# Critical Object Recognition in Millimeter-Wave Images with Robustness to Rotation and Scale

HODA MOHAMMADZADE[1,*], BENYAMIN GHOJOGH[1,+], SINA FAEZI[1,2,+], AND MAHDI SHABANY[1]

[1] Department of Electrical Engineering, Sharif University of Technology, Tehran, Iran
[2] Department of Electrical Engineering, University of California, Irvine, California, US
[*] Corresponding author: hoda@sharif.edu
[+] Benyamin Ghojogh and Sina Faezi contributed equally to this work.

Compiled March 30, 2017

**Locating critical objects is crucial in various security applications and industries. For example, in security applications, such as in airports, these objects might be hidden or covered under shields or secret sheaths. Millimeter-wave images can be utilized to discover and recognize the critical objects out of the hidden cases without any health risk due to their non-ionizing feature. However, millimeter-wave images usually have waves in and around the detected objects, making object recognition difficult. Thus, regular image processing and classification methods cannot be used for these images and additional pre-processings and classification methods should be introduced. This article proposes a novel pre-processing method for canceling rotation and scale using Principal Component Analysis (PCA). In addition, a two-layer classification method is introduced and utilized for recognition. Moreover, a big dataset of millimeter-wave images is collected and created for experiments. Experimental results show that a typical classification method such as SVM can recognize 45.5% of a type of critical objects at 34.2% FAR, which is a drastically poor recognition. The same method within the proposed recognition framework achieves 92.9% recognition rate at 0.43% FAR which indicates a highly significant improvement. The significant contribution of this work is to introduce a new method of analyzing millimeter-wave images based on machine vision and learning approaches, which is not yet widely noticed in the field of millimeter-wave image analysis.**


***OCIS codes:*** *Robust object recognition, millimeter-wave images, image processing, rotation, scale, critical objects.*

http://dx.doi.org/10.1364/ao.XX.XXXXXX

## 1. INTRODUCTION

Millimeter-wave (mmW) images are images produced using electromagnetic waves with frequency of 30 to 300 GHz (1 cm to 1 mm) [1–5]. In better words, when the mmW radio signals are introduced to the objects, they are scattered with a specific pattern according to the shape and material of the object. Either the pattern and characteristics of the scattered signal or the resulted image can be processed and studied in order to figure out the shape and material of the object.

While there are different applications of mmW imaging such as non-destructive testing [6], breast cancer [7] and remote sensing [8], one of the important applications of mmW images is to analyze the objects which are ambiguous and hidden beneath a cover such as clothes [1, 9–11]. These widespread applications are borne due to the intrinsic characteristic of electromagnetic waves in this range of frequency that can penetrate through interior body structures and clothing without bringing harmful ionizing effects to the body. Moreover, millimeter-wave systems provide a very high resolution due to their relatively short wavelength (1-10 mm). The other benefit of mmW images is that they can penetrate through hazy, dusty, snowy, or rainy weather, while these climate conditions make visual or infrared observation difficult [9, 12].

Millimeter-wave imaging systems are divided into two main categories, passive and active systems. Passive systems capture the ambient radiation reflected from the objects [12]. However, active systems emit the mmW signal toward the object and then process the reflected and scattered signal off the object. The advantage of active systems is a higher resolution, sensetivity and speed in acquisition [9]. There are various researches performed for improving both active [1, 9, 13–16] and passive [11, 17, 18] mmW imaging systems. There are also some works in which information of mmW images is combined (fused) with other types of images such as infrared [19]. Most of the works on mmW images deal with the Fourier transform or microwave pattern of signals. Most of Fourier-based approaches are for image reconstruction from holograms and not for object recognition.

After receiving the hologram in active mmW system, first the image reconstruction is performed, and then the image recognition and improvement is applied. This work deals with the image recognition and improvement part. Lack of researches with computer vision and statistical learning vantage point can be sensed in mmW imaging. The reason of this lack lies in the fact that the mmW images are blurry and wavy and therefore, the conventional image processing methods cannot be directly applied to them. For instance, Local Binary Pattern (LBP) [20, 21], Scale-Invariant Feature Transform (SIFT) [20, 22], Histogram of Oriented Gradients (HOG) [20, 23], bag of words [20, 24], and Gabor filters [20, 25] are not directly effective on these images. In addition to image understanding methods, the ordinary classification methods, such as Linear Discriminant Analysis (LDA) [26, 27] and Support Vector Machines (SVM) [27, 28], are not necessarily practical for blurry mmW images because of the wave artifacts on the object and also on its background.

According to the mentioned problems in both processing and classification of wavy mmW images, this work aims to introduce a machine vision and learning approach toward mmW imaging, which has been rarely addressed in literature. The contribution of this work can be summarized as follows. First, using the proposed alignment method, possible rotation and scales of the objects are canceled in order to prepare proper inputs for classification. Then, critical versus non-critical objects are effectively classified using the proposed layered classifier. This classifier is constructed by cascading two simpler classifiers, each of which performs poorly on mmW images by itself. Experimental results show that the recognition rate of a type of critical objects by a typical classification method such as SVM is 45.5% at 34.2% False Alarm Rate (FAR), which is obviously not satisfactory. The proposed classification method, however, achieves 92.9% recognition rate at 0.43% FAR, indicating a highly significant improvement.

The remaining of this article is organized as follows. Section 2 explains the proposed method which includes pre-processing, canceling rotation and scale, and the proposed layered classification method. In Section 3, the prepared dataset is introduced and then the experiments of this method and its comparison with some existing classification methods are reported. Finally, Section 4 concludes the paper.

## 2. METHODOLOGY

This work contributes in two main sections: (I) alignment for canceling rotation and scale, and (II) layered classification. The input of the proposed framework is mmW image which has wave artifacts. The images can be in any size as they are resized and scaled afterwards. The proposed framework can be implemented as a real-time system because after offline training, the images are only required to be pre-processed and projected into the classification projection directions.

The proposed method consists of four steps, i.e., pre-processing, canceling rotation, resizing objects (canceling scale), and classification. Pre-processing, itself, includes resampling images and background removal. Canceling rotation and resizing objects make the method robust to rotation and scale, respectively. In the classification section, the proposed layered classification frameworks, which are specialized for mmW images, are introduced. In the following, these steps are explained in detail.

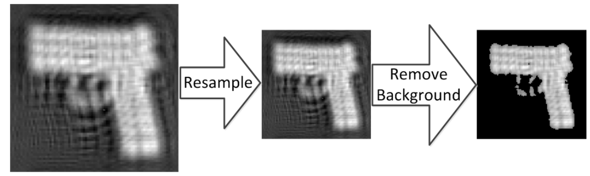


**Fig. 1.** Pre-processing including resampling image and background removal.

### A. Pre-processing

Pre-processing includes resampling images and removing background which are performed as follows.

#### A.1. Resampling Images

The input mmW images might be in different sizes. They may also be in high-resolution quality which is much time consuming for processing. Therefore, the images should be resampled to a fixed acceptable size. In this work, all input images are resampled on a uniform $200 \times 200$ grid (see Fig. 1). Notice that the images are also converted to gray-scale if they are in RGB format.

#### A.2. Background Removal

Millimeter-wave images have artifacts of wave reflection which scatter around the target object. The wave effects have destructive impact specially on the background of images.

There are various methods for removing background in the literature such as rolling ball [29]. However, in mmW images, the target object is bright and the background is dark and therefore the background can be removed by a simple thresholding technique. Suppose that $I$ denotes the maximum intensity. Background can be removed by setting intensities of pixels with intensity less than $0.5 \times I$ to be zero, where half the maximum intensity was found to be a good threshold in this work. See Fig. 1 for an example of this step.

### B. Canceling Rotation

The input images may have different rotations which make the variance of each class bigger and thus decreases the rate of recognition. The rotation of objects can be removed by first finding the principle orientation of them as follows.

The covariance matrix of the coordinates of non-zero pixels (i.e., object pixels) is formed as,

$$C = \sum_{i=1}^{n} (X - \overline{X})(X - \overline{X})^T, \quad \textbf{(1)}$$

where $n$ is the number of non-zero pixels and,

$$X = \begin{bmatrix} x_1 & x_2 & x_3 & \dots & x_n \\ y_1 & y_2 & y_3 & \dots & y_n \end{bmatrix}, \quad \textbf{(2)}$$

is the matrix of their coordinates, where the scanning of pixels can be performed either row-wise or column-wise. Also, $\overline{X}$ is the average of $X$ formulated as,

$$\overline{X} = [\frac{1}{n} \sum_{j=1}^{n} x_j, \quad \frac{1}{n} \sum_{j=1}^{n} y_j]^T. \quad \textbf{(3)}$$

The two eigenvectors of $C$ are the Principle Components (PCs) of bright pixels. These two vectors are depicted in Fig. 2.

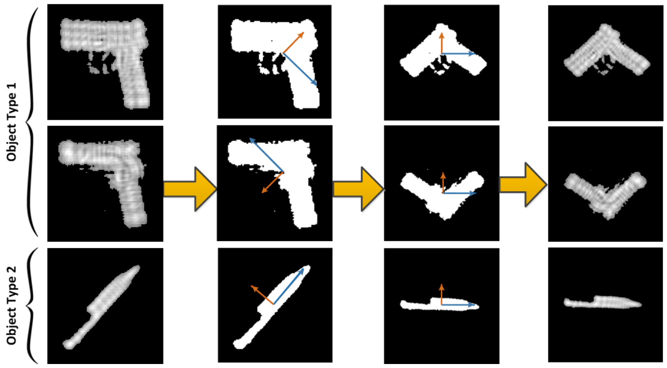


**Fig. 2.** Canceling rotation without noting the density of bright pixels around the second eigenvector.

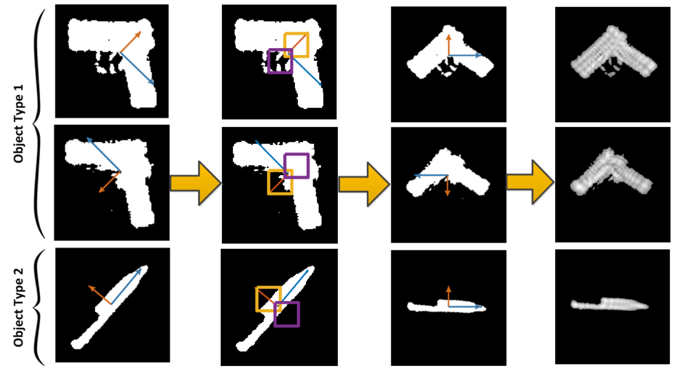


**Fig. 3.** Canceling rotation.

In the second column of this figure, the first eigenvector which has the bigger eigenvalue is depicted with blue color and the second one is shown with red vector.

The first PC is used in order to cancel the rotation. If the first eigenvector is $u_1 = [u_{11}, u_{12}]^T$, then the angle that the image should be rotated on is calculated as,

$$\theta = \tan^{-1}(\frac{u_{12}}{u_{11}}) \qquad , \theta \in [0, 2\pi]. \tag{4}$$

In other words, the image is rotated so that the first PC stands at the right horizon (see the third and forth columns of Fig. 2). Note that after rotation, the images should be again resampled on a $200 \times 200$ uniform grid.

However, as can be seen in the last column in Fig. 2, this rotation may not exactly align the objects of the same type. This happens because some objects have two main directions of bright pixels with similar eigenvalues. This causes the first PC to fall on each of them in different pictures of the same type. Therefore, the rotation needs to be corrected for better alignment. Notice that the second PC always stands at 90° lack phase in comparison to the first one. The two opposite sides of the second PC, i.e., one of them at the direction of it and the other is at the opposite direction, are considered in two squared regions of pixels with contractual similar sizes (see the second column of Fig. 3). It is desired to locate the denser region at the top after rotation. Thus, if the region at the opposite direction of the second PC has more pixels with non-zero intensities than the other region, the angle of rotation will be corrected as,

$$\theta = \tan^{-1}(\frac{u_{12}}{u_{11}}) + \pi \qquad , \theta \in [0, 2\pi]. \tag{5}$$

As can be seen in the third and forth column of Fig. 3, the rotation is corrected.

### C. Resizing Objects

At the first step of this phase, the rectangular region of image which contains the object is cropped from the rotated image (see third column of Fig. 4).

Suppose that the size of the cropped box is $[L_h, L_v]^T$. According to previous explanations, it is expected that the longest part of the object stands horizontally; meaning that $L_h \geq L_v$. Therefore, the cropped box can be resampled to 200 pixels. The vertical part of the cropped box, thus, is resampled to $\frac{L_v}{L_h} \times 200$ in order to keep the aspect ratio of the object unchanged. Afterwards, the resized rectangular box is placed at the center of

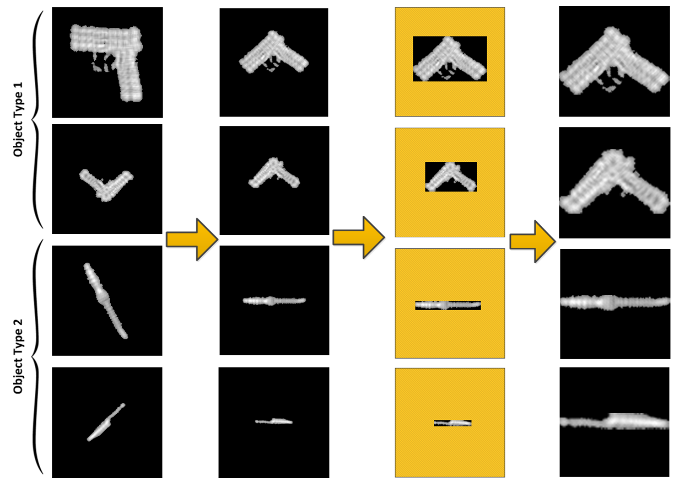


**Fig. 4.** Resizing objects.

a black $200 \times 200$ image (see the last column in Fig. 4). In order to do this, the top-left pixel of box should be placed at pixel $[1, \frac{200}{2} - \frac{L_v}{2}]^T$, if the coordinates start at 1 and from top-left.

It is worth to mention that in very rare cases, the eigenvectors do not satisfy the condition $L_h \geq L_v$. Therefore, the vertical size of the cropped box might become more than 200. Thus, all the images are finally resampled on a $200 \times 200$ grid thereafter to keep the images in a fixed size.

It is also worthy to mention that, for more complicated types of images, on which binary segmentation is performed by other methods rather than simple thresholding, the proposed rotation and scale cancellation is expected to be applicable. We are considering to apply this cancellation on other types of images, as a future work.

### D. Classification

#### *D.1. Classification Structure*

In this article, a holistic-based layered classification method is introduced which is suitable for classifying the mmW images. It can be shown that because of the characteristics of this category of images, a holistic-based approach outperforms feature-based approaches. Eigenfaces [30, 31] and Fisherfaces [32] are well-known examples of holistic-based approaches in which all pixels of the image are directly used for classification. On the other hand, in feature-based approaches, instead of raw pixel intensities, features that are either extracted using filters (e.g., Haar features) or created using histograms (e.g., HOG features) [20]

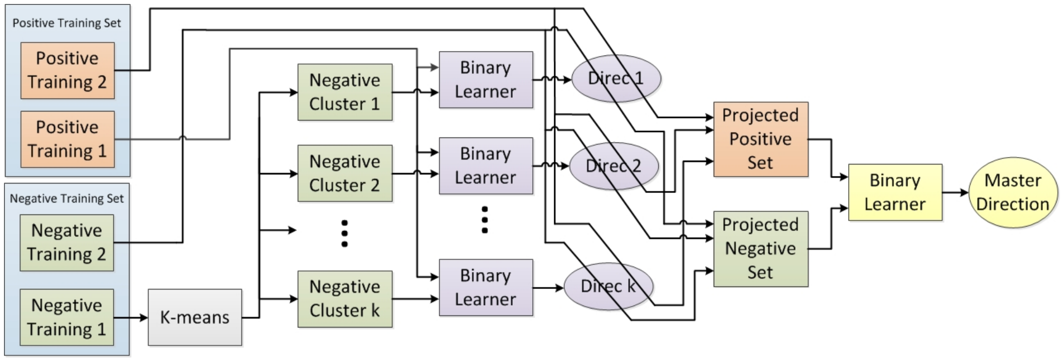


**Fig. 5.** Learning phase.

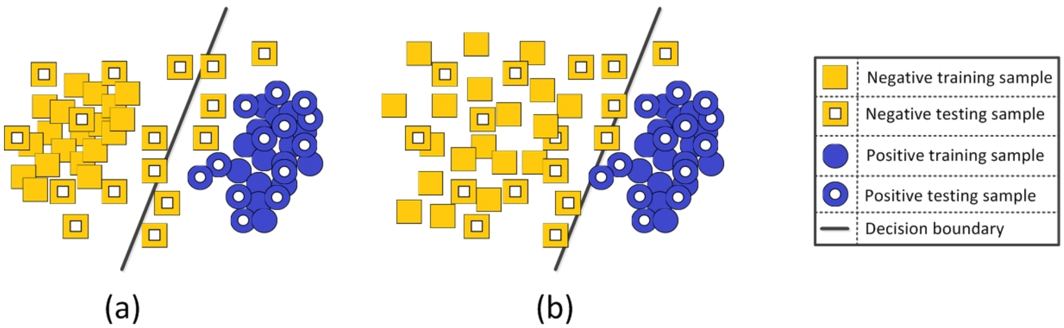


**Fig. 6.** Improvement of recognition by considering imperfect samples for training. This figure illustrates an example of training and testing samples in any of the binary classifiers of the first layer (see column fourth column of Fig. 5). The positive training and positive testing samples have similar distribution. On the other hand, the negative samples have some variances as is shown in this figure. (a) Using purely selected negative samples for training, (b) Training with imperfect negative samples.

are fed into classifiers. The main reason for failure of feature-based methods is the existence of wave artifacts in mmW images which produce strong edges and consequently create misleading features.

Throughout this paper, critical and non-critical objects are respectively referred to as positive and negative class. In the proposed classification framework, the training samples of each class are initially divided into two sets (see the first column in Fig. 5). The first set of them is used for the first layer of classification and the second set is for the second one. Each image sample in the training or test set is converted to an $mn \times 1$ vector by scanning the image either row-wise or column-wise, where $m$ and $n$ are horizontal and vertical size of images, respectively. The details of the two layers and their motivations are explained in the following.

### D.2. Clustering Negative Set

Obviously, small within- and significant between-class variance increases the chance of correct classification. Suppose that the positive set only includes one type of object (this assumption will be retreated later on), and hence its within-class variance is not salient. However, the negative training set contains lots of various objects and therefore, it has a very significant within-class variance which results in poorer classification performance. In order to deal with this problem, the negative training images are clustered into several smaller groups. This will significantly decrease the within-class variance of each negative cluster in comparison to the whole negative set.

As a result, the first set of negative samples, i.e., the negative set of the first layer, are clustered into $k$ groups using an unsupervised clustering method such as K-means algorithm (see the second and third columns in Fig. 5). The best parameter $k$ which results in better recognition can be found by validation and varies for different critical objects.

Note that, categorizing negative images using K-means is more promising than using supervised features or visual selection for this goal. In other words, pure features extracted by visual system or supervised methods are not necessarily good features for preparing training samples for binary classifiers of the first layer. Categorizing by visual selection was experimented in this work but the results were not good enough. The reason is that (perhaps) imperfect samples, categorized by unsupervised clustering methods such as K-means, can result in more precise decision boundaries. In fact, in practice, test samples might not necessarily obey the distribution of the training samples. Note that the positive samples belong to the class of a certain critical object, while the negative samples belong to the class of negative objects which includes a wide range of various non-critical objects. As a result, the estimation of the distribution of the positive samples is rather robust, while the estimation of the distribution of different sub-classes of the negative samples is not. Therefore, showing imperfect negative samples, resulted from K-means clustering, to the classifiers in the training phase improves the decision boundaries.

To better visualize the above point, see Fig. 6 which illustrates an example of this possible event for any of the binary classifiers in the first layer. As shown in this figure, when perfect negative samples are used for training, the negative test samples might not totally obey their distribution, resulting in more errors. However, by showing imperfect training samples

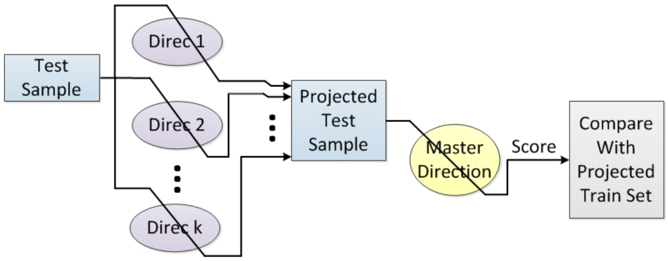


**Fig. 7.** Testing phase.

(obtained from unsupervised clustering) to the system, it performs better in recognizing imperfect test samples.

### D.3. First Layer Of Classification

There are $k$ binary classifiers in this layer of classification. Each of the negative clusters and the first set of positive samples are fed to a separate binary classifier as inputs (see the forth column in Fig. 5). The binary classifiers can be arbitrary classifier such as Fisher Linear Discriminant Analysis (LDA) [26, 27].

If these binary classifiers are LDA, then the discriminative projection directions are formed as follows. First, the dimension of input data is required to be reduced effectively using a dimensionality reduction method such as Principal Component Analysis (PCA). Details and techniques for this step can be found in the well-known Eigenfaces paper [30, 31]. Let $\mu_i$ and $\mu$ denote the mean of the $i^{th}$ class and the total mean, respectively for the new training vectors ($i \in \{1,2\}$). The within-($S_w$) and between-class ($S_b$) scatter matrices are defined as,

$$S_w = \sum_{i=1}^{2} \sum_{x_k \in \text{Class } i} (x_k - \mu_i)(x_k - \mu_i)^T, \tag{6}$$

$$S_b = \sum_{i=1}^{2} N_i(\mu_i - \mu)(\mu_i - \mu)^T, \tag{7}$$

where $N_i$ is the number of samples of the $i^{th}$ class. The eigenvectors of $S_w^{-1}S_b$ construct the discriminative projection direction, in which the within-class scatter is minimized while the between-class scatter is maximized. Since there are only two classes, one discriminative projection direction, i.e., the eigenvector corresponding to the largest eigenvalue of $S_w^{-1}S_b$ is obtained [32]. Then, the feature for an input is obtained by projecting it onto this direction, resulting in a 1-D feature, i.e., a scalar number.

After training the $k$ binary classifiers of the first layer, $k$ projection directions are created, denoted as $\overrightarrow{u_i}$ ($i \in \{1,\ldots,k\}$). Therefore, projecting an input to these $k$ discriminative projection directions results in $k$ features. Concatenating these features in a vector forms the feature vector of the first layer denoted as $\overrightarrow{f}$.

### D.4. Second Layer Of Classification

The second layer of classification consists of one binary classifier whose inputs are the feature vectors generated by the first layer (see the fifth and sixth columns in Fig. 5). Feature vectors of positive samples form the set of positive training samples and those of negative samples form the set of negative training samples for this layer. However, in order to train the classifier of this layer in a more rigorous situation and also to prevent overfitting, the feature vectors of the second set of training samples are used as the training samples for this layer. As mentioned in previous sections, only the first set of training samples is used to train the first layer, and hence, the samples of the second set have not been seen by the first layer and therefore result in a more powerful training here than the first set.

The binary classifier of the second layer can again be an arbitrary powerful classifier such as LDA or SVM [27, 28]. If the classifier of the second layer is LDA, another projection direction is created which is called here the master direction. Then, the resulting feature from this layer is a scalar number.

### D.5. Testing Phase

The testing phase is illustrated in Fig. 7. Every test sample is first projected onto the $k$ projection directions of the first layer. The resulting feature vector is then projected onto the master direction in the second layer. By doing this, a scalar score is obtained.

Now in order to determine the class of the test sample, its score is required to be compared with the score of the training samples. Various techniques can be used to perform this comparison such as K-nearest neighbors or nearest mean technique. The nearest mean method is used here as follows. The distance from the score of the test sample to the mean of the scores of the positive training samples is compared with that to the mean of the scores of the negative training samples. Then, the lower distance determines the class of the test sample.

### D.6. Using Support Vector Machines as Classifier

As previously mentioned, the classifiers of the first and second layers can be Support Vector Machine (SVM) [27, 28]. SVM constructs a hyperplane to separate the feature vectors, instead of creating a projection direction. This hyperplane maximizes the margin, that is, the distance between the nearest training samples and the hyperplanes. The parameters of the hyperplane can be expressed as,

$$\mathbf{w} = \sum_{i=1}^{N} \alpha_i y_i x_i \tag{8}$$

$$b = < y_i - \mathbf{w}^T x_i > \tag{9}$$

where $x_i$'s are the training samples, $\mathbf{w}$ is the normal vector of the hyperplane, $b$ is the bias, $\alpha_i$ is positive Lagrange multiplier, $N$ is the number of all training samples (including samples of both classes), $y_i \in \{+1,-1\}$ is the label of traning data, and $< . >$ denotes average operation. Finally, the decision function for classification can be written as,

$$F(x) = \text{sign}[\mathbf{w}^T x + b], \tag{10}$$

where $x$ is the sample to be classified. The term $(\mathbf{w}^T x + b)$ is actually the distance of the sample $x$ from the hyperplane. As mentioned before, SVM does not prepare a projection direction. Thus, if using SVM as classifier, all the previous explanations work properly; except that the projection directions should be replaced by the concept of hyperplane. In this case, for a binary classifier, instead of projecting $x$ to its discriminative projection direction, $(\mathbf{w}^T x + b)$ is calculated.

### D.7. Two Learning Structures

As was previously mentioned, both first and second layer classifiers can be either LDA, SVM, or any other classification method. In this work, two different structures are used and experimented which are LDA-LDA and SVM-LDA. In LDA-LDA, both layers use LDA classifiers while in SVM-LDA, classifiers of the first and second layers are SVM and LDA, respectively.

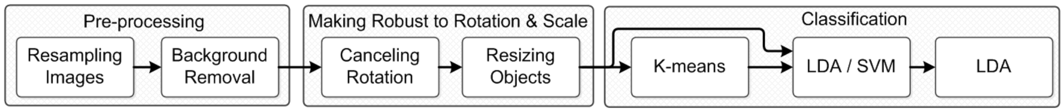


**Fig. 8.** Overall Structure of the proposed framework.

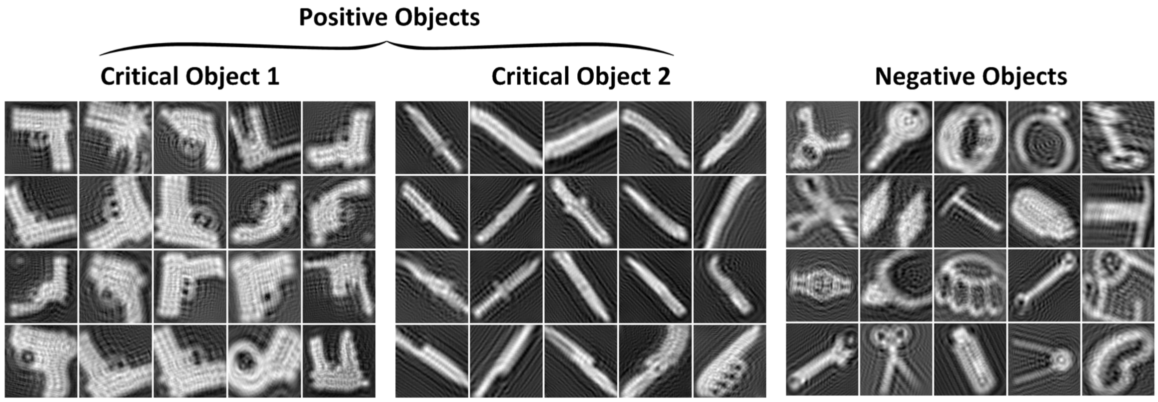


**Fig. 9.** Samples of created dataset.

**Table 1.** Settings of the simulation for creating millimeter-wave images

| | |
|---|---|
| Number Of Antennas | 82 |
| Frequency | 30 GHz |
| Distance B/W aperture plane and object | 0.4 m |
| Aperture size | $(0.48 \times 0.48)$ m$^2$ |
| Input and output image resolution | $(200 \times 200)$ pixels |

### E. Overall Structure of the Proposed Framework

The overall structure of the proposed framework is illustrated and summarized in Fig. 8. As can be seen in this figure, the steps of the proposed method is categorized to three major parts. The first section is pre-processing in which the images are resampled and backgrounds are removed. The second part is for making the method robust to rotation and scale, and includes canceling rotation and resizing objects. The last part is classification which consists of K-means, first layer (SVM or LDA), and second layer (LDA).

## 3. EXPERIMENTAL RESULTS

### A. Created Dataset

Due to the lack of proper mmW image datasets, a large mmW images dataset was created by authors to perform the experiments. This dataset was generated by a mmW image simulator based on the method in [1]. In this method, the RGB raw images are used as inputs and corresponding mmW images are simulated. In this method, the intensity of an input image is assumed to be the inductance characteristic of the imaged object. In other words, the more intense regions on the input image corresponds to the areas that have a higher inductance characteristic and therefore, are more likely to have a metallic substance. The settings for the simulation are reported in Table 1.

The created dataset contains two different types of critical (positive) images and a large set of non-critical (negative) images. There exist 614 images of critical type 1, 243 images of critical type 2, and 1773 images of negative type in this dataset. Figure 9 depicts several samples of positive (critical) and negative (non-critical) sets.

### B. Experiments on simulated dataset

Notice that, if there are multiple different types of critical objects, a separate two-layer classifier is constructed for each critical object. Then, a multiple critical object classifier is created by combining/fusing the the results of each classifier. We experimented our method on two different types of critical objects and a large negative (non-critical) set. As was explained before, the negative images are clustered using K-means. An example of this clustering is illustrated in Fig. 10.

#### *B.1. Experiment on the effect of alignment*

LDA and SVM have been chosen as baseline for comparison, as well as binary classifiers in the proposed classification framework. Obviously, these two methods can be replaced by other more complicated classifiers in the proposed framework to improve the recognition performance. In Tables 2 and 3, the recognition rates and False Alarm Rates (FAR) for the first and second type of critical objects are reported, respectively. Without aligning the objects, LDA recognizes 44.5% of the type 1 critical object with 46% FAR, which is clearly not acceptable. After performing alignment, the same method recognizes 89.7% of the type 1 critical object with 16.1% FAR, which obviously shows highly significant improvement as a result of the proposed alignment method. A similar trend of enhancement can be observed when SVM is used for classification, as well as in Table 3 for the recognition of the type 2 critical object.

#### *B.2. Experiment on the effect of layered classification*

In order to evaluate the performance of the proposed classification frameworks, the recognition rates and corresponding FARs are reported for the two LDA-LDA and SVM-LDA structures, each with three different numbers of clusters in Tables 2 and 3. For instance, with SVM-LDA and eight clusters of negative objects, 92.8% of the type 1 critical objects are recognized at the

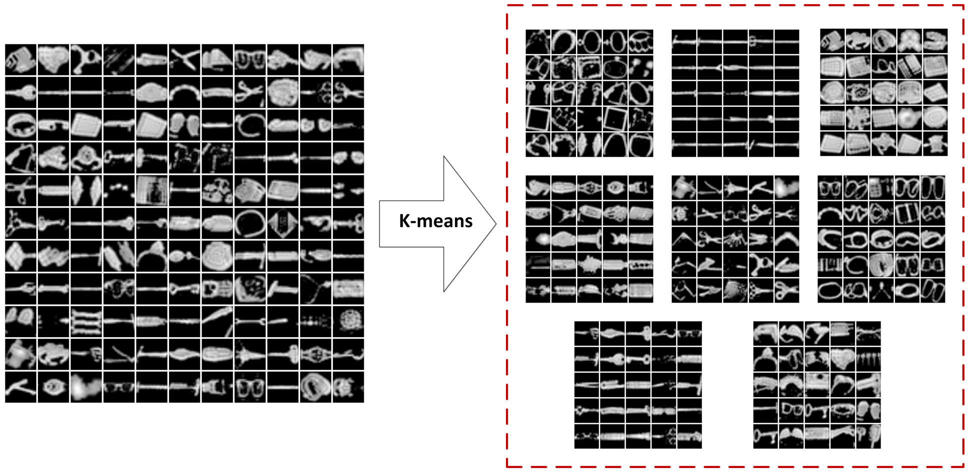


**Fig. 10.** K-means clustering on the negative set (k=8).

**Table 2.** Recognition rate of critical objects of type 1

| | Number of clusters | Recognition Rate | False Alarm Rate (FAR) |
|---|---|---|---|
| **SVM** on **non-aligned** objects | – | 45.45% | 34.2% |
| **LDA** on **non-aligned** objects | – | 44.54% | 46.00% |
| **SVM** on **aligned** objects | – | 67.57% | 12.62% |
| **LDA** on **aligned** objects | – | 89.69% | 16.11% |
| **SVM-LDA** | k = 7 | 93.96% | 0.45% |
| | k = 8 | 92.85% | **0.43%** |
| | k = 9 | 93.01% | 0.85% |
| **LDA-LDA** | k = 7 | 94.44% | 33.40% |
| | k = 8 | 94.60% | 34.56% |
| | k = 9 | **94.92%** | 31.84% |

**Table 3.** Recognition rate of critical objects of type 2

| | Number of clusters | Recognition Rate | False Alarm Rate (FAR) |
|---|---|---|---|
| **SVM** on **non-aligned** objects | – | 28.86% | 40.57% |
| **LDA** on **non-aligned** objects | – | 17.50% | 19.06% |
| **SVM** on **aligned** objects | – | 58.18% | 20.42% |
| **LDA** on **aligned** objects | – | 54.54% | 26.81% |
| **SVM-LDA** | k = 7 | 82.96% | 14.27% |
| | k = 8 | 80.74% | **11.14%** |
| | k = 9 | 83.70% | 17.59% |
| **LDA-LDA** | k = 7 | **94.81%** | 39.18% |
| | k = 8 | **94.81%** | 41.76% |
| | k = 9 | 94.07% | 40.23% |

cost of 0.43% FAR, which shows a great improvement over a single layer classifier. A similar improvement can be seen in Table 3 for the recognition of type 2 critical object. As previously mentioned, the threshold for class label decision is set to midpoint between the means of positive and negative features from training samples. Obviously, depending on the design characteristics, other desired FAR or recognition rates can be achieved by changing the decision borderline.

In order to evaluate the performance of the proposed method on classification of multiple critical objects, the overall rates are reported in Table 4. These rates are obtained by combining the two classifiers of type 1 and 2 objects in a single classification system. In other words, in this experiment, if the input object is estimated as critical object 1 or 2 in each of the

**Table 4.** Recognition rate of total critical objects

| | Number of clusters | Recognition Rate | False Alarm Rate (FAR) |
|---|---|---|---|
| **SVM** on **non-aligned** objects | – | 64.09% | 34.10% |
| **LDA** on **non-aligned** objects | – | 85.00% | 32.29% |
| **SVM** on **aligned** objects | – | 94.54% | 27.90% |
| **LDA** on **aligned** objects | – | 95.45% | 35.01% |
| **SVM-LDA** | k = 7 | 86.66% | 13.83% |
| | k = 8 | 78.88% | **6.27%** |
| | k = 9 | 91.11% | 15.49% |
| **LDA-LDA** | k = 7 | 98.88% | 55.35% |
| | k = 8 | **100%** | 57.74% |
| | k = 9 | **100%** | 60.70% |

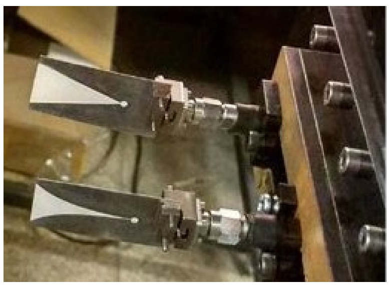

**Fig. 11.** The transmitter and receiver Vivaldi antennas.

two classifiers, it is recognized as critical; and if it is estimated as negative object in both classifiers, it is recognized as non-critical. As it is shown in Table 4, using SVM-LDA structure and eight negative clusters, 78.9% of critical objects are recognized at 6.3% FAR. Degradation in the performance in comparison with the recognition of type 1 object is due to the similarity of type 2 object to negative objects.

It is important to note that when any of the critical objects is similar to the type 2 one in this work, which is very similar to a number of negative objects, the overall performance degrades slightly. This outcome originates from the drawbacks of the current mmW imaging systems where fine characteristics of objects cannot be captured and consequently, small differences are not visible in these images even by eyes.

### C. Experiments on captured images

In spite of the fact that this work only addresses mmW images with signal processing, learning, and vision perspective, in order to test the proposed method in real applications, several mmW images obtained from real experiments were tested.

#### *C.1. The imaging system*

Several numbers of objects were imaged using two microstrip Vivaldi antennas with source power $-10$ dBm and gain 13 dB; one as transmitter and one as receiver. The antennas are shown in Fig. 11. Distance between the two antennas was 10 cm. The operating frequency was in limit $[27, 31]$ GHz, in which 51 uniform frequencies were used for sending and receiving signals. The signals were used to construct a spot of mmW image. The spatial resolutions of imaging (the distance of two neighbor spots) were $\Delta x = 4$ mm and $\Delta y = 4$ mm in this experiment,

**Fig. 12.** The setup of antennas for capturing and reconstructing mmW images.

where 4 mm is half the wavelength. In other words, the antenna was moved to next location by $\Delta x = 4$ mm, and after finishing a row of image, its vertical location was moved by $\Delta y = 4$ mm. After finishing all rows, the scan of target area is completed. The setup of antennas for capturing and reconstructing mmW images is shown in Fig. 12. The method of mmW imaging (constructing mmW images out of the the receiving mmW signals) is beyond the aim of this work and can be found for instance in [1, 9, 12].

#### *C.2. Optical and imaging properties*

Here, the properties of the captured mmW images in terms of the specifications of the imaging system are briefly discussed. The contrast of mmW image is directly affected by the source power. The larger the source power, the higher is the signal-to-noise ratio, and therefore the better contrast is achieved. In experiments of this section, the source power was $-10$ dBm. The contrast is also affected by the measurement and environmental noises. The spatial resolution, on the other hand, which determines the accuracy of capturing mmW imaging, may be categorized into two types, cross range resolution and depth resolution. In this experiment, the images are dominantly flat

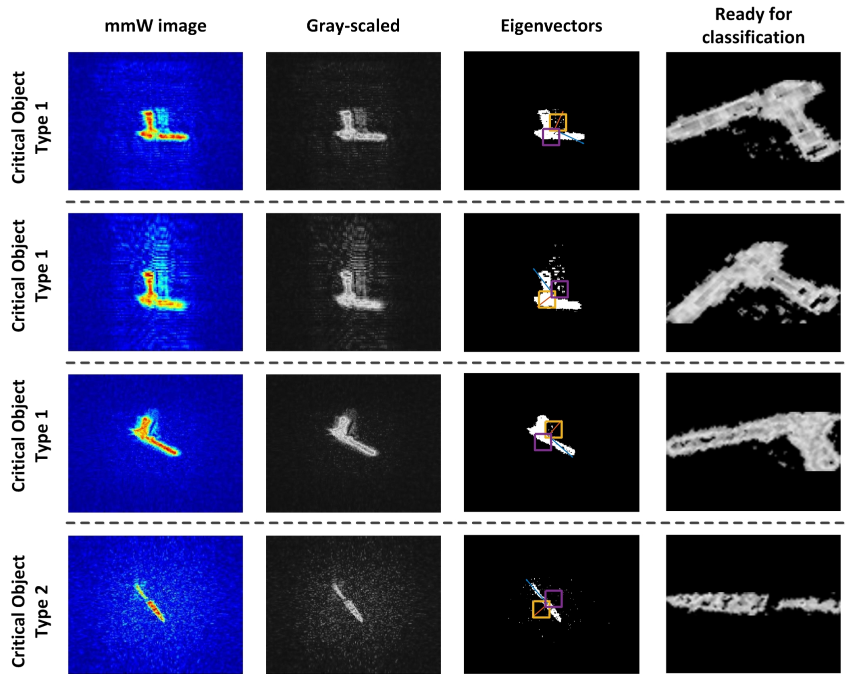


**Fig. 13.** Pre-processing on real mmW images. Columns from left to right respectively represent mmW images, gray-scaled images, eigenvectors for canceling rotation, and the prepared image for classification. (Best viewed in color).

and thus the depth resolution is not an important factor. The cross range resolution is determined by the beam width of antenna, the operating frequency, and the step of moving antenna for scanning which were mentioned in the previous paragraph.

### *C.3. Recognition using the proposed method*

Several metal toy objects were created by hand for this experiment. Objects were hidden behind cloth to simulate the real situations in which objects are probably concealed under cloth. The mentioned system in previous paragraph was used for capturing mmW images from the objects. The corresponding mmW images are shown in the first column in Fig. 13. First, the images were converted to gray scale (second column of figure), and then rotation and scale cancellation were performed on them. Note that for the real mmW images in this experiment, thresholds for background removal, rotation cancellation, and resizing objects were respectively chosen to be $0.5 \times I$, $0.5 \times I$, and $0.66 \times I$, where $I$ denotes the maximum intensity. The threshold for resizing objects was required to be better tuned here compared to the case of simulated images because of the dot noises in the images caused by real conditions and probably the effect of cloth. The resulted images to be used as inputs for the classification stage are shown in the last column in Fig. 13. The images were all recognized correctly by the proposed method.

## 4. CONCLUSION

In this article, a machine learning is introduced for the classification of mmW images, which are difficult to process because of their blurry and wavy patterns. In this work, a complete structure for recognizing critical images out of the mmW images is presented. First, an aligning and resizing method based on PCA is proposed for these images to cancel rotation and scaling of objects. Thereafter, a two-layer holistic-based classification method is introduced for classifying the mmW images. Two different learning structures are proposed and tested. The results showed highly significant improvement compared to the typical image classification methods. LDA and SVM are used as sample linear classification methods in the proposed frameworks. It is expected that, use of other more complicated classifiers can further improve the recognition performance.

## REFERENCES

1. David M. Sheen, Douglas L. McMakin, Thomas E. Hall, "Three-dimensional millimeter-wave imaging for concealed weapon detection," *IEEE Transactions on Microwave Theory and Techniques*, vol. 49, no. 9, pp. 1581-1592, 2001.
2. Roger Appleby, Rupert N. Anderton, "Millimeter-Wave and Submillimeter-Wave Imaging for Security and Surveillance," *Proceedings of the IEEE*, vol. 95, no. 8, 2007.
3. David M. Sheen, H. Dale Collins, Thomas E. Hall, Douglas L. McMakin, R. Parks Gribble, Ronald H. Severtsen, James M. Prince, Larry D. Reid, "For near real-time imaging of a target," *U.S. Patent*, no. 5,557,283, issued September 17, 1996.
4. G. Richard Huguenin, Paul F. Goldsmith, Naresh C. Deo, David K. Walker, "Contraband detection system," *U.S. Patent*, no. 5,227,800, issued July 13, 1993.
5. David Sheen, Douglas McMakin, Thomas Hall, "Near-field three-dimensional radar imaging techniques and applications," *Applied Optics*, vol. 49, no. 19, E83-E93, 2010.
6. Sergey Kharkovsky, Reza Zoughi, "Microwave and millimeter wave nondestructive testing and evaluation –Overview and recent advances," *IEEE Instrumentation & Measurement Magazine*, vol. 10, no. 2, pp. 26–38, 2007.
7. Natalia Nikolova, "Microwave imaging for breast cancer," *IEEE microwave magazine*, vol. 7, no. 12, pp. 78–94, 2011.
8. M.R. Fetterman, J. Grata, G. Jubic, W.L. Kiser Jr, A. Visnansky, "Simulation, acquisition and analysis of passive millimeter-wave images in remote sensing applications," *Optics Express*, vol. 16, no. 25, pp. 20503–20515, 2008.
9. Amir Ashkan Farsaei, Farzad Mokhtari-Koushyar, Seyed Mohammad Javad Seyed-Talebi, Zahra Kavehvash, Mahdi Shabany, "Improved Two-

Dimensional Millimeter-Wave Imaging for Concealed Weapon Detection Through Partial Fourier Sampling," *Journal of Infrared, Millimeter and Terahertz Waves*, vol. 37, no. 3, pp. 267–280, 2016.
10. Hua-Mei Chen, Seungsin Lee, Raghuveer M. Rao, M-A. Slamani, Pramod K. Varshney, "Imaging for concealed weapon detection: a tutorial overview of development in imaging sensors and processing," *IEEE signal processing Magazine*, vol. 22, no. 2, pp. 52–61, 2005.
11. Seokwon Yeom, Dong-Su Lee, Jung-Young Son, Min-Kyoo Jung, YuShin Jang, Sang-Won Jung, Seok-Jae Lee, "Real-time outdoor concealed-object detection with passive millimeter wave imaging," *Optics express*, vol. 19, no. 3, pp. 2530–2536, 2011.
12. R. Appleby, "Passive millimetrewave imaging and how it differs from terahertz imaging," *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 362, no. 1815, pp. 379–393, 2004.
13. Stuart William Harmer, Shawn Edward Cole, Nicholas John Bowring, Nacer Ddine Rezgui, David Andrews, "On body concealed weapon detection using a phased antenna array," *Progress In Electromagnetics Research* vol. 124, pp. 187–210, 2012.
14. Abdalrahman Al-Qubaa, Gui Yan Tian, John Wilson, "Electromagnetic imaging system for weapon detection and classification," *In Fifth International Conference on Sensor Technologies and Applications*, pp. 317–321, 2011.
15. David M. Sheen, Douglas L. McMakin, Thomas E. Hall, Ronald H. Severtsen, "Active millimeter-wave standoff and portal imaging techniques for personnel screening," *In Technologies for Homeland Security, 2009. HST'09. IEEE Conference on, IEEE*, pp. 440–447, 2009.
16. Erich N. Grossman, Aaron J. Miller, "Active millimeter-wave imaging for concealed weapons detection," *In AeroSense 2003, International Society for Optics and Photonics*, pp. 62–70, 2003.
17. Larry Yujiri, "Passive millimeter wave imaging," *In 2006 IEEE MTT-S International Microwave Symposium Digest, IEEE*, pp. 98–101, 2006.
18. Thomas Luthi, Christian Matzler, "Stereoscopic passive millimeter-wave imaging and ranging," *IEEE transactions on microwave theory and techniques*, vol. 53, no. 8, pp. 2594–2599, 2005.
19. Pramod K. Varshney, Hua-Mei Chen, Liane C. Ramac, Mucahit Uner, David Ferris, Mark Alford, "Registration and fusion of infrared and millimeter wave images for concealed weapon detection," *In Image Processing, 1999. ICIP 99. Proceedings. 1999 International Conference on, IEEE*, vol. 3, pp. 532–536, 1999.
20. Simon J.D. Prince, "Computer vision: models, learning and inference," *Cambridge University Press*, 2012.
21. Marko Heikkilä, Matti Pietikäinen, Cordelia Schmid, "Description of interest regions with local binary patterns," *Pattern recognition*, vol. 42, no. 3, pp. 425–436, 2009.
22. David G. Lowe, "Object recognition from local scale-invariant features," *In Computer vision, 1999. The proceedings of the seventh IEEE international conference on, IEEE*, vol. 2, pp. 1150–1157, 1999.
23. Navneet Dalal, Bill Triggs, "Histograms of oriented gradients for human detection," *In 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), IEEE*, vol. 1, pp. 886–893, 2005.
24. Jun Yang, Yu-Gang Jiang, Alexander G. Hauptmann, Chong-Wah Ngo, "Evaluating bag-of-visual-words representations in scene classification," *In Proceedings of the international workshop on Workshop on multimedia information retrieval, ACM*, pp. 197–206, 2007.
25. Ville Kyrki, Joni-Kristian Kamarainen, Heikki Kälviäinen, "Simple Gabor feature space for invariant object recognition," *Pattern recognition letters*, vol. 25, no. 3, pp. 311–318, 2004.
26. R.A. Fisher, "The Use of Multiple Measures in Taxonomic Problems," *Ann. Eugenics*, vol. 7, pp. 179-188, 1936.
27. Trevor Hastie, Robert Tibshirani, Jerome Friedman, "The Elements of Statistical Learning, Data Mining, Inference, and Prediction," *Springer*, Second Edition, 2008.
28. Christopher JC Burges, "A tutorial on support vector machines for pattern recognition," *Data mining and knowledge discovery*, vol. 2, no. 2, pp. 121–167, 1998.
29. Stanley R. Sternberg, "Grayscale morphology," *Computer Vision, Graphics, and Image Processing*, vol. 35, no. 3, pp. 333–355, 1986.
30. Mathew A. Turk, Alex P. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, pp. 72-86, 1991.
31. Mathew A. Turk, Alex P. Pentland, "Face recognition using eigenfaces," *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR '91., IEEE Computer Society Conference on*, pp. 586–591, 1991.
32. Peter N. Belhumeur, João P. Hespanha, and David J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997.