Proceedings of the IIEEJ Image Electronics
and Visual Computing Workshop 2012
Kuching, Malaysia, November 21-24, 2012

A BK SUBPRODUCT APPROACH FOR SCENE CLASSIFICATION

Ekta Vats; Chee Kau Lim and Chee Seng Chan

Center of Image and Signal Processing, Faculty of Comp. Sci. &. Info. Tech.
University of Malaya, 50603 Kuala Lumpur, Malaysia
{ekta.vats;limck}@siswa.um.edu.my; cs.chan@um.edu.my

ABSTRACT

Scene classification has received much attention in the recent past. How images can be efficiently classified becomes an important research topic. Most of the state-of-the-art solutions assumed that scene classes are mutually exclusive. But, different human beings have different perceptions, and different people tend to respond inconsistently even to the same scene images. Hence, a scene image may belong to multiple classes. In this paper, we propose a BK subproduct approach to address this issue. The advantages are two-folds. Firstly, our approach is able to model the non-mutually exclusive data. Secondly, our classification result is not binary. Instead, it classifies each scene image as a combination of different classes using membership function. Empirical results using a standard dataset had shown the effectiveness and robustness of the proposed method.

1. INTRODUCTION

Scene classification is one of the most important issues in computer vision community; and has been studied extensively in the recent past [16]. However, in order to understand and interpret a natural scene is still a challenging task because of the variability, ambiguity and wide range of illumination and scale conditions that may apply. Several image features such as shape, colour and texture have been widely used along with supervised learning methods such as K-nearest neighbour (KNN), support vector machines (SVM) to classify images into several classes (coast, open country, street, etc.).

A scene is composed of several objects often organized in an unpredictable layout [7]. Oliva and Torralba [1] proposed a set of perceptual dimensions (naturalness, openness, roughness, expansion, ruggedness) to represent the dominant spatial structure of a scene. Then, a SVM classifier with Gaussian kernel is employed to classify the scene classes. Fei-Fei and Perona [2] proposed a bayesian hierachical model to learn natural scene categories, which is an extension of latent dirichlet allocation (LDA). This work inspired Bosch et al. [3] to propose probabilistic latent semantic analysis (pLSA) incorporated with KNN to classify scenes. Graphical models were used for scene classification by Kumar et al. [4] for the detection and localization of man-made features in a scene. Vogel and Schiele [5], [6] used occurring frequency of different concepts (water, rocks, etc.) as the intermediate features for scene classification. Recently, Parikh and Grauman [8] proposed a relative attribute which can make a computer to mimic how humans classify an object based on the relative attributes. Boutell et al [9] proposed a cross training on multi-label approach for scene modelling. Then, a SVM with Gaussian kernel is employed to perform classification.

(a) Open Country (b) Ambiguous (c) Coast

Fig. 1 Example of ambiguous scene between open country and coast

Although significant results have been achieved, errors in classification often occur when there is an overlap between classes in the selected feature space. The reason being that they assumed the scene classes to be mutually exclusive, which means that most systems are exemplar-based, learning patterns from a training set and searching for the images similar to it, where low level features are used as a parameter to define similarity.

In this paper, we introduce a BK subproduct framework to address this problem. In contrast to the above, our paper uses a BK subproduct approach for scene classification. BK sub-triangle product or BK subproduct in short, is able to trace relations within sets which may not directly relate if both sets show relations with a third set. In previous, BK subproduct has proved its capability in developing inference engines of medical expert systems [11][12], path finding of autonomous underwater vehicles [13], land evaluation [14] and etc.

[15] verified that BK subproduct is an effective and efficient model in fuzzy reasoning.

Our approach adopted the SIFT to represent the feature space, and then K-means for clustering. Lastly, human annotation is used for scene classification. The advantages are two-folds. Firstly, our approach is able to model the non-mutually exclusive data. It is possible for an image to belong to multiple classes. For example, in Fig. 1 (b) it is unclear that it is an open country scene or a coast scene. Different people will respond inconsistently in providing the presence or absence of the local features for this image. Secondly, our classification result is not binary. Instead, it classifies each scene image as a combination of different classes using membership function.

The rest of the paper is organized as follows. Section 2 reviews the past work related to scene classification. Section 3 discusses the BK subproduct approach in detail. Section 4 presents the empirical results using a standard dataset, and we conclude with suggestions of future work in section 5.

2. METHODOLOGY

2.1. BK Subproduct Revisit

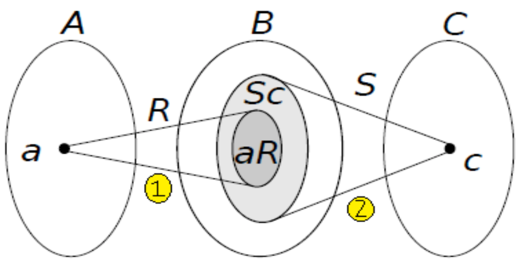

Fig. 2 Relationship between 2 sets which are not directly related using BK subproduct.

We start the discussion with a short revision on the BK subproduct in crisp environment. Assume that there are set A = $\{a_i \mid i=1,...,m\}$ and set B = $\{b_j \mid j=1,...,n\}$. R is defined as a relation from A to B such that $R \subseteq A \times B$. If there is a set C = $\{c_k \mid k=1,...,p\}$ such that A has no direct relation with C, but B is in relation S with C, we can trace the indirect relation of A and C with BK subproduct:

$$R \triangleleft S = \{(a,c) \mid (a,c) \in A \times C \text{ and } aR \subseteq Sc\} \qquad (1)$$

In (1), *aR* is the image of a in set *B* under relation *R* and *Sc* is the image of c under converse relation of *S*, i.e. S^c. One can see that BK subproduct gives all (a,c) couples such that the image of a under relation R in B is among the subset of c under S^c in *B* (Fig. 2). With this composition of relations, one can trace the relation between an object, *A* and target, *C* if a set with common features, *B* appears in the middle. Therefore, when a particular $a_1 \in A$ is concerned, using BK subproduct is helpful in retrieving a set of $c \in C$ where a_1R is subset of *Sc* for all *c*. This make BK subproduct become a tool in retrieving indirect relations, as well as inference:

$$a_1(R \triangleleft S) = \{c \mid c \in C \text{ and } a_1R \subseteq Sc\} \qquad (2)$$

In a fuzzy environment, BK subproduct uses a fuzzy subsethood measurement. As proposed in [10], in the same universe X = $\{x_i \mid i=1,...,n\}$, the degree of subsethood of a fuzzy set A in B, $\pi(A \subseteq B)$ can be expressed as :

$$\pi(A \subseteq B) = \bigwedge_{x \in U} (\mu_A(x) \rightarrow \mu_B(x)) \qquad (3)$$

where:

$\wedge$ can be defined as infimum operator in harsh criterion, or arithmetic mean in mean criterion;

$\mu_A(x)$ and $\mu_B(x)$ are membership function of x in A and B respectively;

$\rightarrow$ is fuzzy implication operator.

With this fuzzy subsethood measurement definition, it is easy to rewrite the expression of BK subproduct in fuzzy environment. For the case of 2 elements $a \in A$ and $c \in C$, the fuzzy relation from a to c is given as:

$$R \triangleleft S\,(a,c) = \bigwedge_{b \in B} (R_{ab} \rightarrow S_{bc}) \qquad (4)$$

2.2. Scene Classification with BK Subproduct

In our proposed algorithm, set *A* is a set of images whereas *C* is a set of scene classes. *A* has no direct relation with *C*, but set *B* which denotes a set of features (codebook) after clustering (eg. water, sky, tree, mountain, etc.). Set *B* is in relation with both *A* and *C*. Our aim is to trace the indirect relationship between *A* and *C* using BK subproduct.

For each image $a \in A$, several local patches are extracted and converted into 128-dimensional numerical vectors ($V_1, V_2, \ldots, V_{128}$) using SIFT. With this, each *a* is represented by a unique set of vectors $a' \in A'$ (Fig. 3). Therefore, instead of using relation $R \subseteq A \times B$ in (4), we replace R with R' where $R' \subseteq A' \times B$.

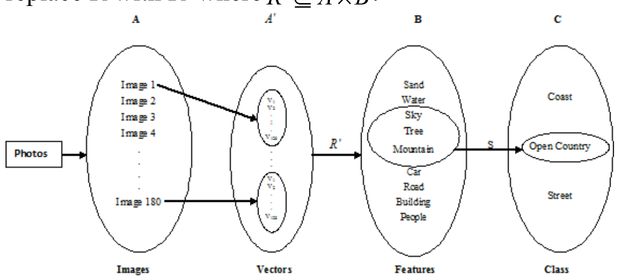

Fig. 3 An example of BK subproduct approach towards scene classification

While performing classification, vectors that representing patches are converted into codewords using

K-means. The distances between these results and clusters that represent objects in set *B* are calculated and normalized to obtain membership functions for relations R'. In the next step, scene classification is done using K-nearest neighbor. Set *C* denotes the set of scene classes (coast, open country and street). Support Vector Machines (SVM) is employed for testing and confusion matrix is produced as output. Lastly, using human annotation the membership degree of an object belonging to different scene classes is defined, as shown in Table 1.

Table 1. An example of membership function

Objects	Coast	Open Country	Street
Sand	0.55	0.35	0.00
Water	1.00	0.10	0.02
Sky	1.00	1.00	0.87
Tree	0.25	0.72	0.17
Mountain	0.45	0.60	0.02
Vehicle	0.12	0.00	0.80
Road	0.00	0.02	0.95
Building	0.05	0.05	1.00
People	0.15	0.10	0.30

3. EXPERIMENTAL RESULTS

In order to test the effectiveness and robustness of our proposed framework, we tested the proposed method using the Outdoor Scene Recognition (OSR) Dataset [1]. A total of 3 classes of the scenes are used throughout the experiments which are coast, open country and street. The examples of those scenes are shown in Fig. 4. Each scene class has 60 images from original image database, there are 180 images totally. For the 60 images in each scene class, the first 40 images are for training, the last 20 images are for testing. Thus, the total number of training images is 120 and the total number of testing images is 60. Furthermore, all observed images are gray scaled.

All the experiments were conducted using MATLAB 2010B, 64bit. As discussed in Section 3, given an image, several local patches are extracted and converted into 128-dimensional numerical vectors SIFT descriptors. Fig. 5 shows the output after execution of SIFT on an image from coast scene class. Then, the vector represented patches are converted into codewords (set of objects) using K-means. The membership function is defined by the distance measure.

3.1. Scene Classification

In these experiments, we used K-nearest neighbour classifier to classify coast, open country and street scenes and SVM algorithm is employed for testing. SVM implementation is based on the LIBSVM which is a MATLAB toolbox. Confusion matrix is produced as output after this step shown in Fig. 6. The experiments were conducted 20 times totally based on the same patches which are gridded by SIFT. As a result of there are some similar elements in different image categories and the size of the data sets, the overall accuracy is found to be 85% (taking the mean of all the results).

(a) Examples of coast scene

(b) Examples of open country scene

(c) Examples of street scene

Fig. 4. Examples of the scenes from three classes

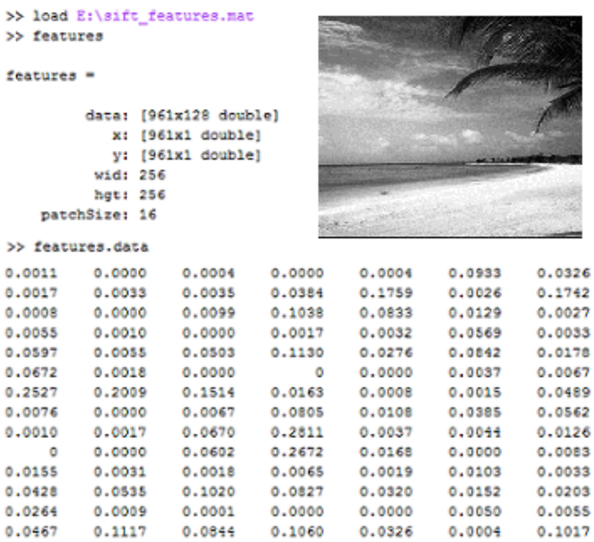

Fig. 5. Output of SIFT descriptor from coast scene

3.2. Human Annotation

Secondly, we used human annotation for scene classification. Different people tend to respond inconsistently even given a same scene image. On conducting a survey on 200 people via FACEBOOK, the most popular social networking site, about how different people can classify objects into various scene

Fig. 6. Confusion matrix after executing SVM

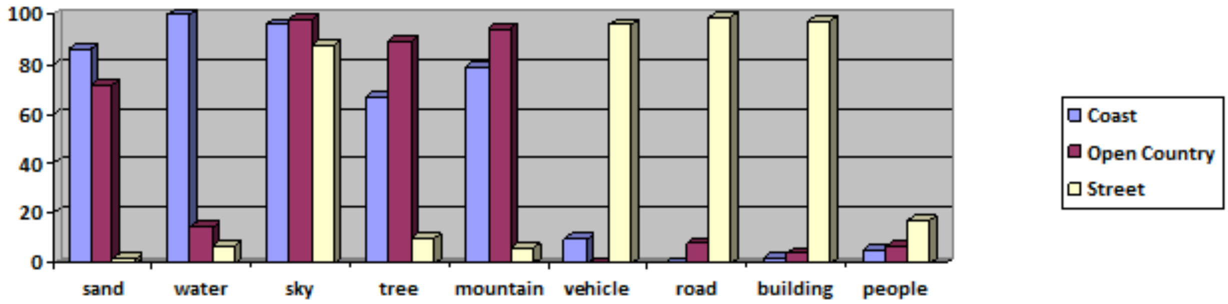

Fig. 7. Bar chart representing the result after FACEBOOK surveys

classes (coast, open country and street), the following results were obtained as shown in Fig. 7, where X-axis denotes the objects and Y-axis denotes the percentage of people.

The confusion matrix (Fig. 6) gives us the relationship between two different classes, whereas the bar chart (Fig. 7) provides us a clearer picture of the relationship. That is, class-to-class relationship is given by confusion matrix, whereas bar chart gives object-to-class relationship.

4. CONCLUDING REMARKS

In this paper, we show the implementation of BK subproduct and the usage of it in natural scene classification. The main achievement of our approach is that it is able to model the non-mutually exclusive data and our classification result is not binary. Instead, it classifies each scene image as a combination of different classes using membership function. Also, a clear picture of object-to-class relationship is presented using human annotation. The experiments show positive results, however, there is more research to be done to fine tune the proposed framework. In the future, we intent to solve the problem using a large dataset of both indoor and outdoor scene images for better classification results.

5. REFERENCES

[1] A. Oliva and A. Torralba. Modelling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, vol. 42, no. 3, pp. 145-177, 2001.

[2] L. Fei-Fei and P. Perona. A Bayesian hierarchical model for learning natural scene categories. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2, no. 15, pp. 524-531, 2005

[3] A. Bosch, A. Zisserman, and X. Mu. Scene classification via psla. *European Conference on Computer Vision (ECCV),* vol. 3954, pp. 517-530, 2006.

[4] S. Kumar and M. Hebert. Discriminative random fields: A discriminative framework for contextual interaction in classification. In *Proceedings of the Ninth IEEE International Conference on Computer Vision (ICCV)*, vol. 2, pp. 1150-1157, 2003.

[5] J. Vogel and B. Schiele. Natural scene retrieval based on a semantic modeling step. *Image and video Retrieval*, vol. 3115, pp. 1950-1950., 2004.

[6] J. Vogel and B. Schiele. Semantic modeling of natural scenes for content-based image retrieval. *International Journal of Computer Vision*, vol. 72, no. 2, pp. 133-157, 2007.

[7] P. Quelhas, F. Monay, J.-M. Odobez, D. Gatica-Perez, T. Tuytelaars, and L. Van Gool. Modeling scenes with local descriptors and latent aspects. In *Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV)*, vol. 1, pp. 883-890, 2005.

[8] D. Parikh and K. Grauman. Relative attributes. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 503-510, 2011.

[9] M. Boutell, J. Luo, X. Shen, and C. Brown. Learning multi-label scene classification. *Pattern Recognition*, vol. 37, no. 9, pp. 1757-1771, 2004.

[10] L. J. Kohout and W. Bandler. Fuzzy Power Sets And Fuzzy Implication Operators. *Fuzzy Sets and Systems*, vol. 4, pp. 13–30, 1980.

[11] K. M. Yew and L. J. Kohout, Interval-based reasoning in medical diagnosis. In *Proceedings Intelligent Information Systems* (IIS), pp. 32–36, 1997.

[12] C.K. Lim and C.S. Chan. Logical Connectives and Operativeness of BK Sub-triangle Product in Fuzzy Inferencing. *International Journal of Fuzzy Systems*, vol. 13, no. 4, pp. 4-12, 2011.

[13] L.-d. Bui and Y.-g. Kim, An Obstacle-avoidance Technique for Autonomous Underwater Vehicles Based On BK-products of Fuzzy Relation. *Fuzzy Sets and Systems*, vol. 157, pp. 560-577, 2006.

[14] R. Groenemans, V. E. Ranst, and E. Kerre, Fuzzy relational calculus in land evaluation. *Geoderma*, vol. 77, pp. 283–298, 1997.

[15] M. Stepnicka and B. Jayaram, On the Suitability of the Bandler Kohout Subproduct as an Inference Mechanism. *IEEE Transactions on Fuzzy Systems*, vol. 18, no. 2, pp. 285-298, 2010.

[16] C.H. Lim and C.S. Chan. A fuzzy qualitative approach for scene classification. In *Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pp. 1-8, 2012.