

Extracting keywords from images: bag-of-visual-words enriched with graph techniques

Gjorgji Madjarov¹ and Sanda Martinčić-Ipšić²

¹ Faculty of Computer Science and Engineering, Ss. Cyril and Methodius University, Skopje, Macedonia

² Department of Informatics, University of Rijeka, Rijeka, Croatia
gjorgji.madjarov@finki.ukim.mk, smarti@inf.uniri.hr

Keywords have become primary means for searching information in documents, images and videos on the WWW. Automatic keyword extraction establishes foundation for various natural language and multimedia processing applications: information retrieval, automatic indexing and classification of a collection of documents, automatic summarization, high-level semantic description, etc. The task of keyword extraction is to automatically identify a set of terms that best describe the document [1].

State-of-the-art keyword extraction approaches are based on statistical methods which require learning from hand-annotated data sets. Lately, the focus of research has shifted toward unsupervised methods, mainly network or graph enabled keyword extraction [2] has attracted researchers attention.

In a network (graph) based keyword extraction the source (document, text, specific data etc.) is transformed into network in a way: words (or units) are nodes of the network and their relations are represented with links. This way, both the statistical properties (frequencies) as well as the structure of source are represented by unique formal representation, hence complex network. Graph formalism beside text can model many different data sources biological, ecological, social relations, transporting infrastructure, etc. We propose extending graph representation model for bag-of-visual words (BoVW) used in image retrieval for extracting the most representative visual parts of image using graph-enabled keyword extraction principles.

State-of-the-art systems for image retrieval use BoVW representation of images. In BoVW models, a vocabulary (or codebook) of visual words is obtained by clustering local image descriptors extracted from images. An image is then represented as a BoVW, which is a sparse vector of occurrence counts of the visual words in the vocabulary.

In our case, we want to represent the images as a complex network of linked visual words: each individual visual word is a node and interactions amongst visual words are links. Co-occurrence networks exploit global location costs of visual words and the adjacency cost of local descriptors in the database [3] [4] as weights of the links between the visual words.

Adjacency cost is defined as the negative logarithm of the normalized histogram of co-occurrences of pairs of visual words in a neighboring region. Only pairs which are in m -neighbor distance are taking into account and their relative positions are using. On the other hand, global location cost is defined as the

negative logarithm of the normalized histogram of the occurrence of a certain visual word at a certain location.

Additionally, network (or graph) enabled keyword extraction exploits different measures for the task of identification and ranking of the most representative features of the source. The keyword extraction enabling network measures are derived [2]: on the node level - degree, strength, selectivity or centrality measures; on the network level - coreness, clustering coefficient, PageRank motivated score Text Rank or HITS-motivated hub and authority score ; on the subnetwork level communities and K-cores .

Using network model and measures used in graph theory, we can represent the images on higher level (e.g. construct a layer with a semantic view of the image). In particular, we expect to identify representative parts of images, patterns or even detect and describe objects in the images. In this situation the keyword extraction and, representation, retrieval, clustering, searching of the images will be improved.

References

1. Mihalcea, R.: Graph-based ranking algorithms for sentence extraction, applied to text summarization. In: Proceedings of the ACL 2004 on Interactive Poster and Demonstration Sessions. (2004)
2. S. Beliga, A.M., Martinčić-Ipšić, S.: An overview of graph-based keyword extraction methods and approaches. *Journal of Information and Organizational Sciences* **39** (2015)
3. Kato, H., Harada, T.: Image reconstruction from bag-of-visual-words. In: Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on. (2014) 955–962
4. Weinzaepfel, P., Jegou, H., Perez, P.: Reconstructing an image from its local descriptors. In: Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition. CVPR '11 (2011) 337–344