

Estado del arte en reconocimiento facial

Martín Adrián Garduño Santana, L. E. Díaz-Sánchez, Israel Tabarez Paz,
Marcelo Romero Huertas

Universidad Autónoma del Estado de México, Toluca, México

martingarduno.94@gmail.com, lediazs@uaemex.mx, israeltabarez@gmail.com,
mromeroh@uaemex.mx

Resumen. En este trabajo se resumen los métodos más utilizados para el reconocimiento facial, incluyendo las ventajas y desventajas de los sistemas desarrollados hasta ahora. También se describen las futuras líneas de investigación y se discute el rumbo del reconocimiento facial en los próximos años. Esta revisión es relevante pues se busca la implementación de un novedoso sistema de reconocimiento facial.

Palabras clave: reconocimiento facial, sistemas biométricos, ciudades inteligentes, imágenes 2D y 3D.

Face Recognition: a Survey

Abstract. In this paper, we summarize the most used methods for face recognition, including advantages and drawbacks of each system that has been developed until now. Also, we describe the future research lines and discuss the direction of face recognition. This survey is relevant for our research, as we are encouraged to develop a novel face recognition system as part of our future work.

Keywords: face recognition, biometrics, smart cities, 2D and 3D images.

1. Introducción

El reconocimiento facial (RF) se ha convertido un área importante para la investigación desde la década de los setentas del siglo XX, puesto que varios científicos e ingenieros alrededor del mundo se han enfocado para establecer algoritmos y métodos cada vez más robustos y precisos para este tipo de sistemas y su aplicación en la vida cotidiana. Además, recientemente, el objetivo de desarrollar aplicaciones biométricas, como el reconocimiento facial, es importante en las ciudades inteligentes.

Actualmente existen diversas áreas dentro del funcionamiento de una ciudad inteligente que van desde el entretenimiento hasta seguridad, en las cuales el RF está involucrado. La Figura 1 muestra las áreas y aplicaciones específicas para el RF.

De manera general, el RF se clasifica en dos tipos: verificación o autenticación e identificación o reconocimiento.

Áreas	Aplicaciones específicas
Entretenimiento	Videojuegos, realidad virtual, Interacción Humano-robot y humano-computadora.
Tarjetas Inteligentes	Licencias de conducir, ID's, pasaportes, inmigración.
Seguridad	Inicio de sesión o desbloqueo en dispositivos personales, encriptamiento de archivos, bases de datos.
Cumplimiento de la ley y vigilancia	Video-vigilancia avanzada, control de circuitos cerrados, control de accesos.

Fig. 1. Principales aplicaciones y áreas de desarrollo de sistemas de reconocimiento facial [34].

La verificación consta de comparar la cara de un sujeto, quien introduce su información personal, con una base de datos, el sistema dice si el sujeto es quien dice ser o no. En otras palabras, es una correspondencia uno a uno.

Por otro lado, la identificación es una búsqueda de uno en muchos, pues un rostro desconocido es comparado con toda la base de datos y el sistema establece si el rostro pertenece a algún sujeto dentro de la base de datos o si definitivamente es desconocido. Ciertos sistemas de identificación son capaces de aprender de manera automática, por lo que es posible identificar un rostro que antes era desconocido luego de que el sistema lo haya visto en previas ocasiones añadiéndolo a la base de datos [1, 2, 21, 32].

Ya sea para la verificación o la identificación, todos los métodos para el reconocimiento facial se clasifican en tres categorías:

- Basados en imágenes 2D,
- Basados en imágenes 3D,
- Híbridos (2D+3D).

En los últimos años, se han publicado varios artículos de reseña literaria [1, 6, 24, 34], teniendo en cuenta los métodos de RF más relevantes, algunos también han incluido enfoques multimodales. En esta revisión, damos una introducción y la aplicación potencial de sistemas RF para ciudades inteligentes.

2. Estado del arte

El reconocimiento facial es un trabajo complicado, incluso para los seres humanos, porque solo es posible recordar e identificar un número limitado de caras. La psicología y neurociencias han estudiado como los seres humanos pueden reconocer diferentes rostros [34], por su parte, investigadores en el área computacional e informática buscan que esta tarea sea realizada de manera eficiente mediante computadoras.

En los setenta, los investigadores de la interacción hombre-máquina intentaron automatizar este proceso. Harmon y Kanade fueron los primeros en implementar un sistema capaz de reconocer caras humanas por una computadora. En sus trabajos, [11, 12], Harmon identificó perfiles del rostro dibujados por un artista. En su tesis doctoral, Kanade desarrolló un sistema para reconocer caras con una alta precisión, detectando 608 caras de una base de datos de 670 imágenes [16].

2.1. Reconocimiento facial en 2D

Al principio el RF fue tratado como un reconocimiento de patrones de imágenes en 2D y con las diversas investigaciones se implementó una subcategoría para estos sistemas basados en 2D, misma que se muestra a continuación.

Los métodos basados en características extraen rasgos o marcas locales de cada cara, como ojos, nariz, boca, etc., y se colocan en un clasificador. Por geometría, los métodos genéricos utilizan bordes, líneas y curvas para reconocer caras diferentes. Otros utilizan plantillas o algunas restricciones en las características [34].

Manjunath et al., utilizaron un enfoque basado en características usando los datos de intensidad de la imagen para adquirir las características faciales. El modelo de detección está biológicamente motivado en la corteza visual de los mamíferos y para obtener la información de las características que utilizan una transformación matemática llamada Gabor wavelet [22].

En la aproximación holística el rostro completo del sujeto se toma como datos de entrada en el sistema de reconocimiento [34]. Uno de los métodos holísticos más utilizados hasta ahora, fue presentado por Kirby y Sirovich en 1990, que implementa un modelo matemático conocido como análisis de componentes principales (PCA por sus siglas en inglés), que toma los $N \times N$ píxeles de una imagen de un rostro y los linealiza para crear un vector de longitud N^2 . Después de eso, con una galería o conjunto de entrenamiento se obtiene un rostro medio y se calcula la matriz de covarianza para encontrar sus eigenvectores y eigenvalores. La principal ventaja del PCA es la reducción de la dimensionalidad, debido a que solo se pueden elegir únicamente los vectores propios más grandes y crear un espacio de dimensión $M < N$.

Para averiguar si un nuevo rostro pertenece a un sujeto de la base de datos, este se proyecta en el mismo espacio del conjunto de entrenamiento y la distancia de la imagen proyectada y las imágenes de la base de datos se calcula, si dicha distancia es menor que un parámetro ϵ_k , el rostro pertenece al k -ésimo sujeto de la base de datos [18][33].

Los vectores propios obtenidos fueron llamados por Turk y Pentland como eigenrostros (ver Figura 2). También presentaron una aproximación utilizando eigenfaces para la detección e identificación de caras, además de un sistema de reconocimiento que sigue a la persona y es capaz de identificar si está en una base de datos conocida [32].

Otro método que incluye eigenvectores fue propuesto por Etemad y Chellappa, a través de un modelo lineal conocido como análisis de discriminante lineal (LDA por sus siglas en inglés) que se basa en el análisis de las diferentes características de la cara en términos de su potencial discriminante, este método proporciona una mejor

discriminación entre clases, tratando de maximizar las diferencias entre clases y minimizarla dentro de la misma clase [8]. En 1998, Chao et al., desarrollaron el análisis discriminante de los componentes principales combinando los métodos PCA y LDA [35].



Fig. 2. Eigenrostros calculados por PCA [18].

En la misma línea, los investigadores habían tratado de resolver los problemas de usar PCA, tales como, que no es discriminante entre clases y problemas con las variaciones de pose y de iluminación. En 1997, Belhumer et. al., propusieron la aproximación de Fisherostros basado en el método de discriminante lineal de Fisher, y demostraron que esta técnica tiene tasas de error más bajas que el método de eigenrostros [4].

En [3], Bartlett presenta una generalización de PCA, conocida como análisis de componentes independientes (ICA por sus siglas en inglés) que da una base de imágenes más local, y autores como Déniz et al. combinan esta técnica y PCA con máquinas de soporte vectorial (SVM por sus siglas en inglés), con una mejor aproximación para el reconocimiento [7]. Una investigación reciente por Olivares hace modificaciones a un método similar usando eigenfaces y SVM [23].

Las redes neuronales también se han utilizado para RF, la idea básica es considerar una red con una neurona para cada píxel en la imagen [1]. Lin et al. proponen un sistema basado en una red neural probabilista basada en la toma de decisiones (PDBNN por sus siglas en inglés) y consiste en tres módulos, un detector de cara, que localiza el rostro humano en una imagen, un localizador de ojos y un reconocedor de rostro a partir de una base de datos [20].

El reto de reconocimiento facial del bueno, el malo y el feo (GBUFRC por sus siglas en inglés) es un trabajo reciente para alentar el desarrollo de algoritmos que son robustos a través de los cambios en imágenes fijas frontales 2D. Consiste en tres particiones: el bueno contiene pares de imágenes consideradas fáciles de reconocer con una tasa de verificación base (VR por sus siglas en inglés) de 98%. El malo contiene

pares de imágenes en promedio difíciles de reconocer con un VR de 80%. El feo son pares de imágenes difíciles de reconocer con VR de 15% [26].

2.2. Reconocimiento facial 3D

En las técnicas basadas en imágenes 2D algunas características se pierden debido a la estructura 3D de la cara. Las variaciones en la pose y la iluminación son los principales problemas de las imágenes 2D, por lo que los investigadores han elegido imágenes 3D para el reconocimiento facial y compensar esta desventaja. Hay dos maneras de representar una imagen en 3D, a través de imágenes de rango o imágenes 2.5D y 3D representadas por mallas poligonales. La imagen de rango es una representación de los puntos 3D establecidos (x , y , z), donde los puntos z son un valor de profundidad y corresponden al valor de intensidad de cada píxel en el plano XY (ver Figura 3). Si una imagen 2.5D se toma desde un único punto de vista, solo una superficie facial podría ser modelada en lugar de la cabeza entera. Para un modelo de cabeza completo se requieren varias tomas desde diferentes puntos de vista. La malla poligonal 3D consiste en una lista de vértices conectados bordes [1].

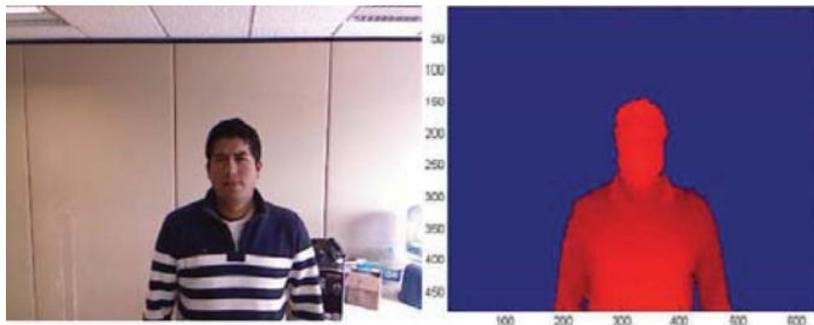


Fig. 3. Imagen de rango tomada mediante el sensor Kinect de Microsoft [28].

La adquisición estéreo, la luz estructural y el sensor láser son tres técnicas de adquisición diferentes para datos 3D. Para la adquisición estéreo, dos o más cámaras se posicionan y calibran para obtener imágenes simultáneas del sujeto.

Entonces los datos de profundidad se calculan a partir de modelos geométricos y con puntos de referencia resuelven el problema de correspondencia. Esta técnica es la más barata y más fácil de usar. La técnica de luz estructural implica un patrón de luz proyectado y la información de profundidad se adquiere a partir de la distorsión del patrón. Es relativamente rápido y barato. La técnica láser es más precisa pero la más cara y la más lenta también. Emplea un sensor láser para escanear.

Generalmente, los datos 3D requieren ser preprocesados después de la adquisición, dependiendo del sensor, hay agujeros, picos y ruido en los datos [2].

Blanz et al., proponen un método basado en el modelo deformable del rostro humano 3D. Su algoritmo simula el proceso de formación de imágenes en el espacio 3D y codifica la forma y la textura en términos de parámetros del modelo. Utilizaron una base de datos de escaneo láser y la identificación se basa en la comparación de

parámetros de coeficientes de una imagen de entrada con los coeficientes de la base de datos [5].

En 2006, Kakadiaris et al., desarrollaron un sistema completamente automático basado en digitalizadores faciales en 3D y reportan una de las más altas precisiones en la base de datos del Face Recognition Grand Challenge (FRGC) [15]. También en 2009, otro sistema 3D fue implementado por Khalid y Lili, el cual extrae rasgos faciales locales de imágenes de rango y calcula la distancia y el ángulo de doce puntos de referencia para la identificación [17]. El uso de Kinect de Microsoft en investigaciones recientes ha facilitado la adquisición de imágenes en profundidad [30]. En su tesis de licenciatura, seis estudiantes de la Universidad Tecnológica de Delft desarrollaron un sistema utilizando el dispositivo Kinect y MATLAB. Dividieron el proceso en tres etapas: adquisición de imágenes, procesamiento de datos y reconocimiento [9, 10, 14].

2.3. Métodos híbridos o multimodales

Los métodos híbridos o biométricos multimodales combinan imágenes en 2D y 3D para obtener un mejor rendimiento [6].

En una investigación reciente se propone el sistema DeepFace, y es posible debido a la combinación de redes neuronales e imágenes 3D, y prueban que este sistema tiene una precisión del 91,4 % [31].

En [6] se muestra una comparación entre sistemas multimodales, 3D y 2D, concluyendo que los multimodales tienen un mejor rendimiento, sin embargo, en estas comparaciones no se controla el número de muestras para cada experimento.

La primera evaluación biométrica multimodal se presentó en el Face Recognition Vendor Test 2006, midiendo los rendimientos en el reconocimiento facial y de iris.

La prueba se enfocó en el reconocimiento de imágenes de alta resolución 2D y 3D tomadas en condiciones controladas y no controladas. Además, por primera vez se comparó la capacidad de reconocimiento entre hombre y máquina, teniendo como resultados que varios algoritmos tienen un mejor rendimiento que las personas [29].

2.4. Base de datos de RF

Como parte importante del desarrollo de sistemas de RF, las bases de datos son utilizadas por los investigadores para evaluar el algoritmo empleado, entre los más utilizados son la Face Recognition Technology (FERET por sus siglas en inglés) [27] y Face Recognition Vendor Test (FRVT) [28, 29]. Face Recognition Grand Challenge es una continuación de FRVT, que introduce pruebas 3D con una base de 50,000 imágenes y aporta cinco conjeturas de diversos criterios de los autores acerca de que método tiene mejor rendimiento [25].

Cada base de datos contiene diferentes imágenes y varios escenarios para evaluar, desde variaciones de iluminación, pose, expresión, cambios de edad o añadir accesorios, como gafas o bufandas. Por otro lado, los investigadores crean sus propias bases de datos y el desempeño del sistema puede ser medido por dos curvas: la curva característica de coincidencia acumulativa (CMC por sus siglas en inglés) y la característica de operación del receptor (ROC por sus siglas en inglés). El primero es

una herramienta para resumir el porcentaje acumulativo de reconocimiento correcto y el segundo resume el espacio de posibles puntos de operación para un sistema de verificación a través de dos tipos de tasas de error [6, 30].

3. Discusión

Los métodos basados en características tienen un problema cuando se trata de detectar características faciales cuando éstas no se encuentran en condiciones normales, es decir, los ojos cerrados, la boca abierta, o con el uso de accesorios. Los métodos basados en imágenes 2D demuestran que son sensibles principalmente a los cambios en la iluminación y la postura. Bajo condiciones controladas, la mayoría de los métodos basados en la imagen 2D alcanzan una precisión superior al 90% pero caen cuando se introducen las condiciones reales.

El método de eigenrostros se ha vuelto una base natural para comenzar el estudio de reconocimiento facial, una vez que se conocen e identifican las deficiencias de este modelo puede pasarse a uno más complejo o buscar las soluciones a sus desventajas.

Las imágenes 3D podrían ser invariantes para los cambios en la pose y la iluminación. En condiciones reales, la mayoría de los sistemas tienen baja precisión, por lo que los investigadores están buscando nuevas técnicas y nuevas tecnologías.

En los próximos años el uso de cámaras de alta calidad y computación de alto rendimiento será una gran ventaja para los sistemas RF y su implementación en ciudades inteligentes.

4. Trabajo a futuro

Estamos interesados en el desarrollo de un sistema de RF por su posible aplicación en ciudades inteligentes. En un trabajo previo, [30], el sensor Kinect se implementó con las bibliotecas Matlab, y esto representa un inconveniente por la licencia requerida de Matlab, por lo que, queremos conectar este dispositivo con OpenSource.

Además, la colaboración con Fiware-México, nos brinda la oportunidad de usar su plataforma y es una manera de implementar un sistema RF en una ciudad inteligente.

Referencias

1. Abate, A., Nappi, M., Riccio, D., Sabatino, G.: 2d and 3d face recognition: A survey. *Pattern recognition letters*, 28(14), pp. 1885–1906 (2007)
2. Akarun, L., Gokberk, B., Salah, A. A.: 3d face recognition for biometric applications. In: *Signal Processing Conference, IEEE, 2005 13th European*, pp. 1–5 (2005)
3. Stewart-Bartlett, M., Movellan, J. R., Sejnowski, T. J.: Face recognition by independent component analysis. *IEEE Transactions on neural networks*, 13(6), pp. 1450–1464 (2002)
4. Belhumeur, P. N., Hespanha, J. P., Kriegman, D. J.: Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on pattern analysis and machine intelligence*, 19(7), pp. 711–720 (1997)

5. Blanz, V., Vetter, T.: Face recognition based on fitting a 3d morphable model. *IEEE Transactions on pattern analysis and machine intelligence*, 25(9), pp. 1063–1074 (2003)
6. Bowyer, K. W., Chang, K., Flynn, P.: A survey of approaches and challenges in 3d and multi-modal 3d+ 2d face recognition. *Computer vision and image understanding*, 101(1), pp. 1–15 (2006)
7. Déniz, O., Castrillon, M., Hernández, M.: Face recognition using independent component analysis and support vector machines. *Pattern recognition letters*, 24(13), pp. 2153–2157 (2003)
8. Etemad, K., Chellappa, R.: Discriminant analysis for recognition of human face images. *JOSA A*, 14(8), pp. 1724–1733 (1997)
9. Fikke, F. S., Gardiner, B. K.: 3d face recognition (image acquisition). Bachelor of Science Thesis, Electrical Engineering (2013)
10. Gerlach, M. M. J., Rooijers, C. T.: 3d face recognition (data processing: Registration and deformation). Bachelor of Science Thesis, Electrical Engineering (2013)
11. Harmon, L. D., Kuo, S. C., Ramig, P. F., Raudkivi, U.: Identification of human face profiles by computer. *Pattern Recognition*, 10(5-6), pp. 301–312 (1978)
12. Harmon, L. D., Hunt, W. F.: Automatic recognition of human face profiles. *Computer Graphics and Image Processing*, 6(2), pp. 135–156 (1977)
13. Heshner, C., Srivastava, A., Erlebacher, G.: A novel technique for face recognition using range imaging. In: *Signal processing and its applications, Proceedings, Seventh international symposium on, IEEE, Vol. 2*, pp. 201–204 (2003)
14. Huijbregts, M. A., Stobbe, B.: 3d face recognition: How to make a fast and reliable database and compare the database with 2d+3d facial input? Bachelor of Science Thesis, Electrical Engineering (2013)
15. Kakadiaris, I. A., Passalis, G., Toderici, G., Murtuza, M. N., Theohari, T.: 3d face recognition. In: *BMVC*, pp. 869–878 (2006)
16. Kanade, T.: *Computer recognition of human faces*, Birkhäuser Basel, Vol. 47 (1977)
17. Khalid, F., Lili, N.: 3d face recognition using multiple features for local depth information. *IJCSNS*, 9(1), pp. 27–32 (2009)
18. Kirby, M., Sirovich, L.: Application of the Karhunen-Loeve procedure for the characterization of human faces. *IEEE Transactions on Pattern analysis and Machine intelligence*, 12(1), pp. 103–108 (1990)
19. Lee, Y., Yi, T.: 3d face recognition using multiple features for local depth information. In: *Video/Image Processing and Multimedia Communications, 4th EURASIP Conference focused on, IEEE, Vol. 1*, pp. 429–434 (2003)
20. Shang-Hung, L., Sun-Yuan, K., Long-Ji, L.: Face recognition/detection by probabilistic decision-based neural network. *IEEE transactions on neural networks*, 8(1), pp. 114–132 (1997)
21. Xiaoguang, L.: Image analysis for face recognition. *Personal notes*, pp. 36 (2003)
22. Manjunath, B. S., Chellappa, R., von der Malsburg, C.: A feature based approach to face recognition. In *Computer Vision and Pattern Recognition, 1992, Proceedings CVPR'92, 1992 IEEE Computer Society Conference on, IEEE*, pp. 373–378 (1992)
23. Olivares-Mercado, J., Toscano-Medina, K., Sánchez-Pérez, G., Nakano-Miyatake, M., Pérez-Meana, H.: Modificaciones del método de eigenphases para el reconocimiento de rostro basado en svm. *Ingeniería, investigación y tecnología*, 17(1), pp. 119–129 (2016)
24. Divyarajsinh, N. P., Brijesh, B. M.: Face recognition methods & applications. arXiv preprint arXiv:1403.0485 (2014)
25. Phillips, P. J.: Face recognition grand challenge. In *Biometric Consortium Conference* (2004)

26. Phillips, P. J., Beveridge, J. R., Draper, B. A., Givens, G., O'Toole, A. J., Bolme, D. S., Dunlop, J., Lui, Y. M., Sahibzada, H., Weimer, S.: An introduction to the good, the bad, & the ugly face recognition challenge problem. In: Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on, IEEE, pp. 346–353 (2011)
27. Phillips, P. J., Grother, P. M., Ross, B., Duane, M., Tabassi, E., Bone, M.: Face recognition vendor test 2002. In: Analysis and Modeling of Faces and Gestures, 2003, AMFG 2003, IEEE International Workshop on, IEEE, pp. 44 (2003)
28. Phillips, P. J., Moon, H., Rizvi, S. A., Rauss, P. J.: The Feret evaluation methodology for face-recognition algorithms. *IEEE Transactions on pattern analysis and machine intelligence*, 22(10), pp. 1090–1104 (2000)
29. Phillips, P. J., Scruggs, W. T., O'Toole, A. J., Flynn, P. J., Bowyer, K W., Schott, C. L., Sharpe, M.: Fvt 2006 and ice 2006 large-scale results. *National Institute of Standards and Technology, NIS- TIR*, 7408(1) (2007)
30. Romero, M., Flores, C., Muñoz, V., Altamirano, L. C.: Face recognition using eigensurface on kinect depth-maps. In: Proceedings of the International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV), The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (World- Comp), pp. 241 (2016)
31. Taigman, Y., Yang, M., Ranzato, M. C., Wolf, L.: Deepface: Closing the gap to human-level performance in face verification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1701–1708 (2014)
32. Turk, M. A., Pentland, A. P.: Face recognition using eigenfaces. In: Computer Vision and Pattern Recognition, proceedings CVPR'91., IEEE Computer Society Conference on, IEEE, pp. 586–591 (1991)
33. Wold, S., Esbensen, K., Geladi, P.: Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1-3), pp. 37–52 (1987)
34. Zhao, W., Chellappa, R., Phillips, P. J., Rosenfeld, A.: Face recognition: A literature survey. *ACM computing surveys (CSUR)*, 35(4), pp. 399–458 (2003)
35. Zhao, W., Krishnaswamy, A., Chellappa, R., Swets, D. L., Weng, J.: Discriminant analysis of principal components for face recognition. In: *Face Recognition*, Springer, pp. 73–85 (1998)