Over-Segmentation Based Background Modeling and Foreground Detection with Shadow Removal by Using Hierarchical MRFs

Te-Feng Su Yi-Ling Chen Shang-Hong Lai

Department of Computer Science, National Tsing Hua University, No. 101, Section 2, Kuang-Fu Road, Hsinchu, Taiwan 30013, R.O.C

Abstract. In this paper, we propose a novel over-segmentation based method for the detection of foreground objects from a surveillance video by integrating techniques of background modeling and Markov Random Fields classification. Firstly, we introduce a fast affinity propagation clustering algorithm to produce the over-segmentation of a reference image by taking into account color difference and spatial relationship between pixels. A background model is learned by using Gaussian Mixture Models with color features of the segments to represent the time-varying background scene. Next, each segment is treated as a node in a Markov Random Field and assigned a state of foreground, shadow and background, which is determined by using hierarchical belief propagation. The relationship between neighboring regions is also considered to ensure spatial coherence of segments. Finally, we demonstrate experimental results on several image sequences to show the effectiveness and robustness of the proposed method.

1 Introduction

Extracting foreground objects from image sequences is a critical task for many computer vision applications, such as video processing, visual surveillance and object recognition. Background subtraction is a core component for video surveillance, whose objective is to discriminate the foreground from the background scene. To achieve this goal, a robust background modeling technique is essential. The basic idea of background modeling is to maintain an estimation of the background image model which represents the scene with no foreground objects. Then, moving objects can be detected by a simple subtraction and thresholding procedure. Hence, the more accurate the background model, the more accurate is the detection of the foreground objects.

Most traditional background modeling techniques are pixel-based, and they usually estimate the probability of the individual pixels belonging to background by using GMMs [1] or to label each pixel as foreground or background by MRFs [2]. However, pixel-based models are less efficient and effective in handling illumination change and dynamic scene such as swaying vegetation, waving trees, fluttering flags, and so on. Even though the background is static, camera jittering and signal noise may still cause non-stationary problems. Several block-based methods were developed to overcome such problems, which partition a background image into sub-blocks to utilize block correlation [3] or to compute block-specific features, such as *local binary pattern* histograms [4]. However, the fix-sized blocks often fail to correctly classify the foreground objects because they are not well fitted to object boundaries, which also results in inaccurate shape of the detected foreground objects.

Motivated by the above issues, we propose a novel over-segmentation based approach for foreground objects detection in this paper. Unlike pixel-based or block-based methods, the proposed method exploits the observation that neighboring pixels are very likely to have the same foreground or background classification if they are appropriately grouped together according to certain similarity measure. Despite the simplicity of dividing an image into blocks, the subdivided regions usually do not fit to the object boundaries well. We thus propose a fast and effective affinity propagation algorithm to obtain the over-segmentation of a background image to facilitate the task of background modeling. By considering color and spatial coherence of neighboring pixels, the proposed method is capable of handling illumination change in a scene effectively. In the following foreground/background classification stage, we produce the over-segmentation of various resolutions on the reference image to form a hierarchy of MRFs and each segment is treated as a node. Hierarchical belief propagation [5] is then utilized to label the MRFs.

Fig. 1 illustrates the flowchart of the proposed background modeling and foreground detection system. In the training phase, an reference background image is selected and its over-segmentation is produced by performing fast AP clustering algorithm. Based on the over-segmented image, the background models are learned via GMMs and a hierarchy of MRFs is constructed. Foreground object detection is accomplished through MRF classification by hierarchical belief propagation when new images are acquired from a test sequence and the corresponding background models are updated accordingly.

The rest of this paper is organized as follows. In Section 2, we first briefly review the related work. Then, the over-segmentation based background modeling method is introduced in Section 3. In Section 4, the detection of foreground objects by MRFs classification is described. In Section 5, we show some experimental results and quantitative comparisons to demonstrate the superior performance of the proposed method over previous methods. Section 6 concludes this paper.

2 Related Work

Many approaches for background subtraction have been proposed over the past decades, but they usually differ in the ways of modeling the background. Most of them can be classified as pixel-based approaches. A well-known method by Grimson and Stauffer [1] proposed to use GMMs for background modeling. It

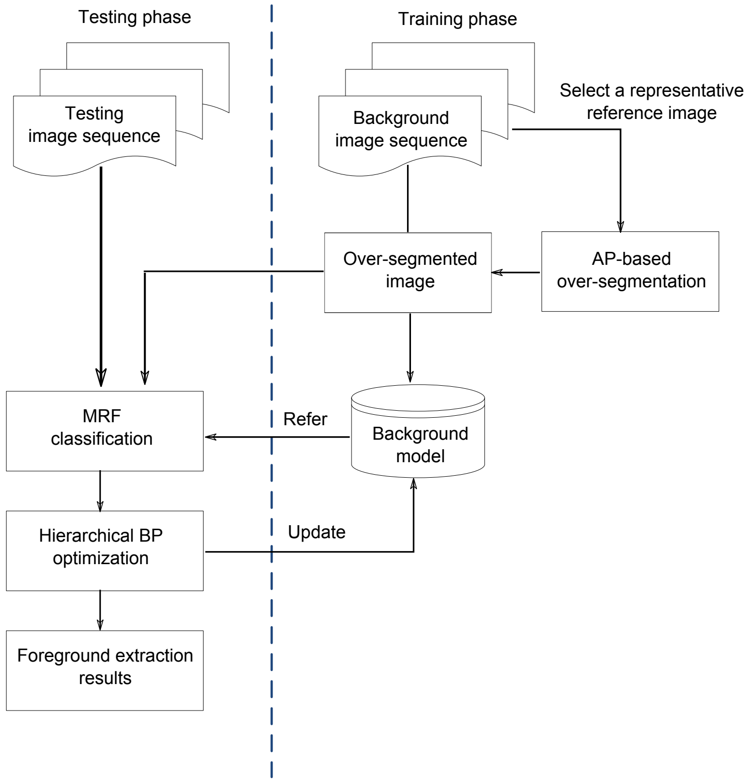


Fig. 1. System flowchart.

describes each pixel as a mixture of Gaussians and updates the models adaptively according to the input image sequence. Zivkovic proposed an improved GMM learning algorithm that estimates the parameters of the GMM and simultaneously selects the number of Gaussians [6]. Elgammal et al. introduced nonparametric estimation method for per-pixel background modeling [7]. They utilized a general nonparametric kernel density estimation technique for building a statistical representation of the background scene.

Spatial and temporal neighboring relationships of pixels are useful information for object segmentation. In [2], Paragios and Ramesh proposed a MRF-based method to deal with change detection for subway monitoring. Migdal and Grimson adopted MRFs to model spatial and temporal relationship of neighborhood pixels [8]. Wang et al. [9] introduced dynamic CRFs for foreground object and moving shadow segmentation in indoor video scene. An approximate filtering algorithm is exploited to update parameters of CRF models according to previous frames. Huang et al. [10] proposed a region-based motion segmentation algorithm to obtain motion-coherence regions. Both spatial and temporal coherence of regions are taken into account to maintain the continuity of segmentation by using MRFs.

Recently, several block-based methods were developed for background modeling and subtraction to more effectively deal with illumination change and dynamic scenes. Generally, block-based algorithms start by dividing a background image into blocks and construct the background models by calculating block-specific features extracted from these blocks. In [3], the correlation between blocks is measured by the *normalized vector distance* to realize robust background subtraction against varying illumination. Heikkila and Pietikinen [4] proposed to model the background scene based on Local Binary Pattern (LBP) histogram and produce coarse detection of foreground object. However, the LBP histogram cannot capture temporal variation in the pattern. Following [4], Chen et al. proposed a contrast histogram measure to describe each block and performed object detection by combining a pixel-level GMMs and block-wise contrast descriptors [11].

Cast shadows are difficult to be correctly detected by most background subtraction methods. It is often misclassified as the foreground region, resulting in inaccurate object shapes and the degradation of model updating. Shadow detection techniques can be classified into two groups: model-based and property-based techniques. Model-based techniques rely on models representing the prior knowledge of the geometry of the scene or objects, and the illumination [12]. Property-based techniques identify shadows by using features, such as brightness [13, 14], geometry [15] or texture [16].

3 Over-segmentation Based Background Modeling

To enable the background model to more effectively handle changes occurred in the scene, it is preferable to divide a background scene into sub-regions and learn the background models accordingly. To this end, we propose a simple and efficient affinity propagation clustering algorithm to over-segment a reference image I_R among an input sequence before the background model learning process. Furthermore, a hierarchy of over-segmentation built over I_R is constructed to facilitate the following foreground object detection by MRF classification (Section 4). In this work, we adopt GMMs [1] as underlying the background models, which are trained for both pixel and segmentation level.

Affinity propagation (AP) [17] is an iterative algorithm that groups data points into clusters by sending *messages* between data points. The pairwise *similarity* $s(i, k)$, which measures how well-suited data point k is to be the *exemplar* (i.e. cluster center) for data point i, is taken as input and AP aims to search for a number of clusters such that the net similarity is maximized. Unlike other clustering techniques, such as the k-means clustering that needs the number of clusters to be explicitly specified, AP takes for each data point k a *preference* value $s(k, k)$ as input that indicates a candidate exemplar's potential of being chosen as an exemplar. Exemplars emerge during the process of message passing and the number of identified exemplars depends on the input preference values.

There are two kinds of messages to be updated during each iteration, i.e. *responsibility* and *availability*, and each accounts for a different kind of competition. Briefly speaking, responsibility update lets all candidate exemplars compete for ownership of a data point while availability update collects evidence from data points reflecting the competence of each candidate exemplar. The message updating procedures are summarized in Algorithm 1.

Algorithm 1 Affinity propagation

Initialization:

$r(i,k) = 0,\ a(k,i) = 0$ for all $i,\ k$

Responsibility updating:

$r(i,k) \leftarrow s(i,k) - \max_{j:j \neq k}\{a(j,i) + s(i,j)\}$

Availability updating:

$a(k,k) \leftarrow \sum_{j:j \neq k} \max\{0,\ r(j,k)\}$

$a(k,i) \leftarrow \min\{0,\ r(k,k) + \sum_{j:j \notin \{k,i\}} \max\{0,\ r(j,k)\}\}$

Exemplar assignments:

$c_i^* \leftarrow \arg\max_k\ r(i,k) + a(k,i)$

The messages are directional. The responsibility $r(i,k)$, sent from data point i to candidate exemplar k, delivers the accumulated evidence for how well-suited it is to assign point i to point k, by considering other potential exemplars' competition for point i. The availability $a(k,i)$, sent from candidate exemplar k to point i, delivers the accumulated evidence for how appropriate it would be for point i to choose point k as its exemplar, by considering the aggregate support of choosing point k to be an exemplar from other points. After convergence, availabilities and responsibilities can be combined to identify exemplars. For point i, it is assigned to the exemplar c_i that maximizes $r(i,k) + a(k,i)$.

In the application of image over-segmentation, neighboring pixels of similar color are grouped together and each pixel competes for exemplarship by exchanging messages with each other. One drawback of AP clustering is its high complexity of message updating, which is $O(N^2)$ if each pixel sends messages to all the other pixels. To achieve efficiency, we exploit the assumption that distant pixels are not possible to be assigned to the same exemplar and thus message exchange between pixels far away is not necessary. To further reduce the amount of messages, we take advantage of a set of *virtual* exemplars, which are responsible for competing for pixels, to form the over-segmentation. We do not define pixel-to-pixel similarity since the image pixels are no longer candidates of exemplars to form the final segmentation. Therefore, the amount of messages to be updated is greatly reduced, leading to an efficient algorithm.

Following the convention of AP, we define the following negative real-valued similarity measure between a pixel p and its nearby virtual exemplar v, taking into account color difference and spatial relationship

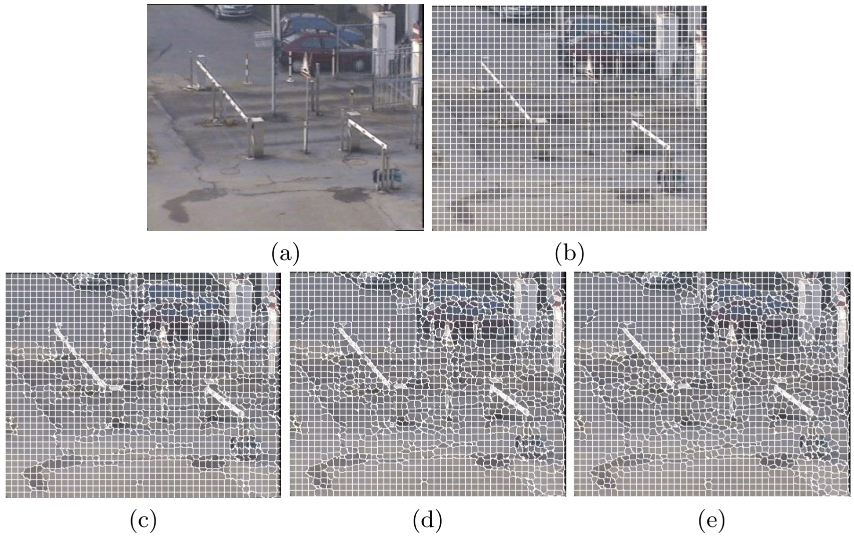


Fig. 2. Over-segmenting a reference image from 8×8 regular grid by our fast AP clustering algorithm. (a) original image (b) 8×8 regular grid image (c) AP based clustering algorithm with 1 iteration (d) AP-based clustering algorithm with 3 iterations (e)final over-segmentation image from AP-based clustering algorithm

$$s(p, v) = -(\lambda_c \|\mathbf{c}_p - \mathbf{c}_v\|^2 + \lambda_s \|\mathbf{u}_p - \mathbf{u}_v\|^2), \tag{1}$$

where λ_c and λ_s are weighting coefficients to balance the two terms. Initially, we obtain an initial over-segmentation by partitioning the input image into a regular grid consisting of fix-sized blocks, e.g. 8×8, and associate each block with a virtual exemplar. The mean color $\mathbf{c}_v$ and position $\mathbf{u}_v$ of each block are then computed for the associated v for message computation. Each pixel only sends messages to the virtual exemplars with the corresponding segments connected to each other. The message updating procedure of AP remains unchanged. After each iteration of AP clustering, the pixels assigned to each segment vary and the exemplar attributes $\mathbf{u}_v$ and $\mathbf{c}_v$ are updated accordingly. Generally, 5$\sim$10 iterations suffice to obtain good segmentation results in our experiments.

4 Markov Random Fields for Classification

4.1 Energy Minimization

In this paper, we formulate the foreground/background classification problem as labeling a MRF with each node corresponding to a pixel or segment in an image. Let $\mathcal{S} = \{s_1, s_2, ..., s_n\}$ be the set of nodes in a graph $\mathcal{G}$ and $\mathcal{E}$ be the set of edges with $(s_i, s_j) \in \mathcal{E}$ indicating that there is an edge connecting s_i and

s_j. We aim to find an optimal *configuration* $\widehat{\Omega}$ of $\mathcal{G}$ that assigns a label $l_i \in \{foreground, background, shadow\}$ for node s_i such that the following energy function is minimized:

$$E(\Omega) = \sum_{s_i \in \mathcal{S}} V_{likelihood}(l_i) + \alpha \sum_{(s_i,s_j) \in \mathcal{E}} V_{prior}(l_i, l_j), \tag{2}$$

where α is a weighting coefficient. The energy function $E(\Omega)$ is the sum of two terms: *likelihood energy* $V_{likelihood}$ and *prior energy* V_{prior}. The likelihood term $V_{likelihood}$ measures the likelihood that a node s_i is classified as one of its three possible states and is composed of the weighted combination of two terms: *color distortion* V^C and *gain* information V^G:

$$V_{likelihood}(l_i) = \lambda \sum_{s_i \in \mathcal{S}} V^C(l_i) + (1-\lambda) \sum_{s_i \in \mathcal{S}} V^G(l_i). \tag{3}$$

Color distortion V^C is the angle between the color vectors associated with a node s_i in the current observed image and the corresponding background model. Note that if s_i corresponds to a segment, the average color vector is used to compute V^C to measure the similarity with its corresponding background model. The *gain* information V^G was designed to handle cast shadow based on the observation that shadow regions are expected to possess lower luminance but similar chromaticity values [14]. It is calculated by the ratio between the brightness change over the corresponding background model,

$$gain = \frac{I_o - I_b}{I_b} \tag{4}$$

where I_b and I_o are the average intensity of background model and the observed region, respectively. The variation of intensity in the shadow regions due to illumination changes should be relatively small.

Let R_i and R_i^b be the colors corresponding to the ith pixels and mean color for segment case in an observed image F_o and background image F_b, respectively. V^C and V^G terms are defined as follows:

$$V^C_{(l_i)} = \begin{cases} 1 - exp(-f_{cd}(R_i, R_i^b)) \ if \ l_i = background \\ exp(-f_{cd}(R_i, R_i^b)) \ otherwise \end{cases} \tag{5}$$

$$V^G_{(l_i)} = \begin{cases} 1 - exp(-f_{gain}(R_i, R_i^b)) \ if \ l_i = background, shadow \\ exp(-f_{gain}(R_i, R_i^b)) \ otherwise \end{cases} \tag{6}$$

where f_{cd} and f_{gain} are functions to calculate color distortion and gain information between R_i and R_i^b, respectively. The definition for f_{cd} and f_{gain} can be formed by:

$$f_{cd}(m, n) = \arccos(\frac{\overrightarrow{m} \cdot \overrightarrow{n}}{|\overrightarrow{m}| \ |\overrightarrow{n}|}) \tag{7}$$

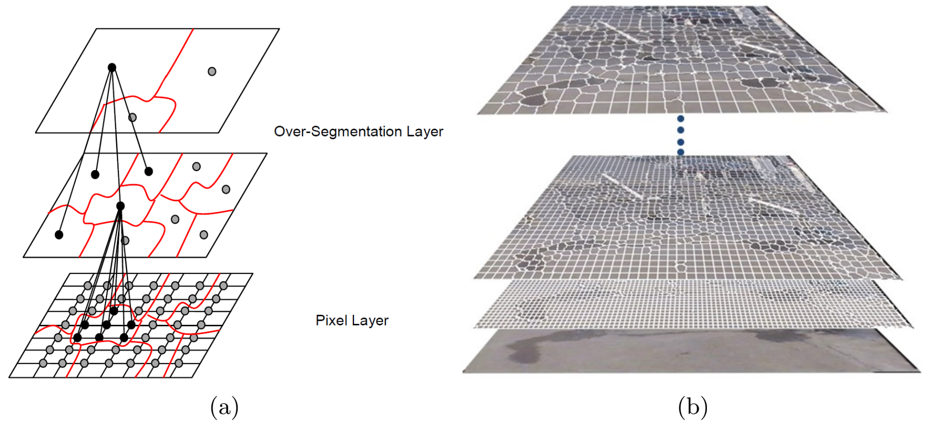


Fig. 3. (a) Hierarchical MRFs built over pixel and segmentation levels. (b) Example of hierarchical over-segmentation.

$$f_{gain}(I_o, I_b) = \frac{I_o - I_b}{I_b} \tag{8}$$

where m and n are two input of color vectors.

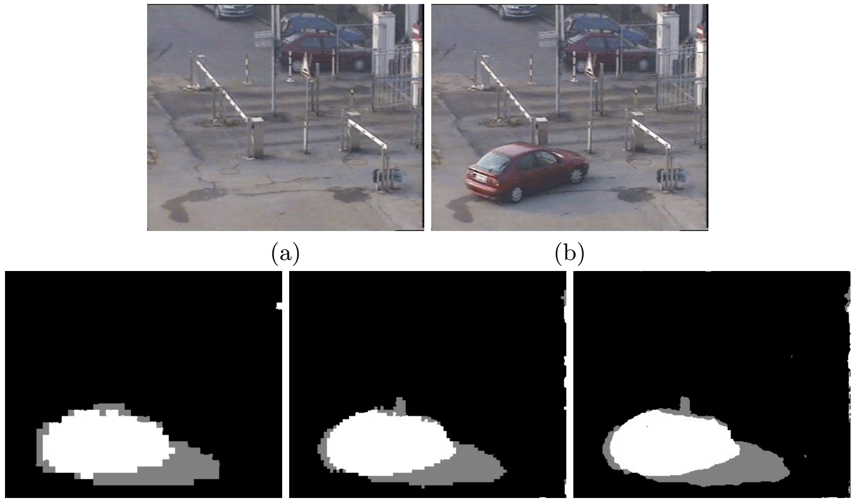
The prior energy V_{prior} captures the spatial continuity between neighboring pixels or segments. It introduces more penalty if two neighboring regions with small color distortion are assigned different labels. Let R_i and R_j be the pixels or segments corresponding two connected nodes s_i and s_j in $\mathcal{G}$. We thus define V_{prior} as follows,

$$V_{prior} = \begin{cases} 1 - \exp(-f_{cd}(R_i, R_j)), & if \;\; l_i \neq l_j \\ 0, & if \;\; l_i = l_j. \end{cases} \tag{9}$$

4.2 Hierarchical belief propagation optimization over MRFs

After the MRF is built, belief propagation (BP) algorithm is employed to find the optimal label assignment of each node by minimizing the energy function defined in Equation (2). In this work, we build a hierarchy of MRFs on both pixel and segmentation levels. By using the proposed fast AP clustering algorithm, we over-segment the reference background image with various initial block sizes. As shown in Fig. 3, a segment in the over-segmented image is viewed as a node in the MRF model. A segment in a coarser level is obtained by merging some neighboring segments in the next finer level. Therefore, we can easily build a hierarchical MRF structure from finest level to coarsest level as shown in Fig. 3.

We exploit the hierarchical BP algorithm [5] to solve the optimization problem defined in the last subsection. The messages at the coarsest level are initialized as zero and the messages after convergence at each level are passed to



(a) (b) (c) (d) (e)

Fig. 4. A coarse to fine foreground and shadow detection results by our proposed method from "campus" image sequence. (a) original image (b) frame 65 of "campus" image sequence. (c)$\sim$(e) show our foreground and shadow detection results from coarse to fine

the successive finer level as the initial guesses for BP message updating. Fig. 4 shows an example of coarse-to-fine foreground and shadow detection by hierarchical MRF classification.

5 Experimental Results

To evaluate the performance of the proposed over-segmentation based background subtraction, three image sequences from public domain are adopted as benchmark.

The three sequences are "*campus*", "*intelligent room*" and "*hall monitoring*", which are taken from various types of scenes, such as outdoor and indoor environments, to demonstrate the robustness of the proposed method. Fig. 6 compares the results by the proposed method to those by [6] and [14]. Obviously, the foreground objects, such as the moving car and pedestrians in the *campus* sequence, detected by our method is more accurate than previous methods. In *hall monitoring* sequence, the proposed method is more robust against the noise due to light fluctuation. Cast shadows caused by the foreground objects are also detected well in all test sequences.

To provide quantitative evaluation, we use the similarity measure presented by Li et al. [18] in this paper. Let A be a detected region and B be the corresponding ground truth. The similarity measure between A and B is defined

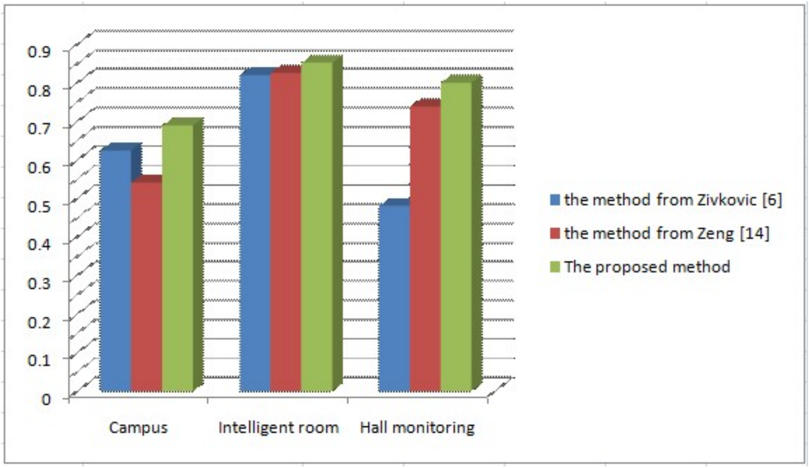


Fig. 5. Quantitative comparison between different methods in the three static scene sequences.

as

$$S(A,B) = \frac{A \bigcap B}{A \bigcup B} \tag{10}$$

$S(A,B)$ reaches its maximal value of 1 if A and B is exactly the same. Otherwise, $S(A,B)$ fluctuates between 0 to 1 depending on their overlapped regions. The ground truth data are obtained from public domain and the residuals are marked manually. The quantitative comparisons shown in Fig. 5 indicates the superior performance of the proposed method over the previous methods.

6 Conclusion

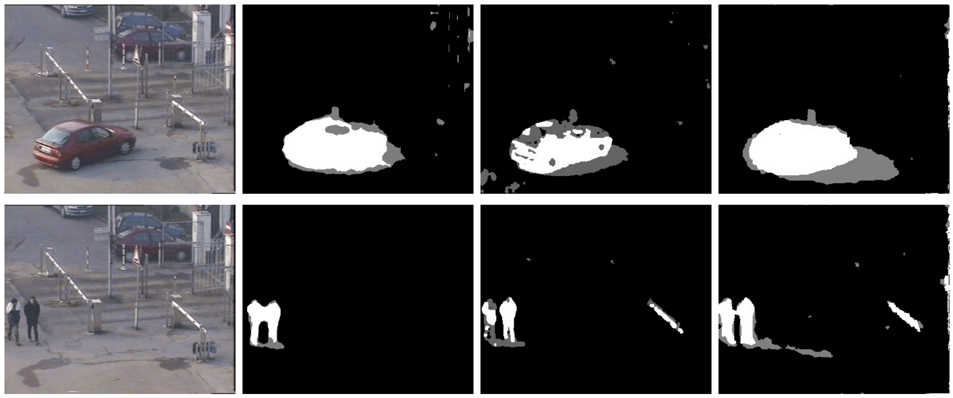
In this paper, a new over-segmentation based background modeling algorithm is presented for foreground and shadow segmentation. The proposed method uses a fast AP clustering algorithm to obtain image over-segmentation of various resolutions. The foreground/background classification is then formulated as an energy minimization problem over the MRFs constructed on the segmented image by using hierarchical belief propagation. Experimental results on several test sequences and quantitative analysis show that the proposed method performs well for foreground object extraction and cast shadow detection.

Acknowledgement. This work was partially supported by National Science Council in Taiwan under the project 98-2220-E-007-029

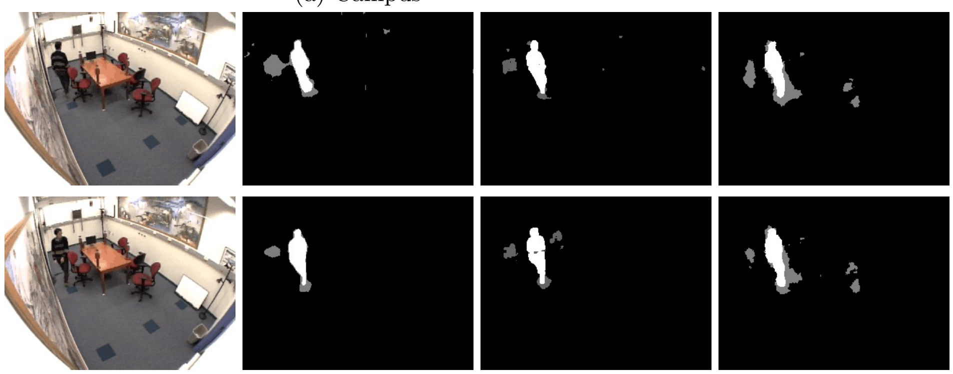
References

1. Stauffer, C., Grimson, W.: Adaptive background mixture models for real-time tracking. In: IEEE Conference on CVPR. Volume 2. (1999) 252

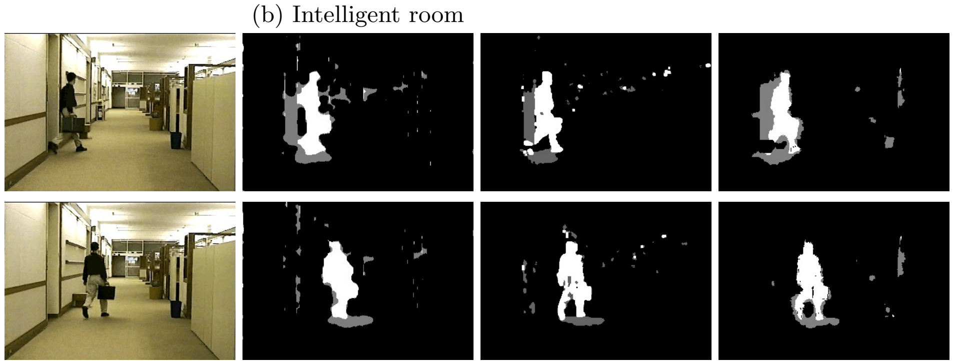
2. Paragios, N., Ramesh, V.: A mrf-based approach for real-time subway monitoring. In: IEEE Conference on CVPR. Volume 1. (2001) I–1034 – I–1040
3. Matsuyama, T., Ohya, T., , Habe, H.: Background subtraction for nonstationary scenes. In: Proceedings of ACCV. (2000) 662–667
4. Heikkila, M., Pietikainen, M.: A texture-based method for modeling the background and detecting moving objects. IEEE Transactions on PAMI **28** (2006) 657 –662
5. Felzenszwalb, P., Huttenlocher, D.: Efficient belief propagation for early vision. International Journal of Computer Vision **70** (2006) 41–54
6. Zivkovic, Z., van der Heijden, F.: Recursive unsupervised learning of finite mixture models. IEEE Transactions on PAMI **26** (2004) 651 –656
7. Elgammal, A.M., Harwood, D., Davis, L.S.: Non-parametric model for background subtraction. In: proceedings of ECCV. (2000) 751–767
8. Migdal, J., Grimson, W.E.L.: Background subtraction using markov thresholds. In: IEEE Workshop on Motion and Video Computing. Volume 2. (2005) 58 –65
9. Wang, Y., Loe, K.F., Wu, J.K.: A dynamic conditional random field model for foreground and shadow segmentation. IEEE Transactions on PAMI **28** (2006) 279 –289
10. Huang, S.S., Fu, L.C., Hsiao, P.Y.: Region-level motion-based background modeling and subtraction using mrfs. IEEE Transactions on IP **16** (2007) 1446 –1456
11. Chen, Y.T., Chen, C.S., Huang, C.R., Hung, Y.P.: Efficient hierarchical method for background subtraction. Pattern Recognition **40** (2007) 2706–2715
12. Martel-Brisson, N., Zaccarin, A.: Learning and removing cast shadows through a multidistribution approach. IEEE Transactions on PAMI **29** (2007) 1133 –1146
13. Cucchiara, R., Grana, C., Piccardi, M., Prati, A.: Detecting moving objects, ghosts, and shadows in video streams. IEEE Transactions on PAMI **25** (2003) 1337 – 1342
14. Zeng, H.C., Lai, S.H.: Adaptive foreground object extraction for real-time video surveillance with lighting variations. In: ICASSP. (2007) I–1201 –I–1204
15. Salvador, E., Cavallaro, A., Ebrahimi, T.: Cast shadow segmentation using invariant color features. Computer Visual Image Understand **95** (2004) 238–259
16. Zhang, W., Fang, X.Z., Yang, X., Wu, Q.: Moving cast shadows detection using ratio edge. IEEE Transactions on Multimedia **9** (2007) 1202 –1214
17. Frey, B.J., Dueck, D.: Clustering by passing messages between data points. Science **315** (2007) 972–976
18. Li, L., Huang, W., Gu, I.Y.H., Tian, Q.: Statistical modeling of complex backgrounds for foreground object detection. IEEE Transactions on IP **13** (2004) 1459 –1472



(a) Campus



(b) Intelligent room



(c) Hall monitoring

Fig. 6. Some background subtraction results of different methods. The white and gray pixels indicate the detected foreground and shadow regions. First column: original image. Second column: the results by Zivkovic [6]. Third column: the results by Zeng [14]. The right-most column: the results by the proposed method.