Movement Recognition with Intelligent Multisensor Analisys, a Lexical Approach

Movement Recognition with Intelligent Multisensor Analisys, a Lexical Approach.

Stefano PINARDI, Roberto BISIANI

Department of Computer Science, University of Milano-Bicocca.

pinardi@disco.unimib.it; bisiani@disco.unimib.it

Abstract. The analysis of movements using inertial sensors represents an interesting alternative to video cameras, or other instrumentation used in posture analysis (treadmills, force plates, pressure plates, EMG). Inertial-sensor based analysis has been shown to be useful to classify Activities of Daily Living for situation assessment, healthcare applications, or to understand human emotions from body posture. We classify movements using a "lexical-like" approach. We use a vector representation of movements using a technique able to extract a great number of generic features, and a methods of classification, inspired by text mining, and machine learning techniques with some modifications, that transform our vector space from the feature-value space into a feature-frequency space. We used this method to classify a set of 21 movements performed by 13 people with good recognition results. Then we tested our method on the public WARD 1.0 database outperforming the results presented in literature on that database. The method we describe also shows to be technologically independent and semantically scalable, uses fast algorithms and appears to be suitable for every practical application where runtime movement analysis with big dictionaries could be a key factor.

Keywords. Action recognition, wearable sensors, similarity measures, ranking algorithms, lexical approach, movement semantic, public database of movements.

Introduction.

Movement recognition with inertial sensors proved to be useful for social surveillance applications [2], in neuroscience [5] and for tracking activities [12,13]. Inertial sensors can also be used for sport analysis, for gait and posture analysis, for human computer interaction and in motion recognition and capture [9,10,12,14,16]. To classify movements with inertial sensors could also be an important step to recognize human emotions from body movements and posture [6,7].

On the one hand, inertial sensors require a certain amount of user cooperation and could be considered invasive and cumbersome. On the other hand, since hardware is becoming smaller and smaller the user acceptability of body-worn sensors has improved and will continue to improve. Moreover, inertial sensors have many advantages since they can be directly placed on specific body segments or in clothes

accessories. This has many implications: we know with certainty to which segment of the body the data collected by the sensors refers to, we do not have to solve "hidden parts" problems created by video cameras, nor solve color and luminance issues. Also, we do not have to interpret/understand the surrounding environment, for example separate the body information from the background information, and identify people [3,4,15].

In this work we focus our attention on movement recognition with inertial sensors for movement classification, using a generic method, semantically flexible, and technological independent. Our method was tested with two very different technologies and vocabularies of actions and in all cases performed with good accuracy.

1. State of the Art.

Many different approaches were used in the movement recognition area with inertial sensors, we can roughly divide them in two different kinds: i) the approaches where researchers used a specific set of features that have heuristically been proven to be suitable for characterizing a chosen set of movements; ii) the approaches where machine learning techniques are used to recognize a movement [12]. Furthermore, other techniques interpret a movement as a sequence of hidden states utilizing a Hidden Markov Model to predict movement from observables [11]. Many different technologies and sensors have been used, both in quality and quantity. Some prefer to use many mono dimensional sensors [12], others a single device mounted in a specific place of the body [2,12]. Others prefer multimodal approaches conjunctly using audio and inertial sensors [9,11].

Recently, a new technique was proposed by A.Y. Yang et al. of the University of Berkeley called Distributed Sparsity Classifier [1]. For this work a public database of movement has been made available, the WARD 1.0 database.

2. The Sensors Architecture.

We used an "ad hoc" architecture and a specific instrumentation to develop and test our method. In particular, we used five MTx inertial sensors of XSens [16]. Each MTx sensors is provided with three devices: an accelerometer, a gyroscope, and a magnetometer; each device has three degrees of freedom, providing information on acceleration (+/- 50 m/sec2), rate of turn (+/-2 rad/sec), and earth-magnetic field (+/-1 normalized) in a three-axial reference system. The sample rate is 50 Hz.

3. Feature Extraction.

Each device – accelerometer, gyroscope, and magnetometer – yields three dimensional data (X, Y, Z). Every datum is also considered in its 2D and 3D norm representation (|XY|, |XZ|, |YZ|, |XYZ|). Subsequently, data is filtered with eight transformation functions (null, smoothing, low pass, mean, variance, variance with low pass, first derivative, second derivative) generating 840 transformations of the original signals. Then, 10 generic features are chosen for each transformation, generating 8400 features.

Feature values are then quantized into 22 intervals for a total of 184.800 intervals. When hit, a specific interval is marked 1, otherwise is left to 0. Hence, every action generates a sparse vector of 184.800 binary values (see Figure 1).

4. Actions-Vocabulary Analysis.

Some features are more frequent within the population, others can be less frequent inside the vocabulary's actions. In order to take into account this aspects, two weights have been introduced: the FF ("Feature Frequency") and the IVFF ("Inverse vocabulary frequency"). Feature Frequency is calculated using the following formula:

$$\mathrm{Ff}_{i,j} = \frac{n_{i,j}}{|P|} \qquad (1.)$$

where $n_{i,j}$ is the number of occurrences of the σ_i feature in the action a_j, and $|P|$ represents the population cardinality. Inverse vocabulary frequency weights features according to their "discriminatory" ability within the dictionary they belong to. Its formula is:

$$\mathrm{IVFf}_i = \log \frac{|A|}{|\{a : \sigma_i \in a\}|} \qquad (2.)$$

where $|A|$ represents the cardinality of the vocabulary, and $|\{a : \sigma_i \in a\}|$ the number of actions a where feature σ_i assumes values. The overall weight of a single feature is given by the multiplication: $W_{i,j} = \mathrm{Ff}_{i,j} * \mathrm{IVFf}_i$.

The FF and IVFF formulas transform the vector space. The FF takes into account how frequent is a feature in the given population rising the importance of the features that appear in the same class of movements (Eq.1). The IVFF takes into account how frequent is a feature in the dictionary. A feature that is present in more actions is considered less discriminative, and its weight is lowered according to the formula (Eq. 2.). We have to note that distances in the feature-frequency space could be very different than in the feature-values space: some dimensions can be canceled or enhanced depending on the role of features in the dictionary.

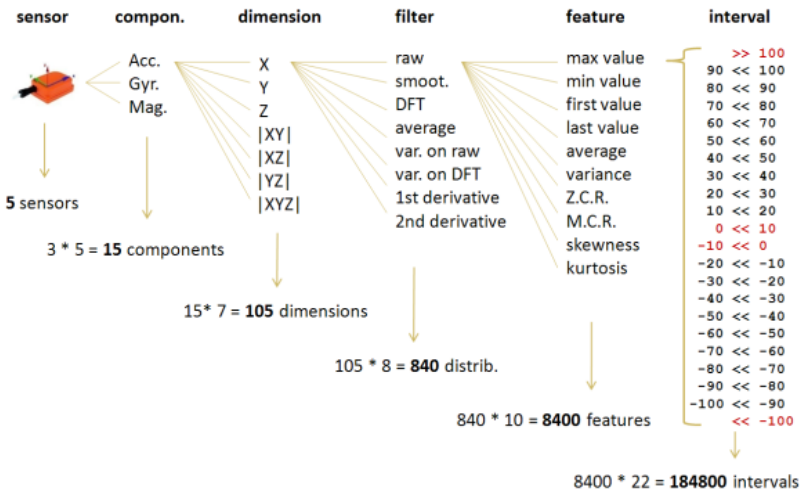

Figure 1. Feature extraction process using an iterative extraction operation.

5. The Evaluation Phase.

Four our first test we used a Test Set called NIDA (Nomadis Internal Database of Actions) of 21 different actions done by 7 different subjects; each action was repeated twice by all subjects but one, for a total of 273 actions. This set is larger and more varied then most test sets we found in literature. Then a more extended test was done on another database, the WARD 1.0.

Once a set of actions suitable for recognition has been defined and samples collected, they are placed into the Feature-Action space and transformed by FFxIVFF during the training phase. Then the recognition phase begins. In order to recognize which action is the most similar to a given action, we measure which is the closest one inside the Feature-Action space using three classification algorithms: a Ranking algorithm (Eq.3), an Euclidean Distance (Eq.4), and a Cosine Similarity (Eq.5). We also used a "Majority Classification" that selects the action "called" by the majority of the three methods. The formulas are the followings:

$$\text{rank}_j = \sum_{i=1}^{n} W_{i,j} \qquad (3.)$$

$$\text{dist}_i = \sqrt{\sum_{j=1}^{n} \left(W_{i,j} - q_{i,j}\right)^2} \qquad (4.)$$

$$\cos\theta = \frac{W_{i,j} \cdot q_{i,j}}{|W_{i,j}||q_{i,j}|} \qquad (5.)$$

where $W_{i,j}$ represents the weight of the σ_i interval of action a_j of the Training-Set, and $q_{i,j}$ is the IVFF value associated to the feature of query.

6. Databases NIDA and WARD

The NIDA 1.0 (Nomadis Internal Database of Actions) database contains movements acquired by the NOMADIS Laboratory of the University of Milano-Bicocca. These acquisitions have been obtained using 5 MTx sensors positioned on the pelvis, on the right and left wrist, and on the right and left ankle. NIDA includes 21 types of actions performed by 7 people (5 males and 2 females) ranging from 19 to 44-years-old, for a total of 273 actions. The database has a rich vocabulary: it contains both the typical movements of Daily Living, and actions like "karate punch", "karate frontal kick", "karate side kick". The complete list is the following:

1. Get up from bed. 2. Get up from a chair. 3. Open a wardrobe. 4. Open a door. 5. Fall. 6. Walk forward 7. Run. 8. Turn left 180 degrees. 9. Turn right 180 degrees. 10. Turn left 90 degrees. 11. Turn right 180 degrees. 12. Karate frontal kick. 13. Karate side kick. 14. Karate punch. 15. Go upstairs. 16. Go downstairs. 17. Jump. 18. Write. 19. Lie down on a bed. 20. Sitting on a chair 21. Heavily sitting on a chair

WARD 1.0 (Wearable Action Recognition Database) was collected at UC Berkeley. Acquisitions have been obtained positioning 5 sensors on the pelvis, on the right and left wrist, and on the right and left ankle [1]. Each sensor contained a 3-axial accelerometer and a 2-axial gyroscope; magnetometers were not present. Data have been calibrated and normalized to their appropriate unit of measure before using them for the training phase. WARD contains 13 types of actions performed by 20 people (7 women and 13 men) ranging from 20 to 79-years-old with 5 repetition per action, for a total o 1200 actions. The complete list of actions is the following:

1.Stand (ST). 2. Sit (SI). 3. Lie down (LI). 4. Walk forward (WF). 5. Walk left-circle (WL). 6. Walk right circle (WR). 7. Turn left (TL). 8. Turn right (TR). 9. Go upstairs (UP). 10. Go downstairs (DO). 11. Jog (JO). 12. Jump (JU). 13. Push wheelchair (PU).

7. Testing techniques

We used a Leave One Out Cross-Validation (LOOCV) method to calculate accuracy. We also used Majority Voting combinations that must be intended as an extension of the LOOCV test, which carries out the results of Majority Voting among all classifiers, varying each time the preference given to a classifier in case of tie.

The classification accuracies of the algorithms using the NIDA database (273 actions of 21 type) are the followings: Ranking 89.74%, Euclidean Distance 95.23%, Cosine similarity 95.23% , Majority Voting 94.7%.

The classification accuracies of algorithms using the WARD database are the followings: Ranking 97.5%, Euclidean Distance 97.74%, Cosine 97.63% , Majority Voting 97.79%.

We give results of the Cosine Similarity also in a synoptic way with a Confusion Matrix (see Figure 2). The columns contain the ground truth, while the rows contain the results of our classification algorithm.

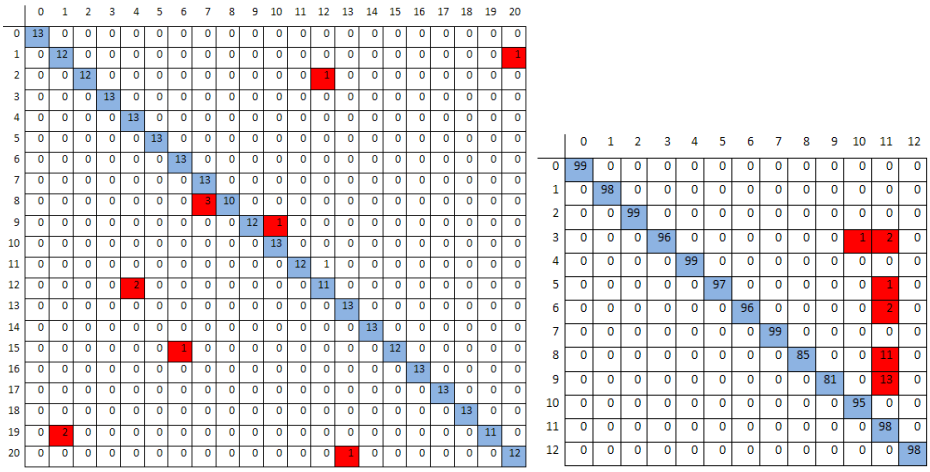

Figure 2. Confusion Matrix. NIDA Cosine Similarity: accuracy 95.23 % (left). WARD Cosine Similarity: accuracy 97.63% (right).

8. Tests results

The tests show that the single classifiers with the highest performing rate on both databases is the Cosine Similarity: the accuracy is 95,23% on the NIDA database and 97,63% on the WARD database. Majority voting gives an accuracy of respectively 94.7% and 97.79%.

To correctly compare the performances of the NIDA and WARD databases we have to weigh the relative accuracies considering the different dimension of the dictionaries. An algorithm that chooses randomly could have an accuracy that is roughly the inverse of the dimension of the dictionary. We weigh the given accuracy with this factor in order to confront the accuracy obtained on both databases. We calculate the ratio with the following formulas :

$$\left.\begin{aligned} N_1 &= Acc_W * |D_w| = 1269.19 \\ N_2 &= Acc_N * |D_N| = 1999.83 \end{aligned}\right\} \quad \frac{N_2}{N_1} = 1.58$$

where AccW refers to WARD accuracy and AccN refers to NIDA accuracy and |Dw| |DN| are the respective cardinality of the dictionaries. We see that the ratio gives us a results of approximately 1.58 in favor of the NIDA database. As NIDA's actions are more numerous – and also more difficult to be discriminated (like for "karate frontal kick" and "karate side kick") – than WARD's dictionary of actions we could say that the accuracy obtained with the NIDA database is higher.

We also tested the algorithms performance using three out of five sensors both for the WARD and the NIDA database, in order to understand the sensitivity of the method to the number and the placement of sensors on the body.

Using cosine similarity with the NIDA database, using 3 sensors on the pelvis, the right wrist and right ankle, we obtain a 93.40% accuracy. With 3 sensors on pelvis, left

wrist and ankle we obtain 92.30% accuracy. We also get interesting results by positioning 3 sensor "diagonally": on the pelvis, the right wrist and left ankle reaching a 94.13%; with 3 sensors on the pelvis, the left wrist and right ankle we obtain 93.77%.

Using cosine similarity with WARD database using 3 sensors on the pelvis, the right wrist and ankle give an accuracy of 97.63%. We obtain an identical accuracy by using 3 sensors on the pelvis, the left wrist and ankle (97.63%). Again, we obtain interesting results by positioning 3 sensors diagonally on pelvis, right wrist and left ankle reaching 96.69% and pelvis left wrist and right ankle reaching 97.48%, just 0.15 % less than accuracy obtained using 5 sensors. Note that we have reached a better accuracy of A.Y.Yang et al. [1] just using three out of five sensors on their database with the same data.

9. Conclusions

A representation of movements as a vector in the relevance of feature space, resembling methods used in text classification with some important modification seems to performs well. Similar actions are discriminated by our method both in a big database (WARD), or with a big dictionary (NIDA), accordingly to the hypothesis that a "lexical-like" approach is well suited for action recognition with inertial sensors.

The results show that this classification method is reliable and does not depend on technology or specific features; also, it does not require any specific "a priori" or biomechanical knowledge about the given movements. Consequently, we do not depend on the application domain technology and we can change and improve the dimension of the vocabulary at will with satisfactory results. As far as we could find, the dimension of these dictionaries are greater than all examples given in the literature. Also, our method does not have a strong dependency on the position of sensors or on their number, and we have a good accuracy using just 3 sensors on the WARD database (having only 3 accelerometers, and 2 gyroscopes per sensor) outperforming similar results present in literature on the same database. Other Machine Learning techniques can be used to classify a greater range of movements, to understand feature dependencies, and to analyze the quality of the movements, extending the method to the domain of gait and posture analysis, and to the area of recognizing human emotion from body movements and postures.

References.

[1] A.Y. Yang, R. Jafari, S.S. Sastry, and R Bajcsy, Distributed Recognition of Human Actions Using Wearable Motion Sensor Networks, *Journal of Ambient Intelligence and Smart Environment*s, 2009.

[2] A. Mileo, D. Merico, S. Pinardi, and R. Bisiani, A Logical Approach to Home Healthcare with Intelligent Sensor-Network Support, The Computer Journal, 2009

[3] J. Cameron, J.Lasenby, Estimating Human Skeleton Parameters and Configuration in Real-Time from Markered Optical Motion Capture, AMD08, 2008

[4] R. Okada, B. Stenger, A Single Camera Motion Capture System for Human-Computer Interaction, ICICE(E91-D), No. 7, pp. 1855-1862, July 2008

[5] S. Ohgi, S. Morita, K.K. Loo, and C. Mizuike, Time Series Analysis of Spontaneous Upper-Extremity Movements of Premature Infants With Brain Injuries, PHYS. THER., Vol. 88, No. 9, September 2008, pp. 1022-1033.

[6] Kleinsmith, A. and Bianchi-Berthouze. Recognizing Affective Dimensions from Body Posture, Proc. ACII07, Lecture Notes In Computer Science, vol. 4738. Springer-Verlag, Berlin, Heidelberg, 48-58, 2007

[7] Castellano, G., Villalba, S. D., and Camurri, Recognising Human Emotions from Body Movement and Gesture Dynamics, Proc. ACII07, Lecture Notes In Computer Science, vol. 4738. Springer-Verlag, Berlin, Heidelberg, 2007

[8] G. Guerra-Filho, Y.Aloimonos, A language for Human Action, IEEE Computer Magazine, 40:60–69, 2007

[9] J. Lester, T.Choudhury, and G. Borriello, A Practical Approach to Recognizing Physical Activities, Pervasive 2006, LNCS 3968, pp. 1 – 16, 2006

[10] E. F. Desserée, Calais, L. R. Legrand, First Results of a Complete Marker-Free Methodology for Human Gait Analysis, Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference Shanghai, China, September 1-4, 2005

[11] T. Choudhury, N. Kern, G. Borriello, B. Hannaford, and J. Lester, A Hybrid Discriminative/Generative Approach for Modeling Human Activities, Proceedings of the Nineteenth IJCAI, pages 766 - 722, Edinburgh, Scotland, 2005.

[12] L. Bao, Physical Activity Recognition from Acceleration, Department of Electrical Engineering and Computer at the Massachusetts Institute of Tecnology, Master Thesis, August 2003.

[13] S. Lee and K. Mase, Activity and location recognition using wearable, IEEE Pervasive Computing, 2002.

[14] J. Himberg, and T. Seppanen J. Mantyjarvi, Recognizing human motion with multiple acceleration sensors, Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, 2001.

[15] T.B. Moeslund, E. Granum, A Survey of Computer Vision-Based Human Motion Capture, Computer Vision and Image Understanding, 2001

[16] XSens. http://www.xsens.com.