Personalized Life Log Media System in Ubiquitous Environment

Ig-Jae Kim, Sang Chul Ahn, and Hyoung-Gon Kim

Imaging Media Research Center, KIST,
39-1, Hawolgokdong, Seongbukgu, Seoul, Korea
{kij,asc,hgk}@imrc.kist.re.kr

Abstract. In this paper, we propose new system for storing and retrieval of personal life log media on ubiquitous environment. We can gather personal life log media from intelligent gadgets which are connected with wireless network. Our intelligent gadgets consist of wearable gadgets and environment gadgets. Wearable gadgets include audiovisual device, GPS, 3D-accelerometer and physiological reaction sensors. Environment gadgets include the smart sensors attached to the daily supplies, such as cup, chair, door and so on. User can get multimedia stream with wearable intelligent gadget and also get the environmental information around him from environment gadgets as personal life log media. These life log media(LLM) can be logged on the LLM server in realtime. In LLM server, learning-based activity analysis engine will process logged data and create meta data for retrieval automatically. By using proposed system, user can log with personalized life log media and can retrieve the media at any time. To give more intuitive retrieval, we provide intuitive spatiotemporal graphical user interface in client part. Finally we can provide user-centered service with individual activity registration and classification for each user with our proposed system.

Keywords: life log system, spatiotemporal interface, activity analysis.

1 Introduction

Recently, a large number of researches have been proposed for recording and retrieval for the information of personalized daily life due to the development of ubiquitous computing devices. We call these personalized media life log media(LLM). Life log media include the thing that one can see, the sound that one can hear, the information where one is, the state how one feels and so forth. Realtime retrieval of continuously captured personalized LLM will assist to enhance user's memory. To do this, we propose the new system with semi-automatic annotation technique based on our speech and activity analysis engine. We can give a cognitive assistants to people who want to organize their activity. To do this, we, first, record events using intelligent gadgets which is composed of wearable gadgets and environmental gadgets. Using captured LLM, we apply learning based activity classification technique based on multimodal analysis. After activity analysis, meta data for retrieval is created automatically. With

F. Stajano et al. (Eds.): ICUCT 2006, LNCS 4412, pp. 20–29, 2007.
© Springer-Verlag Berlin Heidelberg 2007

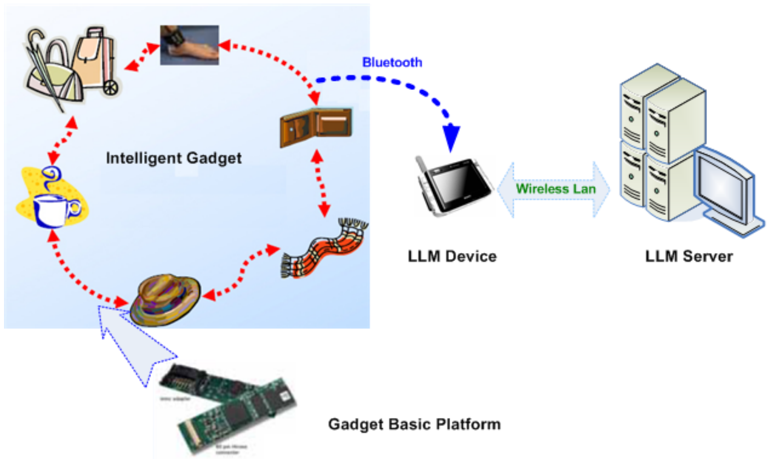

Fig. 1. Our system environment

described techniques, we can provide access to captured life log media at multiple levels of granularity and abstractions, using appropriate access mechanism in representations and terminology familiar to application users.

Our proposed system is composed of several parts, such as intelligent gadget, LLM device, LLM server and LLM browser. Figure 2 shows the relation of each component. In intelligent gadget, there are two components, wearable gadget and environmental gadget. Wearable gadget includes GPS, camera, microphone, body sensors, and HMD which give a functionality of I/O for P-LLM. Environmental gadget is built on the small and low power processing module and has a wireless interface like a zigbee or bluetooth. It is attached to our daily objects to give an information which is given to the LLM device of user. LLM device carried by user can get the information from intelligent gadget and send captured P-LLM to the LLM server at the idle time. In our LLM server, we can analyze the speech signal to identify speaker and classify speech from environment noise. We can, also, analyze the video signal to detect the registered objects and human faces. Besides, we can classify the pre-defined activity using multimodal sensor fusion technique. After analyzing LLM data in the server, Meta data are associated with A/V media data automatically and are used to retrieve. In our LLM client shown in Figure 3, we make a web-based browser with spatiotemporal query interface and tree-based activity search interface, such that user can query intuitively and see the retrieved results at their own LLM device. The rest of this paper is organized as follows. Section 2 reviews the related work in life log system. We present the whole scheme of our proposed system in section 3 and detail explanation about analysis for life log media data in section 4. Finally, we show the experimental results and conclusions in section 5, 6 respectively.

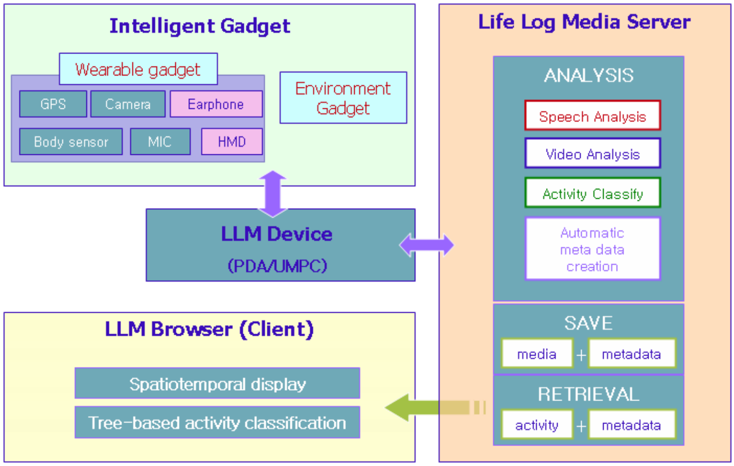

Fig. 2. Whole scheme for our proposed system

2 Related Work

There have been proposed several techniques for personal life log system . Gemmell et al. introduced SenseCam which is a device that combines a camera with a number of sensors[2]. Data from SenseCam is uploaded into a MyLifeBits repository, where a number of features, but especially correlation and relations, are used to manage the data. Mann described EyeTap which facilitate the continuous archival and retrieval of personal experiences, by way of lifelong video capture[3]. Vemuri et al. presented a method for audio-based memory retrieval[6]. They developed a pc based memory retrieval tool allowing browsing, searching, and listening to audio and associated speech-recognizer-generated transcripts. Aizawa et al. used audiovisual information as content to detect the conversation scenes and GPS data was applied as context to extract spatiotemporal key frames from time and distance sampling[1]. Tancharoen et al. extended their previous work including content based talking scene detection and context based key frame extraction using GPS data[9]. Recognizing general human activity or special motions is important key for automatic annotation. Recognizing general user activity has been tried by various authors. Randell et al. have done early investigations of the problem using only single 2-axis accelerometers[8]. Kern et al. presented a hardware platform to use multiple acceleration sensors that are distributed over the body[7]. They could capture 3-dimensional acceleration data from up to 48 positions on the human body. It is especially designed for robustness, allowing for recording even very dynamic activities, such as playing badminton or climbing. Chambers et al. focus on the recognition of complex gestures using Hidden Markov Models[4]. Kern et al. summarizes work on automatically annotating meeting recordings, extracting context from body-worn acceleration sensors alone, and combining context from three different sensors (acceleration, audio, location) for estimating the interruptability of the user[10].

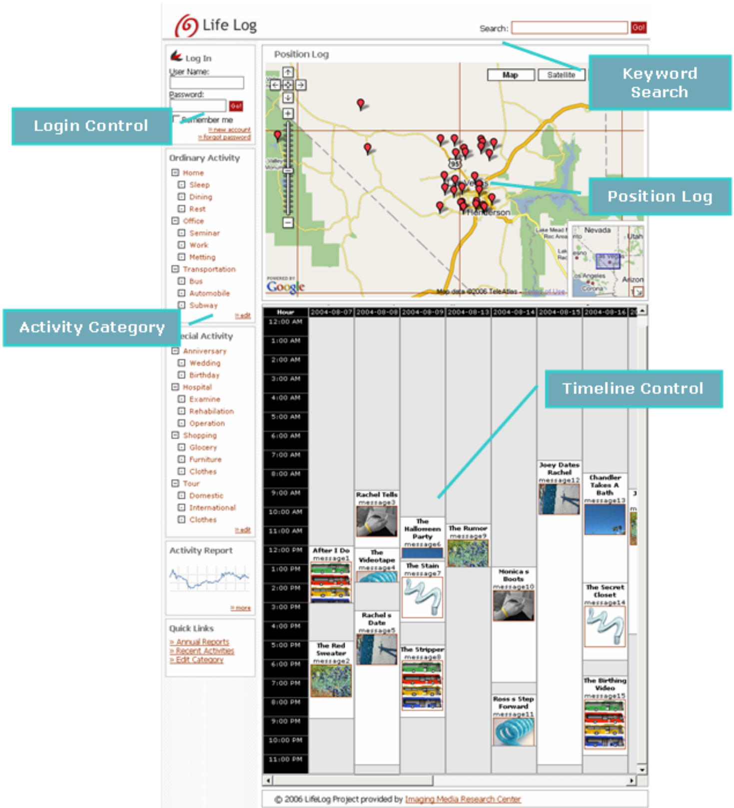

Fig. 3. Our web-based Life log media browser : User can query intuitively with spatiotemporal query interface and activity category

3 System Configuration

As shown in figure 2, our proposed system consists of four components. In intelligent gadgets, there are two modules, wearable gadgets and environment gadgets. User can capture the LLM data and show the retrieval results with wearable gadgets, such as camera, microphone, HMD and body sensors. These wearable gadgets are connected to the LLM device which user always carry. LLM device can also be connected to the environment gadgets which is embedded into the articles for daily use. In figure 1, we show the example of intelligent gadgets. We implemented the information gathering module and wireless communication module, such as zigbee or bluetooth onto the gadget basic platform and attached it to the objects for daily use. These intelligent gadgets are connected to the LLM device with wireless module and then LLM device can gathering the user's information in realtime. In fact, user's logging data must be sent to the LLM server in realtime, however, it is impossible to be connected to the server in everywhere. Practically, we use LLM device as a buffer for storing user's logging

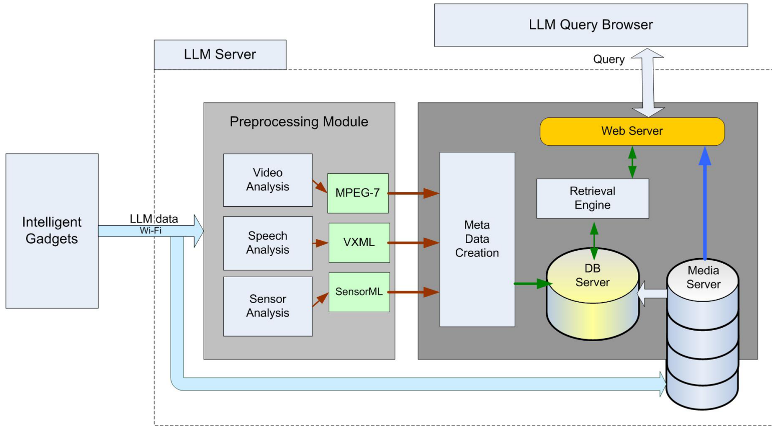

Fig. 4. LLM server system architecture

data temporarily in case the network connection is unavailable. And also, LLM device is used as a terminal for querying and browsing the retrieved results. After the LLM data being transmitted to the server, the LLM server analyze the user's activity.

We implemented the LLM server on the Windows XP. Figure 4 shows the architecture of LLM server. For web server and database server, we use IIS server, ASP.NET framework and PostgreSQL. In LLM server, user-dependent activity analysis can be done using multimodal data fusion technique, such as automatic video, speech and sensor data analysis. User can query and see the retrieved results through the LLM browser, which is served as web service and give user friendly spatiotemporal interface. In the following each section, more detailed explanation for each component will be given.

4 Activity Analysis

As mentioned in previous section, the captured LLM data from intelligent gadgets will be uploaded to the LLM server. After LLM being uploaded, the analysis will be started to classify the activity. Basically, our proposed system is based on the semi-automatic activity analysis. For the activity analysis, we use multimodal sensor fusion technique which is based on the speech analysis, video analysis and pattern classification from various sensors, such as accelerometer, gyro, physiological reaction sensors and environment sensors. In audio analysis, we extract the information of time, sex distinction, the number of speaker and captured environment as a meta data. Those information can be taken from pattern classification using several feature vectors such as MFCC, ZCR, Cepstrum energy and spectral differences. In video analysis, we used machine learning technique to detect the indoor location and registered objects. First, we select some pictures for specific area and familiar objects, such as corridor, meeting room, monitors and so forth, and then put them into the training data set. We also processed on video stream to detect scene change using histogram matching method and identify the registered users using fisher classifier. From audiovisual analysis and action detection from smart sensors, we can transform them to LLM meta data according to the hierarchical meta data structure in figure 5.

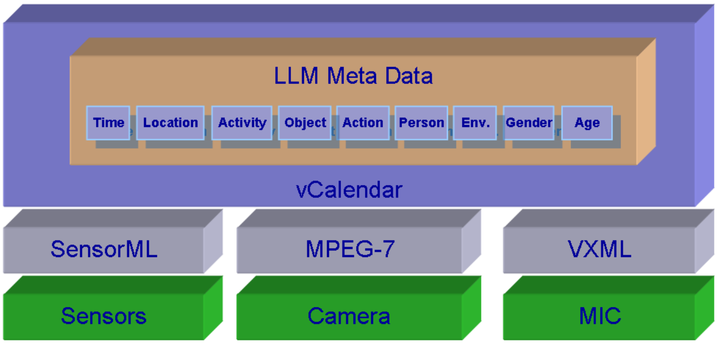

Fig. 5. LLM meta data structure

Multiple sensor data give a user's action information to classify the user's activity, such as lie down, running and fast moving. We use four 3D-accelerometers to detect the user's activity. These are attached above the each knee and on the each wrist of the user. Figure 6 shows the hierarchical sensor fusion process for LLM. We adopted bayesian analysis for multiple sensor fusion to classify the user's activity. An example of meta data creation using sensor fusion process is shown in Figure 7. However, it is impossible to define the user's whole activities automatically because the definition of activity is a sort of subjective evaluation, therefore even though the same sensor value will not be assigned the same category for the different users. For this reason, we define the general category for user's activity in advance, such as ordinary activity and extra-ordinary activity. In ordinary activity is related to the activity in home or office. Generally, the activities occurred outside of those area, they are classified as extraordinary activities. In addition to these pre-defined activities, users can add their own activity through our learning based structure.

To provide the user-dependent service, we need the definition for the individual activity for each user. To do this, as mentioned previous section, we let the user annotate personally at once for the repeated behavior with same objects and same time, as it is called *learning process*. After then, our activity inferring engine can automatically annotate for the same action. Although there are some intervention of user, when the user register the special activity on his/her own browser, in our activity analysis engine, it is more robust way for classify the user's activity in comparison with full-automatic activity classification method.

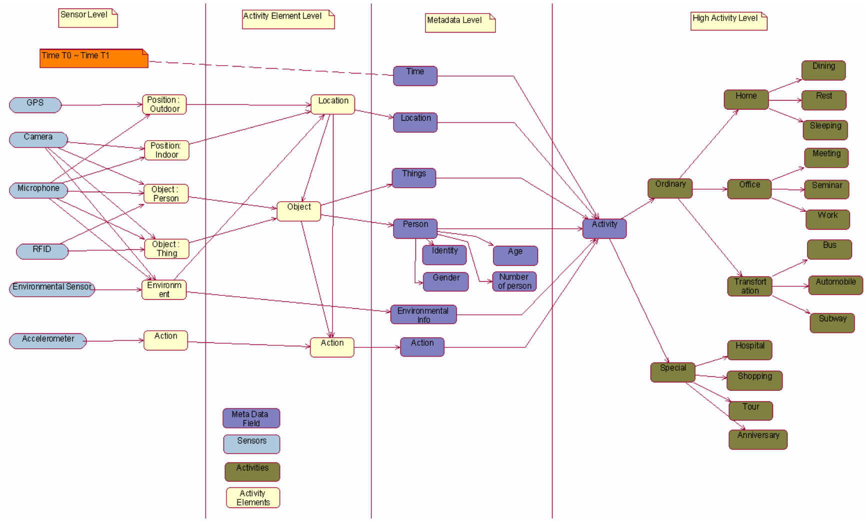

Fig. 6. Hierarchical sensor fusion process for LLM

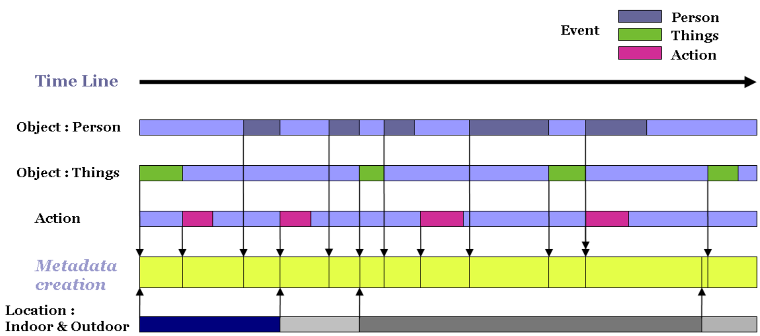

Fig. 7. Meta data creation from sensor fusion process

After learning process, we can calculate the probabilities of each sensors as a hypothesis for maximum likelihood estimation.

5 LLM Browser

After the automatic annotation is completed, users can search the AV data whenever they want through our LLM browser. In our LLM browser, we provide the user friendly graphical interface with spatiotemporal query interface and visualization, tree based menu selection method and categorized activity selection. Besides, users can access whenever/wherever they want as web service. We show

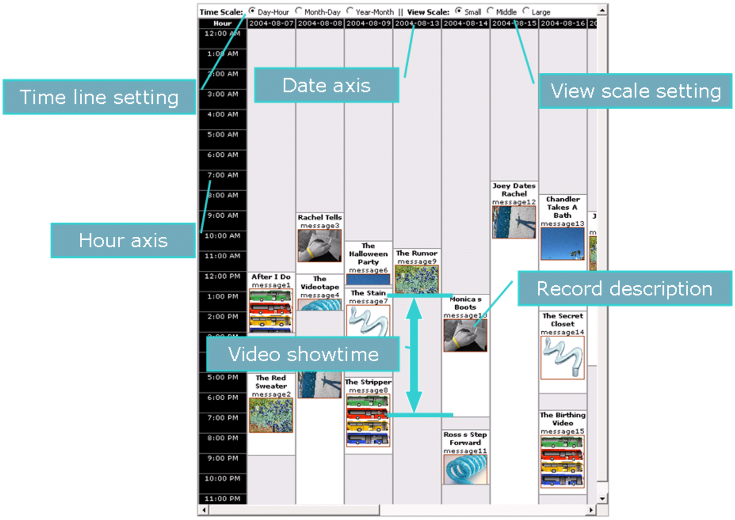

Fig. 8. Timeline control in LLM browser

the whole view of our LLM browser in Figure 3 and explain more specifically in following subsections.

5.1 Timeline Control

In LLM browser, we develop scalable timeline control interface which is shown in figure 8. In timeline setting, user can see the annotated description, corresponded video showtime in scalable view form. If user want to see larger thumnail image in time line windows, user can control the size of windows[Fig. 9]. As time goes by, user's data will increase, so that user cannot find his logged information at a glance. For more efficient search, we developed adaptive timeline control mechanism. User can select Day-hour, Month-day and Year-month pairs according to user's interest.

5.2 Map Control

We made our map control using Google Map API 2.0. There are several controls for zoom, position control, map overview and tooltip. Besides mentioned functionalities, we can display the user's logged position using custom overlay functionality. If the user's logged position can be transferred by LLM device automatically, server writes XML file which include the user's logged information and then our map control module reads the XML file to render points on the map. If user want to find logged information in some area, user can confine geographical area to select the boundary in the map by clicking the mouse button. If user click the title of map, LLM browser can show the record in timeline and

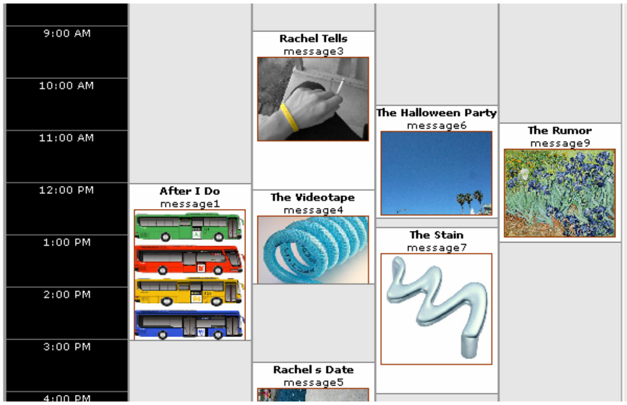

Fig. 9. View scale control in LLM browser

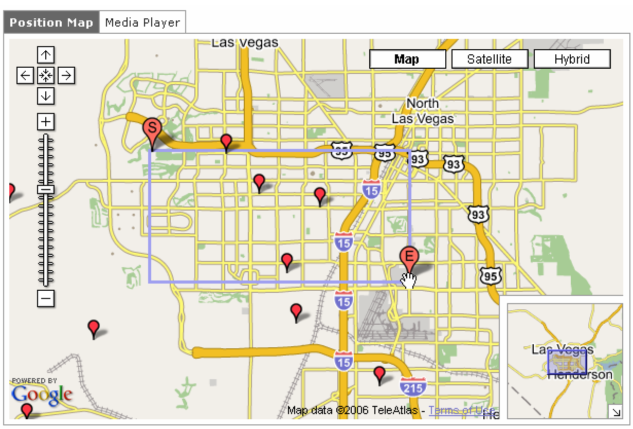

Fig. 10. Example of setting search boundary in the map

highlights it. In the same way, user click the title of timeline finds the record in the map, and also highlights itself as well.

6 Conclusion

In this paper, we present a new system for capturing and searching of life log media on networked environment. Our proposed system has four components, such as LLM device, Intelligent gadgets, LLM server and LLM client, which is connected on wireless network. LLM device is working as a gateway between intelligent gadgets and LLM server to connect them if network service is available. LLM device works as a terminal for querying and viewing the query results.

LLM server analyze the user's activity from stored LLM data. Intelligent gadgets which are attached to the daily supplies provide A/V data and multiple sensor data of user to be used for activity analysis and memory enhancement. LLM browser provides intuitive query interface to the user. For analysis of activity, we developed the learning based activity classification technique to annotate the captured LLM data. This learning based classification is a semi-automatic approach but we find it is more robust and adequate for user-dependent activity analysis. In this paper, we proposed a new platform for personal life log system. In this system, it is important that we can classify the user's activities accurately, therefore we have to investigate more robust activity classification technique in near future. We also, will make more compact and user-friendly LLM device with long battery life, which is basic problem to be solved.

References

1. Aizawa, K., Tancharoen, D., Kawasaki, S., and Yamasaki, T. 2004. Efficient Retrieval of Life Log Based on Context and Content. ACM Workshop CARPE 2004, 22–31.
2. Gemmell, J., Williams, L., Wood, K., Lueder, R., and Bell, G. 2004. Passive Capture and Ensuing Issues for a Personal Lifetime Store. ACM Workshop CARPE 2004, 48–55.
3. Mann, S. 2004. Continuous Lifelong Capture of Personal Experience with EyeTap. ACM Workshop CARPE 2004, 1–21.
4. Chambers, G., Venkatesh, S., West, G., and Bui, H. 2002. Hierarchical recognition of intentional human gestures for sports video annotation. In *In proceeding of IEEE Conference on Pattern Recognition. 2002*, Vol. 2, 2002, 1082–1085.
5. Torralba, A., Murphy, K., Freeman, T., and Rubin, M. 2003. Context-based vision system for place and object recognition. In *In Proceedings of International Conference on Computer Vision 2003*
6. Vemuri, S., Schmandt, C., Bender, W., Tellex, S. and Lassey, B. 2004. An audio-based personal memory aid. In *In Proceedings of Ubicomp 2004* Ubiquitous Computing, 2004, 400–417.
7. Kern, N., Schiele, B., and Schmidt, A. 2003. Multi-Sensor Activity Context Detection for Wearable Computing. In *In European Symposium on Ambient Intelligence 2003* 2003.
8. Randell, C., and Muller, H. 2000. Context awareness by analyzing accelerometer data. In *In proceeding of Fourth International Symoposium on Wearable Computers 2000*, 2000, 175–176
9. Tancharoen, D., Yamasaki, T., and Aizawa, K. 2005. Practical Experience Recording and Indexing of Life Long Video. ACM Workshop CARPE 2005, 61–66.
10. Kern, N., Schiele, B., Junker, H., Lukowicz, P., Troster, G., and Schmidt, A. 2004. Context Annotation for a Live Life Recording. In *In proceeding of Pervasive 2004*, 2004.