Image Tag Assignment, Refinement and Retrieval

Xirong Li*
Renmin University of China
xirong@ruc.edu.cn

Tiberio Uricchio*
University of Florence
tiberio.uricchio@unifi.it

Lamberto Ballan
University of Florence &
Stanford University
lballan@cs.stanford.edu

Marco Bertini
University of Florence
marco.bertini@unifi.it

Cees G.M. Snoek
University of Amsterdam &
Qualcomm Research
Netherlands
cgmsnoek@uva.nl

Alberto Del Bimbo
University of Florence
alberto.delbimbo@unifi.it

ABSTRACT

This tutorial focuses on challenges and solutions for content-based image annotation and retrieval in the context of online image sharing and tagging. We present a unified review on three closely linked problems, i.e., tag assignment, tag refinement, and tag-based image retrieval. We introduce a taxonomy to structure the growing literature, understand the ingredients of the main works, clarify their connections and difference, and recognize their merits and limitations. Moreover, we present an open-source testbed, with training sets of varying sizes and three test datasets, to evaluate methods of varied learning complexity. A selected set of eleven representative works have been implemented and evaluated. During the tutorial we provide a practice session for hands on experience with the methods, software and datasets. For repeatable experiments all data and code are online at `http://www.micc.unifi.it/tagsurvey`.

Categories and Subject Descriptors

H.3.3 [**INFORMATION STORAGE AND RETRIEVAL**]: Information Search and Retrieval; H.3.1 [**INFORMATION STORAGE AND RETRIEVAL**]: Content Analysis and Indexing—*Indexing Methods*

General Terms

Algorithms, Experimentation, Performance

Keywords

Content-based image retrieval, social tagging, tag relevance, tag assignment, tag refinement, tag retrieval

*Equal contribution and corresponding authors.

MM'15 October 26-30, 2015, Brisbane, Australia.
ACM 978-1-4503-3459-4/15/10.
DOI: http://dx.doi.org/10.1145/2733373.2807419.

1. TUTORIAL DESCRIPTION

Several technological developments have spurred the sharing of images in unprecedented volumes. The first is the ease with which images can be captured in a digital format by cameras, cellphones and other wearable sensory devices. The second is the Internet that allows transfer of digital image content to anyone, anywhere in the world. Finally, and most recently, the sharing of digital imagery has reached new heights by the massive adoption of social network platforms. All of a sudden images came with tags, and tagging, commenting, and rating of any digital image has become a common habit. Despite this downpour of images and tags, the problem of annotating and searching a particular image is still largely unsolved. It has instead dilated the problem with the demand of reliable and objective image tags.

In this tutorial we focus on challenges in content-based image annotation and retrieval in the context of social image sharing and tagging, with a unified review on three closely linked problems in the field, i.e., image tag assignment, tag refinement, and tag-based image retrieval.

Existing works in tag assignment, refinement, and retrieval vary in terms of their targeted tasks and methodology, making it non-trivial to interpret them within a unified framework. We reckon that all works rely on the key functionality of *tag relevance*, i.e., estimating the relevance of a specific tag with respect to the visual content of a given image. Given such a tag relevance function, one can perform tag assignment and refinement by sorting tags in light of the function, and retrieve images by sorting them accordingly. We present a taxonomy, which structures the rich literature along two dimensions, namely *media* and *learning*. The media dimension characterizes what essential information the tag relevance function exploits, while the learning dimension depicts how such information is exploited. With this taxonomy, we discuss connections and difference between the many methods, their advantages as well as limitations.

Comparative evaluation of methods and systems is imperative to appreciate progress. In spite of the growing literature in the field, there is a lack of consensus on the performance of the individual methods. This is largely due to the fact that existing works either use homemade data, which are not publicly accessible, or use selected subsets of benchmark data. Consequently we present an open-source test bed, with training sets of different sizes to evaluate methods of varied learning complexity, and three test sets contributed

Table 1: Methods evaluated in this tutorial. The media column characterizes what essential information a specific method exploits, while the learning SemanticField depicts how such information is exploited.

Method	Media	Learning	Code
SemanticField [11]	tag	Instance based	Python
TagCooccur [9]	tag	Instance based	Python
TagRanking [6]	tag + image	Instance based	Python
KNN [7]	tag + image	Instance based	C + Python
TagVote [4]	tag + image	Instance based	C + Python
TagCooccur+ [4]	tag + image	Instance based	Python
TagProp [2]	tag + image	Model based	C + Matlab + Python
TagFeature [1]	tag + image	Model based	C + Python
RelExample [3]	tag + image	Model based	C + Python
RobustPCA [10]	tag + image	Transduction based	C + Matlab + Python
TensorAnalysis [8]	tag + image + user	Transduction based	–[1]

by various research groups. A selected set of eleven representative works, i.e., SemanticField [11], TagRanking [6], KNN [7], TagVote [4], TagProp [2], TagCooccur [9], TagCooccur+ [4], TagFeature [1], RelExample [3], RobustPCA [10], TensorAnalysis [8], have been implemented and evaluated on the test bed for tag assignment, refinement, and/or retrieval. An overview of the methods is given in Table 1. The interested reader is referred to [5] for a comprehensive comparison between these methods.

During the tutorial, we also provide a practice session for hands on experience with the methods, software, and datasets. For each method a front-end pipeline is implemented, allowing users to conduct tag relevance learning from scratch, obtain tag ranks and image ranks accordingly, and report multiple performance metrics including image-centric Mean image Average Precision (MiAP), tag-centric Mean Average Precision (MAP), and Normalized Discounted Cumulative Gain (NDCG). In addition, Python wrappers for C and Matlab code are given for the ease of cross-platform use.

We conclude the course with our perspective on the many challenges and opportunities ahead for the multimedia community.

A survey paper entitled "Socializing the Semantic Gap: A Comparative Survey on Image Tag Assignment, Refinement and Retrieval" (`http://arxiv.org/abs/1503.08248`) [5] completes the material given to the attendees.

Acknowledgments. This research is supported by Natural Science Foundation of China (No. 61303184), the Fundamental Research Funds for the Central Universities and the Research Funds of Renmin University of China (No. 14XNLQ01), SRFDP (No. 20130004120006), the Dutch national program COMMIT, the STW STORY project, Telecom Italia PhD grant funds, and the AQUIS-CH project granted by the Tuscany Region (Italy). L. Ballan acknowledges also the support of the EC's FP7 under the grant agreement No. 623930 (Marie Curie IOF).

2. REFERENCES

[1] L. Chen, D. Xu, I. Tsang, and J. Luo. Tag-based image retrieval improved by augmented features and group-based refinement. *IEEE Transactions on Multimedia*, 14(4):1057–1067, 2012.

[2] M. Guillaumin, T. Mensink, J. Verbeek, and C. Schmid. TagProp: Discriminative metric learning in nearest neighbor models for image auto-annotation. In *Proc. of ICCV*, 2009.

[3] X. Li and C. Snoek. Classifying tag relevance with relevant positive and negative examples. In *Proc. of ACM Multimedia*, 2013.

[4] X. Li, C. Snoek, and M. Worring. Learning social tag relevance by neighbor voting. *IEEE Transactions on Multimedia*, 11(7):1310–1322, 2009.

[5] X. Li, T. Uricchio, L. Ballan, M. Bertini, C. Snoek, and A. Del Bimbo. Socializing the semantic gap: A comparative survey on image tag assignment, refinement and retrieval. *CoRR*, abs/1503.08248, 2015.

[6] D. Liu, X.-S. Hua, L. Yang, M. Wang, and H.-J. Zhang. Tag ranking. In *Proc. of WWW*, 2009.

[7] A. Makadia, V. Pavlovic, and S. Kumar. Baselines for image annotation. *International Journal of Computer Vision*, 90(1):88–105, 2010.

[8] J. Sang, C. Xu, and J. Liu. User-aware image tag refinement via ternary semantic analysis. *IEEE Transactions on Multimedia*, 14(3):883–895, 2012.

[9] B. Sigurbjörnsson and R. van Zwol. Flickr tag recommendation based on collective knowledge. In *Proc. of WWW*, 2008.

[10] G. Zhu, S. Yan, and Y. Ma. Image tag refinement towards low-rank, content-tag prior and error sparsity. In *Proc. of ACM Multimedia*, 2010.

[11] S. Zhu, Y.-G. Jiang, and C.-W. Ngo. Sampling and ontologically pooling web images for visual concept learning. *IEEE Transactions on Multimedia*, 14(4):1068–1078, 2012.

[1]Results provided by the method developers.