Boosted Cannabis Image Recognition

Nianhua Xie, Xi Li, Xiaoqin Zhang, Weiming Hu
Institute of Automation, Chinese Academy of Sciences
{nhxie, lixi, xqzhang, wmhu}@nlpr.ia.ac.cn

James Z. Wang*
The Pennsylvania State University
jwang@ist.psu.edu

Abstract

With the large number of Websites promoting the use of illicit drugs, it has become important to screen these sites for the protection of children on the Internet. Conventional keyword-based approaches are not sufficient because these Websites often have lots of images and little meaningful words than prices. We propose an AdaBoost-based algorithm for cannabis image recognition. This is the first known attempt at computerized detection of illicit drug Web contents using images. The main technical contributions of our work are two-fold. First, we introduce a novel weak classifier which considers the inherently structural property or "self-similarity" of the cannabis plants. The self-correlation structural characteristics of cannabis can be used as a discriminative property for the purpose of cannabis image recognition. Second, we propose a rapid weak classifier finder, which can efficiently select discriminative weak classifiers from the weak classifier space with little degradation to the classification accuracy. Experiments on real world images have demonstrated improved performance of our method over other methods.

1 Introduction

With the rapid development of the digital world and the freedom of speech protected on the Internet, there have been a wide spread of Websites promoting illicit drugs. Like in the filtering of online pornographic content, keyword-based filtering is not sufficient. It is therefore important to develop image-based filters for public schools and libraries. Among images on these sites, drug-type cannabis images are typical. Cannabis images are often used to stimulate interests in drug-type cannabis products, which are considered illegal in many countries. Cannabis image recognition becomes important for the proper development of Web culture.

*J. Wang is also affiliated with the Robotics Institute, Carnegie Mellon University.

According to the Wikipedia, the cannabis leaves are "palmately compound, with serrate leaflets"[1]. Specifically, the cannabis images on the Web has the following properties, which make the task of cannabis image recognition difficult: (1) non-uniform background; (2) different illuminations and occlusions; (3) visual similarity with some other plants in the leaf shape; and (4) the shape deformation of the leaves. Examples are in Fig.1.

Figure 1. Images of cannabis plants collected from illicit drug websites.

(a) (b)

Figure 2. A cannabis leaf and the detected SIFT features.

Three discriminative properties can be utilize in the computerized recognition (as illustrated in Fig. 2(a)): (1) the edge of leaflets is serrated; (2) the shape of leaflets is subulate; and (3) the leaflets are self-similar. How do we effectively encode these properties so that a computer can detect images containing such plants? This is the question we attempt to answer here.

[1] `http://en.wikipedia.org/wiki/Cannabis`

The shape structures of cannabis leaves have strong self-correlations in a cannabis image. This can also be called self-similarity [6]. We can assume that within a cannabis image, many structures are similar to each others. Based on this assumption, we introduce a *self-similarity based weak classifier* (referred to as SSWC).

We propose an AdaBoost-based algorithm for cannabis image recognition. In our algorithm, each feature from positive images can form an SSWC. In each round of boosting, a rapid weak classifier finder (WCF) is used to search for the weak classifier with the lowest error rate in the weak classifier space. All the selected weak classifiers are used to train a strong classifier for cannabis image recognition.

The rest of paper is organized as follows. Related work is discussed in Section 2. In Section 3, we introduce the feature we used and the AdaBoost-based algorithm for recognition. Experimental results are reported in Section 4. We conclude in Section 5.

2 Related Work

Cannabis image recognition can be treated as a special case of object class recognition. Recently, much work has been done in the field of object class recognition. Fergus et al. [1] proposed a probability model for objects built as flexible constellations of parts. They used an EM algorithm to learn the parameters of this model and achieved good accuracy. Csurka et al. [7] proposed the idea of bag of keypoints. Training features were clustered to construct a codebook. Then the number of patches assigned to each cluster was counted to get a bag of keypoints, which could be used to train a multi-class classifier. This method is limited when dealing with images that have non-uniform backgrounds, for large number of local features from the background can distract the clustering. Viola and Jones [2] introduced the AdaBoost model for face detection. Discriminative features were automatically selected by AdaBoost from a large set of background features. But the Haar features can be inefficient in other object recognition. Opelt et al. [3] improved this framework to general object class recognition and yielded results comparable to [1]. However, their method was designed for general object class recognition and didn't have the ability to measure the self-similarity of cannabis images. Zhang et al. [5] proposed a two layer AdaBoost model of object class recognition. Local features and global features were boosted in Layer 1, and the spatial features were boosted in Layer 2. They showed that combining different types of features can significantly improve the performance of classifiers.

3 The Recognition Algorithm

3.1 Feature Extraction

Recently, many local descriptors have been proposed in the literature [4, 8, 9, 10]. Among these, the Scale Invariant Feature Transform (SIFT) [4] has been widely used in object class recognition and detection because of its invariance to image scale and rotation, and its robustness to illumination, affine distortion, and changes in 3D viewpoint. We consider SIFT as an appropriate feature to describe serrate edge and subulate shape. Fig. 2 shows the SIFT features of a leaf. Each arrow represents a SIFT feature, with the length of the arrow stands for the scale.

Since leaflets have similar shape structures, each feature in one leaflet is similar to the feature of another leaflet in the corresponding location. The features within a leaflet also have strong similarities because the leaflet is bilateral symmetric. We consider the self-similarity a discriminative property for cannabis image recognition.

3.2 The Boosting Process

Let $\{I_i, i = 1, \ldots, N\}$ be a set of training images and $\{l_i, i = 1, \ldots, N\}$ be their associated labels, where N is the number of training images. Each image I_i has a set of SIFT features $\{F_{i,j}, j = 1, \ldots, n_i\}$, where n_i is the number of features of I_i. The set $\{f_k, k = 1, \cdots, K\}$ is a set of features from positive (cannabis) training images, where K is the total number of features from positive training images. The main idea is to select effective features from training set, and combine them to get a stronger classifier. The weights of the positive training images are initialized to $\frac{1}{2N_p}$ and those of the negative images to $\frac{1}{2N_n}$, where N_p and N_n are the numbers of positive and negative samples respectively. Next, one feature is selected by WCF for each round of AdaBoost, together with threshold(s) to construct a weak classifier h_t. After that, the weights of training samples are calculated and updated for the next round.

The AdaBoost-based algorithm for each round t of boosting is given as follows:

- Normalize the weights of training samples such that $\sum_{i=1}^{N} w_{i,t} = 1$.
- A WCF is used to automatically find a weak classifier, with the lowest classification error at the situation of $\{w_{i,t}, i = 1, \ldots, N\}$. The classification error is defined as : $\varepsilon_k = \frac{1}{2} \sum_{i=1}^{N} w_{i,t} |h_k - l_i|$.
- Update the weights of training samples: $w^*_{i,t} = w_{i,t} \beta_t^{1-|h_t - l_i|}$, where $\beta_t = \sqrt{\frac{\varepsilon}{1-\varepsilon}}$.

After T round of boosting, the resulting strong classifier is :

$$H = \begin{cases} 1, & \sum_{t=1}^{T} \alpha_t h_t > \Omega \\ -1, & otherwise \end{cases}$$

where $\alpha_t = log(1/\beta_t)$ and threshold Ω can be adjusted to get a desired detection rate.

3.3 The Self-Similarity Based Weak Classifier

Motivated by the the self-similarity property of a cannabis image, we propose a *'self-similarity based weak classifier'* (SSWC) to measure the level of self-similarity. The process is as follows. Let $d(\cdot,\cdot)$ be the Euclidean distance of the SIFT features. The SSWC consists of a feature f_k from positive images, a distance threshold θ_d, and a number threshold θ_n. If a feature $F_{i,j}$ satisfies the function $d(F_{i,j}, f_k) < \theta_d$, we say that $F_{i,j}$ is similar to this weak classifier. Our self-similarity based weak classifier is defined as:

$$h(f_k, \theta_d, \theta_n) = \begin{cases} 1, Card(\{F_{i,j} | d(F_{i,j}, f_k) < \theta_d, j = 1, \ldots, n_i\}) \geq \theta_n \\ -1, \quad otherwise \end{cases}$$

where $Card(\{\cdot\})$ is the cardinality of a set. $Card(\{F_{i,j} | d(F_{i,j}, f_k) < \theta_d, j = 1, \ldots, n_i\})$ stands for the number of features in $\{F_{i,j}, j = 1, \ldots, n_i\}$ that satisfies the function $d(F_{i,j}, f_k) < \theta_d$. This implies we not only concern the existence of similarity of features and the classifier, but also how many these features are. An image can be accepted by the weak classifier only when it has enough features that are similar to the weak classifier.

The two thresholds θ_d and θ_n in SSWC determine a large weak classifier space. It's impossible to use the exhaustive WCF in [3]. Therefore we introduce a rapid WCF with an optimized search strategy.

3.4 Rapid Weak Classifier Finder

The rapid WCF contains two steps:

- **Pre-processing**: For each feature f_k of positive training samples, do the following.
  1. For all images I_i, calculate the first M minimal distances $\{d_{k,i,j}, i = 1, \ldots, N, j = 1, \ldots, M\}$ between f_k and features in I_i.
  2. Let $\pi_k(1), \ldots, \pi_k(N \times M)$ be a permutation set of index such that

     $$d_{k,\pi_k(1)} \leq \cdots \leq d_{k,\pi_k(N \times M)}$$

  3. under the initial weight $\{w_{i,1}\}$ (set to $\frac{1}{2N_p}$ for positive training and $\frac{1}{2N_n}$ otherwise), search for all possible values of thresholds

     $$\{(\theta_d, \theta_n) | \theta_d = \frac{d_{k,\pi_k(s)} + d_{k,\pi_k(s+1)}}{2},$$

     $$s = 1, \ldots, N \times M - 1, \theta_n = 1, \ldots, M\},$$

     and find the lowest error rate ε_k^* and its corresponding thresholds (θ_d^*, θ_n^*), where $\varepsilon_k^* = \frac{1}{2}\sum_{i=1}^{N} w_{i,1} |h_{k,i} - l_i|$ and

     $$h_{k,i}(f_k, \theta_d^*, \theta_n^*) = \begin{cases} 1, Card(\{d_{k,i,j} | d_{k,i,j} < \theta_d^*, j = 1, \ldots, M) \geq \theta_n^* \\ -1, \quad otherwise \end{cases}$$

  After that, for each feature f_k, we obtain the best thresholds (θ_d^*, θ_n^*), weak classifier $h_k(f_k, \theta_d^*, \theta_n^*)$, and its classification result $\{h_{k,i}, i = 1, \ldots, N\}$.

- **Training**: For each round t of boosting, with the current weights $\{w_{i,t}, i = 1, \ldots, N\}$, output the classifier h_x with the lowest error rate ε_x of all weak classifiers $\{h_k | k = 1, \ldots, K\}$, where $\varepsilon_x = \frac{1}{2}\sum_{i=1}^{N} w_{i,t} |h_{x,i} - l_i|$.

The main strategy of the rapid WCF is as follows: pre-calculate the best weak classifier h_k and its threshold(s) for each feature f_k with the initial weights $w_{i,1}$; for each round of boosting, we only search in the weak classifier space $\{h_k | k = 1, \ldots, K\}$ while Opelt et al. [3] search in the space $\{h_k(\theta(s)) | k = 1, \ldots, K, s = 1, \ldots, N-1\}$.

All of the parameters are automatically selected, except for T, Ω, and M. In the experiments below, we will show the influence of different values of T and M. When $M = 1$, our SSWC is degenerated to the weak classifier of [3], while our rapid WCF is an accelerated version of the WCF in [3]. As a result, our method can be used for general object class recognition, but is more suitable for the objects with the self-similarity property.

4 Experimental Results

The dataset used for the experiments contains 1,197 cannabis images and 1,821 images of 28 other plants. 100 cannabis images and 300 images of other plants are randomly selected from this dataset for training. Bilinear interpolation is used to normalize images such that the longer side is 360 pixels.

We conduct a comparison experiment among the following methods: our method (*SSWC*($M = 5, 10, 15$) + *rapid WCF*), Opelt et al. method [3] with our rapid WCF (*Opelt's + rapid WCF*), Opelt et al. method

(*Opelt's + their WCF*), and Csurka et al. method [7]. The performance figures quoted are the ROC equal error rates [1]. For example, a figure of 85% means 85% of the cannabis images are correctly classified and 85% of the images of other plants are correctly classified.

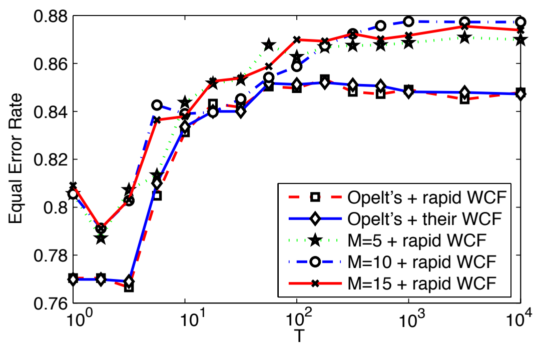

Figure 3. The equal error rates as compared to other methods for object recognition.

Fig.3 shows the equal error rates of our method, Opelt et al. method with our rapid WCF, and Opelt et al. method when the number of training rounds T increases. When T=1, equal error rates of the first weak classifiers selected by these methods achieve about 77% , and the weak classifiers selected by our method obviously outperform the one of Opelt et al. method. This is because our weak classifier contains two thresholds. Compared with the only one threshold of Opelt et al. weak classifier, a better result can be achieved by varying these thresholds.

Moreover, the curve of *Opelt's + rapid WCF* is almost the same as that of *Opelt's + their WCF*. This implies that our rapid WCF achieves similar classification accuracy as Opelt's WCF. However, each iteration of *Opelt's + rapid WCF* after pre-processing requires only 0.025 second computation time on a Pentium-4, 3.2GHz CPU against 65 seconds needed by *Opelt's + their WCF*.

From Fig. 3, we can also see that our method ($SSWC(M = 5, 10, 15)$ + *rapid WCF*) outperforms Opelt et al. method when T and M vary. When M varies, *M=10 + rapid WCF* has the best result. This is probably because a large value of M may bring large values of θ_n, which make the weak classifier high dependence on training set and low generalization; when M is too small, it may prevent WCF from choosing the best number thresholds.

Table 1 shows the best equal error rate[2] of our method ($T = 1000, M = 10, \Omega = -25.5$) compared with the ones of Opelt et al. [3] ($T = 178, \Omega = -3$) and Csurka et al. [7] ($k = 1000$ where k is the number of clusters in K-Means). This shows that our method is more accurate than the other two methods.

[2]The best equal error rate is the best rate when parameters vary.

Ours	Opelt et al. [3]	Csurka et al. [7]
87.3%	85.3%	83.5%

Table 1. The best equal error rates.

5 Conclusions and Future Work

We have presented an AdaBoost-based cannabis image recognition algorithm. As far as we know, this is the first attempt at screening image contents related to illicit drugs. Self-similarity based weak classifier has been proposed to measure the level of self-similarity. We have also introduced a rapid WCF, which can dramatically reduce the training time, with little degradation in the classification accuracy. Future work will seek for better features to effectively represent the structural information of cannabis images and to integrate this into a complete system.

Acknowledgement

This work is partly supported by NSFC (Grant No. 60672040, 60705003) and the National 863 High-Tech R&D Program of China (Grant No.2006AA01Z453). J. Wang would like to thank the K. C. Wong Education Foundation, H.K., for funding his visit to China. His research has been funded by the U.S. NSF.

References

[1] R. Fergus, P. Perona, and A. Zisserman. Object class recognition by unsupervised scale-invariant learning. In *Proc. CVPR*, 2:264-271, 2003.

[2] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Proc. CVPR*, 1:511-518, 2001.

[3] A. Opelt, M. Fussenegger, A. Pinz, and P. Auer. Weak hypotheses and boosting for generic object detection and recognition. In *Proc. ECCV*, 2:71-84, 2004.

[4] D. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91-110, 2004.

[5] W. Zhang, B. Yu, G.J. Zelinsky, D. Samaras. Object class recognition using multiple layer boosting with heterogeneous features. In *Proc. CVPR*, 2:323-330, 2005.

[6] E. Shechtman and M. Irani. Matching local self-similarities across images and videos. In *Proc. CVPR*, 2007.

[7] G. Csurka, C. Dance, L. Fan, J. Williamowski, and C. Bray. Visual categorization with bags of keypoints. In *ECCV Workshop on SLCV*, 2004.

[8] A. C. Berg and J. Malik. Geometric blur for template matching. In *Proc. CVPR*, 1:607-614, 2001.

[9] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE Trans. on PAMI*, 27(10):1615-1630, 2005.

[10] Y. Ke and R. Sukthankar. PCA-SIFT: A more distinctive representation for local image descriptors. In *Proc. CVPR*, 2004.