MVA2005 IAPR Conference on Machine VIsion Applications, May 16-18, 2005 Tsukuba Science City, Japan
8-26

Regional Video Insertion Through Robust Region Tracking In The Presence Of Occluding Objects

Chang-Hwan Lee, Dae-Woong Kim, Ki-Sang Hong
Pohang University of Science and Technology (POSTECH)
{special, dwkim, hongks} @ postech.ac.kr

Abstract

In this paper, we propose a method to solve the occlusion problem for insertion of a synthetic image into the planar region. This paper consists of the two main procedures; a tracking algorithm in the presence of occlusion and a segmentation algorithm to separate the occluding object from the planar region. To track the planar region in a robust way, a new energy function is proposed, and the planar region is tracked by extracting the homography that minimizes the energy function between images using the LM (Levenbert-Marquardt) algorithm. The energy function includes the intensity difference considering illumination change and slow motion prior. Also, it is modeled with the skipped mean estimator to exclude the effect of the occlusion. In each tracked planar region, the occluding object is segmented and replaced with synthetic objects by modeling each pixel with the mixture of Gaussian. Even if the occluding object exists, the planar region is tracked successfully by using the tracking algorithm. And experimental results show that the synthetic images are naturally inserted into the desired planar region.

1 Introduction

The algorithm of image insertion is applied to a wide range of video editing technology. In research on image insertion, the occlusion problem has been a difficult problem.

Many algorithms have been proposed for this problem. Tracking with the template matching method [1][2] can be reliable over a short duration, but it deals poorly with the occluding object. The approach using background subtraction [3][4] is only applied to the static camera. Also, the method using global statistics [5][6] such as color histograms cannot track the planar region exactly in the presence of occlusion. The segmentation algorithm using the color chromakeying [7] cannot segment moving object from the planar region which have a complex color distribution.

These algorithms commonly have problems because no appropriate model representing the occluding objects exists. In this paper, we propose a method to solve the occlusion problem for insertion of a synthetic image into a planar region. Two main algorithms are proposed in this paper; a tracking algorithm of the planar region in the presence of occlusion and a segmentation algorithm to separate the occluding object from the planar region.

To track the planar region in a robust way, a new energy function is proposed, and the planar region is tracked by extracting the homography that minimizes the energy function between images using the LM (Levenbert-Marquardt) algorithm. The energy function includes the intensity difference and slow motion prior. To exclude the effect of the occluding object, the skipped mean estimator is used to model the energy function. Also, it is modeled with the illumination parameter to improve the robustness when the illumination changes.

After the planar region is tracked, the occluding object is segmented from the planar region. Each pixel is modeled with MoG (the mixture of Gaussian) model. The EM (Expectation-Maximization) algorithm is used to estimate the parameters of MoG, and we use the internal parameters of MoG as threshold values to segment the occluding object.

The novelty of the tracking algorithm is that the planar region is tracked successfully even if an occluding object exists. Also, through the proposed segmentation algorithm, the occluding object is segmented exactly even with noisy images by using the mixture of Gaussian model.

The remainder of this paper is organized as follows. In section 2, we briefly describe our tracking, segmentation, and insertion system. In section 3, we show some experimental results, and we conclude with some remarks in section 4.

2 Proposed Algorithm

Fig. 1 shows the block diagram of our algorithms. First, the planar region is tracked in the presence of occlusion. After that, the occluding object is segmented from the planar region. Finally, the synthetic image is inserted into the planar region. More detailed algorithms are presented in section 3, 4.

2.1 Tracking algorithm

In this section, our tracking algorithm is introduced to solve the occlusion problem. To track the planar region robustly in the presence of occlusion, an energy function is modeled using a robust estimator.

The homography relation between the current frame and the previous frame is represented by

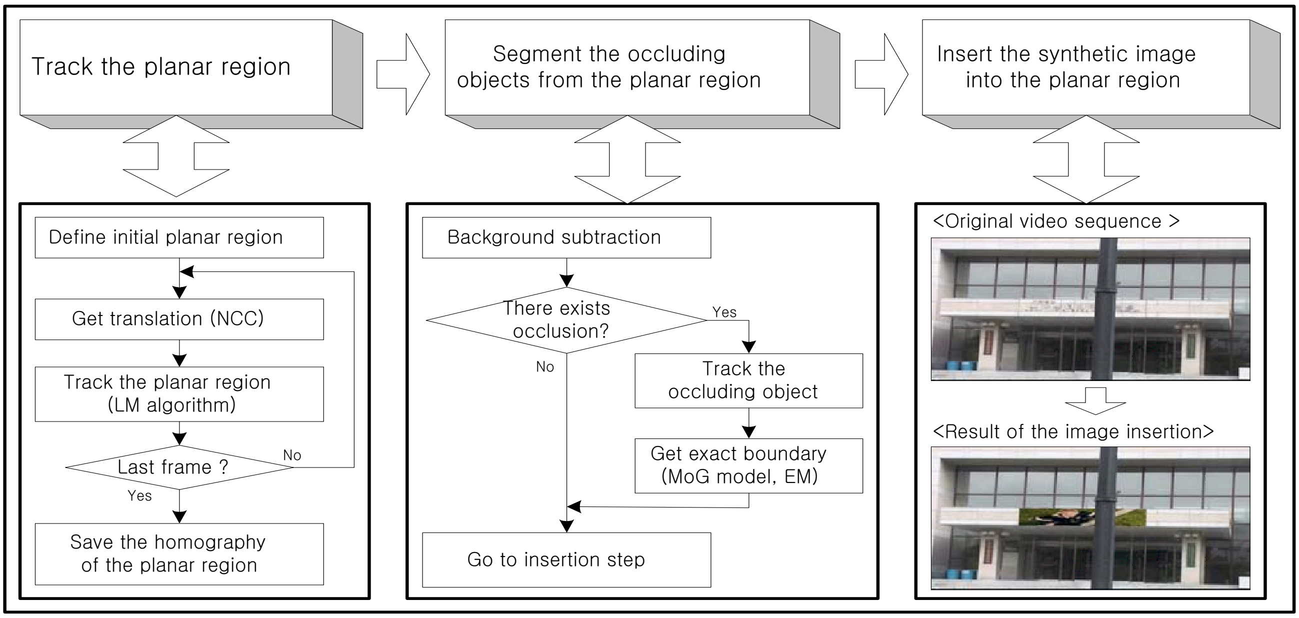


Figure 1. Block diagram of our proposed algorithm.

$$\begin{bmatrix} x_{i,t} \\ y_{i,t} \\ w \end{bmatrix} = \begin{bmatrix} m_0 & m_1 & m_2 \\ m_3 & m_4 & m_5 \\ m_6 & m_7 & 1 \end{bmatrix} \begin{bmatrix} x_{i,t-1} \\ y_{i,t-1} \\ 1 \end{bmatrix}, \quad (1)$$

where $x_{i,t}, y_{i,t}$ is the coordinate of the current frame, $x_{i,t-1}, y_{i,t-1}$ is the coordinate of the previous frame, and $m_0, m_1, \ldots, m_7$ are elements of the homography.

The illumination change is modeled with parameters μ, ε, and the illumination change between the current frame and the previous frame can be represented by

$$I_t(x_{i,t}, y_{i,t}) = \mu I_{t-1}(x_{i,t-1}, y_{i,t-1}) + \varepsilon . \quad (2)$$

Finally, all parameters to be estimated are represented by

$$\mathbf{M_t} = [m_0 \quad m_1 \quad m_2 \quad m_3 \quad m_4 \quad m_5 \quad m_6 \quad m_7 \quad \mu \quad \varepsilon]^T . \quad (3)$$

Before tracking, we get the translation (x_t, y_t) between the current and the previous frame using NCC (Normalized Cross Correlation), and we use it as the initial homography to be estimated.

Among a bulk of noisy data in the overlapping region, there may be outliers such as moving objects which contribute too much to the overall solution in a negative direction. A way of overcoming the outlier problem is by robust statistics. To track the planar region exactly, the energy function for the intensity difference considering the illumination change is represented by

$$E_\rho(\mathbf{M_t}) = \frac{1}{N}\sum_{i\in\Omega} \rho(I_t(x_{i,t}, y_{i,t}) - (\mu I_{t-1}(x_{i,t-1}, y_{i,t-1}) + \varepsilon)) , \quad (4)$$

where ρ is the M-Estimator [8], which is used to minimize the effect of the occluding object, μ, ε is the illumination change parameter, Ω is tracked region of the previous frame, N is the number of pixels in Ω, and $I_t(x, y)$ is the intensity of x, y coordinate at *t*-th frame.

The energy function for the slow motion prior is represented by

$$E_p(\mathbf{M_t}) = (\mathbf{M_t} - \xi)^T \mathbf{C}^{-1} (\mathbf{M_t} - \xi) , \quad (5)$$

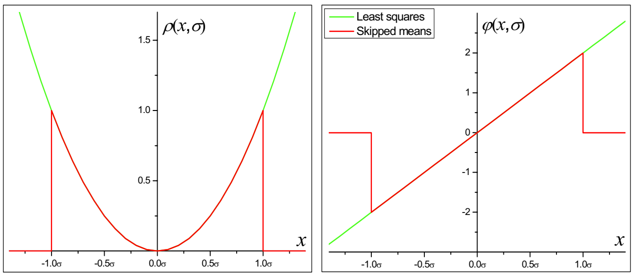
where ξ is the mean vector and $\mathbf{C}$ the covariance matrix. If each parameter strays far from the mean value, the energy function increases. Therefore, it prevents the parameter from changing abruptly.

The final energy function to be minimized is given by

$$E(\mathbf{M_t}) = E_\rho(\mathbf{M_t}) + \lambda E_p(\mathbf{M_t}) , \quad (6)$$

where λ is Lagrange-multiplier.

The planar region is tracked by extracting the parameter $\mathbf{M_t}$ that minimizes the energy function of equation (6) using the LM (Levenberg-Marquardt) algorithm. As we use the robust estimator to model the energy function, the planar region can be tracked exactly even if occluding objects exist. The choice of the ρ-functions results in different robust estimators, and the robustness of a particular estimator refers to its insensitivity to outliers. In this paper, we used a Huber-type skipped means estimator as shown in Fig 2. This rejects everything which is more than 5.2 median deviations away from the median and takes the mean of the remainder, and the second derivative is always positive, which is a very important property for incorporating the robust estimator into LM optimization.



(a)ρ – function (b)φ – function

Figure 2. Outlier process of the least squares estimator and the skipped mean estimator. (a) ρ-function. (b) The derivative of ρ-function.

2.2 Segmentation algorithm

Before starting segmentation, the frame images are aligned to the coordinate of the occluding object as shown in Fig. 3 (a) through the following procedure; 1) get the initial mask representing the occluding object through background subtraction, 2) get the homography of the occluding object by tracking the region corresponding to the initial mask, 3) align the frame images using the homography.

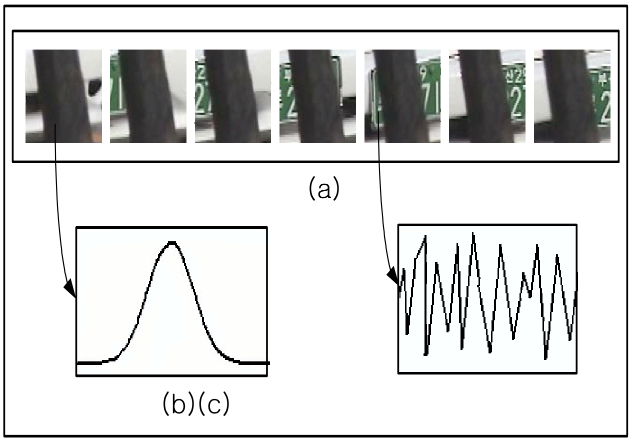


Figure 3. Candidate of the occluding region and expected distribution. (a) Aligned candidate of occluding region. (b) The distribution of the occluding object region. (c) The distribution of the non-occluded region.

After all the frames are aligned, the distribution of one pixel is represented by equation (7) to model the distribution of Fig. 3 (b), (c) in the aligned image of Fig. 3 (a).

$$p(d_j \mid \boldsymbol{\alpha},\boldsymbol{\mu},\boldsymbol{\Sigma}) = \sum_{i=1}^{M} \alpha_i p_i(d_j \mid \mu_i, \sigma_i^2) , \qquad (7)$$

where d_j is the observation of the j-th data, α_i is the mixing probability of the i-th Gaussian distribution, μ_i, σ_i^2 are the mean and variance of the i-th Gaussian distribution, $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, ..., \alpha_M]^T$, $\boldsymbol{\mu} = [\mu_1, \mu_2, ..., \mu_M]^T$, and $\boldsymbol{\Sigma} = [\sigma_1^2, \sigma_2^2, ..., \sigma_M^2]^T$.

To estimate all parameters in equation (7), the likelihood function is defined by

$$L(\mathbf{X} \mid \Theta) = \prod_{j=1}^{K} \sum_{i=1}^{M} \alpha_i p_i(d_j \mid \mu_i, \sigma_i^2) , \qquad (8)$$

where $\Theta = [\alpha_1, ..., \alpha_M, \mu_1, ..., \mu_M, \sigma_1^2, ..., \sigma_M^2]^T$ is the MoG parameter to be estimated, $\mathbf{X}(=[d_1, ..., d_K]^T)$ is the data observation of one pixel, and K is the number of frames with occluding objects.

We used the EM (Expectation-Maximization) algorithm [9] to estimate parameters $\boldsymbol{\alpha},\boldsymbol{\mu},\boldsymbol{\Sigma}$ which maximize the probability of equation (8). Once all the parameters are estimated, the pixel is classified as the occluding object through thresholding (T_1) with $\max(\alpha_1, ..., \alpha_M)$. The occluding object is segmented from the planar region by processing the above procedure for all pixels.

3 Experimental Results

We present the experimental results of our algorithm. We used 5 as the number of the mixture of Gaussian M and 0.8 as threshold T_1. A SONY digital handicam LCM-TRVX is used in acquiring the input sequence, and a Pentium IV 1.8G processor system is used to process the algorithm.

In Fig. 4, the building signboard was tracked exactly in the presence occlusion(the black pillar), the black pillar was segmented from the signboard. Finally, the movie was naturally inserted into the building signboard.

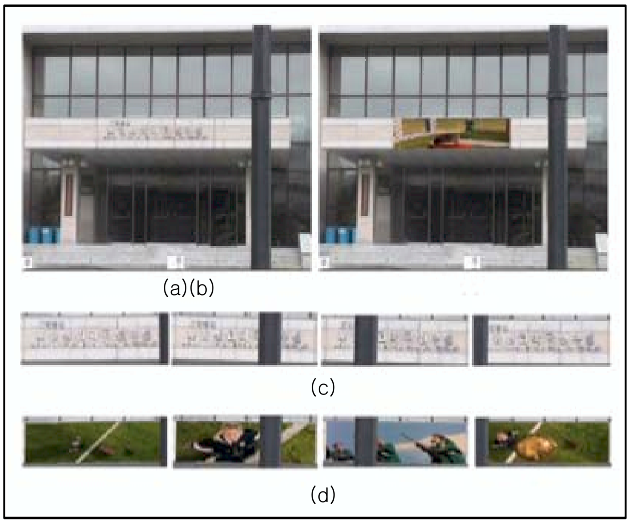


Figure 4. The building signboard tracking and insertion of another image. (a) Original image of initial frame. (b) Insertion result of synthetic image in initial frame. (c) Input sequences of frame 100, 160, 320, and 440. (d) Insertion result of frame 100, 160, 320, and 440.

Fig. 5, the license plate is tracked exactly although the occluding object (tree region) is relatively large. In addition, the tree region is segmented from the license plate as shown in second row in Fig. 5 (c). The same parameter values are used as in Fig. 4.

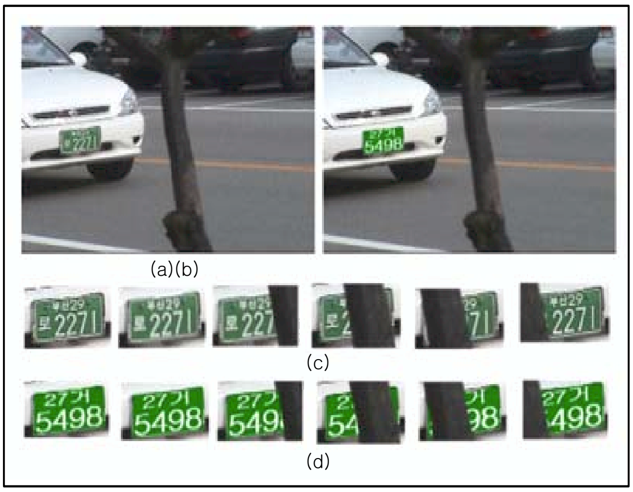


Figure 5. The License plate tracking and insertion of another image. (a) Original image of initial frame. (b) Insertion result of synthetic image in initial frame. (c) Input sequences of frame 0, 23, 46, 56, 66, and 72. (d) Insertion result of frame 0, 23, 46, 56, 66, and 72.

Fig. 6 shows the tracking result when illumination changes exist, and Fig. 6 (d) shows the result. It showed that the planar region was tracked robustly as we used the illumination change parameter.

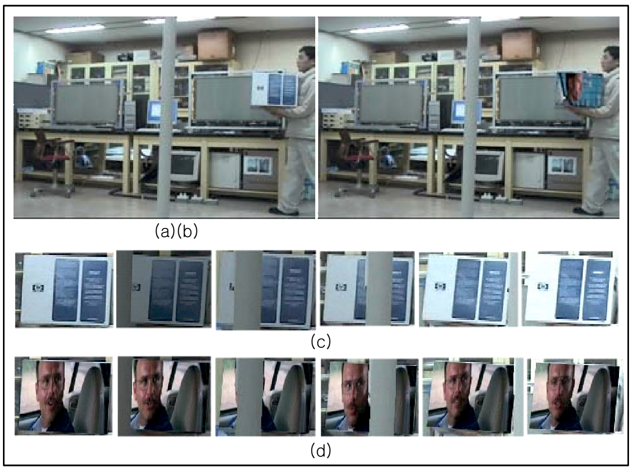


Figure 6. Insertion result in the presence of illumination change. (a) Original image of initial frame. (b) Insertion result of synthetic image in initial frame. (c) Input sequences of frame 120, 126, 132, 139, 146, and 152. (d) Insertion result of frame 120, 126, 132, 139, 146, and 152.

In these experiments, the planar region was tracked in the presence of occluding object, and the rigid-occluding object was segmented from the planar region. Also, we showed that the tracking method proposed in this research was robust to the illumination change.

4 Conclusion

We presented a framework for inserting synthetic images into a planar region. To estimate parameters robust to outliers such as misaligned pixels or occluding moving objects in the overlapping region, we adapted a skipped mean estimator, and incorporated it with LM optimization. The occluding object is segmented from the planar region with MoG (the mixture of Gaussian) model, and EM (Expectation-Maximization) algorithm is used to estimate the parameters of MoG. The occluding object was successfully segmented if the planar region follows a rigid motion. However, we need to improve our segmentation algorithm to handle non-rigid occluding objects. In addition, the computational cost of the segmentation algorithm is relatively high because our segmentation algorithm is a post-processing approach that uses all frame images. More research is required to develop methods to reduce the computational cost.

Acknowledgments

This work was partly supported by the Korea Research Foundation and the Brain Korea 21 Project.

References

[1] S. Niyogi and W. T. Freeman, "Example based head tracking." *In Proceedings 2nd International Conference on Automatic Face and Gesture Recognition, Killington, VT, USA, 1996.*

[2] Frederic Jurie and Michel Dhome, "Real time robust template matching," *BMVC 2002.*

[3] Omar Javed, Khurram Shafique and Mubarak Shah, "A hierarchical approach to robust background subtraction using color and gradient information," *IEEE Workshop on Motion and Video Computing, 2002.*

[4] C. Stauffer and W. Grimson, "Adaptive background mixture models for real-time tracking," *In Proceedings CVPR, pp. 246-252, 1999.*

[5] Dorin Comaniciu, Visvanathan Ramesh, Peter Meer, Senior Member, "Kernel-based object tracking," *IEEE Transactions on PAMI, vol. 25, No. 5, May 2003.*

[6] Tyng-Luh Liu and Hwann-Tzong,"Real-time tracking using trust-region methods," *IEEE Transactions on PAMI, vol. 26, No. 3, March 2004.*

[7] Jinhui Pan, Chia-Wen Lin, Chuang Gu, and Ming-Ting Sun, "A robust video object segmentation scheme with prestored background information," *IEEE International Symposium on Circuits and Systems, 2002.*

[8] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel, "Robust statistics: The approach based on influence function," *New York: Wiley, 1986.*

[9] A.P. Dempster, N.M. Laird, and D.B. Rubin. "Maximum likelihood from incomplete data via the EM algorithm," *J. Royal, Stat. Soc. B, 1977.*