

Identificación de personas aplicando visión por computadora

Ivonne López¹, Jorge de la Calleja², Antonio Benitez², Ma. Auxilio Medina²

²Cuerpo Académico de Sistemas y Cómputo Inteligente, ¹Universidad Politécnica de Puebla, 3er Carril del Ejido "Serrano" S/N, San Mateo Cuanalá, Juan C. Bonilla, Puebla, México. C.P. 72640.

{ivonne.lopez, jorge.delacalleja, antonio.benitez, maria.medina}@uppuebla.edu.mx

Resumen

La visión por computadora es una disciplina de la inteligencia artificial que permite realizar diferentes tareas de forma automática como reconocer, clasificar o identificar objetos. En este trabajo se presenta un estudio experimental para la identificación de personas usando algoritmos de aprendizaje automático y análisis de componentes principales.

1. Introducción

La identificación de personas es un proceso que realiza el ser humano día con día, a través de diferentes medios, como puede ser la voz, huellas dactilares o una tarjeta de identificación. A menudo se utiliza el rostro para realizar el reconocimiento entre una persona y otra, el cual puede presentarse de manera física o en una imagen. Sin embargo al incrementarse el número de personas a identificar esta tarea se vuelve compleja, tediosa y en ocasiones genera confusión para determinar que una persona es quien dice ser. Por lo tanto se han propuesto diferentes soluciones para automatizar el proceso de identificación de personas empleando algoritmos, tanto de visión por computadora como de aprendizaje automático.

En 1997 K. Etemad y R. Chellappa [8], se utilizaron el algoritmo de linear discriminant analysis (LDA), demostrando que no es necesario un rango amplio de características para un reconocimiento competitivo y preciso, aunque en este caso se realizaron experimentos en un conjunto de datos de imágenes pequeño, en el cual se obtuvo un 92% de precisión. Jian Yang et al [7] en 2004 describieron la técnica 2DPCA, para la representación de la imagen. A diferencia de PCA, 2DPCA se basa en matrices de imagen 2D en lugar de vectores 1D por lo que la

imagen matriz no necesita ser transformada en un vector antes de la extracción de características. Para probar 2DPCA y evaluar el rendimiento, se llevó a cabo una serie de experimentos, en donde se usaron tres bases de datos de imágenes de la cara: ORL, AR y bases de datos de cara Yale. El reconocimiento promedio en todas las pruebas fue mayor utilizando 2DPCA que PCA. En 2006, C. Aguerrebere et al [2] presentaron un sistema biométrico de reconocimiento, utilizando como característica biométrica una imagen digital del rostro de una persona. El sistema estuvo basado en características locales, las cuales fueron obtenidas con filtros de Gabor como descriptor de la cara, inspirado en el algoritmo EBGm. Se evaluó el sistema y distintas alternativas estudiadas con la base estándar FERET. El desempeño obtenido fue comparable con los resultados más significativos en esta base, obteniendo un porcentaje de 88.6% de precisión. En 2007 S. Liao et al [9], propusieron una nueva representación de LBP denominada Multiscale Block Local Binary Pattern (MB-LBP) y aplicada en el reconocimiento de rostros, obteniendo mejores patrones de características redefiniendo la uniformidad de los mismos. Utilizaron AdaBoost para clasificar los rostros. MB-LBP se probó en la base de datos de Face Recognition Grand Challenge (FRGC) versión 2.0 obteniendo un porcentaje del 96.07% de precisión. En 2012, J. Shermina y V. Vasudevan [6], realizaron un sistema de reconocimiento facial basado en la oclusión y la expresión parcial. La extracción de características para el reconocimiento del rostro se realizó usando PCA (Principal Component Analysis). Los resultados experimentales mostraron que el método propuesto alcanzó una relación de reconocimiento de más de 97.2% de precisión. En el mismo año G. Günlü [5], mejoró el tiempo de ejecución del reconocimiento de rostros mediante el algoritmo de Viola-Jones y el método de DCT (Discrete Cosine Transform). La selección de las características más representativas de



Figura 1. Ejemplo de pre-procesamiento de imágenes de rostros: a) imágenes originales y b) imágenes recortadas.

las imágenes del rostro se realizó mediante la división de la misma en subregiones. Cantidades cambiantes de coeficientes discriminativos eran seleccionados de cada subregión utilizando el algoritmo de LDA (Linear Discriminant Analysis) y luego empleados para el reconocimiento. Se realizaron pruebas de los métodos combinados y se obtuvo una tasa del 97.4% de precisión en la base de datos de cara FRGC.

En este trabajo se propone un método de clasificación de rostros usando la técnica de análisis de componentes principales para reducir su dimensionalidad y obtener los atributos que las caractericen. Y para la clasificación se aplican métodos de aprendizaje automático: árboles de decisión, k-vecinos más cercanos, clasificador naive Bayes, redes neuronales y máquinas de vectores de soporte.

El resto del artículo está organizado de la siguiente manera: en la sección 2 se presenta la metodología a seguir para realizar la identificación de rostros. En la sección 3 se muestran los resultados y finalmente en la sección 4 algunas conclusiones y trabajo futuro son presentados.

2. Metodología de identificación de rostros

El método que se propone en este trabajo para realizar la clasificación de rostros de los diferentes conjuntos de datos está dividido en cuatro etapas: 1) creación de la base de datos de imágenes, 2) pre-procesamiento de imágenes, 3) caracterización de imágenes y creación de archivos .arff de los diferentes conjuntos y 4) clasificación de rostros con Weka. A continuación en las siguientes subsecciones se describen a detalle cada una de las etapas.

2.1 Creación de la base de datos

Se elaboró una base de datos de rostros de personas con un total de 200 imágenes. Estos datos pertenecen a

cinco personas y por cada una de ellas se tienen 40 imágenes de su rostro.

Las imágenes fueron tomadas de frente y tienen diferentes expresiones, por ejemplo: normal, feliz, enojado y repugnancia; además las condiciones de la iluminación varía en el conjunto de datos. Cabe mencionar que las imágenes no presentan las mismas condiciones en el ambiente. En este caso ambiente se debe de entender, como el fondo, la iluminación, posición y distancia de la cual fue tomada la imagen.

2.2 Pre-procesamiento de imágenes

La razón de realizar el pre-procesamiento es obtener imágenes que contengan únicamente el rostro de la persona, ya que por lo general las imágenes fueron tomadas de los hombros hacia arriba o contiene un fondo. Se sabe que una imagen es una matriz de píxeles, los cuales están agrupados por filas y columnas, en este caso se necesita trabajar únicamente con los píxeles que definan al rostro.

Por lo anterior, se realiza la localización del rostro en la imagen usando OpenCV¹. El algoritmo empleado es una cascada de clasificación ya entrenada con características que distinguen a un rostro: ojos, nariz y boca. Una vez detectado el rostro se extraen las coordenadas de su ubicación y se recorta la imagen, con el objetivo de tener sólo el rostro. La Figura 1 muestra algunos ejemplos del pre-procesamiento para diferentes imágenes de personas.

2.3 Caracterización de imágenes

Una vez que se tienen recortadas las imágenes es necesario realizar la extracción de información o características que diferencien un rostro del otro. Si se considera a cada píxel como una característica, entonces en una imagen se tienen $m \times n$ características y

¹ OpenCV es una biblioteca libre de visión artificial, <http://www.opencv.org>

se incrementan por el número de imágenes totales de cada persona. Al tener gran cantidad de características por individuo se complica el proceso de clasificación. Sin embargo existen algoritmos de reducción de características, los cuales buscan la información más relevante del conjunto, la extraen y posteriormente se puede almacenar.

En este trabajo se utilizó el algoritmo análisis de componentes principales para buscar y extraer las características más relevantes de conjunto de rostros.

2.3.1 Análisis de componentes principales

El análisis de componentes principales (PCA del inglés Principal Components Analysis) es una técnica inventada por Pearson (1901) y Hotelling (1933) implementada originalmente en 1981. El objetivo principal que persigue dicha técnica es reducir la dimensionalidad de un conjunto de datos con una gran cantidad de variables, ayudándose del estudio de la estructura de varianzas-covarianzas entre las variables que componen los datos de entrada. A partir de la proyección de los datos de entrada sobre las direcciones de máxima varianza se obtendrá un nuevo espacio de representación de los datos en el que se puede eliminar fácilmente aquellas componentes con menor varianza, garantizando la mínima pérdida de información [1].

En este trabajo se utilizó la implementación de la técnica de PCA desarrollada por Matthew Dailey . Como entrada se tiene un archivo de texto que contiene los datos necesarios para realizar la recuperación de información y poder aplicar la técnica de PCA. La salida es una matriz que contiene los datos relevantes por imagen; su tamaño corresponde en columnas al número de características y renglones al número de imágenes del conjunto.

2.3.2 Creación de archivos .arff

Para realizar la clasificación de las imágenes por medio de la herramienta de Weka es necesario organizarlas en un formato denominado ARFF. El formato ARFF (Atributo-Relación File Format) es un archivo de texto ASCII que describe una lista de casos que comparten un conjunto de atributos. La estructura de este archivo consta de tres partes:

- **Cabecera:** Nombre del conjunto de datos o relación (@relation).

- **Declaraciones de atributos:** En esta sección se declaran los atributos que comprenden el archivo junto con su tipo (@attribute)
- **Sección de datos:** Se declaran los datos que componen al conjunto en forma de lista. Cada ejemplo se representa en una sola línea, con retornos de carro que denotan el final del ejemplo. Los valores de atributo para cada ejemplo están delimitados por comas (@data).

```
@relation PCArostroPersonal
@attribute 1 real
@attribute 2 real
@attribute 3 real
@attribute 4 real
@attribute Class {PERSONA1 , DESCONOCIDO}

@data
6949.946270,4077.559366,-1402.473279,-38.581878,PERSONA1
5101.604893,-2935.661116,-5191.070432,2565.183115,PERSONA1
5094.491489,4745.198785,-1989.723077,-394.133522,PERSONA1
6671.745916,-603.465212,-3648.977374,600.100709,DESCONOCIDO
...|
```

Figura 2. Ejemplo de archivo ARFF

En la Figura 2 se presenta un ejemplo del formato ARFF. El primer renglón corresponde al nombre del conjunto de imágenes que para este ejemplo es PCArostroPersonal. Enseguida se enlistan los cinco atributos, cuatro de tipo real y uno de tipo nominal. Finalmente se tiene la lista de los ejemplos con su respectivo valor por atributo.

2.4 Clasificación de rostros con Weka

En esta etapa se realizó la clasificación del conjunto de datos adquirido de la sección anterior, usando diferentes algoritmos de aprendizaje automático implementados en el paquete de herramientas de Weka (Waikato Environment for Knowledge Analysis) .

La clasificación se hace de acuerdo a los procesos de cada algoritmo. A continuación se describen los que se utilizaron en este trabajo.

Árboles de decisión: Un árbol de decisión se puede definir como una estructura de datos con la información necesaria para tomar decisiones de alguna tarea u objetivo específico. Están compuestos por nodos internos que son las pruebas o preguntas para cada valor de los atributos, las ramas son los posibles valores que pueden tomar los atributos, mientras que los nodos hoja

¹ <http://www.cs.ait.ac.th/~mdailey/matlab/>

¹ Plataforma de software para el aprendizaje automático y la minería de datos escrito en Java y desarrollado en la Universidad de Waikato. <http://www.cs.waikato.ac.nz/ml/weka/>

- son el valor final de la prueba realizada desde la raíz [3].
- *Naive Bayes*: El clasificador de Bayes es un algoritmo probabilístico que se basa en la suposición de que todos los valores de los atributos del conjunto de datos son condicionalmente independientes para los valores objetivos. Para clasificar un dato nuevo, este algoritmo le asigna la probabilidad más alta de acuerdo con la función objetivo [4].
- *Redes Neuronales Artificiales*: Se trata de un sistema de interconexión de neuronas artificiales que colaboran entre sí para producir un estímulo de salida. Una red neuronal funciona de la siguiente manera: cada nodo produce un valor de salida, que resulta de combinar los valores de entrada de los nodos que le anteceden, de esta forma se alimenta a los siguientes nodos de las capas siguientes hasta producir un valor final en la capa de salida [10].
- *Máquinas de vectores de soporte*: Pertenecen a la familia de métodos basados en funciones de kernel, que se encargan de mapear un conjunto de datos a un espacio de alta dimensionalidad, donde cada coordenada corresponde a una característica, transformándolo a un conjunto de puntos en un espacio euclidiano [2].
- *k-vecinos más cercanos*: Este tipo de métodos simplemente almacenan todos los datos de entrenamiento disponibles, y cuando se requiere clasificar un dato nuevo, se buscan los datos más parecidos a éste para asignarle el valor (clase) correspondiente [10].

3. Resultados

Del conjunto de datos que se describió anteriormente se obtuvieron diferentes números de atributos que representan el 70%, 80% y 90% de componentes principales, adquiridos por medio de la técnica de PCA, siendo estos 5, 9 y 24 componentes, respectivamente. Cabe mencionar que un solo componente principal alcanza un 47% de la información del conjunto de datos.

Como se señaló anteriormente para realizar la clasificación de los datos de acuerdo a los diferentes números de componentes se utilizó la herramienta de Weka. Los algoritmos se encuentran identificados con los siguientes nombres: J48 (árboles de decisión), Naive Bayes (clasificador de Bayes), Multilayer Perceptron (redes neuronales), SMO (máquinas de vectores de soporte) e IBk (k-vecinos más cercanos). Y para realizar la experimentación se usó la técnica de validación cruzada, en particular 10-fold cross-validation.

Los valores para los algoritmos fueron en el caso de k-vecinos más cercanos se usaron 3 vecinos; para redes neuronales artificiales varía el número de capas ocultas de 5 a 15, el rango de aprendizaje van de 0.4 y 0.5 y las épocas de 10, 50 y 300. Y en el caso de máquinas de vectores de soporte se utilizan kernels polinomiales de grado 2, 3 y 4 así como también la construcción de modelos logísticos. Los algoritmos de árboles de decisión y naive Bayes, se dejaron los parámetros por defecto.

N° Atributos	Persona	Algoritmo de clasificación					Promedio
		K-vecinos	RNA	SVM	Bayes	Árboles	
5	1	96.40%	94.40%	91.90%	91.10%	94.70%	93.70%
	2	96.10%	90.00%	85.60%	81.10%	89.40%	88.44%
	3	96.70%	95.10%	95.20%	95.10%	92.80%	94.98%
	4	90.50%	86.90%	80.00%	74.40%	82.70%	82.90%
	5	96.10%	90.70%	80.20%	83.20%	88.00%	87.64%
	Promedio	95.16%	91.42%	86.58%	84.98%	89.52%	
9	1	98.90%	96.30%	97.80%	94.10%	95.80%	96.58%
	2	97.50%	94.90%	96.80%	91.50%	91.30%	94.40%
	3	96.40%	95.80%	96.00%	95.70%	93.20%	95.42%
	4	95.00%	92.40%	94.80%	89.30%	90.60%	92.42%
	5	98.00%	94.10%	94.90%	87.10%	91.90%	93.20%
	Promedio	97.16%	94.70%	96.06%	91.54%	92.56%	
24	1	97.80%	98.20%	98.90%	94.80%	96.40%	97.22%
	2	97.20%	97.00%	98.00%	93.90%	89.80%	95.18%
	3	97.60%	97.20%	96.50%	95.80%	93.20%	96.06%
	4	97.00%	97.20%	98.00%	92.90%	88.90%	94.80%
	5	98.40%	93.80%	96.10%	92.70%	89.60%	94.12%
	Promedio	97.60%	96.68%	97.50%	94.02%	91.58%	

Tabla 1. Resultados de exactitud de los algoritmos de acuerdo al número de componentes para identificar a cada persona.

Los archivos con formato ARFF contienen únicamente dos clases, la primera clase pertenece a la persona a reconocer y la segunda denominada como desconocido, pertenece al resto del conjunto de datos. Teniendo en cada caso 40 ejemplos positivos y 160 negativos.

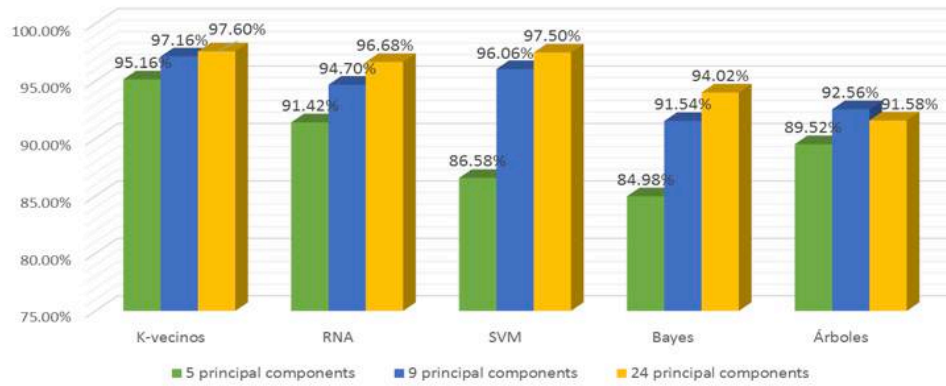


Figura3. Comparativa de algoritmos con mejor exactitud de clasificación

La Tabla 1 muestra la exactitud obtenida de acuerdo al número de atributos y algoritmo utilizado. Esta medida proporciona el porcentaje alcanzado de la clasificación correcta por cada persona. Como se puede observar, los algoritmos que mejor clasificaron los datos fueron, k-vecinos más cercanos, redes neuronales artificiales y máquinas de vectores de soporte.

En la Figura 3 se puede apreciar la comparación entre los tres mejores algoritmos clasificación del conjunto de datos. De esta manera se puede visualizar que entre 9 y 24 componentes principales se obtiene un porcentaje de exactitud que va de 92.4% al 98%. En el caso de k-vecinos más cercanos con este número de atributos la exactitud va del 95% hacia arriba.

4. Conclusiones y trabajo a futuro

Se ha presentado un método de identificación de rostros usando algoritmos de aprendizaje automático y análisis de componentes principales. De acuerdo con los resultados obtenidos se puede concluir que el algoritmo de k-vecinos más cercanos fue el que mejor desempeño presentó. En contraste, el algoritmo de árboles de decisión mostró el resultado más bajo. Cabe mencionar que los algoritmos de redes neuronales artificiales y máquinas de vectores de soporte alcanzan un promedio por encima del 96%, teniendo una diferencia pequeña comparada con el mejor resultado. Como trabajo futuro se experimentará con otros métodos de reducción y obtención de características y

se probará esta misma metodología con el conjunto balanceado.

5. Referencias

- [1] A. Sanchez, Análisis de componentes principales: versiones dispersas y robustas al ruido impulsivo. Proyecto Fin de Carrera. Escuela Politécnica Superior de la Universidad Carlos III de Madrid, 2012.
- [2] C. Aguerrebere, G. Capdehourat, M. Delbracio y M. Mateu, Proyecto Aguará: Reconocimiento de caras, Montevideo, Uruguay, 2006.
- [3] E. Alpaydin, "Introduction to Machine Learning", Cambridge, Massachusetts: The MIT Press, 2004.
- [4] F. R. Bach, G. R. Lanckriet, & M. I. Jordan, "Multiple kernel learning, conic duality, and the SMO algorithm". In Proceedings of the twenty-first international conference on Machine learning. ACM, 2004, p. 6.
- [5] G. Günlü, "Embedded Face Detection", International Journal of Advanced Robotic Systems, vol. 9, pp. 96-112, 28 Junio 2012.
- [6] J. Shermina y V. Vasudevan, "Recognition of the face images with occlusion and expression", International Journal of Pattern Recognition and Artificial Intelligence, World Scientific, vol. 26, n° 03, 2012.
- [7] J. Yang, D. Zhang, A. F. Frangi y J.-y. Yang, "Two-dimensional PCA: a new approach to appearance-based face representation and recognition", Pattern Analysis and Machine Intelligence, IEEE Transactions on IEEE, vol. 26, n° 1, pp. 131-137, 2004.
- [8] K. Etemad y R. Chellappa, "Discriminant analysis for recognition of human face images", J. Opt. Soc. Am. A, vol. 14, n° 8, pp. 1724-1733, 1997.
- [9] S. Liao, X. Zhu, Z. Lei, L. Zhang y S. Z. Li, Learning multi-scale block local binary patterns for face recognition, Seoul, Korea: Springer, 2007, pp. 828-837.
- [10] T. Michell, "Machine Learning". McGraw Hill, 1997.