International Journal of Innovative Computing, Information and Control
Volume **10**, Number **4**, August **2014**
ICIC International ©2014 ISSN 1349-4198
pp. 1267–1274

A ROBOT VISION SYSTEM FOR VISUAL SURVEILLANCE

Huei-Yung Lin[1,2], Chin-Chen Chang[3,*], Meng-Shiun Yu[1,2] and Horng Wu[1,2]

[1]Department of Electrical Engineering
[2]Advanced Institute of Manufacturing with High-Tech Innovations
National Chung Cheng University
No. 168, University Road, Min-Hsiung, Chiayi 62102, Taiwan

[3]Department of Computer Science and Information Engineering
National United University
No. 1, Lien-Da, Kung-Ching Li, Miaoli 36003, Taiwan
[*]Corresponding author: ccchang@nuu.edu.tw

Received July 2013; revised November 2013

Abstract. *In this paper, we present a robot vision system for visual surveillance based on omnidirectional and PTZ (pan-tilt-zoom) cameras. The proposed system provides not only the surveillance of the surrounding area of the robot, but also observes every corner of the indoor environment. Moreover, the proposed system is mounted on a mobile robot platform, which can move around the environment under the user's command. The applications of the proposed system are the detection and tracking of moving objects in the environment. Since the omnidirectional camera can observe the surrounding regions of the robot, the detection and tracking are more stable. When the omnidirectional camera detects the object, the PTZ camera is used to detect the object's face. The mobile robot can then follow the object for navigation.*
Keywords: Visual surveillance, Robot vision, Omnidirectional camera, PTZ camera

1. **Introduction.** In recent years, surveillance systems have become more and more important [5-7,11,13,17]. Most of the surveillance systems are installed in fixed locations and have simple visual processing capabilities [8]. However, once the interested objects are obscured or in the blind pots, it cannot acquire their image information. Moreover, the whole area monitoring for an environment is not feasible. One way to cope with this problem is to increase the number of cameras for multi-view surveillance [1,9,10]. However, the costs for the hardware and the system development will both increase.

Some previous works have introduced the 360-degree panoramic camera for visual surveillance [3,12,20]. It is capable of monitoring the entire interior space by using just a single camera. However, due to the limited resolution of the panoramic cameras, the detailed information of the objects cannot be identified clearly.

In this paper, we propose a surveillance system consisting of two types of cameras for mobile robots. One is an omnidirectional camera and the other is a PTZ (pan-tilt-zoom) camera. Unlike the conventional wall-mounted surveillance systems, the proposed vision system is installed on a mobile platform. It can move freely in the environment to actively monitor and track the objects. Generally speaking, the omnidirectional camera [16,18] can monitor a wider range area compared with the conventional cameras, but suffer from lower resolution image quality. Thus, if we need to monitor a specific area, PTZ cameras can be used to fulfill this requirement [15].

Our system combines the advantages of the two different cameras and the characteristics of the mobile platform to achieve the purpose of moving object detection and human face tracking. We implement the image processing techniques including moving object detection, tracking, and face detection. In our system, OpenCV library is adopted for face detection. The detected object in the image is classified as a human face if it contains similar facial features. Once the system detects a moving object, it will start the PTZ camera and the mobile platform to track the object. The higher-resolution image capture from the PTZ camera is then used for human face detection.

2. System Overview.

2.1. Mobile robot construction.

The hardware equipment consists of a personal computer and two cameras mounted on a mobile platform, as shown in Figure 1. The images captured by the cameras are processed by the computer and the analyzed data are used to provide the instructions for robot motion. The mobile robot is built with acrylic sheets and four motors, which support all hardware devices, including the camera system and the personal computer.

The visual surveillance system is mounted on the mobile robot platform. It is composed of two different cameras, the omnidirectional camera and the PTZ camera. The omnidirectional camera and the PTZ camera are installed at about 34 cm and 60 cm from the ground, respectively. The personal computer is responsible for image information processing and analysis, and the results are then used for various actions for the robot. The inputs to the personal computer are mainly from the two camera images and emphasize on the two original images for processing.

In this paper, the main objective is to track the person in motion as well as the face detection. Also, the analysis of the original input image will provide the information for sending the control signals to the robot mobile platform and PTZ cameras. It can then control the robot's position for human following and performing the face detection with the PTZ camera.

FIGURE 1. The mobile robot and visual surveillance system

2.2. **Image coordinate transformation.** Since the omnidirectional camera provides 360 degrees FOV (field of view), it is relatively easy to detect the moving objects from the panoramic images. In order to derive the high resolution images for face detection, the location of the object must be estimated for the PTZ camera observation based on the panoramic image coordinates. That is, the orientation of the PTZ camera should be given for close-up image captures of the face.

Some researchers have proposed the coordinate transform between the omnidirectional and PTZ cameras. They calculated the angle between the moving object and a pre-defined reference point in the omnidirectional image, and then converted the angle to a directional vector. The PTZ camera then used the vector to control its orientation to face the moving object. This paper presents an alternative way to improve the coordinate transform. We first expand the panoramic image to a rectangular image and then perform the coordinate transform. In this work, the panning angle of the PTZ camera is from -170 to 170 degrees, and the panoramic image size is 320×240. Thus, the moving distance of a pixel in panoramic image is about 1.0625 degrees for the PTZ camera. We use this transformation to calculate the PTZ camera motion vector for further image processing tasks.

3. **Moving Object Tracking and Human Face Detection.** This section introduces the processes for object detection, tracking [14] and face detection. For the object detection, we will explain how to use continuous image subtraction to detect a moving object, remove the noise through the mathematical morphology, and then build the object color model. In the object tracking part, we use the derived object color model and the CamShift algorithm [2] to keep track of the objects. The location of the object in the panoramic image is then converted to the PTZ camera's panning angle and motion direction of the robot.

The visual surveillance system mounted on the mobile robot can detect the moving objects in the environment and provide the information for the robot to track the object. The proposed system has two operation modes: one is the manual mode and the other is the automatic mode. The manual mode can control the PTZ camera and mobile robot platform. The automatic mode can create your own object model and drive the robot to track the moving object as well as detect the human face.

There are two stages in the automatic mode: object detection and object tracking. The main purpose is to build the object color model in object detection, and then use the CamShift algorithm to track the object.

In the object detection part, we use a simple background subtraction to detect the moving objects in the environment. If the system detects the moving objects in the environment, the moving objects captured in the panoramic images are used to create the object color model. In the object tracking step, we use the object color model and the CamShift algorithm to identify the most suitable image coordinates of the object color model for object tracking.

The main purpose of object detection is to create the color model of the object and then use the CamShift algorithm to track the object. The object detection process can be divided into three steps:

Step 1: Image subtraction: detecting moving objects.
Step 2: Image pre-processing: image binarization and noise removal.
Step 3: Object model creation: put the object region into the color histogram.

3.1. **Moving object detection.** Moving object detection is the first part of the visual surveillance system. In this system, the detected moving object will be used to establish an object model for continuous tracking of the objects. There are many fast and efficient ways to detect moving objects.

In our system, we use a continuous image subtraction to detect the obstacles. Generally, the background subtraction must establish the background image first. Because this system is constructed on a mobile platform, the background may change when the platform moves to another place. The white balance mode for the omnidirectional camera is turned on. When the brightness is changed, the colors of some objects are also altered under this illumination change. That is, we cannot use the background method. Thus, we use the continuous image subtraction to detect the moving object.

3.2. **Image pre-processing.** Image pre-processing is for binarization and noise removal based on the results of the previous stage. Due to the illumination condition change or the camera quality, the image noise will usually appear in the real scene. To avoid the errors caused by the noise, we use erosion and expansion of morphology to remove the noise and get the blob of the moving object. We then calculate the object's size and central location to determine whether the object moves or not and establish the object model.

3.3. **Establish the object color model.** Creating the object color model is to find the direction of the moving object and to track the moving object through the CamShift algorithm. Currently, only limited dynamic tracking methods have been proposed. One way is to use the color information to track the moving objects. After detecting the object in the panoramic image coordinates, the object region is converted into a color histogram. The visual surveillance system can track the object based on the color histogram. The CamShift algorithm finds locations of the most similar color of object in the continuous image sequence. The PTZ camera and the mobile platform can then move to the color information to track the moving objects.

3.4. **Object tracking.** In our system, the object tracking is based on the color information, and the CamShift algorithm is used. The CamShift tracking process is given as follows:

Step 1: The search area is set as the entire image.
Step 2: Initialize the size of the search window.
Step 3: Calculate the color probability distribution in the search window.
Step 4: Use the Mean Shift algorithm [4] to calculate the new search window center and size.
Step 5: Use the results form Step 4 to initialize an image and then go back to Step 3 until completing the object tracking.

The first two steps are setting the parameters in the CamShift algorithm. The third step is based on the histogram of the object color to create the color probability distribution. The Mean Shift algorithm is then carried out to calculate the value until convergence, and find the position of the object.

When the system begins to track the object, we preset three possible situations. Different robot actions are carried out based on these conditions.

1) Object found: When the system finds a moving object, it keeps track of the object by the color histogram.

2) Object too close: If the robot is too close to the object, it stops moving, but still keeps tracking the object.

3) Object not found: If the system loses the target, it returns to the initial state.

We use the size of the object in the image to determine the distance between the object and the robot. If the object is closer to the robot, then it is also closer to the omnidirectional camera. If the objects have the same size, then it appears larger as it gets closer to the camera's lens. The surveillance system uses this method to determine whether the moving object is too close to the robot or not. When the moving object is too close to the robot, the robot will halt in order to avoid hitting the moving object.

3.5. **Face detection.** The face detection method is based on the algorithm [19]. They detected the faces based on the human face features. The advantage of this approach is that the computation time is short, and it is able to detect the different sizes of the facial features. The algorithm is robust under color changes or image zooming. In the implementation, we use the OpenCV image processing library. This library provides instructions and example programs about face detection. In this work we have done some modifications and the process is as follows.
Step 1: Load a database of face classifier.
Step 2: Reduce resolution of the original input image.
Step 3: Through a classifier detect whether any facial features are in image.
Step 4: Output the result.

In order to achieve rapid detection of human face, the image resolution in Step 2 is reduced. The more reduction on the image size implies the less computing time. The human face detection accuracy rate is reduced because of the reduction in image resolution. In Step 3, Ada-Boost classifier is used. This classifier consists of many stages. In the first stage the most representative features of human face are reduced to many non-face images. The second stage is usually complementary relationship with the first stage. For example, if the first stage is to detect the eye features, then the second stage no longer detects the characteristics of the eye, but detects the characteristics of the mouth or nose, etc. When the system finds a human face, the PTZ camera will change the focal length to enlarge the human face image and continue to detect human face. In this way we can closely observe the facial features, and provide the users with identification of the face.

4. **Results.** There are three experiments in this work. Experiment one is the manual mode. We manually control the system hardware, including the PTZ camera and the robot mobile platform. Experiment two is the face detection. The purpose of the experiment is to test whether the PTZ camera can detect human face. When the camera detects the human face, the camera's focal length will change. It is equivalent to enlarging the images. Experiment three includes the face detection and the object tracking. After the robot detects an object, it immediately starts the tracking function to follow the moving object in the environment.

The experiments are carried out mainly in the indoor environment. The room is a rectangular space of about 636 cm × 400 cm, as shown in Figure 2. There is only one entrance and no other obstacles in the room. The hardware used in the experiment consists of the personal computer, video capture cards, motor drive circuits and two analog cameras. The personal computer is equipped with the Intel 2GHz Dual Core CPU and 2GB main memory. There is also an ADLINK image capture card.

When using multiple cameras for image capture, we consider the synchronization problem. All cameras have to capture the images at the same time. The ADLINK image capture card provides a software triggered synchronization function. The resolution of the captured image is 320 × 240. In addition to the ADLINK sample programs and libraries, we also used OpenCV image processing library. This library provides many image processing functions and most of these programs are optimized.

For the real-time systems, the performance of the program is very important. In the hardware control, we use RS232 to send commands to control the PTZ camera and a motor

Figure 2. Experimental environment

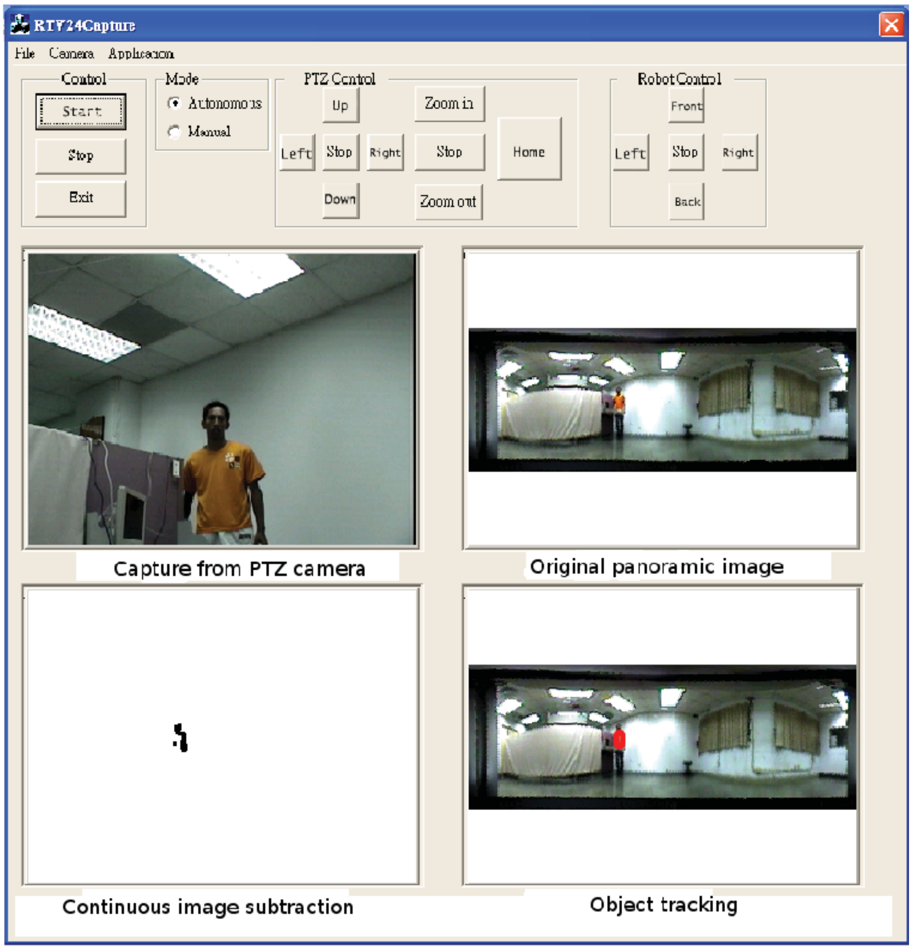

Figure 3. Program interface

drive circuit. The interface of this programming is shown in Figure 3. The functions of the program include start, pause and end modes, selection mode, PTZ control and robot control functions.

In the manual mode, the user can operate the PTZ camera and the robot mobile platform. We can control the PTZ camera's tilting and panning movement and also

FIGURE 4. Robot control

FIGURE 5. The facial detection result

change the focal length. We can also control the robot mobile platform about the direction of movement as shown in Figure 4.

The second experiment is to use the PTZ camera to detect the human face. The size of the real scene is about 636 cm $\times$ 400 cm. The robot is stationary, while the PTZ camera continues to track the moving objects and to detect the human face. After detecting human faces, the focal length of the PTZ camera is changed to provide the detailed observation of facial features, as shown in Figure 5.

Experiment 3 tests whether the robot can drive the motors and track the moving objects in the environment. In the experiment, there is a person moving in environment and staying at the four fixed positions. Then let the robot moves close. If the robot is too close to the moving object, the robot will stop, but PTZ camera will keep tracking the object.

5. **Conclusions.** We have presented a robot vision system which combines the omnidirectional and PTZ cameras. The combination of the omnidirectional and PTZ cameras allows visual surveillance for a wide field of view with detailed observations on the object features. In addition to visual surveillance, we present the applications of moving object detection, tracking, and the facial feature detection.

Our system can achieve the above goals but still need some improvements to increase the system stability. In the future, our system might be integrated to other platforms for more applications and studies, such as robot localization and robot obstacle avoidance.

Acknowledgment. The authors would like to thank the National Science Council of Taiwan for financially supporting this research under Contract No. NSC 98-2221-E-194-041.

REFERENCES

[1] P. Amnuaykanjanasin, S. Aramvith and T. H. Chalidabhongse, Realtime face identification using two cooperative active cameras, *International Conference on Automation, Robotics and Computer Vision*, pp.1-6, 2006.

[2] G. R. Bradski and S. Clara, Computer vision face tracking for use in a perceptual user interface, *Intel. Technology Journal*, 1998.

[3] X. Chen and J. Yang, Toward monitoring human activities using an omni-directional camera, *IEEE International Conference on Multimodal Interfaces*, pp.423-428, 2002.

[4] D. Comaniciu and P. Meer, Mean shift: A robust approach toward feature space analysis, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.24, no.5, pp.603-619, 2002.

[5] A. Elmarhomy and K. Terada, A distance-based method for counting passersby using space-time imagery, *International Journal of Innovative Computing, Information and Control*, vol.9, no.10, pp.4113-4129, 2013.

[6] S. S. Ghidary, Y. Nakata, T. Takamori and M. Hattori, Localization and approaching to the human by mobile home robot, *IEEE International Workshop on Robot and Human Interactive*, 2000.

[7] M. Greiffenhagen, D. Comaniciu, H. Niemann and V. Ramesh, Design, analysis, and engineering of video monitoring systems: An approach and a case study, *Proc. of the IEEE*, vol.89, no.10, pp.1498-1517, 2001.

[8] I. Haritaoglu, D. Harwood and L. S. Davis, W4S: A real-time system for detecting and tracking people in 2 1/2D, *Proc. of the 5th European Conference on Computer Vision*, pp.877-892, 1998.

[9] J. Kang, I. Cohen and G. Medioni, Continuous tracking within and across camera streams, *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, vol.1, pp.267-272, 2003.

[10] H. Kato, S. Inokuchi and A. Nakazawa, Human tracking using distributed vision systems, *International Conference on Pattern Recognition*, vol.1, pp.593-596, 1998.

[11] J. B. Kim and H. J. Kim, Efficient region-based motion segmentation for a video monitoring system, *Pattern Recognition Letters*, vol.24, pp.113-128, 2003.

[12] A. Krishnan and N. Ahuja, Panoramic image acquisition, *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp.379-384, 1996.

[13] C.-Y. Lee, S.-J. Lin, C.-W. Lee and C.-S. Yang, An efficient camera hand-off filter in real-time surveillance tracking system, *International Journal of Innovative Computing, Information and Control*, vol.8, no.2, pp.1397-1417, 2012.

[14] Y. B. Lee, B. J. You and S. W. Lee, A real-time color-based object tracking robust to irregular illumination variations, *IEEE International Conference on Robotics on Automation*, vol.2, pp.1659-1664, 2000.

[15] H. Y. Lin and M. L. Wang, Generalized stereo for hybrid omnidirectional and perspective imaging, *Proc. of the Workshop on Omnidirectional Vision, Camera Networks and Non-Classical Cameras*, Kyoto, Japan, pp.2220-2227, 2009.

[16] V. S. Nalwa, A true omnidirectional viewer, *Technical Report*, Bell Labs, NJ, USA, 1996.

[17] S. Qian, J. K. Tan, H. Kim, S. Ishikawa, T. Morie and T. Shinomiya, Road region estimation and obstacles extraction using a monocular camera, *International Journal of Innovative Computing, Information and Control*, vol.9, no.9, pp.3561-3572, 2013.

[18] S. Shimizu, K. Yamamoto, C. H. Wang, Y. Sato, H. Tanahashi and Y. Niwa, Moving object detection with mobile stereo omnidirectional system (SOS) based on motion compensatory inter-frame depth subtraction, *International Conference on Pattern Recognition*, vol.3, pp.248-251, 2004.

[19] P. Viola and M. Jones, Rapid object detection using a boosted cascade of simple features, *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, vol.1, pp.I-511-I-518, 2001.

[20] M. L. Wang and H. Y. Lin, Object recognition from omnidirectional visual sensing for mobile robot applications, *IEEE International Conference on Systems, Man, and Cybernetics*, San Antonio, TX, USA, pp.2010-2015, 2009.