Proceedings of I-KNOW '08 and I-MEDIA '08
Graz, Austria, September 3-5, 2008

Using Visual Features to Improve Tag Suggestions in Image Sharing Sites

Mathias Lux
(Klagenfurt University, Klagenfurt, Austria
mlux@itec.uni-klu.ac.at)

Oge Marques
(Florida Atlantic University, Boca Raton, USA
omarques@fau.edu)

Arthur Pitman
(Klagenfurt University, Klagenfurt, Austria
apitman@edu.uni-klu.ac.at)

Abstract: Social media sharing sites such as Flickr or YouTube have become immensely popular. Besides sharing actual content, users also share annotations describing or classifying the contents they publish. Although tagging is easy, annotation still is a laborious task that can be made easier by suggesting meaningful additional tags to the user automatically. In this position paper we propose a system architecture and process for supporting annotation by tag suggestion to increase the quality and quantity of social annotations. The goal is not to tag previously untagged images in a completely automatic way, but instead to extend the amount and completeness of annotations by supporting the user in the process of adding further tags.

Keywords: Multimedia, annotation, tag recommendation, social software
Categories: H.3.3, H.5.1

1 Introduction

The amount of visual information available online is increasing rapidly. Flickr alone claims more than 5,000 uploads per minute to be viewed by an audience of millions. Still the task of finding images of interest remains as challenging as ever. A major difficulty in image retrieval is that most successful search engines rely on the presence of text associated with the images to be able to properly retrieve them. In case of social sharing sites, keywords usually appear as *tags* associated with the images. In a perfect world, all images would have a reasonable number of user-generated tags, which would then enable other users to find and retrieve them. In reality, however, only a small share of uploaded pictures are tagged. While in the early days of *content-based image retrieval* (CBIR) researchers thought visual features might tackle this problem, research was hampered by what became known as the 'semantic gap', which refers to the inability of a machine to fully understand and interpret images based on automatically extracted low-level visual features [Smeulders 2000]. The obstacles imposed by such a gap have limited the success of pure CBIR solutions to narrow domains.

Much of current research in visual information retrieval (VIR) is aimed at reducing the 'semantic gap' and incorporating textual information in order to improve the overall quality of the retrieval results [Datta 2008]. For several years, one of the most prominent obstacles for combining visual data and textual metadata was the long-held assumption that manual image annotation is too expensive, subjective, biased, and ultimately, not feasible. This assumption is now being challenged in many ways, from the availability of Semantic Web-related ontologies, to the popularity of image labelling games, to the willingness of users to annotate, tag, rate, and comment on pictures, enabled by social media sharing sites. The latter aspect, namely the availability of user-generated tags, combined with the successful track record of CBIR within narrow domains, has motivated this work: we want to extend annotation quality and quantity (in the number of tags assigned to an image) by tag recommendation. To do this we aim to identify 'narrow domains', select appropriate visual features automatically and use this knowledge to present users a ranked list of possible additional tags.

2 Related work

Tags assigned by users are often ambiguous, available in several languages or declinations and sometimes not even related to the image content at all [Golder 2006]. Despite these shortcomings, social tagging often leads to surprisingly good annotations extracted from a huge amount of annotated content due to the *wisdom of the crowds* effect [Guy 2006]. A central concept in collaborative tagging is that of *folksonomy*: the result of social annotation, a network of users, resources and tags [Mika 2005]. Also research efforts towards semantically-capable visual information retrieval systems have grown exponentially over the past five years. Some of these efforts are tied to Semantic Web standards, languages and ontologies [Hyvonen 2002], while others employ keywords in a loose way (not associated with any ontology or folksonomy) [Rasiwasia 2006]. Still others rely on tags (e.g., [Datta 2007]) and are therefore more closely related to the work proposed in this paper.

Especially relevant to the ideas presented in this position paper is the work described in [Kern 2008] proposing a tag recommendation system for images solely based on tag co-assignment analysis, which we plan to extend further. The approach of [Aurnhammer 2006] is also related to our work to the extent that they also postulate that a combination of content-based image features and tags enhances image management. However, while we focus on supporting the annotation process to improve and extend the quality of annotations, in [Aurnhammer 2006] the focus is put on reducing the negative effects of mistaken tags (typos and false tags), synonymy and homonymy for retrieval in image databases. Moreover, from a CBIR point of view, our proposed architecture extends the approach in [Aurnhammer 2006] towards current state of the art: we focus on automatic identification of best performing metric/feature combinations for narrow domains instead of defining features (and their combination) a priori.

3 Architecture for Semi-automatic Tag Suggestion

Figure 1 provides a general overview of the main user actions as well as the tasks to be performed by the proposed system. In our scenario we assume that a user tags an image with at least one input tag. Based on the input tag, related tags are selected solely based on tag co-occurrence. Based on a visual similarity analysis using low level visual features involving the input image against images tagged with related tags, the list of tag suggestions is then re-ranked to give higher priority to tags associated with visually similar images.

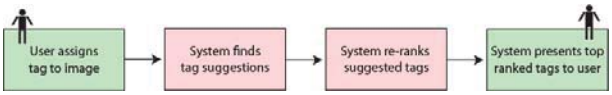

Figure 1: Overall process of the proposed system.

The re-ranking mechanism is the central and novel piece of the proposed architecture. Since this approach is only feasible for narrow domains – and different domains will call for different combination of feature vectors (descriptors) and dissimilarity metrics in order to be successful [Deselaers 2008] – our proposed system employs machine learning to train classifiers to accomplish this goal, namely, to discover the best combination of descriptors and dissimilarity measures for a certain combination of co-occurring tags, which should improve the chances of success of our classifiers.

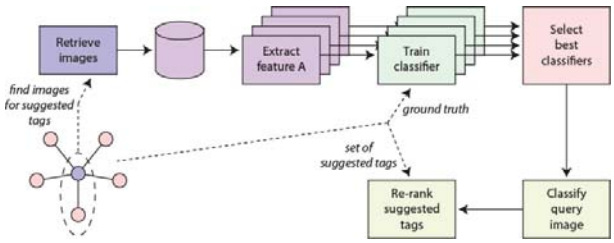

Figure 2: Flow diagram of the re-ranking step in the overall system design.

Figure 2 shows a schematic overview of the proposed re-ranking mechanism and its main steps. We assume that the co-assignment analysis of the input tag(s) results in an affiliation network depicting tag relations. The ego-centred network of the input tag of the user is the input to this process. For each tag related to the input tag, images tagged with *both* the input tag and the related tag (in a pair-wise fashion) are retrieved, resulting in N images per pair, which are stored locally for further visual analysis. For each of those images, visual features (e.g. colour histograms, texture or local features) are extracted by a set of independent feature extraction modules. Treating each tag pair and the corresponding retrieved images as a single topic, we train fuzzy classifiers for feature selection [Wu 2006] for each topic in the following way: for each available visual feature and metric combination a classifier is trained

and evaluated using cross-fold validation. Based on results from the cross-fold validation, the best performing feature/metric combination is used. The result of this procedure is a list of trained fuzzy classifiers. Each of them can be used to find the degree of membership of an image to a certain topic. Next, we return to the input image and use the best performing classifiers to determine its degree of membership. The degree of membership is then used as relevance function for tag suggestions.

4 Research Issues and Contributions

We want to increase quality and quantity of social annotations through the proposed tag suggestion system to support users in the annotation process. The proposed work focuses in particular on the usefulness of the combination of state of the art VIR methods and tag co-assignment analysis. We plan to answer following research questions:

- **Quality of suggestions.** Does our approach enhance the quality of tag suggestions? And if so, by how much?
- **Tag to image content correlation.** Does a significant subset of tags correlate to the image content? And if so, by how much?

As there is no gold standard for social media sharing we plan to evaluate our approach (i) user-centred and (ii) social annotation based, with the social annotations are considered as gold standard.

References

[Aurnhammer 2006] Aurnhammer, M., Hanappe, P., Steels, L.: “Integrating Collaborative Tagging and Emerging Semantics for Image Retrieval”, Proc. Collaborative Web Tagging Workshop (WWW’06), 2006

[Datta 2007] Datta, R., Ge, W., Li, J., Wang, J.: “Toward Bridging the Annotation-Retrieval Gap in Image Search”, IEEE Multimedia, 24-35, 2007

[Datta 2008] Datta, R.; Joshi, D.; Li, J.; Wang, J. Z.: “Image retrieval: Ideas, influences, and trends of the new age”, ACM Computing Surveys 40, 2, p. 1-60, 2008

[Golder 2006] Golder, S. A.; Huberman, B. A.: ”The Structure of Collaborative Tagging Systems”, Journal of Information Science 32, 2, 2006

[Guy 2006] Guy, M., Tonkin, E.: “Folksonomies - Tidying up Tags?” D-Lib Magazine, 12, 1, 2006.

[Hyvonen 2002] Hyvonen, E., Styrman, A., Saarela, S.: “Ontology-based image retrieval”, Proc. XML Finland 2002 Conference, 15-27, 2002.

[Kern 2008] Kern, R.; Granitzer, M.; Pammer, V.: “Extending Folksonomies for Image Tagging”, Proc. WIAMIS 2008, IEEE, Klagenfurt, 2008.

[Mika 2005] Mika, P.: “Ontologies Are Us: A Unified Model of Social Networks and Semantics”, Proc. International Semantic Web Conference, Springer, 522-536, 2005

[Rasiwasia 2006] Rasiwasia, N.; Vasconcelos, N.; Moreno, P.: “Query by Semantic Example', Proc. of the International Conference in Image and Video Retrieval, 51-60, 2006

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

[Wu 2006] Wu, K.; Yap, K.: “Fuzzy SVM for content-based image retrieval: a pseudo-label support vector machine framework”, IEEE Computational Intelligence Magazine, 1, 2, 10-16, 2006