Measurement of Tag Confidence in User Generated Contents Retrieval

Sihyoung Lee, Hyun-Seok Min, Young Bok Lee, Yong Man Ro*
Image and Video Systems Lab., Information and Communications University (ICU),
Munjiro 119, Yuseong, Daejeon, South Korea

ABSTRACT

As online image sharing services are becoming popular, the importance of correctly annotated tags is being emphasized for precise search and retrieval. Tags created by user along with user-generated contents (UGC) are often ambiguous due to the fact that some tags are highly subjective and visually unrelated to the image. They cause unwanted results to users when image search engines rely on tags. In this paper, we propose a method of measuring tag confidence so that one can differentiate confidence tags from noisy tags. The proposed tag confidence is measured from visual semantics of the image. To verify the usefulness of the proposed method, experiments were performed with UGC database from social network sites. Experimental results showed that the image retrieval performance with confidence tags was increased.

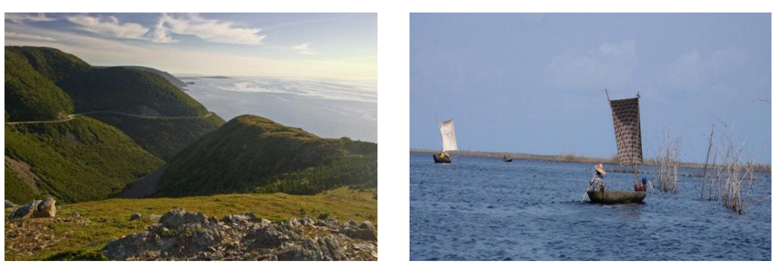
Keywords: Tag confidence, semantic vector, visual perception, tag refinement

1. INTRODUCTION

The amount of user-generated content is increasing rapidly during the recent years due to the popularity of easy-to-use multimedia devices, as well as the availability of cheaper storage and bandwidth [1, 2, 3]. In particular, digital cameras are powerful and popular tool to capture one's daily life. Images are often shared through online image sharing services such as Flickr [4], Webshots [5]. For example, as of July 2007, Flickr is known to serve 12,000 images per second, and over two million images are uploaded every day by 8.5 million registered members [6].

The vast amount of online available images requires a certain amount of descriptive tags to images, because generic online image sharing services are using tags to deal with searching, managing and organizing images. To do that, users have possibility to register appropriate tags to images when uploading these images. Image searching engines can then retrieve images by relying on tag-based search techniques.

However, those tags annotated by users are not always representing the actual image. They are likely to be highly subjective [7]. Sometimes, some tags are not related to the visual perception of the image. For example, the pictures and tags shown in Fig. 1 are retrieved from Flickr. As seen in Fig. 1(a), a user has tagged 'occasions' which is a highly subjective tag. In Fig 1 (b), it is difficult to find any relations between the image and tags such as 'hotel' and 'house'. Since these highly subjective and visually unrelated tags are registered in online image sharing services, current image search engines relying on tags would result in unwanted retrieval results to users.



Tags: hill, holiday, nova, occasions, terrain, ...
(a) highly subjective tag (occasions)

Tags: sea, coast, water, sky, river, hotel, house, …
(b) visually not related tag (hotel, house)

Fig. 1. Example pictures from the Website Flickr.com

* Correspondence: Prof. Yong Man Ro; Email: yro@icu.ac.kr;
Image and Video Systems Laboratory, Information and Communications University (ICU), Munji 103-6, Yuseong, Daejeon, 305-732, South Korea, Phone: +82-42-866-6129, Fax: +82-42-866-6245

Visual Communications and Image Processing 2009, edited by Majid Rabbani, Robert L. Stevenson, Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 7257, 72571P · © 2009 SPIE-IS&T · CCC code: 0277-786X/08/$18 · doi: 10.1117/12.806124

Tag recommendation could be one technique to reduce the above mentioned problem. Supervised tag recommendation not only makes users to annotate tags in a more time efficient way but also assists in maintaining consistency between images and tags. It supports suggestions to users based on analysis of visual similarity between images. Using tag recommendation, candidate keywords are typically produced by an automatic image annotation process, which analyzes the correlation between an image and several keywords. Classification-based image annotation attempts to infer correlations between keywords representing semantics and images by relying on classifiers. Each classifier is used for the detection of a single concept in an image, such as 'architecture', 'sky', 'beach', 'sunset', and so on. Several classification methods have been developed and discussed in the literature, such as techniques based on Support Vector Machines (SVMs) [9] and techniques relying on Bayes Point Machines [10]. Besides classification-based image annotation, it is also possible to make use of probability-based annotation. This technique infers the probability that a particular keyword can be associated with an image. To do so, the similarity is computed between the image to be annotated and a set of training images that are already annotated with tags [11-14]. However, tag recommendation is applicable to images which are newly uploaded. It could not be applied to the images uploaded before, because recommendation operates within the uploading process.

To overcome the aforementioned limitation, this paper proposes a method to measure tag confidence for images in database so that one can differentiate confidence tags from noisy tags. We calculate the confidence value for tag in UGC image database by analyzing the associated visual information and ontology information. Semantic vector is adapted to measure visual similarity, and WordNet is used to obtain similarity between tags. Our goal is that image search engine can provide reliable retrieval performance by only using highly confident tags.

The rest of the paper is organized as follows: Section 2 provides the details of the proposed method. Experiments and results are given in Section 3. Finally, Section 4 concludes the paper and discusses future work.

2. PROPOSED METHOD

2.1 Overall procedure of proposed method

Tag confidence in this paper describes how much the tag is related to associate image. Tag confidence value is defined varying from 0 to 1. If the tag is not related to the image, tag confidence value is close to 0. Likewise, as tag confidence value is close to 1, the tag is highly related to the image.

Fig. 2 illustrates the overall procedure of the proposed tag confidence measurement. The tag confidence is calculated based on the visual information of image. Let us assume the image database in Fig. 2 consist of images from UGC and associated tags. We suppose to measure tag confidence for all tags with associated images in the database. To do that, a target image is selected from the image database and the tag confidence for the target image is measured. All images in the database are selected as target images so that confidence values of all tags in the database are measured.

Semantic vectors for a target image and images in the database are generated. With the generated semantic vectors, visual similarity between the target image and database images is calculated. Based on the visual similarity, likelihood images from the database, which are visually similar to the target image, are selected. The tag confidence of the target image is measured by tags on likelihood images which are visually similar to the target image. This paper employs WordNet [8] ontology knowledge to obtain word similarity among tags. Each module in Fig. 2 is explained as follows:

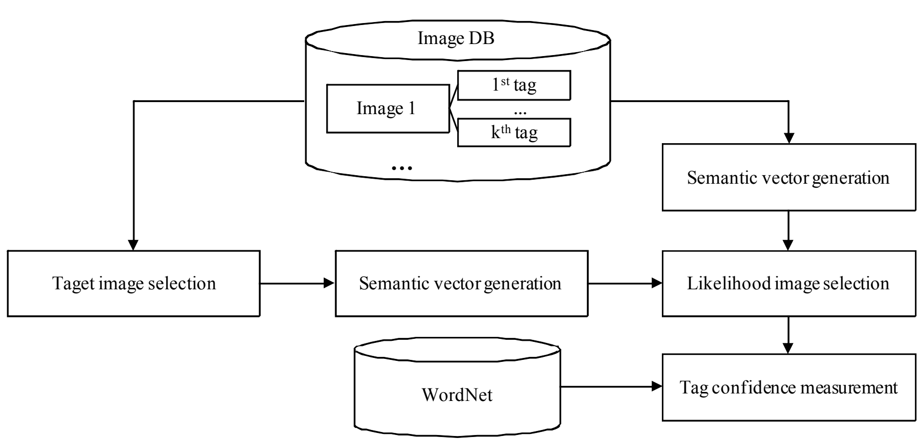


Fig. 2. Overall procedure of proposed tag confidence measurement

2.2 Semantic vector generation

A semantic vector aggregates the confidence values that are the result of applying a series of independent classifiers to an image [14]. Each confidence value represents the degree of certainty in which a particular concept has been detected. Each element of a semantic vector, i.e. the confidence value for a particular concept, is generated from a collection of training images that all contain the same concept.

Since the training set contains diverse kind of images in the same concept, a semantic vector is generally better to represent a concept than a low-level feature vector. For example, the concept ‘bear’ could be learned by using a training database containing images of both polar bears and brown bears. Although the visual difference between polar bear images and brown bear images is significant (white versus brown), the semantic difference is rather small (all images represent the ‘bear’ concept). As such, using low-level features, it is possible to under-estimate the image similarity although the images in question represent the same concept. However, using semantic vectors, we know a bear can sometimes be white or brown (from the training database). Therefore, it is possible to compensate for the under-estimated similarity.

Fig. 3 shows the process for generating semantic vector for an image. If an image I is input, the entire image is divided spatially into sub-regions in the same way as in [9]. Low-level visual features are extracted from the sub-regions. By using the low-level features and semantic learning model described in [9], the semantic vector $\mathbf{V}_I$ is finally generated.

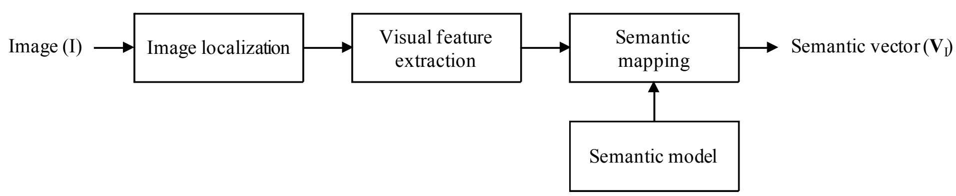


Fig. 3. Semantic vector generation process

2.3 Likelihood image selection

Likelihood images in the database, which are visually similar to the target image, are obtained with the semantic vectors mentioned above. The distance between two vectors of the target image and an image in the database is calculated as

$$d_2(\mathbf{V}_t, \mathbf{V}_d) = \frac{1}{p} \cdot \left(\sum_{k=1}^{p} \left| v_t^k - v_d^k \right|^2 \right)^{\frac{1}{2}}, \tag{1}$$

where $\mathbf{V}_t$ and $\mathbf{V}_d$ are semantic vectors for target image and database images respectably, and v_t^k and v_d^k represent k^{th} elements of $\mathbf{V}_t$ and $\mathbf{V}_d$. p is the dimension of semantic vector.

Since each element of semantic vector represents the probability of the semantic concepts [9], the maximum distance in Eq. (1) is 1. The selected likelihood images are $I_1, \ldots, I_{tho}$ which are represented in order of similarity distance. Image in the database having small similarity distance means it is close to the target image.

2.4 Tag confidence measurement

Each tag confidence in the target image is supposed to be measured. Let n^{th} tag of target image be T_t^n . It is calculated by analyzing the word similarity of tags in likelihood images. In this paper, we adopt WordNet to calculate the similarity among tags. Fig. 4 shows the schematic structure of the tag confidence measurement.

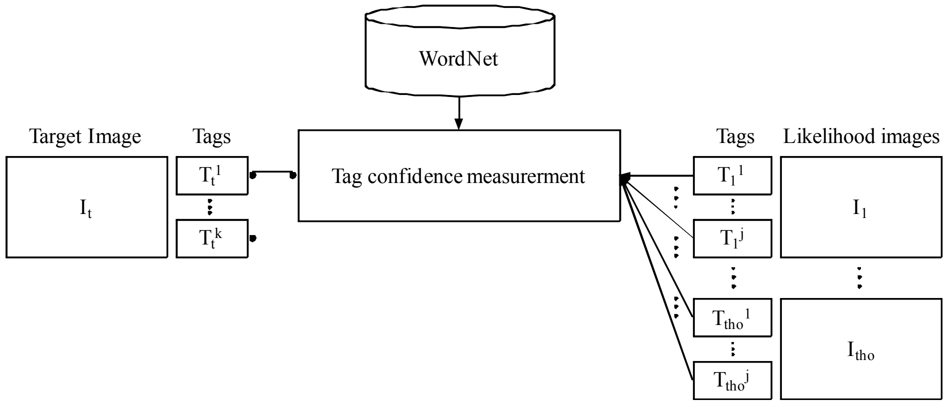


Fig. 4. Proposed tag confidence measure

We use two methods for measuring the tag confidence in this paper. Method 1 considers only the word similarity of likelihood images to measure tag confidence described as,

$$\text{Tag confidence for } \mathrm{T}_t^n = \frac{1}{\text{tho}} \cdot \sum_{m=1}^{\text{tho}} \left[\max\{S(\mathrm{T}_t^n, \mathrm{T}_m^i) \mid i = 1,2,..., j\right]. \tag{2}$$

where *tho* is the number of likelihood images, $\max\{\cdot\}$ returns the maximum element in the set, and $S(\cdot)$ is a word similarity between tags. The similarity of two words is obtained as the cosine of the angle between their vector representations (gloss vectors) as follows [10],

$$S\left(\mathrm{T}^k, \mathrm{T}^i\right) = \cos\left(angle\left(\mathbf{g}^k, \mathbf{g}^i\right)\right) = \frac{\mathbf{g}^k \cdot \mathbf{g}^i}{\left|\mathbf{g}^k\right|\left|\mathbf{g}^i\right|}, \tag{3}$$

where $\mathbf{g}^k$ and $\mathbf{g}^i$ are the gloss vector representation of T^k and T^i, respectively. And $angle(\cdot)$ returns the angle between vectors.

As seen in Eq. (2), the maximum word similarity is found for each likelihood images and summed for all likelihood images. The summed value is normalized to obtain the tag confidence.

Method 2 is similar to Method 1 except that it considers visual similarity as well as the word similarity. It is described as,

$$\text{Tag confidence for } \mathrm{T}_t^n = \frac{1}{\text{tho}} \cdot \sum_{m=1}^{\text{tho}} \left[v_{t,m} \cdot \max\{S(\mathrm{T}_t^n, \mathrm{T}_m^i) \mid i = 1,2,..., j\}\right], \tag{4}$$

where $v_{t,m}$ is the visual similarity between target image and m^{th} likelihood image which is measured by ,

$$v_{t,m} = 1 - d_2\left(\mathbf{V}_t, \mathbf{V}_m\right). \tag{5}$$

In Eq. (4), the word similarity is weighted by the visual similarity. It sets high confidence value to the tag of higher ranked likelihood image. The higher confidence value signifies that the tag is highly related to the image.

3. EXPERIMENTS

We have collected 3108 images and associated tags of the images from Flickr.com through OpenAPI. There are 29,081 tags and 5504 distinct tags. We have used 428 images in the experiment, which have at least two visually related tags. For ground truth, each tag is manually analyzed to see that it is related to the visual perception of the image.

Fig. 5 shows the refinement results according to the tag confidence proposed in this paper. The underlined tags have high confidence and are selected as visually related tags. We can easily recognize that the selected tags, 'hill' and 'terrain' in Fig 5 (a) and 'sea', 'coast', 'water', 'sky' and 'river' in Fig 5 (b), are visually related with the pictures.

(a) <u>hill</u>, holiday, nova, occasions, <u>terrain</u>, …

(b) <u>sea</u>, <u>coast</u>, <u>water</u>, <u>sky</u>, <u>river</u>, hotel, house, …

Fig. 5. The result about tag refinement with proposed tag confidence

To verify the effectiveness of proposed method, we performed an image retrieval using tag and measured recall and precision rate for two methods: Two methods are image retrieval with tag confidence though method 1, and with tag confidence through method 2. Note that tag confidence represents how much the tag is visually related to the image. To measure the recall and precision with method 1 and 2, we selected confident tags by thresholding tag confidence values.

Table 1 shows the results of the experiments. As seen in Table 1, Retrieval performances applying tag confidence method are better that the method without tag confidence. Recall and precision is measured as followed,

$$\text{Recall} = \frac{a_{correct}}{a_{determined}}, \tag{6}$$

$$\text{Precision} = \frac{a_{correct}}{a_{manual}}, \tag{7}$$

where a_{correct} is the number of visually related tags which is retrieved correctly. $a_{\text{determined}}$ is the number of tags selected by thresholding tag confidences. a_{manual} is the number of visually related tags determined by people.

Table 1. Experimental results

	Recall	Precision
Method 1	0.60	0.58
Method 2	0.62	0.61

4. CONCLUSIONS AND FUTURE WORK

In order to enable efficient search in vast collections of images, tags are important feature because current image searching engine retrieve images through the use of tags. However, there exists inappropriate tags (highly subjected tags and visually unrelated tags), and those tags decrease the retrieval performance. Therefore, tag confidence which represent how much the tag is related with the image can provide enhancement to image retrieval. In this paper, the measurement of tag confidence in UGC database was proposed. The proposed tag confidence method was based on visual semantics of image. The purpose of this work is to reduce noisy tags such as highly subjective tags and visually unrelated tags from pool of tags generated by users. Experimental results showed that retrieval performance was increased by using confident tags.

At present, recall rates are lower than expected since it is affected by several tags which represent location information. It is difficult to recognize specific locations such as 'England' or 'Tokyo' with visual information only. A potential improvement to increase the recall rate is to obtain location information by using GPS information in EXIF. Besides, extensive experiments with more scalable database will be performed.

REFERENCES

[1] OECD, "OECD study on the Participative Web: User Generated Content," (2007).

[2] M. Ames, and M. Naaman, "Why We Tag: Motivations for Annotation in Mobile and Online Media", ACM CHI 2007, 971-980 (2007).

[3] R. Ramakrishnan and A. Tomkins, "Toward a People Web," IEEE Computer, 40(8), 63-72 (2007).

[4] Flickr, http://www.flickr.com/.

[5] Webshots, http://www.webshots.com/.

[6] Flickr blog, "We're going down," http://blog.flickr/en/2007/05/29/were-going-down/.

[7] J. Li, James Z. Wang, "Real-Time Computerized Annotation of Pictures," IEEE Trans. PAMI, 30(6), 2008.

[8] T. Pedersen, S. Patwardhan, and J. Michelizzi, "Wordnet::Similarity - Measuring the Relatedness of Concepts," Proc. 9th National Conf. on Artificial Intelligence, 2004.

[9] J. R. Smith, M. Naphade, A. Natsev, "Multimedia semantic indexing using model vectors," Proc. ICME, 2, 445-448 (2003).

[10] C. Cusano, G. Ciocca, R. Schettini, "Image annotation using SVM," Proc. Internet Imaging, (2004).

[11] E. Chang, G. Kingshy, G. Sychay, G. Wu, "CBSA: content-based soft annotation for multimodal image retrieval using Bayes point machines," IEEE Trans. CSVT, 13(1), 26-38 (2003).

[12] P. Duygulu, K. Barnard, "Object recognition as machine translation: learning a lexicon for a fixed image vocabulary," Proc. 7th European Conference on Computer Vision, 4, 97-112 (2002).

[13] J. Jeon, R. Manmatha, "Automatic image annotation of news images with large vocabularies and low quality training data," Proc. ACM Multimedia, (2004).

[14] J. R. Smith, M. Naphade, A. Natsev, "Multimedia semantic indexing using model vectors," Proc. ICME, 2, 445-448 (2003).

[15] S. Yang, S. Kim, Y.M. Ro, "Semantic Home Photo Categorization", IEEE Trans. Circuits and Systems for Video Technology, 17(3), 324-335 (2007).

[16] S. Patwardhan, T. Pedersen. "Using WordNet-based Context Vectors to Estimate the Semantic Relatedness of Concepts," Proc. EACL 2006 Workshop, 1-8 (2006).