An Assessment of MPEG-7 Visual Descriptors for Images of Maize Plagues and Diseases

J. Manjarrez-Sanchez

***Abstract*—Feature description is a fundamental process in the analysis of images for content-based retrieval and classification, among other tasks in which the image feature descriptor should have enough discriminative power to differentiate similar from dissimilar images according with a distance measure. Although several descriptors have been proposed for a variety of images, the challenge is their suitability to solve these tasks efficiently. One proposal is the set of standard MPEG-7 visual descriptors. We address their suitability to efficiently describe plagues and diseases in images of maize plants. The importance of this crop is its worldwide relevance for human and animal consumption. Experiments for similarity search queries using a set of distance measures, show that the Color Structure Descriptor with the Bray-Curtis distance is the most efficient and provides around 68% precision in most cases.**

***Index Terms*—Feature extraction, Image analysis, Machine vision.**

I. Introducción

El combate de plagas y enfermedades en cultivos es prioritario para garantizar el abasto para el consumo humano, animal e industrial. Esto puede hacerse más eficientemente con el apoyo de tecnologías, como las de visión por computadora, que ayuden a su diagnóstico rápido y oportuno. Este es un problema de investigación activo [1][2].

Una etapa fundamental en el reconocimiento automático de plagas y enfermedades en cultivos es la descripción eficiente de los rasgos visuales que las caracterizan. El reto es obtener una descripción visual abstracta, sea global o local, lo suficientemente robusta ante cambios de escala, iluminación, orientación y oclusiones que preserve el poder discriminativo que diferencía una descripción obtenida de una imagen de otra. Para resolver este problema, se ha probado con diferentes descriptores como SIFT, Patrones Locales Binarios, Filtros de Gabor y varios más [3]. Nosotros también probamos SIFT con los mismos experimentos y conjunto de imágenes que describimos en las siguientes secciones y la precisión máxima obtenida fue de 10%, lo cual es un indicativo del reto abierto de encontrar un descriptor que permita determinar la similaridad o disimilaridad entre imágenes que tienen una gran semejanza, como es el caso de las plagas y enfermedades en cultivos de maíz.

This work was supported by Tecnológico Nacional de México under a research grant.

J. Manjarrez-Sanchez. jorgems@acm.org.

Por otro lado, un segundo enfoque aprovecha la creciente capacidad de procesamiento de cómputo para utilizar técnicas computacionalmente más costosas [4][5] que se apoyan en Redes Neuronales y Aprendizaje Profundo y toman provecho de la capacidad automática para la extracción de características, que según el análisis de [6] es una de las principales razones de estos enfoques para ofrecer un mejor rendimiento comparado con otras técnicas. Sin embargo, para que esto sea posible, es necesario contar con una base de imágenes suficientemente grande y de una etapa de entrenamiento computacionalmente costosa.

Siguiendo el primer enfoque, existen diferentes descriptores de imágenes que nos proporcionan la cuantificación de sus características visuales sobresalientes y uno de los desafíos que se tienen es determinar el descriptor más eficiente para realizar la extracción de esas características debido a que existen varios factores que dificultan su selección. Con el fin de encontrar una propuesta para la descripción eficiente de las características visuales en imágenes de plagas y cultivos de maíz, en este trabajo analizamos el rendimiento de los seis descriptores visuales del estándar MPEG-7 (Moving Picture Coding Expert Group) [7][8], los cuales siguen siendo una referencia en la recuperación de imágenes por el contenido e incluso fueron utilizados para describir uno de los conjuntos de imágenes más grandes disponibles [9]. Se identifica cual es el descriptor más eficiente mediante experimentos en 800 imágenes de 16 diferentes enfermedades y plagas en cultivos de maíz, usando diferentes dimensiones para los descriptores y además, 7 distancias de similaridad. Estudios previos relacionados [10][11] hacen una evaluación de estos descriptores pero para un conjunto de imágenes generales y con un conjunto de distancias diferente, su conclusión es que las distancias de la familia de Minkowski son las que permiten mejores resultados. Otros estudios solamente consideran algunos de los descriptores visuales MPEG-7 para ciertos tipos de imágenes, lo que no permite hacer una evaluación exhaustiva. Por ejemplo en [12] se evaluó la pertinencia de usar MPEG-7 para el análisis automatizado de imágenes de endoscopias, encontrándose que los descriptores Color Escalable y Textura Homogénea son los más idóneos.

En las secciones que siguen, primero completamos la revisión de la literatura relacionada y relevante para nuestro trabajo y proporcionamos una explicación de los descriptores estudiados. Enseguida, en la sección III describimos el conjunto de imágenes, las funciones de distancia y los experimentos diseñados para analizar la eficiencia de los descriptores MEPG-7. En la sección IV presentamos los resultados de los experimentos y su correspondiente análisis. Finalmente, en la

sección V, presentamos las observaciones finales y trabajos futuros.

II. Antecedentes

Para los sistemas que analizan semánticamente el contenido de imágenes, básicamente hay dos tipos de enfoques, el aprendizaje estadístico o automático donde las imágenes se procesan para la extracción de descriptores y después esos descriptores se procesan a su vez, para inferir significado. El otro enfoque se basa en aprendizaje profundo donde las características descriptivas se extraen gradualmente en etapas o capas de procesamiento, agregando significado en cada una, siempre y cuando el modelo computacional se haya diseñado de manera adecuada. En nuestro caso, centramos esta exposición en el primer enfoque y particularmente en los trabajos que procesan imágenes de cultivos de maiz. Aunque una combinación de ambos puede significar mejoras en el rendimiento, como sugieren algunos experimentos que discutimos a continuación.

Aunque normalmente se extraen uno o varios descriptores locales o globales de algún tipo (color, textura, forma), se ha dicho que un solo tipo de descriptor no es suficiente y en aras de mejorar la precisión se ha experimentado con combinaciones de varios, por ejemplo, en [13] analizaron enfermedades en cultivos de peras, uvas y frambuesas. Experimentaron con 16 descriptores y los 6 mejores fueron combinados en grupos de 3 para mejorar el valor-F (F-measure, una combinación de precisión y recall que explicamos en la sección que sigue), obteniendo valores máximos de 0.4392 para pera, 0.3348 para uvas y 0.2230 para frambuesa. Entre los descriptores evaluados estaban algunos de MPEG-7 como el Color Escalable, Distribución de color e Histograma de Borde, pero todos usando únicamente la distancia euclidiana y el conjunto de imágenes sólo contenía 3 enfermedades para los frutos mencionados. De manera individual, estos fueron los que mejores resultados proporcionaron, pero su rendimiento fue diferente según el fruto de que se trataba, por ejemplo, el Color Escalable para pera permitió una precisión de 0.620, para uva de 0.469 y para frambuesa de 0.654. Esta variación estuvo presente en los 16 descriptores analizados, lo que enfatiza el hecho de que un descriptor no es eficiente en todos los tipos de cultivos o frutos, hay que buscar una solución de acuerdo al problema.

También, otra propuesta reciente [14] usa una combinación de descriptores, de cada imagen se extraen 22 descriptores: 7 de textura, 6 de color y 9 morfológicas, después de una etapa de preprocesamiento consistente en reducción de ruido, segmentación, transformación de color y binarización. Utiliza la distancia Euclidiana y reporta un 95% de precisión, aunque el conjunto de imágenes utilizado contiene únicamente 3 tipos de enfermedades, lo que reduce la complejidad del espacio de datos. Otro estudio que también combina SIFT, Histograma de Color y una modificación a LBP (Local Binary Pattern) y LGBP (Local Gabor Binary Pattern) es [3] que obtiene en promedio una precisión de 80% en un conjunto de 3 enfermedades en hojas de soya.

Dentro de la popularidad de la redes neuronales y áreas relacionadas, una propuesta reciente es [15], que mediante una red neuronal convolucional profunda logran en promedio un 95% de Top 1, que significa que la primera imagen de un conjunto de resultados es etiquetada correctamente. Sin embargo, las pruebas se hicieron para grupos individuales de 7 enfermedades únicamente de hojas. Pero además, cualquier otra nueva enfermedad o plaga implicaría re-entrenar la red propuesta, siempre y cuando se disponga de un número suficiente de imágenes representativas y de los recursos computacionales para dicho entrenamiento. En esta misma área están [16][17][18]. En [16] se extraen características de color y textura, pero su objetivo es comparar la calidad de la clasificación hecha con SVM y ANN, bajo éstas condiciones SVM fue superior. Trabaja con un conjunto de 5 clases de patologías en cultivos: hongos, bacterias, virus, nematodos y deficiencias. De cada imagen se obtienen varios descriptores para cada canal en el espacio de color RGB: media, varianza, desviación estándar, rango, después se analizan y se descartan los que están fuera de un rango establecido. Luego para extraer los descriptores de textura hace un análisis para la obtención de una matrix de co-ocurrencias de niveles de gris (GLCM, Gray-Leve Co-Ocurrence Matrix) de 30 características para finalmente seleccionar y trabajar con 5. Usando descriptores de color una patología es reconocida con una precisión entre 78% y 85%, mientras que usando textura fue entre 83% y 90%, combinadas entre 86% y 98%. En [17], se analizaron imágenes de 14 clases de 4 frutos, de los cuales 13 son de plagas más uno de frutos sanos, mediante redes neuronales convolucionales profundas (deep CNN) entrenadas con un conjunto de 4483 imágenes originales aumentadas mediante transformaciones para un total de 30880, los resultados obtenidos en los experimentos tienen una precisión entre el 91% y 98% pero para pruebas separadas de clases individuales, es decir un problema de clasificación binaria. Finalmente, en [18] se reconocen 3 tipos de enfermedades en las hojas de plantas de maíz diferenciándolas de las sanas mediante una CNN, con precisión entre 91% y 99% pero las pruebas también fueron para clases separadas.

De esta discusión, lo que se evidencia es que no hay un solo descriptor universal o enfoque de solución que sirva para todo tipo de plagas y enfermedades en cultivos o frutos, y que la combinación de dos o más es sugerida, pero primero deben encontrarse aquellos que de manera individual sean eficientes. En este trabajo, nosotros experimentamos con los descriptores visuales estándares MPEG-7 y analizamos su capacidad descriptiva para imágenes de plagas y enfermedades de cultivos de maíz.

El estándar MPEG-7 es una interfaz de descripción de contenidos multimedia. Para nuestro contexto son importantes los descriptores globales visuales de color y textura para imágenes [19]:

Estructura de color (Color Structure Descriptor, CSD).

Color escalable (Scalable Color Descriptor, SCD)

Distribución de color (Color Layout Descriptor, CLD)

Color dominante (Dominant Color Descriptor, DCD)

Textura homogénea (Homogeneous Texture Descriptor, HTD)

Histograma de borde (Edge Histogram Descriptor, EHD)

El descriptor CSD construye un histograma de colores HMMD (Hue Max Min Diff) presentes en una imagen mediante una ventana deslizante, de este modo, aunque es un descriptor global, los colores identificados son locales, lo que permite la fidelidad de la descripción y aunque dos imágenes tengan los mismos colores, su distribución las diferencia. Cada bin es un color diferente.

Color Escalable construye un histograma pero a partir del espacio de color HSV mediante la transformada Haar. Es escalable en cuanto al número de bins, que quiere decir que es variable o configurable como coeficientes de la ondeleta usada para codificar el histograma.

Distribución de Color. Mediante la Transformada del Coseno Discreto cuantifica pequeños bloques de la imagen que se recorren en zig-zag. En la especificación no se indica el método para seleccionar el color representativo de la región. Trata de capturar con la mayor precisión posible los colores representativos de la imagen.

Color Dominante como sugiere su nombre, permite encontrar un número determinado de colores representativos de una imagen. La cuantización es el porcentaje de la presencia del color en la imagen, es decir, el número de pixeles de la imagen correspondientes a ese color.

Para obtener la descripción de Textura Homogénea de la imagen se descompone en los canales perceptibles mediante la Transformada de Radon y Fourier 1D, cada valor del descriptor es un valor obtenido de un canal. El interés es capturar la direccionalidad, aspereza y las irregularidades en los patrones de las imágenes.

En el Histograma de Borde o llamado también de textura no homogénea se proporciona la distribución espacial de los bordes presentes en la imagen, por lo que es un descriptor global. Este descriptor consiste de un histograma de 80 bins que se construye al dividir la imagen en 16 bloques y obtiene una representación de cada borde presente en cada bloque. Aunque es invariante a traslaciones, por su naturaleza, este descriptor participa sólo en un experimento.

III. Metodología

Para evaluar los descriptores MPEG-7 se construyó una base de imágenes. Después, cada imagen se procesó con cada uno de los descriptores en diferentes dimensiones para obtener una representación numérica de sus propiedades visuales o vector característica. En seguida, mediante la ejecución de búsquedas de los vecinos más cercanos y diferentes funciones de distancia, se mide la calidad de los resultados que es posible obtener con cada combinación de descriptor y distancia, así podemos determinar cuál es la mejor combinación descriptor-distancia. De esta manera, analizamos la pertinencia de usar los descriptores MPEG-7 para imágenes de plagas y enfermedades en cultivos de maíz. A continuación se detallan estas etapas.

A. Conjunto de Imágenes

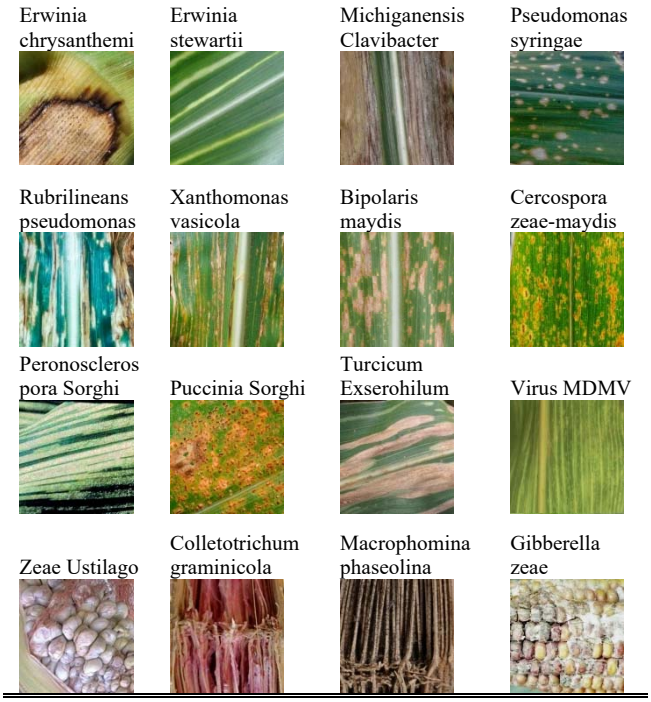
Para realizar los experimentos se recopilaron 800 imágenes, porque, para mejorar la validez de nuestros resultados, requerimos contar con una amplia variedad de enfermedades y plagas comunes en cultivos de maíz y visibles al ojo humano, son 16 y para cada una obtuvimos 50 imágenes y las dividimos en tres clases tomando en cuenta la parte de la planta que más afectan, según se detalla en la tabla I. Las imágenes se transformaron al formato jpg y a un tamaño de 500x500 pixeles.

TABLA I
CLASES DE PLAGAS Y ENFERMEDADES ANALIZADAS

Fruto	Hoja	Tallo
Gibberella zeae Zeae Ustilago	Bipolaris maydis, Cercospora zeae-maydis, Erwinia stewartii, Michiganensis Clavibacter, Peronosclerospora Sorghi, Pseudomonas syringae, Puccinia Sorghi, Rubrilineans pseudomonas, Turcicum Exserohilum, Virus, Xanthomonas vasicola	Colletotrichum graminícola, Erwinia chrysanthemi, Macrophomina phaseolina

Para dar una idea del tipo de problema que representa su correcta descripción visual, se presenta una muestra de las imágenes para cada una de las 16 clases en la tabla II. Como se puede apreciar, el parecido entre algunas de ellas es bastante, esta similaridad inter-clase es lo que complica su análisis aún para el experto y es el origen del reto computacional.

TABLA II
MUESTRA DE LAS PLAGAS Y ENFERMEDADES DE MAÍZ ANALIZADAS



B. Extracción de Descriptores

Cada una de las 800 imágenes se procesó para obtener los descriptores CSD, SCD, CLD, DCD, HTD y EHD. Para permitir agrupar por consultas y realizar una comparación justa entre ellos, se trató de obtener el vector característica de cada imagen en 32, 64, 80, 128 y 256 dimensiones o una cercana

donde era posible debido a la especificación del algoritmo de cada descriptor, con la implementación de MPEG7Fex [20]. Cada descriptor o vector característica también puede ser considerado como un punto en un espacio multidimensional.

C. Consultas de los K Vecinos más Cercanos

Para evaluar la eficiencia de cada descriptor se utiliza la búsqueda secuencial de los k vecinos más cercanos $k$nn, primero porque no es relevante desde el punto de vista de esta investigación la velocidad de procesamiento, es decir, solamente queremos evaluar la calidad de la descripción y hacer una búsqueda mucho más rápida mediante el uso de estructuras de indexación, en principio, sólo acelera la obtención de resultados y de todos modos cualquier estructura de indexación por arriba de cierta dimensionalidad se vuelve ineficiente debido, entre otras condiciones, por el fenómeno conocido como maldición de la dimensionalidad [21]. Ahora bien, usar una consulta de tipo $k$nn permite medir precisión, recall (a veces traducido como *exhaustividad* en la literatura en español), y además en este trabajo presentamos dos nuevas medidas que nos dan una idea más clara de la calidad de los resultados. Pudieran usarse métodos más complejos de búsqueda o calificación pero nuestro interés está únicamente en evaluar la capacidad descriptiva, porque es la etapa fundamental para construir cualquier tipo de aplicaciones.

Consulta knn. Sea M el conjunto de vectores característica $\mathbf{m}$ de dimensionalidad d, obtenidos al procesar la base de imágenes de tamaño $\mathbf{n}$. El objetivo de la consulta $k$nn es encontrar $V \subseteq M$ donde $|V| = k$ $\mathbf{m}$ vectores correspondientes a las imágenes más similares a la de una imagen de vector $\mathbf{q}$ mediante una función de distancia ∂:

$$V = \{\mathbf{m} \in M : \forall \mathbf{m} \in V, \forall \mathbf{z} \in M \setminus V, \partial(\mathbf{q}, \mathbf{m}) < \partial(\mathbf{q}, \mathbf{z})\}$$

□.

En la figura 1 se muestra el algoritmo para encontrar el conjunto V de los k descriptores más similares a uno $\mathbf{q}$ dado.

Algorithm 1: búsqueda kNN secuencial

Input : M, q
Output: V, donde $|V| = k$

```
foreach m_i en M do
    distancias[i] ← δ(q,m_i)
end
sort distancias
V ← select top-k en distancias
```

Fig. 1 Pseudocódigo para el algoritmo de la búsqueda secuencial kNN.

Las funciones de distancia ∂ [22] que usamos para medir la similaridad entre los vectores característica $\mathbf{m}$ y $\mathbf{q}$ se resumen en la tabla III.

D. Métricas de Evaluación

Para saber cuál es el mejor descriptor necesitamos cuantificar qué tan buenos son los resultados que permiten obtener. Las dos medidas tradicionales en un sistema de búsqueda de imágenes por similaridad son precisión P y recall R.

TABLA III
FUNCIONES DE DISTANCIA Y SU FÓRMULA PARA CALCULARLAS

Chebyshev	$\partial(\boldsymbol{q},\boldsymbol{m}) = \boldsymbol{max}_{\boldsymbol{i=1}}^{\boldsymbol{d}} \lvert \boldsymbol{q_i} - \boldsymbol{m_i} \rvert$
Manhattan	$\partial(q,m) = \sum_{i=1}^{d} \lvert q_i - m_i \rvert$
Euclideana	$\partial_2(q,m) = \left[\sum_{i=1}^{d} (q_i - m_i)^2\right]^{1/2}$
Canberra	$\partial(q,m) = \sum_{i=1}^{d} \frac{\lvert q_i - m_i \rvert}{(q_i + m_i)}$
Bray-Curtis	$\partial(q,m) = \frac{\sum_{i=1}^{d} \lvert q_i - m_i \rvert}{\sum_{i=1}^{d} (q_i + m_i)}$
Clark	$\partial(q,m) = \sqrt{\sum_{i=1}^{d} \left(\frac{\lvert q_i - m_i \rvert}{q_i + m_i}\right)^2}$
Pearson Chi Cuadrada	$\partial(q,m) = \sum_{i=1}^{d} \frac{\lvert q_i - m_i \rvert^2}{q_i}$
Neyman Chi Cuadrada	$d(q,m) = \sum_{i=1}^{d} \frac{\lvert q_i - m_i \rvert^2}{m_i}$

La precisión mide la cantidad de descriptores de imágenes relevantes en los resultados, mientras que recall es la fracción de descriptores de imágenes relevantes en la base de datos y que son incluidos en los resultados. Sea $|V_t|$ el número de vectores relevantes recuperados y $|V_f|$ el número de vectores irrelevantes recuperados y $|V_t'|$ el número de vectores relevantes pero que no se recuperaron, entonces:

$$P = \frac{|V_t|}{|V_t| + |V_f|}$$

$$R = \frac{|V_t|}{|V_t| + |V_t'|}$$

son las fórmulas para determinar precisión y recall respectivamente.

La idea es que con un buen descriptor cada imagen recuperada es relevante (P = 100%) y cada imagen relevante es recuperada (R = 100%). Evidentemente, recall está influenciado por el tamaño de k en este tipo de consultas, pero de todos modos, su cuantificación permite establecer un parámetro de evaluación de la pertinencia de los descriptores analizados para este tipo de imágenes. Además, otras medidas pueden derivarse a partir de estas dos. Pero lo estándar para medir la calidad de una búsqueda es la precisión.

Ahora bien, dado que se tiene una variedad de clases de enfermedades y plagas con una similaridad inter-clase considerable, es posible que en los resultados aparezcan imágenes de varias clases. Por esta razón introducimos dos nuevas medidas: *top-1* y *mixture*. Top-1 nos dice si la clase del vecino más cercano es igual a la de la consulta $\boldsymbol{q}$, y mixture nos indica la pureza de los resultados al contabilizar el número de clases involucradas en los $k$nn. Entonces, un descriptor ideal sería aquel que nos permitiría lograr resultados con Top-1 y un mixture=1 para cualquier consulta.

Finalmente, una quinta métrica, y aunque no es relevante

para los fines de este trabajo, medimos el tiempo de procesamiento en segundos *Tseg*, que puede dar una idea de la complejidad de la función de distancia y su relación con la dimensionalidad del descriptor.

IV. Resultados

En esta sección se describen los experimentos realizados y su análisis para evaluar el desempeño de los descriptores visuales del estándar MPEG-7. También se incluye una sección de discusión final, donde se relacionan los resultados obtenidos con trabajos relacionados.

A. Plataforma Experimental

Para las búsquedas *k*nn se seleccionaron aleatoriamente 10 imágenes de cada clase, para tener un total de 160 imágenes consulta. En total, se realizaron 5 experimentos usando las 8 distancias de similaridad de la Tabla III, por lo que los resultados presentados son el análisis de aproximadamente 25,600 consultas para descriptores de diferentes dimensiones de imágenes de 16 diferentes plagas y enfermedades de cultivos de maíz. Se implementó una plataforma parametrizable en Java 11 que permite la ejecución y visualización automática de experimentos para facilitar su análisis en un procesador Intel i7-5500U a 2.40 GHz con 16 GB RAM.

B. Experimentos y Análisis

Los experimentos se agruparon por dimensiones, con algunos descriptores no es posible parametrizar exactamente a una dimensión específica pero se usó el valor más cercano. El detalle de la configuración se muestra en cada experimento junto las gráficas del resumen de los porcentajes de rendimiento promedio alcanzado por cada uno de los descriptores involucrados para las correspondientes distancias y métricas; y después en una tabla, el detalle de los resultados obtenidos para el mejor descriptor. Esta, es suficiente información para apreciar el análisis utilizado para determinar la mejor opción de combinación de descriptor y función de distancia.

Se reportan resultados para k=10, que es con el que mejores valores se obtienen. Para k=16 la precisión promedio decreció de 65% a 60%. Para k=50 bajó a 48%. Esto es normal, la precisión es una métrica sensible al número de imágenes devueltas por la consulta. De todos modos, en una aplicación real, por ejemplo en una de recuperación de imágenes por el contenido (CBIR, Content Based Image Retrieval), las imágenes recuperadas están ordenadas por relevancia. El tamaño de k es dependiente de la aplicación. Presentar un conjunto grande de resultados (valores grandes de k) resulta útil en aplicaciones donde se permita al usuario navegar por los resultados y se asume que los primeros son los de mayor interés para el usuario. Esto puede combinarse, por ejemplo, con esquemas de retroalimentación para refinarlos. Para los fines de este trabajo, donde queremos encontrar los mejores parámetros para la descripción de imágenes basados en MPEG-7, k representa un indicio para el desarrollo de aplicaciones a partir de esto resultados.

Experimento 1 – 32 dimensiones. Los descriptores exactamente parametrizables fueron CSD, SCD y HTD. El CLD se aproximó a 33. En la Figura 2 se muestra el resumen de los porcentajes de precisión, recall y top-1 alcanzados con cada descriptor.

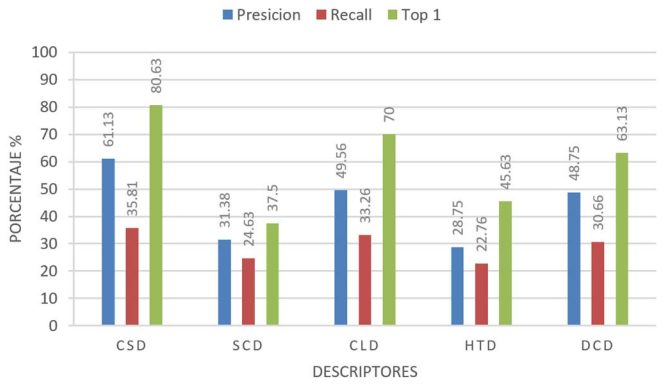


Fig. 2. Resumen de los porcentajes de rendimiento para los descriptores en 32 dimensiones.

Los mejores resultados se obtuvieron con el CSD con una precisión de 61.13%, recall 35.81% y el 80.63% de las veces la búsqueda knn era Top-1 de la misma clase, seguido de CLD que obtuvo una ligera ventaja de 0.81% en precisión comparado con DCD. Sin embargo, como se ve en la tabla IV para mixture el mejor resultado se obtuvo con SCD.

TABLA IV
Mejores Resultados de la Comparación de Descriptores de 32 Dimensiones

	Precisión	Recall	Top	Mixture	Tseg
Resultado	61.13%	35.81%	80.63%	1.6875	0.0003491
Descriptor	CSD	CSD	CSD	SCD	CSD
Distancia	Bray-Curtis	Bray-Curtis	Manhattan	Pearson Chi	Manhattan

Experimento 2 – 64 dimensiones. Los descriptores exactamente parametrizables fueron CSD y SCD. El CLD se aproximó a 63 y HTD a 62. Los resultados se visualizan en la figura 3.

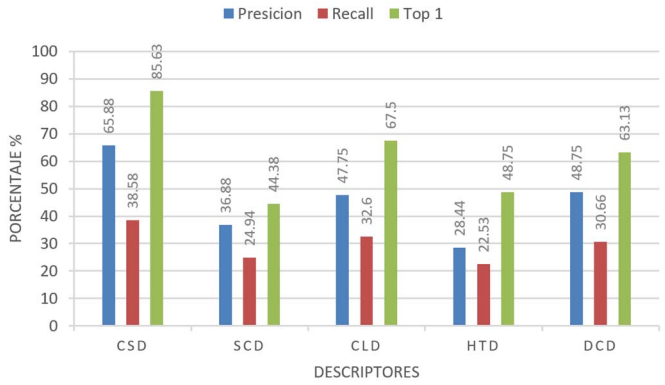


Fig. 3. Resumen de los porcentajes de rendimiento para los descriptores en 64 dimensiones.

CSD obtiene los mejores rendimientos, además en comparación con 32 dimensiones también mejora, ahora permite obtener el top-1 el 85.63% de las veces, con una precisión del 65.88%. Esto es de esperarse, una mayor dimensionalidad mejora la calidad de la descripción. Las

mejores distancias con Bray-Curtis y Manhattan y se muestra en la Tabla V.

TABLA V
MEJORES RESULTADOS DE LA COMPARACIÓN DE DESCRIPTORES DE 64 DIMENSIONES

	Precisión	Recall	Top	Mixture	Tseg
Resultado	65.88%	38.58%	85.63%	2.025	0.0005909
Descriptor	CSD	CSD	CSD	CSD	CSD
Distancia	Bray-Curtis	Bray-Curtis	Bray-Curtis	Manhattan	Manhattan

Experimento 3 – 80 dimensiones. Esta dimensionalidad se realizó para incluir al descriptor EHD que es de 80 dimensiones. Por equidad se compara únicamente con el CLD de 84 dimensiones. También se incluye el DCD.

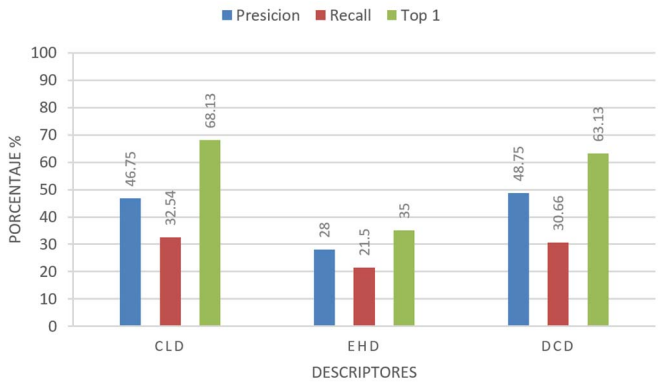


Fig. 4. Resumen de los porcentajes de rendimiento para los descriptores en 80 dimensiones.

TABLA VI
MEJORES RESULTADOS DE LA COMPARACIÓN DE DESCRIPTORES DE 80 DIMENSIONES

	Precisión	Recall	Top	Mixture	Tseg
Resultado	48.75%	32.54%	68.13%	2.5875	0.0006136
Descriptor	DCD	CLD	CLD	EHD	DCD
Distancia	Pearson Chi	Euclidiana	Euclidiana	Neyman Chi	Euclidiana

De la Figura 4 y la Tabla VI el descriptor con mejor rendimiento fue el CLD, sin embargo, comparando los porcentajes alcanzados en este experimento con los del experimento 2, ninguno de los tres supera los obtenidos por el CSD aunque éste sea de 64 dimensiones.

Experimento 4 – 128 dimensiones. Los descriptores analizados son CSD, SCD, CLD y DCD. Con esta dimensionalidad, los resultados entre los 4 descriptores son más uniformes, lo que hace pensar que para este tipo de imágenes, esta sería una opción aceptable para cualquiera de los descriptores. CSD continúa siendo el que mejor rendimiento exhibe con cualquier distancia, superando a SCD por un 8.75% en top-1 y en 7% en precisión. Esto se ilustra en la Figura 5 y Tabla VII.

Experimento 5 – 256 dimensiones. Este experimento es el de mayor dimensionalidad que realizamos e incluye a CSD, SCD y DCD. Con CSD se obtienen los porcentajes más altos de precisión, recall y top con la distancia de Bray-Curtis, de 68.13%, 39.30% y 86.25% respectivamente, aunque es tan sólo un 1.94% en precisión y 1.87% en top-1 por arriba de SCD, seguidos de DCD. Esto se ilustra en la Figura 6 y la Tabla VIII. Los mejores resultados se obtuvieron con la distancia Bray-Curtis.

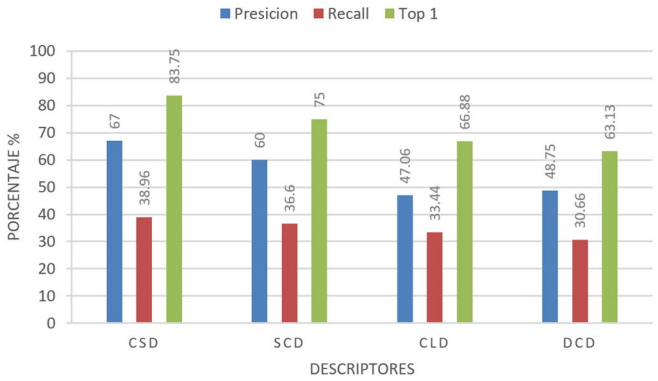


Fig. 5. Resumen de los porcentajes de rendimiento para los descriptores en 128 dimensiones.

TABLA VII
MEJORES RESULTADOS DE LA COMPARACIÓN DE DESCRIPTORES DE 128 DIMENSIONES

	Precisión	Recall	Top	Mixture	Tseg
Resultado	67.00%	38.96%	83.75%	1.9125	0.0006136
Descriptor	CSD	CSD	CSD	CSD	DCD
Distancia	Euclidiana	Bray-Curtis	Euclidiana Bray-Curtis	Neyman Chi	Euclidiana

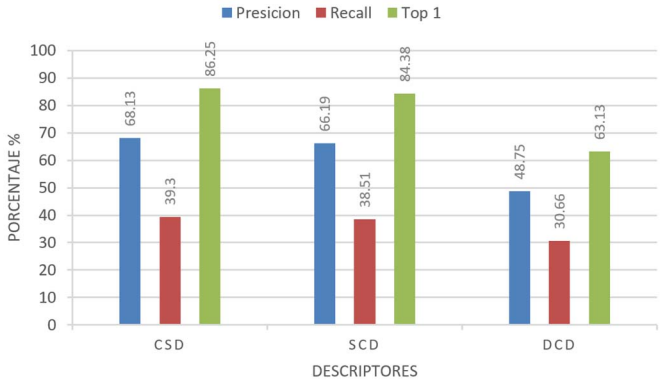


Fig. 6. Resumen de los porcentajes de rendimiento para los descriptores en 256 dimensiones.

TABLA VIII
MEJORES RESULTADOS DE LA COMPARACIÓN DE DESCRIPTORES EN 256 DIMENSIONES

	Precisión	Recall	Top	Mixture	Tseg
Resultado	68.13%	39.30%	86.25%	1.0812	0.0006136
Descriptor	CSD	CSD	CSD	SCD	DCD
Distancia	Bray-Curtis	Bray-Curtis	Bray-Curti	Neyman Chi	Euclidiana

C. *Discusión*

A manera de resumen, presentamos la Figura 7 con los valores obtenidos para precisión, recall y top 1 por el mejor descriptor de la ronda de experimentos: el CSD, en 64, 128 y 256 dimensiones. La diferencia entre 64 y 256 dimensiones en cuanto a precisión es de 2.25%, recall 0.72% y de top-1 2.5%. Aquí la reflexión es si realmente vale la pena invertir 4x más de almacenamiento y el correspondiente incremento en el cómputo de la función de distancia (2.5x aproximadamente para este caso) utilizada para obtener esa ligera ganancia en rendimiento.

Aunque los resultados son más puros: mixture de 1.75 vs 2.02.

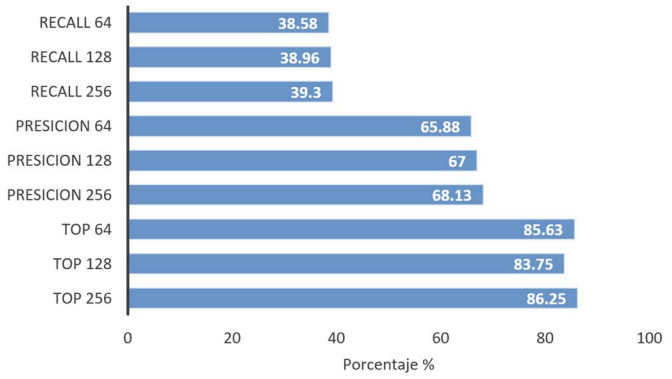


Fig. 7. Mejores resultados de la precisión, recall y top 1 del descriptor CSD de 64, 128 y 256 dimensiones.

Para soporte de éste análisis, en las Tablas IX y X, se muestran los promedios de los resultados obtenidos con CSD para las 160 consultas, primero en 64 y luego en 256 dimensiones. El valor de mixture es mejor si es menor, entonces los valores obtenidos con Bray-Curtis aunque no son los mejores, están muy cerca. De hecho son el segundo y tercer mejor resultado para 64 y 256 dimensiones respectivamente. Sin embargo, para precisión y recall sí fue la mejor distancia.

TABLA IX
RESULTADOS DE LOS EXPERIMENTOS PARA CSD DE 64 DIMENSIONES

	Precisión%	Recall%	Top%	Mixture%	Tseg
Chebyshev	63.94	37.54	80.63	2.09	0.00085
Manhattan	64.25	37.83	81.25	**2.02**	**0.00059**
Euclidiana	65.19	38.36	83.75	2.05	0.00088
Canberra	59.75	34.93	81.25	2.44	0.00164
Bray-Curtis	**65.88**	**38.58**	**85.63**	2.03	0.00061
Clark	59.31	34.34	80.00	2.43	0.00119
Pearson Chi	56.56	32.84	73.75	2.15	0.00102
Neyman Chi	41.44	27.30	62.50	2.35	0.00092

TABLA X
RESULTADOS DE LOS EXPERIMENTOS PARA CSD DE 256 DIMENSIONES

	Precisión%	Recall%	Top%	Mixture%	Tseg
Chebyshev	63.81	36.11	80.63	2.08	**0.00153**
Manhattan	66.06	37.70	85.63	2.01	0.00155
Euclidiana	67.19	38.66	85.63	1.93	0.0016
Canberra	63.44	37.18	83.13	2.22	0.00208
Bray-Curtis	**68.13**	**39.30**	**86.25**	1.95	0.00187
Clark	64.88	37.35	84.38	2.21	0.00252
Pearson Chi	56.31	31.89	76.25	2.23	0.00168
Neyman Chi	29.69	22.50	41.25	**1.75**	0.002

Como se analizó en las secciones I y II, existen otras propuestas de solución, específicamente los trabajos de [14] y [15]. Donde los resultados obtenidos tienen una gran precisión, pero el espacio de búsqueda no es complejo, porque para el caso de la [14] el conjunto de imágenes únicamente contiene 3 clases y el [15] 7 clases. Además el primero usa 3 descriptores combinados junto con Máquina de Soporte Vectorial (SVM, Support Vector Machine) y el segundo redes neuronales. En contraste, el conjunto de imágenes que se recopiló y se usa para evaluación en este trabajo, tiene 16 clases y los resultados presentados se obtienen a un menor costo computacional. Nótese que de cada imagen se extrajeron CSD, SCD, CLD, DCD, HTD y EHD, pero los resultados son obtenidos de manera individual, es decir, sin combinar los descriptores, y ganancias adicionales en la calidad de los resultados pueden venir de otras fuentes, como usar combinaciones de descriptores (posiblemente con otros), usar SVM o aprendizaje profundo, entre otras opciones.

V. CONCLUSIONES

El mejor descriptor para plagas y enfermedades en cultivos de maíz es el CSD, que en conjunto con la función de distancia Bray-Curtis, permite alcanzar los mejores resultados.

A trabajo futuro pudieran evaluarse otras distancias, combinando descriptores entre ellos y con técnicas, quizás más costosas computacionalmente para acelerar la búsqueda o usarse un mecanismo de clasificación, dependiendo de la aplicación, pero lo importante es que con el enfoque directo o más simple hemos podido evaluar la pertinencia de este estándar para la recuperación de imágenes de plagas de maíz. También pueden corroborarse los resultados aquí obtenidos con otras plagas y enfermedades de otros cultivos haciendo uso de otras imágenes[23].

Finalmente, como una contribución adicional y en pro de la publicación de resultados de investigación reproducibles y para permitir futuros desarrollos, el conjunto de imágenes usados en este trabajo está disponibles en DOI:/10.6084/m9.figshare.10314539.v3.

REFERENCIAS

[1] J. G. A. Barbedo, “A review on the main challenges in automatic plant disease identification based on visible range images,” *Biosyst. Eng.*, vol. 144, pp. 52–60, Apr. 2016.

[2] D. . Sena Jr, F. A. . Pinto, D. . Queiroz, and P. . Viana, “Fall Armyworm Damaged Maize Plant Identification using Digital Images,” *Biosyst. Eng.*, vol. 85, no. 4, pp. 449–454, Aug. 2003.

[3] J. K. Patil and R. Kumar, “Analysis of content based image retrieval for plant leaf diseases using color, shape and texture features,” *Eng. Agric. Environ. Food*, vol. 10, no. 2, pp. 69–78, Apr. 2017.

[4] K. P. Ferentinos, “Deep learning models for plant disease detection and diagnosis,” *Comput. Electron. Agric.*, vol. 145, pp. 311–318, Feb. 2018.

[5] E. C. Too, L. Yujian, S. Njuki, and L. Yingchun, “A comparative study of fine-tuning deep learning models for plant disease identification,” *Comput. Electron. Agric.*, Mar. 2018.

[6] A. Kamilaris and F. X. Prenafeta-Boldú, “Deep learning in agriculture: A survey,” *Comput. Electron. Agric.*, vol. 147, pp. 70–90, Apr. 2018.

[7] T. Sikora, “The MPEG-7 visual standard for content description-an overview,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 6, pp. 696–702, Jun. 2001.

[8] B. S. Manjunath, J.-R. Ohm, V. V. Vasudevan, and A. Yamada, “Color and texture descriptors,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 6, pp. 703–715, Jun. 2001.

[9] P. Bolettieri *et al.*, “CoPhIR: a Test Collection for Content-Based Image Retrieval,” May 2009.

[10] H. Eidenberger and Horst, “Distance measures for MPEG-7-based retrieval,” in *Proceedings of the 5th ACM SIGMM international workshop on Multimedia information retrieval - MIR ’03*, 2003, p. 130.

[11] H. Eidenberger, “Evaluation and analysis of similarity measures for content-based visual information retrieval,” *Multimed. Syst.*, vol. 12, no. 2, pp. 71–87, Nov. 2006.

[12] M. T. Coimbra and J. P. S. Cunha, “MPEG-7 visual descriptors - Contributions for automated feature extraction in capsule endoscopy,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 5, pp. 628–636, May 2006.

[13] Z. Piao *et al.*, “Performance analysis of combined descriptors for similar crop disease image retrieval,” *Cluster Comput.*, vol. 20, no. 4,

pp. 3565–3577, Dec. 2017.
[14] E. Alehegn, “Ethiopian maize diseases recognition and classification using support vector machine,” *Int. J. Comput. Vis. Robot.*, vol. 9, no. 1, p. 90, 2019.
[15] X. Zhang, Y. Qiao, F. Meng, C. Fan, and M. Zhang, “Identification of Maize Leaf Diseases Using Improved Deep Convolutional Neural Networks,” *IEEE Access*, vol. 6, pp. 30370–30377, 2018.
[16] J. D.Pujari, R. Yakkundimath, and A. S. Byadgi, “SVM and ANN Based Classification of Plant Diseases Using Feature Reduction Technique | International Journal of Interactive Multimedia and Artificial Intelligence,” *Int. J. Interact. Multimed. Artif. Intell.*, vol. 3, no. 7, pp. 6–14, 2016.
[17] S. Sladojevic, M. Arsenovic, A. Anderla, D. Culibrk, and D. Stefanovic, “Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification,” *Comput. Intell. Neurosci.*, vol. 2016, 2016.
[18] M. Sibiya, M. Sumbwanyambe, M. Sibiya, and M. Sumbwanyambe, “A Computational Procedure for the Recognition and Classification of Maize Leaf Diseases Out of Healthy Leaves Using Convolutional Neural Networks,” *AgriEngineering*, vol. 1, no. 1, pp. 119–131, Mar. 2019.
[19] B. S. Manjunath, T. Sikora, and P. Salembier, *Introduction to MPEG-7 : multimedia content description interface*. Wiley, 2002.
[20] M. Bastan, H. Cam, U. Gudukbay, and O. Ulusoy, “Bilvideo-7: an MPEG-7- compatible video indexing and retrieval system,” *IEEE Multimed.*, vol. 17, no. 3, pp. 62–73, 2010.
[21] K. Beyer, J. Goldstein, R. Ramakrishnan, and U. Shaft, “When Is ‘Nearest Neighbor’ Meaningful?,” Springer, Berlin, Heidelberg, 1999, pp. 217–235.
[22] V. Tyagi, *Content-Based Image Retrieval : ideas, influences, and current trends.* Springer Verlag Singapore, 2018.
[23] J. Garcia Arnal Barbedo *et al.*, “Annotated Plant Pathology Databases for Image-Based Detection and Recognition of Diseases,” *IEEE Lat. Am. Trans.*, vol. 16, no. 6, pp. 1749–1757, Jun. 2018.

Jorge Manjarrez-Sánchez Obtuvo el grado de Doctor en Ciencias de la Computación en la Université de Nantes, Francia, trabajando en la optimización del procesamiento de consultas en bases de datos paralelas. Es profesor-investigador en el Tecnológico Nacional de México, también es profesor PRODEP. Sus áreas de interés son Scalable Machine Learning, Parallel and Distributed Data Management y Programación Paralela y Concurrente.