ROI Tracking and Enhancement in Low-Quality Video Sequences

Shwu-Huey Yen Tai-Kuang Li

PRIA Lab., Department of Computer Science and Information Engineering
Tamkang University, Taipei, Taiwan, R.O.C.
105390@mail.tku.edu.tw 695410141@s95.tku.edu.tw

***Abstract*— Due to cost consideration, most of surveillance systems adopts the low resolution format to record the video sequences. However, the low resolution image quality often results the interested object too vague to be identified. We propose a two folds algorithm- tracking and enhancement- to solve this problem. First, the block-based motion estimation is used for object tracking. Next, a novel image enhancement scheme is used to reconstruct an initial high resolution image acquired from the region of interested (ROI). The error back-projection is further used to improve the quality of the reconstructed high resolution image. The proposed algorithm is tested on surveillance video sequences and some common video sequences. The tracking results are satisfying that it demonstrated the proposed algorithm is illumination invariant and robust to complex background. The image enhancement scheme is also tested. The test results on synthetic images showed that the quality of enlarged images has been improved.**

***Index Terms*— surveillance system, tracking, block-based motion estimation, enhancement, error back-projection**

I. INTRODUCTION

Safety and security have become critical in many public areas or private institutions. Thus, the importance of video surveillance techniques has increased considerably. For example, use video surveillance system to monitor public areas, such as railway stations, airports, highways, hospitals, banks; to monitor elders in nursing home; to monitor the entrance of an apartment building, etc. The current surveillance system can be classified into categories depending on whether the camera is static or mobile. For those static cameras with fixed focus surveillance system, the monitored scene stays fixed. Conversely, the mobile video camera is mounted in a fixed location but rotates regularly. To monitor across a wide area, multi-camera surveillance systems is also of popularity. Techniques that address handover between cameras in configurations with shared or disjoint views are therefore important [3]. However, up to now, the most widely used video surveillance systems is still single static camera such as in small convenient stores, and teller-free ATMs. Thus, in this paper we will focus on the single static video camera surveillance system.

To observe any unusual objects or unusual behaviors is the main purpose of surveillance systems. Therefore, object tracking is an important issue in the field of video surveillance system. Through the tracking on a region of interested (ROI), we can effectively observe movement of the interested object. Due to long time monitoring in video surveillance systems, most of users choose a low resolution (LR) format to record the video for memory space saving. However, when there is a need to extract information from the recorded data (video sequence), it is often difficult to clearly identify the interested object because of the low visual quality. In this paper, we propose a two folds algorithm-tracking and enhancement- to be applied on single static camera surveillance systems. First, the block-based motion estimation is used on the given LR recorded surveillance video sequence for object tracking. Next, a novel image enhancement scheme is used to reconstruct an estimated initial high resolution (HR) image from the LR one. The error back-projection is further used to improve the quality of the reconstructed high resolution image. The aim of this study is to track the suspect and show the suspect's face clearly when there is any suspect

presented in the LR surveillance video. By this way, the proposed algorithm can improve the safety of the monitored area and provide valuable information when needed.

In general, the object tracking problem defines background and foreground objects (i.e., the interested objects). If background can be well-modeled then the foreground objects can easily be tracked. In 2004, Li et al. [8] proposed a statistical method for background modeling to handle complex environment. Through this method, the background appearance is characterized by principal features and their statistics. The test results have shown that the principal features are effective in representing the spectral, spatial, and temporal characteristics of the background. Hsieh et al. [4] proposed a vehicle-classification scheme for estimating traffic parameters from video sequences. In their approach, for robustness consideration, a background updating scheme is used to keep background updated. Then, desired vehicles can be detected through image differencing and then tracked by a Kalman filter. Mustafah et al. [12] proposed a face detection and tracking system that is suitable to implement on the smart camera system. The system consists of a background subtraction stage, a skin color detection stage, and two-step Viola-Jones face detection stage. Cui et al. [2] proposed a method of tracking multiple people in an open area, such as shopping mall and exhibition hall. In their system, they utilized two laser scanners and one camera set on an exhibition hall monitoring visitors' flow.

Another way to track objects is by motion vectors. Applying motion vectors to an image to synthesize the transformation to the next image is called motion compensation. The combination of block-based motion estimation and motion compensation is a key part of video compression as used by MPEG 1, 2 and 4 as well as many other video compressing standards. To improve the full search (FS) algorithm in determining the block-based motion estimation, several researchers contribute their efforts in speeding up the process yet to maintain the results similar to those in the FS algorithm. Three-step search algorithm (TSS) [7] was proposed in 1981. It utilizes three stages to estimate the motion and checks only 25 points. Thus the computation complexity is much reduced, and the accuracy is slightly degraded. Since then, there have been a few improved algorithms proposed based on TSS. To name a few: new three-step search (NTSS) algorithm [9] in 1994, four-step search algorithm [15] and gradient descent search algorithm [11] in 1996, improved three-step search algorithm (ITSS) [17] in 1998, fast and efficient two-step search algorithm [1] in 1999, diamond search algorithm (DS) [19] in 2000, hexagon-based search algorithm (HEXBS) [18] in 2002, efficient three-step search algorithm (E3SS) [6] in 2004, and Orthogonal Logarithmic Search algorithm (OLS) [16] in 2005.

To use the valuable information from the surveillance video, it is important to enhance the LR image acquired from a surveillance system. One way to accomplish the goal is by the iterative back projection (IBP) (also known as recursive error back-projection, EBP) which was first presented by Irani and Peleg on 1991 [5]. In this approach, the HR image is estimated by back projecting the error (difference) between simulated LR images via imaging blur and the observed LR images. The advantage of IBP is that it can be understood intuitively and easily. However, this method has no unique solution due to the ill-posed nature of the inverse problem, and it is difficult to choose an appropriate back-projection kernel that determines the contribution of the error [14]. In 2004, Park et al. [13] proposed a synthesizing HR facial image using IBP. Their focus is on reconstruction of the initial HR facial image. It is based on the morphable face prototype composed of shape and texture components. The input LR is as low as 16×16 or 32×32 and the reconstructed image is 256×256. Their results are similar to the original HR one and clearer than methods like bilinear or bicubic interpolations. The facial model is the basis of the method which is trained by 100 trimmed frontal facial images. A pixel-wise correspondence has to be built between the to-be-enhanced LR image and the reference facial image. However, in the real surveillance system, the acquired facial images may be in different poses, different appearances and under different illumination, etc., that these images can not fit into the facial model in the proposed algorithm. Lin et al. [10] also proposed an IBP related method in resolution enhancement. They first enlarged the

given LR image z×z times by bicubic interpolation, the enlarged image is then decomposed into z^2 LR shifted images. Rebuild the HR images of these z^2 LR shifted images by IBP respectively. Finally, the resulted image is an average of these z^2 high resolution images. As pointed by the authors, the proposed back-projection kernel that determines the contribution of the error is an adaptive one. Depending on the location of a pixel, smooth area or edge area, it takes different weights to reflect the properties. Although they claimed the proposed method outperformed the conventional IBP, the computation complexity is heavy since it needs to compute the weight on every point for every iteration [10].

The remaining of the paper is organized as follows. Section II gives the description of the proposed algorithm. Experimental results are in Section III, and, finally, conclusion is given in the Section IV.

II. Propose Method

The proposed algorithm has two phases. First is the object tracking which is an extended version of the TSS, and the second is the IBP-based image enhancement. By observing the surveillance video, if an observer is interested in any object, he/she can mark a box to indicate the ROI. Then the system automatically tracks the indicated ROI and obtains a series of the tracked ROI images. These images are to be resolution enhanced to obtain better quality images. The following two subsections give the details of the algorithm.

A. Object tracking- the extended three-step search scheme

The TSS scheme is proposed by Koga et al. in 1981 [7]. It has been widely used in block matching motion estimation due to simplicity and effectiveness. For a maximum motion displacement of ±7, the TSS is a coarse-to-fine search mechanism and its step size decrease logarithmically (i.e., 4, 2, 1). As shown in Fig. 1, first, nine checking points (points are 4 pixels away) are to be compared (the point of the previous frame positioned on (0,0) is the one to find the motion vector), the center then moves to the point with the minimum distortion and the step size is halved. The procedure is repeated until the final step, which yields the motion vector.

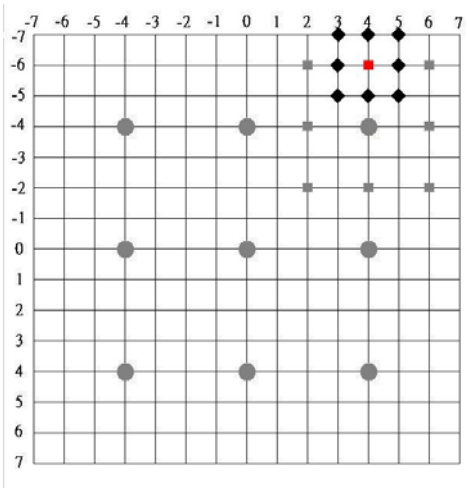


Fig. 1 The example path for convergence of the TSS scheme ● Points for Stage1 ■ Points for Stage2 ◆ Points for Stage3

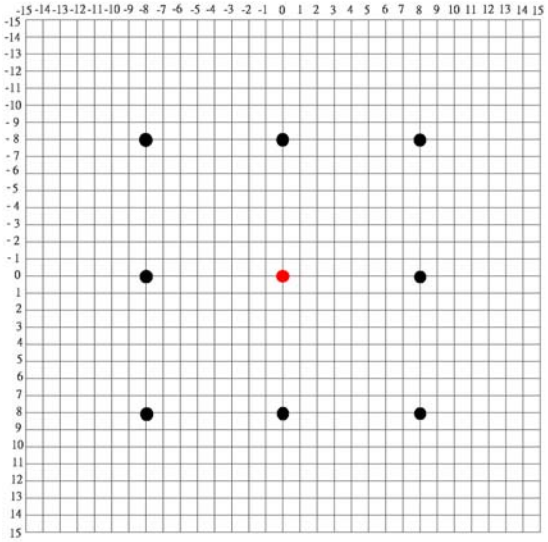
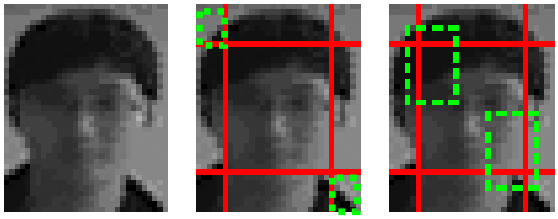


Fig. 2 The first stage in the extended TSS scheme used in our algorithm

In our application, the maximum motion displacement is doubled to be ±15 to accommodate possible large motions. Therefore, the searching scheme has 4 stages with the first stage of nine checking points and step size to be 8 as shown in Fig. 2. The center moves to the point with the minimum distortion and the step size is halved, then the scheme is completed followed by the TSS.

After the ROI is indicated (Fig. 3(a)), the region will be divided into nine blocks by the ratio of 1:3:1 both in height and width (Fig. 3(b)). To be time efficient, we choose only the top-left and bottom-right blocks for motion estimation. By assumption that the center of the ROI usually is important to human vision, the dimensions and loca-

tions of these two blocks are adjusted. Blocks dimensions are doubled (in height and width), and they are shifted to the center horizontally and vertically by 10% of their width and height respectively. As shown in Fig. 3(c), those two green boxes are for motion estimation. In stead of finding motion estimation for every point in the box, we adopt the sub-sampling idea as proposed in [1] to speed up the computation. The final motion vector is determined by voting. After the motion vectors of these two boxes are determined, then the ROI for next frame is determined too and the tracking process repeats again.



(a) (b) (c)

Fig.3 Determine the ROI of the next frame (a) ROI of the current frame; (b) dividing into nine blocks by ratio 1:3:1; (c) two green boxes are for motion estimation to determine the ROI of the next frame

B. Image enhancement

To reconstruct a HR image from a given LR image, first an initial HR image H_0^R is constructed and it is further enhanced by IBP. Figure 4 shows the flowchart of the image enhancement procedure.

- Reconstruction of H_0^R

There are plenty methods for image enlargement. In Fig. 5, two images, *Lena* and *Barbara*, of original 384×384 shown on (a) and (d), down-sampled into 192×192 as the LR images. Now enlarge the images 2 times by duplication and bicubic interpolation shown on (b), (c) and (e), (f). It is clearly that duplication is prone to having blocking effect. In addition, bicubic interpolation is well known that it outperforms interpolation methods such as linear or bilinear in image quality. Thus given the LR image L, we choose the bicubic interpolation to obtain the initial HR image H. Next, the Gaussian filter (size 5×5 and $\sigma = 1$) is applied to H to remove the noise caused by the low resolution quality. Laplacian filter is then applied to enhance the contrast.

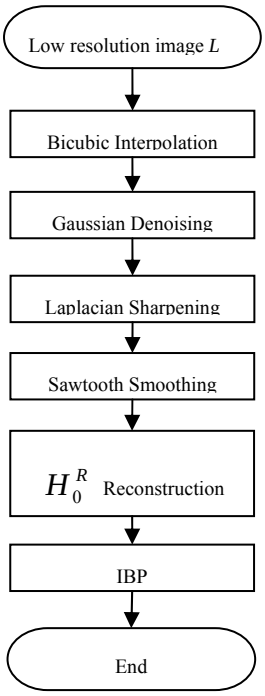


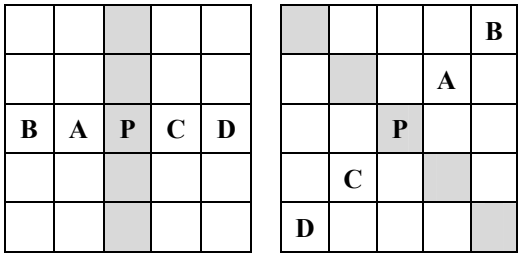
Fig. 4 The flowchart of an HR image reconstruction

One of the disadvantages about interpolation is the sawtooth emerging along slanted edges. Thus we propose a sawtooth smoothing algorithm to reduce this side effect. The main idea is to smooth the neighboring points of an edge pixel according to the gradient of such edge pixel. Figure 6 shows examples such that P is on a vertical (90^o) line or a 135^o slanted line. Then two non-edge neighbors of P which located on the direction perpendicular to the edge are to be smoothed. For example, in 6(a) (or 6(b)), if points A, B are non-edge points, then I(A), the intensity of point A, is replaced by the average of I(A) and I(B). Similarly, if C, D are non-edge points, then I(C) is replaced by $1/2(I(C)+I(D))$. Equation (1) defines the direction of gradient (angle Θ). Equations (2), (3) are for Θ to be 90°(Fig. 6(a)), (4) and (5) are for Θ to be 135°(Fig. 6(b)). The cases of 0^o and 45^o are similar. We adopt the Canny edge detector to identify edge pixels and Θ for every pixel is determined and classified to one of the four major directions whichever is the closest to Θ. After sawtooth smoothing, we obtain a HR image H_0^R and further enhanced by IBP is followed.

(a) (b) (c)

(d) (e) (f)

Fig. 5 The comparison of enlarged results. (a) and (d) are the original images, the rest of images are obtained from down sampled images then enlarged such that (b) and (e) are done by duplication; (c) and (f) are done by the bicubic interpolation.



(a) (b)

Fig. 6. P is an edge pixel such that its gradient direction is (a) 90^o and (b) 135^o

$$\theta = \tan^{-1}\frac{\Delta y}{\Delta x}. \quad (1)$$

$$I(i-1,j) = \frac{I(i-1,j) + I(i-2,j)}{2}, \quad (2)$$

$$I(i+1,j) = \frac{I(i+1,j) + I(i+2,j)}{2}, \quad (3)$$

$$I(i+1,j-1) = \frac{I(i+1,j-1) + I(i+2,j-1)}{2}, \quad (4)$$

$$I(i-1,j+1) = \frac{I(i-1,j+1) + I(i-2,j+2)}{2}. \quad (5)$$

- The IBP algorithm

There are quite a few research works discussing the IBP algorithm, we used the version adopted in [13]. If $H_{t\text{-}1}^{R}$ represents the HR image obtained in the $(t\text{-}1)^{st}$ iteration, we define L_t^E to be the difference of the original LR image L and L_t^R (i.e., $L_t^E = L - L_t^R$) and the distance $D_t = | L_t^E |$ where L_t^R is the downsampled version of $H_{t\text{-}1}^{R}$. Next, to obtain the updated HR image H_t^R, a fraction of the HR of the L_t^E is added to the H_0^R as the error compensation. Equation (6) depicts such error compensation.

$$H_t^R = H_0^R + \omega_t \cdot (\text{reconstructed HR of } L_t^E). \quad (6)$$

The process is repeated until the allowed maximum number T is reached or D_t is small enough, or D_t is very similar to D_{t-1}. The details of the IBP algorithm can be referred to [13]. In our implementation, the reconstruction of HR image of L_t^E is done by bicubic interpolation and ω_t is chosen to be $\sqrt[3]{D+1}/15$ which will be discussed more in the next section.

III. Experimental Results

We discuss the experiments in two parts. First, the tracking robustness on illumination changes and fast movements is tested. Then the effectiveness on image enhancement is examined.

A. Object tracking

The experiments are performed on four different video sequences. The first one is taken in a laboratory with a fixed-light source (Fig. 7), the second (Fig. 8) and the third (Fig. 9) are from IPPR 2007 [20]. The last one is *Mobile and Calendar* (Fig. 10) which is not a LR sequence. We use it to examine the effectiveness of the tracking algorithm under complex background.

Figure 7 shows the result of tracking inside a laboratory with fixed-light source. The man walked in a way of Z-path and his face was labeled as a ROI. The proposed algorithm is effective under fixed-light condition. As in Fig. 8, the man who was tracked walked up stairs and made a right turn towards the lower right corner of the image. There were two sudden illuminant variations as shown in Fig. 8 (b) and (f). Figure 8(c) showed that the tracking was not affected by the sudden strong light. The man passed a woman as in Fig. 8 (d) and (e). This demonstrates that the proposed algorithm is robust to illuminant variation and to the moving background.

Figure 9 shows an outdoor tracking result. The man who was tracked walked from right to left with fast head movements. As seen in Fig. 9 (c) and (d), the man's face became blurred due to fast motion, however, the proposed algorithm still works correctly.

Mobile and Calendar is a video sequence acquired from a moving camera. Besides the complex wall paper, the calendar is moving up and down slowly, the ball is bouncing back and forth, and the train is moving from right to left. In the experiment, the engine is labeled as a ROI. Figure 10 shows the tracking result. The proposed method worked well at the beginning, but the ROI then included part of ball (on left) and part of the cart (on the right) which are not included in the original ROI. The result is acceptable, but it is not as good as the previous experiments. We plan to tackle this problem by additional color information and texture analysis in the future.

B. Image enhancement

- Effectiveness of the algorithm H_0^R

We down sample images to 1/3 of the size and then enlarge the LR images to the original size. Table 1 is the comparison between bicubic and our propose HR reconstruction on H_0^R. According to Table 1, it shows that our method is a little bit better. Figure 12 (a) and (c) are the H_0^R.

Table 1
Image reconstruction comparison between bicubic and ours

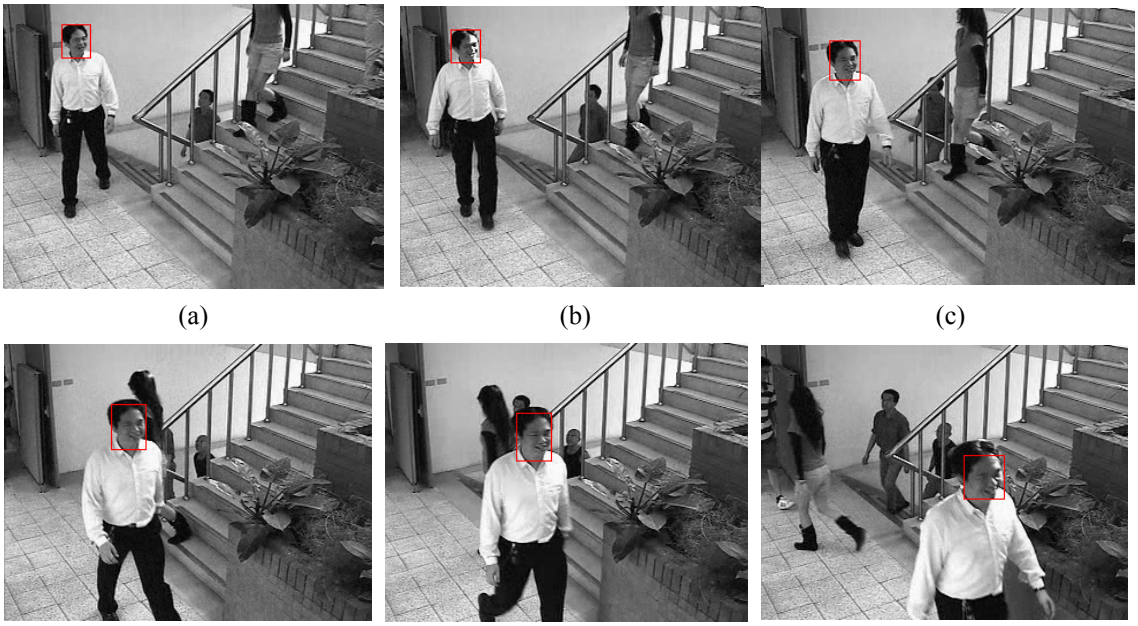
PSNR / Image	Bicubic	Propose method
Lena	23.6092	23.6096
Barbara	16.03	16.2986

- The weight of IBP

The IBP method is mainly adding a fraction of the reconstructed HR of the error to H_0^R, as in Eq. (6), to improve the similarity of reconstructed and the given LR image. How to choose ω_t is important. In general, if the difference D_t is large then H_0^R can be modified more and vice versa. Thus, we need a ω_t that is proportional to D_t. In the following discussion, the subscript t will be omitted for writing convenience. We test some functions for ω : $\ln(D+1)/k$, $\sqrt{D+1}/k$, and $\sqrt[3]{D+1}/k$, k is a constant. Figure 11 summarizes the PSNR values for *Lena*

(a) (b)

Fig. 7 Tracking result of an indoor with fixed-light source and (a) is Frame 21, (b) is Frame 250.



(a) (b) (c)

(d) (e) (f)

Fig. 8 Tracking result of illumination changes in an indoor environment (a) Frame 01: ROI is indicated (b)Frame 06: the man came across a strong light as half of his face turning bright (c)Frame 14: the man left the strong light area (d) Frame 24: the man just passed the woman (e) Frame 27: the man departed from the woman (f)Frame 35: the man came across a strong light again

(a) (b)

(c) (d)

Fig. 9 Tracking result of fast head movement in an outdoor environment (a)Frame 01 (b)Frame17 (c)Frame25 (d)Frame32

(a) (b)

(c) (d)

Fig. 10 Tracking result in *Mobile and Calendar* (a)original video sequence, (b)Frame 01, (c)Frame 10, (d)Frame 20

under different constants k. According to Fig. 11, we choose $\omega = \sqrt[3]{D+1}\,/\,15$ which performs the best. Figure 12 is the result of *Lena* and *Barbara* with PSNR indicated. (b) and (d) look similar to (a) and (c), and their PSNR value also indicate that the effect on applying the IBP is not too much. In the future, we will explore more to improve the effectiveness of IBP. The enhancement is also applied on LR image acquired from video sequence. The resolutions are 28x36, 29x34, and 20x40 shown on Fig. 13 (a~c), and the results are shown on (d~f).

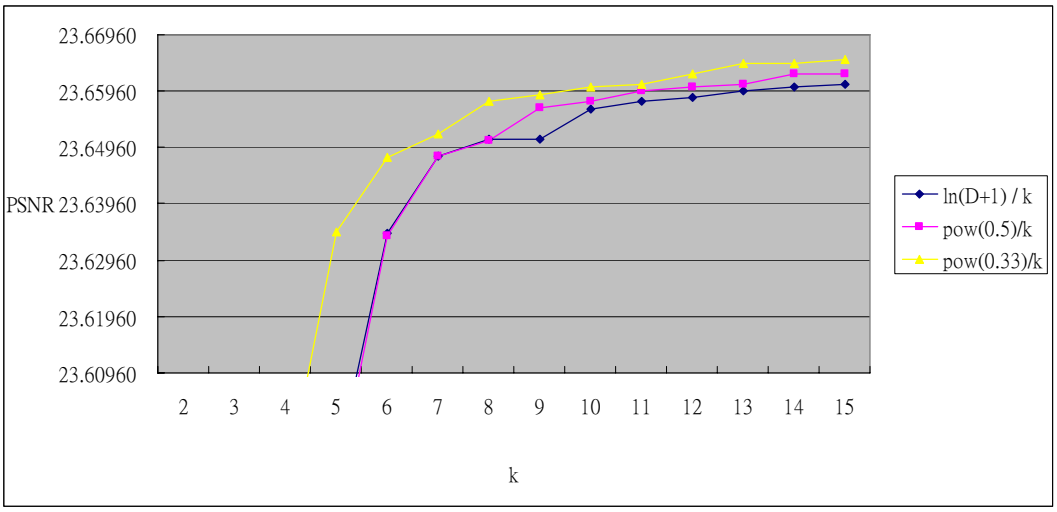
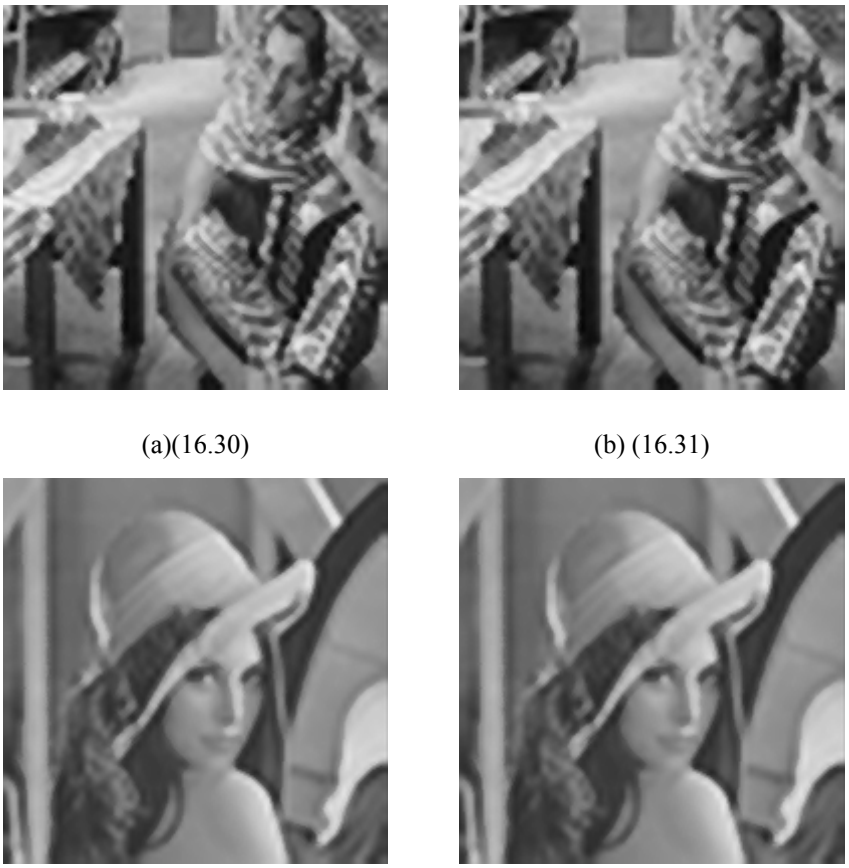


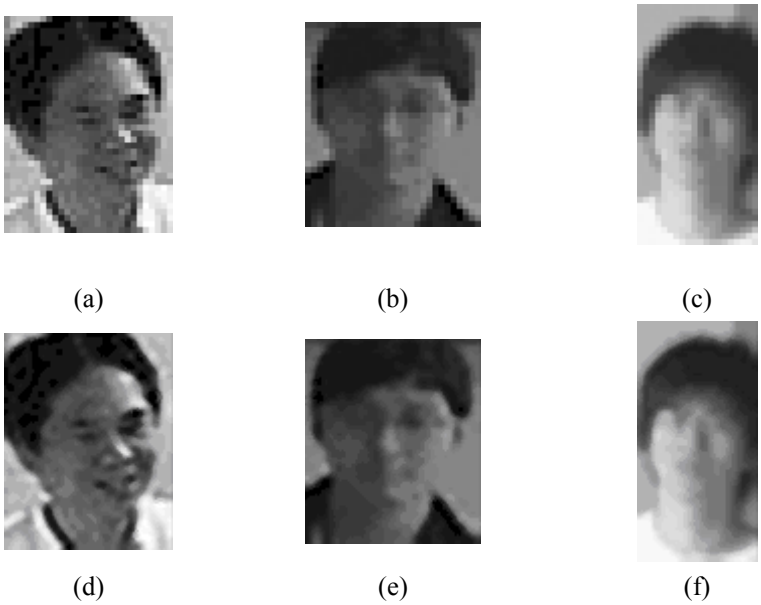
Fig. 11 The PSNRs under different weight functions in *Lena* enhancement



(a)(16.30) (b) (16.31)

(c) (23.61) (d) (23.67)

Fig 12 Enhancement of *Lena* and *Barbara* with PSNR indicated by the parenthesized numbers. (a) and (c) are the result of H_0^R ; (b) and (d) are the result following by IBP.



(a) (b) (c)

(d) (e) (f)

Fig 13 The enhancement results of our propose method in surveillance video sequences. (a)-(c) are acquired LR images from the video (enlarge by duplication for better vision) with resolutions 28x36, 29x34, and 20x40; and (d)-(f) are enhanced and enlarged three times by our method.

IV. CONCLUSION

In this paper, we utilized the extended three-step search algorithm that its search range is ±15 for block motion estimation. The experiment shows that it can track a ROI in real-time, and performs a good tracking result. After tracking, a HR image reconstruction method is used to reconstruct a high resolution image from the acquired low resolution image of the surveillance system. Finally we use the iterative back-projection method to further enhance the image quality.

Our experiment showed that the IBP method has limited improvement. In the future, we will explore more on an adaptive weight function to improve the effectiveness of the IBP. The issues of tracking multiple ROIs in real-time and combining color information and texture analysis in tracking are to be further studied too.

REFERENCE

[1] F. H. Cheng and S. N. Sun, "New Fast and Efficient Two-Step Search Algorithm for Block Motion Estimation," IEEE Trans. Circuits Syst. Video Technol, Vol. 9, pp. 977-983, 1999.

[2] J. Cui, H. Zha, H. Zhao, and R. Shivasaki, "Multi-Modal Tracking of People Using Laser Scanners and Video Camera, " Image and Vision Computing, Vol.26, pp. 240-252, 2008.

[3] G. L. Foresti, C. Micheloni, L. Snidaro, P. Re-magnino, and T. Ellis, "Active Video-Based Surveillance System," IEEE Signal Processing Magazine, Vol.22, pp. 25-37, 2005.

[4] J. W. Hsieh, S. H. Yu, and W. F. Hu, "Automatic Traffic Surveillance System for Vehicle Tracking and Classification," IEEE Trans. Intelligent Transportation Systems, Vol.7, pp. 175-187, Jun. 2006.

[5] M. Irani and S. Peleg, "Improving Resolution by Image Registration," CVGIP: Graphical Models and Image Proc., Vol. 53, pp. 231-239, 1991.

[6] X. Jing and L.P. Chau, "An Efficient Three-Step Search Algorithm for Block Motion Estimation," IEEE Trans. Multimedia, Vol.6, pp. 435-438, 2004.

[7] T. Koga, K. Iinuma, A. Hirano, Y. Iijima, and T. Ishi-guro, "Motion Compensated Interframe Coding for Video Conferencing," Proc. of the National Tele-commu-nications Conference (NTC), pp. G5.3.1-5,

1981.

[8] L. Li, W. Huang, Y.H. Gu, and Q. Tian, “Statistical Modeling of Complex Backgrounds for Foreground Ob-ject Detection,” IEEE Trans. image processing, Vol.13, pp. 1459-1472, 2004.

[9] R. Li, B. Zeng and M. L. Liou, “A New Three-Step Search Algorithm for Block Motion Estimation,” IEEE Trans. Circuits Syst. Video Technol., Vol.4, pp. 438-442, 1994.

[10] G. S. Lin and M. K. Lai, “Enhancing Resolution Using Iterative Back-Projection Technique for Image Se-quences,” Journal of Computers, Vol.19, 2008.

[11] L. K. Liu and E. Feig, “A Block-Based Gradient Descent Search Algorithm for Block Motion Estimation in Video Coding,” IEEE Trans. Circuits Syst. Video Technol, Vol.6, pp. 419-422, 1996.

[12] Y. M. Mustafah, T. Shan, A. W. Azman, A. Bigdeli, and B.C. Lovell, “Real-Time Face Detection and Track-ing for High Resolution Smart Camera System,” Digital Image Computing Techniques and Applications, pp. 387-393, 2007.

[13] J. S. Park, and S. W. Lee, “Resolution Enhancement of Facial Image Using an Error Back-Projection of Ex-ample-Based Learning,” Sixth IEEE International Con-ference Automatic Face and Gesture Recognition, pp. 831-836, 2004.

[14] S. C. Park, M. K. Park, and M. G. Kang, “Su-per-Resolution Image Reconstruction: A Technical Overview,” IEEE Signal Processing Magazine, Vol.20, pp.21-36, 2003.

[15] L. M. Po and W. C. Ma, “A Novel Four-Step Search Al-gorithm for Fast Block Motion Estimation,” IEEE Trans. Circuits Syst. Video Technol, Vol.6, pp. 313-317, 1996.

[16] S. Soongsathitanon, W.L. Woo, and S.S. Dlay, “Fast Search Algorithms for Video Coding using Orthogonal Logarithmic Search Algorithm,” IEEE Trans. Consumer Electronics, Vol.51, pp. 552-559, 2005.

[17] D. Xu, C. Bailey, and R. Sotudeh, “An Improved Three-Step Search Block-Matching Algorithm for Low Bit-Rate Video Coding Applications,” Signals, Systems, and Electronics, ISSSE 98. pp. 178-181, 1998.

[18] C. Zhu, X. Lin, and L. P. Chau, “Hexagon-Based Search Pattern for Fast Block Motion Estimation,” IEEE Trans. Circuits Syst. Video Technol, Vol. 12, pp. 349-355, 2002.

[19] S. Zhu and K.K. Ma, “A New Diamond Search Al-gorithm for Fast Block-Matching Motion Estima-tion,” IEEE Trans. Image Processing, Vol.9, pp.287-290, 2000.

[20] http://www.ippr.org.tw/contest07, Face Detection in Video, IPPR Contest 2007, downloaded on May 2008.