Combining Multiple Spatial Hidden Markov Models in Image Semantic Classification and Annotation

Lihua Wang, Horace H. S. Ip
Image Computing Group, Department of Computer Science
Center for Innovative Applications of Internet and Multimedia Technologies(AIMtech Centre)
City University of Hong Kong, Hong Kong
{wlihua2@student. , cship@}cityu.edu.hk

Abstract

The spatial-hidden Markov model (SHMM) [11] is a two dimensional generalization of the traditional hidden Markov model (HMM), with the capability of block-based semantic annotation as well as classification of images. In this paper, we conduct a sensitivity analysis of SHMM in semantic classification with respect to different block sizes and from this analysis, we propose a novel multi-scales SHMM that combines multiple SHMMs, each classifying the image on a different scale. By regarding each SHMM as distinct classifiers, classifier combination algorithm can be applied to integrate the outputs of the respective SHMMs to improve image classification accuracy. Experiment results demonstrate that the multi-scale SHMM consistently outperforms single SHMM in image semantic classifications. The proposed approach can be extended to other block-based image classification algorithms.

1. Introduction

Research in content-based retrieval of images over the past decades has led to the now widely recognized challenge of bridging the semantic gap between the low-level visual features and the high-level semantic of an image. Since last decade, numerous approaches have been designed primarily to provide annotations of the entire image as a whole [2]. However, for many applications, it is desirable to obtain the correspondence between semantic concepts and image regions. Traditional image segmentation techniques have been adopted in these region-based image annotation systems [1, 10]. In order to pass-by the challenge of semantic image segmentation, block-based algorithms for image retrieval have been proposed [5, 7]. We have previously proposed a block-based semantic annotation approach, the Spatial Hidden Markov Model (SHMM), which extends the Hidden Markov Model (HMM) to learn, classify and then annotate images automatically [11]. Experiments demonstrate that the SHMM-based approach can achieve high classification and annotation accuracy.

As with most block-based algorithms, the performance of SHMM, to a certain extent, depends on the choice of block size for the image concerned. In this paper, we provide a sensitivity analysis of SHMM with respect to the choice of block size. To classify an image in different scales, here we regard a SHMM that works on specific block size as a single image classifier and propose a multi-scale SHMM that employs a Bayes rule based multiple-classifier combination approach to integrate the outputs of distinct SHMM classifiers. The experimental results show that the multi-scale SHMM improves image classification accuracy by approximately 6% on average. The key contributions of this work is first we provide a sensitivity analysis of block size in SHMM, and then we propose a novel multi-scale SHMM by, considering SHMMs that work on different block sizes as distinct classifiers so that the classifier combination algorithm can be adopted to improve the image classification accuracy. In particular, the combination of SHMMs using different block sizes ensures that the classification decision is based upon considering the image features in several scales, which serve to improve the accuracy of classification process. The proposed approach can be extended to other block-based image classification algorithms.

The combination of classifiers is normally conducted to reduce the misclassification error of single classifier and achieve a robust performance. Based on the outputs of several classifiers, there are three possible types of situations for combining these classifiers [8]. The problem we presented in the paper falls into the type-one

978-1-4244-2175-6/08/$25.00 ©2008 IEEE

condition, where the classifier merely outputs a unique class label to indicate the most probable class of input pattern. Many combination approaches have been developed and been employed in various applications, including voting (bagging, boosting), Bayesian methods [9], Dempster-Shafer [9], averaging, SVM, decision trees, etc. There have also been some applications that classify images by combining several commonly used classifiers [3, 4, 6].

Our proposed approach is introduced as follows: the theoretical foundation of SHMM is briefly reviewed in Section 2. Section 3 presents the Bayes rule based combination of multiple SHMMs, followed by the experimental results in Section 4.

2. Spatial Hidden Markov Model

The Spatial Hidden Markov Model (SHMM) is a two dimensional generalization of traditional hidden Markov model (HMM). A SHMM λ is a 4-tuple $\lambda = (H, V, B, \pi)$, which specifies the number of states N, the number of observation symbols M, and the four probability measures: H(horizontal state transition matrix), V(vertical state transition matrix), B (observation symbol probability distribution) and π (initial state probability distribution). In SHMM, each image is divided into uniform blocks, from which image feature vectors are extracted. And each concept class of images is represented by a statistical model. An unknown image is classified to a specific model via estimating its feature vectors within blocks and the spatial relationship across blocks. Moreover, a concept class is associated with several keywords. The image blocks are also subsequently annotated with semantic labels.

The SHMM conducts image classification and annotation based on following two assumptions. First, the feature vector of a block is assumed to follow a Gaussian distribution, whose parameters are decided by the state of the block. This assumption is validated in [5]. The second assumption is an extension of the Markov property of Markov chain. It is assumed that the state of a block merely depends on the state of its upper and left neighboring blocks and is independent of other blocks, which is denoted as follows:

$$P(q_{x,y}|Q_{x,y-1}) = P(q_{x,y}|q_{x-1,y}, q_{x,y-1}) \quad (1)$$

Where $q_{x,y}$ denotes the state of block (x, y), and $Q_{x,y-1}$ denotes the state sequence of $q_{1,1}, q_{1,2}, \ldots, q_{x,y-1}$.

Since the detail derivation of the above equation is not the focus of this paper, the readers are referred to [11] for the details.

Table 1. Classification accuracy of SHMM

Group	**Classification Accuracy**		
	Block size: 32×32	Block size: 48×48	Block size: 64×64
Dinosaur	71.67%	83.33%	68.33%
Beach	80.00%	78.33%	85.00%
Bus	75.00%	83.33%	88.33%
Elephant	71.67%	73.33%	65.00%
Flower	76.67%	93.33%	96.67%
Food	86.67%	71.67%	65.00%
Horse	93.33%	85.00%	88.33%
Mountain	68.33%	66.67%	63.33%
Total	**77.92**%	**77.38**%	**77.50**%

3. A Bayes rule approach for combining multiple SHMMs

3.1. Sensitivity analysis of block size in SHMM

As mentioned in the last section, the SHMM is a block-based image content analysis model. Images to be processed are divided into uniform blocks, from which feature vectors are extracted. Consequently, changing of the block size of an image is likely to affect the outputs of the SHMM. In this section, we present a sensitivity analysis of the performance of SHMM with respect to the choice of block size.

We performed a series of image classification experiments with three different block sizes, (32×32 pixels, 48×48 pixels, and 64×64 pixels), on eight classes of images, each containing 100 images. The results are shown in Table1. We can observe that the accuracy ratios of the different image classes do fluctuate according to the block size. At the same time, it is also seen that there is not an optimal block size that is appropriate for all the tested image classes. Further details of the experiments can be found in section 4.1. Suffice here to conclude that performance of image classification based on SHMM is influenced by the choice of block size.

3.2. Bayes rule combination of multiple SHMMs

As stated in section 3.1, one image is likely to be assigned to different classes by SHMM using different block sizes. In this section, we utilize a Bayes rule approach to improve the classification accuracy ratio by integrating these separate outputs into a final classification decision.

The SHMM can be viewed as classifier that assigns an unknown image to one of the N candidate classes,

$\omega_1, \omega_2, \ldots, \omega_N$. Given an unknown image x, we apply SHMM to do classification on it using K different block sizes, $b_1, b_2, \ldots, b_K$, respectively. The SHMM with different block sizes are regarded as distinct classifiers, which are denoted by $\varphi_1, \varphi_2, \ldots, \varphi_K$ in accordance with block size $b_1, b_2, \ldots, b_K$. Then an event $\varphi_k(x) = i, 1 \leq i \leq N$, denotes that image x is classified to class ω_i by classifier φ_k. Let the classification results of K classifiers to be $r_1, r_2, \ldots, r_K$. The Bayesian decision rule indicates that the optimal rule in classification is to assign an image to the class with largest posterior probability [8], as follow:

$$x \in \omega_I,\ with \tag{2}$$
$$I = \arg\max_i (P(x \in \omega_i \mid \varphi_1(x) = r_1, \ldots, \varphi_K(x) = r_K))$$

In principle, the probabilities $P(x \in \omega_i \mid \varphi_1(x) = r_1, \ldots, \varphi_K(x) = r_K), 1 \leq i \leq N$, are not available and can be estimated from training images. However, this approach requires to build a probability matrix of $P(x \in \omega_i \mid \varphi_1(x) = r_1, \ldots, \varphi_K(x) = r_K)$, consisted of N^{K+1} items. The computation is expensive. To build such a table needs more than N^{K+1} training images. Therefore, we employ a Bayes rule [9] approach to deal with the problem as follows .

Let $P(x \in \omega_i \mid \varphi_k(x) = r_k)$ denotes the probability that image x comes from class ω_i, given the classification result of classifier φ_k on x is r_k. We can easily build the probability matrix of $P(x \in \omega_i \mid \varphi_k(x) = r_k)$ for each block size by estimating training images as the prior knowledge, with each matrix consisting of N^2 items. The probability matrix of block size b_i is called Classification Accuracy Tables of b_i in this paper.

According to Bayes rule, the probability in equation (2) equals to:

$$\frac{P(\varphi_1(x) = r_1, \ldots, \varphi_K(x) = r_K \mid x \in \omega_i) \cdot P(x \in \omega_i)}{P(\varphi_1(x) = r_1, \ldots, \varphi_K(x) = r_K)},$$
$$with\ \ 1 \leq i \leq N \tag{3}$$

Here we suppose classifiers $\omega_1, \omega_2, , \omega_N$ are independent to each other. Since they conduct classification with distinct feature vectors, where classifier ω_i use the feature vectors extracted from blocks with size of b_i. Then the equation (3) is transformed into:

$$\frac{P(x \in \omega_i) \cdot \prod_{j=1}^{K} P(\varphi_j(x) = r_j \mid x \in \omega_i)}{\prod_{j=1}^{K} P(\varphi_j(x) = r_j)} \tag{4}$$

Following the Bayes rule, we get

$$P(x \in \omega_i) \cdot \prod_{j=1}^{K} \frac{P(x \in \omega_i \mid \varphi_j(x) = r_j)}{P(x \in \omega_i)}$$

$$= \frac{\prod_{j=1}^{K} P(x \in \omega_i \mid \varphi_j(x) = r_j)}{P(x \in \omega_i)^{K-1}} \tag{5}$$

In practice, the value of equation (5) can be approximated by [9]:

$$\frac{\prod_{j=1}^{K} P(x \in \omega_i \mid \varphi_j(x) = r_j)}{\eta},\ where\, \eta\, is\, a\, constant$$

$$and,\ \eta = \sum_{i=1}^{N} \prod_{j=1}^{K} P(x \in \omega_i \mid \varphi_j(x) = r_j) \tag{6}$$

Substituting (6) for the right hand part of equation (2), the image x should be assigned to class ω_I, where:

$$I = \arg\max_i \left(\frac{\prod_{j=1}^{K} P(x \in \omega_i \mid \varphi_j(x) = r_j)}{\eta}\right) \tag{7}$$

4. Experimental results

4.1. Sensitivity Analysis of Block Size in SHMM

We conducted experiments on a subset of the COREL standard image database, consisting of eight image categories with 100 images in each. All the images are in the size of either 384×256 pixels or 256×384 pixels. For each category, 40% of the images are used to build the SHMM for that category. The other 60% images (480) are used as testing images to evaluate the classification accuracy of the resulting SHMM.

Initially, experiments are conducted to test the classification accuracy of SHMM with specific block size using all the 480 testing images. The classifications of the 480 images are carried out with block size of 32×32 pixels, 48×48 pixels, and 64×64 pixels respectively. Table 1 shows the percentage of correctly classified images in each category with respect to the different block sizes. It can be seen from the table that the classification accuracy varies with respect to the block sizes. And a single fixed block size cannot always provide the best accuracy ratio for all the semantic categories

4.2. Comparison of the performances of the SHMM and Multi-scale SHMM

Experiments were performed to compare the classification accuracy of SHMM and the multi-scales SHMM presented in this paper. Among the 480 images of the eight semantic classes, we randomly selected n images from each class as the training images to build the Classification Accuracy Tables as prior knowledge, while the remaining images were used for comparison. The

Table 2. Classification accuracies of SHMM (for 3 different block sizes) and multi-scale SHMM with n=40

Block Size	32×32	48×48	64×64	multi-scale SHMM
Accuracy	76.25%	76.25%	71.88%	81.88%

three block sizes considered here are 32×32, 48×48 and 64×64, which are conventionally used in many applications. Table 2 shows the experimental result with 320 training images and 160 testing images, where n takes the value of 40. The Classification Accuracy indicates the proportion of correctly classified images to the entire set of testing images. As shown in the table, the Bayes rule approach improves the classification accuracy by 5.63%, 5.63% and 10.00% respectively. It can be seen from Figure 1 that combination of multiple SHMMs can always achieve higher classification accuracy than the single SHMM irrespective of the block size, regardless of the ratio of training images and testing images. The overall increase of accuracy brought about by the multi-scale SHMM approach is up to 10% and is approximate 6% on average. As the semantic annotation of image blocks depends on the classification result, the accuracy of annotation increases accordingly.

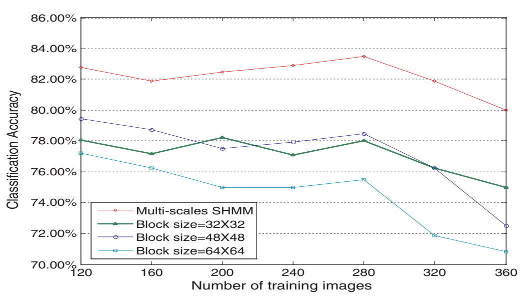

Figure 1. Comparison of classification accuracies

5. Conclusion

In this paper, we propose a multi-scale classification of image by combining multiple SHMMs using the Bayes rule approach. Specifically, we combine the classification results of several SHMM using different block sizes to achieve higher image classification accuracy. As a block-based algorithm, SHMM extracts semantic content based on image feature vectors extracted from these blocks. In this paper, we have shown that the performance of SHMM is, to a certain extent, influenced by the size of the blocks. We further present applying the Bayes rule classifier combination approach to combine the outputs of multiple SHMMs to improve the performance of image classification. Experiments have shown that this SHMM combination approach can significantly improve the classification accuracy of single SHMM for a diverse class of images.

6. Acknowledgment

The work described in this paper is fully supported by a grant from HKSAR Research Grant Council project no: CityU114007.

References

[1] D. M. Blei and M. I. Jordan. Modeling annotated data. In *Proc. 26th Int. ACM SIGIR Research and Development in Information Retrieval*, 2003.

[2] J. Fan, Y. Gao, H. Luo, and R. Jain. Mining multilevel image semantics vis hierarchical classification. *IEEE Trans. Multimedia*, 10(2):167–187, 2008.

[3] G. Giacinto, F. Roli, and L. Bruzzone. Combination of neural and statistical algorithms for supervised classification of remote-sensing images. *Pattern Recognition Letters*, 21:385–397, 2000.

[4] L. Lepisto, I. Kunttu, J. Autio, and A. Visa. Classification of non-homogenous texture images by combining classifiers. *Proc. IEEE Int. Image Processing*, 1:981–984, 2003.

[5] J. Li, R. M. Gray, and R. A. Olshen. Multiresolution image classification by hierarchical modeling with two-dimensional hidden markov models. *IEEE Trans. Pattern Anal. Machine Intell*, 46(5):1826–1841, 2000.

[6] X. Lu, Y. Wang, and A. K. Jain. Combining classifiers for face recognition. *Proc. 2003 Int. Multimedia and Expo*, 3:13–16, 2003.

[7] Y. Mori, H. Takahashi, and R. Oka. Image-to-word transformation based on dividing and vector quantizing images with words. In *First International Workshop on Multimedia Intelligent Storage and Retrieval Management*, 1999.

[8] D. M. J. Tax, M. V. Breukelen, R. P. W. Duin, and J. Kittler. Combining multiple classifiers by averaging or by multiplying? *Pattern Recognition*, 33:1475–1485, 2000.

[9] L. Xu, A. Krzyzak, and C. Y. Suen. Methods of combining multiple classifiers and their applications to handwriting recognition. *IEEE Trans. Systems, Man, and Cybernetics*, 22(3):418–435, 1992.

[10] C. Yang, M. Dong, and F. Fotouhi. Region based image annotation through multiple-instance learning. In *Proc. 13th ACM Int. Multimedia*, pages 435–438, 2005.

[11] F. Yu and H. H. S. Ip. Automatic semantic annotation of images using spatial hidden markov model. In *Proc. Int. Multimedia and Expo*, pages 305–308, 2006.