EMOTION BASED CLASSIFICATION OF NATURAL SCENE IMAGES

Kuldip Kandekar[1], Ankush Lahamge[2], Ajit Tadvi[3], Abhilasha Karpe[4], Prof. P. S. Daware[4]

[1]BE Student, Computer Engineering, Sir Visvesvaraya Institute of Technology, Chincholi, Nashik, India.
[2]BE Student, Computer Engineering, Sir Visvesvaraya Institute of Technology, Chincholi, Nashik, India.
[3]BE Student, Computer Engineering, Sir Visvesvaraya Institute of Technology, Chincholi, Nashik, India.
[4]Assistant Professor, Dept. of Computer Engineering, Sir Visvesvaraya Institute of Technology, Chincholi, Nashik, 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. Thise Emotion Detection from Natural Scene Images is a new concept in an innovative field of Image Processsing domain. Image processing domain has always proven to be a challenging criterion in field of research and development.

Keyword: - *Image Description, Image classification, Content Based Image Retrieval (CBIR).*

---

1. 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 [2] [3]. To predict emotion, we carry out a fairly traditional machine learning method with the addition of feature selection techniques. Specifically, the experiments here use a set of six basic emotions [2].

2. CLASSIFICATION

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

1.2 K-Nearest Neighbour Classifier

The k-Nearest-Neighbour (KNN) classification is one of the most fundamental and simple non-parametric classification methods [3]. For k nearest neighbours, the predicted class of test sample x is set equal to the most frequent 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.

1.3 Support Vector Machine Classifier

Support Vector Machines (SVM) is based on the concept of 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] with the 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 [1].

2. 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 [3].

3. AESTHETICS AND EMOTIONS IN IMAGES

In this study, we define and discuss key aspects of the problem of computational inference of aesthetics and emotion from images [4]. 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 significant number of papers that have 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.

4. CONTENT BASED IMAGE RETRIEVAL AND HIGH LEVEL SEMANTICS

Semantic gap that between the visual features and human semantics has become a bottleneck of contentbased image retrieval [1]. 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.

5. 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 pair wise 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 [1].

6. 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 colour space. These apply to retrieval systems working only with colours [1].

8. 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.

8.1 Color Features

The color feature is one of the most widely used visual features in the image retrieval. In our work we use linear Lab_ color histogram. L_ represents the lightness, a_ the red-green component and b_ the blueyellow component. Colour histogram describes the distribution of color and lightness within the sub regions. The histogram is invariant to rotation, translation and scaling, but does not contain semantic information.

8.2 Edge Direction Features

As the second kind of feature we use edge direction histogram. It is computed by grouping the edge pixels which fall into edge directions and counting the number of pixel sin each direction. We are applying the canny edge operator and consider 4 directional edges (horizontal, vertical and 2 diagonals) and 1 nondirectional edge. Since our sub regions are arbitrary shaped we need to apply simple mirror padding to extend region to a rectangular area [5].

8.3 Texture Features

Texture is another important property of images that helps in the image retrieval. We combine texture features with other visual attribute, because texture on its own does not have the capability of finding similar images. But it can classify textured images from nontextured ones [2].

9. ALGORITHMS

9.1 Color Moment

Color moments are measures that can be used differentiate images based on their features of color. Once calculated, these moments provide a measurement for color similarity between images. These values of similarity can then be compared to the values of images indexed in a database for tasks like image retrieval. The basis of color moments lays in the assumption that the distribution of color in an image can be interpreted as a probability distribution. Probability distributions are characterized by a number of unique moments (e.g. Normal distributions are differentiated by their mean and variance). It therefore follows that if the color in an image follows a certain probability distribution, the moments of that distribution can then be used as features to identify that image based on color. Three central moments of a image's color distribution. They are Mean, Standard deviation and Skewness. A color can be defined by 3 or more values. (Here we will restrict ourselves to the HSV scheme of Hue, Saturation and brightness, although alternative encoding could just as easily be used.) Moments are calculated for each of these channels in an image. An image therefore is characterized by 9 moment’s 3 moments for each 3 color channels.

9.2 Co-Occurrence Matrix

A co-occurrence matrix or co-occurrence distribution (less often co-occurrence matrix or co-occurrence distribution) is a matrix or distribution that is defined over an image to be the distribution of co-occurring values at a given offset. Mathematically, a cooccurrence matrix C is defined over an n x m image I, parameterized by an offset (Δx, Δy), as:

$$C_{\Delta x,\Delta y}(i,j)=\sum_{p=1}^{n}\sum_{q=1}^{m}\begin{cases}1, & \text{if } I(p,q)=i \text{ and } I(p+\Delta x,q+\Delta y)=j\\ 0, & \text{otherwise}\end{cases} \quad \text{……………………(1)}$$

The 'value' of the image originally referred to the gray scale value of the specified pixel. The value could be anything, from a binary on/off value to 32-bit color and beyond [3]. Note that 32-bit colour will yield a $2^{32} * 2^{32}$ co-occurrence matrix. Really any matrix or pair of matrices can be used to generate a cooccurrence matrix, though their main applicability has been in the measuring of texture in images, so the typical definition, as above, assumes that the matrix is in fact an image. It is also possible to define the matrix across two different images. Such a matrix can then be used for colour mapping. Note that the (Δx, Δy) parameterization makes the co-occurrence matrix sensitive to rotation. We choose one offset vector, so a rotation of the image not equal to 180 degrees will result in a different co-occurrence distribution for the same (rotated) image. This is rarely desirable in the applications co-occurrence matrices are used in, so the co-occurrence matrix is often formed using a set of offsets sweeping through 180 degrees (i.e. 0, 45, 90, and 135 degrees) at the same distance to achieve a degree of rotational invariance [5].

Table-1: Confusion Matrix of the SVM Concept Classification (C=8, g=0.125). Classification is in %

Overall 67,8%	**w**	**g**	**r**	**s**	**s**	**f**	**f**	**t**
water	71,8	1,4	7,0	1,4	14,1	4,2	0,0	0,0
grass	9,2	40,0	7,5	0,8	0,0	18,3	23,3	0,8
rock	7,4	0,5	77,5	6,9	0,5	2,0	2,5	2,9
sand	0,0	6,7	23,3	53,3	10,0	3,3	0,0	3,3
sky	8,0	0,0	0,9	1,8	82,1	2,7	4,5	0,0
foliage	3,1	14,8	4,8	0,0	0,0	71,6	3,9	1,7
flowers	0,0	11,6	2,0	1,5	0,0	17,2	67,7	0,0
trunks	1,6	3,1	23,4	7,8	0,0	9,4	1,6	53,1
Precision	**54,26**	**43,24**	**75,24**	**38,109**	**86,7**	**69,2**	**73,631**	**73,9**

9.3 Average RGB

A very simple way of extracting a colour palette from an image, one technique would be to average the colour values within specific areas. Averaging colour values is almost identical to averaging numbers, except with the added initial step of finding the red, green and blue components of the colour [4] [2].

Fig-1: Average RGB.

9.4 Edge Direction Histogram

Edge directions histograms are widely used as an image descriptor for image retrieval and recognition applications. Edges represent textures and are also representative of the image shapes. In this work a histogram of the edge pixel directions is defined for image description. The edges detected with the canny algorithm will be described in two different scales in four directions. In the lower scale the image is divided into 16 sub-images, and a descriptor with 64 bins results. In the higher scale, as no image division is done because only the most important image features will be present, 4 bins result. A total of 68 bins are used to describe the image in scale-space. Images will be compared using the Euclidean distance between histograms. The provided results will be compared with the ones that result from the use of the histogram in the low scale only. Improved classification using the nearest class mean and neural networks will be used. A higher level semantic annotation, based on this low level descriptor that results from the multiscale image analysis, will be extracted [4].

10. RESULTS

This section summarizes the results of the proposed approach as tabulated in folloeing Table-2.

Table-2: Low Level Feature Relevance.

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%

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).

Overlap=area (OR\MD) area (MD)

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

Le ft out = area (OR MD) 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. (See Figure 6) As a next step we tested which low level features are most relevant in classification Process. Results obtained using SVM classifier can be find in Table 2. 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. The Table 3 shows that the SVM classification performs 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. In Table 1 is displayed confusion matrix of the SVM concept classification.

Table-3: KNN and SVM Classification.

	Class size	Classification	
		Accuracy KNN	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 %

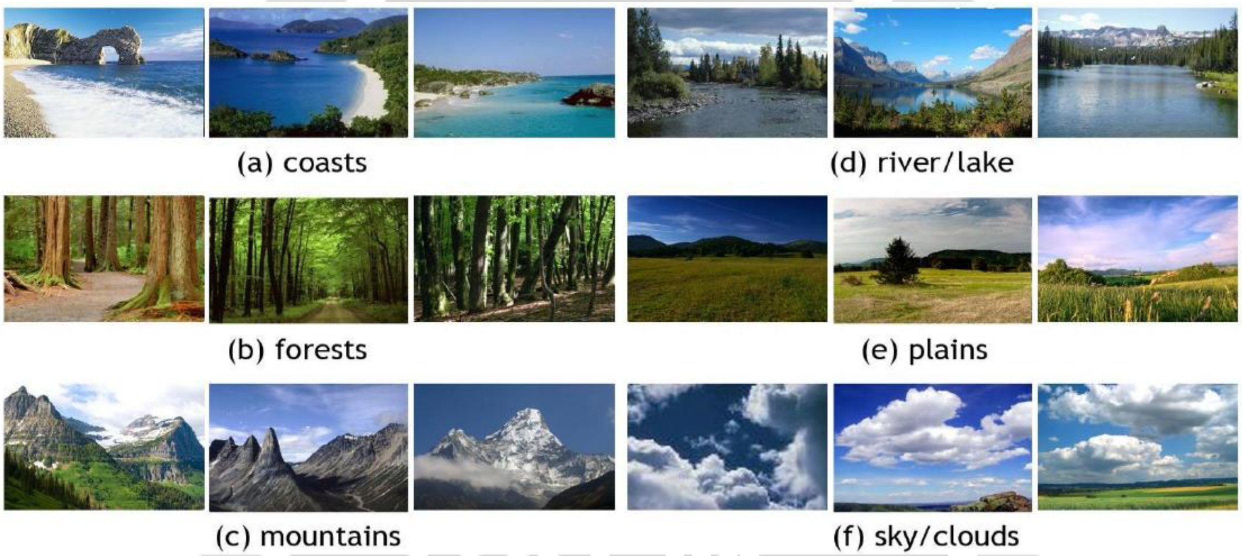

Fig-1: Exemplary Images for Each Category.

11. CONCLUSION

Emotion Detection from Natural Scene Images is a new concept in an innovative field of Image Processing domain. Image processing domain has always proven to be a challenging criterion in field of research and development.

REFERENCES

[1] A. Mojsilovic and B. Gomes, J.and Rogowitz, "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.
[2] E. Sikudova´, "On Some Possibilities of Automatic Image Data Classification" PhD thesis, Comenius University, Bratislava, Slovakia, March 2006.
[3] Jianjiang Lu, Zhenghui Xie, Ran Li, Yafei Zhang, and Jiabao Wang., "A Framework of Cbir System Based On Relevance Feedback.
[4] J. Vogel and B. Schiele, "Semantic Modeling of Natural Scenes For Content-Based Image Retrieval. Int. J. Comput. Vision" 72:133–157, April 2007.
[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.

BIOGRAPHIES

[1]**Mr. Kuldip Kandekar** pursuing Bachelor Degree in Computer Engineering from Sir Visvesvaraya Institute of Technology, Chincholi, Nashik, Savitribai Phule Pune University, Maharashtra, India.

[2]**Mr. Ankush Lahamge** pursuing Bachelor Degree in Computer Engineering from Sir Visvesvaraya Institute of Technology, Chincholi, Nashik, Savitribai Phule Pune University, Maharashtra, India.

[3]**Mr. Ajit Tadvi** pursuing Bachelor Degree in Computer Engineering from Sir Visvesvaraya Institute of Technology, Chincholi, Nashik, Savitribai Phule Pune University, Maharashtra, India.

[3]**Miss. Abhilasha Karpe** pursuing Bachelor Degree in Computer Engineering from Sir Visvesvaraya Institute of Technology, Chincholi, Nashik, Savitribai Phule Pune University, Maharashtra, India.

[4]**Prof. P.S. Daware** Assistant Professor in the Department of Computer Engineering, Sir Visvesvaraya Institute of Technology, Chincholi, Nashik, Savitribai Phule Pune University, Pune, Maharashtra, India.