «offer» (B) sono identici riguardo ai gesti delle mani ma sono diversi riguardo al profilo delle labbra.*

# RICONOSCIMENTO DEL LINGUAGGIO DEI SEGNI

## *Riconoscimento di segni mediante features non associate alle mani*

Di seguito vengono riportati i principali parametri non associati alle mani.

### *Postura della parte superiore del corpo*

Il busto costituisce generalmente un riferimento nello *spazio dei segni*. Le distanze spaziali e gli elementi del contesto possono essere comunicati mediante la postura del busto. Ad esempio i segni di «rifiuto» e di «attrazione» mostrano una leggera inclinazione del busto verso la parte posteriore e in avanti.

### *Posizione della testa*

La posizione della testa supporta la semantica del linguaggio dei segni. Domande, affermazioni, negazioni e proposizioni condizionali sono ad esempio espresse con l'aiuto della posizione della testa. Inoltre, la posizione della testa può permettere di codificare informazioni sul tempo. Ad esempio, alcuni segni che vengono eseguiti in un breve periodo di tempo sono caratterizzati da un cambiamento minimo della posizione della testa mentre altri segni eseguiti in un periodo di tempo più lungo sono caratterizzati da una evidente rotazione della testa in direzione opposta al gesto.

# RICONOSCIMENTO DEL LINGUAGGIO DEI SEGNI

## *Riconoscimento di segni mediante features non associate alle mani*

### *Direzione dello sguardo*

Due persone sorde che comunicano stabiliscono solitamente uno stretto contatto visivo. Tuttavia, un breve cambiamento della direzione dello sguardo può essere utilizzato come riferimento per il significato spaziale di un gesto. Allo stesso modo, la direzione dello sguardo può essere usata in combinazione con la postura del busto.

### *Espressione facciale*

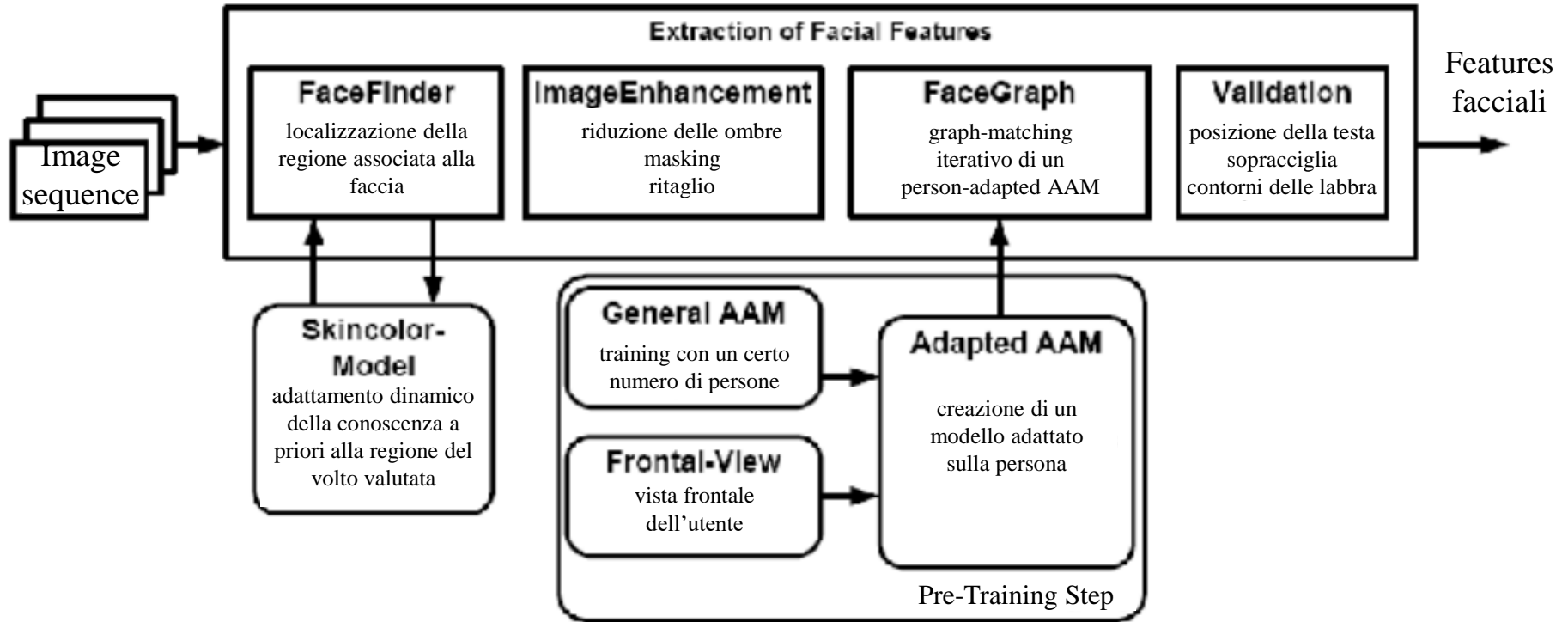
Le espressioni facciali servono essenzialmente alla trasmissione di sentimenti (mimica lessicale). Inoltre, possono essere trasmessi anche elementi grammaticali. Un cambio di posizione della testa combinato con il sollevamento delle sopracciglia corrisponde, ad esempio, a un congiuntivo.

### *Profilo delle labbra*

Il profilo delle labbra rappresenta la caratteristica più evidente non associata alle mani. Il profilo delle labbra risolve le ambiguità tra i segni («brother» vs «sister») e specifica le espressioni («meat» vs «hamburger»). Esso fornisce inoltre informazioni ridondanti rispetto ai gesti per distinguere segni simili.

# RICONOSCIMENTO DEL LINGUAGGIO DEI SEGNI

## *Riconoscimento di segni mediante features non associate alle mani*

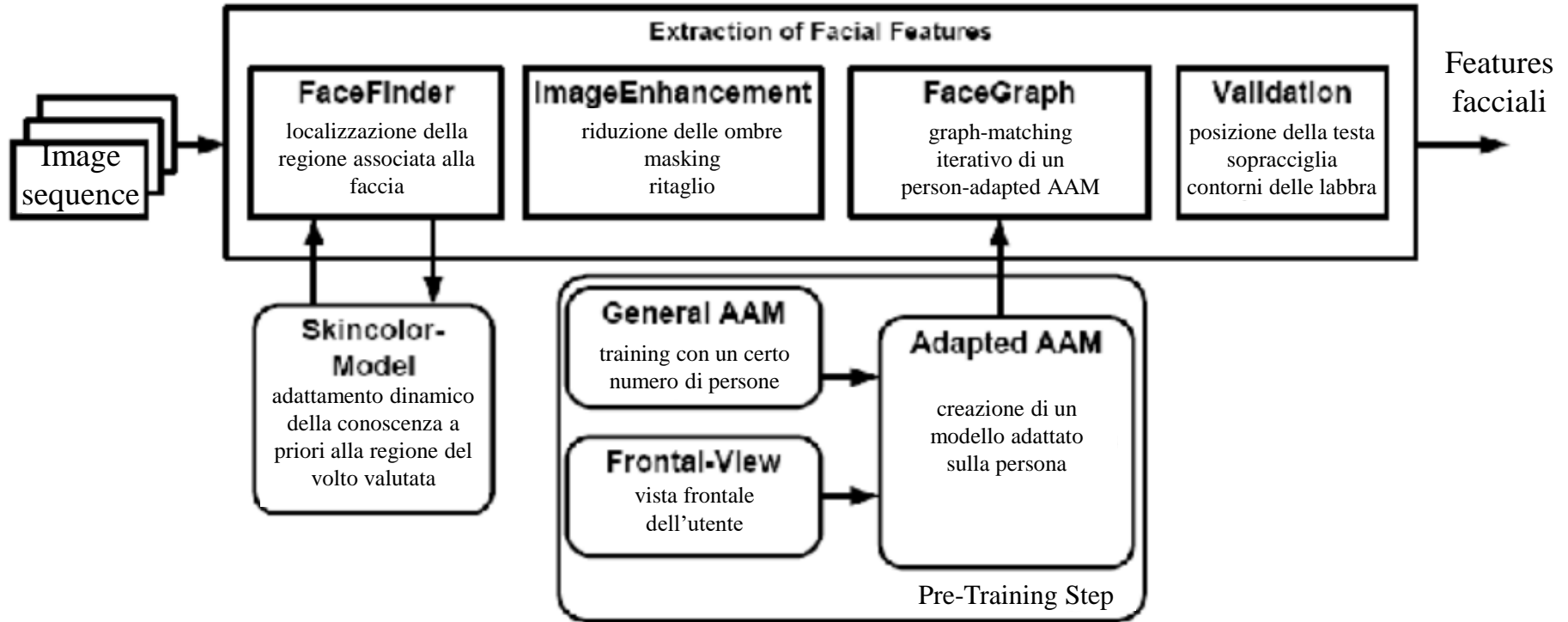


*Procedura di elaborazione per l'estrazione di features non associate alle mani.*

Dopo l'acquisizione delle immagini dalla camera, viene incluso un Face Finder in modo da localizzare la regione della faccia del segnante nell'immagine. In seguito, tale regione viene ritagliata e scalata.

# RICONOSCIMENTO DEL LINGUAGGIO DEI SEGNI

## *Riconoscimento di segni mediante features non associate alle mani*



*Procedura di elaborazione per l'estrazione di features non associate alle mani.*

La fase successiva è il fitting di un grafo associato alla faccia (adattato all'utente) mediante *Active Appearance Models* (AAMs). Tale grafo è utile per l'estrazione di parametri associati al profilo delle labbra, agli occhi e alle sopracciglia (si veda la figura della slide seguente).



# RICONOSCIMENTO DEL LINGUAGGIO DEI SEGNI

## *Riconoscimento di segni mediante features non associate alle mani*

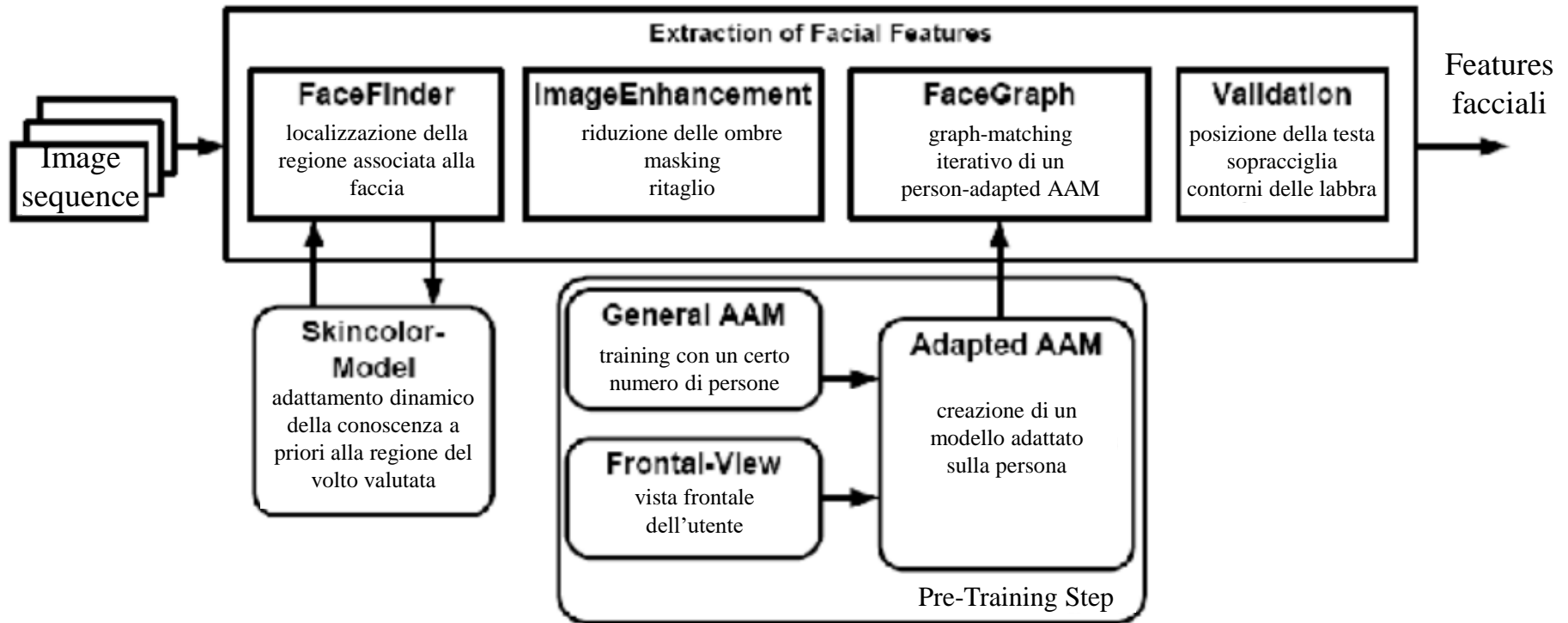


*Procedura di elaborazione per il ritaglio della regione associata alla faccia e per il matching di un grafo associato alla faccia.*



# RICONOSCIMENTO DEL LINGUAGGIO DEI SEGNI

## *Riconoscimento di segni mediante features non associate alle mani*



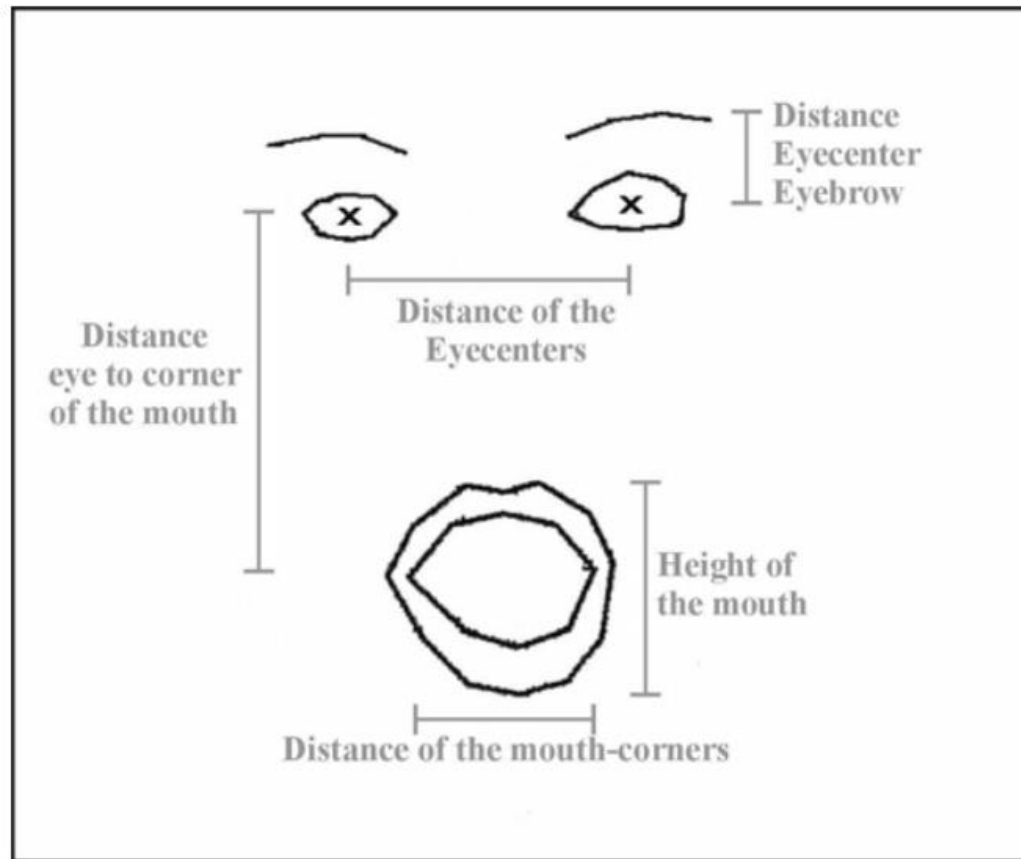
*Procedura di elaborazione per l'estrazione di features non associate alle mani.*

Le features non associate alle mani possono essere divise in tre gruppi (si veda la figura nella slide seguente). Il primo gruppo riguarda il profilo delle labbra, che viene descritto mediante larghezza, altezza e features relative alla forma come i momenti invarianti, l'eccentricità e l'orientamento (definiti per il riconoscimento dei gesti). Le distanze tra gli occhi e gli angoli della bocca sono

# RICONOSCIMENTO DEL LINGUAGGIO DEI SEGNI

## *Riconoscimento di segni mediante features non associate alle mani*

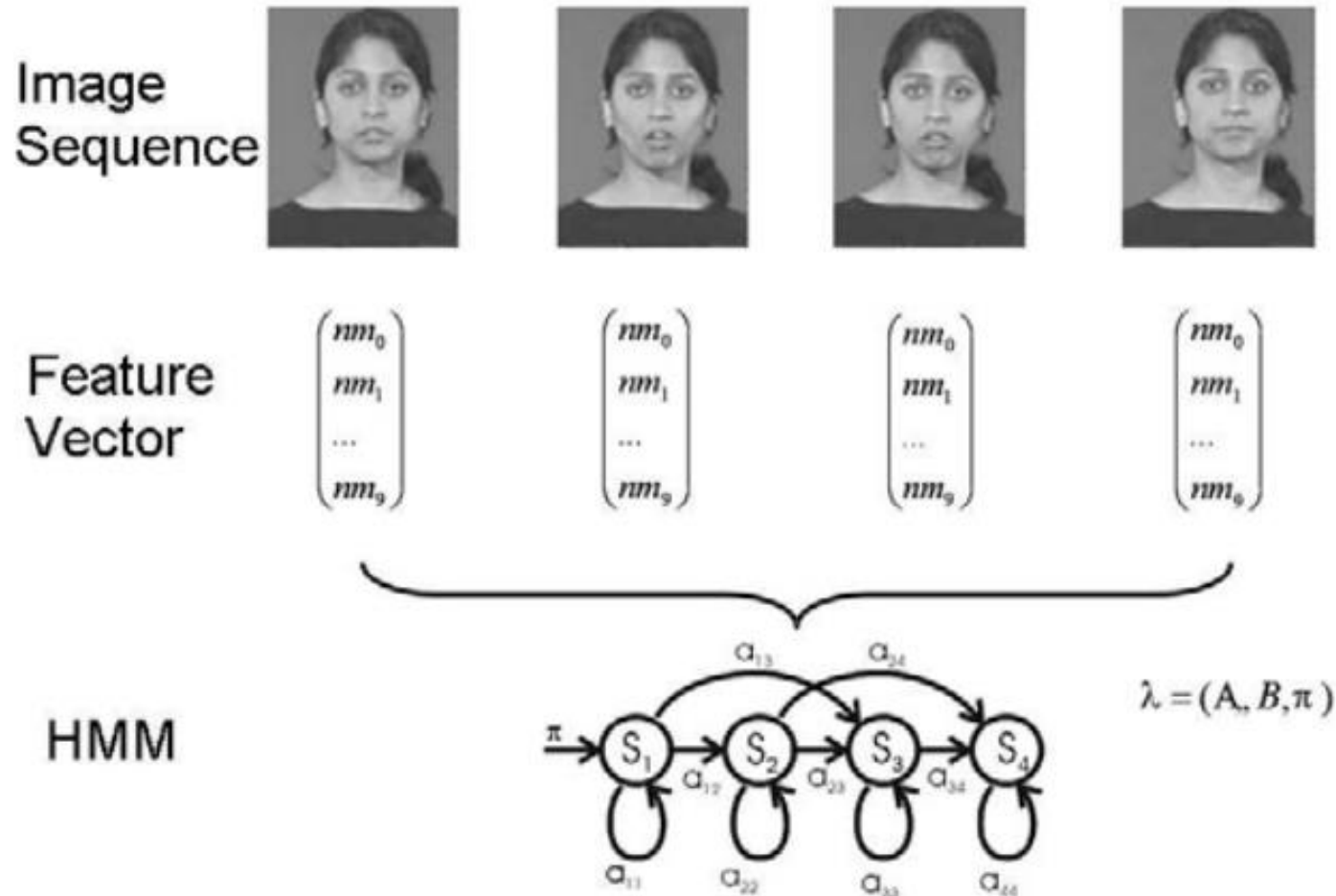
nel secondo gruppo, mentre il terzo gruppo contiene le distanze tra gli occhi e le sopracciglia. Per ogni immagine della sequenza, i parametri estratti vengono inclusi in un *feature vector*, il quale viene utilizzato nella classificazione (si veda la figura nella slide seguente).



*Rappresentazione dei parametri non associati alle mani.*

# RICONOSCIMENTO DEL LINGUAGGIO DEI SEGNI

## *Riconoscimento di segni mediante features non associate alle mani*



*Schema dell'estrazione e della classificazione di features non associate alle mani.*

# RICONOSCIMENTO DEL LINGUAGGIO DEI SEGNI

## *Riconoscimento di segni mediante features non associate alle mani*

In caso di sovrapposizione parziale di mani e faccia, i modelli AAMs non riescono a fornire un fitting accurato. Possono essere adottate alcune soluzioni per questo problema.



*Nella sovrapposizione di mani e faccia, alcune regioni della faccia vengono analizzate separatamente. Se ad esempio un occhio e la bocca sono nascosti (sinistra), non si considera alcuna feature della faccia. Tuttavia, se gli occhi e la bocca sono localizzati, le features calcolate possono essere utilizzate per la classificazione (destra).*

# RICONOSCIMENTO DEL LINGUAGGIO DEI SEGNI

## *Riconoscimento di frasi nel linguaggio dei segni*

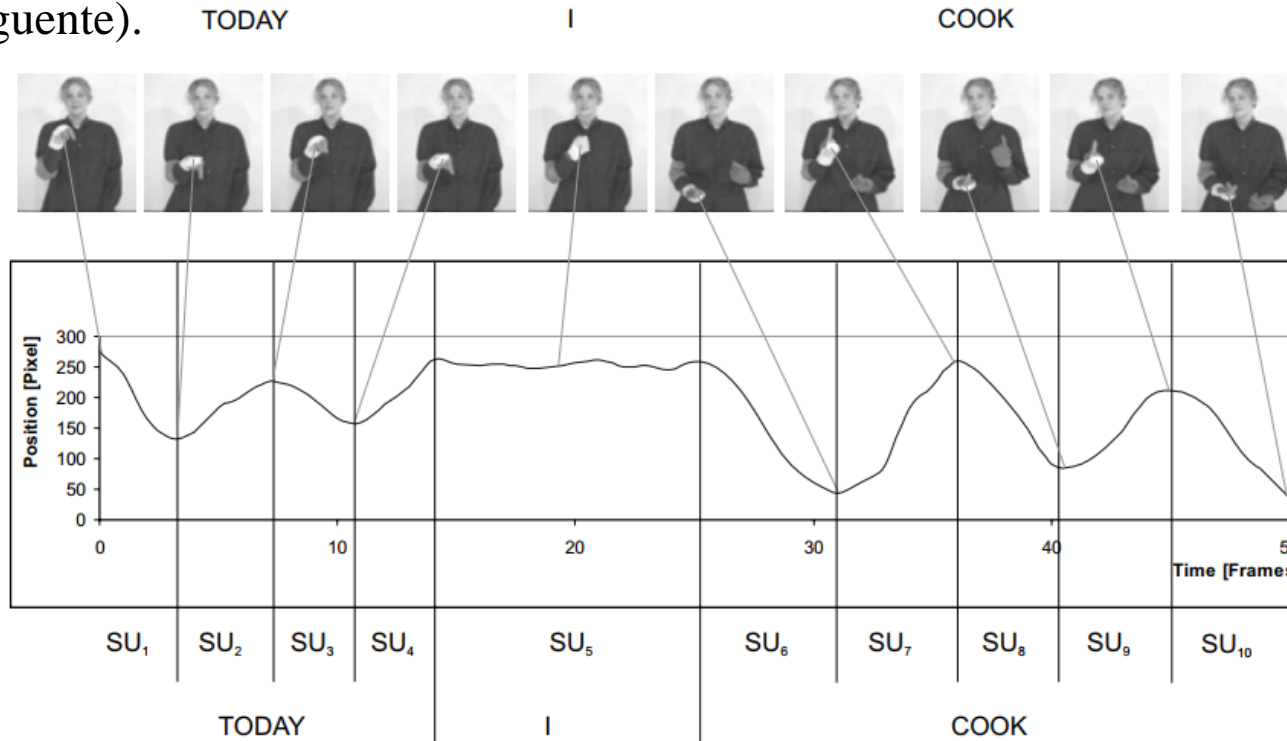
Una problematica tipica nel riconoscimento del linguaggio dei segni è quella di individuare quale parte di una frase di segni rappresenti un solido modello di base. La miglior soluzione dal punto di vista teorico è che ad una intera frase del linguaggio dei segni corrisponda un modello. Grazie alla loro lunga durata, le frasi non portano solitamente a una classificazione errata. Tuttavia, il maggior svantaggio è rappresentato dal fatto che ci sono troppe combinazioni per addestrare tutte le possibili frasi nel linguaggio dei segni. Di conseguenza, molti sistemi di riconoscimento del linguaggio dei segni sono basati su modelli associati a parole dove un segno rappresenta un modello nel database dei modelli. Quindi, il riconoscimento di frasi nel linguaggio dei segni è molto flessibile. Tuttavia, anche questo metodo presenta alcuni inconvenienti:

- La complessità della fase di training aumenta con l'aumentare delle dimensioni del vocabolario.
- Un'estensione futura del vocabolario può essere problematica poiché i nuovi segni vengono di solito inseriti nel contesto di altri segni per la fase di training (embedded training).

# RICONOSCIMENTO DEL LINGUAGGIO DEI SEGNI

## *Riconoscimento di frasi nel linguaggio dei segni*

Un approccio alternativo è rappresentato dalla modellizzazione dei segni con *subunits* (SU) di segni. Tale approccio è simile a quello utilizzato per la modellizzazione del parlato mediante fonemi. Le *subunits* sono segmenti di segni che emergono dalla suddivisione dei segni (si veda la figura seguente).



*La frase (linguaggio dei segni) «TODAY I COOK (HEUTE ICH KOCHEN)» nel German Sign Language (GSL). Nella parte superiore si nota la sequenza dei segni acquisita. Al centro si nota la posizione verticale della mano destra del segnante. Nella parte inferiore sono evidenziate le subunits. Per una migliore visualizzazione, il segnante indossa guanti di cotone colorati.*



# RICONOSCIMENTO DEL LINGUAGGIO DEI SEGNI

## *Riconoscimento di frasi nel linguaggio dei segni*

Le *subunits* devono rappresentare una quantità finita di modelli. Il loro numero dovrebbe essere scelto in modo tale che ogni segno possa essere composto con *subunits*. I vantaggi di tale approccio sono:

- Il periodo di training termina quando tutti i modelli associati alle varie *subunits* hanno eseguito la propria fase di training. Un'estensione del vocabolario può essere ottenuta con la composizione di un nuovo segno mediante concatenazione di modelli adeguati di *subunits*.
- La dimensione del vocabolario generale può essere estesa.
- Il riconoscimento sarà più robusto (in base alla dimensione del vocabolario).

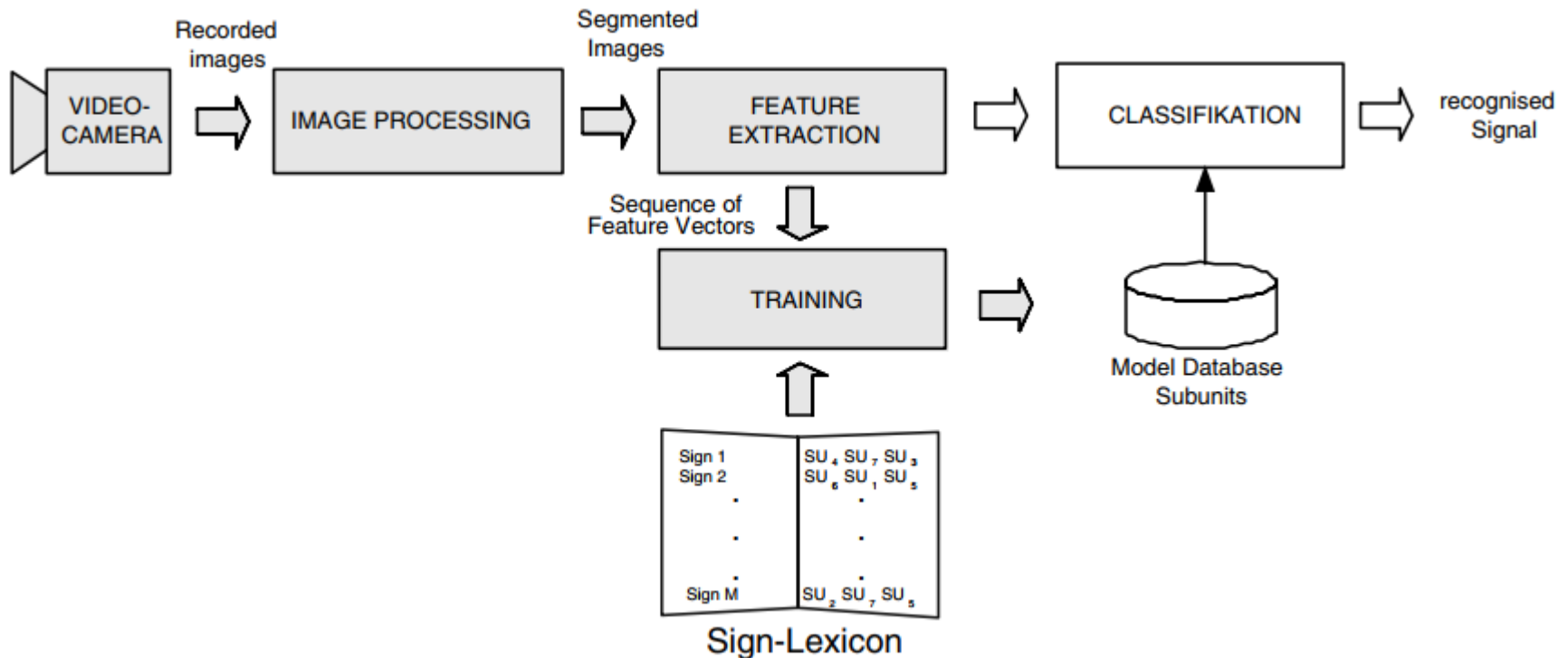
Le fasi di training e classificazione precedentemente descritte necessitano di alcune modifiche al fine di poter essere applicate al riconoscimento mediante *subunits*. Un sistema di riconoscimento basato su *subunits* necessita di un'ulteriore fonte di conoscenza, nella quale deve essere dettagliata la codifica (detta anche trascrizione) del segno considerato in *subunits*. Entrambi i processi di training e classificazione sono basati su un lessico dei segni (sign-lexicon).



# RICONOSCIMENTO DEL LINGUAGGIO DEI SEGNI

## *Riconoscimento di frasi nel linguaggio dei segni*

Lo scopo del processo di training è la stima dei parametri del modello di ogni *subunit*. L'esempio riportato nella figura seguente mostra che il segno 1 è composto dalle *subunits* (SU)  $SU_4$ ,  $SU_7$  e  $SU_3$ . I parametri dei modelli associati a queste *subunits* vengono calcolati mediante una fase di training basata sui dati acquisiti (segno 1). Un esempio di algoritmo che si può utilizzare è l'algoritmo di Viterbi.

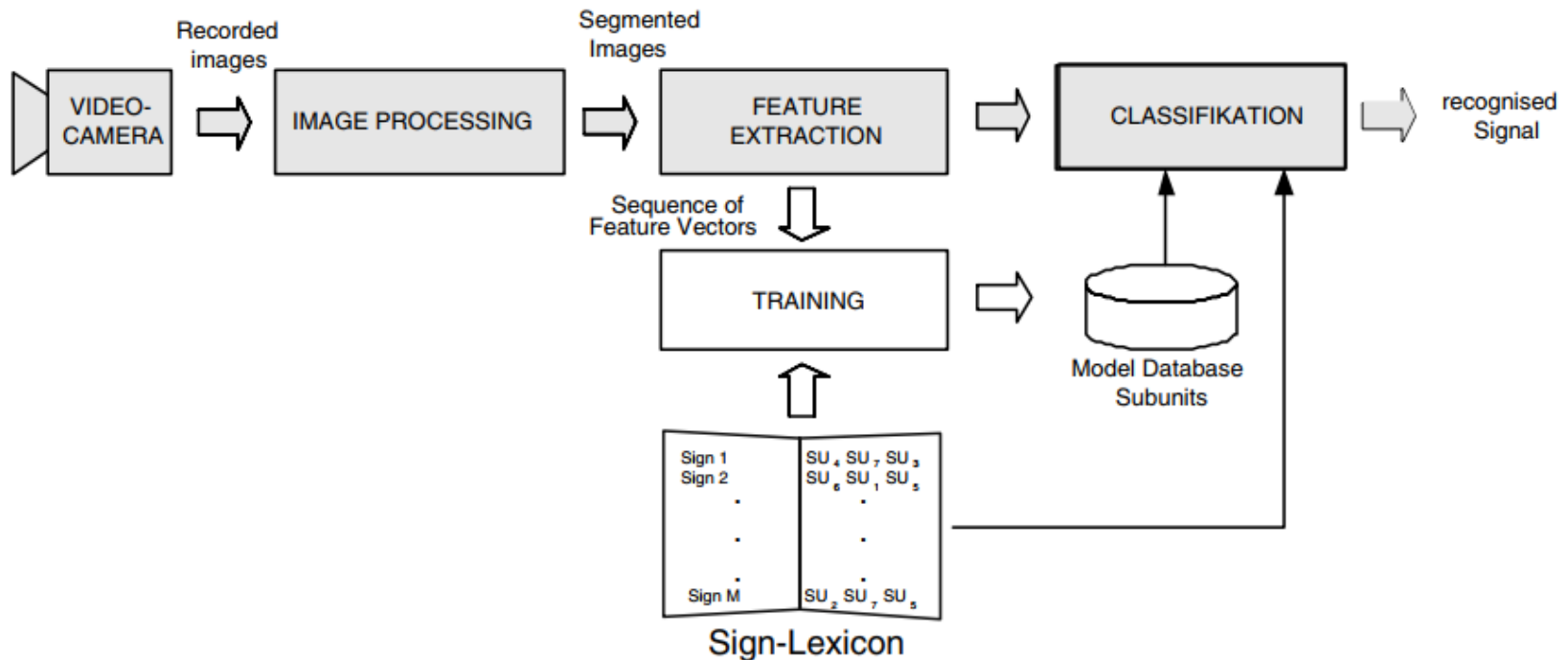


*Componenti del processo di training per l'identificazione delle subunits.*

# RICONOSCIMENTO DEL LINGUAGGIO DEI SEGNI

## *Riconoscimento di frasi nel linguaggio dei segni*

Dopo il completamento della fase di training, viene riempito un database con tutti i modelli delle *subunits*; tale database rappresenta la base per il processo di classificazione. Tuttavia, lo scopo della classificazione non è il riconoscimento delle *subunits* ma è il riconoscimento dei segni. Quindi, anche in questo caso è necessario conoscere la caratterizzazione di ogni segno in termini di *subunits* (da quali *subunits* è composto questo segno?). Tale informazione è contenuta nel lessico dei segni.



*Componenti del processo di classificazione basato su subunits.*

# RICONOSCIMENTO DEL LINGUAGGIO DEI SEGNI

## *Riconoscimento di frasi nel linguaggio dei segni*

Finora si è ipotizzato che sia disponibile un lessico dei segni, cioè che sia nota la composizione dei segni in termini di *subunits*. Tuttavia, tale ipotesi potrebbe non essere verificata. Quindi la suddivisione di un segno in *subunits* opportune rappresenta una problematica da risolvere. Inoltre, occorre determinare la semantica delle *subunits*. In tale contesto, due possibili approcci per la trascrizione del linguaggio dei segni sono: approccio orientato alla linguistica e approccio orientato alla visione.

### *Approccio orientato alla linguistica*

Le *subunits* utilizzate per il riconoscimento del parlato sono perlopiù motivate dal punto di vista linguistico e sono tipicamente sillabe o fonemi. La base di questa suddivisione del parlato è simile al sistema di rappresentazione del parlato: un testo scritto con l'ortografia corrispondente è il sistema di rappresentazione standard del parlato. Esistono moltissimi elementi nel lessico del parlato: essi consistono nella trascrizione del parlato in *subunits*. Tali elementi possono rappresentare la base per il riconoscimento del parlato. Trasferendo questo concetto al riconoscimento del linguaggio dei segni, ci si trova di fronte a una varietà di opzioni per la rappresentazione e inoltre c'è una minore standardizzazione rispetto al parlato.

# RICONOSCIMENTO DEL LINGUAGGIO DEI SEGNI

## *Riconoscimento di frasi nel linguaggio dei segni*

Le *subunits* del linguaggio dei segni vengono definite cheremi (corrispondono ai fonemi del parlato). Nel seguito vengono riportati due sistemi di rappresentazione.

### *Notazione di Stokoe*

Nella notazione di Stokoe sono definiti tre tipi di cheremi. Il primo tipo descrive la configurazione della forma della mano ed è chiamato *dez* (designator). Il secondo tipo è *sig* (signation) e descrive il tipo di movimento del segno eseguito. Il terzo tipo è la posizione del segno eseguito e viene chiamato *tab* (tabula). Stokoe sviluppò un lessico per l'American Sign Language (ASL) mediante i cheremi appena descritti. Il lessico è formato da circa 2500 input, dove i segni sono codificati in 55 differenti cheremi (12 *tab*, 19 *dez*, 24 *sig*). Un esempio di segno codificato nella notazione di Stokoe è riportato nella figura seguente.



3<sup>1</sup>

*Sinistra: il segno THREE in ASL. Destra: rappresentazione del segno nella notazione di Stokoe.*

# RICONOSCIMENTO DEL LINGUAGGIO DEI SEGNI

## *Riconoscimento di frasi nel linguaggio dei segni*

I cheremi impiegati sono difficili da utilizzare nelle applicazioni pratiche di riconoscimento. Non tutti i segni sono inclusi nel lessico della notazione di Stokoe. Inoltre molti cheremi della notazione di Stokoe sono eseguiti in parallelo, mentre un sistema di riconoscimento considera *subunits* in sequenza. Inoltre, nessuno dei cheremi *sig* (i quali codificano il movimento del segno eseguito) sono necessari per un sistema di riconoscimento, poiché i movimenti vengono modellizzati mediante *Hidden Markov Models* (*HMMs*). Quindi, il lessico di Stokoe, senza modifiche manuali, non può essere utilizzato come base per un sistema di riconoscimento.

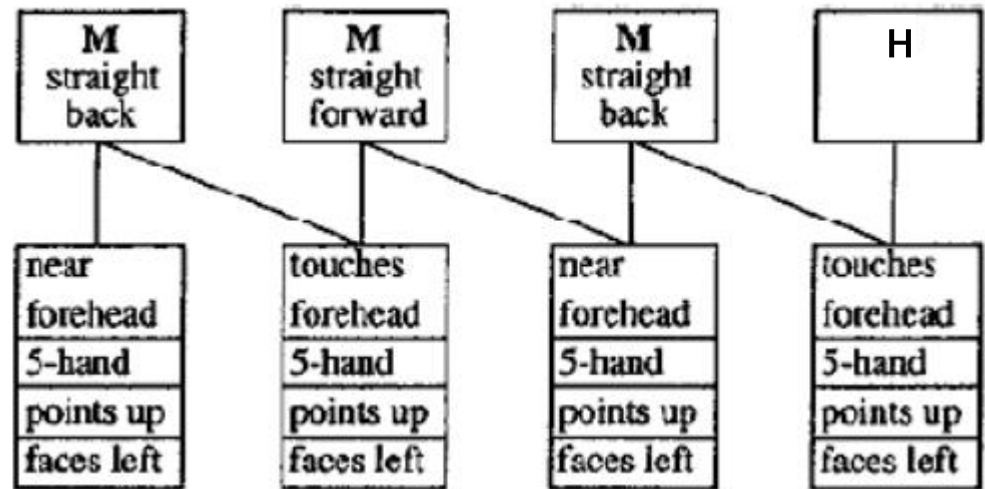
### *Notazione di Liddell e Johnson*

Un'altra notazione fu sviluppata da Liddell e Johnson. Essi hanno suddiviso i segni in cheremi mediante il Movement-Hold-Model. In tale modello, i segni sono divisi in ordine sequenziale in segmenti. Esistono due tipi di segmenti: i *movements* sono segmenti nei quali la configurazione di una mano è ancora in movimento; gli *hold-segments* sono segmenti dove non si verifica alcun movimento, cioè la configurazione delle mani è fissa. Ogni segno può essere modellizzato come una sequenza di *movements* e *hold-segments*. Inoltre, ogni *hold-segment* è composto da features associate alle articolazioni. Tali features

# RICONOSCIMENTO DEL LINGUAGGIO DEI SEGNI

## *Riconoscimento di frasi nel linguaggio dei segni*

descrivono la forma della mano, la posizione e l'orientamento della mano, i movimenti delle dita, e la rotazione e l'orientamento del polso. La figura seguente illustra un esempio di rappresentazione di un segno mediante *movements* e *hold-segments*.



*Rappresentazione del segno FATHER del ASL mediante il Movement-Hold-Model.*

Mentre la notazione di Stokoe è basata su una struttura perlopiù parallela di segni, tale notazione produce una sequenza di segmenti, la quale è più adeguata per un sistema di riconoscimento. Comunque, come accade per la notazione di Stokoe, anche in tale notazione non esiste un lessico tale da codificare



# RICONOSCIMENTO DEL LINGUAGGIO DEI SEGNI

## *Riconoscimento di frasi nel linguaggio dei segni*

tutti i segni. Inoltre, la codifica dettagliata delle features articolari potrebbe causare altri problemi. Una procedura di estrazione delle features video-based da parte del sistema di riconoscimento potrebbe non essere capace di raggiungere un tale livello di dettaglio. Quindi, la notazione di Liddell e Johnson, senza modifiche manuali o anche trascrizioni manuali, non è adatta ad essere impiegata come lessico dei segni in un sistema di riconoscimento.

### *Approccio orientato alla visione*

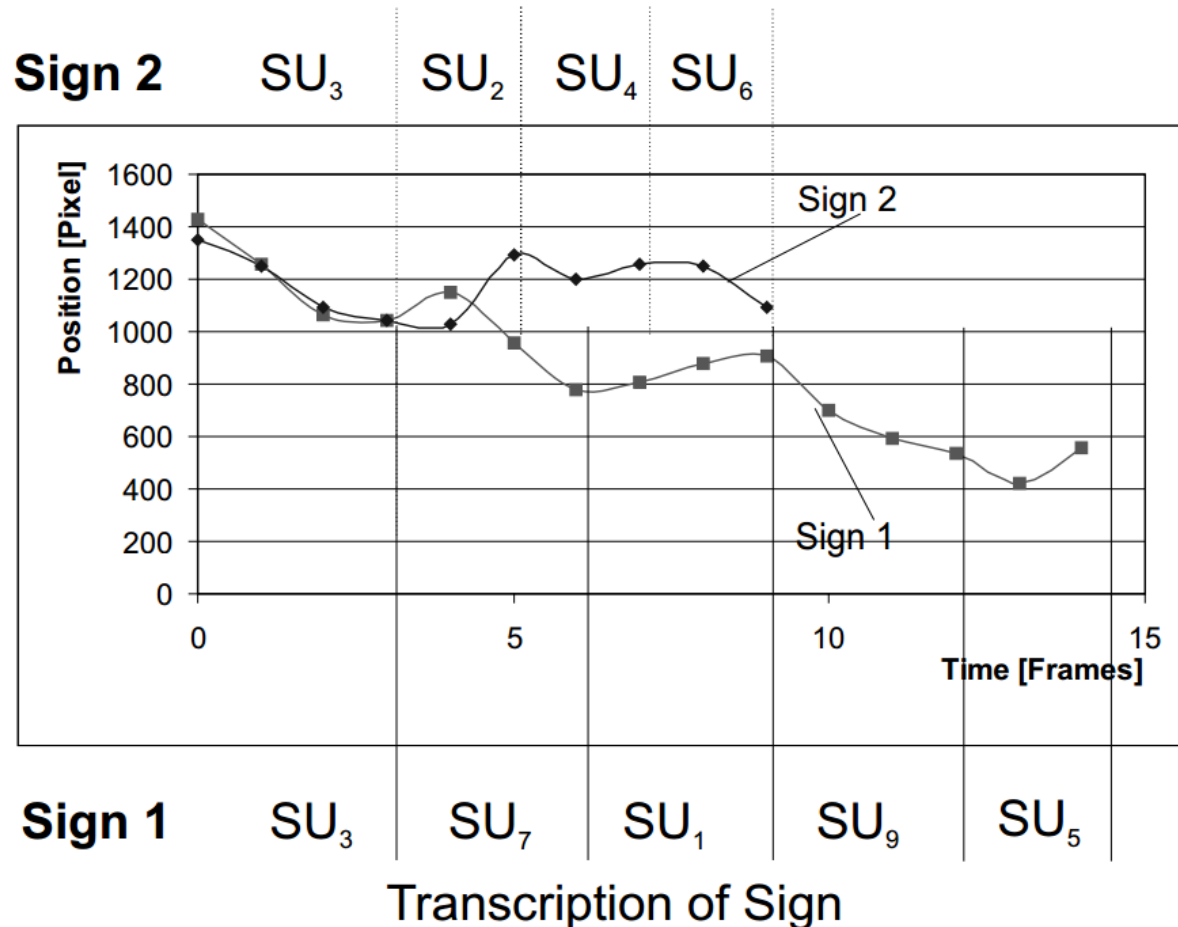
L'approccio orientato alla visione di una notazione o di un sistema di trascrizione per il riconoscimento del linguaggio dei segni non si basa su alcuna conoscenza linguistica relativa al linguaggio dei segni. In tale approccio, i segni vengono suddivisi in *subunits* mediante un processo data-driven. In una prima fase, ogni segno del vocabolario viene diviso in maniera sequenziale in differenti segmenti, i quali non hanno significato semantico. Una fase successiva determina le similarità tra i segmenti identificati. Segmenti simili vengono quindi raggruppati ed etichettati (labeled). Essi vengono considerati come un'unica *subunit*. Questo processo non richiede altre fonti di conoscenza se non i dati stessi. Ogni segno può essere quindi descritto come una sequenza delle *subunits* contenute, le quali



# RICONOSCIMENTO DEL LINGUAGGIO DEI SEGNI

## *Riconoscimento di frasi nel linguaggio dei segni*

vengono distinte mediante le loro etichette (labels). La figura seguente mostra ad esempio la progressione temporale orizzontale (mano destra) di due segni differenti.

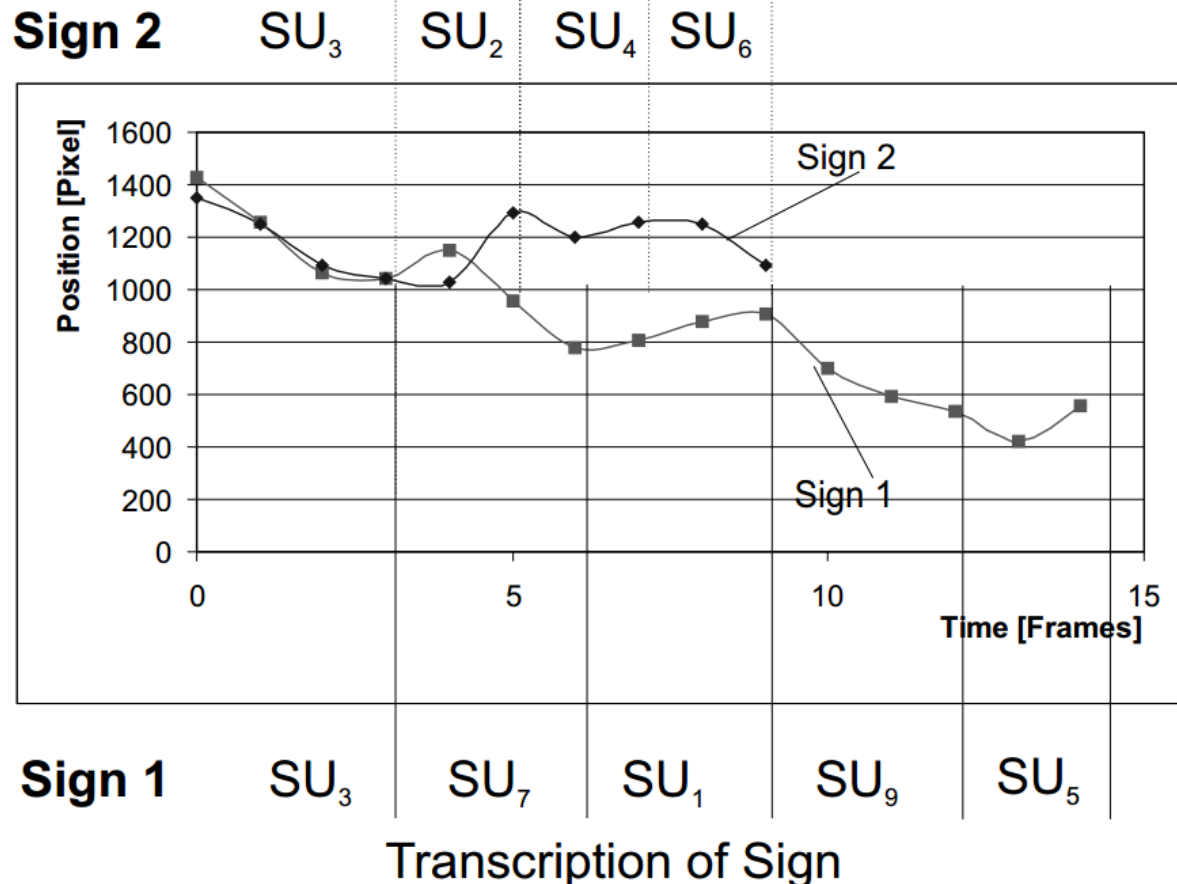


*Esempio di differenti trascrizioni di due segni.*

# RICONOSCIMENTO DEL LINGUAGGIO DEI SEGNI

## *Riconoscimento di frasi nel linguaggio dei segni*

I segni eseguiti sono simili all'inizio. Di conseguenza, all'inizio entrambi i segni vengono assegnati alla stessa *subunit* ( $SU_3$ ). Nei frames successivi i due segni si differenziano, quindi il prosieguo della trascrizione del segno 1 è differente da quello del segno 2.



*Esempio di differenti trascrizioni di due segni.*

# RICONOSCIMENTO DEL LINGUAGGIO DEI SEGNI

## *Riconoscimento di frasi nel linguaggio dei segni*

L'esempio riportato nella figura della slide precedente illustra solo un aspetto del segno eseguito: la progressione orizzontale della mano destra. Tale progressione orizzontale può corrispondere ad esempio alla coordinata associata alla posizione della mano destra sull'asse  $y$ . Per una descrizione completa di un segno, non è sufficiente un'unica feature. Infatti un sistema di riconoscimento deve gestire molte features, le quali possono essere raggruppate in *feature groups*. La composizione dei *feature groups* deve tenere conto dei parametri linguistici del linguaggio dei segni, cioè posizione, forma della mano e orientamento della mano.

Esempio di suddivisione del *feature vector* in *feature groups*:

- *feature group* «pos»: tutte le features riguardanti la posizione («pos») di entrambe le mani vengono raggruppate;
- *feature group* «size»: esso rappresenta tutte le features che descrivono la dimensione della parte visibile delle mani;
- *feature group* «dist»: in tale gruppo sono incluse tutte le features riguardanti le distanze tra le dita.

Gli ultimi due gruppi indicano la forma della mano e l'orientamento della mano.

# RICONOSCIMENTO DEL LINGUAGGIO DEI SEGNI

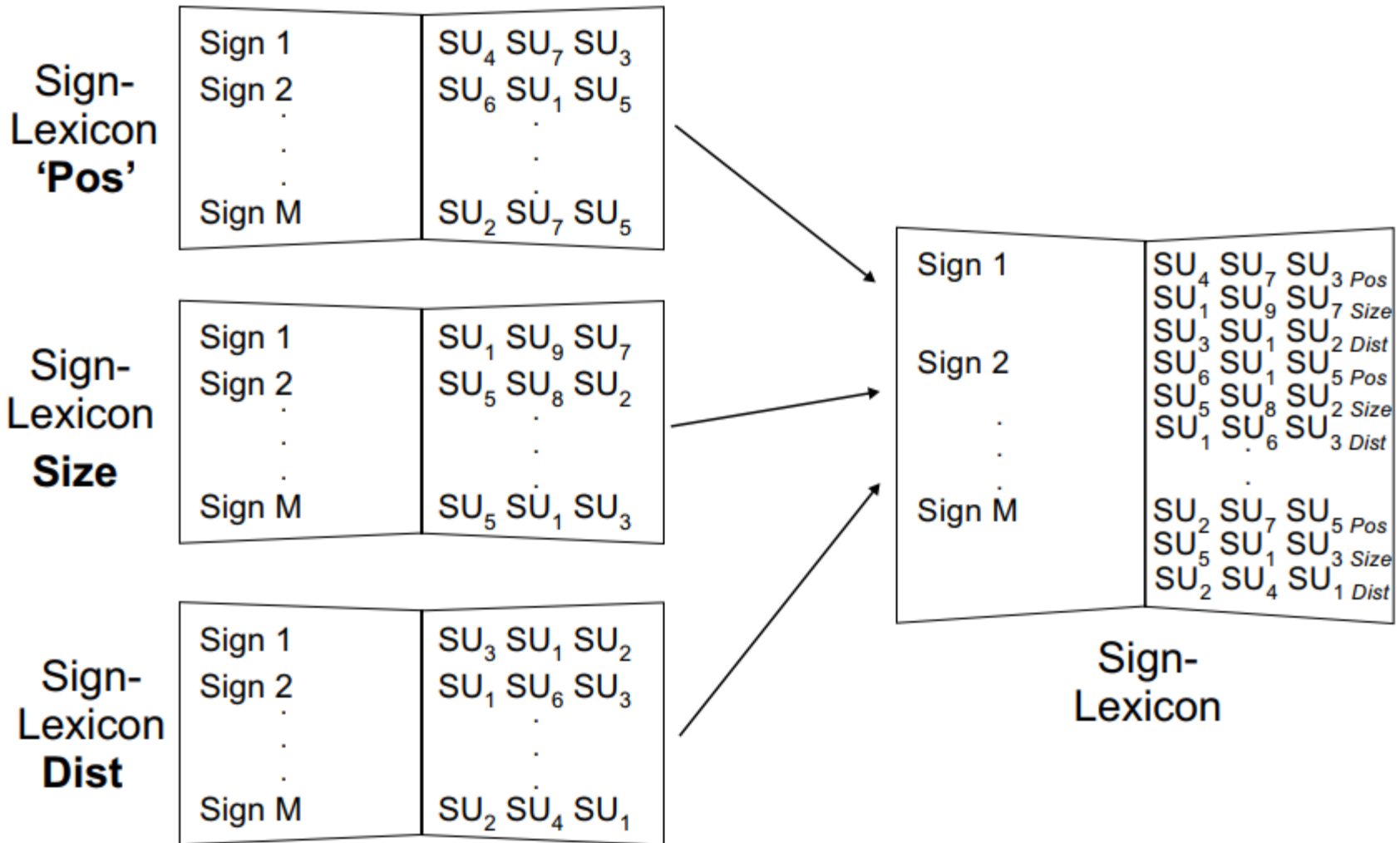
## *Riconoscimento di frasi nel linguaggio dei segni*

Seguendo la struttura parallela di un segno, ogni *feature group* risultante viene segmentato in ordine sequenziale in *subunits*. L'identificazione di segmenti simili non viene eseguita sull'intero *feature vector* ma solo all'interno di ognuno dei tre *feature groups*. Segmenti simili rappresentano le *subunits* di un *feature group*. Ad esempio, i segmenti del *feature group* «pos» rappresentano una certa posizione indipendentemente da ogni forma della mano e orientamento della mano.

La struttura parallela e sequenziale dei segni restituisce tre lessici differenti, i quali vengono uniti in un unico lessico (si veda la figura riportata nella slide seguente).

# RICONOSCIMENTO DEL LINGUAGGIO DEI SEGNI

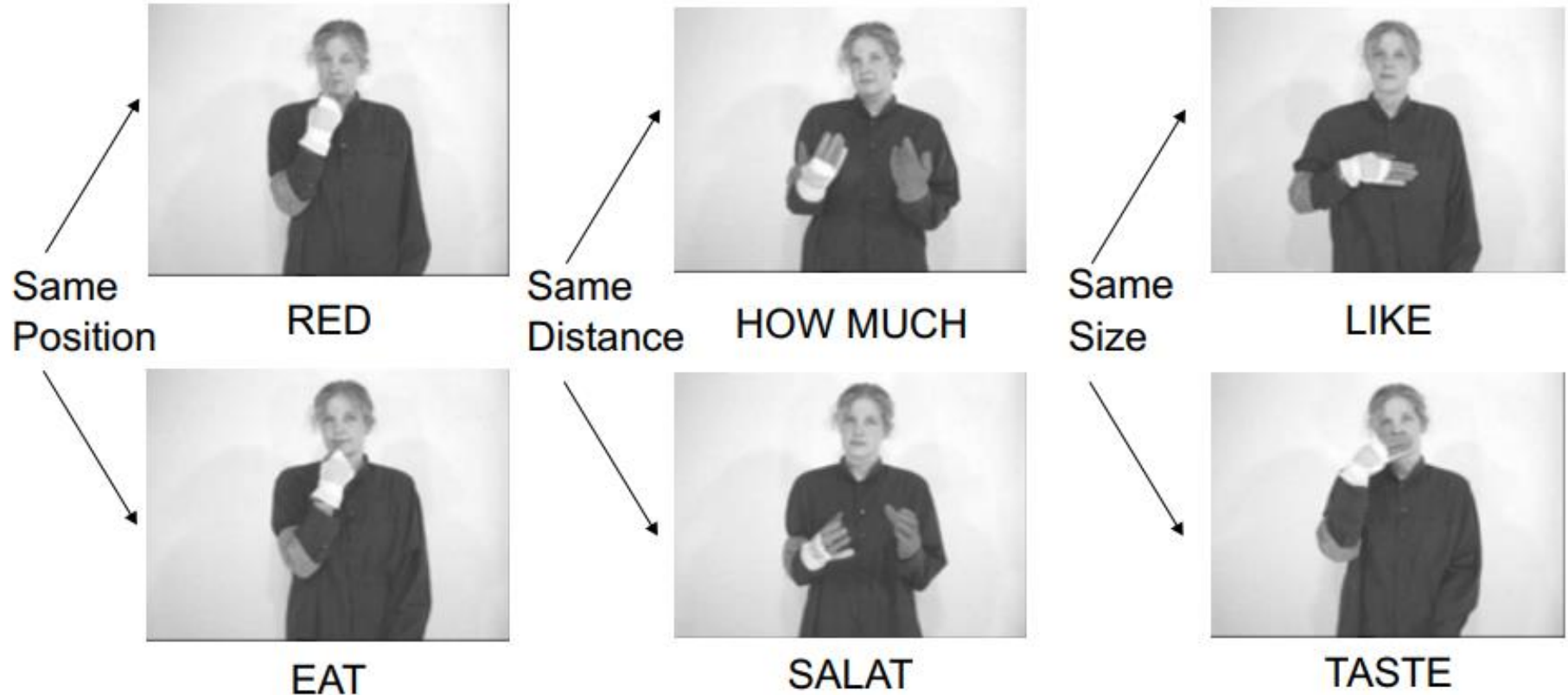
## *Riconoscimento di frasi nel linguaggio dei segni*



*Ogni feature group è caratterizzato da un proprio lessico dei segni; i tre lessici dei segni vengono poi combinati in un unico lessico dei segni.*

# RICONOSCIMENTO DEL LINGUAGGIO DEI SEGNI

## *Riconoscimento di frasi nel linguaggio dei segni*



*Esempi di segmenti simili di segni differenti in accordo a feature groups specifici.*

# RICONOSCIMENTO DEL LINGUAGGIO DEI SEGNI

## *Riconoscimento di frasi nel linguaggio dei segni*

I modelli *Hidden Markov Models* (*HMMs*) convenzionali sono adeguati per gestire sequenze di gesti (come descritto in precedenza). Tuttavia, per quanto riguarda il riconoscimento dei segni, la suddivisione in *feature groups* genera segnali paralleli che possono essere gestiti mediante i modelli *Parallel Hidden Markov Models* (*PaHMMs*). I *PaHMMs* sono *HMMs* utilizzati in parallelo. Ognuno degli *HMMs* combinati in parallelo viene denominato *canale* (*channel*). Ogni *canale* è indipendente dagli altri.

Per il riconoscimento basato su *subunits*, ognuno dei *feature groups* viene modellizzato mediante un *canale* di un *PaHMM*.



# RICONOSCIMENTO DEL LINGUAGGIO DEI SEGNI

## *Esempi di valutazione delle prestazioni*

La tabella mostra il tasso di riconoscimento relativo ad un test con 4 segnanti e diverse risoluzioni video (riconoscimento di segni isolati person-dependent). Il test è stato effettuato considerando «controlled laboratory conditions». La risoluzione durante la fase di training è stata sempre 384x288. Poiché il numero di segni registrati varia leggermente, la dimensione del vocabolario è specificata separatamente per ogni segnante. Sono state considerate solo features associate alle mani. Osservando l'ultima riga, si nota che, utilizzando come features solo le coordinate del centro di gravità (COG), si ottiene un tasso di riconoscimento medio circa uguale a 95% su un vocabolario di circa 230 segni.

Test Video Resolution	Features	Signer, Vocabulary Size				
		Ben 235 signs	Michael 232 signs	Paula 219 signs	Sanchu 230 signs	∅ 229 signs
384 × 288	all	98.7%	99.3%	98.5%	99.1%	98.9%
192 × 144	all	98.5%	97.4%	98.5%	99.1%	98.4%
128 × 96	all	97.7%	96.5%	98.3%	98.6%	97.8%
96 × 72	all	93.1%	93.7%	97.1%	95.9%	94.1%
384 × 288	$x, \dot{x}, y, \dot{y}$	93.8%	93.9%	95.5%	96.1%	94.8%

*Riconoscimento di segni isolati person-dependent in ambienti controllati mediante features associate alle mani (fase di test).*

# RICONOSCIMENTO DEL LINGUAGGIO DEI SEGNI

## *Esempi di valutazione delle prestazioni*

Nelle stesse condizioni elencate nella slide precedente, la posizione della testa, la posizione delle sopracciglia e il profilo delle labbra sono state utilizzate come features non associate alle mani. Il tasso di riconoscimento ottenuto sui vocabolari considerati nella tabella della slide precedente varia tra 49.3% e 72.4% tra i 4 segnanti. Il tasso di riconoscimento medio è pari a 63.6%. Quindi, circa due segni su tre vengono riconosciuti mediante features non associate alle mani: tale risultato enfatizza l'importanza della mimica per il riconoscimento del linguaggio dei segni.

# RICONOSCIMENTO DEL LINGUAGGIO DEI SEGNI

## *Esempi di valutazione delle prestazioni*

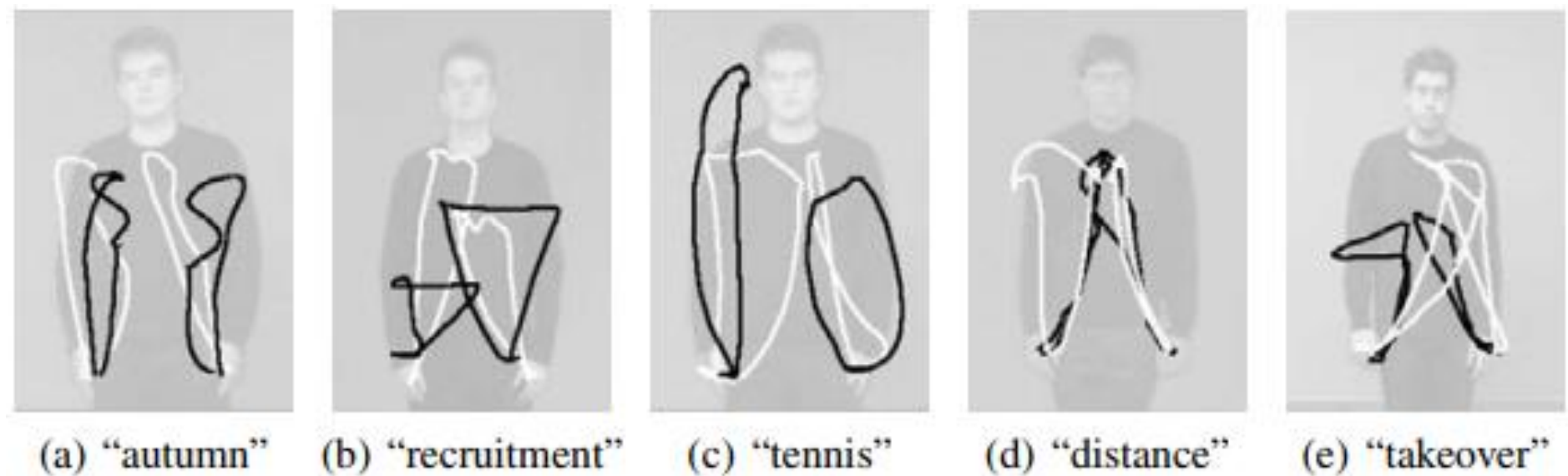
La tabella mostra alcuni risultati relativi al riconoscimento person-independent. Poiché i segnanti hanno utilizzato segni differenti, il vocabolario è stato ottenuto mediante l'intersezione tra i segni della fase di test e l'unione di tutti i segni della fase di training. I risultati sono peggiori rispetto al caso person-dependent. Ciò è causato da un'elevata varianza (*interpersonal variance*) associata all'esecuzione dei segni da parte di persone diverse (si veda la figura della slide seguente, la quale mostra come varia il COG per lo stesso segno eseguito da segnanti diversi).

Training Signer(s)	Test Signer	Vocabulary Size	n-Best Rate		
			1	5	10
Michael	Sanchu	205	36.0%	58.0%	64.9%
Paula, Sanchu	Michael	218	30.5%	53.6%	63.2%
Ben, Paula, Sanchu	Michael	224	44.5%	69.3%	77.1%
Ben, Michael, Paula	Sanchu	221	54.2%	79.4%	84.5%
Ben, Michael, Sanchu	Paula	212	37.0%	63.6%	72.8%
Michael, Sanchu	Ben	206	48.1%	70.0%	77.4%

*Riconoscimento di segni isolati person-independent in ambienti controllati (fase di test). Il n-Best Rate indica la percentuale di risultati per i quali il segno corretto è tra gli n segni ritenuti più simili al segno di input dal classificatore. Per n=1 si ha il tasso di riconoscimento.*

# RICONOSCIMENTO DEL LINGUAGGIO DEI SEGNI

## *Esempi di valutazione delle prestazioni*



*Tracciamento di  $(x_{cog}, y_{cog})$  per entrambe le mani considerando due persone differenti (bianco e nero) che eseguono lo stesso segno. Lo sfondo mostra una delle due persone come riferimento.*

# RICONOSCIMENTO DEL LINGUAGGIO DEI SEGNI

## *Esempi di valutazione delle prestazioni*

I risultati di riconoscimento person-independent in ambienti non controllati sono difficili da valutare poiché essi dipendono da parametri multipli (segnante, vocabolario, sfondo, illuminazione, camera). Rumore e outliers sono introdotti inevitabilmente nelle features quando ci si trova in scenari reali. Inoltre si ha un'elevata varianza associata all'esecuzione dei segni da parte di persone diverse (*interpersonal variance*). Nella tabella sono riportati alcuni risultati riguardo al riconoscimento person-independent per piccoli vocabolari. Ogni segnante ha eseguito i segni in uno scenario reale differente e la selezione dei segni è rappresentativa dell'intero vocabolario (esso contiene segni basati sull'utilizzo di una mano e due mani, sia con che senza sovrapposizione).

Vocabulary Size	Test Signer						
	Christian	Claudia	Holger	Jörg	Markus	Ulrich	∅
6	96.7%	83.3%	96.7%	100%	100%	93.3%	95.0%
18	90.0%	70.0%	90.0%	93.3%	96.7%	86.7%	87.8%

*Tassi di riconoscimento (riconoscimento di segni isolati person-independent) in scenari reali (fase di test).*

## *Riferimenti Bibliografici*

- [1] Kraiss, K. -F. (2006). Advanced Man-Machine Interaction: Fundamentals and Implementation. Springer-Verlag Berlin Heidelberg. ISBN-10: 3-540-30618-8