17th Mediterranean Conference on Control & Automation
Makedonia Palace, Thessaloniki, Greece
June 24 - 26, 2009

Activity recognition using a wrist-worn inertial measurement unit: a case study for industrial assembly lines

Heli Koskimäki, Ville Huikari, Pekka Siirtola, Perttu Laurinen, Juha Röning
Intelligent Systems Group, University of Oulu,
Po Box 4500, 90014 University of Oulu, Finland
Email: {heli.koskimaki, ville.huikari, pekka.siirtola, perttu.laurinen, juha.roning@ee.oulu.fi}

***Abstract*—As wearable sensors are becoming more common, their utilization in real-world applications is also becoming more attractive. In this study, a single wrist-worn inertial measurement unit was attached to the active wrist of a worker and acceleration and angular speed information was used to decide what activity the worker was performing at certain time intervals. This activity information can then be used for proactive instruction systems or to ensure that all the needed work phases are performed. In this study, the selected activities were basic tasks of hammering, screwing, spanner use and using a power drill for screwing. In addition, a null activity class consisting of other activities (moving around the post, staying still, changing tools) was defined. The performed activity could then be recognized online by using a sliding window method to divide the data into two-second intervals and overlapping two adjacent windows by 1.5 seconds. Thus, the activity was recognized every half second. The method used for the actual recognition was the k nearest neighbor method with a specific distance boundary for classifying completely new events as null data. In addition, the final class was decided by using a majority vote to classifications of three adjacent windows. The results showed that almost 90 percent accuracy can be achieved with this kind of setting; the activity-specific accuracies for hammering, screwing, spanner use, power drilling and null data were 96.4%, 89.7%, 89.5%, 77.6% and 89.0%, respectively. In addition, in a case with completely new null events, use of the specific distance measure improved accuracy from 68.6% to 82.3%.**

I. INTRODUCTION

As the practical constraints related to wearable sensors have been solved, it is now possible to utilize the sensors more widely in various real-world applications. For example, they can be used to monitor workers' activities on different industrial assembly lines. In this study, a single wrist-worn inertial measurement unit was attached to a worker's active wrist, and the measured data were used to decide what activities the worker was doing at certain time intervals. This kind of monitoring enables development of proactive systems. For example, the worker can be instructed to perform one task at a time, with the instructions automatically changing according to the performed tasks. In addition, the information can be used to ensure that all the needed phases are performed before the product is sent forward on the assembly line.

In this study four different activities were recognized using data obtained from a triaxial accelerator and a gyroscope. The activities were hammering, screwing with a screwdriver, using a spanner and attaching screws with a power drill. In addition, the angle at which the activities were performed varied between horizontal and vertical. The data were gathered as a sequence, where the data consisted of not only performance of the activities, but also data where the worker moved around the post, changed the tool being used or just stayed still. These extra activities were considered null data, which constituted their own class. The activities were recognized by using the k nearest neighbor method (knn) to classify the data. The method made it possible to use an extra distance boundary, so that events with all their nearest neighbors outside this distance boundary were considered null data. This made it possible to classify completely new events where the worker was not performing the usual sequence, even though there was no training data available from these new events. This boundary was selected so that it did not affect the classification when the data set consisted only of data from the four work activities and basic null data.

Previously, a similar problem using wearable sensors has been tackled from the viewpoint of a mock assembly scenario ([1] and [2]). In these studies the work phases were recognized by combining data from an accelerometer and a microphone. In addition, the best results were achieved when using data from the same person for training and testing, while our system is user independent and the data are gathered from an accelerometer and a gyroscope. Another difference in ([1] and [2]) compared with our study is that, although the data set consisted of null data, these null data points were also classified as some of the actual activities. However, with our approach, using the null data as a separate class gave more accurate overall results. In addition, the selected distance boundary helped in handling data from totally new events.

The accelerometer-based approach to monitoring human movements, in general, is a widely spreading research area. Accelerometers have been utilized in different application areas, including metabolic energy expenditure, physical activity, postural sway, gait, fall detection and activity classification [3]. Although Ward et al. considered activity recognition as a special case of gesture recognition and adapted the Hidden Markov model based on its functionality in gesture recognition studies, the basic activity classification approach does not rely on gestures. Nevertheless, the other activity classification

Fig. 1. Inertial measurement unit on the active wrist of a worker.

Fig. 2. Used tools: bumping mallet (hammer), screwdriver, spanner and power drill.

studies comparable to our approach mainly consider daily activities or sports. For example, good results have been achieved when using a triaxial accelerometer to recognize periods of sit-to-stand and stand-to-sit transitions and walking [4]. On the other hand, to recognize a more complicated set of movements, the amount of inertial measurement units has been increased; in [5] five biaxial accelerometers were attached to the hip, wrist, arm, ankle and thigh.

This article is organized as follows: Section II introduces the inertial measurement unit and the data gathering setup. The sliding window method and basic feature extraction are explained in Section III, and the study and results are covered in Section IV. The results are then discussed in more detail in Section V and the whole study is concluded in Section VI.

II. Data collection

The data were collected using an inertial measurement unit, called Shake. It can be used for wireless recording and collection of context data and it is manufactured by SAMH Engineering Services (Ireland) and is about the size of a matchbox. The device contains the following sensors: 3D accelerometer, 3D magnetometer and 3D gyroscope. In the data collection phase the inertial measurement unit is attached to the worker's active wrist (Figure 1). In this study only the accelerometer and gyroscope data were used and they were collected at a frequency of 100 Hz. Additional video data were also collected with a video camera mounted near the place of work. The video data were not used with the recognition system, but to label the data in order to analyze the results. The labeling can easily be done, for example, with a time series visualization method that augments visualization of a numeric time series stream with an easily perceivable audiovisual stream [6].

The data were gathered as a sequence in which a worker fastened / opened four vertical and four horizontal screws using a power drill, screwed one vertical and one horizontal screw with a screwdriver, hammered in the vertical and horizontal directions and fastened / opened two nuts with a spanner. The tools are shown in Figure 2. In this study fastening and opening were considered to belong to the same class, although in further studies they can be split into separate classes.

In addition, the data set constituted basic limited null data in which the worker moved around the post, changed tools or just stayed still. Altogether four sequences lasting approximately four minutes were measured from four different persons. An additional remark concerning the sequence is that the order in which the tasks were performed was not pre-defined and the number of repetitions of each activity depended heavily on the worker.

III. Feature extraction

In order to utilize the system on a production line, the activity recognition method should be usable online. Thus, a sliding window method was used to divide the time series signals into smaller sequences (Figure 3). A window length of two seconds with a slide of 0.5 seconds between two sequential windows was found to be most suitable for the system. This means the system classified the ongoing activity at intervals of a half a second. Using the sliding window method, altogether 4814 windows were obtained, while a single sequence consisted of approximately 400 windows.

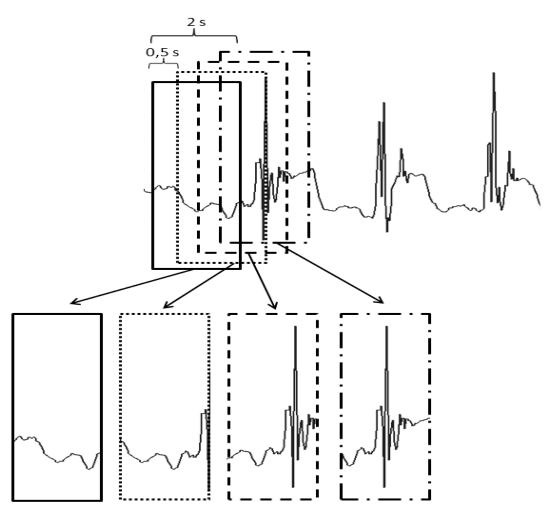

Fig. 3. Principle of the sliding window method.

However, there is a small problem with this kind of approach from the standpoint of result presentation. The data set also includes windows in which the activity changes. For example, a window can consist of a half a second of hammering and one and a half seconds of null data. For this study, it was decided to label these cases according to the majority activity in the window (the previous example would be labeled as null data). Nevertheless, also results in which these kinds of mixed windows were removed from the data set are presented in Section IV. From the standpoint of an actual application, this does not cause any problems, as it is not significant if the activity lasts four or five seconds if it is correctly classified.

For the final recognition system, different features were calculated from the two-second windows to compress the information of interest in the data. The calculated features consisted of

- simple statistical values calculated for every channel of acceleration and angular speed separately
  - mean, standard deviation, minimum, maximum, median, quartiles
- frequency domain features
  - sums of smaller sequences of Fourier-transformed signals
- correlation features between different channels

Altogether 84 different features were calculated. However, there were plenty of correlations between the features, and to ease the calculation load only the 20 most significant features were chosen using a sequential forward selection method.

IV. Study

The idea of the worker monitoring system is to be as imperceptible as possible. The attached measurement units should not disturb the worker, and thus this study was done using only a single inertial measurement unit attached to the active wrist of the worker. In addition, the system was developed to be user-independent, making the results more adaptable to an actual production line environment.

In the study, a k nearest neighbors classifier (knn) was utilized. The idea of knn classification is quite simple: a data point is classified into the class where most of its k nearest neighbors belong [7]. The nearest neighbors are defined using, for example, the Euclidean distance measure [8]. The final classification was based on a majority vote between three adjacent windows, utilizing the fact that in most cases the three adjacent windows are from the same measured activity. If all three adjacent windows belonged to a different class, the window was classified as null data.

In addition, a specific distance boundary was utilized to classify completely new events as null data, as a comprehensive null data set cannot be gathered. This meant that if there were no nearest neighbors inside this distance, the data points were directly classified as null data. In practice, the boundary was specified to be large enough to not have a meaningful effect on the classification in the basic case. The term 'basic case' is used here to describe a data set consisting of performed activities and a limited null data set in which the worker moved around the post, changed tools or just stayed still.

The results were calculated for each person individually using the sequences from other persons as training data. With a setting where the windows could be comprised of two different activities, overall accuracy was 88.2%. Table I presents a confusion matrix of the results. It can be seen that the power drill was most difficult to classify; approximately every fifth two-second window was classified as null data. On the other hand, this is quite obvious, as the acceleration and angular speed do not change much during drilling and thus the activity is easily mixed with null data where the hand is held still. Approximately 90% accuracy was achieved with the other activities, which can be considered very high accuracy when only a single inertial measurement unit is used.

On the other hand, more accurate results were naturally achieved in the case where all the windows consisting of several activities were removed. Total accuracy for all the activities was then 92.2 %. Nevertheless, also in this case the lowest accuracy was achieved in recognizing the power drill (Table II).

As it was mentioned earlier, the extra distance boundary was added to classify abnormal activities into a null class, but in a way that did not affect the actual classification. This means that, although the distance boundary was used when classifying the windows presented in Tables I and II, classification accuracy would be approximately the same also without the use of the boundary. This is not surprising, as our data set consisted of the four activities and basic null actions (changing tools, moving around the post and staying still). Nevertheless, to explain the usefulness of the distance boundary in an actual assembly line setting, a diversified eight-minute null data set was gathered from one subject. This data set was comprised of typing with a keyboard, writing with a pencil, walking and drinking, which were considered actions that a worker could do between actual work activities. When classifying this data set without using the boundary and using the sequences from the three other subjects as training data, 68.6% of the data set were correctly classified as null data. The remaining 31.4% were classified as belonging to one of the four work activities (in most cases screwing with a power drill). However, when the boundary was added to the classifier, accuracy of 82.3% was achieved. This accuracy corresponds better with the accuracy of the whole system and can be considered tolerable from the industrial point of view. Thus, it can be concluded that, although the boundary does not give any extra advantage in an ideal scenario, in real-world assembly lines the boundary is a valuable addition to the system.

V. Discussion

As it was shown in Section IV, the basic activities performed by the worker can be recognized with high accuracy using acceleration and angular speed information from an inertial measurement unit. From Tables I and II it can be seen that very rarely are the actual activities mixed up, but misclassifications occur with null data. Naturally, to improve recognition accuracy, RFID tags and readers could be utilized

TABLE I

CONFUSION MATRIX OF ACTIVITY RECOGNITION ACCURACIES WHEN ALL THE WINDOWS WERE IN THE DATA SET.

true class \ predicted class	null	hammering	screwing	spanner using	power drill	percent
null	**1788**	6	33	74	107	89.0%
hammering	7	**216**	0	1	0	96.4%
screwing	31	0	**849**	67	0	89.7%
spanner using	55	0	53	**925**	0	89.5%
power drill	135	0	0	0	**467**	77.6%
total						**88.2%**

TABLE II

CONFUSION MATRIX OF ACTIVITY RECOGNITION ACCURACIES WHEN THE WINDOWS CONSISTING OF SEVERAL ACTIVITIES WERE REMOVED FROM THE DATA SET.

true class \ predicted class	null	hammering	screwing	spanner using	power drill	percent
null	**1468**	0	12	37	72	92.4%
hammering	6	**179**	0	2	0	95.7%
screwing	21	0	**837**	31	0	94.2%
spanner using	50	0	19	**900**	0	92.9%
power drill	59	0	0	0	**280**	82.6%
total						**92.2%**

to inform when the tools are actually in use. However, as mentioned earlier, at this point the we sought to minimize the amount and size of measurement units, and thus the RFID approach was not studied more carefully.

An interesting observation is that, although sequences from only three different persons were used as training data, classification accuracy was remarkably high. This can be explained with the similar way the activities are performed by different workers. On the other hand, the tasks on production assembly lines are pre-defined and the optimal way of performing the tasks is instructed in advance. Thus, the user-independent approach is suitable for the activity recognition system.

There are two major uses for activity recognition information. In the first one recognition information can be used in the development of proactive instruction systems. Further instructions are provided when the system recognizes that the previous activity has been performed. For example, on an assembly line a screw should be screwed and after that some specific badding knobs should be attached using a rubber hammer. After the activity recognition system has detected that the screwing has ended, the instruction can be automatically changed to describe the attachment of the knobs.

The second use for activity recognition information is related to quality monitoring. As reported in [9], activity recognition information can be used to monitor tasks performed by the worker to ensure that all the needed work phases are done. Nevertheless, in our study we also wanted monitoring to be performed with as few inertial measurement units as possible. With the accuracies achieved in activity recognition and studying the distribution of the misclassified samples, this is also easily achievable with high accuracy using a single inertial measurement unit.

VI. CONCLUSIONS

This article describes an approach for monitoring activities performed by a worker in order to be able to instruct the worker in a specific task at a time or to ensure that all the needed tasks are performed before the product is sent forward on a manufacturing assembly line. The k nearest neighbor method was used to differentiate hammering, screwing with a screwdriver, spanner use, screwing with a power drill and null data. As it was shown, overall accuracy of 88.2% was achieved with online recognition using tri-axial acceleration and angular speed data from a single inertial measurement unit. Online recognition was carried out with a sliding window method comprised of a new classification every half a second. The results proved that the approach is suitable for the problem, as the recognition information can be utilized in concrete industrial tasks and thus benefit the industry.

ACKNOWLEDGMENT

This study was carried out with financial support from the Sixth Framework Programme of the European Community for research, technological development and demonstration activities in an XPRESS (FleXible Production Experts for reconfigurable aSSembly technology) project. It does not necessarily reflect the program's views and in no way anticipates the Commission's future policy in this area. In addition, the authors would like to thank Infotech Oulu and GETA graduate schools.

REFERENCES

[1] J. Ward, P. Lukowicz, and G. Tröster, "Gesture spotting using wrist worn microphone and 3-axis accelerometer," in *Proceedings of Soc-Eusai '05*, Oct. 12-14 2005.

[2] J. A. Ward, P. Lukowicz, G. Tröster, and T. E. Starner, "Activity recognition of assembly tasks using body-worn microphones and accelerometers," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 10, pp. 1553–1567, 2006.

[3] M. Mathie, A. Coster, N. Lovell, and B. Celler, "Accelerometry: Providing an integrated, practical method for long-term, ambulatory monitoring of human movement," *Physiological Measurement*, vol. 25, pp. R1–R20, 2004.

[4] ——, "Detection of daily physical activities using a triaxial accelerometer," *Medical and Biological Engineering and Computing*, vol. 41, no. 3, pp. 296–301, 2003.

[5] M. Ermes, J. Pärkkä, J. Mäntyjärvi, and I. Korhonen, "Detection of daily activities and sports with wearable sensors in controlled and uncontrolled conditions," *IEEE Transactions on Information Technology in Biomedicine*, vol. 12, no. 1, pp. 20–26, 2008.
[6] M. Puhakka, P. Laurinen, and J. Röning, "Time series visualization method with an augmenting audiovisual stream," submitted to IEEE Information Visualization Conference.
[7] E. Fix and J. L. Hodges Jr., "Discriminatory analysis - nonparametric discrimination: Consistency properties," *Technical Report 4, U.S. Air Force, School of Aviation Medicine, Randolph Field, TX*, 1951.
[8] T. Mitchell, *Machine Learning*. The McGraw-Hill Companies, Inc., 1997.
[9] T. Stiefmeier, D. Roggen, G. Tröster, G. Ogris, and P. Lukowicz, "Wearable activity tracking in car manufacturing," *IEEE Pervasive Computing*, vol. 7, no. 2, pp. 42–50, 2008.