IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

A SURVEY ON DIFFERENT WEB IMAGE RE-RANKING TECHNIQUES

Jeeshna P.V.[*], Mr. Chitharanjan K.
[*] Computer Science, SCT College of Engg. Pappanamcode, India

ABSTRACT

Today hundreds and millions of people use YouTube, Tudou, Face book like social networks. So users upload tremendous number of images on the web. Many people search and view images and videos on the web every second. There are many methods to search and retrieve the images and videos: content based search, text based search, visual search and re-ranking techniques. The content based search, is widely used in the past decades.In this method the user give some example images or videos and then similar images and videos are returned by querying a visual representation in a large scale database. In text based multimedia search completely relies on indexing the associated textual information of images, such as image tags, web page file names, and surrounding text. With textual information, well understood search techniques can be applied directly to image search. Web image re-ranking techniques leverages from content and text based search. So it becomes a popular web image re-ranking method. Image re-ranking is an effective way to improve the results of web based image search has been adopted by the current commercial search engines. Image re-ranking is important and effective for improving the performance of a text based image search. Image re-ranking is the most popular research area and recently several researches started on this area. Naturally this is not a complete survey on the entire web image re-ranking techniques. In this focuses some of the web image re-ranking methods.

KEYWORDS: Re-ranking, Active, Bag based, Query specific, Click prediction

INTRODUCTION

Numerous numbers of images are available on the websites. People search and view pictures and videos on the web every second, with the help of search engines such as Google, Yahoo, Bing, AltaVista, Ask etc. There are many methods to retrieve and search the images and videos: content based search, text based search, visual search and re-ranking techniques.

Text based image retrieval is commonly used in the web search engines. The user enter the input keyword as a textual query to retrieve the system. The system returns the ranked relevant data based on the similarity measurements between the adjacent text of the query keyword and the textual features of relevant data. The recovery performance is very low because of the textual features of the web images are sparse and noisy in a high dimensional space. To solve this problem image re-ranking methods are introduced.

Several approaches used to improve the performance of web image search. One approach is image annotation and second is web image search re-ranking. Automatic annotations of large scale web images hardly attain suitable performance due to the semantic gap. Another approach is web image search re-ranking aims to improve the text indexing of web images, re-ranking are applied to directly adjust search results by mining images visual content. Image search re-ranking methods are developed based on two criterias:

1. The final results after re-ranking should not change too much from the original ranking list.
2. Visually similar images should be close in a ranking list.

The performance of text based image retrieval for many searches is acceptable, due to the accuracy and efficiency of the retrieved results. One major problem facing the performance is the mismatches between the original content of images and the textual data on the web page. Main method to solve this problem is image re-ranking, in which both visual and textual information is joined to return improved results to the user. The ranking of images based on a text based search is considered a reasonable. Extracted visual information is then used to re-rank related images to the top of the list.

The main problem of web image search is the mismatch between the image content and the web page text. Image search re-ranking attempts to resolve this issue by relying on both the text information and visual information channels during the image search process. The ranked list of images obtained via search in the text channel is considered a noisy but informative baseline. The images visual content is then deployed to reduce ambiguity in the list and move more of the relevant images toward the top of the list. Re-ranking method to improve web image retrieval by reordering the images retrieved from an image search engine. The re-ranked image retrieval achieved better performance than original web image retrieval, suggesting the effectiveness of the re-ranking method. The re-ranking process should be applicable to any image search engines with little effort.

OVERVIEW OF WEB IMAGE RE-RANKING METHODS

The textual and visual informations combined to form the actual result in web image re-rankiing process. The text based search first generate the ranked list and the extracted visual data will re-rank the related images images to the top of the list. Some effective techniques of web image re-ranking are discussed.

Active re-ranking for web image search

The query term is not clear then we can not satisfies the user intention. So re-ranking with user involvements or active re-ranking is very important to improve the performance of the search. The active re-ranking is used to focus the user's want, that identifies query related important images from an unimportant images. To satisfy this aim to use the structural information based sample selection strategy is to reduce the user's efforts, a local global discriminative dimension reduction algorithm is used to localize the user's wish based on the visual features. It is used to transferring the local geometry and the discriminative information from the labeled images to the whole (global) image database.

Figure 1:

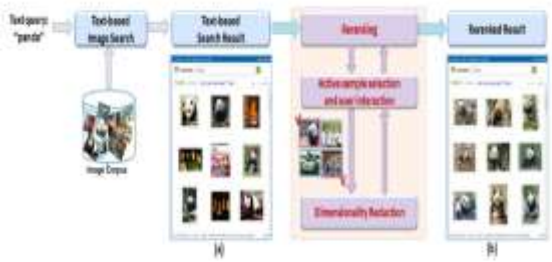

Fig: Framework for active re-ranking in web image search with the query "panda"

Fig:1 shows the general architecture with the text query term "panda" as an example[2]. First enter the term "panda", an initial text based search result is generated, as shown in Fig. 1(a) (some top ranked images are given). This result is not satisfied due to the presence of both animal and person images as the top results. It is caused by the ambiguity of the text query term. Without the user involvement, it is impossible to remove this ambiguity. Based on the user's intention specify animal or person.

In active re-ranking Fig. 1, first picked the four images based on an active sample selection strategy, and then the user mark them. If the user marks the animal panda as query relevant and other images (person, car) as query irrelevant then animal panda is the user's intention. The intention, i.e., the animal panda, a discriminative sub manifold should be separate query relevant images from irrelevant ones. A dimension reduction step is to localize the visual characteristics of the user's intention. Based on the known user's intention, including both the labeling information and the learned discriminative submanifold, the re-ranking process is done and different kinds of animal pandas are get as result, as shown in Fig. 1(b). Sometimes, several rounds are perform to get the satisfactory performance. There are two key steps to learning the user's intention, i.e.,

- Active sample selection strategy
- Dimension reduction algorithm.

These two steps via a new SInfo sample selection strategy and a novel LGD dimension reduction algorithm.

SINFO Active Sample Selection

Structural Information based sample selection strategy is used to understand the user's intention efficiently with two important aspects. One is ambiguity and the second is representativeness.

1. Ambiguity

The ambiguity shows whether an image is important or unimportant. In SInfo, the ambiguity of an image is calculated by the entropy of the relevance probability distribution. In active re-ranking, it is direct and reasonable to measure the ambiguity with the ranking scores obtained in the re-ranking process. There are two reasons. One is the re-ranking problem it is different from classification, thus the ambiguity estimated via conducting classification task may be not as accurate as that directly derived in re-ranking process. The other reason is that additional cost will be introduced if the ambiguity is estimated via other learning methods. In contrast, measuring ambiguity through the ranking scores avoids this additional cost.

For an image I_i, $0 <= r_i <= 1$ is its ranking score, where $r_i = 1$ means I_i is relevant, while $r_i=0$ means I_i is totally irrelevant. Ambiguity of I_i is

$$H_r(I_i) = -r_i log r_i - (1 - r_i)\log(1 - r_i) \quad (1)$$

Ambiguity in the initial text based search result is

$$H_{\bar{r}}(I_i) = -\bar{r}_i log \bar{r}_i - (1 - \bar{r}_i)\log(1 - \bar{r}_i) \quad (2)$$

where $0 <= \bar{r}_i <= 1$ is the initial text based search reranking score for I_i. Therefore, (1) and (2) combine to produce the total ambiguity for I_i is

$$H(I_i) = \alpha H_r(I_i) + (1 - \alpha)H\bar{r}(I_i) \quad (3)$$

where $\alpha(\varepsilon\ [0,1])$ is the trade-off parameter to control the influence of the two ambiguity.

2. Representativeness

The information provided by individual sample can be shared by its neighbours. Therefore more representative samples are preferred for labeling. To measure the representativeness of image I_i via the probability density $p(I_i)$, which can be estimated by using the kernel density estimation(KDE).

$$p(I_i) = {}^{1}/_{N_i} \sum_{I_j \in N_i} k(x_i - x_j) \qquad (4)$$

where N_i is the set of neighbours of I_i, x_i is the visual feature for image I_i. k_x is a kernel function that satisfies both k(x) > 0 and lim k(x)d(x) = 1

3. Active Sample Selection

The structural information of image I_i measured by the product of the two terms, i.e ambiguity and representativeness; Combine (3) and (4) gives the $SI(I_i)$.

$$SI(I_i) = p(I_i)H(I_i) \qquad (5)$$

LGD dimension reduction

Local Glogal Discriminative(LGD) dimension reduction algorithm is used to signify visual characteristics of the user's intention. LGD considers both the local and global informations. Local information considered in the marked images and the global information of the whole image database simultaneously. LGD converts the local information, it contains both the local geometry and the discriminative information in the labeled images, to the global domain. A local area formed by a set of adjacent images are called as one patch. There are three types of images: labeled relevant, labeled irrelevant, and unlabeled. Therefore, 3 types of patches, which are:

1. Local patches for labeled relevant images: To represent the local geometry of them and the discriminative information to separate important images from unimportant ones.

$$I_i^+ = \min ||y_i^+ - \sum_{j=1}^{k_1} (c_i)_j y_{i_j}| \quad |^2 - \beta \sum_{j=k_1+1}^{k_1+k_2} ||y_i^+ - y_{i_j}| \quad |^2 \qquad (6)$$

2. Local patches for labeled irrelevant images: To represent the discriminative information to separate irrelevant images from relevant ones.

$$I_i^- = min - \sum_{j=1}^{k} ||y_i^- - y_{i_j}||^2 = min\ tr(Y_i^- L_i^- (Y_i^-)^T) \qquad (7)$$

3. Global patches for both labeled and unlabeled images: For transferring both the local geometry and the discriminative information from all marked images to the unmarked ones.

$$\max tr\,((y_i - y^m)(y_i - y^m)^T) = \max tr(Y_i L_i^{PCA} Y_i^T) \qquad (8)$$

Web image re-ranking by bag-based

Bag- based re-ranking framework is used for large scale TBIR [3]. First cluster important images using both textual and visual features. Each cluster as a "bag" and the images in the bag are considered as "instances," gives this issue as a multi-instance (MI) learning problem. To identify the ambiguities on the instance labels in the positive and negative bags under GMI setting. GMI [3] is to improve retrieval performance by propagating the labels from the bag level to the instance level. To acquire bag annotations for GMI learning, a bag ranking method to rank all the bags according to the defined bag ranking score. The optimized top ranked bags having similar relevant images are used as pseudo positive training bags, while pseudo negative bags can be a few unimportant images that are not associated with the textual query.

Clustering is the process of grouping similar images together and comparing or matching among clusters instead of individual images. It will reduce the large time complexity. The positive and negative bag clustering is based on the theory of Generalized Multi called as bag based image re-ranking.

In this method first seperates the important images to clusters by visual and textual features. Consider 'bag' as each cluster and 'instances' as the images in the bag with the help of multi-instance(MI) .

Figure 2:

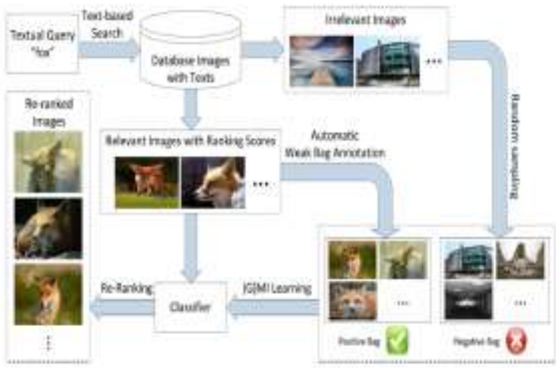

Fig: Bag based image re-ranking framework for large scale TBIR

In MI learning techniques, if a bag contains at least one important instance, this bag is labeled as positive; if the instances in a bag are all unimportant, this bag is negative. Some cases multiple relevant images are clustered in a positive bag while a few relevant images may be clustered with irrelevant images in a negative bag. GMI technique is used in this case i.e at least a certain portion of a positive bag is of positive instances, while a negative bag might contain at most a few positive instances. This case the MI techniques may not be effective to solve the ambiguities. Therefore, GMI learning algorithm effectively re-rank the relevant images by propagating the labels from the bag level to the instance level.

GMI learning method in bag based framework, weak bag annotation process automatically find positive and negative bags for classifiers. First, introduce an instance ranking score defined by the similarity between the textual query and each important image. Averaging the instance ranking scores of the instances in this bag is obtain with the

ISSN: 2277-9655
(I2OR), Publication Impact Factor: 3.785

help of ranking score of each bag. Finally, rank all bags with the bag ranking score. Automatic bag annotation technique, the top ranked bags are used as the pseudo positive bags, and pseudo negative bags are obtained by a few irrelevant images that are not associated with the textual query. Then these bags are used to train a classifier that is then used to re-rank the database images.

Web image re-ranking using query-specific semantic signatures

Query specific semantic image re-ranking framework has two parts: offline and online part. At the offline stage, get different semantic spaces for different query keywords. These semantic signatures are get based on projecting the visual feature of images to the semantic spaces specified by the query keyword. At the online part, comparing the semantic signatures are aquired from the semantic space and images are re-ranked. The query-specific semantic signatures significantly increase both the accuracy and effectiveness of image re-ranking.

Figure 3:

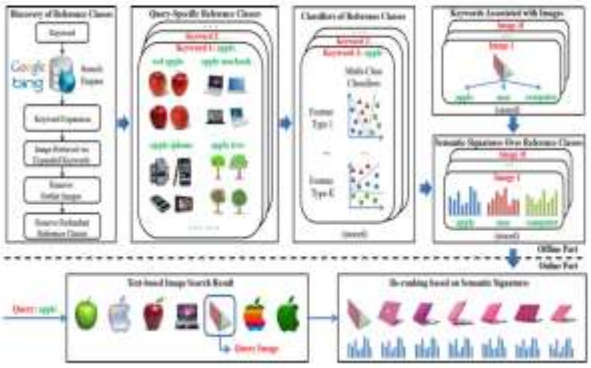

Fig: Web image re-re-ranking using query specific semantic signatures

In the Fig 3: having two stages: Online part and offline part [4], [5]. At the offline stage, the reference classes related to query keywords are automatically exposed and collected the images. The steps are :

1. Keyword: Enter the query keyword in search engines (Eg: Apple).
2. Keyword Expansions: A set of keyword expansions automatically selected with the help of textual and visual information and defines the reference classes. It is used to retrieve images by the search engine based on textual information again.(E.g, red apple apple macbook, apple tree and apple iphone.)
3. Images retrieved by the keyword expansion are much less varied than those retrieved by the original keyword.
4. Automatically removing outliers, the retrieved top images are used as the training examples of the reference class.

Click prediction for web image re-ranking hypergraph base sparse coding

User click information [6] is used in image re-ranking, so clicks are to be used to show more accurate images to search queries. But the main problem of click-based methods is the lack of click data, since users click small number of web images. To solve this issue by predicting image clicks. The image clicks prediction, and apply the resultant click data to the re-ranking of images. A hyper graph is used to build different features through a group of weights. A hyper graph that has an edge between two vertices, a hyper edge which connects a set of vertices, and helps conserve the local smoothness of the constructed sparse codes. Then the optimization procedure and the weights of different modalities and the sparse codes are simultaneously got. Atlast, a voting strategy explains the predicted click as a binary event i.e a click or no click, from the images' corresponding sparse codes.

Figure 4:

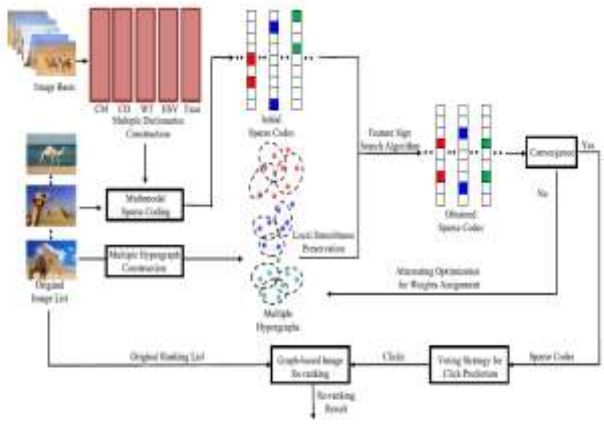

Fig: Multimodal hyper graph learning based sparse coding for click prediction

The Fig 4. shows the architecture of multimodal hyper graph learning based sparse coding for click prediction. First, multiple features are extracted to describe web images. Second, from these features, construct multiple hyper graph, and perform sparse coding based on the integration of multiple features [7]. Local smoothness of the sparse codes is preserved by using manifold learning on the hyper graphs. The sparse codes of images, and the weights for different hyper graphs, are obtained by simultaneous optimization using an iterative two stage procedure. A voting strategy is adopted to predict the click as a binary event from the obtained sparse codes. The non-zero positions in sparse code represent a cluster of images, which are used to reconstruct the images. If more images have clicked, then the image is predicted as clicked otherwise not clicked. Finally, a graph-based schema is used with the predicted clicks to obtain image re-ranking.

1. *Definition of Hypergraph-Based Sparse Coding*

The spare coding method, the web images are represented independently and similar images can be described totally different sparse codes. To preserve locality informtion, the hypergraph Laplacian is utilized. The pairwise distance between the sparse code within each hyperedge by w(e)/δ(e), the hypergraph based sparse coding formulated as

$$min_{c_1 \ldots\ldots c_n} \sum_i ||x_{i-}Ac_i||^2 + \alpha \sum_i ||c_i||_1 + \frac{\beta}{2}\sum_{e\in E}\sum_{(p,q)\in E} w(e)/\delta(e)||c_p - c_q||^2$$

where c_i is a vector of sparse coding.

2. *Multimodal Feature Combinations*

Images are described by multimodal features. The multiple features: $X = \{X^i = [x_1^i \ldots\ldots x_n^i] \in R^{m_i * n}\}_{i=1}^t$, in which each representation X^i is a feature matrix from view i, L^j is the constructed hypergraph Laplacian matrix for the j[th] view, and λ_j is the weight. $A^j = [a_1^j, a_2^j, \ldots\ldots\ldots. a_j^s], (A^j \in R^{d*s})$ is a specified code for j[th] view.

3. *Implementations for Sparse Codes.*

Algorithm:

Input: Basis matrix A; k[th] image: x_k; sparse codes C; the weights σ=[1/t,……………..,1/t]; the L^, α and β.

Output: Sparse codes for image x_k : c_k.c_k is the k[th] column in C.

Step1: For the vector c_k, ϑ, $\hat{L_k}$ and γ_{ck}, adopts c_k^r, ϑ_r, $L_k^{\hat{r}}$ and γ_{ck}^r describe the r[th] entries.

Step2: Initialize ϑ : = sign(c_k), and find the active set of c_k : $(c_k)_{active}$: = Find $(c_{k\neq 0})$. $\vartheta_r \in$ {-1,0,1} is obtained by sign (c_k^r) , which is the sign of the r[th] entry of c_k.

Step3: From zero coefficients of c_k, select $p :=$ arg $max_r|\gamma_{c_k}^r|$.

3.1 If $\gamma_{c_k}^p > \alpha$, set $\vartheta_p := -1$, active set:={p}∪ active set.

3.2 If$\gamma_{c_k}^p > -\alpha$, set $\vartheta_p := 1$, active set:={p}∪ active set.

Step 4: The details for this step are:

4.1 Denote A and $\gamma_{c_k c_k}$ be the sub matrix of A and $\gamma_{c_k c_k}$that preserve the columns corresponding to the active set.

4.2 Denote $\bar{c_k}$ and $\bar{\vartheta}$ be the sub vectors of c_k and ϑ corresponding to the active set.

4.3 Obtain the solution by solving the unconstrained QP problem : $\underline{min_{c_k}} Q(\bar{c}_k) + \alpha\bar{\vartheta}^T c_k$.

The analytical solution is:

$\bar{c}_{k_{new}} = \bar{\gamma} c_k c_k)^{-1}[2\bar{A}^T x_k - 2\beta(C_{-k}L_{k,-k})_{active} - \alpha\bar{\vartheta}]$

4.4 A discrete line search is conducted on the closed line segment from $\bar{c_k}$to $\overline{c_{k_{new}}}$.

4.5 The zero coefficients are removed from the active set, and ϑ := sign(c_k) is updated.

Step 5: The optimality conditions are justified:

5.1 Check the optimality condition for non zero coefficients: $\gamma_{c_k}^j + \alpha\, sign(c_k^j) = 0, \forall c_k^j \neq 0.$ If the condition is false , Jump to step 5; else condition 5.2.

5.2 Check the optimality condition for zero coefficients : $|\gamma_{c_k}^j| < \alpha, \forall c_k^j \neq 0$. If condition 5.2 is not satisfied, jump to 4; otherwise return c_k as the solution, and update the sparse codes C.

4. *Implementations for Obtaining Weights*

Algorithm: Obtaion Optimization

Input: The multimedia image sets X; the multimodal bases matrix A; the initialization of C, $\hat{L}, \alpha, \beta$.

Output: The sparse codes C and the optimal weights λ

Step 1: Initialize the weights λ = [1/t,………1/t].

Step 2: Repeat

Step 3: Calculate sparse codes C according to the sparse code algorithm.

Step 4: Calculate λ_j as

$$\lambda_j = \frac{({}^1/_{tr}\,(CL^jC^T))^{1/(z-1)}}{\sum_{j=1}^t ({}^1/_{tr}\,(CL^{(j)}C^T))^{1/(z-1)}}$$

Step 5: Until convergence.

5. Time Complexity Analysis

The time complexity Ω of efficient sparse coding is lower than some state-of-art sparse coding methods. Therefore click prediction method obtains state-of-art performance of time complexity.

Table 1. Comparison table for web image re-ranking

METHODS	METHODOLO-GIES	APPRO-ACHES	ADVAN-TAGES
Active	-Collect labeling information from user to obtain specified semantic space. -Localize the visual characteristicsof the user intentions inspace.	-SInfo Based Sample Selection. -LGD Reduction Algorithm.	-Reduce user labeling efforts. -To satisfies the user's intention. -Learn the user's intention more extensively and completely.
Bag Based	-Partition images into clusters using textual and visual features. -Generalized multiinstance(GMI) framework. -Treats each cluster as Bag and images as instances.	-MI, GMI. -Weakbag Annotation -Clustering Algorithm.	-Textual and visual features are efficiently extacted. -Effective to address the ambiguities

Query-Specific Semantic Signatures.	-At offline, learns different semantic spaces for different query keyword. -At online, images reranked by comparing the semantic signatures obtained from semantic space specified by the query keyword.	-Query specific semantic signature. -Visual/textual features at online and offline stage.	-Improve accuracy and efficiency. -At online stage, efficient computational cost. -At offline stage, accuracy at the cost of storage.
Click prediction using multimodal sparse coding.	-MHGSC for image clickprediction, re-ranking images. -Hypergraph and hyperedges preserve the local smoothness of the constructed sparse codes. -Optimization is performed. -Voting strategies used to describe the predicted clicks. -Graph based schema.	-MHGSC. -Early and late fusion. -Optimize. -Voting strategy.	-Optimized web image reranking. -Improved and fast web image re-ranking. -Highly satisfies the user intention. -Minimize reconstruction errors.

CONCLUSION

Web images are associated with rich semantic textual descriptions. The semantic textual descriptions of web images has been widely used in all search engines. The retrieval performance can be very poor, so web image re-ranking methods are used.

Active re-ranking to target the user's intention effectively and efficiently. Active sample selection strategy and a dimension reduction algorithm, to reduce labeling efforts and to learn the visual characteristics of the intention respectively. The SInfo active sample selection strategy takes both the ambiguity and the representativeness. The LGD dimension reduction algorithm transfers the local information in the domain of the labeled images domain to the whole image database.Bag based framework partitioned the relevant images into clusters. To address the ambiguities on the instance labels in both positive and negative bags, GMI to enhance retrieval performance, in which the labels from the bag level have been propagated to the instance level. An automatic bag annotation method to automatically find positive and negative bags for training classifiers and achieve the best performance.

In query-specific semantic spaces, the visual features of images are projected into their related semantic spaces automatically learned through keyword expansions. Multimodal hyper graph learning based sparse coding method for the click prediction of images. The obtained sparse codes can be used for image re-ranking by integrating them with a graph-based schema. The hyper edge in a hyper graph helps to preserve the local smoothness of the constructed sparse codes. Then, optimization procedure is performed and the weights of different modalities and sparse codes are simultaneously obtained using this optimization strategy. Finally, a voting strategy is used to predict the click from the corresponding sparse code.

ACKNOWLEDGEMENTS

I would like to express my gratitude to all those who helped me to complete this survey. I would like to thank the faculty members , friends for giving me the opportunityand facilities to undertake the study of this topic and for its successful completion.

I am deeply indebted to my Guide, Asst Prof. Mr. Chitharanjan K, whose help, stimulating suggestions and encouragement helped me throughout the duration of the study. His technical knowledgeand valuable hints were of immense help to me for carrying out the study.

REFERENCES

1. D. T. K. L. MengWang, Hao Li and X.Wu, "Multimodal graph-based reranking for web image search," IEEE Trans on Image Process., vol. 21, No. 11, Nov 2012.
2. X.-S. H. X. W. XinmeiTian, Dacheng Tao, "Active reranking for web image search," IEEE Trans on Image Process., vol. 19, No. 3, March 2010.
3. I. W. T. L. Duan, W. Li and D. Xu, "Improving web image search by bag based re-ranking," IEEE Trans. Image Process., vol. 20, no. 11, pp. 3280–3290, Nov. 2011.
4. K. L. X. T. Xiaogang Wang, Shi Qiu, "Web image re-ranking using query-specific semantic signatures," IEEE Trans on Pattern Analysis and machine intelligence, Vol. 36,No. 4, April2014.
5. A. H. Linjun Yang, "Learning to rerank web images,"IEEE Computer Society, 2013.
6. D. T. Jun Yu, Yong Rui, "Click prediction for web image reranking using multimodal sparse coding," IEEE Trans on Image Process., vol. 23, no. 5, May 2014.
7. Y. Liu and T. Mei, "Optimizing visual search reranking via pairwise learning," Trans on Multimedia, vol. 13, NO.2, April 2011.