



EVALUACIÓN DE LA CALIDAD DE LAS IMÁGENES DE ROSTRO: HOMBROS RECTOS Y PRESENCIA DE SOMBREROS

EVALUATION OF THE FACE IMAGES QUALITY PARAMETERS: ALIGNED SHOULDERS AND HAT PRESENCE

Jesús Pérez-Martín¹, Yenisel Plasencia Calaña^{1,2}, Heydi Méndez-Vázquez¹

1 División de Investigaciones CENATAV, Datys Soluciones Tecnológicas (DATYS), La Habana,
Cuba, CP 12200

2 Faculty of Electrical Engineering, Mathematics and Computer Sciences, Delft University of Tech-
nology, The Netherlands

{jmartin¹, yplasencia^{1,2}, hmendez¹}@cenatav.co.cu

RESUMEN: *La evaluación automática de la calidad de las imágenes de rostros es un tema de gran importancia dentro del desarrollo de sistemas biométricos de reconocimiento de rostros. En este trabajo se describen dos nuevos métodos para evaluar si las imágenes de rostros cumplen con dos de los parámetros establecidos en la norma ISO/IEC 19794-5, que permiten determinar si la persona se encuentra (o no) con los hombros alineados y si lleva (o no) algún tipo de sombrero. Esto contribuye a establecer si la imagen tiene valor identificativo y si puede utilizarse en documentos de identificación personal. Los resultados experimentales muestran que se logra una alta eficacia para la determinación de los parámetros de calidad con los métodos propuestos.*

Palabras Clave: Bordes, agrupamiento, BSM, SVM, Bolsa de palabras visuales, SIFT, SURF, DSIFT, PHOW

ABSTRACT: *The automatic evaluation of face image quality is a very important topic in the development of face recognition systems. This paper presents two new methods to evaluate if face images fulfill two parameters set by the ISO/IEC 19794-5 norm, that allow one to determine if the person have the shoulders aligned and if the person have some type of hat. This contributes to establish if a given image possesses identification value and if it can be used in personal identification documents. The experimental results showed a high accuracy of the proposed methods.*

KeyWords: Edges, clustering, BSM, SVM, bag of visual words, SIFT, SURF, DSIFT, PHOW

1. INTRODUCCIÓN

Producto del gran número de aplicaciones prácticas en las que se necesita el reconocimiento automático de personas, la determinación de la identidad de una

persona a partir de una imagen de su rostro es uno de los problemas en los que se ha centrado el procesamiento digital de imágenes y el reconocimiento de patrones en los últimos años. Por tal motivo, este tema ha sido abordado en la literatura desde varias

dimensiones. El rostro es la característica biométrica más comúnmente utilizada por los seres humanos para identificar a otras personas. Desde hace varias décadas, las imágenes de rostros han sido utilizadas, por ejemplo, para identificar a los individuos en los documentos personales. Los avances tecnológicos han permitido que las imágenes digitales del rostro sean además la base de sistemas biométricos que implementan la verificación e identificación de personas, la video-vigilancia y el análisis de las acciones humanas. Con el aumento de los sistemas biométricos basados en imágenes de rostros, ha surgido la necesidad de definir un formato de datos estándar, que establezca los requisitos indispensables para que las imágenes que se utilizan tengan valor identificativo y que permita la interoperabilidad entre diferentes sistemas.

La ICAO (del inglés International Civil Aviation Organization) establece que las imágenes de rostro deben cumplir con los estándares de la norma ISO/IEC 19794-5 [1], la cual enumera los distintos parámetros de calidad que deben ser verificados, entre los que se encuentra:

- El solicitante estará de frente a la cámara, no mirando sobre el hombro (estilo de retrato artístico), así como tendrá los hombros a la misma altura.
- No se aceptarán sombreros excepto en circunstancias en que la autoridad estatal competente los apruebe específicamente. Tales circunstancias pueden ser religiosas, médicas o culturales.

En este trabajo se propondrán nuevos métodos para la evaluación automática de estos parámetros.

Para el problema de la alineación de los hombros, en la etapa de detección se exponen y comparan dos estrategias diferentes que usan: técnicas de detección de bordes, agrupamiento jerárquico de píxeles mediante distancia mínima, extracción de rasgos BSM (del inglés Blurred Shape Model) y clasificación usando máquinas de vectores de soporte (SVM). De esta manera se obtiene un detector de hombros robusto y eficaz.

Para lograr la detección de sombreros se propone la creación de un método de reconocimiento de patrones basado en un modelo de Bolsa de Palabras Visuales para representar las imágenes. Se realizaron pruebas con descriptores de rasgos SURF (del inglés Speeded Up Robust Features), SIFT (del inglés Scale Invariant Feature Transform), y los novedosos DSIFT (del inglés Dense SIFT) y PHOW (del inglés Pyramid Histogram of visual Words) que brindaron los mejores resultados. Luego se propone el uso de k-Means para construir el diccionario visual y finalmente ejecutar la clasificación mediante máquinas de vectores de soporte usando el kernel de intersec-

ción de histogramas. Este kernel, que hace relativamente poco tiempo se descubrió que cumple las propiedades necesarias para ser usado en el contexto de estos clasificadores, hace el papel de una similitud y es apropiado para tipos de datos como los basados en histogramas.

2. ALINEACION DE LOS HOMBROS

En la solución que se plantea en este trabajo para determinar la alineación de los hombros, primeramente se propone detectar los hombros y luego calcular su alineación. Notar que sobre la alineación de los hombros no existe ninguna regla que establezca cuando se está en presencia de hombros alineados o no, por tanto este criterio se definió en base a la opinión de expertos en reconocimiento facial. Tampoco se encontró en la literatura ningún método que aborde este problema.

El método propuesto consta de cuatro etapas fundamentales:

1. Preprocesamiento de la imagen, donde se aplican un conjunto de transformaciones y técnicas sobre la imagen. Esto se hace con el fin de estandarizar las imágenes, resaltar regiones de interés y detectar los bordes para alcanzar un máximo rendimiento en las siguientes etapas.
2. Agrupamiento de los puntos de bordes de manera que queden curvas alargadas o segmentos de líneas rectas, buscando extraer información semántica que permita luego determinar con buena precisión cuáles verdaderamente forman parte de los hombros.
3. Detección de los hombros, determinar a partir del resultado de la etapa anterior cuáles grupos de píxeles representan verdaderamente a los hombros de la persona retratada.
4. Finalmente, calcular la alineación. La propuesta en este sentido fue resultado de la elaboración de una heurística propia, ya que no existe una regla para esto.

2.1. Preprocesamiento de la imagen

En primer lugar, la imagen se transforma a escala de grises. Puesto que en las imágenes de rostro existen criterios que influyen decisivamente en la capacidad de análisis de cualquier método que las procese, como por ejemplo, la distancia entre la persona y la cámara, la centralización del rostro y la inclinación del mismo, se propone, antes de continuar el proceso, llevar a cabo una normalización de las imágenes del siguiente modo:

1. Escalar la imagen a una altura de 400 píxeles.
2. Intentar detectar el rostro de la persona, así como su inclinación a partir de la posición de

sus ojos, usando un algoritmo que sigue la idea del conocido detector Viola-Jones¹[10].

3. Escalar la imagen a un tamaño para el cual la distancia entre los ojos sea de 60 píxeles.
4. Si la inclinación del rostro no excede los 30 grados, se asume que el rostro no está encima de ninguno de los hombros y se procede entonces al centrado del mismo.
5. Por último se realiza un suavizado de la imagen buscando reducir el ruido o los detalles artificiales producidos al tomar la imagen.

Luego de normalizar la imagen se procede a su segmentación buscando extraer las regiones de interés para el problema, en este caso, dos regiones independientes que contengan a los hombros por separado. A partir de la nueva posición del rostro centrado en la imagen, se determinan dos regiones de igual alto y ancho, una a cada lado del mismo que se extienden desde los bordes laterales de la imagen hasta los bordes del rostro y desde la mitad del rostro hasta el borde inferior de la imagen. En la Figura 1 se muestra el proceso para una imagen de ejemplo. A partir de este punto el procesamiento que se hace a cada una de las dos regiones es análogo.

2.2. Agrupamiento

Debido a que los hombros mayormente quedan representados por una secuencia de píxeles bien definida (un contorno o parte de alguno), los bordes detectados en cada una de las subregiones son procesados y filtrados con el objetivo de que los que pertenecen a los hombros queden separados del resto. Para esto se propone un agrupamiento jerárquico aglomerativo con distancia mínima [11].

2.3. Detección de los hombros

Luego del agrupamiento, se tienen varios grupos en cada región, quedando cada uno de los dos hombros incluidos en alguno de los grupos detectados en la región correspondiente. Sin embargo, en estos pueden existir además píxeles que no pertenecen a los hombros, es necesario entonces limpiar los grupos de estos píxeles. Por tanto, para cada grupo se determina el conjunto de píxeles que está en su parte superior (los píxeles más cercanos al borde superior de la imagen).

Como resultado del paso anterior puede suceder que algunos de los grupos se separen producto de la eliminación de los píxeles que no formaban parte de las fronteras. Debido a esto se decidió realizar nuevamente un agrupamiento pero solo con los píxeles que forman parte de alguna frontera, usando el

mismo límite mínimo de distancia antes mencionado y con similar fin que el agrupamiento anterior.

Los subconjuntos de píxeles obtenidos finalmente son clasificados, con el propósito de escoger de ellos los que verdaderamente representan a los hombros, encontrándonos entonces en un problema de clasificación de patrones de dos clases: hombro y no hombro.

Para llevar a cabo la clasificación se propone explorar las capacidades que tiene SVM como clasificador en problemas de dos clases. Se pretende distinguir entre patrones de una clase o la otra a partir de la *forma* que presentan; ya que los hombros generalmente poseen una forma bien definida a pesar de que en muchos casos su determinación se encuentre afectada por la presencia de detalles del vestuario, el cabello y otros. La estrategia seguida se basa en [2] donde se expone un reconocedor de objetos mediante descriptores de forma.

Para el aprendizaje se entrenan dos modelos SVM (destinados a la clasificación de los grupos de la derecha e izquierda) y se le determina el descriptor BSM (del inglés, Blurred Shape Model) [2] a cada uno de los grupos de puntos (los que representan hombros y los que no) extraídos de un conjunto de imágenes de entrenamiento etiquetadas debidamente.

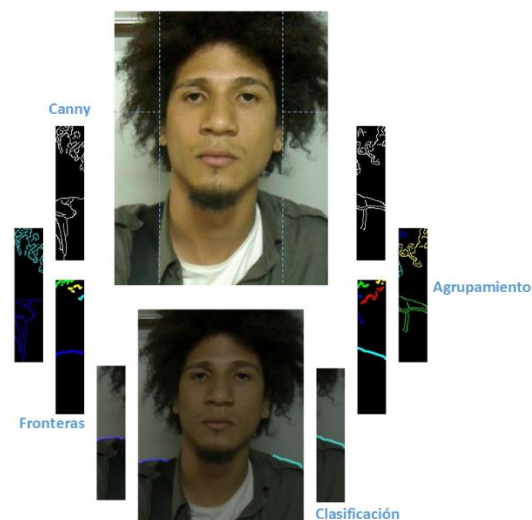


Figura. 1: Etapas del proceso de detección de los hombros: detección de bordes, agrupamiento de los píxeles, selección de fronteras y clasificación entre las categorías hombro y no hombro.

2.4. Cálculo de la alineación de los hombros

Luego de la detección se procede a determinar la alineación de los hombros mediante la diferencia absoluta entre la coordenada vertical de los centroides de

¹ Técnica de detección de rostros desarrollada

por Paul Viola y Michael Jones.

cada uno de los grupos seleccionados como hombros, correspondiente a la diferencia de alturas entre ellos. Mediante el entrenamiento antes mencionado, se determinó que este valor se encuentra en el intervalo [0,30]. En caso de que ninguno de los grupos determinados cumpla con los requisitos mínimos para ser considerado como hombro, esto significa que hubo una mala detección.

3. DETECCIÓN DE SOMBREROS

Para dar solución a este problema, en el presente trabajo se propone la creación de un método de reconocimiento de patrones, basado en un modelo de Bolsa de Palabras Visuales (BoVW) [5], donde cada imagen va a estar representada por un conjunto de vectores en vez de por un solo vector de características como sucede en los enfoques clásicos de representación.

3.1. Modelo de Bolsa de Palabras Visuales

BoVW es actualmente un método popular para el reconocimiento de objetos y escenas en visión por computadoras. A una imagen se le extraen los rasgos locales y pasa a ser considerada como una *bolsa de rasgos (bag of features)*, es decir, ignorando las relaciones espaciales entre ellos. Como desventaja podemos mencionar que este no cuenta con un mecanismo eficiente y efectivo de codificación de la información espacial que existe para los rasgos. Un método basado en el BoVW clásico consiste en las siguientes etapas:

- **Extracción de rasgos:** Los rasgos locales y sus descriptores correspondientes se extraen de regiones de la imagen. Los dos descriptores visuales más usados son SIFT [3] y SURF [4]. Algunos métodos los extraen en ciertos puntos de interés detectados y otros obtienen los rasgos locales densamente, en posiciones regulares de la imagen como es el caso del PHOW [4].
- **Generar un diccionario y mapear los rasgos a palabras visuales:** Un diccionario visual es un método que divide el espacio de descriptores visuales en varias regiones. Los rasgos de una región corresponden a la misma palabra visual. Entonces, una imagen se codifica como un histograma de la frecuencia de ocurrencia de cada palabra visual. Esto se hace asignando a cada vector de rasgos de la imagen su región más cercana, de manera que al terminar el proceso se tenga la cantidad de vectores asignados a cada región y se asigna esa cantidad a la componente correspondiente a esa palabra visual en el histograma.

- **Clasificación:** Varios métodos de aprendizaje por computadora pueden aplicarse para la representación de las imágenes. SVM es frecuentemente usado como clasificador en modelos BoVW para el reconocimiento de objetos y escenas. Este fue el clasificador escogido para resolver el problema planteado, en conjunto con el kernel aditivo de intersección de histogramas debido a su utilidad y buen desempeño para representaciones basadas en histogramas.

3.2. Preprocesamiento

Como el problema se centra solo en imágenes de rostro de fondo uniforme y los sombreros siempre se encuentran en una misma región relativa a las personas, se decidió, con el fin de reducir el área de búsqueda, seleccionar de la imagen la región en la que debe estar el sombrero. Esto se hace convirtiendo la imagen a escala de grises y a partir de la detección del rostro de la persona, se realiza un escalado de manera que sus ojos queden a una distancia de 20 píxeles y finalmente se selecciona una región de la imagen que se extiende desde la mitad del rostro hacia arriba, con un ancho y alto no mayores de 100 píxeles. Si el rostro no fuese detectado, entonces la imagen es escalada a una altura de 200 píxeles. En la figura 2 se muestra el resultado de aplicar este proceso para una imagen de ejemplo.



Figura. 2: Preprocesamiento de la imagen, selección de la región de interés a partir de la detección del rostro y el escalado de la imagen.

Con esto se busca eliminar la influencia que puede tener la distancia a la que fue tomada la imagen y se garantiza lo mejor posible que el sombrero sea segmentado completamente del resto de la imagen.

3.3. Extracción de características

Como parte de la investigación desarrollada en este trabajo, en la búsqueda de la mejor solución al problema de la detección de sombrero se probaron varios de los métodos que se mencionan en la literatura para la extracción de las características de las imágenes y que han mostrado buenos resultados, estos son SIFT, SURF, DSIFT y su variante, PHOW.

3.4. Construcción del diccionario visual

Una vez detectados los puntos claves y extraídas las características con su descriptor, BoVW propone la creación de un diccionario visual, un conjunto de palabras visuales, con el fin de describir posteriormente las imágenes mediante la detección de la ocurrencia de estas palabras en ellas. La calidad del diccionario visual tiene un impacto significativo sobre el éxito de los métodos basados en BoVW. Muchos métodos para la categorización de objetos y escenas emplean métodos de aprendizaje no supervisado (por ejemplo, el agrupamiento k-Means) para obtener dicho diccionario visual, tomando como palabras visuales los centroides obtenidos para cada grupo en este proceso.

Se hicieron pruebas con agrupamiento k-means usando distancia Euclidiana y con agrupamiento jerárquico aglomerativo con distancia media (average linkage, en inglés) usando distancia χ^2 , con el fin de determinar cuál se ajustaba mejor al problema. Un aspecto fundamental en el rendimiento de estos métodos es el tamaño del diccionario, puesto que esto influye en la capacidad representativa y discriminativa de las palabras visuales sobre las clases. A pesar de existir varios estudios sobre este tema, no existe ninguna regla para determinar a priori qué tamaño dará los mejores resultados, y al parecer esto depende del problema que sea abordado. En artículos como [5,6] se logran buenos resultados con una cantidad de palabras en el rango de entre las 100 y las 800.

3.5. Descriptores de las imágenes

Para describir las imágenes se utilizó un esquema de *Emparejamiento Piramidal Espacial (Spatial Pyramid Matching)*, donde se plantea el cálculo de histogramas de frecuencias de las palabras visuales a distintas resoluciones de la imagen. En algunos artículos como [7,8] las diferentes resoluciones se determinan mediante repetidos submuestreos de la imagen y computan un histograma global de los valores de los píxeles para ese nivel, se varía la resolución a la que los rasgos son calculados (valores de los píxeles), pero la resolución del histograma (escala de intensidad) se mantiene fija. En artículos como [5,9], se plantea un enfoque opuesto, fijar la resolución a la que se determinan los rasgos, pero variar la resolución espacial en la que son agregados. En este trabajo se propone un Emparejamiento Piramidal Espacial variando ambas resoluciones.

3.6. Clasificación

El resultado de este proceso es la decisión sobre la clase a la que pertenece la imagen. Cada imagen es

reconocida como perteneciente a uno de los siguientes tipos: Personas con sombrero, Personas sin sombrero. Existen varias técnicas de clasificación que han sido probadas para el modelo BoVW y han mostrado buenos resultados. En este trabajo se compararon fundamentalmente dos enfoques, uno basado en la aplicación de una red Bayesiana de tres niveles y otro basado en un modelo SVM.

4. EXPERIMENTOS Y DISCUSION

En esta sección se exponen los experimentos realizados para evaluar la eficacia de cada uno de los métodos propuestos.

4.1. Alineación de los hombros

Primeramente, es necesario destacar que debido a la novedad del problema, no se cuenta con ninguna base de datos de imágenes de rostro clasificadas de acuerdo a la alineación de los hombros de las personas. Por tanto, fue necesaria la creación de una base de datos de imágenes bien clasificada que permitiera realizar los experimentos y comprobar la eficacia y eficiencia del sistema. Para la creación de esta base de datos se llevó a cabo un proceso de clasificación de 150 imágenes de rostro con fondo uniforme, donde para cada una de las imágenes se le asignó una de las dos clases: hombros alineados y hombros desalineados.

Además, a las 150 imágenes se les realizó el mismo proceso de clasificación de los grupos detectados descrito en la subsección 2.3, extrayéndole también a cada grupo un descriptor BSM de 64 componentes. Con esto se busca crear un conjunto de entrenamiento para el detector de hombros. Se generan dos bases de datos de grupos de píxeles etiquetados como hombro y no hombro; una para los hombros derechos y otra para los izquierdos. El total de grupos de píxeles determinados fue aproximadamente de 800, con un promedio de 5 grupos por imagen, entre los que se etiquetaron 300 como hombro (2 por cada imagen de la base de datos) y el resto como no hombro.

Los experimentos realizados fueron dirigidos a determinar la precisión del detector de hombros, ya que esta es la etapa fundamental de la solución. Una buena detección es la base para poder determinar la alineación de los hombros correctamente.

Para determinar la precisión de la clasificación se siguió el siguiente protocolo para cada una de las dos bases de datos: realizar 10 divisiones aleatorias sucesivas de los datos en 65% para entrenar y 35 % para probar con 5 modelos SVM, uno con kernel lineal para $C=10$ y 4 con kernel RBF de parámetros $C=0.01, 0.1, 1, 10$, siempre garantizando que los dos

subconjuntos queden balanceados en cuanto a la cantidad de grupos por cada una de las dos clases (hombro / no hombro). Los niveles de error obtenidos en cada uno de las divisiones se muestran en las figuras 3 para los descriptores de la región derecha e izquierda respectivamente.

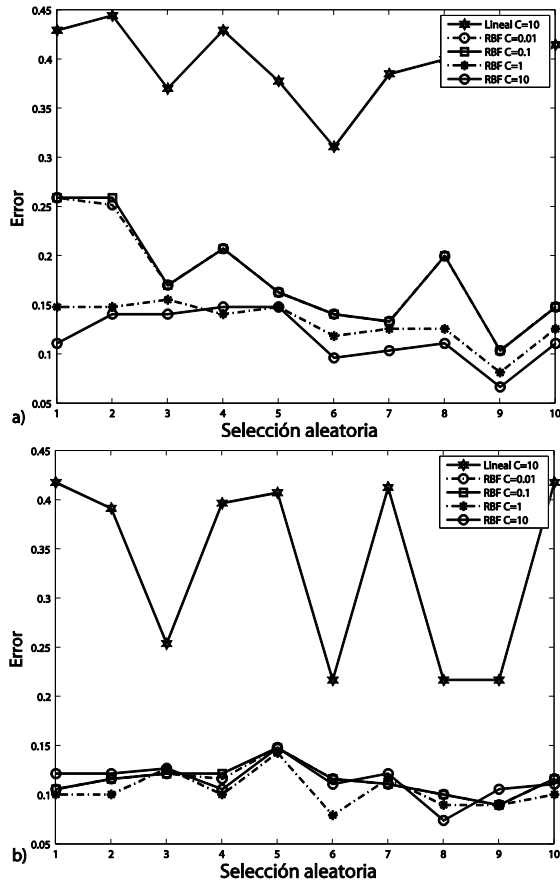


Figura 3. Error de la clasificación con SVM sobre los descriptores BSM de la región izquierda (a) y de la derecha (b).

Al analizar los gráficos mostrados en las figuras 3 y 4 puede verse como el SVM con kernel lineal no produce resultados satisfactorios para el problema planteado, mientras el uso de un kernel RBF disminuye en gran medida los índices de error, siendo el de parámetro $C=10$ el que muestra la mayor estabilidad y precisión, mostrando un error promedio de 0.1148 y 0.1178 en la detección de los hombros derechos e izquierdos respectivamente; dicha diferencia entre los kernels da una medida de la complejidad y no linealidad de las representaciones generadas mediante el descriptor BSM.

A partir del análisis de los resultados, se propone como solución final al problema planteado la creación de un método de cuatro etapas fundamentales que se basa en:

1. Detección del rostro y extracción de dos regiones de interés de la imagen.
2. Detección de bordes mediante el algoritmo de Canny y agrupamiento de píxeles con una estrategia jerárquica aglomerativa con distancia mínima entre los grupos.
3. Reconocer o detectar el patrón que más se asemeja a un hombro en cada una de las dos regiones mediante la extracción de un descriptor BSM a cada uno de los grupos detectados y su clasificación mediante dos modelos SVM previamente entrenados (destinados a la clasificación de los grupos de la derecha e izquierda)
4. Determinar la alineación.

4.2. Detección de sombreros

Para crear los diccionarios visuales se seleccionaron 30 imágenes aleatorias del conjunto de entrenamiento, 20 de la clase *persona con sombrero* y 10 de la clase *persona sin sombrero*. A estas se le extrajeron los rasgos usando SURF, SIFT, DSIFT y PHOW, tomando por cada tipo de rasgo hasta un total de 100000 descriptores aleatorios para construir el diccionario visual mediante el algoritmo de agrupamiento *k-means*, fijando como cantidad de grupos (palabras visuales) a generar, $K=100,200,\dots,600$, para un total de $4 \times 6 = 24$ diccionarios.

A partir de los 24 diccionarios generados, es necesario determinar cuál se ajusta mejor al problema, seleccionando así, el método de extracción de características y la cantidad de palabras que usará la solución finalmente. Luego se realizó otro experimento para determinar el valor de la constante C para el modelo SVM a generar y la cantidad de niveles a tener en cuenta en el análisis espacial de modo que se minimice el porcentaje de error. Para conseguir esto se realizaron las siguientes etapas:

- Se seleccionaron 170 imágenes con un balance entre la cantidad por clase.
- De las imágenes se extraen 4 descriptores por cada uno de los 24 diccionarios, los cuales representan los niveles de profundidad en el análisis espacial, $l=0, 1, 2, 3$. Para un total de $4 \times 4 = 16$ representaciones de las 170 imágenes.
- Para cada uno de ellos se realizaron 10 divisiones aleatorias sucesivas al 50% usando una mitad para entrenar 4 modelos SVM con kernel de intersección de histogramas con valores distintos del parámetro $C=0.1, 1, 10, 100$ y la otra para probar, siempre garantizando que las dos mitades queden balanceadas en cuanto a la cantidad de imágenes por clases.

En el gráfico mostrado en la figura 4 se presentan las combinaciones que mejor precisión alcanzaron para cada uno de los métodos de extracción de características.

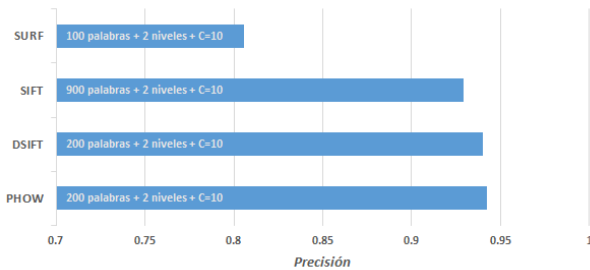


Figura 4. Mejores resultados de los experimentos para los métodos de extracción de características SURF, SIFT, DSIFT y PHOW.

Un resumen de los resultados obtenidos de estos experimentos para los diccionarios generados mediante el método PHOW hasta nivel 2 de profundidad se representa en la figura 5, donde se puede observar que se logra alcanzar hasta aproximadamente un 6% de error como promedio en las 10 divisiones aleatoria en algunos casos. Se observa también que para diccionarios pequeños es mejor seguir un enfoque espacial mientras que para los más grandes analizar la imagen completa es lo más conveniente.

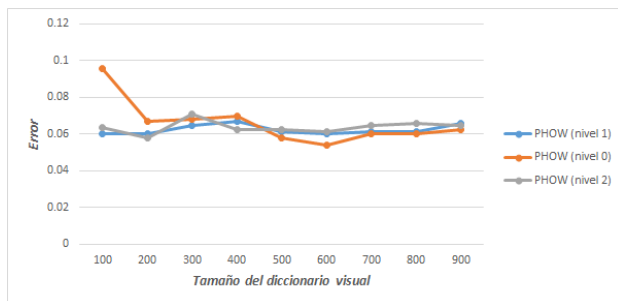


Figura 5. Menor error obtenido de los cuatro modelos SVM probados para cada uno de los tamaños del diccionario generados mediante la variante PHOW.

En el gráfico que se muestra en la figura 5 puede verse que el nivel 1 fue el más estable, reportando el segundo valor mínimo de error general con un tamaño del vocabulario visual de $K=200$. A partir de esto se justifica proponer como solución final un método de reconocimiento de patrones basado en BoVW con la siguiente configuración:

1. PHOW como método de extracción de rasgos de las imágenes.
2. Diccionario visual de tamaño $K=200$.
3. Describir las imágenes mediante vectores de 1000 componentes, formados a partir de la concatenación de los histogramas locales obtenidos de la división de la imagen en 4 regiones a

lo sumo (descender en el análisis espacial hasta un nivel de profundidad $l=1$), para dar una respuesta lo más rápido posible sin perder demasiado en precisión.

4. Finalmente clasificar las nuevas imágenes a partir de un modelo SVM con kernel de intersección de histogramas con parámetro de holgura $C=10$.

5. CONCLUSIONES

Pocas veces nos percatamos de la complejidad de los procesos que entraña la visión humana. Cuando distinguimos unas letras de otras, reconocemos a personas cercanas a nosotros o determinamos el género de una persona con una simple mirada, nuestro cerebro tiene que realizar un trabajo que implica procesos muy complejos. En esencia, un método de reconocimiento de patrones debería imitar estos criterios. Este concepto no es trivial y es lo que en cierto sentido tratamos de imitar con los enfoques automáticos usados para abordar los dos problemas tratados en este trabajo. Con esta investigación no se pretende obtener resultados mejores que la percepción humana, pero si aproximarse a ella.

Ambos problemas representan un ámbito novedoso, por lo cual ha sido necesario adaptar técnicas utilizadas en problemas similares y realizar modificaciones para explotar distintos enfoques de cómo abordarlos.

Esta investigación brinda dos soluciones completas para los problemas de la determinación de la alineación de los hombros y la detección de sombreros en imágenes de rostro con fondo uniforme, que adquiere gran importancia debido a la necesidad de determinar el valor identificativo de una imagen, a partir de la comprobación de las normas de calidad establecidas en la norma ICAO para los documentos de identificación personal.

Estas soluciones ya han sido implementadas en lenguaje C++ debido a las buenas prestaciones de este lenguaje para el procesamiento digital de las imágenes, que cobra mayor importancia cuando se busca que la evaluación de la calidad de la imagen sea un paso precursor a la identificación de las personas y por tanto un proceso que debe realizarse lo más eficientemente posible.

6. AGRADECIMIENTOS

Agradecemos a Dr. Leonardo Chang por varias discusiones interesantes de las que surgieron ideas de gran ayuda para encaminar nuestra investigación.

7. REFERENCIAS BIBLIOGRÁFICAS

1. **Ferrara, Matteo and Franco, Annalisa and Maio, Dario and Maltoni, Davide:** "Face Image Conformance to ISO/ICAO Standards in Machine Readable Travel Documents", IEEE Transactions on Information Forensics and Security, 2012.
2. **Sanz Cuenca, Julián and Escalera Guerrero, Sergio:** "Reconocimiento de Objetos por Descriptores de Forma", Departamento de Matemáticas Aplicada y Análisis, Universidad de Barcelona, 2008.
3. **Lowe, David G.:** "Distinctive Image Features from Scale-Invariant Keypoints", Int. J. Comput. Vision, pp. 91 – 110, 2004.
4. **Vedaldi, Andrea and Fulkerson, Brian:** "Vlfeat: An Open and Portable Library of Computer Vision Algorithms", Proceedings of the International Conference on Multimedia, pp. 1469 – 1472, ACM, 2010.
5. **Lazebnik, Svetlana and Schmid, Cordelia and Ponce, Jean:** "Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories", Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2, pp. 2169 – 2178, 2010.
6. **Bosch, A. and Zisserman, A. and Munoz, X.:** "Image Classification using Random Forests and Ferns", IEEE International Conference on Computer Vision, 2007.
7. **Lara, Arnaldo Camara and Jr., Roberto Hirata:** "Combining Features to a Class-Specific Model in an Instance Detection Framework", SIB-GRAPI, pp. 165-172, 2011.

8. **Hadjidemetriou, E. and Grossberg, M. and Nayar, S.:** "Multiresolution Histograms and Their Use for Recognition", Pattern Analysis and Machine Intelligence, IEEE Transactions on, Vol. 26, pp. 831 – 847, 2004.

9. **Kristen Grauman and Trevor Darrell:** "The pyramid match kernel: Discriminative classification with sets of image features", IN ICCV, pp. 1458 – 1465, 2005.

10. **Viola, Paul and Jones, Michel:** "Robust Real-time Object Detection", International Journal of Computer Vision, 2001.

11. **Duda, Richard O. and Hart, Peter E. and Stork, David G.:** Pattern Classification (2nd Edition), 2000.

8. SÍNTESIS CURRICULARES DE LOS AUTORES



Jesús Pérez-Martín: La Habana, Cuba, 27 de agosto de 1990. Graduado de Ciencia de la Computación, en la Universidad de La Habana en 2015. Desde entonces se ha desempeñado como Investigador en el departamento de Biometría del Centro de Aplicación de Tecnologías de Avanzada (CENATAV), Cuba. Ha dirigido su área de investigación hacia el campo de la visión por computadoras, el procesamiento de imágenes digitales, la biometría y el

reconocimiento de rostros. Se encuentra desarrollando su proyecto de doctorado en la selección y actualización de plantillas de rostro para el trabajo con grandes volúmenes de datos.