

An Efficient Algorithm to Reduce the Semantic Gap between Image Contents and Tags

Grishma Y. Bobhate
Sinhgad Technical Education
Society's, Smt. Kashibai
Navale College of Engg
Vadgaon (BK), Pune.

Prof. Usha A. Jogalekar
Sinhgad Technical Education
Society's, Smt. Kashibai
Navale College of Engg,
Vadgaon (BK), Pune.

ABSTRACT

With the rapid growth in the new era of Internet Technology, image retrieval is an active and traditional method for searching the images by keywords or by images from the large amount of image database. As tags gives the descriptive information of an image on the web. Due to noisy nature in tags, it becomes necessary to correlate both image content and tag information for retrieval purposes. However, semantic gap is a major problem in the image processing concept. Therefore, our presented research is going to reduce the problem of semantic gap by applying techniques to extract low level features of an image such as color, texture and edge. Then, construction of a mixed graph between image and tag to perform random walk on graph for getting accurate results in an efficient way. Experimental results show the effectiveness of our approach.

General Terms

Image retrieval, Image Processing.

Keywords

Random Walk, Semantic Gap.

1. INTRODUCTION

Image retrieval [7] is the field of study related with searching and surfing the images from collection of the database. These researches exist since the 1990s and produced emerging techniques in various methods. In the internet, large numbers of images are uploaded, due to these more attraction has been developed among researchers and experts in the application areas of image processing, computer recognition, database application.

Content based-image processing [4] has been studied and explored in image processing concept. It extracts low level features of images such as color, texture, shape to retrieve the images from a large collection of images. As there is a rapid increase of digital images on the web, the automatic image annotation has become an important concept in image retrieval approach. Image annotation [9] was described as adding keywords automatically for images to represent the semantic content for the image.

Semantic gap [8] is the relationship between the information that one can extract low level features from the visual data and the human interpretation. This gap can be balanced by using high level features of information to index and retrieve the images.

Most studies in this research work [8] [9] [12] have concentrated on how to form groups of images into semantically related group or indexes the images in the database based on low-level visual features of the images. When user entered a query such as keywords, images then system will return images which are semantically relevant to that query image. Moreover, users contribute tagging to images for efficient searching which degrades the performance of retrieval system.

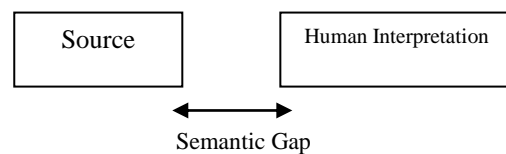


Fig 1: Structure of the relationship of the semantic gap of an image.

Generally, Tags [6] are associated with images contain valuable and meaningful information for relevant image search and retrieval processes. Specifically, it has not been yet studied properly about the task of ranking images for perfect matching with a given query tag. Therefore, image retrieval task produces complex solution for searching by tag.

By taking the consideration of the visual information of images, an image search engine has been improved because visual information balances features of images for searching purpose. Human cognition and interpretation is sensitive to this information which gives 80 % efficiency.

Content based image retrieval produce such complex solution. Hence, data for tag is an important and simple method for improving the image analysis and retrieval procedure. When users shared their photos on social networking sites with tag or comment are semantically meaningful for that images. Therefore, this paper focuses on collected images to reduce the lack of relationship between low-level image features and high-level semantic concepts by constructing mixed graph between images and tags. Then top N images are grouped and compared with limited number of relevant images to rank the results, which provides accurate approach for collection of various data and modeling the procedure.

The remainder of this paper is designated as follows. The next section describes the Proposed System Design. Section III explains the Advantages with the system. Section IV shows the experimental results. Section V concludes the paper with the future work.

2. PROPOSED SYSTEM DESIGN

The main aim of the system is to reduce the gap between image and verbal tag in web application. It also provides the user with an efficient and effective image retrieval system. When the user gives a query to the system to perform better search to find image from tag or vice-versa. It is designated as follows:

- Extraction of the universal low level feature from every image.
- Construction of the mixed graph between image-image and image-tag.
- Perform a random walk concept through the constructed graph to retrieve the results.

For performing this step, an efficient algorithm is used to retrieve the top ranked relevant results. This gives improvement in the ranking performance of the input query.

2.1 Universal Low Level Feature Extraction of Image

The techniques used for feature extraction [1] are widely used in pattern recognition, computer vision and feature detection. These techniques extract the low level feature such as color, texture, pattern, edge to provide meaningful information of the image. These features are likely efficient for computation. In this paper, extraction of four different kinds of features from an image is done as follows.

2.1.1. Color

Color [14] is an important feature for image searching and representation which is widely used in image retrieval. Color feature can be calculated using color moment: mean, variance, skewness. Color moments are measures that can be used to distinguished images based on their local features of color. When these moments are calculated, it provides a measurement for color similarity between images.

$$\text{Mean: } E_{ij} = \frac{1}{N} \sum_{j=1}^N (P_{ij}) \quad (1)$$

$$\text{Variance: } \delta_{ij} = \sqrt{\left[\frac{1}{N} \sum_{j=1}^N (E_{ij} - P_{ij})^2 \right]} \quad (2)$$

$$\text{Skewness: } \sigma_{ij} = \sqrt[3]{\left[\frac{1}{N} \sum_{j=1}^N (E_{ij} - P_{ij})^2 \right]} \quad (3)$$

2.1.2. Pattern

Pattern [17] is defined as repetitive pattern used for classification of data. Here each pixels of an image are first calculated by taking the difference between the center pixel and its neighboring pixels. Using a function $u(x)$, $u(x) = 1$ when x is greater than or equal to 0, otherwise 0. The patterns are said to be consistent if the movement between 0 and 1 are less than or equal to two. In this way, local binary pattern effectively and reliably capture the pattern information of the images.

2.1.3. Texture

Texture is another property of an image. It is a regular repetitive or visual pattern of an image. Texture analysis [10] plays a vital role in image processing and pattern recognition tasks such as remote sensing, medical imaging and robot vision and query by content in large image databases. In this section to extract texture features, multi-resolution filtering techniques such as Gabor wavelet transform to categorize texture elements by the statistical distribution of the image is introduced. The Gabor wavelet transform [10] are collection of wavelet or signals which captures energy of every wavelet at specific frequency for specific direction. Scale and orientation are transitional parameters for each wavelet. It is then applied on the image with five levels and eight orientations, which results into 40 sub images. After these three kinds of moments such as mean, variance, and skewness are calculated for each sub images.

2.1.4. Edge

An edge is change in intensity of the images. Here, first extract histogram for each image and convert each image into a grayscale image, then apply a Sobel edge detector to obtain the edge map of an image for computing histogram. The histogram of edge is divided into 36 bins of 10 degrees each. An extra bin is used to compute the total number of pixels present in the image. Thus, these feature vectors are used to represent all the universal features for each image in the dataset and normalized.

2.2 Construction of Mixed Graph between Image and Tag

When the universal low level features are extracted, an image similarity graph is constructed using k nearest neighbor graph, expo-weighted NN graph. Let n be total number of image dataset and v be the feature vectors of an images. For query image, calculate similarity measure based on nearest feature vectors of image and construct similarity graph. By using formula for similarity measure:

$$\text{Sim}(p, q) = \frac{p \cdot q}{|p||q|} \quad (4)$$

Where, p and q are two images, $\text{Sim}(p, q)$ is the similarity measure between two images p and q .

Above formula explains that if q is nearest to p then it creates an edge based on computed similarities. Here, Similarity measure is done by using Euclidean distance method. These distance methods calculates the distance between two nodes which is nearer to each other and then form a cluster which is group of nodes. In this, cluster technique is used in the form of knn graph. K is used to set the path between two nodes. It is observed that K nearest neighbor graph is not well-formed graph because if query node form an edge with second node, then it is not necessary that second node form an edge with first node. However, the nearest neighbor graph based method [2] [9] is used to efficiently build image tag dualistic graph. This graph cannot exactly understand the relationship between image and tag because it is undirected graph. To overcome this issue, this graph is converted into directed graph. Then a mixed graph is constructed by combining these two graphs: image similarity and image-tag directed graph. The given figure shows that left part denote as image node and right part dente as tag node. The edge from image to tag gives total number of times images is tagged and the edge from tag to

image gives total number of times tag has been allowed to assign to a particular image.

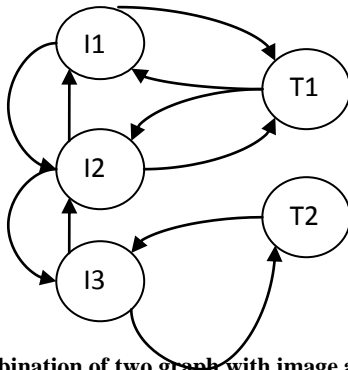


Fig 2: Combination of two graph with image and tag node

This combined graph based method is used to improve the quality of tag by performing random walk model that will be introduced further.

2.3. Perform Random Walk Method on Mixed Graph

A novel random walk model [18] is proposed on combined graph to smoothly employ the visual and tag information for image retrieval task. It can be used in many ways such as random walk on line, plane, and pixel or on the graph. It is proposed on combined image-tag graph to clearly understand the visual and tag information for the process image retrieval task to give reliable solution. Using transition probabilistic matrix random walk can be performed easily on the combined graph. To provide more flexibility on the random walk model, the parameter λ plays an important role in the transition probability matrix to define how random walk diffuses on two sub graphs: image-image and image-tag graph. When $\lambda = 1$ the random walk performs on the image-tag sub graph. On the other hand, when $\lambda = 0$ random walk will not diffuse on the image-tag sub graph, and it only diffuses on the image similarity graph. In the intermediate case, when λ is relatively large, the diffusion on the image-tag sub graph is faster than the diffusion on the image similarity sub graph. As a result, the random walk will depend more on the image-tag information. If λ is relatively small, the results will depend more on the image visual information. The random walk sums the probabilities of all paths of length t between the two nodes. It gives a measure of the volume of paths between these two sub graphs.

Algorithm:

1. Give a query and obtain n feature vectors for all nodes in the database.
2. Perform a first t -steps random walk and get a new vector value.
3. Rank the relevant list of the nodes in the highest order.
4. Make a new feature vector value according to the entries with highest rank=1 otherwise 0.
5. Again perform a new t -steps random walk and get the retrieve results.

2.4. Retrieval by Relevance Feedback Rules

Relevance feedback [15] is a feature of some information retrieval systems. For a given query, the system first retrieves a list of ranked images according to accurate similar matching measures. Then, the user marks the retrieved images as relevant to the query or not relevant. The system will refine the retrieval results based on the feedback method and present a new list of images to the user. It is a specific method to balance the gap between low-level features and semantic concepts in image retrieval process. This method satisfies the concept of retrieval procedure and greatly increases user interpretation.

Pseudo relevance feedback [5] [16] is an important technique for improving the accuracy of retrieval method. The proposed system takes the basic idea of pseudo-relevance feedback: assume a small number of top-ranked images in the initial retrieval results that are relevant and select from these results related to the query. Thus it leads to improve the representation of query by expanding query. Also, it provides an efficient method to automatic analysis and reduces complexity problem. This improves the retrieval performance and exploits the strategy of the feedback.

3. ADVANTAGES OF THE PROPOSED SYSTEM:

The Proposed system design will provide more user interaction to improve retrieval performance. These gives advance in computer vision domain for encouraging image processing concept. Hence, this will remove image annotation problems noisy tags. Thus it will lead to improve image quality and tags. Here the disadvantage is that we use machine learning techniques which require more memorization and thus increases cost for computation.

Applications:

1. Content Based Image Retrieval: CBIR stands for content-based image retrieval means images are retrieved based on image content such as color, texture. When user gives an image, find images which are relevant based on visual information of an image. The relevant documents should be ranked highly based on similarity measurement.

2. Image-to-Tag Searching: This is also called as image annotation. Automated image annotation is an active and broad research topic in computer vision and pattern recognition. When user gives an image, it will search for relevant tag based on semantic relation between image and tag.

3. Text Based Image Retrieval: When user gives a tag as a query, search a list of images which relevantly matched and associated with this tag. This is more likely called as text based image retrieval.

4. Tag-to-Tag Searching: when user gives a tag as query, it will search for most related tag that are associated with it. This is also known as tag recommendation problem.

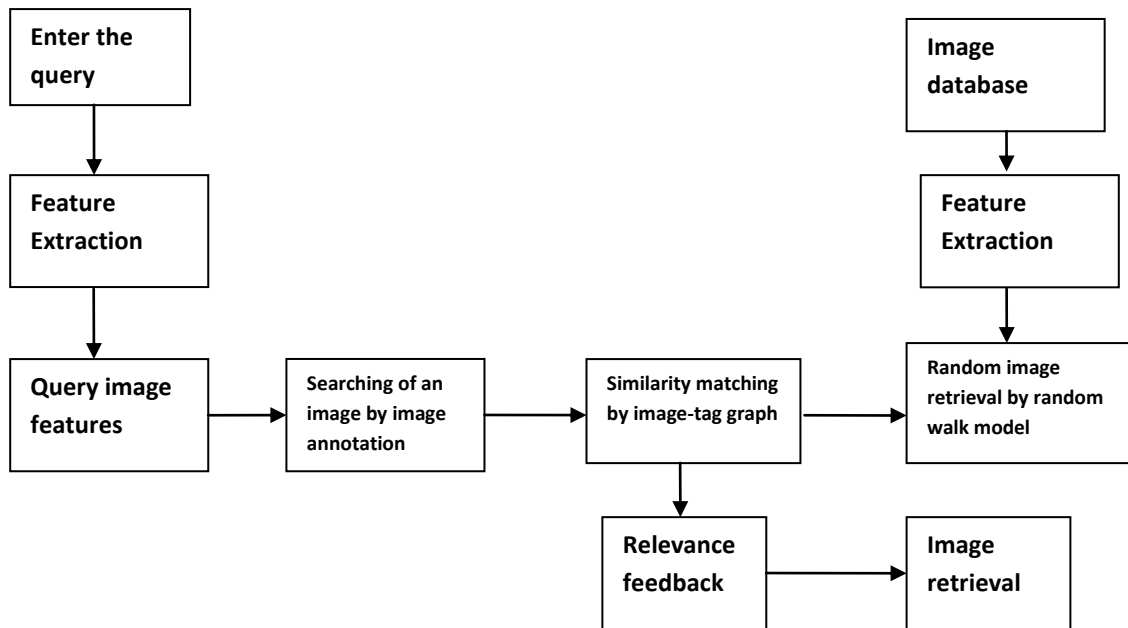


Fig 3: The Proposed system Design

4. EXPERIMENTAL RESULT

In this section, the proposed framework is performed with the help of Flickr dataset on content-based image retrieval, tag-based image retrieval. Flickr is an image hosting Web site and popular image search engine tool. It creates famous sites for users to share their personal photographs, tag photographs, and communicate with other users. Hence, Flickr is an ideal source for the enquiring and efficient image-retrieval related research.

In the content-based image retrieval (CBIR) tasks, above method calculates the low level features of the images such as color, texture, pattern and edges. Before compiling, every image is resized up to 120 by 120. Further feature extraction technique obtains the values of appropriate features for an image. For color feature extraction, three different color moments are calculated for query image. Pattern feature is used for mapping function on center pixel of an image and edge detector obtains 37 bins of a histogram to count the total number of pixel. Gabor wavelet transforms first partitioned the image into 64 by 64 pixel computes for five scales and eight orientations for each subpart. At each scales, various filtering techniques is employed for scales for calculating all the frequencies of an image. As these values are calculated for all orientation it improves the process of texture discrimination. It is clear that Gabor wavelet texture is more effective for texture classification. Hence this method can be widely used for texture analysis and segmentation.

Then mixed graph method is applied for constructing image-tag bipartite graph. This method is used for synchronization between images and tags, after that start the random walk at an image node. In addition to the fusion parameter λ , set another two parameters: the parameter k for building NN image similarity graph and the parameter t for the random walk steps. After a t -step random walk, retrieve the top-ranked images as the retrieval results. In this experimental result set $k=40$ means that for every image has 40 most

similar neighbors. Hence, the out degree for every image node in the image similarity sub graph is 40 and $t=10$, this parameter t determines the resolutions of the Markov random walk. When the input image is given to the system, retrieve top ranked images with parameter λ value. It has been observed that all this images are semantically relevant to the input image.



Fig 4: An input image given to the system



Fig 6: Images are retrieved by using fusion by Random walk method

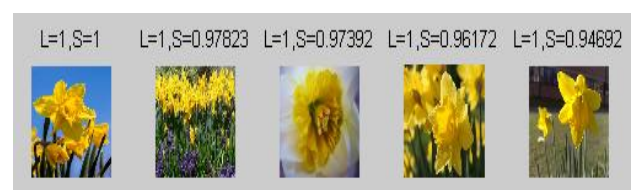


Fig 7: Images are retrieved by using fusion by Random walk pseudo relevance feedback method

To show the performance improvement, a new method Fusion by random walk and Fusion by random walk Pseudo relevance feedback method is introduced. To evaluate this method we have conducted Precision @N.

$$P@N = \frac{1}{|I|} \sum_{i \in I} \frac{R_i}{N} \quad (5)$$

Where, I is the total number of all testing query images, Ri is number of relevant images retrieved for ith testing query image and N is number of top images retrieved for every testing query images.

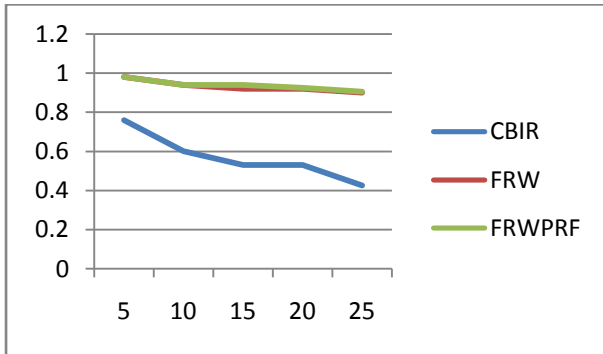


Fig 8: Graph with FRW and FRWPRF method v/s accuracy

Above figure shows the final performance output for ten testing query images using this method. The X-axis represents number of top images retrieved for every testing query images and Y-axis represents the accuracy with respect to above formula described for every testing query images. Blue line shows the CBIR result without using random walk method decreases the efficiency as P@N increases. It reaches up to accuracy of 0.42, when 25 top ranked images are retrieved. Similarly green line shows the result of fusion by random walk method and accuracy of 0.9. Whereas, red line gives the efficient result for fusion by random walk with pseudo relevance feedback method and shows the accuracy of 0.905.

Finally calculate FIAR ratio and FIRR ratio with respect to ten testing query images. FIAR means False image acceptance ratio and FIRR means False image rejection ratio. Formula for calculation is given as:

$$FIAR = \frac{\text{No.of irrelevant images accepted}}{\text{Total no.of images accepted}} \quad (6)$$

$$FIRR = \frac{\text{No.of relevant images rejected}}{\text{Total no.of images rejected}} \quad (7)$$

This formula computes the value for those irrelevant images accepted and relevant images rejected from the database for every testing query input images.

Following figure gives FIAR graph and FIRR graph for Content based image retrieval, Fusion by random walk and Fusion by random walk with Pseudo relevance feedback method for all number of retrieval outcomes such as 5,10,15,20 and 25.

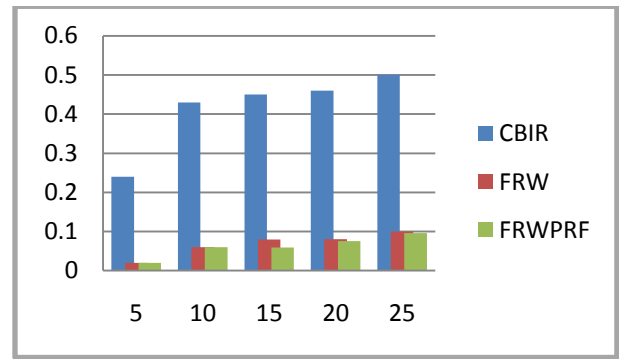


Fig 9: FIAR ratio for CBIR, FRW and FRWPRF

Above figure shows the FIAR value changes for every respective top retrieval outcomes. For Content based image retrieval, values ranges from 0.24 when top 5 images are retrieved, 0.43 for 10, 0.45 for 15, 0.46 for 20 and finally 0.5 for top 25 images retrieved. However for Fusion by random walk values ranges from 0.02 for 5, 0.06 for 10, 0.079 for 15, 0.8 for 20 and 0.1 for 25. Similarly, for Fusion by random walk with pseudo relevance feedback values ranges from 0.02 for 5, 0.06 for 10, 0.059 for 15, 0.075 for 20 and finally 0.096 for top 25 images retrieved.

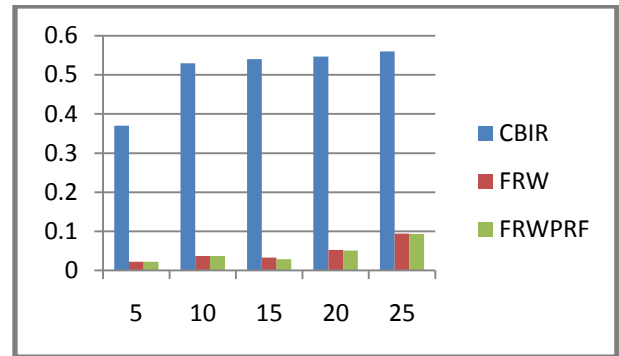


Fig 10: FIRR ratio for CBIR, FRW and FRWPRF

Above figure shows the FIRR value changes for every respective top retrieval outcomes. For Content based image retrieval, values ranges from 0.37 when top 5 images are retrieved, 0.53 for 10, 0.54 for 15, 0.547 for 20 and finally 0.56 for top 25 images retrieved. However for Fusion by random walk values ranges from 0.022 for 5, 0.037 for 10, 0.033 for 15, 0.052 for 20 and 0.094 for 25. Similarly, for Fusion by random walk with pseudo relevance feedback values ranges from 0.022 for 5, 0.037 for 10, 0.029 for 15, 0.051 for 20 and finally 0.093 for top 25 images retrieved.

Therefore, False image acceptance ratio and False image rejection ratio gives ratio values for relevant and irrelevant images for various retrieval techniques. Hence, this value concludes the performance outcomes for FRW and FRWPRF method in accurate manner. Thus, more minimum values of FRW and FRWPRF method for all outcomes more will be efficient. Efficient random walk algorithm enhances the performance of the retrieval and gives accurate identical values for every pairs. Hence, Relevance feedback method plays a vital role in content-based image retrieval method as well as image processing domain.

5. CONCLUSION

This paper utilizes the universal features of the images, constructs the directed mixed graph between image-tag and retrieves the result by random walk method. In this paper, filtration method is used by Gabor transform for every image to avoid the disturbance occur in accessing the system. By considering relevance feedback mechanism, it helps to fill the gap between the semantic searching and low level features that gives efficient and accurate result. This shows the advancement in image processing concept and in networking technologies. Furthermore, this can be designed in a flexible way to include all the information pieces to improve the retrieval performance.

6. REFERENCES

- [1] Sridhar, Gowri. 2012. Color and Texture Based Image Retrieval ARPN Journal of System and Software, Vol. 2, No. 1, pp 1–6.
- [2] P.Nagarani, R.VenkataRamanaChary and Dr.D.Rajya Lakshmi 2012. Semantic Based Image Annotation Using Descriptive Features and Retagging approach, International Journal of Multimedia Its Applications (IJMA), Vol.4, No.1, pp 15–26.
- [3] Yin-Hsi Kuo, Wen-Huang Cheng, Member, IEEE, Hsuan-Tien Lin, Member, IEEE, and Winston H. Hs. 2012. Unsupervised Semantic Feature Discovery for Image Object Retrieval and Tag Refinement, Transactions on Multimedia, Vol. 14, No. 4, pp 1079–1090.
- [4] Sowmya Rani, Rajani N., and Swathi Reddy. 2012. Comparative Study on Content Based Image Retrieval, International Journal of Future Computer and Communication, Vol. 1, No. 4, pp 366–368.
- [5] Hao Ma, Jianke Zhu, Michael Rung-Tsong Lyu, and Irwin King. 2010. Bridging the Semantic Gap Between Image Contents and Tags, IEEE Transactions On Multimedia, Vol. 12, No. 5, pp 462–473.
- [6] L. Wu, L. Yang, N. Yu, and X.-S. Hua. 2009. Learning to Tag, In Proceedings of ACM of the 18th International Conference on World Wide Web (WWW 2009), pp 361–370.
- [7] R. Datta, D. Joshi, J. Li, and J. Z. Wang. 2008. Image Retrieval: Ideas, Influences, and Trends Of The New Age, ACM Computing Survey. Vol. 40, No. 2, pp 1– 60.
- [8] C.Wang, L. Zhang, and H.-J. Zhang. 2008. Learning to reduce the semantic gap in web image retrieval and annotation, In Proc. SIGIR, pp. 355–362.
- [9] G. Carneiro, A. B. Chan, P. J. Moreno, and N. Vasconcelos. 2007 Supervised learning of semantic classes for image annotation and retrieval, Transaction on pattern analysis and machine Intelligence (PAMI), pp 394–410.
- [10] S. Arivazhagan, L. Ganesan, and S. P. Priyal. 2006. Texture classification using Gabor wavelets based rotation invariant features, Pattern recognition letters, Vol. 27, Issue No. 16, pp 1976–1982.
- [11] Leo Grady. 2006. Random Walks for Image Segmentation, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 28, No. 11, pp 1–17.
- [12] Najlae Idrissi, Jose Martinez and Driss Aboutajdine, 2009. Bridging the Semantic Gap for Texture-based Image Retrieval and Navigation, Journal of Multimedia, Vol. 4, No. 5, pp 277–283.
- [13] Dengsheng Zhang, Aylwin Wong, Maria Indrawan, Guojun Lu. 2006. Content-based Image Retrieval Using Gabor Texture Features, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 28, No. 11.
- [14] Yining Deng, B. S. Manjunath, Charles Kenney, Michael S. Moore and Hyundoo Shin. 2001. An Efficient Color Representation for Image Retrieval, IEEE Transactions On Image Processing, Vol. 10, No. 1, pp 140–147.
- [15] Yuanhua Lv and ChengXiang Zhai, 2010. Positional Relevance Model for Pseudo-Relevance Feedback ACM, Special Interest Group on Information Retrieval (SIGIR), pp 579–586.
- [16] Xiang Sean Zhou, Thomas S. Huang. 2003. Relevance Feedback in Image Retrieval: A Comprehensive Review Multimedia Systems, Springer-Verlag, Vol. 8, No. 6, pp 536–544.
- [17] Shu Liao and Albert C. S. Chung. 2007. Texture Classification By Using Advanced Local Binary Patterns And Spatial Distribution Of Dominant Patterns, IEEE International Conference on Acoustics, Speech and Signal Processing, Vol. 1, pp I–1221 – I–1224..
- [18] Li, J. and J. Wang. 2003. Automatic linguistic indexing of pictures by a statistical modeling approach, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 25, No. 9, pp 1075–1088.
- [19] Jeon, J., V. Lavrenko, and R. Manmatha. 2003. Automatic image annotation and retrieval using cross media relevance models, In: Proc. ACM SIGIR Conf. Research and Development in Information Retrieval. New York, NY, USA, pp. 119 –126.
- [20] P. Duygulu, K. Barnard, J. F. G. de Freitas, and D. A. Forsyth. 2002. Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary, In ECCV 2002: Proceedings of the 7th European Conference on Computer Vision-Part IV, Springer-Verlag, pp 97–112.
- [21] Blei, D. M. and M. I. Jordan. 2003. Modeling annotated data. In: Proc. ACM SIGIR. pp. 127–134.
- [22] S. Arya, D. M. Mount, N. S. Netanyahu, R. Silverman, and A. Y. Wu. 1998. An optimal algorithm for approximate nearest neighbor searching fixed dimensions, Journal of ACM, 45(6): pp 8914-923.