Raja, Muneeba; Hughes, Philip A.; Xu, Y.; Zarei, P.; Michelson, D.; Sigg, Stephan

Wireless Multifrequency Feature Set to Simplify Human 3D Pose Estimation

Published in:
IEEE Antennas and Wireless Propagation Letters

DOI:
10.1109/LAWP.2019.2904580

Published: 01/05/2019

Document Version
Peer reviewed version

Please cite the original version:
Wireless Multi-frequency Feature Set to Simplify Human 3D Pose Estimation

Muneeba Raja, Aidan Hughes, Yixuan Xu, Parham Zarei, David G. Michelson, Senior Member, IEEE and Stephan Sigg

Abstract—We present a multi-frequency feature set to detect driver’s 3D head and torso movements from fluctuations in the Radio Frequency (RF) channel due to body movements. Current features used for movement detection are based on time-of-flight, received signal strength and channel state information, and come with the limitations of coarse tracking, sensitivity towards multi-path effects and handling corrupted phase data, respectively. There is no standalone feature set which accurately detects small and large movements and determines the direction in 3D space. We resolve this problem by using two radio signals at widely separated frequencies in a monostatic configuration. By combining information about displacement, velocity and direction of movements derived from the Doppler Effect at each frequency, we expand the number of existing features. We separate Pitch, Roll and Yaw movements of head from torso and arm. The extracted feature set is used to train a K-Nearest Neighbor classification algorithm which could provide behavioral awareness to cars while being less invasive as compared to camera-based systems. The training results on data from 4 participants reveal that at 1.8GHz, the classification accuracy is 77.4%, at 30GHz it is 87.4%, and multi-frequency feature set improves the accuracy to 92%.

Index Terms—Class membership, dynamic features, doppler Effect, multi-frequency, wireless sensors, 3D posture recognition

I. INTRODUCTION

HUMAN behaviour and attention monitoring holds great importance in medical, automotive and human computer interaction (HCI) systems. Primary indicators used to predict attention levels include eye lid movements, head orientations, postures and physiological changes.

Previous techniques for human activity indicators detection are cameras or wearable and ambient sensors [1]–[4]. Camera based techniques are able to detect fine grained visual features; however, it comes with the challenge of privacy intrusion and in accuracy in low ambient luminosity. Acceleration and physiological based wearable sensors detect orientation and heart rate/ breathing rate respectively; however, they rely on users carrying or wearing equipment, and changing sensor positions alters system performance. These limitations can be overcome by RF sensing.

RF sensing techniques allow ubiquitous, contact-free, and less privacy intrusive detection. Analyzing changes in RF channel characteristics due to body movements enable human activity/gesture/movement detection [5]–[11]. Features like Micro-Doppler signature and phase variations have proven to be reliable in large movement detection, such as gait, running, walking [12], small periodic movements such as vital signs [13], [14], and gesture and activity recognition [8], [15]. Doppler sensors are used by [16] for eye blinking detection, keeping sensors 6cm from the eyes using a carrier frequency of 5.8GHz. In addition, time-of-flight (TOF) with frequency modulated carrier wave (FMCW) performs coarse tracking of only one large body part (arm or leg) [17]. RSSI fluctuation is a low resolution and less predictable feature due to high sensitivity to multi-path effects and is not sufficient to obtain small-scale fading and direction of movements [18], [19]. Channel State Information (CSI), on the other hand, gives amplitude and phase information over multiple sub carriers, but phase correction processing makes it computationally intensive. Existing methods are mostly based on one frequency or a set of similar frequencies to detect a single type of movement. There is no stand-alone feature set which can accurately separate small and large movement within a system, and predict its direction in 3 dimensions. This brings the need to cater for natural behavioral situations, such as in a car, where we need not only to separate these movements but also to classify them.

In this paper, we introduce a multi-frequency based feature set to distinguish between directional big and small movements which are further separated into classes by applying appropriate machine learning algorithms. Our choice of frequencies is based on required wavelengths, size of the body part performing movement and displacement. Detectability or Radar Cross Section (RCS) of an object by a radar depends on its size and wavelength. The size is inversely proportional...
to wavelength [20], which means that \( \lambda \ll \text{size of the body part} \) in order to obtain the complete re-radiated signal. With manual measurements, we observe that head (small) movements typically cover shorter displacements \((d = 10 - 15\text{cm})\) as compared to movements of the large torso \((d = 40 - 50\text{cm})\) and arm \((d = 70 - 80\text{cm})\) movements. At 1.8GHz, \( \lambda = 16.7\text{cm} \), we can capture the movements equal or larger than \( \lambda \) which is the case where average sized subjects in our experiments perform torso and arm movements. At 30GHz, the \( \lambda \) is 1cm, which is much smaller than a proper head movement, therefore it is possible to capture it with high accuracy. The reason of using specifically 1.8GHz and 30GHz frequencies is that these met our criteria of wide separation, and ability to highlight large and small movements, and there were complete equipment facilities in our lab to use them for experiment set up and evaluate our technique.

We make the following contributions in this paper to analyze behaviour from naturally occurring body movements.

C1: Defining our classes based on 3D pitch, roll and yaw movements of head and torso which indicate human behaviour (c.f Fig. 1).

C2: Separating 3D rotational movements from non-invasive radio signals.

C3: Introducing a stand-alone multi-frequency based feature set for separating small and large movements.

C4: Detecting non-periodic movement and direction using Doppler spread spectrum analysis.

C5: Comparison of individual vs. combined frequency features with regards to accuracy.

The technical details of our system are explained in section II. Hardware setup and methodology is described in section III. As the detection methodology is same for pitch, roll and yaw, experiment outcomes for pitch movements are presented in section IV and conclusions are listed in section V.

<table>
<thead>
<tr>
<th>Pitch</th>
<th>( \phi = 0^\circ, \theta \in [-60^\circ + 60^\circ] )</th>
<th>forward, backward</th>
</tr>
</thead>
<tbody>
<tr>
<td>Roll</td>
<td>( \phi = 90^\circ, \theta \in [-60^\circ + 60^\circ] )</td>
<td>clockwise, anti-clockwise</td>
</tr>
<tr>
<td>Yaw</td>
<td>( \phi = 0^\circ, \theta = 0^\circ )</td>
<td>left, right</td>
</tr>
</tbody>
</table>

Fig. 2: Pitch for torso explained.

II. TECHNICAL CONCEPT

Our approach for estimating 3D head and torso movements is focused on scenarios where human body movements are partially restricted, as in the case of a driver with fastened seat belt. Pitch, roll or yaw movements are of great interest because they provide information about attention levels and behaviour [21]. We separate pitch, roll and yaw movements using multi-frequency RF features, as the movements influence the signal propagation path, resulting in channel quality fluctuation. In addition, we also detect arm movements. In this section, we first explain the technical definitions of pitch, roll or yaw, then our features and finally classification.

1) Pitch, Roll or Yaw: As shown in Fig. 2, the subject is modelled as a vector \( \vec{S} \) from the origin \( O_s \) at the waist (pointing towards the head).

To measure pitch, the receiver is placed such that \( \vec{R}_a \) is parallel to the ground (x-axis). Pitch is defined as a change in \( \theta \) in the \( \phi = 0^\circ \) direction:

\[
\vec{\omega}_p = (\omega_p \cdot \hat{\theta} + 0 \cdot \hat{\phi} + 0 \cdot \hat{r})
\]

Projecting the angular velocity \( \vec{\omega}_p \) onto the x-axis shows that the component of the angular velocity in direction of \( \vec{R}_a \) results in a measurable change in the Doppler frequency at the receiver (c.f Table I). Roll \( \vec{\omega}_r \) is defined in the same manner as a change in \( \phi \) in the \( \phi = 90^\circ \) direction and \( \vec{R}_a \) is parallel to the y-axis (receiver placed beside the subject). Yaw \( \vec{\omega}_y \) is defined as a change in \( \phi \) when \( \theta = 0^\circ \); however, this results in no significant components projected onto either the \( x \) or \( y \) axes, therefore we cannot distinguish between a left or right-hand yaw.

2) RF Features: In order to best understand the variations in propagation due to head and torso movements, we use the following frequency-time domain features computed at two frequencies and provide them as input for classification in our analysis.

\( a) \) Doppler Spread: The Doppler Effect is the primary feature for detecting body movements. A velocity in the radial direction of the receiver will cause a frequency shift. The total distribution of frequency shift due to the Doppler Effect is the Doppler spread \( B_D \).

\[
f_D = 2 \cdot v_m \cdot f_c / c\ [22],
\]

where \( v_m \) is movement velocity, \( c \) is speed of light and \( \lambda = c / f_c \) is the carrier wavelength. We calculate \( B_D \) by computing the Fast Fourier transform (FFT) of the received time-domain channel response and by measuring bandwidth above a defined power threshold [23].

\( b) \) Movement Velocity, Displacement and Direction: The movement velocity \( v \), is calculated from \( B_D \) to distinguish between head and torso movements. The system’s steady state is when torso and head are upright and static. A scalar displacement \( d \) from the steady state \( d = \frac{\vec{R}_a \cdot \vec{S}}{||\vec{R}_a||} \).

Different body movements have limited ranges of possible displacements from the steady state, depending upon the body part. For example, pitching the head forward will result in a smaller \( d \) than the torso. The range of a movement can be computed by integrating the measured velocity over time. In our case, the interesting movements are those which could be translated to distracted behavior. Using frequency as high as 30GHz, and a high sampling rate, we can capture extremely small displacements but our range and accuracy of displacements is conditioned on significant head turns and torso movements. The accuracy and precision, therefore, is not derived in terms of absolute values of displacement or velocity values but from class separation (e.g. visible from its confusion matrix).
TABLE II: VNA values for 30GHz and 1.8GHz.

<table>
<thead>
<tr>
<th>Sweep mode</th>
<th>Points</th>
<th>Time</th>
<th>IF BW</th>
<th>Power</th>
<th>Freq1</th>
<th>Freq2</th>
</tr>
</thead>
<tbody>
<tr>
<td>CW</td>
<td>1201</td>
<td>5.1s</td>
<td>10KHz</td>
<td>-20dBm</td>
<td>1449MHz</td>
<td>1750MHz</td>
</tr>
</tbody>
</table>

TABLE III: Measurement values for 30GHz and 1.8GHz.

<table>
<thead>
<tr>
<th>Frequency setup</th>
<th>TX beamwidth/gain</th>
<th>RX beamwidth/gain</th>
<th>Distance TX-RX</th>
<th>Distance RX-Subject</th>
</tr>
</thead>
<tbody>
<tr>
<td>30GHz</td>
<td>horn 18.3°/20dBdBi</td>
<td>horn 11.5°/23dBdBi</td>
<td>1.5m</td>
<td>1.6m</td>
</tr>
<tr>
<td>1.8GHz</td>
<td>directional CPE 60°/13dBdBi</td>
<td>1.5m</td>
<td>1.6m</td>
<td></td>
</tr>
</tbody>
</table>

c) Time-Frequency Transforms: Using the Vector Network Analyzer (VNA) parameters shown in Table II, the continuous wave (CW) signal is sampled at 234 complex samples/sec (Hz) with a sweep time of 5sec. We use a short time Fourier transform (STFT), which helps identifying the direction, time of occurrence, and duration of movements. The STFT applies a windowing function \( w(n) \) to an input signal \( x(n) \), which splits this signal into \( m \) sections.

3) Classification: We use the K-Nearest Neighbor algorithm on the calculated feature set, with 10-fold cross validation, prediction speed of 230obs/sec and training time of 0.5sec to classify movements. We explain our results in Section IV.

III. Methodology

Experiments are taken in an anechoic chamber located at the Radio Science lab at the University of British Columbia. Experiment setups are monostatic (TX-RX placed 1.5m apart) with subjects performing activities at 1.6m away from TX/RX. Two setups are deployed separately for 30GHz and 1.8GHz as shown in Fig. 3. The training data are captured for 4 subjects (height =157cm - 187cm), where measurements of each subject are taken at different times and days to avoid any environment bias in the results.

For the experiment at 30GHz, a signal block diagram to upconvert the VNA signal is shown in Fig. 3 and equipment details are given in Table III. For the experiment at 1.8GHz, the directional TX and RX antennas are connected directly to the VNA. The VNA is connected to a laptop via a TCP connection to remotely control the operations for experiments (cf. Table II).

For each experiment, subjects perform pitch, roll and yaw with alternate directions. Ground truth details are recorded manually with each measurement for comparison.

We propose this technique for static environments where all the physical entities remain stable including positioning of objects, signal-to-noise ratio, and clutters [24]. The signal changes are solely due to the human body and not the environment. Interference from movements, such as, due to multiple persons is expected in realistic environments and will need to be resolved in future work.

IV. Results

Based on our calculated features, Fig. 4 depict the results obtained for \( B_D \) at 1.8GHz and 30GHz. At 1.8GHz, we can only detect the large movements due to much larger range of displacement than the \( \lambda \) of 16.7cm.

At 30GHz, the wavelength, \( \lambda \) of 1cm is much smaller than the ranges of both head and torso movement displacements so that both head and torso movements can be detected. Doppler spread width determines the type of movement while positive/negative frequency shift indicates the orientation (forward/backwards for pitch). Fig. 4 shows that the Doppler shift in either direction can be detected for head movements. The time-frequency transform helps to determine the real time changes in frequency as well as the direction (Fig. 5). Confusion matrices, recall and precision values of the computed features on training data are shown in Fig. 6 and Fig. 7. For the separate frequency data feature set for 1.8GHz, we achieve a classification accuracy \(^2\) of 77.8%, for 30GHz it is 87.4% while the combined frequency feature set reaches the improved accuracy of 92%.

Arm movements cannot be classified as pitch, roll, and yaw, therefore additional features are required. Arm movements could be filtered out using signal processing techniques before the feature calculation step. Alternatively, a separated class is defined on arm movements for enhanced behavior and activity recognition. This would require techniques such as modelling the head, torso and arm as separate point scatters which can be divided in the delay domain, similar to the fast time processing techniques shown by [25].

V. Conclusions

We presented an RF-based multi-frequency feature set to separate large (torso) and small (head) movements in 3D within the same system. Existing feature sets, such as time of flight, RSSI and CSI, are not standalone and cannot accurately detect small and large movements at the same time nor determine body direction in 3D space. Our feature set is based on dynamic features derived from time-varying body movements, such as Doppler spread, displacement, velocity and direction to classify the 3D pitch, roll and yaw of head and torso. Wide separation of frequencies ensures that large movements cause visible deep fading on low frequencies and small movements are detectable on high frequencies due to their wavelengths. This feature set widely separates the classes when trained over a K-Nearest Neighbor algorithm, which could provide human behavioral awareness to scenarios such as, in-car personal assistant systems so that more personalized feedback can be provided to the driver. The training results on data from a set of 4 participants\(^1\) reveal that at 1.8GHz the accuracy is 77.4% and at 30GHz it is 87.4%, while using a combined feature set gives an accuracy of 92%. In the future, we intend to carry out experiments outside of the anechoic chamber, in realistic environments. Furthermore, the feature set could be enhanced to include the detection of complex arm movements using a motion capture device and techniques such as modelling the head, torso and arm as separate point scatters which can be isolated in the delay domain.

\(^1\)TF=tfosoro forward, TB=torsor backward, HF=head forward, HB=head backwards, AM=arm movements, Prec.=precision, Recl.=recall.

\(^2\)Classification accuracy = \( \frac{TP + TN}{TP + TN + FP + FN} \)
Fig. 3: Left: signal block diagram for 30GHz frequency measurements; right: 30GHz and 1.8GHz setup in the anechoic chamber.

Fig. 4: Doppler spread visualization in frequency domain graph for torso and head movements.

(a) Torso forwards and backwards at 1.8GHz.
(b) Head forwards and backwards at 30GHz.

Fig. 5: Spectrograms for 30GHz experiments for Pitch.

(a) Head forwards
(b) Head backwards
(c) Torso forwards
(d) Torso backwards

Fig. 6: Confusion matrix for 1.8GHz and 30 GHz frequency data.

Fig. 7: Confusion matrix for multi-frequency feature set.
REFERENCES


