International Journal of Image and Graphics

A Multi-histogram Clustering Approach Toward Markov Random Field for Foreground Segmentation

Wenjia Yang
School of Automation, Beijing Institute of Technology, Beijing, 100081,
Beijing Key Laboratory of Automatic Control System (Beijing Institute of Technology),
Beijing City, 100081, China,
ywjowen@yahoo.com.cn

Lihua Dou
School of Automation, Beijing Institute of Technology, Beijing, 100081,
Beijing Key Laboratory of Automatic Control System (Beijing Institute of Technology),
Beijing City, 100081, China
doulihua@bit.edu.cn

Juan Zhang
School of Automation, Beijing Institute of Technology, Beijing, 100081,
Beijing Key Laboratory of Automatic Control System (Beijing Institute of Technology),
Beijing City, 100081, China
zhjuan@bit.edu.cn

Received (19 June 2009)
Revised (3 March 2010)
Accepted (13 October 2010)

This paper presents a Bayesian approach for foreground segmentation in monocular image sequences. To overcome the limitations of background modeling in dealing with pixel-wise processing, spatial coherence and temporal persistency are formulated with background model under a maximum a posterior probability (MAP)-MRF framework. Fuzzy clustering factor was introduced into the prior energy of MRFs for the new implementation scheme, where contextual constraints can be adaptively adjusted in terms of feature cues. Experimental results for several image sequences are provided to demonstrate the effectiveness of the proposed approach.

Keywords: background modeling; markov random field; clustering; foreground segmentation.

1. Introduction

Foreground segmentation in image sequences is very important in many application areas such as video surveillance, object tracking, and human computer interaction. A typical method is background subtraction involving calculating a reference image, subtracting each new frame from this image and thresholding the result. However, accurate foreground segmentation could be difficult due to potential variability, such as illumination or moving background, and camouflage. Many approaches for background subtraction have been proposed over the past decades, but usually differ in the ways of modeling the background. Pfinder[1] uses a single Gaussian background model per pixel, the pixel intensity is updated recursively by a linear adaptive filter. Stauffer and Grimson[2] modeled each pixel as a K mixture of Gaussian distributions to represent various background appearances. Elgammal[3] proposed the nonparametric estimation method for modeling background. They used kernel density estimation method to establish local membership of a pixel. But these methods can lead to misclassification of foreground if the background scenes are complex, for example, if the colors of the pixels in the region change widely over time, foreground objects with similar colors could easily be misclassified as background. Many researchers took the consideration of combing spatial and temporal information to solve the problem. Li et al.[4] proposed a method to employ the statistics of color co-occurrence to model the dynamic features associated with a background object. Sheikh and Shah[5] have exploited spatial correlation among pixels by using the position of a pixel along with the color. On the other hand, the pixel-pair relationship is attracting more and more researchers because of its advantage in propagating neighborhood information, which incorporate long-range dependencies between pixels. A dynamic framework of topology free HMM capable of dealing with sudden or gradient illumination changes is proposed in [6]. In his paper, the topology and parameter estimation is posed as a model selection problem with an MDL prior. Spatial color distribution[7] can be used to characterize background and foreground objects within dynamic scenes. Wang[8] has considered the background, shadow, and foreground to be stochastic processes, their spatial interaction constraint of neighboring pixels can be modeled by markov random fields. Another dynamic conditional random field model[9] was proposed to exploit temporal and spatial dependencies of consecutive segmentation fields. The work in [10] has started an important line of research in using motion and its spatial context. They use new features to capture the visual context and filling-in missing, motionless regions. When training examples are available, statistical learning methods can be used to segment foreground and background[11]. Stereo-based segmentation[12] seems to achieve the most robust results as background objects are correctly separated from the foreground independently from their motion- stasis characteristics. Hong Yang et al.[13] [14] proposed a spatial-temporal smooth model based on conditional random field, data dependencies are encoded into the model for accurate foreground segmentation.

In fact, a primary difficulty in foreground segmentation lies in modeling the relationship between pixel pairs precisely and robustly. In this letter, we present a clustering approach to represent the state of foreground and background at each pixel in the markov random fields (MRFs) statistical framework. Gaussian mixture background model is used to model likelihood energy. Besides that, a fuzzy clustering factor was introduced to form the prior energy whose feature cue is based on multi-histogram and fuzzy c-means clustering algorithm, which can enable us to exploit the spatial and temporal coherence to maintain the continuity of our segmentation. Our feature cue utilizes the spatial and temporal information adequately by analyzing characteristics among different intensity levels in an image sequence.

The rest part of the paper is organized as follows. In Section II we introduce the adaptive background modeling algorithm to build the likelihood energy. Section III introduces fuzzy clustering factor based on multi-histogram and fuzzy c-means clustering. The foreground segmentation process based on MRFs statistical framework is described in Section IV. In Section V, we demonstrate the effectiveness of our approach by providing some appealing experimental results. Finally, the conclusion is reached in Section VI.

2. Adaptive Background Model for Segmentation

Color is the most direct evidence used in the process of foreground segmentation. When dealing with dynamic background, a robust background model plays an important role in improve spatial coherence. However, all the observed data are not of the same importance when given the video sequence. Our approach must first model the background of the image sequence which is also referred in the work of [8]. Each pixel in the scene is modeled by a mixture of K Gaussian distributions. The probability that a certain pixel has a value of $g_k(x)$ can be written as

$$p(g_k(x)) = \sum_{i=1}^{K} w_{i,k}\eta(g_k(x), \theta_{i,k}) \tag{1}$$

where $w_{i,k}$ is the weight parameter of the corresponding Gaussian component. $\eta(g_k(x), \theta_{i,k})$ is the Normal distribution of the component. So our observation model for background is denoted as [8]

$$b_k(x) = \mu_k(x) + n_k(x) \tag{2}$$

where $b_k(x)$ is the intensity of a single pixel x at time k, and $\mu_k(x)$ is the intensity mean, $n_k(x)$ is the independent zero mean additive noise with variance $\sigma_k^2(x)$ at time k. Our entire background is expressed by $(\mu_k(x), \sigma_k^2(x))$, which is denoted as θ. An online EM approximation is used to train the GMMs based on the ideas from Stauffer and Grimson. Because the Gaussians are ordered by the value of $\frac{w}{\sigma}$, the first Gaussian distribution that has the highest value is chosen as the background model for each pixel.

$$\theta_k(x) = (\mu_{e,k}, \sigma_{e,k}^2(x))^T \tag{3}$$

where $e = argmax_i \frac{w_{i,k}(x)}{\sigma_{i,k}(x)}$

3. Fuzzy Clustering Factor on Multi-histogram

In order to provide the fuzzy clustering factor for the prior model in our statistical framework, we introduced fuzzy c-means (FCM) algorithm based on multi-histogram. FCM, an unsupervised fuzzy clustering technique, is a new multi-threshold selection method based on artificial intelligence and suits for the uncertain and ambiguous characteristics in image processing. But the conventional FCM clustering algorithm considers[15] only the intensity levels of the pixels and ignores the spatial distribution of pixels. A lot of methods are put forward in order to conquer this limitation [16] [17]. But all these methods increased the computation load when introducing spatial information into clustering, which may become uncontrollable in our foreground segmentation framework. To overcome this limitation, we extend the conventional histogram to a broader one and propose a new method to get our clustering factor. By analyzing different features among different intensity levels, spatial information can be used adequately. In [18] multi-histogram fuzzy clustering method is proposed, which the theory is the same as ours. But it is used for different application, and we extended the conventional histogram to a broader one and proposed a new method to get our clustering factor.

The FCM algorithm is an iterative partitioning method that produces an optimal c-partition which can be referred in [17]. In fact, besides frequency distribution of intensity levels, there are many other kinds of statistical information, such as the whole position and standard deviations of intensity levels, which can reflect the global spatial distribution. Given a grayscale image I with n intensity levels, we suppose

$$x_i^{'} = (mx_i, my_i, s_i, f_i, i) \tag{4}$$

where $i = 1, 2, ..., n$, $mx_i = \frac{\sum_{j=1}^{f_i} px_j}{f_i}$, $my_i = \frac{\sum_{j=1}^{f_i} py_j}{f_i}$,$p_j = (px_j + py_j)^{\frac{1}{2}}$, $s_i = \frac{\sum_{j=i}^{f_i}[p_j - \frac{\sum_{j=1}^{f_i} p_j}{f_i}]^2}{f_i}$. mx_i, my_i, s_i are the horizontal mean coordinates, vertical mean coordinates and standard deviations of all the pixels in level i respectively and present the spatial information of an image. Obviously the problem is translated into a multidimensional feature space while the problem does not become more complex. Suppose the image is 256 intensity levels and $256 * 256$ size, then the number of feature vectors of is $256 * m$, where m, the number of the extracted features, is chosen to five. So the number of feature vectors of our method is only $256 * 5$, while that of FCM algorithm is $256 * 256 * 1$. Apparently, the computation cost is far less than that of FCM algorithm. Apparently, it can not only decrease the computation cost by increasing several dimension, but also the spatial information used in clustering is increased greatly.

On the other hand, the weights of the feature vector have different dimensions, if we classify the image data directly, it will fail. We should normalize these weights to eliminate the negative effects of different dimensions. Its normalized features are

(a) car (b) clustering

Fig. 1. original frame and segmentation mask

listed as follows

$$\begin{aligned} x_{i1} &= \frac{mx_i}{max\{[(mx_i)^2 + (my_i)^2]^{\frac{1}{2}}\}}, \\ x_{i2} &= \frac{my_i}{max\{[(mx_i)^2 + (my_i)^2]^{\frac{1}{2}}\}}, \\ x_{i3} &= \frac{s_i}{max(s_i)}, \\ x_{i4} &= \frac{f_i}{max(f_i)}, \\ x_{i5} &= \frac{i}{n}, \end{aligned} \tag{5}$$

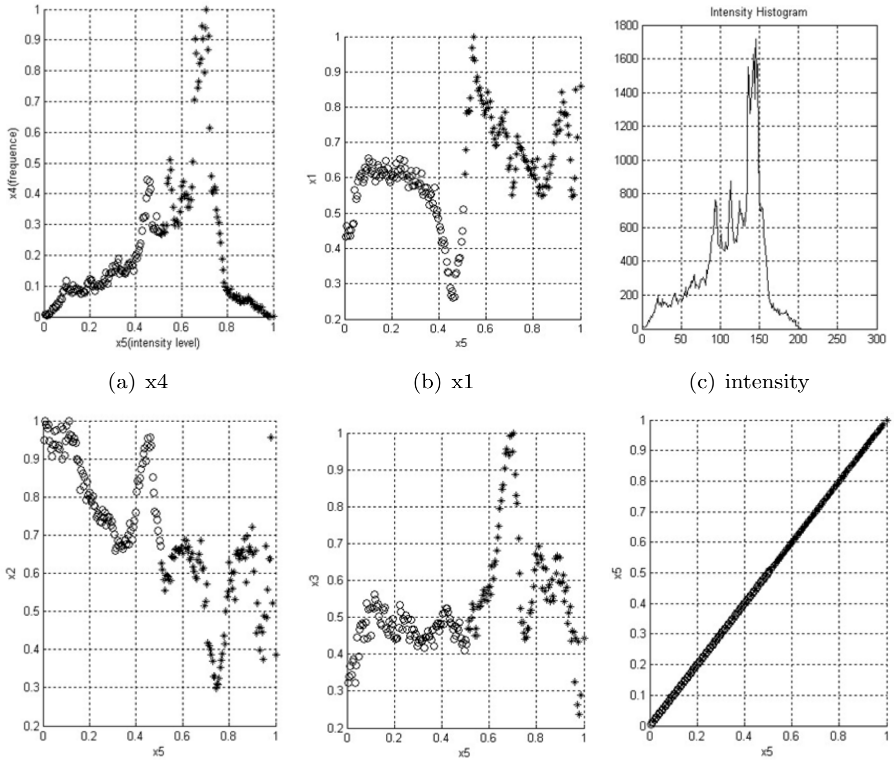
where $i = 1, 2, ..., n$, All of the above features compose the feature vector $x_i = \{x_{i1}, x_{i2}, x_{i3}, x_{i4}, x_{i5}\}$, and then the finite set $X = \{x_i\}$ of n data patterns is obtained. A looser normalized condition was substituted in the conventional FCM, which is to make the sum of the membership degree of all data samples to each class be n, that is,

$$\sum_{j=1}^{c} \sum_{i=1}^{n} \mu_j(x_i) = n \tag{6}$$

Under this condition, our membership function is represented as follows:

$$\mu_j(x_i) = n * (\frac{1}{||x_i - m_j||^2})^{\frac{1}{b-1}} (\sum_{p=1}^{c} \sum_{q=1}^{n} (\frac{1}{||x_i - m_j||^2})^{\frac{1}{b-1}})^{-1} \tag{7}$$

where $i = 1, 2, ..., n,\ j = 1, 2, ..., c$. The relationships among different feature cues based on sequence can be seen in Fig.1 and Fig.2.



(a) x4 (b) x1 (c) intensity

(d) x2 (e) x3 (f) x5

Fig. 2. Results of multi-histogram based FCM on sequence

4. Bayesian foreground segmentation

To extract the foreground given the current frame, background, feature cues, we wish to compute the maximum a posterior (MAP) estimation of the segmentation field. We take $F = \{f_s\}_{s \in S}$ to represent a set of image data where f_s stands for the gray level at pixel s. s is the site of a pixel and S is the set of all pixels' site. The segmentation problem can be converted to find the labeling w which maximizes $p(w|F)$. Bayes theorem tells us that $p(w|F) = \frac{p(F|w)p(w)}{p(F)}$. Actually $p(F)$ does not depend on the labeling and we have the assumption that conditional independence exists among spatially distinct observations. It is then easy to see that the global labeling which we are trying to find is given by

$$w = max_{w \in \Omega} \prod_{s \in S} p(f_s|w_s) \prod_{s \in S} p(w_s) = max_{w \in \Omega} \prod_{s \in S} p(f_s|w_s) \prod_{z \in Z} exp(-V_k(x, y|x, y \in S)) \tag{8}$$

It is obvious from this expression that a posterior probability also derives from a MRF. The energies of cliques directly reflect the probabilistic modeling of labels without context, which would be used for labeling the pixels independently. Let us assume that the likelihood model $p(f_s|w_s)$ is Gaussian, its parameter can be represented by $\theta_k(x)$ mentioned above. We get:

$$p(f_s|w_s) = \frac{exp\{\frac{-|f_s - \mu_{w_s}(x)|^2}{2\sigma^2_{w_s}(x)}\}}{\sqrt{2\pi}\sigma_{w_s}(x)} \tag{9}$$

The prior model $p(w)$ represents the prior probability of the segmentation field. According to the Hammersley-Clifford theorem, the density is given with the above form in [19 20]. where V_k is the potential function at time k, Z is the set of all cliques z . Only two pixel cliques are used in our work. Spatial connectivity can be imposed by the feature cue in the form of fuzzy clustering factor. The two pixel clique potential can be defined as:

$$V(x, y) = \frac{\beta||x - y||}{2} \tag{10}$$

where $||x - y|| = \sum_{l=1}^{m} |\mu_{xl} - \mu_{yl}| * ||x - y||$ denotes the distance of the fuzzy set, m is the number of classification, and $||x - y||$ is the Eculidian distance. Thus two neighboring pixels are more likely to belong to the same class than to different classes. Also, if the cliques potential lose its fuzziness, our potential will degrade to the classical Pottes model in Markov random fields. β controls the influence of the prior term. In maximum likelihood estimation, it must deal with the partition function, which brings a large computation load in the parameter estimation. We use pseudo-likelihood to substitute the maximum likelihood to estimate the parameter β.

To Potts model, we can give that its maximum pseudo-likelihood estimation is the implicit solution decided by this equation as follows (see Appendix A):

$$t(f) = \sum_{s \in S} E_p[n(s, f_s)] \tag{11}$$

where $E_p[n(s, f_s)]$ is mean of $n(s, f_s)$ to local conditional probability of Markov, $n(s, f_s)$ is the number of pixel which has different intensity value in the third-order neighborhood system. Thus, the prior parameter can be denoted as $\lambda t(f)$, it can be adaptively changed with time. Combining the above models, the Bayesian MAP estimation is obtained by minimizing the objective function, we get:

$$w = min_{w \in \Omega}(\sum_{s \in S}(log(\sqrt{2\pi}\sigma_{w_s}) + \frac{(f_s - \mu_{w_s})^2}{2\sigma^2_{w_s}})) + \lambda \sum_{z \in Z} exp(-V_k(x, y|x, y \in Z)) \tag{12}$$

Obviously, there is no simple method of performing the optimization; furthermore, the objective function does not have a unique minimum since it is no convex. The global inference in the MAP-MRF framework is worked out by using max flow min cut algorithm proposed recently by Boykov et al. in [21].

5. Experiment results and discussion

The proposed approach has been tested in various environments, including both indoor and outdoor scenes. Here, we describe five different sequences that represent typical situations critical for foreground segmentation, and present qualitative results obtained with the proposed method. The walking sequence and armored car sequence were video clips taken by non-professional users with handheld consumer digital cameras. Laboratory sequence can be available from $http://cvrr.ucsd.edu/aton/shadow/index.html$. Our C program can process about one to two frames per second, of size $320 * 240$ on a Pentium4 2.8GHz PC. The parameters of the proposed model are set manually without great difficulty. The local model of GMM consists of five Gaussians, the parameters of initial w and σ^2 can be set in 0.05 and 30. The model parameter β was adaptively estimated by Pseudo-likelihood. It changes with the segmentation model, which controls the influence of the prior term. λ is set to 1 to balance the prior and likelihood.

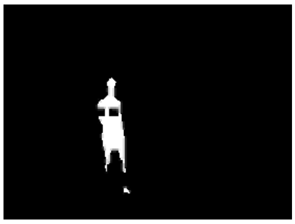
For color image sequences, they are first converted into grayscales ones. The proposed technique is compared to the Gaussian Mixture Model, the Pfinder and Liyuan Li[4] method. The manually segmented "ground truth" foreground images are also shown. Fig.3 shows the segmentation results for the outdoor walking sequence. In this sequence, a man is walking on balcony of a building. It represents an example of easy sequence, in that its background is quite stable, but the lighting conditions will change according to time. Compared with the other methods, we can get more details in the foreground. Fig.3(c) shows the similar result of the foreground with us but in Fig.3(c) the detail of the foot could not be segmented accurately. Li's method can segment the foreground globally. But it exists empty hole in the part of body segmentation.

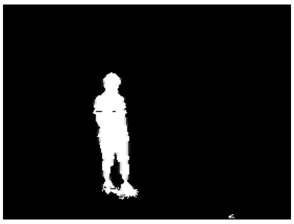
Sequence "laboratory" consists a people walking past a table of the laboratory, it is obviously that this scene is more complicated than the walking man. This is an example of hard sequence because the light condition is much worse than in the previous, and the moving person is disrupted by the table. Fig.4(c) segmented the foreground and the table together, it makes us fail to recognize the people in the image sequence. Fig.4(d) improved this but it also segmented no use background into the foreground. In Fig.4(e) the foreground segmentation result demonstrates better. Our proposed method presents a more accurate result in details. Fig.5 shows another segmentation results from laboratory sequence. When opening and closing the door of the cabinet, foreground segmentation was not performed well in other methods. In Fig.5 (f), fuzzy clustering factors were exploited to enhance the spatial and temporal coherence, which enables us to maintain the continuity of our segmentation.

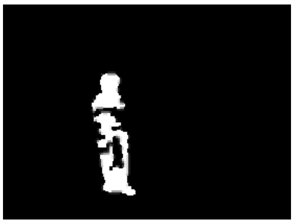
In Fig.6 we used anther outdoor sequence, which is available on the net pro-


(a) walking sequence


(b) ground truth


(c) pfinder


(d) GMM

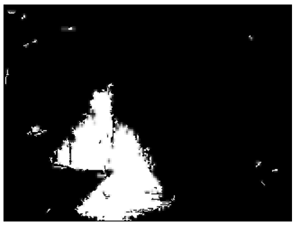

(e) Li

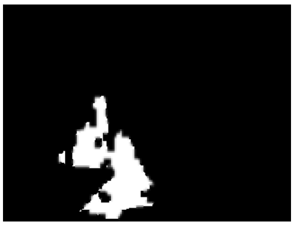

(f) Proposed algorithm

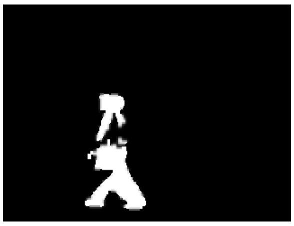
Fig. 3. Results of foreground segmentation on outdoor walking sequence


(a) laboratory sequence


(b) ground truth


(c) pfinder

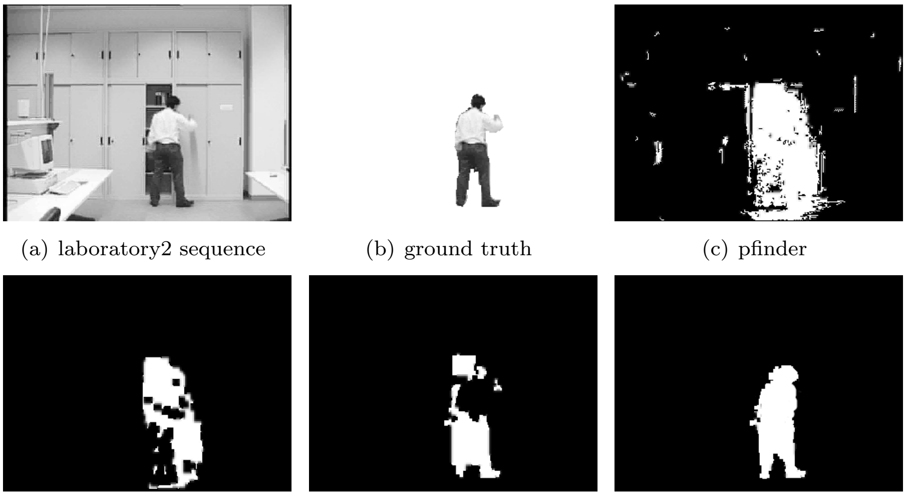

(d) GMM


(e) Li


(f) Proposed algorithm

Fig. 4. Results of foreground segmentation on laboratory sequence

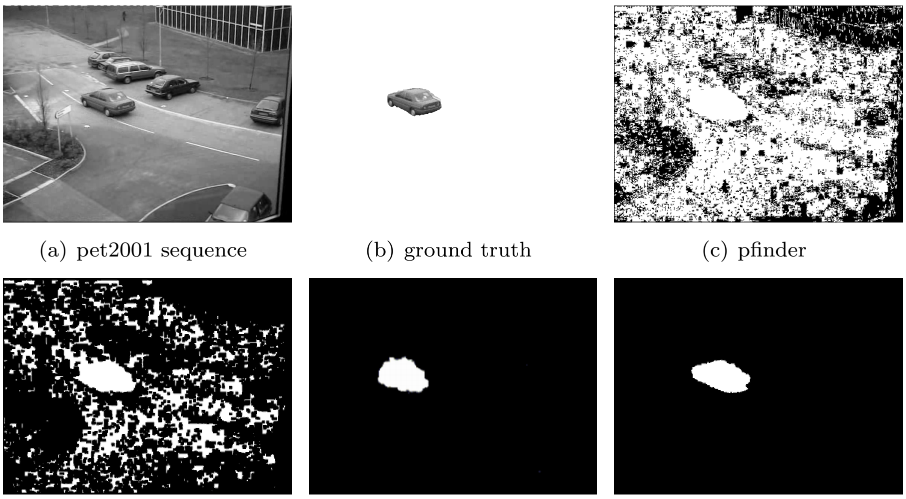
vided by Reading University's pet2001data. The error of the segmentation results appeared large in that a lot of flickering background was segmented as foreground, while this effect is successfully overcome in Fig.6(f). Because the spatial and temporal coherence between pixels which is combined with feature cues will globally decrease the false segmentation. It helps locate structure changes of the scene and



(a) laboratory2 sequence (b) ground truth (c) pfinder

(d) GMM (e) Li (f) Proposed algorithm

Fig. 5. Results of foreground segmentation on laboratory2 sequence



(a) pet2001 sequence (b) ground truth (c) pfinder

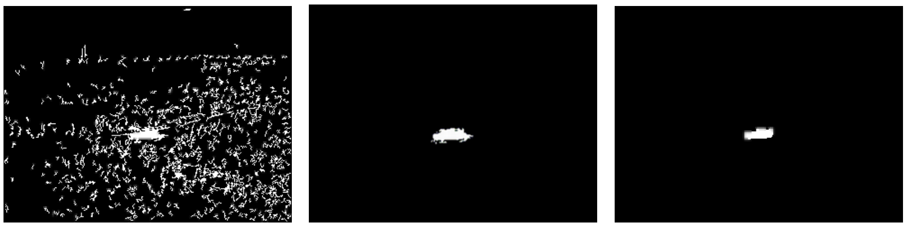
(d) GMM (e) Li (f) Proposed algorithm

Fig. 6. Results of foreground segmentation on pet2001 sequence

improves the reliability of foreground segmentation.

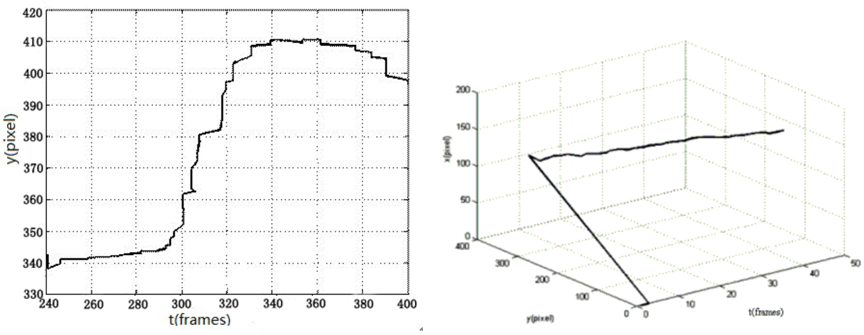
Fig.7(a) shows us the most complicated scene in the outdoor. It contains light changes, moving background, and camouflage and so on. The other two methods are affected by moving background and can not segment the foreground accurately. But our proposed algorithm shows a similar result with Li's method. Besides

(a) armored car sequence (b) ground truth (c) pfinder



(d) GMM (e) Li (f) Proposed algorithm

Fig. 7. Results of foreground segmentation on armored car sequence



(a) 2D tracks in 160 frames (b) 3D tracks in 50 frames

Fig. 8. tracks of moving foreground

that, Fig.8(a) shows 2D moving tracks of segmented foreground in 160 frames and Fig.8(b) shows 3D moving tracks of the foreground in 50 frames, it demonstrates the continuity and consistency in spatial-temporal field of our foreground segmentation algorithm.

For measuring accuracy we adopted two different metrics[22]. Detection rate gives the percentage of detected true positives as compared to the total number of the true positives in the ground truth. Positive Prediction gives the percentage of detected true positives as compared to the total number of items detected by the method.

$$DR = \frac{tp}{tp + fn} \tag{13}$$

$$PP = \frac{tp}{tp + fp} \tag{14}$$

where tp is the total number of true positives, fn is the total number of false negatives, and $tp + fn$ indicates the total number of items present in the ground truth. fp is the total number of false positives, $tp + fp$ indicates the total number of detected items. We use these two metrics to demonstrate the effectiveness of our method.

Table 1. Segmentation accuracy of different methods

Video sequence	Detection rate %	Positive prediction %
Pfinder	55.65	4.18
GMM	57.97	4.31
Li	65.33	87.94
Our method	67.82	95.12

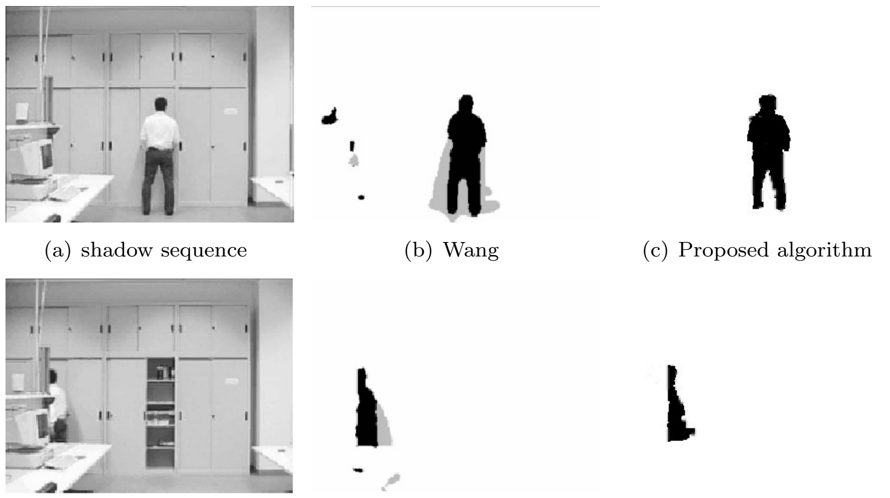
Fig.9 shows the segmentation results in laboratory sequence with camouflage and shadow environment. Wang's method deals with foreground and shadow segmentation in indoor sequence. while our method lies in the construction of segmentation model and feature cues utilization to segment foreground both indoor and outdoor. So we segmented the shadow into the background together. In Fig.9 (d), we can eliminate the local camouflage by spatial and temporal constraints in our proposed framework. When dealing with global camouflage, it also performs not very well as lacking of global information. The effects of shadow and camouflage in the video sequence are the next work of our research, because we can add our shadow model in the MAP-MRF framework and camouflage is decreased by encouraging the information of continuous segmentation regions.

6. Conclusion

In this paper, we proposed a clustering approach for foreground segmentation in image sequences. In our work, fuzzy clustering factor was introduced into the statistical maximum a posterior probability (MAP)-MRF framework. It combines spatial information, temporal information and clustering information in feature space. Experimental results show that our method successfully deals with various dynamic backgrounds. How to further decrease the computation load and deal with shadow and camouflage in our model is the topic of our future study.

Appendix A. Appendices

The maximum pseudo likelihood function is defined as



(a) shadow sequence (b) Wang (c) Proposed algorithm

(d) camouflage sequence (e) Wang (f) Proposed algorithm

Fig. 9. Results of foreground segmentation on shadow and camouflage enviroment

$$PL(f) = \prod_{s \in S} p(f_s | f_{N_s}) = \prod_{s \in S} \frac{exp(-\beta n(s, f_s))}{\sum_{f_s \in L} exp(-\beta n(s, f_s))} \tag{A.1}$$

the Log-pseudo likelihood function is computed as

$$logPL(f) = \sum_{s \in S} log[\frac{exp(-\beta n(s, f_s))}{\sum_{f_s \in L} exp(-\beta n(s, f_s))}] = \sum_{s \in S} [-\beta n(s, f_s) - log(\sum_{f_s \in L} exp(-\beta n(s, f_s)))] \tag{A.2}$$

Also

$$= \frac{\partial}{\partial \beta}(log[p(f_s | f_{N_s})]) \tag{A.3}$$

$$= -n(s, f_s) + \frac{n(s,1)exp[-n(s,1)\beta] + \ldots + n(s,k)exp[-n(s,k)\beta]}{exp[-n(s,1)\beta] + \ldots + n(s,k)exp[-n(s,k)\beta]} \tag{A.4}$$

$$= -n(s, f_s) + n(s,1)p(f_s = 1 | x_{N_s}) + \ldots + n(s,k)p(f_s = k | x_{N_s}) \tag{A.5}$$

$$= -n(s, f_s) + E_p[n(s, f_s)] \tag{A.6}$$

So

$$\frac{\partial}{\partial \beta}(log[PL(f)]) = \sum_{s \in S} -n(s, f_s) + E_p[n(s, f_s)] \tag{A.7}$$

Let

$$\frac{\partial}{\partial \beta}(log[PL(f)]) = 0 \tag{A.8}$$

We can get

$$t(f) = \sum_{s \in S} E_p[n(s, f_s)] \tag{A.9}$$

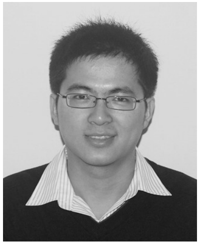
Which is the maximum pseudo likelihood estimation of β.

References

1. C. Wren, A. Azarbayejan, and T. Darrell, Pfinder: Realtime tracking of the human body, *IEEE Trans. Pattern Anal. Mach. Intell.* (1997), p. 780.
2. C. Stauffer and W. Grimson, Adaptive background mixture models for real-time tracking, *in Proc. CVPR* (1999).
3. A. Elgammal, D. Harwood, and L. Davis, Non-parametric model for background subtraction, *in Proc. 6th Eur. Conf. Computer Vision (ECCV)* (2000).
4. L. Li, W. Huang, I. Gu, and Q. Tian, Statistical modeling of complex backgrounds for foreground object detection, *IEEE Trans. Image Processing.*(2004), p. 1459.
5. Y. Sheikh and M. Shah, Bayesian object detection in dynamic scenes, *in Proc. CVPR*(2005).
6. Stenger B, Ramesh V, Paragios N, Coetzee F, Topology free hidden Markov models: application to background modeling, *In Proc. of the 8th ICCV*(2001).
7. Zhou Y, Xu W, Tao H, Gong Y H, Background segmentation using spatial-temporal multi-resolution MRF, *In Proc. Workshop on Motion and Video Computing*(2005).
8. Y. Wang, T. Tan, K.-F. Loe, and J.-K. Wu, A probabilistic approach for foreground and shadow segmentation in monocular image sequences, *Pattern Recognition*(2005), p. 1937.
9. Y. Wang, Dynamic conditional random field for foreground and shadow segmentation, *IEEE Trans. Pattern Anal. Mach. Intell.*(2006), p. 279.
10. Yin Pei,Criminisi A,Winn J, Tree-Based classifiers for bilayer video segmentation, *In Proc. of IEEE CVPR.*(2007).
11. J. Zhou, D. Gao, and D. Zhang, Moving vehicle detection for automatic traffic monitoring, *IEEE. Trans. Veh. Technol.*(2007), p. 51.
12. V. Kolmogorov, A. Criminisi, A. Blake, Bi-layer segmentation of binocular stereo video, *In Proc. of IEEE CVPR.*(2005), p. 407.
13. Hong Yang, Jinwen Tian, Spatiotemporal smooth models for moving object detection, *IEEE Trans. Signal Processing Letters.*(2008), p. 497.
14. S. Geman and D. Geman, Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, *IEEE Trans. Pattern Anal. Mach. Intell.*(1984), p. 721.
15. C. Bezdek, N. R. Pal, J. Keller, and R. Krishnapuram, *Fuzzy models and algorithms for pattern recognition and image processing.*(1999).
16. Mahmoud Ramze, A multiresolution image segmentation technique based on pyramidal segmentation and fuzzy clustering, *IEEE Trans. Image Processing.*(2000), p. 1238.
17. Sotirios P. Chatzis, Theodora A. Varvarigou, A fuzzy clustering approach toward hidden markov random field models for enhanced spatially constrained image segmentation, *IEEE Trans. Fuzzy System.*(2008), p. 1351.
18. Zhenhua Wang, Jie Chen, Lihua Dou, A new multi-threshold segmentation method based on MHFFCM, *2006 IEEE International Conference on Information Acquisition.*(2006), p. 195.
19. Huawu Deng, David A. Clausi, Unsupervised image segmentation using a simple MRF model with a new implementation scheme, *Pattern Recognition.*(2004), p. 2323.
20. S. Z. Li, *Markov Random Field Modeling in Image Analysis.*(1995).
21. Y. Boykov and V. Kolmogorov, An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision, *Pattern Recognition.*(2004), p. 1124.
22. Lucia Maddalena, Alfredo Petrosino, A Self-organizing approach to background subtraction for visual surveillance applications, *IEEE Trans. Image Processing.*(2008),

p. 1168.

Photo and Bibliography



Wenjia Yang received the B.S. degree in automatic control from Beijing Institute of Technology in 2005. He was a visiting student of Laboratory GTI of Universidad Politcnica de Madrid (UPM) in 2010. He is currently a Ph.D. candidate of control science and engineering, Beijing Institute of Technology, P.R. China. His research interests include image processing, computer vision and pattern recognition.

Lihua Dou received the B.S. degree, the M.S. degree and the Ph.D. degree in control theory and control engineering in 1979, 1987 and 2001, respectively, from Beijing Institute of Technology. She is currently a professor of control science and engineering, Beijing Institute of Technology, P.R. China. Her research interests include multi-objective optimization and decision, pattern recognition, and image processing.

Juan Zhang received the B.S. degree and the Ph.D. degree in control theory and control engineering in 1997 and 2002, respectively, from Beijing Institute of Technology. She is currently an associate professor of control science and engineering, Beijing Institute of Technology, P.R. China. Her research interests include operations research, intelligent decision-makings, distributed simulation, and image processing.