Available online at www.sciencedirect.com
ScienceDirect
Pattern Recognition 41 (2008) 2777–2788
PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY
www.elsevier.com/locate/pr

Applying the multi-category learning to multiple video object extraction

Yi Liu[a], Yuan F. Zheng[b,d,*], Xiaotong Shen[c]

[a]PIPS Technology, A Federal Signal Company, Knoxville, TN 37932, USA
[b]Department of Electrical and Computer Engineering, The Ohio State University, Columbus, OH 43210, USA
[c]School of Statistics, University of Minnesota, Minneapolis, MN 55455, USA
[d]School of Electronic, Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

Received 3 May 2006; received in revised form 9 January 2008; accepted 19 February 2008

Abstract

Video object (VO) extraction is of great importance in multimedia processing. In recent years approaches have been proposed to deal with VO extraction as a classification problem. This type of methods calls for state-of-the-art classifiers because the performance is directly related to the accuracy of classification. Promising results have been reported for single object extraction using support vector machines (SVM) and its extensions. Multiple object extraction, on the other hand, still imposes great difficulty as multi-category classification is an ongoing research topic in machine learning. This paper introduces a new scheme of multi-category learning for multiple VO extraction, and demonstrates its effectiveness and advantages by experiments.

Keywords: VO extraction; Multiple object tracking; ψ-Learning; Support vector machines (SVM); Multi-class classification

1. Introduction

Video object (VO) extraction, the process of segmenting and tracking semantic entities with pixel-wise accuracy [1], is an important yet challenging task for content-based video processing. For this purpose a great deal of approaches have been proposed [2–10], which provide satisfactory results for extracting VOs of homogeneous motion characteristics. Unfortunately, dealing with VOs with abrupt motions or occlusions remains a challenge. In recent years classification-based approaches have been proposed to meet the challenge by handling object tracking as a classification problem [11–13]. Each VO is considered as a class, and VO extraction is achieved by classifying every pixel to one of the available classes. By doing so, temporal associations of objects between frames are automatically maintained through correct classifications which is therefore motion-assumption free. As a result, the approaches are more robust to complicated motion fluctuations.

What learning algorithm to use is key to the success of the classification-based approaches. By using powerful classifiers high classification accuracy can be achieved which leads to better performance for VO extraction. However, most of the results reported are limited to single object scenarios. In other words, only binary classification between the object and the background has been tackled. At the first glance, the extension from single object to multiple object extraction is straightforward since conceptually one only needs to replace the binary classifier with a multi-class classifier. Unfortunately, the implementation of such an extension is far more difficult than it appears because multi-category classification is still an ongoing and immature research topic itself in machine learning. Only recently have works emerged to offer new tools that can help tackle the multi-object problem. This work presents an attempt of such.

Over the last decade, margin-based classification technologies for which the best known example is support vector machines (SVM) [14] have drawn tremendous attention due to their theoretical merits and practical success. Instead of directly estimating the conditional probabilities, the margin-based classifiers focus on the decision boundary which; however, makes it difficult to generalize their applications from binary to multi-class scenarios.

"Single machine" and "error correcting" are two mainstreams for multi-class margin-based classification. As its name

* Corresponding author. Tel.: +1 614 292 8039; fax: +1 614 292 7596.
E-mail addresses: yi.liu@pipstechnology.com (Y. Liu), zheng@ece.osu.edu (Y.F. Zheng), xshen@stat.umn.edu (X. Shen).

0031-3203/$30.00 © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2008.02.007

suggests, the "single machine" type of approaches attempts to construct a multi-class classifier by solving just a single optimization problem [15–19]. On the contrary, the "error correcting" type of approaches [20,21] works with a collection of binary classifiers, for which the primary goal is to determine what binary classifiers should be chosen to train and how to combine their classification results to make the final decision. Among all the methods published in the literature, "one-against-all", "one-against-one" and directed acyclic graph (DAG) [22] are most popular choices in solving real-world problems. A good overview of multi-class classification can be found in Refs. [23,24].

As a natural extension of binary large margin classification, the "single machine" type of approaches is intuitively appealing. It has drawn even more attention when certain formulations are reported to yield classifiers with consistency approaching the optimal Bayes error rate in the large sample limit [25]. Multi-class ψ-learning is such a learning algorithm [26]. Moreover, ψ-learning aims directly at minimizing the generalization error (GE), which is the reason why its binary version has shown significant advantage over SVM in terms of generalization both theoretically and experimentally [27]. The extended multi-class ψ-learning retains the desirable properties of its binary counterpart. In addition, a computational tool based on the recent advance in global optimization has been developed to reduce the time of training for the "single machine" [28].

The purpose of this paper is twofold. First, it introduces multi-category ψ-learning [26] to tackle the multiple VO extraction problem. Secondly, it reports the performance of the new learning algorithm on several MPEG-4 standard video sequences instead of synthetic data on which many multi-class learning algorithms are tested.

The rest of the paper is organized as follows. Section 2 gives an introduction of multi-class ψ-learning. Then a multiple VO extraction method using this new learning methodology is explained in Section 3. Section 4 provides the experimental results which are followed by conclusions in Section 5.

2. Multi-category ψ-learning

We first introduce the notations that will be used for the rest of the paper. In the framework of multi-category ψ-learning, the class label is coded as $y \in \{1, 2, \ldots, M\}$, and for a sample $x \in \mathbb{R}^d$ the decision rule is

$$y = \arg\max_{i=1,\ldots,M} f_i(x), \tag{1}$$

where M is the number of classes and f_i is the decision function of class i for $i = 1, \ldots, M$. For the linear classifier, we have $f_i(x) = w_i^{\mathrm{T}} x_i + b_i$ with $w_i \in \mathbb{R}^d$ and $b_i \in \mathbb{R}$. Conventionally, the classifier is represented as $\mathbf{f} = (f_1, f_2, \ldots, f_M)$.

As a characteristic of multi-class problems, multiple comparisons between classes need to be performed. In order to simplify the notations an $(M-1)$-dimensional vector-valued function $g(x, y)$ and a multivariate sign function $\mathrm{sign}(u)$ where $u = (u_1, \ldots, u_{M-1})$ are defined as follows:

$$\begin{aligned} g(x, y) = (&f_y(x) - f_1(x), \ldots, f_y(x) - f_{y-1}(x),\\ &f_y(x) - f_{y+1}(x), \ldots, f_y(x) - f_M(x)), \end{aligned}$$

$$\mathrm{sign}(u) = \begin{cases} 1 & \text{if } u_{\min} = \min(u_1, u_2, \ldots, u_{M-1}) \geqslant 0, \\ -1 & \text{if } u_{\min} < 0. \end{cases} \tag{2}$$

As mentioned before, the most prominent feature of ψ-learning is the direct consideration of GE. Defined as the probability of misclassification, GE yielded by an M-class classifier is

$$\mathrm{Err}(\mathbf{f}) = P\left[Y \neq \arg\max_{i=1,\ldots,M} f_i(X)\right].$$

It can be shown that with the notations of $g(x, y)$ and $\mathrm{sign}(u)$ GE can be rewritten as

$$\mathrm{Err}(\mathbf{f}) = \tfrac{1}{2} E[1 - \mathrm{sign}(g(X, Y))].$$

2.1. Multi-category ψ-learning

Seeking a vector $\mathbf{f}$ to minimize GE is the ultimate goal for any learning algorithm. For example, in the coding system described above,[1] the cost function of the well-known linear SVM can be rewritten as [26]

$$\begin{aligned} &\text{minimize} \quad \frac{1}{2}\sum_{j=1}^{2} \|w_j\|^2 + C \sum_{i=1}^{N} F_{\mathrm{SVM}}(f_{y_i}(x_i) - f_{3-y_i}(x_i)),\\ &\text{subject to} \quad \sum_{j=1}^{2} f_j(x) = 0 \quad \text{for } \forall x, \end{aligned} \tag{3}$$

where N is the number of training samples and the sum-to-zero constraint is invoked to eliminate the redundancy in (f_1, f_2). The parameter C is a regularizer that controls the relative importance between the separating margin and the training error which are reflected in the quantities $\frac{1}{2}\sum_{j=1}^{2}\|w_j\|^2$ and $\sum_{i=1}^{N} F_{\mathrm{SVM}}$, respectively. Here the so-called hinge loss $F_{\mathrm{SVM}}(u) = 0$ if $u \geqslant 1$, and $2(1-u)$ if $u \leqslant 1$ is a convex upper envelope of $F_{\mathrm{GE}} = (1 - \mathrm{sign}(u))$. However, as shown in Fig. 1(a) and (b) there is significant difference between this convex envelope and $(1 - \mathrm{sign}(u))$ itself especially when $u < 0$, which corresponds to the inevitable misclassifications in non-separable cases. Motivated by this consideration, Shen et al. [26,27] proposes to replace F_{SVM} with a non-convex ψ function as

$$\begin{aligned} &\text{minimize} \quad \frac{1}{2}\sum_{j=1}^{2} \|w_j\|^2 + C \sum_{i=1}^{N} \psi_b(f_{y_i}(x_i) - f_{3-y_i}(x_i)),\\ &\text{subject to} \quad \sum_{j=1}^{2} f_j(x) = 0 \quad \text{for } \forall x. \end{aligned} \tag{4}$$

[1] Conventionally, the formulation of SVM is expressed in the coding system where the class label $y \in \{-1, 1\}$.