LONG TERM HUMAN ACTIVITY RECOGNITION WITH AUTOMATIC ORIENTATION ESTIMATION

Blanca Florentino-Liaño, Niamh O’Mahony, Antonio Artés-Rodríguez

Department of Signal Theory and Communications, Universidad Carlos III de Madrid, Avenida de la Universidad, 30, Leganés, 28911, Spain
Email: {blanca, niamh, antonio}@tsc.uc3m.es

ABSTRACT
This work deals with the elimination of sensitivity to sensor orientation in the task of human daily activity recognition using a single miniature inertial sensor. The proposed method detects time intervals of walking, automatically estimating the orientation in these intervals and transforming the observed signals to a “virtual” sensor orientation. Classification results show that excellent performance, in terms of both precision and recall (up to 100%), is achieved, for long-term recordings in real-life settings.

Index Terms—Activity recognition, wearable sensors, orientation estimation.

1. INTRODUCTION
Human activity recognition using wearable inertial sensors has recently become a popular topic of research interest due to the growth of applications based on context-aware monitoring. In healthcare applications, knowing the activity being carried out by a patient provides context for biomedical measurements, such as heart rate or blood pressure, allowing the medical professional to infer the underlying health status of the patient and to make a more accurate analysis of these measurements than in a stand-alone monitoring system. This context-awareness can help to overcome the limitations associated with the use of self-reporting in medical assessment and, also, can reduce the frequency of patients’ visits to medical centers, improving their quality of life and reducing medical costs. For these reasons, automatic recognition of context is essential for many healthcare applications.

In recent years, inertial sensor devices have become compact and portable enough to be unobtrusively attached to the human body. For this reason, wearable miniature inertial sensors have become the ideal platform in monitoring applications, such as human movement monitoring [1] and tele-rehabilitation [2]. One major challenge in activity recognition is the development of a system that is independent of the sensor orientation, since, in real life situations, control of the exact orientation of the sensor is not feasible, due to variations in body shape, clothing and other factors. This can significantly affect the accuracy of automatic activity recognition.

This work describes an activity recognition algorithm based on a single inertial sensor that allows recognition of day to day activities, such as sitting, lying, standing, walking, running and jumping, in a “plug-and-play” manner. “Plug-and-play” here refers to the fact that the user is simply required to attach the sensor to a belt, with no specific requirements for its exact placement and orientation. Thus, the algorithm is ideal for personalized medicine applications and home-based monitoring. This feature is achieved by making the algorithm insensitive to the orientation of the sensor through automatically recognising when the subject is in an upright pose, but without requiring any prior information about the subject’s initial pose. The proposed algorithm can achieve high classification results on long term recordings of day to day activities.

The paper is organized as follows: in Section 2, background theory is presented. Section 3 describes the proposed solution to make the activity recognition algorithm insensitive to sensor orientation. The system validation is outlined in Section 4, whilst, in Section 5, the results obtained with our activity recognition system are presented. Finally, in Section 6, conclusions and future lines of work are discussed.

2. BACKGROUND

2.1. Inertial sensors
Inertial measurement units (IMUs), combining triaxial accelerometers and gyroscopes, are often used for activity recognition [3–5]. Another common approach is based on triaxial accelerometers alone [6,7], with many applications in smartphones, which have embedded accelerometers [8–10]. In this work, an APDM Opal [11] IMU, consisting of both
accelerometers and gyroscopes, is used. A single device is placed on the hip for data collection.

The accelerometers measure, at time, \( t \), the total inertial force, \( \alpha_t \), acting on the sensor, in m/s\(^2\). This inertial force includes both linear accelerations, \( a_t \), in each of the three sensor axes and a gravitational force component, \( g_t \), in each axis. The gyroscopes measure the angular velocity, \( \omega_t \), of the sensor, in rad/s.

The signals recorded by the inertial sensors are measured in a three-dimensional coordinate system which is fixed to, and moves with, the sensor. This frame is referred to as the sensor frame (\( S \)) and is defined by a mutually orthogonal set of unit vectors, \( \{ \hat{x}_S, \hat{y}_S, \hat{z}_S \} \). A fixed frame (\( F \)) can also be defined, in which the gravitational component of the force of acceleration is constant; for example, in the frame, \( \{ \hat{x}_F, \hat{y}_F, \hat{z}_F \} = \{ \text{North, West, Up} \} \), acceleration due to gravity is given by \( G^F \approx [0, 0, 9.81] \) m/s\(^2\).

The body frame (\( B \)) is fixed to, and moves with, the center of mass of the subject’s body (approximately located at the waist). This is the frame of a “virtual sensor” to which all of the sensor measurements will be transformed. The direction of each axis, relative to the subject’s body, in a standing position, can be described as: \( \{ \hat{x}_B, \hat{y}_B, \hat{z}_B \} = \{ \text{Forward, Left, Up} \} \). It should be remembered that as the subject moves, these directions will change with respect to the fixed frame.

The final coordinate system to be introduced is a reference frame (\( R \)), which is defined as the body frame when the subject is standing still. This reference system is made up of a set of unit vector defined by \( \{ \hat{x}_R, \hat{y}_R, \hat{z}_R \} = \{ \text{Forward, Sideward, Vertical} \} \). Every time the patient is upright the reference frame and the body frame are aligned. The vertical axis of the reference frame is always the same, in terms of the fixed frame, since it corresponds to the direction of the gravity. The directions of Forward and Sideward, relative to North and West of the fixed frame can change.

### 2.2. Activity recognition

Most of the published work in activity recognition follows a common approach, consisting of two steps: a processing step and a classification step. The processing step, typically, focuses on the construction of a feature vector derived from the raw sensor signals. In the literature, a large number of different features have been reported as being suitable for activity recognition; [12] provides a comparison of the most popular features. Once the feature vector has been computed, the next step is the development of a model that is able to discriminate among activities. The most popular methods used are batch supervised learning algorithms and dynamic Bayesian networks (DBN). In [4], a comparison of classification results using various batch supervised learning algorithms, including Bayesian decision making (BDM), least-squares method (LSM), \( k \)-nearest neighbor (\( k \)-NN), support vector machines (SVM) and artificial neural networks (ANN), can be found.

In the case of DBNs, hidden Markov models (HMM) are the most frequently used [5, 6, 10].

### 2.3. Hierarchical dynamic model with HMM

Recently, the authors presented a novel algorithm for the classification of human activities based on a hierarchical dynamic model (HDM) [13]. This method was shown to give competitive classification results, compared to state-of-the-art methods, whilst avoiding the computational bottleneck of traditional feature extraction methods, by basing the entire algorithm on the raw sensor signals. This section briefly reviews the HDM with HMM; for details see [13].

The HDM with HMM constructs a dynamic model, taking two levels of dynamics into account: inter-activity and intra-activity. The inter-activity dynamics refer to the temporal dependency among activities, modeling, by transition probabilities, the dependency of the current activity on the previous one. On the other hand, the intra-activity dynamics model the amplitude of the signals and how they evolve in time, by constructing a HMM for each of the activities.

Various different intra-activity models are proposed, each with a different topology, according to the type of activity being modeled. For example, for stationary activities like standing, sitting and lying, a left-right model with three states is defined. The first and the last states are transient states and the state in the middle models the permanent state of, for example, being seated.

Due to using raw signals, the HDM is more sensitive to the orientation of the sensor than feature extraction methods. Thus, it is essential to make the signals invariant to orientation. This invariance is achieved by a transformation, which will be described in Section 3.1. Even if a phase of feature extraction does exist, the transformation of the signals to a common reference frame can make the algorithm more robust.

### 2.4. Orientation invariance

Several works in the activity recognition literature have faced the problem of sensitivity to sensor orientation. The most popular approach is to estimate the gravity vector in the sensor frame, in order to align the sensor frame with the vertical axis of the reference frame.

Mizell’s work [14] proposed a method for determining the vertical axis and the horizontal plane by estimating the gravity vector from acceleration signals (\( \alpha^2_t \)) averaged over a time interval. This approximation has been used in activity recognition to compute robust features from the transformed raw signals [9, 10] and, also, in [7, 15] but, there, only during time intervals where the linear acceleration is negligible (i.e. static activities). However, these approaches do not take account of the fact that different static activities exist, like lying down, in which the gravity vector does not correspond to the vertical axis.
In [8, 16, 17], dynamic portions of the signals are used to estimate the gravity vector, under the assumption that most dynamic activities are bipedal locomotion, ensuring that the vertical axis is aligned with the gravity vector. The gravity vector is estimated using an acceleration signal which contains both gravity and linear acceleration due to locomotion. Thus, the estimate is inaccurate and leads to non-alignment of the vertical axis and, hence, misclassification of activities [9].

Recently, the authors have presented a method to make the activity recognition algorithm insensitive to sensor orientation [18]. The presented method requires information about the activity recognition algorithm insensitive to sensor orientation consists of, firstly, identifying walking epochs, since, in these epochs, the body frame is aligned to the reference frame, orientation estimation is shown in Fig. 1. The inputs of the algorithm in [13] appears that all measurements have been recorded from the vertical sensor frame to the body frame, such that the $z$-component of the transformed measurement is aligned with the $z^B$-axis and the $x^S$-$y^S$ plane is aligned with the $x^R$-$y^R$ plane. The rotation matrix is calculated by:

$$ R(\theta_x, \theta_y) = \begin{bmatrix} C(\theta_y) & S(\theta_y) & C(\theta_x)S(\theta_y) \\ 0 & C(\theta_x) & -S(\theta_x)S(\theta_y) \\ -S(\theta_y) & C(\theta_y) & C(\theta_x)C(\theta_y) \end{bmatrix} $$

(3)

where $C$ and $S$ denote cosine and sine, respectively.

Using (3), the measured signals are all transformed by the same constant rotation at each time instant, such that it appears that all measurements have been recorded from the virtual sensor position. The transformation of the acceleration is given by:

$$ \alpha^B_t = R(\theta_x, \theta_y) \alpha^S_t, $$

(4)

where the frame, $B^t$, denotes the body frame with an arbitrary yaw angle. The gyroscope signals are transformed in the same way.

The $x^B$- and $y^B$-components remain dependent on the initial yaw. However, the modulus of the acceleration in the $x^B$-$y^B$ plane, $\alpha^B_{x-y} = ||\alpha^B_x, \alpha^B_y||$ is independent of the initial yaw. Thus, the transformed signals are two-dimensional signals given by $\alpha^T = \{\alpha^T_{x-y}, \alpha^T_{z} \}$ and $\omega^T = \{\omega^T_{x-y}, \omega^T_{z} \}$. This part of the algorithm was published by the authors in [18].

### 3.2. Automatic orientation estimation

In order to estimate the orientation, as outlined in the previous section, the pitch and roll are estimated using the acceleration signals when the body frame and the reference frame are aligned. This can be done using known epochs of low linear acceleration, while the subject is standing. However, these epochs are not known to the algorithm. The automatic orientation estimation algorithm consists of a walking detector, for finding segments when the subject is upright, and a fine-tuning process, for adjusting the pitch and roll angles, in order to eliminate the effects of the linear acceleration on the estimation of the angles.

#### 3.2.1. Walking detector

The magnitude of the signals ($||\alpha^S||$, $||\omega^S||$) is insensitive to sensor orientation and, during dynamic activities, characteristic dynamic patterns can be observed in $||\alpha^S||$ and in $||\omega^S||$. For these reasons, the magnitude is used to detect walking.

A HDM with HMM [13] is used to discriminate between walking, running and static activities. Here, walking must be identified with high precision (100%) and, so, only segments of, at least, 20 seconds of inferred activity, “walking”, are
used for the orientation estimation step. Averaging the acceleration signals across these segments, initial estimates of roll, \( \hat{\theta}_x \), and pitch, \( \hat{\theta}_y \), are calculated, using (1) and (2), respectively. These estimates are not accurate, since the acceleration signals contain not only gravity but also linear acceleration. Therefore, fine-tuning of these angles is needed.

3.2.2. Fine-tuning

The inputs of this part of the algorithm are \( \hat{\theta}_x \), \( \hat{\theta}_y \) and the three-axis acceleration signal, \( A^S = \alpha^S_{i:\text{walking}} \), of the walking segment from which these angles were calculated. Roll-pitch pairs, \( \{ \theta^k_x, \theta^k_y \} \), are calculated, such that, \( \theta^k_x = \hat{\theta}_i + k\delta \), for \( i \in \{x,y\}, k = -4, \cdots, 4 \) and \( \delta \) is a small adjustment angle, e.g. \( \delta = 0.05 \text{ rad} \). The transformed walking segment, \( A^{T}_k \), is calculated, using equation (4), for each pitch-roll pair, identified by \( k \). The likelihood, \( L_k \), that \( A^{T}_k \) was generated with the walking model of the HDM of the recognizer algorithm is computed for each \( k \). The roll-pitch pair that produces the highest likelihood is selected, using \( k_{\text{opt}} = \arg \max_k L_k \), and the final orientation estimate is \( \{ \theta_x, \theta_y \} = \{ \theta^k_{x,\text{opt}}, \theta^k_{y,\text{opt}} \} \).

3.2.3. Orientation update

The automatic orientation estimation process is carried out for every walking segment detected. Therefore, the transformation can be updated each time walking is detected. This allows our algorithm to deal with changes in sensor orientation while recording.

4. ALGORITHM VALIDATION

The model was tested using activity sequences recorded by sensors placed on the hip in a random orientation and the subject was told to remain standing still for the first 5 seconds. Two experiments were performed. In the first experiment, the algorithm was given prior information about the 5 seconds of standing and the transformation of the signals was computed using roll and pitch angles estimated during this interval. In the second experiment, no prior information was given to the algorithm, so the automatic orientation estimation algorithm, proposed in previous section, was used.

4.1. Database description, training and testing

To train the activity recognition algorithm, the database from [3], on which the results of our previous work [13, 18] were based, has been used. This database consists of 4 hours and 30 minutes of activity data from 16 subjects (6 females and 10 males) aged between 23 and 50 years. Data were recorded in semi-naturalistic conditions. The IMU was placed either on the right or left hip, providing 3-axis acceleration and 3-axis angular velocity signals at a sampling rate of 100 Hz. More details of the data collection and labeling can be found in [3].

With this database, two HDMs were constructed. The first one, HDM1, was used in the recognizer algorithm (Fig. 1). The inputs of this model are the transformed signals \((\alpha^T, \omega^T)\). In the training phase, the signals were transformed using true segments of standing (i.e. the activity was known). The second one, HDM2, was used in the walking detector algorithm. The inputs of this model are the magnitude of the signals \((||\alpha^S||, ||\omega^S||)\), so no transformation is required. More details on HDM construction and training can be found in [13].

In the test phase, signals were recorded by the authors using APDM sensors [11]. A test database, consisting of six long term recordings (> 40 mins) of day-to-day activities from six different subjects was constructed. The sensor was attached to the belt of the subject without any restriction on the exact place or the orientation. The sequences were recorded in different naturalistic scenarios: having a coffee at the workplace, working at the desk and watching a football...
match at home.

5. RESULTS AND DISCUSSION

Table 1 shows the total precision and recall values of each activity for the test sequences, using the angles computed from labeled epochs of standing (Automatic Orientation: NO) and using the proposed automatic orientation estimation with updates (Automatic Orientation: YES). The precision of activity, $Z$, is measured as number of samples classified correctly as activity, $Z$, divided by the total number of samples with inferred label equal to $Z$. The recall parameter is the number of samples correctly classified as activity, $Z$, divided by the number of samples whose true label is $Z$.

<table>
<thead>
<tr>
<th>Activity</th>
<th>Precision</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Running</td>
<td>100%</td>
<td>90%</td>
</tr>
<tr>
<td>Walking</td>
<td>96%</td>
<td>95%</td>
</tr>
<tr>
<td>Standing</td>
<td>97%</td>
<td>95%</td>
</tr>
<tr>
<td>Sitting</td>
<td>100%</td>
<td>96%</td>
</tr>
<tr>
<td>Lying</td>
<td>68%</td>
<td>100%</td>
</tr>
<tr>
<td>Jump</td>
<td>60%</td>
<td>91%</td>
</tr>
<tr>
<td>Average</td>
<td>87%</td>
<td>95%</td>
</tr>
</tbody>
</table>

As shown in Table 1, similar results are obtained with and without prior information (labeled standing epochs), showing that our automatic orientation estimation algorithm works very well. This is a significant improvement in activity recognition, as no restrictions on the data collection protocol are required. The results also show high recall and precision in sitting and standing. These activities are very easily confused using only one inertial sensor placed on the hip since the difference in hip orientation between these activities is very subtle. This shows that our automatic algorithm estimates the orientation of the sensor very accurately.

The low precision achieved in lying is, in part, a result of inferring lying while the true activity was sitting on the sofa, in a slouched pose. Such a pose is subject to interpretation as either sitting or lying. The label in this case was given as sitting.

The activity, jumping, also has low precision because the initial samples of running have, at times, been inferred as jumping. Arguably, running, by its nature, contains an element of jumping. Nevertheless, every activity achieves high recall.

6. CONCLUSIONS AND FUTURE WORK

The major novelty of the work lies in the ability of the proposed system to accurately determine a subject’s activity sequence, using one single IMU, placed in any arbitrary orientation and position on a belt, without any requirement for prior information. Monitoring can be carried out over long durations of time and in natural settings and the algorithm can successfully distinguish between the static activities, lying, sitting and standing, a task which is not achieved by many existing single sensor algorithms.

The activity models in the method could be easily tuned for an individual subject, tailoring them to the way the subject tends to walk or run, to potentially provide higher accuracy for that subject than the general models designed in this work. This remains as future work.

7. REFERENCES


