ISSN 0868 - 3980

Tạp chí KHOA HỌC & CÔNG NGHỆ CÁC TRƯỜNG ĐẠI HỌC KỸ THUẬT

JOURNAL OF SCIENCE & TECHNOLOGY TECHNICAL UNIVERSITIES

No.95
2013

CONTENTS

SCENE CLASSIFICATION FOR ADVERTISING SERVICE BASED ON IMAGE CONTENT

NHẬN DẠNG KHUNG CẢNH ỨNG DỤNG TRONG HỆ THỐNG QUẢNG CÁO DỰA TRÊN HÌNH ẢNH

Thi-Lan Le, Hai Vu, Thanh-Hai Tran

International Research Institute MICA, HUST - CNRS/UMI 2954 - Grenoble INP

Received February 28, 2013; accepted April 26, 2013

ABSTRACT

Scene classification is a fundamental problem in image understanding. Scene classification has a high potential for improving the performance of other computer vision applications such as browsing, retrieval and object recognition. However, scene classification is not an easy task owing to their variability, ambiguity, and the wide range of illumination and scale conditions that may apply. A number of works have been proposed for scene classification. However, it lacks a quantitative comparison. This paper has two mains contributions. Firstly, we provide an analysis of different types of feature for scene representation as well as a quantitative comparison of these features for scene classification. Secondly, we present a new application of scene classification: advertising service based on image content.

Keywords: scene recognition, scene analysis, image analysis

TÓM TẮT

Nhận dạng khung cảnh là một bài toán quan trọng trong phân tích và hiểu nội dung của ảnh. Kết quả của bài toán nhận dạng khung cảnh có thể được sử dụng để cải thiện hiệu quả của các bài toán khác như tìm kiếm ảnh theo nội dung, nhận dạng đối tượng. Bài toán nhận dạng khung cảnh là bài toán khó do sự thay đổi lớn của các điều kiện chiếu sáng, tỉ lệ và sự đa dạng các khung cảnh trong cùng một lớp. Mặc dù đã có nhiều đề xuất được đưa ra cho nhận dạng khung cảnh nhưng theo hiểu biết của chúng tôi chưa có một đánh giá định lượng nào được đưa ra cho bài toán này. Trong bài báo này, đầu tiên chúng tôi sẽ cung cấp một phân tích tổng quan về các phương pháp nhận dạng khung cảnh và một đánh giá định lượng về các phương pháp này. Sau đó chúng tôi sẽ giới thiệu một ứng dụng mới của bài toán nhận dạng khung cảnh: dịch vụ quảng cáo ảnh dựa trên nội dung của ảnh.

1. INTRODUCTION

Ads services in an online sharing system is becoming a very attractive topic to scientists as well as businesses. Since many years, Google has developed text based advertising services that is a great success of Google and is one of its main sources of revenue. One interesting question that appears to us is a part from text, why do not advertise based on other types of data such as images or video while these data are much more meaning and even very attractive to users than text. For instance, when the user looks at an image of beach, an advertisement on tourism promotion could be pushed on. From this idea, we propose to lead our research on image categorization, in particular scene image classification that will be integrated in an advertising service.

To the best of our knowledge, this is a novel application that no one has conducted to before. The problem of scene recognition is defined as follows: Given an image of scene, identify the class label to which this scene belongs to. Our main contributions in this paper are: (1) an analysis on different types of feature representing a scene image including global and local features; (2) a novel application of scene recognition for advertising service in online sharing system.

2. RELATED WORKS

A general framework of image recognition consists of two main phases: learning and recognition. For each phase, features representing a scene need to be extracted for learning scene model (in learning phase) or for categorizing an image of scene (in recognition phase).

In [1], the authors proposed to represent a scene image by its "GIST". GIST presents a brief observation or a report at the first glance of an outdoor scene that summarizes the quintessential characteristics of an image. Authors in [1] shown that the GIST may be reliably estimated using spectral and coarsely localized information. With a limitation number of scenes, precisions of outdoor scene classification using GIST features can obtain in range from 75% to 82%. Recently, in [2], the authors have indicated that GIST is good for natural scene (outdoor) but as GIST ignores the object details in the scene, it can not deal with indoor scenes. Therefore, these authors have proposed a descriptor Centrist (Census Transform Histogram) that combines both local and global information in the image. The method obtained 84.96% in term of classification rate for 15 scene classes. However, the descriptor Centris is not invariant to rotation. In addition, it ignores color information of the scene. GIST and Centrist are global features. HOG (Histogram of Oriented Gradient) and SSIM (Self Similarity Descriptor) that provide the spatial ranging of the scene and grayscale histogram have been studied in [3]. For all types of descriptor, SVM (Support Vector Machine) is used for classification. With 15 scenes database, using only SIFT descriptor gives 81.4% in term of classification while combining features improve to 88%.

3. PROPOSED APPROACH

3.1. General description

The works proposed for scene classification distinguish each other by the techniques chosen for feature extraction and classification. As we present in the related works, a number of features and classification methods have been selected for scene classification. Among various features proposed for the scene classification, we choose three features: color histogram and color moment, Local Dominant Orientation (LDO) and GIST. Concerning classification method, we use classic methods such as SVM (Support Vector Machine) and kNN (k Nearest Neighbor). In the next section, we describe in detail the extracted features and classification methods.

3.2. Extracted features

a. Color histogram and color moment

Color histogram is a widely used feature in image indexing and retrieval. A color histogram represents the number of pixels that have colors in each of a fixed list of color ranges. For one color image, we can extract one histogram of each color component of the image or convert the color image to gray image and compute only one color histogram for this image. Histogram of gray image is a vector of 256 elements corresponding to 256 level of gray. Normally, the dimension of the histogram is reduced in order to increase the speed of classification. In our work, we reduce the dimension of histogram to 24.

Besides color histogram, the color moments give summary information of image content. We can compute three color moments for an image containing N pixel as follows:

$$E_i = \frac{1}{N}\sum_{j=1}^{N} p_{ij}$$

$$\delta_i = \left(\frac{1}{N}\sum_{j=1}^{N}\left(p_{ij} - E_i\right)^2\right)^{1/2}$$

$$s_i = \left(\frac{1}{N}\sum_{j=1}^{N}\left(p_{ij} - E_i\right)^3\right)^{1/3}$$

where p_{ij} is color intensity of pixel (i,j).

b. Local Dominant Orientation (LDO) Features

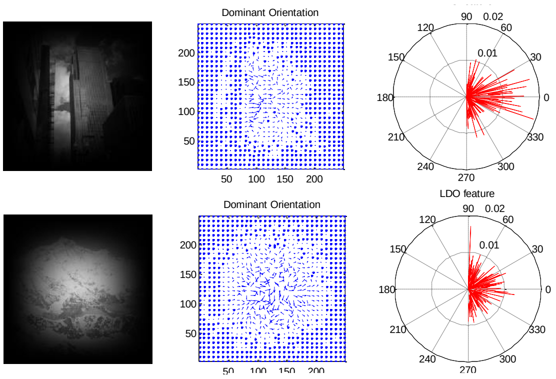
Global structure of outdoor scenes is highly relevant to distributions of edge orientations [4,5]. To evaluate these characteristics, Local Dominant Orientations (LDO) features are extracted based on works in [5].

An input image *I(x,y)* is partitioned in 8x8 pixels non overlapped blocks. For each block $B_k(x, y)$, two features that are the dominant orientation of the block θ_k and the strength of the dominant orientation A_k are calculated as below:

$$\tan(\theta_k) = \sum_{u=1,u=u+2}^{7} D(k,0) / \sum_{v=1,v=v+2}^{7} D(0,v)$$

$$A_k = \sum_{u=1}^{7} D_k^2(u,0) + \sum_{v=1}^{7} D_k^2(0,v) + \sum_{u=1}^{7}\sum_{v=1}^{7} D_k^2(u,v)$$

with $D_k(u,v)$ is Discrete Cosine Transformation (DCT) coefficients of 8x8 block $B_k(x,y)$.



(a) (b) (c)

Fig. 1. LDO feature extractions. Top: high-building image; Low: mountain image. (a) Left: input image. (b) LDO feature extraction for each block 8x8 pixels. (c) Histogram of LDO with d=256 bins

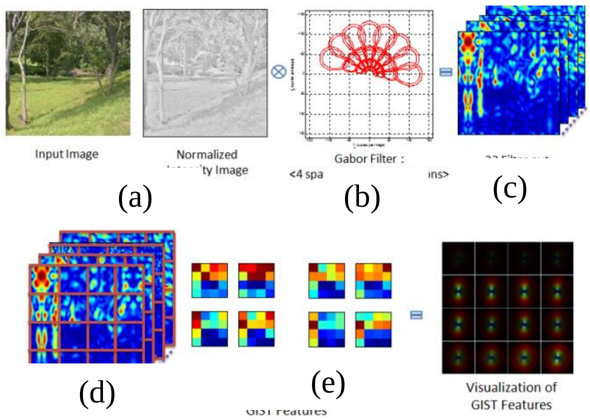
Figure 1(b) shows local dominant orientations of high building (Fig.1-(a)-top) vs. mountain scene (Fig.1(a) bottom). To represent holistic of the whole image I, we then create a polar histogram H=[$f_{\theta_1}, f_{\theta_2}, ..., f_{\theta_d}$], of LDO, containing *d-directions*, with $f_{\theta_i} = \frac{\sum_{A_k \in D} \log(A_k)}{SN}$, *SN* is the normalization constant. Figure 1(c) show polar histograms *H* of LDO with pre-determined *d = 256 bins*. As shown, the main directions of high building are biased toward horizontal orientations (perpendicular with tall building, Fig. 1(a) top), whereas domain orientation of mountain image are spanning uniformity (Fig. 1(a)-bottom). The feature vector (*256-dimensions*) therefore is separable two different scenes.

c. GIST of scene

An equivalent of model in[4] for extracting GIST of scene is shown in Figure 2. Firstly, an original image (Fig. 2(a)-left) is converted and normalized to gray scale image I(x,y) (Fig. 2(a)-right). We then apply a pre-filtering to I(x,y) in order to reduce illumination effects and to prevent some local image regions to dominate the energy spectrum. The filtered image I(x,y) then is decomposed by a set of Gabor filters. A 2-D Gabor filter is defined as follows:

$$h(x,y) = e^{-\frac{1}{2}\left(\frac{x^2}{\delta_x^2} + \frac{y^2}{\delta_y^2}\right)} e^{-j2\pi(u_0 x + v_0 y)}$$

As shown in Fig. 2b, configuration of gabor filters contains 4 spatial scales and 8 directions. At each scale (δ_x, δ_y), by passing the image I(x,y) through a Gabor filter *h(x,y)*, we obtain all those components in the image that have their energies concentrated near the spatial frequency point (u_0, v_0). Therefore, the GIST vector is calculated using energy spectrum of 32 responses (Fig. 2c). To reduce dimensions of feature vector, we calculated averaging over grid of 4x4 on each response, as shown in Fig. 2(d). Consequently, the GIST feature vector is reduced to 512 dimensions. (Fig. 2(e).)



(a) (b) (c)

(d) (e)

Fig. 2. GIST feature extraction

3.2. Classification

a. Support Vector Machine (SVM)

Support Vector Machine, a supervised learning method. The goal of SVM is to produce a model (based on the training data) which predicts the target values of the test data given only the test data attributes. When using SVM for scene classification, we compute color histogram and color moments from training images dataset and train SVM model. For an image, in order to determine scene label of this image, we predict this label by using trained SVM model.

b. k-NN

K-Nearest neighbor (K-NN) classifier is selected for classification using LDO and GIST feature because they are high dimensional descriptors. Given a testing image, we found K cases in the training set that is minimum distance between the feature vectors (LDO and GIST) of the input image and those of the training set. A decision of the label of testing image was based on majority vote of the K label found. In this work, we select Euclidian distance that is usually realized in the context of image retrieval.

4. EXPERIMENTAL RESULTS AND DISCUSSIONS

4.1. Application and Database

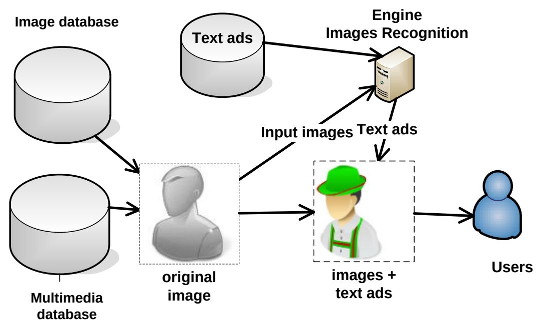


Fig. 3. The overall of the system for ads service based on image content

The work presented in this paper is developed in the context of advertising service application based on image content. This system contains: (1) Images and multimedia database; (2) Text ads database contains advertising texts; (3) Image recognition engine that analyze images and return the content of image in term of recognized objects or scenes; (4) Text ads generation module that puts ads text on the image in function of recognized objects and scenes (Fig. 3).

The system works as follows: When user opens a webpage containing image and moves the mouse on the image, the image recognition engine will be activated and the output of this module will be used to generate the ads string. The ads string will be put on the original image in a certain time. This ads string can link to other webpages. Since an image can contain objects or represent a specific scene. The image recognition engine is divided into two main parts: object recognition and scene recognition. Our object recognition work has been presented in [6]. In this paper, we focus on scene classification part.

In order to evaluate scene classification methods, we need to prepare database of scene images. While working with the company that is interested in this application, we define to work with 5 types of scenes: *beach*, *pagoda*, *mountain*, *street* and *forest*. Since, there are no available image database containing all these scenes, we need to prepare ourselves this database. Images that we choose to build database comes from two sources. The first source is upanh.com website. This is the website of our partner company allowing users to upload photos for sharing on the Internet. The second source is a free scene database (http://people.csail.mit.edu/torralba/code/spatial envelope/). Then we build a dataset of 2500 images of 5 scenes (each scene has 500 images) (see Tab.1). We divide the database into 2 parts: training and testing, each contains 1250 images.

4.2. Evaluation results

We use Precision as an evaluation measure. This measure is defined as: TP/TP+FP where TP is true positive and FP is false positive. Table 2 shows the obtained result of scene classification when using color histogram and color moment as descriptor and SVM as classification method, LDO as descriptor and kNN as classification and GIST as descriptor and kNN as classification respectively. Concerning SVM, after testing with differnt type of SVM and kernel, we choose nu-SVC with polynomial kernel. The chosen value of k in KNN is 5. The obtained results show that GIST is robust feature for scene classification.

For histogram of LDO features, although [5] reported that it is efficient for discriminating natural and artificial scenes (appr. 90% of precisions), performances of LDO are quickly decreased with multi scenes classification. In term of implementations, major different between LDO and GIST is impacts of spatial scales. Therefore, in our opinion, scale-space is an important factor to increase performance of scene classifications. Based on results in Table 2, we decide to choose GIST and kNN for our application (advertising service based on image content). This application has been implemented and deployed in http://quangcaoanh.com/. Concerning computation time, the classification method takes 96.5 ms/image.

Table 1. Examples of scene images

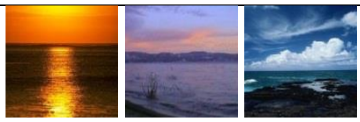
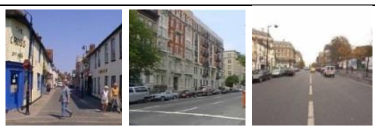
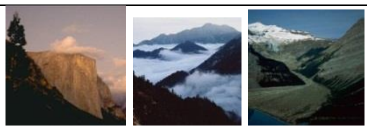
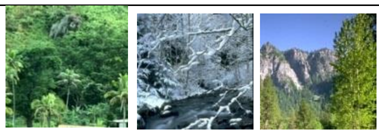
Beach	
Pagoda	
Street	
Mountain	
Forest	

Table 2. Scene classification result

Scene	Precision (%) of color histogram and SVM	Precision (%) of LDO and kNN	Precision (%) of GIST and kNN
Beach	64.2	51	86.8
Pagoda	68.2	55	93
Mountain	60.4	48	82.9
Street	66.2	59	92.4
Forest	80.1	64	84.7
Average	**67.82**	**55**	**87.96**

V. CONCLUSIONS

This paper gives a quantitative comparison of performance of three methods for sence classification. The obtained results prove that GIST is a robust descriptor for scene classification. The paper has two mains contributions. Firstly, we provide an analysis of different types of feature for scene representation as well as a quantitative comparison of these features for scene classification. Secondly, we present a new application of scene classification: advertising service based on image content

Acknowledgements: This work is supported by the project "Some Advanced Statistical Learning Techniques for Computer Vision" funded by the National Foundation of Science and Technology Development, Vietnam under grant number 102.01-2011.17.

REFERENCES

1. A. Oliva and A. Torralba "Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope." International Journal of Computer Vision **42**(3): 145–175, 2001.
2. J. Wu and J. M. Rehg, "CENTRIST: A Visual Descriptor for Scene Categorization." IEEE TRANS. PAMI, 2009.
3. Lazebnik, C. Schmid, et al., "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories". Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2006.
4. A. Guerin-Dugue, A. Oliva, "Classification of scene photographs from local orientation features", Pattern Recognition Letters, Vol. 21, pp. 1135-1140, 2000
5. S. Battiato, G. M. Farinella, G. Gallo, E. Messina, "Naturalness Classification of Images into DCT Domain", Proc. SPIE 7250, Digital Photography V, 72500F (January 19, 2009)
6. Quoc-Hung Nguyen, Thi Thanh- Hai Tran, Thi-Lan Le, Quang-Hoan Nguyen, "Object detection and classification for advertising service based on image content", 7th SEATUC , 2013.

Author's address: Le Thi Lan – Tel.: (+84) 904 412 844, Email: Thi-Lan.Le@mica.edu.vn
Hanoi University of Science and Technology
No.1 Dai Co Viet Str., Ha Noi, Viet Nam