

Sistema Inteligente de Visión por Computadora para el Reconocimiento de Expresiones Faciales

Victor E. Anchondo Trejo

Montana 2038, Chihuahua, Chihuahua, México

Abstract

En el sector comercial, el análisis y el impacto publicitario de nuevos productos es de mucha importancia para las empresas, porque sin un análisis o estudio de mercado apropiado puede derivar a que el lanzamiento de un producto o servicio sea una total catástrofe.

Debido a esto las grandes empresas destinan una cantidad considerable de recursos para la elaboración de estos análisis y así poder predecir el impacto de su próximo lanzamiento.

Al aplicar tecnologías actuales como el Procesamiento de Lenguaje Natural se puede automatizar, estandarizar y obtener resultados más detallados del impacto que puede tener el nuevo producto sobre un público en específico al analizar las expresiones faciales de cada persona para así determinar el impacto que hubo sobre el público y compararlo con el impacto esperado.

Keywords: Computer Vision, Natural Language Processing, Machine Learning, Facial Expression Recognition, Survey, Emotion Detection

1. Introduction

La finalidad es crear un sistema que pueda ser utilizado por entidades de gobierno, empresas privadas o partidos políticos que necesitan presentar algún producto nuevo, imagen, comercial o discurso y poder obtener

las expresiones de la muestra de personas a quienes se lo presentan antes de salir al mercado para saber el impacto que puede llegar a tener y si es el deseado.

Actualmente esto se realiza haciendo cuestionarios a cada una de las personas pertenecientes al grupo muestra, con esta aplicación se pretende aumentar la cantidad de información recolectada para así ayudar a que las empresas puedan tomar una decisión final mas informada.

Email address: eduardoanchondo@gmail.com
(Victor E. Anchondo Trejo)

URL: www.vanchondo.com (Victor E. Anchondo Trejo)

Ya que no se espera la misma reacción para diferentes situaciones, por ejemplo cuando se presenta el tráiler de una nueva película de terror, o el nuevo anuncio de la consola de video juegos de ultima generación, o del discurso presidencial para anunciar nuevos impuestos.

Es por eso por lo que debemos recolectar información de las expresiones faciales del público muestra durante la presentación y contar con los resultados de impacto para cualquier momento/segmento de la presentación para así determinar si el impacto obtenido en las personas es el que se desea.

2. Related Works

Con el auge de los algoritmos de Inteligencia Artificial y los algoritmos de visión por computadora, ya existen otros estudios similares al presentado, tales como:

- Deep Facial Expression Recognition: A Survey[1].
- The first facial expression recognition and analysis challenge[2].
- Facial expression recognition from video sequences: temporal and static modeling.[3].
- Facial expression recognition based on Local Binary Patterns: A comprehensive study[4].
- Facial Expression Recognition by De-Expression Residue Learning[5].
- Facial expression recognition using facial movement features[6].
- Facial expression recognition with identity and emotion joint learning[7].
- Annotation: Development of facial expression recognition from childhood to adolescence: Behavioural and neurological perspectives[8].
- Does facial expression recognition provide a toehold for the development of emotion understanding?[9].
- Emotion-modulated attention improves expression recognition: A deep learning model[10].
- Deep-emotion: Facial expression recognition using attentional convolutional network[11].
- Laplacian Nonlinear Logistic Step-wise and Gravitational Deep Neural Classification for Facial Expression Recognition[12].
- A facial expression recognizer using modified ResNet-152[13].
- Evoker: Narrative-based Facial Expression Game for Emotional Development of Adolescents[14].

Estos estudios cuentan con bases similares al aquí presentado, la gran diferencia es que estos no generan resultados para un lapso de tiempo, ni tampoco un resultado final sobre el impacto de lo que está viviendo o viendo la persona.

3. Methodology

La metodología utilizada es CRISP-DM.

1. Comprensión del negocio (Business Understanding).

Los objetivos de este proyecto son:

- Poder reconocer e interpretar las expresiones faciales de un grupo de personas durante la presentación de un discurso, comercial o producto nuevo.
- Una vez procesada la información, poder entregar al cliente un

2. Comprensión de datos (Data Understanding)

Para poder comprender e interpretar la información recolectada, se diseñará un clasificador el cual será entrenado utilizando un data set existente llamado AffectNet[15], el cual cuenta con las siguientes clasificaciones:

Table 1: Número de imagenes clasificadas en cada categoría. [15]

Expression	Number
Neutral	80,276
Happy	146,198
Sad	29,487
Surprise	16,288
Fear	8,191
Disgust	5,264
Anger	28,130
Contempt	5,135
None	35,322
Uncertain	13,163
Non-Face	88,895

3. Preparación de datos (Data Preparation).

Para preparar los datos se realizan los siguientes pasos:

- La información será capturada con una o mas cámaras las cuales tomaran fotos al grupo muestra de personas durante la presentación.
- Se procesan las fotos para separar los rostros detectados en una foto en imágenes diferentes.

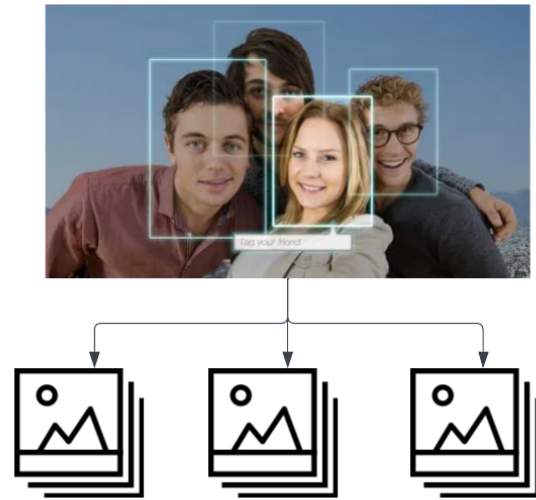


Figure 1: Imagen ilustrativa de la detección y separación de los rostros en multiples imagenes.

Source: <https://computerhoy.com/noticias/life/reconocimiento-facial-ventajas-peligros-revolucion-69441>

- Se aplican técnicas de limpieza sobre la imagen para eliminar la existencia de ruido.

4. Modelado (Modeling)

Las imágenes se categorizan por lapsos

de tiempo, por ejemplo, cada 5 segundos, este tiempo es personalizable, para así poder obtener los cambios de los rostros faciales del publico de una manera promediada y saber mejor el impacto que se esta obteniendo debido a la presentación.

5. Evaluación (Evaluation)

Para la evaluación, se van a ejecutar múltiples experimentos con grupos de personas diversos, por ejemplo, un grupo de estudiantes de primaria, secundaria y prepa en donde se les presente el tráiler de una nueva película.

Al finalizar, se realiza una encuesta para obtener información básica sobre que fue lo que sintieron durante la presentación. De esta forma podremos verificar si los resultados obtenidos por el sistema son similares al obtenido de la encuesta.

Algo a tener en mente es que para las personas es mas fácil quedarse con el sentimiento de lo ultimo que vieron, entonces si el inicio de la presentación les pareció muy interesante y el final muy aburrido, podríamos llegar a obtener resultados diferentes, dicho esto, se recomienda que para las evaluaciones se utilicen presentaciones de periodos cortos y que no abusen de las emociones del publico.

6. Implementación (Deployment)

Para la implementación se planea hacer uso de servicios en la nube para realizar todo el procesamiento de la información. Y el uso de un equipo moderado para la captura de las imágenes tales como cámaras y una computadora para poder subir esta información a nuestro servicio

para iniciar el procesamiento.

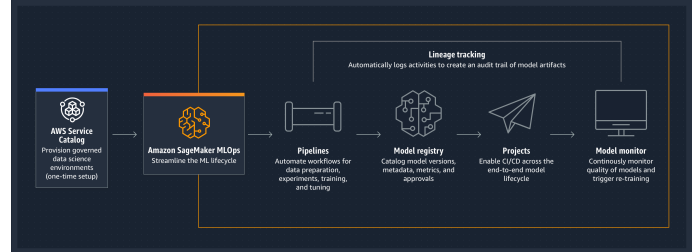


Figure 2: Amazon SageMaker

Source: <https://aws.amazon.com/sagemaker/mlops/>

4. Results

En esta primera etapa del software, se realizaron pruebas con personas de diferentes edades, las categorías que vamos a manejar son:

- Infantes: Entre 6 y 11 años.
- Adolescentes: Entre 12 y 17 años.
- Jóvenes: Entre 18 y 26 años.
- Adultos: Entre 27 y 59 años.

La cantidad de personas que participaron de las diferentes categorías son:

Table 2: Número de personas que participaron por categoría.

Categoría	Cantidad
Infantes	55
Adolescentes	53
Jóvenes	34
Adultos	27

Después de haber realizado pruebas con diferentes personas de las diferentes categorías en condiciones apropiadas (iluminación adecuada, poco movimiento, y fotografías de alta calidad) se obtuvieron los siguientes resultados:

Table 3: Número de personas que participaron por categoría.

Categoría	Cantidad	Aciertos	Errores	% de Aciertos	% de Errores
Infantes	55	37	18	67.27%	32.72%
Adolescentes	53	40	13	75.47%	24.52%
Jóvenes	34	26	8	76.47%	23.52%
Adultos	27	22	5	81.48%	18.51%
Total Promedio:	42.25	31.25	11	73.96%	26.03%

Podemos observar que se obtuvo un promedio global de 73.96% de aciertos y de 26.03% de errores. Estos resultados son aceptables para esta primera etapa del sistema.

En posteriores versiones del sistema, después de seguir entrenando la red con más personas de distintas edades y de distintas regiones se espera lograr un porcentaje de aciertos cercano al 90%.

Para lograr esto, se está planificando realizar visitas a escuelas de diferentes niveles que nos permitan recopilar la información sin perturbar las actividades diarias de los alumnos. Así mismo también se está colaborando con empresas del sector privado para realizar más captura de información.

5. Conclusion

Para ser esta la primera etapa del sistema nos genera buenos resultados y nos da áreas de oportunidad para mejorar la captura de información y el procesamiento de las imágenes

para garantizar el anonimato y la seguridad de la gente que nos ayuda para realizar el entrenamiento y las pruebas.

Para la captura de información tenemos las siguientes áreas de oportunidad:

- Aumentar los requisitos de la calidad de las cámaras que se utilizaran para la captura de los rostros.

Esto debido a que si se utilizan cámaras de baja resolución/calidad es más difícil para el sistema poder reconocer el rostro y la emoción que hay en él.

- Mejorar los algoritmos de filtrado. Esto porque cuando las fotografías son tomadas a una persona en movimiento, la limpieza y el filtrado de la imagen juega un papel muy importante para poder obtener una imagen del rostro de buena calidad.

- Capacitar al personal que acude a escuelas de nivel básico.

Al tratar con infantes es más difícil mantener la concentración del grupo y evitar que se muevan demasiado de sus lugares, por lo tanto, se requiere llevar personal capacitado para tratar con infantes.

- Mejorar la compresión de las imágenes. Esto es porque después de capturar una cantidad considerable de fotografías, el manejo y el procesamiento de estas se vuelve difícil y se debe de considerar que pueden existir localizaciones donde el acceso a internet sea un poco más limitado.

References

- [1] S. Li, W. Deng, Deep facial expression recognition: A survey, *IEEE transactions on affective computing* (2020).
- [2] M. F. Valstar, B. Jiang, M. Mehu, M. Pantic, K. Scherer, The first facial expression recognition and analysis challenge, in: 2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG), IEEE, 2011, pp. 921–926.
- [3] I. Cohen, N. Sebe, A. Garg, L. S. Chen, T. S. Huang, Facial expression recognition from video sequences: temporal and static modeling, *Computer Vision and image understanding* 91 (1-2) (2003) 160–187.
- [4] C. Shan, S. Gong, P. W. McOwan, Facial expression recognition based on local binary patterns: A comprehensive study, *Image and vision Computing* 27 (6) (2009) 803–816.
- [5] H. Yang, U. Ciftci, L. Yin, Facial expression recognition by de-expression residue learning, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 2168–2177.
- [6] L. Zhang, D. Tjondronegoro, Facial expression recognition using facial movement features, *IEEE transactions on affective computing* 2 (4) (2011) 219–229.
- [7] M. Li, H. Xu, X. Huang, Z. Song, X. Liu, X. Li, Facial expression recognition with identity and emotion joint learning, *IEEE Transactions on affective computing* 12 (2) (2018) 544–550.
- [8] C. Herba, M. Phillips, Annotation: Development of facial expression recognition from childhood to adolescence: Behavioural and neurological perspectives, *Journal of Child Psychology and Psychiatry* 45 (7) (2004) 1185–1198.
- [9] P. S. Strand, A. Downs, C. Barbosa-Leiker, Does facial expression recognition provide a toehold for the development of emotion understanding?, *Developmental Psychology* 52 (8) (2016) 1182.
- [10] P. Barros, G. I. Parisi, C. Weber, S. Wermter, Emotion-modulated attention improves expression recognition: A deep learning model, *Neurocomputing* 253 (2017) 104–114.
- [11] S. Minaee, M. Minaei, A. Abdolrashidi, Deep-emotion: Facial expression recognition using attentional convolutional network, *Sensors* 21 (9) (2021) 3046.
- [12] M. B. Kumari, D. B. Sivagami, Laplacian nonlinear logistic stepwise and gravitational deep neural classification for facial expression recognition, Available at SSRN 4096801.
- [13] W. Xu, R. S. Cloutier, A facial expression recognizer using modified resnet-152, *EAI Endorsed Transactions on Internet of Things* 7 (28) (2022) e5–e5.
- [14] S. Hong, Y. Choi, Y. Sung, Y. Jin, Y. Y. Doh, J. Lee, Evoker: Narrative-based fa-

cial expression game for emotional development of adolescents, in: CHI Conference on Human Factors in Computing Systems Extended Abstracts, 2022, pp. 1–8.

- [15] A. Mollahosseini, B. Hasani, M. H. Mahoor, Affectnet: A database for facial expression, valence, and arousal computing in the wild, *IEEE Transactions on Affective Computing* 10 (1) (2017) 18–31.