

UniversidadeVigo

DISEÑO E IMPLEMENTACIÓN DE UN SISTEMA
DE RECONOCIMIENTO FACIAL PARA LA
CARACTERIZACIÓN DE PERSONAS

Elvira García Mariño

Trabajo de Fin de Grado
Escuela de Ingeniería de Telecomunicación
Grado en Ingeniería de Tecnologías de Telecomunicación

Tutores
Laura Docío Fernández

2017

DISEÑO E IMPLEMENTACIÓN DE UN SISTEMA DE RECONOCIMIENTO FACIAL PARA LA CARACTERIZACIÓN DE PERSONAS

Autora: Elvira García Mariño

Tutora: Laura Docío Fernández

Curso: 2016/2017

I. Introducción

En el mundo actual, el rostro juega un papel fundamental en las relaciones que establece el ser humano con el entorno que le rodea: constituye una herramienta fundamental en la diferenciación de otros sujetos y es uno de los pilares en los que se asienta la comunicación no verbal. Es por ello que el reconocimiento facial y sus aplicaciones derivadas se han convertido en las últimas décadas en un campo de investigación creciente por sus múltiples aplicaciones en las nuevas tecnologías, entre las que se encuentran la realidad virtual o la interacción entre humanos y robots [1]. Además, también es destacable su uso en otras aplicaciones más específicas como pueden ser los sistemas de detección de fatiga, sistemas de detección de mentiras, sistemas de tutoría inteligente (ITS) [6], sistemas de video vigilancia en tiempo real, estudios de mercado y marketing (diseño y ajuste de publicidad dirigida a un público concreto), para la indexación y búsqueda de contenidos multimedia o para el etiquetado de las expresiones faciales en las redes sociales [9].

Por lo tanto, el desarrollo de sistemas y algoritmos de reconocimiento facial que sean capaces de detectar el género, predecir la edad, predecir el estado emocional y, en general, cualquier característica personal han cobrado una gran importancia en los últimos años. Muestra de ello son los numerosos artículos que presentan trabajos para la clasificación demográfica de imágenes faciales [2-5] (detección de género, clasificación de personas dependiendo del grupo de edad al que pertenecen o clasificación humana distinguiendo los distintos grupos étnicos), reconocimiento de expresiones faciales [5-8]...

Este trabajo fin de grado se centra en la caracterización de personas a partir de su cara. Dado que extraer, a partir de técnicas de reconocimiento facial, todas las posibles características que describen y diferencian a las personas supondría desarrollar una investigación muy extensa del problema, se ha acotado esta caracterización a la detección del género y al reconocimiento de emociones a partir de la expresión facial. Para ello se plantea el diseño, desarrollo y prueba de un sistema que permita caracterizar a las personas según su género y su expresión facial.

En la literatura existente sobre el tema a tratar se han propuesto múltiples técnicas de implementación. Actualmente, y tras la popularización del *deep learning* en la última década se pueden distinguir dos tipos de enfoques a la hora de resolver los problemas relacionados con el reconocimiento facial de personas. Por una parte, se pueden encontrar trabajos recientes [5, 7, 8] que implementan el reconocimiento facial, de

detección de género y de expresiones faciales mediante el uso de técnicas de *deep learning*. En el otro extremo se encuentran las técnicas de reconocimiento facial anteriores a la aplicación del *deep learning* en reconocimiento facial.

Si bien las técnicas basadas en *deep learning* proporcionan buenos resultados, demandan una gran cantidad de recursos para implementarlas, requiriendo de grandes bases de datos y del uso de GPUs para acelerar los algoritmos de entrenamiento [10]. Es por esta razón que en este trabajo se implementará un sistema de reconocimiento de género y expresiones faciales utilizando técnicas estado del arte anteriores a la aparición del *deep learning*.

Para ello, se seguirán los pasos habituales de los enfoques de reconocimiento facial tradicionales: preprocesado de la imagen, extracción de características y clasificación [7].

II. Objetivos

El objetivo final de este trabajo fin de grado es presentar un sistema que permita caracterizar personas dependiendo de su género y su expresión facial o emoción. En el primer caso se decide utilizar una muestra de catorce sujetos de cada género de la base de datos FERET [12, 13] dado que esta recopilación de imágenes cuenta con un gran número de individuos de ambos sexos. Para el reconocimiento de expresiones faciales, y puesto que el número de clases a reconocer aumenta considerablemente con respecto a las dos existentes en la detección de género (masculino y femenino), se considera necesario probar el sistema con expresiones diferentes en dos bases de datos cuya etiquetación las convierte en idóneas para su uso en reconocimiento de expresiones faciales: Yale Face Database [32] y JAFFE (*Japanese Female Facial Expression*) Database [14]. En la primera de ellas el número de clases a distinguir son seis (dormido, guiño, normal, felicidad, sorpresa y tristeza), mientras que en los experimentos en los que se utiliza la base de datos JAFFE se distinguirán 7 clases de expresiones faciales (enfado, desagrado, miedo, felicidad, tristeza, sorpresa y neutral).

Por las razones expuestas en el apartado anterior, esta caracterización se implementará siguiendo las etapas utilizadas tradicionalmente en los sistemas de la caracterización de personas: preprocesado, extracción de características o descriptores y clasificación. El sistema se desarrollará en C++ usando las librerías de OpenCV.

El primer bloque del sistema de caracterización consistirá en un preprocesado previo de la imagen de la persona a caracterizar. Con este preprocesado de la imagen se llevará a cabo la conversión a escala de grises de la imagen y la detección de la cara en la imagen objeto de estudio para posteriormente centrarnos en ella mediante el recorte de esta región de interés, ROI, (figura 1). Este preprocesado es importante ya que interesa descartar aquellas zonas de la imagen que no son relevantes para la caracterización de la persona y quedarse solo con las áreas clave o más relevantes de la imagen.



Figura 1: ROI con el resultado de la detección de la cara

En este proyecto, para el reconocimiento de expresiones faciales, se analizarán 3 posibles estrategias de definición o selección de las áreas relevantes. La primera de ellas consiste en utilizar toda el área que devuelve el bloque de detección facial. Las otras dos se basan en utilizar áreas en las que se localizan rasgos que resultan especialmente característicos en el reconocimiento de expresiones faciales como son los ojos y la boca. Por lo tanto, una vez realizada la detección facial se obtienen las regiones de la imagen que contienen los ojos y la boca. De este modo se obtienen dos subimágenes por cada imagen original cada una conteniendo un área relevante (figura 2).



Figura 2: ROI con el resultado de la detección de la boca y los ojos

La segunda estrategia consiste en utilizar estas dos subimágenes, procesando cada una de ellas de forma independiente y por lo tanto extrayendo de cada una de ellas un conjunto de descriptores que posteriormente se pueden combinar. Por último, la tercera estrategia que se analizará consiste en unir las dos regiones anteriores (boca y ojos) en una sola imagen (figura 3).

Tras la detección de las diferentes regiones de interés, o áreas relevantes de las que se extraerán los descriptores, se extraen estas regiones y se reescalan para que todas tengan el mismo tamaño. El tamaño utilizado en la primera estrategia (extracción de descriptores de las regiones de ojos y de boca de forma independiente para posteriormente concatenar los descriptores en un vector único) ha sido de 100 cols x 20 filas para ojos y de 80 cols x 40 filas para boca en Yale Face Database. En el caso de JAFFE, se han reescalado las imágenes a 128 cols x 64 filas para ojos y 64 cols x 32 filas para boca. En las otras dos estrategias, en Yale Face Database, las regiones se reescalan a un tamaño de 80 cols x 80 filas, mientras que en JAFFE Database el tamaño de la imagen luego del reescalado es de 128 cols x 128 filas.

La detección de rostro, ojos y boca se realizará siguiendo el proceso presentado por Viola y Jones en 2001 [15] y su procedimiento básico se describe en el estado del arte de esta memoria (anexo 1).



Figura 3: Tercer tipo de preprocesado

Una vez concluida la fase de preprocesado anterior, la siguiente etapa consistirá en la extracción de características o descriptores de la región seleccionada de la imagen. En este trabajo se utilizarán y analizarán dos tipos de descriptores diferentes. El primero que se implementará será el Histograma de Gradientes Orientados (HOG, *Histogram of Oriented Gradients*), el cual se describe de forma breve en el estado del arte (anexo 1). Para la implementación de este descriptor en el sistema de caracterización de personas se optará por utilizar el código ya desarrollado en las librerías OpenCV para C++. El segundo

descriptor a implementar será LBP (*Local Binary Patterns*), un descriptor de textura utilizado ampliamente en la literatura existente sobre técnicas de reconocimiento facial y cuyo concepto básico puede consultarse en el estado del arte (anexo 1).

La última etapa en el sistema de caracterización de personas que se describe en este trabajo consistirá en la clasificación de las imágenes dependiendo de la clase a la que pertenezcan, para lo que se ha decidido utilizar como único clasificador el SVM (*Support Vector Machines*), del que también se hace una pequeña descripción en el anexo 1 de esta memoria.

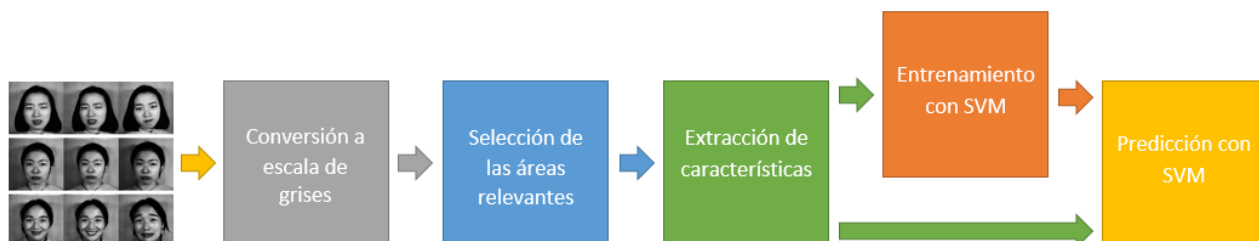


Figura 4: Diagrama de bloques del sistema

III. Resultados

Como ya se ha adelantado en los apartados anteriores, el sistema descrito en esta memoria implementará un sistema capaz de caracterizar a una persona según su género y su expresión facial (emoción). Para ello se implementarán dos clasificadores, uno para cada una de esas tareas.

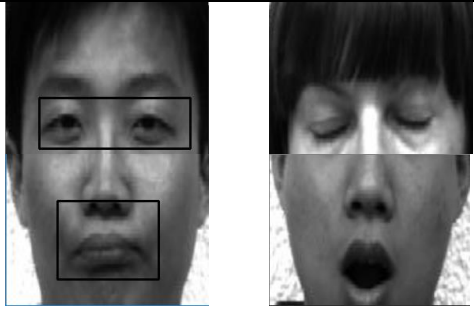
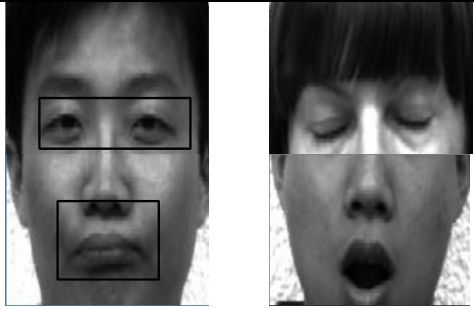
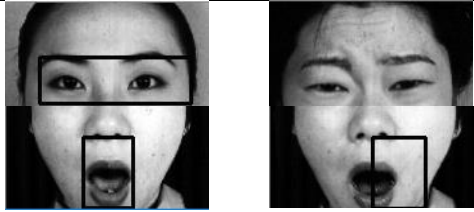
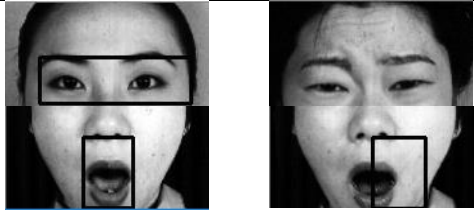
Con el fin de diseñar y desarrollar un sistema que proporcione las mejores prestaciones posibles, se han llevado a cabo diferentes experimentos en los que se comparan las prestaciones que se obtienen al variar distintos parámetros de configuración en la extracción de los dos descriptores analizados en este trabajo: HOG y LBP. Concretamente se ha evaluado la influencia del tamaño de las celdas o bloques, así como también diversas combinaciones y variantes de ellos.

El procedimiento para obtener el resultado final de clasificación que se utilizará es muy similar para los dos tipos de reconocimientos que son objeto de estudio en este trabajo, por lo que se ha decidido realizar todo este análisis y toma de decisiones en la tarea de reconocimiento de expresiones faciales para posteriormente desarrollar el detector de género utilizando aquellas configuraciones que hayan dado mejores resultados.

El principal motivo de implementarlo sobre el reconocimiento de expresiones faciales es que en este caso, como se ha indicado en la sección anterior, se han estudiado y analizado tres estrategias para la definición de regiones o áreas de interés de las imágenes, las cuales se utilizarán para extraer los descriptores y realizar la clasificación, mientras que en la detección de género esa área o región consistirá exclusivamente en el resultado de la detección facial.

Así pues, con el fin de reducir el porcentaje de fallo en el sistema final de reconocimiento de expresiones faciales, en la fase de preprocesado se han descartado todas aquellas imágenes en las que no ha sido posible detectar de forma correcta alguna de las regiones de interés. En la tabla I se muestra el porcentaje de fallo para las dos bases de datos de prueba, así como una muestra de lo que se ha considerado como un reconocimiento de rasgos faciales correcto e incorrecto. Es necesario destacar que en dicha tabla no se han incorporado los porcentajes de error para la detección facial ya que tanto en Yale Face Database como en JAFFE todas las caras han sido identificadas correctamente.

Tabla I: Resultados de la detección de rasgos faciales

		Porcentaje de fallo	Ejemplo de acierto	Ejemplo de fallo
Yale Face Database Original	Detección de ojos	3,33%		
	Detección de bocas	2,22%		
	Total	5.6%		
JAFPE Database	Detección de ojos	11.27%		
	Detección de bocas	0.47%		
	Total	11.74%		

Una vez se han eliminado de ambas bases de datos las imágenes en las que la detección de rasgos faciales no es exitosa y que por lo tanto reducirían la tasa de éxito al entrenar el clasificador con esos datos, se procede a la extracción de características en cada una de las estrategias de definición de ROI consideradas en el reconocimiento de expresiones faciales. Para ello, se han utilizado dos tipos de descriptores diferentes, HOG y LBP, en los que es crucial la elección adecuada de diferentes parámetros como el tamaño de las celdas o bloques que se utilizan ya que con ello se define el número de regiones en las que se dividen las imágenes, y de las que se extraen dichos descriptores, lo que a su vez constituye un factor importante en la obtención de buenas prestaciones. Debido a lo anterior, los experimentos que se han realizado en este trabajo se centran en mejorar progresivamente la calidad de los descriptores obtenidos, no solo modificando el tamaño de celda o bloque sino que también se comparan diferentes variantes en la implementación de los mismos.

El principal motivo para ello es que, aunque para obtener buenos resultados en el reconocimiento de expresiones faciales también es de especial importancia una elección óptima de los parámetros utilizados en el clasificador, en este sistema se ha usado el código desarrollado en las librerías OpenCV para C++ en la implementación del clasificador SVM. Entre sus funciones se incluye la función *trainAuto*, que busca de forma automática los parámetros óptimos para el clasificador por lo que se usa esta función para la búsqueda automática de los parámetros adecuados. Además, facilita la identificación del mejor tipo de kernel, que en este problema en concreto ha resultado ser el RBF (*Radial Basis Function*) en vez del lineal.







Por todo lo anterior, a continuación se explicarán los experimentos realizados, así como los resultados obtenidos en cada uno de ellos.

Para el desarrollo del primer experimento se han dividido cada una de las bases de datos en tres grupos diferentes con el objetivo de realizar una validación cruzada (*cross validation*) de 3 grupos. En cada grupo se incluirán a los mismos sujetos para cada una de las emociones a clasificar debido a que se ha decidido implementar un sistema de clasificación independiente de las personas incluidas en el conjunto de entrenamiento, es decir, los conjuntos de entrenamiento y test son disjuntos en cuanto a sujetos presentes

en ellos. De este modo, para cada experimento se realizan varios entrenamientos diferentes, tres en este primer caso, en el que dos de los grupos de imágenes se utilizan para el entrenamiento del clasificador SVM y sobre el grupo restante se realiza la predicción de las emociones.

Además, en esta primera aproximación se ha optado por implementar un sistema en el que apenas se divida la imagen en bloques o ventanas de menor tamaño. De hecho, para la extracción de características LBP se ha decidido no dividir la imagen en ventanas más pequeñas obteniendo en consecuencia 256 descriptores para cada imagen, por lo que se espera que el error en el reconocimiento de expresiones faciales sea elevado. Pese a ello, este primer experimento se ha considerado relevante ya que su comparación con experimentos posteriores permite percatarse de la gran influencia de la división de la imagen en ventanas a la hora de conseguir un buen reconocedor de expresiones faciales. Luego de obtener los resultados, como se puede comprobar de forma resumida en la tabla II, se ha verificado que este tipo de configuración no permite alcanzar unos resultados que puedan llevar a un correcto reconocimiento de expresiones faciales.

Tabla II: Resultados del primer experimento sobre imágenes de prueba (promedio)

		Número descriptores HOG	Número descriptores LBP	Aciertos HOG+SVM	Aciertos LBP+SVM	Ejemplo de imagen usada
Yale Face	Primer preprocesado (descriptores ojos y boca)	648	512	46.26%	48.18%	
	Segundo preprocesado (composición ojos y boca)	324	256	63.78%	34.19%	
	Tercer preprocesado (imagen completa)	324	256	61.32%	14.31%	
JAFPE	Primer preprocesado (descriptores ojos y boca)	648	512	57.19%	41.15%	
	Segundo preprocesado (composición ojos y boca)	324	256	54.62%	35.16%	
	Tercer preprocesado (imagen completa)	324	256	68.33%	17.09%	

Por otra parte, y siempre dentro de las condiciones establecidas previamente para el desarrollo de este primer ensayo, en la extracción de características HOG se utilizaron ventanas de gran tamaño de modo que se obtenga un número reducido de descriptores para cada imagen (324 en ambas bases de datos). Los resultados, como se puede observar en la tabla II, mejoran considerablemente con respecto a los obtenidos con LBP sin dividir la imagen en regiones de menor tamaño y consigue clasificar la expresión facial de forma correcta para aproximadamente el 60% de las imágenes de prueba en el preprocesado con mejores resultados. Si se comparan los resultados obtenidos con HOG y LBP en este primer experimento, aun tratándose de descriptores diferentes, se puede intuir la importancia de dividir la imagen en ventanas durante el reconocimiento de expresiones faciales dado que las diferencias obtenidas son lo suficientemente grandes como para concluir que la principal razón de ellas es que en HOG se ha dividido la imagen en ventanas, al contrario que en LBP. De hecho, si se observa la tabla II se puede comprobar que en el único caso en el que los resultados LBP se asemejan a los obtenidos con HOG es en el que se obtienen los descriptores para las zonas de ojos y bocas por separado y posteriormente se concatenan, por lo que el número de descriptores total es el doble (512 descriptores) que en los dos casos restantes.

Puesto que en el ensayo anterior no se han obtenido unos resultados que conduzcan a un reconocimiento de expresiones faciales aceptable utilizando la extracción de características LBP y considerando toda la región de interés como ventana de análisis sobre la que calcular el histograma, en este segundo experimento se calculan los descriptores LBP utilizando ventanas o bloques más pequeños para luego obtener un único vector cuyo número de valores sea mayor que el obtenido con anterioridad. A la hora de dividir la imagen se ha optado por fraccionar tanto las filas como las columnas en 6 o más ventanas. Con la configuración usada, como se puede observar en la tabla III, se aumenta el tamaño del vector de descriptores (aumenta del orden de cientos de valores a miles) y con ello se logran unos resultados que permiten hablar de una clasificación exitosa en más del 50% de las imágenes de prueba para todos los tipos de preprocesado. Además, en la extracción de características utilizando HOG se reducirá también el tamaño de la celda o ventana para comprobar en qué medida mejoran los resultados obtenidos anteriormente cuando el número de descriptores aumenta (1764 descriptores para las imágenes completas y en torno a 1500 para la concatenación de ojos y boca). Si se comparan las tablas II y III se observa que los resultados mejoran para la base de datos Yale Face Database pero que empeoran para JAFFE en el reconocimiento de expresiones en la imagen completa y cuando se usa como descriptor la concatenación de los descriptores obtenidos de la ROI de los ojos y de la ROI de la boca, por lo que se deduce que no siempre un aumento de descriptores conlleva una mejora de los resultados.

Tabla III: Resultados del segundo experimento sobre imágenes de prueba (promedio)

	Número descriptores HOG	Número descriptores LBP	Yale Face		JAFFE	
			Aciertos HOG+SVM	Aciertos LBP+SVM	Aciertos HOG+SVM	Aciertos LBP+SVM
Primer preprocesado (descriptores ojos y boca)	1512 (JAFFE) 1620 (Yale)	9216	62.07%	54.80%	57.12%	51.64%
Segundo preprocesado (composición ojos y boca)	1764	12544	66.35%	59.09%	57.95%	52.23%
Tercer preprocesado (imagen completa)	1764	12544	61.64%	54.78%	62.88%	58.28%

Una vez que se ha comprobado que un aumento de descriptores no siempre implica que mejoren los resultados y ya que el número de imágenes de entrenamiento usadas para cada clase en los experimentos anteriores es bajo (aproximadamente 10 para Yale Face Database y 25 para JAFFE), en el tercer ensayo se utiliza un conjunto mayor de imágenes de entrenamiento en las pruebas realizadas, concretamente se realiza un experimento de tipo “dejando uno fuera” (*leave-one-out*). De este modo, en cada iteración que se ejecuta se predice la expresión facial sobre un único sujeto de cada base de datos mientras que los restantes son usados en el entrenamiento del clasificador SVM. Así pues, se han realizado 10 entrenamientos diferentes en la Japanese Female Facial Expression y 15 en Yale Face Database, es decir, uno por cada persona que conforma la base de datos. Se debe destacar que el tamaño de las imágenes y de las ventanas usadas es el mismo que en el experimento anterior, y por lo tanto el número de descriptores con los que se entrena el sistema también se mantiene.

Así, comparando los resultados de las tablas III y IV se demuestra que el aumento del número de imágenes utilizadas en el entrenamiento del sistema repercute positivamente en los resultados obtenidos en la predicción de expresiones faciales.

Además, se considera conveniente probar la influencia de diferentes variantes para la extracción de descriptores LBP sobre el sistema de reconocimiento final. Para ello, se comparan el Extended LBP (usado en todos los experimentos anteriores) con el LBP uniforme y se obtienen unos resultados muy similares, por lo que no se puede concluir que se obtenga una mejora significativa con el uso de ninguno de ellos.

Finalmente, en el marco de este último experimento realizado con respecto al reconocimiento de expresiones faciales se comprueba si el uso conjunto de los descriptores LBP y HOG produce una mejora significativa de los resultados. Puesto que en este tercer ensayo se han obtenido los mejores resultados cuando se considera la cara completa como área o región relevante, esta última prueba se implementa únicamente en esta situación. Como se puede comprobar en la tabla IV y de forma más detallada en el anexo 4, con este sistema se consigue aproximar el resultado al obtenido con el mejor de los dos descriptores para cada caso particular, pero no se consigue una mejora en términos globales del sistema de reconocimiento de expresiones faciales.



Tabla IV: Resultados del tercer experimento sobre imágenes de prueba (promedio)

		Aciertos HOG+SVM	Aciertos LBP+SVM	Aciertos LBP Uniforme+SVM	Aciertos LBP Uniforme+ HOG+SVM
Yale Face	Primer preprocesado (descriptores ojos y boca)	56.44%	60.88%	59.77%	
	Segundo preprocesado (composición ojos y boca)	69%	60.89%	60.89%	
	Tercer preprocesado (imagen completa)	69.89%	59.78%	60.89%	69.89%
JAFFE	Primer preprocesado (descriptores ojos y boca)	61.19%	56.64%	57.06%	
	Segundo preprocesado (composición ojos y boca)	61.58%	57.60%	57.60%	
	Tercer preprocesado (imagen completa)	65.38%	59.69%	59.44%	64.66%

Una vez diseñado y analizado el sistema de reconocimiento final sobre la clasificación de expresiones faciales, se ha decidido probar la tercera configuración de las analizadas anteriormente en la clasificación de género puesto que es la que mejores resultados proporciona en el reconocimiento de emociones. Como ya se ha mencionado, solo se obtendrán los descriptores para la imagen completa del rostro ya que para los rasgos faciales involucrados en la detección de género no se pueden identificar de forma sencilla zonas que sean especialmente características de ese rasgo.

Para ello, antes de comenzar el proceso de reconocimiento facial deberán descartarse aquellas imágenes en las que durante la fase preprocesado la detección de la cara no haya concluido de forma exitosa. Si se consulta la tabla V, el porcentaje de imágenes en las que no se ha identificado el rostro correctamente es muy elevado (del 29%, aproximadamente). Sin embargo, esto se debe a que no se ha descartado ninguna de las imágenes de los sujetos seleccionados de entre la gran gama de individuos disponibles en FERET, por lo que se incluyen instantáneas en las que los sujetos posan de perfil. Es en ellas donde el sistema de detección de la cara falla.

Tabla V: Resultados de la detección facial en FERET

	Porcentaje de fallo	Ejemplo de acierto	Ejemplo de fallo
Detección de caras	28.91%		

Por último, para el desarrollo del sistema de clasificación por género se han utilizado los mismos procedimientos utilizados en el tercer experimento desarrollado para el reconocimiento de expresiones faciales. De este modo, puesto que se han usado 14 sujetos para cada una de las dos clases (masculina y femenina), también se necesitan 14 entrenamientos para la predicción de género. Es decir, en cada una de las iteraciones se entrenará el clasificador SVM con 26 individuos y se predecirá el género en los dos sujetos restantes, que siempre pertenecerán a clases diferentes. Observando la tabla VI y comparando los resultados con los obtenidos en la tabla IV, se concluye que el porcentaje de éxito es similar en el reconocimiento de expresiones faciales y en la clasificación facial por género, por lo que los resultados obtenidos dependen sobre todo del sistema de reconocimiento implementado y no de las características personales que se desean reconocer.

Tabla VI: Resultados de la clasificación facial por género

	Aciertos HOG+SVM	Aciertos LBP Uniforme+SVM	Aciertos LBP Uniforme+ HOG+SVM
Sobre imágenes de prueba (promedio)	67.01%	58.79%	67.71%

IV. Conclusiones

A continuación, se presentan las conclusiones que se deducen a partir de los resultados presentados en el apartado anterior:

1. Para que la extracción de características proporcione buenos resultados en la fase de clasificación es importante realizar una extracción de descriptores de forma local, es decir, dividir la imagen en ventanas de menor tamaño y obtener los descriptores de cada una de ellas para luego concatenarlos en un vector de características único. Con este proceso, a pesar de que el tamaño del vector de descriptores aumenta de forma considerable, se consigue una gran mejora en las prestaciones del clasificador.
2. Aunque fraccionar la imagen en bloques o ventanas y, en consecuencia, aumentar el número de descriptores se traduce en una mejora de los resultados en comparación con los obtenidos si se utilizase una sola ventana del tamaño de la imagen, no siempre aumentar el número de descriptores conlleva una mejora significativa en el sistema de reconocimiento. Esto se debe a que para trabajar con descriptores de alta dimensionalidad son necesarios también una mayor cantidad de datos si se quiere que el clasificador entrenado sea capaz de generalizar y no se incurra en un sobreajuste del mismo a los datos de entrenamiento.
3. En este trabajo los resultados se han obtenido para dos sistemas de caracterización diferentes: reconocimiento de expresiones faciales (emociones) y clasificación del sujeto por género. Del hecho de que el primero sea un sistema multiclase (se deben catalogar los sujetos en 6 o 7 clases dependiendo de la base de datos analizada), se podría intuir inicialmente que conllevaría una peor clasificación final con respecto al segundo sistema, en el que solo se cuenta con dos clases diferentes. Sin embargo, luego de analizar los resultados obtenidos en el apartado 3 de esta memoria se concluye que la tasa éxito depende sobre todo de la configuración del sistema en cada una de sus etapas (preprocesado, extracción de características y clasificación) y no tanto del tipo o tarea de reconocimiento facial que se implemente.

En el anexo 3 se plantean aquellas líneas futuras en las que se debería seguir trabajando para conseguir mejores resultados en la caracterización de personas según su género y expresión facial.

V. Bibliografía

- [1] Zhao, W., Chellappa, R., Phillips, P. J., & Rosenfeld, A. (2003). Face recognition: A literature survey. *ACM computing surveys (CSUR)*, 35(4), 399-458.
- [2] Wu, B., Ai, H., & Huang, C. (2004, August). Facial image retrieval based on demographic classification. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on* (Vol. 3, pp. 914-917). IEEE.
- [3] Shewaye, T. N. (2013). Age group and gender recognition from human facial images. *arXiv preprint arXiv:1304.0019*.
- [4] Mousa Pasandi, M. E. (2014). *Face, Age and Gender Recognition using Local Descriptors* (Doctoral dissertation, Université d'Ottawa/University of Ottawa).
- [5] Dehghan, A., Ortiz, E. G., Shu, G., & Masood, S. Z. (2017). DAGER: Deep Age, Gender and Emotion Recognition Using Convolutional Neural Network. *arXiv preprint arXiv:1702.04280*.
- [6] Kumari, J., Rajesh, R., & Pooja, K. M. (2015). Facial expression recognition: A survey. *Procedia Computer Science*, 58, 486-491.
- [7] Mollahosseini, A., Hasani, B., Salvador, M. J., Abdollahi, H., Chan, D., & Mahoor, M. H. (2016). Facial Expression Recognition from World Wild Web. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (pp. 58-65).
- [8] Alizadeh, S., & Fazel, A. (2017). Convolutional Neural Networks for Facial Expression Recognition. *arXiv preprint arXiv:1704.06756*.
- [9] Roychowdhury, S., & Emmons, M. (2015). A survey of the trends in facial and expression recognition databases and methods. *arXiv preprint arXiv:1511.02407*.
- [10] Liu, Y., Li, Y., Ma, X., & Song, R. (2017). Facial Expression Recognition with Fusion Features Extracted from Salient Facial Areas. *Sensors*, 17(4), 712.
- [11] Izard, C. E. (1994). Innate and universal facial expressions: evidence from developmental and cross-cultural research.
- [12] Phillips, P. J., Wechsler, H., Huang, J., & Rauss, P. J. (1998). The FERET database and evaluation procedure for face-recognition algorithms. *Image and vision computing*, 16(5), 295-306.
- [13] Phillips, P. J., Moon, H., Rizvi, S. A., & Rauss, P. J. (2000). The FERET evaluation methodology for face-recognition algorithms. *IEEE Transactions on pattern analysis and machine intelligence*, 22(10), 1090-1104.
- [14] Lyons, M., Akamatsu, S., Kamachi, M., & Gyoba, J. (1998, April). Coding facial expressions with gabor wavelets. In *Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on* (pp. 200-205). IEEE.
- [15] Viola, P., & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on* (Vol. 1, pp. I-I). IEEE.
- [16] Zafeiriou, S., Zhang, C., & Zhang, Z. (2015). A survey on face detection in the wild: past, present and future. *Computer Vision and Image Understanding*, 138, 1-24.

- [17] Ojala, T., Pietikainen, M., & Harwood, D. (1994, October). Performance evaluation of texture measures with classification based on Kullback discrimination of distributions. In *Pattern Recognition, 1994. Vol. 1-Conference A: Computer Vision & Image Processing, Proceedings of the 12th IAPR International Conference on* (Vol. 1, pp. 582-585). IEEE.
- [18] Wang, L., & He, D. C. (1990). Texture classification using texture spectrum. *Pattern Recognition*, 23(8), 905-910.
- [19] Ahonen, T., Hadid, A., & Pietikäinen, M. (2004). Face recognition with local binary patterns. *Computer vision-eccv 2004*, 469-481.
- [20] Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3), 273-297.
- [21] Freund, Y. & Schapire, R. (1997). A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *Journal of Computer and System Sciences*, 55(1), pp.119-139.
- [22] Dalal, N., & Triggs, B. (2005, June). Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on* (Vol. 1, pp. 886-893). IEEE.
- [23] Arranz Aranda, F., Yin, L., Qi López Cámara, J. M., & Martín de la Calle, P. J. (2011). *Interacción persona-computador basada en el reconocimiento visual de manos* (Doctoral dissertation, Tesis Universidad Complutense de Madrid).
- [24] Huang, D., Shan, C., Ardebilian, M., & Chen, L. (2011). Facial image analysis based on local binary patterns: A survey. *IEEE Transactions on Image Processing*.
- [25] Vasanth, P. C., & Nataraj, K. R. (2015). Facial Expression Recognition Using SVM Classifier. *Indonesian Journal of Electrical Engineering and Informatics (IJEI)*, 3(1), 16-20.
- [26] Ghimire, D., Jeong, S., Lee, J., & Park, S. H. (2016). Facial expression recognition based on local region specific features and support vector machines. *arXiv preprint arXiv:1604.04337*.
- [27] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444.
- [28] Liu, P., Han, S., Meng, Z., & Tong, Y. (2014). Facial expression recognition via a boosted deep belief network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1805-1812).
- [29] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097-1105).
- [30] *Face Recognition Technology (FERET)*. (2017). NIST. Recuperado el 9 de Julio de 2017, a partir de <https://www.nist.gov/programs-projects/face-recognition-technology-feret>
- [31] Ojansivu, V., & Heikkilä, J. (2008, July). Blur insensitive texture classification using local phase quantization. In *International conference on image and signal processing* (pp. 236-243). Springer Berlin Heidelberg.
- [32] *Yale Face Database*. (2017). [cvc.cs.yale.edu](http://cvc.cs.yale.edu/cvc/projects/yalefaces/yalefaces.html). Recuperado el 11 de Julio de 2017, a partir de <http://cvc.cs.yale.edu/cvc/projects/yalefaces/yalefaces.html>

Anexo 1: Estado del arte

Los primeros estudios sobre el reconocimiento facial y detección de expresiones faciales se remontan a la década de los años 60 [9] y constituyen una de las primeras líneas de investigación que se llevaron a cabo en el campo de la visión artificial. En los años 90 se produce un aumento significativo de trabajos que versan sobre este tema, y en esta década se publican por primera vez los descriptores LBP [17] o los clasificadores SVM [20] y AdaBoost [21].

Presentados por primera vez en 1994 por Timo Ojala, Matti Pietikainen y David Harwood [17], los descriptores LBP (*Local Binary Patterns*) fueron definidos como una medida para analizar la textura de una imagen. Los LBP originales son una adaptación de dos niveles del descriptor de textura propuesto en 1990 por Li Wang y Dong-Chen He [18], y con ello consiguen reducir el número de unidades de textura de 6561 a 256. Esto es debido a que para etiquetar los píxeles de la imagen se define una región de 3x3 vecinos, siendo el píxel que es objeto de evaluación el central y el que servirá de umbral para asignar un valor binario a los píxeles vecinos dependiendo de si son mayores o menores que el píxel central. El valor resultante de concatenar todos los valores asignados a los píxeles vecinos se considera un número binario que servirá para codificar el píxel central (figura 5). Por último, una vez codificados los píxeles, se calcula el histograma de estos y se utiliza como un descriptor de textura .

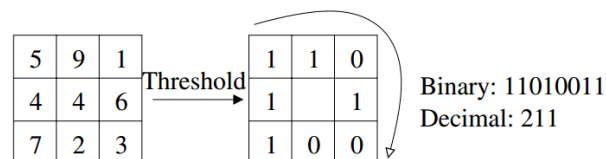


Figura 5: Funcionamiento del LBP [19].

Aunque fueron diseñados como descriptores de textura de una imagen, el uso de LBP en reconocimiento facial se debe al hecho de que la imagen puede ser representada como un conjunto de micropatrones invariantes a transformaciones en escala de grises [19]. Sin embargo, la extracción de características utilizando un único descriptor LBP para toda la imagen solo nos proporciona información sobre la distribución de los micropatrones existentes sobre la imagen. Si se divide la imagen en ventanas de menor tamaño, se calculan sobre ellas los descriptores LBP y por último se concatenan todos los descriptores LBP en un único vector global para toda la imagen, se consigue una descripción local de la textura y una descripción global de la forma de la imagen [19]. Desde la publicación de los LBP originales hasta la actualidad, se han publicado numerosas variantes del primer estudio como son el ILBP (*Improved LBP*), ELBP (*Extended LBP*) o el MLBP (*Modified LBP*) [24].

El clasificador SVM fue descrito por Corinna Cortes y Vladimir Vapnik en 1995 [20] y es una de las técnicas de aprendizaje más populares en los procesos de clasificación [4]. En el proceso de entrenamiento el SVM necesita un conjunto de muestras etiquetadas atendiendo a la clase a la que pertenecen. Estos datos son representados como puntos en el espacio para luego encontrar el hiperplano de separación entre clases que minimice el error de clasificación. El SVM básico fue diseñado para situaciones en las que solo se cuenta con dos tipos de clases linealmente separables. Para ampliar esta técnica a conjuntos de datos no divisibles linealmente, se proponen distintos tipos de kernel (RBF, polinómico, sigmoide...).

Por su parte, AdaBoost es un algoritmo de aprendizaje presentado por Yoav Freund y Robert Schapire [21]. El algoritmo presentado propone un método de clasificación iterativo fundamentado en una serie de clasificadores débiles. La idea es que cada nuevo clasificador débil que se introduzca se centre en los errores de los clasificadores anteriores, obteniendo así un clasificador que devuelva unos buenos resultados.

A pesar de todos los avances logrados en la década de los 90, no es hasta 2001 cuando Paul Viola y Michael Jones [15] describen el primer entorno de trabajo que proporciona buenos resultados en la detección de rostros en tiempo real [16].

En este trabajo, Viola y Jones [15] señalan que son tres las principales contribuciones de su entorno de trabajo. La primera es la combinación de las características Haar Like con la *imagen integral*, con lo que consigue un procesamiento rápido de características. En segundo lugar, reducen el elevado número de características Haar Like resultantes del paso anterior utilizando AdaBoost. Por último, desarrollan un método en el que combinan clasificadores cada vez más complejos en un sistema en cascada con el que aumentar la velocidad de detección centrándose en las zonas de la imagen en las que es más probable que se detecte una cara.

En 2005, con la publicación de Navneet Dalal y Bill Triggs [22] en la que hacen uso de los descriptores HOG (*Histogram of Oriented Gradients*) para la detección de personas, se populariza el uso de HOG en visión artificial. Este descriptor se basa en la idea de que la apariencia y forma local de un objeto representado en una imagen puede ser definida por la orientación de los gradientes de intensidad local. Para ello, se divide la imagen en regiones más pequeñas (celdas) y para cada una de ellas se calcula el histograma local sobre las orientaciones que toma el gradiente de cada píxel. Para que el método sea robusto frente a variaciones de iluminación, se normalizan los valores de cada celda utilizando para ello la acumulación de valores de los Histogramas de Gradientes Orientados de varias celdas. Esta agrupación de celdas se denomina bloque y el valor acumulado sirve para normalizar todo el bloque. Para calcular el HOG final de la imagen, la normalización de histogramas se realiza solapando los bloques, de forma que en cada nueva iteración se eliminan la fila izquierda y la columna superior del bloque anterior [23].

Los descriptores y clasificadores expuestos en los párrafos anteriores son solo una pequeña muestra del extenso trabajo de investigación llevado a cabo con el propósito de definir métodos de extracción de descriptores y de clasificación que proporcionen los mejores resultados posibles. Fundamentándose en ellos, en la última década se han publicado numerosos sistemas de reconocimiento de expresiones faciales que combinan diferentes descriptores y clasificadores [4, 10, 25, 26].

Con la aparición del uso de *deep learning* en visión artificial, se han publicado numerosos trabajos en los que se usan las DNN (*Deep Neural Networks*) en sistemas de reconocimiento facial. El término *deep learning* o aprendizaje profundo engloba diversas técnicas cuya característica común es que cuentan con varias capas de procesamiento con el objetivo de aprender conjuntos de datos con múltiples niveles de abstracción [28]. En esta categoría se encuentran, entre otras, las redes neuronales convolucionales (CNN, *Convolutional Neural Network*) o las redes de creencia profunda (DBN, *Deep Belief Network*), de las que podemos encontrar trabajos en el ámbito del reconocimiento facial [5, 7, 8, 28].

Hasta 2012, la aplicación de redes neuronales en trabajos relacionados con el ámbito de la visión artificial no era frecuente. En la competición *ImageNet* de 2012 se presentó un sistema de entrenamiento basado en redes neuronales convolucionales y que fue aplicado a un conjunto de datos de alrededor de un millón de imágenes pertenecientes a 1000 clases diferentes [29], obteniendo tasas de error mucho menores que el resto de trabajos (del orden de la mitad con respecto a los mejores competidores). Esto supuso una revolución en el campo de la visión artificial, surgiendo numerosos estudios basados en redes neuronales [10].

Sin embargo, aunque los trabajos basados en *deep learning* proporcionan unos buenos resultados, el hecho de que para su entrenamiento se necesiten grandes bases de datos y el uso de GPUs promueve la publicación de nuevos estudios basados en técnicas de reconocimiento facial tradicionales [10].

Anexo 2: Bases de datos

El trabajo de recogida y anotación de imágenes atendiendo a la expresión facial es una tarea compleja y con un coste de tiempo elevado. Por ello, surgen numerosas bases de datos en las que las imágenes están etiquetadas por expresión facial como la JAFFE (*Japanese Female Facial Expression*) Database [14] o la Yale Face Database [9, 32]. La mayoría de estas bases de datos recogen todas o algunas de las expresiones faciales básicas (enfado, desagrado, miedo, felicidad, tristeza y sorpresa) [10, 11], aunque también pueden diferenciar otras como la expresión neutral [7], dormido o guiño. Otras bases de datos, como la FERET Database [12,13], resultan más adecuadas para implementar la clasificación de imágenes atendiendo al género humano dado que, además de contar con un gran número de sujetos de ambos sexos, contienen una anotación detallada sobre el género y el tipo de expresiones individuales de cada uno de ellos. Dado que en este trabajo se utilizan como bases de datos FERET, Yale Face Database y JAFFE, a continuación se hará un breve resumen de sus características.

La base de datos FERET [12, 13] (*Face Recognition Technology*) fue desarrollada durante la década de los noventa dentro un programa desarrollado por el Departamento de Defensa de Estados Unidos cuyo objetivo final consistía en la implementación de un sistema de reconocimiento facial automático que sirviese de ayuda a las fuerzas de seguridad en el desempeño de sus funciones. Esta base de datos contiene 1564 conjuntos de imágenes, con un total de 14126 imágenes. De ellos, 1199 contienen imágenes de sujetos diferentes mientras que los 365 conjuntos restantes son los denominados "conjuntos duplicados de imágenes" dado que sus imágenes corresponden a personas ya retratadas en la base de datos pero que es común que se tomasen en días diferentes a las imágenes originales [30].

Por su parte, Yale Face Database [9, 32] es una base de datos con 165 imágenes en escala de grises pertenecientes a 15 individuos diferentes, de los cuales se dispone de una imagen para cada configuración o expresión que se recoge en la base de datos: iluminación central, con gafas, iluminación derecha, iluminación izquierda, sin gafas, feliz, normal, triste, dormido, sorprendido y guiño. Para el propósito de este trabajo, se han descartado las configuraciones que no están relacionadas con el reconocimiento de expresiones faciales, por lo que el número de expresiones para cada sujeto se verá reducida a seis (figura 6).



Figura 6: Ejemplo de expresiones faciales en Yale Face Database

La base de datos JAFFE (*Japanese Female Facial Expression*) surge a raíz de los progresos llevados a cabo por Michael Lyons, Miyuki Kamachi, y Jiro Gyoba [14] y sus fotos fueron tomadas por el equipo del departamento de psicología de la universidad de Kyushu. Cuenta 213 imágenes de 10 sujetos de género femenino diferentes, de las que se han tomado entre tres y cuatro imágenes por cada una de las siete expresiones faciales contempladas en la base de datos. En este caso, se han considerado las seis expresiones básicas (enfado, desagrado, miedo, felicidad, tristeza y sorpresa) y, de forma adicional, la expresión neutral (figura 7).



Figura 7: Ejemplo de expresiones faciales en JAFFE Database

Anexo 3: Líneas futuras

A continuación se enumeran algunas de las posibles líneas futuras a implementar con el objetivo de mejorar el sistema presentado:

1. Análisis e implementación de otros tipos de métodos de extracción de características a mayores de HOG y LBP, como pueden ser LPQ (*Local Phase Quantification*) [31] o las características Haar Like [15].
2. Catalogación utilizando otro tipo de clasificadores como, por ejemplo, AdaBoost [21].
3. Entrenamiento con un mayor número de imágenes.
4. Implementación del sistema usando una red neuronal para comprobar si es posible mejorar los resultados en las bases de datos utilizadas en este trabajo.

Anexo 4: Tablas de resultados





Yale Face Database

A continuación se presentan todas las tablas con los resultados obtenidos en los experimentos realizados con la base de datos *Yale Face Database*.

Tras la detección de las diferentes regiones de interés, o áreas relevantes de las que se extraerán los descriptores, se extraen estas regiones y se reescalan para que todas tengan el mismo tamaño. El tamaño utilizado en la primera estrategia (extracción de descriptores de las regiones de ojos y de boca de forma independiente para posteriormente concatenar los descriptores en un vector único) ha sido de 100 cols x 20 filas para ojos y 80 cols x 40 filas para boca. En las otras dos estrategias, las regiones se reescalan a un tamaño de 80 cols x 80 filas.

Porcentaje de fallo en la detección de rasgos faciales

Tabla VII: Resultados de la detección de rasgos faciales en Yale Face Database Original

	Porcentaje de acierto	Porcentaje de fallo	Ejemplo de acierto	Ejemplo de fallo
Detección de ojos	96,67%	3,33%		
Detección de bocas	97,78%	2,22%		
Total	94.4%	5.6%		

Primer experimento

En este primer experimento se separan los distintos sujetos de la base de datos en tres grupos. En HOG se utiliza un tamaño de ventana grande, de modo que el número de descriptores es bajo. En la extracción de características LBP la imagen no se divide en ventanas, obteniendo 256 descriptores para cada imagen.

Tabla VIII: Resultados del reconocimiento de expresiones faciales concatenando los descriptores obtenidos para las regiones de ojos y boca.

			HOG+SVM	LBP+SVM
Configuración utilizada	Ojos	Tamaño ventana	Blocksize: 50x10 Cellsize: 25x5	No se utilizan
		Número descriptores	324	256
	Boca	Tamaño ventana	Blocksize: 50x10 Cellsize: 25x5	No se utilizan
		Número descriptores	324	256
Porcentaje de éxito	Sobre imágenes de entrenamiento		69.41%	77.65%
	Sobre imágenes de prueba (1)		37.5%	50%
	Sobre imágenes de prueba (2)		50%	45.83%
	Sobre imágenes de prueba (3)		51.28%	48.72%
	Promedio		46.26%	48.18%

Tabla IX: Resultados del reconocimiento de expresiones faciales concatenando ojos y boca en una imagen única.

		HOG+SVM	LBP+SVM
Configuración utilizada	Tamaño ventana	Blocksize: 64x64 Cellsize: 32x32	No se utilizan
	Número descriptores	324	256
Porcentaje de éxito	Sobre imágenes de entrenamiento	98.82%	47.05%
	Sobre imágenes de prueba (1)	62.5%	37.5%
	Sobre imágenes de prueba (2)	75%	29.17%
	Sobre imágenes de prueba (3)	53.84%	35.90%
	Promedio	63.78%	34.19%

Tabla X: Resultados del reconocimiento de expresiones faciales para imagen total (recortando solo cara).

		HOG+SVM	LBP+SVM
Configuración utilizada	Tamaño ventana	Blocksize: 64x64 Cellsize: 32x32	No se utilizan
	Número descriptores	324	256
Porcentaje de éxito	Sobre imágenes de entrenamiento	98.82%	34.12%
	Sobre imágenes de prueba (1)	58.33%	12.5%
	Sobre imágenes de prueba (2)	66.67%	12.5%
	Sobre imágenes de prueba (3)	58.97%	17.95%
	Promedio	61.32%	14.31%

Segundo experimento

En este segundo experimento, en la extracción de características LBP se dividirá la imagen en ventanas y en HOG se utilizarán ventanas de menor tamaño con el objetivo de mejorar los resultados obtenidos en el primer experimento. Del mismo modo que en el anterior experimento, la base de datos se divide en tres grupos de imágenes diferentes.

Tabla XI: Resultados del reconocimiento de expresiones faciales concatenando los descriptores obtenidos para las regiones de ojos y boca.

			HOG+SVM	LBP+SVM
Configuración utilizada	Ojos	Tamaño ventana	Blocksize: 20x4 Cellsize: 10x4	10x5
		Número descriptores	810	6912
	Boca	Tamaño ventana	Blocksize: 16x8 Cellsize: 8x8	8x8
		Número descriptores	810	9216
Porcentaje de éxito	Sobre imágenes de entrenamiento		97.65%	98.82%
	Sobre imágenes de prueba (1)		66.67%	54.16%
	Sobre imágenes de prueba (2)		70.83%	66.67%
	Sobre imágenes de prueba (3)		48.72%	43.59%
	Promedio		62.07%	54.80%

Tabla XII: Resultados del reconocimiento de expresiones faciales concatenando ojos y boca en una imagen única.

			HOG+SVM	LBP+SVM
Configuración utilizada	Tamaño ventana		Blocksize: 20x20 Cellsize: 10x10	10x10
	Número descriptores		1764	12544
Porcentaje de éxito	Sobre imágenes de entrenamiento		98.82%	98.82%
	Sobre imágenes de prueba (1)		66.67%	58.33%
	Sobre imágenes de prueba (2)		70.83%	66.67%
	Sobre imágenes de prueba (3)		61.54%	52.28%
	Promedio		66.35%	59.09%

Tabla XIII: Resultados del reconocimiento de expresiones faciales con imagen total (recortando solo cara).

			HOG+SVM	LBP+SVM
Configuración utilizada	Tamaño ventana		Blocksize: 20x20 Cellsize: 10x10	10x10
	Número descriptores		1764	12544
Porcentaje de éxito	Sobre imágenes de entrenamiento		98.82%	98.82%
	Sobre imágenes de prueba (1)		45.83%	50%
	Sobre imágenes de prueba (2)		75%	62.5%
	Sobre imágenes de prueba (3)		64.10%	51.28%
	Promedio		61.64%	54.78%

Tercer experimento, entrenando sistema con N-1 sujetos

En este último experimento realizado para el reconocimiento de expresiones faciales, se utiliza un conjunto mayor de imágenes de entrenamiento, concretamente se realiza un experimento de tipo “dejando uno fuera” (*leave-one-out*). Además, se añade una nueva variante para la extracción de características LBP: el LBP uniforme. Por último, se comprueba si el uso conjunto de LBP y HOG mejora significativamente los resultados.

Tabla XIV: Resultados del reconocimiento de expresiones faciales concatenando los descriptores obtenidos para las regiones de ojos y boca.

				HOG+SVM	LBP+SVM	LBP Uniforme+SVM
Configuración utilizada	Ojos	Tamaño ventana		Blocksize: 20x4 Cellsize: 10x4	10x5	10x5
		Número descriptores		810	6912	1593
	Boca	Tamaño ventana		Blocksize: 16x8 Cellsize: 8x8	8x8	8x8
		Número descriptores		810	9216	2124
Porcentaje de éxito		Sobre imágenes de entrenamiento	de	97.65%	98.82%	98.82%
		Sobre imágenes de prueba (1)		83.33%	50%	50%
		Sobre imágenes de prueba (2)		80%	60%	60%
		Sobre imágenes de prueba (3)		66.67%	83.33%	83.33%
		Sobre imágenes de prueba (4)		20%	60%	60%
		Sobre imágenes de prueba (5)		33.33%	50%	50%
		Sobre imágenes de prueba (6)		50%	33.33%	33.33%
		Sobre imágenes de prueba (7)		50%	83.33%	83.33%
		Sobre imágenes de prueba (8)		66.67%	66.67%	66.67%
		Sobre imágenes de prueba (9)		66.67%	83.33%	66.67%
		Sobre imágenes de prueba (10)		66.67%	50%	50%
		Sobre imágenes de prueba (11)		80%	60%	60%
		Sobre imágenes de prueba (12)		66.67%	83.33%	83.33%
		Sobre imágenes de prueba (13)		50%	50%	50%
		Sobre imágenes de prueba (14)		66.67%	50%	50%
		Sobre imágenes de prueba (15)		66.67%	50%	50%
		Promedio		56.44%	60.88%	59.77%

Tabla XV: Resultados del reconocimiento de expresiones faciales concatenando ojos y boca en una imagen única.

Configuración utilizada	Tamaño ventana	HOG+SVM	LBP+SVM	LBP Uniforme+SVM
		Blocksize: 20x20 Cellsize: 10x10	10x10	10x10
	Número descriptores	1764	12544	2891
Porcentaje de éxito	Sobre imágenes de entrenamiento	98.82%	98.82%	98.82%
	Sobre imágenes de prueba (1)	83.33%	66.67%	66.67%
	Sobre imágenes de prueba (2)	60%	60%	60%
	Sobre imágenes de prueba (3)	83.33%	83.33%	83.33%
	Sobre imágenes de prueba (4)	40%	40%	40%
	Sobre imágenes de prueba (5)	66.67%	33.33%	33.33%
	Sobre imágenes de prueba (6)	83.33%	50%	50%
	Sobre imágenes de prueba (7)	83.33%	83.33%	83.33%
	Sobre imágenes de prueba (8)	66.67%	50%	50%
	Sobre imágenes de prueba (9)	66.67%	83.33%	83.33%
	Sobre imágenes de prueba (10)	66.67%	50%	50%
	Sobre imágenes de prueba (11)	60%	80%	80%
	Sobre imágenes de prueba (12)	66.67%	66.67%	66.67%
	Sobre imágenes de prueba (13)	75%	50%	50%
	Sobre imágenes de prueba (14)	66.67%	50%	50%
	Sobre imágenes de prueba (15)	66.67%	66.67%	66.67%
	Promedio	69%	60.89%	60.89%

Tabla XVI: Resultados del reconocimiento de expresiones faciales con imagen total (recortando solo cara).

Configuración utilizada	Tamaño ventana	HOG+SVM	LBP+SVM	LBP Uniforme+SVM	Aciertos LBP Uniforme+HOG+SVM
		Blocksize: 20x20 Cellsize: 10x10	10x10	10x10	
	Número descriptores	1764	12544	2891	4655
Porcentaje de éxito	Sobre imágenes de entrenamiento	98.82%	98.82%	98.82%	98.82
	Sobre imágenes de prueba (1)	83.33%	66.67%	83.33%	83.33%
	Sobre imágenes de prueba (2)	80%	40%	40%	80%
	Sobre imágenes de prueba (3)	83.33%	83.33%	83.33%	83.33%
	Sobre imágenes de prueba (4)	80%	80%	80%	80%
	Sobre imágenes de prueba (5)	50%	33.33%	33.33%	50%
	Sobre imágenes de prueba (6)	50%	50%	50%	50%
	Sobre imágenes de prueba (7)	50%	50%	50%	50%
	Sobre imágenes de prueba (8)	50%	66.67%	66.67%	50%
	Sobre imágenes de prueba (9)	100%	66.67%	66.67%	100%
	Sobre imágenes de prueba (10)	66.67%	100%	100%	66.67%
	Sobre imágenes de prueba (11)	80%	60%	60%	80%
	Sobre imágenes de prueba (12)	83.33%	50%	50%	83.33%
	Sobre imágenes de prueba (13)	75%	50%	50%	75%
	Sobre imágenes de prueba (14)	33.33%	50%	50%	50%
	Sobre imágenes de prueba (15)	83.33%	50%	50%	66.67%
	Promedio	69.89%	59.78%	60.89%	69.89%





JAFFE Database

A continuación se presentan todas las tablas con los resultados obtenidos en los experimentos realizados con la base de datos *JAFFE Database*.

Tras la detección de las diferentes regiones de interés, o áreas relevantes de las que se extraerán los descriptores, se extraen estas regiones y se reescalan para que todas tengan el mismo tamaño. El tamaño utilizado en la primera estrategia (extracción de descriptores de las regiones de ojos y de boca de forma independiente para posteriormente concatenar los descriptores en un vector único) ha sido de 128 cols x 64 filas para ojos y 64 cols x 32 filas para boca. En las otras dos estrategias, las regiones se reescalan a un tamaño de 128 cols x 128 filas.

Porcentaje de fallo en la detección de rasgos faciales

Tabla XVII: Resultados de la detección de ojos en JAFFE Database

	Porcentaje de acierto	Porcentaje de fallo	Ejemplo de acierto	Ejemplo de fallo
Detección de ojos	88.73%	11.27%		
Detección de bocas	99.53%	0.47%		
Total	88.26%	11.74%		

Primer experimento

Del mismo modo que en el primer experimento realizado para Yale Face Database, en la extracción de características LBP, la imagen no se divide en ventanas. Por otra parte, en la extracción de descriptores HOG se dividirá la imagen en ventanas de gran tamaño, por lo que el número de descriptores será bajo (del orden de cientos).

Tabla XVIII: Resultados del reconocimiento de expresiones faciales concatenando los descriptores obtenidos para las regiones de ojos y boca.

			HOG+SVM	LBP+SVM
Configuración utilizada	Ojos	Tamaño ventana	Blocksize: 64x32 Cellsize: 32x16	No se utilizan
		Número descriptores	324	256
	Boca	Tamaño ventana	Blocksize: 64x32 Cellsize: 32x16	No se utilizan
		Número descriptores	324	256
Porcentaje de éxito	Sobre imágenes de entrenamiento		99.47%	72.34%
	Sobre imágenes de prueba (1)		52.5%	45%
	Sobre imágenes de prueba (2)		62.5%	40.28%
	Sobre imágenes de prueba (3)		56.58%	38.16%
	Promedio		57.19%	41.15%

Tabla XIX: Resultados del reconocimiento de expresiones faciales concatenando ojos y boca en una imagen única.

		HOG+SVM	LBP+SVM
Configuración utilizada	Tamaño ventana	Blocksize: 64x64 Cellsize: 32x32	No se utilizan
	Número descriptores	324	256
Porcentaje de éxito	Sobre imágenes de entrenamiento	99.46%	43.09%
	Sobre imágenes de prueba (1)	47.5%	35%
	Sobre imágenes de prueba (2)	61.11%	38.89%
	Sobre imágenes de prueba (3)	55.26%	31.58%
	Promedio	54.62%	35.16%

Tabla XX: Resultados del reconocimiento de expresiones faciales con imagen total (recortando solo cara).

		HOG+SVM	LBP+SVM
Configuración utilizada	Tamaño ventana	Blocksize: 64x64 Cellsize: 32x32	No se utilizan
	Número descriptores	324	256
Porcentaje de éxito	Sobre imágenes de entrenamiento	99.46%	17.02%
	Sobre imágenes de prueba (1)	77.5%	17.5%
	Sobre imágenes de prueba (2)	72.22%	16.67%
	Sobre imágenes de prueba (3)	55.26%	17.11%
	Promedio	68.33%	17.09%

Segundo experimento

En este experimento, para la extracción de características LBP, se divide la imagen en ventanas. En el caso de la extracción de descriptores con HOG el tamaño de las ventanas utilizado es menor que en el primer experimento. Igual que en el experimento anterior, la base de datos se divide en tres grupos diferentes, que estarán formados por sujetos distintos de la base de datos.

Tabla XXI: Resultados del reconocimiento de expresiones faciales concatenando los descriptores obtenidos para las regiones de ojos y boca.

			HOG+SVM	LBP+SVM
Configuración utilizada	Ojos	Tamaño ventana	Blocksize: 32x32 Cellsize: 16x16	21x16
		Número descriptores	756	4608
	Boca	Tamaño ventana	Blocksize: 16x6 Cellsize: 8x8	8x16
		Número descriptores	756	2304
Porcentaje de éxito	Sobre imágenes de entrenamiento		100%	100%
	Sobre imágenes de prueba (1)		52.5%	52.5%
	Sobre imágenes de prueba (2)		58.33%	45.83%
	Sobre imágenes de prueba (3)		60.53%	56.58%
	Promedio		57.12%	51.64%

Tabla XXII: Resultados del reconocimiento de expresiones faciales concatenando ojos y boca en una imagen única.

		HOG+SVM	LBP+SVM
Configuración utilizada	Tamaño ventana	Blocksize: 32x32 Cellsize: 16x16	16x16
	Número descriptores	1764	12544
Porcentaje de éxito	Sobre imágenes de entrenamiento	100%	100%
	Sobre imágenes de prueba (1)	55%	47.5%
	Sobre imágenes de prueba (2)	58.33%	50%
	Sobre imágenes de prueba (3)	60.53%	59.21%
	Promedio	57.95%	52.23%

Tabla XXIII: Resultados del reconocimiento de expresiones faciales con imagen total (recortando solo cara).

		HOG+SVM	LBP+SVM
Configuración utilizada	Tamaño ventana	Blocksize: 32x32 Cellsize: 16x16	16x16
	Número descriptores	1764	12544
Porcentaje de éxito	Sobre imágenes de entrenamiento	100%	100%
	Sobre imágenes de prueba (1)	60%	60%
	Sobre imágenes de prueba (2)	69.44%	56.94%
	Sobre imágenes de prueba (3)	59.21%	57.89%
	Promedio	62.88%	58.28%

Tercer experimento, entrenando el sistema con N-1 sujetos

En este tercer y último experimento realizado para el reconocimiento de expresiones faciales se realiza un experimento de tipo “dejando uno fuera” (*leave-one-out*). Además, aunque el número de descriptores utilizados en cada caso es el mismo que en el experimento anterior, se añaden los resultados obtenidos utilizando LBP uniforme, que es una nueva variante para la extracción de características LBP. Por último, se comprueba si el uso conjunto de LBP y HOG mejora significativamente los resultados para la imagen completa, que es el tipo de preprocesado con el que se obtienen los mejores resultados en este último experimento.

Tabla XXIV: Resultados del reconocimiento de expresiones faciales concatenando los descriptores obtenidos para las regiones de ojos y boca.

			HOG+SVM	LBP+SVM	LBP Uniforme+SVM
Configuración utilizada	Ojos	Tamaño ventana	Blocksize: 32x32 Cellsize: 16x16	21x16	21x16
		Número descriptores	756	4608	1602
	Boca	Tamaño ventana	Blocksize: 16x6 Cellsize: 8x8	8x16	8x16
		Número descriptores	756	2304	531
Porcentaje de éxito	Sobre imágenes de entrenamiento		100%	100%	100%
	Sobre imágenes de prueba (1)		52.17%	47.83%	47.83%
	Sobre imágenes de prueba (2)		64.71%	52.94%	52.94%
	Sobre imágenes de prueba (3)		68.75%	56.25%	56.25%
	Sobre imágenes de prueba (4)		47.37%	47.37%	47.37%
	Sobre imágenes de prueba (5)		86.67%	80%	80%
	Sobre imágenes de prueba (6)		52.38%	42.86%	42.86%
	Sobre imágenes de prueba (7)		55%	50%	50%
	Sobre imágenes de prueba (8)		65%	60%	70%
	Sobre imágenes de prueba (9)		31.58%	52.63%	52.63%
	Sobre imágenes de prueba (10)		88.24%	76.47%	70.58%
	Promedio		61.19%	56.64%	57.06%

Tabla XXV: Resultados del reconocimiento de expresiones faciales concatenando ojos y boca en una imagen única.

		HOG+SVM	LBP+SVM	LBP Uniforme+SVM
Configuración utilizada	Tamaño ventana	Blocksize: 32x32 Cellsize: 16x16	16x16	16x16
	Número descriptores	1764	12544	2891
Porcentaje de éxito	Sobre imágenes de entrenamiento	100%	100%	100%
	Sobre imágenes de prueba (1)	52.17%	52.17%	52.17%
	Sobre imágenes de prueba (2)	58.82%	47.06%	47.06%
	Sobre imágenes de prueba (3)	62.5%	50%	50%
	Sobre imágenes de prueba (4)	42.11%	31.58%	31.58%
	Sobre imágenes de prueba (5)	86.67%	86.67%	86.67%
	Sobre imágenes de prueba (6)	52.38%	42.86%	42.86%
	Sobre imágenes de prueba (7)	55%	60%	60%
	Sobre imágenes de prueba (8)	70%	70%	70%
	Sobre imágenes de prueba (9)	42.11%	47.37%	47.37%
	Sobre imágenes de prueba (10)	94.12%	88.24%	88.24%
	Promedio	61.58%	57.60%	57.60%



Tabla XXVI: Resultados del reconocimiento de expresiones faciales con imagen total (recortando solo cara).

		HOG+SVM	LBP+SVM	LBP Uniforme+SVM	LBP Uniforme+ HOG+SVM
Configuración utilizada	Tamaño ventana	Blocksize: 32x32 Cellsize: 16x16	16x16	16x16	
	Número descriptores	1764	12544	2891	4685
Porcentaje de éxito	Sobre imágenes de entrenamiento	100%	100%	100%	100%
	Sobre imágenes de prueba (1)	52.17%	47.83%	43.47%	56.52%
	Sobre imágenes de prueba (2)	64.71%	64.71%	64.71%	64.70%
	Sobre imágenes de prueba (3)	81.25%	62.5%	56.25%	75%
	Sobre imágenes de prueba (4)	73.68%	52.63%	47.36%	73.68%
	Sobre imágenes de prueba (5)	73.33%	66.67%	80%	73.33%
	Sobre imágenes de prueba (6)	52.38%	57.14%	57.14%	57.14%
	Sobre imágenes de prueba (7)	45%	55%	45%	40%
	Sobre imágenes de prueba (8)	75%	70%	80%	70%
	Sobre imágenes de prueba (9)	42.11%	26.32%	26.32%	42.11%
	Sobre imágenes de prueba (10)	94.12%	94.12%	94.12%	94.12%
	Promedio	65.38%	59.69%	59.44%	64.66%

FERET Database

Porcentaje de fallo en la detección de rasgos faciales

Tabla XXVII: Detección facial en FERET

	Porcentaje de acierto	Porcentaje de fallo	Ejemplo de acierto	Ejemplo de fallo
Detección de caras	71.09%	28.91%		

Detección de género

Tabla XXVIII: Detección de género para FERET con imagen total (recortando solo cara)

			Aciertos HOG+SVM	Aciertos LBP Uniforme+SVM	Aciertos LBP Uniforme+HOG+SVM
Configuración utilizada	Tamaño ventana		Blocksize:32x48 Cellsize: 16x24	16x24	
	Número descriptores		1764	2891	4655
Porcentaje de éxito	Sobre imágenes de entrenamiento	de	71.09%	71.09%	71.09%
	Sobre imágenes de prueba (1)		66.67%	46.67%	66.67%
	Sobre imágenes de prueba (2)		52.63%	44.74%	47.37%
	Sobre imágenes de prueba (3)		65%	60%	65%
	Sobre imágenes de prueba (4)		80%	70%	80%
	Sobre imágenes de prueba (5)		60%	64%	60%
	Sobre imágenes de prueba (6)		70%	70%	70%
	Sobre imágenes de prueba (7)		40%	50%	50%
	Sobre imágenes de prueba (8)		88.90%	77.78%	88.90%
	Sobre imágenes de prueba (9)		80%	30%	80%
	Sobre imágenes de prueba (10)		65%	60%	70%
	Sobre imágenes de prueba (11)		80%	80%	80%
	Sobre imágenes de prueba (12)		70%	70%	70%
	Sobre imágenes de prueba (13)		70%	50%	70%
	Sobre imágenes de prueba (14)		50%	50%	50%
	Promedio		67.01%	58.79%	67.71%