ES

Clasificación del lenguaje de señas mexicano con SVM generando datos artificiales

Classification of Mexican Sign Language SVM generating data

Jair Cervantes*
Farid García-Lamont**
José H. Santiago***
Josue Espejel Cabrera****
Adrian Trueba****

Fecha de recepción: 16 de marzo de 2013
Fecha de aprobación: 30 de abril 30 de 2013

Resumen

El desarrollo de herramientas que faciliten la comunicación de personas sordas es un reto de investigación actual muy importante. Una línea de investigación es el desarrollo de sistemas de visión con un gran poder de generalización. Obtener una buena precisión de generalización requiere un conjunto de datos muy grande durante el entrenamiento, y el incremento de datos muchas veces solo añade información repetitiva y no representativa. En este artículo describimos el desarrollo de un sistema de reconocimiento de lenguaje de señas. El sistema propuesto permite identificar con una alta precisión el Lenguaje de Señas Mexicano introduciendo datos muy representativos generados artificialmente, que permiten mejorar la capacidad de generalización del cla-

* Posgrado e Investigación UAEMEX-Texcoco jcervantesc@uaemex.mx
** Posgrado e Investigación, UAEMEX-Texcoco, fglamont@yahoo.com.mx
*** Posgrado e Investigación, UAEMEX-Texcoco, josehsantiago@hotmail.com
**** Posgrado e Investigación, UAEMEX-Texcoco jec0309@hotmail.com
**** Posgrado e Investigación, UAEMEX-Texcoco atruebae@gmail.com

sificador. Los resultados obtenidos muestran que el algoritmo propuesto mejora la precisión de generalización de las SVM al utilizar la metodología propuesta.

Palabras clave

Lenguaje de Señas Mexicano, Clasificación, SVM

Abstract

The development of tools to facilitate communication of deaf people is a current research challenge very important. One line of research is the development of vision systems with a great power of generalization. Get a good generalization accuracy requires a very large data set during training, and increase data just adds information often repetitive and not representative. In this paper we describe the development of a system of sign language recognition. The proposed system can identify with high accuracy the Mexican Sign Language very representative introducing artificially generated data, which can improve the generalization ability of the classifier. The results show that the proposed algorithm enhances the accuracy of the SVM to widespread use methodology.

Keywords

Mexican Sign Language, classification, SVM

1. Introducción

Los recientes avances en poder de cómputo, miniaturización de componentes y aprendizaje de máquinas han permitido el desarrollo de múltiples herramientas y dispositivos que ayuden a mejorar las condiciones de vida del ser humano en diferentes campos. Los avances realizados en reconocimiento de patrones, minería de datos e inteligencia artificial han impactado notoriamente en disciplinas tan diversas como medicina, agronomía, finanzas, aeronáutica, etc. Tan solo en medicina se han realizado investigaciones que van desde detección de cáncer, movimiento anormal de los ojos, diagnóstico de enfermedades, detección de tumores hasta dispositivos que determinen la cantidad de radiación por emitir para tratar a personas con cáncer.

El desarrollo de dispositivos que ayuden a personas con ciertas limitaciones como pérdida o disminución de sus facultades físicas, intelectuales o sensoriales, para realizar sus actividades connaturales no ha sido la excepción. Uno de los campos de investigación activa es el reconocimiento de lenguaje de señas, ya sea para establecer comunicación entre un emisor-receptor (persona sordomuda) mientras se generan

animaciones que realizan señas a partir de texto o también entre un emisor (persona sordomuda)-receptor mediante reconocimiento de lenguaje. El objetivo de reconocimiento de lenguaje es proveer un mecanismo eficiente y preciso que interprete el lenguaje de señas en texto.

La dificultad de las personas sordas para comunicarse disminuye considerablemente su desarrollo educativo, profesional y humano. Los principales problemas de las personas con discapacidad son el desempleo, la discriminación y una enorme dependencia, que tienen una relación directa con su capacidad de interacción social. Aunado a lo anterior, a pesar de ser un problema muy grande en México existen alrededor de 400 mil personas con discapacidad auditiva, según datos del Censo de Población y Vivienda 2010; además, muy poca gente tiene un buen conocimiento del lenguaje de señas. La Lengua de Señas Mexicana (LSM) es considerada como el lenguaje utilizado por la comunidad sorda teniendo valor como cualquier otro lenguaje, pues posee vocabulario propio, gramática y sintaxis especial, siendo completamente capaz de expresar tan amplia gama de pensamientos y emociones como cualquier otra lengua, ademas de tener expresiones idiomáticas propias.

El lenguaje de señas representa el principal medio de comunicación para personas sordas. Sin embargo, la mayoría de las veces es necesario un traductor para tener comunicación con una persona sorda, esto representa un problema muy grande no solo en situaciones de emergencia, sino también en problemas o situaciones cotidianas, lo cual, a su vez, limita a las personas sordas por la enorme cantidad de restricciones que les impiden contar con mecanismos de interacción y relación con la sociedad. El desarrollo de herramientas prácticas que faciliten la comunicación empleando el LSM para personas sordas y para personas oyentes, es de vital importancia a fin de facilitar la interacción, el diálogo y la información, tanto en lo social como en lo privado, además de posibilitar el acceso a la educación y al empleo, entre muchos otros espacios de la vida.

El desarrollo de herramientas que ayuden a la comunicación total emisor ps-receptor [4, 10, 27], emisor-receptor ps [5, 6, 23, 24] es pues, un reto de investigación actual muy importante. La investigación de varios autores se ha centrado en la comunicación en ambas direcciones, permitiendo no solo generar señas en video al introducir frases en una PC disminuyendo la barrera de comunicación que existe entre una persona que no conoce el lenguaje de señas y una persona sordomuda, sino también traducir un diálogo o identificar palabras del lenguaje de señas emitidas por una persona sordomuda y traducidas por sistemas de reconocimiento de lenguaje de señas.

Traducir un diálogo al lenguaje de señas no es un reto fácil debido a que las señas utilizadas en el LSM son movimientos o detenciones en algún punto del cuerpo o del espacio. Cada seña puede estar compuesta de cinco o incluso siete rasgos; configuración de la mano, ubicación, dirección, orientación, rasgos no manuales, lugar de articulación y punto de contacto. Además, cada persona realiza ligeras variaciones al utilizar el LSM. La figura 1muestra diferentes variaciones de las señas que representan las letras "C" y "P". Varios autores han explorado diversas técnicas para resolver este problema durante los últimos quince años. En [25] los autores emplean seis momentos geométricos para reconocer setenta señales del Lenguaje de Señas Japonés, segmentando cara y manos. En [10] Dreuw desarrolla un sistema de traducción que obtiene texto a partir de sentencias en lenguaje de señas. Dreuw utiliza señales de video como entrada y obtiene diferentes características obtenidas a partir de varias técnicas incluido el PCA, velocidad y trayectoria de la mano para reconocimiento. En sus experimentos muestrean y reducen a imagen original, lo que podría llevar a perder información importante.

En [27], los autores utilizan una combinación de características geométricas inclu-

Figura 1. Variaciones al describir vocales

Letra P **Letra C**

Fuente: elaboración propia.

yendo intensidad de la imagen, intensidad de la textura y diferentes tipos de derivadas de primero y segundo orden para reconocer imágenes segmentadas. Los resultados reportados por Zadehi muestran errores en los datos de prueba del 30 %. Bauer [4] desarrolla un sistema de reconocimiento de señas alemán, obteniendo el área de las manos como característica; en sus experimentos reporta precisiones del 91 % al 94 %. El trabajo presentado en este artículo da un paso en el desarrollo de sistemas automáticos de traducción de señas de la LSM al implementar un método de generación de puntos artificiales. En reconocimiento de patrones es necesario que el conjunto de datos con el que se va a entrenar sea lo más completo posible, sin embargo, es bien sabido que es imposible obtener un conjunto de datos que contenga todas las variantes posibles con que una palabra puede expresarse mediante señas. La cantidad de variantes y la singularidad con que cada persona puede expresar una palabra lo hace prohibitivo. En este artículo se propone un método alternativo de generación de puntos artificiales, los puntos generados permiten disminuir la incertidumbre entre clases, lo que permite mejorar la capacidad de generalización. El sistema propuesto reconoce veintiocho señas del LSM empleando imágenes. En la metodologí a propuesta empleamos técnicas de extracción de características geométricas invariantes a escalado, traslación y rotación y utilizamos SVM para realizar el trabajo de reconocimiento.

2. Estado del arte

Varios investigadores han trabajado en los últimos años en reconocimiento de lenguaje de señas. Las diferentes líneas que se han desarrollado podrían delimitarse a la forma como son obtenidas las características de entrada, que podrían ser:

1. Métodos basados en obtener impulsos generados por guantes. Los diferentes sensores colocados en el guante van generando impulsos a partir de la orientación, posición y ángulos de la mano. Estos impulsos son trasladados a mediciones que en conjunto determinan el tipo de seña realizada.

2. Métodos basados en visión. A partir de video son obtenidas las imágenes, que pasan por diversos procesamientos hasta obtener las características que definen a cada seña mediante forma, longitud de la mano, posición, exión y ángulos tanto de la mano como de los dedos.

Los métodos basados en visión han recibido más atención debido a la necesidad de métodos automáticos de traducción. En [15], los autores desarrollan un sistema de reconocimiento del lenguaje de señas arábigo; el método se basa en detectar las posiciones de las yemas de los dedos y las muñecas. Los autores emplean un guante con seis diferentes colores para determinar las posiciones de las yemas de los dedos y muñecas, estas distancias son calculadas y sirven como características de entrada para una red neuronal difusa que determina la seña realizada. Dreuw [10] obtiene características basadas en apariencia para reconocer señas a partir de secuencias de video. En sus experimentos utiliza 35 señas de 20 personas distintas, las imágenes que utiliza son imágenes de 32x 32, que podría llevar a una gran pérdida de información.

Dreuw [10] presenta un enfoque de reconocimiento automático de señas y traducción a un lenguaje oral. Dreuw usa una combinación de características en el reconocimiento como posición, velocidad y trayectoria de la mano, además de utilizar PCA para reducir la dimensión del conjunto obtenido.

Por su parte,[1] presenta un sistema de reconocimiento de lenguaje de señas arábigo basado en modelos ocultos de Markov (hidden Markov models). En el método que propone emplear una transformada de coseno discreta para extraer características y representa la imagen como una suma de sinusoidales de magnitud y frecuencia variable. En sus simulaciones emplea treinta palabras y al igual que [15] emplea guantes marcados con seis colores en diferentes regiones. Nam *et al.* [22] describen un método basado en modelos ocultos de Markov para reconocimiento de patrones de movimiento espacio-tiempo de la mano. En sus experimentos obtienen 300 patrones para cada movimiento. Habili [13] propone una técnica de segmentación empleando color y movimiento para realizar una representación del contenido de las secuencias de video. La técnica consiste de tres etapas; segmentación de color, detección de cambios y segmentación de la mano y cara. Cui [7] emplea partición recursiva y utiliza PCA y análisis de discriminantes en cada nodo de un árbol para automáticamente derivar el mejor subconjunto de características a partir de imágenes. Durante el entrenamiento, su método es capaz de llevar a cabo clasificación no lineal en el espacio de características. Vogler emplea tres cámaras posicionadas en una configuración ortogonal para explotar información obtenida a partir de parámetros 3D y utiliza modelos ocultos de Markov para reconocimiento.

Cada uno de los autores mencionados antes desarrollan modelos con el fin de mejorar la precisión de reconocimiento o disminuir el tiempo de detección. Una desventaja muy importante en reconocimiento es la cantidad de ejemplos de entrenamiento. Un conjunto de entrenamiento pequeño puede degradar severamente la precisión del reconocimiento y en alguna ocasiones, cuando el modelo obtenido durante el entrenamiento es probado en conjuntos de datos con imágenes de una persona sordomuda diferente, la precisión cae considerablemente [2,11,17,26]. Algunos investigadores han propuesto diversos métodos para enfrentar esta desventaja que van desde sobre ejemplificado [3] reejemplificado [19], generando nuevas imágenes y agrandando el conjunto de entrenamiento [8], modificando la iluminación en las imágenes e insertando las imágenes modificadas en el conjunto de entrenamiento [12] e incluso generando datos artificiales. [16] desarrolla un sistema para generar datos artificiales a partir de datos obtenidos mediante impulsos eléctricos de un guante.

Los resultados mostrados en todas las investigaciones anteriores muestran que al incrementar el conjunto de datos de entrenamiento se mejora la precisión de reconocimiento. Sin embargo, es claro que en muchas ocasiones anexar datos al conjunto de entrenamiento se puede duplicar la información existente sin anexar información importante al clasificador. Por otro lado, aunque el conjunto de datos sea grande, algunos casos requieren delimitar perfectamente la frontera entre datos de una y otra clase. Otro problema importante en clasificación es la fase de entrenamiento. En esta fase se requiere obtener un conjunto de datos con las características más discriminantes, de forma que el modelo obtenido en la fase de entrenamiento facilite y optimice la capacidad de generalización. En clasificación de imágenes una gran desventaja radica en que estas características pueden verse afectadas de varios procesos previos, como son preprocesamiento, segmentación y extracción de características, cada uno con una complejidad diferente, pero muy importantes y necesarias. Una mala segmentación, preprocesamiento o extracción de características puede provocar estragos en el reconocimiento de la imagen.

Sin embargo, aunque se realicen todos estos pasos de manera adecuada, no es posible

dotar al clasificador de un conjunto de datos con las mejores características para obtener un modelo completamente discriminante, esto debido a que se ocupará de una cantidad de imágenes enorme, con muchos datos repetidos y poca o nula información importante adherida en cada imagen después de cierta cantidad de imágenes base. El método propuesto permite encontrar las mejores características en un conjunto de datos, y a partir de ahí generar nuevos datos con características muy similares a los mejores datos, disminuyendo la incertidumbre en regiones vitales y mejorando la habilidad discriminativa del modelo resultante. Aunque el método propuesto no sustituye los métodos de preprocesamiento de la imagen, si mejora la capacidad de generalización del clasificador.

3. Preliminares

En esta sección se muestran los métodos empleados para desarrollar el sistema propuesto.

3.1 Extracción de características

3.1.1 Características básicas

Las características que se emplearon describen las propiedades básicas de la región a reconocer; estas son: área de la región, redondez de a mano, longitud del borde de la mano, elongación de la mano definida por la longitud y ancho de la mano, las coordenadas x e y del centro de gravedad, densidad, definida por la longitud de los bordes de la mano y el área de esta. En total nueve características fueron calculadas. La mayoría de estas son bien conocidas, por razones de espacio solo se definen las dos últimas

1. Densidad.

$$Densidad = \frac{(longitud\_región\_mano)^2}{A}$$

2. Distancia centroide. Distancia de los bordes de la mano al centroide (x_c; y_c)

$$r(t) = [(x(t) - x_c)^2 + (y(t) - y_c^2)^{1/2}]$$

donde

$$x_c = \frac{1}{L}\sum_{t=0}^{L-1} x(t), y_c = \frac{1}{L}\sum_{t=0}^{L-1} y(t)$$

donde x_c es el centroide i; y_c es el centroide j, (x_c; y_c) es el centroide del objeto.

3.1.2 Momentos

Los momentos son muy empleados en reconocimiento de imágenes, estos permiten reconocer imágenes independientemente de su rotación, traslación o inversión.

Los momentos de orden (p + q) son definidos como:

$$r(t) = [(x(t) - x_c)^2 + (y(t) - y_c^2)^{1/2}]$$

$$m_{pq} = \sum\sum x^p y^q \rho(x, y). \quad (1)$$

donde (x; y) es definida por la región segmentada. Los momentos de orden pequeño describen la forma de la región. Por ejemplo m_{00} describe el área de la región segmentada, mientras que m_{01} y m_{10} definen las coordenadas x e y del centro de gravedad. Sin embargo, los momentos m_{02}, m_{03}, m_{11}, m_{12}, m_{20}, m_{21} y m_{30} son invariantes a traslación, rotación e inversión. Los momentos centrales son invariantes a desplazamiento y pueden ser calculados mediante

Los momentos de Hu pueden ser obtenidos mediante:

$$\mu_{pq} = \sum_{i,j \in R} (i - \bar{i})^p (j - \bar{j})^q \quad (2)$$

Donde *p,q* pertenecen a la región segmentada y el centro de gravedad de la región es definido por:

$$\bar{i} = \frac{m_{10}}{m_{00}}, \bar{j} = \frac{m_{01}}{m_{00}} \quad (3)$$

Los momentos Hu pueden ser obtenidos mediante:

Otras caracter sticas empleadas fueron descriptores de elipse, convexidad de region, momentos de Flusser y orientacion, en total 57 caracter sticas fueron extraidas de cada imagen.

3.2 Máquinas de Vectores Soporte (SVM)

Las SVM son una de las tecnicas de clasi cacion mas uti-lizadas en los ultimos a~nos. El sistema de reconocimiento propuesto emplea SVM para clasi car las se~nas una vez obtenidas sus caracter sticas. Formalmente las SVM puede ser de nidas de la siguiente manera:

Asumiendo que un conjunto de datos de entrenamiento X es dado como:

$$\phi_1 = \eta_{20}+\eta_{02} \quad (4)$$

$$\phi_2 = (\eta_{20}-\eta_{02})^2 + 4\eta^2_{11} \quad (5)$$

$$\phi_3 = (\eta_{30}-3\eta_{12})^2 + (3\eta_{21}-\eta_{03})^2 \quad (6)$$

$$\phi_4 = (\eta_{30}-3\eta_{12})^2 + (3\eta_{21}-\eta_{03})^2 \quad (7)$$

$$\phi_5 = (\eta_{30}-3\eta_{12})(\eta_{30}+\eta_{12})[(\eta_{30}+\eta_{12})^2 - 3(\eta_{21}+\eta_{03})^2] (\eta_{21}-\eta_{03})(\eta_{21}+\eta_{03})[3(\eta_{30}+\eta_{12})^2 - (\eta_{21}+\eta_{03})^2] \quad (8)$$

$$\phi_6 = (\eta_{20}-\eta_{02})[(\eta_{30}+\eta_{12})^2 - (\eta_{21}+\eta_{03})^2]+ 4(\eta_{11}(\eta_{30}-\eta 12)(\eta_{21}+\eta_{03}) \quad (9)$$

$$\phi_7 = (3\eta_{21}-3\eta_{03})(\eta_{30}+\eta_{12})[(\eta_{30}+\eta_{12})^2 - 3(\eta_{21}+\eta_{03})^2] (\eta_{30}-\eta_{12})(\eta_{21}+\eta_{03})[3(\eta_{30}+\eta_{12})^2 - (\eta_{21}+\eta_{03})^2] \quad (10)$$

Donde $\eta_{pq}=\frac{\mu rs}{\mu 00^t}$, $t=\frac{p+q}{2^t}+1$

Otras características empleadas fueron descriptores de elipse, convexidad de región, momentos de Flusser y orientación, en total 57 características fueron extraídas de cada imagen.

3.2 Máquinas de Vectores de Soporte (SVM)

Las SVM son unas técnicas de clasificación más utilizadas en los últimos años. El sistema de reconocimiento propuesto emplea SVM para clasificar las señas una vez obtenidas sus características. Formalmente, las SVM pueden ser definidas de la siguiente manera:

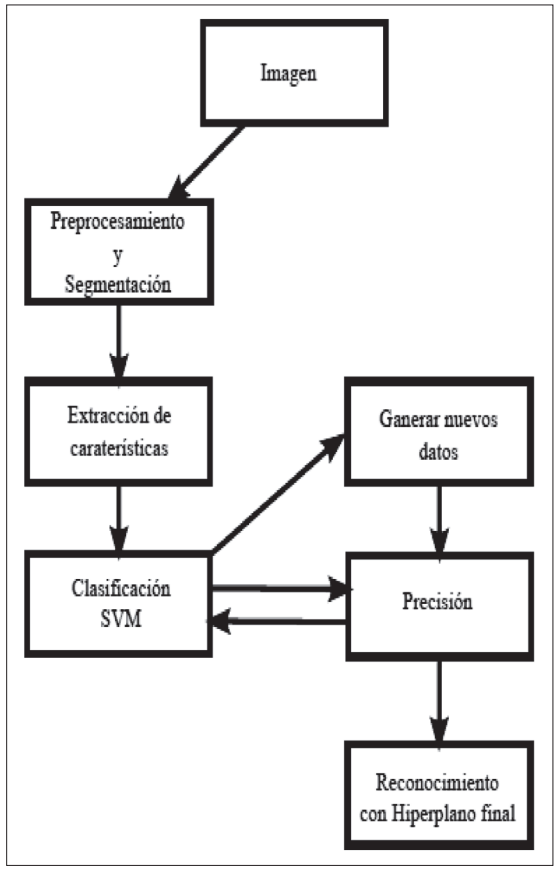
Asumiendo que un conjunto de datos de entrenamiento X es dado como:

$$(x_1, y_1), (x_2, y_2),...., (x_n, y) \quad (11)$$

i.e.$X = \{x_i, y_i\}_i^n = 1$ donde $x_i \in R^d$ y $y_i \in (+1,-1)$. Entrenar una SVM permite resolver un problema de programación cuadratica como sigue:

$$\max_{\alpha_i} -\frac{1}{2}\sum_{i,j=1}^{l}\alpha_i y_i \alpha_j y_j \mathbf{K}\langle x_i \cdot x_j\rangle + \sum_{i=1}^{l}\alpha_i$$
$$\text{sujeto a: } \sum_{i=1}^{l}\alpha_i y_i = 0, \quad C \geq \alpha_i \geq 0, i = 1, 2, \ldots l \quad (11)$$

Figura 2. Metodología propuesta



Fuente: elaboración propia.

donde C >0, α_i =[α_1, α_2,... α_l]^T, $\alpha_i \geq 0$,i=1, 2, ...l son coeficientes que corresponden x_i, $x_{i'}$ con α_i diferentes a cero que son llamados Vectores Soporte (SV). La función K es una función, que debe satisfacer las condiciones de Mercer.

Sea S el conjunto de SV obtenidos después del entrenamiento, entonces el hiperplano óptimo es dado por:

$$\sum_{i \in S} (\alpha_i y_i) K(\mathbf{x}_i, \mathbf{x}_j) + b = 0 \qquad (13)$$

Y la función de decisión óptima es definida como:

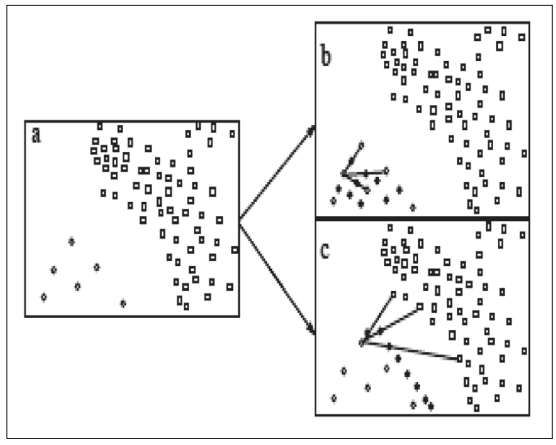
$$f(x) = sign(\sum_{i \in S} (\alpha_i y_i) K(\mathbf{x}_i, \mathbf{x}_j) + b) \qquad (14)$$

donde x=[x_1, x_2,... α_l]s son los datos de entrada, α_i y y_i son los multiplicadores de Lagrange. Un nuevo objeto x puede ser clasificado si se emplea (14). El Vector x_i es dado en la forma de producto punto. Existe un multiplicador de Lagrange para cada punto de entrenamiento. Cuando el máximo margen del hiperplano es encontrado, solamente los puntos más cercanos al hiperplano satisfacen α> 0: Estos puntos son los SV:

4. Metodología

El objetivo de la investigación en reconocimiento automático de señas tiene como fin desarrollar sistemas con un buen desempeño. Sin embargo, esto requiere que los datos empleados en el entrenamiento contengan enormes cantidades de información, bajo diferentes condiciones y de distintas personas. Es decir, se requiere que la información sea tan representativa como sea posible. En la mayoría de los casos, esto no es posible por dos razones, la cantidad de imágenes de señas siempre estar limitada a un conjunto de personas y los tiempos de entrenamiento de los clasificadores. Por otro lado, al incrementar la cantidad de imágenes en el conjunto de entrenamiento, anexamos datos en su mayoría innecesarios. En este artículo desarrollamos un método para generar datos artificiales para mejorar el desempeño de los sistemas de clasificación. El método propuesto obtiene un modelo de clasificación con SVM a partir de un conjunto de datos de entrenamiento inicial, de los SV obtenidos, genera nuevos datos. Los SV son los datos más importantes de conjunto de datos de entrenamiento, esto es, los datos más discriminantes del conjunto de datos entero. Tal es su capacidad discriminativa de los SV que la solución es dada mediante únicamente estos datos.

Figura 3. Técnicas de generación de datos sintéticos



Fuente: elaboración propia.

Ya que los SV son los datos con mayor capacidad discriminativa, al generar nuevos datos modificando los SV es posible mejorar el desempeño del clasificador, en otras palabras, estamos añadiendo información representativa al modelo. Con el objetivo de anexar información importante a los clasificadores, algunas técnicas han sido propuestas en clasificación de señas, En [16] se mencionan dos métodos para generar datos internamente y externamente (ver figura 3). Sin embargo, generar datos internos a partir de SV no mejora su capacidad discriminativa, generar datos externos pareciera ser la solución, pues estos poseen mayor información, pero se debe tener especial cuidado al generar datos externos. En este artículo se desarrolla un método para generar datos externos.

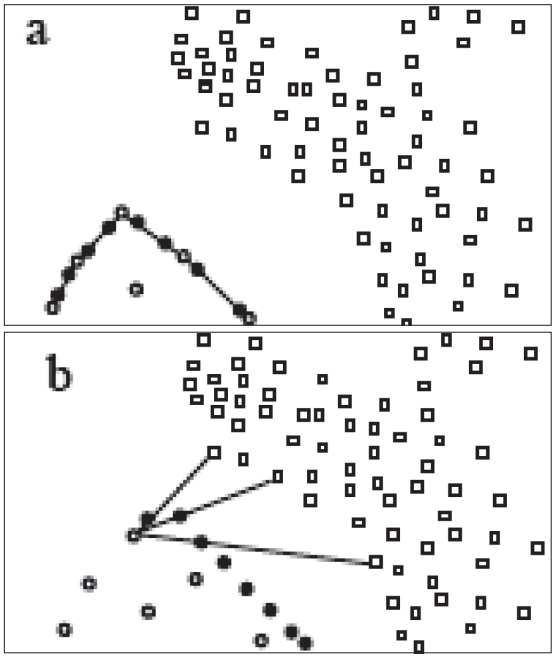
La metodología del sistema propuesto es mostrada en la figura 2. Los primeros pasos son los habituales en cualquier sistema de

reconocimiento de señas a partir de características geométricas. Debido a las condiciones de espacio se definen brevemente los pasos que son comunes en los sistemas de clasificación y que son llevados a cabo también por el método propuesto, pero se da especial énfasis en la variante original propuesta en este artículo. Los algoritmos 1 y 2 describen cada uno de los pasos empleados en el método propuesto.

Primero, las imágenes son preprocesadas y segmentadas. Regularmente en preprocesamiento se emplea una máscara Gaussiana para obtener una buena segmentación. En las simulaciones realizadas se aplicó este tipo de máscaras. La región de la mano en cada imagen fue segmentada empleando los siguientes pasos: 1) cálculo de alto contraste en escala de grises a partir de combinación lineal óptima de los componentes de color en RGB; 2) estimar una frontera óptima empleando momentos acumulativos de orden cero y de primer orden; 3) operaciones morfológicas para rellenar posibles espacios vacíos en la imagen segmentada. Todo esto con el objetivo de obtener una buena segmentación aun cuando existan cambios en las condiciones globales de brillo. Al segmentar la imagen, el sistema propuesto puede utilizar únicamente la región del ademán, determinar sus bordes y calcular las propiedades mediante la extracción de características.

Una vez segmentada la región se extraen sus características, para ello empleamos las técnicas de extracción descritas en el segundo apartado del presente escrito. La extracción de características nos permite representar la imagen mediante un conjunto de valores numéricos con gran poder discriminativo, eliminando características redundantes y reduciendo la dimensionalidad de la imagen. Las características obtenidas son capaces de asociar rangos muy similares a imágenes similares, asociar rangos diferentes a imágenes diferentes, además de ser invariantes a escalado, rotación y traslación, lo que le permite al clasificador reconocer objetos a pesar de tener diferente tamaño, diferente posición y orientación. Todas estas características desempeñan un papel importante en el desempeño del algoritmo y le permiten al clasificador discriminar de una forma apropiada entre distintas clases.

Figura 4. Técnicas propuestas de generación de datos sintéticos



Fuente: elaboración propia.

4.1 Generar datos artificiales

Realizada la extracción de características, se separa el conjunto de entrada en dos subconjuntos (entrenamiento y prueba) y se entrena una SVM. Los vectores soporte son los datos más importantes del conjunto de datos de entrenamiento y el modelo de clasificación es de nido por estos, este modelo puede modificarse solo si existen mejores datos en el conjunto de datos de entrenamiento, esto es, si alguno de los vectores soporte es sustituido por otro mejor.

Debido a que los vectores soporte definen un margen de separación resulta lógico que cualquier vector que mejore la precisión del modelo de existir, debe encontrarse dentro de ese margen de separación. En este artí-

Fuente: elaboración propia.

culo, los vectores soporte obtenidos son empleados para obtener nuevos datos que mejoren a los SV obtenidos en la primera fase y mejoren la precisión al generalizar. El método propuesto trabaja excitando los vectores soporte y desplazando los SV para encontrar un hiperplano más discriminante, la dirección del movimiento del SV es elegida de acuerdo con las Ecs.(15) y (17). Los algoritmos 1 y 2 representan el proceso completo del algoritmo propuesto.

$$\nu_i = \frac{x_{svi}^{+} - x_{ij}^{-}}{\left\| x_{ij}^{-} - x_{svi}^{+} \right\|_2}, \; i = 1, ..., \mathsf{X}_r^{-} \quad (15)$$

Entrada
X : Conjunto de datos(Matriz de características)
Salida
Sistema de reconocimiento LSM
Begin
$X_r^+ \leftarrow \{x_i \in X : y_i = +1\}$, i=1,...,p
$X_r^- \leftarrow \{x_i \in X : y_i = -1\}$, i=1,...,n
$H_1 \leftarrow$ trainSVM X_r^+, X_r^- /*Entrenar la SVM*/
$SV \leftarrow getSV(X_r^+, X_r^-)$ /*Obtener los SV*/
repeat
  Crear X_{sr}^+, X_{sr}^- nuevos datos con Algorithm 2
  $H_2 \leftarrow$ TrainSVM $X_r^+, X_r^- \cup X_{sr}^+, X_{sr}^-$ /*Calcular SVM con hiperplano mejorado*/
  $SV \leftarrow getSV(X_r^+, X_r^- \cup X_{sr}^+, X_{sr}^-)$ /*Obtener SV*/
  Acc(t)$\leftarrow TestSVMH_2(X_{rt}^+, X_{tr}^-)$ /*Probar precisión*/
  if (Acc(t)-Acc(t-1) > 0
    $H_f(X_{RD}^+, X_{RD}^-) = H_2(X_{rt}^+, X_{tr}^-)$
  end
while (Acc(t)-Acc(t-1) > 0)
return $H_f(X_{RD}^+, X_{RD}^-)$
End

Algorithm 1: Algoritmo de Sistema de Reconocimiento LSM

$$\nu_i = \frac{x_{svi}^{+} - x_{ij}^{+}}{\left\| x_{ij}^{+} - x_{svi}^{+} \right\|_2}, \; i = 1, ..., \left| X_r^{+} \right| \quad (16)$$

$$x_{ij}^{-} = jmin \left\| x_{svi} - x_{svj} \right\|_2, \; j = 1, ..., \mathsf{X}_r^{+} \quad (17)$$

$$x_{ij}^{+} = min \left\| x_{svi} - x_{svj} \right\|_2, \; i, j = 1, ..., \left| X_r^{+} \right|, i \neq j \quad (18)$$

$$x_{svi}^{-} \in SV \text{ de } X_{tr}^{-} \quad (19)$$

$$x_{svj}^{+} \in SV \text{ de } X_{tr}^{+} \quad (20)$$

La magnitud de desplazamiento es definido por $\in$, los mejores valores que se encontraron de $\in$están entre $1 \square X \square 10^{-1}$ y $1 \square X \square 10^{-3}$.

$$x_{svg} = x_{svi} + \in \cdot v_i \quad (21)$$

Los nuevos datos son generados por los SV, pero en base a la cercanía de los vectores SV con clase contraria y en base a los SV de la misma clase más cercanos. La figura 4 muestra cómo son generados los nuevos puntos a partir de los SV mediante dos técnicas, buscando los vectores cercanos más cercanos de diferente clase (figura 4b) y buscando los vectores soporte más cercanos de la misma clase (figura 4a).

En los dos casos, una vez encontrados los Sv más cercanos se genera un nuevo dato entre estos (en la figura los puntos marcados en azul 4). El algoritmo guarda en memoria el mejor hiperplano y obtiene esta configuración de SV hasta que se cumple un criterio de paro, en nuestras simulaciones el algoritmo para después de ciertas generaciones en que no se ha mejorado la precisión.

Input
SV_+, SV_-: Conjunto de SV
CT : Criterio de paro, numero de nuevos puntos generados
Output
$Data_{new}(X_{sr}^+, X_{sr}^-)$: Datos generados
Begin
  $Data_{new}(X_{sr}^+, X_{sr}^-) \leftarrow 0$ /* Vaciar vector */
  Obtener SV SV_+, SV_- de H_1
  Calcular distancias entre vectores con distinta clase usando ec.(17) y distancia entre vectores misma clase ec.(18)
  Calcular desplazamiento ν_i empleando ecs. (15) y (16)
repeat
  Obtener nuevos puntos $Data_{new}(X_{sr}^+, X_{sr}^-)$ usando ecs.(21)
while $(size(Data_{new}) < CT)$
  return $Data_{new}(X_{sr}^+, X_{sr}^-)$
End

Algorithm 2: Datos artificiales

5. Resultados experimentales

En esta sección se muestra la técnica de selección de parámetros, normalización de datos y los resultados experimentales, obtenidos con el sistema propuesto.

5.1 Conjunto de datos

En los experimentos realizados, las imágenes fueron tomadas de 40 diferentes voluntarios, obtuvimos 20 cuadros por segundo de los videos tomados, con un tamaño de 640 X 480 pixeles en color RGB y formato JEPG. Un total de 15.015 imágenes fueron obtenidas, alrededor de 500 imágenes para cada seña de 27 sílabas algunas de las cuales son mostradas en la figura 5. Cada seña realizada por diferentes personas, esto es debido a que cada persona describe una sílaba con ligeras variaciones; además de ello, existe una variación cuando estas son descritas por personas zurdas, como describe la figura 1.

El conjunto de datos inicial con 15.015 imágenes y resolución de 640 X 480 forma una matriz de entrada de 15.015 X 57. Es decir, cada imagen posee un total de 57 características que permiten definirla. Las características fueron almacenadas en una matriz de *T* de *m x n* que corresponde a una matriz de *m* imágenes con *n* características. Estas últimas fueron normalizadas como:

$$f_{ij} = \frac{T_{ij} - \mu_j}{\sigma_j}$$

Donde *i=1,…, m* y *j=1,…,n* y μj y ${}_{oj}T_{ij} = \mu j$ representan la media y la desviación estándar de *j-ésima* característica del *i-ésimo* vector, *m* es el número de imágenes y *n* es el número de características. Las características normalizadas tienen media cero y desviación estándar igual a 1.

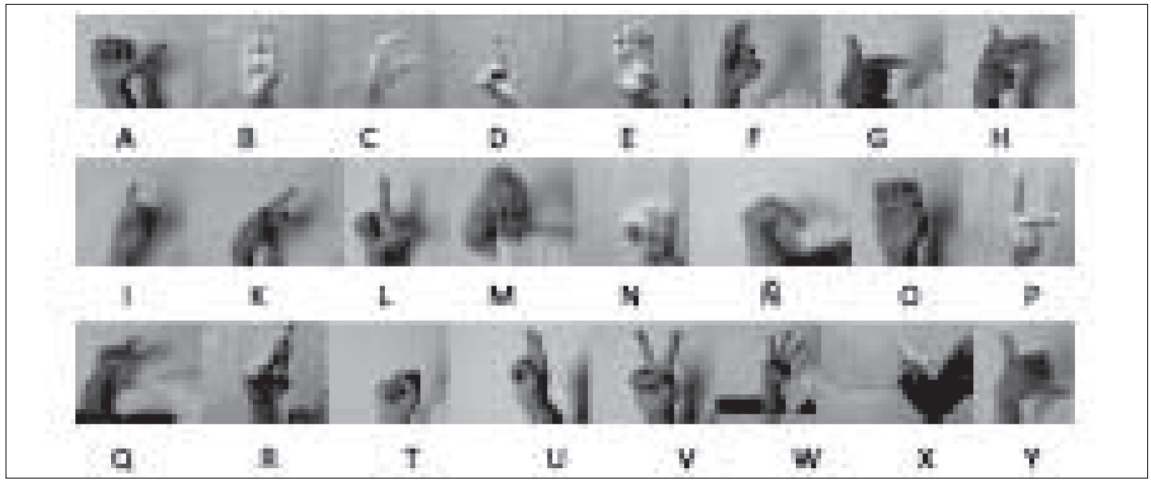
5.2 Selección de parámetros

Seleccionar un conjunto de parámetros para el clasificador empleado es de gran importancia, ya que una buena selección de parámetros tiene un efecto considerable en el desempeño del clasificador. En todos los experimentos realizados empleamos funciones radiales base como kernel, que es definida como:

$$K(x_i - x_j) = \exp\left(-\gamma \left\|x_i - x_j\right\|^2\right), \gamma > 0$$

para obtener los parámetros óptimos se utilizó validación cruzada y búsqueda de malla. En los experimentos todos los conjuntos de datos fueron normalizados y se utilizó validación cruzada con k = 10; 30 corridas fueron realizadas y el promedio de estas es el que se reporta. Para crear el conjunto de datos de entrenamiento y prueba, se seleccionaron aleatoriamente 80% y 20% de los elementos de cada conjunto de datos de entrada. La métrica usada para evaluar la SVM es el número de ejemplos correctamente cla-

Figura 5. Señas del LSM



Fuente: elaboración propia.

Tabla 1. Desempeño de la SVM al clasificar señas del LSM

Clases 1-14					Clases 15-28				
TP_SVM	*TP_P*	AUC_SVM	*AUC_P*	*Clase*	*TP_SVM*	TP_P	*AUC_SVM*	*AUC_P*	*Clase*
0.928	0.962	0.963	0.97	1	0.878	0.92	0.937	0.94	15
0.789	0.85	0.9	0.92	2	0.857	0.88	0.925	0.94	16
0.804	0.83	0.906	0.92	3	0.87	0.89	0.932	0.94	17
0.791	0.84	0.907	0.91	4	0.788	0.83	0.889	0.90	18
0.78	0.82	0.887	0.90	5	0.752	0.78	0.871	0.89	19
0.862	0.90	0.929	0.93	6	0.752	0.80	0.87	0.90	20
0.814	0.85	0.904	0.92	7	0.914	0.93	0.885	0.90	21
0.822	0.86	0.915	0.93	8	0.859	0.88	0.926	0.93	22
0.768	0.82	0.88	0.92	9	0.887	0.92	0.942	0.96	23
0.615	0.70	0.803	0.86	10	0.919	0.94	0.958	0.97	24
0.735	0.76	0.861	0.89	11	0.968	0.97	0.932	0.94	25
0.824	0.85	0.909	0.92	12	0.918	0.95	0.958	0.96	26
0.804	0.84	0.897	0.90	13	0.765	0.79	0.879	0.90	27
0.829	0.86	0.911	0.91	14					

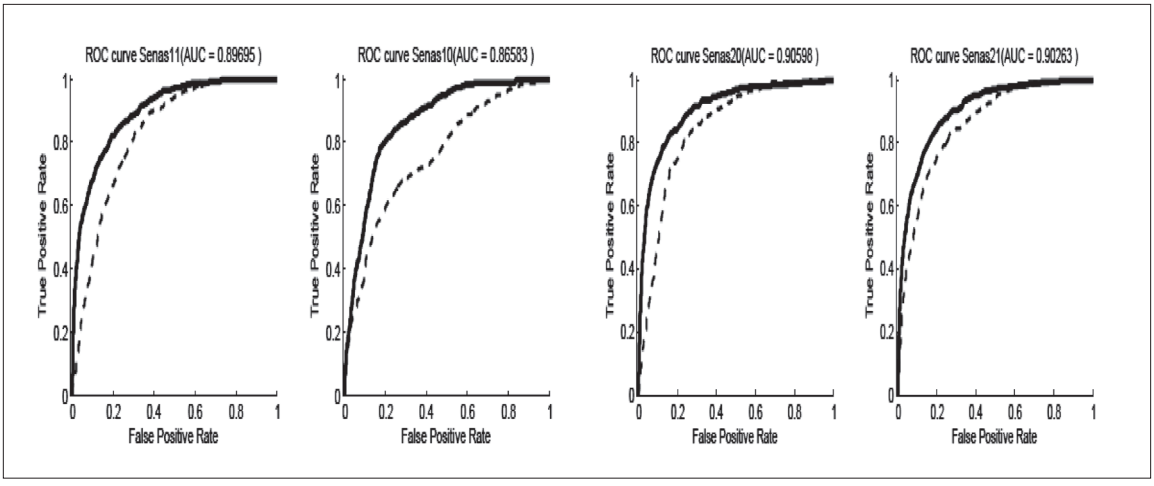
Fuente: elaboración propia.

sificados sobre el número total de ejemplos y el área bajo la curva ROC.

5.3 Resultados

En las simulaciones entrenamos las SVM con 12.012 datos y probamos el clasificador con las 3003 imágenes restantes. En la tabla 1 se muestran los resultados obtenidos, TP_SV M representa el porcentaje de verdaderos positivos encontrados con la SVM, TP_P representa el porcentaje de verdaderos positivos obtenido con el método propuesto, AUC_SV M y AUC_P representan el área bajo la curva obtenida con la SVM y con el método propuesto respectivamente. Los resultados fueron obtenidos al clasificar una cada clase contra el resto. La media de TP obtenida solo con SVM es de 0,8256 y el área bajo la curva es de 0,9065, mientras que la precisión general del clasificador propuesto es de 86 %, mientras que el área bajo la curva con el método propuesto es de 0,9207. Las dos silabas mejor clasificadas fueron la silaba "a" y la sílaba "w".

Figura 6. Curva ROC de vocales a, b, c y d



Fuente: elaboración propia.

El hiperplano de separación obtenido por la SVM en esta fase de entrenamiento es el optimo para el conjunto de datos con el que se entreno, sin embargo, algunos autores [2, 11, 17, 26] han mencionado que al probar el modelo obtenido en personas sordas de las que no se obtuvieron imágenes; la precisión cae considerablemente, esto debido en gran parte a que datos muy representativos no fueron aprendidos durante la fase de entrenamiento. En los resultados obtenidos es posible apreciar que al introducir datos muy parecidos a los SV es posible disminuir la incertidumbre de las regiones vitales y mejorar la habilidad discriminativa del modelo resultante. La figura 6 muestra el área bajo la curva obtenida para las señas 10, 11, 20, 21. En rojo se muestra el área bajo la curva obtenida con las SVM y la línea negra muestra la obtenida con el método propuesto. En todos los casos, el método propuesto mejora considerablemente el área bajo la curva y la sensitividad del clasificador.

7. Reconocimientos

Los autores agradecen al Programa para el Mejoramiento del Profesorado (PROMEP) por el apoyo al Cuerpo Académico UAEM-CA-259 para su consolidación con el proyecto de investigación (103:5=12=4586). Y al Consejo Mexiquense de Ciencia y Tecnología (COMECYT), por el apoyo complementario para extender y llevar a cabo estancias de investigación.

8. Referencias

[1] M. Al-Roussan, K. Assaleh and A. Talaa, "Video-based signer-independent Arabic sign language recognition using Hidden Markov Models". *Appl. Soft Comput,* vol. 9, num. 3. 2009.

[2] M. Assa, K. Grobel, "Video-based sign language recognition using Hidden Markov Models". In: *Proc. Gesture Workshop*, pp. 97-109. 1997.

[3] H. Baird, State of the art of document image degradation modeling. In: Proc. 4th IAPR Workshop on Document Analysis Systems, pp. 1-16. 2000.

[4] B. Bauer, H. Hienz, Relevant features for video-based continuous sign language recognition. In: FG'00. IEEE Computer Society, Washington, DC, USA, pp. 440-450. 2000.

[5] R. Cole, *et al.*, New tools for interactive speech and language training: using animated conversational agents in the classrooms of profoundly deaf children. In: Proc. ESCA/SOCRATES Workshop on Method and Tool Innovations for Speech Science Education, London, pp. 45-52. 1999.

[6] Cole, R., Van Vuuren, S., Pellom, B., Hacioglu, K., Ma, J., Movellan, J., Schwartz, S., Wade-Stein, D., Ward, W., Yan, J., 2003. Perceptive animated interfaces: rst steps toward a new paradigm for human computer interaction. IEEE Transactions on Multimedia: Special Issue on Human Computer Interaction 91 (9), 1391-1405.

[7] Y. Cui, J. Weng, Appearance-based hand sign recognition from intensity image sequences, Comput. Vis. Image Understand. 78 (2) (2000) 157-176.

[8] S. C. Chen, D. Q. Zhang, Z. H. Zhou, "Enhanced (PC)2A for face recognition with one training image per person". *Pattern Recognition Lett.*, vol. 25, num. 10, 1173-1181. 2004.

[9] P. Dreuw, P. Appearance-Based Gesture Recognition. Diploma Thesis, RWTH Aachen University, Aachen, Germany. 2005.

[10] P. Dreuw, D. Stein, T. Deselaers, D. Rybach, D. Zahedi, J. Bungeroth, H. Ney, "Spoken language processing techniques for sign language recognition and translation". In: Technology and Disability, vol. 20, Amsterdam. 2008.

[11] G. L. Fang, W. Gao, D. B. Zhao, "Large vocabulary sign language recognition based on fuzzy decision trees". *IEEE*

Trans. Systems Man Cybernet, vol. 34, núm. 3, 305-314. 2004.

[12] W. Gao, S. G. Shan, X. G. Chai, "Virtual face image generation for illumination and pose insensitive face recognition". In: Internat. Conf. on Acoustic, Speech and Signal Processing, pp. 776-779. 2003.

[13] N. Habili and A. Moini, "Segmentation of face and hands in sign language video sequences using color and motion cues". *IEEE Trans Circ Syst Video Technol*: 1086-1097. 2004.

[14] J. Henk Haarmann, A. C. Katherine, D. S. Ruchkin, "Short-term semantic retention during on-line sentence comprehension. Brain potential evidence from llergap constructions", *Cognitive Brain Research*, vol. 15, núm. 2: 178-190. 2003.

[15] M. A. Hussain, Automatic recognition of sign language gestures, Master Thesis, Jordan University of Science and Technology, Irbid, 1999.

[16] Feng Jiang, Wen Gao, Hongxun Yao, Debin Zhao and Xilin Chen, Synthetic data generation technique in Signer-independent sign language recognition, Pattern Recognition Letters 30: 513-524. 2009.

[17] M. W. Kadous, Machine recognition of Auslan signs using PowerGloves: Towards large-lexicon recognition of sign language. In: Proc. Workshop Integration of Gestures in Language and Speech, pp. 165-174. 1996.

[18] I. Kononenko, "Machine learning for medical diagnosis: history, state of the art and perspective", *Artificial Intelligence in Medicine*, vol. 23, pp. 89-109, 2001.

[19] X. Lu, Jain, A. K., Ber. resampling for face recognition chem. In: 4th Internat. Conf. on Audio and Video based Biometric Person Authentication, pp. 869-877. 2003.

[20] D. MacKenzie, "Electronic device allows the blind to see", *New Scientist*, vol. 208, núm. 2785, 6 November 2010: 10.

[21] Farid Melgani and Yakoub Bazi, Classi cation of Electrocardiogram Signals With Support Vector Machines and Particle Swarm Optimization, IEEE Transactions on Information Technology in Biomedicine, Vol. 12, No. 5, September 2008

[22] Y. Nam and K. Whon. "Recognition of space^e"time hand-gestures using Hidden Markov Model". In: ACM symposium on virtual reality software and technology Hong Kong; p. 51-58. 1996.

[23] R. San-Segundo, R. Barra, R. Córdoba, L. F. D'Haro, F. Fernandez, J. Ferreiros, J. M. Lucas, J. Macias-Guarasa, J. M. Montero, J. M. Pardo, Speech to sign language translation system for Spanish. Speech Commun. 50, 1009-1020. 2008.

[24] R. San-Segundo, J.M. Pardo, J. Ferreiros, V. Sama, R. Barra-Chicote, J.M. Lucas, D. Sánchez, A. García, "Spoken Spanish generation from sign language", *Interacting with Computers*, vol. 22, núm. 2: 123-139. 2010.

[25] N. Tanibata, N. Shimada, Y. Shirai, "Extraction of hand features for recognition of sign language words". In *Internat. Conf. on Vision Interface:* 391-398. 2002.

[26] P. Vamplew and A. Adams, "Recognition of sign language gestures using neural networks". *Australian J. Intell. Information Process. Systems,* vol. 5, num. 2: 94-102. 1998.

[27] M. Zahedi, P. Dreuw, D. Rybach, T. Deselaers and H. Ney, Using geometric features to improve continuous appearance-based sign language recognition. In: British Machine Vision Conference (BMVC), Vol. 3, Edinburgh, UK. 2006.