Ambient and Wearable Sensor Fusion for Activity Recognition in Healthcare Monitoring Systems

Julien Pansiot[1,2], Danail Stoyanov[1,2], Douglas McIlwraith[1], Benny P.L. Lo[1,2] and G. Z. Yang[1,2]

[1] Royal Society/Wolfson MIC Lab, Department of Computing, Imperial College London, United Kingdom
[2] Institute of Biomedical Engineering, Imperial College London, United Kingdom

Abstract— **The use of wearable sensors for home monitoring provides an effective means of inferring a patient's level of activity. However, wearable sensors have intrinsic ambiguities that prevent certain activities to be recognized accurately. The purpose of this paper is to introduce a robust framework for enhanced activity recognition by integrating an ear-worn activity recognition (e-AR) sensor with ambient blob-based vision sensors. Accelerometer information from the e-AR is fused with features extracted from the vision sensor by using a Gaussian Mixture Model Bayes classifier. The experimental results showed a significant improvement of the classification accuracy compared to the use of the e-AR sensor alone.**

Keywords— **blob sensor, wearable sensor, sensor fusion, activity recognition**

I. Introduction

Monitoring the status of the elderly or chronically ill patients in their own homes is an essential requirement for delivering more effective pervasive healthcare. By continuous monitoring of key physiological parameters of the patients, wearable sensors can provide a rich source of information about their current status of health [1]. There is also increasing evidence to suggest that major episodes are often preceded by changes in behaviour and domicile activity, which may be detected from detailed information about the posture, gait and general activity of the patient based on ambient sensing [2]. To achieve truly pervasive health monitoring, it is necessary to integrate wearable/ implantable sensors provided by a body sensor network (BSN) with data acquired from ambient environment sensors.

Recently, various systems for home monitoring have been developed based on either wearable or ambient sensors. Frameworks using wearable sensors are typically based on accelerometers [15,16], ECG sensors [18], pulse oximeters ($Sp0_2$) [16], temperature [15], and bend sensors [12,13,18]. Other wearable sensors include humidity, acoustic and light sensors [14]. By the use of wearable sensors, it provides an effective means of monitoring the biophysical status of the patient. Due to the lack of a global reference, however, it can be difficult to use this data to infer certain physical activities. For example, a wearable accelerometer positioned on the head can detect local motion but not whether the subject is standing or sitting. Either ambient or additional wearable sensors (typically positioned on the joints) need to be used for achieving the required body posture differentiation.

For monitoring environments based on ambient sensors, current systems include the use of cameras [4,11,21,24], IR sensors, ambient sound [21], heat, as well as contact sensors mounted on furniture [19]. These systems can provide information about the spatial location and general activity of the subject within the environment. The weakness of ambient sensing is that it is often too ambiguous to differentiate detailed information about the subject, which in many cases can only be derived from a wearable system.

Existing research has shown that there is a complementary relationship between the two sensing paradigms. Effective sensor fusion can be used to combine the strengths of ambient and wearable sensors by fusing sensory data at hardware, raw data, feature, or decision levels [22]. At the hardware level, it can be achieved by using simple thresholds [15]. At the data level, dimensionality reduction such as Principal Component Analysis (PCA) is often deployed before further pattern classification techniques are applied. At this level of sensor fusion, modelling methods such as Gaussian mixtures [14], Bayes networks [21] or Hidden Markov Models (HMM) methods are common.

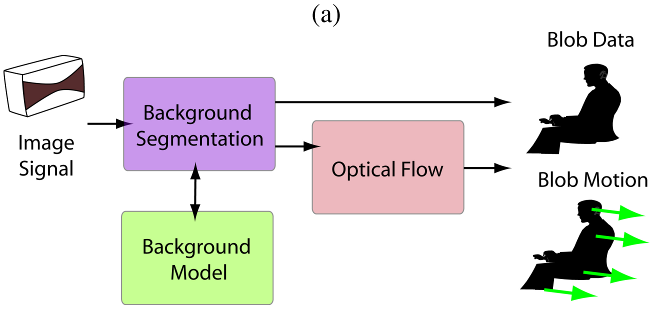
The purpose of this paper is to develop a framework for improved activity recognition by integrating an ear-worn activity recognition (e-AR) sensor with ambient blob sensors. Data independently obtained by each sensor is preprocessed for dimensionality reduction before the application of a Bayesian classifier. To assess the improved accuracy of the proposed method, we evaluated the classifier against a lab-based home monitoring scenario. Significant improvements in the recognition rates of all activities have been achieved when compared to using wearable or ambient sensors alone.

II. System Design

A. Wearable e-AR Sensor

The e-AR sensor [16] is based on the BSN platform [20] that consists a Texas Instruments MSP430 processor, Chipcon CC2420 radio transceiver, Atmel 512KB EEPROM, MCC ChipOX SpO_2 module and a 3-axis accelerometer. The integrated e-AR sensor used for this study is shown in Fig. 1(a). For activity recognition in this study, the main information used was derived from the 3-axis accelerometer whereas SpO_2 signals of the e-AR sensor were not used.

(a)



(b)

Fig. 1 (a) The e-AR sensor used in this study [16] and (b) the data processing pipeline for the ambient blob sensor.

B. Ambient Blob Sensors

The ambient sensor proposed in this study is a self-contained module consisting of a video sensor, on-board processor, wireless communication and battery [4]. It has a wall mount design and can be integrated into the home environment similar to a PIR security device. Video data observed by the device is processed on board in real-time and the sensor communicates only derived signal metrics such as the silhouette of a moving object and its local motion in the form of optical flows. Communication between ambient nodes is used to provide large scale tracking and improves the overall system robustness. The ambient sensor being under development, CCTV cameras were used in this experiment. Under this sensing paradigm, blobs representing the monitored subject are first extracted from the video signal using a background statistical model, where every pixel is represented as a Gaussian mixture distribution maintained over time as proposed by Lee [10]. Incoming signals are compared with the existing background model and segmented into a binary map of foreground and background. The use of normalized RGB colour space reduces the sensitivity of the algorithm to shadows. Post-processing of the foreground object based on mathematical morphology is used for noise removal.

In addition to the extraction of blob profiles, the optical flow within the blob is also extracted, which is based on the classical technique proposed by Horn and Schunck [8]. Optical flow can be considered as a natural extension of the blob silhouette as it also captures the motion of the limbs. This information has been used previously for gesture recognition [25] and activity recognition in a multi-resolution framework [9]. The complete data processing work flow is summarized in Fig 1(b).

It is important to note that the silhouette and optical flow extracted by each ambient sensor do not carry any appearance information and no image data is transmitted to other devices. This is important for home care environments where privacy is of high priority.

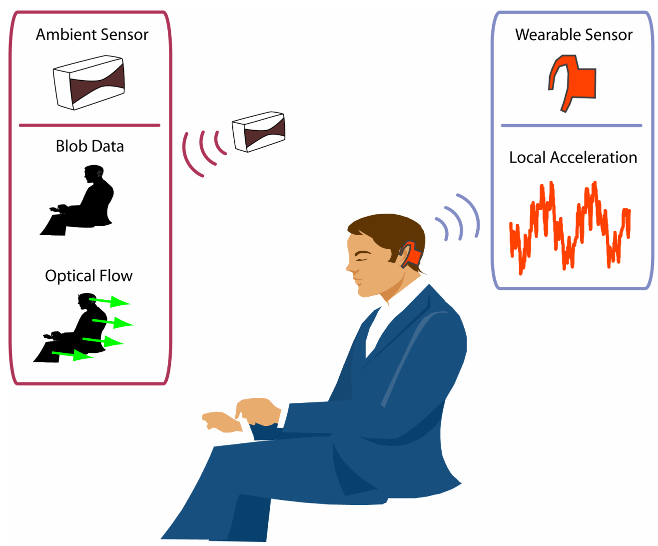


Fig. 2 A schematic diagram of the proposed ambient and wearable monitoring system.

C. Overall System Integration

For the motion data provided by the e-AR sensor, the accelerometer signal is intrinsically linked to the patient's movement. On its own, it is capable of differentiating activities such as walking, standing, and sleeping. For certain activities, however, the readings from the e-AR sensor are ambiguous, *e.g,* standing and sitting still. In these cases, the e-AR sensor cannot correctly classify the patient's activity as it cannot obtain a global perspective of the body's position just from the head motion alone. It is expected that by fusing the e-AR data with the ambient sensors at the data level, it is a possible to obtain a much more reliable activity classification result for a wider range of activities. Fig. 2 illustrates a schematic diagram outlining the proposed system.

III. Ambient and Wearable Sensor Fusion

For effective sensor fusion, two types of features are extracted from the e-AR accelerometers: tilt and movement frequency spectrum. To derive the tilt information, the accelerometers are pre-calibrated such that the acceleration due to gravity can be evaluated. A record of the total acceleration in the three axes allows the calculation of the gravity constant component on each of the accelerometers. It is therefore possible to separate the relative head acceleration and gravity acceleration to the tilt with respect to the vertical axis. Moving window Fast Fourier Transform (FFT) was also computed on the acceleration data to the intrinsic frequency of the movement.

Table 1 Features used in classification

Sensor	Feature	Size
e-AR	Acceleration X axis	1
e-AR	Acceleration Y axis	1
e-AR	Acceleration Z axis	1
e-AR	FFT acceleration X axis	11
e-AR	FFT acceleration Y axis	11
e-AR	FFT acceleration Z axis	11
e-AR	Head tilt X	1
e-AR	Head tilt Z	1
Blob	Blob speed estimation	1
Blob	Blob aspect ratio	1
Blob	Subject height estimation	1
Blob	Subject optical flow intensity	1
Blob	Subject optical flow correlation	1
Blob	Subject optical flow aspect ratio	1

From the ambient sensor, the derived features used for sensor fusion include the aspect ratio and mean velocity of the blob. The calculation of the optical flow is based on the iterative application of the following equation:

$$V^{k+1} = \overline{V^k} - I\frac{I_x\overline{V_x^k} + I_y\overline{V_y^k} + I_t}{\alpha^2 + I_x^2 + I_y^2}$$

for iteration k for a given pixel. In this equation V^k is the optical flow vector, I , I_x , I_y and I_t the image intensity and its partial derivatives and $\overline{V^k}$ the average of V^k in its neighbourhood. In the above equation, α is a regularization constant to ensure the smoothness of the result derived. After noise filtering, the main moving elements of the field-of-view are extracted and the bounding box is calculated from the the eigenvectors of the covariance matrix of the blobs as proposed by Lahanas *et al.* [23]. A complete list of the features used for the classifier is summarised in Table 1.

Sensor fusion is performed based on a Gaussian Bayes EM classifier based on the e-AR and the blob sensor data. A Gaussian Mixture Model (GMM) is used to model each activity class. For the implementation of the classifier, we used the Bayes Net Toolkit (BNT) [5] and a total of nine classes were modelled to describe different activities. The activities used for classification in this study include walking, standing still, standing with head tilted on the side, sitting at the dining table, reading at the table, eating, sitting on the sofa, lounging on the sofa and eventually lying down.

For each of the activities considered, three quarters of the data were used for training of the inference system and the rest for validation. For training, an Expectation-Maximisation (EM) iterative method was used to compute the maximum likelihood fit [6]. Given a dataset $\{x_{1..R}\}$ to be classified in c classes, and assuming that the conditional probability density function (PDF) $P(X=x)$ for each of these classes is Gaussian, we try to find the best fit of:

$$\lambda_t = \{\mu_{1..c}(t), \Sigma_{1..c}(t), p_{1..c}(t)\}$$

where $\mu_i(t)$, $\Sigma_i(t)$ and $p_i(t) = P(w_i)(t)$ are the mean, the covariance and the estimates of the weights of the mixtures at the iteration t, respectively. The expectation and maximization steps are performed iteratively until convergence. The expectation step for each class i, based on Bayes' law can be represented as the following:

$$P(w_i|x_k,\lambda_t) = \frac{p(x_k|w_i,\lambda_t)P(w_i|\lambda_t)}{p(x_k|\lambda_t)} = \frac{p(x_k|w_i,\mu_i(t),\Sigma_i(t))p_i(t)}{\sum_{j=1}^{c} p(x_k|w_j,\mu_j(t),\Sigma_j(t))p_j(t)}$$

Because the Gaussian PDF is differentiable, the maximization of the likelihood step for each class i can be expressed as:

$$\mu_i(t+1) = \frac{\sum_{k=1}^{R} P(w_i | x_k, \lambda_t) x_k}{\sum_{k=1}^{R} P(w_i | x_k, \lambda_t)}$$

$$\Sigma_i(t+1) = \frac{\sum_{k=1}^{R} P(w_i | x_k, \lambda_t)(x_k - \mu_i(t+1))(x_k - \mu_i(t+1))^T}{\sum_{k=1}^{R} P(w_i | x_k, \lambda_t)}$$

$$p_i(t+1) = \frac{1}{R} \sum_{k=1}^{R} P(w_i | x_k, \lambda_t) x_k$$

Once the model is computed through EM, the remaining data is used to evaluate the accuracy of the classifier based on the marginal probability of every activity. The highest probability was chosen for the final classification.

IV. Experiments and Results

To evaluate the proposed technique, a purpose built simulated home environment was used. Data was recorded for two actors wearing an e-AR sensor. A total of 9 activities were performed by the actors with each activity lasted for approximately one minute. The classification results by using the proposed method as compared those with the e-AR sensor alone are presented in Table 2. A detailed analysis of inter-class misclassification as illustrated as confusion matrices with and without sensor fusion is provided in Figure 3.

From the results shown in Table 2, it is evident that by incorporating ambient sensing with the e-AR sensor, activity classification is improved significantly. This is particularly obvious for classes where the e-AR sensor was ambiguous due to a lack of global information. For example, reading is not easily classified with the e-AR sensor only, as very little temporal and global orientation information is available. With the use of the blob sensor, this improves significantly because of the strong difference of the appearance cue. The same effects are visible in differentiating classes such as standing with the head tilted and sitting activities such as eating and reading, which are not well classified by the e-AR sensor alone. In these cases, the optical flow features provide a good clue about the type of activity, which significantly improves the sensitivity and specificity of the system. In the current implementation, however, the differentiation between sitting on the chair and on the sofa is relatively low, as evident from the confusion matrix shown in Figure 3.

Table 2 Comparison of activity classification rates between using a wearable sensor alone and with the proposed combined system

Class	**Activity**	**e-AR sensor alone**	**e-AR + ambient sensing**
1	Walking	79%	100%
2	Standing	83%	75%
3	Standing (head tilted)	65%	80%
4	Sitting	73%	47%
5	Reading	55%	80%
6	Eating	39%	81%
7	Sitting (sofa)	84%	90%
8	Lounging	77%	92%
9	Lying down	100%	100%

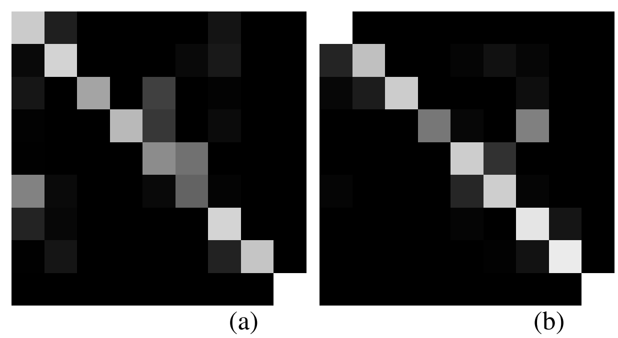


Fig. 3 The confusion matrices showing how the algorithm differentiates eating and reading activities with (a) the e-AR sensor only and (b) after sensor fusion. The main non-diagonal element in (b) is the confusion between sitting on a chair and on the sofa.

III. Conclusions and Future Work

In this paper, we have proposed a sensor fusion framework for integrating ambient and wearable e-AR sensors. Our experiments illustrate the practical value of the method by improving classification rates for most activities investigated in this study. This clearly demonstrates the fact that ambient environment sensors can be used to overcome some of the ambiguities in activity recognition by using wearable sensing alone. This is a desirable feature for the effective deployment of future pervasive patient monitoring systems.

In the current system, we did not explicitly handle the spatial dependency between the ambient sensor and the patient. This projective relationship can influence the ambient sensor readings. We are currently looking into resolving this issue by using multi-view geometry to derive pose invariant 3D representations. Other areas for further improvement include the development of robust learning capabilities of the ambient sensors and real-time implementation of the proposed sensor fusion paradigm directly on the sensor nodes.

References

1. Yang G-Z (2006) Body Sensor Networks
2. Verghese J, Lipton R B, Hall C B, Kuslansky G, Katz M J, Buschke H (2002) Abnormality of gait as a predictor for non-Alzheimer's dementia. N Engl J Med, 347:1761-1768
3. Stauffer C, Grimson W E L (2000) Learning patterns of activity using real-time tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22:747-757
4. Lo B P L, Wang J L, Yang G Z (2005) From imaging networks to behavior profiling: Ubiquitous Sensing for Managed Homecare of the Elderly. Proc. International Conference on Pervasive Computing (PERVASIVE 2005), pp. 101-104
5. Murphy K. Bayes Net Toolkit. http://bnt.sourceforge.net
6. Dempster A P, Laird N M, Rubin D B (1977) Maximum likelihood from incomplete data via the EM algorithm, by Journal of the Royal Statistical Society. Series B (Methodological) Royal Statistical Society
7. McKenna S J, Nait-Charif H (2004) Summarising contextual activity and detecting unusual inactivity in a supportive home environment. Pattern Analysis and Applications 7(4), 386-401, Springer-Verlag, December 2004
8. Horn B K P, Schunck B G (1981) Determining optical flow. Artificial Intelligence, vol 17, pp 185-203.
9. Wactlar H, Yang, Jie, Chen, D, Hauptmann A, Christel M (2004) Infrastructure for Machine Understanding of Video Observations in Skilled Care Facilities: Implications of Early Results from CareMedia Case Studies. UbiComp 2004: The 2nd International Workshop on Ubiquitous Computing for Pervasive Healthcare, Applications, Nottingham, U.K., September 6-7, 2004
10. Lee D-S (2005) Effective Gaussian Mixture Learning for Video Background Subtraction. IEEE Transactions on Pattern Analysis and Machine Intelligence ,vol. 27, no. 5, pp. 827-832, May, 2005.
11. Pansiot J, Stoyanov D, Lo B P L, Yang G-Z (2006) Towards image-based modeling for ambient sensing. . International Workshop on Wearable and Implantable Body Sensor Networks (BSN 2006), 3-5 April 2006
12. Amft O, Junker H, Lukowicz P, Troster G, Schuster C (2006) Sensing muscle activities with body-worn sensors. International Workshop on Wearable and Implantable Body Sensor Networks (BSN 2006), 3-5 April 2006
13. Krebs D E, Huddleston J I, Goldvasser D, Scarborough D M, Harris W H, Malchau H (2006) Biomotion community-wearable human activity monitor: total knee replacement and healthy control subjects. International Workshop on Wearable and Implantable Body Sensor Networks (BSN 2006), 3-5 April 2006
14. Grajales L, Nicolaescu I V (2006) Wearable multisensor heart rate monitor. International Workshop on Wearable and Implantable Body Sensor Networks (BSN 2006), 3-5 April 2006
15. Laerhoven K V, Gellersen H W, Malliaris Y G (2006) Long term activity monitoring with a wearable sensor node. International Workshop on Wearable and Implantable Body Sensor Networks (BSN 2006), 3-5 April 2006
16. Aziz O, Lo B, King R, Darzi A, Yang G-Z (2006) Pervasive body sensor network: an approach to monitoring the post-operative surgical patient. International Workshop on Wearable and Implantable Body Sensor Networks (BSN 2006), 3-5 April 2006
17. Giorgino T, Quaglini S, Lorassi F, De Rossi D (2006) Experiments in the detection of upper limb posture through kinestetic strain sensors. International Workshop on Wearable and Implantable Body Sensor Networks (BSN 2006), 3-5 April 2006
18. Park C, Pai H, Chou, Bai Y, Matthews R, Hibbs A (2006) An Ultra-Wearable, Wireless, Low Power ECG Monitoring System, to appear in Proc. IEEE BioCAS, Nov 29 - Dec 1, 2006. The British Library, London.
19. Reeves A A, Ng J W P, Brown S J, Barnes N M (2006) Remotely supporting care provision for older adults. International Workshop on Wearable and Implantable Body Sensor Networks (BSN 2006), 3-5 April 2006
20. Lo B, Yang G-Z (2005) Key Technical Challenges and Current Implementations of Body Sensor Networks. IEE Proceedings of the 2nd International Workshop on Body Sensor Networks (BSN 2005), 1-5, April 2005
21. Chen D, Malkin R, Yang J (2004) Multimodal detection of human interaction events in a nursing home environment. Proceedings of the 6th international Conference on Multimodal interfaces (State College, PA, USA, October 13 - 15, 2004). ICMI '04. ACM Press, New York, NY, 82-89.
22. McCullough C L, Dasarathy B V, Lindberg P C (1996) Multi-level sensor fusion for improved target discrimination, Decision and Control. Proceedings of the 35th IEEE , vol.4, no.pp.3674-3675 vol.4, 11-13 Dec 1996
23. Lahanas M, Kemmerer T, Milickovic N, Karouzakis K, Baltas D, Zamboglou N (2000) Optimized bounding boxes for three-dimensional treatment planning in brachytherapy. Medical Physics 27 (10): 2333-2342 October 2000
24. Wang J, Lo B P L, Yang G-Z (2005) Ubiquitous Sensing For Posture/Behavior Analysis. IEE Proceedings of the 2nd International Workshop on Body Sensor Networks (BSN 2005), pp.112-115, April 2005.
25. Cutler R, Davis L (1998) View-based detection and analysis of periodic motion. International Conference on Pattern Recognition, Brisbane, Australia, August 1998

Address of the corresponding author:

Author:	Julien Pansiot
Institute:	Department of Computing, Imperial College London
Street:	180 Queen's Gate, South Kensington
City:	London
Country:	United Kingdom
Email:	julien.pansiot@imperial.ac.uk