

A NOVEL OBJECT OF INTEREST EXTRACTION METHOD FOR MOBILE MAPPING SYSTEMS

Peng Li ^{a,b}, Cheng Wang ^{b,*}, Yunhui Yi ^b, Hanyun Wang ^{a,b}, Ming Cheng ^b, Lun Luo ^c, Yinsheng Liu ^d

^a School of Electronic Engineering and Science, National University of Defense Technology, 410073, Changsha, China
- peng.li.scholar@gmail.com

^b School of Information Science and Technology, Xiamen University, 361005, Xiamen, China - cwang@xmu.edu.cn

^c China Transport Telecommunication & information Center, Beijing, China

^d Hunan Provincial Transport Technical Information Center, Changsha, China

Commission III, Working Group III/5

KEY WORDS: Mobile , Mapping, GIS, object, Extraction, Tracking, Segmentation

ABSTRACT:

In this paper, we present a novel object of interest (OOI) extraction scheme that can work robustly for digital measurable image (DMI) sequences collected by mobile mapping systems (MMS). The proposed method integrates tracking and segmentation in a unified framework. We incorporate a new object-shaped kernel with the scale invariant mean shift algorithm to track the OOI through the DMI sequence and thus keep the temporal consistency. The well-known GrabCut approach for static 2D image segmentation is generalized to the DMI sequence for OOI segmentation. Experimental results on real DMI sequence collected by VISATTM MMS demonstrate that the proposed approach is robust to the challenges such as low frame rate, large inter-frame displacement of the OOI and background clutter.

1. INTRODUCTION

Mobile mapping system (MMS) has been recognized as the optimal tool for making digital navigation map, map amendment and GIS data collection. Digital measurable image (DMI) sequences, gathered by MMS, covering both the geo-information and street view images, have been widely applied to provide location-based service (Petrie, 2010; Wang, C. *et al.*, 2008a). Extraction and update of GIS features, such as fire horns, guideboards, traffic lights, is the key post-processing for MMS. Up to now, it is highly dependent on human annotation and hence the bottleneck of MMS-based applications. Due to the complex background and the large number of GIS feature categories, computer-aided strategy is a promising tactical solution to improve the productivity. Therefore, many researchers and practitioners try to extract GIS object of interest (OOI) in DMI sequences with as less human labor as possible (Hinz, 2008). Here the OOI refers to the selected GIS feature. DMI sequence can be regarded as a subset of video sequence but with much lower frame rates, known 0.5-2 fps, and much larger size, typically larger than 1600×1200, and collected by moving surveying vehicle in natural scenes with no constraints.

Although noticeable success has been achieved in the video object extraction, the state-of-art video OOI extraction can not directly applied to the task of OOI extraction from DMI sequences because of the new challenges brought by the DMI sequences. First, the low frame rate of the DMI sequences often results in large scaling of the OOI through the sequence and this requests more robust estimation of the scale factor, which are often assumed to be relatively small in successive frames (Bae *et al.*, 2009; Wang, C.H. *et al.*, 2008b). Second, low frame rate of DMI sequence often cause large inter-frame displacement of the OOI and this violates the tight assumption of the spatiotemporal continuity of the OOI in previous work (Criminisi *et al.*, 2006; Wang, J. *et al.*, 2005). Third, the imaging platform, known as the surveying vehicle are moving and the static background assumption (Li *et al.*, 2005; Wu *et al.*, 2009) is no longer available. This challenge is compound with

the complexity of the background. These challenges are the aim of our work in this paper.

In this paper, we propose a novel GIS OOI extraction approach that integrates tracking and segmentation in a unified framework. The diagram of the proposed method is how in Figure 1. The manual interaction is limited to a bounding box around the OOI in a keyframe. Two major contributions are included in the method, first, a scale invariant mean shift with an object-shaped kernel is proposed for OOI tracking; second, a generalized GrabCut approach is proposed for object segmentation through DMI sequence. The experimental results demonstrate the effectiveness and robustness of the proposed method.

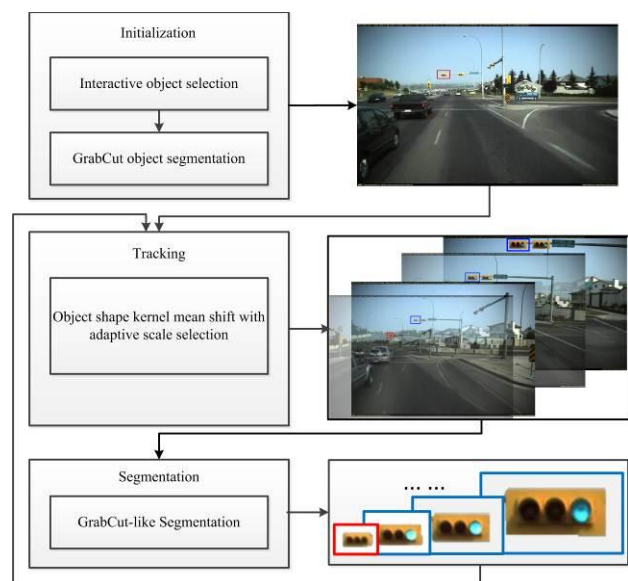


Figure 1. Diagram of the proposed scheme.

The rest of the paper is organized as follows: Section 2 presents a brief review on the previous work; Section 3 proposes a object-shaped kernel mean shift tracking scheme. The proposed generalized GrabCut video object segmentation method will be presented in Section 4. Experimental results are provided in Section 5. Section 6 concludes the paper.

2. RELATED WORK

We firstly review previous graph cuts-based video OOI extraction approaches (Boykov, Y. *et al.*, 2006; Boykov, Y.Y. *et al.*, 2001) that are directly related to our work. Then, we revisit the mean shift-based tracking algorithms.

2.1 Graph Cuts-based OOI extraction

Video OOI extraction methods can be categorized into 3D volume-based and 2D frame-by-frame based groups.

The 3D spatiotemporal volume (STV)-based approach unifies the analysis of spatial and temporal information by constructing a volume of spatiotemporal data in which consecutive images are stacked to form a third temporal dimension (Boykov, Y. *et al.*, 2006). Wang *et al.* (Wang, J. *et al.*, 2005) present an interactive video cutout system, which provides a novel user interface letting user painting clue strokes of the foreground and background, to cut out dynamic foreground objects from a video sequence. The system proposed by Li *et al.* (Li *et al.*, 2005) first oversegments each frame into atomic regions using watershed transform and then performs a 3D-graph cuts to bi-label each atomic region as foreground or background. Such methods is quite limited in real-world applications for requirement of quite long preprocessing time for oversegmentation and labor-intensive user interface.

In 2D single frame-based methods, objects are segmented in image/video sequences frame by frame, making these methods more feasible in practical applications. In general, pixels in a frame are labelled foreground or background according to color, intensity, texture, temporal consistency, motion or a mixture of them (Bae *et al.*, 2009; Malcolm *et al.*, 2007). State-of-art methods integrate such information to the energy term of the Graph Cuts and attempt to seek the global optimal segmentation of the object. Then the segmentation results, often in the form of labeling trimap, would be propagated to the following frames as the prior of the object. Once some pixels are segmented and labeled incorrectly, the error will be propagated through the sequence and there is no chance to correct the errors.

In order to keep temporal consistency, tracking is often employed (Comaniciu *et al.*, 2003) and to be discussed below.

2.2 Mean shift object tracking

The mean shift algorithm is a nonparametric method based on kernel density estimation (KDE) for mode seeking (Comaniciu *et al.*, 2002). It has been widely applied for visual object tracking in video and image sequence successfully due to its efficiency and simplicity (Mena, 2003; Petrie, 2010; Yi *et al.*, 2008). However, the original mean shift algorithm is also well known for its limitations of background clutter in radically symmetric kernel, incapability to handle scale change and large displacement of objects. To overcome the first limitation, Yilmaz propose an asymmetric level set kernel to exclude background pixels (Yilmaz, 2007). To overcome the second limitation, Collins (Collins, 2003) proposed to compute the bandwidth of the kernel in the scale space. A better alternation

is to simultaneously estimates the position and scale (Yilmaz, 2007).

3. OBJECT-SHAPED KERNEL SCALE INVARIANT MEAN SHIFT

In mean shift tracking (Comaniciu *et al.*, 2003), the appearance of the object is represented by its probability density function in a given feature space, such as weighted color histogram $\hat{\mathbf{h}}$ because of its weak independence of scaling and rotation, robustness to partial occlusion and low-computational cost. Having the weighted color histogram of $\hat{\mathbf{h}}_m$ and $\hat{\mathbf{h}}_c$ generated from the object model and candidate region within kernel K in image I , each pixel \mathbf{x}_i in the candidate region is given a weight:

$$w(\mathbf{x}_i) = \sqrt{\hat{\mathbf{h}}_m(\mathbf{x}_i) / \hat{\mathbf{h}}_c(\mathbf{x}_i)}. \quad (1)$$

Let the initial hypothesized position of the object candidate region be $\hat{\mathbf{y}}_{old}$, the new position be $\hat{\mathbf{y}}_{new}$, the pixels inside the candidate region be $\{\mathbf{x}_i\}$. $k(\bullet)$ is the profile of K , i.e.

$K(\mathbf{x}) = k(\|\mathbf{x}\|^2)$, and $k'(\bullet)$ is the derivative of the kernel $k(\bullet)$.

Then using the weights in Eq. (1), the mean shift vector $\Delta\mathbf{y} = \hat{\mathbf{y}}_{new} - \hat{\mathbf{y}}_{old}$ is calculated as

$$\Delta\mathbf{y} = \frac{\sum_i k'(\|\mathbf{x}_i - \hat{\mathbf{y}}_{old}\|^2) w(\mathbf{x}_i) (\mathbf{x}_i - \hat{\mathbf{y}}_{old})}{\sum_i k'(\|\mathbf{x}_i - \hat{\mathbf{y}}_{old}\|^2) w(\mathbf{x}_i)}. \quad (2)$$

3.1 The Object-Shaped Kernel:

In traditional mean shift object tracking (Comaniciu *et al.*, 2003), the kernel is commonly chosen as a primitive geometric shape and can't describe the object's shape accurately. The histogram calculated inside the kernel is not accurate for considering the non-object regions residing inside the kernel as a part of the object.

To overcome this problem, we propose an object-shaped kernel generated by accurate object shape mask. The object shape mask is given by the segmentation results. The centroid of the object is positioned at, any point $\mathbf{x}_i = (x_i, y_i)$ inside the object region has a distance of $\delta(\alpha_i)$ from the centroid, where α_i is the angle the point makes with the x -axis. In a similar way, the maximal distance from \mathbf{x}_c in the direction α_i is defined as the radius in α_i : $r(\alpha_i) = \|\mathbf{x}_b(\alpha_i) - \mathbf{x}_c\|_2$, where $\mathbf{x}_b(\alpha_i)$ is the point on the boundary in direction α_i . Then the normalized distance $d_{norm}(x_i)$ in $[0,1]$ from any \mathbf{x}_i to \mathbf{x}_c is defined as

$$d_{norm}(x_i) = \frac{\delta(\alpha_i)}{r(\alpha_i)} = \frac{\|\mathbf{x}_i - \mathbf{x}_c\|_2}{\|\mathbf{x}_b(\alpha_i) - \mathbf{x}_c\|_2} \quad (3)$$

In this formalism, the density mode is sought by mean shift in spatial dimension using the Epanechnikov kernel

$$K_E(\mathbf{x}) = \begin{cases} \frac{1}{2}c_d^{-1}(d+2)(1-\|\mathbf{x}\|^2) & \|\mathbf{x}\| = d_{norm}(\mathbf{x}) < 1 \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

where c_d is the volume of unit d -dimensional sphere. In the mean shift procedure in spatial dimension, we set $d=2$. Compared to the level set kernel proposed by Yilmaz (Yilmaz, 2007), where the value of the kernel is the distance from the boundary of the mask and may not guarantee a convex and monotonically decreasing profile in some situation as shown in Figure 2(b), the object-shaped Epanechnikov kernel generated from Eq. (4) has a convex and monotonically decreasing profile. Thus strict convergence is ensured (Comaniciu *et al.*, 2002). Moreover, the object-shaped kernel has a very simple derivative and the mean shift vector can be derived by Eq. (5) for the use of Epanechnikov kernel whose profile's derivative is uniform anywhere and therefore we obtained:

$$\Delta \mathbf{y} = \frac{\sum_i w(\mathbf{x}_i)(\mathbf{x}_i - \hat{\mathbf{y}}_{old})}{\sum_i w(\mathbf{x}_i)}. \quad (5)$$

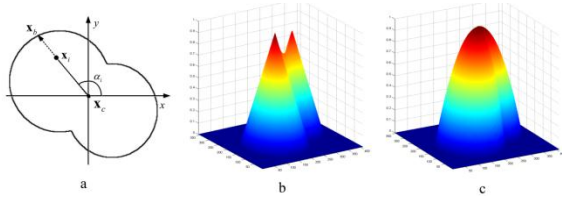


Figure 2. Object and object-shaped kernel. (a) An irregular-shaped object. (b) The level set kernel. (c) The proposed kernel.

3.2 Object-Shaped Scale Invariant Mean Shift

During motion of surveying vehicle, tracked object may be viewed at different sizes (referred to as the object scale) and orientations change is negligible. So we simultaneously estimates the position and scale of the object. The scale of an object pixel is represented by a linear transformation of the image coordinates to the scale dimension $s(\mathbf{x}_i)$ of same value as the normalized distance by (3). We employ the same definition of the scale mean (Yilmaz, 2011), \hat{s} , which is observed to be $\sqrt{2}$, a constant for any type of the object shape, and differentiates the scale dimension from the image coordinates when conducting the mean shift iterations. An update in the scale dimension should take relation between the bandwidth and the scale into consideration.

The scale change can be estimated by setting the dimensions in the Epanechnikov kernel in (4) to $d=1$ and replacing the distance d_{norm} with scale s . Let the mean shift iteration provide a scale update of Δs , the new bandwidth $r_{new}(\alpha)$ will be updated by

$$r_{new}(\alpha) = (1 + \sqrt{2}\Delta s)r(\alpha). \quad (6)$$

To run the mean shift iterations in the joint search space $\Gamma=(x, y, s)$, a 3D Epanechnikov kernel consisting of the product of the spatial object based kernel from and the kernel for the scale dimension

$$K_{3d}(x, y, s) = K_E(x, y)K_E(s), \quad (7)$$

is defined based on the assumption that the incremental updates to each dimension is independent to each other. Using the kernel given in Eq. (7) and let $w(\mathbf{x}_i, s_i)$ be the weight of image pixel at position \mathbf{x}_i and scale s_i the mean shift vector which maximizes the density iteratively is computed by:

$$\Delta \Gamma = \frac{\sum_i w(\mathbf{x}_i, s_i)(\Gamma_i - \Gamma_{old})}{\sum_i w(\mathbf{x}_i, s_i)}. \quad (8)$$

4. GENERALIZED GRABCUT FOR OBJECT SEGMENTATION FROM LFR SEQUENCES

4.1 Graph Cuts-Based Video Object Segmentation and Its Limitations

Graph Cuts-based interactive image segmentation methods (Boykov, Y.Y. *et al.*, 2001) first generate a trimap $T = \{T_F, T_B, T_U\}$, which partitions the image into three sets: T_F and T_B comprises of pixels selected by the user as either foreground or background respectively, and T_U is the remaining set of unknown pixels. The segmentation is formulated as estimating the global minimum and solved by the max-flow min-cut algorithm. The segmentation result is a mask with pixel labelled as the foreground and the background defined as $S = L_F, L_B$.

For object segmentation from video sequences, graph cuts-based video object segmentation approaches generally treat the segmentation results in the previous frame as the label of foreground in current frame to guide the segmentation (Boykov, Y. *et al.*, 2006), i.e., the trimap in the i^{th} frame $T(i) = \{T_F(i), T_B(i), T_U(i)\}$ is determined by the segmentation result $S_{i-1} = L_F_{i-1}, L_B_{i-1}$ in the $(i-1)^{th}$ frame. As shown in Figure 3 (b), and (c), pixels around the boarder of the object mask are generally assigned in $T_U(i)$; other pixels in L_F_{i-1} and L_B_{i-1} are assigned in $T_F(i)$ and $T_B(i)$ respectively. During the segmentation of the i^{th} frame, only pixels in $T_U(i)$ will be partitioned once while the labels of the other pixels remain unchanged.



Figure 3. (a) The yellow brush pot as an instance of OOI, (b) OOI mask, (c) The trimap of the graph cuts-based method, white, gray, and black indicate foreground, the unlabelled and background respectively, and (d) The dual map of the proposed method.

The graph cuts-based methods are hindered by the limitations: (1) It requires that the frame rate of the sequences is high enough to ensure the spatiotemporal consistency of the object appearance in consecutive frames (Bae *et al.*, 2009). Unfortunately, the OOIs undergo large displacement and scaling in the DMI sequences. Direct propagation of the object segmentation results may lay foreground labels on background pixels and vice-versa and cause error segmentation. (2) Segment error will be propagated through the following sequence and there is no chance to correct the errors. (3) Initializations of the graph cuts-based methods need to specify both the foreground and background by trivial sketches. A more flexible labelling and propagation strategy is desired for the OOI extraction through DMI sequences in MMS applications.

4.2 Generalized GrabCut for Video Object Segmentation

Inspired by the GrabCut approach (Rother *et al.*, 2004) for segmenting static images, we propose a generalized GrabCut framework for video object segmentation.

In contrast to the graph cuts-based approaches which build trimap T , we construct a dual map $D_{Map} i = D_U i, D_B i$ for the current i^{th} frame, where D_U and D_B are pixels selected as unlabelled and background respectively, by inheriting the segmentation result $S_{i-1} = L_F i-1, L_B i-1$ in the previous $i-1^{th}$ as prior information for current frame. As shown in Figure 3(d), no foreground pixels are specified. Since there is no mandatory labelling of the foreground, the mislabelling error propagation can be avoided by setting the D_U properly larger than original object mask.

Although OOI in the DMI sequence changes their locations and sizes dramatically in consecutive frames, their appearance remain relatively invariant. Therefore the dual map $D_U i$ and $D_B i$ in frame i can inherit the segment results $S_{i-1} = L_F i-1, L_B i-1$ as well as their color distribution $G_F i-1$ and $G_B i-1$ in frame $i-1$, as initialization for segmentation of the next frame, i.e.

$$G_D i = G_D i | S_D i-1, D \in F, B. \quad (9)$$

For the i^{th} frame, the image consists of N pixels z_n in RGB color space and the color distribution is modelled by two Gaussian mixture models (GMM), $G_F i$ and $G_B i$ for the unlabelled pixels and background respectively. The GMMs are taken to be a full-covariance with K components (typically

$K=5$). In order to deal with the GMM tractably, in the optimization framework, an additional vector $\mathbf{k} = \{k_1, \dots, k_n, \dots, k_N\}$ is introduced, with $k_n = \{1, \dots, K\}$, assigning, to each pixel, a unique GMM component, one component either from G_B or the G_F , according to $\alpha_n = 0$ or 1.

The Gibbs energy, which is formulated as the sum of region term R and boundary term B for segmentation now becomes

$$E(\mathbf{a}, \mathbf{k}, \boldsymbol{\theta}, \mathbf{z}, i) = R(\mathbf{a}, \mathbf{k}, \boldsymbol{\theta}, \mathbf{z}, i) + B(\mathbf{a}, \mathbf{z}). \quad (10)$$

The data term R is now defined, taking account of the color GMM models, as

$$R(\mathbf{a}, \mathbf{k}, \boldsymbol{\theta}, \mathbf{z}, i) = \sum_n D(\alpha_n, k_n, \boldsymbol{\theta}, z_n | S_{i-1}), \quad (11)$$

where $D(\alpha_n, k_n, \boldsymbol{\theta}, z_n | S_{i-1})$ is posterior of appearance distribution of the i^{th} frame based on the segmentation results S_{i-1} of the $i-1^{th}$ frame.

The boundary term B is computed using Euclidean distance in color space:

$$B(\mathbf{a}, \mathbf{z}) = \gamma \sum_{(m,n) \in C} [\alpha_n \neq \alpha_m] \exp\left(-\frac{\|z_m - z_n\|^2}{\beta}\right). \quad (12)$$

The constant β is taken to be the expectation over an image sample.

Having all the parameters defined, the binary segmentation $\hat{\mathbf{a}}$ in the i^{th} frame can be formulated as estimation of a global minimum and is solved by max-flow min-cut algorithm.

$$\hat{\mathbf{a}} = \arg \min_{\mathbf{a}} E(\mathbf{a}, \mathbf{k}, \boldsymbol{\theta}, \mathbf{z}, i). \quad (13)$$

In the proposed scheme, the energy minimization in (13) is run and the color distribution $G_F i$ and $G_B i$ in frame i are updated iteratively until convergence. We adopt morphological post-processing to eliminate the isolated segments and propagate $G_F i$ and $G_B i$ to next frame on-line. By these means, the generalized GrabCut approach has been endowed the ability to segment objects from dynamic image sequences.

5. EXPERIMENTAL RESULTS ON MMS DATA

In this section, we evaluate our OOI extraction method with real-world LFR image sequences collected in Calgary, Canada, by the VISATTM MMS. The size of the frames is 1600×1238, and the frame rate is 1 fps. Only some most significant results are displayed here for the limit of the paper length. To further

demonstrate the proposed approach. The MATLAB demos and more experimental results for this paper are available at http://chwang.xmu.edu.cn/works/demo_ooi_extraction.htm.

5.1 Tracking Performance Evaluation

To evaluating the performance of the proposed object-shaped kernel scale invariant mean shift algorithm, we have performed comparisons with the standard adaptive mean shift algorithm (Comaniciu *et al.*, 2003) and a newest scale adaptive mean shift algorithm, SOAMST (Ning *et al.*, 2011), which adaptively selects scale by compute the moment. In all experiments, the RGB color space was taken as feature space and quantized to $8 \times 8 \times 8$ bins. The OOI is initialized by bounding box drawn by the users in the first frame and then automatically segmented by the GrabCut algorithm (Rother *et al.*, 2004) to generate the object mask for our improved mean shift tracker. The accuracy is evaluated by the distance between the tracked object centroid and the ground truth.

Figure 4 shows the results of tracking a brown rectangular signboard, the tracking details are magnified in the left-top rectangle. Tracking accuracy result of the brown signboard is shown in Figure 5. It can be seen that our method outperforms the other two methods especially when the OOI is moving fast and changing its size dramatically.

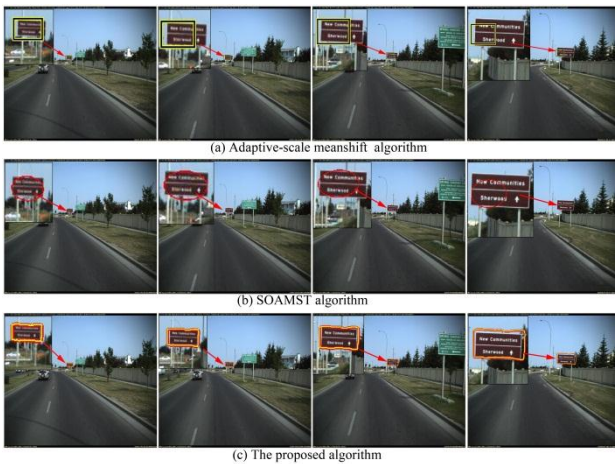


Figure 4. Tracking results of the brown signboard. The frames 3,6,9 and 12 are displayed.

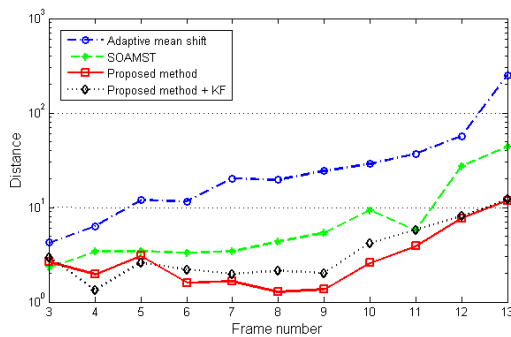


Figure 5. Tracking accuracy of the brown signboard.

We have introduced a Kalman filter module and test its performance. As shown in Figure 10, the accuracy was about the same of that without a Kalman filter. However, in practical

MMS applications, a more accurate initial state of the Kalman filter module can be estimated from the measurable GIS information, which is beyond the discussion of this paper. So the introduction of Kalman filter is essential and promising.

5.2 Segmentation Performance Evaluation

We obtain the candidate region of the OOI in current frame as the result of tracking. Then we segment the object in the candidate region using the proposed generalized GrabCut approach. We present some segmentation results of the traffic light, red guide board and brown guide board sequence in Figure 6.

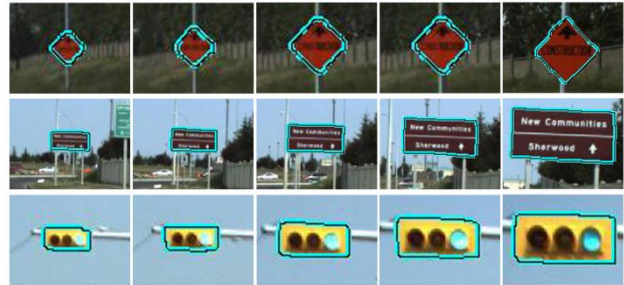


Figure 6. OOI segmentation results for real-world DMI sequence in MMS.

In order to perform objective evaluation of our approach, we first manually segmented the reference objects (or ground truths) for the test DMI sequences. The segmentation performance is measured by the precision, recall and F-score defined Eq. (14) and compared to graph cuts-based methods (Boykov, Y.Y. *et al.*, 2001). We haven't performed matting as the authors do in GrabCut approach (Rother *et al.*, 2004).

$$\text{precision} = \frac{\text{true area} \cap \text{segmented area}}{\text{segmented area}},$$

$$\text{recall} = \frac{\text{true area} \cap \text{segmented area}}{\text{true area}}, \quad (14)$$

$$F = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}.$$

The expectation is to observe the precision, recall and F-score values close to 1 for both measures. The average precision, recall and F-score are shown in Table 1, 2 and 3.

Table 1. Average precision, recall and F-score of the brown guideboard sequence

	Graph Cuts-based method	Our generalized GrabCut
Average precision	0.9773	0.9939
Average recall	0.9660	0.9821
Average F-score	0.9719	0.9879

Table 2. Average precision, recall and F-score of the traffic light sequence

	Graph Cuts-based method	Our generalized GrabCut
Average precision	0.9568	0.9749
Average recall	0.9460	0.9561
Average F-score	0.9513	0.9651

Table 3. Average precision, recall and F-score of the red guideboard sequence

	Graph Cuts-based method	Generalized GrabCut
Average precision	0.9306	0.9329
Average recall	0.9540	0.9569
Average F-score	0.9426	0.9432

From Table 1, 2, 3, our method performs better and more stable segmentation than the graph cuts-based methods. However, our method reports a better average recall and the two other measures differ very little. Moreover, the total interactions involved in our method are much less than in graph cuts-based methods.

In a word, our approach can achieve satisfactory tracking and segmentation results in most sequences. For some extreme cases, where the object undergoes full occlusion, our method may fail. However, introducing some prior geo information is a promising measure to solve this problem and will be researched in future.

6. CONCLUSIONS

In this paper, we propose a novel method to extract OOI in DMI sequence collected by the MMSs. The method integrates object-shaped kernel scale invariant mean shift tracking and the generalized GrabCut segmentation in a unified framework and needs only one bounding box in the keyframe as initialization. The proposed object-shaped kernel is applied to mean shift tracking with adaptive scale selection to keep the temporal consistency of the OOI through the DMI sequence. We generalize the GrabCut method to object segmentation from image sequence by propagating the distribution through sequence. The effectiveness and robustness of the proposed OOI extraction method for DMI image sequences is demonstrated by experiments on real-world DMI sequences collected by the VISATTM MMS.

ACKNOWLEDGEMENTS

This paper is supported by National Natural Science Foundation of China Project 40971245.

REFERENCE

Bae, G., Kwak, S., and Byun, H. (2009), Interactive Object Segmentation System from a Video Sequence, *Lect Notes Comput Sc*, 5618, pp.221-228.

Boykov, Y., and Funka-Lea, G. (2006), Graph cuts and efficient N-D image segmentation, *Int J Comput Vision*, 70(2), pp.109-131.

Boykov, Y. Y., and Jolly, M. P. (2001), Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images, paper presented at ICCV.

Collins, R. T. (2003), Mean-shift blob-tracking through scale space, paper presented at CVPR.

Comaniciu, D., and Meer, P. (2002), Mean shift: A robust approach toward feature space analysis, *IEEE T PAMI*, 24(5), pp.603-619.

Comaniciu, D., Ramesh, V., and Meer, P. (2003), Kernel-based object tracking, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(5), pp.564-577.

Criminisi, A., Cross, G., Blake, A., and Kolmogorov, V. (2006), Bilayer Segmentation of Live Video, paper presented at Computer Vision and Pattern Recognition, 17-22 June 2006.

Hinz, S. (2008), Automatic object extraction for change detection and GIS update, *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXVII(B4), pp.277-284.

Li, Y., Sun, J., and Shum, H. Y. (2005), Video object cut and paste, *Acm T Graphic*, 24(3), pp.595-600.

Malcolm, J., Rathi, Y., and Tannenbaum, A. (2007), Multi-object tracking through clutter using graph cuts, paper presented at International Conference on Computer Vision.

Mena, J. B. (2003), State of the art on automatic road extraction for GIS update: a novel classification, *Pattern Recogn Lett*, 24(16), pp.3037-3058.

Ning, J., Zhang, L., Zhang, D., and Wu, C. (2011), Scale and Orientation Adaptive Mean Shift Tracking, *IET Computer Vision*, unpublished.

Petrie, G. (2010), An Introduction to the Technology Mobile Mapping Systems, *GeoInformatics* 13(1), pp.32-43.

Rother, C., Kolmogorov, V., and Blake, A. (2004), "GrabCut" - Interactive foreground extraction using iterated graph cuts, *Acm T Graphic*, 23(3), pp.309-314.

Wang, C., Hassan, T., El Sheimy, N., and Lavigne, M. (2008a), Automatic Road Vector Extraction for Mobile Mapping Systems, in *ISPRS Congress 08*, edited.

Wang, C. H., and Guan, L. (2008b), Graph Cut video object segmentation using Histogram of Oriented Gradients, *IEEE Int Symp Circ S*, pp.2590-2593.

Wang, J., Bhat, P., Colburn, R. A., Agrawala, M., and Cohen, M. F. (2005), Interactive video cutout, *Acm T Graphic*, 24(3), pp.585-594.

Wu, X. Y., Yang, L., and Yang, C. (2009), Monocular Real-time Foreground Cut Based on Multiple Cues, *Proceedings of the 2009 2nd International Congress on Image and Signal Processing, Vols 1-9*, pp.1836-1840.

Yi, K. M., Ahn, H. S., and Choi, J. Y. (2008), Orientation and Scale Invariant Mean Shift Using Object Mask-Based Kernel, *ICPR*, 1-6, pp.3121-3124.

Yilmaz, A. (2007), Object tracking by asymmetric kernel mean shift with automatic scale and orientation selection, paper presented at CVPR.

Yilmaz, A. (2011), Kernel-based object tracking using asymmetric kernels with adaptive scale and orientation selection, *Mach Vision Appl*, 22(2), pp.255-268.