

# A Probabilistic Approach to Handle Missing Data for Multi-Sensory Activity Recognition

**Hesam Sagha**  
EPFL STI CPN CNBI  
1015 Lausanne, Switzerland  
hesam.sagha@epfl.ch

**José del R. Millán**  
EPFL STI CPN CNBI  
1015 Lausanne, Switzerland  
jose.millan@epfl.ch

**Ricardo Chavarriaga**  
EPFL STI CPN CNBI  
1015 Lausanne, Switzerland  
ricardo.chavarriaga@a3.epfl.ch

## ABSTRACT

Context and activity recognition in complex scenarios is prone to data loss due to disconnections, sensor failure, transmission problems, etc. This generally implies significant changes in the recognition performance. In the case of classifier fusion architecture faulty sensors can be removed from the recognition chain to overcome this issue. Alternatively, we can try to compensate or impute data to replace the missing signals. In this paper we proposed a probabilistic method for imputation of missing data based on conditional Gaussian distribution. Our method exploits the correlation among classifier outputs to infer missing values in a probabilistic manner. We assess the method performance using two datasets (car manufacturing and a daily activities scenarios) with three different configurations of sensors. Results show the advantages of the probabilistic estimation at the classifier decision level.

## Author Keywords

Activity recognition, missing values, multiple sensors, classifier fusion, data reconstruction.

## ACM Classification Keywords

L.7.0 Wireless/Pervasive Computing: Miscellaneous

## General Terms

Algorithms, Performance, Reliability.

## INTRODUCTION

Sensor networks are prone to data loss due to disconnections, sensor failure, and transmission problems. This is particularly relevant for wearable and wireless sensors deployed in real-life scenarios. Indeed, as sensing devices as well ambient intelligence environments are more and more available the common assumption of a known, static, well characterized sensor configuration is less and less valid. Ideally, in activity recognition scenarios, as persons perform their daily activities, they can move across environments with different sensor configurations, moreover on-body sensors can fail or

suffer from communication problems. In order to cope with these issues, activity recognition systems should be robust enough to recover from these changes without requiring re-training or calibration processes; a quality some groups have termed *Opportunistic* [8]. They define Opportunistic systems where sensors can be added or removed from the network at any time, without drastic changes in the recognition performance.

The problem of classification with missing data has been extensively studied in the fields of machine learning, speech recognition and microarray data, but less studied in the case of activity recognition. For speech recognition applications, there exists a wealth of studies on spectrogram missing (unreliable) data estimation [3, 6] and there is a comparison study of statistical value imputation using conditional and marginal distributions with a Gaussian classifier [4]. Similarly, several methods have been proposed to handle missing gene expression when analyzing micro-array data [2]. In addition, Saar-Tsechansky et al. provide a survey of different methods for handling missing data when classifiers take the responsibility of missing values, such as C4.5 trees [9]. In activity recognition applications, high level predictive tree-structures have been proposed for classification using available deficient data and the last available complete data classification result [1]. Setz et al. provided a case study of missing data in emotion recognition using physiological signals and they found that reduced-feature models are competitive or even slightly better than mean-value imputation [5]. A distance measure is computed between the current available sensor values and patterns previously stored in a ‘pair dataset’. In the case of missing data the closest pattern is imported. The performance of their system fully relies on the quality and the size of the dataset.

The rationale of these methods can also be applied for continuous signals from networked sensor systems. In this case, missing sensor readings can be estimated based on previous samples or a-priori knowledge about the system. Once the missing values have been imputed, the classification process is performed normally. Alternatively, in the frame of classifier fusion –where decisions are made by combination of several classifiers, each of them corresponding to a subset of sensors– the classifier of the faulty sensor can be taken out from the fusion process. On the other hand, instead of estimating the signal values, the method can be used to estimate the output of the corresponding classifiers based on the correlation between them.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

*UbiComp '10*, Sep 26-Sep 29, 2010, Copenhagen, Denmark.

Copyright 2010 ACM 978-1-60558-843-8/10/09...\$10.00.

In this paper we proposed a probabilistic method based on conditional Gaussian distribution to estimate the values of the classifier outputs correspond to disconnected sensors in a multisensor scenario. Method can be applied with any type of classifiers that provide class probability and is independent of the fusion technique. We compare performance over two databases on realistic activity recognition scenarios for car manufacturing and daily home activities with different configuration of sensors. In the next section we present the imputation method then we present the simulation of the methods and discuss these results and its implications in the design of robust activity recognition chain.

## METHODS

Our aim is to handle missing values during activity recognition in multi-sensory networks. We assume that an ensemble of classifiers has been previously trained, and each classifier corresponds to an individual or a subset of sensors. This approach allows for the modular design of the ensemble with respect to the type and number of sensors and it does not depend on the type of classifiers used. There are three levels for handling missing values in such a network; raw data, feature level and classifier. Estimation of missing data at the level of the raw data and feature level based on the available channels may face problems in multimodal scenarios since channels may not lie on the same space. The proposed method acts at the last level, where classifiers have made their decisions and values are in the same space (they all correspond to class probabilities). Here, we suppose that each classifier has the ability to provide a vector with the probability of the input pattern to belong to each class. These vectors form a  $N \times C$  matrix called Decision Profile (DP), where  $N$  is the number of classifiers and  $C$  is the number of classes. In the case of data missing from one or more sensors, the corresponding classifier yields no output and we want to estimate the missing vector of probabilities,  $DP^{Test}(miss)$ , using the available ones,  $DP^{Test}(av)$ , based on the correlation between each element. To this end, a Gaussian distribution is estimated using the DPs of training data, and the missing values are replaced by the mean values of the conditional distribution, taking into account the current values of the available vectors. To do so, first we re-order DP to be a column vector  $\overline{DP}$  of size  $N * C$ , and we estimate the covariance between each element. Then, when encountering missing values, we infer and impute the mean  $\mu_{a|b}$  of conditional distribution,

$$\overline{DP}^{Test}(miss) = \mu_{a|b} = \mu_a + \Sigma_{ab}\Sigma_{bb}^{-1}(x_b - \mu_b) \quad (1)$$

where  $\mu$  and  $\Sigma$  are the mean vector and covariance matrix of data computed on training data.  $\mu_a$  and  $\mu_b$  are subvectors of  $\mu$  and are mean values corresponding to missing values and available values, respectively.  $x_b$  is the vector of values of the available classifiers,  $\overline{DP}^{Test}(av)$ ,  $\Sigma_{ab}$  is the covariance between missed and available values, and  $\Sigma_{bb}$  is the covariance between available values. Hereafter we call the method *Decision Profile using Conditional Distribution (DPCD)*. When a sensor is removed, the inverse matrix  $\Sigma_{bb}^{-1}$  should be computed once and the computation cost for each pattern is three matrix multiplication with the sizes  $N_a \times N_b, N_b \times N_b$  and  $N_b \times 1$ , plus one matrix addition with

size  $1 \times N_a$  and one matrix subtraction with size  $N_b \times 1$  where finally results to  $O(N_a N_b^3)$ .

We evaluate the proposed method by comparing it with the methods described below.

1. *Removal*, removing the classifiers associated with the faulty sensors from the classifier fusion process.
2. *Cluster*, we enhanced the method described in [5] by clustering all training DPs using K-means, and while testing we replace the values from the nearest cluster into the missing part of the DP. The Euclidean distance is computed based on the available part of DP.
3. *Raw Data Conditional Distribution (RDCD)*, in contrast with the mentioned methods, this method tries to estimate the missing raw signal. We used the same approach as DPCD at the raw data level; In this case the correlation between all the sensors are estimated.

## EXPERIMENT

We evaluate the classification performance of the different imputing methods on two activity recognition datasets, using several sensor configurations for each. For each case we design an ensemble of classifiers whose decisions are combined using a classifier fusion technique. Gaussian Mixture Model (GMM) with two units per class is chosen for classification and Dempster-Shafer is used as classifier fusion because empirically it is shown that they yields acceptable performance than other techniques on the datasets. For Cluster method we empirically set the number of clusters to 60.

The first dataset (*Skoda dataset*) corresponds to a car manufacturing scenario [10]. We use data from 8 subjects performing 10 recording sessions each (except one subject who recorded 8 sessions only). There are 20 classes corresponding to the activities like Open hood, Close hood, Open door, Close door and etc. The second dataset (*Opportunity dataset*) contains data for daily home activities in a breakfast scenario [7]. The data were recorded in a highly instrumented environment set up in a room with three doors, a kitchen and a table in the center. For the present simulations we performed classification based on 16 low-level actions for one subject. The subject performed activities in 5 sessions.

Each dataset contains different modules of sensors, e.g. 3-axis accelerometer, 3-axis rate gyro, 3-axis magnet sensors. For each simulation, we use different combination of the sensors, accelerometer, accelerometer + rate gyro, and accelerometer + rate gyro + magnet sensors. We group the sensors that are physically together into a package of sensors and set a classifier per package. This is more realistic since when the connection is lost the whole package is not able to transfer data. So for the first configuration, each classifier –corresponding to one package– uses data from one accelerometer (3 inputs), for the second configuration one accelerometer and one rate gyro correspond to each classifier (6 inputs) and finally one accelerometer, one rate gyro and one magnet sensor are associated to each classifier for the third configuration (9 inputs). The number of pack-

ages in the Skoda dataset is seven corresponding the sensors mounted on hand, torso, lower and upper arm of both hands. For the Opportunity dataset we use five packages mounted on back, lower and upper arm of both hands. Here we assume that there is an algorithm which differentiates actions and no-actions and just sends the data to classifiers when an action is detected. 30 repetitions have been done for each simulation and at each repetition we remove a number of packages randomly selected. To show the robustness against different sessions, data are evaluated based on one-session-out cross validation.

## RESULTS

Figures 1 and 2 illustrate the classification accuracy on Skoda and Opportunity datasets, respectively. The proposed method yielded consistently less performance degradation for both datasets and all sensor configurations, specially when a large number of packages fail. For the Skoda dataset DPCD and Cluster method have better performance with respect to Removal, while in the Opportunity dataset DPCD outperforms other methods. In contrast RDCD performs poorly on the Skoda dataset while for the Opportunity dataset up to some point it performs better than Removal. Overall, depending of the number of clusters for clustering method and consequent expected accuracy, DPCD method is more effective and has a low computational cost. Figures 3 and 4 illustrate the percentage of improvement using the proposed method with respect to Removal method. Fall in the ratio in the Opportunity dataset may be due to the low number of sensors in the network, which does not allow a reasonable estimation for other missed values.

In addition, estimation at the fusion level allows us to use a limited number of sensors in order to reduce the energy consumption in the network with graceful degradation of performance. As in Skoda dataset using only 4 packages out of 7 does not degrade dramatically the performance as well as using 4 out of 5 packages for Opportunity dataset.

## CONCLUSION

We have introduced a probabilistic approach for estimating missing data at the classification level in multisensory activity recognition systems. It exploits classifier output correlations to reliably estimate the missing inputs. The proposed method performs better than clustering decision profiles, removal of classifiers and raw data reconstruction. In fact, as non-homogeneous sensors may have a low correlation this may decrease the estimation accuracy at the raw data level; In contrast, at the fusion level all values are in the same space leading to a more accurate estimation. Finally, depending on the application, the trade-off between reconstructing missing data and performance improvement with respect to sensor removal should be investigated. Furthermore, we will also investigate the effect of using different fusion techniques after compensating missing vectors.

## ACKNOWLEDGEMENTS

This work has been supported by the EU Future and Emerging Technologies (FET) contract number FP7-Opportunity-225938. This paper only reflects the authors' views and

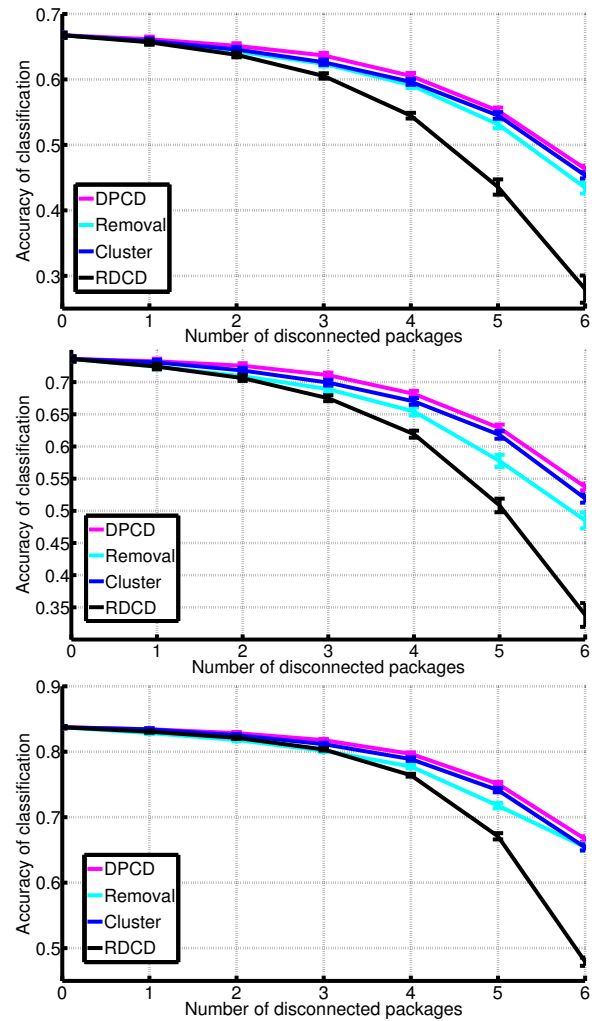


Figure 1. Result on Skoda dataset averaged on 7 subjects. Used packages: (top) accelerometer, (middle) accelerometer + gyro. (bottom) accelerometer + gyro + magnet sensors.

funding agencies are not liable for any use that may be made of the information contained herein.

## REFERENCES

1. I. Avdoueviski, R. Kerminen, J. Makkonen, and A. Visa. Missing values in user activity classification applications utilizing wireless sensors. In *MobiWac '08: Proceedings of the 6th ACM international symposium on Mobility management and wireless access*, pages 151–155, New York, NY, USA, 2008. ACM.
2. M. Celton, A. Malpertuy, G. Lelandais, and A. de Brevern. Comparative analysis of missing value imputation methods to improve clustering and interpretation of microarray experiments. *BMC Genomics*, 11(1):15, 2010.
3. C. Cerisara, S. Demange, and J.-P. Haton. On noise masking for automatic missing data speech recognition: A survey and discussion. *Computer Speech and*

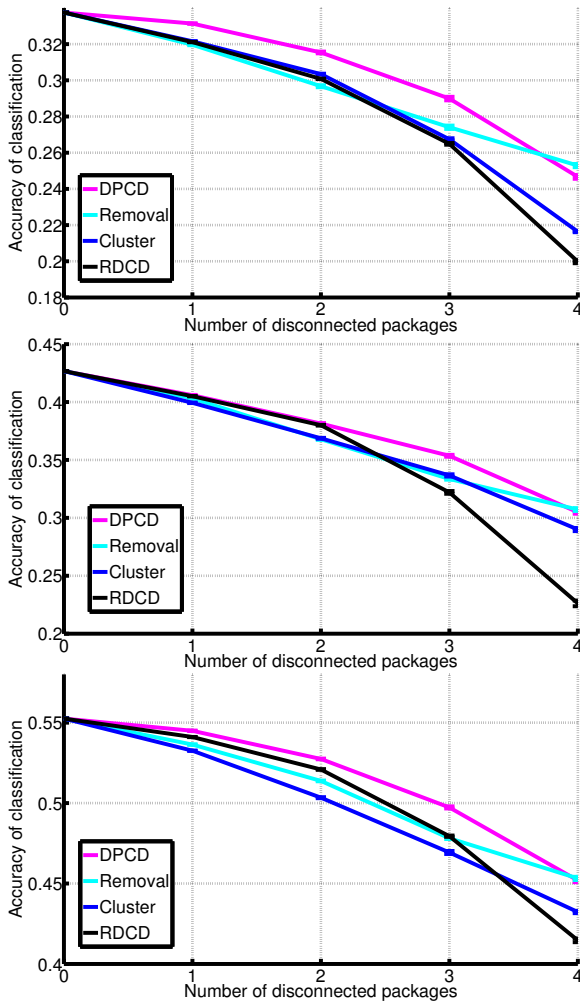


Figure 2. Result of simulations on Opportunity dataset. Used packages: (top)accelerometer, (middle) accelerometer+ gyro (bottom)accelerometer + gyro + magnet sensors.

Language, 21(3):443 – 457, 2007.

- M. Cooke, A. Morris, and P. Green. Missing data techniques for robust speech recognition. In *ICASSP*

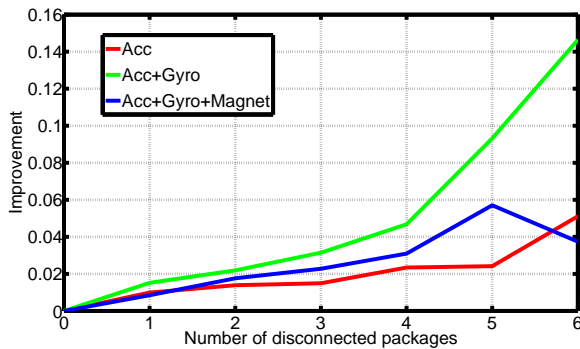


Figure 3. Percentage of improvement achieved with the proposed method with respect to *Removal* on Skoda dataset for three configurations.

'97: *Proceedings of the 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '97)-Volume 2*, page 863, Washington, DC, USA, 1997. IEEE Computer Society.

- K. Murao, T. Terada, Y. Takegawa, and S. Nishio. A context-aware system that changes sensor combinations considering energy consumption. In *Pervasive '08: Proceedings of the 6th International Conference on Pervasive Computing*, pages 197–212, Berlin, Heidelberg, 2008. Springer-Verlag.
- B. Raj, M. L. Seltzer, and R. M. Stern. Reconstruction of missing features for robust speech recognition. *Speech Communication*, 43(4):275 – 296, 2004.
- D. Roggen, A. Calatroni, M. Rossi, T. Holleczeck, K. Förster, G. Tröster, P. Lukowicz, D. Bannach, G. Pirkl, A. Ferscha, J. Doppler, C. Holzmann, M. Kurz, G. Holl, R. Chavarriaga, H. Sagha, H. Bayati, M. Creatura, and J. R. Millán. Collecting complex activity data sets in highly rich networked sensor environments. In *Seventh International Conference on Networked Sensing Systems*, 2010.
- D. Roggen, K. Förster, A. Calatroni, T. Holleczeck, Y. Fang, G. Tröster, P. Lukowicz, G. Pirkl, D. Bannach, K. Kunze, A. Ferscha, C. Holzmann, A. Riener, R. Chavarriaga, and J. Millán. OPPORTUNITY: Towards opportunistic activity and context recognition systems. In *Third IEEE WoWMoM Workshop on Autonomic and Opportunistic Communications*, 2009.
- M. Saar-Tsechansky and F. Provost. Handling missing values when applying classification models. *J. Mach. Learn. Res.*, 8:1623–1657, 2007.
- T. Stiefmeier, D. Roggen, G. Ogris, P. Lukowicz, and G. Tröster. Wearable activity tracking in car manufacturing. *IEEE Pervasive Computing*, 7:42–50, 2008.

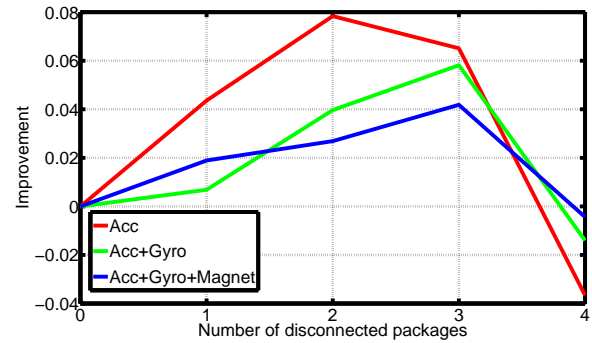


Figure 4. Percentage of improvement achieved with the proposed method with respect to *Removal* on Opportunity dataset for three configurations