ANALES DE LA
Academia de Ciencias de Cuba

Nuevos métodos para el reconocimiento automático de rostros en video

ENTIDAD EJECUTORA PRINCIPAL: Centro de Aplicaciones de Tecnologías de Avanzada; División de Investigaciones CENATAV, Datys Soluciones Tecnológicas

AUTOR PRINCIPAL: Dr. C. Yoanna Martínez Díaz[1]

Otros autores: Dr. C. Heydi Méndez Vázquez[1], Dr. C. Noslen Hernández González[2], Dr.C. Leonardo Chang Fernández[3]

Colaboradores científicos: Dr. C. Luis Enrique Sucar Succar[4], Dr. C. Rolando Biscay[5], Dr. C. Edel García Reyes[1], Leyanis Lopez[1], Zhenhua Chai[6]

Filiación: [1]Centro de Aplicaciones de Tecnologías de Avanzada. División de Investigaciones CENATAV, Datys Soluciones Tecnológicas; [2]Universidad de Sao Paulo, Brasil; [3]Tecnológico de Monterrey, México; [4]INAOE, México; [5]Centro de Investigación en Matemáticas, Guanajuato, México; [6]Laboratorio Nacional de Reconocimiento de Patrones, NLPR, China

Palabras clave

reconocimiento automático; rostros; video

RESUMEN

El desarrollo de métodos eficientes para el reconocimiento de rostros en videos es fundamental para aplicaciones prácticas de videoprotección en entornos no controlados. En la presente investigación se proponen métodos para el reconocimiento automático de rostros en videos, los cuales permiten que todo el proceso sea más eficiente y que tenga niveles de eficacia similares a los de los métodos documentados en la literatura. Los métodos propuestos son evaluados en bases de datos de pruebas internacionales y han sido validados experimentalmente en escenarios reales de videoprotección de Cuba. Los resultados obtenidos muestran la factibilidad de su uso en estos escenarios, y revisten gran importancia para la defensa y el mantenimiento del orden interior del país.

El reconocimiento de rostros en videos es de suma importancia para diversas aplicaciones como la videoprotección, el monitoreo y el control de accesos [1]. Requiere de diferentes pasos que se ejecutan a partir de la entrada al sistema de una secuencia de imágenes o cuadros que contienen los rostros de las personas que se desea reconocer (Fig. 1). Primero se determina, mediante la detección y el seguimiento, la ubicación de los rostros en cada uno de los cuadros del video; luego se procede a la representación de las secuencias, y finalmente, mediante la comparación, se obtiene una respuesta de la identificación/verificación de los rostros analizados.

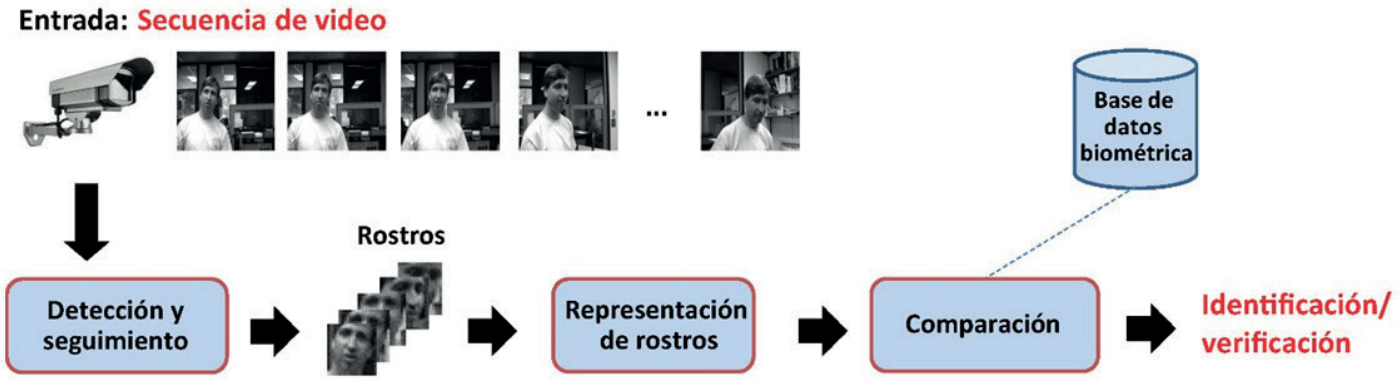

Fig. 1. Esquema general de reconocimiento de rostros en videos.

La mayoría de los métodos para el reconocimiento de rostros en videos propuestos en la literatura [1-3] se centran en obtener altos niveles de eficacia en la etapa de identificación/verificación, descuidando el costo computacional y suponiendo que los rostros han sido detectados y localizados previamente. Para aplicaciones prácticas es necesario contar con algoritmos que sean eficaces y a la vez eficientes, que permitan procesar la mayor cantidad posible de flujos provenientes de las cámaras existentes. En la actualidad, los mejores resultados en cada una de las etapas se han obtenido utilizando métodos basados en redes neuronales profundas [3,4], pero estos tienen varias limitantes para su uso práctico, entre ellas su costo computacional [5].

Para la detección de rostros en video, los métodos basados en técnicas para imágenes fijas son los más utilizados [6]. Entre estos, el método de detección desarrollado por Viola y Jones [7] es uno de los más populares y extendidos, ya que permite detectar rostros con una complejidad computacional muy baja. Sin embargo, poca atención se le ha prestado a la integración de la información temporal y espacial de la apariencia facial para la detección de rostros en videos cuando se ha visto que los rostros presentan una estructura con una distribución espacial que se logra conservar en el tiempo [8].

En el caso del seguimiento, los enfoques basados en el seguimiento mediante la detección [9] han mostrado ser los más apropiados, ya que pueden tratar con escenas complejas y son más robustos ante diferentes variaciones. La idea consiste en aplicar un detector en cada cuadro del video y luego, asociar las detecciones obtenidas para construir la trayectoria correspondiente a cada rostro. Dado que los detectores existentes aún están lejos de ser perfectos, los criterios de decisión para asignar un resultado de detección a una trayectoria determinan el éxito del proceso de seguimiento.

Para la representación de las secuencias de rostros, se han utilizado diferentes tipos de características [2]. A pesar de que algunas de ellas han permitido obtener altos valores de eficacia, la mayoría son costosas computacionalmente. Recientemente, descriptores binarios [10] han sido propuestos para diferentes aplicaciones con el objetivo de minimizar los costos de cómputo y almacenamiento. Este tipo de rasgos resulta una solución atractiva para aplicaciones prácticas, en las que se requiera un procesamiento rápido de grandes volúmenes de datos y donde los recursos de cómputo sean limitados.

Para medir el desempeño de los métodos de reconocimiento, un factor fundamental es el diseño de protocolos de evaluación apropiados, definidos en bases de datos de referencia. No obstante, han sido publicadas pocas bases de datos que cuenten con protocolos que logren capturar los requisitos de los escenarios reales y explotar todos los datos disponibles. Es por esto que, en ocasiones, el desempeño de los métodos actuales en bases de datos de referencia parece estar saturado [2, 3]. En este sentido, los esfuerzos deberían estar centrados en el diseño de nuevos protocolos y no en la recopilación de nuevos datos, que es más costosa y menos factible.

Sobre la base de lo explicado anteriormente y los problemas detectados en la literatura relacionada, en el marco de esta investigación se desarrollaron nuevos métodos que tributan al mejoramiento de la eficiencia del reconocimiento de rostros en aplicaciones de videoprotección. Se extendió el método de detección de Viola y Jones al dominio espaciotemporal, manteniendo sus niveles de eficiencia en el procesamiento y con una mayor eficacia para el caso de la detección de rostros en videos. La estructura del detector propuesto consta de un nuevo descriptor [11] que codifica la información espacial y temporal para representar los patrones facia-

les en un conjunto de cuadros consecutivos, clasificadores *boosting* para seleccionar y aprender de manera automática las características más discriminativas y una cascada de dichos clasificadores *boosting* para acelerar el proceso de la detección. Los experimentos realizados en videos de bases de datos internacionales mostraron que el detector propuesto alcanza una eficacia superior a la del método de Viola y Jones, manteniendo sus niveles de eficiencia, con un menor número de operaciones. Específicamente, se logró aumentar la eficacia en un 10 % en comparación con el uso de características Haar y en un 5 % con respecto al uso de descriptores LBP. La novedad científica de este método está avalada por cuatro artículos publicados [11-14].

Se desarrolló un nuevo método de seguimiento basado en la asociación de detecciones espaciotemporales el cual, al contrario de otros enfoques en los cuales para obtener los fragmentos hay que procesar las detecciones obtenidas cuadro a cuadro, considera la salida del detector espaciotemporal propuesto como fragmentos de trayectorias iniciales confiables. Para vincular los fragmentos y formar las trayectorias finales de cada rostro, que sean consistentes en cuanto a movimiento y apariencia, se resuelve un problema de asociación de datos. Para modelar el movimiento se propuso un filtro de Kalman adaptativo, que ajusta de manera dinámica sus parámetros sobre la base de la confiabilidad del detector. La apariencia de cada fragmento se modela usando el histograma de la representación espaciotemporal computada para esa detección. En los experimentos realizados se observó que, mediante la asociación basada en detecciones espaciotemporales, se obtienen trayectorias de rostros más largas y un menor número de falsas trayectorias que cuando se usan detectores cuadro a cuadro. La novedad científica de este método queda avalada en una publicación [15].

Se diseñó una nueva representación de secuencias de rostros, basada en la codificación vector de Fisher de rasgos binarios, con el fin de disminuir el costo computacional de la representación. Específicamente, se utiliza el descriptor BRIEF como rasgo binario, debido a su marcada simplicidad y almacenamiento compacto en memoria. Una vez que los rasgos BRIEF son extraídos de manera densa y a múltiples escalas de todos los cuadros del video, se utiliza un método de análisis de componentes principales logístico para proyectar dichos descriptores binarios a un espacio vectorial de valores reales y de esta forma poder usar la formulación clásica vector de Fisher basada en GMM, reducir la dimensión de los descriptores locales e incluir la información temporal agregando sus coordenadas espaciales al vector de rasgos proyectado. Los experimentos llevados a cabo en bases de datos internacionales mostraron que la representación propuesta permite disminuir los tiempos de ejecución con una eficacia similar a los métodos relacionados en la literatura para la identificación/verificación de rostros. En particular, se mostró en la evaluación experimental que la codificación vector de Fisher de rasgos binarios BRIEF logra una aceleración de alrededor de tres veces con respecto al tiempo de cómputo de la codificación vector de Fisher de rasgos SIFT, manteniendo una eficacia muy similar. La novedad científica de esta representación está validada por tres artículos publicados [16-18].

Se diseñaron nuevos protocolos de evaluación para la base de datos internacional YouTube Faces que reflejan condiciones similares a las existentes en las aplicaciones prácticas. Específicamente, se propuso un nuevo protocolo de verificación que hace uso completo de la base de datos y proporciona un mayor número de comparaciones, lo que permite la evaluación más real de los métodos, considerando bajos índices de falsos aceptados. Además, se diseñaron protocolos de identificación tanto en conjuntos abiertos como cerrados y teniendo en cuenta diferentes tamaños de galería, así como comparaciones de videos contra videos y contra imágenes. La evaluación de diferentes métodos de la literatura bajo los protocolos propuestos muestra una percepción contraria a cuando se usa el protocolo estándar. Los mejores resultados alcanzados evidencian que la eficacia de reconocimiento aún tiene mucho camino por recorrer. No obstante, el desempeño de los métodos evaluados establece una línea de base para la comparación de futuras investigaciones en el reconocimiento de rostros en video. El aporte científico de los protocolos diseñados queda avalado en una publicación [19].

Para validar la aplicabilidad de los resultados de esta investigación en escenarios reales, se realizaron pruebas en videos provenientes de cámaras de videoprotección de Cuba, lo cual fue parte de los resultados de un trabajo de tesis de doctorado en Ciencias Técnicas [20]. En este trabajo se desarrolló una plataforma experimental que permitió realizar las pruebas en videos reales. Por otro lado, el aporte teórico de este trabajo radica esencialmente en el desarrollo de un conjunto de métodos que permiten llevar a cabo el reconocimiento de rostros en videos de manera eficiente y con una eficacia al nivel de los métodos actuales. Estos métodos no solo poseen novedad científica en el país, sino que además son contribuciones en esta área del conocimiento a nivel internacional, como lo avalan las publicaciones en revistas y congresos de alto prestigio en este tema [11-19].

Estos nuevos métodos son la base teórica del conocimiento para resolver el problema del reconocimiento de rostros en distintos escenarios de videoprotección, lo cual es una necesidad del Ministerio del Interior (MININT). En este sentido, la significación práctica de este trabajo viene dada en primer

lugar, por la posibilidad de usar los métodos propuestos en el desarrollo de sistemas propios para el reconocimiento de rostros en aplicaciones de videoprotección. Precisamente, por la importancia de este tipo de aplicaciones para la defensa y el orden interior de cualquier país, gran parte de los sistemas de videoprotección no aparecen libremente disponibles en el mercado. Además, los sistemas de reconocimiento de rostros existentes en el mercado internacional tienen elevados precios, sin contar que en ocasiones existen restricciones para su adquisición y despliegue en Cuba. En especial, esta investigación básica fue dirigida a buscar soluciones para los problemas relacionados con la videoprotección, donde una de las necesidades es determinar o verificar de manera automática la identidad de las personas a partir de su rostro, según el escenario donde se desarrolle la vigilancia.

Se prevé que las soluciones propuestas en esta investigación sean introducidas en el sistema Xyma Safe Vision que desarrolla la empresa Datys para la videoprotección, con lo cual se le agregaría el valor agregado de reconocer rostros en los escenarios que se analizan, y complementaría las herramientas con las que cuenta el MININT, desarrolladas por la empresa. Otras posibles aplicaciones de interés de los resultados obtenidos están enmarcadas en el reconocimiento de rostros en dispositivos móviles o empotrados, donde se necesitan algoritmos con bajas demandas computacionales debido a las características de este tipo de dispositivos.

Referencias bibliográficas

1. J. R. Barr, K. W. Bowyer, P. J. Flynn, and S. Biswas, "Face Recognition from Video: A Review," IJPRAI, 26, 2012.
2. O. M. Parkhi, K. Simonyan, A. Vedaldi, and A. Zisserman. "A compact and discriminative face track descriptor". In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1693-1700, 2014.
3. J. Yang, P. Ren, D. Chen, F. Wen, H. Li, and G. Hua. "Neural aggregation network for video face recognition". In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 4362-4371, 2017.
4. S. S. Farfade, M. J. Saberian, and Li-Jia Li. "Multi-view face detection using deep convolutional neural networks". In Proceedings of the 5th ACM on International Conference on Multimedia Retrieval, 643-650, 2015.
5. Chen, Jun-Cheng, Rajeev Ranjan, Swami Sankaranarayanan, Amit Kumar, Ching-Hui Chen, Vishal M. Patel, Carlos D. Castillo, and Rama Chellappa. "Unconstrained Still/Video-Based Face Verification with Deep Convolutional Neural Networks." International Journal of Computer Vision 126, no. 2-4 (2018): 272-291.
6. L. Wael and K. N. Plataniotis. "Co-occurrence of local binary patterns features for frontal face detection in surveillance applications". EURASIP Journal on Image and Video Processing, 2011(1):1-17, 2011.
7. Viola, P., Jones, M. Rapid object detection using a boosted cascade of simple features. IEEE Computer Society Conference on CVPR (2001) 511-518.
8. M. Castrillón, O. Déniz, C. Guerra, and M. Hernández. "Encara2: Real-time detection of multiple faces at different resolutions in video streams". Journal of Visual Communication and Image Representation, 18(2):130-140, 2007.
9. B. Wang, G. Wang, K.L. Chan, and L. Wang. "Tracklet association by online target-specific metric learning and coherent dynamics estimation". IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(3):589-602, 2017.
10. M. Calonder, V. Lepetit, Ch. Strecha, and P. Fua. "Brief: Binary robust independent elementary features". In 11th European Conference on Computer Vision (ECCV), 778-792, 2010.
11. H. Méndez-Vázquez, Y. Martínez-Díaz, and Z. Chai. "Volume structured ordinal features with background similarity measure for video face recognition". (ICB 2013), Lectures Notes in Computer Computer Science: Advances in Biometrics, 1-6, 2013.
12. Y. Martínez-Díaz, H. Méndez-Vázquez, and E. García-Reyes, "Detección de rostros: una revisión crítica". Memorias del IX Congreso Nacional de Reconocimiento de Patrones (Compumat), 2011.
13. Y. Martínez-Díaz, H. Méndez-Vázquez, N. Hernández, E. García-Reyes, "Improving Faces/Non- Faces Discrimination in Video Sequences by Using a Local Spatio-Temporal Representation", (ICB 2013), Lectures Notes in Computer Computer Science: Advances in Biometrics, pp. 1-5, 2013.
14. Y. Martínez-Díaz, H. Méndez-Vázquez, N. Hernández, "Face detection in video using local spatio-temporal representations", (CIARP 2014), Lecture Notes in Computer Science: Progress in Pattern Recognition, Image Analysis and Applications, 8827, 860-867, 2014.
15. Y. Martínez-Díaz, N. Hernández, H. Méndez-Vázquez. "Multi-face tracking based on spatio- temporal detections". Intelligent Data Analysis, 20, s1, 141-154, 2016.
16. Y. Martínez-Díaz, L. Chang, N. Hernández, H. Méndez-Vázquez, L.E. Sucar, "Efficient Video Face Recognition by using Fisher Vector of Binary Features", (ICPR 2016), IEEE Computer Society, pp. 1436-1441, 2016.
17. Y. Martínez-Díaz, L. Chang, and H. Méndez-Vázquez, "Reconocimiento de rostros en video mediante la codicación Fisher Vector de rasgos binarios". Memorias del XIV Congreso Nacional de Reconocimiento de Patrones (RECPAT 2016), 2016.
18. Y. Martínez-Díaz, N. Hernández, R.J. Biscay, L. Chang, H. Méndez-Vázquez, L.E. Sucar. "On Fisher Vector Encoding of Binary Features for Video Face Recognition". Journal of Visual Communication and Image Representation, 51: 155-161, 2018.
19. Y. Martínez-Díaz, H. Méndez-Vázquez, L. López-Avila, L. Chang, L.E. Sucar, M. Tistarelli. Toward More Realistic Face Recognition Evaluation Protocols for the YouTube Faces Database. (CVPR 2018), 2018.
20. Y. Martínez-Díaz. "Método para el reconocimiento automático de rostros en aplicaciones de video-protección." Tesis en opción al grado de Doctor en Ciencias Técnicas, Cuba, 2018.

AUTOR PARA LA CORRESPONDENCIA

Dr. C. Yoanna Martínez Díaz. *7a A, núm. 21406 e/ 214 y 216, Reparto Siboney, Playa. La Habana, C.P. 12200. Correo electrónico: ymartinez@cenatav.co.cu*