Pertanika J. Sci. & Technol. 21 (1): 161 - 168 (2013)

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Content-based Image Retrieval Using Colour and Shape Fused Features

Mas Rina Mustaffa*, Fatimah Ahmad, Ramlan Mahmod and Shyamala Doraisamy

Department of Multimedia, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia

ABSTRACT

Multi-feature methods are able to contribute to a more effective method compared to single-feature methods since feature fusion methods will be able to close the gap that exists in the single-feature methods. This paper presents a feature fusion method, which focuses on extracting colour and shape features for content-based image retrieval (CBIR). The colour feature is extracted based on the proposed Multi-resolution Joint Auto Correlograms (MJAC), while the shape information is obtained through the proposed Extended Generalised Ridgelet-Fourier (EGRF). These features are fused together through a proposed integrated scheme. The feature fusion method has been tested on the SIMPLIcity image database, where several retrieval measurements are utilised to compare the effectiveness of the proposed method with few other comparable methods. The retrieval results show that the proposed Integrated Colour-shape (ICS) descriptor has successfully obtained the best overall retrieval performance in all the retrieval measurements as compared to the benchmark methods, which include precision (53.50%), precision at 11 standard recall levels (52.48%), and rank (17.40).

Keywords: CBIR, colour, feature fusion, and shape

INTRODUCTION

Content-based Image Retrieval (CBIR) is developing amazingly due to the increase needs to show, share, organise, search, and retrieve digital imagery. The CBIR technology has been useful in many areas, such as crime prevention, medicine, law, science, fashion and interior design, and a few others. As compared to conventional image retrieval techniques that use image annotations or keywords, CBIR has been able to reduce the many problems associated with retrieving images based on

Article history:
Received: 31 March 2012
Accepted: 31 August 2012

E-mail addresses:
masrina@fsktm.upm.edu.my (Mas Rina Mustaffa),
fatimah@fsktm.upm.edu.my (Fatimah Ahmad),
ramlan@fsktm.upm.edu.my (Ramlan Mahmod),
shyamala@fsktm.upm.edu.my (Shyamala Doraisamy)
*Corresponding Author

ISSN: 0128-7680

text, such as manual image annotations, differences in perceptions and interpretations, and language barrier, where image annotations are usually presented in one language, by retrieving images on the basis of automatically derived low-level features, middle-level features, or high-level features (Datta *et al*., 2008; Vassilieva, 2009; Veltkamp & Tanase, 2002). Among these features, the low-level features are the most popular due to their simplicity compared to the other levels of features plus automatic object recognition and classification which are still among the most difficult problems in image understanding and computer vision.

Apart from utilising the low-level features separately, they can also be integrated whereby in most cases, a single feature will not be suffice to represent the pictorial content of an image and fusing the low-level features will lead to a better performance. According to Forczmański and Frejlichowski (2007), the fusing features can be done either through the sequential approach or parallel approach. Based on the sequential approach, an image will first be processed by one feature descriptor and the output will then be used as the input for the following feature descriptor, and so on until a final feature vector is obtained. The parallel approach, on the other hand, works by allowing an image to be processed by different feature descriptors separately, where each of these descriptors will generate its respective coefficients and these coefficients will be integrated based on a certain scheme to produce the final integrated feature vectors. The sequential approach allows for a filtering procedure, where images which are depicted as very dissimilar by the first feature descriptor, can be omitted for subsequent steps. This can lead to a decrease in future computation time and computation complexity. However, allowing for images to be filtered out at the initial stage may result in false elimination of similar images, thus influencing the end result as well as the effectiveness of the approach. On the contrary, the parallel approach allows for computation time reduction where different methods can be calculated at the same time independently. The integration of the features is usually performed at the final stage where equal or different weights are assigned to each feature. Determining the best weight for the respective feature can be a complex task. However, assigning different weights to each feature will allow a method to stress the contribution of one set of features which may be considered as more important in the application than the rest. Weight distribution is definitely easier to be done in a parallel-based approach. Unlike the sequential-based feature fusion, to stress out a feature is to modify the process flow all together. The sequential and parallel approaches have their own strengths and weaknesses. In practice, the selection between a sequential-based or a parallel-based approach is very much dependent upon the application needs and requirements. There are quite a number of efforts done related to the feature fusion in CBIR to overcome the shortcomings associated with the single-feature methods. Some of the many works can be found in the following references (see Choraś *et al*., 2007; Lee & Yin, 2009; Prasad *et al*., 2001; Schaefer, 2004; J.Y. Wang & Zhu, 2010; Yue *et al*., 2011).

Therefore, this paper aims to contribute to a parallel-based feature fusion method, namely ICS descriptor, which combines colour and shape features to improve the performance of a CBIR system. The remainder of the paper is organised as follows: The methods to extract the respective colour and shape features are first described. Then, the combining strategy is explained following by the description on the experimental setting and discussion of results. Conclusions and future works can be found at the end of the paper.

COLOUR FEATURE BASED ON MJAC

The proposed MJAC is an improved framework by adopting ideas from three previous works in CAC including those by Huang *et al*. (1997), Moghaddam *et al*. (2005) and Williams and Yoon (2007) by implementing the Joint Auto Correlograms in a multi-level resolution. This provides an enhanced method which allows for pixels correlation of several local image features, such as colour, gradient magnitude, rank, and texturedness to be captured at different image scales in the frequency domain. At the beginning, the collected images will go through an extraction process to separate the RGB colour space and the grey-scale colour space. This is followed by the implementation of the Ridgelet transform on both of the RGB colour space and the grey-scale colour space by first calculating the Radon transform, followed by the implementation of the one-dimensional discrete Wavelet transform on the Radon transform coefficients up to four sub-bands. Next, focusing only on scale-3 and scale-4 of the one-dimensional discrete Wavelet transform coefficients, the local image features such as colour, gradient magnitude, rank, and texturedness are extracted, where each of the local image features is treated independently. The colour feature is extracted in the RGB colour space, while the gradient magnitude, rank, and texturedness are extracted in the grey-scale colour space. In order to reduce the feature dimension, a compact representation, as well as ease of coefficient management, these local image features is then quantised. Auto correlogram is then performed on each of the quantised local image features. At the end of this process, feature vectors are generated to represent the images, which are stored in the feature database. The similarity between the query image and the images in the database are measured by comparing the feature vectors of the query image with the feature vectors of all the images in the database using the L_1-norm distance function.

SHAPE FEATURE BASED ON ENRF

ENRF is an extension and improvement to the work done by Chen *et al*. (2006), where the extended descriptor is now resulting in a rotation, scaling, and translation invariant Ridgelet transform for images of various sizes. The first step is by putting the collected images through an initialisation process. During this process, all the images in the collection are assigned with certain constraints. The images will then be made translation and scaling invariant. Only pixels of the translation and scaling invariant images that fall within the ellipse template centred at ($M/2$, $N/2$) are considered for the next steps (note that M and N represents the width and height of an image, respectively). A template option is made available after implementing the ellipse template setting, where it allows the algorithm to provide an option in choosing between proceeding with the ellipse template or considering the square template instead when processing the translated and scaled pixels of an image. A square template is used if the percentage of the pixels of the translation and scaling invariant image that fall outside of the ellipse area is more than the threshold value. Otherwise, the ellipse template is utilised. Radon transform is then performed on the shape that has been processed according to the suitable template using 128 points for both the number of samples taken on each radial line as well as the number of orientations. After normalising the Radon transform, the next step is to apply the one-dimensional Discrete Wavelet transform on each of the Radon slices to obtain the Ridgelet coefficients. In order to make the descriptor invariant to rotation, the one-dimensional Discrete Fourier transform is performed along the angular direction of the scale 3 and scale 4 of the wavelet decomposition levels. For each of the mentioned wavelet decomposition levels, only 15 Fourier magnitude spectrums are captured to represent the shape. These coefficients will be used as the feature vectors for the image which is stored in the feature database. As for the distance function, the L_1-norm is utilised.

SIMILARITY MEASURE

Fig.1 shows the scheme to combine the colour and shape features. According to the L_1- norm distance function concept, the smaller the similarity value between an image and the query image, the more similar the image is to the query image. Meanwhile, between the similarity values generated for both the colour and shape features, it is easier to identify whether an image is similar or not to a query image based on the shape feature similarity values rather than the colour feature similarity values since the shape feature similarity values are in wider range. Due to this reason, the shape feature is being made the benchmark in determining the initial similarity as shown in 'Line 1' of Fig.1 below.

As shown in 'Line 1' of Fig.1, the shape feature similarity values are being multiplied with a threshold value, *Threshold_Value*. The *Threshold_Value* is used to identify the limit of which the *Shape_Similarity_Value* is small enough to make it noteworthy to be retained. Through the experiments, it has been identified that the best *Threshold_Value* is 3000. If the *Shape_Similarity_Value* multiplied by 3000 is below the value of 1, the *Shape_Similarity_Value* is noteworthy enough to be retained. The retention process is done by multiplying the *Shape_Similarity_Value* (range 0.000248 to 0.761901) with the *Colour_Similarity_Value* (range 1.195548 to 9.146877). Since the range of the *Colour_Similarity_Value* is more than 1, when it is multiplied with the *Shape_Similarity_Value* of less than 1, it will always return a sum of less than 1, and hence retaining the small similarity value. If the *Shape_Similarity_Value* multiplied by 3000 is more than the value of 1, the similarity value is no longer noteworthy to be retained, hence an addition is performed.

Based on the literature review, the colour feature is usually superior in image representation compared to other features. Due to this reason, more weights are given to the colour feature compared to the shape feature, as shown in 'Line 2' and 'Line 4' of Fig.1 above. A weight of 0.2 is given to the shape feature while a weight of 0.8 is given to the colour feature ($a + b = 1$). These weights are obtained experimentally.

Line 1: If (*Shape_Similarity_Value* * *Threshold_Value*) < 1
Line 2: (*a* * *Shape_Similarity_Value*) * (*b* * *Colour_Similarity_Value*)
Line 3: Else
Line 4: (*a* * *Shape_Similarity_Value*) + (*b* * *Colour_Similarity_Value*)
Line 5: End If

Fig.1: Algorithm for Combining the Colour and Shape Features

EVALUATION AND ANALYSIS OF THE RESULTS

The objective of this experiment was to evaluate the retrieval effectiveness of the proposed feature fusion descriptor, namely, ICS compared to the proposed colour-only descriptor (MJAC) and the proposed shape-only descriptor (ENRF) based on several retrieval measurements. The retrieval experiments were conducted on an Intel Pentium Dual-Core 2.5 GHz desktop. These methods were tested on 100 SIMPLIcity image dataset where the first 10 images from each of the image classes made up the 100 images (for example, the images labelled 0-9 for the first class, 100-109 for the second class, 200-209 for the third class, etc.) (Wang *et al*., 2001). For each class, all of the 10 images were selected as image queries. In total, there were 100 query images. There are 10 similar images in each image class, which provide the ground truth. The Query-by-Example (QBE) paradigm is employed. Samples of the SIMPLICITY image dataset is shown in Fig.2.

The retrieval effectiveness of the MJAC, ENRF, and ICS was evaluated using three different retrieval measurements, which were the precision at 10, 11 standard precision-recall at 10, and rank at 100.

Table 1 shows a summary of the overall average retrieval effectiveness of the proposed MJAC, ENRF, and ICS for 100 SIMPLIcity image dataset based on three retrieval measurements. It should be highlighted that the proposed feature fusion descriptor (ICS) is able to achieve the best overall performance for all the retrieval measurements. This indicates that the proposed ICS is better in differentiating between similar and dissimilar images when given a query image as well as able to retrieve similar images at higher rank compared to the colour-only MJAC and the shape-only ENRF. The best overall retrieval performance achieved by the ICS in all three retrieval measurements indicates its capability to maintain a good performance across different retrieval measurements compared to the benchmark methods. The consideration of colour and shape information in one descriptor allows or the ICS to perform better compared to the colour-only descriptor and shape-only descriptor. It is common where one feature may not be enough to represent an image since different images may contain colour information but not shapes, and vice versa. Generally, by taking a few features together, they can complement each other, hence contributing to a better descriptor rather than a single feature descriptor.

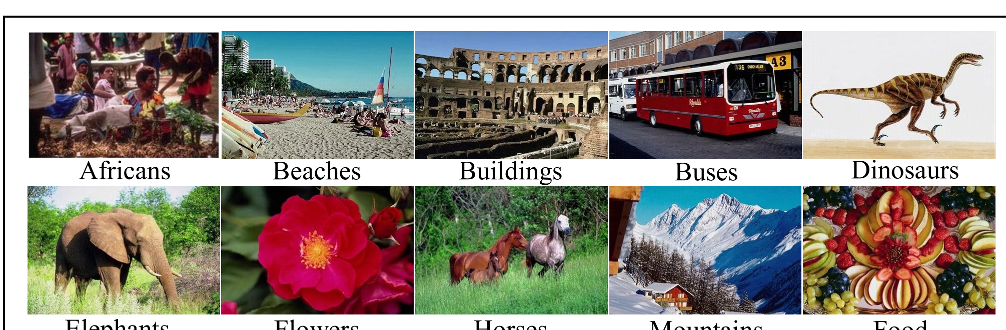

Fig.2: Samples of the SIMPLIcity image dataset

This work has successfully integrated the colour and shape features and the retrieval results have shown the implication of the proposed feature fusion descriptor (ICS) in improving the retrieval effectiveness.

TABLE 1: Retrieval effectiveness summary between the ICS and its benchmark methods based on various retrieval measurements

Method	Average Precision (p-10)	Average 11 Standard Precision-Recall (p-10)	Average Rank (p-100)
MJAC	0.508000	0.491669	18.409000
ENRF	0.312000	0.325884	36.790000
ICS	0.535000	0.524829	17.403000

From Table 1, it can be concluded that the proposed MJAC is suitable for images that have various colour information with inadequate shape characteristics, while the ENRF performs better on images with obvious shapes or regions but less colour variations. On the contrary, ICS is able to achieve higher precision rate for images with obvious colour and shape information. In general, the SIMPLIcity image dataset contains images with complex shapes and backgrounds. Therefore, it can be observed that the proposed shape descriptor is not able to perform equally to that of the proposed colour descriptor since colour information is definitely more obvious in such images. This has led to a relatively big gap in retrieval performance between the proposed colour and shape descriptors. Since the feature fusion method makes use of these two descriptors, the proposed ICS will generally hold a retrieval rate which gives a balance between these two descriptors. Nevertheless, it can be said that the proposed feature fusion method has successfully achieved a reasonable precision rate for all image classes with either achieving the highest precision rate or the second highest (with mostly achieving small gaps compared to the highest precision rate) in most of the image classes.

CONCLUSION AND FUTURE SUGGESTIONS FOR WORK

In this paper, a new feature fusion approach is proposed. The proposed ICS integrates colour and shape features, where the colour feature is extracted based on the MJAC while the shape information is obtained through the ENRF. These features are combined through a proposed integrated strategy. The feature fusion method is introduced to boost the retrieval performance compared to single-only methods. The results have shown that retrieval effectiveness has been improved by integrating features for image representation. Based on 100 SIMPLIcity images, where all 100 images have been chosen as query images, ICS has successfully obtained the best overall retrieval performance in three different retrieval measurements, which include precision at 10 with 53.50%, 11 standard precision-recall at 10 with 52.48%, and rank at 100 with 17.40. This has proven that the proposed integrated descriptor, which combines the colour and shape features, has contributed to a better retrieval and rank of similar images. Achieving the best overall retrieval performance in all three different retrieval measurements has also proven the capability of the proposed ICS to maintain a good performance across different retrieval measurements as compared to the benchmark methods. Future work will include providing relevance feedback as well as combining the colour-shape methods with other features such as texture.

ACKNOWLEDGMENTS

Financial support provided by the Ministry of higher Education, Malaysia and Universiti Putra Malaysia during the duration of this research is gratefully acknowledged.

REFERENCES

Chen, G. Y., Bui, T. D., & Krzyżak, A. (2006). Rotation invariant feature extraction using Ridgelet and Fourier transforms. *Pattern Analysis and Applications*, *9*(1), 83-93.

Choraś, R. S., Andrysiak, T., & Choraś, M. (2007). Integrated colour, texture, and shape information for content-based image retrieval. *Pattern Analysis and Applications*, *10*(4), 333-343.

Datta, R., Joshi, D., Li, J., & Wang, J. Z. (2008). Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys, 40*(2), 1-60.

Forczmański, P., & Frejlichowski, D. (2007). Strategies of shape and colour fusions for content-based image retrieval. *Computer Recognition Systems*, *2*, 3-10.

Huang, J., Kumar, S. R., Mitra, M., Zhu, W. J., & Zabih, R. (1997). Image indexing using colour correlograms. In *Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition (CVPR '97)* (pp. 762-768). Washington, DC, USA: IEEE Computer Society.

Lee, X. F., & Yin, Q. (2009). Combining colour and shape features for image retrieval. In *Universal Access in HCI, Part III, HCII 2009, 5616 of LNCS* (pp. 569-576).

Moghaddam, H. A., Khajoie, T. T., Rouhi, A. H., & Tarzjan, M. S. (2005). Wavelet correlogram: A new approach for image indexing and retrieval. *Pattern Recognition*, *38*(12), 2506-2518.

Prasad, B. G., Biswas, K. K., & Gupta, S. K. (2001). Colour and shape index for region-based image retrieval. In Arcelli, C., Cordella, L. P., & Di Baja, G. S. (Eds.), *Proceedings of the 4th. International Workshop on Visual Form (IWVF-4)*, 28-30 May 2001, Capri, Italy (716-725). London, UK: Springer-Verlag.

Schaefer, G. (2004). CVPIC colour/shape histograms for compressed domain image retrieval. In *DAGM 2004, 3175 of LNCS*, Tubingen, Germany (pp. 424-431).

Vassilieva, N. S. (2009). Content-based image retrieval methods. *Programming and Computing Software, 35*(3), 158-180.

Veltkamp, R. C., & Tanase, M. (2002). *Content-based image Retrieval Systems: A Survey*. Technical Report UU-CS-2000-34, Institute of Information and Computing Sciences, Utrecht University.

Wang, J. Y., & Zhu, Z. (2010). Image retrieval system based on multi-feature fusion and relevance feedback. In *Proceedings of the Ninth International Conference on Machine Learning and Cybernetics*, 11-14 July 2010, Qingdao (pp. 2053-2058).

Wang, J. Z., Li, J., & Wiederhold, G. (2001). SIMPLIcity: Semantics-sensitive integrated matching for picture libraries. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *23*(9), 947-963.

Williams, A., & Yoon, P. (2007). Content-based image retrieval using joint correlograms. *Multimedia Tools and Applications*, *34*(2), 239-248.

Yue, J., Li, Z., Liu, L., & Fu, Z. (2011). Content-based image retrieval using colour and texture fused features. *Mathematical and Computer Modelling, 54*(3-4), 1121-1127.