http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 12 (21): 6258-6262, 2013
ISSN 1812-5638 / DOI: 10.3923/itj.2013.6258.6262

Man-made Object Detection Based on Latent Dirichlet Allocation

[1]Xiaojun Xu and [2]Yingli Lv
[1]School of information and communication engineering,
Beijing University of Posts and Telecommunications, Beijing, China
[2]Hebei Institute of Architecture and Civil Engineering, Zhang Jiakou, P.R. China

Abstract: With the rapid development of multimedia technologies, man-made object detection is one of the important applications. An improved LDA approach was used to learn and recognize man-made and natural scene categories. It represent the image of a scene by a collection of local regions, denoted as codewords, each region is represented as part of a "theme". It learns the theme distributions as well as the codewords distribution over the themes. At last Support Vector Machine (SVM) classifier was used to image database for the man-made object detection. We report satisfactory categorization performances on a large set of image database.

Key words: Man-made object, LDA, SIFT, SVM

INTRODUCTION

Man-made object detection is one of the important applications of image processing. With the rapid progress of multimedia technologies, image retrieval and classification (Chen and Wang, 2012) from massive amounts of multimedia contents have been a hot topic among the multimedia research and applications (Datta *et al*., 2008; Xu and Ding, 2012). Due to the variability of image data, the relevant research aim at handling the huge amount of multimedia information intelligently and automatically for providing the functions like annotation and indexing by letting the machine comprehend the human concept. Novel methods like Content-Based Image Retrieval (CBIR) techniques have been proposed for image retrieval (Lew *et al*., 2006). Furthermore, various methods like SVM, k-means, decision tree and association rules-based approach have been studied for image classification by utilizing the image features like Color, Shape and Texture. In fact, the semantics of a whole image cannot be represented only by its low-level features. The methods previously of classification only use the low-level features of image. To get closer to human perception, approaches have been developed to map low-level features into high-level semantics. In general, image classification is involved with the following main issues: extraction of image features, organization and representation of image features, building of effective classifier, semantic structure modeling.

We introduce the SIFT (Scale-invariant feature transform) image feature in Section 2. Section 3 describes improved LDA (latent Dirichlet allocation) classifier for classification. Section 4 introduces the system of SVM classification use LDA model and k-mean algorithm. Section 5 illustrates the experimental results. We discuss in Section 6 our results.

THE FEATURE OF IMAGE

The feature of image retrieval system previously is color, textual and shape etc (Li and Zhang, 2012). But these features are non-context, they are pixel-level. It isn't suitable for denotation of image context (Lowe, 2004).

The SIFT is a local descriptor which was proposed by Lowe and used to match between different views of objects or scenes. The descriptor was invariance to image rotation and scaling and partially invariance to the different in illumination and viewpoint. This characteristic make itself has good performance than other local descriptors in the fields of object recognition and matching, robotic mapping and navigation, image stitching, 3D modeling, Gesture recognition, video tracking and match moving etc. Improving matching properties and reducing the computational complexity are two aspects of SIFT developing. Now, there was a distracted things is one image having more than two objects could be considered to belong to at least two different image categories. So it has an important work that filtering out those truly meaningful SFIT features in dependence on the category they belong to.

David G. Lowe proposed constructing a DOG (Difference Of Gaussian) scale-space to detect stable keypoint locations in scale space, The DOG space can be

Corresponding Author: Xiaojun Xu, School of information and communication engineering, Beijing University of Posts and Telecommunications, Beijing, China

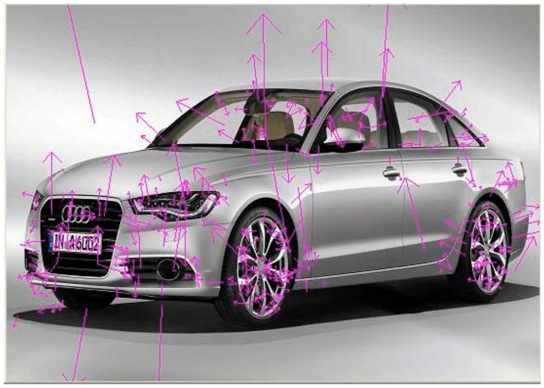

Fig. 1: An example of SIFT feature on image

produced from the difference between two convolution functions of variable-scales Gaussian with the original image. Subtracting Gaussian-smoothed images in eight directions can compute DOG descriptor. Comparing the eight neighbors in the current image and nine neighbors in the forward and backward DOG scale-space can get local minimum or maximum. It can get the extreme of the scale-space in local. The location and scale of the keypoints can be accurately localized by fitting a 3D quadratic function to the local sample points. For improve the stability of matching, we can discarded the keypoints with low contrast points. The SIFT descriptor is inherently invariant to rotation due to underlying histogram structure and gradient information. The experiments in Low's study shows the best results are achieved with 4×4 array of histograms with 8 orientation bins in each histogram. Hence, a 4×4×8 = 128 dimension feature is used for each keypoint. Figure 1 is an example of SIFT feature extraction of an image from car classification (Huang *et al*., 2009).

INSTRUCTION OF LDA

Different with HMM (Hidden Markov Model) (Sendi *et al*., 2012), LDA (latent Dirichlet allocation) is a generative model. It is an example of a topic model and was first presented as a graphical model for topic. It allows sets of observations to be explained by unobserved groups that explain why some parts of the data are similar. LDA is a hierarchical nonparametric Bayesian approach to create topic model in text processing. Each document θ_i from a topics is obeys a multinomial distribution Multi (θ) and each topic φk in the corpus obeys a multinomial distribution p(w|zi). It develops from pLSA and LSA model.

Plate notation can capture the dependencies among the many variables concisely. The boxes are "plates" representing replicates. Figure 2 is graphical model

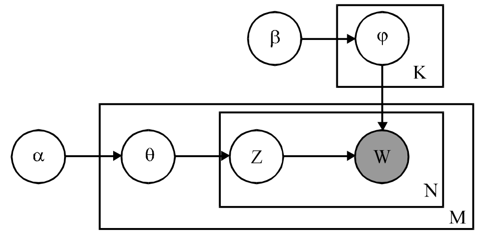

Fig. 2: Plate notation for smoothed LDA

representation of smoothed LDA. It illustrated the plate notation (Fei-Fei and Pietro, 2005). The outer plate represents documents, while the inner plate represents the repeated choice of topics and words within a document. M denotes the number of documents, N the number of words in a document. Thus:

α is the parameter of the Dirichlet prior.
β is the parameter of the Dirichlet prior.
θ_i is the topic distribution for document.
φ_k is the word distribution for topic,
z_{ij} is the topic for the jth word in document i,
w_{ij} is the specific word.

A k-dimensional Dirichlet random variable θ can take values in the (k-1)-simplex (a k-vector θ lies in the (k-1)-simplex if $\theta_i = 0$:

$$\sum_{i=1}^{k} \theta_i = 1$$

and has the following probability density on this simplex:

$$p(\theta \mid \alpha) = \frac{\Gamma(\sum_{i=1}^{k} \alpha_i)}{\prod_{i=1}^{k} \Gamma(\alpha_i)} \theta_1^{\alpha_1 - 1} \ldots \theta_k^{\alpha_k - 1} \quad (1)$$

where the parameter α is a k-vector with components $\alpha_i > 0$ and where $\Gamma(x)$ is the Gamma function.

Given the parameters α and β, the joint distribution of a topic mixture θ, a set of N topics z and a set of N words w is given by:

$$p(\theta, Z, W \mid \alpha, \beta) = p(\theta \mid \alpha) \prod_{n=1}^{N} p(z_n \mid \theta) p(w_n \mid z_n, \beta) \quad (2)$$

where, $p(z_n|\theta)$ is simply θ_i for the unique i such that $z_i^n = 1$. Integrating over θ and summing over z, we obtain the marginal distribution of a document:

$$p(W \mid \alpha, \beta) = \int p(\theta \mid \alpha) \left(\prod_{n=1}^{N} \sum_{z_n} p(z_n \mid \theta) p(w_n \mid z_n, \beta) \right) d\theta \quad (3)$$

Only the wij are the observable variables that are shown as words in the document and the other variables are all latent variables. The basic LDA model can be extended to a smoothed LDA. Where K denotes the number of topics in the model and φ is a K*V Markov matrix each row of which denotes the word distribution of a topic. V is the dimension of the vocabulary.

The generative process behind is that documents are represented as random mixtures over latent topics, where each topic is characterized by a distribution over words. LDA assumes the following generative process for each document in a corpus D:

- Choose θ_i~Dirichlet(α), where i{1, …, M}
- Choose φ_k~Dirichlet(β), where k{1,…,K}
- For each of the words w_{ij}, where j{1,…,Ni}
  - Choose a topic $z_{i,j}$~Multi (θ_i)
  - Choose a word $w_{i,j}$ ~Multi (φ_z (i,j))

When uses LDA to system, the inference of LDA must be obtained.

SYSTEM IMPLEAMENT

We use k-means and SVM for the system impleament. K-means clustering (Boser *et al*., 1992) is a method of cluster analysis (Wu and Yuan, 2012). It seperates n observations into k clusters in which each observation belongs to the cluster with the nearest mean. The term "k-means" was first used by MacQueen (1967). This results into a partitioning of the data space into Voronoi cells. The k-means algorithm is a NP-hard problem. Used efficient fast algorithms k-means converge fast to a local optimum. It is similar to the EM (expectation- maximization) algorithm for mixtures of Gaussian distributions via an iterative refinement approach. The expectation- maximization mechanism allows clusters to have different shapes. But k-means clustering tends to find clusters of comparable spatial extent. The k-means algorithm is shown like this. Given a d-dimensional real vector set of observations (x_1, x_2, …, x_n), k-means clustering algorithm partition the n observations into k sets (k = n) S = {S_1, S_2, …, S_k}. In the k sets, the sum of squares within the cluster is minimized,

$$\arg\min_S \sum_{i=1}^{k} \sum_{x_j \in S_i} \left\| x_j - u_i \right\|^2 \quad (4)$$

where, μ_i is the mean of points in S_i.

In the study, SVM algorithm used to classify the images that belong to diffierent classes. The original SVM algorithm was invented by Vladimir N. Vapnik and the

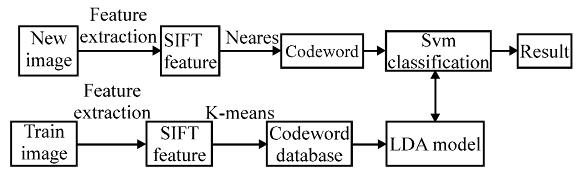

Fig. 3: Impleament of system

current standard description was proposed by Cortes and Vapnik (1995). The structured support vector machine is a machine learning algorithm that generalizes the Support Vector Machine classifier. The SVM classifier can use to binary classification, multiclassfication and regression. the structured SVM allows training of a classifier for general structured output labels (Ding and He, 2004).

We extraction the SIFT feature from train image set, select 100,000 SIFT descriptor for k-means, produce 3000 descriptor regard as the codeword. Use codeword as the word and use image as the topic come into being LDA model. When give a new image, extract the SIFT feature and produce codeword by nearest algorithm. Then send the codeword to SVM classifier get the result of classification. The process of the system implement is showed in Fig. 3.

EXPERIMENTS

In this section, we explain how to construct our experiments and show the performance of classification on man-made and nature object.

Experiments setup: ImageNet (Deng *et al*., 2009) is an image database organized according to the WordNet hierarchy, in which each node of the hierarchy is depicted by hundreds of images. Currently it has an average of over five hundred images per node. ImageNet will become a useful resource for researchers, educators, students and all of you who share our passion for pictures. Each meaningful concept in WordNet, possibly described by multiple words or word phrases, is called a "synonym set" or "synset". There are more than 100,000 synsets in WordNet, majority of them are nouns (80,000+). Images of each concept are quality-controlled and human-annotated. In its completion, it will offer tens of millions of cleanly sorted images for most of the concepts in the WordNet hierarchy.

We developed our experimental program uses SIFT-VC for SIFT feature extraction and matching, it uses gsl and opencv. The GibbsLDA++ is used to compute LDA model parameters.

In our experiments, we used 8 classifications. The mammal, mountain, bird, flower are belong to nature

Fig. 4: The match between car to flower

Table 1: Performance compare experimens

	Object class	Performance (%)
Nature classification	mammal	62.5
	Bird	75.4
	Flower	76.7
	Mountain	81.3
Man-made classification	Airplane	72.9
	Car	76.3
	Boat	78.7
	house	64.8

Table 2: Performance for increased number of train image

No. of training image	Mountain class (%)	Car class (%)
20	68.1	66.2
50	72.3	69.5
100	76.9	72.4
500	81.3	76.3
1000	83.7	78.2

classification and the airplane, car, boat and house are belong to man-made classification. The Fig. 4 shows the match between car to car and car to flower. It illustrated the same classification has more SIFT descriptor than the different.

Results of the experiments: In our study, we construct 8 datasets from for experiments. The Table 1 gives the result of our experiments. It shows our method has good performance about most image classification. From Table 1 we can look out the different of the every classification result is varying. We think is about the different characters of object.

When the numbers of train images is increased, the performance is enhanced. Table 2 shows this phenomenon. But the increased of numbers will cost more times for computation of model produced.

SUMMARY

We use SIFT descriptor for feature extraction and matching for image classification system. Although, SIFT descriptor have been used in a variety of studies. In this study, SIFT shows significant performance and stable on the classification evaluation. The proposed method can be appended to most existing classification systems and improve their accuracy and efficiency. The current scheme presented in this study does not use to multiclass. For next work, we will also concentrate on how to use to multiclass for good performance.

ACKNOWLEDGMENTS

This study was partially supported by National Natural Science Foundation of China under Grant No. 61005004, 61175011 and 61171193, the Next-Generation Broadband Wireless Mobile Communications Network Technology Key Project under Grant No. 2011ZX03002-005-01, the 111 project under Grant No. B08004 and Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry.

REFERENCES

Boser, B.E., I.M. Guyon and V.N. Vapnik, 1992. A training algorithm for optimal margin classifiers. Proceedings of the 5th Annual Workshop on Computational Learning Theory, July 27-29, 1992, Pittsburgh, Pennsylvania, USA., pp: 144-152.

Chen, X.J. and Q. Wang, 2012. Tour route recommendation begins with multimodal classification. J. Multimedia, 7: 21-30.

Cortes, C. and V. Vapnik, 1995. Support-vector networks. Machine Learn., 20: 273-297.

Datta, R., D. Joshi, J. Li and J.Z. Wang, 2008. Image retrieval: Ideas, influences and trends of the new age. ACM Comput. Surv., Vol. 40. 10.1145/1348246.1348248

Deng, J., W. Dong, R. Socher, L.J. Li, K. Li and L. Fei-Fei, 2009. ImageNet: A large-scale hierarchical image database. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June 20-25, 2009, Miami, FL., USA., pp: 248-255.

Ding, C. and X.F. He, 2004. K-Means clustering via principal component analysis. Proceedings of the 21st International Conference on Machine Learning, August 4-8, 2004, Banff, Alberta, Canada, pp: 225-232.

Fei-Fei, L. and P. Pietro, 2005. A bayesian hierarchical model for learning natural scene categories. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, July 20-26, 2005, Washington DC., USA., pp: 524-531.

Huang, S.L., C. Cai and Y. Zhang, 2009. Wood image retrieval using SIFT descriptor. Proceedings of the International Conference on Computational Intelligence and Software Engineering, December 11-13, 2009, Wuhan, China, pp: 1-4.

Lew, M.S., N. Sebe, C. Djerba and R. Jain, 2006. Content-based multimedia information retrieval: State of the art and challenges. ACM Trans. Multimedia Comput. Commun. Appl., 2: 1-19.

Li, J. and X.W. Zhang, 2012. Geometrically invariant watermarking scheme based on local feature points. J. Multimedia, 7: 231-238.

Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vision, 60: 91-110.

MacQueen, J., 1967. Some methods for classification and analysis of multivariate observations. Proc. 5th Berkeley Symp. Math. Statist. Prob., 1: 281-297.

Sendi, A.S., M. Dagenais, M. Jabbarifar and M. Couture, 2012. Real time intrusion prediction based on optimized alerts with hidden Markov model. J. Networks, 7: 311-322.

Wu, Q. and L.X. Yuan, 2012. A new filled function method for smooth clustering. J. Comput., 7: 491-498.

Xu, S. and J.F. Ding, 2012. Palmprint image processing and linear discriminant analysis method. J. Multimedia, 7: 269-276.