

Article

Joint Infrared Target Recognition and Segmentation Using a Shape Manifold-Aware Level Set

Liangjiang Yu ¹, Guoliang Fan ^{1,*}, Jiulu Gong ² and Joseph P. Havlicek ³

¹ School of Electrical and Computer Engineering, Oklahoma State University, Stillwater, OK 74078, USA; E-Mail: liangjiang.yu@okstate.edu

² School of Mechatronical Engineering, Beijing Institute of Technology, Beijing 100081, China; E-Mail: lujiugong@bit.edu.cn

³ School of Electrical and Computer Engineering, University of Oklahoma, Norman, OK 73019, USA; E-Mail: joebob@ou.edu

* Author to whom correspondence should be addressed; E-Mail: guoliang.fan@okstate.edu; Tel.: +1-405-744-1547.

Academic Editor: Felipe Jimenez

Received: 13 February 2015 / Accepted: 22 April 2015 / Published: 29 April 2015

Abstract: We propose new techniques for joint recognition, segmentation and pose estimation of infrared (IR) targets. The problem is formulated in a probabilistic level set framework where a shape constrained generative model is used to provide a multi-class and multi-view shape prior and where the shape model involves a couplet of view and identity manifolds (CVIM). A level set energy function is then iteratively optimized under the shape constraints provided by the CVIM. Since both the view and identity variables are expressed explicitly in the objective function, this approach naturally accomplishes recognition, segmentation and pose estimation as joint products of the optimization process. For realistic target chips, we solve the resulting multi-modal optimization problem by adopting a particle swarm optimization (PSO) algorithm and then improve the computational efficiency by implementing a gradient-boosted PSO (GB-PSO). Evaluation was performed using the Military Sensing Information Analysis Center (SENSIAC) ATR database, and experimental results show that both of the PSO algorithms reduce the cost of shape matching during CVIM-based shape inference. Particularly, GB-PSO outperforms other recent ATR algorithms, which require intensive shape matching, either explicitly (with pre-segmentation) or implicitly (without pre-segmentation).

Keywords: infrared ATR; level set; shape modeling; particle swarm optimization

1. Introduction

We consider automatic target recognition (ATR) systems that detect and recognize extended targets by processing a sequence of images acquired from a passive imaging infrared (IR) sensor [1,2]. Our interest is primarily in sensors operating in the traditional 3–5 μm mid-wave IR (MWIR) or 8–12 μm long-wave IR (LWIR) bands, although our results are also applicable to those operating in the near, short-wave or far-IR bands, as well. The main functions typically performed by practical IR ATR systems of these types include detection, segmentation, feature extraction, tracking and recognition [3,4]. While these functions have historically been implemented sequentially, there is a growing recent interest in performing them jointly, so that tracking and recognition are not delayed by ambiguities in the detection process and so that inferences made by the track processor can be leveraged for both recognition and detection.

The infrared ATR problem presents significant challenges. Growth and processing techniques for IR detector materials, such as HgCdTe and InSb, are less mature than those for silicon, and hence, imaging IR sensors are typically characterized by higher noise and poor uniformity compared to their visible wavelength counterparts. The imagery acquired under practical field conditions often exhibits strong, structured clutter, poor target-to-clutter ratios and poor SNR. In important surveillance, security and military applications, the targets of interest may be non-cooperative, employing camouflage, decoys, countermeasures and complex maneuvers in an effort to evade detection and tracking. These difficulties are often exacerbated by the strong ego-motion of the sensor platform relative to the target. Depending on the operational waveband of the sensor, environmental conditions, such as smoke, haze, fog and rain, can result in degraded target signatures, as well as partial or full occlusions. All of these factors contribute to substantial appearance variability of the target thermal signature observed by the sensor, thereby limiting the effectiveness of approaches based on, e.g., stored libraries of static *a priori* signatures. A few examples of MWIR signature variability from the Military Sensing Information Analysis Center (SENSIAC) ATR Algorithm Development Image Database [5] are shown in Figure 1. Moreover, one would ideally like the ATR system to be capable of generalizing on the fly, so that both unknown target types and previously unseen views of known target types can be detected, tracked and recognized, at least to within an appropriate target class.

A very large number of ATR algorithms have been proposed in recent decades [3,4,6,7]. Some have been based primarily on the computation of certain types of features, such as PCA [8], edge and corner descriptors [9], wavelets [10] or deformable templates [11], while others have been driven more by a particular classification scheme, e.g., neural networks [12], support vector machines (SVM) [13] or sparse representations [14]. In the closely-related fields of computer vision and visual tracking, there have been significant developments in object detection and recognition based on visual features, including the histogram of oriented gradients (HOG) [15,16], the scale-invariant feature transform (SIFT) [17], spin images [18], patch features [19], shape contexts [20], optical flow [21] and local binary patterns [22]. Several feature point descriptors for long-wave IR data applications were evaluated in [23], including SIFT, speeded up robust features (SURF), binary robust invariant scalable keypoints (BRISK),

binary robust independent elementary features (BRIEF), fast retina keypoint (FREAK), oriented features from accelerated segment test (FAST) and rotated BRIEF (ORB) features. Certain geometric, topological and spectral descriptors have been widely used, as well [24,25]. Active contour methods [26,27] and level set algorithms have also been widely used in shape-based segmentation algorithms [28–30]. Shape priors were incorporated into both active contours and level set methods to handle cases of complicated background/foreground structure or objects in [31,32]. In [33], a couplet of identity and view manifolds (CVIM) was proposed for shape modeling by generalizing nonlinear tensor decomposition in [34]. CVIM explicitly defines view and identity variables in a compact latent space and was used with particle filtering for IR tracking and recognition in [33]. Gaussian process latent variable models (GPLVMs) were also used to learn a shape prior in order to accomplish joint tracking and segmentation in [35,36], and GPLVM was further extended for IR ATR application in [37].

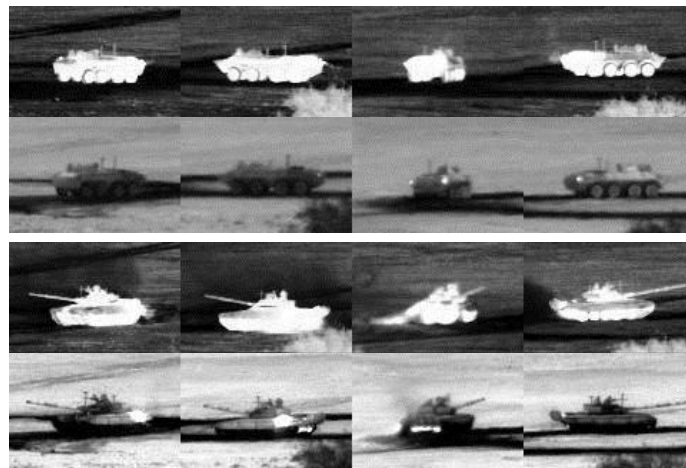


Figure 1. Examples of target signature variability from the Military Sensing Information Analysis Center (SENSIAC) ATR database [5]. The first and second rows show diurnal and nocturnal mid-wave IR (MWIR) images of a BTR70 personnel carrier, respectively. The third and fourth rows are diurnal and nocturnal images of a T72 main battle tank. Targets in each column are under the same view.

In this paper, we propose a new shape-constrained level set algorithm that incorporates the parametric CVIM model in a probabilistic framework for integrated target recognition, segmentation and pose estimation. The objective energy function of the level set is defined by associating CVIM with observations via a graphical model. To cope with the multi-modal property of CVIM for implicit shape matching, we first develop a particle swarm optimization (PSO) strategy [38] to optimize the energy function with respect to CVIM parameters, and then, we further propose a gradient-boosted PSO (GB-PSO) to improve the computational efficiency by taking advantage of the analytical nature of the objective function. There are two main contributions. The first one is a unified probabilistic level set framework that integrates CVIM-based implicit shape modeling and naturally supports multiple ATR tasks in one computational flow. The second one is an efficient GB-PSO algorithm that combines both gradient-based and sampling-based optimization schemes for CVIM-based implicit shape matching. Experimental results on the SENSIAC ATR database demonstrate the performance and

computational advantages of the proposed GB-PSO over other CVIM-based implementations, as well as recent ATR algorithms.

The remainder of the paper is organized as follows. Related ATR methods are reviewed in Section 2. Section 3 presents the proposed recognition and segmentation framework, including the analytical formulation and development of the two PSO-based algorithms. The experimental setup and empirical performance comparison are described in Section 4, while conclusions appear in Section 5.

2. Related Work

We first categorize several recent ATR algorithms into two main groups as shown in Figure 2 and then present the motivation and contributions of our present work relative to the existing methods.

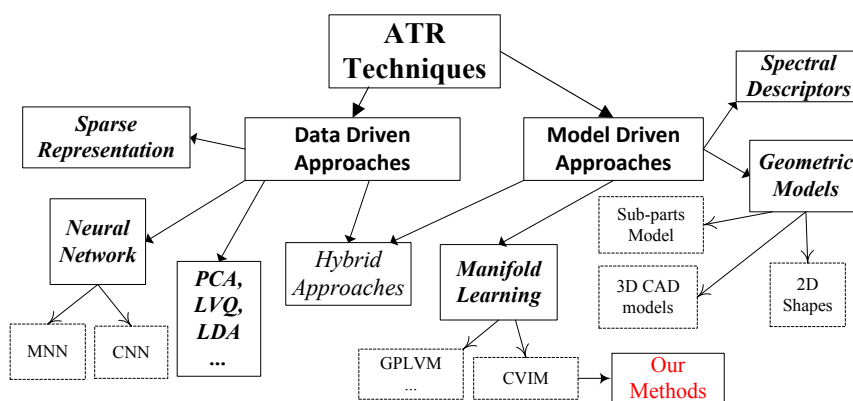


Figure 2. A taxonomy of ATR methods. MNN (modular neural network); CNN (convolutional neural network); LVQ (learning vector quantization); GPLVM (Gaussian process latent variable model); CVIM (couplet of view and identity manifolds).

2.1. Data-Driven Approaches

Data-driven approaches are typically based on learning from a set of labeled real-world training data. Neural networks (NN) [39–43] are an important exemplar. With the NN-based approach, the images acquired from the sensor are treated as points in a high-dimensional vector space. The objective is then to train a large multi-layer perceptron to perform the required mapping from this space to the space of labeled training images [39]. Convolutional neural networks (CNNs) generalize the basic idea by incorporating local receptive fields, weight sharing and spatial sub-sampling to accommodate some degree of translation and local deformation [40]. Modular neural networks (MNNs) are another important generalization where a collection of several independently trained networks each make a classification decision based on local features extracted from a specific region of the image [41]. These individual decisions are then combined to arrive at the overall final ATR classification decision. Another data-driven approach is the vector quantization (VQ)-based method developed in [42,43], where each target class is trained by the learning vector quantization (LVQ) algorithm and multi-layer perceptrons (MLPs) are used for recognition. A related architecture combining several individual ATR classifiers was proposed in [12]. A K-nearest-neighbor (KNN) data-driven approach for animal recognition using IR sensors was

proposed in [44]. Recently, sparse representation-based classification (SRC) methods have shown great promise in face recognition [45,46] and have also been applied to IR data for target detection [47], tracking and recognition [14,48]. The main drawbacks of these data-driven approaches are that they require large sets of training data, especially in the IR ATR applications considered here, and that the profound appearance variability of the observed target thermal signatures expected under practical field conditions tends to make dictionary selection extremely difficult.

2.2. Model-Driven Approaches

The model-driven approaches are based on computer-generated models (e.g., CAD models) with or without real-world sensor data for model learning. CAD models have been widely used for object segmentation, tracking and recognition [49–51]. Modern model-based ATR approaches generate target hypotheses and match the observed sensor data to the hypothesized signatures or appearance models [3]. The main idea is that a better interpretation of the scene and target can be achieved by applying intelligent reasoning while preserving as much target information as possible [52]. For example, in [53], radar features were extracted from the sensor data and used to construct a 3D model of the observed target that was compared with known models of the objects of interest to find the best match. There are also hybrid techniques [54–56] that combine both CAD models and data-driven 2D image features for model learning and inferencing. Indeed, 2D image features play an important role in many IR ATR algorithms [57], and a variety of such shape features were evaluated in [58]. One example of a hybrid technique is the multi-view morphing algorithm that was used in [59] to construct a view morphing database in an implicit way.

A number of manifold learning methods have also been proposed for shape modeling and have been applied recently in object tracking and recognition [60]. Elliptic Fourier descriptors were used in [35] to model shapes as sums of elliptic harmonics, and the latent space of target shapes was learned through GPLVMs. In [61], a level set framework was developed to optimize a pixel-wise posterior in the shape latent space in order to achieve simultaneous segmentation and tracking. A similarity space was added in [36] to handle multi-modal problems where an efficient discrete cosine transform (DCT)-based shape descriptor was used for manifold learning. A shape model called the couplet of view and identity manifolds (CVIM) that represents the target view and identity variables on a coupled pair of manifolds was proposed in [33] for joint target tracking and recognition in IR imagery. In [37], a probabilistic level set framework with shape modeling was proposed for target tracking and recognition, where a motion model was used in a particle filter-based sequential inference process. Sampling was performed in a local area predicted by the motion model, thereby alleviating the multi-modal optimization problem.

2.3. Research Motivation

Motivated by [33,36,37], our focus here is on developing a model-driven approach by combining relatively simple CAD models with advanced manifold learning for robust ATR to reliably segment and recognize target chips in the sequence of images acquired from an imaging IR sensor. More specifically, our goal is to incorporate CVIM into a probabilistic level set framework with shape-constrained latent space to achieve joint and seamless target segmentation, recognition and pose estimation with an

analytical formulation that facilitates efficient sampling-based or gradient-based global solution of the multi-modal optimization problem. This leads to a new approach that does not require labeled training data and is free from the need of any explicit feature extraction technique. Unlike many ATR algorithms, including [33], it is also free from the need for auxiliary background rejection or pre-segmentation processing; with our proposed methods, target segmentation instead becomes a useful byproduct of the joint ATR inference process.

3. Proposed Methods

In this section, we first introduce a probabilistic formulation of the proposed shape-constrained level set framework. This is followed by a brief review of CVIM for shape modeling. We then propose two PSO algorithms for joint ATR optimization. The first is a standard PSO that involves CVIM-based sampling for shape interpolation, while the second is a gradient-boosted PSO (GB-PSO) that implements a gradient-based search in the CVIM latent space.

3.1. Problem Formulation

Implicit contour and level set methods have been proven effective for image segmentation by optimizing an energy function, which represents the contour of an object appearing in the scene. A common approach is to compute the segmentation by optimizing the shape embedding function Φ [62]. The basic idea is to initialize a shape contour and then minimize the energy function related to Φ along the gradient direction. A probabilistic level set segmentation framework was proposed in [61], where, as illustrated in Figure 3a, an energy function called the pixel-wise posterior was defined to represent the image as a bag of pixels with the background and foreground models obtained from Φ [63].

Here, we extend the model from [61] to obtain a new shape-constrained level set segmentation method by incorporating the CVIM shape model parameterized by $\Lambda = [\Theta^T, \alpha]^T$, which explicitly represents the target identity variable α and azimuth/elevation view angles $\Theta = [\theta, \phi]^T$, thus inherently supporting joint target recognition, segmentation and pose estimation. We derive a new joint probability density function:

$$P(\mathbf{x}, \mathbf{y}, \Lambda, \mathbf{p}, M) = P(\mathbf{x}|\Lambda, \mathbf{p}, M)P(\mathbf{y}|M)P(M)P(\Lambda)P(\mathbf{p}) \quad (1)$$

where M is the foreground/background model, \mathbf{p} is the location of the target centroid in image coordinates, $P(\mathbf{p})$ is the prior probability of the target centroid location, which is assumed uniform. \mathbf{x} is a pixel location in image coordinates and \mathbf{y} is the pixel intensity. The intensity is usually scalar-valued for the case of an imaging MWIR or LWIR sensor, but may generally be vector-valued in our framework and formulation.

By marginalizing over the foreground/background model M [61] and using the logarithmic opinion pool [64], we formulate a new pixel-wise posterior:

$$P(\Lambda, \mathbf{p}|\Omega) = \prod_{i=1}^N \left\{ P(\mathbf{x}_i|\Lambda, \mathbf{p}, \mathbf{y}_i) \right\} \cdot P(\Lambda)P(\mathbf{p}) \quad (2)$$

where $\Omega = \{\mathbf{x}, \mathbf{y}\}$ is a small IR chip cropped from the IR frame acquired by the sensor, N is the number of pixels in the chip and i is the pixel index. Because we are focused on small IR chips that contain a

target, we localize the target centroid \mathbf{p} after segmentation and recognition. Therefore, \mathbf{p} is omitted in the following. As in [61], $P(\mathbf{x}_i|\Lambda, \mathbf{y}_i)$ in Equation (2) may be expressed according to:

$$P(\mathbf{x}_i|\Lambda, \mathbf{y}_i) = H_\epsilon(\Phi_{\mathbf{x}_i})P_f + (1 - H_\epsilon(\Phi_{\mathbf{x}_i}))P_b \quad (3)$$

where Φ is the shape embedding function generated from CVIM given Λ (in the form of a signed distance function, as shown in Figure 3b), $H_\epsilon(\cdot)$ is the smoothed Heaviside step function and $\Phi_{\mathbf{x}_i}$ is the value of Φ at pixel location \mathbf{x}_i . In Equation (3),

$$P_f = \frac{P(\mathbf{y}_i|M_f)}{\eta_f P(\mathbf{y}_i|M_f) + \eta_b P(\mathbf{y}_i|M_b)} \quad (4)$$

and

$$P_b = \frac{P(\mathbf{y}_i|M_b)}{\eta_f P(\mathbf{y}_i|M_f) + \eta_b P(\mathbf{y}_i|M_b)} \quad (5)$$

where η_f and η_b are the number of pixels belonging to the foreground and background regions respectively and where $P(\mathbf{y}|M_f)$ and $P(\mathbf{y}|M_b)$ are the foreground and background appearance models, which are represented by histograms.

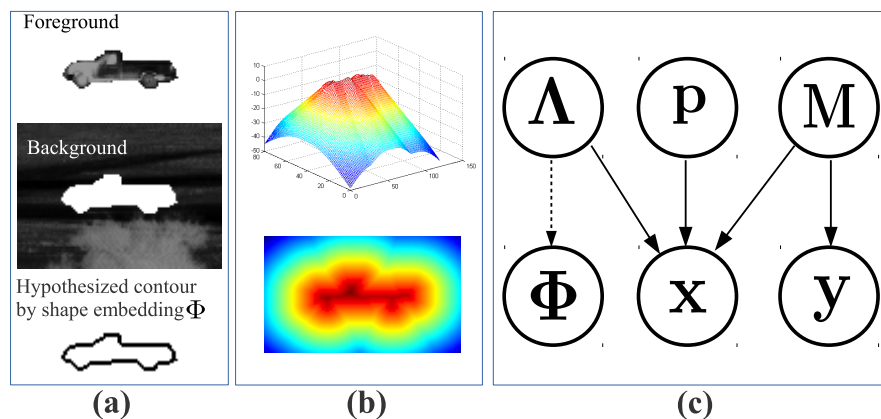


Figure 3. (a) Representation of an IR target by a hypothesized shape contour that separates the foreground and background regions; (b) shape embedding function Φ represented by the signed distance transform. Φ is generated by CVIM given the parameter vector Λ , which contains the view angles Θ and identity variable α ; (c) The proposed probabilistic level set framework, where \mathbf{p} is the target centroid in image coordinates, M is the foreground/background model, \mathbf{x} is a pixel location in image coordinates and \mathbf{y} is a pixel intensity value. The dashed line represents the CVIM-based mapping from the latent shape space Λ to the shape embedding function Φ .

The goal of the shape-constrained level set optimization is then to maximize Equation (2) with respect to Λ according to:

$$\Lambda^* = \arg \max_{\Lambda} P(\Lambda|\Omega) \quad (6)$$

The calculus of variations could be applied to compute the derivative of Equation (2) with respect to Λ . However, due to the multi-modal nature of the CVIM-based shape modeling, we develop a PSO-based optimization framework to search for the optimal latent variable Λ^* that maximizes Equation (2). To

further enhance the efficiency, we then develop a gradient-boosted PSO (GB-PSO) method that provides faster optimization by taking advantage of the parametric nature of CVIM.

3.2. DCT-Enhanced CVIM Shape Modeling

In this section, we briefly review the CVIM [33] and then extend it to accommodate DCT-based shape descriptors for learning and inference. The CVIM can be learned from a set of 2D shape silhouettes [65] created by a series of 3D CAD models by a nonlinear kernelized tensor decomposition, as shown in Figure 4. Here, we use six models for each of six target classes. The CVIM consists of a hemisphere-shaped view manifold and a closed-loop identity manifold in the tensor coefficient space. Two practical considerations lead to this heuristic simplification of the identity manifold. First, the SENSIAC targets of interest in this work are all man-made vehicles that exhibit distinct inter-class appearance similarities. Second, these similarities can be leveraged to judiciously order the classes along a 1D closed loop manifold in order to support convenient identity inference, as shown in Figure 4. In [33], a class-constrained shortest-closed-path method was proposed to deduce an optimal topology ensuring that targets of the same class or of similar shapes remain close along the identity manifold (*i.e.*, armored personnel carriers (APCs) → tanks → pick-ups → sedans → minivans → SUVs → APCs).

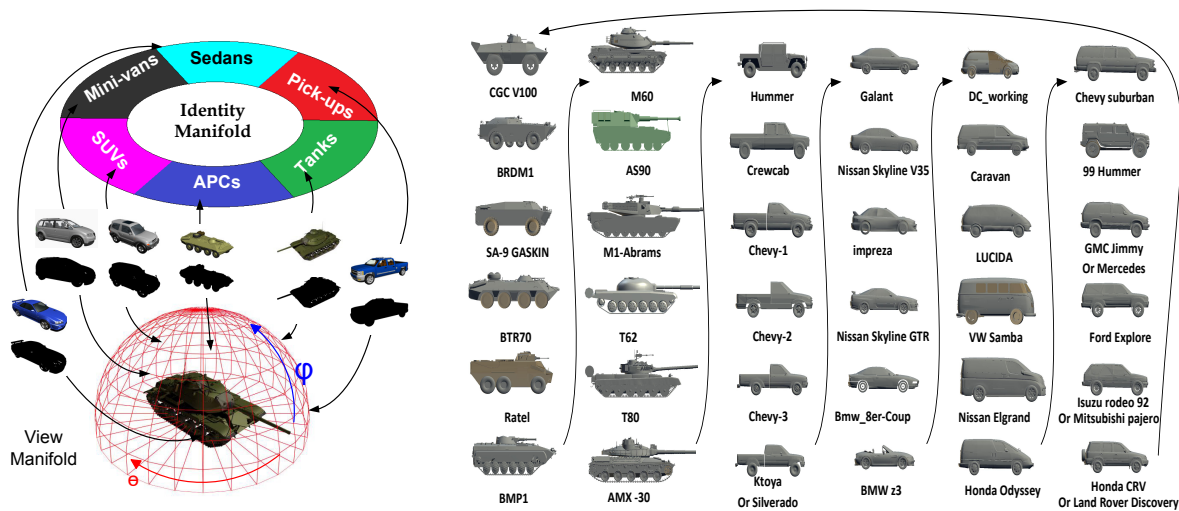


Figure 4. CVIM learning. (Left) The couplet of view and identity manifold (CVIM) shape generative model [33]; (Right) six CVIM CAD training models for each of six target classes.

The original CVIM is learned as follows: (1) given a set of silhouettes (represented by the signed distance transform) from N_m target types under N_c views, a mapping from a conceptual hemispherical view manifold Θ to the high dimensional data is learned using radial basis functions (RBFs) $\Psi(\Theta)$ for each target shape. (2) by stacking the collection of these mappings for all target shapes and applying the high-order singular value decomposition (HOSVD), we obtain a core tensor \mathbf{A} and N_m identity vectors for all training types in the tensor coefficient space \mathbf{i}_m ($m = 1, 2, \dots, N_m$); (3) a mapping from the coefficient vector space to a 1D closed loop identity manifold α is then constructed using the optimal identity manifold topology, where each training target type \mathbf{i}_m is represented by an identity vector $\mathbf{i}(\alpha_m)$ associated with a point along the identity manifold. For any arbitrary $\alpha \in [0, 2\pi)$, we can then obtain

a corresponding identity vector $\mathbf{i}(\alpha)$ from the two closest training identity vectors $\mathbf{i}(\alpha_m)$ and $\mathbf{i}(\alpha_{m+1})$ by applying cubic spline interpolation along the identity manifold. The CVIM model was tested against the SENSIAC database [5] for target tracking and recognition in [33], where the experimental results validated its efficacy both qualitatively and quantitatively.

Here, we reduce the inference complexity of the original CVIM method [33] by replacing the silhouettes used for training with the simple, but efficient DCT-based shape descriptor proposed in [36]. Thus, each training shape (in the form of the signed distance transform) is represented by a small set of 2D DCT coefficients reshaped into a column vector, where, e.g., only the top 10% of the DCT coefficients that are largest in magnitude are retained. The CVIM can then be learned using the same process as before to represent the sparse DCT coefficient vectors \mathcal{S}_{DCT} of the training targets by $\mathbf{\Lambda} = [\Theta^T, \alpha]^T$ according to:

$$\mathcal{S}_{\text{DCT}}(\mathbf{\Lambda}) = \mathcal{S}_{\text{DCT}}(\alpha, \Theta) = \mathbf{A} \times_3 \mathbf{i}(\alpha) \times_2 \Psi(\Theta) \quad (7)$$

where \mathbf{A} is a core tensor obtained by tensor decomposition, \times_n is the mode- n tensor multiplication, α and $\Theta = [\theta, \phi]^T$ are the identity and view latent variables on the identity and view manifolds, respectively, and θ and ϕ are the azimuth and elevation angles. For an arbitrary α , the associated $1 \times N_m$ identity (row) vector $\mathbf{i}(\alpha)$ in Equation (7) can be interpolated as:

$$\mathbf{i}(\alpha) = \mathbf{a}_m(\alpha - \alpha_m)^3 + \mathbf{b}_m(\alpha - \alpha_m)^2 + \mathbf{c}_m(\alpha - \alpha_m) + \mathbf{d}_m, \quad \alpha \in [\alpha_m, \alpha_{m+1}] \quad (8)$$

where \mathbf{a}_m , \mathbf{b}_m , \mathbf{c}_m and \mathbf{d}_m are the piecewise polynomial coefficient row vectors obtained by applying cubic spline interpolation in the tensor coefficient space between the closest two adjacent training target types $\mathbf{i}(\alpha_m)$ and $\mathbf{i}(\alpha_{m+1})$, as depicted in Figure 5. Let $\Psi(\Theta)$ be the RBF mapping along the view manifold given by:

$$\Psi(\Theta) = [\kappa(\|\Theta - \mathbf{S}_1\|), \dots, \kappa(\|\Theta - \mathbf{S}_{N_c}\|)] \quad (9)$$

where $\kappa(\|\Theta - \mathbf{S}_i\|) = e^{-c(\Theta - \mathbf{S}_i)^T(\Theta - \mathbf{S}_i)}$, c is the RBF kernel width, \mathbf{S}_i is a training view on the view manifold and N_c is the number of training views. One major advantage of this DCT-based shape representation over the original silhouette-based one is that it naturally provides reconstruction of a shape at arbitrary magnification factors by appropriately zero-padding the DCT coefficients prior to inverse DCT (IDCT). This feature is desirable to deal with various IR targets at different ranges.

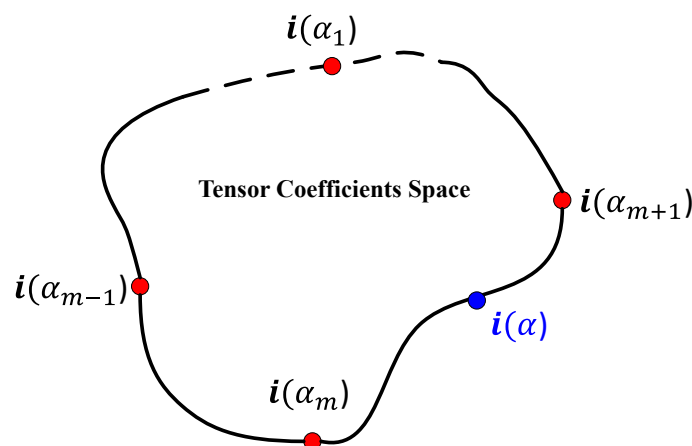


Figure 5. Cubic spline interpolation along the identity manifold in CVIM.

We represent the shape embedding function Φ (referred to in Figure 3c and Equation (3)) in terms of the CVIM parameter Λ by:

$$\Phi(\Lambda) = \text{IDCT}(\mathcal{S}_{\text{DCT}}(\Lambda)) \quad (10)$$

where $\text{IDCT}(\cdot)$ is the IDCT with two reshape operations. The first is for the input (from 1D to 2D) prior to IDCT, and the second is for the output (from 2D to 1D) after IDCT. Note that the derivative of the IDCT of a matrix may be computed as the IDCT of the derivative of that matrix [36]. Therefore, the DCT-shape presentation can easily be incorporated into the above optimization framework without major modifications. Through this CVIM model, target shapes corresponding to arbitrary α can readily be interpolated along the view and identity manifolds.

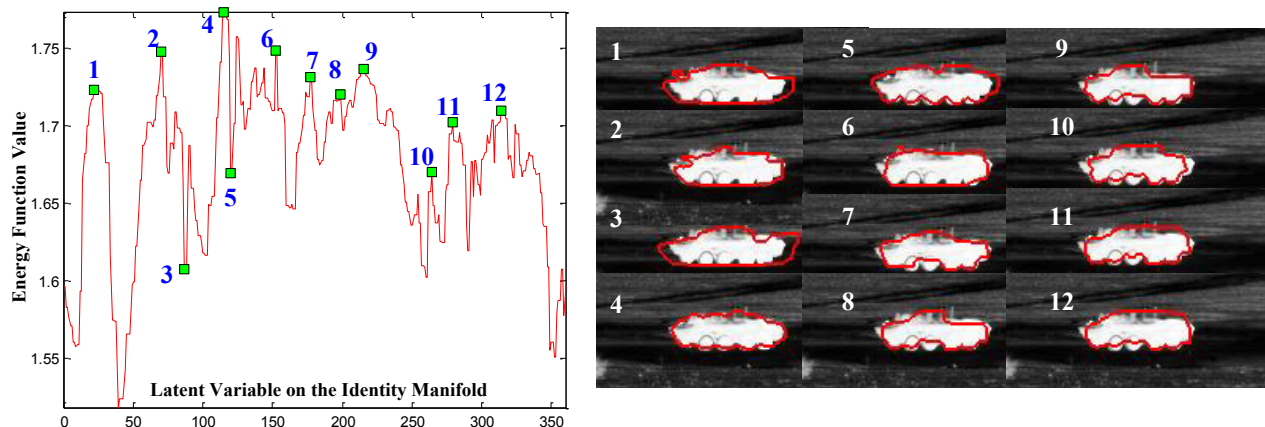


Figure 6. Effectiveness of the energy function. (Left) Plot of the energy function with respect to the identity latent variable; (right) a group of CVIM-generated shapes corresponding to the latent variables labeled in the plot on the left are superimposed onto the original IR data chips.

Optimizing the posterior Equation (2) may be thought of as finding a contour that maximizes the histogram difference between the foreground and background in the region of interest. This consideration is based on the assumption that the target-of-interest has different intensity values compared with the background. Intuitively, if the shape contour of the target is correctly hypothesized in terms of the target type (recognition), view angle (pose estimation) and location (segmentation), then the foreground and background defined by this contour will have maximum histogram divergence and, therefore, maximize the energy function Equation (2) as illustrated in Figure 6. For a given observation, in Figure 6, we calculate the value of the energy function with respect to almost all possible values along the circularly-shaped identity manifold $\alpha = 1, 2, 3, \dots, 360^\circ$ with the view angle Θ known for simplicity. The figure shows several CVIM interpolated shapes superimposed on the original mid-wave IR image data. As seen in the left part of the figure, the maximum value of the energy function is attained by the contour (numbered 4) that is best in the sense of being closest to the actual boundary of the target in the right part of the figure. However, the multi-model nature of the energy function as shown in Figure 6

(left part), which is typical, represents significant challenges for CVIM-based shape optimization and motivates the PSO and GB-PSO algorithms that we develop below in Sections 3.3 and 3.4.

3.3. PSO-Based Optimization

We first implement a standard PSO algorithm due to its simplicity and effectiveness in dealing with multi-modal optimization problems. PSO algorithms were originally developed in [38] inspired by the observation of bird flocking and fish schooling, which are also related to genetic algorithms, and they are widely used in both scientific research [66,67] and engineering applications [68,69]. PSO optimizes a problem by moving solution hypotheses around in the search-space according to the current hypothesis and velocity computed to the present local and global optima. Our energy function is defined in Equation (2). Since we assume $\Lambda = [\theta, \phi, \alpha]^T$ (CVIM parameters) to be uniformly distributed (*i.e.*, no prior knowledge) with the registration of the object frame \mathbf{p} omitted, the energy function Equation (2) is rewritten as:

$$f(\Lambda) = P(\Lambda|\Omega) \propto \prod_{i=1}^N \left\{ H_{\epsilon}(\Phi_{\mathbf{x}_i}(\Lambda))P_f + (1 - H_{\epsilon}(\Phi_{\mathbf{x}_i}(\Lambda)))P_b \right\} \quad (11)$$

where $\Phi_{\mathbf{x}_i}(\Lambda)$ is defined as the value of shape embedding (in the form of the signed distance transform) in Equation (10) at pixel location \mathbf{x}_i . During the PSO optimization process, particles are updated as flying in the latent space of CVIM, Λ , based on the velocity \mathbf{V} :

$$\Lambda_j(k+1) = \Lambda_j(k) + \mathbf{V}_j(k+1) \quad (12)$$

where j and k are the particle and iteration indexes, respectively. Velocity \mathbf{V} is a randomly-weighted average of the best position evaluated by that particle so far and the global best position among all particles:

$$\mathbf{V}_j(k+1) = \mathbf{V}_j(k) + \Upsilon_1 \cdot (\mathbf{L}_j^{best}(k) - \Lambda_j(k)) + \Upsilon_2 \cdot (\mathbf{G}^{best}(k) - \Lambda_j(k)) \quad (13)$$

where $\mathbf{V}_j(k)$ is the velocity for particle $j = 1 : ps$ at optimization step $k = 1 : K$ and ps is the population size. Υ_1 and Υ_2 are random vectors, where each entry is uniformly distributed between $[0, 1]$. $\mathbf{L}_j^{best}(k)$ is the best position in the latent space found by particle j evaluated by Equation (11), while $\mathbf{G}^{best}(k)$ is the global best position found among all particles. It is worth mentioning that the direction of each particle move is determined by comparing the current energy with the present local/global optima. Thus, while the magnitude of the move is chosen randomly, the direction is not. By doing so, PSO discourages the solution from becoming trapped in local optima by moving each particle in a way that considers both the local and global best solutions from among all current particles. All particle hypotheses are clipped to be within the range of the CVIM latent space, and the maximum velocity is restricted within $\pm 10\%$ of the range of the latent space [70]. We summarize the PSO algorithm in Algorithm 1.

3.4. GB-PSO-Based Optimization

The PSO algorithm is simple, straightforward and robust, but it suffers high computational load due to CVIM-based shape interpolation, as well as the large number of iterations that are typically needed

to obtain convergence. In some applications, the gradient is incorporated in sampling optimization to achieve a higher convergence rate [71]. In this section, we take advantage of the parametric nature of CVIM and incorporate a gradient-ascent step in the PSO to obtain a gradient-boosted PSO (GB-PSO) that overcomes these limitations by balancing between exploration and convergence with a deterministic and fast local search. Thus, GB-PSO is expected to be both more efficient and effective than the basic PSO in Algorithm 1.

Algorithm 1 PSO method.

```

1: Initialization
2: • do level-set segmentation to initialize the target location  $\mathbf{p}$ 
3: • draw particles  $\Lambda_j(0), j = 1 : ps$  randomly distributed in the latent space  $\Lambda$ , where  $ps$  is the population size
4: • evaluate  $\Lambda_j(0)$  by Equation (11) to get  $f(\Lambda_j(0))$ , set  $\mathbf{G}^{best}(0) = \arg \max_{\Lambda_j(0)} f(\Lambda_j(0))$  and  $\mathbf{L}_j^{best}(0) = \Lambda_j(0), (j = 1 : ps)$ 
5: PSO algorithm
6: for each iteration ( $k = 0 : K - 1$ ) do
7:   for each particle ( $j = 1 : ps$ ) do
8:     • calculate velocity  $\mathbf{V}_j(k + 1)$  and new particle  $\Lambda_j(k + 1)$  by Equations (13) and (12)
9:     • compute  $f(\Lambda_j(k + 1))$  by Equation (11)
10:    if  $f(\Lambda_j(k + 1)) > f(\mathbf{L}_j^{best}(k))$  then
11:      • set  $\mathbf{L}_j^{best}(k + 1) = \Lambda_j(k + 1)$ 
12:      if  $f(\Lambda_j(k + 1)) > f(\mathbf{G}^{best}(k))$  then
13:        • set  $\mathbf{G}^{best}(k + 1) = \Lambda_j(k + 1)$ 
14:      end if
15:    end if
16:  end for
17: end for
18: • obtain the final result,  $\Lambda^*$ , by selecting from  $\mathbf{G}^{best}(K)$ .
  
```

A classical gradient ascent method starts from an initial hypothesis in the search space, *i.e.*, the parameter space of CVIM denoted by Λ ; then, by computing the local gradient direction, small steps are made toward the maximum iteratively. Due to the smooth and continuous nature of CVIM, which generates the shape embedding function Φ , $f = P(\Lambda|\Omega)$ can be differentiated with respect to Λ . Beginning from some initial guesses Λ_0 , we will then update our guess iteratively along the gradient direction:

$$\Lambda^{(k+1)} = \Lambda^{(k)} - r \cdot (-\nabla f|_{\Lambda=\Lambda^{(k)}}), = \Lambda_k + r \cdot \nabla f|_{\Lambda=\Lambda^{(k)}} \quad (14)$$

where r is the learning rate that determines the step size and $\nabla f|_{\Lambda=\Lambda_k}$ is the gradient of f evaluated at the old guess. To compute $\nabla f|_{\Lambda=\Lambda_k}$, we take the derivative of f with respect to Λ by the chain rule as:

$$\frac{\partial f}{\partial \Lambda} = \left(\left(\frac{\partial f}{\partial \Phi} \right)^T \cdot \frac{\partial \Phi}{\partial \Lambda} \right)^T = \left(\frac{\partial \Phi}{\partial \Lambda} \right)^T \frac{\partial f}{\partial \Phi} \quad (15)$$

Similar to [61], the first term in Equation (15) can be written as:

$$\frac{\partial f}{\partial \Phi} = \frac{\delta_\epsilon(\Phi)(P_f - P_b)}{H_\epsilon(\Phi)P_f + (1 - H_\epsilon(\Phi))P_b} \quad (16)$$

where $\delta_\epsilon(\cdot)$ is the derivative of the Heaviside step function and P_f and P_b are defined in Equations (4) and (5). Since the latent variable $\Lambda = [\Theta^T, \alpha]^T$, so the second term in Equation (15) may be written as:

$$\frac{\partial \Phi}{\partial \Lambda} = \begin{bmatrix} \text{IDCT}(\frac{\partial \mathcal{S}_{\text{DCT}}}{\partial \Theta})^T \\ \text{IDCT}(\frac{\partial \mathcal{S}_{\text{DCT}}}{\partial \alpha})^T \end{bmatrix}^T \quad (17)$$

The CVIM-based DCT generation of \mathcal{S}_{DCT} is defined in Equation (7). From the properties of the tensor multiplication [72], we can rewrite Equation (7) as:

$$\mathcal{S}_{\text{DCT}}(\alpha, \Theta) = \mathbf{A} \times_3 \mathbf{i}(\alpha) \times_2 \Psi(\Theta) = \mathbf{A} \times_2 \Psi(\Theta) \times_3 \mathbf{i}(\alpha) \quad (18)$$

where both $\mathbf{i}(\alpha)$ and $\Psi(\Theta)$ are row vectors. The steepest ascent optimization may then be performed based on the gradients along the view and identity manifolds.

(I) Gradient along the view manifold

Let $\mathbf{B} = \mathbf{A} \times_3 \mathbf{i}(\alpha)$. From the tensor multiplication and flattening properties [72], it then follows that:

$$\mathcal{S}_{\text{DCT}}(\alpha, \Theta) = \mathbf{B} \times_2 \Psi(\Theta) = \mathbf{B}_{(2)}^T \cdot \Psi^T(\Theta) \quad (19)$$

where $\mathbf{B}_{(2)}$ is the mode-two flattened matrix of \mathbf{B} . Hence,

$$\frac{\partial \mathcal{S}_{\text{DCT}}}{\partial \Theta} = \mathbf{B}_{(2)}^T \cdot \frac{\partial \Psi^T(\Theta)}{\partial \Theta} \quad (20)$$

It then follows from Equation (9) that:

$$\frac{\partial \Psi^T(\Theta)}{\partial \Theta} = -2c[\kappa(\|\Theta - \mathbf{S}_1\|)(\Theta - \mathbf{S}_1), \dots, \kappa(\|\Theta - \mathbf{S}_{N_c}\|)(\Theta - \mathbf{S}_{N_c})]^T \quad (21)$$

where $\kappa(\|\Theta - \mathbf{S}_i\|)$ is defined in Equation (9). For the first term in Equation (17), we then have:

$$\frac{\partial \mathcal{S}_{\text{DCT}}}{\partial \Theta} = -2c \cdot \mathbf{B}_{(2)}^T \cdot [\kappa(\|\Theta - \mathbf{S}_1\|)(\Theta - \mathbf{S}_1), \dots, \kappa(\|\Theta - \mathbf{S}_{N_c}\|)(\Theta - \mathbf{S}_{N_c})]^T \quad (22)$$

(II) Gradient along the identity manifold

Let $\mathbf{C} = \mathbf{A} \times_2 \Psi(\Theta)$. From Equation (18), we have then that:

$$\mathcal{S}_{\text{DCT}}(\alpha, \Theta) = \mathbf{C} \times_3 \mathbf{i}(\alpha) = \mathbf{C}_{(3)}^T \cdot \mathbf{i}^T(\alpha) \quad (23)$$

so:

$$\frac{\partial \mathcal{S}_{\text{DCT}}}{\partial \alpha} = \mathbf{C}_{(3)}^T \cdot \frac{\partial \mathbf{i}^T(\alpha)}{\partial \alpha} \quad (24)$$

Since $\mathbf{i}(\alpha)$ is the piecewise polynomial interpolation function, which is differentiable between any two given data points, it follows from Equation (8) that:

$$\frac{\partial \mathbf{i}^T(\alpha)}{\partial \alpha} = 3\mathbf{a}_m^T(\alpha - \alpha_m)^2 + 2\mathbf{b}_m^T(\alpha - \alpha_m) + \mathbf{c}_m^T, \quad \alpha \in [\alpha_m, \alpha_{m+1}) \quad (25)$$

Thus, we obtain finally:

$$\frac{\partial \mathcal{S}_{\text{DCT}}}{\partial \alpha} = \mathbf{C}_{(3)}^T \cdot [3\mathbf{a}_m^T(\alpha - \alpha_m)^2 + 2\mathbf{b}_m^T(\alpha - \alpha_m) + \mathbf{c}_m^T], \quad \alpha \in [\alpha_m, \alpha_{m+1}) \quad (26)$$

which together with Equation (22) provides the explicit formulation for both terms in Equation (17).

(III) Gradient in the latent space

From Equations (16) and (17), $\nabla f|_{\Lambda=\Lambda_k}$ may be rewritten as:

$$\nabla f|_{\Lambda=\Lambda_k} = \begin{bmatrix} \text{IDCT}\left(\frac{\partial S_{\text{DCT}}}{\partial \Theta_k}\right)^T \\ \text{IDCT}\left(\frac{\partial S_{\text{DCT}}}{\partial \alpha_k}\right)^T \end{bmatrix} \frac{\delta_\epsilon(\Phi_k)(P_f - P_b)}{H_\epsilon(\Phi_k)P_f + (1 - H_\epsilon(\Phi_k))P_b} \quad (27)$$

where $\alpha \in [\alpha_m, \alpha_{m+1})$, $\Phi_k = \Phi(\Lambda_k)$ and $P(\mathbf{x}|\Lambda_k, \mathbf{y})$ are defined in Equations (3) and (10), while $\frac{\partial S_{\text{DCT}}}{\partial \Theta_k}$ and $\frac{\partial S_{\text{DCT}}}{\partial \alpha_k}$ are defined in Equations (22) and (26).

Algorithm 2 The gradient-boosted (GB)-PSO method.

```

1: Initialization
2: • refer to Algorithm 1 Lines 1 ~ 4
3: GB-PSO step
4: for each iteration step ( $k = 0 : K - 1$ ) do
5:   for each particle ( $j = 1 : ps$ ) do
6:     • refer to Algorithm 1 Lines 8 ~ 15
7:   end for
8:   Gradient Ascent Local Search
9:   • set  $\Lambda^0 = \mathbf{G}^{best}(k + 1)$  for gradient ascent local search;
10:  for each gradient ascent step ( $l = 1 : pl$ ) do
11:    • calculate  $\nabla f|_{\Lambda=\Lambda^{l-1}}$ 
12:    if the gradient is not significant then
13:      break
14:    end if
15:    • draw a sample for step size  $r$ 
16:    • update  $\Lambda^l = \Lambda^{l-1} + r \cdot \nabla f|_{\Lambda=\Lambda^{l-1}}$  according to Equation (27)
17:  end for
18:  • set  $\mathbf{G}^{best}(k + 1) = \Lambda^l$ ;
19: end for
20: • obtain the final result,  $\Lambda^*$ , by selecting from  $\mathbf{G}^{best}(K)$ .

```

As suggested in [73], a uniformly-distributed random step size (r) could be used for the steepest ascent method, which turned out to be effective in practice. In practice, r is uniformly distributed between $[\frac{\pi}{90}, \frac{\pi}{15}]$. In the GB-PSO method, the standard PSO is involved as the first step, then the global optimum ($\mathbf{G}^{best}(k + 1)$) is updated by the gradient ascent method, which helps the next round PSO converge fast by improving velocity estimation. Thus, in the GB-PSO method, the total number of iterations required could be dramatically reduced compared with PSO. The computational load of the additional steps in GB-PSO is negligible due to two reasons: (1) the analytical nature of the energy function makes the gradient computation very efficient for the present global solution $\mathbf{G}^{best}(k + 1)$ that is to be shared by all particles in the next iteration; (2) the update along the gradient direction is done analytically according to Equation (27), and there is a maximum number of gradient ascent iterations (*i.e.*, $pl = 20$ in this

work) and a check of the current gradient value to determine if additional moves are necessary. In our experiment, we found that the actual number of the steps along the gradient direction is often much less than pl (around 10), which confirms that the solution of gradient-based search is in the proximity of a local optimum. We summarize the GB-PSO method in Algorithm 2.

4. Experimental Result

In this work, our interest is to develop a general model-driven ATR algorithm where no IR data are used for training and no prior feature extraction is needed from the IR data, unlike most traditional methods that heavily rely on the quality of the training data, as well as feature extraction. We have conducted two comparative studies to evaluate the performance of the proposed algorithms. First, we have involved five comparable algorithms, including the proposed PSO and GB-PSO algorithms, all of which apply CVIM for shape modeling. Second, we also compared our algorithms with several recent ATR algorithms, including two SRC-based approaches [45,46] and our previously-proposed ATR algorithm, which involves a joint view-identity manifold (JVIM) for target tracking and recognition [37]. The purpose of the first study is to validate the advantages of “implicit shape matching” over “explicit shape matching”, as well as the efficiency of GB-PSO over PSO. That of the second study is to demonstrate the effectiveness of our new ATR algorithms compared with the recent ones in a similar experimental setting. It was reported in [14] that SRC-based methods can achieve state-of-the-art performance. In the following, we will first discuss the experimental setup shared by two comparative studies along with the metrics used for performance evaluation. Then, we present the two comparative studies one-by-one in detail.

4.1. Experimental Setup

Similar to [33], we selected six 3D CAD models for each of the six target classes for CVIM training (36 models in total; Figure 4): APCs (armored personnel carriers), tanks, pick-ups, sedans, vans and SUVs. We considered elevation angles in $0^\circ \sim 40^\circ$ and azimuth angles in $0^\circ \sim 360^\circ$, with 10° and 12° intervals along the elevation and azimuth angles, respectively, on the view manifold, giving 150 multi-view shapes for each target. We also adopted a DCT-based shape descriptor [36], which facilitates CVIM learning and shape inference. All experiments were performed against the SENSIAC database [5], which provides a large collection of mid-wave IR and visible data depicting seven military targets and two civilian vehicles. We used 24 mid-wave (23 night-time and 1 day-time) IR sequences captured from 8 targets (Figure 7) at 1 km, 2 km and 3 km. In each sequence, there is a civilian or military vehicle traversing a closed-circular path with a diameter of 100 m. We selected 100 frames from each sequence by down-sampling each sequence that has 1800 frames originally, where the aspect angle ranges from 0° to 360° with around a $5^\circ \sim 10^\circ$ interval; so in total, there are 2400 frames used for evaluation. The SENSIAC database also provides a rich amount of metadata, which can be used for performance evaluation, such as the aspect angle of the target, the field of view and the 2D bounding box of the target in each frame. Since we mainly focus on the recognition rather than detection, we also generated our target chips with the help of target 2D locations from this metadata (averaging around

$50 \times 30 = 1500$ pixels at 1 km, $25 \times 14 = 350$ pixels at 2 km and $15 \times 10 = 150$ pixels at 3 km) in our experiments.



Figure 7. All eight targets we used for algorithm evaluation were from the SENSIAC database.

4.2. The First Comparative Study: CVIM-Based Approaches

This study compares five CVIM-based algorithms, which involve different segmentation and optimization techniques. Specifically, Method I uses background subtraction [74] to pre-segment a target-of-interest. This method is only suitable for a stationary sensor platform. Method II applies level set segmentation without a shape prior [61]. Both Method I and Method II need explicit shape matching, which involves Markov Chain Monte Carlo (MCMC)-based CVIM inference after segmentation to accomplish ATR [75]. Method III applies a multi-threaded MCMC-based inference technique to jointly optimize over CVIM in a level set by involving implicit shape matching without target pre-segmentation. It was shown in [75] that Method III significantly outperforms the first two, but it suffers from high computational complexity due to the MCMC-based shape inference. PSO and GB-PSO are referred to as Methods IV and V, respectively. The computational time for each ATR chip (50×30 pixels) for five algorithms is around 10, 14, 22, 15 and 6 s, respectively, using an un-optimized MATLAB code on a PC with a Quad-core CPU (2.5 GHZ).

We evaluate these five algorithms with respect to: (1) the accuracy of pose estimation (*i.e.*, the aspect angle); (2) the 2D pixel location errors between the segmented shape and the ground truth bounding box; (3) the recognition accuracy in terms of six major target classes; and (4) the sensor-target distance (*i.e.*, range, computed by scaling factors) errors in meters. To examine the robustness of our algorithms, we analyze (5) the recognition accuracy *versus* three related factors, *i.e.*, the contrast of image chips [76], the foreground/background χ^2 histogram distance [77] based on the segmentation results and the aspect angle. The chip contrast and χ^2 histogram distance indicate the IR image quality and the target visibility, respectively. Similar to [78–81], we also evaluate the overlap ratio between the estimated bounding box (derived from the segmentation result) and the ground truth bounding box (available from ground-truth data) (6), which is a simple, yet effective and widely-accepted way to quantify the segmentation performance. Furthermore, we manually created the ground truth segmentation masks from five randomly-selected frames per IR sequence, so that we can compute the overlap ratio between the segmentation results with the ground-truth masks (7). Moreover we will show the capability of the

proposed algorithm (GB-PSO) for sub-class recognition, *i.e.*, the specific target type within a class, even if the exact target type is not in the training data.

4.2.1. Pose Estimation Results

Table 1 reports aspect angle error (pose estimation) results for all five tested methods along with the 2D pixel error and 3D range error in the predefined 3D camera coordinate system (given in the metadata). We can see clearly that both Methods IV and V can achieve moderate, significant and slight improvements over Method I (background subtraction for segmentation), Method II (level set segmentation without shape prior) and Method III (MCMC-based CVIM inference), respectively. Although Methods IV and V do not provide a significant improvement in pose and location estimation performance compared to Method III, they provide similar performance at a greatly reduced computational complexity. Numerical results from PSO and GB-PSO are comparable to each other. However, Figure 8 shows that GB-PSO converges nearly three-times faster than the PSO, demonstrating the value of gradient boosting in the CVIM latent space.

Table 1. Pose and location estimation errors for all five tested methods averaged over SENSIAC mid-wave IR sequences for each target to sensor distances depicting eight difference targets (Method I/Method II/Method III/Method IV/Method V).

	2D Pixel Error (pixels)	Aspect Angle Error (°)	Range Error (m)
1 km	2.8/3.1/1.9/2.1/ 1.9	17.2/17.9/15.1/ 14.8 /15.1	25.1/27.8/24.2/ 24.1 /24.5
2 km	2.9/3.4/2.3/2.4/ 2.1	21.2/25.2/18.7/18.2/ 17.1	39.1/38.2/33.8/32.8/ 32.6
3 km	2.5/3.8/2.2/ 1.8 /2.0	26.1/27.5/21.7/21.9/ 20.5	43.5/48.3/40.2/ 40.1 /41.1

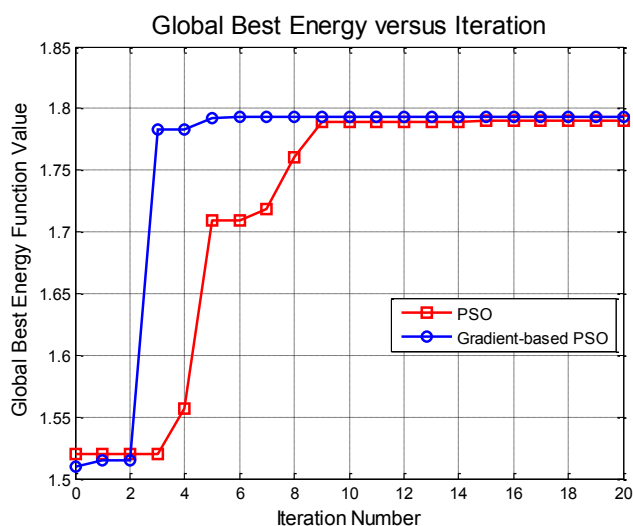


Figure 8. Convergence plots for PSO and GB-PSO in one example.

4.2.2. Target Recognition Results

The recognition results are computed based on the percentage of frames where the target class is correctly classified. As shown in Table 2, both PSO (Method IV) and GB-PSO (Method V) generally

achieve modest performance gains over Method I–III, while GB-PSO does so with a significantly reduced computational complexity compared to all four of the other methods. Furthermore, Figure 9 shows some sub-class recognition results for eight 1-km IR images. The sub-class recognition can be achieved via CVIM by finding the two closest training target types along the identity manifold. Since the training data only have the BTR70 model, we find that we can recognize the BTR70 at the sub-class level most of the time. Interestingly, we can see that T72, BMP2 and 2S3 are also recognized as T62, BMP1 and AS90, respectively, which are the closest sub-class target types available in our training data.

Table 2. Overall recognition and segmentation results of five methods (I/II/III/IV/V).

	Average Recognition Accuracy (%)	Bounding Box Overlap (%)
1 km	81/78/85/ 86 /85	85.2/82.9/88.1/88.6/ 88.9
2 km	71/64/73/75/ 76	75.6/74.1/79.5/ 79.8 /79.2
3 km	69/62/70/72/ 73	67.7/65.5/70.1/70.9/ 71.6
	Segmentation Mask Overlap (%)	Fore/Background χ^2 Histogram Distance
1 km	79.3/83.2/83.5/ 83.8 /83.6	0.30/0.32/0.34/0.34/ 0.35
2 km	72.3/73.9/79.8/79.6/ 80.1	0.26/0.25/0.29/0.28/ 0.31
3 km	63.8/67.2/68.8/ 69.3 /69.1	0.20/0.23/0.25/ 0.26 /0.25

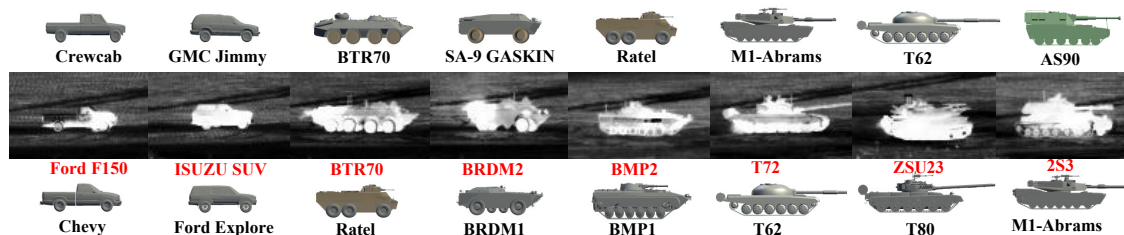


Figure 9. Some sub-class recognition results under 1000 m. The first row and third row show the closest training vehicles along the identity manifold in the CVIM, and the middle row presents the original IR chips.

We also summarize recognition results from GB-PSO vs. the chip contrast, foreground/background histogram distance and aspect angle in Figure 10. It is shown in Figure 10a that our algorithm performs well for most chips with reasonable contrast and tends to deteriorate for chips with a very low contrast, which is usually associated with poor image quality (e.g., day-time IR imagery). As illustrated in Figure 10b, the foreground/background χ^2 histogram distance is strongly related to the recognition accuracy. This is because the χ^2 distance is related to the target visibility and can also quantify the segmentation quality. When segmentation results are good with large background/foreground separation (large χ^2 distance values), the recognition accuracies are usually high, which also imply good target segmentations. Furthermore, the aspect angle is a key factor that affects the recognition performance. As shown in Figure 10c, the highest accuracy occurs around the side views (90° and 270°) when the targets are most recognizable. Most failed cases are around 0° (or 360°) (frontal views) and 180° (rear views), when it is hard to differentiate different targets due to the high shape ambiguity. Since we only

use the shape information here, a more advanced and informative target appearance representation that involves intensity and other features could make ATR performance more robust to aspect angles.

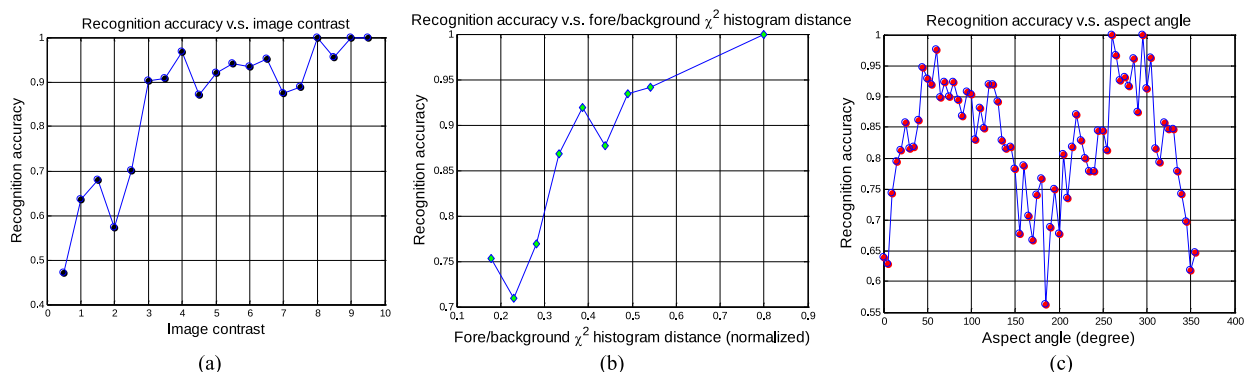


Figure 10. Robustness analysis of GB-PSO over all 2400 chips. (a) Recognition accuracy *versus* the chip contrast; (b) recognition accuracy *versus* foreground/background histogram distances; (c) recognition accuracy *versus* the aspect angles.

4.2.3. Target Segmentation Results

Table 2 shows the target segmentation results in terms of the bounding box overlap, the segmentation mask overlap and the foreground/background χ^2 histogram distance. Both PSO and GB-PSO outperform Methods I and II, while performing comparably to Method III at a lower computational complexity. Figure 11 shows some snapshots of the original IR imagery of eight targets under the 1-km range, along with the manually-cropped segmentation masks, the results of the background subtraction segmentation, the level set segmentation without a shape prior and the PSO method, respectively. It may be seen that the CVIM shape prior drives the segmentation to a semantically more meaningful shape compared to Methods I and II, where a shape prior is not involved.

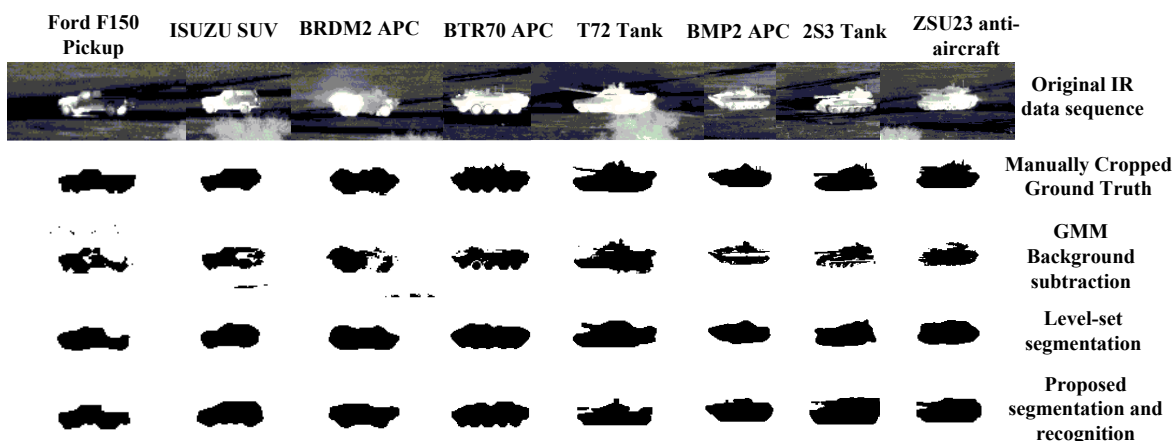


Figure 11. Snapshot of the segmentation results. From the first row to the last: original IR frame, manually-cropped segmentation masks, results of background subtraction segmentation, level set segmentation without shape prior and the final segmentation and recognition results with CVIM shape prior (Method IV) interpolated from the CVIM.

Some snapshots of segmentation results along with pose estimation and recognition results of Method V (GB-PSO) are shown in Figure 12. It is found that GB-PSO is robust to background clutter and engine smoke, and the frontal/rear views may pose some challenge. For BRDM2 APC, we see a heat spot near the tail, which changes the target appearance significantly, but we can still recognize it as an APC. However, for targets of near front/rear views, although the segmentation results are still acceptable, the recognition results are often wrong. For example, the BMP2 APC was misrecognized as the M60 tank in the front view. By closely observing the IR appearances of BMP2 APC, we find that this particular APC does indeed look similar to a tank when viewed frontally. This can also be explained by our identity manifold topology learned by *class-constrained shortest-closed-path*, where the BMP2 stays closest to tanks along the identity manifold among all APCs, as shown in Figure 4.

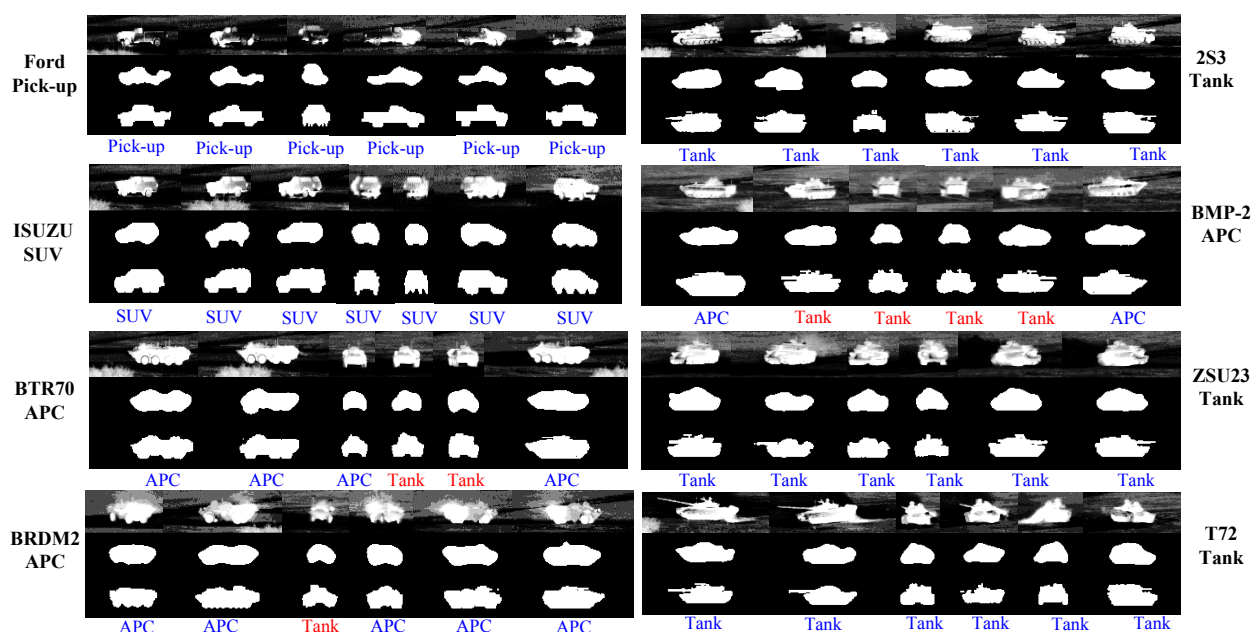


Figure 12. Some GB-PSO results for pose estimation and recognition results of eight targets under a 1-km target-sensor range from the SENSIAC database. For each target, from the first row to the third: original IR frame, level set segmentation-based initialization and the final pose estimation and recognition results from GB-PSO. Recognition results are denoted in blue if correct and red if wrong below each chip.

A similar case happens to BTR70 APC. Moreover, the proposed algorithm as realized in Methods IV and V performs poorly against long-range day-time IR data (3 km), where the foreground/background contrast is low and the target is small. This is illustrated in Figure 13, where the 2S3 tank is misclassified as an APC in several frames. As we already mentioned, a more powerful appearance representation is needed to handle challenging cases of this type.

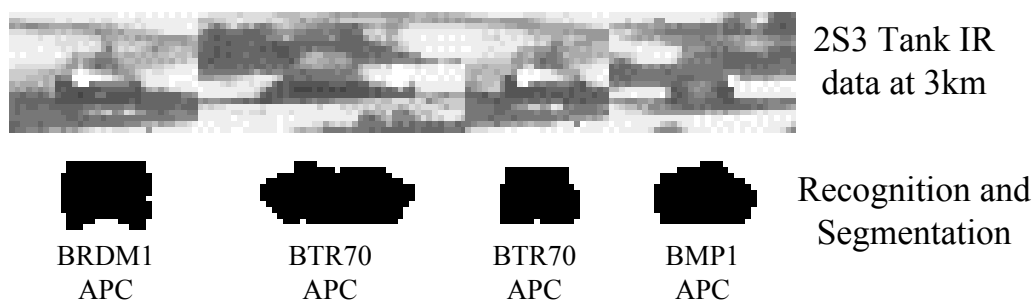


Figure 13. Some failed cases of the 2S3 tank in a day-time sequence at 3 km.

4.3. Comparative Study: Recent ATR Methods

This comparative study includes three recent ATR algorithms that are compared against PSO and GB-PSO. Specifically, we applied the gradient-based optimization technique discussed in [37] to apply JVIM (learned from the same set of training shapes as CVIM) for image-based ATR where level set segmentation without a shape prior is used for initialization. In addition, we have also implemented the SRC-based algorithm [14] and the multi-attribute Lasso with group constraint (MALGC) [46], which is an extended SRC approach by taking advantage of the attribute information (*i.e.*, angles) during sparse optimization. Both SRC algorithms require a dictionary that includes training shapes also used for CVIM learning. The input is the level set segmentation output from an IR chip. We also use the segmentation results from GB-PSO for SRC-based ATR (namely SRC-GB-PSO) to see the effect of good target segmentation. The computational time (in the same computational setting as before) for four implementations is around 4 (JVIM), 16 (SRC), 20 (MALGC) and 18 (SRC-GB-PSO) s, compared with 15 and 6 s for PSO and GB-PSO, respectively. We compare six ATR algorithms in Table 3.

Table 3. The performance comparison with recent ATR methods in terms of the recognition accuracy and the aspect angle error (joint view-identity manifold (JVIM)/sparse representation-based classification (SRC)/multi-attribute Lasso with group constraint (MALGC)/SRC-GB-PSO/PSO/GB-PSO).

Ranges	Average Recognition Accuracy (%)	Aspect Angle Error (°)
1 km	82/83/83/85/ 86 /85	16.9/15.6/15.4/15.3/ 14.8 /15.1
2 km	69/72/73/75/75/ 76	22.3/20.1/19.9/17.8/18.2/ 17.1
3 km	65/69/70/72/72/ 73	26.8/24.5/23.9/21.1/21.9/ 20.5

We have the following observations and discussion according to Table 3.

- The JVIM model unifies the view and identity manifolds in one latent space for more accurate shape modeling than CVIM, which involves separate view and identity manifolds [82], and it is especially suitable for target tracking due to the unified and smooth shape manifold [37]. However, JVIM has a similar multi-modal problem as CVIM that makes the gradient-based optimization often trapped in local minima, as reflected by relatively poor results in Table 3. It may not be

efficient to apply sample-based approaches (e.g., PSO or MCMC) to JVIM optimization due to the fact that its joint manifold structure will require a large number of samples to ensure effective sampling. The reason that JVIM shows promising results in target tracking is because the dynamic modeling involved greatly facilitates sequential state estimation. In the case of image-based ATR, CVIM shows some advantages over JVIM due to its simpler structure.

- Both SRC and MALGC methods show reasonable performance by only using shapes segmented by level set for ATR. Especially for the range of 1 km when target segmentation is likely more accurate, all algorithms are very comparable. It is observed that MALGC is slightly better than SRC by utilizing the angle information in sparse optimization. More interestingly, we can see that better target segmentation results via GB-PSO moderately improve the SRC's performance (the fourth algorithm, GB-PSO-SRC). The computational complexity of SRC and MALGC is comparable with PSO and much higher than that of GB-PSO.
- It is shown that the proposed PSO and GB-PSO algorithms are comparable, both of which are better than the others in all cases. Although the improvement of GB-PSO over the other two SRC approaches is moderate, it does offer several advantages: (1) it is computationally efficient with only a 30%–40% computational load; (2) target segmentation is not required and can be considered as a byproduct of ATR; and (3) the proposed GB-PSO algorithm is a model-driven approach that does not require real-world training data, and it has potential to be combined with other data-driven approaches to further improve ATR performance.

5. Conclusions

In this paper, we have integrated a shape generative model (CVIM) into a probabilistic level set framework to implement joint target recognition, segmentation and pose estimation in IR imagery. Due to the multi-modal nature of the optimization problem, we first implemented a PSO-based method to jointly optimize CVIM-based implicit shape matching and level set segmentation and then developed the gradient-boosted PSO (GB-PSO) algorithm to further improve the efficiency by taking advantage of the analytical and differentiable nature of CVIM. We have conducted two comparative studies on the recent SENSIAC ATR database to demonstrate the advantages of the two PSO-based algorithms. The first study involves five methods where CVIM is optimized by either explicit shape matching or MCMC-based implicit shape matching. GB-PSO and PSO are shown to be more effective than other methods. Moreover, GB-PSO was also shown to be more efficient than PSO with a much improved convergence rate due to the gradient-driven technique. The second study includes a few recent ATR algorithms for performance evaluation. It is shown that the proposed GB-PSO algorithm moderately outperforms other recent ATR algorithms at a much lower computational load. The proposed framework could be further extended to incorporate new appearance features or effective optimization to deal with more challenging ATR problems.

Acknowledgments

This work was supported in part by the U.S. Army Research Laboratory and the U.S. Army Research Office under Grant W911NF-08-1-0293 and the Oklahoma Center for the Advancement of Science and Technology (OCAST) under Grant HR12-30.

Author Contributions

Liangjiang Yu developed the algorithm, conducted the experiments and participated in writing the paper. Jiulu Gong contributed to the theoretical analysis and algorithm development and contributed to the writing. Guoliang Fan and Joseph Havlicek planned and supervised the research, analyzed the results and participated in writing the paper.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Bhanu, B. Automatic target recognition: State of the art survey. *IEEE Trans. Aerosp. Electron. Syst.* **1986**, *AES-22*, 364–379.
2. Bhanu, B.; Jones, T. Image understanding research for automatic target recognition. *IEEE Aerosp. Electron. Syst. Mag.* **1993**, *8*, 15–23.
3. Dudgeon, D.E.; Lacoss, R.T. An overview of automatic target recognition. *Linc. Lab. J.* **1993**, *6*, 3–10.
4. Ratches, J.A. Review of current aided/automatic target acquisition technology for military target acquisition tasks. *Opt. Eng.* **2011**, *50*, 072001-1–072001-8.
5. Military Sensing Information Analysis Center (SENSIAC), 2008. Available online: <http://www.sensiac.org/external/index.jsf> (accessed on 28 April 2015).
6. Yilmaz, A.; Javed, O.; Shah, M. Object tracking: A survey. *ACM Comput. Surv.* **2006**, *38*, doi:10.1145/1177352.1177355.
7. Srinivas, U.; Monga, V.; Riasati, V. A comparative study of basis selection techniques for automatic target recognition. In Proceedings of the 2012 IEEE Radar Conference, Atlanta, GA, USA, 7–11 May 2012; pp. 711–713.
8. Bhatnagar, V.; Shaw, A.; Williams, R. Improved automatic target recognition using singular value decomposition. In Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, Seattle, WA, USA, 12–15 May 1998; Volume 5, pp. 2717–2720.
9. Olson, C.; Huttenlocher, D. Automatic target recognition by matching oriented edge pixels. *IEEE Trans. Image Process.* **1997**, *6*, 103–113.
10. Casasent, D.P.; Smokelin, J.S.; Ye, A. Wavelet and Gabor transforms for detection. *Opt. Eng.* **1992**, *31*, 1893–1898.
11. Grenander, U.; Miller, M.; Srivastava, A. Hilbert-Schmidt lower bounds for estimators on matrix Lie groups for ATR. *IEEE Trans. Pattern Anal. Mach. Intell.* **1998**, *20*, 790–802.

12. Wang, L.C.; Der, S.Z.; Nasrabadi, N.M.; Rizvi, S.A. Automatic target recognition using neural networks. *Proc. SPIE* **1998**, *3466*, 278–289.
13. Xiong, W.; Cao, L. Automatic target recognition based on rough set-support vector machine in SAR images. In Proceedings of the International Joint Conference on Computational Sciences, Optimization, Sanya, Hainan, China, 24–26 April 2009; Volume 1, pp. 489–491.
14. Patel, V.M.; Nasrabadi, N.M.; Chellappa, R. Sparsity-motivated automatic target recognition. *Appl. Opt.* **2011**, *50*, 1425–1433.
15. Felzenszwalb, P.; Girshick, R.; McAllester, D.; Ramanan, D. Object detection with discriminatively trained part-based models. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 1627–1645.
16. Dong, L.; Yu, X.; Li, L.; Hoe, J. HOG based multi-stage object detection and pose recognition for service robot. In Proceedings of the International Conference on Control, Automation Robotics Vision (ICARCV), Singapore, 7–10 December 2010; pp. 2495–2500.
17. Lowe, D. Object recognition from local scale-invariant features. In Proceedings of the 7th IEEE International Conference on Computer Vision, Kerkyra, Greece, 20–27 September 1999; Volume 2, pp. 1150–1157.
18. Johnson, A.; Hebert, M. Using spin images for efficient object recognition in cluttered 3D scenes. *IEEE Trans. Pattern Anal. Mach. Intell.* **1999**, *21*, 433–449.
19. Opelt, A.; Pinz, A.; Zisserman, A. Learning an alphabet of shape and appearance for multi-class object detection. *Int. J. Comput. Vis.* **2008**, *80*, 16–44.
20. Belongie, S.; Malik, J.; Puzicha, J. Shape matching and object recognition using shape contexts. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 509–522.
21. Mo, Q.; Draper, B.A. Semi-nonnegative matrix factorization for motion segmentation with missing data. In *Computer Vision—ECCV 2012*; Lecture Notes in Computer Science; Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C., Eds.; Springer: Berlin Heidelberg, Germany, 2012; Volume 7578, pp. 402–415.
22. Heikkila, M.; Pietikainen, M. A texture-based method for modeling the background and detecting moving objects. *IEEE Trans. Pattern Anal. Mach. Intell.* **2006**, *28*, 657–662.
23. Ricaurte, P.; Chilán, C.; Aguilera-Carrasco, C.A.; Vintimilla, B.X.; Sappa, A.D. Feature point descriptors: Infrared and visible spectra. *Sensors* **2014**, *14*, 3690–3701.
24. Nguyen, H.V.; Porikli, F. Support vector shape: A classifier-based shape representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 970–982.
25. Saidi, M.N.; Toumi, A.; Hoeltzner, B.; Khenchaf, A.; Aboutajdine, D. Aircraft target recognition: A novel approach for features extraction from ISAR images. In Proceedings of the Radar International Conference—Surveillance for a Safer World, Bordeaux, France, 12–16 October 2009; pp. 1–5.
26. Kass, M.; Witkin, A.; Terzopoulos, D. Snakes: Active contour models. *Int. J. Comput. Vis.* **1988**, *1*, 321–331.
27. Xu, C.; Prince, J. Snakes, shapes, and gradient vector flow. *IEEE Trans. Image Process.* **1998**, *7*, 359–369.
28. Paragios, N.; Deriche, R. Geodesic active contours and level sets for the detection and tracking of moving objects. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 266–280.

29. Unal, G.B.; Krim, A.H. Segmentation and target recognition in SAR imagery using a level-sets-multiscale-filtering technique. *Proc. SPIE* **2001**, *4391*, 370–379.
30. Chen, M.; Cai, J. An on-line learning tracking of non-rigid target combining multiple-instance boosting and level set. *Proc. SPIE* **2013**, doi:10.1117/12.2031170.
31. Leventon, M.; Grimson, W.; Faugeras, O. Statistical shape influence in geodesic active contours. In Proceedings of the IEEE Conference on Computer Vision, Pattern Recognition, Hilton Head Island, SC, USA, 13–15 June 2000; pp. 316–323.
32. Tsai, A.; Yezzi, A., J.; Wells, W.; Tempany, C.; Tucker, D.; Fan, A.; Grimson, W.; Willsky, A. A shape-based approach to the segmentation of medical imagery using level sets. *IEEE Trans. Med. Imag.* **2003**, *22*, 137–154.
33. Venkataraman, V.; Fan, G.; Yu, L.; Zhang, X.; Liu, W.; Havlicek, J. Automated target tracking and recognition using coupled view and identity manifolds for shape representation. *EURASIP J. Adv. Signal Process.* **2011**, *2011*, 1–17.
34. Lee, C.; Elgammal, A. Modeling View and Posture Manifolds for Tracking. In Proceedings of the IEEE International Conference on Computer Vision, Rio de Janeiro, Brazil, 14–21 October 2007.
35. Prisacariu, V.; Reid, I. Nonlinear shape manifolds as shape priors in level set segmentation and tracking. In Proceedings of the IEEE Conference on Computer Vision, Pattern Recognition, Colorado Springs, CO, USA, 20–25 June 2011; pp. 2185–2192.
36. Prisacariu, V.; Reid, I. Shared shape spaces. In Proceedings of the IEEE International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; pp. 2587–2594.
37. Gong, J.; Fan, G.; Yu, L.; Havlicek, J.P.; Chen, D.; Fan, N. Joint target tracking, recognition and segmentation for infrared Imagery using a shape manifold-based level set. *Sensors* **2014**, *14*, 10124–10145.
38. Kennedy, J.; Eberhart, R. Particle swarm optimization. In Proceedings of the IEEE International Conference on Neural Networks, Perth, WA, USA, 27 November–1 December 1995; Volume 4, pp. 1942–1948.
39. Hornik, K.; Stinchcombe, M.; White, H. Multilayer feedforward networks are universal approximators. *Neural Netw.* **1989**, *2*, 359–366.
40. Mirelli, V.; Rizvi, S.A. Automatic target recognition using a multilayer convolution neural network. *Proc. SPIE* **1996**, *2755*, 106–125.
41. Wang, L.; Der, S.; Nasrabadi, N. Automatic target recognition using a feature-decomposition and data-decomposition modular neural network. *IEEE Trans. Image Process.* **1998**, *7*, 1113–1121.
42. Chan, L.; Nasrabadi, N.; Mirelli, V. Multi-stage target recognition using modular vector quantizers and multilayer perceptrons. In Proceedings of the IEEE Conference on Computer Vision, Pattern Recognition, San Francisco, CA, USA, 18–20 June 1996; pp. 114–119.
43. Chan, L.A.; Nasrabadi, N.M. Automatic target recognition using vector quantization and neural networks. *Opt. Eng.* **1999**, *38*, 2147–2161.
44. Christiansen, P.; Steen, K.A.; Jørgensen, R.N.; Karstoft, H. Automated detection and recognition of wildlife using thermal cameras. *Sensors* **2014**, *14*, 13778–13793.
45. Wright, J.; Yang, A.Y.; Ganesh, A.; Sastry, S.S.; Ma, Y. Robust face recognition via sparse representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2009**, *31*, 210–227.

46. Chiang, C.K.; Su, T.F.; Yen, C.; Lai, S.H. Multi-attribute sparse representation with group constraints for face recognition under different variations. In Proceedings of the 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), Shanghai, China, 22–26 April 2013; pp. 1–6.
47. Li, Z.Z.; Chen, J.; Hou, Q.; Fu, H.X.; Dai, Z.; Jin, G.; Li, R.Z.; Liu, C.J. Sparse representation for infrared dim target detection via a discriminative over-complete dictionary learned online. *Sensors* **2014**, *14*, 9451–9470.
48. Li, X.; Guo, R.; Chen, C. Robust pedestrian tracking and recognition from FLIR video: A unified approach via sparse coding. *Sensors* **2014**, *14*, 11245–11259.
49. Srivastava, A. Bayesian filtering for tracking pose and location of rigid targets. *Proc. SPIE* **2000**, *4052*, 160–171.
50. Rosenhahn, B.; Brox, T.; Weickert, J. Three-dimensional shape knowledge for joint image segmentation and pose tracking. *Int. J. Comput. Vis.* **2007**, *73*, 243–262.
51. Prisacariu, V.; Reid, I. PWP3D: Real-time segmentation and tracking of 3D objects. In Proceedings of the 20th British Machine Vision Conference, London, UK, 7–10 September 2009.
52. Sadjadi, F.A. Automatic object recognition: critical issues and current approaches. *Proc. SPIE* **1991**, *1471*, 303–313.
53. Aull, A.; Gabel, R.; Goblick, T. Real-time radar image understanding: A machine intelligence approach. *Linc. Lab. J.* **1992**, *5*, 195–222.
54. Liebelt, J.; Schmid, C.; Schertler, K. Viewpoint-independent object class detection using 3D Feature Maps. In Proceedings of the IEEE Conference on Computer Vision, Pattern Recognition, Anchorage, AK, USA, 23–28 June 2008.
55. Khan, S.; Cheng, H.; Matthies, D.; Sawhney, H. 3D model based vehicle classification in aerial imagery. In Proceedings of the IEEE Conference on Computer Vision, Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; pp. 1681–1687.
56. Toshev, A.; Makadia, A.; Daniilidis, K. Shape-based object recognition in videos using 3D synthetic object models. In Proceedings of the IEEE Conference on Computer Vision, Pattern Recognition, Miami Beach, FL, USA, 20–25 June 2009; pp. 288–295.
57. Paravati, G.; Esposito, S. Relevance-based template matching for tracking targets in FLIR imagery. *Sensors* **2014**, *14*, 14106–14130.
58. Chari, S.K.; Halford, C.E.; Jacobs, E. Human target identification and automated shape based target recognition algorithms using target silhouette. *Proc. SPIE* **2008**, *6941*, 69410B-1–69410B-9.
59. Xiao, J.; Shah, M. Automatic target recognition using multi-view morphing. *Proc. SPIE* **2004**, *5426*, 391–399.
60. Chahooki, M.; Charkari, N. Learning the shape manifold to improve object recognition. *Mach. Vis. Appl.* **2013**, *24*, 33–46.
61. Bibby, C.; Reid, I. Robust real-time visual tracking using pixel-wise posteriors. In *Computer Vision—ECCV 2008*; Lecture Notes in Computer Science; Forsyth, D., Torr, P., Zisserman, A., Eds.; Springer: Berlin Heidelberg, Germany, 2008; Volume 5303, pp. 831–844.
62. Cremers, D.; Rousson, M.; Deriche, R. A review of statistical approaches to level set segmentation: Integrating color, texture, motion and shape. *Int. J. Comput. Vis.* **2007**, *72*, 195–215.

63. Jebara, T. Images as bags of pixels. In Proceedings of the IEEE International Conference on Computer Vision, Nice, France, 13–16 October 2003; pp. 265–272.
64. Cremers, D. Dynamical statistical shape priors for level set-based tracking. *IEEE Trans. Pattern Anal. Mach. Intell.* **2006**, *28*, 1262–1273.
65. Li, C.; Xu, C.; Gui, C.; Fox, M. Distance regularized level set evolution and its application to image segmentation. *IEEE Trans. Image Process.* **2010**, *19*, 3243–3254.
66. De Falco, I.; Della Cioppa, A.; Tarantino, E. Evaluation of particle swarm optimization effectiveness in classification. In *Fuzzy Logic and Applications; Lecture Notes in Computer Science*; Bloch, I., Petrosino, A., Tettamanzi, A.G., Eds.; Springer: Berlin Heidelberg, Germany, 2006; Volume 3849, pp. 164–171.
67. Omranpour, H.; Ebadzadeh, M.; Shiry, S.; Barzegar, S. Dynamic particle swarm optimization for multimodal function. *IAES Int. J. Artif. Intell.* **2012**, *1*, 1–10.
68. Ye, L.; Yang, M.; Xu, L.; Zhuang, X.; Dong, Z.; Li, S. Nonlinearity analysis and parameters optimization for an inductive angle sensor. *Sensors* **2014**, *14*, 4111–4125.
69. Han, W.; Xu, J.; Wang, P.; Tian, G. Defect profile estimation from magnetic flux leakage signal via efficient managing particle swarm optimization. *Sensors* **2014**, *14*, 10361–10380.
70. Evers, G.; Ben Ghalia, M. Regrouping particle swarm optimization: A new global optimization algorithm with improved performance consistency across benchmarks. In Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, San Antonio, TX, USA, 11–14 October 2009; pp. 3901–3908.
71. Noel, M.M. A new gradient based particle swarm optimization algorithm for accurate computation of global minimum. *Appl. Soft Comput.* **2012**, *12*, 353–359.
72. Vasilescu, M.A.O.; Terzopoulos, D. Multilinear analysis of image ensembles: TensorFaces. In *Computer Vision—ECCV 2002; Lecture Notes in Computer Science*; Heyden, A., Sparr, G., Nielsen, M., Johansen, P., Eds.; Springer: Berlin Heidelberg, Germany, 2002; Volume 2350, pp. 447–460.
73. Raydan, M.; Svaiter, B.F. Relaxed steepest descent and Cauchy-Barzilai-Borwein method. *Comput. Optim. Appl.* **2002**, *21*, 155–167.
74. Zivkovic, Z.; van der Heijden, F. Efficient adaptive density estimation per image pixel for the task of background subtraction. *Pattern Recogn. Lett.* **2006**, *27*, 773–780.
75. Yu, L.; Fan, G.; Gong, J.; Havlicek, J. Simultaneous target recognition, segmentation and pose estimation. In Proceedings of the IEEE International Conference on Image Processing, Melbourne, Australia, 15–18 September 2013; pp. 2655–2659.
76. Peli, E. Contrast in complex images. *J. Opt. Soc. Am.* **1990**, *7*, 2032–2040.
77. Varma, M.; Zisserman, A. A statistical approach to material classification using image patch exemplars. *IEEE Trans. Pattern Anal. Mach. Intell.* **2009**, *31*, 2032–2047.
78. Arbelaez, P.; Maire, M.; Fowlkes, C.; Malik, J. Contour detection and hierarchical image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *33*, 898–916.

79. Babalola, K.O.; Patenaude, B.; Aljabar, P.; Schnabel, J.; Kennedy, D.; Crum, W.; Smith, S.; Cootes, T.F.; Jenkinson, M.; Rueckert, D. Comparison and evaluation of segmentation techniques for subcortical structures in brain MRI. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2008*; Lecture Notes in Computer Science; Metaxas, D., Axel, L., Fichtinger, G., Székely, P., Eds.; Springer: Berlin Heidelberg, Germany, 2008; Volume 5241, pp. 409–416.
80. Ge, F.; Wang, S.; Liu, T. Image-segmentation evaluation from the perspective of salient object extraction. In *Proceedings of the IEEE Conference on Computer Vision, Pattern Recognition*, New York, NY, USA, 17–22 June 2006; Volume 1, pp. 1146–1153.
81. Carreira, J.; Sminchisescu, C. CPMC: Automatic object segmentation using constrained parametric min-cuts. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 1312–1328.
82. Gong, J.; Fan, G.; Yu, L.; Havlicek, J.P.; Chen, D.; Fan, N. Joint View-Identity Manifold for Infrared Target Tracking and Recognition. *Comput. Vis. Image Underst.* **2014**, *118*, 211–224.

© 2015 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).