

Métodos para el reconocimiento automático de objetos combinando modelos de apariencia y relaciones espaciales y jerárquicas

Autora principal

Annette Morales González-Quevedo¹.

Otro autor

Edel García Reyes¹.

Colaboradores

Dr.C Luis Enrique Sucar Succar², Ms.C. Niusvel Acosta Mendoza¹, Dr.C. Andrés Gago Alonso , Dr.C. José Eladio Medina Pagola¹, Lic. Eric Javier Hernández Saura¹.

Entidad ejecutora principal

¹Centro de Aplicaciones de Tecnologías de Avanzada (División de Investigaciones CENATAV, ¹ Datys Soluciones Tecnológicas).

Entidades participantes

²INAOE, México.

AutorU para correspondencia

Annette Morales González-Quevedo

Dirección: 7^aA #21406 e/ 214 y 216, Reparto Siboney, Playa, C.P. 12200

Fax: (+) 537.273.0045

correo electrónico: amorales@cenatav.co.cu

Resumen

Esta investigación teórica aborda el problema del reconocimiento automático de objetos en imágenes digitales bidimensionales (2D). Esta temática se enmarca dentro del campo de la Visión por Computadora, y comprende el reconocimiento de objetos genéricos o específicos presentes en las escenas para lograr una comprensión del contenido visual de las mismas de manera automatizada. En particular, en esta investigación se propone una nueva representación jerárquica de la escena, combinando información visual y espacial, mediante pirámides irregulares con atributos en los vértices y en las aristas. Esta combinación de información extraída de las imágenes para representarlas, es un nuevo aporte al conocimiento en este campo de investigación. A partir de esta representación, se proponen dos nuevos métodos de reconocimiento de objetos en escenarios simples y un nuevo método de reconocimiento de objetos en escenarios complejos. El primer método plantea el reconocimiento de clases de objetos como un problema de correspondencia de grafos, el segundo enfoca el reconocimiento de objetos específicos como un problema de clasificación utilizando el paradigma de bolsa de palabras y el tercero, como un problema de re-etiquetado relajado mediante Campos Aleatorios de Markov. Cada uno de estos métodos va encaminado a resolver diferentes problemas que se presentan en este campo de investigación y los tres constan de novedad científica en este ámbito. Los experimentos se realizaron sobre bases de datos de comparación internacionales para mostrar los beneficios de la nueva representación y de los nuevos métodos propuestos, con respecto a los métodos ya existentes en la literatura. Los resultados obtenidos en este trabajo de tesis están dirigidos a la creación de sistemas propios para la video-protección inteligente, la recuperación de imágenes por contenido y el análisis de flujos de información, fundamentalmente encaminados a la defensa y orden interior de nuestro país, dentro del MININT. La novedad científica de este trabajo está avalada principalmente por 2 artículos publicados en revistas de impacto internacional, por 5 artículos publicados en memorias de eventos de impacto internacional y especializados en el tema, así como por la Tesis de Doctor en Ciencias Técnicas de la autora principal.

Comunicación corta

1. Descripción de la problemática existente

El reconocimiento automático de objetos es una parte crucial en el proceso de análisis de las escenas y consiste en identificar instancias de una categoría de objetos en una imagen. En esta área de investigación los esfuerzos han estado encaminados a desarrollar representaciones y algoritmos que permitan reconocer objetos específicos o genéricos en imágenes ante distintas condiciones.

El problema principal en la interpretación de imágenes y en el reconocimiento de objetos es la alta variabilidad intra-clase (objetos pertenecientes a la misma clase son muy diferentes entre sí) y la baja variabilidad inter-clases (objetos de distintas clases son muy similares), causadas por varios factores como condiciones de iluminación, pose de los objetos, localización de la cámara, oclusiones parciales y fondos no relacionados con el objeto o ruidosos. De manera general, el reconocimiento automático de objetos debe seguir dos pasos fundamentales [4]: (1) la extracción de rasgos visuales de bajo nivel y (2) la interpretación de estos rasgos como conceptos de alto nivel. La falta de correspondencia entre los rasgos de bajo nivel y los conceptos de alto nivel es conocida como brecha semántica y es el problema fundamental en esta área de investigación [8].

De manera general, el reconocimiento de objetos se puede dividir en dos grandes grupos en cuanto a la forma de reconocimiento [5]. Uno es el reconocimiento de objetos específicos, que busca identificar un objeto en particular (ej. la Torre Eiffel, el rostro del Che, el carro de Pepe) y la otra es el reconocimiento de clases de objetos, que pretende reconocer instancias de una categoría genérica pertenecientes a una misma clase conceptual (ej. edificios, tazas, carros, tomates).

Existen aplicaciones de reconocimiento visual de objetos en las cuales se combinan enfoques de reconocimiento de objetos en escenarios simples (solo hay un objeto de interés en una imagen) y de objetos en escenarios más complejos (como imágenes naturales donde interactúan varios objetos a la vez), para obtener una mayor eficacia en la comprensión de la escena.

Entre los enfoques más populares en esta área se encuentran los basados en representaciones regulares de la imagen, sobre las cuales extraen rasgos y realizan operaciones de clasificación. En este caso se encuentran las ventanas deslizantes [3] y las bolsas de palabras visuales [1]. También están los enfoques basados en partes, que intentan crear modelos de los objetos basados en sus partes y sus relaciones espaciales, pero los objetos y sus partes están delimitados por regiones rectangulares. No obstante, una región rectangular que delimite un objeto no constituye una forma efectiva de localización y representación de los mismos para muchas tareas.

La segmentación automática de regiones presenta aún muchas deficiencias para separar un objeto del fondo, por lo cual la mayoría de los enfoques de reconocimiento de objetos no utilizan segmentaciones irregulares. Sin embargo, la ausencia de segmentación no permite la localización de objetos individuales que apunten a un contexto más específico. Los objetos que difieren en cuanto a forma y no en apariencia no pueden ser modelados sin segmentación. Además, cuando la identidad de un objeto es ambigua son importantes las pistas contextuales ofrecidas por objetos cercanos [4]. Los algoritmos más recientes han abandonado las ventanas deslizantes y han incluido la segmentación de la imagen entre los pasos de la detección. Los tres ganadores de la competencia ILSVRC 2013 [2] en la categoría de detección de objetos utilizan segmentación para generar posibles localizaciones de los objetos que se buscan. Para mejorar las representaciones basadas en segmentación, se ha propuesto mayormente la utilización de diferentes segmentaciones para una misma imagen [6].

Otro aspecto a considerar para abordar el problema de la brecha semántica, es cómo tener en cuenta el contexto de cada objeto o región de la imagen. Se ha planteado que las relaciones espaciales entre los componentes de la imagen juegan un rol fundamental en el proceso de interpretación de las mismas que realizan los humanos [7].

La gran mayoría de los algoritmos actuales se centran en atacar uno de estos problemas a la vez, descartando la posibilidad de combinar estos distintos tipos de información que pueden extraerse de las imágenes.

2. Novedad Científica

Sobre la base de lo explicado anteriormente y los problemas detectados en la literatura relacionada, en el marco de esta investigación, se diseñó una nueva representación de las imágenes, y tres algoritmos de reconocimiento de objetos que hacen uso de la misma. Específicamente:

- 1- Se diseñó una nueva representación jerárquica de las imágenes, basada en pirámides irregulares de grafos. Las pirámides irregulares de grafos están formadas por un Grafo de Adyacencia de Regiones por nivel. En estos grafos, los vértices representan regiones en la imagen y las aristas representan relaciones de adyacencia entre cada par de regiones. Una pirámide irregular es, por tanto, un conjunto de grafos, donde cada grafo es construido reduciendo el grafo del nivel anterior. En la representación propuesta, a cada grafo se le añaden atributos en los vértices y en las aristas para lograr mayor representatividad de las regiones y relaciones subyacentes. Los vértices reciben como atributos rasgos visuales que describen la apariencia de cada región. Como atributo de las aristas se propuso un nuevo descriptor espacial que consiste en un vector binario que recepciona la presencia o no de un conjunto de relaciones básicas entre regiones. La combinación de la información visual y espacial en una estructura jerárquica de la imagen, es un aspecto novedoso en la representación propuesta. Se propusieron también medidas de similitud para hallar aproximaciones entre estructuras de este tipo. La novedad científica de esta representación está validada por los artículos publicados [9-17].

- 2- Se propuso un nuevo algoritmo de reconocimiento de objetos en escenarios simples (condiciones controladas), basado en correspondencia de grafos. Para podar el espacio de búsqueda se utilizan las medidas de similitud visual y espacial propuestas para la nueva representación, para discriminar vértices y aristas que sean muy distintos entre sí. Además, con el fin de reducir el tiempo de procesamiento, se propuso un nuevo método para seleccionar los niveles de la pirámide irregular cuya segmentación subyacente preserve la mayor cantidad de bordes presentes en la imagen. Con este nuevo método, es posible aplicar el algoritmo de correspondencia de grafos solamente en los mejores niveles de la pirámide. En los experimentos realizados sobre bases de datos de competencia internacional, este algoritmo mostró mejor desempeño para la tarea de reconocimiento de clases de objetos, pues la correspondencia de estructuras empleando aproximaciones visuales y espaciales permite una mejor generalización de los objetos. Con este algoritmo se logró obtener un 90.2% de eficacia global en el experimento de reconocimiento de clases de objetos, que superó en 2.6% al mejor resultado reportado en la literatura. El aporte científico de este resultado queda validado por las publicaciones [9, 11,16].
- 3- Se propuso un nuevo algoritmo de reconocimiento de objetos en escenarios simples (condiciones controladas), basado en bolsa de palabras visuales. La nueva propuesta es tener en cuenta las relaciones espaciales en la creación del vocabulario visual. Concretamente, las palabras visuales, en lugar de ser creadas a partir de rasgos visuales individuales (como lo hacen los enfoques encontrados en la literatura relacionada), son creadas a partir de subgrafos frecuentes en las imágenes. Para encontrar un vocabulario de subgrafos se decidió utilizar técnicas de minería de datos, específicamente la minería de Subgrafos Frecuentes Aproximados (FAS, por sus siglas en inglés), pues tiene en cuenta posibles distorsiones de los datos y no hace una correspondencia exacta, sino aproximada de los mismos a la hora del descubrimiento de los subgrafos frecuentes. Luego de obtenido el vocabulario compuesto de FAS, se representa cada imagen mediante las ocurrencias de los mismos. Se utiliza un clasificador para entrenar y clasificar las imágenes utilizando estos rasgos. En los experimentos realizados, este algoritmo mostró mejor desempeño para la tarea de reconocimiento de objetos específicos, pues los subgrafos frecuentes caracterizan ocurrencias de configuraciones visuales particulares de cada objeto. En los experimentos sobre bases de datos de competencia internacional, este algoritmo mostró mejor eficacia (99.4 %) que los encontrados en el estado del arte para dicha tarea. Además, este algoritmo explora un campo de investigación muy poco desarrollado actualmente, que consiste en la combinación de técnicas de clasificación de imágenes con técnicas de minería de datos. La novedad científica queda avalada por las publicaciones [10, 12].

- 4- Se desarrolló un nuevo algoritmo para el reconocimiento de objetos en escenarios complejos (condiciones no controladas), aprovechando la representación propuesta para introducir relaciones jerárquicas en el etiquetado de regiones mediante el uso de Modelos Gráficos Probabilísticos, específicamente, Campos Aleatorios de Markov (MRF, por sus siglas en inglés). Cada nivel de la pirámide irregular es representado por un MRF y se propone un esquema para resolver cada MRF, empleando información jerárquica de la pirámide. Posteriormente este algoritmo fue extendido, proponiendo un nuevo esquema que incluye mejorar la segmentación inicial de la pirámide, utilizando la información semántica brindada por el etiquetado resultante de los MRFs en cada nivel. De esta forma se combinan la segmentación y anotación semántica de las imágenes de forma simultánea e iterativa, lo cual es un aspecto novedoso con respecto a la literatura, donde solo se encuentran enfoques dirigidos a resolver estos problemas de forma independiente. Esta nueva propuesta obtuvo una eficacia que mejoró en 6.1 % al mejor resultado reportado en la literatura internacional. El aporte científico de este nuevo método queda validado por las publicaciones [13-15].

3- Impacto del resultado

Para validar la aplicabilidad de los resultados de este trabajo en situaciones reales, se realizaron pruebas en videos provenientes de cámaras de videovigilancia de nuestro país, lo cual fue el resultado de un trabajo de diploma en Licenciatura en Ciencias de la Computación [17]. En este trabajo se desarrolló una plataforma experimental para hacer pruebas en videos reales.

El aporte teórico de este trabajo en la rama de Visión por Computadora radica esencialmente en la nueva representación creada para las imágenes, donde se emplean relaciones jerárquicas y espaciales entre las partes que conforman las mismas. Tanto esta nueva representación como los tres nuevos algoritmos de reconocimiento automático de objetos desarrollados para maximizar las potencialidades de la misma, además de poseer novedad científica en el país, son contribuciones en esta área del conocimiento a nivel internacional, como lo avalan las publicaciones en revistas y congresos de alto prestigio internacional en este tema [9-16].

Estos nuevos algoritmos son la base teórica del conocimiento para resolver el problema de la detección y reconocimiento de objetos en distintos escenarios de videovigilancia, lo cual es una necesidad del Ministerio del Interior. En este sentido, la significación práctica de este trabajo viene dada en primer lugar, por la posibilidad de usar los algoritmos propuestos en el desarrollo de sistemas propios que precisen el reconocimiento de objetos en imágenes o videos. Las aplicaciones que utilizan reconocimiento de objetos son muy costosas, además de estar limitadas en muchos casos a dominios específicos. Precisamente, por la importancia de este campo para la defensa y el orden interior de cualquier país, gran parte de los sistemas de videovigilancia no aparecen libremente disponibles en el mercado. En especial esta investigación básica fue dirigida a buscar soluciones para los problemas relacionados con videovigilancia que se presentan en el MININT, donde una de las necesidades es reconocer automáticamente objetos como vehículos, peatones, señales, carreteras, edificaciones, animales, aviones, embarcaciones, etc., según el escenario donde se desarrolle la vigilancia. Las soluciones propuestas serán introducidas en el módulo Analytics del sistema Xyma Vision que desarrolla la empresa Datys para la videovigilancia, con lo cual se le agregaría al sistema el valor agregado de reconocer y/o recuperar automáticamente objetos de interés en los escenarios que se analizan.

Otra necesidad ministerial que se beneficia de estos nuevos resultados, es el análisis de flujos de información, en particular, de correos electrónicos, donde es preciso conocer el contenido de las imágenes que son enviadas o recibidas. Esto puede ser introducido en el sistema GEMA, también desarrollado por la empresa Datys, donde actualmente solo se realiza análisis de flujos de información en forma de texto. La posibilidad de realizar un análisis de flujos de imágenes es un valor agregado importante para este sistema.

Otras posibles aplicaciones de interés de estos resultados pueden ser el reconocimiento de escenas, la recuperación de imágenes y videos a partir de una base de datos almacenada y la anotación de imágenes y videos por su contenido semántico, para facilitar la indexación y búsquedas.

Referencias Bibliográficas

- [1] G. Csurka, C. R. Dance, L. Fan, J. Willamowski, and C. Bray. "Visual categorization with bags of keypoints". In ECCV, 1-22, 2004.
- [2] <http://www.image-net.org/challenges/LSVRC/2013/results.php> (Consultado Enero/2014).
- [3] P. Viola and M. J. Jones. "Robust real-time face detection". *Int. J. Comput. Vision*, 57(2):137–154, 2004.
- [4] Sven J. Dickinson, Ales Leonardis, Bernt Schiele, and Michael J. Tarr. "Object Categorization: Computer and Human Vision Perspectives", Cambridge, University Press, New York, NY, USA, 1st edition, 2009.
- [5] Kristen Grauman and Bastian Leibe. "Visual Object Recognition. Synthesis Lectures on Artificial Intelligence and Machine Learning". Morgan & Claypool Publishers, 2011.
- [6] Caroline Pantofaru, Cordelia Schmid, and Martial Hebert. "Object recognition by integrating multiple image segmentations". In *Proceedings of the 10th European Conference on Computer Vision: Part III, ECCV '08*, pages 481–494, Berlin, Heidelberg, Springer-Verlag, 2008.
- [7] Alper Aydemir, Kristoffer Sjöö, John Folkesson, Andrzej Pronobis, and Patric Jensfelt, "Search in the real world: Active visual object search based on spatial relations", In *ICRA*, pages 2818–2824. IEEE, 2011.
- [8] Anne-Marie Tousch, St'ephane Herbin, and Jean-Yves Audibert, "Semantic hierarchies for image annotation: A survey". *Pattern Recogn.*, 45(1):333–345, January 2012.
- [9] A. Morales-González, E. García-Reyes: "Simple object recognition based on spatial relations and visual features represented using irregular pyramids". *Multimedia Tools and Applications*. 63(3): 875-897, 2013.
- [10] A. Morales-González, N. Acosta-Mendoza, A. Gago-Alonso, E. B. García-Reyes, J. E. Medina-Pagola. "A new proposal for graph-based image classification using frequent approximate subgraphs". *Pattern Recognition*, 47(1): 169-177, 2014.
- [11] A. Morales-González, E. García-Reyes, "Assessing the Role of Spatial Relations for the Object Recognition Task". 15th Iberoamerican Congress on Pattern Recognition (CIARP 2010), Springer LNCS 6419, pp. 549-556, November 08-11, São Paulo, Brazil, 2010.

- [12] N. Acosta-Mendoza, A. Morales-González, A. Gago-Alonso, E. B. García-Reyes, J. E. Medina-Pagola, "Image Classification Using Frequent Approximate Subgraphs", In Proceedings of 17th Iberoamerican Congress on Pattern Recognition, CIARP 2012, Springer LNCS 7441, pp. 292–299, 2012.
- [13] A. Morales-González, E. García-Reyes, L. E. Sucar, "Hierarchical Markov Random Fields with Irregular Pyramids for Improving Image Annotation", In Proc. of 13th Iberoamerican Conference on Artificial Intelligence, IBERAMIA 2012, Springer LNAI 7637, pp. 521–530, 2012.
- [14] A. Morales-González, E. García-Reyes, L. E. Sucar. "Improving Image Segmentation for Boosting Image Annotation with Irregular Pyramids". In: Proceedings of 18th Iberoamerican Congress on Pattern Recognition, CIARP 2013, Springer LNCS, Vol. 8258, Part I, 2013. Berlin Heidelberg, Springer-Verlag, 2013, pp. 399-406. 2013.
- [15] A. Morales-González, E. García-Reyes, L. E. Sucar. "Unsupervised Segmentation Evaluation for Image Annotation". In Proceedings of the 10th International Conference on Computer Vision Theory and Applications, VISAPP 2015, Berlin, Germany, Vol 2, pp 148-155, ISBN: 978-989-758-090-1, 2015.
- [16] A. Morales-González, E. García-Reyes, L.E. Sucar, "Segmentation based on level combination of irregular pyramids". Proceedings of the Automatic Image Annotation and Retrieval Workshop (AIAR'2010), Puebla, Mexico, volume 1, issue 1, pp. 1-10, 2010.
- [17] Hernández Saura, Eric J.: "Recuperación de imágenes y videos por contenido utilizando MatchPyr". Tesis de Diploma Curso 2012-2013, Licenciatura en Ciencias de la Computación (UH), Asesor: Ing. Annette Morales González-Quevedo. 2013.