Design of Face Tracking System Using Environmental Cameras and Flying Robot for Evaluation of Health Care

Veerachart Srisamosorn[1(✉)], Noriaki Kuwahara[2], Atsushi Yamashita[1],
Taiki Ogata[3], and Jun Ota[3]

[1] Department of Precision Engineering, Graduate School of Engineering,
The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku 113-8656, Tokyo, Japan
`veera.sr@race.u-tokyo.ac.jp`, `yamashita@robot.t.u-tokyo.ac.jp`
[2] Department of Advanced Fibro-Science, Kyoto Institute of Technology,
Kyoto-shi 606-8585, Japan
`nkuwahar@kit.ac.jp`
[3] Research into Artifacts, Center for Engineering (RACE), The University of Tokyo,
5-1-5 Kashiwanoha, Kashiwa-shi 277-8568, Chiba, Japan
`{ogata,ota}@race.u-tokyo.ac.jp`

Abstract. This paper presents a face tracking system for evaluation of
health care service for elderly people in a care house. As face can show
patient's smile and emotional response, it can be used for evaluation of
the quality of health care and treatment provided to each patient, and
therefore can be used to improve the quality of care. The conceptual sys-
tem consists of cameras fixed in the environment to provide information
about each person's location and face direction, and moving cameras for
tracking the faces. To prove the concept, a system with 5 fixed Kinects
and a quadrotor was set up to cover the area and track one person The
experiment shows that the system can control the quadrotor to follow
the movements by the person. By attaching a wireless camera to the
quadrotor, facial images can be obtained from the system, proving the
validity of tracking.

Keywords: Health care evaluation · Face tracking · UAV · Quadrotor

1 Introduction

In health care practice, the quality of the health care should be regularly assessed
in order to maintain the quality of the care as well as adjusting measures to
improve the performance. According to Donabedian Model [1], care can be eval-
uated according to their structure (organization, facility, and staff), processes
(activities), and outcomes (symptoms, rate of reoccurence, etc.) Therefore, by
observing the activities during health care and treatment as well as the outcome
on the patient's expression, quality of the care can be evaluated. An example of
health care provided to elderly people in care house is taken into consideration.

Traditional method for evaluating the cares quality involves caretakers observing the patient's face for the smile and emotional responses. As the number of patients per staff is high, staff have difficulty in continuously observing patients while providing care. It is also necessary to record the activities and interactions between patients and caregivers as well as among patients to understand their social relationships. An example of Media Therapy is practiced with dementia patients, in which patient, family members, care managers and caregivers sit together around the table and pictures of important events in the past (for example wedding ceremony) are shown on the screen and they have discussion on the topic. Facial expression on the patient's face is recorded and evaluated by a camera placed on the table. Without movements, placing camera on the table is enough for recording the patient's face, but not interaction with others. There are also other therapies which include movements of patients and placing camera on the table cannot guarantee that face will always be recorded. Therefore, a system for tracking and recording person's position and face, both for the patients and the caregivers, in an area is important. The recorded video and images can be used for further analysis and as a part of the report by caregivers to transfer their observation on the patients to other caregivers in different shifts, giving them a hint on what needs to be carefully observed from the patients in the next shift. Care managers can also use the video to judge each caregiver if they did the right practice or not.

In order to achieve the record of people's faces and positions, a tracking system is required. There are a number of publications about systems for tracking people's positions, such as W^4 [2], and a system of multiple stereo cameras [3], without the effort on tracking people's faces. Face tracking at a distance sometimes uses the technique of multi-camera active vision system, in which wide field-of-view (WFOV) cameras detect and locate people while narrow field-of-view (NFOV) cameras are actively controlled to capture high-resolution faces by using pan-tilt-zoom (PTZ) commands. Examples of systems utilizing this method for face tracking are described in [4,5]. Face tracking can be achieved, but with the NFOV camera set in one direction, tracking is limited to the case of person walking in one direction, i.e. towards the camera. There are also robots tracking human beings and their faces [6–8], but the person to be tracked must be in front of the robot before tracking can begin. We also proposed a system of one depth-sensor and a flying quadrotor for tracking a person's face in [9]. This work expands the system to utilize multiple depth-sensors to enable tracking of the person in wider area and more various movements.

2   Problem Statement

Considering the application of recording the patient's face while receiving health care, a camera is required to be in front of each patient. The camera should be at appropriate angle and distance to ensure that the obtained face can be used in evaluation. The following assumptions are applied to the system:

- The environment (room and cameras' positions) is not changing during the tracking process.
- There is only one person in the area.
- The movement of the person being tracked is smooth and not too fast at standard speed (approximately around 1 m/s).
- The behavior of the person's face looking up and down is not considered.
- The person being tracked turns with the whole body, not by turning only his/her head.

3   System Design

3.1   Utilization of Cameras and Camera Configuration

Tracking position in indoor environment by using cameras was chosen as no device needs to be carried by the person or object to be tracked. The price is also relatively cheap. Accuracy is not so high but enough for the application. Depth camera was selected as it can provide 3D information of the positions.

Cameras can be configured for tracking in various methods. Utilizing only environmental cameras fixed in the environment is simple to implement, but requires large number of cameras in order to completely cover the whole area. Utilizing only moving cameras that follow the movement of people can ideally reduce the number of cameras down to one camera per person. However, searching for the person is required before tracking can begin, and it is necessary to search every time tracking is lost. Therefore, we propose to use the combination of both environmental cameras and moving cameras. Environmental cameras provide the information about the location and direction of each person as well as the position of each moving camera, while moving cameras use the information to move to the position where they can capture facial images at better quality. This method reduces the number of required cameras, as the environmental cameras do not need to see the face. Searching is also replaced by the use of position information from the environmental cameras. Moving cameras can also get closer to the faces and therefore give images with higher resolution.

3.2   System Overview

The system uses the combination of environmental cameras, depth cameras placed at fixed locations and orientations, and moving cameras, small cameras attached on moving robots. Environmental cameras provide information about each person's position and direction, as well as each moving camera's position (Fig. 1a). This information is used to set up the goal for each moving camera where the face of each person can be captured, i.e. in front of the person at an appropriate distance (Fig. 1b), and control the moving cameras so that they move to the goal position (Fig. 1c).
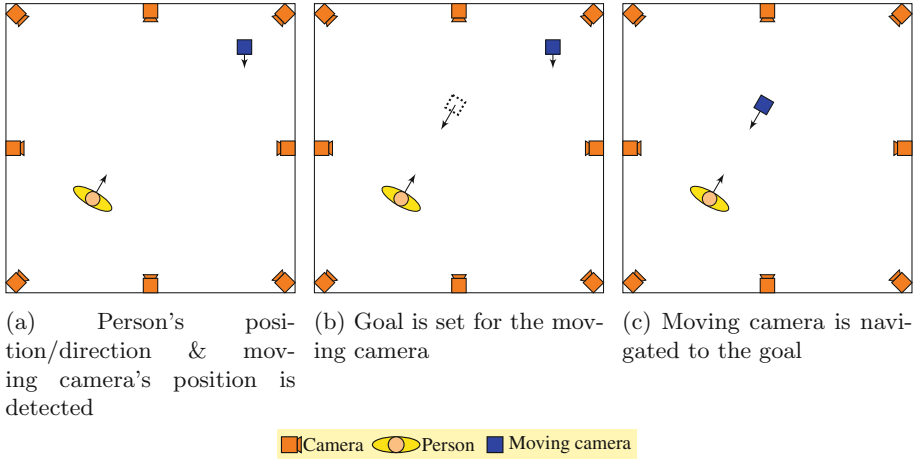
(a) Person's posi- (b) Goal is set for the mov- (c) Moving camera is navi-
tion/direction & mov- ing camera   gated to the goal
ing camera's position is
detected

■Camera ⬭Person ■ Moving camera

Fig. 1. Concept and steps of the system

4   Experiments and Results

4.1   System Implementation

The system was constructed in our laboratory as a test for the validity of the
design. Xbox Kinect sensors were chosen as the sensor for acquiring depth infor-
mation in the role the environmental cameras. Aerial robot was picked up for the
choice of the robot carrying moving camera as its workspace does not overlap
with human's moving space so it is more agile. Bitcraze's Crazyflie 2.0 quadro-
tor [10], shown in Fig. 2, was selected from among other flying robots for the task
of moving camera in this experiment due to its small size and programmability.

Kinects were set up in the selected environment to cover the desired area of
3.0 m by 3.5 m, from the height of 0.7–2.5 m from the floor for both detection
of the person and the quadrotor. Position and orientation of each Kinect was
determined by optimization using the experimental space's dimension, possible
location of cameras, and model of camera's field of view (FOV). Simulation was
done to minimize the number of cameras and maximize the coverage of the whole
area by adding one camera at a time. The best result uses 5 Kinects according
to Fig. 3.

The system runs on Robot Operating System (ROS) [11]. The program is
based on the package for Crazyflie control by Oliver Dunkley [12], which pro-
vides the control of the quadrotor by using joystick controller or inputting goal
position via graphical user interface (GUI), obtaining the position of the quadro-
tor by background subtraction on the depth image from a single Kinect. Our
modifications include multiple-Kinect integration, data fusion, human tracking
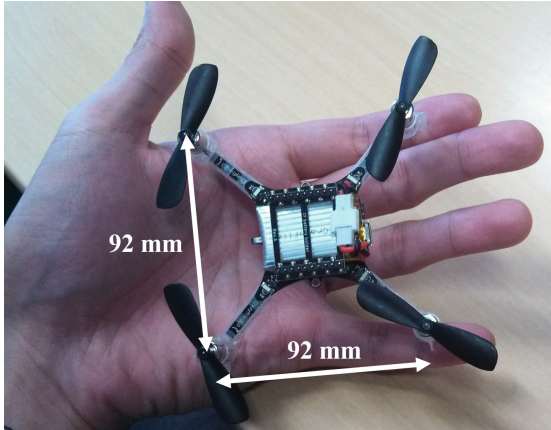and controlling based on human's position and direction.

Fig. 2. Crazyflie 2.0

Human detection and tracking are done by OpenNI library, using ROS package openni_tracker [13]. The package provides approximation of position and orientation of each joint of the body, and the head's position and orientation are used. Data from multiple Kinects are fused together as the head of the same person if they are close together. The position and orientation of the fused head are used for setting up the goal for the moving camera to track each person, 1.5 m in front and 0.6 m above for safety and avoiding too direct observation of a person.

Positions of the quadrotor from different Kinects are also fused together when they are close together. Due to the size of the quadrotor, it is prone to false detection. To prevent the system from this false detection, the information about the number of Kinects observing the same object is used. There is higher chance that the detected object is real quadrotor and not the noise if there is more than one Kinect observing this object. Therefore, at the first time an object is detected, the number of sensors seeing that object is also obtained. If the number is more than one, it is considered as a real quadrotor, and tracking starts. If the number is one, it may be a noise. In the next observation, if there is no observation close to this object, there is high chance that it is a noise and it is removed from tracking. However, if there is more than one Kinect seeing it in the next observation, it is a real quadrotor and tracking starts. The number of quadrotors being tracked is limited to the number of quadrotors being used, which is known by the user beforehand.

As Kinect sensor utilizes unmodulated infrared light pattern for calculations of the depth [14], when multiple Kinects are used together in the same area, patterns overlap each other and pattern from one Kinect interferes with the patterns of other Kinects, resulting in confusion and loss of depth data in the intersected area. A vibration unit consisting of a DC motor and an unbalanced weight, as proposed in [15,16], is added to each Kinect in order to blur the

patterns from other Kinects and keep its own pattern clear, as the pattern projector and receiver synchronously move together. The unit can solve the interference problem, as shown in Fig. 4 but also creates some disturbing noise. However, this will be ignored at this moment.
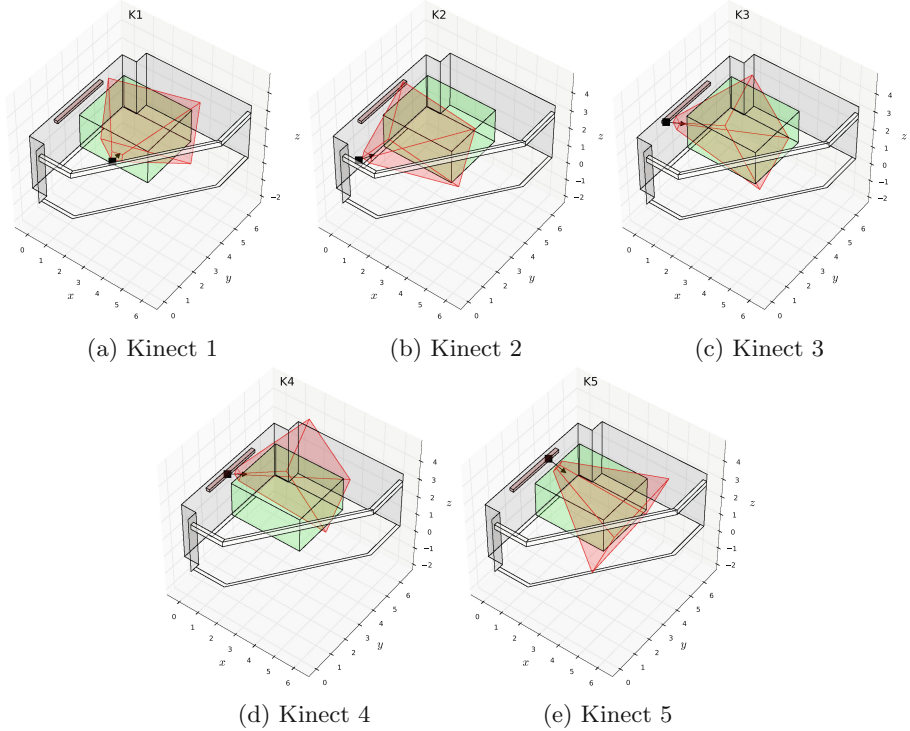

(a) Kinect 1   (b) Kinect 2   (c) Kinect 3

(d) Kinect 4   (e) Kinect 5

Fig. 3. 3D FOV of each Kinect and the region of interest

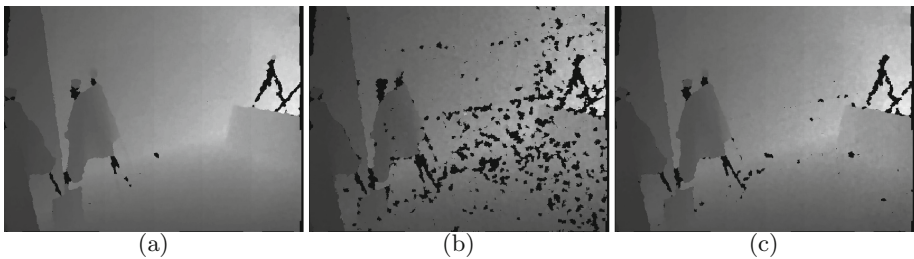(a)   (b)   (c)

Fig. 4. Result of interference removal by vibration: (a) without interference, (b) with interference from other Kinects, (c) using vibration with interference from other Kinects
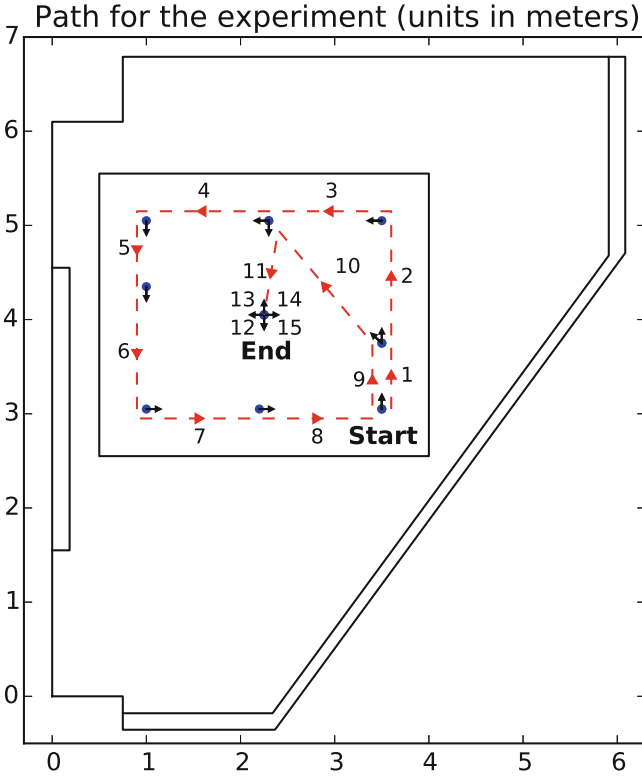
Fig. 5. Path for the experiment with the boundary of the setup area

4.2 Experimental Setup

To evaluate the tracking ability of the system, an experiment with a person, assuming the role of a patient, moving inside the area according to the path shown in Fig 5 was performed. The person walked along the numbered path, stopped at the markers on the floor (denoted by dots in the figure), facing in the direction of the arrows. The path ends in the center of the area, with the person turning around the point, stopping at around $-\frac{\pi}{2}$, $-\pi$, $\frac{\pi}{2}$, and 0 radian respectively, before finishing at $-\frac{\pi}{2}$ radian.

4.3 Results

Figure 6 shows the snapshots of the tracking experiment (quadrotor in the circle). The video can be found at http://youtu.be/OdvLoFQu5gk. From the video, we can confirm that the system can control the quadrotor to move and follow the motion of the person inside the designed area.

Fig. 6. Snapshots of quadrotor following the movement of a person

Fig. 7. A snapshot of the video taken by the camera on the quadrotor

By adding a small wireless camera to the quadrotor and testing the system again with random path, real facial images could be obtained from the on-board camera as shown in Fig. 7. Vibration and transmission noises were present so the quality of the video was not so high.

5   Conclusion and Future Works

In order to record elderly person's position and facial images for his/her facial expression in response to care and treatment provided in health care facility, a face tracking system utilizing environmental cameras and moving cameras is presented. The system is implemented by using multiple Kinect sensors, placed in positions and orientations obtained by optimization, and a small quadrotor. The experiment showed that with the system, the moving camera could move to follow the movement of the person inside the designed area. With a wireless camera attached to the quadrotor, facial images could be obtained by the proposed tracking system. The concept was proven to be effective for tracking of people's position and face in indoor environment.

Using quadrotors to move cameras has some drawbacks. Small quadrotors have quite short battery life (7 min without any loads for Crazyflie 2.0) and even larger quadrotors cannot fly longer than half an hour. Moreover, noise from the continuously rotating propellors are quite disturbing and can create the fear of falling and hitting the elderly people. This may have effects on the facial expression obtained and therefore alter the result of health care evaluation. In the next development, quieter, less power consuming, and safer helium-filled blimp will replace the noisy quadrotor. Kinect sensors would also be replaced by the new technology of 360-degree cameras so that activities and interaction of elderly people and caregivers can also be recorded.

References

1. Donabedian, A.: The quality of care: how can it be assessed? JAMA **260**(12), 1743–1748 (1988). http://dx.doi.org/10.1001/jama.1988.03410120089033
2. Haritaoglu, I., Harwood, D., Davis, L.: W^4: real-time surveillance of people and their activities. IEEE Trans. Pattern Anal. Mach. Intell. **22**(8), 809–830 (2000)
3. Zhao, T., Aggarwal, M., Kumar, R., Sawhney, H.: Real-time wide area multi-camera stereo tracking. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005, vol. 1, pp. 976–983, June 2005
4. Stillman, S., Tanawongsuwan, R., Essa, I.: A system for tracking and recognizing multiple people with multiple cameras. In: Proceedings of Second International Conference on Audio-Visionbased Person Authentication, pp. 96–101 (1998)
5. Wheeler, F., Weiss, R., Tu, P.: Face recognition at a distance system for surveillance applications. In: 2010 Fourth IEEE International Conference on Biometrics: Theory Applications and Systems (BTAS), pp. 1–8, September 2010

6. Ali, B., Qureshi, A., Iqbal, K., Ayaz, Y., Gilani, S., Jamil, M., Muhammad, N., Ahmed, F., Muhammad, M., Kim, W.Y., Ra, M.: Human tracking by a mobile robot using 3d features. In: 2013 IEEE International Conference on Robotics and Biomimetics (ROBIO), pp. 2464–2469, December 2013

7. Bellotto, N., Hu, H.: Multisensor-based human detection and tracking for mobile service robots. IEEE Trans. Syst. Man Cybern. Part B: Cybern. **39**(1), 167–181 (2009)

8. Vadakkepat, P., Lim, P., De Silva, L., Jing, L., Ling, L.L.: Multimodal approach to human-face detection and tracking. IEEE Trans. Ind. Electron. **55**(3), 1385–1393 (2008)

9. Srisamosorn, V., Kuwahara, N., Yamashita, A., Ogata, T., Ota, J.: Automatic face tracking system using quadrotors: Control by goal position thresholding. In: 2014 IEEE International Conference on Robotics and Biomimetics (ROBIO), pp. 1314–1319, December 2014

10. Bitcraze AB Company: Bitcraze. http://www.bitcraze.io/. Accessed 25 Aug 2015

11. ROS.org — Powering the world's robots. http://www.ros.org/. Accessed 25 Aug 2015

12. Dunkley, O.: GitHub omwdunkley/crazyflieROS, downloaded branch joyManager. http://github.com/omwdunkley/crazyflieROS. Accessed 15 April 2014

13. Field, T.: openni_tracker - ROS Wiki. http://wiki.ros.org/openni_tracker. Accessed 07 July 2015

14. Khoshelham, K., Elberink, S.O.: Accuracy and resolution of kinect depth data for indoor mapping applications. Sensors **12**(2), 1437 (2012). http://www.mdpi.com/1424-8220/12/2/1437

15. Maimone, A., Fuchs, H.: Reducing interference between multiple structured light depth sensors using motion. In: Virtual Reality Short Papers and Posters (VRW), pp. 51–54. IEEE, March 2012

16. Butler, A., Izadi, S., Hilliges, O., Molyneaux, D., Hodges, S., Kim, D.: Shake'n'sense: reducing interference for overlapping structured light depth cameras. In: Proceedings of the 2012 ACM Annual Conference on Human Factors in Computing Systems, pp. 1933–1936. ACM (2012). http://research.microsoft.com/apps/pubs/default.aspx?id=171706