ISSN 0868 - 3980

Tạp chí KHOA HỌC & CÔNG NGHỆ

CÁC TRƯỜNG ĐẠI HỌC KỸ THUẬT

JOURNAL OF SCIENCE & TECHNOLOGY

TECHNICAL UNIVERSITIES

C

No.95
2013

CONTENTS

OBJECT CLASSIFICATION: A COMPARATIVE STUDY AND APPLYING FOR ADVERTISEMENT SERVICES BASED ON IMAGE CONTENT

ĐÁNH GIÁ, SO SÁNH MỘT SỐ PHƯƠNG PHÁP PHÂN LỚP ĐỐI TƯỢNG ỨNG DỤNG TRONG CÁC DỊCH VỤ QUẢNG CÁO DỰA TRÊN NỘI DUNG ẢNH

Quoc Hung Nguyen[1], Thanh Hai Tran[1], Thi Lan Le[1], Hai Vu[1], Ngoc Hai Pham[1]
Quang Hoan Nguyen[2]
(1) Hanoi University of Science and Technology
(2) Hung Yen University of Technology and Education
Received February 25, 2013; accepted April 26, 2013

ABSTRACT

In this paper, we address the problem of automatic object classification from images. Specifically, we develop three approaches: (1) Haar-like feature based combined with Cascaded Adaboost Classifier; (2) Histogram of Oriented Gradient combined with Support Vector Machine and (3) Gist descriptor using K- Nearest Neighbors. Currently, each method has been shown to be efficient for specific kinds of objects (1 - hand posture, 2 - standing person, 3 - natural scene). We then evaluate the performance (in term of computational time and precision) of each method on a predefined and challenge database containing 10 types of object classes. The experimental results allow us to choose the most robust one to deploy an image content based ads system.

Keywords: object recognition, object classification, image analysis

TÓM TẮT

Trong bài báo này, chúng tôi trình bày các nghiên cứu về phân lớp đối tượng ảnh (1) phương pháp dựa trên đặc trưng Haarlike và bộ phân lớp đa tầng Adaboost; (2) phương pháp dựa trên đặc trưng về lược đồ hướng các vector gradient (HOG) và SVM; (3) phương pháp dựa trên bộ mô tả toàn thể Gist và giải thuật K người hàng xóm gần nhất KNN. Các phương pháp này được chứng minh là hiệu quả đối với một số loại đối tượng cụ thể. Chúng tôi đánh giá thực nghiệm ba phương pháp này (dựa trên thời gian tính toán và độ chính xác phân lớp) trên một CSDL ảnh của 10 lớp đối tượng. Các kết quả so sánh thực nghiệm sẽ cho phép chúng tôi lựa chọn ra phương pháp phù hợp để triển khai hệ thống quảng cáo dựa trên nội dung ảnh.

1. INTRODUCTION

This paper focuses on studying generic object classification methods that means they must work with many types of object classes, not only one specific class. The main objective is to build visual content based ads services in an online sharing system. Regarding ads services over www, Google Inc., has developed a great product, named AdWord, that is a text based advertising service. Approaching this motivation, in this paper, we describe a new advertising system, whose ads-functions are based on image contents instead of traditional text. Our main contributions in this paper are:

- An comparative study on object classification methods;
- A demonstrator of ads services based on image contents that works in real-time.

The paper is organized as follows. Section II presents related works on object classifications. Section III describes three generic methods that we develop in this context for evaluation. Section IV shows experimental results and application. Section V concludes and gives some ideas for future works.

2. RELATED WORKS

In this section, we briefly survey related works for object recognition and internet multimedia advertising services.

2.1. Object Recognition

The works proposed for object recognition can be divided into two approaches. The first approach performs object detection before object classification while the second approach executes object detection and object

classification at the same time. The works belonging to the first approach try firstly to separate object regions from background in the input image and then classify the object regions into classes. In this approach, object regions and background are separated by image segmentation techniques [1] or using objectness measures [2].

The works of the second approach do not perform explicitly object detection. In these works, binary classifiers are built for each object category of interest. In order to detect and recognize object, these approaches usually use sliding windows scanning techniques through an image. The binary classifier will be checked at each window in order to determine the presence of the interested object. A huge number of low level features are proposed for object recognitions. In this paper, we do not try to give an exhaustive overview of object recognition method, but point out some important works. In [3], the authors proposed an object recognition method based on HOG feature (Histogram of Oriented Gradient) and SVM (Support Vector Machine). This method has been proved effective for human detection. HOG feature allows to capture the local object appearance and shape within an image by using the distribution of intensity gradients or edge directions. Recently, Haar and Adaboost have been proposed for object detection/recognition in general and face detection in particular [4].

2.2 Internet multimedia advertising services

Recent technologies and systems of internet multimedia advertising are surveyed in [5], [6], [7]. Particularly, Microsoft has introduced ImageSense, VideoSense, GameSense, so on , that are modern multimedia advertising schemes. For example, ImageSense utilizes visual content detection in an image, it automatically embeds a relevant product logo in the most nonintrusive region in each image. Although these products are undergoing development, they have shown to be effective and efficient advertising on multimedia content.

3. PROPOSED APPROACH

As we can see in the related works section, many feature types and machine learning algorithms have been proposed for object classification. However, each method seems to be specific for one type of object class. For example, HOG-SVM is designed for human detection, GIST-KNN is for scene classification, Haar-Adaboost is widely used for face detection. In this paper, we would like to investigate the performance of each method for multiclass object recognition problem. The main idea is to discover the most robust one for ads service based on image content. Comparing to the paper [8] that uses only Haarlike feature and Adaboost Classifier for object classification, this paper makes a comparative study on several methods and choose the best one for applying in the frame work of our project

3.1. Object classification problem

The object classification module takes an image as input and returns a vector of N elements of value 0 or 1 corresponding to the fact that the image has or has not an object of interest among N predefined object classes. The Fig.1 shows the input and output of this module.

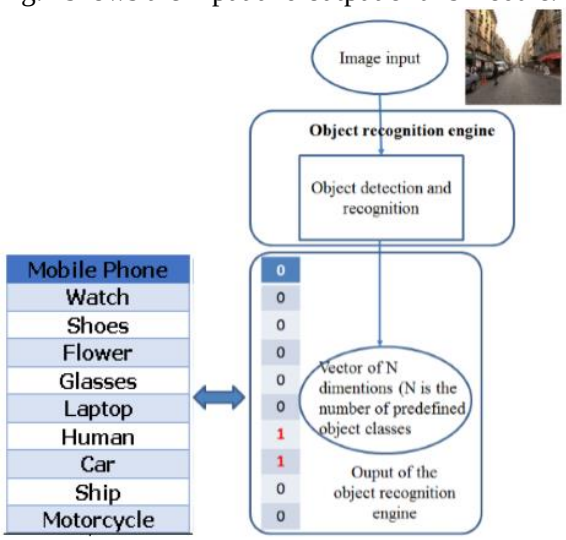

Fig.1. Illustration of the input and output of the object recognition module

3.2. Comparative study for object classification

Various vision-based approaches, which utilize different types of low level features and classifiers have presented in the literature. In this section, we evaluate three different approaches for object classifications. This study is to select the best one which is a reliable

approach for object classifications in problem of Ads service application.

Three candidates below are selected:

- Haar and Adaboost
- HoG and SVM
- GIST and K-NN

As aforementioned, these methods are suggested due to their effectiveness in each classification problems. For each method, we propose to learn each object class by a binary classifier. At classification phase, sliding window technique is used to scan the whole image; each window candidate will be passed through feature extraction module then the computed descriptor will be passed into the corresponding binary classifier (Fig.2).

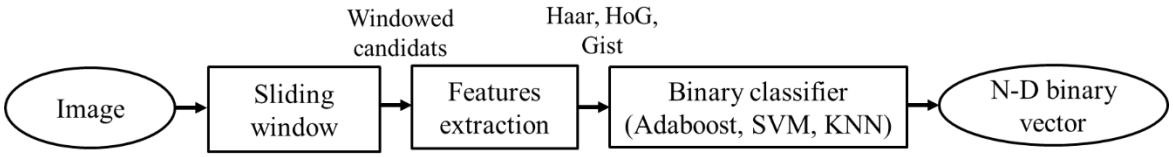

Fig.2. Common framework of three methods

3.2.1. Haar and Adaboost

For object detection and recognition, we follow the framework proposed by [9] that uses Haarlike features for object representation and Cascaded Adaboost Classifier for object classification. The reasons for which we use this framework are : i) it has been shown to be efficient for object recognition in general ; ii) it can work in real time ; iii) it is easy to expand the number of class of objects to be recognized.

- *Haar-like feature* is a feature built from equal-sized rectangles, used to calculate the difference between the values of adjacent image in the region. The value of a Haarlike feature is computed as difference between the sum of all image pixels values in the dark and light rectangular areas. The value of Haarlike feature is computed fast by integral image technique.

- *Cascaded Adaboost Classifier* is a degenerated decision tree where at each stage a classifier is trained to detect almost all objects of interest while rejecting a certain fraction of the non-object patterns (Fig.3).

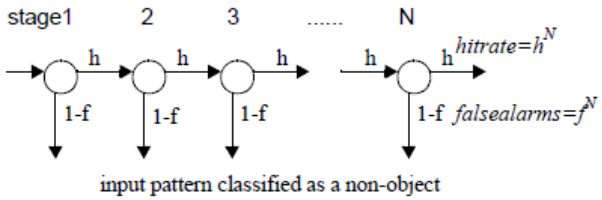

Fig.3. Cascade of Classifiers with N stages.

Each stage was trained using the Discrete Adaboost algorithm. Discrete Adaboost is a powerful machine learning algorithm. It can learn a strong classifier based on a (large) set of weak classifiers by re-weighting the training samples. Weak classifiers are only required to be slightly better than chance. At each round of boosting, the feature-based classifier is added that best classifies the weighted training samples.

3.2.2. HOG and SVM

Among image descriptors, HOG has been proved robust for object detection and object identification. In our work, we propose to study HOG for object identification based on object information. The object identification method based on HOG consists of three steps. Firstly, HOG features are computed for all images in the database. In the second step, we propose to use Maximum Margin Criterion (MMC) to reduce the descriptor dimension. Finally, for object classification, we apply SVM.

HOG descriptor is proposed in [3]. In order to compute HOG, firstly image is divided into squares cells and blocks. Then, the gradient and its direction of pixels in cells are calculated. After this step, a histogram is created. Each bin of this histogram contains the number of pixels in the same direction. The histogram of each block is created by accumulating histogram of its cells. Finally, all histogram are arranged to from HOG descriptor of an image. Since the

dimension of HOG descriptor is relatively high (with an image of 320*240 of resolution, the size of cell is 16*16 pixels and the size of block is 2×2 cells in Fig.4, HOG descriptor dimension is 9576) and not all element of HOG descriptor are relevant to object representation, before applying classification method, we have to reduce the dimension of HOG.

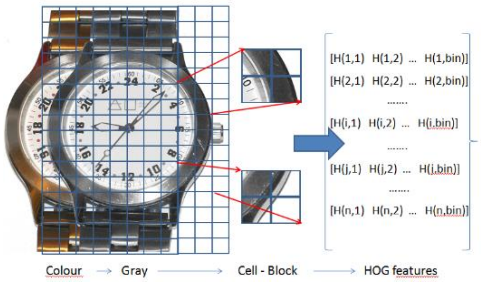

Fig.4. HOG features extraction

We then use MMC technique to reduce dimension of features to 100 because experiment shows that only 100 first eigenvalues are significant. Concerning the classification method, SVM was selected for classification in our research due to its high accuracy work with high dimensional data and its ability to generate non-linear and well as high dimensional classifier.

3.2.3. Gist and Knn

Results of variety of state of the art scene recognition algorithms [10] shown that GIST features[1] [11] obtains an acceptable result of outdoor scene classification (appr. 73 – 80 %). Therefore, in this study, we would like to investigate if GIST features are still good for object classification. In this section, we briefly describe procedures of GIST feature extractions proposed in [11].

To capture remarkable/considering of a scene, Oliva et al in [11] evaluated seven characteristics of a outdoor scenes such as naturalness, openness, roughness, expansion, ruggedness, so on. The authors in [11] suggested that these characteristics may be reliably estimated using spectral and coarsely localized information. Steps to extract GIST features are explained in Fig.5. Firstly, an original image is converted and normalized to gray scale image I(x,y) (Fig.5 (a) – (b)). We then apply a pre-filtering to reduce illumination effects and to prevent some local image regions to dominate the energy spectrum. The image I(x,y) is decomposed by a set of Gabor filters. The 2-D Gabor filter is defined as follows:

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

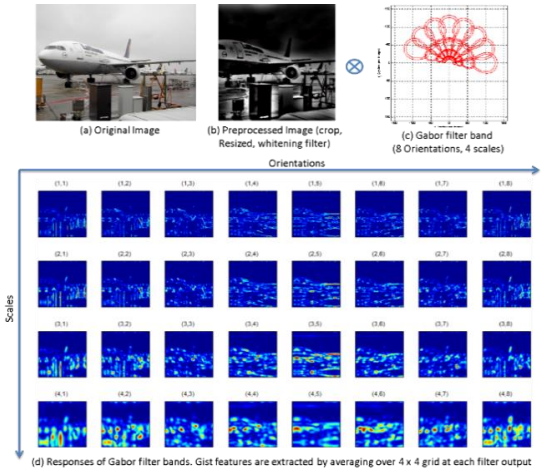

Fig.5. GIST Feature Extraction Procedures

The parameters (δ_x , δ_y) are the standard deviation of the Gaussian envelope along vertical and horizontal directions; (u_0 , v_0) refers to spatial central frequency of Gabor filters. As shown in Fig.5 (c), configurations 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. We calculated averaging over each grid of 16 x 16 pixels on each response, as shown in Fig.5 (d).

Totally, a GIST feature vector is reduced to 512 dimensions. After feature extraction procedure, K-Nearest neighbor (K-NN) classifier is selected for classification. Given a testing image, we found K cases in the training set that have minimum distance between the gist vectors of the input images and those of the

[1] *Gist feature present a brief observation or a report at the first glance of a outdoor scene that summarizes the quintessential characteristics of an image*

training set. A decision of the label of testing image was based on majority vote of the K label found. The fact that no general rule for selected appropriate dissimilarity measures (Minkowsky, Kullback-Leibler, Intersection..). In this work, we select Euclidian distance that is usually realized in the context of image retrieval.

4. EVALUATION AND APPLICATIONS

4.1. Database

To train the classifiers and test the recognition system as well as the ads service based on image content, we need to prepare database of images. Building a dataset for training a classifier is not a simple problem because we need to ensure that the image number to be big enough and representative to represent one object class and discriminative from other classes. Images that we choose to build database comes from two sources:

- One is from www.upanh.com website. This is a most used Vietnamese website allowing users to upload photos for sharing on the Internet. Number of images is 4250.

- One is from the dataset of PASCAL VOC [12], a famous contest on object detection and classification for computer vision community. Number of images is 750

Then we build a dataset of 10 object classes: Mobile Phone, Wacth, Shoes, Flower, Glasses, Laptop, Human, Car, Ship, and Motorcycle. Each object class has 500 images that make a total number of images in the database is 5000 images. In addition, our database is very challenge because of very invariant instances of a class. And objects are taken from different views point and lighting conditions in the Fig.6. The images in the database is annotated manually and organized in the directory as.

Fig.6. Some images in our database

We divide the database into 2 parts; each contains 2500 images of 10 object classes, one for training and one for testing. As we have 10 object classes, we will train 10 Classifier

4.2. Integration of the recognition module in the overall system

In order to develop ads service based on image content, we have worked with Naiscorp company. This company has developed a website for uploading and sharing images www.upanh.com and a website for ad service based on image content www.quangcaoanh.com. The object recognition module is dependently developed and called as a web service. If website quangcaoanh.com needs to generate an ads string, it sends the URL of this image to our object recognition module. The object recognition module does its work and returns a list of recognized objects in this image.

4.3. Evaluation measure

There are many measures for evaluating recognition performance such as Recall, Precision, and Accuracy. In our context, as we know the distribution of positive and negative examples (the ratio between positive and negative is 1/9) so we propose to evaluate our system by Precision criterion. The precision is defined as True Positive /(True Positive + False Positive) [13].

4.4. Experimental Results

We do experiments with 2500 images in the test part of our database. One image is positive of one object class but negative of the others classes. The following table gives the precision for each object class and average precision.

Table. 1. Experimental results

N_0	Object Classes	Haarlike Adaboost	HOG SVM	Gist KNN
1	Mobile Phone	0.97	0.67	0.88
2	Watch	0.98	0.95	0.81
3	Shoes	0.34	0.67	0.73
4	Flower	0.9	0.76	0.75
5	Glasses	0.91	0.87	0.98
6	Laptop	0.62	0.78	0.99
7	Human	0.91	0.9	0.77
8	Car	1.00	0.85	0.91
9	Ship	1.00	0.78	0.92
10	Motorcycle	0.56	0.88	0.96
	Average	**0.82**	**0.81**	**0.87**

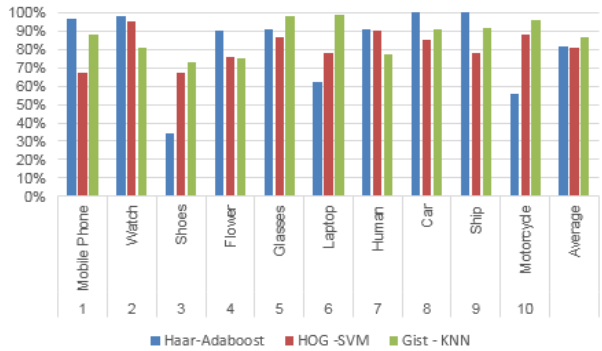

Fig.7. Comparison of three methods for object recognition

Table.2. Computational time

N_0	Classification	Times (ms)
1	Haarlike and Adaboost	**95.60**
2	HOG and SVM	**150.00**
3	Gist and KNN	**88.12**

There methods run on a computer with the following configurations *(CHIP Intel(R) Core(TM) i5-2520M CPU @ 3.2 GHz x 2, RAM 8GB)*. The average resolution of images is 600x400.

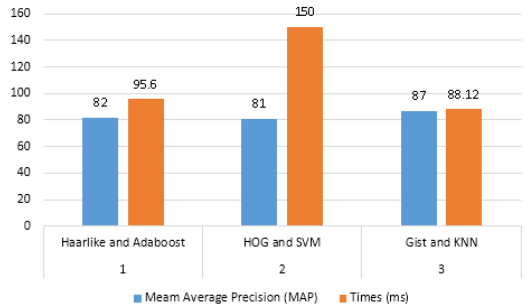

Fig.8. Performance evaluation chart

Look at Table. 1, we found that no method is good for all types of objects. Haar-Adaboost is the best method for Mobile Phone, Watch, Flower, Car and Ship recognition. This is because Haarlike features represent very well the detail of these objects. GIST– KNN is the best method for Shoes, Glasses, Laptop, Motorbike recognition. The reason is that in these images, scene background plays an important role so the gist is powerful for representing scene. In average, GIST-KNN promisingly gives the best results. That means that GIST – KNN is not only good for scene classification but also still good for object classification. Concerning computational time, Table.2 shows that GIST-KNN the most rapid method. The Fig.9 shows some results of object classification using GIST-KNN. We see that our system can give more than one answer for an image (for example image of watch and flower). In most cases, object are correctly detected and classified (watch, flower, glasses), but there are also some false alarms (for example in image of water lily flower, we detect also human and or motorcycles). The reason for false alarms is at a certain scale, theses image regions look similar to positive samples of some object classes.

Fig.9. *Several recognition results*

4.4 Ads service based on image content

We has integrated our object recognition engine into the whole system and built a demo for ads service based on image content. This demo is on quangcaoanh.com from this webpage (Fig.10), when the user can click to see an image of interest, the URL of this image will be sent to the object recognition module and the ad string generated based on image content will be displayed.

Fig.10. Object recognized in image and the ads string is generated at the end of images

5. CONCLUSIONS

This paper presented a comparative study on object classification and a framework for ads service based on image content. We have studied 3 methods: Haar - Adaboost; HOG - SVM; GIST - KNN. The experimental results have shown that "Gist of scene" and KNN is the best combination for object classification in

both term of precision and computational time. This fact shows that when the number of object classes is big, the gist that represents the global perception of the scene is still a good feature for discriminating object classes. Therefore, GIST - KNN has been selected to be integrated into the ads service system based on image content. In the future, we will test with more object classes and release the ads services based on image content for and users.

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. Winn, J. and A. Criminisi, Object Class Recognition at a Glance. Proc.Conf. Computer Vision and Pattern Recognition (CVPR) - Video Track, 2006.
2. B. Alexe, et al., What is an object? CVPR 2010, 2010.
3. Navneet Dalal, B.T., Histograms of Oriented Gradients for Human Detection. Computer Vision and Pattern Recognition, 2005. **1**.
4. Lowe, D.G., Distinctive Image Features from Scale-Invariant Keypoints. Int. J. Comput. Vision, 2004. **60**(2): p. 91-110.
5. Xian-Sheng Hua, et al., Online Multimedia Advertising: Techniques and Technologies. Igi Global, 2011.
6. Wei Jiang, et al., An online advertisement platform based on image content bidding, in Proceedings of the 2009 IEEE international conference on Multimedia and Expo2009, IEEE Press: New York, NY, USA. p. 1234-1237.
7. Mei, T., R. Zhang, and X.-S. Hua, Internet multimedia advertising, in Proceedings of the 19th ACM international conference on Multimedia2011, ACM: Scottsdale, Arizona, USA. p. 627-628.
8. Quoc- Hung Nguyen, T.T.-H.T., Thi-Lan Le, Quang - Hoan Nguyen, Object detection and Classification for Advertising service Based on image content. 7th SEATUC 2013: p. 35.
9. Viola, P. and M.J. Jones, Robust Real-Time Face Detection. Int. J. Comput. Vision, 2004. **57**(2): p. 137-154.
10. Quattoni, A. and A.Torralba, Recognizing Indoor Scenes. In Proceeding of the International Conference on Computer Vision and Pattern Recognition, 2009: p. 1-8.
11. Oliva, A. and A. Torralba, Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope. Int. J. Comput. Vision, 2001. **42**(3): p. 145-175.
12. Everingham, M., et al., The Pascal Visual Object Classes (VOC) Challenge. Int. J. Comput. Vision, 2006. **88**(2): p. 303-338.
13. Everingham, M., et al., The PASCAL Visual Object Classes (VOC) Challenge. International Journal of Computer Vision, 2009.

Author's address: Nguyen Quoc Hung - Tel: (+84) 912251253
Email: Quoc-Hung.Nguyen@mica.edu.vn
Hanoi University of Science and Technology
No.1 Dai Co Viet Str., Ha Noi, Viet Nam