ISSN No. 0976-5697

Volume 5, No. 5, May-June 2014

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

Classification of Natural Scene Images Based on Detection of Emotions

Shweta Wagh
Dept. of Computer Engineering
AVCOE, Sangamner Sangamner, India

Savita Darekar
Dept. of Computer Engineering
AVCOE, Sangamner Sangamner, India

Mayuri Gawali
Dept. of Computer Engineering
AVCOE, Sangamner Sangamner, India

Poonam Nikam
Dept. of Computer Engineering
AVCOE, Sangamner Sangamner, India

Abstract: Emotion in natural scene images plays an important role in the way humans perceive an image. Based on the emotion (happiness, sadness, fear, anger etc.) of any human being the images that are viewed by that person can have a significant impact in a sense that if the person is for example in happy mood and he/she views an image that is pleasing then he/she would have a better sense of attachment towards that image and would not accept an image that depicts sadness as an emotion. Although different people may interpret the same image in different ways, we still can build a universal classification for different emotions.Emotion detection in scene images basically means that the scene image should be classified properly based on image semantics. Any image can be classified into three levels of semantics that is low level, medium level, high level. Our task is to bridge the gap between the different levels.

Keywords: Emotion Recognition, Facial Expression, Image Classification, CBIR, Support Vector Machine.

I. INTRODUCTION

The problem of emotion detection poses interesting questions from a research point of view; for instance: how to model the text for the detection task, what features offer the best prediction/detection power, and to what extent it is even possible to accurately distinguish subjective labels such as emotions from a given source text.To predict emotion, we carry out a fairly traditional machinelearning method with the addition of feature selection techniques.Specfically, the experiments here use a set of six basic emotions:happiness, sadness, anger, surprise, fear and disgust.

This work deals with the color component of an image. The tasks to be performed are: Build Classifiers for every mood, retrieve low-level semantic [5] information of the chosen image and accordingly classify the image mapping its retrieved information. Although different people may interpret the same image in different ways, we still can build a universal classification for different emotions. However, it is a challenging task for any machine to recognize emotion in any scene image we still can build classifiers [8][4] which can help the machine to adequately classify images according to different emotions.

We can use the scene images in generating greeting cards. According to the occasion for which we are sending the card we can directly generate the card by just entering the mood for which it being generated.

Content-Based Image Retrieval (CBIR) [4] has been discussed in the technical literature as a method that may develop into an efficient image search and retrieval technique. Prior work on medical image retrieval has mainly focused on extracting low-level visual features (e.g., color, texture, shape, spatial layout) and then using them directly to compute image similarity. Extensive experiments have shown, however, that low-level image features cannot always capture the biomedical semantic concepts [9] in the image.

This poses a serious shortcoming in applying CBIR to routine clinical use, where image content is defined in terms of biomedical concepts. In general, it is challenging to link high-level semantic concepts and automatically-extracted, low-level image features. Therefore, to support query by semantic concept, there is a compelling need for CBIR systems to provide maximum support towards bridging the semantic gap between the low-level visual features and the semantics in biomedical concepts.

II. RELATED WORK

A. Product Perspective:

The retrieval accuracy of content-based image retrieval systems, research focus has been shifted from designing sophisticated low-level feature extraction algorithms to reducing the ‘semantic gap’ between the visual features and the richness of human semantics. Attempts to provide a comprehensive survey of the recent technical achievements in high-level semantic-based image retrieval [2] has been done. Major recent publications are included in this survey covering different aspects of the research in this area, including low-level image feature extraction, pattern recognition, similarity measurement, and deriving high-level semantic features.

B. Product Features:

The Intended product would typically identify the emotion from a scene image contained in a database. According to the input of the user for the emotion, these images would be scanned and a suitable image would be presented in return. These images would contain only scenes excluding human intervention. Emotion detection in scene images basically means that the scene image should be classified properly based on image semantics. Any image can be classified into three levels of semantics (low level, medium level, high level).Our task is to bridge the gap between the different levels. This project deals with the color component of an image.

C. Fuzzy GIST Emotion Detection from Scene Image:

Emotion modeling evoked by scenes is challenging issue. In this paper, we propose a novel scheme for analyzing the emotion reflected by a scene, considering the human emotional status. Based on the concept of original GIST, we developed the fuzzy-GIST [8] to build the emotional feature space. According to the relationship between emotional factors and the characters of image, L*C*H* color and orientation information are chosen to study the relationship between human's low level emotions and image characteristics. And it is realized that we need to analyze the visual features at semantic level, so we incorporate the fuzzy concept to extract features with semantic meanings. Moreover, we treat emotional electroencephalography (EEG) using the fuzzy logic [11] based on possibility theory rather than widely used conventional probability theory to generate the semantic feature of the human emotions. Fuzzy-GIST consists of both semantic visual information and linguistic EEG feature, it is used to represent emotional gist of a scene in a semantic level. The emotion evoked by an image is predicted from fuzzy-GIST by using a support vector machine, and the mean opinion score (MOS) is used for performance evaluation for the proposed scheme. The experiments results show that positive and negative emotions can be recognized with high accuracy for a given dataset.

III. CLASSIFICATION

We used two methods for the classification of the local semantic concepts, k-Nearest Neighbour and Support Vector Machine classifiers. Same classification methods wereused in the initial method.

A. K-Nearest Neighbour Classifier:

The k-Nearest-Neighbour (KNN) classification is oneof the most fundamental and simple non-parametric classification methods. For k nearest neighbours, the predicted class of test sample x is set equal to the mostfrequent true class among k nearest training samples.In our work we used the mat lab implementation of KNN classifier. We tested several values of k. Best results were obtained by k = 10.

B. Support Vector Machine Classifier:

Support Vector Machines (SVM) is based on the conceptof decision hyper plane. The SVM finds a linear separating hyper plane with a maximal margin in the higher dimensional space.For our experiments, the LIBSVM package [1] withthe radial basis function (RBF) kernel was employed.LIBSVM implements the "one-against-one" approach for multi-class classification. For n = 8 classes there are n(n 1) 2 =28 single classifier and each one trains data from two classes. Each binary classification is considered to be a voting.

IV. METHODS

A. Architecture:

A system architecture or systemsarchitecture is the conceptual model that defines the structure, behavior, and more views of a system. An architecture description is a formal description and representation of a system, organized in a way that supports reasoning about the structure of the system which comprises system components, the externally visible properties of those components, the relationships (e.g. the behavior) between them, and provides a plan from which products can be procured, and systems developed, that will work together to implement the overall system.

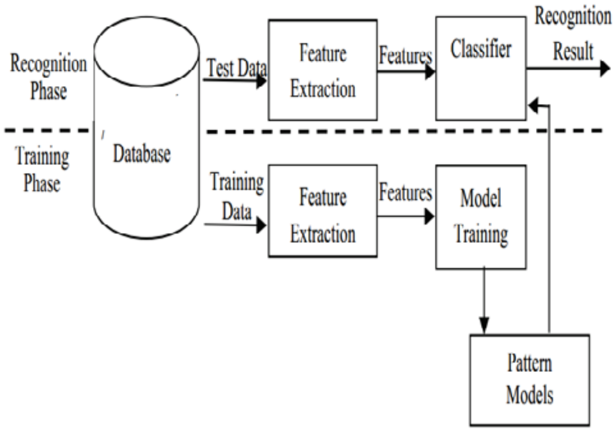


Figure 1. Architecture of a system

B. Emotion Detection From Natural Scene Image:

Emotion modelling evoked by natural scenes is challenging issue. In this paper, we propose a novel scheme for analysing the emotion reflected by a natural scene, considering the human emotional status. Based on the concept of original GIST, we developed the fuzzy-GIST to build the emotional feature space. According to the relationship between emotional factors and the characters of image, L*C*H* colour and orientation information are chosen to study the relationship between human's low level emotions and image characteristics. And it is realized that we need to analyse the visual features at semantic level, so we incorporate the fuzzy concept to extract features with semantic meanings. Moreover, we treat emotional electroencephalography (EEG) using the fuzzy logic based on possibility theory rather than widely used conventional probability theory to generate the semantic feature of the human emotions. Fuzzy-GIST consists of both semantic visual information and linguistic EEG feature, it is used to represent emotional gist of a natural scene in a semantic level. The emotion evoked by an image is predicted from fuzzy-GIST by using a support vector machine, and the mean opinion score (MOS) is used for performance evaluation for the proposed scheme. The experiments results show that positive and negative emotions can be recognized with high accuracy for a given dataset.

C. Aesthetics And Emotions In Images:

In this study, we define and discuss key aspects of the problem of computational inference of aesthetics andemotion from images. We begin with a background discussion on philosophy, photography, paintings, visual arts, and psychology. This is followed by introduction of a set of key computational problems that the research community has been striving to solve and the computational framework required for solving them.

We also describe data sets available for performing assessment and outline several real-world applications where research in this domain can be employed.

A significantnumber of papers thathave attempted to solve problems in aesthetics and emotion inference are surveyed in this tutorial. We also discuss future directions that researchers can pursue and make a strong case for seriously attempting to solve problems in this research domain.

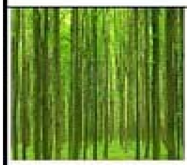
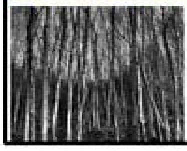
Images	Output Expected	Emotion Detected by a Particular Age Group(10 people in each group) 12-25		25-40		41-60		Majority say
		Happy	Sad	Happy	Sad	Happy	Sad	
	Happy Face	8	2	7	3	8	2	Happy
	Sad Grass	3	7	1	9	2	8	Sad
	Happy Sky	10	0	9	1	8	2	Happy
	Happy Tree	6	4	7	3	7	3	Happy
	Sad Tree	3	7	2	8	1	9	Sad

Fig 2. Testing Result based on Human Perception

D. Content Based Image Retrieval And High Level Semantics:

Semantic gap that between the visual features and human semantics has become a bottleneck of content-based image retrieval. The need for improving the retrieval accuracy of image retrieval systems and narrowing down the semantic gap is high in view of the fast growing need of image retrieval. In this paper, we first introduce the image semantic description methods, and then we discuss the main technologies for reducing the semantic gap, namely, object-ontology, machine learning, and relevance feedback. Applications of above-mentioned technologies in various areas are also introduced. Finally, some future research directions and problems of image retrieval are presented.

E. Mapping Low-Level Image Features To Semantic Concepts:

In this study, a novel offline supervised learning method is proposed to map low-level visual features to high-level semantic concepts for region-based image retrieval. The contributions of this study lie in three folds. (1) For each semantic concept, a set of low-level tokens are extracted from the segmented regions of training images. Those tokens capture the representative information for describing the semantic meaning of that concept; (2) a set of posteriors are generated based on the low-level tokens through pairwise classification, which denote the probabilities of images belonging to the semantic concepts. The posteriors are treated as high-level features that connect images with high-level semantic concepts. Long-term relevance feedback learning is incorporated to provide the supervisory information needed in the above offline learning process, including the concept information and the relevant training set for each concept; an integrated algorithm is implemented to combine two kinds of information for retrieval: the information from the offline feature-to-concept mapping process and the high-level semantic information from the long-term learned memory. Experimental evaluation on 10,000 images proves the effectiveness of our method.

F. Image Segmentation:

Automatic image segmentation is a difficult task. A variety of techniques have been proposed in the past, such as curve evolution, energy diffusion, and graph partitioning. Many existing segmentation techniques work well for mages that contain only homogeneous colour regions, such as direct clustering methods in colourspace. These apply to retrieval systems working only with colours.

G. Low-Level Image Features:

Low-level image feature extraction is the basis of CBIR systems. To performance CBIR, image features can be either extracted from the entire image or from regions. As it has been found that users are usually more interested in specific regions rather than the entire image, most current CBIR systems are region-based. Global feature based retrieval is comparatively simpler. Representation of images at region level is proved to be more close to human perception system.

V. RESULTS

This section summarizes the results of the proposed approach.

We measured the quality of human body detection by comparing the obtained results with manual detection. We calculated the overlap and left-out feature. Overlap feature determines what percentage of the manual detection (MD) is covered by the obtained result (OR).

$_{\text{Overlap=}}\text{area}(OR\backslash MD)\text{area}(MD)$

Left-out feature determines what percentage of the obtained result is not covered by the manual detection.

$_{\text{Le ft out=}}\text{area } (OR\ MD)\ \text{area}(OR)$

Our method for human body detection was tested on 15 images and we achieved average Overlap 92; 06% and average Left-out 15; 42%. The method works well if the person is standing straight. It is a typical pose on holiday pictures. If person is sitting or lying some errors may occur.

As a next step we tested which low level features are most relevant in classificationProcess. Results obtained using SVM classifier can be find in Table 1. It is obvious that color feature give a good result, but its combination with texture feature leads in even better accuracy.

The ground truth for sub region membership to one of the eight semantic concepts was annotated manually. To-gather we annotated 1028 sub regions. The class sizes vary from 54 (trunks) up to 192 (sky), because sky appears more often in the images than trunks. The classifiers are challenged with the inequality in the class sizes and the visual similarity of image regions that belong to different classes.

Table 1: Low level feature relevance

	Class size	Classification accuracy KNN	Classification Accuracy SVM
Sky	192	77,2%	82,1%
Water	139	53,4%	71,8%
grass	111	20,7%	40,0%
trunks	54	43,8%	53,1%
foliage	166	66,7%	71,6%
Sand	103	47,6%	53,3%
rocks	171	66,0%	77,5%
flowers	94	57,7%	67,7%

Table 2. KNN and SVM classification accuracies

Color	52,3%
Co-occurrence matrix	41,2%
Gabor feature	43,4%
Edge direction	25,3%
Color+Co-ocurance matrix	59,8%
Color+Gabor feature	62,5%
Color+Edge direction	56,7%
All features	67,8%

The Table 2 shows that the SVM classification per-forms better than the KNN classification. We can see a correlation between the class size and the classification result. Sky, foliage, and rocks are the largest classes and they are also classified with the highest accuracy.

VI. CONCLUSION

Emotion of an image depicting scene is identified correctly by our project. The features of the images were successfully extracted which were then given to the classification algorithm which is a Support Vector machine algorithm to segregate the image into classes and therefore detect the emotion of an image respectively.

Classification of emotions based on scene images is a new concept in an innovative field of Image Processing domain. Image processing domain has always proven to be challenging criteria in field of research and development.

This project demands a thorough study of every concept related to Image retrieval, emotion detection and CBIR (Content Based image Retrieval) technique. Emotion in scene images plays an important role in the way humans perceive an image. Based on the emotion (happiness, sadness, fear, anger etc.) of any human being the images that are viewed by that person can have a significant impact in a sense that if the person is for example in happy mood and he/she views an image that is pleasing then he/she would have a better sense of attachment towards that image and would not accept an image that depicts sadness as an emotion. Although different people may interpret the same image in different ways, we still can build a universal classification for different emotions.

VII. REFERENCES

[1] Ch. Chang and Ch. Lin. LIBSVM: A library for sup-port vector machines. ACM Transactions on Intel-ligent Systems and Technology, 2:27:1–27:27, 2011.Software available at http://www.csie.ntu. edu.tw/-cjlin/libsvm.

[2] R. S. Choras. Image feature extraction techniques and their application for cbir and biometrics systems. International Journal of Biology and Biomedical En-gineering, 1:6–16, 2007.

[3] D. Comaniciu, P. Meer, and Senior Member. Mean shift: A robust approach toward feature space analy-sis. IEEE Transactions on Pattern Analysis and Ma-chine Intelligence, 24:603 – 619, 2002.

[4] H. B. Kekre and D. Mishra. Cbir using upper six fft sectors of color images for feature vector generation. International Journal of Engineering and Technol-ogy, 2:49–54, 2010.

[5] Y. Liu, D. Zhang, G. Lu, and W. Ma. A survey of content-based image retrieval with high-level seman-tics. Pattern Recognition, 40(1):262 – 282, 2007.

[6] Jianjiang Lu, ZhenghuiXie, Ran Li, Yafei Zhang, and Jiabao Wang. A framework of cbir system based on relevance feedback. In Proceedings of the 3rd international conference on Intelligentinfor-mation technology application, IITA'09, pages 175– 178, Piscataway, NJ, USA, 2009. IEEE Press.

[7] V. Mezaris, I. Kompatsiaris, and M.G. Strintzis. An ontology approach to object-based image retrieval. In Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on, volume 2, pages II – 511–14 vol.3, sept. 2003.

[8] A. Mojsilovic and B. Gomes, J.andRogowitz. Semantic-friendly indexing and quering of images based on the extraction of the objective semantic cues. International Journal of Computer Vision, 56:79–107, 2004.

[9] E. Sikudova´. On some possibilities of automatic im-age data classification.PhD thesis, Comenius Uni-versity, Bratislava, Slovakia, March 2006.

[10] P. Viola and M. J. Jones.Robust real-time face detec-tion. Int. J. Comput. Vision, 57:137–154, May 2004.

[11] J. Vogel and B. Schiele.Semantic modeling of nat-ural scenes for content-based image retrieval. Int. J. Comput. Vision, 72:133–157, April 2007.

[12] J. Wu, Z. Lin, and M. Lu.Asymmetric semi-supervised boosting for svm active learning in cbir. In Proceedings of the ACM International Conference on Image and Video Retrieval, CIVR '10, pages 182– 188, New York, NY, USA, 2010. ACM.