

Automatic Landform Clasification of Uplands Based on Haralick's Texture

Diego Patiño
Escuela de Sistemas
Facultad de Minas
Universidad Nacional de Colombia
Medellín, Antioquia
Email: dapatinoco@unal.edu.co

Domingo Mery
Departamento de
Ciencias de la Computación
Universidad Católica de Chile
Santiago de Chile
Email: dmery@ing.puc.cl

Verónica Botero Fernandez
Escuela de Geociencias y
Medio Ambiente
Facultad de Minas
Universidad Nacional de Colombia
Medellín, Antioquia
Email: vbotero@unal.edu.co

John Willian Branch
Departamento de Ciencias de la Computación
y toma de decisiones
Facultad de Minas
Universidad Nacional de Colombia
Medellín, Antioquia
Email: jwbranch@unal.edu.co

Abstract—This paper presents the results of applying Haralick's textures in the upland landform classification process. In environmental and landuse projects is important to know what kind of landforms are present in a geographic area. The objective of this work is to improve the performance of such classifications, adding information to the commonly used data features in such problems. The texture information was extracted using the method of Haralick, parameterized with moving windows of geo-referenced maps in *raster* format. Those maps contains information of the terrain morphology such as elevation, slopes, etc. Several tests were performed using different classifiers and cross-validation over a dataset with a total of 203401 samples. It was shown that Haralick's textures features are useful for the problem, because its performance is higher (about 97 %) in comparison with the achieved when using only morphological information of the terrain.

Keywords—Haralick's texture, Landform classification, Co-occurrence matrix, Feature selection, Cross-validation.

I. INTRODUCCIÓN

La clasificación automática de tipos de relieve surge de la necesidad de conocer la forma de un territorio evitando recorrerlo por completo; ya sea por ser de gran tamaño, que requiera un esfuerzo considerable o que sea de difícil o imposible acceso [1], [2]. Este tipo de clasificaciones consituyen un insumo importante para procesos de toma de decisiones en áreas ambientales o de ordenamiento territorial; además, sirven como información base para diversos modelos hidrológicos y geológicos [3], [4].

En este trabajo se presenta un enfoque de clasificación de tipos de relieve, específicamente aquellos que corresponden a zonas de alta, media y baja montaña. Las dos clases objetivo

del problema de clasificación planteado son: cañones (valles de quebradas y ríos jóvenes) y vertientes (Laderas de montañas que contribuyen al cauce de los ríos y quebradas).

El reconocimiento de patrones es el proceso a través del cual se construye un modelo que aproxima la forma en que los individuos de un conjunto de datos son asignados a clases (patrones) previamente definidas o desconocidas. La clasificación de patrones es la etapa final del proceso de reconocimiento y consiste en asignar una etiqueta a un vector n -dimensional donde cada dimensión representa una característica (descriptor o *feature*) de la información que se está analizando, y cada etiqueta representa una de las clases.

Esta investigación propone un mecanismo para encontrar un conjunto de descriptores que ofrezcan un considerable grado de separabilidad de las clases objetivo haciendo uso de información sobre la textura del terreno. Una textura de terreno puede ser definida como el patrón de comportamiento que se percibe de la forma de una región geográfica.

Esta información es calculada a partir del análisis de diferentes mapas georeferenciados de variables geomorfológicas en formato *raster*; e.g: MDE (Modelo Digital de Elevación), pendientes, curvaturas, etc. El conjunto de datos para la clasificación propuesta se construye entonces como la unión entre las características geomorfológicas y las de características de textura derivadas de ellas.

Una vez consolidados los datos, se seleccionaron aquellos descriptores que ofrecían la mejor separabilidad entre las clases. Para ello fue utilizado el algoritmo de selección de características SFS [5] con diferentes conjuntos de parámetros. Posteriormente se utilizaron varios clasificadores conocidos con el fin de encontrar aquel que obtiene el mejor desempeño.

La principal contribución con el desarrollo de este trabajo es la consolidación de un conjunto de descriptores apto para el problema de la clasificación de tipos de relieve; de manera que el desempeño obtenido con los descriptores seleccionados sea considerablemente mayor que los reportados en la literatura en aproximaciones de solución al mismo problema.

En la sección II se discute la problemática asociada a la clasificación de tipos de relieve, y se dan algunos ejemplos de diferentes enfoques utilizados en la literatura para estudiar el problema. En la sección III se define el conjunto de datos utilizado. En la sección IV se presentan los detalles del proceso de extracción de características a partir de la metodología de análisis de texturas. En la sección V se ofrecen algunos detalles del proceso de clasificación y de pruebas. En la sección VI se exponen los resultados obtenidos con la metodología propuesta, y se compara su desempeño con una clasificación utilizando solo los descriptores geomorfológicos. Finalmente, en la sección VII se presentan algunas conclusiones y se sugieren algunos posibles focos de trabajo futuro.

II. CLASIFICACIÓN DE TIPOS DE RELIEVE

La clasificación automática de tipos de relieve es un tema de interés en la comunidad científica desde hace alrededor de 15 años, como lo demuestran los estudios de Irvin et al. [6], MacMillan et al. [7] y Burrough et al. [2]. Estas clasificaciones se realizan a partir de información que describe las características del terreno, como su forma, geometría, propiedades del suelo, apariencia, etc.

Conocer los tipos de relieve presentes en un territorio es un problema de especial interés, puesto que este tipo de información constituye un insumo importante para diversos modelos de toma de decisión en proyectos de índole ambiental e ingenieril, e.g proyectos de ordenamiento territorial; cálculos de amenazas por deslizamientos de tierra, inundaciones; modelos hidrológicos e hidroclimáticos, entre otros [3].

En la comunidad científica existen numerosos trabajos que hacen referencia a este tipo de clasificaciones con diversos enfoques de reconocimiento de patrones para diferentes clases objetivo y en varias regiones geográficas. Una gran cantidad de estos trabajos sobresalen por el uso de técnicas no supervisadas para resolver el problema; empleando algoritmos tales como K-Means [8], ISODATA [9], fuzzy-K-Means [10] y redes neuronales auto-organizativas SOM [11], [12], [2], [13], [6]. Es importante resaltar que no existe una definición estándar de tipos de relieve, por lo tanto, cada clasificación lleva consigo un componente de subjetividad inherente a la geografía del caso de estudio para el que fue desarrollado.

En otros estudios sobre el tema abunda principalmente el uso de lógica difusa en conjunción con reglas heurísticas estrechamente ligadas a conocimiento experto. Entre ellos se encuentra el de Arell et al. [13] y el trabajo de Dragut et al. [14]. Este último realiza una clasificación basada en objetos con características geométricas del terreno como la concavidad y la orientación. De esa manera definen nueve clases, que son el resultado de las posibles combinaciones de los valores de curvatura plana y perfil de curvatura.

En la literatura, los autores que investigan el problema suelen presentar siempre un conjunto de descriptores común a todos los estudios. Entre ellos se encuentran algunos con información derivada de modelos digitales del terreno, tales como: mapas de elevación, mapas de pendientes, mapas de curvaturas, etc. Además de esta información, también varias investigaciones en donde se estudia la generación de nuevos descriptores que tal que estos ofrezcan mejores resultados al momento de realizar la clasificación. Para dar algunos ejemplos se pueden relacionar los trabajos de Acciani et al. [11] y Ardiansyah et al. [15] y Renno et al. [16].

Sin embargo, los descriptores anteriormente mencionados no ofrecen una separabilidad satisfactoria de las clases planteadas; de manera tal que es necesario explorar alguna otra técnica que permita extraer algunos descriptores nuevos más adecuados para la clasificación de cañones y vertientes en zonas de alta, media y baja montaña. La figura 1 muestra los histogramas de algunos de estos descriptores comunes en la literatura para cada una de las 2 clases, en donde se aprecia la baja separabilidad que estos ofrecen.

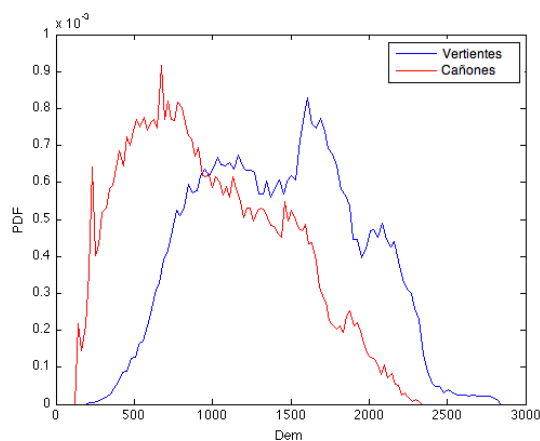
Cuando la clasificación de tipos de relieve se realiza de manera manual, un experto geomorfológico hace uso de su conocimiento a través de la observación de las características de la forma del terreno para llevar a cabo dicha clasificación. Teniendo esto en cuenta, se propone la extracción de nuevos descriptores útiles para el problema aquí presentado, que estén basados en información de las texturas morfológicas; análogamente a como se realiza en el caso manual.

III. CASO DE ESTUDIO

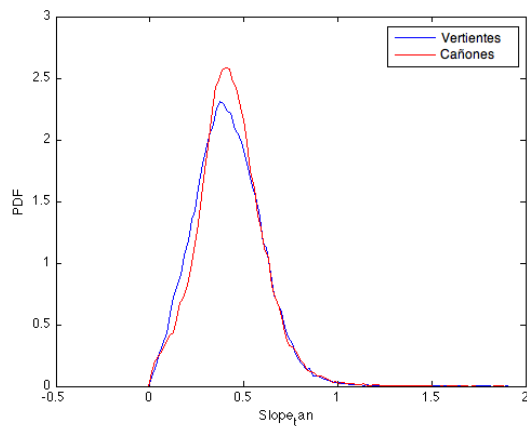
La información fuente para los experimentos de este trabajo corresponde a una sección del modelo de elevación digital de la cuenca hidrográfica del río Sinú. Esta cuenca está ubicada en Colombia entre los departamentos de Antioquia y Córdoba y cuenta con un área de alrededor de 13.700 km^2 . El río Sinú discurre de sur a norte, comenzado su recorrido en el “Nudo del Paramillo” (Antioquia) a 3.000 msnm en la zona alta montaña de la Cordillera de lo Andes central [17].

Pese a que existen diferentes tipos de relieve en la geografía de la cuenca, la zona de estudios fue elegida de manera que las geoformas predominantes fueran dos: cañones y vertientes; las cuales son las clases objetivo del problema de clasificación planteado. La figura 2 muestra la cuenca del río Sinú por completo (izquierda), y resalta el lugar donde se ubica la zona de estudio con su respectiva clasificación (Azul: Vertientes, Rojo: Cañones).

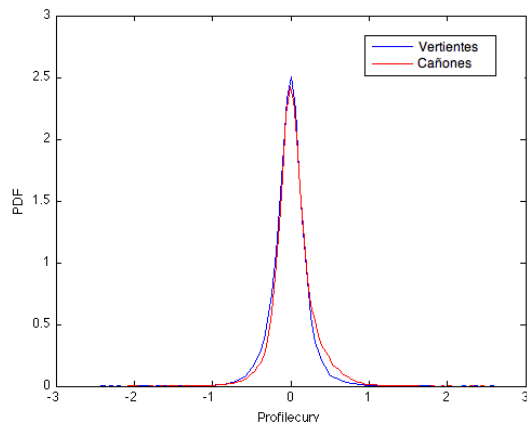
El modelo de elevación de la zona de estudio consta de siete mapas Raster geográficamente referenciados que corresponden a las diferentes características morfológicas del terreno. Cada mapa tiene un tamaño de 3.207 filas por 1.751 columnas que abarcan un poco más del área de toda la cuenca del Sinú. Cada celda de estos mapas representa un área de 92 m^2 . Los siete mapas con una pequeña descripción de cada uno se resumen en la tabla I. El conjunto de datos elegido para la clasificación, que solo es una porción del modelo de elevación digital, cuenta



a) Histograma de Elevaciones en msnm.



b) Histograma de los valores de la tangente de la pendiente.



c) Valores de perfil de curvatura.

Figura 1. Histogramas para tres de los descriptores morfológicos.

entonces con un total de $n = 203,401$ muestras y $m = 7$ variables.

IV. EXTRACCIÓN DE CARACTERÍSTICAS USANDO INFORMACIÓN DE TEXTURAS

La hipótesis fundamental de este trabajo es que se puede mejorar el desempeño de la clasificación de tipos de relieve,

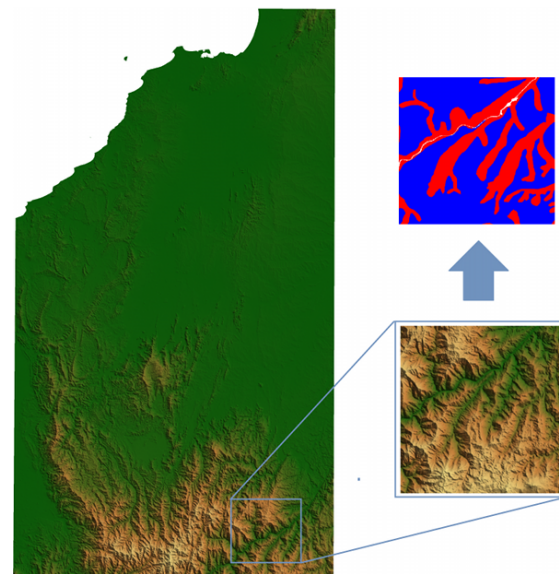


Figura 2. Cuenca del río Sinú.

Cuadro I
MAPAS DEL MODELO DE ELEVACIÓN DE LA CUENCA DEL RÍO SINÚ

f	Nombre	Unidades	Descripción
1	Elevaciones	msnm	Alturas en metros sobre el nivel del mar.
2	Perfil de curvatura	NA	Medida de la convexidad, concavidad o planaridad de una región en la dirección de máxima pendiente. Se corresponde con la segunda derivada.
3	Curvatura plana	NA	Medida de la convexidad, concavidad o planaridad de una región en la dirección perpendicular a la de máxima pendiente. Se corresponde con la segunda derivada.
4	Pendientes en grados	Grados °	Inclinación máxima del terreno en un punto.
5	Tangente de las pendientes	NA	Valor de la tangente para el ángulo de inclinación máxima del terreno.
6	Índice topográfico compuesto (CTI)	NA	Métrica que relaciona el tamaño del área de drenaje de una cuenca en un punto del mapa, con la pendiente en ese mismo punto.
7	Height Above Nearest Neighbor (HAND)	msnm	Altura del terreno en metros calculada con respecto al punto de drenaje más cercano [18], [16].

explicada en la sección II, a través del cálculo de características basadas en la textura del terreno; estas texturas son extraídas a partir de los descriptores morfológicos del terreno descritos en la sección III.

Existen en la literatura varias metodologías de extracción de texturas ampliamente utilizadas en áreas como visión por computador, procesamiento digital de imágenes, reconocimiento de voz, etc. Entre las más destacadas se encuentran los métodos de extracción de texturas de Gabor, Local Binary Pattern y las texturas de Haralick [19].

		(-1,-1)	(0,-1)	(+1,-1)
		(-1,0)	(0,0)	(+1,0)
		(-1,+1)	(0,+1)	(+1,+1)

Figura 3. Posibles direcciones de cálculo de la matriz de co-ocurrencia.

IV-A. Descriptores de textura de Haralick

En la década de los 70's, Haralick desarrolló un modelo para analizar texturas en una imagen por medio del uso de la matriz de co-ocurrencia (COM) de los valores de intensidad de la imagen.

La matriz de co-ocurrencia que se denotará como P_{kl} , donde el elemento $P_{kl}(i, j)$ otorga el valor de frecuencia (divido por NT) de ocurrencia de los valores de color i y j en dos pixeles ubicados en una posición relativa dada por el vector (k, l) . La variable NT significa el número de pixeles que fueron necesarios para calcular P_{kl} , con esto se normaliza la matriz de co-ocurrencia ya que la suma de todos sus elementos es uno.

Si la variable de color tiene por ejemplo una resolución de 256 (de 0 a 255), el tamaño de la matriz de co-ocurrencia P_{kl} será 256x256. Ya que el cálculo sucesivo de una matriz de este tamaño implica un costo computacional elevado, es común que se utilicen matrices más pequeñas empleando sólo los bits más significativos de la variable de color [20]. A manera de ejemplo, se puede tener una matriz de co-ocurrencia de 8x8 agrupando el valor de la variable de color en rangos como: [0, ..., 31], [32, ..., 63], ... [224, ..., 255].

La matriz de co-ocurrencia contiene en su estructura información sobre los patrones de textura contabilizando cuantas veces se repite una combinación de valores dentro de una imagen. Este conteo se realiza en una dirección dada de 8 posibles como se muestra en la figura 3. Haralick presenta 14 métricas que son obtenidas realizando diferentes operaciones sobre la COM. La tabla II indica el nombre de cada uno de las métricas propuestos por Haralick.

Pese a que estas métricas se desarrollaron para analizar texturas en imágenes, su conceputalización es trasladable al caso de los mapas raster georeferenciados, puesto que al tener estos últimos la forma de una matriz bidimensional coincide con la definición discreta de una imagen, acomodándose a los cálculos y mecanismos de la metodología. Antes de aplicar la metodología de texturas a los mapas raster, es necesario

Cuadro II
LISTA DE LAS MÉTRICAS DE TEXTURA PROPUESTAS POR HARALICK

Id	Nombre
1	Angular Second Moment
2	Contrast
3	Correlacion
4	Sum of squares
5	Inverse Difference Moment
6	Sum Average
7	Sum Entropy
8	Sum Variance
9	Entropy
10	Difference Variance
11	Difference Entropy
12	Information Measures of Correlation 1
13	Information Measures of Correlation 2
14	Maximal Correlation Coefficient

discretizar previamente los datos de cada mapa en valores enteros para que puedan ser tratados como si fueran niveles de gris en una imagen.

IV-B. Extracción de texturas a partir de descriptores geomorfológicos

Para cada una de las 14 métricas de Haralick de la tabla II se generó un nuevo mapa (descriptor) del mismo tamaño que el mapa original, donde el valor de cada celda es la métrica calculada en una ventana cuadrada con $lado = 2r + 1$ centrada en esa celda. La variable r es un parámetro del algoritmo de extracción de texturas que representa un determinado número de celdas.

Cada uno de los siete descriptores morfológicos contienen valores reales, con excepción de las elevaciones y el HAND; por lo tanto, para poder aplicar la metodología de texturas tal y como fue definida para el caso de imágenes, es necesario discretizar sus valores y escalarlos. Este escalamiento tuvo además el propósito de reducir dimensiones de las matrices de co-ocurrencia evitando al máximo pérdida de información. Por lo anterior, los valores de todos los descriptores morfológicos se escalaron y discretizaron en dos diferentes rangos:

$$d_1 = [0, 7] , d_2 = [0, 15] \quad (1)$$

Esta situación da como resultado matrices de co-ocurrencia de 8x8 y 16x16 respectivamente por cada una de las 14 métricas de textura, por cada uno de los descriptores morfológicos.

Análogamente se eligieron varios valores para el tamaño de la ventana sobre la que se extrajo la información de texturas. El rango de valores de r fue:

$$r = [10, 15, 25, 30, 50] \quad (2)$$

Dado que cada celda en los descriptores representa un área de $92 m^2$, se entiende entonces que las ventanas representan áreas de alrededor de $3,5 km^2$, $7,6 km^2$, $21 km^2$, $30 km^2$ y $82,6 km^2$ respectivamente.

La ejecución del algoritmo de extracción bajo cada combinación de los parámetros descritos anteriormente entregó un total de 980 descriptores adicionales. Con el fin de facilitar el análisis de los resultados de la clasificación con los nuevos descriptores en etapas posteriores de este trabajo, se estableció en (3) una nomenclatura para denominar a cada uno de ellos.

$$H_{\mathbf{f}\#, \mathbf{r}\#, \mathbf{d}\#}^{(\#)} \quad (3)$$

En dicha nomenclatura el superíndice indica una de las 14 métricas de textura de la tabla II, f se refiere al descriptor morfológico de la tabla I al que se le extrajo la información de textura, r representa el tamaño de la ventana utilizada y d indica el tamaño de la matriz de co-ocurrencia. Esta nomenclatura será utilizada en el resto del documento.

Para ejemplificar los resultados obtenidos la figura 4 muestra algunas de las características de Haralick para una ventana con $r = 10$, $d = 8$, que corresponden al mapa de elevaciones. Se puede observar en la figura que algunos de ellos resaltan adecuadamente la división de la zona que representa en las diferentes clases objetivo.

Finalmente, el algoritmo básico de Haralick fue optimizado para este problema gracias a una propiedad de la matriz de co-ocurrencia que permite que su cálculo para una ventana centrada en una celda (i, j) se realice a partir de pequeñas modificaciones a la matriz de co-ocurrencia para una ventana de igual tamaño y forma centrada en la celda $(i, j - 1)$ como se muestra a continuación:

$$C_{(i,j)} = C_{(i,j-1)} - C_{\text{izquierda}(i,j-1)} + C_{\text{derecha}(i,j)} \quad (4)$$

Donde $C_{(i,j)}$ es la matriz de co-ocurrencia de una ventana centrada en la celda (i, j) del mapa, $C_{\text{izquierda}(i,j-1)}$ es la matriz de co-ocurrencia de la primera columna de la ventana centrada en $(i, j - 1)$, y $C_{\text{derecha}(i,j)}$ es la matriz de co-ocurrencia de la última columna de la ventana centrada en la celda (i, j) . De esta manera se eliminan operaciones redundantes, y se agiliza el cálculo.

V. CLASIFICACIÓN AUTOMÁTICA

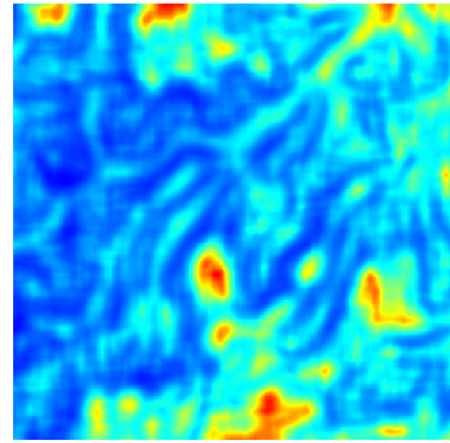
V-A. Selección de características

El conjunto de datos para la clasificación queda definido como la unión de los descriptores morfológicos con los descriptores basados en texturas de Haralick. Una vez consolidados los datos, el proceso de clasificación continúa seleccionando aquellas características que ofrecen una mejor separación de las clases objetivos. Para seleccionar tales características se usó el algoritmo de selección conocido como SFS [21], [5] (Sequential Forward Search por su nombre en inglés).

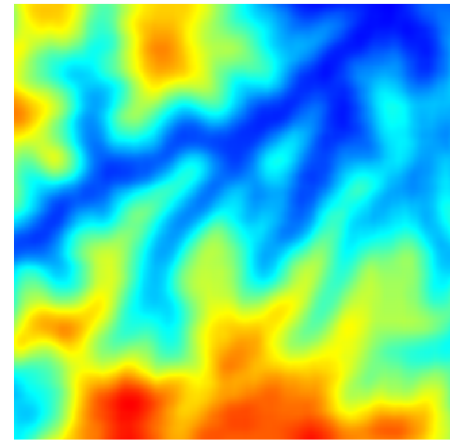
El objetivo del algoritmo SFS es encontrar los k primeros descriptores que maximizan la separabilidad de las clases de acuerdo a un criterio establecido (criterios wrapper[22] y filter[23]). Los descriptores seleccionados se entregan en orden de mayor a menor grado de separabilidad; teniendo en cuenta que la separabilidad conjunta de los p primeros es menor que la separabilidad de los $p+1$ descriptores entregados.



a) Conjunto de etiquetas para el modelo



b) Descriptor $H_{\mathbf{f}1, \mathbf{r}10, \mathbf{d}08}^{(1)}$



c) Descriptor $H_{\mathbf{f}1, \mathbf{r}10, \mathbf{d}08}^{(6)}$

Figura 4. Ejemplo de algunos de los descriptores de textura extraídos.

Para este trabajo se usó la implementación del algoritmo SFS que se encuentra en el toolbox Balú [24], y se estableció un valor de $k = 40$ como máximo número de características para cuatro criterios de separabilidad: Fisher,

Cuadro III
DESCRIPTORES SELECCIONADOS CON VARIAS METODOLOGÍAS DE
SELECCIÓN DE CARACTERÍSTICAS

	Criterio	Descriptores seleccionados (Solo los primeros 3)	Total
S1	Fisher	$H_{f_2, r_{25}, d_{16}}^{(6)}, H_{f_7, r_{15}, d_{16}}^{(2)}, H_{f_7, r_{50}, d_{16}}^{(12)}$	40
S2	LDA	$H_{f_2, r_{25}, d_{16}}^{(6)}, H_{f_3, r_{30}, d_{08}}^{(1)}, H_{f_7, r_{25}, d_{08}}^{(14)}$	27
S3	Knn, $k = 5$	$H_{f_2, r_{25}, d_{16}}^{(6)}, H_{f_7, r_{50}, d_{08}}^{(2)}, H_{f_4, r_{25}, d_{08}}^{(7)}$	21
S4	Knn, $k = 7$	$H_{f_2, r_{25}, d_{16}}^{(6)}, H_{f_7, r_{15}, d_{16}}^{(2)}, H_{f_6, r_{25}, d_{08}}^{(7)}$	16

LDA, KNN con $k = 5$ y KNN con $k = 7$ [25], [26].

En la tabla III se muestran los diferentes subconjuntos de descriptores entregados por el algoritmo SFS. Solo se muestran los 3 primeros descriptores seleccionados por motivos de estilo y legibilidad; sin embargo, se reporta el total entregado por el algoritmo. Dos aspectos se resaltan de los resultados de la selección: 1) Para cada una de las selecciones, el primer descriptor es siempre el mismo; es decir, aquel descriptor que ofrece una mejor separabilidad de las clases fue persistente en cada una de las cuatro selecciones realizadas. 2) El número total de descriptores seleccionados en S4 y S3 es significativamente menor que los obtenidos en S1; donde de hecho, el algoritmo SFS no converge, entregando el máximo número de descriptores pedido.

V-B. Elección de clasificadores y pruebas de validación

Para llevar a cabo los experimentos de este trabajo se utilizaron nueve clasificadores de tipo supervisado, de manera que constituyeran un conjunto suficientemente representativo de los diferentes enfoques de reconocimiento de patrones. Entre los clasificadores utilizados se encuentran: K-Nearest-Neighbor, SVM (Support Vector Machines), Redes Neuronales, LDA (Linear Discriminant Analysis), entre otros. La implementación de cada uno de ellos está incluida en el toolbox Balú[24] que está disponible en la web para su uso en ambientes académicos.

La calidad de los resultados de clasificación fue evaluada mediante la metodología de validación cruzada con un total de 10 divisiones de los datos. Esta validación reportó el desempeño de cada uno de los nueve clasificadores, y construyó intervalos con el 95 % de confianza para ayudar a analizar la estabilidad de los resultados obtenidos.

VI. RESULTADOS

El proceso de cálculo fue exhaustivo y costoso computacionalmente debido principalmente al tamaño del conjunto de datos y a la gran cantidad de combinaciones de parámetros. Sin embargo, en la etapa de selección de características los datos se redujeron a aproximadamente el 4 % de su tamaño inicial; lo que facilitó el ejercicio de clasificación y de las pruebas.

Cuadro IV
DESEMPEÑO DE LA CLASIFICACIÓN EN EL CONJUNTO DE PRUEBAS

Clasificador	Sólo descriptores morfológicos	S1	S2	S3	S4
knn, $k = 5$	0.7210	0.8985	0.6936	0.9682	0.9674
knn, $k = 7$	0.7314	0.8921	0.7028	0.9657	0.9651
knn, $k = 9$	0.7345	0.8873	0.6992	0.9637	0.9638
LDA	0.7198	0.8672	0.8168	0.7875	0.7770
QDA	0.6989	0.8677	0.8223	0.8077	0.7801
Red Neuronal	0.7195	0.8683	0.8190	0.7883	0.7791
SVM, kernel = RBF	0.7283	0.9558	0.9607	0.9729	0.9737
Mahalanobis	0.6846	0.8695	0.8396	0.8240	0.7969
Mínima distancia	0.6705	0.6243	0.5018	0.6127	0.5730

VI-A. Resultado de la clasificación

En la tabla IV se resumen los resultados de la validación cruzada para cada uno de los clasificadores y para cada uno de los subconjuntos de características obtenidas en la sección V. En la tabla IV se incluye también la validación cruzada de los nueve clasificadores con el conjunto de características que corresponde solamente a los descriptores morfológicos.

En general el desempeño de la clasificación con información de texturas arrojó resultados satisfactorios con índices de clasificación por encima del 80 %, y llegando a un máximo de 97,37 % de desempeño con un intervalo de confianza de

$$97,07\% \leq p \leq 97,66\%$$

Este resultado fue alcanzado con la selección S4 en el clasificador SVM, donde se usó como kernel una función de base radial. Existen algunas excepciones donde el desempeño estuvo por debajo del 70 %, particularmente en el caso del clasificador de mínima distancia.

VI-B. Discusión

En la tabla IV se resalta el desempeño del clasificador SVM, el cual posee la mayor tasa de aciertos en la clasificación de todos los usados en el conjunto de pruebas. Además, su desempeño a través de todas las selecciones realizadas siempre se mantuvo entre las más altas con respecto al resto de clasificadores, con excepción de la clasificación realizada solo con descriptores morfológicos.

Por otro lado, ningún clasificador obtuvo resultados por encima del 80 % con la selección de características que no incluía información de texturas. De lo anterior se infiere entonces la utilidad y efectividad de la metodología propuesta frente al enfoque típico para el problema estudiado.

VII. CONCLUSIONES Y TRABAJO FUTURO

VII-A. Conclusiones

La metodología de extracciones de características a partir de texturas de Haralick entrega una gran cantidad de descriptores

nuevos, donde algunos de ellos aportan una considerable mejora en el desempeño de la clasificación cuando son incluidos. Según revelan los desempeños en las clasificaciones realizadas, aquella que se realiza con información de texturas es superior a la realizada solo con descriptores morfológicos con alrededor de 20 % más de precisión al separar las clases.

A partir de los análisis realizados a los descriptores morfológicos iniciales se puede deducir que ninguno de ellos ofrece una separabilidad aceptable para discriminar los tipos de relieve; un ejemplo de ellos es el descriptor “curvatura” que se muestra en la figura 1. Sin embargo, los resultados de la extracción de características revelan que el descriptor de textura que mejor separa los datos en cada una de las selecciones realizadas, es $H_{f_2, r_{25}, d_{16}}^{(6)}$, que fue extraído a partir del descriptor curvatura.

La contribución principal de este trabajo es el establecimiento de un conjunto de descriptores apto para el problema de la clasificación de tipos de relieve. Este conjunto ofrece desempeños de clasificación mayores que los reportados en la literatura en aproximaciones de solución al mismo problema.

VII-B. Trabajo futuro

Para la realización de los experimentos de este trabajo se utilizó un número razonable de clasificadores de manera que fuera un conjunto suficientemente representativo de los diferentes enfoques de reconocimiento de patrones. Sin embargo, los parámetros de cada uno de ellos se eligieron teniendo en cuenta su efectividad en otros problema de clasificación. Un foco importante de trabajo futuro es encontrar un conjunto de parámetros para cada clasificador que sea más adecuado para el problema presentado; e incluso, utilizar clasificadores híbridos que puedan ayudar a mejorar el desempeño de la clasificación.

Los algoritmos utilizados para la selección de características fueron limitados, y se eligieron por ser los más usados y recomendados en la literatura. Sin embargo, existe una gran cantidad de algoritmos de selección de características que aplicados al problema planteado podrían mejorar los resultados de la clasificación o mejorar la eficiencia computacional de la misma.

Una vez que ha sido comprobada la eficacia de extraer información de texturas para clasificar tipos de relieve, una dirección de trabajo futuro es la exploración de otras técnicas para extraer dicha información y aportar más descriptores útiles para la clasificación.

REFERENCIAS

- [1] B. Bue and T. Stepinski, “Automated classification of landforms on Mars,” *Computers & Geosciences*, vol. 32, no. 5, pp. 604–614, Jun. 2006. [Online]. Available: <http://linkinghub.elsevier.com/retrieve/pii/S0098300405001925>
- [2] P. Burrough, “High-resolution landform classification using fuzzy k-means,” *Fuzzy Sets and Systems*, vol. 113, no. 1, pp. 37–52, Jul. 2000. [Online]. Available: <http://linkinghub.elsevier.com/retrieve/pii/S0165011499000111>
- [3] E. d. g. y. m. a. U. Nacional, “Implementación de la metodología de zonificación de amenazas por inundaciones para tres grandes cuencas del país (colombia),” Universidad Nacional de Colombia - Sede Medellín, Tech. Rep., 2011.
- [4] J. C. Castro, “Metodología para la obtención de modelos digitales de terreno hidrológica y geomorfológicamente coherentes,” Master’s thesis, Escuela de geociencias y medio ambiente, Universidad Nacional de Colombia, 2011.
- [5] A. Jain and D. Zongker, “Feature Selection: Evaluation, Application, and Small Sample Performance,” vol. 19, no. 2, pp. 153–158, 1997.
- [6] B. Irvin, “Fuzzy and isodata classification of landform elements from digital terrain data in Pleasant Valley, Wisconsin,” *Geoderma*, vol. 77, no. 2-4, pp. 137–154, Jun. 1997. [Online]. Available: <http://linkinghub.elsevier.com/retrieve/pii/S0016706197000190>
- [7] R. MacMillan, “A generic procedure for automatically segmenting landforms into landform elements using DEMs, heuristic rules and fuzzy logic,” *Fuzzy Sets and Systems*, vol. 113, no. 1, pp. 81–109, Jul. 2000. [Online]. Available: <http://linkinghub.elsevier.com/retrieve/pii/S0165011499000147>
- [8] J. A. Hartigan and M. A. Wong, “Algorithm as 136: A k-means clustering algorithm,” *Journal of the Royal Statistical Society*, 1979.
- [9] G. H. Ball, “ISODATA: A novel method for data analysis and pattern classification,” Stanford Research Institute, Tech. Rep., 1965.
- [10] J. C. Bezdek, “FCM : THE FUZZY c-MEANS CLUSTERING ALGORITHM 1 ; yk E Y ~ l,” vol. 10, no. 2, pp. 191–203, 1984.
- [11] G. Acciani, E. Chiarantoni, G. Fornarelli, and S. Vergura, “2003 Special Issue A feature extraction unsupervised neural network for an environmental data set,” *Neural Networks*, vol. 16, pp. 427–436, 2003.
- [12] F. Bacao, V. Lobo, and M. Painho, “The self-organizing map, the Geo-SOM, and relevant variants for geosciences,” *Computers & Geosciences*, vol. 31, no. 2, pp. 155–163, Mar. 2005. [Online]. Available: <http://linkinghub.elsevier.com/retrieve/pii/S0098300404001918>
- [13] K. Arrell, P. Fisher, N. Tate, and L. Bastin, “A fuzzy c-means classification of elevation derivatives to extract the morphometric classification of landforms in Snowdonia, Wales,” *Computers & Geosciences*, vol. 33, no. 10, pp. 1366–1381, Oct. 2007. [Online]. Available: <http://linkinghub.elsevier.com/retrieve/pii/S0098300407000933>
- [14] L. Dragut and T. Blaschke, “Automated classification of landform elements using object-based image analysis,” *Geomorphology*, vol. 81, no. 3-4, pp. 330–344, Nov. 2006. [Online]. Available: <http://linkinghub.elsevier.com/retrieve/pii/S0169555X06001589>
- [15] O. Prima, a. Echigo, R. Yokoyama, and T. Yoshida, “Supervised landform classification of Northeast Honshu from DEM-derived thematic maps,” *Geomorphology*, vol. 78, no. 3-4, pp. 373–386, Aug. 2006. [Online]. Available: <http://linkinghub.elsevier.com/retrieve/pii/S0169555X06000791>
- [16] C. Renno, a. Nobre, L. Cuartas, J. Soares, M. Hodnett, J. Tomasella, and M. Waterloo, “HAND, a new terrain descriptor using SRTM-DEM: Mapping terra-firme rainforest environments in Amazonia,” *Remote Sensing of Environment*, vol. 112, no. 9, pp. 3469–3481, Sep. 2008. [Online]. Available: <http://linkinghub.elsevier.com/retrieve/pii/S003442570800120X>
- [17] M. Ruiz-Ochoa, G. Bernal, and J. Polanía, “Influence of sinú river and the caribbean sea over the cispata lagoon system,” *Boletín de Investigaciones Marinas y Costeras - INVEMAR*, 2008.
- [18] a.D. Nobre, L. Cuartas, M. Hodnett, C. Rennó, G. Rodrigues, a. Silveira, M. Waterloo, and S. Saleska, “Height above the Nearest Drainage, a hydrologically relevant new terrain model,” *Journal of Hydrology*, vol. 404, no. 1-2, pp. 13–29, Apr. 2011. [Online]. Available: <http://linkinghub.elsevier.com/retrieve/pii/S0022169411002599>
- [19] T. Randen, “Filtering for Texture Classification : A Comparative Study,” vol. 21, no. 4, pp. 291–310, 1999.

- [20] K. Castleman, *Digital image processing*, ser. Prentice-Hall signal processing series. Prentice Hall, 1996. [Online]. Available: <http://books.google.cl/books?id=0YUYAQAIAAJ>
- [21] K. Koutroumbas and S. Theodoridis, *An Introduction to Pattern Recognition: A Matlab Approach*, Elsevier, Ed., 2009. [Online]. Available: [http://www.unhas.ac.id/tahir/ebook/600 electrical engineering books/An Introduction to Pattern Recognition - Michael Alder.pdf](http://www.unhas.ac.id/tahir/ebook/600%20electrical%20engineering%20books/An%20Introduction%20to%20Pattern%20Recognition%20-%20Michael%20Alder.pdf)
- [22] R. Kohavi and H. John, "Artificial Intelligence Wrappers for feature subset selection," vol. 97, no. 97, pp. 273–324, 2011.
- [23] M. A. Hall and L. A. Smith, "Practical Feature Subset Selection for Machine Learning," 1997.
- [24] D. Mery, "BALU: A toolbox Matlab for computer vision, pattern recognition and image processing (<http://dmery.ing.puc.cl/index.php/balu>)," 2011.
- [25] S. B. Kotsiantis, "Supervised Machine Learning : A Review of Classification Techniques," *Informatica*, vol. 31, pp. 249–268, 2007.
- [26] a.K. Jain and P. Duin, "Statistical pattern recognition: a review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, pp. 4–37, 2000. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=824819>