

Tipo de artículo: Artículo original
Temática: Reconocimiento de patrones
Recibido: 10/10/2015 | Aceptado: 15/12/2015

Aprendizaje de métrica para el reconocimiento de rostros a partir de imágenes de baja resolución

Metric Learning for Low-Resolution Face Recognition

Mairelys Hernández-Durán ^{1*}, Yenisel Plasencia-Calaña ¹

¹Centro de Aplicaciones de Tecnologías de Avanzada (CENATAV). 7^a A #21406 esq. 216, Playa. C.P. 12200, La Habana, Cuba. {mhduran,yplasencia@cenatav.co.cu}

* Autor para correspondencia: mhduran@cenatav.co.cu

Resumen

El reconocimiento de rostros a partir de imágenes de baja resolución es un problema muy difícil. En esta situación, la galería o base de datos contiene imágenes de alta resolución, pero la imagen a ser reconocida es de baja resolución. En consecuencia, se trata de un problema de desajuste de resolución para las imágenes de entrenamiento y prueba. Los métodos estándar de reconocimiento facial fallan en este contexto, sugiriendo que los enfoques de representación de características actuales no son suficientes para hacer frente a este problema. Por lo tanto, se propone el uso de representaciones de disimilitud como alternativa al uso de representación de características. El siguiente trabajo es una extensión a un artículo previo en el que se utilizó el espacio de disimilitudes para el reconocimiento de rostros usando imágenes de baja resolución. En el presente proponemos reemplazar una distancia euclidiana calculada sobre los vectores de características por una distancia de Mahalanobis aprendida automáticamente optimizando un criterio de clasificación en el conjunto de entrenamiento. Se propone también reemplazar la propia distancia Euclidiana en el espacio de disimilitud por una aprendida automáticamente. Los experimentos en dos conjuntos de datos faciales estándar demuestran que el uso del aprendizaje de métricas supera la distancia euclidiana inicial para el reconocimiento de rostros de baja resolución. Se utilizó la mejor estrategia obtenida en el trabajo previo para resolver el problema del desajuste en la resolución que consiste en submuestrear y luego escalar las imágenes de entrenamiento y escalar las de prueba.

Palabras clave: espacio de disimilitud, baja resolución, reconocimiento de rostros, superresolución, selección de prototipos.

Abstract

Low-resolution face recognition is a very difficult problem. In this setup, the database or gallery contains high-resolution images but the image to be recognized is a low-resolution one. Thus we are dealing with a resolution mismatch problem for training and test images. Standard face recognition methods fail in this setting, which suggests that current feature representation approaches are not adequate to cope with this problem. Therefore, we propose the use of dissimilarity representations as an alternative to the use of feature representations. This work is an extension of a previous one, in which the dissimilarity space was used for low-resolution face recognition. In this paper we propose to replace a Euclidean distance computed over the vector features for a Mahalanobis distance, which is a metric automatically learned by optimizing a classification criterion in the training set. We also propose to replace the Euclidean distance in the dissimilarity space by a metric automatically learned. Experiments on two standard face datasets demonstrate that the use of metric learning outperforms the initial Euclidean distance for low-resolution face recognition. To solve the mismatch problem, the best strategy obtained in previous work was used, which consist on subsample and then scale the training images and scale test.

Keywords: *dissimilarity space, low-resolution, face recognition, super-resolution, prototype selection.*

Introducción

El reconocimiento de rostros ha sido estudiado por décadas debido a su amplia gama de aplicaciones. A pesar de que se han obtenido altas tasas de reconocimiento en ambientes controlados, para los sistemas de reconocimiento de rostros a partir de imágenes de baja resolución (BR) los resultados siguen siendo insatisfactorios. Estos sistemas tratan de identificar imágenes de rostro de BR donde la galería está compuesta por imágenes de alta resolución (AR). La BR de las imágenes afecta el desempeño de los sistemas de reconocimiento de rostros tradicionales (Hennings, 2008). Los enfoques actuales incluyen principalmente la representación de vectores de características para facilitar el poder discriminativo entre diferentes rostros y enfrentar el problema del reconocimiento de rostros a partir de imágenes de BR. Algunos métodos como el vecino más cercano (1-NN) y la interpolación bicúbica constituyen la forma más sencilla de incrementar la resolución a partir de una imagen de BR (Wang, 2014). Otros enfoques han considerado los métodos de representación de características de resolución robustas para el caso de la BR (Li, 2010). Sin embargo, para el reconocimiento de rostros a partir de imágenes de BR esto es difícil de encontrar porque la mayoría de las características eficaces que se utilizan en los métodos tradicionales que emplean imágenes de AR (como textura y color) pueden fallar en el caso de la BR. Por lo tanto, la mayoría de los enfoques tradicionales exitosos no pueden ser empleados de forma eficaz para el caso de la BR (Wang, 2014).

Una solución alternativa es la representación por disimilitudes entre los objetos basada en la idea propuesta en (Pekalska, 2005). Una representación basada en disimilitudes ofrece ventajas en situaciones en las que es difícil definir rasgos discriminativos suficientes y es más fácil definir similitudes (Pekalska, 2005). Dicha representación permite explotar grandes conjuntos de entrenamiento, aumentando la precisión, mientras la complejidad sigue siendo la misma (Orozco, 2007). Basado en el éxito de enfoques previos (Pekalska, 2005), en este trabajo se utilizó la representación por disimilitudes para resolver el problema de clasificación que aborda a los autores. Se considera que la representación por disimilitudes puede ser adecuada para el reconocimiento de rostros a partir de imágenes de BR porque ha sido usada para resolver problemas complejos como: situaciones de tamaño de la muestra pequeño (Orozco, 2007) o problemas en los que los resultados del 1-NN sobre los rasgos siguen siendo insatisfactorios. Además, este tipo de representación ha sido utilizada con éxito en múltiples tareas como la reidentificación de personas (Satta, 2012) y la clasificación de objetos (Carli, 2010).

El k-NN es un algoritmo sencillo pero eficaz para la clasificación. Un elemento clave en el mismo es la elección de una función de distancia o métrica que capture el tipo de invarianza usada para medir similitudes entre pares de objetos (Ramanan, 2011). Algunos investigadores han demostrado que la clasificación del 1-NN se puede mejorar en gran medida por el aprendizaje de una métrica de distancia apropiada para objetos etiquetados (Chopra, 2005). Específicamente, las métricas que son aprendidas superan el desempeño de otras métricas no aprendidas automáticamente, tienen la ventaja de mejorar la distancia euclidiana original en el sentido de la clasificación y proveer un embedding a un espacio de menor dimensión (Bar-Hillel, 2005). En (Weinberger, 2009) aprenden la distancia métrica de Mahalanobis para mejorar la clasificación de 1-NN utilizando varios conjuntos de datos de muestras etiquetadas. Dicha solución ofrece las ventajas de convergencia a un óptimo local y optimización de un criterio de clasificación en el 1-NN. Siguiendo este enfoque y los resultados exitosos de otros autores (Bar-Hillel, 2005), en el presente trabajo se utiliza el aprendizaje de métrica para mejorar los resultados obtenidos con la distancia euclidiana en un artículo anterior de los autores para el problema de clasificación que les ocupa.

Se decidió mejorar los resultados de la distancia euclidiana en el trabajo anterior utilizando aprendizaje de métricas por las ventajas que ofrece este enfoque. Siguiendo el trabajo de (Weinberger, 2009) se sustituyó el uso de la distancia euclidiana por la distancia de Mahalanobis. Esta distancia se puede ver como una transformación lineal del espacio de entrada que precede a la clasificación 1-NN usando distancias euclidianas (Weinberger, 2009). El algoritmo trata de incrementar el número de muestras de entrenamiento aprendiendo una transformación lineal del espacio de entrada, que precede la clasificación k-NN usando distancias euclidianas. Esta transformación se obtiene minimizando una función de coste que consiste en dos términos. Al minimizar estos términos se obtiene una transformación lineal del

espacio de entrada de manera que las muestras de entrenamiento quedan más cercanas a sus k -vecinos más cercanos. La distancia euclidiana en este espacio transformado puede ser equivalente a la distancia de Mahalanobis en el espacio original (Weinberger, 2009). En el presente artículo se utiliza el aprendizaje de métrica como un problema de optimización convexa a partir del método propuesto en (Weinberger, 2009).

En este trabajo, se presenta además una alternativa a la representación basada en rasgos para el reconocimiento de rostros a partir de imágenes de BR, con el uso de la representación en espacio de disimilitudes (DS). La misma se comparó con la representación basada en rasgos para imágenes de rostro de BR. El presente trabajo constituye una extensión a un artículo previo de los autores en el que se compararon diferentes estrategias para lidiar con el problema de desajuste en la resolución usando el espacio de disimilitudes para el reconocimiento de rostros de baja resolución. A pesar de haber obtenido resultados prometedores en el trabajo anterior, en el presente se reemplaza la distancia euclidiana en el espacio de características y en el espacio de disimilitudes por una distancia aprendida utilizando aprendizaje de métricas, específicamente la distancia de Mahalanobis. Los experimentos muestran mejores resultados a los reportados con anterioridad y la métrica aprendida supera los resultados en comparación con otras métricas no aprendidas de forma automática. Para enfrentar el problema del desajuste en la resolución de las imágenes de entrenamiento y prueba se utilizó la estrategia de baja-alta que fue la que mejores resultados arrojó en el trabajo anterior, donde las imágenes de entrenamiento son submuestreadas y luego escaladas mientras que las imágenes de prueba son escaladas. Los experimentos muestran que la representación DS supera a la representación en espacio de rasgos (FS). Se propone finalmente el uso de representación en espacio de disimilitudes reducido (EDR) usando selección de prototipos.

Materiales y métodos

La representación por disimilitudes ha sido estudiada en un gran número de problemas (Orozco, 2007- Bunke, 2008); sin embargo, su aplicación para resolver el problema del desajuste en la resolución aún no ha sido analizada. Consideramos que en el contexto de comparaciones con otros objetos es posible lidiar mejor con la falta de información característica de este tipo de imágenes y con la poca discriminabilidad de la representación basada en características cuando se usan imágenes de BR.

Siendo X el espacio de objetos, y $R = \{r_1, r_2, \dots, r_k\}$ el conjunto de prototipos de manera que $R \in X$, y siendo $d: X \times X \rightarrow R^+$ una medida de disimilitud adecuada para el problema; para un conjunto de entrenamiento $T = \{x_1, x_2, \dots, x_l\}$

de forma que $T \in X$, un mapeo $\Phi_R^d: X \rightarrow R^k$ define la incorporación de objetos de entrenamiento y prueba en el DS por las disimilitudes con los prototipos:

$$\Phi_R^d(x_i) = [d(x_i, r_1) \ d(x_i, r_2) \ \dots \ d(x_i, r_k)]$$

Una métrica satisface algunas propiedades como la simetría y la desigualdad triangular. Se puede obtener una familia de métricas sobre X computando distancias euclidianas luego de realizar una transformación lineal $v = Lv$. Estas métricas computan distancias cuadráticas que pueden ser expresadas en términos de la matriz cuadrada: $M = L^T L$. Cualquier matriz M formada de esta forma a partir de una matriz de valor real L es semidefinida positiva, referida a la métrica de Mahalanobis (Weinberger, 2009). Esta distancia trata de mejorar la distancia euclidiana representada como una matriz en el medio de la multiplicación entre un vector y su traspuesta para darle diferentes pesos a cada una de las coordenadas de los vectores de referencia ($v^T M v$). Tiene las ventajas de converger a un óptimo global y maximizar el criterio de cercanía de objetos de la misma clase.

En la situación que aborda el presente trabajo las imágenes de prueba son de BR, por lo que se necesita decidir cómo enfrentar el problema del desajuste en la resolución. Se utilizó la estrategia de baja-alta (imágenes de prueba escaladas, imágenes de entrenamiento submuestreadas y luego escaladas y prototipos de AR). El mismo conjunto de entrenamiento se puede usar como conjunto de prototipos. Sin embargo, para conjuntos de entrenamiento de moderado a gran tamaño, se necesita una selección de los mejores prototipos. Diferentes enfoques han dedicado sus estudios a este propósito (Bunke, 2008). En (Plasencia, 2014), se propuso un algoritmo genético (AG), que mostró ser muy rápido y eficaz en la selección de un conjunto de prototipos. Este trabajo utiliza la estrategia de selección de prototipos supervisada de (Plasencia, 2014), para encontrar un conjunto de prototipos adecuado para una cardinalidad deseada o dada de un DS.

Resultados y discusión

Se utilizaron dos bases de datos de pruebas internacionales para los experimentos: Olivetti Research Laboratory (ORL) (Samaria, 1994), y Labeled Faces in the Wild (LFW) (Huang, 2007). Las imágenes de prueba se obtuvieron submuestreando las imágenes originales usando el método de interpolación bicúbica. Todas las imágenes fueron normalizadas geométricamente por el centro de los ojos a un tamaño de BR de 10x12 píxeles. Se utilizó la interpolación bicúbica en el escalado de las imágenes para obtener imágenes de AR de 64x80 píxeles.

Olivetti Research Laboratory (ORL): Contiene 400 imágenes en escala de grises de 40 individuos, 10 imágenes por cada uno. Algunas imágenes fueron tomadas con cierto intervalo de tiempo de diferencia. Presentan variaciones en la expresión facial (ojos abiertos/cerrados), cambios de iluminación, diferentes detalles en el rostro (con/sin espejuelos) y diferencias en la escala, entre otras. La figura 1 muestra algunos ejemplos de variaciones en dicha base de datos.



Figura. 1. Algunos ejemplos de la base de datos ORL

Labeled Faces in the Wild (LFW): Contiene 13233 imágenes etiquetadas de 5749 personas. Para 1680 personas están disponibles dos o más imágenes de rostro. Es una base de datos moderna y difícil que representa un reto porque las imágenes de rostro fueron detectadas en la web. Las imágenes presentan variaciones incluyendo cambios en la escala, pose, fondo, estilo del cabello, ropa, expresión, resolución de la imagen, enfoque y otros. Durante los experimentos se utilizó un subconjunto de la base de datos consistente en 3832 imágenes pertenecientes a 178 clases, seleccionando las clases que tuvieran 8 o más imágenes. Algunos ejemplos se muestran en la figura 2.



Figura. 2. Algunos ejemplos de la base de datos LFW

Los conjuntos de datos se dividieron de forma aleatoria en conjunto de entrenamiento y conjunto de prueba, realizando cinco pruebas. Los clasificadores y los prototipos se entrenaron usando el conjunto de entrenamiento y se computaron errores de clasificación para el conjunto de prueba. Se consideraron dos espacios de representación: FS y DS; y dos clasificadores: análisis discriminante lineal LDC (del inglés *Linear Discriminant Classifier*) y el 1-NN. Se computaron patrones binarios locales en bloques para obtener la representación de rasgos; obteniendo histogramas en cada bloque y concatenándolos. Para el AG se utilizaron parámetros similares a los presentados en (Plasencia, 2014). La cantidad de prototipos usados fue de 100 en todos los casos, conllevando a espacios de cardinalidad 100. El espacio a partir de la transformación lineal hallada por el método de aprendizaje métrico también es de 100 dimensiones.

Se considera que las imágenes de BR se benefician de la representación relacional debido a que los rasgos pueden no capturar información relevante para el proceso discriminativo. La clasificación con LDC en el DS mostró resultados prometedores en comparación al clasificador 1-NN. Se encontró que mientras la resolución de las imágenes de prueba aumenta, los resultados de la clasificación con LDC son prometedores. En (Hennings, 2008) encontraron que el cotejo en el dominio de BR es mejor que aplicar superresolución cuando las imágenes son de muy BR. Se encontró además que un pequeño conjunto de prototipos adecuadamente seleccionado es suficiente para alcanzar errores de clasificación más bajos. Esto es altamente beneficioso porque implica que en tiempo de prueba solo se necesita medir las disimilitudes con los conjuntos pequeños de prototipos. Se quiere destacar el hecho de que no se comparan métodos de extracción de rasgos, ya que requeriría el cálculo de diferencias con todos los prototipos antes de realizar la reducción para objetos de prueba entrantes. Esto supone un coste computacional adicional que se evita con la presente propuesta.

Para los dos clasificadores utilizados se usó la distancia euclidiana en el FS y en el DS. Estos resultados también se compararon con el uso del aprendizaje de métricas en el espacio de rasgos. En las tablas 1 y 2 se muestran los resultados obtenidos en la base de datos ORL y LFW, respectivamente. A partir de las tasas de error obtenidas se puede ver que la representación en DS es superior a la representación en el FS para el clasificador LDC y que el aprendizaje de métrica en el espacio de rasgos mejora aún más los resultados anteriores para ambos clasificadores.

Tabla 1. Tasas de error obtenidas utilizando la distancia euclidiana en la base de datos ORL

Clasificador	Espacio de rasgos	Espacio de disimilitudes	Aprendizaje de métrica en el espacio de rasgos
1-NN	0.19	0.38	0.09
LDC	0.15	0.12	0.06

Tabla 2. Tasas de error obtenidas utilizando la distancia euclidiana en la base de datos LFW

Clasificador	Espacio de rasgos	Espacio de disimilitudes	Aprendizaje de métrica en el espacio de rasgos
1-NN	0.83	0.89	0.79
LDC	0.78	0.76	0.73

Para el caso de la base de datos LFW, que consideramos una base de datos más compleja, se intentó usar además medidas de distancia más robustas. Específicamente se utilizó el aprendizaje de métrica sobre el espacio de disimilitud creado a partir de distancias chi-cuadrado con el objetivo de mejorar la distancia euclidiana sobre ese espacio. Los resultados obtenidos se muestran en la tabla 3. A partir de estos resultados se pudo constatar que el

aprendizaje de una métrica de Mahalanobis para reemplazar la distancia euclidiana en el espacio de disimilitud superó significativamente los resultados de la clasificación en el espacio de rasgos y en el espacio de disimilitudes.

Tabla 3. Tasas de error obtenidas utilizando la distancia euclidiana en la base de datos LFW

Clasificador	Espacio de disimilitudes usando distancia chi-cuadrado	Aprendizaje de métrica sobre espacio de disimilitudes creado usando distancia chi-cuadrado
1-NN	0.85	0.73
LDC	0.68	0.63

Conclusiones

En este trabajo se presentó el espacio de disimilitudes reducido para enfrentar el problema del desajuste de resolución que afecta el reconocimiento de rostros a partir de imágenes de BR. Se compararon representaciones basadas en disimilitudes con representaciones basadas en rasgos.

Se propone transformar también las imágenes de la galería para asemejarlas a las imágenes de prueba de BR. Los experimentos mostraron que se puede obtener más información discriminativa para la clasificación si las imágenes de BR son analizadas en el contexto de disimilitudes con otras imágenes. Se reportaron resultados prometedores para el clasificador LDC en relación al 1-NN. En este trabajo proponemos el uso de métricas aprendidas para el reconocimiento de rostros de BR combinado con una representación previamente propuesta basada en espacio de disimilitud. Se utiliza la estrategia del aprendizaje de métrica para superar el desempeño de la distancia euclidiana en el espacio de disimilitud, haciendo uso de la distancia de Mahalanobis. Se mostró que el aprendizaje de métricas supera tanto los resultados de la distancia euclidiana en el espacio de rasgos como los resultados de la misma en el espacio de disimilitudes. Las futuras investigaciones estarán dedicadas al estudio de otros métodos de aprendizaje de métricas que optimizan criterios diferentes al usado en este trabajo.

Agradecimientos

Agradecemos de forma muy especial a la Dra. Veronika Cheplygina, perteneciente al Grupo de Imágenes Biomédicas del Centro Médico Erasmus y al Laboratorio de Reconocimiento de Patrones de la Universidad de Tecnología Delf, en Holanda por su colaboración en algunos de los temas tratados en el presente artículo.

Referencias

- HENNINGS-YEOMANS, P.H., BAKER, S., KUMAR, B.V.: Simultaneous super-resolution and feature extraction for recognition of low-resolution faces. In: Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, IEEE (2008) 1-8
- WANG, Z., MIAO, Z., WU, Q.J., WAN, Y., TANG, Z.: Low-resolution face recognition: a review. The Visual Computer 30(4) (2014) 359-386
- PEKALSKA, E., DUIN, R.P.W.: The Dissimilarity Representation for Pattern Recognition: Foundations and Applications (Machine Perception and Artificial Intelligence). World Scientific Publishing Co., Inc., River Edge, NJ, USA (2005)
- LI, B., CHANG, H., SHAN, S., CHEN, X.: Low-resolution face recognition via coupled locality preserving mappings. Signal Processing Letters, IEEE 17(1) (2010) 20-23
- OROZCO-ALZATE, M., DUIN, R.P., CASTELLANOS-DOMÍNGUEZ, C.G.: On selecting middlelength feature lines for dissimilarity-based classification. In: XII Simposio de Tratamiento de Señales, Imágenes y Visión Artificial, STSIVA 2007. (2007)
- SATTA, R., FUMERA, G., ROLI, F.: Fast person re-identification based on dissimilarity representations. Pattern Recognition Letters 33(14) (2012) 1838-1848
- CARLI, A., CASTELLANI, U., BICEGO, M., MURINO, V.: Dissimilarity-based representation for local parts. In: Cognitive Information Processing (CIP), 2010 2nd International Workshop on, IEEE (2010) 299-303
- OROZCO-ALZATE, M., CASTELLANOS-DOMÍNGUEZ, C.: Nearest feature rules and dissimilarity representations for face recognition problems. Face Recognition; International Journal of Advanced Robotic Systems, Vienna, Austria (2007) 337-356
- BUNKE, H., RIESEN, K.: Graph classification based on dissimilarity space embedding. In: N. da Vitoria Lobo et al., editor, SSSPR, LNCS 5342. (2008) 996-1008
- PLASENCIA-CALAÑA, Y., OROZCO-ALZATE, M., MÉNDEZ-VÁZQUEZ, H., GARCÍA-REYES, E., DUIN, R.P.W.: Towards scalable prototype selection by genetic algorithms with fast criteria. In Franti, P., Brown, G., Loog, M., Escolano, F., Pelillo, M., eds.: S+SSPR. Volume 8621 of LNCS. Springer Berlin Heidelberg (2014) 343-352

PLASENCIA-CALAÑA, Y., CHEPLYGINA, V., DUIN, R.P.W., GARCÍA-REYES, E.B., OROZCO- ALZATE, M., TAX, D.M.J., LOOG, M.: On the informativeness of asymmetric dissimilarities. In: Proceedings of the Second International Conference on Similarity-Based Pattern Recognition. SIMBAD'13, Berlin, Heidelberg, Springer-Verlag (2013) 75-89

DINH, V.C., DUIN, R.P.W., LOOG, M.: A study on semi-supervised dissimilarity representation. In: International Conference on Pattern Recognition, IEEE (2012) 2861-2864

WEINBERGER, KILIAN Q., AND LAWRENCE K. SAUL. Distance metric learning for large margin nearest neighbor classification. The Journal of Machine Learning Research 10 (2009): 207-244.

RAMANAN, DEVA, AND SIMON BAKER. "Local distance functions: A taxonomy, new algorithms, and an evaluation." Pattern Analysis and Machine Intelligence, IEEE Transactions on 33.4 (2011): 794-806.

BAR-HILLEL, A., HERTZ, T., SHENTAL, N., & WEINSHALL, D. Learning a mahalanobis metric from equivalence constraints. Journal of Machine Learning Research, 6(6), (2005) 937-965.

CHOPRA, S., HADSELL, R., & LECUN, Y. (2005, June). Learning a similarity metric discriminatively, with application to face verification. In Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on (Vol. 1, pp. 539-546). IEEE.

SAMARIA, F.S., *T, F.S.S., HARTER, A., Site, O.A.: Parameterisation of a stochastic model for human face identification (1994)

HUANG, G.B., RAMESH, M., BERG, T., Learned-Miller, E.: Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical report, Technical Report 07-49, University of Massachusetts, Amherst (2007)