2011 3rd International Conference on Information and Financial Engineering
IPEDR vol.12 (2011) © (2011) IACSIT Press, Singapore

Integrating Qualitative Features Selection for Semantic Image Classification with Support Vector Machine

Nutchanun Chinpanthana
Dhurakij Pundit University, Faculty of Information Technology,
110/1-4 Prachachuen rd. Laksi, Bangkok 10210, Thailand

Abstract. Semantic image classification is an active problem in multimedia image retrieval. Many researchers have attempted to improve semantic models by using high-level concept based on keyword annotation. The retrieval process of such approaches is done by keyword searching. The model is rather rudimentary and it does not specific enough for representing the meaning of images. In this paper, we present a technique of the semantic image classification by using integrating qualitative feature selection. The structural skeleton is used to extract the object components and image meaning. The feature selection methods are introduced to select the essential features from existing features. The experimental results indicate that our proposed approach offers significant performance improvements in the interpretation of semantic image classification, compare with other features, with the maximum of 93.8%

Keywords: semantic image classification, support vector machine, Bayesian probability, feature extraction, feature selection, chi-squared, gain ratio, oneR.

1. Introduction

Semantic classification has been playing an important role in image processing research, since the digital camera technology has been extremely successful developed over the last decade. People can take photos quickly and easily, so the number of digital images has increased significantly [1]. When people need to find a desired image, they often spend time searching the images in a large database. Researchers attempt to make various methods for semantic classification. However, the classification results are not directly to the semantic images. For this reason, semantic image classification has become an important field of research.

In this section, we briefly discuss some significant examples and categorize them into two main research directions: low-level feature based approaches and high-level semantic based approaches. A majority of prior researches has focused on low-level features [1] such as color, texture, and shape, which can be extracted using image processing algorithms. The regions are recognized into the predefined objects with multiple features [2]. This causes a limitation, since most images have several complicated regions, and elements are often composed of various parts. In many cases, the results are obtained from the similarity objects that are inadequate due to the whole-semantic images. Next, researches have attempting to discriminate the recognition low-level feature extraction process by labelling the objects with keywords [3]. There are ongoing in a high-level representation. Researchers have begun to combine wording and visual feature maps into high-level concepts for actual semantic images. Some researchers have addressed the issues of learning of term similarity matrix and word grouping for intelligent query expansion [4]-[6]. They construct more meaningful concept clusters of co-occurring keywords technique based on dictionary [6]. But the consistency of keyword techniques depends on their contents. The latter group [7][8] applied the multiple layers to classify the higher semantic. Although results are relevant to the set of keywords, it does not represent whole image meaning. The set of keyword in the image is not related to the semantic of images directly.

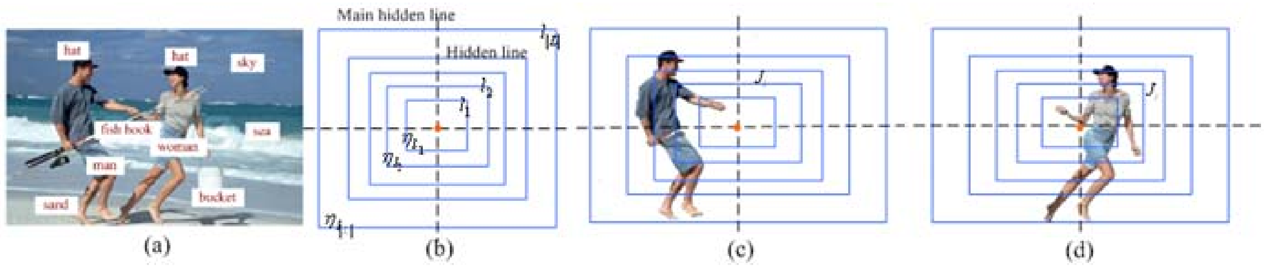

Fig. 1: (a) sample image of object annotation (b) the structural skeleton framework (c)-(d) mapping the main objects into (b)

In this paper, we discuss a new approach to classify the images based on their semantic. We focus on the concept of extracting the essential features. The main idea is to observe the extracting visual features from structural skeleton framework as these features carry important meaning of the images. The structural skeleton framework is representing the pattern of object perception from psychology study [9]. The Bayesian and Support vector machine is then used in classification process. The outline of the paper is as follows. The next section will give the extraction of the visual features with purposed method. Subsequently, section 3 shows the best features by comparing 3 methods: chi-squared, oneR and gain ratio. Section 4 overviews the mostly used classification methods for the next task. Experiment results are presented in section 5. Finally, section 6 concludes this work

2. Proposed Method

The proposed approach takes the extraction of qualitative features with the structural skeleton framework. The extraction process is divided into 2 steps as follow:

2.1. Extracting the image contents

Each object is annotated to an element of the keyword archive. The objects have identified that are corresponding between each image regions and annotated words. The keyword archive is based on a collection of groups of keywords which have already been used for testing automated image annotation algorithms. Let the collection of annotated images denoted by T, and the size of the collection denoted by $|T|$. Each annotated image $J_i \in T$ is represented by $\beta_{i,j} = \sum_{k,z}^{|\beta_{i,j}|} pixel(x_k, y_z)$. Let $\beta_{i,j}$ be the size's object of jth that appears in the ith image; $\vartheta_{i,j}$ is a keyword that indicates the jth word appears in the ith image. Feature vector of annotating image objects with words can be expressed as follows: its image objects and annotated words, i.e., $J_i = \{\vec{\beta}_i; \vec{\vartheta}_i\} = \{\beta_{i,1}, \beta_{i,2}, \ldots, \beta_{i,|\beta|}; \vartheta_{i,1}, \vartheta_{i,2}, \ldots, \vartheta_{i,|\vartheta|}\}$. Each object has to be mapped into one of the annotated words. Figure 1 (a) shows an example, where ϑ_j includes man, woman, sea, hat, sand, sky, and bucket. After that objects are mapped for fitting into structure skeleton in next section.

2.2. Mapping the contents into structural skeleton framework

In the psychology of the perception indicates that a dominant object is primary content as essence of the semantic image and should be given the priority concern [9]. Therefore, we focus on the position of objects in the images. The object's position is an essential content for extracting the semantic images. The structural skeleton framework is the pattern of object perception from psychology study [9] as shown in Figure 1 (b). Let $l_1 \subset l_2 \ldots \subset l_{|L|}$ be hidden lines which are in the form of nested rectangles corresponding of image edge. Each hidden line is attached with a predefined weight η_l. The ϑ_i in the central position has a higher η_l than other ϑ_j at the corner of the image as shown in Figure 1(c) and (d). The object woman has a higher weight than man. Thus, we calculate the total score of β_i as the following function: $\rho_j = \sum_{\beta_{i,j} \in l, j=1}^{|\vartheta|,|L|} \eta_{l_j}$ where ρ_j is a position in each object region.

3. Finding the qualitative features

This section describes the techniques of features selection. Feature selection is necessary to find the best combination of features that are suitable in constructing a classifier with high performance. However, with a large number of features, it is intractable to find the best combination from all possible combinations of features due to the complexity of the problem. In general, selection feature has two fundamentally different approaches. One is to make an independent assessment based on general characteristics of the data; called the filter method The attribute set is filtered to produce the most promising subset before learning commences. Some typical methods of this approach are chi-squared [10], oneR [11], and gain ratio [12]. In the chi-squared, χ^2 method [10], the χ^2 value of an attribute is defined as:, $\chi^2 = \sum_{i=1}^{m}\sum_{j=1}^{k}\frac{\left(A_{ij}-X_{ij}\right)^2}{X_{ij}}$, where $X_{ij} = \frac{R_i \times C_j}{N}$, where m is the number of intervals, k the number of classes, A_{ij} the number of samples in the ith interval, jth class, R_i the number of objects in the ith interval, C_j the number of objects in the jth class, N the total number of objects. Therefore $X_{ij}$the expected frequency of A_{ij}. The oneR [11] known as a one-level decision tree, selects only one feature that generates the lowest error rate, as opposed to the entropy-based measures used in decision tree. It treats all numerically-valued attributes as continuous and uses a straightforward method to divide the range of values into several disjoint intervals. Applied in the general decision-tree induction, the gain ratio [12] is the expected reduction in entropy reduction from partitioning the dataset objects according to a particular feature. For each attribute, the considered data is split based on the formula of $Split(S,A) = -\sum_{i=1}^{c}\frac{|s_i|}{|S|}log_2\frac{|s_i|}{|S|}$, where each S_i is a subset of objects generated by partitioning S with the c-valued attribute A. The gain ratio is defined by $GR(S,A) = \frac{IG(S,A)}{Split(S,A)}$, where $IG(S,A)$ is the information gain [12].

4. Semantic Classification

To compare our proposed method, we used Bayesian probability and support vector machine classification as following:

4.1. Bayesian Probability Theory

Bayesian theory [13] is the basis of statistical classification methods that provides the fundamental probability model for classification procedures. It requires all assumptions be explicitly built into models which are constructed by using the training data to estimate the probability of each class. Consider a general N -group classification problem in which each object has an associated attribute vector x of d dimensions. Let c denote the member that takes a value of c_i if an object is belong to group i. Define $p(c_i)$ as the prior probability of class i and $f(X|c_i)$ as the probability density function. According to the Bayes rule: $p(c_i|X) = \frac{f(X|c_i)p(c_i)}{f(X)}$,where $p(c_i|X)$ is posterior probability of class c_i and $f(X)$ is the probability density function: $f(X) = \sum_{i=1}^{M} f\left(X\middle|c_j\right)p(c_i)$.The Bayes classifier learns the conditional probability of each attribute x_i (i = 1, 2, …, N) of X given the class label c_i, denoted as: $p(c_i|X) = \prod_{i=1}^{N} p(x_i|c_i)$. Therefore, the image classification can be stated as given a set of observed features $x_i$from an image X, classify X into one of the classes c_i.

4.2. Support Vector Machine

Support Vector Machine (SVM) [13] is another simple classifier. SVM is designed for binary classification. That is, to separate a set of training vectors which belong to two different classes. Let $(x_i, y_i)_{1\le i\le N}$ be a set of training examples, each example $x_i \in \mathbb{R}^d$, d being the dimensional of input feature space, belongs to a class labeled by $y_i \in \{-1, 1\}$. During the SVM model generation, the input vectors, are mapped into a new higher dimensional feature space. Then, an optimal separating hyperplane in the new feature space is constructed by a kernel function which products between input vectors x and y, $K(x, y)$. Two most used kernel functions are Polynomial and Gaussian Radial Basis Function (RBF) kernel functions which are: $K_{poly}\left(x_i, y_j\right) = (x_i \cdot y_j + 1)^p$, where p is the degree of polynomial and $K_{gaussian}\left(x_i, y_j\right) = e^{\frac{\|x_i - y_j\|}{2\sigma}}$, σ is Gaussian sigma respectively. All vectors lying on one side of the hyperplane are labelled as -1, and all vectors lying on another side are labeled as +1. The training instances that lie closest to the hyperplane in the transformed space are called support vectors. The number of these support vectors is usually small compared to the size of the training set and they determine the margin of the hyperplane, and thus the decision surface.

5. Experimental Results and Discussion

TABLE I. SUMMARY OF EACH FEATURE (J_i) WITH SELECTION METHODS

Feature	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9	f_{10}	f_{11}	f_{12}
ChiSquared	69.38	119.82	39.30	156.31	162.19	13.37	455.54	213.99	59.56	91.51	90.36	32.25
OneR	39.68	31.66	37.27	35.67	38.48	35.27	49.30	45.69	32.67	43.09	39.68	30.66
InfoGain (10^2)	9.67	14.92	5.29	23.03	23.54	1.97	65.7	28.01	8.95	12.6	11.99	4.29

TABLE II. COMPARISON CLASSIFICATION RESULS (%) OF EACH FEATURE (J_i)

No. of feature	1	2	3	4	5	6	7	8	9	10	11	12
Bayesian	43.09	52.71	61.72	54.11	54.11	55.51	55.71	57.11	63.93	64.93	64.93	65.73
SVM	50.50	62.93	55.91	59.92	59.92	59.92	74.75	65.93	67.54	97.94	72.14	70.74

5.1. Data Sets and Evaluation Methods

In our experiments, we selected the probe images from two data sources. The dataset contains approximately 1,500 images, where 750 images of natural scene selected from the Corel Image Database [14] and the Corbis Image Database [15]. We focus on five categories: *beach*, *city*, *factory*, *indoor* and *landscape*. The database was setup to cover a variety of main image contents that included background, and foreground. The background was scoped into two semantics scene types: indoor and outdoor. The foreground included outstanding peripheral objects that were semantically assigned with labels. Example street, sky, sand, tree, kid, garden etc. In order to make labeling consistency, we used a vocabulary set taken from the NIST TRECVID 2003 development set [16]. The set of labels was selected from a subset of NIST lexicon.

In this work, we performed classification using the Bayesian and support vector machine (SVM). We used precision (Pr) and recall (Re) for evaluation. In order to get a single measure of effectiveness, we employed F1 that is a combination of precision and recall. All experiments were conducted using 5-fold cross validation [14]. The images for each cross-validation round were randomly split the data set into 80% for training and the remaining 20 % for testing.

5.2. Semantic Classification with Selected Features

In this section, we evaluate the semantic classification results by comparing with three selection features. The methods are chi-squared, gain ratio, and oneR. We test our method with two classifiers: Bayesian probability and SVM. We used the filter-typed feature selection function and classification in Weka. In the preliminary stage, we make an experiment to find the best features from a single type of features. As stated above, this work focuses on name's object of features. It is essential content to classify. We compare the three methods described in Section 3. The feature selection results of each feature are shown in Table I. We can see that f_7 from J_i has the highest value: Chi-squared provides 455.54, oneR provides 49.30, and infogain provides 65.7. The highest value reflects the most discriminative features. The essential features are the same therefore we classify these feature order as shown in Table II. The classification of SVM can achieve the better result of 74.7% accuracy when the number of feature is 7. Whereas the Bayesian probability gains only 55.7%. Therefore, we combined the 7 features (ϑ^*): $f_7, f_8, f_5, f_4, f_2, f_{10}, f_{11}$ to consider in the next experiment. We examine four different feature settings, including ϑ^* (set 1), ϑ^* and position of object (ρ) (set 2), ϑ^* and size of object (β) (set 3) and ϑ^*, β and ρ (set 4) as shown in Table III and VI. As general observation, considering all feature types allow us to obtain better accuracy than the cases of single feature type. Comparing the results, we can observe that the accuracy of set 4 in SVM provide 93.80%. Whereas the set 2 and 3 with SVM obtains only 73.55% and 87.95%, Bayesian provides 66.53% and 83.20%. However, the accuracy of ϑ^* in SVM obtains up to 74.7% but Bayesian gains only 55.7%.

6. Conclusion

In this paper, we have presents a new technique for semantic classification. The major components are extracted from qualitative features by using the structural skeleton framework. These features including with name, size and position. The size and position of object are extracting from structural skeleton framework. Then, semantic images are classified into the high-level semantics by using supporting vector machine. Our results indicate that the proposed approach offers good interpretation of semantic images. This technique is an important theory for various applications in digital library and other annotation systems.

TABLE III. CLASSIFICATION RESULTS OF COMBINING FEATURE GROUPS WITH BAYESIAN

Classes	ϑ^*			ϑ^*, ρ			ϑ^*, β			ϑ^*, β, ρ		
	Pr	**Re**	**F1**	**Pr**	**Re**	**F1**	**Pr**	**Re**	**F1**	**Pr**	**Re**	**F1**
Beach	54.1	53.0	53.5	67.5	79.8	73.1	94.1	96.0	95	95.0	96.0	95.5
City	40.8	60.0	48.6	58.5	72.0	64.6	82.2	83.0	82.6	81.2	82.0	81.6
Factory	71.6	58.0	64.1	78.9	60.0	68.2	74.3	78.0	76.1	79.0	83.0	81.0
Indoor	73.1	68.0	70.5	71.3	72.0	71.6	73.9	85.0	79.1	80.4	86.0	83.1
Landscape	48.1	39.0	43.1	59.8	49.0	53.8	96.1	74.0	83.6	91.9	79.0	84.9
Average	55.60			66.53			83.20			85.20		

TABLE IV. CLASSIFICATION RESULTS OF COMBINING FEATURE GROUPS WITH SVM

Classes	ϑ^*			ϑ^*, ρ			ϑ^*, β			ϑ^*, β, ρ		
	Pr	**Re**	**F1**	**Pr**	**Re**	**F1**	**Pr**	**Re**	**F1**	**Pr**	**Re**	**F1**
Beach	72.3	73.0	72.6	71.8	89.9	79.8	96.0	96.0	96.0	97.1	99.0	98.0
City	74.2	89.0	80.9	62.4	88.0	73.0	86.4	95.0	90.5	92.5	99.0	95.6
Factory	78.4	87.0	82.5	78.2	79.0	78.6	81.1	78.6	79.8	92.1	94.0	95.7
Indoor	80.7	67.0	73.2	82.3	65.0	72.6	85.3	81.0	83.1	93.8	90.0	91.8
Landscape	67.1	57.0	61.6	85.2	46.0	59.7	90.8	89.0	89.9	93.5	87.0	90.2
Average	74.60			73.55			87.95			93.80		

7. References

[1] P.Sankara Rao.et. al., "An Approach for CBIR System Through Multi Layer Neural. Network," International Journal of Engineering Science and Technology, vol. 2(4), 2010, pp. 559-563.

[2] Julia Vogel and Bernt Schiele, "Semantic Scene Modeling and Retrieval for Content-Based Image Retrieval," *Int. J. of Computer Vision*. vol. 72, no. 2, pp. 133-157, April 2007.

[3] Benitez, A.B., and S.-F. Chang, "Semantic Knowledge Construction From Annotated Image Collections", *International Conference On Multimedia & Expo (*ICME-2002), Lausanne, Switzerland, Aug 26-29, 2002.

[4] Lawrence Reeve and Hyoil Han, "Survey of semantic annotation platforms," *Proceedings of the 2005 ACM symposium on Applied computing*, Santa Fe, New Mexico, 2005.

[5] Dawei Ding, et al.: A Semantic Model. for Flash Retrieval Using Co-occurrence Analysis. Proc. ACM Int. Conf. on Multimedia, 2003.

[6] Zhao T, et al., "Visual Keyword Image Synergetic Neural Network for Web-Based Image Search", *Journal of Real-Time Systems*, Vol.21, Issue 1/2 , 2001.

[7] Seungji Yang, et al., "Semantic Home Photo Categorization," IEEE Transactions on In Circuits and Systems for Video Technology, Vol.17, No.3, 2007.

[8] Nicolas Tsapatsoulis, "A Hierarchical Classification Scheme for Semantic Image Annotation," First International Conference on Advances in Multimedia, 2009.

[9] R. Arnheim, *Art and Visual Perception A Psychology of the Creative Eye*, University of California Press Ltd., 1974.

[10] H. Liu and R. Setiono, "Chi2: Feature selection and discretization of numeric attributes," In Proc. of the 7th IEEE Int.Conf. on Tools with Artificial Intelligence, 1995.

[11] R.C. Holte. Very simple classification rules perform well on most commonly used datasets. Machine Learning, Vol. 11, No. 1, pp. 63–90. 1993.
[12] E. Hunt,etal.,“Stepwise Feature Selection by Cross Validation for EEG-based Brain Computer Interface,” Proc. of IEEE World Congress on Computational Intelligence, 2006.
[13] Richard O. Duda, Peter E. Hart and David G. Stork, Pattern classification, New York, Wiley, 2nd, 2001.
[14] The Corel Corporation home page: http://www.corel.com/
[15] The Corbis Corporation home page: http://pro.corbis.com/
[16] C.-Y. Lin, et al., “Video collaborative annotation forum: Establishing ground-truth labels on large multimedia datasets ,” TRECVID Workshop, USA, 2003.