Pattern Recognition 52 (2016) 287–305

Contents lists available at ScienceDirect

Pattern Recognition

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

Two-tier image annotation model based on a multi-label classifier and fuzzy-knowledge representation scheme

Marina Ivasic-Kos [a,*], Miran Pobar [a], Slobodan Ribaric [b]

[a] Department of Informatics, University of Rijeka, Rijeka, Croatia
[b] Faculty of Electrical Engineering and Computing, University of Zagreb, Zagreb, Croatia

ARTICLE INFO

Article history:
Received 31 October 2014
Received in revised form
31 August 2015
Accepted 22 October 2015
Available online 6 November 2015

Keywords:
Image annotation
Knowledge representation
Inference algorithms
Fuzzy Petri Net
Multi-label image classification

ABSTRACT

Automatic image annotation involves automatically assigning useful keywords to an unlabelled image. The major goal is to bridge the so-called semantic gap between the available image features and the keywords that people might use to annotate images. Although different people will most likely use different words to annotate the same image, most people can use object or scene labels when searching for images.

We propose a two-tier annotation model where the first tier corresponds to object-level and the second tier to scene-level annotation. In the first tier, images are annotated with labels of objects present in them, using multi-label classification methods on low-level features extracted from images. Scene-level annotation is performed in the second tier, using the originally developed inference-based algorithms for annotation refinement and for scene recognition. These algorithms use a fuzzy knowledge representation scheme based on Fuzzy Petri Net, KRFPNs, that is defined to enable reasoning with concepts useful for image annotation. To define the elements of the KRFPNs scheme, novel data-driven algorithms for acquisition of fuzzy knowledge are proposed.

The proposed image annotation model is evaluated separately on the first and on the second tier using a dataset of outdoor images. The results outperform the published results obtained on the same image collection, both on the object-level and on scene-level annotation. Different subsets of features composed of dominant colours, image moments, and GIST descriptors, as well as different classification methods (RAKEL, ML-kNN and Naïve Bayes), were tested in the first tier. The results of scene level annotation in the second tier are also compared with a common classification method (Naïve Bayes) and have shown superior performance. The proposed model enables the expanding of image annotation with new concepts regardless of their level of abstraction.

1. Introduction

Image retrieval, search and organisation have become a problem on account of the huge number of images produced daily. In order to simplify these tasks, different approaches for image retrieval have been proposed that can be roughly divided into those that compare visual content (content-based image retrieval) and those that use text descriptions of images (text-based image retrieval) [27,7].

Image retrieval based on text has appeared to be easier, more natural and more suitable for people in most everyday cases. This is because it is much simpler to write a keyword-based query than to provide image examples, and it is likely that the user does not have an example image of the query. Besides, images corresponding to the same keywords can be very diverse. For example, if a person has an image of a town, but wants a different view, content-based retrieval would not be the best choice because most of the results would be too similar to the image he already has. On the other hand, very diverse images can be retrieved with a keyword query, e.g. for the keyword *Dubrovnik,* different views of the town can be retrieved (Fig. 1.).

To be able to retrieve images using text, they must be labelled or described in the surrounding text. However, most images are not. Since manually providing image annotation is a tedious and expensive task, especially when dealing with a large number of images, automatic image annotation has appeared as a solution.

* Correspondence to: R. Matejčić 2, 51000 Rijeka, Croatia. Tel.: +385 51584710.
E-mail address: marinai@uniri.hr (M. Ivasic-Kos).

http://dx.doi.org/10.1016/j.patcog.2015.10.017
0031-3203/

Fig. 1. Part of the image search results for the keyword "Dubrovnik".

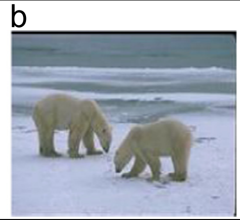
	a 	b 
Object labels	*tracks, train, cloud, sky, trees,*	*snow, polar bear*
Scene label	*SceneTrain, Transportation*	*ScenePolarbear, WildLife, Arctic*

Fig. 2. Examples of images and their annotation at object and scene levels.

Automatic annotation methods deal with visual features that can be extracted from the raw image data, such as colour, texture and structure, and can automatically assign metadata in the form of keywords from a controlled vocabulary to an unlabelled image. The major goal is to bridge the so-called semantic gap [12] between the available features and the keywords that could be useful to people. In the paper [15] an image representation model is given to reflect the semantic levels of words used in image annotation.

This problem is challenging because different people will most likely annotate the same image with different words that reflect their knowledge of the context of the image, their experience and cultural background. However, most people when searching for images use two types of labels: object and scene labels. Object labels correspond to objects that can be recognised in an image, like *sky, trees, tracks* and *train* for the image in Fig. 2a. Scene labels represent the context of the whole image, like *SceneTrain* or more general *Transportation* for Fig. 2a. Scene labels can be either directly obtained as a result of the global classification of image features [23] or inferred from object labels, as proposed in our approach.

Since object and scene labels are most commonly used, in this paper we focus on automatic image annotation at scene and object levels.

We propose a two-tier annotation model for automatic image annotation, where the first tier corresponds to object annotation and the second tier to scene-level annotation. An overview of the proposed model is given after the sections with the related work, in Section 3. The first assumption is that there can be many objects in any image, but an image can be classified into one scene. The second assumption is that there are typical objects of which scenes are composed. Since many object labels can be assigned to an image, the object-level annotation was treated as a multi-label problem and appropriate multi-label classification methods RAKEL and ML-kNN were used. The details of object level annotation are given in Section 4. The scene-level annotation relies on originally developed inference based algorithm defined for fuzzy knowledge representation scheme based on Fuzzy Petri Net, KRFPNs. The KRFPNs scheme and data-driven algorithms for knowledge acquisition about domain images are defined in Section 5. The inference approach for annotation refinement at the object level is given in Section 6. The conclusions about scenes are drawn using inference-based algorithms on facts about image domain and object labels obtained in the first tier, as detailed in Section 7. The major contributions of this paper can be summarised as follows:

1. The adaptive two-tier annotation model in which each tier can be independently used and modified.
2. The first tier uses multi-label classification to suit the image annotation at object level.
3. The second tier uses inference-based algorithms to handle the recognition of scenes and higher-level concepts that do not have specific low-level features.
4. The definition of the knowledge representation scheme based on Fuzzy Petri Nets, to represent knowledge about domain images that is often incomplete, uncertain and ambiguous.
5. Novel data-driven algorithms for acquisition of fuzzy knowledge about relations between objects and about scenes.
6. Novel inference based algorithm for annotation refinement to reduce the error propagation through the hierarchical structure of concepts during the inference of scenes.
7. Novel inference based algorithm for automatic scene recognition.
8. A comparison of inference based scene classification with an ordinary classification approach.

The performance of the proposed two-tier automatic annotation system was evaluated on outdoor images with regard to different feature subsets (dominant colours, moments, GIST descriptors [23] and compared to the published results obtained on the same image database as detailed in Section 8. The inference based scene classification is also compared with an ordinary classification approach and has shown better performance. The paper ends with a conclusion and directions for future work, Section 8.