Electronic Letters on Computer Vision and Image Analysis 14(3):11-12, 2015

Toward a perceptual object recognition system

Dounia AWAD

L3i Laboratory, La Rochelle University, La Rochelle, France
Advisor/s: Vincent Courboulay and Arnaud Revel
Date and location of PhD thesis defense: 5 September 2014, University of La Rochelle

Received 24th February 2015; accepted 27th July 2015

Abstract

[1] demonstrated that humans are easily able to recognize an object in less than 0.5 seconds. Unfortunately, object recognition remains one of the most challenging problems in computer vision. Many algorithms based on local approaches have been proposed in recent decades. Local approaches can be divided in 4 phases: region selection, region appearance description, image representation and classification [2]. Although these systems have demonstrated excellent performance, some weaknesses remain. The first limitation is in the region selection phase. Many existing techniques extract a large number of points/regions of interest. For instance, dense grids contain tens of thousands of points per image while interest point detectors often extract thousands of points. Furthermore, some studies have demonstrated that these techniques were not designed to detect the most pertinent regions for object recognition. There is only a weak correlation between the distribution of extracted points and eye fixations [3]. The second limitation mentioned in the literature concerns the region appearance description phase. The techniques used in this phase typically describe image regions using high-dimensional vectors [4]. For example, SIFT, the most popular descriptor for object recognition, produces a 128-dimensional vector per region [5].

The main objective of this thesis is to propose a pipeline for an object recognition algorithm based on human perception which addresses the object recognition system complexity: query run time and memory allocation. In this context, we propose a filter based on a visual attention system [6] to address the problems of extracting a large number of points of interest using existing region selection techniques. We chose to use bottom-up visual attention systems that encode attentional fixations in a topographic map, known as a saliency map. This map serves as basis for generating a mask to select salient points according to human interest, from the points extracted by a region selection technique [7]. Furthermore, we addressed the problem of high dimensionality of descriptors in region appearance phase. We proposed a new hybrid descriptor representing the spatial frequency of some perceptual features, extracted by a visual attention system (color, texture, intensity [8]. This descriptor consist of a concatenation of energy measures computed at the output of a filter bank [9], at each level of the multi-resolution pyramid of perceptual features. This descriptor has the advantage of being lower dimensional than traditional descriptors.

The test of our filtering approach, using Perreira da Silva system [10] as a filter on VOC2005, demonstrated that we can maintain approximately the same performance of an object recognition system by selecting only

Correspondence to: <dounia.awad@gmail.com>

Recommended for acceptance by Jorge Bernal
DOI http://dx.doi.org/10.5565/rev/elcvia.714
ELCVIA ISSN:1577-5097
Published by Computer Vision Center / Universitat Autònoma de Barcelona, Barcelona, Spain

40% of extracted points (using Harris-Laplace [11] and Laplacian [12]), while having an important reduction in complexity (40% reduction in query run time). Furthermore, evaluating our descriptor with an object recognition system using Harris-Laplace and Laplacian interest point detectors on VOC2007 database showed a slight decrease in performance (5% reduction of average precision) compared to the original system based on the SIFT descriptor, but with a 50% reduction in complexity. In addition, we evaluated our descriptor using a visual attention system as the region selection technique on VOC2005. The experiment showed a slight decrease in performance (3% reduction in precision), but a drastically reduced complexity of the system (with 5% reduction in query run-time and 70% in complexity).

In this thesis, we proposed two approaches to manage the problems of complexity in object recognition system. In future, it would be interesting to address the problems of the last two phases in object system: image representation and classification, by introducing perceptually plausible concepts such as deep learning techniques.

References

[1] H. Kirchner and S. J. Thorpe, "Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited," *Vision research*, vol. 46, no. 11, pp. 1762–1776, 2006.

[2] K. Chatfield, V. Lempitsky, A. Vedaldi, and A. Zisserman, "The devil is in the details: an evaluation of recent feature encoding methods," in *Proceedings of the British Machine Vision Conference*. BMVA Press, 2011, pp. 76.1–76.12, http://dx.doi.org/10.5244/C.25.76.

[3] A. Dave, R. Dubey, and B. Ghanem, "Do humans fixate on interest points?" in *Pattern Recognition (ICPR)*, 2012, pp. 2784–2787.

[4] L. Ledwich and S. Williams, "Reduced sift features for image retrieval and indoor localisation," in *In Australian Conference on Robotics and Automation*, 2004.

[5] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vision*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[6] S. Frintrop, "Computational visual attention," in *Computer Analysis of Human Behavior*. Springer, 2011, pp. 69–101.

[7] D. Awad, V. Courboulay, and A. Revel, "Saliency filtering of sift detectors: Application to cbir," in *ACIVS*, 2012, pp. 290–300.

[8] ——, "A new hybrid texture-perceptual descriptor: Application cbir," in *Pattern Recognition (ICPR), 2014 22nd International Conference on*, Aug 2014, pp. 1150–1155.

[9] M. Unser, "Local linear transforms for texture measurements," *Signal Process.*, vol. 11, no. 1, pp. 61–79, Jul. 1986.

[10] M. Perreira Da Silva, "Modèle computationnel d'attention pour la vision adaptative," THESE, Université de La Rochelle, Dec. 2010. [Online]. Available: http://hal.archives-ouvertes.fr/tel-00573844

[11] K. Mikolajczyk and C. Schmid, "Scale & affine invariant interest point detectors," *Int. J. Comput. Vision*, vol. 60, no. 1, pp. 63–86, Oct. 2004. [Online]. Available: http://dx.doi.org/10.1023/B:VISI.0000027790.02288.f2

[12] T. Lindeberg, "Feature detection with automatic scale selection," *Int. J. Comput. Vision*, vol. 30, no. 2, pp. 79–116, Nov. 1998. [Online]. Available: http://dx.doi.org/10.1023/A:1008045108935