International Journal of Advances in Image Processing Techniques– IJIPT
Volume 4: Issue 1 [ISSN : 2372-3998]
Publication Date : 06 April, 2017

Automatic Image Annotation using PHOW Features

Shereen A. Hussein, Howida Youssry Abd El Naby, Aliaa A. A. Youssif

Abstract

Image annotation is a task of assigning a set of semantic tags or keywords to unlabeled image based on a training set of data. This machine learning process depends on extraction and clustering the low-level features of images then mapping them to the semantic which is high-interest image retrieval. Many annotation techniques with reasonable performance have proposed in the last decade. The proposed algorithm for automatic annotation in this paper depends on similarity computation and label transferring to the query image. Similarity computing uses low-level image features and significant distance estimates. The feature vector of the test image compared with the feature matrix of similar and dissimilar image pairs in the training data sets. Then the label or keyword transfer from the similar image pair to the test image is performed by counting the local frequency of neighbor's keywords. Performance is evaluated using precision and recall.

Keywords: ***Automatic Image Annotation, Similarity, Feature Matrix, PHOW.***

1. Introduction

Automatic image annotation is an interesting topic in current searches due to its effective impact in image retrieval systems. It concerns in labeling the digital images with keywords express their contents. As a machine learning process, [1- 4] the correlation between extracted image features and the training annotation word list are used translate the textual vocabulary to visual vocabulary. The huge challenge in this process is to connect the various types of low-level features (colors, textures, and boundaries) to a high-level concept.

Automatic image annotation is regarded as a kind of multiclass image classification [5] and also known as linguistic indexing or automatic image tagging. Converting automatic image annotation into classification problem, make the annotation process [6– 8] easy under a standard semantic label. Therefore, feature extraction and selection is a vital strategy to guarantee reliable and meaningful results for data classification alongside different advantages such less data storage and computation cost. First, the extracted features are modeled in pyramid histogram of visual words then converted to form feature matrix then compute the similarity between image pairs to facilitate query image automatically annotation. The rest of the paper organized as follows. Section 2 gives a short review of automatic image annotation. Section 3 presents the proposed annotation method using PHOW features and similarity image pairs;
Section 4 shows the experimental results and analysis of the annotation on Corel5k and IAPR TC-12 dataset. Section 5 contains the conclusions and future work.

2. Related Work

Mori et al. [9] proposed the first attempts at image auto-annotation as transforming image to word based on word co-occurrence. The word co-occurrence is a linguistic term used in the natural image processing which refers to the often used words together. Duygulu et al. [10] proposed the image auto-annotation in a machine translation model. Focusing on region-based image annotation instead of the global image annotation because of its insufficiency in determine which part in image relate to which label. Jeon et al. [11] introduced the Cross-Media Relevance Model (CMRM) where visual information of an image denoted as blob set. It helps in clarification the image semantic information obviously. Putthividhya D. et al. [12] used Probabilistic latent semantic analysis (PLSA) model and latent Dirichlet allocation (LDA) to uncover the hidden themes in documents collections. Feng S et al. [14] proposed Bernoulli relevance model to improve CMRM and CRM accuracy. It is a generative statistical model that uses an annotated training set for the annotation process. Tianxia Gong, Shimiao Li, Chew Lim Tan et al. [15] used probabilistic models to present a framework to represent the word-to-word relation. Test images prediction tags depends on the classifier of visual features that is trained by the discriminative model. S. Deerwester et al. [16] proposed Latent Semantic Indexing (LSI) which deals with the problems of referring the same object with multiple words or some word with multiple meanings. So it overcomes the traditional lexical matching techniques shortcomings. Hofmann et al. [17] used the probabilistic LSI (PLSI) model, as an alternative to LSI. The PLSI is a more

Shereen A. Hussein / PHD Student
Faculty of Computer Science /Helwan University
Cairo, Egypt

Howida Youssry Abd El Naby / Assistant Professor and head
Information Technology faculty / Misr University for Science and Technology
Cairo, Egypt

Howida Youssry Abd El Naby / professor and dean
Faculty of Computers and Information, Helwan University
Cairo, Egypt

robust technique for automatic document indexing where each document represented by its word frequency. S. Zhang et al., [18] proposed group sparsity as a technique for automatic image annotation. It solves annotation as a retrieval problem and handles the features selection issue. Images pair similarity is evaluated by a specific value if it is positive means pair are similar otherwise are dissimilar. The overall performance is highly affected by pair similarity computations [19 – 28].

3. Proposed Framwork

An efficient method for automatic image annotation is to adopt labels from similar images in the dataset with keywords. Doing so has two sub-problems first, computing the image similarity and second, choosing the keywords from the similar images. Simple methods for these sub-problems proposed in the framework of creating feature matrix of n feature vector for each image in the training set and pairs of similar & dissimilar images. Image similarity measure has built from analysis on the histograms of a bag of visual words features matrix. The similarity within the image pair considers the similarity of keywords. Positive image pairs share some keywords, while negative ones do not. Based on the weights calculated the most similar images are found out from pairs. The quality of these pairs highly influences the overall performance. However, the ground truth of similarity is not available.

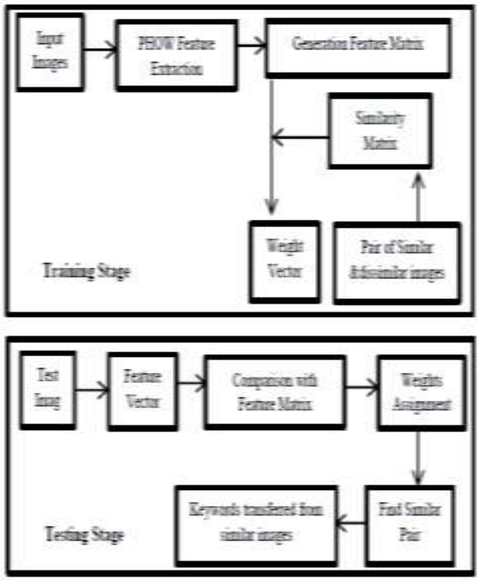


Figure 1 Proposed Framework Architecture

3.1 Feature Extraction (The PHOW Descriptors)

A Feature matrix for the images dataset created with the PHOW feature vectors. Image description by Pyramid of Histograms of Visual Words (PHOW) method is an extension to the bag-of-words (BOW) model in which the extracted SIFT image features treated as words. It considers the local information feature of the image [29-32]. The method implemented as in figure 2. PHOW [33, 34] overcomes the drawback of BOW of unavailability of spatial image features information by dividing the image into fine sub-regions (pyramids) and concatenating the histogram of each of these regions to the histogram of the original image with a suitable weight. When color images were processed, they are converted from RGB space to HSV color space with the SIFT feature extracted from each channel. For grayscale images, only the intensity is used. Hence, the resulting SIFT feature dimension is 128*3 for color images and 128 for grayscale images. Once the SIFT features obtained, "bag of words" model is used to quantize them into visual words by k-means clustering. Thus the image is represented by a histogram of visual word occurrences.

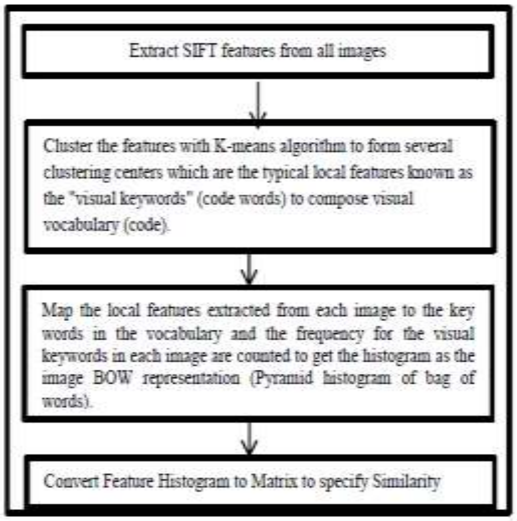

Extract SIFT features from all images
Cluster the features with K-means algorithm to form several clustering centers which are the typical local features known as the "visual keywords" (code words) to compose visual vocabulary (code).
Map the local features extracted from each image to the key words in the vocabulary and the frequency for the visual keywords in each image are counted to get the histogram as the image BOW representation (Pyramid histogram of bag of words).
Convert Feature Histogram to Matrix to specify Similarity

Figure 2 the PHOW Feature Matrix Steps

3.2 Obtaining Image Pairs

In this framework, the objective function is a similarity function [39-43]. Similar and dissimilar image pairs are required for training purposes. Since the goal is to assign relevant keywords, A. Makadia [7] proposed an approach to discover these pairs based on the keyword similarity [44-46]. Any pair of images in the training set that shares more than four keywords is considered as a positive training example, while a pair without any common keyword is a negative one. Obviously, the distances within positive pairs are expected to be smaller than the ones within negative pairs, since images of similar pairs should be much closer to each other than dissimilar ones as shown in figure 3.

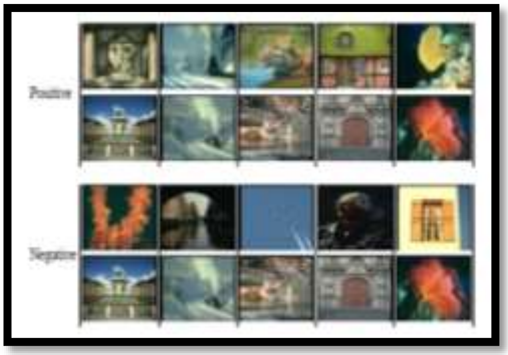


Figure 3 Positive & Negative Image Pairs [Makadia, Pavlovic and Kumar[7]]

3.3 Weight Vector Calculation

The weight of each feature vectors calculated by using feature matrix obtained using the PHOW method and set of pairs of similar and dissimilar images. The most efficient method to compute the weight vector is weighted least square (1) due to its ability to provide different types of easily statistical interpretation for estimation and prediction.

$W= \arg\min \| Xw - Y\|^2_2 \qquad w \in Rp \qquad (1)$

In the image annotation testing stage, this weight is used for computing similarity array by the iterative equation (2) for each row in the feature matrix. $S_i = (t_i - f_i) * wi \qquad (2)$

Where: - t is feature vector of test image.
f is feature value of feature matrix of training images.
w is the weight vector.

Then a comparison between the similarity array and the pairs of similar and dissimilar images is performed to choose a predefined number of values from s with the highest local frequency keywords which are transferred to annotate the image.

4 Experiments & Discussion

Here is a description of used benchmarks (Corel5K, IAPR TC-12) and evaluation of image annotation method.

4.1 Dataset

In Figure 4, there is some samples from Corel5K and IAPR datasets.

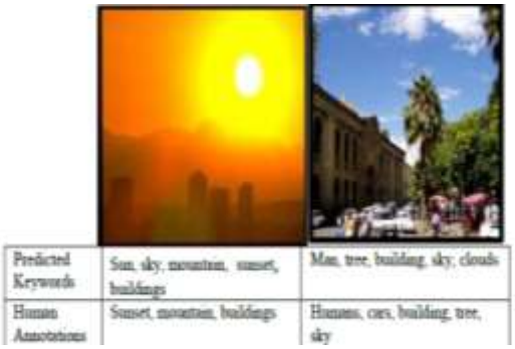
Figure 4 Image Samples of Corel 5k in the left, IAPR TC12 in the Right

– Corel5K [47, 48] has become a de-facto evaluation benchmark in the image annotation community. It contains 5,000 images collected from the larger Corel CD set, split into 4,500 training and 500 test examples. Each image is annotated with an average of 3.5 keywords, and the dictionary contains 260 words that appear in both the train and the test set.

– IAPR TC-12 [49] is a collection of 19,805 images of natural scenes that include different sports and actions, photographs of people, animals, cities, landscapes and many other aspects of contemporary life. Unlike other similar databases, images in IAPR TC-12 are accompanied by free-flowing text captions. While this set typically used for cross-language retrieval, we have concentrated on the English captions and extracted keywords (nouns) using the Tree Tagger part-of-speech tagger. That resulted in a dictionary size of 291 and an average of 4.7 keywords per image. 17,825 images used for training and the remaining 1,980 for testing.

4.2 Discussion

Understanding the scene is the key to solving the image annotation problem at the human level. However; objects identification and giving them keywords based on the given scene is still a hot topic in research. The goal of this work was not to develop a new annotation method but enhance the performance with presenting multiple techniques of similarity computation between image features. Experiments [50-55] on the two different datasets Corel 5K and IAPR TC12 aim to bridge the gap between the low-level representations of images which concerned with color, shape and texture and the high-level one with the semantic meanings. It is clear that a simple combination of basic distance measures over some new feature extraction methods can effectively serve in the performance of image annotation methods.



Predicted Keywords	Sun, sky, mountain, sunset, buildings	Man, tree, building, sky, clouds
Human Annotations	Sunset, mountain, buildings	Humans, cars, building, tree, sky

Figure 5 Predicted Keywords versus Human Annotation for Corel5k in left and IAPR TC12 in right

4.3 Performance Evaluation

The performance of annotation task calculated by F-score value (5) which uses the value of precision (3) and recall (4). The proposed algorithm achieves F-score 0.45 for Corel5K dataset and 0.42 for IAPR TC12 dataset. Figure 5 shows an example of the annotation results in comparison to the human labeling and Table 1 presents a list of F-score values for different annotation models versus the proposed method.

$$\text{Precision} = \frac{\text{no.of correct annotated label}}{\text{Total annotated label}} \qquad (3)$$

$$\text{Recall} = \frac{\text{no.of correct annotated label}}{\text{Total label in test set}} \qquad (4)$$

$$\mathrm{F-score} = \frac{2*\text{Precision}*\text{Recall}}{(\text{precision}+\text{Recall})} \qquad (5)$$

Table 1 Annotation results of the proposed method versus the previous published algorithms on Corel 5k & IAPR TC12 Datasets [ZHANG et al.[5]]

Model	F-Score	
	Corel5K	IAPR TC12
Co-occurrence	0.02	-
Translation Model	0.05	-
CMRM	0.09	-
Max. Entropy	0.10	-
CRM	0.17	-
MBRM	0.22	0.23
GS	0.31	0.30
Tag-Prop	0.36	0.39
Proposed Method	0.45	0.42

5 Conclusion & Future Work

In the proposed framework, pyramid histogram of visual words description used in extracting image features to solve the automatic image annotation problem. Since the BOW image model for a feature, representation is the base for image annotation and classification. Weighting these features using weight least square method and obtaining accurate similar and dissimilar pairs of images is enhancing the power of assigning correct labels for the image that help in understanding its meaning. We tried in this algorithm to get used of PHOW advantages in formulating certain features. Also combining similarities computation between training image pairs with the keywords information gives high results for the annotation task in comparison to other existing methods. In future work, Try using ontologies which are the integration of images analysis improvement knowledge and images interpretation, instead of using contextual knowledge for semantic image annotation by building semantic hierarchies.

6 References

[1]Shalini, K., Kharkate, Nitin, J., Janwe, "A Novel Approach For Automatic Image Annotation Using Color Saliency", International Journal of Innovative Research in Computer and Communication Engineering, 2015.

[2]S., Hossein; H., Maryam. "A Novel Semantic Statistical Model for Automatic Image Annotation Using Ontology". Majlesi Journal of Multimedia Processing, 2015, 4.2.

[3] Siddiqui A, Mishra N, Verma JS. A Survey on Automatic Image Annotation and Retrieval. International Journal of Computer Applications. 2015.

[4] B., Riadh; M., Abir; A., Jalel. "Using a bag of words for automatic medical image annotation with a latent semantic". arXiv preprint arXiv:1306.0178, 2013.

[5] T., Dongping, "Support Vector Machine for Automatic Image Annotation", 2015.

[6] Datta, R., Joshi, D., Li, J., Wang, J.Z. "Image retrieval: Ideas, influences, and trends of the new age". ACM Computing Surveys, 2008.

[7] A. Makadia, V. Pavlovic, and S. Kumar, "A new baseline for image annotation," in Proc. Eur. Conf. Comput. Vis., 2008, 316–329.

[8] S. Yang, J. Bian and H. Zha, "Hybrid generative / discriminative learning for automatic image annotation", In Proceedings of UAI, 2010, 1-8.

[9] Y. Mori, H. Takahashi, R. Oka, "Image-to-word transformation based on dividing and vector quantizing images with words", Neural Networks in Boltzmann machine, 1999.

[10] P. Duygulu, K. Barnard and N. de Freitas, "Object recognition as machine translation: learning a lexicon for a fixed image vocabulary", Proceedings of 7th European Conference on Computer Vision, Copenhagen, Denmark, 2002, pp. 97-112.

[11] Jeon J, Lavrenko V, Manmatha R.," Automatic image annotation and retrieval using cross-media relevance models". Proc. of Int. ACM SIGIR Conf. On Research and Development in Information Retrieval, Toronto, Canada, 2003, 119-126.

[12] Lavrenko V, Manmatha R, Jeon J." A model for learning the semantics of pictures". Proceedings of the Seventeenth Annual Conference on Neural Information Processing Systems, 2003, 119-126.

[13] Putthividhya,D., Attias,H.T., Nagarajan,S.S.", Supervised topic model for automatic image annotation",IEEE International Conference on Acoustics Speech and Signal Processing , 2010.

[14] Feng S, ManmathaR ,Lavrenko V. "Multiple bernoulli relevance models for image and video annotation." Proc. of IEEE Int. Conf. on Computer Vision and Pattern Recognition (CVPR‟04), 2004.

[15] Tianxia Gong, Shimiao Li, Chew Lim Tan,"A Semantic Similarity Language Model to Improve Automatic image annotation",22nd International Conference on Tools with Artificial Intelligence, 2010, 197-203.

[16] S. Deerwester, S. T. Tumais, T. K. Landauer, G. W. Furnas, and R. A. Harshman, "Indexing by latent semantic analysis", J. Soc. Inform. Sci. 41, 1990, 391_407.

[17] T. Hofmann, "Probabilistic Latent Semantic Indexing," Proc.22ndInt'l Conf. Research and Development in Information Retrieval SIGIR'99, 1999.

[18] S. Zhang, J. Huang, H. Li, and D. N. Metaxas, "Automatic Image Annotation and Retrieval Using Group Sparsity", IEEE, 2012.

[19]Yang, C., Dong, M., Hua, J., "Region-based image annotation using asymmetrical support vector machine-based multiple-instance learning". In: Proceedings of the IEEE

International Conference on Computer Vision and Pattern Recognition, 2006.

[20] Carneiro, G., Chan, A.B., Moreno, P.J., Vasconcelos, N.,"Supervised learning of semantic classes for image annotation and retrieval". IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007.

[21] Blei, D.M., Jordan, M.I. "Modeling annotated data". In: Proc. ACM SIGIR, 2003, 127–134.

[22] Wang, L., Liu, L., Khan, L. "Automatic image annotation and retrieval using subspace clustering algorithm". In: ACM Int'l Workshop Multimedia Databases, 2004.

[23] Monay, F., Gatica-Perez, D. "On image auto-annotation with latent space models". In: ACM Int'l Conf. Multimedia, 2003.

[24] Metzler, D., Manmatha, R. "An inference network approach to image retrieval". In: Image and Video Retrieval, Springer, 2005.

[25] Frome, A., Singer, Y., Sha, F., Malik., J." Learning globally-consistent local distance functions for shape-based image retrieval and classification". In: Proceedings of the IEEE International Conference on Computer Vision, Rio de Janeiro, Brazil, 2007.

[26] Jin, R., Chai, J.Y., Si, L. "Effective automatic image annotation via a coherent language model and active learning". In: ACM Multimedia Conference, 2004, 892–899.

[27] Yavlinsky, A., Schofield, E., Ruger, S., "Automated Image Annotation Using Global Features and Robust Nonparametric Density Estimation". In: CIVR. 2005.

[28] Lei Ye, Philip Ogunbona and Jianqiang Wang" Image Content Annotation Based on Visual Features", Proceedings of the Eighth IEEE International Symposium on Multimedia (ISM'06), 2006.

[29] D Dongjian He & Yu Zheng, Shirui Pan, Jinglei Tang, "Ensemble Of Multiple Descriptors For Automatic Image Annotation" -2010, 3rd International Congress on Image and Signal Processing

[30] A. Argyriou, T. Evgeniou, and M. Pontil, "Multi-task feature learning," in Proc. Annual Conf. Neural Inf. Process. Syst., 2007.

[31] H. Zhang, A. C. Berg, M. Maire, and J. Malik. "SVM-KNN: Discriminative nearest neighbor classification for visual category
Recognition". In CVPR, pages 2126–2136, 2006.

[32] M. Guillaumin, T. Mensink, J. Verbeek, and C. Schmid, "TagProp: Discriminative metric learning in nearest neighbor models for image autoannotation," in Proc. Int. Conf. Comput. Vis., 2009, 309–316.

[33] Lazebnik Svetlana, Schmid Cordelia and Ponce Jean, "Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories," in Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference, Illinois, 2006.

[34] Bosch Anna, Zisserman Andrew and Munoz Xavier, "Representing shape with a spatial pyramid kernel," in CIVR, Amsterdam, 2007.

[35] T. Hertz, A. Bar-hillel, and D. Weinshall, "Learning distance functions for image retrieval," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2004, II-570–II-577.

[36] J. Liu, M. Li, Q. Liu, H. Lu, and S. Ma, "Image annotation via graph learning," Pattern Recognit., vol. 42, no. 2, 2009, 218–228

[37] W. J. Fu, "Penalized regressions: The bridge versus the lasso," J. Comput. Graph. Stat., vol. 7, no. 3, 1998, 397–416.

[38] R. Tibshirani, "Regression shrinkage and selection via the LASSO," J. Royal Stat. Soc., vol. 58, no.1, 1994, 267–288.

[39] Hare, J.S., Lewisa, P.H., Enserb, P.G.B., Sandomb, C.J."Mind the gap: Another look at the problem of the semantic gap in image retrieval". Multimedia Content, Analysis, Management and Retrieval, 2006.

[40] Gao, Y., Fan, J. "Incorporating concept ontology to enable probabilistic concept reasoning for multi-level image annotation". In: 8th ACM International Workshop on Multimedia Information Retrieval, 2006, 79–88.

[41] Li, J.,Wang, J.,"Automatic linguistic indexing of pictures by a statistical modeling approach". IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003.

[42] Yunhee Shin, Youngrae Kim, Eun Yi Kim," Automatic textile image annotation by predicting emotional concepts from visual features". Image and Vision Computing, 2010.

[43] O. Yakhnenko and V. Honavar, "Annotating images and image objects using a hierarchical dirichlet process model," in Proc. Int. Conf. Mobile Data Manage., 2008, 1–7.

[44] S. Gao, L.-T. Chia, and I. W.-H. Tsang, "Multi-layer group sparse coding for concurrent image classification and annotation," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2011, 2809–2816.

[45] J. Li and J. Wang, "Real-time computerized annotation of pictures," IEEE Trans. Pattern Anal. Mach. Intell., vol. 30, no. 6, 2008.

[46] D. Putthividhya, H. Attias, and S. Nagarajan, "Topic-regression multimodal latent dirichlet allocation for image annotation," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2010, 3408–3415.

[47] K. Barnard, M. Johnson, "Word sense disambiguation with pictures". Artificial Intelligence, 2005, 13–30.

[48] K. Barnard, P. Duygulu, D. Forsyth, N. D. Freitas, D. M. Blei, J. Kandola, T. Hofmann, T. Poggio, and J. Shawe-Taylor, "Matching words and pictures," J. Mach. Learn. Res., vol. 3, 2003, 1107–1135.

[49] M. Grubinger, P. D. Clough, H. Muller, and T. Deselaers, "The IAPR TC-12 benchmark—A new evaluation resource for visual information systems," in Proc. Int. Workshop OntoImage, 2006, 13–23.

[50] K. van de Sande, T. Gevers, and C. Snoek, "Evaluating color descriptors for object and scene recognition," IEEE Trans. Pattern Anal.Mach. Intell., vol. 32, no. 9, 2010, 1582–1596.

Shereen A. Hussein field of interests includes Computer Vision, pattern recognition, AI researches, and medical imaging.

Howida Youssry Abd El Naby field of interests includes pattern recognition, AI researches, and medical imaging.

Dr. A. Youssif field of interests includes pattern recognition, AI researches, and medical imaging. She published more than 70 papers in different fields.