# One-Class Multiple Instance Learning via Robust PCA for Common Object Discovery

Xinggang Wang[1]*, Zhengdong Zhang[2], Yi Ma[2], Xiang Bai[1],
Wenyu Liu[1], and Zhuowen Tu[2,3]

[1]Huazhong University of Science and Technology
[2]Visual Computing Group, Microsoft Research Asia
[3]Lab of Neuro Imaging and Department of Computer Science, UCLA
{wxghust,zhangzdfaint}@gmail.com, mayi@microsoft.com,
{xbai,liuwy}@hust.edu.cn, ztu@loni.ucla.edu

**Abstract.** Principal component analysis (PCA), as a key component in statistical learning, has been adopted in a wide variety of applications in computer vision and machine learning. From a different angle, weakly supervised learning, more specifically multiple instance learning (MIL), allows fine-grained information to be exploited from coarsely-grained label information. In this paper, we propose an algorithm using the robust PCA (RPCA) [1] in a iterative way to perform simultaneous common object discovery and model learning under a one-class multiple instance learning setting. We show the advantage of our method on common object discovery and model learning, which needs no fine/coarse alignment in the input data; in addition, it achieves comparable results with standard two-class MIL learning algorithms but our method is learning from one-class data only.

## 1 Introduction

Principal component analysis (PCA) has been adopted in a wide variety of domains [2], enjoying its simplicity and effectiveness. A robust principal component analysis model (RPCA) [1] was recently proposed along the line of increasingly popular sparsity and robust measures (e.g. the $\ell_1$ norm) [3]. Unlike the $\ell_2$ norm used in the standard PCA approach, RPCA encourages a low-rank part in the data matrix while having the $\ell_1$ norm on the residual, allowing the robust handling of data corruption and missing entries. The general assumption of PCA and RPCA though depends on well-aligned input data. However, this requirement is often too strong, especially for data of high dimension, which is particularly problematic in computer vision; for example, even well-studied frontal faces are hard to be perfectly aligned due to their intrinsic ambiguity. A so-called robust alignment by sparse and low-rank decomposition (RASL) algorithm [4] was very recently developed based on RPCA to deal with the local transformation/alignment. However, RASL only works on image data with small deformations; it is hard to apply RASL in more general cases without high quality initializations.

* This work was done while the author was an intern in Microsoft Research Asia.

From a different angle, weakly supervised learning, more specifically multiple instance learning (MIL) [5–8], allows fine-grained information to be exploited from coarsely-grained label supervision. In MIL, a training set consists of many bags (images in our case); each bag consists of a number of instances (patches in our case); only bag-level labels are given in training; the instance-level labels are therefore unknown in the training stage; the training algorithm then automatically explores instance-level and bag-level models to best fit the given bag labels. One promising aspect of MIL is that it allows for the automatic model learning and instance-level label prediction at the same time. In the end, a discriminative classifier is learned with the simultaneous label predictions on the instances. Thus, MIL seems to be on the complementary side of PCA and RPCA in removing the restrictions on having well-aligned input data. However, existing MIL methods are mostly focused on learning discriminative models requiring both the positive and negative data; essentially, the instance–level labels for the negative bags are known to us since we assume the presence of positive instances only in the positive bags. Here, we assume no given negative bags and we want to learn a PCA-like generative model for the instance-level data of interest; this represents many practical situations which are hard to be handled by the existing MIL methods.

Another recent active research area in computer vision is unsupervised/weakly-supervised object discovery [8–11]. However, the existing approaches either separate the task of object discovery from model learning or are formulated in a standard MIL setting. Different the other approaches try to discovery multi-class objects, e.g. [10], we focus on *common* object discovery. Thus, we requires all images come from the same class; no negative/irrelevate images are needed.

In this paper, we propose a new algorithm using robust principal component analysis (RPCA) to perform simultaneous object discovery and model learning within a one-class multiple instance learning framework. In the experiments, we show the advantage of our method on several applications to discover e.g. frontal faces of large variations; it also achieves comparable results as the standard two-class MIL learning algorithms with models learned from one-class data only.

## 2 Related Work

A robust principal component analysis (RPCA) was recently proposed in [1] for video surveillance and face recognition; there has been also immediate work adopting RPCA: further optimization approach was engaged to enhance the results of RPCA [12]; in [4], robust alignment by sparse and low-rank decomposition (RASL) was applied for face alignment. RASL aims to align multiple images of an object class of interest to a canonical template and it assumes that the degree of initial misalignment is not too large. In our problem, as stated before, we allow for objects of unknown locations and scales with possibility in severe occlusions.

Multiple instance learning (MIL) has recently received a lot of attentions. The diverse density (DD) method [13] tackles MIL by finding regions in the instance space with instances from many different positive bags and few instances from negative bags. In [6] learning algorithm of DD is refined using expectation maximization (EM). MI-SVM and MILBoost are proposed in [14] and [8] in which they train SVM and boosting classifier for instances respectively. Our method only models the positive instances without the negative bags. We use a EM-like algorithm to learn our generative model which is similar to [6]. Similar to MI-SVM and MIL-Boost, our model maintains a latent selection of most positive instance with a bag. However, our model is generative studying the instances of interest directly and explicitly.

Object discovery is a recent active research area [15–20]; although their results on benchmark datasets are promising, these existing methods are for specific purposes built with complicated systems. Here, we focus on a simple but general framework to discover objects and learn a PCA model from images known to contain an object class of interest. Therefore, we only focus on rigid objects which can be modeled by a PCA-like model. It alleviates the burden in having negative bags, as required by many MIL approaches. Our method shows its particular robustness in handling occlusions and outliers. Other methods such as 'co-segmentation' method in [21] and the detecting and sketching the common method in [22] do not require negative images for detection. However, they work on two (or a few) images only with no explicit model learning in an integrated framework.

## 3 Notation and Problem Formulation

In this section, we first give a brief introduction to the notation that will be used throughout this paper. Then a detailed discussion about the formulation of our problem will follow.

### 3.1 Notation

Suppose we are given $N$ bags of instances. Each instance is represented by a $d$-dimensional vector $x \in \mathbb{R}^d$, and the $k$-th bag contains $n_k$ instances. We name all the instances for the $k$-th bag as $x_1^k, x_2^k, \ldots, x_{n_k}^k$, and by putting them together we get a representing matrix $X_k = \left[x_1^k, x_2^k, \ldots, x_{n_k}^k\right] \in \mathbb{R}^{d \times n_k}$ for each of the bags. Each instance $x_i^k$ belongs to either the positive or the negative category. So we label it with a binary variable $z_i^k \in \{0, 1\}$, where $z_i^k = 1$ indicates positiveness and vice versa. Each bag is also associated with a binary label $Z_k$ based on the labels of its instances: $Z_k = \bigvee_{i=1}^{n_k} z_i^k$. Intuitively speaking, a bag is positive if and only if some of its instances is positive.

For convenience we define a new operator $x \circ z$ as follows:

$$x \circ z = \begin{cases} x & \text{if } z = 1 \\ 0 & \text{otherwise} \end{cases}$$

Moreover, we generalize this operation to the bag level:

$$X^k \circ Z^k = \left[x_1^k \circ z_1^k, \ldots, x_{n_k}^k \circ z_{n_k}^k\right].$$

Following the convention, $\|\cdot\|_*$ stands for nuclear norm of a matrix(sum of the singular values), and $\|\cdot\|_1$ means $l_1$-norm(sum of the magnitude of entries) for both vectors and matrices, $\|\cdot\|_0$ counts the number of non-zero entries in a vector and matrix. Moreover, $[n]$ denotes the set of positive integers less than or equal to $n$: $\{1, 2, \ldots, n\}$.

### 3.2 One-Class Multiple Instance Learning via Robust Rank Minimization

As has stated in Section 1, traditional settings of the Multiple Instance Learning problem requires both positive bags and negative bags to be available. Also in the training stage we must know exactly which bags are positive and which are not. In this paper, we will study how to tackle this challenging problem under a totally different setting. Basically it is assumed that we only have access to the positive bags, *without* any touch on the rest negative bags. Specifically, in our notation, $\forall k \in [N]$, we have $Z_k = \bigvee z_i^k = 1$.

Hence, by throwing away negative bags, we also disable ourselves from seeking discriminative information to separate positive and negative bags. Therefore, to make the problem tangible, some special assumptions on the intrinsic structure of positive and negative bags must be made. Below is the one of our choice.

> *Assumption 1: All the positive instances lie in a subspace $\Omega$ with extremely low dimensionality. Meanwhile, all the negative instances lies in another high-dimensional subspace that is incoherent with $\Omega$.*

This assumption is in fact pretty reasonable in practice. For example, let us examine the scenario of single common object discovery in images. If we align the common objects together, they actually form a rank 1 subspace $\Omega$. Background patches and other uncommon objects naturally lie on another subspace which, compared with $\Omega$, is of much higher dimensionality, since they are by definition uncommon between images.

Under this assumption, we have turned our task into the following form:

> *From each bag, pick out several positive instances, such that when we put all these instances together as a whole into a matrix, that matrix is of the lowest-rank possible.*

Mathematically, we are trying to solve the following optimization.

$$\min_{z_i^k \in \{0,1\}} rank\left([X_1 \circ Z_1 | X_2 \circ Z_2 | \ldots | X_N \circ Z_N]\right) \quad s.t. \quad \forall k \in [N], Z_k = \bigvee_{i=1}^{n_k} z_i^k = 1 \tag{1}$$

For simplicity, we abbreviate $([X_1 \circ Z_1 | X_2 \circ Z_2 | \ldots | X_N \circ Z_N])$ into $X \circ Z \in R^{d \times (n_1 + \ldots + n_N)}$. Unfortunately, even though the ground-truth positive instances may satisfy this strict low-rank assumption, the observed versions of them seldom meet this requirement. One cause of this is due to quantization errors, changes on illumination, noise and even occlusions. Apart from these, a small fraction of the positive instances may turn out to be wrongly labeled, i.e., they come from negative categories. To handle these in a uniform framework, we model all of the corruption and outliers as sparse error added to the clean data. In other words, the observation $X \circ Z$ is a superposition of a low-rank component $L$ and sparse error matrix $S$:

$$X \circ Z = L + S$$

Here $\Omega = span(L)$. In the following sections of the paper, by a slight abuse of notation, we will not distinguish between $\Omega$ and $L$. Thus (1) is reshaped into:

$$\min_{z_i^k \in 0,1,L,S} rank(L) + \lambda_0 \|S\|_0, \quad s.t. \;\; X \circ Z = L + S, \forall k \in [N], Z_k = \bigvee_{i=1}^{n_k} z_i^k = 1 \quad (2)$$

$\lambda_0$ here is a weight that balances the low-rankness of $L$ and the sparsity of $S$.

## 4 Solution via Iterative Robust PCA

Notice that the highly combinatorial nature of (2) on binary variables $z_i^k \in \{0, 1\}$ makes it difficult to tackle. So we borrow the idea of iterative minimization from k-means to design an approximate solution. Specifically we would like to fix the guess of instance labels $Z$ and estimate the low-dimensional subspace $L$ despite corruption $S$. Then with the estimated $L$ and $S$ we update the instance labels $Z$ under certain strategy. We keep iterating the above two steps until convergence. The algorithm is summarized in Algorithm 4.2.

### 4.1 Estimate the Low-rank Subspace by Robust Principal Component Analysis

With $Z$ fixed, the constraints of (2) is already linear with respect to $L$ and $S$. So we just need to address the non-convex function $rank(\cdot)$ and $\|\cdot\|_0$. As proposed in [1], replacing the intangible operator $rank$ and $\|\cdot\|_0$ with their convex surrogate nuclear norm($\|\cdot\|_*$) and $l_1$-norm($\|\cdot\|_1$) actually will not affect the global optimal solution under mild conditions. Based on this fact, we transform (2) into the following form:

$$\min_{L,S} \|L\|_* + \lambda \|S\|_1, \quad s.t. \;\; X \circ Z = L + S \qquad (3)$$

Notice that here $\lambda = 1/\sqrt{\min(d, N)}$ guarantees the global convergence to the desired solution under reasonable assumptions [1]. This convex optimization problem exactly obeys the form of the Robust PCA and can be solved efficiently utilizing the Augmented Lagrangian Multiplier method proposed in [12].

### 4.2 Update Instance Labels through $l_1$ Regression

Once the low-rank subspace $L$ is retrieved, to update the guess of labels of each instance $x$, we need to test how well $x$ fits into the subspace $L$. This can be measured by the $l_1$ regression error $e$ of $x$ over $L$ , which is defined as follows:

$$e = \min_w \|x - Lw\|_1 \qquad (4)$$

This regression can also get efficiently solved via [23]. Because each bag contains at least one positive instance, we sort the instances $x_i^k$ by $e_i^k$ in an ascent order, pick the best $\rho$ instances to be positive and set the rest negative.

**Algorithm 1** Iterative RPCA for One-Class Multiple Instance Learning

**Input**: Positive bags $X$, initialized instance labels $Z_0$, weight $\lambda$, parameter $\rho$.
**Initialize**: $L = 0$, $Z = Z_0$.
**While** not converged **Do**
  **Step 1.** Fix labels $Z$, update $L$ via Robust PCA:
    $(L^*, S^*) \leftarrow \arg\min_{L,S} \|L\|_* + \lambda\|S\|_1, \quad s.t. \quad X \circ Z = L + S.$
    $L \leftarrow L^*$
  **Step 2.** Update the label based on $L$.
    **For** each bag $X_k$
      **For** each instance $x_i^k$ in $X_k$
        Get the reconstruction error by $l_1$-regression:
          $e_i^k = \min_w \|x_i^k - Lw\|_1$
      **EndFor**
      **If** $e_j^k$ is within the $\rho$-th smallest among all $e_i^k, i \in [n_k]$,
        Set $z_j^k = 1$
      **Else**
        Set $z_j^k = 0$
      **EndIf**
    **EndFor**
**EndWhile**
**Output**: The learned low-dimensional subspace $L$ and the instance labels $Z$.

### 4.3 Implementation Details

*Construction of bags/instances* In the common object discovery task, we run saliency detection method in [24] on all the images to get a set of salient patches, each with a score indicating the saliency degree. Each image is considered as a bag, and the salient patches detected by saliency detector described using HoG feature in [25] are considered as instances; number of instances is determined by the output of [24].

*Initialization of Labels* Different strategies applies to different scenarios. For the task of common object discovery in images, we choose the patch with the highest saliency score to be positive and set the rest negative. For other tasks such as multiple instance learning on existing online published datasets, the saliency based method could not apply since we only have access to the well-prepared instance points and bags. In this situation, we just randomly pick out a few instances from each bag as positive. Then we turn to RANSAC, repeating the estimation independently a few times and selecting out the best model. Often the bags in these datasets do not contains many instances, thus this random initialization strategy has a fairly large chance of success provided repeated enough times.

*Choice of $\lambda$ in* (3) Although $\lambda = 1/\sqrt{\min(d, N)}$ has already given (3) a lot of nice properties, in practice we sometimes still need to tune it to further improve the results. For instance, in single common object discovery in images, we lower $\lambda$ to $1/\sqrt{2\min(d, N)}$ to make sure that $rank(L) = 1$. However, setting $\lambda$ out of the range $\left[1/2\sqrt{\min(d,N)}, 2/\sqrt{\min(d,N)}\right]$ will not make the algorithm produce anything meaningful at least empirically.

*Choice of* $\rho$ For single common object discovery, since the positive subspace has the property that $rank(L) = 1$, $\rho = 1$ is definitely the best choice. And typically $\rho = 1$ will not make the algorithm go wrong in most of the situations. However, if there are multiple objects that presents simultaneously in the same image or the common object is represented by multiple instances that almost do not overlap, then we have to set $\rho$ to larger values. In the experiments on MIL benchmark datasets, we don't know number of positive instances in each bag, so we find the best value of $\rho$ by running cross validation on training data.

## 5 Experiments

In this part, we carry out the object discovery experiments on image datasets and test our RPCA-based one-class MIL algorithm on standard MIL benchmark. As the baseline of comparison, we would like to slightly change our method in Algorithm 4.2 by replacing the Robust PCA component to classical PCA which is not robust to corruptions but is optimal provided no outliers exist. By slight abuse of notation, we denote the original algorithm using RPCA by *RPCA-based learning method* (for short, RPCA method) and the modified version is named *PCA-based learning method* (for short, PCA method). To compare PCA with RPCA fairly, we set the number of projection dimension of PCA to the rank of $L$ in RPCA. In the following experiments, we will demonstrate the advantages of RPCA method over the PCA method. We also compare RPCA method to other related state-of-the-art methods. We do not aim at developing a system to over-perform the state-of-the-art methods. Instead, we just want to highlight that RPCA model truthfully reflect the existing outliers or corruptions that is massively existing in the data of real world.

### 5.1 Occluded Face Discovery

We collect a face image dataset which contains 50 face images with many occluded faces at different sizes from web and the LFW image dataset [26]. Some of the images in the dataset are shown in Fig. 1. As is shown there, faces are occluded by different kinds of objects, ranging from sunglass, tennis, to hands etc. Aside from this, expressions on faces and background of faces in images also vary a lot.

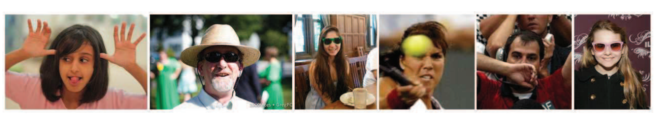

**Fig. 1.** Some face images in the face image dataset.

In Fig. 2, we show the image patches for initialization (in the first column among each group), face discovery results of PCA method (in the second column

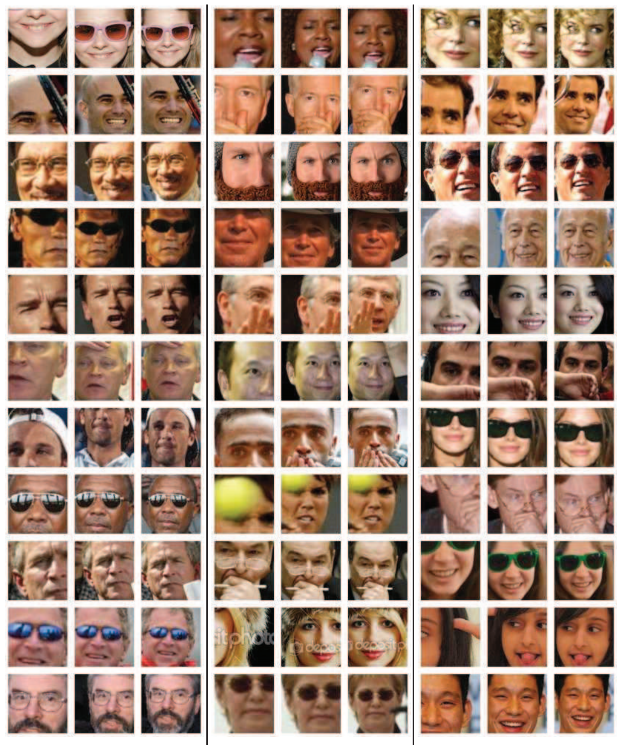

**Fig. 2.** Face discovery results: the first column among each group shows image patches for initialization, the second column shows results of PCA method, the third column shows results of RPCA method.

among each group) and face discovery results of RPCA method (in the third among each column) for 33 of all 50 images. Fig. 2 shows that image patches used for initialization are extremely challenging. Only part of the faces are present in each patch. What's worse is that even the present patches are not consistent across different images. Notice that here we do not use the raw pixels but rather extract some HoG features from each patch to represent every instance. As is observed in Fig. 2: Faces discovered by PCA method are not well aligned, most of which shift away from centers due to occlusion; while RPCA method align these discovered faces pretty well. Quantitative results are in Table 1 which also shows that RPCA can significantly outperform PCA in this occluded face discovery experiment.

Our conclusion of this experiment is that RPCA method outperforms PCA method on this occluded face discovery task due to the fact that RPCA in [1] is designed to handle large sparse error on data, and in this case, the large sparse error corresponds to the occlusion on face images.

Fig. 3 visualizes the learned low-rank subspace in HoG feature space in every iteration for both PCA method and RPCA method. The visualization method is from [25]. It shows that RPCA method can iteratively get the sketch of face, while PCA method can not converge to a good face model.

**Table 1.** The overlap percentages between ground-truth and initialized box, predicted box by PCA, and predicted box by RPCA.

| | Initialization | PCA | RPCA |
|---|---|---|---|
| Overlap with ground-truth | 40.94% | 65.88% | **79.28%** |

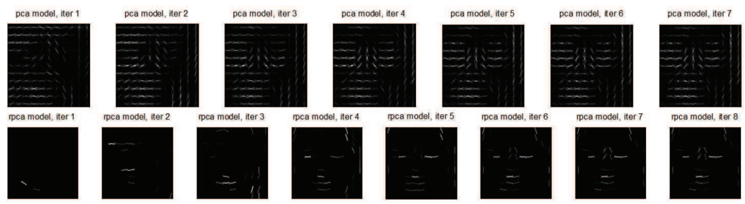


**Fig. 3.** Visualization of the PCA model (above) and the RPCA model (below) in every iteration. For better viewing, please see the original pdf file.

### 5.2 Common Object Discovery on ETHZ Dataset

In this experiment, we use RPCA method and PCA method for object discovery on the challenging ETHZ dataset [27] which is widely used for supervised objection detection. We perform object discovery on the *applelogos* and *bottles* classes separately. There are 40, 48 images in the *applelogos* and *bottles* classes respectively. Images in the two classes have significant intra-class variation, scale change, and illumination difference; some of images have very clustered background. Because HoG template cannot handle large deformation in the other three classes in ETHZ dataset, we don't work on them.

A discovered window is correct if it intersects with a groundtruth object by more than half of their union (PASCAL criteria). Object discovery performance is evaluated by 1) precision-recall curves, generated by varying the score threshold, 2) average precision (AP), computed by averaging multiple precisions corresponding to different recalls at regular internals and 3) detection rate against the number of false-positives averaged over all images with the class (FPPI).

We first compare RPCA method to PCA method, and the salient object detection (SD) method in [24]. Precision-recall curves and average precision in Fig. 4 illustrate the performance of RPCA method, PCA method, and SD method. Both RPCA method and PCA method outperform SD method significantly, and RPCA method works better then PCA method. We then compare RPCA method to a supervised object detection method [28] in which half of the images with bounding boxes in each class are used for training. Detection rates at 0.3/0.4 FPPI of [28] and RPCA method are listed in Table 2. It illustrates

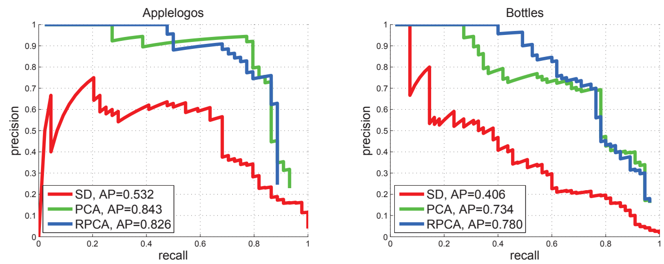


**Fig. 4.** Precision-recall curves for RPCA method (in blue), PCA method (in green) and SD method in [24] (in red) on ETHZ *applelogos* class (left) and *bottles* class (right).

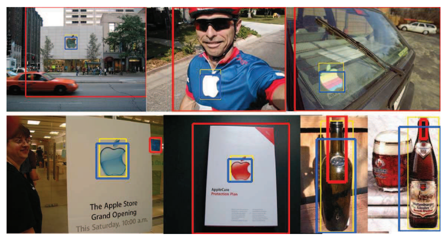


**Fig. 5.** The most confident detection hypothesises given by SD method in [24] (in red) and RPCA method (in blue), groundtruth objects are in yellow on ETHZ dataset.

that RPCA method is comparable to the supervised object detection method [28] on *applelogos* and *bottles* classes. Fig. 5 shows the most confident detection hypothesis given by SD method and RPCA method in some of images in the dataset. As shown in this figure, using the objects in red boxes as initialization, our RPCA method can iteratively find the true object locations marked in blue. The salient object detection result on the other three classes of ETHZ dataset are too bad, so we have not tested the performance of the proposed method on the other three classes.

**Table 2.** Comparison of detection rates of the supervised object detection method [28] and RPCA method at 0.3/0.4 FPPI on ETHZ *applelogos* class and *bottles* class.

| classes | *applelogos* | *bottles* |
|---|---|---|
| Ferrari et al. [28] | 0.777/0.832 | 0.798/0.816 |
| RPCA | 0.800/0.864 | 0.709/0.763 |

### 5.3 Classification on MIL Benchmark Dataset

Until now, we have demonstrated a lot about the power of Robust PCA method for solving one class Multiple Instance Learning Problem without any information about negative bags. In this experiment, we will show that utilizing the learned model, with simple modification, our method can actually do the same two classes bag classification tasks. Moreover, we will show that indeed this simple modification would grant our algorithm with similar performance compared with the popular discriminative MIL method, e.g., the mi-SVM method.

Specifically, suppose we have learned the low-rank subspace $L$ for positive instances from the given positive training data. But after this we now have additional access to some other negative bags. Utilizing the new negative bags and $L$, we can train a SVM classifier as follows: Upon each bag $X_k$, no matter its positive or negative, for each instance $x_i^k$ in this bag, we build a histogram $h_k$ to show the distribution of the $l_1$ reconstruction error $e_i^k$, and use $h_k$ as the final representation for $X_k$. Then we train a simple linear SVM classifier using $h_k$ as training bags to accomplish the bag classification task. To compare the proposed RPCA method with standard two-class MIL learning algorithms, we evaluate RPCA method on four benchmark datasets [14] that are very popularly in studies of multiple instance learning, including *Musk1, Elephant, Fox and Tiger*. For each dataset, first we use the random initialization strategy described in the previous sections to set up the algorithm. Then RPCA model and PCA model are learned only using positive bags according to Algorithm 4.2. Then classifiers are trained based on all these. Following the standard verification convention, experiments are performed in a 10-fold cross-validation manner and per-fold average test classification performance is reported in Table 3.

**Table 3.** Results on MIL benchmark datasets. Bag classification accuracies (%) of RPCA method and PCA method on four MIL benchmark datasets compared to the state-of-the-art. The results of the upper part are taken from respective papers.

| Datasets | *Musk1* | *Elephant* | *Fox* | *Tiger* |
|---|---|---|---|---|
| MI-SVM [14] | 77.9 | 81.4 | 59.4 | 84.0 |
| mi-SVM [14] | 87.4 | 82.0 | 58.2 | 78.9 |
| EM-DD [6] | 84.8 | 78.3 | 56.1 | 72.1 |
| PPMM Kernel [29] | **95.6** | 82.4 | 60.3 | 80.2 |
| MIGraph [30] | 90.0±3.8 | 85.1±2.8 | 61.2±1.7 | 81.9±1.5 |
| miGraph [30] | 88.9±3.3 | **86.8±0.7** | 61.6±2.8 | **86.0±1.6** |
| MI-CRF [7] | 87.0 | 85.0 | **65.0** | 79.5 |
| PCA | 85.7±1.4 | 73.0±1.5 | 60.8±1.4 | 75.8±2.0 |
| RPCA | 82.9±2.8 | 78.3±1.1 | 61.0±1.4 | 76.9±0.9 |

In Table 3, we have compared RPCA method to PCA method, some popular MIL methods and the state-of-the-art methods, such as [14, 6, 29, 30, 7] are also listed. RPCA method outperforms PCA method in 3 of the 4 datasets (mark in red color), which shows that RPCA method is more practical than PCA method in general data. Only positive bags are used for learning the model in our proposed RPCA method. However, the performance of RPCA method is comparable to discriminative mi-SVM method in [14]. This good property makes the proposed RPCA method can be more widely used, such as unsupervised object detection without any negative training images in section 5.1. In the state-of-the-art face detection approach in [31], it need about 10000 non-face images for training.

## 6 Conclusion and Future Work

In this paper we proposed a new one-class multiple instance learning method based on Robust PCA [1] without negative bags. The algorithm achieves comparable robustness to both corruption on data and wrongly categorized instances, thus can work in some situations that PCA doesn't work well. We also show that with slight modification our method can achieve comparable performance to some popular methods that leverage discriminative information. In the future, we will develop composition model for object representation, rather than the current simple HoG template, to discover more complex objects in images.

*Acknowledgement:* The work was supported by NSF CAREER award IIS-0844566, NSF award IIS-1216528, and by the National Natural Science Foundation of China (NSFC) Grants 60903096, 61173120 and 61222308.

## References

1. Candes, E., Li, X., Ma, Y., Wright, J.: Robust principal component analysis? Journal of the ACM **58** (2011)
2. Jolliffe, I.T.: Principal component analysis. Springer-Verlag (1986)
3. Candes, E., Tao, T.: Near-optimal signal recovery from random projections: universal encoding strategies. IEEE Trans. Inform. Theory **52** (2005) 5406–5425
4. Peng, Y., Ganesh, A., Wright, J., Xu, W., Ma, Y.: Rasl: Robust alignment by sparse and low-rank decomposition for linearly correlated images. In: CVPR. (2010) 763–770
5. Dietterich, T.G., Lathrop, R.H.: Solving the multiple-instance problem with axis-parallel rectangles. Artificial Intelligence **89** (1997) 31–71
6. Zhang, Q., Goldman, S.A.: Em-dd: An improved multiple-instance learning technique. In: Advances in Neural Information Processing Systems, MIT Press (2001) 1073–1080
7. Deselaers, T., Ferrari, V.: A conditional random field for multiple-instance learning. In: Proceedings of the 26th International Conference on Machine Learning. (2010)
8. Viola, P., Platt, J.C., Zhang, C.: Multiple instance boosting for object detection. In: Advances in Neural Information Processing Systems, MIT Press (2006) 1419–1426
9. Russell, B.C., Efros, A.A., Sivic, J., Freeman, W.T., Zisserman, A.: Using multiple segmentations to discover objects and their extent in image collections. In: IEEE Conference on Computer Vision and Pattern Recognition. (2006)
10. Lee, Y.J., Grauman, K.: Shape discovery from unlabeled image collections. In: IEEE Conference on Computer Vision and Pattern Recognition. (2009)
11. Deselaers, T., Alexe, B., Ferrari, V.: Localizing objects while learning their appearance. ETHZ TR No 276, Eidgenossische Technische Hochschule Zurich (2011)
12. Lin, Z., Chen, M., Wu, L., Ma, Y.: The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. UIUC Technical Report UILU-ENG-09-2215 (2009)

13. Maron, O., Lozano-Prez, T.: A framework for multiple-instance learning. In: Advances in Neural Information Processing Systems, MIT Press (1998) 570–576
14. Andrews, S., Tsochantaridis, I., Hofmann, T.: Support vector machines for multiple-instance learning. In: Advances in Neural Information Processing Systems, MIT Press (2003) 561–568
15. Fergus, R., Perona, P., Zisserman, A.: Object class recognition by unsupervised scale-invariant learning. In: IEEE Conference on Computer Vision and Pattern Recognition. (2003)
16. Chum, O., Zisserman, A.: An exemplar model for learning object classes. In: IEEE Conference on Computer Vision and Pattern Recognition. (2007)
17. Vijayanarasimhan, S., Grauman, K.: Keywords to visual categories: Multiple-instance learning for weakly supervised object categorization. In: IEEE Conference on Computer Vision and Pattern Recognition. (2008)
18. Lee, Y.J., Grauman, K.: Object-graphs for context-aware category discovery. IEEE Transactions on Pattern Analysis and Machine Intelligence. TPAMI (2011)
19. Zhu, L., Lin, C., Huang, H., Chen, Y., Yuille, A.L.: Unsupervised structure learning: Hierarchical recursive composition, suspicious coincidence and competitive exclusion. In: European Conference on Computer Vision. (2008)
20. Wu, Y.N., Si, Z., Gong, H., Zhu, S.C.: Learning active basis model for object detection and recognition. International Journal of Computer Vision **90** (2010) 198–235
21. Rother, C., Minka, T.P., Blake, A., Kolmogorov, V.: Cosegmentation of image pairs by histogram matching - incorporating a global constraint into mrfs. In: IEEE Conference on Computer Vision and Pattern Recognition. (2006) 993–1000
22. Bagon, S., Brostovski, O., Galun, M., Irani, M.: Detecting and sketching the common. In: IEEE Conference on Computer Vision and Pattern Recognition. (2010)
23. Yang, A., Ganesh, A., Sastry, S., Ma, Y.: Fast l1-minimization algorithms and an application in robust face recognition: A review. Technical Report UCB/EECS-2010-13, EECS Department, University of California, Berkeley (2010)
24. Feng, J., Wei, Y., Tao, L., Zhang, C., Sun, J.: Salient object detection by composition. In: International Conference on Computer Vision. (2011)
25. Felzenszwalb, P., Girshick, R., McAllester, D., Ramanan, D.: Object detection with discriminatively trained part based models. IEEE Transactions on Pattern Analysis and Machine Intelligence **32** (2010)
26. Huang, G.B., Ramesh, M., Berg, T., Learned-Miller, E.: Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst (2007)
27. Ferrari, V., Tuytelaars, T., Gool, L.J.V.: Object detection by contour segment networks. In: European Conference on Computer Vision. (2006)
28. Ferrari, V., Jurie, F., Schmid, C.: From images to shape models for object detection. International Journal of Computer Vision **87** (2010) 284–303
29. Wang, H., Yang, Q., Zha, H.: Adaptive p-posterior mixture-model kernels for multiple instance learning. In: Proceedings of the 26th International Conference on Machine Learning. (2008)
30. Zhou, Z., Sun, Y., Li, Y.: Multi-instance learning by treating instances as non-i.i.d. samples. In: Proceedings of the 26th International Conference on Machine Learning. (2009)
31. Viola, P., Jones, M.: Robust real-time face detection. International Journal of Computer Vision **57** (2004) 137–154