A New Approach for Automatic Object Labeling

Joanna Isabelle Olszewska
University of Huddersfield, UK
{j.olszewska@hud.ac.uk}

In order to semantically label visual objects in a large amount of images, we propose a new approach which is fast and accurate. The developed automatic object labeling technique relies on the embedded double matching of local visual descriptors detected in the example and candidate images to semantically tag the candidate image with the label associated to the example image, leading thus to the automatic annotation of the candidate image. Our system has been successfully tested on different standard datasets and is compatible with online applications.

Object Labeling plays an important role for applications such as scene understanding [1], object recognition [2], image retrieval [3], [4], or automatic image annotations [5], [6]. Most of the existing object labeling systems such as LabelMe [7] are manual and thus intrinsically time consuming. Consequently, they are not appropriate for processing large image datasets. For this purpose, some object labeling techniques have been developed recently [8], [9]. However, they are only semi-automatically labeling the objects and they rely on region growing segmentation approaches which require user's prior knowledge.

In this paper, we propose to automatically label an object in an image of a given database by comparing it with a given example image. The main steps of our system are (i) the detection of image features, (ii) their matching and (iii) the label inheritance. The comparison between the query image and the candidate one is performed in the feature space in order to be more computationally effective and to be more robust towards noise and affine transformation, and is followed by the computation of an associated similarity measure. Moreover, our approach does not require any training and thus is online compatible.

In particular, the first step consists in computing for both the example and candidate images, the Scale Invariant Feature Transform (SIFT) descriptors which are robust to rotation, translation, scale changes as well as some viewpoint variations. Secondly, these two sets of extracted SIFT descriptors are matched by means of the Embedded Double Matching Algorithm [5], [6]. Finally, once the decision that a candidate image contains similar content to the query image one has been taken, the label of the candidate image is automatically mapped with the predefined label of the example image.

Hence, by using different objects' examples, all the images from the database are indexed with the relevant semantic labels based on their visual content similarity, while the objects of interest are automatically labeled and localized within these images.

The approach was validated for 3 databases, namely, the CalTech dataset - 101 categories, the OxfordFlower dataset and the Caltech-UCSD Birds 200 database that we have merged to obtain a broad database of 23468 images with an average resolution of 500x500 pixels and containing about 400 classes. In this way, the images of our resulting database have different size and resolution as well as large inter-class similarities and intra-class variations.

The mean average accuracy of our approach is 95% and its computational time is in the range of few milliseconds on an Intel Core 2 Duo Pentium T9300, 2.5 GHz, 2Gb RAM, using MatLab software.

Compared to the work of [8], our approach has a c.10%-better confidence on the label results and does not need the use of arbitrary parameters for the image segmentation. In comparison with [9], our approach does not face issues with non-objects regions and is much more computationally efficient (up to 40 times faster).

In conclusion, the paper presents a fully automatic object labeling method which semantically tags a candidate image based on the results of the matching between the visual descriptors of an example image and the candidate one. By applying recursively this process for different objects of interest, any image could thus be automatically annotated with multiple labels.

REFERENCES

[1] J. Olszewska, "Spatio-temporal visual ontology", In *EPSRC Workshop on Vision and Language (VL 2011)*, Brighton, UK, September 2011**.**

[2] T. Alqaisi, D. Gledhill, and J. Olszewska, **"**Embedded double matching of local descriptors for a fast automatic recognition of real-world objects", In *Proceedings of the IEEE International Conference on Image Processing (ICIP 2012)*, pp. 2385-2388, Orlando, Florida, USA, October 2012.

[3] J. Olszewska and D. Wilson, "Hausdorff-distance enhanced matching of Scale Invariant Feature Transform descriptors in context of image querying", In *Proceedings of the IEEE International Conference on Intelligent Engineering Systems (INES 2012)*, pp. 91-96, Lisbon, Portugal, June 2012.

[4] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1349-1380, December 2000.

[5] F. Wang, "A survey on automatic image annotation and trends of the new age", *Procedia Engineering*, 23:434-438, 2011.

[6] D. Zhang, M. M. Islam, and G. Lu, "A review on automatic image annotation techniques", *Pattern Recognition*, 45(1):346-362, January 2012.

[7] B. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman, "LabelMe: A database and web-based tool for image annotation", *International Journal of Computer Vision*, 77(1-3):157-173, May 2008.

[8] T. Athanasiadis, P. Mylonas, Y. Avrithis, and S. Kollias, "Semantic image segmentation and object labeling", *IEEE Transactions on Circuits and Systems for Video Technology*, 17(3):298-312, March 2007.

[9] W. Wu and J. Yang, "Semi-automatically labeling objects in images", *IEEE Transactions on Image Processing*, 18(6):1340-1349, June 2009.