

ASAI 2015, 16° Simposio Argentino de Inteligencia Artificial.

# Identificación de hojas de plantas usando Vectores de Fisher

Javier A. Redolfi<sup>1,2,3</sup>, Jorge A. Sánchez<sup>2,3</sup> y Julián A. Pucheta<sup>2</sup>

<sup>1</sup> Centro de Investigación en Informática para la Ingeniería, Universidad Tecnológica Nacional, Facultad Regional Córdoba, Maestro López S/N, Córdoba, Argentina,

<sup>2</sup> Universidad Nacional de Córdoba, Haya de la Torre S/N, Córdoba, Argentina,

<sup>3</sup> CONICET, Haya de la Torre S/N, Córdoba, Argentina

[jredolfi@frc.utn.edu.ar](mailto:jredolfi@frc.utn.edu.ar)

**Resumen** En este trabajo se propone el uso de descriptores locales codificados con el método de Vectores de Fisher para resolver el problema de clasificación de hojas de plantas usando imágenes. Se demuestra que con este enfoque se obtienen resultados de estado del arte en conjuntos de datos públicos, sin la necesidad de recurrir a la creación de nuevos descriptores visuales específicos para el problema, ni la necesidad de realizar ningún preprocesamiento de las imágenes a analizar.

**Keywords:** identificación de plantas, vectores de Fisher, clasificación de imágenes, agricultura de precisión

## 1. Introducción

El problema de clasificación de especies de plantas es un tema al cual se le está prestando una creciente atención en los últimos años [2,8,10,12,17]. Algunas de las razones son la gran cantidad de especies en peligro de extinción y las altas tasas de deforestación debida al corrimiento de la frontera agropecuaria y a una mala planificación urbana. Las plantas tiene un rol crucial para la vida en la tierra y su descuido trae aparejados problemas irreversibles para la sociedad, como son el calentamiento global, la pérdida de la biodiversidad y el daño ambiental [2,18].

También en el ámbito académico, el problema presenta un desafío muy interesante por su dificultad, incluso para personas entrenadas resulta muy difícil la identificación de especies de plantas. Esto se puede explicar por la gran cantidad de especies de plantas, contabilizadas aproximadamente en 200000, también por la gran variabilidad dentro de una misma clase y las similitudes entre distintas clases. Esto último cataloga al problema como uno de clasificación de grano fino [17].

Además de la importancia que tiene la identificación de plantas para el cuidado del medio, también puede ser útil en aplicaciones de agricultura de precisión, como puede ser la detección de malezas para la pulverización selectiva [5].

## 2. Trabajos Relacionados

En la literatura se han propuesto una gran cantidad de algoritmos de preprocesamiento, extracción de descriptores visuales y clasificación de estos descriptores, específicos para la identificación de especies de plantas.

En algunos de los trabajos los autores se concentran en la definición de nuevos descriptores visuales útiles para la descripción de hojas, en este sentido uno de los primeros trabajos es el de Wu et. al. [19], en el cual se calculan descriptores geométricos y morfológicos básicos, los cuales son clasificados usando redes neuronales probabilísticas (PNN). Este algoritmo es semiautomático porque para el cálculo de uno de los descriptores propuestos un usuario debe marcar los dos puntos terminales de la nervadura principal de la hoja. Siguiendo con esta modalidad, en [9] al igual que en el trabajo anterior se usa como clasificador una PNN, pero se propone el uso de momentos de color, descriptores de forma, de nervaduras y de textura para mejorar la precisión. Los autores de [13] definen otro tipo de descriptores basados en el contorno de la hoja, en el espectro del mismo, en su forma y geometría. Para clasificar usan un método basado en los rangos de valores de cada descriptor. Los autores concluyen que el sistema propuesto no puede ser usado en imágenes con fondos complejos, esto se debe a que antes del cálculo de los descriptores se debe computar el contorno de la hoja, lo cual es difícil cuando en la imagen hay más de una hoja. En [10] los autores proponen el uso de descriptores geométricos y momentos invariantes, pero además definen dos nuevos descriptores que miden cuan rugoso es el contorno de la hoja. Estos nuevos descriptores como los anteriores dependen fuertemente de un correcto preprocesamiento de las imágenes, el cual debe remover el ruido y extraer el contorno. También deben usar un algoritmo para remover el tallo. Los descriptores resultantes son clasificados usando discriminantes lineales. Otro de los trabajos [18], propone la combinación de descriptores de forma basados en el contorno y definen un nuevo descriptor de textura basado en el modelo de intersección cortical que trata de imitar la respuesta de la corteza visual. Su extracción también requiere un preprocesamiento de la imagen que incluye segmentación, suavizado y realce de contraste y textura. Para la clasificación utilizan Máquinas de Soporte Vectorial (SVM). Los autores de [8], proponen el uso de descriptores generados desde la matriz de co-ocurrencia de niveles de grises (GLCM), el concepto de lagunaridad para modelar textura y descriptores de Shen para modelar la forma. Estos descriptores son clasificados con un esquema bayesiano.

En otra clase de trabajos los autores se concentran en proponer nuevos métodos de clasificación, por ejemplo en [17] se plantea el problema como uno de recuperación de imágenes. El algoritmo es semi-automático porque retorna las clases más probables ordenadas en forma descendente y el usuario debe desambiguar cual es la clase verdadera, además para el cálculo de los descriptores un usuario debe elegir la base y el ápice de la hoja. El esquema de recuperación contiene dos etapas, la primera es un árbol de búsqueda y la segunda un discriminante lineal. En el trabajo de Hsiao et. al. [6], se propone un nuevo esquema de clasificación basado en bolsa de palabras (BoW) y codificación rala. Como descriptores vi-

suales usan SIFT calculado en forma densa. Una de las ventajas de este trabajo es que no necesita la extracción del contorno para una correcta extracción de los descriptores. En [12] los autores proponen un nuevo método de identificación de plantas basado en descriptores de kernel (KDES). Este método se basa en el clásico esquema de bolsa de palabras visuales aplicado sobre descriptores de kernel. Para clasificar usan SVM en un esquema de uno contra todos (OvA).

### 3. Método Propuesto

Como se puede ver en la literatura analizada anteriormente, en la mayoría de los trabajos se propone el uso de nuevos descriptores diseñados específicamente para clasificar hojas ([19], [9], [13], [10], [18]) y en otros se diseñan nuevos algoritmos para clasificar estos descriptores ([13], [17], [6]).

En este trabajo sostenemos que para resolver el problema de clasificación de imágenes de hojas, no es necesario el diseño de nuevos descriptores específicos ni de nuevos esquemas de clasificación, sino que es suficiente el uso de esquemas ya probados en clasificación de imágenes en general y de grano fino en particular.

El esquema elegido es similar al que se describe en [16], el cual presenta varias ventajas con respecto a los trabajos analizados, por ejemplo no necesita el uso de preprocesamiento de las imágenes para una correcta extracción del contorno de la hoja, necesario para la extracción de descriptores geométricos y de forma; al usar descriptores locales calculados en forma densa, no es necesario que las imágenes sean de una hoja con un fondo fácilmente distinguible, lo que hace que pueda ser usado en clasificación de imágenes naturales.

El sistema contiene cuatro etapas, la primera es la extracción de descriptores visuales de forma densa en la imagen, luego estos descriptores son reducidos en dimensionalidad usando Análisis de Componentes Principales (PCA) [1], estos nuevos descriptores son codificados usando el método conocido como vectores de Fisher (FV) y por último estos vectores son clasificados usando SVM [3]. En la figura 1 se puede ver un diagrama completo y a continuación se explican cada una de las partes.

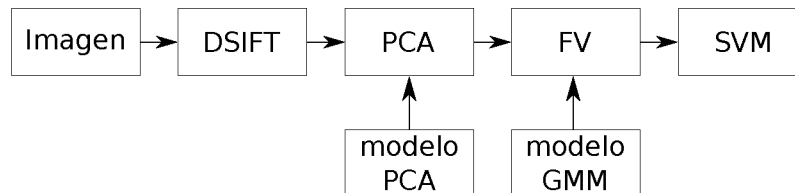
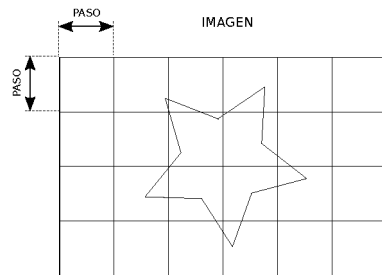


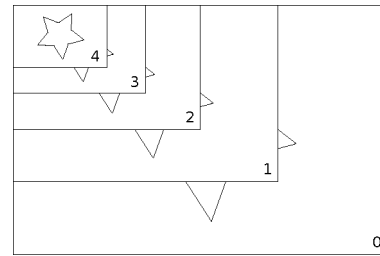
Figura 1. Diagrama en bloques del sistema propuesto.

### 3.1. Descriptores

Los descriptores son extraídos en forma densa, sobre una grilla regular con el mismo paso en ambas direcciones como se puede ver en la figura 2. Además estos se calculan sobre la imagen original y en 4 escalas de la misma, con un escalado de  $\frac{1}{\sqrt{2}}$ , esto se puede ver en la figura 3, en donde la escala 0 representa la imagen original. El descriptor utilizado es SIFT [14].



**Figura 2.** Grilla densa para el cálculo de descriptores.



**Figura 3.** Escalas de la imagen sobre la cual se calculan las grillas densas de descriptores.

### 3.2. Análisis de Componentes Principales

Para reducir la complejidad del problema se aplica la técnica de PCA. El principio de PCA es transformar un conjunto de observaciones posiblemente correlacionadas en un conjunto de variables no correlacionadas las cuales son conocidas como componentes principales. Normalmente el número de componentes principales es menor que el número de variables originales, lo cual nos permite reducir la dimensionalidad de los descriptores y como consecuencia la del problema.

### 3.3. Vectores de Fisher

La representación usando vectores de Fisher [16] es actualmente una de las más robustas en clasificación de imágenes [11] y en clasificación de grano fino [4]. Para calcular esta representación, los descriptores son modelados como muestras sampleadas de una mezcla de distribuciones Gaussianas (GMM) [1]. Luego una imagen es codificada como un vector gradiente que caracteriza la distribución de las muestras con respecto a los parámetros del modelo probabilístico.

Supongamos una imagen  $I$ , en la que extraemos descriptores en forma densa,

$$I = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$$

en donde  $\mathbf{x}_i \in \mathbb{R}^D$  representa a los descriptores calculados en la imagen  $I$  después de haber sido reducidos a  $D$  dimensiones con PCA.

Definamos a  $\Theta = \{\mu_k, \Sigma_k, \pi_k : k = 1, \dots, K\}$  como los parámetros de una distribución mezcla de  $K$  Gaussianas definida como:

$$p(\mathbf{x}_i | \Theta) = \sum_{k=1}^K \pi_k p(\mathbf{x}_i | \mu_k, \Sigma_k)$$

$$p(\mathbf{x}_i | \mu_k, \Sigma_k) = \frac{1}{\sqrt{(2\pi)^D |\Sigma_k|}} \exp \left[ -\frac{1}{2} (\mathbf{x}_i - \mu_k)^T \Sigma_k^{-1} (\mathbf{x}_i - \mu_k) \right]$$

en donde  $\mu_k$ ,  $\Sigma_k$  y  $\pi_k$  representan al vector de medias, la matriz de covarianza y el coeficiente de mezcla de la Gaussiana  $k$ . Además supongamos que la matriz de covarianza es diagonal.

Sean los gradientes con respecto a la media ( $\mu_k$ ) y a la covarianza ( $\Sigma_k$ ), evaluados en el descriptor  $i$  y en la dimensión  $j$ , respectivamente:

$$\mu_{jk} = \frac{1}{N \sqrt{\pi_k}} \sum_{i=1}^N q_{ik} \frac{x_{ji} - \mu_{jk}}{\sigma_{jk}}$$

$$v_{jk} = \frac{1}{N \sqrt{2\pi_k}} \sum_{i=1}^N q_{ik} \left[ \left( \frac{x_{ji} - \mu_{jk}}{\sigma_{jk}} \right)^2 - 1 \right]$$

en donde  $q_{ik}$  esta definido como:

$$q_{ik} = \frac{\exp \left[ -\frac{1}{2} (\mathbf{x}_i - \mu_k)^T \Sigma_k^{-1} (\mathbf{x}_i - \mu_k) \right]}{\sum_{t=1}^K \exp \left[ -\frac{1}{2} (\mathbf{x}_i - \mu_t)^T \Sigma_t^{-1} (\mathbf{x}_i - \mu_t) \right]}$$

y  $\sigma_{jk}^2$  es la componente  $(j, j)$  de la matriz de covarianza de la gaussiana  $k$ .

El FV de la imagen  $I$  es la concatenación de los gradientes  $\mu_k$  y  $\mathbf{v}_k$  para cada una de las  $K$  componentes:

$$FV(I) = [\mu_0 \quad \mathbf{v}_0 \quad \dots \quad \mu_K \quad \mathbf{v}_K]^T$$

Usando descriptores robustos como SIFT, la codificación con FV permite obtener resultados estado del arte en problemas de clasificación y reconocimiento.

Como no es el objetivo de este trabajo una derivación completa de la codificación con FV, para una mejor y más detallada derivación de la misma ver [16].

### 3.4. Clasificadores

Para la clasificación de los FV se utilizaron Máquinas de Soporte Vectorial (SVM) con kernel lineal, debido a que es la elección normal para este tipo de problemas. El uso de kernels de tipo no lineal resulta problemático por la gran dimensionalidad de los FV. Como el problema a resolver es uno del tipo multiclase, se entreno un SVM para cada clase con un esquema del tipo OvA y el score para cada imagen se obtuvo con una función del tipo argmax.

## 4. Experimentos

Para demostrar la viabilidad del esquema propuesto se realizaron experimentos en diferentes datasets comúnmente usados para la evaluación de sistemas de clasificación de imágenes de hojas. La ventaja de usar datasets públicos es que nos permite comparar de una forma rápida y fácil nuestros resultados con los de los otros algoritmos.

### 4.1. Datasets

El primero de los datasets usados es el presentado en [19], más conocido como Flavia, el cual consiste en 1907 imágenes de hojas de 32 árboles distintos, con un mínimo de 50 muestras por clase y un máximo de 72. El procedimiento normal de evaluación es dejar 10 muestras de cada clase para test y entrenar sobre las restantes. Como la elección de las muestras para test no está definida, en los experimentos sobre Flavia se decidió mostrar los resultados de 10 corridas con conjuntos de entrenamiento y test generados en forma aleatoria, siempre respetando que el tamaño del conjunto de test sea de 10 muestras.

El otro dataset usado es conocido como Foliage [7], el cual contiene 120 muestras para cada una de 60 especies diferentes de árboles. El procedimiento recomendado para la evaluación es tomar 100 muestras para entrenamiento y 20 para test por cada clase. En este dataset solo se realiza una corrida porque los conjuntos ya están definidos en el mismo dataset.

### 4.2. Configuración Experimental

Como ya se dijo, como descriptor local se usó SIFT calculado sobre una grilla regular con un paso en ambas direcciones de 6 pixels y en 5 escalas de la imagen original. Para la reducción de la escala se utilizó un factor de  $\frac{1}{\sqrt{2}}$ . Estos descriptores tienen una dimensionalidad de 128, la cual fue reducida a 64 dimensiones usando PCA. Sobre estos descriptores se ajustó una GMM de 64 componentes, que luego fue usada para calcular los FV. Por último se entrenó un SVM en una configuración OvA.

## 5. Resultados

En la segunda columna del cuadro 1, se muestra la precisión (ecuación 1) del método propuesto, identificado como FV, sobre el dataset Flavia y los resultados de los distintos métodos analizados. La precisión está calculada como el promedio de 10 corridas.

$$\text{precisión} = \frac{\text{número de muestras bien clasificadas}}{\text{número total de muestras}} \quad (1)$$

En la última columna del cuadro 1 se compara la precisión del método propuesto con la del método presentado en [8], sobre el dataset Foliage. La precisión

**Cuadro 1.** Comparación de distintos algoritmos de clasificación en los datasets Flavia y Foliage.

| Método    | Flavia       | Foliage      |
|-----------|--------------|--------------|
| PNN [19]  | 90.31        | -            |
| GLC [10]  | 93.00        | -            |
| SCTF [9]  | 93.75        | -            |
| SC [6]    | 95.47        | -            |
| KDES [12] | 97.00        | -            |
| CS [17]   | 97.00        | -            |
| GLS [8]   | 97.19        | 95.00        |
| ICM [18]  | 97.82        | -            |
| <b>FV</b> | <b>99.22</b> | <b>98.83</b> |

sigue siendo muy buena sobre este dataset, aunque el mismo contiene casi el doble de clases que el anterior.

Como se puede ver en el cuadro de comparación (1), usando FV se obtienen resultados que son estado del arte en los 2 datasets analizados, superando la precisión de los métodos propuestos en la literatura actual.

## 6. Conclusiones

El objetivo principal de este trabajo era mostrar que con un método clásico de clasificación de imágenes y con descriptores robustos ya probados en diferentes problemas de clasificación y reconocimiento se pueden obtener resultados que son estado del arte en el problema de clasificación de especies de hojas. Como se puede ver en la sección 5, con la codificación usando vectores de Fisher, se obtienen resultados de estado del arte en 2 datasets. Además con este método no se necesita realizar preprocesamiento de los datos ni un algoritmo de segmentación o extracción de contorno, aumentando la robustez del mismo. Esta última característica permite que el sistema sea usado en imágenes naturales con ruido de fondo.

## 7. Trabajo a Futuro

A futuro se planea evaluar dos modificaciones al esquema actual. La primera es utilizar el sistema en imágenes con fondo natural en el cual haya mucho ruido de fondo, oclusiones y desorden, para poder corroborar la capacidad de generalización del sistema. La segunda es calcular FV sobre los descriptores visuales que se obtienen como salida de una red neuronal convolucional (CNN) como se plantea en [15], en donde se demuestra que este tipo de descriptor es muy poderoso para diversos problemas relacionados a visión por computadora.

## Referencias

1. Bishop, C.M.: Pattern recognition and machine learning. springer (2006)
2. Chaki, J., Parekh, R., Bhattacharya, S.: Plant leaf recognition using texture and shape features with neural classifiers. *Pattern Recognition Letters* 58, 61–68 (2015)
3. Fan, R.E., Chang, K.W., Hsieh, C.J., Wang, X.R., Lin, C.J.: Liblinear: A library for large linear classification. *The Journal of Machine Learning Research* 9, 1871–1874 (2008)
4. Gosselin, P.H., Murray, N., Jégou, H., Perronnin, F.: Revisiting the fisher vector for fine-grained classification. *Pattern Recognition Letters* 49, 92–98 (2014)
5. Hong, S., Minzan, L., Qin, Z.: Detection system of smart sprayers: Status, challenges, and perspectives. *International Journal of Agricultural and Biological Engineering* 5(3), 10–23 (2012)
6. Hsiao, J.K., Kang, L.W., Chang, C.L., Lin, C.Y.: Comparative study of leaf image recognition with a novel learning-based approach. In: *Science and Information Conference (SAI)*, 2014. pp. 389–393. IEEE (2014)
7. Kadir, A., Nugroho, L., Susanto, A., Santosa, P.: Neural network application on foliage plant identification. *International Journal of Computer Applications* 29(9), 15–22 (2011)
8. Kadir, A.: A model of plant identification system using glcm, lacunarity and shen features. arXiv preprint arXiv:1410.0969 (2014)
9. Kadir, A., Nugroho, L.E., Susanto, A., Santosa, P.I.: Leaf classification using shape, color, and texture features. arXiv preprint arXiv:1401.4447 (2013)
10. Kalyoncu, C., Toygar, Ö.: Geometric leaf classification. *Computer Vision and Image Understanding* (2014)
11. Ken Chatfield, Victor Lempitsky, A.V., Zisserman, A.: The devil is in the details: an evaluation of recent feature encoding methods. In: *Proc. BMVC*. pp. 76.1–76.12 (2011), <http://dx.doi.org/10.5244/C.25.76>
12. Le, T.L., Tran, D.T., Pham, N.H.: Kernel descriptor based plant leaf identification. In: *Image Processing Theory, Tools and Applications (IPTA)*, 2014 4th International Conference on. pp. 1–5. IEEE (2014)
13. Lee, K.B., Hong, K.S.: An implementation of leaf recognition system using leaf vein and shape. *International Journal of Bio-Science and Bio-Technology* 5(2), 57–66 (2013)
14. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *International journal of computer vision* 60(2), 91–110 (2004)
15. Razavian, A.S., Azizpour, H., Sullivan, J., Carlsson, S.: CNN features off-the-shelf: an astounding baseline for recognition. *CoRR* (2014)
16. Sánchez, J., Perronnin, F., Mensink, T., Verbeek, J.J.: Image classification with the fisher vector: Theory and practice. *International Journal of Computer Vision* 105(3), 222–245 (2013)
17. Sfar, A.R., Boujemaa, N., Geman, D.: Confidence sets for fine-grained categorization and plant species identification. *International Journal of Computer Vision* pp. 1–21 (2014)
18. Wang, Z., Sun, X., Ma, Y., Zhang, H., Ma, Y., Xie, W., Zhang, Y.: Plant recognition based on intersecting cortical model. In: *Neural Networks (IJCNN)*, 2014 International Joint Conference on. pp. 975–980. IEEE (2014)
19. Wu, S.G., Bao, F.S., Xu, E.Y., Wang, Y.X., Chang, Y.F., Xiang, Q.L.: A leaf recognition algorithm for plant classification using probabilistic neural network. In: *Signal Processing and Information Technology*, 2007 IEEE International Symposium on. pp. 11–16. IEEE (2007)