Capture and Access Tools for Event Annotation and Visualisation

R. Hunter, M. P. Donnelly, D. D. Finlay, G. Moore
Computer Science Research Institute
University of Ulster
Newtownabbey, Northern Ireland
hunter-r9@email.ulster.ac.uk,
{mp.donnelly, d.finlay, g.moore}@ulster.ac.uk

N. Booth
Parents Education as Autism Therapists (PEAT),
Belfast, Northern Ireland
nichola@peatni.org

***Abstract* – With the increasing rate in diagnosis of developmental disorders in children, the availability of therapists does not satisfy demand. Parents are themselves taking on the role of the therapist as they turn to home-based therapy to manage their childrens' behaviour. Typically, this approach is supported by a Behaviour Analyst (BA) who periodically visits, assesses progress, and set new targets for development. Having access to accurate data is critical for determining suitable targets. This model presents several challenges such as maintaining a record of behavioural events over time, and providing suitable methods of representing such records. The current paper presents a suite of conceptual tools. Mobile Annotation of eVents *In Situ* (MAVIS) is a novel mobile-based tool for parents to use during home-based therapy sessions, allowing them to annotate significant behaviours *in situ*. Using a service called VISualisation of Annotated eVEnts (VISAVE), a BA could remotely access the annotations alongside synchronised environmental sensor data from the home to aid their understanding of specific behavioural events. This would enable them to provide succinct feedback to parents remotely, and to make recommendations for the future course of therapy.**

Keywords-event annotation; sensor data capture and storage; data visualisation; mobile; interface design.

I. Introduction

Behaviour monitoring is the process of observing people's interactions within an environment. The goal is to be able to model particular behaviours, or to observe changes in behaviour over time [1]. Behaviour monitoring is particularly useful for observing the behaviour of children with developmental disorders such as autism. One application of behaviour monitoring takes place within the context of home-based autism intervention. This involves BAs providing parents with the resources to conduct a tailored course of therapy at home. Children undergo intensive one-to-one sessions with a parent in order to increase or reduce particular behaviours. Traditionally, behaviour monitoring in this instance has been conducted using pen-and-paper, and stored in a paper-based file. This traditional method is prone to error particularly when an instance of behaviour is missed while the parent's attention is diverted as they write their notes from the previous behavioural event. Paper-based records also make retrieving past data a time-consuming challenge. One approach to overcome this challenge could be to introduce technology for adding annotations *in situ* [2]. This would be well suited to situations where behaviour monitoring already involves the presence of both an observer (parent) and a person being observed (child) such as in autism intervention. Due to the nature of observing and recording events *in situ*, it is postulated that mobile-technologies could offer a specific advantage that would be applicable in behavioural event annotation. Leveraging upon developments in touch screen technology, and the portability of these devices offers opportunities to augment traditional pen-and-paper annotation approaches. However, marking the occurrence of behavioural events is not enough on its own. For a BA to identify ways in which to address behaviours, they must understand why they occur in the first place. For this, they need to recognise what precedes each instance of behaviour, and the subsequent consequence. These details are seldom included in handwritten notes, and as a result there exists a real challenge to find improved ways to help augment traditional approaches for monitoring behaviour. One method for detecting the additional events surrounding specific behaviours is to embed sensors in the home to collect complementary data. In particular, video sensors have been used in recent research to support behaviour monitoring [3-5]. Video offers unprecedented insight into a person's movement within an environment, making it well suited to the application of behaviour monitoring. Using event annotation data from an intervention session, video data can be given context by adding visual markers to a session timeline to highlight significance segments of data for review.

This paper presents a conceptual suite of tools designed to support *in situ* mobile-based event annotation and the visualisation of synchronised sensor data. The MAVIS tool supports *in situ* annotation of behavioural events on a mobile device, and stores that data alongside sensor data gathered from heterogeneous sources within the home-therapy setting. VISAVE provides a visualisation service for BAs to remotely access therapy session data so they can observe changes, and make adjustments to a course of therapy. The remainder of the paper is structured as follows: Section II reviews the current state-of-the-art that is reported in the literature. Section III presents the opportunities and challenges that have been identified. Section IV describes a conceptual capture and access system before a discussion on the direction of the future work in Section V.

II. Related Work

There has been an increase in recent research into the design of capture and access tools, specifically for behavioural data. Here, we introduce some of the related work, from early studies in digital annotation to recently developed tools.

An early annotation tool was the Experimental Video Annotator (EVA) [2]. EVA led the way for digital, text-based, annotations; improving upon the previous system for writing on paper, which required later transcription. Marquee [6] was released five years later, and took the digital annotation process one step further by providing a platform for real-time pen-based annotation on a touchscreen display. Perhaps the closest to traditional paper-based annotation methods, pen-based annotation requires a user to draw or write on a touch sensitive surface with an augmented pen. This presented a more familiar interface to help those used to traditional annotation methods to transition to a digital system. It allowed users to not only write notes, but also draw symbols and sketches, allowing for more diverse styles of annotation.

In more recent years, the Notelook [7] client application runs on wireless pen-based notebook computers. Six microphone feeds across a room are amalgamated into a single audio stream, and video is captured at 15 or 30 fps with a resolution of 640 by 480 pixels. Both the audio and video streams are stored on the server and sent across a wireless network to be accessed on a notebook computer. With the video displayed on the touch-sensitive screen, the user writes freeform notes with a stylus. In an evaluation study, four subjects (each having undergone 20 minutes of training with the application) took notes in 13 sessions over a six-week period. Each of the subjects utilised the range features provided by the application and they reported that they would use Notelook again. Future considerations for the application include improving image quality and the generation of an electronic copy of the notes made.

Wearable sensors have also been explored in capture and access tools. Walden Monitor (WM) [8] is a system composed of a head-mounted camera and a tablet PC. It is designed to allow a user to observe behaviours in 10-second intervals, each time noting the child's behaviour through buttons on the interface. This is repeated 20 times, and all of the data (including video captured during each interval) is displayed in a timeline with the behaviour annotations given at the start of each interval. While WM effectively supports the gathering of behavioural data, it does not support customisation based on previously gathered data. A main strength of WM is that the first-person view of the video footage is able to provide rich data. However, an initial trial with the system indicated users' apprehension towards wearing a head-mounted camera. As a result of the trial, feedback expressed that the tablet PC was more cumbersome to carry around when compared to a clipboard for paper-based notes.

To overcome the apprehension towards a head-mounted camera, a non-invasive wearable device is used by the Multiple Perspective Behaviour Analyser (MuPerBeAn) [3]. This platform allows the review of footage captured from miniaturised cameras embedded within glasses. Footage is captured at multiple perspectives – capturing the personal view of the child, as well as the situation's context. The videos are then synchronised by aligning them at a frame which shows an object, for example – a watch displaying the time – that has been captured at the same time at multiple views. The videos can then be reviewed on a frame-by-frame basis, and annotated accordingly.

A more recent tool, Continuous Recording and Flagging Technology (CRAFT) [4], supports the *in situ* and *post hoc* 'flagging' of behavioural events within the home environment. Parents flag incidents of problem behaviour using a custom wireless device throughout the recording session. A BA then reviews and annotates the footage *post hoc* based on their professional criteria of problem behaviour. One study with CRAFT produced an average of 12 hours of video footage for each of the eight participants. It required 10 BAs to review the footage, and to annotate it accordingly. The main strength of this system is that the two sets of annotations can be compared to note agreements and disagreements. This allows the BA to train the parent in identifying what specifically constitutes problem behaviour. It also presents several weaknesses; the vast amount of video data requires extensive storage, and the system requires many hours to review the entirety of footage.

Another recent annotation tool is the Dynamic ANnotation Tool for smart Environments (DANTE) [5], which uses a pair of stereo cameras to monitor the movements of objects within a scene by the placement of markers and sensors. Using video technology, it was reported that activities of daily living (ADLs) could be verified when the video footage was reviewed alongside the sensor data. The DANTE system provides an activity recognition platform, reducing the need for individuals to keep manual records. Three subjects, with experience in assistive technology, were involved in a trial with DANTE. They each performed two ADLs three times, resulting in 18 video clips. The objects they interacted with were fitted with sensors and markers. DANTE detected the movement of the objects through the movement of the markers within the scene. At the end of the recording session, the video files were annotated, first based on the raw sensor data, and then supported by DANTE. The results showed that using DANTE for annotation took 6.5 minutes, compared to 12 minutes without using the tool.

Based on a review of the existing literature, there does not appear to have been a mobile solution that provides a suite of tools to support both the annotation and visualisation of behavioural events in synchrony with corresponding sensor data. MAVIS has been conceptualised to draw on the strengths of previously developed event annotation tools, while addressing their limitations. VISAVE is designed to complement the mobile-based MAVIS tool, by adding a presentation layer to the sensor data for efficient review of data over time.

III. Opportunities and Challenges

While the opportunity for the capture and retrieval of sensor data is clear, there is a challenge of storage. With many hours, days, and even several weeks' worth of sensor data, there is a clear challenge in deciding on a suitable storage medium for all data generated by a capture and access system. Traditional SQL databases are commonly used for sensor data storage and present a stable platform on which to store and query large volumes of data. However, with on-going challenges of scalability and performance, NoSQL databases are gaining more attention in the area of data storage [9]. NoSQL databases are more tolerant to network partitioning, making it easier to add servers as necessary, instead of increasing the capacity of a single server as would be the case for an SQL database, which is less tolerant to network partitions.

With a broad range of sensors being used to efficiently monitor behaviour within the home, it is inevitable that the resulting output is a vast quantity of heterogeneous data, which often lack interoperability [10]. The processes for the aggregation and synchronisation of sensor data are both common challenges in data handling [11]. The installation of a Network Time Protocol (NTP) time server provides a reliable and secure time synchronisation service for a network. With a NTP installed in a capture and access system, each segment of data can have an accurate timestamp, allowing efficient review of sequential events.

A further challenge exists surrounding the understanding of sensor data. Even when it is aggregated and synchronised, the ability to extract useful information from it can be a laborious manual task requiring many hours of review. Indeed, with video sensors alone there is a continuing challenge is to find ways to annotate and analyse a large volume of video data without having to review many hours of footage [12]. Szewcyzk *et al.* describe the challenge of annotating data by sensor data analysis as time consuming and subject to error, and outline that it is an on-going challenge. [13].

Traditionally, event annotation is completed *post hoc*. It requires a complete review of all the data, which is time-consuming and laborious. While *in situ* annotation offers significant advantages, it still presents a series of challenges. During *in situ* annotation, the observer who is making annotations may have a restricted view of events, or may be have their attention averted from observing live events as they focus on entering a previous annotation [14]. This means that the annotations may be incomplete, and would require a *post hoc* review. With additional data to support annotations, it is easy to confirm or reject events based on a review of the complementary data. In addition, real-time annotation causes a tendency to annotate as quickly as possible so as to not miss any events, which may results in annotations lacking sufficient detail.

There are several methods for adding annotations to data. The most popular is text entry, where a user has to type the details of the events. Audio annotation has also been utilised in some studies. While audio presents clear advantages of speed and attention required, it requires transcription to be turned in to useful information. The transcription process is one which remains inaccurate due to the variability of voice tones, accents and languages.

There is a clear opportunity for a novel capture and access suite of tools. By providing a mobile platform for the annotation of event data, and a visualization tool for illustrating annotations and sensor data, the process of behaviour analysis can be improved both from time and cost perspectives. This paper presents a conceptual system that aims to address the challenges and opportunities that have been discussed.

IV. System Description

The conceptual use of MAVIS and VISAVE are shown in Figure 1. What follows is a detailed discussion of the system components.

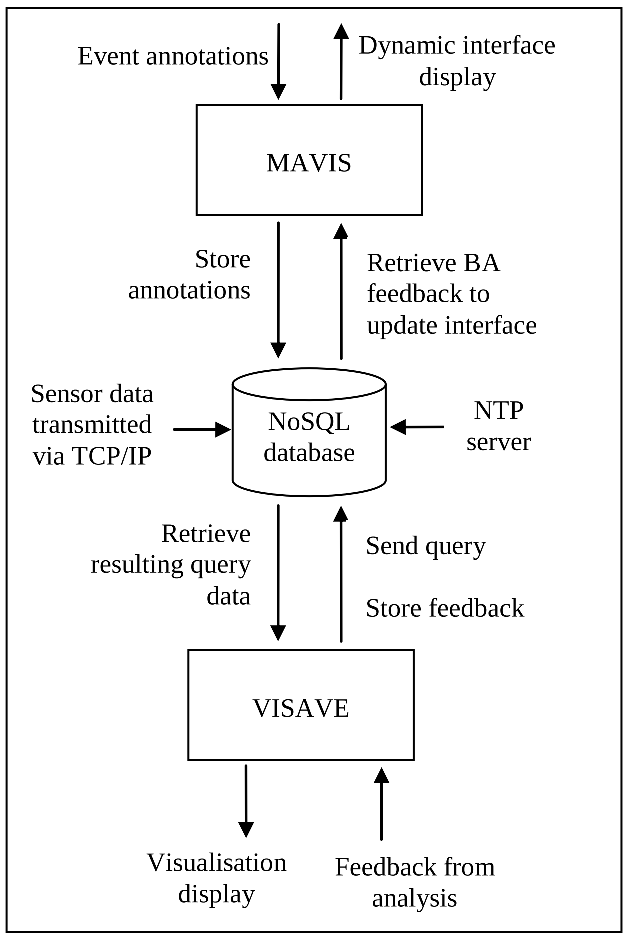

Figure 1 - A sequence diagram showing how data flows through the system.

A. Data Capture

Annotations are captured through interaction with the MAVIS tool in real-time. A parent loads the MAVIS tool through a web browser on their mobile device, and logs in with their username and password credentials. They are then presented with a dynamically generated interface with pre-defined buttons based on previous BA feedback. The buttons are labelled with specific events or behaviours being monitored at any one point throughout the course of therapy. When a parent touches a button on screen, the click event is detected and an annotation is initiated. The timestamp for the click event is retrieved from the NTP server, and the type of behaviour is determined based on the corresponding button label. This annotation data is added in real-time to an XML file stored on the server. At the end of the session, the XML file and the details of the session (including start and end time, and the child's client number) are sent to be stored in a database for later access.

In addition to annotations, the capture of sensor data is required to provide additional information about events. As sensor data is captured it is transmitted to the server via TCP/IP. The use of video sensors to record events as they occur is the most convenient way to provide a method for reviewing and analyzing behaviour events accurately. With the ability to capture a scene from several viewpoints simultaneously and store that data for later retrieval, video offers valuable data, which no other sensor type can provide [15]. Depending on the specific requirements for each child, an array of sensors can be used with the system such as physiological sensors to allow the automatic annotation of physiological events which may be unseen, such as a change in heart rate. Sensor-augmented toys can provide insight into how a child interacts with certain objects. For example, a parent may note that a child has thrown a toy, and with additional accelerometer data, it would be possible for the BA to determine the force with which it was thrown. When the session is initiated through the MAVIS tool, a message is sent to the server to signal the sensors to begin to record sensor data. Sensor data is captured and stored directly to the database through the network.

B. Data Storage

For a novel approach to the storage of sensor data and annotations, the NoSQL database CouchDB [16] will be implemented for the system. This will give the advantages of scalability and performance over traditional relational databases (RDBMS). CouchDB will allow the storage of heterogeneous sensor data within the database without having performance issues that would commonly be caused by video data, for example, when stored in an RDBMS.

The data stored in the database consists of parental data including login credentials and contact information; session data including start and end time, and data about the child; annotation data including event type and a timestamp; sensor data including sensor type, sensor data and a timestamp; and feedback data indicating the progress of each child.

C. Data Access

With the vast amount of sensor data gathered from heterogeneous sources, there is a requirement to have a service in place for retrieving and visualising relevant data to produce useful information. VISAVE is a web-based application that has access to the database containing session data archives for each user. The service is accessed by the BA remotely on a desktop computer. They access the data by logging in with their credentials. A list of MAVIS users with which they are associated is given, and they select the name of the child requiring review. Depending on the specified course of therapy for that individual child, the BA identifies the events they wish to review. They can customise their selection by behavioural event type, by session date, and by timescale. Depending on their selection, they will be presented with a single-screen interactive visualisation illustrating the annotations and synchronised sensor data. They are able to navigate quickly to areas of importance as highlighted by the annotation markers along a timeline, and observe the surrounding sensor data. This can help the BA to understand why certain behaviours may occur; including observation of environmental factors or other events which may be triggering repeat behaviours. After reviewing the data, they decide if a change in the course of therapy is to be recommended, or if they require the parents to annotate additional events during future sessions. They can input this information as feedback for the parents to review when they access the tool.

D. Interface Considerations

The prototype interface for MAVIS (Figure 2) has been designed to allow even novice users to efficiently utilise the tool *in situ*. As mentioned in the description, parents do not need to set up the tool for each session; they simply have to follow instructions from the BA, and note events as they occur through on screen buttons. The configuration is carried out by the BA, who specifies the parameters for the interface through the VISAVE tool, ahead of a therapy session. The simple button interface allows a parent to maintain their attention on conducting the therapy, as a touch of a button will not avert their attention nor prove to be a distraction for the child.

The key design consideration for the VISAVE interface (Figure 3) is to consider ways in which to display a large volume of heterogeneous data in a single visualisation. The data that a BA requests is to be presented full screen so in a single glance they can tell if the current course of therapy is altering a child's behaviour, and the rate at which that change is occurring. It is envisaged that one or more timelines will display markers for annotations and significant sensor data changes, allowing a BA to quickly browse to a segment of interest. The surrounding event data is played back in synchrony in windows on the interface. The interface layout will adjust accordingly for multiple sensor streams. VISAVE will also have a feedback facility where a BA can type notes to the parents, to guide them in a suitable course of therapy. It will support a range of personalisation features to allow the BA to update the MAVIS interface with suitable data for a specific therapy session. This feedback would be sent to the server for MAVIS to retrieve when next accessed.

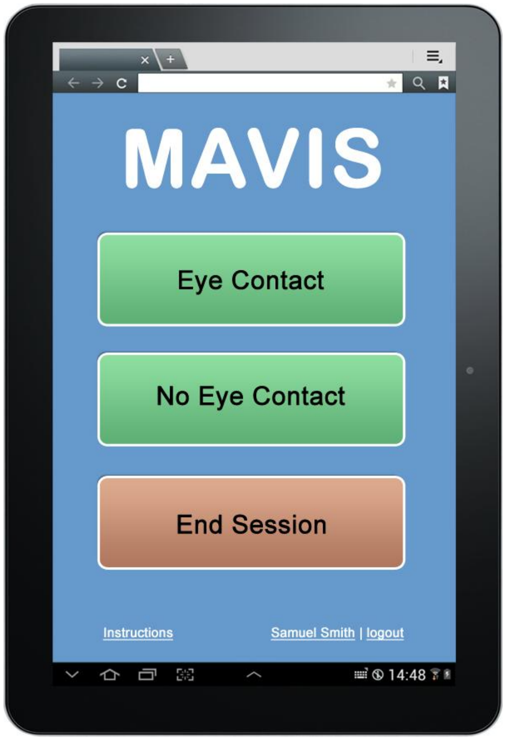

Figure 2 – The MAVIS prototype interface displayed on a touchscreen tablet computer.

Figure 3 – The conceptual VISAVE interface displayed on a desktop computer.

E. Case Study Presentation

To help contextualise the potential utilisation of the proposed system, an autism intervention scenario is presented in the following case study.

John is an 8-year old boy who has been diagnosed with autism. He has impairments in social interaction and communication; notably struggling with maintaining attention, eye contact, and speech development. His father, Samuel, has been researching his condition, and has decided that he wants to take on the role of a therapist at home, and guide John through a program of interventions designed for him by Laura, a qualified BA.

Laura attends the home to meet John, and prepares a series of interventions to address his impairments. She designs an intervention to improve John’s eye contact, and gives Samuel a program to follow once a day. Samuel is to sit across from John at a table, and say “John, look at me”. If he follows the command within 10 seconds, he receives verbal praise and a reward. If he does not, Samuel must repeat the phrase again. This process should continue for 5 minutes during each intervention.

Laura specifies the text for event buttons via VISAVE on her laptop. For the eye contact intervention, she wants to specifically see positive and negative responses, so adds two buttons. She shows Samuel how to use MAVIS on his personal tablet, and points out the button interface that has been set up to indicate the behaviours he should annotate.

The dining room in their house is set up with two cameras pointing at the table, and Laura has provided a set of sensor-augmented objects to use during the therapy. At the start of the session, Samuel presses a ‘Start Session’ button, which triggers the start of the capture of sensor data. During the intervention, he uses the MAVIS tool on his tablet to annotate John’s responses.

The following week, Laura logs on to the VISAVE service and reviews the data. At a glance, she is able to observe how often John followed Samuel’s command during each intervention session, and whether or not his attention remained on the task or if he was interacting with the sensor-augmented objects. The visualisation shows the video and other sensor data in synchronisation with Samuel’s annotations. By clicking through the annotations on the video, Laura can observe the eye contact taking place, and the surrounding conditions. She is able to identify that John is showing an improvement in eye-contact, and can recommend a slight change to the intervention to make further improvement. The following week, Samuel is to repeat the intervention, but now present John with a toy at the table before repeating the “look at me” command.

The process of regularly adjusting the course of intervention helps John’s impairments to improve. Samuel continues his role as a therapist at home, and Laura’s remote input ensures the interventions are effective.

 ISBN: 978-1-61208-289-9

V. CONCLUSION

MAVIS is currently in prototype phase. As development continues, the next step will be to utilise MAVIS to annotate pre-recorded events post hoc. By comparing the processes for online and offline annotation, the effectiveness of the MAVIS tool will be realised. A usability study will follow where users will experiment with the tool for testing and analysis purposes. The interface will be updated based on the feedback provided. A further study will explore the methods for storing heterogeneous sensor data alongside annotations, as well as effective query and retrieval methods. The VISAVE service will need to be developed with user needs in mind. It will be adapted to suit any range of queries that a BA may require. With testing complete, the entire MAVIS system will be trialled across a suitable cohort.

REFERENCES

[1] T. Ploetz et al., "Automatic Assessment of Problem Behaviour in Individuals with Developmental Disabilities," Proceedings of the 14th ACM International Conference on Ubiquitous Computing, Pittsburgh, USA, Sep. 2012, pp. 391-400.

[2] W. E. Mackay and G. Davenport, "Virtual Video Editing in Interactive Multimedia Applications," Communications of the ACM, vol. 32, no. 7, Jul. 1989, pp. 802-810.

[3] P. Alt, B. Pfleging, A. Bungert, A. Schmidt, and M. Havemann, "Supporting Children with Special Needs through Multi-Perspective Behaviour Analysis," Proceedings of the 10th International Conference on Mobile and Ubiquitous Multimedia, Beijing, China, Dec. 2011, pp.81-84.

[4] N. Nazneen et al., "Supporting parents for in-home capture of problem behaviours of children with developmental disabilities," Personal and Ubiquitous Computing, vol. 16, no. 2, Feb. 2012, pp. 193-207.

[5] M. Donnelly, T. Magherini, C. Nugent, F. Cruciani, and C. Paggetti, "Annotating Sensor Data to Identify Activities of Daily Living," Toward Useful Services for Elderly and People with Disabilities - Proceedings of the 9th International Conference on Smart Homes and Health Telematics, Montreal, Canada, Jun. 2011, pp. 41-48.

[6] K. Waber and A. Poon, "Marquee: A Tool For Real-Time Video Logging," Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 94), Boston, USA, Apr. 1994, pp.58-94.

[7] P. Chiu, A. Kapuskar, S. Reitmeier, and L. Wilcox, "NoteLook: Taking Notes in Meetings with Digital Video and Ink," Proceedings of the 7th ACM International Conference on Multimedia, Orlando, USA, Nov. 1999, pp. 149-158.

[8] G. R. Hayes, J. A. Kientz, K. N. Truong, D. R. White, and G D. Abowd, T. Pering, "Designing Capture Applications to Support the Education of Children with Autism," Proceedings of the 6th international conference on Ubiquitous Computing, Nottingham, UK, Sep. 2004, pp. 161-178.

[9] J. S. van der Veen, B van der Waaij, and R. J. Meijer, "Sensor Data Storage Performance: SQL or NoSQL, Physical or Virtual," Proceeding of the 5th IEEE International Conference on Cloud Computing (CLOUD), Honolulu, USA, Jun. 2012, pp. 431-438.

[10] H. A. McDonald, C. D. Nugent, G. Moore, and D. D. Finlay, "An XML Based Format for the Storage of Data Generated both inside and outside of a smart home environments," in Proceedings of the 10th IEEE International Conference on Information Technology and Applications in Biomedicine, Corfu, Greece, Nov. 2010, pp.1-4.

[11] H. Lee, K. Park, B. Lee, J. Choi, and R. Elmasri, "Issues in data fusion for healthcare monitoring," in Proceedings of the 1st International Conference on Pervasive Technologies Related to Assistive Environments (PETRA), Athens, Greece, Jul. 2008, Article no.3.

[12] O. Duchenne, I. Laptev, J. Sivic, F. Bach, and J. Ponce, "Automatic annotation of human actions in video," in Proceedings of IEEE 12th International Conference on Computer Vision, Kyoto, Japan, Oct. 2009, pp. 1491 - 1498.

[13] S. Szewcyzk, K. Dwan, B. Minor, B. Swedlove, and D. Cook, "Annotating Smart Environment Sensor Data for Activity Learning," Technology and Healthcare, vol. 17, no. 3, Aug. 2009, pp. 161-169.

[14] P. Alt, B. Pfleging, A. Bungert, A. Schmidt, and M. Havemann, "Supporting Children with Special Needs through Multi-Perspective Behaviour Analysis," in Proceedings of the 10th International Conference on Mobile and Ubiquitous Multimedia (MUM 11), Beijing, China, Dec 2011, pp. 81-84.

[15] D. Ding, R. A. Cooper, P. F. Pasquina, and L. Fici-Pasquina, "Sensor technology for smart homes," Maturitas, vol. 69, no. 2, Mar. 2011, pp. 131-136.

[16] J. Lennon, "Exploring CouchDB:A document-oriented database for Web applications", 31 Mar 2009 [Online]. Available: http://www.ibm.com/developerworks/opensource/library/os-couchdb/ [Accessed on 2013-03-01]

 ISBN: 978-1-61208-289-9