What are you doing while answering your smartphone?

Andrea F. Abate[1], Silvio Barra[2], Maria De Marsico[3], Michele Nappi[1]

[1] *Department of Computer Science, University of Salerno, 84084, Salerno, Italy*

[2] *Department of Mathematics and Computer Science, University of Cagliari, 09124 Cagliari, Italy*

[3] *Department of Computer Science, Sapienza University of Rome, 00185 Roma, Italy*

ICPR2018

In Beijing, China, August 20th-24th 2018

ICPR will be an international forum for discussions on recent advances in the fields of Pattern Recognition, Machine Learning and Computer Vision, and on applications of these technologies in various fields.

Abstract

- Ambient Intelligence requires Context Awareness
- Use of smartphones: increasing number and types of sensors to record ambient and user events
- ***From Action Recognition to Action Context Recognition***

System Outline and Dataset

The prototype system is implemented on Android OS.

- When a phone ring arrives, the user is expected to perform the arm gesture of lifting the phone. The phone ring activates the frontal phone camera and the acquisition of dynamic movement data. The camera acquisition stops as soon as a ear image is detected.
- The arm movement acquisition pertains **accelerometer** and **gyroscope** signals, each represented as three time series over X, Y, and Z coordinates.
- A sampling frequency of $50ms$ has been chosen to read the sensor values. A maximum of 400 triplets is acquired

The dataset consists of 129 acquisitions from 38 subjects.

- All of the acquisitions include the signals captured by accelerometer, gyroscope and GPS during the arm gesture performed to answer a phone call.
- Acquisitions were carried out in 4 states: standing, sitting, walking and running.
- Only smartphone motion data (accelerometer and gyroscope) have been used for experiments: the combination with ear recognition is left for a future development phase.

Classifiers

- **Fine Gaussian SVM**:nonlinear SVM classifier with Gaussian kernel function; the specific option used (fine) provides high model flexibility, and allows making finely detailed distinctions between classes; it was used as a binary classifier.
- **Fine KNN** provides finely detailed distinctions between classes, with Euclidean metric and number of neighbors set to 1.
- **Medium KNN** and **Cubic KNN** provide medium distinction between classes. For both the number of neighbors is set to 10. While the first uses a Euclidean Distance metric, the second uses the Minkovski distance metric.
- **Weighted KNN** provides a medium distinctions between classes, using a distance weight and the number of neighbors set to 10; it was used with Euclidean distance, with Distance Weight equal to Squared Inverse.
- **Bagged Trees** provides a high model flexibility and uses Breiman's Random Forest algorithm [1]; it was used with 30 learners and Learning Rate 0.1.

Experiments' Settings

In the experiments reported below, the 3D corresponding points in the two sensor signals are combined into a 6D point. The space required by sensor signals storage is possibly reduced by a preprocessing step. It is worth underlining that this is not equivalent to setting up a coarser sampling rate. Dense data are rather processed and summarized. Accuracy is defined as the percentage of the correct recognition of the action per the total number of states of the same person. The threshold value t is experimentally set to 0.2.

The Signals

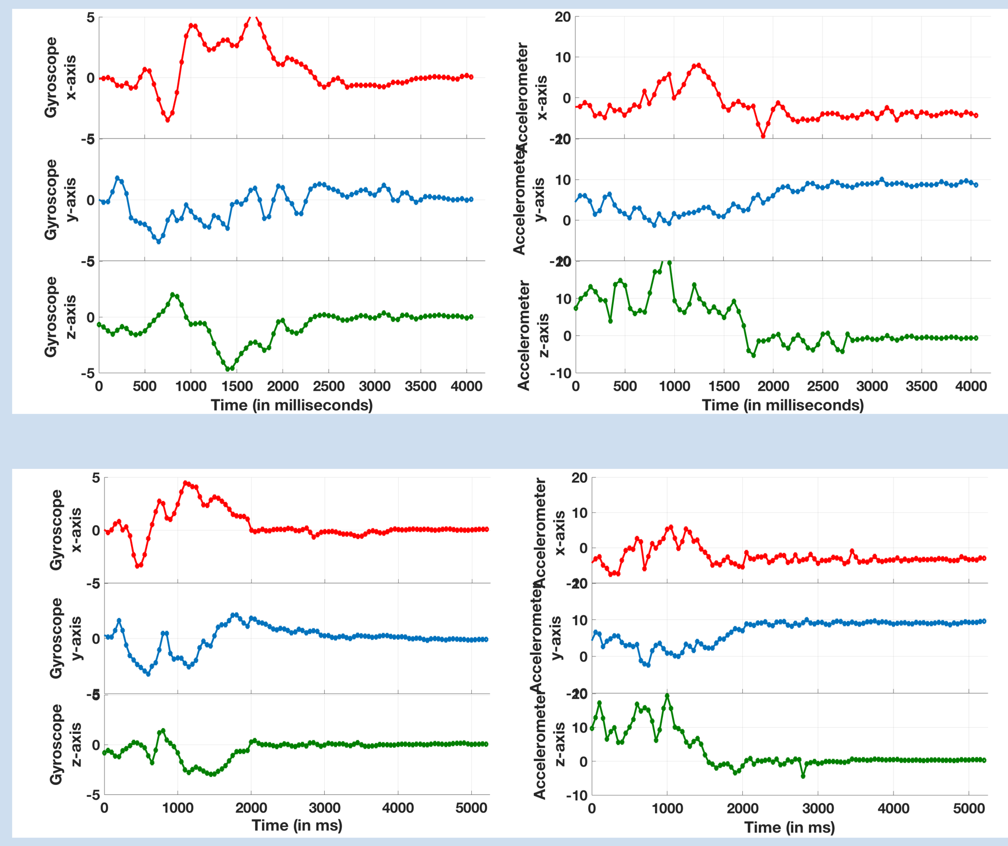


Figure 1. In the topmost, the recording related to the gyroscope (the leftmost plot) and the accelerometer(the rightmost) for the *standing* state; in bottommost image, the plots describe the same for the *sitting* state. The signals on the x, y, and z axes are plotted in red, blue and green respectively. Time is expressed in milliseconds. .

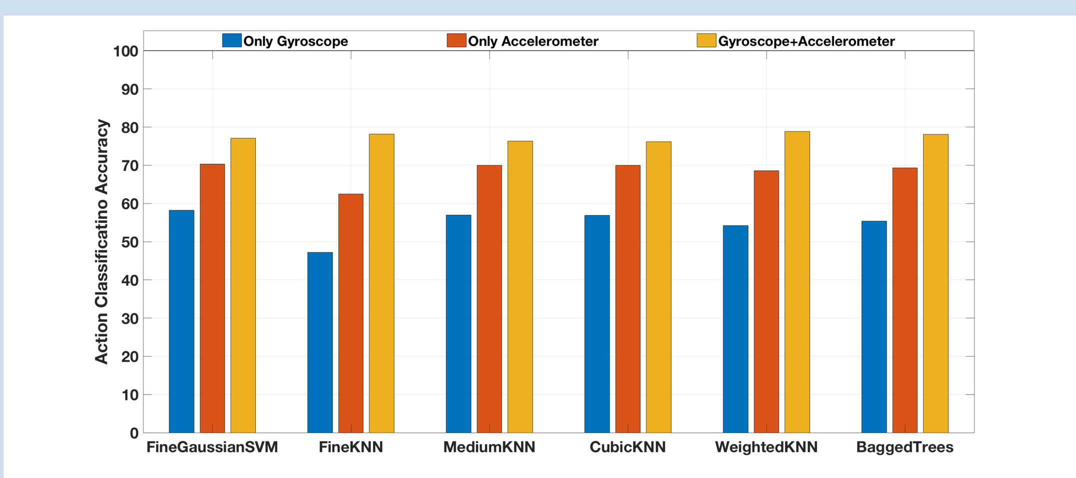
Data Reduction Strategies

DRS 1: $\forall$ a (either ACC. or GYR. signal) composed of N action points in 3D space (considering x,y,and z coordinates), the Euclidean distance $ED(xyz(i), xyz(i+1))$. If $ED < t$, then $xyz(i)$ and $xyz(i+1)$ are added to a set P. As soon as a point $j+1$ following a point j is found, such that $ED > t$, the centroid c of points in P is computed; all points in P are substituted by c. The process restarts with an empty P until the whole signal has been processed. Classification is carried out using the new sets of points.

DRS 2: Instead of substituting each set of points with the centroid c, the point in the set with the minimum ED from c is rather used.

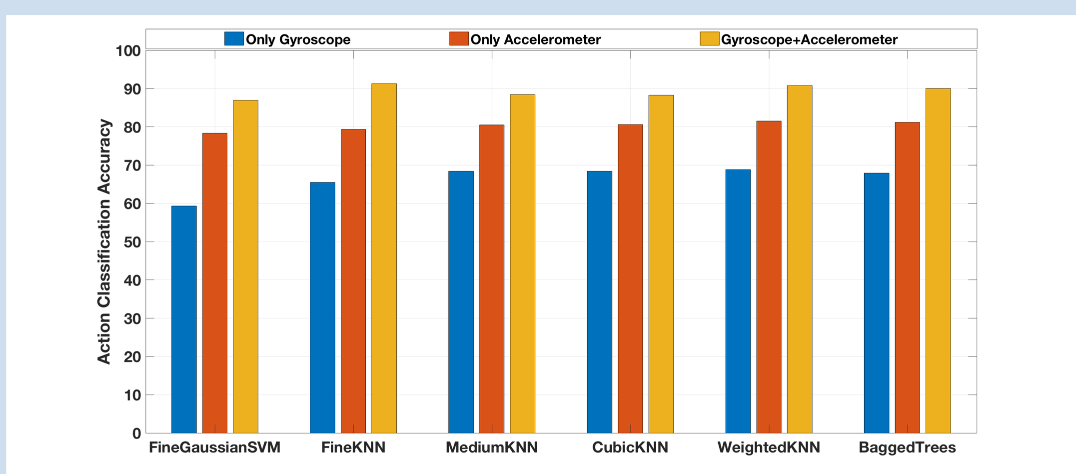
Experiment 1

No compression of sensor data was carried out. Each action was classified by using raw information.



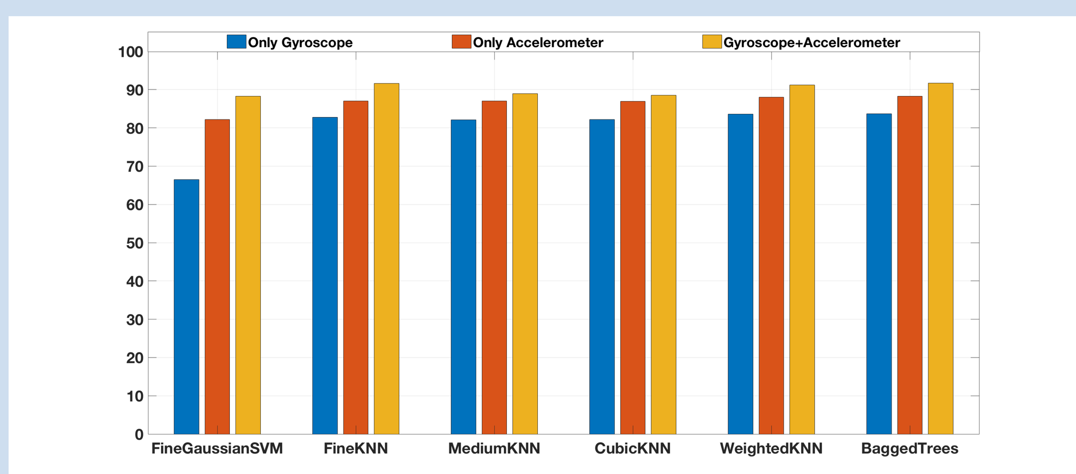
Experiment 2

The amount of data for each action was reduced according to the Data Reduction Strategy number 1 (previously named **DRS 1**)



Experiment 3

The amount of data for each action was reduced according to the Data Reduction Strategy number 2 (previously named **DRS 2**)



Results' Discussion and Conclusion

This work presents a preliminary investigation on action state recognition by the embedded sensors of the user smartphone. Differently from works in literature that recognize the general bodily state [2, 3](e.g., standing, sitting, walking, or running) the proposed approach carries out recognition during a determined action, to provide a more complete context awareness. The complete system will also include user, location, and speed recognition, to allow a more detailed adaptation of the services triggered in the smart ambient. The achieved results are related to different learning techniques and different data configurations, with and without compression. Data cleaning results in a significant improvement of the accuracy. The present results demonstrate the general advantage of using both sensors, but further experiments are needed to get a deeper insight into the possibilities of the proposed approach.

References

[1] Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.
[2] Schmidt, A. (2000). Implicit human computer interaction through context. Personal technologies, 4(2-3), 191-199.
[3] Abate, A. F., Nappi, M., & Ricciardi, S. (2017). I-Am: implicitly authenticate me person authentication on mobile devices through ear shape and arm gesture. IEEE Transactions on Systems, Man, and Cybernetics: Systems, (99), 1-13.