An Intelligent Multi-Sensor Surveillance System for Elderly Care

Sou-Young Jin¹, Young-Seob Jeong¹, Chanky Park², KyoJoong Oh¹ and Ho-Jin Choi¹

¹ Dept. of Computer Science, KAIST, 291 Daehak-ro, Yuseong-gu, Daejeon 305-701, Korea
² Dept. of Robot/Cognitive System Research, Electronics & Telecommunication Research Institute (ETRI), Daejeon, Korea

* Corresponding Author: Ho-Jin Choi

Received May 27, 2012; Revised July 29, 2012; Accepted August 6, 2012; Published August 31, 2012

Abstract: This paper is an overview of our on-going project that proposes a monitoring system based on various sensors to detect risky situations for the elderly. From the standpoint of the end-user, a video surveillance system equipped with many other sensors can relieve caregivers from the need to keep a vigilant eye on each patient’s movements, while such technology can be effectively used for monitoring elderly people with dementia. Since a camera surveillance system has limits to classify complex human actions, this project aims to design an intelligent healthcare surveillance system, which extends the conventional automated video surveillance system with various additional sensors, to improve the performance of surveillance. The main contributions of our proposed system will be: (i) minimize human intervention; (ii) detect more complex activities and situations using various sensors and improved sensor fusion techniques; and (iii) design a novel classifier that identifies risky situations with the collected information.

Keywords: Surveillance System, Elderly Care, Sensor Fusion, Data Mining, Activity Recognition

Introduction

In a u-healthcare society, patient monitoring and surveillance systems based on information and communication technology are attractive for their potential to reduce the burden and cost of giving care to elderly people while maintaining safety and autonomy. Especially, a video surveillance system equipped with many other sensors can relieve caregivers from the need to keep a vigilant eye on their patients’ movements, and such technology can be effectively used...

† This work was supported by the National Research Foundation (NRF) grant (No. 2012-0001001) of Ministry of Education, Science and Technology (MEST) of Korea.
‡ A preliminary version of this paper was awarded as the Outstanding Paper at the 2012 International Conference on Smart Convergence Technologies and Applications (SCTA2012), held on 7-9 August 2012 in Gwangju, South Korea.

DOI: 10.6029/smartcr.2012.04.006
for monitoring elderly people. Thus, there exists good potential for surveillance technologies to assist and improve the quality of life for those elderly people. This project aims to design an intelligent healthcare surveillance system which augments a conventional automated video surveillance system with additional sensors such as audio sensors, wireless sensors, and so on. The key idea is to fuse these various kinds of sensor information with normal image tracking information to improve the performance of surveillance.

An aging population created by great advancements in medical science has brought about many important problems to the human race. Dementia is one such problem, and surveillance technologies have a great potential to assist elderly people with dementia. Providing 24 hour supervision by human caregivers is extremely expensive, if ever arranged, and an automated video monitoring and surveillance system should be considered as a viable alternative. In this project, we will develop surveillance technologies for monitoring and analyzing behavioral patterns of elderly people with dementia to detect potentially risky situations. When successful, the results of this project will make contributions to the field of ubiquitous healthcare.

Our aim is to develop a distributed, intelligent multi-sensor surveillance system for monitoring and caring for elderly people in a hospital or in a nursing home. A camera-based surveillance system has limits to complex classify human actions [1]. From a technical standpoint, this means designing an intelligent healthcare surveillance system that extends the conventional automated video surveillance system by fusing the additional information gained from various additional sensors to improve the performance of surveillance. From an application standpoint, we will aim to apply a surveillance system for caring for elderly people who require continuous monitoring (e.g., those with dementia). In the process, we will use data mining, machine learning, and other artificial intelligence techniques to understand and reason about human behavioral patterns for the early detection of emergency situations.

Related Work

Related Projects

As the number of elderly people living alone increases, the elderly care industries does as well. The Ambient Assisted Living (AAL) project is representative of this trend. It was established by the European community, and is comprised of more than 60 on-going projects (2008-2013). The objective of ALL is to enhance the quality of life of elderly people and strengthen the industrial base in Europe through the use of Information and Communication Technologies (ICT). One of AAL Project’s issues is to prevent and manage the chronic conditions of elderly people. To prevent and manage, they monitor patients or environments and alert to the caregivers, which is similar to the required functionalities of our proposed project. In the remaining parts, we describe several AAL projects and discuss the comparison to our project.

Neurodegenerative diseases like dementia are one of the most common diseases in elderly people. It is known that early detection of the symptoms of disease is critical. There are three AAL projects that deal with these diseases: ALADDIN, BEDMOND, and ROSETTA. The general goal of ALADDIN [2] is to make a piece of middleware with two main roles: distance monitoring of patient status and social networking between patients and caregivers. This social networking is to exchange experiences with persons in similar situations and to enhance their social relations. However, the ALADDIN system depends too much on constant observation by caregivers. For example, caregivers fill in the ALADDIN questionnaire for neuropsychological assessment from home. As an extension for social networking, they offer educational tools, but the services are limited to providing guidelines. In the BEDMOND system [3], data is gathered by a non-intrusive sensor network installed at the user’s home to track the behavior of elderly people. The data is then processed to recognize patterns of daily activities to be arranged through a rule-based engine. Medical experts use the processed data to determine whether or not the behavior changes mean the beginning of a cognitive decline. The BEDMOND system uses presence/motion sensors, reed switches, temperature sensors, power plug sensors, power consumption sensors, and smoke sensors in its surveillance network. Nonetheless, the roles of these sensors are limited. This issue also happens in the ROSETTA project [4]. The ROSETTA system detects the location of the subject and records how long the subject stays in place. The decision is made by comparing the duration of staying with a pre-set threshold, which is determined by the data observed during the first two weeks. However, the process is so simple and almost impossible to recognize complex activities. Hence, the quality and quantity of the information that can be acquired from these sensors is not enough. Meanwhile, the BEDMOND system offers GUI tools that will help the caregivers to be informed about the performance of the last activities and medical experts to address diagnosis. We can say that the system relies too much on human input as with the ALADDIN system.

As with existing on-going projects, we propose a monitoring system based on various sensors designed to detect risky situations for elderly patients. The main contributions of our proposed system will be to: (i) minimize human intervention; (ii) detect more complex activities and situations using various sensors and improved sensor fusion techniques; and (iii) design a novel classifier that determines a risky situation with the collected information. Note that most other systems
determine risk by comparing the numerical values for the current situation with normal or prefixed thresholds. Unlike them, we will propose a learning approach to classify risky and non-risky situations.

Related Literature

The development of a dementia monitoring system requires a variety of technologies that can be classified into three categories: activity recognition technologies, sensor fusion technologies, and risky situation technologies. For each category, this section presents related work including some of our previous studies.

Before the launch of Microsoft’s XBOX Kinect, only RGB cameras were used to deal with a vision-based surveillance system. Two common approaches, global representation and local representation, have been taken in order to represent videos captured by RGB cameras. Both approaches are not enough to represent activities in a video. Fortunately, the advent of the Kinect and research about estimating the 3D locations of human body joints enables computers to recognize humans in a video as a skeleton. Shotton et al. designed a classifier to estimate body parts invariant to pose, body shape, and clothing with a large and highly-varied training dataset. With the predicted body joints and skeletons, Sung et al. extracted three main features: body pose, hand position, and motion information [5]. They also designed the system to recognize twelve complex activities, such as brushing teeth, which are modeled with sub-activities using a hierarchical Markov model. Though the recognition accuracy is only 64.2% (when the subjects in the test video were not seen in the training video), their work is meaningful, since much more complex activities are dealt with.

As Rose and Wagner mentioned, a surveillance system using only cameras has limits to classify complex human activities [1]. Audio sensors (microphones) are also commonly used. Kim and Ko [6] detected and classified abnormal acoustic events occurring in an elevator environment. Wi-Fi, GPS, and Bluetooth sensors are also applied to a surveillance system for the purpose of patient authentication and location tracking. Han et al. [7] designed a system which estimates the indoor location of the device by capturing set of AP signals through a smartphone after collecting preliminary location information related to Wi-Fi fingerprints.

To generate a risky situation detector, we investigate psychiatric studies on behaviors in people with dementia [8]. Hope et al. studied the range and prevalence of behavior changes in a group of people with dementia. Among thirty types of selected behaviors, they reported the eight most common behavioral changes such as moving and mislaying objects and verbal aggression. Hope et al. also worked on whether or not robust behavioral syndromes can be identified from the widely heterogeneous behavioral changes which occur in dementia. Meanwhile, some researchers tried to design a personalized risky situation detector by recognizing abnormal activities by comparing the activities to his/her accumulated daily activities.

Healthcare Service Scenario

Target Domain

The target domain for the proposed applications is where a large number of elderly people require help from a relatively small number of caregivers. Accordingly, we will set an experimental environment inside a hospital or a nursing home. Figure 1 depicts such an environment.

Normal vision-based cameras, depth sensors, and microphones are installed in a room. With a normal camera, it is not easy to track a person because of the background subtraction problem. By installing additional depth sensors, the system can easily track a patient. It may also alleviate the issue of privacy invasion, since the depth sensors use the depth map images instead of RGB images.

In this setting, every elderly person carries a smartphone or a device that supports standard communication signals. Through the smartphone, the system can get to where the person is located and who the person is. Meanwhile, there is an agent for a certain place that monitors elderly people in the place and determines risky situations. We strongly assume that the experiment environment supports Wi-Fi and GPS networks. Wi-Fi fingerprints will be captured in the preliminary environment setting stage.

Scenario Description

With the proposed system, two scenarios are available: in a room and in more than two rooms. We assume that an agent is only concerned about one room or a certain place. Hence, the more than two rooms scenario deals with how the communication between agents improves the performance of the surveillance system.

When an elderly person is within the range of an agent, his smartphone tells both his exact location and personal information to the corresponding agent through near field communication, such as Bluetooth. Personal information is
composed of medical history, BMI, and an identification number. The identification number is used to identify and find a certain person among many elderly people. After receiving it, the agent conveys the location information to depth sensors and normal cameras located close to the person so that they can track the person. The depth sensors and cameras begin to track the person and send the observed data to the agent. The data is composed of RGB video as well as the depth map associated with the video. The agent finally analyzes the behaviors or activity patterns of the person through the data. When the preliminary conditions of dangerous situations occur to the elderly person, the agent notifies the caregiver. Microphones also send the observed data to the agent. Data observed from the microphones is especially helpful when a person is located in a blind spot that is hard to observe by cameras and depth sensors. The agent will recognize risky situations by detecting abnormal sounds such as yelling.

![Figure 1. Proposed generic environment](image)

The second scenario covers larger areas, which require at least two agents. Suppose an elderly person is in a room. When they leave the room, the agent responsible for the room may fail to track him. At first, the agent tries to locate the person roughly using the Wi-Fi fingerprint and GPS signals. It then advertises the patient’s geographical information to the local network around the rough location. By communicating via smartphones and various sensors, all the agents included in the network search for him. Once an agent finds the patient, that agent is in charge of that patient.

**Architecture of Multi-Sensor Surveillance System**

The development of such a dementia monitoring system requires a variety of enabling technologies. Figure 2 displays the architecture of a multi-sensor surveillance system.

#### Four Layers

In Figure 2, there are four layers in the architecture: sensor, client, repository, and infrastructure. First, the sensory layer is composed of various sensors. They are depth sensors, cameras, fall detectors, smartphones, and microphones. More sensors can be added to improve the performance of our surveillance system. In the client layer, we have a medical expert, caregiver, and elderly person. Knowledge from the medical experts helps to construct the database of predefined risky situations, while the system contacts caregivers when there are risky situations involving the elderly patients. Our main client in this project is elderly people. They will be asked either to wear a fall detector or to carry a smartphone. The repository layer has three databases necessary to operate our infrastructure. Finally, the infrastructure layer consists of three components: activity recognition, sensor fusion and risky situation detection. The three major components in the infrastructure layer are described in more detail below.

#### Activity Recognition
As a basic step for the system monitoring, activity recognition technologies should be studied in order to understand a patient’s complex behaviors. These technologies are generally referred to as vision-based human activity recognition in the area of artificial intelligence. The goal of these technologies is to generate a human activity recognizer. This recognition research includes gesture recognition and fall detection. In the proposed system, both conventional cameras and depth sensors, such as Microsoft’s XBOX Kinect, are used.

The issue to challenge is the level of complexity of activities. In order to cover various human activities while keeping higher recognition accuracy, the data should be expressed with more distinguishable features. The Kinect offers SDK software that estimates 3D positions of body joints [5]. In this component, the video image representation approach by processing the estimated 3D positions can be studied. In addition, the automatic method of modeling and learning activities will be studied. Moreover, we will research the basic units for human activities which may solve the issue of segmenting continual activity data.

![Figure 2. Architecture of multi-sensor surveillance system](image)

### Sensor Fusion

Sensor fusion is the combination of sensory data or data derived from sensory data from disparate sources so that the resulting information is in some sense better than when these sources were used individually [9]. By installing additional
sensors, the amount of data to be processed increases. This data can be meaningful information and may even create a synergy effect when it is processed properly. Therefore, sensor fusion is the most important technology in the proposed system in order to improve overall performance.

The challenging issue is how well the proposed system utilizes auxiliary data from various sensors. For example, when a person is located in the blind spot that is not easy to see by vision-based sensors, the microphones can supplement them by catching audio data. In addition, it is assumed that every subject in this experiment carries a device like a smartphone, which supports Wi-Fi, Bluetooth, GPS, and an accelerometer. Whenever a patient enters a certain range, his smartphone will send Bluetooth signals to an agent for patient authentication. When the subject moves further away from a particular camera location, wireless sensors or GPS can be used to trace the moving subject and/or to coordinate with many other camera locations in the distributed network. These technologies will naturally involve the issue of synthesizing distributed information gained from the network of multiple cameras and sensors deployed over various spots in a hospital or in a nursing home. To improve recognition accuracy, we can also add accelerometers that are widely used as well as vision sensors.

Meanwhile, abundant data captured from various sensors accompany noises which may slow down the computation speed and degrade the performance of the system. Hence, various data mining techniques should be adapted to filter abundant data captured from various sensors.

### Risky Situation Detection

To detect potentially dangerous situations, Artificial Intelligence technology should be used. After analyzing data from various sensors, the system determines whether or not the situation is potentially risky. The issues to challenge are how well the system constructs the database and the accuracy of its predictions.

In the proposed system, a database for risky situations will be constructed by predefining risky situations with the help of domain experts or knowledge as in [8]. A risky situation recognizer will be generated as well. Every second the recognizer calculates the likelihood of being risky. If the likelihood scores more than a threshold, it will recognize the situation as a risky situation and alarm the caregiver.

Since life patterns such as activity priority and conditions, differ from person to person, the recognizer can be personalized. The surveillance system extracts important features from a huge volume of video, audio, and other sensory data to learn behavioral rhythms and deviations for an individual, which is used for detecting unusual behavior.

### Research Plans and Progress

Our long-term research target is to develop a surveillance system for elderly people by fusing sensory information. As the project needs to deal with many difficult problems in computer vision and artificial intelligence, we propose performing the research in three phases.

In the first phase, the focus is to develop the basic activity recognition modules for the surveillance system. All the research will be performed at a campus laboratory (not in a hospital) with ordinary university life scenarios familiar to the students participating in this project. Conventional video cameras and depth sensors are used. During this phase, which enhance the background subtraction ability.

The second phase will start with a field study in a hospital or nursing home environment to elicit the practical requirements and scenarios for monitoring the behaviors of elderly people with dementia. These scenarios will assume to take place within an indoor environment, but nonetheless be diversified by including audio and wireless sensors in addition to video-based subject tracking and action recognition. Sensor fusion techniques for combining heterogeneous types of information from multiple sources, and machine learning or data mining techniques for analyzing individual behavioral patterns, will also be investigated.

In the third phase, the project will study the possibility to extend the system to support even more complex scenarios for outdoor environments. A distributed network of video cameras, audio sensors, Bluetooth, Wi-Fi, and GPS sensors will have to be coordinated, for which an architectural framework needs to be developed to fuse and synthesize an even wider range of heterogeneous information in the complex decision-making process.

We have performed research to tackle some of the challenges discussed above. In this section, we will summarize our achievements.

### Clustering Space-time Interest Points for Action Representation
As a basic module for video image recognition, we studied space-time interest points (STIP), well known for action video image representation. These interest points are locations where the image values have significant local variants in space and time in a video [10].

STIPs were widely used as features for action recognition. However, they are not sufficient to fully describe an action in a video. Each interest point exists without having any geometric relevance to other points. To enhance the performance of action recognition, we have extended our research to add geometric relationships to STIPs [11]. In particular, we propose a novel approach to represent the geometric relationships using hexahedrons where each hexahedron represents a cluster of local interest points in a video. We have also considered the relationship between each hexahedron in the video and use it as features for action recognition. Figure 3 shows the overview of the proposed approach.

The main contribution is that we successfully increase the accuracy of the action recognizer compared to results using local representation. In addition, the experiments show that the simple characteristic of the hexahedrons are powerful enough to recover global information and good at classifying the actions.

![Figure 3. Overview of the proposed video representation approach.](image)

### Essential Body Part Detection for Human Action Recognition

While studying various papers on human action recognition, we observed that a hand wave is not easily recognized when it is learned from various types of hand waving actions. For example, a man can wave his hand to a friend while sitting on a chair. On the other hand, a woman can wave her hand while standing on the floor. If these two different types of hand waving actions are learned separately, each action can be recognized with high accuracy. However, if these two actions are learned as one action, the performance of the action recognizer is not good.

According to this observation, we created a hypothesis that if an action recognizer concentrates on the movements of essential body parts, the performance of the action recognizer would be improved. To detect essential body parts, we encoded skeleton data, predicted by the Kinect SDK, into a symbolic sequence representation. This novel encoding algorithm allows us to detect essential body parts and the corresponding action pattern through the longest common subsequence algorithm [12]. Figure 4 displays the overview of the proposed approach.

The main contribution of this research is that the proposed approach is able to detect essential body parts automatically. Hence, by learning an action with only characteristic body parts, the required human effort to achieve the goal of human action recognition will be minimized.

### A Wrist-mounted Fall Detector with Statistical Classifier for Elderly Care

One of the most critical accidents for elderly people is a fall. In [13], we proposed a statistical classifier which is based on a wrist-mounted fall detector as in Figure 5. With the development of wireless networks and a low-power microcontroller unit, a wearable fall detector has been released. It contains inertia sensors and can automatically send an alarm to caregivers. Although the wrist-mounted fall detector is acceptable to elderly people, the false positive rates of fall detection were too high due to the complexity of arm and hand movements. We were also motivated by the fact that the fall detection algorithm should be simple and not computationally expensive so that it could be applied to real-time operations.

In order to acquire higher accuracy and reduce computational costs, we proposed the efficient method to reduce a lot of computations by performing feature reduction techniques on large datasets and high dimensions of feature vectors. In the experiments, we achieved a recognition rate higher than 90%.
Context Awareness of Social Groups by Topic Mining Visiting Logs of Mobile Users in Two Dimensions

As one possible filtering method, we studied a topic mining technique, which was able to find hidden meanings, i.e. topics, through unsupervised learning. Most opinions for a specific product on the Internet come from the group of people familiar with Internet technology. In [14], our motivation was learning how to derive group preferences automatically. We assumed that each individual carries a smartphone. The visit log, i.e. location information from GPS signals and time intervals that is captured from a smartphone, is sent to our server. The server first decides whether or not to filter the visit logs for relevant information. It then constructs topic cubes, as in Figure 6, to offer useful group preferences in multi-dimensional ways.

We experimented with five smartphones and eight specific spots inside a campus. With this simple experiment, we could obtain successful group preferences. This research can answer the questions like “In which month did patients go to the hospital most frequently last year?”
To recognize human activities, we used sensory information captured by a mobile environment, i.e. smartphones, as well. Nowadays, most individual people carry a smartphone on themselves, and a smartphone is a collection of sensors. We were motivated by the observation that location information - GPS signals and Wi-Fi signals – has limits to understand human activities in detail. Specifically, we studied a novel method to recognize hand movements by analyzing 3D accelerometer data captured by a smartphone, which eventually helps to understand one’s activities [15].

Eight directions were set to states as in Figure 7. We also defined nine character symbols with these states as in Table 1. For example, the word toilet is abbreviated as the letter T. Then, by applying the Parametric Hidden Markov Model, we defined each gesture model efficiently. As smartphones can draw spots and vectors, we used starting points, direction, and significance in a vector to define the models.

Although all character symbols should be first defined in the form of parametric values, we contributed to propose a simple and effective method to recognize human activities using accelerometer data by analyzing predefined character symbols. We experimented with two letters, A and T, and achieved acceptable results.

**Figure 6.** Our inference system model using topic cube

**Figure 7.** Eight directions
Conclusion

An aging population achieved by great advancements in medical science has brought about many important problems to the human race. Dementia is one such problem, and surveillance technologies have a large potential to assist those elderly people with dementia. Providing 24 hour supervision by human caregivers requires huge labor cost, if ever arranged, and an automated video monitoring and surveillance system should be considered as a viable alternative. In this project, we are developing surveillance technologies to recognize risky behavioral patterns and situations of elderly people with dementia. When successful, the results of this project will make contributions to the field of ubiquitous healthcare.

As further research, we will study sensor fusion to enhance a vision-based surveillance system. With additional information such as Wi-Fi fingerprints, the performance of our system will improve. In addition, we consider research using Microsoft’s Xbox Kinect 2, which even may support lip reading technology.

Regarding activity recognition research, we need to collect a more reliable dataset. Recently, the Graphics Lab at Carnegie Mellon University has published a Motion Capture Database that includes a sixteen daily activity dataset captured by a Kinect device. We are going to apply our novel algorithms into this dataset to improve the activity recognition system.

References

KyoJoong Oh is currently a PhD student in the Dept. of Computer Science at KAIST, Daejeon, Korea. He received a BS in Computer Science from KAIST in 2011, and entered the Master’s program in Computer Science, then changed his status to the combined MS/PhD Renaissance Program in Computer Science at KAIST. His current research interests include data mining, machine learning, context awareness, and big data.

Chankyu Park received an MS in Electronics Engineering from Kyungpook National University in 1997, an MS in Software Engineering from Carnegie Mellon University in 2005, and is currently a PhD candidate in Electric & Electronics engineering at KAIST. He joined the Electronics and Telecommunications Research Institute (ETRI) in 1997 and has been a Senior Member of the Engineering Staff in the Intelligent Robot Research Division since 2004. His research interests are in pattern recognition, machine learning, computer vision, and healthcare robotics.

Young-Seob Jeong is currently a PhD student in the Dept. of Computer Science at KAIST, Daejeon, Korea. In 2010, he received a BS in Computer Science from Hanyang University, Ansan, Korea. Her current research interests include data mining, topic mining, and pattern recognition.

Sou-Young Jin is currently a PhD student in the Dept. of Computer Science at KAIST, Daejeon, Korea. She received her MS from KAIST in 2012 and her BS from Dongguk University in 2010. Her research interests include vision-based activity recognition, data mining, and object recognition.

Ho-Jin Choi is currently an associate professor in the Dept. of Computer Science at KAIST. In 1982 he received a BS in Computer Engineering from Seoul National University, Korea. In 1985 he got an MSc in Computing Software and Systems Design from Newcastle University, UK. And in 1995, he got a PhD in Artificial Intelligence from Imperial College, London, UK. From 1982 to 1989, he worked for DACOM, Korea, and between 1995 and 1996 worked as a post-doctoral researcher at Imperial College. From 1997 to 2002, he served as a faculty member at Korea Aerospace University, Korea. He moved to Information and Communications University (ICU), Korea, from 2002 to 2009. And since 2009 he has been with the Dept. of Computer Science at KAIST. Between 2002 and 2003 he visited Carnegie Mellon University, Pittsburgh, USA, and served as an adjunct professor the Master of Software Engineering (MSE) program. Between 2006 and 2008, he served as the Director of the Institute for IT Gifted Youth at ICU. Since 2010, he has been participating in the Systems Biomedical Informatics National Core Research Center at the Medical School of Seoul National University. Currently, he serves as a member of the board of directors for the Software Engineering Society of Korea, for the Computational Intelligence Society of Korea, and for Korean Society of Medical Informatics. His current research interests include artificial intelligence, data mining, software engineering, and biomedical informatics.