J Ambient Intell Human Comput
DOI 10.1007/s12652-015-0328-1

ORIGINAL RESEARCH

Vision-based patient monitoring: a comprehensive review of algorithms and technologies

Supriya Sathyanarayana[1] · Ravi Kumar Satzoda[2] · Suchitra Sathyanarayana[2] · Srikanthan Thambipillai[1]

Received: 7 July 2015 / Accepted: 3 November 2015

Abstract Vision-based monitoring for assisted living is gaining increasing attention, especially in multi-modal monitoring systems owing to the several advantages of vision-based sensors. In this paper, a detailed survey of some of the important vision-based patient monitoring applications is presented, namely (a) fall detection (b) action and activity monitoring (c) sleep monitoring (d) respiration and apnea monitoring (e) epilepsy monitoring (f) vital signs monitoring and (g) facial expression monitoring. The challenges and state-of-art technologies in each of these applications is presented. This is the first work to present such a comprehensive survey with the focus on a set of seven most common applications pertaining to patient monitoring. Potential future directions are presented while also considering practical large scale deployment of vision-based systems in patient monitoring. One of the important conclusions drawn is that rather than applying generic algorithms, use of the application context of patient monitoring can be a useful way to develop novel techniques that are robust and yet cost-effective.

Keywords Survey · Patient monitoring · Elderly care · Remote monitoring · Computer vision

✉ Supriya Sathyanarayana
supriya001@e.ntu.edu.sg

Ravi Kumar Satzoda
rsatzoda@eng.ucsd.edu

Suchitra Sathyanarayana
ssathyanarayana@ucsd.edu

Srikanthan Thambipillai
astsrikan@ntu.edu.sg

1 Nanyang Technological University, Singapore, Singapore

2 University of California, San Diego, USA

1 Introduction

In the recent past, assisted living systems have been gaining increasing importance (Rashidi and Mihailidis 2013). The main factors contributing to this trend are (a) shortage of nursing staff for continuous monitoring of patients (Needleman et al. 2002) and the elderly, and (b) the increasing rate of growth of the elderly population (UN 2012). Around 89 % of the elderly prefer to live independently and age in their homes. Studies on age-related diseases reinforce the need for developing technologies to assist the elderly (Rashidi and Mihailidis 2013).

One of the major components of assisted living systems is patient monitoring. Functions of patient monitoring systems (PMS) include measuring and monitoring physiological medical parameters (Paradiso 2003), remote monitoring of patients' activities, breathing, sleep etc. (Rashidi and Mihailidis 2013).

There exist a variety of PMS and most of them are particularly designed for tracking the physiological parameters (Pantelopoulos and Bourbakis 2010), which mostly require intrusive and contact-based sensing technologies. While such monitoring systems are necessary because they directly monitor the physiological parameters, computer vision offers a complementary monitoring system using non-contact based vision sensors or cameras. With decreasing costs and increasing miniaturization of cameras, vision-based PMS are being explored to assist medical staff to monitor patients and their activities. This is motivated further because of the increasing pervasiveness of smart phones and tablets that are equipped with high resolution cameras and high speed computing systems. Therefore, such systems can be deployed to monitor the patients remotely as well (Mihailidis et al. 2004).

Vision-sensors can be deployed to play an important role in monitoring a patient. Typically, a patient is either

Published online: 26 November 2015

under observation or undergoing treatment or under post-operative care. In these scenarios, the patient's vital signs are recorded at intervals and monitored (Johnson et al. 2014). Yet, continuous monitoring can play a crucial role in detecting and alerting in case of any abnormalities or emergencies during the intervals between the vital-signs' recording. Detecting patient's conditions based on visual cues such as body movements or facial expressions using an automated vision based system makes monitoring a more efficient process.

In this paper, we present a detailed literature survey on different vision-based systems that have been designed for patient monitoring (or for elderly care) in the recent past, mainly in the past decade. The vision-based sensors considered in this study include video cameras, RGB-D cameras (RGB-depth cameras), infrared (IR) sensors and time-of-flight cameras. Data collected by any (or a combination) of these sensors are processed aiding in the various patient monitoring applications. Considering that there are a variety of parameters that can be monitored using vision sensors, the scope of this survey is limited to the following seven main applications: (a) fall detection, (b) action and activity monitoring (c) sleep monitoring (d) Respiration and apnea monitoring (e) Epilepsy monitoring (f) vital signs monitoring and (g) facial expression monitoring. For each of these applications, a detailed survey of the challenges and techniques will be presented.

The main contributions of the paper can be summarized as follows: (a) This is the first work to present a comprehensive survey with the focus on a set of seven most common applications in vision-based patient monitoring; (b) In this survey, for some patient monitoring applications such as fall detection, sleep monitoring etc., a survey of technologies and challenges is presented, which is a different way of reviewing algorithms as compared to other existing surveys for these topics; (c) Though topics such as facial expression monitoring and activity monitoring have been studied previously in a generic context of computer vision and machine learning, in this work, we are specifically surveying them in the context of patient monitoring; (d) This is the first survey which comprehensively reviews some topics such as vital signs monitoring and depression monitoring; (e) This survey highlights the extent to which there has been progress in certain applications with regards to vision-based patient monitoring and outlines possible future directions and challenges that are yet to be addressed.

2 Overview of the proposed survey

In this paper, we present a detailed survey of seven important applications in patient monitoring as mentioned in Sect. 1. An overview of the technologies/techniques in each of the applications and the related works are summarized in Table 1. Fall detection, sleep monitoring and epilepsy monitoring are categorized based on the sensor-based technologies that are employed. Breathing monitoring and sleep apnea detection are classified based on the type of approaches that are used in the proposed methods. Facial expression monitoring is divided further as pain monitoring, depression detection, facial expression detection for medical condition detection and monitoring the after-effects of drugs. Pain and depression detection are classified based on the approaches used. The section on facial expression detection for medical condition detection will comprise of techniques for different medical conditions that have been addressed in literature. Similarly, vital signs monitoring is classified based on approaches used. Activity monitoring is divided into: pose monitoring, monitoring activities of daily living and unusual activity detection. Pose monitoring is classified based on approaches used and monitoring activities of daily living is classified based on sensing modalities that are used.

3 Fall detection

Fall detection and accident monitoring have been of great interest among the remote patient monitoring applications. This is because they are a major cause for fatal injury, especially for the elderly (Mubashir et al. 2013). While various sensor based approaches have been explored for fall detection, vision based approaches have attracted special interest due to the advantages of vision based sensors over others (Mubashir et al. 2013). Surveys on vision-based fall detection can be found in Mubashir et al. (2013), Hijaz et al. (2010), Yu (2008), Noury et al. (2007), Igual et al. (2013). In this section, we shall review the technologies and the extent to which the major challenges have been addressed, and scope for future work in fall detection.

3.1 Technologies

In this section, a survey of fall detection technologies proposed in the last decade are discussed and classified based on the vision-sensor system used.

3.1.1 Multi-camera systems

Multi-camera systems for fall detection in single room scenarios are designed to provide multi-view information (Zambanini et al. 2010; Auvinet et al. 2011; Anderson et al. 2009; Fleck and Strasser 2008; Rougier et al. 2011b; Nicolas Thome 2008; Tabar et al. 2006; Hazelhoff et al. 2008; Cucchiara et al. 2007; Auvinet et al. 2008). Actions

Table 1 Overview of the survey paper

PMS applications	Technologies/approaches	
Fall detection	Multi-camera systems	Zambanini et al. (2010), Auvinet et al. (2011), Anderson et al. (2009), Fleck and Strasser (2008), Rougier et al. (2011b), Nicolas Thome (2008), Tabar et al. (2006), Hazelhoff et al. (2008), Cucchiara et al. (2007), Nicolas Thome (2008), Auvinet et al. (2008), Wang et al. (2013)
	Monocular systems	Nait-Charif and McKenna (2004), Toreyin et al. (2006), Nasution and Emmanuel (2007), Mirmahboub et al. (2013), Yu et al. (2012), Makantasis et al. (2012), Rougier et al. (2013)
	Infrared and range sensor based systems	Grassi et al. (2008), Diraco et al. (2010), Leone et al. (2011), Rougier et al. (2011a), Mastorakis and Makris (2012), Ni et al. (2012)
	Bio-inspired vision sensor based systems	Humenberger et al. (2012), Belbachir et al. (2012), Belbachir et al. (2011), Fu et al. (2008)
Sleep monitoring	Infrared camera based systems	Yu et al. (2013), Chen et al. (2014), Liao and Yang (2008), Chen et al. (2014), Peng et al. (2006), Yang et al. (2003), Kayyali et al. (2008), Espa et al. (2000)
	Pressure-sensor based systems	Malakuti and Albu (2010)
	CCD video camera based systems	Nakajim et al. (2001)
Breathing monitoring	Thermal imaging of nostril region	Murthy et al. (2004), Zhu et al. (2005), AL-Khalidi et al. (2010), Fei and Pavlidis (2006), Johnson et al. (2007)
	Tracking movement of chest and abdomen	Kuo et al. (2010), Martinez and Stiefelhagen (2012), Kroutil et al. (2011), Sato and Nakajima (2005), Nakajima et al. (2000), Frigola et al. (2002), Wiesner and Yaniv (2007), Aoki et al. (2005), Sato and Nakajima (2005), Martinez and Stiefelhagen (2012), Bai et al. (2010), Bai et al. (2012), Kroutil et al. (2011)
Sleep apnea detection	Tracking movement of chest and abdomen	Takemura et al. (2005), Wang et al. (2006), Wang et al. (2014), Rodrigues et al. (2007), Falie et al. (2009), Falie et al. (2008), Wang et al. (2010), Wang et al. (2007)
	Thermal imaging of nostril region	Fei et al. (2009)
Epilepsy monitoring	Video-camera based methods	Li et al. (2002), Cuppens et al. (2009), Cuppens et al. (2010), Kalitzin et al. (2012), Pediaditis et al. (2012), Lu et al. (2013), Liu et al. (2004)
	3D vision based methods	Cuppens et al. (2012)
Facial expression monitoring		
Pain detection and monitoring	Appearance based	Littlewort et al. (2007), Gholami et al. (2009), Nanni et al. (2010), Sikka et al. (2014)
	Appearance and geometry based	Ashraf et al. (2007), Lucey et al. (2011a), Hammal and Cohn (2012), Hammal and Cohn (2012), Lucey et al. (2012), Werner et al. (2013), Sikka et al. (2014), Khan et al. (2013)
Depression monitoring	Subject-dependent approach	McIntyre et al. (2009), Alghowinem et al. (2013b), Alghowinem et al. (2013a), Cohn et al. (2009)
	Subject-independent approach	Joshi et al. (2012), Joshi et al. (2013b), Joshi et al. (2013a)
Facial expression detection for medical condition detection	Different techniques to detect the various medical conditions	Bevilacqua et al. (2011), Hamm et al. (2011), Wang et al. (2008), Alvino et al. (2007), Wang et al. (2008), Bevilacqua et al. (2011), Hamm et al. (2011), Naufal Mansor et al. (2010), Bin Mansor et al. (2010), Dai et al. (2001)
Monitoring the after-effect of drug	3D model and optical flow based methods	Rogers et al. (2007), Brusco and Paviotti (2005)
Vital signs monitoring	Thermal imaging based methods	Sun et al. (2005), Chekmenev et al. (2005), Gault et al. (2010), Gault and Farag (2013)
	Color image processing based methods	Lewandowska et al. (2011), Kwon et al. (2012), Poh et al. (2010), Li et al. (2014), Poh et al. (2010), Kwon et al. (2012), Balakrishnan et al. (2013)
	Motion based methods	Balakrishnan et al. (2013)
Activity monitoring		
Pose monitoring	Blob based approaches	Brulin et al. (2012), Jansen et al. (2007)

Table 1 continued

PMS applications	Technologies/approaches	
	Skeletonization based approach	Obdrzalek et al. (2012)
Activities of daily living	Multi-sensor based systems	Zouba et al. (2007), Bieber et al. (2009), Junior et al. (2012), Amoretti et al. (2011), Matic et al. (2010)
	Depth sensor based systems	Zhang and Tian (2012), Banerjee et al. (2014)
	Video camera based systems	Cheng et al. (2011), Banerjee et al. (2014), Zhan et al. (2012), Olivieri et al. (2012)
Unusual activity detection	Techniques	Zhong et al. (2004), Khan and Sohn (2011)

appear different depending on the viewpoint and angle of the camera and hence multi-camera systems have the advantage of cameras complementing each other for a more robust detection. In most multi-camera systems, the simple image features that are extracted or the independently processed decisions output at the individual camera nodes are then combined using a fusion unit.

3.1.2 Monocular digital camera systems

In order to address the concerns of multi-camera systems such as calibration and synchronization, monocular systems have been proposed in the recent past, that achieve 3D space measures (Mirmahboub et al. 2013; Makantasis et al. 2012; Rougier et al. 2013). In Mirmahboub et al. (2013), silhouette area is used to define view-invariant features for classifying falls and Yu et al. (2012) defines features from the ellipse fitted around the extracted silhouette, and projection histograms. Rougier et al. (2013) views the head as a 3D ellipsoid, which is projected on to the 2D image plane, and tracked with particle filter. In Makantasis et al. (2012), the 3D space measures are achieved through camera calibration and inverse perspective mapping. In methods such as Toreyin et al. (2006), the camera is mounted to the side of the wall and the dimensions of the bounding box enclosing the individual in the image are measured before and after the fall and compared. In other monocular systems, the camera is usually mounted on the ceiling (Nait-Charif and McKenna 2004) or higher in the room (Mirmahboub et al. 2013; Rougier et al. 2013) in order to obtain a wide field of view and avoid occlusions.

3.1.3 Infrared and time-of-flight camera systems

Grassi et al. (2008), Diraco et al. (2010), Leone et al. (2011) use 3D time of flight range cameras for fall detection. The depth camera built based on the TOF (time-of-flight) principle uses a near infrared illuminator in these methods.

In the RGBD camera systems (Rougier et al. 2011a; Mastorakis and Makris 2012; Ni et al. 2012), the depth images generated by the Kinect's infrared camera are processed. Mastorakis and Makris (2012) detects falls based on the change in height and width of a 3-D bounding box. Rougier et al. (2011a) analyzes the 3D trajectory of the person's centroid to detect falls. Diraco et al. (2010) combines the information on the centroid with the orientation of the body spine.

3.1.4 Bio-inspired vision sensor based systems

Humenberger et al. (2012), Fu et al. (2008), Belbachir et al. (2012), Belbachir et al. (2011) use bio-inspired stereo vision where two optical detector chips with event-driven pixels that are sensitive to relative light intensity changes are used. The approach used in Humenberger et al. (2012) is that as there is insufficient explanation to exactly describe a fall and hence define it through an algorithm, a fall is learned using a neural network. In Fu et al. (2008) a single asynchronous temporal contrast vision sensor that reports a fall with ten times higher temporal resolution.

3.2 Discussion

We now visit the challenges in fall detection and survey the technologies with respect to these challenges.

3.2.1 Invariance to viewpoint

Monocular systems (Nait-Charif and McKenna 2004; Toreyin et al. 2006; Nasution and Emmanuel 2007; Mirmahboub et al. 2013; Yu et al. 2012; Makantasis et al. 2012; Rougier et al. 2013) for fall detection have evolved over time with improvements in robustness to changes in view-point. View-invariant features are proposed in Yu et al. (2012) and Mirmahboub et al. (2013). Multi-camera systems are inherently designed to handle the view point invaraince. Nicolas Thome (2008) shows that a two-camera system is sufficient to achieve viewpoint invariance. In bio-inspired stereo vision fall detection systems (Humenberger et al. 2012; Belbachir et al. 2012, 2011), this challenge is not specifically addressed, while in other TOF camera based systems (Diraco et al. 2010; Leone et al. 2011), this challenge is addressed.

3.2.2 Handling occlusion

Handling partial or full occlusions is one of the main challenges faced by fall detection algorithms when applied to real-life scenarios. Some methods address partial occlusions (Nicolas Thome 2008), whereas some are designed to be occlusion-resistant (Auvinet et al. 2011; Rougier et al. 2011a). Multi-camera based methods (Auvinet et al. 2011; Anderson et al. 2009) extract information from the scene from different view-points and hence aid in coping with occlusions. Nait-Charif and McKenna (2004) use ceiling mounted wide angle cameras to avoid interference of occlusions. However, the authors do not discuss how the activity patterns have been defined. Depth-based methods have a primary advantage of the ability to handle occlusions (Grassi et al. 2008; Diraco et al. 2010; Leone et al. 2011). Rougier et al. (2011a) uses centroid velocity of the person to even detect occluded falls (person is totally occluded at the end of the fall).

3.2.3 Trade-off between real-time performance, detection rate and cost of implementation

Efforts have been made to propose fall detection methods that are real-time, low-cost and that also achieve the required detection rate. Zambanini et al. (2010) and Hazelhoff et al. (2008) use network of low resolution cameras and perform 3D reconstruction, the emphasis being on low computational effort and fast processing. Zambanini et al. (2010) achieves a high detection rate, while Hazelhoff et al. (2008) has relatively lesser detection rate. Belbachir et al. (2012) uses a stereo set-up with on-board processing for real-time fall detection. Rougier et al. (2011b) reduces costs by involving inexpensive IP cameras. TOF cameras provide more accurate depth map than stereo-vision systems, but are currently very expensive (Rougier et al. 2011a). On the other hand, Kinect sensor based systems provide lesser accurate depth information at the boundary of objects and for objects that are very far away from the sensor, though they are lesser expensive than TOF cameras (Rougier et al. 2011a). While 3D range cameras are expensive, bio-inspired stereo sensors have an added advantage of being low cost and well-suited for embedded realization. Humenberger et al. (2012), Belbachir et al. (2012), Belbachir et al. (2011) also have on-board processing for real-time performance. A computationally less complex algorithm leading to a real-time, low cost method can be found in Fu et al. (2008).

3.2.4 Night and low light conditions

Significant percentage of falls take place during night time (Zhang et al. 2011) and hence it is imperative to address the challenge of detecting falls during night time and low lighting conditions. Rougier et al. (2011a), Mastorakis and Makris (2012), Zhang et al. (2011) use infrared sensor, which generate depth maps irrespective of the lighting conditions. Leone et al. (2011) uses a range camera that provides metric information under all illumination conditions.

3.2.5 Privacy

The output of conventional-video camera based automated fall detection systems (Mirmahboub et al. 2013; Yu et al. 2012; Makantasis et al. 2012; Rougier et al. 2013) is only an alarm to indicate the occurrence of fall, and not the video frames themselves. For instance, since fall detection is performed in real-time (Zambanini et al. 2010), the frames are not stored, and when a fall occurs, an anonymous snapshot showing only the silhouette of the person in the scene is taken. Yet, video-based approaches are more prone to privacy issues (Igual et al. 2013) and must ensure that the privacy of the users is protected. However, with the advancements in infrared imaging and 3D vision sensor technology, handling privacy issues has been a lot more effective than using the conventional 2D camera based systems since they do not process any chromatic information that may carry chances of revealing the identity of the individual (Humenberger et al. 2012; Rougier et al. 2011a; Belbachir et al. 2011). 3D imaging methods (Rougier et al. 2011a) protect the privacy of the individual owing to the fact that only target depth is processed in such methods. In bio-inspired vision sensor based systems (Belbachir et al. 2012), the privacy of the person is systematically ensured since the event-driven dynamic vision sensor based detector does not produce real images such as classic video sensors. In Fu et al. (2008), detailed visual appearance of the patient are filtered out and no image snapshot is taken.

3.2.6 Datasets

Most of the evaluation of fall detection algorithms is done on simulated datasets and the challenges of real-life scenarios are yet to be fully tackled (Igual et al. 2013). Since most fall detection methods are designed specific to scenarios, Igual et al. (2013) explains the importance of a generic fall detection algorithm. The need for comprehensive fall detection

Table 2 Summary of sensor-systems used in vision-based fall detection

Type of system		Features	CM	Det. rate	Platform	Real-time	Occlusions	View-point changes	Real-world set-up
Network of cameras	Zambanini et al. (2010)	Bounding-box aspect ratio, orientation, axis ratio and motion speed extracted from voxel space	Fuzzy logic	AUC 0.974	NS	yes	NS	Yes	No
	Auvinet et al. (2011),	3D volume (vertical volume distribution ratio)	Threshold crossed by the VVD (vertical volume distribution)	99.7 % sensitivity	GPU	16 fps	Yes	Yes	Yes
	Auvinet et al. (2008)	Centroid of silhouette	Blob analysis	100 %	Laptop with GPU	0.3 seconds to detect fall	Yes	Yes	Yes
	Fleck and Strasser (2008)	Appearance, position features, other features	SVM	NS	Xilinx Spartan 3 FPGA, PowerPC processor	8–10 fps	NS	Yes	Yes
	Anderson et al. (2009)	Voxel	Fuzzy logic	100 %	NS	NS	Yes	Yes	Yes
	Cucchiara et al. (2007)	Projection histograms	HMM (hidden Markov model)	97.23 %	NS	Real-time	Yes	Yes	Yes
	Rougier et al. (2011b)	Shape	GMM (Gaussian mixture model) and shape matching	0.978 AUC	PC	5 fps	Yes	Yes	Yes
	Nicolas Thome (2008)	Ellipse fitted around silhouette	Fuzzy logic and LHMM (layered hidden Markov model)	100 %	PC	36 ms	Partial	Yes	Yes
	Hazelhoff et al. (2008)	Principal component, variance ratio	Gaussian multi-frame classifier	More than 85 %	NS (low cost)	15 fps	Yes	Yes	Yes
	Tabar et al. (2006)	Edges, color	Threshold based	94 %	Inexpensive hardware	NS	NS	Yes	Yes
Infrared sensor	Mastorakis and Makris (2012)	Dimensions of 3D bounding box, its first derivatives	Threshold-based	NS	PC	0.3–0.4 ms	NS	Yes	Yes
	Rougier et al. (2011a)	Centroid position and velocity	Threshold-based	98.7 % success rate	PC	Real-time if implemented in C/C++	Yes	Yes	Yes
Time-of-flight cameras	Leone et al. (2011)	Position and velocity vectors of body's centroid	Threshold-based	97.3 % specificity and 80 % sensitivity	PC	8 fps	Partial	Yes	Yes
	Grassi et al. (2008)	Distance of centroid of blob from floor	Threshold-based	97.3 % reliability & 80 % efficiency for threshold of 0.4 m	Linux-based architecture	NS	Partial	No	No

Table 2 continued

Type of system		Features	CM	Det. rate	Platform	Real-time	Occlusions	View-point changes	Real-world set-up
	Ni et al. (2012)	MHI, HOG, HOF features	Multiple kernel learning (MKL) framework	98.76 %	PC	10 fps	NS	No	Yes
	Jansen and Deklerck (2006)	Log-polar histograms of vectors between edge pixels	Chi-square distance	NS	NS	NS	Yes	NS	NS
	Planinc and Kampel (2013)	Orientation of person's major axis and height of spine	Threshold-based	0.958 Accuracy, 0.925 Recall	PC	30 fps	NS	NS	No
Bio-inspired Vision sensors	Humenberger et al. (2012)	3D point cloud	Neural network	More than 96 %, FP < 5 %	FPGA and DSP	Yes	NS	NS	Yes
	Fu et al. (2008)	Motion centroid	Threshold based	NS	PC	Yes	No	No	No
	Belbachir et al. (2012)	4D data	Neural network	> 90 % detection rate	FPGA and DSP	Yes	NS	Yes	NS
Monocular Vision sensors	Toreyin et al. (2006)	Wavelet transform	HMM-based classifier	100 % (if audio input also included)	PC	Yes	No	No	No
	Nait-Charif and McKenna (2004)	Ellipse model	Particle filter and Bayesian classifier	100 %	NS	NS	Yes	Yes	Yes
	Nasution and Emmanuel (2007)	Projection histograms	k-Nearest neighbor algorithm and evidence accumulation technique	90 %	PC	12–16 fps	No	No	No
	Mirmahboub et al. (2013)	Motion speed based on silhouette area	SVM	Accuracy 94.01 %, Error 4.8 %	PC (MATLAB)	Real-time	Occlusion handling	Yes	Real-world set-up
	Yu et al. (2012)	Ellipse and projection histograms	Directed acyclic graph support vector machine (DAGSVM)	97.08 %	PC	NS	No	Yes	No
	Makantasis et al. (2012)	Projected Height-width ratio of bounding box	A feed forward neural network	Over 90 %	NS	20 fps	Yes	No	Yes
	Rougier et al. (2013)	Ellipsoid projection	Particle filter	Error 5 %	PC	Quasi-real time (130 ms/frame)	Yes	Yes	Yes

AUC Area under curve, *CM* classifier/Model, *NS* not specified, *fps* frames per second, *FP* false positive rate datasets for both training and evaluation is also emphasised. Currently, limited datasets are publicly available, such as Auvinet et al. (2010), Charfi et al. (2013), Kwolek and Kepski (2014), Charfi et al. (2012).

With a practical view that remote monitoring systems will need to meet large scale demands in the near future, the computational cost, robustness in realistic scenarios and real-time performance must be regarded (Mubashir et al. 2013). Recent studies (Hawley-Hague et al. 2014) show that the technologies developed for fall detection and monitoring need to be simple, reliable, effective and tailored to the individual needs of the aging individuals. A summary of the sensor systems in fall detection is presented in Table 2.

4 Sleep monitoring

Sleep is an important indicator of patient health and state of wellness. The quality of sleep is influenced by sleep-related disorders such as apnea (Somers et al. 2008) or restlessness due to other health conditions, more prevalent with patients with cardiovascular disease (Somers et al. 2008; Calhoun 2010). Short sleep durations could also be an independent indicator of either stroke events in elderly hypertensive patients (Eguchi et al. 2010) or complications due to certain types of medications (Neubauer 1999).

Sleep monitoring involves recording any of (or a combination of) the following: wake/sleep status, duration of sleep, body movement patterns, heart-rate variations, rate of breathing, sleep position etc. Intrusive and wearable sensor based technologies like polysomnography (PSG) (Malakuti and Albu 2010) or actigraphy are used to study sleep-related disorders. These involve contact-based sensors, which are usually limited for use in hospitals and sleep clinics.

Among the non-intrusive sensor based sleep monitoring systems, vision sensors and pressure sensors have been used.

4.1 Technologies

Existing non-intrusive sleep monitoring techniques usually involve infrared sensors (Yu et al. 2013; Chen et al. 2014), monocular vision sensors (Nakajim et al. 2001) and pressure-based sensors (Malakuti and Albu 2010).

Infrared based are commonly used for sleep monitoring, considering that sleep monitoring includes night time monitoring. In Liao and Yang (2008), motion information is used to infer the wake/sleep status using near infrared (NIR) cameras. Multi-modal sensing is employed in Chen et al. (2014) and Peng et al. (2006) to monitor sleep patterns. For example, in Peng et al. (2006), audio and heart rate (HR) sensors along with infrared night-vision camera are used to infer sleep/awake status. Motion amplitude histogram and SVM (support vector machines) classifier are used in determining the sleep/awake status. In Yu et al. (2013), the breathing rate, body movement, and sleep position of the user are monitored using the Kinect's infrared camera. The depth images are processed and connected component analysis is used to detect the head and torso of the patient and the changes in average depth values are used for breath measurement. In Yang et al. (2003), the body orientation features are identified from video frames that are captured by a night vision infrared camera. A follower neural network is then used to determine the body direction. Kayyali et al. (2008) and Espa et al. (2000) also use an infrared camera, but for the purpose of real-time patient visualization for all-night monitoring. The video is not processed on chip, and is instead transferred to a remote laboratory for monitoring.

In pressure-sensor based systems, variations in periodicity of the pressure signals induced by sleep irregularities are acquired from a bed of pressure sensors. These variations are represented as image patterns in the inter-frame similarity matrices which are obtained from the pressure maps where the periodic changes due to breathing are induced Malakuti and Albu (2010). The patterns represent breathing, motion and apnea.

In video camera based methods such as Nakajim et al. (2001), frequency of changes in posture and respiration rate are inferred using an optical flow method to detect movements of blanket or chest. However, the method faces the limitation of the sensing technology itself, while trying to detect subtle movements of the chest.

Table 3 summarizes the surveyed vision based techniques for sleep monitoring and the associated challenges.

4.2 Discussion

Although most existing sleep monitoring systems explore different techniques to develop improve robustness and effectiveness, there are open challenges which need to be addressed. We list them briefly here:

– *Noise* This is particularly important in pressure sensor based systems such as Malakuti and Albu (2010) where noisy pressure readings can introduce errors in pressure maps. The challenge arises from low SNRs (signal to noise ratios) of the pressure sensor that in-turn affects the interframe similarity matrices derived from pressure maps. Similarly, noisy motion vectors can play spoilsport in methods like (Liao and Yang 2008), which are motion based methods to detect sleep patterns. However, homomorphic filtering is used to filter such noisy readings in this method.

Table 3 Overview of the surveyed sensor-systems used in vision-based sleep monitoring

	Technology used	Presnece of blanket	Parameters monitored	Features used	Sources of error
Yu et al. (2013)	IR camera	Yes (thin)	Breathing rate, sleep movement and position	3D points, average depth values	Occlusion by hand, movement of user while measuring breath
Liao and Yang (2008)	IR camera	Yes	Wake/sleep states	Motion History Image (MHI)	Not discussed
Yang et al. (2003)	IR camera	Yes (limited study)	Sleep motion (body orientation and direction)	Edges, Image projections	Presence of quilt and variation in hairstyle, clothing
Peng et al. (2006)	Heart rate monitor, audio and IR night vision camera	Not discussed	Body movement, heart rate	Motion vector amplitude histogram	Imperfect contact with sensor belt
Malakuti and Albu (2010)	Pressure sensors	Not discussed	Distinguish between normal breathing, sleep apnea nd body motion	Interframe similarity matrices from pressure maps	Low spatial and temporal resolutions and SNR of bed sensor
Nakajim et al. (2001)	CCD camera	Yes	Frequency of changes in sleeping posture, changes in respiration rate	Optical flow vectors	Non-uniform lighting

- *Presence of blanket* Although it may seem trivial, presence of a quilt or blanket has been found to be challenge in sleep monitoring. Some of the IR-camera based methods have considered presence of a quilt or a blanket (Yu et al. 2013; Yang et al. 2003; Nakajim et al. 2001; Liao and Yang 2008). A thin quilt is considered in (Yu et al. 2013) and the results of the experiments performed with and without quilt are compared. The accuracy of detection is limited by the thickness of the quilt. Presence of a quilt is a bigger challenge in monocular vision based works, which is currently less addressed.
- *Low-cost systems* In the view of developing low-cost systems for sleep monitoring, low cost and low resolution NIR cameras have been used. This introduces a trade-off between the cost and the robustness of the system. For example in NIR camera based systems (Yang et al. 2003; Liao and Yang 2008; Yu et al. 2013; Peng et al. 2006), the intensity of the image is found to be a function of distance between camera and patient. Therefore, developing robust and accurate systems with adequate amount of visual information using low cost sensing technologies remains a challenge that needs to be effectively addressed.

5 Breathing monitoring

Breathing or respiration monitoring includes monitoring breathing rate, breathing depth, breathing stability, inhalation time, exhalation time, inhalation/exhalation ratio, and symptoms of sleep apnea (Yu et al. 2013). Infrared vision sensors are mainly used in breath or respiration monitoring, and two types of approaches to breath-monitoring using infrared cameras have been mainly explored.

5.1 Approaches for respiration monitoring

AL-Khalidi et al. (2011) provides a review of the respiration monitoring methods, including both contact-based and non-contact-based methods. In this section, the main approaches to vision-based respiration monitoring are discussed.

5.1.1 Thermal imaging of nostril region

Murthy et al. (2004) propose a method based on the observation that air breathed out has a higher temperature than the background of indoor environments. Hence, the expired air is observed to be at a distinct thermal signature in the infrared imagery, and a statistical algorithm is used to compute this signature. In Zhu et al. (2005), the region below the nose is tracked using mean shift localization

(MSL)-based particle filtering, thereby tracking the breathing signal and capable of handling occlusion and significant head movement. The profile view of the subject is considered in this method from a distance of 6-8 feet. In AL-Khalidi et al. (2010) and Alkali et al. (2013), facial features from the thermal images are extracted and eventually the respiration rate is computed. A method for the quantification of breath flow utilizing thermal images is proposed Johnson et al. (2007). In Fei and Pavlidis (2006), the region under the nostril is selected as the region of interest (ROI) and the mean dynamic thermal signal of the breath is computed. The breathing patterns are characterized based on observations across breathing cycles and abnormalities are detected in the breathing patterns.

5.1.2 Tracking movement of chest and abdomen

The second approach for monitoring breathing is by detecting the periodic rise and fall of the chest and abdomen (Kuo et al. 2010; Martinez and Stiefelhagen 2012; Kroutil et al. 2011; Sato and Nakajima 2005). In Kuo et al. (2010), the respiratory region (RR) of the subject who may change position and posture while sleeping, is dynamically located, after which the respiration of the subject is analyzed until the subject again moves or changes position. In Nakajima et al. (2000), Frigola et al. (2002), the breath rate is computed by measuring the movement of chest by analyzing the optical flow vectors. Wiesner and Yaniv (2007) proposes a system to track the motion of color fiducials placed on the subject's abdomen using a single low-cost camera. Aoki et al. (2005) and Kroutil et al. (2011) use a fiber grating vision sensor and the shift in the projected light pattern over the thorasic abdominal part of the subject during respiration is detected. Sato and Nakajima (2005) detects change in volume using fiber-grating 3D vision sensors and infers the movement of the outside body.

In Martinez and Stiefelhagen (2012), an infrared dot pattern is used to detect breathing patterns. The dot pattern is captured using a camera and a matching IR-filter. The dots are tracked over time and their trajectories are filtered using a PCA based method; the respiratory rate is estimated using autoregressive spectral analysis.

In Bai et al. (2010), double webcams are used to capture the images and transmit them to an embedded board, where the chest movement is detected through temporal differencing of the images, and the breath rate is calculated. Additionally, an infrared illuminator is used to automatically irradiate the human body when the light is insufficient.

5.2 Sleep apnea detection

Sleep apnea is characterized by repetitive interruption or cessation of ventilation during sleep (Somers et al. 2008). Sleep apnea detection involves approaches similar to breathing monitoring, viz. tracking movement of chest and abdomen, and thermal imaging of nostril region. Below is a review of some of the recent works that have specifically addressed detection of sleep apnea.

In methods based on tracking movement of chest region such as Takemura et al. (2005), fiber-grating sensor based method is used to detect and also distinguish between the events of obstructive and central sleep apnea. In Wang et al. (2006), an infrared video camera is used and a 2D breath-activity template that is continuously updated is used. The method caters to varying postures of the subject and if partially covered by a sheet or blanket. Wang et al. (2014) is an extension of Wang et al. (2006), where improved models for motion detection and activity recognition have been introduced. The method runs in real-time and is robust to occlusions by a blanket or bed cover, variations in appearance of the subject and substantial changes in camera view relative to the subject. Rodrigues et al. (2007) describes a system that uses audio and image data along with oximetry data to detect events of sleep apnea. Visually capturable signs such as head movements, mouth kept open, etc. that precede the apnea event are intended to be recorded by the camera. However, there is insufficient information regarding the experimentation and evaluation of this method. A TOF camera is used to record the movements of abdomen and thorax of the subject in Falie et al. (2008, 2009). The reflected signals from the abdomen and thorax are normally in phase, and when they are out of phase, it indicates the occurrence of a sleep apnea event. However, the method is constrained that it requires the patient to be sleeping facing the camera. In order to tackle the challenges by approaches that detect the changes in the chest and abdominal movement, techniques to estimate the posture of the subject in varying poses and handle the occlusion by the blanket covering the subject have been proposed Wang et al. (2010), Wang et al. (2007).

In methods based on thermal imaging of the nostril region (Fei et al. 2009), tracking of nostril region is followed by wavelet analysis to detect the occurrence of an apnea event.

5.3 Discussion

The thermal imaging based method faces the challenge of patient covering his nose while sleeping or when he sleeps laterally. Murthy et al. (2004) deals only with profile views of the subject, whereas Zhu et al. (2005), AL-Khalidi et al. (2010) can detect in front-facing views of the subject as well. Also, Zhu et al. (2005) considers occlusions to the region of interest and significant head movement. However, it is a challenge for these methods to work if the patient's nose is covered by a blanket.

In Johnson et al. (2007), the subject is required to place his head inside a container that has the system set-up to capture the image data. This method is not suitable for a continuous monitoring system in a hospital or a home environment.

Methods based on tracking movement of chest and abdomen require the subject to be lying down on a bed (Kroutil et al. 2011). Some methods incorporate an ROI detection step (Kuo et al. 2010) whereby they can cater to varying postures of the subject. Bai et al. (2012), Kuo et al. (2010), Nakajima et al. (2000) incorporate the ability to monitor breathing irrespective of changes in posture or for a limited set of postures (Sato and Nakajima 2005). Also, most of the methods except methods such as Sato and Nakajima (2005), Wiesner and Yaniv (2007) consider the subject to be covered by blanket.

In most of the methods, the subject is considered to be within a distance of 6–8 feet from the camera system (Murthy et al. 2004; AL-Khalidi et al. 2010; Alkali et al. 2013; Sato and Nakajima 2005; Fei and Pavlidis 2006). Considering that the system needs to detect subtle movements and thermal changes, the detection system needs to be placed within a certain maximum distance from the subject. Change in sensitivity of an IR-camera with respect to the distance from the subject is shown in Bai et al. (2012). Certain recent methods (Martinez and Stiefelhagen 2012) are designed to extract maximum information, and hence can estimate very small movements of the order of 2 mm from a distance of 3 m.

6 Epilepsy monitoring

Epilepsy is a medical condition arising out of an abnormal neural activity in the brain, and accompanied by repeated seizures or convulsions over time (Iasemidis 2003). Convulsions or seizures are characterized by uncoordinated movement of the body or certain body parts (Lu et al. 2013). Classification of epileptic seizures can be found in Engel and ILAE (2001), Noachtar and Peters (2009). Visual indications of epileptic seizure onset include changes in posture, eye, mouth and head movements (Pediaditis et al. 2012). Though video-EEG methods are being used for clinical practice, they do not involve automatic analysis of the video for detecting the seizure related movements. Automatic detection of epileptic motion from video have drawn interest over the past decade.

The epileptic motion detection methods are classified as: marker-based and marker-free methods, and as: methods dealing with motion detection, motion analysis and motion recognition. A systematic review of vision-based detection of human motion in epileptic seizures is provided by Pediaditis et al. (2010). Here, we provide a review of the approaches and challenges in vision-based epilepsy monitoring.

6.1 Technologies

Techniques have been proposed to track and measure the movement of the body for monitoring epileptic seizures. One of the commonly used approaches is to employ optical flow for detection of specific bodily movements. In Kalitzin et al. (2012), optical flow, extraction of global group transformation velocities, and band-pass temporal filtering is used to detect clonic movements and high detection rate is achieved. In Cuppens et al. (2012), spatio-temporal features are used to detect motion in myoclonic epileptic seizures. This is a marker-free approach and also does not require the patient not to be covered by a blanket. In this method, epileptic motion is studied using HoF (histograms of optical flow) features. Lu et al. (2013) utilizes a single camera mounted on the ceiling and uses color-based video analysis in order to quantify limb movements in epileptic seizures. The system can estimate quantitative measurements such as the jerking frequency and duration, which neither EEG nor visual inspection can provide. In Pediaditis et al. (2012), absence seizures are detected using a video camera. Time-varying signals are extracted at the eye and mouth regions using optical flow computation and averaging background subtraction techniques and a classification algorithm is applied to classify the signal as indicative of a seizure or otherwise. Sathyanarayana et al. (2015a) propose a technique to detect absence seizure by detecting the occurrence of blank stares. The technique is based on weighted accumulation of intensities and gradient magnitudes in the eye region, along with a template based approach.

Marker-based approaches (Li et al. 2002) involve infrared reflective markers attached to landmark points on the body. An infrared sensitive CCD camera is used to detect the markers, which are seen as high-intensity blobs in the video.

A stereo camera system is used in Cuppens et al. (2009), Cuppens et al. (2010), and optical flow algorithm is applied in order to detect myoclonic jerks in epileptic patients during sleep. In Cuppens et al. (2010), optimal thresholds to distinguish movement from non-movement epochs, which were tested using a three-fold cross-validation are determined. In Liu et al. (2004), an illumination invariant change detection method is proposed and applied to detect epileptic seizure movements.

6.2 Discussion

Some of the challenges encountered in vision-based epilepsy monitoring are occlusion by body parts of the subject, occlusion by objects such as blankets or other persons, low lighting conditions and night conditions, variations in posture of the subject, the various types of epileptic movements which need to be addressed, and unavailability of sufficient clinical test data.

Handling low lighting conditions Lu et al. (2013) fails to work if lighting conditions are not sufficient or if the colors are not clearly distinguishable and hence the adaptability of the technique for a subject covered with a blanket is yet to be explored. Lu et al. (2013) works for detecting seizures with large limb movements, and not for seizures with only subtle movements. Exploring the use of infrared based methods can make the method more sensitive to detect the movements under low-lighting conditions.

Handling occlusions The methods proposed so far are mostly 2D vision-based techniques, while 3D vision-based methods are less explored (Pediaditis et al. 2010). Depth based methods have the potential to tackle many of the occlusion related challenges.

Addressing the various types of seiures Given that there are various types of epileptic seizures that can occur and only a limited set of seizure types have been addressed (Pediaditis et al. 2010), exploring the detection of the various other types of seizures is a potential and important area of research.

7 Facial expression monitoring

Face and facial expressions are an important source of information while monitoring the wellness of a patient. Examples of such applications include pain monitoring, vital signs monitoring, monitoring the basic facial expressions of a patient, depression, drowsiness and stress monitoring. It is observed that pain detection has been relatively the most researched area among the above mentioned applications

7.1 Pain detection and monitoring

Pain detection is important in post-operative care (Gilbert et al. 1999), regulating medications, long-term monitoring and guaging the effectiveness of a treatment and hence it is considered critical to detect pain in patients (Sikka et al. 2014). Pain detection from facial images is done by capturing the pain-related facial actions (Prkachin 1992). Ekman and Friesen (1978) provides further information on facial actions and the facial action encoding system (FACS). A common facial response to pain is observed across age groups of people (Prkachin 2009). Facial actions such as orbit tightening, cheeks raising, eyebrow lowering are indicative of pain (Prkachin 2009). The main challenge in pain detection methods is the frame-by-frame labeling of the FACS actions in terms of action units in the ground truth data, which is very time-consuming. Methods have been proposed to address this challenge. Other challenges include the variation in the way individuals express pain, and hence the adaptability of the pain detection methods to individuals. The approaches for pain detection can be categorized under two classes viz., appearance based methods and appearance and geometry based methods.

7.1.1 Appearance-based methods

One of the first attempts to automatically detect pain was carried out by Littlewort et al. (2007). Action units are recognized using gabor filters, adaboost and SVM and were used to differentiate between genuine pain and fake pain. Methods for neonatal pain detection have been proposed (Gholami et al. 2009; Nanni et al. 2010). Images of the COPE database are classified as pain or non-pain. Gholami et al. (2009) uses the column stacking of intensity values as feature extractors and RVM (relevance vector machine) as a classification technique, whereas, elongated ternacy pattern (ELTP) are used as feature descriptors and SVM is used in Nanni et al. (2010). Sikka et al. (2014) propose a multiple-segment multiple instance learning framework for detecting spontaneous expressions of pain in videos. Multiple instance learning is used in order to use sequence-level ground truth, or in other words, weakly labeled data, as against most of the previous methods that rely on frame-level labeling. This method also includes the temporal dynamics in the proposed method of detecting pain.

7.1.2 Appearance and geometry based methods

Geometry based features such as shape and distance along with appearance based features are used to add robustness to changes in head pose. AAM-based (Active Appearance model based) shape and appearance features are extracted in Ashraf et al. (2007). A rigid representation of face appearance and frame-level labels to train classifiers is used to provide a sequence-level detection of pain. In Werner et al. (2012), distance and gradient features are extracted and an SVM classifier to distinguish pain from non-pain. A comparative labeling and learning model is proposed for assessing pain expression intensity. Further, pain detection at frame-level in a video was explored in Ashraf et al. (2009), Lucey et al. (2011a).

AAM-based shape and appearance features are used to make the method more robust to major facial deformation and head movement that is common in pain (Lucey et al. 2011a; Hammal and Cohn 2012). While most of the methods consider head pose as a source of registration error, it is a potential source of information for pain and pain intensity indication (Hammal and Cohn 2012). Levels of pain are classified based on the Prkachin and Solomon Pain Intensity Metric (PSPI) metric. Other methods that use head pose information for classification in addition to facial expression was proposed in Lucey et al. (2012), Werner et al. (2013). In Lucey et al. (2012), 3D parameters derived from the AAM are used along with facial expressions to detect pain and assess the level of pain based on the PSPI scale. In Werner et al. (2013), a method for fully automatic detection of pain based on facial expressions and head pose information is proposed. Depth and color features are extracted at the frame level and a time window descriptor is calculated from them. Werner et al. (2013) also contribute a new database, the BioVid Heat Pain database.

The above methods except (Sikka et al. 2014) are also static classification based methods and do not consider the dynamic change in the deformation of the facial features during the sequence of an expression. Hammal and Kunz (2012) propose a temporal and multi-cues modeling process for automatic and dynamic pain recognition.

All of the above discussed methods except (Sikka et al. 2014) are based on coding every frame of the video based on FACS. Considering real-time applications, this could be a time-consuming process and hence be a possible bottleneck. Also, most of these methods are based on AAM-based features, which require face alignment as a necessary step. Recently, Khan et al. (2013) proposed a framework which is not based on FACS, nor is face alignment required. Shape information is extracted using pyramid histogram of orientation gradients (PHOG) and appearance information using pyramid local binary pattern (PLBP).

Studies related to pain assessment in patients who could verbally not communicate the pain due to dementia show that facial expressions among other indications such as vocalizations are accurate measures of assessing the presence of pain, but not its intensity (Manfredi et al. 2003; Horgas and Miller 2008). In Kunz et al. (2007), the FACS is applied for assessing pain from the recorded video of the patients with dementia.

An overview of the surveyed papers in pain detection is presented in presented in Table 4.

7.2 Depression monitoring

Facial expressions can be used to detect depression, based on the FACS action units (AUs) (Ellgring 1989). In Alghowinem et al. (2013b), the approach used is that positive emotions are less expressed in depressed subjects and that recognizing depression by detecting positive emotions is more accurate than detecting negative emotions. In McIntyre et al. (2009), the number of occurrences of positive (such as happiness) and negative prototypical expressions (such as anger, fear) and non-prototypical expressions within a certain duration of time are noted, based on which depression is detected. Local shape and texture features from the facial image are measured using AAM, which are then used to classify regions using a multiboost classifier. In Cohn et al. (2009), the facial behavior of the subject in response to certain questions were noted. The facial actions through both manual coding and by using AAM were measured and compared.

Head pose and movement provide effective cues while recognizing depression as a binary classification task (depressed versus non-depressed) (Alghowinem et al. 2013b). The head pose and movement features are extracted by projecting the 2D points of face AAM into a 3D face model. A hybrid classifier based on Guassian Mixture Models (GMM) for each subject and SVM is used for classification. A recognition rate of 71.2 % was achieved just based on the head pose and movement cues, while a combination of these cues along with the frame-by-frame features results in an average recall rate of 76.8 %.

Alghowinem et al. also propose a technique for detecting depression by analyzing eye movement and blinking (Alghowinem et al. 2013a). Eye movement features extracted using AAM models and hybrid classifier based on GMM and SVM is used which results in 70 % accuracy in detecting depression. Eye movement by themselves can be used as a complementary cue along with other cues such as speech, for recognizing depression.

McIntyre et al. (2009), Alghowinem et al. (2013b), Alghowinem et al. (2013a), Cohn et al. (2009) use person-specific AAM models, and hence a new AAM model needs to be trained for every new subject, and this can be time-consuming and complex. Joshi et al. propose a framework that is subject-independent (Joshi et al. 2012, 2013b). In Joshi et al. (2012), facial dynamics are analyzed using features based on Local binary pattern (LBP) and upper body movements are analyzed using features based on space-time interest points (STIP). A maximum accuracy of 88.6 % was obtained.

An extension of their work in Joshi et al. (2013b), upper body expressions and gestures are used for automatic depression analysis, and the effect of upper body parts versus face versus head movements are compared. STIP and BoW based framework is proposed; bag of body expressions and bag of facial dynamics are created. A nonlinear SVM was used for classification and 71 % accuracy was seen in both the cases of using only head movements for analysis and using facial dynamics alone.

Intra-facial muscle movements and the head and shoulder movements are combined with audio features and then fused to result in a multi-modal framework (Joshi et al. 2013a). The authors of this work conclude that only a multi-modal system will be suitable for achieving the robustness needed for real-world applications.

Table 4 Overview of the surveyed pain detection methods

	Level of classification	Type of features	Features	Classifier	Database	Classification accuracy
Ashraf et al. (2007)	Sequence-level detection of pain/no-pain	Shape and appearance features	AAM-based shape and appearance features are extracted; Rigid representation of face appearance and frame-level labels to train classifiers	SVM	UNBC-McMaster	Hit-rate—81.21 %
Ashraf et al. (2009)	Frame-level labeling	Shape and appearance features	AAM-based shape and appearance features	SVM	UNBC-McMaster	Hit-rate—82 %
Littlewort et al. (2007)	Sequence level labeling	Appearance based	A subset of Gabor filters selected using Adaboost	SVM	FACS-101, MMI Facial Expression Database, Cohn-Kanade DFAT-504 dataset, Ekman and Hager	72 %
Lucey et al. (2009)	–	Shape and appearance features	AAM-based shape and appearance features are extracted with respect to specific AU targets	SVM	UNBC-McMaster	AUC = 78.37
Gholami et al. (2009)	–	Appearance based	Column stacking of intensity values	RVM	Infant Classification of Pain Expressions (COPE) database	Accuracy—82–88 % (pain or no pain), 85 %— assessing level of pain
Nanni et al. (2010)	–	Appearance based	ETLP (Elongated Ternary Pattern) features	SVM	Infant Classification of Pain Expressions (COPE) database	AUC = 93
Lucey et al. (2011a)	Frame-level classification	Shape and appearance based features	AAM-based features	SVM	UNBC-McMaster	AUC=84.7
Werner et al. (2013)	Sequence level	Distance and appearance features	Distances, gradient features and a linear intensity model	SVM	Hi4D-ADSIP database (3D posed pain sequences)	TPR = 92.9 %
Hammal and Cohn (2012)	–	Shape and appearance features	AAM, log-normal filters	SVM	UNBC-McMaster	Average Classification rate = 96 %
Lucey et al. (2012)	Detect different levels of pain intensity at frame-level	Shape and appearance features	AAM, 3D head pose features	SVM	UNBC-McMaster	AUC = 84
Hammal and Kunz (2012)	Sequence level	Shape features	Shape deformations of facial features	TBM (Transferable Belief Model) based model	STOIC and spontaneous database	Recall of 85 %

Table 4 continued

	Level of classification	Type of features	Features	Classifier	Database	Classification accuracy
Khan et al. (2013)	Sequence level	Shape and appearance features	PHOG and PLBP	SVM, random forest, 2Nearest neighbor, decision tree	UNBC-McMaster	Pain recognition rate—between 90.2 to 96.9 % (for the different classifiers used)
Werner et al. (2013)	Frame level + time window level	Distance and appearance features (features extracted from head pose and expression)	Color, depth, distance, gradient magnitude features	SVM	BioVid Heat Pain database	7.6 % error
Sikka et al. (2014)	Frame-level classification	Appearance based	Spatial pyramid BoW (Bag of Words) framework	Multiple-segment multiple instance learning	UNBC-McMaster (Lucey et al. 2011b)	83.7 %

While the methods proposed have been tested on real-world clinical data, based on subjects with depression, a major challenge faced is that the size of the dataset on which the methods are tested are limited, based on a few subjects. This indicates the need to create larger datasets (Alghowinem et al. 2013b).

7.3 Facial expression detection for medical condition detection

Methods that evaluate facial movements for health monitoring is also being explored, a review of which is provided in Mishima and Sugahara (2009). Mishima and Sugahara (2009) discusses the importance of evaluating facial movements in the diagnosis and treatment of conditions such as facial paralysis, cleft lip and palate, trauma etc. It is observed that most of the methods are marker-based methods, which is time-consuming and may also inhibit the natural or spontaneous facial motion.

Dai et al. propose a technique to monitor patients on bed based on analyzing expressions such as happiness, easiness, uneasiness, disgust, suffering, and surprise (Dai et al. 2001). Optical flow and associate model are used for classification of the facial expressions. Health condition detection based on facial changes other than pain and depression detection that are being explored are sleep, drowsiness, fatigue detection and monitoring, stress recognition, and monitoring vital parameters (vital signs monitoring is discussed separately Sect. 8).

Initial work on monitoring of coma patients or patients under post-surgical care in an intensive care unit (ICU), based on facial changes such as eyes and mouth movements was explored in Bin Mansor et al. (2010) and Naufal Mansor et al. (2010). Features of interest (distance between eyelids and lips, areas of eyes and mouth) are extracted and passed through a fuzzy classifier and eventually the 'awake' state of the patient is detected .

Liao et al. proposed a multi-modal stress recognition system (Liao et al. 2005), that includes information from facial expressions, eye movements and head movements that are extracted using vision-based sensors. A dynamic Bayesian network (DBN) framework is used to model the stress. Eye detection and tracking, gaze estimation, facial feature tracking and face pose estimation are the main components of the visual tracking system.

Facial expression recognition has been used to study deficits in emotional expression and social cognition (Mandal et al. 1998) and automated vision-based expression detection has been explored in the context of neuropsychiatric disorders, primarily schizophrenia and dementia (Bevilacqua et al. 2011; Hamm et al. 2011; Wang et al. 2008; Alvino et al. 2007).

Automated computational framework for analyzing facial expressions from video data of Asperger's and schizophrenia patients was proposed by Wang et al. (2008) and it was shown that healthy subjects show the intended emotion better than the subjects with Aspergers and schizophrenia. Facial expressions images of patients with schizophrenia and healthy subjects are taken and the extent to which the expressed emotion matches the intended emotion (Alvino et al. 2007). Regional volumetric difference (RVD) functions of the expression changes are used to train classifiers and pattern classification techniques are applied. Neurological distress expresses as loss of expressiveness in the face, and facial expression recognition based on the FACS is used to diagnose neurological disorders such as Alzheimer's disease or depression (Bevilacqua et al. 2011). The flatness and inappropriateness in the expressions of subjects with schizophrenia was studied in Hamm et al. (2011).

7.4 Monitoring the after-effect of drug

Monitoring the after-effects of drug is vital in patient monitoring and can help prevent the occurrence medical emergencies due to unexpected reactions of a patient to certain drugs. Vision-based monitoring of after-effects of drug based on facial changes of a patient have gained interest lately. The effectiveness of certain drugs or clinical procedures are monitored by monitoring the temporary evolution of facial edemas. A vision-based method to quantify the edema was proposed by Brusco and Paviotti (2005). 3D model of the patient's face is constructed and the difference in facial surfaces indicate any facial volume changes due to edema. A challenge faced by this method is the alignment of the 3D models of the same person taken at different times. After giving Botox injections to patients with facial nerve disorders, the subtle changes in their facial motion was detected using automated facial image analysis (Rogers et al. 2007). The types of abnormal facial activity following facial paralysis have been listed in Rogers et al. (2007), and feature points of interest were marked in the first frame and then tracked based on the optical flow algorithm.

7.5 Discussion

Facial expression monitoring is still an active research area because of variety of challenges that are yet to be fully resolved. One of the main challenges is related to face detection itself, which is most often the first step in facial expression monitoring. Therefore, the challenges involved in face detection will also affect in detecting the expressions. Detecting varying scales of faces, robustness to orientation changes and occlusions are key challenges in detecting faces. Additionally, in the context of patient monitoring face detection is further affected by the presence medical accessories on faces such as tubes etc. Detection of facial features using the action units (AUs) is affected by such artefacts.

Another challenge is with respect to datasets in the context of patient monitoring. There are a number of datasets for generic facial expressions monitoring such as AM-FE datasets (McDuff et al. 2013) and Cohn-Kanade dataset (Kanade et al. 2000). However, they may not fully represent the facial expressions that one would see in the case of patients. Additionally, the above mentioned datasets are meant for evaluating seven basic emotions (Kanade et al. 2000). They do not cater to specific medical conditions that a computer vision system for patient monitoring demands. Also, the perspectives in such datasets are usually different from what one would see while monitoring a patient using an overhead camera or front camera.

The major challenge in creating datasets is the issue of medical privacy. Given that it involves faces of patients under medical examination, it becomes even more difficult to create such datasets. However, considering that certain medical conditions can be simulated or enacted by normal subjects, creation of datasets in such a manner will aid in developing and evaluating techniques for proof-of-concept. For example, initial work on dataset creation for patient wellness monitoring based on facial analysis is presented in Sathyanarayana et al. (2015b), and involves simulation of the facial expressions indicating certain "unwell" conditions by healthy subjects. Also, most of the studies are usually conducted on proprietary datasets, which cannot be used for comparative analysis, and hence publicly available datasets are needed.

Another critical challenge is related to the trade-off between computational efficiency and robustness of face detection. Although Viola-Jones Haar-Adaboost classifiers (Viola and Jones 2001) give real-time performance in detecting faces, increasing the complexity by including multiple perspectives and scales decreases the performance of these classifiers in terms of both performance and robustness. Increasing the algorithmic complexity to improve robustness towards achieving better face detection results also decreases the performance in terms of frames per second. Techniques such as search space reduction (Sathyanarayana et al. 2014) are being employed to generate hypothesis windows for face detection, thereby improving the computational performance of face detection.

Another interesting challenge is the detection of true facial expressions. In the context of patient monitoring, detecting fake pain as against true pain is a commonly studied topic (Littlewort et al. 2007). Statistical modeling is done to model fake pain and differentiating it from true pain. However, the accuracy of such models is still an active area of research. Also, it is important to note that it is still a challenge to estimate the pain if the person's pain may not reflect evidently as changes in their facial expressions or if there is no facial reaction to the pain (Werner et al. 2013).

Apart from pain detection and monitoring, other applications such as fatigue and drowsiness monitoring are areas to be explored further in the context of patient monitoring. Bringing in the context-awareness with respect to patient will aid in designing cost-effective techniques for these applications. Monitoring the after-effect of drug has currently been explored with respect to very few drugs and is an important direction of research. Monitoring a patient under anaesthesia is a critical area of patient monitoring and detecting pathological events such as hypovolaemia, sympathetic responses which are indicative of inadequacy of pain relief, etc. based on visual cues is a promising area of research (Gohil et al. 2007).

8 Vital signs monitoring

Vision-based human vital signs monitoring has been an upcoming field of research that has gained increasing interest in the recent past. Vital signs include heart rate, pulse rate, blood pressure, body temperature and respiration rate. While we have elaborately discussed respiration related techniques, we will now survey techniques for detecting the other vital signs using vision based methods. The methods proposed in literature are mainly either based on thermal imaging or color-image processing.

8.1 Thermal imaging based methods

One of the first attempts to extract human cardiac pulse from thermal video sequences was made by Sun et al. (2005). The method is based on the assumption that temperature on the blood vessel is modulated by the pulsative blood flow within it and this produces the strongest thermal signal on a superficial blood vessel. The method exploits the quasi-periodic properties of the cardiac pulse through Fourier signal analysis. Similarly, a thermal imaging based method of the face and neck areas to detect breathing and heart rate was proposed in Chekmenev et al. (2005). In Chekmenev et al. (2005), wavelet analysis of the long-wave IR is performed, whereas the method proposed in Sun et al. (2005) uses mid-wave IR. Gault et al. also use long-wave IR for calculating heart rate from thermal video frames with an average accuracy of around 93 % (Gault et al. 2010). Vascular maps are used to identify the vessel of interest, which is tracked through time and the arterial pulse waveform is recovered. An improvement of this work towards a more accurate, sensitive and automatic method was proposed (Gault and Farag 2013). A review of thermal imaging based and RGB imaging based methods for heart rate measurement is provided in Kranjec et al. (2014).

8.2 Color image processing based methods

Color-based methods to detect vital parameters are based on the idea that the color of the skin/face changes when there are movements of blood in particular areas. In Lewandowska et al. (2011), principal component analysis (PCA) is applied to the color channels and pulse measurements are made considering a very small part of the facial image. The color channels carry information about the color changes corresponding to the blood volume pulse. In Kwon et al. (2012), a mobile phone placed at 30 cm from the face of a person was used to extract heart rate. A cardiac pulse reflects as a subtle color change of the skin. The raw trace signal from the green channel of the image is taken and independent component analysis (ICA) is applied on it to extract a more accurate cardiac signal. Frequency analysis is applied on both these signals to compute the heart rate. Low error rates were observed. Poh et al. also propose a method for cardiac pulse measurement based on color information (Poh et al. 2010). The method is based on blind source separation of the color channels into independent components using ICA. They propose a low-cost, accurate, video-based, automated and motion-intolerant method designed for detecting the heart rate of more than one individual at a time, unlike the other methods which are limited by movement of the individual and are catered to one individual at a time. However, the motion considered is controlled slow head swings. Li et al. consider spontaneous movements and illumination changes as in the case of realistic scenarios in their technique for heart rate measurement from video (Li et al. 2014). Face tracking is used to tackle the challenge of rigid head movements and the green channel of the background is used as a reference to reduce interference due to illumination variations. The method is tested on a public dataset MANHOB-HCI, unlike the other methods that are tested on proprietary datasets that are not made publicly available. The authors also compare the performance of their method with Poh et al. (2010), Kwon et al. (2012) and (Balakrishnan et al. 2013) and show the improvement in performance of their method over the previous methods when realistic dataset was considered.

8.3 Motion based methods

Balakrishnan et al. (2013) propose a technique for detecting heart rate from subtle head movements that are caused by cardiovascular pulse. PCA is applied on trajectories of tracked feature points on the head. However, this method requires the subjects to not to perform any voluntary movements.

8.4 Discussion

- Vital signs monitoring methods are shown to achieve detection rates over 90 %. However, certain challenges faced in these methods: artefacts due to body movement and other heat distortions from the surrounding or the body in thermal based vital signs measurement. RGB-based methods are also limited by movement artefacts and also require the patient to be illuminated throughout the duration of recording of the signals. In the case of motion-based methods such as (Balakrishnan et al. 2013), the subjects are required to not perform any voluntary movements. The authors also mention that detecting the heart rate on moving subjects (which is the case of realistic scenarios) is a future direction of study.
- It is observed that heart rate or pulse rate and respiration rate are the two parameters that are mostly considered in the vital signs measurement methods. Hence, exploring vision-based measurement and monitoring of the other vital signals is a potential area of research Poh et al. (2010).
- Creation of publicly available datasets, making the methods robust to noise arising from patient movement, for higher reliability and precision are important in this application area.

9 Activity monitoring

Activity monitoring includes monitoring the posture, action (a sequence of postures), or an interaction (between 2 individuals) (Vishwakarma and Agrawal 2012) depending on the level of complexity of the activity. Activity recognition techniques that have been proposed specific to patient monitoring or elderly care will be briefly reviewed in this section. We will first review the techniques in posture recognition, followed by daily activity monitoring and unusual activity detection.

9.1 Pose monitoring

Current pose recognition systems in literature are mostly limited to classifying poses into standing, sitting and lying down classes, which are eventually intended for fall detection, for instance, (Brulin et al. 2012; Jansen et al. 2007). In Jansen et al. (2007), 3D vision is used. Variations in posture were considered in the test sequences, but the method was tested on a single subject.

In Brulin et al. (2012), after the human is detected, principal axis and volume occupied by the human body are extracted as features, and a fuzzy classifier is used for classification. In Obdrzalek et al. (2012), accurate joint localization for pose estimation in the context of coaching the elderly population to perform exercises using the kinect camera has been evaluated and compared with certain other methods. Robustness to orientation and occlusions and accuracy of joint localization has been evaluated.

9.2 Vision-based daily activity and unusual activity monitoring

9.2.1 Activities of daily living (ADL)

Monitoring the activities of daily living (ADL) such as eating, toileting, etc. is commonly considered for assessing the well-being of subjects, and has been mostly catered to elderly home monitoring.

Most of the vision based techniques proposed for action or activity detection are limited to recognition of the action primitives rather than the actions or activities themselves (Van Kasteren 2011). The CareMedia coding manual has listed out 19 actions that are of interest to doctors, that includes standing, walking, attempting to exit, talking to a staff, object placed on a table, etc. (Stevens and Bharucha 2003). Most of the work on activity recognition is confined to recognition of a limited subset of these activities.

Video camera based systems Crispim-Junior et al. present a video-based event recognition system for older people, using only 2D RGB cameras owing to their larger field of view compared to RGB-D cameras (Crispim et al. 2013). A set of daily activities of subjects with Alzheimer's disease and controlled subjects are monitored. In Cheng et al. (2011), classification into four classes such as eating, walking, no action and other action are considered. Spatio-temporal interest points (STIP) features are extracted and a feature filtering technique to eliminate the interest points that belong to other people is proposed. Cheng et al. (2011) have also released the senior home activity dataset (SAR). In Olivieri et al. (2012), dense optic flow is extracted from video sequences based on motion velocity flow instance (MVFI) templates to detect activities such as falls, walking, bending over, lying down etc. A gait assessment tool is proposed in Wang et al. (2009). In this method, two video cameras are used and silhouettes are extracted and the voxel centroid change between video frames is noted and is used to compute the walking speed, step time and length. Activity detection is formulated as a search and classification problem rather than applying temporal segmentation in Chen (2010). MoSIFT features are extracted from the video segment and represented by the bag-of-features representation and classified using the cascade SVM classifier. Recognition of basic activities such as drinking, walking, going upstairs and downstairs using a wearable camera which can aid in assisting those with disabilities or the elderly is proposed in Zhan et al. (2012). Optical flow features are extracted and k-Nearest Neighbour algorithm (k-NN), LogitBoost (on Decision Stumps) and SVM are used to classify the sequence of activities.

Multi-sensor systems Multi-sensor systems with video and contact-based sensors to monitor the activities of elderly patients have been proposed (Zouba et al. 2007; Bieber et al. 2009; Junior et al. 2012; Amoretti et al. 2011). In Zouba et al. (2007), vision based detection and tracking of the persons is done considering the human as the moving object. The scenario is also recognized. This information is then combined with the contact sensor-based information for the overall inference of the event or activity. Junior et al. (2012) propose a system with video cameras and accelerometers, for fall detection and monitoring of self-maintenance activities. In Bieber et al. (2009), the grayscaled images are divided into cubes and the difference in sum of the gray level values in corresponding cubes across frames is extracted as the features to detect motion. Actions such as a person falling or waving his hand are detected. It has been shown in these works that a high detection rate is achieved in the multi-sensor based systems compared to systems based on individual sensors. In Matic et al. (2010), a vision and RFID based technique for monitoring the dressing activity of a patient is proposed. The face of the individual is blurred to protect the privacy of the individual.

Depth sensor based methods In Amoretti et al. (2011), time-of-flight cameras are deployed for gesture/posture recognition, which is then merged with information based on usage of user-controlled devices for a higher level inference about the activity. In Zhou et al. (2009), background is continuously updated and the human silhouettes are extracted. Based on location and moving speed of the silhouette, actions such as standing, walking are detected with a low error rate of 7.3 %. RGB-D cameras are used for detecting falls and 13 activities of daily living (Zhang and Tian 2012). Structure and motion features of the human are extracted and BoW model is used to represent them. The features are trained using SVM based classifiers. Banerjee et al. (2014) proposes a system with web cameras, web cameras using infrared lighting and Microsoft Kinect cameras for recognition of activity states such as sitting, being upright and being on the floor during both day and night times. In this method, fuzzy clustering methods are used. In Jansen et al. (2008), the position of the patient in the room is detected and the activity levels are derived based on the amount of time the patient spends in the different parts of the room, on the chair, bed etc.

9.2.2 *Unusual activity*

An unsupervised technique for detecting unusual activity a patient monitoring scenario among other scenarios has been proposed Zhong et al. (2004). The unusual events are detected based on a large number of observations that make it easy to identify an avent is indeed unusual, and unusual events are detected by comparing two events and their similarity is measured. Simple features such as spatial histograms of objects in the scene detected, based on motion and K-means is applied to detect the important features from among them. In Khan and Sohn (2011), silhouette features are extracted from the video and R-transform and kernel discriminant analysis (KDA) are applied. Hidden markov model (HMM) is used for training and recognition. Abnormal activities such as faint, fall, vomiting, headache etc. were detected by detecting the key postures in each of these activities and a 95.8 % detection rate is achieved.

Summarizing the findings on activity detection, it is found that methods for pose estimation mainly focus on assisting in fall detection. Techniques for activity detection are designed to detect a certain well-defined set of actions or activities. The datasets are collected mostly in an experimental and controlled set up, while some methods have considered a real-world set up.

9.3 Behavior monitoring

Behavior is defined as human motion patterns with high level descriptions of actions and interactions (Chaaraoui et al. 2012). Here, we present a brief review of techniques involved in human behavior analysis for patient monitoring. Certain researchers take the view point that detecting changes from the expected daily routine of a patient, or in other words detecting abnormal behavior is more meaningful than monitoring the daily activities. A method for detecting abnormal behavior of patients based on unusual patterns of inactivity was proposed (Dickinson and Hunter 2008). The method is based on unsupervised learning of a spatial inactivity map represented as a 2D mixture of Gaussians, which is combined with a Hidden Markov Model (HMM) framework to construct models for normal patterns of behavior. In Kosmopoulos et al. (2008), multiple cameras are used for monitoring patients and alerting during abnormal behavior. Optical flow features from foreground objects are extracted and combined with view-specific features to infer the short term action as perceived by each camera, and the decisions from all the cameras are then fused. If the short term action or motion trajectory is detected as ‘abnormal’, it is highlighted. Hierarchical Context Hidden Markov Model (HC-HMM) for behavior monitoring of patients in a nursing center is proposed in

Chung and Liu (2008). Abnormal behavior such as a faint, patient standing on the bed, sleeping for too long are inferred based on spatial, activities and temporal contexts. An omni-directional vision sensor as an important part of an elderly monitoring system has been explored in Seki (2009). The vision sensor automatically learns the behavioral patterns and detects the unusual behavior patterns based on fuzzy inference approach.

Monitoring the behavior of patients diagnosed with or suspected to be affected by certain health disorders has been considered to be of tremendous value, where the monitoring system can be programmed to detect certain behavior characteristic to the health disorder. In Hirayama et al. (2008), the behavior of a Parkinson's disease patient during a sleep attack, where the patient's head would sag on the onset of a sleep attack was recorded on a video camera, which the patient was unaware about. In Joumier et al. (2011), an automatic video based activity recognition system to detect the psycho-behavioral disorders associated with Alzheimer's disease was proposed.

Systems to monitor behavior of dementia patients have been proposed (Foo Siang Fook et al. 2007; Karaman et al. 2010). A technique for monitoring the agitation behavior in patients with dementia is proposed, based on tracking limb movements (Foo Siang Fook et al. 2007). However, this method is highly dependent on the accuracy of skin segmentation and also requires the subject not to be covered by a blanket. The work presented in Karaman et al. (2010) involves wearable cameras to monitor the activities and behavior of dementia patients.

10 Discussion and conclusions

Having surveyed some of the main applications in vision-based patient monitoring systems so far, this section summarizes the overall observations and draws attention to the key challenges, and the extent to which they have been addressed.

Certain applications have gained more interest than the others in the recent past. For instance, the applications such as fall detection and pain monitoring have seen a wide range of vision-based techniques. The vital signs monitoring is gaining increased importance over the last few years, especially with the recent technological developments in mobile computing. Other applications such as sleep and breathing monitoring, epileptic seizure detection, monitoring facial expressions (other than pain) and patient-specific activity and behavior monitoring are in the relatively early phases of research.

In the following paragraphs, we list the key challenges in the context of vision-based PMS, intended to light on future research.

Invariance to view-point This is one of the key challenges in many PMS applications. Considering that patient monitoring involves the detection of various parts of the body such as face, chest and abdomen, arms etc., the techniques must be invariant to the changes in the positions of these parts of the body. This would directly be affected by the ability of the algorithms to handle invariance to view-point. This is particularly important in applications like fall detection and activity monitoring using fixed or mounted cameras, where the pose of the subject can change drastically with respect to the camera during the fall. Sensing the scene using multiple cameras is commonly deployed in fall-detection. However, applications like sleep and breathing monitoring have not fully accounted for variances in patient's postures. The sensing modalities, i.e the IR cameras need to be further explored to incorporate pose invariances in such applications. For on-bed monitoring of patient's facial expressions, using multi-camera systems or depth-based camera systems are potential ways of handling view-point invariance.

Handling occlusions Similar to view-point invariance, handling occlusions is another critical item that needs to be addressed in future related research. Although depth sensors are used in some of the fall detection and activity recognition methods, the limited range of the depth sensors like Kinect can constrain their use. Stereo vision techniques are also explored in some applications but they bring with them additional challenges like calibration and increase computational costs in terms of registration. Also, such sensors may not be ideal for detecting occlusions during night time where there is limited data for registration. In sleep and breathing monitoring, occlusion due to blanket can cause interference in monitoring, especially because very subtle movement of the chest needs to be detected. In epilepsy monitoring techniques, occlusions due to body parts and blanket are yet to be addressed. Pain and expression monitoring methods mostly have a prerequisite that face and facial features need to be visible (although some methods allow head pose variations), and do not usually discuss the practical issues of occlusions that are possible in real-life scenarios.

Trade-off between real-time performance, detection rate and cost of implementation This is a classic trade off that is always a point of discussion, especially when the vision solution is meant for real-time applications. One of the main reasons for low availability of efficient monitoring systems in the open market is because of the costs involved in manufacturing them. Maintaining lower costs, and yet providing high levels of reliability and dependability is usually hard to achieve. Many fall and pain detection techniques have achieved high detection rates and some methods have also considered real-time performance. Low resolution cameras have been employed in fall detection,

Table 5 Summary of challenges in the various PMS applications

Challenges	Fall	Sleep	Breathing	Epilepsy	Pain	Other facial expressions	Vital signs	Activity
Invariance to viewpoint	• • •	•	•	○	○	○	○	• •
Handling occlusions	• •	•	•	•	○	○	○	•
Real-time performance	• •	• •	• •	○	•	○	•	•
Detection rate	• • •	• •	• •	•	• • •	•	• •	• •
Cost of implementation	• •	•	•	•	○	○	• •	•
Night and low lighting conditions	•	• •	• •	•	○	○	•	○
Realistic scenarios-	•	•	•	○	•	○	•	•
Availability of datasets for evaluation	•	○	○	○	• • •	○	•	•
Limitations due to proximity of camera to patient	• • •	•	•	•	•	•	•	• • •
Privacy	•	• • •	• • •	•	○	○	• • •	•

'• • •' - Addressed by most methods/fairly well, '• •' - Addressed by certain methods/addressed to a certain extent, '•' - Addressed by very few methods/to a less extent, '○' - Yet to be addressed

sleep and breathing monitoring, vital signs monitoring in order to reduce overall cost of the system. However, the accuracy and robustness could be affect by reducing the resolution below acceptable limits. With respect to sleep and breathing monitoring techniques, many of them rely on subtle motion detection, and achieving good detection rates is a challenge. In the context of patient facial expression monitoring, challenges of cost of implementation are relatively less discussed in the current literature.

Night and low lighting conditions Given that patient monitoring is equally important during night time, use of vision sensors for night and low lighting conditions is always questionable. Although IR cameras do provide a possible sensing solution, processing IR images to get acceptable levels of robustness is a challenge in itself. In applications such as sleep and breathing monitoring, techniques are mostly based on infrared cameras which are inherently designed to handle low lighting/night conditions. Certain fall detection methods use 3D range cameras to enable detection even during night time. There is a huge potential in using infrared and RGB-D based sensing in the other application areas such as epileptic seizure detection and facial expression monitoring, although some initial work has been carried out in infrared camera based expression recognition (Jiang and Kang 2007). There is potential in exploring infrared based technology due to the reducing hardware costs, although infrared cameras are more expensive than visual cameras (Gade and Moeslund 2013).

Handling realistic scenarios Although many fall detection methods have been developed on simulated real-life scenarios, they are limited in scope currently. With regards to activity recognition techniques, most of the techniques address a very small subset of generic poses and activities, and hence they need to be developed specifically for patient monitoring. Similarly, pain detection is evaluated on datasets in which the subjects are in seated posture facing the camera with some variations in head pose. But, if case of a realistic scenario of monitoring a patient, e.g. on bed monitoring, then the perspective to the camera, etc. need to be considered accordingly. Certain techniques for sleep, breathing and epileptic seizure detection require the patient not to be covered by a blanket, which may not be met many times, when realistic scenarios are considered.

Availability of datasets for evaluation With respect to datasets for patient monitoring, obtaining real-patient based datasets is challenging, mainly due to possible privacy issues. Publicly available datasets are available for fall detection, pain detection and activity recognition. There is a need for creation of publicly available datasets, especially in the other applications of patient monitoring. Also, datasets in which the real patient scenarios are simulated need to be developed to provide more avenues for benchmarking.

Limitations due to proximity of camera to patient Unlike fall detection and activity recognition methods, where the patient's activity is monitored within the indoor scenario, sleeping, breathing and vital signs monitoring require the patient to be in close proximity to the vision-based sensing device. This is because they rely on detecting subtler features such as blood flows, arterial color changes etc. Similarly, facial expression monitoring techniques, require the face and facial features to be visible such that the movements or appearance on the face is detectable. Such arrangement of the sensors in close proximity to the patients could be a logistic challenge in such applications, which needs to be addressed effectively.

Privacy issues While privacy is a common concern with respect to video-camera based monitoring, many techniques employ either depth based, bio-inspired sensors or IR-based sensing technologies. Video-camera based monitoring can be deployed such that privacy issues are addressed, through restricted information being passed on to the medical personnel. Certain pre-processing techniques such as blurring the face or using infrared cameras (Jiang and Kang 2007) can help in tackling privacy related issues, although very few such methods exist currently (Gade and Moeslund 2013).

Table 5 lists the different challenges and surveys how these challenges have been addressed in different PMS applications. It can be seen that there are many open areas that are yet to be fully explored in the area of vision-based patient monitoring. Given that there is an increasing innovation in the type of vision sensors and novel computer vision and machine learning algorithms, deploying them in the right proportion to this application area can yield consumer-centric solutions. Additionally, rather than applying generic algorithms, use of the application context of patient monitoring can be a useful way to develop novel techniques that are robust and yet cost-effective.

References

AL-Khalidi F, Saatchi R, Burke D, Elphick H (2010) Tracking human face features in thermal images for respiration monitoring. In: 2010 IEEE/ACS Intl. Conf. on Computer Systems and Applications (AICCSA), pp 1–6

AL-Khalidi F, Saatchi R, Burke D, Elphick H, Tan S (2011) Respiration rate monitoring methods: a review. Pediatr Pulmonol 46(6):523–529

Alghowinem S, Goecke R, Wagner M, Parker G, Breakspear M (2013a) Eye movement analysis for depression detection. In: 2013 20th IEEE Intl Conf on Image Processing (ICIP), pp 4220–4224

Alghowinem S, Goecke R, Wagner M, Parkerx G, Breakspear M (2013b) Head Pose and Movement Analysis as an Indicator of Depression. In: 2013 Humaine Association Conf. on Affective Computing and Intelligent Interaction (ACII), pp 283–288

Alkali A, Saatchi R, Elphick H, Burke D (2013) Facial Tracking in Thermal Images for Real-Time Noncontact Respiration Rate Monitoring. In: Modelling Symposium (EMS), 2013 European, pp 265–270

Alvino C, Kohler C, Barrett F, Gur RE, Gur RC, Verma R (2007) Computerized measurement of facial expression of emotions in schizophrenia. J Neurosci Methods 163(2):350–361

Amoretti M, Copelli S, Wientapper F, Furfari F, Lenzi S, Chessa S (2011) Sensor data fusion for activity monitoring in the PERSONA ambient assisted living project. J Ambient Intell Human Comput 4(1):67–84

Anderson D, Luke RH, Keller JM, Skubic M, Rantz M, Aud M (2009) Linguistic Summarization of Video for Fall Detection Using Voxel Person and Fuzzy Logic. Comput Vis Image Underst 113(1):80–89

Aoki H, Koshiji K, Nakamura H, Takemura Y, Nakajima M (2005) Study on respiration monitoring method using near-infrared multiple slit-lights projection. In: 2005 IEEE Intl. Symposium on Micro-NanoMechatronics and Human Science, pp 291–296

Ashraf AB, Prkachin K, Chen T, Lucey S, Solomon P, Ambadar Z, Cohn JF, Theobald BJ (2007) The painful face—pain expression recognition using active appearance models. In: In 9th Intl Conf on Multimodal Interfaces ICMI, pp 9–14

Ashraf AB, Lucey S, Cohn JF, Chen T, Ambadar Z, Prkachin KM, Solomon PE (2009) The painful face—pain expression recognition using active appearance models. Image Vis Comput 27(12):1788–1796

Auvinet E, Reveret L, St-Arnaud A, Rousseau J, Meunier J (2008) Fall detection using multiple cameras. In: 30th Annual Intl Conf of the IEEE Engineering in Medicine and Biology Society, 2008. EMBS 2008, pp 2554–2557

Auvinet E, Rougier C, Meunier J, St-Arnaud A, Rousseau J (2010) Multiple cameras fall dataset. DIRO-Université de Montréal, Tech Rep 1350

Auvinet E, Multon F, Saint-Arnaud A, Rousseau J, Meunier J (2011) Fall detection with multiple cameras: an occlusion-resistant method based on 3-D silhouette vertical distribution. IEEE Trans Inf Technol Biomed 15(2):290–300

Bai YW, Li WT, Chen YW (2010) Design and implementation of an embedded monitor system for detection of a patient's breath by double Webcams. In: 2010 IEEE Intl Workshop on Medical Measurements and Applications Proceedings (MeMeA), pp 171–176

Bai YW, Tsai CL, Wu SC (2012) Design of a breath detection system with multiple remotely enhanced hand-computer interaction devices. In: 2012 IEEE 16th Intl. Symposium on Consumer Electronics (ISCE), pp 1–5

Balakrishnan G, Durand F, Guttag J (2013) Detecting Pulse from Head Motions in Video. In: 2013 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pp 3430–3437

Banerjee T, Keller J, Skubic M, Stone E (2014) Day or night activity recognition from video using fuzzy clustering techniques. IEEE Trans Fuzzy Syst 22(3):483–493

Belbachir A, Schraml S, Nowakowska A (2011) Event-driven stereo vision for fall detection. In: 2011 IEEE Computer Society Conf on Computer Vision and Pattern Recognition Workshops (CVPRW), pp 78–83

Belbachir A, Litzenberger M, Schraml S, Hofstatter M, Bauer D, Schon P, Humenberger M, Sulzbachner C, Lunden T, Merne M (2012) CARE: A dynamic stereo vision sensor system for fall detection. In: 2012 IEEE Intl Symposium on Circuits and Systems (ISCAS), pp 731–734

Bevilacqua V, D'Ambruoso D, Mandolino G, Suma M (2011) A new tool to support diagnosis of neurological disorders by means of facial expressions. In: 2011 IEEE Intl. Workshop on Medical Measurements and Applications Proc. (MeMeA), pp 544–549

Bieber G, Hoffmeyer A, Gutzeit E, Peter C, Urban B (2009) Activity Monitoring by Fusion of Optical and Mechanical Tracking Technologies for User Behavior Analysis. In: Proc. of the 2Nd Intl. Conf. on PErvasive Technologies Related to Assistive Environments, ACM, PETRA '09, pp 45:1–45:6

Bin Mansor MN, Yaacob S, Nagarajan R, Hariharan M (2010) Detection of facial changes for hospital ICU patients. In: 2010 6th Intl. Colloquium on Signal Processing and Its Applications (CSPA), pp 1–5

Brulin D, Benezeth Y, Courtial E (2012) Posture recognition based on fuzzy logic for home monitoring of the elderly. IEEE Trans Inf Technol Biomed 16(5):974–982

Brusco N, Paviotti A (2005) 3d quantification of facial edemas. In: Proc. of the 4th Intl. Symposium on Image and Signal Processing and Analysis, 2005. ISPA 2005, pp 191–196

Calhoun DA (2010) Sleep and hypertension. CHEST J 138(2):434

Chaaraoui AA, Climent-Prez P, Flrez-Revuelta F (2012) A review on vision techniques applied to human behaviour analysis for ambient-assisted living. Expert Syst Appl 39(12):10873–10888

Charfi I, Miteran J, Dubois J, Atri M, Tourki R (2012) Definition and Performance Evaluation of a Robust SVM Based Fall Detection Solution. In: 2012 Eighth Intl. Conf. on Signal Image Technology and Internet Based Systems (SITIS), pp 218–224

Charfi I, Miteran J, Dubois J, Atri M, Tourki R (2013) Optimized spatio-temporal descriptors for real-time fall detection: comparison of support vector machine and Adaboost-based classification. J Electron Imaging 22(4):041106

Chekmenev SY, Rara H, Farag AA (2005) Non-contact, wavelet-based measurement of vital signs using thermal imaging. In: The first Intl. Conf on graphics, vision, and image processing (GVIP), pp 107–112

Chen LL, Chen KW, Hung YP (2014) A sleep monitoring system based on audio, video and depth information for detecting sleep events. In: 2014 IEEE Intl. Conf. on Multimedia and Expo (ICME), pp 1–6

Chen MY (2010) Long Term Activity Analysis in Surveillance Video Archives. Ph.D. thesis, Carnegie Mellon University, Pittsburgh, PA, USA

Cheng H, Liu Z, Zhao Y, Ye G (2011) Real world activity summary for senior home monitoring. In: 2011 IEEE Intl. Conf. on Multimedia and Expo (ICME), pp 1–4

Chung PC, Liu CD (2008) A daily behavior enabled hidden Markov model for human behavior understanding. Pattern Recognit 41(5):1572–1580

Cohn J, Kruez T, Matthews I, Yang Y, Nguyen MH, Padilla M, Zhou F, De la Torre F (2009) Detecting depression from facial actions and vocal prosody. In: 3rd Intl. Conf. on Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009, pp 1–7

Crispim C, Bathrinarayanan V, Fosty B, Konig A, Romdhane R, Thonnat M, Bremond F (2013) Evaluation of a monitoring system for event recognition of older people. In: 2013 10th IEEE Intl. Conf. on Advanced Video and Signal Based Surveillance (AVSS), pp 165–170

Cucchiara R, Prati A, Vezzani R (2007) A multi-camera vision system for fall detection and alarm generation. Expert Syst 24(5):334–345

Cuppens K, Lagae L, Vanrumste B (2009) Towards automatic detection of movement during sleep in pediatric patients with epilepsy by means of video recordings and the optical flow algorithm. In: 4th European Conference of the International Federation for Medical and Biological Engineering, Springer, pp 784–789

Cuppens K, Lagae L, Ceulemans B, Huffel SV, Vanrumste B (2010) Automatic video detection of body movement during sleep based on optical flow in pediatric patients with epilepsy. Med Biol Eng Comput 48(9):923–931

Cuppens K, Chen CW, Wong KBY, Van de Vel A, Lagae L, Ceulemans B, Tuytelaars T, Van Huffel S, Vanrumste B, Aghajan H (2012) Using Spatio-Temporal Interest Points (STIP) for myoclonic jerk detection in nocturnal video. Conf Proc IEEE Eng Med Biol Soc 2012:4454–4457

Dai Y, Shibata Y, Ishii T, Hashimoto K, Katamachi K, Noguchi K, Kakizaki N, Cai D (2001) An associate memory model of facial expressions and its application in facial expression recognition of patients on bed. In: IEEE Intl. Conf. on Multimedia and Expo, 2001. ICME 2001, pp 591–594, 00014

Dickinson P, Hunter A (2008) Using Inactivity to Detect Unusual behavior. In: IEEE Workshop on Motion and video Computing, 2008. WMVC 2008, pp 1–6

Diraco G, Leone A, Siciliano P (2010) An active vision system for fall detection and posture recognition in elderly healthcare. In: Design, Automation Test in Europe Conf Exhibition (DATE), 2010, pp 1536–1541

Eguchi K, Hoshide S, Ishikawa S, Shimada K, Kario K (2010) Short sleep duration is an independent predictor of stroke events in elderly hypertensive patients. J Am Soc Hypertens 4(5):255–262

Ekman P, Friesen W (1978) Facial Action Coding System: A Technique for the Measurement of Facial Movement. CA: Consulting Psychologists Press, Palo Alto, CA

Ellgring H (1989) Non verbal Communication in Depression. Cambridge University Press, Cambridge, England

Engel J, Epilepsy Intl League Against, (ILAE), (2001) A proposed diagnostic scheme for people with epileptic seizures and with epilepsy: report of the ILAE Task Force on Classification and Terminology. Epilepsia 42(6):796–803

Espa F, Ondze B, Deglise P, Billiard M, Besset A (2000) Sleep architecture, slow wave activity, and sleep spindles in adult patients with sleepwalking and sleep terrors. Clin Neurophysiol 111(5):929–939

Falie D, Ichim M, David L (2008) Respiratory Motion Visualization and the Sleep Apnea Diagnosis with the Time of Flight (TOF) Camera

Falie D, David L, Ichim M (2009) Statistical algorithm for detection and screening sleep apnea. In: Intl. Symposium on Signals, Circuits and Systems, 2009. ISSCS 2009, pp 1–4

Fei J, Pavlidis I (2006) Analysis of breathing air flow patterns in thermal imaging. In: Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE, IEEE, pp 946–952

Fei J, Pavlidis I, Murthy J (2009) Thermal vision for sleep apnea monitoring. Med Image Comput Comput Assist Interv 12(Pt 2):1084–1091

Fleck S, Strasser W (2008) Smart camera based monitoring system and its application to assisted living. Proc IEEE 96(10):1698–1714

Foo Siang Fook V, Thang PV, Htwe TM, Qiang Q, Wai AAP, Jayachandran M, Biswas J, Yap P (2007) Automated Recognition of Complex Agitation Behavior of Dementia Patients Using Video Camera. 2007 9th Intl. Conf. on e-Health Networking, Application and Services, pp 68–73

Frigola M, Amat J, Pagès J (2002) Vision Based Respiratory Monitoring System. In: Proceedings of the 10th Mediterranean Conference on Control and AutomationMED2002 Lisbon, Portugal

Fu Z, Delbruck T, Lichtsteiner P, Culurciello E (2008) An address-event fall detector for assisted living applications. IEEE Trans Biomed Circuits Syst 2(2):88–96

Gade R, Moeslund TB (2013) Thermal cameras and applications: a survey. Mach Vis Appl 25(1):245–262

Gault T, Farag A (2013) A Fully Automatic Method to Extract the Heart Rate from Thermal Video. In: 2013 IEEE Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW), pp 336–341

Gault T, Blumenthal N, Farag A, Starr T (2010) Extraction of the superficial facial vasculature, vital signs waveforms and rates using thermal imaging. In: 2010 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW), pp 1–8

Gholami B, Haddad WM, Tannenbaum AR (2009) Agitation and pain assessment using digital imaging. Conf Proc IEEE Eng Med Biol Soc 2009:2176–2179

Gilbert CA, Lilley CM, Craig KD, McGrath PJ, Court CA, Bennett SM, Montgomery CJ (1999) Postoperative pain expression in preschool children: validation of the child facial coding system. Clin J Pain 15(3):192–200

Gohil B, Gholamhhosseini H, Harrison MJ, Lowe A, Al-Jumaily A (2007) Intelligent monitoring of critical pathological events during anesthesia. Conf Proc IEEE Eng Med Biol Soc 2007:4343–4346

Grassi M, Lombardi A, Rescio G, Malcovati P, Malfatti M, Gonzo L, Leone A, Diraco G, Distante C, Siciliano P, Libal V, Huang J, Potamianos G (2008) A hardware-software framework for high-reliability people fall detection. In: 2008 IEEE Sensors, pp 1328–1331

Hamm J, Kohler CG, Gur RC, Verma R (2011) Automated Facial Action Coding System for Dynamic Analysis of Facial Expressions in Neuropsychiatric Disorders. J Neurosci Methods 200(2):237–256

Hammal Z, Cohn JF (2012) Automatic Detection of Pain Intensity. In: Proc. of the 14th ACM Intl. Conf. on Multimodal Interaction, ACM, pp 47–52

Hammal Z, Kunz M (2012) Pain monitoring: a dynamic and context-sensitive system. Pattern Recognit 45(4):1265–1280

Hawley-Hague H, Boulton E, Hall A, Pfeiffer K, Todd C (2014) Older adults perceptions of technologies aimed at falls prevention, detection or monitoring: a systematic review. Intl J Med Inf 83(6):416–426

Hazelhoff L, Han J, With PH (2008) Video-Based Fall Detection in the Home Using Principal Component Analysis. In: Proceedings of the 10th Intl Conf on Advanced Concepts for Intelligent Vision Systems, Springer-Verlag, Berlin, Heidelberg, ACIVS '08, pp 298–309

Hijaz F, Afzal N, Ahmad T, Hasan O (2010) Survey of fall detection and daily activity monitoring techniques. In: 2010 Intl Conf on Information and Emerging Technologies (ICIET), pp 1–6

Hirayama M, Nakamura T, Hori N, Koike Y, Sobue G (2008) The video images of sleep attacks in Parkinson's disease. Mov Disord 23(2):288–290

Horgas A, Miller L (2008) Pain assessment in people with dementia. AJN Am J Nurs 108(7):62–70

Humenberger M, Schraml S, Sulzbachner C, Belbachir A, Srp A, Vajda F (2012) Embedded fall detection with a neural network and bio-inspired stereo vision. In: 2012 IEEE Computer Society Conf on Computer Vision and Pattern Recognition Workshops (CVPRW), pp 60–67

Iasemidis L (2003) Epileptic seizure prediction and control. IEEE Trans Biomed Eng 50(5):549–558

Igual R, Medrano C, Plaza I (2013) Challenges, issues and trends in fall detection systems. BioMed Eng OnLine 12(1):66

Jansen B, Deklerck R (2006) Context aware inactivity recognition for visual fall detection. Pervasive Health Conf Workshops 2006:1–4

Jansen B, Temmermans F, Deklerck R (2007) 3d human pose recognition for home monitoring of elderly. In: 29th Annual Intl Conf of the IEEE Engineering in Medicine and Biology Society, 2007. EMBS 2007, pp 4049–4051

Jansen B, Rebel S, Deklerck R, Mets T, Schelkens P (2008) Detection of activity pattern changes among elderly with 3d camera technology. In: Photonics Europe, International Society for Optics and Photonics, pp 70,000O–70,000O

Jiang G, Kang L (2007) Character Analysis of Facial Expression Thermal Image. In: IEEE/ICME Intl. Conf. on Complex Medical Engineering, 2007. CME 2007, pp 824–827

Johnson KD, Winkelman C, Burant CJ, Dolansky M, Totten V (2014) The factors that affect the frequency of vital sign monitoring in the emergency department. J Emerg Nurs 40(1):27–35

Johnson ML, Price PA, Jovanov E (2007) A new method for the quantification of breathing. In: Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE, IEEE, pp 4568–4571

Joshi J, Dhall A, Goecke R, Breakspear M, Parker G (2012) Neural-net classification for spatio-temporal descriptor based depression analysis. In: 2012 21st Intl. Conf. on Pattern Recognition (ICPR), pp 2634–2638

Joshi J, Goecke R, Alghowinem S, Dhall A, Wagner M, Epps J, Parker G, Breakspear M (2013a) Multimodal assistive technologies for depression diagnosis and monitoring. J Multimodal User Interfaces 7(3):217–228

Joshi J, Goecke R, Parker G, Breakspear M (2013b) Can body expressions contribute to automatic depression analysis? In: 2013 10th IEEE Intl. Conf. and Workshops on Automatic Face and Gesture Recognition (FG), pp 1–7

Joumier V, Romdhane R, Bremond F, Thonnat M, Mulin E, Robert P, Derreumaux A, Piano J, Lee J (2011) Video Activity Recognition Framework for assessing motor behavioural disorders in Alzheimer Disease Patients. In: International Workshop on Behaviour Analysis and Video Understanding (ICVS 2011), p 9

Junior CFC, Bremond F, Joumier V (2012) A Multi-Sensor Approach for Activity Recognition in Older Patients. In: Second International Conference on Ambient Computing, Applications, Services and Technologies

Kalitzin S, Petkov G, Velis D, Vledder B, Lopes da Silva F (2012) Automatic segmentation of episodes containing epileptic clonic seizures in video sequences. IEEE Trans Biomed Eng 59(12):3379–3385

Kanade T, Cohn J, Tian Y (2000) Comprehensive database for facial expression analysis. In: Fourth IEEE Intl. Conf. on Automatic Face and Gesture Recognition, 2000. Proc., pp 46–53

Karaman S, Benois-Pineau J, Mgret R, Dovgalecs V, Dartigues JF, Gastel Y (2010) Human Daily Activities Indexing in Videos from Wearable Cameras for Monitoring of Patients with Dementia Diseases. In: 2010 20th Intl. Conf. on Pattern Recognition (ICPR), pp 4113–4116

Kayyali HA, Weimer S, Frederick C, Martin C, Basa D, Juguilon JA, Jugilioni F (2008) Remotely attended home monitoring of sleep disorders. Telemed J E Health 14(4):371–374

Khan R, Meyer A, Konik H, Bouakaz S (2013) Pain detection through shape and appearance features. In: 2013 IEEE Intl. Conf. on Multimedia and Expo (ICME), pp 1–6

Khan Z, Sohn W (2011) Abnormal human activity recognition system based on R-transform and kernel discriminant technique for elderly home care. IEEE Trans Consum Electron 57(4):1843–1850

Kosmopoulos D, Antonakaki P, Valasoulis K, Katsoulas D (2008) Monitoring Human Behavior in an Assistive Environment Using Multiple Views. In: Proc. of the 1st Intl. Conf. on PErvasive Technologies Related to Assistive Environments, ACM, New York, NY, USA, PETRA '08, pp 32:1–32:6

Kranjec J, Begu S, Gerak G, Drnovek J (2014) Non-contact heart rate and heart rate variability measurements: a review. Biomed Signal Process Control 13:102–112

Kroutil J, Laposa A, Husak M (2011) Respiration Monitoring During Sleeping. In: Proc. of the 4th Intl. Symposium on Applied Sciences in Biomedical and Communication Technologies, ACM, New York, NY, USA, ISABEL '11, pp 33:1–33:5

Kunz M, Scharmann S, Hemmeter U, Schepelmann K, Lautenbacher S (2007) The facial expression of pain in patients with dementia. PAIN 133(1–3):221–228

Kuo YM, Lee JS, Pc Chung (2010) A visual context-awareness-based sleeping-respiration measurement system. IEEE Trans Inf Technol Biomed 14(2):255–265

Kwolek B, Kepski M (2014) Human fall detection on embedded platform using depth maps and wireless accelerometer. Comput Methods Programs Biomed 117(3):489–501

Kwon S, Kim H, Park KS (2012) Validation of heart rate extraction using video imaging on a built-in camera system of a smartphone. In: 2012 Annual Intl. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBC), pp 2174–2177

Leone A, Diraco G, Siciliano P (2011) Detecting falls with 3d range camera in ambient assisted living applications: a preliminary study. Med Eng Phys 33(6):770–781

Lewandowska M, Ruminski J, Kocejko T, Nowak J (2011) Measuring pulse rate with a webcam; A non-contact method for evaluating cardiac activity. In: 2011 Federated Conf. on Computer Science and Information Systems (FedCSIS), pp 405–410

Li X, Chen J, Zhao G, Pietikainen M (2014) Remote Heart Rate Measurement from Face Videos under Realistic Situations. In: Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on, IEEE, pp 4264–4271

Li Z, da Silva A, Cunha J (2002) Movement quantification in epileptic seizures: a new approach to video-EEG analysis. IEEE Trans Biomed Eng 49(6):565–573

Liao W, Zhang W, Zhu Z, Ji Q (2005) A Real-Time Human Stress Monitoring System Using Dynamic Bayesian Network. In: IEEE Computer Society Conf. on Computer Vision and Pattern Recognition—Workshops, 2005. CVPR Workshops, pp 70–70

Liao WH, Yang CM (2008) Video-based activity and movement pattern analysis in overnight sleep studies. In: 19th Intl Conf on Pattern Recognition, 2008. ICPR 2008, pp 1–4

Littlewort GC, Bartlett MS, Lee K (2007) Faces of Pain: Automated Measurement of Spontaneousallfacial Expressions of Genuine and Posed Pain. In: Proc. of the 9th Intl. Conf. on Multimodal Interfaces, ACM, New York, NY, USA, ICMI ’07, pp 15–21

Liu Q, Sun M, Sclabassi R (2004) Illumination-invariant change detection model for patient monitoring video. In: 26th Annual Intl. Conf. of the IEEE Engineering in Medicine and Biology Society, 2004. IEMBS ’04, vol 1, pp 1782–1785

Lu H, Pan Y, Mandal B, Eng HL, Guan C, Chan DWS (2013) Quantifying limb movements in epileptic seizures through color-based video analysis. IEEE Trans Biomed Eng 60(2):461–469

Lucey P, Cohn J, Lucey S, Matthews I, Sridharan S, Prkachin KM (2009) Automatically detecting pain using facial actions. Int Conf Affect Comput Intell Interact Workshops 2009:1–8

Lucey P, Cohn J, Matthews I, Lucey S, Sridharan S, Howlett J, Prkachin K (2011a) Automatically detecting pain in video through facial action units. IEEE Trans Syst Man Cybern Part B: Cybern 41(3):664–674

Lucey P, Cohn JF, Prkachin KM, Solomon PE, Matthews I (2011b) Painful data: The unbc-mcmaster shoulder pain expression archive database. In: Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on, IEEE, pp 57–64

Lucey P, Cohn JF, Prkachin KM, Solomon PE, Chew S, Matthews I (2012) Painful monitoring: automatic pain monitoring using the UNBC-McMaster shoulder pain expression archive database. Image Vis Comput 30(3):197–205

Makantasis K, Protopapadakis E, Doulamis A, Grammatikopoulos L, Stentoumis C (2012) Monocular Camera Fall Detection System Exploiting 3d Measures: A Semi-supervised Learning Approach. In: Fusiello A, Murino V, Cucchiara R (eds) Computer Vision ECCV 2012. Workshops and Demonstrations, no. 7585 in Lecture Notes in Computer Science, Springer Berlin Heidelberg, pp 81–90

Malakuti K, Albu A (2010) Towards an Intelligent Bed Sensor: Non-intrusive Monitoring of Sleep Irregularities with Computer Vision Techniques. In: 2010 20th Intl Conf on Pattern Recognition (ICPR), pp 4004–4007

Mandal MK, Pandey R, Prasad AB (1998) Facial expressions of emotions and schizophrenia: a review. Schizophr Bull 24(3):399–412

Manfredi PL, Breuer B, Meier DE, Libow L (2003) Pain assessment in elderly patients with severe dementia. J Pain Symptom Manag 25(1):48–52

Martinez M, Stiefelhagen R (2012) Breath rate monitoring during sleep using near-ir imagery and PCA. In: 2012 21st Intl. Conf. on Pattern Recognition (ICPR), pp 3472–3475

Mastorakis G, Makris D (2012) Fall detection system using Kinects infrared sensor. J Real-Time Image Proc pp 1–12

Matic A, Mehta P, Rehg J, Osmani V, Mayora O (2010) AID-ME: Automatic identification of dressing failures through monitoring of patients and activity Evaluation. In: Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2010 4th Intl. Conf. on-NO PERMISSIONS, pp 1–8

McDuff D, El Kaliouby R, Senechal T, Amr M, Cohn J, Picard R (2013) Affectiva-MIT Facial Expression Dataset (AM-FED): Naturalistic and Spontaneous Facial Expressions Collected In-the-Wild. In: 2013 IEEE Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW), pp 881–888

McIntyre G, Gocke R, Hyett M, Green M, Breakspear M (2009) An approach for automatically measuring facial activity in depressed subjects. In: 3rd Intl. Conf. on Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009, pp 1–8

Mihailidis A, Carmichael B, Boger J (2004) The use of computer vision in an intelligent environment to support aging-in-place, safety, and independence in the home. IEEE Trans Inf Technol Biomed 8(3):238–247

Mirmahboub B, Samavi S, Karimi N, Shirani S (2013) Automatic monocular system for human fall detection based on variations in silhouette area. IEEE Trans Biomed Eng 60(2):427–436

Mishima K, Sugahara T (2009) Analysis methods for facial motion. Jpn Dental Sci Rev 45(1):4–13

Mubashir M, Shao L, Seed L (2013) A survey on fall detection: principles and approaches. Neurocomputing 100:144–152

Murthy R, Pavlidis I, Tsiamyrtzis P (2004) Touchless monitoring of breathing function. Conf Proc IEEE Eng Med Biol Soc 2:1196–1199

Nait-Charif H, McKenna S (2004) Activity summarisation and fall detection in a supportive home environment. In: Proceedings of the 17th Intl Conf on Pattern Recognition, 2004. ICPR 2004, vol 4, pp 323–326 Vol.4

Nakajim K, Matsumoto Y, Tamura T (2001) Development of real-time image sequence analysis for evaluating posture change and respiratory rate of a subject in bed. Physiol Meas 22(3):N21–28

Nakajima K, Matsumoto Y, Tamura T (2000) A monitor for posture changes and respiration in bed using real time image sequence analysis. In: Proc. of the 22nd Annual Intl. Conf. of the IEEE Engineering in Medicine and Biology Society, 2000, vol 1, pp 51–54 vol.1

Nanni L, Brahnam S, Lumini A (2010) A local approach based on a local binary patterns variant texture descriptor for classifying pain states. Expert Syst Appl 37(12):7888–7894

Nasution A, Emmanuel S (2007) Intelligent Video Surveillance for Monitoring Elderly in Home Environments. In: IEEE 9th Workshop on Multimedia Signal Processing, 2007. MMSP 2007, pp 203–206

Naufal Mansor M, Yaacob S, Nagarajan R, Che LS, Hariharan M, Ezanuddin M (2010) Detection of facial changes for ICU patients using KNN classifier. In: 2010 Intl. Conf. on Intelligent and Advanced Systems (ICIAS), pp 1–5

Needleman J, Buerhaus P, Mattke S, Stewart M, Zelevinsky K (2002) Nurse-staffing levels and the quality of care in hospitals. N Engl J Med 346(22):1715–1722

Neubauer DN (1999) Sleep problems in the elderly. Am Fam Phys 59(9):2551–2558, 2559–2560

Ni B, Dat NC, Moulin P (2012) RGBD-camera based get-up event detection for hospital fall prevention. In: 2012 IEEE Intl Conf on Acoustics, Speech and Signal Processing (ICASSP), pp 1405–1408

Nicolas Thome SM (2008) A real-time, multi-view fall detection system: a LHMM-based approach. IEEE Trans Circuits Syst Video Technol 18:1522–1532

Noachtar S, Peters AS (2009) Semiology of epileptic seizures: a critical review. Epilepsy Behav 15(1):2–9

Noury N, Fleury A, Rumeau P, Bourke A, Laighin G, Rialle V, Lundy JE (2007) Fall detection—Principles and Methods. In: 29th Annual Intl Conf of the IEEE Engineering in Medicine and Biology Society, 2007. EMBS 2007, pp 1663–1666

Obdrzalek S, Kurillo G, Ofli F, Bajcsy R, Seto E, Jimison H, Pavel M (2012) Accuracy and robustness of Kinect pose estimation in the context of coaching of elderly population. In: 2012 Annual Intl. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBC), pp 1188–1193

Olivieri DN, Gmez Conde I, Vila Sobrino XA (2012) Eigenspace-based fall detection and activity recognition from motion templates and machine learning. Expert Syst Appl 39(5):5935–5945

Pantelopoulos A, Bourbakis NG (2010) A survey on wearable sensor-based systems for health monitoring and prognosis. IEEE Trans Syst Man Cybern Part C-Appl Rev 40(1):1–12

Paradiso R (2003) Wearable health care system for vital signs monitoring. In: 4th Intl. IEEE EMBS Special Topic Conf. on Information Technology Applications in Biomedicine, 2003, pp 283–286

Pediaditis M, Tsiknakis M, Vorgia P, Kafetzopoulos D, Danilatou V, Fotiadis D (2010) Vision-based human motion analysis in epilepsy—Methods and challenges. In: 2010 10th IEEE Intl. Conf. on Information Technology and Applications in Biomedicine (ITAB), pp 1–5

Pediaditis M, Tsiknakis M, Koumakis L, Karachaliou M, Voutoufianakis S, Vorgia P (2012) Vision-based absence seizure detection. Conf Proc IEEE Eng Med Biol Soc 2012:65–68

Peng YT, Lin CY, Sun MT, Feng MW (2006) Sleep condition inferencing using simple multimodality sensors. In: Circuits and Systems, 2006. ISCAS 2006. Proceedings. 2006 IEEE International Symposium on, IEEE, pp 4–pp

Planinc R, Kampel M (2013) Introducing the use of depth data for fall detection. Pers Ubiquit Comput 17(6):1063–1072

Poh MZ, McDuff DJ, Picard RW (2010) Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. Opt Express 18(10):10762–10774

Prkachin KM (1992) The consistency of facial expressions of pain: a comparison across modalities. Pain 51(3):297–306

Prkachin KM (2009) Assessing pain by facial expression: facial expression as nexus. Pain Res Manag 14(1):53–58

Rashidi P, Mihailidis A (2013) A survey on ambient-assisted living tools for older adults. IEEE J Biomed Health Inf 17(3):579–590

Rodrigues J, Estevao M, Malaquias J, Santos P, Gouveia G, Simoes J (2007) SleepAtHome—Portable Home Based System for Pediatric Sleep Apnoea Diagnosis. In: IEEE Intl. Conf. on Portable Information Devices, 2007. PORTABLE 07, pp 1–4

Rogers CR, Schmidt KL, VanSwearingen JM, Cohn JF, Wachtman GS, Manders EK, Deleyiannis FWB (2007) Automated facial image analysis: detecting improvement in abnormal facial movement after treatment with botulinum toxin A. Ann Plast Surg 58(1):39–47

Rougier C, Auvinet E, Rousseau J, Mignotte M, Meunier J (2011a) Fall Detection from Depth Map Video Sequences. In: Abdulrazak B, Giroux S, Bouchard B, Pigot H, Mokhtari M (eds) Toward Useful Services for Elderly and People with Disabilities, no. 6719 in Lecture Notes in Computer Science, Springer Berlin Heidelberg, pp 121–128

Rougier C, Meunier J, St-Arnaud A, Rousseau J (2011b) Robust video surveillance for fall detection based on human shape deformation. IEEE Trans Circuits Syst Video Technol 21(5):611–622

Rougier C, Meunier J, St-Arnaud A, Rousseau J (2013) 3d head tracking for fall detection using a single calibrated camera. Image Vis Comput 31(3):246–254

Sathyanarayana S, Satzoda RK, Sathyanarayana S, Thambipillai S (2014) Block-Based Search Space Reduction Technique for Face Detection Using Shoulder and Head Curves. In: Image and Video Technology, no. 8333 in Lecture Notes in Computer Science, Springer Berlin Heidelberg, pp 385–396

Sathyanarayana S, Satzoda R, Sathyanarayana S, Thambipillai S (2015a) Identifying epileptic seizures based on a template-based eyeball detection technique. In: 2015 IEEE Intl Conf on Image Processing (ICIP), p [in Press]

Sathyanarayana S, Satzoda R, Sathyanarayana S, Thambipillai S (2015b) WellCam: Dataset for Vision-based Patient Wellness Monitoring. CVPR Workshop on The Future of Datasets in Vision

Sato I, Nakajima M (2005) Non-contact Breath Motion Monitor ing System in Full Automation. In: Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual Intl. Conf. of the, pp 3448–3451

Seki H (2009) Fuzzy inference based non-daily behavior pattern detection for elderly people monitoring system. Conf Proc IEEE Eng Med Biol Soc 2009:6187–6192

Sikka K, Dhall A, Bartlett MS (2014) Classification and weakly supervised pain localization using multiple segment representation. Image Vis Comput 32(10):659–670

Somers VK, White DP, Amin R, Abraham WT, Costa F, Culebras A, Daniels S, Floras JS, Hunt CE, Olson LJ, Pickering TG, Russell R, Woo M, Young T (2008) Sleep apnea and cardiovascular disease. J Am Coll Cardiol 52(8):686–717

Stevens S, Bharucha A (2003) Caremedia: Automated video and sensor analysis for geriatric care. Carnegie Mellon University

Sun N, Garbey M, Merla A, Pavlidis I (2005) Imaging the cardiovascular pulse. In: IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, 2005. CVPR 2005, vol 2, pp 416–421 vol. 2

Tabar AM, Keshavarz A, Aghajan H (2006) Smart Home Care Network Using Sensor Fusion and Distributed Vision-based Reasoning. In: Proceedings of the 4th ACM Intl Workshop on Video Surveillance and Sensor Networks, ACM, New York, NY, USA, VSSN '06, pp 145–154

Takemura Y, Jy Sato, Nakajima M (2005) A respiratory movement monitoring system using fiber-grating vision sensor for diagnosing sleep apnea syndrome. OPT REV 12(1):46–53

Toreyin B, Dedeoglu Y, Cetin A (2006) HMM Based Falling Person Detection Using Both Audio and Video. In: Signal Processing and Communications Applications, 2006 IEEE 14th, pp 1–4

UN (2012) World population ageing 1950–2050. Department of Economic and Social Affairs, Population Division, United Nations, http://www.un.org/esa/population/publications/worldageing19502050/

Van Kasteren TLM (2011) Activity Recognition for Health Monitoring Elderly Using Temporal Probabilistic Models. ASCI

Viola P, Jones M (2001) Robust real-time object detection. Int J Comput Vis 4:51–52

Vishwakarma S, Agrawal A (2012) A survey on activity recognition and behavior understanding in video surveillance. Vis Comput 29(10):983–1009

Wang CW, Ahmed A, Hunter A (2006) Vision analysis in detecting abnormal breathing activity in application to diagnosis of obstructive sleep apnoea. Conf Proc IEEE Eng Med Biol Soc 1:4469–4473

Wang CW, Ahmed A, Hunter A (2007) Locating the upper body of covered humans in application to diagnosis of obstructive sleep apnea. In: World Congress on Engineering, pp 662–667

Wang CW, Hunter A, Gravill N, Matusiewicz S (2010) Real time pose recognition of covered human for diagnosis of sleep apnoea. Comput Med Imaging Graph 34(6):523–533

Wang CW, Hunter A, Gravill N, Matusiewicz S (2014) Unconstrained video monitoring of breathing behavior and application to diagnosis of sleep apnea. IEEE Trans Biomed Eng 61(2):396–404

Wang F, Stone E, Dai W, Banerjee T, Giger J, Krampe J, Rantz M, Skubic M (2009) Testing an in-home gait assessment tool for older adults. Conf Proc IEEE Eng Med Biol Soc 2009:6147–6150

Wang P, Barrett F, Martin E, Milanova M, Gur RE, Gur RC, Kohler C, Verma R (2008) Automated video based facial expression analysis of neuropsychiatric disorders. J Neurosci Methods 168(1):224–238

Wang S, Xu Z, Yang Y, Li X, Pang C, Haumptmann AG (2013) Fall Detection in Multi-camera Surveillance Videos: Experimentations and Observations. In: Proceedings of the 1st ACM Intl Workshop on Multimedia Indexing and Information Retrieval for Healthcare, ACM, New York, NY, USA, MIIRH ’13, pp 33–38

Werner P, Al-Hamadi A, Niese R (2012) Pain recognition and intensity rating based on Comparative Learning. In: 2012 19th IEEE Intl. Conf. on Image Processing (ICIP), pp 2313–2316

Werner P, Al-Hamadi A, Niese R, Walter S, Gruss S, Traue HC (2013) Towards Pain Monitoring: Facial Expression, Head Pose, a new Database, an Automatic System and Remaining Challenges. In: Proceedings of the British Machine Vision Conference

Wiesner S, Yaniv Z (2007) Monitoring patient respiration using a single optical camera. Conf Proc IEEE Eng Med Biol Soc 2007:2740–2743

Yang FC, Kuo CH, Tsai MY, Huang SC (2003) Image-based sleep motion recognition using artificial neural networks. In: 2003 Intl. Conf on Machine Learning and Cybernetics, vol 5, pp 2775–2780 Vol.5

Yu M, Rhuma A, Naqvi S, Wang L, Chambers J (2012) A posture recognition-based fall detection system for monitoring an elderly person in a smart home environment. IEEE Trans Inf Technol Biomed 16(6):1274–1286

Yu MC, Wu H, Liou JL, Lee MS, Hung YP (2013) Multiparameter Sleep Monitoring Using a Depth Camera. In: Gabriel J, Schier J, Huffel SV, Conchon E, Correia C, Fred A, Gamboa H (eds) Biomedical Engineering Systems and Technologies, no. 357 in Communications in Computer and Information Science, Springer Berlin Heidelberg, pp 311–325

Yu X (2008) Approaches and principles of fall detection for elderly and patient. In: 10th Intl Conf on e-health Networking, Applications and Services, 2008. HealthCom 2008, pp 42–47

Zambanini S, Machajdik J, Kampel M (2010) Detecting falls at homes using a network of low-resolution cameras. In: 2010 10th IEEE Intl Conf on Information Technology and Applications in Biomedicine (ITAB), pp 1–4

Zhan K, Ramos F, Faux S (2012) Activity recognition from a wearable camera. In: 2012 12th Intl. Conf. on Control Automation Robotics Vision (ICARCV), pp 365–370

Zhang C, Tian Y (2012) Rgb-d camera-based daily living activity recognition. J Comput Vis Image Process 2(4):12

Zhang Z, Kapoor U, Narayanan M, Lovell N, Redmond S (2011) Design of an unobtrusive wireless sensor network for nighttime falls detection. In: 2011 Annual Intl Conf of the IEEE Engineering in Medicine and Biology Society,EMBC, pp 5275–5278

Zhong H, Shi J, Visontai M (2004) Detecting unusual activity in video. In: Proc. of the 2004 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, 2004. CVPR 2004, vol 2, pp II–819–II–826 Vol.2

Zhou Z, Dai W, Eggert J, Giger J, Keller J, Rantz M, He Z (2009) A real-time system for in-home activity monitoring of elders. In: Annual Intl. Conf. of the IEEE Engineering in Medicine and Biology Society, 2009. EMBC 2009, pp 6115–6118

Zhu Z, Fei J, Pavlidis I (2005) Tracking human breath in infrared imaging. In: Fifth IEEE Symposium on Bioinformatics and Bioengineering, 2005. BIBE 2005, pp 227–231

Zouba N, Bremond F, Thonnat M, Vu VT (2007) Multi-sensors Analysis for Everyday Activity Monitoring. Proc of SETIT pp 25–29