

# Research on Perception-Oriented Image Scene and Emotion Categorization

Yang Zhuo\*

*Department of Electromechanical Engineering, Changzhou Textile Garment Institute, Changzhou, China*

**Abstract:** With the development of multimedia technology and computer network, the number of available images increases with an explosive speed. But the technology also brings some trouble to its users, sometimes it's very difficult for us to find some details that very important from a huge amount of available data. At this time, image scene and emotion categorization technologies are required urgently. The purpose of emotional image classification is that we hope the computer can express the emotion reaction when observing the image, and classify the images into the different emotional categories automatically. The process of scene image classification is that how to make computer systems to classify the image sets automatically which contain semantic information, according to the visual perception mechanism of human. Here, for the scene categorization problem based on the visual words, the dissertation presents a novel learning framework to design discriminating semantic visual words. At last, for the emotion categorization of natural scene images, the dissertation presents an emotion categorization using Affective-probabilistic Latent Semantic Analysis model based on the visual cognitive theory.

**Keywords:** Context feature, emotion categorization, emotional image classification, image scene, perception-oriented image scene.

## INTRODUCTION

With the development of multimedia technology and computer network, the number of available images increase with an explosive speed. Images play a very important role in people's lives. There are hundreds of GB or TB of digital image production, publishing and share [1]. Vast amounts of data bring people a variety of convenient, but it also brings a great deal of problems at the same time: it is easy to get lost in the multitude of data and difficult to find the information people really need. Faced with such a large-scale image resources, how they can be effective classification organization has become an urgent problem to be solved.

Scene classification has become an active research topic in the computer vision area, which provides important environmental clues. Usually, the user focuses on the semantic meanings reflected by image content during his classifying process. Due to the limitation on image understanding, it is quite difficult to infer the image semantics directly from its visual features. From the perspective of cognitive psychology, our work narrows down the well known semantic gap between low-level visual features and high-level semantic concepts with the hierarchical structure, following the routine of "Bag of Visual Words-Semantic Topic Model-Emotional Mapping Function" for object recognition and other computer vision tasks model based on the visual cognitive theory [2].

As an important image understanding research, scene classification is an effective way to achieve organize and manage image based on content. Makes the growing amount of image datas quickly and efficiently browsed possible, and has very important applications in the following fields [3]: (1) Image Retrieval. Image Retrieval is very close contacted to scene classification. (2) Video Retrieval. In recent years, multimedia information collection, storage and transmission technology is developing fast, video retrieval requires effective organization and description of vast amounts of video data. (3) Computer vision tasks. Image of the scene category includes not only the whole understanding of peoples' about an image, but also provide critical environmental cues for object recognition, computer vision tasks and other behavioral detection.

## The Theoretical Basis of Image Emotional Classification

Image emotion classification is a process that computer assign the images into different categories based on the emotional reaction of humans'. Scene image classification, in a short is a process that use computer to simulate humans' visual perception principles, to achieve the scenes image that contains several semantic information automatically marked [4]. Scene classification is different from general problem of image understanding, emphasizing the overall semantic representation of different distribution relationships between images, classification can be achieved without getting a full understanding of the target. Image sentiment classification is widely used, such as images and video retrieval, automatic image filtering, advertising design, etc. Scene image classification lay the foundation for achieving reasonable

and effective management of vast amounts of image data, and become a key issue in computer vision research field.

**The Basic Steps of Image Emotional Classification**

Image sentiment classification is emotion recognition in the database, and assign the image to the correct emotion category according to the emotional attributes, based on the emotional attributes that computer image expressed. Emotional attributes include happy, sad, scared, awe, excitement, satisfaction, nausea, fear and so on [5].

The basic steps are: (1) Extracte emotional characteristics of the image and clustering generate semantic visual dictionary. (2) Build pyramids description for the image space. (3) Training classifier using the training image. (4) Using classification to achieve the category determination of test emotion image (Fig. 1).

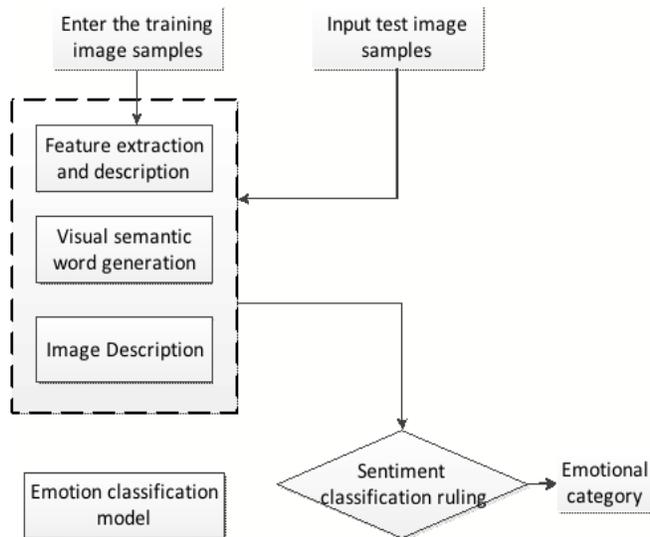


Fig. (1). Image sentiment classification ruling.

**VISUAL FEATURES**

**LBP Feature**

Loeal Binary Pattern (LBP) is used to describe the partial similarity between pixels. The basic idea of the algorithm is based on a certain pixel as the center, compare the gray value of the center pot with the pixel gray value that near to the center point, to obtain a set of binary numbers as binary pattern of center point, and further converted to decimal value of the point to get LBP. LBP algorithm is simple and not sensitive to the changes in rotation and gray value, so it is widely used by researchers [6].

Make the basic LBP algorithm defined as:

$$LBP_{R,N}(x,y) = \sum_{i=0}^{N-1} s(p_i - p_c) \times 2^i, s(x) = \begin{cases} 1, x \geq 0 \\ 0, x < 0 \end{cases} \quad (1)$$

$s(x)$  is a sign function. It represents gray comparison value between adjacent pixels and the center point.  $p_c$  is the gray value of local neighborhood center point  $c$ .  $p_i$  is the

gray value of  $N$  pixels which are  $R$  distance far away from the center point in neighborhood area. In calculating the LBP characteristic value of the image pixel, there is a high correlation between adjacent pixels.

**SIFT Feature Extraction**

Koendetink and Lindeberg and other person proved that the only linear transformation of nuclear scaling is Gaussian convolution kernel. The convolution between image  $I(x,y)$  and variable scale Gaussian function  $G(x,y,\sigma)$  can get its Scale Space  $L(x,y,\sigma)$ :

$$L(x,y,\sigma) = G(x,y,\sigma) * I(x,y) \quad (2)$$

\* is the sign of convolution, variable scale Gaussian function  $G(x,y,\alpha)$  is:

$$G(x,y,\sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \quad (3)$$

Lowe proposed that the use of different scales Gaussian kernel function and image convolution generated difference of Gaussian scale space:

$$D(x,y,\sigma) = (G(x,y,k\sigma) - G(x,y,\sigma)) * I(x,y) = L(x,y,k\sigma) - L(x,y,\sigma) \quad (4)$$

In the base of summary, Lowe proposed Scale Invariant Feature Transform (SIFT), a description operator based on local image feature of scale space. It has very good invariance in image translation, lighting transform, image scaling, rotation and even affine transform [7].

**SUMO Top-Level Architecture of Ontology Library**

The Spatial entities of SUMO Ontology library is the part of the physical world, and is described by physical entity subclass. In the child node contained by the physical entity, Region and Self Connected Object are closely contacted to Space concept. In the space of ontology library, the meaning of the two child node region and self connected object is consistent with SUMO [8]; region represents location and space on geography, such Asia and Uranus. The difference between the top space ontology concepts, terminology SUMO name and classification can be seen in Table 1.

**PLSA Model Principle**

PLSA model principle is a generating model that proposed by Hoffman for latent semantic analysis, its' main idea is to analyze the documentation set word symbiosis, for text processing at the early time. Generation process of the whole document is (1) chose a document, the probability is  $P(d)$ ; (2) chose an implicit theme  $z$ , make  $z$  satisfies polynomial distribution; (3) with knowing implicit theme, the conditional probability of the word  $w$  appears  $P(w|z)$  satisfies polynomial distribution [9]. By the generation process, the joint distribution of co-occurrences of words and documents represented as:

Table 1. The difference between sumo and spatial ontology in the top level.

The First Layer Concept	The Second Layer Concept	The Third Layer Concept	The Fourth Layer Concept	SUMO	Our Way	
Entity	Physical	Process		√		
		Object	Region		√	√
			Self Connected Object		√	√
			Collection		√	
			Agent		√	
	Abstract	Proposition		√	√	
		Attribute		√	√	
		Relation		√	√	
		Quantity		√		

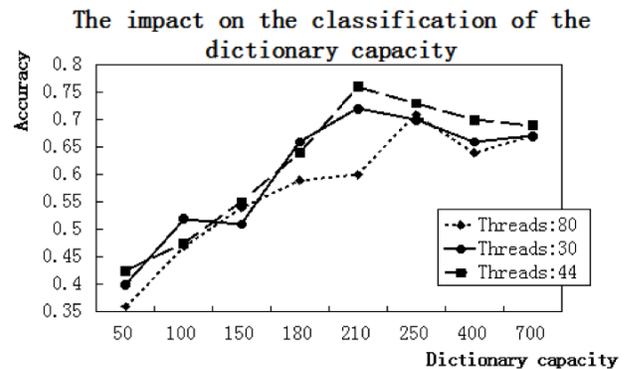
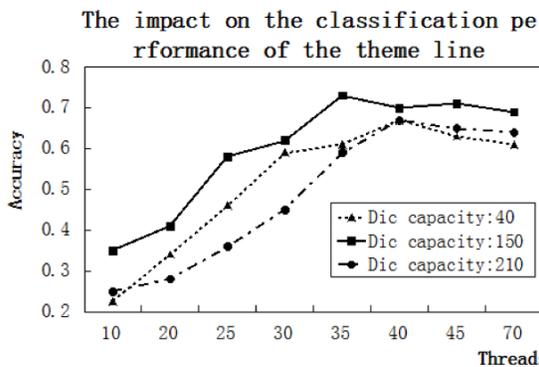


Fig. (2). The influence of the topic number and the codebook size on the classification accuracy.

$$P(w, d) = \sum_{z \in Z} P(w, d, z) = P(d) * \sum_{z \in Z} P(w|z)P(z|d) \quad (5)$$

As  $P(w, d) = P(d)P(w|d)$ , according to the formula (5),  $P(w|d)$  can be written as:

$$P(w|d) = \sum_{z \in Z} P(w|z)P(z|d) \quad (6)$$

**Markov Random Model**

Markov random model consider each pixel conditional distributionon. Its definition is: at the grid S, the point  $s = \{(i, j) | 1 \leq i, j \leq n\}$  has a state  $Z_{i,j}$ :

$$P(Z_{i,j} | Z_{S-(i,j)}) = p(Z_{i,j} | Z_{N(i,j)}) \quad (7)$$

According to Hamersley-Clifford theorem, the joint probability distribution of Markov random equals to Gibbs:

$$P(Z_{i,j}) = Z^{-1} \times \exp(-U(Z_{i,j})) \quad (8)$$

**ANALYSIS OF EXPERIMENTAL RESULTS**

**Factors Affecting the Classification Accuracy**

We can use the strategy one-against-all to construct multiple SVM classifiers, calculated each scene of the class

classifier output confidence image, select the highest confidence classifier categories as an image category [10]. The number of topic size and the capacity of the dictionary size impact on the average of all scenarios class classification accuracy can be seen in Fig. (2).

Two images are shown the trend "significantly improved from the smooth to slowly lower", the reason may be too small dictionary capacity or topics will loose too semantic and oversized dictionary capacity or topics can resulted in excessive semantic redundancy, and then affects the average classification performance.

**The Performance Analysis of Image Scene Classification Algorithm Based on Multiscale Contextual Semantic Information**

In order to verify the performance of this algorithm, it would be compared with several existing classification algorithms, including Lazebnik space pyramids model (SPM), Feifei and othe LDA model (V-LDA), SPM- pLSA, pLSA+MRF (Gudivada V N. 1998). We can see clearly in Table 2.

The results can be seen from the table, the average correct rate of classification algorithm is better than several other methods, in the case of SVM classifier, the algorithm achieved the highest rate of 88.4% correct classification, the correct rate increased 1.8 percentage points compared to

Table 2. The average correct classification rate comparison of algorithms.

Classification Algorithm	Classifier	Image Library		
		OT Library	FP Library	LSP Library
SPM	SVM	63.3%	61.6%	58.8%
	KNN	65.1%	58.3%	53.5%
V-LDA	SVM	73.5%	68.7%	64.6%
	KNN	71.2%	66.5%	60.2%
SPM-pLSA	SVM	82.5%	75.4%	69.8%
	KNN	78.8%	70.3%	65.1%
pLSA+MRF	SVM	86.6%	83.2%	79.6%
	KNN	85.1%	80.4%	77.5%
Our way	SVM	88.4%	84.5%	81.3%
	KNN	83.2%	81.6%	79.8%

pLSA+MRF method. It shows that the multi-scale information of the image has a positive effect to improve scene classification performance, and further proved that multistate information and contextual semantic information are effective to improve classification performance.

**CONCLUSION**

Image of the scene and sentiment classification are hotspots in multimedia information processing. So improve the emotion image classification recognition rate and make the ability of computer understand image close to people’s level. This must play a promoting role in the developing of image classification and image retrieval research. For the emotion categorization of natural scene images, the dissertation presents an emotion categorization using Affective-probabilistic Latent Semantic Analysis. Spatial conceptual model is a fundamental problem in the text-to-scene conversion tasks. The research on effective automatic spatial conceptual model which is described in natural language is of great theoretical and practical significance. Be similar to the words in text data, modeling scene images with visual vocabulary could form a middle representation, which describes the semantic information of scene images effectively.

**CONFLICT OF INTEREST**

The author confirms that this article content has no conflict of interest.

**ACKNOWLEDGEMENTS**

Declared none.

**REFERENCES**

- [1] A. Hanjalic, “Video and Image Retrieval beyond the Cognitive Level: The Needs and Possibilities,” In: *Proc. of SPIE Storage and Retrieval for Media Database*, San Jose, CA, vol. 4315, pp. 130-140, 2001.
- [2] B. Leibe, A. Leonardis, and B. Schiele, “Robust Object Detection with Interleaved Categorization and Segmentation,” *International Journal of Computer Vision*, vol. 77, no. 1-3, pp. 259-289, 2008.
- [3] F. F. Li, and P. Persona, “A Bayesian hierarchical model for learning natural scene categories,” USA: IEEE Computer Society Press, 2005.
- [4] T. Linderberg, “Scale-space Theory: a basic tool for analyzing structures at different scales,” *Journal of Applied Statistics*, vol. 21, no. 2, pp. 224-270, 1994.
- [5] P. Quelhas, F. Monay, and J. M. Adobes, “A Thousand Words in a Scene,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 9, pp. 1575-1589, 2007.
- [6] Z. L. Stan, “Markov random field modeling in image Analysis,” *Germany: Springer*, 2001.
- [7] V. N. Gudivada, “A Geometry-Based Representation for Efficient and Effective Retrieval of Images by Spatial Similarity,” *IEEE Trans. On Knowledge and Data Engineering*, vol. 10, no. 3, pp. 504-512, 1998.
- [8] D. M. Blei, A. Ng, and M. Jordan, “Latent dirchlet allocation,” *Journal of Machine Learning Research*, vol. 2, no. 3, pp. 993, 2003.
- [9] A. Vailaya, A. Figueiredo, A. Jain, and H. J. Zhang, “Image classification for content-based indexing,” *IEEE Transactions on Image Processing*, vol. 10, no. 1, pp. 117-130, 2001.
- [10] A. Bosch, “Image Classification for Large Number of Object Categories,” University of Girona. PhD dissertation, pp. 1-12, 2007.

Received: June 10, 2015

Revised: July 29, 2015

Accepted: August 15, 2015

© Yang Zhuo; Licensee Bentham Open.

This is an open access article licensed under the terms of the (<https://creativecommons.org/licenses/by/4.0/legalcode>), which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.