

Métricas estructurales para evaluar la similitud de texturas naturales

Alma Alheli Pedro Pérez¹, Raúl Pinto Elias², Jasiel Hassan Toscano Martinez¹

¹ Universidad de la Sierra Juárez, Oaxaca,
México

² Centro Nacional de Investigación y Desarrollo Tecnológico, Cuernavaca Morelos,
México

almaalheli@unsj.edu.mx, toscano@unsj.edu.mx, rpinto@cenidet.edu.mx

Resumen. En este artículo se evalúa la similitud de imágenes de texturas naturales con cambios de escala y rotación, mediante la aplicación de métricas estructurales. Las métricas aplicadas son: *Complex-Wavelet Structural Similarity Metric (CW-SSIM)* y *Structural Texture Similarity Metric (STSIM)*. Primero se realizó una descomposición de las imágenes en sub-bandas. Con la información generada se extrajeron los descriptores y se aplicaron las fórmulas de comparación y las estrategias de unión para obtener un valor de similitud total. En los experimentos con imágenes de texturas con cambios por rotación, las métricas CW-SSIM y STSIM reportan valores promedio de desempeño de: 0.4857 y 0.9735 respectivamente. Con respecto a las pruebas con imágenes de texturas con cambios de escala, las métricas CW-SSIM y STSIM reportan valores promedio de: 0.7002 y 0.9840. Con base en los resultados se determinó que la métrica STSIM es la más recomendable para valorar la similitud de imágenes de texturas.

Palabras clave: Similitud de imágenes, métrica de similitud, texturas naturales, escala, rotación.

Structural Metrics for Evaluating Natural Textures Similarity

Abstract: This article evaluates the similarity of images of natural textures with changes of scale and rotation, through the application of structural metrics. The applied metrics are: *Complex-Wavelet Structural Similarity Metric (CW-SSIM)* and *Structural Texture Similarity Metric (STSIM)*. First a decomposition of the images in sub-bands was realized. With the generated information the descriptors were extracted and the comparison formulas were applied and the strategies of union to obtain a value of total similarity. In experiments with texture images with rotation changes, the CW-SSIM and STSIM metrics showed average values of: 0.4857 and 0.9735 respectively. With respect to texture image tests with scale changes, the CW-SSIM and STSIM metrics showed average values of: 0.7002

and 0.9840. Based on the results it was determined that the STSIM metric is the most advisable to evaluate the similarity of texture images.

Keywords: Images similarity, similarity metric, natural textures, scale, rotation.

1. Introducción

En el campo de la visión artificial existen aplicaciones que requieren evaluar la similitud entre imágenes de texturas para realizar tareas de clasificación, recuperación de imágenes con base en el contenido, compresión de imágenes, entre otras aplicaciones [1]. Estas aplicaciones usan métricas o funciones que indican en forma numérica qué tan similar o disímil es una imagen en comparación con otras. En el caso de la clasificación y recuperación, la comparación se realiza entre imágenes contenidas en una misma base de datos [2]. Por ejemplo, existen sistemas médicos que realizan la recuperación de imágenes tomográficas y las clasifican de acuerdo a su modelo de enfermedad [3].

En cuanto a la compresión, se compara la imagen original contra la imagen comprimida [2]. Un ejemplo claro, se observa con el aumento de la cantidad de imágenes tomográficas en un hospital. El aumento de los datos genera nuevas formas de archivar la información, por esta razón, se aplica la compresión de las imágenes para reducir la cantidad de datos. Sin embargo, al comprimir la imagen, se corre el riesgo de deteriorar la calidad. Es por ello, que se requieren métricas de similitud que evalúen el grado de pérdida de información de una imagen a diferentes grados de compresión, para asegurar que la imagen mantenga la calidad necesaria [4].

En la literatura existen investigaciones que abordan la temática de las métricas de semejanza para texturas [5], [6], éstas demuestran un alto nivel de eficacia, en el tema de las métricas desde el punto de vista objetivo y subjetivo, analizan diversas características en función de la estructura de la imagen. Cabe mencionar que estos trabajos se concretan en la clasificación y recuperación de imágenes [7], tareas en las cuales es de suma importancia describir objetos o regiones presentes en las imágenes empleando la textura [8].

Las texturas naturales suelen compartir características similares lo cual provoca que una imagen sea catalogada como de otra clase y no a la que pertenece, es el caso de la textura “árbol” que es semejante a la del “pasto”. Además de estos inconvenientes inherentes a las texturas naturales, existen otros aspectos no controlados, como cambios de escala, rotación, traslación, cambios de contraste e intensidad, entre otros [8]. Estos aspectos se presentan de forma prominente en las texturas naturales porque están constituidas por patrones espaciales cuasi-aleatorios difíciles de preservar ante diversos cambios. Para mejorar algunas de las tareas en el campo de la visión artificial tales como la clasificación de imágenes, es importante contar con métricas de semejanza para la valoración de texturas naturales que consideren las invariancias a diversos cambios que se presentan en las imágenes.

Este trabajo está estructurado de la siguiente manera: en la sección 2 se mencionan algunos trabajos relacionados al tema propuesto, en la sección 3 se describen de forma breve las métricas de similitud utilizadas, en la sección 4 se describe la metodología de la propuesta de solución, en la sección 5 se presentan los resultados obtenidos y finalmente en la sección 6 se presentan las conclusiones derivadas de este trabajo.

2. Trabajos relacionados

Las métricas de similitud de imágenes se clasifican de acuerdo al enfoque de evaluación en tradicionales y estructurales. Las métricas tradicionales evalúan la similitud entre dos imágenes en una base punto a punto. Estas usan modelos explícitos de bajo nivel de la percepción humana y se aplican en el dominio espacial o en el dominio de la frecuencia. Las métricas estructurales utilizan información estructural para realizar comparaciones de la imágenes [6].

En [9] Zhou implementa la métrica tradicional MSE mostrando sus ventajas y desventajas. Esta métrica ha sido criticada por graves deficiencias cuando se emplea en la evaluación de imágenes. El mal rendimiento de la métrica MSE se debe a que emplea mediciones punto a punto y no considera los aspectos que causan distorsiones en las imágenes. Debido al mal rendimiento de la métrica, surgen alternativas interesantes como las métricas SSIM (Structural Similarity Metric) y CW-SSIM (Complex-Wavelet Structural Similarity Metric).

Zujovic [5] realiza una evaluación del rendimiento de las métricas de similitud de texturas aplicadas a la compresión de imágenes sin pérdida estructural. Para realizar la evaluación, se generaron distorsiones en las texturas, estas modelaron las variaciones que ocurren en las texturas naturales. Se evaluaron métricas tradicionales como PSNR, así como métricas con un enfoque estructural como: SSIM, CW-SSIM, STSIM (Structural Texture Similarity Metric) y STSIM-2 Structural Texture Similarity Metric-2). De acuerdo a la experimentación se observó que las métricas estructurales son las más adecuadas para evaluar la similitud de imágenes de texturas sobre todo cuando se trabaja con imágenes de texturas con diversas rotaciones y deformaciones.

Finalmente en [1], Pappas presenta algunas métricas de semejanza de texturas y sus aplicaciones haciendo énfasis en la compresión de imágenes y en la recuperación basada en el contenido, en particular de imágenes de texturas naturales [10]. Expone el uso de métricas tradicionales como PSNR (Peak Signal-to-Noise Ratio) y métricas que incorporan el estudio de la estructura de la imagen, por ejemplo la métrica SSIM. Esta sirve como base para desarrollar la métrica STSIM. En la experimentación las métricas estructurales presentaron un mejor rendimiento en la recuperación de imágenes.

3. Métricas de similitud estructural

Una métrica de similitud estructural incorpora información estructural en las comparaciones de la imagen. Dicha información se basa en estadísticas locales extraídas de las imágenes o de las sub-bandas de estas imágenes [6]. En el estudio de la semejanza de imágenes, las métricas de similitud por lo regular se aplican al análisis de la textura.

Una de las cualidades más importantes de las métricas de similitud estructural, es su invariancia a cambios no estructurales, como el contraste y la intensidad. Así mismo, invariantes a pequeñas traslaciones, rotaciones y cambios de escala que son detectables pero que no afectan la calidad percibida de la imagen. Estas métricas se aplican en el dominio espacial o de la frecuencia. Cuando se implementan en el dominio espacial se obtiene invariancia a cambios de luminancia y contraste. Cuando se implementan en el dominio de la frecuencia son tolerantes a pequeños desplazamientos espaciales,

rotaciones o cambios de escala [6]. A continuación se describen cada uno de los elementos que componen a las métricas CW-SSIM y STSIM.

3.1. Métrica de semejanza estructural (CW-SSIM)

La métrica CW-SSIM (Complex-Wavelet Structural Similarity Metric) requiere que la imagen X y Y a comparar sean descompuestas en sub-bandas. Esto se realiza aplicando una pirámide orientable con determinado número de orientaciones y escalas [11]. Los descriptores locales de esta métrica se extraen de los coeficientes de las sub-bandas (ecuación 1,2 y 3) y después se calculan los términos de luminancia, contraste y estructura (ecuación 4,5 y 6). Finalmente los valores de estas fórmulas se combinan y se obtiene un valor de semejanza total [5] mediante la ecuación 7.

Media de los coeficientes de sub-banda:

$$\mu_x^m = \frac{1}{W} \sum_{(i,j) \in W} X^m(i,j). \quad (1)$$

Desviación estándar de los coeficientes de sub-banda:

$$\sigma_x^m = \sqrt{\frac{1}{W-1} \sum_{(i,j) \in W} (X^m(i,j) - \mu_x^m)^2}. \quad (2)$$

Covarianza de los coeficientes de sub-banda:

$$\sigma_{xy}^m = \frac{1}{W-1} \sum_{(i,j) \in W} (X^m(i,j) - \mu_x^m)(Y^m(i,j) - \mu_y^m), \quad (3)$$

donde x, y y son las dos imágenes a comparar. i, j son los índices de los coeficientes en el dominio transformado. m es el índice de sub-bandas. W es el tamaño de la sub-banda. μ_x^m y μ_y^m son las medias de las dos sub-bandas. σ_x^m y σ_y^m son desviaciones estándar de las dos sub-bandas, estas se elevan al cuadrado para obtener las varianzas.

Luminancia de los coeficientes de sub-banda:

$$l_{x,y}^m = \frac{2\mu_x^m \mu_y^m + C_0}{(\mu_x^m)^2 + (\mu_y^m)^2 + C_0}. \quad (4)$$

Contraste de los coeficientes de sub-banda:

$$c_{x,y}^m = \frac{2\sigma_x^m \sigma_y^m + C_1}{(\sigma_x^m)^2 + (\sigma_y^m)^2 + C_1}. \quad (5)$$

Estructura de los coeficientes de sub-banda:

$$s_{x,y}^m = \frac{\sigma_{xy}^m + C_2}{\sigma_x^m \sigma_y^m + C_2}. \quad (6)$$

Estrategía de unión:

$$q_{CW-SSIM}^m(x,y) = (l_{x,y}^m)^\alpha (c_{x,y}^m)^\beta (s_{x,y}^m)^\gamma. \quad (7)$$

Las constantes α , β y γ toman el valor de 1. Los valores de $C_0=6.5025$, $C_1=58.5225$ y $C_2=C_1/2$ se definen con base en los experimentos realizados por [12]. Los descriptores y fórmulas de comparación toman valores de 0 a 1, excepto la estructura, ésta toma valores de -1 a 1. Los valores que se obtienen de aplicar la estrategia de unión van de 0 a 1. Cuando existe mayor similitud entre las imágenes, este valor es próximo a 1.

3.2. Métrica de semejanza estructural de textura (STSIM)

La métrica STSIM (Structural Texture Similarity Metric) es una medida que elimina por completo las comparaciones punto a punto que dependen de los descriptores locales de la imagen. Esto se refiere a reemplazar el término estructura empleado en la métrica CW-SSIM por autocorrelaciones horizontal y vertical. Las imágenes se descomponen en sub-bandas utilizando una pirámide orientable. Los descriptores (ecuación 1,2) y los términos para comparar (ecuación 4,5, 8, 9) se calculan en toda la sub-banda (ventana global) o a través de una pequeña ventana deslizante que recorre la sub-banda. Al final se genera un promedio de la misma [6]. Una vez calculados los términos de comparación, se aplica la ecuación 10 y 11 para obtener las autocorrelaciones horizontal y vertical. Finalmente se emplea la ecuación 12 para obtener un valor de similitud total.

Coefficiente de autocorrelación horizontal:

$$p_x^m(0,1) = \frac{\frac{1}{W} \sum_{(i,j) \in W} (X^m(i,j) - \mu_x^m)(X^m(i,j+1) - \mu_x^m)}{(\sigma_x^m)^2} \quad (8)$$

Coefficiente de autocorrelación vertical:

$$p_x^m(1,0) = \frac{\frac{1}{W} \sum_{(i,j) \in W} (X^m(i,j) - \mu_x^m)(X^m(i+1,j) - \mu_x^m)}{(\sigma_x^m)^2}, \quad (9)$$

donde x , y y son las dos imágenes a comparar. i, j son los índices de los coeficientes en el dominio transformado. m es el índice de sub-bandas. W es el tamaño de las dimensiones de la sub-banda. μ_x^m y μ_y^m son las medias de las dos sub-bandas. σ_x^m y σ_y^m son desviaciones estándar de las dos sub-bandas, estas se elevan al cuadrado para obtener las varianzas.

Autocorrelación horizontal:

$$c_{x,y}^m(0,1) = 1 - 0.5 |\rho_x^m(0,1) - \rho_y^m(0,1)|^p. \quad (10)$$

Autocorrelación vertical:

$$c_{x,y}^m(1,0) = 1 - 0.5 |\rho_x^m(1,0) - \rho_y^m(1,0)|^p. \quad (11)$$

Los valores (0,1) y (1,0) representan el recorrido en la dirección horizontal y vertical de la sub-banda respectivamente. Los valores que se obtienen a partir de la estrategia de unión están en un rango de 0 a 1. Un valor de similitud alto estará más cercano a 1.

Estrategia de unión:

$$Q_{STSIM,x,y}^m = (I_{x,y}^m)^{1/4} (c_{x,y}^m)^{1/4} (c_{x,y}^m(0,1))^{1/4} (c_{x,y}^m(1,0))^{1/4}. \quad (12)$$

4. Metodología

La metodología propuesta involucra los siguientes pasos: selección de un conjunto de imágenes de una base de datos, una descomposición de las imágenes mediante la aplicación de una pirámide orientable compleja, de esta descomposición se procede a extraer la información de cada una de las sub-bandas generadas para posteriormente aplicar las métricas y obtener los valores de similitud. En la Fig. 1 se detalla el esquema general con las etapas y los elementos que intervienen.

1) Base de datos de imágenes. En el paso uno, se eligen de una base de datos las imágenes a comparar. La dimensión de las imágenes debe ser de 128 x 128 píxeles con extensión .JPG. No existe restricción en el tamaño de las imágenes. Sin embargo, éstas deben cumplir con la repetitividad de los patrones. Por convención se determinó usar la dimensión empleada por [2].

2) Descomposición en sub-bandas. En este paso, cada una de las imágenes se descompone en un conjunto de sub-bandas, empleando una pirámide orientable compleja. Primero se realiza una división de la imagen en sub-bandas paso alto y paso bajo, aplicando filtros de paso alto y paso bajo, después la sub-banda paso bajo se divide en un conjunto de sub-bandas de paso de banda orientadas y una sub-banda de paso bajo la cual es sub-muestreada por un factor de dos en las direcciones X y Y. Este proceso se sigue de forma recursiva hasta completar los cuatro niveles de escala y las cuatro orientaciones como se muestra en la Fig. 2.

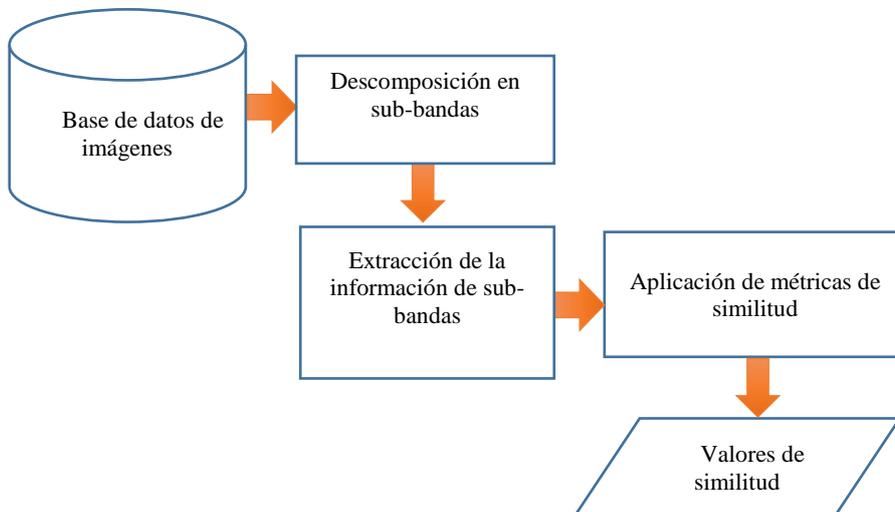


Fig. 1. Arquitectura general del sistema, describe las etapas y los elementos que intervienen.

3) Extracción de la información de sub-bandas. En este paso, cada una de las sub-bandas extraídas es almacenada de forma individual en archivos de texto, por cada imagen se genera un archivo que contiene el nombre de los 18 archivos de texto correspondiente a las 18 sub-bandas en las que se descompuso la imagen. Los archivos

que contienen la información de sub-bandas se almacenan en matrices para la aplicación del siguiente paso.

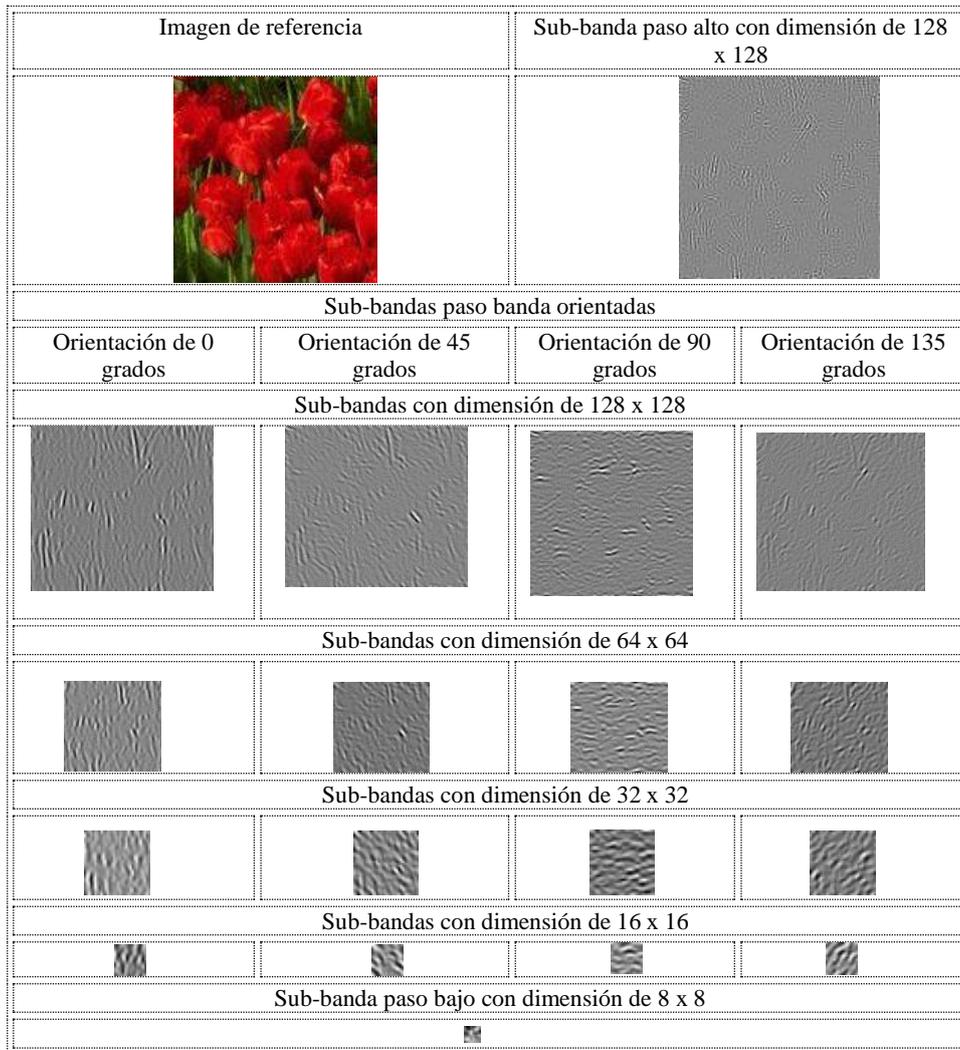


Fig. 2. Descomposición de una imagen en sub-bandas.

4) Aplicación de métricas de similitud. Primero se realiza la extracción de descriptores de las matrices, éstos dependerán de la métrica implementada. Una vez extraídos los descriptores, se aplican fórmulas de comparación de descriptores entre matrices de sub-bandas, de la imagen de referencia y la imagen a comparar.

Por último, se aplica una estrategia de unión para obtener un valor de similitud total, este valor se obtiene al multiplicar los resultados de las fórmulas de comparación de descriptores. En la sección 3.1 y 3.2 descritas anteriormente, se describen las métricas implementadas.

5. Experimentos y resultados

5.1. Descripción del conjunto de datos

La base de imágenes se integró de imágenes de texturas naturales de las categorías: flores, montañas (descargadas del sitio web Corbis corporation) y césped (obtenidas mediante captura personal). El número de imágenes por cada categoría es de 50, a estas imágenes se les denomina imágenes fuente, estas imágenes se procesaron para obtener distintas representaciones de cada una con cambios de escala y rotación.

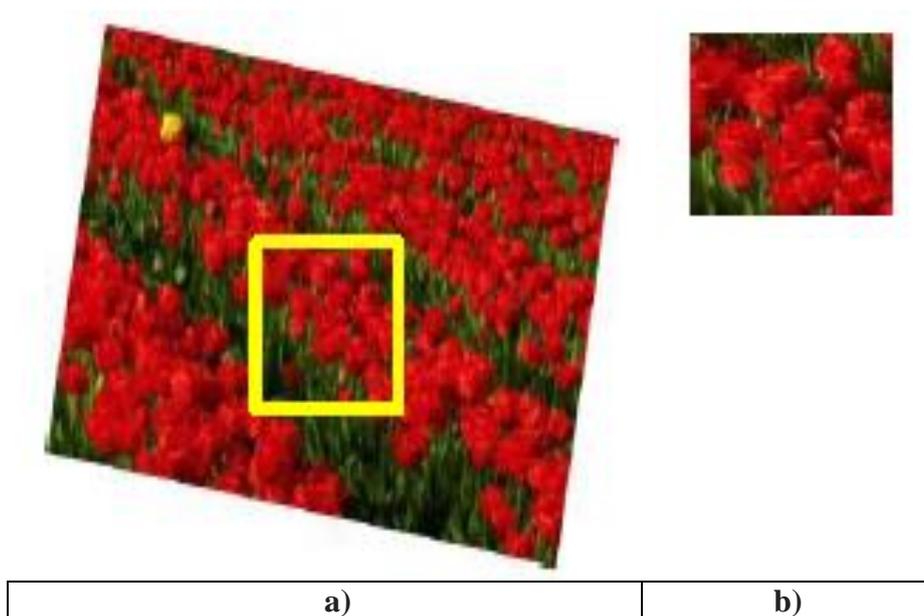


Fig. 3. Proceso de rotación y extracción de un segmento de la imagen fuente. a) Imagen fuente con cambio por rotación y b) Segmento extraído de la parte central de la imagen Fuente.

En los cambios de rotación por cada imagen fuente se tienen 21 representaciones con diferentes grados de rotación, generando así 1050 imágenes por cada categoría y haciendo un total de 3150 imágenes. Para obtener estas imágenes, primero se extrae un segmento de la imagen fuente de 128 x 128 píxeles (este segmento se denomina imagen de referencia), después se rota la imagen y se extrae un segmento de 128 x 128 píxeles como se muestra en la Fig. 3. Los segmentos extraídos forman un conjunto de 21 imágenes (Fig. 4).

En los cambios de escala, por cada imagen fuente, se tienen 11 representaciones con diferentes niveles de escalado, generando así 550 imágenes por cada categoría y haciendo un total de 1650 imágenes. Para obtener estas imágenes, primero se realiza un proceso de extracción de un segmento de la imagen fuente de 128 x 128 píxeles (imagen de referencia), después se escala la imagen fuente y se extrae un segmento de 128 x 128 píxeles como se muestra en la Fig. 5, los segmentos extraídos forman un conjunto de

11 imágenes. En la Fig. 6 se muestra un ejemplo del conjunto de imágenes generado de la categoría flores.

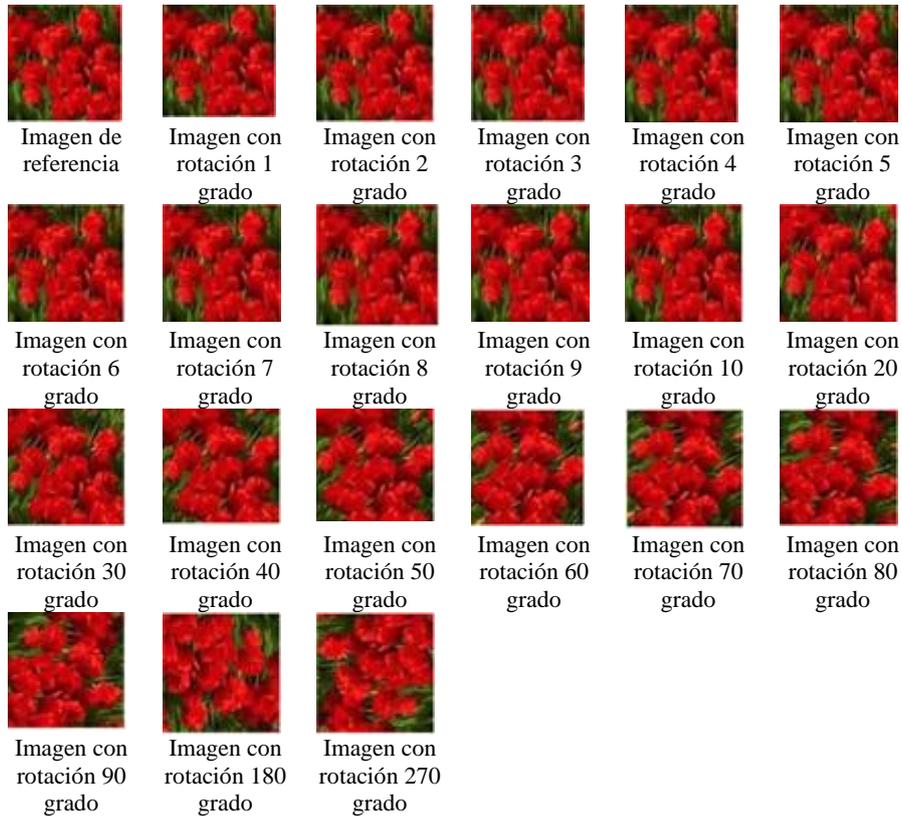


Fig. 4. Imágenes de ejemplo correspondientes a cada magnitud de rotación de la categoría flores.

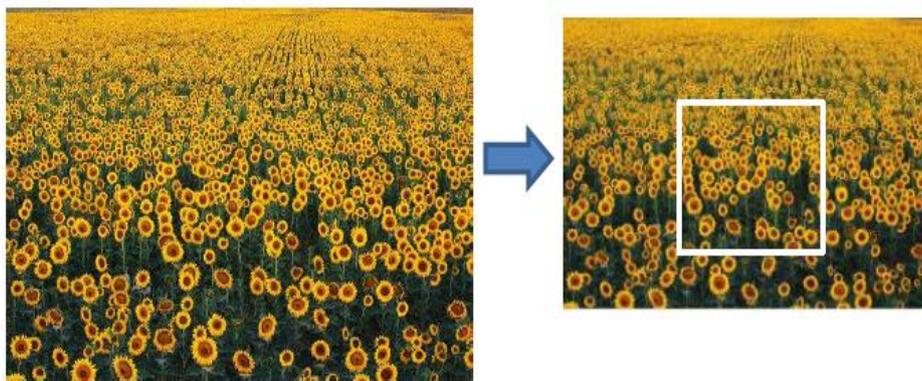


Fig. 5. Proceso de escalado y extracción de un segmento de la imagen fuente.

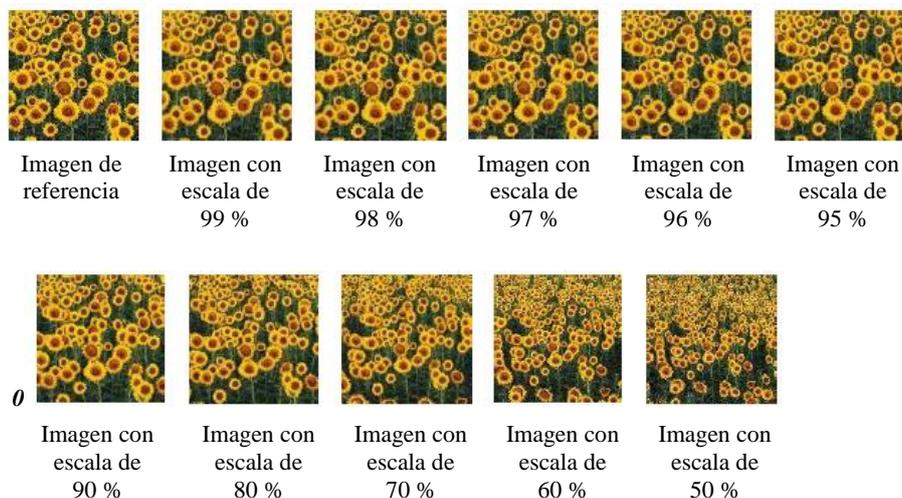


Fig. 6. Imágenes de ejemplo correspondientes a cada escala de la categoría flores.

5.2. Pruebas y resultados

En la parte experimental se realizaron cuatro pruebas. En la primera se utilizaron imágenes de las tres categorías con cambios de rotación aplicando la métrica CW-SSIM. En la segunda prueba, se emplearon las mismas imágenes de la prueba anterior aplicando la métrica STSIM. En la tercera prueba se aplicó la métrica CW-SSIM, para ello se emplearon imágenes de las tres categorías con cambios de escala. Finalmente en la última prueba se utilizaron imágenes de la prueba tres aplicando la métrica STSIM. A partir de los datos obtenidos se realizó un promedio por cada magnitud de rotación y de escala, con la finalidad de obtener un valor de rendimiento total de cada una de las métricas como se muestra en la Tabla 1.

Tabla 1. Resultados de la aplicación aplicación de las métricas CW-SSIM y STSIM a imágenes de texturas naturales de las categorías: flores, césped y montañas.

Cambios en las imágenes	Métrica	Categoría Flores	Categoría Césped	Categoría Montañas	Promedio
Rotación	CW-SSIM	0.5126	0.3948	0.5499	0.4857
Rotación	STSIM	0.9843	0.9854	0.9508	0.9735
Escala	CW-SSIM	0.7008	0.6541	0.7458	0.7002
Escala	STSIM	0.9881	0.9893	0.9748	0.9840

En los experimentos con imágenes de texturas naturales con cambios por rotación, las métricas CW-SSIM y STSIM reportan valores promedio de desempeño de: 0.4857 y 0.9735 respectivamente. En los experimentos con imágenes de texturas naturales con cambios de escala, las métricas CW-SSIM y STSIM reportan valores promedio de desempeño de: 0.7002 y 0.9840 respectivamente.

6. Conclusiones

Con base en los resultados obtenidos, se determinó que la métrica STSIM es la más recomendable para valorar la similitud de imágenes de texturas naturales con cambios de escala y rotación, porque presenta valores promedio de similitud más altos en todas las categorías de imágenes de texturas naturales empleadas.

La métrica STSIM se basa en una descomposición en sub-bandas. Éstas se emplean para la medición de los elementos de luminancia y contraste utilizados en la métrica CW-SSIM y se reemplaza el término estructura por autocorrelaciones horizontal y vertical. Si bien, la luminancia y el contraste siguen comparando las medias y las varianzas, las autocorrelaciones horizontal y vertical comparan las autocovarianzas. Éstas proporcionan información de la direccionalidad horizontal y vertical para mejorar la comparación de texturas.

En este trabajo, la aplicación de las métricas de similitud se sitúa en un nivel de descripción ya que los valores obtenidos describen la semejanza entre las imágenes. Únicamente se analiza el desempeño de las métricas con base en los valores de semejanza obtenidos. Sin embargo, si se requiere realizar tareas de reconocimiento en el área de visión artificial, los valores de semejanza pueden ser empleados por algoritmos de clasificación y segmentación, entre otros.

Las métricas de similitud se aplican en diversas áreas científicas, entre ellas destaca la medicina, en donde se requieren sistemas para clasificar, recuperar y comprimir imágenes.

Referencias

1. Pappas, T. N., Neuhoff, D. L., de Ridder, H., Zujovic, J.: Image Analysis: Focus on Texture Similarity. *Proceedings of the IEEE*, Vol. 101, No. 9, pp. 2044–2057 (2013)
2. Zujovic, J.: *Perceptual Texture Similarity Metrics*. Northwestern University (2011)
3. La Serna-Palomino, N., Contreras, W., Ruiz, M. E.: Procesamiento Digital de Textura: Técnicas utilizadas en Aplicaciones actuales de CBIR. *Revista de Investigación de Sistemas e Informática*, Vol. 7, No. 1, pp. 57–64 (2010)
4. Kim, K. J., Kim, B., Mantiuk, R., Ritcher, T., Lee, H., Kang, H. S., Seo, J., Lee, K. H.: A Comparison of Three Image Fidelity Metrics of Different Computational Principles for JPEG2000 Compressed Abdomen CT Images. Vol. 29, No. 8, pp. 1496–1503 (2010)
5. Zujovic, J., Pappas, T. N., Neuhoff, D. L., van Egmond, R., de Ridder, H.: Subjective and objective texture similarity for image compression. *Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1369–1372 (2012)
6. Zujovic, J., Pappas, T. N., Neuhoff, D. L.: Structural Texture Similarity Metrics for Image Analysis and Retrieval. *IEEE Transactions on Image Processing*, Vol. 22, No. 7, pp. 2545–2558 (2013)
7. Patil, S., Talbar, S.: *Content Based Image Retrieval Using Various Distance Metrics*. Lecture Notes in Computer Science, Vol. 6411, Springer-Verlag, Berlin Heidelberg (2012)
8. Avellaneda-Avellaneda, D.: *Caracterización de Texturas Naturales*. Centro Nacional de Investigación y Desarrollo Tecnológico (2009)
9. Zhou, W., Bovik, A.C.: Mean squared error: Love it or leave it? A new look at Signal Fidelity Measures. *Signal Processing Magazine*, Vol. 26, No. 1, pp. 98–117 (2009)

Alma Alheli Pedro Pérez, Raúl Pinto Elias, Jasiel Hassan Toscano Martinez

10. Troncoso-Rey, P. A.: *Indexado y Recuperación de Imágenes por Contenido*. Centro Nacional de Investigación y Desarrollo Tecnológico (2007)
11. An, J., Zhang, X.: *Robust Image Matching Method Based on Complex Wavelet Structural Similarity*. *Advances in Computer Science*, Springer-Verlag, Berlin Heildeberg, Vol. 215 (2011)
12. Zhou, W., Bovik, A. C., Sheikh, H. R., Simoncelli, E. P.: *Image Quality Assessment: From Error Visibility to Structural Similarity*. *IEEE Transactions on Image Processing*, Vol. 13, No. 4, pp. 600–612 (2004)