平成 30 年度　博士論文（要約）

東京大学
THE UNIVERSITY OF TOKYO

Development of Face Tracking System using Fixed Cameras and Flying Robot in Indoor Environment

(飛行ロボットと固定環境カメラを用いた顔追跡システムの開発)

指導教員　太田　順　教授

東京大学大学院　工学系研究科　精密工学専攻
学生証番号 37-157261

Veerachart Srisamosorn

シーサモーソーン　ウィーラシャート

Human localization in indoor environment has been an active field of research, and can be used in many applications such as surveillance, behavior and interaction analysis. Faces can provide more information and can be used for personal identification, emotion prediction, or human-robot interaction. This thesis presents the design and construction of face tracking system using the combination of fixed and moving cameras. The system utilizes a combination of fixed cameras installed in the environment to track the location of the person and control the moving camera to follow the movement, in order to obtain facial images from the person. There should be no limit in the type of sensors nor robots to use as well as the environment of implementation.

The configuration of using both fixed cameras and moving cameras is proposed so that the moving camera can start tracking a person even without the face observation inside its frame by using the information from the fixed cameras to approach to the face directly. This can also be very useful when the person turns, which requires the robot to sweep around the person at the same rate of rotation. When the robot cannot keep up with the rotation, face can be lost from the camera and tracking would have been stopped if only the on-board camera is used. With the help of the fixed cameras, the robot can still know where it should go in order to recover the face and continue tracking. Two systems using two different combinations of robot-sensor were designed, with different characteristics and for different applications.

Kinect-quadrotor system was proposed with multiple Kinect sensors for the duty of environmental cameras and a miniature quadrotor to carry moving camera. An experiment of tracking a person within a designated area was conducted, with the person moving and rotating around, stopping from point to point. Even without a camera attached to the quadrotor, it could track and follow the movement of the person and keep flying in front, rotating its front side toward the person's face, mimicking pointing the camera to capture the facial images.

Fisheye-blimp system using multiple fisheye cameras as environmental cameras and a blimp as moving camera was also proposed for the same task of face tracking of a person moving and rotating inside the area. Firstly, human detection and direction estimation system is required. Based on some methods available in perspective cameras, modifications were made to combine human tracking and direction estimation methods to work in fisheye images. Descriptor using histogram of oriented gradients (HOG) was modified for human detection inside fisheye camera, and was named "FEHOG". FEHOG is also used in the method for estimating head direction, adapted from the method used with perspective camera. This is believed to be the first system utilizing fisheye camera for estimation of head direction. Tested with a publicly available dataset and manually marked direction ground truth, the system could provide the direction estimates with errors less than 90 degrees for about 91 percent and less than 45 degrees for 70 percent. The results on a recorded dataset with ground truth from motion capture system were also similar. This estimation can be used to guide the blimp toward the front side of the person.

A blimp was constructed with a wireless camera attached to the front. Navigation of the blimp in the case that face cannot be seen in the on-board camera is considered. Information from the fisheye cameras is used to set a goal in front, and the blimp is navigated toward the goal position by direct waypoint navigation. When the face can be observed, it is first matched with the existence of the person in the fisheye camera system

to avoid tracking false positives. After the confirmation, visual servoing can be used to track the face. An office environment was set and an experiment was conducted with the person not visible by the camera at the first place. The system could use the information from the fisheye camera to set the goal for the blimp and waypoint navigation was used to control the blimp to the goal. When waypoint navigation succeeded, the blimp could navigate closer to the person, and then detect the face to use it for adjusting and following the movement.

The two systems were analyzed for their differences from various points of view and the suitable situations for each system. Selecting the right system to solve the problem in the desired application can be done by studying the requirements of that application and picking the system which can satisfy these requirements properly.

In Chapter 1, motivation and background of the researches as well as some related works about human and face tracking systems and human tracking robots are discussed. The objectives of this research together with the challenging points to solve are listed, as well as the approach to each point.

Chapter 2 discusses the definition of tasks and the problem to solve. System configuration of fixed environmental cameras and moving cameras for the task of face tracking is explained. Fixed cameras can provide information about the person's and robot's positions and directions, and then the robot can be controlled toward the person's face without even seeing the face from the beginning. Two systems are proposed with different sensors and robots: one with depth-sensor-quadrotor combination and the other with fisheye-camera-blimp combination. System requirements are also determined, with assumptions to be applied to simplify the problem.

Chapter 3 describes the design and construction of the system implementing depth sensors and quadrotor. Kinect sensors are used as environmental cameras for providing the position and direction of the person as well as the position of the quadrotor. Small-size *Crazyflie 2.0* is used for the task of moving camera. An experiment of tracking a person moving and rotating is conducted for evaluation in Section 3.3, with the quadrotor controlled to fly in front of the person with its front turning toward the person to mimic face tracking behavior.

System for tracking human and estimating direction in fisheye images, to be used in the fisheye-blimp system, is explained in Chapter 4. With the information, the blimp can be controlled to move and follow the person's movement as in the depth-sensor-quadrotor system. Histogram of oriented gradients (HOG) descriptor was modified from perspective image domain to be usable in human detection in fisheye images. This is called "FEHOG". Human's body and head positions are tracked and the head area is used for head direction estimation. The method used in perspective human head images is adapted to work with fisheye images by changing the original HOG descriptor to the proposed modified HOG descriptor. Experiment of human tracking and direction estimation is performed and evaluated on a dataset. Another dataset is recorded together with motion capture ground truth for further evaluation.

Blimp construction and its navigation for face tracking task are discussed in Chapter 5. Planning algorithm based on slower dynamics of the blimp is considered using both information from the fisheye cameras and the on-board camera, depending on their availability and certainty. The blimp's camera should have a better view of a person but only when he/she is in front of the camera. There is also some possibility of false positive detection. Both information of human position from fisheye cameras and face detected by the blimp's camera will be matched together to confirm the correctness of the detected face. When it cannot observe any person, information from the fisheye cameras will be used to move the blimp toward the target until it is in the range to see the person. When the face is found and it corresponds to the tracked person in the fisheye cameras, face direction is also estimated from the blimp's camera, and used to adjust the blimp's position accordingly. An experiment was conducted in an office environment, moving the blimp to track the person moving inside the area, using information from both fisheye cameras and on-board camera.

Chapter 6 provides the discussions about the whole thesis. Two proposed systems are compared for their benefits and limitations. Scalability and possible uses of the systems are also discussed. The thesis is concluded in Chapter 7 with suggestions of some future topics available for research.