ESTIMATING HUMAN LOCATION AND MOTION USING CAMERAS IN SMART HOME

*Quoc Cuong Nguyen, Dongil Shin, Dongkyoo Shin**

Department of Computer Engineering, Sejong University
98 Kunja-Dong, Kwangjin-Gu, Seoul 143-747, Korea
cuongnqc@gce.sejong.ac.kr, {dsin, shindk}@sejong.ac.kr

ABSTRACT

The ubiquitous smart home is the home of future that takes advantage of context information from user and home environment and provides automatic home services for the user. User's location and motion are the most important contexts in the ubiquitous smart home. This paper presents a method positioning user's location using four cameras and some home context parameter together with user preferences provided. Some geometry math problems would be raised to figure out approximately which area is monitored by cameras. The moving object is detected within the image frames and then simulated in a 2D window to present visually where user is located and show his moving. The moving ways are statistically recorded and used for predicting user's future moving.

Keywords: smart home, human location, home context, camera monitoring

1. INTRODUCTION

The ubiquitous computing research was found by Mark Weiser [15]. Since then, many efforts have been made on this field. Recently, researchers focus on the smart homes and these assure will bring the intelligence service to wide range of function from energy management, access monitoring, alarms, response systems, medical emergency and so on.

The human location and tracking are those of the important issues in the smart home. Some researches use the sensor system [5, 6] in the analysis process and for tracking motions. But this paper presents the human location estimation and moving direction prediction in the smart home using multi-camera system. The home context is being estimated approximately using the parameters supplied of the room and the view angle of each camera. The application of this research could be the energy management, the furniture adjusting or some services adaptation in place where installing devices does not require much of work.

We designed and implemented the estimation process of effective area where can be monitor from system camera view angle, this process requires to be supplied the two sizes of room and the horizontal view angle of the cameras. We use a classification algorithm on estimating the human location and predicting the human's moving direction base on the frames collected from the multi-camera system and correspondence position over the view angle of the camera.

The paper consists of five sections. After the introduction in this first section, section 2 gives some brief view in some related research works on data mining algorithm to smart home research. Section 3 addresses the detail data mining algorithm to estimating location and predicting moving direction. Section 4 presents the implementation and experimental result, the conclusion finally in Section 5.

2. BACKGROUND

Human location and human tracking were studied by various methods; Gao [4] proposed the curve segmentation which was told that curve is the basic sharp for objects and human. He tracked down the objects through raster scanning of frames, detected the potential curves and then tracked the motion invoked from edge tracker. Sajal Das et al [5, 6] in their researches used the sensor systems and track the moving by those sensors, they proposed the predicting in their MavHome by their stochastic model, giving some abilities of predicting the habit of human in moving between different rooms in house.

Further complex sensor using in smart home was introduced by Yamazaki [8], that research installed some floor sensor to detect the furniture position, the infrared sensor to detect the movement, the microphone and cameras installed in each room to perceive the audiovisual information.

Eric Manley and Jitender Deogun [9] used the ID3 and and Neural Network in learning location through beacon signals getting from map of sensor in the building. The research opened the way besides of infrared sensor, sound and camera to detect the objects, and to be expected of detecting the sitting, standing, etc by learning from the intensity of signal.

Other researches [10, 11] used multi-cameras combining with floor sensors and color, texture to track the motion of moving human; these systems have ability to track multiple peoples.

Abhishek Roy, Sajal Das and Kalyan Basu [14] developed predictive framework in which location and series of

*Corresponding author
This study was supported by the Seoul R&BD Program (BU070131).

movements, waiting time, has been mined in order to learn the rule and predict in next action. In this research energy management also introduced, an example of usage of heat extraction and air conditional with properties of location, temperature and time had been considered as a factor in which predictive framework would controlled.

In contrast of using a recordable sensor, other research [12] used small and simple state-change reflected sensors which installed on everyday objects in doors, refrigerators, stove and so on, the collection of activities had been recorded, each sensor with attributes of on/off then being the input to run yielding the clustering algorithm. Some rich density areas should be considered in the result of process and users customized the application by modifying some parameters related to learning rate, size of neighborhood.

Our system uses the multi-camera system in learning the location of human and predicting the moving intention. The result may attend in some public areas, and installing the multi-camera does not require many efforts or modifying too much on the current facility.

3. ESTIMATION OF HUMAN LOCATION AND MOTION IN HOME CONTEXT

3.1 Camera calibration to generate the home context

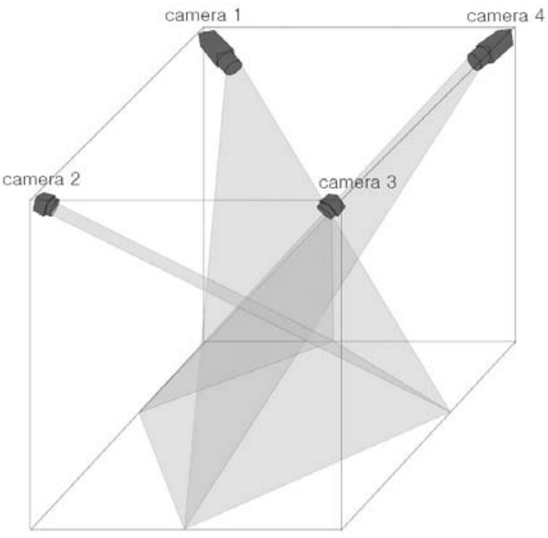

Fig. 1: Structure of Camera systems

Figure 1 shows arrangement of the camera system, there are four network cameras installed in the room, these cameras were placed in the four corners at the top of the room which is in this research the home context was fixed to rectangle shape. With some user's preferences this process would estimate approximately the position in which view angle of camera point to going to predict position and moving action further.

Figure 2 below presents the pictures captured from network camera, the total monitored area includes the four corners and the area at the center of room. The process of estimating home context does not require any participation in taking picture in empty room or picture with furniture, the estimation process is divided into sub problems at the special case and the real result would be calculated base on the ratio to those special cases.

a) first corner image

b) second corner image

c) third corner image

d) fourth corner image

Fig. 2: sampled images from the network camera system

The camera system provided the images from four corners of the room, there is a need to specify the range of covered view in order to know approximate position in the room that the camera point to.

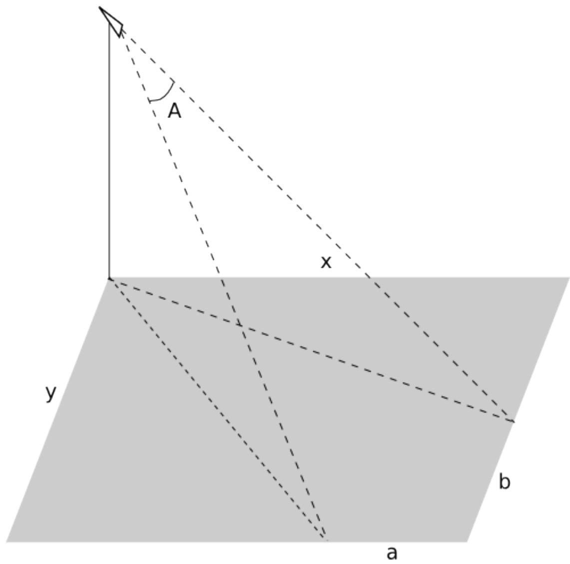

Fig. 3: Process of estimation the scale factor

Figure 3 shows home context of the experiment, x and y represent the size of the room. There are four cameras installed in the four corners of the room, this figure represents only one of them, all of the others are similar. The camera would point to the opposite corners which a and b are two sides which actually be captured by the camera. The A angle is the horizontal view angle of that camera. This parameter was provided by the operator, also with the x and y size of room. Note that there are three types of camera angle: horizontal view angle, vertical view angle and diagonal view angle. In this research we use the horizontal view angle.

The following problems are introduced in order to solve the problems of identifying the effective area that can be monitor be the camera.

Problem 1: The camera lens is by circle shape but actually the view area is the rectangle and those factors used to measure the camera lens are view angle by horizontal, vertical and diagonally as illustrating in Figure 4 below.

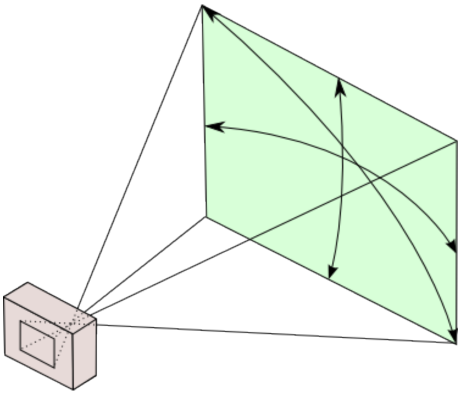

Fig. 4: camera view angle

The camera is installed in the corner of the room, if the moving is fixing to only up and down, the result getting from these are three lines parallel as showing in Figure 5 below.

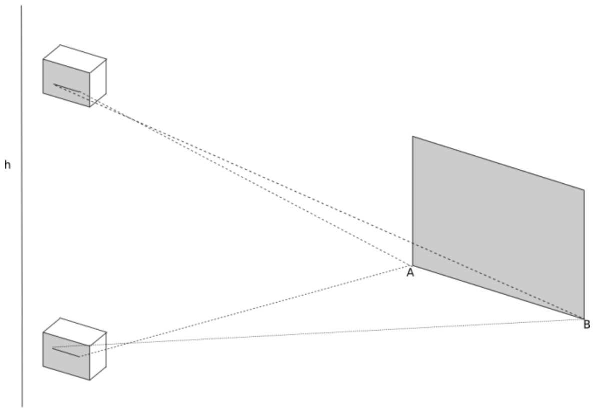

Fig. 5: Camera context

The h parameter presents for the height of room, A and B are two virtual lines produced by the horizontal view angle of the camera. If the camera is fixed to vertical and moving from up to down, the position of A and B virtual lines are not changed, that means the horizontal view angle is fixed by moving camera to the vertical lines.

So from Figure 3 where camera is installed in the roof of room, known the size of room, the horizontal view angle. The computation on effected area a and b are similar by acting as camera is install in the floor. That means the 3D mathematic problems become the 2D.

Problem 2: 2D context of camera view angle and effective area should be calculated.

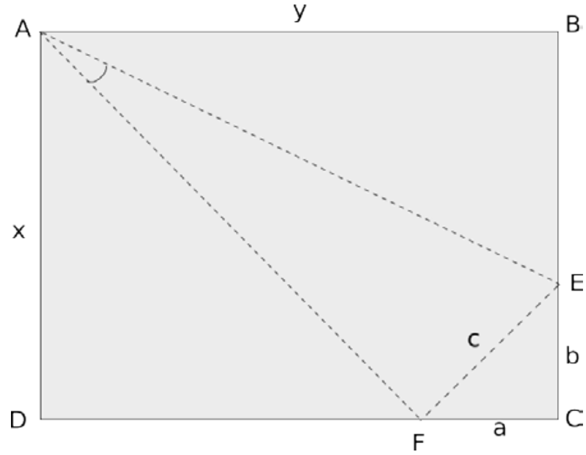

Fig 6. 2D context of camera view angle and effective area.

From the result of problem 1, where 3D problem of horizontal view angle is similar to 2D as shows Figure 6 above. In this case the angle EAF is supplied by user preferences and to be assume as 53^o, the problem 2 focuses on estimating the a and b in the 2D context.

Sub problem 2-1: computing the length of a and b when the horizontal view angle divided the corner into two equaled parts, or a and b in the Fig 6 are equal.

Applying the law of cosines [16] to the triangle AEF, here listing the result:

$$\cos(EAF)*2*AE*AF = AE^2 + AF^2 - c^2$$

The angle EAF originally is horizontal view angle as proving in problem 1. Replace the AE and AF to the formula, the c can be compute, from c the a and b can be achieved.

Sub problem 2-2: there is a case when one size of room is many times larger than the other. The view angle focuses right on the corner of opposite, estimate the length of the effected size.

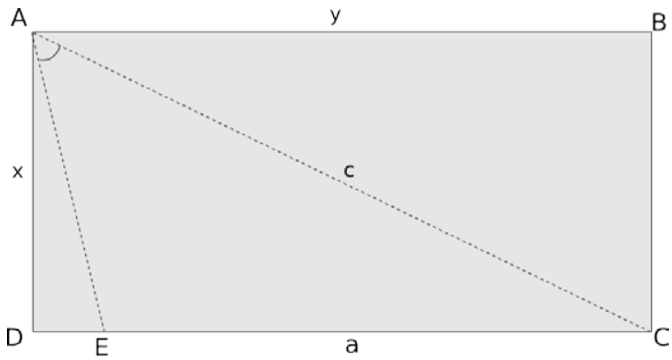

Fig. 7: Extra case of horizontal view angle of camera

This case happens in the room which large different in the size, exactly that the horizontal view angle cannot cover the whole size of longer size as show in Figure 7, the CAF triangle then called the obtuse triangle. There is a need to calculate the length a. We get the following result:

$$\frac{(a-c)}{(a+c)}=\frac{(tangent(\frac{1}{2}(AEC-EAC)))}{(tangent(\frac{1}{2}(AEC+EAC)))}$$

All of the value has known, it is easily to compute the a, which is the needed value.

Problem 3: estimate the real horizontal view angle from real images.

This problem requires the user preference take part in the real image. Figure 2 shows the real image be captured from network cameras. These steps below illustrate the process of specifying real effective area:

- The operator is required to specify three points correspond to the left horizontal point, the right horizontal point and the central corner of the image.
- From those three points, estimate the ratio between left horizontal line and right horizontal line from the central corner.
- Compute the left most value and right most value, from ratio in second step to estimate approximately the real effective area.
- Repeat these steps above to the rest of the images.

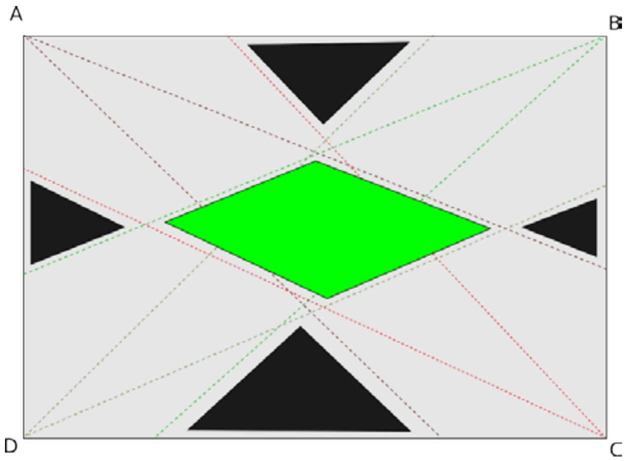

Fig. 8: Effective area

This problem solves the area actually be covered by the camera and can be drawled in Figure 8 above. The lines from each corner shows the area that cover by that camera, there are four cover areas and between them existed the *unmonitored area* as color by black. The green area shows the area actually can be seen by more than one camera, this is *general area*.

3.2 Location estimation

Location estimation is the process of identifying the position in the room in which person or things are standing. From previous part we extract approximately the area which cover by the view angle, the following classification algorithm identify the problems of specifying human position:

- Drawn the 2D diagram in a window with appropriate scale to the real size entered by operator.
- Label each camera with the appropriate positions in the diagram.
- Record the empty human image in each network camera.
- Operator specifies the three points in each camera correspond to the left most view angle line, right most view angle line of the camera and the corner of the room.
- Specify the border lines in each image correspond to operator preferences in showing the corner point and lines from that point to the boundary of images.
- The human object is extract by minus operation between recorded images at the real-time to the original. The position of human is extract by the lowest position of object; correspond to the foot in of that object.
- The distance from detected object in real image to the border lines in step 4 above correspond the real scale distance.
- If object appears in the left most means the real position correspond to unmonitored area at the left of that camera.
- If object appears in the right most means the real position point to the unmonitored area at the right of that camera.
- If the object were discovered at the general area, if the similar area also had seen on the other camera, then there are only one object, the real position then should be considered as average computation.

The captured objects depend on how height the camera had been positioned, if the camera is installed so high then the general area is limited.

3.3 Estimation of human moving direction

In this paper the moving is consider with the continuous frames minus to the original empty human images. We construct the classification algorithm as follow:

- Record a point with starting time to record the prediction of the human direction. Record the lowest position of human object at the starting time, corresponding to foot of human.
- Identify the stepping time in seconds to monitor moving action; this should be less than a second. Record the lowest position of human object at the stepping time, corresponding to foot of human.
- Drawn the virtual line from first point to the second point to predict the next positions. This process is useful at the place near the unmonitored area.
- Record line when moving is not by predicted way.
- Count the number of times at correct and incorrect prediction.

This algorithm is tracked and predicted the lines from where human is moving, this reflects the appearance of the furniture of room. We are building up the simple recorded engine for better performances as follow:

- Specify the time buffer recorded in days when moving action is kept at database.
- Record lines in room where people keep moving.

These are results from previous algorithm of tracking. Corresponding to each line, recorded the number of times that were predicted correctly.

- When predicting, if detected the old line in database were used, use the largest recorded line with a specific day to predict the moving directions.

4. IMPLEMENTATION AND EXPERIMENTS

The positioning algorithm approximately identify the position where people standing base on the result from home context were identified in part three, the advantage of this process is to reflect the relation of human position to furniture without any mapping to the real furniture objects. This research useful in case of large room which includes complicated furniture and not suitable for installing any sensor, and because the relation between cameras in network is not closed, these could expand to more than four cameras installed in the two lines run parallel, that case would required to specify the distance between cameras, not the distance in the room. The method has the disadvantage of does not specify correctly the foot of object as in condition of being hide after the furniture, this case leads to the inaccuracy in the final result, but depending on how large the room, the error is not much effect.

In our case we used the WebEye camera network. There are four cameras and one of them is main camera, the system can be access through network and the frame rate is around 10 frames per second, these are because of the network transferring data. To manipulate with the image data we used OpenCV library. The library includes technique for creating tracking windows, mouse event, image operations, and so on. The main program uses MFC library provided by camera manufactory to capture the network frames. On each initial point of opening the camera, the user preference takes part in specifying the three points in image; this result is recorded for future works.

5. CONCLUSION AND FUTURE WORK

There are much of works studied in the ubiquitous smart home but the contexts are diversified in their own way. This paper mainly focuses on the visual problems, specifying the location; the way of moving and simple rule to predict is generated. Using this result, future work for further complicated human oriented services building are in the way of studying.

6. REFERENCES

[1] J. Choi, D.K. Shin, DI. Shin, “Research and implementation of the contextaware middle ware for controlling home appliances”, Consumer Electronics, IEEE Transactions. Vol 51, Issue 1, 2005, pp 301-306.

[2] M. Tominaga, H. Hongo, H. Koshimizu, Y. Niwa, K. Yamamoto, “Estimation of human motion from multiple cameras for gesture recognition”, *Pattern Recognition*, 2002, pp 401-404.

[3] J. Krumm, S. Harris, B. Meyers, M. Hale, S. Shafer, “Multi-camera multi person tracking for easy living”, Visual Surveillance. Proceedings, Third IEEE International Workshop, 2000, pp 3-10.

[4] Q. Gao, A. Parslow, M. Tan, “Perceptual motion tracking from image sequences” *IEEE*, 2001, pp 389-392

[5] S. K. Das, D.J. Cook, A. BhattaCharya, E.O. Heireman III, T. Lin, “The role of prediction algorithms in the mavhome smart home architecture”, *IEEE Wireless Communication*, Dec 2002. Pp 77-82

[6] A. Roy, S. K. Das Bhaumik, A. Bhattacharya, K. Basu, D.J. Cook, S.J. Das, “Location aware resource management in Smart Homes”, *Proceedings of the First IEEE International Conference on Pervasive Computing and Communications*, 2003, pp. 481–488.

[7] D. Wang, X. Ma, “Multimedia data mining for building rule-based image retrieval systems”. *IEEE International Conference on Multimedia and Expo*, 2005.

[8] T. Yamazaki, “Beyond the Smart Home”, *IEEE International Conference on Hybrid Information Technology*, 2006.

[9] E. D. Manley, J. S. Deogun, “Location learning for Smart Homes”, *IEEE 21st International Conference on App workshops*, 2007.

[10] V. Takala, M. Pietikainen, “Multi-object tracking using Color, Texture and motion”, *IEEE Conference on Computer vision and Pattern Recognition*, 2007.

[11] C. Yu, C. Wu, C. Lu, L. Fu, “Human Localization via Multi-Camera and Floor sensors in Smart Home”, *IEEE International Conference on Systems, Man, and Cybernetics*, Oct 2006, pp 3822- 3827.

[12] H. Zheng, H. Wang, N. Black, “Human activity detection in Smart Home Environment with Self-Adaptive neural networks”, *IEEE International Conference on networking, sensing and control*, 2008, pp. 1505–1510.

[13] S. Li, Robert Lee, S. Lang, “Mining distance-based outliers from Caterorical data”, *7th IEEE International Conference on Data Mining*, 2007, pp. 225–230.

[14] A. Roy, S. K. Das, K. Basu, “A predictive framework for location-aware resource management in Smart Homes”, *IEEE Transaction on Mobile computing*, Nov 2007, pp. 1270-1283.

[15] M. Weiser, “The computer for the 21st Century”, Scientific American, Sept 1992, pp 99-104.

[16] Weisstein, Eric W. "Law of Cosines." From *MathWorld* -- A Wolfram Web Resource. http://mathworld.wolfram.com/LawofCosines.html

[17] Weisstein, Eric W. "Inverse Cotangent." From *MathWorld*-- A Wolfram Web Resource. http://mathworld.wolfram.com/InverseCotangent.html

[18] Weisstein, Eric W. "Law of Tangents." From *MathWorld* -- A Wolfram Web Resource. http://mathworld.wolfram.com/LawofTangents.html