A Review of Methods for The Image Automatic Annotation

Myasar Mundher Adnan¹,², Mohd Shafry Mohd Rahim¹,³, Mohammed Hasan ali¹, Karrar Al-Jawaheri², Karrar Neamah¹

¹School of Computing, Faculty of Engineering, University Technology of Malaysia, Johar Bahru, Malaysia
²Islamic University, Najaf, Iraq
³UTM-IRDA Digital Media Institute of Human-Centred Engineering, University Technology of Malaysia, Johar, Bahru, Malaysia
⁴Computer Techniques Engineering Department, Faculty of Information Technology, Imam Ja’afar Al-sadiq University, Najaf, Iraq

Email: maiserlove06@gmail.com

Abstract. Nowadays, image annotation has attracted extensive attention due to the explosive growth of image data. Large amount of researches on AIA have been proposed, mainly including classification-based methods and probabilistic modeling methods. In this paper, a detailed study on state-of-the-art of image annotation was presented devoted to a detailed study of image annotation methods. Differences between manual, semi-automatic and automatic annotation were completely distinguished. The criteria for evaluating annotation systems are also presented in this study. In conclusion, a synthesis of methods of automatic image annotation were shown by presenting the pros and cons of each. This synthesis allowed us to examine our choice for automatic image annotation and the importance of integrating user feedback and a semantic. Finally, we participated in our perspective on the issues and challenges in AIA as well as research tendency in the future.

1. Introduction

Nowadays, image annotation has attracted extensive attention due to the explosive growth of image data. Large amount of researches on IA have been proposed, mainly including classification-based methods and probabilistic modelling methods. Image annotation allows processes of indexing and searching large collections of images in an easy and fast way. In this paper, a detailed study of image annotation methods were demonstrated. There are three types of annotation including manual annotation, semi-automatic annotation and automatic annotation. The method of automatic annotation is our main focus for this study. Various methods have recently been proposed on AIA systems, giving rise to the development of several AIA algorithms. In this work, the attention was given to five techniques based on AIA; these techniques are Relevance Return based on AIA, Inter-domain semi-supervised learning based on AIA, Metadata Use based on AIA, Semantic Hierarchies based on AIA and Hollow coding models based on AIA. We concluded this paper with a synthesis of the methods of automatic image annotation by presenting the pros and cons of each techniques. The remaining part of this work is organized thus: Section 2 described image feature extraction while Section 3 described various machine learning-based AIA techniques. In Section 4 we presented the summary and conclusion of the review.
2. Features Extraction

In this section, feature extraction (FE) which is an important step in an AIA model during the conversion of raw images into features was discussed. The automatic image annotation (AIA) process was based on many factors, which consist of the feature extraction process, use of suitable features in the AIA, the selected mathematical transform for determining the feedback usage, effective features, etc. Any annotation systems are effective if they can improve a few of the characteristic factors. In this study, the researchers used the low and high-level image features like texture, shape and colour for assembling all the information from the image to its recuperation.

3. Automatic image annotation techniques

In this section, a brief review of the techniques for AIA was conducted. These techniques are classified into Relevance Return based on AIA, Inter-domain semi-supervised learning based on AIA, Metadata Use based on AIA, Semantic Hierarchies based on AIA and Hollow coding models based on AIA.

3.1 Relevance Return

Researchers studied the use of user feedback to improve image annotation, as it is difficult to construct an appropriate classifier to automatically annotate images with text labels. Several relevance return models have been incorporated into the image annotation models. In 2013, [1] presented an interactive tool, called IGAnn, in the form of a semi-automatic image annotation procedure for users. The annotation approach in IGAnn is based on the return of relevance. The user identifies several images that relevance to a given label to improve the learning of the classifier which determines images that are most likely to be associated with that label. The learning of the hierarchical classifier is carried out by a semi-supervised approach that involves both labeled and unlabelled images to obtain a good annotation result. [2] proposed a multi-directional search technique (MDS) for the propagation of image annotation. This approach dynamically analyzes the user's return of relevance and considers multiple neighborhoods by breaking down the initial query into separate sub-requests. Unlike KPPV techniques based on a return of relevance that seek in a single group of neighbors, MDS provides a remedy for this weakness since it covers distinct groups of images. It can therefore simultaneously annotate images with different visual characteristics, but with the same semantic concept. In [3], the system presented chooses appropriate words from a vocabulary such as labels for a given image, and refines the labels using user feedback. For a given image, the user return to the labels corrects the outputs of the auxiliary classifiers and the system recommends more appropriate labels for future iterations. [4] have proposed a structured model for image labelling that takes into account the dependencies between image labels explicitly. This model is interactive where the user provides the value of certain labels, which leads to more accurate forecasts.

Annotation methods using relevance feedback significantly improve automatic image annotation results. However, they are time-consuming for the user.

3.2 Inter-domain semi-supervised learning

Inter-domain learning methods [5] have recently been proposed for annotation. The basic idea of these methods is to use data labeled from auxiliary domains. Labeled learning data is often fixed and quite small, but the amount of unlabelled data is vast and rapidly growing. Therefore, both unlabelled data in the target domain and data labeled in an auxiliary domain should be used to increase image annotation performance.[6] used a new semi-supervised learning algorithm (SSLKDE) for automatic video annotation. This algorithm uses labeled and unlabelled data to estimate a class’s conditional probability density. [7] proposed a new method of image annotation using the deduction of semantic concepts from
community images and their associated noisy labels. To deduce concepts more accurately, they propose a semi-supervised learning approach based on a hollow graph to exploit the data labeled and not labeled simultaneously. [8] proposed a semi-supervised learning method with hollow groups for image annotation called S2CLGS. Both labeled and unlabelled learning data were used with their diverse structural information in the target area to make semi-supervised learning. S2CLGS' performance for image annotation is enhanced using data labeled from auxiliary domains. Other approaches using semi-supervised inter-domain learning have been proposed in [9]. These approaches solve the problem of the availability of large collections of labeled data and thus allow for more accurate models than those achieved by purely supervised learning methods. However, these approaches required minimal difference in data distribution between auxiliary and target domains, which is not always simple.

3.3 Metadata Use

Web images are usually surrounded by multiple pieces of information (text description, URL, date, HTML,...). So we can use this information to annotate images. Several techniques [10], [11], [12] and [13] have been developed for web image annotation, most of these techniques incorporate visual features and metadata. [14] proposed an automatic system that annotates images using both web descriptions and visual features. The system needs at least one correct initial keyword and example image to start the process. The keyword is used for a web search to find images and their web descriptions. Then, the visual features are used to select a number of images similar to the example image. Web descriptions of selected images are grouped using a special text aggregation algorithm. Words in clusters are used for annotations. A similar method of annotation is presented in [11], [15], [16] proposed a method for predicting labels, tags and groups for images from the metadata recovered from the social network Flickr. Labels are provided by human annotators outside Flickr, which provide annotations based solely on the content of the image. Tags can be provided by a number of commenters, and may include hard-to-detect information from visual content alone, such as the camera's mark and image location. Groups are similar to labels, with the difference that the groups to which an image is assigned are chosen only by the author of the image. [17] analyzed the co-occurring images and their textual data in press articles to generate possible annotations to these images. Textual data in the briefing materials describe the central point of reporting based on the images and titles given. They used textual data as a resource to assign annotations to images using WordNet's TF (Term Frequency) and WUP [18] similarities. The proposed method has shown that text analysis is another possible technique for automatically annotating images. An example of an annotation of an image by analyzing the text around it is shown in Figure 1.
3.4 Semantic Hierarchies

In most image annotation approaches, semantics is limited to its perceptual manifestation by learning a matching function combining low-level features with higher-level visual concepts. Semantic. The performance of these approaches varies depending on the number of concepts and the nature of the targeted data. Recently, several studies have focused on the use of semantic hierarchies to overcome these problems.

Semantic hierarchies can be grouped into four classes [19], [20] taxonomy, thesaurus, ontology and semantic network. Taxonomy is used to describe any kind of hierarchy between objects. It is used for hyperonymy/hyponymy-type relationships. A thesaurus is an organized list of controlled and standardized terms to represent the concepts of a domain. It can be considered an extension of a taxonomy. An ontology is a model for the formal description of concepts. It is defined by a set of concepts, properties and types of relationships between concepts. The formal model should be understandable by a machine. Semantic networks can be defined [21] as intermediaries between thesaurus and ontologies. They describe more relationships than thesaurus but less formal than ontologies.

Several approaches were based on the WordNet lexical basis for the construction of semantic hierarchies [22], [23] and [24]. The use of semantic hierarchies is increasingly interesting for the task of annotating images. They allow for better performance when the number of categories increases. However, building a semantic hierarchy from a given vocabulary is complex.

3.5 Hollow coding models

Hollow coding is the process of learning a sparse representation of a signal in terms of the coefficients of a set of basic signals or predictive variables [25]. Recently, motivated by the success of hollow coding for image classification [26], [27] and [28], hollow coding methods have been applied to solve image annotation problems. The annotation system proposed in [25] uses two layers of hollow coding.
processing to treat learning images as predictors and test images as the target signal. Learning images are grouped into themes based on textual and visual similarity. The first layer takes advantage of this group structure and identifies themes relevant to the test images. The next layer of hollow coding processes the reduced set of learning and vocabulary, containing images and words belonging to relevant themes only. In 2016 [29], the authors proposed a new approach to image annotation. This method is based on learning the multi-label dictionary with regularization of label consistency and the integration of partially identical labels. Dictionary learning aims to learn a complete visual dictionary from the space of learning images. [30] introduced the Semantic Label Embedding Dictionary (SLED) hollow-coding method to automatically annotate images. In this method, the learning data is divided at the beginning into a set of superimposed groups. Then, several dictionaries specific to each label are trained in order to explicitly decorate the representation of each label. To discover the context information hidden in co-occurrence labels, the semantic relationship between visual words in dictionaries and labels is explored in a multitasking learning method versus coefficients reconstruction of learning data. Finally, a test image is annotated through a robust label propagation method based on RankSVM [31].

Other approaches to image annotation using hollow coding have been proposed in [32], [33], [34] and [35].

4. Discussion

In this paper, the different methods of annotating images were discussed. Although manual image annotation is better at accuracy for image recovery, due to keywords chosen by the user to describe the semantic content of the images, however it is an intensive work and a tedious process. Automatic annotation methods require a large number of examples to perform learning. As a matter of fact, the available annotated images are not always sufficient. A good compromise would therefore be to choose an automatic annotation method and try to reduce the learning set. In this brief, we place ourselves in the perspective of automatic annotation.

| methods of annotation | pros | cons |
|------------------------|------|------|
| KPPV                   | -Do not require a learning phase | -Semantic gap |
| (Verma and Jawahar, 2012)[36] | -Good performance | -Unbalanced label frequencies |
| (Makadia, Pavlovic and Kumar, 2010)[37] | | |
| (Guillaumin et al., 2009)[38] | | |
| (Li, Snoek and Worring, 2009) [39] | | |
| (Zhang et al., 2011) [40] | | |
| (Kalayeh et al., 2014)[41] | | |
| Return of relevance    | -significantly improves results | - expensive for users |
| (Chiang, 2013)[1]      | | |
| (Yu, Hua and Cheng, 2012)[2] | | |
| (S Zhang et al., 2010)[3] | | |
| (Mensink, Verbeek and Csurka, 2011)[4] | | |
In Table 1, we attempted to synthesize the majority of the techniques of automatic annotation of images. We focused on the pros and cons of each of these techniques. In state-of-the-art works, there are several automatic annotation techniques, including generative models, discriminating models and graphic models[46]. Overall, automatic image annotation is a difficult search area where several major locks are to be lifted. The first is difficulty in extracting textual characteristics and the second lock is Semantic gap.
5. Conclusion

In this paper, a detailed state-of-the-art study of image annotation was presented. The previous techniques have been cited and discussed, therefore the criteria for evaluating annotation systems were obtained in this study. The databases used in this brief were also presented. In conclusion, a synthesis of techniques of automatic image annotation were successfully determined by presenting the pros and cons of each techniques. Finally, some of the researches on image Annotation systems tend to achieve high precision and low recall [44], on one hand some models of image annotation are locks. The first is difficulty in extracting textual characteristics and the second lock is Semantic gap. On other hand some models of image annotation are difficulty in linking visual features to textual features and unbalanced label frequencies.

6. References

[1] Chiang, C. C. (2013) ‘Interactive tool for image annotation using a semi-supervised and hierarchical approach’, Computer Standards and Interfaces, 35(1), pp. 50–58. doi: 10.1016/j.csi.2012.05.002.
[2] Yu, N., Hua, K. A. and Cheng, H. (2012) ‘A Multi-Directional Search technique for image annotation propagation’, Journal of Visual Communication and Image Representation, 23(1), pp. 237–244. doi: 10.1016/j.jvcir.2011.10.004.
[3] Zhang, S. et al. (2010) ‘Automatic image annotation using group sparsity’, in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 3312–3319. doi: 10.1109/CVPR.2010.5540036.
[4] Mensink, T., Verbeek, J. and Csurka, G. (2011) ‘Learning structured prediction models for interactive image labeling’, in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, pp. 833–840. doi: 10.1109/CVPR.2011.5995380.
[5] Csurka, G. (2019) ‘Domain Adaptation for Visual Applications: A Comprehensive Survey’, arXiv preprint arXiv:1702.05374. Available at: http://arxiv.org/abs/1702.05374 (Accessed: 13 October 2019).
[6] Wang, M. et al. (2006) ‘Automatic video annotation by semi-supervised learning with kernel density estimation’, in Proceedings of the 14th Annual ACM International Conference on Multimedia, MM 2006, pp. 967–976. doi: 10.1145/1180639.1180855.
[7] Tang, J., Yan, S., et al. (2011) ‘Inferring Semantic Concepts from Community-Contributed Images and Noisy Tags’, In Proceedings of the 17th ACM international conference on Multimedia, pp. 223–232. Available at: http://www.flickr.com/ (Accessed: 12 October 2019).
[8] Yuan, Y. et al. (2013) ‘Image annotation by semi-supervised cross-domain learning with group sparsity’, Journal of Visual Communication and Image Representation. Academic Press Inc., 24(2), pp. 95–102. doi: 10.1016/j.jvcir.2012.02.007.
[9] Duan, L. et al. (2009) ‘Domain adaptation from multiple sources via auxiliary classifiers’, in Proceedings of the 26th Annual International Conference on Machine Learning - ICML ’09. New York, New York, USA: ACM Press, pp. 1–8. doi: 10.1145/1553374.1553411.
[10] Jin, Y. et al. (2005) ‘Image Annotations By Combining Multiple Evidence & WordNet’, In Proceedings of the 13th annual ACM international conference on Multimedia, pp. 706–715.
[11] Li, X. et al. (2006) ‘Image Annotation by Large-Scale Content-based Image Retrieval Image enhancement and restoration’, In Proceedings of the 14th ACM international conference on Multimedia, (ACM), pp. 607–610. doi: 10.1145/1180639.1180764.
[12] Wang, C. et al. (2007) ‘Content-based image annotation refinement’, in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. doi: 10.1109/CVPR.2007.383221.
[13] Choi, D. and Kim, P. (2012) ‘Automatic image annotation using semantic text analysis’, in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), pp. 479–487. doi: 10.1007/978-3-642-32498-7_36.

[14] Wang, X. J. et al. (2010) ‘ARISTA - Image search to annotation on billions of web photos’, in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 2987–2994. doi: 10.1109/CVPR.2010.5540046.

[15] Wang, X. J. et al. (2006) ‘AnnoSearch: Image auto-annotation by search’, in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1483–1490. doi: 10.1109/CVPR.2006.58.

[16] McAuley, J. and Leskovec, J. (2012) ‘Image labeling on a network: Using social-network metadata for image classification’, in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), pp. 828–841. doi: 10.1007/978-3-642-33765-9_59.

[17] Choi, D. and Kim, P. (2012) ‘Automatic image annotation using semantic text analysis’, in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), pp. 479–487. doi: 10.1007/978-3-642-32498-7_36.

[18] Wu, Z. and Palmer, M. (1994) VERB SEMANTICS AND LEXICAL SELECTION. Yakhnenko, O. and Honavar, V. (2008) ‘Annotating images and image objects using a hierarchical Dirichlet process model’, in Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1–7. doi: 10.1145/1509212.1509213.

[19] Garshol, L. M. (2004) ‘Metadata? Thesauri? Taxonomies? Topic maps! Making sense of it all’, Journal of Information Science, 30(4), pp. 378–391. doi: 10.1177/0165551504045856.

[20] Tousch, A. M., Herbin, S. and Audibert, J. Y. (2012) ‘Semantic hierarchies for image annotation: A survey’, Pattern Recognition. Elsevier Ltd, 45(1), pp. 333–345. doi: 10.1016/j.patcog.2011.05.017.

[21] Garshol, L. M. (2004) ‘Metadata? Thesauri? Taxonomies? Topic maps! Making sense of it all’, Journal of Information Science, 30(4), pp. 378–391. doi: 10.1177/0165551504045856.

[22] Aslandogan, Y. et al. (1997) ‘Using semantic contents and WordNet in image retrieval’, SIGIR 97 Philadelphia PA, USA, pp. 286–295. Available at: http://cake.fiu.edu/Publications/Aslandogan+al-97-US.Using_semantic_contents_and_Wordnet_in_image_retrieval.published-scanned.pdf (Accessed: 12 October 2019).

[23] Marszalek, M., Schmid, C. and Marszalek, M. (2007) ‘Semantic Hierarchies for Visual Object Recognition’, In 2007 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–7. doi: 10.1109/CVPR.2007.383272i.

[24] Li, L.-J. et al. (2009) ‘ImageNet: a Large-Scale Hierarchical Image Database’, In 2009 IEEE conference on computer vision and pattern recognition, pp. 248–255. doi: 10.1109/CVPR.2009.5206848.

[25] Tariq, A. and Forooosh, H. (2017) ‘Image Annotation using Multi-Layer Sparse Coding’, arXiv preprint arXiv, p. 1705.02460. Available at: http://arxiv.org/abs/1705.02460 (Accessed: 13 October 2019).

[26] ShenghuaGao, I. and Liang-TienChia, P. (2010) ‘Local features are not lonely–Laplacian sparse coding for image classification’. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.180.6821&rep=rep1&type=pdf (Accessed: 13 October 2019).

[27] Chen, X. et al. (2011) ‘Multi-label visual classification with label exclusive context’, in Proceedings of the IEEE International Conference on Computer Vision, pp. 834–841. doi: 10.1109/ICCV.2011.6126323.

[28] Gao, S., Chia, L. T. and Tsang, I. W. H. (2011) ‘Multi-layer group sparse coding - For concurrent image classification and annotation’, in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, pp. 2809–2816. doi: 10.1109/CVPR.2011.5995454.
[29] Jing, X. Y. et al. (2016) ‘Multi-Label Dictionary Learning for Image Annotation’, IEEE Transactions on Image Processing. Institute of Electrical and Electronics Engineers Inc., 25(6), pp. 2712–2725. doi: 10.1109/TIP.2016.2549459.
[30] Cao, X. et al. (2015) ‘SLED: Semantic label embedding dictionary representation for multilabel image annotation’, IEEE Transactions on Image Processing. Institute of Electrical and Electronics Engineers Inc., 24(9), pp. 2746–2759. doi: 10.1109/TIP.2015.2428055.
[31] Sun, F. et al. (2014) ‘Multi-label image categorization with sparse factor representation’, IEEE Transactions on Image Processing, 23(3), pp. 1028–1037. doi: 10.1109/TIP.2014.2298978.
[32] Wang, M., Li, F. and Wang, M. (2012) ‘Collaborative visual modeling for automatic image annotation via sparse model coding’, Neurocomputing, 95, pp. 22–28. doi: 10.1016/j.neucom.2011.04.049.
[33] Gao, S. et al. (2014) ‘Concurrent single-label image classification and annotation via efficient multi-layer group sparse coding’, IEEE Transactions on Multimedia. Institute of Electrical and Electronics Engineers Inc., 16(3), pp. 762–771. doi: 10.1109/TMM.2014.2299516.
[34] Liu, W. et al. (2014) ‘Multiview Hessian discriminative sparse coding for image annotation’, Computer Vision and Image Understanding, 118, pp. 50–60. doi: 10.1016/j.cviu.2013.03.007.
[35] Lu, Z. et al. (2015) ‘Semantic sparse recoding of visual content for image applications’, IEEE Transactions on Image Processing. Institute of Electrical and Electronics Engineers Inc., 24(1), pp. 176–188. doi: 10.1109/TIP.2014.2375641.
[36] Verma, Y. and Jawahar, C. V. (2012) ‘Image annotation using metric learning in semantic neighbourhoods’, in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), pp. 836–849. doi: 10.1007/978-3-642-33712-3_60.
[37] Makadia, A., Pavlovic, V. and Kumar, S. (2010) ‘Baselines for image annotation’, International Journal of Computer Vision, 90(1), pp. 88–105. doi: 10.1007/s11263-010-0338-6.
[38] Guillaumin, M. et al. (2009) ‘TagProp: Discriminative metric learning in nearest neighbor models for image auto-annotation’, in Proceedings of the IEEE International Conference on Computer Vision, pp. 309–316. doi: 10.1109/ICCV.2009.5459266.
[39] Li, X., Snoek, C. G. M. and Worring, M. (2009) ‘Learning social tag relevance by neighbor voting’, IEEE Transactions on Multimedia, 11(7), pp. 1310–1322. doi: 10.1109/TMM.2009.2030598.
[40] Zhang, S. et al. (2010) ‘Automatic image annotation using group sparsity’, in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 3312–3319. doi: 10.1109/CVPR.2010.5540036.
[41] Kalayeh, M. et al. (2014) ‘NMF-KNN: Image annotation using weighted multi-view non-negative matrix factorization’, cv-foundation.org, pp. 1–8. Available at: https://www.cv-foundation.org/openaccess/content_cvpr_2014/html/Kalayeh_NMF-
[42] Lixin Duan et al. (2009) ‘Domain Transfer SVM for video concept detection’, in 2009 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, pp. 1375–1381. doi: 10.1109/CVPR.2009.5206747.
[43] Bouzaeni, A., & Tabbone, S. (2018, August). Image Annotation Using a Semantic Hierarchy. In Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR) (pp. 3-13). Springer, Cham.
[44] Mundher, Myasar, et al. (2014) "Digital watermarking for images security using discrete slantlet transform." Applied Mathematics & Information Sciences 8.6 2823.
[45] Waheed, Safa Riyadh, et al.(2020) "Fuzzy Logic Controller for Classroom Air Conditioner." Journal of Physic, IOP Publishing: Conference Series. Vol. 1484. No. 1.
[46] Mundher, Myasar, et al. " A Survey Automatic Image Annotation Based on Machine Learning Models." Journal of Engineering and Applied Sciences2018 14 (20), 7627-7635.