SEMI-SUPERVISED OBJECT BASED DIGITAL MEASURABLE IMAGE SEQUENCE SEGMENTATION FOR MMS

Peng LI[a], Cheng WANG[a,b,] *, Hanyun WANG[a]

[a] School of Electrical Science and Engineering, National University of Defence Technology, 47 Yanwachi Street, Changsha, Hunan Province, China, 410073, - peterlee20058@hotmail.com, why860314@126.com
[b] Department of Computer Science, School of Information Science and Tech. , Xiamen University, Xiamen City, Fujian Province, China, 361005, - chwang_nudt@263.net

Commission I, ICWG V/1

KEY WORDS: MMS, DMI, Image Sequence Segmentation, Object, Color, Keypoint

ABSTRACT:
Mobile mapping system (MMS) is the trend of state of art mapping technique and digital measurable image sequence is the main data type of Mobile mapping system. As one of the most important pre-processing for mobile mapping system, it is necessary to segment digital measurable image sequences in to manageable and meaningful objects such as guideposts. This paper proposed a novel semi-supervised object based image sequence segmentation method. Firstly, the object of interest such as guideposts or mailbox is specified by the user in a key frame. Secondly, color information is employed to find the candidate region of the object of interest in the new frame. Then local invariant features such as SIFT are employed to describe the object of interest. A novel matching is done in the following frames to segment the region of the object of interest frame by frame. The object template is revised as the matching going on. The segmentation and modelling method we developed is applied to different digital measurable image sequences. Experimental results demonstrate the robustness and effectiveness of our method.

1. INTRODUCTION

As the dramatic developing of GIS, Mobile Mapping System (MMS) as one of important way of producing GIS data has got great progress. MMS is the trend of state of art mapping technique for large scale digital road map drawing and digital measurable image (DMI) sequence is the main data type of MMS. MMS based road GIS data generation is typically divided into three stages: data acquisition, DMI sequence production and GIS data extraction and update.

Through the last two decades, data acquisition and DMI sequence producing for MMS have been solved and successfully used in many applications. However, automatic GIS data extraction and update is still an unsolved puzzle and the bottle neck of the MMS (Wang et al, 2008). Most concerned information in MMS is about objects, such as guideposts, central lines and so on. Effectively segmentation of these objects in DMI sequences is useful for further processing. This paper proposes a novel semi-supervised object-based DMI sequence segmentation method for MMS.

Segmentation and tracking of objects in a 2D image sequence is an important and challenging research area, which has many important applications including change detection, object-based video coding (MPEG-4), video postproduction, content-based indexing and retrieval, surveillance, and 3D scene reconstruction for 3D TV (Correia and Pereira, 2004; Babacan and Pappas, 2006). In real-world application, object based image sequence segmentation suffer from change of illumination, viewpoint and so on. Partly occlusion, as well as background clutter, is also a realistic challenge for segmentation (Correia and Pereira, 2004; Apostoloff and Fitzgibbon, 2006).

State-of-art object based image sequence segmentation technique can be grouped into different categories, for instance, supervised, unsupervised or semi-supervised; region based or boundary based; high level or low level; local information based or global information based; segmentation of mobile objects or static objects. Under the hypothesis of static camera, epipolar-plane image can be used for spatiotemporal volume segmentation of object in image sequences. T-junctions in epipolar image plane (EPI) are used as indicators of boundaries of spatiotemporal volume (Apostoloff and Fitzgibbon, 2006). However, the static camera hypothesis hinders the usage of this kind of methods. Active contours of different kinds have been widely utilized for semi-automatic video object segmentation and tracking (Sun et al, 2003). Vaswani et al proposed that particle filter combined with level set based active contour are used to segment moving and deforming object(s) from a sequence of images (Vaswani et al, 2009). Region based method are often implemented in the graph or hypergraph structure (Huang et al, 2009; Borenstein and Malik, 2006; Brendel and Todorovic, 2009). Region-based methods were developed by performing the clustering operation or regional splitting and growing on the feature space, which is usually formed by motion vectors and some spatial features, like color, texture, and position. However, accurate region boundary is difficult to achieve. Probabilistic issues are often involved in region based methods (Ahmed et al, 2005). Keypoint features are frequently employed to localize the object of interest in an image sequence (Lowe, 2004).

Although the aforementioned method can achieve segmentation of image sequence in different ways, most of them do well only when the camera is relative static or the object is moving. In the

* Correspondence Author: Cheng WANG, Department of Computer Science, School of Information Science and Tech., Xiamen University, Xiamen City, Fujian Province, China, 361005 - chwang_nudt@263.net

application of MMS, both object in DMI and camera may be in motion.

To address the above-mentioned problems, we propose a novel semi-supervised object based image sequence segmentation method in this paper. The two main contributions of this paper are: 1) the proposed method works in a user interactive way, which has been proved can benefit the segmentation result (Zin and Hama, 2007; Correia and Pereira, 2004); 2) we segment the object of interest by combining both their local visual features such as SIFT and global visual features as HS color histogram. Different from common image sequence segmentation methods focused on mobile object segmentation, the proposed method pays special attention on the segmentation of objects in DMI where objects captured by moving camera. In this regard, simultaneous motion of both interest object and background is a rough challenge for our work because the motion information is not that useful. Experiment results demonstrate the proposed method capable to segment object of interest in DMI sequences and its robustness to background clutter, variance of illumination and viewpoint.

This remainder of this paper is organized as follows: In Section 2 discusses the local invariant descriptor used to localize the object. In Section 3, matching by color histogram back projection is described. Our method of combination of SIFT and color information is delineated in Section 4. An analysis of the experimental results is given in Section 5. Section 6 is the conclusion.

2. SIFT FOR OBJECT LOCALIZATION

The Scale Invariant Feature Transform (SIFT) extracts point features from the image and creates a high dimensional description vector (descriptor) for the local image content (Lowe, 2004). The description of the feature is then used to look for matching features in a model image and a test image. The extraction and description step are both invariant to rotational, scaling and illumination effects, as well as the addition of noise. The technique also works for a substantial range of affine distortion and change in viewpoint.

The features are strong extremal points in a Difference of Gaussians (DoG) pyramid (scaling invariance). After extraction, a relative coordinate system (rotational invariance) is assigned to each feature, based on local gradient information extracted at the scale at which the feature point is found. The descriptor is then computed based on local gradient information aligned with the new coordinate system.

For image matching and recognition, SIFT features are first extracted from a set of reference images and stored in a database. A new image is matched by individually comparing each feature from the new image to this previous database and finding candidate matching features based on Euclidean distance of their feature vectors.

We have used an approximate nearest neighbour algorithm, called the Best-Bin-First (BBF) algorithm to match the feature sets of the object image and the new image. This is approximate in the sense that it returns the closest neighbour with high probability. The BBF algorithm uses a modified search ordering for the k-d tree algorithm so that bins in feature space are searched in the order of their closest distance from the query location.

Having got the matched features, we adopt RANSAC (random sample consensus) to reject the outlier and get the fundamental matrix to decide the transformation model.

The SIFT matching results can be seen in Figure 1. As can be seen, the performance of SIFT is not robust enough in the application of MMS. There are two reasons for the decrease of performance: first, object of interest such as guideposts in DMI sequence doesn’t contain enough blob-like structures, which are easily detected by the SIFT keypoint detector, as in common image sequences; second， lots of mismatched points (outliers) are caused by background clutter and viewpoint change in different frames of DMI sequence. After the outlier elimination process by RANSAC, there maybe not enough inliers (less than four) left for calculating the fundamental matrix in some extreme conditions.

(a) (b) (c)

Figure 1. This example shows SIFT keypoints in an image and the matching stage. (a) Template image of object of interest and SIFT keypoints detected as well as descriptors. (b) Matching keypoints between object template and another frame to localize the object. (c) Matching results after elimination outliers by RANSAC.

As shown in the example in Figure 1, there are 53 keypoints detected by the SIFT keypoint detectors in the object template image of a guidepost. However, only 12 of the total 53 keypoints find their matched counterpart in the new frame. After RANSAC process to eliminate the outliers, only 9 inliers left. More than half of the detected keypoints in object template can not find their correspondence keypoints in the new frame, because there are so many keypoints in the new frame that the keypoints in frame produce many mismatches and lead to a degeneration of the performance of SIFT. Another reason for the unexpected low performance is that objects of interest in MMS such as guideposts are not rich in blob-like structures which can be robustly matched by SIFT.

To overcome the challenge in the application of MMS, we must induce some extra information of other features to enhance the robustness of SIFT for object localization and segmentation.

3. COLOR HISTOGRAM BACK PROJECT BASED CANDIDATE REGION OF OBJECT SELECTION

Color, one of the global features, is a feature of the great majority of content based retrieval systems. The color histograms are used to represent the color distribution in an image or a video frame. For digital images, it is basically the number of pixels that have colors in each of a fixed list of color ranges, which span the image's color space, the set of all possible colors.

The color histogram can be built for any kind of color space, although the term is more often used for three-dimensional spaces like RGB or HSV. The color histogram of an image is relatively invariant with translation and rotation about the viewing axis, and varies only slowly with the angle of view (Huang et al, 1999). Furthermore, compared to other invariant features, color histogram is faster to compute. By comparing histograms signatures of two images and matching the color content of one image with the other, the color histogram is particularly well suited for the problem of recognizing an object of unknown position and rotation within a scene.

Usually chi-square, histogram intersection, correlation or Bhattacharyya distances can be employed for the calculation of the images' histograms similarity rating (Allen et al, 2003; Huang et al, 1999). Any of these values does not reflect the similarity rate of two images in itself. It is useful only with comparison to other similar values. Equation (1)-(4) are different methods for measuring similarity of two histograms noted as H_1 and H_2 (Bradski and Kaehler, 2008).

Correlation:

$$d_{correl}(H_1, H_2) = \frac{\sum_i H_1^{'}(i).H_2^{'}(i)}{\sqrt{\sum_i H_1^{'2}(i) \cdot H_2^{'2}(i)}} \tag{1}$$

Chi-square:

$$d_{chi-square}(H_1, H_2) = \sum_i \frac{(H_1(i) - H_2(i))^2}{H_1(i) - H_2(i)} \tag{2}$$

Intersection:

$$d_{intersection}(H_1, H_2) = \sum_i \min(H_1(i), H_2(i)) \tag{3}$$

Bhattacharyya distance:

$$d_{Bhattacharyya}(H_1, H_2) = \sqrt{1 - \sum_i \frac{\sqrt{H_1(i) \cdot H_2(i)}}{\sqrt{\sum_i H_1^2(i) \cdot H_2^2(i)}}} \tag{4}$$

Histogram back projection is a primitive operation that associates the pixel values in the image with the value of the corresponding histogram bin. Color histogram back projection is a low complexity, active vision algorithm for finding objects in complex scenes and is little affected by the movement of camera.

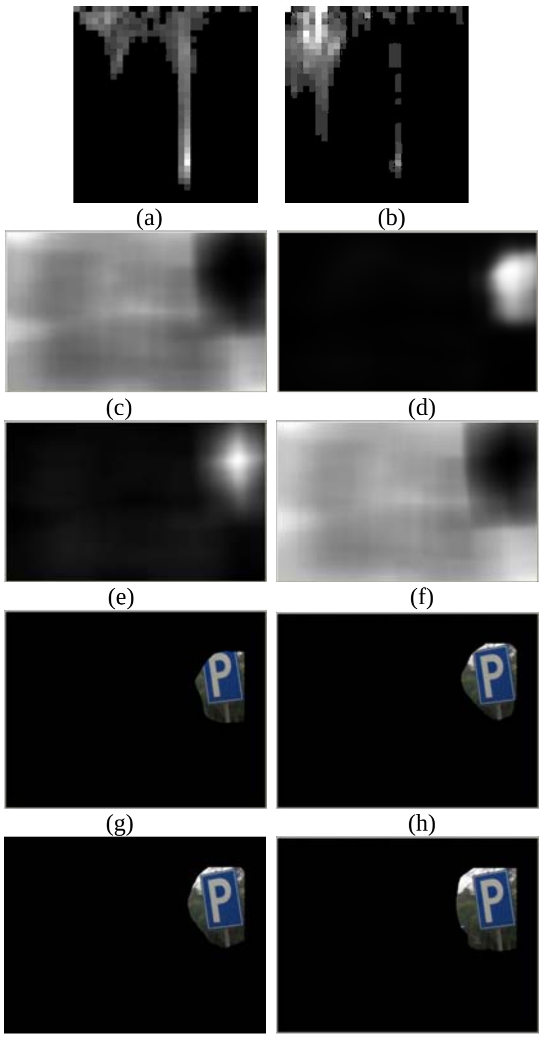
The color histogram back projection method computes the ratio of the color histogram of the object example image (O_i) and the image color histogram (I_i):

$$R_i = \frac{O_i}{I_i} \tag{5}$$

For each image pixel, the probability of it belonging to the sought object is computed as the count of the R_i bin the pixel indexes into. Replacing each pixel with the corresponding probability, a greyscale image regarded as probability image is computed. Object location is established as the pixel with the maximum spatial average over a circular neighbourhood centred on the pixel.

Given a priori knowledge of the size of the sought object, we can choose an appropriate search template/neighborhood. Then we employ the patch-based back projection method to approximately localize the object of interes. We use H-S color histogram in this paper because it is insensitive to viewpoint change, affine transform and illuminance change.

We test the four methods aforementioned in equation (1)-(4) one by one for the application of localizing the object of interest approximately with real-world images. The result is shown in Figure 2.



(a) (b)
(c) (d)
(e) (f)
(g) (h)
(i) (j)

From Figure 2 (g) - (j), we can see that patch-based back projection method using correlation criteria is incompetent to find full candidate region of object of interest. While applying the Chi-square and the Bhattacharyya distance criteria, the threshold of the probability image must be chosen carefully to get the candidate region. The intersection is the most convenient correlation to use.

As can be seen in Figure 2, the employed color histogram based method can select a reasonable candidate region for the object of interest. Furthermore, patch-based color histogram back projection is insensitive to translations such as rotation, scale invariance, of the object template image. Even if we convert the object template 90 degree, we can find the candidate region as well. There are to main drawbacks of the proposed method: firstly, the proposed method is likely to fail to locate non-compact object. However, the non-compact objects are not common in MMS. Secondly, the proposed method can't identify different object that have the same color histogram or it may fail when the background has some regions whose color histogram is similar to the ones of object of interest. Consequently, we combine the color histogram back projection method with SIFT to form a brand new object based image sequence segmentation method, which will be discussed in detail in next section.

4. OBJECT BASED IMAGE SEQUENCE SEGMENTATION BY COMBINING COLOR HISTOGRAM AND SIFT FEATURE IN IMAGE SEQUENCE

The flowchart of the proposed object based image sequence segmentation method is shown in Figure 3.

As shown in Figure 3, the proposed method is a semi-supervised method because of consideration for higher accuracy. Furthermore, automatic indexing of interesting objects is a time-consuming process while the indexing result is of ambiguity. User interaction can significantly alleviate this side effect. The method can be divided in following steps.

Step 1, the object of interest such as guidepost or mailbox is specified by bounding box manually drawn by the user in a key frame of the DMI sequence to form the object template.

Step 2, we compute the H-S color histogram of the object template. Then, the patch-based histogram back projection method is employed to get the candidate region of the object template in a new frame of the DMI sequence. Here we choose the intersection as the similarity measurement for comparing the two histograms.

Step 3, SIFT keypoint detector is employed to detect the keypoints in both the object template and the candidate region. The 128-dimensional SIFT descriptor of each keypoint is calculated. Then, we employ an approximate algorithm, called the Best-Bin-First (BBF) algorithm, to match the keypoints detected to according to the Euclidean distance.

Step 4, we use RANSAC algorithm to eliminate the mismatched keypoints and calculate the fundamental matrix. Once the fundamental matrix successfully computed, we specify the object in current frame by a bounding box and refresh the object template by the region specified by the bounding box in current region. If there are not enough inliers left to calculate the fundamental matrix, we think there is no object of interest exists in current frame. Then we skip in to the next frame and the object template remains unchanged.

Finally, iterate step2 to step4 until the last frame of the DMI sequence.

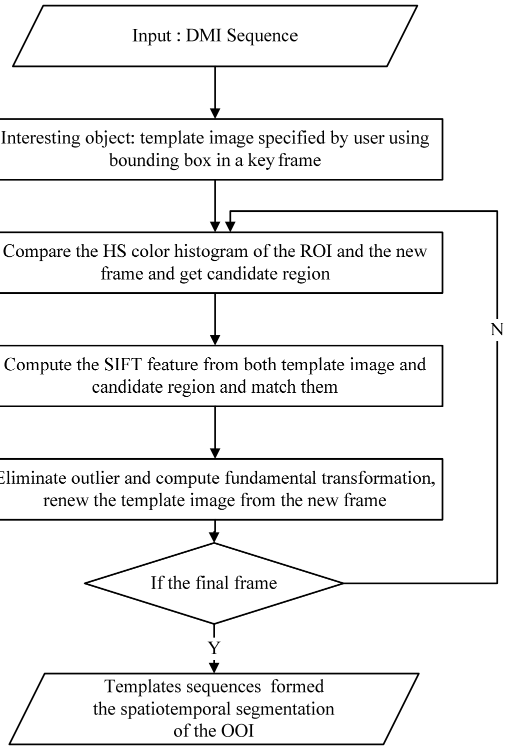
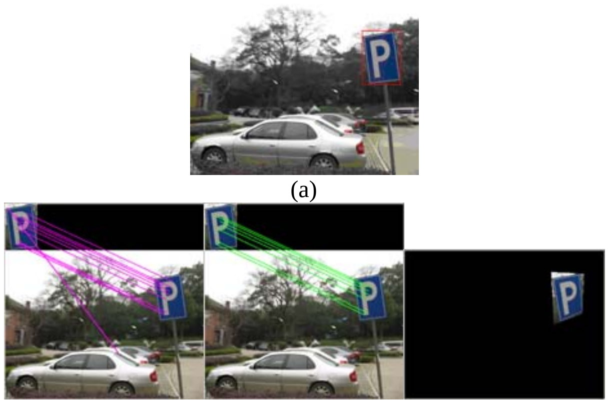


Figure 3. Flowchart of the proposed method

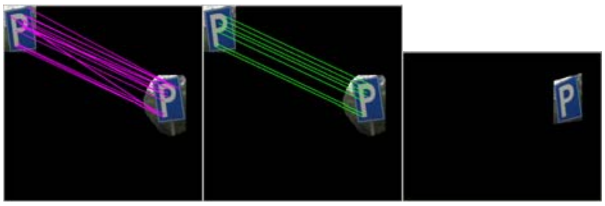
5. EXPERIMENTAL RESULTS

In this section, we present the experimental results of the proposed method. The effectiveness of the proposed method is demonstrated via the experiment results using DMI sequences in real MMS system. The guidepost sequence for experiment contains a 'P' guidepost embedded in a moving clutter background (see Figure 4(a)), the size of each frame is 640× 480.The segmentation result can be seen in Figure 4.



(a)

(b) (c) (d)



(e) (f) (g)

Figure 4. Object based DMI sequence segmentation result. (a) Specified object of interest by a bounding box. (b)- (d) SIFT alone for matching and segmenting the object of interest in image sequence. (e)- (f) Object localization and segmentation result by the proposed method. Compare the segment results in (d) and (g), the object segmented by the proposed method is more accurate.

Compared with using SIFT alone in Figure 4(b), number of matched points in the same frame has been increased from 12 to 17 by the proposed method (in Figure 4(e)). And the overall promotion of numbers of matched points is about 30%. The time for detecting the keypoints without any optimization decreases from more than 7 second per frame to 3 second per frame because the candidate region is much smaller than the original frame so that less keypoints needed to detect and less time consumed. The time consumed for matching is about 250ms per frame (SIFT only) vs about 47ms per frame (our method). Obviously, the method proposed show great improvement in working speed.

6. CONCLUSIONS

We have delineated a novel method for the object-based segmentation of DMI sequences. This methods adopt patch-based color histogram back projection as a pre-process of the SIFT to get a candidate region for SIFT keypoints detecting. As a result of the improvement, more matched keypoints can be detected and the segmentation of the object of interest is more accurate. Experimental results on real-world DMI sequences demonstrate that the proposed methods are successful in difficult scenes with significant background clutter with a great decrease in time consumed. The proposed method can be applied in various MMS.

REFERENCE

Ahmed, R., Karmakar, G. C. and Dooley, L. S., 2005. Probabilistic Spatio-Temporal Video Object Segmentation Incorporating Shape Information. *ICASSP*.

Allen, J. G., Xu, R. Y. D. and Jin, J. S., 2003. Object Tracking Using CamShift Algorithm and Multiple Quantized Feature Spaces. *Pan-Sydney Area Workshop on Visual Information Processin.*

Apostoloff, N. and Fitzgibbon, A., 2006. Automatic video segmentation using spatiotemporal T-junctions. *BMVC.*

Babacan, S. W. and Pappas, T. N.,2006. Spatiotemporal algorithm for joint video segmentation and foreground detection. *EUSIPCO*.

Borenstein, E. and Malik, J., 2006. Shape Guided Object Segmentation. *CVPR.*

Bradski, G. and Kaehler, A., 2008. *Learning OpenCV.* O'Reilly Media Inc., Sebastopol, pp.194-221

Brendel, W. and Todorovic, S., 2009. Video Object Segmentation by Tracking Regions. *ICCV.*

Correia, P. L., and Pereira, F., 2004, Classification of Video Segmentation Application Scenarios. *IEEE Trans. on Circuits and Systems for Video Technology*, 14(5), pp. 735-741.

Huang, J., Ravikumar, S., Mitra, M., Zhu, W.J. and Zabih, R., 1999. Spatial Color Indexing and Applications. *International Journal of Computer Vision*, 35(3), pp. 245–268

Huang, Y. C., Liu, Q.S. and Metaxas, D., 2009. Video Object Segmentation by Hypergraph Cut. *CVPR.*

Lowe, D., 2004. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2), pp. 91–110.

Sun, S.J., Haynor, D. R. and Kim, Y. M., 2003. Semiautomatic Video Object Segmentation Using VSnakes. *IEEE Trans. on Circuits and Systems for Video Technology*. 13(1), pp. 75-82,

Vaswani, N., Rathi, Y., Yezzi, A. and Tannenbaum, A., 2009. Deform PF-MT: Particle Filter with Mode Tracker for Tracking Non-Affine Contour Deformations. *IEEE Trans. Image Processing*, 19(4), pp. 841-857.

Wang, C., Hassan, T., El-Sheimy, N. and Lavigne, M., 2008. Automatic Road Vector Extraction for Mobile Mapping Systems. *XXI Congress,ISPRS.*

Zin, T. T. and Hama. H., 2007. A Method Using Morphology and Histogram for Object-based Retrieval in Image and Video Databases. *International Journal of Computer Science and Network Security*, 7(9),123-129.

ACKNOWLEDGEMENTS

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