

RECONOCIMIENTO DE EXPRESIONES FACIALES UTILIZANDO ANÁLISIS DE COMPONENTES PRINCIPALES KERNEL (KPCA)

Facial Expression Recognition using kernel principal component analysis (KPCA)

RESUMEN

Este artículo presenta una metodología para el reconocimiento de expresiones faciales con análisis de componentes principales kernel, la base de datos utilizada es la Carnegie Mellon University como herramienta de prueba. El método utiliza una función kernel que mapea los datos del espacio característico original a uno de mayor dimensionalidad, de esta forma un problema de origen no lineal se traslada a uno lineal y puede resolverse linealmente, además los métodos basados en kernel pueden reducir el número de parámetros usados para la clasificación, este método es comparado con el análisis de componentes principales y es puesto a discusión donde los porcentajes de acierto encontrados con la base de datos son mayor al 90%.

PALABRAS CLAVES: Reconocimiento de patrones, PCA, Kernel, KPCA, escalamiento multidimensional.

ABSTRACT

This paper presents a methodology on the recognition of facial expressions with kernel principal component analysis using the Carnegie Mellon University database as a testing tool. This method uses a kernel function to map data from the original feature space to a higher dimensional space, through which a non-linear problem is translated into a linear one and is to be solved in a linear way, besides a kernel based method can reduce the number of parameters used by the classifier, this method compares with principal component analysis and discussed where the percentages of success found with the database is greater than 90%.

KEYWORDS: Pattern Recognition, PCA, kernel, KPCA, multidimensional scaling.

1. INTRODUCCIÓN

Obtener sistemas que puedan adaptarse a las condiciones de trabajo y aprendan de la experiencia, tienen una gran aceptación por los investigadores en diferentes campos como: ciencias de la computación, ingeniería, matemáticas, física, neurociencia y ciencia cognitiva.

En la tarea de reconocimiento de rostros uno de los métodos más utilizados es el de "autocaros" ("eigenfaces") o PCA aplicado a imágenes [1], donde el objetivo es transformar las imágenes de rostros en un grupo pequeño de características, denominadas eigenfaces o autocaras, que son los componentes principales del entrenamiento inicial del conjunto de las imágenes. El reconocimiento se lleva a cabo proyectando una nueva imagen dentro del subespacio formado por las autocaras, para su clasificación el rostro se compara por la posición en el espacio característico con las posiciones de individuos conocidos. Esta aproximación tiene ventajas frente a otros esquemas de reconocimiento facial en su simplicidad, velocidad, capacidad de aprendizaje y relativa insensibilidad a pequeños cambios en la imagen del rostro [2]. Cabe resaltar que este método toma

características globales y no locales del rostro previo preprocesamiento de la imagen.

En términos matemáticos se desea encontrar los componentes principales de la distribución de los rostros o los autovectores de la matriz de covarianza de un grupo de imágenes. Estos autovectores son un grupo de características que presentan variaciones entre las imágenes. Cada imagen contribuye a cada autovector, así se puede tener un autovector como una clase de rostro fantasma denominada autocara. De esta forma cada imagen en el entrenamiento se puede representar exactamente como una combinación lineal de las autocaras. El número de posibles autocaras es igual al número de imágenes del conjunto de entrenamiento. Los rostros pueden aproximarse usando solo las mejores autocaras que son las que tienen mayores autovectores y por tanto tienen la mayor varianza dentro del conjunto de imágenes [3].

En el reconocimiento de expresiones faciales [4] utilizan PCA y análisis discriminante lineal (LDA) para clasificar considerando una máscara facial con acierto de 85% en promedio, en [5] se hace una revisión de las diferentes

ÁLVARO ÁNGEL OROZCO G.

Ingeniero Electricista, M. Sc.
Profesor titular
Universidad Tecnológica de Pereira
aaog@utp.edu.co

MAURICIO ÁLVAREZ

Ingeniero Electrónico, M. Sc.
Profesor titular
Universidad Tecnológica de Pereira
malvarez @utp.edu.co

JUAN GABRIEL FETECUA VALENCIA

Ingeniero Electricista.
Profesor Catedrático
Universidad Tecnológica de Pereira
jgabrielval@ohm.utp.edu.co

técnicas utilizadas para el análisis automático de expresiones faciales con métodos holísticos y locales para extracción de características y clasificación, en [6] se desarrolla un sistema de reconocimiento facial con supervisión lineal licalizada embebida (SLLE) con detección de rostro, extracción de características y clasificación con resultados superiores a los de PCA, en [7] se implementan wavelets Gabor y cuantización de aprendizaje lineal (LVQ) para reconocer expresiones faciales con porcentajes de acierto hasta el 88% para diferentes dimensiones del vector característico, en [8] se utiliza aprendizaje híbrido mejorado para la detección de pose y reconocimiento de expresiones faciales con aciertos del 93% en promedio.

Estudios con métodos kernel [9-11] presentan nuevas alternativas en diferentes tareas para el análisis y procesamiento de señales específicamente en la eliminación de ruido, reducción de dimensionalidad y reconocimiento de patrones. Básicamente se realiza un mapeo de los datos del espacio de entrada \mathbf{X} a otro espacio característico φ . La función kernel $k(\cdot, \cdot)$ es equivalente al producto interno del mapeo en otro espacio $\varphi(\mathbf{x}_i) \cdot \varphi(\mathbf{x}_j)$ con $\mathbf{x}_i, \mathbf{x}_j \in \mathbf{X}$, lo que permite una generalización no lineal de la mayoría de algoritmos lineales, por ejemplo: máquinas de soporte vectorial (SVM), métodos de agrupamiento [12-13], análisis de componentes independientes kernel [14] y kernel PCA [13].

El objetivo de esta investigación es utilizar los kernels en combinación al método de eigenfaces [15] método conocido como KPCA en el reconocimiento de expresiones faciales y comparar los resultados frente al método convencional PCA. Donde los datos de entrada son mapeados del espacio característico original a uno de mayor dimensionalidad, de esta forma un problema de origen no lineal se traslada a uno lineal y se resuelve linealmente, donde la matriz de covarianza es calculada a partir del mapeo de los datos y se encuentran los vectores y valores propios, como este proceso requiere de un alto costo computacional debido a la alta dimensionalidad de los datos se resuelve el problema dual que depende de la matriz kernel de los datos y sus valores propios.

Es posible desarrollar KPCA de forma equivalente a [16] y mostrar que matricialmente se llega a la misma solución, donde el objetivo es encontrar la distancia entre la proyección de los datos de prueba y el mapeo de los datos de entrenamiento para determinar la clase a la cual pertenece cada nuevo patrón, lo cual se explicará a continuación.

2. MARCO TEÓRICO

El desarrollo de métodos de aprendizaje basados en kernel ha resultado como la combinación de teoría de aprendizaje de máquina, algoritmos de optimización y

técnicas kernel. Un kernel k se presenta como una medida de similitud, la cual puede ser vista como un producto punto en un espacio característico H , utilizando un mapeo φ , para un conjunto \mathbf{X} de entrada se tiene:

$$\begin{aligned} \varphi: \mathbf{X} &\rightarrow H \\ \mathbf{x} &\rightarrow \varphi(\mathbf{x}) \end{aligned} \quad (1)$$

donde la función kernel es:

$$k(\mathbf{x}, \mathbf{x}') = \langle \varphi(\mathbf{x}), \varphi(\mathbf{x}') \rangle \quad (2)$$

Se observa que un kernel es equivalente al producto interno del mapeo de los datos de entrada, siempre y cuando la función kernel cumpla con la condición de kernel definido positivo [17].

Dado un grupo de patrones $\{\mathbf{x}_1, \dots, \mathbf{x}_N\} \in \mathbf{R}^d$. Kernel PCA lleva a cabo el PCA tradicional en el espacio característico, en este caso se utiliza una función kernel en el problema de descomposición en componentes principales, solucionando

$$\mathbf{H}\mathbf{K}\mathbf{H} = \mathbf{U}\mathbf{E}\mathbf{U}' \quad (3)$$

donde \mathbf{K} es la matriz kernel con entradas $K_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$,

$$\mathbf{H} = \mathbf{I} - \frac{1}{N} \mathbf{1}\mathbf{1}' \quad (4)$$

es la matriz centrada, \mathbf{I} es la matriz identidad de $N \times N$, $\mathbf{1} = [1, 1, \dots, 1]'$ es un vector unitario, $\mathbf{U} = [\mathbf{a}_1, \dots, \mathbf{a}_N]$ donde $\mathbf{a}_i = [a_{i1}, \dots, a_{iN}]'$ es la matriz que contiene los vectores propios y $\mathbf{E} = \text{diag}(\lambda_1, \dots, \lambda_N)$ contiene los valores propios correspondientes. Para los patrones mapeados φ se define la media como $\bar{\varphi} = \frac{1}{N} \sum_{i=1}^N \varphi(\mathbf{x}_i)$ y el mapeo centrado como:

$$\mathbf{O}(\mathbf{x}) = \varphi(\mathbf{x}) - \bar{\varphi} \quad (5)$$

si $\mathbf{k}_x = [k(\mathbf{x}, \mathbf{x}_1), \dots, k(\mathbf{x}, \mathbf{x}_N)]'$, se define:

$$\tilde{\mathbf{k}}_x = \mathbf{H} \left(\mathbf{k}_x - \frac{1}{N} \mathbf{K}\mathbf{1} \right) \quad (6)$$

Por tanto la proyección de $\varphi(\mathbf{x})$ en el subespacio de los primeros K vectores propios es:

$$P_{\varphi}(\mathbf{x}) = \mathbf{O} \mathbf{M} \tilde{\mathbf{k}}_x + \bar{\varphi} \quad (7)$$

donde $\mathbf{M} = \sum_{k=1}^K \frac{1}{\lambda_k} \mathbf{a}_k \mathbf{a}_k'$ es simétrica.

Para el kernel Gaussiano $k(\mathbf{x}, \mathbf{y}) = e^{-\|\mathbf{x}-\mathbf{y}\|^2/c}$, es posible encontrar para dos puntos \mathbf{x}_i y \mathbf{x}_j la distancia Euclidiana en el espacio de entrada $d(\mathbf{x}_i, \mathbf{x}_j)$ y la distancia en el espacio característico entre los datos mapeados $\bar{d}(\varphi(\mathbf{x}_i), \varphi(\mathbf{x}_j))$. Se toma un patrón \mathbf{x} que se desea clasificar, el mapeo $\varphi(\mathbf{x})$ para ser proyectado en el espacio característico $P_\varphi(\mathbf{x})$, donde para cada patrón de entrenamiento \mathbf{x}_i , existirá una distancia entre el dato proyectado y el patrón de entrenamiento mapeado $\varphi(\mathbf{x}_i)$ en el espacio característico $\bar{d}(P_\varphi(\mathbf{x}), \varphi(\mathbf{x}_i))$.

Sean dos patrones \mathbf{x} y \mathbf{x}_i , la distancia cuadrática en el espacio característico entre la proyección $P_\varphi(\mathbf{x})$ y $\varphi(\mathbf{x}_i)$ se da por:

$$\bar{d}^2(P_\varphi(\mathbf{x}), \varphi(\mathbf{x}_i)) = \|P_\varphi(\mathbf{x})\|^2 + \|\varphi(\mathbf{x}_i)\|^2 - 2P_\varphi(\mathbf{x})' \varphi(\mathbf{x}_i) \tag{8}$$

Reemplazando $\tilde{\mathbf{k}}_x$ y $P_\varphi(\mathbf{x})$ en $\bar{d}^2(P_\varphi(\mathbf{x}), \varphi(\mathbf{x}_i))$ se obtiene:

$$\begin{aligned} \bar{d}^2(P_\varphi(\mathbf{x}), \varphi(\mathbf{x}_i)) = & \left(\mathbf{k}_x + \frac{1}{N} \mathbf{K} \mathbf{1} - 2\mathbf{k}_{xi} \right)' \mathbf{H}' \mathbf{M} \mathbf{H} \left(\mathbf{k}_x - \frac{1}{N} \mathbf{K} \mathbf{1} \right) \\ & + \frac{1}{N^2} \mathbf{1}' \mathbf{K} \mathbf{1} + K_{ii} - \frac{2}{N} \mathbf{1}' \mathbf{k}_{xi} \end{aligned} \tag{9}$$

donde $K_{ii} = k(\mathbf{x}_i, \mathbf{x}_i)$.

4. MATERIALES

Se utiliza la base de datos de la CMU facial expression que consta de 13 personas, con 75 imágenes por persona para 5 expresiones faciales (alegría, enojo, neutral, sorpresa y tristeza) la cual ha sido previamente procesada, se considera un conjunto total de 20 imágenes para cada tipo de expresión facial independiente de la persona, para un total de 100 imágenes de 64 x 64 píxeles. Las expresiones faciales (alegría, enojo, neutral, sorpresa y tristeza) se observan en las figuras 1-5 respectivamente.



Figura 1. Base de datos CMU, alegría.



Figura 2. Base de datos CMU, enojo.



Figura 3. Base de datos CMU, neutral.



Figura 4. Base de datos CMU, sorpresa.



Figura 5. Base de datos CMU, tristeza.

5. METODOLOGÍA

Se implementaron los algoritmos de PCA y KPCA (con kernel Gaussiano) en MATLAB, considerando diferentes pruebas, la primera para determinar el mejor parámetro kernel y la segunda determinando para que número de vectores propios el algoritmo obtenía el mejor resultado utilizando la base de datos de la CMU, para lo cual se varían los parámetros entre un mínimo y un máximo a un paso determinado.

Se toma aleatoriamente un grupo de entrenamiento y uno de validación (normalizando los datos), se define el parámetro del kernel (variable) y el número de vectores propios (variable), se calcula la matriz kernel total tanto del grupo de entrenamiento como del grupo de validación, se halla la descomposición en componentes principales, se calcula M , se toma como criterio de clasificación las distancias en el espacio característico entre cada imagen nueva y el grupo de entrenamiento, la asignación de la clase se presenta a la distancia más cercana a la imagen de prueba.

El parámetro del kernel c con la mejor discriminación de datos encontrado es $c = 500$, el mejor resultado de clasificación se presentó para 50 vectores propios. Se utiliza validación sencilla, se selecciona un porcentaje de

la base de datos en forma aleatoria para el entrenamiento (70% - 70 imágenes) y para validación (30% - 30 imágenes). Se realiza la prueba para 20 casos y se obtiene el porcentaje de acierto para cada expresión facial.

6. RESULTADOS

En las figuras 6-7, se muestra un conjunto de validación y la clasificación de expresiones faciales con KPCA respectivamente, en uno de los 20 casos de prueba. La tabla 1 muestra los porcentajes de acierto entre PCA [1] convencional y KPCA para cada expresión facial.



Figura 6. Grupo de validación aleatorio para la base de datos CMU facial expression.



Figura 7. Resultado de la clasificación con KPCA, base de datos CMU facial expression.

Expresión facial	PCA	KPCA
alegría	66,67+- 5,22%	96,67, +- 2,35%
enojo	50,00 +- 8,65%	85,00 +- 3,02%
neutral	16,67 +- 5,21%	76,67 +- 4,26%
sorpresa	83,33 +- 3,19%	95,00 +- 2,49%
tristeza	83,33+- 3,19%	85,00 +- 3,02%

Tabla 1. Porcentaje de acierto para cada una de las expresiones faciales en la prueba para la base de datos CMU facial expression.

Se observa que el resultado con KPCA es mucho mejor que PCA convencional, esto se debe principalmente a que PCA es limitado respecto al análisis lineal de los datos en el espacio característico original, por otro lado KPCA mapea los datos de entrada llevando el problema a un espacio en teoría n-dimensional donde es posible una solución lineal, además la función kernel vista como un producto interno entre el mapeo de los datos y su característica de ser definido positivo hace que cualquier función que cumpla estas condiciones pueda ser utilizada, en este caso para un kernel Gaussiano el resultado es bueno, aunque esta metodología es más utilizada para eliminación de ruido y caracterización de patrones, se demuestra que los métodos basados en kernel como KPCA, es una alternativa para el reconocimiento de expresiones faciales.

5. CONCLUSIONES Y TRABAJO FUTURO

El método kernel PCA para la tarea de reconocimiento de expresiones faciales tiene mejores resultados que el PCA convencional bajo las mismas condiciones.

La obtención de modelos estocásticos para la selección del kernel se plantea como una alternativa para mejorar los porcentajes de aciertos, lo cual se logra al obtener la matriz kernel basado en modelos estadísticos y de optimización considerando los patrones de entrenamiento.

AGRADECIMIENTOS

Se agradece al grupo de investigación en Control e Instrumentación por los recursos físicos e intelectuales para el desarrollo de esta investigación. Este trabajo se desarrolló en el marco del proyecto: "IMPLEMENTACIÓN Y EFECTIVIDAD DE UN SISTEMA BASADO EN INTELIGENCIA ARTIFICIAL COMO HERRAMIENTA PARA EL TRATAMIENTO PSICOLÓGICO DE PERSONAS CON TRASTORNO DE ESTRÉS POSTRAUMÁTICO", financiado por Colciencias. Código 111037019600 y la Universidad Tecnológica de Pereira. Código 511-3-243-08.

Se agradece al ingeniero Ricardo Henao por su colaboración en el desarrollo de esta investigación.

6. BIBLIOGRAFÍA

- [1] M. Turk and A. Pentland, "Eigenfaces for Recognition," *Journal of Cognitive Neuroscience* vol. 3, no. 1, pp. 71-86, Mar. 1991.
- [2] Xiaoguang Lu, *Image Analysis for Face Recognition*, Dept. of Computer Science & Engineering Michigan, State University, East Lansing, MI, 48824. 2003.
- [3] M. Turk and A. Pentland, "Face recognition using eigenfaces". *Proc. IEEE Conference on Computer Vision and Pattern Recognition*: 586-591. 1991.
- [4] S. Dubuisson, F. Davoine and M. Masson, A solution for facial expression representation and recognition, *Signal Processing: Image Communication*, Volume 17, Issue 9, October 2002, Pages 657-673.
- [5] B. Fasel and Juergen Luetin, Automatic facial expression analysis: a survey, *Pattern Recognition*, Volume 36, Issue 1, January 2003, Pages 259-275.
- [6] Dong Liang, Jie Yang, Zhonglong Zheng and Yuchou Chang, A facial expression recognition system based on supervised locally linear embedding, *Pattern Recognition Letters*, Volume 26, Issue 15, November 2005, Pages 2374-2389.
- [7] Shishir Bashyal and Ganesh K. Venayagamoorthy, Recognition of facial expressions using Gabor wavelets and learning vector quantization, *Engineering Applications of Artificial Intelligence*, In Press, Corrected Proof, Available online 14 January 2008.
- [8] Hsiuao-Ying Chen, Chung-Lin Huang and Chih-Ming Fu, Hybrid-boost learning for multi-pose face detection and facial expression recognition, *Pattern Recognition*, Volume 41, Issue 3, March 2008, Pages 1173-1185
- [9] Kwok, J.T.-Y. and Tsang, I.W.-H. "The Pre-Image Problem in kernel methods", *IEEE Transactions on Neural Networks*, Nov. 2004, Vol 15, pp. 1517-1525.
- [10] T.F. Cox and M.A.A. Cox, *Multidimensional Scaling*, Monographs on Statistics and Applied Probability 88. Chapman & Hall / CRC, second edition, 2001.
- [11] C.K.I. Williams, On a connection between kernel PCA and a metric multidimensional scaling, in *Advances in Neural Information Processing Systems 13*, T. Leen, T. Dietterich, and V. Tresp, Eds., Cambridge, MA, 2001, MIT Press.

- [12] M. Girolami, "Mercer kernel-based clustering in feature space," *IEEE Transactions on Neural Networks*, vol. 13, no. 3, pp. 780–784, May 2002.
- [13] B. Schölkopf, A. Smola, and K.R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Computation*, vol. 10, pp. 1299–1319, 1998.
- [14] F.R. Bach and M.I. Jordan, "Kernel independent component analysis," Tech. Rep. UCB/CSD-01-1166, Department of Computer Science, University of California, 2001.
- [15] B. Moghaddam and A. Pentland. Probabilistic visual learning for object recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7): 696–710, 1997.
- [16] S. Mika, B. Schölkopf, A. Smola, K.R. Müller, M. Scholz, and G. Rätsch, "Kernel PCA and de-noising in feature spaces," in *Advances in Neural Information Processing Systems*, 1998.
- [17] B. Schölkopf and A.J. Smola, *Learning with Kernels*, MIT Press, Cambridge, MA, 2002.