

Face-to-Face Social Activity Detection Using Data Collected with a Wearable Device

Pierluigi Casale^{1,2}, Oriol Pujol^{1,2}, and Petia Radeva^{1,2}

¹ Computer Vision Center, Campus UAB, Edifici O, Bellaterra, Barcelona, Spain

² Dep. of Applied Mathematics and Analysis, University of Barcelona, Spain
pierluigi@cvc.uab.es

<http://www.cvc.uab.es>, <http://www.maia.ub.es>

Abstract. In this work the feasibility of building a socially aware badge that learns from user activities is explored. A wearable multisensor device has been prototyped for collecting data about user movements and photos of the environment where the user acts. Using motion data, *speaking* and other activities have been classified. Images have been analysed in order to complement motion data and help for the detection of social behaviours. A face detector and an activity classifier are both used for detecting if users have a social activity in the time they worn the device. Good results encourage the improvement of the system at both hardware and software level.

Keywords: Socially-Aware Sensors, Wearable Devices, Activity Classification, Social Activity Detection.

1 Introduction

Computation is packaged in a variety of devices. Nowadays, personal organizers and mobile phones are really networked computers. Interconnected computing devices using various sensing technologies, from simple motion sensors to electronic tags to videocameras are invading our personal and social activities and environments. This kind of technologies is moving the site and style of human-computer interaction from desktop environments into the larger real world where we live and act. Nevertheless, they are not yet able to understand social signaling and social context.

At MIT Media Labs, three socially aware communications systems incorporating social signaling measurements have been developed (Pentland, 2005 [2]). The *Uberbadge* is a badge-like platform, *GroupMedia* is based on the Sharp Zaurus PDA and *Serendipity* is based on the Nokia 6600 mobile telephone. In each system the basic element of social context is the identity of people in the users immediate presence. The systems use several sensors, including Bluetooth-based proximity detection, infrared or radio-frequency tags, and vocal analysis. At Microsoft Research, Hodges *et al.*, 2006 [3] presented a sensor augmented wearable stills camera, the *SenseCam*, designed to capture a digital record of

the wearers day, by recording a series of images and capturing a log of sensor data.

In this work, we explore the feasibility of building a socially aware device that learns from user’s activities. The novelty of the work consists in presenting a device able to acquire multi-modal data, recognize several activities and if a social interaction occurs. The wearable multisensor device, *TheBadge*, has been prototyped to collect data for detecting face-to-face social activities of people wearing it. Using data collected by a LIS3LV02DQ accelerometer, *speaking* and others coarse activities as “walking”, “climbing stairs” or “moving” have been classified. In addition, a face detector working on the photos automatically taken by a C628 Camera Module, is used in combination with the activity classifier for helping the detection of a social activity.

This document is organized as follows. In Section 2, a description of TheBadge at hardware level is given. In Section 3, the classification techniques, the overall classification system architecture and the single components of the system are explained. Section 4 describes which features have been used for classifying data and Section 5 report the results of the experiments we perform. Conclusions and improvements are discussed in Section 6.

2 TheBadge

TheBadge prototype consists of a PIC16F680 managing a LIS3LV02DQ triaxial accelerometer and a C628 Enhanced JPEG Camera Module. The device can be worn by a lanyard around the neck as shown in Figure 1.a. The block diagram of TheBadge is shown in Figure 1.b. The digital camera takes photos automatically, without user intervention, whilst the device is being worn. In addition, the tri-axial accelerometer estimates the acceleration of the user movements along the x, y, z axes. Photos and acceleration values are stored in a Secure Digital (SD) Card for offline processing. Communication with the camera via UART protocol has been programmed directly in PIC Assembler in order to properly manage the baudrate and overcome the differences between the clock of the microcontroller and the clock of the camera. Communication with accelerometer and SDcard via SPI bus has been written in C. Every 100 *ms*, the values of the acceleration on the x, y, z axis are read and stored in a buffer in the RAM. At most 240

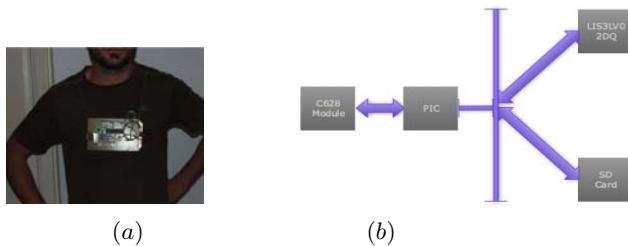


Fig. 1. a) TheBadge worn by an Experimenter; b) Block Diagram of TheBadge

bytes of data can be stored in the PIC RAM, that means 4 seconds of motion data. When the buffer is full, the command for taking a photograph is sent to the camera and a routine for managing data transfer runs. With taking a photo, motion data are complemented with a visual information of the environment.

3 Features Extraction

In order to classify data related to user's activity, a sliding windows technique with 50% of overlap has been used, as described in Dietterich, 2002 [4]. Overlapping demonstrated success in the classification of accelerometer data (Bao and Intille, 2004 [7]). Features for activity classification have been computed on a motion data frame composed by 40 acceleration samples of each of x, y, and z axes, with 20 samples overlapping between consecutive windows. Each window represents 4 seconds of motion data. Features have been extracted using wavelet analysis. Wavelets have advantages over traditional windowed Fourier transforms in analyzing physical situations where the signal contains discontinuities and sharp spikes. From wavelet decomposition using Haar bases, features have been computed as the subband energy of the wavelet coefficients, for data in each axes of the accelerometer. In addition, standard deviation and covariance between all pairwise combinations of the three axes have been computed. Features measuring the degree of coupling between axes can help to discriminate between activities like walking and climbing stairs, having same energy but different behaviours in the axes.

Gabor filters based features have been used for detecting faces. Gabor filters have been considered as a very useful tool in computer vision and image analysis due to its optimal localization properties in both spatial analysis and frequency domain. The Gabor filter banks, whose kernels are similar to the 2D receptive fields of the mammalian cortical simple cells, exhibit desirable characteristics of spatial locality and orientation selectivity. Gabor filter based features have been used for face recognition on different illumination variations, facial expression classification etc. and reported to achieve great successes. For the face detector, Gabor filters with seven orientations on three scales have been used, obtaining 21 filters. The face image is convolved with each filter. Each filtered image is subdivided in 4x4 no-overlapped regions and the integral image of every region is computed obtaining a 336 dimensional feature vector.

4 The Classification System

Data collected with TheBadge have been used for learning user's social activity. A face-to-face social activity might not be evaluated taking into account only one single source of data. For instance, a person speaking at phone could exhibit the same motion patterns than a person speaking with another one. On the other side, a face detector should detect a face of unknown people staying in front of us who is not speaking with us. The block diagram of the software system developed is shown in Figure 2.

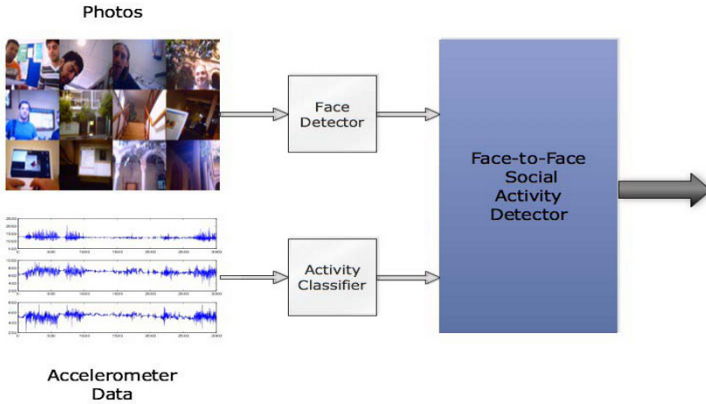


Fig. 2. System Architecture

The overall system has been built as a stacked architecture, as proposed by Wolpert, 1992 [5]. The activity classifier and the face detector are the Level-0 classifiers. The activity classifier uses motion data to classify user’s activities. The face detector detects if people are present in the photos. The output of both classifiers, with in addition position and distance of faces in consecutive images, is given as input of the Level-1 classifier, that perform the final prediction about the social activity. In next subsections, every block of the classification system is described in detail.

Activity Classifier: A multiclass GentleBoost classifier has been used for activity classification. GentleBoost performs better than AdaBoost on the noisy data collected with accelerometer. The multiclass extension of GentleBoost has been performed via ECOC technique as proposed by Allwein *et al.*, 2001 [6]. Activity with strong movements on two or three axis, such as working at computer or moving on chair, have been taken into account together and labeled as a “Moving” activity. On the other side, activities involving slow movements on two axis have been labeled as a “Stopped” activity. Standing still or waiting for the elevator are examples of activities belonging to this class. A social activity has been labeled into a “Speaking” class. It has been chosen to keep together climbing up or down stairs activities in a general “Climbing Stairs” class. Finally, a “Walking” class has been also taken into account.

Face Detector: Images from the real world have been collected with The-Badge. In order to detect faces in such images, a technique able to overcome problems like strong luminance contrast or images quite moved is needed. A face detector has been trained using an AdaBoost classifier working on Gabor filters based features, as proposed by Huan *et al.*, 2004 ([8]) able to detect faces with different poses and illumination conditions assuring, at the same time, very high detection rates and very low false positive rate. The face detector gives as output a boolean value and the position of the face when a face is found in the image.

Face-to-Face Social Activity Detection: A GentleBoost classifier takes care of the detection of a face-to-face social activity. The detector works on windows of three consecutive outputs of both the activity classifier and the face

detector. Three outputs has been arbitrarily chosen from taking into account that a face-to-face social activity does not last less than 15 seconds. In addition, when a face in consecutive images is found, the euclidean distance between the position of the face is computed and passed as input to the final classifier.

5 Experiment and Results

TheBadge has been worn by five persons acting in two circumscribed environments. Each person was asked to perform a sequence of activity with, at least, activities like walking, climbing stairs and speaking with a person. Experimenters performed activities in a random order selected by themselves. The experiments had a duration of 10/15 minutes. Another person annotating the time sequence of the activity, always accompanied who wore TheBadge. At the end of each session, data were downloaded from SD memory and labeled. A total of thirteen data sequences have been collected. Ten of thirteen motion data sequences have

Table 1. 10-fold Cross Validated Confusion Matrix for Activity Classifier

—	Climbing	Speaking	Moving	Stopped	Walking
Climbing	211	1	6	5	30
Speaking	1	287	13	15	25
Moving	0	14	506	90	28
Stopped	14	9	75	663	13
Walking	33	19	25	12	387

Table 2. Confusion Matrix for Activity Classifier evaluated by 10-fold Cross Validation on the Exploitation Set

—	Climbing	Speaking	Moving	Stopped	Walking
Climbing	39	0	0	0	21
Speaking	0	98	2	4	2
Moving	0	0	13	13	6
Stopped	0	0	0	158	5
Walking	5	4	8	2	276

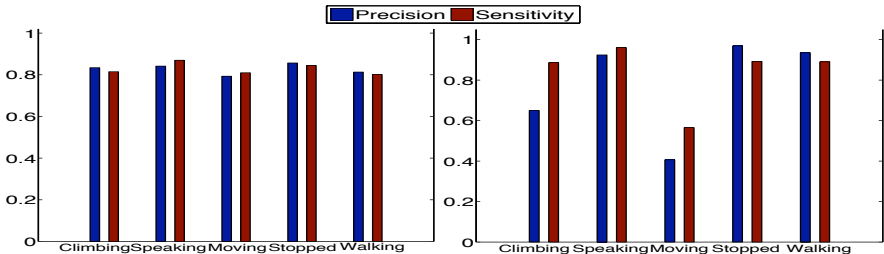


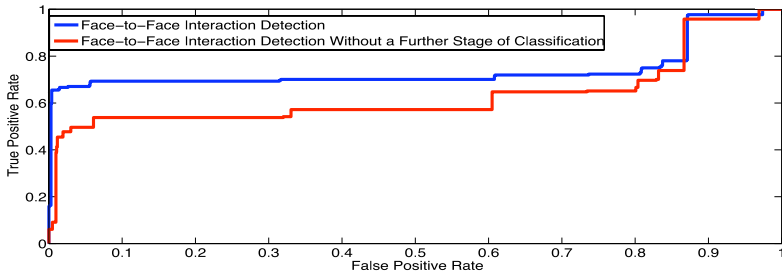
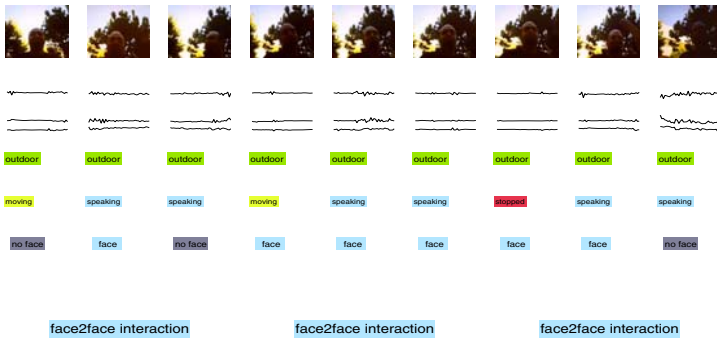
Fig. 3. Precision and Sensitivity for Activity Classifier evaluated on the Training Set, on the right, and on the Exploitation Set, on the left

Table 3. Performaces of Face Detector

Confusion Matrix			Performance Metrics		
—	Face	No Face	—	Precision	Sensitivity
Face	289	7	Face	0.9763	0.9829
No Face	5	474	No Face	0.9895	0.9854

Table 4. Performaces of Social Activity Classifier

Confusion Matrix			Performance Metrics		
—	Interaction	No Interaction	—	Precision	Sensitivity
Interaction	197	57	Interaction	0.6502	0.8239
No Interaction	67	976	No Interaction	0.9151	0.9471

**Fig. 4.** Roc Curve of Face-to-Face Social Activity Classifier, in blue, and Roc Curve obtained taking the logic “and” of the outputs of the Level-0 classifiers without a further stage of classification, in red**Fig. 5.** Example of Classification Results on Data Collected with TheBadge

been used for training the activity classifier, three sequences have been used for exploitation. Table 1 reports the confusion matrix obtained by 10-fold cross validation on the training set. Table 2 reports the confusion matrix obtained by 10-fold cross-validation on the exploitation set. Performance metrics evaluated on both training set and exploitation set are reported in Figure 3. The face

detector has been trained on a set of 294 faces taken from the Yale Face Database ([9]) and 481 no face images. Confusion matrix obtained by 10-fold cross validation on the training set and performance metrics are reported in Table 3. The detector has a false positive rate of 0.01%. The detector has been tested on 100 images taken with TheBadge, 50% containing faces and 50% not containing faces. True positive detections occur in 32 images. False negatives occur in 16 images. In only four images, false positives have been detected.

The Face-to-Face Social Activity Classifier has been trained using the entire dataset and validated by 10-fold cross validation. Confusion matrix and performances of the detector are shown in Table 4. The detector has a classification rate of 0.928. Figure 4 shows the ROC curve of the face-to-face social activity detector, in blue. The area under ROC curve is 0.7398. In the same figure, it has been shown that the detection of a Face-to-Face Social activity using only the logic “and” between the outputs of the Level-0 classifiers, without a further stage of classification, is worst than using a subsequent stage of classification. In this case, the area under the red ROC curve equals to 0.6309.

In Figure 5, a sequence of classified data is shown. The activity classifier correctly classifies six of nine motion data frame. The interaction detector operating on windows, despite some confusion of the base classifiers, correctly detects that an interaction is occurring.

6 Conclusions and Future Works

In this work, the feasibility of building a socially-aware wearable device has been evaluated. Results show that a system being aware of a user’s social activity is feasible using a camera and an accelerometer. Data collection from two sensors has been useful for avoiding the lack of information due to store data after every acquisition frame. From motion data, coarse activities as speaking, walking and climbing stairs have been successfully detected. In particular, it has been shown that a speaking activity can be recognized and separated from general moving patterns using only data related to the user’s motion. Analyzing the images, the presence of a person in front of the user confirms that user had a social activity in that time. Furthermore, using a second level of classification enhances the prediction of a social activity.

Many improvements have to be taken into account in our system. At hardware level, passing from the prototyped version to a more robust system is necessary for optimizing power consumption and operating time with the aim to perform experiments on long time and to use TheBadge in many real life situations. In addition, social activities involve communication. For that reason, adding audio capabilities able to get data about user’s conversations must be considered a necessary and immediate evolution of TheBadge. At software level, using different sequential machine learning techniques will improve the detection of activities based on the inherently sequential motion data. Finally, some non technical issues related to privacy and safeness of data collected have to be taken into account. In our experiments, all the people wearing TheBadge give their consent

for using data collected. The aim of future works will be to develop a system able to process data in real-time without the need to store personal data. In this way, the system will not invade privacy of people wearing TheBadge.

Acknowledgments. This work is partially supported by a research grant from projects TIN2006-15308-C02, FIS-PI061290 and CONSOLIDER-INGENIO 2010 (CSD2007-00018), MI 1509/2005.

References

1. Moran, T.P., Dourish, P.: Introduction to Special Issue on Context-Aware Computing. *Human-Computer Interaction (HCI)* 16(2-3), 87–96 (2001)
2. Pentland, A.: Socially Aware Media. In: *MULTIMEDIA 2005: Proceedings of the 13th annual ACM international conference on Multimedia*, vol. 1, pp. 690–695 (2005)
3. Hodges, S., Williams, L., Berry, E., Izadi, S., Srinivasan, J., Butler, A., Smyth, G., Kapur, N., Wood, K.: SenseCam: A Retrospective Memory Aid. In: Dourish, P., Friday, A. (eds.) *UbiComp 2006*. LNCS, vol. 4206, pp. 177–193. Springer, Heidelberg (2006)
4. Dietterich, T.G.: Machine Learning for Sequential Data: A Review. In: Caelli, T.M., Amin, A., Duin, R.P.W., Kamel, M.S., de Ridder, D. (eds.) *SPR 2002 and SSPR 2002*. LNCS, vol. 2396, pp. 15–30. Springer, Heidelberg (2002)
5. Wolpert, D.H.: Stacked generalization. *Neural Netw.* 5(2), 241–259 (1992)
6. Allwein, E.L., Schapire, R.E., Singer, Y.: Reducing multiclass to binary: a unifying approach for margin classifiers. *JMLR* 1, 113–141 (2001)
7. Bao, L., Intille, S.S.: Activity recognition from user-annotated acceleration data. In: *Proc. Pervasive*, Vienna, Austria, vol. 1, pp. 1–17 (2004)
8. Huang, L., Shimizu, A., Kobatake, H.: Classification-based face detection using Gabor filter features. *Automatic Face and Gesture Recognition* 21(9), 849–870 (2004)
9. Georghiades, A.S., Belhumeur, P.N., Kriegman, D.J.: From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose. *IEEE Trans. Pattern Anal. Mach. Intell.* 23(6), 643–660 (2001)