Available online at www.sciencedirect.com
ScienceDirect
Digital Signal Processing 18 (2008) 375–390
Digital Signal Processing
www.elsevier.com/locate/dsp

On statistical approaches to target silhouette classification in difficult conditions

Conrad Sanderson [a,b,*], Danny Gibbins [c], Stephen Searle [d]

[a] *NICTA, 300 Adelaide Street, Brisbane, QLD 4000, Australia*
[b] *Australian National University, Canberra, ACT 0200, Australia*
[c] *School of Electrical and Electronic Engineering, University of Adelaide, SA 5005, Australia*
[d] *Department of Electrical and Electronic Engineering, University of Melbourne, VIC 3010, Australia*

Available online 6 June 2007

Abstract

In this paper we present a methodical evaluation of the performance of a new and two traditional approaches to automatic target recognition (ATR) based on silhouette representation of objects. Performance is evaluated under the simulated conditions of imperfect localization by a region of interest (ROI) algorithm (resulting in clipping and scale changes) as well as occlusions by other silhouettes, noise and out-of-plane rotations. The two traditional approaches are holistic in nature and are based on moment invariants and principal component analysis (PCA), while the proposed approach is based on local features (object parts) and is comprised of a block-by-block 2D Hadamard transform (HT) coupled with a Gaussian mixture model (GMM) classifier. Experiments show that the proposed approach has good robustness to clipping and, to a lesser extent, robustness to scale changes. The moment invariants based approach achieves poor performance in advantageous conditions and is easily affected by clipping and occlusions. The PCA based approach is highly affected by scale changes and clipping, while being relatively robust to occlusions and noise. Furthermore, we show that the performance of a silhouette recognition system subject to mismatches between training and test angles of silhouettes (caused by an out-of-plane rotation) can be considerably improved by extending the training set using only a few angles which are widely spaced apart. The improvement comes without affecting the performance at "side-on" views.

Keywords: Automatic target recognition; Silhouette classification; Adverse conditions; Statistical models; Local features; Object parts

1. Introduction

Infra-red (IR) images (also known as thermal images) provide information regarding the emitted and/or reflected IR radiation by objects present in a scene [1,2], and are often used to facilitate operation of automated target recognition (ATR) systems in various environmental conditions, such as hazy weather and the absence of visible light [3–5]. The intensity of the radiation (i.e., temperature) is encoded as the magnitude of each pixel in the image [4,6]. As such, there can be quite a lot of variability in the greyscale representation of objects; these variabilities can stem from the object's history (e.g., due to variable use of engines), sea and air temperature (partly dependent on the time of day), lo-

* Corresponding author.
E-mail address: conradsand@ieee.org (C. Sanderson).

1051-2004/$ – see front matter
doi:10.1016/j.dsp.2007.05.008

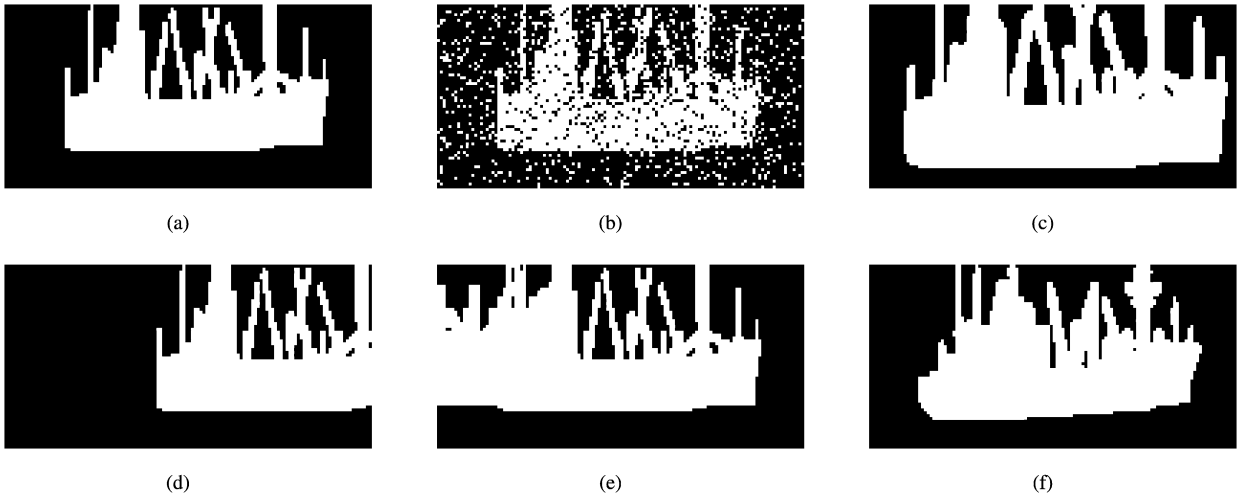

Fig. 1. (a) Silhouette representation of a ship, (b) corrupted by speckle noise, (c) scaled, (d) shifted, (e) overlapped with another silhouette, (f) rotated in depth.

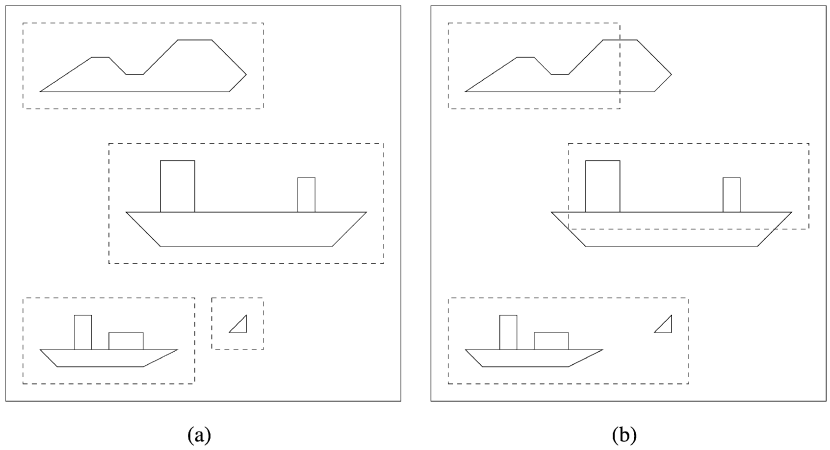

Fig. 2. Conceptual example of (a) scene with perfect ROI localization, (b) scene with faulty ROI localization.

cation and range [4,7–9]. To reduce the effects of intensity variations, binary versions of images can be used [10–12]. When only one solid object is present in a binary image, a *silhouette* of the object is visible; an example is shown in Fig. 1.

We consider part of the problem of localizing a *specific* object based on its silhouette representation. In particular, we assume that we are given a region of interest (ROI), on which we have to perform a two-class recognition task: either the presented ROI contains the silhouette that we are interested in[1] (referred to as a *true silhouette*), or it does not (referred to as an *impostor silhouette*). The recognition task described above (often known as a verification task) is in contrast to the closed set identification task, where a given object is assigned into one of K object classes (where K is the number of *known* objects). The closed set identification task represents a controlled environment and as such is not representative of our problem where any object could be encountered [13] (i.e., in real life situations one cannot assume that all objects to be encountered are going to be known).

The overall performance of an ATR system can highly depend on the quality of the ROI localization algorithm [14]. The algorithm may provide improperly located regions, e.g., objects at incorrect scale and/or partially shifted out view. Examples are shown in Fig. 2. The overall performance can also be affected by the quality of the IR images (e.g., due

[1] For example, the silhouette of a specific ship.

to environmental conditions) as well as the observed state of the located objects (e.g., the objects may have been rotated from the point of view of the ATR system).

A pattern recognition system can be generally split into two distinct parts: feature extraction and classification. Approaches to feature extraction from images containing objects can be roughly divided into three areas: boundary descriptions, local features,[2] and holistic features. Holistic approaches extract information from the entire image in one hit while local feature extractors divide the image (or object) into several parts (possibly nonuniformly sized parts) and extract features for each part.

Examples of boundary descriptions include distances of each boundary pixel from the centroid [15] and Fourier descriptors [16]; examples of parts based approaches include decomposition into distinct surfaces [17], regional principal component analysis (PCA) [14,18], and 2D discrete cosine transform (DCT) [16,19]; examples of holistic descriptions include moment invariants [10,16,20–22], and holistic PCA [11,23].

A possible disadvantage of local features is that relatively complex classifiers may be required in order to take advantage of the spatial relation between object parts (e.g., 2D hidden Markov models used in face classification [24]). Holistic representations, on the other hand, represent the entire image of an object using only one feature vector; the classifier can thus be relatively simple, e.g., an object's features can be considered to have a Gaussian distribution. Further examples and discussion of object descriptors can be found in [16,25,26].

Several studies [9,23,27] have compared the performance of various approaches for recognizing objects in IR images. However, these studies assumed good quality ROI localization or used a closed set identification setup; furthermore, in [9,23] silhouette representations of objects was not used. An initial evaluation of the effects of angle mismatches (due to out-of-plane rotations) was presented in [10], though the dataset used was rather limited (it comprised five ships).

This paper has two main aims: (i) To methodically evaluate, on a common dataset, several combinations of feature extractors and statistical classifiers for the purposes of silhouette verification, subject to imperfect ROI localization and other adverse conditions. In particular, the performance is evaluated for the following conditions: clean (i.e., no impediments), corrupted by speckle noise, partially shifted out of view (clipped), partially occluded by another object, and finally corrupted by a scale change. The speckle noise corruption is a representative of imaging artefacts, a boundary corruption process (other than occlusions by another object), as well as a process which adds *previously unseen elements* to the image (i.e., elements of a different nature than what is present in training images). (ii) To evaluate how sensitive the approaches are to out-of-plane rotations of silhouettes when only "side-on" views are available for training, and to see the degree of possible improvement by extending the training set to include silhouettes at various degrees of rotation.

We consider two holistic silhouette classification approaches (based on coupling moment invariants and PCA based feature extraction with a Gaussian classifier) as well as a local feature silhouette recognition approach based on coupling a block-by-block 2D Hadamard transform (HT) [16] feature extractor with a Gaussian mixture model (GMM) classifier.

Overall, the main contributions of this paper are as follows:

- A methodical evaluation of performance of several approaches to silhouette classification subject to imperfect ROI localization and other adverse conditions. Furthermore, the evaluation is performed using an experiment setup where silhouette classes that were not seen during training are encountered.
- Proposal of a local feature silhouette recognition approach based on a block-by-block 2D Hadamard transform coupled with a Gaussian mixture model classifier.
- Demonstration that deliberately disregarding the relative location of component parts of each silhouette results in good robustness to translations and, to a lesser extent, robustness to scale changes. This is advantageous, as mistakes in the location and scale of an object can be caused by a ROI localization algorithm [14].
- Performance of a silhouette classification system subject to mismatches between training and test angles of silhouettes (caused by an out-of-plane rotation) can be considerably improved by extending the training set using only a few angles which are widely spaced apart. The improvement comes without affecting the performance at "side-on" views.

[2] Also known in the literature as *feature-based* and *parts-based* approaches.

The balance of this paper is structured as follows. In Section 2 we overview the three feature extraction techniques, followed by a brief description of two statistical classifiers in Section 3. Section 4 is devoted to experiments and discussions. The main findings and avenues for future work are summarized in Section 5.

2. Feature extraction

In the following sections it is assumed that the binary image is described by $\{f(x, y) \in \{0, 1\} \mid x = 0, 1, \ldots, N_X - 1, y = 0, 1, \ldots, N_Y - 1\}$, where N_X is the number of columns and N_Y the number of rows. In Sections 2.1 and 2.2 we describe the two holistic feature extraction techniques, while in Section 2.3 we describe how to obtain local features.

2.1. Hu's moment invariants

In this feature extraction method, Hu's moment invariants are derived from normalized central moments of a given image; it is often stated that these moments are invariant to translations, scale changes and in-plane rotations [16]. A moment of $(p + q)$th order is defined as

$$m_{pq} = \sum_{x=0}^{N_X-1} \sum_{y=0}^{N_Y-1} x^p y^q f(x, y) \tag{1}$$

for $p, q = 0, 1, 2, \ldots$. Central moments are in turn defined as

$$c_{pq} = \sum_{x=0}^{N_X-1} \sum_{y=0}^{N_Y-1} (x - \mu_x)^p (y - \mu_y)^q f(x, y), \tag{2}$$

where $\mu_x = m_{10}/m_{00}$ and $\mu_y = m_{01}/m_{00}$. Normalized central moments are in turn defined as: $n_{pq} = c_{pq}/c_{00}^{\gamma}$, where $\gamma = (p + q)/2 + 1$ and $(p + q) \geqslant 2$. Central moments address translations of an object, while normalized central moments in turn address the scale of the object (as c_{00} describes the number of white pixels in a binary image). Seven moment invariants (Φ_1–Φ_7) can then be obtained as nonlinear combinations of second and third order normalized central moments [16,20]; the combinations address the problem of in-plane rotations. As the values of the resultant features can be very small, it is customary to utilize the logarithm of their absolute values in order to avoid precision problems [21]. Formally, a feature vector based on the seven features is formed using:

$$\mathbf{x}^T = \left[\log |\Phi_i|\right]_{i=1}^{7}. \tag{3}$$

2.2. Principal component analysis

In principal component analysis (PCA) based feature extraction [28,29], the first step is to concatenate all the columns of the binary image to form a high dimensional vector; let us denote the resultant vector as $\mathbf{r}$. A new feature vector, usually with a lower dimensionality, is then obtained using:

$$\mathbf{x} = \mathbf{U}^T(\mathbf{r} - \mathbf{r}_\mu). \tag{4}$$

$\mathbf{U}^T$ and $\mathbf{r}_\mu$ are found as follows. A set of training pixel feature vectors is collected from all training images; let us define this set as

$$R = \{\mathbf{r}_i\}_{i=1}^{N_A}. \tag{5}$$

The mean vector of set F is then found using $\mathbf{r}_\mu = \frac{1}{N_A} \sum_{n=1}^{N_A} \mathbf{r}_i$. A covariance matrix is then calculated:

$$\mathbf{C} = \frac{1}{N_A} \sum_{i=1}^{N_A} (\mathbf{r}_i - \mathbf{r}_\mu)(\mathbf{r}_i - \mathbf{r}_\mu)^T. \tag{6}$$

Matrix U is then formed:

$$\mathbf{U} = [\mathbf{e}_1 \quad \mathbf{e}_2 \quad \ldots \quad \mathbf{e}_D], \tag{7}$$

where $\mathbf{e}_n$ is the nth eigenvector of $\mathbf{C}$; the eigenvectors are ordered in a descending manner according to their corresponding eigenvalues; doing so defines orthogonal directions that account for the highest amount of variance (i.e., highest amount of information). D has the following constraints: $D \leqslant N_A$ and $D \leqslant N_X N_Y$. If $D = N_X N_Y$ then no dimensionality reduction occurs; in that case, vector $\mathbf{x}$ represents a decorrelated version of the pixel vector $\mathbf{r}$. A method for choosing the dimensionality is given in Section 4.5.

As this feature extraction technique basically produces dimensionality reduced versions of binary images, we would expect it to be affected by scale changes, clipping and rotations; the onus of any robustness to these changes would be taken by the classifier.

2.3. Local features via the 2D Hadamard transform

As opposed to the holistic feature extraction methods presented in Sections 2.1 and 2.2, where analyzing one image results in one feature vector, in the 2D Hadamard transform (HT) based approach we obtain a set of feature vectors from one image. In the literature this type of feature extraction is known as a *local feature* approach and as a *parts based* approach [24,30].

The 2D HT is similar in nature to the 2D DCT [16] (which is the heart of the JPEG compression algorithm [31]); we shall interpret the HT as a "black and white" version of the DCT, and postulate that it is a good candidate for processing binary images.

In a similar manner to DCT based feature extraction [19], a given image is analyzed on a block-by-block basis (see Fig. 3 for an example) Each block overlaps neighbouring blocks by a configurable amount of pixels.[3] Each image block $\alpha(k,l)$, where $k,l = 0, 1, \ldots, N_P - 1$ and $N_P = 2^n$, is decomposed in terms of orthogonal 2D Hadamard basis functions (see Fig. 4). The result is an $N_P \times N_P$ matrix $H(u,v)$ containing 2D Hadamard coefficients:

$$H(u,v) = \frac{1}{N_P} \sum_{k=0}^{N_P-1} \sum_{l=0}^{N_P-1} \alpha(k,l)(-1)^{\beta(k,l,u,v)}, \tag{8}$$

where

$$\beta(k,l,u,v) = \sum_{i=0}^{n-1} \big[b_i(k)\, p_i(u) + b_i(l)\, p_i(v) \big], \tag{9}$$

$$b_i(k) = i\text{th bit in the binary representation of } k, \tag{10}$$

$$p_i(k) = \begin{cases} b_{n-i}(k) & \text{for } i = 0, \\ b_{n-i}(k) + b_{n-(i+1)}(k) & \text{for } i \in [1, n-1]. \end{cases} \tag{11}$$

The summations in Eqs. (9) and (11) are performed in modulo 2 arithmetic. As the 2D HT is similar in nature to the 2D DCT, the basis functions can be thought of as representing different "frequency" components. If we assume low frequency information to be dominant, we can order the resulting coefficients according to a zig-zag pattern [16], reflecting the amount of information stored in each coefficient (see Fig. 5). Formally, for a block located at (k,l), the Hadamard feature vector is composed of

$$\mathbf{x}^{(k,l)} = \big[h_0^{(k,l)} \quad h_1^{(k,l)} \quad \ldots \quad h_{M-1}^{(k,l)} \big]^T, \tag{12}$$

where $h_n^{(k,l)}$ denotes the nth Hadamard coefficient and M is the number of retained coefficients. By truncating the number of coefficients we are effectively throwing out high frequency information. The vectors from the entire image can then be collected in a set, $X = \{\mathbf{x}_i\}_{i=1}^{N_V}$ where the superscript indicating the location of each source block has been replaced by the subscript i, which indicates the ith vector of the image. The degree of overlap has two main effects:

1. With a large overlap, the HT coefficients from a set of (horizontally or vertically) consecutive blocks will not vary abruptly.

[3] A similar overlapping approach is used in processing of speech signals [32,33].

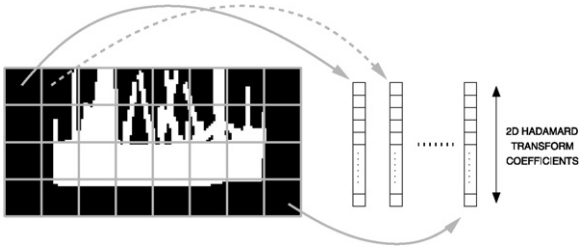

Fig. 3. Conceptual example of block-by-block feature extraction.

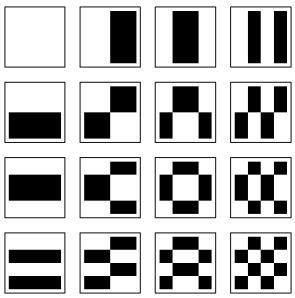

Fig. 4. 2D Hadamard basis functions for $N_P = 4$; white areas indicate 1 while black areas indicate -1.

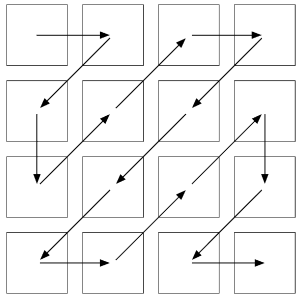

Fig. 5. Zig-zag ordering of 2D Hadamard coefficients $H(u, v)$ for $N_P = 4$.

2. When using a large overlap, the parts of each silhouette are in effect "sampled" at various degrees of translations, resulting in models which should be robust to minor translations of the silhouettes. This is in *addition* to the translation robustness provided by the GMM classifier (described in Section 3.2), where the location of each block has little influence. By itself, GMM's built-in robustness only works when the size of the translation is equivalent to an integral multiple of the block size (see the following section for more information).

3. Statistical classifiers

As our task is to look for a *specific* silhouette, we are only concerned with two possible outcomes: either the given ROI contains the specific silhouette, or it does not. As such, we need two models; the first model describes the silhouette we are looking for, while the second model approximately describes other silhouettes, or interpreted alternatively, a "generic" silhouette.

Let us assume that the class for a specific silhouette is denoted by S while the class for any other silhouette is denoted by G. Furthermore, let us describe the specific silhouette model[4] by λ_S and the generic silhouette model by λ_G. The class membership of a set of feature vectors, $X = \{\mathbf{x}_i\}_{i=1}^{N_V}$, is found using:

$$\text{class membership} = \begin{cases} S, & \text{if } \Lambda(X|\lambda_S, \lambda_G) > t, \\ G, & \text{if } \Lambda(X|\lambda_S, \lambda_G) \leqslant t, \end{cases} \tag{13}$$

where t is a tunable decision threshold and $\Lambda(X|\lambda_S, \lambda_G)$ is a log-likelihood ratio, defined as

$$\Lambda(X|\lambda_S, \lambda_G) = \log P(X|\lambda_S) - \log P(X|\lambda_G) \tag{14}$$

$$= \log \prod_{i=1}^{N_V} P(\mathbf{x}_i|\lambda_S) - \log \prod_{i=1}^{N_V} P(\mathbf{x}_i|\lambda_G) \tag{15}$$

$$= \sum_{i=1}^{N_V} \log P(\mathbf{x}_i|\lambda_S) - \sum_{i=1}^{N_V} \log P(\mathbf{x}_i|\lambda_G). \tag{16}$$

Note that above it has been assumed (for mathematical convenience) that each vector in X is independent and identically distributed and that prior probabilities of the two classes are noninformative [28]. The value of $\Lambda(X|\lambda_S, \lambda_G)$ can be interpreted as the *opinion* on the given ROI; a high opinion suggests that the ROI contains the required silhouette, while a low opinion suggests a different silhouette.

3.1. Gaussian based

In the Gaussian based classifier each class is assumed to have a Gaussian distribution, i.e.,

[4] For convenience, we use the terms "model" and "parameter set" interchangeably.

$$P(\mathbf{x}|\lambda) \triangleq \mathcal{N}(\mathbf{x}|\mu, \mathbf{\Sigma}) = \frac{\exp[-\frac{1}{2}(\mathbf{x}-\mu)^T \mathbf{\Sigma}^{-1}(\mathbf{x}-\mu)]}{(2\pi)^{D/2}|\mathbf{\Sigma}|^{1/2}}, \tag{17}$$

where $\lambda = \{\mu, \mathbf{\Sigma}\}$ is the parameter set, D is the dimensionality of the feature vectors, $\mathbf{\Sigma}$ is the covariance matrix, and μ is the mean.

Assuming the training dataset for a specific class is large enough, the mean and covariance matrix for the class is taken to be the sample mean and sample covariance matrix of the training vectors for that class, respectively. However, in our case there is only a few training images for each silhouette class (see Section 4), resulting in only a few training vectors when using holistic feature extraction; this in turn makes the estimation of covariance matrices for each object ill advised [28]. The following alternative training strategy is used. First, we obtain the mean vector and covariance matrix for the generic model by using vectors representing all training silhouettes (excluding impostor silhouettes, see Section 4). Due to the relatively small amount of training data, the covariance matrix is restricted to be diagonal. The mean for a particular silhouette is taken to be the mean of that silhouette's vectors; the covariance matrix for each silhouette is then inherited from the generic model [28].

3.2. Gaussian mixture model based

The Gaussian mixture model (GMM) based classifier [28,29,33] can model distributions much more precisely compared to the single Gaussian classifier, as each class is now assumed to have a distribution comprised of a mixture of Gaussians:

$$P(\mathbf{x}|\lambda) \triangleq \sum_{g=1}^{N_G} w_g \mathcal{N}(\mathbf{x}|\mu_g, \mathbf{\Sigma}_g). \tag{18}$$

Here $\lambda = \{w_g, \mu_g, \mathbf{\Sigma}_g\}_{g=1}^{N_G}$, N_G is the number of Gaussians and w_g is the weight for Gaussian g (with constraints $\sum_{g=1}^{N_G} w_g = 1$ and $\forall g\colon\ w_g \geqslant 0$). The higher the N_G, the more precise the model (assuming a large enough training dataset). Moreover, even if diagonal covariance matrices are utilized, it is possible to model correlated data using $N_G \geqslant 2$ [33].

In a similar manner to the Gaussian classifier described in Section 3.1, training is accomplished as a two-stage process. First, the parameters for the generic model are estimated using the expectation maximization (EM) algorithm [34]; data from all training silhouettes is utilized (excluding impostor silhouettes, see Section 4). A model for each class is then obtained by adapting the generic model using a form of maximum a posteriori adaptation [33,35].

Note that when using 2D HT based feature extraction, Eq. (16) indicates that the ordering of the blocks is not important and that the spatial relations between the blocks are lost (i.e., the relative locations of the component parts of a silhouette are disregarded). This indicates that translations of the silhouette should have relatively little influence.

4. Evaluation

4.1. Dataset and associated experiment protocols

The dataset is comprised of 64 synthetically generated objects (32 of which are ships); each object is rotated (out-of-plane rotation) 180° in 5° steps, resulting in 36 images per object. Each image has a resolution of 512×512 pixels and contains the silhouette of the object.

In this paper three configurations of the dataset were used. In all configurations, the silhouettes were distributed into three types: *true silhouettes*, *evaluation impostor silhouettes*, and *test impostor silhouettes*. 48 silhouettes were assigned to the true silhouettes section, while 16 silhouettes were assigned to the impostor sections (eight to the evaluation section and eight to the test section). Apart from forcing half the silhouettes in each section to be ships, the assignments were otherwise random. The silhouettes for the objects were further split into three sections: training, evaluation and test. The three configurations of the dataset are shown in Tables 1–3.

In configuration A, it was assumed that images for 0°, ±5° and ±10° are various realizations of the "side-on" view. Images for 0° and 5° were assigned to the training section, images for −5° were assigned to the evaluation section, and finally images for −10° and 10° were assigned to the test section. Configuration B is similar to configuration A;

Table 1
Dataset configuration A

Angle	True silhouettes (48)	Evaluation impostor silhouettes (8)	Test impostor silhouettes (8)
0° and 5°	Training data	–	–
−5°	Evaluation data	Evaluation data	–
−10° and 10°	Test data	–	Test data

Table 2
Dataset configuration B

Angle	True silhouettes (48)	Evaluation impostor silhouettes (8)	Test impostor silhouettes (8)
0° and 5°	Training data	–	–
−5°	Evaluation data	Evaluation data	–
One of even angles within ±90°	Test data	–	Test data

Table 3
Dataset configuration C

Angle	True silhouettes (48)	Evaluation impostor silhouettes (8)	Test impostor silhouettes (8)
Subset of odd angles within ±90°	Training data	–	–
Same as above	Evaluation data	Evaluation data	–
0° or one of even angles within ±90°	Test data	–	Test data

the difference being that instead of using only test images for ±10°, images corresponding to even angles upto ±90° are used (i.e., ±10°, ±20°, ..., ±90°). In configuration C, a subset of odd angles within ±90° were assigned to the training section (e.g., {±5°, ±45°, ±85°}). The evaluation section used the same angles as the training section. For the test section, images from 0° or even angles within ±90° were used.

For all experiments the training section was utilized as a source of data for training the silhouette models. Wherever applicable, the evaluation section was used for tuning feature extractors and classifiers (e.g., to choose the optimum dimensionality and/or the decision threshold). Once the optimum parameters are found, the test section is used for final performance measurement.

This dataset partitioning is necessary in order to avoid optimistic biases in the final performance evaluation (i.e., parameters for the feature extractors and classifiers are never selected based on the performance on the test section). Moreover, where possible, the evaluation section is used to avoid overfitting (i.e., to promote good generalization capability) [28,29]. However, it should be pointed out that some bias may still exist due to the limited nature of the dataset.

4.2. Performance measures

There are two types of errors that can occur in a verification task: a false acceptance (FA), which occurs when the system accepts an impostor silhouette and a false rejection (FR), which occurs when the system refuses a true silhouette. The performance of verification systems is generally measured in terms of *false acceptance rate* (FAR) and *false rejection rate* (FRR), defined as

$$\mathrm{FAR} = \frac{\text{number of FAs}}{\text{number of impostor silhouette presentations}}, \tag{19}$$

$$\mathrm{FRR} = \frac{\text{number of FRs}}{\text{number of true silhouette presentations}}. \tag{20}$$

To aid the interpretation of performance, the two error measures are often combined using the half total error rate (HTER[5]), defined as

[5] The HTER can be thought of as a special case of the decision cost function (DCF) [36,37].

$$\mathrm{HTER}(t, \mathcal{D}) = \frac{\mathrm{FAR}(t, \mathcal{D}) + \mathrm{FRR}(t, \mathcal{D})}{2}, \tag{21}$$

where $\mathrm{FAR}(t, \mathcal{D})$ and $\mathrm{FRR}(t, \mathcal{D})$ are the FAR and FRR, respectively, for a decision threshold t and dataset $\mathcal{D}$. A particular case of the HTER, known as the equal error rate (EER), occurs when the threshold is adjusted so that FAR = FRR on a particular dataset $\mathcal{D}'$ (which could be different from $\mathcal{D}$).

4.3. Feature and classifier combinations

Due to the limited number of training images and the nature of the feature extractors based on PCA and Hu's moment invariants (where processing one image results in one feature vector), there is not enough data for estimating more than one Gaussian. As such, the Gaussian classifier (Section 3.1) was coupled with Hu's moment invariants and PCA based features. The resulting holistic classification approaches will be referred to as Hu + G and PCA + G, respectively. The GMM classifier was coupled with 2D HT based feature extraction, where it is possible to obtain a large set of feature vectors from each image. The resulting local feature classification approach will be referred to as HT + GMM.

4.4. Image and feature pre-processing

Prior to feature extraction or applying any image corruption, all objects are normalized in size. The size normalization stage corresponds to a size normalized image being provided by a ROI algorithm. We find the mean (μ) and standard deviation (σ) of object pixel positions along the x and y dimensions of the image; the object is then assumed to be contained within the area specified by $\mu \pm 2.5\sigma$ in each dimension; 2.5 was chosen so that all objects in the dataset fit into the specified area. The area is extracted and normalized to have a size of 128 columns and 64 rows. As we have no prior information as to which image dimensions are optimal, the image size was selected to be as small as possible while still allowing the objects to be easily distinguished visually.

Preliminary experiments with the HT + GMM combination showed that using all extracted blocks led to poor performance. The problem was traced back to a few dominant Gaussians which tended to model blocks with little or no discriminative information. These blocks represented the black background, completely solid areas of silhouettes and elongated sections of silhouettes (e.g., the bottom of a ship, as shown in the bottom row of blocks in Fig. 3). The black background and completely solid areas of silhouettes are common to all images of silhouettes and thus carry little discriminative information. The elongated sections of silhouettes resulted in a large subset of feature vectors which were the same or very similar to each other. This in turn biased the EM algorithm to emphasize the Gaussians which modelled the elongated sections.

To reduce the above modelling problems, the following feature pre-processing heuristic was introduced. Firstly, any duplicate feature vectors are removed, thus disallowing the emphasis on any particular section of a silhouette. Secondly, any feature vector which comes from a completely empty block (i.e., containing only black pixels) or a completely solid block (i.e., containing only white pixels) is ignored. This feature pre-processing heuristic causes the number of vectors extracted from an image to be dependent on the content of the image, hence opinions resulting from Eq. (16) will not be comparable between different objects. To address this side-effect, Eq. (16) is modified to

$$\Lambda(X|\lambda_S, \lambda_G) = \frac{1}{N_V} \sum_{i=1}^{N_V} \log P(\mathbf{x}_i|\lambda_S) - \frac{1}{N_V} \sum_{i=1}^{N_V} \log P(\mathbf{x}_i|\lambda_G), \tag{22}$$

i.e., the effect of variable number of feature vectors is reduced by using average log-likelihoods.

4.5. Experiments with noise, clipping, scale changes and occlusions

In the first set of experiments we have used dataset configuration A (to largely discount the effects of out-of-plane rotations). The amount of training and evaluation data was artificially increased by using mirrored versions of images. In each experiment the classifier was given the model of the silhouette we are interested in, as well as test images of that silhouette and impostor silhouettes; each given image was classified as either containing the silhouette we want (i.e., a true silhouette), or containing a different silhouette (i.e., an impostor silhouette). This procedure was repeated

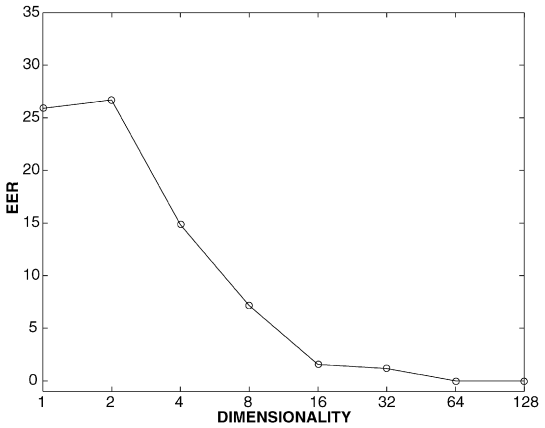

Fig. 6. EER performance on evaluation data of PCA + G approach for varying dimensionality of feature vectors.

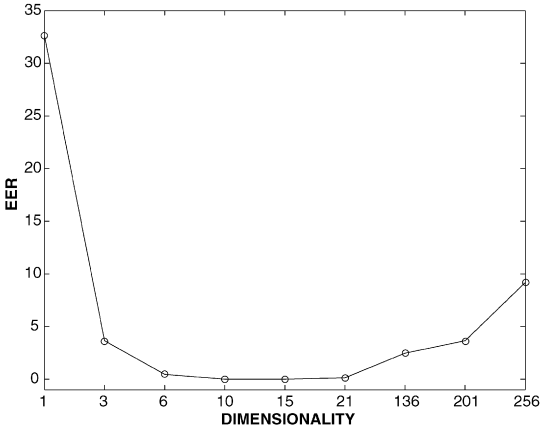

Fig. 7. EER performance on evaluation data of HT + GMM approach for several dimensionalities of feature vectors extracted from 16×16 blocks.

for all silhouettes in the *true silhouettes* section, resulting in a total of $48 \times 2 = 96$ true silhouette presentations and $48 \times 8 \times 2 = 760$ impostor silhouette presentations.

The evaluation section of the dataset was used to jointly optimize the performance of the feature extractors and classifiers (e.g., dimensionality of feature vectors, number of Gaussians, the decision threshold). The optimization was performed in terms of minimum EER on evaluation data.

For PCA based feature extraction, the dimensionality was varied from 1 to 128, doubling the number of dimensions in each step (i.e., 1, 2, 4, ..., 128). The optimum dimensionality, according to performance on evaluation data, was found to be 64 (i.e., the dimensionality of binary images was reduced from 8192 to 64, representing a reduction of 99.2%). Figure 6 shows the performance of the PCA + G approach for varying dimensionality of the feature vectors.

For HT based feature extraction, each possible dimensionality was based on the cumulative amount of coefficients along the diagonals traced by the zig-zag pattern (see Fig. 5). The number of Gaussians in the GMM approach was varied[6] from 1 to 64, doubling the number of Gaussians in each step (i.e., 1, 2, 4, ..., 64). Three block sizes were evaluated: 4×4, 8×8, and 16×16. The optimum performance on the evaluation set was obtained using 16×16 blocks, feature vector dimensionality of 10 and models with 64 Gaussians. This configuration was used in further experiments. Fig. 7 shows the performance of the HT + GMM approach for several dimensionalities of feature vectors extracted from 16×16 blocks.

Once the optimum parameters were found on the evaluation section, the resulting systems were tested on clean images (i.e., noncorrupted) from the test section. Results presented in Table 4 show that the HT + GMM approach achieves performance comparable to that of the PCA + G approach. The Hu + G approach obtains by far the worst performance. Its poor performance can be attributed to inadequate representation of the silhouettes (i.e., the features are not discriminative enough), and as the moments are correlated [38], partly to the diagonal covariance matrix assumption used in the Gaussian classifier.

The results for test silhouettes corrupted by speckle noise, clipping, scale changes and occlusions are shown in Figs. 8–11. For the speckle noise results, the noise level indicates the percentage of pixels which were randomly set to either zero or one; the location of the pixel to be corrupted was randomly selected according to a uniform distribution. For clipping experiments, the shift level indicates the fraction of columns by which the test silhouettes have been shifted to the right. For scale experiments, the scale level indicates the size multiplier of each silhouette. For occlusion experiments, the overlap level indicates the fraction of columns which have been corrupted by a secondary silhouette, moving in from the left; the secondary silhouette was taken to be one of the test impostor silhouettes.

The results show that all three approaches are affected by noise, with the HT + GMM approach the most sensitive. The sensitivity of moment invariants stems from the fact that as more noise is present, the more affected μ_x, μ_y and c_{00} are (see Section 2.1). Analysis shows that the sensitivity of the HT + GMM approach is partially due to the

[6] Computational and time limitations prevented us from utilizing more than 64 Gaussians.

Table 4
Performance using clean silhouettes. EER represents performance on evaluation data, while HTER represents performance on test data

Approach	EER	HTER
HU + G	22.14	18.95
PCA + G	0.00	3.19
HT + GMM	0.00	4.88

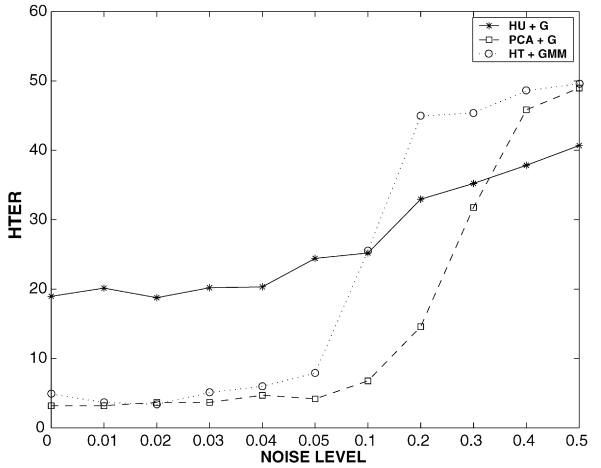

Fig. 8. HTER for noise corrupted silhouettes.

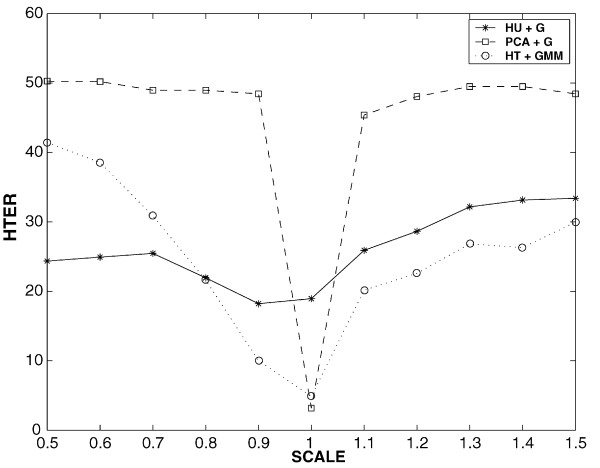

Fig. 9. HTER for scaled silhouettes.

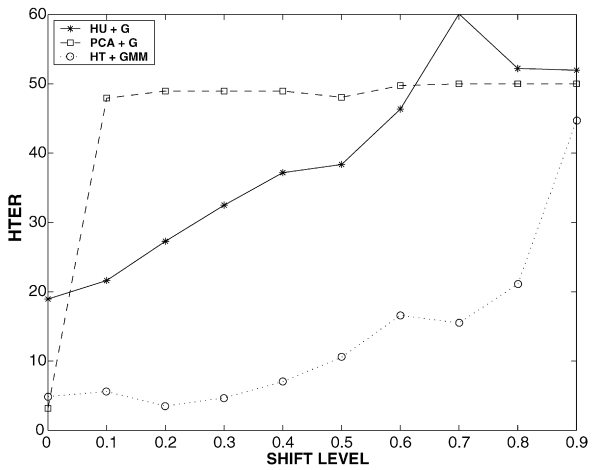

Fig. 10. HTER for clipped silhouettes.

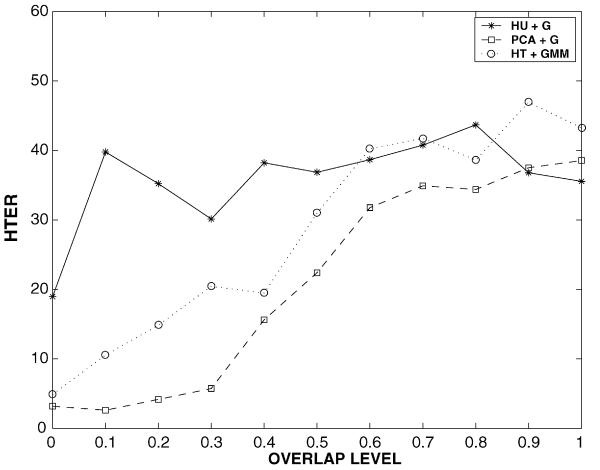

Fig. 11. HTER for occluded silhouettes.

feature pre-processing heuristic, which utilizes only rudimentary rules for determining whether a given feature vector is discriminative. For noisy images, vectors from blocks containing noise (e.g., only one white pixel) are passed to the classifier, which in turn contribute to the averages in Eq. (22). Moreover, due to the nature of the block-by-block analysis, vectors from blocks which contain noise-like patterns (e.g., a block containing one white pixel in the bottom right corner) are used during the modelling stage. This causes the object models to represent noise-like patterns, which in turn means that vectors from noisy blocks tend to give relatively high likelihoods, from both the silhouette specific and generic models. The differences in the averages in Eq. (22) are thus reduced, leading to worse performance.

The results further show that the PCA + G approach is very sensitive to translations and scale changes. The Hu + G approach is somewhat less affected, due to the explicit translation and scale normalization steps used in obtaining the moment invariants. However, as soon as part of each silhouette is moved out of view, the performance of the Hu + G

approach rapidly decays. This is due to the number of white pixels decreasing and the center of mass now being in a different location. In stark contrast, the performance of the HT + GMM approach stays relatively constant until shift level is more than 40% of the image width. The relative robustness of the HT + GMM approach to translations can be attributed to the use of the parts based representation and the loss of spatial relation between each part (i.e., the location of each part of a silhouette has little influence).

The HT + GMM approach is also affected by scale changes, but considerably less than the two holistic approaches. We conjecture that the relative robustness to scale changes stems from some object attributes remaining relatively constant across a small range of scale changes. An example of this would be the mast of a ship joining the ship's hull, represented by the shape "⊥"; if we split this shape into two parts, we obtain "⌟" and "⌞"; while scale changes would in effect move the two parts, their shape would stay relatively constant.

Occlusion experiments show that the Hu + G approach is easily affected. Again, this is due to the number of white pixels increasing (due to the presence of another object), and the center of mass now being in a different location. The PCA + G approach tolerates overlaps upto approximately 30% before its performance starts to significantly degrade. The relative robustness of the PCA approach can be partly attributed to utilizing little or no information from the sides of the image, where the overlap is most pronounced. More formally, the elements, representing the sides of the image, in the eigenvectors of $\mathbf{U}$ in Eq. (4), are relatively small compared to elements describing, say, the middle of the image. This is due to most training objects not extending all the way to the edges of the image (see Fig. 1 for an example). While the HT + GMM approach is not as robust to occlusions as the PCA + G approach, its performance is considerably better than the performance of the HT + G approach. When the overlap is small, only a small subset of the blocks is affected.

4.6. Experiments with out-of-plane rotations

In this set of experiments, dataset configuration B was used, where training images are restricted to come from "side-on" views while the test images cover a wide range of angles (i.e., $\pm 10°, \pm 20°, \ldots, \pm 90°$). Performance subject to mismatches between training and test angles will indicate how sensitive the approaches are to out-of-plane rotations of silhouettes.

The systems were optimized on clean evaluation data as described in Section 4.5. Rather than using test images from all angles at the same time, images from each angle were considered separately; this lead to, for each angle, a total of 48 true object presentations and $48 \times 8 = 384$ impostor object presentations. The results presented in Fig. 12 show that all three approaches are affected by angle changes, with the performance getting worse as the difference between the training and test angles increases.

Assuming images for other angles are available, one way to address the performance degradation would be to directly utilize the images during training, resulting in models which represent the silhouettes at multiple views. However, this may result in the models being too broad (i.e., not discriminative enough), thus reducing performance. If so, this would imply that an angle determination step is necessary prior to classification, in order to select the most appropriate model. A related question is: how many discrete angles are necessary to adequately cover a wide range of angles? Lastly, what is the range of angles beyond which the performance deteriorates too much?

To address these concerns, experiments using configuration C of the dataset were performed. Here four different subsets of training angles were used:

- Subset $\alpha = \{-5°, 5°\}$.
- Subset $\beta = \{-45°, -5°, 5°, 45°\}$.
- Subset $\gamma = \{-85°, -45°, -5°, 5°, 45°, 85°\}$.
- Subset $\delta = \{-85°, -75°, \ldots, -15°, -5°, 5°, 15°, \ldots, 75°, 85°\}$.

Results for the PCA + G and HT + GMM approaches are shown in Figs. 13 and 14, respectively. While training with subset δ provides the best overall performance, using the much smaller subset γ provides similar performance, suggesting that only a few angles, which are widely spaced apart, are required. When using subset δ for training, relatively good performance is achieved between $-70°$ and $70°$; for subset γ, the range of angles is $-60°$ and $60°$. The good performance for angles in between the angles specified by subset γ (e.g., between $45°$ and $85°$) suggests that each training angle in effect covers neighbouring angles, and/or the models in effect "interpolate" between angles.

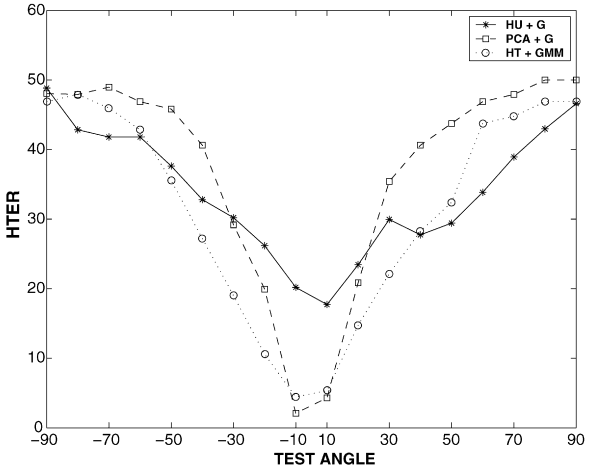

Fig. 12. HTER performance for out-of-plane rotated objects (dataset configuration B).

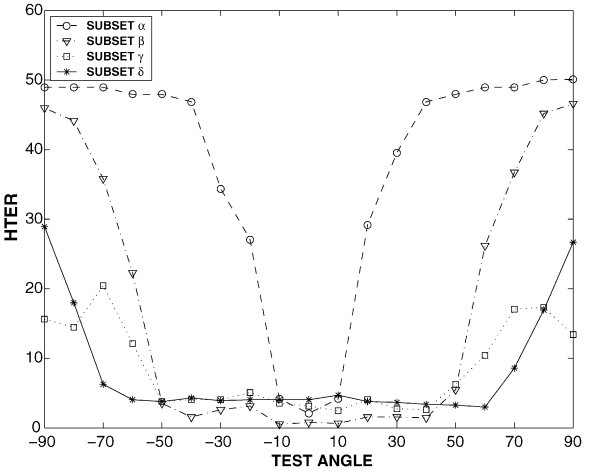

Fig. 13. HTER performance for PCA + G approach using out-of-plane rotated objects (dataset configuration C).

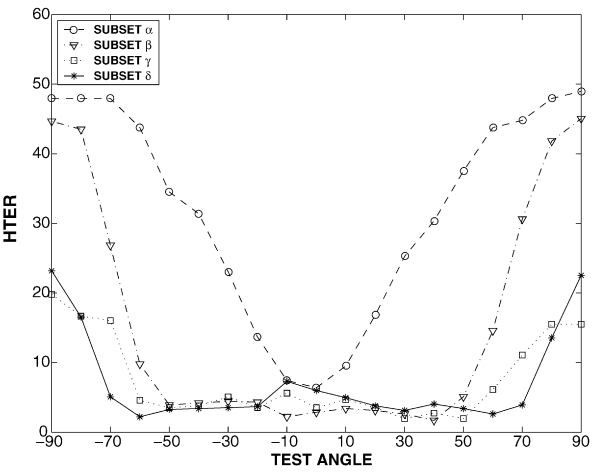

Fig. 14. HTER performance for HT + GMM approach using out-of-plane rotated objects (dataset configuration C).

For "side-on" views, the performance of systems trained with subsets β, γ, and δ is equal or better than the system trained with subset α, indicating that extending the training set does not cause a performance degradation. For test angles $\pm 90^\circ$ (i.e., the fronts and ends of objects), the use of subsets γ and δ results in considerably better performance than using subsets α and β. However, at $\pm 90^\circ$ the performance is still quite poor, suggesting that in this case it would be more appropriate to use models specifically trained for either -90° or 90°; note that this would require being able to determine the angle of the silhouette, which introduces additional complexity.

5. Main findings and future work

In this paper we evaluated the performance of several approaches to automatic target recognition (ATR) based on silhouette representation of objects. Performance was evaluated under the simulated conditions of imperfect localization by a region of interest (ROI) algorithm (resulting in clipping and scale changes) as well as occlusions by other silhouettes, noise and out-of-plane rotations. Furthermore, we demonstrated that performance degradation due to mismatches between training and test angles of silhouettes (caused by an out-of-plane rotation) can be considerably reduced by extending the training set using only a few angles which are widely spaced apart. The improvement comes without affecting the performance at "side-on" views.

Two traditional approaches (based on Hu's moment invariants and principal component analysis (PCA)) were evaluated as well as a proposed local feature approach comprised of a block-by-block 2D Hadamard transform (HT) coupled with a Gaussian mixture model (GMM) classifier. The first two methods are holistic in nature while the proposed method (HT + GMM) decomposes a silhouette into many parts prior to classification.

In the HT + GMM approach the relative location of component parts of each silhouette are disregarded, resulting in good robustness to translations and, to a lesser extent, robustness to scale changes. This is advantageous, as mistakes in the location and scale of a silhouette can be caused by a ROI localization algorithm [14].

On the other hand, the HT + GMM approach in its current form is sensitive to noise (or imaging artefacts). The results further show that the moment invariants based approach achieves poor performance in advantageous conditions and is easily affected by clipping and occlusions. The PCA based approach is highly affected by scale changes and clipping, while being relatively robust to occlusions and noise.

In terms of a silhouette based ATR system, there are several implications from the above observations. The performance can be *highly dependent* on the performance of the region of interest (ROI) localization algorithm (i.e., the algorithm's ability to accurately locate an object, with no clipping or scale problems). As such, the pattern classification stage following the ROI localization stage must be designed to handle imperfectly located ROI. Moreover, out of the considered combinations of feature extractors and classifiers, no approach is immune to all evaluated conditions.

The sensitivity of the HT + GMM approach can be addressed by further research. For example, the feature preprocessing heuristic, described in Section 4.4, could be improved by using a block occupation measure: only vectors, from blocks where the number of white pixels falls within upper and lower limits, could be used. This would reduce training with noise-like patterns and passing many noisy patterns to the classifier. Alternatively, a "bag of words" technique [39,40] could be used, where a given ROI would be described in terms of "visual words." Here, each object part would be quantized into one of a large number of pre-set object parts ("words") prior to classification. Furthermore, instead of using a direct probabilistic similarity measure (see Eq. (14)), a histogram of "visual word" occurrences could be built and then classified by a kernel based technique such as the support vector machine (SVM) [28,29]. Due to regularization as well as the margin constraint, SVMs may have the advantage of being able to use (in effect) only the most stable and/or discriminative features[7] (where, in our case, the features would be object parts). Lastly, we note that it is possible to use other local feature extraction methods—for example, Zernike moments [21,22] as well as SIFT descriptors [42] could be evaluated in future work.

Acknowledgments

We thank the anonymous reviewers for helping to improve the presentation of this paper. This paper is based in part on research performed under a CSSIP commercial research contract. NICTA is funded by the Australian Government's *Backing Australia's Ability* initiative, in part through the Australian Research Council.

References

[1] N. Nandhakumar, J. Michel, D. Arnold, G. Tsihrintzis, V. Velten, Robust thermophysics-based interpretation of radiometrically uncalibrated IR images for ATR and site change detection, IEEE Trans. Image Process. 6 (1) (1997) 65–78.
[2] A. Yilmaz, K. Shafique, M. Shah, Target tracking in airborne forward looking infrared imagery, Image Vision Comput. 21 (2003) 623–635.
[3] Chambers Science and Technology Dictionary, Chambers, England, 1991.
[4] J. Ratches, C. Walters, R. Buser, B. Guenther, Aided and automatic target recognition based upon sensory inputs from image forming systems, IEEE Trans. Pattern Anal. Machine Intell. 19 (9) (1997) 1004–1019.
[5] A. Waxman, M. Seibert, A. Gove, D. Fay, A. Bernardon, C. Lazott, W. Steele, R. Cunningham, Neural processing of targets in visible, multispectral IR and SAR imagery, Pattern Recogn. 8 (7–8) (1995) 1029–1051.
[6] J. Mitzel, Multitarget tracking applied to automatic target recognition with an imaging infrared sensor, in: Y. Bar-Shalom (Ed.), Multitarget-Multisensor Tracking: Advanced Applications, vol. 1, Artech House, Norwood, MA, 1990, pp. 297–320.
[7] P. Bharadwaj, P. Runkle, L. Carin, Infrared-image classification using expansion matching filters and hidden Markov trees, in: Proc. Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP), vol. 3, 2001, pp. 1553-1556.
[8] S. Der, R. Chellappa, Probe-based automatic target recognition in infrared imagery, IEEE Trans. Image Process. 6 (1) (1997) 92–101.
[9] S. Rizvi, N. Nasrabadi, Fusion of FLIR automatic target recognition algorithms, Inform. Fusion 4 (2003) 247–258.

[7] A demonstration of this ability is shown in [41], in the context of cancer classification.

[10] J.A. Alves, Recognition of ship types from an infrared image using moment invariants and neural networks, MS thesis, US Naval Postgraduate School, 2001.
[11] V. Gouaillier, L. Gagnon, Ship silhouette recognition using principal components analysis, in: Applications of Digital Image Processing XX, Proc. SPIE, vol. 3164, 1997.
[12] D.N. Kato, R.D. Holben, A.S. Politopoulos, B.H. Yin, Ship classification and aimpoint maintenance, in: Infrared Systems and Components II, Proc. SPIE, vol. 890, 1988.
[13] B. Kamgar-Parsi, B. Kamgar-Parsi, A. Jain, J. Dayhoff, Aircraft detection: A case study in using human similarity measure, IEEE Trans. Pattern Anal. Machine Intell. 23 (12) (2001) 1404–1414.
[14] S. Rizvi, N. Nasrabadi, A modular clutter rejection technique for FLIR imagery using region-based principal component analysis, Pattern Recogn. 35 (2002) 2895–2904.
[15] S.-G. Sun, H.W. Park, Automatic target recognition using boundary partitioning and invariant features in forward–looking infrared images, Opt. Eng. 42 (2) (2003) 524–533.
[16] R.C. Gonzales, R.E. Woods, Digital Image Processing, Addison–Wesley, Reading, MA, 1992.
[17] D. Nair, J.K. Aggarwal, Bayesian recognition of targets by parts in second generation forward looking infrared images, Image Vision Comput. 18 (2000) 849–864.
[18] A. Pentland, B. Moghaddam, T. Starner, View-based and modular eigenspaces for face recognition, in: Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition, Seattle, 1994, pp. 84–91.
[19] C. Sanderson, S. Bengio, Y. Gao, On transforming statistical models for non-frontal face verification, Pattern Recogn. 39 (2) (2006) 288–302.
[20] S. Dudani, K. Breeding, R. McGhee, Aircraft identification by moment invariants, IEEE Trans. Comput. 26 (1) (1977) 39–45.
[21] A. Khotanzad, J.-H. Lu, Classification of invariant image representations using a neural network, IEEE Trans. Acoust. Speech Signal Process. 38 (6) (1990) 1028–1038.
[22] J. Wood, Invariant pattern recognition: A review, Pattern Recogn. 29 (1) (1996) 1–17.
[23] B. Li, R. Chellappa, Q. Zheng, S. Der, N. Nasrabadi, L. Chan, L. Wang, Experimental evaluation of FLIR ATR approaches—A comparative study, Comput. Vision Image Understand. 84 (2001) 5–24.
[24] F. Cardinaux, C. Sanderson, S. Bengio, User authentication via adapted statistical models of face images, IEEE Trans. Signal Process. 54 (1) (2006) 361–373.
[25] S. Loncaric, A survey of shape analysis techniques, Pattern Recogn. 31 (8) (1998) 983–1001.
[26] D. Zhang, G. Lu, Review of shape representation and description techniques, Pattern Recogn. 37 (2004) 1–19.
[27] C.J.S. deSilva, G. Lee, R. Johnson, All-aspect ship recognition in infrared images, in: Proc. Electronic Technology Directions to the Year 2000, Adelaide, 1995, pp. 194–198.
[28] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, second ed., Wiley, New York, 2001.
[29] A. Webb, Statistical Pattern Recognition, Wiley, UK, 2002.
[30] S. Lucey, T. Chen, A GMM parts based face representation for improved verification through relevance adaptation, in: Proc. IEEE Conf. Computer Vision and Pattern Recognition, vol. 2, 2004, pp. 855–861.
[31] G.K. Wallace, The JPEG still picture compression standard, IEEE Trans. Consum. Electron. 38 (1) (1992) xviii–xxxiv.
[32] J. Picone, Signal modeling techniques in speech recognition, Proc. IEEE 81 (9) (1993) 1215–1247.
[33] D. Reynolds, T. Quatieri, R. Dunn, Speaker verification using adapted Gaussian mixture models, Digital Signal Process. 10 (1–3) (2000) 19–41.
[34] A.P. Dempster, N.M. Laird, D.B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, J. R. Statist. Soc. B 39 (1) (1977) 1–38.
[35] J.-L. Gauvain, C.-H. Lee, Maximum a posteriori estimation for multivariate Gaussian mixture observations of Markov chains, IEEE Trans. Speech Audio Process. 2 (2) (1994) 291–298.
[36] S. Bengio, J. Mariéthoz, S. Marcel, Evaluation of biometric technology on XM2VTS, IDIAP Research Report 01-21, Martigny, Switzerland, 2001.
[37] G.R. Doddington, M.A. Przybycki, A.F. Martin, D.A. Reynolds, The NIST speaker recognition evaluation—Overview, methodology, systems, results, perspective, Speech Commun. 31 (2–3) (2000) 225–254.
[38] A. Khotanzad, Y. Hong, Invariant image recognition by Zernike moments, IEEE Trans. Pattern Anal. Machine Intell. 12 (5) (1990) 489–497.
[39] G. Csurka, C.R. Dance, L. Fan, J. Willamowski, C. Bray, Visual cetegorization with bags of keypoints, in: Workshop on Statistical Learning in Computer Vision (in conjunction with ECCV'04), Prague, 2004.
[40] J. Sivic, A. Zisserman, Video Google: A text retrieval approach to object matching in videos, in: Proc. 9th International Conference on Computer Vision (ICCV), vol. 2, 2003, pp. 1470–1477.
[41] I. Guyon, J. Weston, S. Barnhill, V. Vapnik, Gene selection for cancer classification using support vector machines, Machine Learn. 46 (2002) 389–422.
[42] D.G. Lowe, Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vision 60 (2) (2004) 91–110.

Dr. Conrad Sanderson is a researcher at NICTA and an adjunct research fellow at the Australian National University. He received the Ph.D. degree in 2003 from Griffith University (Queensland, Australia), in the area of information fusion applied to speech- and face-based biometric person recognition. He has worked on a number of applied research projects, including robust speech recognition at the Advanced Telecommunication Research Laboratories (Japan), audio-visual biometrics at the IDIAP Research Institute (Switzerland), ship classification in infra-red images at the Centre for Sensor Signal and Information Processing

(CSSIP), as well as natural language processing (author identity deduction) and bioinformatics (feature selection for cancer classification) at NICTA. His current research interests are in the areas of machine learning, pattern recognition and computer vision, with applications such as intelligent surveillance.

Dr. Danny Gibbins is a senior research fellow with the Sensor Signal Processing Group at the University of Adelaide. He received the Ph.D. degree in machine vision from Flinders University in South Australia in 1995. Between 1995 and 2005 he was a research fellow with the CRC for Sensor Signal and Information Processing (CSSIP) where he was involved in the development of surveillance, classification and recognition systems for radar and electro-optical systems. Since 2005 he has focussed on video surveillance and 3D terrain modelling applications for unmanned air vehicles (UAVs). His research areas include signal and image processing, object classification and automatic target recognition and 3D reasoning using LADAR.

Stephen Searle completed the B.Sc. (Hons) at Flinders University of South Australia in 1991 and the M.App.Sc. at the University of Adelaide in 1997. His thesis entitled "Matched Doppler Processing" examined the modelling of Doppler effect in matched field processing solutions for moving target localisation with acoustic measurements. Since 1992 he has worked for government, academia and industry in the broad area of digital signal processing research. Applications have included sonar signal processing and array shape estimation, communications signal processing (in particular baseband CDMA simulation), motion detection and tracking with visual-band and infrared imagery, target detection and parameter estimation. Since 2006 Stephen has been employed as a research fellow by the University of Melbourne. His current research interests include target motion analysis, radar signal processing and waveform design.