VIDEO BASED 3D RECONSTRUCTION USING SPATIO-TEMPORAL ATTENTION ANALYSIS

Xian Xiao[1, 2], Changsheng Xu[1, 2], Yong Rui[3]

[1]National Lab of Pattern Recognition, Institute of Automation, CAS, Beijing 100190, China
{xxiao, csxu}@nlpr.ia.ac.cn
[2]China-Singapore Institute of Digital Media
[3]Microsoft China R&D Group, Beijing, 100080, China
yongrui@microsoft.com

ABSTRACT

3D reconstruction has been widely used in many important applications. While extensive research has been done in 3D reconstruction, several key issues are still open and the precision of the recovered regions is still far from satisfaction. In this paper, we propose a novel approach to selecting regions of interest in video frames by analyzing multiple spatio-temporal characteristics and reconstruct 3D objects based on the selected regions. Firstly, the static, location and motion attention are extracted from video frames to generate saliency maps. Then, all the video frames are clustered and a candidate set of key frames is extracted based on the saliency maps, where the key frames are extracted according to the constraints in terms of geometry and visibility. Finally, the 3D structure of the attention region is recovered using the selected key frames and the generated saliency maps. The experiments on real-world indoor and outdoor scenes demonstrate that the proposed approach is both more accurate (better attention regions) and computationally more efficient.

***Keywords*—**Visual attention, Video analysis, Key frames selection, 3D reconstruction

1. INTRODUCTION

With the development of digital photography, high quality videos become abundant. Since both geometric accuracy and visual quality can be improved by exploiting video data redundancy, video based 3D reconstruction has become a popular research topic in the communities of computer vision, image process and multimedia analysis.

Generally, video/image based 3D reconstruction systems can be classified into two categories: non-calibration based and self-calibration based. The non-calibration based systems need both images and camera parameters to reconstruct 3D objects, e.g., the patch-based multi-view stereo software (PMVS) [1] which enforces local photometric consistency and global visibility constraints to recover 3D structure of an object or a scene being visible in the images. In contrast, the self-calibration based systems firstly estimate the camera parameters from input images via camera self-calibration methods and then recover the 3D points, e.g., Bundler [2]. However, the existing methods only provide 3D structure of a whole scene whereas people only pay attention to the regions which attract their interest in most of the situations. These methods waste much computation power on reconstructing the regions of un-interest and the reconstructed 3D models cannot give prior to the favorite regions.

People always pay more attention to the visually salient regions [3] which can be obtained by visual attention analysis. Much research on visual attention analysis has been studied and widely used in computer vision, artificial intelligence and multimedia processing [4-8]. Most of the pioneer work focused on still images [4][5], which mainly utilized static information. Recently, video attention analysis attracts much more attention. Abdollahian and Delp [6] combined static and location saliency maps to find regions of interest in key frames of home videos. Besides static and location attention, motion, which attracts much human attention, has been widely used to detect attention regions based on spatio-temporal cues [7][8]. Motion vector can be obtained by several methods such as optical flow. However, a critical issue is that the estimation of motion vector under moving camera is still a challenging problem and the motion attention analysis only from the video viewer's perspective is not enough.

Considering the extensive applications of visual attention analysis on region of interest detection, we propose an approach for spatio-temporal attention region generation to enhance the video-based 3D reconstruction. Our approach is tailored to the characteristics of video-based 3D reconstruction including erratic camera motion and the emergence of unexpected objects. In contrast with the traditional 3D reconstruction, our enhanced approach is able to obtain better 3D model accuracy and lower computational cost. The framework overview of the proposed approach is illustrated in Fig.1.

As shown in Fig.1, the flow of our proposed method consists of three steps: 1) video based visual attention analysis, 2) key frames selection and 3) enhanced 3D reconstruction. Firstly, the static, location and motion attention are combined to generate the frame saliency map for each frame. Then, all frames are represented with GIST descriptors and clustered by K-means [9]. For each category of the clustering, a category saliency map is calculated by averaging the involved frame saliency maps. According to the distance between the frame saliency map and the category saliency map, a number of frames are selected from each category to generate a candidate set of key frames. Geometric and visibility constraints are considered for the final key frames selection. Finally, the 3D structure is recovered using the improved PMVS only on the regions of interest corresponding to the frame saliency maps.

978-1-4244-7493-6/10/$26.00 ©2010 IEEE

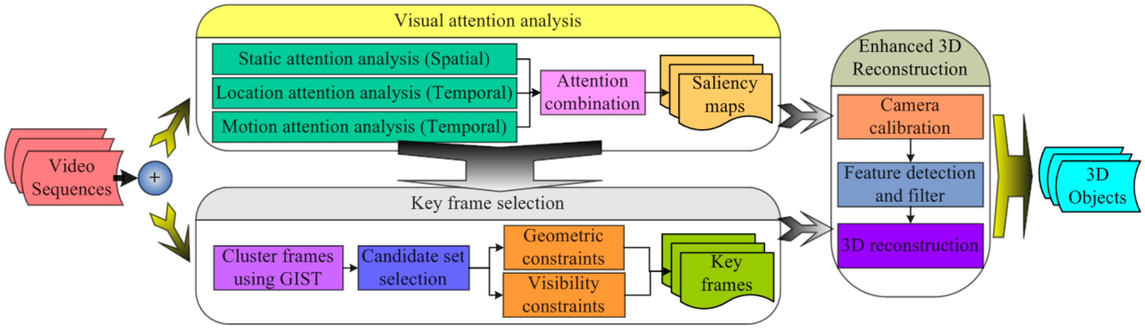

Visual attention analysis
Static attention analysis (Spatial)
Location attention analysis (Temporal)
Motion attention analysis (Temporal)
Attention combination
Saliency maps
Video Sequences
+
Key frame selection
Cluster frames using GIST
Candidate set selection
Geometric constraints
Visibility constraints
Key frames
Enhanced 3D Reconstruction
Camera calibration
Feature detection and filter
3D reconstruction
3D Objects

Fig.1. Framework of video based 3D reconstruction with spatio-temporal attention analysis

Compared with the existing approaches, the contributions of our work can be summarized as two perspectives: (1) we propose a novel approach for video attention region detection, which generates a combined saliency map to represent the regions of interest. Particularly, the integration of static, location and motion attention analysis improves the accuracy of attention region detection; (2) we propose a key frame extraction approach according to the saliency map and geometric and visibility constraints to enable the enhanced 3D reconstruction from video sequence.

The rest of the paper is organized as follows. The details of video attention analysis, video frames clustering and key frame extraction, and 3D reconstruction are described in Section 2, 3, and 4 respectively. Experimental results are reported in Section 5. We conclude the paper with future work in Section 6.

2. VIDEO ATTENTION ANALYSIS

Since the visually salient regions always attract human's attention, we employ the visual attention analysis to extract the regions of interest in the video. The attention analysis is comprehensively performed in terms of static, location and motion in video sequence.

2.1. Static attention analysis

Static objects may attract human attention, which is referred as the static attention. Contrast based attention analysis [4-8] takes the notion that the center-surround structure of receptive field provides human visual system (HVS) sensitivity to feature contrast. Information theory based methods [10] adopt the premise that visual attention proceeds entirely by maximizing the information sampled from an image. Contrast and information sampling are two factors used to evaluate saliency in computational visual attention. Motivated by [7], we integrate the contrast and information to calculate saliency map as follows:

$$Map_{static}(x,y) = Con(x,y) \times ID(x,y) \qquad (1)$$

where *Con(x, y)* and *ID(x, y)* are contrast and information density of *point(x, y)* and normalized to [0, 1].

2.2. Location attention analysis

Location is another important factor affecting human attention. Considering the common sense of photography, the photographers always lay the content of interest on the central part of the still images, and the motion of the camera always follows the photographers' attention.

We utilize the feature in terms of horizontal (*H*), vertical (*V*) and radial (*R*) properties to represent the camera motion which can be estimated using Integral Template Matching technique [11]. With the 3-parameter motion model, the calculation of the three maps for *H*, *V* and *R* directions can be formulated as equations (2), (3), and (4), respectively.

$$Map_H(i,j) = \max(0, 1 - \frac{|j - width/2 - k_H \times H|}{width/2}) \qquad (2)$$

$$Map_V(i,j) = \max(0, 1 - \frac{|i - height/2 - k_V \times V|}{height/2}) \qquad (3)$$

$$Map_R(i,j) = \begin{cases} 1 - r/r_{\max} & \mathrm{R} \geq 0 \\ -k_r \times r/r_{\max} & \mathrm{R} < 0 \end{cases} \qquad (4)$$

where (i, j) is the pixel location, r represents the distance of pixel from the center of the frame and r_{max} is the maximum value of r in the frame. k_H, k_V and k_r are constants whose values were experimentally found to be optimum at 16,12 and 0.5 respectively in our work.

The final location saliency map is obtained as follows:

$$Map_{loc} = Map_H + Map_V + Map_R \qquad (5)$$

2.3. Motion attention analysis

In previous video-based motion attention analysis [7-9], the region containing a moving object was considered to attract more viewers' attention. However, motion attention analysis only from the video viewers' perspective is not rigorous for the interest region detection in the videos captured by moving camera. Our approach analyzes the motion attention from the perspectives of both the video viewers and the photographers. From the perspective of viewers, we analyze which region attracts more viewers' attention. From the perspective of photographers, we

978-1-4244-7493-6/10/$26.00 ©2010 IEEE

investigate which area the photographers prefer to record from the real-world.

In the scene videos, from the viewers' perspective, HVS is more sensitive to the region with high motion intensity than the others. However, from the point of view of photographers, the moving objects with intensive motion destroy the view of the whole scene and only appear for a very short time. This is an irreconcilable conflict between photographers and viewers. In addition, the static objects in the video will be detected to be moving in some situations because of the motion of the camera. The detected motion intensity of a static object far from the moving camera is bigger than that of a nearer one and relies on the distance between the static object and the camera. However, the photographers generally would like to show the objects closed to the camera instead of the farthest or the nearest ones and the viewers always pay more attention to the nearby objects rather than the farthest ones.

In our approach, the regions which attract both the photographers and the viewers' attention regarding the moving camera are motion attention regions. Moreover, the motion attention regions belong to neither the maximal nor the minimal motion intensity region and the visual saliency is inversely proportional to the motion intensity.

We utilize optical flow to detect the motion intensity and represent it with *UV*. The limitation of optical flow under moving camera is that it may detect the moving objects with low-texture to be still, such as white wall or sky.

The mean and standard deviation of motion intensity for each frame are a kind of important description. The motion saliency map is generated as follows:

$$Map_{motion}(x,y)=\begin{cases}0 & UV(x,y)>Mean+\delta\times SD\\ 0 & UV(x,y)<\max(Mean-\delta\times SD,UB)\\ 1-UV(x,y) & Others\end{cases} \quad (6)$$

where *Mean* and *SD* represent the mean and standard deviation motion intensity respectively, δ is the loosing coefficient which is empirically set to be 1.0 in the experiments, *UB* is the upper bound of the error for optical flow detection for the distant low-textured regions whose value is experimentally set to be 0.1.

2.4 Attention fusion

Static saliency map represents the static object which attracts user's attention. The location saliency map describes the distribution of human visual sensibility. The visual salient regions with high human visual sensibility attract more human attention than the lower one. Therefore we generate a location enhanced static saliency map utilizing static saliency multiply the location saliency on each point of the frame. Motion saliency map describes the movement in video sequences to which human vision system is sensitive.

We propose a dynamic fusion technique and the weights of static and motion saliency is determined by the ratio between the mean of the static and motion saliency map for each frame. The final saliency map for each frame is obtained by fusing three saliency maps as follows:

$$Map_{fusion}=Map_{motion}\times\lambda+Map_{loc}.\times Map_{static}\times(1-\lambda) \quad (7)$$

$$\lambda=Mean_{motion}/(Mean_{motion}+Mean_{static}) \quad (8)$$

where λ is the dynamic weight for the motion attention, $Mean_{static}$ and $Mean_{motion}$ are the mean of the static and motion saliency map, respectively.

3. ATTENTION BASED KEY FRAME SELECTION

To select frames for 3D reconstruction, we propose a novel key frame extraction method including three steps. We firstly cluster all the frames into *k* categories based on GIST descriptors. Then for each category, a category saliency map is calculated by averaging the involved frame saliency maps. According to the distance between frame saliency map and category saliency map, we select frames from each category with a predetermined ratio to generate a candidate set of key frames. Any *k* frames coming from the candidate set form a frame group if they belong to different categories. We finally sort the entire possible frame groups with geometric and visibility constraints and determine the key frame group.

3.1 GIST clustering

The goal of clustering is to represent the video content by identifying a set of iconic views corresponding to the dominant aspects in 3D scene. If there are many frames belonging to very similar viewpoints, some of them will at least have a similar image appearance, which can be efficiently matched using a low-dimensional global description of their pixel patterns. We utilize K-means with the global descriptor GIST which was found to be impactful for grouping images by perceptual similarity [9] to cluster frames.

3.2 Candidate set of key frames generation

We generate an average saliency map for each category based on the frame saliency maps which belong to the same category, namely category saliency map. The Euclidean distance between the frame saliency map of category member and the category map is used to rank frames. We select a predetermined rate of frames from each category that are closer to the category saliency map to constitute a candidate set of key frames and at least one frame is selected from each category. The final key frames come from the set. We calculate the rate as follows:

$$\eta=1/(n/k) \quad (9)$$

where η is the rate, *n* is the total number of frames in the video sequence and *k* is the category number.

For each category, we calculate the number of selected frames as follows:

$$S_i=\lceil n_i\times\eta\rceil \quad (10)$$

where S_i is the number of selected frames for the *ith* category, n_i is the total number of frames in the *ith* category.

3.3 Key frame selection

To select the key frame group for 3D reconstruction, we sort all frame groups using geometric and visibility constraints.

The geometric constraints perform verification of each key frames group to confirm whether frames in each group share a common 3D structure. We extract SIFT features [12] and use QDEGSAC algorithm [13] to estimate a fundamental matrix. For a specific frame group, each frame has a number of inliers to the others in the same group. The sum of inliers of a frame group is a new measure for the group namely the geometric constraints score. We rank the frame groups with the score by descending order and the results will affect the final key frame selection.

Frames in different frame groups correspond to different viewpoints. The visibility constraints describe the viewpoints from which a real-world point is visible. For the frames in the candidate set, we recover their viewpoints order using the method in [14] and then obtain their viewpoints rankings. Given a group, we define the visibility loss (*VL*) as follows:

$$VL = \sum_{i=2}^{k-1} |(O_{i-1} + O_{i+1})/2 - O_i| \qquad (11)$$

where *k* is the category number and O_i represents the viewpoint ranking for the *ith* frame of the group. The *VL* is called visibility constraints score for a given frame group. We rank the possible frame groups with *VL* by ascending order and the results will be another influence factor for the final key frame selection.

For each frame group, we add the ranks of geometric and visibility constraints, and determine the key frame group with the smallest sum-rank. If several frame groups own the same smallest sum-rank value, each of them can be selected as the key frame group.

4. ATTENTION ENHANCED 3D RECONSTRUCTION

We propose an attention enhanced 3D reconstruction method which improves the patch-based multi-view stereo algorithm (PMVS) [1] to recover 3D information. Our method is a non-calibration based algorithm. Compared with the previous 3D reconstruction, our enhanced 3D reconstruction not only gives the prominence to the regions of interest but also relieves the computational cost in 3D reconstruction

Firstly, the camera parameters are estimated from the key frames automatically with the structure-from-motion approach in Bundler [2]. Then we detect blob and corner features in each key frame using the Difference-of-Gaussians (DOG) and Harris operators. For each key frame, the regions of interest consist of the pixels with high visual saliency. According to the frame saliency maps, we remove the detected image features which distribute on the regions of un-interest. Finally, the retained image features instead of the whole detected features are supplied to recover the 3D information through a simple match, expand, and filter procedure: (1) initial feature matching: the retained image features are firstly matched across multiple frames with the epipolar consistency, yielding a sparse set of patches associated with salient frame regions. Given these initial matches, the following two steps are repeated n times (*n=3* in our experiments); (2) patch expansion: we utilize a technique similar to [15] to spread the initial matches to nearby pixels and obtain a dense set of patches; (3) patch filtering: filter rely on the visibility consistency are employed to eliminate erroneous matches.

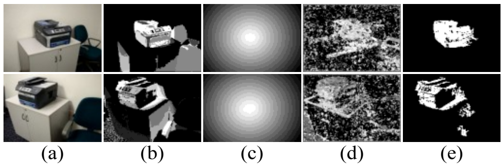


Fig.2. Examples of visual attention analysis results for indoor scene. (a) Original frame, (b) Static attention, (c) Location attention, (d) Motion attention, (e) Fused attention.

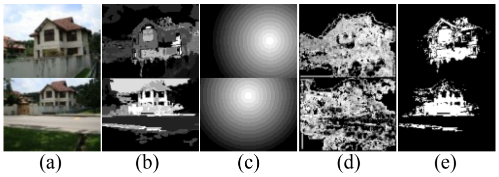


Fig.3. Example of visual attention analysis results for outdoor scene. (a) Original frame, (b) Static attention, (c) Location attention, (d) Motion attention, (e) Fused attention.

5. EXPERIMENTAL RESULTS

We conduct three groups of experiments: visual attention analysis, frame clustering and key frames selection, and 3D reconstruction. Our approach is evaluated on 12 real-world videos including both indoor and outdoor scenes. We present two instances for illustration as shown in Fig. 2 and Fig. 3. The first video is captured in office environment (Fig. 2(a)), and the second one is taken from outdoors scene (Fig. 3(a)). The two videos include 441 and 261 frames, respectively.

5.1 Visual attention analysis

Fig. 2 shows an example of indoor scene. We select two frames from different viewpoints arbitrarily as examples. Location attention in Fig. 2(c) focuses on the center of the frames, Fig. 2(b) and Fig. 2(d) show that the static and motion attention analysis results contain the target area, but they all include additional noise information. Compared with the two saliency maps, the combined saliency maps in Fig. 2(e) show the printer region more accurate.

Meanwhile, the example in Fig. 3 shows an outdoor building. As shown in Fig. 3(b) and Fig. 3(d), both the static and motion saliency map almost fail to describe the target region especially for the second row. However, the final result is very exciting even though many unexpected areas are detected as regions of interest in Fig. 3(b) and Fig. 3(d). The saliency maps in Fig. 3(e) show that our method succeeds in removing most of the unexpected regions such as sky, road and trees etc. Although the second map in Fig. 3(e) contains a long significant region which is unexpected, Fig. 3(e) is much better than Fig. 3(b) and Fig. 3(d). It is accurate enough for the next 3D reconstruction procedure.

5.2 Key frame selection

The key frame selection results are shown in Fig. 4 and Fig. 5. The iconic image selection result using the method in [9] is also provided as the comparison.

In Fig. 4 and Fig. 5, some of the frames selected by our approach and the method in [9] are the same. The saliency maps which cannot well represent the target object region are signed with a yellow bounding box. In Fig. 4, there is only one yellow box in Fig. 4(b) for our results compared with the three in Fig. 4(d) for the results using the method in [9]. In Fig. 5, the number of yellow box for our approach and the method in [9] are one and five, respectively. Since the saliency map will directly affect our 3D reconstruction, the two experiments prove that our key frame selection algorithm is better than the iconic images selection method [9].

Key frames are from variant viewpoints and the disparity between two adjacent frames is not too large. Moreover, the camera motion is complicated. Therefore, although the key frames are sorted according to the visibility constraints, the order is not particularly accurate.

Not all of the saliency maps under key frames appear to be accurate. However, it is worth noting that there are not many unexpected regions simultaneously owned by distinct saliency maps. That means when we match frames, since the extracted points only distribute on the visually salient regions, it is difficult to find a point corresponding to the point from the unexpected region. From this perspective, accurate saliency maps are not prerequisite as long as they do not share the same unexpected region.

5.3 Evaluation of 3D reconstruction

For the indoor scene in Fig. 6, we provide three groups of frames to reconstruct 3D models: iconic images results [9], our key frames extraction results and our improved PMVS results. The time-consuming for our key frames extraction is 2.5 hours compared with the iconic images extraction is 47 minutes. The computational costs for the three 3D reconstruction processes are 3.5 hours, 3.5 hours and 1 hour respectively. Therefore, the total computational cost for our enhanced 3D reconstruction is lower than the other two methods. Here the second column of Fig. 6(c), Fig. 6(d) and Fig. 6(e) is the views from the right-side viewpoint. From the right-side viewpoint, it is obvious that the 3D model in Fig. 6(d) is more accurate than that in Fig. 6(c) at the region with a yellow rectangle frame which illustrates that our key frames extraction approach is effective. Although the saliency map in Fig. 6(b) is not very accurate, the 3D model in Fig. 6(e) shows that only the user attention region is reconstructed and the accuracy is similar to the results in Fig. 6(d).

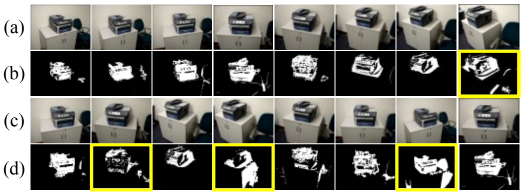


Fig.4. The frames are clustered into eight categories, the selected frames and saliency maps are shown. (a) is the selected key frames with the proposed algorithm; (b) is saliency maps of (a); (c) is the iconic images using the method proposed in [9]; (d) is saliency maps of (c). We mark the unsatisfactory saliency maps with a yellow bounding box.

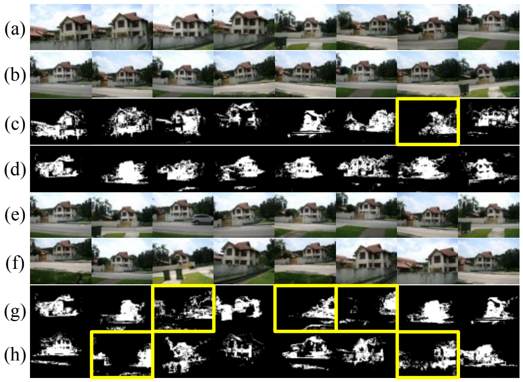
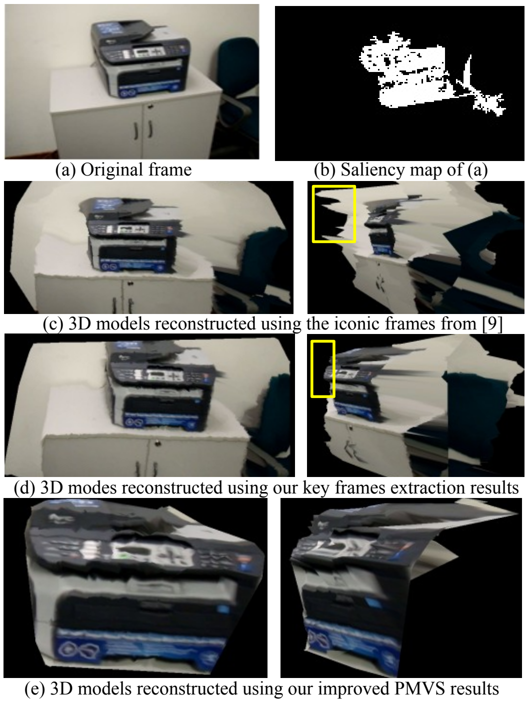


Fig.5. The frames are clustered into sixteen categories, the selected frames and saliency maps are shown. (a) and (b) are the selected key frames with the proposed algorithm; (c) and (d) are saliency maps of (a) and (b); (e) and (f) are the iconic images using the method proposed in [9]; (g) and (h) are saliency maps of (e) and (f). We mark the unsatisfactory saliency maps with a yellow bounding box.



(a) Original frame

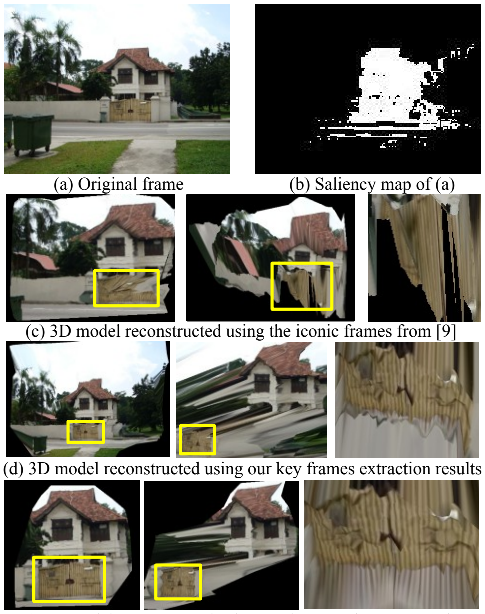
(b) Saliency map of (a)

(c) 3D models reconstructed using the iconic frames from [9]

(d) 3D modes reconstructed using our key frames extraction results

(e) 3D models reconstructed using our improved PMVS results

Fig.6. An example for the printer's 3D reconstruction. The first column of (c) (d) and (e) is views from the forward viewpoint; the second column is views from a right side viewpoint.



(a) Original frame (b) Saliency map of (a)

(c) 3D model reconstructed using the iconic frames from [9]

(d) 3D model reconstructed using our key frames extraction results

(e) 3D model reconstructed using our improved PMVS results

Fig.7. An example for the old building's 3D reconstruction. The first column of (c) (d) and (e) is views from the forward viewpoint; the second column of (c) is a vertical view and of (d) and (e) is views from a right side viewpoint; the third column is the detail of the "door" region.

In the outdoor video, the scene is much more complicated than the indoor scene. We also provide three groups of frames to reconstruct 3D models and the computational costs are 8 hours, 8 hours and 1.5 hours respectively. We spent 2 hours for our key frames extraction and 40 minutes for iconic images extraction [9]. Compared with the first two groups of frames our approach has lower computational cost. In Fig. 7, the first column of Fig. 7(c) and Fig. 7(d) is the views from the forward viewpoint and the 3D model in Fig. 7(d) is more integrated than the one in Fig. 7(c). From the second column of Fig. 7, it is apparent that the "door" region with a yellow rectangle frame in Fig. 7(c) was not reconstructed well, and the position of the door is totally wrong. However, the result in Fig. 7(d) is better. The recovered 3D points on the "door" region in Fig. 7(d) basically share a same plane which is in accordance with the practical situation and the position of the "door" region is right. This proves that our key frames extraction approach is effective. The enhanced 3D model shown in Fig. 7(e) focuses on the visual attention region only and the accuracy is similar to Fig. 7(d).

6. CONCLUSION

In this paper, we have presented a novel approach to enhancing video based 3D reconstruction. Relying on visual attention analysis, we are able to make 3D reconstruction focus on the user attention regions which relieves the computational cost. In addition, we consider geometric and visibility constraints for key frames extraction and improve the reconstruction accuracy. The experimental results validate that the proposed approach is an efficient and robust solution for user attention regions 3D reconstruction on both indoor and outdoor scenes. Our approach is also able to be used in many other applications such as video coding, summary generation, and adaptive browsing on small screens.

For the future work, we will investigate reconstructing objects from more complicated videos and extend our method to wider range of applications.

7. ACKNOWLEDGEMENT

The research is supported by supported by NSFC (National Natural Science Foundation of China) under the contract of 60970092

8. REFERENCES

[1] Y. Furukawa and J. Ponce, "Accurate, dense, and robust multi-view stereopsis," IEEE Conference on Computer Vision and Pattern Recognition, June.2007.

[2] N. Snavely, S.M. Seitz, R. Szeliski, "Bundler.", [Online], Available: http://phototour.cs.washington.edu/bundler/

[3] J. Duncan and G.W. Humphreys, "Visual search and stimulus similarity," Psychological review, 1989.

[4] L. Itti, C. Koch and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, no. 11, 1998

[5] Y.F. Ma and H.J. Zhang, "Contrast-based image attention analysis by using fuzzy growing," Proceedings of the eleventh ACM international Conference on Multimedia, pp. 374–381, 2003.

[6] G. Abdollahian and E. J. Delp, "Finding regions of interest in home videos based on camera motion," IEEE International Conference on Image Processing, pp.545-548, Sept-Oct 2007.

[7] H.Y. Liu, S.Q. Jiang, Q.M. Huang and C.S. Xu, "A generic virtual content insertion system based on visual attention analysis," Proceedings of the sixteenth ACM international Conference on Multimedia, pp. 379-388, Oct.2008.

[8] Y.F. Ma, X.S. Hua, L. Lu, H.J. Zhang, "A generic framework of user attention model and its application in video summarization," IEEE Transactions on Multimedia, vol.7, no.5, pp.907-919, Oct. 2005.

[9] X.W. li, C.C. Wu, C. Zach, S. Lazebnik and J.M. Frahm, "Modeling and recognition of landmark image collections using iconic scene graphs," The 10th European Conference on Computer Vision, pp. 427-440, Oct.2008.

[10] N. D. B. Bruce, "Features that draw visual attention: an information theoretic perspective," Neurocomputing, vol. 65-66, pp. 125-133, 2005.

[11] D. Lan, Y. Ma, and H. Zhang, "A novel motion-based representation for video mining," Proceedings of IEEE International Conference on Multimedia and Expo, July 2003.

[12] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," International Journal of Computer Vision, vol.60, no.2, pp. 91-110, 2004.

[13] J.M. Frahm, and M. Pollefeys, "RANSAC for (quasi-) degenerate data (QDEGSAC)," IEEE Conference on Computer Vision and Pattern Recognition, pp.453-460, June.2006.

[14] R. Basri, P.F. Felzenszwalb, R.B. Girshick, D.W. Jacobs and C.J. Klivans, "Visibility constraints on features of 3D objects," IEEE Conference on Computer Vision and Pattern Recognition, June.2009.

[15] M. Lhuillier and L. Quan, "A quasi-dense approach to surface reconstruction from un-calibrated images," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 3, pp. 418–433, 2005.