

Tipo de artículo: Artículo de revisión  
Temática: Tecnologías de la información y las telecomunicaciones  
Recibido : 22/02/2022 | Aceptado: 31/05/2022

## **Revisión de los métodos de reconocimiento facial en imágenes RGB-D adquiridas mediante un sensor Kinect**

Review of facial recognition methods in RGB-D images acquired using a Kinect sensor

Alejandro Alvarez Navarro <sup>1\*</sup> <https://orcid.org/0000-0001-8426-2638>

Enrique Juan Marañón Reyes <sup>1</sup> <https://orcid.org/0000-0002-6640-5487>

Rubén Orozco Morales <sup>2</sup> <https://orcid.org/0000-0002-6240-1569>

<sup>1</sup> Universidad de Oriente. Ave. De Las Américas SN. Santiago de Cuba. Cuba. {alejandroa, enriqueem}@uo.edu.cu

<sup>2</sup> Universidad Central "Marta Abreu. Carretera a Camajuaní Km. 5 y 1/2. Santa Clara. Villa Clara. Cuba. rorozco@uclv.edu.cu

\* Autor para la correspondencia. ([alejandroa@uo.edu.cu](mailto:alejandroa@uo.edu.cu))

---

### **RESUMEN**

El reconocimiento facial en una imagen o video, es un tema que ha sido investigado y desarrollado en los últimos años. Sin embargo, continúa siendo una tarea abierta con grandes complejidades, desde el punto de vista científico y económico. Se investiga para desarrollar o mejorar técnicas y algoritmos que permitan tener mayor efectividad que las existentes y presentar menores costos computacionales. Se plantean tres enfoques, los que se basan exclusivamente en imágenes RGB, de profundidad, o la fusión de ambas modalidades. El objetivo de este trabajo es ofrecer una panorámica de las propuestas de reconocimiento

facial utilizando un sensor Kinect para adquirir las imágenes, realizar un análisis crítico y comparativo de las principales propuestas y analizar los resultados obtenidos. A partir de este análisis veremos los principales aportes y evaluaremos los principales problemas y retos de investigación por resolver. En los artículos revisados en el período 2012-2021 se proponen diversos descriptores para la extracción de características y un conjunto de clasificadores.

**Palabras clave:** Reconocimiento facial; Kinect; descriptores; extracción de características; clasificadores.

## ABSTRACT

Facial recognition in an image or video is a topic that has been researched and developed in recent years. However, it continues to be an open task with great complexities, from the scientific and economic point of view. Research is being carried out to develop or improve techniques and algorithms that are more effective than the existing ones and have lower computational costs. Three approaches are proposed, those based exclusively on RGB, depth images, or the fusion of both modalities. The objective of this work is offer a panoramic view of the proposals of face recognition using a Kinect sensor to acquire the images, realize a critical and comparative analysis of the main proposals and analyze the results obtained. From this analysis we will see the main contributions and evaluate the main research problems and challenges to be solved. Several descriptors for feature extraction and a set of classifiers are proposed in the articles reviewed in the period 2012-2021.

**Keywords:** Facial recognition; Kinect; descriptors; feature extraction; classifiers.

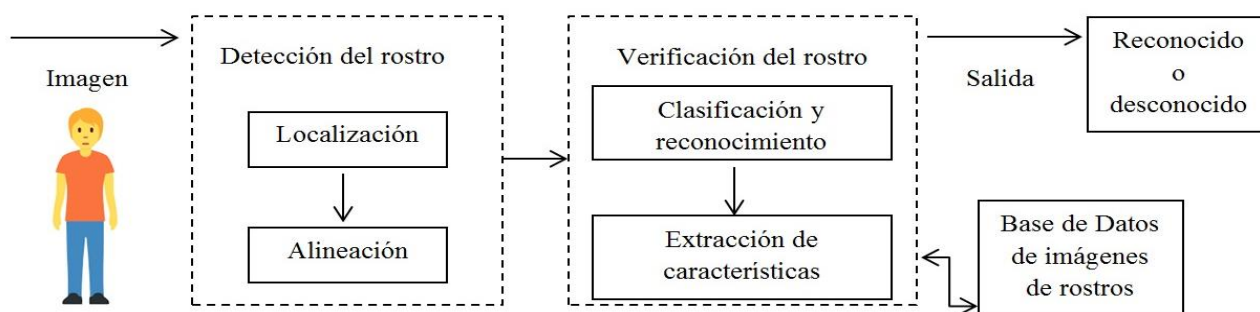
---

## Introducción

Un sistema de reconocimiento facial es una aplicación que emplea un algoritmo para identificar automáticamente a una persona en una imagen o video. En la actualidad es de gran importancia ya que

agiliza los trámites y garantiza mayor seguridad. Las principales investigaciones se desarrollan en aplicaciones de video vigilancia o control de acceso, estas aplicaciones se han extendido en procesos de autenticación en teléfonos móviles y otros dispositivos electrónicos. El reconocimiento facial se puede dividir en dos tipos: verificación facial e identificación facial. La verificación facial es una coincidencia 1 a 1 en la que solo detecta a partir de dos imágenes, si ambas imágenes son de la misma persona o no. La identificación facial es una coincidencia 1 a muchos, donde se necesita determinar quién es esta persona en la imagen entre todas N imágenes posibles.

De manera general podemos ver en la figura 1 un esquema de las principales etapas que tiene un sistema de reconocimiento facial. Luego de obtenida la imagen, el primer bloque se encarga de detectar el rostro y pre procesarlo, buscando generalmente una alineación frontal del rostro encontrando la punta de la nariz. En una segunda etapa se verifica si el rostro se puede identificar, para esto se extraen características que servirán para compararse con los rostros almacenados en una base de datos y determinar la similitud de los rostros que se intenta identificar con los que están almacenados en el sistema. A partir de esa medida de similitud se determina si el rostro es reconocido o no y en caso de ser reconocido, se informa a quien pertenece. En este trabajo se considerarán los métodos encontrados en la revisión bibliográfica usados en la etapa de verificación de rostro donde se utiliza el sensor Kinect para adquirir las imágenes.



**Fig. 1** – Esquema general de un sistema de reconocimiento facial.

Una gran parte de las investigaciones realizadas han hecho uso de imágenes RGB (2 dimensiones), las cuales se obtienen con facilidad. A pesar de los resultados positivos obtenidos, ciertas condiciones comunes, como cambios en la iluminación, ángulos de visión y fondos no uniformes, así como cambios en la

aparición humana debido al envejecimiento, las emociones y las oclusiones, aún limitan el correcto reconocimiento del rostro en aplicaciones del mundo real. (M. Wang & Deng, 2021)

La aparición de nuevos tipos de sensores de imagen con cámaras RGB-D (rojo, verde, azul y de profundidad) ha abierto nuevas fronteras para los sistemas de reconocimiento facial como es el caso del Kinect con un precio económico en el mercado. Este sensor ha hecho posible y rentable capturar simultáneamente datos de intensidad a color y profundidad registrados conjuntamente de una escena. (Cai et al., 2017) La información de profundidad puede ser fundamental para el reconocimiento facial, ya que proporciona información geométrica sobre el rostro, muestreando la superficie de los componentes faciales (Goswami et al., 2013a), permitiendo a los modelos representaciones faciales más sólidas, lo que mejora los métodos de reconocimiento. Además, éstos métodos son menos sensibles a las variaciones de pose e iluminación (Hayat et al., 2016).

Existe un test de evaluación para los algoritmos comerciales y prototipos de técnicas de reconocimiento facial, llamado FaceRecognitionVendor Test (FRVT) realizado por The National Institute of Standards and Technology (NIST); de los análisis realizados en los últimos años se concluye que el actual estado del arte del reconocimiento facial 2D es insuficiente para la alta demanda biométrica de las aplicaciones, a partir de lo cual intentar usar información 3D se ha vuelto la dirección de búsqueda emergente con la esperanza de hacer que el reconocimiento facial sea más exacto y robusto (*National Institute of Standards and Technology / NIST*, s. f.).

Una tarea determinante es extraer determinadas medidas sobre el rostro que evidencien los rasgos más relevantes en una determinada persona. Estas medidas pueden dividirse en dos parámetros de información: el primer parámetro está relacionado con la estructura morfológica del rostro. Esta, a menudo es llamada información de forma y se caracteriza por estar representada mediante una serie de puntos característicos del rostro a analizar como son la distancia entre los ojos, anchura de la nariz, forma de la barbilla, pómulos, ancho de la boca, entre otras (Abudarham et al., 2019). El segundo parámetro está relacionado con la información de textura presente en el rostro a ser analizado. En esta parte, se utilizan los valores de intensidad y color que se encuentran ubicados dentro del rostro del sujeto a analizar (Minami et al., 2018).

Resulta difícil realizar una clasificación de los métodos existentes de reconocimiento facial, pero se pueden identificar tres clases a un nivel alto:

1. Las técnicas holísticas, donde se utiliza toda la región del rostro como dato de entrada.
2. Las técnicas de extracción de características locales, donde se utilizan solo regiones de interés del modelo facial
3. Los métodos híbridos que integran información local y holística, la fusión de ambos tipos de características aporta más información discriminante.

Cuando se toma toda la región del rostro como dato de entrada presenta como ventaja que se dispone de mayor información, pero a su vez cuando las imágenes no presentan condiciones de adquisición uniformes esto se convierte en una desventaja. Cuando la extracción de características se define solo en regiones parciales del rostro, presenta mayor robustez cuando las condiciones de adquisición no son uniformes pues, por ejemplo, las condiciones de iluminaciones no afectan por igual a todas las regiones del rostro o cuando se presentan partes ocluidas del mismo, pero presenta la desventaja de contar con menor información para el modelo de este. Por lo general, estos algoritmos explotan el mapa de profundidad proporcionado por el sensor, sin embargo, en el caso de Kinect no es tan preciso como el proporcionado por otros sensores 3D, a causa del ruido en el momento de la captura de los datos (Wasenmüller & Stricker, 2017). Es por eso que la mayoría de los métodos desarrollados y mejorados en los últimos años van encaminados a un enfoque de fusión. (Soltanpour et al., 2017) (Naz et al., 2017)

A pesar de la importancia de hacer uso de estos recursos las empresas encuentran limitantes a la hora de utilizar estas técnicas, la razón principal es el costo económico elevado en el que podrían incurrir a la hora de emplear este tipo de tecnología ya que tanto el escáner 3D como el software propietario se venden a un precio muy alto. Es por eso la importancia de buscar soluciones más económicas y factibles para su implementación.

En un estudio realizado por Allied Market Research se estimó que el mercado global de reconocimiento facial se valoró en 3.400 millones de USD en 2019 y se prevé que se expanda a una tasa compuesta anual del 14,5% de 2020 a 2027 (*Facial Recognition Market Analysis by Technology, Application*, s. f.). La tecnología está mejorando, evolucionando y expandiéndose a un ritmo explosivo. Las tecnologías como la biometría se utilizan ampliamente para mejorar la seguridad. Se utilizan en varias aplicaciones, como control de acceso y seguimiento de asistencia. Los datos biométricos son universales, únicos y medibles y,

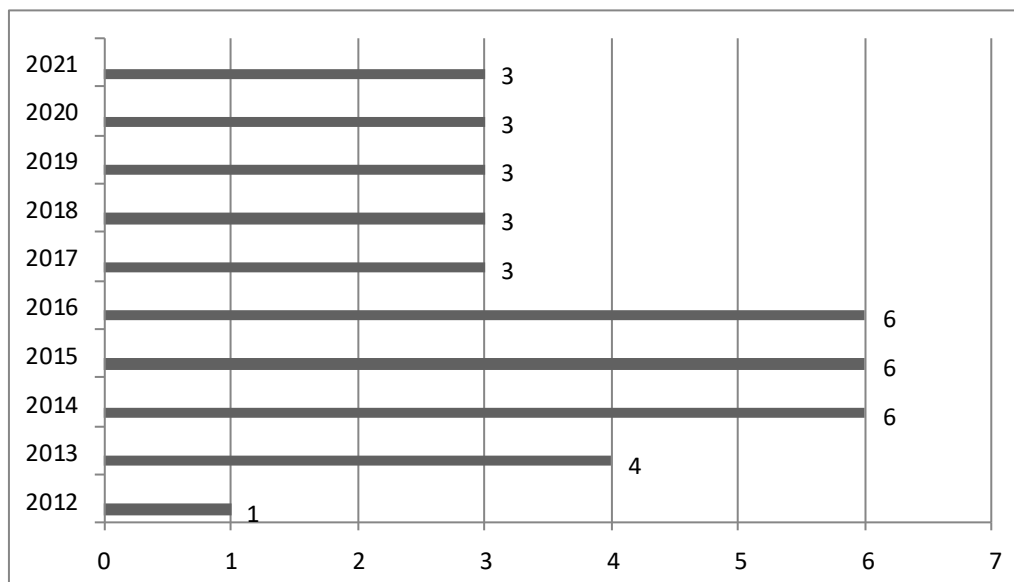
por lo tanto, se pueden utilizar para proporcionar soluciones de seguridad. Este estudio muestra la importancia económica que adquirirá el reconocimiento facial y los ingresos económicos que esta línea de investigación podría generar en los próximos años. Así como la importancia y demanda que se está generando a nivel internacional.

En la literatura consultada aparecen varios estudios realizados sobre las diferentes técnicas y algoritmos utilizados para reconocimiento facial. Pero se detectó la falta de un estudio profundo enfocado al sensor de adquisición de las imágenes o video Kinect, por lo que resulta insuficiente para determinar cuáles son las técnicas y algoritmos mayormente usados en el mismo y dónde se obtienen los mejores resultados. Por lo tanto, el objetivo de este trabajo es ofrecer una panorámica de las propuestas de reconocimiento facial utilizando un sensor Kinect para adquirir las imágenes, comparar y analizar los resultados que se obtienen de los diferentes trabajos consultados, para determinar los principales aportes y evaluar los problemas y retos de investigación pendientes de resolver.

## **Métodos o Metodología Computacional**

Existe gran cantidad de trabajos enfocados en el reconocimiento facial, donde se han desarrollado diversas técnicas y se ha realizado gran cantidad de mejoras a los algoritmos existentes. Se analizaron las publicaciones comprendidas en los últimos 10 años, en el periodo comprendido entre 2012-2021. Se encontraron amplias contribuciones a esta área de investigación, tanto en español como en inglés, con importantes aportes en el último año, lo que demuestra el gran interés de los investigadores sobre este tema. Por esto es útil realizar una evaluación, identificación e interpretación de las investigaciones más relevantes hasta la fecha. Para este estudio se utilizaron como fuentes principales Google Scholar, para identificar otras fuentes de interés científico, con énfasis en las bases de datos ACM Digital Library, IEEE Explorer, ScienceDirect, National library of medicine (PubMed), Researchgate, Scopus y Springer. No se tomaron en cuenta resúmenes o publicaciones parciales. Se presentan de manera especial aquellas contribuciones relacionadas con el reconocimiento facial que hicieran uso del sensor Kinect para la adquisición de las imágenes. Los términos de búsqueda utilizados para encontrar estudios relativos al tema fueron: “Facial

recognition”, “Kinect”, “LBP”, “HOG”, “Deep Learning”, “RGB”, “RGB-D”. Se acotaron los términos entre comillas, y se empleó el operador AND para relacionar los diferentes términos. Se seleccionaron los trabajos a partir de las coincidencias, reduciendo el volumen de trabajos considerando la relevancia de los aportes a partir del análisis del contenido. Las búsquedas realizadas permitieron obtener un total de 38 artículos donde se analizan los diferentes métodos, mejoras y aportes presentados por los investigadores para el reconocimiento facial. En la Fig. 2 puede observarse la distribución por año de los trabajos consultados y revisados.



**Fig. 2** – Cantidad de publicaciones sobre reconocimiento facial usando sensor Kinect en el período 2012-2021.

Kinect en su primera versión tiene una resolución en imágenes RGB de 640×480 a 30 fotogramas por segundos (frames por segundos, fps) y 1280×960 a 12 fps y en los mapas de profundidad es de 320×240 a distancias de 0.8 a 4 metros y 640×480 a distancias de 0.4 a 3 metros. La segunda versión de Kinect tiene una resolución en imágenes RGB de 1920×1080 a 30 fps y en los mapas de profundidad de 512×424 en distancias 0.5 a 4.5 metros. (Al-Naji et al., 2017)

Hay que destacar que, en el período revisado, se proponen técnicas y algoritmos para reconocimiento facial, en imágenes RGB y mapas de profundidad, que corresponden a otro tipo de sensor de profundidad. Estos sensores tienen una mayor resolución que las obtenidas mediante el Kinect y son más costosos en el mercado. A pesar de que estos trabajos también utilizan imágenes RGB-D no fueron tenidos en cuenta en este estudio, tal como fue explicado anteriormente. Es importante destacar que los artículos revisados en el año 2021 son solo hasta el mes de junio, encontrándose 3 artículos publicados. Del total de artículos revisados el 60,53 % se encuentra publicado entre los últimos 10 a 5 años y el 39,47 % pertenece a los últimos 5 años. Si analizamos los artículos que utilizan métodos de DL observamos que de los 11 artículos revisados el 81,82 % se encuentran publicados en los últimos 5 años, demostrándose el interés reciente y la relevancia de utilizar esta técnica para el reconocimiento facial. El 36,84 % utilizan solo mapas de profundidad, solo el 7,8 % utilizan únicamente imágenes RGB y más de la mitad de los métodos consultados, representando el 55,36 %, combinan RGB y mapas de profundidad para obtener mayor información discriminante. En este trabajo hemos agrupado estos métodos según la técnica empleada y hacemos un análisis las contribuciones y deficiencias detectadas.

## Resultados y discusión

Los algoritmos estudiados fueron agrupados en dos categorías: primeramente, los que no utilizan redes neuronales artificiales con aprendizaje profundo (Deep Learning, DL) (Tabla 1). Estos a su vez, por su importancia, fueron divididos en otras dos sub-categorías: los que se basan en Histogramas de Gradiente (Histogram of Gradient, HOG) (Tabla 2) y los que lo hacen en el descriptor Patrón Binario Local (Local Binary Pattern, LBP) (Tabla 3). En la segunda categoría se agruparon los que utilizan DL (Tabla 4).

**Tabla 1** - Métodos que no utilizan DL.

Ref	Año	Imagen	Extracción de características	Clasificador
(Li, 2013)	2013	RGB y profundidad	Punto iterativo más cercano (Iterative Closest Point) ICP; Espacio de color discriminado (Discriminat	Clasificador de representación de colores dispersos (Color Sparse



			Color Space) DCS	Representation Classifier) SRC.
(Ben Amor et al., 2014)	2014	RGB y profundidad	Campo escalar denso (Dense Scalar Field).	Modelos ocultos de Markov (Hidden Markov Models)
(Zhou et al., 2015)	2015	RGB y profundidad	7 puntos de características faciales: la parte superior, inferior, izquierda, derecha del rostro, ojo izquierdo, ojo derecho y nariz. Se calcula la longitud y el ancho del rostro, la distancia entre dos ojos y la distancia media entre la nariz y los ojos.	Cascada de dos niveles: <ul style="list-style-type: none"> <li>• Árbol de decisión</li> <li>• Distancia euclidiana mejorado</li> </ul>
(Elaiwat, 2015)	2015	RGB y profundidad	Función local de curvatura.	Modelo semántico basado en RBM Restricted Boltzmann machine (Máquina de Boltzmann restringida) Algoritmo de divergencia contrastiva cuatripartita
(Romero et al., 2016)	2016	Mapas de profundidad	Eigensurface	Mahalanobis distance.
(J. Zhang et al., 2016)	2016	Mapas de profundidad	ICP	Función de distancia firmada (Signed Distance Function) SDF.
(Hayat et al., 2015)	2016	RGB y profundidad	Representación de matriz de covarianza.	SVM.
(Novamizanti et al., 2019)	2019	Mapas de profundidad	ICP	distancia euclidiana KNN y la distancia de la manzana de la ciudad.
(Ouannes et al., 2020)	2020	RGB	Descriptor de característica robusta acelerada (speeded-up-robust-feature descriptor) (SURF).	KNN + árbol k-dimensional (k-dimensional tree).

**Tabla 2 - Métodos que no utilizan DL agrupados en el uso del descriptor HOG.**

Ref	Año	Imagen	Extracción de características	Clasificador
(Segundo et al., 2013)	2013	RGB y profundidad	HOG	Clasificador en cascada potenciado (boosted cascade classifiers).
(Goswami et al., 2013b)	2013	RGB y profundidad	HOG	Árbol de decisiones aleatorio (Random Decision Forest) (RDF)
(Kishan Anapu & Rao Peri, 2014)	2014	RGB y profundidad	Análisis del Discriminante Lineal (Lineal Discriminant Analysis) LDA, HOG, LBP.	Vecino más cercano.
(Goswami et al., 2014)	2014	RGB y profundidad	Prominencia y entropía de la imagen RGB-D (RGB-D Image Saliency and Entropy) RISE, HOG.	RDF

(Boumedine et al., 2021)	2021	Mapas de profundidad	HOG	Clasificador de representación colaborativa (Collaborative Representation Classifier) (CRC).
--------------------------	------	----------------------	-----	--

**Tabla 3 - Métodos que no utilizan DL agrupados en el uso del descriptor LBP.**

Ref	Año	Imagen	Extracción de características	Clasificador
(Huynh et al., 2013)	2012	RGB y profundidad	LBP, 3DLBP y Gradient-LBP.	Máquina de soporte vectorial (Support Vector Machines) (SVM).
(Ciaccio et al., 2013)	2013	RGB	LBP y descriptor de covarianza (covariance descriptor) (COV).	Medida probabilística.
(Min et al., 2014)	2014	RGB y profundidad	<ul style="list-style-type: none"> <li>• Análisis de Componentes Principales (Principle Component Analysis) PCA,</li> <li>• LBP,</li> <li>• Transformación de características de escala invariable (Scale-Invariant Feature Transform) SIFT,</li> <li>• Patrones binarios locales de Gabor (Local Gabor Binary Patterns) LGBP</li> </ul>	k vecinos más próximos (Nearest-Neighbor) kNN.
(J. B. C. Neto & Marana, 2014)	2014	Mapas de profundidad	3DLBP	SVM
(Mantecon et al., 2014)	2014	Mapas de profundidad	Depth Local Quantized Pattern (DLQP).	SVM
(Naik & G.N., 2015)	2015	RGB y profundidad	LBP y cálculo de histograma.	SVM
(Aissaoui & Martinet, 2015)	2015	RGB y profundidad	LBP y DLBP.	SVM
(Cardia Neto & Marana, 2015)	2015	Mapas de profundidad	3DLBP e Histograma de gradientes orientados promediados (Histogram of Averaged Oriented Gradients) HOAG.	SVM
(Li et al., 2016)	2016	RGB y profundidad	Espacio de color discriminante (Discriminant Color Space) (DCS) en RGB y LBP, Haar, Gabor en profundidad. Análisis discriminante lineal (Linear discriminant	Vecino más cercano.

			analysis) (LDA) para reducir dimensionalidad.	
(Cardia & Marana, 2016)	2016	Mapas de profundidad	Red de distribución de contenido (Content Distribution Network) con una modificación P (ACDN+P) y 3DLBP.	SVM
(Zohra & Gavrilova, 2017)	2017	Mapas de profundidad	LBP	SVM
(Saleh, 2017)	2017	Mapas de profundidad	LBP y Eigenfaces: Detector de características densas (Dense Feature Detector).	Funciones elementales independientes robustas binarias (Binary Robust Independent Elementary Features) (BRIEF). Se usa para hacer coincidir puntos similares que se comparan.
(Mahdi Barrah et al., 2017)	2017	Mapas de profundidad	3DLBP y Valor singular de descomposición (singular value decomposition) (SVD).	SVM

**Tabla 4 - Métodos que utilizan DL.**

Ref	Año	Imagen	Extracción de características	Clasificador
(Fernández-Cervantes et al., 2015)	2015	RGB y profundidad	Red neuronal difusa (Histograma información de color + fuzzy pattern)	función softmax
(Chowdhury et al., 2016)	2016	RGB y profundidad	Auto-encoder	función softmax
(Borghini et al., 2018)	2018	Mapas de profundidad	Siamese Red Neuronal Convolutiva (Convolutional Neural Networks) (CNN)	función softmax
(H. Zhang et al., 2018)	2018	RGB y profundidad	9 Capas CNN + Inception	función softmax
(Sepas-Moghaddam et al., 2018)	2018	RGB y profundidad	Fine-tuned VGG-Face	función softmax
(Mu et al., 2019)	2019	Mapas de profundidad	14 Capas CNN + Attention	función softmax
(Zafar et al., 2019)	2019	RGB	Red convolutiva bayesiana (Bayesian convolutional networks)	función softmax
(J. Chen et al., 2020)	2020	Mapas de	10 Capas CNN	función softmax

2019)		profundidad		
(Xiao & Xie, 2020)	2020	RGB y profundidad	GaborDCNN	función softmax
(Jiang et al., 2021)	2021	RGB y profundidad	ResNet18 + módulo de vectorización de atención espacial	función softmax
(Uppal et al., 2021)	2021	RGB y profundidad	Extractor de características convolucional VGG-16	función softmax

En la Tabla 5 se muestra la eficacia de varios de los métodos revisados, la base de datos sobre la que fueron evaluados y el sitio donde fueron publicados, observándose un predominio en IEEE, Springer y ACM Digital Library. Se han incluido solamente aquellos métodos con una eficacia reportada por sus autores superior al 90 %. Como se puede observar, los experimentos se realizaron sobre diferentes bases de datos, y los resultados varían según la complejidad de las imágenes escogidas. Las bases de datos mayormente usadas fueron: IIIT-D, EURECOM y CurtinFaces. Los mejores resultados fueron obtenidos para las bases de datos IIIT-D y EURECOM y para expresiones faciales neutrales, que presentan menor nivel de complejidad al momento de aplicar el reconocimiento facial. Los mejores resultados se encontraron en los métodos que fusionan: HOG y DL; LBP, 3DLBP y Gradient-LBP; ICP y Espacio de color discriminado, mostrando resultados de eficacia superiores al 98 %. Es importante tener en cuenta que estos resultados pueden variar según las imágenes escogidas y la complejidad de la base de datos sobre la que se realice el experimento.

**Tabla 5** - Comparación entre los métodos de reconocimiento facial con mejor desempeño.

Ref	Eficacia (Accuracy)	Base de Datos donde se evaluó	Publicación
(Li, 2013)	98.40 %	CurtinFaces	Tesis Doctoral, Curtin University, Australia
(Ben Amor et al., 2014)	93,83	BU-4DFE	IEEE
(Zhou et al., 2015)	90,5 con Distribución uniforme 95,9 con Distribución Normal	Propia del autor	IEEE
(Elaiwat, 2015)	93.74% Para expresiones faciales no neutrales 97.43% Para expresiones faciales	FRGC 2.0	Tesis Doctoral, The University of Western, Australia

	neutrales		
(Hayat et al., 2015)	93,56 %	CurtinFaces, Biwi Kinect y UWA Kinect	Neurocomputing
(Novamizanti et al., 2019)	97,56%	Propia del autor	Revista de tecnología de la información y ciencias de la computación
(Ouannes et al., 2020)	95,51 % 80,40 %	kinect face database EUROCOM	IEEE
(Goswami et al., 2013b)	91.6 % 98.1 %	IIIT-D EURECOM	IEEE
(Goswami et al., 2014)	95.3 % 98.5 %	IIIT-D EURECOM	IEEE
(Boumedine et al., 2021)	82,53% Para expresiones faciales no neutrales 98,87% Para expresiones faciales neutrales	CurtinFaces	Springer
(Huynh et al., 2013)	98.08 % Kinect 1 90.38 % Kinect 2 95.30 % Ambos	EURECOM	Springer
(Ciaccio et al., 2013)	94.2% 30 <sup>o</sup> posición 84.6% 60 <sup>o</sup> posición 75.0% 90 <sup>o</sup> posición	CurtinFaces	IEEE
(J. B. C. Neto & Marana, 2014)	96% usando 64 regiones	EURECOM	Conferencia Internacional
(Aissaoui & Martinet, 2015)	93,7 %	FRGC 2.0	ACM Digital Library
(Cardia Neto & Marana, 2015)	97 %	EURECOM	ACM Digital Library
(Zohra & Gavrilova, 2017)	96.15% oclusión por mano 90.38% oclusión por papel	EURECOM	Springer
(Chowdhury et al., 2016)	67.77 % 98.71 %	KaspAROV IIIT-D	IEEE
(Borghi et al., 2018)	81,42 % 98,77 %	Pandora HRRFaceD	Conferencia Internacional
(H. Zhang et al., 2018)	98.6% 90.8%	IIIT-D VAP	IEEE

	96.3% 90.8% 97.3%	EUROCOM BUAA Propia del autor	
(Sepas-Moghaddam et al., 2018)	98.1%	EUROCOM	IEEE
(Zafar et al., 2019)	98.1 %	EUROCOM	IEEE
(Xiao & Xie, 2020)	97.76 %	EUROCOM	International Journal of Network Security
(Uppal et al., 2021)	99,1 % 99,7 % 95,3 %	CurtinFaces IIIT-D KaspAROV	IEEE

El trabajo presentado en (Li, 2013) tiene como objetivo mejorar la precisión y robustez del reconocimiento de rostros utilizando color e información de profundidad para imágenes adquiridas con el sensor Kinect. Para la información de color en primer lugar, se propone el Espacio de color discriminante por bloques (Block-wise Discriminant Color Space). Este método aprende el espacio de color discriminativo basado en parches locales de una imagen de rostro humano en lugar de la imagen holística. Esto es debido a que los rostros humanos muestran diferentes colores en sus diferentes partes. En segundo lugar, se observa que la mayoría de los espacios de color existentes constan de tres componentes de color como máximo, mientras que se puede encontrar información complementaria en múltiples componentes de color en múltiples espacios de color y, por lo tanto, se propone el modelo de fusión de múltiples colores (Multiple Color Fusion model) para buscar y utilizar múltiples componentes de color de manera efectiva. Por último, se proponen dos algoritmos robustos de reconocimiento de rostros de color. El método codificación escasa de color (Color Sparse Coding), que puede tratar las imágenes faciales con ruido y oclusión; y el método discriminante de tensor de color multilineal (Multi-linear Color Tensor Discriminant), que utiliza una técnica multilineal para manejar datos no lineales. Para la información de profundidad se utiliza el algoritmo ICP para extraer las características. Con la información de color y profundidad extraída finalmente se realiza la clasificación mediante SRC.

En (Ben Amor et al., 2014) se usa un campo escalar denso (Dense Scalar Field) para la extracción de características y modelos ocultos de Markov (Hidden Markov Models, HMM) como clasificador, éste presenta limitaciones en la detección de la punta de la nariz en caso de vistas no frontales y en presencia de oclusiones (por gafas, mano, cabello, etc.). También presenta problemas al procesar videos en 3D con baja resolución, como es el Kinect v1, debido, además, a la presencia de ruido en proceso de captura de la información. Este método pudiera mejorar su comportamiento usando Kinect v2, que presenta mayor resolución. En (Zhou et al., 2015) se propone una técnica que integra los datos de profundidad con los datos RGB para generar datos en bruto de rostro 3D y luego extrae puntos de características. Luego se identifica el objetivo a través de un clasificador en cascada de dos niveles. Los resultados experimentales indican que el algoritmo mejora la precisión de reconocimiento en comparación con los algoritmos de reconocimiento de rostro 2D y 3D existentes cuando el rostro está exactamente delante del sensor Kinect, también puede aumentar en 9.3% la precisión de reconocimiento en comparación con el algoritmo PCA-3D cuando no está frente a la cámara.

En (Elaiwat, 2015) se presenta el método de función local de curvatura, que propone la extracción de características locales en lugar de globales. Este trabajo investiga la capacidad de la transformación de la curva para extraer características globales robustas para la tarea de reconocimiento facial 3D en diferentes expresiones faciales y los beneficios de fusionar las características de la curva 3D y 2D para lograr la identificación de rostros multimodal. En (Romero et al., 2016) aplican el algoritmo Eigensurface para el reconocimiento facial de los mapas de profundidad obtenidos del sensor Kinect. La eficiencia de la detección facial obtenida no fue buena y se detectaron problemas por la baja resolución de los mapas de profundidad del Kinect y la mala precisión en la segmentación del rostro.

En (Segundo et al., 2013) el sistema propuesto usa un detector facial 3D basado en clasificadores en cascada potenciados (boosted cascade classifiers) para ubicar rostros bajo variación de pose que son normalizados con el algoritmo ICP, y con HOG se extraen tres diferentes regiones faciales. Para cada cuadro de la secuencia de video, solo se utiliza la región menos afectada por el ruido. En (Goswami et al., 2013b) se calcularon mapas de entropía correspondientes a las imágenes RGB de entrada y de profundidad junto con

un mapa de prominencia visual correspondiente a la imagen RGB. A continuación, se aplicó el descriptor HOG, extrayendo así características de estos mapas. Las características extraídas se concatenaron finalmente para ser utilizadas como entrada a un clasificador de Bosque de decisión aleatoria (Random Decision Forest, RDF) para reconocer la identidad.

El método de reconocimiento facial RGB-D propuesto en (Goswami et al., 2014) calculó un nuevo descriptor basado en mapas de prominencia y entropía, llamado descriptor RISE. Las características extraídas de diferentes mapas se concatenaron y luego se usaron descriptores HOG para proporcionar un clasificador RDF con las características de la textura. El uso de descriptores Hog requiere muchos menos recursos computacionales (memoria, potencia de procesamiento y duración de la batería) en comparación con otras técnicas, considerándolo adecuado para las limitaciones del hardware portátil (L. B. Neto et al., 2017). En (Kishan Anapu & Rao Peri, 2014) se usó HOG para la extracción de características de las imágenes RGB y LDA para las de profundidad, de la misma manera Patrones binarios locales (Local Binary Patterns, LBP) para RGB y LDA para profundidad.

Para superar las limitaciones de algunos algoritmos, como PCA, LDA y LPP cuando se aplican sobre bases de datos que tienen diferentes variaciones faciales, las características locales pueden lograr resultados prometedores (Dong Li et al., 2015). La extracción de estas características locales se puede hacer de forma automática utilizando herramientas adecuadas o dividiendo el rostro en varias partes. Los métodos más comunes utilizados para extraerlas son LBP y la Transformada Wavelet de Gabor (Gabor Wavelet Transform) GWT. LBP se aplicó originalmente a la representación y clasificación de textura, aunque se comprobó su utilidad para el reconocimiento facial (Ahonen et al., 2006). El uso de LBP en el paso de extracción de características hace que el sistema sea más robusto a las variaciones faciales en iluminación y rotación. Las representaciones faciales GWT también son más robustas a las variaciones faciales en la iluminación y las expresiones, siendo muy similares a los del sistema visual humano, por lo que muchos investigadores muestran interés en éstos para capturar características discriminatorias. (Liu et al., 2021).



El descriptor 3DLBP (Huang et al., 2006) codifica las diferencias de profundidad además de la información de textura capturada por el LBP tradicional. Se menciona que 3DLBP posee varias deficiencias: gran longitud de las características, codificación altamente sensible y pérdida de información de los signos asociados con los valores de profundidad. Por lo que en su lugar el autor propone el descriptor Gradient-LBP (G-LBP) para abordar estas limitaciones, el cual fue también utilizado en (Huynh et al., 2013). En (J. B. C. Neto & Marana, 2014) se propone el uso de 3DLBP aplicado a la profundidad de mapa obtenido de un sensor Kinect. En (Naik & G.N, 2015) se busca un reconocimiento facial y seguimiento usando Kinect y auxiliándose de la potencia de cálculo de la Unidad de Procesamiento Gráfico (Graphics Processing Unit) (GPU). Un vector de características de LBP modificado se calcula utilizando la información de fusión de la profundidad y la escala de grises en la GPU. Se usa la imagen de profundidad de Kinect para aumentar la robustez y reducir el costo computacional del reconocimiento de rostros basado en LBP convencional. En (Min et al., 2014), el rendimiento de diferentes extractores de características como son PCA, LBP, SIFT y LGBP se compararon para el reconocimiento facial RGB-D, donde el descriptor LBP obtuvo el mejor desempeño. En (Cardia & Marana, 2016) se propone un método de fusión entre Active Content Distribution Network con una modificación P (ACDN+P) y 3DLBP para imágenes capturadas con el sensor Kinect. Para el preprocesamiento se propone el uso de Symmetric Filling, Iterative Closest Point, y Savitzky-Golay Filter, mostrándose que a pesar de que hay una gran variación de expresión y oclusión en la base de datos donde se probó el método, éste funcionó bien, aumentando en gran medida la tasa de reconocimiento. En (Mahdi Barraah et al., 2017) se propone la fusión entre 3DLBP y singular value decomposition (SVD). Aunque el descriptor SVD tiene el rendimiento más bajo en varios casos, cuando se fusiona con 3DLBP mejora su rendimiento, convirtiéndose en una buena alternativa para aumentar el rendimiento de reconocimiento facial cuando se utiliza el sensor Kinect para la adquisición de las imágenes. En (Aissaoui & Martinet, 2015) se propone la fusión de LBP para las imágenes de intensidad y DLBP para las de profundidad. Las modalidades 2D y 3D pueden verse afectadas de manera similar por ciertas condiciones de adquisición, como la oclusión o las variaciones de postura. Esto muestra que las imágenes de profundidad e intensidad son de alguna manera dependientes. Esta es una cuestión a tener en cuenta cuando se pretenda utilizar técnicas que integren información de intensidad y profundidad.

Está demostrado que algoritmos basados en LBP y Eigenfaces son capaces de proporcionar un alto nivel de precisión en el reconocimiento facial. Esto es debido a la resolución significativamente alta de las imágenes del mapa de profundidad generado por la última versión del dispositivo Kinect (Saleh, 2017). En este trabajo se propone un algoritmo llamado Dense Feature Detector el cual demuestra ser eficaz en el reconocimiento facial utilizando imágenes de mapas de profundidad, en particular en condiciones de buena iluminación. En (Cardia Neto & Marana, 2015) se propone la fusión de 3DLBP and Histogram of Averaged Oriented Gradients (HAOG) utilizando el sensor Kinect. Los resultados de este trabajo muestran que los datos generados por Kinect son lo suficientemente discriminativos como para permitir el reconocimiento facial y que 3DLBP funciona mejor que los otros métodos propuestos en la literatura. Este método, que utiliza solo mapas de puntos, presentó mejores resultados que el método de Entropy and Saliency Map, que utiliza mapas de profundidad e información RGB. Como preprocesamiento se utilizó el método de Symmetric Filling mejorando la calidad del reconocimiento facial en presencia de obstrucciones. Como clasificador se usó SVM. La fusión de 3DLBP y HOAG muestra robustez en imágenes con variaciones en pose, iluminación y con algún grado de obstrucción. En este método al usar solo los mapas de profundidad y obtener resultados favorables, se podría suponer que usar solo los mapas de profundidad sería suficiente para obtener resultados satisfactorios, aunque esto no queda claro en la literatura revisada pues existe contradicción entre los autores consultados, como se muestra más adelante.

En (Mantecon et al., 2014) se usa el descriptor Depth Local Quantized Pattern (DLQP) con un clasificador SVM. Se realizó un experimento en el que se compararon los resultados de los algoritmos LBP y SIFT, mostrando que la técnica DLQP-SVM dio resultados mucho mejores que el método LBP, y resultados moderadamente superiores a SIFT. Sin embargo, el autor plantea que los mapas de profundidad solo se utilizaron para el reconocimiento de rostros, por lo que no se aprovecha la información adicional que se puede proporcionar a partir de las imágenes RGB correspondientes. En (Li et al., 2016) se propuso un algoritmo que extrae diferentes características de RGB e imágenes de profundidad tomadas de Microsoft Kinect v1 y las fusiona utilizando el método Finer Feature Fusion desarrollado por los autores, principalmente para eliminar información redundante y para preservar características importantes extraídas. Los métodos y técnicas existentes logran una alta precisión en cuanto al uso de RGB y datos de profundidad

conjuntamente, los autores concluyen que la información de profundidad del Kinect puede ayudar a mejorar significativamente el reconocimiento facial.

En (Zohra & Gavrilova, 2017) las características locales se extraen de las imágenes de profundidad utilizando operadores LBP y se utiliza un clasificador SVM no lineal para detectar el rostro frontal y el rostro ocluido. Para localizar regiones ocluidas en la imagen facial, se presenta un enfoque basado en el umbral para identificar el área ocluida. Los resultados experimentales muestran que las imágenes faciales ocluidas pueden localizarse y detectarse de manera efectiva a partir de las imágenes de profundidad. El sistema de reconocimiento descarta las regiones ocluidas de las imágenes faciales y combina solo la parte facial no ocluida con las imágenes de la galería para encontrar la mejor coincidencia posible. Por lo tanto, el método propuesto mejora el rendimiento de reconocimiento en presencia de oclusión en las imágenes faciales. Se plantea para trabajo futuro calcular la calidad de la imagen facial bajo oclusión determinando la proporción de la región ocluida en las imágenes faciales. En función de la proporción del área ocluida, se puede asignar un puntaje de calidad a la imagen facial. Luego, el puntaje de calidad se puede fusionar con el puntaje correspondiente en la fusión de puntaje para determinar la confianza del sistema de reconocimiento facial bajo oclusión.

La utilización de redes neuronales convolucionales (CNN) con 3DLBP intentan dar solución a los problemas de oclusión y otros tipos de desafíos para el reconocimiento facial. (Cardia Neto & Marana, 2018) Como se puede observar en la Tabla 2, se han empezado a usar métodos de aprendizaje profundo desde 2015 para el reconocimiento facial usando RGB-D. Se han utilizado varias estrategias para aprovechar al máximo la información de profundidad proporcionada por los sensores RGB-D. En (Chowdhury et al., 2016), se propuso un método de reconocimiento facial RGB-D basado en una arquitectura de codificador automático para aprender una función de mapeo entre las modalidades RGB y de profundidad, generando así una representación de características más rica. Se propuso una nueva estrategia de entrenamiento en el contexto del reconocimiento facial RGB-D (Xu et al., 2015), aprovechando la información de profundidad para mejorar el aprendizaje de una distancia métrica durante el entrenamiento de una CNN. En (H. Zhang et al., 2018), se utilizó una nueva arquitectura para aprender de

las modalidades RGB y de profundidad, introduciendo una capa compartida entre dos redes correspondientes a las dos modalidades, permitiendo así la interferencia entre modalidades en las primeras capas. En (Sepas-Moghaddam et al., 2018), las imágenes RGB, de disparidad y profundidad se utilizaron de forma independiente como entradas a una arquitectura VGG-16 para ajustar el modelo VGG-Face. Las incrustaciones obtenidas finalmente se fusionaron para alimentar un clasificador SVM para realizar el reconocimiento facial.

En (Fernández-Cervantes et al., 2015) se presenta un trabajo de un sistema basado en una red neuronal difusa que combina histogramas de información de color y profundidad basado en los algoritmos de teoría de patrón difuso (fuzzy pattern) para el reconocimiento facial en tiempo real utilizando el sensor Kinect. La identificación de rostros se realiza en dos fases. En la primera se calcula la hipótesis de patrón facial de los puntos faciales, se configura cada forma de punto, la ubicación relacionada en las áreas y las líneas del rostro. Luego, en la segunda fase, el algoritmo realiza una búsqueda en estas configuraciones de puntos frontales. En (Zafar et al., 2019) se propone un algoritmo de red convolucional bayesiana para que los sistemas de vigilancia sean más robustos, mejorando su efectividad en el tratamiento de falsos positivos mediante el uso de la incertidumbre del modelo. Los resultados de los experimentos mostraron una mejora del 3-4% en la precisión con la incertidumbre del modelo sobre las DCNN (Deep Convolutional Neural Network) y las técnicas convencionales de aprendizaje automático.

La arquitectura del modelo CNN propuesto en (J. Chen et al., 2019) consta de una capa de entrada de datos, cuatro capas de convolución, tres capas de grupo y dos capas completamente conectadas. Además, hay seis capas a las que se añaden unidades lineales rectificadas. El análisis de los resultados sugiere que la arquitectura CNN propuesta tiene un mejor rendimiento de reconocimiento que algunos métodos tradicionales de extracción manual de características, como HOG y LBP. En (Xiao & Xie, 2020) se propone un enfoque denominado GaborDCNN aplicado a la tecnología de reconocimiento facial de imágenes RGB-D de dos modalidades, que puede extraer las características a través de la transformación de Gabor de imágenes y una red neuronal convolucional profunda. Esta investigación proporciona un método eficaz para el reconocimiento facial modal múltiple en condiciones complejas.

En (Jiang et al., 2021) se propone un método de fusión multimodal de extremo a extremo basado en la atención espacial y de canal para fusionar de manera efectiva dos modalidades de imagen, RGB y profundidad, para mejorar el reconocimiento facial RGB-D. Se establecen tres ramas combinadas con los módulos de atención espacial y de canal a partir de ResNet18, de manera que se obtienen las características de tres modalidades: RGB, mapa de profundidad y sus modalidades de fusión, respectivamente. Luego, las características bajo estas tres modalidades se fusionan, se alimentan a una capa compartida, y las características fusionadas se aprenden para características discriminatorias más profundas. Finalmente, son procesadas a través de un módulo de vectorización de atención espacial. El rendimiento del método se basa en la contribución de cada componente obteniéndose buenos resultados. En (Uppal et al., 2021) se utilizan dos extractores de características convolucionales VGG-16 cuyas salidas son los mapas convolucionales de profundidad y RGB. Estas dos redes se combinan para formar el módulo extractor de características convolucionales, se crea un mapa de características agrupando estas funciones y se envían al clasificador para el reconocimiento. En este trabajo también se realizaron experimentos adicionales con imágenes térmicas, en lugar de imágenes de profundidad, mostrando la capacidad de generalización del método propuesto.

En los últimos años, se han propuesto varios métodos de aprendizaje profundo para el reconocimiento facial, lo que ha redundado en un desarrollo y mejoramiento de esta tarea. El reconocimiento facial se beneficia de la arquitectura jerárquica de los métodos de aprendizaje profundo para aprender la representación facial discriminativa. A pesar de esto, continúan siendo un desafío los problemas de oclusión. En los trabajos revisados se pudo observar que los descriptores más usados, y con los cuales se alcanzan buenos resultados, son las variantes desarrolladas a partir de HOG y LBP. Del análisis de los trabajos revisados se puede inferir que el rendimiento de las redes de aprendizaje profundo puede mejorarse aún más a partir de conjuntos de datos y algoritmos novedosos, mediante fusiones con técnicas de extracción de características que permitan mejorar la correcta clasificación de un rostro y sirvan de apoyo al momento de tomar la decisión en la identificación. Más de la mitad de los métodos consultados combinan RGB y mapas de profundidad aprovechando la información que puede obtenerse de cada una de estas. Ante

el costo computacional que puede producirse por el uso de estas técnicas y el uso de estas redes con alta información de características del rostro, puede deducirse del estudio realizado, que al trabajar en métodos que reduzcan la información discriminante, pudiera alcanzarse mejoras en la velocidad de procesamiento sin que se vea afectada la efectividad método.

Los mejores resultados de eficacia según el análisis de los autores en los trabajos revisados se obtienen en los métodos que usan HOG y DL. Del estudio realizado se observa una alta eficacia de estos métodos cuando los rostros presentan baja o ninguna oclusión, cuando la posición del rostro se encuentra frontal al sensor y con expresiones faciales neutrales, a medida que estas condiciones no se cumplen empeora el desempeño de los métodos. Por esto se considera necesario continuar investigando en esta área para desarrollar algoritmos que mejoren su desempeño ante imágenes que presenten un nivel de complejidad alto. Para aminorar el problema de la baja resolución que presentan las imágenes de profundidad adquiridas mediante el sensor kinect y la posibilidad de múltiples rostros en la misma imagen, se podría crear un conjunto de entrenamiento más grande y diverso para incluir estas variaciones durante el entrenamiento y ayudar al clasificador con una mejor generalización. (García Pinilla, 2020) Para intentar atenuar el problema de la baja resolución en las imágenes obtenidas por el sensor Kinect puede tenerse en cuenta la adaptabilidad de las redes profundas a imágenes de baja resolución mediante el aprendizaje de múltiples escalas (Multi-scale Deep Learning) (Karnewar & Wang, 2020) (Y. Wang et al., 2020).

Las soluciones también se pueden adoptar en el contexto de diferentes tareas, como la segmentación de escenas, la detección de objetos y el reconocimiento de acciones. También pueden añadirse atributos biométricos como, el sexo, el color de ojos y cabello, la altura y el peso (Jiang et al., 2020) (Das & Dantcheva, 2018) para impulsar el reconocimiento o apoyarse en un sistema de reconocimiento de la marcha (Fumio, 2020).

El aprendizaje automático adverso (Adversarial machine learning) ha sido eficaz para generar imágenes faciales realistas y de alta calidad, y también ofrece control sobre el estilo de las imágenes generadas con

diferentes niveles de detalle mediante la variación de los vectores de estilo y el ruido (Karras et al., 2019) (Karras et al., 2017) (Vorobeychik & Kantarcioglu, 2018)

## Conclusiones

En esta revisión se logró agrupar y evaluar varios trabajos en el área del reconocimiento facial en imágenes RGB-D obtenidas a partir del sensor Kinect. El principal objetivo de este artículo fue identificar estas propuestas relevantes en los últimos 10 años. La revisión realizada de la literatura científica sobre esta temática permitió resumir las técnicas empleadas para la extracción de características tanto en imágenes RGB como en los mapas de profundidad, así como las fusiones que permiten mejorar el rendimiento de estos sistemas. Además, se estudiaron los clasificadores usados en los métodos que, a partir de la información obtenida por los descriptores, intentan identificar correctamente el rostro.

A pesar de los avances logrados en esta área, muchas no utilizan el sensor Kinect, presumiblemente por su baja resolución. Teniendo en cuenta los trabajos revisados que utilizan Kinect los aportes encontrados son de gran importancia, pero se consideran aún insuficientes pues aún quedan problemas por resolver tales como cuando la imagen presenta oclusión, la posición del rostro no se encuentra frontal al sensor, las expresiones faciales no son neutrales y los cambios fisiológicos que puede presentar el rostro al envejecer. Continúa siendo determinante el conjunto de entrenamiento al utilizar redes de aprendizaje profundo y el tratamiento que se da para resolver el costo computacional.

Este trabajo constituye un nuevo punto de partida para promover el desarrollo de nuevas técnicas e incentivar futuras investigaciones donde se mejoren los algoritmos o se busquen nuevas fusiones que mejoren los resultados del reconocimiento facial en las imágenes que se adquieren mediante el sensor Kinect, que en la práctica es más económico y portable que otros sensores comerciales. A partir de este estudio se evidencia que mediante el sensor Kinect es posible alcanzar resultados satisfactorios y con posibilidad de ser mejorados mediante métodos que combinen información del rostro de las imágenes RGB y los mapas de profundidad, observándose que en los últimos años las investigaciones van encaminadas a

un mayor uso de redes neuronales profundas, las que fusionadas con métodos clásicos pueden ayudar a mejorar aún más la eficacia de los sistemas de reconocimiento facial.

## Referencias

- Abudarham, N., Shkiller, L., & Yovel, G. (2019). Critical features for face recognition. *Cognition*, 182, 73-83. <https://doi.org/10.1016/j.cognition.2018.09.002>
- Ahonen, T., Hadid, A., & Pietikainen, M. (2006). Face Description with Local Binary Patterns: Application to Face Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(12), 2037-2041. <https://doi.org/10.1109/TPAMI.2006.244>
- Aissaoui, A., & Martinet, J. (2015). Bi-modal Face Recognition—How combining 2D and 3D Clues Can Increase the Precision: Proceedings of the 10th International Conference on Computer Vision Theory and Applications, 559-564. <https://doi.org/10.5220/0005359605590564>
- Al-Naji, A., Gibson, K., Lee, S.-H., & Chahl, J. (2017). Real Time Apnoea Monitoring of Children Using the Microsoft Kinect Sensor: A Pilot Study. *Sensors*, 17(2), 286. <https://doi.org/10.3390/s17020286>
- Ben Amor, B., Drira, H., Berretti, S., Daoudi, M., & Srivastava, A. (2014). 4D Facial Expression Recognition by Learning Geometric Deformations. *IEEE Transactions on Cybernetics*, 44(12), 2443-2457.
- Borghini, G., Pini, S., Grazioli, F., Vezzani, R., & Cucchiara, R. (2018, septiembre 3). Face Verification from Depth using Privileged Information. *BMVC 2018 - 29th British Machine Vision Conference*. *BMVC 2018 - 29th British Machine Vision Conference*, Newcastle, UK.
- Boumedine, A. Y., Bentaieb, S., Ouamri, A., & Mallek, A. (2021). 3D Face Identification Using HOG Features and Collaborative Representation. En M. R. Senouci, M. E. Y. Boudaren, F. Sebbak, & M. Mataoui (Eds.), *Advances in Computing Systems and Applications* (Vol. 199, pp. 3-13). Springer International Publishing. [https://doi.org/10.1007/978-3-030-69418-0\\_1](https://doi.org/10.1007/978-3-030-69418-0_1)
- Cai, Z., Han, J., Liu, L., & Shao, L. (2017). RGB-D datasets using microsoft kinect or similar sensors: A survey. *Multimedia Tools and Applications*, 76(3), 4313-4355. <https://doi.org/10.1007/s11042-016-3374-6>



- Cardia, J. B., & Marana, A. N. (2016). Score Fusion of 3DLBP and ACDN+P for 3D Face Recognition. *Journal on Advances in Theoretical and Applied Informatics*, 2(1), 34. <https://doi.org/10.26729/jadi.v2i1.1660>
- Cardia Neto, J. B., & Marana, A. N. (2015). 3DLBP and HAOG fusion for face recognition utilizing Kinect as a 3D scanner. *Proceedings of the 30th Annual ACM Symposium on Applied Computing - SAC '15*, 66-73. <https://doi.org/10.1145/2695664.2695807>
- Cardia Neto, J. B., & Marana, A. N. (2018). Utilizing Deep Learning and 3DLBP for 3D Face Recognition. En M. Mendoza & S. Velastín (Eds.), *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications* (Vol. 10657, pp. 135-142). Springer International Publishing. [https://doi.org/10.1007/978-3-319-75193-1\\_17](https://doi.org/10.1007/978-3-319-75193-1_17)
- Chen, J., Zhang, Z., Yao, L., Li, B., & Chen, T. (2019). Face Recognition Using Depth Images Base Convolutional Neural Network. *2019 International Conference on Computer, Information and Telecommunication Systems (CITS)*, 1-4. <https://doi.org/10.1109/CITS.2019.8862099>
- Chowdhury, A., Ghosh, S., Singh, R., & Vatsa, M. (2016). RGB-D face recognition via learning-based reconstruction. *2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, 1-7. <https://doi.org/10.1109/BTAS.2016.7791199>
- Ciaccio, C., Wen, L., & Guo, G. (2013). Face recognition robust to head pose changes based on the RGB-D sensor. *2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, 1-6. <https://doi.org/10.1109/BTAS.2013.6712718>
- Das, A., & Dantcheva, A. (2018, octubre 1). Mitigating Bias in Gender, Age, and Ethnicity Classification: A Multi-Task Convolution Neural Network Approach.
- Dong Li, Huiling Zhou, & Kin-Man Lam. (2015). High-Resolution Face Verification Using Pore-Scale Facial Features. *IEEE Transactions on Image Processing*, 24(8), 2317-2327. <https://doi.org/10.1109/TIP.2015.2412374>
- Elaiwat, S. (2015). Engineered and learned features for face and facial expression recognition. <https://research-repository.uwa.edu.au/en/publications/engineered-and-learned-features-for-face-and-facial-expression-re>

Facial Recognition Market Analysis by Technology, Application. (s. f.). Recuperado 17 de octubre de 2018, de <https://www.alliedmarketresearch.com/facial-recognition-market?facial-recognition-market>

Fernández-Cervantes, V., García, A., Ramos, M. A., & Méndez, A. (2015). Facial Geometry Identification through Fuzzy Patterns with RGBD Sensor. *Computación y Sistemas*, 19(3), 529-546. <https://doi.org/10.13053/cys-19-3-2015>

Fumio, O. (2020). Gait Recognition Using Convolutional Neural Network with RGB-D Sensor Data. 2020 IEEE/SICE International Symposium on System Integration (SII), 213-218. <https://doi.org/10.1109/SII46433.2020.9025838>

García Pinilla, P. (2020). Reconocimiento facial y extracción de información mediante Deep Learning a partir de datos obtenidos en red social. <https://academica-e.unavarra.es/xmlui/handle/2454/37576>

Goswami, G., Bharadwaj, S., Vatsa, M., & Singh, R. (2013a). On RGB-D face recognition using Kinect. 2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS), 1-6. <https://doi.org/10.1109/BTAS.2013.6712717>

Goswami, G., Bharadwaj, S., Vatsa, M., & Singh, R. (2013b). On RGB-D face recognition using Kinect. 2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS), 1-6. <https://doi.org/10.1109/BTAS.2013.6712717>

Goswami, G., Vatsa, M., & Singh, R. (2014). RGB-D Face Recognition With Texture and Attribute Features. *IEEE Transactions on Information Forensics and Security*, 9(10), 1629-1640. <https://doi.org/10.1109/TIFS.2014.2343913>

Hayat, M., Bennamoun, M., & El-Sallam, A. (2015). An RGB-D based Image Set Classification for Robust Face Recognition from Kinect Data. *Neurocomputing*, 171. <https://doi.org/10.1016/j.neucom.2015.07.027>

Hayat, M., Bennamoun, M., & El-Sallam, A. A. (2016). An RGB-D based image set classification for robust face recognition from Kinect data. *Neurocomputing*, 171, 889-900. <https://doi.org/10.1016/j.neucom.2015.07.027>

Huang, Y., Wang, Y., & Tan, T. (2006). Combining Statistics of Geometrical and Correlative Features for 3D Face Recognition. *BMVC*, 879-888.

Huynh, T., Min, R., & Dugelay, J.-L. (2013). An Efficient LBP-Based Descriptor for Facial Depth Images Applied to Gender Recognition Using RGB-D Face Data. En J.-I. Park & J. Kim (Eds.), *Computer Vision—*

ACCV 2012 Workshops (Vol. 7728, pp. 133-145). Springer Berlin Heidelberg. [https://doi.org/10.1007/978-3-642-37410-4\\_12](https://doi.org/10.1007/978-3-642-37410-4_12)

Jiang, L., Zhang, J., & Deng, B. (2020). Robust RGB-D Face Recognition Using Attribute-Aware Loss. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(10), 2552-2566. <https://doi.org/10.1109/TPAMI.2019.2919284>

Jiang, L., Zhang, J., Li, C., & Zhou, J. (2021). RGB-D Face Recognition via Spatial and Channel Attentions. 2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), 5, 2037-2041. <https://doi.org/10.1109/IAEAC50856.2021.9391062>

Karnewar, A., & Wang, O. (2020). MSG-GAN: Multi-Scale Gradients for Generative Adversarial Networks. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 7796-7805. <https://doi.org/10.1109/CVPR42600.2020.00782>

Karras, T., Aila, T., Laine, S., & Lehtinen, J. (2017). Progressive Growing of GANs for Improved Quality, Stability, and Variation.

Karras, T., Laine, S., & Aila, T. (2019). A Style-Based Generator Architecture for Generative Adversarial Networks. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 4396-4405. <https://doi.org/10.1109/CVPR.2019.00453>

Kishan Anapu, S., & Rao Peri, S. (2014). Fusion of RGB and Depth Images for Robust Face Recognition using Close-Range 3D Camera. *International Journal of Innovative Research in Computer and Communication Engineering*, Volume 2. <http://www.ijircce.com/volume-2-issue-7.html#>

Li, B. Y. L. (2013). Robust Face Recognition based on Color and Depth Information [Thesis, Curtin University]. <https://espace.curtin.edu.au/handle/20.500.11937/1334>

Li, B. Y. L., Mian, A. S., Liu, W., & Krishna, A. (2016). Face recognition based on Kinect. *Pattern Analysis and Applications*, 19(4), 977-987. <https://doi.org/10.1007/s10044-015-0456-4>

Liu, H., Li, D., Wang, X., Liu, L., Zhang, Z., & Subramanian, S. (2021). Precise head pose estimation on HPD5A database for attention recognition based on convolutional neural network in human-computer interaction. *Infrared Physics & Technology*, 116, 103740. <https://doi.org/10.1016/j.infrared.2021.103740>

- Mahdi Barraah, E., Ahdid, R., Safi, S., & Malaoui, A. (2017). 3DLBP and SVD Fusion for 3D Face Recognition Using Range Image. *Computer Science and Information Technology*, 5(2), 61-65. <https://doi.org/10.13189/csit.2017.050203>
- Mantecon, T., del-Bianco, C. R., Jaureguizar, F., & García, N. (2014). Depth-based face recognition using local quantized patterns adapted for range data. 2014 IEEE International Conference on Image Processing (ICIP), 293-297. <https://doi.org/10.1109/ICIP.2014.7025058>
- Min, R., Kose, N., & Dugelay, J.-L. (2014). KinectFaceDB: A Kinect Database for Face Recognition. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 44(11), 1534-1548. <https://doi.org/10.1109/TSMC.2014.2331215>
- Minami, T., Nakajima, K., & Nakauchi, S. (2018). Effects of Face and Background Color on Facial Expression Perception. *Frontiers in Psychology*, 9, 1012. <https://doi.org/10.3389/fpsyg.2018.01012>
- Mu, G., Huang, D., Hu, G., Sun, J., & Wang, Y. (2019). Led3D: A Lightweight and Efficient Deep Approach to Recognizing Low-Quality 3D Faces. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 5766-5775. <https://doi.org/10.1109/CVPR.2019.00592>
- Naik, N., & G.N, R. (2015). Robust Real Time Face Recognition and Tracking on GPU Using Fusion of RGB and Depth Image. *Computer Science & Information Technology ( CS & IT )*, 01-08. <https://doi.org/10.5121/csit.2015.50601>
- National Institute of Standards and Technology | NIST. (s. f.). Recuperado 16 de octubre de 2018, de <https://www.nist.gov/>
- Naz, F., Rehman, A., MIS Department CBA Salman bin Abdulaziz University Alkharj KSA, & Department of Computer Science, Comsats Institute of Information technology WahCantt. (2017). Face Recognition: A Survey. *Journal of Engineering Science and Technology Review*, 10(2), 166-177. <https://doi.org/10.25103/jestr.102.20>
- Neto, J. B. C., & Marana, A. N. (2014). Face Recognition Using 3DLBP Method Applied to Depth Maps Obtained from Kinect Sensors. *x workshop computer vision WVC*, 168-172.
- Neto, L. B., Grijalva, F., Maíke, V. R. M. L., Martini, L. C., Florencio, D., Baranauskas, M. C. C., Rocha, A., & Goldenstein, S. (2017). A Kinect-Based Wearable Face Recognition System to Aid Visually Impaired

- Users. IEEE Transactions on Human-Machine Systems, 47(1), 52-64.  
<https://doi.org/10.1109/THMS.2016.2604367>
- Novamizanti, L., Lima, N., & Susatio, E. (2019). Sistem Pengenalan Wajah 3D Menggunakan ICP dan SVM. Jurnal Teknologi Informasi dan Ilmu Komputer, 6, 601. <https://doi.org/10.25126/jtiik.2019661609>
- Ouannes, L., Khalifa, A. B., & Ben Amara, N. E. (2020). Facial Recognition in Degraded Conditions Using Local Interest Points. 2020 17th International Multi-Conference on Systems, Signals & Devices (SSD), 404-409. <https://doi.org/10.1109/SSD49366.2020.9364124>
- Romero, M., Flores, C., Muñoz, V., & Altamirano, L. C. (2016). Face recognition using Eigensurface on Kinect depth-maps. Proceedings of the International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV), 241.
- Saleh, Y. (2017). Face recognition enhancement through the use of depth maps and deep learning [Ph.D., Loughborough University]. <https://dspace.lboro.ac.uk/2134/26606>
- Segundo, M. P., Sarkar, S., Goldgof, D., Silva, L., & Bellon, O. (2013). Continuous 3D Face Authentication Using RGB-D Cameras. 64-69. <https://doi.org/10.1109/CVPRW.2013.17>
- Sepas-Moghaddam, A., Correia, P. L., Nasrollahi, K., Moeslund, T. B., & Pereira, F. (2018). LIGHT FIELD BASED FACE RECOGNITION VIA A FUSED DEEP REPRESENTATION. 2018 IEEE 28th International Workshop on Machine Learning for Signal Processing (MLSP), 1-6. <https://doi.org/10.1109/MLSP.2018.8516966>
- Soltanpour, S., Boufama, B., & Jonathan Wu, Q. M. (2017). A survey of local feature methods for 3D face recognition. Pattern Recognition, 72, 391-406. <https://doi.org/10.1016/j.patcog.2017.08.003>
- Uppal, H., Sepas-Moghaddam, A., Greenspan, M., & Etemad, A. (2021). Depth as Attention for Face Representation Learning. IEEE Transactions on Information Forensics and Security, 16, 2461-2476. <https://doi.org/10.1109/TIFS.2021.3053458>
- Vorobeychik, Y., & Kantarcioglu, M. (2018). Adversarial Machine Learning. Synthesis Lectures on Artificial Intelligence and Machine Learning, 12(3), 1-169. <https://doi.org/10.2200/S00861ED1V01Y201806AIM039>
- Wang, M., & Deng, W. (2021). Deep face recognition: A survey. Neurocomputing, 429, 215-244. <https://doi.org/10.1016/j.neucom.2020.10.081>

- Wang, Y., Cheung, S. W., Chung, E. T., Efendiev, Y., & Wang, M. (2020). Deep multiscale model learning. *Journal of Computational Physics*, 406, 109071. <https://doi.org/10.1016/j.jcp.2019.109071>
- Wasenmüller, O., & Stricker, D. (2017). Comparison of Kinect V1 and V2 Depth Images in Terms of Accuracy and Precision. En C.-S. Chen, J. Lu, & K.-K. Ma (Eds.), *Computer Vision – ACCV 2016 Workshops* (pp. 34-45). Springer International Publishing.
- Xiao, Y., & Xie, X. (2020). Application of Novel Gabor-DCNN into RGB-D Face Recognition. *Int. J. Netw. Secur.*
- Xu, X., Li, W., & Xu, D. (2015). Distance Metric Learning Using Privileged Information for Face Verification and Person Re-Identification. *IEEE Transactions on Neural Networks and Learning Systems*, 26(12), 3150-3162. <https://doi.org/10.1109/TNNLS.2015.2405574>
- Zafar, U., Ghafoor, M., Zia, T., Ahmed, G., Latif, A., Malik, K. R., & Sharif, A. M. (2019). Face recognition with Bayesian convolutional networks for robust surveillance systems. *EURASIP Journal on Image and Video Processing*, 2019(1). <https://doi.org/10.1186/s13640-019-0406-y>
- Zhang, H., Han, H., Cui, J., Shan, S., & Chen, X. (2018). RGB-D Face Recognition via Deep Complementary and Common Feature Learning. 2018 13th IEEE International Conference on Automatic Face Gesture Recognition (FG 2018), 8-15. <https://doi.org/10.1109/FG.2018.00012>
- Zhang, J., Huang, D., Wang, Y., & Sun, J. (2016). Lock3DFace: A large-scale database of low-cost Kinect 3D faces. 2016 International Conference on Biometrics (ICB), 1-8. <https://doi.org/10.1109/ICB.2016.7550062>
- Zhou, W., Chen, J., & Wang, L. (2015). A RGB-D face recognition approach without confronting the camera. 109-114. <https://doi.org/10.1109/CompComm.2015.7387550>
- Zohra, F. T., & Gavrilova, M. (2017). KINECT Face Recognition Using Occluded Area Localization Method. En *Transactions on Computational Science XXX* (pp. 12-28). Springer.

### **Conflicto de interés**

El autor autoriza la distribución y uso de su artículo.

### **Contribuciones de los autores**

1. Conceptualización: Alejandro Alvarez Navarro, Enrique Juan Marañon Reyes, Rubén Orozco Morales
2. Curación de datos: Alejandro Alvarez Navarro
3. Análisis formal: Alejandro Alvarez Navarro, Enrique Juan Marañon Reyes, Rubén Orozco Morales
4. Adquisición de fondos:
5. Investigación: Alejandro Alvarez Navarro
6. Metodología: Alejandro Alvarez Navarro, Enrique Juan Marañon Reyes, Rubén Orozco Morales
7. Administración del proyecto: Enrique Juan Marañon Reyes
8. Recursos:
9. Software: Alejandro Alvarez Navarro
10. Supervisión: Enrique Juan Marañon Reyes, Rubén Orozco Morales
11. Validación: Alejandro Alvarez Navarro
12. Visualización: Alejandro Alvarez Navarro
13. Redacción – borrador original: Alejandro Alvarez Navarro
14. Redacción – revisión y edición: Alejandro Alvarez Navarro