

DESCRIPCIÓN DE MÓDULOS QUE INTEGRAN LA HERRAMIENTA DE APOYO PARA IDENTIFICAR EXPRESIONES FACIALES.

M. en C. Monserrat Gabriela PÉREZ VERA
ESCOM-IPN

mperezve@ipn.mx

Ing. José Antonio ORTIZ RAMÍREZ
ESCOM-IPN

jaortiz@gmail.com

M. en C. Sandra Mercedes PÉREZ VERA
ESCOM-IPN

sperezv@ipn.mx

Escuela Superior de Cómputo del Instituto Politécnico Nacional

La propuesta del presente trabajo consiste en el describir los módulos más importantes que integran una herramienta de reconocimiento de expresiones faciales. Dichas expresiones serán disparadas a través de la influencia de estímulos externos (auditivos, visuales, etc.) presentadas al usuario y serán reconocidas por medio de una implementación ligeramente modificada del Sistema de Codificación Facial propuesto por Carl- Herman Hjortsjö y mejorado posteriormente por Paul Ekman en el desarrollo de su Modelo de Emociones Universales. Se requiere de una fotografía inicial del usuario y fotografías posteriores a la presentación del estímulo que se presente al mismo.

Introducción

Las expresiones faciales, en conjunto con el lenguaje corporal y la mirada, son los medios más efectivos para la comunicación de nuestro estado de ánimo y de las emociones que ciertas situaciones no hacen sentir. Es decir, a través de la interpretación y comprensión de las expresiones faciales, es posible tener una mejor idea de lo que la gente alrededor de uno quiere expresarnos.

Actualmente se han desarrollado sistemas para el reconocimiento facial por medio de distintos algoritmos, entre los cuales destacan las redes neuronales, llegando a tener un 95% de efectividad en el reconocimiento facial.

De igual forma, en el campo del reconocimiento de emociones, se utilizará el modelo de Emociones Universales propuesto por el Dr. Paul Ekman, el cual define 5 emociones primarias (también llamados por el Dr. Paul Ekman como continentes emocionales), en primera instancia:

- Felicidad
- Ira
- Miedo
- Tristeza
- Asco

Metodológica

La herramienta de apoyo está constituida de por cinco módulos principales; adquisición de la imagen, tratamiento de la imagen, detección de rostro, extracción de puntos característicos principales y un algoritmo de reconocimiento de expresiones faciales.

El primer módulo se encuentra constituido por una cámara web como interfaz de adquisición de imagen para su posterior análisis.

El apartado de tratamiento de la imagen consta de dos módulos secundarios que es la transformación de imagen a escala de grises y la implementación de CLAHE (Contrast Limited Adaptive Histogram Equalization) esto con la finalidad de mejorar el contraste de la imagen (Zuiderveld, 1994) buscando una mayor precisión a la hora de marcar los puntos principales en el rostro.

El módulo de detección de rostro el cual únicamente consiste en la extracción de coordenadas en el cual se encuentra el rostro del sujeto a estudiar para que el área restante sea ignorada.

Para el bloque de extracción de puntos característicos principales consiste en la implementación de una pirámide de imágenes, la cual ampliará o disminuirá las dimensiones de la imagen para apoyar al proceso de extracción de objetos para lo cual se utiliza un histograma de gradientes orientados.

Para el algoritmo de reconocimiento de emociones se hace uso de una red neuronal del tipo backpropagation, dicha red permitirá a la herramienta, clasificar las expresiones que se ingresen al sistema, haciendo una aproximación a los patrones con los cuales sea entrenada.

A. Tratamiento de imágenes

1) Características de validación de imagen: La imagen capturada por la cámara web debe poseer ciertos criterios para ser válida como lo son, la posición y distancia del rostro así como el control de la iluminación.

2) Conversión de imagen a escala de grises: La imagen se convertirá a escala de grises para disminuir la cantidad distorsiones que las sombras de la iluminación podrían tener sobre la imagen adquirida (Gaytan, 2013).

3) Implementación de CLAHE: utilizado para aumentar el contraste de la imagen, lo que permite que los detalles del rostro sean más claros para la detección de los puntos característicos.

B. Proceso de adquisición de imágenes. Para la adquisición de las imágenes, será necesario solicitar al usuario en cuestión se registre dentro de la herramienta, lo cual incluye la toma de una foto, la cual funcionará dentro del sistema como imagen preliminar.

Una vez el usuario se haya registrado, será capaz de realizar las pruebas, las cuales son la presentación de estímulos, ya sean visuales o auditivos, es de vital importancia, pedir al usuario en cuestión no gire o agache su rostro, de lo contrario las pruebas tendrán que repetirse, al no ser identificado ningún rostro.

C. Entrenamiento de la RNA. La propagación hacia atrás de errores o retro propagación (del inglés backpropagation) es un algoritmo de aprendizaje supervisado que se usa para entrenar redes neuronales artificiales. El algoritmo emplea un ciclo propagación –adaptación de dos fases. Una vez que se ha aplicado un patrón a la entrada de la red como estímulo, este se propaga desde la primera capa a través de las capas superiores de la red, hasta generar una salida. La señal de salida se compara con la salida deseada y se calcula una señal de error para cada una de las salidas.

La importancia de este proceso consiste en que, a medida que se entrena la red, las neuronas de las capas intermedias se organizan a sí mismas de tal modo que las distintas neuronas aprenden a reconocer distintas características del espacio total de entrada. Después del entrenamiento, cuando se les presente un patrón arbitrario de entrada que contenga ruido o que esté incompleto, las neuronas de la capa oculta de la red responderán con una salida activa si la nueva entrada contiene un patrón que se asemeje a aquella característica que las neuronas individuales hayan aprendido a reconocer durante su entrenamiento. Esto último hace referencia al patrón general de una red neuronal artificial el cual no permite un desconocimiento total siempre se encarga de buscar semejanza de patrones con la finalidad de obtener una salida (Rojas, 1996).

D. Puntos característicos. La investigación de reconocimiento facial "Reconocimiento de caras con características locales", toma 15 puntos característicos para reconocer las características locales, los cuales se empalman con los puntos a tomar en cuenta por las micro expresiones. Dichos puntos son listados a continuación:

□ Ojo izquierdo:

1. Izquierda (LEyeOutside)
2. Arriba (LEyeBrowPeak)
3. Centro (LEye)
4. Derecha (LEyeInside)

□ Ojo derecho:

5. Izquierda (REyeInside)
6. Arriba (REyeBrowPeak)
7. Centro (REye)
8. Derecha (REyeOutside) □ Nariz:
9. Izquierda (LNoseBottom)
10. Abajo (CNoseTip)
11. Derecha (RNoseBottom) □ Boca:
12. Izquierda (LMouthCorner)

13. Arriba (CMouthTop)
14. Derecha (RMouthCorner)
15. Abajo (CMouthBottom)

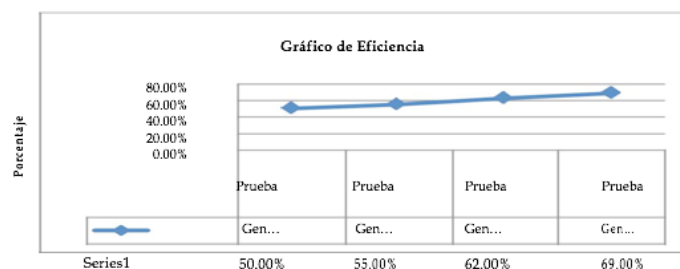
Los puntos anteriores son los mismos puntos analizados por las micro expresiones, lo cual posteriormente ayudará en la detección de los cambios entre expresiones faciales (Aguerrebere y Capdehourat, 2006).

Resultados

Cabe mencionar que una vez terminadas las pruebas, fue necesario realizar ajustes nuevamente a la red neuronal, teniendo un total de 4 versiones de la RNA y realizando después de cada ajuste una sesión de pruebas.

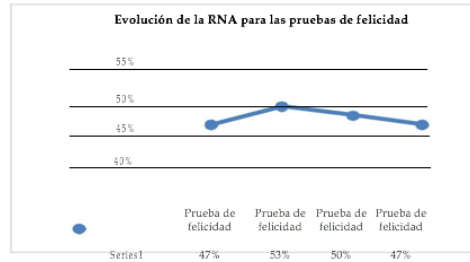
Se realizaron pruebas para las cuatro versiones de la RNA. Para las primeras dos versiones de la RNA, se hicieron pruebas a 21 personas, tomando estímulos aleatorios para cada una de ellas, tomando solo dos imágenes posteriores a los estímulos mostrados, el rango de edad de las personas que se tomaron en cuenta para las pruebas se encuentra entre los 25 a 35 años de edad. De las 21 personas seleccionadas, de esas 21 personas, 15 de ellas fueron hombres y 6 mujeres. El cambio entre la primera versión de la RNA y la segunda, fue la disminución de la tolerancia de error de un 5% a 1%, dejando la mínima cantidad posible de error que tomarían los patrones de la red neuronal. Para el caso de la segunda versión respecto a la tercera se realizó un aumento de patrones pasando de 40 patrones a 100 en una distribución de 35 de felicidad, 35 de enojo y 30 de tristeza.

Finalmente, de la tercera versión a la cuarta versión se volvieron a incrementar los patrones de 100 a 440 en una distribución de 146 de felicidad, 146 de enojo y 148 de tristeza. Como se observa a continuación, la red neuronal tuvo una mejora del 19% respecto de la primera versión con la cuarta y última.



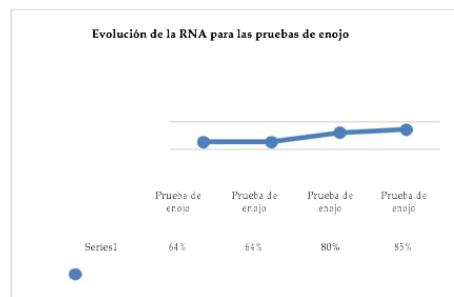
Gráfica 1. Gráfico de eficiencia en los resultados generales de las pruebas por versión.

Respecto a las pruebas de felicidad, la red neuronal tuvo la siguiente evolución, siendo el patrón de expresión facial con menor desempeño en la red neuronal:



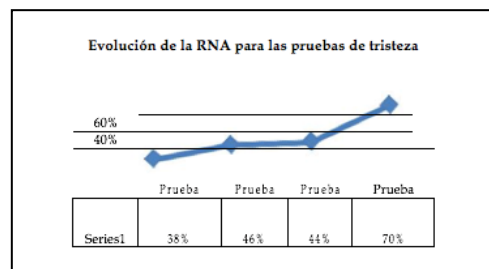
Gráfica 2. Gráfico de la evolución de la RNA en cuanto a los patrones de felicidad.

Para el caso de las pruebas de enojo, el desempeño de la red neuronal fue el siguiente:



Gráfica 31. Gráfico de la evolución de la RNA en cuanto a los patrones de enojo.

Finalmente, para las pruebas de tristeza, el desempeño de la red neuronal fue el siguiente:



Gráfica 4. Gráfico de la evolución de la RNA en cuanto a patrones de tristeza.

Cabe mencionar que la efectividad de la RNA. Puede aumentar entre más clasificaciones se tengan y el número de patrones incrementa.

Conclusiones

El objetivo del trabajo terminal es la identificación de las expresiones faciales a través de los trabajos y avances del Dr. Paul Ekman, lo cual se lleva a cabo mediante la toma de una imagen preliminar en la cual se marcan los puntos marcados por el Dr. Ekman como la base de su Sistema Universal de Emociones que se encuentran de igual forma mapeados dentro del FACS. De igual forma, se realizaron un par de cambios respecto al modelo original del Dr. Ekman, ya que al realizar pruebas con el sistema, ciertos puntos no eran detectados por la

herramienta, por lo cual se reemplazaron los puntos que rodean a la nariz, por los puntos relativos a las cejas, lo cual permitió a la herramienta hacer una mejor discriminación entre expresiones.

Un aspecto a tomar en cuenta es mejorar las características de los equipos en que se ejecutan, ya que la herramienta tarda aproximadamente entre 1 minuto y minuto y medio en realizar el análisis y clasificación de las expresiones faciales, por lo que para optimizar tiempos en cuanto a resultados, un equipo con mayores características sería necesario.

Referencias

Aguerreberre, C. y Capdehourat, G. (2006). Reconocimiento de caras con características locales. Proyecto Final Reconocimiento de Patrones. Uruguay: Instituto de Ingeniería Eléctrica Universidad de la Republica

Ekman, P. (2017). Paul Ekman. Recuperado de:
https://es.wikipedia.org/wiki/Paul_Ekman#cite_note-5

Gaytan, S. (2013). La importancia de la escala de grises. Recuperada de:
<http://hometech.com.mx/la-importancia-de-la-escala-de-grises/>

Rojas, R. (1996). Neural Networks: A Systematic Introduction. E:U: Springer.

Zuiderveld, K. (1994), Contrast limited adaptive histogram equalization. Graphics gems IV, Academic Press Professional, Inc., pp. 474-48.