

Tags Re-ranking Using Multi-level Features in Automatic Image Annotation

Forogh Ahmadi¹, Vafa Maihami²

1- Department of Computer Engineering, Islamic Azad University, Sanandaj Branch, Sanandaj, Iran.

2- Department of Computer Engineering, Islamic Azad University, Sanandaj Branch, Sanandaj, Iran. (maihami@iausdj.ac.ir)

Received (2019-07-06)

Accepted (2019-08-22)

Abstract: Automatic image annotation is a process in which computer systems automatically assign the textual tags related with visual content to a query image. In most cases, inappropriate tags generated by the users as well as the images without any tags among the challenges available in this field have a negative effect on the query's result. In this paper, a new method is presented for automatic image annotation with the aim at improving the obtained tags, as well as reducing the effect of unrelated tags. In the proposed method, first, the initial tags are determined by extracting the low-level features of the image and using neighbor voting method. Afterwards, each initial tag is assigned by a degree based on the neighbor image features of the query image. Finally, they will be ranked based on summing the degrees of each tag and the best tags will be selected by removing the unrelated tags. The experiments conducted on the proposed method using the NUS-WIDE dataset and the commonly used evaluation metrics demonstrate the effectiveness of the proposed system compared to the previous works.

Keywords: Automatic image annotation, Low level feature, Tag ranking, Neighbor voting.

How to cite this article:

Forogh Ahmadi, Vafa Maihami. Tags Re-ranking Using Multi-level Features in Automatic Image Annotation. J. ADV COMP ENG TECHNOL, 5(4) Autumn 2019 : 255-265

1. INTRODUCTION

With the ever-increasing development of image production technologies, in particular, mobile devices, the amount of image data has increased significantly. Typically, the images are stored without classification and additional information about their content. Therefore, organizing information and analyzing contents based on users need is becoming a challenge more than ever [1–4]. This process requires the existence of effective search technologies for large volumes of images in image retrieval applications [5]. In the content-based images retrieval using visual features, the goal is

to find images that are similar to the input image, which has challenges such as the inaccessibility of an appropriate image for a query, as well as a semantic gap between low-level features and high-level concepts [6]. Automatic image annotation is one of the most effective and practical ways to solve this problem. Automatic image annotation is the process of adding metadata as keywords to the image by a computer system that directly describes its content. When an image is retrieved, the user's textual query is compared to the annotation described for each image and as the result images similar to the query are selected. This approach, besides providing text-based search capability, organizes a large volume of information logically without users intervention [7, 8].



This work is licensed under the Creative Commons Attribution 4.0 International Licence.

To view a copy of this licence, visit <https://creativecommons.org/licenses/by/4.0/>

Although in recent years, many studies have been done in the field of automatic image annotation, there still remain challenges such as assigning some weak tags to images [9]. In that sense, the relation between the descriptive words and the image content is low. In the post tagging process, it is possible to re-rank them, select the most relevant ones and remove unrelated ones. The studies carried out in this area used the low-level features of the image in only one step. The proposed method in this paper is based on two steps, obtaining initial tags and re-ranking tags. Neighbor voting method is used in the process of obtaining initial tags [10]. In this method, first, low-level features are extracted from the images, then based on these features, the similarity of the query image and the dataset images are obtained. The initial tags are determined based on the highest number of repetition of the image tags with the highest similarity of content to the query image. At the re-ranking step of tags, to improve the accuracy of the results of initial tags and to avoid uncertainty in their priorities, especially the same repetitions of the voting stage, other visual contents besides the features used in the earlier stage are extracted from the neighbor image. After combining the features based on the amount obtained from each neighbor image an important score is assigned to each one of them. Initial tags will be weighted according to the importance score of the neighbors. Finally, the relevance level of each tag to the image is determined based on their total weight. The final tags will be selected after re-ranking.

The main contributions of this paper to maximize exploitation of neighbor image information using visual features are as follows (1: Extract low-level features from neighboring images and combine them together to achieve a stronger feature vector; (2 assigning importance score to each neighbor based on the combination of the new feature to determine their significance; 3) Weighting each initial tag according to the importance of neighbors and finally re-rank them. The structured as follows: the related works are introduced in section 2. In section 3, the proposed method is described in details. Section 4 is devoted to the implementation and results of the proposed method, and in section 5 the conclusion and future works are presented.

2. RELATED WORKS

In recent years, the discussion of image annotation has become a popular topic in the field of image processing, computer vision, and multimedia. Given the advantages of these systems, they can become one of the best ways to retrieve images on the internet [3, 10, 11]. Therefore, this subject has attracted the attention of many researchers in this field.

Based on the classification raised in [13], five categories were considered for automatic image annotation(AIA): 1- Generative model-based AIA method, which predicts the image tags based on a joint probabilistic model of image features and word available in the training dataset. 2- In the nearest neighbor-based image annotation method, the query image tags are performed based on visual similarity calculation and the selection of image tags having contents highly similar to the query image. Distance metric learning (DML) is one of the subsets of this category, which has many uses in the fields of machine learning and data mining. 3- Discriminative model-based image annotation method selects the tags for each of theme based on the independent binary learning action and using the result of binary classification. 4- Tag completion-based image annotation. The difference between this category and the former is in that it is possible to fill in the missing tags automatically, no need for training processes, as well as to correct noisy tags. Some methods used in this category are Matrix completion-based methods, Linear space reconstruction-based methods, and subspace clustering-based methods. 5- In Deep Learning-based image annotation, the process of tag assigning has got two general stages. In the first stage image features are generated using convolutional neural network, and in the second information such as semantic tags relationship is used.

Considering the challenges in this area, there has been a large number research carried out in recent years. Several types of research will be mentioned in the following:

The basis of the instant-based algorithms is to compare each input with the entire training set. This method is a component of non-parametric algorithms in which all the hypotheses are constructed based on training

set [14]. The conducted works have often used the hybrid models in order to define a common distribution on image attributes and tags [15, 16]. [17] Addresses automatic image annotation through transferring the tag to a semantic space. Accordingly, the semantic gap problem can be reduced by constructing a semantic space which includes a combination of visual and textual information. In [18] a voting directional graph-based framework is presented for retrieving related tags. [19] Presents a model called ML-RANK, which aims at ranking the tags related to an image based on visual similarities and semantic relations. [20] Given that some images have unclear content and their detection is somewhat difficult, other images and metadata are likely to exist in the neighborhood of them, which help to identify and recognize the content of the desired image. [21] Describes the problem of learning the relationship between labels as a basis for improving their descriptive power. Specifically, it defines a supervised neighbor voting stage that labels relations obtained by visual neighbors. The structure of this system consists of two main parts: 1- the formulation of relationship between tags and 2- ranking oriented learning. [22] Uses a two-step algorithm to retrieve tags. First, the initial score of the associated tag is calculated for each of tag using kernel density estimation (Gaussian), and second, a random motion on the tag graph is done by which the edges of a related tag are weighted based on similarity. [23] Proposes a framework for multi-query expansion to retrieve semantically robust landmarks of a user's query using Latent Dirichlet Allocation (LDA) technique for incomplete and poorly queries from the user. In [24] uses a data-driven approach as well as knowledge extracted from the generated tags by users, available resources on the web, images uploaded by users and visual similarity of key frames to refine video tags in order to increase the number of initial tags provided by users.

Some good studies whose main ideas are based on neighbor voting have been carried out [27-30]. In paper [27], the tags are clustered together based on the graph connectivity (communities) and the most important community is selected and corresponding tags are assigned to the query image. In paper [28], first adopt a genetic-based prototyping algorithm to obtain optimal

prototype from images. Then, for a given query image, its neighbor images are retrieved from the optimal prototype gained, and to generate its candidate tags some methods such as voting are used.

3. PROPOSED METHOD

In this section, the method which is automatic image annotation with the aim of ranking the image tags will be proposed.

3.1. Methodology

The purpose of this research is to low-level features at two steps in order to the image's content more precisely to improve retrieved labels. The process is showed in figure 1.

Neighbor voting method which is based on most repeated labels in neighboring images is used in step 1, to determine primary labels [10]. In the second step, content features of neighbor images are used and to determine the relevance existing among the labels and the query image. Each tag will take different weights according to the level of content features of the neighbor images. Finally, based on the weights assigned to each tag, re-ranking is performed.

As shown in Figure 1, in first step, determining the initial tags of the image are carried out neighbor voting method. The neighbor voting method predicts query image labels based on the most repeated tags in visual neighbors. Therefore, one of the most effective parameters in this method is the number of selected neighbors for voting [10]. The whole process of choosing initial tags using neighbor voting method is shown in Figure 2.

In the first step and according to equation (1), the feature vector of query image is compared to all feature vectors of annotated images based on the cosine similarity and by extracting the content features of the query image (QI) and the annotated image (AI).

$$\text{Cosine}(QI, AI) = \frac{\sum_{i=1}^n QI_i \times AI_i}{\sqrt{\sum_{i=1}^n QI_i^2} \times \sqrt{\sum_{i=1}^n AI_i^2}} \quad (1)$$

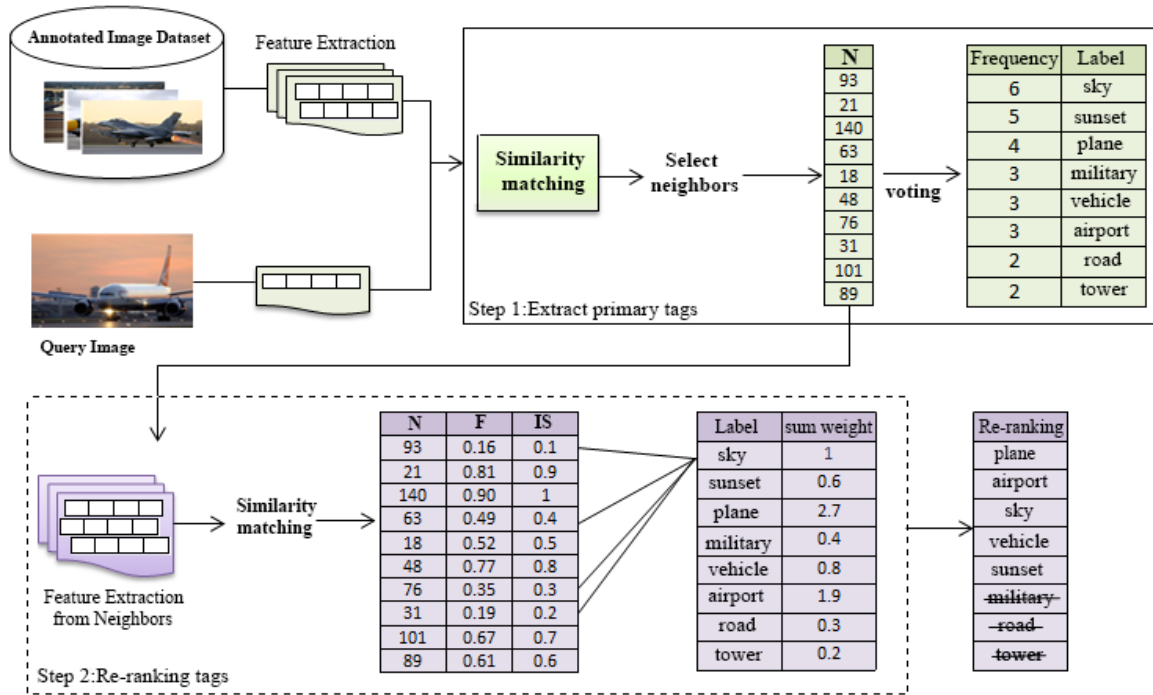


Fig.1 Proposed method: step1 neighbor voting method, step 2 Tag re-ranking based on combining visual features of neighbor images

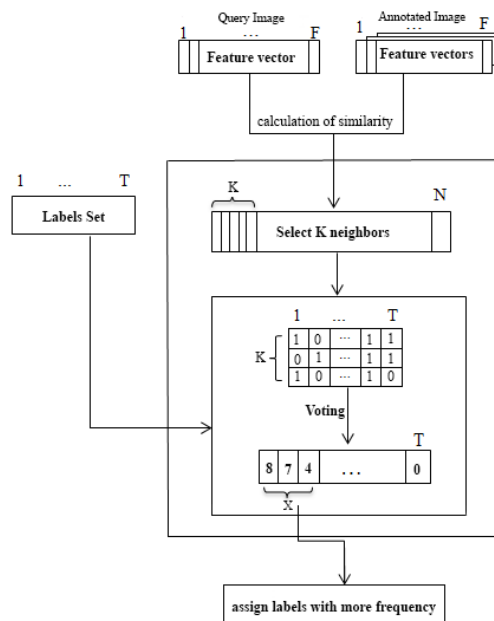


Fig.2 Neighbor voting method (Extract initial tags)

In order to calculate the similarity of the query image with any annotated images in training set, a value which indicates the contents similarity of two compared images will be obtained. Afterwards, based on the obtained values, the number of K images having the highest value will be selected as the query neighbor images. After choosing the neighbor images, they will be checked for all tags they have. By counting the tags in these images, the x tag that has the most frequency in the neighbor images is considered as the initial tag of the query image.

3.2. Tag re-ranking based on combining visual features of neighbor images

As it was previously mentioned in the first step, k neighbors having the most similar contents are chosen for the query image. In this stage, the same neighbors are used and low-level features

(CORP, WT, CH) are extracted in addition to the features used in the previous step. As shown in Figure 3, the features obtained from neighbor images are combined, which is a single feature vector of different features obtained for each image. Since the number of neighboring images is very lower than the total number of images, then the feature extraction step will not have a high time complexity.

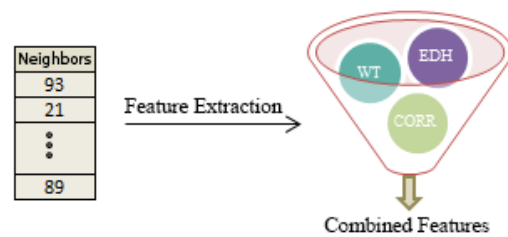


Fig.3 Combine content features of neighbor images

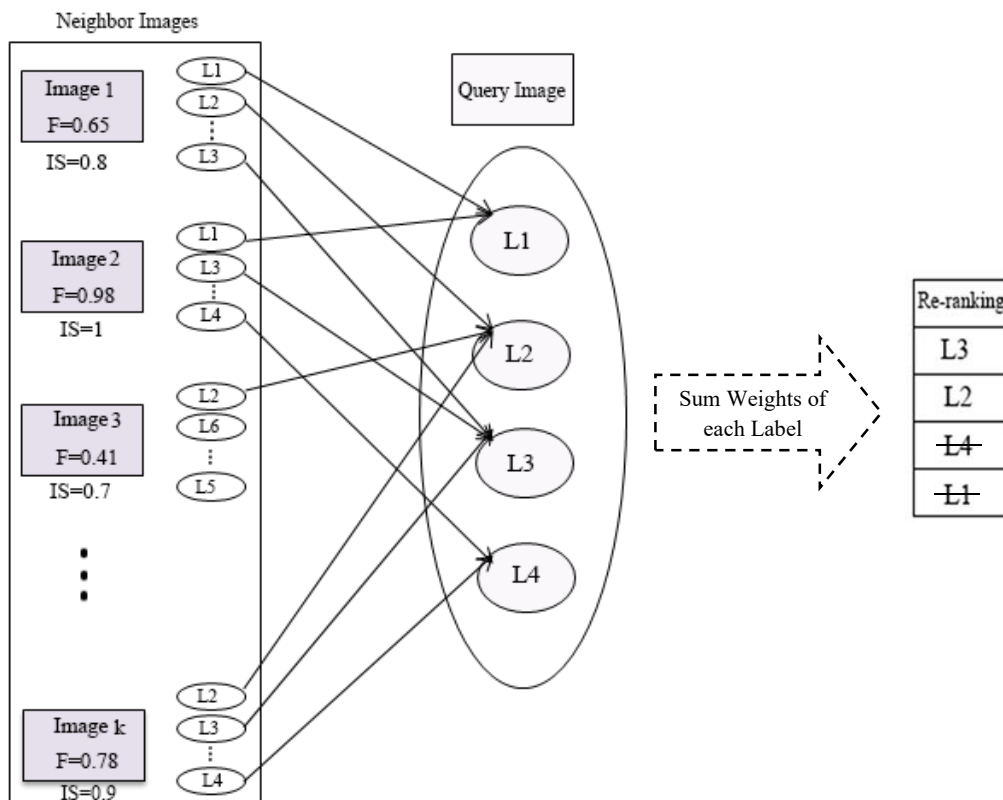


Fig.4 Determining the importance of retrieved labels according to the content feature of neighboring images

In this step, the similarity of the query image to all its neighbors is recalculated based on new features and using the cosine similarity measure. According to the value obtained for each neighbor in the similarity calculation step, an importance score (IS) in the range of [0,1] is assigned.

Now, each initial tag obtained for the query image from step 1 is checked if it belongs to the set of the selected labels of neighbor. This comparison among retrieved tags is done for all the neighbors of the query image one by one as shown in Figure 4.

If the tag exists in the Neighbor Image1 tag set, the score will receive the importance score of that image, otherwise, it will not be assigned any score. This process will be repeated for k neighbors. Finally, the weights each tag receives in accordance to the importance scores of its neighbors are aggregated based on equation (2) and then will be re-ranked. At a new ranking, the tags with highest weight will be selected, and unrelated tags will be deleted.

$$W_{tag} = \sum IS_{neighbors} \quad (2)$$

4. RESULTS AND EXPERIMENTS

In this section, the evaluation of the proposed method is discussed. First, the dataset used will be introduced along with the evaluation metrics. Afterwards the results will be expressed using the proposed method.

4.1. Data set and features

In the proposed method, a real image data set called NUS-WIDE [25] is used. NUS-WIDE data set is one of the large-scale image data sets provided by the Multimedia Information Laboratory at Singapore National University. This data set contains 269,648 images collected from Flickr website having 5018 tags created by users. By doing a preprocessing step, it turned out that images having more than two tags were 77,486 and were used to evaluate the work. Six

types of feature including 64_D color histogram, 144_D color correlogram, 73_D edge direction histogram, 128_D wavelet texture, 225_D block-wise color moments, and 500_D bag of words based on SIFT descriptions are calculated for each image in this dataset. Semantic meanings should be specified in the images in order to evaluate the search results based on the tag. For this purpose, NUS-WIDE data set, selects 81 semantic contents and uses it as ground-truth. To evaluate the efficiency of the proposed method, we compare the results of the experiments using low-level features of images in two steps to the results of one-step experiments.

The effectiveness of image retrieval system depends on the type of shape representation used. In this paper, the selection features that have improved the ranking of labels, include combination CORR, EDH and WT. Among the color features of the image, color histogram (CH) easy to calculate but captures only the color distribution in an image, in addition it's very sensitive to noise. One of the other color method used in feature extraction is a color moment (CM) that not enough to describe all colors and it has not spatial information. So we decided to use a feature that is more powerful than other image color features. The highlights of color correlogram (CORR) includes expresses that how the spatial co-relation of pairs of colors changes with distance. A CORR feature for an image is described as a table indexed by color pairs, where the d^{th} entry at location (i,j) is calculated by counting number of pixels of color j at a distance d from a pixel of color i in the image, divided by the total number of pixels in the image. It thus rectifies the major drawbacks of the CH and CM method. The wavelet texture (WT) is also an important cue for the analysis of many types of images. Actually, the term is used to point to intrinsic characteristic of surfaces contains intuitive properties like roughness, granulation and regularity. Some of the benefits of WT comparison to other low-level features are Meaningful, Easy to understand and can be extracted from any shape without losing information. While color describes only one pixel, the texture is described by a group of pixels. In common, Region based approach may not be easy and reliable for a diverse collection of images due to the unavailability of fully automated generalized approach, edge detection as more reliable and

it contains rich texture and shape information. Edge-detection based feature extraction method uses perimeter, curvature, edge direction etc. it describes the features based on their orientations and correlation between neighboring edges, and it is invariant to translation, viewing position and small rotation.

As explained above we have used three types of features for capturing color and shape information. First normalize all measure of the features used by max_min normalization. Then all three features used as a single feature vector to compare the test and train images. This enables simultaneous analysis of several features side by side from an image. Analytics include color variations, image curves, and shapes which can extract conceptual information from the image.

4.2. Metrics evaluation

To evaluate the results of re-ranking tags, three known evaluation metrics in the field of information retrieval namely F-measure, Recall, Precision are used. N_c is the number of correctly retrieved tags, N_s total number of retrieved tags, and N_r is the total number of available real tags. The F-measure is a measure of a query's accuracy and is defined as the weighted harmonic mean of the precision and recall of the query.

$$\text{Precision}(v_i) = \frac{N_c}{N_s} \tag{3}$$

$$\text{Recall}(v_i) = \frac{N_c}{N_r} \tag{4}$$

$$\text{F1} = \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \tag{5}$$

4.3. Results

According to the research hypothesis the use of the content features of neighbor images and their combination together to the ranking of initial tags can improve the precision of annotation. Figures 5 and 6 proves the precision of the proposed method based on Precision and Recall values in comparison with RWLabel [22] and LSLabel [26].

As shown in Figure 5, the precision of the proposed method is improved in comparison with previous methods. In the RWLabel method, tags are assigned to the image based on the random walking algorithm. The LSLabel method uses the inner correlation to get more information from tags. In the above mentioned methods the use of visual features of the image is low. In the other words, there is no way to measure the relevance of each tag with the image and refinement of final tags before final tag assignment. In the proposed method, the goal is to maximize the content features of the image. In one step experiment the features in the entire data set are used to find similar images to the query image. In the next step, after determining the initial tags, more features are extracted and combined to specify the relevance of each tag with the query image.

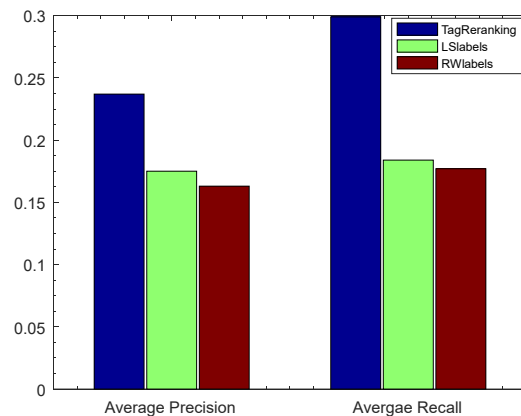


Fig.5 The comparison of average precision and recall

Another way to test the proposed method is to calculate the precision per tag used for annotation. Figure 6 illustrates the precision of partial tags in the NUS-WIDE data set. The Figure shows that some tags such as {“ dog”, “fish”, ”ocean”, ”water”} are more precise than others. One main reason for achieving these results are the clarity of features

Table.1 Impact of different number of neighbors on result

Number of neighbors	Number of tags			
	evaluation metrics	5	7	10
K=20	Precision	0.204	0.210	0.218
	Recall	0.166	0.198	0.310
	F_measure	0.181	0.203	0.247
K=35	Precision	0.181	0.188	0.196
	Recall	0.160	0.219	0.318
	F_measure	0.167	0.201	0.241
K= 50	Precision	0.178	0.198	0.210
	Recall	0.158	0.211	0.308
	F_measure	0.166	0.202	0.249
K=70	Precision	0.169	0.214	0.228
	Recall	0.155	0.204	0.301
	F_measure	0.160	0.208	0.258
K=90	Precision	0.178	0.219	0.237
	Recall	0.151	0.224	0.298
	F_measure	0.161	0.221	0.263
K=120	Precision	0.162	0.199	0.199
	Recall	0.148	0.238	0.288
	F_measure	0.151	0.215	0.235
K=200	Precision	0.155	0.178	0.179
	Recall	0.140	0.219	0.280
	F_measure	0.145	0.193	0.217
K=500	Precision	0.141	0.164	0.156
	Recall	0.132	0.196	0.239
	F_measure	0.134	0.177	0.183
K=1000	Precision	0.136	0.158	0.142
	Recall	0.129	0.180	0.241
	F_measure	0.132	0.165	0.178

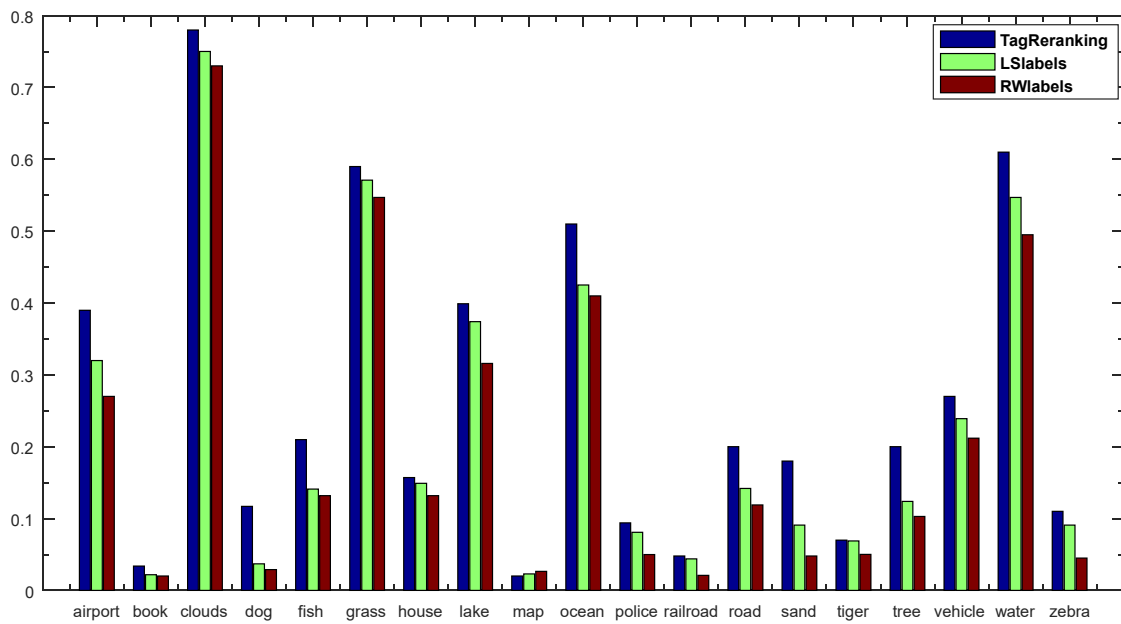


Fig.6 The average Precision of the partial labels

such as wavelet texture of these images, which is extracted from the query image neighbors after the neighbor voting stage, and the significance of the image will change in relation to the value of this feature. The fluctuation of importance scores in neighbor images makes it more prominent in the image and uses its result to weight the initial tags to determine the relevance of each tag to the query image.

Given that the selection of primary tags is carried out by the neighbor voting method and the number of selected neighbors directly affect its result, in the table 1, test to is goal the proposed method using a different number of neighbors and examine its impact on the precision, recall and f-score after extracting new features from neighbor images of the query image. The minimum selected neighbors are 20 and maximum is 1000, where the highest precision is k=90.

In order to achieve maximum precision, different similarity metrics are used in the experiments. In figure 7 the comparison of three similarity metrics including Cosine, Euclidean, Manhattan and their effects on precision is shown and defined by Eq. (1), (6) and (7) respectively.

$$d_E(QI, AI) = \sqrt{\sum(QI_i - AI_i)^2} \quad (6)$$

$$d_M(QI, AI) = \sum|(QI_i - AI_i)| \quad (7)$$

Where, QI denotes feature vector of query image and AI denotes feature vector of annotated image. Fig.7 indicates that Cosine distance is better than the other two metrics.

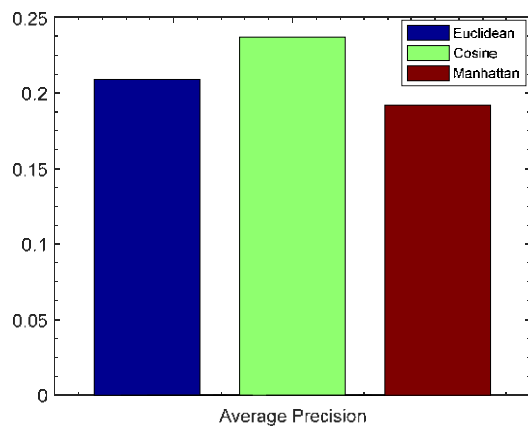


Fig.7 The comparison of the performance of different similarity measures

5. CONVLUSION

The purpose of systems automatic image annotation is to retrieve images based on a text query. In this paper, a method was presented for re-ranking of automatic image annotations based on the extraction of content features from neighbor images. As almost all previous works used content features of the image in one step, this paper attempts to use low-level features of the image to obtain more information in two steps. In the first step, based on the image content feature, the image similarity is measured for two sets of the train and test using the neighbor voting method and then the initial tags are determined. In the second step, neighbors of each image in test set obtained from the previous step is used to acquire more information in order to match the query image based on low level feature and then they are combined. The main difference is that low level features are only used in one step experiments.

To carry out further works in the field of image annotation and based on the proposed method, the application of efficient algorithms in the field of extracting the appropriate features will improve the precision of automatic image annotation. Of course, it should be noted that annotation time is a very important parameter that needs to be addressed.

REFERENCES

1. R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Image retrieval: Ideas, influences, and trends of the new age," *ACM Comput. Surv.*, vol. 40, no. 2, pp. 1–60, Apr. 2008.
2. X. Chang, H. Shen, S. Wang, J. Liu, and X. Li, "Semi-supervised Feature Analysis for Multimedia Annotation by Mining Label Correlation," pp. 74–85, 2014.
3. X. Li, T. Uricchio, L. Ballan, M. Bertini, C. G. M. Snoek, and A. Del Bimbo, "Socializing the semantic gap: A comparative survey on image tag assignment, refinement, and retrieval," *ACM Comput. Surv.*, vol. 49, no. 1, pp. 1–39, Jun. 2016.
4. M. Wang, B. Ni, X.-S. Hua, and T.-S. Chua, "A survey of multimedia tagging with human-computer joint exploration," *ACM Comput. Surv.*, vol. 44, no. 4, pp. 1–24, Aug. 2012.
5. Yue Gao, Meng Wang, Zheng-Jun Zha, Jialie Shen, Xuelong Li, and Xindong Wu, "Visual-Textual Joint Relevance Learning for Tag-Based Social Image Search," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 363–376, Jan. 2013.
6. Y. Liu, D. Zhang, G. Lu, and W.-Y. Ma, "A survey of content-based image retrieval with high-level semantics," *Pattern Recognit.*, vol. 40, no. 1, pp. 262–282, Jan. 2007.
7. D. Zhang, M. M. Islam, and G. Lu, "A review on automatic image annotation techniques," *Pattern Recognit.*, vol. 45, no. 1, pp. 346–362, Jan. 2012.
8. A. E. Brito, D. Kletter, M. Singhal, and M. Bern, "Benchmark study of automatic annotation of MALDI-TOF N-glycan profiles," *J. Proteomics*, vol. 129, pp. 71–77, Nov. 2015.
9. S. Protasov, A. M. Khan, K. Sozykin, and M. Ahmad, "Using deep features for video scene detection and annotation," *Signal, Image Video Process.*, pp. 1–9, Jan. 2018.
10. X. Xirong Li, C. G. M. Snoek, and M. Worring, "Learning Social Tag Relevance by Neighbor Voting," *IEEE Trans. Multimed.*, vol. 11, no. 7, pp. 1310–1322, Nov. 2009.
11. D. Tian and Z. Shi, "Automatic image annotation based on Gaussian mixture model considering cross-modal correlations," *J. Vis. Commun. Image Represent.*, vol. 44, pp. 50–60, Apr. 2017.
12. K. Akhilesh and R. R. Sedamkar, "Automatic image annotation using an ant colony optimization algorithm (ACO)," in *2016 IEEE 7th Power India International Conference (PIICON)*, 2016, pp. 1–4.
13. Q. Cheng, Q. Zhang, P. Fu, C. Tu, and S. Li, "A survey and analysis on automatic image annotation," *Pattern Recognit.*, vol. 79, pp. 242–259, Jul. 2018.
14. S. Lee, W. De Neve, and Y. M. Ro, "Visually weighted neighbor voting for image tag relevance learning," *Multimed. Tools Appl.*, vol. 72, no. 2, pp. 1363–1386, Apr. 2013.
15. G. Carneiro, A. B. Chan, P. J. Moreno, and N. Vasconcelos, "Supervised Learning of Semantic Classes for Image Annotation and Retrieval," *IEEE Trans. Pattern Anal.*

Mach. Intell., vol. 29, no. 3, pp. 394–410, Mar. 2007.

16. T. Uricchio, L. Ballan, M. Bertini, and A. Del Bimbo, “An evaluation of nearest-neighbor methods for tag refinement,” in 2013 IEEE International Conference on Multimedia and Expo (ICME), 2013, pp. 1–6.

17. T. Uricchio, L. Ballan, L. Seidenari, and A. Del Bimbo, “Automatic image annotation via label transfer in the semantic space,” *Pattern Recognit.*, vol. 71, pp. 144–157, Nov. 2017.

18. X. Zhu, W. Nejdl, and M. Georgescu, “An adaptive teleportation random walk model for learning social tag relevance,” in Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval - SIGIR '14, 2014, pp. 223–232.

19. Z. Li, J. Liu, C. Xu, and H. Lu, “MLRank: Multi-correlation Learning to Rank for image annotation,” *Pattern Recognit.*, vol. 46, no. 10, pp. 2700–2710, Oct. 2013.

20. J. Johnson, L. Ballan, and L. Fei-Fei, “Love Thy Neighbors: Image Annotation by Exploiting Image Metadata,” in 2015 IEEE International Conference on Computer Vision (ICCV), 2015, pp. 4624–4632.

21. C. Cui, J. Shen, J. Ma, and T. Lian, “Social tag relevance learning via ranking-oriented neighbor voting,” *Multimed. Tools Appl.*, vol. 76, no. 6, pp. 8831–8857, Mar. 2017.

22. D. Liu, X.-S. Hua, L. Yang, M. Wang, and H.-J. Zhang, “Tag ranking,” in Proceedings of the 18th international conference on World wide web - WWW '09, 2009, p. 351.

23. Y. Wang, X. Lin, L. Wu, and W. Zhang, “Effective Multi-Query Expansions: Collaborative Deep Networks for Robust Landmark Retrieval,” *IEEE Trans. Image Process.*, vol. 26, no. 3, pp. 1393–1404, Mar. 2017.

24. L. Ballan, M. Bertini, G. Serra, and A. Del Bimbo, “A data-driven approach for tag refinement and localization in web videos,” *Comput. Vis. Image Underst.*, vol. 140, pp. 58–67, Nov. 2015.

25. T.-S. Chua, J. Tang, R. Hong, H. Li, Z. Luo, and Y. Zheng, “NUS-WIDE: a real-world web image database from National University of Singapore,” in Proceeding of the ACM International Conference on Image and Video Retrieval - CIVR '09, 2009.

26. F. Tian, X. Shen, and X. Liu, “Multimedia automatic annotation by mining label set correlation,” *Multimed. Tools Appl.*, vol. 77, no. 3, pp. 3473–3491, Feb. 2018.

27. Maihami V, Yaghmaee F. Automatic image annotation using community detection in neighbor images. *Physica A: Statistical Mechanics and its Applications*. 1;507:123-32, 2018.

28. Maihami V, Yaghmaee F. A genetic-based prototyping for automatic image annotation. *Computers & Electrical Engineering*. 1;70:400-12, 2018.

29. Lotfi A, Maihami V, Yaghmaee F. Wood image annotation using gabor texture feature. *Int. J. Mechatronics, Electr. Comput. Technol.*;4:1508-23, 2014.

30. Maihami V, Yaghmaee F. A review on the application of structured sparse representation at image annotation.

Artificial Intelligence Review. 1;48(3):331-48, 2017.