Clasificación de texturas por algoritmos de nearest neighbor y random forest

Juan Sebastián Cuéllar
Universidad de los Andes, Bogotá, Colombia
js.cuellar169@uniandes.edu.co

1. Introduction

En este artículo se describen y se evalúan dos métodos de clasificación para solucionar uno de los problemas fundamentales de la visión computacional: extracción y clasificación de descriptores de texturas. Para evaluar el desempeño de los métodos se hará uso del data set de texturas del grupo Ponce introducido en [3], se dividirá en grupos de test y entrenamiento sobre los cuales se emplearán los algoritmos de clasificación y posteriormente se definirá su rendimiento con matrices de confusión.

El data set del grupo Ponce está constituido por 1000 imágenes, 25 categorías de textura y 40 muestras por cada categoría. Todas las imágenes se encuentran en escala de grises, en formato JPG y presentan una resolución de 640 x 480 pixeles [3]. Se incluyeron 10 imágenes aleatorias de cada categoría para conformar la base de test y el resto de las imágenes se tomaron como base de entrenamiento.

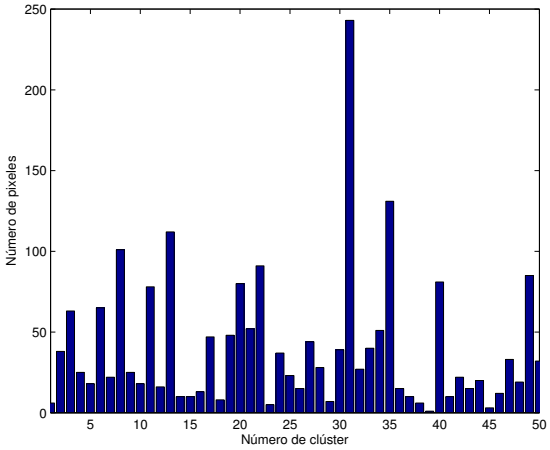

Figure 1. Histograma de los pixeles de la muestra representativa de las texturas dentro de cada clúster o texton.

2. Metodología

2.1. Creación del diccionario de textones

La creación del diccionario de textones inicia con la creación del banco de filtros. El banco de filtros se constituye de un total de 32 filtros dentro de los cuales se incluyen 8 orientaciones, 2 tamaños y 2 simetrías (par o impar). El banco de filtros se corre sobre la concatenación de 25 imágenes representativas de cada categoría (una imagen por categoría) y tras hallar la respuesta de todos los pixeles para cada filtro se reorganizan todos los vectores asociados (en dimensión 32) y se extraen 2000 pixeles aleatorios como muestra representativa de todas las texturas de la base. Estos 2000 pixeles se clusterizan en 50 grupos usando el método de kmeans para encontrar los centroides de la respuesta de los pixeles al banco de filtros, es decir, los textones. El número de textones no debe ser reducido puesto que limita la distinción entre diferentes texturas, es decir, representa dos o más texturas bajo un mismo centroide o texton, pero tampoco muy grande para evitar grandes gastos de memoria computacional y tiempo.

¿Qué filtros discriminan más? Los filtros que tienen una respuesta más fuerte presentan valores altos en los centroides de los clúster que logran agrupar la mayor cantidad de pixeles de la muestra representativa de las texturas. La cantidad de pixeles de la muestra representativa en cada clúster se puede observar en el histograma de la fig 1 y muestra que el clúster 31 es el que más logra discriminar entre texturas. Así mismo, el clúster 31 presenta valores altos de los filtros 10 y 26, los cuales son filtros correspondientes a los patrones verticales en las dos escalas disponibles, sin embargo, se debe tener en cuenta que el clúster es una combinación de todos los filtros.

2.2. Descripción de los métodos de clasificación

Para definir los descriptores de texturas de una imagen primero se corre el banco de filtros sobre la imagen, se reorganiza la imagen en una matriz de vectores de cada pixel con las 32 respuestas, se buscan las distancias de cada pixel con respecto a los textones y a cada pixel se le asigna la etiqueta del texton más cercano resultando en el mapa de textones de la imagen. Se calcula el histograma del mapa de textones y se define el descriptor como la ocurrencia normalizada de cada texton dentro de la imagen, es decir, el

número de pixeles que pertenece a cada texton sobre el total de pixeles.

Tras definir los descriptores de cada imagen se definen los métodos de clasificación que recibirán como entrada dichos descriptores. Para el caso particular de este estudio se usarán los métodos Nearest neighbor (NN) y Random forest.

Nearest neighbor (NN) busca los puntos más cercanos o más similares. En este caso se busca el histograma de la base de entrenamiento más similar a un histograma de entrada, para definir cuantitativamente dicho parámetro de similaridad se utiliza la distancia chi-cuadrado, d (H1, H2)[4].

$$d(H_1, H_2) = \sum_I \frac{(H_1(I) - H_2(I))^2}{H_1(I)} \qquad (1)$$

Donde H_1 es la ocurrencia de la imagen de entrenamiento, H_2 es la ocurrencia de la imagen de entrada o de test e I es el número del texton. El histograma de test se compara con todos los 750 histogramas de entrenamiento y se busca la menor distancia chi-cuadrado de todas las comparaciones. Se busca la imagen de entrenamiento que pertenece a dicha distancia y se asigna su etiqueta a la imagen de test. Teniendo en cuenta que es una sumatoria del error relativo entre los descriptores de los histogramas se espera lograr una discriminación adecuada entre categorías.

Random forest (RF) es un ensamble de varios modelos de predicción del tipo árboles de decisión que utilizan construcciones lógicas para clasificar una serie de condiciones que ocurren sucesivamente [2]. Los random forest son una forma de promediar las respuestas de múltiples árboles de decisión, cada uno entrenado aleatoriamente en diferentes partes de la misma base de entrenamiento con el fin de evitar un sobreajuste de dicho set [1]. Para los propósitos de este trabajo las entradas a clasificar de dichos árboles son el conjunto de atributos representativo de la imagen de test, es decir, el histograma de textones con sus 50 descriptores. Debido a que el algoritmo de random forest exige asignar las etiquetas a cada conjunto de descriptores se incluye una columna adicional con el número de categoría (1-25) al que pertenece cada conjunto.

Cada árbol utiliza la raíz cuadrada del número total de descriptores para ajustarse y entrenarse en base a esta submuestra. Para la muestra de histogramas de 50 descriptores se tiene una submuestra de 8 descriptores que se eligen aleatoriamente para cada árbol. Es claro el interés por incluir todos los descriptores al menos una vez en el bosque, por tal motivo se utiliza un número de 200 árboles y se ejecuta la opción que permite calcular el error de clasificación 'fuera de la bolsa', que permite ver el error asociado a dejar por fuera de la decisión algunos descriptores como consecuencia de utilizar pocos árboles de decisión. Como se

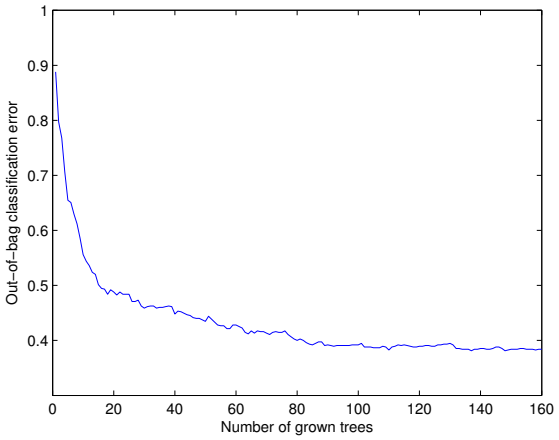

Figure 2. Error 'fuera de la bolsa' del random forest creado a partir de los descriptores de la base de entrenamiento.

muestra en la fig. 2 después de utilizar 160 árboles el error no muestra variación apreciable, por esta razón se considera prudente utilizar 160 árboles de decisión para clasificar todas las imágenes de la base de test. La profundidad de cada árbol juega un papel importante en el desempeño del RF porque establece la especificidad hacia los descriptores de entrenamiento y finalmente define la sobrespecialización a estos, de modo que para evitar dicha sobrespecialización se reducen los nodos de todos los árboles a 41. Con la asignación de este parámetro se permite clasificar imágenes diferentes al set usado en este estudio.

El desempeño de los métodos se puede evaluar usando una matriz de confusión, donde se pueden observar los porcentajes de las imágenes de test que han sido clasificadas en cada categoría. Las casillas de la diagonal son los aciertos de los algoritmos.

3. Results

Como se puede observar en la fig. 3 el método de clasificación que tiene mayor exactitud global en la base de test es el nearest neighbor con un 68%, es decir que hay mayor probabilidad de clasificar cualquiera de las imágenes en la categoría correcta usando dicho algoritmo. Se puede observar también que la clasificación de la base de entrenamiento tiene puntaje perfecto en NN, resultado esperado puesto que la menor distancia se encuentra con la misma imagen, pero exactitud de 78.93% en RF. Esto se esperaba puesto que los nodos de los árboles se redujeron para evitar la sobrespecialización de los descriptores de entrenamiento aceptando el riesgo de reducir la exactitud global de clasificación.

La categoría que causa más confusión es la categoría 23 donde se observan los menores porcentajes de aciertos (RF-acc = 10; NN-acc = 30). Esto indica que con el diccionario

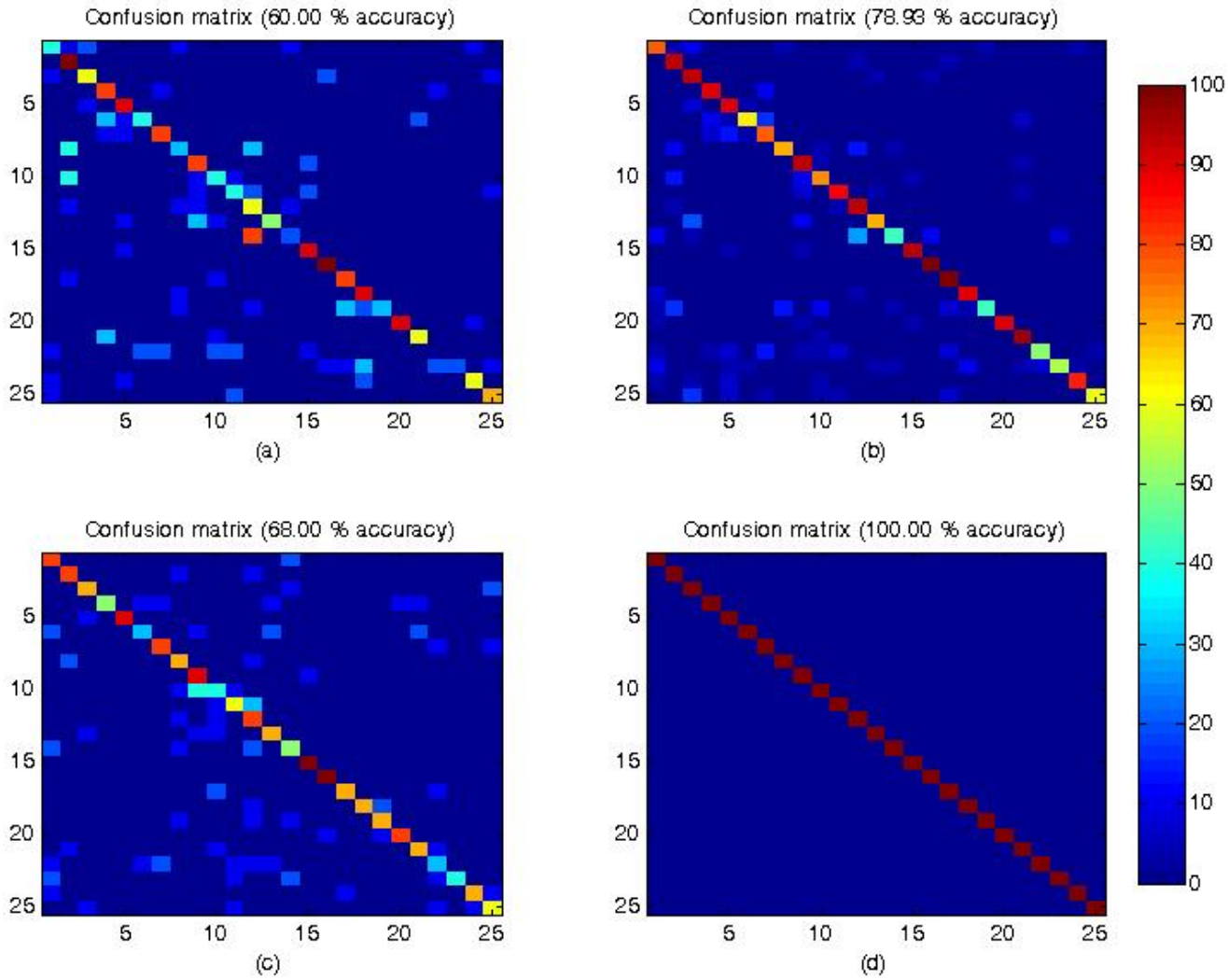

Figure 3. Matrices de confusión para (a) Random forest sobre la base de test (b) Random forest sobre la base de entrenamiento (c) Nearest neighbor sobre la base de test y (d) Nearest neighbor sobre la base de entrenamiento. Arriba de cada matriz se muestran los promedios de aciertos de la diagonal, valores en porcentajes.

de textones usado el clasificador no puede discriminar los descriptores de las imágenes de la categoría 23 y las confunde con los descriptores de la categoría 1, 18, 22 y 25 usando RF, y 1, 11, 14, 17 y 22 usando NN.

3.1. Tiempos de ejecución

Los tiempos empleados por los algoritmos se muestran acontinuación

Para Random forest se utilizó un tiempo de entrenamiento para los árboles de decisión de 4.55 segundos y un tiempo de clasificación de la base de test de 76.21 segundos.

Para Nearest neighbor se utilizó un tiempo de clasificación de la base de test de 0.57 segundos.

La diferencia en tiempos de procesamiento entre los dos métodos es muy apreciable y evidencia la rapidez del método de NN en comparación con el RF. Aunque la diferencia entre los métodos para bases de imágenes mucho más grandes puede ser altamente significativa, la elección del método depende de la base de test a clasificar y por ello se necesita una evaluación posterior, es preferible emplear la cantidad de tiempo requerido para clasificar las imágenes por los dos métodos en cada base de test a evaluar y seleccionar según el desempeño.

4. Discusión

En conjunto, y por más bueno que se configuren los métodos siempre existe la probabilidad de que se califique

mal alguna categoría. Debido a que los descriptores utilizados (textones) representan cambios de patrones repetitivos en diferentes orientaciones, la clasificación de cierta forma tiende a clasificar en la misma categoría las texturas que muestran patrones alineados bajo el mismo ángulo, por ejemplo, en ciertas ocasiones se clasifica en el mismo grupo dos texturas que tienen patrones diagonales pero que pertenecen a diferentes categorías. Este fenómeno afecta principalmente el método de NN porque la distancia chicuadrado pondera todos los descriptores al mismo tiempo mientras que en RF lo hace en varios subsets para cada árbol. Además se debe tener en cuenta que en este estudio no se utilizaron filtros circulares lo que pudo limitar una correcta clasificación de texturas con patrones redondos.

Por otro lado es probable que el número de pixeles de las respuestas a los filtros (2000) tomado aleatoriamente haya sido insuficiente para conformar un tamaño de muestras representativo de todas las texturas y por ende se halla clasificado el diccionario de forma incongruente con la dispersión total de los pixeles utilizados.

Adicionalmente, en el método RF se debe elegir el número de nodos de los árboles teniendo en cuenta que existe una compensación entre la exactitud y la sobrespecialización de la clasificación, lo que en aplicaciones prácticas es difícil de definir. Finalmente es importante elegir tan pocos árboles de decisión como sea posible porque el algoritmo consume mucha memoria computacional.

La base de datos solo permite obtener un diccionario de textones representativo para 25 texturas lo que impediría un diccionario de textones práctico en la vida real. Además las fotografías utilizadas como entrenamiento están tomadas en un ámbito profesional para este tipo de usos y no son representaciones aleatorias de texturas presentes en fotografías comunes, esto posiblemente evitaría una correcta clasificación al ingresar fotografías del día a día a los algoritmos. Por otro lado, toda la base de datos se limita a presentar texturas sin variaciones de iluminación severas y esto impide evaluar los métodos de forma completa.

Adicionalmente se cuenta con una base de test un poco limitada puesto que se quisiera contar con muchas imágenes que incluyeran las texturas evaluadas en distintas presentaciones, tamaños y como parte de objetos o conjunto de objetos.

Los métodos solo usan patrones repetitivos de líneas como descriptores. Para mejorar la clasificación de este set de imágenes en particular se propone utilizar más filtros dentro del banco de filtros, recordando que en este estudio no se utilizaron filtros circulares, y a partir de los nuevos filtros definir una mayor cantidad de textones que logren discriminar mejor entre texturas.

Para mejorar la clasificación en general se puede ampliar el espacio de representación utilizando descriptores de color y de posiciones relativas entre pixeles, además de métodos que ponderen mayores pesos a ciertos descriptores que incluyan información con mayor poder de discriminación entre texturas.

References

[1] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An Introduction to Statistical Learning*. Springer, 2013.

[2] Lior Rokach and Oded Maimon. *Data mining with decision trees: theory and applications*. World Scientific, 2008.

[3] Svetlana Lazebnik, Cordelia Schmid, and Jean Ponce. A Sparse Texture Representation Using Local Affine Regions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8):1265–1278, Aug. 2005.

[4] Y. Rubner, C. Tomasi, and L.J. Guibas. The Earth Mover's Distance as a Metric for Image Retrieval. *Technical Report STAN-CS-TN-98-86, Department of Computer Science, Stanford University*, Sept. 1998.