

One-class Label Propagation Using Local Cone Based Similarity

Takumi Kobayashi and Nobuyuki Otsu

Abstract—In this paper, we propose a novel method of label propagation for one-class learning. For binary (positive/negative) classification, the proposed method simultaneously measures the pair-wise similarity between samples and the negativity at every sample based on a cone-based model of local neighborhoods. Relying only on positive labeled samples as in one-class learning, the method estimates the labels of unlabeled samples via label propagation using the similarities and the negativities in the framework of semi-supervised learning. In the proposed method, unlike standard label propagation methods, it is not necessary to prepare negative labeled samples since the measured negativity works as an alternative of such labeling negative samples. In experiments on target detection in still images and motion images, the proposed method exhibits the favorable performances compared to the other methods.

I. INTRODUCTION

Classifying pattern vectors is a fundamental procedure for pattern recognition and various classification methods have been developed along with the advances in machine learning, such as SVM [18] and kernel-based methods [17]. The classifier is generally learnt by using given training (labeled) samples; e.g., for two-class (binary) classification, we have to prepare both positive and negative labeled samples. Although the performance of the classifier depends on the amount of such labeled samples, the task of labeling samples by hand requires heavy human labor and thus, in practice, it is difficult to prepare plenty of labeled samples.

To avoid such difficulty, the framework of semi-supervised learning is useful. In that framework, classification, i.e., label estimation, is performed by incorporating a large amount of unlabeled samples in addition to a small number of labeled ones. The method of label propagation (LP) [20] is frequently adopted in the semi-supervised learning, such as for image matting (segmentation) [15], [7], image annotation [9] and image classification [5]. The LP method effectively integrates the labeled and unlabeled samples through pair-wise similarities defined between samples, and it copes with binary classification problems.¹

On the other hand, the method of one-class learning is also effective to reduce the required labeled samples for the binary problem, especially for a target detection problem. In the one-class learning, a certain model, such as a subspace [16] and a cone [11], is fitted to the sample distribution in the target class (positive class). Thus, we do not need to prepare the negative labeled samples for training. This eases the users to collect the labeled samples.

In this study, we propose a novel method to combine the one-class learning and semi-supervised learning. The proposed method estimates the binary class labels by using only positive labeled samples (one-class learning) and unlabeled samples (semi-supervised learning), and basically performs in the framework of the label propagation. In the proposed method, we employ the cone-based model to define the similarities in local neighborhood samples. At the same time to construct the similarities, the negativity is also measured at every sample, and it works as an alternative of labeling negative samples. By using these negativities as well as the similarities, the proposed method is applicable to one-class learning in the framework of semi-supervised learning, and is quite effective in practice since users are only required to label a few positive samples in a large amount of unlabeled

A. Related Work

We briefly mention the related methods of label propagation and those of one-class learning.

The main concern in the label propagation is how to construct effective similarities between samples. In the case that we have no prior knowledge about the samples, a model is assumed to measure the similarities. The commonly used similarity formulation is the Gaussian kernel similarity (GKS) [3] which assumes a Gaussian distribution centered at each sample. The sparsity induced similarity (SIS) [5] employs the linear model (subspace) in the local neighborhood samples, and the similarity is computed based on the coefficients in the linear model. In addition, sparseness in the similarities is explicitly induced by minimizing the L^1 -norm of the coefficients. In an image matting [15], the similarities between pixels are derived from the local linear model of alpha values on the pixels.

In the one-class learning, the distribution of the samples belonging to a target class (positive class) is approximated by using a certain model. The intuitive method is the subspace method [16] which extracts the low-dimensional subspace from the sample distribution. The subspace method is recently further developed to the cone-restricted subspace method [11]. The cone-restricted subspace method utilizes the cone-based model which imposes the non-negativity constraint on the linear subspace in order to approximate the sample distribution more precisely. The method of one-class SVM [17] is also frequently used in the one-class learning. In the one-class SVM, the hyper-plane that separates the samples from the origin as much as possible is obtained

T. Kobayashi and N. Otsu are with National Institute of Advanced Industrial Science and Technology, Umezono 1-1-1, Tsukuba, Japan {takumi.kobayashi,otsu.n}@aist.go.jp

¹In the case of multi-class problems, the problems are divided into several binary-class problems and the LP is respectively applied to them.

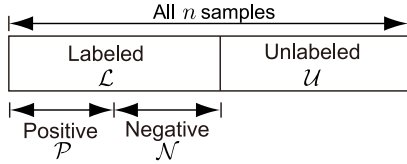


Fig. 1. All n samples are divided into labeled (\mathcal{L}) and unlabeled (\mathcal{U}) ones, and labeled samples (\mathcal{L}) are further split into positive (\mathcal{P}) and negative (\mathcal{N}) ones. The proposed method utilizes only the labels assigned to \mathcal{P} .

based on the large margin criterion in a manner similar to SVM [18].

As above, most methods have been developed along the line of either semi-supervised or one-class learning. However, the method that combines these two directions has been rarely proposed. In this study, we effectively utilize the cone-based model in the local neighborhood samples and propose a novel method for one-class semi-supervised learning. The combination of semi-supervised and one-class learning is important from the viewpoint of practice for easing users' labeling task.

II. ONE-CLASS SEMI-SUPERVISED LEARNING

In this section, we derive the method of one-class label propagation for one-class semi-supervised learning. The one-class label propagation can estimate the binary labels by using only the positive labeled samples without the negative labeled ones. We begin with the standard label propagation method which requires both the positive and the negative labeled samples, and then introduce a *negativity* as an alternative for such negative labeled samples.

A. Standard Label Propagation

First, we briefly review the method of label propagation (LP) [20]. The LP method integrates labeled and unlabeled samples for estimating binary labels in the framework of semi-supervised learning.

We consider a binary (two-class) label estimation problem and denote the label value of the i -th sample by $\alpha_i \in [0, 1]$ in which positive and negative labels are indicated as 1 and 0, respectively. Suppose, in all n samples, some of samples are labeled as positive ($\alpha = 1$) and negative ($\alpha = 0$), and the others are not assigned any labels. Let \mathcal{L} be the index set of those labeled samples and \mathcal{U} be that of the remaining unlabeled samples ($\mathcal{L} \cap \mathcal{U} = \phi$, $|\mathcal{L} \cup \mathcal{U}| = n$), as shown in Fig. 1. Given a symmetric similarity $s_{ij} = s_{ji}$ between the i, j -th samples ($1 \leq i, j \leq n$), the method of LP estimates the labels $\alpha_{\mathcal{U}}$ of the unlabeled samples by minimizing the following cost function with the fixed $\alpha_{\mathcal{L}}$:

$$\begin{aligned}
 J &= \sum_{i,j=1}^n s_{ij}(\alpha_i - \alpha_j)^2 \\
 &= \sum_{i,j \in \mathcal{U}} s_{ij}(\alpha_i - \alpha_j)^2 - 2 \sum_{i \in \mathcal{U}, j \in \mathcal{L}} s_{ij}(\alpha_i - \alpha_j)^2 + Const \\
 &= \alpha'_{\mathcal{U}}(\mathbf{D}_{\mathcal{U}\mathcal{U}} - \mathbf{S}_{\mathcal{U}\mathcal{U}})\alpha_{\mathcal{U}} - 2\alpha'_{\mathcal{U}}\mathbf{S}_{\mathcal{U}\mathcal{L}}\alpha_{\mathcal{L}} + Const, \quad (1)
 \end{aligned}$$

where \mathbf{D} is a diagonal matrix, $D_{ii} = \sum_{j=1}^n s_{ij}$. The optimum labels are obtained in a closed form:

$$\alpha_{\mathcal{U}}^* = (\mathbf{D}_{\mathcal{U}\mathcal{U}} - \mathbf{S}_{\mathcal{U}\mathcal{U}})^{-1}\mathbf{S}_{\mathcal{U}\mathcal{L}}\alpha_{\mathcal{L}}. \quad (2)$$

B. One-class Label Propagation

The above standard LP operates only when both positive and negative labeled samples are given. If the given labels are all identical, a trivial solution (uniform labeling) is produced.

In the following, we describe a role of the labeled samples, especially in a negative class, to establish one-class label propagation. Let the index sets of the positive and negative labeled samples be \mathcal{P} and \mathcal{N} , respectively ($\mathcal{P} \cup \mathcal{N} = \mathcal{L}$, see Fig. 1). By taking into account the given negative label values $\alpha_{\mathcal{N}} = 0$, Equation (2) actually results in

$$\alpha_{\mathcal{U}}^* = (\mathbf{D}_{\mathcal{U}\mathcal{U}} - \mathbf{S}_{\mathcal{U}\mathcal{U}})^{-1}\mathbf{S}_{\mathcal{U}\mathcal{P}}\alpha_{\mathcal{P}}, \quad (3)$$

and the similarities connected to the negative labeled samples \mathcal{N} appear only in \mathbf{D} :

$$D_{ii} = \sum_{j=1}^n s_{ij} = \sum_{j \in \mathcal{U} \cup \mathcal{P}} s_{ij} + \underbrace{\sum_{j \in \mathcal{N}} s_{ij}}_{\delta_i}. \quad (4)$$

Therefore, in this framework, only the summation $\delta_i \triangleq \sum_{j \in \mathcal{N}} s_{ij}$ is required without explicitly providing the negative labeled samples \mathcal{N} . This summation δ_i of the similarities connected to the negative labeled samples can be regarded as a *negativity* at the i -th sample; the negativity indicates how far the i -th sample is from the target (positive). We do not know a priori which samples are positive, but some suggestion for the negativity can be statistically given, such as by outliers. By introducing the negativity, the LP is applicable to one-class semi-supervised learning which uses only the positive labeled samples.

In the next section, we describe the proposed method that simultaneously measures the pairwise similarities s_{ij} and the negativities δ_i on the basis of a cone-based model of local neighborhoods.

III. CONE-BASED SIMILARITY AND NEGATIVITY

In the proposed method, both the similarities and the negativity are simultaneously measured at each sample by fitting a cone-based model [11] to local neighborhood samples.

A. Similarity

Let the i -th sample be represented by the feature vector \mathbf{x}_i . The commonly used Gaussian kernel similarity (GKS) [3] models the local sample distribution by a Gaussian distribution; $s_{ij} = \exp(-\gamma\|\mathbf{x}_i - \mathbf{x}_j\|^2)$. The sparsity induced similarity (SIS) [5] assumes linear model, i.e., subspace, in the local neighborhood samples:

$$\mathbf{p}_i^* = \arg \min_{\mathbf{p}_i} \|\mathbf{p}_i\|_1 \quad (\mathbf{p}_i = [p_{i1}, \dots, p_{in}]) \quad (5)$$

$$s.t. \quad \mathbf{x}_i = \sum_{j \in k\text{NN}(i)} p_{ij}\mathbf{x}_j, \quad p_{ij} = 0 \text{ for } j \notin k\text{NN}(i),$$

$$\hat{p}_{ij} = \frac{\max(p_{ij}^*, 0)}{\sum_{k=1, k \neq i}^n \max(p_{ik}^*, 0)}, \quad s_{ij} = \frac{\hat{p}_{ij} + \hat{p}_{ji}}{2}, \quad (6)$$

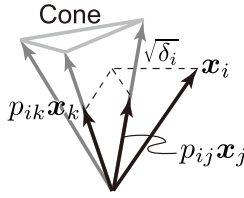


Fig. 2. The i -th sample vector \mathbf{x}_i is projected onto the cone consisting of its local neighborhoods, and the projected vector is represented by the the basis sample vectors \mathbf{x}_j and \mathbf{x}_k with the non-negative coefficients p_{ij} and p_{ik} . The similarities s_{ij} to the i -th sample are defined by using those coefficients, and the distance of the projection is utilized for measuring the negativity δ_i .

where $k\text{NN}(i)$ indicates the index set of k nearest neighbor samples around the i -th sample. In the SIS, sparsity in the similarities is induced by minimizing the L^1 -norm of the coefficients p_{ij} in the linear model.

Recently, the cone-restricted subspace method has been proposed and produced superior performances to the subspace method [11]. In that method, the sample distribution is modeled as a *cone* which imposes non-negativity constraints on the coefficients in the linear model:

$$\text{Cone} : \left\{ \mathbf{x} \mid \mathbf{x} = \sum_i p_i \mathbf{x}_i, p_i \geq 0 \text{ for } \forall i \right\}. \quad (7)$$

In this study, we apply the cone to model local neighborhood samples and then derive the similarities as follows:

$$\mathbf{p}_i^* = \arg \min_{\mathbf{p}_i} \left\| \mathbf{x}_i - \sum_{j \in k\text{NN}(i)} p_{ij} \mathbf{x}_j \right\|^2 \quad (8)$$

$$\text{s.t. } p_{ij} \geq 0 \text{ for } \forall j, p_{ij} = 0 \text{ for } j \notin k\text{NN}(i),$$

$$s_{ij} = \frac{p_{ij}^* + p_{ji}^*}{2}, \quad (9)$$

where the feature vectors are normalized to unit in L^2 norm ($\|\mathbf{x}\|_2 = 1$). Equation (8) is efficiently solved by applying the fast non-negative least square (FNNLS) method [4]. In this formulation, the i -th sample \mathbf{x}_i is approximated by using the cone composed of the neighborhood samples \mathbf{x}_j . In (8), we calculate the projection from \mathbf{x}_i onto the cone and obtain the non-negative coefficients p_{ij}^* such that the projection is represented by $\sum_{j \in k\text{NN}(i)} p_{ij}^* \mathbf{x}_j$, as shown in Fig. 2. The sample of the larger coefficient significantly contributes to the projection and thus is considered to be highly similar to \mathbf{x}_i . Therefore, the coefficients reflect the similarities between \mathbf{x}_i and its neighbors. We simply take an average of those coefficients for symmetric similarities in (9).

Due to the normalized feature vectors, the similarities are fairly measured across the neighborhood samples based on the angles between the vectors, disregarding their magnitudes. In case that the feature vector is scaled by the factor s , $\hat{\mathbf{x}}_j \leftarrow s \mathbf{x}_j$, the coefficient is also (inversely) scaled by $\hat{p}_{ij} \leftarrow \frac{1}{s} p_{ij}$ ($\because p_{ij} \mathbf{x}_j = \frac{1}{s} p_{ij} \hat{\mathbf{x}}_j$). This indicates that the samples with smaller magnitudes would have larger similarity values. To avoid such unfavorable side-effects, we disregard the magnitude information and employ the angle between sample vectors which is sufficiently discriminative [11]. In addition, the feature vectors are often (inherently) normalized, such as in histogram-based features (HOG [6] and GLAC [12]) and

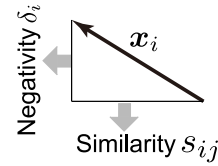


Fig. 3. The information of the i -th sample is split into the similarities along projection on the cone and the negativity based on the distance of the projection. We fully extract the information, although the most methods use only the projected information.

image vectors, so as to increase robustness to environmental changes.

As in SIS [5], the formulation defined above induces sparsity into similarities due to the non-negativity constraint without explicitly imposing a sparseness constraint. The non-zero coefficients are assigned only to the basis samples of the cone which are close to the sample \mathbf{x}_i (Fig. 2). As a result, the number of non-zero similarities is small.

B. Negativity

Next, we describe how the negativity δ_i is measured at the same time to construct the similarities. Model parameters, such as coefficients in the linear model in (5), are usually exploited to construct the similarities under the assumed model, although the deviations from the model are ignored. In this study, we make use of such deviations to measure the negativity at every sample. The deviation manifests how the sample is outlier in the (local) sample distribution. While the positive (target) samples, e.g., faces or pedestrians to detect, must obey certain distributions such as clusters, the negative (non-target) samples lie out of such distributions and could be outliers. The deviation from the cone, i.e., the negativity δ_i , is directly calculated by squared Euclidean distance:

$$\begin{aligned} \delta_i &= \min_{\mathbf{p}_i} \left\| \mathbf{x}_i - \sum_{j \in k\text{NN}(i)} p_{ij} \mathbf{x}_j \right\|^2 \\ &= \sum_{j, k \in k\text{NN}(i)} p_{ij}^* p_{ik}^* \mathbf{x}_j' \mathbf{x}_k - 2 \sum_{j \in k\text{NN}(i)} p_{ij}^* \mathbf{x}_j' \mathbf{x}_i + 1 \\ &= 1 - \sum_{j \in k\text{NN}(i)} p_{ij}^* \mathbf{x}_j' \mathbf{x}_i. \end{aligned} \quad (10)$$

where we use $\sum_{j, k \in k\text{NN}(i)} p_{ij}^* p_{ik}^* \mathbf{x}_j' \mathbf{x}_k = \sum_{j \in k\text{NN}(i)} p_{ij}^* \mathbf{x}_j' \mathbf{x}_i$ corresponding to the squared norm of the projected vector from \mathbf{x}_i onto the cone.

As shown in Fig. 3, the information that the sample contains is fully exploited by both the similarities (along the projection) and the negativity (along the perpendicular to the projection). By using these similarities and negativities, all of which are derived from the cone-based model in the local neighborhood, we can perform one-class semi-supervised learning via label propagation in (3) and (4).

C. Kernel-based Method

In the above descriptions, we mentioned the method in the (linear) space of feature vectors. The formulation in (8) is easily extended to the non-linear case via kernel tricks. We introduce the non-linear functions $\phi(\mathbf{x}_i)$ such that $k(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)' \phi(\mathbf{x}_j)$. The cone-based model in

the local neighborhood can also be defined in the kernel space [14]:

$$\begin{aligned} \mathbf{p}_i^* &= \arg \min_{\mathbf{p}_i} \|\phi(\mathbf{x}_i) - \sum_{j \in k\text{NN}(i)} p_{ij} \phi(\mathbf{x}_j)\|^2 \quad (11) \\ &= \sum_{j, k \in k\text{NN}(i)} p_{ij} p_{ik} k(\mathbf{x}_j, \mathbf{x}_k) - 2 \sum_{j \in k\text{NN}(i)} p_{ij} k(\mathbf{x}_i, \mathbf{x}_j) + 1 \\ &\text{s.t. } p_{ij} \geq 0 \text{ for } \forall j, \quad p_{ij} = 0 \text{ for } j \notin k\text{NN}(i), \end{aligned}$$

where we assume a normalized kernel function ($k(\mathbf{x}, \mathbf{x}) = 1$). Most kernel functions are inherently normalized, such as in RBF kernel $k(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2)$. This is also solved by the FNNLS method [4]. Thus, the similarities and the negativities are calculated as is the case with the above-mentioned linear method, by substituting the kernel function $k(\mathbf{x}_i, \mathbf{x}_j)$ for the inner product $\mathbf{x}_i' \mathbf{x}_j$.

By using the kernel version of the proposed method, the one-class label propagation is applicable in the case that not the explicit form of the feature vectors but only the kernel function is obtained, such as in graph kernel [10] and string kernel [8].

IV. EXPERIMENTAL RESULTS

First, we illustrate how the proposed method works on a toy problem. Then, we apply the proposed method to practical problems of one-class learning; face/pedestrian detection and specific action detection. In those tasks, only a few positive labels for targets are provided and the proposed method estimates the binary labels (target or non-target) in the framework of semi-supervised learning.

A. Toy Problem

We collected the samples that are concentrically distributed on a unit sphere in the three-dimensional space, as shown in Fig. 4. Only two samples around the center of the distribution are labeled as positive. The proposed method favorably estimates the label values; the circular distributed samples around the center are classified as positive and, meanwhile, the samples of the surrounding ring distribution are classified as negative (Fig. 4(a)). The negativities are also shown in Fig. 4(b), and the samples that lie out of the distribution in negative class have high negativity values. Thus, based on such negativities, the proposed method can estimate the labels correctly by using only a few positive labeled samples.

B. Object Detection

In object detection problems, the samples of the target object belong to a positive class, while the non-target samples are regarded as negative. This corresponds to a binary class classification.

1) *Experimental Setup*: We compared the proposed method to several other methods of one-class learning and of semi-supervised learning; one-class SVM [17], cone-restricted subspace method [11], and the LP methods using GKS [3] and SIS [5]. Whilst the proposed method (one-class LP), one-class SVM and cone-restricted subspace utilize only positive (target) labeled samples, the GKS and SIS

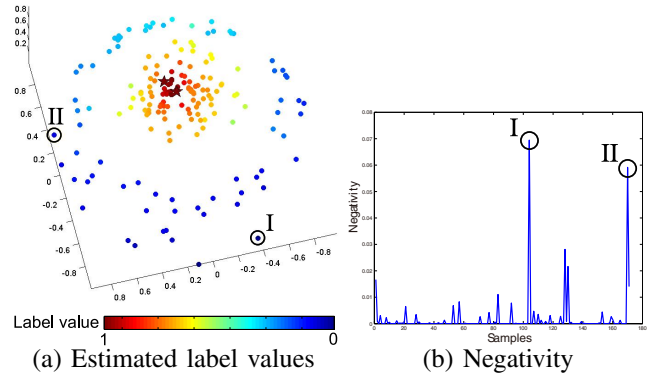


Fig. 4. The proposed method is applied to toy samples on a unit sphere in the three-dimensional space. (a) Only two samples denoted by star points are labeled positive samples and the estimated label values ($[0, 1]$) are denoted by the pseudo colors. (b) The measured negativities are shown for all unlabeled samples. The high negativity values are produced at the negative samples that lie out of the distribution. This figure is best viewed in color.

methods require negative (non-target) labeled samples as well as the positive labeled ones. In these experiments, the number of the negative labeled samples is the same as that of the positive ones. The parameters in those methods are empirically determined as follows: In the proposed method and SIS, the number of nearest neighbors k is determined as half of the dimensionality of feature vectors. The ν parameter in one-class SVM is set to 0.5, and the γ parameter in GKS is determined as the inverse of the sample variance. We also applied the kernel version of the proposed method (one-class kernel LP), one-class kernel SVM [17] and cone-restricted kernel subspace method [14]. In these kernel-based methods, we employed the Gaussian kernel for image vectors in Sec.IV-B.2 and the chi-square kernel [19] for histogram-base feature vectors in Sec.IV-B.3 and Sec.IV-B.4. The parameter value of the kernel is simply determined based on the sample variance in the same manner to GKS. All these methods are implemented by using MATLAB on 3.3GHz Xeon PC; we use L^1 -magic toolbox [2] for SIS and mex-C for the proposed method.

The performances are measured by the DET (Detection Error Trade-off) curve based on the false positive and false negative rates which contains the same information as the ROC curve. The AUC (area under the curve) of the ROC curve is computed for evaluation as well.

2) *Face Detection*: We utilized MIT face dataset [1] to evaluate the face detection performance of the proposed method. The dataset contains 2,429 face images (19×19 pixels) for training and 472 face and 23,573 non-face images for test (see Fig. 5(a)). The image vector ($\mathbf{x} \in \mathbf{R}^{361}$) was simply extracted as the feature vector. The DET curves are shown in Fig. 5(b), and the proposed method exhibits superior performances to the others. The kernel version of the proposed method is inferior to the linear version, indicating that the cone-based model is not necessarily proper in this kernel (nonlinear) space, and the kernel function employed in this experiment is based on GKS, which might degrade the performance. Then, we reduced the number of the positive

(face) labeled samples by ranging the ratio of the positive samples from 0.1 to 1. The AUC-based scores (1–AUC) are calculated for respective cases, as shown in Fig. 5(c). The proposed method produced much favorable performances across all amounts of the labeled samples, compared to the other methods. Even when the small amount of the positive labeled samples is given, the proposed method works well by effectively incorporating unlabeled samples via the similarities. In addition, the computation time for calculating similarities is significantly low compared to SIS, as shown in Fig. 5(d).

3) *Pedestrian Detection*: Next, the methods were applied to a pedestrian detection task according to the same experimental protocol as above. We used INRIA person dataset [6], containing 2,416 person images (128×64 pixels) for training and 1,132 person and 13,590 non-person images for test (see Fig. 6(a)). The GLAC features [12] ($x \in \mathbf{R}^{3888}$) were extracted from those images. In this case, due to the computational issue, the number of nearest neighbors k in the method of SIS is set to $k = 500$ and the proposed method with such parameter value was additionally applied for fair comparison. The performance results are shown in Fig. 6(b) and Fig. 6(c), demonstrating that the proposed method is superior to the others as in Sec.IV-B.2. The proposed method can effectively perform even for the GLAC features which have high dimensionality but are highly discriminative.

4) *Action Detection*: Finally, we applied the methods to detect a specific action from motion images. The CHLAC motion features [13] ($x \in \mathbf{R}^{1251}$) were employed to characterize the action (motion). We conducted the experiment on detecting gymnastics sport actions in horizontal bar and parallel bar events. The motion images (320×240 pixels) were captured by a static camera in gymkhana. We collected nine sequences and 17 sequences in parallel bar and in horizontal bar events, respectively, containing the target acrobatic actions to detect as well as the other non-target actions, as shown in Fig. 7(a). The performances were evaluated in the following scheme; for training, we randomly selected only three sequences (one in parallel bar and two in horizontal bar events) which contain about seven acrobatic actions, i.e., seven positive labeled samples, and for test, we drew the samples extracted from the remaining sequences (roughly speaking, there are 26 positive acrobatic samples and 1,500 negative samples). The target actions (positive samples) are sub-sequences of about 20 frames, and similarly we extracted sub-sequences of 20 frames for negative samples. This trial was repeated three times and the averaged performances were reported. Fig. 7(b,c) show the DET curve and the AUC-based evaluation scores. The proposed method produces clearly superior performances even in such case of quite a few labeled positive samples.

In the above all experimental results, the linear version of the proposed method is superior to the other methods of one-class learning and semi-supervised learning, and requires a little computation time. These results show that the similarities and the negativities derived from the cone-

based model of local neighborhoods in linear feature space are effective in the framework of one-class semi-supervised learning. However, the kernel version yields inferior performances. The cone-based model could be improper in the kernel (nonlinear) space, although the performance might be improved by carefully choosing the kernel function and tuning the parameter values in the kernel function.

V. CONCLUSION

We have proposed a novel method of one-class semi-supervised learning. Based only on positive labeled samples, the proposed method estimates the class labels, incorporating unlabeled samples via similarities, without using negative labeled samples. In the method, the similarities between samples are computed by exploiting the cone-based model in local neighborhoods, and at the same time, the negativity is measured at every sample. The negativities work as alternatives for labeling negative samples, which frees us from collecting those samples. In the various experiments on detecting targets, the proposed method produced the favorable performances compared to the other methods of one-class learning and of semi-supervised learning.

The proposed method is quite effective in practice as users are only required to label a few positive samples in a lot of unlabeled samples, and is so general as to be applicable to the other applications, such as object tracking in the semi-supervised framework.

REFERENCES

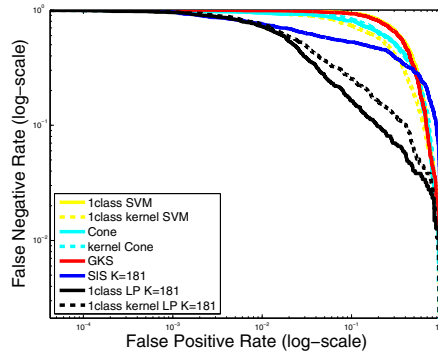
- [1] CBCL Face Database, www.ai.mit.edu/projects/cbcl.
- [2] L1-magic. <http://www.acm.caltech.edu/l1magic/>.
- [3] M. Belkin and P. Niyogi. Laplacian eigenmaps and spectral techniques for embedding and clustering. In *Advances in neural information processing systems*, pages 346–358, 2002.
- [4] R. Bro and S. Jong. A fast non-negativity-constrained least squares algorithm. *Journal of Chemometrics*, 11:393–401, 1997.
- [5] H. Cheng, Z. Liu, and J. Yang. Sparsity induced similarity measure for label propagation. In *International Conference on Computer Vision*, 2009.
- [6] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 20–25, 2005.
- [7] L. Grady, T. Schiwietz, and S. Aharon. Random walks for interactive alpha-matting. In *Visualization, Imaging and Image Processing*, 2005.
- [8] H.Lodhi, C.Saunders, J.Shawe-Taylor, N.Cristianini, and C.Watkins. Text classification using string kernels. *Journal of Machine Learning Research*, 2:419–444, 2002.
- [9] F. Kang, R. Jin, and R. Sukthankar. Correlated label propagation with application to multi-label learning. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2006.
- [10] H. Kashima, K. Tsuda, and A. Inokuchi. Marginalized kernels between labeled graphs. In *International Conference on Machine Learning*, 2003.
- [11] T. Kobayashi and N. Otsu. Cone-restricted subspace methods. In *International Conference on Pattern Recognition*, 2008.
- [12] T. Kobayashi and N. Otsu. Image feature extraction using gradient local auto-correlations. In *European Conference on Computer Vision*, pages 346–358, 2008.
- [13] T. Kobayashi and N. Otsu. A three-way auto-correlation based approach to motion recognition. *Pattern Recognition Letters*, 30(3):185–192, 2009.
- [14] T. Kobayashi, F. Yoshikawa, and N. Otsu. Cone-restricted kernel subspace methods. In *International Conference on Image Processing*, 2010.
- [15] A. Levin, D. Lischinski, and Y. Weiss. A closed form solution to natural image matting. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 30(2):228–242, 2008.



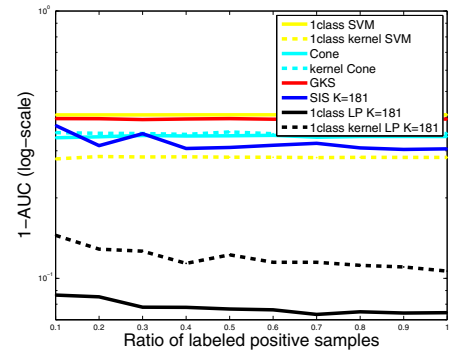
(a) Dataset

(d) Similarity comp. time per sample (msec)

SIS K=181	66.2
1class LP K=181	2.6



(b) DET curve



(c) 1-AUC

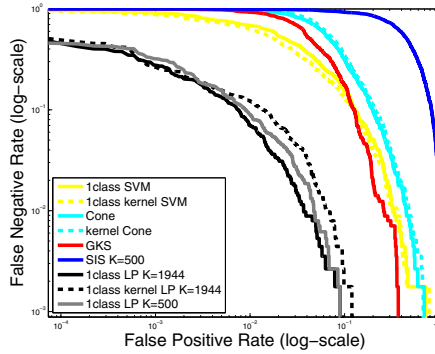
Fig. 5. Performances for face detection compared to the other methods. (a) MIT face dataset [1]. (b) DET (Detection Error Trade-off) curve in log-scale, based on the false positive and false negative rates. The lower-left curve indicates a better performance. (c) AUC-based evaluation scores (1-AUC) in log-scale for various ratios of the positive labeled samples. The lower value indicates a better performance. (d) Similarity computation time per sample (millisecond). These figures are best viewed in color.



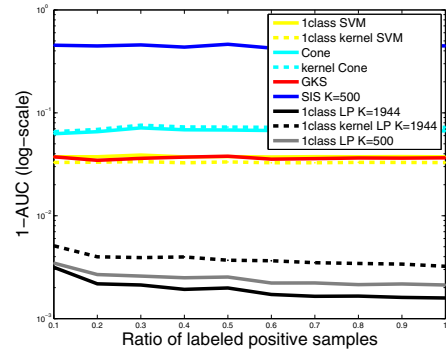
(a) Dataset

(d) Similarity comp. time per sample (msec)

SIS K=500	1227.1
1class LP K=500	6.3
1class LP K=1944	63.6

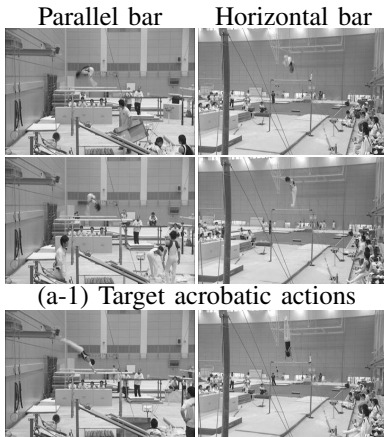


(b) DET curve



(c) 1-AUC

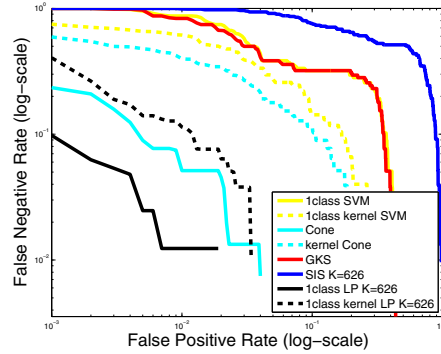
Fig. 6. Performances for pedestrian detection compared to the other methods. (a) INRIA person dataset [6]. (b) DET (Detection Error Trade-off) curve in log-scale. (c) AUC-based evaluation scores (1-AUC) in log-scale for various ratios of the positive labeled samples. (d) Similarity computation time per sample (millisecond). These figures are best viewed in color.



(a-1) Target acrobatic actions

(a-2) Non-target actions

(a) Dataset



(b) DET curve

(c) 1-AUC

1class SVM	0.115505
1class kernel SVM	0.051159
Cone	0.001572
kernel Cone	0.032906
GKS	0.112777
SIS K=626	0.451276
1class LP K=626	0.000373
1class kernel LP K=626	0.002574

(d) Similarity comp. time per sample (msec)

SIS K=626	1819.3
1class LP K=626	2.6

Fig. 7. Performances for action detection compared to the other methods. (a) Examples of actions in parallel and horizontal bar events. Target acrobatic actions jumping out of bars are positive (a-1) and the other non-target actions are negative (a-2). (b) DET (Detection Error Trade-off) curve in log-scale, averaging all three random trials. (c) Averaged AUC-based evaluation scores (1-AUC) in log-scale. (d) Similarity computation time per sample (millisecond). These figures are best viewed in color.

[16] E. Oya. *Subspace methods of pattern recognition*. Research Studies Press, 1984.
 [17] B. Schölkopf and A. Smola. *Learning with Kernels*. MIT Press, 2001.
 [18] V. Vapnik. *Statistical Learning Theory*. Wiley, 1998.
 [19] J. Zhang, M. Marszalek, S. Lazebnik, and C. Schmid. Local features and kernels for classification of texture and object categories:

A comprehensive study. *International journal of computer vision*, 73(2):213–238, 2007.
 [20] X. Zhu, Z. Ghahramani, and J. Lafferty. Semi-supervised learning using gaussian fields and harmonic functions. In *International Conference on Machine Learning*, 2003.