International Journal of Research in Advent Technology, Vol.6, No.6, June 2018
E-ISSN: 2321-9637
Available online at www.ijrat.org

A Review on Relevance Feedback in CBIR

K. Srinivasa Reddy[1] P. Swaminathan[2] R. Anandan[3] K. Kalaivani[4]
[1]*Ph.D. Research Scholar, Vels Institute of Science, Technology & Advanced Studies (VISTAS), Chennai, India,*
E-mail: ksrmlgnlgap@yahoo.co.in
[2]*Vice Chancellor (I/C), Vels Institute of Science, Technology & Advanced Studies (VISTAS), Chennai, India,*
E-mail: vc@velsuniv.ac.in
[3]*Associate Professor, Vels Institute of Science, Technology & Advanced Studies (VISTAS), Chennai, India,*
E-mail: anandan.se@velsuniv.ac.in
[4]*Professor & HOD, Vels Institute of Science, Technology & Advanced Studies (VISTAS), Chennai, India,*
E-mail: hodcse@velsuniv.ac.in
Corresponding author: ksrmlgnlgap@yahoo.co.in

Abstract- Relevance Feedback (RFB) is one of the features which are used in some information retrieval systems. RFB is used an interactive process to build the bridge between users and search engine in content based image Retrieval (CBIR) systems. It is used to improve retrieval precision over time by communicating relevant and irrelevant features to retrieval system implicitly. RFB is a supervised learning technique which is developed initially to improve the effectiveness of information retrieval systems. RFB technique is used to bridge gap between low-level image features and semantic concepts which can also be used as query expansion technique. In CBIR, Relevance feedback is used as powerful query modification technique by utilizing feedback information from user query to retrieval system for improving retrieval performance. It uses previous results obtained from given query, which also collects user feedback information and uses this information to classify whether those results obtained are relevant or not relevant to perform new query. In CBIR, two main important characteristics are less computational complexity and high retrieval efficiency. The key issue in RFB is how to utilize feedback information effectively to improve performance of retrieval systems. Here, we are proposed to give a review on RFB for improving performance of retrieval systems.

Index Terms- CBIR; query expansion; Relevance Feedback; Retrieval efficiency; Retrieval system

1. INTRODUCTION

CBIR is one of the methods to retrieve similar images from image database in response to query. CBIR process is usually based on comparison of low-level features such as Color, texture and shape extracted from images themselves. The performance of the CBIR systems in two aspects such as response time and retrieval accuracy is still limited. Limited retrieval accuracy is biggest problem in CBIR because of large gap between low-level image features and semantic concepts. Different types of features have different significance in different queries. There is a lack of universal formula for all queries used to derive a weighting scheme which is used to balance different feature types. One of the techniques to bridge the gap is to use Relevance Feedback (RFB).

To improve accuracy of intermediate feature based image retrieval, artificial neural network based relevance feedback method is used. Instead of ANN, distance based method can also used in relevance feedback stage. Relevance Feedback is used in retrieval systems to prompt the user, to give feedback on retrieval results and then utilize this feedback for subsequent retrievals to achieve increased retrieval performance. After retrieving set of images, the user is able to mark as “relevant” or “irrelevant” for each image. User ratings are fed back to retrieval system, which then used to retrieve images in the database which are more similar to user. A set of images are retrieved, and same process repeats until user satisfies with feedback results. It is also used to increase recall rate thereby increases retrieval performance. Relevance Feedback uses positive and negative feedback provided by user to improve performance of retrieval system. For given query, retrieval system retrieves list of ranked images according to predefined similarity metrics. From retrieved images, user interprets set of positive as well as negative feedback. Then retrieval system refines query by retrieving new list of ranked images. The key issue in RFB technique is how to use positive and negative feedback mechanism to refine query and the method of adjusting similarity measurement according to given feedback. Positive feedback is more valuable than negative feedback and many systems allow only positive feedback.

In Relevance Feedback, user can go through one or more iterations. Image search also provides an example of relevance feedback.

Relevance Feedback retriever should possess limited feedback per iteration, only one or two iterations and be fast enough for online implementation to improve retrieval performance. In CBIR, query point movement is also used as RFB technique to achieve better performance and accuracy. Query expansion and term reweighting are used to improve initial query formulation. The approaches are based on user feedback information, information obtained from initial retrieved documents and global information retrieved from document collection. RF algorithms are divided in to three different types based according to the retrieval model adopted namely probabilistic approach, distance-based approach, and machine learning based approach.

Relevance feedback usually improves average precision by increasing positive feedback in query 10-15% as compared to traditional information retrieval search. Thesaurus expands query to increase accuracy. Thesaurus is generated by finding similar terms. Terms that co-occur with each other over a threshold are considered similar. To expand query using term Co-occurrence, for given term t_i, top t similar terms are picked and these words can be used for query expansion. Relevance feedback, thesaurus and word net are used as the query expansion techniques. The effectiveness of relevance feedback depends upon similarities between relevant and irrelevant documents are small.

The advantages of query expansion technique are, there is no need to query documents two or three times, suitable for systems whose response times are small and also improve retrieval performance in the range of 20%. Because of long respose times for user, high cost for retrieval system and partial solution, long queries are inefficient for typical IR Search engine.

The organization of a review on RFB in CBIR is presented in different parts. Part 2 presents detailed information and general review of RFB such as operations, methods, approaches, modifications and types etc. The method of retrieval of images using RFB and its implementation is presented in part 3. The discussion about Literature survey is presented in part 4. The last part 5 gives conclusions about RFB mechanism and its performance in retrieval systems and scope of future work.

2. GENERAL REVIEW ON RELEVNCE FEEDBCK

2.1. Basic Operations in RFB:

Two basic operations used in RFB are Query expansion and term reweighting. Query expansion is used to add new terms from relevant document where as term reweighting is used to modify term weights based on user relevance feedback.

2.2. Methods of RFB:

Initially, RFB methods are used for short-term learning and long-term learning.

2.2.1. Short-*term Learning*: It is also known as intra query learning. It deals with current feedback session. It ignores all other users' previous historical data which results a high loss of useful information.

2.2.2. Long-*term Learning:* It is also known as inter-query learning. It uses feedback information by collecting and recording over a different variety of query sessions from different users. In multimedia searching, in the last few years it has played an important role increasingly. It can also be used to improve retrieval performance more effectively and efficiently.

2.3. Relevance Feedback approaches

RF approaches can be classified in two main categories such as learning-based method and model-based method.

2.3.1. *Learning-based method:* It includes methods which are based on support vector machine to train a classifier to distinguish between positive and negative feedback.

2.3.2. *Model-based method:* It includes those that attempt to model statistical distribution of feedback mechanisms in feature space.

2.4. RFB Modifications:

Relevance feedback process can be improved by using various techniques such as number of top ranked documents, number of feedback terms, term weightings, feedback term selection techniques, iterations, phrase versus single term, document clustering, term frequency cutoff points, RFB thresholding, query expansion using a thesaurus.

2.5. Types of RFB:

There are three types of Relevance feedback namely explicit feedback, implicit feedback and pseudo (blind) feedback.

2.5.1. Explicit RFB: This type of feedback is used to indicate relevance of a document retrieved for a query and defined as explicit only when users of a system know that provided feedback can be interpreted as relevance judgments. Binary or graded relevance systems are used to indicate relevance explicitly by users. A document is either relevant or irrelevant for a given query is indicated by binary relevance feedback. The relevance of a document to a query on a scale using letters, numbers or descriptions.

2.5.2. Implicit RFB: It is inferred from user behavior. This type of feedback is used to collect information about documents which are used to select or not to select for viewing, time duration for viewing document or browsing a page or for scrolling actions. During search process, there are many signals, one can use for implicit feedback.

2.5.3. Pseudo (blind) RFB: This feedback method is used for automatic local analysis without user intervention. In this method, the user gets improved retrieval performance without an extended interaction, automatically using manual part of relevance feedback. It retrieves initial set of relevant documents and assumes that top ranked documents are relevant. It is one of the mechanisms used for improving performance of retrieval systems.

Rocchio algorithm is used to interpolate relevance feedback information with original query and improves performance of retrieval systems. It is used to implement Relevance Feedback.

2.6. Interaction between User and Search Engine using RFB Mechanism

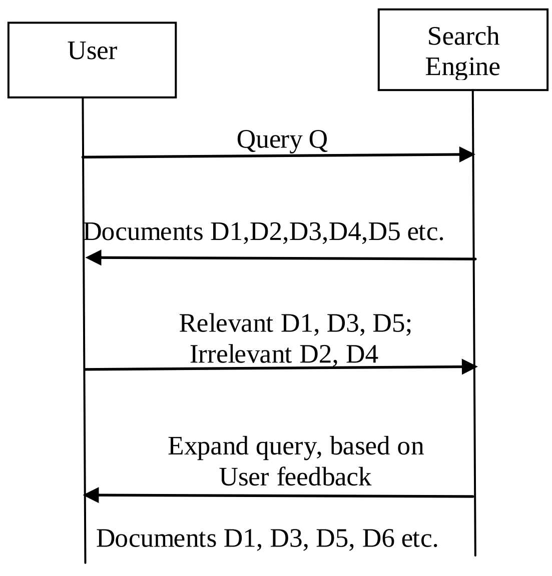

Fig. 2.1. Flow diagram of interaction between user and search engine.

The interaction between user and search engine using relevant and irrelevant feedback is shown in figure 2.1. Sequence of operations used to retrieve relevant documents is as follows:

- User sends Query Q to Search engine.
- In response to query, Search engine retrieves all documents related to query.
- User provides feedback information of documents as relevant or non-relevant to Search engine from retrieved documents.
- Search engine expands query based on feedback from user and returns refined search results.

2.7. Query Processing:

It is an important process used by search engine**.** The steps used in query processing are as below:

- Query can be represented by $q = (t_1, \ t_2, \ t_{3,\ldots\ldots})$
- Find all document terms similar to query terms based on thesaurus and given as $(k_1, \ k_2, \ k_{3,\ldots\ldots})$ are similar to terms $(t_1, \ t_2, \ t_{3,\ldots\ldots})$
- For each similar term, we compute similarity of $(k_1, \ k_2, \ k_{3,\ldots\ldots})$ to whole query (q).
- Expand query with terms which are most similar to original query.

2.8. General Retrieval Process Using RFB

In Relevance feedback, users gives additional input i.e., relevant or non-relevant on documents.

- Collect the results from top 10 to 50 in most cases obtained by initial query as relevant.
- Select only top 20 to 30 terms from documents using instance term frequency (tf) - inverse document frequency (idf) weights.
- Perform Query expansion i.e., users gives additional input on words or phrases and then add these terms to query which is very effective in increasing recall.
- Match the returned or retrieved documents for this query.
- Returns or retrieves the most relevant documents.

3. RETRIEVAL OF IMAGES OR DOCUMENTS USING RFB

RFB is used to expand queries with vector model, to reweight query terms with probabilistic model and also with a variant of probabilistic model.

3.1. Rocchio Vector space RFB Model

To pick a relevance feedback query, Rocchio algorithm uses vector space model. Based on

relevant documents, it adds important terms to original query. Similarly, based on irrelevant documents, it removes unimportant terms from original query. Vector space model is an algebraic model used to represent documents.
Vector representation of documents is

$$V_d = [W_1, W_2, W_3, W_4, \ldots\ldots]$$

Where W_i= Weight of a term 'i' in a document

The different steps used in this method are to run query and show user results. Based on user feedback, increase query term weights or add new terms to query. Otherwise, decrease term weights or remove terms to query in order to increase query accuracy.

$$Q' = \alpha Q + \beta \sum_{i+1}^{n_1} R_i - \gamma \sum_{i=1}^{n_2} S_i$$

Where Q′ = New Query Vector; Q= Original query Vector; R= Set of relevant document vectors;
S= Set of non-relevant or irrelevant document vectors and α.β.γ are Constants known as Rocchio weights.

Variations in this model are given as follows:

- $\alpha = \beta = \gamma = 1$;
- $\gamma < \beta$;
- $\gamma = 0$; allows only positive feedback.

Use only first n documents from R and S. Among that, use only first document of S. Do not use S when γ=0.

Example: 3.1.1. Original Query Q: (1, 0, 3, 2, 1)
Relevant documents R1&R2:
(1, 2, 0, 1, 3) (2, 0, 1, 2, 1)
Non-Relevant document S : (2, 1, 3, 0, 1)
If α = 1; β= 0.5; γ =0.2;
Then, Q′ = Q + 0.5 (R1+R2) – 0.2 S
Q′= (1, 0, 3, 2, 1) + 0.5 (3, 2, 1, 3, 4) – 0.2 (2, 1, 3, 0, 1)
Q′= (2.1, 0.8, 2.9, 3.5, 2.8)

3.2. Probabilistic model in RFB

In Probabilistic model, some query terms are more important than others which are less important. This model is used to recalculate weights of query terms. It does not remove or add terms. To identify relevant or irrelevant documents, this model needs user interaction and determines with high precision. In order to rank matching documents according to their relevance to given search query, probabilistic relevance feedback uses search engines and web search engines to derive ranking functions.

This model estimates the probability of finding that a document d_j is relevant to a query q, and assumes that probability of relevance depends on representation of document and query. It also assumes that there is portion of all documents that is preferred by user as answer set for query q called as R which should maximize overall probability of relevance to that user. The documents which are present in this set R are relevant to query and those documents which are not present in this set R are non-relevant.

$$sim(d_j, q) = \frac{P(\frac{R}{\vec{d}_j})}{P(\frac{\bar{R}}{\vec{d}_j})}$$

The limitations in this model are there is no accurate estimate for first run probabilities, index terms are not weighted and terms assumed are mutually independent. To overcome and solve these problems, new developed one such as Binary Independence Model is used.

3.3. Semantic Networks in RFB Model

In Relevance feedback, mismatch problem is resolved by using semantic networks which measures semantic distance, instead of matching document terms and query terms. Semantic networks are used to build a network that is used for each word which shows its relationships to other words may be phrases. In semantic networks, to expand query, find the word in network and follow various arcs to other related words. To compute distance from one word in network to another, different distance measures can be used.

3.4. Retrieval of Images or documents using RFB from Search Engine

Use of RFB for searching images from a search engine is shown in fig. 3.1. RFB is to involve user in retrieval process to improve refined final result. From initial set of results, user gives feedback on relevance of documents. Steps used in retrieval process are as follows. Initially, user issues a short or simple query "bike" for search engine. Then, retrieval system returns an initial set of results and user mark relevant and irrelevant images and gives these feedback to the retrieval system as shown in fig.3.1 (a). Retrieval system then computes better representation of information and displays revised set of retrieval results as shown in fig.3.1 (b). In this type of sort, RFB uses one or more iterations.

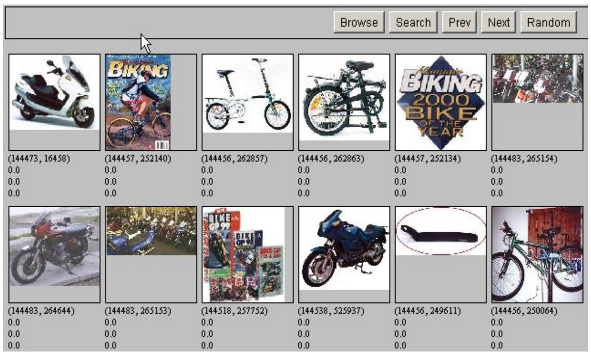

Fig. 3.1.(a). Marking first, third and fourth results in top row and fourth result in bottom row as relevant images for initial query results and submits this feedback information

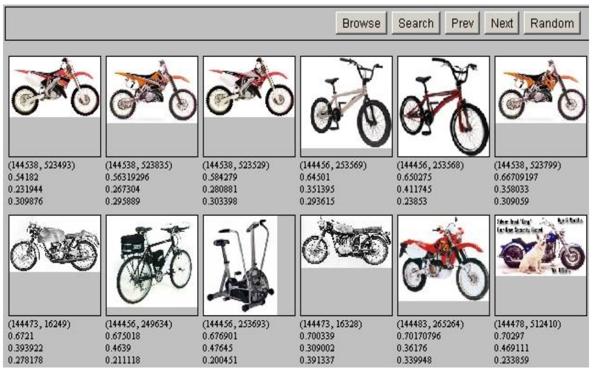

Fig. 3.1.(b). Refined result for user after improving precision using Relevance Feedback.

Fig. 3.1 Use of RFB for searching images with Search Engine

The schematic diagram for retrieval of images is shown in figure.3.2.

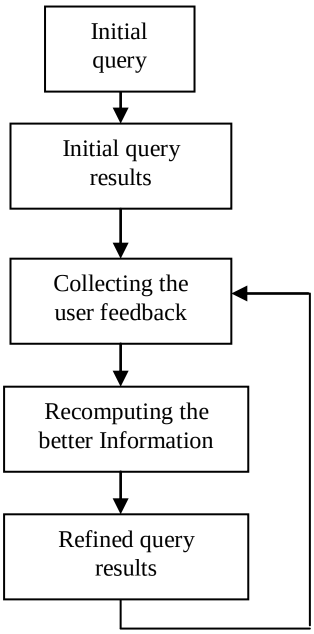

Fig. 3.2 Schematic diagram for retrieval of images using RFB mechanism

3.5 Retrieval of Images using Positive and Negative feedback:

Query Image

Fig. 3.3. (a) Query Image

Initial query results

Fig. 3.3. (b) Retrieval of images after Initial query

User Relevance Feedback

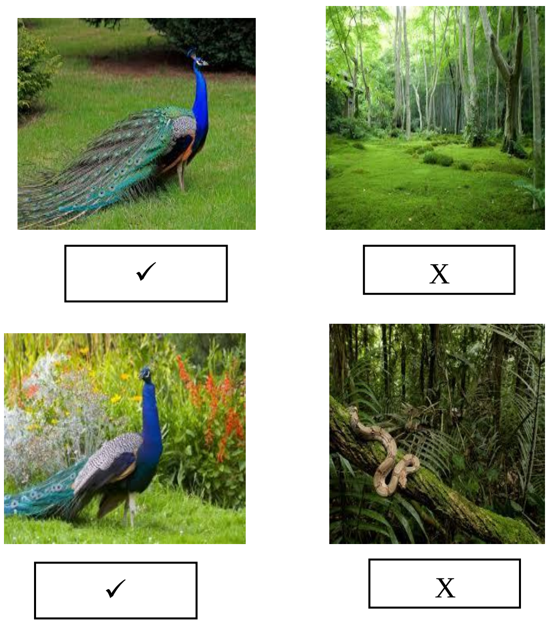

Fig. 3.3. (c) User feedback "√" for positive and "X" for negative feedback

Query results after user feedback

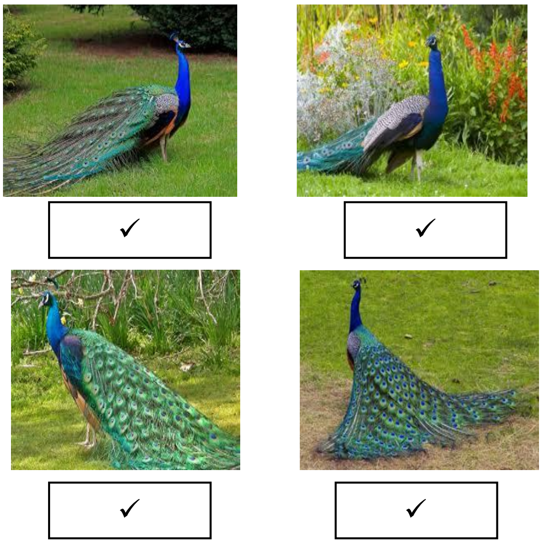

Fig. 3.3 (d) Refined retrieval results using feedback

Fig. 3.3. Retrieval of relevant images using user feedback

Retrieval of images using positive and negative feedback is shown in fig.3.3. Query image is shown in fig.3.3 (a). According to predefined similarity metrics, retrieval system retrieves list of ranked images for the given query as shown in fig.3.3 (b). The user marks positive feedback as "√" for relevant images and negative feedback as "X" to non-relevant images from retrieved images in response to query as shown in fig.3.3(c). Based on feedback given by user, retrieval system will refine results and presents new list of relevant retrieved images as shown in fig.3.3 (d). In this process, user provides binary information for both positive and negative feedback. But, some systems provides degree of relevance and irrelevance by specifying as "relevant", "high-relevant", "non-relevant", "highly non-relevant" or 50% relevant.

4. LITERATURE REVIEW

The earliest paper using query expansion technique is reported in [1]. In this paper, the authors proposed one of the techniques to improve precision and recall rate. Carbonell et al., [2] proposed method of reordering and reranking of documents and presented brief summary. Sean D. MacArthur et al., [3] proposed RFB in retrieval systems, for each retrieval iteration using a decision tree which is useful for differentiating relevant and irrelevant instances and also implemented the HRCT images of lung.

Kelly et al., [4] discussed that how implicit feedback used for inferring user preference. Zhong Su et al., [5] used Bayesian classifier to provide positive and negative feedback with different techniques to improve retrieval speed by reducing required memory and also significant retrieval accuracy. Jing Xin et al., [6] presented that, RFB mechanism as powerful tool in image retrieval using Bayesian belief network as the relevant image adoption model.

Jansen B.J. et al., [7] discussed methods of evaluation and effectiveness and also patterns of interactions with automated assistance of information retrieval systems. Apostolos Marakakis et al., [8] describes relevance feedback approach using Gaussian Mixture models in probabilistic manner and comparative experimental results are presented after evaluation of probabilistic framework. Stephen Robertson et al., [9] describe probabilistic relevance framework and assumptions of probability dependence factors. Yuanhua Lv et al., [10] specified significance of pseudo-relevance feedback and also positional relevance model.

Manish Chowdhury et al., [11] implemented CBIR system using Ripplet Transform technique and also used fuzzy RFB mechanism to improve retrieval results from large databases effectively and efficiently with experiments. Shanmuga priya. N et al., [12] extracted texture based images using Gaussian mixture models and also proposed that query point movements is also one of the technique used as a RFB in CBIR systems to achieve better accuracy and performance in retrieving images.

Zhijiang Li et al., [13] used Visual attention models with experimental results to extract features more effectively and to optimize learning process of feedback model based on support vector machine. Particle swarm optimization is also used to boost the accuracy and efficiency of the retrieval systems in retrieval of multiple objects in one image. Yong Rui et al., [14] described interactive retrieval approach which is used to reduce user effort for composing more precisely such as query and capturing user information and also introduced human-computer interaction approach based on relevance feedback in CBIR.

Nazli Gaharian [15] described use of RFB, thesaurus and semantic networks as query expansion techniques to increase retrieval performance by increasing precision range of 20%. Jing Li et al., [16] discussed short-term and long-term learning approaches in RFB used for CBIR systems. Dipankar Hazra et al., [17] applied Artificial Neural Network in relevance feed back stage and Structure. Query Language is used in initial stage to construct intermediate texture based image retrieval to improve accuracy.

5. CONCLUSIONS AND FUTURE SCOPE

Here, we are focused on use of RFB and retrieval of images or documents using user feedback methods and also discussed about a review of RFB mechanism in CBIR. This paper also suggested that modification for Relevance Feedback using different techniques and concluded that Relevance Feedback is used to increase precision range of 20%. This paper is also given direction for future researchers to increase precision range in order to increase retrieval performance. Our future work will be directed to retrieve the medical images by improving retrieval speed and accuracy in order to increase retrieval performance.

REFERENCES

[1]. Jinxi Xu and W. Bruce Croft: Query expansion using local and global document analysis: Proceedings of the 19th annual international ACM SIGIR conference on Research and development in information retrieval. (SIGIR), 1996.

[2]. Carbonell and Goldstein: The use of MMR, diversity-based reranking for reordering documents and producing summaries: SIGIR 21, 1998.

[3].Sean, D. MacArthur; Carla, E. Brodley; Chi-Ren Shyu: Relevance Feedback Decision Trees in Content-Based Image Retrieval: Proceedings of the IEEE workshop on Content-Based access of Image and video libraries, 2000.

[4]. Kelly; Diane, and Jaime Teevan: Implicit feedback for inferring user preference: a bibliography: ACM SIGIR Forum, Vol.37, No.2, ACM, 2003.

[5]. Zhong Su; Hongjiang Zhang; Stan Li and shoaping Ma: Relevance Feedback in Content-Based Image Retrieval: Bayesian Framework, Feature Subspaces, and Progressive Learning: IEEE Transactions on Image Processing, Vol.12, No.8, August, 2003.

[6].Jing Xin; Jesse; S.Jin:Relevance Feedback for Content-Based Image Retrieval using Bayesian Network: Conferences in Research and practice in information technology, Vol.36, 2004.

[7]. Jansen, B.J. and McNeese: Evaluating the effectiveness of and patterns of interactions with automated assistance in IR systems: Journal of the American Society for information Science and Technology, 56(14), 2005, 1480-1503.

[8].Apostolos Marakakis; Nikolaos Galatsanos; Artistidis Likas and Andreas: Application of Relevance Feedback in Content Based Image Retrieval Using Gaussian Mixture Models: 20th IEEE International Conference on Tools with Artificial Intelligence, 2008.

[9]. Stephen Robertson and Hugo Zaragoza: The probabilistic relevance framework: BM 25 and beyond: Trends inf., 2009.

[10]. Yuanhua Lv and Cheng Xiang Zhai: Positional relevance model for pseudo-relevance feedback: Proceedings of the 33rd International ACM SIGIR conference on Research and development in information retrieval (SIGIR), 2010.

[11].Manish Chowdhury; Sudeb Das and Malay Kumar Kundu: Interactive Content-Based Image Retrieval Using Ripplet Transform and Fuzzy Relevance Feedback: PerMIn 2012, LNCS7143, 2012, pp. 243-251.

[12]. Shanmugapriya, N., and Nallusamy, R. : A New Content Based Image Retrieval System using GMM and Relevance Feedback: Journal of Computer Science 10(2), 2014, 330-340.

[13]. Zhijiang Li; Jiaxian Long; Chuan Dong: Visual attention model and relevant Feedback based Image Retrieval: IS&T International Symposium on Electronic Imaging, Visual Information Processing and Communication VII, Society for Imaging Science and Technology, 2016.

[14].Yong Rui; Thomas, Huang, S. ; Michael Ortega and Sharad Mahrotra: Relevance Feedback: A powerful tool for Interactive Content-Based Image Retrieval: IEEE Transactions on Circuit and Video Technology.

[15].Nozli Goharian: Query Expansion techniques (COSC 488): Relevance Feedback, Thesaurus, Semantic Network.

[16]. Jing Li and Nigel, M. Allinson: Relevance Feedback in Content-Based Image Retrieval: A survey, Chapter 13.

[17]. Dipankar Hazra; Debnath Bhattacharyya; Tai-hoon Kim: Artificial Neural Network based Relevance Feedback for Intermediate Feature Based Image Retrieval: Advances in Information Science and Computer Engineering.