

## Análisis e Invariancias en Gestos Faciales Primarios

MONTECILLO-PUENTE, Francisco†, SAMANO-ABONCE, Obed, LOPEZ-ENRIQUEZ, Renato & MEDINA-REZA, Alejandro

*Instituto Tecnológico Superior de Salvatierra*

Recibido 27 de Abril, 2015; Aceptado 10 de Junio, 2015

### Resumen

El objetivo de este trabajo es realizar un análisis de gestos faciales humanos primarios a través del uso del sensor visión, la deformación dinámica temporal de señales y la mezcla de Gaussianas. Además, se presenta una base de datos de gestos faciales para participantes en México. También, se presenta una metodología para extraer invariancias en gestos faciales, así como un procedimiento para su clasificación. El sensor visión es utilizado para capturar información de gestos faciales primarios, estos son modelados mediante curvas generadas por un conjunto de puntos de interés. Estas curvas para ser comparadas entre si se normalizan a través del uso de la deformación dinámica temporal. Posteriormente, para extraer la propiedad de invariancia utilizamos la mezcla de Gaussianas, este mismo procedimiento nos permite desarrollar un mecanismo de clasificación de gestos. Las principales contribuciones de este trabajo son la generación de una base de datos de gestos faciales en México y la determinación de sus curvas invariantes.

**Gestos Faciales, Visión por Computadora, Curvas Invariantes, Deformación Dinámica Temporal, Mezclas de Gaussianas**

### Abstract

In this work, an analysis of primary facial gestures is performed through the use of visual information, bsplines curves, dynamic time warping and Gaussian mixture models. Also, a procedure to build a database of facial gestures with Mexican people is presented. The facial gesture analysis provides insight to present a method to extract invariances in gestures. These invariances are used for gesture classification. First, the visual sensor is used to capture visual information of human facial gestures, which are modeled as curves generated from connected feature points. Next, dynamic time warping is used to normalize the curves in order to be compared. Then, a Gaussian Mixture procedure is used to extract and model invariances among the database of facial gestures. This model is used to classify new facial information into a primary gesture. The main contributions of this work are the creation of a new facial gesture database for Mexico and the presentation of a procedure to extract invariants curves from facial gesture information.

**Facial Gestures, Computer Visión, Invariance Curves, Dynamic Time Warping, Gaussian Mixtures**

**Citación:** MONTECILLO-PUENTE, Francisco, SAMANO-ABONCE, Obed, LOPEZ-ENRIQUEZ, Renato & MEDINA-REZA, Alejandro, Yedid. Análisis e Invariancias en Gestos Faciales Primarios. Revista de Tecnología e Innovación 2015, 2-3:476-486

† Investigador contribuyendo como primer autor.

## Introducción

El análisis de gestos faciales es uno de los tópicos más importantes en la actualidad debido a su amplio espectro de aplicaciones. Entre estas se encuentran el desarrollo de nuevas interfaces humano computadora, la generación de gestos faciales sintetizados por computadora, la robótica de servicio, así como la asistencia en cirugías medicas.

En la literatura se han reportado algunos resultados sobre el análisis de gestos faciales. Ekman and Friessen (1977) del área de psicología realizaron un amplio estudio sobre gestos faciales y establecieron un sistema de codificación de acciones faciales, FACS, para catalogar expresiones faciales. Este sistema se basa en cambios específicos en los músculos de la cara que son llamados Unidades de Acción (AU). Ekman and Friessen definieron en su estudio seis emociones básicas: felicidad, tristeza, enojo, sorpresa, disgusto y temor. Sin embargo, Parrot (2000) estableció que los seres humanos son capaces de expresar 136 estados emocionales. Estos estados emocionales son clasificados como emociones primarias, emociones secundarias y emociones terciarias. Cada emoción primaria se puede clasificar dentro de una emoción secundaria, y a su vez en una terciaria. Las emociones propuestas por Ekman corresponden a las emociones primarias definidas por Parrot. En ambos trabajos, los estados emocionales fueron extraídos mediante observación y su clasificación fue realizada por personas con un grado elevado de expertis en la materia. En la comunidad que estudia gestos faciales, el sistema FACS y los 136 estados emocionales se han establecido como un estándar.

Por otra parte, las áreas de visión, robótica, animación por computadora y la industria del entretenimiento han realizado esfuerzos orientados hacia la detección de rostros, clasificación de gestos faciales, desarrollo de nuevas interfaces humano robot y generación automática de expresiones faciales en actores digitales. Además, se han creado diferentes bases de datos de movimiento humano.

El sistema de reconocimiento de expresiones faciales espontaneas presentado en (Pfister et al., 2011) utiliza un modelo de interpolación temporal para la detección de expresiones, además crea una base de datos de expresiones a 100 cuadros por segundo. Se utiliza hardware especializado para capturar expresiones que tienen duraciones muy cortas, ya que con cámaras de video tradicionales no sería posible detectarlas. Los autores se enfocan al reconocimiento de expresiones primarias.

Actualmente, los robots son una realidad en la industria. Sin embargo, existen muchos esfuerzos para acercar la robótica a la sociedad, llamada robótica de servicio. Dentro de las áreas a explorar están: la creación expresiones faciales en los robots humanoides. En este caso, la dificultad radica en transformar movimientos faciales previamente capturados en señales eléctricas que alimentan los motores de la cabeza del robot, (Wu et al., 2009). En el trabajo desarrollado por Dalibard, Magnenat-Thalmann & Thalmann (2012) se realiza un estudio sobre los avances en la creación movimientos autónomos, emocionales y con apariencia humana en actores virtuales y robots de servicio.

En animación por computadora para crear expresiones faciales en actores digitales se utilizan diferentes métodos. La utilización de escáneres tridimensionales sobre el rostro en el que se marcan los puntos de interés han sido utilizados, (Bickel et al., 2007). Estos métodos se utilizan para crear geometrías de alta resolución, sin embargo el tiempo y la carga computacional requeridos son muy elevados. Li, Weise, & Pauly (2010) desarrollaron un método para transferir poses faciales a diferentes caracteres a través de un método de entrenamiento. Por otro lado, facciones características de los Coreanos fueron transferidas exitosamente a modelos tridimensionales del rostro, (Lee & Hur, 2007). Las etapas que se generalmente se siguen para la transferencia de gestos faciales a actores digitales son: la captura de datos, la transformación a una morfología específica y su transformación a una geometría final, (Orvalho, Parke & Alvarez, 2012). En trabajos más recientes se enfocan a transferir emociones a actores digitales mientras hablan, (Jia et al., 2014).

En el trabajo desarrollado por Mavadati et al. (2013) se hace énfasis en que para desarrollar sistemas de reconocimiento facial automático es importante contar con bases de datos confiables. Las bases de datos disponibles las categorizan principalmente en expresiones espontáneas y expresiones fingidas. Yana et al. (2014) realizan un análisis sobre los aspectos que se deben considerar para crear este tipo de bases de datos. En general, las bases de datos consisten de segmentos de videos etiquetados con un gesto. También, presentan la problemática de propiciar en los participantes las emociones de forma natural, así como las dificultades en su clasificación. Con la disponibilidad de sensores de profundidad se han creado bases de datos de gestos tridimensionales, (Zhanga et al., 2014).

Finalmente, en el área de investigación de medios es de interés estudiar las preferencias de los tele-espectadores. Por lo cual, las empresas crean paneles de personas para observar las expresiones de los mismos y determinar sus preferencias, Kodra et al. (2013). En medicina, los gestos faciales se han utilizado para determinar dolor en pacientes (Sikkaa, Dhallb & Stewart, 2014).

En el presente trabajo se reporta la creación de una base de datos siguiendo las recomendaciones mencionadas en la literatura, y el análisis de los datos a fin de extraer patrones característicos para los gestos primarios. Como objetivos se pretende extender la base de datos a todo México, a fin de incluir los distintos rasgos faciales representativos de nuestra sociedad ya que las bases que existen son para Asia, Europa o Norte América.

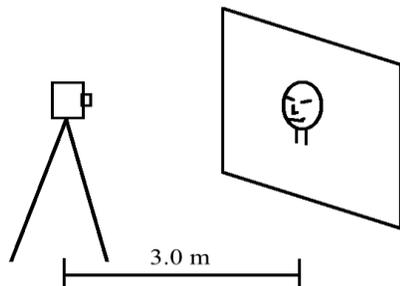
La problemática principal abordada en este artículo es determinar si en la población en México las expresiones primarias están presentes y si éstas presentan alguna invariancia. En la primera sección del artículo se presenta el protocolo seguido para crear la base de datos y un procesamiento preliminar de los datos. Posteriormente, en la segunda Sección se presenta un método para el análisis de la información de gestos faciales con la intención de verificar la existencia de curvas invariantes. Luego, en la Sección 3 se presentan un método para clasificar la información facial en un gesto primario. Los resultados obtenidos y su discusión se presentan en la Sección 4. Finalmente, las conclusiones del trabajo se presentan en la última Sección.

### Datos Visuales de Gestos Faciales.

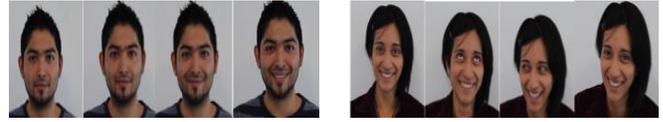
En esta sección se describe el protocolo a través del cual se capturaron los gestos faciales. Para este fin, se utilizó un cámara de video, a varios sujetos se les pidió que participaran para que fueran tomados en video realizando ciertas expresiones faciales. Luego, los videos fueron clasificados y algunos puntos de interés seleccionados de forma manual.

### Información Visual de Gestos Faciales.

Para capturar la información visual de gestos faciales se preparó un espacio confortable para los participantes y con iluminación adecuada. Dicho lugar contaba con una cámara de video Canon HD con una capacidad de 30 fps y una resolución de 1280x720. Cada participante fue grabado durante el tiempo requerido para realizar los gestos solicitados, siempre a una misma distancia y tratando de que la cabeza del participante permaneciera en una misma orientación y ubicación. Luego, a cada participante se le dieron indicaciones de realizar cuatro gestos faciales primarios: feliz, enojado, triste y asustado. Para esto se solicitó que pensarán en algo relacionado con la expresión facial y que trataran de reflejarlo en su rostro. En la Figura 1, se ilustra cómo fueron capturadas los gestos, también se muestra una secuencia exitosa y una secuencia fallida.



(a)



(b)

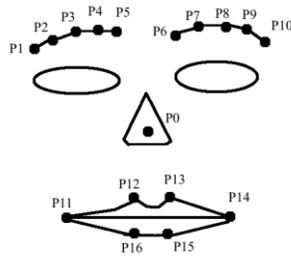
(c)

**Figura 1** Captura de expresiones faciales: a) Espacio de captura de gestos faciales, b) secuencia de captura correcta, c) secuencia de captura fallida, la persona gira su cabeza.

Para crear nuestra base de datos participaron 30 personas quienes recibieron las indicaciones antes mencionadas. Luego, estos datos fueron mostrados a 7 personas para clasificarlas y etiquetarlos. Las secuencias donde el movimiento de la cabeza era grande fueron descartadas.

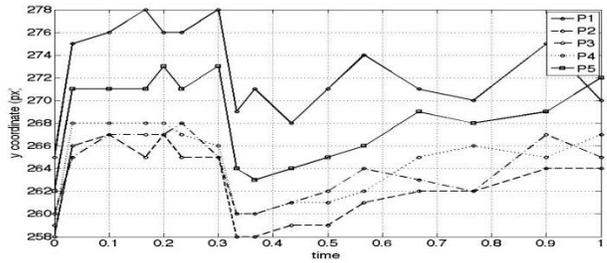
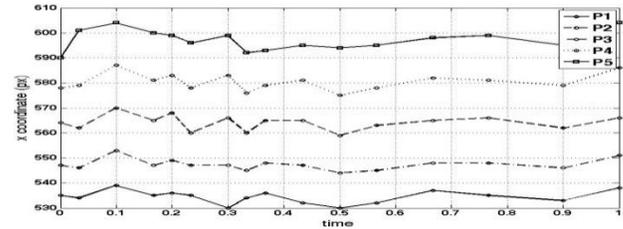
### Puntos de Interés de Gestos Faciales.

De la revisión bibliográfica que se consultó, determinar con información visual las Unidades de Acción definidas por Ekman (1977) es una tarea muy difícil, por lo que en su lugar se utilizan puntos de interés. Además, para detectar las emociones primarias en diferentes trabajos los puntos de interés que tienen en común están localizados en las cejas, la boca y la nariz. En este trabajo, se eligieron 17 puntos para realizar el análisis. Estos puntos consisten en un punto de referencia localizado en la nariz, 5 puntos para la ceja izquierda, 5 puntos para la ceja derecha y 6 puntos para los labios, estos puntos se muestran en la Figura 2.



**Figura 2** En esta imagen se muestra la enumeración de los puntos que se utilizan para el análisis de gestos faciales.

La localización de estos puntos fue extraída manualmente para cada segmento de video y por 5 participantes diferentes, esto con la finalidad de crear redundancia e incluir el factor humano en su apreciación. Dado que un gesto dura aproximadamente 2 segundos, la cantidad de puntos a seleccionar manualmente son  $17 \times 2 \times 30 = 1020$  por gesto, por lo que se seleccionaron los cuadros donde se aprecia un cambio importante, luego los demás puntos se obtendrán por interpolación basada en curvas splines. La cantidad de cuadros por gesto tiene un rango  $m = [9,15]$ , por lo que cada gesto consiste de un vector  $m \times 34$  ( $m \square m$ ). En la Gráfica 1, se muestra la curva típica de una secuencia para el gesto correspondiente a la emoción de enojo de un participante sin pre procesamiento.



**Gráfico 1** Patrón típico del gesto enojado para la caja. Se puede observar que algunos de los puntos su desplazamiento es muy pequeño.

**Análisis de Información Visual de Gestos Faciales**

La información capturada de los gestos faciales de cada participante no son comparables directamente entre sí debido a que tienen diferente duración y selección de cuadros. En esta sección se presenta la metodología empleada para normalizar los datos. Primero, se utiliza un interpolación spline y luego se aplica una deformación dinámica temporal.

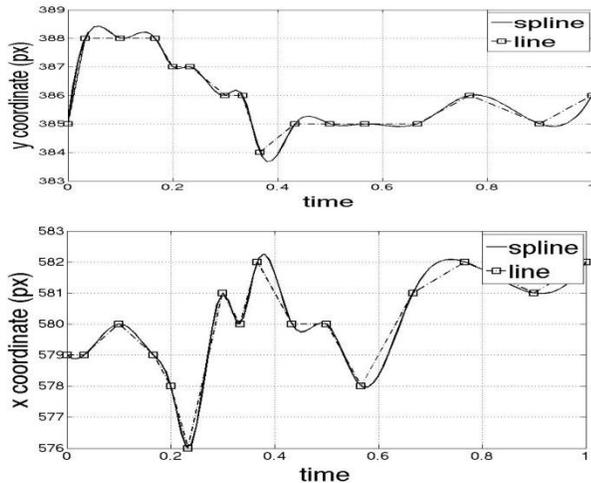
**Interpolación basada en Curvas bsplines**

Debido a que para cada participante los gestos fueron extraídos manualmente por varias personas, sus curvas pueden no coincidir. Para hacer comparables las curvas se realiza un interpolación, en particular se utilizan curvas bsplines. La elección de estas curvas es porque tienen las características de producir curvas suaves y que tienen una cantidad fija de puntos que influyen localmente.

La interpolación spline se realiza utilizando la siguiente ecuación,

$$S(x) = \sum_{i=-\infty}^{\infty} A_i B_{i-k}^k(x) \quad (1)$$

donde  $B_m^k$  es la función base  $m$  de orden  $k$ ,  $A_i$  son los coeficientes de interpolación y  $S(x)$  es la función interpolada. Los coeficientes de interpolación  $A_i$  se determinan a partir de las condiciones sobre las funciones base y los puntos de control, (Prautzsch, Boehm, & Paluszny 2002). En la Gráfica 2, se muestran la interpolación bspline del punto P<sub>11</sub> para el gesto feliz.



**Gráfico 2** Forma de la curva interpolada del punto P<sub>11</sub>, punto localizado en los labios. En la gráfica se muestran los puntos de control marcador con un cuadrado.

### Deformación Dinámica Temporal.

Cuando se tienen señales que se desean comparar y estas se encuentran en el mismo intervalo temporal, una correlación puede ser suficiente. Sin embargo, hay situaciones donde el rango temporal no es el mismo, por tanto otras técnicas deben emplearse. Por ejemplo, en el reconocimiento de voz (Rabiner & Juang, 1993), en la escritura (Huang et al., 2010) y transferencia de movimiento humano, (Zhou, De la Torre, 2012). La deformación dinámica temporal consiste en lo siguiente: dadas dos señales temporales, una señal  $S_1$  de longitud  $n$  y una señal  $S_2$  de longitud  $m$ . La alineación de ambas se construye a partir de una matriz de  $n \times m$ , donde el elemento  $(i,j)$  de la matriz corresponde a la distancia cuadrada euclidiana entre los puntos  $S_1^i$  y  $S_2^j$ . La mejor correspondencia entre las dos señales, consiste en el camino en la matriz que minimiza la distancia total acumulada. En particular, el camino óptimo es el camino que minimiza el costo de deformación siguiente,

$$DWT(S_1, S_2) = \min\left\{ \sqrt{\sum_{k=1}^K W_k} \right\} \quad (2)$$

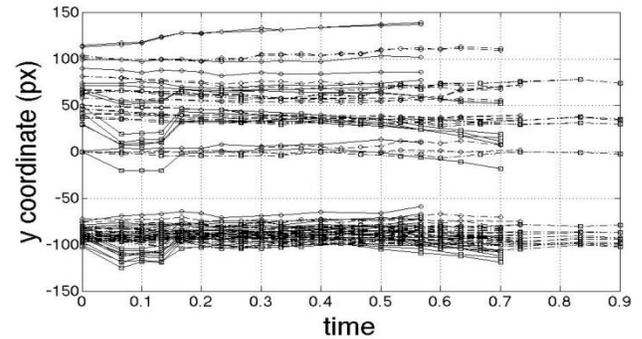
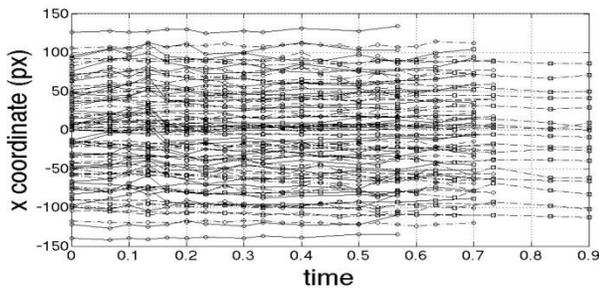
donde  $W_k$  es el elemento de la matriz de distancias  $(i,j)_k$ . Un conjunto contiguo de elementos de la matriz representa el mapeo entre  $S_1$  y  $S_2$ .

### Detección de Invariencias en los Gestos Faciales.

Para determinar la existencia de invariencias en Gestos faciales con la misma etiqueta, se procede a analizar experimentalmente la información recopilada utilizando la mitad de los participantes. Cabe recordar que cada participante fue analizado por 7 personas. Por lo que la cantidad total de información para caracterizar un gesto son  $30/2 \times 7 \times 17 \times 2 = 1575$ .

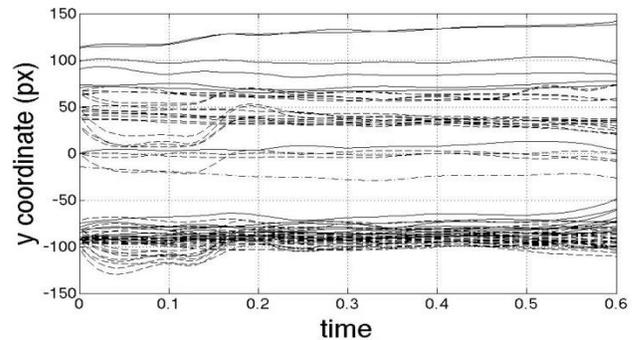
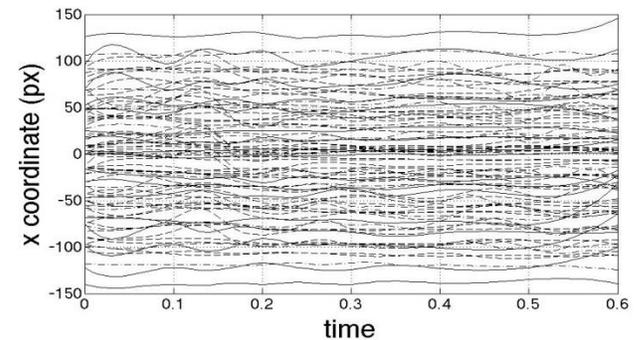
El análisis realizado consiste en interpolar cada punto  $P_i$  para cada gesto facial con la misma etiqueta. Luego, cada curva interpolada se ajusta temporalmente utilizando la deformación dinámica temporal. La deformación dinámica temporal se aplica utilizando como señal de referencia  $S_1$  la señal más larga, y la segunda señal  $S_2$  es cada una de las otras señales restantes. Posteriormente, las curvas de deformación temporal obtenidas son analizadas, si existe una curva de deformación cuya distancia  $\max \{D_{min}\}$  es pequeña, entonces se ha determinado una curva invariante. En las Gráficas 3, 4, 5 se muestran el procedimiento desarrollado para el gesto facial feliz.

En la Grafica 3, se presentan las curvas de los 17 puntos de la expresión facial feliz de algunos participantes. Se puede observar algunas curvas tienen diferente duración, es decir, los participantes realizan el gesto con diferente duración. Además, se puede observar que se sigue un patrón en sus curvas, regiones donde se aprecia más densidad. Este patrón es la evidencia de que en los gestos faciales existen invariantes.



**Gráfico 3** Conjunto de puntos para el gesto feliz, se muestran las coordenadas x y de cada punto.

Las gráficas interpoladas y deformadas en el tiempo son mostradas en la Gráfica 4, se puede observar suavidad en estas.



**Gráfico 4** Conjunto de puntos para el gesto feliz, con interpolación bspline y a justadas en tiempo.

## Clasificación de Gestos Visuales utilizando Mezclas de Gaussianas .

El procedimiento presentado en la Sección anterior tiene como objetivo mostrar la existencia o no de invariancias de forma experimental. Estas invariancias sirven de base, para implementar un clasificador supervisado basado en las mezclas de Gaussianas, (Gelman et al., 2013).

### Mezcla de Gaussianas.

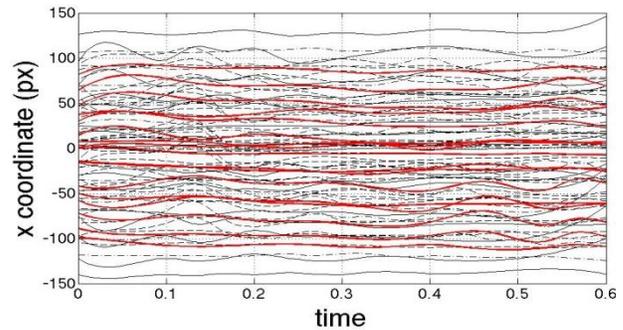
La distribución de una variable aleatoria  $y \in R^d$  puede modelarse utilizando una mezcla de Gaussianas como,

$$p(y|\theta) = \sum_{k=1}^K w_k \cdot N(y; u_k, \Sigma_k) \quad (3)$$

donde  $\theta = \{w_1, u_1, \Sigma_1, \dots\}$ ;  $w_k, u_k, \Sigma_k$  son el peso, media y matriz de covarianza para la  $k$ -ésima componente Gaussiana unimodal  $N(\bullet)$ . Dada una cantidad de  $n$  puntos  $\{y_1, \dots, y_n\}$  en el espacio  $d$ -dimensional,  $R^d$ , el algoritmo EM estima en términos probabilísticos  $w_k, u_k, \Sigma_k, k=1 \dots K$ .

### Modelado de Invariancias.

Dadas una cantidad  $n$  de curvas, donde cada curva es representada por un punto en el espacio  $d$ -dimensional, el algoritmo EM puede ser utilizado para modelar este conjunto de curvas como mezcla de Gaussianas. Por tanto, todas las curvas que representan gestos primarios pueden modelarse mediante la Mezcla de Gaussianas. Usando  $K$  componentes Gaussianas cada componente modela un gesto primario. Por tanto, las invariancias de los gestos son representadas por un componente unimodal Gaussiano. En la Gráfica 5, se muestra la curva invariante para el gesto Feliz, para una coordenada.



**Gráfico 5** Gráfica de invariancia para el gesto feliz, la curva invariante se muestra en rojo.

### Clasificación de Gestos

Dado que la mezcla de Gaussianas es una función de probabilidad multimodal, cuando un nuevo dato se evalúa se puede determinar el componente al que pertenece este nuevo dato. Análogamente, si las invariancias de los Gestos Faciales han sido modeladas con el uso de mezclas de Gaussianas, cuando se presente un nuevo gesto, este puede ser evaluado para determinar el gesto primario al que pertenece. Esto es, el nuevo gesto pertenece al componente con mas probabilidad.

Para evaluar un nuevo gesto las curvas de los puntos de interés son primero interpolados, luego deformados dinámicamente respecto a la curva de deformación de cada gesto. Finalmente, el gesto representado por un punto  $N$ -dimensional es clasificado utilizando las probabilidades de la mezcla de Gaussianas.

### Resultados y Discusión

Como resultados se ha creado una base de datos formada por secuencias de video de gestos faciales etiquetados como feliz, enojado, asombrado y triste. Además, para los gestos de cada participante se tiene la información de los 17 puntos de interés seleccionados para este trabajo.

El etiquetado y la selección de puntos fue realizado por otros 7 participantes de forma manual. En resumen, se cuenta con 120 secuencias de video con gestos faciales etiquetados y sus 17 puntos de interés marcados por 7 personas diferentes. *“Esta base de datos puede ser solicitada enviando un e-mail a alguno de los autores”*. Dentro de las dificultades encontradas para la creación de esta base de datos, fue que no todos los participantes pueden expresar los estados emocionales fácilmente. En general, los gestos de felicidad y enojo fueron realizados sin dificultad. De acuerdo a lo que se pudo notar, la razón por la cual algunos de los participantes no expresaban las emociones es porque se sentían un poco intimidados o nerviosos por que iban a ser grabados. Sin embargo, otros métodos para propiciar emociones espontaneas deben ser estudiados. Algunos de los videos que se capturaron fueron desechados debido a que no cumplían con los requerimientos que se establecieron.

Respecto a la detección de invariancias y clasificación de gestos, en los gestos faciales primarios Feliz, Enojado y Sorprendido se observo que sus curvas en general son diferentes. Sin embargo, la emoción de tristeza es difícil de detectar con los puntos de interés utilizados. De la observación de los videos, los participantes expresan tristeza sea inclinando la cabeza, cambiando su mirada o mediante otras partes del cuerpo como los hombros. Recordando que se utilizo solo una parte de los datos para entrenamiento, la otra parte se utilizo para prueba. Los resultados que se obtuvieron para el reconocimiento exitoso de gestos fue de 55%, 60%, 62% y 40% para los gestos feliz, enojado, sorprendido y triste respectivamente. Estos resultados son comparables con los reportados en la literatura, (Pfister, Li, Zhao, & Pietikainen 2011).

Para la comparación con otros trabajos, es difícil establecer un marco comparativo ya que se utilizan diferentes puntos de interés, diferentes información visual, además de diferentes infraestructuras de hardware. Este un área de oportunidad a explorar en un futuro. De aquí la importancia de contar con una base de datos común para que sea el medio de comparación entre los diferentes métodos.

### Conclusiones

En este trabajo se presentó una nueva base de datos para el análisis de gestos faciales. Sin embargo, esta base de datos debe ser actualizada y mejorada. Los puntos a mejorar son contar con mayor cantidad de participantes, que incluya diferentes características de hardware, los movimientos de la cabeza, mirada y cuerpo sean considerados. Además, se deben incluir nuevas métodos para propiciar gestos faciales genuinos y espontáneos. Con la disponibilidad del sensor visual de profundidad y cámaras termo graficas, es posible incluir otro tipo de información para crear sistemas confiables autónomos de detección en línea, sobre un amplio rango de emociones. Ya que hasta el momento en la literatura existente solo se ha logrado detectar y clasificar automáticamente las emociones primarias y algunas secundarias. Por lo que nuevos algoritmos para la detección automática de emociones secundarias y terciarias es un área de oportunidad a explorar. Adicionalmente, se debe considerar el desarrollo de nuevas aplicaciones que incluyan el uso de sistemas autónomos de detección de emociones, tales como interfaces hombre-maquina para la robótica de servicio o interfaces que faciliten el aprendizaje en línea.

## Referencias

- Bickel, B., Botsch, M., Angst, R., Matusik, W., Otaduy, M., Pfister, H. & Gross M. (2007). *Multi-Scale Capture of Facial Geometry and Motion*. ACM Transactions on Graphics (Proc. SIGGRAPH 2007), 26.
- Dalibard, S., Magnenat-Thalmann, N., & Thalmann, D. (2012), Workshop on Autonomous Social Robots and Virtual Humans at the 25th Annual Conference on Computer Animation and Social Agents.
- Ekman, P., & Rosenberg, E.L. (1997). *What the face reveals: basic and applied studies of spontaneous expression using the facial action coding system (FACS)*. Illustrated Edition, Oxford University Press.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian Data Analysis*. Third Edition Chapman & Hall/CRC Texts in Statistical Science.
- Huang, G., Zhang, D., Zheng, X. & Zhu, X. (2010). *An EMG-based handwriting recognition through dynamic time warping*. Annual International Conference of the IEEE, Engineering in Medicine and Biology Society.
- Jia, J., Wu, Z., Zhang, S., Meng, H. M. & Cai, L. (2014). Head and facial gestures synthesis using PAD model for an expressive talking avatar. *Multimedia Tools and Applications*. 73(1), 439-461
- Kodra, E., Senechal, T., McDuff, D. & El Kaliouby, R. (2013). *From Dials to Facial Coding: Automated Detection of Spontaneous Facial Expressions for Media Research*. 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG).
- Lee, H.-C. & Hur, G.-T. (2007). *3D Face Deformation Using Control Points and Vector Muscles*. *International Journal of Computer Science and Network Security*, 7(4).
- Li, H., Weise, T., & Pauly M. (2010). *Example-Based Facial Rigging*. ACM Transactions on Graphics, Proceedings of the 37th ACM SIGGRAPH Conference and Exhibition.
- Mavadati, S.M., Mahoor, M.H., Bartlett, K., Trinh, P., Cohn, J.F. (2013). *DISFA: A Spontaneous Facial Action Intensity Database.*, *IEEE Transactions on Affective Computing*, 4 (2), 151 – 160.
- Orvalho, V., Parke, F. & Alvarez X. (2012). *A Facial Rigging Survey*. *Proc. Eurographics*, 32, 10 – 32.
- Parrott, W.G. (2000). *Emotions in Social Psychology*. Philadelphia: Psychology Press.
- Pfister, T., Li, X., Zhao, G. & Pietikainen, M. (2011), *Recognising spontaneous facial micro-expressions*. *IEEE International Conference on Computer Vision (ICCV)*. 1449 – 1456.
- Prautzsch, H., Boehm, W. & Paluszny, M. (2002), *Bézier and B-Spline Techniques*. Mathematics and Visualization. Berlin Heidelberg: Springer-Verlag.
- Rabiner, L. & Juang, B. (1993). *Fundamentals of speech recognition*. Prentice Hall.
- Wu, T., Butko, N. J., Ruvulo, P., Bartlett, M. S., & Movellan J. R. (2007). *Learning to Make Facial Expressions*. 2009 Ieee 8th International Conference On Development And Learning.
- Sikkaa, K., Dhallb, A. & Stewart M. (2014). *Classification and weakly supervised pain localization using multiple segment representation*. *Image and Vision Computing*. Volume 32(10), 659–670.

Yana, W.-J., Wanga, S.-J., Liuc, Y.-J., Wud, Q. & Fu, X. (2014). *For micro-expression recognition: Database and suggestions*. Neurocomputing, 136, 82–87.

Zhanga, X, Yina, L., Cohnb, J. F., Canavana, S., Realea, M., Horowitza, A., Liua, P., Girardb, J. M. (2014). *BP4D-Spontaneous: a high-resolution spontaneous 3D dynamic facial expression database*. Image and Vision Computing. 32(10), 692–706.

Zhou, F. & De la Torre, F. (2012). *Generalized time warping for multi-modal alignment of human motion*. IEEE Conference on Computer Vision and Pattern Recognition.