Performance evaluation of foreground modeling in moving foreground segmentation

Xiang Zhang,[a] Jie Yang,[a] and Zhi Liu[b]
[a]Shanghai Jiaotong University, Institute of Image Processing and Pattern Recognition, Shanghai 200240, China
[b]Shanghai University, School of Communication and Information Engineering, Shanghai 200072, China
E-mail: hover_chang@sjtu.edu.cn

Abstract. Nonparametric statistical modeling of background and foreground has been widely used for moving foreground segmentation from video sequences. In this work, a simple metric is presented to evaluate the performance of various foreground models. The proposed metric allows us to test the robustness of the foreground model to the motion and deformation of the moving foreground. Experiments are performed on five typical foreground models, showing that the proposed metric is effective.
[DOI: 10.1117/1.3094946]

Subject terms: performance evaluation; foreground segmentation; foreground modeling.

Paper 080856LR received Nov. 4, 2008; revised manuscript received Jan. 21, 2009; accepted for publication Jan. 23, 2009; published online Mar. 6, 2009.

1 Introduction

Foreground segmentation plays an important role in a wide range of computer vision applications. Foreground modeling[1,2] has been recently used in conjunction with background modeling[3] for segmentation. Foreground and background models can be created in a consistent fashion, and the nonparametric statistical model[4] is the frequently used model now.

To compare the performance of different segmentation algorithms, a few metrics are presented. Precision and recall[1] are the standard measures used in current literatures. The two measures compare segmentations with the ground truth in a pixel-level way, ignoring region-level information. Nascimento and Marques[5] proposed a region-level method to classify segmentation errors into detection failures, false alarms, splits, merges, and split/merges. The method presented in Ref. 6 is also a pixel-level approach, which is designed to compare the ground truth with detected silhouettes used in gait recognition.

We have known that segmentation performance is largely dependent on foreground modeling. Although some metrics have been presented for the comparison of segmentation performance, no metrics are reported for the comparison of model performance. We present a novel metric to compare the performance of different foreground models. Further, the proposed metric is capable of explaining the difference in segmentation performance of different algorithms from the perspective of foreground modeling. This metric is also helpful in developing new foreground models.

This work is organized as follows. The proposed metric is described in Sec. 2. Experimental results are given in Sec. 3, followed by conclusions in Sec. 4.

2 Proposed Metric

Some nonparametric methods use multiple features as statistical variables.[7,8] Although the performance improvement of segmentation is distinct by the use of multiple statistical variables, it is still difficult to get full segmentation, because the statistical analysis cannot resolve the uncertainty of the foreground, such as the motion and deformation of a moving object. Thus we do not consider those models taking multiple features for statistical analysis, but only those models taking advantage of multiple features in a way different from statistical analysis. For example, the model proposed in Ref. 9 uses the color histogram to select the most suited samples for foreground modeling from all historical segmentations.

Let I^t be the input image at time instant t, and I_n^t be the color vector of a pixel in position n (for fair comparison, the YUV color space is used for all models). All nonparametric statistical foreground models of I^t can be denoted in the form of $\phi^t=\{Y^1, Y^2, \ldots, Y^R\}$, where each element in ϕ^t consists of all pixels labeled foreground at certain time instants, and R is the frame length of ϕ^t. Let X^t be the binary ground truth of image I^t, with 1 and 0 denoting foreground and background pixels, respectively. Let X^r be the binary mask of Y^r, with 1 and 0 denoting pixels labeled foreground in Y^r and all other pixels, respectively. Let $Q^{t,r}$ be the XOR image of X^t and X^r, where each pixel $Q_n^{t,r}$ of $Q^{t,r}$ is defined as

$$Q_n^{t,r}=\begin{cases}1, & \text{if } X_n^t=X_n^r=1\\ 0, & \text{if } X_n^t=X_n^r=0\\ -1, & \text{if } X_n^t\neq X_n^r\end{cases}. \tag{1}$$

The performance of ϕ^t can be measured with the proposed metric as

$$M(\phi^t)=\frac{\Sigma_r\Sigma_n Q_n^{t,r}}{R^*\Sigma_n X_n^t}. \tag{2}$$

According to the definition of M, the most desirable foreground model should be such that each element X^r of the model is the same as X^t. In other words, each element in the most desirable model is a segmentation in which the moving object shows the same shape in the same place as the moving object in the current frame. In the previous definitions, Y is a set of all pixels classified as foreground in a certain frame, where each pixel is a color vector, and X is a binary image with the same size as the original image I.

To test the robustness of the model to an object's motion, we can compute the position distance of corresponding objects between X^t and X^r. To test the robustness of the model to an object's deformation, we can compute the shape distance of the moving object between X^t and X^r using various shape descriptors. However, to find corresponding features is a very labor intensive task and error

0091-3286/2009/$25.00 © 2009 SPIE

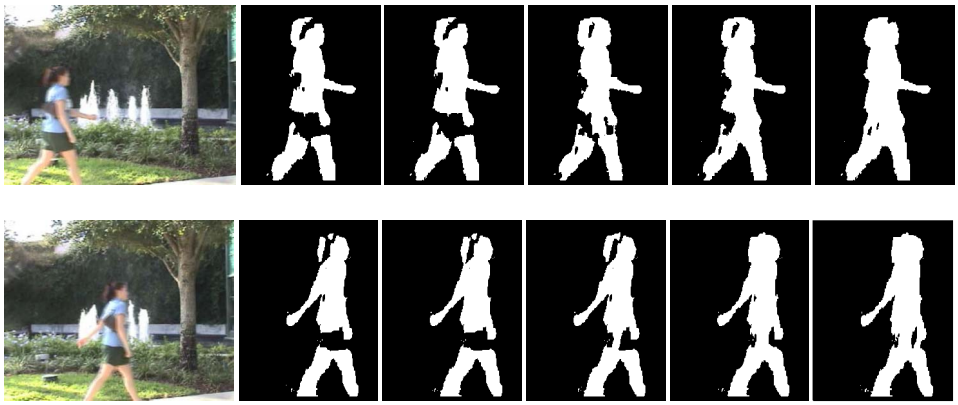

Fig. 1 Two images of the first sequence and corresponding segmentations with different models. See text for details.

prone, and X^r is often corrupted with splits and defects, making shape similarity measurement unreliable. The advantage of the proposed metric is that the computation of distance is avoided, and the uncertainty of the moving object is implicitly highlighted in the XOR operation.

3 Experimental Results

The proposed metric is applied to characterize the performance of five foreground models. Each foreground model is used in conjunction with a background model to classified pixels based on energy minimization, as in Ref. 1. The five foreground models are simply described as follows.

The first foreground model, the general foreground model ϕ_G^t,[1] can be denoted as $\phi_G^t = \{G^{t-1}, \ldots, G^{t-r}, \ldots, G^{t-R}\}$, where G^{t-r} is a set of all pixels labeled foreground at time instant $t-r$. Then we consider two ways of using motion information for foreground modeling. For simplicity, only single object detection is considered. The centroid of the moving object in the current frame is predicted by a Kalman filter. Then we move all elements in ϕ_G^t from their centroids to the predicted centroid, resulting in the second foreground model ϕ_P^t.[10] A substitution of prediction is tracking. The moving object is tracked from one frame to the next by the mean-shift tracker. All elements in ϕ_G^t are shifted from their centroids to the centroid of the tracking window in the current frame, leading to the third foreground model ϕ_T^t.

The fourth foreground model[9] takes advantage of the shape and motion information for foreground modeling. First, predetection is carried out on the current frame with ϕ_G^t. Then we align all historical segmentations to the presegmentation based on the centroid of the moving object. The shape similarity of the moving object between presegmentation and each aligned segmentation is measured based on the color histogram. The R frames of aligned segmentations, which have the largest similarity values, are chosen to form the fourth foreground model ϕ_H^t. The fifth foreground model ϕ_U^t consists of the same historical segmentations as ϕ_H^t, but with each element unaligned, which means the motion information is ignored.

The first column of Fig. 1 shows two typical images of the first test sequence with serious color similarity between foreground and background. Detected foreground by ϕ_G^t, ϕ_U^t, ϕ_P^t, ϕ_T^t, and ϕ_H^t is shown from the second to the sixth columns, respectively. Segmentations of 32 frames are compared with the ground truth in terms of recall,[1] which is

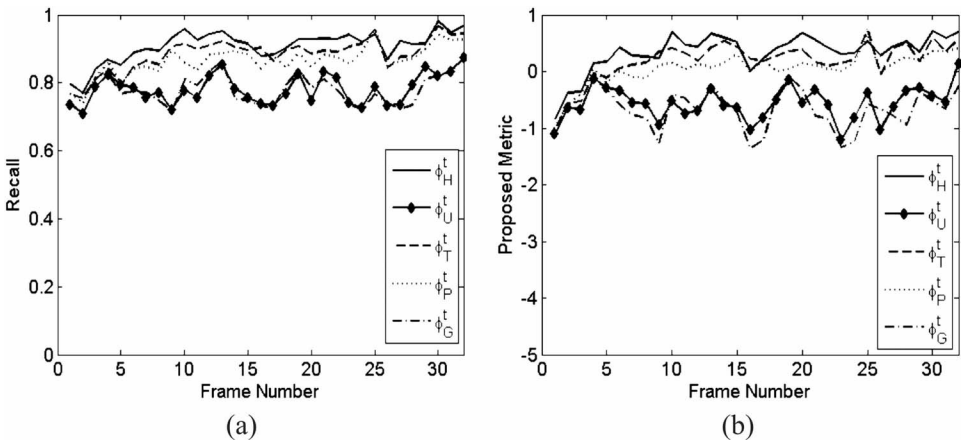

Fig. 2 Performance test of the first sequence: (a) is recall and (b) is the propose metric.

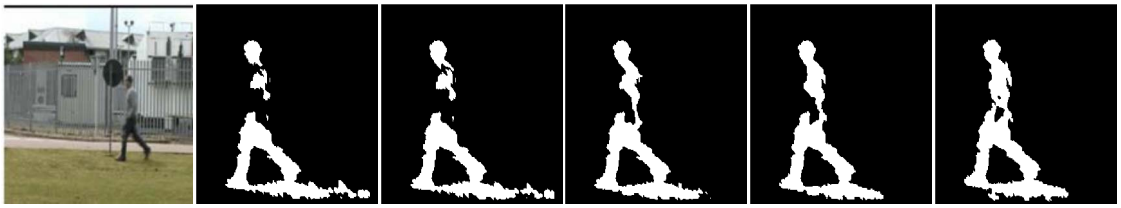

Fig. 3 The second test sequence. See text for details.

able to characterize the robustness of the segmentations to splits and defects due to the color similarity problem. The performance test is shown in Fig. 2.

Figure 2 shows that the segmentation performance is certainly dependent on the model performance. By the use of motion information, the model accuracy is largely improved compared with ϕ_G^t; as a result, segmentations with much better recall are derived by ϕ_P^t and ϕ_T^t. The segmentations by ϕ_T^t are a little better than the segmentations by ϕ_P^t, because of the better model performance of ϕ_T^t. The reason for this is that the information in the current frame is used by the tracker but not by the predictor.

The shape alone cannot provide a notable improvement in modeling and segmentation. However, combining motion and shape, as ϕ_H^t, displays more obvious improvement than using motion or shape alone. Some foreground pixels still cannot be detected in the last of Fig. 1. This suggests the use of finer features for shape representation, such as the Zernike moment descriptor and the Fourier descriptor. The performance of different shape descriptors for foreground modeling also can be identified by the proposed metric.

Experimental results on the second test sequence are shown in Figs. 3 and 4. This sequence also can be seen in Ref. 11. The second column to the last column in Fig. 3 are segmentations by ϕ_G^t, ϕ_U^t, ϕ_P^t, ϕ_T^t, and ϕ_H^t, respectively. We can think that the same curve as Fig. 2(a) can be obtained by observing Fig. 3. By comparing Figs. 3 and 4, the same conclusions can be obtained as those from the first test sequence.

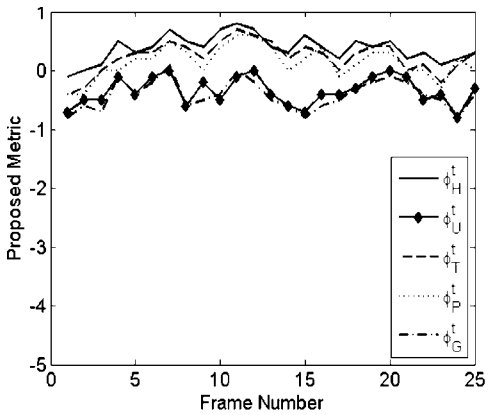

Fig. 4 Performance test of the second sequence with the proposed metric.

4 Conclusion

We propose a metric to check the robustness of different foreground models with the uncertainty of the moving foreground. This metric is able to explain the difference in segmentations by different algorithms from the perspective of foreground modeling. Our future work is to develop new foreground models to more effectively take advantage of the motion and shape information of the moving object based on the proposed metric.

Acknowledgments

The authors are grateful to the anonymous reviewers for their comments, which have helped us to improve this work. This study is supported by the China 863 High Tech. Plan (number 2007AA01Z164), and supported by the National Natural Science Foundation of China (numbers 60602012, 60772097, and 60675023).

References

1. Y. Sheikh and M. Shah, "Bayesian modeling of dynamic scenes for object detection," *IEEE Trans. Pattern Anal. Mach. Intell.* **27**(11), 1778–1792 (2005).
2. K. A. Patwardhan, G. Sapiro, and V. Morellas, "Robust foreground detection in video using pixel layers," *IEEE Trans. Pattern Anal. Mach. Intell.* **30**(4), 746–751 (2008).
3. C. Stauffer and W. E. L. Grimson, "Learning patterns of activity using real-time tracking," *IEEE Trans. Pattern Anal. Mach. Intell.* **22**(8), 747–757 (2000).
4. A. Elgammal, R. Duraiswami, D. Harwood, and L. Davis, "Background and foreground modeling using non-parametric kernel density estimation for visual surveillance," *Proc. IEEE* **90**, 1151–1163 (2002).
5. J. Nascimento and J. Marques, "Performance evaluation of object detection algorithms for video surveillance," *IEEE Trans. Multimedia* **8**(4), 761–774 (2006).
6. Z. Liu, L. Malave, and S. Sarkar, "Studies on silhouette quality and gait recognition," *IEEE Conf. Computer Vision Patt. Recog.*, Vol. 2, pp. 704–711 (2004).
7. A. Mittal and N. Paragios, "Motion-based background subtraction using adaptive kernel density estimation," *IEEE Conf. Computer Vision Patt. Recog.*, Vol. 2, pp. 302–309 (2004).
8. A. Criminisi, G. Cross, A. Blake, and V. Kolmogorov, "Bilayer segmentation of live video," *IEEE Conf. Computer Vision Patt. Recog.* Vol. 1, pp. 53–60 (2006).
9. X. Zhang and J. Yang, "Foreground segmentation based on selective foreground model," *IEEE Electronics Letters* **44**(14), 851–852 (2008).
10. X. Zhang and J. Yang, "A novel algorithm to segment foreground from a similarly colored background," *AEU, Int. J. Electron. Commun.* (in press).
11. A. Loza, L. Mihaylova, D. Bull, and N. Canagarajah, "Structural similarity-based object tracking in multimodality surveillance videos," *Mach. Vision Appl.* **20**(2), 71–83 (2009).