

University of Wollongong

Research Online

Faculty of Engineering and Information
Sciences - Papers: Part A

Faculty of Engineering and Information
Sciences

1-1-2013

An image-based approach for classification of human micro-doppler radar signatures

Fok Hing Chi Tivive

University of Wollongong, tivive@uow.edu.au

Son Lam Phung

University of Wollongong, phung@uow.edu.au

Abdesselam Bouzerdoum

University of Wollongong, bouzer@uow.edu.au

Follow this and additional works at: <https://ro.uow.edu.au/eispapers>



Part of the [Engineering Commons](#), and the [Science and Technology Studies Commons](#)

Recommended Citation

Tivive, Fok Hing Chi; Phung, Son Lam; and Bouzerdoum, Abdesselam, "An image-based approach for classification of human micro-doppler radar signatures" (2013). *Faculty of Engineering and Information Sciences - Papers: Part A*. 1596.

<https://ro.uow.edu.au/eispapers/1596>

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: research-pubs@uow.edu.au

An image-based approach for classification of human micro-doppler radar signatures

Abstract

With the advances in radar technology, there is an increasing interest in automatic radar-based human gait identification. This is because radar signals can penetrate through most dielectric materials. In this paper, an image-based approach is proposed for classifying human micro-Doppler radar signatures. The time-varying radar signal is first converted into a time-frequency representation, which is then cast as a two-dimensional image. A descriptor is developed to extract micro-Doppler features from local time-frequency patches centered along the torso Doppler frequency. Experimental results based on real data collected from a 24-GHz Doppler radar showed that the proposed approach achieves promising classification performance.

Keywords

radar, signatures, micro, human, classification, doppler, approach, image

Disciplines

Engineering | Science and Technology Studies

Publication Details

F. Tivive, S. Phung & A. Bouzerdoun, "An image-based approach for classification of human micro-doppler radar signatures," in Proceedings of SPIE - The International Society for Optical Engineering, 2013, pp. 873406-1-873406-12.

An Image-based Approach for Classification of Human Micro-Doppler Radar Signatures

Fok Hing Chi Tivive, Son Lam Phung and Abdesselam Bouzerdoum

School of Electrical, Computer and Telecommunications Engineering, University of Wollongong, Northfields Avenue, Wollongong NSW 2522, Australia

ABSTRACT

With the advances in radar technology, there is an increasing interest in automatic radar-based human gait identification. This is because radar signals can penetrate through most dielectric materials. In this paper, an image-based approach is proposed for classifying human micro-Doppler radar signatures. The time-varying radar signal is first converted into a time-frequency representation, which is then cast as a two-dimensional image. A descriptor is developed to extract micro-Doppler features from local time-frequency patches centered along the torso Doppler frequency. Experimental results based on real data collected from a 24-GHz Doppler radar showed that the proposed approach achieves promising classification performance.

Keywords: Spectrogram, Human micro-Doppler radar signature, Micro-Doppler descriptor, Log-Gabor filters, Two-directional two-dimensional principal component analysis, Support vector machines

1. INTRODUCTION

Identification of human gaits using Doppler radar is emerging as a very important research field due to its potential civilian and military applications, including surveillance, search and rescue, and health care monitoring. In contrast to many optical-based systems, Doppler radar provides several advantages in sensing human motions. It can operate in all weather conditions, during the day or at night. It has the capability to penetrate most non-metallic materials, making it operationally robust against clothing variations and capable of detecting humans behind opaque obstacles, such as walls and foliage. Furthermore, it can be placed at a distance from the target area and mounted unobtrusively for through-the-wall monitoring.

A Doppler radar can sense not only the translational motion of the moving target, but also any micro-motions exhibited by the moving structures attached to the target. The translational motion induces a Doppler frequency shift related to the target velocity, whereas the vibration or rotation of the target components induce additional frequency modulations on the radar returns. For example, when a human is walking, the periodic motions of the arms and legs produce sidebands next to the main Doppler frequency, creating a unique micro-Doppler signature, which can be used for identification. Since this type of radar signal is non-stationary, a time-frequency analysis method such as Short-Time Fourier Transform (STFT) is widely used for visualizing the radar signal in its joint time-frequency space and depicting the micro-Doppler signatures.

Various approaches have been proposed for classification of moving targets using their micro-Doppler signatures. Kim and Ling developed a parametric approach to distinguish seven human activities: running, walking, walking without moving arms, crawling, boxing, boxing while moving forward, and sitting relatively still.¹ In their approach, the spectrogram is firstly segmented using a threshold computed from that of the background scene devoid of any moving targets. Then, six features are extracted from the segmented spectrogram: the torso Doppler frequency, the total bandwidth of the Doppler signal, the offset of the total Doppler, the bandwidth without micro-Dopplers, the normalized standard deviation of the Doppler signal strength, and the period of the limb motion. Support vector machines (SVM) are used to classify radar signals from 12 human subjects. Kim

Further author information: (Send correspondence to F. H. C. Tivive)

F. H. C. Tivive: E-mail: tivive@uow.edu.au, Telephone: +612 4221 4814

S. L. Phung: E-mail: phung@uow.edu.au, Telephone: +612 4221 3407

A. Bouzerdoum: E-mail: a.bouzerdoum@uow.edu.au, Telephone: +612 4221 3397

and Ling showed that their approach can successfully classify all seven human activities. Based on the periodicity of the arm motion, Orovic and his colleagues employed Hermite S-method to convert the radar signal into a high-resolution time-frequency representation and developed an envelope detection method to capture the evolution of the arm swing.² From the segmented time-frequency representation, the envelope of the micro-Doppler signature is extracted for classification. Their approach assumes that the arm motion is more pronounced than the leg motion, i.e., the person sensed by the radar is marching with full arm swing. Mobasseri and Amin, on the other hand, used principal component analysis (PCA) for feature extraction.³ Data samples are defined as either time slices or frequency slices of the spectrogram. Time slices contain the Doppler frequencies, whereas frequency slices comprise the time points. By taking either the time or frequency slices as data samples, PCA is applied to generate three subspaces characterizing three different types of human motion. Samples from a test spectrogram are projected onto these subspaces. A minimum distance classifier and a majority-voting scheme are used to classify the test spectrogram. Mobasseri and Amin showed that frequency slices, which capture the periodicity of the arms and legs, give a higher classification rate than the time slices. Speech processing techniques have also been adopted for classification of micro-Doppler signatures. Mel-frequency cepstral coefficients, linear predictive coding, and cepstrum coefficients all have been employed for characterizing the micro-Doppler signatures of different types of target: humans, vehicle, truck and background clutter.⁴ Fairchild and Narayanan proposed the use of empirical mode decomposition (EMD) to decompose the Doppler signal into a set of intrinsic mode functions (IMFs) that capture the different oscillatory modes embedded in the signal.⁵ Using the IMF energies as features and SVM as a classifier, they classify four different human activities: a person breathing while standing still, a person picking up an object, a person standing up from crouching position, and a person swinging arms.

In this paper, we propose an image-based approach for classifying human micro-Doppler radar signatures. A micro-Doppler descriptor based on a bank of log-Gabor filters is developed to extract features from time-frequency patches centered along the torso frequency of the spectrogram. The proposed approach can be applied to classify signal frame of arbitrary length. The remainder of the paper is organized as follows. Section 2 describes the proposed image-based classification approach. Section 3 presents the experimental results, and Section 4 gives the conclusion.

2. PROPOSED MICRO-DOPPLER SIGNATURE CLASSIFICATION APPROACH

In the proposed approach, the input Doppler signal is first converted into a time-frequency representation, using the short-time Fourier Transform. Then, a micro-Doppler descriptor is developed to extract features from a local time-frequency patch centered along the torso frequency. After feature extraction, the size of the feature map is reduced using a matrix-based principal component analysis. Figure 1 depicts the proposed image-based classification approach.

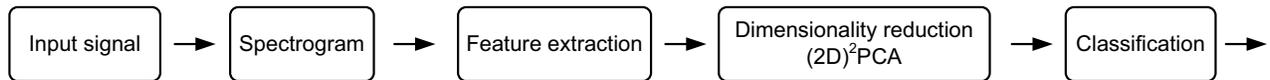


Figure 1. A schematic diagram of the proposed image-based classification system.

2.1 Time-Frequency Patch

When a Doppler signal scatters back from a moving person, the radar returns comprise multiple time-varying frequencies, which are induced by different moving body parts of the person. To capture these time-varying Doppler frequencies, the radar signal is transformed into a time-frequency representation. The short-time Fourier transform (STFT) is commonly used to convert a discrete-time Doppler signal $x(n)$ into its time-frequency representation. Mathematically, the STFT is given by

$$X(n, k) = \sum_{m=-\infty}^{\infty} x(m) w(m-n) e^{-j2\pi mk/N}, \quad k = 0, 1, \dots, N-1, \quad (1)$$

where n is the discrete-time index, k is the discrete frequency index, N is the number of frequency points, and $w(m)$ is a window function. The size of the window function is determined empirically so as to achieve a good classification. The spectrogram is obtained by taking the magnitude of the STFT, i.e., $S(n, k) = |X(n, k)|$. Figure 2 shows the spectrogram of a person walking with both arms swinging. The time-varying Doppler frequency around 200 Hz is induced by the oscillating torso, and it is considered the spine of the spectrogram. The main sidelobe represents the micro-Doppler of the leg swing and one of the sidelobes within it represents that of the arm swing.

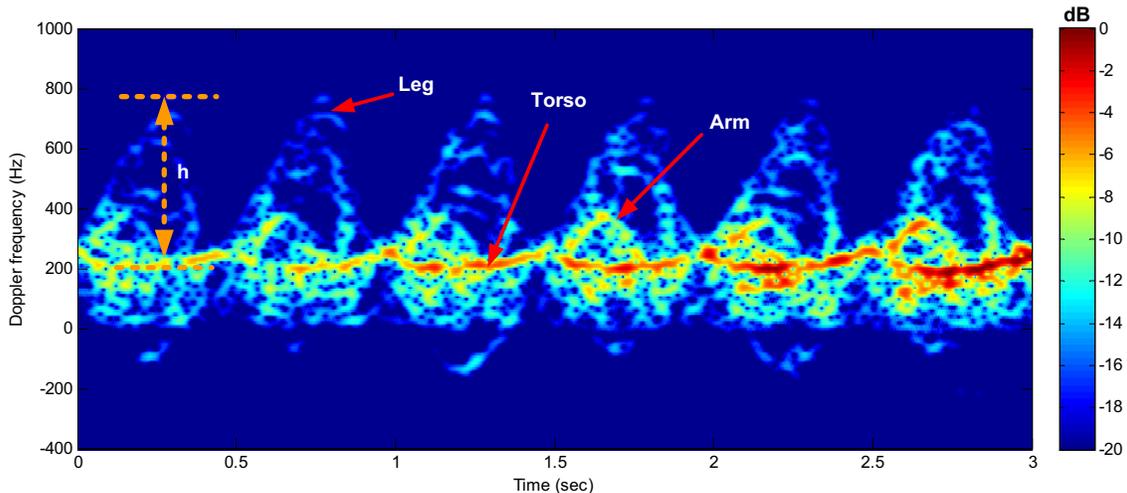


Figure 2. Spectrogram of a person walking toward the radar while swinging both arms. For visualization, the spectrogram is displayed in log scale and is thresholded at -20 dB.

Most micro-Dopplers exhibited by the moving body parts of a person are located around the spine of the spectrogram. Thus, instead of processing the entire spectrogram, we extract two-dimensional (2-D) time-frequency patches centered along the spine of the spectrogram and use them as inputs to the feature extraction stage. The spine of the spectrogram has the highest energy. It is determined by finding the maximum value at each column of the spectrogram. The extent in frequency of the patch is estimated from the height of the main sidelobe, which is due to either the motion of the leg when a person is walking or the motion of the arm when a person is marching. Before estimating the height of the sidelobe, we convolve each column of the spectrogram with a 1-D median filter and segment the spectrogram using a threshold computed from the spectrogram of the background scene. Then, we compute the frequency profile as

$$F_p(k) = \sum_{n=1}^M S(n, k), \quad (2)$$

where M is the total number of time points. The frequency profile, $F_p(k)$, is normalized to have a range $[0, 1]$. The maximum frequency of the main sidelobe peak is determined by satisfying:

$$\frac{\sum_{i=k_t}^K F_p(k)}{\sum_{i=k_t}^{N/2} F_p(k)} \geq \theta. \quad (3)$$

where θ is a given threshold and k_t denotes the frequency coordinate of the spine. Subsequently, the height of the main sidelobe is computed as $h = K - k_t$. The frequency range representing the height of the patch is computed as $[y_t - h, y_t + h]$. On the other hand, the width of the patch is chosen so as to capture at least one cycle of the full motion; the optimum size will be determined experimentally. To compensate for variations in the frequency and time, the extracted patch is resized to a predefined height, N_y , and width, N_x , before feature extraction. Since the torso pixels have higher intensity value than those of the micro-Doppler components, a contrast enhancement

technique is applied to improve the contrast of the micro-Doppler patch. Let $S(i, j)$, $i \in [1, N_y]$, $j \in [1, N_x]$ denote a micro-Doppler patch. For contrast enhancement, we use Naka-Rushton function given by

$$\tilde{S}(i, j) = \frac{S(i, j)}{S(i, j) + \mu}, \quad (4)$$

where μ is the mean intensity value of the patch:

$$\mu = \frac{1}{N_y N_x} \sum_{i=1}^{N_y} \sum_{j=1}^{N_x} S(i, j). \quad (5)$$

Figure 3 shows a time-frequency patch before and after applying the contrast enhancement technique. The micro-Doppler signature in Fig. 3(b) is more pronounced than that shown in Fig. 3(a).

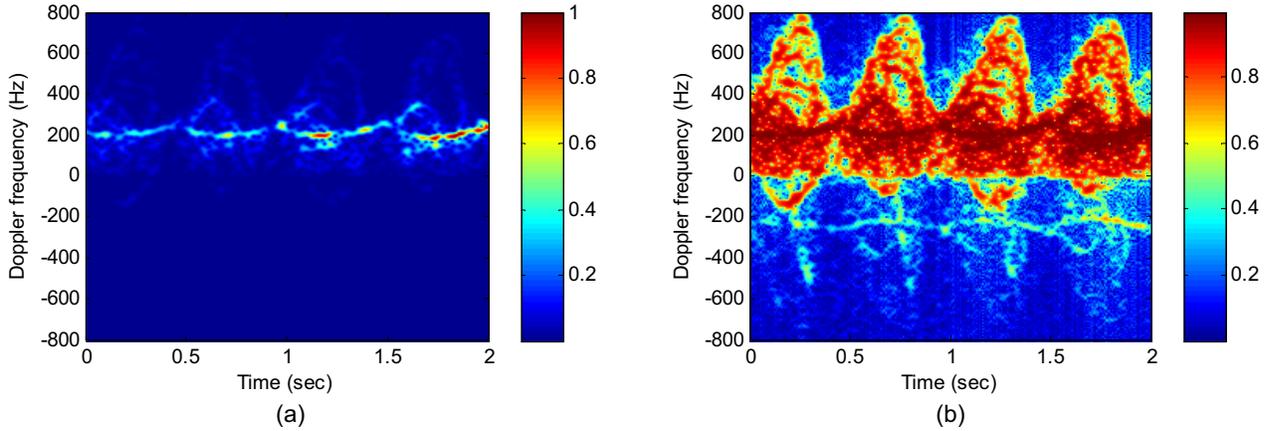


Figure 3. A time-frequency patch before and after contrast enhancement: (a) input patch and (b) contrast-enhanced patch. The patches are displayed in the linear scale.

2.2 Feature Extraction

To extract features from the time-frequency patch, we develop a micro-Doppler descriptor, which is based on the principles of the gist descriptor proposed by Oliva and Torralba.⁶ The proposed micro-Doppler descriptor employs log-Gabor filters to capture the global-level information of the patch. These Gabor filters, which were proposed by Field,⁷ have no DC component and no bandwidth limitation, compared to the normal Gabor filters. Therefore, a small set of log-Gabor filters is adequate to cover the desired frequency spectrum. On the linear frequency scale a Gabor filter has the shape of a Gaussian function, whereas log-Gabor filter has similar shape to a Gabor filter on the logarithmic frequency scale with an extended tail in the high frequency region. Due to the singularity of the log-Gabor function at the origin, log-Gabor filters are designed in the frequency domain. In a polar coordinate system, log-Gabor filters can be expressed as

$$G_k(f, \theta) = \exp\left(\frac{-(\ln(f/f_i))^2}{2(\ln(\sigma_f/f_i))^2}\right) \exp\left(\frac{-(\theta - \theta_j)^2}{2\sigma_\theta^2}\right), \quad i \in [1, N_f], j \in [1, N_\theta], k \in [1, R], \quad (6)$$

where f_i is the i -th center frequency of the filter, θ_j is the j -th orientation angle of the filter, N_f is the number of scales, N_θ is the number of orientations, $R = N_f N_\theta$ is the total number of log-Gabor filters, σ_f and σ_θ are the radial and angular bandwidth, respectively. Figure 4 illustrates some of log-Gabor filters employed by the micro-Doppler descriptor in frequency domain.

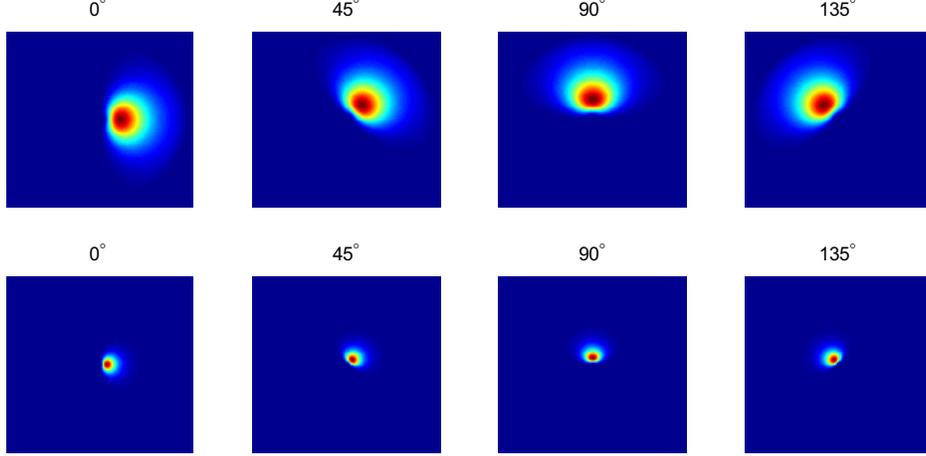


Figure 4. Examples of log-Gabor filters in frequency domain.

The first step in computing the micro-Doppler features is to convolve the input image with a set of R log-Gabor filters, $\{G_1, \dots, G_k, \dots, G_R\}$. Let I denote the Fourier transform of the contrast-enhanced patch \tilde{S} . The real, $Z_{k,1}$, and imaginary, $Z_{k,0}$, parts of the output of the k -th log-Gabor filter are given by

$$Z_{k,1} = \Re\{IFFT2(I \cdot G_k)\}, \quad (7)$$

and

$$Z_{k,0} = \Im\{IFFT2(I \cdot G_k)\}, \quad (8)$$

where IFFT2 denotes 2-D inverse Fast Fourier transform, and $\Re\{\cdot\}$ and $\Im\{\cdot\}$ are the real and imaginary operators, respectively. A half-wave rectification operation is performed to decompose each filtered image into an on- and an off-map:

$$\begin{cases} \text{on-map} : Z_{k,1,1} = \max(Z_{k,1}, 0) \\ \text{off-map} : Z_{k,1,0} = -\min(Z_{k,1}, 0) \end{cases}, \quad (9)$$

This thresholding operation is also applied to the imaginary part of the filtered images, i.e., $Z_{k,0}$, $k \in [1, R]$. The Naka-Rushton transform is then applied to the on-map and off-map, separately,

$$\tilde{Z}_{k,l,q}(i, j) = \frac{Z_{k,l,q}(i, j)}{Z_{k,l,q}(i, j) + \mu_l}, \quad l, q = 1, 0, \quad (10)$$

where $\mu_l = 1/(N_y N_x) \sum_{i=1}^{N_y} \sum_{j=1}^{N_x} |Z_{k,l}(i, j)|$.

The last step of feature extraction is to arrange the extracted features into a matrix. Each contrast-enhanced on-map or off-map is partitioned into d non-overlapping sub-regions along each direction, (horizontally and vertically), and the mean of each sub-region is taken as a feature. The mean values computed from the on-map of the real part of the k -th filtered image are stored in a vector $\boldsymbol{\mu}_{k,1,1} = [\mu_1, \dots, \dots, \mu_L]^T$, $L = d^2$, with its l -th element given by

$$\mu_m = \frac{L}{N_y N_x} \sum_{i=1+(r-1)N_y/d}^{N_y/d+(r-1)N_y/d} \sum_{j=1+(c-1)N_x/d}^{N_x/d+(c-1)N_x/d} \tilde{Z}_{k,1,1}(i, j), \quad r, c = 1, \dots, d, \quad m = 1, \dots, L. \quad (11)$$

After computing the features from all filtered images, we arrange the features into a matrix $A \in \mathbb{R}^{L \times 4R}$:

$$A = [\boldsymbol{\mu}_{1,1,0}, \boldsymbol{\mu}_{1,0,0}, \boldsymbol{\mu}_{1,1,1}, \boldsymbol{\mu}_{1,0,1}, \dots, \boldsymbol{\mu}_{R,1,0}, \boldsymbol{\mu}_{R,0,0}, \boldsymbol{\mu}_{R,1,1}, \boldsymbol{\mu}_{R,0,1}]. \quad (12)$$

After the feature extraction stage, the size of the feature matrix A is reduced using a matrix-based principal component analysis, which will be briefly described in the following section.

2.3 Two-Directional Two-Dimensional Principal Component Analysis

After feature extraction, a dimensionality reduction technique is used to compress the number of features. Two-directional two-dimensional principal component analysis ((2D)²-PCA) is applied for dimensionality reduction. It can be regarded as a two dimensional version of PCA with two projection matrices to reduce the number of rows and columns of an image simultaneously. The advantage of using (2D)²-PCA is that it is less costly to compute the covariance and projection matrices. In addition, (2D)²-PCA has been shown to achieve superior results than traditional PCA.⁸ Let a time-frequency patch denoted by a matrix $A \in \mathbb{R}^{n_r \times n_c}$. The compressed feature map $D \in \mathbb{R}^{m_r \times m_c}$, ($m_r \leq n_r, m_c \leq n_c$), can be rewritten as

$$D = \Phi^T A \Psi, \quad (13)$$

where Φ and Ψ are the projection matrices with orthonormal components. Then, the columns of matrix D are concatenated to form a feature vector, which is used as an input to the classifier. Let $Y = A\Psi$. The projection matrix Ψ can be determined by maximizing the following criterion:

$$\begin{aligned} J(\Psi) &= \text{trace}\{\text{E}[(Y - \text{E}(Y))(Y - \text{E}(Y))^T]\} \\ &= \text{trace}\{\text{E}[(A\Psi - \text{E}(A\Psi))(A\Psi - \text{E}(A\Psi))^T]\} \\ &= \text{trace}\{\Psi^T \text{E}[(A - \text{E}(A))^T(A - \text{E}(A))]\Psi\}, \end{aligned} \quad (14)$$

where we have used $\text{trace}(AB) = \text{trace}(BA)$, assuming the product BA exists. Define the image covariance matrix $G_c = \text{E}[(A - \text{E}(A))^T(A - \text{E}(A))]$, which is an n_c by n_c nonnegative definite matrix. Suppose that the training set consists of P patches $\{A_1, \dots, A_P\}$. The image covariance matrix G_c can be computed as

$$G_c = \frac{1}{P} \sum_{i=1}^P (A_i - \bar{A})^T (A_i - \bar{A}), \quad (15)$$

where \bar{A} is the average patch given by

$$\bar{A} = \frac{1}{P} \sum_{i=1}^P A_i. \quad (16)$$

The criterion in (14) can be expressed as

$$J(\Psi) = \text{trace}(\Psi^T G_c \Psi). \quad (17)$$

The criterion $J(\Psi)$ is maximized when Ψ is composed of eigenvectors of G_c corresponding to the m_c largest eigenvalues, i.e., $\Psi = [\psi_1, \dots, \psi_{m_c}]$. The other optimal projection matrix Φ comprises the eigenvectors $\{\phi_1, \dots, \phi_{m_r}\}$ of G_r associated with the first m_r dominant eigenvalues of the image covariance matrix G_r given by

$$G_r = \frac{1}{P} \sum_{i=1}^P (A_i - \bar{A})(A_i - \bar{A})^T. \quad (18)$$

The number of eigenvectors m_c is selected so that the corresponding eigenvalues of G_c satisfy the relation:

$$\frac{\sum_{i=1}^{m_c} \lambda_i}{\sum_{i=1}^{n_c} \lambda_i} \geq \beta, \quad (19)$$

where λ_i denotes the i -th eigenvalue and β is a given threshold. This scheme is also applied to determine the size of the projection matrix Ψ , i.e. the number of eigenvectors.

After feature extraction and dimensionality reduction, the last stage of the proposed method is the classification in which support vector machines (SVMs) are used as a classifier.

2.4 Classification

The classification stage employs support vector machines (SVMs) to classify the input patches into different types of motion. The key concept of SVM is to determine an optimal hyperplane used as a decision boundary to separate data points between two different classes. The hyperplane is considered to be optimal if it maximizes the distance between the hyperplane and the nearest input vector with the smallest error or without any error. The minimum distance between the hyperplane and a training vector is known as the margin; any input vector that lies on the plane defined by the margin is called a support vector. Suppose that there is a training data set $\{\mathbf{x}_i, y_i\}_{i=1}^P$, with $\mathbf{x}_i \in \mathbb{R}^n$ being the i -th input vector and $y_i \in \{1, -1\}$ being its corresponding class label. To train a SVM classifier, we need to maximize the margin by solving the following optimization problem:

$$\min_{\mathbf{w}, b, \xi} \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^P \xi_i, \quad \text{subject to} \quad y_i(\mathbf{w}^T \phi(\mathbf{x}_i) + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad i = 1, \dots, P, \quad (20)$$

where $\phi(\mathbf{x}_i)$ maps the training vector \mathbf{x}_i into a higher-dimensional feature space such that the classification problem becomes simpler in that feature space and $C > 0$ is a regularization constant to control the trade-off between the large margin and misclassification error. Due to the possible high dimensionality of the vector \mathbf{w} , the following dual problem is solved:

$$\max_{\alpha} \sum_{i=1}^P \alpha_i - \frac{1}{2} \sum_{i,j=1}^P \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \quad \text{subject to} \quad 0 \leq \alpha_i \leq C, \quad \sum_{i=1}^P \alpha_i y_i = 0, \quad (21)$$

where $K(\mathbf{x}_i, \mathbf{x}_j) \equiv \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j)$ is the kernel function. After Eq. (21) is solved, the vector \mathbf{w} in the primal problem is related to the variable α_i , $i \in [1, P]$ in the dual problem by

$$\mathbf{w} = \sum_{i=1}^P y_i \alpha_i \phi(\mathbf{x}_i). \quad (22)$$

Once the optimal α_i has been calculated, the decision function is given by

$$y_p(\mathbf{x}) = \text{sgn}\left(\sum_{i=1}^P y_i \alpha_i K(\mathbf{x}_i, \mathbf{x}) + b\right) = \text{sgn}(\mathbf{w}^T \phi(\mathbf{x}) + b). \quad (23)$$

The bias parameter b is computed as

$$b = \frac{1}{N_{sv}} \sum_{i=1}^{N_{sv}} \mathbf{w}^T \mathbf{x}_i - y_i, \quad (24)$$

where N_{sv} is the number of support vectors. The kernel K is a symmetric, positive definite function. Here, we use a linear kernel given by $K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j$.

Since SVM was designed for solving two-class classification problems, there are two approaches that can be used to solve multi-class problems: one-versus-one and one-versus-all. In this paper, we employ one-versus-one approach. For k classes, the one-versus-one approach constructs $k(k-1)/2$ two-class SVM classifiers; each classifier is trained on data from two classes. To combine all two-class SVM classifiers, a voting strategy is used. The output of each binary classifier is considered as a vote for one of the two classes in consideration. After all binary classifiers have voted, the class with the most votes is designated as the final predicted output class. We employ the LIBLINEAR SVM toolbox developed by Fan and his colleagues⁹ to implement the SVM classifier. The optimal parameter C for the linear kernel SVM is obtained by designing several SVM classifiers with different C values; the SVM classifier with the highest classification performance, using a cross-validation procedure, is chosen to obtain the optimum C value. In the next section, the proposed image-based classification method is evaluated on real Doppler radar signals.

3. EXPERIMENTAL RESULTS

The proposed image-based classification approach is tested on real radar signals. To collect Doppler signals, we first describe the experimental setup. Then, we investigate the optimum signal length associated with the time-frequency patch. The proposed approach is compared with other existing approaches on the classification of three types of human micro-Doppler radar signature.

3.1 Experimental Setup

A 24-GHz continuous-wave Doppler radar was used to collect radar signals for evaluating the proposed micro-Doppler signature classification approach. The database comprises radar signals from 18 subjects, including 7 females and 11 males. Each subject performed three types of motion: walking while holding a heavy object with both hands, walking while carrying an object in one hand, and walking with both arms swinging. Each subject repeats these arm motions three times while walking towards the radar. The radar signal was recorded at a sampling rate of 7,812 Hz in an outdoor environment for 10 seconds. Figure 5 presents images of a subject walking with different arm motions. The recorded radar signal from each motion is converted into its time-frequency representation. Figure 6 depicts the three spectrograms of signals obtained in the corresponding experiments of Fig. 5.



Figure 5. Images showing three different types of human motions: (a) walking while holding a heavy object with both hands, (b) walking while carrying an object in one hand, and (c) walking with both arms swinging.

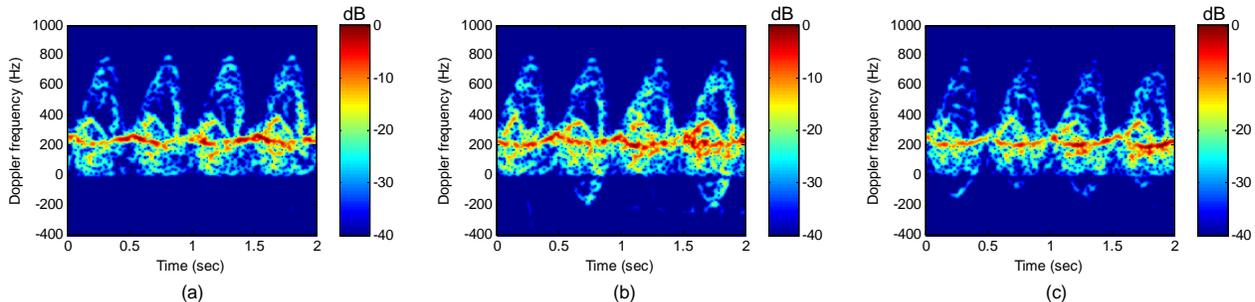


Figure 6. Spectrograms of three different types of human walking with: (a) no arms swinging, (b) one arm swinging, and (c) both arms swinging.

To generate the spectrogram, a Hamming window is selected as the window of the STFT. The extracted time-frequency patches are resized to 128×128 before the feature extraction. The micro-Doppler descriptor consists of a set of 36 log-Gabor filters, i.e., nine orientations ($N_\theta = 9$) and four center frequencies, ($N_f = 4$), for feature extraction. Each filtered image is partitioned into 64 non-overlapping sub-regions, i.e., $d = 8$. Consequently, the feature map generated by the micro-Doppler descriptor has a size of 64×144 .

3.2 Analysis of the Length of Input Signal

Most micro-Doppler signature classification approaches use a long radar signal. In our approach, time-frequency patches are extracted from the spectrogram for classification. To determine the signal length associated with the time-frequency patch, we investigate different lengths of signal, starting from 0.5 s to 2.5 s at a step size of 0.5 s. Figure 7 illustrates the classification rate of the proposed method as a function of the input signal length. At a signal length of 0.5 s, which covers one leg swing, the proposed method achieves a classification rate of 77.1%. With a signal length of 1.5 s, the classification rate increases to 92.5%. Further increasing the signal length slightly decreases the classification rate. This is because the time-frequency patch is severely downsampled to have a predefined size of 128×128 .

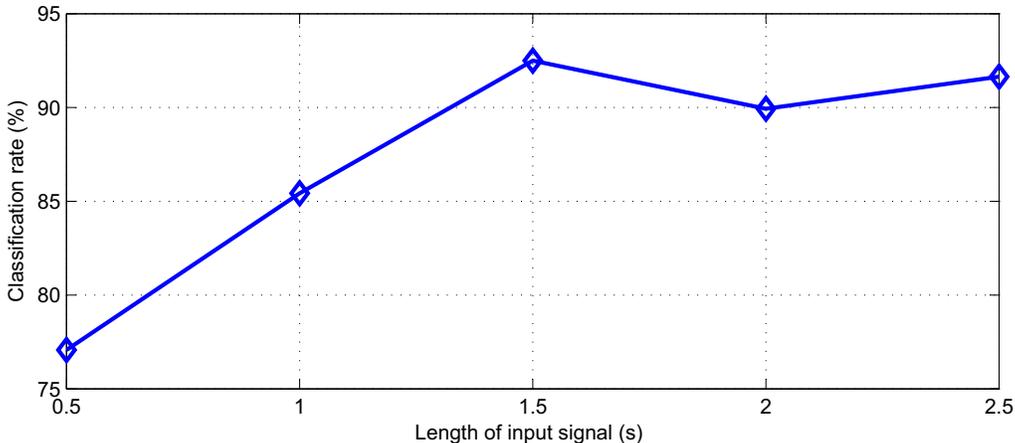


Figure 7. Classification rate as a function of the length of input signal.

3.3 Analysis of the Classification Performance

To examine the classification rate of each micro-Doppler radar signature, the confusion matrix is shown in Table 1. In the table, the off-diagonal entries list the misclassification rates, whereas the main diagonal entries give the correct classification rates of the micro-Doppler radar signatures. For example, the percentage in the second row and third column, i.e. (2, 3), indicates that the proposed method misclassifies one-arm swing as two-arm swing with an error rate of 6.5%. At the entry (3, 2), it shows there is 7.3% of the radar signals belonging to two-arm swing incorrectly classified as one-arm swing. It seems that the proposed method has difficulty distinguishing between one-arm swing and two-arm swing. One way to improve the classification results is to use a large time-frequency patch covering several leg swings.

Table 1. Confusion matrix listing the classification rates of three human motions.

	Predicted no-arm swing	Predicted one-arm swing	Predicted two-arm swing
Labeled no-arm swing	92.7%	5.2%	2.1%
Labeled one-arm swing	6.9%	86.6%	6.5%
Labeled two-arm swing	1.6%	7.3%	91.1%

Next, we investigate the robustness of the proposed classification method against noise. Therefore, the radar signals are corrupted by white Gaussian noise of different levels. Figure 8 illustrates the classification rate as a function of the input signal SNR. At SNR = 40 dB, the proposed method achieves a classification rate of 90%. Its classification rate remains unchanged until the SNR of the input signal drops below 20 dB. At low SNR = 5 dB, the classification rate is 50%.

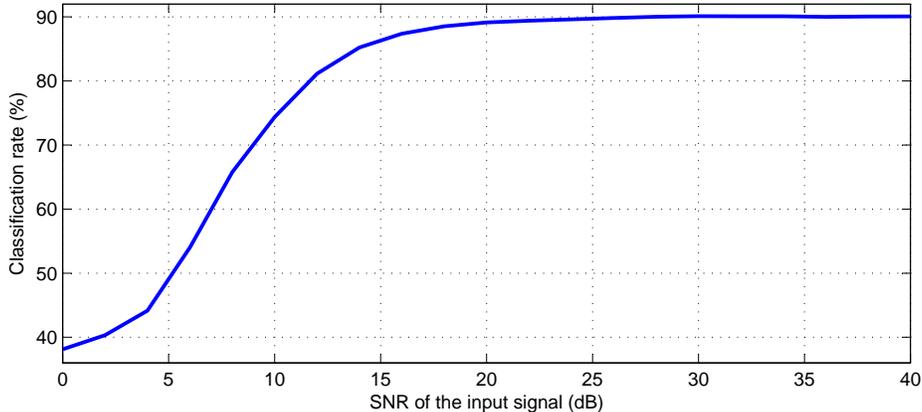


Figure 8. Classification rate of the proposed method as a function of SNR of the input signal.

3.4 Comparison of Classification Methods

For comparison of the different classification methods, a cross-validation procedure is adopted. As the database contains radar signals from 18 subjects, the entire dataset is divided into 6 equal subsets so that radar signals used to generate the time-frequency patches for training and testing are from different subjects. Five subsets are used for training and the last subset is kept for testing. This process is repeated six times and the classification rate is computed as the total number of correctly classified samples divided by the total number of test samples. The length of the signal for classification is set to 1.5 s, which is the same signal length as that of the time-frequency patch. The proposed approach is compared with four other methods for classification of micro-Doppler radar signatures: MFCC, Gabor wavelet filter, Cadence frequency, and EMD. In the first method, MFCC features are extracted from a signal frame using a set of 40 triangular bandpass filters to produce 64 mel-scale cepstral coefficients. The analysis window size and the overlapping size are set to 0.98 s and 0.49 s, respectively. Then, $(2D)^2$ -PCA method is applied for dimensionality reduction. Lei and Lu used Gabor wavelet filters for extracting micro-Doppler features from the spectrogram.¹¹ In our comparison, we use the same number of Gabor wavelet filters as that of the proposed method, i.e., 36 filters, and downsample each filtered image to a size of 16×16 . Arranging the extracted features into a matrix, we obtain a feature map of size 256×36 , which is then compressed using $(2D)^2$ -PCA. Another method developed by Bjorklund and his colleagues¹⁰ is also tested. Instead of spectrogram, the cadence frequency diagram, which is obtained by applying Fourier transform along the horizontal direction of the spectrogram, is used for feature extraction. At the second and third harmonic frequencies, the frequency profiles, i.e., the columns of the cadence frequency diagram are extracted as features. The resulting feature vector comprises the two harmonic frequencies and their corresponding frequency profiles. For dimensionality reduction, normal PCA is used. When PCA is applied for dimensionality reduction, the threshold β is set to 0.95. Fairchild and Narayanan employed EMD for feature extraction.⁵ In our experiment, each radar signal is decomposed into a set of 11 IMFs using EMD. Then, the energies of the IMFs are used as features. In the classification stage of these methods, we use SVMs with a linear kernel as a classifier.

Table 2 presents the classification performances and the feature dimensions produced by the various approaches. The proposed method achieves a classification rate of 90.1%. With Gabor wavelet filters, we obtain a classification rate of 71.5%. The method that used MFCC features to characterize human micro-Doppler signals achieves a classification rate of 70.8%. The cadence frequency based method gives a classification rate of 64.4%. Among the four approaches, the EMD-based method achieves the lowest classification rate. One possible reason why EMD performs poorly in this experiment is that the difference among the three types of human motion is simply the arm swing, in contrast to the experiments described in reference⁵ where the dataset consists of radar signals from different human activities.

Table 2. Classification rates of several micro-Doppler signature classification approaches.

Method	Feature dimension	Classification rate \pm standard deviation (%)
Proposed micro-Doppler descriptor	2,323	90.1 ± 3.1
Gabor wavelet filter	4,067	71.5 ± 4.0
MFCC	42	70.8 ± 0.9
Cadence frequency based method ¹⁰	799	59.8 ± 5.6
EMD	11	37.6 ± 2.3

3.5 Classification of Long Signal Frame

In practice, it is common to record a long micro-Doppler signal for analysis. Therefore, to classify a signal frame of arbitrary length, overlapping local time-frequency patches centered along the torso frequency are extracted and processed by the proposed classification method. The amount of overlap between two consecutive windows, in time, is varied from 50% to 90%. The classification scores of the different patches are aggregated and a majority voting is used to predict the class label of the signal frame. Different signal lengths are investigated, ranging from 3 s to 10 s. Figure 9 depicts the classification rate of the proposed method as a function of the signal length. Increasing the amount of overlap between two time-frequency patches improves the classification rate. Using the entire 10-second radar signal as input, we obtain a classification rate of 99.2%.

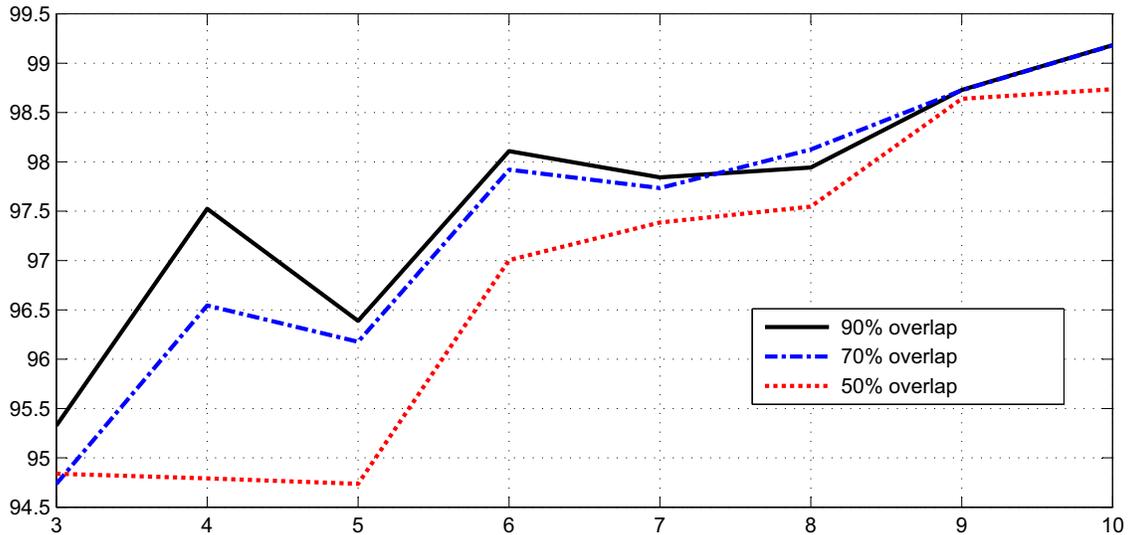


Figure 9. Classification rate of the proposed method as a function of the length of the input signal. The amount of overlap between consecutive temporal windows is varied from 50% to 90%.

4. CONCLUSION

In this paper, an image-based classification approach was proposed for classifying micro-Doppler radar signatures of human motions. The input time-varying radar signal is converted into a spectrogram, where local time-frequency patches centered along the torso frequency are extracted. A micro-Doppler descriptor was developed to convert the extracted patches as feature maps, which are subsequently compressed using the matrix-based PCA technique. The proposed approach can be employed to classify signal frames of arbitrary length. Experimental results showed that the proposed method achieved higher classification performance than some of the existing classification approaches.

REFERENCES

- [1] Y. Kim and H. Ling, "Human activity classification based on micro-Doppler signatures using a support vector machine," *IEEE Transactions on Geoscience and Remote Sensing*, **47**(5), 1328–1337 (2009).
- [2] I. Orović, S. Stanković, and M. Amin, "A new approach for classification of human gait based on time-frequency feature representations," *Signal Processing*, **47**(91)6, 1448–1456 (2011).
- [3] B. G. Mobasser and M. Amin, "A time-frequency classifier for human gait recognition," *Proc. SPIE 7306, Optics and Photonics in Global Homeland Security V and Biometric Technology for Human Identification VI*, 730628 (May 05, 2009); doi:10.1117/12.819060.
- [4] D. Yessad, A. Amrouche, M. Debyeche, and M. Djeddou, "Micro-Doppler classification for ground surveillance radar using speech recognition tools," vol. 7042 of *Lecture Notes in Computer Science*, 280–287, Springer Berlin Heidelberg (2011).
- [5] D. P. Fairchild and R. M. Narayanan, "Classification and modeling of human activities using empirical mode decomposition with S-band and millimeter-wave micro-doppler radars," *Proc. SPIE 8361, Radar Sensor Technology XVI*, (May 1, 2012), doi: 10.1117/12.922448.
- [6] A. Oliva and A. Torralba, "Building the gist of a scene: the role of global image features in recognition," *Progress in Brain Research: Visual perception* **155**, 23–36 (2006).
- [7] D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *Journal of the Optical Society of America A*, **4**(12), 2379–2394 (1987).
- [8] D. Zhang and Z.-H. Zhou, "Two-directional two-dimensional PCA for efficient face representation and recognition," *Neurocomputing* **69**(1-3), 224–231 (2005).
- [9] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin, "LIBLINEAR: A library for large linear classification," *Journal of Machine Learning Research*, **9**, 1871–1874 (2008).
- [10] S. Bjorklund, T. Johansson, and H. Petersson, "Evaluation of a micro-Doppler classification method on mm-wave data," *Proc. IEEE Radar Conference*, 0934–0939 (2012).
- [11] J. Lei and C. Lu, "Target classification based on micro-doppler signatures," *Proc. IEEE Radar Conference*, 179–183 (2005).