SISTEMA DE VERIFICACIÓN FACIAL EN DISPOSITIVOS MÓVILES

Ing. Honorio Candelario-Emigdio[1], Dr. José A. Montero-Valverde[2], Dr. Eduardo de la Cruz-Gámez[3], MC. José F. Gazga-Portillo[4], Dra. Miriam. Martínez-Arroyo[5]

***Resumen*- En este trabajo, se muestran los resultados obtenidos del desarrollo de un sistema de verificación facial para la seguridad biométrica mediante. El sistema realiza un proceso de verificación facial, el cual se basa en detectar el rostro, alinearlo, extraer las características y realizar la verificación facial. Asimismo, se utilizan las redes neuronales convolucionales como método para la extracción y representación de vectores de características. Los vectores extraídos mediante la red son representativos de cada rostro y permiten realizar una comparación que devuelva una distancia entre las imágenes almacenadas en la base de datos. La red busca maximizar la distancia entre los vectores de rostros distintos y minimizar la distancia entre los de la misma persona. La distancia entre los vectores almacenados es evaluada tomando en cuenta un punto de corte para determinar si son de la misma persona.**

***Palabras clave*- Redes neuronales convolucionales; Reconocimiento facial; Dispositivos móviles; Biometría.**

Introducción

La seguridad es un tema que ha sido de gran interés para la comunidad científica, el reconocimiento y la verificación de la identidad de las personas es uno de los aspectos fundamentales. Los ataques terroristas ocurridos en los últimos años han demostrado la necesidad de establecer métodos más confiables para verificar la identidad de las personas. Los sistemas biométricos surgen como una solución real a los problemas de verificación. La biometría consiste de un conjunto de métodos automatizados para la verificación de individuos mediante el uso de características físicas o del comportamiento de la persona [3]. Esta tecnología se basa en la premisa de que cada persona es única y posee rasgos distintivos que pueden ser utilizados para identificarla.

El procesamiento automático de imágenes para extraer contenido semántico es una tarea que ha ganado mucha importancia durante los últimos años debido al número cada vez mayor de fotografías digitales en Internet o que se almacenan en computadoras personales. La necesidad de organizarlos de forma inteligente utilizando técnicas de indexación y recuperación de imágenes requiere un análisis efectivo y eficiente y algoritmos de reconocimiento de patrones que sean capaces de extraer información semántica relevante.

Especialmente las caras contienen una gran cantidad de información valiosa en comparación con otros objetos o elementos visuales en las imágenes. Por ejemplo, reconocer a una persona en una fotografía, en general, dice mucho sobre el contenido general de la imagen.

El objetivo principal del análisis de rostros es extraer información valiosa de las caras, como su posición en la imagen, las características, las expresiones, el género o la identidad de la persona.

La visión computacional es el estudio de los procesos de reconocer y localizar objetos usando el procesamiento de imágenes de tal forma que se logre un mayor entendimiento de estos Brandon Amos y otros [1]. Para esto se busca construir tecnologías con dichas capacidades. Asimismo, estos autores plantean que la visión computacional se centra en la extracción de características de la cara para que estas sean entendidas por una computadora. Mediante el entendimiento de estas características, las computadoras pueden determinar la localización de ciertos objetos dentro de una imagen, reconocerlos, clasificarlos o descomponerlos.

El presente trabajo se apoya en el campo de la visión computacional, en especial en los métodos de verificación facial para la creación de un sistema online con dichas funcionalidades (análisis de las características faciales del sujeto extraídas de la imagen). La verificación facial en este trabajo de investigación consiste en que se va a mostrar la

[1] Alumno de la Maestría en Sistemas Computacionales del I.T. de Acapulco
honorio_30@hotmail.com
[2] Profesor Adscrito al Depto. De Sistemas y Computación del I.T. de Acapulco
[3] Profesor Adscrito al Depto. De Sistemas y Computación del I.T. de Acapulco
[4] Profesor Adscrito al Depto. De Sistemas y Computación del I.T. de Acapulco
[5] Profesor Adscrito al Depto. De Sistemas y Computación del I.T. de Acapulco

identidad del rostro en el dispositivo móvil con el fin de tener una mayor aplicación con pocos recursos, además se plantea la utilización de redes neuronales convolucionales ya que mejores resultados están dando en la actualidad [2].

El proceso a utilizar empieza con la obtención de la imagen del rostro por medio del dispositivo móvil. Luego se realiza la alineación del rostro para que se encuentre lo menos rotado posible. A partir de esto, se extraen las características de los rostros alineados y finalmente se logra la verificación facial realizando la búsqueda en una base de datos para la comparación de las características de varias imágenes.

En el presente trabajo, las redes neuronales convolucionales permiten extraer características representativas y diferenciales de dos rostros de tal forma que sean comparables entre sí. Esta tiene como salida un vector de características representativo por cada imagen.

Trabajos relacionados

La identificación biométrica en teléfonos móviles/inteligentes es una de las áreas de investigación activa en sistemas de información seguros e inteligentes. Se han realizado diferentes estudios de investigación sobre las diferentes técnicas biométricas disponibles para teléfonos móviles/inteligentes. Estas técnicas incluyen, reconocimiento de huellas dactilares, reconocimiento de rostro, geometría de la mano, reconocimiento de iris, reconocimiento de voz, reconocimiento de firma y pulsación de teclas, etc.

M. Gargi y otros [4] propusieron un método para brindar seguridad a los teléfonos inteligentes Android que utilizan la función biométrica del iris. El estudio proporciona resultados prometedores en un dispositivo Android con procesador de 1 GHz y 4 GB de memoria interna, y con un tiempo total de 80 a 90 segundos para autenticar a un solo usuario móvil de la base de datos de 75 personas que contienen 5 imágenes de iris de cada persona.

Santo Sierra y otros [5] propusieron un sistema biométrico basado en la geometría de la mano, que está orientado a dispositivos móviles. Los autores afirman que el sistema puede proporcionar resultados precisos en la identificación individual. La investigación muestra la implementación de la biométrica manual en una PC con 2,4 GHz y una plataforma móvil Android con procesador de 1 GHz y 576 Mb de RAM. El resultado de la investigación mostró un buen rendimiento con FAR = 0.089% y FRR = 5.89%. La implementación móvil tardó menos de 3 segundos en proporcionar una identificación de una base de datos de 120 individuos diferentes. Una de las limitaciones de la geometría de la mano es que la geometría de la mano no es muy única y no se puede utilizar para la identificación dentro de una gran población.

Guillaume Dave y otros [6] investigaron el rendimiento de diferentes algoritmos de reconocimiento facial en teléfonos inteligentes. Los autores analizan el rendimiento de los algoritmos aplicándolos a un teléfono Android con procesador de 600 MHz y 256 Mb de RAM. Las pruebas se realizaron con 134 imágenes de caras de 10 personas diferentes. Los resultados indican que logró una tasa de reconocimiento del 94% con algoritmos de fisher-face y no tomó más de 1,6 segundos.

Vázquez-Fernández y otros [7] presentan una aplicación inteligente para compartir fotos para dispositivos móviles basada en el motor de reconocimiento facial. El sistema se implementa en la plataforma Android y se prueba en dos teléfonos inteligentes de diferentes fabricantes, HTC Desire con procesador de 1 GHz y 576 MB de RAM y Samsung Galaxy Tab con procesador de 1 GHz y 512 MB de RAM. Las pruebas se realizaron para 50 contactos con 4 caras por contacto. Los resultados mostraron que la aplicación tardó 0.35 segundos en HTC Desire y 0.47 segundos en el Samsung Galaxy para reconocer la cara.

Estos estudios proporcionan la implementación de diferentes técnicas de identificación biométrica en los teléfonos inteligentes, pero no indican su rendimiento en grandes conjuntos de datos.

Metodología

La verificación facial se basará en seguir el proceso de detectar los rostros, alinearlos, representarlos y clasificarlos [2][1]. Asimismo, se utilizará como apoyo el desarrollo OpenSource de OpenFace [1] el cual incluye un modelo pre-entrenado de red neuronal convolucional basado en FaceNet [8].

Este módulo está relacionado con el algoritmo de red neuronal convolucional y se encuentra en el dispositivo móvil. A este se le hacen consultas de las imágenes faciales para que devuelva resultados de la comparación.

El proceso que se desarrolla en este módulo se explicará a continuación y se resume en la Figura 1.

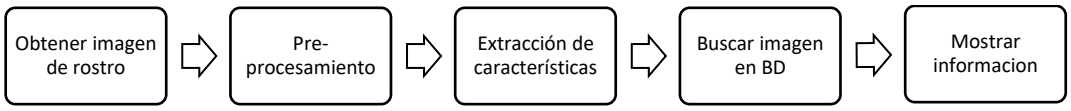

Figura.1. Etapas del procesamiento de imágenes

Obtener imagen

La obtención de la imagen del rostro se hace a través del dispositivo móvil y de manera sencilla dado que, al tomarse de forma vertical, se puede obtener la imagen rotada correctamente sin problemas.

Pre-procesamiento

Reducir el entorno que no es de interés para el problema. Fondo, ruido, etc.

Extracción de características

Seleccionar y extraer "características" apropiadas para la identificación de los objetos deseados. Posteriormente, las imágenes procesadas son enviadas una por la red neuronal convolucional la cual devuelve un vector de 128 dimensiones para cada imagen. La red neuronal convolucional fue extraída de OpenFace [1], está en su mayoría basada en la red establecida por FaceNet [9].

Buscar imagen en la base de datos

En primer lugar, se obtienen las imágenes de una base de datos. Finalmente, se convierte cada archivo en matrices de pixeles para que puedan ser procesados por el resto del servicio.

Mostrar información

Una vez obtenidos los resultados de la red neuronal convolucional, se comparan ambos vectores de características obteniendo la distancia euclidiana entre ambos. En este caso, esta se halla mediante la siguiente fórmula:

$$Distancia\ Euclidiana = \sqrt{\sum_{i=1}^{n}(X_{1i} - X_{2i})^2}$$

Donde X1 y X2 son los dos vectores de características, uno por cada rostro y n es el número de elementos de cada vector. En este caso n es 128 dado que la red neuronal convolucional tiene como salida un vector de características de 128 valores por cada rostro. El resultado es enviado de vuelta al prototipo.

El prototipo da una respuesta al usuario dependiendo del resultado aceptado/rechazado. En caso la distancia sea mayor al punto de corte, determina que son rostros de distintas personas y viceversa.

Una red neuronal convolucional se utiliza para la extracción de características. Es de gran importancia extraer características locales en vez de centrarse en los pixeles de manera específica en una imagen. Muchos objetos pueden aparecer distorsionados o en distintas posiciones haciendo que sea necesario características generales que describan la imagen en su conjunto o por áreas. Por esta razón, se debe dar importancia a las regiones de la imagen para así detectar características en diversos tamaños y posiciones. Este comportamiento puede ser replicado en una red neuronal forzando a las capas ocultas a combinar fuentes de información local de la imagen. De esta forma, distintas características especiales pueden aparecer en distintos lugares de la imagen y ser detectados de igual manera [9].

Las redes neuronales convolucionales son una extensión de las redes neuronales clásicas, pero con más dimensiones al recibir valores matriciales de imágenes en más de un canal. Asimismo, estas redes tienen varias características que

las diferencian entre las que resaltan el compartir pesos entre neuronas y el uso de pesos matriciales por cada neurona [10].

Las redes neuronales convolucionales están subdividas en capas. Las capas más comunes son las siguientes [1]:

- Capas de convolución que deslizan un filtro sobre los valores de características de entrada.
- Capas totalmente conectadas que hallan la sumatoria de los valores de entrada considerando ciertos pesos.
- Capas de submuestreo que suelen obtener el máximo o promedio de regiones espaciales del mapa de características.

La red planteada devuelve un vector de características por imagen. En la fase de verificación facial, la distancia entre ambos vectores es evaluada tomando en cuenta un punto de corte determinando para establecer si se trata de dos fotos de la misma persona.

Resultados parciales

El experimento realizado fue mediante pruebas preliminares del sistema. Donde se pueden presentar cuatro situaciones posibles a la hora de realizar una verificación facial en función de cuál es la clase del usuario genuino o impostor y de la decisión tomada por el sistema verificador dado un umbral de decisión aceptación o rechazo, como se muestra a continuación en la tabla 1.

	Aceptación	Rechazo
Genuino	Verdadero Positivo	Falso Negativo
Impostor	Falso Positivo	Verdadero Negativo

Tabla 1: Situaciones posibles en la verificación facial.

TP (True Positive, traducido como verdaderos positivos) número de predicción correctas de un ejemplo positivo.

FP (False Positive, traducido como falsos positivos) número de predicciones incorrectas de un ejemplo positivo.

TN (True Negative, traducido como verdadero negativo) número de predicciones correctas de un ejemplo negativo.

FN (False Negative, traducido como falso negativo) número de predicciones incorrectas de un ejemplo negativo.

La población del estudio es la población estudiantil del instituto tecnológico mayor a 18 años. La muestra de la investigación es de 336 comparaciones de 42 personas.

A continuación, se discutirán los resultados encontrados en la investigación. A partir de las imágenes de los rostros, se hizo también una comparación entre el rostro de cada persona con todos. De esta forma se obtuvieron resultados en los que se esperaba un rechazo o aceptación.

	Aceptación	Rechazo
Genuino	TP(0.334)	FN(0.666)
Impostor	FP(0)	TN(1)

Tabla 2. Matriz de confusión de la clasificación Genuino e Impostor con Umbral de 0.60

	Aceptación	Rechazo
Genuino	TP(0.762)	FN(0.238)
Impostor	FP(0)	TN(1)

Tabla 3. Matriz de confusión de la clasificación Genuino e Impostor con Umbral de 0.70

	Aceptación	Rechazo
Genuino	TP(0.905)	FN(0.095)
Impostor	FP(0.073)	TN(0.927)

Tabla 4. Matriz de confusión de la clasificación Genuino e Impostor con Umbral de 0.80

	Aceptación	Rechazo
Genuino	TP(0.952)	FN(0.048)
Impostor	FP(0.342)	TN(0.658)

Tabla 5. Matriz de confusión de la clasificación Genuino e Impostor con Umbral de 0.99

Se mostraron los resultados generales obtenidos en los cuales se consideró los puntos de corte en 0.60, 0.70, 0.80 y 0.99. Donde se determino que en el punto de corte 0.99 es el que obtuvo el mayor porcentaje total de resultados correctos o exactitud de todos los casos fue de 95.2%.

A partir de los resultados obtenidos, es posible establecer un punto de corte óptimo al minimizar falsos positivos y falsos negativos. La distancia euclidiana influye drásticamente al momento de tomar la decisión. Esta relación se debe principalmente a un punto de corte establecido. Finalmente, cabe destacar que se obtuvieron resultados aceptables dentro del sistema considerando que solo se contaba con dos muestras de foto por persona de las cuales una era una fotografía extraída a baja resolución del Documento de Identidad.

Conclusión

La gente utiliza múltiples dispositivos móviles como teléfonos inteligentes, tablets o portátiles. Replicar los sistemas de autenticación biométrica a través de los diferentes dispositivos es un punto débil en la experiencia del usuario. Una sola inscripción debe ser suficiente para acceder a diferentes dispositivos y servicios, con el fin de lograr una mejor experiencia de usuario. Desafortunadamente, la operación de un sistema de reconocimiento facial generalmente varía debido al uso de diferentes cámaras y ópticas (dispositivo de captura). Esto señala una pregunta muy relacionada con el punto anterior: el análisis del rendimiento entre dispositivos y cómo el rendimiento puede verse afectado si se utilizan diferentes dispositivos para la inscripción y para la autenticación. Es necesario realizar más investigaciones sobre la autenticación multimodal entre dispositivos para lograr una mejor experiencia de autenticación biométrica móvil.

Como trabajo futuro tenemos la elaboración de una base de datos de imágenes más amplia con diferentes condiciones de iluminación y algunas posiciones del rostro (ejem., frente al dispositivo de captura y con algún ángulo de rotación de la cabeza). Estas bases de datos de imágenes servirán para ver y valorar la repetibilidad del algoritmo y si tiene alguna variación con los porcentajes de autenticación y reconocimiento facial.

Referencias

[1] Amos, B., Ludwiczuk, B., & Satyanarayanan, M. (2016). OpenFace: A general-purpose face recognition library with mobile applications. 1-18.

[2] Taigman, Y., Yang, M., Ranzato, M. A., & Wolf, L. (2014). DeepFace: Closing the Gap to Human-Level Performance in Face Verification. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1701-1708.

[3] P. Reid. BIOMETRICS for Networks Security. Prentice Hall, 2004.

[4] Gargi M, J. Jasmin Sylvia Rani, Madhu Ramiah, N. T. Naresh Babu, A. Annis Fathima and V. Vaidehi. "Mobile Authentication Using Iris Biometrics". Published by Springer Berlin Heidelberg, Networked Digital Technologies, Vol. 294. pp 332-341, 2012.

[5] De Santos Sierra. A, C. Sanchez Avila, A. Mendaza Ormaza, J. Guerra Casanova. Towards Hand Biometrics in Mobile devices. In Proceeding of BIOSIG, Darmstadt, ISBN: 978-3-88579-285-7, 2011..

[6] Dave G, Chao, X., & Sriadibhatla, K. "Face Recognition in Mobile Phones". Department of Electrical Engineering Stanford University, USA, 2010.

[7] Vazquez-Fernandez, Esteban, et al. "Built-in face recognition for smart photo sharing in mobile devices." Multimedia and Expo (ICME), 2011 IEEE International Conference on. IEEE.

[8] Schroff, F., Kalenichenko, D., & Philbin, J. (2015). FaceNet: A Unified Embedding for Face Recognition and Clustering. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 815-823.

[9] LeCun, Y., Boser, B., Denker, J. S., Howard, R. E., Hubbard, W., & Jacket, L. D. (1989). Backpropagation Applied to Handwritten Zip Code Recognition. Neural Computation, 1(4), 541-551. **[10]** L. F. Nicolas-Alonso and J. Gomez-Gil, "Brain Computer Interfaces, a Review", Sensors, vol. 12, no. 2, año 2012, pages 1211-1279.

[10] Karpathy, A., Johnson, J. & Fei-Fei, L (2016). CS231n Convolutional Neural Networks for Visual Recognition. Stanford University. Extraído desde: http://cs231n.stanford.edu/