Heterogeneous Structure Fusion for Target Recognition in Infrared Imagery

Guangfeng Lin[1], Guoliang Fan[2], Liangjiang Yu[2], Xiaobing Kang[1], Erhu Zhang[1]
Department of Information Science, Xi'an University of Technology, Xi'an Shaanxi, China[1]
School of Electrical and Computer Engineering, Oklahoma State University, Stillwater, OK, USA [2]
lgf78103@xaut.edu.cn, {guoliang.fan,liangjiang.yu}@okstate.edu, {kangxb,eh-zhang}@xaut.edu.cn

Abstract

We study Automatic Target Recognition (ATR) in infrared (IR) imagery from the perspective of feature fusion. The key to feature fusion is to take advantage of the discriminative and complementary information from different feature sets, which can be represented as internal (within each feature set) or external structures (across different feature sets). Traditional approaches tend to preserve either internal or external structures via certain feature projection. Some early attempts consider both structures implicitly or indirectly without revealing their relative importance and relevance. We propose a new unsupervised heterogeneous structure fusion (HSF) algorithm that is able to jointly optimize two kinds of structures explicitly and directly via a unified feature projection. The objective function of HSF integrates two feature structures in a closed form which can be optimized alternately via linear programming and eigenvector methods. The HSF solution provides not only the optimal feature projection but also the weight coefficients that encode the relative importance between two kinds of structures and among multiple feature sets. The experimental results on the COMANCHE IR dataset demonstrate that HSF outperforms state-of-the-art methods.

1. Introduction

Automatic target recognition (ATR) in infrared (IR) imagery is important in many civilian and military applications. It is still a challenging problem due to the lack of texture information and highly populated natural and manmade distractions such as heavy background clusters, dust, exhaust smoke, etc. Moreover, various atmospheric conditions (lighting, weather, temperature), size of targets, aspect angle and sensor movement could influence the target's signatures significantly. Typical ATR systems may consists of four steps [10], i.e., target detection, clutter rejection, feature extraction and recognition. In this work we focus on feature fusion-based ATR after detection, where only small IR chips with targets located near the center are considered.

There are a variety of features used for vision-based ATR, such as edge and corner descriptors [22], wavelets [6], deformable templates [12], histogram of oriented gradient (HOG) [15], shape silhouettes [38]. ATR shares many similarities to general computer vision techniques where many more visual features have been proven useful, including optical flow [21], local binary patterns (LBP) [14], shape contexts [4]. Since different feature sets can capture different characteristics of the same pattern, feature fusion is an effective approach to enhance the discriminability between different patterns and to reduce the feature redundancy via dimensionality reduction. Most efforts study the pattern complexity caused by some factors such as background clutter and nonlinear variations in the object appearance due to occlusions or the change of pose and illumination. Several fusion techniques have been applied in ATR applications on synthetic aperture radar (SAR) data [5], sonar imagery [2], aircraft object recognition [40], facial expression recognition [32], airborne tracking and identification using electro-optical (EO), IR and SAR data.

Traditionally, feature fusion has three steps: *selection*, *extraction* and *combination*. First, useful feature sets are selected according to some quantitative measure [18]. Second, salient features are extracted by space projection to reduce the redundancy and to improve the discriminability, for example, principal component analysis (PCA) [34], independent component analysis (ICA) [7], linear discriminant analysis (LDA), locality preserving projection (LPP) [13] and canonical correlation analysis (CCA) [23]. Third, multiple feature sets are combined in a serial or parallel fashion [34]. Recent efforts tend to integrate those three steps into a unified formulation for joint optimization, for example, several CCA variants [39], subspace learning [11], discriminant learning [25] and multiview spectral embedding [33]. Those methods try to preserve some feature structures during fusion, including the local structure among different feature sets [39], intra-class and inter-class neighborhood geometries [25]. Specifically, inter and intra-variability of feature sets are considered in [19], where the relationship between those variabilities is learned implicitly without providing their relative importance and relevance.

In this paper, we propose a new unsupervised heterogeneous structure fusion (HSF) algorithm which explicitly preserves and balances the two kinds of feature variabilities by finding a unified feature projection. In the proposed HSF algorithm, we refer inter and intra-variability of feature sets as the *external* and *internal* feature structures, respectively, which are jointly formulated in one optimization framework. The objective function of HSF combines two features structures in a closed form which can be optimized alternately via linear programming and eigenvector methods. The HSF solution provides not only the optimal feature projection but also the weight coefficients that encode the relative importance and relevance between two kinds of structures and among multiple feature sets. *The main contribution of this work is to explicitly and directly mine the relationship between internal and external feature structures by finding a unified feature projection that not only preserves the two kinds of structures and but also allows them to complement each other in an optimal way.* We evaluate the proposed algorithm on the COMANCHE IR dataset provided by the U.S. Army Research Lab (ARL), in which there are ten IR target classes taken under 72 azimuth angles ($0°$,$5°$,$10°$, ..., $355°$). To further demonstrate the effectiveness and robustness of the proposed HSF algorithm, we compare both target recognition and pose estimation results with a sparse representation based (SRC) recognition method [24], which reportedly outperforms nearly all existing ATR algorithms on the COMANCHE IR dataset.

2. Related Works

Feature fusion can be formulated in the context of supervised or unsupervised learning. Although this work is focused on unsupervised feature fusion, HSF is inspired by supervised methods in terms of how to characterize the relationship between different feature sets. We provide a brief review of feature fusion in two learning paradigms.

Supervised learning methods tend to balance the effect between the label information and the original feature distribution. However, the label information is the main cue for fusing features by maximizing the class separability and minimizing the class divergence ,like classic LDA. For example, the discrimination information is learned by considering inter-class "local" and intra-class "global" correlation of feature sets for visual recognition [26]. The class-dependent neighborhood information is incorporated into CCA in order to consider the local structure for multi-view dimensionality reduction [31]. In [36], both intra-class and inter-class geometries are taken into consideration in multi-view dimensionality reduction for scene classification where neighboring samples with different labels are used to preserve feature discriminability. In [37, 35], local patches are constructed for cartoon synthesis by fusing multiple features guided by labeling information.

Unsupervised learning methods focus on the original feature distribution and try to capture and preserve certain discriminant structure. For example, local structure description representing the relationship among neighbors, a kind of internal structure, is introduced in [33]. the maximum correlation among feature sets that is an often used to describe the external structure in [28]. In [16], multiple feature sets are projected into a unified low-dimensional space while preserving the internal or external feature structures but without explicitly revealing their relationship. In [19], the projection matrix of ICA is updated by CCA which implicitly considers the intra-variability and indirectly describes the inter-variability among features. From different perspectives of feature analysis, these methods perform well in many pattern recognition problems, including classification, visualization, pose estimation, face recognition, image retrieval, video annotation, document clustering, and brain function analysis. However, these methods do not explicitly represent the relationship between the internal and external structures. In other words, it is unclear what is the relative importance and relevance between two different structures in feature fusion.

3. Heterogeneous Structure Fusion (HSF)

3.1. Research Overview

The main idea of HSF is shown in Fig. 1. To the best of our knowledge, this work is the first attempt to explicitly represent both the internal and external feature structures and to jointly optimize them in a unified framework. In the following, we first introduce the metrics used to characterize two kinds of feature structures. Then we optimize feature projection by exploring and exploiting their relationship via linear programming and eigenvector methods. We also provide the pseudo code of the HSF algorithm.

3.2. Structure Metrics

Given M data samples each of which is represented by N-channel D-dimensional feature vectors, we can encapsulate the input data in a matrix $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_N]^T$ $(ND \times M)$, where $\mathbf{X}_k = [\mathbf{x}_{k1}, \mathbf{x}_{k2}, ..., \mathbf{x}_{kM}]^T$ $(k = 1, 2, ..., N)$ is the kth feature channel where $\mathbf{x}_{kl} \in R^D$ is the kth-channel feature of the lth sample. Similar to CCA or LPP, a feature projection of $\mathbf{X}_k$ is represented by $\mathbf{Y}_k = [\mathbf{y}_{k1}, \mathbf{y}_{k2}, ..., \mathbf{y}_{kM}]^T$ where $\mathbf{y}_{kl} \in R^d$ $(d << D)$ and which is expected to preserve certain feature structure, i.e., feature correlation in CCA or data similarity in LPP. Given a projection matrix $\mathbf{A}$ $(ND \times d)$, the input data $\mathbf{X}$ is projected to the fused data $\mathbf{Y}$ $(d \times M)$ via $\mathbf{Y} = \mathbf{A}^T\mathbf{X}$. The goal of HSF is to find the optimal $\mathbf{A}$ that preserves internal and external feature structures with appropriate weighting coefficients. We will discuss some structure metrics used for two kinds of feature structures below.

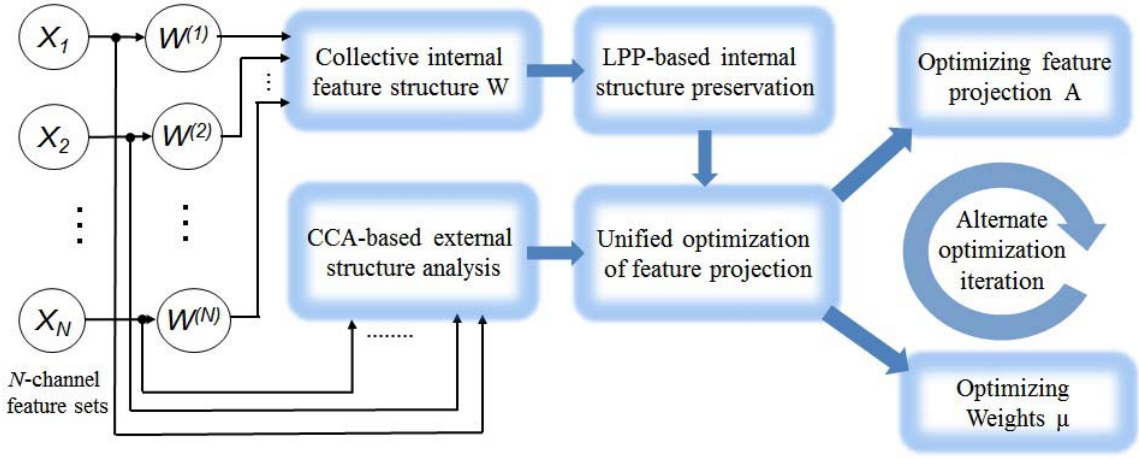

Figure 1: The illustration of our main research idea, where the internal and external feature structures are represented by LPP-based and CCA-based approaches and unified in one projection optimization framework.

The internal structure is represented by the similarity between data samples within the same feature set. Although the Euclidean metric is an often distance measurement for computing the data similarity, it is not suitable here due to the lack of the consideration of possible disconnected distribution in the feature space. On the other hand, the χ^2 metric is more appropriate to capture the internal structure because it involves a normalization factor to cope with different distribution scales. Thus we adopt the χ^2 metric as one option here due to easiness and convenience [17].The similarity matrix for the kth feature channel denoted by $\mathbf{W}^{(k)} = \{\mathbf{W}_{ij}^{(k)} | i, j = 1, 2, ..., M\}$ is computed as follows:

$$\mathbf{W}_{ij}^{(k)} = \begin{cases} e^{\frac{-dis(\mathbf{x}_{ki}, \mathbf{x}_{kj})}{\sigma_1}}, & \mathbf{x}_{ki} \in N_{\mathbf{x}_{kj}} \\ 0 \quad \text{else}, \end{cases} \tag{1}$$

where $dis(\cdot)$ is measured by χ^2 depending the nature of feature extraction; $\mathbf{x}_{ki}$ and $\mathbf{x}_{kj}$ are the kth-channel feature vectors of the ith and jth data samples, respectively; $N_{\mathbf{x}_{kj}}$ is the neighborhood of $\mathbf{x}_{kj}$; and σ_1 adjusts the sensitivity of similarity measure. The set of similarity matrices, $\{\mathbf{W}^{(k)} | k = 1, ..., N\}$, is used collectively as the measurement of the internal structure of N feature sets. This internal structure describes the distributional characteristics of each feature set, and it plays an important role in many pattern recognition tasks.

On the other hand, the structure represents the correlation and the distributional relationship among N feature sets. Specifically, we define $S(\mathbf{X}_p, \mathbf{X}_q)$ $(p, q = 1, 2, ..., N)$ to quantify the correlation between $\mathbf{X}_p = [\mathbf{x}_{p1}, \mathbf{x}_{p2}, ..., \mathbf{x}_{pM}]^T$ and $\mathbf{X}_q = [\mathbf{x}_{q1}, \mathbf{x}_{q2}, ..., \mathbf{x}_{qM}]^T$, which can be obtained by finding feature relevance among different feature sets as introduced in [11] as:

$$S(\mathbf{X}_p, \mathbf{X}_q) = \text{Tr}(\mathbf{Q}_{pq}^T \mathbf{P}_p^T \mathbf{P}_q \mathbf{Q}_{qp}), \tag{2}$$

where two orthonormal basis matrices $\mathbf{P}_p$ and $\mathbf{P}_q$ are introduced. To obtain these orthonormal basis matrices, we decompose the feature set matrices $\mathbf{X}_p \mathbf{X}_p^T = \mathbf{P}_p \mathbf{\Lambda}_p \mathbf{P}_p^T$ and $\mathbf{X}_q \mathbf{X}_q^T = \mathbf{P}_q \mathbf{\Lambda}_q \mathbf{P}_q^T$ where $\mathbf{\Lambda}_p$ and $\mathbf{\Lambda}_q$ respectively are the diagonal matrices of the corresponding eigenvalues. To measure the correlation of the feature sets, the SVD of $\mathbf{P}_p^T \mathbf{P}_q$ is $\mathbf{Q}_{pq} \mathbf{\Lambda}_{pq} \mathbf{Q}_{qp}^T$, where $\mathbf{\Lambda}_{pq}$ is the diagonal matrix of singular values.

3.3. Structure Fusion

Similar to [13, 3], if the data samples are close in the high-dimensional data space, we want the projected samples are still close in the low dimensional subspace. This idea can be further extended to the fusion of the internal structure extracted from multiple feature sets where we assume a linear relationship among them (represented a set of normalized weights $\{\omega_1, \omega_2, ...\omega_N\}$). Thus the projection matrix $\mathbf{A}$ that only considers the internal feature structure can be optimized via:

$$\begin{aligned} (\mathbf{A}, \omega) = \arg \min_{\mathbf{A}^T \mathbf{X} \mathbf{X}^T \mathbf{A} = I} \sum_{i,j} [\| \mathbf{A}^T \mathbf{x}_i - \mathbf{A}^T \mathbf{x}_j \|^2 \\ \cdot (\omega_1 \mathbf{W}_{ij}^{(1)} + ... + \omega_N \mathbf{W}_{ij}^{(N)})], \end{aligned} \tag{3}$$

where $\mathbf{x}_i$ and $\mathbf{x}_j$ are two column vectors of $\mathbf{X}$ representing the concatenated feature vectors $(ND \times 1)$ for ith and jth data samples, respectively. We factorize (3) into a matrix representation as

$$(\mathbf{A}, \omega) = \arg \min_{\mathbf{A}^T \mathbf{X} \mathbf{X}^T \mathbf{A} = I} \text{Tr}(\mathbf{A}^T \mathbf{X} \mathbf{L} \mathbf{X}^T \mathbf{A}), \tag{4}$$

where $\mathbf{L} = \mathbf{V} - \mathbf{W}$, $\mathbf{V}_{ii} = \sum_j \mathbf{W}_{ij}$; $\mathbf{W} = \{\mathbf{W}_{ij}|i,j = 1,...,M\}$ and $\mathbf{W}_{ij} = \omega_1\mathbf{W}_{ij}^{(1)} + ... + \omega_N\mathbf{W}_{ij}^{(N)}$. $\mathbf{W}$ encodes the fusion of internal structure from N feature channels. We further convert (4) into a maximization problem as:

$$(\mathbf{A}, \omega) = \arg \max_{\mathbf{A}^T\mathbf{X}\mathbf{X}^T\mathbf{A}=I} \mathrm{Tr}(\mathbf{A}^T\mathbf{X}\mathbf{W}\mathbf{X}^T\mathbf{A}). \quad (5)$$

For convenience, we define an objective function related to the internal structure fusion as:

$$J_{int} = \mathrm{Tr}(\mathbf{A}^T\mathbf{X}\mathbf{W}\mathbf{X}^T\mathbf{A}), \quad (6)$$

which will be further combined with the external structure in the following.

In (2), $\mathbf{P}_p$ and $\mathbf{P}_q$ are normalized by $\mathbf{P}_p\mathbf{R}_p^{-1}$ and $\mathbf{P}_q\mathbf{R}_q^{-1}$ ($\mathbf{R}_p$ and $\mathbf{R}_q$ respectively are the upper triangular matrixes of $\mathbf{A}^T\mathbf{P}_p$ and $\mathbf{A}^T\mathbf{P}_q$). We represent the correlation of two projected feature sets via $\mathbf{A}$ by extending (2) as:

$$S_{\mathbf{A}}(\mathbf{X}_p, \mathbf{X}_q) = \mathrm{Tr}(\mathbf{Q}_{pq}^T\mathbf{P}_p^T\mathbf{A}\mathbf{A}^T\mathbf{P}_q\mathbf{Q}_{qp}), \quad (7)$$

which is used to optimize the feature correlation among N feature sets as

$$\begin{aligned}\mathbf{A} &= \arg \max \sum_{p=1}^{N}\sum_{q=1}^{N} S_{\mathbf{A}}(\mathbf{X}_p, \mathbf{X}_q) \\ &= \arg \max \mathrm{Tr}(\mathbf{A}^T\mathbf{O}\mathbf{A}),\end{aligned} \quad (8)$$

where $\mathbf{O}$ represents the correlation among all N features sets (i.e., the external structure) as defined below:

$$\mathbf{O} = \sum_{p=1}^{N}\sum_{q=1}^{N}[(\mathbf{P}_p\mathbf{Q}_{pq} - \mathbf{P}_q\mathbf{Q}_{qp})(\mathbf{P}_p\mathbf{Q}_{pq} - \mathbf{P}_q\mathbf{Q}_{qp})^T], \quad (9)$$

from which we define an objective function pertaining to the external structure as:

$$J_{ext} = \mathrm{Tr}(\mathbf{A}^T\mathbf{O}\mathbf{A}). \quad (10)$$

Based on the non-negativity in (5), (6), (8) and (10), The HSF objective function is constructed by unifying J_{int} and J_{ext} together as:

$$\begin{aligned}(\mathbf{A}, \omega, \eta) &= \arg \max_{\mathbf{A}^T\mathbf{X}\mathbf{X}^T\mathbf{A}=I} (\eta_1 J_{ext} + \eta_2 J_{int}) \\ &= \arg \max_{\mathbf{A}^T\mathbf{X}\mathbf{X}^T\mathbf{A}=I} (\eta_1 J_{ext} + \eta_2 \mathrm{Tr}(\mathbf{A}^T\mathbf{X}\mathbf{W}\mathbf{X}^T\mathbf{A})) \\ &= \arg \max_{\mathbf{A}^T\mathbf{X}\mathbf{X}^T\mathbf{A}=I} \mathrm{Tr}(\mathbf{A}^T(\eta_1\mathbf{O} + \eta_2\mathbf{X}\mathbf{W}\mathbf{X}^T)\mathbf{A}) \\ &= \arg \max_{\mathbf{A}^T\mathbf{X}\mathbf{X}^T\mathbf{A}=I} \mathrm{Tr}(\mathbf{A}^T(\eta_1\mathbf{O} + \eta_2\mathbf{X}(\omega_1\mathbf{W}^{(1)} \\ &\quad + ... + \omega_N\mathbf{W}^{(N)})\mathbf{X}^T)\mathbf{A}) \\ &= \arg \max_{\mathbf{A}^T\mathbf{X}\mathbf{X}^T\mathbf{A}=I} \mathrm{Tr}(\mathbf{A}^T(\eta_1\mathbf{O} \\ &\quad + \eta_2\omega_1\mathbf{X}\mathbf{W}^{(1)}\mathbf{X}^T ... + \eta_2\omega_N\mathbf{X}\mathbf{W}^{(N)}\mathbf{X}^T)\mathbf{A}),\end{aligned} \quad (11)$$

where η_1 and η_2 two weights to balance two objection functions. $\omega_{1:N}$ and $\eta_{1:2}$ are independent and can be merged into a new set of weights that encode the relative importance and relevance between two kinds of structures and among multiple feature sets, as shown below:

$$\mu = \eta \circ \omega, \quad (12)$$

where $\mu = [\mu_1, \mu_2, ..., \mu_{N+1}]^T$, "$\circ$" is the Hadamard product, $\eta = [\eta_1, \eta_2, ..., \eta_2]^T \in R^{(N+1)}$, and $\omega = [1, \omega_1, \omega_2, ..., \omega_N]^T \in R^{(N+1)}$). Then (11) is converted to:

$$\begin{aligned}(\mathbf{A}, \mu) = \arg_A \max_{\mathbf{A}^T\mathbf{X}\mathbf{X}^T\mathbf{A}=I} &\mathrm{Tr}(\mathbf{A}^T(\mu_1 O + \mu_2\mathbf{X}\mathbf{W}^{(1)}\mathbf{X}^T \\ &+ ... + \mu_{N+1}\mathbf{X}\mathbf{W}^{(N)}\mathbf{X}^T)\mathbf{A}).\end{aligned} \quad (13)$$

If μ is fixed, then (13) is reduced to:

$$\mathbf{A} = \arg_{\mathbf{A}} \max_{\mathbf{A}^T\mathbf{X}\mathbf{X}^T\mathbf{A}=I} \mathrm{Tr}(\mathbf{A}^T\mathbf{Z}\mathbf{A}), \quad (14)$$

where

$$\mathbf{Z} = \mu_1\mathbf{O} + \mu_2\mathbf{X}\mathbf{W}^{(1)}\mathbf{X}^T + ... + \mu_{N+1}\mathbf{X}\mathbf{W}^{(N)}\mathbf{X}^T, \quad (15)$$

where $\mathbf{Z}$ shows the non-linear relationship between the internal structure ($\{\mathbf{W}^{(1)}, \mathbf{W}^{(2)}, ..., \mathbf{W}^{(N)}\}$) and the external structure ($\mathbf{O}$). The optimization of (14) is a generalized eigenvalue problem where the eigenvectors corresponding to the largest eigenvalues are used to construct the projection matrix $\mathbf{A}$. If the value of $\mathbf{A}$ is fixed, (13) becomes

$$\mu = \arg_{\mathbf{A}} \max_{\mathbf{A}^T\mathbf{X}\mathbf{X}^T\mathbf{A}=I} \mu^T\mathbf{H}, \quad (16)$$

where

$$\begin{aligned}\mathbf{H} = &[\mathrm{Tr}(\mathbf{A}^T\mathbf{O}\mathbf{A}), Tr(\mathbf{A}^T\mathbf{X}\mathbf{W}^{(1)}\mathbf{X}^T\mathbf{A}), \\ &..., Tr(\mathbf{A}^T\mathbf{X}\mathbf{W}^{(N)}\mathbf{X}^T\mathbf{A})]^T.\end{aligned} \quad (17)$$

The optimization of (16) is based on linear programming. Therefore, the solution to HSF is obtained by iteratively alternating between the eigenvector (14) and linear programming (16) methods.

3.4. HSF Algorithm

The pseudo code of the proposed HSF algorithm is presented in Algorithm 1. N is the number of feature sets or channels, and $\mathbf{X}_p, \mathbf{X}_q$ $(p, q = 1, 2, ..., N)$ respectively is any feature set. Firstly, the similarity matrix of each feature set is computed to describe the internal structure from step 1 to step 3. Secondly, the orthonormal basis matrixes is calculated to encode the following the external structure from step 4 to step 9. At last, the weight of structures and projection matrix is solved by alternately iterative optimization from step 10 to step 17.

Dimensions	60	65	70	75	80	85	90	95	100
Train/test(%)									
10/90	79.38	79.73	**79.86**	79.77	79.52	79.26	78.92	78.40	78.07
20/80	88.41	88.60	88.86	**88.92**	88.88	88.75	88.46	88.18	87.89
30/70	92.87	93.03	93.16	**93.24**	93.13	93.14	93.02	92.77	92.58
40/60	95.16	95.19	95.39	**95.48**	95.46	95.45	95.28	95.16	94.98
50/50	96.74	96.84	96.96	**96.98**	96.96	96.92	96.97	96.79	96.67
60/40	97.51	97.50	97.67	**97.76**	97.72	97.73	97.71	97.64	97.51
70/30	98.16	98.08	98.26	98.30	**98.34**	98.29	98.34	98.27	98.15
80/20	98.58	98.60	98.69	98.75	98.67	98.71	**98.79**	98.75	98.66
90/10	99.03	98.99	99.08	99.08	**99.18**	99.05	99.12	99.09	99.08
LOOCV	99.10	99.06	99.17	99.19	**99.22**	99.17	99.21	99.22	99.14

Table 1: The recognition accuracy on different training/testing and different dimension.

Algorithm 1 The pseudo code of the HSF algorithm

Input: $\mathbf{A} = I$ and $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_N]^T$
Output: fused feature $\mathbf{Y} = \mathbf{A}^T\mathbf{X}$

for $0 < i < N$ **do**
  Compute $\mathbf{W}^{(i)}$ according to (1)
end for
for $0 < p < N$ **do**
  for $0 < q < N$ **do**
    Compute $\mathbf{P}_p$ from $\mathbf{X}_p\mathbf{X}_p^T = \mathbf{P}_p\mathbf{\Lambda}_p\mathbf{P}_p^T$
    Compute $\mathbf{P}_q$ from $\mathbf{X}_q\mathbf{X}_q^T = \mathbf{P}_q\mathbf{\Lambda}_q\mathbf{P}_q^T$
  end for
end for
for $0 < i < T$ (T is the iteration number) **do**
  $\mathbf{R}_p$ and $\mathbf{R}_q$ respectively are computed by the upper triangular matrixes of $\mathbf{A}^T\mathbf{P}_p$ and $\mathbf{A}^T\mathbf{P}_q$
  $\mathbf{P}_p$ and $\mathbf{P}_q$ are normalized by $\mathbf{P}_p\mathbf{R}_p^{-1}$ and $\mathbf{P}_q\mathbf{R}_q^{-1}$
  $\mathbf{Q}_{pq}\mathbf{\Lambda}_0\mathbf{Q}_{qp}^T$ is the SVD of $\mathbf{P}_p^T\mathbf{P}_q \in R^{M\times M}$
  Compute $\mathbf{O}$ according to (9)
  Solve μ by optimizing (16)
  Solve $\mathbf{A}$ by optimizing (14)
end for

4. Experiments

In this section, we evaluate the performance of HSF in the ATR task on the Comanche IR database. In this database, there are 10 different military targets, and there are 72 orientations for each target ($0°, 5°, ..., 355°$). In addition, the database includes 874 to 1518 IR chips (40×75) for each target class, totaly 13859 chips.In Fig. 2, some chips is shown. The rows and columns are respectively the different targets and orientations. The experimental analysis has three aspects. First, the performance of different feature fusion methods are compared regarding the recognition accuracy to show their advantages to enhance the discrimination of features by mining their intrinsic structures. Second, we compare the HSF algorithm with the SRC-based methods which are considered as the-state-of-the-art ones in the field. Third, we further conduct the detailed analysis on the accuracy of pose estimation between HSF and SRC methods. Specifically, we extract two kinds of features from each IR chip, HOG (Histogram of Oriented Gradients) [8] and LBP (local binary pattern) [1]. More features are possible, but these two are found to be more effective ones.

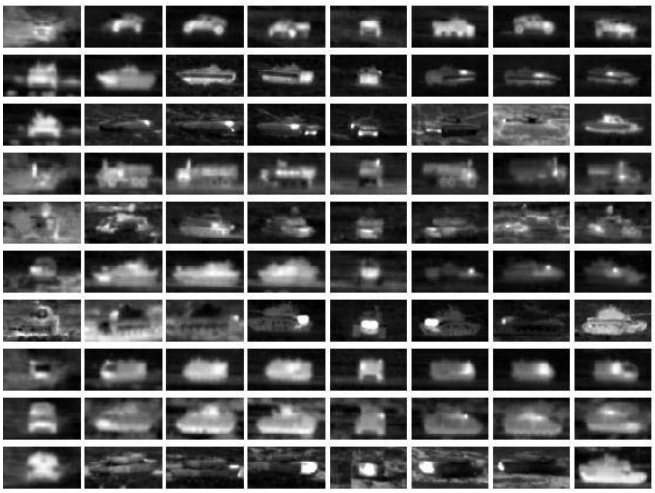

Figure 2: IR chips of 10 targets (row-wise) in 8 orientations (column-wise, $0°, 45°, 90°, 135°, 180°, 225°, 270°, 315°$) in the Comanche database

4.1. Comparison of Feature Fusion Methods

We first compare HSF with other feature fusion methods, including SFLPP [16], CCA [28], and MSE [33] in terms of their effectiveness on ATR. Classification is done by the nearest neighbor (NN) classifier. The experimental results show that the recognition accuracy of HSF can reach

99.23% under leave-one-out cross-validation (LOOCV). In contrast, those of SFLPP, CCA and MSE are respectively 97.81%, 96.19%, and 87.08%. In Fig. 3, the horizontal axis shows the feature dimension feature in the low-dimension manifold, and the vertical axis is the recognition accuracy. We can obtain the following observations.

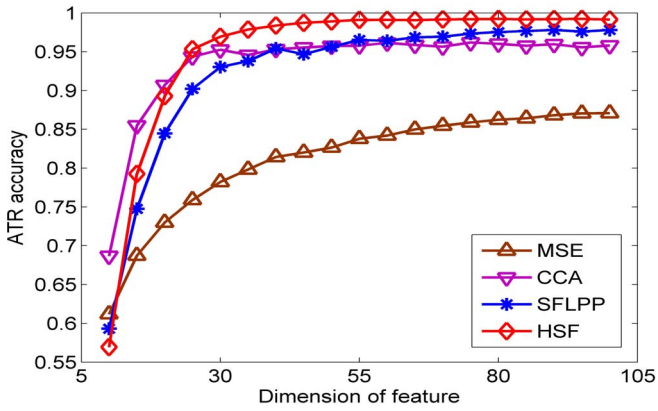

Figure 3: The recognition accuracy on the different fusion methods, which are HSF, SFLPP, CCA and MSE.

- When the dimension is low (< 25), the performance of HSF is slightly worse than those of the others. Otherwise, HSF is best among all methods. The reason is that a higher dimension is necessary to capture the feature structure for discriminating the different targets.
- SFLPP is better than CCA in the case of higher dimensions. This is because the internal structure is more important than the external structure for recognizing the different targets.
- Although MSE considers the internal feature structure, it does not involve feature projection which makes MSE less stable due to possible missing information during feature fusion.
- HSF has shown great promise in the experiment due to its explicit consideration and joint optimization of both internal and external structures for feature fusion.

4.2. Comparison between HSF and SRC Methods

To evaluate the robustness of HSF and confirm its optimal dimension of feature fusion in this work, we partition the database into different train/test sets 10%/90%, 20%/80% ,...,90%/10% in each class and under each angle. Each partition has 10 random trails, and we obtain the average recognition accuracy for each partition. In Table 1, we find that the 80-dimension feature results the best (the largest mean) and the most stable (the smallest stdev) recognition accuracy over 10 different partitions. Therefore, we select the 80-dimension feature in the following comparative analysis. We compare HSF with two SRC algorithms SparseLab [9] and spectral projected gradient (SPGL1) [30, 29]. In these SRC algorithms, we choose 'lasso' for optimization due to its better performance than others (for example, 'nnlasso' or 'OMP') [27]. Therefore, we name two SRC methods as SparseLab-lasso and SPGL1-lasso. As shown in Table 3, HSF outperforms others in the case of sufficient training samples ($> 30\%$).

Method	HSF	SparseLab-lasso	SPGL1-lasso
Train/test(%)			
10/90	79.52	82.64	**83.06**
20/80	88.88	89.17	**90.26**
30/70	93.13	92.37	**93.57**
40/60	**95.46**	93.95	95.22
50/50	**96.96**	95.37	96.60
60/40	**97.72**	95.94	97.20
70/30	**98.34**	96.83	97.94
80/20	**98.67**	97.30	98.28
90/10	**99.18**	97.79	98.80
LOOCV	**99.22**	97.94	98.79

Table 3: The recognition accuracy on different training/testing partitions by three methods.

4.3. Comparison of Target Pose Estimation

We also examine each of the three methods regarding their effectiveness on estimating the target's pose. We compute the accuracy of pose recognition under different feature dimensions and under different tolerable angle deviations ($\Delta = 5°, 10°, ..., 60°$). In Table 2, again we find the 80-dimension feature is the most suitable one. In Table 4, we compare HSF (dimension 80) with SparseLab-lasso and SPGL1-lasso. The performance of HSF is superior to that of both SRC algorithms under all 12 tolerable angle deviations. It shows that internal and external feature structures can be fused by HSF effectively for robust and accurate ATR. It is worth mentioning that the major computational load of HSF is for the one-time unsupervised learning. HSF-based classification is very efficient due to a relatively low feature dimension. On the other hand, two SRC algorithms are computationally costly due to the optimization involving all training data for each classification.

5. Conclusions

We have presented a new unsupervised feature fusion algorithm for ATR in IR imagery, called Heterogenous Structure Fusion (HSF), which jointly and explicitly takes advantage of heterogenous structures among multiple feature sets, namely the internal and external structures. Specifically, the former one characterizes the distribution structure in each feature channel, and the latter one represents the correlation

Dimensions	60	65	70	75	80	85	90	95	100
$\Delta(^\circ)$									
5	86.30	86.32	86.37	86.42	**86.44**	86.36	86.42	86.39	86.39
10	95.98	95.97	96.06	96.09	**96.13**	96.05	96.09	96.10	96.08
15	97.88	97.87	98.00	98.03	**98.08**	98.00	98.03	98.05	98.01
20	98.48	98.47	98.59	98.62	**98.67**	98.59	98.63	98.65	98.61
25	98.69	98.67	98.79	98.83	**98.88**	98.80	98.84	98.86	98.81
30	98.77	98.75	98.87	98.90	**98.96**	98.88	98.92	98.94	98.89
35	98.81	98.80	98.92	98.94	**99.01**	98.92	98.96	98.98	98.93
40	98.83	98.82	98.93	98.96	**99.02**	98.93	98.98	99.00	98.94
45	98.85	98.83	98.96	98.97	**99.04**	98.96	98.99	99.02	98.96
50	98.85	98.84	98.96	98.98	**99.04**	98.96	99.00	99.03	98.96
55	98.85	98.86	98.98	98.01	**99.07**	98.99	99.03	99.06	98.99
60	98.88	98.86	98.98	99.01	**99.07**	98.99	99.03	99.06	98.99

Table 2: The pose recognition accuracy of HSF under different dimensions and different tolerable angle deviations.

Methods	HSF	SparseLab-lasso	SPGL1-lasso
$\Delta(^\circ)$			
5	**86.44**	78.00	79.10
10	**96.13**	91.16	92.83
15	**98.08**	94.99	96.07
20	**98.67**	96.31	97.35
25	**98.88**	96.81	97.84
30	**98.96**	97.04	98.06
35	**99.01**	97.19	98.18
40	**99.02**	97.33	98.27
45	**99.04**	97.43	98.34
50	**99.04**	97.48	98.37
55	**99.07**	97.52	98.40
60	**99.07**	97.54	98.43

Table 4: The pose recognition accuracy of three methods under different tolerable angle deviations.

among all feature channels. The objective function is constructed in two parts. First, the internal structures across all feature channels are accumulated under a linear assumption due to their homogeneity. Second, the internal and external structures are combined with a non-linear relationship that reveals the heterogeneity of two kinds of feature structures. The HSF solution is obtained by iteratively alternating linear programming and eigenvector methods, which includes not only the optimal feature projection but also the weight coefficients that encode the relative importance and relevance between two kinds of structures and among multiple feature sets. The proposed HSF algorithm is evaluated thoroughly in the context of ATR where both LBP and HOG features are considered. Experimental results demonstrate that HSF not only outperforms recent feature fusion methods but also achieves very competitive ATR performance when compared the state-of-the-art methods.

Acknowledgment

The authors would like to thank the anonymous reviewers for their insightful comments that helped improve the quality of this letter. This work was supported by NSFC (Program No.61073092), Natural Science Basic Research Plan in Shaanxi Province of China (Program No.2014JM2-6111), Scientific Research Program Funded by Shaanxi Provincial Education Department (Program No.14JK1256) and Xian university of technology PhD project started (Program No. 104-211308)

References

[1] T. Ahonen, A. Hadid, and M. Pietikainen. Face description with local binary patterns: Application to face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(12):2037–2041, 2006.

[2] T. Aridgides and M. Fernandez. Automatic target recognition algorithm for high resolution multi-band sonar imagery. In *Proceedings of OCEANS 2008*, pages 1–7, Sept 2008.

[3] M. Belkin and P. Niyogi. Laplacian eigenmaps and spectral techniques for embedding and clustering. In *Proceedings of Advances in Neural Information Processing Systems*, 2001.

[4] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24(4):509–522, 2002.

[5] Z. Cao, Z. Cui, Y. Fan, and Q. Zhang. SAR automatic target recognition using a hierarchical multi-feature fusion strategy. In *Proceedings of IEEE Globecom Workshops*, pages 1450–1454, Dec 2012.

[6] D. P. Casasent, J. S. Smokelin, and A. Ye. Wavelet and gabor transforms for detection. *Optical Engineering*, 31(9):1893–1898, 1992.

[7] Z. Cataltepe, H. M. Genc, and T. Pearson. A pca/ica based feature selection method and its application for corn fungi detection. In *Proceedings of EUSIPCO*, 2007.

[8] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Proceedings of Computer Vision and Pattern Recognition*, 2005.

[9] D. Donoho, V. Stodden, and Y. Tsaig. Sparselab, March 2007. http://sparselab.stanford.edu.

[10] D. E. Dudgeon and R. Lacoss. An overview of automatic target recognition. *The Lincoln Laboratory Journal*, 6(1), 1993.

[11] Y. Fu, L. Cao, G. Guo, and T. S. Huang. Multiple feature fusion by subspace learning. In *Proceedings of the 2008 international conference on Content-based image and video retrieval*, 2008.

[12] U. Grenander, M. Miller, and A. Srivastava. Hilbert-schmidt lower bounds for estimators on matrix lie groups for atr. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 20(8):790–802, 1998.

[13] X. He and P. Niyogi. Locality preserving projections. In *Proceedings of Advances in Neural Information Processing Systems 16*, 2003.

[14] M. Heikkila and M. Pietikainen. A texture-based method for modeling the background and detecting moving objects. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(4):657–662, 2006.

[15] M. Khan, G. Fan, D. Heisterkamp, and L. Yu. Automatic target recognition in infrared imagery using dense hog features and relevance grouping of vocabulary. In *Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 293–298, June 2014.

[16] G. Lin, H. Zhu, X. Kang, C. Fan, and E. Zhang. Multi-feature structure fusion of contours for unsupervised shape classification. *Pattern Recognition Letters*, 34(11):1286–1290, 2013.

[17] H. Ling and D. W. Jacobs. Shape classification using the inner-distance. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(2):286–299, 2007.

[18] H. Liu and L. Yu. Toward integrating feature selection algorithms for classification and clustering. *IEEE Trans. Knowl. Data Eng.*, 17(4):491–502, 2005.

[19] J. Liu, G. Pearlson, A. Windemuth, G. Ruano, N. I. Perrone-Bizzozero, and V. Calhoun. Combining fmri and snp data to investigate connections between brain function and genetics using parallel ica. *Human brain mapping*, 30(1):241255, 2009.

[20] Q. Mo and B. Draper. Semi-nonnegative matrix factorization for motion segmentation with missing data. In *Computer Vision ECCV 2012*, volume 7578 of *Lecture Notes in Computer Science*, pages 402–415. 2012.

[21] C. Olson and D. Huttenlocher. Automatic target recognition by matching oriented edge pixels. *IEEE Trans. Image Processing*, 6(1):103–113, 1997.

[22] B. Paskaleva, M. Hayat, Z. Wang, J. Tyo, and S. Krishna. Canonical correlation feature selection for sensors with overlapping bands: Theory and application. *IEEE Trans. Geosci. Remote Sens*, 46(10):3346–3358, 2008.

[23] V. M. Patel, N. M. Nasrabadi, and R. Chellappa. Sparsity-motivated automatic target recognition. *Applied optics*, 50(10):1425–1433, 2011.

[24] Y. Su, Y. Fu, X. Gao, and Q. Tian. Discriminant learning through multiple principal angles for visual recognition. *IEEE Trans. Image Process.*, 21(3):1381–1390, 2012.

[25] Y. Su, Y. Fu, X. Gao, and Q. Tian. Discriminant learning through multiple principal angles for visual recognition. *IEEE Trans. Image Process.*, 21(3):1381–1390, 2012.

[26] J. Sun, G. Fan, L. Yu, and X. Wu. Concave-convex local binary features for automatic target recognition in infrared imagery. *EURASIP Journal on Image and Video Processing*, 2014(1), 2014.

[27] N. Sun, Z. hai Ji, C. Zou, and L. Zhao. Two-dimensional canonical correlation analysis and its application in small sample size face recognition. *Neural Computing and Applications*, 19(3):377–382, 2010.

[28] E. van den Berg and M. P. Friedlander. SPGL1: A solver for large-scale sparse reconstruction, June 2007. http://www.cs.ubc.ca/labs/scl/spgl1.

[29] E. van den Berg and M. P. Friedlander. Probing the pareto frontier for basis pursuit solutions. *SIAM Journal on Scientific Computing*, 31(2):890–912, 2008.

[30] F. Wang and D. Zhang. A new locality-preserving canonical correlation analysis algorithm for multi-view dimensionality reduction. *Neural Processing Letters*, 37(2):135–146, 2013.

[31] Z. Wang and S. Wang. Spontaneous facial expression recognition by using feature-level fusion of visible and thermal infrared images. In *Proceedings of IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, pages 1–6, Sept 2011.

[32] T. Xia, D. Tao, T. Mei, and Y. Zhang. Multiview spectral embedding. *IEEE Trans. Syst., Man, Cybern. Part B: Cybernetics*, 40(6):1438–1446, 2010.

[33] J. Yang, J.-Y. Yang, D. Zhang, and J. Lu. Feature fusion: parallel strategy vs. serial strategy. *Pattern Recognition*, 36(6):1369–1381, 2003.

[34] J. Yu, D. Liu, D. Tao, and H. S. Seah. On combining multiple features for cartoon character retrieval and clip synthesis. *IEEE Trans. Syst., Man, Cybern. Part B: Cybernetics*, 42(5):1413–1427, 2012.

[35] J. Yu, D. Tao, Y. Rui, and J. Cheng. Pairwise constraints based multiview features fusion for scene classification. *Pattern Recognition*, 46(2):483–496, 2013.

[36] J. Yu, M. Wang, and D. Tao. Semisupervised multiview distance metric learning for cartoon synthesis. *IEEE Trans. Image Process.*, 21(11):4636–4648, 2012.

[37] L. Yu, G. Fan, J. Gong, and J. Havlicek. Simultaneous target recognition, segmentation and pose estimation. In *Proceedings of IEEE International Conference on Image Processing (ICIP)*, pages 2655–2659, Sept 2013.

[38] Y.-H. Yuan, Q.-S. Sun, and H.-W. Ge. Fractional-order embedding canonical correlation analysis and its applications to multi-view dimensionality reduction and recognition. *Pattern Recognition*, 47(3):1411–1424, 2014.

[39] J. Zhao, Y. Fan, and W. Fan. Fusion of global and local feature using kcca for automatic target recognition. In *Proceedings of Fifth International Conference on Image and Graphics*, pages 958–962, Sept 2009.