2014 IEEE International Conference on Robotics & Automation (ICRA)
Hong Kong Convention and Exhibition Center
May 31 - June 7, 2014. Hong Kong, China

Simultaneous Prototype Selection and Outlier Isolation for Traffic Sign Recognition: A Collaborative Sparse Optimization Method

Huaping Liu, Yulong Liu, Yuanlong Yu and Fuchun Sun

Abstract— **Video-based traffic sign recognition is one of the most important task for unmanned autonomous vehicle. However, there always exists unavoidable outliers in the practical scenario. Therefore, robust prototype extraction from the noisy sample set is highly expected to help traffic sign recognition in video sequence. In this paper, we propose a novel approach for simultaneous prototype extraction and outlier isolation through collaborative sparse learning. The new model accounts for not only the reconstruction capability and the sparsity, but also the robustness. To solve the optimization problem, we adopt the Alternating Directional Method of Multiplier (ADMM) technology to design an iterative algorithm. Finally, the effectiveness of the approach is demonstrated by experiments on GTSRB dataset.**

I. Introduction

The traffic sign recognition from video sequence is a multi-class classification problem that has become a real challenge for unmanned autonomous vehicle. However, most of the existing work focus on recognizing the traffic sign from a single image frame (please refer to [17] for state-of-the-art). Very recently, Ref.[19] utilized the supervised low-rank decomposition technology to exploit the temporal information and obtained promising results. However, it did not consider the influence of outlier image frame and the redundancy which exists in the image sequence (see Fig.1 for an example). In practice, to saving computational time and memory requirements, the simultaneous prototype selection and outlier isolation for traffic sign recognition is highly expected. This motivates us to investigate the automatic prototype extraction, which aims to select a small set of the representatives for a specific sample set. Although prototype extraction has been extensively studied during the past years, how to effectively select certain samples while preserving the essential characterization of the original sample set and isolating the outliers is still a challenging problem. Recently, sparse coding[2][3][20] was successfully utilized for prototype extraction in [5] and [7], which utilized the self-expressiveness property of the samples. An important merit of such methods is that they can produce desired number of prototypes and supply ranked output prototypes, which is highly expected for practical applications. As a result, it does not incur additional complexity cost when changing configurations such as the number of the prototypes[5].

Because the traffic sign video usually includes outliers, a method that robustly finds prototypes is of particular importance, as it reduces the redundancy of the data and removes points that do not really belong to the data set. In [7], the authors discussed how their sparse modeling method can deal with outliers and robustly find prototypes for data sets. Their method is based on the fact that outliers are often incoherent with respect to the collection of the true data. They defined the row-sparsity-index of each candidate prototype to detect the outlier, and a threshold is required to be prescribed. But this index is very sensitive to noise and even the outlier sample may be selected as the prototypes. In addition, when we perform the sparse coding, the outlier is not isolated, and therefore the obtained solution may be seriously influenced. This will further deteriorate the extraction of prototypes and the isolation of the outliers.

Fig. 1. An example of a track in GTSRB dataset (all images are normalized to the the same size). The 1-15th images are shown in the first row and the 16-30th images are shown in the second row. From this example we can see that the consecutive images in one single track are indeed similar. In addition, the content changes dramatically at the 2th image due to the reason that the camera is bumped strongly and recovers at once.

In this paper, we address the problem of robust prototype extraction from noisy sample set. The main contributions are listed as follows:

1) A collaborative sparse coding model is proposed to describe the representativeness and the robustness, which are formulated into row sparsity and column sparsity optimization problems, respectively. Compared with existing work[7], the prototype extraction and outlier isolation can be performed simultaneously. In [7], the sparse coding is first performed, and a so-called RSI value is calculated for each sample to judge if it is outlier or not. Since the two stages are performed individually, the obtained solution is sub-optimal.
2) The original non-convex model is relaxed to be convex model using $||\cdot||_{\infty,1}$ and $||\cdot||_{1,\infty}$ norms. An iterative optimization algorithm is proposed to solve the optimization problem. Since the model is convex, the solution can be obtained efficiently.
3) We use the proposed model on the GTSRB dataset to reliably find some outliers and obtain promising recognition performance.

The rest of the paper is organized as follows: Section 2

Huaping Liu, Yulong Liu and Fuchun Sun are with Department of Computer Science and Technology, Tsinghua University, State Key Laboratory of Intelligent Technology and Systems, TNLIST, Beijing, P.R.China. E-mail: hpliu@tsinghua.edu.cn. Yuanlong Yu is with College of Mathematics and Computer Science, Fuzhou University, P.R.China.

reviews some related work. In Section 3, we present the problem formulation and the proposed model. Then, we introduce the optimization algorithm in Section 4. Subsequently, we describe the experiments and evaluation results by comparing our method to the state-of-the-art. Finally, we summarize the work of this paper.

Notations. Let $\mathbf{M} \in R^{r\times c}$. We use superscripts for the rows of $\mathbf{M}$, i.e., $\mathbf{M}^{(i)}$ denotes the i-th row; and subscripts for the columns of $\mathbf{M}$, i.e., $\mathbf{M}_{(j)}$ denotes the j-th column. We will use various matrix norms, here are the notations we use: $||\mathbf{M}||_F$ is the Froubenius norm, which is also equal to $\sqrt{Tr(\mathbf{M}^T\mathbf{M})}$; $||\mathbf{M}||_{2,1}$ is the sum of the $\mathcal{L}_2$ norm of the rows of $\mathbf{M}$: $||\mathbf{M}||_{2,1} = \sum_{i=1}^{r} ||\mathbf{M}^{(i)}||_2$; and $||\mathbf{M}||_{1,2}$ is the sum of the $\mathcal{L}_2$ norm of the columns of $\mathbf{M}$: $||\mathbf{M}||_{1,2} = \sum_{j=1}^{c} ||\mathbf{M}_{(j)}||_2$. Similarly, $||\mathbf{M}||_{\infty,1}$ and $||\mathbf{M}||_{1,\infty}$ are defined as $||\mathbf{M}||_{\infty,1} = \sum_{i=1}^{r} ||\mathbf{M}^{(i)}||_\infty$ and $||\mathbf{M}||_{1,\infty} = \sum_{j=1}^{c} ||\mathbf{M}_{(j)}||_\infty$, respectively.

II. Related work

A straightforward application of prototype extraction is the key-frame selection, which plays important roles for robotics. Ref.[6] used the selected key-frames to form a visual path and a 2D visual servo step allowed the mobile robot to navigate from its current position to the next key image. In [23], key-frames constituted the images amongst the sequence for which the pose was computed and were used to estimate the geometry of the scene. In Ref.[13], the robot detected loop closures for simultaneous localization and mapping[10] by matching data frames against a subset of previously collected frames. In addition, key-frame is very useful to construct 3D model and augmentation reality systems. Ref.[16] selected key-frames for doing full bundle adjustment for augmented-reality applications. Ref.[21] selected a smaller number of frames to form a skeletal set that still spanned the whole dataset and produced reconstructions of comparable accuracy. Ref.[15] described a parallel tracking and mapping system which simultaneously applied full bundle adjustment to key-frames selected from a video stream, while performing robust real-time pose estimation on intermediate frames. In all of these applications, the key-frame served as the representatives of the corresponding dataset and played important role to dramatically reduce the computation burden. Key-frame selection is the basis of video summarization. Very recently, Ref.[25] proposed a strategy to summarize the wireless capsule endoscopy video clips. Ref.[11] investigated the problem of summarizing observations made by a mobile robot on a trajectory. Such a work was extended to on-line scenario in [12]. Ref.[18] described how a subset of images can be selected to summarise the robot's cumulative visual experience. All of these work do not consider the influence of outliers.

III. Problem Formulation

The prototypes can be regarded as a set consists of a collection of representatives extracted from the underlying data set. Therefore, prototype extraction is equivalent to how to select an optimal subset from the entire data set under certain constraints. Consider a matrix $\mathbf{B} = [\mathbf{b}_1, \mathbf{b}_2, \cdots, \mathbf{b}_N] \in R^{d\times N}$, where each column vector denotes a sample represented as a feature vector. The task is to find an optimal subset $\bar{\mathbf{B}} = [\mathbf{b}_{i_1}, \mathbf{b}_{i_2}, \cdots, \mathbf{b}_{i_n}] \in R^{d\times n}$, where $i_1, i_2, \cdots, i_n \in \{1, 2, \cdots, N\}$, such that the inliers of the original set can be approximately reconstructed and the outliers of the original set can be isolated. In addition, the value of n is expected to be as small as possible. Therefore, the extracted prototypes should be representatives of all of the samples, i.e., the cost to use such prototypes to reconstruct the whole data sample set should be small. That is to say, for any $k \in [1, N]$, the sample $\mathbf{b}_k$ should be well approximated by

$$\mathbf{b}_k \doteq w_{k,i_1}\mathbf{b}_{i_1} + w_{k,i_2}\mathbf{b}_{i_2} + \cdots + w_{k,i_n}\mathbf{b}_{i_n} \tag{1}$$

where $w_{k,i_1}, \cdots, w_{k,i_n}$ are the reconstruction coefficients for $\mathbf{b}_k$.

To characterize this reconstruction capability, the following cost should be minimized to use such prototypes to reconstruct the whole sample set:

$$\min_{\bar{\mathbf{B}},\bar{\mathbf{W}}} ||\mathbf{B} - \bar{\mathbf{B}}\bar{\mathbf{W}}||_F^2, \tag{2}$$

where $\bar{\mathbf{W}} \in R^{n\times N}$ is the coefficient matrix.

The problem is that the coefficients $\bar{\mathbf{W}}$, as well as the index set $\{i_1, i_2, \cdots, i_n\}$, are unknown. Hence, one starts by using all columns of $\mathbf{B}$ to describe $\mathbf{B}$ itself, i.e., $\mathbf{W} \in R^{N\times N}$ should be found to satisfy

$$\mathbf{B} = \mathbf{B}\mathbf{W}.$$

Such an optimization problem is of no sense since a trivial solution $\mathbf{W} = \mathbf{I}$ can always be obtained. Fortunately. a prior that the number of the selected prototypes should be as small as possible can be utilized.

Our intuition is that we want few atoms to participate in the approximation, but we want each atom to contribute to as many columns of the sample matrix as possible. In other words, most rows of the coefficient matrix should be zero, but the nonzero rows should have many nonzero entries. Therefore, a straightforward approach is to minimize the following objective function,

$$\min_{\mathbf{W}} ||\mathbf{W}||_{\text{row}-0} + \lambda||\mathbf{B} - \mathbf{B}\mathbf{W}||_F^2, \tag{3}$$

where $\mathbf{W} \in R^{N\times N}$ is the pursuit coefficient matrix; the term $||\mathbf{B}-\mathbf{B}\mathbf{W}||_F^2$ is used to evaluate the reconstruction error; and the parameter λ is used to balance different penalty terms. The symbol $||\mathbf{W}||_{\text{row}-0}$ counts the number of nonzero rows of $\mathbf{W}$. By adding such a term into the objective function, the trivial solution can be avoided and the obtained solution will be row sparse, i.e., most of its rows are zero vectors.

In the above model (3), the sparsity is imposed on the rows of the matrix $\mathbf{W}$, while the reconstruction error is evaluated using the Frobenious norm. It is well known that such a reconstruction error is easily influenced by the outliers. Ref.[7] proposed a row sparsity index to detect the outlier after the sparse coding stage. Such a method has two major

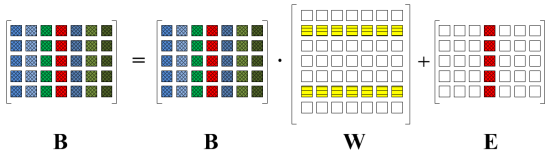

Fig. 2. Illustration of the proposed collaborative sparse coding method. $\mathbf{B}$ is composed of 7 samples of 5 dimensions. The textured squares represent the data elements in $\mathbf{B}$; The blank squares represent zero elements and yellow squares represent the non-zero elements. The row-sparsity of $\mathbf{W}$ helps to extract the prototypes (the 2nd and the 6th samples) and the column-sparsity of $\mathbf{E}$ helps to isolate the outliers (the 4th sample).

shortcomings: First, since the outlier detection is performed after the sparse coding stage, the obtained coding vector itself is influenced by the outliers and therefore may be inaccurate. Second, the proposed row sparsity index is sensitive to the noise (as we shall see in the experimental section).

To isolate the outlier, we exploit the fact that the outliers should not be well reconstructed by the prototypes and therefore the corresponding reconstruction error should be large. That is to say, if the i-th sample is outlier, then the $i-$th column of the reconstruction error matrix $\mathbf{B}-\mathbf{BW}$ will be non-zero and may admit large values. Otherwise, if the i-th sample is inlier, then the $i-$th column of the reconstruction error matrix $\mathbf{B}-\mathbf{BW}$ will be close to zero. On the other hand, it is frequently observed that the outliers are usually a minority in the sample set, i.e., the number of the outliers is usually not too large. We make use of this property and formalize it as the column sparsity regularization term

$$||\mathbf{B}-\mathbf{BW}||_{\text{column}-0},$$

which counts the number of the non-zero column of $\mathbf{B}-\mathbf{BW}$.

Therefore we modify the original model (3) as:

$$\begin{aligned}&\min_{\mathbf{W},\mathbf{E}} ||\mathbf{W}||_{\text{row}-0}+\lambda||\mathbf{E}||_{\text{column}-0}\\ &\text{s.t.}\quad \mathbf{B}-\mathbf{BW}=\mathbf{E}.\end{aligned} \tag{4}$$

In such a model, the row-sparsity is imposed on $\mathbf{W}$ to detect the prototypes and the column sparsity is imposed on the error matrix $\mathbf{E}$ to isolate the outliers. By this way, the extracted prototypes will focus on reconstructing the inliers only. This method is illustrated in Fig.2. Since the two sparsity optimization problems are simultaneously solved, we call it as collaborative sparsity.

However, such an approach is of little practical use, since the optimization problem is NP-hard as its solution requires a combinatorial search which grows faster than polynomial as the dimension N grows. A natural alternative is to use the $\mathcal{L}_{\infty,1}$-norm and $\mathcal{L}_{1,\infty}$-norm to replace the $||\cdot||_{\text{row}-0}$, and $||\cdot||_{\text{column}-0}$, respectively. This results in the following convex optimization problem:

$$\begin{aligned}&\min_{\mathbf{W},\mathbf{E}} ||\mathbf{W}||_{\infty,1}+\lambda||\mathbf{E}||_{1,\infty}\\ &\text{s.t.}\quad \mathbf{B}-\mathbf{BW}=\mathbf{E}.\end{aligned} \tag{5}$$

Since the outliers are absorbed by the error term $\mathbf{E}$, the weight $||\mathbf{W}^{(i)}||_\infty$ can be reliably utilized to extract prototypes. After deriving $\mathbf{W}$, we use the weight $||\mathbf{W}^{(i)}||_\infty$ to rank the samples. The larger $||\mathbf{W}^{(i)}||_\infty$ is, the more important this sample is. We can either extract a fixed number of the most important samples or set a threshold and select the samples of which $||\mathbf{W}^{(i)}||_\infty$ is larger than this value. In addition, if one is interested in the outliers, he can even use the similar method on column vectors of the obtained solution $\mathbf{E}$ to extract the outliers for further analysis.

Remark 3.1: In some recent work[5][7], the authors used to $||\mathbf{W}||_{2,1}$ to relax $||\mathbf{W}||_{\text{row}-0}$ in the model (3). Such a work also encourages the row sparsity and can be used to extract the prototypes. However, it neglects the influences of the outliers. In fact, [7] developed extra post-processing modules to reject the outliers. Since the sparse coding stage and the outlier rejection stage are performed individually, the original structure of the sparse coding may be destroyed and the obtained solution is just sub-optimal. In this work, the prototype extraction and the outlier isolation are performed simultaneously and therefore the obtained solution is expected to better.

Remark 3.2: In this work, we use the norm $||\mathbf{W}||_{\infty,1}$ to relax $||\mathbf{W}||_{\text{row}-0}$ to encourage row sparsity and use the norm $||\mathbf{E}||_{1,\infty}$ to relax $||\mathbf{E}||_{\text{column}-0}$ to encourage column sparsity. This is not the unique choice. In fact, other relaxations are certainly possible. However, both [24] and [9] indicate that there are few theoretical results available for these other relaxations. Therefore we prefer to the ∞ norm relaxation.

IV. Optimization Algorithm

Considering both $||\cdot||_{\infty,1}$ and $||\cdot||_{1,\infty}$ are convex norms, we can optimize them by Alternating Directional Method of Multipliers (ADMM)[1]. Correspondingly, equation (5) is equivalently converted to the following form:

$$\begin{aligned}&\min_{\mathbf{W},\mathbf{X},\mathbf{E}} ||\mathbf{X}||_{\infty,1}+\lambda||\mathbf{E}||_{1,\infty}\\ &\text{s.t.}\quad \mathbf{B}-\mathbf{BW}=\mathbf{E},\\ &\qquad \mathbf{W}-\mathbf{X}=0.\end{aligned} \tag{6}$$

By this way, the optimization over $\mathbf{W}$ and $\mathbf{X}$ are performed individually.

The augmented Lagrangian associated with the above optimization problem is given by

$$\begin{aligned}&\mathcal{L}(\mathbf{X},\mathbf{W},\mathbf{E},\mathbf{Y}_1,\mathbf{Y}_2)=||\mathbf{X}||_{\infty,1}+\lambda||\mathbf{E}||_{1,\infty}\\ &+Tr(\mathbf{Y}_1^T(\mathbf{B}-\mathbf{BW}-\mathbf{E}))+\tfrac{\mu}{2}||\mathbf{B}-\mathbf{BW}-\mathbf{E}||_F^2\\ &+Tr(\mathbf{Y}_2^T(\mathbf{W}-\mathbf{X}))+\tfrac{\mu}{2}||\mathbf{W}-\mathbf{X}||_F^2,\end{aligned} \tag{7}$$

where $\mathbf{Y}_1$ and $\mathbf{Y}_2$ are the dual variables (i.e., the Lagrangian multipliers), μ is a positive scalar. In order to find a minimizer of the constrained problem, the ADMM algorithm uses a sequence of iterations

$$\begin{cases}\mathbf{X}^{(k+1)}=\text{argmin}\ \ \mathcal{L}(\mathbf{X},\mathbf{W}^{(k)},\mathbf{E}^{(k)},\mathbf{Y}_1^{(k)},\mathbf{Y}_2^{(k)})\\ \mathbf{W}^{(k+1)}=\text{argmin}\ \ \mathcal{L}(\mathbf{X}^{(k+1)},\mathbf{W},\mathbf{E}^{(k)},\mathbf{Y}_1^{(k)},\mathbf{Y}_2^{(k)})\\ \mathbf{E}^{(k+1)}=\text{argmin}\ \ \mathcal{L}(\mathbf{X}^{(k+1)},\mathbf{W}^{(k+1)},\mathbf{E},\mathbf{Y}_1^{(k)},\mathbf{Y}_2^{(k)})\\ \mathbf{Y}_1^{(k+1)}=\mathbf{Y}_1^{(k)}+\mu(\mathbf{B}-\mathbf{BW}^{(k+1)}-\mathbf{E}^{(k+1)}))\\ \mathbf{Y}_2^{(k+1)}=\mathbf{Y}_2^{(k)}+\mu(\mathbf{W}^{(k+1)}-\mathbf{X}^{(k+1)})),\end{cases} \tag{8}$$

until $||\mathbf{B} - \mathbf{B}\mathbf{W}^{(k+1)} - \mathbf{E}^{(k+1)}||_F \leq \epsilon$ and $||\mathbf{W}^{(k+1)} - \mathbf{X}^{(k+1)})||_F \leq \epsilon$, where ϵ is the tolerance error. The superscript (k) represents the iteration number.

In the following we explain how to solve the optimization problems in (8).

First, the optimization over $\mathbf{X}$ is equivalent to

$$\min_{\mathbf{X}} \mathcal{L}(\mathbf{X}) = \frac{1}{\mu}||\mathbf{X}||_{\infty,1} + ||\mathbf{X} - \mathbf{V}^{(k)}||_F^2, \tag{9}$$

where

$$\mathbf{V}^{(k)} = \frac{1}{\mu}\mathbf{Y}_2^{(k)} + \mathbf{W}^{(k)}. \tag{10}$$

The solution of the optimization problem (9) is the proximity operator of the norm $||\cdot||_{\infty,1}$[4]. According to [9], the solution can be obtained as

$$\mathbf{X} = \mathbf{V}^{(k)} - \mathcal{P}_{C_{\frac{1}{2\mu}}}(\mathbf{V}^{(k)}), \tag{11}$$

where $\mathcal{P}_{\mathcal{C}_{\frac{1}{2\mu}}}(\mathbf{V}^{(k)})$ orthogonally projects each row of $\mathbf{V}^{(k)}$ onto the set $\mathcal{C}_{\frac{1}{2\mu}}$, which is defined as $\mathcal{C}_{\frac{1}{2\mu}} = \{\mathbf{c} \in R^N \mid ||\mathbf{c}||_1 \leq \frac{1}{2\mu}\}$.

Secondly, the optimization over $\mathbf{W}$ is equivalent to

$$\begin{aligned}\min_{\mathbf{W}} \quad & \mathcal{L}(\mathbf{W}) = Tr(\mathbf{Y}_1^T(\mathbf{B} - \mathbf{B}\mathbf{W} - \mathbf{E})) \\ & + \tfrac{\mu}{2}||\mathbf{B} - \mathbf{B}\mathbf{W} - \mathbf{E}||_F^2 + Tr(\mathbf{Y}_2^T(\mathbf{W} - \mathbf{X})) \\ & + \tfrac{\mu}{2}||\mathbf{W} - \mathbf{X}||_F^2,\end{aligned}$$

which is equivalent to

$$\begin{aligned}\min_{\mathbf{W}} \quad & \mathcal{L}(\mathbf{W}) = ||\mathbf{B} - \mathbf{B}\mathbf{W} - \mathbf{E} + \tfrac{1}{\mu}\mathbf{Y}_1||_F^2 \\ & + ||\mathbf{W} - \mathbf{X} + \tfrac{1}{\mu}\mathbf{Y}_2||_F^2.\end{aligned}$$

The solution can be obtained as

$$\mathbf{W} = (\mathbf{I} + \mathbf{B}^T\mathbf{B})^{-1}(\mathbf{B}^T\mathbf{B} - \mathbf{B}^T\mathbf{E} + \mathbf{X} + \frac{1}{\mu}\mathbf{B}^T\mathbf{Y}_1 - \frac{1}{\mu}\mathbf{Y}_2). \tag{12}$$

Finally, the optimization over $\mathbf{E}$ is equivalent to

$$\min_{\mathbf{E}} \quad \mathcal{L}(\mathbf{E}) = \tfrac{\lambda}{\mu}||\mathbf{E}||_{1,\infty} + ||\mathbf{E} - \mathbf{U}^{(k)}||_F^2, \tag{13}$$

where

$$\mathbf{U}^{(k)} = \frac{1}{\mu}\mathbf{Y}_1^{(k)} + \mathbf{B} - \mathbf{B}\mathbf{W}^{(k)}. \tag{14}$$

The solution of the optimization problem (13) is the proximity operator of the norm $||\cdot||_{1,\infty}$. Similar to the solution of $\mathbf{X}$, we have

$$\mathbf{E} = \mathbf{U}^{(k)} - \mathcal{P}_{C_{\frac{\lambda}{2\mu}}}(\mathbf{U}^{(k)}), \tag{15}$$

where $\mathcal{P}_{\mathcal{C}_{\frac{\lambda}{2\mu}}}(\mathbf{U}^{(k)})$ orthogonally projects each column of $\mathbf{U}^{(k)}$ onto the set $\mathcal{C}_{\frac{\lambda}{2\mu}}$, which is defined as $\mathcal{C}_{\frac{\lambda}{2\mu}} = \{\mathbf{c} \in R^N \mid \ ||\mathbf{c}||_1 \leq \frac{\lambda}{2\mu}\}$.

A whole algorithm summary which includes the above optimization procedures is given in Algorithm 1.

Algorithm 1

initialize Data set $\mathbf{B} \in R^{d\times N}$

Ensure: Solutions $\mathbf{W} \in R^{N\times N}$ and $\mathbf{E} \in R^{d\times N}$

1: Initialization: Set $\mathbf{X}$, $\mathbf{W}$, $\mathbf{E}$, $\mathbf{Y}_1$ and $\mathbf{Y}_2$ as zero matrices with appropriate dimensions.
2: **while** Not convergent **do**
3: Update $\mathbf{X}, \mathbf{W}, \mathbf{E}, \mathbf{Y}_1, \mathbf{Y}_2$ according to Eq.(8).
4: Update μ as $\min(1.1\mu, 10^{10})$.
5: **end while**

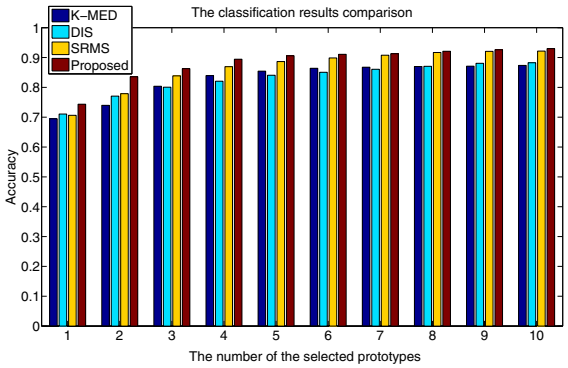

Fig. 3. Classification accuracy for various prototype extraction methods.

V. Experimental Results

We compare the proposed method with several baselines or state-of-the-art methods: K-Medoids, SRMS[7], and DIS[8] methods, on the recent German Traffic Sign Recognition Benchmark (GTSRB)[22]. This data set consists of 43 classes with 39209 training images and 12630 testing images in total. It is strongly unbalanced since the number of training samples in each class varies between 210 (for the first class) and 2250 (for the third class). In fact, the training sample set includes 1307 tracks and keeps multiple (about 30) images per track. Consecutive images in one track are similar and therefore the strong data redundancy is introduced. Here we investigate how the prototype extraction can be used to reduce the data redundancy. For each class, we use the above-mentioned methods to extract r representative training samples, where r is set to be within the interval [10, 100] with the space 10. In all the experiments, we use the HoG2 feature which is provided with the GTSRB dataset as the feature vector.

In Fig.3 we report the classification accuracy using SVM classifier. Since K-Medoids depends on initialization, we use 50 restarts of the algorithm and take the best result. This figure shows that when the number of the prototypes is small, our method significantly outperforms other ones. The main reason is that our model explicitly isolates the outliers. The role of the outlier isolation becomes weak when the number of the prototypes increases, as we see that the performance gap between the proposed method and SRMS becomes small when r is larger than 70.

In the following we present a result to show the parameter

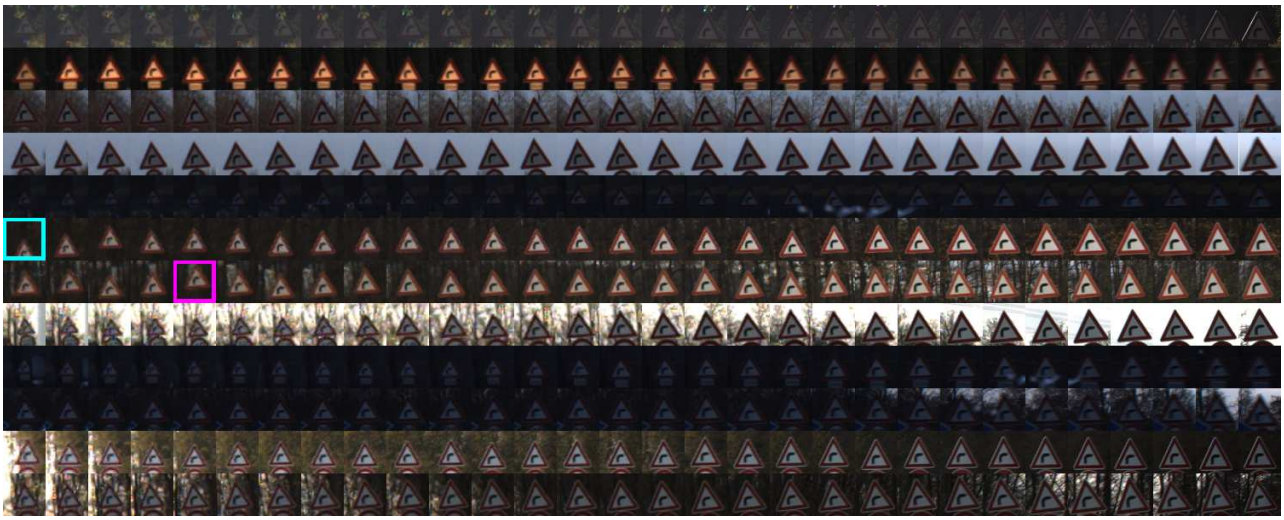

Fig. 7. The image sample set of the 21st class. Each row represents a track with 30 consecutive samples. All sample images are resized into the resolution of 50×50. The samples which are surrounded with a *celeste* box are the sample which obtains the largest value on the $||\mathbf{E}_{(j)}||_\infty$ curve in Fig.5. The samples which are surrounded with a *magenta* box are the sample which obtains the second largest value on the $||\mathbf{E}_{(j)}||_\infty$ curve in Fig.5. For clear illustration, please enlarge it in electronic format.

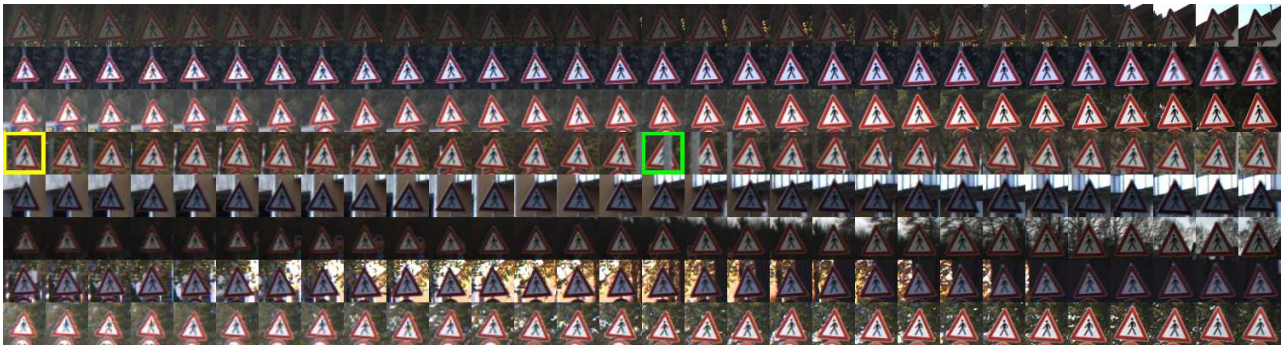

Fig. 8. The image sample set of the 28th class. Each row represents a track with 30 consecutive samples. All sample images are resized into the resolution of 50×50. The samples which are surrounded with a *green* box are the sample which obtains the largest value on the $||\mathbf{E}_{(j)}||_\infty$ curve in Fig.6. The samples which are surrounded with a *yellow* box are the sample which obtains the second largest value on the $||\mathbf{E}_{(j)}||_\infty$ curve in Fig.6. For clear illustration, please enlarge it in electronic format.

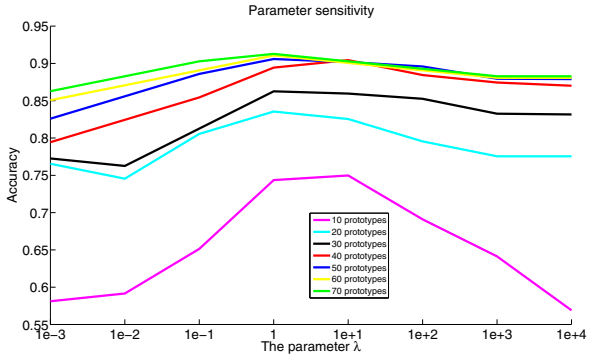

Fig. 4. Parameter sensitivity.

sensitivity. We change the regularization parameter λ from 10^{-3} to 10^4, and run the proposed algorithm on the dataset with different values of r. The classification accuracies using SVM are illustrated in Fig.4, from which we find that as λ increases, the classification performance does not monotonically increases. The best λ all appear in the range of [1,10], which means that both the reconstruction error term and the outlier isolation terms play important roles in finding good solution.

Finally we give more detailed results to show the outlier isolation capability. Fig.5 shows the results for the 21-st class. From the lower part (the $||\mathbf{E}_{(j)}||_\infty$ curve) we find there are two obvious peaks which are marked with color *celeste* and *magenta*, respectively. The corresponding samples are show in Fig.7, from which we see that the two samples which are surrounded with boxes are indeed outliers. In fact, the reason for the dramatic content change is that the camera is bumped and recovers at once.

Another example is shown in Figs.6-8, which are related to the 28-th class. From Fig.6 we also find two peaks (marked with the color *green* and *yellow*) in the $||\mathbf{E}_{(j)}||_\infty$ curve. The corresponding samples are show in Fig.8, from which we see that the two samples which are surrounded with

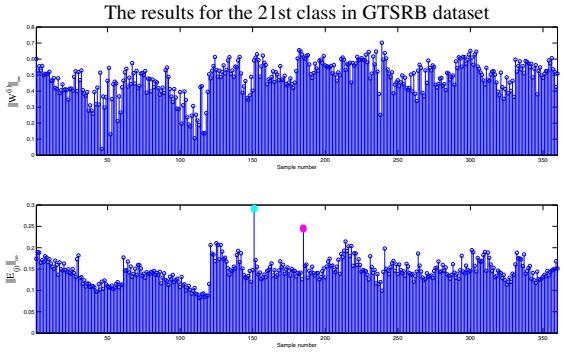

Fig. 5. Results for the 21st class. The upper panel shows the $||\mathbf{W}^{(i)}||_\infty$ curve and the lower panel shows the $||\mathbf{E}_{(j)}||_\infty$ curve.

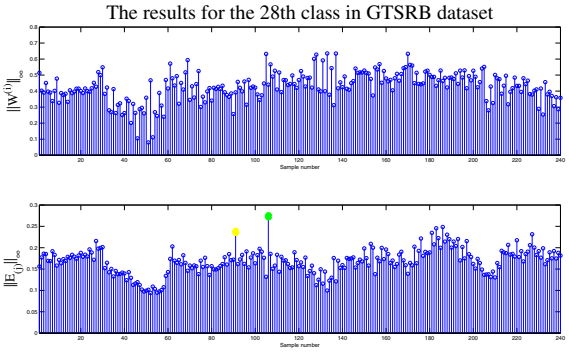

Fig. 6. Results for the 28st class. The upper panel shows the $||\mathbf{W}^{(i)}||_\infty$ curve and the lower panel shows the $||\mathbf{E}_{(j)}||_\infty$ curve.

boxes are indeed outliers: The sample with *green* box is seriously occluded and the sample with *yellow* box is slightly occluded. Both of them can be reliably isolated.

VI. Conclusions

In this paper, a collaborative sparse coding model is proposed to simultaneously characterize the representativeness and robustness which should be satisfied in robust prototype extraction task. The $\mathcal{L}_{1,\infty}$ norm is introduced to isolate the outlier and improve the robustness. Such a model admits convex characterization and we design an iterative ADMM algorithm to solve it. Finally, we present some validations on GTSRB dataset and show that the proposed model obtains promising results.

VII. Acknowledgements

This work is jointly supported by the National Key Project for Basic Research of China (2013CB329403), the National Natural Science Foundation of China (Grants No: 91120011, 61210013, 61105102) and the Tsinghua Self-innovation Project (Grant No:20111081111).

References

[1] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers, *Foundations and Trends in Machine Learning*, vol.3, no.1, pp.1-122, 2011

[2] H. Cheng, Z. Liu, L. Yang, Sparsity induced similarity measure for label propagation, in: *Proc. of Int Conf. on Computer Vision (ICCV)*, pp.317-324, 2009

[3] H. Cheng, Z. Liu, L. Hou, J. Yang, Sparsity induced similarity measure and its applications, *IEEE Trans. on Circuits and Systems for Video Technology*, In press (DOI: 10.1109/TCSVT.2012.2225911)

[4] P. L. Combettes, J. C. Pesquet, Proximal splitting methods in signal processing, *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, Springer Optimization and Its Applications, pp.185-212, 2011

[5] Y. Cong, J. Yuan, J. Luo, Towards scalable summarization of consumer videos via sparse dictionary selection, *IEEE Trans. on Multimedia*, vol.14, pp.66-75, 2012

[6] A. Dame, E. marchand, Using mutual information for appearance-based visual path following, *Robotics and Autonomous Systems*, vol.61, pp.259-270, 2013

[7] E. Elhamifar, G. Sapiro, R. Vidal, See all by looking at a few: Sparse modeling for finding representative objects, in: *Proc. of Computer Vision and Pattern Recognition (CVPR)*, pp.1600-1607, 2012

[8] E. Elhamifar, G. Sapiro, R. Vidal, Finding prototypes from pairwise dissimilarities via simultaneous sparse recovery, in: *Proc. of Advances in Neural Information Processing Systems (NIPS)*, pp.1-9, 2012

[9] E. Esser, M. Moller, S. Osher, G. Sapiro, J. Xin, A convex model for nonnegative matrix factorization and dimensionality reduction on physical space, *IEEE Trans. on Image Processing*, vol.21, no.7, pp.3239-3252, 2012

[10] F. Fraundorfer, D. Scaramuzza, Visual odometry, *IEEE Robotics and Automation Magazine*, pp.78-90, 2012

[11] Y. Girdhar, G. Dudek, Offline navigation summaries, in: *Proc. of Int. Conf. on Robotics and Automation (ICRA)*, pp.5769-5775, 2011

[12] Y. Girdhar, G. Dudek, Efficient on-line data summarization using extremum summaries, in: *Proc. of Int. Conf. on Robotics and Automation (ICRA)*, pp.3490-3496, 2012

[13] P. Henry, M. Krainin, E. Herbst, X. Ren, D. Fox, RGB-D mapping: using depth cameras for dense 3D modeling of indoor environments, *Int. J. of Robotics Research*, pp.1-17, 2012

[14] M. Hjelm, C. H. Ek, R. Detry, H. Kjellstrom, D. Kragic, Sparse summarization of robotic grasping data, in: *Proc. of Int. Conf. on Robotics and Automation (ICRA)*, pp.1074-1079, 2013

[15] G. Klein, D. Murray, Parallel tracking and mapping for small AR workspaces, in: *Proc. of Int. Symp. on Mixed and Augmented Reality (ISMAR)*, pp.225-234, 2007

[16] G. Klein, D. Murray, Improving the agility of keyframe-based SLAM, in: *Proc. of European Conf. on Computer Vision (ECCV)*, pp.802-815, 2008

[17] H. Liu, Y. Liu, F. Sun, Traffic sign recognition using group sparse coding, *Information Sciences*, in press

[18] R. Paul, D. Rus, P. Newman, How was your day? Online visual workspace summaries using incremental clustering in topic space, in: *Proc. of IEEE Int. Conf. on Robotics and Automation (ICRA)*, pp.4058-4065, 2012

[19] D. Pei, F. Sun, H. Liu, Supervised low-rank matrix recovery for traffic sign recognition in image sequences, *IEEE Signal Processing Letters*, vol.20, no.3, pp.241-244, 2013

[20] C. Slaughter, A. Y. Yang, J. Bagwell, C. Checkles, L. Sentis, S. Vishwanath, Sparse online low-rank projection and outlier rejection (SOLO) for 3-D rigid-body motion registration, in: *Proc. of IEEE Int. Conf. on Robotics and Automation (ICRA)*, pp.4414-4421, 2012

[21] N. Snavely, S. M. Seitz, R. Szeliski, Skeletal graphs for efficient structure from motion, in: *Proc. of Computer Vision and Pattern Recognition (CVPR)*, pp.1-8, 2008

[22] J. Stallkamp, M. Schlipsing, J. Salmena, C. Igel, Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition, *Neural Networks*, vol.32, pp.323-332, 2012

[23] J. Tardif, Y. Pavlidis, K. Daniilidis, Monocular visual odometry in urban environments using an omnidirectional camera, in: *Proc. of Int. Conf. on Intelligent Robots and Systems (IROS)*, pp.2531-2538, 2008

[24] J. A. Tropp, Algorithms for simultaneous sparse approxiamtion Part II: Convex relaxation, *Signal Processing*, vo.86, no.3, pp.589-602, 2006

[25] Q. Zhao, M. Q. H. Meng, A strategy to abstract WCE video clips based on LDA, in: *Proc. of IEEE Int. Conf. on Robotics and Automation (ICRA)*, 2011, pp.4145-4150, 2011