Automated Multi-Camera System for Long Term Behavioral Monitoring in Intensive Care Units

Manuel Martinez and Rainer Stiefelhagen
Institute of Anthropomatics, Karlsruhe Institute of Technology
Karlsruhe, Germany
manuel.martinez@kit.edu and rainer.stiefelhagen@kit.edu

Abstract

Automated vision-based systems have been suggested for complementary monitoring in Intensive Care Units (ICU) due to their ability to detect behavioral cues used in sedation delivery and accident detection. However, data acquisition is a major limitation: single-camera systems are reliable but not very capable where complex systems are unable to work unattendedly. This has prevented the development of complex behavioral analysis algorithms. In this paper we present a Medical Recording Device (MRD) developed for long term ICU monitoring, including three major vision components: stereo, depth and hi-res, together with a number of secondary sensors. Unlike current approaches which require controlled environments, image markers and optimal lighting conditions, the MRD is capable of registering behavioral cues autonomously regardless of the environment conditions.

1 Introduction

Determining a correct sedation protocol for Intensive Care Unit (ICU) patients is a complex and individualized procedure. Excessive sedation can be dangerous while insuficient sedation increases the risk of excessive anxiety and agitation. The process is iterative and depends on the patient feedback, but common vital signs monitored in ICUs are not enough by themselves; behavioral cues, which are annotated by the nursing staff, need to be considered too.

Therefore Computer Vision (CV) systems had been suggested to avoid the undesirable subjective nature of those behavioural annotations by providing strong objective measurements. The most used behavior cue is agitation [4], as it is meaningful, robust to occlusions, and easy to measure, but other considered cues are: facial changes [2, 8], awaken/sleep state [11] and breathing patterns [1, 9].

Accident detection is the second main use of CV systems in ICU's [6]. Disoriented patients can hurt themselves inadvertently. While the most common event is simply falling, Accidental Catheter Removal [7] (ACR) is common and can have severe consequences if it not detected quicky by the staff.

Although there have been some CV attemps at those problems in the past, they mostly deal with synthetic benchmarks [2, 4, 8, 11]. The longest field test [6] (6 days, 1 bed) failed to record any accident and therefore accident detection was evaluated in a simulated scenario.

Those short field tests and simulated scenarios fail to sample the extreme diversity of conditions that happen in ICU rooms. Accidents are, luckyly, very rare

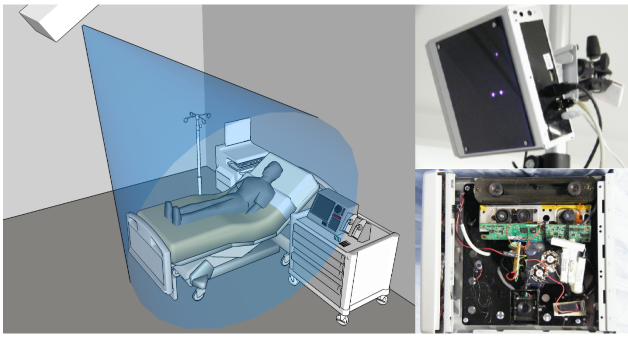

Figure 1. The Medical Recording Device monitors the patient and the ICU environment. Stereo and depth cameras cover the entire body, while a hi-res camera covers the face.

(the mean ACR rate has been observed of 1.5 per 100 days [7]). To overcome this scarcity of events, observational studies are lengthy, with monitoring periods as long as 300000 hours (around 2000 patients) not unusual [7].

The lack of a CV corpus of this magnitude hinders the development of advanced CV algorithms, and makes the evaluation of current approaches impossible.

In this paper we present a novel system for automated CV monitoring of ICU rooms, with the goal of long-term data recording. The resulting dataset will be used to develop and evaluate behavioral algorithms that could be of assistive use for sedation control and accident avoidance.

The capabilities of the system were tested on current ICU challenges, and their autonomy and robustness were throughly evaluated before installing them in an hospital.

2 The Medical Recording Device

A data collection of this magnitude has several challenges: it must be secure, robust and ethical. But at the same time must be autonomous and adjust automatically to the varying conditions: lighting, noise, occlusions, sleeping positions, etc.

Finally, a dataset is useless if the correct data modality is not captured. From our analysis of the previous work, we see that some systems require a single camera [6] while stereo cameras are used in several action recognition frameworks [12]. Depth cameras based on Kinect are increasingly being used [10]. Finally the field of view (FoV) recorded is also crucial. Although some systems use only face information [2, 8, 11], our medical advisors suggested us to monitor the full body as in [6]. We decided to monitor all information we found possible while keeping a low profile, which is an important factor as the monitoring system should be unobstrusive in order to not interfere with the normal ICU enviroment.

We designed our Medical Recording Device (MRD) around three main vision systems: One stereo camera, one depth camera, and one camera for face analysis. The MRD is meant to work during day and night, therefore all cameras capture only infrared images. A coupled IR illuminator is attached to the device, and a control and communications center manages all devices and sends the information to the file server using a single Gigabit Ethernet link. Secondary sensors are provided to capture several environmental conditions and complement the monitoring task.

2.1 Architecture

The Stereo and Face cameras are USB cameras provided by Ecomunicat Electronics, allowing raw access to the image sensors. Stereo camera images are downsampled to QWVGA@10fps to save bandwidth, and the face camera is a 1.3MP model configured as a flying Region of Interest to follow the face in realtime and capture it at 256x256@50fps.

The Depth camera is based on Kinect, and it is configured at VGA@30fps.

The Control and Communications Center: A 1.2GHz Marvell Kirkwood processor (based on ARM) was selected as it was designed for efficient USB to Gigabit Ethernet transfers. The chosen operative system was Debian Linux. PC based systems were ruled out for budgetary and ambiental reasons.

2.2 Low-profile

The monitoring must not hinder the normal functioning of the ICU, therefore modifying the environment, adding markers or using custom colored clothing was strictly forbidden. The MRD itself is packaged as a single box of 192x210x62mm. However IR illuminator is placed externally to avoid overheating. As a long term monitoring solution, it was designed to be easy to maintain, and the full system is attached using an standard Vesa mount allowing simple cleaning and replacement.

Finally the visual profile was also reduced by hiding all cameras behind an IR-bandpass filter, as it is known that visible cameras can increase the anxiety to the recorded subject.

2.3 Robustness

The MRDs have been tested against a series of physical and technical threats. The software running in the MRD deals with communication problems: hot plugging, network congestion, data corruption, etc. It also manages all sensors, monitoring their status, restarting them if required, altering the parameters to adapt to the current environment, etc. Finally it continuously checks the system integrity including disk, time synchronization and checking for upgrades.

2.4 Performance

As the MRD is mainly a data transmission system, most performance problems deal with data throughput. Memory interfaces in embedded processors are several orders of magnitude slower than PCs. Therefore custom drivers for the USB cameras and Kinect were developed where unnecessary active memory transfers were avoided and using DMA channels where available. The communication protocol was designed in parallel to even further minimize memory transfers.

2.5 Storage limitations

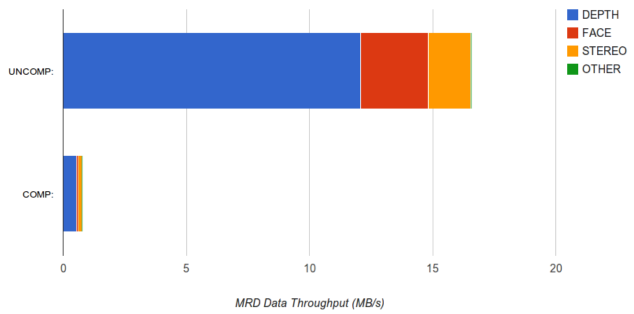

Figure 2. Hybrid lossless/lossy compression achieves a ratio of 21:1.

The artifacts caused by lossy video compression algorithms are designed to be barely perceptible to the human eye, but some Computer Vision algorithms are susceptible to them and thus its use is discouraged. However, the uncompressed data rate of a MRD is around 16.5MB/s, which equals to around 1.4TB/day. Storing a dataset of around 300000 hours would use 16.5 PetaBytes: equivalent to 26 full-size racks of storage servers even without considering backups. Lossless algorithms do not provide a significant improvement as their compression rate is small (around 3:1). Knowing that in most cases there is almost no motion in ICUs, we developed a hybrid lossless/lossy compression algorithm that compresses images in one second packs, storing one lossless image and several lossy images per pack using inter-frame compression. When the system detects activity in the image, it switches to a full lossless compression system. In our experiments we obtain a compression rate of 21:1 using our compression system to achieve a global rate of 800KB/s (Fig. 2).

2.6 Privacy

Privacy is a great concern when dealing with medical data, and it covers the patient as well as the ICU staff. Therefore we have several mechanisms in place in order to preserve privacy. Including filesystem and communications encryption.

The suggested procedure to deal with private events (such as when the patient is being washed), is to turn off the visual recordings using a remote control. Alternatively there is a 3 days safety buffer where the ICU staff can mark a sequence as compromised, and will be not be permanently stored.

3 Multi-Camera Calibration

A tool was designed to simplify the multi-camera calibration procedure by calibrating all the cameras simultaneously. Once the procedure has started, a calibration pattern is placed firmly in the field of view of the system, then we take a snapshot using the remote control. The snapshot takes a few seconds as all cameras activate iteratively setting themselves at maximum resultion and adjusting to the best ratio of gain and exposure before taking the picture. The Kinect is configured first as a IR camera to capture the pattern, and then switches to depth to capture a disparity image. When the sequence finishes the computer notifies the user and the calibration pattern can be moved to the next position. This method allows the system to be recalibrated in-place in less than 5 minutes. The snapshots are used to get the intrinsic parameters of all the cameras using [14]. To get the extrinsic parameters, the face camera acts as the master, and the other cameras are pairwise calibrated to it.

By using Kinect IR image for extrinsic calibration, we avoid the complexity of novel depth-to-color calibration methods like [5].

4 Face camera and Head Detection

Figure 3. The face camera allows to record hi-res, hi-speed images of the face.

The most important information we can get about the behavior of a person lies in his face. However, algorithms like facial grimacing recognition [2] require high resolution images, while microexpression recognition requires a medium-high framerate [3].

The face is not in a fixed position in the FoV, but sending the full FoV image at high framerates is not possible due to bandwidth constraints (1.3MP@50fps=62.5MB/s).

Pan-and-tilt cameras are not advisable as they are larger, noisier, and its calibration is less precise than a fixed camera.

We resolved to use a fixed camera with a FoV covering all the scene, but configured with a flying ROI of 256x256@50fps that is actively aimed at the face. This limits the bandwidth requirement to 3.125MB/s.

The ROI is aimed by triangulating the face position as detected by the stereo cameras, and projecting the 3D position into the image plane of the face camera (see Fig. 3).

5 Framework

The main idea behind the MRD is to fuse the information from multiple image sensors to improve environment awareness. To this end we have developed a camera fusion framework. It allows us to project data from one camera to another, triangulate objects to estimate its 3d position, and analyze the scene as a pointcloud.

The user interface of the framework is 3d enabled, and using a 3d mouse it is possible to navigate through the environment and thus simulate virtual point of views of the scene, as shown in Fig. 4.

6 Performance Evaluation

Figure 5. We can simultaneously visualize images from the all sensors alongside debugging information. Note the accurate detection of the bed in the depth field (blue mask).

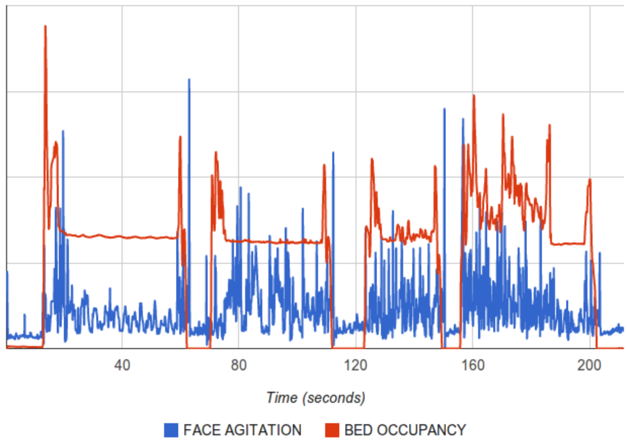

Figure 6. This sequence simulates 4 scenarios. 10s-60s: Sleeping relaxed shows an almost flat bed occupancy indicator and low agitation levels in the face. 70s-110s: Sleeping with pain expressions is not reflected in the volumetric information, but it is detected by the face agitation levels. 120s-145s: Being restless in bed is reflected by an clear response in both indicators. 145s-200s: Strong compulsions ending with an accident and sudden loss of consciousness.

We evaluate the performance of the MRD and the Framework on current ICU challenges to ensure that the device will record relevant data for the development of future algorithms. To detect the bed position, we assume it is roughly centered in the field of view, and we estimate the bed plane by region growing from the depth map, results of the obtained ROI can be seen in Fig. 5. Then two indicators are extracted from the image information: the *bed occupancy* indicator is estimated from the depth camera and is a rough indicator of the volume over the bed. It can be used to detect the events corresponding to entering and exiting the bed, body agitation and even breathing. The second indicator used is the *face agitation*, which is obtained from the face camera. Each image is resized to 32x32 pixels

Figure 4. The framework and integrated interface allows us to navigate through the environment in 3d.

and then Bayesian Surprise is calculated using a Gaussian window of mean 25 frames ($500ms$). Bayesian Surprise has been used to detect salient events [13], and the *face agitation* indicator displays activity when the patient shows discomfort.

Results of this detectors can be seen in a short simulated test in Fig. 6, there we aimed to register the behavior of the scene using both indicators.

Due to the subjective behavior of the measurements and the lack of a public database or even a common methodology for evaluation, it is not possible to directly compare our results to alternative approaches. We focused instead on showing how a reasonable behavioral description of an scenario can be obtained in a non-invasive way, without using markers, and is robust to changes in illumination. In this regard our approach proves to be superior to state of the art systems [2,4,6].

7 Conclusions and Future Work

We have shown a monitoring system for ICUs designed to capture a long term multimedia data set for behavioral analysis. We have described several problems associated with long term medical monitoring, and a set of possible solutions to them.

The combination of multiple vision modalities on our Medical Recording Device will allows us to work with a large variety of ICU monitoring algorithms, as we will have data of the whole body of the patient, high speed and high resolution data of the patients face, and a depth data provided by Kinect. To show the effectiveness of the system we implemented a simple algorithm that provides indicators for *bed occupancy* and *face agitation*. These indicators can be obtained regardless of the illumination conditions and without human interaction, and are used to register a continuous behavioral profile of the monitored patient.

Once the dataset is available, we plan to use it to develop accident detecting algorithms and advanced behavior recognition algorithms able to better aid in ICU rooms.

Acknowledgements: This work is supported by the German Federal Ministry of Education and Research (BMBF) within the VIPSAFE project.

References

[1] H. Aoki, Y. Takemura, K. Mimura, and M. Nakajima, *Development of non-restrictive sensing system for sleeping person using fiber grating vision sensor*, Micromechatronics and Human Science, 2001.

[2] P. Becouze, C.E. Hann, J.G. Chase, and G.M. Shaw, *Measuring facial grimacing for quantifying patient agitation in critical care*, Computer Methods and Programs in Biomedicine, 2007.

[3] T. Gehrig and H.K. Ekenel, *A common framework for real-time emotion recognition and facial action unit detection*, Computer Vision and Pattern Recognition Workshops (CVPRW), 2011.

[4] J. Geoffrey Chase, F. Agogue, C. Starfinger, Z.H. Lam, G.M. Shaw, A.D. Rudge, and H. Sirisena, *Quantifying agitation in sedated icu patients using digital imaging*, Computer Methods and Programs in Biomedicine, 2004.

[5] D. Herrera C, J. Kannala, and J. Heikkilä, *Accurate and practical calibration of a depth and color camera pair*, Computer Analysis of Images and Patterns, 2011.

[6] P. Kittipanya-Ngam, O.S. Guat, and E.H. Lung, *Computer vision applications for patients monitoring system*, Information Fusion (FUSION), 2012.

[7] L. Lorente, M.S. Huidobro, M.M. Martín, A. Jiménez, M.L. Mora, et al., *Accidental catheter removal in critically ill patients: a prospective and observational study*, Crit Care, 2004.

[8] M.N. Mansor, S. Yaacob, R. Nagarajan, L.S. Che, M. Hariharan, and M. Ezanuddin, *Detection of facial changes for icu patients using knn classifier*, Intelligent and Advanced Systems (ICIAS), 2010.

[9] M. Martinez and R. Stiefelhagen, *Breath rate monitoring during sleep using near-ir imagery and pca*, International Conference on Pattern Recognition (ICPR), 2012.

[10] V. Megavannan, B. Agarwal, and R.V. Babu, *Human action recognition using depth maps*, Signal Processing and Communications (SPCOM), 2012.

[11] M. Naufal Bin Mansor, S. Yaacob, R. Nagarajan, and M. Hariharan, *Patient monitoring in icu under unstructured lighting condition*, Industrial Electronics & Applications (ISIEA), 2010.

[12] L. Rybok, S. Friedberger, U.D. Hanebeck, and R. Stiefelhagen, *The kit robo-kitchen data set for the evaluation of view-based activity recognition systems*, Humanoid Robots (Humanoids), 2011.

[13] B. Schauerte, B. Kuhn, K. Kroschel, and R. Stiefelhagen, *Multimodal saliency-based attention for object-based scene analysis*, Intelligent Robots and Systems (IROS), 2011.

[14] Z. Zhang, *A flexible new technique for camera calibration*, Pattern Analysis and Machine Intelligence, 2000.