International Journal of Multimedia and Ubiquitous Engineering
Vol.9, No.3 (2014), pp.117-128
http://dx.doi.org/10.14257/ijmue.2014.9.3.12

A Survey of Refining Image Annotation Techniques

Dong ping Tian[1, 2]

[1] *Institute of Computer Software, Baoji University of Arts and Sciences, Baoji, Shaanxi, 721007, China*
[2] *Institute of Computational Information Science, Baoji University of Arts and Sciences, Baoji, Shaanxi, 721007, China*

{tdp211}@163.com

Abstract

Automatic image annotation has been an active research topic in recent years due to its potential impact on both image understanding and semantic based image retrieval. However, the results of the state-of-the-art image annotation methods are still far from satisfaction due to the existence of semantic gap. Thus refining image annotation (RIA) has become one of the core research topics in computer vision and multimedia areas, whose purpose is to reserve the highly correlated annotations whereas remove the irrelevant or weakly relevant annotations by fully exploring the correlations of annotation keywords. RIA, to some extent, can effectively mitigate the semantic gap between low-level visual features and high-level semantic concepts. So in this paper, we focus on the latest development in image retrieval and provide a comprehensive survey on refining image annotation techniques. In particular, we analyze the key aspects of various RIA methods, including their original intentions and annotation models. Finally, we draw some important conclusions and highlight the potential research directions for the future.

Keywords: *Refining image annotation, Graphical model, Random field model, Manifold ranking, Semi-supervised learning*

1. Introduction

With the rapid explosion of images available from various multimedia devices, effective technologies for organizing, searching and browsing these images are urgently required by common users. Ideally, those images should be indexed by semantic descriptions so that traditional information retrieval techniques may be adopted for precise image search. However, as it is impossible to manually annotate so many images, automatic image annotation (AIA) might be a promising solution. The goal of AIA is to automatically assign some keywords to an image that can well describe the content in it. Figure 1 illustrates a typical system of automatic image annotation. Given an image collection and a dictionary of keywords, a computer assigns keywords to each image automatically. In recent years, a significant amount of researches have focused on automatic image annotation. Early work by Duygulu *et al.* [1] propose the translation model (TM) to treat AIA as a process of translation from a set of blob tokens, obtained by clustering image regions, to a set of keywords. Jeon *et al.* [2] put forward cross-media relevance model (CMRM) to annotate image, assuming that

ISSN: 1975-0080 IJMUE

the blobs and words are mutually independent given a specific image. Subsequently, CMRM is improved through continuous-space relevance model (CRM) [3] and multiple-Bernoulli relevance model (MB- RM) [4]. Recently, the dual cross-media relevance model (DCMRM) [5] which calculates the expectation over words in a pre-defined lexicon is also proposed. In addition, Carneiro *et al.* [6] come up with the supervised multi-class labeling (SML), which utilizes optimal principle of minimum probability of error and treats annotation as a multi-class classification problem. As latent aspect models, probabilistic latent semantic analysis (PLSA) [7], latent semantic analysis (LSA) [8] and layered pictorial structures (LPS) [9] have also been successfully applied in automatic image annotation. In [10], Fergus *et al.* extend the PLSA model by adding spatial information based on the visual words. Subsequently, Monay and Gatica-Perez have proposed the classical PLSA-WORDS and PLSA-FEATURES models [11].

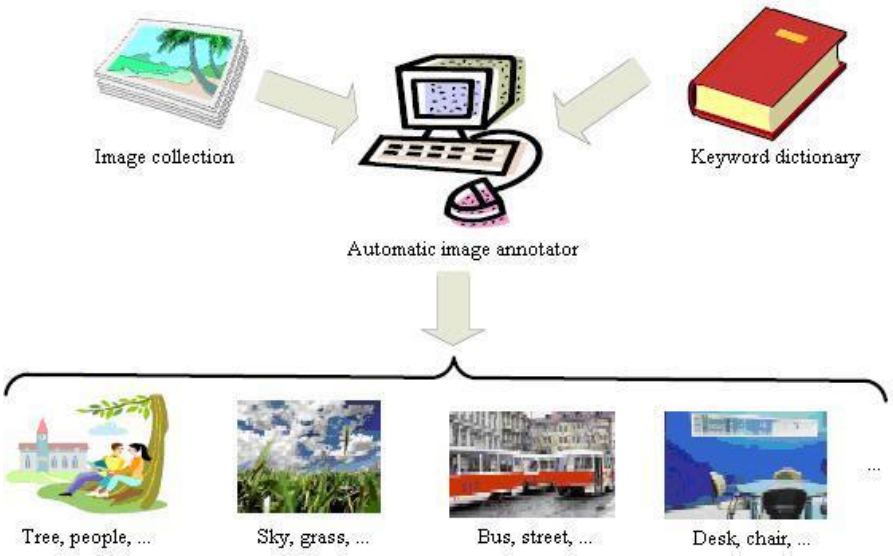

Figure 1. An illustration of a typical system for automatic image annotation

All of the annotation models aforementioned, to some extent, can achieve better annotating performance than that of the early manual annotations. However, their results are still far from satisfaction due to the existence of semantic gap as well as the little consideration of relations among annotation keywords. Confronted with these problems, refining image annotation (RIA) has been proposed, which aims to reserve the highly correlated annotations and remove the non-correlated or weakly-correlated annotations based on the information of candidate annotations generated by some existing annotation methods. As a pioneer work, Jin *et al.* [12] utilize a generic knowledge-based word-net to refine image annotation by pruning the irrelevant annotations. The basic assumption is that highly correlated annotations should be reserved and non-correlated annotations should be removed. In their work, however, only global textual information is used, and the refinement process is independent of the target image, which means that different images with the same candidate annotations would obtain

the same refinement annotation results. So in this paper, we review the various RIA methods, including their original intentions and annotation models adopted.

The rest of the paper is organized as follows. Section 2 elaborates various refining image annotations, including their original intentions and annotation models adopted. In Section 3, we draw some important conclusions and highlight the potential research directions for the future.

2. Refining Image Annotation Techniques

Since the pioneer work of refining image annotation done by Jin *et al.* [12], many approach- es have emerged up subsequently. Most of them can be roughly classified into three categori- es, *i.e.*, graphical model based RIA, random field model based RIA, manifold ranking based RIA and other hybrid refining image annotation approaches. In the following, we will elaborate some representative RIA approaches belonged to each category as well as their pros and cons.

2.1. Graphical model based RIA

Graphical model (GM) is a marriage between probability theory and graph theory [13], which provides a natural tool for dealing with two problems that occur throughout applied mathematics and engineering, *i.e.*, uncertainty and complexity and in particular, GM is playing an increasingly important role in the design and analysis of machine learning algorithms. In the most recent years, graphical model has been attracting significant research attention in multimedia and computer vision area, especially in refining image annotation. As the represe- ntative work, Pan *et al.* [14] propose a graph-based approach for refining image annotation (GCap). To be specific, they first represent an image as a set of regions, each of which is described by a visual feature vector. A graph is constructed on the whole training data. Then they define three types of node in this graph, viz., image node representing an image, region node representing an image region and word node representing a textual keyword. The links between nodes represent the relationship between different units (image, region and words). Finally, the problem is to capture the correlation between image features and caption terms according to their known association so as to implement image annotation. This method has the advantages of being domain independent and simple parameter tuning, which are strong points shared by general graph model method. However, region-based visual features are sampled from continuous sources and annotations are sampled from discrete sources of finite alphabet, so it is difficult to weight these two types of nodes from different modalities in one graph. The basic flowchart of the GCap is illustrated in Figure 2, in which three sample images, their captions and their regions are depicted step by step respectively.

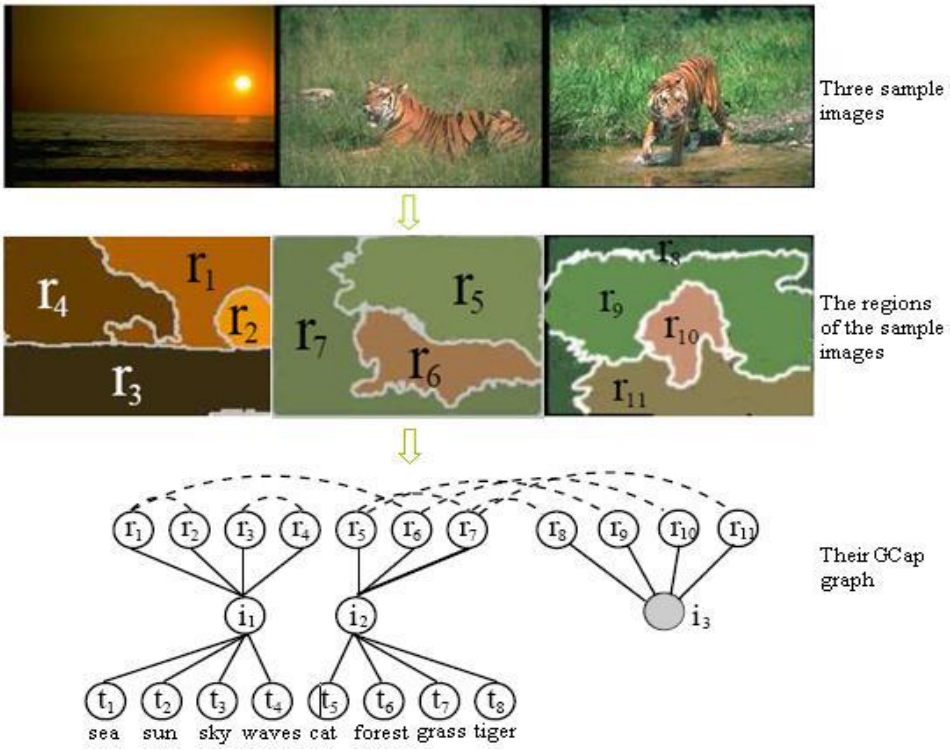

Figure 2. An illustration of the GCap model for automatic image annotation

In addition, Wang *et al.* [15], first of all, get the candidate annotations from both web and non-web images, and then adopt an algorithm based on random walk with restarts to re-rank the candidate annotations in which the corpus information as well as confidence scores of original annotations is leveraged. The experimental results demonstrate its effectiveness. However, it is still implicitly based on the assumption *majority should win* and the refinement process is still independent of the original query image. Subsequently, they [16] propose another refining image annotation method, in which CMRM is first used to obtain the candidate annotations, and then they formulate the annotation refinement process as a Markov process and define the candidate annotations as the states of a Markov chain. Recently, Jin *et al.* [17] put forward knowledge-based image annotation refinement (KBIAR) approach. They reformulate KBIAR into weighted max-cut problem satisfied with the 0.87856 ratio with the optimal solution as well as polynomial running time. In addition, Liu *et al.* [18] propose a NSC-based method to calculate image similarities on visual features and propagate annotations from training images to their similar test images. Exactly speaking, they develop a novel method to estimate the word correlation based on the improved nearest spanning chains, which can extract more informative and reasonable relations among keywords. After obtaining the enhanced word correlation, a word-based graph is constructed and used to refine the candidate annotations for an untagged image. More recently, Liu *et al.* [19] present a graph-based appro- ach to automatically refine image annotation. Similar to other refining methods, a set of candidate annotations for an unseen image is first extracted by some existing image annota- tion methods. Then, each candidate annotation is converted to vertex of a graph and the semantic similarity between two candidate annotations is used as edge weight. Finally, a rank- two relaxation heuristics approximation algorithm is used to solve the weighted MAX-CUT problem and obtain the refined annotations by the decision. Alternatively, Tian *et*

al. [20] put forward a two-stage refining image annotation method. They first exploit a probabilistic latent semantic analysis (PLSA) model with asymmetric modalities to accomplish the initial seman- tic annotation, and then implement random walk process over the constructed label similarity graph to refine the candidate annotations generated by the PLSA. Followed by they propose a very similar two-stage refining image annotation method [21], in which the refining annotation has been viewed as a graph partitioning problem and the max-bisection rather than the random walk over the label similarity graph is implemented based on the rank-two relaxa- tion heuristics in the secondary refining stage to further mine the correlation among the candidate annotations. Figure 3 illustrates the generic framework for refining image annotation proposed in these two literatures. Alternatively, Table 1 simply summarizes the graphical model based refining image annotation methods mentioned above.

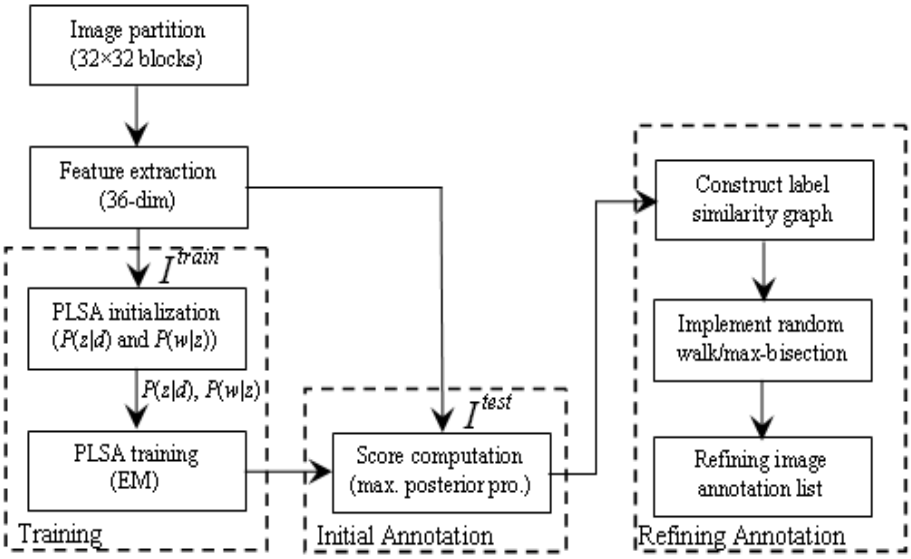

Figure 3. The generic framework for refining image annotation

As reviewed above, most of the refining image annotation methods can get relatively ideal annotating results compared to the traditional approaches. The reason lies in two-fold. First, the semantic relevance of annotating concepts is incorporated into the refining image annotation process. Second, the graphical model based refining image annotation, in general, comprises initial annotation and refining annotation stages, which can further prune the noisy annotations from the candidate ones effectively.

Table 1. Summary of different graphical model based refining annotation methods

Sources	Methods adopted	Image datasets applied
Pan et al. [14]	*k*-nearest neighbors, random walk with restarts	Corel dataset
Wang et al. [15]	Cross media relevance model, random walk with restarts	Corel dataset, web images of photo forum sites
Wang et al. [16]	Cross media relevance model, query bias- ed Markov chain	Corel dataset
Jin et al. [17]	Dempster-Shafer evidence theory, rando- mized approximation weighted max-cut	Corel dataset
Liu et al. [18]	Multiple-Bernoulli relevance model, near- est spanning chain	Corel dataset

Liu et al. [19]	Rank-two relaxation heuristics algorithm	Web images from PhotoSIG
Tian et al. [20]	Probabilistic latent semantic analysis, random walk	Corel, Mirflickr dataset
Tian et al. [21]	Probabilistic latent semantic analysis, max-bisection	Corel dataset

2.2. Random field model based RIA

Random field (RF) is a generalization of a stochastic process such that the underlying parameter need no longer be a simple real or integer valued time, but can instead take values that are multidimensional vectors, or points on some manifold. Random field theory is a recent body of mathematics defining theoretical results for smooth statistical maps. The theory has been versatile in dealing with many of the threshold problems that we encounter in functional imaging. Over the years, RF has been widely utilized in multimedia and computer vision field, especially the Markov random field and conditional random field. In the following subsections we will elaborate their applications in refining image annotation.

2.2.1. Markov random field based refining image annotation: Markov random field (MRF) is a probabilistic model which combines a priori knowledge given by some observations and knowledge given by the interaction with neighbors. MRF is also referred to as a Gibbs random field in case the probability distribution is positive accord- ing to the Hammersley-Clifford theorem, so it then can be represented by a Gibbs measure. MRF is appealing in automatic image annotation for the following reasons [22]. First, one can systematically develop algorithms based on sound principles rather than on some ad-hoc heuristics for a variety of problems. Second, it makes it easier to derive quantitative performance measures for characterizing how well the image analysis algorithms work. Third, MRF models can be used to incorporate various prior contextual information or constraints in a quantitative way, and last but not the least, the MRF-based algorithms tend to be local, and tend themselves to parallel hardware implementation in a natural way. Escalante *et al.* [23] propose an approach for refining image annotation based on the fact that accuracy of current image annotation methods is low if the most confident label is considered only. Instead, accuracy can be improved if the correct labels within the set of the top-*k* candidate labels are taken into account. They capture spatial dependencies between connected regions through MRF model with iterated conditional modes and simulated annealing as optimization strategies. In addition, semantic information between labels is also incorporated using word co-occurrences to improve the performance of annotation systems. And the experimental results of the proposed method together with a *k*-nearest neighbor classifier as annotation method show the important error reductions. Hernandez-Gracidas *et al.* [24] come up with an approach based on Markov random fields to represent the information about the spatial relations among the regions in an image, so the probability of occurrence of a certain spatial relation between each pair of labels could be used to obtain the most probable label for each region, *i.e.*, the most probable configuration of labels for the whole image. The spatial relations considered in this work are shown in Table 2 and they are divided into three groups: topological relations, horizontal relations and vertical relations. Meanwhile, the spatial information is fused with "expert" knowledge to represent the information coming from the neighbors. The experiment conducted on Corel dataset shows that the proposed approach is feasible to apply spatial relations and MRF to improve automatic image annotation systems.

Table 2. Spatial relations among the image regions employed in [21]

Relation types		No.	Directed	Undirected
Topological relations		1	/	Adjacent
		2	/	Disjoint
Order relations	Horizontal relations	3	/	Beside(left or right)
		4	/	Horizontally aligned
	Vertical relations	5	Above	/
		6	Below	/
		7	/	Vertically aligned

More recently, Llorente *et al.* [25] propose a direct image retrieval framework based on Markov random fields (MRFs) that exploits the semantic context dependencies of the image. The main novelty lies in the use of different kernels in the non-parametric density estimation together with the utilization of configurations that explore semantic relationships among concepts at the same time as low-level features, instead of just focusing on correlation between image features like in previous formulations. The following Figure 4 shows a graph representing the dependencies explored in [25]. The left side of the image illustrates the clique configura- tions considered in the research which contemplates cliques of up to third order. A 2-clique (*r*-*w*) consisting of a query node *w* and a feature vector *r*, followed by a 2-clique (*w*-*w'*) representing the dependencies between words *w* and *w'*, and finally a 3-clique (*r*-*w*-*w'*) capturing the relation between a feature vector *r* and two word nodes *w* and *w'*.

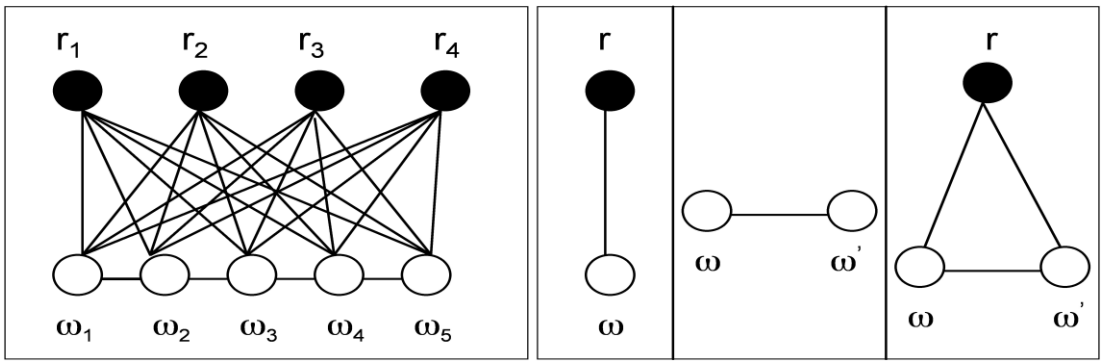

Figure. 4 Markov random fields graph model employed in Ref. [25]

2.2.2. Conditional random field based refining image annotation: Conditional random field (CRF) [26] is a probabilistic framework for labeling and segmen- ting structured data, such as sequences, trees and lattices. The underlying idea is that of defin- ing a conditional probability distribution over label sequences given a particular observation sequence, rather than a joint distribution over both label and observation sequences. CRF is a type of discriminative undirected probabilistic graphical model. The primary advantage of CRF over hidden Markov models (HMM) is its conditional nature, resulting in the relaxation of the independence assumptions required by HMM in order to ensure tractable inference. Additionally, CRF avoids the label bias problem, a weakness exhibited by maximum entropy Markov models (MEMM) and other conditional Markov models based on directed graphical models. Due to its good property, CRF has been extensively applied in multimedia processing in recent years. As a representative work of CRF for refining image annotation, Wang *et al.* [27] present a method by incorporating semantic relations between annotation words using a conditional random field model. Similar to other refining annotation methods, a candidate set of annotation words with confidence scores by the relevance vector machine is first achieved.

Followed by the semantic relationship between candidate annotations are modeled by a conditional random field, where each vertex indicates the final decision on a candidate annotation word. Finally, the refined annotation can be obtained by inferring the most likely states of these vertexes. Li *et al.* [28] formulate the image annotation problem as a joint classification task based on two dimensional conditional random fields together with semi-supervised learning, in which the 2D CRF is used to effectively capture the spatial dependency between the neighboring labels while the semi-supervised learning technique is employed to exploit the unlabeled data to improve the joint classification performance. In [29], an integration of CRF and SVM is utilized for automatic image region annotation, whose main goal is to exploit the spatial context constraints based on the conditional random field for boosting the image region annotation performance. More recently, Huang *et al.* [30] present a hierarchical two-stage CRF model to deal with the problem of labeling images of street scenes, which combines the ideas used in both parametric and nonparametric image labeling methods. All in all, many existing image annotation approaches based on the CRF model are comparable to, and in many cases superior to, those previous traditional methods.

2.3. Manifold Ranking based RIA

The manifold ranking algorithm [31] is initially proposed to rank the data points or to predict the labels of unlabeled data points along their underlying manifold by analyzing their relationship in Euclidean space. In recent years, it has been successfully applied in image annotation and retrieval community [32, 33, 34, 35]. As a representative work of employing manifold ranking for RIA, Liu *et al.* [32] propose a novel automatic image annotation method based on manifold ranking learning, in which the visual and textual information are well integrated. On the one hand, they employ nearest spanning chain to generate an adaptive similarity graph. On the other hand, the word-to-word correlations obtained from word-net and the pairwise co-occurrence are taken into consideration to expand the annotations and prune irrelevant annotations for each image, which make the manifold ranking efficient for refining image annotation. Figure 5 illustrates a toy example of the NSC, in which the left figure presents the data distribution, in which the numbers outside of bracket and in the bracket represent the index and coordinate for each point respectively. The right one gives nine examples of NSC denoted by the indexes of data. In addition, a novel semi-supervised multi-instance multi- label learning algorithm is put forward for the task of refining image annotation [34], in which the manifold ranking algorithm is applied to propagate the corresponding labels from the positive bags to unlabeled bags directly. Experiments on the Corel dataset validate its effectiveness and efficiency.

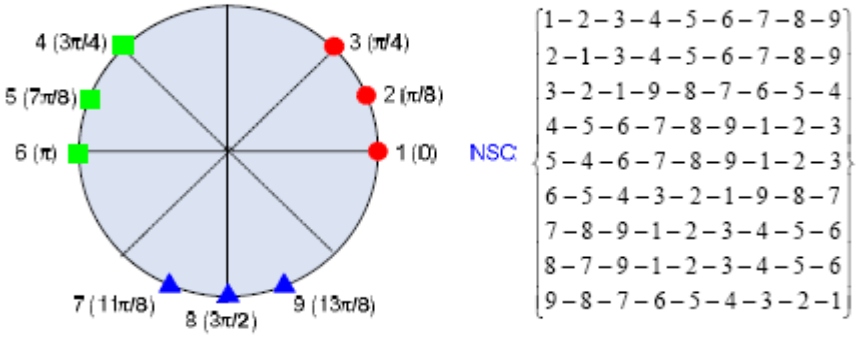

Figure. 5 Toy example of NSC in Ref. [32]

2.4. Other Refining Approaches

In addition to the aforementioned refining image annotation approaches, there are other types of RIA, which can also capture better annotation results compared to the traditional methods. Zhu *et al.* [36] develop a novel approach to automatically refine the initial annotation of images. In their method, the candidate annotations are first obtained by a step-up model- based algorithm using perceptual visual characteristic. Then, a refine algorithm, fast random walk with restart is used to re-rank the candidate annotations and the top ones are reserved as the final annotations. Recently, Zhu *et al.* [37] formulate the tag refinement problem as a decomposition of the user-provided tag matrix into a low-rank refined matrix and a sparse error matrix, targeting the optimality measured by four aspects, *i.e.*, low-rank, content consist- ency, tag correlation and error sparsity. All these components constitute a constrained yet convex optimization problem and an efficient convergence provable iterative procedure is proposed for the optimization based on accelerated proximal gradient method for refining annotation. In addition, a similar work is proposed by Jia *et al.* in [38], in which the textual similarities of tags and visual similarities of images are fused in a multi-graph reinforcement framework. In [39], Xu *et al.* propose to do tag refinement from topic modeling point of view. A new graphical model named as regularized latent Dirichlet allocation (rLDA) is presented to jointly model the tag similarity and tag relevance.

Alternatively, in the scenario of image annotation, an image is usually described by multi-ple semantic labels and these labels are often highly related to respective regions rather than the entire image (see Figure 6). As a result, image annotation is modeled as a multi-label multi- instance learning problem [40]. In order to utilize the unlabeled data to achieve more promis- ing performance, Feng *et al.* [41] recently present a transductive multi-instance multi-label (TMIML) learning algorithm for refining image annotation, which aims at taking full advan- tage of both labeled and unlabeled data to address the annotation problem. Here we only give a brief introduction to multi-instance learning (MIL) and multi-label learning in the applica- tion of refining image annotation. For more details of them please refer to [42, 43].

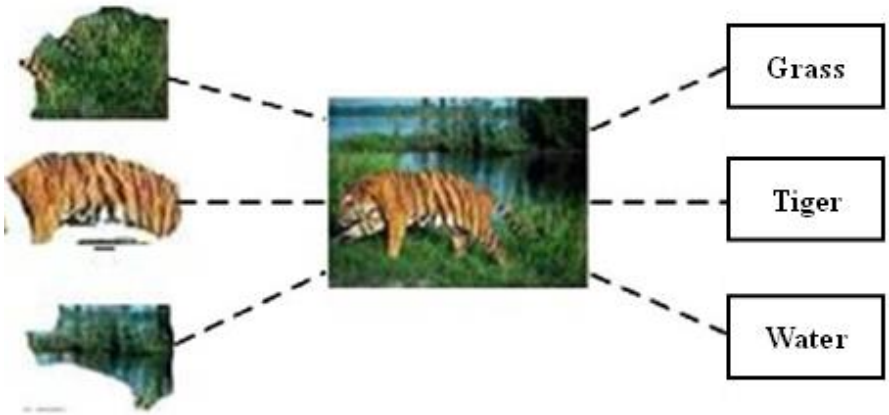

Figure 6. The example of the multi-instance multi-label learning framework in image annotation

3. Discussion and Conclusions

We have made a comprehensive review on the state-of-the-art refining image annotation techniques in literature. RIA, here, is summarized from three main aspects, *i.e.*, graphical

model, random field model and manifold ranking. All of them can shoot for better annotating performance from different point of view, such as the integration of visual similarities of images and textual similarities of tags, the two-stage annotation including the initial and refining image annotations, *etc.* However, there are still several major issues in RIA research to be explored.

The first issue is how to extract ideal image features to reflect the inherent content of images as complete as possible. Currently, all existing features have limitations of describing images and none of existing features is powerful enough to represent the large variety of images in nature. Common practice is to combine several types of features to represent as many images as possible. However, the processing and analyzing of high dimensional image features is a very complex issue.

The second issue is how to build an effective refining image annotation model. Most existing RIA models are learned from both low level visual features and high level semantic information or from the hybrid two-stage annotation methods. However, due to the labeled images are hard to obtain enough compared to the unlabeled ones, which are required to guarantee the feasibility of the annotating model. So the semi-supervised learning can be employed to improve the refining image annotation accuracy under the conditions that there are only a few labeled but a large amount of unlabeled images to implement automatic annotation.

The third issue is the lack of commonly acceptable image database for RIA training and evaluation. All RIA methods require a certain number of labeled images for training the model. At this moment, different RIA methods use different image datasets for training and testing, thus making it difficult to evaluate their performance. Therefore, some standard image databases are expected to be created for researches in the future shared by people.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (No.610350- 03, No.61072085, No.60933004, No.60903141) and the National Program on Key Basic Res- earch Project (973 Program) (No.2013CB329502).

References

[1] P. Duygulu, K. Barnard, N. de Freitas and D. Forsyth, "Object recognition as machine translation: learning a lexicon for a fixed image vocabulary", In Proc. ECCV, **(2002)**, pp. 97-112.
[2] L. Jeon, V. Lavrenko and R. Manmantha, "Automatic image annotation and retrieval using cross- media relevance models", In Proc. SIGIR, **(2003)**, pp. 119-126.
[3] R. Manmatha, V. Lavrenko and J. Jeon, "A model for learning the semantics of pictures", In Proc. NIPS, **(2003)**, pp. 553-560.
[4] S. Feng, R. Manmatha and V. Lavrenko, "Multiple Bernoulli relevance models for image and video annotation", In Proc. ICPR, **(2004)**, pp. 1002-1009.
[5] J. Liu, B. Wang, M. Li, *et al.*, "Dual cross-media relevance model for image annotation", In Proc. ACM MM, **(2007)**, pp. 605-614.
[6] G. Carneiro, A. Chan, P. Moreno and N. Vasconcelos, "Supervised learning of semantic classes for image annotation and retrieval", IEEE Trans. PAMI, vol. 29, no. 3, **(2007)**, pp. 394-410.
[7] T. Hofmann, "Unsupervised learning by probabilistic latent semantic analysis, Machine Learning", vol. 42, no. 1, **(2001)**, pp. 177-196.
[8] Y. Chen, F. Tsai and K. Chan, "Machine learning techniques for business blog search and mining", Expert Systems with Applications, vol. 35, no. 3, **(2008)**, pp. 581-590.
[9] M. Kumar, P. Torr and A. Zisserman, "OBJ CUT", In Proc. CVPR, **(2005)**, pp. 18-25.
[10] R. Fergus, F. Li, P. Perona and A. Zisserman, "Learning object categories from Google's image search", In Proc. ICCV, **(2005)**, pp. 1816-1823.

[11] F. Monay and D. Gatica-Perez, "Modeling semantic aspects for cross-media image indexing", IEEE Trans. on PAMI, vol. 29, no. 10, **(2007)**, pp. 1802-1817.
[12] Y. Jin, L. Khan, L. Wang and M. Awad, "Image annotations by combining multiple evidence and word-net", In Proc. ACM MM, **(2005)**, pp. 706-715.
[13] M. Jordan, "Graphical Models", Statistical Science, vol. 19, **(2004)**, pp. 140-155.
[14] J. Pan, H. Yang, C. Faloutsos and P. Duygulu, "GCap: Graph-based automatic image captioning", In Proc. Workshop on CVPR, **(2004)**, pp. 146-149.
[15] C. Wang, F. Jing, L. Zhang and H. Zhang, "Image annotation refinement using random walk with restarts", In Proc. ACM MM, **(2006)**, pp. 647-650.
[16] C. Wang, F. Jing, L. Zhang and H. Zhang, "Content-based image annotation refinement", In Proc. CVPR, **(2007)**, pp.1-8.
[17] Y. Jin, L. Khan and B. Prabhakaran, "Knowledge based image annotation refinement", Journal of Signal Processing Systems, vol. 58, **(2010)**, pp. 387-406.
[18] J. Liu, M. Li, Q. Liu, H. Lu and S. Ma, "Image annotation refinement using NSC-based word correlation", In Proc. ICME, **(2007)**, pp. 799-802.
[19] Z. Liu, "A novel graph-based image annotation refinement algorithm", In Proc. FSKD, **(2009)**, pp. 330-334.
[20] D. Tian, X. Zhao and Z. Shi, "Refining image annotation by integrating PLSA with random walk model", In Proc. MMM, **(2013)**, Part I, LNCS 7732, pp. 13-23.
[21] D. Tian, W. Zhang, X. Zhao and Z. Shi, "Employing PLSA model and max-bisection for refining image annotation", In Proc. ICIP, **(2013)**, pp. 3996-4000.
[22] S. Z. Li, "Markov random field models in computer vision", In Proc. ECCV, **(1994)**, pp. 361-370.
[23] H. Escalante, M. Montes and L. Sucar, "Word co-occurrence and Markov random fields for improving automatic image annotation", In Proc. BMVC, **(2007)**.
[24] C. Hernandez-Gracidas and L. Sucar, "Markov random fields and spatial information to improve automatic image annotation", In Proc. PSIVT, **(2007)**, pp. 879-892.
[25] A. Llorente, R. Manmatha and S. Rüger, "Image retrieval using Markov random fields and global image features", In Proc. CIVR, **(2010)**, pp. 243-250.
[26] J. Lafferty, A. McCallum and F. Pereira, "Conditional random fields: probabilistic models for segmenting and labeling sequence data", In Proc. ICML, **(2001)**, pp. 282-289.
[27] Y. Wang and S. Gong, "Refining image annotation using contextual relations between words", In Proc. CIVR, **(2007)**, pp. 425-432.
[28] W. Li and M. Sun, "Semi-supervised learning for image annotation based on conditional random fields", In Proc. CIVR, **(2006)**, pp. 463-472.
[29] J. Yuan, J. Li and B. Zhang, "Exploiting spatial context constraints for automatic image region annotation", In Proc. ACM MM, **(2007)**, pp. 595-604.
[30] Q. Huang, M. Han, B. Wu and S. Ioffe, "A hierarchical conditional random field model for labeling and segmenting images of street scenes", In Proc. CVPR, **(2011)**, pp. 1953-1960.
[31] D. Zhou, J. Weston, A. Gretton, O. Bousquet and B. Schölkopf, "Ranking on data manifolds", In Proc. NIPS, **(2003)**, pp. 169-176.
[32] J. Liu, M. Li, W. Ma, Q. Liu and H. Lu, "An adaptive graph model for automatic image annotation", In Proc. MIR, **(2006)**, pp. 61-69.
[33] J. He, M. Li, H. Zhang, H. Tong and C. Zhang, "Generalized manifold-ranking-based image retrieval", IEEE Transactions on Image Processing, vol. 15, no. 10, **(2006)**, pp. 3170-3177.
[34] S. Feng, D. Xu, C. Lang and B. Li, "Automatic image annotation using semi-supervised multi- instance multi-label learning algorithm", Chinese Journal of Electronics, vol. 17, no. 4, **(2008)**, pp. 602- 606.
[35] B. Xu, J. Bu, C. Chen, D. Cai, X. He, W. Liu and J. Luo, "Efficient manifold ranking for image retrieval", In Proc. SIGIR, **(2011)**, pp. 525-534.
[36] S. Zhu and Y. Liu, "Image annotation refinement using semantic similarity correlation", In Proc. ICPR, **(2008)**, pp. 1-4.
[37] G. Zhu, S. Yan and Y. Ma, "Image tag refinement towards low-rank, content-tag prior and error sparsity", In Proc. ACM MM, **(2010)**, pp. 461-470.
[38] J. Jia, N. Yu, X. Rui and M. Li, "Multi-graph similarity reinforcement for image annotation refinement", In Proc. ICIP, **(2008)**, pp. 993-996.
[39] H. Xu, J. Wang, X. Hua and S. Li, "Tag refinement by regularized LDA", In Proc. ACM MM, **(2009)**, pp. 573-576.
[40] Z. Zhou and M. Zhang, "Multi-instance multi-label learning with application to scene classifica- tion", In Proc. NIPS, **(2006)**, pp. 1609-1616.
[41] S. Feng and D. Xu, "Transductive multi-instance multi-label learning algorithm with application to automatic image annotation", Expert Systems with Applications, vol. 37, no. 1, **(2010)**, pp. 661 -670.

[42] C. Yang, M. Dong and F. Fotouhi, "Region based image annotation through multiple-instance learning", In Proc. ACM MM, **(2005)**, pp. 435-438.
[43] J. Tang and P. Lewis, "A study of quality issues for image auto-annotation with the Corel dataset", IEEE Trans. CSVT, vol. 17, no. 3, **(2007)**, pp. 384-389.

