Illumination and Motion-Based Video Enhancement for Night Surveillance

Jing Li[1], Stan Z.Li[2], Quan Pan[1], Tao Yang[1]
[1]College of Automatic Control, Northwestern Polytechnical University, Xi'an, China , 710072
[2]National Lab of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, China, 100080
jinglinwpu@163.com , szli@nlpr.ia.ac.cn , quanpan@nwpu.edu.cn, yangtaonwpu@163.com

Fig.1 Left: Night input image. Right: Enhancement result.

ABSTRACT

This work presents a context enhancement method of low illumination video for night surveillance. A unique characteristic of the algorithm is its ability to extract and maintenance the meaningful information like highlight area or moving objects with low contrast in the enhanced image, meanwhile recover the surrounding scene information by fusing the daytime background image. A main challenge comes from the extraction of meaningful area in the night video sequence. To address this problem, a novel bidirectional extraction approach is presented. In evaluation experiments with real data, the notable information of the night video is extracted successfully and the background scene is fused smoothly with the night images to show enhanced surveillance video for observers.

1. INTRODUCTION[1]

Night video enhancement [1,2] is one of the most important and difficult component of video security surveillance system. Recent conflicts have again highlighted the crucial requirement forever more sophisticated night vision systems for sea, land and air forces. The increasing use of night operations requires that effective night vision systems are available for all platforms.

However, the performance of most surveillance cameras are not satisfied at low light or high contrast situations. Low light generates noisy video images, and bright lights (like from car head lights) overexpose the electronics in the camera, such that all detail is lost and the low signal-to-noise image limits the amount of information conveyed to the user with the computer interface. The electronics in a standard surveillance camera are just too simple to compensate for that, so it is now viable to consider digitally enhancing the night image before presenting it to the user, thus increasing the information throughput [3].

As mentioned above, the difficulties of night image problem mainly contain two aspects.

- The first is that the obtained night image appears much noise, due to reasons of sensor noises or very low luminance.
- The second is the high light or dark areas in which the scene information cannot be seen clearly by the observers.

In this paper, we address the problems of generating a more context included descriptions of night video for surveillance based on extracting and fusing techniques. And enlighten by [1], which presents a new idea of fusion daytime and nighttime image for image context enhancement, we present a novel illumination and motion based extraction approach to extract the meaningful information in nighttime video, and propose a motion based background modeling method to acquire the surrounding scene information under various illumination

[1] The work presented in this paper was sponsored by the Foundation of National Laboratory of Pattern Recognition (#1M99G50) and National Natural Science Foundation of China (#60172037) .

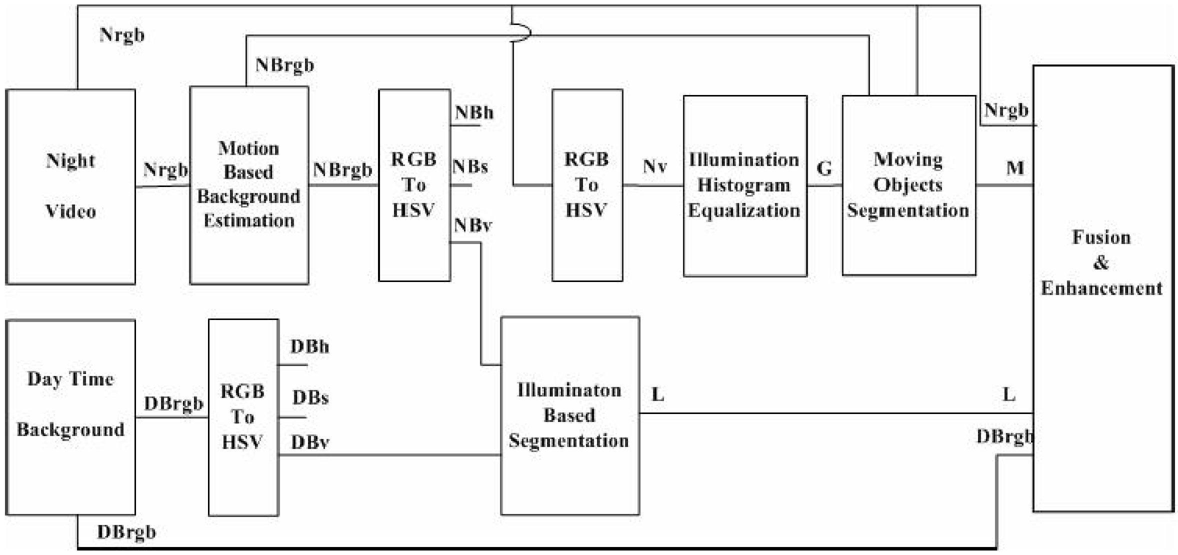


Fig.2 Framework of the algorithm

level. The objective of our method is to guarantee most of the important contexts in the scene are synthesized to create a much clearer video for observers. Extensive experiments performed using video sequences under various scenes demonstrated that our algorithm is fast and efficient for night video enhancement.

The paper is organized as follows. Section 2 introduces the framework of the algorithm. Section 3 explains the details of the extraction and fusion step of the enhancement algorithm. Section 4 and 5 presents extensive results and conclusion.

2. OUTLINE OF THE ALGORITHM

The system consists of five parts (Shown in Fig.2). (1)Motion based background estimation, (2) Illumination based segmentation, (3) Illumination histogram equalization, (4) Moving objects segmentation and (5) Fusion and enhancement.

In part one, a dynamic background is created on line. In part two its illumination will be contrasted to the reference background of daytime to acquire the high light and low light area. Pixels in high light area will be directly sent to the final fusion module. Meanwhile, the illumination of the current and background night image are transformed into several levels in part three and various thresholds are used in each level to segment moving objects in part four. In part five, combining the extracting result of moving and light area, a multi-resolution based fusion method is presented to get the final enhancement result.

3. NIGHT VIDEO ENHANCEMETN ALGORITHM

3.1. Motion base background model estimation

Background maintenance in video sequences is a basic task in many computer vision and video analysis applications [4,5,6,7]. The basic idea of our background estimate method comes from an assumption that the pixel value in the moving object's position changes faster than those in the real background. Fortunately, this is a valid assumption in most application fields such as traffic video analysis, people detection and tracking in intelligent surveillance. Under this assumption, we develop a pixel level motion detection method which could identifies each pixel's changing character over a period of time by frame-to-frame difference and analyzes a dynamic matrix $D(k)$ presented in this paper.

Let $I(k)$ denotes the input frame at time k, and the subscript i,j of $I_{i,j}(k)$ represent the pixel position. Equation (1) and (2) show the expression of frame-to-frame difference image $F(k)$ and the dynamic matrix $D(k)$ at time k.

$$F_{i,j}(k)=\begin{cases}0 & \left|I_{i,j}(k)-I_{i,j}(k-\gamma)\right|\le Tf\\ 1 & otherwise\end{cases} \quad (1)$$

$$D_{i,j}(k)=\begin{cases}D_{i,j}(k-1)-1 & F_{i,j}(t)=0, D_{i,j}(k-1)\neq 0\\ \lambda & F_{i,j}(t)\neq 0\end{cases} \quad (2)$$

Where γ represent the interval time between the current frame and the old one, Tf is the threshold to make a decision whether the pixel is changing at time k or not, and λ is the time length to record the pixel's moving state, once the $D_{i,j}(k)$ equates to zero, the pixel update method will make a decision that this pixel should be updated into the background B . Fig.3 represents the results of the background estimation of day and night video separately.

Fig.3 Background estimation. The first column a), c) contains the input video of day and night. The second column b),d) contains the estimated background.

3.2. Illumination based segmentation

Extracting meaningful context can enhance the low quality night videos, such as the ones obtained for security surveillance. In this paper, the meaningful context of the night video is defined as area with high illumination or moving objects. And for the daytime reference background, the scene information like building, road, trees are considered important.

The problem here is how to segment the high light area, which is easy for observer to see, and the moving objects which are important for visual-based surveillance fields. In this section, we will present a real time high light area segmentation algorithm. Considering the background images of daytime and night are images of the same scene captured under different illumination, we may draw an assumption that only in the man-made high light, the illumination of pixel in night image maybe higher than its corresponding point in the daytime. Fortunately, this is a valid assumption in many night video surveillance scenes, and based on it we develop the following illumination area segmentation algorithm.

After background model estimate, the background image of day and night (DB and NB) are transformed from RGB color space to HSV (Hue-Saturation-Value) color space. An illumination segmentation map $L_{(i,j)}$ can be computed as (3)

$$L_{(i,j)} = \begin{cases} 1 & (NB_{(i,j)}(V) - DB_{(i,j)}(V)) \geq 0 \\ 0 & (NB_{(i,j)}(V) - DB_{(i,j)}(V)) < 0 \end{cases} \quad (3)$$

where $DB_{(i,j)}(V)$ and $NB_{(i,j)}(V)$ denote the luminance value of background image DB and NB separate at position (i, j) .

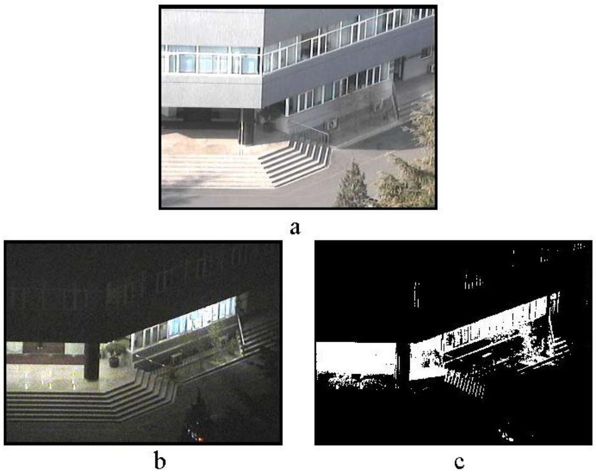


Fig.4 a) Daytime background image DB . b) Night background image NB . c) Illumination segmentation result L

Fig.4 shows an example of illumination segmentation result with (3). The high light area is accurately segmented (Shown in Fig.4(c)), and it will be used to direct the final fusion step. One problem of this technique is that illumination segmentation result does not include the moving objects in the dark area, which is important especially for security surveillance. To address this problem, we develop a multiple level moving objects segmentation method. The following section describes the details.

3.3. Moving objects segmentation

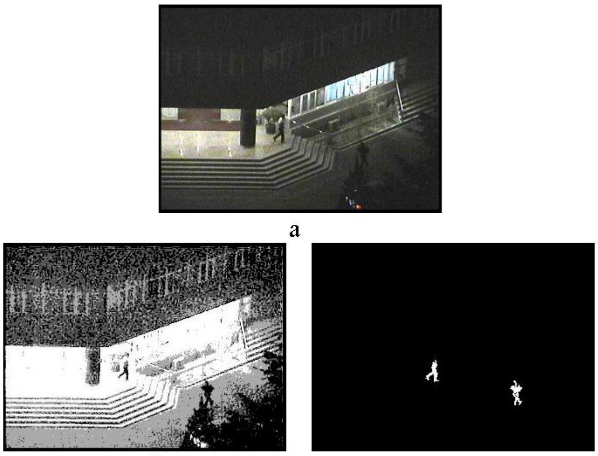
Consider the man-made lights, the illumination intensity in night image changes a lot (Shown in Fig.4.b) and the contrasts between the foreground and background are quite different in those areas. Thus it's not suitable to use the same threshold in moving objects segmentation. One popular method which uses various threshold for each pixel is to model the probability of observing the current pixel value as a mixture of K Gaussian distribution [5]. Although the performance of K Gaussian background model is satisfied in theory, the proceeding algorithm is computationally intensive to real-time use, especially the step of fitting K Gaussians to the data for each pixel and every frame. In our experiment, the processing speed of K Gaussian model is less than 10 fps for image size of 320x240.

To achieve real-time and accurate moving objects segmentation, we first use illumination histogram equalization in the night video $N_{(i,j)}(V)$. Pixels will be classified to M levels according to their luminance. After that, different thresholds will be assigned for different classes in the background subtraction. Let $p(i)$ denotes the ratio of pixels, which luminance equals to i in $N_{(i,j)}(V)$, G denotes the equalized image, and it can be computed through equation (4).

$$G_{(i,j)} = M \cdot f(m),\ m = 1,...,M \tag{4}$$

where $f(m) = \sum_{i=0}^{i=m} p(i)$ and $G_{(i,j)}$ will be modified to nearest integral number. For the high light area has already been exacted in the formal section 3.3. The motion map M_d can be computed by (5).

$$M_{(i,j)} = \begin{cases} 1 & \begin{array}{l} \left|N_{(i,j)}(R) - NB_{(i,j)}(R)\right| > T(m), or \\ \left|N_{(i,j)}(G) - NB_{(i,j)}(G)\right| > T(m), or \\ \left|N_{(i,j)}(B) - NB_{(i,j)}(B)\right| > T(m) \end{array} \\ 0 & otherwise \end{cases} \tag{5}$$



a

b c

Fig.5 a) Night image. b) Illumination equalization image. c) Moving objects segmentation result.

Where $T(m)$ represents the threshold at luminance level m and $m = G_{(i,j)}$.

In Fig.5 we divide the input image to four luminance levels (displayed with different gray values in Fig.5(b)). Different thresholds are used at those four levels. Fig.5(c) shows the moving segmentation result. In this image, small noises are rejected through morphologic filtering. Note that the running person which has low contrast to the background in the dark area is accurately segmented (Fig.5(c)).

3.4. Image Fusion

Many techniques can be used in the final fusion step. However the DWT and Laplacian image pyramid fusion sequences exhibited flickering distortions due to the shift variance of the decomposition process. So in our experiments, we selected the SIDWT [8] (Shift-Invariant Wavelet Transform) based method to overcome the shift dependency. It consists of three main steps. First, each source image is decomposed into a decomposed into their shift invariant wavelet representation. Then a composite multiscale representation is constructed from the source representations and a fusion rule. Finally the fused image is obtained by taking an inverse SIDWT transform of the composite multiscale representation.

The fusion rule we used is choosing the maximum value of the coefficients of the night input image and daytime reference background image for the high frequency band. For the low frequency band, the coefficients of the images are weighted according to the motion and illumination map. Let $EN_{(i,j)}$ and $EDB_{(i,j)}$ represent the coefficients of input image $N_{(i,j)}$ and daytime reference background $DB_{(i,j)}$, the fused image $EF_{(i,j)}$ can be computed by (6).

$$EF_{(i,j)}^{high} = \max(EN_{(i,j)}, EDB_{(i,j)}) \tag{6}$$

$$EF_{(i,j)}^{low} = \begin{cases} \alpha \cdot EN_{(i,j)}^{low} + (1-\alpha) \cdot EDB_{(i,j)}^{low} & if\ L_{(i,j)} = 1 \\ EN_{(i,j)}^{low} & if\ L_{(i,j)} = 0 \\ & \& M_{(i,j)} = 1 \end{cases} \tag{7}$$

where $EF_{(i,j)}^{low}$ and $EF_{(i,j)}^{high}$ denote coefficients of fused image in the low and high frequency band.

4. EXPERIMENT RESULTS

A real time night video enhancement system based on the presented algorithm has been developed. The system is implemented on standard PC hardware (Pentium IV at 3.0GHz). The algorithm has been tested in various environments, and the performance is satisfied. We shown an example of outdoor scene combined from a daytime background and a night picture (see Fig.6). Notice that a running people in dark area is correctly extracted and fused in the final result (see Fig.6(c,d)). The enhanced video sequence may found in Fig.7. What's more, we do many experiments in different scenes (see Fig.8), and the results show that this algorithm does well.

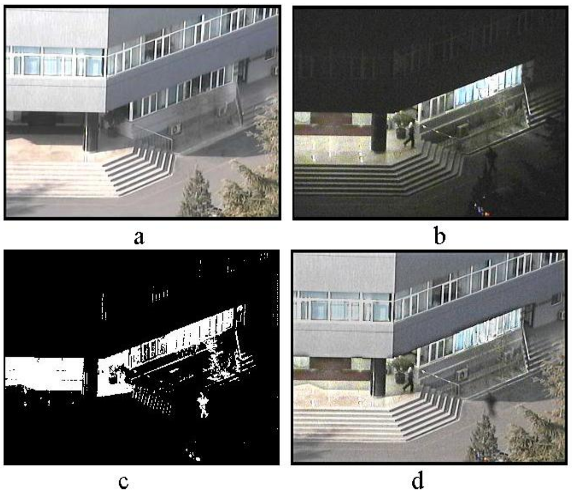


Fig.6 Image enhancement result. a) Daytime background. b) Night input video. c) High illumination and motion map. d)Enhanced result.

5. CONCLUSION

A night video illumination and motion based enhancement algorithm is presented which could extract and fusion meaningful information from multiple images. A real time night video enhancement system based on the presented has been developed and tested with long time video in various environments. Experiment results demonstrate that the system is highly computationally cost effective. Moreover, the enhanced video is visually significant and contains more information than the original night vision images.

6. REFERENCES

[1] Ramesh Raskar, Adrian Ilie and Jingyi Yu, "Image Fusion for Context Enhancement and video surrealism", *The 3rd International Symposium on Non-Photorealistic Animation and Rendering* (NPAR), Annecy, France, 2004.

[2] Sale, D, Schultz, R.R. Szczerba, R.J, "Super-resolution enhancement of night vision image sequences", *Systems, Man, and Cybernetics, 2000 IEEE International Conference* , Vol: 3 , Pages: 8-11 Oct. 2000.

[3] Chek K. Teo,Digital Enhancement of Night Vision and Thermal Images, Master's thesis, Naval Postgraduate school Monterey.

[4] Collins, Lipton, Kanade, Fujiyoshi, Duggins, Tsin, Tolliver, Enomoto, and Hasegawa ," A System for Video Surveillance and Monitoring. VSAM Final Report," *Technical report CMU-RI-TR-00-12*, Robotics Institute, Carnegie Mellon University, May, 2000.

[5] P.KawTraKulPong, R.Bowden, "An improved adaptive background mixture model for real-time tracking with shadow detection," *Proceedings of Second European Workshop on Advanced Video-based Surveillance Systems*, 2001.

[6] Stauffer, C, Grimson, W.E.L., "Learning patterns of activity using real-time tracking", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol: 22 , Issue: 8, Pages:747 –757, Aug. 2000.

[7] Tao Yang, Stan Z.Li , Quan Pan, Jing Li, "Real-time Multiple Object Tracking with Occlusion Handling in Dynamic Scenes," *Proceedings of IEEE Computer Vision and Pattern Recognition Conference (CVPR'05),* Vol I, Pages:970-975, June 20-26, San Diego, CA, USA.

[8] Chi-Man Pun, Moon-Chuen Lee, "Extraction of shift invariant wavelet features for classification of images with different sizes", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol: 26, Issue: 9. Pages: 1228-1233, 2004.

#135 Original image

#135 Enhanced image

#170 Original image

#170 Enhanced image

#178 Original image

#178 Enhanced image

Fig.7 The first column contains the night video sequence.
The second column contains the enhanced result.

Original image (street)

Enhanced image (street)

Original image (playground)

Enhanced image (playground)

Original image (backyard)

Enhanced image (backyard)

Fig.8 The first column contains the night video of different scenes.
The second column contains the enhanced result.