Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Engineering 30 (2012) 542 – 545

Procedia Engineering

www.elsevier.com/locate/procedia

International Conference on Communication Technology and System Design 2011

Image retrieval based on visual attention model

Satrajit Acharya[2], M.R.Vimala Devi[1], 1*

[2]*School of Electrical Sciences, VIT University, Vellore.*
[1]*Department of ECE, SASTRA University, Thanjavur*

Abstract

Image Retrieval has always been an area of extensive research. Many efficient retrieval algorithms have already been proposed. Text based image retrieval system got replaced and Content Based Image Retrieval (CBIR) came into being. CBIR, more efficient than text based image retrieval, has already been implemented by several techniques and researches are going on to increase the efficiency. This paper proposes a new technique of incorporating visual attention model to segment and extract the ROI from an image and then use the result for image retrieval purposes. The main advantage of this concept lies in the improvement of the performance of this retrieval scheme in terms of two parameters: Precision and Recall.
Keywords: image retrieval; precision; recall.

1. Introduction

Over the past few years much progress have been made in the field of image retrieval. The increase in the volume of digital images that continuously gets uploaded and with the invention of inexpensive hardware and software for image acquisition, storage and distribution fostered the growth for CBIR. The model consists of a search engine which translates the visual information imbibed in the user's query image and retrieves the similar images from the database which are similar to it in some respect.
Many techniques have already been implemented but none of them has satisfactorily solved CBIR. The problems include Some of the open problems include the gap between the image features that can be extracted using image processing algorithms and the semantic concepts to which they may be related (the well-known *semantic gap* problem which can often be translated as "the discrepancy between the query a user ideally *would* and the one it actually *could* submit to an information retrieval system".The architecture is inspired by computational model of human visual attention [1, 2,3], which provides important cues about the location of the most salient ROIs within an image. These ROIs, once extracted, are used to obtain feature vectors which are used to find similar ROIs that may have appeared in other images in the existing database.The paper is structured as follows: Section 2 of this paper discusses in brief the overall methodology. Section 3 deals with the results obtained. Section 4 contains the concluding remarks.

2. Methodology

2.1 Saliency map

The Itti-Koch model [2] of visual attention considers the task of attention selection from a purely bottom-up perspective. The model generates a map of the most salient points in an image, which will be used as cues for identifying ROIs. The image is subdivided into Gaussian pyramids of color, intensity and orientation [2]. These are used to form feature maps [5] by taking centre surround differences. Finally, the conspicuity maps are obtained by across scale addition and normalization [5] whose mean gives the saliency map.

2.2 Visual attention map

* Vimala Devi R. Tel.: +91-9003365001.
E-mail address; madhumrvd@yahoo.co.in.

1877-7058 © 2011 Published by Elsevier Ltd.
doi:10.1016/j.proeng.2012.01.896

The model of visual attention proposed by Stentiford [3,6] functions by suppressing areas of the image with patterns that are repeated elsewhere. A random pixel along with its neighborhood is compared with random pixels elsewhere in the image for a match. As a result flat surfaces and textures are suppressed while unique objects are given prominence. Regions are marked as high interest if they possess features not frequently present elsewhere in the image.

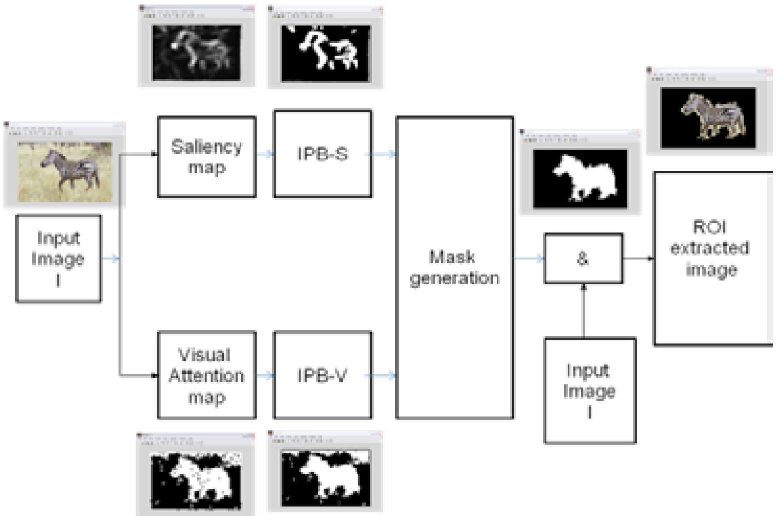

Fig 1. Step by step implementation of ROI extraction process

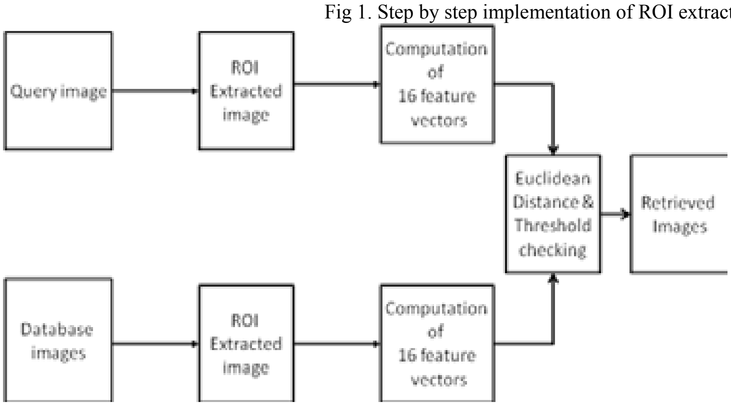

Fig 2. Block diagram of image retrieval system using ROI extraction results

2.3 Mask generation

Using the results from the two saliency and visual attention map, the mask is generated [1]. It consists of thresholding the two maps followed by certain morphological operations [1] to remove the spurious pixels and fill the holes. Using the regions common to both the maps a mask is generated which is used to extract the ROI

2.4 Feature vectors

Feature vectors are calculated for the ROIs extracted from the database as well as the query image. In this paper color has been considered as the feature vector and mean, standard deviation, entropy and energy as the feature parameters. The ROI is subdivided into red, green and blue channels. Each channel is further divided into two sub images, one containing pixels with values greater than the mean and the other containing pixels value less than the mean. For each such sub image mean and standard deviation is calculated forming 3*2*2=12 feature parameters. Furthermore mean, standard deviation, entropy and energy are calculated for the whole ROI. Thus, 16 feature parameters are constructed, in total.

2.5. Similarity measure

Euclidean distance is used to find the similarity between the feature parameters of query image's ROI and database images' ROI. A threshold of 14 out of 16 feature parameters is chosen, by trial and error, to conclude whether an image is similar or not. ROIs passing the threshold criteria are retrieved as similar ones.

3. Experiments and results

This section contains representative results from our experiments and discusses the performance of the proposed approach on a representative dataset. The model has been implemented in MATLAB 7.0.

3.1 Results

The composition of the image database is of primary importance for the success of this technique. The images should be carefully chosen so that they are appropriate to the context and yield relevant results. Images should have semantically well defined ROIs (regions that are salient by design) with the background mostly uniform and the objects distinct from the rest of the image. The database images consist of meaningful, distinct objects against a more or less uniform background. All the images are naturally occurring photography scenes. The database [8,9] consists of a total of 100 images with 6 categories namely Leopard(22), Zebra(19), Bear(15) and Flower(20). The Bear and Flower categories have two different sets of images with the colors different from each other. Thus they are further divided into Bear 1(7), Bear 2(8), Flower 1(9), and Flower 2 (11). There are 24 other images which have been introduced to estimate the performance of the proposed model. The images have wide range of images in terms of the difficulty to extract the ROIs.

The following section reports the performance of the proposed model.

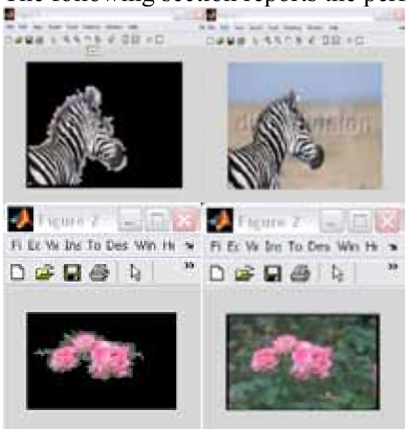

Fig 3. ROI extraction results

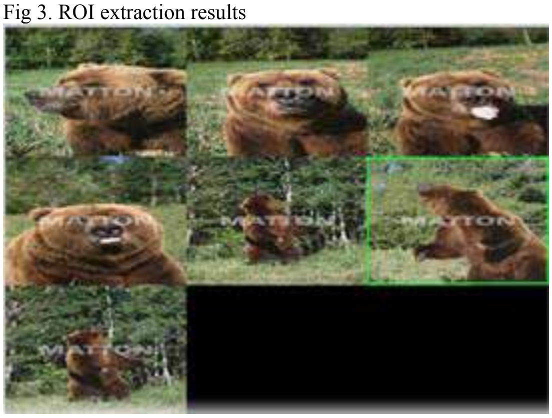

Fig 4. Image retrieval result (image enclosed in green box is the input image)

3.2 Image retrieval

The performance of image retrieval is determined in terms of two parameters: *precision* and *recall*. Precision is defined as the fraction of relevant images retrieved from all the retrieved images. Recall is fraction of relevant images which has been retrieved from all the relevant images in the database. The following table shows the precision and recall value of the different image categories.

Table 1: Precision and Recall Values for different image categories in the database

Category	Precision	Recall
Leopard	.9	.81
Zebra	.85	.89
Bear 1	1	1
Bear 2	.88	.88
Flower 1	.9	1
Flower 2	.77	.91

3.3 Discussions

The Performance measure shows that the proposed model has improved the efficiency of CBIR. Unlike the existing CBIR technique the proposed one has performed better in terms of the two performance parameters: precision and recall. Both the parameters have achieved high values and the performance of one parameter has not affected the other one. However, there are still certain shortcomings that need to be addressed which invariably fit into one of the following categories*: false negatives* (meaningful ROIs not extracted), *false positives* (extraneous ROIs extracted), threshold.

The first two belong to the ROI extraction stage. The false positives affect the feature vector extraction process thus making it essential to include a threshold value at a later stage. False negatives are completely lost and cannot be recovered at a later stage

The third shortcoming, threshold, belong to both ROI extraction as well as image retrieval stage. In case of ROI extraction stage, threshold determines which ROIs to keep and which to discard. The threshold value should be such that it minimizes false positives while maximizing false negatives. In image retrieval stage the threshold value determines whether the feature vectors are similar or not. Thresholds are set purely based on trial and error and hence a scope for further improvement is always there.

4. Conclusion

Visual attention model has thus been successfully implemented using computational methods. Visual attention model has also proved to be extremely useful in segmenting the image and extracting the Region of Interest (ROI) with less number of unwanted pixels. The incorporation of both the Itti-Koch model and Stentiford model has been found to be fruitful in eliminating the individual demerits of the two models. The application of Visual Attention model has also been successfully implemented for Image retrieval Purposes. It has been found that the performance parameters of the retrieval system have been enhanced due the incorporation of visual attention. The existing disadvantage of one performance parameters being enhanced at the cost of the other has also been nullified with both parameters being found to be more or less same and of high value.

References

[1] Oge Marques, “An attention-driven model for grouping similar images with image retrieval applications”, *Eurasip Journal on Advances in Signal Processing*, Special Issue on Image Perception, 2007, Vol. 2007, Article ID 43450, 17 pages.

[2] U. Rutishauser, D. Walther, C. Koch, and P. Perona, “Is bottom-up attention useful for object recognition?”, in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR ’04),* Washington, DC, USA, June-July 2004, vol. 2, pp. 37–44.

[3] A. Bamidele, F. W. M. Stentiford, and J. Morphett, “An attention-based approach to content-based image retrieval”, *BT Technology Journal, 2004,* vol. 22, no. 3, pp. 151–160.

[4] Ying-Hua Lu, Xiao-Hua Zhang, Jun Kong , Xue-Feng Wang, “A Novel Content-Based Image Retrieval Approach Based On Attention-Driven Model”, *Proceedings of International Conference on Wavelet Analysis and Pattern recognition, Beijing, China,*2-4 Nov. 2007.

[5] L. Itti and C. Koch, “Feature combination strategies for saliency-based visual attention systems”, *Journal of Electronic Imaging*, 10(1):161–169, 2001.

[6] Fred Stentiford, “A Visual Attention Estimator Applied to Image Subject Enhancement and Colour and Grey Level Compression”, *Proceedings of the 17th International Conference on Pattern Recognition (ICPR’04*), 2004.

[7] W. Freeman and E. Adelson, “The design and use of steerable filter”, *IEEE Trans. Pattern Anal. Machine Intell.,* 1991, vol. 13, pp. 891–906.

[8] http://ilab.usc.edu/imgdbs/

[9] http://elib.cs.berkeley.edu/photos/blobworld