ISSN (Online): 2456-5717

International Journal of Advanced Research in Basic Engineering Sciences and Technology (IJARBEST)
Vol.3 Issue.6 June 2017

Efficient Online Multi-modal Distance Metric Learning with Application to Image Retrieval Applications

V. Karthika[1], Dr. M. Pushparani[2]
[1]M.Phil. Research Scholar, [2]Professor & Head,
[1, 2]Department of computer Science, Mother Teresa Women's University, Madurai, India.

Abstract— **Distance metric learning (DML) is an important technique to improve similarity search in content-based image retrieval. Despite being studied extensively, most existing DML approaches typically adopt a single-modal learning framework that learns the distance metric on either a single feature type or a combined feature space where multiple types of features are simply concatenated. Such single-modal DML methods suffer from some critical limitations: (i) some type of features may significantly dominate the others in the DML task due to diverse feature representations; and (ii) learning a distance metric on the combined high-dimensional feature space can be extremely time-consuming using the naive feature concatenation approach. To address these limitations, in this paper, we investigate a novel scheme of online multi-modal distance metric learning (OMDML), which explores a unified two-level online learning scheme: (i) it learns to optimize a distance metric on each individual feature space; and (ii) then it learns to find the optimal combination of diverse types of features. To further reduce the expensive cost of DML on high-dimensional feature space, we propose a low-rank OMDML algorithm which not only significantly reduces the computational cost but also retains highly competing or even better learning accuracy. We conduct extensive experiments to evaluate the performance of the proposed algorithms for multi-modal image retrieval, in which encouraging results validate the effectiveness of the proposed technique.**

Index Terms – **Distance Metric Learning (DML), online multi-modal distance metric learning (OMDML). Image retrieval**

I. INTRODUCTION

Multi-modal data is ubiquitous in web. Especially, with the vast prosperity of social media, data with multiple modalities is enjoying explosive growth. For example, in user-centric social networks (e.g., Facebook, Google Plus), users possess blogs, photos, friends circles. In photo sharing websites (e.g., Flickr, Pinterest), photos can be described by image contents, text tags and meta information like albums and groups. In video sharing website (e.g., Youtube), videos can be characterized by image frames, audio, and user comments. In music social network (e.g., iTunes Ping), songs are accompanied by acoustic features (e.g., rhythm and timbre), semantic features (e.g., tags, lyrics) and social features (e.g., artist reviews) [McFee and Lanckriet, 2011]. Information from dif ferent sources (text, image, video, audio, meta and social information) jointly reveal the fundamental characteristics of the study subjects from different views. Choosing a proper distance function or similarity measure for multi-modal data is crucial for many applications, including retrieval [Zhang et al., 2011; Zhen and Yeung, 2012], clustering [Bekkerman and Jeon, 2007; Qi et al., 2012],

classification [Nishida et al., 2012] and recommendation [Aizenberg et al., 2012; Baluja et al., 2008]. While various metric learning methods [Xing et al., 2002; Globerson and Roweis, 2006; Weinberger et al., 2006; Davis et al., 2007] defined on single data modality have been proposed, learning distance in the presence of multiple modalities remains largely unexploited. McFee and Lanckriet [2011] applied the multiple kernel learning technique for integrating heterogeneous feature modalities into a single unified similarity space. An ensemble of kernel transformations are learned given the labeled relative similarity comparisons. To our best knowledge, this is the only work regarding multi-modal distance metric learning. However, their method enjoys very limited scalability.

One of the core research problems in multimedia retrieval is to seek an effective distance metric/function for computing similarity of two objects in content-based multimedia retrieval tasks [1], [2], [3]. Over the past decades, multimedia researchers have spent much effort in designing a variety of low-level feature representations and different distance measures [4], [5], [6]. Finding a good distance metric/function remains an open challenge for content-based multimedia retrieval tasks till now. In recent years, one promising direction to address this challenge is to explore distance metric learning (DML) [7], [8], [9] by applying machine learning techniques to optimize distance metrics from training data or side information, such as historical logs of user relevance feedback in content-based image retrieval (CBIR) systems [6], [7].

Although various DML algorithms have been proposed in literature [7], [10], [11], [12], [13], most existing DML methods in general belong to single-modal DML in that they learn a distance metric either on a single type of feature or on a combined feature space by simply concatenating multiple types of diverse features together. In a real-world application, such approaches may suffer from some practical limitations: i) some types of features may significantly dominate the others in the DML task, weakening the ability to exploit the potential of all features; and (ii) the naive concatenation approach may result in a combined high-dimensional feature space, making the subsequent DML task computationally intensive.

II. LITERATURE SURVEY

Paper Title	Objective	Limitations
Online multi-modal distance learning for scalable multimedia retrieval.	We present a novel online learning framework for learning distance functions on multimodal data through the combination of multiple kernels	The limitations of MKPOE, this paper proposes a novel Online Multi-modal Distance Learning (OMDL) scheme, which exploits the local dependency of underlying data distributions to enhance the learning efficacy of Multiple Kernel Partial Order Embedding(MKPOE).

Learning to name faces: a multimodal learning scheme for search-based face annotation	We tackle this open problem by investigating a search-based face annotation (SBFA) paradigm for mining large amounts of web facial images freely available on the WWW.	Model-based face annotation techniques suffer from some common drawbacks, e.g., being difficult and expensive to collect large high-quality training data and being non-trivial for adding new training data
Libol: A library for online learning Algorithms.	We develop LIBOL as an easy-to-use online learning tool that consists a large family of existing and recent state-of-the-art online learning algorithms for large-scale online classification tasks. In contrast to many existing software for large-scale data classification, LIBOL enjoys significant advantages for massive-scale classification in the era of big Data nowadays, especially in efficiency, scalability, parallelization, and adaptability.	LIBOL is not only a machine learning toolbox, but also a comprehensive experimental platform for conducting online learning research.
View generation for multi view maximum disagreement Based active learning for hyper spectral image classification.	AL method, this study investigates the principles and capability of several approaches for the view generation for hyper spectral data classification, including clustering, random selection, and	The clustering is more complicated in small training set of data.

	uniform subset slicing methods, which are then incorporated with dynamic view updating and feature space bagging strategies.	
Online multiple kernel learning: Algorithms and mistake bounds	Online Multiple Kernels Learning (OMKL) that aims to learn a kernel based prediction function from a pool of predefined kernels in an online learning fashion. OMKL is generally more challenging than typical online learning because both the kernel classifiers and their linear combination weights must be learned simultaneously	Compared to the exiting methods for multiple kernel learning, online multiple kernel learning is computationally advan-tageous in that it only requires going through training examples once

LIBOL is an open-source library for large-scale online learning, which consists of a large family of efficient and scalable state-of-the-art online learning algorithms for large-scale online classification tasks. We have offered easy-to-use command-line tools and examples for users and developers, and also have made comprehensive documents available for both beginners and advanced users. LIBOL is not only a machine learning toolbox, but also a comprehensive experimental platform for conducting online learning research.

Disadvantage:

Easy-to-use command-line tools and examples for users and developers, and also have made comprehensive documents available for both beginners and advanced users.

III. IMAGE RETRIEVAL ANALYSIS

a) Content based image retrieval (CBIR)

It has gained much attention in the past decade. However, the gap between low-level features and high-level semantic meanings usually leads to poor performance for CBIR. Relevance feedback is a powerful tool to involve the user in the loop to enhance CBIR's performance. Recently, many RF methods have been introduced. Feature selection- based methods adjust weights associated with various dimensions of the feature space to enhance the importance of those dimensions that help in retrieving the relevant images and reduce the importance of those dimensions that hinder image retrieval. Alternatively, features can be selected by the boosting technique in which a strong classifier is obtained as a weighted sum of weak classifiers along the different feature dimensions.

Probabilistic model-based methods use entropy to minimize the expected number of iterations. discriminant analysis-based methods either find row dimensional subspace of the feature space, such that the positive and negative samples are well seperated after projection to this subspace or define a (1+x)-class problem (biased discriminant analysis and find a subspace within which to discriminant the one positive class and the unknown number of negative sample classes. More recently the direct kernel based discriminant analysis was developed and reported to outperform the BDA in both linear space and the nonlinear kernel space. Support vector machine (SVM)-based methods either estimate the density of positive instances or regard RF as a classification problem with the positive and negative samples as training sets. SVM active learning, which plays an important role in CBIR RF research, selects the samples near the SVM boundary and queries the use for labels, then, after training; the points near the SVM boundary are regarded as the most-informative images while the most-positive images are the farthest ones from the boundary on the positive side. Recently, SVM active learning is also combined with a multimodal concept-dependent process for CBIR, constrained similarity measure (CSM). CSM-based SVM learns a boundary that seperates all the images in the database into two clusters and the image inside the boundary are ranked by their euclidean distances to the query image.

Derived from one-class SVM in a biased SVM is proposed, which can better model the relevance feedback problem and reduce the performance degradation caused by the imbalanced data set problem i.e., the number of the positive feedback samples is much less than the number of the negative feedback samples. These conventional schemes have been successfully in solving some samples of the problems in CBIR RF.

b) Shape based image retrieval (SBIR):

In recent years, content based image retrieval has been studied with more attention as huge amounts of image data accumulate in various fields, e.g., medical images, satellite

images, art collections, commercial images and general photographs. Image databases are usually very big, and in most cases, the images are indexed only by keywords given by a human.

Although keywords are the most useful in retrieving images that a user wants, sometimes the keyword approach is not sufficient. Instead, Query-by-example or pictorial-query approaches make the system return similar images to the example image given by a user. The example images can be a photograph, user-painted example, or line-drawing sketch. In this method, images are retrieved by their contents: color, texture, shape, or objects.

Thus, the degree of similarity between query images and images in databases can be measured by color distribution, texture distribution, shape similarity, or object presence between the two images. There have been many works done with color and texture property. Searching for images using shape features has attracted much attention. Shape representation and description is a difficult task. This is because when a 3-D real world object is projected onto a 2-D image plane, one dimension of object information is lost.

As a result, the shape extracted from the image only partially represents the projected object. To make the problem even more complex, shape is often corrupted with noise, defects, arbitrary distortion and occlusion.

There are many shape representation and description techniques in the literature. Marr and Nishihara and Braddy have thoroughly discussed representation and sets of criteria for the evaluation of shape. Soffer and Samet proposed a pictorial query specification technique that enables the formulation of complex pictorial queries including spatial constraints between query- image objects which are predefined symbolic images and contextual constraints which specify how many objects should be in the target Image.

The predefined symbolic query- images are represented by shape feature, e.g., moment, circularity, eccentricity, rectangularity, etc. Folkers and Samet extended this tool to permit query-images that have spatial extent such as ellipses, rectangles, polygons, and B-splines.

The query-images are represented by Fourier descriptors which serve powerful boundary- shape representation tools because of invariance property in affine transformation. However, there is a limit to expressing an object by its boundary because the boundary itself does not represent inside shape feature of the object. In and, shapes were represented using a Fourier expansion of the function of their tangent angle and their arc length. The lower- order Fourier coefficients were then used to represent the shape. Lie determined points of high curvature of a shape and represented them in polar form. These methods were invariant to translation and scale. However, the Fourier descriptor has several shortcomings in shape representation.

Bernier and Landry use a polar transformation of the shape points about the geometric center of object, the distinctive vertices of the shape are extracted and used as comparative parameters to minimize the difference of shape distance from the center. But it was not designed to be tolerant to occlusion.

Seldom works have been done with shape similarity since we need to have good representation and description algorithm to use shape similarity in retrieving images and the state of the algorithms are still primitive. In this paper, shape-based image retrieval problem is handled especially in a colour image database. A program that extracts the proposed shape features from database images, compares these features with query image features and retrieves the image based on similarity between these two feature sets was implemented.

The shape feature representation's performance was evaluated in terms of how effective the above goal was met by testing the results from queries performed on a colour image database.

The basic idea is to use the centroid-radii model to represent shapes. In this method, lengths of the shape's radii from centroid to boundary are used to represent the shape. If a shape is used as feature, edge detection might be the first step to extract that feature. In our work, the canny edge detector is used to determine the edge of the object in the scene.

After the edge has been detected the important step is tracing the contour of the object in the scene. For this the edge image is scanned from four directions (right to left, left to right, top to bottom, bottom to top) and the first layer of the edge occurred is detected as image contour. To avoid discontinuities in object boundary the contour image is then re-sampled. After the object contour has been detected the first step in shape representation for an object is to locate the central point of the object

Fig 2.3 shows the retrieval results obtained using shape features for the query image 'aero'. All the 20 aero plane images are retrieved in the top 21 retrieved images. Other images retrieved with ranking 19, 22, 23, 24 and 25 are Watch and Barbie images. When top 25 retrieved images are considered for the query image aero plane16, all the 20 aero plane class images are retrieved within the 25 retrieved images with the retrieval efficiency of 100%.

Fig. 3.1 Image Retrieval in shape & Content based Approaches

Sl.no	Query image	Retrieval efficiency (%)		
1	Aero plane	89.00	94.25	97.50
2 2	Apple	49.50	54.75	60.75
3	Barbie	66.25	73.00	76.25
4	Bike	88.50	94.25	96.50
5	Book	66.25	71.00	73.75
6	Coin	80.00	88.50	93.25
7	Hammer	73.25	83.00	86.75
8	Pen	73.00	77.50	81.25
9	T- shirt	79.75	85.50	87.75
10	Watch	45.75	49.00	52.50
	Average retrieval Efficiency	71.17%	77.07%	80.625%

The above Table shows the retrieval efficiency, considering the Top 20, Top 25 as well as Top 30 retrieved images for each query image. It is important to note that 80.625% retrieval efficiency is obtained when Top 30 images are considered. The retrieval efficiency is calculated as follows: For example when top 25 retrieved images are considered for the query image Bike.0000, all the 20 Bike class images are retrieved within the 25 retrieved images. When the color image database is considered for comparison 71.17% retrieval efficiency is obtained for top 20 retrievals. For top 30 retrieval 80.625% of average retrieval efficiency is achieved.

Gabor wavelet:

This approach describes an image retrieval technique based on Gabor texture feature. Texture is an important feature of natural images. A variety of techniques have been developed for measuring texture similarity. Most techniques rely on comparing values of what are known as second-order statistics calculated from query and stored images. These methods calculate Measures of image texture such as the degree of contrast,

coarseness, directionality and regularity or periodicity, directionality and randomness. Alternative methods of texture analysis for image retrieval include the use of Gabor filters and fractal.

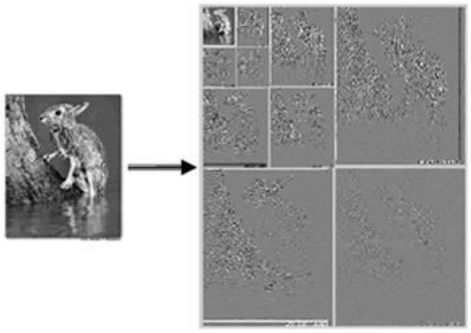

Fig: 3.2 Gabor Wavelet Analysis

Gabor filter (or Gabor wavelet) is widely adopted to extract texture features from the images for image retrieval, and has been shown to be very efficient. Manjunath and Ma have shown that image retrieval using Gabor features outperforms that using pyramid-structured wavelet transform (PWT) features, tree-structured wavelet transform (TWT) features and multi resolution simultaneous autoregressive model features.

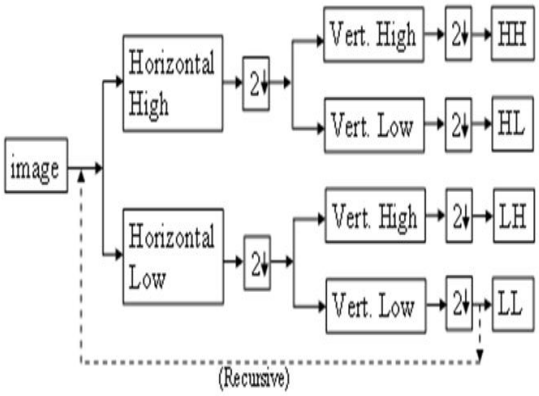

Fig. 3.3 Wavelet decomposition

Basically, Gabor filters are a group of wavelets, with each wavelet capturing energy at a specific frequency and a specific direction. Expanding a signal using this basis provides a localized frequency description, therefore capturing local features/energy of the signal. Texture features can then be extracted from this group of energy distributions. The scale frequency) and orientation tunable property of Gabor filter makes it especially useful for texture analysis. Experimental evidence on human and mammalian vision supports the notion of spatial-frequency (multi-scale) analysis that maximizes the simultaneous localization of energy in both spatial and frequency domains. Currently, most techniques make an explicit or implicit assumption that all the images are captured under the same orientations. In many

practical applications such as image retrieval, objects recognition etc, such an assumption is unrealistic. Some other techniques carry out rotation normalization, but they are computationally demanding. In this technique we propose a rotation normalization method that achieves rotation invariance by a circular shift of the feature elements so that all images have the same dominant direction.

In this approach, we describe texture representation based on Gabor transform, texture similarity calculation and rotation Normalization. After applying Gabor filters on the image with different orientation at different scale, we obtain an array of magnitudes:

A feature vector f (texture representation) is created using μmn and σmn as the feature components. Five scales and 6 orientations are used in common implementation and the feature vector is given by:

$$f = (\mu 00, \sigma 00, \mu 01, \sigma 01 \ldots \mu 45, \sigma 45).$$

Shows the energy map of the mean feature elements μmn for a straw texture image.

IV. Proposed Implementation

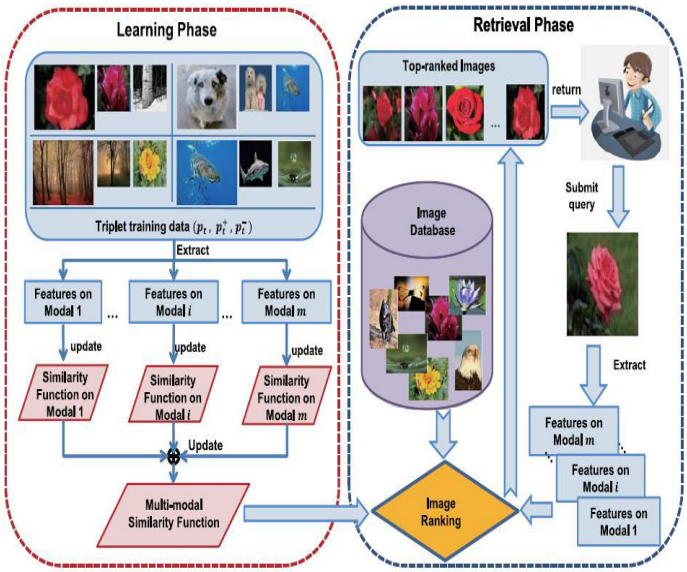

Fig. 4.1 System Architecture

We investigated a novel family of online multi-modal distance metric learning (OMDML) algorithms for CBIR tasks with the exploitation of multiple types of features. We pinpointed the serious limitations of traditional DML approaches in practice, and presented the online multi-modal DML method which simultaneously learns both the optimal distance metric on each individual feature space and the optimal combination of the metrics on multiple types of features. We further proposed the low-rank online multi-modal DML algorithm (LOMDML), which not only runs more efficiently and scalably, but also attains the state-of-the-art performance among all the competing algorithms as observed from our

extensive set of experiments. Our future work will extend the proposed framework for learning non-linear distance functions.

Advantage:

Which not only runs more efficiently and scalable, but also attains the state-of-the-art performance among all the competing algorithms as observed from our extensive set of experiments.

Diagrams:

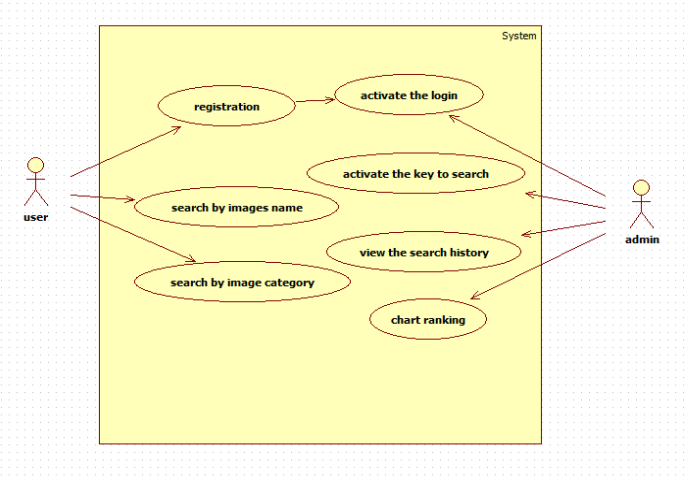

Fig. 4.2 Use Case Diagram for propose implementations

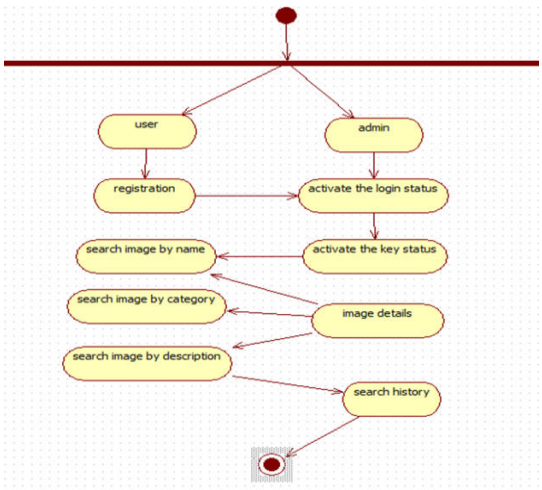

Fig. 4.3 Activity Diagram

ISSN (Online): 2456-5717

International Journal of Advanced Research in Basic Engineering Sciences and Technology (IJARBEST)
Vol.3 Issue.6 June 2017

V. RESULTS & DISCUSSIONS

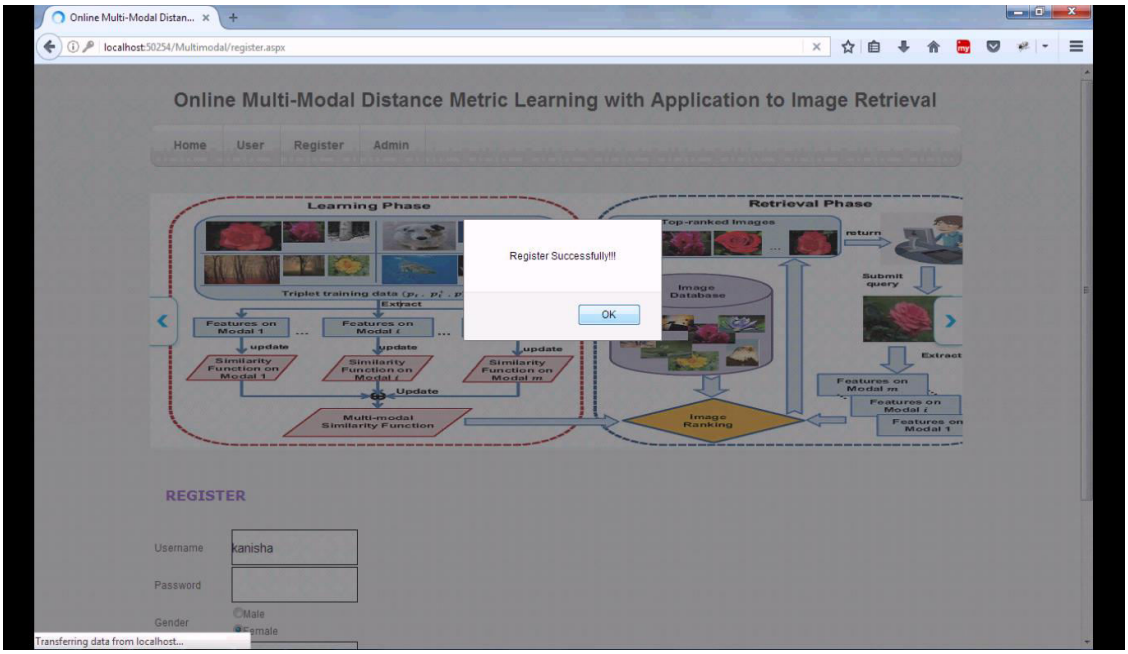

Fig. 5.1 Registration process completed

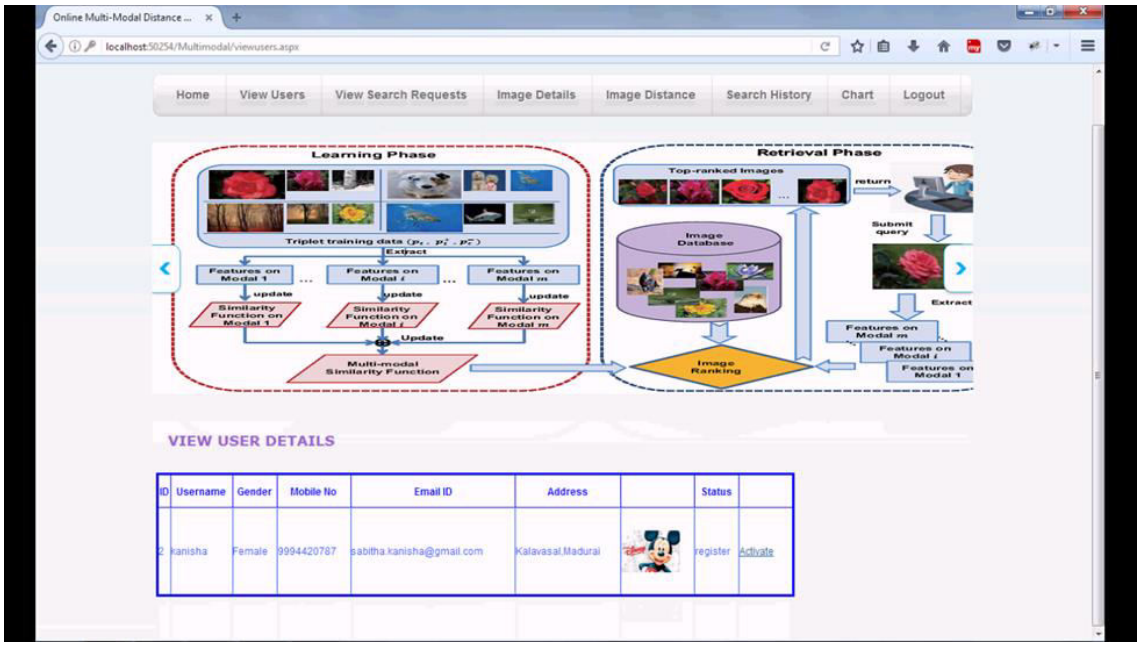

Fig.5.2 User image added

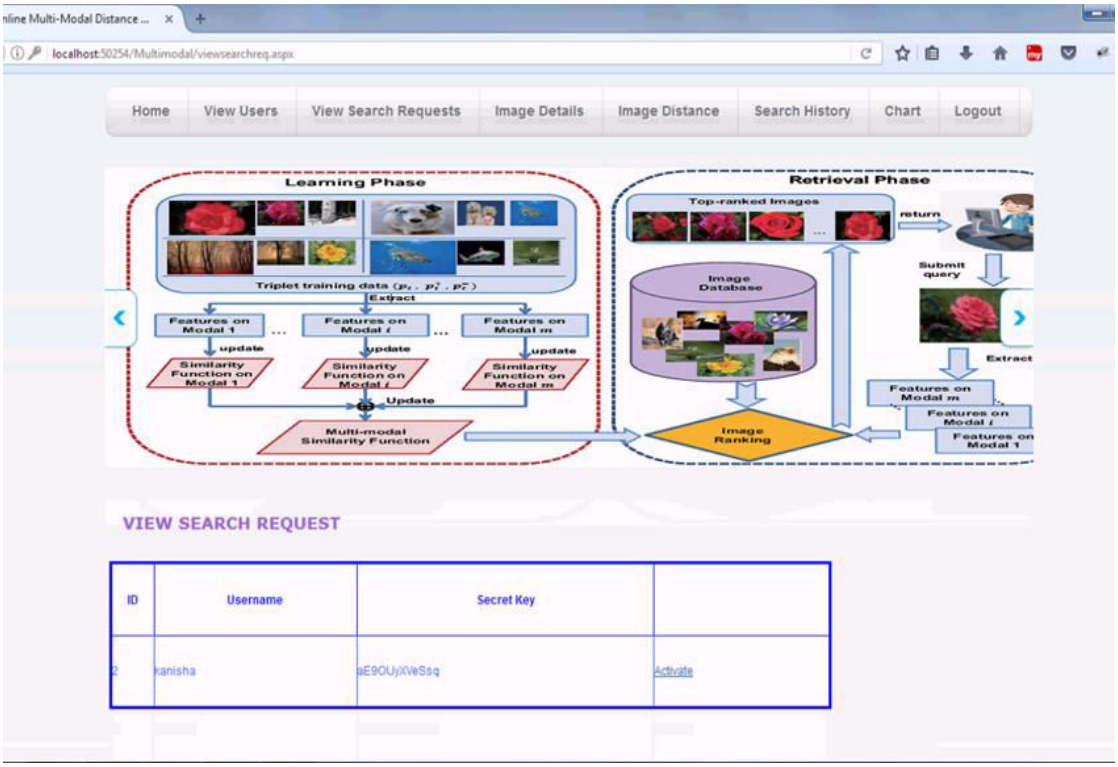

Fig. 5.3 View Search Request

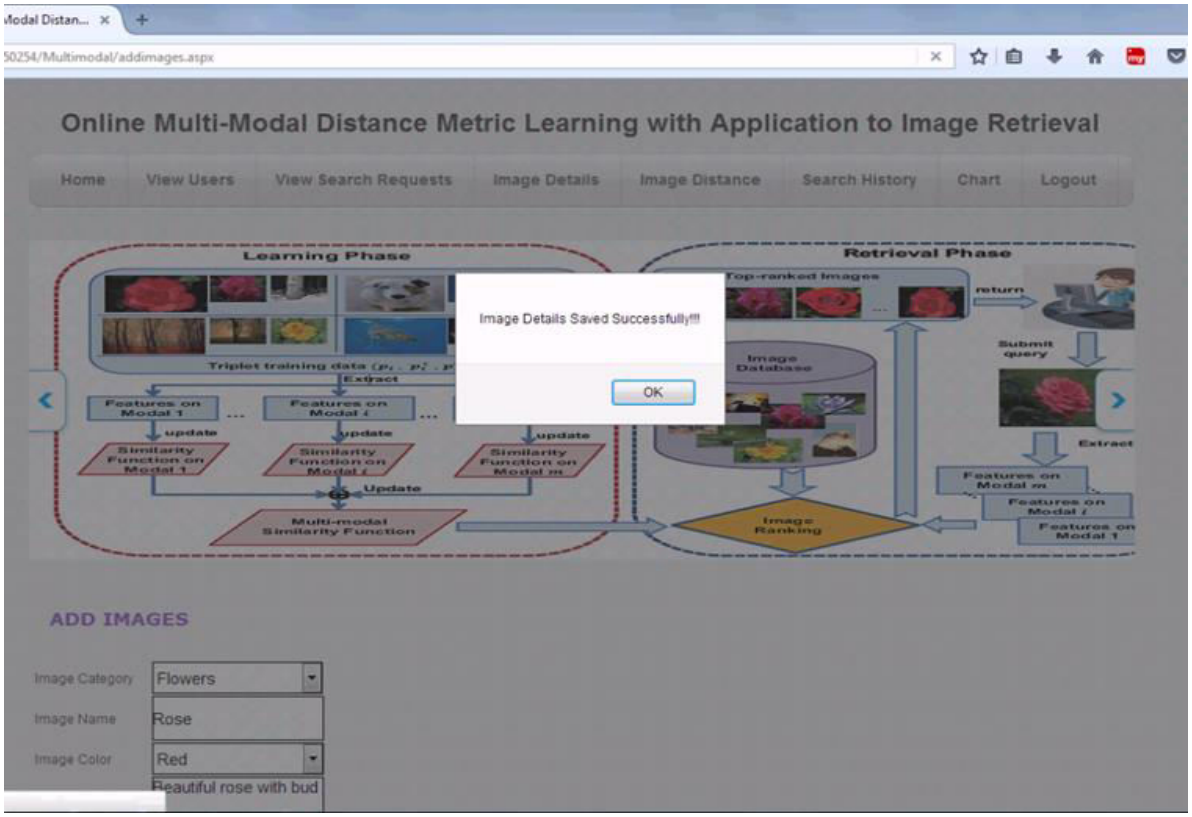

Fig. 5.4 Image Saved to DB

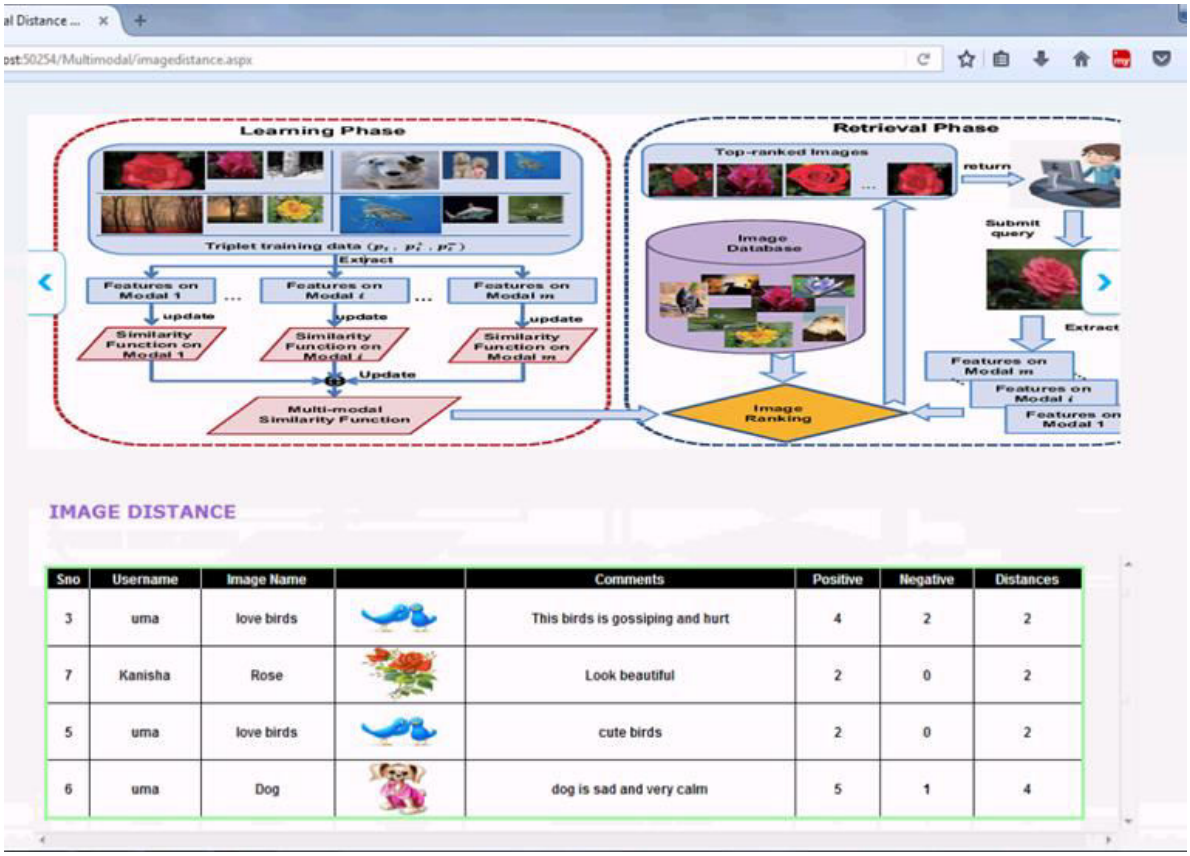

Fig. 5.5 Image distance

VI. CONCLUSION AND DISCUSSIONS

Finally we investigated a novel family of online multi-modal distance metric learning (OMDML) algorithms for CBIR tasks ion of multiple types of features. We pinpointed the serious limitations of traditional DML approaches in practice, and presented the online multi-modal DML method which simultaneously learns both the optimal distance metric on each individual feature space and the optimal combination of the metrics on multiple types of features. We further proposed the low-rank online multi-modal DML algorithm (LOMDML), which not only runs more efficiently and scalable, but also attains the state-of-the-art performance among all the competing algorithms as observed from our extensive set of experiments. Our future work will extend the proposed framework for learning non-linear distance functions.

REFERENCES

[1] M. S. Lew, N. Sebe, C. Djeraba, and R. Jain, "Content-based multimedia information retrieval: State of the art and challenges," Multimedia Computing, Communications and Applications, ACM Transactions on, vol. 2, no. 1, pp. 1–19, 2006.

[2] Y. Jing and S. Baluja, "Visualrank: Applying pagerank to large-scale image search," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 30, no. 11, pp. 1877–1890, 2008.

[3] D. Grangier and S. Bengio, "A discriminative kernel-based approach to rank images from text queries," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 30, no. 8, pp. 1371–1384, 2008.
[4] K. Jain and A. Vailaya, "Shape-based retrieval: a case study with trademark image database," Pattern Recognition, no. 9, pp. 1369–1390,1998.
[5] Y. Rubner, C. Tomasi, and L. J. Guibas, "The earth movers distance as a metric for image retrieval," International Journal of Computer Vision, vol. 40, p. 2000, 2000.
[6] W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 22, no. 12, pp. 1349–1380, 2000.
[7] S. C. Hoi, W. Liu, M. R. Lyu, and W.-Y. Ma, "Learning distance metrics with contextual constraints for image retrieval," in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, New York, US, Jun. 17–22 2006, dCA.
[8] L. Si, R. Jin, S. C. Hoi, and M. R. Lyu, "Collaborative image retrieval via regularized metric learning," ACM Multimedia Systems Journal, vol. 12, no. 1, pp. 34–44, 2006.
[9] S. C. Hoi, W. Liu, and S.-F. Chang, "Semi-supervised distance metric learning for collaborative image retrieval," in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Jun. 2008.
[10] G. H. J. Goldberger, S. Roweis and R. Salakhutdinov, "Neighbourhood components analysis," in Advances in Neural Information Processing Systems, 2005.
[11] K. Fukunaga, Introduction to Statistical Pattern Recognition. Elsevier, 1990.
[12] Globerson and S. Roweis, "Metric learning by collapsing classes," in Advances in Neural Information Processing Systems, 2005.
[13] L. Yang, R. Jin, R. Sukthankar, and Y. Liu, "An efficient algorithm for local distance metric learning," in Association for the Advancement of Artificial Intelligence, 2006.
[14] [14] A. K. Jain and A. Vailaya, "Image retrieval using color and shape," Pattern Recognition, vol. 29, pp. 1233–1244, 1996.
[15] B. S. Manjunath and W.-Y. Ma, "Texture features for browsing and retrieval of image data," Pattern Analysis and Machine Intelligence,
[16] IEEE Transactions on, vol. 18, no. 8, pp. 837–842, 1996.
[17] J. Sivic, B. C. Russell, A. A. Efros, A. Zisserman, and W. T. Freeman, "Discovering objects and their location in images," in IEEE Conference on Computer Vision and Pattern Recognition, 2005.
[18] J. Yang, Y.-G. Jiang, A. G. Hauptmann, and C.-W. Ngo, "Evaluating bag-of-visual-words representations in scene classification," in ACM International Conference on Multimedia Information Retrieval, 2007, pp. 197–206.
[19] D. G. Lowe, "Object recognition from local scale-invariant features," in IEEE International Conference on Computer Vision, 1999, pp. 1150– 1157.
[20] R. S. Mohammad Norouzi, David Fleet, "Hamming distance metric learning," in Advances in Neural Information Processing Systems, 2012.
[21] G. Chechik, V. Sharma, U. Shalit, and S. Bengio, "Large scale online learning of image similarity through ranking," Journal of Machine Learning Research, vol. 11, pp. 1109–1135, 2010.
[22] H. Chang and D.-Y. Yeung, "Kernel-based distance metric learning for content-based image retrieval," Image and Vision Computing, vol. 25, no. 5, pp. 695–703, 2007.
[23] R. Salakhutdinov and G. Hinton, "Semantic hashing," International Journal of Approximate Reasoning, vol. 50, no. 7, pp. 969–978, Jul. 2009. [Online]. Available: http://dx.doi.org/10.1016/j.ijar.2008.11.006
[24] H. Jegou, F. Perronnin, M. Douze, J. Sanchez, P. Perez, and C. Schmid, "Aggregating local image descriptors into compact codes," IEEE Trans. Pattern Anal. Mach. Intell., vol. 34, no. 9, pp. 1704–1716, Sep. 2012. [Online]. Available: http://dx.doi.org/10.1109/TPAMI.2011.235
[25] K. Chatfield, V. S. Lempitsky, A. Vedaldi, and A. Zisserman, "The devil is in the details: an evaluation of recent feature encoding methods," in BMVC, 2011, pp. 1–12.