A Hierarchical Description-based Video Monitoring System for Elderly

Mochamad Irwan Nari, Agung Wahyu Setiawan and Widyawardana Adiprawita
School of Electrical Engineering and Informatics
Institut Teknologi Bandung
Bandung, Indonesia
m.irwan.nari@gmail.com, awsetiawan@stei.itb.ac.id, wadiprawita@stei.itb.ac.id

***Abstract*—The increase in the number of elderly motivates academic researchers to develop technologies that can ensure self-sufficiency in their lives. In this research, prototype of an inexpensive video monitoring system for the elderly using a single RGB camera proposed. In the process is divided into two, namely vision and event recognition module. For event recognition, we use a hierarchical description-based approach with three attributes, namely posture (*e.g.*, stand, sit and lie), location (*e.g.*, walking zone, relaxing zone and toilet zone) and duration (*e.g.*, short and long). Output this system is description activity recognized in the text. The experiment result shows our system can provide the effectiveness of the context description.**

Index Terms*—Video monitoring system, hierarchical description, video processing, elderly.

I. Introduction

The elderly population in the world has increased significantly wherein 2015 reached 15% of total world population and estimated to reach 20% in 2020 [1]. According to WHO and Law No. 13 of 1998 mentioned that the elderly are the age group over 60 years. While according to [2], the elderly can be classified into three, namely young old (age between 65-74 years), middle old (aged between 75-84 years) and old (aged over 85 years).

Indonesia is one country in Asia that has the number of elderly is large enough. Based on data from Badan Pusat Statistik (BPS-Statistics Indonesia) survey in 2014, the population aged 60 years and older reached 20.24 million [3]. From these data, the status of the elderly who live alone is approximately 9.66%.

With the increasing number of elderly, we need a technology that can support the development of independence in their lives [4]. Therefore, the monitoring system for everyday activities is needed to monitor the conditions in real-time. Due to human activities are monitored by the system, if there's activity is not common, this can be known early.

In this context, we propose a video monitoring system with a video camera on to recognize activities undertaken by the elderly. This research was inspired by the work [5] where they introduce automatic video interpretation. One advantage of using a camera is the detection and location of people in the same time and easy to install.

Research using the camera for recognition of daily activities have been carried out. Duong *et al.* [6] use four cameras to capture the scene from different angles. They extract the location of the subject to activities recognized. Dubois *et al.* [7] using the RGB-D camera to capture objects. For the activity classification using Hidden Markov Model (HMM). They did eight activities like sitting, walking, climbing, squatting, lying on the couch, falling, bend and lie down.

Some of the methods used for people monitoring with the camera like Cao *et al.* [8] do the extraction of the human body perform context (*e.g.*, sitting, standing, walking and lying) and environment context in the activities model. They use a description-based approach and rule-based activity reasoning to generate context description. Other methods are the hierarchical description-based. Zuoba *et al*. [9] do the extraction of physical attributes of the object (*e.g.*, person, equipment) for the identification of activities of daily living of older people in an apartment. Joumier *et al.* [10] conducted extraction of attributes (*e.g.*, duration, walking speed) to distinguish between healthy and Alzheimer's group. Carlos *et al.* [11] also perform extraction features such person and event zone for recognition of 29 Alzheimer's participants.

The paper is organized as follows: Section I describes the introduction and some algorithms of the video monitoring system in the previous study. Section II describes the proposed method of video monitoring system. Section III discuss the experiments and results. Finally, in Section IV describes the conclusions and future work of the study.

II. Video Monitoring System

Video monitoring systems are generally divided into three parts that are input, process and output. On the input use a camera to capture the object. In the general process is divided into two, namely the vision module and recognition module. The process of the vision module is the person and the posture detected on the scene. In recognition module event involving models and semantics zone. In this study, the event follows the model proposed by [11] that a declarative and intuitive ontology-based language. On the output consists of a collection of events that have been

978-1-5386-0549-3/17/$31.00 ©2017 IEEE

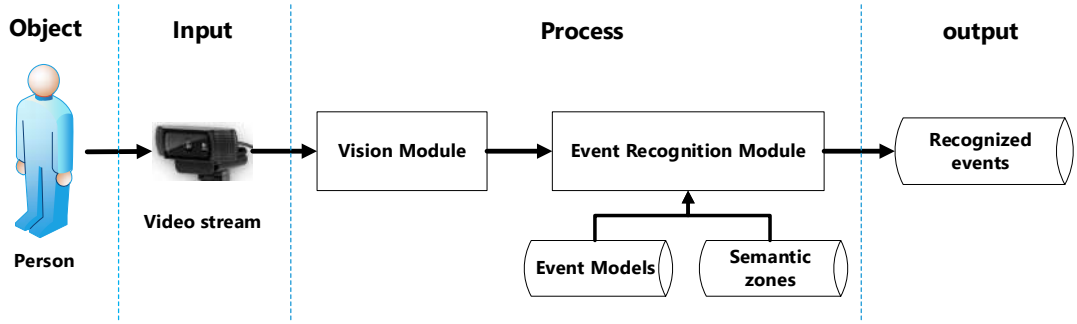


Fig. 1. Architecture of the Video Monitoring System

recognized in the form of context description. In Fig. 1 is shown the architecture of the system of a monitoring system with an RGB camera as input.

A. Vision Module

Vision module is a module that allows trying different algorithms of each stage in computer vision are like object detection, object segmentation and object classification. Vision module consists of the detection process and the posture of people.

Object detection process is carried out by extracting the foreground contained object with the background using frame differencing method proposed by [12]. Before determining the region of interest (ROI), some process is performed that is converted from RGB image to grayscale, reducing noise using morphological operations, changes into a binary image by threshold and contours detection. To mark the detected object then do make bounding box. To distinguish the object being detected by others then we added shape features. In Fig. 2 are shown the results of object detection using vision algorithm. Rectangular shape with a green line represents the object has been detected.

Fig. 2. Scene image of a person detected

Once the object is detected, then the posture detection. The first step is to track the center of mass of the person. Because we think people carry mobile information point (*e.g.*, width, height), then the ratio of the value can distinguish several postures of people such as stand, sit and lie.

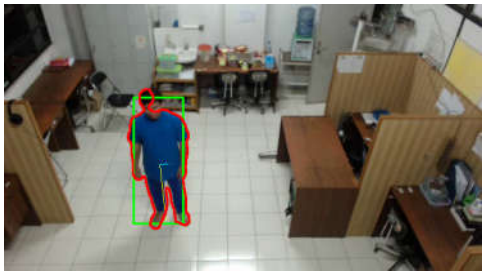
In addition to that feature, we added features of the orientation of the object using the Principal Components Analysis (PCA) [13]. To get the angle value object orientation, calculation as in [16]. In Fig. 3 are shown the results orientation of an object using PCA. The yellow line is the largest principal component first. While the blue line is the second largest principal component.

Fig. 3. Scene image of result orientation of an object

In Fig. 4 are shown the results of the posture detection feature using a combination of width, height and angle of orientation. The red line surrounding the object representing the results of contour detection.

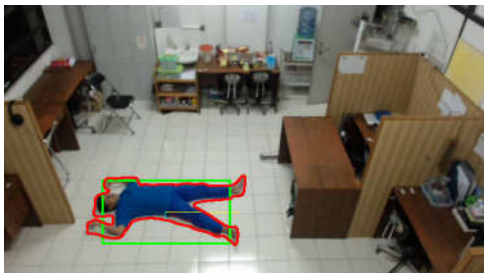
B. Event Recognition Module

In this study, the event recognition similar to [11]. This module provides a framework model of activities and semantic zones. Models made by considering the activities prior knowledge of the scene is done and the nature of the moving object is detected. This activity model is declarative and intuitive ontology based on language which uses the generic term to allow users to easily add and change the model. A Priori knowledge is the beginning of knowledge possessed by the system and provides a floor plan of scenes in spatial zones which has information about the activities of semantics. Semantics is a sign or symbol. Semantics zone is the zone that has been marked with certain information such as the toilet zone, relaxing zone and walking zone.



(a)

(b)



(c)

Fig. 4. Scene image of a posture detected: (a) stand, (b) sit and (c) lie

The model consists of the activities of the six components [11]. Model activities proposed in the study consisted of four, namely physical objects, components, constraints and alerts. Physical objects are associated objects in the scene covers the activities undertaken and the context of the zone (*e.g.*, walk zone). The component is part of the activities related to the scenes activities include primitive state and the composite event. Where primitive state indicates the nature of the object that moment in the scene of activities such as the sitting posture and the composite event shows the combined temporally related activities such as the transition of people stand to sit before people get relaxed zone. Constraints are certain conditions of physical objects or components in the scenes activities such as the activities associated temporally. Alert is a warning to give specific information about the activities undertaken, such as a person being in the walking zone. In Fig. 5 are shown an example of a primitive state model. This model represents the posture detected. This attribute value is obtained from the combination of width, height and angle of object orientation.

PrimitiveState (Person_standing,
PhysicalObjects ((p : person))
Constraints ((P -> posture = stand)))

Fig. 5. Description of Primitive State "Person is standing"

In Fig. 6 are shown an example of a composite event model. This model checking posture detected attribute and spatial zones that have semantic information. Where the posture value obtained from the combination of width, height and angular orientation of an object and spatial zones obtained from the initialization by the user in advance of the zone.

CompositeEvent (Person_sitting_relaxing zone,
PhysicalObjects ((p : person), (z : zone)
)
Components ((c1 : CompositeEvent
P_inrelaxingzone (p,z))
(c2 : PrimitiveState
P_sitting(p))
)
Constraints ((c2 and c1))
Alert (Atext ("Person is sitting in a relaxing zone"))

Fig. 6. Description of Composite Event "Person is sitting in a relaxing zone"

In addition to the features posture and zones, in this study added extra feature is the duration. This duration is calculated when activities are carried out continuously detected [14]. The duration can also be defined as the accumulated time of occurrence of the same activities. The duration is divided into two, namely short and long duration. We assume that a short duration is defined as the time the activity occurred during less than 1 minute. While the long duration is defined when activity occurs continuously more than 1 minute.

III. Experiments and Results

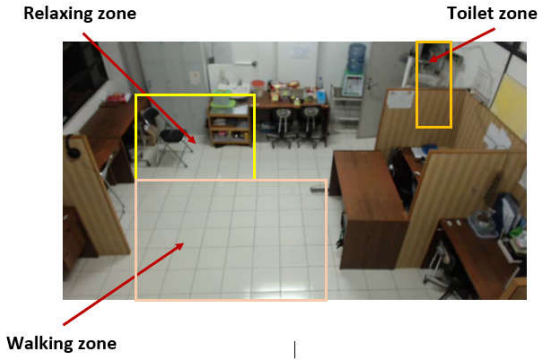
In this section, we will explain in detail the design of experiments and experiments results.

A. Experiments Design

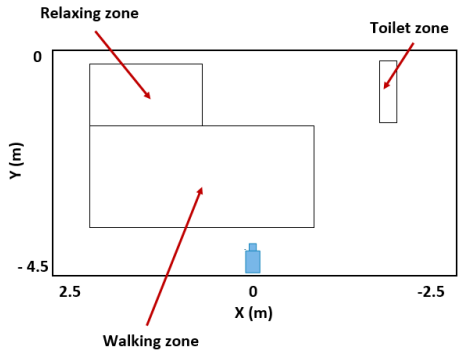
The initial stage of the implementation of video monitoring system is to prepare the test environment. The location used is room 22 m2. In Fig. 7 (a) are shown the capture of the rooms were used. We divide the space into three zones, namely the walking zone, relaxing zone and toilet zone. In Fig. 7 (b) are shown the layout used this experiment. RGB camera in installed on the wall with a distance of 3.5 meters from the floor. The camera is used to capture the human movement of the subject and location. This camera has a specification of 30 fps, field of view 78 and resolution 1920x1080 [15]. In the implementation, the resolution of the camera is lowered to 640x360 and fps to 6 fps.

Notebook used to collect data and run activity recognition in real time. Specifications of the notebook is the Intel® Core ™ i3 and 10 GB RAM memory. For video processing, we use Visual Studio 2015 (C++ with OpenCV library 3.1). The program created for processing the data from the camera and activity recognition. Because the signal processing is done in discrete we conducted the election activity was detected most frequently in a period of 60 frames. This may imply that each of these periods obtained the data. Data is collected every 60 frames with consideration of the action transition time has been covered. For the duration of the information obtained from the accumulation of activities that occur continuously. For example of people doing activities in the toilet for a duration of more than 20 times the data. This may imply that happened is a long duration. The results of the program execution are then displayed on the monitor screen.

Before the experiments were conducted, the human subject is given information about the semantic zones and posture that are recognized by the system. Then human subjects are given 10 minutes to perform daily activities randomly. In this case we assume that the human subject to wear clothing with colours that contrast with the surrounding environment.



(a)



(b)

Fig. 7. Views of experiment environment: (a) picture of the real environment (b) apartment map and position camera

After the experiments were conducted, the results obtained dataset recorded video with a duration of 10 minutes. Next is the system testing performed by looking at the output in the form of a description of the context of the system according to the scene.

B. *Experiments Results*

The experiments were performed with one male subject. This is because it is still in the prototype stage. In Fig. 8 are shown five input image and its context in the form of text description. The images are the result of an experiment with a short duration of activity. In the future, this description could provide insight into the human being syntax definition is accurate.

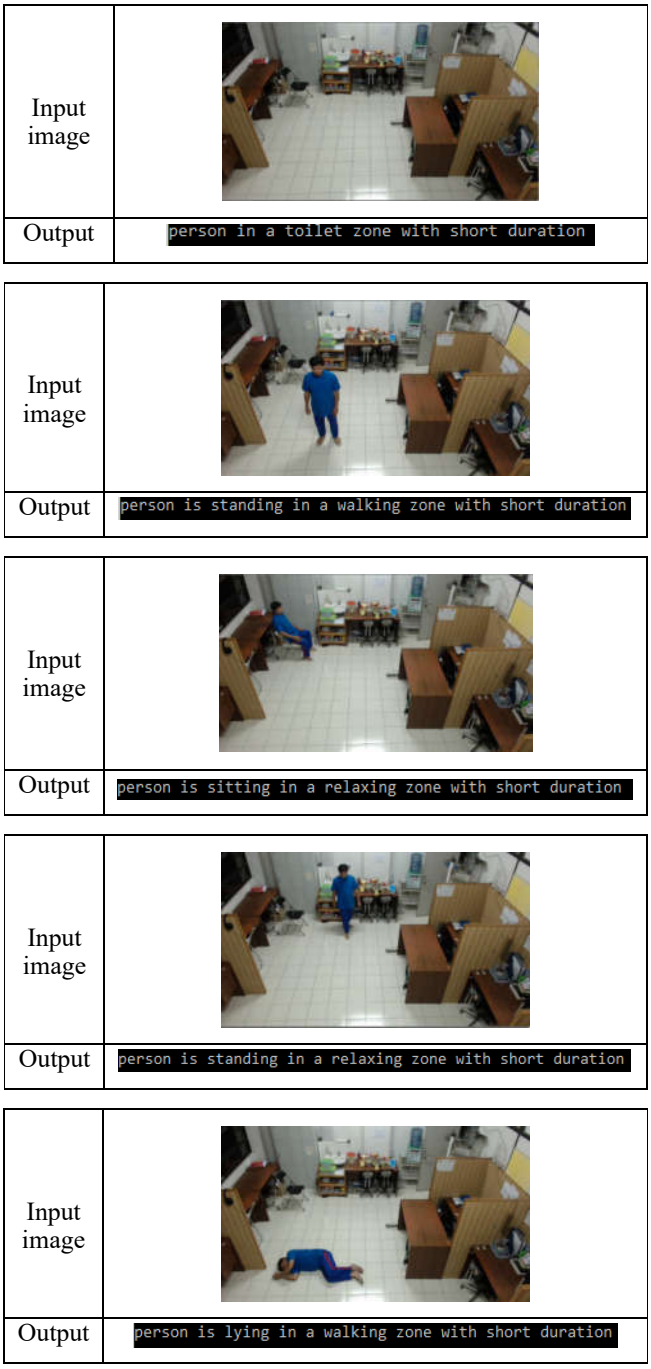


Fig. 8. Input images and output in text

Table 1 shows the results of activity recognition testing. The data used is a video recording activity of four people. Based on the table above, the three data are the recording of the activity of three people whose poses have been identified and included in the dataset. While one data is the recording of the activity of one person who has never recognized posture. The total duration of the activity test video is 41 minutes 3 seconds. Based on the table below it can be seen that the average accuracy of the four experimental videos is 92.53%. The calculation of accuracy is based on the results of the truth of the manual observation of each data.

TABLE I. TEST RESULTS FOR ACTIVITY RECOGNITION

Video Name	Number of observation data	Correct	Incorrect	Accuracy (%)
act_irw_6.avi	189	172	17	91.01
act_adt_6.avi	203	186	17	91.63
act_sen_6.avi	218	198	20	90.83
act_era_6.avi	208	201	7	96.63
Average Accuracy (%)				**92.53**

IV. CONCLUSION

In this paper has described the monitoring system for elderly to use a video camera. In experiments used single RGB camera which has the advantages of the larger field of view. For the detection of use frame differencing method. While posture detection using the feature such as width, height and orientation of the object. We propose a prototype for event recognition with a hierarchical description-based. Two types of physical object that have been in posture and contextual zones. Posture consists of the stand, sit and lie. The contextual zone consists of a walking zone, relaxing zone and toilet zone. To provide information activities becomes more complete, we add the duration time of the event. The output of the system is the activity that has been recognized in the form of text. There are ten scenarios that activity has been carried out. Based on the experiment, the average accuracy result for the activity recognition test of four videos was 92.53%. Hence, our future work is focused on the addition of a physical object such as a contextual object which can provide additional context a complete description.

REFERENCES

[1] Tonchev. Krasimir, Strahil Sokolov, Yulian Velchev, Georgy Balabanov and Vladimir Poulkov, "Recognition of Human daily activities," IEEE International Conference on Communication Workshop (ICCW), 2015.

[2] Zhang. Shuai, Sally McClean, Bryan Scotney, Priyanka Chaurasia and Chris Nugent, "Using duration to learn activities of daily living in a smart home," 4th International Conference on Pervasive Computing Technologies for Healthcare, 2010.

[3] Statistik Penduduk Lanjut Usia 2014. Available: http://www.bappenas.go.id

[4] Shimokawara. Eri, Tetsuya Kaneko, Toru Yamaguchi, Makoto Mizukawa and NobutoMatsuhira, "Estimation of Basic Activities of Daily Living using ZigBee 3D Accelerometer Sensor Network," International Conference on Biometrics and Kansei Engineering, 2013.

[5] V. Vu, F. Bremond and M. Thonnat, "Automatic video interpretation: A novel algorithm based for temporal scenario recognition," The Eighteenth International Joint Conference on Artificial Intelligence (IJCAI), 2003.

[6] Duong. T., H. Bui, D. Phung and S. Venkatesh, "Activity recognition and abnormality detection with the switching hidden semi-markov model," IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 2005.

[7] Dubois, Amandine and Francois Charpillet, "Human Activities Recognition with RGB-Depth Camera using HMM," 35th Annual International Conference of the IEEE EMBS, 2013.

[8] Cao. Y., L. Tao and G. Xu, "An Event-Driven Context Model in Elderly Health Monitoring," Proceedings of Symposia and Workshop on Ubiquitous, 2009.

[9] Zouba. N., F. Bremond and M. Thonnat, "An Activity Monitoring System for real Elderly at Home: Validation Study," 7th IEEE International Conference on Advanced Video and Signal-Based Surveillance, 2010.

[10] Joumier. J., R. Romdhane, F. Bremond, M. Thonnat, E. Mulin, P.H. Robert, A. Derreumeaux, J. Piano and L. Lee, "Video Activity Recognition Framework for Assessing Motor Behavioral Disorders in Alzheimer Disease Patients," International Workshop on Behavioral Analysis, 2011.

[11] Carlos F. C-Junior, Vasanth Bathrinarayanan, Baptiste Fosty, Alexandra Konig, Rim Romdhane, M. Thonnat and F. Bremond, "Evaluation of a Monitoring System for Event Recognition of Older People," 10th IEEE International Conference on Advanced Video and Signal Based Surveillance, 2013.

[12] Stalin. D. Alex and Amitabh Wahi, "BSFD: Background Subtraction Frame Difference Algorithm for Moving Object Detection and Extraction," Journal of Theoretical and Applied Information Technology, 2014.

[13] M.Z. Uddin and M.A. Yousuf, "A New Method for Human Posture Recognition Using Principal Component Analysis and Artificial Neural Network," Journal of Scientific Research, 2015.

[14] Zhu. C., Wiuhua Sheng and Meiqin Liu, "Wearable Sensor-Based Behavioral Anomaly Detection in Smart Assisted Living Systems," IEEE Transactions on Automation Science and Engineering, 2015.

[15] Webcamera C920. Available: www.logitech.com/en-us/product/hd-pro-webcam-c920

[16] Mudrova, M. and Prochazka A., "Principal Component Analysis in Image Processing," Proceedings of MATLAB Technical Computing Conference, 2005.