Survey on Multiple Query Content Based Image Retrieval Systems

Abeer Almohamade *amu.mohammadi@gmail.com*
College of Computer and Information Sciences,
Computer Science Department, King Saud University,
Riyadh, Saudi Arabia

Ouiem Bchir *ouiem.bchir@gmail.com*
College of Computer and Information Sciences
Computer Science Department, King Saud University,
Riyadh, Saudi Arabia

Abstract

This paper reviews multiple query approaches for Content-Based Image Retrieval systems (MQIR). These are recently proposed Content-Based Image Retrieval systems that enhance the retrieval performance by conveying a richer understanding of the user high-level interest to the retrieval system. In fact, by allowing the user to express his interest using a set of query images, MQIR bridge the semantic gap with the low-level image features. Nevertheless, the main challenge of MQRI systems is how to compute the distances between the set of query images and each image in the database in a way that enhances the retrieval results and reflects the high-level semantic the user is interested in. For this matter, several approaches have been reported in the literature. In this paper, we investigate existing multiple query retrieval systems. We describe each approach, detail the way it computes the distances between the set of query images and each image in the database, and analyze its advantages and disadvantages in reflecting the high-level semantics meant by the user.

Keywords: Content Based Image Retrieval, Multiple Query, Semantic Gap, User Interest.

1. INTRODUCTION

The proliferation of social networks along with the widespread of smart devices yields an exponential growth of digital image databases. This massive increase triggered the challenge of mining specific image among huge collections. Thus, image retrieval has become an active field of research [1]. Content Based Image Retrieval (CBIR) allows the user to express his interest by providing a query image that reflects the semantics he is looking for. The database is then mined according to the query content. For CBIR systems, the visual properties of an image are described using low-level feature descriptors [2] . More specifically, the descriptors can be the color, the texture and the shape properties of the image. These low-level features translate the visual content of the image into numerical vectors that allow quantitative estimation of the similarity between two images [3],[4],[5],[6],[7]. However, there is a gap between the semantic interest of the user and the extracted visual feature descriptor. For instance, as shown in Figure 1, if the user provides an image containing a red apple as a query, the retrieved images may contain red rose, red balloon or green apple depending on the visual descriptor used by the CBIR. Another issue in single image retrieval appears when the query image contains several objects. In fact, the retrieved images may not be relevant to the specific object meant by the user. Figure 2 shows a bike query image with green background. When retrieving this query image using Google image search engine, it returns images with green background but without a bike. This means that the "bike" semantics meant by the user are not conveyed to the retrieval system.

Recently, the multiple query retrieval system has been proposed in order to fill this semantic gap, and enhance the retrieval performance [8]. For multiple query retrieval system, the user expresses his interest using a set of query images. This yields a richer understanding of the user high-level interest to the retrieval system and bridges the semantic gap with the low-level image features. Unlike single query based CBIR system where the distances between the visual feature descriptor of the query image and the visual feature descriptor of each image in the database are computed and sorted in order to submit the most relevant images to the user, for multiple query image based CBIR systems, the challenge is how to compute the distances between the set of query images and each image in the database in a way that enhances the retrieval results and reflects the high-level semantic the user is interested in.

(a) (b) (c) (d)

FIGURE 1: Illustrative example: Red apple query. **(**a) Query image: red apple (b) Retrieved image: red rose (c) Retrieved image: red balloon (d) Retrieved image: green apple.

(a) (b)

FIGURE 2: Illustrative example: Bike query with green background image (a) Query image b) Images retrieved using a Google image search engine.

In this paper, we investigate existing multiple query retrieval systems. The rest of the paper is organized as follows. In section 2, we highlight the background topics related to retrieval systems. In section 3, we describe multiple image query approaches. Finally, in section 4, we conclude and outline future works.

2. BACKGROUNDS

Low-level feature transforms the visual content of an image to numerical values. According to the type of the extracted feature vector (color, texture, shape, etc.), a specific information about the image content is translated into a numerical vector. The resulting numerical vector is then conveyed as input to any computer based algorithm to represent the image. Thus, the choice of the visual descriptor is related to the type of information that we want to represent the image with. Various visual descriptors extracted to represent image content have been reported in the literature[9], [10], [11].

Despite the availability of advanced image low-level features, expressing the high-level semantic meaning of images remains a challenging task. For CBIR system, capturing user interest is even more challenging using one single query image. Multiple query images help to better express the user high-level interest and narrow the gap between the image semantics and its low-level visual descriptors [12][13]. In this section, we first describe the basic single query content based image retrieval system. Then, we present the multiple query content based image retrieval system.

In the single query content based image retrieval, a single query image is provided by the user as shown in figure 3. Its feature is extracted and the distance, usually Euclidean, is computed between the obtained feature and the feature extracted from all the images in the database. We should mention here that extracting features from the image database is done offline before the user provides the query image.

Once the pair wise distances are computed, they are sorted in an increasing order and the top images of the database with the smallest distances are returned as the retrieved ones. We notice that the result depends mainly on the choice of the feature and also on the type of the distance used to compute the pair wise distances.

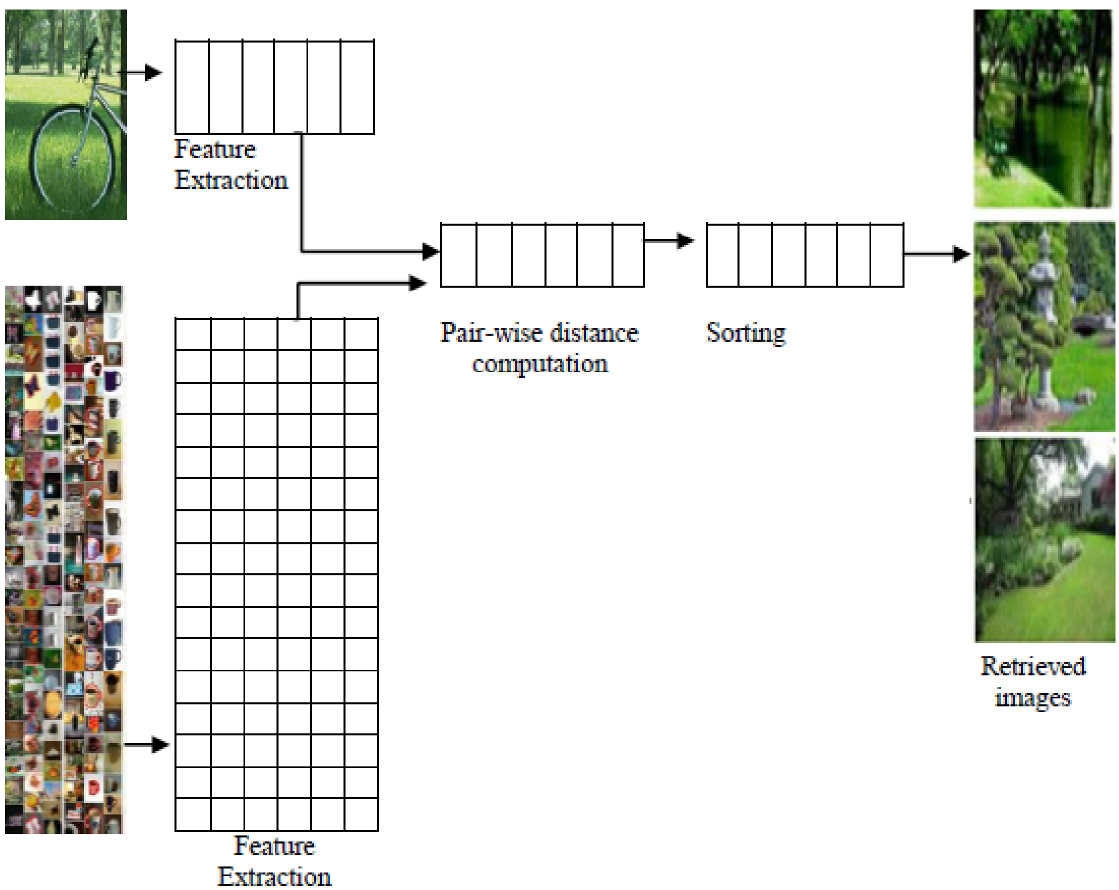

FIGURE 3: Single Query Content Based Image Retrieval.

For multiple query based CBIR, several query images are provided by the user. As shown in Figure 4, the low-level features are extracted from all query images. The visual descriptors from the image database are also extracted offline before the user provides the query images. The key idea of a multiple query system is the way the pairwise distances between the features extracted from the query images and the database image features are computed. In fact, for multiple query based image retrieval system, the distance is not computed between two feature vectors, but between a set of query feature vectors and feature vectors representing the database images. For this matter, several approaches have been reported in the literature. Once the pairwise distances are computed, similar steps, as for single query based retrieval system, are performed. In fact, the distances are sorted and images from the database corresponding to the lowest distances are returned as retrieval results. The multiple query based CBIR intends to provide more information to the retrieval engine than a single query based system does. However, the choice of the appropriate visual descriptor remains a critical task that affects drastic the retrieval results.

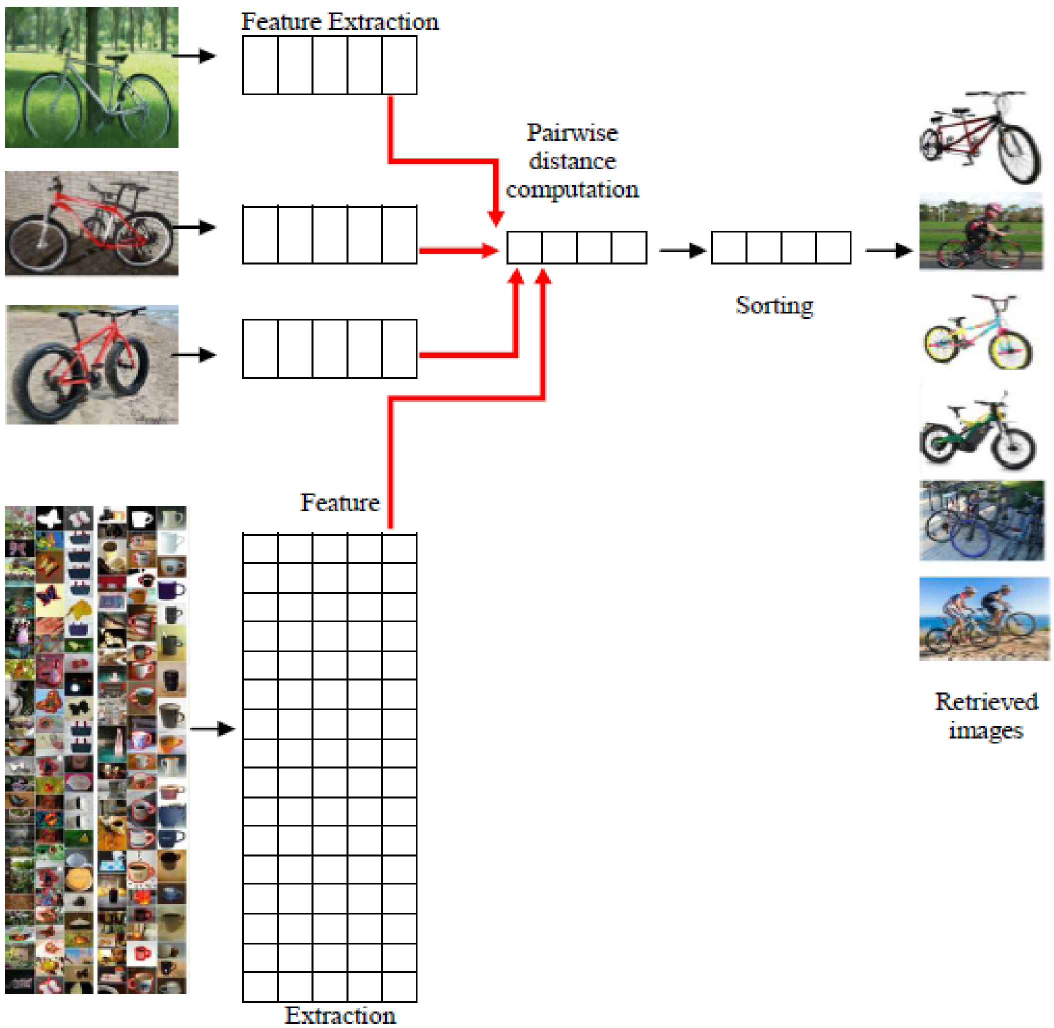

FIGURE 4: Multiple Query Content Based Image Retrieval.

3. MULTIPLE QUERY CONTENT BASED IMAGE RETRIEVAL

In the context of multiple query retrieval, a set of query images, is provided by the user in order to express his interest. As shown in Figure 4, the set of query images is then compared to the images in the database according to a defined distance that encapsulates the information conveyed by low-level features of the set of query images. Let $I_Q = \{I_Q^i (i = 1, \dots, M)\}$ be the set of M query images. Similarly, let $I_D = \{I_D^j (j = 1, \dots, N)\}$ be the image dataset of size N.

As mentioned in section 2, the key idea of a multiple query system is the way the pairwise distances between the features extracted from the query images and the database image features are computed. Once the pairwise distances are computed, similar steps, as for single query based retrieval system, are performed. For this matter, several approaches have been reported in the literature. In the following, we review these approaches.

3.1 Weighted Color and Texture Histograms

In [14], the authors propose an algorithm for CBIR using multiple query approach. The algorithm relies on multi-histogram intersection method that measures the similarity between the query images and each image in the database using color and texture histogram features. Given a set of query images where some of them reflect the texture information while the others reflect the color information, the distance between the set of query images $I_Q$and an image $\mathrm{I}_\mathrm{D}^\mathrm{j}$ in the database is a weighted combination of the texture and color distance. It is computed as in (1)

$$\mathrm{D}\left(\mathrm{I}_\mathrm{Q}^1, \dots, \mathrm{I}_\mathrm{Q}^\mathrm{M}, \mathrm{I}_\mathrm{D}^\mathrm{j}\right) = \mathrm{w}_1 \mathrm{D}_\mathrm{C}\left(\mathrm{I}_\mathrm{Q}^1, \dots, \mathrm{I}_\mathrm{Q}^\mathrm{M}, \mathrm{I}_\mathrm{D}^\mathrm{j}\right) + \mathrm{w}_2 \mathrm{D}_\mathrm{T}\left(\mathrm{I}_\mathrm{Q}^1, \dots, \mathrm{I}_\mathrm{Q}^\mathrm{M}, \mathrm{I}_\mathrm{D}^\mathrm{j}\right) \quad (1)$$

where $\mathrm{D}_\mathrm{C}(\mathrm{I}_\mathrm{Q}^1, \dots, \mathrm{I}_\mathrm{Q}^\mathrm{M}, \mathrm{I}_\mathrm{D}^\mathrm{j})$ is the distance based on color histogram similarity using multiple histogram intersection [15], and $\mathrm{D}_\mathrm{T}(\mathrm{I}_\mathrm{Q}^1, \dots, \mathrm{I}_\mathrm{Q}^\mathrm{M}, \mathrm{I}_\mathrm{D}^\mathrm{j})$ is the distance based on texture histogram similarity using multiple histogram intersection. w_1 and w_2 are two parameters that indicate the relative importance of each feature. The authors in [14] suggest those cross-validation technique on labeled data to learn these two parameters. However, this implies that these weights do not represent the user interest, but rather the intrinsic content characteristics of the images in the considered database.

3.2 Image Grouper

In [16], the authors outline an image retrieval algorithm using multiple query images. It is based on expressing the query using one set of images that are relevant to the user semantic, and another set of images that are irrelevant to the user semantic. These two sets are called multiple positive groups and multiple negative groups, respectively. The distance between the set of query images I_Q and an image $\mathrm{I}_\mathrm{D}^\mathrm{j}$ from the database is estimated as the distance between the images in the database and the mean of each positive group using the Fisher's Discriminant analysis (FDA) [17]. More specifically, it is estimated as follows

$$\mathrm{D}\left(\mathrm{I}_\mathrm{Q}^1, \dots, \mathrm{I}_\mathrm{Q}^\mathrm{M}, \mathrm{I}_\mathrm{D}^\mathrm{j}\right) = \mathrm{A}^\mathrm{T}\left(\mathrm{mean}_{\mathrm{i} \in \mathrm{PG}_\mathrm{t}}(\mathrm{F}_\mathrm{Q}^\mathrm{i}) - \mathrm{F}_\mathrm{D}^\mathrm{j}\right)\mathrm{A} \quad (2)$$

where PG_t is the positive group t, $\mathrm{F}_\mathrm{Q}^\mathrm{i}$ and $\mathrm{F}_\mathrm{D}^\mathrm{j}$are the feature vectors of the query image $\mathrm{I}_\mathrm{Q}^\mathrm{i}$, and the database image $\mathrm{I}_\mathrm{D}^\mathrm{j}$, respectively. A is the discriminating transformation matrix. In order to determine A, first the sum of the within scatter matrix S_W of each positive group $\mathrm{PG}_\mathrm{t}$and the scatter matrix $\mathrm{S}_\mathrm{PN}$of positive-negative groups PN_t are computed. Then, as in Fisher Discriminant Analysis FDA[17] , the discriminating transformation matrix A is solved as the eigenvectors of the largest eigenvalues of $\mathrm{S}_\mathrm{PN}^{-1}\mathrm{S}_\mathrm{W}$. As stated above, the user interest is expressed using a set of positive group images and a set of negative group images. This multiple query formulation is tedious, time consuming, and may not be practical for the user.

3.3 Minimum Weighted Distance Combination

In [18], the authors combine structure, color, and texture features using linear weighted summation. The corresponding weights are learned in order to enhance the retrieval results. The distance between the set of query images I_Q and imageI_D^j from the database is defined as the minimum distance between each query image $\mathrm{I}_\mathrm{Q}^\mathrm{i}$and the image I_D^j. It is expressed as

$$D\left(I_Q^1, \dots, I_Q^M, I_D^j\right) = \min_i\left(w_1 D_C\left(I_Q^i, I_D^j\right) + w_2 D_T\left(I_Q^i, I_D^j\right) + w_3 D_S\left(I_Q^i, I_D^j\right)\right) \tag{3}$$

where $D_C\left(I_Q^i, I_D^j\right)$, $D_T\left(I_Q^i, I_D^j\right)$, and $D_S\left(I_Q^i, I_D^j\right)$ are the distances between query image I_Q^i and image I_D^j using color, texture and structure features, respectively. w_1, w_2, and w_3 are their respective weights.

As stated above the combination weights are learned in such a way that they increase the retrieval performance. However, as in [14], the learned weights depend on the considered dataset. Thus, a learning process is required for each dataset. Besides, these weights do not reflect accurately the user semantic interest but rather the visual properties of the dataset.

3.4 Standard Deviation Based Weights

In [19], the distance between a query image $I_Q^i$and an image $I_D^j$is expressed as a linear weighted combination, and the distance between the set of query images $I_Q$and the an image I_D^j is defined as the minimum distance between each query image $I_Q^i$and the imageI_D^j. It is formulated as

$$D\left(I_Q^1, \dots, I_Q^M, I_D^j\right) = \min_i\left(\sum_k w_{ik} D_k\left(I_Q^i, I_D^j\right)\right) \tag{4}$$

where $D_k\left(I_Q^i, I_D^j\right)$ is the distance between the query image I_Q^i, and image I_D^j with respect to feature k, and w_{ik} is the corresponding relevance weight.

Since the standard deviations is expected to be small for similar images, and large for dissimilar ones, the weights w_{ik} are defined in [19] as

$$w_{ik} = \frac{1}{\sigma_{ik}^{\beta}}\left(\sum_{lm} \sigma_{lm}^{-\beta}\right)^{-1} \tag{5}$$

where β is a parameter that tunes the weights importance. We should notice here that in order for the standard deviation to be significant, it has to be computed over an important size of query images set. However, this makes the querying process non practical for users.

3.5 MindReader

The researchers in [4] propose to learn the optimal query image I_Q^t and a Mahalanobis distance [20] based on the query images set $I_Q = \left\{I_Q^i (i = 1, \dots, M)\right\}$ and their corresponding goodness scores $v_i (i = 1, \dots, M)$. The distance between an image query $I_Q^i$and an image in the database I_D^j is defined as

$$D\left(I_Q^t, I_D^j\right) = \left(F_Q^t - F_D^j\right)^T A\left(F_Q^t - F_D^j\right) \tag{6}$$

where F_Q^t and $F_D^j$are the feature vectors of the optimal query I_Q^t, and the database image I_D^j respectively, and A is a matrix defining the Mahanolobis distance. In order to learn the optimal feature vector F_Q^t, and the Mahanalobis matrix A, the authors in [4] minimize the following objective function

$$\min_{i, F_Q^t} \sum_{j=1}^{N} v_i \left(F_Q^t - F_D^j\right)^T A\left(F_Q^t - F_D^j\right) \tag{7}$$

subject to

$$\det(A) = 1 \tag{8}$$

Using the Lagrange multiplier technique [21], the minimization of the considered objective function gives

$$F_Q^t = \frac{\sum_{j=1}^{N} v_j \ F_D^j}{\sum_{j=1}^{N} v_j} \tag{9}$$

and

$$A = \det(C)^{\frac{1}{n}} C^{-1} \tag{10}$$

In (10), C is the covariance matrix of the image database feature vectors F_D^j. As stated above, the user has to express his interest as a set of query images scored according to their goodness. Moreover, the user has to provide a large number of query images so that the learned Mahalanobis matrix reflects accurately the user high level sematics. Besides, when the dimension of the considered feature is large, the Mahalanobis matrix computation gets highly expensive in terms of time complexity.

3.6 Logic AND Based Distance

The authors in [22] adopt the Euclidean distance in order to mine the image database. They assume that the relation between the set of query images and a given image in the database should an AND logical operation because retrieved images have to be similar to all query images. This implies the following definition of the distance between a set of query images I_Q, and image I_D^j:

$$D\left(I_Q^1, \dots, I_Q^M, I_D^j\right) = \max_i \left(ED\left(I_Q^i, I_D^j\right)\right) \tag{11}$$

where $ED\left(I_Q^i, I_D^j\right)$ is the Euclidean distance between image query I_Q^i and the database image I_D^j. The proposed approach in [22] does not consider feature weighting and assumes that all features are equally important. This means that all query images are assumed to provide the same kind of information.

3.7 Multi-Feature Query

In [23], the authors propose a multiple query retrieval system based on multiple low-level features. This CBIR system is based on a combined logic AND and logic OR distance. In fact, it considers an OR relation between the distances of a given query image I_Q^i and a given database image I_D^j with respect to the different features. Also, it adopts an AND relation between the distances of the different query images to a given image in the database. The distance proposed in [23] is expressed as

$$D\left(I_Q^1, \dots, I_Q^M, I_D^j\right) = \max_i \left(\min_s D_s\left(I_Q^i, I_D^j\right)\right) \tag{12}$$

where $D_s\left(I_Q^i, I_D^j\right)$ is the distance between image query I_Q^i and the database image I_D^j according to the feature s. The approach proposed in [23] does not consider feature weighting and assumes that only one feature is important by discarding the other descriptors.

3.8 Distance Combination

The linear distance combination approach in [24] assumes that the user interest is expressed using a set of query images and their respective weights or score of goodness [4]. The distance between the set of query images I_Q^i and any image I_D^j is a weighted sum over all distances between query images and image I_D^j. It is formulated as

$$D\left(I_Q^1, \ldots, I_Q^M, I_D^j\right) = \sum_{i=1}^{M} v_i \, D\left(I_Q^i, I_D^j\right)^t \qquad (13)$$

where v_i is the score of goodness of image query I_Q^i , $D\left(I_Q^i, I_D^j\right)$ is the distance between image query I_Q^i and the image database I_D^j, and t is the power of the distance which is equal or greater than 1. We should mention here that the score of goodness v_i are set by the user in order to reflect his interest.

3.9 Discussion

Approach	Feature weighting	Manually set feature weight	Manually set image query goodness score	Feature weights learning	Large number of query images	Color histogram and texture histogram only	Positive and negative groups required
Weighted color and texture histograms [14]	yes	Yes	No	no	no	yes	no
Image Grouper [16]	yes	No	No	yes	yes	no	yes
Minimum weighted distance combination [18]	yes	Yes	No	no	no	no	no
Standard deviation based weights [19]	yes	No	No	yes	yes	no	no
MindReader algorithm [20]	no	No	Yes	no	yes	no	no
Logic AND based distance algorithm [22]	no	No	No	no	no	no	no
Multi-feature query algorithm [23]	no	No	No	no	no	no	no
Distance combination algorithm [24]	no	No	Yes	no	no	no	no

TABLE 1: Characteristics of existing multiple query content based image retrieval approaches.

Table 1 summarizes the characteristics of existing multiple query content based image retrieval approaches. As stated above, some multiple query image retrieval systems do not consider weighting the features or the query images ***Logic AND based distance algorithm*** [22] ***Multi-feature query algorithm*** [23] and rather one of the query images is selected as being the more informative among the set of query images and the information provided by the remaining query images is not used. Other approaches like in ***MindReader algorithm*** [20] and in ***Distance combination algorithm*** [24] are based on the user scoring of the query images which is unpractical for the user of the retrieval system. Another unpractical aspect of some existing multiple query retrieving systems **Weighted color and texture histograms** [14] ***Image Grouper*** [16] ***Minimum weighted distance combination*** [18] is the need to provide an important number of query images in order to learn appropriate feature weights. Moreover, the feature weights learned in **Weighted color and texture histograms** [14] and in ***Minimum weighted distance combination*** [18] are based on cross-validation on a specific dataset. Thus, a learning process is required for each dataset and the learned weights do not reflect the user semantic interest but rather the visual properties of the dataset.

4. CONCLUSIONS AND FUTURE WORKS

In this paper, we reviewed existing multiple query content based image retrieval systems. We concluded that the key point of MQIR system is the choice of the distance between the set of query images and each image in the database and we see how the computing distance is effect on retrieval . In order to enhance the retrieval results by filling the gap between the semantics meant by the user and the low level features, several distances have been designed. They focus either on selecting and/ or weighting the set of query images and/or the low level features. Although several interesting approaches have been reported in the literature that enhances the retrieval results and reflects the high-level semantic the user is interested in, the practical aspect for the user has to be considered further. In fact, we observe that 2 to 3 query images should be enough to express the high level semantics meant by the user. From this small set of query images, weights for both the image queries and the low level features should be learned automatically. Thus, as future work, we intend designing same MQRI system structure but a novel distance Equation then compare it with the existing system with using a non cumbersome number of query images.

5. REFERENCES

[1] Ben Ismail M. M. "A Survey on Content-based Image Retrieval." International Journal of Advanced Computer Science and Applications, Vol. 8, No. 5, pp. 159-170, 2017.

[2] J. Komal , V. Akhilesh, G. Savita, Swati Goel. "A Survey on Recent Image Indexing and Retrieval Techniques for Low-Level Feature Extraction in CBIR Systems," in IEEE International Conference on Computational Intelligence & Communication Technology (CICT), 2015.

[3] Jing-Ming Guo J. M, H. Prasetyo, J. H. Chen. "Content-Based Image Retrieval Using Error Diffusion Block Truncation Coding Features," IEEE Transactions on Circuits and Systems for Video Technology, vol.25 (3), pp. 466-481, 2015.

[4] S. D. Thepade and Y. D.Shinde "Robust CBIR using sectorisation of hybrid wavelet transforms with Cosine-Walsh, Cosine-Kekre, Cosine-Hartley combinations," in International Conference on Pervasive Computing (ICPC), 2015.

[5] N. Gupta, S. Das, S.Chakraborti ."Extracting information from a query image, for content based image retrieval," in Eighth International Conference on Advances in Pattern Recognition (ICAPR), 2015.

[6] P.P. Hassekar, R.R.Sawant. "Experimental analysis of perceptual based texture features for image retrieval," in InternationalConference onCommunication,Information & Computing Technology (ICCICT), 2015.

[7] N. S. Patil and S. D.Sawarkar" Semantic Concept Detection in Video Using Hybrid Model of CNN and SVM Classifiers," International Journal of Image Processing (IJIP), vol.13(2), pp. 13-28, 2019.

[8] K. J. Hsiao, J. Calder, A.O.Hero. "Pareto-Depth for Multiple-Query Image Retrieval," IEEE Transactions on Image Processing, vol.24(2), pp. 583-594, 2015.

[9] C. S. Won, D. K. Park, S. J.Park " Efficient use of MPEG-7 edge histogram descriptor," Etri Journal, vol .24, no. 1, pp. 23-30, 2013.

[10] H.Alraqibah, M. M. Ben Ismail, O.Bchir . "Empirical Comparison of Visual Descriptors for Content based X-ray Image Retrieval," in International Conference on Image and Signal Processing, Cherbourg, 2014.

[11] S.Otaiba, S.Qassim, O.Bchir, M. M.Ben Ismail." Empirical comparison of visual descriptors for multiple bleeding spots recognition in wireless capsule endoscopy video," in International Conference on Computer Analysis of Images and Patterns (CAIP), York, UK, 2014.

[12] S. M. Tahaghoghi, J. A.Thom , H. E. Williams. "Are two pictures better than one?," In Proceedings of the 12th Australasian database conference ,IEEE Computer Society, January, 2001, pp. 138-144).

[13] S. M.Tahaghoghi , J.A.Thom , H. E. Williams."Multiple example queries in content-based image retrieval," In International Symposium on String Processing and Information Retrieval (pp. 227-241). Springer, Berlin, Heidelberg, September 2002.

[14] J.Tang, S. Acton. "An Image Retrieval Algorithm Using multiple query Images," in Seventh International Symposium on Signal Processing and its Applications, Paris, France,2013.

[15] M. J. Swain and D. H. Ballard. "Color indexing," IJCV,vol. 7(1), pp.11–32,1991.

[16] M. Nakazato and T. S. Huang. "Extending Image Retrieval with Group-Oriented Interface," in the IEEE International Conference on Multimedia & Expo (ICME), Lausanne, Switzerland, 2002.

[17] S. Mika, G. Rätsch, J. Weston, B.Scholkopf, K. R. "Fisher Discriminant Analysis with Kernels," in IEEE Conference on Neural Networks for Signal Processing,1999.

[18] Q.Iqbal and J. K.Aggarwa. "Feature Integration, Multi-image Queries and Relevance Feedback in Image Retrieval," in International Conference on Visual Information Systems, Miami, Florida, 2003.

[19] R.Brunelli and O.Mich. "Image Retrieval by Examples, IEEE Transactions On Multimedia," vol. 2(3), pp. 164-170,2000.

[20] M.Hazewinkel. "Mahalanobis distance," Encyclopedia of Mathematics, Springer,2001.

[21] M.Hazewinkel. "Lagrange multipliers," Encyclopedia of Mathematics, Springer,1998.

[22] J.Simily and B.Kannan. "Multi Query Image Retrieval System with Application to Mammogram Images", International Journal of Advanced Research in Computer Science, vol.3(3), pp. 469-474, 2012.

[23] S.Nepal and M.V.Ramakrishna. "MultiFeature Query by Multiple Examples in Image Databases, Advances In Data Management," McGrawHill Publishing Company Ltd,2000.

[24] L.Zhu and A.Zhang. "Supporting multi-example image queries in image databases," IEEE International Conference on Multimedia and Expo, New York, NY,2000.