INFRARED-IMAGE CLASSIFICATION USING EXPANSION MATCHING FILTERS AND HIDDEN MARKOV TREES

Priya Bharadwaj, Paul Runkle and Lawrence Carin

Department of Electrical and Computer Engineering
Duke University
Durham, NC 27708-0291

ABSTRACT

Forward-looking infrared (FLIR) images of targets are characterized by the different target components visible in the image, with such dependent on the target-sensor orientation and target history (i.e., which target components are hot). We define a target class as a set of contiguous target-sensor orientations over which the associated image is relatively invariant, or statistically stationary. Given an image from an unknown target, the objective is proper target-class association (target identity and pose). Our principal contribution is an image classifier in which a distinct set of templates is designed for each image class, with templates linked to the object sub-components, and the associated statistics are characterized via a hidden Markov model. In particular, we employ expansion matching (EXM) filters to identify the presence of the target components in the image, and use a hidden Markov tree (HMT) to characterize the statistics of the correlation of the image with the various templates. We achieve a successful classification rate of 92% on a data set of FLIR vehicle images, compared with 72% for a previously developed wavelet-feature-based HMT technique.

1. INTRODUCTION

Classification of images is a problem of long-standing interest, because of applications in target identification, medical diagnosis, character recognition, etc. In this paper we propose a new technique, employing expansion matching and a hidden Markov tree (EXM-HMT), to classify two-dimensional forward-looking infrared (FLIR) images of three-dimensional targets. As we move around the target, certain parts in the two-dimensional projection (image) of the object become visible, and certain others remain hidden, depending on the target-sensor orientation. The images of an target therefore vary depending on the orientation of the sensor with respect to the object, while also being a function of the target history (e.g., how long the target engine has been on or off). For example, if the object under consideration is a car, the images of the front of the car are often dramatically different from the images of the sides. Moreover, there is not simply one realization of the FLIR signature of a vehicle at a given orientation, but rather an ensemble of such accounting for variable target history.

The classification problem involves assigning each image to a class, where a class is defined as a set of object-sensor orientations, for a given target, over which the images remain relatively invariant or stationary (with respect to target-sensor variation and target history). There is a *set* of classes for each of multiple targets.

The fundamental idea behind the image classification scheme introduced in this paper is that images can be classified by identifying the parts of the object that are visible in each class of images, and by considering the *relative* position of the various parts in the image. We represent the target parts by a set of templates, and use expansion matching (EXM) filters [5] instead of the more commonly used matched filters, to correlate the image with the templates. The response of the EXM filters has sharper peaks, which facilitates the process of locating the template in an image.

Since the images belonging to a particular class are statistically stationary, the feature vectors of the images can be characterized by a single statistical model. A two-state model is used to represent each coefficient of the feature vector, and the statistics of the coefficient within each such state is modeled via a distinct Gaussian density [3]. Further, the states sampled by successive coefficients of the feature vector are modeled as a Markov process. This formulation results in a hidden Markov tree (HMT): 'hidden' because the states sampled by the coefficients are unknown. The feature vector is arranged in a tree [3,4]. The performance of the HMT based on EXM filters, tied to target parts, is compared to HMT performance based on a Haar-wavelet decomposition [4].

2. FEATURE EXTRACTION: EXPANSION MATCHING (EXM) FILTERS

We derive the templates for each target class by partitioning the images into several subimages. We have an additional template for the overall image, to characterize the global target shape and size. In Fig. 1, for example, an image is divided into six subimages, numbered 2-7 in the figure, and the template of the entire image is indexed as 1. It can be seen from Fig. 1 that subimages 2 and 3 represent the body of the car, and subimages 4, 5, 6 and 7 represent the tires and the lower half of the car.

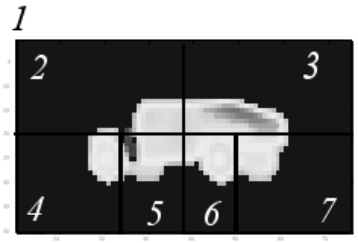

Figure 1. Image divided into several subimages

For a given feature template, the matched filter is an optimal filter in the sense that the SNR is maximized, with SNR defined as the ratio of the filter's response at the center of the pattern to the variance of the filter's response to noise. However, one of the drawbacks of a matched filter [5] is that the response off the center of the feature can be high (as the matched filter is optimized only with respect to the response at the center of the template); as a result the response has a broad peak, and it is difficult to locate the feature in the image, especially if the image has several similar features close to each other.

This limitation of a matched filter is alleviated by the Expansion Matching (EXM) filter [5] which maximizes a criterion called Discriminative SNR (DSNR, [5]), by seeking to minimize the off-center response of the filter; EXM filters generate sharper peaks, enhancing the localization of features in an image. The EXM filter obtained by maximizing DSNR is the same as the Wiener filter [5] formulation for restoring images in the presence of noise and blurring effects. In this context, the feature template corresponds to the blurring function, and a delta function is to be restored. Hence, the EXM filter of a template is given as

$$\Theta_{EXM}(\varpi_1,\varpi_2)=\frac{\overline{\Phi}(\varpi_1,\varpi_2)}{\left|\Phi(\varpi_1,\varpi_2)\right|^2+\dfrac{S_{\lambda\lambda}(\varpi_1,\varpi_2)}{S_{cc}(\varpi_1,\varpi_2)}} \qquad (1)$$

$\overline{\Phi}(\varpi_1,\varpi_2)$ is the complex conjugate of the Fourier transform of the feature template $\Phi(\varpi_1,\varpi_2)$, and $S_{\lambda\lambda}$ and S_{cc} are the power spectral densities of the noise and the input sequence that is to be estimated, respectively.

2.1 Eigen Detectors Using Karhunen-Loeve Transform

Assume there are M image classes, and let N_m denote the number of training images associated with class m ($1 \le m \le M$). Assume P feature templates (e.g., P=7 for the image in Fig. 1) are derived from each of the N_m images belonging to the training set of class m. Therefore, there are N_m realizations of each of the P templates of class m. In order to correlate the feature templates with each image, EXM filters are generated from each template using (1).

By using N_m EXM filters for each of the P feature detectors, we incorporate the variations in the templates of the images belonging to the same class into the feature detectors. We use the Karhunen-Loeve transform (KLT) [6] to reduce the computational complexity of correlating the image with N_m filters. The KLT produces an orthonormal set of basis functions for the N_m realizations of template p of class m. The eigenvectors are arranged in the descending order of eigenvalues; MSE can be minimized by using the top N_{eig} eigenvalues as a truncated basis to represent the entire set of N_m filters. In general, $N_{eig} < N_m$. It should be noted that N_{eig} is not a fixed value: the value of N_{eig} depends on the EXM filter set under consideration.

2.2 Feature Vector

Each image is reduced to a feature vector by correlating the image with the eigen detectors of the EXM filters of P templates of a particular class m, summing the responses from the respective EXM filters, and determining the maximum value of the correlation in a particular neighborhood in the image. Since P feature detectors characterize class m, the length of the feature vector equals P. The feature vector for image n, class m is

$$C_m^n = [corr_1^{m,n}, corr_2^{m,n}, ..., corr_P^{m,n}] \qquad (2)$$

where

$$corr_p^{m,n} = \max_{neighborhood}\left[\sum_{j=1}^{N_{eig}} \frac{\left\langle I^{m,n}, \phi_j^{m,p}\right\rangle}{\left\langle I_{exm}^{m,n}, I_{exm}^{m,n}\right\rangle^{\frac{1}{2}}\left\langle \phi_j^{m,p}, \phi_j^{m,p}\right\rangle^{\frac{1}{2}}}\right] \qquad (3)$$

$\phi_j^{m,p}$ is the j^{th} eigenvector of the EXM filter set of template p, belonging to image class m. One of the limitations of both matched filters and EXM filters is that they produce a high response at certain locations that have high amplitudes, despite the absence of the template at those locations. In order to offset the effect of high amplitude regions, we use correlation as the feature, and not energy extracted by the template from the image, as correlation is a better indicator of the 'match' between templates. We thereby nullify the effects of high-amplitude regions by dividing the inner product in (3) by the energy in the image over the support of the filter under consideration.

We do not search for the maximum value of the correlation in (3) over the entire image, rather we restrict our search to a prescribed neighborhood. Since the subimages are formed by dividing the image into parts, and all the images of the training set are located at a known reference point and oriented at a particular angle, for a given class we know where each template should be located approximately. We look for maximum correlation only in the neighborhood of the corresponding image component.

It is not necessary, however, to know the location and the orientation of the test images. The test images can be centered by correlating the image with the set of EXM filters derived from the entire image, and then shifting the image such that maximum value of correlation lies in the center of the image (or at any other

reference point). Similarly, we can use the training set to develop EXM filters of the images oriented at different angles, and determine the orientation of the test image by correlating it with the rotated set of EXM filters. The orientation of the test image corresponds to the orientation of the EXM filter set for which the maximum value of correlation is obtained, and the test image can be oriented as the training images by rotating it through this angle. The EXM-HMT scheme is, therefore, approximately shift and rotation invariant.

3. STOCHASTIC MODEL: HIDDEN MARKOV TREE (HMT)

The value of $corr_p^{m,n}$ in (3) can be either 'high' or 'low' [3,4] depending on whether that particular feature is present or occluded in the FLIR image being considered. For example, if a given target part is cool, it will have a low value in the FLIR image, with the opposite true for hot target parts. Occlusions can also play a role in the strength of a given target component. We call the 'high' and 'low' correlation values 'high' and 'low' states, respectively, of a feature. The statistics of the 'high' and 'low' states, corresponding to each element of the feature vector, are modeled via a distinct Gaussian density (or, possibly, a Gaussian mixture). Also, if the $corr_p^{m,n}$ is 'high', it is still possible that $corr_{p+1}^{m,n}$ is 'low'. Such interactions between the states of different elements, for a given class of images, are modeled as a Markov process [5]. This formulation results in a hidden Markov tree, since the state of the coefficient being sampled is 'hidden', and the tree nature of the feature vector C_m^n.

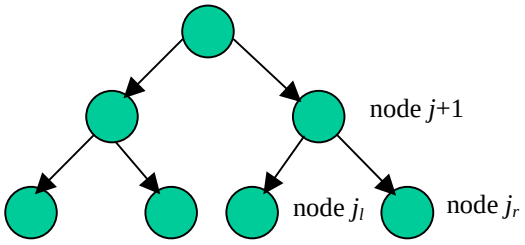

Figure 2. Hidden Markov Tree

The feature vectors of the images can be cast into a tree structure, similar to the wavelet coefficients for which the HMT was developed in [4]. Figure 2 shows a 3-level HMT used to classify the images belonging to the same class as the one shown in Fig. 1, and the index p in each node of the HMT indicates that the node is occupied by $corr_p^{m,n}$. The correlation with the EXM filter of the entire image, $corr_1^{m,n}$, occupies the position at the top of the tree. Subsequent levels are occupied by the correlation values with templates 2 to P, which for the image shown in Fig.1 correspond to the body and the tires of the car (for this example).

The EXM-HMT scheme is demonstrated on FLIR images of vehicles (Sec. 5), with the intensity of these images a function of the temperature of the vehicles. As discussed, in such images 'high' and 'low' states correspond to whether a particular part of the object is 'hot' or 'cold'. The model in Fig. 2 is compatible with our understanding of the physical nature of infrared images. Referring to Figs. 1 and 2, the state of nodes 2 and 3, i.e., correlation with the templates corresponding to the body of the car, are dependent on the state of the correlation with the entire image, i.e. node 1. For example, if node 1 is in the 'high' state, it means that the vehicle is generally 'hot', and therefore it is likely that the nodes 2 and 3 are also in the 'high' state. Since parts 4 and 5 are close to 2, states of 4 and 5 are likely to be influenced by the state of 2, and similarly states of 6 and 7 are dependent on the state of 3. We note, however, that there are multiple ways of devising the tree structure. The goal is to link the decomposition of the Markov tree to the physical (thermal) characteristics of the target.

All but the lowest HMT nodes are connected with two "children" at a lower level. Referring to Fig. 2, let j_l and j_r represent the "children" nodes to the left and right of node j+1. Each node of the HMT, as mentioned, is characterized by a two-state Gaussian model. Let H and L represent the "high" and "low" states of node j+1, with H_l and L_l similarly defined for j_l. There are four possible state transitions from j+1 to j_l: the node j+1 could be H and the element at j_l could be H_l, [H,H_l]; similarly we could have [H,L_l], [L,H_l] or [L,L_l]. Each state transition, listed above, is characterized by an associated probability. A similar set of state transitions is defined for transition from j+1 to j_r. The initial-state probability for the top node is defined as the probability that element $corr_1^{m,n}$ is in the "high" or "low" state. The hidden Markov tree is completely characterized by the dual-state Gaussian model for each element, the state-transition probabilities, and the initial-state probability for the top node. The HMT construct developed here is motivated by [4], in which it was applied to a wavelet decomposition.

4. COMPARISON WITH WAVELET-HMT

Since the HMTs were first developed [4] to characterize wavelet coefficients, we compare the classification results obtained via the EXM-HMT algorithm with results from the wavelet-HMT scheme. The resulting wavelet-HMT structure is a quadtree [4], in which each parent node is connected to four child nodes (in the HMT model discussed in Sec. 3 each parent is connected to two children). We here employ a decomposition based on the Haar wavelet, although the study of HMT performance with alternative wavelets will also be considered. The wavelet decomposition of the FLIR images is performed to the coarsest level, and quadtree HMTs are developed for the sequence of high-high, high-low and low-high images, using the coarsest and two subsequent finer levels (a total of three levels). Due to the fact that the FLIR images are not spatially stationary, we do not perform tying [4]. Consequently, with the finite imagery available for training, we cannot accurately estimate HMT parameters for more than three wavelet levels.

As indicated, there is a wavelet quadtree for the sequence of high-high, high-low and low-high FLIR imagery (for three levels), with these here taken as statistically independent, for simplicity. Therefore, the total likelihood that a given image is

associated with a given class is computed as the product of the likelihoods of the three associated wavelet-quadtree HMTs.

5. CLASSIFICATION RESULTS

We employ the EXM-HMT classification technique to classify FLIR images of four distinct vehicles: three tanks and one truck. We observe that the images, formed at 5^o intervals around the vehicle, vary as a function of the target-sensor orientation (and as a function of target history). We identify two sets of angular regions (classes) for each vehicle over which the images are relatively unchanged (stationary). Let 0^o be defined as looking at the front end of the vehicle. Class 1 of a target type is defined as FLIR images of the front and rear of the vehicle (angles 0-15^o, 345-360^o and 165-195^o), and class 2 comprises images of the sides of the vehicle (angle 20-160^o and 200-340^o). There is not sufficient resolution and training data to separately distinguish the front and back of the targets. Since there are two classes for each vehicle, there are a total of M=8 image classes (four vehicles with 2 classes per vehicle). The data was provided by the US Army Research Laboratory [7], with example FLIR images shown in Fig. 3. For vehicle 1, class 1 and class 2, a set of N_m=260 images are used to train the HMT. For the other image classes, N_m=152. Seven EXM filters are developed for each image: one for each subimage (see Fig. 1), and one EXM filter for the entire image. We perform KLT, and N_{eig} is set such that 90% of the energy in the original set of filters can be extracted by the eigen-detectors. For vehicle 1, class 1 and class 2, N_{eig} =30, and for the rest N_{eig} =50.

The average correct classification of the EXM-HMT was 92% (the associated confusion matrix is shown in Table 1), while the wavelet-based HMT yielded 72% correct classification (Haar wavelets). The testing and training data was completely independent.

6. CONCLUSIONS

We have designed a hidden Markov tree (HMT) for target classification, based on expansion-matching filters. Such a model has been developed previously based on a wavelet decomposition. The principal contribution reported here is an extension of the HMT to more general filters, in particular to EXM (Wiener) filters [5] matched to fundamental components of the targets of interest. The method was tested on FLIR data from similar targets [7].

REFERENCES

[1] L.R. Rabiner and B.H. Juang, " An introduction to hidden Markov models," *IEEE Acoust. Speech Signal Proc. Magazine*, vol. 3, pp. 4-16, Jan. 1986

[2] P.K. Bharadwaj *et al.* , "Target identification with wave-based matching pursuits and hidden Markov model", *IEEE Trans. Antenna Propagation*, vol.47, no. 10, pp. 1543-1554, Oct. 1999.

[3] M.S. Crouse, *et al.*, "Wavelet-based statistical signal processing using hidden Markov models," *IEEE Trans. Signal Processing*, vol. 46, pp. 886-902, April 1998.

[4] N. Dasgupta, *et al.*, "Dual hidden Markov model for characterizing the wavelet coefficients from multi-aspect scattering data," to appear in *Signal Processing.*

[5] J. Ben-Arie, K.R. Rao, " A novel approach for template matching by nonorthogonal image expansion," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 3, no. 1, pp.71-84, February 1993.

[6] M.L. C. and M.I. Miller, "Information measures for object recognition accommodating signature variability," *IEEE Trans. on Information Theory*, vol. 46, no. 5, August 2000.

[7] L.-C. Wang; S.Z. Der, N.M. Nasrabadi, "Automatic target recognition using a feature-decomposition and data-decomposition modular neural network, *IEEE Trans. Image Proc.*, vol. 7, pp. 1113-1121, Aug. 1998.

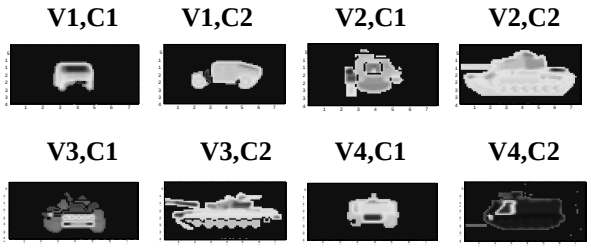

Fig. 3. Example FLIR imagery from targets V1-V4, with two classes (C1 and C2) per target.

	V1 C1	V1 C2	V2 C1	V2 C2	V3 C1	V3 C2	V4 C1	V4 C2
V1C1	**95.5**	0	0.38	0	2.31	0	2.31	0
V1 C2	0	**96.54**	0.77	0	1.92	0	0.38	0.38
V2 C1	0.67	2.63	**82.24**	1.32	12.50	0	0.66	0
V2 C2	0	0	0.66	**98.03**	0.66	0	0.66	0
V3 C1	1.97	4.61	7.89	0	**84.87**	0	0.66	0
V3 C2	0	0	1.32	0.66	0	**98.03**	0	0
V4 C1	1.32	1.32	1.97	0	7.89	0	**85.53**	1.97
V4 C2	0	0	2.63	0	0	0	1.97	**95.39**

Table 1. Confusion matrix of EXM-HMT algorithm for four vehicles (V1, V2, V3 and V4), with two classes per target (C1 and C2). Testing and training data is independent.