

Clasificación Supervisada y No-Supervisada de Imágenes

Santiago Villalba
Universidad de los Andes
Cra. 1 18a 12, Bogotá, Colombia
sh.villalba@uniandes.edu.co

Nicolás Garnica
Universidad de los Andes
Cra. 1 18a 12, Bogotá, Colombia
pn.garnicar@uniandes.edu.co

1. Introducción

La clasificación de imágenes es el procedimiento de definir límites precisos entre las intensidades de unos píxeles de otros píxeles, siendo esto la acción de extraer las clases de la imagen. La clasificación se puede dividir en dos tipos, Clasificación supervisada y Clasificación no supervisada[1]. La primera, es aquella en la que se tiene un conocimiento previo de lo que se quiere agrupar o ya se tiene los modelos clasificados de los objetos con características comunes, esta se puede dividir en dos fases, en la fase inicial se posee el conjunto de entrenamiento y test o validación, y en la fase final es el proceso de clasificar los objetos de los cuales no se conoce la clase. La segunda, es aquella en la cual no tenemos conocimiento previo de lo que se quiere clasificar por lo que el objetivo es obtener los diferentes grupos de objetos cuyas características nos permitan separar las diferentes clases[2].

De igual manera, para este informe se tiene como objetivo realizar la clasificación de las escenas con un método no supervisado. Se utilizarán descriptores de color de las imágenes y k-means como clasificador y se realizará una clasificación con k-means según la definición de clusters. Así mismo, se posee un dataset para el desarrollo de este proyecto con tres carpetas, test, la cual posee 12 imágenes, train, la cual posee 48 imágenes y validación, con 12 elementos, todas las imágenes con dimensiones 150x150 píxeles y en espacio de color RGB.

2. Metodología - Histogramas de Color y K-means

2.1. Histograma conjunto de color

El primer descriptor que utilizamos fueron los histogramas conjuntos de color. Estos consisten en obtener la información de los tres canales de color, de forma que se pueda obtener un histograma donde se represente la distribución de cada triada de valores de intensidad sobre la imagen. Para obtenerlos, nos basamos en los 3 canales que contienen información de una imagen a color independientemente de

su espacio de color correspondiente. En la Figura 1. se presentan los pasos para obtener el histograma conjunto de color

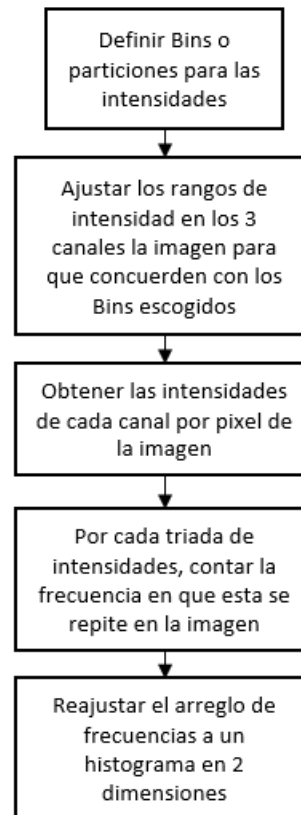


Figure 1. Procedimiento para obtener el histograma conjunto de color.

2.2. Histograma concatenado

Este descriptor es útil para la descripción de las imágenes a color ya que contiene la información de cada píxel y la frecuencia con la que dicho píxel se repite en la imagen, lo que puede ayudar a establecer similitudes o diferencias entre imágenes de acuerdo únicamente a la distribución de las

intensidades por triadas. Algunas de las ventajas de usar estos histogramas es que son invariables a modificación en tamaño, rotaciones o traslaciones de la imagen, mientras que una de sus mayores desventajas es ignorar la información espacial de la distribución de las intensidades.

Posteriormente utilizamos los histogramas concatenados como segundo descriptor de color de las imágenes. Para obtener estos histogramas, se utiliza la información de los tres canales de color. Inicialmente se especifica el número de bins que deberá tener cada canal, luego se crea un arreglo que tendrá el largo de la sumatoria de los 3 bins especificados. Posteriormente se modifican las intensidades de la imagen para que concuerden con los bins correspondiente de acuerdo a los rangos que fueron determinados para cada canal. Finalmente se rellena el arreglo vacío con los valores de la imagen de los 3 canales en orden y se halla el histograma de dicho arreglo con los valores modificados; de manera que los valores de un canal no se confundan ni modifiquen la información de frecuencia de las intensidades de los otros canales. Este histograma es útil para diferenciar imágenes no solo por la frecuencia de los píxeles sino en las diferencias que existen en los tres canales de la imagen, es decir, que se puede tener una percepción más amplia de los 3 canales de la imagen, donde uno solo de estos puede tener la variación necesaria para diferenciar el descriptor de color de una imagen, de otra, únicamente utilizando el histograma concatenado.

2.3. Agrupamiento por K-Means

El método de clasificación o agrupamiento por K-Means, está basado en la clasificación no supervisada, donde el algoritmo de clasificación no necesita de etiquetas para poder realizar una predicción. Con la metodología de K-Means, se asignan etiquetas a los datos de entrenamiento y son estas etiquetas las que el algoritmo utilizará posteriormente para realizar la predicción. El algoritmo de agrupamiento consiste en formar grupos de acuerdo a la similitud de los datos. Entre más parecidos sean estos se formarán grupos, en los cuales dentro de estos grupos dichos datos son más similares entre sí, en comparación con los datos de los otros grupos. Para esto se inicializan unos centros de grupos llamados centroides, al azar, que serán el punto de partida para la comparación de las distancias entre los datos a estos puntos para formar los grupos. Posteriormente se realizan iteraciones en donde los centroides son escogidos de nuevo aleatoriamente sin tener en cuenta los que ya se usaron, y se vuelven a formar los grupos. El algoritmo realizará estas iteraciones hasta que la sumatoria de las distancias al cuadrado de los centroides y los elementos de su grupo correspondiente sean la mínima, por lo tanto el algoritmo habrá convergido y se tendrán tantos grupos como centroides. La cantidad de centroides será elegida de acuerdo a la cantidad de clases que se deseen separar.[3]

3. Experimentos - Histogramas de Color y K-means

Los parámetros que modificamos para nuestros experimentos, fueron los *bins* de los histogramas, el espacio de color en el que estaría representada la imagen y la cantidad de clusters o centroides que se utilizarían para el entrenamiento y clasificación con ayuda de la función de K-Means. Para nuestros experimentos, decidimos utilizar únicamente los espacios de color RGB y HSV ya que la visualización por medio de estos espacios de color generaba una mayor diferencia entre las imágenes de manera visual. En cuanto al número de *bins* utilizados, decidimos utilizar 150 y 10 *bins* ya que con 150 *bins* aun se conserva gran parte de la variedad de intensidades posibles para los canales de la imagen, evitando pérdida sustancial de información, sin tener que procesar las 250 intensidades posibles, ahorrando tiempo de procesamiento. Y utilizamos 10 *bins*, por que sería una representación de la distribución del color muy resumida que nos ayudaría en la comparación con otras imágenes y su diferenciación en caso de que las diferencias en intensidad fueran demasiado pronunciadas, dándonos un indicativo de la diferencia en el color de las imágenes teniendo en cuenta la clase a la cual pertenecen. Finalmente, el parámetro de centroides k fue fijado en 6 ya que eran 6 clases las que debíamos identificar, sin embargo, para mejorar probar el desempeño del algoritmo disminuimos el valor de k para ver si con menor cantidad de clases el desempeño de clasificación era mejor.

3.1. Resultados

En la Figura 2. se pueden observar los diferentes experimentos que realizamos, teniendo en cuenta los parámetros y sus variaciones utilizadas.

Teniendo en cuenta los resultados, el mejor desempeño del algoritmo nos dio cuando utilizamos el descriptor del histograma conjunto de color, con las imágenes en el espacio de color RGB y con un número de 3 centroides (k). Observando los resultados que obtuvimos, podemos notar que el espacio de color HSV tiene una medida F1 promedio mayor que el espacio de color RGB al utilizar el descriptor de color del histograma concatenado. Por lo que podemos identificar que tanto el descriptor como el espacio de color son posibles factores de variabilidad en los resultados. Podemos notar también por el promedio de las medidas F1, que el descriptor que en promedio dio mejores resultados fue el histograma conjunto, independientemente del espacio de color. En cuanto a las variaciones en el parámetro k (Clusters), podemos notar que este parámetro no genera una diferencia significativa en los resultados, por lo que el desempeño de la clasificación estaría siendo afectado en su mayoría por el descriptor utilizado y en una menor proporción por el espacio de color.

Experimentos						
Descriptor de color	Espacio de color	Bins del histograma	Número de clusters	Presición	Cobertura	F1
Histograma conjunto	RGB	150	6	0.5	0.45	0.47
Histograma conjunto	RGB	10	6	0.36	0.26	0.28
Histograma conjunto	HSV	150	6	0.5	0.45	0.47
Histograma conjunto	HSV	10	6	0.33	0.25	0.28
Histograma concatenado	RGB	150	6	0.19	0.179	0.184
Histograma concatenado	RGB	10	6	0.11	0.1	0.107
Histograma concatenado	HSV	150	6	0.11	0.083	0.095
Histograma concatenado	HSV	10	6	0.6	0.39	0.46
Histograma conjunto	RGB	150	5	0.5	0.3	0.38
Histograma conjunto	RGB	150	4	0.3	0.45	0.41
Histograma conjunto	RGB	150	3	0.5	0.45	0.47

Figure 2. Experimentos de clasificación, modificando los parametros de los descriptores y del clasificador K-Means.

4. Discusión - Histogramas de Color y K-means

Consideramos que la diferencia se debe a que el histograma conjunto diferenciaba de mejor manera los histogramas obtenidos ya que los píxeles de las imágenes de una misma clase debían de ser muy parecidos entre cada clase, esta similitud puede ser capturada al cuantificar las frecuencias de las triadas de valores de intensidad que conforman cada pixel, lo que generaba que las respuestas fueran mas diferenciables en comparación con el histograma concatenado, es decir, la combinación de los canales generaba una respuesta mas diferenciable en comparación con la separación de los canales.

4.1. Mejor resultado

Escogimos el mejor resultado al histograma conjunto de color en el espacio de color RGB con 3 clusters, ya que al tener la imagen inicialmente en RGB se evitaba tiempo adicional en el procesamiento para convertirla al espacio HSV, además de que al utilizar unicamente 3 clusters se ahorra tiempo de procesado y se obtenían los mismos resultados que los mostrados en el primer y tercer experimento.

4.2. Color

En cuanto al color, podríamos decir que si es un buen descriptor de la imagen mientras cada clase tenga tonalidades muy distintas entre si. El problema está cuando las imágenes de varias clases tienen colores muy similares, y dado que los descriptores de color utilizados no contemplan la distribución espacial del color, este descriptor no sería muy bueno para el entrenamiento del algoritmo, y creemos que esa es una de las razones del bajo desempeño de la clasificación. Esto podría mejorarse si se implementaran descriptores de color regionales de la imagen, lo que podría

ayudar a mejorar la diferenciación entre las clases.

4.3. Desempeño de K-Means

Según los resultados que obtuvimos con K-Means, pensamos que no es un método óptimo para realizar la clasificación de las imágenes, específicamente por que no tiene en cuenta las etiquetas que estas tienen asignadas. Debido a esto el método de agrupación asignaba etiquetas que no correspondían con las originales, haciendo que de una sola clase, hubieran etiquetas de otras clases, por lo tanto para asignarle una única etiqueta a las imágenes pertenecientes a una clase, decidimos sacar la moda de las etiquetas asignadas por K-Means a una clase determinada, teniendo en cuenta el parámetro k . Creemos que esto es el resultado de utilizar hiperplanos en la separación por diagramas de *Voronoi* en el algoritmo de K-Means, ya que los datos podrían haber sido mejor separados en grupos por elementos mas dinámicos que por hiperplanos.

4.4. Posibles mejoras

Teniendo en cuenta que la comparación de histogramas puede hacerse por medio de distancias euclídeas, podríamos utilizar el espacio de color Lab que es perceptualmente uniforme donde sus distancias medidas son también distancias euclídeas. También podríamos implementar un método de clasificación supervisado como SVM para mejorar la clasificación de las imágenes.

5. Textones y SVM

Una vez realizadas las funciones (I), (II) Y (III) de la sección 7.1.1 e incluirlas dentro del archivo main.py, fue posible realizar una serie de experimentos donde se utilizaron dos tamaños de clusters diferentes (50 y 100) y dos

diferentes tipos de Kernels (Polinomio grado 2 y Generalized Histogram Intersection). Los resultados se muestran en la siguiente tabla:

Experimentos	Clusters	kernel	Precision	Cobertura	F1
1	100	Polinomio grado 2	0,027	0,16	0,04620321
2	50	Polinomio grado 2	0,11	0,25	0,15277778
3	100	Generalized Histogram Intersection	0,77	0,66	0,71076923
4	50	Generalized Histogram Intersection	0,88	0,83	0,85426901

Table 1. Resultados experimentos Textones-SVM.

Los resultados demuestran que al utilizar una menor cantidad de clusters, se obtienen mejores resultados para Precision, Recall y F1- score. Sin embargo, al utilizar el Kernel Generalized Histogram Intersection se obtienen unos resultados muy cercanos a la idealidad de la clasificación de imágenes (precisión=1 y recall=1). Esto se debe a que el Kernel Generalized Histogram Intersection nos permite un mejor ajuste del modelo del Kernel por medio de la comparación de histogramas [4].

6. Descriptor HOG y Random Forest

Una de las mayores desventajas de utilizar Random Forest como clasificador es que esta metodología tiene muchos parámetros a controlar, y con cada modificación pequeña a estos, los resultados cambian drásticamente. La Tabla 2 muestra los parámetros que variamos para hacer nuestros experimentos, encontrando que la mejor metodología fue la de descriptores HOG con 5 orientaciones y distancia L1, y el clasificador RF con 100 estimadores y una profundidad de 4.

Experimento	Clasificador RF		Descriptor HOG		Precision	Cobertura	F1
	# Estimadores	Profundidad	Norma	Orientaciones de HOG			
1	100	4	L1	5	0,75	0,66	0,67
2	100	8	L2	10	0,19	0,33	0,24
3	100	4	L1	5	0,48	0,5	0,46
4	200	8	L1	10	0,45	0,5	0,43
5	200	8	L2	5	0,401	0,5	0,41
6	100	4	L2	10	0,43	0,5	0,433

Table 2. Resultados experimentos HOG y random forest.

Teniendo en cuenta estos resultados, concluimos que la forma no es un buen descriptor para la clasificación de estas imágenes, ya que la forma de las estructuras que estas contienen pueden ser muy similares, como lo serían un glaciar y una montaña o una jungla y una ciudad.

7. Experimentos con clasificadores

En la experimentación con los clasificadores, encontramos que el mejor resultado nos dio con SVM y textones, como se puede apreciar en la Tabla 3. Sin embargo se tuvo también un resultado alto con el descriptor de forma HOG y K-Means. Esto puede deberse a que varias de las imágenes de las distintas clases, presentan morfologías similares entre ellas como fue descrito en la sección anterior. Al disminuir la cantidad de clusters a 3 en el algoritmo de K-Means, este pudo agrupar de una mejor manera las imágenes que tenían formas similares entre sí.

Descriptor	Clasificador	Precision	Cobertura	F1
Histograma concatenado	K-Means	0.5	0.45	0.47
Histograma concatenado	SVM	0.47	0.58	0.521
Histograma concatenado	RF	0.27	0.25	0.26
Textones	K-Means	0.44	0.38	0.41
Textones	SVM	0.88	0.83	0.85
Textones	RF	0.72	0.58	0.64
HOG	K-Means	0.83	0.83	0.83
HOG	SVM	0.5	0.5	0.5
HOG	RF	0.75	0.66	0.7

Table 3. Resultados experimentos cambiando los clasificadores con los tres descriptores con los parámetros óptimos.

Concluimos que el mejor descriptor para la clasificación de las imágenes es el de Textones ya que es posible diferenciar las imágenes más por su textura que por las demás características que estas posean.

8. Ensayo en base de datos Test

Los resultados del análisis con nuestro mejor modelo en la base de datos de train no fue tan bueno como esperábamos, a pesar de que las imágenes con respecto a la base de datos de validación no son sustancialmente diferentes, es posible que las texturas de la imagen cambiaran drásticamente en algunas imágenes, como las imágenes de ciudades, donde se tenían imágenes más alejadas o que creemos que causa una diferencia de texturas con la clase de edificios entreteñada. Para resolver este problema, se podrían utilizar distintos tamaños de las imágenes para poder entrenar mejor el algoritmo.

```

Mejor modelo
Modelo: Textones + SVM
Parametros SVM: K = 50, Kernel = GeneralizedHistogramIntersection
Convenciones: 1-Building, 2-Forest, 3-Glacier, 4-Mountains, 5-Sea, 6-Street
Ground truth: 1,1,2,2,3,3,4,4,5,5,6,6
Predicciones: [1 3 2 2 3 5 5 4 4 2 2]
Métricas: Precision = 0.3333333333333333, Cobertura = 0.3333333333333333, F1 = 0.3333333333333333

```

Table 4. Resultados de análisis de las imágenes de test.

Como mejoras, podríamos utilizar un método combinado

que incluyera tanto el color como la textura para tener dos criterios de clasificacion ya que a simple vista, por el color se pueden diferenciar algunas de las clases. La inclusion de las textura ayudaria a diferenciar aquellas clases con tonalidades similares como los glaciares y el oceano o los bosques y las montañas.

References

- [1] A. HELP, “¿qué es la clasificación de imagen?” 2016. [Online]. Available: <https://desktop.arcgis.com/es/arcmap/10.3/guide-books/extensions/spatial-analyst/image-classification/what-is-image-classification-.htm>
- [2] A. G. de Tecnologías Avanzadas en Computación, “Clasificación supervisada y no supervisada,” 2008. [Online]. Available: <https://advancedtech.wordpress.com/2008/04/14/clasificacion-supervisada-y-no-supervisada/>
- [3] D. Pham, S. Dimov, and C. Nguyen, “Selection of k in k-means clustering,” 2005. [Online]. Available: <https://journals.sagepub.com/doi/abs/10.1243/095440605x8298>
- [4] S. Boughorbel, J.-P. Tarel, and N. Boujemaa, “Generalized histogram intersection kernel for image recognition,” *Proc. IEEE Int. Conf. Image Process.*, vol. 3, pp. III – 161, 10 2005.

Realizado en L^AT_EX