AUTOMATIC TARGET RECOGNITION BASED ON SIMULTANEOUS SPARSE REPRESENTATION

Vishal M. Patel[1], *Nasser M. Nasrabadi*[2] *and Rama Chellappa*[1]

[1]Department of Electrical & Computer Engineering
Center for Automation Research,
UMIACS, University of Maryland,
College Park, MD 20742.
{pvishalm,rama}@umiacs.umd.edu

[2]U.S. Army Research Laboratory
2800, Powder Mill Road,
Adelphi, MD 20783.
nnasraba@arl.army.mil

ABSTRACT

In this paper, an automatic target recognition algorithm is presented based on a framework for learning dictionaries for simultaneous sparse signal representation and feature extraction. The dictionary learning algorithm is based on class supervised simultaneous orthogonal matching pursuit while a matching pursuit-based similarity measure is used for classification. We show how the proposed framework can be helpful for efficient utilization of data, with the possibility of developing real-time, robust target classification. We verify the efficacy of the proposed algorithm using confusion matrices on the well known Comanche forward-looking infrared data set consisting of ten different military targets at different orientations.

Index Terms— Automatic Target Recognition, Forward-Looking Infrared (FLIR) Imagery, Simultaneous orthogonal matching pursuit (SOMP), Sparse representation.

1. INTRODUCTION

In automatic target recognition (ATR), the objective is to classify each target image into one of a number of classes. However, the presence of high clutter background, sensor noise, the large number of target classes, and the computational load involved in processing all the sensor data has often hampered the development of real-time robust ATR algorithms. The recognition algorithm usually consists of several stages such as detection of target, background noise removal, feature extraction and classification. In this paper, we mainly focus on the last two stages. Target recognition using forward-looking infrared (FLIR) imagery of different targets in natural scenes is difficult due to high variation in the thermal signatures of targets. Many ATR algorithms have been proposed for FLIR imagery. In [1], an ATR algorithm for FLIR imagery based on modular neural network was proposed. Wavelet based vector quantization was used for FLIR ATR in [2]. See [3] for an excellent survey of papers and experimental evaluation of FLIR ATR.

Recently, Wright *et al.* [4] introduced a sparse representation based robust face recognition algorithm, which outperformed many state of the art algorithms. Extensions based on [4] for FLIR ATR were recently presented in [5]. However, one of the main limitations of this approach is that for good recognition performance, the training images are required to be extensive enough to span the conditions that might occur in the test set. This may not be the case in many practical scenarios. Another limitation of this approach is that the large size of the matrix due to the inclusion of large number of gallery images can tremendously increase the computational complexity which can make the real-time processing very difficult.

To overcome the aforementioned limitations, in this paper, we propose an ATR algorithm based on learning class supervised dictionaries for simultaneous sparse signal representation and classification.

1.1. Paper Organization

In Section 2, we discuss details about the proposed framework for recognition using simultaneous orthogonal matching pursuit and a dissimilarity measure. In Section 3, we show some initial recognition results on a FLIR data set, and present the concluding remarks and future work in Section 4.

2. SIMULTANEOUS SIGNAL REPRESENTATION

In this section, we show how simultaneous orthogonal matching pursuit can be used for ATR.

2.1. Simultaneous Orthogonal Matching Pursuit (SOMP)

Let D be a redundant dictionary with K atoms in $\mathbb{R}^n$. The elements of the dictionary are indexed by $\gamma \in \Gamma$, i.e

$$D = \{\phi_\gamma : \gamma \in \Gamma\} \subset \mathbb{R}^n.$$

The atoms have unit Euclidean norm i.e., $\|\phi_\gamma\|_2 = 1, \forall \gamma \in \Gamma$. Let $X = [x_1, ..., x_s]$ be a set of training signals, where $x_i \in \mathbb{R}^n$ denotes the ith signal of X. Given D and X, SOMP attempts to approximate these signals at once as a linear combination of a common subset of atoms of cardinality much smaller than n [6]. Under the assumption that these signals belong to a certain class, SOMP extracts their common internal structure [6]. In fact, by keeping the sparsity low enough, one can eliminate the internal variation of the class which can lead to more accurate recognition while being robust to noise [6],[7],[8]. The SOMP algorithm is summarized in Fig. 1. In what follows, we show that after adding a discriminative term into SOMP, how we can use the coefficients of sparse representation together with the residual, over a class specific learned dictionary for recognition.

This work was partially supported by an ARO STIR Grant W911NF0910408.

Input: Dictionary D, signal matrix X, sparsity level T.
Output: A set Λ_T containing T indices, approximation A and residual matrix R.
Procedure:
1. Initialize the residual $R_0 = X$, $\Lambda = \emptyset$, and $t = 1$.
2. Find index γ_t, which solves the optimization problem

$$\arg\max_{\gamma \in \Gamma} \|R_t^T \phi_\gamma\|_1.$$

3. Set $\Lambda_t = \Lambda_{t-1} \cup \{\gamma_t\}$.
4. Determine the orthogonal projector P_t onto the span of the atoms indexed in Λ_t.
5. Compute the new approximation and residual:

$$A_t = P_t X,$$
$$R_t = (I - P_t)X.$$

6. If $t = T$, then stop. Otherwise, increment $t = t + 1$, and go to step 2.

Fig. 1. SOMP algorithm.

2.2. Separability based SOMP (SSOMP)

To further increase the discriminative power of SOMP, we adapt a supervised learning algorithm based on linear discriminant analysis (LDA). Note that LDA-based basis selection and feature extraction algorithm for classification using wavelet packets was proposed by Etemand and Chellappa in [9]. Recently, similar algorithms for simultaneous representation and discrimination have also been proposed in [7], [8], and [10].

Let us denote the number of classes by c and assume that

$$X = [X^{(1)}, \cdots, X^{(c)}] \in \mathbb{R}^{n \times m},$$

where $X^{(j)} = [x_1^{(j)}, \cdots, x_{n_i}^{(j)}] \in \mathbb{R}^{n \times n_i}$ denotes the samples that belong to the jth class that has n_i samples and $m = c.n_i$. To obtain a supervised atom selection algorithm, we modify the SOMP algorithm by adding a separability constraint that captures within-class and between-class variations. Define the within-class scatter matrix S_w as

$$S_w = \sum_{i=1}^{c} S_i, \tag{1}$$

where

$$S_i = \sum_{k=1}^{n_i} (x_k^{(i)} - \mu^{(i)})(x_k^{(i)} - \mu^{(i)})^T, \tag{2}$$

and $\mu^{(i)} = \frac{1}{n_i} \sum_{k=1}^{n_i} x_k^{(i)}$. One can also define the between-class scatter matrix S_b as

$$S_b = \sum_{i=1}^{c} n_i (\mu^{(i)} - \mu)(\mu^{(i)} - \mu)^T, \tag{3}$$

where $\mu = \frac{1}{cn_i} \sum_{i=1}^{c} n_i \mu^{(i)}$ is the total mean vector. In order to achieve good separability for classification, one needs to have large between-class scatter and small within-class scatter simultaneously. This can be achieved by introducing various cost functions [7],[8],[9]. In this paper, we use the following cost function

$$J(X) = Tr(S_w^{-1} S_b) \tag{4}$$

but similar results can be obtained by any of the other cost functions defined in [7],[8],[9].

For a dictionary D and a set of indices Λ, let $\Phi_\Lambda \in \mathbb{R}^{n \times |\Lambda|}$ be the matrix induced by the restriction of the dictionary elements whose indices are the elements of Λ. Then, the sparsity coefficients are given by $\alpha_k^{(j)} = (\Phi_\Lambda^T \Phi_\Lambda)^{-1} \Phi_\Lambda^T x_k^{(j)}$. From this observation, one can show that

$$Sb(\alpha) = (\Phi_\Lambda^T \Phi_\Lambda)^{-1} \Phi_\Lambda^T Sb(X) \Phi_\Lambda (\Phi_\Lambda^T \Phi_\Lambda)^{-1}$$
$$Sw(\alpha) = (\Phi_\Lambda^T \Phi_\Lambda)^{-1} \Phi_\Lambda^T Sw(X) \Phi_\Lambda (\Phi_\Lambda^T \Phi_\Lambda)^{-1}.$$

Hence, we can write the optimization problem that we want to solve in step 2 of the SOMP algorithm (to get the supervised SOMP) as follows

$$\arg\max_{\gamma \in \Gamma} \left(\|R_t^T \phi_\gamma\|_1 + \lambda J(\alpha) \right), \tag{5}$$

where $\lambda \geq 0$ controls the trade-off between discrimination and reconstruction. We call the resulting algorithm supervised SOMP (SSOMP).

2.3. Classification Using SOMP and SSOMP

Once the dictionaries are learned for each class, one can design a classifier based on either residuals (i.e. approximation error) or coefficients. For instance, SOMP (or SSOMP) approximations of the test sample g can be found using the learned dictionaries. The test sample can then be assigned the label of the class whose dictionary gives the best approximation of g (i.e. the smallest residual). However, a test signal may find an economic representation in many dictionaries. Hence, the approximation error by itself may not be the most reliable measure for classification.

The approach of comparing coefficient vectors of projected and original objects have also been proposed for classification [11]. Also, in [8] and [10, 7], nearest neighbor (NN) and SVM classifiers are used on the coefficient vectors for classification, respectively.

Since, the matching pursuit approximation defines a signal s in terms of its projection, the coefficient vector and the residual, we propose to use these for classification. Let P_s be the projection operator defined by the dictionary learned for the class containing s. Let $\alpha(s, P_s)$ be the coefficient vector and $R(s, P_s)$ be the residual. Then, in order to compare two signals g and s, we project g onto the projection P_s of s and noting the coefficient vector $\alpha(g, P_s)$ and residual $R(g, P_s)$. Based on these, the matching pursuit dissimilarity measure (MPDM)[12] has been defined as

$$\delta(g, s) = \sqrt{\theta F_R(g, s) + (1 - \theta) F_\alpha(g, s)}, \tag{6}$$

where $\theta \in [0, 1]$ determines the importance of the residuals and coefficients in δ, $F_R(g, s)$ is the difference between the residuals of g and s when both samples are projected onto the projection P_s of s

$$F_R(g, s) = \|R(g, P_s) - R(s, P_s)\|^2 \tag{7}$$

and $F_\alpha(g, s)$ compares their corresponding coefficient vectors

$$F_\alpha(g, s) = \|\alpha(g, P_s) - \alpha(s, P_s)\|^2. \tag{8}$$

Note that MPDM is a dissimilarity measure as small values indicate similar signals, while large values indicate dissimilar signals (see [12] for details). Once the class specific dictionaries are learned, the classification is accomplished using the NN classification rule in the MPDM sense. To further increase the recognition performance, one can also perform the k-NN in terms of MPDM.

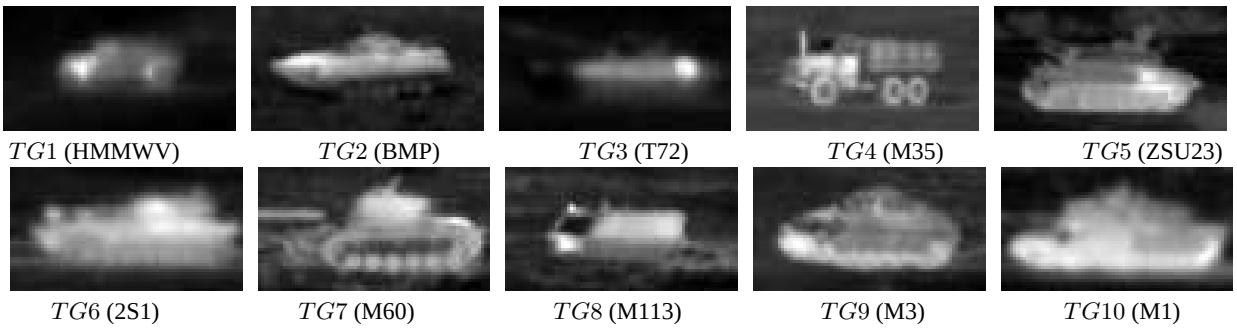

Fig. 2. Side view of all 10 targets present in the SIG data set.

It is also simple to introduce a reject threshold using the MPDM; if the value is too big, then the sample is considered not to belong to any class and should be rejected.

3. EXPERIMENTAL RESULTS

In this section, we present some preliminary results of our proposed algorithm on the Comanche FLIR data set consisting of different military targets at different orientations.

3.1. Dataset

The data set contains 10 different vehicle targets. We will denote these targets as $TG1, TG2, \cdots, TG10$. For each target, there are 72 orientations, corresponding to aspect angles of $0^\circ, 5^\circ, \cdots, 355^\circ$ in azimuth. The data consists of a training set and a test set. We will refer to the training set as the SIG set and the test set as the ROI set. The SIG data set has about 13,816 image chips, while there are 3,353 images in the ROI data set. The SIG data set consists of the images that were collected under very favorable conditions. The SIG data set contains 874 to 1468 images per target class. The ROI set consists of only five targets, namely $TG1, TG2, TG3, TG4$ and $TG7$. The target images for the ROI set were taken under less favorable conditions, such as targets with different weather conditions, in different background, in and around clutter; hence, these data are very challenging. There are 577 to 798 images for each of these five target classes. The images are of size 40×75 pixels. All the images in the SIG and ROI sets were normalized to a fixed range with the target put approximately in the center. The orientation in the ROI set was given very coarsely; every 45°. In Fig. 2 we show side view of all the 10 targets present in the SIG set.

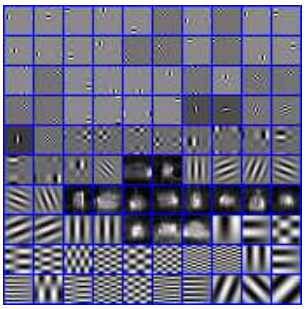

Fig. 3. A few 16×16 atoms from the dictionary, D.

3.2. Dictionary

In our experiments, the dictionary, D, contained about 1500 elements. It consisted of 2-D DCT atoms, 2-D Daubechies (4-taps) wavelet atoms, Gabor atoms and a few target chips. Fig. 3 shows some of the atoms from our dictionary.

3.3. Results

In the first set of experiments, we randomly selected 11 targets per aspect angle from the SIG data set for training, called TRAIN-SIG, and another set of 1000 targets disjoint from the training data for testing, called TEST-SIG. We used SOMP and SSOMP for training class specific dictionaries with 10 atoms. The value for λ in (5) was chosen to be 0.2 and the θ value in (6) was fixed to 0.5. In all the experiments, the target chips size was reduced from 40×75 to 16×16. Given c target classes, $\omega^1, \cdots, \omega^c$, each represented by its own separate dictionary, the classification rule we use is the following

$$\text{if } \delta(x, x_k^{(j)}) < \delta(x, x_k^{(l)}), \forall j \neq l, \ \forall k = 1, \cdots, n_i$$

then classify x into ω^j. The probabilities of correct classification for this experiment are 93.60 and 94.80 percent for the SOMP and SSOMP, respectively. The confusion matrices for this experiment are shown in Fig. 4 (a) and (c) for SOMP and SSOMP, respectively.

In the second set of experiments, we again randomly selected 11 targets per aspect angle from the SIG data set for training. We randomly chose a set of 1000 targets from the ROI data set for testing, called TEST-ROI. Again, we used SOMP and SSOMP for training class specific dictionaries with 10 atoms. The same values for λ and θ were used as before. The probabilities of correct classification for this experiment are 71.89 and 76.19 percent for the SOMP and SSOMP, respectively. The confusion matrices for this experiment are shown in Fig. 4 (b) and (d) for SOMP and SSOMP, respectively.

Table 1. Recognition rates (in $\%$) for different methods.

Methods	CNN4	MNN	LVQ	SOMP	SSOMP
TRAIN-SIG	95.16	95.49	99.72	100	100
TEST-SIG	-	90.53	-	93.60	94.80
TEST-ROI	59.25	75.58	75.12	71.89	76.19

From the above experiments, it is clear that introducing a discriminative term into SOMP generally improves the classification

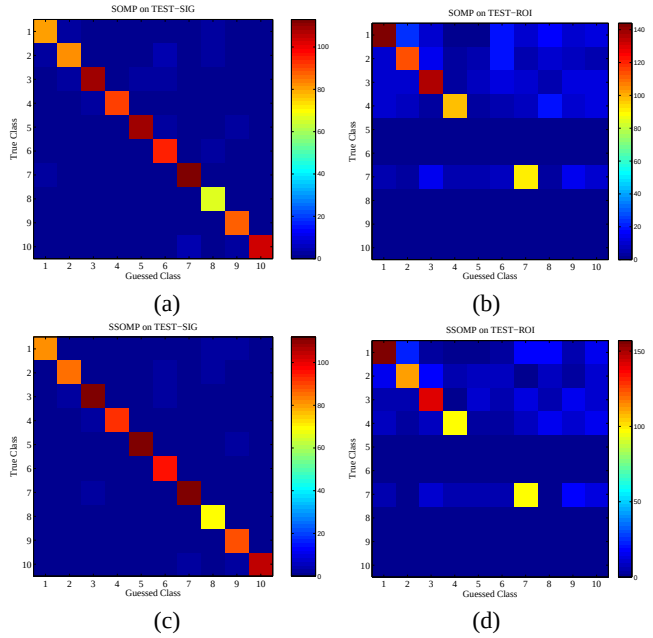

Fig. 4. Confusion matrices. (a) SOMP on TEST-SIG. (b) SOMP on TEST-ROI. (c) SSOMP on TEST-SIG. (d) SSOMP on TEST-ROI.

performance over SOMP. Also, note that our method is more general than the methods presented in [1] and [2]. In their methods, to deal with the background artifacts, they use several rectangular windows of different size based on the ground truth silhouette computer-aided design models. As a result, their performance significantly depends on the choice of windows [1],[2],[3]. In contrast, the method presented here does not require any windowing. Results obtained using different techniques are compared in Table. 1, where CNN4, MNN and LVQ stand for 4 layered convolutional neural network [3], modular neural network [1] and learning vector quantization [2], respectively.

4. DISCUSSION AND CONCLUSION

In this paper, we presented a framework for simultaneous sparse signal representation for robust ATR. Supervised SOMP was proposed to learn discriminative class specific dictionaries. The classification rule was based on a dissimilarity measure that combined both the coefficient vector and the residuals. Promising preliminary results were obtained on a difficult FLIR target data set.

Several future directions of inquiry are possible considering our new approach to ATR. For instance, in our proposed method, the dictionary, D, was predetermined. However, it has been observed that the choice of the dictionary that sparsifies the signals is crucial for signal representation and in some cases for classification [13],[14],[15]. We are currently investigation this possibility of optimizing the dictionary for the FLIR target images using the K-SVD like algorithms [13]. Also, the sparsity motivated methods for ATR presented here for FLIR images can be easily extended to the other object recognition problems such as the one based on synthetic aperture radar imagery and face recognition.

5. REFERENCES

[1] L. Wang, S. Z. Der and N. M. Nasrabadi, "Automatic Target Recognition using a Feature-decomposition and Data-decomposition Modular Neural Network," *IEEE Transactions on Image Processing*, vol. 7, No. 8, pp. 1113–1121, Aug. 1998.

[2] L. A. Chan and N. M. Nasrabadi, "An Application of Wavelet-based Vector Quantization in Target Recognition," *International Journal on Artificial Intelligence Tools*, vol. 6, No. 2, pp. 165–178, 1997.

[3] B. Li, R. Chellappa, Q. Zheng, S. Der, N. M. Nasrabadi, L. Chan and L. Wang, "Experimental Evaluation of FLIR ATR approaches - A Comparative Study," *Computer Vision and image understanding*, vol. 84, pp. 5–24, 2001.

[4] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust Face Recognition via Sparse Representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 31, no. 2, pp. 210–227, 2009.

[5] V. M. Patel, N. M. Nasrabadi and R. Chellappa, "Sparsity Inspired Automatic Target Recognition," SPIE Defense and Security Symposium, Automatic Target Recognition XX, Orlando, Florida, April 2010.

[6] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit", *Signal Processing*, vol. 86, no. 3, pp. 572-588, March 2006.

[7] F. Rodriguez and G. Sapiro, "Sparse representations for image classification: Learning discriminative and reconstructive non-parametric dictionaries," Tech. Report, University of Minnesota, Dec. 2007.

[8] E. Kokiopoulou and P. Frossard, "Semantic Coding by Supervised Dimensionality Reduction," *IEEE Transactions on Multimedia*, vol. 10, no. 5, pp. 806–818, Aug. 2008.

[9] K. Etemand and R. Chellappa, "Separability-Based multiscale basis selection and feature extraction for signal and image classification," *IEEE Transactions on Image Processing*, vol. 7, no. 10, pp. 1453–1465, Oct. 1998.

[10] K. Huang and S. Aviyente, "Sparse representation for signal classification," *NIPS*, vol. 19, pp. 609–616, 2007.

[11] P. J. Phillips, "Matching pursuit filters applied to face identification," *IEEE Transactions on Image Processing*, vol. 7, no. 8, pp. 150-164, 1998.

[12] R. Mazhar, P. D. Gader and J. N. Wilson, "Matching-Pursuits Dissimilarity Measure for Shape-Based Comparison and Classification of High-Dimensional Data," *IEEE Transactions on Fuzzy Systems*, vol. 17, no. 5, pp. 1175–1188, Oct. 2009.

[13] M. Aharon, M. Elad, and A. M. Bruckstein, "The K-SVD: an algorithm for designing of overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Process.*, vol. 54, no. 11, pp. 4311–4322, 2006.

[14] J. Mairal, F. Bach, J. Ponce, G. Sapiro and A. Zisserman., "Discriminative Learned Dictionaries for Local Image Analysis", *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008.

[15] J. Mairal, M. Leordeanu, F. Bach, M. Hebert and J. Ponce, "Discriminative Sparse Image Models for Class-Specific Edge Detection and Image Interpretation", *Proceedings of the European Conference on Computer Vision (ECCV)*, 2008.