ISSN:0975-9646

A Survey on Automatic Annotation and Annotation Based Image Retrieval

Himali Chaudhari , Prof D.D.Patil

Shri Sant Gadge Baba College of Engineering and Technology (SSGBCOET), Bhusawal

Computer Science Department, North Maharashtra University Jalgaon,India

Abstract—**With the advances in multimedia technologies collection of digital images is growing rapidly. In the last few decades Content Based Image retrieval is very important area of research in the field of image retrieval. Research shows that main challenge in the CBIR systems is the semantic gap. Image annotation is the effective way to bridge a semantic gap between low level features and high level semantics.Many researchers develop and use lots of approaches towards image annotation. Automatic image annotation is the process of automatically assigning semantic labels to images. This paper presents the survey of different approaches for automatic annotation and annotation based image retrieval. This paper aims to cover the latent space and generative approaches for automatic image annotation.**

Keywords— **Automatic Image Annotation, Content Based Image Retrieval, Semantic Gap, Annotation Based Image Retrieval**

I Introduction

Now a days due to increase in digital media like camera, mobile phones collection of digital images is growing rapidly .So there is need to efficiently store and retrieve theses images from a large collection of image databases. In the recent years many image retrieval systems have been developed to browse, search and retrieve images from large databases. Current State of the art in image retrieval has two approaches: content-based image retrieval (CBIR) and annotation based image retrieval (ABIR).They mainly differ in the way a query is formulated. CBIR systems search images using low level features such as color, texture, shape, spatial layout etc. which can be automatically extracted and used to index images. Humans tend to associate images with keywords rather than query image. The initial requirement of CBIR systems is to provide query similar image to the retrieval system. The CBIR systems fail to meet user expectations because those systems are unable to index images according to the high level features (keywords, text descriptors etc) as perceived by the user. The main challenge in the CBIR is the two gaps namely semantic gap and sensory gap.

Smeulders et al [1] define the semantic gap as the “lack of coincidence between the information that one can extract from the visual data and the interpretation that the same data have for a user in a given situation”. The aim of content-based retrieval systems must be to provide maximum support in bridging the semantic gap between low level features extracted from images and the high level information need of the user.

Smeulders et al [1] also mention another gap of relevance to content based retrieval, the sensory gap, which they define as “the gap between the object in the world and the information in a (computational) description derived from a recording of that scene.

While the former gap brings in the issue of users' interpretations of images and how it is inherently difficult to capture them in visual content, the latter gap makes recognition from image content challenging due to limitations in recording and description capabilities[27].

Image annotation, the task of associating text to the semantic content of images, is a good way to reduce the semantic gap and can be used as an intermediate step to image retrieval. It enables users to retrieve images by text queries and often provides semantically better results than content-based image retrieval. In recent years, it is observed that image annotation has attracted more and more research interests.

When images are retrieved using these annotations, such retrieval is known as annotation-based image retrieval (ABIR). Annotation-Based Image Retrieval (ABIR) systems are an attempt to incorporate more efficient semantic content into both text-based queries and image captions. As can be seen in many of today's image retrieval systems, ABIR is considered more practical. Consequently, textual information should play a central role in visual information retrieval. However, CBIR has been researched far more than ABIR [2].

This paper presents a survey of the research related to the automatic annotation and annotation based image retrieval. The rest of the paper is organised as follows: Section II reviews related to CBIR and document retrieval .Section III reviews automatic image annotation including comparison of different approaches of automatic image annotation. Section IV concludes the paper.

II Related Work

A. Content Based Image Retrieval

In the image retrieval, content-based image retrieval (CBIR) is one of the most important research topics. In the last decade more than 200 Content-based image retrieval (CBIR) systems have been studied and explored [11]. Some of the CBIR systems like QBIC [3], Photobook [4], Virage[5], Visual Seek [6], WebSeek [6], Netra[7], Cypress[8],are attracting attention but they are still not very common. Since CBIR systems mainly depend on low level features for retrieving images, user need to provide the retrieval system with query similar images which is a challenging need of today. Several surveys on CBIR research in literature can be found in [1, 9-13].

B. Document Retrieval

Annotation based image retrieval is based on the theory of text retrieval systems. Many document retrieval and indexing techniques were incorporated into ABIR systems. In this section we will discuss some of the important document retrieval techniques.

Latent Semantic Indexing (LSI) was first introduced by S. Deerwester [14] as a document retrieval technique to address the some of the shortcomings inherent in traditional lexical matching techniques.LSI deals with the problems of synonymy (Many words refer to same object) and polysemy (Many words have multiple meanings). LSI tries to search for something that is closer to representing the underlying semantics of a document, rather than just by matching specific keywords. Latent semantic indexing method starts with the creation of terms by document matrix. Then this high dimensional matrix is further decomposed into a reduced dimension matrix called Singular Value Decomposition (SVD). This filters out the noise found in a document, such that two documents that have same semantics will be located close to one another in a multi-dimensional space [15].However LSI has some drawbacks such as reduced dimensions are difficult to interpret , SVD is computationally expensive, performance and speed level degrades when applied to large scale collection.

As LSI has number of deficits due to its unsatisfactory and incomplete theoretical foundation, Hofmann [16] presented the probabilistic LSI (PLSI) model, as an alternative to LSI. The roots of PLSI go back to the LSI. Like LSI, PLSI also deals with synonymous as well as polysemous words. PLSI is an automated document indexing technique, in which each document is represented by its word frequency. PLSI is also known as the aspect model which is the Latent variable model in which latent variables are associated with observed variables. Consequently, it has a more robust statistical foundation, and is able to provide a proper generative data model. PLSI is based on the Expectation Maximization (EM) algorithm. While PLSI is one of the good text analysis technique it has some drawbacks such as it is incomplete since provide no probabilistic model at the level of documents,leads to over fitting problems if there are too many parameters in the model and it's not clear how to assign how to assign probability to a document outside of the training data.

To address the limitations of pLSI, Blei et al. [17] proposed a unsupervised, generative model called Latent Dirichlet Allocation (LDA).It is closely related to PLSI. It is a powerful generative probabilistic model developed for modeling words in a document. In LDA each document is a mixture of a small number of latent topics, here each topic is characterized by a distribution over words.

III AUTOMATIC IMAGE ANNOTATION

In some scenarios most of the times desired pictorial information can be efficiently described by means of keywords. The process of assigning a set of keywords (or text) to an image is called as annotation. Image Annotation systems attempt to reduce the semantic gap. The task of automatically assigning semantic labels to images is known as automatic image annotation (AIA). Automatic image annotation is also known as auto-annotation or linguistic indexing [11].In the last decade Automatic image annotation (AIA) is a highly popular topic in the field of information retrieval research. The main idea of AIA is to automatically learn semantic descriptors from large number of image samples, and use the concept models to label new images. Once images are annotated with semantic labels, images can be retrieved by keywords.

The problem of automatic image annotation is closely related to that of content-based image retrieval. In the recent year a variety of learning methods have been actively researched for automatically annotating images. The main purpose of these methods is to assign a set of keywords to each image. Different strategies including co-occurrence model [19], machine translation model [20], latent space approaches [21][22], classification approaches [27] - [29] and relevance language models [30] [31] have been proposed in the literature and each strategy tries to improve previous one.

Automatic image annotation can be approached with a variety of machine learning methods, from supervised classification to probabilistic to clustering.

Trong-Ton Pham [21] et al proposed a model to study the effect of LSA on multimedia document retrieval (MDR) and automatic image annotation (AIA).The model studies the effect of LSA on multimedia document indexing and retrieval on a significant number of documents. Also show that fusion of several image representation methods improves the results on AIA task. Compared to the previous methods of multimedia document retrieval this model uses significantly larger number of documents. This model shows improvements to the results of the effect of LSA on image retrieval system on large scale database and results of AIA using combination of different image representations.

Florent Monay and Daniel GaticaPerez [22] propose and compare automatic image annotation strategies for latent space models such as Latent Semantic Analysis and Probabilistic latent semantic analysis (PLSA).They have discussed three annotation strategies for direct match, LSA and PLSA. Annotation by direct match and LSA are based on comparison and annotation by propagation while annotation with PLSA is based on statistical inference. Results of annotation by propagation (LSA and direct match) are better than annotation by inference (PLSA).

Florent Monay and Daniel GaticaPerez et al[23] proposed probabilistic latent space models for automatic image annotation called PLSA words. The model constrains the latent space by focusing on the textual features .The model consist of two steps learning parameters and annotation by inference respectively. This PLSA based annotation model is shown to outperform previous latent space models

David M. Blei and Michael I. Jordan[24] extended the Latent Dirichlet Allocation (LDA) Model and proposed a model for modeling annotated data of multi-type , called correspondence latent Dirichlet allocation (Corr-LDA).They have described three models namely Gaussian multinomial mixture model (GM-Mixture), Gaussian Multinomial LDA (GM-LDA) and Correspondence LDA (Corr-LDA).Corr -LDA consists of GM-Mixture and GM – LDA. Corr-LDA model can be used for automatic image annotation, automatic region annotation, and text-based image retrieval. It can also be applied to any kind of annotated data such as video/closed-captions, music/text, and gene/functions.

Kobus Barnard Pinar Duygulu et al [25] presented Multi-Modal Hierarchical Aspect Model and Mixture of Multi-Modal Latent Dirichlet Allocation model for image annotation. Multi-Modal Hierarchical Aspect Models is based on Hofmann's hierarchical model for text while second model is the extension of LDA.

Recently Konstantinos A. Raftopoulos et al [26] introduced a novel probabilistic approach for automatic annotation, indexing and annotation-based retrieval of images called as Markovian Semantic Indexing (MSI).This method is suitable when per image sparse keyword annotation is limited. The authors compare MSI with the existing methods such as LSI and PLSI under two scenarios when markovian annotation is available and using external annotation. A comparison to LSI on 64 images gathered from the Google Image Search and annotated in a transparent way, revealed certain advantages for the MSI method, mainly in retrieving images with deeper dependencies than simple keyword co-occurrence. MSI is shown to outperform both LSI and PLSI.

IV CONCLUSION

A wide variety of researches have been made on image retrieval in multimedia databases. Each work has its own technique, contribution and limitations. In CBIR main challenge is the semantic gap. In order to overcome the problem in semantic gap, automatic image annotation is the solution. In this paper, we attempted to provide a comprehensive survey on automatic image annotation techniques. As a survey paper, we might not include each and every aspect of individual works; however we have focused on the latent space approaches and generative models of automatic image annotation.

TABLE I
SUMMARY OF DOCUMENT RETRIEVAL METHODS

Methods	Document Representation	Problems addressing	Limitations	Applications
LSI [14]	Term document matrix	Polysemy , Synonymy	Dimensions are difficult to interpret , computationally expensive,storage ,efficiency	Information retrieval, [15], Information Filtering [34] , Cross language retrieval [33], Spam filtering [32]
PLSI [16]	Word frequency	Using probability, Automated Document Indexing	Over fitting, Generalization	Automatic easy grading [36], multi-criteria recommender system[37], classification[38] , online event analysis[39]
LDA [17]	mixture of topics with a probability distribution	Exchangeability	Incapable to model relations among topics	Automatic easy grading [35], Automatic labelling [39], Word sense disambiguation[40]

TABLE II
COMPARISON OF AUTOMATIC IMAGE ANNOTATION METHODS

Author	Methods	Image representation	Dataset	Performance measure
Trong-Ton Pham [21]	LSA	Bags of Visterms	Corel	Precision and recall
Florent Monay and Daniel GaticaPerez [22]	LSA and PLSA	Vector Space	Corel	normalized score
Florent Monay and Daniel GaticaPerez et al[23]	PLSA	RGB,Blobs	Corel	annotation accuracy, normalized score
David M. Blei and Michael I. Jordan[24]	LDA	Blobs	Corel	perplexity
Konstantinos A. Raftopoulos et al [26]	MSI	Probability vector	Google image search	Precision and recall

REFERENCES

[1] A.W.M. Smeulders, M. Worring, A. Gupta, R. Jain, Content-based image retrieval at the end of the early years, IEEE Trans. Pattern Anal. Mach. Intell. 22 (12) (2000) 1349–1380.

[2] M. Inoue. On the need for annotation-based image retrieval. *Workshop on Information Retrieval in Context*, 2004.

[3] Flickner, M., Sawhney, H., Niblack, W., Ashley, J., Huang, Q., Dom, B., Gorkani, M., Hafner, J., Lee, D., Petkovic, D., Steele, D.,And Yanker, P. 1995. Query by image and video content: The QBIC system. IEEE Comput. 28, 9, 23–32.

[4] A. Pentland, R. Picard, and S. Sclaroff. Photobook: Tools for content-based manipulation of image databases. *IJCV*, 18(3):233–254, 1996.

[5] A. Gupta, R. Jain, Visual information retrieval, Commun. ACM 40(5) (1997) 70–79.

[6] J. R. Smith and S.-F. Chang. Visualseek: a fully automated content-based image query system. In *Proc.ACM Multimedia '96*, 1996.

[7] Ma, W. And Manjunath, B. 1997. Netra: A toolbox for navigating large image databases. In Proceedings of the IEEE International Conference on Image Processing (ICIP).

[8] D. A. Forsyth, J. Malik, M. M. Fleck, H. Greenspan ,T. Leung, S. Belongie, C. Carson, and C. Bregler. Finding pictures of objects in large collections of images.In *Proc. ECCV 96 Workshop on Object Rep.*, 1996.

[9] M.S. Lew, N. Sebe, C. Djeraba, R. Jain, Content-based multimedia information retrieval: state of the art and challenges, ACM Transactions on Multimedia Computing, Communications and Applications 2 (1) (2006) 1–19.

[10] N. Vasconcelos, From pixels to semantic spaces: advances in content-based image retrieval, Computer 40 (7) (2007) 20–26.

[11] R. Datta, D. Joshi, J. Li, J.Z. Wang, Image retrieval: ideas, influences and trends of the new age, ACM Computing Surveys 40 (2) (April 2008).

[12] F. Long, H.J. Zhang, D.D. Feng, Fundamentals of content-based image retrieval, in: D.D. Feng, W.C. Siuandg, H.J. Zhan (Eds.), Multimedia Information Retrieval and Management, Springer, 2003.

[13] Y. Rui, T.S. Huang, S.F. Chang, Image retrieval: current techniques, promising directions and open issues 10 (1999) 39–62 Journal of Visual Communication and Image Representation 10 (1999) 39–62.

[14] 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, 6 (1990), 391_407.

[15] M.W. Berry, S.T. Dumais, and G.W. O'Brien, "Using Linear Algebra for Intelligent Information Retrieval," SIAM Rev., vol. 37, no. 4, pp. 573-595, 1995.

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

[17] D.M. Blei and A.Y. Ng, and M.I. Jordan, "Latent Dirichlet Allocation," J. Machine Learning Research, vol. 3, pp. 993-1022, 2003.

[18] Z. Guo, S. Zhu, Y. Chi, Z. Zhang, and Y. Gong, "A Latent Topic Model for Linked Documents," Proc. 32nd Int'l ACM SIGIR Conf.Research and Development in Information Retrieval (SIGIR), 2009.

[19] Y Mori, H Takahashi and R Oka, Image-to-word transformation based on dividing and vector quantizing images with words, In MISRM99 First Intl. Workshop on Multimedia Intelligent Storage and Retrieval Management,1999.

[20] Duygulu P et al., Object Recognition as Machine Translation: Learning a Lexicon for a Fixed Image

[21] Vocabulary. in proceedings of 7th European Conference on Computer Vision, 2002.

[22] T.-T. Pham, N.E. Maillot, J.-H. Lim, and J.-P. Chevallet, "Latent Semantic Fusion Model for Image Retrieval and Annotation," Proc. 16th ACM Conf. Information and Knowledge Management(CIKM), 2007.

[23] Monay F and D Gatica-Perez, On Image Auto-Annotation with Latent Space Models, in proc. of ACM International Conference on Multimedia, 2003.

[24] Monay, F., Gatica-Perez, D.: Plsa-based image auto-annotation: constraining the latent space .In: Multimedia '04: Proceedings of the 12th annual ACM international conference on Multimedia, New York, NY, USA, ACM Press (2004) 348–351

[25] Blei D and M Jordan, Modeling Annotated Data, in proceedings of 26th International Conference on Research and Development in Information Retrieval (SIGIR), 2003.

[26] Barnard, K., Duygulu, P., Forsyth, D., de Freitas, N., Blei, D., Jordan, M.: Matching words and pictures. Machine Learning Research 3 (2003) 1107–1135

[27] Konstantinos A. Raftopoulos, Member, IEEE, Klimis S. Ntalianis, Dionyssios D. Sourlas, and Stefanos D. Kollias, Member, IEEE Mining User Queries with Markov Chains: Application to Online Image Retrieval IEEE Transactions On Knowledge And Data Engineering, Vol. 25, No. 2, February 2013

[28] Chang E et al., Cbsa: Content-Based Soft Annotation for Multimodal Image Retrieval Using Bayes Point Machines, CirSysVideo, 13(1): pp. 26-38, 2003.

[29] Cusano C, G Ciocca and R Schettini, Image Annotation Using SVM, in proceedings of Internet Imaging IV, Vol. SPIE 5304, 2004.

[30] Li J and J Z Wang, Automatic Linguistic Indexing of Pictures by a Statistical Modeling Approach, IEEE Trans.on Pattern Analysis and Machine Intelligence, 25(19): pp.1075-1088.

[31] Jeon J, V Lavrenko and R Manmatha, "Automatic Image Annotation and Retrieval Using Cross-Media Relevance Models", in Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval, 2003.

[32] Lavrenko V, R Manmatha and J Jeon, "A Model for Learning the Semantics of Pictures", in Proceedings of Advance in Neutral Information Processing, 2003.

[33] Gansterer, W., Janecek, A., and Neumayer, R., "Spam filtering based on latent semantic indexing", Survey of Text Mining II: Clustering, Classification, and Retrieval, 2008,165-183.

[34] Automatic Cross-Language Retrieval Using Latent Semantic Indexing by Michael Littman, Susan T. Dumais, Thomas K. Landauer - Cross-Language Information Retrieval, chapter 5 , 1998

[35] Foltz, P. W. (1990) Using Latent Semantic Indexing for Information Filtering. In R. B. Allen (Ed.) Proceedings of the Conference on Office Information Systems, Cambridge, MA, 40-47.

[36] Kakkonen, T., Myller, N., and Sutinen, E., "Applying latent Dirichlet allocation to automatic essay grading", Lecture Notes in Computer Science, 4139, 2006, 110-120.

[37] Ahrendt, P., Goutte, C., and Larsen, J., "Co-occurrence models in music genre classification", IEEE International workshop on Machine Learning for Signal Processing, 2005,247-252.

[38] Yin Zhang, Yueting Zhuang, Jiangqin Wu, Liang Zhang,"Applying probabilistic latent semantic analysis to multi-criteria recommender system AI communications" 2009

[39] Chou, T.-C., and Chen, M.C., "Using Incremental PLSI for Threshold-Resilient Online Event Analysis", Knowledge and Data Engineering, IEEE Transactions on, 20 (3), 2008,289-299.

[40] Magatti, D., Calegari, S., Ciucci, D., and Stella, F., "Automatic Labeling Of Topics", Ninth International Conference on Intelligent Systems Design and Applications, 2009, 1227-1232.

[41] Boyd-Graber, J., Blei, D., and Zhu, X., "A topic model for word sense disambiguation", 2007, 1024-1033.