A Study on Wide Area Situation Recognition by Aerial Image Analysis

Ganwen JIANG

Kagoshima University

March 2014

CONTENTS

Chapter 1

Introduction

In this paper, an automatic broad surveillance system is proposed to understand wide area situation. In our study, “wide area situation recognition” is not a fixed board field of view due to a camera which is attached to balloon, but it is intended to cover dynamically wide space from above to track the movement of a person. Specific contents of the state recognition contain suspicious person detection and tracking, abnormalities detection, confirmation of crop growth state, life log auto-created, and so on. To achieve automatic processing of target tracking and state monitoring, a flying robot that is equipped with camera is used for getting moving images and a new image processing method by confusing multiple modules is proposed for analyzing aerial image and calculating the position and moving direction of the subject. Then, the position and moving direction are used to control flying robot to track the subject automatically.

With development of image processing technology and UAV (Unmanned Aerial Vehicles) technology, wide area situation recognition is concerned by more and more researchers. By understanding wide area situation, we can grasp the important information of the target, such as size, color, position, direction, what is doing, and so on. It can be applied to many fields, such as the surveillance, human life log, artificial satellite, medical science, and so on. Therefore, object detection is very important. The image processing technology is used to detect and analyze the characteristics of the target. In addition, due to these characteristics, the target can be tracked, and it can help us to continuously understand information of the target in detail. Above said is a general procedure for monitoring system. In view of this, the image processing technology is very important, so we must understand the development of image processing technology.

The history of the image processing technology has been over 50 years. Meanwhile, a lot of technologies and theories have been developed, and various application systems have been commercialized [1]. As shown in Table.1, the digital image processing by using computer was beginning around 1960. Image quality improvement of satellite images and the binary image expressed by the character recognition have been attempted [2]. The computer vision technology started from the late 1960s. Studies on computer vision are not only for characters, symbols, and graphics recognition in two-dimension, but also for object recognition in three-dimension. Therefore, From 1970s, developments on the industrial vision processing, satellite image processing and medical imaging processing are prosperous. In 1972, Computer tomography developed by EMI is widespread concerned. In the same year, digital images of resource data of the earth's surface are transmitted to the ground by the Space Resources Satellite Landsat traveling around the track, and remote sensing field is flourished. About industrial vision processing, studies of industrial usage of vision technology begin in early 1970s, and major field of robotics is formed in late 1970s.

Table.1 Historical changes of image processing technology

	Technology	**Specific Examples**
1960s	Image processing in science	satellite photo processing
	Computer vision	Processing of track photo of the bubble chamber
	Character recognition	Reading apparatus of handwritten postal code
1970s	Remote sensing	Processing of Landsat image
	Medical imaging	CT tomography
	Industrial vision	Semiconductor assembly
1980s	Standardization of image processing software	ETL "SPIDER" published
	Hardware for image processing	Development of dedicated processors
	Practical technology	Application for drawing recognition, facsimile, copy machine
1990s	Sophisticated processing	Expert system
	Faster processing	Parallel processing, pipeline processing
	Spread of applications	Dedicated image processing system
2000s~	Non-image information→ high concrete media	Information fusion
	Robot vision	Human cooperative
	Multimedia database	Purposive

Especially, the contribution of visual technology on the semiconductor assembly is remarkable. Because the automation, efficiency, and labor saving in the production process are produced, it is a significant contribution to improve quality of semiconductor devices. In addition, entering 1970s, the interactive image processing system by using a mini computer is more and more popular. In the 1980s, as above studies developed, technologies of drawing recognition, document understanding, facsimile, and copier are spread. By promoting the development of dedicated processor, practical application of remote sensing and industrial applications has become popular. Entering 1990s, use of knowledge, the introduction of image processing expert system, parallel processing for high-speed processing, and pipeline processing are rapidly progressed [3].

In the 21st century, four fundamental technologies are quickly developed. Firstly, object of image processing is from non-image information to high concrete media, and image processing and application technology is becoming the leading role. Furthermore, robot vision also has been developed rapidly. It use two-dimensional information to restore the actual scene of three-dimension, and a number of conventional approaches were performed. In 21st century, according to understanding the realistic situation, it becomes more and more important to close to the cognitive ability of human vision. In addition, technologies of multimedia database and visual interface are also flourished [4].

As a monitoring system, it contains three important parts, monitoring section, transmitting section and control section. In monitoring section, moving images are got. Then, images are transmitted to control section by transmitting section. In control section, images are analyzed by using image processing technology, and then control command is transferred to monitoring section. This procedure is shown in Fig.1.

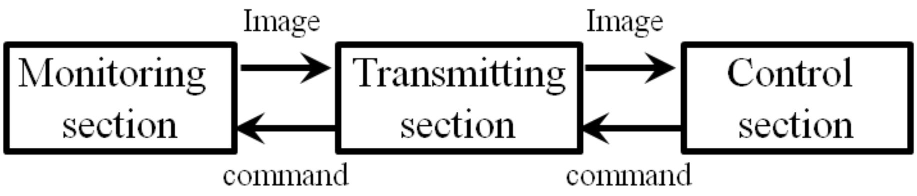

Fig.1 Working process of a monitoring system

For a monitoring system, the Image Processing Technology is not enough. We must consider about the capability and cost of monitoring section, so it can help me to choose the monitoring equipment. For example, if we want to carry out the fixed point surveillance, we can choose one or multiple surveillance cameras. If we want to track the target always, we can choose running robot or aerial robot that has tracking capability. In addition, if we want to monitor the target in wide area, it is important to keep wide visual field, so we can choose aerial robot. In the Table.2, differences between the surveillance camera [5], running robot [6], and flying robot [7] are shown. The visual field of surveillance camera is limited, and it cannot track the target always, so it is not suitable for wide area monitoring. Due to running robot, the view field is also limited, and it may be affected by obstructions, so it also not suitable for wide area monitoring system. About flying robot, it can fly in air, so the visual field is larger than surveillance camera and running robot. Also, flying robot is almost not affected by obstructions. However, it may be affected by outside influences, such as wind, rain, airflow, and so on.

Table.2 Differences of surveillance equipments

Property / Equipment	Visual field	Tracking capability	Influence of obstacles	Cost	Speed
Surveillance camera	Limited	No	Yes	Cheap	No
Running robot	Limited	Yes	Yes	Expensive	Slow
Flying robot	Wide	Yes	Little	Expensive	Fast

Because of this, aerial robot is chosen for the wide area monitoring system. The development of aerial robot is rapid, especially in the Unmanned Aerial Vehicle (UAV). As shown in Table.3, the history of UAV is very long. In 1896, the flight test of the Model Aircraft named Aerodrome had been performed by Langley [8]. In 1903, the Manned Power Aircraft is tested by Wright brothers [9]. Around 1914, the development of autopilot system by using Gyro is successful. During the First World War, the UAV integrating wireless control and autopilot is developed. However, because peripheral technology is immature, it cannot be for practical realization [10]. In 1928 and 1935, the Biplane UAV with wireless control is successful in America and Britain respectively. During the Second World War, the development of UAV system is very quickly. For example, Reginald Denny Industries of Hollywood star produced Radio Controlled (RC) target planes for the military from 1942 to 1945 [11]. After the Second World War, the Military UAVs are researched in many countries. In addition, after the Vietnam War, the Large-scale UAVs are developed in America. In 1987, Pioneer UAV of Israel is successfully tested [12]. In 1995, Civilian UAV named Aerosonde is successfully developed in Australia [13]. Entering 2000, as the technologies of GPS, Sensors, property of computer improving, the Military UAV and Civilian UAV are both rapidly developed in many countries, such as America, Japan, China, and so on.

Nowadays, there are about 300 types UAVs in the world [14]. A UAV is the prominent part of a whole system that is necessary to fly the aircraft. Even though there is no pilot physically present in the aircraft, this doesn’t mean that it flies by itself autonomously. In many cases, the crew responsible for a UAV is larger than that of a conventional aircraft. The aircraft is controlled from the ground (the Ground Control Station or GCS), so it needs reliable communication links to and from the aircraft, but also to the local Air Traffic Control (ATC) authorities if required (usually when flying higher than 150-200 m above the ground). The GCS provides a working space for a pilot, navigator, instrument operator and usually a mission commander [15].

Table.3 Historical changes of Unmanned Aerial Vehicle

Year	Event
1896	Flight Test of Aerodrome by Langley
1903	Test for Manned Power Aircraft by Wright brother
1914	Development of autopilot system by using Gyro
First World War	Development of UAV with autopilot and wireless control by Elmer A. Sperry (but practical use is difficult)
1928	Wireless control of biplane UAV is successful in America
1935	Wireless control of biplane UAV is successful in Britain
Second World War	Practical realization of the Radio plane with wireless control
After Second World War	Development of Military UAV
After Vietnam War	Development of Large-scale UAV in America
1987	The Pioneer UAV of Israel is successful
1995	Civilian UAV in Australia
2000~	As the technologies of GPS, Sensors, computer developing, development of the Military UAV and the Civilian UAV is executed in many countries

As above said, the aircraft is controlled from the ground (the Ground Control Station or GCS), so it needs reliable communication links from the aircraft. For achieving a safe and secure society, we would build an autonomous wide area monitoring system (or autonomous flight system).

The flight system contains two parts. One is image processing, and another is flying robot control. To accurately determine the position and traveling direction of the subject, a new visual control method by integrating multiple modules is proposed in image processing part. In addition, we also use the Kalman Filter to predict the position and moving direction of the subject when he is lost. By using accurate data calculated from image processing part of position and moving direction, flying robot can be controlled exactly in flying robot control part. There are four contributions in this study.

(1) To avoid the problem of shielding and camera installation, aerial image is used.

(2) To solve the problem of unable to track the subject out of view, flying robot that has good tracking capability is utilized.

(3) Since viewpoint is different from conventional automatic monitoring system, a new moving image processing technology by integrating multiple modules is proposed.

(4) To control flying robot accurately, the predictive control is used. In addition, it is unique to let flying robot go up to widen visual field and search target again when the subject is lost, so it can prevent the subject lost.

The final goal of our study is to build a fully automatic wide area monitoring system to accurately track the particular people in daytime by using aerial robot. Conceptual diagram of the study is shown in Fig.2. As shown in Fig.2, we must correctly detect the particular people firstly. Then, information of position and moving direction has to be calculated by analyzing motion of the subject. In addition, the information of position and direction of the subject are utilized to exactly control aerial robot. In addition, to prevent target disappeared, the information of position and direction have to be predicted by using the predictive control when human is lost, and then search specified people again. If the subject cannot be found again, flying robot will go up to widen visual field and search the subject in second time.

Therefore, there are three objectives must be achieved.

(1) The particular people must be detected correctly, and then the information of position and moving direction must be calculated by analyzing motion of the subject.

(2) Flying robot must be controlled exactly by using information of position and direction of the subject.

(3) When the subject is lost, the information of position and direction of the target must be predicted accurately by using predictive control algorithm.

By achieving this three aims, the fully automatic monitoring system can be built. This system is expected to be applied in general case, and human autonomous tracking can be achieved.

Fig.2 Conceptual diagram of our study

(* When movement of the subject is detected, the position and moving direction will feedback to aerial robot, and then aerial robot will track the target)

Chapter 2
Related Works

2.1 Aerial Robot Systems

Aerial Robotics as the basis of aerial shoot system is roughly classified into the Unmanned Aerial Vehicle (UAV) type and the lighter-than-air aircraft (LTA) type [16]. In addition, the former is classified for the fixed-wing type (FWT) and the rotary wing type (RWT). With the camera attached on it, it is possible to capture aerial image as the viewpoint position changing. The latter is classified to the airship type (AST) with motive power and the balloon type (BT) without motive power. Table.4 shows the differences between FWT, RWT, AST, and BT aerial robot from six viewpoints.

About the fixed-wing type, it belongs to the Unmanned Aerial Vehicle type. The FWT has tracking capability, and fast speed. However, it cannot hover in air, so it cannot carry out sentinel surveillance. With regard to the rotary wing type, it also belongs to the UAV type. In addition, the RWT not only has tracking capability and fast speed, but also has hovering capability. Therefore, it not only can execute dynamic monitoring, but also can execute sentinel surveillance. The airship type robot belongs to the lighter-than-air aircraft type. Its property is same with the RWT type robot, but flight speed is slower than the RWT type, and the volume is larger than the RWT type. Therefore, if it is as the monitoring equipment, it is easy to be discovered. On the other hand, because of its slow speed, it cannot track the fast-moving objects, such as car, train, and so on. About the balloon type robot, because it has not tracking capability, the range of its application is narrower than the top three.

From related studies, the UAV accounted for a large proportion because of its good properties. According to its property, it can be divided into the Military UAVs, and the Civilian UAVs. The Civilian UAVs are mainly used in civil aviation pictures, Geological Survey, and the Military UAVs mainly used for reconnaissance, combat. The purpose of the Civilian UAVs is for civil uses. Because the Civilian UAVs has superiorities of low cost, no casualties risk, ability to survive, well flexibility, and easy to use, it has been applied for the field of aviation photos, geological survey, high-voltage transmission line inspection, oil pipeline inspection, highway management, forest fire safety inspections, gas prospecting, emergency rescue, ambulance, and so on.

Table.4 Differences of each aerial robot

	Tracking	Hovering	Cost	Speed	Influence of outside
FWT	Yes	No	High	Fast	△
RWT	Yes	Yes	High	Fast	△
AST	Yes	Yes	High	Slow	○
BT	No	Yes	Low	No	○

Next, we will analyze the aerial robot in place to use, automation, and application respectively. According to aerial robot's size, weight, and property, aerial robot can be used to different place and for different application.

About place to use, aerial robot can be used in indoors environment and outdoors environment. In indoors environment, there is little outside influence, such as wind, rain, airflow, and so on. Therefore, size of aerial robot is smaller than aerial robot's in outdoors. In addition, requirement for shape is not so high, because air resistance is small. Due to aerial robot, the streamlined shape is better, but it is not absolute. Table.5 shows the studies are for each type aerial robot indoors and outdoors. Because the fixed wing type UAV has not capability of hovering, UAV must carry out emergency avoidance when there is obstacle. Therefore, obstacle detection is very important for fixed wing type UAV. For example, in the study of Jean-Christophe Zufferey et al. [17], they aim at developing autonomous microflyers capable of navigating within houses or small indoor environments using vision as the principal source of information. They use optical flow method to detect the wall to prevent collisions. The weight of this UAV is only 30 gram, and image of it is shown in Fig.3.

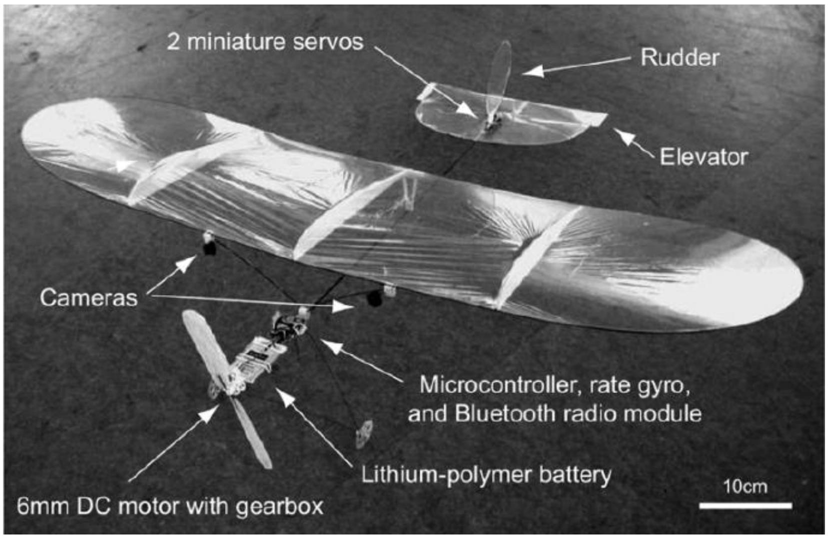

Fig.3 30-g aircraft with description of electronic components, sensors, and actuators

Table.5 Place to use of each type aerial robot

Place / Type	Indoors	Outdoors
FWT	Fly-Inspired Visual Steering of an Ultralight Indoor Aircraft [17]	Autonomous Searching and Tracking of a River using an UAV [18]
RWT	Small automatic surveillance flying robot [19]	Amazon Prime Air [20]
AST	Developing a Low-Cost Autonomous Indoor Blimp [21]	Advanced Virtual Reality Technologies for Surveillance and Security Applications [22]
BT	Research of Broad Area Monitoring by Simplified Aerial Image Acquisition System [23]	Development of the balloon-cable driven robot for information collection from sky and proposal of the search strategy at a major disaster [24]

Nowadays, the target is detected and tracked by aerial robot by means of image processing method, GPS, various sensors, and so on. In other words, image processing technology on UAV development has played a significant role. For example, in the study of Sivakumar Rathinam et al. [18], they use an UAV to detect and track a river by using GPS based tracking and vision based tracking. From the aerial image, because the river component happens to be the longest region crossing the entire image, they detect the river by using the pyramid-linking image segmentation algorithm suggested by Burt et al. [25] [26], but a search area have to be specified by the user.

The autonomous small surveillance flying robot made by SECOM Company has been reported as the news on 26th, December, 2012 [19]. The SECOM Company is the corporation to perform security service for business and home security for individuals in Japan. This study is for indoors surveillance system. It is to detect the intrusion target by laser sensors. Then, it is to decide whether the intrusion target is the suspicious individual. If he is, he will be tracked by this autonomous small surveillance flying robot shown in Fig.4.

Fig.4 Autonomous small surveillance flying robot

As shown in Table.6, there are some differences with our study. SECOM's study uses flying robot to check the suspicious person at front viewpoint, so I think it is not very good. There are many reasons. For example, it would cause dangerous to people, it is easy to be found, and the noise is large. Therefore, the concealing property is bad. Also, it would cause interfere to other people.

About our study, because the flying robot is in air, it would not cause dangerous to people, the noise is small, and it is difficult to be found. Also, it would not cause interfere to other people. Therefore, the concealing property of our system is good. In addition, the system of our study can be used in indoors and outdoors. However, the system of SECOM is only can be used in indoors environment. Because there is a light attached at the flying robot in SECOM's study, so it can be used not only in daytime, but also at night. If attach a light to our flying robot, it also can be used in any time. Therefore, the system we proposed is better than the SECOM's.

Table.6 The differences with our study

	Our study	SECOM study
Viewpoint	Upside	Front
Riskiness	Low	High
Noise	Small	Big
Concealing property	good	Bad
Environment	Daytime	Any time
Other aspects	Not interfere	Interfere

Also, there is an interest for UAV in outdoors environment. The UAVs for delivering products is made by the Amazon Company. This system is tested on the December 2, 2013, and this service is named “Amazon Prime Air”. The system is to deliver the products to customer by using the UAVs made by eight propellers, and the size of the UAVs is similar with the RC Airplane [20]. The delivery process is shown in Fig.5. First step is to prepare the products. Then, second step is to set the products at the UAV made by the Amazon Company. In addition, it is to send the products to customer in step three. Furthermore, the products are delivered in step four. It is possible to achieve this service after four or five years.

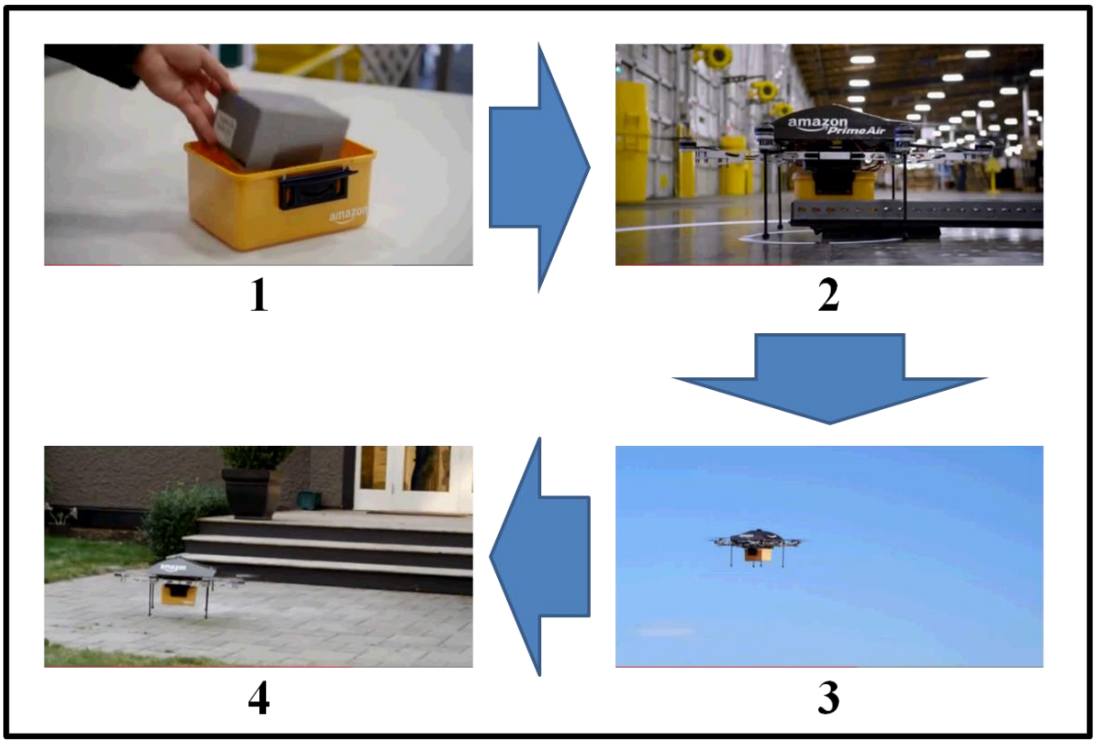

Fig.5 The delivery process of the Amazon Prime Air

On the other hand, automation of the aerial robot is very important for today's research. Table.7 shows related works in recent years on aerial robot automation. Nowadays, aerial robot is applied for more and more fields, so its automation is more and more perfect. About the UAV, there are more than 300 types in the world. A UAV is the prominent part of a whole system that is necessary to fly the aircraft. Even though there is no pilot physically present in the aircraft, this doesn't mean that it flies by itself autonomously. In many cases, the crew responsible for a UAV is larger than that of a conventional aircraft. The aircraft is controlled from the ground (the Ground Control Station or GCS), so it needs reliable communication links to and from the aircraft, but also to the local Air Traffic Control (ATC) authorities if required (usually when flying higher than 150-200 m above the ground). The GCS provides a working space for a pilot, navigator, instrument operator and usually a mission commander [27].

Automatic system is usually to effectively combine image processing technology, GPS, and sensor technology. In the study of Sivakumar Rathinam et al. [18], they make use of the vision tracking and the GPS tracking technology to automatically detect and track the river by using an UAV. However, in the studies of Casau et al. [28] and Eisenbeiss et al. [29], UAVs are controlled by the Ground Control station. In order to achieve artificial intelligence, more and more researchers care about the field of automatic aerial shoot system.

In addition, some additional researches also exist. For example, the study of monitoring forest fire by Unmanned Aircraft System (UAS) has been done. Unmanned Aircraft System can play an important role for forest fire response. They have been already successfully demonstrated for fire detection, localization and observation [30] [31] [32] [33]. Furthermore, the study of traffic monitoring by Unmanned Aircraft System (UAS) has been carried out [34] [35] [36] [37].

Table.7 Automation of each aerial robot

Place / Type	Automatic	Semi-automatic
FWT	Autonomous Searching and Tracking of a River using an UAV [18]	Transition Control for a fixed-wing Vertical Take-Off and Landing Aircraft [28]
RWT	Small automatic surveillance flying robot [19]	APPLICATIONS OF PHOTOGRAMMETRIC PROCESSING USING AN AUTONOMOUS MODEL HELICOPTER [29]
AST	Developing a Low-Cost Autonomous Indoor Blimp [21]	Advanced Virtual Reality Technologies for Surveillance and Security Applications [22]
BT	Research of Broad Area Monitoring by Simplified Aerial Image Acquisition System [23]	Development of the balloon-cable driven robot for information collection from sky and proposal of the search strategy at a major disaster [24]

Furthermore, from the viewpoint of application, aerial robot can be used for the military and civilian. Table.8 shows examples for the military UAV and the civilian UAV. Because the performance limitations of airship type aerial robot and balloon type, they are not generally used for military purposes. In the study of Yuichiro NISHI [38], he has studied the development of UAV for military in the America. UAV is classified in detail, and the situation of UAV in America is analyzed in detail. In addition, in the study of Campoy et al. [39], they make use of the image processing technology and tracking algorithm to detect and track the feature point of images, such as the Harris Corner detector [40], the SIFT method [41], the Hough Transform method [42], and so on. After that, they can use this feature points to analyze and understand the wide area situation by using an UAV. It can be applied for many fields in our lives.

Table.8 Application of the UAV

Application	Military	Civilian
UAV	Current State of UAV R&D and Deployment in the USA [38]	Computer vision onboard UAVs for civilian tasks [39]

2.2 Tracking Method

To understand wide area situation, we must detect and track the target in any time, so we have to know about the tracking method. Originally, there are many methods for human tracking, such as the Background Subtraction method, the Mean Shift method, the Particle Filter method, and so on. Here, tracking technologies are analyzed from principle, application range, and for our study. Table.9 shows the basic tracking method, such as the Feature Matching method, Bayesian Tracking method, Mean Shift method, Optical Flow method, Template Matching method, and Contour Based method.

- **Method based on contrast analysis:**

This method makes use of the contrast differences between target and background to extract, identify and track the subject. There are some typical methods, such as the Background Subtraction method [44] and the Frame Difference method [45]. However, it is not suitable to track the subject in complex background or changing background. Therefore, it is not suitable to our study.

- **Method based on feature matching:**

This method has to pick up the main feature of the subject at first. And then, this main feature is found in every frame. This process is feature matching [46]. The feature contains geometric shape, subspace feature, contour, feature points, and so on. In our study, because main features are few and environment changes at any time, this method is not suitable for our study.

- **Method of Bayesian Tracking**

This method contains the Kalman Filter method, the Particle Filter method, the Hidden Markov models, and Dynamic Bayesian networks. The principle is as follows.

In order to describe tracking problem, we must definite target state space model like this:

$$\begin{cases} x_k = f_k(x_{k-1}, v_k) \\ y_k = h_k(x_k, w_k) \end{cases} \tag{1}$$

In equation (1), $x_k \in R^n$ is system state at time k, $y_k \in R^n$ is measurement state, and $v_k \in R^n$, $w_k \in R^n$ are system noise and measurement noise respectively. From the Bayesian Estimation, tracking problem is to estimate the value of state variables x_k at time k from measurement information $y_k = \{y_1, y_2, \ldots, y_k\}$, and it is the posterior probability estimates $P(x_k|y_k)$.

Table.9 Tracking methods [43]

	Principle	Application range	For our study
Contrast analysis	Contrast difference between the target and background	Constant background	Background change at any time
Feature matching	Feature extraction → Image matching	Invariant feature	Feature is less
Bayesian tracking	State estimation in time series $P(x_k \mid y_{k-1}) = \int P(x_k \mid x_{k-1}) P(x_{k-1} \mid y_{k-1}) dx_{k-1}$ $P(x_k \mid y_k) = \frac{P(y_k \mid x_k) P(x_k \mid y_{k-1})}{P(y_k \mid y_{k-1})}$	Gaussian and linear problem	PF→Global KF→Prediction
Mean Shift	Color Histogram $M_h(x) = \frac{1}{k} \sum_{x_i \in s_h} (x_i - x)$	Invariant color histogram	Color of head and clothes is invariant
Optical flow	Instantaneous velocity of the pixel $I_x \cdot U_x + I_y \cdot U_y + I_t = 0$ $\vec{U} = (U_x, U_y)^T$	Motion estimation problem in two-dimensional	Human moves together with the flying robot
Template matching	Template image → Image comparison → similar region	No deformation	Body deformation
Contour-based (snakes)	Initial contour →object contour	Not complex shapes	Clothes extraction

Due to the Bayes Theorem [47] [48], x_k obeys an order Markov process, and measurement sequence y_k is independent for each other, then the prior distribution of initial state x_0 is $P(x_0 \mid y_0) = P(x_0)$. Therefore, the state prediction equation is

$$P(x_k \mid y_{k-1}) = \int P(x_k \mid x_{k-1}) P(x_{k-1} \mid y_{k-1}) dx_{k-1} \tag{2}$$

In addition, the state update equation is

$$P(x_k \mid y_k) = \frac{P(y_k \mid x_k) P(x_k \mid y_{k-1})}{P(y_k \mid y_{k-1})} \tag{3}$$

In equation (3),

$$P(y_k \mid y_{k-1}) = \int P(y_k \mid x_k) P(x_k \mid y_{k-1}) dx_k \tag{4}$$

After getting posterior probability estimates $P(x_k|y_k)$, the state value of the subject can be calculated by using some criterions, such as the Maximum likelihood estimation, Minimum mean square error estimation, Maximum posteriori estimation, and so on.

The Kalman Filter (KF) [49] is a recursive algorithm of noisy linear dynamic System state estimation, and it is a process of continuously predicting and correcting. When the system state model and observation model are linear and obeying Gaussian distribution and noise obeys Gaussian distribution, Kalman Filter is the optimal filter. However, there are some limits on it. It cannot solve the nonlinear and non-Gaussian problems. In our study, it can be used to predict the moving direction of the subject. The Particle Filter (PF) has been introduced in previous chapter. Except KF and PF, the Hidden Markov models (HMMs), and Dynamic Bayesian networks (DBNs) are also important visual tracking method. In addition, the differences between KF, PF, HMMs and DBNs are shown in Table.10.

Table.10 Comparisons between KF, PF, HMMs and DBNs

Algorithm	Scope	State representation	Topology
KF	Linear, Gaussian distribution	A random variable(vector)	Fixed
PF	Nonlinear, Arbitrary distribution	A random variable(vector)	Fixed
HMMs	Nonlinear, Arbitrary distribution	A random variable(vector)	Fixed
DBNs	Nonlinear, Arbitrary distribution	Collection of random variables (vector)	Changeable

■ **Mean Shift method**

This method use color histogram as the matching feature. About the Mean Shift tracking algorithm [50], data points move toward to the Mean Shift vector direction, and eventually converge to extreme point of a probability density function. In Mean Shift algorithm, the similarity function is used to describe the similarity between the kernel histogram of the target template and the kernel histogram of candidate region, and the Bhattacharyya coefficient is used. Therefore, this method convert tracking problem to Mean Shift pattern matching problem. There are many merits on Mean Shift method. For example, feature histogram is enough to determine the target location, sufficiently robust, and insensitive to other motions. Also, it can avoid complexity modeling of target shape, appearance, or motion, and can establish the relationship between the statistical measures of similarity and successive optimization. However, this method is not suitable to rotation and scale movement.

■ **Method based on motion detection**

It is to find the target area by detecting different motion of the subject and background in sequence images, and to achieve target tracking. The Optical Flow algorithm [51] is a typical method based on motion detection. Optical flow [52] is the instantaneous velocity of pixel motion of moving objects in the image plane. Optical flow vector is the grayscale instantaneous rate of change of image plane coordinates. About the optical flow calculation, it makes use of the changes in the time domain and correlation of pixel gray level distribution in image sequence to determine the movement of each pixel position, and to research the relationship between grayscale changes of image in time and the object structure and motion in the scene. Optical flow constraint equation (shown in equation (5)) is introduced by contacting two-dimensional velocity field with the grayscale. Then, the basic algorithm of optical flow calculation can be obtained.

$$I_x \cdot U_x + I_y \cdot U_y + I_t = 0 \tag{5}$$

$$\vec{U} = (U_x, U_y)^T \tag{6}$$

In equation (5), I_x and I_y are light intensity changes of image plane relative to x, y-axis direction, and I_t is the intensity change of the same pixel in adjacent moment. $\vec{U}$ is optical flow.

Optical flow method can be good for two-dimensional motion estimation. But, when the target and background both moves, this method may be failed.

In our study, because the subject and background both changes in every frame and features are few, the Mean Shift method is used for human tracking. In addition, the Kalman Filter is used to predict the moving direction and position of the subject.

■ **Template matching method**

At first, template image has to be prepared. Then, it is to find the similar part with the template image in whole image. Methods of similarity calculation are the SSD (Sum of Squared Difference) and the SAD (Sum of Absolute Difference) [53].

<1> SSD method:

$$SSD = \sum_{x=0}^{X-1}\sum_{y=0}^{Y-1}(I(x,y) - T(x,y))^2 \tag{7}$$

<2> SAD method:

$$SSD = \sum_{x=0}^{X-1}\sum_{y=0}^{Y-1}| I(x,y) - T(x,y) | \tag{8}$$

In equation (7) and (8), I(x,y) is the brightness value of the image, and T(x,y) is the brightness value of the template image. This method is suitable to track the subject that is no deformation. In our study, posture of body is change at any time, so this method is not suitable for our study.

■ **Contour based method**

Energy minimizing active contour models for representing controls in an image in a form that allow interaction, with high level processes have been proposed by Kass et al. in 1987. Snakes are active contour models: they lock onto nearby edges, localizing them accurately. Moreover, snakes provide unified account of a number of visual problems by incorporating additional energy terms [54]. By this method, body contour can be extracted by setting the suitable initial contour in our study.

In summary, several researches for aerial robot system have been executed, but problems of limitations for place, incomplete automation and applications still exist. For example, some aerial robots only can be used in indoors environment, because its weight is light and its size is small. Conversely, some aerial robots only can be used in outdoors environment, because its size is big and control accuracy is low. Many reasons cause it cannot use for not only in indoors but also in outdoors. In addition, many aerial robots are controlled by the Ground Control Station, so they cannot track the subject until it getting command send by the user. Furthermore, when aerial robot is used to track the target, there are two ways, vision tracking and GPS tracking. Sometimes, they use a combination of both or plus some sensors. In this way, the cost would increase. Moreover, because the speed of airship type aerial robot is slow and balloon type robot has not tracking capability, their applications are fewer than the UAVs. Also, because airship and balloon type robot are susceptible to outside influence, when there is strong wind, it cannot fly in normal.

Therefore, previous aerial robot system has three disadvantages: limitations for place, incomplete automation, and high cost. In order to solve these problems, we proposed a new visual control method to achieve moving direction automatic detection and automatic human tracking for flying robot. The method consists of integrating multiple modules to detect moving direction exactly and performing autonomous control of flying robot.

About the tracking algorithm, because aerial robot and the target moves together, we must choose methods are not affected by changes in background. Therefore, we choose the Hough Transform method to detect the head part at first, and then use the Active Contour Model (ACM) method to extract the body contour to calculate the clothes information. After that, the subject is tracked by integrating the Mean Shift method and the Particle Filter method. Also, we would use the Kalman Filter method to predict the position and direction of the target.

Chapter 3
Proposed Method

In our study, we propose a new visual control method to achieve moving direction automatic detection and automatic human tracking for flying robot. The method consists in integrating multiple modules to detect the moving direction exactly and performing autonomous control of the flying robot. These modules are not simply combined, but integrated each other effectively. From aerial image, we can see the shape and color of head is invariant within a certain period of time (maybe wear a hat or take off a cap). Then, the color of cloth is invariant (less likely to change clothes). Other features are changed at any time. Therefore, head part and clothes part as important features are used to make sure the position and direction of the subject. At first, the Hough Transform method is used to detect head part of the subject. Then, clothes part of the subject is extracted by using the Active Contour Model (ACM). Because color information of head part and clothes part can be used only, the subject is tracked by using color histogram of head part and clothes part. Here, the Mean Shift method and the Particle Filter method have been used. Then, we calculate the moving direction and position of the subject, and make use of the Kalman Filter method to correct the information of moving direction and position calculated and predict the direction and position when the subject is lost by using time series statistics data. Such effective integration (or this workflow) is unique, and it is the novelty of our study.

3.1 Aerial Shoot System

In our study, aerial shoot system for wide area situation recognition is proposed, and it is divided into two parts. One is image processing thread, and another is flying robot control thread. Fig.6 shows the overall flow of aerial shoot system.

In image processing part, missing times, current position and moving direction of the subject are calculated. Then, the data is processed in middle part. According to the position, moving direction and missing times, state value will be changed. In addition, state value is fed back to flying robot control part. Then, flying robot is controlled by state value.

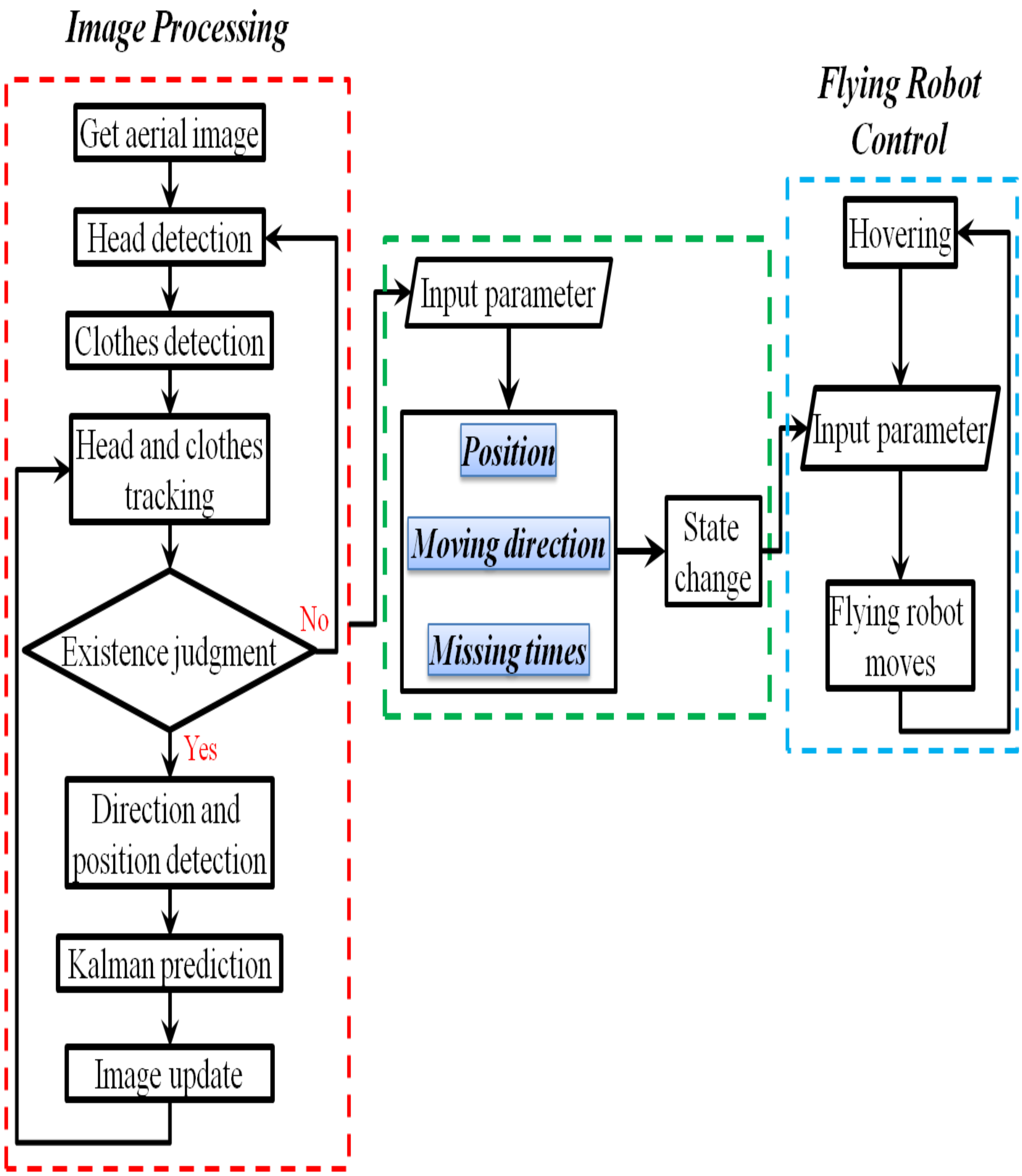

Fig.6 Overall flow of aerial shoot system [55]

3.1.1 Image Processing Thread

Fig.7 shows the flowchart of image processing thread. In image processing part, firstly, the aerial robot is used to capture aerial image. Then, head part of the subject is detected according to its shape and brightness histogram. After that, ACM method is utilized to get body contour and calculate the Hue histogram of clothes part around the head. In addition, the Mean Shift method is used to track the head part and clothes part respectively. Furthermore, we have to judge the existence of the subject. If the subject exists, position and direction of the subject are detected and predicted. If the subject is lost, the target will be detected again.

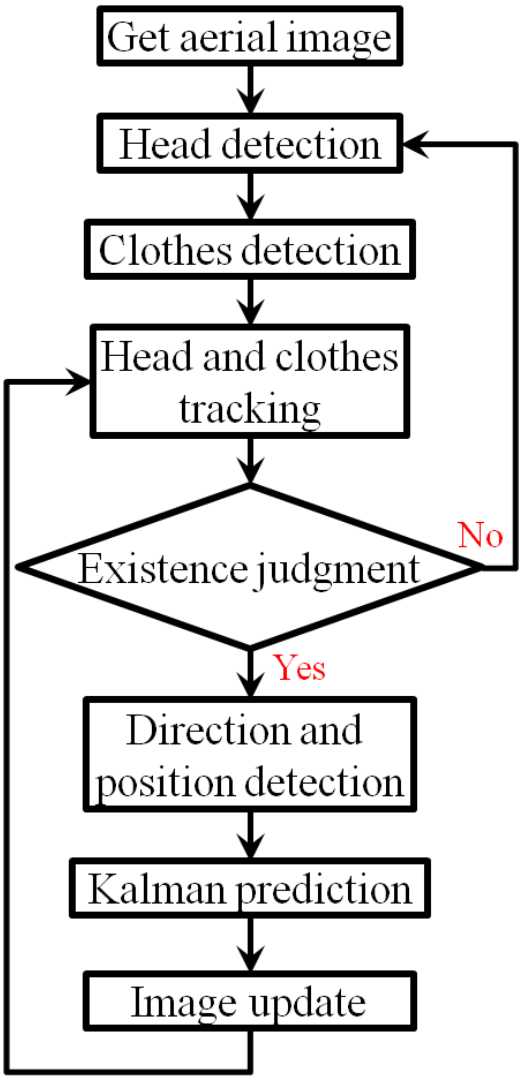

Fig. 7 Flowchart of image processing thread

3.1.1.1 Head Detection

In our study, head part is detected due to the shape of head and color histogram. When the subject is not wearing the cap or hat, head part detection is carried out as shown in Fig.8. Above all, aerial image is captured, and edge processing is performed. Then, we use the Hough Transform method to detect round objects from edge image. After that, the brightness histogram of this round object is calculated. In addition, because head part is black in general, we would calculate the similarity of this round object and black color by comparing the brightness histogram of each other. If it is near to black, we can make sure this round object is head of the subject. If not, it is to search other round object again.

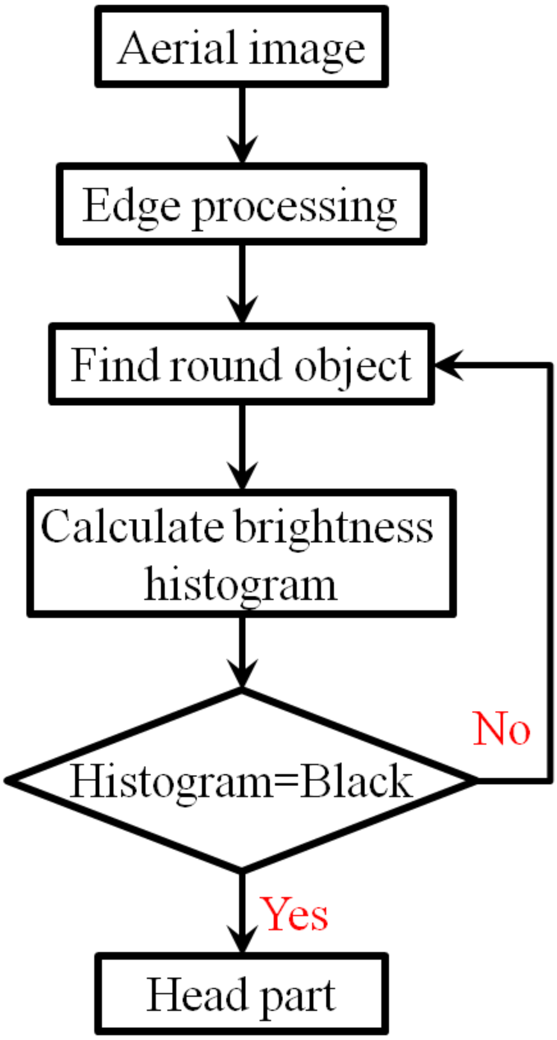

Fig. 8 Head detection when target is not wearing cap or hat

When the target is wearing cap or hat, head detection is carried out according to the Fig.9. At first, the preliminary experiment is necessary. About this preliminary experiment, we will talk about it in detail in chapter four. In preliminary experiment, the hue (or brightness) histogram of cap (or hat) is learned. Then, cap or hat can be detected the same with above said.

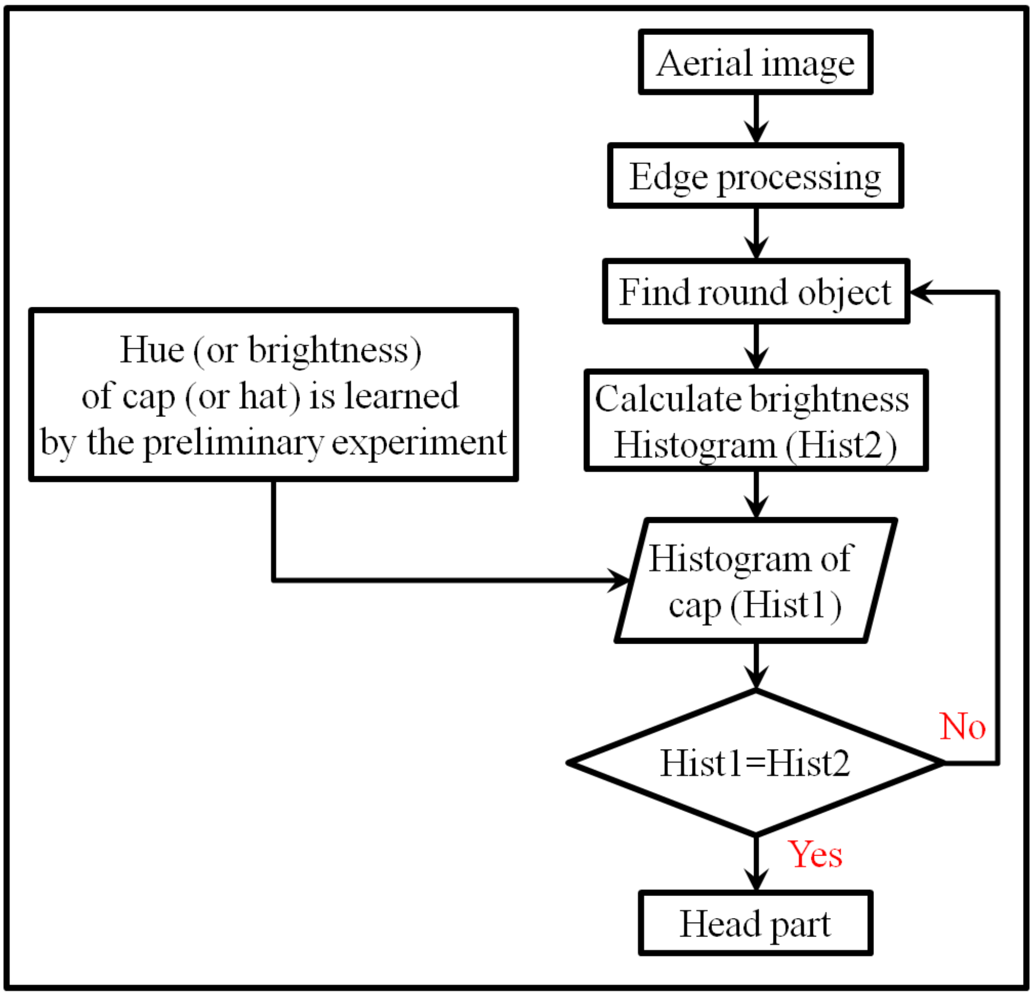

Fig. 9 Head detection when target is wearing cap or hat

The Hough Transform method is a basic method of image processing. The basic principle of the Hough Transform method [56] [57] is shown in Fig.10. Originally, the Hough Transform method is for line detection.

Line equation is “y=kx+b”, k is the slope, b is the Intercept. For example, there are four points in this line in the image space. When image space is transformed into the parameter space, these four points correspond to four lines in the parameter space. Also, these four lines intersect at one point, and this point is the line we want to seek in image space. By this theory, the Hough Transform method can be used for the line detection, curve detection, circle detection, ellipse detection, and so on. About the circle detection, it is divided into two cases. One is that r is known, and another is that r is unknown. If r is known, circle can be detected in two dimensions in parameter space. On the contrary, if r is unknown, circle will be detected in three dimensions as shown in the lower right corner of Fig.10. In fact, Hough Transform method can detect arbitrary curve which equation is known, and how to choose the parameter space is very important. For example, equation of circle is $(x-a)^2+(y-b)^2=r^2$, so when we use Hough Transform method to detect a circle, we can choose a space the same with original space is as the parameter space. Therefore, a circle in original space corresponds to a point in parameter space. The parameter of points on the same circle in original space is same. In other words, a is same to b. Therefore, their corresponding circles in parameter space pass the same point (a, b). After changing all points in original space to them in parameter space, according to degree of aggregation of points in parameter space, we can judge whether there are similar circular pattern. If so, this parameter is the circle parameters.

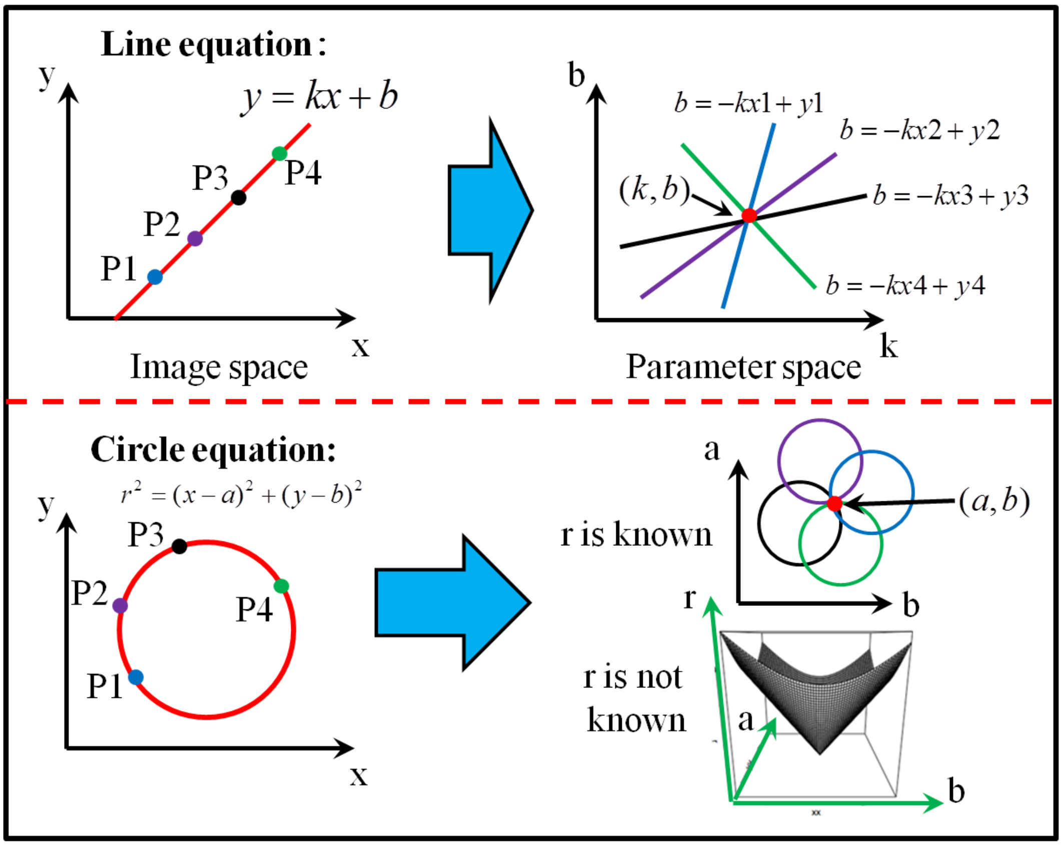

Fig. 10 Principle of the Hough Transform method

3.1.1.2 Clothes Detection

When head part has been detected in previous chapter, clothes part is detected by using the Active Contour Model (ACM), and hue histogram of clothes part can be calculated. Then, the regions of the clothes can be decided by finding the most appropriate rectangle around the clothes. The flowchart of clothes detection is shown in Fig.11.

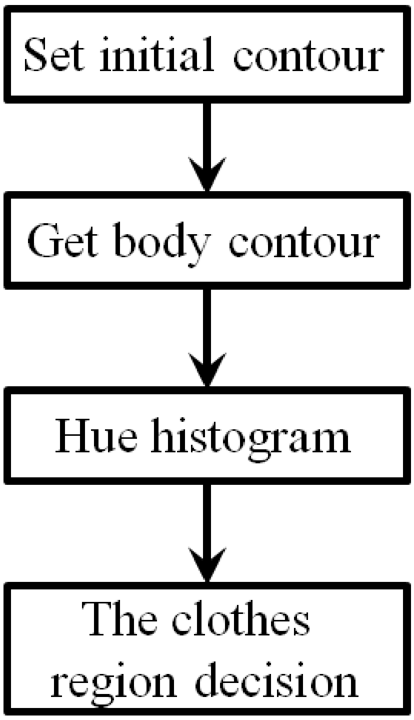

Fig. 11 The clothes detection from aerial image

Above all, initial contour is set by the center of head and thickness of the body. Then, body contour is got by initial contour converging. After that, the Hue histogram of clothes can be calculated from body contour. In addition, the region of clothes can be determined by finding the most appropriate rectangle around the body part.

The ACM (Active Contour Model) method [58] [59] is an important method of image processing. This method makes use of the energy minimizing theory to extract body contour. The energy equation is shown as follows.

$$E_{snake} = \int_0^1 E_{snake}(v(s))ds = \int_0^1 \{E_{\text{int}}(v(s)) + E_{image}(v(s)) + E_{con}(v(s))\}ds \tag{9}$$

In equation (9), E_{int} is the internal energy, E_{image} is the image energy, and E_{con} is the external energy. The s is parameter of route.

The calculation formula of internal energy is shown in equation (10). In this equation, α is the coefficient on the contraction of the contour, and β is the coefficient for bending. From the Fig.12 we can see, the first item of equation (10) represents the length of the closed curve, and the second item represents the difference vector of the displacement vector between adjacent.

$$E_{\text{int}}(v(s)) = \int_0^1 (\alpha(s)\,|\,v_s(s)\,|^2 + \beta(s)\,|\,v_{ss}(s)\,|^2)ds \tag{10}$$

The calculation formula of image energy is shown in equation (11). In this equation, I is brightness of the image. E_{image} represents edge part of image.

$$E_{image}(v(s)) = -|\,\nabla I(v(s))^2\,| \tag{11}$$

By using the ACM method, body contour of the subject can be extracted, and we can get the Hue histogram of clothes part from the body contour.

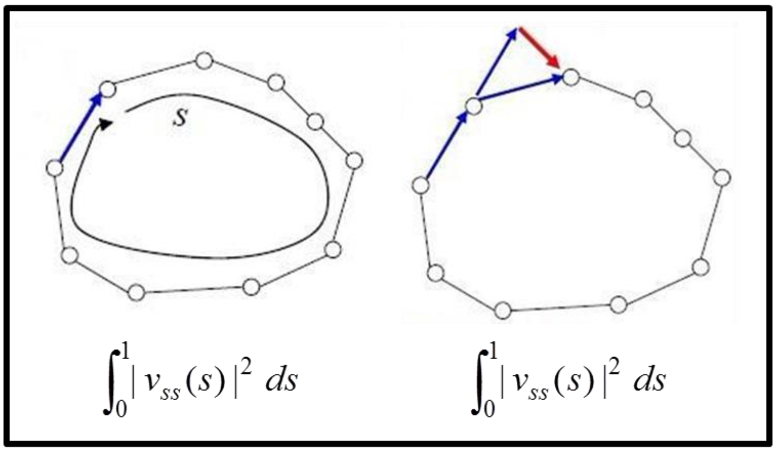

Fig. 12 Internal energy of ACM method

3.1.1.3 Human Tracking

It is assumed that there are N sample points in d-dimensional space R^d, Therefore, the Mean Shift vector at point x can be expressed as follows. S_h is a high-dimensional ball area, the radius is h while k is the number of samples in S_h area.

$$M_h(x) = \frac{1}{k} \sum_{x_i \in S_h} (x_i - x) \tag{12}$$

In equation (12), the $(x_i\text{-}x)$ is offset vector between x_i and x, and $M_h(x)$ is average offset of all samples in S_h area. Intuitively, if sample x_i is determined from a probability density function f(x), because non-zero probability density gradient point to the direction of density probability increasing, Mean Shift vector $M_h(x)$ points to direction of probability density gradient [60] (as shown in Fig. 13).

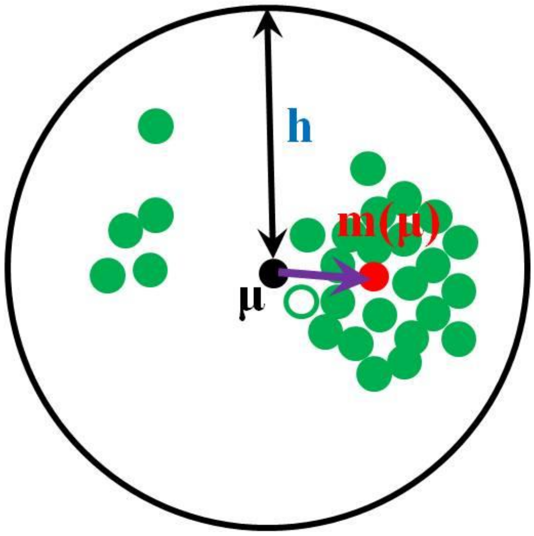

Fig. 13 The Mean Shift algorithm

In our study, the hue information of clothes is as the sample S = $\{x_i\}$. The T(μ) indicates the number of sample points in the tracking area, and μ is the center of tracking area. The m(μ) is the Mean Shift vector.

$$m(\mu) = \frac{1}{|T(\mu)|} \sum_{x \in T(\mu)} x_i \tag{13}$$

Iteration step s(μ) is the difference vector between m(μ) and μ.

$$s(\mu) = m(\mu) - \mu \tag{14}$$

In addition, by μ converging to m(μ) repeatedly, current position μ will converge to mode point near the initial value. By this way, head part and clothes part of the target can be tracked respectively.

The above said is only one person in the scene. However, this situation is not the universally case.

3.1.1.4 Existence Judgment

In previous chapter, we used the Mean Shift method to track the head part and clothes part. Then, it is necessary to judge whether the subject is lost. Because head part and clothes part are tracked respectively, the subject can be decided by using color of head and clothes. If there is little color of head in head region or little color of clothes in clothes region, we can think that the subject is lost. On the contrary, the subject exists.

In other words, we can explain it by using equation (15) and (16). In equation (15), x_h is brightness value of image, V is brightness value of head part, and TH_h is threshold of brightness value. If equation (15) is established, head part exists. In equation (16), x_c is Hue value of image, H is Hue value of clothes part, and TH_c is threshold of Hue value. If equation (16) is established, clothes part exists.

$$x_h - V < TH_h \tag{15}$$

$$x_c - H < TH_c \tag{16}$$

Therefore, if equation (15) and (16) are both established, we can decide the subject exists. On the contrary, if any equation is not established, we will decide the target does not exist.

Above said is the case for only one people in the scene. When there are two, three or multiple persons in the scene, then how to judge who particular people is, we have to consider it.

If the color of clothes part is not same, it is easy to distinguish them by using color information. Here, we only tell about the case of multiple people wearing the similar color clothes. As shown in Fig.14, in first frame, only the particular person (p1) is in the scene. In second frame, new person (p2) enters. At this time, because the distant between the particular people in the first frame and the second frame is near, we can judge who the particular people is in the second frame. Even if they are very close, because the moving direction of each other is different, we also can use direction information to separate them. The processing flow is shown in Fig.15.

A particular person is judged by using the Labeling theory [61] as shown in Fig.15. According to color of head and clothes, gray image will be got. In other words, pixel value which is similar with color of head and clothes is set to 255 (white), and the others will be set to 0 (black). Then, every white part is labeled with the label number. In first frame, because there is only the particular person in the scene, we can get the labeling image of particular people. Capturing next frame image, image is labeled in same method as before. After that, it is to compare each label part in the second frame to the label part in the first frame, and calculate each distant. If p1 is far from p2, we will judge which is bigger between d1 and d2. If d1 is smaller one, p1 is the particular people. If p1 is near to p2, we will judge whose moving direction is same with direction 1. If the direction of p1 is same, p1 is the particular people.

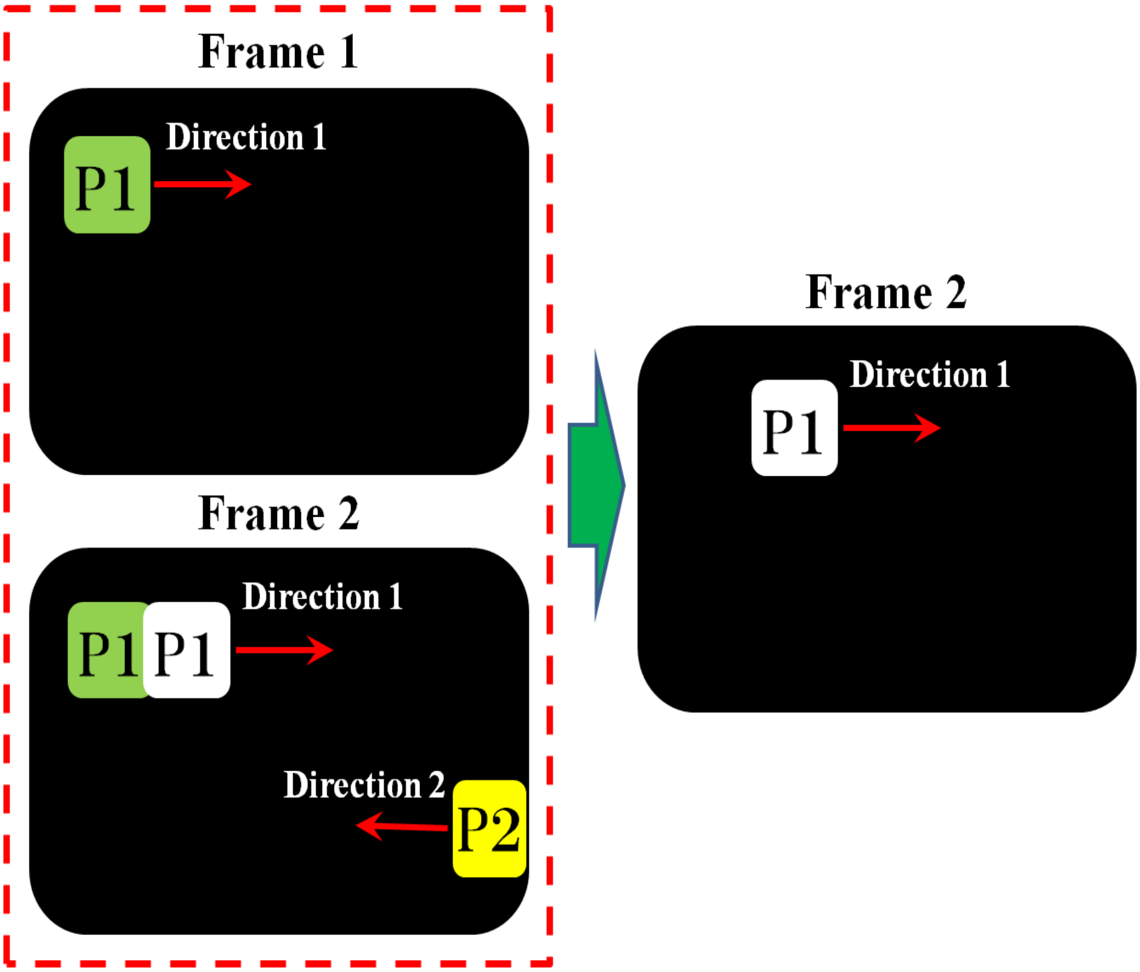

Fig. 14 Judge particular people when multiple persons exist

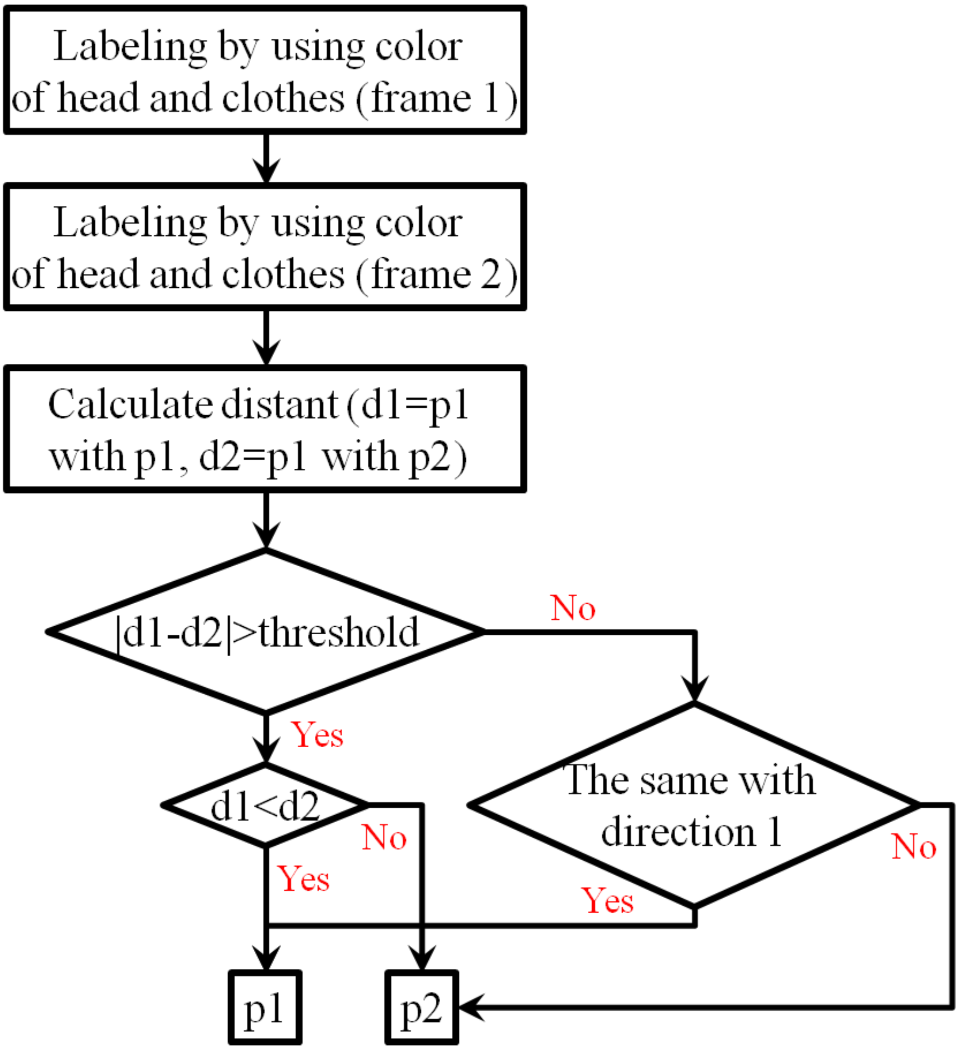

Fig. 15 Processing flow of particular people judgment

3.1.1.5 Target Redetection

Once the subject is lost judged in previous chapter, we will detect the head part and clothes part of the target. Head part is to be detected by using the Hough Transform method, and the flowchart is same with above. However, clothes part is to be detected by using the Particle Filter method.

Clothes part can be rediscovered due to hue histogram of clothes by using the Particle Filter algorithm. It is necessary to build the state transition function and the likelihood function [62]. The states are position (X_t, Y_t) and velocity (U_t, V_t) shown in Fig.16.

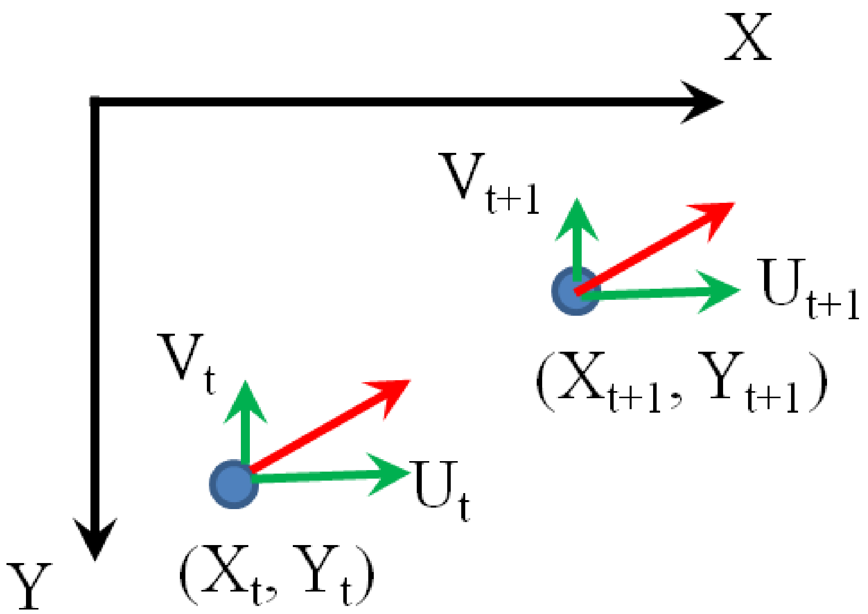

Fig. 16 The state vector of particle

The coordinate (X_t, Y_t) is the position of particle at time t, U_t is x-direction velocity component, and V_t is y-direction velocity component at time t. The state transition function is shown in equation (17).

$$S_{t+1} = F \times S_t \tag{17}$$

F is the transition matrix of n×n, and it is used to associate the state at time t and the state at time t+1. It is indicated in equation (18). DynamMatr [] is dynamics of the state vector.

$$F = \begin{pmatrix} \text{DynamMatr[0]} & \text{DynamMatr[1]} & \text{DynamMatr[2]} & \text{DynamMatr[3]} \\ \text{DynamMatr[4]} & \text{DynamMatr[5]} & \text{DynamMatr[6]} & \text{DynamMatr[7]} \\ \text{DynamMatr[8]} & \text{DynamMatr[9]} & \text{DynamMatr[10]} & \text{DynamMatr[11]} \\ \text{DynamMatr[12]} & \text{DynamMatr[13]} & \text{DynamMatr[14]} & \text{DynamMatr[15]} \end{pmatrix} \tag{18}$$

In our program, state transition is assumed as linear prediction model (uniform linear motion). Therefore, F can be simplified to

$$F = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Furthermore, the state equation can be indicated as shown in equation (19).

$$\begin{pmatrix} X_{t+1} \\ Y_{t+1} \\ U_{t+1} \\ V_{t+1} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X_t \\ Y_t \\ U_t \\ V_t \end{pmatrix} \tag{19}$$

In addition, likelihood of pixel value in particle location is calculated according to equation (21), and the Euclidean distance (Ed) with observation value is calculated in equation (20). In equation (20), the value of b, g and r is observation value, and the value of b_i, g_i and r_i is pixel value. In equation (21), L(Ed) is the likelihood function, σ is variance for normal distribution.

$$Ed = \sqrt{(b-b_i)^2 + (g-g_i)^2 + (r-r_i)^2} \tag{20}$$

$$L(Ed) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{Ed^2}{2\sigma^2}) \tag{21}$$

In order to calculate σ, color histograms of clothes and head are necessary. The calculation process is shown as follows.

Theoretically, the equation of the normal distribution is

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(x-\mu)^2}{2\sigma^2}) \tag{22}$$

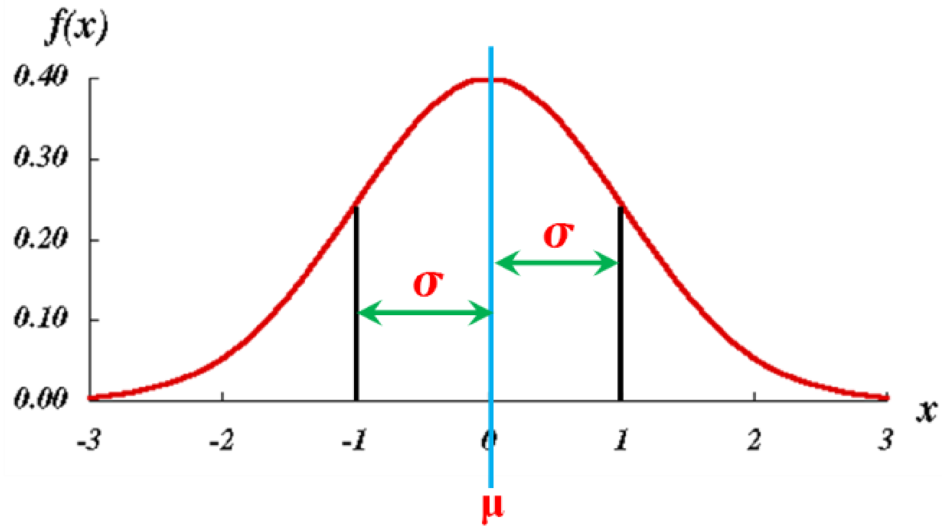

Fig. 17 Normal distribution analysis

In equation (22), μ is mathematical expectation, and σ is variance (shown in Fig.17). Then, μ and σ can be calculated in equation (23) and (24). In addition, N is the number of samples, and X_i is sample.

$$\mu = \sum X_i / N \tag{23}$$

$$\sigma = \sqrt{\sum (X_i - \mu)^2 / (N-1)} \tag{24}$$

When the head part is determined, the color histogram of clothes can be calculated around the head. From color histogram, we can choose the maximum probability partial, and approximate it as the normal distribution function by μ and σ calculated previously. Furthermore, particles with low likelihood are deleted. In the remaining particles, by calculating particle distribution, human area can be determined. Furthermore, particles with low likelihood are deleted. In the remaining particles, by calculating particle distribution, human area can be determined.

3.1.1.6 Direction and Position Detection

Just now, it is said that head part and clothes part are tracked respectively by the Mean Shift method. Therefore, we use the relationship between head and clothes to detect moving direction and position of the subject. Fig.18 shows the process of direction and position detection, and moving direction is expressed as the vertical direction of bounding rectangle.

Above all, head part and clothes part are tracked by the Mean Shift method in previous chapter. Then, the optimal bounding rectangle in clothes part has to be calculated, and it is shown in the second image of Fig.18. In addition, according to the relationship between head part and clothes part, direction can be decided, and it is shown in the fourth image of Fig.18. Furthermore, Judgment method of moving direction is shown in Fig.19.

In Fig.19, green point represents the center of gravity of head part, and blue line represents the connection line between two endpoints of the shoulder. Then, relationship between green point and blue line can be judged. The coordinate of green point is (x_h, y_h), and the coordinates of two endpoints of blue line are (x_l, y_l) and (x_r, y_r) respectively. Line equation is

$$y = kx + b \tag{25}$$

According to line equation, about blue line, k and b can be calculated.

$$k = \frac{y_r - y_l}{x_r - x_l} \tag{26}$$

$$b = y_r - kx_r \quad \text{or} \quad b = y_l - kx_l \tag{27}$$

Therefore, if

$$y_h > kx_h + b \tag{28}$$

Head is below the shoulder.
If

$$y_h < kx_h + b \tag{29}$$

Head is on the top of the shoulder.
Therefore, moving direction can be decided by the red arrows shown in Fig.19.

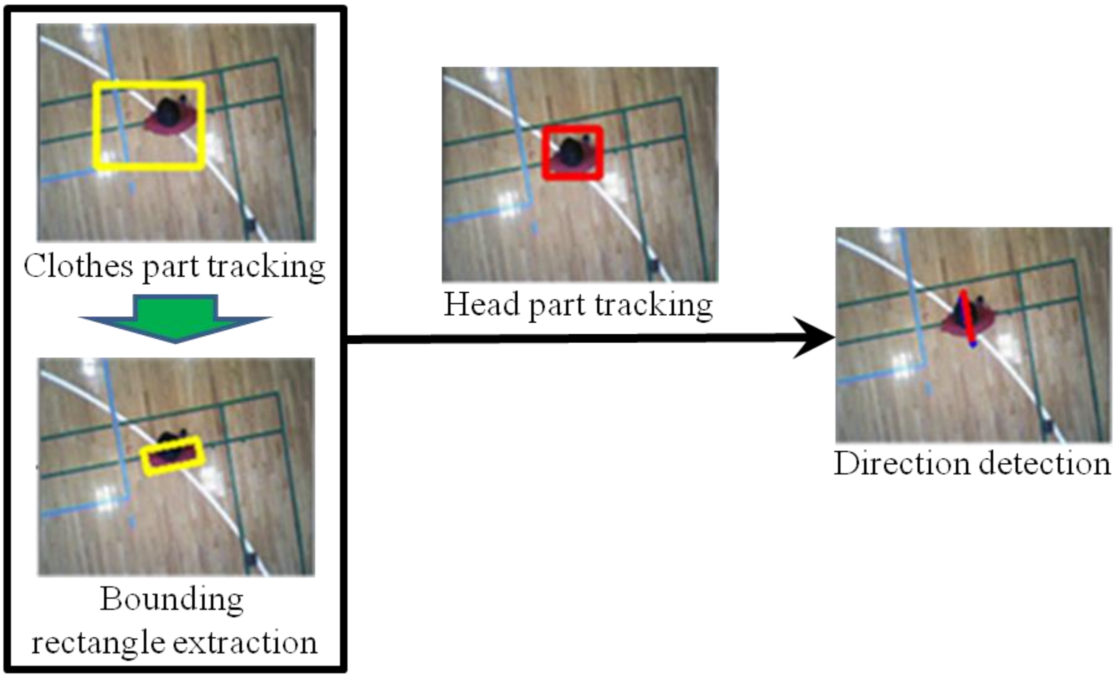

Fig. 18 Process of direction and position detection

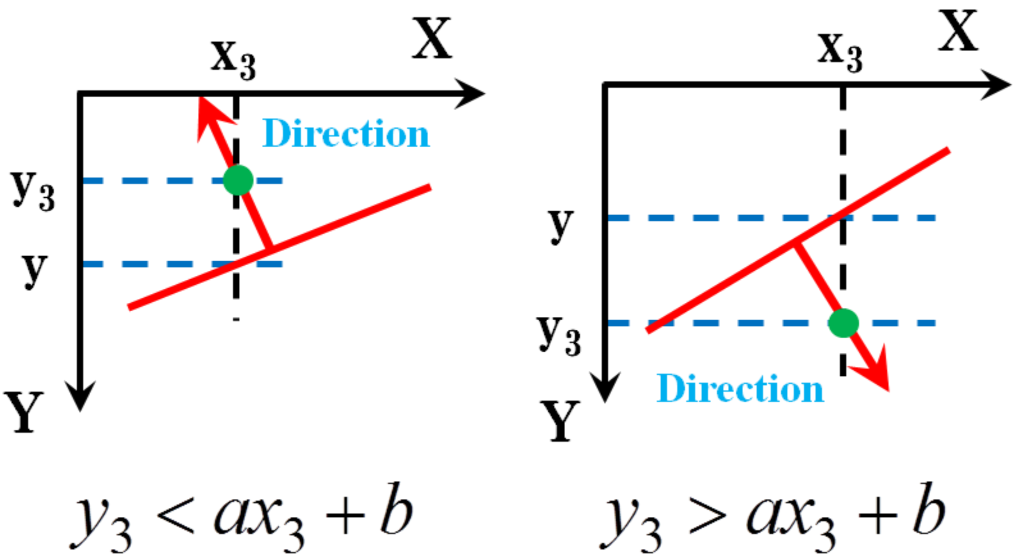

Fig. 19 Judgment method of moving direction

3.1.1.7 Kalman Filter Prediction

Generally, there are some deviations caused by noise perhaps in the process of calculating the moving direction and position. Therefore, we must use some algorithm to correct these deviations. In our study, to solve this problem, the Kalman Filter is utilized to predict the correct value of moving direction [63]. About Kalman Filter method, the state model of the system is designed, and state equation is shown in equation (30).

$$S_{k+1} = A \times S_k \tag{30}$$

$$\begin{pmatrix} x_{k+1} \\ y_{k+1} \\ dx \\ dy \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_k \\ y_k \\ dx \\ dy \end{pmatrix}$$

In equation (30), S is the state matrix, k is the time, and A is the transition matrix. It is to predict states (position) at time k+1 from states at time k. in other words, S_k is the priori state estimate, and S_{k+1} is the posteriori state estimate. Then, the posteriori state estimate M_{k+1} can be predicted (shown in equation 31).

$$M_{k+1} = S_{k+1} + K(Z_{k+1} + H \times S_{k+1}) \tag{31}$$

Z is the observation values matrix, H is observation matrix, and K is the Kalman gain. Then, by constantly updating S_{k+1} and K, current state (M_{k+1}) can be predicted continuously. K can be calculated as shown in equation (32).

$$K = \overline{P}H^T(H\overline{P}H^T + R)^{-1} \tag{32}$$

P is a posteriori error covariance, and R is covariance of the observation error.

In our study, states are position of the subject (x, y) and moving direction (θ). The state equations are designed as follows. Equation (33) shows states are predicted at k+1 time by using average values of previous 5 frames (from k-4 to k).

$$S_{k+1} = A \times S_{avg} \tag{33}$$

$$\begin{pmatrix} x_{k+1} \\ y_{k+1} \\ \theta_{k+1} \\ dx \\ dy \\ d\theta \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_k \\ y_k \\ \theta_k \\ d_{xavg} \\ d_{yavg} \\ d_{\theta avg} \end{pmatrix}$$

In equation (33), S_{avg} is the average value of 5 frames, k is the time, and A is the transition matrix. Then, the average change of position (d_{xavg}, d_{yavg}) and direction ($d_{\theta avg}$) of previous 5 frames can be calculated.

$$M_{k+1} = S_{k+1} + K(Z_{k+1} + H \times S_{k+1}) \tag{34}$$

$$H = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}$$

From equation (34), by constantly updating S_{k+1} and K, current state (M_{k+1}) can be predicted from Z_{k+1} (observation value at current frame) continuously. H is observation matrix. P is a posteriori error covariance and R is covariance of the observation error. The S_{avg} is average value of n frame, and the Z_{k+1} is observation value of current frame. The prediction process is shown in Fig.20.

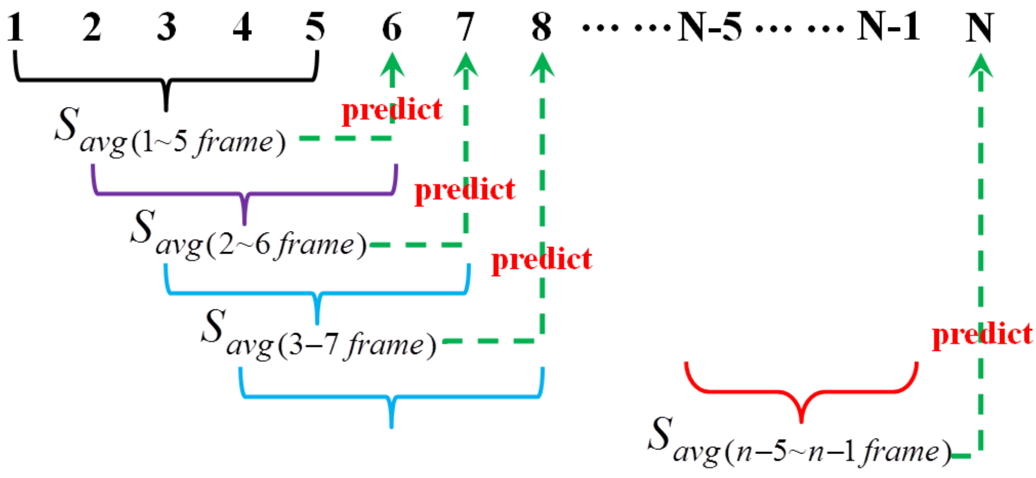

Fig. 20 Prediction process of the Kalman Filter

3.1.2 Flying Robot Control

This part is very important for our study. In this part, flying robot AR.Drone is controlled by using the position and moving direction calculated in previous chapter.

Fig.21 shows the flowchart of flying robot control. At the present stage, in order to verify the effectiveness of robot vision part, AR.Drone is controlled manually. From now on, automatic control of AR.Drone will be carried out. In image processing part, position of the subject (P_x, P_y) and moving direction (θ) are calculated. N is the missing times, and (x_0, y_0) is the coordinate of image center. Then, the data is utilized to control AR.Drone automatically.

At first, AR.Drone is hovering in the air. When the subject enters, the distance of (P_x, P_y) to (x_0, y_0) is calculated. If distance is less than the threshold value (Th_p), it is to check whether moving direction (θ) is more than threshold value (Th_θ). If θ is less than Th_θ, AR.Drone is hovering. Otherwise, AR.Drone is to rotate. In addition, if distance is more than Th_p, it is to make (P_x, P_y) and (x_0, y_0) consistent by AR.Drone moving. Furthermore, when the subject is lost, the missing times N is to be accumulating. If N is less that threshold value (Th_N), AR.Drone is to go forward due to position predicted by Kalman Filter, and search the subject again. But if N is more than Th_N, AR.Drone is to rise and search people once again. If person does not exist, AR.Drone is to be landing.

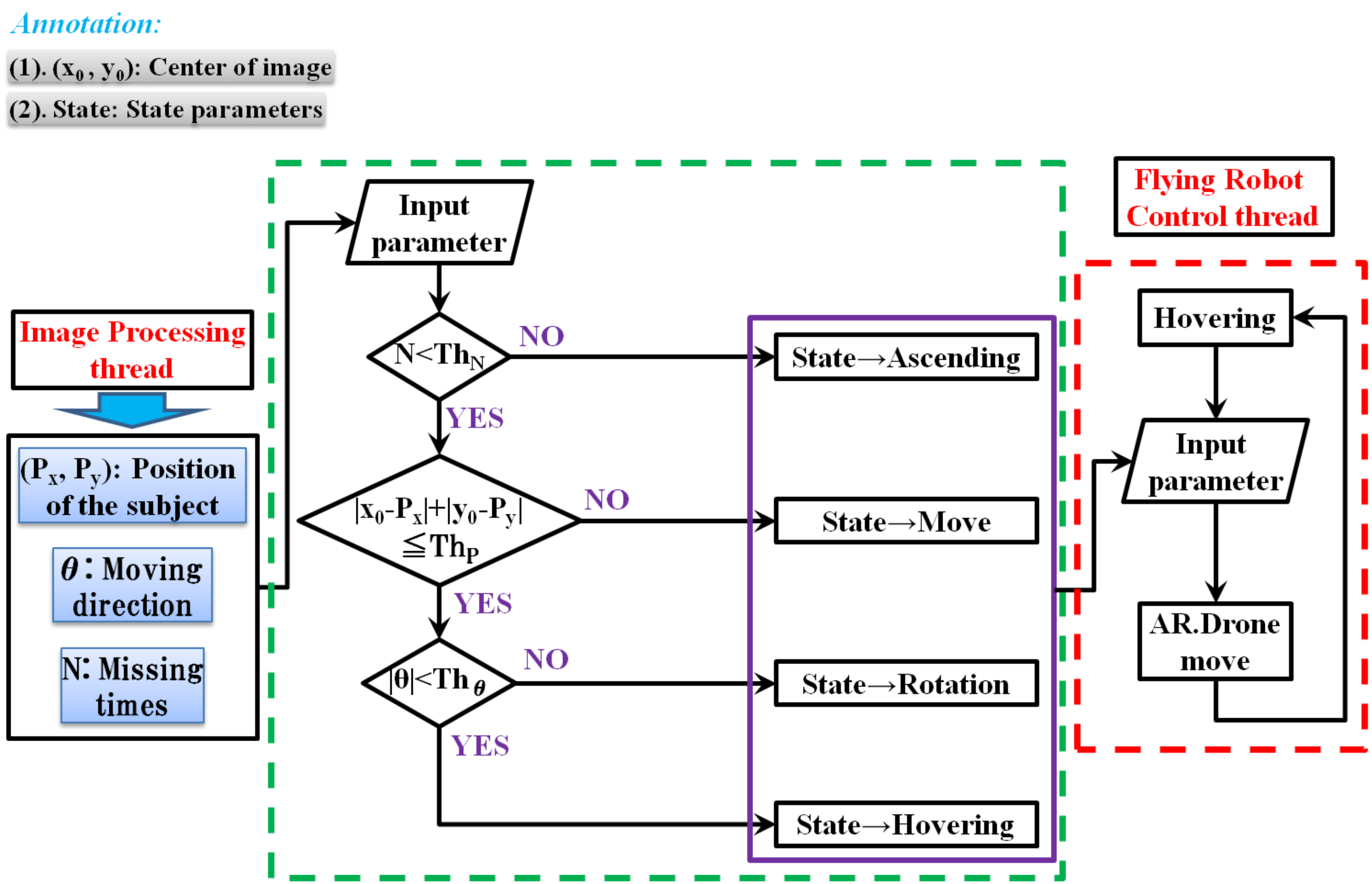

Fig. 21 Flying robot control thread [55]

AR.Drone is a quadrotor. In Fig.22, the mechanical structure comprises four rotors attached to the four ends of a crossing to which the battery and the RF hardware are attached. Each pair of opposite rotors is turning the same way. One pair is turning clockwise and the other anti-clockwise [64].

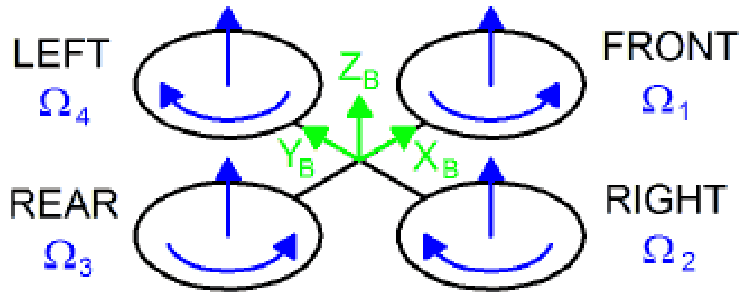

Fig. 22 AR.Drone control

As shown in Fig.23, varying left and right rotors speeds the opposite way yields roll movement. This allows to go forth and back. Varying front and rear rotors speeds the opposite way yields pitch movement. Varying each rotor pair speed the opposite way yields yaw movement. This allows turning left and right. Manoeuvres are obtained by changing pitch, roll and yaw angles of the AR.Drone.

About (a) of Fig.23, when four propellers accelerate together, it will cause an upward force, and AR.Drone will go up.

In (b) of Fig.23, when left propeller accelerates and right propeller decelerate, but front propeller and rear propeller do not change, it will cause a leftward force, and AR.Drone will move to left.

In (c) of Fig.23, when front propeller accelerates and rear propeller decelerate, but left propeller and right propeller do not change, it will cause a forward force, and AR.Drone will move to front.

In (d) of Fig.23, when left propeller and right propeller accelerates, and front propeller and rear propeller decelerate, it will cause a rotational force, and AR.Drone will do a rotary movement.

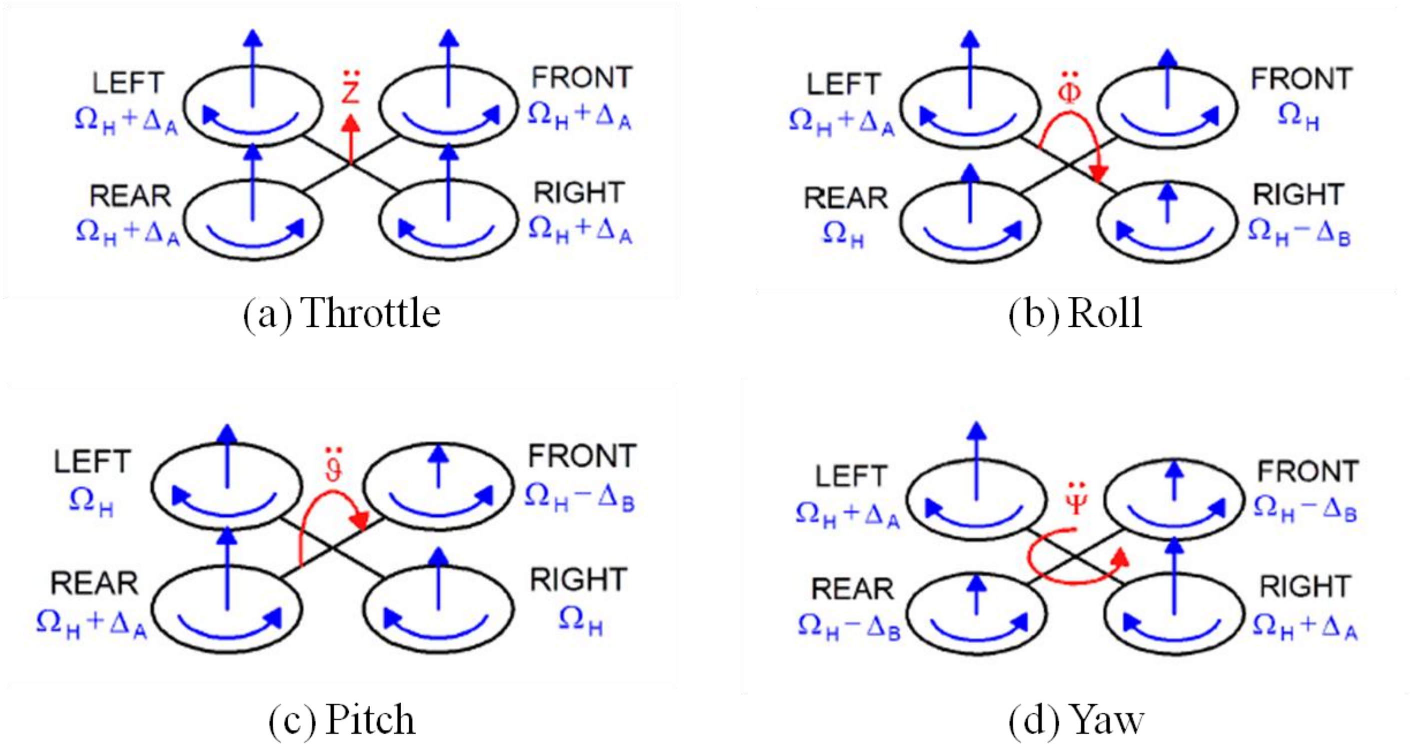

Fig. 23 AR.Drone movements [64]

Chapter 4
Evaluation Experiments

4.1 Experiment Environment

The problem of detecting the target from air is identified by the change of texture and edge in floor and road. In general, the active area of people can be roughly divided into indoors environment and outdoors environment. In addition, indoors environment can be divided into simple environment (a monochromatic space by spreading the blue sheet on the floor of the gymnasium that is used for our experiments) and cluttered environment (a state of that there are many patterns on the ground). Simple environment means the color of the ground is monochrome. Cluttered environment means the color of the ground is not monochrome or there are other patterns on the ground. Outdoors environment also can be divided into simple environment and cluttered environment, and the means of them is the same with indoors environment. However, the important different between indoors environment and outdoors environment is the outside influence, such as wind, airflow, atmospheric pressure, and so on. In order to make our study closer to the actual situation, we carry out experiments in simple environment and cluttered environment of indoors and outdoors environment. Because we need wide area to carry out our experiments for indoor environment for flying robot, so we choose the gymnasium to perform our experiments.

Because there are some lines and other patterns on the ground of gymnasium, we use a blue cloth to put on the floor to make ground become monochrome. In this case, we expect to verify the effectiveness of the proposed method if the subject can be detected and tracked correctly. In the cluttered environment of indoors, this case is near to general situation, so we have to solve effects of background, and let the subject can be detected and tracked correctly. In outdoors environment, this case is the general case, and we have to consider the effects of outside (wind, airflow, shadow, and so on). The purpose of this case is to solve effects of outside and make the subject can be detected and tracked accurately. In each case, there is a purpose for the experiment, and it is shown in Table.11.

Table.11 The purpose of experiment in each environment

Background	*Sample Image*	*Purpose*
Simple (indoors)		Correctly detect the subject and track him
Cluttered (indoors)		Solve the effects of background, and correctly detect the subject and track him
Outdoors		Solve effects of outside, and correctly detect the subject and track him

For the experiment, the quadrotor named AR.Drone made by Parrot Company is utilized. There are two types for indoors and outdoors respectively. There are two CMOS cameras attached at the front and bottom of AR.Drone. Camera resolution of each is 640×480 (30fps). Because cameras attached on AR.Drone are visible cameras, the experiment is carried out in day time. Also, there are Ultrasonic Sensor, Gyro Sensor, and Acceleration Sensor with it. In addition, AR.Drone 2 is sold on July, 2012. Table.12 shows the specification of AR.Drone 1 and AR.Drone 2. AR.Drone 1 is lighter than AR.Drone 2, but the camera performance and sensor performance of AR.Drone 2 are better than AR.Drone.

Table.12 The specification of AR.Drone 1 and AR.Drone 2

	AR.Drone 1	**AR.Drone 2**
For indoors		
For outdoors		
Speed	18km/h	18km/h
Weight	420g(indoors)	455g(indoors)
	380g(outdoors)	425g(outdoors)
Camera	480p, 30fps	720p, 30fps
Sensors	Ultrasonic sensor 3-axis accelerometer (2-axis gyroscope and 1-axis yaw precision gyroscope)	Ultrasonic sensor 3-axis Magnetic sensor 3-axis gyroscope Pressure sensor

Moreover, Table.13 shows the specification of PC our used. The CPU of PC for experiment is Intel Core i5, the memory is 2GB, and the OS is Microsoft Windows XP Professional. Programming language is Microsoft Visual C++. Development environment is Microsoft Visual Studio 2010. The functions of OpenCV2.3 are used for a part of the processing. Experiments are carried out in gymnasium and outdoors, and executed with AR.Drone manual control.

Table.13 the specification of PC our used

OS	Microsoft Windows XP professional
CPU	Intel Core i5
Library	OpenCV2.3
Development Environment	Microsoft Visual Studio 2010

4.2 Preliminary Experiment

4.2.1 Visual Field of Camera

In order to understand the visual field of camera on the bottom of AR.Drone, we do the experiment to test it. Firstly, we paste four red tapes on the floor (as shown in left image of Fig.24). Then, we make AR.Drone fly at 1m height, and adjust the position of tapes. When four tapes are just at the border, positions of tapes are noted. According to these data, visual field can be calculated (as shown in right image of Fig.24). AR.Drone is flying at about 3m. In this case, area size is 2.91m×2.1m, and the area of region is 6.111m^2.

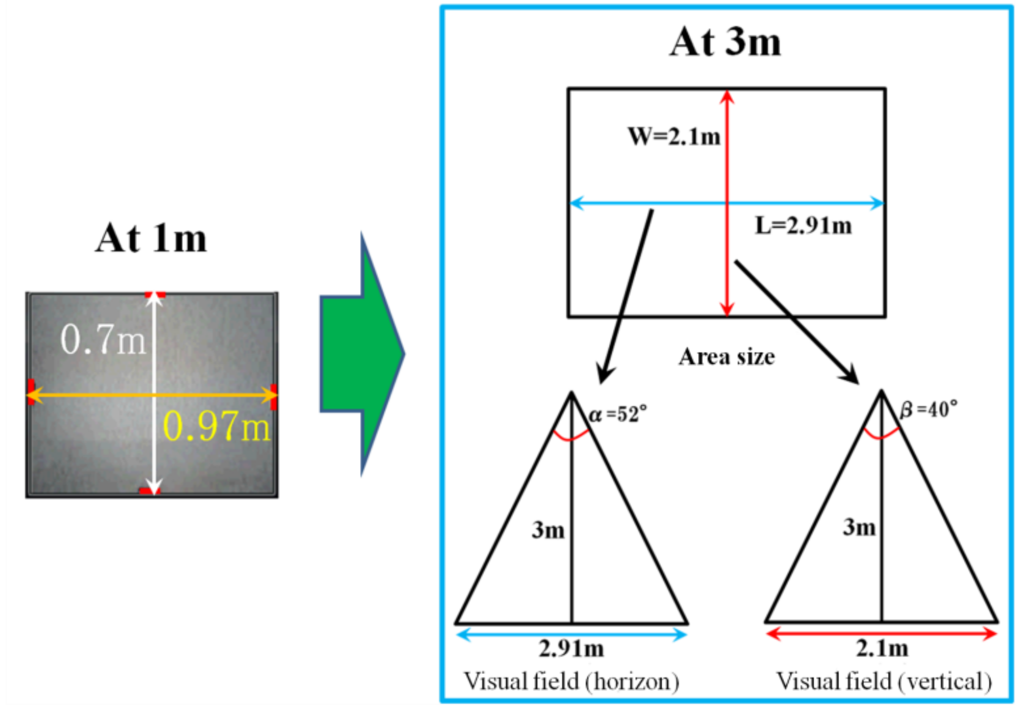

Fig.24 Visual field of AR.Drone

4.2.2 Cap Detection

In previous chapter, we have said how to detect the head part of the subject. However, when the subject wears cap or hat, prior learning of cap or hat is necessary. Beforehand, multiple patterns (color histogram of head or cap) are learned, and procedure is shown in Fig.25. Next, some head images (reference images) are prepared. Then, the head part can be detected by using the Hough Transform method (mask images). After that, Hue (or Brightness) histogram can be calculated. In addition, these patterns (histogram) are learned. Furthermore, because only head part is used, there is no constraint for background (both with single background and complex background).

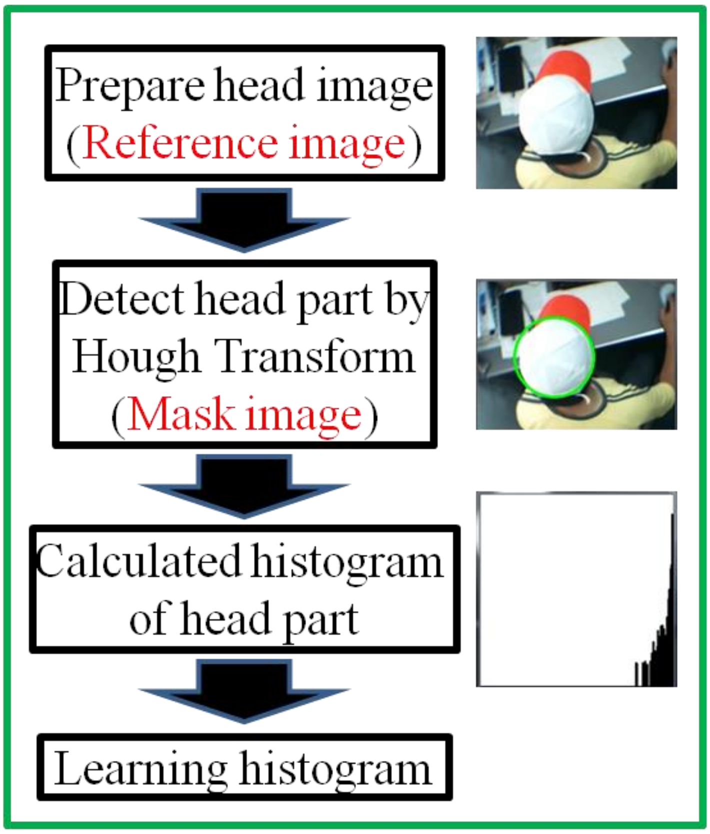

Fig. 25 Procedure of prior learning

After that, edge image can be attained by calculating quadratic differential of aerial image. Then, circular objects are detected by Hough transform method. In addition, brightness histogram of each circular object (Hist2) can be calculated. Finally, cap or hat part can be determined by comparing Hist1 and Hist2 (as shown in Fig.26).

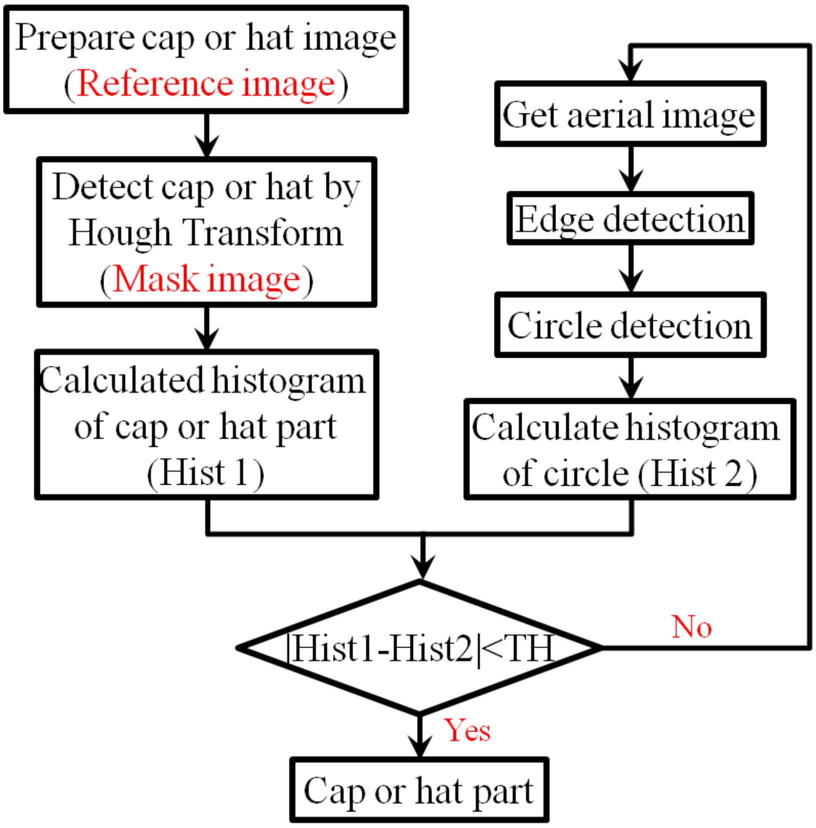

Fig. 26 Flowchart of cap detection

4.2.3 Threshold of Head Detection Setting

In head detection part, we use the Hough Transform method to detect the head part of the subject. In order to correctly detect the head part, we have to set the threshold value of head radius for the Hough Transform method. In previous chapter, the Hough Transform method is introduced. If head radius is unknown, the number of space dimensions will increase, and calculation amount also increases. Therefore, we have to test the head radius in different altitude for threshold automatically setting.

In this experiment, because we carry out the experiments at about 3m, we test the head radius in 0.5m, 1m, 1.5m, 2m, 2.5m, 3m, and 3.5m. The graph is shown in Fig.27. From the Fig.27, we can get the maximum value of head radius and minimum value of head radius at any altitude.

The equation of the maximum value is

$$y = 1E - 12x^4 - 1E - 08x^3 + 4E - 0.5x^2 - 0.0715x + 59.571 \tag{35}$$

The equation of the minimum value is

$$y = 1E - 12x^4 - 1E - 08x^3 + 5E - 0.5x^2 - 0.0742 + 56.000 \tag{36}$$

In equation (35) and (36), x is flight altitude of AR.Drone, and y is the head radius of the subject in different altitude.

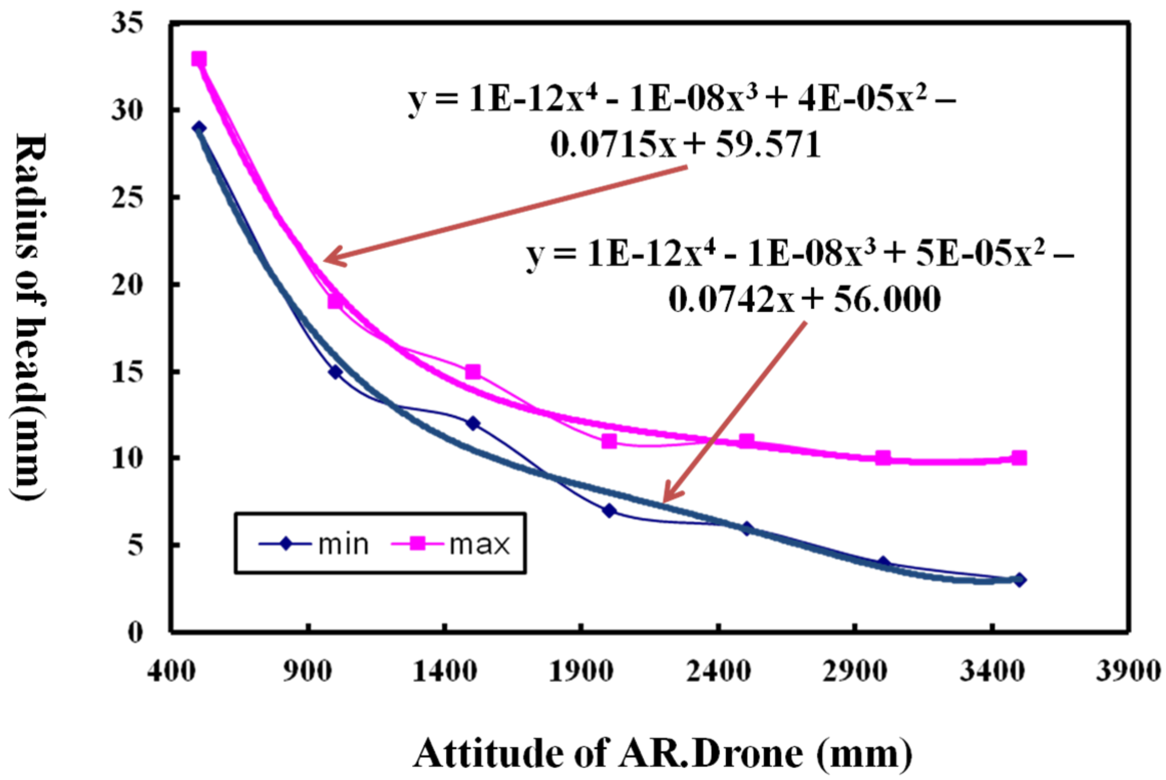

Fig. 27 Threshold automatically setting for head detection

4.2.4 AR.Drone Performance Test

In this part, properties of AR.Drone are to be tested. Properties of AR.Drone contain the minimum moving speed in front and rear direction, the minimum rotation speed, the minimum moving speed in left and right direction, and the minimum moving speed in vertical direction. Test data is shown in Table.14, Table.15, Table.16 and Table.17.

Table.14 Moving speed in front and rear direction

Time	Distant
10s	8.5m
10s	8.3m
20s	15.7m
20s	15.3m
30s	26.2m

Table.15 Rotation speed of flying robot

Time	Angle
5s	45.6 degree
10s	91.2 degree
15s	135.9 degree
20s	182.1 degree
30s	272.5 degree

Table.16 Moving speed in left and right direction

Time	Distant
1s	0.6m
5s	3.4m
5s	3.75m
10s	8.3m
10s	8.15m

Table.17 Moving speed in vertical speed

Time	Distant
1s	10cm
3s	30cm
3s	35cm
3s	30cm
3s	27.5cm

Therefore, the minimum moving speed in front and rear direction of AR.Drone is 82.06 cm/s in average and the minimum rotation speed of AR.Drone is 9.10 degree/s in average. The minimum moving speed in left and right direction of AR.Drone is 73.5cm/s in average and the minimum moving direction in vertical direction of AR.Drone is 10.17cm/s in average. The maximum moving speed of AR.Drone is 18km/h. Fig.28 shows the example image of AR.Drone performance test. In Fig.28, images on the left are start position of the AR.Drone, and images on the right are end position of the AR.Drone. Then, we can know speed due to flight time and flight distant (or rotation angle).

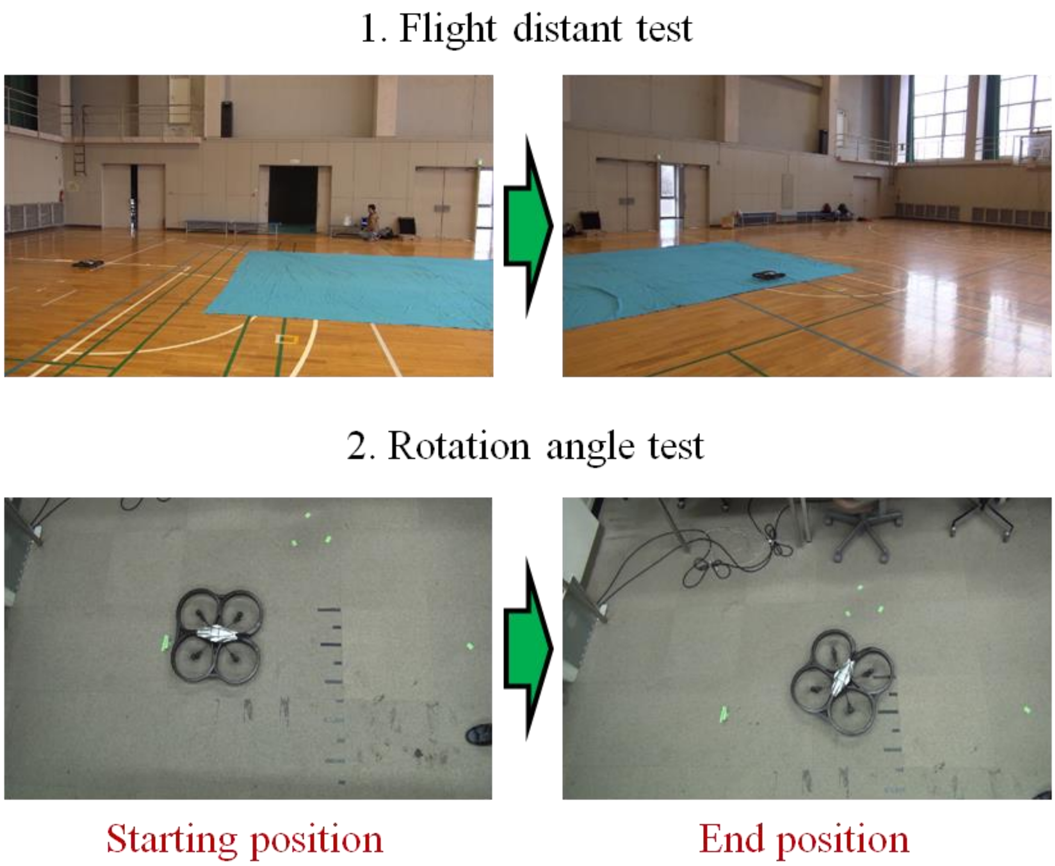

Fig.28 The experiment of AR.Drone performance test

When we got these values of speed, we can use the Proportional (P) Control to automatically set AR.Drone control thresholds. The equation is as follows.

$$\begin{pmatrix} phi \\ theta \\ gaz \\ yaw \end{pmatrix} = \begin{pmatrix} 73.5 & 0 & 0 & 0 \\ 0 & 82.06 & 0 & 0 \\ 0 & 0 & 10.17 & 0 \\ 0 & 0 & 0 & 9.10 \end{pmatrix} \begin{pmatrix} V_{lr} \\ V_{fr} \\ V_{\mathrm{var}} \\ V_{rot} \end{pmatrix} \tag{37}$$

In equation (37), "phi" is threshold of speed in front and rear direction, "theta" is threshold of speed in left and right direction, "gaz" is threshold of speed in vertical direction, and "yaw" is threshold of rotation speed. V_{lr} is speed in left and right direction of the subject, V_{fr} is speed in front and rear direction of the subject, V_{var} is speed in vertical direction of the subject, and V_{rot} is rotation speed of the subject. When the speed in each direction of the target is known, threshold of AR.Drone control can be automatically set by using this method.

4.3 Experiment Results

4.3.1 Human Detection

<1> Head detection

When the subject is not wearing the cap or hat, head part can be detected by the Hough Transform method and the Brightness Histogram of black color, because the most of people's head is similar with black. The detection results are shown in Fig.29.

Experiments of head detection are carried out in simple environment, cluttered environment, and outdoors environments respectively. Also, experiments are carried out at about 3m and 4m. Fig.29 shows the successful cases of head detection. The success rate is different in different environment. In simple environment, because the background is monochrome, there is little noise in this case. However, the altitude would affect the results. When we do experiments in 3m, we can easily detect the head. On the contrary, when flying robot is at 4m, head slightly more difficult to be detected, but the gap is small. Therefore, the success rate of head detection in simple environment is over 95%. In cluttered environment, because there are other patterns in the ground, the success rate of head detection would be influenced. Therefore, the success rate of head detection is about 85%. In outdoors environment, head detection is not only affected by background but also by outside influence (such as wind, airflow, and so on). Therefore, the success rate of head detection is low, about 74%. Head detection is not only influenced by background, outside influence, but also by flight altitude of flying robot.

Fig.30 shows the failure cases of head detection in different environment and different altitude respectively. In simple environment, there is almost no failure detection. In cluttered environment, because of background influence, there are some failure cases of head detection. In outdoors environment, because of outside influence and background influence, the success rate would be affected.

As shown in Table.18, the success rate of head detection in simple environment is 97.2% (at 3m) and 95.8% (at 4m) respectively. The success rate of head detection in cluttered environment is 86.6% (at 3m) and 82.7% (at 4m) respectively. The success rate of head detection in outdoors environment is 77.4% (at 3m) and 72.6% (at 4m) respectively.

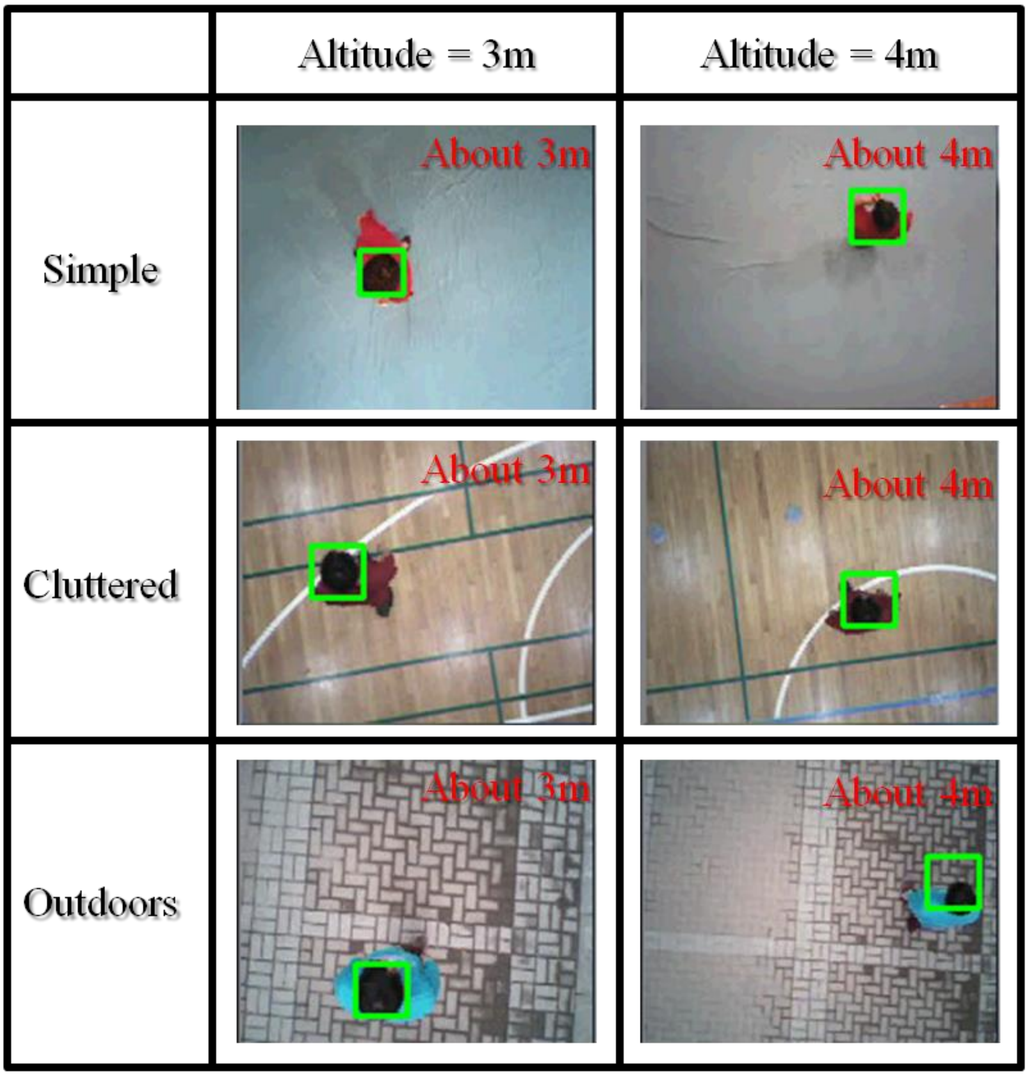

Fig.29 Successful cases of head detection

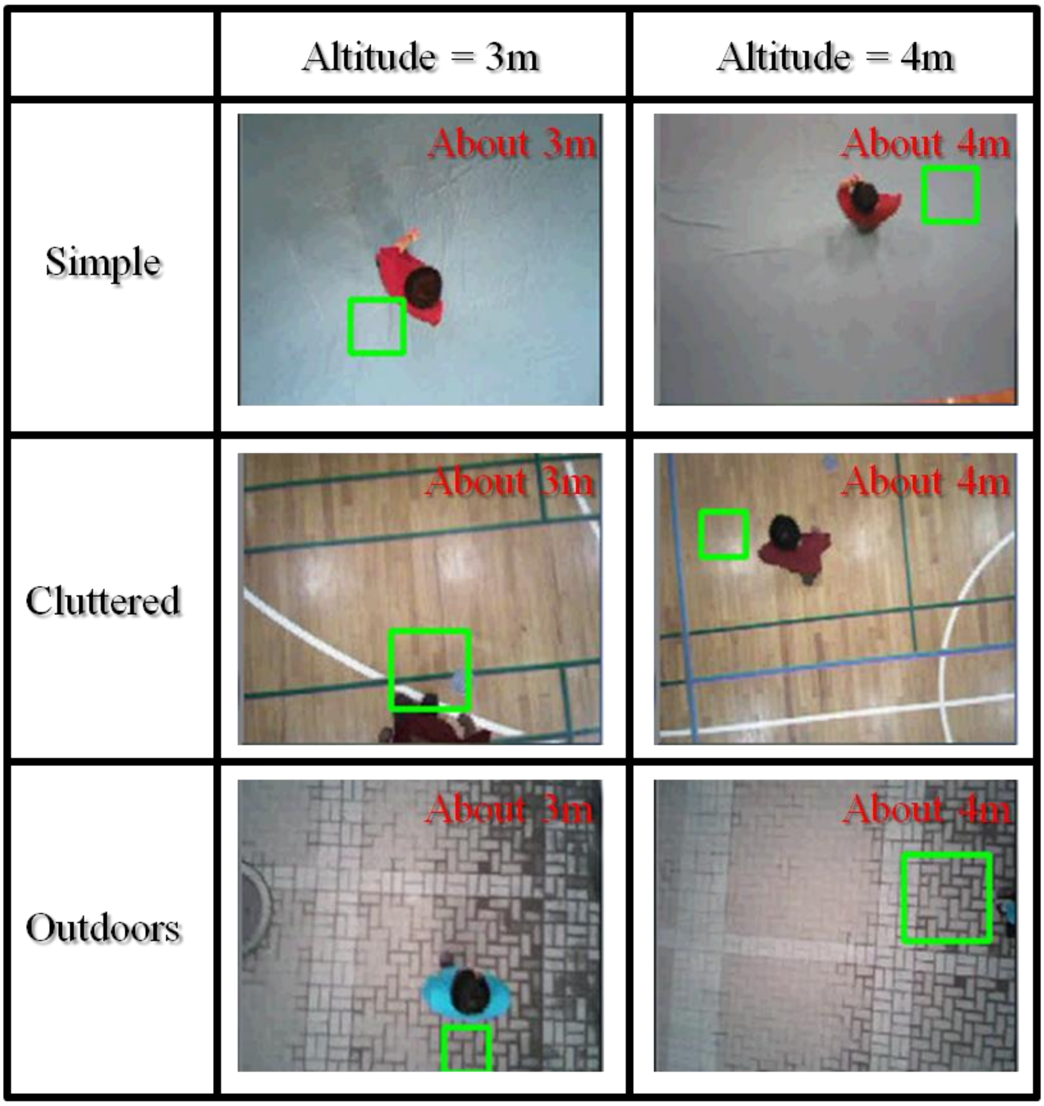

Fig.30 Failure cases of head detection

Success rate (sr) is calculated according to equation (38).

$$sr = n_{\det ect} / N \tag{38}$$

In equation (38), the sr is success rate, n_{detect} is a number of frames in which human is successfully detected, and N is the total frames.

Table.18 The success rate of head detection

	Altitude = 3m	Altitude = 4m
Simple	97.2%	95.8%
Cluttered	86.6%	82.7%
Outdoors	77.4%	72.6%

<2> Clothes detection

When head is detected in previous chapter, the clothes part of the subject would be detected in this part. About clothes part detection, head information is necessary. Firstly, the initial contour is set by using head center and head radius. Then, control points will be chosen. By converging of control points, body contour can be extracted. In addition, clothes information can be calculated. Therefore, factors of initial contour setting and environment situation are very important. We do experiments in the simple environment, cluttered environment, and outdoors environment respectively. Also, we do experiments in about 3m and about 4m respectively, because flight altitude will affect detection results.

Fig.31 and Fig.32 show results of the clothes detection. In simple environment, because background is monochrome, as long as the head can be detected, body contour can be extracted correctly. In cluttered environment, because background is complex, it would affect the results of getting body contour. In this case, the success rate of clothes detection is low. In outdoors environment, because head detection is difficult, it will influence the initial contour setting, so the clothes detection would be affected. Also, there are influences on wind, airflow, and so on. Therefore, it would cause flying robot instability, so the results of clothes detection would be affected.

Fig. 31 shows the successful cases of clothes detection in the simple environment, cluttered environment, and outside environment. In addition, Fig.32 shows some failure cases of clothes detection in different environments and different flight altitude.

In our study, body contour only be carried out at the beginning of the experiment. Therefore, head detection will affect the result of clothes detection directly. If head part can be detected correctly, clothes part also can be detected at the beginning of the experiment.

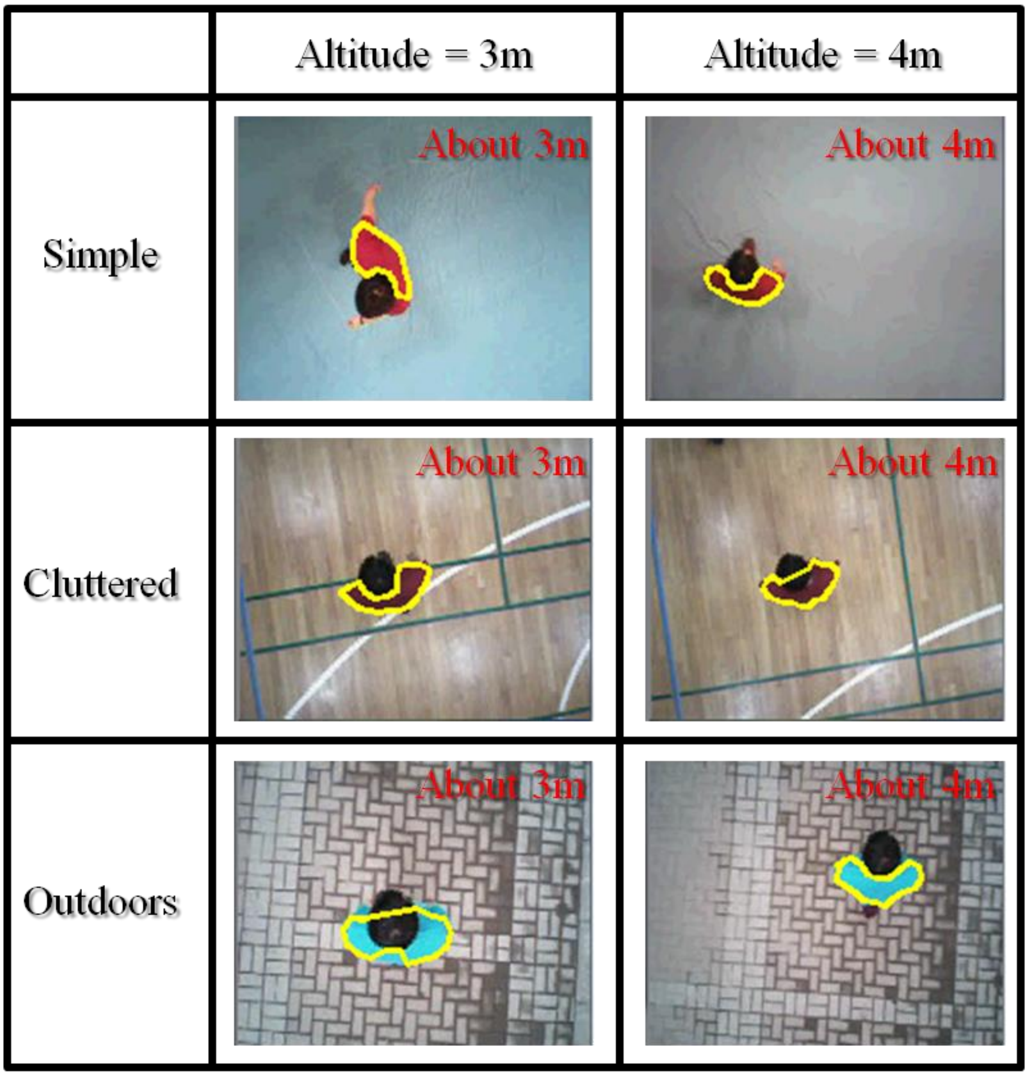

Fig.31 Successful cases of clothes detection

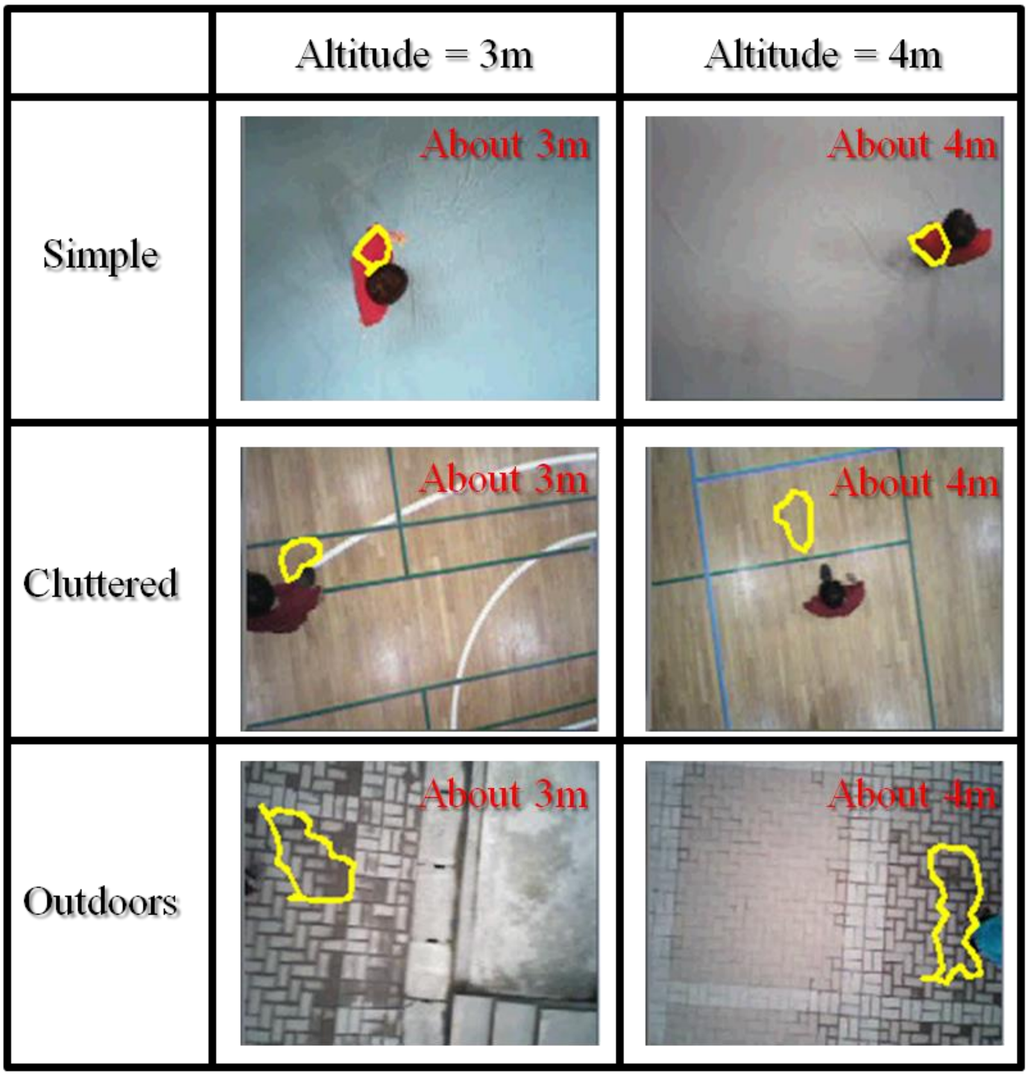

Fig.32 Failure cases of clothes detection

4.3.2 Human Tracking

In our study, human tracking is the very important part of image processing thread. We test the tracking properties in the simple environment, cluttered environment, and outdoors environment.

<1> One person

In simple environment, because background is monochromatic, when only one person is in the scene, the subject can be correctly tracked in simple background. In cluttered environment, because the Mean Shift method uses the color histogram to track the subject and there are many patterns on the background, success rate of human tracking would be affected by complex background. In addition, in outdoors environment, because of outside influences (such as wind, airflow, and so on), flying robot is instability. It would influence the result of human tracking. In addition, because the background is complex, success rate of human tracking would be affected.

Fig.33 shows the successful cases of human tracking when only one person in indoors environment and outdoors environment. Left part shows results of head tracking, and right part shows results of clothes tracking. Because there are many influences, there are some failure cases in our experiments. Therefore, Fig.34 shows the failure cases of human tracking in every environment and situation. The standard of success is head part is tracked successfully and clothes part is tracked successfully. The success rate is calculated due to equation (38), N is 4000 frames. Because success rate is calculated when the subject exists, in 4000 frames, the target always exists in the scene.

As shown in Table.19, when there is one person in the scene, the success rate of human tracking in simple environment is 98.7%, in cluttered environment is 94.3%, and in outdoors environment is 90.2%.

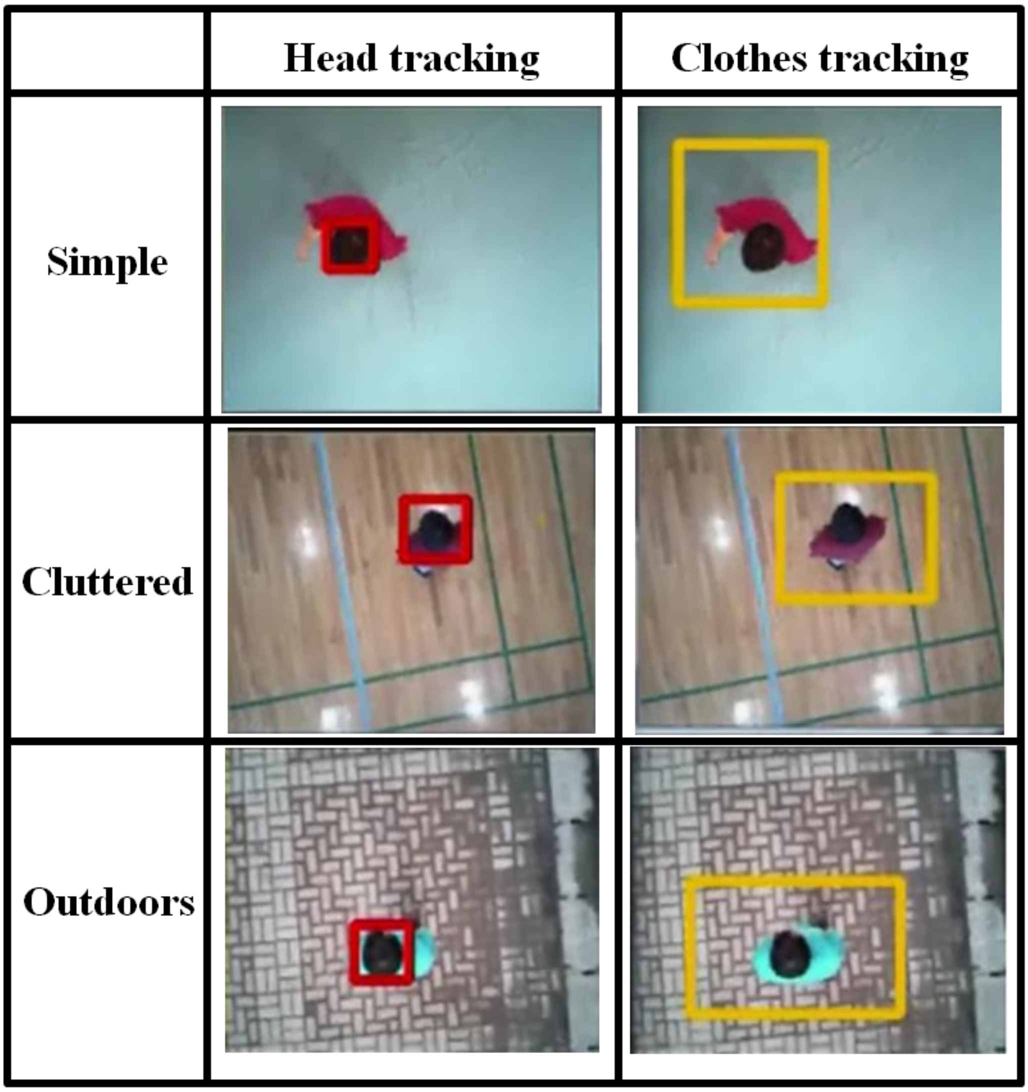

Fig.33 Successful cases of human tracking

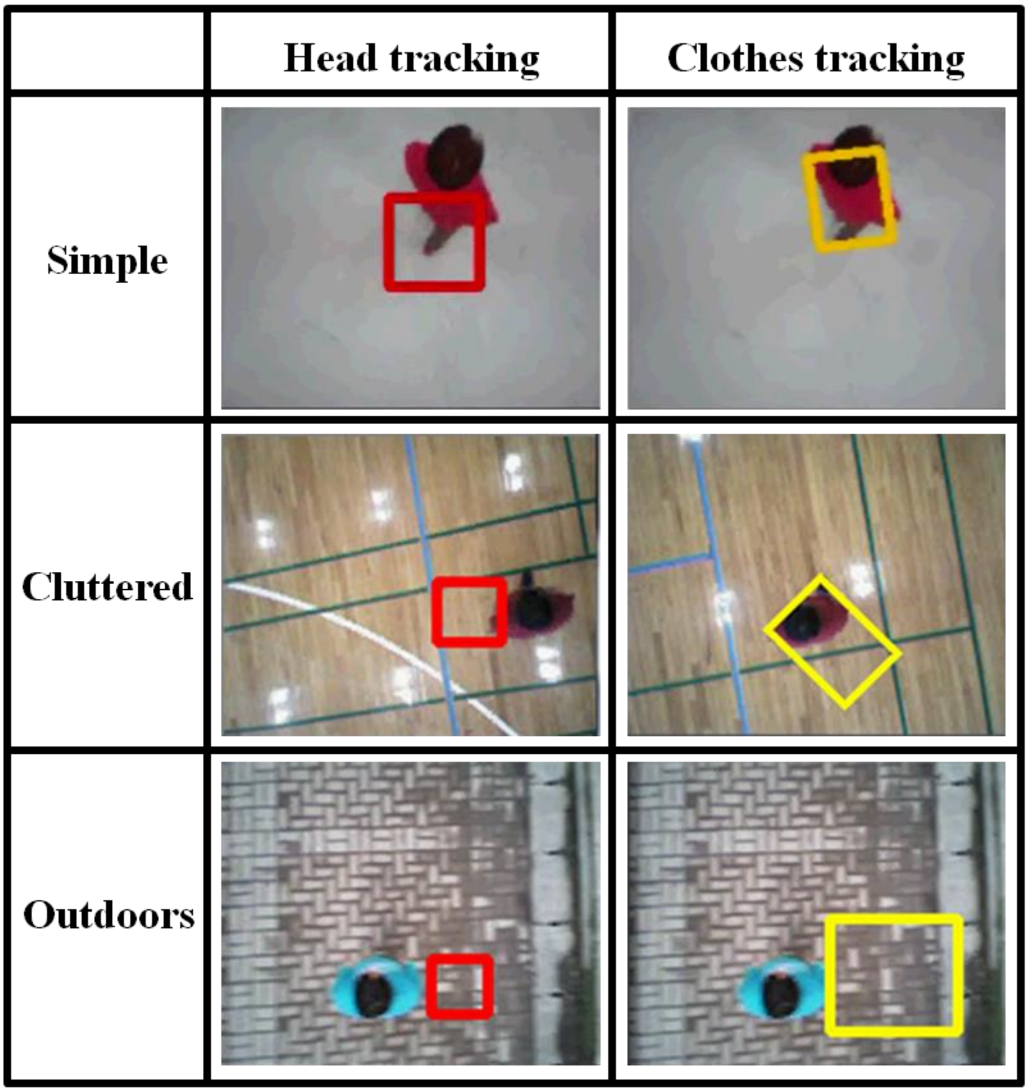

Fig.34 Failure cases of human tracking

Table.19 Success rate of one person tracking

	Frame number	**Success rate**
Simple	4000	98.7%
Cluttered	4000	94.3%
Outdoors	4000	90.2%

<2> Multiple persons

When there are multiple persons in the scene and the color of clothes are similar, the success rate of human tracking is tested. In simple environment, when there are multiple people in the scene and these people are wearing same color clothes with the subject, because the Mean Shift method uses the color histogram to track the subject, result of human tracking would be influenced. To separate the subject from these multiple persons, the Labeling method is used.

In cluttered environment, influences are not only the same color clothes but also the cluttered background, and the stability of the flying robot. In outdoors environment, firstly, because of outside influences (such as wind, airflow, and so on), flying robot is instability. It would influence the result of human tracking. In addition, because the background is complex and the same color clothes, success rate of human tracking would be affected.

Fig.35 shows the successful cases of human tracking when multiple persons are in indoors environment and outdoors environment. In Fig.35, there are five rows. First row shows the head tracking image, second row shows the clothes tracking image, third row shows the labeling image of current frame, fourth row shows the labeling result of last frame, and fifth row shows the labeling result of current frame. Because there are many influences, there are some failure cases in our experiments. Therefore, Fig.36 shows the failure cases of human tracking in every environment and situation. The standard of success is head part is tracked successfully and clothes part is tracked successfully. The success rate is calculated due to equation (38), N is 4000 frames. Because success rate is calculated when the subject exists, in 4000 frames, the target always exists in the scene.

As shown in Table.20, when there are multiple persons in the scene, the success rate of human tracking in simple environment is 97.8%, in cluttered environment is 91.2%, and in outdoors environment is 88.3%.

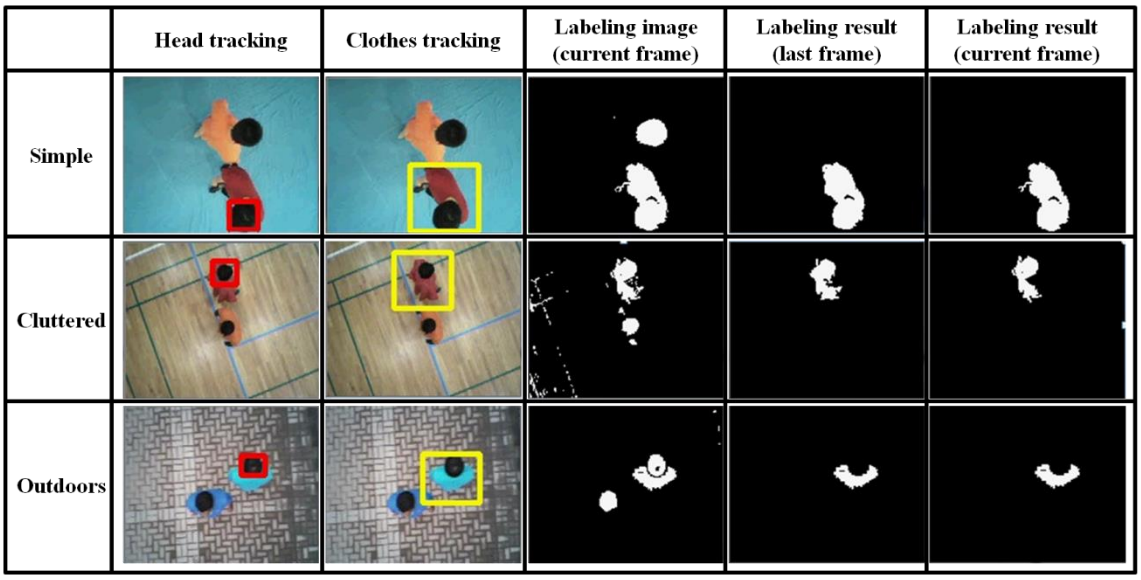

Fig.35 Successful cases of human tracking

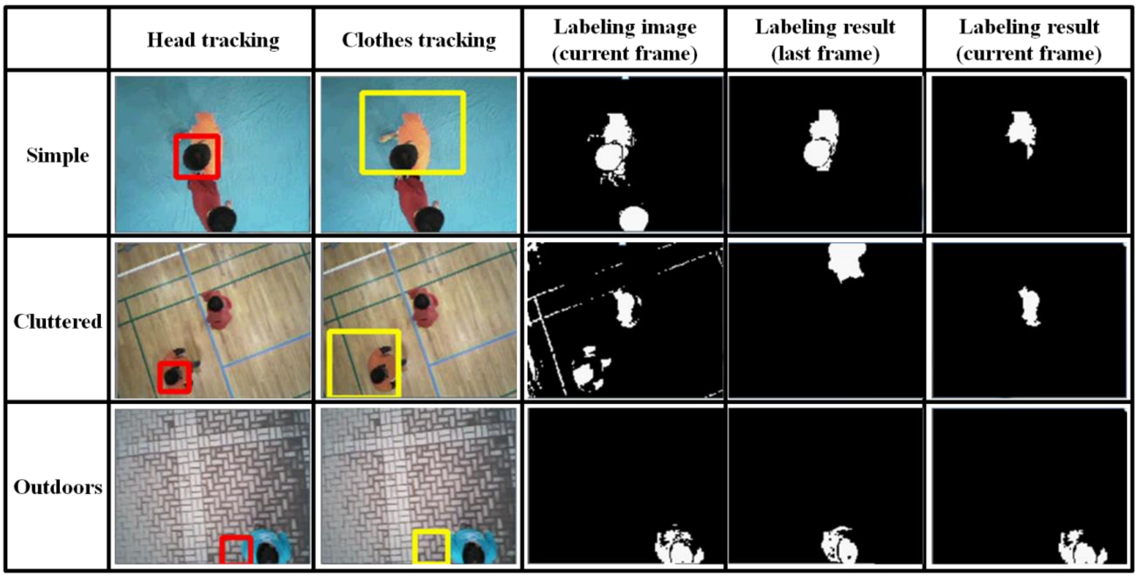

Fig.36 Failure cases of human tracking

Table.20 Success rate of target tracking (multiple persons)

	Frame number	Success rate
Simple	4000	97.8%
Cluttered	4000	91.2%
Outdoors	4000	88.3%

4.3.3 Target Redetection

Once the subject is lost, the target must be correctly detected again. Firstly, we have to detect the head part of the subject again by using the Hough Transform method and the color histogram, and the processing flow is same with the previous chapter. Then, we have to detect the position of the clothes part of the subject again. However, we make use of the Particle Filter method for the clothes area redetection. The results are shown in Fig.37 and Fig.38. Because the target redetection is carried out only when the subject is lost, it is not necessary to perform it in every frame, so the success rate is not very important. We only expect the target can be correctly detected when he or she is lost.

Fig.37 shows the successful cases of the target redetection, and Fig.38 shows the failure cases of the target redetection. From the results, when experiments are carried out in simple environment of indoors, the subject can be almost correctly redetected without no background influence and no outside influence. Therefore, there is no failure image of simple environment in Fig.38.

When experiments are performed in cluttered environment, because color of background is similar with the clothes part, this reason would cause the false detection. In Fig.38, background in the field of view is detected as the clothes part, so this result is wrong, but most cases still correct.

In outdoors environment, the influence is not only color of background but also wind, airflow, and so on, so the difficulty of the subject redetection increases in this case. Because of outside influence, flying robot is instability, so it would cause flying robot flying around and shaking. Therefore, the target would be lost usually and the aerial images captured also are unstable. Therefore, false detection is more than it in simple environment and in cluttered environment.

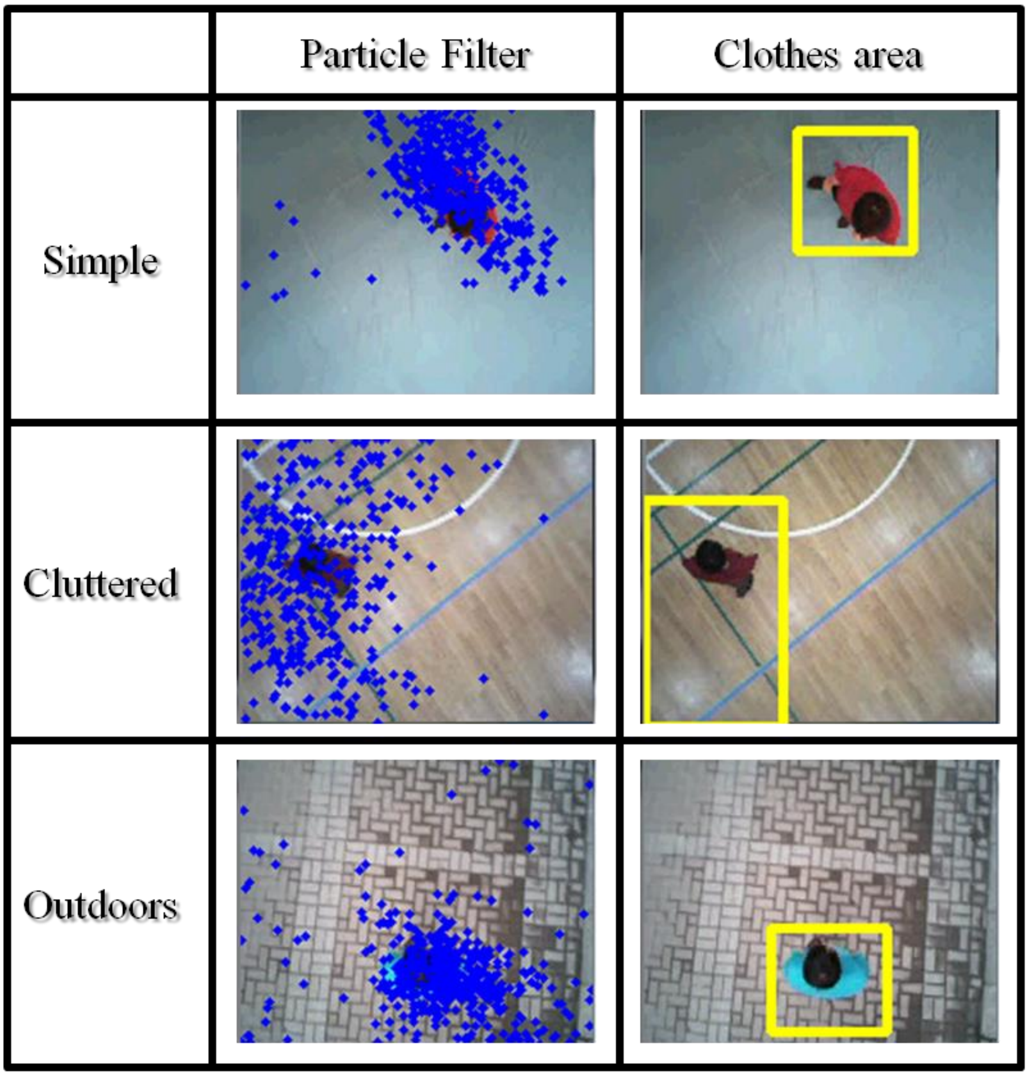

Fig. 37 Successful cases of the target redetection

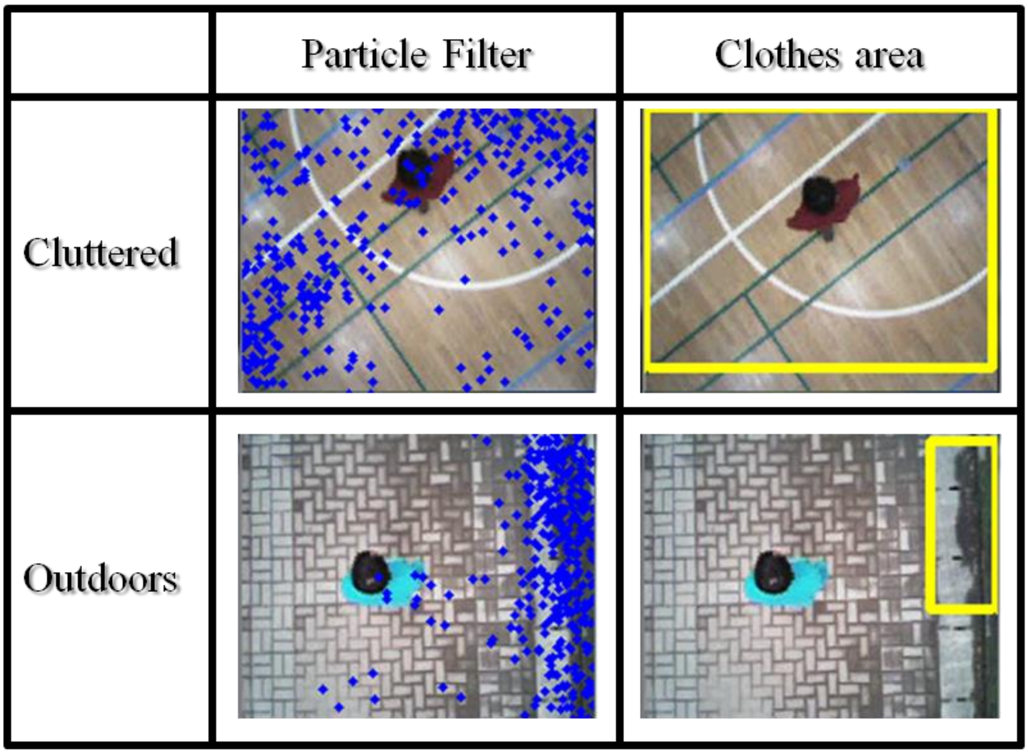

Fig. 38 Failure cases of the target redetection

4.3.4 Moving Direction Detection

To correctly calculate the moving direction of the target, the positions of the head part and clothes part have to be known. This is a new idea for direction calculation. In many previous studies, they use the center of gravity moving to calculate the moving direction. There are two problems cannot be solved. Firstly, sometimes the center of gravity is not correct. Secondly, the moving direction of the current frame cannot be known. By our method, these problems can be solved well. Experiments are performed in simple environment of indoors, cluttered environment of indoors, and outdoors environment. Fig.39 shows the successful cases of moving direction detection. A group images in the left of Fig.39 are head detection results, and the center of red rectangle shows the position of head part. A group images in the middle of Fig.39 are clothes detection results, and the inclination of the bounding rectangle represents the direction of the body. Then, according to the position relationship between head part and the clothes part, moving direction of the subject can be decided. The results are shown in the group images in the right of Fig.39.

Fig.40 shows failure cases of direction detection, because head part or clothes part is failed to be detected. In outdoors environment, because of strong wind or airflow, it will cause the object to leave the field of view.

As shown in Tabel.21, when there is one person in the scene, the success rate is 98.1% in simple environment, 95.7% in cluttered environment, and 93.6% in outdoors environment respectively. In addition, when there are multiple persons in the scene, the success rate is 96.5% in simple environment, 93.2% in cluttered environment, and 91.4% in outdoors environment respectively.

Calculation method of success rate is same with head detection. However, which is the correct detection and which is wrong detection is a problem.

In our study, a standard value (SV) will be set at first, and then a threshold value (TH) is set. Successful detection is determined by equation (39).

$$CV - SV < TH \tag{39}$$

In equation (38), CV is current detected value, SV is standard value, and TH is the threshold (in our study, threshold is 10 degree, because rotation angle of AR.Drone in one second is 9.10 degree/s in average).

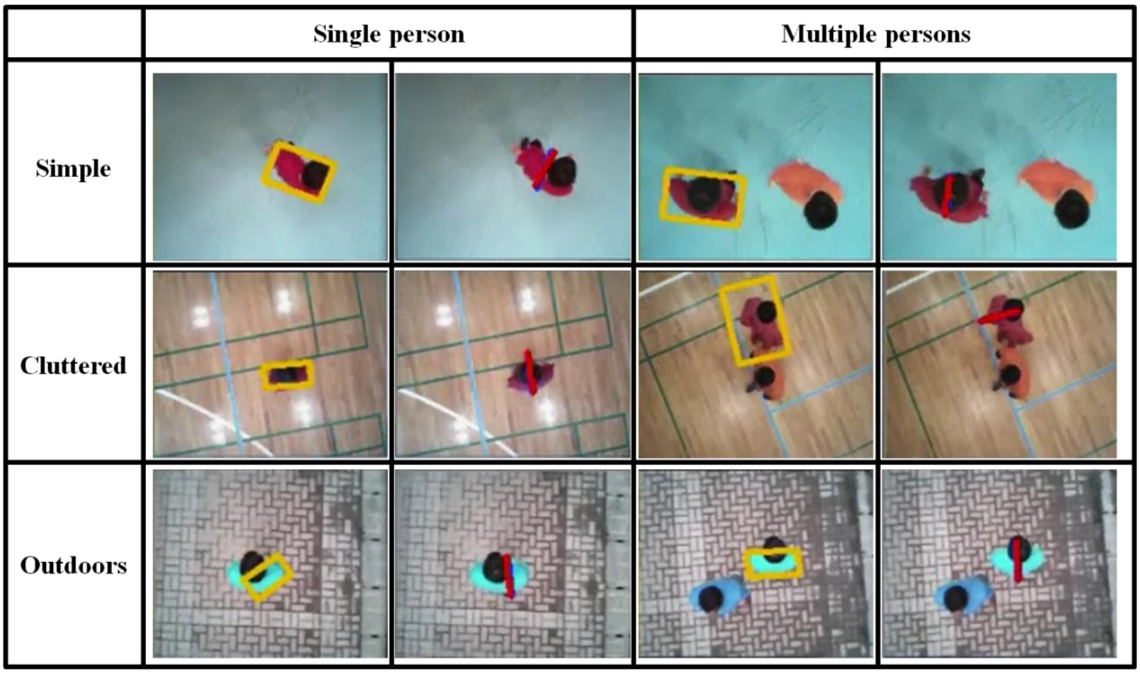

Fig.39 Successful cases of the direction detection

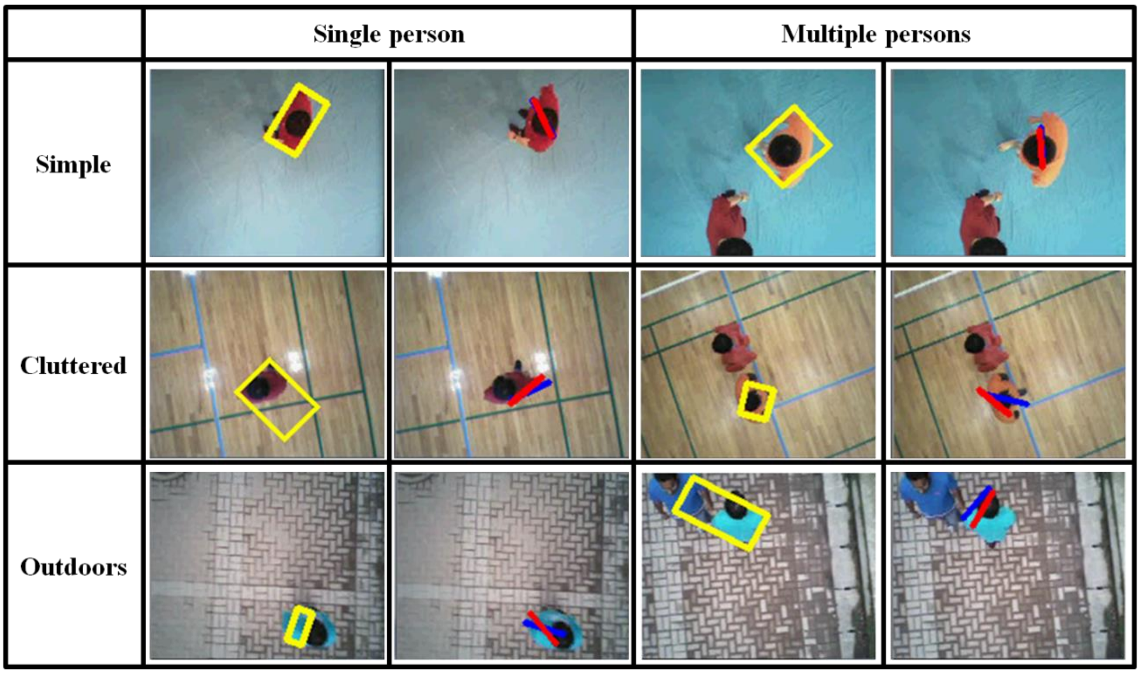

Fig.40 Failure cases of the direction detection

Table.21 The success rate of direction detection

	Frame number	Success rate (one person)	Success rate (multiple persons)
Simple	4000	98.1%	96.5%
Cluttered	4000	95.7%	93.2%
Outdoors	4000	93.6%	91.4%

4.3.5Flying Robot Control

<1> Kalman Filter prediction

In this part, to judge the correctness of prediction data of moving direction by the Kalman Filter algorithm, we draw the trajectory by using prediction data. If it is with the same direction with the subject, we think it is correct. Otherwise, it is wrong.

Fig.41 shows the estimated trajectory of AR.Drone by using prediction data. In fact, the target walks in straight. From the estimated trajectory, if the prediction data is right, AR.Drone will also go straight. Therefore, the prediction data is correct.

Fig.42 shows the case when the target walks in curve. In other words, the subject turns to the left in halfway. Estimated trajectory is a curve bending to the left, and this estimated trajectory is similar with the trajectory of the subject. Therefore, the prediction data calculated by the Kalman Filter method is correct.

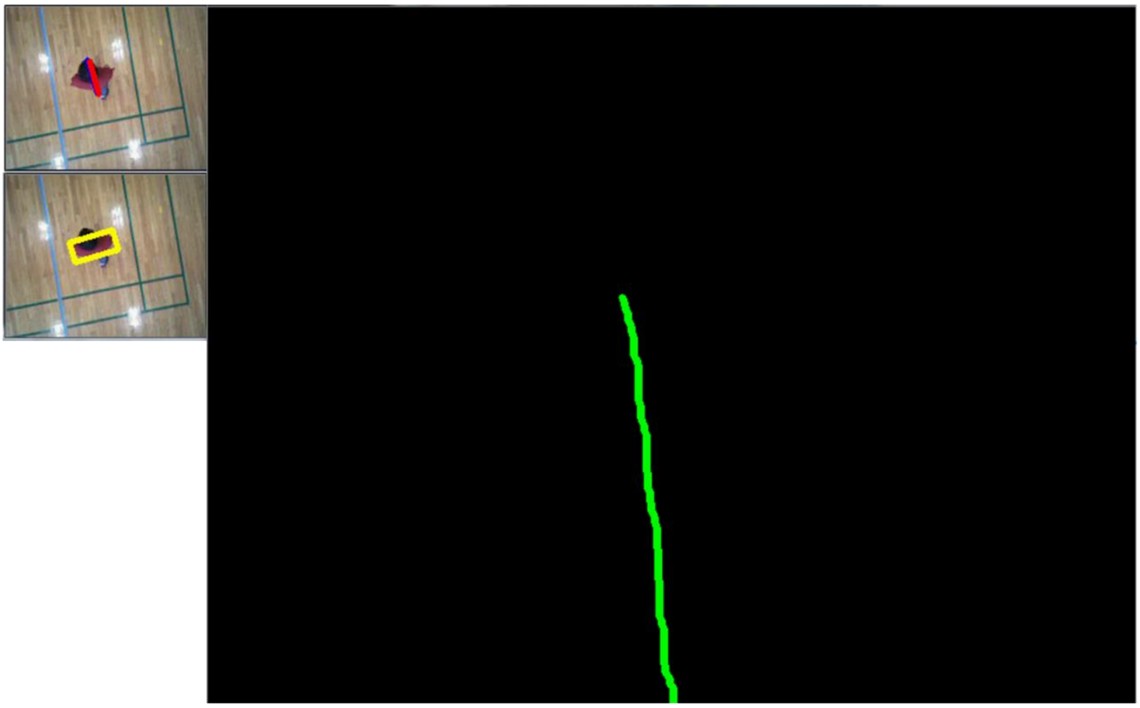

Fig.41 The Kalman prediction (go straight)

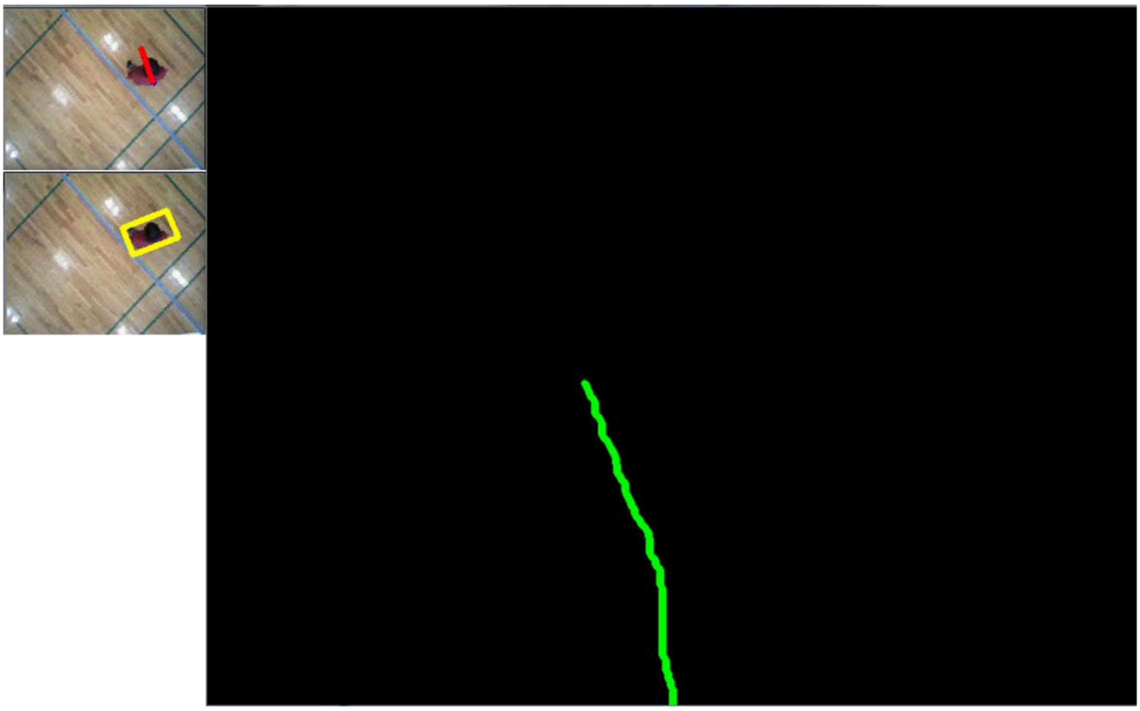

Fig. 42 The Kalman prediction (turn left)

<2> AR.Drone automatically control

In this experiment, AR.Drone will be automatically controlled by position and moving direction of the subject.

We have tested the automatic tracking performance in many cases.

When there is only one person in the scene, tracking performance is tested in simple environment, cluttered environment, short distance and long distance respectively.

When we use AR.Drone to automatically track the subject, the standard of success is as follows. In advance, the goal location is designated. In the beginning of the experiment, AR.Drone is hovering in air, and starting point is that the subject enters the field of view. From start point to goal location, if the subject can be tracked without missing, it is successful. Also, if the target is missing in midway, but AR.Drone can be automatically controlled to rise up and the target can be rediscovered, it is also successful. If not, it is fail.

I. Simple Environment

Firstly, we have carried out experiments in simple environment. The area is from one end to the other end of the blue sheet (about 10m). Fig.43 shows the Successful cases of automatic tracking in simple environment. In Fig.43, front view image is captured by video camera, and it shows the tracking process of AR.Drone. In addition, there are two top view images. The image on the top shows the head part tracking, and the another shows the clothes part tracking. Fig.44 shows failure cases of automatic tracking in simple environment. In Fig.44, for one frame image, we also use front view image and top view images to describe it. The following cases are also same. From Fig.43, target can be correctly tracked in simple environment, because head part and clothes part can be tracked in high accuracy. However, when the subject is missing from the field of view and the speed of the subject is fast, AR.Drone cannot find target again and it is fail.

The success rate of automatic tracking in simple environment is about 80% (as shown in Tabel.22).

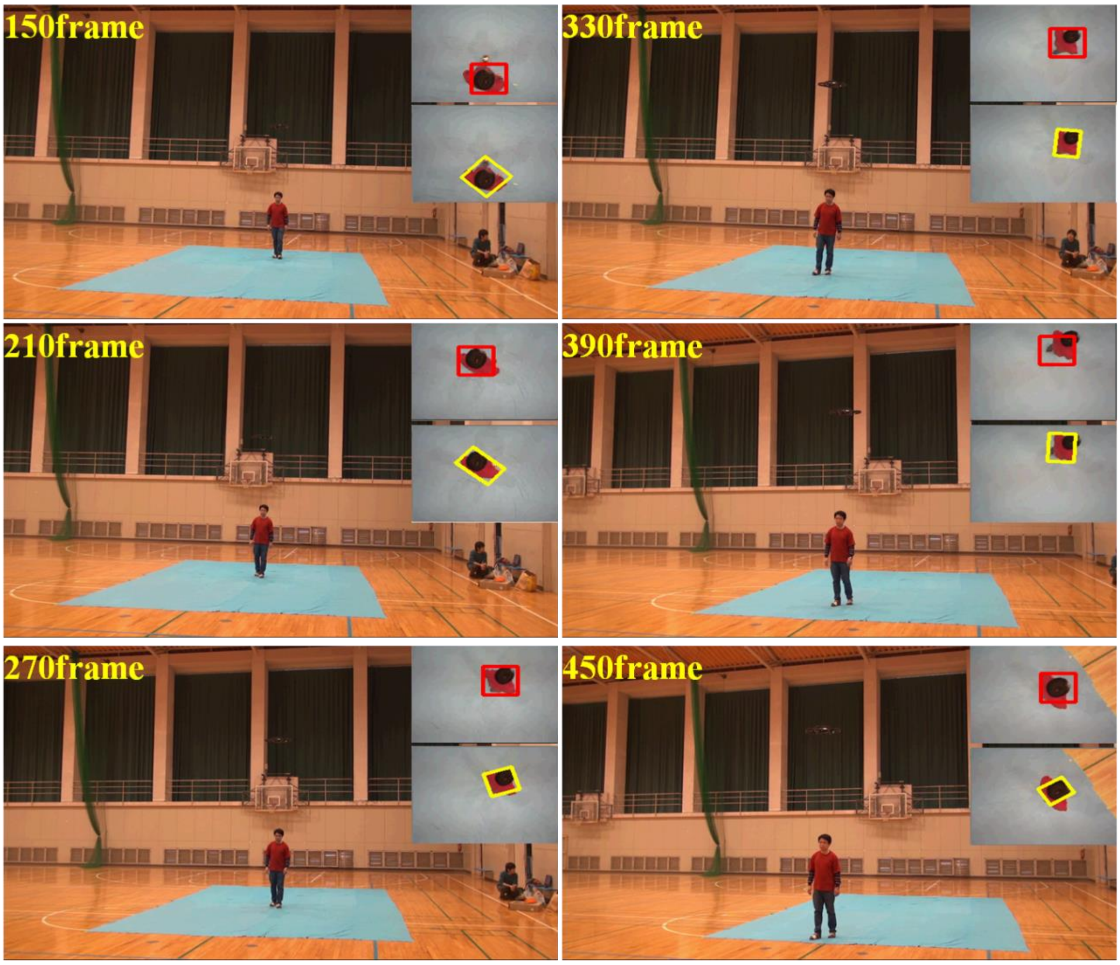

Fig.43 Successful cases of automatic tracking in simple environment

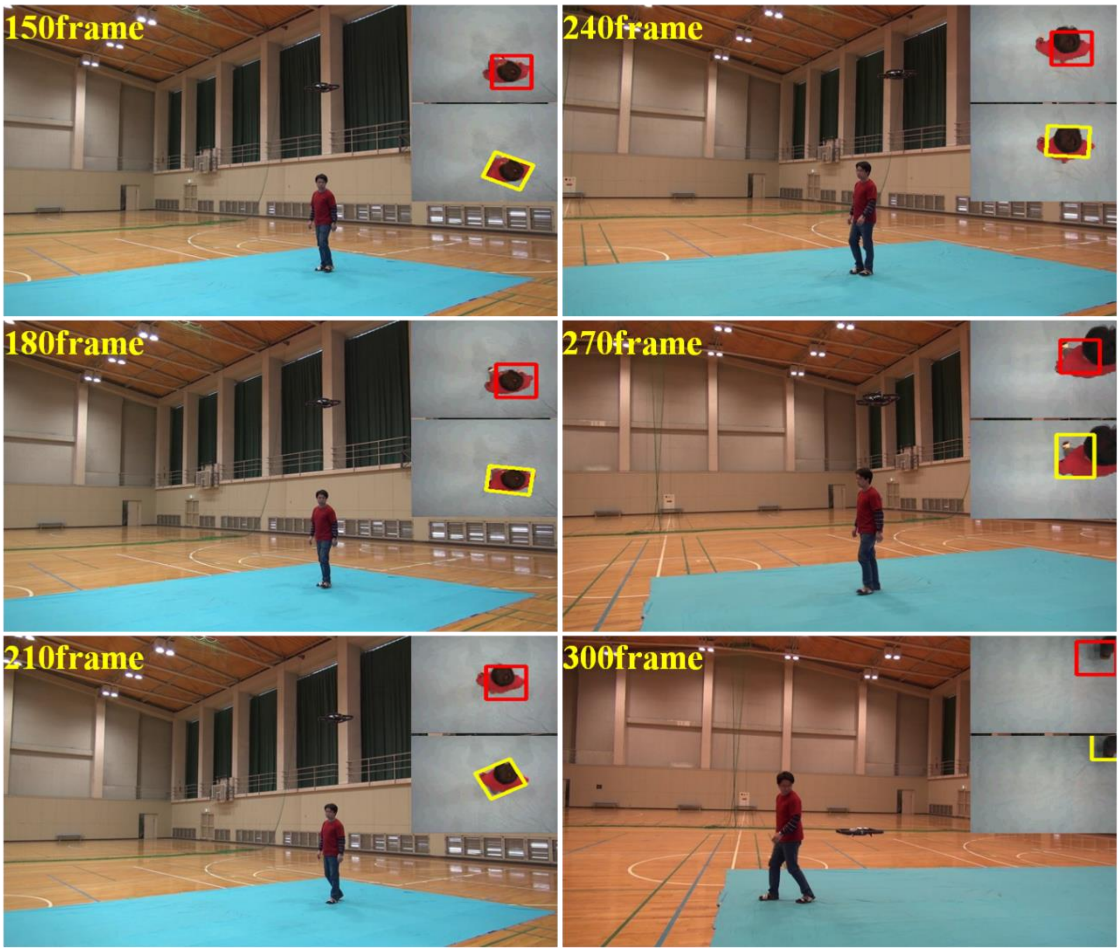

Fig.44 Failure cases of automatic tracking in simple environment

II. Cluttered Environment

In addition, experiments have been carried out in the cluttered environment in indoors. The starting point is the subject enters the field of view, and the goal location is the place of N meters from the starting point (N is about 10, 20, 30…).

Fig.45 shows successful cases of automatic tracking in cluttered environment. In Fig.45, the subject can be tracked successfully, because the speed of the subject is not fast and the subject is in the field of view always. If the speed is fast or quickly turn to other direction, the subject would be lost. Fig.46 shows failure cases of automatic tracking in cluttered environment. In Fig.46, we can see the subject change his direction quickly, so the subject is away from the field of view. Therefore, it cause the experiment failed to track the subject. The success rate of automatic tracking in simple environment is about 68% (as shown in Tabel.22).

Fig.45 Successful cases of automatic tracking in cluttered environment

Fig.46 Failure cases of automatic tracking in cluttered environment

Table.22 Success rate of automatic tracking (simple and cluttered environment)

	Number of experiments	Number of success	Success rate
Simple	50	40	80%
Cluttered	40	27	68%

III. AR.Drone rises automatically

As previously mentioned, when target was lost in midway, AR.Drone would rise up and search the subject again. Therefore, “rise up” is very important for flying robot automatic control. Fig.47 shows AR.Drone rise up to find the subject again when the subject is lost in midway. As to this experiment, there are three occasions, going straight, turning, and AR.Drone rising up. From frame 1 to frame 273, the subject goes straight and AR.Drone tracks the subject correctly. From frame 274 to frame 405, the subject turns to right and the subject was tracked in correct. From frame 406 to frame 495, the subject was missing and AR.Drone rises up automatically to find the target again. From frame 496 to frame 531, the target cannot be found again, so AR.Drone automatically goes down and the program is finished. When the target walk in straight, AR.Drone can track the target in correct. When moving direction of the subject changes, it maybe causes the target away from the field of view. Therefore, when the subject turns slowly, he can be tracked by AR.Drone. If the subject turns rapidly, he may be away from the field of view of the AR.Drone.

From frame 406 to frame 495, we can see AR.Drone rises up and search the subject again. Because we set AR.Drone can rise up 10 times as maximum, the altitude of rising up is not high and the field of view is not very wide, so the subject is not rediscovered. In future, we will change the rising times and make the predicted value more and more correct. In addition, we will remove the altitude restriction.

Fig.47 AR.Drone automatically rising when target is lost in midway

IV. Multiple persons

When there are multiple persons in the scene, we would use the Labeling method to separate the subject from these persons explained in previous chapter. Fig.48 shows the subject is tracked automatically by AR.Drone when multiple persons exist. In Fig.48, when another person wearing same color T-shirt is near to the subject, the subject also can be tracked correctly. Because number of experiments is twice, the calculation of success rate is not enough. In future, we will increase number of experiments and collect more data, and then the success rate can be calculated.

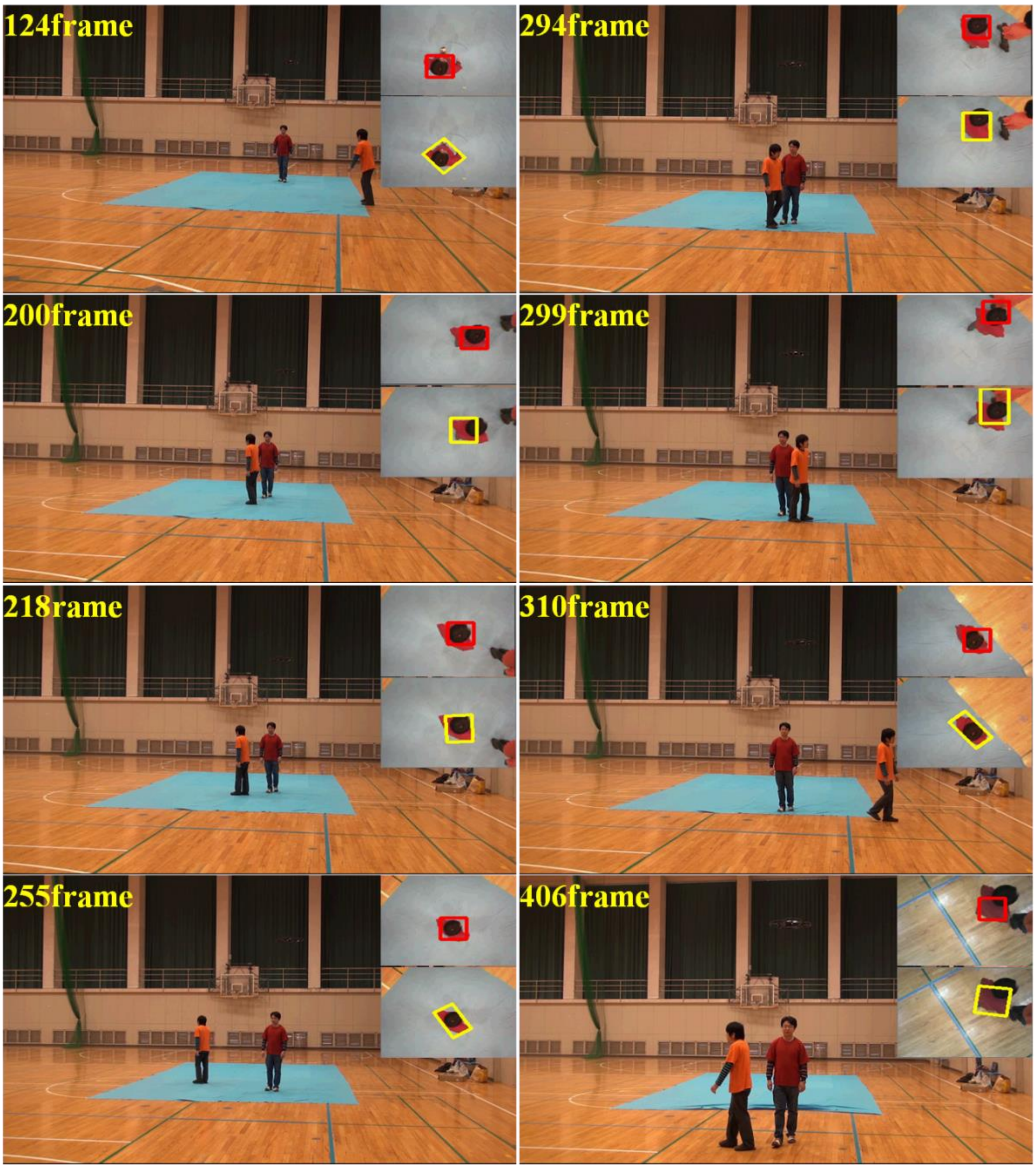

Fig.48 Automatic tracking when multiple persons exist

In this section, we perform automatic tracking experiments in simple environment and in cluttered environment. In each environment, there are short distance cases and long distance. Short distance case is that walking distance is smaller than 20 meters. Long distance case is that walking distance is more than 20 meters. Cases of walking in simple environment are all short distance case, because blue sheet is about 10 meters long. Cases of walking in cluttered environment can be divided into long distance case and short distance case. Results are shown in Table.23.

In Table.23, because cases of walking in simple environment are all short distance case, success rate is about 80.0%. In cluttered environment, we perform experiments of short distance 32 times, and the numbers of success are 22 times, so the success rate is 68.8%. Also, we carry out experiments of long distance 8 times, and the numbers of success are 5 times, so the success rate is 62.5%. When there are multiple persons in the scene, we perform this experiment 2 times, and the numbers of success are 2 times. Because number of experiments is few, the calculation of success rate is not enough. In future, we will collect more data to calculate the success rate in correct.

Table.23 Success rate of automatic tracking (short distance and long distance)

Number of people	Environment	Distance	Number of experiments	Number of success	Success rate
One person	Simple	Short	50	40	80.0%
		Long	0	0	0.0%
	Cluttered	Short	32	22	68.8%
		Long	8	5	62.5%
Multiple persons	Simple	Short	2	2	Not enough
		Long	0	0	0.0%

Chapter 5

Discussions

In our study, the purpose of image processing part is to correctly detect the target and accurately track him at any time. The purpose of flying robot control part is to precisely control aerial robot by using position and moving direction of the subject detected in image processing part. Due to the results of evaluation experiment, the effectiveness of proposed method is verified, and moving direction can be detected automatically and accurately. Also, the Kalman Filter prediction is used to predict the position and moving direction of the subject, and the accuracy is high from the simulation image. In addition, the subject can be automatically tracked in simple environment and cluttered environment by the AR.Drone from results of AR.Drone automatically control. In the experiments, the flight altitude of AR.Drone is about 3 meters in average, and the weight of AR.Drone is light (about 400g). In indoors, AR.Drone can be stabilized flight without outside influence. However, AR.Drone cannot be stable in air with the influence of wind in outdoors. It is difficult to achieve autonomously flight in strong wind, so the experiment is best to be executed in weak wind (below about 4.0m/s in average) condition.

About head detection of the target, to avoid false detection, we use the Hough Transform method and the brightness distribution of detected circle to detect the head part of the subject. Even though there are many circle-like objects with different color to head part on the background, head part of the subject also can be detected correctly. The success rate of head detection in simple environment is 97.2% (at 3m) and 95.8% (at 4m) respectively. The success rate of head detection in cluttered environment is 86.6% (at 3m) and 82.7% (at 4m) respectively. The success rate of head detection in outdoors environment is 77.4% (at 3m) and 72.6% (at 4m) respectively. From failure cases of head detection, we can see head detection is affected by many factors. With the different altitude, the detection accuracy is also different. Moreover, with height increasing, head detection will become difficult, because the radius of head will become smaller. In addition, when we perform experiments in outdoors environment, because of wind, airflow, and so on, aerial robot will become unstable. Therefore, it will cause head detection failed. From results of experiments, success rate in indoors is over 80%, and about 75% in outdoors. Because head is detected in the first stages of experiment if the subject is not lost, this success rate is enough for our experiments. Also, we will try to adjust the threshold to make it more perfect.

As to clothes part detection, we make use of head information detected (such as radius and center) to set initial contour for clothes contour extraction by using the ACM method. From failure cases of clothes detection, it is affected by head detection, initial contour setting and background. If head part is detected incorrectly, it will cause initial contour is wrong, so clothes detection will fail. Also, the size of initial contour will affect the result of clothes detection. If background is simple, after head part is detected correctly, clothes part can be extracted in high accuracy. Conversely, assuming background is cluttered, after head part is detected correctly, if the size of initial contour can be set accurately, clothes part also can be detected correctly. If not, clothes detection will fail. In our study, clothes part of the subject is detected in simple background (blue sheet is on the ground) at the beginning of the experiments. In this case, we can get correct clothes information for human tracking and automatic control of AR.Drone.

About human tracking part, it is an important section of image processing. In this part, we make use of the Mean Shift method to track the head part and clothes part of the subject by using color histogram of each part and position of each part. If any object is near to the subject and the color is similar to head part or clothes part, human tracking maybe fail. When there is only one person in the scene, the success rate in simple environment is 98.7%, in cluttered environment is 94.3%, and in outdoors environment is 90.2%. However, there is some false detection in experiments. In Fig.34, some samples of false detection are shown. One reason is the stability of AR.Drone is not good and communication time of AR.Drone is affected by the distance from AR.Drone to PC (communication distance is about 50m in common), so result is affected by AR.Drone shaking or long communication time. Another is setting of the position threshold (the difference between position in last frame and position in current frame), and we set the position threshold is 30 pixels. It means if the difference between position of last frame and position of current frame is smaller than 30 pixels and color histogram has high similarity, we will say human tracking is correct. Therefore, we will solve these malpractices in future works.

When there are multiple persons in the scene, we use the Labeling method to separate the subject from these persons. The success rate in simple environment is 97.8%, in cluttered environment is 91.2%, and in outdoors environment is 88.3%. From failure cases of human tracking when multiple persons exist, success rate is affected not only by environmental aspects but also by other people. Here, I will give you an example.

In our study, we use the Labeling Method to separate multiple people. On this basis, we use the Labeling Method to find the target with the similar color with clothes part and head part of particular people. After that, we make use of position and direction information in last frame to judge who is the target. Then, the target can be determined. In Fig.36, the labeling image is correct, but human separation is affected by another person, so it will fail. Therefore, we must consider about how to solve the problem of human separation when multiple persons are very near. Also, we can use time series data to solve these problems.

About moving direction detection, when only one person is in the scene, if head part and clothes part can be detected in correct and they can be tracked accurately, moving direction can be calculated correctly. In this part, we use the optimum external rectangular of clothes part and the position of head part to decide the moving direction of the subject (as shown in Fig.19). Therefore, if the external rectangular is wrong or the position of head is wrong, moving direction will be incorrect. In these two factors, the former is more important. The optimum external rectangular of clothes part is got from the result of clothes tracking and influenced by many factors. If the target always exists in the field of view, clothes part can be tracked as shown in Fig.33 or Fig.35. However, if clothes part is tracked failure as shown in Fig.34 or Fig.36, The optimum external rectangular will be wrong. In addition, if the target is lost in midway, we will use the Particle Filter method to detect the clothes part again as shown in Fig.37, because we use particle distribution to decide which the clothes part is, so the size of yellow rectangle in Fig.37 is not same in every frame. In the area of rectangle, if color of some object is similar with clothes, it will be added together to calculate the optimum external rectangular, so it may be wrong (such as cases of single person in cluttered environment and multiple persons in outdoors in Fig.38). Also, the optimum external rectangular would be influenced by viewpoint. Because if viewpoint is different, the acreage of the clothes part seen is also different. Therefore, human tracking and target redetection will affect moving direction detection directly, and we will improve the accuracy of human tracking and target redetection in future. Furthermore, success rate in outdoors is lower than it in indoors, because strong wind will cause AR.Drone shake and unstable. When moving direction is detected, we use the Kalman Filter method to correct it and to predict the moving direction in next time. In our study, we use time series data to correct moving direction and result is shown in Fig.41 and Fig.42. In future, we will improve the accuracy of moving direction detection and Kalman Filter prediction.

When the correct data of moving direction and position is got, AR.Drone would be controlled automatically by these data. This section is very important for our study. At present, AR.Drone I is used for our study. The sampling time of AR.Drone I is about 33.3ms, and image processing time is about 46.6ms in average. According to the position and direction of the subject, AR.Drone can be controlled by adjusting the speed of four propellers. In present, because threshold values of speed of four propellers are not achieved automatic obtaining by the Proportional Control, speed value of four propellers cannot be changed at any time.

Experiments are carried out in simple environment and cluttered environment in indoors respectively. In each environment, it also can be divided in short distance case (<20m) and long distance case (>20m). The standard of success is explained in previous chapter. In simple environment, because background is monochromatic, head part and clothes part can be tracked correctly (as shown in Fig.43). However, if the subject speeds up, he might be away from the field of view. At this time, AR.Drone will search the target again, but the target walks continuously, so it cannot find the target again and it will land automatically. In cluttered background indoors, in general, the subject can be tracked correctly. If the subject is lost in midway, the clothes part will be detected again by using the Particle Filter method. At this time, erroneous detection results might be got (as shown in Fig.38), so it would cause position and moving direction of the subject wrong and AR.Drone cannot track the subject correctly. Therefore, the success rate in simple background is higher than it in cluttered background. Also, in Fig.46, when the target changes his direction suddenly, he might be lost from the field of view. In addition, the communication distance of AR.Drone I is 50 meters. In other words, the distance between AR.Drone and PC is farther, and the transfer time will be longer. It also affects the result of automatic tracking.

In simple environment, because the subject walks on the blue sheet, the distance is the same as the length of blue sheet. Therefore, they are all short distance case. We perform experiments 50 times, and there are 40 times successful, so success rate is 80.0%. In cluttered environment, we perform experiments of short distance case 32 times, and there are 22 times successful, so success rate is 68.8%. In addition, experiments of long distance case are carried out 8 times, and there are 5 times successful, so the success rate is 62.5%. In experiments of long distance case, the longest distance is about 50 meters, and the shortest distance is about 20 meters. Because the speed value of AR.Drone is same (direction is different) and the speed of the subject is different at any time, it also causes the subject lost and automatic tracking fail. In future, we will use the Proportional Control to automatically obtain the threshold value of speed of propellers, and make results of automatic tracking more and more accurate.

In addition, experiments when there are multiple persons exist in the scene are carried out. In simple environment, we perform experiments twice, and there are twice successful, but we cannot say the success rate is 100.0%, because the number of experiments is few. In these two experiments, the speed of the subject is slower than normal walking speed, and the residence time of second person is not long in the scene. In future, we will do a lot of experiments to calculate the success rate of this case, and to test and verify the correctness and effectiveness of automatic tracking when there are multiple persons in the scene.

In the automatic tracking process, if the subject is lost, he cannot know he has been away from the field of view, because AR.Drone is in air and there is not any alarm to prompt the subject. Therefore, when the subject is away from the field of view, AR.Drone cannot find the subject again, because the subject will walk in same direction continuously or change his direction. Therefore, it is better to use mechanical voice or alarm bell to make the subject know he has been away from the field of view in future.

Chapter 6
Conclusions

In our study, a new aerial shoot system for wide area monitoring is proposed, and it is important to achieve practical realization in the future. In addition, a new visual control method by confusing multiple modules to achieve automatic human tracking is proposed. Multiple modules contain the Hough Transform algorithm, the Mean Shift algorithm, the Particle Filter algorithm, the Active Contour Model algorithm, and the Kalman Filter algorithm. They are not simply combined, but effectively confused. Also, a new method for moving direction is proposed, and experiments show that using this method the moving direction can be detected in correct.

In our study, there are three new contributions. Firstly, flying robot is utilized to ensure effective field of view and improve tracking capability. Secondly, moving direction automatic detection is devised. It is possible to detect human moving direction not only in simple environment but also in cluttered environment. Thirdly, prediction control is used to correct moving direction to ensure the correctness of moving direction.

By results of evaluation experiments, when there is only one person in the scene, he can be tracked in high accuracy, and the success rate is 98.7% in simple environment, 94.3% in cluttered environment, and 90.2% in outdoors. Then, his moving direction can be detected correctly, and the success rate is 98.1% in simple environment, 95.7% in cluttered environment, and 93.6% in outdoors environment. In view of this, performances of human tracking and moving direction detection are well. Also, the subject can be tracked automatically by AR.Drone effectively. The success rate of automatic tracking is 80.0% in simple environment and 67.5% in cluttered environment (68.8% in short distance case and 62.5% in long distance case). Automatic tracking is a comprehensive result for the evaluation experiment, and any step might lead automatic tracking to failure. When there are multiple persons in the scene, the subject also can be tracked correctly, and the success rate is 97.8% in simple environment, 91.2% in cluttered environment, and 88.3% in outdoors environment. In addition, his direction can be calculated accurately, and the success rate is 96.5% in simple environment, 93.2% in cluttered environment, and 91.4% in outdoors environment. Furthermore, we have carried out experiments of automatic tracking in simple environment twice, and twice are all successful. Because number of experiments is few, so we cannot say the success rate is 100%, but the subject can be tracked when there are multiple persons exist in the scene.

According to experiment results above, the moving direction of a particular person can be detected automatically and accurately. In addition, the effectiveness of aerial shoot system and proposed visual control method are verified. Furthermore, flying robot can be automatically controlled to track the target, and it further proved correctness and effectiveness of our aerial shoot system.

As to future works, problems stated in previous chapter are to be solved firstly. Then, we will improve the accuracy of flying robot automatic control in simple environment and cluttered environment in indoors. In addition, we will carry out automatic tracking experiments in outdoors environment. In order to use the result of the image recognition to control flying robot smoothly, it is better to improve the capability of flying robot. In present, all experiments are carried out by AR.Drone I. From now on, the AR.Drone II with high resolution cameras and many sensors will be used to carry out evaluation experiments continuously. Furthermore, in order to track fast walking people, the processing time is to be shorten. In addition, we will improve the accuracy of image processing to achieve moving direction accurate detection indoors and outdoors for practical realization. Also, the Kalman Filter theory is utilized to predict moving direction exactly for flying robot control. In present, we just put people as objects in wide area. In future, we will put motionless objects and moving objects all as target, and we will try our best to achieve autonomous monitoring system in wide area as quickly as possible.

References

[1] Takashi Matsuyama, “image processing technology and new deployment”, J.IEE Japan, Vol.120, No .7, pp 402-403, 2000
[2] Takashi Matsuyama, “Past, present and future of visual recognition research”, Eizojoho Industrial, Vol. 30, No.1, pp. 30-32, 1998
[3] http://www.jpo.go.jp/shiryou/s_sonota/map/denki02/1/1-3.htm
[4] Sakauchi Masao, “Dream in the 21st Century of image processing and image application technology”, the Journal of the Institute of Television Engineers of Japan, Vol. 44, No. 11, pp1479, 1990
[5] OAMI Ryoma, ISHIDERA Eiki, UEKI Kazuya, MIYANO Hiroyoshi, FUJITA Mitsuhiro, IHARA Yasuyuki, “Extraction of Human Behavior and Attribute from Surveillance Videos and Its Applications”, IEICE, Vol. 95, No. 5, pp 452-456, 2012
[6] Kohoutek, Tobias. "Analysis and processing the 3d‐range‐image‐data for robot monitoring." Geodezija ir kartografija Vol. 34, No. 3, pp 92-96, 2008
[7] Wang, Zheng, et al., "Design and implementation of multiple-rotorcraft-flying-robot testbed", Robotics and Biomimetics (ROBIO), 2011 IEEE International Conference on. IEEE, 2011.
[8] John D. Anderson, "Langley's Aeronautical Research: A Modern Critique and Reassessment", AIAA Journal, Vol. 35, No. 3, pp. 409-419, 1997
[9] M Leroy Spearman, “SOME AVIATION GROWTH EVENTS”, AIAA 2002-0172, 2002
[10] Shinji Suzuki, “Expectation and Issues for Unmanned Aerial Vehicles”, Journal of Mechanical Systems for Transportation and Logistics (JMTL), Vol. 16, pp 23-28, 2007
[11] www.modelaircraft.org/files/DennyReginald.pdf
[12] Louis J. Rodrigues, “UNMANNED AERIAL VEHICLES”, GAO/T-NSIAD-97-138, 1997
[13] Marius Niculescu, “LATERAL TRACK CONTROL LAW FOR AEROSONDE UAV”, AIAA 2001–0016, 2001
[14] Dang Fen, Wang Minfang, Wang Yinghui, “Development Status and Trends of UAVs”, Land-Based Air Defence Weapons (3), pp 49-54, 2005
[15] J. Everaerts, “THE USE OF UNMANNED AERIAL VEHICLES (UAVS) FOR REMOTE SENSING AND MAPPING”, the International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 37 (2008): 1187-1192.

[16] IWATA Kakuya, on special issue “Aeronautical Robotics”. Journal of Robotics Society of Japan 24(8): 889, 2006
[17] Jean-Christophe Zufferey, Dario Floreano, “Fly-Inspired Visual Steering of an Ultralight Indoor Aircraft”, IEEE TRANSACTIONS ON ROBOTICS, VOL. 22, NO. 1, FEBRUARY 2006
[18] Sivakumar Rathinam, Pedro Almeida, ZuWhan Kim, Steven Jackson, Andrew Tinka, William Grossman and Raja Sengupta, “Autonomous Searching and Tracking of a River using an UAV”, Proceedings of the 2007 American Control Conference Marriott Marquis Hotel at Times Square New York City, USA, July 11-13, 2007
[19] http://www.secom.co.jp/corporate/release/2012/nr_20121226.html
[20] http://blog.livedoor.jp/vanilla7777/archives/35349458.html
[21] P. González, W. Burgard, R. Sanz and J.L. Fernández, “Developing a Low-Cost Autonomous Indoor Blimp”, JOURNAL OF PHYSICAL AGENTS, VOL. 3, NO. 1, JANUARY 2009
[22] Renaud Ott, Mario Guti´ errez, Daniel Thalmann, Fr´ ed´ eric Vexo, “Advanced Virtual Reality Technologies for Surveillance and Security Applications”, Proceedings of the 2006 ACM international conference on Virtual reality continuum and its applications. ACM, 2006.
[23] Kei Terabaru, Masayuki Kashima, Kiminori Sato, Mutsumi Watanabe, “Research of Broad Area Monitoring by Simplified Aerial Image Acquisition System”, IEEJ Transactions on Electronics, Information and Systems Vol.133 No.8 (2013)
[24] Fumiaki Takemura, Masaya Enomoto, Takuma Tanaka, Kazuya Denou, Yasuyuki Kobayashi and Satoshi Tadokoro, “Development of the balloon-cable driven robot for information collection from sky and proposal of the search strategy at a major disaster”, Proceedings of the 2005 IEEE/ASME International Conference on Advanced Intelligent Mechatronics Monterey, California, USA, 24-28 July, 2005
[25] T. H. H. P. J. Burt and A. Rosenfeld, “Segmentation and estimation of image region properties through cooperative hierarchical computation,” IEEE Transactions of Systems, Man and Cybernetics., pp. 802–809, 1981.
[26] I. Corporation, “Open source computer vision library reference manual,” 2001.
[27] J. Everaerts, “THE USE OF UNMANNED AERIAL VEHICLES (UAVS) FOR REMOTE SENSING AND MAPPING”, the International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 37 (2008): 1187-1192, 2008
[28] Casau, Pedro, David Cabecinhas, and Carlos Silvestre, "Transition control for a fixed-wing vertical take-off and landing aircraft." World Congress. Vol. 18. No. 1. 2011

[29] Eisenbeiss, Henri, "Applications of photogrammetric processing using an autonomous model helicopter." Int Arch Photogramm Remote Sens Spat Inf Sci 185 (2006): pp 51-56, 2006
[30] Ambrosia, V., Wegener, S., Sullivan, D., Buechel, S., Brass, S.D.J., Stoneburner, J., Schoenung, S., "Demonstrating UAV-Aquired Real-Time Thermal Data over Fires", Photogrametric Engineering and Remote Sensing 69(4), 391–402, 2003
[31] Merino, L., Caballero, F., Mart´ınez de Dios, J., Ferruz, J., Ollero, A., "A cooperative perception system for multiple UAVs: Application to automatic detection of forest fires", Journal of Field Robotics 23(3-4), 165–184, 2006
[32] Maza, I., Caballero, F., Capitan, J., de Dios, J.M. Ollero, A., "Experimental results in multi-UAV coordination for disaster management and civil security applications", Journal of Intelligent and Robotic Systems 61(1), 563–585, 2011
[33] Maza, I., Caballero, F., Capitan, J., de Dios, J.M., Ollero, A., "A distributed architecture for a robotic platform with aerial sensor transportation and self-deployment capabilities", Journal of Field Robotics 28(3), 303–328, 2011
[34] Suman Srinivasan, Haniph Latchman, John Shea, Tan Wong, Janice McNair, "Airborne traffic surveillance systems: video surveillance of highway traffic", Proceeding VSSN '04 Proceedings of the ACM 2nd international workshop on Video surveillance & sensor networks, 131-135, 2004
[35] Frew, E., McGee, T., Kim, Z.W., Xiao, X., Jackson, S., Morimoto, M., Rathinam, S., Padial, J., and Sengupta, R., "Vision-Based Road-Following Using a Small Autonomous Aircraft" Proc. IEEE Aerospace Conference, 3006-3015, 2004
[36] Seiler, M., and Voorsluijs, G., "Future Surveillance Using Autonomous Unmanned Helicopters", Collect Tech Pap AIAA ICAS Int Air Space Symp Expo Next 100 Years Vol 2, 1773-1782, 2003
[37] Mirchindani, P., Hickman, M., Angel, A., and Chandnani, D., "Application of Aerial Video for Traffic Flow Monitoring and Management", Pecora 15/Land Satellite Information IV/ISPRS Commission I/FIEOS 2002 Conference Proceedings, 346-353, 2002
[38] Yuichiro NISHI, "Current State of UAV R&D and Deployment in the USA", Space Japan Review, No. 74, June / July 2011
[39] Campoy, Pascual, et al, "Computer vision onboard UAVs for civilian tasks." Unmanned Aircraft Systems. Springer Netherlands, pp. 105-135, 2009
[40] Harris, C.G., Stephens, M.: A combined corner and edge detection. In: Proceedings of the 4th Alvey Vision Conference, pp. 147-151 (1988)
[41] Lowe, D.G.: Distintive image features from scale-invariant keypoints. Int. J.

Computer Vision 60(2), 91-110 (2004)
[42] Duda, R.O., Hart, P.E.: Use of the hough transformation to detect lines and curves in pictures. Commun. ACM 15(1), 11-15 (1972)
[43] ZHANG Juan, MAO Xiao-bo, CHEN Tie-Jun, “Survey of moving object tracking algorithm”, Application Research of Computers, Vol. 26, No. 12, 2009
[44] Satoshi Ueno, Hiroshi Sankoh, Sei Naito, “A study on background subtraction method using stereo camera for human extraction”, Forum on Information Technology, Vol. 11, No. 3, pp 155-156, 2012
[45] SHIINA Yuhi , IKENAGA Takeshi, “Difference Information based Particle Filter for Object Tracking in Complex Environment”, Technical report of IEICE. PRMU 111(430), 173-178, 2012
[46] Bo Gao, Zhencheng Hu, Jan Duchacek, Thomas Feraud, Roland Chapuis, “Feature matching and self-calibration strategy for stereo-camera”, IPSJ. ITS, 2013-ITS-55(15), 1-6, 2013
[47] Horiuchi Shuma, Yoshida Shuhei, Ushiyama Zenta, Yamada Takefumi, Yamamoto Manabu, “Improvement of the optical MTF of Rod Lenses using Bayes' Theorem”, Proceedings of the Society Conference of IEICE, 2011-08-30
[48] Horiuchi Shuma, Yoshida Shuhei, Ushiyama Zenta, Yamamoto Manabu, “Image Processing Method of Optical System using Bayes' Theorem”, Proceedings of the IEICE General Conference, 2011-02-28
[49] Tagawa Norio, “Kalman Filter”, The journal of the Institute of Image Information and Television Engineers 67(4), 320-322, 2013-04-01
[50] TRIEU DangBaKhac, MARUYAMA Tsutomu, “An Implementation of Mean Shift Filter on FPGA”, IEICE technical report 111(31), 13-18, 2011-05-05
[51] Amemiya Tomohiro, Hirota Koichi, Ikei Yasushi, “Tactile Apparent Motion on Seat Pan Modulates Perceived Forward Velocity of Self-Motion”, Transactions of the Virtual Reality Society of Japan 18(2), 121-130, 2013-06-30
[52] KAWASHIMA Yuki, FUKUDA Kazuho, UCHIKAWA Keiji, “Determination characteristics for vection yielded by two optical flows”, ITE Technical Report 36(52), 17-23, 2012-11-25
[53] DO Khanh N., OHYA Jun, “Object Tracking Using Kinect and its Application in Tracking Dynamical Behaviors of Botanical Tree Leaves”, Technical report of IEICE. PRMU 111(499), 159-164, 2012-03-22
[54] MATSUMOTO Takuya, HAYASHI Tatsuro, HARA Takeshi, KATSUMATA Akitoshi, MURAMATSU Chisako, ZHOU Xiangrong, IIDA Yukihiro, MATSUOKA Masato, KATAGI Kiyoji, FUJITA Hiroshi, “Automatic method for measuring

mandibular cortical thickness by using active contour model on dental panoramic radiographs”, IEICE technical report. 111(127), 1-5, 2011-07-05
[55] Ganwen Jiang, Masayuki Kashima, Kiminori Sato and Mutsumi Watanabe, "A study on state recognition in wide area by aerial image analysis", Artif Life Robotics(2013), Vol.18, No.3-4, pp.187-195 (2013)
[56] Xu Rong, Ohya Jun, Zhang Bo, Sato Yoshinobu, Fujie Masakatsu G, “Improving Iterative Randomized Hough Transform for Automatic Detection of Fetal Head from Ultrasound Images”, IEICE technical report. Communication systems 111(348), 127-132, 2011-12-08
[57] Rong Xu, Jun Ohya, Bo Zhang, Yoshinobu Sato, Masakatsu G.Fujie, “Improving Iterative Randomized Hough Transform for Automatic Detection of Fetal Head from Ultrasound Images”, IPSJ SIG Notes 2011-AVM-75(24), 1-6, 2011-12-08
[58] CHOI Seokyoon, KIM Hochul, CHOI Khanghee, KIM Mingi, “Improvement of snake model for ventricle segmentation in MRI”, IEICE technical report. 108(385), 165-168, 2009-01-12
[59] HASSANIEN Aboul-Ella, NAKAJIMA Masayuki, “Feature-Specification Algorithm Based on Snake Model for Facial Image Morphing”, IEICE transactions on information and systems E82-D(2), 439-446, 1999
[60] Okada Kazunori, Mean Shift: from Theory to Application. CVIM 2008(27): 401-414, 2008
[61] Yosuke Araya, Yukari Abe, Joji Maeno, “An analysis on labeling digital learning materials in special needs education”, IPSJ SIG Notes 2012-CE-115(2), 1-5, 2012
[62] Kato Takekazu, “Introduction of Particle Filter and Its Implementation”, CVIM 2007(1): 161-168, 2007
[63] LI Jin, KOSUGE Yoshio, MATSUNAMI Isamu, “Tracking Method Using Multiple Kalman Filters without Process Noises”, Technical report of IEICE. SANE 111(407), 55-60, 2012
[64] AR.Drone Developer Guide for SDK 1,7

Acknowledgements

Thanks for Professor Watanabe teaching me various knowledge on computer vision, image processing, pattern recognition, and so on. Also, thanks for Professor Watanabe guiding my thesis, teaching me ways of learning and many life skills. In addition, thank you for Professor Watanabe taking me to attend the International Conference to do oral presentation or poster presentation, and then giving me many suggestion and advice.

Thanks for Professor Kawasaki and Professor Yu guiding my thesis, and teaching me many knowledge not only on professional knowledge and but also on life. Also, thanks for Professor Kawasaki and Professor Yu giving me many suggestion and advice on my study at meeting of seminal in every year. Moreover, thanks for Professor Kawasaki giving me job information to help me search the job.

Thank you for Professor Sato and Assistant Professor Kashima giving me a lot of advice and suggestion for my research, helping me to correct the thesis, and giving me many ideas. In addition, thanks for two Professors taking me to attend conference to discuss my study with other teachers and students. Also, Professor Sato and Assistant Professor Kashima teach me how to learn and how to live, thank you very much.

In addition, I want to thanks for members of Watanabe Group to help my study, teach me the Japanese Culture, and give me a lot of suggestion. Especially, I would like to thank you for Mr. MOHD NORZALI BIN HAJI MOHD, Mr. Koga, Mr. Iwakiri, Mr. Yoshida, Mr. Sakamoto, Ms. Atumari, Mr. Gatayama, Mr. Terabaru, Mr. Fukumoto, and Mr Umetani helping me to carry out experiments. And then, thank you for Mr. Wang and Mr. Fu in Kawasaki Laboratory giving me a lot of suggestion and advice.

Finally, I would like to thanks for Professor Obara taking me to attend SPP Activity in every year, and let me assume TA duties. In SPP activity, I learn much knowledge about electrical and electronic. Also, I want to thanks for my wife-Ms Chen. She has been taking care of my life. When my leg burns, she wholeheartedly takes care of me, and I want to say thank you very much. Moreover, thank for my wife giving me lots of ideas and advice for my study.