

Identificación de áreas donde se siembran variedades tradicionales de frijol común usando remote sensing y algoritmos de clasificación

Harold Armando Achicanoy Estrella, *Maestría en Ingeniería con Énfasis en Ingeniería de Sistemas y Computación, Universidad del Valle*

Abstract—El objetivo del presente artículo es exponer una revisión del estado del arte sobre la identificación de zonas donde se cultivan variedades criollas de frijol a través de algoritmos de clasificación por medio de imágenes satelitales. Se examinaron las investigaciones más relevantes en el uso de imágenes multi-espectrales e hiper-espectrales en el reconocimiento de áreas de cultivo, así como los principales algoritmos de clasificación de imágenes que aprovechan la información espacio-temporal proveniente de los sensores remotos y la fenología de los cultivos para brindar resultados más precisos. Finalmente se exponen las preguntas de investigación a desarrollar y los correspondientes desafíos a afrontar.

Keywords—*Mapeo de zonas de cultivo, remote sensing, algoritmos de clasificación de imágenes, imágenes multi-espectrales, imágenes hiper-espectrales, fenología de cultivos, variedades criollas de frijol común.*

I. INTRODUCCIÓN

Desde el punto de vista de la seguridad alimentaria y conservación de la biodiversidad, la identificación de zonas dedicadas al cultivo es un tema de amplio interés. Por un lado, el constante aumento de la tasa de crecimiento de la población a nivel global impone un reto en función de la disponibilidad de alimentos a producir para cubrir la demanda generada, por ende, se requiere de más de alimentos para producir pero se debe recordar que el uso del suelo tiene condiciones limitadas.

Por otro lado, la conservación de variedades silvestres y variedades criollas de cultivos, es decir, las que no han tenido la influencia del hombre o que la han tenido a partir de procesos no tecnificados, respectivamente; es de vital interés desde el punto de vista de conservación de recursos genéticos. Esto debido a que dichas variedades cuentan con características que los han hecho resistentes a condiciones climáticas adversas, volviéndolos deseables desde el punto de vista genético para hacer cruces y así poder mejorar las variedades cultivadas con el fin de obtener mayores rendimientos y contribuir a la seguridad alimentaria.

Específicamente las variedades tradicionales o criollas, también conocidas en inglés como *landraces*, corresponden a variedades silvestres que fueron intervenidas por el hombre realizando su correspondiente domesticación, de tal manera que dichas variedades puedan ser usadas para la posterior ingesta. Las variedades criollas son importantes desde el punto de vista nutricional, cultural y de adaptación debido a que

son de uso local, han sido mejoradas únicamente realizando la selección de los mejores granos y se encuentran adaptadas a condiciones climáticas locales con climas específicos. Debido a estas razones, el interés del presente estudio consiste en identificar, utilizando imágenes provenientes de sensores remotos satelitales, los lugares donde se cultivan variedades criollas de cultivos, particularmente hablando variedades criollas de frijol común.

La identificación de áreas donde se siembran variedades tradicionales de frijol es un tema relevante, en primer lugar porque es un tema que poco se ha abordado desde el enfoque de imágenes satelitales, donde se ha hecho principal énfasis en los cultivos con la mayor producción a nivel mundial como arroz, maíz, trigo, soya [1] [2] [3] [4]. Por otro lado, los principales desarrollos en la diferenciación de variedades de frijol se han realizado a nivel local con sensores hiper-espectrales manuales obteniendo resultados prometedores [5] [6]. Por tanto, un estudio que permita realizar la diferenciación entre variedades criollas vs variedades de alto rendimiento de frijol usando imágenes satelitales es de vital interés para el área del conocimiento de conservación de recursos genéticos.

El contenido del presente documento desarrolla el estado del arte de la siguiente manera: la primera sección exhibe los principales fundamentos en los cuales se sustenta el uso de imágenes satelitales y sus diferentes aplicaciones en la agricultura, seguidamente se presenta una sección enfocada en los algoritmos usados en la clasificación de imágenes satelitales mostrando las diferentes metodologías disponibles y su aplicación en casos de estudio. Luego se presentan los estudios centrados en el cultivo de frijol común mostrando los avances en clasificación de variedades y finalmente las conclusiones obtenidas donde se plantean las preguntas de investigación del presente estudio.

II. REMOTE SENSING PARA LA IDENTIFICACIÓN DE ÁREAS DE CULTIVO

Remote sensing - definiéndolo como el medio a través del cual se obtiene información de un objeto sin realizar contacto físico con el, sino mediante la toma de imágenes que miden diferentes características del espectro electromagnético - ha tenido un papel importante en la agricultura. Diversas aplicaciones que van desde la identificación de zonas donde se cultiva permitiendo diferenciarlas de otros usos del suelo

hasta la agricultura de precisión donde se puede determinar si las plantas de un lote específico carecen de los nutrientes necesarios para llevar a cabo su adecuado desarrollo, muestran su inmenso potencial.

No obstante, al ser un área de amplio impacto, tiene muchos conceptos que necesitan ser enfocados específicamente al tema de interés que se está abordando. Por esta razón, a continuación se presentan dos artículos que exponen los conceptos teóricos básicos en los que se sustentan los índices de vegetación así como los tipos de imágenes a disposición con las cuales se pueden desarrollar este y futuros estudios.

Basso et al. [7], presenta una revisión de los principios biofísicos a partir de los cuales se construyen los índices de vegetación. Los cuales se basan en procesos como la fotosíntesis y la evapotranspiración de las plantas, que pueden ser medidas a través del espectro electromagnético por los sensores remotos. Se distinguen los diversos tipos de sensores: satelitales, en aviones, remotamente tripulados o drones, manuales y en vehículos de trabajo como tractores. Dado que las plantas toman la luz solar como fuente de energía, esto hace que se generen diversos procesos químicos y físicos que pueden ser capturados por los índices de vegetación permitiendo determinar tanto el estado fenológico de un cultivo como si el cultivo requiere de nutrientes en un momento dado.

Adicionalmente se presenta una clasificación de los tipos de índices en: índices intrínsecos, los cuales están relacionados con la medición de características de la vegetación propias del cultivo en sí, y los índices que combinan información del suelo, con los que se puede medir características de sequía y déficit de nutrientes en el cultivo.

Por otro lado y como insumo de los futuros análisis, Sahoo et al. [8], hace una revisión de las imágenes provenientes de sensores remotos particularmente de los hiper-espectrales y sus usos en agricultura. En términos generales, existen dos tipos de imágenes que miden información a través del espectro electromagnético, los cuales son: imágenes multi-espectrales e imágenes hiper-espectrales. Las imágenes multi-espectrales miden componentes del espectro en rangos amplios, mientras, las imágenes hiper-espectrales se miden en rangos más estrechos y por ende adquieren más información.

El artículo destaca la importancia de las imágenes hiper-espectrales en la identificación y clasificación de propiedades biofísicas y bioquímicas de las plantas o cultivos bajo análisis, permitiendo realizar la discriminación de variedades de un mismo cultivo, detección temprana de estrés por sequía y afectación por plagas y enfermedades. Se identifican cuáles son los rangos óptimos del espectro al momento de evaluar las condiciones anteriores y se nombran las metodologías estadísticas para hacer su identificación usando índices de vegetación hiper-espectrales.

III. ALGORITMOS PARA LA CLASIFICACIÓN DE IMÁGENES SATELITALES

Desde el punto de vista estadístico y de las ciencias de la computación existen diversas metodologías para el análisis de

información, específicamente en el área de clasificación. Aquí se destaca la existencia de algoritmos de aprendizaje supervisado y no supervisado, con las cuales se pueden desarrollar tareas de clasificación y segmentación en imágenes, respectivamente. Sin embargo, no todas las metodologías tienen la capacidad de incorporar la componente temporal.

Gracias al avance en las tecnologías de la información se pueden tener mediciones en intervalos de tiempo relativamente cortos de las imágenes provenientes de sensores remotos satelitales, razón por la cual se tiene información a nivel tanto espacial como temporal. De este modo, a continuación se presenta un conjunto de estudios de clasificación de imágenes satelitales bajo diferentes metodologías algunas de las cuales tienen en consideración la componente temporal.

Clasificadores automáticos basados en reglas: Teluguntla et al. [3], tuvieron como objetivo principal la aplicación de una metodología de clasificación automática basada en reglas fenológicas que buscaba: 1) diferenciar entre zonas de cultivo vs pasturas, 2) clasificar áreas con sistemas secano vs sistemas irrigados, 3) clasificar las anteriores zonas en función de las intensidades de siembra, y 4) determinar cambios en la extensión de las anteriores áreas año a año. Como fuente de datos satelitales se usaron las series de tiempo de MODIS 250-m como producto agregado de 16 días para el periodo 2000-2015 para el continente Australiano. Entre los principales resultados, se destacan los altos niveles de precisión en los algoritmos de clasificación automática utilizando diferentes conjuntos para validación, esto gracias a que para la mayoría de los cultivos sembrados solo rige un patrón unimodal del ciclo de cultivos.

Xiong et al. [4], tuvo como objetivo principal realizar una clasificación automática de áreas de cultivo para todo el continente africano, permitiendo diferenciar diferentes sistemas de cultivo (riego y secano), intensidades de cultivo, tipos de cultivos y pasturas. Para esto se utilizaron series de tiempo de 16 días provenientes de MODIS a una resolución de 250-m para todo Africa, para las cuales se calculó el índice de vegetación NDVI para un periodo de análisis de 2014, pero utilizando los resultados obtenidos sobre el periodo 2003-2013. El algoritmo utilizado está basado en reglas aplicadas sobre arboles de decisión, que en función de los valores que toman los índices de vegetación se puede definir la categoría a la que pertenece cada pixel (pastura o área de cultivo, sistema secano o irrigado, un solo ciclo o más ciclos). No obstante para todo esto, previamente se requiere del uso de máscaras de áreas de cultivo (provenientes de trabajos de campo o imágenes de alta resolución) y un mapa de zonas agroecológicas. La plataforma sobre la cual se desarrollaron los análisis fue Google Earth Engine.

Clasificadores basados en análisis fractal: Shen et al. [9], buscaron determinar las principales etapas dentro del ciclo fenológico del cultivo de maíz para tres sitios en Estados Unidos (Iowa, Illinois y Nebraska) utilizando información proveniente de los sensores MODIS (250-m, series de tiempo de 10 días) y SPOT-VGT (1 km) calculando el índice de vegetación NDVI para el periodo 2002-2011. De esta manera,

el proceso consistió en segmentar inicialmente las imágenes utilizando una capa de áreas de cultivo con el fin de excluir del análisis las áreas que no sembraban maíz, posteriormente para las áreas seleccionadas se calcularon una serie de métricas de dispersión y dimensión fractal, para identificar los principales eventos dentro del ciclo de cultivo o etapas fenológicas, destacándose que las métricas de dimensión fractal son las que mejor representan los eventos ocurridos durante las épocas de cultivo, como las etapas de emergencia de plantas y las etapas de cosecha.

Clasificadores que toman en consideración la componente temporal: Belgiu & Csillik [1], aplicaron la metodología Time-Weighted Dynamic Time Warping (TWDWTW) para la clasificación de zonas de cultivos haciendo uso dos enfoques: el primero basado en píxeles y el segundo basado en objetos previamente segmentados; realizando una comparación de sus resultados con el algoritmo de clasificación Random Forest. Se escogieron tres zonas de prueba en Rumania, Italia y Estados Unidos, las cuales poseían diferentes climas, tipos de parcelas y cultivos. En general, la metodología TWDWTW superó en desempeño al modelo Random Forest en los tres sitios evaluados y de igual manera el enfoque basado en objetos tuvo mejores resultados en comparación con el enfoque basado en píxeles.

Por otro lado, Maus et al. [10], realizaron un estudio de patrones de uso de suelo en la Amazonía brasilera haciendo uso de la metodología Time-Weighted Dynamic Time Warping (TWDWTW) usando series de tiempo del índice de vegetación mejorado de MODIS 250-m 16 días, para el periodo que abarcó de Julio de 2000 hasta Junio de 2013. En cuanto a los resultados obtenidos, se lograron alcanzar niveles de precisión en la caracterización de los usos de suelo de 70-87 % logrando captar las estacionalidades en las series y los ciclos fenológicos de los cultivos.

Finalmente se presenta un enfoque de análisis donde se utilizaron redes neuronales con memoria de corto y largo plazo. RuBwurm & Korner [2], aplicaron una metodología de redes neuronales con memoria para la clasificación del uso de suelo en el norte de Munich (Alemania), usando datos provenientes del satélite Sentinel 2 para el periodo Diciembre de 2015 a Septiembre de 2016. Entre los principales resultados se destaca la clasificación con un nivel de precisión cercano al 100% de nubes, agua y nieve, mientras áreas de cultivo como maíz, remolacha y zonas de ruinas, también alcanzaron altos niveles de precisión.

IV. ESTUDIOS ENFOCADOS EN FRIJOL COMÚN

Pocos estudios se han llevado a cabo específicamente hablando para el cultivo de frijol común (*Phaseolus vulgaris*) en términos de identificación de áreas de cultivo y discriminación varietal, no obstante a continuación se presentan los principales avances en la materia, enfocados en trabajos de campo a nivel experimental.

Rajah et al. [5], abordó el tema de clasificación de variedades de frijol común. El objetivo principal del artículo consistió

en realizar una clasificación de tres variedades de frijol común en la población de Ukulinga (South Africa) de acuerdo a tres etapas fenológicas del cultivo y dos condiciones de riego (riego y seco) haciendo uso de imágenes hiper-espectrales tomadas con un espectro-radiómetro a 50 cm por arriba de la planta. Las tres variedades bajo consideración fueron Caledon, Ukulinga y Gadra. Una vez obtenidas las imágenes se aplicó un Análisis Discriminante de Mínimos Cuadrados Parciales para realizar la clasificación, donde se logró alcanzar un nivel de precisión en la clasificación del 68% en la etapa final correspondiente a formación de vainas y floración.

Por otro lado, Rajah et al. [6], abordó la identificación de la etapa fenológica más óptima para clasificación, para esto se usó la misma configuración de variedades y condiciones de campo que en [5]. No obstante, el objetivo aquí era identificar la etapa fenológica óptima del cultivo para predecir los rendimientos. Para esto se hizo uso de imágenes hiper-espectrales y un modelo de regresión de mínimos cuadrados parciales dispersos. En cuanto a los resultados obtenidos, las métricas de precisión en la clasificación fueron más altas en los lotes donde se aplicó riego en comparación con los lotes que no tuvieron sistema de riego, por otro lado, para la variedad Gadra la mejor etapa para estimar el rendimiento óptimo del cultivo fue la segunda etapa principal de crecimiento, mientras para las variedades Ukulinga y Caledon fue la séptima.

V. CONCLUSIONES

En términos generales la investigación concerniente a la identificación de áreas de cultivo de frijol desde el enfoque de uso de imágenes satelitales, es un área donde se puede aportar desde el punto de vista de la ingeniería de sistemas a partir del uso de algoritmos de clasificación que combinen la componente temporal e imágenes satelitales multi-espectrales o hiper-espectrales. Como se pudo evidenciar a lo largo de la revisión del estado del arte, no se ha llevado a cabo el primer estudio que este completamente dedicado a la identificación de zonas de cultivo de frijol y menos a la identificación de variedades tradicionales, siendo este un trabajo de gran interés para el área.

Las preguntas de investigación a desarrollar son:

- Cómo se pueden identificar variedades criollas de frijol respecto a las variedades de alto rendimiento en diferentes sistemas de cultivo usando información de sensores remotos satelitales?
- Cuáles son los principales elementos para clasificar correctamente las variedades criollas vs las variedades de alto rendimiento usando información proveniente de sensores remotos satelitales y algoritmos de clasificación?
- Cuál sería una metodología exacta y precisa para la identificación de áreas donde se siembran variedades criollas de frijol?
- Cuál sería un nivel de precisión aceptable para predecir la distribución de áreas donde se siembran variedades tradicionales de frijol en sitios de prueba?

Finalmente, los retos que se plantean consisten en: seleccionar una zona de prueba a partir de conocimiento de experto

donde se sepa que hay tanto variedades tradicionales como variedades de alto rendimiento de frijol (preferiblemente en Colombia), para dicha zona realizar la búsqueda de imágenes satelitales de uso libre o pensar en el uso de drones para tomar las primeras imágenes y desarrollar análisis preliminares, por otro lado, seleccionar la metodología de análisis de datos que más se ajuste a las necesidades o probar varias de ellas para evaluar su desempeño.

AGRADECIMIENTOS

El autor agradece al curso: Introducción a la Investigación (Segundo semestre, 2017), por las pautas brindadas para el desarrollo exitoso del presente trabajo.

REFERENCES

- [1] M. Belgiu and O. Csillik, "Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis," *Remote Sensing of Environment*, pp. –, 2017.
- [2] M. Ruwurm and M. Krner, "MULTI-TEMPORAL LAND COVER CLASSIFICATION WITH LONG SHORT-TERM MEMORY NEURAL NETWORKS," vol. XLII-1/W1, pp. 551–558.
- [3] P. Teluguntla, P. Thenkabail, J. Xiong, M. Gumma, R. Congalton, A. Oliphant, J. Poehnelt, K. Yadav, M. Rao, and R. Massey, "Spectral matching techniques (smts) and automated cropland classification algorithms (accas) for mapping croplands of australia using modis 250-m time-series (20002015) data," *International Journal of Digital Earth*, vol. 10, no. 9, pp. 944–977, 2017.
- [4] J. Xiong, P. S. Thenkabail, M. K. Gumma, P. Teluguntla, J. Poehnelt, R. G. Congalton, K. Yadav, and D. Thau, "Automated cropland mapping of continental africa using google earth engine cloud computing," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 126, no. Supplement C, pp. 225 – 244, 2017. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0924271616301575>
- [5] P. Rajah, J. Odindi, E. Abdel-Rahman, O. Mutanga, and A. Modi, "Varietal discrimination of common dry bean (*phaseolus vulgaris* l.) grown under different watering regimes using multitemporal hyperspectral data," *Journal of Applied Remote Sensing*, vol. 9, no. 1, 2015, cited By 5.
- [6] P. Rajah, J. Odindi, E. Abdel-Rahman, and O. Mutanga, "Determining the optimal phenological stage for predicting common dry bean (*phaseolus vulgaris*) yield using field spectroscopy," *South African Journal of Plant and Soil*, pp. 1–10, 2017, cited By 0; Article in Press.
- [7] B. Basso, D. Cammarano, and P. De Vita, "Remotely sensed vegetation indices: theory and application for crop management," vol. 1, pp. 36–53, 01 2004.
- [8] R. Sahoo, S. Ray, and M. R., "Hyperspectral remote sensing of agriculture," vol. 108, pp. 848–859, 03 2015.
- [9] Y. Shen, X. Liu, and X. Yuan, "Fractal dimension of irregular region of interest application to corn phenology characterization," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 4, pp. 1402–1412, April 2017.
- [10] V. Maus, G. Cmara, R. Cartaxo, A. Sanchez, F. M. Ramos, and G. R. d. Queiroz, "A time-weighted dynamic time warping method for land-use and land-cover mapping," vol. 9, no. 8, pp. 3729–3739.