Reconocimiento del Alfabeto Dactilológico Boliviano Bajo Tecnología de Visión por Computador

Evert I. Pocoma Copa
evertlsmael@gmail.com

Jorge A. Nava Amador
jorgeantonio.navaamador@gmail.com

Abstract

En este documento se propone un sistema, en base a técnicas de Visión por Computador, que permita el reconocimiento del Alfabeto Dactilológico Boliviano, para facilitar, en cierta medida, la comunicación de personas con dificultad auditiva para con el resto de la población boliviana. Para tal fin, se combinaron ideas y aportes de diversos autores, para definir el procedimiento del que se constituye el sistema, mismo que se representa mediante tres componentes: Adquisición de Datos, Entrenamiento, e Intérprete. Los datos adquiridos mediante un sensor Kinect se combinan para reducir la dimensión de la imagen a tratarse, la cual contiene la forma de la mano, misma que está asociada a una letra del Alfabeto Dactilológico Boliviano, de donde se excluye del conjunto de estudio a las letras "LL", "J", "Ñ", y "Z". Los datos son recogidos y almacenados en una Base de Datos de entrenamiento previa segmentación por intensidad, mediante el algoritmo de Otsu. Los datos de entrenamiento son empleados para el Entrenamiento de las máquinas de aprendizaje, previa extracción de características a través del descriptor "Histograms of Oriented Gradients" (HOG), para el cual se realizan algunas variaciones al método original, con objeto de conseguir un vector de características de longitud pequeña. Las máquinas de aprendizaje empleadas son las "Support Vector Machine" (SVM), junto con el kernel de transformación "Radial Basis Function" (RBF). Finalmente se constituye una base de conocimientos de 24 SVM, una por cada letra del Alfabeto Dactilológico Boliviano. Finalmente se presentan los resultados y un análisis de los mismos, obtenidos a partir de un prototipo que recoge las funcionalidades mínimas del sistema.

Keywords* – Segmentación por umbral de intensidad - Otsu, Histogram of Orientes Gradients - HOG, Support Vector Machine - SVM, Radial Basis Function - RBF.

I. Introducción

En Bolivia, el número de personas con diferentes discapacidades, hasta el año 2013 ascendía aproximadamente a 45,600 personas, de las cuales el 15 % corresponde a personas con dificultades sensoriales; según datos proporcionados por el Instituto Nacional de Salud Ocupacional (INSO).

Este grupo de personas presentan varios inconvenientes en su diario vivir, desde el trasladarse de un lugar a otro, hasta el simple hecho de comunicarse con otra persona. Este último conflicto relacionado a las personas sordas conforma parte del escenario sobre el cual se enfoca el presente documento. En Bolivia la temática de la discapacidad, en el contexto Social, con el transcurso del tiempo ha evolucionado, desde un servicio meramente asistencialista, de reclusión, exclusión y abandono, en algunos casos hasta llegar a la extrema pobreza; actualmente esta temática ha logrado ser el marco de políticas y acciones que van en beneficio de esta población, como por ejemplo lo es el Decreto Supremo N° 238, que dispone, se incorpore el Lenguaje de Señas Boliviano (LSB) en el sistema educativo del Estado Plurinacional de Bolivia, y en medios de comunicación audiovisual.

Es evidente la dificultad que la población sorda afronta al tratar de comunicarse con personas que se expresan regularmente mediante lenguaje oral, y por si fuera poco, se tiene que no existe, o no se emplea, un lenguaje de señas universal y único, sino que éste varía de región en región, inclusive de país en país.

Por consiguiente, la propuesta del presente trabajo es un sistema, basado en técnicas de Visión por Computador, que permite el reconocimiento de las letras del Alfabeto Dactilológico Boliviano, para facilitar, en cierta medida, la comunicación de personas con dificultad auditiva para con el resto de la población boliviana. Por lo que, se combinaron ideas y aportes de diversos autores, para definir el procedimiento del que se constituye el sistema, mismo que se representa mediante componentes: Adquisición de Datos, Entrenamiento e Intérprete; En el

desarrollo del presente documento se detallan los algoritmos y técnicas implementadas en cada una de estas etapas.

Los datos adquiridos mediante un sensor Kinect se combinan para reducir la dimensión de la imagen a tratarse, la cual contiene la forma de la mano, misma que está asociada a una letra del Alfabeto Dactilológico Boliviano, de donde se excluye del conjunto de estudio a las letras "LL", "J", "Ñ", y "Z". Los datos son recogidos y almacenados en una Base de Datos de entrenamiento previa segmentación por intensidad, mediante el algoritmo de Otsu [1]. Los datos de entrenamiento son empleados para el entrenamiento de las máquinas de aprendizaje, previa extracción de características a través del descriptor "*Histogram of Oriented Gradients*" (HOG)[2], para el cual se realizan algunas variaciones al método originals, con objeto de conseguir un vector de características de longitud pequeña. Las máquinas de aprendizaje empleadas son las "*Support Vector Machine*" (SVM) [3], junto con el *kernel* de transformación "*Radial Basis Function*" (RBF)[4]. Finalmente se constituye una base de conocimientos de 24 SVM, una por cada letra del Alfabeto Dactilológico Boliviano, excluyendo las letras ya mencionadas anteriormente. Finalmente se presentan los resultados y un análisis de los mismos, obtenidos a partir de un prototipo que recoge las funcionalidades mínimas del sistema.

II. Planteamiento del problema

Si bien según las normas y leyes del Estado Plurinacional de Bolivia, como ser la Ley 045 – "Ley contra el racismo y toda forma de discriminación", la que establece un trato igualitario entre personas sin importar raza, procedencia o dificultades físicas; éstas últimas sufren de un cierto aislamiento debido a que la mayor parte de la población no conoce el lenguaje de señas. Lo cual dificulta la comunicación de las personas con discapacidades auditivas con el resto de la población. En algunos casos la comunicación que se puede establecer entre una persona con alguna de estas dificultades, es parcial, como en el caso de las personas mudas las cuales pueden oír, estableciendo una comunicación en un solo sentido, pero se dificulta la comunicación entre personas con dificultad en el habla y/o en el sentido del oído y la población que regularmente utilizan el lenguaje oral.

Por consiguiente, se reconocen los siguientes aspectos en el presente contexto que se abordar en cuanto a la problemática de imposibilidad de comunicación interpersonal:

- Existe una falta de entendimiento, en cuanto a comunicación, entre las personas que se expresan mediante el lenguaje de señas y el resto de la población.
- No se conoce de la existencia en Bolivia de sistema computacional que facilite la comunicación entre personas con dificultades en el habla y/o discapacidades auditivas. Por si fuera poco, se cuentan con muy pocos trabajos desarrollados en cuanto a la implementación de algún sistema de reconocimiento del Lenguaje de Señas que aborde el reconocimiento de letras del alfabeto o el reconocimiento de palabras, más aun, alguno referente al reconocimiento del Lenguaje de Señas Boliviano

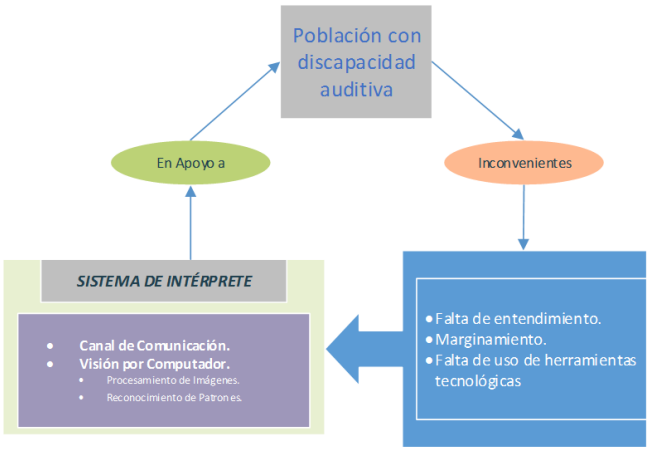

Figura 1: Problemática abordada.

Si bien esta problemática no es reciente, la aplicación de mecanismos tecnológicos para buscar soluciones que mejoren la comunicación ha sido trabajada en los últimos años. En el campo de la Visión por Computador se han logrado desarrollar descriptores mucho más complejos, que si bien modelan muy bien el objeto de interés, lo hacen en desmedro del tiempo de procesamiento o coste computacional. En el campo del Reconocimiento de Patrones se han alcanzado varias mejoras en los métodos de entrenamiento y evaluación de máquinas de aprendizaje.

III. Alcance Propuesto

Se han planteado los siguientes objetivos específicos para la elaboración del trabajo.

1. Constituir una base de datos de entrenamiento representativa, cuya información sea suficiente para un entrenamiento satisfactorio de las SVM; dichos datos, conformados por imágenes, son restringidos a una naturaleza bidimensional, sin apoyarse en la trama de profundidad que pudiera adquirirse mediante el sensor kinect.
2. Analizar y describir las técnicas relacionadas al Procesamiento de Imágenes y el Reconocimiento de Patrones, que se emplean en cada uno de los componentes o etapas del sistema. Se pone en claro que no se realizan comparaciones entre diferentes técnicas, que pudieran llegar a la misma solución, pero de diferente manera; más aún el presente trabajo se limita a realizar un análisis del desempeño en el

Figura 2: Alfabeto Dactilológico Boliviano

empleo de las técnicas, con algunas consideraciones, aplicadas al Reconocimiento del Lenguaje de Señas Boliviano.

3. Elegir los parámetros finales de las SVM, en base a un universo de decisión (superficies de decisión),en base a indicadores de desempeño propias de las máquinas de aprendizaje, con más interés en el índice de generalización y el error de clasificación.

4. Implementar un prototipo de la solución, haciendo uso de un sensor kinect, MatLab, etc. Se aclara que el ambiente para el cual se realiza la adquisición y análisis de los datos, es de naturaleza controlada, manipulando principalmente el fondo y la iluminación del ambiente.

Para el tratamiento de estos objetivos específicos y tal como se mencionó en el apartado introductorio, en este trabajo se toma como caso de estudio al Alfabeto Dactilológico Boliviano, figura , del cual se excluyen las letras "LL", "J", "Ñ" y "Z", ya que estas letras involucran la ejecución de un movimiento.

IV. Descripción del Sistema

La arquitectura del sistema se constituye de tres componentes las cuales son: Adquisición de Datos, Entrenamiento, Intérprete. Cada uno de estos componentes hace uso de técnicas, aportados por otros autores, aplicados en otros campos o aplicaciones similares a la de este documento. A continuación se muestra un esquema general del sistema.

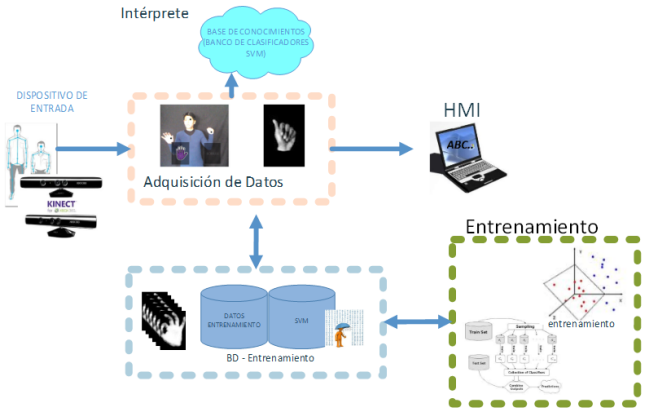

Figura 3: Esquema del sistema

La arquitectura lógica general propuesta para el Sistema, mostrada en la siguiente ilustración (Figura 4), viene dispuesta en capas de servicios, según su naturaleza, y estratificada en cada capa en bloques, según su funcionalidad.

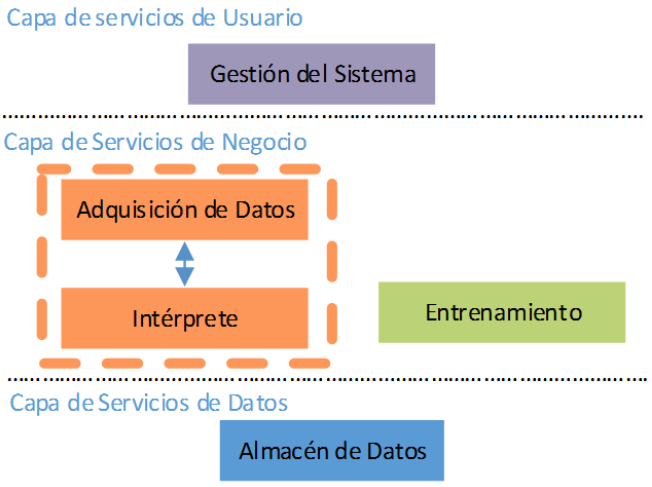

Figura 4: Diagrama en Bloques del sistema

A continuación se detalla cada uno de los componentes de los que se conforma el sistema propuesto.

V. Adquisición de datos

El componente de adquisición de datos es tiene interacción directa con el sensor Kinect, para lo cual, en la implementación del prototipo, este componente fue programado en Visual Studio 2012, haciendo uso de librerías descritas en el texto: "*Programming with the Kinect for Windows Software Development Kit*"[5].

A. Consideraciones del Ambiente

El ambiente para el cual se desarrolló el prototipo del sistema es de naturaleza controlada, es decir, que se especificaron algunas condiciones que el ambiente y el sujeto

de prueba deben cumplir antes de tomar los datos, o de iniciar un reconocimiento de nuevos datos; como por ejemplo:

- Fondo y ropa del sujeto de prueba, deben ser negro y oscura respectivamente, con objeto de maximizar el desempeño del algoritmo de segmentación. En cuanto a la ropa del sujeto de prueba, esta debe ser oscura y uniforme al menos en las proximidades de la mano derecha.
- El sensor Kinect debe estar ubicado a una distancia de entre 1.20 y 1.50 metros del usuario, y a la altura del pecho del usuario; con la finalidad de asegurar un buen funcionamiento del sensor, y llevar a cabo un buen reconocimiento.
- La iluminación se realiza mediante dos reflectores ubicados a nivel del suelo a una distancia igual a la del sensor Kinect, separados por 1.50 metros.

B. *Responsabilidades*

El componente de Adquisición de Datos debe encargarse de las siguientes responsabilidades:

- Obtención y Retención temporal de las Tramas provenientes del sensor Kinect. Se emplean solo las tramas color y *Skeleton*, con una resolución para la trama de color de 640x480 pixeles, y una latencia de arribo de 30 tramas por segundo.
- Reducción de la dimensión de la imagen válida. Se obtiene la posición de la mano derecha, a partir de la trama *Skeleton*, y se recorta una imagen de 64x96 pixeles, tomando como centro del rectángulo la posición de la mano derecha.
- Procesamiento de la imagen. Para esta imagen de menor dimensión se aplica una transformación a escala de grises, incluyendo una maximización de contraste, y finalmente aplicando la segmentación por el método de Otsu, para eliminar el fondo.

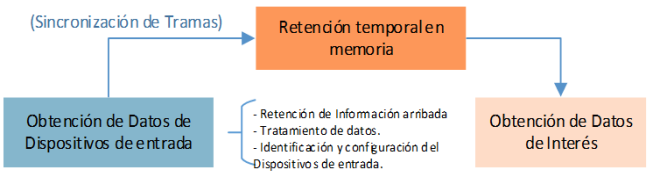

Figura 5: Componente de Adquisición de Datos

C. *Base de Datos de Entrenamiento*

Se conforma estrictamente de imágenes segmentadas y etiquetadas con la respectiva letra a la que representan, tomando 180 imágenes por letra; lo que conforma, para las 24 letras del conjunto de estudio, una Base de Datos de Entrenamiento de 4320 imágenes de 64x96 pixeles.

D. *Segmentación por umbral de intensidad - Otsu*

Este método emplea el histograma de intensidades determinando el umbral que mejor separa las dos clases. Para lo cual emplea técnicas estadísticas, maximizando la varianza entre clases mediante una búsqueda exhaustiva. A continuación se presenta solo un resumen de los aspectos más importantes de los que se comprende este método.

La probabilidad de ocurrencia pi del nivel de gris i en la imagen con N pixeles y L posibles diferentes niveles de grises posibles se denota por:

$$p_i = \frac{f_i}{N} \tag{1}$$

Para el caso de segmentación de dos clases ($C1$ y $C2$), donde las distribuciones de probabilidad de ambas clases son:

$$C_1 = \frac{p_1}{\omega_1(t)}, ..., \frac{p_t}{\omega_1(t)}; C_2 = \frac{p_{t+1}}{\omega_2(t)}, \frac{p_{t+2}}{\omega_2(t)}, ..., \frac{p_L}{\omega_2(t)} \tag{2}$$

Donde: $\omega_1(t) = \sum_{i=1}^{t} p_i; \quad \omega_2(t) = \sum_{i=t+1}^{L} p_i$

Considerando que los valores medios de cada clase como:

$$\mu_1 = \sum_{i=1}^{t} \frac{i \cdot p_i}{\omega_1(t)}; \qquad \mu_2 = \sum_{i=t+1}^{L} \frac{i \cdot p_i}{\omega_2(t)} \tag{3}$$

Y teniendo en cuenta que:

$$\omega_1 \cdot \mu_1 + \omega_2 \cdot \mu_2 = \mu_T; \qquad \omega_1 + \omega_2 = 1 \tag{4}$$

Para el método de Otsu se define la varianza entre dos clases como.

$$\sigma_B^2 = \omega_1 \cdot (\mu_1 - \mu_T)^2 + \omega_2 \cdot (\mu_2 - \mu_T)^2 \tag{5}$$

Por lo que el umbral de Otsu es aquel que maximice la varianza entre dos clases, es decir.

$$T = Max\sigma_B^2(t); \qquad para\ 1 \leq T \leq L \tag{6}$$

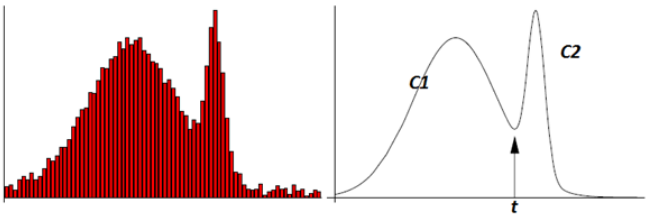

Figura 6: Histograma de intensidades bimodal – "t" umbral óptimo

VI. Entrenamiento

El componente de entrenamiento es donde se realiza la mayor parte del procesamiento de los datos, con el objetivo de elegir los mejores parámetros de los clasificadores

SVM. Por tal motivo, este componente fue implementado en MatLab, haciendo uso del *toolbox de machine learning*.

A. Responsabilidades

El componente de Entrenamiento es el que entrena los clasificadores SVM, y estima su desempeño a partir de índices estimados, este componente debe encargarse de las siguientes responsabilidades:

- Generación del banco de características. Las imágenes de la Base de Datos de entrenamiento son llevadas a un espacio de características a través del descriptor HOG, para el cual se eligen los parámetros adecuados en función del resultado estimado para dichos parámetros.
- Generación Universo de Decisión. Dicho universo de decisión es representado mediante superficies, las cuales muestran la variación del error de clasificación y generalización en función de los parámetros del descriptor y el *kernel* de transformación.

Una vez terminado el entrenamiento, y elegidos los parámetros finales, toda la información referente a la base de conocimientos, conformada por 24 clasificadores SVM, son almacenados en la Base de Datos.

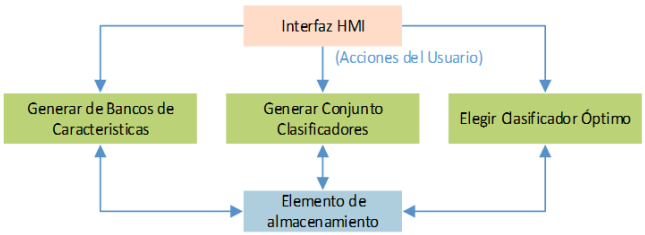

Figura 7: Componente de Entrenamiento

B. Extracción de características - HOG

El descriptor HOG fue desarrollado para la detección de peatones, pero las características y ventajas que ofrece ha hecho que este descriptor sea usado en otras aplicaciones como en el reconocimiento de manos, abordados en los trabajos [6] y [7].

A continuación se mencionan los aspectos que se consideran más relevantes del descriptor HOG dentro del contexto de este documento, así mismo se mencionan consideraciones acerca del empleo de este descriptor como por ejemplo la ausencia de solapamiento.

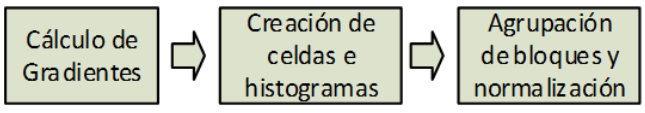

Figura 8: Proceso de extracción de características

Los pasos que se necesitan para la extracción de las características de una imagen mediante el descriptor HOG, son: Cálculo de los gradientes, creación de celdas e histogramas de orientación, creación de bloques y normalización. Estos son pasos son explicados a continuación así como las variaciones propuestas para alguno de los pasos correspondientes:

1) Cálculo de los gradientes

Se extraen los bordes de la imagen, mediante convolución para extraer los gradientes, según la ecuación (7), obteniéndose una matriz en el sentido de X, y otro en el sentido de Y. Ambas matrices se combinan para calcular la dirección del gradiente y su magnitud, mediante la ecuación (8). Gradientes:

$$G_Y = Im \cdot [1, 0, -1]; G_X = Im \cdot [1, 0, -1]^{'} \quad (7)$$

$$Dirección = atan\left(\frac{G_Y}{G_X}\right); Magnitud = \sqrt{G_Y^2 + G_X^2} \quad (8)$$

2) Creación de celdas e histogramas de orientación

La imagen resultante del paso anterior es dividida en celdas de una longitud determinada, para cada celda se computa un histograma de ángulos o direcciones de los gradientes. Dicho histograma divide el intervalo de [0,180] grados en un número determinado de grupos, denominados *bins*.

3) Creación de bloques y normalización

Los bloques son agrupaciones de celdas, para la cual se admiten solapamientos. Para esta aplicación en particular los bloques están comprendidos de una celda, es decir que no existe solapamiento de bloques, esto con la finalidad de conseguir un vector de características de longitud pequeña. Para cada bloque se realiza una normalización de los datos, mediante la ecuación (9).

$$v^* = \frac{v}{\sqrt{\|v\|_2^2 + \epsilon}} \quad (9)$$

Donde:
"v" es uno de los bloques a normalizarse, "v^*" es el bloque normalizado, y "ϵ" es una constante pequeña cuyo valor no tiene gran impacto en los resultados.

A continuación se muestra la representación gráfica del descriptor HOG, para diferentes valores de la longitud de la celda y cantidad de *bins*.

C. Entrenamiento de las SVM

Las SVM son máquinas de base lineal y solución única, las cuales tienen su base en el aprendizaje estadístico. En términos generales las SVM, tienen como objetivo la búsqueda, a través de todas las superficies en el espacio

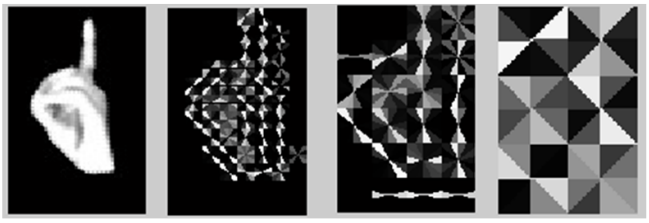

Figura 9: Características HOG para diferentes valores de parámetros–representación gráfica.

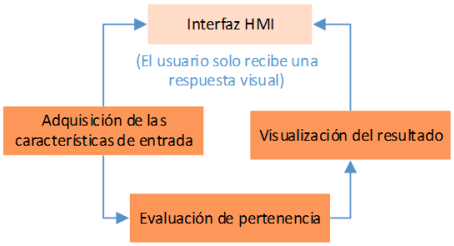

Figura 10: Componente Intérprete

T-dimensional, aquel que separa las dos clases de datos, por el más amplio margen posible. El *kernel* de transformación, proyecta los datos sobre otro espacio de mayor dimensión, calculando el producto interno sin transformar los datos a dicho espacio, con la finalidad de conseguir la separación de los datos a través de un hiperplano. A continuación se mencionan las ecuaciones más importantes de las SVM, sin adentrar en detalle la concepción de las mismas.

El vector de características se denota por x, la función de decisión se define por $f(x)$, el hiperplano de decisión es la frontera de la función de decisión $f(x) = 0$, y la distancia del alguno de los datos al hiperplano se denota por $d(x)$, todas ellas se expresan matemáticamente por:

$$f(x) = w^T x + b; \qquad w^T x + b = 0$$
$$d(x) = \frac{\mid w^T x + b}{\|w\|} \tag{10}$$

El *kernel* de transformación empleado en este documento es el *kernel RBF* el cual se define mediante:

$$k(x, x^{'}) = exp(-\frac{\|x - x^{'}\|_2^2}{2\sigma^2}) \tag{11}$$

Para lograr una distancia $d(x)$ máxima, se afronta un problema de minimización de w. A través de los multiplicadores de Lagrange y trasladando el problema a un dual de maximización, se debe resolver:

$$Maximizar\ \alpha : L_D(\alpha) = \sum_{i=1}^{\tau} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{\tau} \alpha_i \alpha_i y_i y_j k(x_i^T x_j) \tag{12}$$

$$Sujeto\ a : \sum_{i=1}^{\tau} y_i \alpha_i = 0;\ 0 \leq \alpha_i \leq C \quad para\ i = 1, \ldots \tau$$

Donde: "C" es denominado el boxConstraint, "α_i" son los multiplicadores de Lagrange, los cuales son diferentes de cero para conjunto reducido de "x_i" los que se denominan *support vectors*.

Las ecuaciones mostradas en (12) representan un problema de programación cuadrática, la cual, para este trabajo, se resuelve mediante el algoritmo *Sequential Minimal Optimization* (SMO). Por simplicidad el entrenamiento cada clasificador se empleó el *toolbox* de *machine learning* de MatLab.

Por consiguiente la función de decisión resulta en:

$$f(x) = \sum_{i=1}^{numSV} \alpha_i y_i k(x_i^T x) + b \tag{13}$$

Como se observa la evaluación de la función de decisión está sujeta únicamente a los *support vectors*, es decir que el hiperplano de separación está definido solo por estos vectores, más aún, el número de estos *support vectors* y el total de los datos de entrenamiento son los que determinan el índice de generalización de la SVM en cuestión.

VII. Intérprete

A. *Responsabilidades*

El componente de Intérprete debe encargarse de las siguientes responsabilidades:

- Extracción de características para nuevos datos. Los nuevos datos que arriban del sensor Kinect se llevan al mismo espacio de características especificado por los parámetros finales del clasificador SVM seleccionado.
- Evaluación de Pertenencia. Con el vector de características conformado, se consulta a la base de conocimientos si dicho vector pertenece a alguna clase (letra).

B. *Toma de decisión*

Para cada uno de los clasificadores SVM finales, se conocen los parámetros del descriptor para el que fueron seleccionados sus parámetros. Los parámetros del descriptor pueden no ser los mismos para todos los clasificadores, por lo que para un clasificador SVM en cuestión, se debe recalcular el vector de características de la nueva imagen válida, antes de evaluar la pertenencia en dicho clasificador SVM. Este procedimiento puede demandar demasiado tiempo de computo, ya que si todos los clasificadores SVM finales tienen parámetros del descriptor diferentes, se debe recalcular el vector de características 24 veces.

Para agilizar la evaluación de la base de conocimientos, se propone la inclusión de un índice de parámetros para el cual se agrupan aquellos clasificadores de parámetros

iguales. Así de esta manera se evita recalcular el vector de características innecesariamente.

Una vez que el componente de Intérprete encuentra una respuesta de pertenencia positiva, suspende la búsqueda y se acepta el resultado de la Base de Conocimientos

VIII. Resultados Obtenidos

Los resultados que se muestran a continuación son resultado del procesamiento de los datos a través de los componentes de Adquisición de datos y Entrenamiento, aplicados para una base de datos de entrenamiento de 4320 imágenes. Para estimar el desempeño de las máquinas de aprendizaje se recolectó además, una Base de Datos de Prueba conformada por 100 imágenes por letra, haciendo un total de 2400 datos de prueba, para los cuales solo se realiza una medición del error de clasificación.

A. Variaciones en los parámetros del descriptor HOG

Para realizar un análisis a fondo del efecto del descriptor HOG empleado, sobre el resultado obtenido, en la aplicación para el reconocimiento del Lenguaje de Señas de Alfabeto Dactilológico Boliviano; se trabajó para diferentes valores de los parámetros del descriptor, variando entre ellos: el número de bins del histograma y la dimensión de las celdas. No se considera solapamiento entre los bloques adyacentes debido a que cada bloque está conformado por una celda, con objeto de lograr un vector de características de longitud pequeña. Debido a las dimensiones de las imágenes de la Base de Datos de Entrenamiento, la longitud de la celda solo puede adoptar los valores de 8, 16 y 32 pixeles.

B. Variaciones en los parámetros de la SVM

En las SVM, se trabajó para diferentes valores en el parámetros, como por ejemplo: el parámetro "σ" del *kernel* empleado, por otro lado, se pudo trabajar para diferentes valores del *box constraint* "C", pero por simplicidad se decidió trabajar para un valor de 15, siendo este el máximo valor que los multiplicaciones de Lagrange pueden adoptar.

C. Universo de decisión

El universo de decisión está conformado por 9 superficies las cuales son representaciones del cómo se comportan tres indicadores de las SVM: error esperado, error de clasificación, índice de generalización.

El error esperado es calculado mediante validación cruzada, es decir que se estima un error para nuevos datos en función de los datos de entrenamiento, el error de clasificación es una cantidad que representa la cantidad de falsos positivos y falsos negativos obtenidos para los datos de la Base de Datos de Prueba, y finalmente, el índice de generalización viene dado por la cantidad de *support vectors* en relación al total de datos de entrenamiento, siendo el índice de generalización más alto cuantos menos *support vectors* existan.

Cada una de estas superficies viene graficada en función de tres parámetros, dos relacionados con el descriptor HOG, los que son: longitud de celda, y número de *bins*; y uno relacionado con el *kernel* de transformación empleado en las SVM, coeficiente sigma del *kernel* RBF.

D. Base de Conocimientos

Para cada uno de los clasificadores que conforman la base de conocimientos se generaron las 9 superficies de decisión, a partir de las cuales se eligieron los parámetros finales del descriptor y el *kernel*. Por consiguiente, la tabla de la figura 12 muestra un resumen de los parámetros finales que se seleccionaron para cada uno de los 24 clasificadores que conforman la Base de Conocimientos.

Letra	Parámetros Finales			Letra	Parámetros Finales		
	Celda	*Bin*	*Sigma*		*Celda*	*Bin*	*Sigma*
A	32	5	5	N	32	7	5
B	32	5	5	Ñ	No implementado		
C	32	5	5	O	32	7	5,35
D	32	5	5	P	32	7	5,35
E	32	6	5	Q	32	5	5
F	32	5	5	R	32	4	5
G	32	8	5,35	S	32	8	5,35
H	32	5	4,7	T	32	8	5,35
I	32	7	3,57	U	32	8	5,35
J	No implementado			V	32	7	5,35
K	32	7	5	W	32	7	5,35
L	32	5	5,35	X	32	8	5,35
LL	No implementado			Y	32	7	5,35
M	32	7	4,7	Z	No implementado		

Figura 12: Parámetros finales de la Base de Conocimientos

Para los parámetros finales de cada uno de los clasificadores que conforman la Base de Conocimientos se puede observar los índices de desempeño. Por lo tanto, en la tabla de la figura 13 se muestra un resumen de dichos índices, donde E_{Esp}, E_{Clas} y Gen son: el error esperado en base a la Base de Datos de Entrenamiento, error de clasificación en base a la Base de Datos de Pruebas, e índice de Generalización; respectivamente.

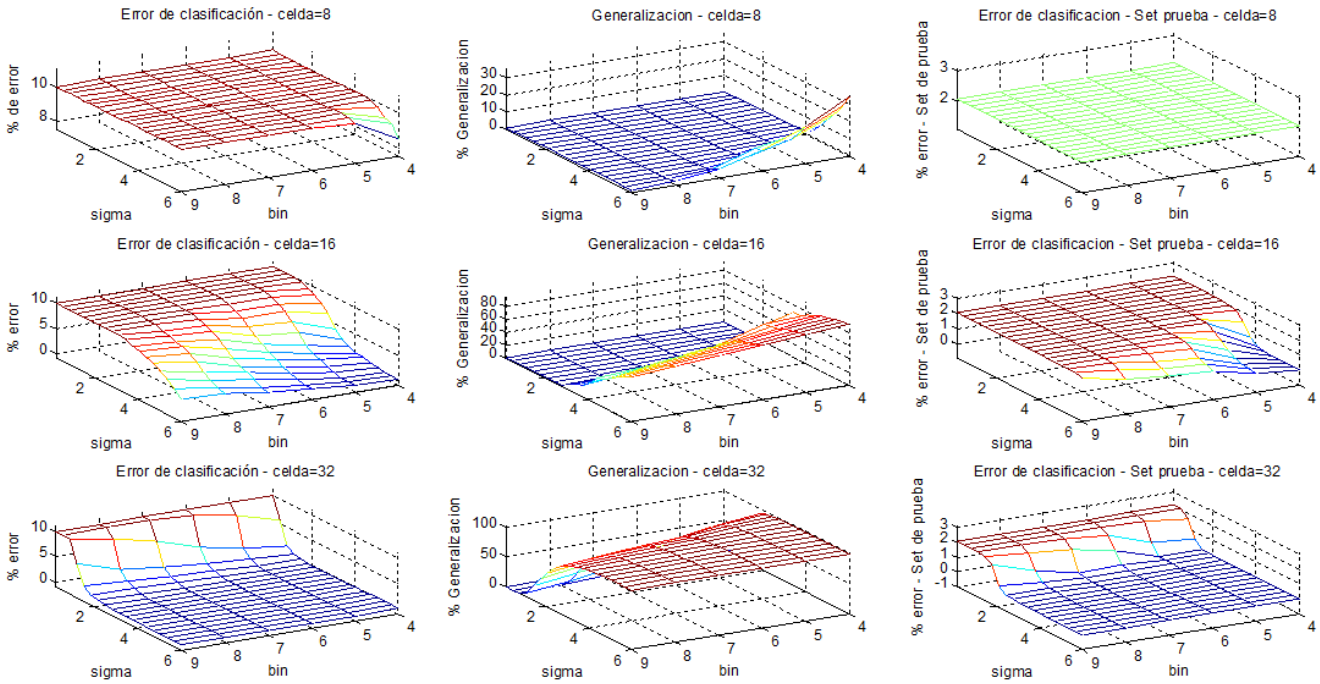

Figura 11: Superficies de Decisión – letra "F"

Letra	Parámetros Finales			Letra	Parámetros Finales		
	E_{Esp}	*Gen*	E_{Clas}		E_{Esp}	*Gen*	E_{Clas}
A	0,1042	98,84	0,2708	**N**	1,146	96,13	0,4167
B	0	99,33	0	**Ñ**	No implementado		
C	0	99,21	0	**O**	0	98,43	0
D	0,06947	99,07	0,1042	**P**	0,1274	98,43	0,0416
E	0,4515	98,15	0,25	**Q**	0,2893	97,69	0
F	0	99,38	0	**R**	0,4051	96,67	0,6947
G	0,2084	97,36	0	**S**	0,544	96,81	0,7083
H	0,2664	98,52	0,02083	**T**	0,1042	97,62	0,2708
I	0	98,13	0,04167	**U**	0,1157	98,59	0,1458
J	No implementado			**V**	0,2779	98,59	0,1458
K	0,0347	98,73	0,0625	**W**	0,02313	99,05	0,04167
L	0	99,28	0	**X**	0,0694	97,78	0,02083
LL	No implementado			**Y**	0,06947	98,29	0,0625
M	1,065	96,18	0,1458	**Z**	No implementado		

Figura 13: Índices de desempeño - Base de Conocimientos

IX. Análisis de Resultados

El análisis de resultados se centra estrictamente en los valores obtenidos para los índices de desempeño, y en la longitud del vector de características, todas ellas en función de la longitud de las celdas, la cantidad de *bins* y el coeficiente *sigma* del *kernel* de transformación RBF.

A. *Longitud del vector de características*

Con la finalidad de obtener un vector de características de longitud pequeña, fue que se omitió el solapamiento entre los bloques del descriptor HOG, y en base a las superficies de decisión, se observó que el mejor desempeño de las SVM fue para los parámetros del descriptor de 32 pixeles de longitud de celda, lo que conlleva a un vector de características mínimamente de 24 componentes, para el caso de 4 divisiones en el histograma (*bins*); y un vector de características de 54 componentes, para el caso de 9 divisiones en el histograma (*bins*).

Si bien este vector de características es de longitud pequeña, fue en desmedro de la robustez ante la iluminación del descriptor HOG. Este inconveniente se solventó señalando precondiciones acerca del Ambiente en el cual se extraen los datos, y sobre el cuál se vaya a implementar el prototipo, es decir que las condiciones de iluminación no deben cambiar o por lo menos variar mínimamente, entre la etapa de toma de datos, procesamiento y prueba del prototipo.

B. *Base de Conocimientos*

El análisis de resultados para la Base de Conocimientos en este documento se lo realiza en base a los promedios en los índices de desempeño que se consideraron para las máquinas de aprendizaje SVM.

1) Error Esperado

Este valor fue calculado mediante el algoritmo de validación cruzada, y es un indicador del error que se espera que obtener para nuevos datos, en función de los datos de entrenamiento. Para la base de conocimientos, en promedio se obtuvo un error esperado menor al 0,5 %.

2) Error de Clasificación

Este valor fue obtenido evaluando la Base de Conocimientos para una Base de Datos de Prueba, dichos datos fueron obtenidos para el mismo ambiente de trabajo, en el cual se adquirieron los datos de entrenamiento. Para la base de conocimientos, en promedio se obtuvo un error de Clasificación de menor al 0,5 %.

3) Indice de generalización

Este valor se encuentra estrechamente relacionado con la cantidad de vectores de soporte, de los que se comprende cada clasificador, y son un indicador del desempeño del clasificador SVM para datos no considerados en los de entrenamiento. Para la base de conocimientos, en promedio se obtuvo un índice de generalización mayor al 98 %.

X. Conclusiones

Se describió un sistema que permite el reconocimiento automático del Alfabeto Dactilológico Boliviano, estructurado sobre técnicas de Visión por Computador, detallando los tres componentes de los que se compone. Para cada componente se describió las técnicas que se emplearon junto con las variaciones en la implementación referentes a las mismas; obteniendo un vector de características de longitud pequeña, bajos errores de clasificación para diferentes bases de datos (de entrenamiento y prueba), y altos índices de generalización.

Para la concepción del sistema intérprete descrito en el presente documento se emplearon técnicas e ideas previamente propuestas y descritas en trabajos relacionados a la temática abordada, por otros autores.

En la implementación del prototipo se emplearon herramientas para facilitar las pruebas y el desarrollo del mismo, como por ejemplo el empleo del *toolbox* de *machine learning* de MatLab, el cual fue empleado en el entrenamiento de los clasificadores SVM.

Referencias

[1] *Segmentación por Umbralización, Método de Otsu.* Ingeniería en Automatización y Control Industrial, Universidad Nacional de Quilmes, 2005 (vid. pág. 2).

[2] N. Dalal y B. Trigss. “Histograms of Oriented Gradients for Human Detection”. Francia: INRIA (vid. pág. 2).

[3] C. Burges. *A Tutorial on Support Vector Machines for Pattern Recognition.* Bell Laboratories, Lucent Technologies, 1998 (vid. pág. 2).

[4] G. Shaklnarocivh. “SVM and Kernels”. Weizman Institute of Science, 2011 (vid. pág. 2).

[5] D. Catuche. *Programming with the Kinect for Windows Software Development Kit.* Microsoft, 2012 (vid. pág. 3).

[6] L. Cheung y C. Medina. “Implementación y Análisis de un Detector de Manos Basado en Visión Artificial”. Panamá: Facultad de Ingeniería Eléctrica, Universidad Tecnológica de Panamá, 2013 (vid. pág. 5).

[7] N. Pugeault H. Cooper E. Ong y R. Bowden. “Sign Language Recognition using Sub - Units”. En: 2012 (vid. pág. 5).

[8] L. Enrique. *Visión Computacional.* Puebla - México.: Instituto Nacional de Astrofísica, Óptica y Electrónica.

[9] M. Gonzalez. “Aprendizaje Estadístico, Redes Neuronales y Support Vector Machines: Un Enfoque Global”. Xalapa, México: Intelligen Transportation Systems Research Group, Universidad Veracruzana, 2005.