

Combined Approach using Artificial Vision and Neural Networks for Facial Recognition

William Gutiérrez Pezoa
Department of Electric Engineering.
Universidad de Santiago de Chile
Santiago, Chile
william.gutierrez.pezoa@gmail.com

Marcela Jamett Domínguez
Department of Electric Engineering
Universidad de Santiago de Chile, Member of IEEE.
Santiago, Chile
marcela.jamett@usach.cl

Abstract—An application of artificial vision and artificial neural networks techniques in face recognition, is presented. In order to do that, a set of images (frontal face photos) with different lighting conditions, gestures, accessories and distances is used. A stepwise algorithm allows to achieve a satisfactory results, obtaining the correct identification of images inside and outside the data set.

Keywords— Computer Vision, Artificial Neural Networks, Pattern Recognition, Face Identification.

I. INTRODUCCIÓN

La visión artificial (VA), es la encargada de adquirir imágenes con el fin de procesarlas, adaptarlas (dimensión de imagen, color de imagen, etc.), realizar un análisis de las diversas cualidades que posee el rostro de una persona y extraer esta información para realizar alguna clasificación específica.

El trabajo se compone de un estudio del estado del arte sobre la visión artificial (VA) y redes neuronales artificiales. Seguido de esto, se realiza el procesamiento de un set de imágenes de prueba, extrayendo las características del rostro a través del algoritmo de Viola Jones (V-J) [1]. Luego, se programa una red neuronal artificial (RNA) del tipo *backpropagation* [2] para reconocer el rostro de las personas para finalmente probar el desempeño de la RNA al ingresar un set de imágenes diferentes a las usadas en el entrenamiento por la red neuronal.

Este procesamiento se realizó mediante pruebas de imágenes estáticas tomadas desde una base de datos, que fueron procesadas a través de una interfaz hombre-máquina (HMI), por lo que queda propuesto como futuro trabajo, el entrenamiento y verificación de imágenes en movimiento.

En la presente investigación, se hace uso de un sistema de reconocimiento facial combinando las técnicas de VA y RNA, evaluando su funcionamiento y desempeño a través del porcentaje de similitud entre la imagen de prueba y la imagen encontrada.

II. ESTADO DEL ARTE

La VA, incluye métodos para adquirir, procesar, analizar y comprender las imágenes del mundo real, con el fin de producir información numérica o simbólica para que puedan ser tratados

por una computadora [3]. Esta, se caracteriza por poseer diversas etapas de procesamiento de imágenes, las cuales se pueden representar en la Fig. 1.

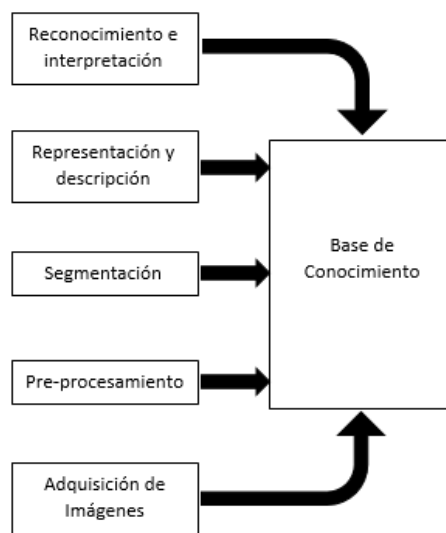


Fig. 1. Etapas generales de la visión artificial [1].

Con respecto a la extracción del rostro de la imagen total, encontramos el algoritmo de V-J [4] propuesto por Viola P. y Jones M. el 2001, el cual describe una máquina de aprendizaje visual, capaz de localizar objetos de forma muy rápida, logrando una alta tasa de detección. El algoritmo está dividido en tres etapas claves. La primera, consiste en una nueva representación de la imagen llamada “imagen integral”, que permite que las funciones utilizadas por nuestro computador se ejecuten rápidamente. La segunda etapa, consiste en un algoritmo de aprendizaje basado en AdaBoost [4], que selecciona un pequeño número de características visuales críticas de un conjunto más grande, convirtiéndose en un clasificador muy eficiente. La tercera fase, es un método llamado “cascada” que permite que las regiones de fondo de la imagen sean descartadas rápidamente, mientras que el costo computacional se centra en las regiones más prometedoras. Sus resultados muestran que pueden detectar el rostro de una persona 15 veces más rápido que cualquier sistema previo a esa fecha.

Adrian Wong [5], junto a su equipo de investigación, en el año 2015 presentan la aceleración del algoritmo de V-J en la detección de la cara frontal de un grupo de personas, logrando acelerar la velocidad de detección hasta 18 veces más que lo logrado con el algoritmo convencional.

Yongzheng Xu [6], dos años más tarde, desarrolla un método avanzado de detección de vehículos, que mejora el esquema de detección de objetos de V-J original, debido a que dicho algoritmo tiene dificultades para detectar vehículos con orientaciones desconocidas. Sus resultados muestran que lograron una detección del 81,36% en ambientes con iluminación variable.

Housseem Lahiani [7], junto a su equipo de investigación, en el mismo año, investiga la estimación de poses de la mano, proponiendo un sistema que reconoce gestos de la mano en tiempo real para interactuar con teléfonos móviles. Los resultados muestran que lograron un promedio de 91% de aciertos en la estimación.

Con respecto al reconocimiento de imágenes de una parte del rostro, Aisyah Kumala [8], junto a su equipo de investigación, usan una RNA *backpropagation* para detectar la producción creciente del ácido del estómago debido a dietas incorrectas. Para ello, diseñaron un sistema que analiza el iris del ojo humano a través de una cámara. La imagen obtenida es procesada por el algoritmo de análisis de componentes principales (PCA) y lo clasifican por medio de RNA. Sus resultados muestran que el sistema es muy eficiente al realizar una clasificación con un 87.5% de precisión a partir de 40 datos de imágenes de iris.

Mohammed Saaidia [9], propone una red neuronal para detectar y clasificar las expresiones faciales de un grupo de personas. Para ello, entrenaron una RNA *backpropagation* usando los momentos de Zernike y simulaban los rostros agrupando a las personas en categorías de estados de ánimo. Su mejor resultado fue de un 89.5% de precisión en el reconocimiento.

En el trabajo de Swati Jado [10], se proponen 4 técnicas de extracción de características del rostro en una imagen. Utiliza una RNA del tipo SOM para construir una base de datos y compara los resultados obtenidos mediante las cuatro técnicas que corresponden a la transformada de coseno discreta direccional, transformada discreta de wavelets, transformada de coseno discreta y detección de bordes SOBEL de MatLab. Sus resultados muestran que se obtuvo un reconocimiento del rostro desde un 84,33% hasta un 100%.

Por último, Ke Shan [11], junto a su equipo de investigación, en el año 2017, proponen una RNA que reconoce las expresiones del rostro de un grupo de personas, obteniendo un desempeño del 76.7%.

De acuerdo a estos resultados previos, en el presente trabajo se propone resolver un problema común que es el reconocimiento de rostros (o cualquier otro patrón) en forma robusta frente perturbaciones, tales como: cambios de gestos, iluminación y otras perturbaciones.

III. DISEÑO DEL EXPERIMENTO

El reconocimiento del rostro está dividido en dos etapas. La primera, corresponde a las técnicas de VA que contiene el procesamiento de imágenes y detección del rostro. La segunda involucra el reconocimiento del rostro a través de una RNA.

La adquisición de imágenes se realizó a partir de una base de datos que contenía 50 imágenes de 5 personas diferentes (10 imágenes de cada una), sin aplicación de filtros compensadores de luz y movimiento. Las imágenes fueron extraídas de una base de datos de libre acceso, proporcionado por la universidad de Yale [12]. Estas imágenes fueron llevadas al programa MatLab las cuales se guardaron en forma de matriz, presentada en la siguiente figura.

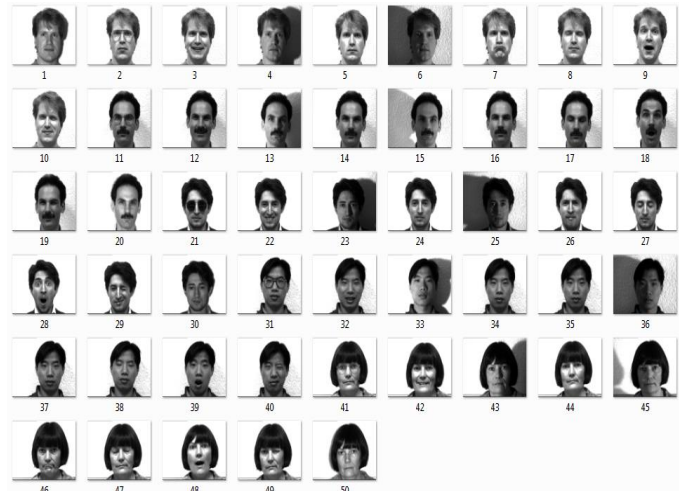


Fig. 2. Imágenes originales de la base de datos de Yale U. [12].

A. Técnicas de Visión Artificial

El pre-procesamiento con VA es el encargado de acondicionar las imágenes guardadas en la matriz. Comienza con la transformación de las imágenes en tres dimensiones (en color) a imágenes en escala de grises. Seguido de esto, se transforman las imágenes resultantes en imágenes binarias (en blanco y negro) para poder facilitar el reconocimiento del rostro [13].

La segmentación de las imágenes, consiste en la extracción del rostro de las imágenes pre-procesadas. Para ello, se utiliza el algoritmo de V-J [1], que realiza la extracción de características específicas del rostro, considerando diferentes zonas de interés, como pueden ser el contorno, los ojos, la nariz, entre otras.

En esta investigación, se seleccionó una aproximación del contorno del rostro en cada imagen, fijando una serie de puntos los cuales permiten la extracción del rostro. Una vez extraído el rostro de las 50 imágenes, se redimensionan las imágenes con un mismo tamaño y se procede a crear una nueva matriz con todos los rostros, de acuerdo a la fig. 3.

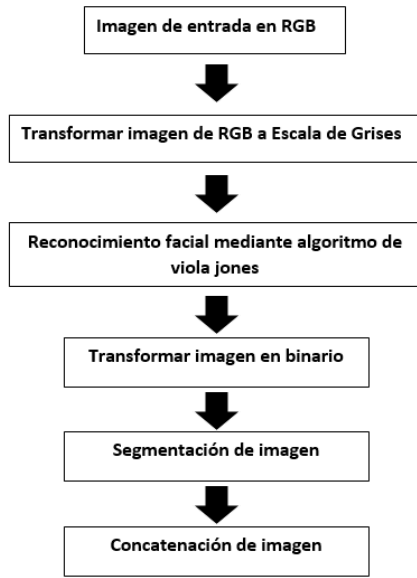


Fig. 3. Diagrama de funcionamiento del procesamiento de imágenes [3].

B. Red Neuronal Artificial

Previo al uso de la RNA, se debe acondicionar los datos de entrada, los cuales son la matriz de entrada y el target.

Los datos de la matriz de entrada, corresponden a la matriz de características con las imágenes ya segmentadas, obtenidas en la última etapa de visión artificial. Del universo de imágenes (50) ya procesadas por la VA, se seleccionan aleatoriamente 35 de ellas, correspondientes al 70%. Para ser incluidas en la fase de entrenamiento, mientras que el resto se deja para las pruebas. A estas 35 imágenes, se le agregan dos señales de ruido para aumentar la tolerancia en la etapa de reconocimiento. Luego de esto, se crea una nueva matriz de características que contiene la concatenación de las imágenes originales y las imágenes con ruido. Este procedimiento se esquematiza en la figura 4.

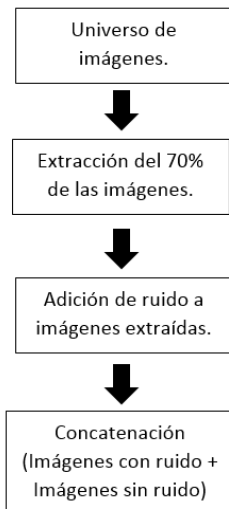


Fig. 4. Etapas del acondicionamiento de set de imágenes para la RNA.

La etapa de RNA está dividida en dos fases. La primera, corresponde al entrenamiento de la red neuronal y la segunda a su simulación.

Dentro de la primera fase, la RNA utilizada es del tipo *backpropagation*, el cual es un algoritmo de aprendizaje supervisado. Este tipo de red neuronal emplea un ciclo propagación-adaptación de dos fases [4]. Se aplica un patrón a la entrada de la red como estímulo, propagándose desde la primera capa hacia las capas siguientes hasta generar una salida. Esta señal se compara con la deseada y se calcula una señal de error para cada una de ellas. Las salidas de error parten de la capa de salida y se propagan hacia atrás, para luego llegar a las neuronas de la capa oculta que contribuyen directamente a la salida. Sin embargo, las neuronas de la capa oculta solo reciben una fracción de la señal total del error, basándose aproximadamente en la contribución relativa que haya aportado cada neurona a la salida original. Este proceso se repite, capa por capa, hasta que todas las neuronas de la red hayan recibido una señal de error que describa su contribución relativa al error total.

El algoritmo de entrenamiento utilizado es el gradiente descendente con rampa de aprendizaje adaptiva, esto debido a su efectividad y capacidad de anticiparse a los sucesos pasados, permitiendo una mayor adaptación frente a los cambios de rostros que se presentan. En la figura 5 se muestra la arquitectura de la red diseñada.

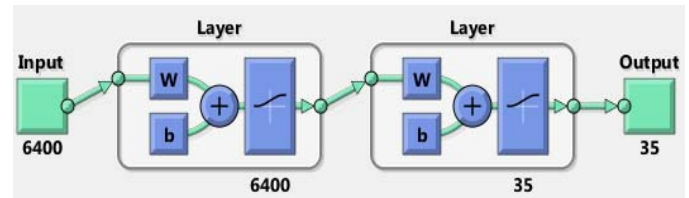


Fig. 5. Red neuronal *backpropagation* simulada en MatLab.

IV. RESULTADOS

La VA y la RNA *backpropagation*, es implementada mediante la utilización programa MatLab. Ambas técnicas fueron programadas en un computador marca Lenovo con un procesador Intel Core i7 y una memoria RAM de 8GB. La adquisición de las imágenes se realizó mediante un algoritmo que seleccionaba cada imagen de la base de datos. En este estudio en particular, las imágenes ya estaban en escala de grises, por lo que los pasos de la VA se reducen a los que se presentan a continuación.

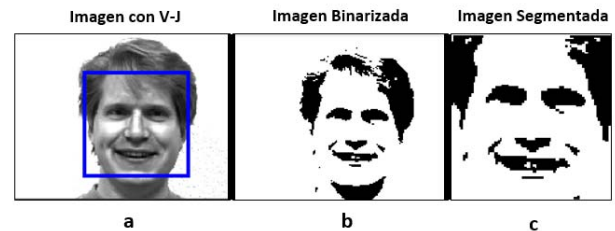


Fig. 6. Pasos de VA para una persona del set de imágenes. El proceso está dividido en tres etapas: a) Algoritmo V-J aplicado a la imagen, b) Imagen binarizada y c) Imagen segmentada.

El resultado de este procedimiento aplicado en las imágenes utilizadas por la RNA en su entrenamiento (incluyendo la adición de ruido y desordenamiento), se muestra en la figura 7.

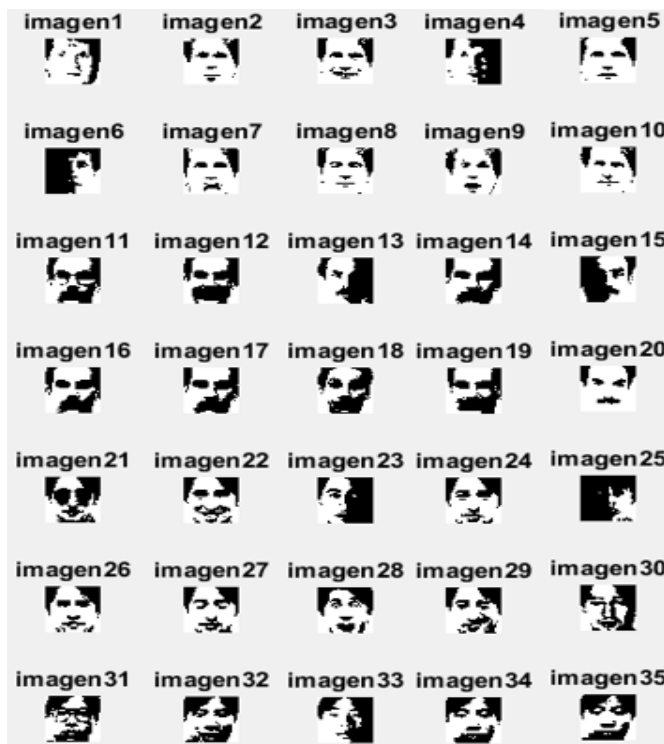


Fig. 7. Imágenes preprocesadas concatenadas en matriz.

El entrenamiento de la red neuronal tardó un poco más de una hora en finalizar, tiempo en el cual realizó un total de mil iteraciones y el performance alcanzó un valor de 0.00132. Este resultado se considera satisfactorio, por lo que se procede a simular la RNA y presentar su salida mediante una HMI.

V. HMI

Se programó una interfaz gráfica para la adquisición, visualización y procesamiento de imágenes, como lo indica la figura 8. Está compuesta por tres etapas. La primera, ubicada en el lado izquierdo, está diseñada para la adquisición de imágenes, ya sea a través de una cámara web o por una imagen guardada previamente en una base de datos. La segunda etapa está ubicada en el sector central de la interfaz gráfica, donde encontramos el procesamiento de imágenes, mostrándose la imagen seleccionada, permitiendo procesar imágenes en diversos formatos, ya sea en escala de grises, binarización, negativo y la imagen segmentada. También encontramos un botón de procesamiento de imágenes que activa la simulación de la red neuronal para encontrar la imagen con mayor similitud. Finalmente, en la tercera etapa, ubicada en el sector derecho de la interfaz gráfica, visualizamos los resultados obtenidos, donde aparece la imagen reconocida y el porcentaje con mayor similitud.



Fig. 8. Interfaz gráfica de adquisición de imágenes.

Para analizar el funcionamiento del sistema, se realizaron 15 pruebas con las imágenes que no participaron en la etapa de entrenamiento del presente sistema de identificación. El umbral para que pueden ser reconocidos es de un 5%. Por lo tanto, bajo este valor aparecerá como desconocido. A continuación, se muestran los casos más representativos de este conjunto de imágenes. De las 15 pruebas se reconocieron 12 imágenes, teniendo el sistema una confiabilidad de un 80%. La figura 9, muestra el resultado de la similitud más alta, la cual corresponde a un 25.6%.

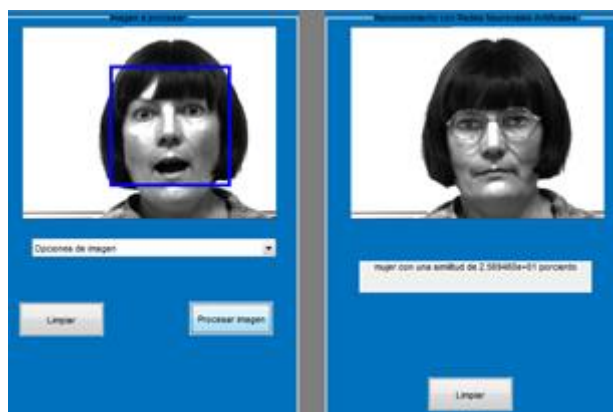


Fig. 9. Reconocimiento de persona con mayor similitud.

La figura 10, muestra el resultado de la similitud más baja encontrada, la cual corresponde a un 0.95 %



Fig. 10. Reconocimiento de persona con menor similitud.

También se realizó una prueba extra con una persona externa al set de entrenamiento y de pruebas, pudiéndose apreciar en la figura 11 que el porcentaje de similitud fue de un 4.3%. Por lo tanto, al no superar el umbral exigido, no se reconoce.

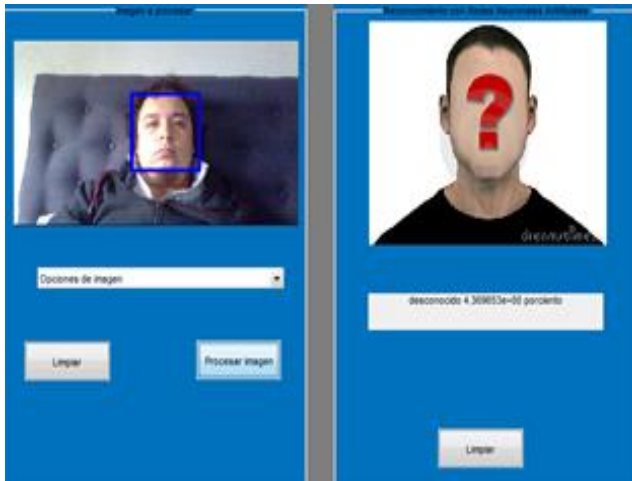


Fig. 11. Reconocimiento de persona distinta a las estudiadas.

La tabla I muestra los resultados del reconocimiento de 15 imágenes que no participaron en la etapa de entrenamiento, agregando una última imagen tomada desde una cámara web.

TABLA I
RESULTADOS DE RECONOCIMIENTO DE IMÁGENES SET DE PRUEBAS

Imagen	% de similitud	Reconocimiento Exitoso	Reconocimiento Fallido.
	7.91	X	
	1.42		X
	14.4	X	
	1.09		X
	6.5	X	
	5.02	X	
	0.95		X
	5.17	X	
	5.23	X	
	14.7	X	
	8.97	X	
	8.01	X	
	25.69	X	
	7.13	X	
	5.16	X	
	4.3	X	

De acuerdo a esto, la tase de acierto corresponde a un 81,25%.

VI. CONCLUSIONES

El reconocimiento facial a partir de un algoritmo por etapas, que combina técnicas de VA (preprocesamiento) y RNA (identificación) se ha desarrollado con éxito, permitiendo la validación del sistema con imágenes desconocidas (tanto con

fotos de las mismas personas del set de entrenamiento como con imágenes de personas ajenas a éste.

Para comparar los resultados obtenidos con investigaciones ya realizadas, nos remitiremos a los presentados en el estado del arte, donde se alcanza una tasa de acierto entre 76.6% y 100%. Sin embargo, estos resultados corresponden a distintas aplicaciones y no especifican la robustez frente a cambios inesperados en las imágenes, la cual es la principal contribución del presente *paper*.

Se espera en un futuro próximo, incluir también mayores complejidades en el sistema de identificación: mayor cantidad de imágenes, incluyendo perturbaciones tales como giros, movimientos, entre otras.

VII. REFERENCIAS.

- [1] Viola P, Jones M. "Robust real-time object detection". In: Second International Workshop on Statistical and Computation Theories of Vision-Modeling, Learning, Computing and Sampling. Vancouver, Canada, pp. 1-4, 2001.
- [2] Seniman, D. Arisandi, R. Fadillah, William and E. Budhiarti, "Chinese chess character recognition using direction feature extraction and backpropagation", IEEE International Conference on Data and Software Engineering (ICoDSE), Indonesia, pp. 1-6, 2017.
- [3] R. Szeliski, *Computer Vision: Algorithms and Applications*, University of Washington, Washington, 2010, pp. 1-50.
- [4] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features", Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Kauai, USA, pp. 1-8, 2003.
- [5] A. Wong and S. Mohd, "GPU acceleration of real time Viola-Jones face detection", IEEE International Conference on Control System, Computing and Engineering (ICCSCE), Malaysia, pp. 1-6, 2015.
- [6] Y. Xu, G. Yu and X. Wu, "An Enhanced Viola-Jones Vehicle Detection Method From Unmanned Aerial Vehicles Imagery", USA, IEEE Transactions on Intelligent Transportation Systems, pp. 1-12, 2017.
- [7] H. La hiani, M. Kherallah and M. Neji, "Hand pose estimation system based on Viola-Jones algorithm for Android devices", IEEE/ACS 13th International Conference of Computer System and Applications (AICCSA), Morocco, pp. 1-6, 2017.
- [8] A. Kumala, A. Novianty and T. Waluyo, "Stomach disorder detection trough the iris image using backpropagation neural network", IEEE International Conference on Informatics and Computing (ICIC), Indonesia, pp. 1-6, 2016.
- [9] M. Saaidia, A. Zermi and M. Ramdani, "Facial expression recognition using neural network trained with Zernike moments", IEEE International Conference on Artificial Intelligence with Applications in Enginnering and Technology, Kota Kinabalu, pp. 1-6, USA, 2015.
- [10] S. Jadon, M. Kumar and Y. Rathi, "Face recognition using SOM neural network with DDCT facial feature extraction techniques", IEEE International Conference on Communications and Signal Processing (ICCS), India, pp. 1-5, 2017.
- [11] K. Shan, J. Guo and W. You, "Automatic facial expression recognition based on a deep convolutional-neural-network structure" IEEE 15th International Conference on Software Engineering Research, Management and Applications (SERA), London, pp. 1-6, 2017.
- [12] D. Radev, Yale Face Database [online], EEUU: Universidad de Yale, 2006. Disponible en <http://www.cad.zju.edu.cn/home/dengcai/Data/FaceData.html>.
- [13] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms", IEEE Transactions on Systems, Man, and Cybernetics, Vol. 9, No. 1, pp. 62-66, 1979.