Electronic Letters on Computer Vision and Image Analysis 14(3):21-23, 2015

Image Retrieval: Modelling Keywords via Low-level Features

Zenonas Theodosiou

Department of Communication and Internet Studies, Cyprus University of Technology, 30 Archbishop Kyprianou Str. 3036, Limassol, Cyprus
Advisor: Nicolas Tsapatsoulis
Date and location of PhD thesis defense: 29 April 2014, Cyprus University of Technology

Received 28th February 2015; accepted 14th May 2015

Abstract

With the advent of cheap digital recording and storage devices and the rapidly increasing popularity of online social networks that make extended use of visual information, like Facebook and Instagram, image retrieval regained great attention among the researchers in the areas of image indexing and retrieval. Image retrieval methods are mainly falling into content-based and text-based frameworks.

Although content-based image retrieval has attracted large amount of research interest, the difficulties in querying by an example propel ultimate users towards text queries. Searching by text queries yields more effective and accurate results that meet the needs of the users while at the same time preserves their familiarity with the way traditional search engines operate. However, text-based image retrieval requires images to be annotated i.e. they are related to text information. Much effort has been invested on automatic image annotation methods [1], since the manual assignment of keywords (which is necessary for text-based image retrieval) is a time consuming and labour intensive procedure [2].

In automatic image annotation, a manually annotated set of data is used to train a system for the identification of joint or conditional probability of an annotation occurring together with a certain distribution of feature vectors corresponding to image content [3]. Different models and machine learning techniques were developed to learn the correlation between image features and textual words based on examples of annotated images. Learned models of this correlation are then applied to predict keywords for unseen images [4].

In the literature of automatic semantic image annotation, proposed approaches tend to classify images using only abstract terms or using holistic image features for both abstract terms and object classes. The extraction and selection of low-level features, either holistic or from particular image areas is of primary importance for automatic image annotation. This is true either for the content-based or for the text-based retrieval paradigm. In the former case the use of appropriate low-level features leads to accurate and effective object class models used in object detection while in the latter case, the better the low- level features are, the easier the learning of keyword models is.

The intent of the image classification is to categorize the content of the input image to one of several keyword classes. A proper image annotation may contain more than one keyword that is relevant to the image content, so a reclassification process is required in this case, as well as whenever a new keyword class is added to the classification scheme. The creation of separate visual models for all keyword classes adds a significant value

Correspondence to: zenonas.theodosiou@cut.ac.cy

Recommended for acceptance by Jorge Bernal
DOI http://dx.doi.org/10.5565/rev/elcvia.725
ELCVIA ISSN:1577-5097
Published by Computer Vision Center / Universitat Autònoma de Barcelona, Barcelona, Spain

in automatic image annotation since several keywords can be assigned to the input image. As the number of keyword classes increases the number of keywords assigned to the images also increases too and there is no need for reclassification. However, the keyword modeling incurred various issues such as the large amount of manual effort required in developing the training data, the differences in interpretation of image contents, and the inconsistency of the keyword assignments among different annotators.

This thesis focuses on image retrieval using keywords under the perspective of machine learning. It covers different aspects of the current research in this area including low-level feature extraction, creation of training sets and development of machine learning methodologies. It also proposes the idea of addressing automatic image annotation by creating visual models, one for each available keyword, and presents several examples of the proposed idea by comparing different features and machine learning algorithms in creating visual models for keywords referring to the athletics domain.

The idea of automatic image annotation through independent keyword visual models is divided into two main parts: the training and automatic image annotation. In the first part, visual models for all available keywords are created, using the one-against-all training paradigm, while in the second part, annotations are produced for a given image based on the output of these models, once they are fed with a feature vector extracted from the input image. An accurate manually annotated dataset containing pairs of images and annotations is prerequisite for a successful automatic image annotation. Since the manual annotations are likely to contain human judgment errors and subjectivity in interpreting the image, the current thesis investigates the factors that influence the creation of manually annotated image datasets [5]. It also proposes the idea of modeling the knowledge of several people by creating visual models using such training data, aiming to significantly improve the ultimate efficiency of image retrieval systems [6].

Moreover, it proposes a new algorithm for the extraction of low level features. The Spatial Histogram of Keypoints (SHiK) [7], keeps the spatial information of localized keypoints, on an effort to overcome the limitations caused by the non-fixed and huge dimensionality of the SIFT feature vector when used in machine learning frameworks. SHiK partitions the image into a fixed number of ordered sub-regions based on the Hilbert space-Filling curve and counts the localized keypoints found inside each sub-region. The resulting spatial histogram is a compact and discriminative low-level feature vector that shows significantly improved performance on classification tasks.

References

[1] D. Zhang, M. M. Islam, G. Lu, “A review on automatic image annotation techniques”, *Pattern Recognition*, 45:346-362, 2012.

[2] A. Hanbury, “A survey of methods for image annotation”, *Journal of Visual Languages & Computing*, 19(5):617-627, 2008.

[3] K. Athanasakos, V. Stathopoulos, J. Jose, “A framework for evaluating automatic image annotation algorithms”, *Lecture Notes in Computer Science*, 5993:217-228, 2010.

[4] R. Zhang, Z. Zhang, M. Li, H. J. Zhang, “A probabilistic semantic model for image annotation and multimodal image retrieval”, Multimedia Systems, pages 12(1):27-33, 2006.

[5] Z. Theodosiou, N. Tsapatsoulis, “Semantic Gap Between People: An Experimental Investigation based on Image Annotation”, *Proc. of the 7th International Workshop on Semantic Media Adaptation and Personalization*, Luxembourg, 73-77, 2012.

[6] Z. Theodosiou, N. Tsapatsoulis, “Modelling Crowdsourcing Originated Keywords within the Athletics Domain”, *Artificial Intelligence Applications and Innovations, IFIP Advances in Information and Communication Technology*, 381:404-413, 2012.

[7] Z. Theodosiou, N. Tsapatsoulis, “Spatial Histogram of Keypoints (SHiK)”, *Proc. of the IEEE International Conference on Image Processing*, Melbourne 2924-2928, 2013.