Automatic Construction of Image Data Set Using Query Expansion Based on Tag Information

Yuki Mizoguchi Takahiro Yoshimura Yuji Iwahori Ryosuke Yamanishi

Department of Computer Science
Chubu University
Kasugai, 487-8501 Japan
Mail:mizoguchi@cvl.cs.chubu.ac.jp
Mail:yoshimura@cvl.cs.chubu.ac.jp
Mail:iwahori@cs.chubu.ac.jp

College of Information Science and Engineering
Ritsumeikan University
Kusatsu, 525-8500 Japan
Mail:ryama@media.ritsumei.ac.jp

***Abstract*—Although some automatic construction methods of image data set have been proposed based on the web data, these methods used tag search for the image collection. The problem is that only images which have a target tag can be collected. This paper focuses on the tag given to the image of the data set constructed by the previous method. Tags which have been added to the image of the data set with high frequency and have high relevance to the target label are selected as new search queries. Furthermore, images collected based on the new search queries are classified by classifiers learned with the data set of the target labels constructed by the previous method. The goal is to construct and to expand image data set including images without the target label.**

Keywords–Image Data Set; Query Expansion.

I. INTRODUCTION

In the research of general object recognition, the computer recognizes images without constraints, and researches have focused on feature extraction and machine learning methods. A huge image data set with various unbiased objects is required for the learning. A manual collection of huge image data set causes biased collection and a lot of human costs.

Recent researches for image data set try to automatically or semi-automatically generate image data set from Web image sharing services, such as Flickr, i.e., Web image mining [1]. Web image mining enables makes it possible to obtain a large amount of images including daily scenes with low cost [2]. One of the problems in Web image mining is that the image data set generated with Web image mining naturally includes noise images. The method for noise removal using image features has been reported [3].

However, in collecting images from the web by most previous methods, tag search is used with the target label as a query, and the collection range is limited to images in which target words are included in meta data given to images, such as tags, titles, and descriptions.

Therefore, this paper proposes extending the data set by selecting appropriate queries and collecting images that can not be collected by the previous method.

In Section 2, proposed method is given. In Section 3, peformance is evaluated in the experiments. In Section 4, conclusion and future work are discussed.

II. PROPOSED METHOD

The proposed method consists of the five steps: 1) Collecting images from web, 2) Noise images removal with visual features, 3) Query expansion based tag of image, 4) Recollecting images from web with new query, and 5) Noise images removal with visual features.

The detailed procedure of the proposed method is as follows.

Step1.	Collect images by tag search using target label as query.
Step2.	Removal of noise image by visual information and set the remaining image as "correct image".
Step3-1.	Calculate the appearance frequency from the tag given in "correct image".
Step3-2.	Calculate the cosine similarity between the tag given to "correct image" on fastText and the target label.
Step3-3.	Select tags that matches "high occurrence frequency" and "high similarity with target label" among the tags as new queries.
Step4.	Collect images by tag search using query selected by Step3.
Step5.	Removal of noise image by visual information as learning by "correct image".

A. Construction of Image Data Set

In this subsection, the Steps 1) and 2) are explained. The original idea of Steps 1) and 2) is from our previous paper [4]. To generate a data set as the basis of query expansion, images are collected according to target labels, and noise removal by visual information.

First of all, images are collected from the web as a query using a target label. Next, among the obtained images, images having minority image features are eliminated as noise images according to automatically determined threshold values.

Image features are feature vectors representing SIFT features as Fisher Vectors. In this paper, SIFT features are extracted by grid sampling, extraction interval is $8px$, extraction window size is $16px$. The number of visual words K when calculating Fisher Vector from the image was set to 512. Get Euclidean distance $FD^l(i)$ between each image feature amount $G(i)$ and the centroid vector M^l of all images. We classify images collected by query l based on $FD^l(i)$ as follows.

$$VI^l = \begin{cases} correct\_image & (FD^l(i) \leq VT^l) \\ noise\_image & (FD^l(i) > VT^l) \end{cases} \quad (1)$$

 ISBN: 978-1-61208-612-5

Here, VI^l is a threshold value, which is the average value of $FD^l(i)$ for each NI images.

$$VI^l = \frac{1}{NI}\sum_{i=1}^{NI} FD^l(i) \tag{2}$$

B. Query Expansion

In this subsection, the Step 3) will be shown. Query expansion based on data set constructed by previous method. We focus on the tag of the image inside the data set as the basis of the query extension.

It is conceivable that the query to be selected is included in the tag inside the data set constructed by the previous method. Therefore, the appearance frequency is used as the first criterion of query expansion.

Here, the appearance frequency $hist$ of a certain label is expressed as in (3) when $count$ is the number of images given the label among the number of images num .

$$hist = \frac{count}{num} \tag{3}$$

In addition, tags have high frequency of appearance, but there are tags with low relevance to target labels. For example, words, such as 'Canon' representing a device that photographed the picture and 'London' representing the point where the picture was taken are often tags that the user gives to the image in Flickr. In order to filter such tags with a general relationship, cos similarity on fastText is used as a second criterion.

Here, the cos similarity sim in a fastText of a label and a target label is expressed by the expression (4) when M is the vector of the target label on fastText and V is the vector of a label on fastText .

$$sim = \frac{M \cdot V}{|M||V|} \tag{4}$$

Among tag given to the image in data set of target label, a tag having "high occurrence frequency" and "high similarity" is set as a new search query. In order to express "high occurrence frequency" and "high similarity degree", the criterion $Score$ of tag selection is assumed to be multiplied by $hist$ and sim, and expressed as in (5) . Here, exclude tags that sim can not calculate because there is no word on fastText.

$$Score = hist * sim \tag{5}$$

Finally, rank tags based on $Score$. However, we exclude the target labels themselves and the conjugative form, the plural forms of target labels.

C. Expansion of Image Data Set

In this subsection, the Steps 4) and 5) will be shown. First, images are collected from the web by the query selected in Step 3. Second, convert the collected images to image features according to step 2. Finally, the image outside the threshold is removed as a noise image by using the threshold of the image feature amount used in step 2. The remaining image is added to the data set as a correct image of the target label, and the data set is expanded.

III. EXPERIMENTS

A. Query Expansion Experiments

In query expansion experiments, we show that the proposed method can expand the query. We collect 2,500 images from Flickr for each label and experiment on 5 target labels of dog, cat, kick, throw, jump. Queries selected by query expansion on the proposed method are shown in Tables I-V. Although the target label in the table is not $hist = 1$, this is because Flickr's search algorithm gathers capital letters and lowercase letters without distinction.

As shown in Tables I-V, tags that are not suitable as queries like 'Canon' or 'London' have been removed.

TABLE I. TAGS SELECTED BY QUERY EXPANSION TO DOG

tag	hist	sim	Score
dog	0.411302	1.000000	0.411302
Dog	0.435850	0.497419	0.216800
Competicion	0.327466	0.230190	0.075379
Elechas	0.327466	0.201274	0.065910
Agility	0.327466	0.188627	0.061769
can	0.328393	0.175213	0.057539
Agility Cantabria	0.327466	0.174608	0.057178
deporte	0.327466	0.147234	0.048214
dogs	0.072256	0.659097	0.047624
puppy	0.085225	0.402321	0.034288

TABLE II. TAGS SELECTED BY QUERY EXPANSION TO CAT

tag	hist	sim	Score
cat	0.603858	1.000000	0.603858
cats	0.201884	0.707840	0.142902
kitten	0.291611	0.415140	0.121059
Cat	0.165545	0.469826	0.077777
catsagram	0.157470	0.475644	0.074900
catsoftwitter	0.157470	0.474278	0.074684
catsofinstagram	0.159264	0.413597	0.065871
kittens	0.157918	0.396114	0.062554
kitty	0.202333	0.295444	0.059778
cute	0.221175	0.246822	0.054591

TABLE III. TAGS SELECTED BY QUERY EXPANSION TO KICK

tag	hist	sim	Score
kick	0.491503	1.000000	0.491503
Kick	0.495528	0.539903	0.267537
Referee	0.294275	0.421585	0.124062
Tackle	0.294275	0.382568	0.112581
Soccer	0.297853	0.368589	0.109785
Football	0.294723	0.315795	0.093072
Pass	0.294275	0.302313	0.088963
Team	0.294275	0.301781	0.088807
Header	0.288014	0.279592	0.080526
Shot	0.294275	0.254012	0.074750

TABLE IV. TAGS SELECTED BY QUERY EXPANSION TO THROW

tag	hist	sim	Score
throw	0.804951	1.000000	0.804951
Throw	0.161758	0.801584	0.129663
ball	0.163892	0.512881	0.084057
throwout	0.096458	0.840592	0.081081
sports	0.288092	0.259825	0.074854
pitcher	0.170294	0.418618	0.071288
sport	0.183099	0.320428	0.058670
catcher	0.105847	0.489021	0.051761
catch	0.091336	0.548228	0.050073
baseball	0.152796	0.327365	0.050020

 ISBN: 978-1-61208-612-5

TABLE V. TAGS SELECTED BY QUERY EXPANSION TO JUMP

tag	hist	sim	Score
jump	0.853669	1.000000	0.853669
volleyball	0.220582	0.363222	0.080120
Jump	0.110725	0.633634	0.070159
team	0.260096	0.251929	0.065526
jumping	0.095962	0.630887	0.060541
spike	0.196266	0.306962	0.060246
sport	0.123752	0.334476	0.041392
block	0.148502	0.272300	0.040437
School	0.184542	0.205032	0.037837
High	0.225792	0.156843	0.035414

B. Evaluation

It is investigated if the image of the data set expanded by the proposed method matches the target label. Queries are superior ranked Score, except for themselves, the conjugative forms and plural forms of target labels from the Tables I-V. These are shown in Table VI. Based on these queries, 5000 images are collected for each, remove noise and build a data set. 200 images were randomly extracted among the constructed data set and the evaluators evaluated for the question whether the image of the data set constructed matches the target label or not.

In this experiment, the image selected that 70% or more of the evaluators match the target label in the questionnaire is taken as the correct image. Among the images of the data set constructed by the proposed method, the ratio of the correct image by subjective evaluation is taken as accuracy. The results of the evaluation experiment are shown in Table VI. Also, an example of the image of the data set constructed by the proposed method is shown in Figures 1-5.

TABLE VI. TARGET LABEL, SELECTED TAG AND ACCURACY

target label	selected tag	accuracy[%]	*1	*2
dog	Agility	93.2	187	174
cat	kitten	91.8	182	167
kick	soccer	28.8	46	13
throw	picher	69.9	97	66
jump	volleyball	14.7	34	5

*1 : Number of correct images by subjective evaluation
*2 : Among the images judged as correct by the proposed method, the number of correct images by subjective evaluation

In dog-Agility, cat-kitten, these accuracy are as high as 90% or more. Most images collected by the query selected by the proposed method are that match the target labels. Thus, proposed method selected useful queries.

Although some other kind of images were obtained from ”pitcher”, noise removel could remove the images except baseball pitcher and its accuracy was around 70%.

The accuracy for some images with kick-soccer and jump-volleyball was low. This is because the number of correct answer images were also small. On Frickr, when collecting images using selected tags as queries, the images clearly matching target labels were significantly smaller than other labels.

Figure 1. Ex.:dog-Agility

Figure 2. Ex.:cat-kitten

Figure 3. Ex.:kick-soccer

Figure 4. Ex.:throw-pichar

Figure 5. Ex.:jump-volleyball

IV. CONCLUSION AND FUTURE WORK

This paper proposes an image data set expansion method that selecting appropriate queries and collecting images that can not be collected in previous methods.

In the experiment, it is possible to select a useful query for data set extension by query expansion from the image tag of the data set constructed by the previous method.

As a future task, in order to select more useful queries, we consider not only combinations of frequencies and relevance, but also combinations of ”occurrence frequency of tags of images removed as noise” or TF-IDF.

ACKNOWLEDGEMENT

Iwahori’s research is supported by Japan Society for the Promotion of Science (JSPS) Grant-in-Aid Scientific Research(C)(#17K00252) and Chubu University Grant.

REFERENCES

[1] 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 Proceedings of the ACM international conference on image and video retrieval. ACM, 2009, p. 48.

[2] X. Li, C. G. Snoek, and M. Worring, “Unsupervised multi-feature tag relevance learning for social image retrieval,” in Proceedings of the ACM International Conference on Image and Video Retrieval. ACM, 2010, pp. 10–17.

[3] R. Fergus, P. Perona, and A. Zisserman, “A visual category filter for google images,” Computer Vision-ECCV 2004, 2004, pp. 242–256.

[4] S. Otani, R. Yamanishi, and Y. Iwahori, “Generation of web image database based on hybrid noise removal method of visual and semantic feature,” Journal of Japanese Society for Artificial Intelligence, vol. 32, no. 1, 2017, pp. WII–N_1–10.

 ISBN: 978-1-61208-612-5