International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 02 Issue: 07 | Oct-2015 www.irjet.net p-ISSN: 2395-0072

Automatic Image Annotation and Retrieval Using Contextual Information

Pranay Tiwari[1], P.M.Kamde[2]

[1]Student, Computer Engineering Department, Sinhgad College of Engineering, Pune, Maharashtra, India
[2]Professor, Computer Engineering Department, Sinhgad College of Engineering, Pune, Maharashtra, India

---

Abstract - With the advent of time, the number of images being captured and shared online has grown exponentially. The images which are captured are later accessed for the purpose of searching, classification and retrieval operation. Hence these images must be labeled with appropriate words, phrases or keywords so that the requisite operation can be performed efficiently. Automatic Image Tagging is such a technique which associates an appropriate keyword from a given set of words or phrases based on the relevance to the content of the image. Web images consist of valuable contextual information located in nearby region or within the image itself. This contextual information is nothing but the information related to or in context of image that can be used for indexing the images and also in image retrieval system. The current existing system involves human interference and is time consuming besides inconsistency. The current existing system also deals with the problem of detecting overlapped contextual information accurately which leads to incorrect assigning of contextual information and also decreases the accuracy and efficiency of the system. All this drawbacks of existing system are removed in proposed system by automatic extraction of contextual information and by text processing. This system can be helpful in areas such as webpage segmentation and image retrieval system. The system uses images or web pages as input from which contextual information can be extracted. This extracted information then is used for image annotation and is used for retrieval purpose.

Keywords: *Automatic Image Tagging, Web Images, Contextual Information, Image Retrieval System, Annotation.*

1. INTRODUCTION

A picture is a resemblance of past memories which is cherished by every individual all their life. Over the years the numbers of pictures being captured and shared have grown exponentially. There are several factors responsible for this growth. Firstly, in present days the digital cameras allow people to capture, edit, store and share high quality images with great ease compared to the old film cameras. Secondly, the availability of low cost of memory and hard disk drives. Thirdly, the popularity of social networking sites like Facebook, MySpace have given the user an additional interest to share photos online with their friends across the globe.

With this rapid growth, arises the need to perform effective manipulation (like searching, retrieval etc...) on images. Several search engines retrieve relevant images by text-based searching without using any content information. However, recent research shows that there is a semantic gap between content based image retrieval and image semantics understandable by humans. Semantic gap can be described 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" [11]. As a result, research in this area has shifted to bridge the semantic gap between low level image features and high level semantics. Thus, assigning relevant keywords is significant and can improve the image search quality. This is known as Image Annotation. It is defined as technique of assigning semantically relevant keywords to an image.

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 image retrieval systems, ABIR is considered more practical. Consequently, textual information should play a central role in visual information retrieval.

Tags are Meta data used to describe a piece of data like a web page, a digital photo, or another type of digital document. It can be defined as a non- hierarchical keyword or term assigned to a piece of information. The term "tagging" is used in context of organizing digital photos. Adobe's Photoshop Album digital photo organizer software brought the tagging concept to the mainstream for digital photography, and the popular online photo sharing service Flickr also helped to spur the trend.

Automatic image tagging can be done based on the visual content of the image or by using contextual information. Contextual information is nothing but the text related to or in context of image. This information is extracted from various sources related to the image such as title, ALT and

text. Contextual information is taken from web page of the image. The contextual information is rich with high-level semantic concepts. It contains both indirect and direct information about the image (i.e. immediate objects, attributes, events, high level knowledge, etc.)

The main objective of Automatic Image Annotation technique is to bridge the semantic gap between the image retrieval and the image semantics understandable by humans. Therefore, such information can prove very useful for annotating the image and also in image retrieval system.

2. Related Work

Nowadays image annotation & retrieval have been a very popular area for research. Many researchers were attracted by the benefits of Web image context in the past. As a result, a variety of context extraction methods, ranging from simple heuristics-based approaches to complex DOM and vision-based extractors have been proposed. In image retrieval and tagging, text annotation act as an important role. Approach [9], [7] uses contextual knowledge for automatic image annotation, whereas [9] uses CMRM techniques and [7] uses DOM based technique. While this technique are simple and fast but are prone to errors, i.e. when the text is overlapping.

Chee et. al [8] and [4] uses text mining technique for automatic image annotation. The selection of appropriate tag is implemented using the classification logic of the text mining algorithm that assigns the given set of keywords or tags to some predefined classes. In this way the most relevant tags can be assigned to the given image.

A.Hanbury [3] and [10] discusses various approaches and algorithms that can be used for image annotation. A detailed literature gap is presented based on the literature that helps to find a new method of image annotation. Approach [5] discusses various techniques for extracting contextual information and a novel approach called as WISE for image annotation that uses contextual information to tag the images.

Daniel et. al [6] used five re-ranking algorithm and rank aggregation algorithm in order to exploit contextual information for improving the effectiveness of CBIR. Re-ranking algorithm exploits contextual information encoded in the relationship among collection images, while the rank aggregation approaches are used to combine result produced by different image descriptors.

Chan et. al [1] discussed a novel approach for image retrieval by using contextual information. The author showed that how contextual information can prove beneficial in image retrieval system and can be used for annotating the image. D.Tian [2] propose a system RIA (Refining Image Annotation) that focuses on the latest development in image retrieval and provided a comprehensive survey on refining image annotation techniques. There are various other approaches [12], [13], [14] that describe various techniques for image annotation and retrieval.

TABLE I: Comparison of Image Annotation Technique

Technique	Advantages	Disadvantages
Manual	Reliable and Accuracy in Extracting semantic Information at several levels.	Tedious, Requires a lot of time and efforts, Costly
Semi-automatic	Efficient than manual annotation More accurate , useful for dynamic Database	Requires User interfaces refinements to improve the feedback process
Automatic	Speed (saves time)	Less Reliable than Manual, More error-prone, Produces more general (less detailed) annotation as compared to manual method, Less accurate than semi-automatic annotation.

Proposed System

The proposed method is a novel automatic image annotation system along with image retrieval, basically designed for tagging the web images and retrieving them. It extracts contextual information from web pages. Then this information is text processed to extract the important keywords from the document. Then each word is assigned a weight depending on its frequency in document. Then the terms with lowest weights are selected and allotted to the corresponding image.

The architecture of system is shown in Fig 1 it consists of four modules (a) Contextual Information Extraction (b) Text Processing (c) Term weighting (d) Image Retrieval. In

the first module contextual information is extracted from the web page. The extracted information is in the form of paragraph of words (document). The noise such as stop words and irrelevant words are removed with the help of second module. Third module assigns weight to the words from the second module depending on their frequency in the document. The words with lowest weights are assigned to the image (i.e. image is annoted with this words). In fourth module, annoted images are stored in database and further they can be retrieved from the database.

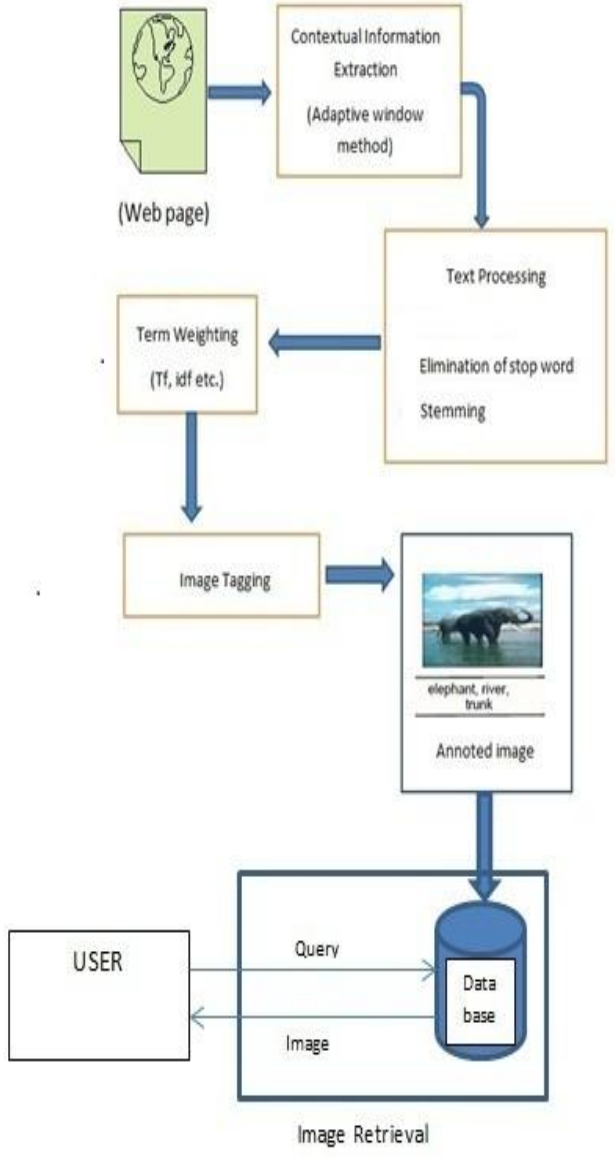

Fig1: Architecture of System

2.1. Information Extraction Module

Contextual information is the information related to or in context of image situated near it. There are various locations on webpage form where we can extract this information. Some of the locations are mentioned below:

- Image file name text that is located in the SRC attribute of the image tag, < IMG >.

- Image ALT text that is located in the ALT attribute of the image tag, < IMG >,

- Page title text between the < TITLE > tag pair in the header of an HTML document.

- Hyper link (link) text between the anchor tags < A > and </A>.

- Uniform Resource Locator (URL).

-Web addresses that are located either in the HREF attribute value of the anchor tag <A> or in the SRC attribute value of the image tag < IMG >.

- Surrounding text appearing before and after the image, ranging from 10 words, 20 words, a paragraph or selection of words to entire page.

There are different techniques or methods to extract contextual information form source, such as Window-based context extraction, Structure-based Wrappers and Context by Page segmentation.

A fast and therefore commonly used method to estimate the context of web images is the window-based method. There are two methods, one is N-Term Window method and another is Adaptive Window method. The N-Term Window method suffers from the fixed-length of the window and thus cannot deal with varying context sizes among web pages. Therefore, we use adaptive window technique for extraction of information.

1) ***Adaptive Window Method****:* The window based extraction technique mentioned above suffers from the fixed-length problem of the window and thus cannot deal with varying context sizes among webpages. To overcome this problem, [17] introduced a more dynamic method, which involves structural information contained in HTML. It was recognized that there are certain structural tags that are used by web designers in order to separate the contents. Some of the separator tags that **{<hr>,
, <p>, <table>, <tbody>, <tr> and <td>}**. These tags can be used as clues for context borders and better precision for the context can be estimated. The extraction algorithm needs an image and web document for input as the size of window is 64 at maximum.

Algorithm 1: ***Adaptive Window Method***

Data: Web document *d,* image *i*

Result: Set *T* of terms surrounding the image

```
1.  S ← get Tags and Terms (d);
2.  i ← index Of (I , S);
3.  t = 0
4.  k = i+1
5.  While t < 32 do
6.      if S [k] instance of TERM then
7.          T.add ( S[k ];
8.          t=t+1;
9.      else
10.         if S[k] is SEPRATOR then
11.             terminate loop;
12.         end
13.     end
14.     k = k+1;
15. end
16. t = 0;
17. k = i – 1;
18. while t < 32 do
19.     if S[k] instance Of TERM then
20.         T.add (S[k]);
21.         t= t+1;
22.     else
23.         if S[k] is SEPRATOR then
24.             terminate loop;
25.         end
26.     end
27.     k=k -1;
28. end
```

The first difference as compared to *n-term window* extractor is that all tags are additionally defined while initialization of the sequence *S*, tags are ordered in *S* by their position in HTML codes. In second line, index *i* of image *I* in Sequence *S* is estimated. Line 3-4 shows the initialization of the index and the terms counter for traversing *S*. The first while loop from line5-14 iterates over the element of sequence *S* that is after the image and checks whether they are terms or separator tags. If the current element is the term, the element is added to the context set *T*. The loop comes to an end, when the 32 terms are collected or either the element is a separator tag. The loop from line 17-28 is repeated in same manner for elements that are before the image in sequence *S*.

This method has been successfully applied by [16] and also [15] has applied a variant for this method in image annotation application.

Time complexity for the all the variants of this method remains linear, as the tags are included in sequence *S*, it depends on both number of tags and text elements.

3.2. Text Processing

The extracted contextual information is to be transformed into set of words that can be used for indexing. To convert the information we need to process it.

1) Elimination of stop words: Words are the most basic objects of text that describes information in a natural language. However, the information value can vary highly from one word to another and thus has an impact on the application using it. Stop words are those words that are frequently used in a language for sentence construction. This words are insignificant and do not carry any information content. Typically articles, preposition, conjunctions and some frequently occurring pronouns are considered as stop words. For example, the following words might be considered as stop words in English language:

a, about, an, are, as, at, by, for, from, how, in, is,...

It is necessary to filter out the stop words before passing it to other processing steps such as indexing.

2) Stemming: In a natural language a word has different syntactical forms. Context of the sentence defines the usage of the words. Pronouns, nouns and adjectives are declined depending on number, case and gender while verbs are conjugated in terms of tense or person. For example, the word stronger and strongest has the same ground form strong. However, in a retrieval scenario where a document contains the first flexive and query contains the second, the document would not appear in the result set, since the words do not match exactly. Stemming tries to avoid such problem by reducing different flexive of a word to a common root (stem).

A stem refers to a segment of word which is obtained after cutting its suffixes or prefixes. For example the word computing, computes and computation are reduced to compute which is an artificial word. In such a way many words can be matched and thus recall can be improved and size of the index can also be further reduced. There are various stemming algorithm for different languages. The most commonly used stemmer algorithm for English is the Porters stemming algorithm, which is based on a set of substitution rules.

3.3. Term Weighting

In this module weights are assigned to each term depending on their frequency in the document.

1) Term Frequency (tf): It is a method for assigning weight to terms in a document. It is the most simplest

approach for calculating weight. Let t be term in a document d, then weight of t in d can be calculated as***: tf = (frequency of term t in document d) / (total number of terms in document d).*** This term frequency suffers from a critical problem, that all terms are considered equally important when it comes to assessing relevancy on a query. Also, if some important word is there and it has fewer occurrences in the document than weight assigned to the word would be less and thus that word can get neglected.

2) Inverse document frequency (idf): It is also one of the techniques for assigning weight to the terms in documents. In this technique the words having the lowest occurrence in the documents are assigned higher weights. If *idf* is calculated for term *t* in document *d*, then ***idf = log_e (total number of documents / number of documents having term t)***. It also suffers from drawback that if some important word have high occurrence then lesser weight would be assigned to it.

3) Term frequency and inverse document frequency (tf-idf): This method removes drawback of both the techniques that are described above. It is developed by combining tem frequency and inverse document frequency. For term t in a document d the tf-idf can be; **tf-idf = (term frequency) * (inverse document frequency)**. It is more accurate and better method then the above two method. This method will be used for calculating weight of terms in our proposed system.

3.4. Image Annotation and Retrieval

After weighting terms, the terms with lowest weight are used for annotating the images (generally 5 terms with lowest weight are selected for annotation). The images are then tagged with these terms. After image annotation, the images are stored in database along with their annoted terms, so that the stored images can be retrieved just by giving the annoted term as query to the system.

4. MATHEMATICAL MODEL

Set Theory Analysis

1. Let S be the system
   $S = \{I, O, F, F_s, F_l, \Phi\}$

2. Identify the inputs
   $I = \{W_i\}$
   Where:-
   W_i = Web page

3. Identify the Outputs
   Let O be the set of outputs
   $O = \{I_a\}$
   Where:-
   I_a = Annoted Images
   I_r = Retrieved Images

4. Identify set of Function
   Let F be the set of Functions
   $F = \{F_1, F_2, F_3, F_4\}$
   Where:-
   F1 = Information Extraction Module
   F2 = Text Processing Module
   F3 = Term Weighting
   F4 = Image Annotation and Retrieval

5. Final State
   $F_s = \{F_a, F_r\}$
   Where:-
   F_a = Successful annotation of images with relevant terms.
   F_r = Successful retrieval of images.

6. Failure case
   Let Fl be the set of Failure case
   $F_l = \{F_d\}$
   Where:-
   F_d = image annoted with irrelevant terms or non annoted image.

7. Constraints
   Let Φ be the set of constraints
   $\Phi = \{C_l\}$
   Where:-

   C_l = Web page must contain images and information given in web page source code related to image is relevant.

5. RESULT AND DISCUSSION

In this section we present evaluation of proposed system .Firstly, in section 5.1 we define evaluation parameters. In section 5.2 experimental results.

5.1. Evaluation Framework

The functionality of the proposed framework has been tested with the common web image context extraction methods, namely, the N-Terms window (NT) extractor, the paragraph (PAR) extractor, the VIPS-based extractor (VIPS), the Monash (MON) extractor, and – as a baseline to clarify the performance gain of using extraction methods – the Full-Text (FULL) extractor.

The Full Text extractor does not contain the complete web document but only its plain text without any code. For the N-Terms window

algorithm, we have implemented the standard fixed-window version as presented in Algorithm 1 with a window size of 64 terms. The VIPS algorithm is a web page segmentation method that divides a web page into blocks. After that, the extraction-by-page-segmentation method can be applied.

All extraction methods were executed and evaluated in the same environment under the same conditions in respect to input and output specification and document preprocessing.

5.2. Time Analysis

We have measured the average running time needed to process a single page for each extraction method on a system with a Core 2 Duo processor at 3.2 GHz and 4 GB RAM memory. Since the input documents are to be taken from internet, loading time for the document source is included. The results of the time analysis is depicted in Figure 5.1.The first observation is that the running time of N-Term window method is significantly less as compared to the other methods.

The more interesting observation is that all extraction methods are slower than the baseline method. This encourages us to replace the full text extractor by proposed methods in real world applications that need image descriptors and used full text until now.

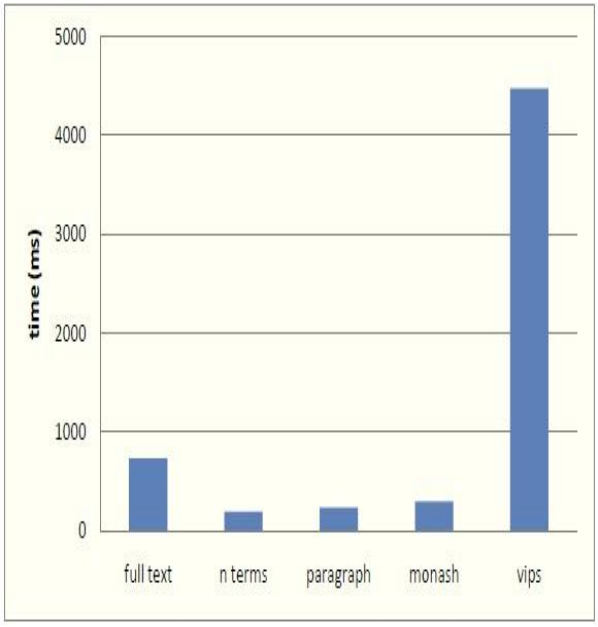

Fig 2: Average running times for different web image context methods

6. EXPERIMENTAL RESULTS

In this section, performance of proposed system is tested by using different window size for term extraction. For each web page, we have compared results by changing the window size of the proposed algorithm. We also evaluate our system by using different web pages with different number of images as dataset. All the experiments are performed on a single core of our common PC with Core2 Quad CPU (3.2 GHz) and 4 GB memory.

6.1. Changing the Window Size

As stated earlier, we are evaluating the proposed system by changing the window size of the proposed algorithm. Different window size is set (i.e. 30, 40, 64 and 80) for same web page and the time required for annotation is calculated along with the relevant terms with which images are annotated. The result of evaluation is shown in figure below.

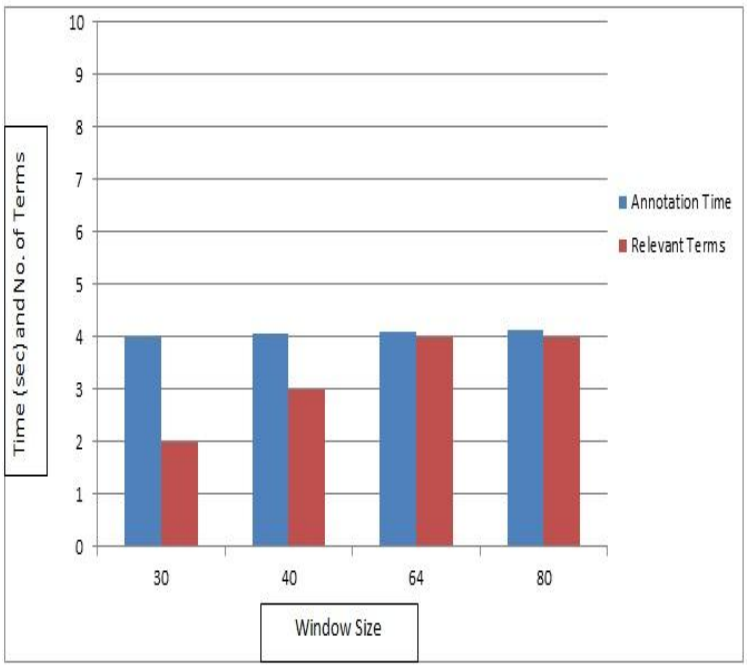

Fig 3: Average time and No. of terms for different window size

From the figure 3, we can easily state that, both the time required and number of relevant terms is directly proportional to the window size. That means, if window size is increased then the time required for annotation also increases (but increase in time is very minimal) and so as the number of relevant terms. We also found that for window size 64 and 80 the number of relevant terms is equal.

6.2. Using web pages with different number of images

We have also evaluated our system with different web pages having different number of images. While evaluating the system we calculated the time required by the system to extract the terms and annotate the images. For different web pages we got different results. The result for different web pages is shown in figure below.

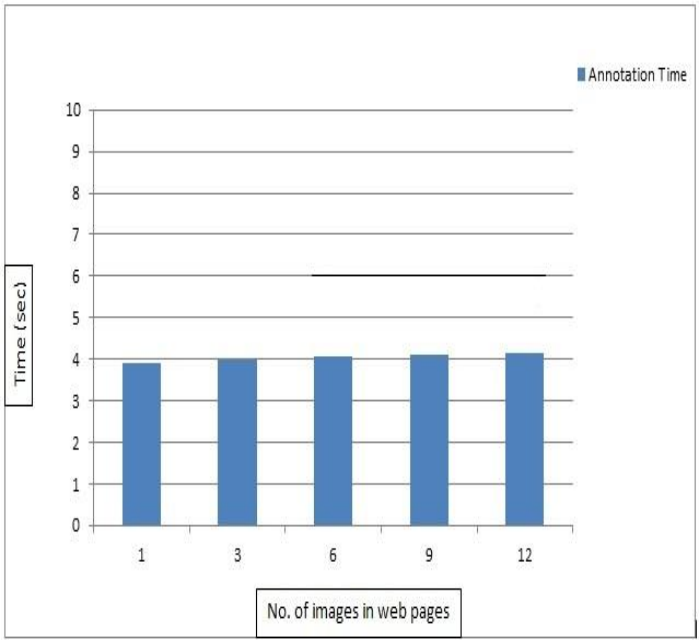

Fig 4: Average time for annotating images of web pages with different no. of images

From the figure above we come to know that as the number of images increases in web pages the time for annotating the images also increases. But it is also observed, that the increase in time between two web pages with different number of images is not more than few milliseconds. Thus, it shows that algorithm used in the proposed system is robust and can perform in different conditions.

7. SCOPE OF SYSTEM

The proposed image annotation system is designed for only web images annotation purpose. The system is able to annotate the web images with the help of contextual information. The information given between the tags near image is considered to be correct and relevant to the image. For contextual information extraction, proper HTML tags are to be considered. The system is then further used as image retrieval systems for retrieving images. Images which are in pop-up or in some external links whose HTML code is not available or dynamically generated such images cannot be annoted

8. FUTURE WORK

The proposed image annotation system is very easy to extend. There are a number of different ways of image annotation by contextual information. To use this system for extracting information surrounding video will be our further research in order to annotate videos. We are also trying to extract information from web pages of other languages such as JAVA Script.

9. CONCLUSION

As conclusion, the proposed work shows that contextual information located nearby the image can be used for image annotation. It also shows that how with the help of contextual information we can overcome the drawbacks of the existing system such as semantic gap and erroneous annotation. It also overcomes the word sparseness flaw that occur in other systems while image annotation. The importance of contextual information in image annotation and retrieval is shown in the given the system. Though, there are many flaws in the usage of contextual information such as:

1) It is subjective to the author's point of view, knowledge, culture and experience.

2) Usage of such information is quiet challenging.

Despite these drawbacks, it can be observed that the contextual information remains one of the most indispensible features for image annotation system.

ACKNOWLEDGMENT

I like to thanks my guide ***Prof.P.M.Kamde*** for their valuable guidance in preparing this paper. Author also likes to thank ***Prof. C.A.Laulkar*** and ***Prof. P.R.Futane*** for their support in this work and also to the unseen authors of various articles on the internet helping author to become aware of the research currently ongoing in this field.

REFERENCES

[1]. Chee Sheen Chan, Adel Johar, Jer Lang Hong, "Contextual Information for Image Retrieval Systems", 10th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), 2013.

[2].Dong ping Tian, "A Survey of Refining Image Annotation Techniques, International Journal of Multimedia and Ubiquitous Engineering Vol.9, No.3 (2014).

[3]. Shalini K.Kharkate,Nitin J.Janwe, "Automatic Image Annotation: A Review ", The International Journal of Computer Science and Applications (TIJCSA), Feb-2013.

[4]. Sayantani Ghosh, Samir Kumar Bandy opadhyay, "A TUTORIAL REVIEW OF AUTOMATIC IMAGE TAGGING TECHNIQUE USING TEXT MINING, IJRET: International Journal of Research in Engineering and Technology, Mar- 2013.

[5]. Sadet Alcic, Leskovac, "Web Image Context Extraction: Methods and Evaluation", Google.

[6]. Daniel Carlos Guimaraes Pedronette, Ricardo da S. Torres, "Exploiting Contextual Information in Image Retrieval Tasks", Recod Lab - Institute of Computing University of Campinas (UNICAMP) Campinas, SP - Brazil, 2012.

[7]. Thomas Scholzc,Stefan Conrad, "Automatic Annotation of Web Images combined with Learning of Contextual Knowledge", Google.

[8]. Chee Wee Leong and Rada Mihalcea and Samer Hassan, "Text Mining for Automatic Image Tagging", Coling 2010: Poster Volume, pages 647655, Beijing, Aug- 2010.

[9]. Yong Wang, TaoMei, Shaogang Gong, Xian-Sheng Hua, "Combining global, regional and contextual features for automatic image annotation", journal homepage: www.elsevier.com/locate/pr, 2007

[10]. Allan Hanbury, "A Survey of Methods for Image Annotation", Pattern Recognition and Image Processing Group (PRIP), Institute of Computer-Aided Automation, Favoritenstrae9/1832, A-1040 Vienna, Austria, May-2007.

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

[12]. Monay, F., Gatica-Perez, D, "Plsa-based image auto-annotation: constraining the latent space", Multimedia 04: Proceedings of the 12th annual ACM international conference on Multimedia, New York, NY, USA, ACM Press 348351, 2014.

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

[14]. Florent Monay and Daniel GaticaPerez, "Latent Semantic Fusion Model for Image Retrieval and Annotation", Proc.16th ACM Conf. Information and Knowledge Management (CIKM), 2007

[15]. Hua Zhigang, Wang Xiang-Jun, Liu Qingshan, and Lu Hanqing. "Semantic knowledge extraction and annotation for web images". In MULTIMEDIA'05: Proceedings of the 13th annual ACM international conference on Multimedia, pages 467–470, New York, NY, USA, 2005. ACM.

[16]. Huamin Feng, Rui Shi, and Tat-Seng Chua. "A bootstrapping framework for annotating and retrieving WWW images". In Proceedings of the 12th annual ACM international conference on Multimedia, MULTIMEDIA'04, pages 960–967, New York, NY, USA, 2004. ACM.

[17]. Lexin Pan. "*Image8: an image search engine for the internet*". Honors year project report, School of Computing, National University of Singapore, April. 2003.

[18]. Vundavalli Srinivas Rao, Vasudeva Varma, "Web Image Annotation Using an Effective Term Weighting", Intelligent Text Processing and Computational Linguistics, 13th International Conference, 2012

[19]. Reena Pagare, Anita Shinde, "A Study on Image Annotation Techniques." International Journal of Computer Applications (0975 8887) Volume 37 No.6,January 2012

[20]. Chandrashekar V, Shailesh Kumar, and C V Jawahar, "Image Annotation in Presence of Noisy Labels",International Journal of Computer Vision 90(1) (2010)

[21]. Barnard, K., Duygulu, P., Forsyth, D., de Freitas, N., Blei, D., Jordan,"Matching words and pictures", Machine Learning Research3 11071135, 2003

[22]. Hui Hui Wang, Dzulkifli Mohamad, N.A. Ismail, "Approaches, Challenges and Future Direction of Image Retrieval System," *JOURNAL OF COMPUTING*, volume 2, issue 6, JUNE 2010, ISSN 2151-9617.

[23]. John R. Smith and Shih-Fu Chang. "*Visually searching the web for content"*, IEEE Multimedia, 4:12–20, July 1997.

[24]. Sadet Alcic. "Improvement of Content-based Image Retrieval Quality using Automatic Annotation", April 2007. Master Thesis.

[25]. Tatiana Almeida Souza Coelho, P´avel Pereira Calado, Lamarque Vieira Souza, Berthier Ribeiro-Neto, and Richard Muntz. *Image Retrieval Using Multiple Evidence Ranking*. IEEE Transactions on Knowledge and Data Engineering, 16(4):408–417, 2004.

[26]. Fariza Fauzi, Jer-Lang Hong, and Mohammed Belkhatir. *Webpage segmentation for extracting images and their surrounding contextual information*. In ACM Multimedia, pages 649–652, 2009.

[27]. Deepayan Chakrabarti, Ravi Kumar, and Kunal Punera. "A graph theoretic approach to webpage segmentation". In Proceeding of the 17th international conference on World Wide Web, WWW '08, pages 377– 386, New York, USA, 2008.

[28]. Parag Mulendra Joshi, Sam Liu, "Web Document Text and Images Extraction using DOM Analysis and Natural Language Processing", HP Laboratories,HPL-2009-187.

[29]. G. Vineel. Web page DOM node characterization and its application to page segmentation. In Internet Multimedia Services Architecture and Applications (IMSAA), 2009 IEEE International Conference on, pages 1 –6, 2009.

[30]. Gen Hattori, Keiichiro Hoashi, Kazunori Matsumoto, and Fumiaki Sugaya. "Robust web page segmentation for mobile terminal using content distances and page layout information". In Proceedings of the 16th international conference on World Wide Web, WWW '07, pages 361–370, New York, NY, USA, 2007. ACM.

[31]. Xiaoguang Rui, Mingjing Li, Zhiwei Li, Wei-Ying Ma, and Nenghai Yu. "Bipartite graph reinforcement model for web image annotation". In *Proceedings of the 15th international conference on Multimedia*, MULTIMEDIA'07, pages 585–594, New York, NY, USA, 2007. ACM.

[32]. Xiaofei He, Deng Cai, Ji-Rong Wen, Wei-Ying Ma, and Hong-Jiang Zhang. "Clustering and searching www images using link and page layout analysis". ACM Trans. Multimedia Comput. Commun. Appl., 3(2):10, 2007

[33]. Huamin Feng, Rui Shi, and Tat-Seng Chua. "A bootstrapping framework for annotating and retrieving WWW images". In Proceedings of the 12th annual ACM international conference on Multimedia, MULTIMEDIA'04, pages 960–967, New York, NY, USA, 2004. ACM.

[34]. Stan Sclaroff, Marco La Cascia, and Saratendu Sethi. "Unifying textual and visual cues for content-based image retrieval on the World Wide Web". Computer Vision and Image Understanding, 75(1-2):86–98, 1999.