Review: Automatic Semantic Image Annotation

Shereen A. Hussein

Computer Science Department, Helwan University, Cairo, Egypt

E-mail: eng_shereen@hotmail.com

Howida Youssry Abd El Naby

Computer Science Department, Misr University for Science & Technology, Cairo, Egypt

E-mail: howida@hotmail.com

Aliaa A. A. Youssif

Computer Science Department, Helwan University, Cairo, Egypt
E-mail: aliaay@helwan.edu.eg

Abstract

There are many approaches for automatic annotation in digital images. Nowadays digital photography is a common technology for capturing and archiving images because of the digital cameras and storage devices reasonable price. As amount of the digital images increase, the problem of annotating a specific image becomes a critical issue. Automated image annotation is creating a model capable of assigning terms to an image in order to describe its content. There are many image annotation techniques that seek to find the correlation between words and image features such as color, shape, and texture to provide an automatically correct annotation words to images which provides an alternative to the time consuming work of manual image annotation. This paper aims to cover a review on different Models (MT, CRM, CSD-Prop, SVD-COS and CSD-SVD) for automating the process of image annotation as an intermediate step in image retrieval process using Corel 5k images.

Keywords: - Image indexing, Visual Features, Content-Based Image Retrieval, Image annotation, Semantic Image annotation, Semantic Gap.

I.   Introduction

Nowadays, image databases are becoming very large and can be used in many areas. These databases may include public domain applications such as social networks and specific domain such as medical applications [19], [25]. Many decisions depend on the process of finding a specific image of set of images so there is an important need for efficient techniques for storage, indexing and retrieval of multimedia information. In particular, image retrieval is still a big challenge despite 20 years of research. The image retrieval field is applying the set of techniques for browsing, searching and retrieving images from a large collection of digital images. Usually, such systems operate in two steps:

1-   Image indexing: which could be defined as the process of extracting, modeling and storing the content of the image, the image data relationships, or other patterns not explicitly stored.[1]

2-   Image search: This consists in executing a matching model to evaluate the relevance of previously indexed images with the user query.

In the text based image retrieval the images are annotated and the data base retrieves them same way as text documents. Because of needing the human intervention it is very difficult and time consuming task. So, content based image retrieval using low level image feature such as color, texture, shape was proposed. Some image processing techniques are used to search for images relevant to a query such as features extraction. Unfortunately the naive user may not be familiar with low level visual features so a problem in semantics has been raised. Semantic gap between low level image content and high level content is main drawback of CBIR system. Image annotation is the next improvement in process of image retrieval. It associates one or more multiple concepts with objects and images therefor it is the professional way for content based image retrieval.

The main contribution of this work is a close review on the different techniques of automating the process of image annotation that used in image retrieval process. In Section II, the Content-Based Image Indexing and Retrieval with its drawbacks and the automated image annotation and the problem of semantic gap are discussed. In Section III, the different approaches in automated image annotation and their pros and cons are listed. In Section IV, a comparison of the annotation methods using the COREL dataset is presented. In Section V, experimental results are explained, and the conclusion in Section VI.

II.   Content Based Image Indexing and Retrieval

Images manually annotated by text descriptors or tags which are then used by an image retrieval system. This process is called 'iconography'. With respect to image retrieval field, 'iconography' means the process of human annotation of images. The iconography [13] may be of prime importance for image retrieval systems since it provides valuable information about image content. It allows disposing of the visual content description of an image, and its subjective dimensions which are priceless for understanding image semantics. However, the iconography requires a considerable level of human labour and it cannot be considered for large image databases.

Current approaches for automatic image indexing and retrieval can be classified into text-based approaches or content-based approaches, according to the used content (or modality) to index images as shown in Figure 1:

In the text-based approaches, images are indexed by a set of text descriptors which are extracted from the surrounding context but it suffers from the following problems:

• The surrounding context is not always relevant with the image content, or sometimes only, a small part of it describes its content. Consequently, the surrounding context is not always relevant for indexing images. Moreover, these methods do not consider the image content during the indexing process, and therefore there is no guarantee that the provided annotation is relevant with respect to image content.

• Text-based approaches are subject to the subjectivity of image semantics, also known as the problem of subjectivity of human perception, which occurs when the one who provided the image description has a different background, and/or he want to express a different semantics in the image content, compared to the one who is searching the image.

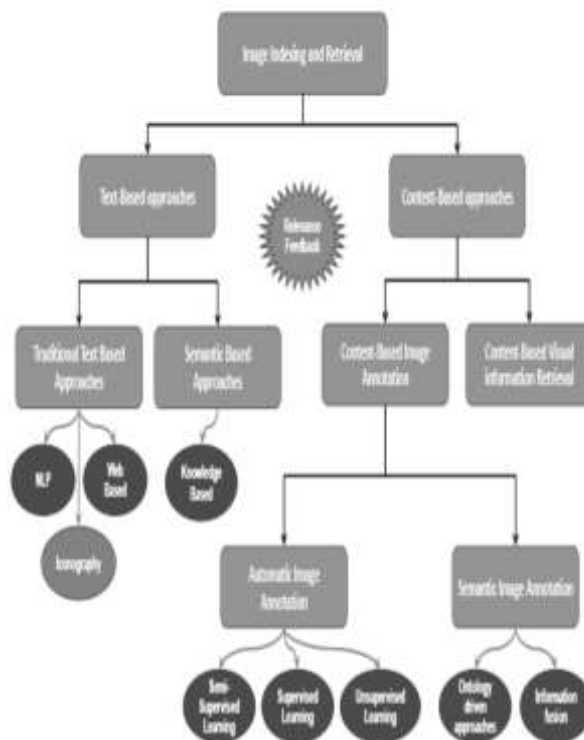• Currently, many enormous image databases are generated daily without any surrounding context.

Figure 1 Proposed taxonomy of image retrieval approaches [13]

A.   Content-Based Image Annotation

We focus on Content-based image annotation which was introduced in the early 1980s to overcome the problems of text-based retrieval which are dependency on the surrounding context or subjectivity of human perception and explained previously in detail. In this type of approaches [19], images are indexed and retrieved using their content as explained in Figure 2.
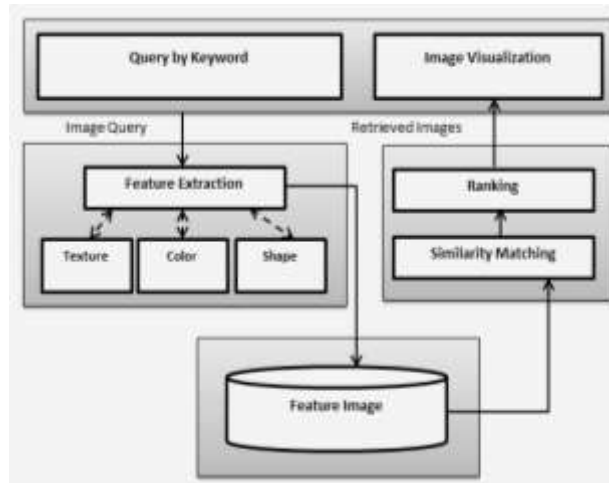
Figure 2 Typical CBIR System [5]

Several algorithms have been proposed to extract and store efficiently the low-level or mid-level visual features. The Low-level visual properties are the basis on which subsequent perceptual organization takes place where the High-level properties arise from specific ways of organizing low-level properties [30]. It is hard to extract semantically meaningful entities using the low-level features of images. This is known as the semantic gap problem, which is defined 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". Figure 3 illustrates this problem with an example.
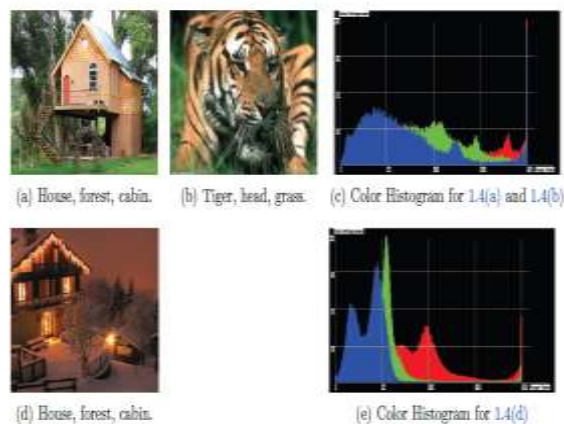

Figure 3 Semantic gap problem. images (a) and (b) have similar color histograms but different meanings. images (a) and (d) have different color histograms but the same meaning [13]

The main issue in image retrieval field is how to relate the visual content of images (low-level or mid-level visual features) to its semantic content (or high-level concepts). Consequently, the development of new approaches allowing narrowing the semantic gap [24] has been a core research topic since already ten years. Many approaches have focused on the problems of recovering effective image descriptors on one hand, and on the other hand on developing efficient machine learning algorithms in order to provide robust methods that allow mapping visual features into semantic concepts. These approaches are also known as image classification or image categorization approaches [13].

Automatic image annotation approaches seem to be also insufficient to bridge the semantic gap as they face the scalability problem when the concept number is high and depend on the targeted datasets as well. They are also subject to many types of uncertainty introduced by machine learning algorithms:

1.   Uncertainties in input data, i.e. images are subject to noise, outliers, and errors, but also the representation of these data introduces some form of uncertainty.

2.   Uncertainties in model parameters. Indeed, machine learning algorithms are sensitive to parameter setup.

3.   Uncertainties due to the lack of a perfect model structure.

7292 | P a g e
S e p t e m b e r   2 0 1 6
c o u n c i l   f o r   I n n o v a t i v e   R e s e a r c h
w w w . c i r w o r l d . c o m

Moreover, these approaches do not adapt to the user background, nor to the specific semantics of the information sought by him in an image retrieval system, i.e. the meaning sought by a given user in the image content with respect to his background and within (or not) a specific domain application.

B. *Automated Semantic Image Annotation*

Many image annotation approaches are based on the automatic association between low-level or mid-level visual features and semantic concepts using machine-learning techniques. Nevertheless, the only use of machine learning seems to be insufficient to bridge the well-known semantic gap problem [13], [19] and therefore to achieve efficient systems for image annotation.

The starting point for most annotation algorithms is a training set of images that have already been annotated by a human annotator with simple keywords that describe the content depicted in them as in Figure 4. Most image annotation systems follow a simple process, which is characterized by three steps:

1. Image analysis techniques are used to extract features from the image pixels, such as color, texture and shape.
2. Models that link the image features with the annotation terms are built.
3. The same feature information is extracted from unseen images in order to assess the validity of the models generated at the previous step to produce a probability value associated to each image.
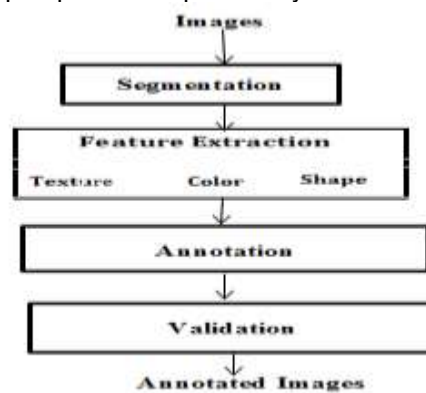

Figure 4 Typical Iamge Annotation Framework

There are three types of image annotation approaches [14], [15], [24]:

▪ **Manual Annotation Technique: -** it needs from humans to enter some descriptive keywords. The result annotated image is so accurate regardless of its problems (time consuming, expensive, difficult, subjective and inconsistent).

▪ **Semi-automatic Annotation Technique: -** it needs user interaction to provide initial query and feedback for image annotation and browsing. The result annotated image is less accurate than manual but can be improved in the interactive manner after correlation.

▪ **Automatic Annotation Technique: -** it needs no user interaction as it automatically detects and labels semantic contents of images with a set of keywords. The result annotated is the least accurate, least time consuming and most efficient.

III. Automated Image Annotation Approaches

Here are some approaches for automatic image annotation from it starts until the latest research finding. Using the earlier literature reviews, and evaluation results as guidelines, an attempt is made to outline the automatic image annotation which is combination of image analysis and statistical learning approaches. Probabilistic models are popular approaches in image retrieval including automatic image annotation. They estimate the joint probability of an image with a set of words and the probabilities of words given an images or specific region of image to annotate it. Then a ranking process is applied on words according to their probabilities. These models have the problem of high computational cost of parameter estimation such as the learning process [16], [25].

Lei Ye, Philip Ogunbona and Jianqiang Wang (2006) [6] recognized that some concepts (atomic concepts and collective concepts) can be characterized by structural visual features of images, other concepts are not directly characterized by structural visual features of images but can be described by certain concepts that can be characterized by structural visual features of images, and some concepts cannot even be derived from image visual features without further information or domain knowledge. So, the annotation problems are formulated as feature classification and concept inference problems. Images are first annotated with atomic concepts that can be characterized by visual features and the collective concepts that cannot be directly characterized by visual features but can be characterized by atomic concepts are then inferred from the

atomic concepts. Features are first mapped to atomic concepts and atomic concepts are then mapped to collective concept.

A graph learning framework (2009) [3] for image annotation, which contained the image-based graph learning and the word-based graph learning. The image-based graph learning generated the candidate annotations, and the word-based graph learning refined the candidate annotations to output the final results. To better capture the complex distribution of the image data, the NSC-based technique was proposed to construct the image-based graph .Graphs edge-weights are derived from the chain-wise statistical information instead of the traditional pairwise similarities. The word-based graph learning was performed by exploring three kinds of word correlations. One is the word co-occurrence in the training set, and the other two are derived from the web context. Extensive experiments on the Corel dataset and the web image dataset demonstrated the effectiveness of the proposed method.

Tianxia Gong, Shimiao Li, Chew Lim Tan (2010) [1] proposed a framework that represents the word-to-word relation utilizing probabilistic models. Also propose the semantic similarity language model to estimate the semantic similarity among the annotation words so that annotations that are more semantically coherent will have higher probability to be chosen to annotate the image. Probabilistic model Machine Translation Model (TM) and Cross Media Relevance Model (CMRM) [17] issues word probability conditioned on image but language model use image probability conditioned on annotation word to reduce bias.

Yunhee Shin, Youngrae Kim, Eun Yi Kim (2010) [4] suggested a method to annotate textiles using emotional concepts such as romantic, classic, cute to reduce the semantic gap between low-level features and the high-level perception of users, and develop a method that automatically predicts the emotional concepts associated with images using machine learning algorithms. The performance of the proposed method was tested with 3600 textile images, and when the results were compared with those for other methods, the proposed prediction method achieved the best annotation accuracy of above 92% thus can be used for image retrieval.

Dongjian He, Yu heng, Shirui Pan, Jinglei Tang (2010) [7] proposed a novel algorithm, Ensemble of Multiple Descriptors for Automatic Image Annotation Model (EMDAIA) for automatic image annotation based on ensemble of descriptors in order to boost annotation performance and to show one to one correspondence between image region and keyword. EMDAIA regards the annotation process as a multi-class image classification. First; each image is segmented into a collection of image regions. For each region, a variety of low-level visual descriptors are extracted. All regions are then clustered into k categories with each cluster associated with an annotation keyword. Moreover, for an unlabeled instance, distance between this instance and each cluster center is measured and the nearest category's keyword is chosen to annotate it. But accuracy is dependent on selection of segmentation algorithm.

Ran Li, YaFei Zhang, Zining Lu, Yu (2010) [12] proposed a novel approach of multi-label image annotation for image retrieval based on annotated keywords. A novel annotation refinement approach based on PageRank is also proposed to further improve retrieval performance. Multi-label annotation contains two main stages: training and annotation. At training stage, considering different importance of features and removing redundancy, a bi-coded genetic algorithm is employed to select optimal feature subsets and corresponding optimal weights for every pair of classes in training set. At annotation stage, after unlabeled image is segmented, a set of pre-trained classifiers are used to vote and annotate each region, the final label of image is merged through all the region labels. Annotation refinement based on PageRank is employed to rank the candidate annotations and to deselect irrelevant labels which lowers down there recall. Therefore require more accurate refinement algorithm.

Rami AlbatalL, Philippe mulhem Yves chiaramella proposed model (2011) [2] where the regions of Interest (ROI) are successfully used in automatic image annotation through Bag of Visual Words (BoVW) Different models. The obtained results indicate that this method outperforms (+6.2 % of MAP) the CBOVW method. These results encourage us to do further analysis on the topological Visual Phrases in order to find interesting patterns for object classes.

S. Hamid Amiri, Mansour Jamzad (2011) [8] proposed an annotation approach which follows the ALIPR structure. To describe the image contents, authors proposed an approach which extracts two discrete distributions as signatures for color and texture. These signatures are determined by applying clustering algorithms to the color and texture content of images. Major advantage of the signature extraction is that the number of segments for color and texture contents is determined automatically. The similarity of two nodes is defined based on the Mallows distance which provides more robust clusters. The required time for training the model of one concept was reduced substantially.

Golnaz Abdollahian, Murat Birinci (2011) [9] used combination of low level features and local descriptors to obtain similarity between images. To identify the large homogeneous areas in the images as "texture areas" and exclude them from the key point detection and matching process. The texture areas are matched between images based on their overall color and texture properties. The non-texture areas are handled by

local descriptor matching. The matching scores from different areas of the images are then combined to obtain an overall matching score for an image.

In automatic image annotation, systems that rely on a single classifier cannot achieve satisfactory results. Therefore, the combination of multiple classifiers for annotation has been an active research area. This kind of combination, which is also known as ensemble classifiers.

Yinjie Lei Wilson Wong (2011) [10] combined the results of individual classifiers to obtain the final annotation without consideration for the potential errors that may be introduced by each individual classifier. To address this problem, a novel AIA System based on a Two-Stage Feature Mapping (TSFM) model and a Training Set Construction (TSC) module using term extraction and image localization techniques is proposed.

Hua Wang, Heng Huang and Chris Ding (2011) [11] proposed "Image Annotation Using Bi-Relational Graph of Images and Semantic Labels", Birelational Graph (BG) approach performs random walks on the BG that comprises both image vertices and class vertices, the resulted equilibrium distributions measure the relevances not only between class and image but also between class and class. They applied the proposed approaches in automatic image annotation and semantic image retrieval tasks. Encouraging results in extensive experiments demonstrated their effectiveness.

Ning Yu, Kien A. Hua, Hao Cheng (2012) [18] proposed a novel Multi-Directional Search framework for semi-automatic annotation propagation. In this system, the user interacts with the system to provide example images and the corresponding annotations during the annotation propagation process. In each iteration, the example images are clustered and the corresponding annotations are propagated separately to each cluster: images in the local neighborhood are annotated. Furthermore, some of those images are returned to the user for further annotation. As the user marks more images, the annotation process goes into multiple directions in the feature space. The query movements can be treated as multiple path navigation. Each path could be further split based on the user's input.

Fei Shi, Fangfang Yang, Jiajun Wang (2012) [23], image annotation is formulated as a multi-class classification problem, which deals with the weak annotation problem and works with image-level ground truth training data. The relationship between low-level visual features and semantic concepts is found by supervised Bayesian learning. For each region in the test image, a posterior probability for each concept is calculated from class densities estimated from the training set and then the probability is modified using relevance with the other regions in the image. The image-level posterior probabilities are obtained by combining the regional posterior probabilities and keywords are selected according to their ranks.

IV. Experimental Work

A. Dataset

To provide a meaningful comparison with previously-reported results, the same dataset is used without any modification. This allows comparing the performance of the model in a strictly controlled manner. The Corel Image set provided by [26] is the widely used in image annotation. It separated into a training set with 4500 images and a test set with 500 images. Most of the images have 4 word annotations, while a few have 1, 2, 3 or 5. The vocabulary size of the whole set is 374 and that of the test set is 263. We note that in fact, the crucial vocabulary size is that of the training set since no other words are accessible for the auto annotation process. The vocabulary size of the Corel training set is 371. It will be shown later that the simple color structure descriptor (CSD)-based propagation method achieves good results, compared with some state-of-the-art methods for this image set [27].

B. Precision and Recall

Precision and recall which are the most popular metrics for comparing CBIR, are also widely used for evaluating the effectiveness of automatic image annotation approaches. Precision is defined as the ratio of the number of words that correctly retrieved to the total number of words retrieved in every image search. While recall is the ratio of the number of words that retrieved correctly to the number of words. The Mean Per-word Precision and Recall and Keyword Number with Recall>0, as used by previous researchers [26], [27], [28], [29] are adopted for evaluating annotation effectiveness. Per-word precision is defined as the number of images correctly annotated with a given word, divided by the total number of images annotated with this word. Per-word recall is defined as the number of images correctly annotated with a given word, divided by the total number of images having this word in its ground-truth or manual annotations. Per-word precision and recall values are averaged over the set of test words to generate the mean per-word precision and recall. A keyword has recall>0 if it is predicted correctly once or more, otherwise not. We also introduce Mean Per-image Precision and Recall and Cumulative Correct Annotations for evaluation. Per-image precision is the number of correctly predicted words for a given image divided by the total number of words predicted for that image, and per-image recall is the number of correctly predicted words divided by the number of manual annotations for that image. Per-image precision and recall are then averaged over all the test images to get the mean per-image precision and recall. Cumulative Correct Annotations is the total number of correct annotations.

V.   Results & Discussion

After applying the previously described annotation algorithms to the Corel set and predict 5 words for each test image. Table I compares the Color Structure Descriptor-Propagation Model (CSD-Prop), Singular value decomposition-cosine Model (Svd-Cos) and Color Structure Descriptor-Support Vector Machine Model (CSD-SVM) methods with the results of some state-of-the-art methods taken from the literature when the Corel training set is trained to annotate the Corel test set. These methods are the Machine Translation model [26], the CRM model [28] and the Multiple Bernoulli Relevance Model (MBRM model) [29].

It is interesting to note in Table 1 that CSD-Prop method achieves results almost as good as the best results from the more advanced methods. CSD-SVM method performs reasonably well in the experiments when compared with the other methods considered. Although it gets a slightly lower number of words with recall > 0 than the CSD-Prop method, overall the CSD-SVM achieves better results than CSD-Prop in view of the higher precision and recall measures.

Models	MT	CRM	MBRM	CSD-Prop	Svd-Cos	CSD-SVM
Words with recall >0	49	107	122	130	102	127
Results on 49 best words						
Mean Per-Word Recall	0.34	0.70	0.78	0.80	0.59	0.84
Mean Per-Word Precision	0.20	0.59	0.74	0.58	0.51	0.74
Results on all 263 words						
Mean Per-Word Recall	0.04	0.19	0.25	0.27	0.15	0.28
Mean Per-Word Precision	0.06	0.16	0.24	0.20	0.15	0.25

Table 1 Comparison between CSD-PROP, SVDCOS, CSD-SVD and some other state – of – art methods using corel images [27].

VI.   Conclusion

This paper overviewed the content-based image retrieval as a part of indexing and retrieval of images. Also, it mentioned detailed literature about various image annotation algorithms and the main problem of semantic gap. These various image annotation techniques have their own advantages and disadvantages. Results of various image auto-annotation methods that have been used to annotate the Corel Dataset have been discussed. The three image auto-annotation methods, CSD-Prop, Svd-Cos and CSD-SVM, have been used to annotate the Corel test set. Through the experiments described in this paper, some issues about the data-sets for image annotation have been demonstrated and discussed. Firstly, we show how the simple propagation method CSD-Prop achieves fairly good results on the Corel set due to the global similarity between images in the training and test sets. So the effectiveness of annotation evaluation may be affected by this cons of the Corel data sets especially it is very popular in image annotation experiments. Secondly, Because of the redundancy information in the training set, the Corel test images can still be annotated well even when only 25% of the training information is used. As a result it is needed to choose the best sub-set for training for computational efficiency. Training set reduction techniques are of potential use for reducing the size of training sets and simultaneously filtering out the noise. New method of image annotation can be found based on this literature.

VII.   References

7296 | P a g e
September 2016
c o u n c i l f o r I n n o v a t i v e R e s e a r c h
w w w . c i r w o r l d . c o m

[1] Gong, Tianxia, Shimiao Li, and Chew Lim Tan. "A Semantic Similarity Language Model to Improve Automatic Image Annotation." *Proceedings of the 2010 22nd IEEE International Conference on Tools with Artificial Intelligence*, Vol 01. IEEE Computer Society, 2010.

[2] Albatal, Rami, Philippe Mulhem, and Yves Chiaramella. "A new ROI grouping schema for automatic image annotation." *Multimedia and Expo (ICME),* 2011*, IEEE International Conference on. IEEE*, 2011.

[3] JingLiua,,Mingjing Lib, Qingshan Liua, Hanqing Lua, Songde Ma, *Image Annotation via graph learning*, Pattern Recognition 42, 2009, pp. 218-228.

[4] Shin, Yunhee, Youngrae Kim, and Eun Yi Kim. "*Automatic textile image annotation by predicting emotional concepts from visual features*" Image and Vision Computing, vol.3, pp.526-537,2010.

[5] Syaifulnizam Abd Manal and Md Jan Nordin "Review on statistical approaches for automatic image annotation, 2009 international conference on electrical engineering and informatics, IEEE, 2009.

[6] Ye, Lei, Philip Ogunbona, and Jianqiang Wang. "*Image content annotation based on visual features*" Multimedia, 2006. ISM'06. Eighth IEEE International Symposium on. IEEE, 2006.

[7] He, Dongjian, et al. "*Ensemble of multiple descriptors for automatic image annotation*." *Image and Signal Processing (CISP), 2010 3rd International Congress on.* Vol. 4. IEEE, 2010.

[8] Hamid ansari, Mansour Jamzad "*Large-Scale Image Annotation using Prototype-based Models*", 7th International Symposium on Image and Signal Processing and Analysis (ISPA), 2011.

[9] Abdollahian, Golnaz, et al. "*A region-dependent image matching method for image and video annotation*." *Content-Based Multimedia Indexing (CBMI), 2011 9th International Workshop on.* IEEE, 2011.

[10] Lei, Yinjie, et al. "*Integrating visual classifier ensemble with term extraction for Automatic Image Annotation*." *Industrial Electronics and Applications (ICIEA), 6th IEEE Conference on.* IEEE, 2011.

[11] Hua Wang, Heng Huang and Chris Ding, "*Image Annotation Using Bi-Relational Graph of Images and Semantic Labels*", pp. 126–139, 2011.

[12] Ran Li, YaFei Zhang, Zining Lu, Jianjiang Lu, Yulong Tian "Techniqu of Image Retrieval Based on Multi-label Image Annotation", *2nd International Conference on MultiMedia and Information Technology*, 2010.

[13] Hichem Bannour, "Building and Using Knowledge Models for Semantic Image Annotation", Ph.D. dissertation, Ecole Centrale Paris, MAS Laboratry, March, 2013.

[14] L. Wenyin, S. Dumais, Y. Sun, H. J. Zhang, M. Czerwinski and B.Field, "Semi-Automatic Image Annotation", *8th IFIP T.C 13 Conference on Human-Computer Interaction*, pp. 326-333, 2002.

[15] C. F. Tsai and C. Hung, "*Automatically Annotating Images with Keywords: A Review of Image Annotation Systems*" Recent Patents on Computer Science, Vol. 1, pp. 55-68, Jan., 2008.

[16] A. Han bury, "*A Survey of Methods for Image Annotation*," J. Vis.Lang. Computer, vol. 19, pp. 617-627, Oct. 2008.

[17] J. Jeon, V. Lavrenko and R. Manmatha, "Automatic Image Annotation and Retrieval Using Cross-Media Relevance Model," in Proc. 26th annual international ACM SIGIR, 2003.

[18] Ning Yu, Kien A. Hua, Hao Cheng," *A Multi-Directional Search technique for image annotation propagation*", J. Vis. Commun. ImagaR.23, pp. 237–244, 2012.

[19] T. Jiayu, "Automatic Image Annotation and Object Detection," Ph.D. dissertation, Southampton Univ., United Kingdom, May 2008.

[20] Andreas Walter, and Gabor Nagypal, " The combination techniques for automatic semantic image annotation generation in the imagination application", in Proc. 4th Internet imaging , Vol. SPIE 5304. 2008.

[21] Feichao Wang, "*A Survey on automatic image annotation and trends of the new age"*, Elsevier, 2011.

[22] Datta, Ritendra, et al. "*Image retrieval: Ideas, influences, and trends of the new age*." ACM Computing Surveys (CSUR), Vol. 40, No.2, 2008.

 [23] Fei Shi, Fangfang Yang, Jiajun Wang," Supervised Semantic Image Annotation Using Region Relevance", International Conference On Medical Physics and Biomedical Engineering, 2012.

[24] Shalini K.Kharkate, Nitin J.Janwe, "*Automatic Image Annotation: A Review*", the International Journal of Computer Science & Applications (TIJCSA), Vol. 1, No. 12, 2013.

[25] T. Sumathi, C.Lakshmi Devasena, and M.Hemalatha, "*An Overview of Automated Image Annotation Approaches*", International Journal of Research and Reviews in Information Sciences, vol. 1, No. 1, 2011.

[26] P. Duygulu, K. Barnard, J. de Freitas, and D. Forsyth., "Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary," 7th European Conference on Computer Vision, Copenhagen, Denmark, 2002.

[27] Jiayu Tang, Paul H. Lewis, "*A Study of Quality Issues for Image Auto-Annotation with the Corel Data-Set*", IEEE transactions on circuits and systems for video technology, Vol. 17, No.3, 2007.

[28] Jeon, Jiwoon, Victor Lavrenko, and Raghavan Manmatha. "Automatic image annotation and retrieval using cross-media relevance models." *Proc. 26th annual international ACM SIGIR conference on Research and development in informaion retrieval. ACM, 2003.*

[29] S. L. Feng, R. Manmatha, and V. Lavrenko., "Multiple Bernoulli relevance models for image and video annotation," *Proc. International Conference on Pattern Recognition (CVPR 2004), vol. 2, 2004.*

[30] Rouw R., Kosslyn SM., Hamel R., "*Detecting high-level and low-level properties in visual images and visual percepts*", Cognition, Vol. 63, No. 2, 1997.