ISSN 1870-4069

Detección de vida usando características de textura invariantes de Haralick

Oswaldo Vázquez, Ariel Alexis Placido Cabrera, Pedro Arguijo

Tecnológico Nacional de México, campus Misantla,
División de Ingeniería en Sistemas Computacionales, Veracruz,
México

{bigosvaap, arielalexisplacidocabrera}@gmail.com,
pedro_arguijo@excite.com

Resumen. La detección de vida dentro de los sistemas biométricos es una técnica cuyo objetivo es determinar si la biometría que se está capturando es una medición real de la persona viva autorizada que está presente en el momento de la captura. En este trabajo se realiza la detección de vida a través de las características de textura invariantes de Haralick en el dataset de "Spoof in the wild" sobre la región del rostro de los sujetos de prueba. Para la clasificación de las características se usaron cuatro clasificadores: arboles de decisión, random forests, support vector machine y gradient boosting tree teniendo los mejores resultados en random forests y gradient boosting tree con una exactitud general del 96% con un conjunto de entrenamiento del 20%.

Palabras clave: Detección de vida, textura, Haralick.

Liveness Detection Using Haralick Invariant Texture Features

Abstract. Life detection within biometric systems is a technique whose objective is to determine if the biometric being captured is a true measurement of the authorized living person present at the time of capture. In this work, life detection is carried out through Haralick invariant texture features in the Spoof in the wild dataset over the face region of the test subjects. Four classifiers were used for feature classification: decision trees, random forests, support vector machine and gradient boosting tree, obtaining the best results with random forests and gradient boosting tree. The overall accuracy of 96% with a training set of 20%.

Keywords: Live detection, texture, Haralick.

1. Introducción

Junto a la digitalización de diferentes servicios como la banca se ha creado una necesidad de medidas de seguridad contra los ataques de engaño. La biometría es una de estas medidas de seguridad que se ha adoptado para combatir dichos ataques. Algunas de las técnicas conocidas para identificación son el reconocimiento facial, el

reconocimiento de huellas dactilares, la verificación de la escritura, la geometría de la mano, escáner de retina e iris. Entre estas técnicas, la que se ha desarrollado rápidamente en los últimos años es la tecnología de reconocimiento facial por ser más directa, fácil de usar y conveniente en comparación con otros métodos. Por lo tanto, se ha aplicado a varios sistemas de seguridad.

Pero, en general, los algoritmos de reconocimiento facial no son capaces de diferenciar si la persona realmente se encuentra *in situ* y el rostro detectado corresponde a una persona que se encuentra en vivo en el lugar (live) o es alguna fotografía de dicha persona que intenta engañar al sistema haciendo creer que la persona se encuentra en el sitio (spoof), esto es un problema de seguridad importante. Es una forma fácil de falsificar los sistemas de reconocimiento facial por medio de imágenes del rostro. Con el fin de evitar este tipo de falsificación, un sistema seguro necesita detección de vida (*liveness detection*) [1].

La biometría es la tecnología que permite establecer la identidad de un individuo basándose en atributos físicos característicos de la persona. La importancia de la biometría en la sociedad moderna ha sido reforzada por la necesidad de sistemas de gestión de la identidad en gran escala cuya funcionalidad depende de la deducción exacta de la identidad de un individuo en el marco de varias aplicaciones.

Algunos ejemplos de estas aplicaciones son el intercambio de recursos informáticos en red, conceder acceso a las instalaciones sensibles, realizar transacciones financieras a distancia o abordar un vuelo comercial [2]. La principal tarea de un sistema de seguridad es la verificación de la identidad.

La razón principal de esto es evitar que los impostores accedan a recursos protegidos. Las técnicas generales para fines de seguridad son las contraseñas o los mecanismos de tarjetas de identificación, pero estas técnicas de identidad pueden perderse, obstaculizarse o ser robadas fácilmente, lo que perjudica a la seguridad prevista. Con la ayuda de las propiedades físicas y biológicas de los seres humanos, un sistema biométrico puede ofrecer más seguridad para un sistema de seguridad [1].

El problema de la falsificación debe ser resuelto antes de que los sistemas de reconocimiento facial puedan ser ampliamente aplicado en nuestra vida diaria. Para distintos tipos de métodos de detección de vida, en la interacción hombre-computadora es casi indispensable detectar el movimiento biológico de los usuarios.

Los movimientos más utilizados incluyen el parpadeo de los ojos [3, 4], la rotación de la cabeza [4, 5], y el movimiento de la boca [6]. Uno de los principales problemas de estos es que los usuarios deben ser altamente cooperativos y la duración de la detección de la vida es relativamente larga, lo que hacen que los usuarios se sientan incómodos al usar dichos sistemas.

En los sistemas biométricos, el objetivo de las pruebas de vida es determinar si la biometría que se está capturando es una medición real de la persona viva autorizada que está presente en el momento de la captura. Si bien los sistemas biométricos pueden tener un excelente rendimiento y mejorar la seguridad, estudios previos han demostrado que no es difícil engañar a los dispositivos biométricos mediante dedos falsos, imágenes o vídeo de alta resolución, lentes de contacto, etc.

Aunque los dispositivos biométricos utilizan información fisiológica para fines de identificación/verificación, estas mediciones rara vez indican la vida útil. La detección de la vida útil reduce el riesgo de falsificación al requerir una firma de vida útil además de la información biométrica correspondiente. Los métodos pueden incluir mediciones

médicas como la oximetría de pulso, un electrocardiograma o el olor. En unos pocos casos, la información sobre la vida es inherente a la propia biometría, es decir, esta medida no puede capturarse a menos que el usuario esté vivo, por ejemplo, el electrocardiograma es un método biométrico que solo puede usarse si la persona está viva, ya que este registra la actividad eléctrica del corazón.

Si bien el algoritmo de liveness dificulta la suplantación de identidad, es necesario considerarlo como un componente de un sistema biométrico que trae consigo características de rendimiento, así como factores como la facilidad de uso, la aceptación del usuario, la universalidad, la posibilidad de suplantación de identidad, la permanencia, etc.

Ningún sistema es perfecto en cuanto a su capacidad para prevenir los ataques de engaño. Sin embargo, los algoritmos de liveness pueden reducir esta vulnerabilidad para reducir al mínimo el riesgo de falsificación [7].

El uso del reconocimiento facial para la autenticación es cada vez más frecuente, especialmente en los dispositivos móviles. Pero entre el fácil acceso a las imágenes en los medios sociales y los avances en la resolución de imágenes digitales e impresas, los sistemas biométricos tienen lagunas de seguridad que los estafadores pueden aprovechar para falsificar con éxito un sistema de reconocimiento facial.

Para que la biometría facial pueda realmente ser adoptada por la mayoría como un mejor modo de autenticación, es esencial determinar si el rostro presentado es auténtico o si se trata de un intento de falsificar el sistema presentando una representación artificial del mismo.

Así pues, la detección automatizada de los ataques de presentación y, concretamente, la detección de la autenticidad se ha convertido en un componente necesario de cualquier sistema de autenticación que se base en la biometría facial para su verificación. El reconocimiento de vida facial ha surgido como una forma de detener el fraude y asegurar la integridad de la biometría facial como medio de autenticación. Mientras que el reconocimiento facial para la autenticación puede responder con precisión a la pregunta "¿Es esta la persona correcta?" no responde a la pregunta "¿Es esta una persona real?". Esta es la función de la detección de la vida.

La mayoría de las tecnologías actuales de detección de la vida facial son "activas", y requieren que los usuarios parpadeen, giren la cabeza o muevan el teléfono de un lado a otro. Esto da lugar a tres problemas: En primer lugar, los estafadores pueden presentar una foto recortada con agujeros en los ojos, utilizar una máscara o mostrar un vídeo para engañar al sistema. En segundo lugar, las técnicas de desafío-respuesta ponen a los atacantes en alerta de que están siendo revisados.

Y, por último, los métodos activos crean fricciones que ralentizan el proceso de autenticación, aumentan las tasas de abandono y disminuyen la experiencia general del usuario. En este trabajo proponemos determinar la detección de vida en imagenes del rostro considerando las matrices de coocurrencia de niveles de gris como una función de densidad de probabilidad discreta, tal como lo propuso Löfstedt et al. [8]. Evaluamos el desempeño de las características extraídas de la región de interés con árboles de decisión, bosques aleatorios (random forest, RF), máquinas de soporte vectorial (support vector machines, SVM) y potenciación del gradiente (gradient boosting tree, GBT).

El artículo se distribuye de la siguiente manera: en la Sección 2 describimos la metodología empleada, en la sección 3 se presentan los resultados y en la sección 4 se incluyen las conclusiones de esta investigación.

2. Trabajos previos

La detección en vivo es crucial a la hora de sistemas biométricos como la detección de rostro, permitiendo garantizar que la persona reconocida se encuentra realmente en el sitio, para la detección en vivo se han propuesto diferentes enfoques como el uso de características de textura, movimiento como parpadeo, boca u ojos, forma, reflejos, color y análisis en el dominio de la frecuencia. Los rasgos de textura se extraen de las imágenes de los rostros bajo el supuesto de que los rostros impresos producen ciertos patrones de textura que no existen en los reales.

La textura es probablemente la evidencia más fuerte de la falsificación, ya que más del 69% de las obras utilizan la textura sola o la combinan con otros descriptores en sus contramedidas [9]. El análisis de microtextura fue implementado por Jukka y otros [10]. La idea clave es enfatizar las diferencias de micro textura en el espacio de los rasgos. Los autores adoptan los patrones binarios locales (LBP), que es un poderoso operador de texturas, para describir las micro texturas y su información espacial. Los vectores en el espacio del rasgo se dan entonces como entrada a un clasificador SVM que determina si los patrones de microtextura caracterizan una imagen falsa o una imagen de persona viva.

Los descriptores de movimiento son los segundos en importancia para la detección de la suplantación de la cara, y hay dos formas diferentes de considerar el movimiento para este propósito. Una forma es detectar y describir las variaciones interfaciales, como el parpadeo de los ojos, las expresiones faciales y la rotación de la cabeza [9]. La técnica basada en el análisis del movimiento de los ojos fue introducida por Hyung-Keun Jee y otros para el sistema de reconocimiento facial incorporado [11]. Los autores propusieron un método para detectar los ojos en imágenes de entrada secuenciales y luego se calcula la variación de cada región ocular y se determina si la cara de entrada es real o no.

La suposición básica es que, debido al parpadeo y a los movimientos incontrolados de las pupilas en los ojos humanos, debería haber grandes variaciones de forma [12]. A su vez también se encuentran trabajos donde se combinan dos o más métodos como en [13], donde utilizaban un método para detectar ataques de "spoofing" facial combinando la detección de texturas de HSV (Tono, Saturación, Valor) y la detección del movimiento del parpadeo de los ojos.

El algoritmo antispoofing consiste en tres módulos principales: el detector de parpadeo, el detector de HSV y el módulo de puntuación combinada. La visión general del flujo de trabajo es que, en primer lugar, la imagen del rostro capturada por la cámara se comprueba de antemano si el rostro que se ha obtenido es el verdadero o un falso rostro resultante de un ataque de falsificación de rostro. La imagen del rostro es una captura cuadro a cuadro donde cada cuadro es el resultado de una captura de cámara de 30 fps (cuadros por segundo).

Los resultados de esta foto se analizan a través del módulo detector de parpadeo para producir un valor de puntuación de parpadeo.

(a) (b)

Fig 1. Ejemplos del dataset. a) Vivo b) Engaño. Tomado de [15].

Fig 2. RGB a Escala de grises.

3. Metodología

3.1. Dataset

El dataset considerado en este trabajo es el reportado por Liu, et. al. [14], el cual contiene 165 videos de personas reales y videos con imágenes falsas de las mismas considerando diversas variaciones. Para cada sujeto, se tienen 8 videos en vivo y hasta 20 falsos, en total 4,478 videos. Todos los videos están en 30 cuadros por segundo, aproximadamente 15 segundos de duración y resolución HD de 1080P.

Los videos en vivo se recogen en cuatro sesiones con variaciones de distancia, pose, iluminación y expresión. Los videos falsos se recopilan con varios ataques, como papel impreso y reproducción [15]. Todos los archivos de video se identifican como SubjectID_SensorID_TypeID_MediumID_SessionID .mov (o *.mp4). SubjectID varía de 001 a 165. SensorID representa el dispositivo de captura. TypeID representa el tipo de falsificación del video. MediumID y SessionID registran detalles adicionales del video, en la Fig. 1 se puede ver un ejemplo del dataset.

También se proporciona un archivo del cuadro delimitador de caras con el mismo nombre del video correspondiente (es decir, SubjectID_SensorID_Ty peID_MediumID_ SessionID.face). En cada archivo *.face, contiene una matriz de 4 por n, donde n es la longitud del marco. Cada fila registra las ubicaciones (x, y) de la

esquina superior izquierda y la esquina inferior derecha del cuadro delimitador actual, como [785 425 1070 710]. [0,0,0,0] significa que no se ha detectado ningún rostro [10].

3.2. Regiones de interés

Nuestra región de interés corresponde al área de rostro, la cual está descrita por las ubicaciones (x, y) descritas en la sección anterior. Para la extracción del rostro se itero en cada fotograma del video y posteriormente se extrajo el rostro usando las coordenadas correspondientes obtenidas del archivo *.face con el que cuenta cada video.

3.3. Matriz de coocurrencia de niveles de gris

Una vez obtenida nuestra región de interés es necesario calcular la matriz de coocurrencia de niveles de gris (GLCM, por sus siglas en inglés) para la cual se debe transformar la imagen RGB a una imagen de intensidad o de niveles de gris, para lo cual se utilizó la siguiente expresión

$$gray = 0.229R + 0.587G + 0.114B,$$

donde R, G y B corresponden a los componentes del rojo, verde y azul, respectivamente de las imágenes a color. En la Fig. 2 se muestra el efecto de transformar una imagen de RGB a escala de grises. La matriz GLCM se crea calculando la frecuencia con la que un píxel con el valor de intensidad (nivel de gris) i aparece en una relación espacial específica con un píxel con el valor j. Cada elemento (i, j) en la matriz GLCM resultante es simplemente la suma del número de veces que el píxel con valor i ocurrió en la relación espacial especificada con un píxel con valor j en la imagen de entrada.

Para calcular la matriz, es necesario definir una distancia y una dirección, además de los pares de píxeles separados esa distancia, tal como se muestra en la Fig. 3. A partir de esta matriz se pueden obtener distintas características de segundo orden: contraste (variación local de intensidad en una imagen), correlación (medida de la dependencia lineal de los niveles de gris), autocorrelación (evaluación tanto de la regularidad como de la tosquedad de la textura), cluster prominence y cluster shade (ofrecen información sobre el grado de simetría), energía (mide la repetición del píxel y expresa la regularidad de la textura), entropía (representa la irregularidad en la distribución de los valores de intensidad y es inversa a la energía), diferencia de entropía, diferencia de la varianza, solo por mencionar algunas.

3.4. Características invariantes de Haralick

Debido a la carga computacional para obtener las matrices de co-ocurrencia, los niveles de gris se cuantifican. En consecuencia, las características resultantes dependen en gran medida de la cuantificación. Si la cuantificación es idéntica, se puede garantizar la reproducibilidad de las características.

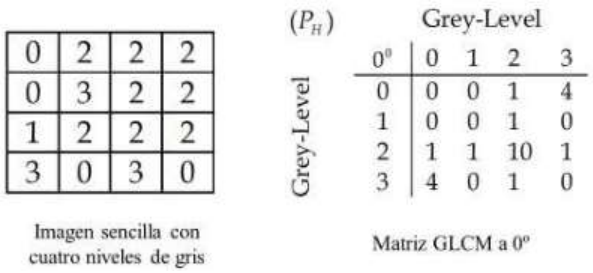

Fig. 3. Ejemplo para obtener la matriz GLCM con distancia par y dirección 0º.

Al redefinir la matriz de co-ocurrencia como una función de densidad de probabilidad discretizada [8], definió GLCM que son asintóticamente invariantes al número de niveles de gris. Con este enfoque, demostraron que las características invariantes tienen una mayor precisión, y tienen rendimientos similares incluso si las imágenes de entrenamiento y de prueba tenían una cuantización bastante diferente.

Una descripción detallada de las 21 características invariantes de textura de Haralick y su cálculo se puede encontrar en [8]. Entre estas están: autocorrelación, contraste, correlación, energía, entropía, homogeneidad, entre otras. Este conjunto de características ha mostrado un mejor desempeño en los clasificadores que las características originales.

3.5. Clasificadores

Para la clasificación se usó cuatro tipos de clasificadores Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM) y Gradient Boosted Tree (GDT). Del dataset [15] se extrajeron un total de 100817 imágenes divididas en dos clases: live and spoof. Todos los clasificadores indicados se probaron con diferentes proporciones en los datos para entrenamiento y prueba: 20/80, 30/70, 40/60, 50/50, 60/40, 70/30 y 80/20.

Para el algoritmo de DT se usó el algoritmo de CART, RF usa un conjunto de tres decisiones para predecir la clase. Cada árbol de decisión ha sido entrenado en un subconjunto aleatorio del conjunto de entrenamiento y solo usa un subconjunto aleatorio de las características, el número de árboles es de 100.

SVM separa los datos de entrenamiento en dos clases utilizando un hiperplano de margen máximo, el kernel utilizado es Radial Basis Function, en cuanto al clasificador GBT está formado por un conjunto de árboles de decisión individuales, entrenados de forma secuencial, de forma que cada nuevo árbol trata de mejorar los errores de los árboles anteriores. La predicción de una nueva observación se obtiene agregando las predicciones de todos los árboles individuales que forman el modelo.

4. Resultados

En la Tabla 1, se muestras los resultados de exactitud obtenidos con los diversos clasificadores considerando las razones de entrenamiento y pruebas mencionados anteriormente. En la tabla podemos ver los resultados con el conjunto de datos de

Tabla 1. Resultados obtenidos por diferentes conjuntos de entrenamiento y prueba, y distintos clasificadores.

Train / Test	DT		RF		SVM		GBT	
	Train	Test	Train	Test	Train	Test	Train	Test
20 / 80	0.925	0.931	0.966	0.970	0.954	0.958	0.966	0.962
30 / 70	0.945	0.943	0.975	0.974	0.951	0.957	0.966	0.962
40 / 60	0.950	0.949	0.980	0.978	0.962	0.960	0.977	0.972
50 / 50	0.952	0.952	0.982	0.981	0.962	0.961	0.974	0.974
60 / 40	0.957	0.957	0.984	0.984	0.960	0.961	0.977	0.975
70 / 30	0.966	0.961	0.987	0.987	0.962	0.961	0.975	0.976
80 / 20	0.961	0.961	0.988	0.987	0.965	0.963	0.977	0.974

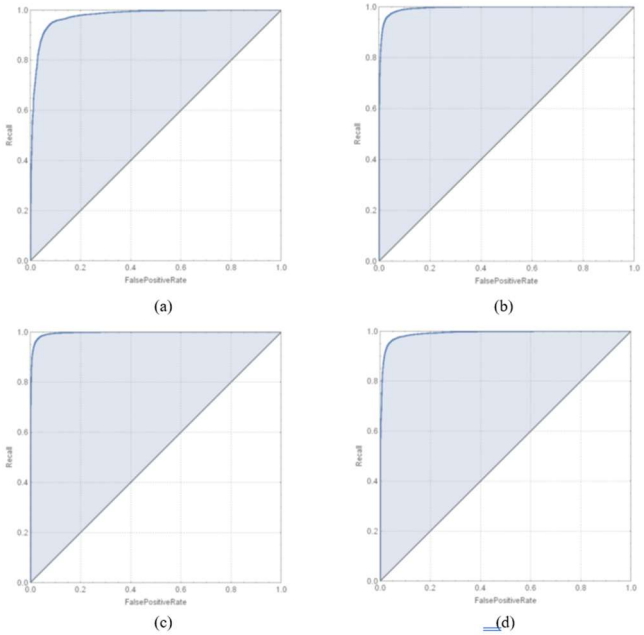

Fig 4. Curvas ROC. a) Decision Tree, b) Random Forest, c) Support Vector Machine, d) Gradient Boosting Tree.

entrenamiento y prueba de los algoritmos de Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM) y Gradient Boosting Tree (GBT), vemos que los resultados de entrenamiento son muy similares a los resultados en la parte de prueba, por lo tanto, el algoritmo no se sobre ajustan al conjunto de entrenamiento y generalizan de forma aceptable el conjunto de datos de prueba. Un ejemplo de las curvas ROC obtenidas con los clasificadores mencionados se muestra en la Fig. 4, la proporción considerada fue la de 20/80.

5. Conclusiones y trabajo futuro

En este trabajo se buscó la detección de vida en el rostro a través del análisis de textura, existen diferentes métodos que se han usado para la detección de vida como el parpadeo de los ojos, hacer que la persona haga ciertos movimientos cada vez que se

desea comprobar esto, lea algún texto para confirmar que es ella realmente el problema con estos enfoques es que solo funcionan teniendo al sujeto frente a la cámara por cierto periodo de tiempo y no son implementables en situaciones donde se necesite comprobar que una fotografía fue tomada realmente a la persona, además de que el método de análisis de textura es implementable en sistemas de video dado que un video se puede ver como una serie de fotografías a alta velocidad, el objetivo principal se cumplió al poderse comprobar que con este enfoque se pueden obtener buenos resultados. Como trabajo futuro se busca la optimización del algoritmo y creación de una librería o API para la fácil implementación en diversos sistemas.

Referencias

1. Chakraborty, S., Das, D.: An overview of face liveness detection. arXiv preprint arXiv:1405.2227 (2014) doi: 10.48550/arXiv.1405.2227
2. Jain, A. K., Flynn, P., Ross, A. A.: Handbook of biometrics. Springer (2008)
3. Pan, G., Sun, L., Wu, Z., Lao, S.: Eyeblink-based Anti-spoofing in face recognition from a genericwebcamera. In: Proceedings of 11th IEEE International Conference on Computer Vision, pp. 1–7 (2007) doi: 10.1109/ICCV.2007.4409068
4. Kollreider, K., Fronthaler, H., Bigun, J.: Verifying liveness by multiple experts in face biometrics. IEEE Computer Vision and Pattern Recognition Workshops, Anchorage, pp. 1–6 (2008) doi: 10.1109/CVPRW.2008.4563115
5. Kollreider, K., Fronthaler H., Bigun, J.: Evaluating liveness by face images and the structure tensor. Fourth IEEE Workshop on Automatic Identification Advanced Technologies, pp. 75–80 (2005) doi: 10.1109/AUTOID.2005.20
6. Chetty, G., Wagner, M.: Liveness verification in audio-video speaker authentication. In: Proceedings of 10th Australian Int. Conference on Speech Science and Technology (2004)
7. Li S. Z., Jain A. K.: Encyclopedia of biometrics. Springer, Boston (2009)
8. Löfstedt, T., Brynolfsson, P., Asklund, T., Nyholm, T., Garpebring, A.: Gray-level invariant Haralick texture features. PloS one, vol. 14, no. 2, pp. e0212110 (2019) doi: 0.1371/journal.pone.0212110
9. Souza, L., Oliveira, L., Pamplona, M., Papa, J.: How far did we get in face spoofing detection? Engineering Applications of Artificial Intelligence, vol. 72, pp. 368–381 (2018) doi: 10.1016/j.engappai.2018.04.013
10. Maatta, J., Hadid, A., Pietikainen, M.: Face spoofing detection from single images using microtexture analysis. In: Proceedings of. International Joint Conference on Biometrics, pp. 1–7 (2011) doi: 10.1109/IJCB.2011.6117510
11. Jee, H. K., Jung, S. U., Yoo, J. H.: Liveness detection for embedded face recognition system. International Journal of Biological and Medical Sciences, vol. 1, no. 4, pp. 235–238 (2006)
12. Hadiprakoso, R. B.: Face Anti-Spoofing Method with Blinking Eye and HSV Texture Analysis. In: Proceedings of IOP Conference Series: Materials Science and Engineering, IOP Publishing, vol. 1007, no. 1, pp. 012034, (2020)
13. Liu, Y., Jourabloo, A., Liu, X. Learning deep models for face anti-spoofing: Binary or auxiliary supervision. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 389–398 (2018)