Context Enhancement of Nighttime Surveillance by Image Fusion

Yinghao Cai, Kaiqi Huang, Tieniu Tan
National Laboratory of Pattern Recognition,
Institute of Automation, Chinese Academy of Sciences.
P.O.Box 2728, Beijing, 100080, China
{yhcai, kqhuang, tnt}@nlpr.ia.ac.cn

Yunhong Wang
School of Computer
Science and Engineering,
BeiHang University
wangyh@nlpr.ia.ac.cn

Abstract

In this paper, we propose a novel method of automatically combining images of a scene at different time intervals by image fusion. All the important information of the original low quality nighttime images is combined with the context from a high quality image of the daytime at the same viewpoint. The fused image contains a comprehensive description of the scene which is more useful for human visual and machine perception. Experimental results show that the proposed method is robust and effective.

1. Introduction

Visual surveillance as an active topic in computer vision has received much attention in recent years. Visual surveillance in dynamic scenes aims to detect, recognize and track objects such as people and cars from image sequences at low level. Moreover, it analyses and understands objects' behavior at a high level.

However, most of the previous work has focused on daytime surveillance. Understanding nighttime video is a challenging problem because of the following reasons: Firstly, due to low contrast, we can not clearly extract moving objects from the dark background. Most color-based methods will fail on this matter if the color of the moving objects and that of the background are similar. Secondly, the signal to noise ratio is usually very low due to high ISO (ISO is the number indicating camera sensors sensitivity to light). Using a high ISO number can produce visible noise in digital photos. But low ISO number means less sensitivity to light. Thirdly, environment information, in other words, the context information of a scene, affects the way people perceive and understand what has happened. Due to the limited information in nighttime video, understanding behavior of people in video becomes more difficult.

Adrian et al. [3] propose a new concept of image and video enhancement technique, which is called CER (context enhanced rendering). In [3], the goal of CER is to incorporate context information of a scene from one image with other important features from another image of the same scene. A typical example of CER is enhancing nighttime video with an image of the daytime at the same viewpoint. In this application, it is supposed that we are capable of capturing scenes of day and night at the same viewpoint by a fixed camera. Therefore, we can make use of the high quality background of the day to enhance the context of nighttime images. Raskar et al. in their work [6] propose a gradient domain technique to combine daytime image and nighttime image together. Extra processes are necessary to deal with observable color shift which is a common problem of gradient-based approaches. Li et al. [5] combine regions of interest together by a multi-resolution based fusion method: shift-invariant discrete wavelet transform (SIDWT). Various thresholds are used which need adjustment when the scene is changed.

(a) Daytime Background (b) Nighttime Image

Figure 1. Example of daytime and nighttime image

As we are combining daytime image and nighttime image together, as shown in Figure 1, we must determine what information should be preserved in the fused image. Since the camera is fixed, both daytime image and nighttime image include static parts of the scene such as buildings, trees, roads, etc. The low quality static parts of the nighttime image can be replaced by the high quality counterpoint of the

Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06)
0-7695-2521-0/06 $20.00 © 2006 IEEE

daytime image. Moving objects are essential to surveillance, so moving objects must be maintained in the fused image. Furthermore, high light areas of the nighttime image are regarded as important information in our method. Moving objects and high light areas preserve the fidelity of important information of the nighttime video.

In this paper, we propose a new method to enhance the context of nighttime surveillance by image fusion. The contrast and signal to noise ratio are improved in the fused image. Furthermore, we could exploit more information from the fused image than a single nighttime image. The fused image increases the information density and provides good input to high-level behavior analysis and understanding.

The main algorithm can be divided into four parts: (1) Enhancement of nighttime video, (2) Motion detection, (3) Estimation of illumination characteristics. (4) Image fusion. Figure 2 shows the overall framework.

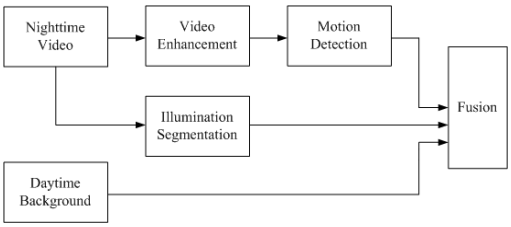


Figure 2. Framework of the algorithm

The paper is organized as follows. Section 2 describes the proposed method. Experimental results are shown in Section 3. Conclusions are made in Section 4.

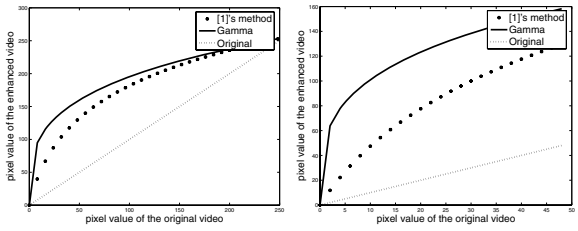
2. Context Enhancement of Nighttime Surveillance by Image Fusion

2.1. Enhancement of Nighttime Video

As we mentioned in Section 1, due to the low contrast of the nighttime video, it becomes more difficult to accurately extract moving objects from the dark background. How to improve the contrast of nighttime video while suppressing the noise becomes a crucial problem. Bennett et al. [1] present an adaptive Spatio-Temporal Accumulation (ASTA) filter for reducing noise in low dynamic range (LDR) videos. In addition, a tone mapping function is proposed to enhance LDR videos. The whole process takes approximately one minute per frame of size 640×480. In this paper, we demonstrate a successful application of this tone mapping function to the nighttime video enhancement. The tone mapping function is given by [1]:

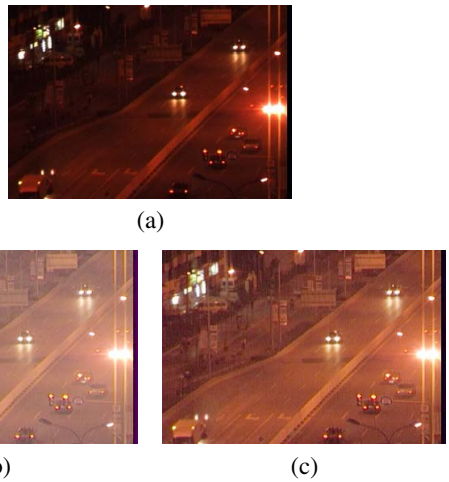
$$y = 255 \cdot \frac{\log(\frac{x}{255}(\Psi - 1) + 1)}{\log(\Psi)} \qquad (1)$$

where x is the pixel value of the original nighttime video, y is the value of the enhanced video, Ψ is a parameter. The pixels of the original video are adjusted in an adaptive manner based on the local luminance. This mapping function exhibits a similar characteristic as the traditional Gamma Correction. As we can see from Figure 3 (a) and Figure 3 (b), this mapping function presents a better performance in dark areas compared with gamma correction. The traditional gamma correction exhibits a steep slope near the origin. Originally dark pixels will be all mapped to medium values, which may result in loss of details in dark areas [1]. Figure 4 (a) shows the original nighttime video, comparison of enhancement between gamma correction and [1] 's method is shown in Figure 4 (b) and Figure 4 (c).



(a) (b)

Figure 3. (a) Mapping function, (b) Mapping function near the origin



(a)

(b) (c)

Figure 4. Enhancement of nighttime video. (a) Original nighttime video, (b) Enhanced nighttime video by Gamma Correction, (c) Enhanced nighttime video by [1] 's method

2.2. Motion Detection

In this section, mixed Gaussian model [2] is adopted to motion detection. Mixed Gaussian model is an effective solution to real time motion detection due to its self learning capacity. In addition, this method is robust to variations in lighting, moving scene clutter, multiple moving objects compared with other methods. The mixed Gaussian model for real time motion detection can be briefly summarized as follows [2]:
(1) Every new pixel value is checked against the existing K Gaussian distributions. A match is defined as a pixel value within 2.5 standard deviations of a distribution.
(2) Sort Gaussians and determine if the Gaussian is background.
(3) Adjust the prior weights of the distributions.
(4) Adjust the mean and standard deviation for matched distributions.
(5) If there is no match, replace least probable Gaussian and set mask pixel to background.

Figure 5 shows the comparison of the result of motion detection before and after video enhancement. Some parts of the moving objects are missing in Figure 5(c), which is the direct motion detection result of the original nighttime video. In Figure 5(d), moving objects are accurately extracted, which demonstrates the positive effect of the proposed enhancement.

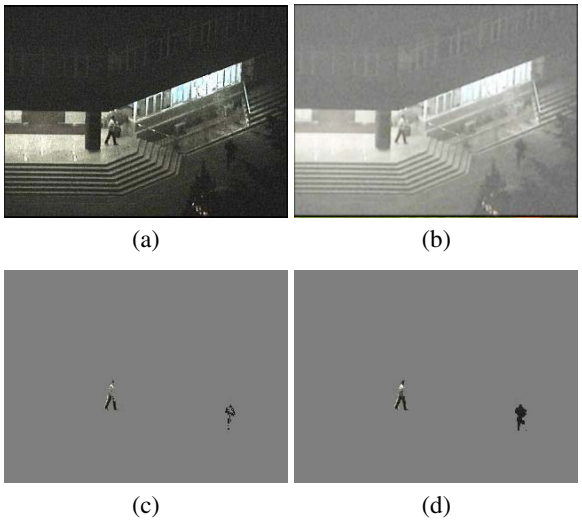


Figure 5. (a) Original nighttime video, (b) Enhanced nighttime video, (c) Motion detection result of original nighttime video, (d) Motion detection result of enhanced nighttime video

2.3. Estimation of Illumination Characteristics

The problem is how to represent the illumination characteristics of nighttime images precisely. Here, we propose to make use of Retinex theory [4] to obtain the illumination characteristics of nighttime images.

There are many successful applications of Retinex theory in dynamic range compression and face recognition to deal with the variation of illumination. We assume that the image I can be represented by the product of reflectance of the scene R and illumination coming from the light source L, as shown in equation 2:

$$I = R \times L \tag{2}$$

Many methods have been put forward to separate a reflection image R (the albedo) and an illumination image L at each point from a single image or image sequences. In Retinex theory, the illumination L can be considered as the low frequency component of image I, which has been theoretically justified by spherical harmonics analysis, and hence can be estimated by using a low-pass filter. The illumination coming from the light source is modeled as:

$$L = G * I \tag{3}$$

where $*$ is the convolution operation, G is the Gaussian smoothing kernel.

In this paper, we represent the illumination characteristics of the nighttime image as the smoothed version of the original image. Using Retinex theory to separate reflectance image and illumination image has several advantages: First, the reflectance image and illumination image can be obtained from a single image instead of a sequence of images. Secondly, this method does not require any learning, so no training images are needed. Finally, there is no assumption about lighting sources and shadow.

2.4. Image Fusion

Now we consider how to combine the high quality daytime image and important information of nighttime image together. We obtain the result of motion detection and illumination characteristics as in Section 2.2 and 2.3 respectively. The weighted sum technique is used here. We first normalize the illumination image L to $[0, 1]$ by:

$$L(x,y) = \frac{L(x,y) - min(L(x,y))}{max(L(x,y)) - min(L(x,y))} \tag{4}$$

The result of motion detection M is a binary mask. We get the weight of nighttime image by:

$$W(x,y) = \begin{cases} 1 & M(x,y) + L(x,y) \geq 1; \\ L(x,y) & M(x,y) + L(x,y) < 1. \end{cases} \quad (5)$$

The weight $W(x,y)$ is in the range $[0,1]$. The weights of the moving objects are set to 1 to highlight the moving objects. The final image can be obtained by:

$$F = W \cdot N + (1 - W) \cdot D \quad (6)$$

where F is the final image, N is the nighttime image, D is the daytime background. Figure 6 shows an example.

(a) (b)

Figure 6. Context Enhancement of a nighttime video. (a) Original nighttime video, (b) Result of context enhancement

3. Experiments

Extensive experiments have been carried out under various scenes. Some of the videos were captured through a Panasonic NV-MX500 digital video (Figure 4). Others were captured by a common CCD camera Panasonic WV-CW860A with the size of 320×240.

Experimental results show that our method is robust and effective (see Figure 6 for an example). The result of the context enhancement maintains the fidelity of important information of both daytime and nighttime images. Figure 7 shows the comparison between our method and Li et al's method [5]. Figure 7(a) is taken directly from [5], not by our implementation. As we can see from Figure 7(b), our method produces visually more pleasant result. In addition, moving objects are clearly represented.

(a)Li's method [5] (b)Our method

Figure 7. Comparison between Li 's method [5] and our method.

4. Conclusions

In this paper, we have presented a novel approach to enhancing the context of nighttime surveillance by image fusion. The proposed method provides a real time and robust solution to front-end image pre-processing in nighttime video surveillance. In addition, our method is flexible in adapting to different scenarios. The resultant image contains a more accurate and comprehensive description of the scene which is more useful for human visual and machine perception.

Acknowledgement

This work is partly supported by NSFC (Grant No. 60121302, 60335010), National Basic Research Program of China (Grant No. 2004CB318110) and MST of PRC (Grant No.2004DFA06900).

References

[1] E. P. Bennett and L. McMillan. Video enhancement using per-pixel virtual exposures. In *Proc. of ACM SIGGRAPH*, volume 24, July 2005.

[2] C.Stauffer and W.E.L.Grimson. Adaptive background mixture models for real-time tracking. In *Proc. of the 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 246–252, June 1999.

[3] A. Ilie, R. Raskar, and J. Yu. Gradient domain context enhancement for fixed cameras. In *Proc. of ACCV*. Jeju Island, Korea, January 2004.

[4] E. Land. The retinex theory of color vision. *Sci.AMer.*, (237):10–128, 1997.

[5] J. Li, S. Z.Li, Q. Pan, and T. Yang. Illumination and motion-based video enhancement for night surveillance. In *Proc. of the 2nd Joint IEEE International Workshop on VS-PETS*, pages 169–175. Beijing, China, October 2005.

[6] R. Raskar, A. Ilie, and J. Yu. Image fusion for context enhancement and video surrealism. In *Proc. of the 3rd international symposium on Non-photorealistic animation and rendering(NPAR)*, pages 85–152. Annecy, France, June 2004.