Content Based Image Retrieval using Localized Multi-Texton Histogram

Muhammad Younas Qazi
Department of Computer Science
National University of Computer and Emerging Sciences
Lahore, Pakistan
Email: ganifbros@gmail.com

Muhammad Shahid Farid
Dipartimento di Informatica
Università Degli Studi di Torino
Corso Svizzera 185, 10149 Torino, Italy
Email: farid@di.unito.it

Abstract—This paper presents a simple yet efficient image retrieval technique that defines image feature descriptors using localized multi-texton histogram. The proposed technique extracts a unique feature vector for each image in the image database based on its shape, texture and color. First, the image is divided into smaller equal size blocks and then for each block texture orientation is computed independently. Second, each block is filtered with a set of predefined textons and finally, a co-occurrence relation is established from the orientation and the filtered texton image. This relationship in turn provides a powerful feature vector. To retrieve similar images, the feature vector of the query image is computed and compared with the feature vectors of the stored images using Euclidean distance measure. The proposed algorithm is tested on standard image dataset Corel 1000 for accuracy and recall with favorable results. It is also compared with existing state of the art Context Based Image Retrieval algorithm and showed convincing results.

Index Terms—Image retrieval, image query, multi texton histogram, query by image content

I. Introduction

Last couple of decades has witnessed excessive and rapid development in multimedia tools and technologies like image capturing devices and other electronic gadgets. The use of these devices has flooded the image and videos repositories on the internet and these repositories contain rich information in for of images. People often are in need to find a particular image or set of images from image repositories located locally or globally for their various tasks. Searching for image manually is not feasible, an automatic image retrieval technique is required that can search these huge repositories for images with particular color, shape or concept. In early days, the images were stored and indexed with tags describing the details of the image like its color and theme. But with exponentially increasing size of the image repositories on the internet, tagging is not a desired option. Secondly, all tags are not well defined; people tag their photos with ambiguous and sometimes misleading text, so searching on the basis of tags may not produce required results. Thirdly, users desire to search typical themes or concepts which cannot be done with tagging, special image search algorithm that can evaluate the semantics of the image are required.

Users from different domain of fields prefer efficient image searching and browsing tools for example in crime prevention, medical imaging and diagnosis, art and publicity and in military. To get accurate and efficient images every time, many techniques have been proposed which usually are categorized into three major categories: text-based, content-based and semantic-based retrieval. The text-based image retrieval system depends on the human categorization of the image. The image is categorized according to its appearance which may vary from one user to user according to their perception about the image. It may increase the chances of error for other user or may produce uncertain and unwanted results as discussed earlier. The text-based approach is pretty simple in which images are annotated with labels and other information and tags that are assigned to it according to its categorization. The text description may not cover the information present in the image, which will result in inefficient and expensive image searching. So this technique is not only very uncertain but also involves the human labor which is undesirable and costly.

Content-based Image Retrieval (CBIR) is considered a better approach for image search. The term 'Content' refers to the information present in the image itself in terms of color, shape and texture. The basic difference between text-based and content-based image is; in text-based human provides a high level description regarding the image like keywords, tags, categorization, etc. whereas, the content-based directly deals with the meta-data and extracts the low level information present in the image which is unique and purely based on content of image and is stored in database as the index. Since the index is directly derived from the image hence, no semantic labeling is required.

The human visual system's perception, categorization and recognition regarding a specific view have no match. Human recognize an image and categorize them with an ideal accuracy and efficiency. They can differentiate each and every minor detail visually present in the image. The researchers are in a daily quest to achieve the desired accuracy and speed to be close to the human relative accuracy and efficiency. Semantic based search is close to human way to perceive the images. Semantic based search requires intelligence in algorithms to process the image and extract the theme, concept and other information present in the image. The image searching algorithm proposed in this paper belongs to second category, content based image retrieval.

Color and texture are very pre-attentive to human visual per-

ception. Julesz originated the Texton Theory which states that textons, composed of local image features, are 'the putative units of pre-attentive human texture perception' [1], [2]. The theory strongly supports the idea of making a strong bound between color and texture. Gupta et al. [3] introduced the edge direction histogram for trademark images retrieval which uses edges that makes it invariant to image translation, rotation and scaling. Only edge information is not as powerful as expected. Gevers et al.[4] came up with a technique that combined the power of a color and shape invariant features. Their technique is invariant to the translation, rotation, scaling, color and small viewpoint variations but is not appropriate for texture based image retrieval. Liu et al. [5] proposed Texton Co-occurrence Matrices (TCM) that describes the correlation of textons for image retrieval which discriminates its power in color, texture and shape features. Liu et al. [6] proposed an extension of TCM; Multi Texton Histogram (MTH) which discriminates TCM over color, texture and shape features and provides better results. Many CBIR content based search engines have been proposed and developed so far like QBIC [7], Photobook [8], Virage [9], VisualSEEK [10], Netra [11], SIMPLIcity [12]. A comprehensive discussion on past, present and future trends and challenges in CBIR can be found in [13], [14], [15], [16], [17].

The performance of any CBIR technique relies highly on the image orientation. The proposed Localized Multi Texton Histogram (LMTH) is an extension in MTH technique and targets to enhance the orientation detection for better retrieval. Secondly, the precedence of orientation values is increased which increases the ratio of image retrieval based on its orientation. Moreover, instead of keeping the global information targeted the local information present in the image is focused resulting in better performance. Section II presents the details of proposed technique. In section III, experimental results are presented and compared with existing techniques. Future work is discussed in sec. IV and conclusions are drawn in sec. V.

II. PROPOSED LOCALIZED MULTI TEXTON HISTOGRAM

The core of LMTH is Julesz's proposed Textons theory that says the local orientation differences between the elements that constitute the two texture images and it is possible to describe the differences between the texture images globally. If the first-order statistics of two texture images are identical, the second-order statistics may also differ greatly [1]. The proposed algorithm is different with existing CBIR techniques in two aspects. First, instead of computing the a single global vector for whole image, we exploit the local details by diving it into smaller sub-images and feature vectors for each block are computed and then integrated into a single feature vector. Second, we observed that texture plays more important role than color in retrieving best match images. In proposed approach, we allocate more space to texture orientations compared to exiting histogram based techniques. Increasing the space for texture orientation means we record the more accurate orientations with an interval of $10°$ degrees. This leads to preserve the texture in better way that eventually

results in good texture match. Third, to compute orientations, instead of using simple operators like Sobel or Prewitt, we use first derivative of Gaussian, resulting in better texture detection.

The first step in proposed technique is *image localization*. Let I be an image of size $m \times n$ and it is segmented into ρ smaller blocks of equal size. Without loss of generality, we assume m and n are divisible by $\frac{\rho}{2}$. All the blocks are distinct and are independent. The purpose of localizing the image is to enhance the importance of each block respectively and concentrate on its color, shape and texture independently. The value of ρ must be chosen carefully, large values results in better retrieval performance but increases the vector size too. Small values results in poor performance but consumes less space. From experiments, it is found that $\rho = \{4, 9\}$ is a good choice. Fig. 1 shows an image and its localized version. Once localization is applied, texture orientation in each block is computed.

(a)   (b)
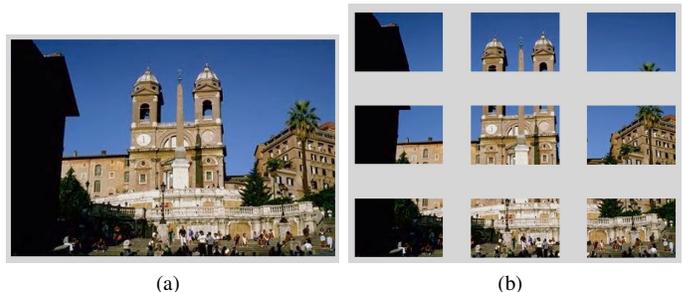
Fig. 1: (a) Original Image, (b) segmented image in 9 equal size blocks.

A. Texture orientations and color quantization

Orientation defines the direction of change in the texture of an image. There are many different ways to compute the orientation like using simple and fast Sobel or Prewitt operators in x and y direction of the input image to compute the gradients and then dividing them and taking inverse tangent gives the orientation angle. In proposed approach it is computed by finding gradient G_x and G_y in x and y directions using first derivative of Gaussian with radius σ (details are discussed in sec. III-A). In contrast to most existing image retrieval technique that compute gradient in each color channel, in proposed algorithm it is carried out in gray scale version of the input image. The orientation is then computed as:

$$\Theta = arctan(\frac{G_y}{G_x}) \tag{1}$$

The value of Θ gives us the direction angle and $0 \leq \Theta < 360$. To limit the vector size to reasonable bound and to minimize the quer computational cost, the texture orientation values are quantized in 36 (0–35) levels. Fig. 2 shows an image with its orientation.

The next step is to quantize the color of the image. We assume the input images are in RGB model that is close to the

Fig. 2: (a) An input image, (b) Texture orientations of (a).

digital medium and is quite straightforward and also simplifies the manipulations. Each color component is quantized into 4 levels, resulting in 64 different levels for all three color channels. The quantized color image is then filtered with the four pre-defined Textons as in MTH algorithm: horizontal, vertical, diagonal and off-diagonal Texton. The purpose of the filtering is to remove the unwanted data and preset the item for the next stage. Textures are formed only if the adjacent elements lie within the neighborhood. Hence, a small element size, such as 2×2 is used for close neighbor information. The texture discrimination can be increased as the texton gradients exist only at texture boundaries. These four textons are traversed over the filter quantized color image. While traversing if the texton is detected, the whole grid matrix is copied resulting in a new texton image. After computing the orientation and texton image, these are used to construct the feature vector.

B. Composing the feature vector and searching mechanism

In the proposed algorithm 36 values have been utilized for the orientation variation and 64 values for color variation for each block respectively, resulting 100 histogram values for each block. Hence, for ρ number of blocks, total feature vector size will be $100 \times \rho$. By taking only the histogram of the values, the image will lose the spatial relation amongst the image neighboring spatial data. For this purpose the predefined textons are used. For each texton, the obtained orientation and quantized textoned images are parsed and if both lie in the direction defined by the texton only then the frequency is added. This way the non-matching neighboring elements information is discarded and only the information related to the matching neighbors is saved which results in as a robust feature vector based on shape, texture and color.

To search an image, its feature vector is computed and compared with all the feature vectors of the stored images to compute the similarity index. Various ways has been proposed to compute the similarity of two feature vectors starting from simple and easy sum absolute difference (SAD) to more complex techniques like Mahalanobis distance (MD), dynamic time warping (DTW) [18], Bhattacharyya distance (BD) [19] and Quadratic distance (QD). In proposed approach Euclidean distance (ED) has been used to compute the similarity between the feature vectors. For two feature vectors of size n, $\alpha = [\alpha_1, \alpha_2, \alpha_3, \cdots \alpha_n]$ and $\beta = [\beta_1, \beta_1, \beta_1, \cdots \beta_n]$, the
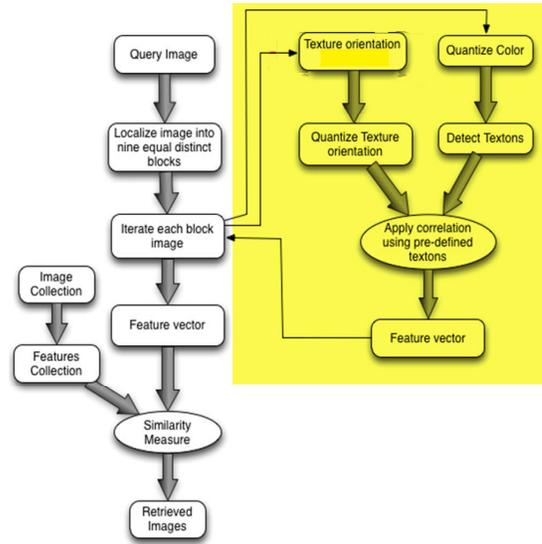
Fig. 3: Flowchart of the the propose algorithm. The area is yellow represents the iterative process carried out on each block.

similarity is given by the following equation:

$$sim(\alpha, \beta) = \sqrt{\sum_{i=0}^{n} (\alpha_i - \beta_i)^2} \qquad (2)$$

The resultant similarity indices are sorted in decreasing order to find the top k best matches. Complete flow of the proposed algorithm is shown in Fig. 3.

III. EXPERIMENTAL RESULTS AND COMPARISON

The proposed algorithm has been implemented in MATLAB and tested on standard Corel 1000 dataset. This database contains 1000 random images with 100 images of same categories are provided. In all experiments, ρ is set to 9 and in orientation detection σ is set to 1.4. To compute the performance of the algorithm, we use *precision* and *recall* measures. For Precision measure, the threshold is set to 12 images and for recall measure; the threshold is set to 100 as there are 100 images in each category in the database. The Matlab code of the proposed technique is available at *https://github.com/yunas/LMTH* with the dataset. Figures 4 and 5 show the results of two queries to the database, to save the space only first five outputs are shown here.


Fig. 4: Results returned against the query of flower image (image 622.jpg in dataset).
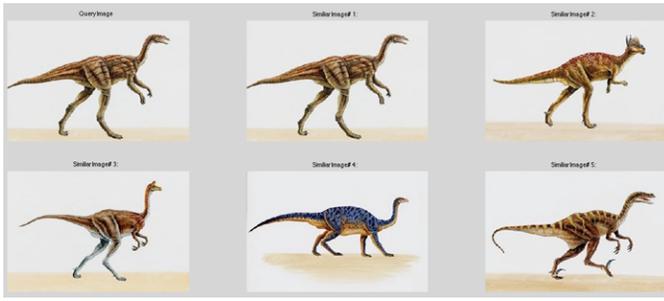
Fig. 5: Results returned against the query of dinosaur image (image 442.jpg in dataset).

The proposed technique is also compared with MTH approach. To compare the precision and recall measurements, 20 random experiments were carried out for the proposed and MTH algorithm. The proposed approach has average precision of 0.8042 compared to MTH which is 0.7417. The average recall of the proposed approach is 0.483 compared to 0.475 of MTH. The statistics of all the experiments for precision and recall have been summarized in Fig. 6, 7. These statistics reveal the better performance of the proposed technique with an average increase of 6.25% in precision. Figures 8, 9, and 10 show the results returned by proposed algorithm and MTH against three image queries. The results of the proposed technique are better due to representing texture in a stronger way in the feature vector. In each figure, image (a) is the result of MTH and (b) shows the output of proposed approach.
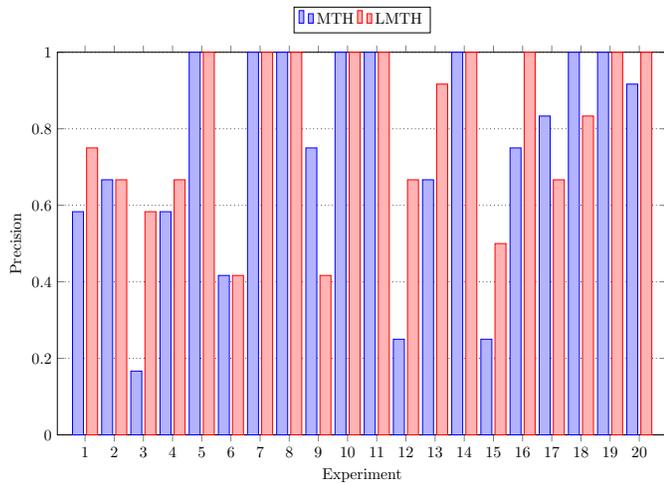

Fig. 6: Precision comparison

A. Performance parameter analysis

The performance of the proposed LMTH algorithm is driven by input parameters ρ and σ. The image is divided into ρ number of blocks and then for each block texture orientation is computed and color channels are quantized. Increasing the number of blocks results in better retrieval the performance but increases the vector size that may increase the query response time. The algorithm is tested against different values of ρ and
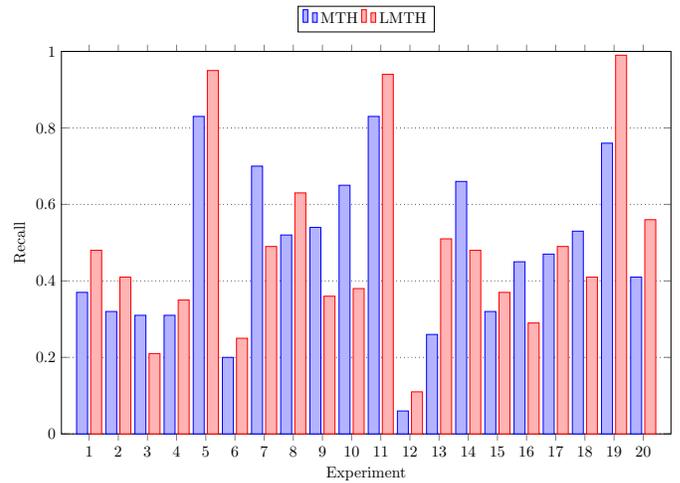

Fig. 7: Recall comparison

it is found that $\rho \in \{4, 9\}$ produces optimal results. The other parameter is σ used in Gaussian filter to compute gradient. Change in σ changes filter size as it is computed from σ by the following equation:

$$q = 2 \left\lfloor \sqrt{-log(\tau) * 2 * \sigma^2} \right\rfloor + 1 \qquad (3)$$

where τ is the threshold and in experiments its value is set to 0.1. Size of the filter is then $q \times q$. In experiment setup the value of σ is set to 1.4, resulting in filter of size 7×7.

IV. FUTURE WORK

The proposed technique only focuses on localized information of the image, as the image is segmented into smaller images and then a single feature vector is constructed using the orientation and color information of these smaller blocks. This results in larger vector size as compared to MTH algorithm which in turn affects the response time of the proposed technique. In future work, we are working to translate the proposed approach into a hierarchical retrieval algorithm in which instead of comparing the whole vectors to find the similarity, comparison will be done in hierarchical fashion, block by block taking into account the spatial correlation of the blocks. This would greatly improve the execution time while maintaining the same quality of retrieval.

Another extension in investigation is to modify this approach to decrease the feature vector size and to utilize the global information of the whole image separately and to decide some precedence to compute the result as collaboration of global and local information. This may result in a fast and more accurate algorithm that exploits both local and global information in the image. Thirdly, the possibility of using dynamic textons instead of using predefined and fixed size textons will also be investigated. Dynamic textons may help to record spatial information at specific angles instead of specific directions.
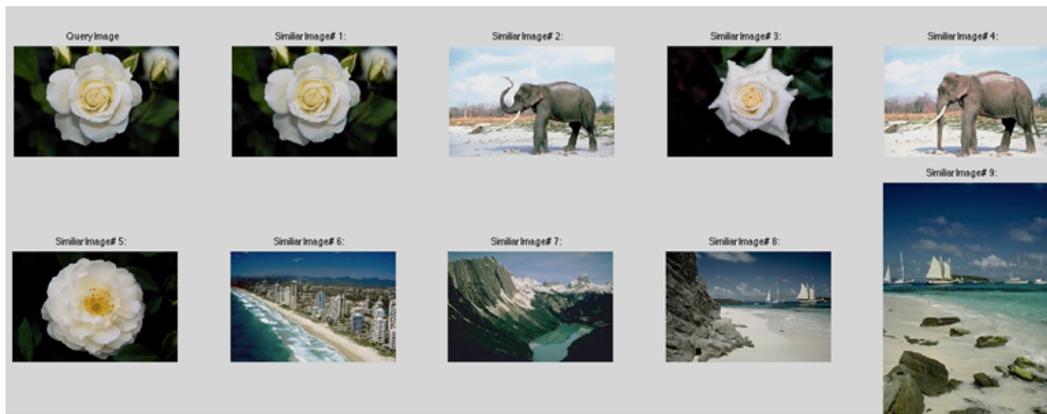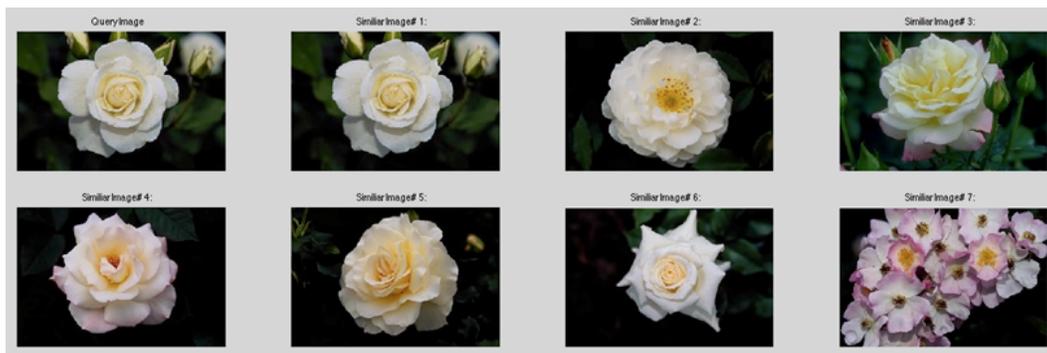
Fig. 8: (a) Result of MTH, (b) Result of proposed LMTH

V. CONCLUSION

A content base image retrieval algorithm is presented in this manuscript. The algorithm holds the spatial correlation between the image orientation and image color based on the textons at localized level. It uses powerful orientation detector which provides more room to capture the texture details present in the image. Further, the algorithm divides the image in smaller chunks and compute features at block level which later are integrated to form a full feature vector. In feature vector, more space is given to texture as compared to its existing techniques which empowers it to find better matches based on the textures of the objects. Due to this, its performance is excellent when the images contain rich information. The similarity between the vectors is evaluated by simply computing the Euclidean distance. The results and comparisons show the effectiveness of the proposed algorithm.

REFERENCES

[1] Bela Julesz, "Texton gradients: The texton theory revisited," *Biological Cybernetics*, vol. 54, no. 4-5, pp. 245–251, 1986.

[2] Bela Julesz, "Textons, the elements of texture perception, and their interactions.," *Nature*, 1981.

[3] A.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.

[4] Fariborz Mahmoudi, Jamshid Shanbehzadeh, Amir-Masoud Eftekhari-Moghadam, and Hamid Soltanian-Zadeh, "Image retrieval based on shape similarity by edge orientation autocorrelogram," *Pattern Recognition*, vol. 36, no. 8, pp. 1725 – 1736, 2003.

[5] Guang-Hai Liu and Jing-Yu Yang, "Image retrieval based on the texton co-occurrence matrix," *Pattern Recognition*, vol. 41, no. 12, pp. 3521 – 3527, 2008.

[6] Guang-Hai Liu, Lei Zhang, Ying-Kun Hou, Zuo-Yong Li, and Jing-Yu Yang, "Image retrieval based on multi-texton histogram," *Pattern Recogn.*, vol. 43, no. 7, pp. 2380–2389, July 2010.

[7] C. Faloutsos, R. Barber, M. Flickner, J. Hafner, W. Niblack, D. Petkovic, and W. Equitz, "Efficient and effective querying by image content," *J. Intell. Inf. Syst.*, vol. 3, no. 3-4, pp. 231–262, July 1994.

[8] A. Pentland, R. W. Picard, and S. Sclaroff, "Photobook: content-based manipulation of image databases," *Int. J. Comput. Vision*, vol. 18, no. 3, pp. 233–254, June 1996.

[9] Amarnath Gupta and Ramesh Jain, "Visual information retrieval," *Commun. ACM*, vol. 40, no. 5, pp. 70–79, May 1997.

[10] John R. Smith and Shih-Fu Chang, "Visualseek: a fully automated content-based image query system," in *Proceedings of the fourth ACM international conference on Multimedia*, New York, NY, USA, 1996, MULTIMEDIA '96, pp. 87–98, ACM.

[11] Wei-Ying Ma, *NETRA: a toolbox for navigating large image databases*, Ph.D. thesis, University of California, Santa Barbara, CA, USA, 1998, UMI Order No. GAX98-00472.

[12] James Z. Wang, Jia Li, and Gio Wiederhold, "Simplicity: Semantics-sensitive integrated matching for picture libraries," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 9, pp. 947–963, Sept. 2001.

[13] Ritendra Datta, Dhiraj Joshi, Jia Li, and James Z. Wang, "Image retrieval: Ideas, influences, and trends of the new age," *ACM Comput. Surv.*, vol. 40, no. 2, pp. 5:1–5:60, May 2008.

[14] Yong Rui, Thomas S. Huang, and Shih-Fu Chang, "Image retrieval: Current techniques, promising directions, and open issues," *Journal of*

Fig. 9: (a) Result of MTH, (b) Result of proposed LMTH
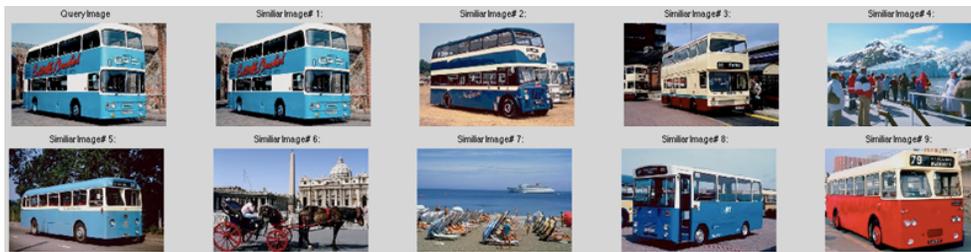
Fig. 10: (a) Result of MTH, (b) Result of proposed LMTH

Visual Communication and Image Representation, vol. 10, no. 1, pp. 39 – 62, 1999.

[15] Michael S. Lew, Nicu Sebe, Chabane Djeraba, and Ramesh Jain, "Content-based multimedia information retrieval: State of the art and challenges," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 2, no. 1, pp. 1–19, Feb. 2006.

[16] Ying Liu, Dengsheng Zhang, Guojun Lu, and Wei-Ying Ma, "A survey of content-based image retrieval with high-level semantics," *Pattern Recognition*, vol. 40, no. 1, pp. 262 – 282, 2007.

[17] CeesG.M. Snoek and Marcel Worring, "Multimodal video indexing: A review of the state-of-the-art," *Multimedia Tools and Applications*, vol. 25, no. 1, pp. 5–35, 2005.

[18] Eamonn Keogh and Chotirat Ann Ratanamahatana, "Exact indexing of dynamic time warping," *Knowl. Inf. Syst.*, vol. 7, no. 3, pp. 358–386, Mar. 2005.

[19] A. Bhattacharyya, "On a measure of divergence between two statistical populations defined by their probability distributions," *Bulletin of the Calcutta Mathematical Society*, vol. 35, pp. 99–109, 1943.