Automatic Zooming Mechanism for Capturing Clear Moving Object Image Using High Definition Fixed Camera

Hsien-Chou Liao*, Po-Yueh Chen**, Zi-Jun Lin**, Zi-Yi Lim*

*Department of Computer Science and Information Engineering, Chaoyang University of Technology, 168, Jifeng E. Rd., Wufeng District, Taichung, 41349 Taiwan, R. O. C.

**Department of Computer Science and Information Engineering, National Changhua University of Education, No.1, Jinde Rd., Changhua City, Changhua County, 500 Taiwan, R. O. C.

seafoodliao@gmail.com, pychen@cc.ncue.edu.tw, beiwue2006@gmail.com, joslimzy@gmail.com

***Abstract*—High definition (HD) camera is widely used in surveillance systems. An HD camera with optical zoom is useful for monitoring a large area. However, it is inconvenient for a user to manually control the optical zoom for a long time. To exploit the functionality and extend the application domains of a HD camera, the zooming should be controlled automatically. Therefore, an automatic zooming mechanism is proposed in this paper. When the number of an object is small in the field of view (FOV) of the camera and an object is moving through the FOV, the zoom is controlled for capturing the object as clear as possible. A clear object image is useful for related image-based services, such as face recognition. In order to achieve the above goal, a Gaussian Mixture Model (GMM), temporal image differencing, a CamShift tracking method, and a Kalman filter are utilized for object detection and tracking. Then, an adaptive neuro-fuzzy inference system (ANFIS) is used to learn and determine a suitable value for adjusting the zoom. According to the experimental study of the prototype, the results show that the proposed mechanism is useful to capture the clear images of moving objects in a practical environment. A face detection algorithm is also used to demonstrate the feasibility of the captured clear images.**

***Keywords*—Object tracking, Surveillance system, Intelligent video surveillance, Neural network**

I. INTRODUCTION

Surveillance systems are widely adopted for security and safety consideration. In most of cases, surveillance cameras are connected to the internet and called IP (Internet Protocol), or network cameras. The development of hardware technology also increases the image quality of a camera to high definition (HD). An HD quality image is useful for many related image-based functions, such as face recognition. There are mainly two types of camera commonly used in a surveillance system, fixed cameras and active cameras. A fixed camera can only be used to monitor a fixed area. On the other hand, an active camera is equipped for pan/tilt/zoom operations. Such a camera can be used to monitor a large area. However, an active camera is inconvenient to be manually controlled for a long time period. These two types of cameras are shown in Fig. 1. A control mechanism must be designed for specific applications, e.g., moving object tracking, smoke/flame detection.

(a) (b)

Fig. 1. Two types of cameras (a) fixed camera (b) active camera

In this study, an automatic zooming mechanism is proposed for tracking a moving person and capturing clear images, using a HD fixed camera with an optical zooming lens. The mechanism enables the facility to capture a clear image of a moving person during the tracking period. An example shown in Fig. 2 is used to illustrate the implemented system. A fixed HD camera is used to monitor a square at a distance. The distance is about 100 meters in the experimental site. When a person is moving through the square, the person is blurred and difficult to identify from his face if the zoom is also fixed at wide-angle. If the zooming function can be controlled to track the person as shown in Fig. 2(a), a series of images as shown in Fig. 2(b) can be easily captured. For these six captured images, the leftmost and rightmost images are still blurred. However, two clear images can be captured as shown in the series of the images. The clear images are useful for the related image-based functions or services, such as face recognition. Basically, the HD camera is controlled to zoom in when a person is moving toward to the center of the square. Oppositely, the camera is controlled to zoom out when the person is moving away from the center of the square.

In order to achieve the above purpose, an adaptive neuro-fuzzy inference system (ANFIS) is used to design the zoom controller in the zooming mechanism. The zoom value and location of the moving person in a real-time image are input to the ANFIS. Accordingly, a suitable value for adjusting the zoom is derived and submitted to the camera. The zoom is controlled to be as large as possible

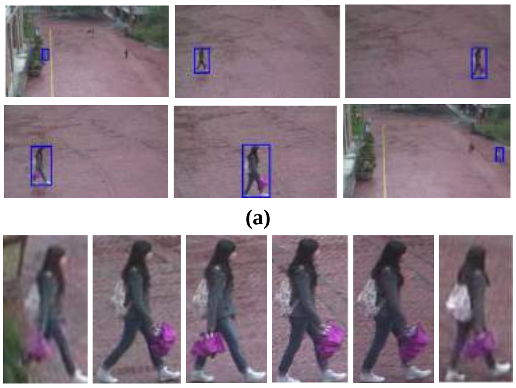
Manuscript received March 7, 2016. This work is supported by Ministry of Science and Technology, Taiwan, R.O.C., under Grant MOST 103-2221-E-324-016.

Hsien-Chou Liao is with the Department and Graduate Institute of Computer Science and Information Engineering, Chaoyang University of Technology, Taichung, Taiwan, R.O.C. Post code: 413. (Corresponding author, phone: +886-4-2332-3000~4211; fax: +886-4-2374-2375; e-mail: seafoodliao@gmail.com).

Zi-Yi Lim is with Chaoyang University of Technology, Taichung, Taiwan, R.O.C. Post code: 413. (phone: +886-4-2332-3000~7713; fax: +886-4-2374-2375; e-mail: joslimzy@gmail.com).

Po-Yueh Chen and Zi-Jun Lin are with the Department and Graduate Institute of Computer Science and Information Engineering, National Changhua University of Education, Taiwan, R.O.C. Post code: 500. (e-mail: pychen@cc.ncue.edu.tw, beiwue2006@gmail.com).

and to capture the image as clear as possible.



(a)

(b)

Fig. 2. The demonstration of the automatic zooming mechanism (a) moving person tracking (b) captured images

The remainder of this paper has been organized as follows: Section 2 reviews related work; Section 3 presents the method of the proposed zooming mechanism; Section 4 describes the experimental study; and Section 5 presents the conclusion and the need for future work as indicated by the results of this study.

II. Related Works

Several previous studies related to HD cameras are discussed here. Methods or systems were proposed for various requirements under some specific environments. These studies are helpful to increase the functionalities of HD cameras. For example: S. C. Chan et al. proposed an intelligent video surveillance (IVS) system in a multi-HD camera environment [1]. Every HD camera was connected to a proposed FPGA (Field Programmable Gate Array) board. This board was designed to handle some basic functions of pattern recognition, such as background model establishment, foreground object detection and tracking. The results were then transmitted to a server for handling advanced analysis functions with high-computing needs, such as consistent labeling of objects cameras or computing of image depth information.

G. Scotti et al. proposed a pedestrian classification system, called panoramic scene analysis (PSA) system, by integrating omni-directional cameras and HD PTZ (Pan-Tilt-Zoom) cameras [2]. The PTZ camera can monitor a large area by using its rotation and zooming functions. It can also provide HD images of pedestrians. Firstly, PSA system transformed the polar image to a wide-view panoramic image. Then, the system classified the moving objects into pedestrians or vehicles. The PSA system controlled the PTZ camera to track all the pedestrians

Besides, M. S. Sayed and J. G. R. Delva proposed an efficient intensity correction algorithm for HD images [3]. The trend of surveillance systems is toward low cost, high efficiency, and high resolution. Many intelligent surveillance functions, such as intrusion detection, should be built in the HD camera for providing efficient response or alarm. However, most of the function can only perform well under fixed light condition. Therefore, the proposed intensity correction algorithm is useful to increase the robustness of the intelligent functions. The correction can be either global or local. An apparent gain factor is defined for intensity correction. The algorithm is also realized in Xilinx Spartan3A digital signal processor (DSP) XC3SD3400A device. The results showed that the algorithm can process 1080p (1920×1080) images at 30 frames per second (FPS).

According to the above studies related to HD cameras, the method or system design must take into account the high computing load on processing HD image. In this paper, the design of zooming mechanism not only speeds up the processing of HD images, but also integrates the tracking technique and fuzzy controller for an HD camera with optical zoom in order to increase the feasibility of HD camera in the practical environment.

III. Method

In order to design an automatic zooming mechanism for a HD fixed camera, two basic steps are performed. One is foreground object detection and tracking, the other is zoom control. After several attempts to achieve the desired results, the final process is depicted in Fig. 3. Firstly, the foreground moving objects are detected. When an object is detected, its template is extracted for tracking. CamShift tracker is used in this process. Then, the location and moving direction of the object is estimated and input to the ANFIS zoom controller. A suitable delta zoom value for zoom-in or zoom-out is submitted to the HD camera. These main steps are presented in the following subsections.

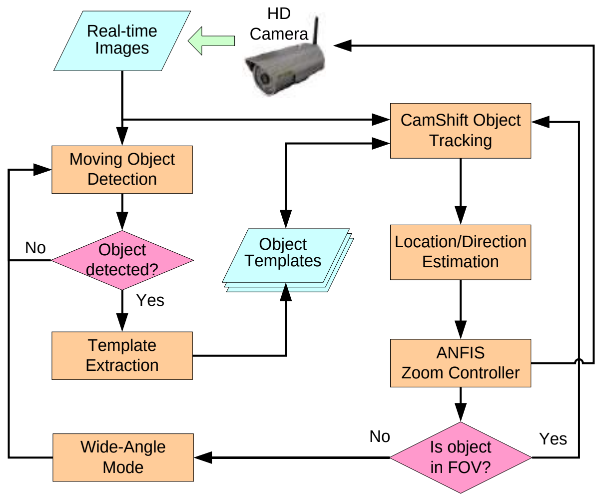


Fig. 3. The process of the proposed automatic zooming mechanism

A. Object detection phase

Foreground object detection is the first step for zooming control. Two popular methods are used in this step, background subtraction and temporal differencing. In this paper, the object detection is performed only in the wide-angle mode. In order to avoid the influence of the environmental brightness, the Gaussian Mixture Model (GMM) is used to construct the background model. Then, the real-time image is subtracted with the background image to detect the foreground objects. Several morphological operations are also used to generate clear foreground objects.

In GMM, a set of Gaussian distributions is used to represent the distribution of pixel values [4]. The model can be represented using Eq. (1).

$$p(x_N|\lambda) = \sum_{i=1}^{M} w_i g_i(x_N) \tag{1}$$

Here, $p(x_N|\lambda)$ is formed by M probability density functions of GMM. Notation x represents the input data and N is the number of input data. λ is formed by $\{w_i, \mu_i, \Sigma_i\}, i = 1,2, \ldots, M$, where w_i is mixture weighted value, μ_i is the average vector, Σ_i is covariance matrix. $g_i(x)$ represents the i-th density function of the Gaussian distribution. The initial values of M, μ, Σ, and w are determined by using the K-means algorithm. Then, Maximum likelihood estimation (MLE) method is used for updating the values to obtain the best results. The density function of GMM is represented in Eq. (2).

$$E(\lambda) = \ln(\prod_{i=1}^{n} P(x_i)) = \sum_{i=1}^{n} \ln P(x_i) \tag{2}$$

Then, expectation maximization (EM) algorithm is used to find the maximal value iteratively, as shown in Eq. (3).

$$\beta_m(x) = p(m|x) = \frac{w_m g(x|\mu_m, \Sigma_m)}{\sum_{j=1}^{M} w_j g(x|\mu_j, \Sigma_i)} \tag{3}$$

$\beta_m(x)$ is the generating function of m-th Gaussian distribution. w_m, μ_m, and Σ_m are the parameters to be estimated and can be obtained by Eq. (4)-(6):

$$\mu_m = \frac{\sum_{i=1}^{n} \beta_m(x_i) x_i}{\sum_{i=1}^{n} \beta_m(x_i)} \tag{4}$$

$$\Sigma_m = \frac{\sum_{i=1}^{n} \beta_m(x_i)(x_i - \mu_m)(x_i - \mu_m)^T}{\sum_{i=1}^{n} \beta_m(x_i)} \tag{5}$$

$$w_m = \frac{1}{n} \sum_{i=1}^{n} \beta_m(x_i) \tag{6}$$

Finally, the results of the above three parameters are applied to Eq. (3) to check whether or not the convergence condition is satisfied. If it is not satisfied, the iteration of EM algorithm is repeated.

An example of foreground object detection is shown in Fig. 4. The background image constructed by using GMM is shown in Fig. 4(a). The real-time image shown in Fig. 4(b) is subtracted by the background image and the result is shown in Fig. 4(c), which contains some noise or fragments. After the processing of morphological operations, including erosion and dilation, the final result is shown in Fig. 4(e), which is a clear detection.

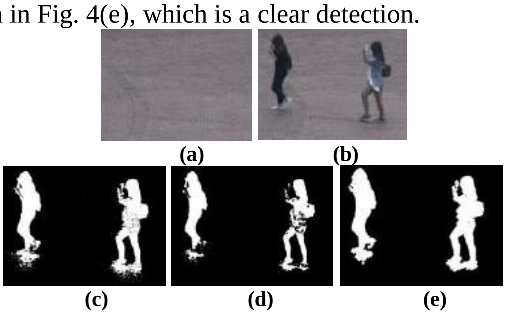


Fig. 4. An example of foreground object detection (a) GMM background image (b) real-time image (c) background subtraction (d) erosion (e) dilation

B. Object tracking phase

After an object is detected, the next step is object tracking. CamShift tracking algorithm is used in this step. CamShift algorithm combines basic mean-shift algorithm with an adaptive region sizing step [5]. In general, the size of a moving object is varied during tracking. CamShift algorithm can reduces the error caused by the change of object size in the mean-shift algorithm. It searches the most likely region, which is most similar to the template, in an area. Assume the template image at time t is denoted as F_t. The probability distribution of F_t is denoted as $P(t)$. The centroid of F_{t-1} and $P(t)$ is used to search the new location of the template in the CamShift algorithm. In order to achieve the above goal, the zero order moments (M_{00}) and the first order moments (M_{01} and M_{10}) are estimated by using Eq. (7)-(10):

$$M_{00} = \sum_y^h \sum_x^w I(x, y) \tag{7}$$

$$M_{10} = \sum_y^h \sum_x^w x I(x, y) \tag{8}$$

$$M_{01} = \sum_y^h \sum_x^w y I(x, y) \tag{9}$$

$$X_c = \frac{M_{10}}{M_{00}}, \; Y_c = \frac{M_{01}}{M_{00}} \tag{10}$$

$I(x, y)$ is the pixel intensity at the location (x, y) of F_t. h and w represent the height and width of the search area, respectively. X_c and Y_c represent the moving distance in the X and Y direction, respectively. The above process is repeated until the convergence of the location (X_c, Y_c) is achieved. An example of such a convergence is shown in Fig. 5.

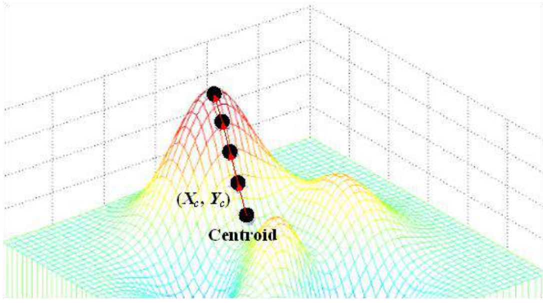


Fig. 5. Convergence of the location (X_c, Y_c) in the CamShift algorithm

A series of (X_c, Y_c) can be obtained in the above CamShift algorithm. However, the location is unstable due to the influence of noise or environment brightness. So, a Kalman is used to smooth the tracking results. The filter is described in the following equations [6].

$$\hat{x}_k^- = A\hat{x}_{k-1} + Bu_{k-1} \tag{11}$$

$$P_k^- = AP_{k-1}A^T + Q \tag{12}$$

$$K_k = P_k^- H^T (HP_k^- H^T + R)^{-1} \tag{13}$$

$$\hat{x}_k = \hat{x}_k^- + K_k(z_k - H\hat{x}_k^-) \tag{14}$$

$$P_k = (I - K_k H)P_k^- \tag{15}$$

where $\hat{x}_k^-$ is a priori state estimate at step k given knowledge of the process prior to step k, $\hat{x}_k$ is a posteriori state estimate at step k given measurement. z_k The matrix A relates the state at the previous time step k-1 to the state at the current step k, and matrix B relates the optional control input u_k to the state x. P_k is estimate error covariance, K_k is Kalman gain, Q is process noise covariance, and R is the measurement noise covariance. Eqs. (11) and (12) are time update equations. Eqs. (13)-(15) are measurement update equations. An application of Kalman filter is shown in Fig. 6. A series of y coordinates is processed by the Kalman filter. The original coordinates in blue are smoothed into the curve in red.

Fig. 6. An application of Kalman filter (parameters Q=0.001 and R=0.0001)

C. Zooming mechanism

For the zoom control of an HD fixed camera, the camera should be zoomed in when the object is moving toward the center of FOV. Oppositely, the camera should be zoomed out when the object is moving away from the center of FOV. Therefore, a fuzzy controller is used for the zooming mechanism. In advance, the settings of fuzzy rules and membership functions are the main challenge on designing the fuzzy controller. The adaption of a fuzzy controller from one environment to another is also an issue. According to the above considerations, an adaptive neuro-fuzzy inference system (ANFIS) is used on the design of automatic zooming mechanism [7]. The first-order Sugeno fuzzy model is used in ANFIS. The structure of ANFIS is shown in Fig. 7.

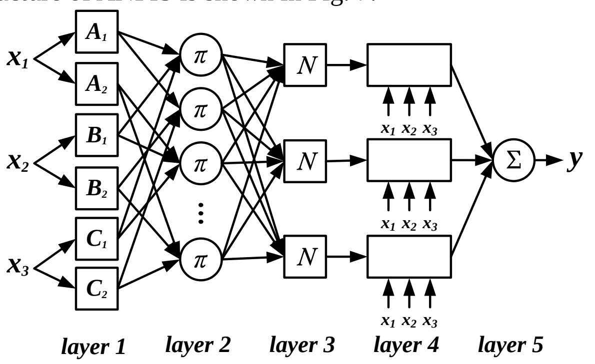


Fig. 7. The structure of ANFIS

ANFIS is a five-layered structure. Every layer is described as follows:

1) The first layer is input layer. Every node in this layer is an adaptive node with a node function represented in Eq. (16).

$$O_{1,i} = \mu A_i(x), \text{for } i = 1,2, \text{or}$$
$$O_{1,i} = \mu B_{i-2}(y), \text{for } i = 3,4 \quad (16)$$

where $O_{1,i}$ is the membership value of the input x or y corresponding to the fuzzy set A_i or B_i. The parameters in this layer are called premise parameters.

2) The second layer is rules. All the nodes are fixed node, denoted as π. Its output equals the multiplication of all the input as represented in Eq. (17). The output represents the strength of the rule.

$$O_{2,i} = w_i = \mu A_i(x)\mu B_i(y), i = 1,2 \quad (17)$$

3) The third layer is the normalization layer. The nodes are denoted as N. In a node i, the ratio of the strength of i-th rule to the strength of all the rules are computed using Eq. (18). The output is the normalized firing strength.

$$O_{3,i} = \overline{w_i} = \frac{w_i}{w_1+w_2}, i = 1,2 \quad (18)$$

4) The forth layer is inference layer of consequence. The node i in this layer is also an adaptive node with the node function represented in Eq. (19). $\overline{w_i}$ is the output and $\{p_i, q_i, r_i\}$ is the parameter set. All the parameters are called consequence parameters.

$$O_{4,i} = \overline{w_i} f_i = w_i(p_i x + q_i y + r_i), i = 1,2 \quad (19)$$

5) The fifth layer is output layer. There is only one node in this layer denoted as Σ. The output is the summation of all the outputs in the previous layer as represented in Eq. (20).

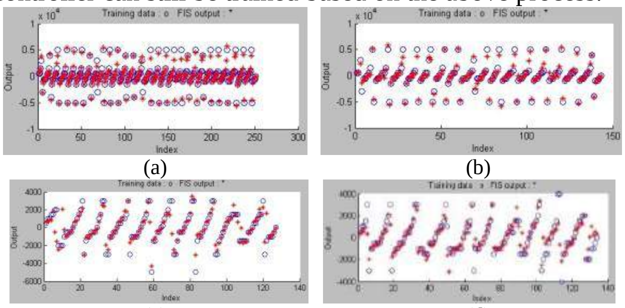
$$O_{5,1} = \sum_i \overline{w_i} f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \quad (20)$$

Although ANFIS is more complicated than the conventional fuzzy inference system, it cannot be applied to any fuzzy inference applications unless the following criteria are satisfied:

1) The ANFIS structures must be the first-order or zero-order Sugeno models.
2) Singleton output: A weighted average method is used for defuzzification.
3) Different rules cannot be applied to the same membership function.
4) The weight of all the rules is the same.

Sugeno models apply polynomial functions to construct the consequent values. A weighted average method defuzzification method is used to generate the final result.

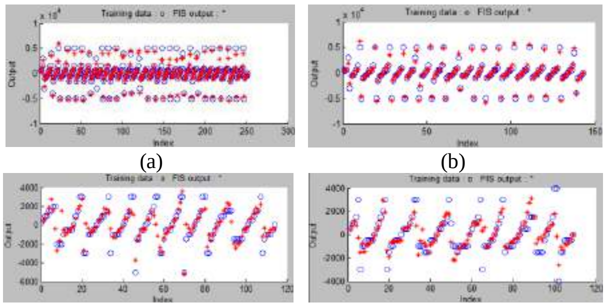
The construction of the fuzzy inference system (FIS) is based on the subtractive clustering method. The parameters, including range of influence, squash factor, accept ratio, and reject ratio are set to 0.2, 1.25, 0.5, and 0.15, respectively. When the FIS is initialized, a set of a series of zoom values is collected. During the period of an object moving through the monitoring area of HD camera, the zoom values of manual control are recorded. Then, a fuzzy error backpropagation algorithm is used to learn the fuzzy control rules membership functions. The errors are propagated backward from layer to layer through the node links. A steepest descent method is used to decrease the error by updating the weights in the neurons of the different layers. The parameters, error tolerance and epochs, are set to 0.001 and 10000, respectively. According to the moving directions of foreground objects, four fuzzy controllers are learning separately. The inputs of the controller include the current zoom value and location. The training results are shown in Fig. 8. The blue circles represent the training data and the red circles represent the training results. There are 600 training data, including 200, 200, 100, and 100 data for left, right, up and down directions separately. When an HD camera is installed at a different environment, a fuzzy controller can still be trained based on the above process.



(a) (b) (c) (d)

Fig. 8. The training data and results based on zoom value and location (a) up (b) down (c) left (d) right

Another set of fuzzy controllers are also trained using the zoom value and moving speed of an object as inputs. The training data and results are shown in Fig. 9. There are four controllers for different directions. However, the zoom value is not close related to the moving speed. Therefore, the errors of these controllers are larger than the previous controllers shown in Fig. 8. The maximal zoom value can be achieved during the moving period is smaller than the previous controllers. It causes the captured image is also less clear than that captured by using the previous controllers. So, the previous controllers are chosen in the design of zooming mechanism.



(a) (b) (c) (d)

Fig. 9. The training data and results based on zoom value and moving speed (a) up (b) down (c) left (d) right

IV. Experimental Study

An experiment is designed to evaluate the performance of the automatic zooming mechanism. The experiment design and results are presented in the following subsections.

A. Experiment Design

The experiment is conducted on a commercial HD camera (AXIS Q1755) and a prototype implemented using Visual C# 2012, MATLAB R2012a and the EmguCV. The screenshot of the prototype is shown in Fig. 10. The prototype is executed on a PC with i3-550 3.2GHz CPU and 6GB RAM.

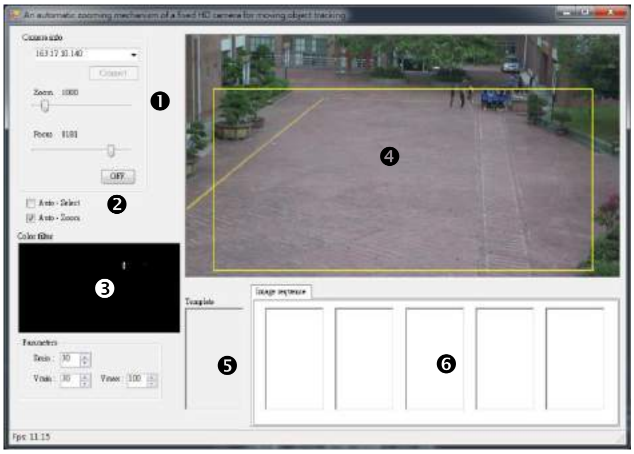


Fig. 10. The screenshot of the prototype

The area marked with ❶ is the basic control of the HD camera. The area marked with ❷ is the functions of the automatic zooming mechanism. The checkbox “Auto-Select” enables the system to select a moving object automatically. The checkbox “Auto-Zoom” enables the automatic zooming mechanism. The area marked with ❸ is the result of the foreground object detection and the setting of parameters. The area marked with ❹ is the real-time image of HD camera. The area marked with ❺ is the tracking template of an object. The area marked with ❻ is the captured images of a moving object under different zooms.

There are totally 70 tests in the experiment. 50 tests are objects moving in horizontal direction and 20 tests are objects moving in vertical direction. The tests of horizontal moving are larger than those of vertical moving because most of the objects are moving in horizontal direction. A moving object is selected randomly in a square and the automatic zooming mechanism is enabled. The zooming mechanism adjusts the zoom every 600 and 800ms for horizontal and vertical moving, respectively. For the vertical moving of objects, the frequency of zoom adjustment is lower because the object size and location is more stable, compared with the horizontal moving. For further analysis, the images of the tracking object are captured before the zoom adjustment as well.

B. Experiment Results

When the automatic zooming mechanism of the prototype is enabled, the zoom of the HD camera is controlled to capture the object image as clear as possible. Two examples of the captured images with two different moving directions are shown in Fig. 11 and Fig. 12. The image resolutions are increasing when the object is moving toward the center and then decreasing when leaving the center.

(a) (b) (c) (d) (e) (f)

Fig. 11. A series of images captured when the object is moving left (a) 0s (b) 3.8s (c) 5.2s (d) 5.3s (e) 8.9s (f) 11.8s

(a) (b) (c) (d) (e) (f) (g)

Fig. 12. A series of images captured when the object is moving down (a)0s (b)1s (c)4s (d)5.1s (e)6.6s (f)7.7s (g)9.7s

The results are analyzed based on the captured images collected in the experiment. There are three results described below:

1) The processing time: The average processing time of all steps are computed. An HD image is usually resized to shorten the processing time.
2) The maximal zoom value: For every tracking, the maximal zoom value is recorded and counted to realize the performance of the automatic zooming mechanism. A higher zoom value means that the zooming mechanism is effective to control the optical zoom of an HD camera.
3) The maximal height of an object: The larger the height of an object is, the clearer is the object quality. Therefore, the maximal height of an object is recorded and counted.

Firstly, the processing time of an image in each step is recorded. In order to reduce the computing load of an HD image, the image is resized to 640×480. The average time of the foreground object detection, CamShift tracking, and zoom control is 7.96 ms, 3.11 ms, and 2.65 ms, respectively. It shows that the proposed zooming mechanism is efficient.

The largest zoom value of the HD camera is 10,000. The maximal zoom values of object tracking are recorded for horizontal and vertical moving separately. These values are summarized by counting the ratio of the maximal values larger than a specific value. The results are depicted in Fig. 13. For the curve of the horizontal moving, it shows that all the maximal zoom values are larger than 3,000 (100%). The ratio remains high (92%) for the zoom value 6,000. For the zoom value 7,000, the ratio is decreased to 80%. For the vertical moving, the object in the image is stable. The zooming mechanism is able to achieve higher zoom than the horizontal moving. So, the ratio of vertical moving is always higher than the ratio of horizontal moving. The results show that the zooming mechanism can achieve a quite high zoom value for capturing clear images.

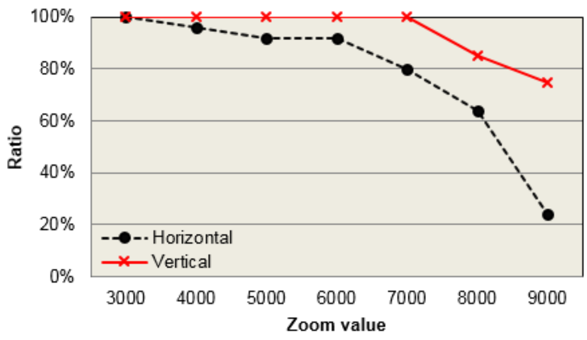


Fig. 13. The percentage of the maximal zoom values

The maximal height of an object is also recorded. The distributions of the maximal heights of two kinds of moving directions are shown in Fig. 14. For the horizontal moving, most of the heights are within 350 to 500 pixels. Oppositely, most of the heights are within 500 to 600 pixels for the vertical moving. It is caused by the same reason that vertical moving is more stable.

Some examples of the captured images with the maximal height are shown in Fig. 15(a) and (b). They illustrate that the object images can be identified clearly.

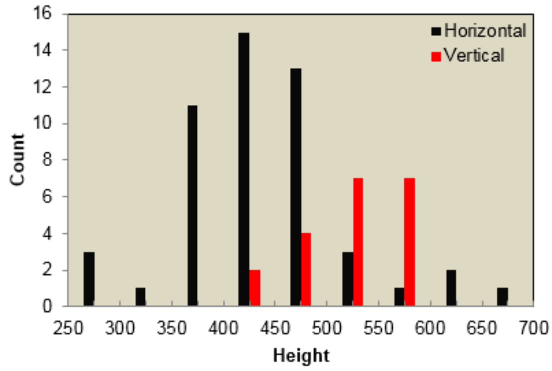
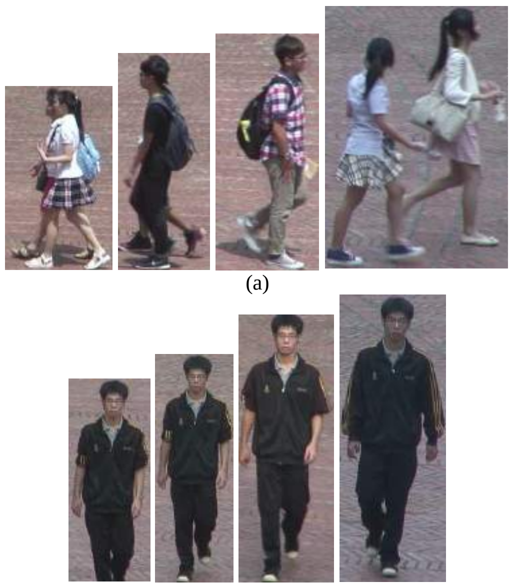


Fig. 14. The distributions of the maximal heights



(a)

(b)

Fig. 15. The captured images with various heights (a) horizontal (b) vertical

Besides, 20 captured images with maximal zoom value of the vertical moving are processed for face detection algorithm. The detection algorithm is based on the skin regions and the geometric constraints, such as ratio. The detection algorithm is implemented by using OpenCV library. All of them can be detected successfully as shown in Fig. 16 since their zoom value is larger than 6,000. Several examples are also shown in Fig. 17. When the face is detected, the detection result is marked by a red circle. These results demonstrate that the proposed automatic zooming mechanism is useful to capture clear images for related image-based services, such as face recognition.

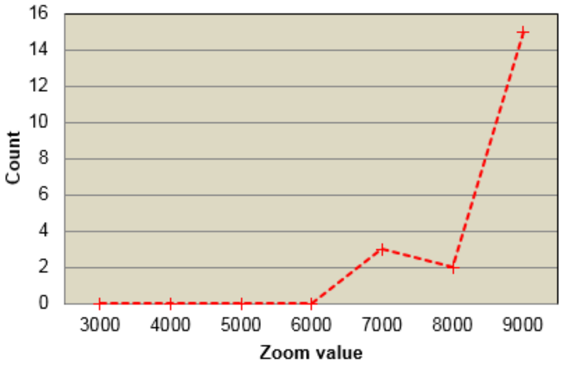


Fig. 16. The zoom values of successful face detection

Fig. 17. Examples of the captured images applied to face detection

Besides, the time spans between two successive zoom adjustments are also recorded. The distributions of time spans for horizontal and vertical directions are shown in Figure 18. The x-axis represents the time span in milliseconds and the y-axis represents the counts of adjustments. The time spans of horizontal direction are about 1,400 to 1,600 milliseconds. The time spans of vertical directions are about 1,700 to 1,900 milliseconds. The moving of an object in the vertical directions is stable and causes the time spans can be longer than that of horizontal direction. The time span is also related to the training data of ANFIS. If the time spans of manual zoom adjustments can be shorten, the time spans of the automatic zooming mechanism can be shorten, too. The maximal zoom values will be possibly higher than the current mechanism.

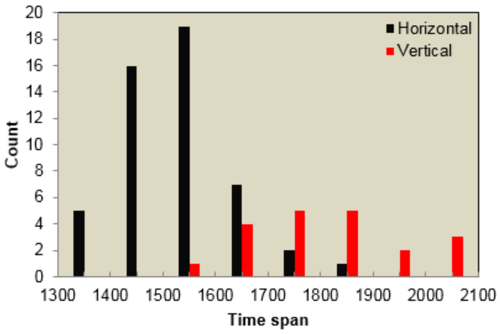


Fig. 18. The distribution of time span between zoom adjustments

V. Conclusion

HD cameras are widely used in a surveillance system. In this paper, an automatic mechanism is proposed for the HD fixed camera with zooming capability. When an object is moving through the FOV of the camera, a clear image can be captured and preserved. The mechanism consists of the object detection, CamShift tracking, ANFIS algorithm, and fuzzy controller. ANFIS is used to help the fuzzy controller learn the fuzzy rules and membership functions from the manual control data. The mechanism can be used in a new environment based on the same learning process.

According to the experiment results, the maximal zoom values of 80% tests can reach 7,000. It ensures that a high resolution image of the moving object can be captured. Furthermore, the average processing time of the mechanism is efficient, about 13.7ms. The face can also be detected successfully in the captured images. It demonstrates that the images captured by the automatic zooming mechanism are feasible for image-related services. Many related services, such as license plate recognition (LPR), can be applied to such HD fixed cameras.

In the experimental study, it is found that the zooming mechanism is directly influenced by the object tracking. The tracking accuracy will be enhanced in the future to increase the performance of the mechanism.

References

[1] S. C. Chan, S. Zhang, J. F. Wu, H. J. Tan, J. Q. Ni, and Y. S. Hung, "On the Hardware/Software Design and Implementation of a High Definition Multiview Video Surveillance System," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, Vol. 3, No. 2, June 2013, pp. 248-262.

[2] G. Scotti, A. Cuocolo, C. Coelho, and L. Marchesotti, "A Novel Pedestrian Classification Algorithm for a High Definition Dual Camera 360 Degrees Surveillance System," *The IEEE International Conference on Image Processing* (*ICIP 2005*), Genoa, Italy, 11-14 Sept. 2005, Vol. III, pp. 880-883.

[3] M. S. Sayed and J. G. R. Delva, "An Efficient Intensity Correction Algorithm for High Definition Video Surveillance Applications," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 21, No. 11, Nov. 2011, pp. 1622-1630.

[4] S. C. Jeng, "A GMM-based Method for Dynamic Background Image Model Construction with Shadow Removal," *Master Thesis, National Chiao-Tung University, ECE*, June, 2005, pp. 83.

[5] G. R. Bradski, "Real Time Face and Object Tracking as a Component of a Perceptual User Interface," *The Fourth IEEE Workshop on Applications of Computer Vision* (*WACV'98*), 1998, pp. 214-219.

[6] G. Welch and G. Bishop. "An Introduction to the Kalman Filter," *Technical Report, TR95-041, University of North Carolina, Chapel Hill*, 2004, 16 pages.

[7] J. S. R. Jang, "ANFIS: Adaptive-Network-Based Fuzzy Inference System," *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 23, 1993, pp. 665-685.

Hsien-Chou Liao received the B.S. and Ph.D. degrees in Computer Science and Information Engineering from National ChiaoTung University in 1991 and 1998, respectively. He is a professor, Department of Computer Science and Information Engineering, at Chaoyang University of Technology, Taiwan, R.O.C. He is also a senior member of IEEE. His research interests include location-based service, smart technology, mobile computing, computer vision, and pattern recognition.

Po-Yueh Chen received his B.S, M.S., and Ph.D. degrees in EE from National Taiwan University in 1988, National Chiao Tung University in 1990 and University of Maryland at College Park in 1997 respectively. Currently, Dr. Chen is a faculty

member in Department of Computer Science and Information Engineering, National Changhua University of Education, Changhua, Taiwan. His research interests include image processing, FPGA implementations, VLSI architecture design and digital IC design.

Zi-Jun Lin received the B.S. and M.S. in Computer Science and Information Engineering from the Chaoyang University of Technology in 2014 and National Changhua University of Education in 2016, respectively. His current research interests are in the computer vision.

Zi-Yi Lim received the B.S. and M.S. in Computer Science and Information Engineering from the Chaoyang University of Technology, Taiwan, Republic of China, in 2014 and 2016. His current research interests are in the area of computer vision.