

# An HMM-SVM-Based Automatic Image Annotation Approach

Yinjie Lei, Wilson Wong, Wei Liu, and Mohammed Bennamoun

School of Computer Science and Software Engineering  
University of Western Australia  
35 Stirling Highway, Crawley WA 6009  
{yinjie,wilson,wei,bennamou}@csse.uwa.edu.au

**Abstract.** This paper presents a novel approach to **Automatic Image Annotation (AIA)** which combines both **Hidden Markov Model (HMM)** and **Support Vector Machine (SVM)**. Typical image annotation methods directly map low-level features to high-level concepts and overlook the importance to mining the contextual information among the annotated keywords. The proposed HMM-SVM based approach comprises two different kinds of HMMs based on image color and texture features as the first-stage mapping scheme and an SVM which is based on the prediction results from the two HMMs as a so-called high-level classifier for final keywording. Our proposed approach assigns 1-5 keywords to each testing image. Using the Corel image dataset, Our experiments have shown that the combination of a discriminative classification and a generative model is beneficial in image annotation

## 1 Introduction

The modern developments of the Internet make it the most efficient platform for obtaining and sharing various kinds of information from anywhere. For this reason the research into search engines for retrieving and managing multimedia data has become very important and attractive [1]. Existing search engines are well-developed in the case of textual data. However, more research is still required for image search and retrieval due to the so-called *semantic-gap*. At the early stage of research, image retrieval was performed by relying on manually assigned keywords. The manual labeling of images however is tedious and difficult for large image collections. To address these drawbacks, content-based image retrieval using low-level image features such as color, texture and shape is proposed [2]. These low-level features representing visual content of an image can be used to measure the similarity between images. This allows images from datasets to be automatically indexed and searched. To improve the process of retrieval, this line of research based on low-level features was soon replaced by the use of the approach of AIA which associates multiple keywords with objects in images. Some researchers argued that if we can associate multiple keywords with the identified object in the image, the retrieval of images could become much easier and more straightforward [3, 4, 5, 6].

For this reason, AIA has become a focus in the area of content-based image retrieval to bridge the semantic gap [7]. In recent years, the classifier ensembles approach has attracted much more attention. Some results report that it is more reliable than most one-level classifier in performing automatic image annotation [8]. Another trend of classification is the fusion with other techniques to enhance the performance [9]. In practice, the intrinsic advantages of generative model have been widely accepted and used in the area of automatic image annotation. Recently, one representation of generative models, namely the Hidden Markov Model has been utilized to resolve automatic image annotation problems [10]. However, there are still opportunities to improve the quality of automatic image annotation for two reasons. First, images which are semantically similar often contain different low-level features. Therefore the direct mapping of the low-level features to high-level concepts may lead to errors. Second, most existing approaches overlook the significance of keyword correlation in image retrieval. For instance, ‘boat’ and ‘water’ tend to co-occur much more often in one image than ‘boat’ and ‘grass’. This suggests the correlation information among keywords can be of great help to improve the performance of AIA.

In this paper, we present a two-stage mapping AIA technique based on both Support Vector Machine and Hidden Markov Model. The first stage comprises two HMMs constructed separately from color and texture features of images for mapping the low-level features to mid-level features. Co-occurrence based keyword correlation is also constructed to enhance the mapping precision. In the second stage we employ support vector machine to map the so-called mid-level features to high-level concepts. The proposed scheme fuses both a discriminative classification and a generative model to avoid the two problems discussed above.

The outline of this paper is as follows: Recent image annotation methods based on both SVM and HMM are briefly reviewed in Section 2. Our proposed SVM-HMM based annotation approach is explained in Section 3. Section 4 presents the experimental results and the performance analysis of the proposed method. The paper is then concluded in Section 5.

## 2 Related Work

Automatic image annotation techniques first appeared about two decades ago. Below is a review of some selected milestones in AIA using SVM and HMM.

Support Vector Machine was first introduced into this area during the last decade. As a very strong data mining technique, one of the first SVM based image classification system paper is [11]. However they only use global color features to solve a small scale classification problem. With the aim to improve the classification accuracy based on a single classifier, a sophisticated classifier system called “classifier ensembles” was introduced to further improve AIA precision. Gao et al. [12] use a combination of multiple SVM classifiers. These classifiers are obtained by combining the output of several effective weak classifiers using a

Boosting technique. Subsequently, Qi and Han [13] also use a combination of two sets of SVMs which relies on the regional image features found using **Multiple Instance Learning (MIL)** and global image features respectively. Tsai et al. [8] present an image indexing and classification system called **CLAIRE**. Their system is based on a **Two Stage Mapping Model (TSMM)** [14]. In their system, three SVMs are constructed as low-level feature classifiers focusing on classifying color and texture features respectively. Another SVM called high-level classifier is constructed based on the outputs of the first low-level classifiers. This system avoids the direct mapping of the low-level features to high-level concepts, and the results show a promising way to assign keywords to images.

As one representative work of generative models, HMM has also been adopted by some researchers to perform AIA. In [15], a one-dimensional hidden Markov model (HMM) was trained on vector-quantized color histograms of image blocks. However, their system can only be used to solve a binary image classification problem. Li and Wang [16] proposed a system called ALIP which is based on a two-dimensional multi-resolution HMM fed by regional image features. Modestino and Zhang [17] use a Markov random field model to capture the spatial relationships between regions and apply a maximum posteriori rule to interpret images. Ghoshal et al. [10] use an HMM for image and video annotation based on two datasets individually, which are COREL and TRECVID. A novel TSVM-HMM based annotation scheme is proposed in [18]. Compared with previous annotation methods, the proposed TSVM-HMM based annotation scheme can achieve better annotation performance with less labeled training images as demonstrated.

### 3 Proposed Approach

In order to overcome the problems discussed above, we propose a Two-Stage Mapping Model to perform Automatic Image Annotation (AIA). An overview of the proposed approach is shown in Fig.1. The first module is composed of two Hidden Markov Models which are responsible for classifying low-level color and texture features respectively. The second module is an SVM classifier which serves as a high-level classifier as in the work of [8] to determine the final annotation results. Unlike that paper, our approach substitutes the SVM with an HMM during the first stage aiming at mining keyword correlations. Meanwhile, we directly use the category names to define the output of the HMMs. This is to overcome the difficulty of only using twelve colors to describe a large number of images. Moreover, all the image regions of our training set are used as opposed to using only the central region. Therefore the image regions used in our approach may contain different objects. This change also adds difficulty to the process of describing image regions with only twelve color names. Once these two stages are completed, five keywords corresponding to the five sub-blocks can automatically be assigned to a test image. Below is a description of each module of our approach in Fig. 1.

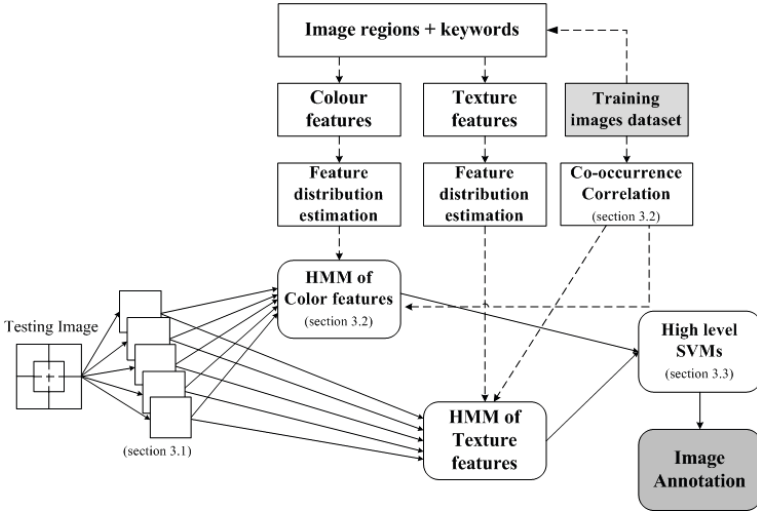


Fig. 1. A block diagram of our proposed approach

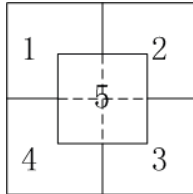


Fig. 2. The tilting scheme

### 3.1 Image Sub-blocking and Feature Extraction

It is well known that automatic image segmentation is a hard task and no approach can achieve perfect results. Moreover, some results show that those models using a sub-blocking scheme perform better than those using object-based segmentation [19]. We use a tilt scheme which was proposed in [8] to divide images into five regions. The original image size in our dataset has  $384 \times 256$  pixel resolution, and the region size is  $192 \times 128$  as shown in Fig.2. The regions include four quadrants. The one in the center is used to increase the weight of the object of interest.

We extract color features and texture features as image descriptors. We do not consider other features such as shape for two reasons. First it is well known that image shape feature extraction is difficult to achieve and computationally expensive. In addition it should be noted that image regions whether containing homogeneous objects or not is not a focus in our approach. Therefore it is meaningless even if image shape feature extraction is applied.

The color features include the mean and standard deviation of every region in the RGB and Lab color spaces. It has been proved that Garbor filter performs

well on extracting image texture features. Therefore we apply a set of Garbor filters with 12 orientations (i.e.  $0^\circ, 30^\circ, 60^\circ, \dots, 270^\circ$ ) on the luminance component of image regions. We then extract the mean and standard deviation values of the 12 filtered images and use them as texture features. This results in a feature vector of length 36 for each region (i.e. 12 color features and 24 texture features).

### 3.2 Hidden Markov Model for Low-Level Annotation

**Hidden Markov Model for AIA.** According to HMM's definition, it is easy to provide a density function to model image features of image regions which belong to the same keyword. By introducing keyword correlation, the context-dependent HMM can improve its accuracy for image annotation.

For the sake of brevity, let  $T_i = \{I_{i1}, I_{i2}, \dots, I_{in}\}$  be the feature set of image regions obtained from our training set for the  $i$ th keyword, where  $n$  is the total number of image regions. The keyword set  $K = \{k_1, k_2, \dots, k_i\}$  represents all the keywords appearing in the whole training set. Given an image, it will be divided into five regions as described above, where the regions are ordered according to the quadrants as shown in Fig. 2. The upper-left one is considered as the first while the center one as the last. Meanwhile, we use  $I_r = \{I_{r1}, I_{r2}, \dots, I_{r5}\}$  to denote its region feature set and  $I_c = \{I_{k1}, I_{k2}, \dots, I_{k5}\}$  to denote its keyword set. We propose to model the AIA task as a Hidden Markov process. Thus, by combining  $I_r$  and  $I_c$ , the joint likelihood function can be formulated as

$$f(I_{r1}, I_{r2}, \dots, I_{r5}, I_{k1}, I_{k2}, \dots, I_{k5} | k_0) = \sum_{k_t \in I_k} \prod_{t=1}^5 f(I_t | k_t) p(k_t | k_{t-1}) \quad (1)$$

According to Eq. (1), an HMM model is mainly affected by the emission density function  $f$ . This function corresponds to the image region feature distribution of one keyword. The transition probability function  $p$  on the other hand reflects the keyword correlations. The problem then becomes how can we formulate the emission density function and transition probability function.

**Low-level Feature Distribution.** As discussed in the last subsection, we should establish a useful emission density function for each keyword [10]. Gaussian Mixture Model (GMM) is one of the most statistically mature methods for density estimation [18]. GMM is the weighted average of Gaussians, and each Gaussian has its own mean and covariance matrix which has to be estimated separately. In the proposed approach, we use GMM to model the low-level image features via relevant image regions. Let  $T_c = \{I_{c1}, I_{c2}, \dots, I_{cn}\}$  and  $T_t = \{I_{t1}, I_{t2}, \dots, I_{tn}\}$  denote the extracted color and texture features of image regions assigned with keyword  $k$ . We then employ a Gaussian mixture model with three components to construct the color and texture feature distribution functions  $f_c(T_c | c)$  and  $f_t(T_t | c)$  as follow,

$$f_{c,t}(T_{c,t}|k) = \alpha_1 g(T_{c,t}; \mu_1, \Sigma_1) + \alpha_2 g(T_{c,t}; \mu_2, \Sigma_2) + \alpha_3 g(T_{c,t}; \mu_3, \Sigma_3) \quad (2)$$

$$g(T_{c,t}; \mu_i, \Sigma_i) = \frac{1}{\sqrt{(2\pi)^{d_{c,t}} |\Sigma_i|}} \exp\left[-\frac{1}{2}(T_{c,t} - \mu_i)^T \Sigma_i^{-1} (T_{c,t} - \mu_i)\right] \quad (3)$$

where  $\alpha_1, \alpha_2, \alpha_3$  represent the weight of each Gaussian component respectively, and  $\alpha_1 + \alpha_2 + \alpha_3 = 1$ .  $\mu$  and  $\Sigma$  denote the mean and covariance matrix respectively.  $d$  denotes the dimension of an image region feature. Here  $d_c = 12$  for color features and  $d_t = 24$  for texture features.

**Keyword Correlation.** The transition probability  $p(k_t|k_{t-1})$  reflects the correlation between the keywords  $k_t$  and  $k_{t-1}$ . Here we use  $k_i$  and  $k_j$  to replace  $k_t$  and  $k_{t-1}$  respectively. The keyword correlation  $p(k_i|k_j)$  is measured by counting the frequency of paired words assigned to each image. We can estimate conditional and joint probabilities of  $p$  if we take:  $p(k_i|k_j) = \frac{p(k_i, k_j)}{p(k_j)}$ , and  $p(k_i, k_j) = \frac{N(k_i, k_j)}{|D|}$ , where  $N(k_i, k_j)$  indicates the number of times  $k_i$  and  $k_j$  appear together in one image, and  $|D|$  is the total number of image regions in the training set. If we repeat this process for each pair of words in the keyword set we can obtain an  $i \times i$  conditional probability matrix  $P_M$ , which reflects the keyword correlation.

A problem with the  $P_M$  matrix is that some of the keywords never appear in the same image. Thus some  $p(k_i, k_j)$  may take a value of zero. We apply a widely used smoothing technique known as ‘‘interpolation smoothing’’ to solve this problem. It can be summarized by Eq. 4.

$$p(k_i|k_j) = \beta * \frac{N(k_i, k_j)}{N(k_j)} + (1 - \beta) * \frac{N(k_j)}{|D|} \quad (4)$$

where  $\beta$  is an interpolation parameter and  $|D|$  is the number of words in the collection. This formula is an interpolation between the empirical estimate  $\frac{N(k_i, k_j)}{N(k_j)}$  and the empirical distribution of the term  $k_j$ . Therefore even if two keywords never appear together, we will not have a zero value in  $P_M$ . It should be noted that both the color and texture HMMs share the same  $P_M$ . It is easy to understand that although the feature sets are different, they should have the same keyword correlation.

**Predictions of HMM.** The objective of AIA is to find the optimal hidden keyword sequence for regions with learnt HMM. Once the density estimation of  $f_{c,t}(T_{c,t}|k)$  for color and texture features of all keywords and transition probabilities have been estimated, given a test image, we perform the Balm-Welch algorithm to compute the posterior probability of each prediction as the first-stage annotation. The posterior probability  $d_j(I_t)$  of being predicted with  $k_j$  is iteratively achieved using:

$$d_j(I_t) = f(I_t|k_j) \sum_{i=1}^M d_i(I_{t-1}) p(k_i|k_j) \quad (5)$$

In Eq. 5, the posterior probabilities of  $M$  keywords, i.e.  $d_j(I_t), j = 1, \dots, M$ , are acquired through the association with a visible region  $I_t$ . The color and texture predictions  $j'_{c,t}$  of the hidden keyword for the region  $I_t$  can be gained based on the following criterion:

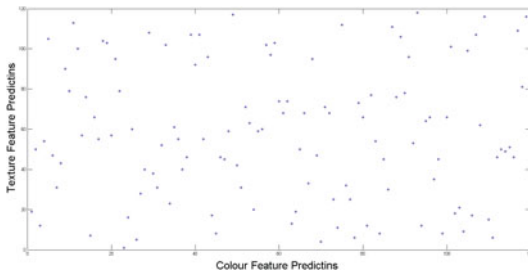
$$j' = \operatorname{argmax}_j(d_j(I_t)) \quad (6)$$

### 3.3 High-Level Concept Classifier

**Training Set for High-level concept classifier.** Unlike [8], our approach directly extracts the predictions from the colour and texture Hidden Markov Model. Let  $T_i = \{I_{i1}, I_{i2}, \dots, I_{in}\}$  be all the feature set of one keyword. We then apply the constructed color HMMs and textures HMMs to the set and collect the color and texture predictions. After this process, we collect the prediction set  $T_i = \{(c_{i1}, t_{i1}), (c_{i2}, t_{i2}), \dots, (c_{in}, t_{in})\}$ .

Let us take the concept 'grass' as an example. After applying the HMMs to the training set, we collect all predictions belonging to the image regions which are labelled 'grass'. Let  $T_{grass} = \{I_{grass1}, I_{grass2}, \dots, I_{grassN}\}$  be all the image regions under the keyword 'grass'. We assume the color and texture predictions for  $I_{grass1}$  as (*tree - color, grass - texture*) and for  $I_{grass2}$  as (*grass-color, sky-texture*). If we repeat the application of the HMMs for all the image regions belonging to the keyword 'grass', we can collect all the predictions for that keyword. Then we use this kind of predictions as mid-level features for each keyword. Let  $M_{grass}$  denotes the output of the HMM for the keyword 'grass', and  $M_{grass} = \{(tree - color, grass - texture), (grass - color, sky - texture), \dots\}$ .

In fact, every image region for all the keywords is mapped into a space that we call the HMM prediction space. This space maps color feature predictions on the X-axis and texture feature predictions on Y-axis. According to the total number of keywords over all the training set, which is 120, both the X- and Y-axis will take values ranging from 1 to 120. Fig. 3 shows an example of an HMM prediction space.



**Fig. 3.** An example of HMM prediction space to illustrate the use of SVMs for learning predictions based on different keywords

**High-level concept SVM.** For classifier design, support vector machines (SVMs) are chosen for the image classification task because of their generalization performance superiority. For the selection of the kernel function for the nonlinear mapping, a degree-2 polynomial kernel is used. We use the predictions of HMMs as input to the SVM, and when the number of keywords is  $i$ , therefore this method constructs  $i$  SVM classifiers. Similar to the multi class classification employed in this paper, the one-against-all method is used. Each input is classified into one positive (+1) and  $C - 1$  negative (-1) classes.

### 3.4 Summary of the Training Procedure

- *Step 1.* Given training set image regions for all keywords. These training sets are composed of 12 color features and 24 texture features.
- *Step 2.* Use the extracted color and texture features to construct the color and texture feature distribution functions  $f_c$  and  $f_t$  used as density estimates for HMMs as shown in Equations 2 and 3.
- *Step 3.* Investigate the keyword correlation based on the labelled regions to obtain the co-occurrence matrix  $P_M$  as transition probability function as described in Equation 4.
- *Step 4.* Return to the training sets in Step 1. Use the constructed HMMs to collect the color and texture predictions of all the image regions. By this time, the training set belonging to each keyword would have been mapped to the prediction space. At the end of this step, the first mapping stage would have been generated.
- *Step 5.* Use the prediction space of each keyword to obtain the high-level concept SVMs. At the end of this step, the second mapping stage would have been generated.

## 4 Experiments

### 4.1 Dataset

We tested the proposed AIA approach on the Corel dataset with 5600 images. A selection of 3456 images in the dataset was initially divided into five regions, and all regions were grouped into 120 keywords. Since a region may contain different objects, if one object occupies more than half in the region, the object name will be assigned to this region. We also discarded some regions which are difficult to label. During the training process, every keyword contained around 54 to 810 regions, with a total of 13754 training regions. Next, another 635 images which had no regions appeared in the training set. They were randomly chosen from the dataset and used as testing images. The proposed approach comprises one color, one texture HMMs and 120 SVMs. The color and texture names are the same as the 120 keywords predicted using our HMMs. During the testing process, 5 keywords were automatically assigned by the proposed approach to the testing images.





Fig. 4. Some region samples of four keywords used for training

Table 1. Performance comparison with other methods in terms of average precision and recall for all keywords

	HMM	CLARIE	HMM-SVM
Words with recall > 0	59	0.76	<b>102</b>
Average words recall	0.21	0.34	<b>0.47</b>
Average word precision	0.19	0.32	<b>0.45</b>




## 4.2 Comparison with Other Methods

The contribution of the proposed HMM-SVM based annotation scheme is to integrate both the discriminative classification and the generative model so as to take full advantage of their combined merits. To evaluate its effectiveness, we compared our HMM-SVM based approach with other two related approaches, namely CLAIRE [8] and HMM-based image annotation [10]. For each method, we assess the annotation performance using the average precision and recall, over all testing images. The precision and recall values are defined in Eq. (7):

$$precision(c) = \frac{num_c}{num_{ca}}; \quad recall(c) = \frac{num_c}{num_{cm}} \quad (7)$$

where  $num_c$  denotes the number of image correctly annotated with keyword  $c$ ,  $num_{ca}$  denotes the number of images automatically annotated with keyword  $c$  and  $num_{cm}$  denotes the number of images manually annotated with keyword  $c$ .

Table 1 shows the average annotation precision and recall over the total 120 keywords. Clearly, we can see that the proposed HMM-SVM based annotation method achieves a significant improvement on our experimental dataset. Compared to the other two methods, it shows an improvement of about 26% and 0.13% in recall and 26% and 13% in precision. Moreover, the number of keywords with positive recalls has increased by 43 and 26. Fig. 5 presents some examples of the annotations produced by the proposed approach. The potential reasons for this improvement can be associated to the following: (1) with the

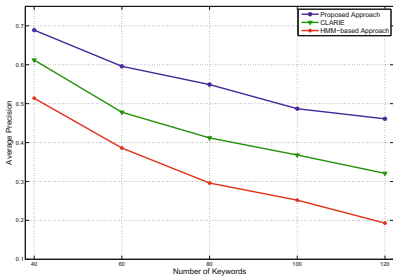
	Ground truth Annotation	HMM-SVM Annotation
	Sky, sky, grass people church	Sky, sky, grass, temple, people
	Sky, mountain, grass, grass, elephant	Sky, sky, grass, grass, elephant
	Cloud, cloud, water, water, castle	Cloud, cloud, water, water, castle

**Fig. 5.** Examples of image annotation

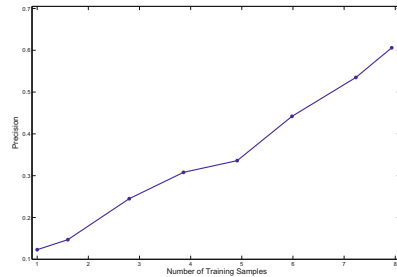
two-stage mapping scheme involved, it is believed that HMM-SVM can outperform those use only single-mapping approaches such as HMM method; (2) the enhanced keyword correlation is also introduced into the proposed AIA approach and hence keyword semantics is capable of being modelled well compared to other methods that do not consider such correlation (e.g. CLAIRE). More details are provided in the next section.

### 4.3 Effectiveness of the Proposed Approach

To construct a reliable generative model, i.e. HMM, our approach employs a keyword correlation with an interpolation smoothing technique and further promotes the performance of HMM. Three schemes are used to obtain the keyword correlation, i.e. the co-occurrence based keyword correlation, the co-occurrence based keyword correlation without interpolation smoothing, and without keyword correlation which is set to be uniform. As shown in Fig. 6, the performance of annotation is greatly improved when taking into consideration keyword correlations, a concept not used in previous annotation approaches. By combining



**Fig. 6.** Evaluation of the effectiveness of keyword correlation based on three schemes described in section 4.3



**Fig. 7.** The annotation precision of eight keywords which contain different number of training region samples

co-occurrence correlation measurements and the interpolation smoothing technique, it can provide more reliable keyword correlation to avoid zero values.

We further examine the relationship between the precision results and the number of training samples. The ten keyword samples randomly selected from the training set, as in Fig. 7, shows an approximately linear relationship between the number of training samples and the annotation precision. We see that a larger number of training samples is a major factor for a better annotation. Therefore, performance should be heavily dependent on the low-level feature representations which employ visual feature distribution functions.

## 5 Conclusion

In this paper, we proposed an approach for Automatic Image Annotation based on the concept of two-stage mapping. Unlike existing two-stage mapping models, the proposed approach combines the advantages of two-stage mapping and keyword correlation. This two-stage mapping scheme avoids the direct mapping of low-level features to high-level concepts. The keyword correlation mechanism is able to capture to a certain extent the meaning of words to improve the performance of AIA. Our experimental results using the Corel image dataset show that, in the case of annotating images with few words, the combination of the discriminative classification and the generative model can improve annotation performance. Thus, the combination of HMM and SVM provides a promising way to perform and improve automatic annotation of images.

**Acknowledgement.** This research was supported by the Australia Research Council grant and the University of Western Australian and the China Scholarship Council Joint Scholarship.

## References

1. Li, J., Wang, J.: Real-time computerized annotation of pictures. In: Proceedings of the ACM Multimedia Conference, pp. 911–920 (2006)
2. Smeulders, A., Worring, M., Santini, S., Gupta, A., Jain, R.: Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22, 1349–1380 (2000)
3. Duygulu, P., Barnard, K., de Freitas, N., Forsyth, D.: Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) *ECCV 2002*. LNCS, vol. 2353, pp. 97–112. Springer, Heidelberg (2002)
4. Barnard, K., Duygulu, P., Forsyth, D., de Freitas, N., Blei, D., Jordan, M.: Matching words and pictures. *Journal of Machine Learning Research* 3, 1107–1135 (2003)
5. Blei, D., Jordan, D.: Modeling annotated data. In: 26th Annual International ACM SIGIR Conference, pp. 127–134 (2003)
6. Jeon, J., Lavrenko, V., Manmatha, R.: Automatic image annotation and retrieval using cross-media relevance models. In: 26th Annual International ACM SIGIR Conference, pp. 119–126 (2003)

7. Rui, Y., Huang, T., Chang, S.: Image retrieval current techniques, promising directions and open issues. *J. Visual Commun. Image Representation* 10, 39–62 (1999)
8. Tsai, C., McGarry, K., Tait, J.: Claire: A modular support vector image indexing and classification system. *ACM Transactions on Information Systems* 24, 353–379 (2006)
9. Wong, W., Hsu, S.: Application of svm and ann for image retrieval. *European Journal of Operational Research* 173, 938–950 (2006)
10. Ghoshal, A., Ircing, P., Khudanpur, S.: Hidden markov models for automatic annotation and contentbased retrieval of images and video. In: *ACM Conference on Special Interest Group on Information Retrieval (SIGIR)*, Brazil (2005)
11. Chapelle, O., Haffner, P., Vapnik, V.: Support vector machines for histogram-based image classification. *IEEE Transactions on Neural Networks* 10, 1055–1064 (1999)
12. Gao, Y., Fan, J., Xue, X., Jain, R.: Automatic image annotation by incorporating feature hierarchy and boosting to scale up svm classifiers. In: *Proceedings of the 14th Annual ACM International Conference on Multimedia, MULTIMEDIA 2006*, New York, NY, USA, pp. 901–910 (2006)
13. Qi, X., Han, Y.: Incorporating multiple svms for automatic image annotation. *Pattern Recognition* 40, 728–741 (2007)
14. Tsai, C.: Stacked generalization: A novel solution to bridge the semantic gap for contentbased image retrieval. *Online Inf. Rev.* 27, 442–445 (2003)
15. Yu, H., Wolf, W.: Scenic classification methods for image and video database. In: *Proceedings of the SPIE International Conference on Digital Image Storage and Archiving Systems*, pp. 363–371 (1995)
16. Li, J., Wang, J.: Automatic linguistic indexing of pictures by a statistical modeling approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25, 1075–1088 (2003)
17. Modestino, J., Zhang, J.: A markov random field model-based approach to image interpretation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14, 606–615 (1992)
18. Zhao, Y., Zhao, Y., Zhu, Z.: Tsvm-hmm: Transductive svm based hidden markov model for automatic image annotation. *Expert Systems with Applications* 36, 9813–9818 (2009)
19. Feng, S., Manmatha, R., Lavrenko, V.: Multiple bernoulli relevance models for image and video annotation. In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, vol. 2, pp. 1002–1009 (2004)