

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

Selección no supervisada de bandas basadas en descomposición de valores singulares para el análisis de imágenes hiperespectrales

Cristian Felipe David Bedoya

Ingeniería de Sistemas

María C. Torres Madroñero

INSTITUTO TECNOLÓGICO METROPOLITANO

25/07/2018

	<p style="text-align: center;">INFORME FINAL DE TRABAJO DE GRADO</p>	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

RESUMEN

El objetivo de este trabajo de grado fue evaluar un algoritmo basado en descomposición de valores singulares, para la selección de bandas discriminantes en imágenes hiperespectrales. Para la evaluación se usaron imágenes hiperespectrales reales, capturadas sobre dos regiones: AP Hill e Indian Pines. A partir de estos resultados se evaluó el algoritmo y su capacidad para seleccionar bandas discriminantes, que permitan mejorar la clasificación. Finalmente, se usó el algoritmo para seleccionar bandas en imágenes multiespectrales capturadas sobre úlceras cutáneas. Los productos de este trabajo de grado incluyen el código en MATLAB del algoritmo de selección de bandas, los resultados de clasificación y sus matrices de confusión para las imágenes de AP Hill e Indian Pines, y las bandas más relevantes para las imágenes de úlceras cutáneas.

Palabras clave: SVD, SVM, imágenes hiperespectral, selección de bandas.

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

RECONOCIMIENTOS

La realización de esta investigación fue posible, en primer lugar, gracias a la cooperación brindada por la Profesora María Constanza Torres Madroño, Profesora del instituto tecnológico metropolitano, quien en su desempeño como Integrante del laboratorio de máquinas inteligentes y reconocimiento de patrones brindo de su tiempo y conocimiento para llevar a cabo la investigación.

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

ACRÓNIMOS

SVD : [Singular value decomposition](#) (Valor singular de descomposición)

SVM: Support vector machines (Soporte de máquinas virtuales)

PCA: análisis de componentes principales

PP: Proyección Persecución

COV: Covarianza

EIG: Eigenvectors

QR: Orthogonal-triangular decomposition (Descomposición triangular ortogonal)

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

TABLA DE CONTENIDO

TABLA DE CONTENIDO5

1. INTRODUCCIÓN6
 2. MARCO TEÓRICO8
 3. METODOLOGÍA10
 4. RESULTADOS Y DISCUSIÓN13
 5. CONCLUSIONES, RECOMENDACIONES Y TRABAJO FUTURO16
- REFERENCIAS17

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

1. INTRODUCCIÓN

Las imágenes hiperespectrales y multiespectrales miden la radianza reflejada y emitida por una superficie en cientos o decenas de bandas espectrales a lo largo del espectro electromagnético [1]. La forma de la firma espectral permite determinar los materiales presentes en una escena. Las aplicaciones de este tipo de imágenes son muchas, van desde la agricultura de precisión, minería, estudio del clima, estudios ambientales, defensa y salud. En esa última área, actualmente el laboratorio de máquinas inteligentes y reconocimiento de patrones está participando en un proyecto de investigación financiado por Colciencias, para el desarrollo de un dispositivo portátil para apoyar el diagnóstico y monitoreo de úlceras cutáneas causadas por leishmaniosis. El dispositivo a desarrollar se basa en imágenes multiespectrales [2]. Para llegar al desarrollo del dispositivo se requiere de un estudio detallado de la firma espectral que caracteriza la piel sana, la úlcera y su borde. Este estudio incluye la identificación de bandas espectrales discriminantes que permitan discriminar la piel sana de la úlcera, y también reconocer el efecto de diferentes tratamientos médicos sobre la úlcera. Para llegar a este desarrollo, se utiliza un estudio comparativo de técnicas de reducción de dimensiones, las cuales incluye técnicas no supervisadas para la selección de bandas espectrales.

El objetivo de este trabajo de grado en el laboratorio es explorar métodos de selección de bandas no supervisados basados en descomposición en valores singulares – SVD. Para ello en el proyecto se desarrollarán las siguientes actividades:

1. Lectura y revisión del artículo “Unsupervised Feature Extraction and Band Subset Selection techniques based on Relative Entropy Criteria for Hyperspectral Data Analysis” [3].

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

2. Implementación en Matlab del algoritmo SVD based Band Subset Selection

3. Evaluación del algoritmo implementado usando imágenes hiperespectrales reales (AP Hill e Indian Pines) usando clasificadores SVM

Los objetivos específicos de este producto fueron:

1. Explorar métodos de selección de bandas no supervisados basados en descomposición en valores singulares.
2. Evaluar el desempeño del método de selección de bandas usando clasificadores SVM

Este trabajo de grado hace parte de los productos comprometidos en el Proyecto 115077757186 “Desarrollo y evaluación de un sistema portátil no invasivo basado en imágenes multiespectrales para el diagnóstico y seguimiento de tratamientos de úlceras cutáneas por Leishmaniasis”.

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

2. MARCO TEÓRICO

El presente trabajo analiza una de las técnicas de análisis de imágenes hiperespectrales propuestas en el artículo: “Unsupervised Feature Extraction and Band Subset Selection techniques based on Relative Entropy Criteria for Hyperspectral data Analysis “. El mencionado artículo propone métodos para extracción de características y la selección de subconjuntos de bandas en base a los criterios de entropía relativa o desorden de los datos. El objetivo principal de la extracción de características es reducir la dimensionalidad de los datos, manteniendo la capacidad de discriminar objetos de interés. Estos métodos logran el objetivo descrito maximizando la diferencia entre la distribución de datos del subespacio dimensional inferior y la distribución gaussiana estándar.

A medida que aumenta el número de bandas espectrales en los sensores hiperespectrales, se agrega una cantidad significativa de información, Por lo tanto, reducir la dimensionalidad de los datos hiperespectrales sin perder información relevante sobre objetos de interés es un tema muy importante para la comunidad científica de percepción remota. Los algoritmos de extracción / selección de características utilizan diferentes enfoques para reducir la dimensionalidad. Entre esas técnicas se encuentran el análisis de componentes principales (PCA), la selección de subconjuntos de bandas y proyección persecución (PP) que busca proyecciones lineales que optimizan cierto índice de proyección [4].

La técnica que fue propuesta para este proyecto se basa en SVD. Para observar el desempeño de dicha técnica se programó una función en Matlab y se usaron las imágenes hiperespectrales de uso abierto AP Hill y Indian Pines para validar su funcionalidad. A continuación, se describe el comportamiento de la función:

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

1. Se reciben los valores de l y p , donde l es la variable que contiene la imagen y p corresponde al número de bandas a seleccionar
2. Se redefinen dimensiones de la variable l para organizar la imagen en una matriz de numero de muestras por numero bandas
3. Se calcula la covarianza, C , de la variable l
4. Se calculan los valores propios y vectores propios de C
5. Se almacenan en V_i los p primeros vectores propios resultantes del proceso anterior,
6. Se hace uso de la descomposición ortogonal triangular sobre la transpuesta de la matriz V
7. $X = l$ multiplica P , siendo P el pivote resultante del proceso del paso anterior,
8. Se capturan los primeros p elementos de X en IF , como bandas seleccionadas.
9. $B =$ Se hallan las bandas seleccionadas

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

3. METODOLOGÍA

3.1 Base de datos

Para este proyecto se emplearon 2 imágenes hiperespectrales que se describen a continuación.

- Indian Pines: imagen hiperespectral capturada por el sensor AVIRIS de la NASA y se encuentra disponible en la página web [[http://www.ehu.es/ccwintco/index.php/Hyperspectral Remote Sensing Scenes](http://www.ehu.es/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes)]. La imagen tiene 145x145 pixeles, 224 bandas espectrales y 16 clases se presentan en la Figura 1.

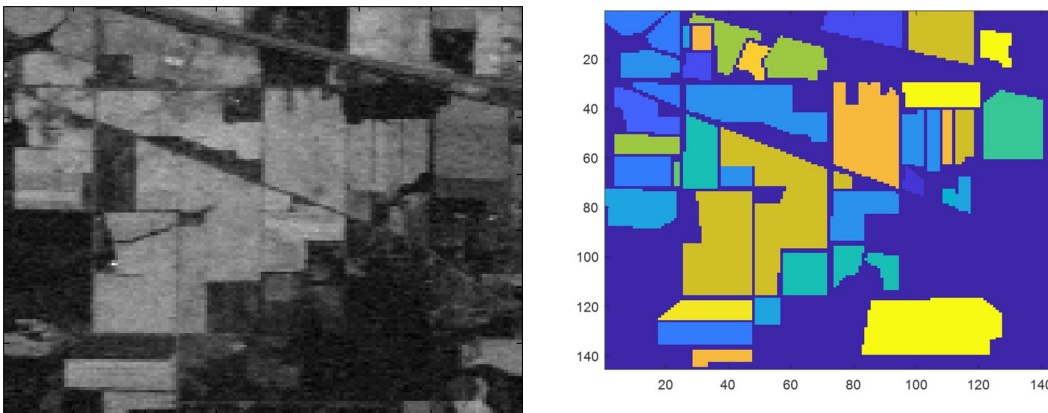


Figura 1. Banda 40 de la imagen Indian Pines y Mapa de Clasificación.

- AP Hills: imagen capturada por AVIRIS. Cuenta con 512 x 540, 224 bandas espectrales y 14 clases que se presentan en la Figura 2.

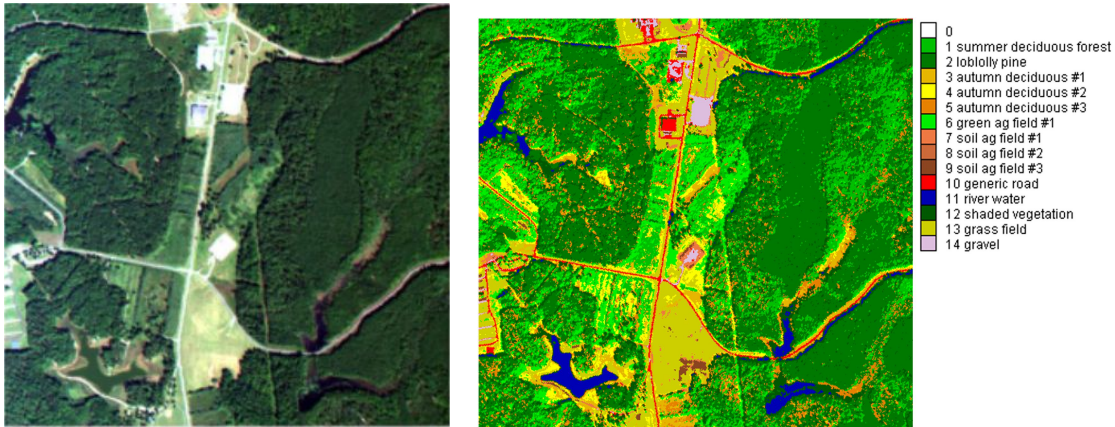


Figura 2. Composición RGB de la imagen AP Hill y Mapa de Clasificación.

3.2 Implementación

Para la implementación del método de descomposición se creó una función que recibiera la imagen hiperespectral y el número de bandas a evaluar. Dicha función es usada en conjunto con una máquina de soporte vectorial, que tiene como insumos el número de bandas de la imagen anteriormente ingresado y 10% de muestras para entrenamiento de la herramienta, dicha implementación retorna: Nro. de clases que contiene la imagen, precisión, banda de muestra del conjunto de datos de Indian Pines. A continuación, se listan los pasos que sigue la función creada.

Lo primero que realiza la función es capturar la imagen hiperespectral en una variable I , se redefinen las dimensiones de la matriz en la variable para aplicar la transpuesta. Teniendo esto se aplica covarianza usando la función COV y se almacena el resultado en la variable C . Luego se extraen los valores propios usando la función EIG , en V : vectores propios y D : valores propios. Posteriormente, se forma la matriz V_i usando los primeros P eigenvectors de la matriz V . Se aplica la función QR sobre la transpuesta de la matriz V y se multiplica la matriz original I con la matriz de permutación P , se almacena este valor en la variable X . Para terminar, se toman los primeros p eigenvectors de la matriz X y se buscan los 1s en la matriz de permutación para ser almacenados en la variable b .

	<p style="text-align: center;">INFORME FINAL DE TRABAJO DE GRADO</p>	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

3.3 Clasificación

Una máquina de soporte vectorial se usó para clasificar usando 10% de muestras para entrenamiento y usando 15,20,30 y 40 bandas de la imagen. Esto con el objetivo de obtener diferentes muestras y así seleccionar la más óptima.

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

4. RESULTADOS Y DISCUSIÓN

A continuación, se presenta los resultados obtenidos con las diferentes clasificaciones suministradas a la implementación para Indian Pines. La Figura 3 presenta el mapa de clasificación obtenido para la imagen Indian Pines. La Tabla 1 presenta el porcentaje promedio de precisión obtenido para 15, 20, 30 y 40 bandas. Dado que se corrió el experimento 10 veces, para cada experimento se presenta el porcentaje de precisión general en la Tabla 2.

Dado los resultados obtenidos por la máquina de soporte vectorial se puede apreciar como a medida que se aumenta el número de bandas seleccionadas, disminuye la precisión de la herramienta de clasificación, se tomaron diez muestras para las cuatro categorías de bandas seleccionadas y luego se promedió para conseguir un resultado más preciso.



Figura 3. Banda 40 de la imagen Indian Pines procesada por algoritmo

Tabla 1. Bandas seleccionadas y porcentaje de precisión promedio

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO			Código	FDE 089
				Versión	03
				Fecha	2015-01-22

Bandas seleccionadas	15	20	30	40
Promedio de precisión	54%	56%	58%	67%

Tabla 2. Precisión por cada iteración tomada para bandas seleccionadas

Muestra	1	2	3	4	5	6	7	8	9	10
Precisión	53%	54%	53%	54%	54%	55%	54%	55%	54%	54%
	56%	56%	56%	56%	57%	56%	56%	56%	56%	55%
	58%	58%	59%	58%	59%	57%	57%	59%	57%	57%
	66%	67%	66%	66%	62%	66%	67%	68%	67%	67%

Para el caso de la imagen hiperespectral AP Hill la ejecución de las funciones fue más lenta; debido a que esta imagen tiene más píxeles. Así que se logró obtener la precisión solo para 15 bandas, la cual fue de 64%

Finalmente, se aplicó el algoritmo de selección de bandas basado en la descomposición en valores singulares a 3 imágenes de úlceras. Las imágenes fueron capturadas con el sistema multiespectral asclepios y tienen un total de 32 bandas. La Tabla 3 presenta el número de bandas seleccionadas para cada una de las 3 imágenes.

Tabla 3. Bandas seleccionadas para imágenes úlceras

Paciente	Bandas									
1	34	1	13	18	16	9	26	6	30	21
3	34	1	7	13	5	16	10	26	30	19
5	34	1	32	16	7	26	4	13	29	19

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

5. CONCLUSIONES, RECOMENDACIONES Y TRABAJO FUTURO

De la evaluación que se realiza al método de selección de bandas no supervisados basados en descomposición en valores singulares SVD propuesto en el artículo “Unsupervised Feature Extraction and Band Subset Selection techniques based on Relative Entropy Criteria for Hyperspectral Data Analysis”. usando las imágenes hiperespectrales Indian Pines y Ap Hill con diferentes clasificadores SMV se concluye que el desempeño conseguido es aceptable y satisface lo buscado con las imágenes hiperespectrales implementadas en la evaluación del algoritmo desarrollado en matlab, de los diferentes resultados que se obtienen con la variación de clasificadores SMV, se evidencia que mientras mayor sea el numero bandas seleccionadas mayor será la precisión obtenida, La velocidad del algoritmo desarrollado es óptima para imágenes cuya resolución no supera los doscientos pixeles, el algoritmo obtenido está muy simplificado y es de fácil comprensión.

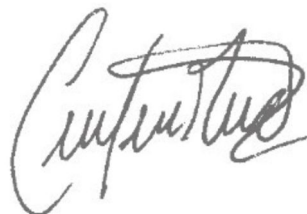
En trabajos futuros se puede hacer uso de imágenes hiperespectrales de mayor tamaño, esto con el objetivo de buscar optimizar el funcionamiento del algoritmo, darle un desempeño optimo con imágenes hiperespectrales de gran tamaño.

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

REFERENCIAS

1. Imagen Multiespectral. (s.f.). En Wikipedia. Recuperado el 25 de Julio de 2018 de https://es.wikipedia.org/wiki/Imagen_multiespectral
2. Imagen Hiperespectral. (s.f.). En Wikipedia. Recuperado el 22 de Marzo de 2018 de <https://es.wikipedia.org/wiki/Hiperespectral>
3. Emmanuel Arzuaga-Cruz, Luis O. Jiménez Rodríguez y Miguel Vélez-Reyes presentan un metodo de selección de bandas no supervisados basados en descomposición en valores singulares SVD en el artículo (Unsupervised Feature Extraction and Band Subset Selection techniques based on Relative Entropy Criteria for Hyperspectral Data Analysis, 2003).
4. Se puede usar la definición y uso de imágenes hiperespectrales como se propone en (Análisis de imágenes hiperespectrales de Avid Roman Gonzalez & Natalia Indira Vargas Cuentas, 2013).

FIRMA ESTUDIANTES _____



FIRMA ASESOR _____

FECHA ENTREGA: _____

FIRMA COMITÉ TRABAJO DE GRADO DE LA FACULTAD _____

RECHAZADO__

ACEPTADO__

ACEPTADO CON MODIFICACIONES__

ACTA NO. _____

FECHA ENTREGA: _____

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

FIRMA CONSEJO DE FACULTAD _____
ACTA NO. _____
FECHA ENTREGA: _____