Computer Vision and Image Understanding 124 (2014) 42–60

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

journal homepage: www.elsevier.com/locate/cviu

Image tag completion via dual-view linear sparse reconstructions

Zijia Lin [a,*], Guiguang Ding [b], Mingqing Hu [c], Yunzhen Lin [b], Shuzhi Sam Ge [d]

[a] Department of Computer Science and Technology, Tsinghua University, Beijing 100084, PR China
[b] School of Software, Tsinghua University, Beijing 100084, PR China
[c] Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, PR China
[d] Department of Electrical and Computer Engineering, The National University of Singapore, Singapore 117576, Singapore

ARTICLE INFO

Article history:
Received 15 August 2013
Accepted 27 March 2014

Keywords:
Image tag completion
Linear sparse reconstruction
DLSR
Image tagging
Tag refinement

ABSTRACT

User-provided textual tags of web images are widely utilized for facilitating image management and retrieval. Yet they are usually incomplete and insufficient to describe the whole semantic content of the corresponding images, resulting in performance degradations of various tag-dependent applications. In this paper, we propose a novel method denoted as DLSR for automatic image tag completion via **D**ual-view **L**inear **S**parse **R**econstructions. Given an incomplete initial tagging matrix with each row representing an image and each column representing a tag, DLSR performs tag completion from both views of image and tag, exploiting various available contextual information. Specifically, for a to-be-completed image, DLSR exploits image-image correlations by linearly reconstructing its low-level image features and initial tagging vector with those of others, and then utilizes them to obtain an image-view reconstructed tagging vector. Meanwhile, by linearly reconstructing the tagging column vector of each tag with those of others, DLSR exploits tag-tag correlations to get a tag-view reconstructed tagging vector with the initially labeled tags. Then both image-view and tag-view reconstructed tagging vectors are combined for better predicting missing related tags. Extensive experiments conducted on benchmark datasets and real-world web images well demonstrate the reasonableness and effectiveness of the proposed DLSR. And it can be utilized to enhance a variety of tag-dependent applications such as image auto-annotation.

1. Introduction

Recently with the prevalence of social network and digital photography, numberless images have been posted to various photo sharing communities, *e.g.* Flickr. Apart from the shared visual information, such large-scale and rapidly-increasing social images are usually associated with user-provided textual tags for describing their corresponding semantic content, which are widely utilized for facilitating kinds of tag-based image applications like text-based image retrieval, *etc*. However, as the manual labeling process can be time-consuming and arbitrary, the user-provided tags probably contain imprecise ones and are usually incomplete, as also revealed in [1,2]. Fig. 1 gives an illustration of the user-provided tags with an exemplary image downloaded from Flickr. From the illustration we can see that the user-provided tags may not only contain misspelling or imprecise ones (*e.g.* "mtn"), but also miss other semantically related ones (*e.g.* "sea", "water", "sky" and "grass").

The imprecision and incompleteness of user-provided tags can lead to performance degradations of various tag-dependent applications. Taking tag-based image retrieval as an example, imprecision of tags will lower the retrieval precision while incompleteness will lower the recall. Therefore, in recent years, tag refinement, including tag denoising and completion, has become an attractive subject of many ongoing researches and has been attracting much attention from both academia and industry. However, previous work on tag refinement, as referred to in related work with details, focused more on denoising but less on completion. As our experiments will show, incompleteness of image tags can bring serious negative effects to tag-dependent applications. And thus we propose that tag completion still deserves further attention and researches, and more effective tag completion methods are expected to be developed.

Given an incomplete initial tagging matrix, with each row representing an image and each column representing a tag, tag completion is to fill it up by identifying more correct associations between images and tags. Specifically, each entry of the initial tagging matrix is either 1 or 0, with 1 indicating that the

* Corresponding author.
E-mail addresses: linzijia07@tsinghua.org.cn (Z. Lin), dinggg@tsinghua.edu.cn (G. Ding), humingqing@ict.ac.cn (M. Hu), lin-yz12@mails.tsinghua.edu.cn (Y. Lin), samge@nus.edu.sg (S. Sam Ge).

http://dx.doi.org/10.1016/j.cviu.2014.03.012
1077-3142/

Fig. 1. An exemplary image downloaded from Flickr, with its initially labeled tags (black & red) and several missing related ones (blue). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

corresponding image contains the corresponding tag and 0 otherwise. Then tag completion is essentially to correct false 0 entries into 1 entries.

In this paper, we propose a novel method denoted as DLSR to perform automatic image tag completion via **D**ual-view **L**inear **S**parse **R**econstructions. Specifically, given the initial tagging matrix, the proposed DLSR completes it from both views of image and tag, exploiting various available contextual information. For any to-be-completed image, DLSR exploits image-image correlations by linearly reconstructing its low-level image features and initial tagging vector with those of others, under constraints of sparsity. Then the obtained reconstruction weights are utilized for obtaining an image-view reconstructed tagging vector. Meanwhile, by linearly reconstructing the tagging column vector of each tag with those of others, DLSR exploits tag-tag correlations to get a tag-view reconstructed tagging vector with the initially labeled tags. Then both image-view and tag-view reconstructed tagging vectors are normalized and combined with effective strategies in the field of meta-search to predict the relevance of unlabeled tags to the to-be-completed image. And those with higher relevance are then selected and added.

Instead of performing global refinement for the initial tagging matrix, DLSR performs tag completion via reconstructing each image (*i.e.* row) and each tag (*i.e.* column) separately. And thus it can be utilized to perform tag completion for an unseen image (*i.e.* inductive method) or an existing dataset (*i.e.* transductive method). Specifically, for an unseen image, DLSR only exploits the completely or partially labeled images in the training set to perform tag completion, and thus is used as an inductive method. And for an existing dataset, DLSR performs tag completion for each to-be-completed image in it with all other images, including other to-be-completed images and the training images, since all the to-be-completed images are already observed and also partially labeled, which can probably provide extra helpful information. In this case DLSR is used as a transductive method. DLSR is evaluated with extensive experiments conducted on benchmark datasets and real-world web images. Experimental results well demonstrate its reasonableness and effectiveness. And it can be utilized for enhancing a variety of tag-dependent applications like image auto-annotation, *etc*.

The main contributions of our work can be summarized as follows.

- We propose a novel effective tag completion method via dual-view linear sparse reconstructions, considering and exploiting various available contextual information.
- We propose to perform tag completion via reconstructing each image and each tag separately, instead of performing global refinement for the initial tagging matrix, which enables DLSR to be used as either an inductive method or a transductive one.

This paper is an extension and improvement of our previous work presented in [3]. And we enhance it to be more effective and practical. Specifically, in image-view reconstruction, here we propose to utilize the same reconstruction weights for concurrently reconstructing the low-level features and the initial tagging vector of a to-be-completed image with those of others, in order to simplify model tuning with less parameters while keeping similar performance. Moreover, to prevent the reconstruction weights from being dominating in only images containing an identical initial tagging vector to that of the to-be-completed image, which will provide no information about the missing tags and thus make the image-view reconstruction not work for tag completion, we introduce a "diversity regularizer" in the objective function, as will be elaborated later. Furthermore, to better combine the image-view and the tag-view reconstructed tagging vectors, we propose to treat both as the tag retrieval results from two distinct "search engines", and resort to effective normalization and combination strategies in the field of meta-search for performance improvement. Experimental results demonstrate that the introduced model enhancements here can generally help to gain performance improvement for the proposed method.

The remainder of this paper is organized as follows. Section 2 gives an overview of related work. Section 3 elaborates on the proposed DLSR, presenting formula details. Then detailed description of experiments, including experimental settings, results and analyses, is given in Section 4. And in Section 5 we investigate various applications that DLSR can be used for. Finally we conclude the paper in Section 6.

2. Related work

As tag completion is to add tags with higher relevance to a given image, it would be natural to compare it with image auto-annotation and tag recommendation. Image auto-annotation [4–10] is to automatically and objectively associate unlabeled images with semantically related tags. Feng et al. [4] proposed a generative learning approach for auto-annotation based on multiple Bernoulli relevance model. Liu et al. [6] built multiple graphical models with various correlations between images and tags, and then performed auto-annotation with manifold learning processes. Makadia et al. [5] proposed a widely-used auto-annotation baseline termed JEC, which is a straightforward greedy algorithm propagating labels from nearest visual neighbors to a to-be-annotated image. And Guillaumin et al. [7] proposed to adopt discriminative metric learning methods in nearest neighbor models, putting forward a state-of-the-art auto-annotation model termed TagProp. In [8,9], Ma et al. further proposed effective methods to exploit the original feature space, via sparsity-based feature selection or uncovering shared subspace, to improve the performance of image auto-annotation. Tag recommendation [2,11–15] is a trade-off between auto-annotation and manual tagging, which is to recommend semantically related tags to a user while he is annotating an image online. Sigurbjörnsson and Zwol [2] proposed a generic tag recommendation method exploiting the collective knowledge residing in images. Wu et al. [12] proposed a learning-based multi-modality recommendation algorithm by considering both tag and visual correlations. And Lee et al. [13] formulated tag recommendation as a maximum a posteriori (MAP) problem using a visual folksonomy.

When comparing tag completion and image auto-annotation, the former can be seen as a special case of the latter. However,