A Kinect-Based Human Micro-Doppler Simulator

Barış Erol
Cesur Karabacak
TOBB University of Economy and Technology
Ankara, Turkey

Sevgi Zübeyde Gürbüz
TOBB Univ. of Economy and Technology
and TUBITAK Space Technologies Research Institute
Ankara, Turkey

INTRODUCTION

Until recently, human surveillance has primarily been accomplished using video cameras. However, radar offers unique advantages over optical sensors, such as being able to operate at far distances, under adverse weather conditions, and at nighttime, when optical devices are unable to acquire meaningful data. Radar is capable of recognizing human activities by classifying the micro-Doppler signature of a subject. Micro-Doppler is caused by any rotating or vibrating parts of a target, and results in frequency modulations centered about the main Doppler shift caused by the translational motion of the target [1]. Thus, the rotation of a helicopter blade, wheels of a vehicle, or treads of a tank all result in micro-Doppler. In the case of humans, the complex motion of the limbs that occur in the course of any activity all result in a micro-Doppler signature visually distinguishable from other targets, even animals [2]–[3], which can then be exploited for human detection [4]–[5], automatic target recognition (ATR) [6]–[7], and activity classification [8].

The development and testing of classification algorithms, however, requires a vast amount of data that reflects the wide variety of target characteristics, such as type, size, build, and gait, as well as environmental factors, such as signal-to-noise ratio (SNR), and physical parameters, such as aspect angle, range, and velocity. Acquisition of such data experimentally is oftentimes expensive and impractical, especially for scenarios difficult to reproduce in a laboratory environment. Thus, the ability to generate accurate, simulated data that closely matches real-world measurements is a critical necessity to develop novel signal processing algorithms. To date, two main approaches have been taken in simulating human micro-Doppler: kinematic modeling and video motion capture (MOCAP). The simplest kinematic model [9] that has been proposed in the literature considers just the oscillation of the torso, which is represented by a sinusoidal function. Slightly expanding upon the torso model, the lower-body model [10] is comprised of three points, the torso and two legs, where the legs are animated similar to the swing of a pendulum. Although these models visually present some similarity to measure micro-Doppler signatures, much of the nuances in the real signature are entirely missed, rendering them unsuitable for most classification studies. Perhaps the most comprehensive and widely used model, however, is the Boulic model [11]. The Boulic model is based on the results of a detailed experimental study of gait analysis, and uses a combination of equations and charts to animate 17 different points and joints on the human body. The main disadvantage of the Boulic model, however, is that it is only valid for walking. Although models have also been put forward for running [12], there are no other full-body models for other types of human motion. Moreover, even if models covering a wider range of activities did exist, such models are inadequate in capturing the variable nature of human motion, which rarely follows a single kinematic pattern.

MOCAP data, on the other hand, is capable of representing almost any sequence of motions one can imagine. For example, the Carnegie Mellon University Graphics Laboratory [13] has collected MOCAP data for walking, running, crawling, jumping, boxing, dancing, and a variety of sports by placing infrared sensors at key joints on the body and recording their time-varying positions. MOCAP data has been exploited in several studies to simulate human micro-Doppler signatures. U.S Air Force Research Laboratory [14]–[15] used the Poser™ system to animate human figures, whose electromagnetic properties were modeled with the aide of the Xpatch software. S. S. Ram et al. [16] used data from a MOCAP database to simulate not just human micro-Doppler signatures, but also that of animals, such as dogs and horses. Although several freely available MOCAP databases exist, such as [13], [17], these databases contain a limited number of data sets with no control over the data collection duration, or the characteristics, speed, duration, and trajectory of the test subject.

Generating high-quality MOCAP data at will requires possession of a professional MOCAP system, such as OpenStage2 from Organic Motion [18], or MVN/BIOMECH from Xsens [19]. Unfortunately, such MOCAP systems are quite expensive (~ $40,000 – $60,000). In contrast, the Kinect™ sensor is priced at approximately $100. Although originally designed to facilitate the interaction between users and a computer game environment [20], the Kinect sensor has been exploited for a wide range of

Authors’ address: TOBB University of Economy and Technology, Dept. of Electrical and Electronics Engineering, Sogutozu Cad. No 43, Sogutozu, Ankara, 06560 Turkey. E-mail: (szgurbuz@etu.edu.tr; sevgi.gurbuz@tubitak.gov.tr). This work was supported in part by the Scientific and Technological Research Council of Turkey (TUBITAK) Project 113E105 and the EU FP7 Project PIRG-GA-2012-268276. Manuscript received July 5, 2014, revised February 6, 2015, and ready for publication March 3, 2015. DOI No. 10.1109/MAES.2015.140121. Review handled by E. Blasch. 0885/8985/15/$26.00 © 2015 IEEE
unique applications, such as robotics and the design of a virtual fitting room.

In this work, a low-cost, Kinect-based radar simulator is developed for generating synthetic micro-Doppler signatures. The Kinect sensor is used as a markerless system for capturing the time-varying coordinate information of human joints. This position information used is then used to compute the expected return for a pulsed Doppler radar, from which the micro-Doppler signature is formed. The accuracy of the Kinect sensor’s depth measurement is not as accurate as specially designed MOCAP systems. Nevertheless, it is shown that spectrograms generated from Kinect data are comparable to that generated from kinematic models and high quality MOCAP data, and thus exploitable for studies of human micro-Doppler. A database of human micro-Doppler signatures is compiled for 16 people engaged in five different activities: walking, running, jumping, boxing, and random motion.

**THE KINECT SENSOR**

The Kinect sensor was originally released in November 2010 for the purpose of enabling interactive gaming, but has also become popular in the area of hands-free control of electronic devices through gesture recognition [21]. Kinect actually comprises two sensors: a red-green-blue (RGB) video camera as well as an infrared (IR) camera. The IR camera operates in conjunction with an IR projector, which casts a grid of IR dots (structured light) over objects residing within the projector’s line-of-sight. Although the human eye is unable to see these dots, the complimentary metal oxide semiconductor (CMOS) camera is outfitted with an IR filter, thereby enabling it to sense IR returns. Kinect processes the way the pixels deform upon striking a surface using classic computer vision algorithms (depth from focus and depth from stereo) to calculate the depth and surface information of objects in the scene. The actual depth values are distances from the camera-laser plane rather than distances from the sensor itself. Hence, the depth sensor simply can be considered as a device that provides the coordinates of 3D objects. The depth accuracy of Kinect has been experimentally shown to be +/- 4 cm [22].

Kinect simultaneously captures depth and color images at a frame rate that depends upon both image resolution and the control software used. For example, MATLAB can capture data at a rate of up to 20 frames per second (fps), while Processing™, supplied by Open National Instruments (NI), is limited to 29 fps maximum. In this work, MATLAB was used to capture images at a resolution of 460 x 840 pixels at a frame rate of 18 Hz. The Kinect’s color camera has a relatively low resolution 640 by 480 pixels with a Bayer color filter, but the hardware is capable of generating resolutions up to 1280 x 1024 at a lower frame rate and other color formats such as UYVY. Technical details of the Kinect hardware are given in Table 1.

Although the Kinect sensor has many advantages, the primary being the capability of measuring the depth objects at distances of 0.4 m – 7 m fairly accurately in real time at low cost, it also possesses several limitations. Naturally, the Kinect sensor is sensitive to external sources of IR energy, which can lead to increased

<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angular field-of-view</td>
<td>57° horizontal, 43° vertical</td>
</tr>
<tr>
<td>Frame rate</td>
<td>≈ 30Hz</td>
</tr>
<tr>
<td>Nominal spatial range</td>
<td>640 x 480 (VGA)</td>
</tr>
<tr>
<td>Nominal spatial resolution (at 2m distance)</td>
<td>3 mm</td>
</tr>
<tr>
<td>Nominal depth range</td>
<td>0.8 m – 3.5 m</td>
</tr>
<tr>
<td>Nominal depth resolution (at 2m distance)</td>
<td>1 cm</td>
</tr>
<tr>
<td>Device connection type</td>
<td>USB (+external power)</td>
</tr>
</tbody>
</table>
A Kinect-Based Human Micro-Doppler Simulator

measurement errors. Moreover, Kinect cannot detect crystalline or highly reflective objects, and can only generate an IR depth map for objects located at least 0.7 m away. Furthermore, Kinect is susceptible to the effects of shadowing, i.e., the failure of IR energy to reach areas blocked by objects (Figure 1). As a result, no depth information can be obtained on objects located in shadowed regions [23].

The shadowing effect is an important limitation in applying Kinect to human micro-Doppler simulation because it is possible for the motion of parts of the body to shield, or hide, the motion of other parts (self-shadowing) that normally would affect the micro-Doppler signature. Indeed, experimental trials showed that this was the case for movements such as crawling. However, through the use of multiple Kinect sensors such self-shadowing limitations may be overcome.

In the proposed radar micro-doppler simulator, the 3D position measurements obtained from the Kinect sensor are used in lieu of the time-varying range measurements typically obtained from radar. First, the Kinect data for a given motion recording are stored in memory. Then, in postprocessing, the Kinect-derived range measurements are then used in a mathematical model for the radar return to simulate the response expected from a human in motion to the transmitted radar signal, as detailed in the section below.

**COMPUTING THE HUMAN RADAR RETURN**

In general, the received radar signal is a time-delayed version of the transmitted signal, while the roundtrip time-delay \( t_d \) is dependent upon the target range \( R \), as \( t_d = 2R/c \). Thus, the return from a point target for a linear frequency modulated (LFM) pulse-Doppler radar may be represented as

\[
s(n,t) = a_se^{j\frac{i-t}{\tau}} e^{j2\pi f_{c}t + \gamma (i-t)^{2}},
\]

where the time \( t \) is defined as \( t = T(n-1) + \hat{t} \) in terms of the pulse repetition interval (PRI) \( T \), the pulse number \( n \), and the time relative to the start of each PRI \( \hat{t} \); \( a_s \) is the amplitude as given by the radar range equation; \( r \) is the pulse width; \( c \) is the speed of light; \( \gamma \) is the chirp slope; and \( f_c \) is the transmitted center frequency.

In addition to the target return, the radar also receives returns from illuminated nontarget surfaces, known as clutter. The effects of clutter as well as other sources of noise and interference may be included in the simulator by simply adding the relevant mathematical models to the expression in (1). As the focus is human micro-Doppler simulation, such effects are not included in this work.

Humans represent a much more complex target as compared with point targets. However, it has been shown in work by Geisheimer [24] and Van Dorp [25] that representing the total human radar return as the superposition of returns from a finite number of point targets located on the body results in spectrograms that closely match experimental measurements. Thus, the total return for a subdivision into \( K \) parts is

\[
s_k(n,t) = \sum_{i=1}^{K} a_{ki} e^{j\frac{i-t_{d,i}}{\tau}} e^{j2\pi f_{c}t + \gamma (i-t_{d,i})^{2}},
\]

where \( a_{ki} \) and \( t_{d,i} \) are the amplitude and time delay of the return of each part. The amplitude \( a_{ki} \) is computed from the range equation as

\[
a_{ki} = \frac{G \lambda^{2} \sqrt{P_{t} \sigma_{i}}}{(4\pi)^{3} R_{i}^{3} L_{s}^{3} L_{d}},
\]

and includes several factors that vary with target range \( R \) and geometry. For instance, the antenna gain \( G \) varies according to angle of incidence, and the atmospheric losses \( L_{s} \) vary with range. Because micro-Doppler signatures are derived from target phase history and the impact of amplitude variations are minor, in this work it is assumed that these parameters are constant, along with the transmitted signal power \( P_{t} \), the wavelength \( \lambda \), and the system loss \( L_{d} \). The radar cross section (RCS) \( \sigma_{i} \) is modeled according to the approximate shape of the body parts. Thus, the RCS of the head is computed from the scattering amplitude and phase of an ellipsoid.

Evaluation of (2) requires knowledge of the time-varying position of each point designated on the human body. In this work, the Kinect sensor is used to measure the required time-varying positions by developing a MATLAB-based skeleton tracking program, which provides the 3D coordinates of points defined upon the skeleton, updated for each frame captured. For a frame rate of 18 Hz, Kinect provides temporal sampling at an interval of 0.055 s. For a person walking at 2 m/s, this would give only 18 samples over an interval of 2 m, which is much too sparse to derive an accurate spectrogram. Thus, the range data provided by Kinect is interpolated to achieve a temporal sampling rate of 2400 Hz prior to any further processing.

Next, the time-varying ranges for each point target on the body are computed according to the position of the radar relative...
to the human target and the target direction of motion. Since the Kinect measurements are provided relative to the hip center, for any desired target path, the relative distance between radar and body point target can be computed in the 3D coordinate system. Thus, the effects of elevation and orientation on the micro-Doppler signature are included in the simulation.

Once the required ranges are estimated from the Kinect data, (2) can be computed for any given human activity and any desired radar system parameters, such as center frequency, bandwidth, chirp rate, and PRI. Although there are many different possible time-frequency representations of micro-Doppler, in this work we use the most ubiquitous representation: spectrograms. The spectrogram of the human radar return can be computed by first applying pulse compression to the data so that a peak is observed at the target location. Considering just the data at the target location, the pulse compressed radar return may be written as

\[
x[n] = \sum_{i=1}^{K} a_i R_{d,i} e^{-j\frac{2\pi}{\lambda} f_{d,i} t}
\]

where \(R_{d,i}\) is the range from the antenna to the center of each body part. The spectrogram is then computed by just taking the short-time Fourier transform of (4).

**Skeleton Tracking with Kinect**

Human pose estimation and skeleton tracking is an important problem with many applications in areas such as human computer interaction (HCI) and computer vision. It is also one of the basic building blocks of markerless MOCAP technology. However, skeleton tracking remains a very difficult and very important problem. There are challenges due to the large parameter space and constraints involved. Other challenges involve the variation of the IR camera’s field of vision, positioning of the Kinect sensor to collect accurate data, external constraints (environmental obstacles), variability in lighting condition and human physique, and the loss of 3D information when extracting depth from observations of the human skeletal structure in a 2D planar image projection.

The skeleton tracking algorithm implemented in this work begins when the user enters the camera’s field of vision. As the subject stands still at a given initial pose, the algorithm detects the pose and superimposes a skeleton comprising 20 points upon the detected human figure. Figure 2 illustrates the skeletal structure for the pose, i.e., position of a person standing, legs slightly apart and hands raised upward bent at the elbows, as well as joint definitions specified in the software.

**Algorithms Implemented by Commercial Software**

Two different software programs for accomplishing skeleton tracking with Kinect were experimented with in the course of this work. The first is a package called Processing, which can be used to record the locations of each joint. Although skeleton tracking generally seems to work fine, this package fails to track the motion of the subject’s feet. Without the range information of the feet, however, the micro-Doppler signatures derived failed to exhibit the Doppler spread typically expected as if the leg micro-Doppler was totally absent. Moreover, the periodicities typically seen in spectrograms of a walking person were also not observed.

Thus, an alternative software package was explored: the Image Acquisition Toolbox Support Package for Kinect, a specific package for MATLAB software that allows users to import Kinect data into MATLAB, access the acquired depth map, and perform functions such as gesture recognition, face, and skeletal tracking. The MATLAB functions for pose detection exploit a method recently proposed [26] for estimating 3D positions of the body joints in a depth image using an object recognition approach, designing an intermediate body part representation to map the difficult pose estimation problem into a simpler per-pixel classification problem. Body part information is inferred from the Kinect depth map using a randomized decision forest, learned from over 1 million training examples that are derived from a vast database of MOCAP data supplied within the Microsoft software accompanying Kinect.

In MATLAB, unlike Processing, users can define the structure of the skeleton to be tracked, thus enabling inclusion of the feet in the skeletal model. The skeleton defined for this work is shown in Figure 2 and comprises twenty different points on the human body. However, when conducting the radar simulations, the time-varying positions of just 17 of the measured points are included to be consistent with the skeletal structure utilized in the widely used Boulic kinematic model. The right wrist, left wrist, and spine point targets are not used in the simulation. Thus, the value of \(K=17\) is used in (2) and (4).

**Software Modifications**

Improvements to MATLAB’s skeleton tracking procedures were made to enable more accurate extraction of information important to generate accurate spectrograms. First, a user interface was designed to control the Kinect during data collection, as illustrated in Figure 3. In this interface, the user can either preset the duration of data collection, or can simply start and stop the processes with a click of the mouse. After termination of the program, the measured duration and exact sampling frequency are reported by the software. Duration and sampling frequency are both important parameters affecting the accuracy of simulated micro-Doppler signatures.
A key modification was also devised to correct for the instability and measurement errors involving tracking of the feet. As previously mentioned, self-shadowing is an important problem that has detrimental effects on the tracking accuracy required for micro-Doppler simulations. When a body part is partially or fully blocked by an object other than that being tracked, position cannot be accurately estimated.

This is due to several reasons. First, the obstruction results in a high-frequency vibration of the tracked position. Secondly, obstruction sometimes results in a body part being incorrectly identified, resulting in instabilities that could result in a loss of tracking information in the form of disconnects and hopping between correct and incorrect positions.

In particular, such vibrations and instability were observed in distance measurements of the left and right foot. When stable 3D measurements of these positions are not made, explicit deviations that are easily noticed from shaky criss-crossing in the skeleton can be seen. As correct tracking of the feet was seen to be critical in accurately capturing the true Doppler spread and periodicities of the micro-Doppler signature, an algorithm for compensating and correcting for highly deviant foot measurements was devised.

The logic of the algorithm is based on the observation that unstable tracks primarily reveal themselves as outliers in the data. Thus, the goal of the enhanced foot tracking algorithm was to identify such outliers and apply a correction on the range measurement.

Outliers were identified by examining histograms of the $x$ and $y$ coordinates of the foot position. Data points observed to deviate significantly from the mean, i.e., the difference between coordinate and the mean lie above an adaptive threshold, are identified as outliers. Threshold is determined by examining the data histogram. The coordinates of the outliers are corrected by forming an estimate of where the foot should be based on past measurements. This is accomplished by first determining whether the foot is moving forwards and backwards, and accordingly updating the current location with the previous location plus the average distance moved each frame.

![Figure 4.](image1.png)

Corrected foot coordinates. (a) Deviations along $x$-axis. (b) Deviations along $y$-axis.

The angle $\theta$ can be computed from Kinect measurements as follows. First define the displacement vectors $v_1$ and $v_2$ as the difference between the knee position $x_1$ and the ankle position $x_0$, and the difference between the foot position $x_2$ and ankle position $x_0$, respectively:

$$v_1 = x_1 - x_0,$$

$$v_2 = x_2 - x_0.$$
\[ v_2 = x_2 - x_0. \]  

(6)

The acute angle between the two vectors can then be calculated by their dot product

\[ \theta = \frac{(v_1 \cdot v_2)}{|v_1||v_2|}. \]  

(7)

which is plotted for example data in Figure 6. Note that in generally \( \theta \) does seem to be continuous as well as periodic in time but that there are certain points that deviate from the mean amplitude, another indication of outliers that may need compensation.

An idea of how well skeletal tracking is performed was also obtained by computing the length of the feet based upon the ankle and foot positions computed by Kinect. Computing the foot length at each frame, the average foot length measured of the course of the data collect was found. Prior to collecting Kinect data for a given test subject, the subject’s actual foot length was measured and compared with the experimentally obtained average foot length. In this way, it was determined whether extremely erroneous foot tracking measurements were obtained. Corrections to the foot position were made for outliers by adjusting the foot length to match the true foot length of the test subject.

**EXPERIMENTAL SET-UP**

The Kinect MOCAP data required for simulating human micro-Doppler signatures were collected with the aid of a treadmill located at the TOBB University of Economics and Technology (ETU) Remote Sensing Laboratory Ankara, Turkey. As capturing the entire human body is required for skeleton tracking, test subjects were required to remain within a region of 135 cm to 312 cm from the Kinect sensor. Since this is a quite restrictive region, data collection over extended dwell times was accomplished with the aid of a treadmill, so that the subject always remained within the field of view of the sensor, as shown in Figure 7. In this configuration, the Kinect sensor was placed upon a tripod of height 92 cm. In the case of activities that were done while standing, such as boxing, or short distances, such as leaping, data collection was done without the treadmill. For this configuration, the Kinect sensor was place upon a table 80 cm in height. Operating within these constraints, skeleton tracking was accomplished based on the Kinect data and used to generate a computer animation in MATLAB, as shown in Figure 8. As described in the previous section, it is the range data of this animation that is used to compute the distances required by (4).

**SIMULATED HUMAN MICRO-DOPPLER SIGNATURES**

Micro-Doppler signatures for any desired frequency, bandwidth, antenna pattern, transmitter power, and scenario may be simulated using (4). As an example, simulated spectrograms are generated for a pulse-Doppler radar with a center frequency of 15 GHz and bandwidth of 150 MHz. Measurements were made for five different human activities: walking, running, leaping, boxing, and random motion (arbitrary arm and leg motion while on treadmill).
Each activity was recorded five times for a duration of 3 to 21 s, depending on activity, for 16 different subjects, yielding a total of 96 data collects per activity. The test subjects posed heights ranging from 1.61 to 1.87 m, weights ranging from 55 to 135 kg, and included 1 female and 11 males.

**VISUAL OBSERVATIONS**

First, the Kinect-based spectrograms for walking and running are compared with spectrograms generated from the Carnegie Mellon University (CMU) Motion Capture Library [27]–[29], as shown in Figure 9, while the boxing and leaping spectrograms are compared in Figure 10. The walking and running spectrograms share a number of important critical features. In both CMU- and Kinect-derived spectrograms the torso response represents the strongest return and exhibits a sinusoidal oscillation. The periodic motion of the legs causes the greatest amplitude and highest frequency oscillation, followed by that of the...
arms, which appear to be more distinct in the Kinect-derived spectrograms. Moreover, the leg oscillation frequency and amplitude for running is greater than that of walking for both data sources, as would be expected. Thus, visually the spectrograms appear to be similar.

There are however several important differences. Notice that the Kinect spectrograms are symmetric in Doppler, while the CMU spectrograms have minimal negative Doppler. This difference is due to the rotation of the treadmill, which causes the legs to move away from the radar during part of the walking cycle, an effect not normally observed when the subject walks on a fixed surface.

The Kinect-based spectrograms for boxing and leaping also compare well with the CMU MOCAP-based spectrograms. The chief difference between the boxing spectrograms, namely, the torso response of the CMU MOCAP-based spectrograms being stronger and having a larger Doppler spread, is due to the differences in the enactment of the boxing motion itself. The Kinect data recordings were made for the subjects planting their feet on the ground and exaggeratedly boxing with their arms only, while video records of the CMU data reveal that the CMU test subjects enacted more limited arm motions while slightly hopping and moving about on their feet, as would a boxer in the ring. These differences are directly observable in the difference between the spectrograms, and is another indication of the Kinect system being able to capture differences in human motion.

Spectrograms for the random motion enacted upon the treadmill are shown in Figure 11. Note that the random motion signature is a mixed motion that bears much similarity with both walking and running, so that we would expect correct classification between walking, running, and random motion to be more difficult than discriminating between the other activities.

Figure 10.
Comparison of Kinect-based and CMU-based spectrograms for boxing and leaping. (a) Kinect-based spectrogram for boxing. (b) Kinect-based spectrogram for leaping. (c) CMU MOCAP-based spectrogram for boxing. (d) CMU MOCAP-based spectrogram for leaping.
A quantitative comparison of CMU- and Kinect-derived spectrograms was made by computing the statistics of seven features extracted from the simulated spectrograms, and are given in Table 2. The upper and lower envelopes are defined as a curve connecting the maxima and minima of the spectrogram, respectively. The mean and variance of features extracted to classify the signatures were seen to be comparable, with the exception of the mean torso velocity and minimum of lower envelope, whose variance in Kinect spectrograms was found to be significantly lower than in CMU-based spectrograms. This difference in average velocity variance is primarily due to the treadmill. Because the treadmill rotates at a fixed speed, the average velocity of the subject is more constrained than what would normally be possible in freestyle walking on a fixed surface. More specifically, averaging over 96 Kinect data collects of 16 different people and 212 CMU walking recordings of 14 different people for a duration of 1.5 s, the variance of the average torso velocity for CMU spectrograms was found to be approximately 100 times greater than that found for Kinect-based spectrograms. In the case of running, the statistics were computed for 80 Kinect records of 16 people and 71 CMU records of 14 people.

Table 2. Statistical properties of features extracted from various sources of simulated spectrograms

<table>
<thead>
<tr>
<th>Feature</th>
<th>Kinect-Based</th>
<th>CMU-Based</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Variance</td>
</tr>
<tr>
<td><strong>WALKING</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 Average torso radial velocity</td>
<td>1.0946</td>
<td>8.6997x10^-5</td>
</tr>
<tr>
<td>2 Bandwidth of torso oscillation</td>
<td>0.5909</td>
<td>0.0041</td>
</tr>
<tr>
<td>3 Maximum of upper envelope</td>
<td>3.7199</td>
<td>0.0813</td>
</tr>
<tr>
<td>4 Minimum of lower envelope</td>
<td>0.0116</td>
<td>1.4282x10^-5</td>
</tr>
<tr>
<td>5 Total Doppler bandwidth</td>
<td>3.7315</td>
<td>0.0831</td>
</tr>
<tr>
<td>6 Average of upper envelope</td>
<td>2.6620</td>
<td>0.0459</td>
</tr>
<tr>
<td>7 Average of lower envelope</td>
<td>-0.4420</td>
<td>0.0446</td>
</tr>
<tr>
<td><strong>RUNNING</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 Average torso radial velocity</td>
<td>1.9366</td>
<td>3.1715x10^-5</td>
</tr>
<tr>
<td>2 Bandwidth of torso oscillation</td>
<td>0.6100</td>
<td>0.0021</td>
</tr>
<tr>
<td>3 Maximum of upper envelope</td>
<td>4.5660</td>
<td>0.1817</td>
</tr>
<tr>
<td>4 Minimum of lower envelope</td>
<td>-0.0044</td>
<td>1.3449x10^-5</td>
</tr>
<tr>
<td>5 Total Doppler bandwidth</td>
<td>4.5704</td>
<td>0.1847</td>
</tr>
<tr>
<td>6 Average of upper envelope</td>
<td>3.3132</td>
<td>0.0638</td>
</tr>
<tr>
<td>7 Average of lower envelope</td>
<td>0.4783</td>
<td>0.0705</td>
</tr>
</tbody>
</table>
records of 8 different people over a period of 1.5 s to yield a 1000-fold difference in speed variance.

Another way to visualize the statistical differences between CMU and Kinect data is to consider scatter plots of the features for different activities. Figure 12 (a), (b) shows the distribution of four features (average of lower envelope, average of upper envelope, bandwidth of torso oscillation, and average torso radial velocity) for discriminating walking versus running with Kinect-based spectrograms. Figure 12 (c), (d) shows the same plots, but this time for features extracted from CMU-based spectrograms. The scatter plots provide a basis for visually ascertaining which features are more discriminatory in nature. For example, comparing Figure 12(a) and Figure 12(b) for Kinect data, as well as Figure 12(c) and Figure 12(d) for CMU data, it may be observed that the average of the lower envelope provides more separation between classes than that of the minimum of the lower envelope.

In general, the scatter plots show that the features can be used to define more or less separable sets, indicating that classification based on these features can be accomplished, but that the CMU-based features enable greater distance between classes. Furthermore, the plots show that the Kinect-based spectrograms represent the underlying human motion well enough that activity classification can be accomplished.

**HUMAN ACTIVITY RECOGNITION FROM KINECT SPECTROGRAMS**

The human micro-Doppler spectrograms simulated from Kinect data were classified for the following activities: walking, running, boxing, leaping, and random motion. Sample Kinect spectrograms for boxing, leaping, and random motion are shown in Figure 12. The data was classified using k-nearest neighbors (kNN) (k=5) using a subset of 4 features selected using a wrapper approach [30]. As the wrapper method tests each possible feature combination in a brute force manner, the classification results yielded by wrapper-selected features represents an upper limit in the classification performance attainable for the five-neighbor kNN classifier.

The classification performance achieved is summarized by the confusion matrix given in Table 3. These results were obtained using 96 walking, 96 running, 96 boxing, 96 leaping, and 96 random records of a duration 6, 6, 1.5, 3, and 6 s, respectively.

---

**Figure 12.**
Scatter plots for Kinect-based features. (a) average of lower envelope, average of upper envelope, and bandwidth of torso oscillation. (b) average of torso radial velocity, maximum of upper envelope and minimum of lower envelope, and CMU-based features. (c) average of lower envelope, average of upper envelope, and bandwidth of torso oscillation.
60% of the data was used for training, and 40% for testing. The largest number of misclassifications occurred between walking and random motion (20%), as well as between running and random motion (10%). This result is not unexpected considering that the random motion only really differs from walking or running in terms of speed and pattern of limb motion. Although it may be surprising that boxing records are confused with walking and random records, this is understandable due to the similarities between leg oscillations in walking/random motion and the higher frequency arm movements seen in boxing.

In summary, the classification success rates achieved illustrate that the simulated micro-Doppler signatures generated are of sufficient quality to represent the nuances of varying activities and form a basis for activity recognition. Indeed, it was seen that the Kinect data were sufficiently accurate that it could even be used to classify CMU database-derived data. Consider the classification results shown in Table 4, in which walking, running, and leaping are classified using Kinect-derived spectrograms for training, and CMU-derived spectrograms for testing.

In the training process, Kinect-derived records composed of 96 walking records of 6 s duration, 96 running records of 6 s duration, and 96 leaping records of 1.5 s duration were utilized to classify CMU-derived records including 93 1.5 s walking records, 32 1.5 s running records, and 140 1.5 s leaping records.

The results in Table 4 show that Kinect-derived training data were useful to correctly classify, using the 5 neighbor kNN classifier, 80% of walking, 95% of running, and 91% of leaping records.

Table 3.

<table>
<thead>
<tr>
<th>Activity</th>
<th>Class</th>
<th>Walking</th>
<th>Running</th>
<th>Box Exercise</th>
<th>Leaping</th>
<th>Random</th>
</tr>
</thead>
<tbody>
<tr>
<td>Walking</td>
<td>Walking</td>
<td>0.80</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0.20</td>
</tr>
<tr>
<td>Running</td>
<td>Running</td>
<td>0.05</td>
<td>0.85</td>
<td>—</td>
<td>—</td>
<td>0.10</td>
</tr>
<tr>
<td>Leaping</td>
<td>Box Exercise</td>
<td>0.02</td>
<td>—</td>
<td>0.93</td>
<td>—</td>
<td>0.05</td>
</tr>
<tr>
<td>Random</td>
<td>Leaping</td>
<td>—</td>
<td>—</td>
<td>0.10</td>
<td>0.90</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>Random</td>
<td>0.10</td>
<td>—</td>
<td>0.03</td>
<td>—</td>
<td>0.87</td>
</tr>
</tbody>
</table>

Table 4.

<table>
<thead>
<tr>
<th>Classes</th>
<th>Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Walking</td>
</tr>
<tr>
<td>Walking</td>
<td>0.80</td>
</tr>
<tr>
<td>Running</td>
<td>0.05</td>
</tr>
<tr>
<td>Leaping</td>
<td>0.07</td>
</tr>
</tbody>
</table>

Figure 13.
This photo was taken on the beach in Akbuk, Turkey by the author. The inset shows a representative micro-Doppler signature of what the radar would see of the boy’s running. Basically, it shows how the happy-go-lucky running boy would look from the eyes of the radar.
cords. Not unexpectedly, the most confusion was observed between walking and running.

**CONCLUSION**

In this work, a low-cost, practical method for simulating human micro-Doppler signatures was developed using the Kinect sensor. Qualitative and quantitative comparisons of the Kinect-based simulated signatures were made against signatures simulated from other sources of MOCAP data, such as the CMU Motion Capture Library. Classification with a five neighbor kNN classifier yielded a correct classification rate of 80% for walking, 85% for running, 93% for boxing, 90% for leaping, and 87% for random motion. Furthermore, it was shown that Kinect-derived spectrograms were of sufficient quality that they could be used as training data to classify other sources of micro-Doppler data. Thus, it was validated that the Kinect-based system indeed generated simulated signatures that could be used for algorithm development related to human activity recognition.

**REFERENCES**


