

# Clasificación de Expresiones Faciales Utilizando Métodos de Aprendizaje Supervisado

Centro de Investigación en Matemáticas. Unidad Monterrey

Marcelo Alberto Sanchez Zaragoza

marcelo.sanchez@cimat.mx

**Resumen**—El proyecto toma una serie de imágenes de distintas mujeres japonesas para intentar identificar la expresión facial que lleva cada una de ellas. El proceso que se lleva a cabo es tomar cada imagen y aplicar un filtro para después obtener en un vector las componentes de esta imagen (píxeles), después buscar una representación adecuada de los datos tal que se encuentren patrones interesantes y finalmente obtener una porción de datos clasificados correctamente basados en métodos de aprendizaje supervisado.

## I. INTRODUCCIÓN

Las expresiones faciales son de suma importancia en las interacciones humanas, una tarea importante es encontrar estas características que las personas identifican para decidir que expresión está mostrando una persona. El procesamiento automático de información facial es un campo importante y muy activo en reconocimiento de patrones.

Existen diferentes tipos de procesamiento de imágenes tal que nos ayudan a obtener ciertas características y dadas estas características encontrar diferencias entre una imagen y otra. Algunos preprocesamientos de imágenes son de mucha utilidad ya que facilitan la aplicación de ciertos modelos, por ejemplo, encontrar rostros en una imagen, encontrar un rostro en específico, encontrar cierto patrón en una serie de imágenes o ciertas características de una expresión facial.

Entre estos procesamientos tenemos el *filtro de Gabor*. Los *filtros de Gabor* proporcionan una nueva imagen que nos ayudan a encontrar ciertas características resaltándolas o disminuyendo algunas. Basado en este procesamiento el objetivo de este proyecto es analizar y clasificar las *expresiones faciales* de un conjunto de imágenes de mujeres japonesas utilizando distintos métodos de aprendizaje supervisado, aunado a este objetivo, también, se pretende identificar que método de clasificación, según las características de las imágenes procesadas y su representación, es un mejor clasificador y por qué.

En la primera sección se describen brevemente la función de la transformada de Gabor y los filtros digitales. En la segunda sección se reportan los resultados de ocupar el filtro de Gabor y las representaciones que se consideraron y por último, se presentan las conclusiones, por último agregar que las implementaciones que se realizaron fueron desarrolladas en el lenguaje de programación *Python*, no se mencionará en el desarrollo del reporte pero se anexa los respectivos códigos junto con el reporte.

## II. FUNCIÓN DE LA TRANSFORMADA DE GABOR

El *filtro de Gabor* es un filtro lineal cuya respuesta de impulso es una función sinusoidal multiplicada por una función Gaussiana. Una ventaja notable que se obtiene al introducir la envolvente Gaussiana es que las funciones de Gabor están localizadas tanto en el dominio espacial como en el de la frecuencia, a diferencia de lo que ocurre con las funciones sinusoidales, que están perfectamente localizadas en el dominio frecuencial y completamente deslocalizadas en el espacial. Por lo que son funciones más adecuadas para representar una señal conjuntamente en ambos dominios.

La función de la transformada de Gabor es la siguiente:

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left(-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right)\right) \exp(2\pi j u_0 x) \quad (1)$$

Donde  $\sigma_x$  y  $\sigma_y$  denotan la envolvente Gaussiana a lo largo de los ejes  $x$  y  $y$ ,  $u_0$  define la frecuencia radial de la función de Gabor.

Se puede obtener una familia de funciones Gabor mediante la dilatación y rotación apropiadas de  $f(x, y)$  a través de la función de generación.

$$f_{pq}(x, y) = \alpha^{-p} f(x', y') \quad (2)$$

Donde:

$$\begin{aligned} x' &= \alpha^{-p}(x \cos \theta_q + y \sin \theta_q) \\ &= \alpha^{-p}(-x \sin \theta_q + y \cos \theta_q) \end{aligned}$$

$$\alpha > 1; \quad p = 1, 2, \dots, S; \quad q = 1, 2, \dots, L.$$

Los subíndices enteros  $p$  y  $q$  representan el índice de escala (dilatación) y orientación (rotación), respectivamente.  $S$  es el número total de escalas y  $L$  es el número total de orientaciones en el banco de filtros de Gabor. Para cada orientación  $q$ , el ángulo  $\theta_q$  es dado por:

$$\theta_q = \frac{\pi(q-1)}{L}; \quad q = 1, 2, \dots, L. \quad (3)$$

Para más referencias de la función de Gabor se puede consultar en [4].

### II-A. Filtros de Imágenes

Un filtro de imagen es un procedimiento que se aplica a una imagen para resaltar o mejorar algunas características de la misma.

Los filtros digitales constituyen uno de los principales modos

de operar en el procesamiento de imágenes digitales. Pueden usarse para distintos fines, pero en todos los casos, el resultado sobre cada pixel depende de los pixeles de su entorno. Entre los distintos objetivos encontramos:

- Suavizar la imagen: reducir las variaciones de intensidad entre píxeles vecinos
- Eliminar ruido: modificar aquellos píxeles cuyo nivel de intensidad es muy diferente al de sus vecinos.
- Realzar la imagen: aumentar las variaciones de intensidad.
- Detectar bordes: detectar aquellos píxeles donde se produce un cambio brusco en la función intensidad.

Los filtros se pueden clasificar en: Filtros en el dominio del espacio y Filtros en el dominio de la frecuencia.

En la figura 1 se ilustra como funciona un filtro en el dominio del espacio.

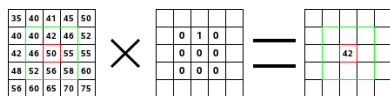


Figura 1. Ejemplo de filtro en el dominio del espacio

En la figura 1 se puede observar que al ocupar dicho filtro vamos a obtener una nueva representación con la misma cantidad de píxeles pero con el detalle que ahora cada pixel tiene la contribución de los que estan a su alrededor. Estos filtros son de mucha ayuda ya que mejoran mucho ciertas imagenes.

El *filtro de Gabor* es un filtro pasabanda en 2D, si le asignamos una determinada escala y orientación obtenemos una reducción del ruido a la vez que preserva una dirección de la imagen original. En la figura 2 se muestra un ejemplo de las familias de Gabor y en la figura 3 se ilustra en un plano de tres dimensiones.

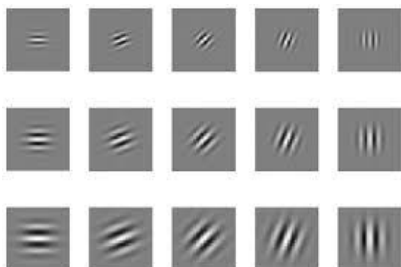


Figura 2. Familia de filtros de Gabor

En la figura 2 se muestra como se puede ir moviendo este filtro de forma que los resultados de aplicar dicho filtro sean diversos. En la figura 3 se muestra como van cambiando estas funciones y proporcionan una mejor idea de como son los filtros de Gabor.

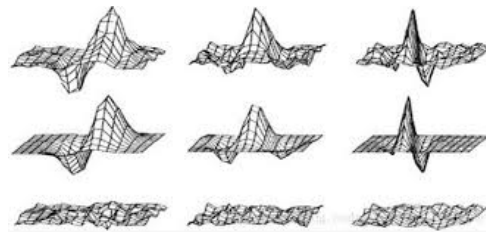


Figura 3. Filtros de Gabor en tres dimensiones

Los *filtro de Gabor* se ocuparon en este proyecto para tratar de encontrar ciertas características en las distintas imágenes y poder definir un patrón para cada expresión facial.

En la figura 4 se muestra otro ejemplo de un filtro de Gabor. Para obtener una imagen a la cual se le aplicó el filtro de Gabor se realiza un procedimiento llamado convolución.



Figura 4. Ejemplo de filtro de Gabor

Al realizar dicha convolución se tiene como posibles resultados imagenes como en la figura 5.

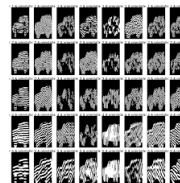


Figura 5. Ejemplo de imagen con filtro de Gabor

En la figura 5 se observa que muchas imagenes no regresan algo interesante pero en otras notamos que guarda ciertas características que nos pueden ayudar a identificar el objeto en que queremos encontrar, por si fuera el caso.

### III. CLASIFICACIÓN DE IMAGENES

El conjunto de imagenes con las que se trabajo son de la base de datos JAFFE. Se cuenta con 10 mujeres japonesas, donde cada una de ellas mostro las siguientes expresiones: Enojo, Descepción, Felicidad, Neutral, Triste y Sorprendido. Se tiene un total de 180 imagenes con un tamaño de 256x256 píxeles cada una. La liga y referencia para encontrar dichas imagenes se anexan al final del reporte.

En la siguiente tabla se muestra la cantidad de imagenes que se tiene por cada expresión facial:

Expresión Facial	Frecuencia
Enojado	30
Decepción	29
Feliz	20
Neutral	30
Triste	31
Sorprendido	30

Y en la siguiente tabla se muestra la cantidad que tiene cada una de las modelos que proporcionaron las expresiones:

Iniciales del Nombre	Número de imágenes
KA	19
KL	19
KM	19
KR	17
MK	18
NA	18
NM	17
TM	18
UY	18
YM	17

En la figura 6 se muestra como son las imágenes de nuestra base de datos. En algunas imágenes notamos que hay muchos cambios entre una expresión y otra, estos cambios se observan desde las cejas, ojos y boca. Estos rasgos son los que se desean encontrar para definir cierto patrón en una expresión facial.

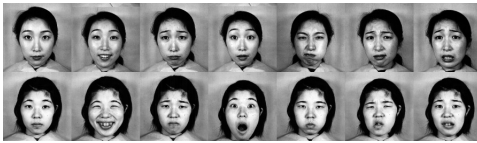


Figura 6. Ejemplo de imágenes de las expresiones faciales de la base de datos

A cada una de las imágenes se les aplico el *filtro de Gabor* para obtener cierta representación de la misma, un ejemplo de este procedimiento se muestra a continuación:



Figura 7. Ejemplo de imagen Original



Figura 8. Ejemplo de filtro de Gabor

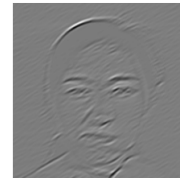


Figura 9. Ejemplo de imagen con filtro de Gabor

Una vez que se aplico el filtro de Gabor a cada una de las imágenes se busco obtener un vector con todos los píxeles que contiene la imagen, el procedimiento nos llevo a tomar 65,536 píxeles para cada imagen. Como resultado de obtener un vector por cada una de las imágenes se tuvo como resultado una matriz de 180 renglones que corresponden a cada imagen por 65,536 columnas que corresponden a cada uno de los píxeles.

En la figura 10 se muestra el resultado de aplicar el filtro en el cual nos vamos a basar en este proyecto. Observe que se logra rescatar características importantes que pueden ayudar a los modelos a identificar mejor cada expresión.



Figura 10. Imagen con Filtro de Gabor

La matriz que se obtuvo era muy grande por lo que se procedio a obtener una representación más pequeña de los datos ya que muchos de los píxeles no aportan mucha información. Se puede observar en la figura 10 que los píxeles que representan el fondo, el cabello y la ropa que llevan puesta no proporciona información que sea de ayuda para la clasificación de emociones.

Las representaciones que se ocuparon fueron las siguientes: PCA, Kernel-PCA, Spectral Embedding y T-SNE. En el caso de PCA y Kernel PCA se busco el número de componentes principales necesarios tal que conservaran cerca del 80 % de la varianza acumulada.

Entre las representaciones que proporcionaron resultados positivos para la clasificación de expresiones faciales, fueron PCA y Kernel-PCA. Las demás representaciones se consideraron pero al analizarlos con distintos métodos de aprendizaje supervisado no encontramos resultados favorables por lo que se obtuvo por no agregarlo al siguiente reporte pero si mencionarlos.

### III-A. Representación PCA

La primer representación que arrojo resultados positivos fue PCA(Análisis de componentes principales). PCA es una técnica utilizada para describir un conjunto de datos en términos de nuevas variables. Dichas nuevas variables se analizaron para observar que proporción varianza acumulada de los datos preservan. Al aplicar el método nos resulto un total de 180 valores propios y 180 representaciones para cada una de las imagenes. En la figura 11 se muestra en el eje x el número de variables y en el eje y la propoción de varianza acumulada

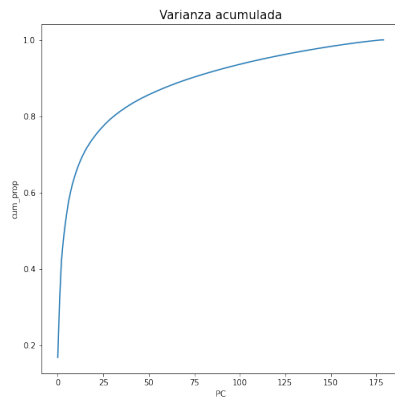


Figura 11. Proporción de Varianza Acumulada

En la figura mostramos una representación de las imágenes en los primeros componentes principales.

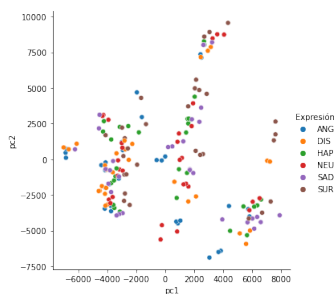


Figura 12. Representación de los datos con PCA-Expresión

Se observa que al hacer esta representación las imagenes con el mismo sentimiento se encuentran cerca una de la otra, es un resultando interesante. La figura 13 se agrego solo con el fin de mostrar como esta representación nos ayuda a identificar las imagenes de cada una de las mujeres con su respectivo grupo.

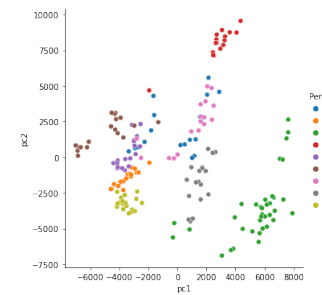


Figura 13. Representación de los datos con PCA-Personas

Se tomaron las primeras 100 representaciones ya que nos conservan cerca del 90 % de la varianza acumulada. La nueva matriz contaba con 180 renglones correspondientes a las imagenes y 100 columnas con las nuevas variables. Los métodos de aprendizaje supervisado que se ocuparon fueron los siguientes: Redes Neuronales, Maquinas de Soporte Vectorial, Arboles de Clasificación, AdaBoost, LDA y QDA. La nueva representación de los datos se dividio en dos grupos, un grupo de entrenamiento tomando el 80 % de los datos y un grupo de prueba con el 20 % restante. Mostraremos los métodos con mejores resultados así como algunos gráficos y solo se mencionara aquellos que no dieron un buen resultado.

**III-A1. Redes Neuronales:** En la figura 14 se muestra la matriz de confusión resultante de haber ocupado Redes Neuronales.

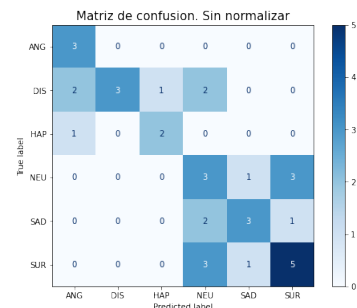


Figura 14. Matriz de Confusión-Redes Neuronales

Una matriz de confusión nos muestra el número de datos clasificados correctamente en la diagonal principal mientras que en valores fuera de la diagonal principal muestra que valores fueron clasificados erróneamente en grupos distintos a los que pertenecen.

Para la figura 14 hay que observar que el conjunto de prueba que se tomo no tiene el mismo número de imagenes en todas las expresiones por lo que a simple vista las emociones de Felicidad y Enojo pudieran parecer las peores no lo son, por ejemplo, la expresión de sorprendido de todo el conjunto de prueba(9 imagenes) tiene un porcentaje de 55.55 % de imagenes clasificadas correctamente mientras que la expresión de Enojo(5 imágenes) tiene el 60 % de imagenes clasificadas correctamente por lo que hay que prestar mucha atención en

estos detalles.

La proporción de datos asignados correctamente con el grupo de prueba fue del 53 %. Los parámetros que se ocuparon fueron los siguientes: una capa con 10 nueronas y un  $\alpha = 0.01$ .

**III-A2. Maquinas de Soporte Vectorial:** También se agrego la matriz de confusión para observar que diferencia hubo entre este método y el anterior. En la figura 15 se ilustra la matriz de confusión.

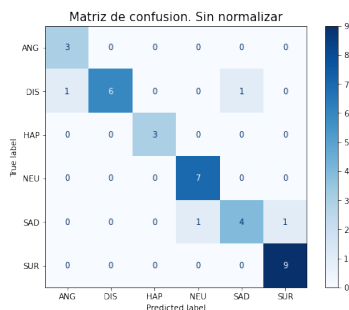


Figura 15. Matriz de Confusión-Maquinas de Soporte Vectorial

Para la matriz de confusión en la figura 15 observamos un mejor comportamiento ya que se redujo el número de imagenes clasificadas erroneamente.

El resultado de la proporción de datos asignados correctamente fue del 89 %. Este fue el mejor de los resultados después de varias en repetidas ocasiones los parámetros. Los parámetros que ocupamos en este método fueron los siguientes: kernel Gaussiano y un valor de  $C=9.5$ (parámetro regularizador).

**III-A3. LDA(Análisis de Discriminante lineal):** Finalmente agregamos la matriz de confusión del método LDA, en la figura 16 se observa el resultado.

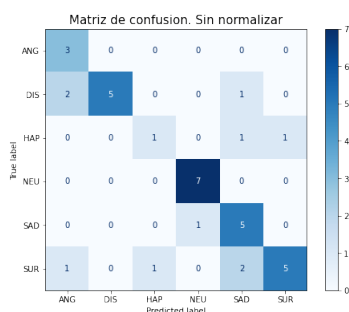


Figura 16. Matriz de Confusión-LDA

Obtuvimos una proporción de datos asignados correctamente del 72 % que no es muy lejano al resultado de Maquinas de Soporte Vectorial.

Para los demás métodos encontramos los siguientes resultados correspondiente a la proporción de datos asignados correctamente:

Método	Proporción
Arboles	44.4 %
AdaBoost	64 %
QDA	17 %

### III-B. Representación Kernel-PCA

Para la representación con kernel-PCA se tomaron en cuenta todas las componentes principales que nos arrojo el método, tenemos un total de 180 nuevas variables para las mismas imagenes. La figura 17 muestra la representación en los dos primeros componentes, observamos que varios imágenes correspondientes al mismo expresión se encuentran cerca mientras en otros casos se encuentran distantes.

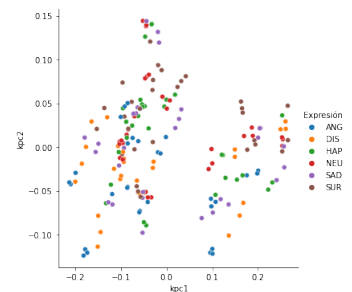


Figura 17. Representación de los datos con Kernel-PCA(Expresión)

Al igual que en PCA agregamos la figura 18 donde observamos una mejor representación de las personas de las mujeres japonesas, se observa que hay grupos completos con todas las imagenes con su respectiva persona.

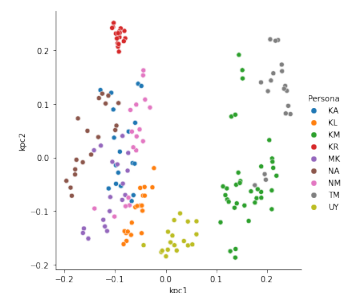


Figura 18. Representación de los datos con Kernel-PCA(Persona)

Los parámetros que se ocuparon para encontrar esta representación fueron los siguientes: un kernel Cosine. Para Kernel-PCA solo se presenta los resultados más interesantes que se obtuvieron en cada método de aprendizaje supervisado.

**III-B1. Redes Neuronales:** En la figura 19 se muestra la matriz de confusión resultante de haber ocupado Redes Neuronales. En la figura se muestra la representación de las imagenes en los dos primeros componentes.

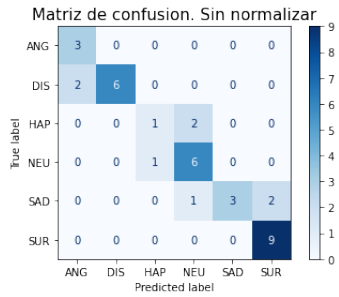


Figura 19. Matriz de Confusión-Redes Neuronales

El resultado que obtuvimos en la proporción de datos asignados correctamente fue del 78 %, comparado con PCA si hay un mejor resultado. Los parámetros que se ocuparon fueron los siguientes: Una capa con 10 neuronas y un  $\alpha = 0.5$ .

Los resultados para los demás métodos no fueron tan buenos comparados con PCA a excepción de Maquinas de Soporte Vectorial, en este caso se obtuvo el mismo porcentaje de datos asignados correctamente. En la siguiente tabla se muestran los resultados de cada método.

Método	Proporción(Kernel-PCA)
Maquinas	78 %
Arboles	30.6 %
AdaBoost	53 %
LDA	14 %
QDA	33 %

### III-C. Otras Representaciones

**III-C1. Distintos Filtros:** Para evaluar que tan bueno fue el *filtro de Gabor* en algunos métodos se tomaron en cuenta otros tipos de filtros para imágenes, fueron los siguientes: Mediana, Promedio y paso alto. Entre estos tres solo se mostraran los resultados de este paso alto y la representaciones tomando en cuenta la expresión y la persona.

Método	Proporción(Filtro Paso Alto)
Redes	69 %
Maquinas	78 %
Arboles	36.1 %
AdaBoost	56 %
LDA	61 %
QDA	30 %

En la figura 20 se ilustra la varianza acumulada y utilizando el mismo criterio que se menciono se tomaron 100 componentes principales.

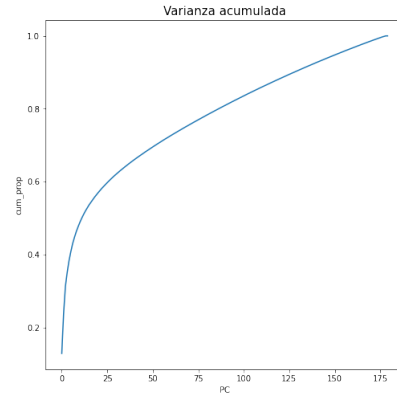


Figura 20. Proporción de Varianza Acumulada

En la figura 21 y figura 22 nos muestra las representaciones en las dos primeras componentes tomando en cuenta la expresión y la persona respectivamente.

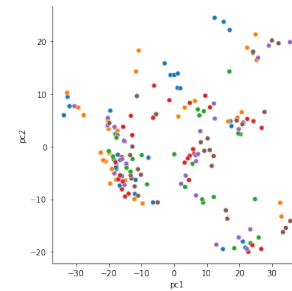


Figura 21. Representación de los datos con PCA-Expresión

Para ambos gráficos observamos un comportamiento similar al obtenido con el *filtro de Gabor*.

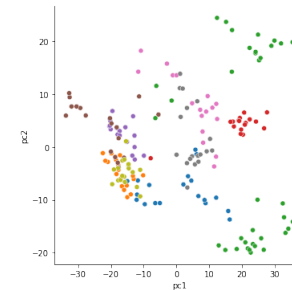


Figura 22. Representación de los datos con PCA-Persona

**III-C2. Clasificación por Persona:** Al observar que el mejor resultado para la proporción de datos clasificados correctamente fue del 81 % se propuso entrenar los mismos modelos pero ahora solo trabajando por grupo de imágenes, es decir, tomar las imágenes de una persona y evaluar que tan bueno es el proceso que se menciono reconociendo las emociones solo de una persona.

Los resultados se mostraran en una tabla. Se tomo en cuenta

lo siguiente: el filtro de Gabor, la representación con PCA, los mismos métodos y parámetros.

Iniciales del Nombre	Proporción
KA	50 %
KL	50 %
KM	50 %
KR	50 %
MK	75 %
NA	75 %
NM	75 %
TM	75 %
UY	25 %
YM	100 %

Fue interesante observar que hay una persona que obtuvo como resultado una proporción igual al 25 % y una igual al 100 %. La figuras figuras 23 y 24 corresponde a la persona que obtuvo el porcentaje del 25 %.



Figura 23. Imagen de UY-Triste



Figura 24. Imagen de UY-Sorprendido

Las figuras 25 y 26 se muestran imagenes de la persona que obtuvo el 100 % en la clasificación de expresiones.



Figura 25. Imagen de YM-Triste



Figura 26. Imagen de YM-Sorprendido

Al observar este resultados encontramos que las expresiones de UY no son tan diferentes comparados con YM ya que al observar la expresión de sorprendido entre ambas mujeres observamos que una es más expresiva que otra y esos movimientos marcados del rostro son los que mejor ayudan a identificar la expresión.

Como ejemplo la figura 24 no es muy marcada la expresión de una persona sorprendida ya que se puede tomar como expresión neutral y esto confundir al método.

*III-C3. Varios Filtros de Gabor:* Una idea extra al procedimiento fue tomar ahora una familia grande de filtros de Gabor para cada imagen. En este procedimiento se tomaron 16 nuevas representaciones para cada imagen, el unico cambio fue que se realizo una suma con cada componente de la imagen(píxel) para finalmente obtener un vector con 16 valores para cada imagen. Al obtener esta representación para las 180 imágenes se procedio a trabajar de la misma forma que con un solo filtro de Gabor, los resultados de realizar dicha tarea no nos genero buenos resultados por lo que se omiten los resultados y solo se menciona.

#### IV. CONCLUSIONES

Los resultados del proceso que se planteo al principio fueron buenos e interesantes, uno de los más importantes es la proporción de datos asignados correctamente, fue del 89 % para una representación(PCA) y el mejor comparado con distintos filtros que se mencionaron. Las representaciones con PCA y Kernel-PCA nos mostraron que el *filtro de Gabor* a parte de dar buenos resultados en las expresiones tiene posiblemente otro buen resultado para indentificar personas, la figura 18 fue la mejor representación para mostrar esta idea ya que divide bien las imágenes con su respectivo grupo.

Al tomar distintos filtros observamos que muchos de ellos no nos ayudaron a mejorar la proporción de datos asignados correctamente comparado con el *filtro de Gabor*.

Finalmente observamos que al tomar cada grupo de imagenes de cada persona y aplicar el mismo procedimiento nos dio resultados interesantes y entre estos resultados encontramos que hay personas que no muestran mucho cambio entre una expresión y otra e incluso se puede llegar a interpretar erróneamente, esto comparado con las personas que son más expresivas ayudan más al modelo y se muestra con la figura 26 donde YM es muy expresiva para sorprenderse ya que no solo mueve los ojos sino también las cejas y boca. Estos detalles ayudan mucho al procedimiento ya que facilitan la obtención de características para definir un patrón de acuerdo a cierta expresión.

Una implementación que no se pudo realizar pero que quiza hubiera ayudado a mejorar el procedimiento era tomar distintas rotaciones de las imagenes y aumentar el conjunto de prueba ya que eso posiblemente ayudara al procedimiento a identificar mejor cada expresión. Otra implementación que no se alcanzo a realizar fue tomar un conjunto de imagenes distintas al conjunto de prueba y observar que tan bien identifica las emociones para cualquier tipo de imagen ya sea dibujo o

fotografía de una persona, incluso comparar las expresiones que se realizan entre poblaciones de Latinoamérica y Asia. Las implementaciones desarrolladas en *Python* para este proyecto se anexan junto con este reporte.

#### REFERENCIAS

- [1] Matthew N., Carrie J., Michael J. Lyons, Miyuki K., Hanae I. y Jiro(2010). *Evidence and a computational explanation of cultural differences in facial expression recognition*.
- [2] Michael J. Lyons, Julien Budynek y Shigeru A.(1999). *Automatic Classification of Single Facial Images*.
- [3] Michael J. Lyons y Shigeru Akamatsu(1998). *Coding Facial Expressions with GaborWavelets*.Proceedings, Third IEEE International Conference on Automatic Face and Gesture Recognition, April 14-16 1998, Nara Japan, IEEE Computer Society, pp. 200-205.
- [4] Mohammad Hag., Saman Zon. y Mohamed Abdel(2015). *Trustworthy cloud-based and cross-enterprise biometric identification*.Proceedings, Third IEEE International Conference on Automatic Face and Gesture Recognition, April 14-16 1998, Nara Japan, IEEE Computer Society, pp. 200-205.
- [5] Libin Yang, William Rea y Alethea Rea(2017). *Financial Insights from the Last Few Components of a Stock Market PCA* .Centre for Applied Statistics, University of Western Australia, 35 Stirling Hwy, Crawley 6009, Australia; alethea.rea@uwa.edu.au