A Pen Based Tool for Annotating Planar Objects

Satoshi Yonemoto

Graduate School of Information Science, Kyushu Sangyo University, Japan
yonemoto@is.kyusan-u.ac.jp

Abstract. In recent augmented reality (AR) application, marker-less tracking approaches are often used. Most marker-less tracking approaches force user to capture the front view of a target object during the initial setup. We have recently proposed two image rectification methods for non-frontal view of a planar object. These methods can be applied to reference image generation in marker-less AR. This paper describes a pen based tool for annotating planar objects. Our tool builds upon several interactive image rectification methods, and supports registration of AR Annotations, marker-less tracking and annotation overlay.

Keywords: image rectification, marker-less tracking, AR annotation.

1 Introduction

In recent augmented reality (AR) application, robust tracking algorithm is used, which does not use any artificial markers. It is known as marker-less tracking [1]. In general, most planar objects in natural scenes are observed in non-frontal view. Most marker-less tracking approaches force user to capture the front view of a target object during the initial setup. In online use, user cannot always capture a front view image (e.g., large sign-board and building surface). We have recently proposed two image rectification methods for non-frontal view of a planar object [9] [10]. These methods can be applied to reference image generation in marker-less AR. Now, we have developed a pen based tool for annotating planar objects. This tool enables on-line registration of AR annotations, marker-less tracking and annotation overlay. The following functions are realized: live image view, image rectification (reference image generation) and registration of AR annotations. In live image view mode, registered objects are tracked and then the corresponding AR annotations are overlaid.

Image rectification method is often used in text detection in natural scene [2] [3]. In most text detection methods, two approaches are taken. The first is vanishing point estimation [4] [5]. The second is quadrangle estimation [6]. A quadrangle is warped into a rectified text area (i.e., a rectangle), correcting any perspective distortion. In document images used in OCR, text lines are regularly aligned in paragraph, so rich horizontal lines are stably extracted. Two text lines and the vertical stroke boundaries are often used as a quadrangle clue [4]. Our method is one of quadrangle estimation methods extended to general planar objects.

Image rectification methods should be used in dependence on appearance of each target object. Therefore, our tool is equipped with several interactive methods.

R. Shumaker and S. Lackey (Eds.): VAMR 2014, Part I, LNCS 8525, pp. 418–427, 2014.

2 Overview

2.1 Image Rectification Algorithm

In the proposed method, it is assumed that target image has rich horizontal and vertical lines. If a target object has a rectangular shape such as book and poster, then their corner points are clearly observed. In this case, the reference image is easily acquired from their points (Fig. 3). The proposed method can also be applied to non-rectangular objects such as circular objects. This idea is realized by using horizontal (or vertical) lines with target object, that is, these lines always become horizontal, correcting any perspective distortion. We assume that at least two horizontal lines and two vertical lines would be observed. Our method is one of quadrangle estimation methods [9]. A quadrangle with a target object is constructed by picking up 4 points or 4 lines in an image. The projection of a rectangle is a quadrangle. Our approach to find such quadrangle is summarized as follows (see Fig. 1): First, specify the region of interest in an interactive way. Next, extract horizontal and vertical line segments in the target image by using probabilistic Hough transform [11]. Then, make a quadrangle hypothesis from their line segments. Finally, warp the quadrangle into a rectangle and then, evaluate whether re-projected line segments will be horizontal (vertical) or not. The goodness function is defined by total error with the warped horizontal and vertical lines. The important point is whether extracted lines will be transformed horizontal (vertical) or not. As a result, the underlying problem is equal to a search problem for a combination of 4 line segments. Fig. 2 shows an example of several quadrangle hypotheses generated from the extracted line segments. The line segments are extracted only in the region of interest. The quadrangle hypotheses q1, q2 and q3 are shown (each goodness is also described in parentheses). Quadrangle hypothesis indicates a combination of different 4 line segments that are selected from among the extracted line segments. In Fig. 2, q3 is the best quadrangle with the max goodness 327. 4 line segments that are part of the quadrangle q3 are also shown.

<u>Image Rectification Algorithm</u>

Step 1	Extract horizontal / vertical line segments (check the proper baseline if box based)
Step 2	Select 4 line segments & generate a quadrangle hypothesis q_i
Step 3	Warp q_i into a rectangle & evaluate the goodness
Step 4	Find the best quadrangle with max goodness

Fig. 1. Our image rectification algorithm

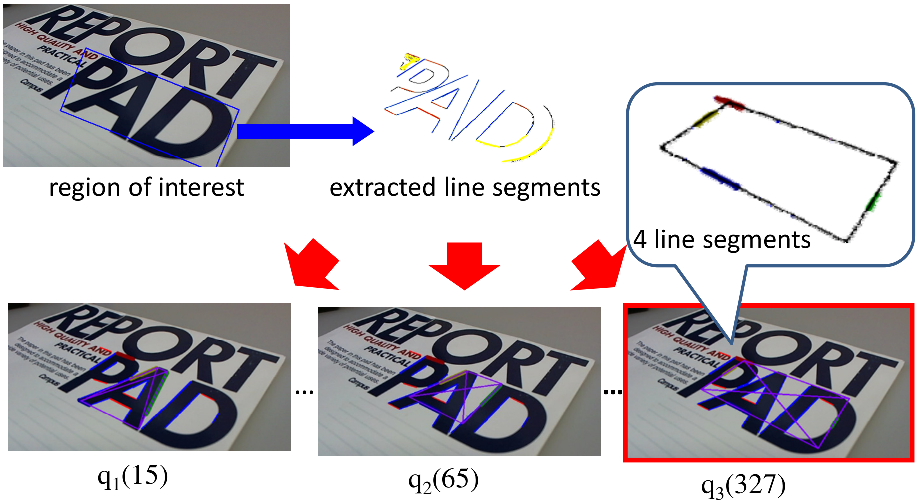


Fig. 2. Quadrangle hypotheses

2.2 Interactive Methods for Image Rectification

To acquire a fronto-parallel image, we can use any one of 4 interactive methods: (A) 4 corner points based, (B) line selection based, (C) user-stroke based and (D) box based. That is, user can select a region of interest in different way. Fig. 3 shows our interactive methods. Method A is used for only rectangular objects. From 4 corner points (which are manually selected by user), the rectified image is generated by warping. Method B is used when user can observe at least two horizontal lines and two vertical lines in the region of interest. User draws their lines on the screen. Note that our method can also use invisible lines existing virtually in the image. In method C, user can directly paint a stroke(s) on a region of interest. Pen size can be selected in advance. Multiple strokes are also supported. In this method, it is assumed that user stroke direction is parallel to a horizontal baseline. In method D, user can specify a bounding box (or freehand area selection) as a region of interest. In this case, user does not clearly define the baseline. To find valid baseline, we take a hypothesis and test approach. Thus, method B, C and D are semi-automatically controlled. After selecting a region of interest, our image rectification algorithm is executed. Image cropping is performed by user (auto-cropping is also supported). The cropped image is registered as a reference image.

Fig. 4 shows two examples of method A (4 corner points based). Fig. 5 shows the results of method B (line selection based). Two horizontal lines (invisible lines) and two vertical lines are picked up by user. The best quadrangle is overlaid in the figure. The rectified image is shown as a fronto-parallel image. Fig. 6 shows the results of method C (user-stroke based). User-strokes are represented by red circles. The baseline is parallel to the bounding box of the strokes. Fig. 7 shows the results of method D (box-based). Valid baseline is automatically determined.

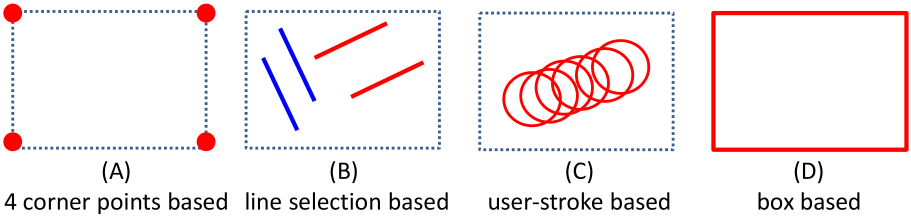


Fig. 3. 4 Interactive methods for image rectification

Fig. 4. 4 corner points selection based method. (left) book (right) poster.

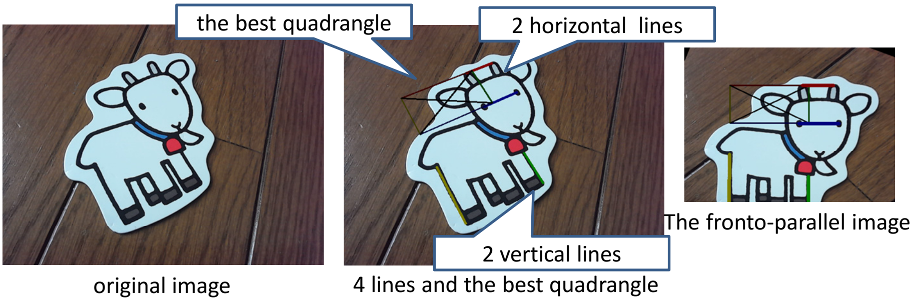


Fig. 5. Line selection based method

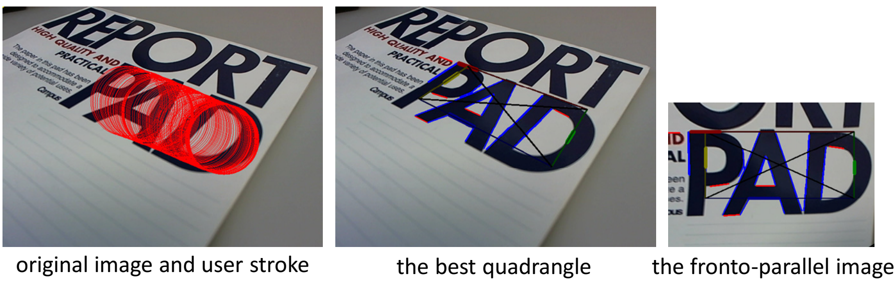
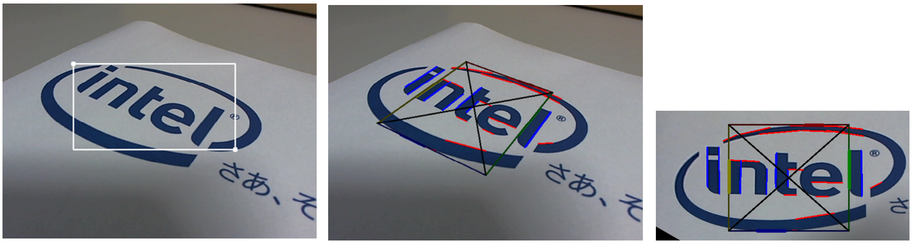


Fig. 6. User-stroke based method



original image and the box | the best quadrangle | the fronto-parallel image

Fig. 7. Box based method

2.3 Registration of AR Annotations

After image rectification, AR annotations are registered for the cropped reference image. Our tool supports the following AR annotations: image texture (virtual graffiti), text, url and transparent marker. Virtual graffiti annotation can be created online, directly painting on the target image (i.e., the preview image) or the reference image. It can be painted on the extended area around the target region. Fig. 8 shows an example of a virtual graffiti (yellow and pink color). When the virtual graffiti is directly painted on the preview image, it is warped into a rectified texture. Given 4 point correspondences with a target region, a homography is computed [12] [13]. The computation is proceeded by the same rectification process as method A. Text and url are also registered. Transparent marker is realized by a see-through overlay. Fig. 9 shows two examples of marker-less tracking and annotation overlay. Target objects were cropped and then registered in advance. Virtual graffiti (top) is created by painting on the reference image. Virtual graffiti (bottom) is created by directly painting on the preview image (see Fig. 8). In this case, AR annotation is overlaid and displayed on the extended area around the target region.

2.4 Marker-Less Tracking

We have implemented marker-less tracking algorithm based on keypoint based descriptors and the trackers (SURF) [7]. It is one of the state of the art trackers, and it is often used at the base of many visual tracking problems. In our marker-less tracking algorithm, the tracker can use any kind of planar object as long as sufficient texture information is available. 4 corner points can be estimated as the tracking result. Thus, AR annotation (i.e., virtual graffiti) can be overlaid and displayed on the area corresponding to the target object on the screen. In Fig. 9, the tracking results are shown. Using 4 corner points of a planar object, real-virtual camera pose parameters can be estimated [2] [12]. Although our tool is capable of displaying any 3D annotations such as 3D text and CAD model, our tool supports only 2D annotations to provide a way that is convenient for a user.

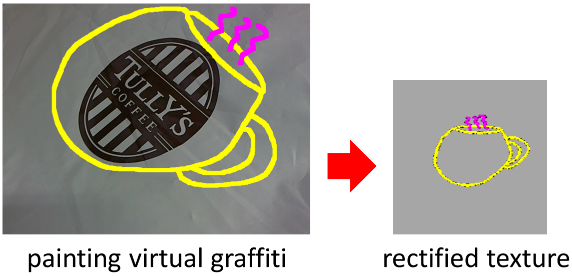


Fig. 8. Virtual graffiti creation

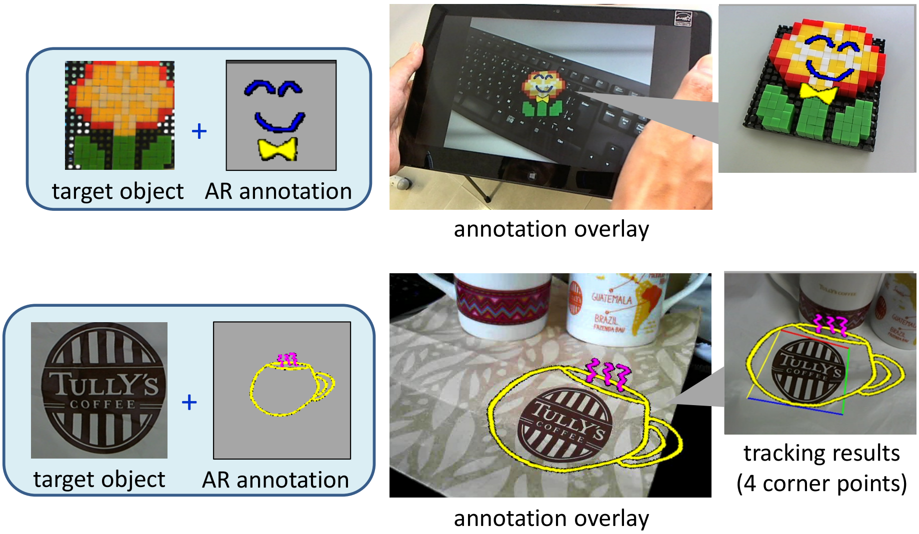


Fig. 9. Marker-less tracking and annotation overlay

3 Experiments

3.1 Implementation

We have implemented our tool on a tablet PC with digitizer pen (OS: windows 8, CPU: Intel Core i7, memory: 8GB and screen: 11.6-inch). Real images are captured using a rear facing camera device. Marker-less tracking is implemented with OpenCV library. Virtual graffiti creation is implemented with OpenGL library.

3.2 Experimental Evaluation

In experiments, we compared and analyzed the performance of 3 interaction methods (B, C and D). We found that the following conclusions. Although method B takes relatively much time for line selection process, the desired results can be easily acquired. Invisible lines are often selected for figures have a symmetric structure. In method C, user strokes determine the baseline that is used to extract horizontal and vertical line segments, so use of multiple strokes affects the accuracy. Although method D takes only a little time, estimation of false baseline affects the accuracy.

3.3 Applications

At present, our tool is used in our practical AR applications. We have implemented prototype versions of ”diminished reality for marker-less AR”, ”marker-less AR with table-top interface”.

Diminished Reality for Marker-Less AR. Diminished reality technique is realized by removing an object or collection of objects and replacing it with an appropriate background image. After object detection, projecting AR annotation is built to replace the target object. In our system, inpainting technique is used to remove the target object [8]. Our tool is used in generating the mask image of a target region. And then, the mask image is warped in conjunction with the object motion. Fig. 10 shows our diminished reality application. First, the mask image is painted by user. And then it is registered as rectified texture by the similar way of creating virtual graffiti (top). Using the extended mask image, the target object is removed (i.e., virtually disappeared). And then, AR annotation is overlaid on the background image that the corresponding object region has been removed. AR annotation can be displayed in marker hiding. The warped mask image (middle) and the removed background image (bottom-left) are shown in Fig. 10.

Marker-Less AR with Table-Top Interface. In table-top display style, user can share AR annotations with the other people. Using marker-less tracking technique, AR annotations can be displayed on table-top display that puts a target object. That is, they are displayed below (around) the moving object. Fig. 11 shows our table-top display application. Target object is a teddy. Both table-top display and the target object are constantly captured by a web camera (left). According to the object location, AR annotation should be projected on the display screen. Using both the tracking results and the display position (4 corner points) in the web camera view, AR annotation is projected and then displayed on the screen (right). If user has an AR glasses (i.e., wearable display), the prototype system can provide the mix world. That is, AR annotations are displayed in both the front and back of the target object [15].

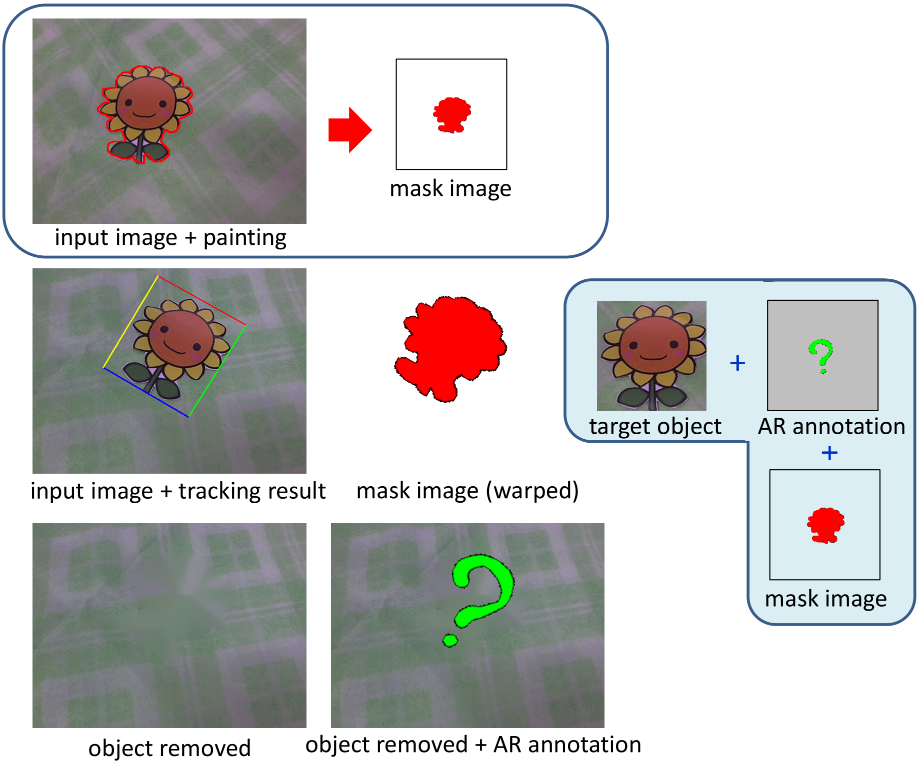


Fig. 10. Diminished reality application

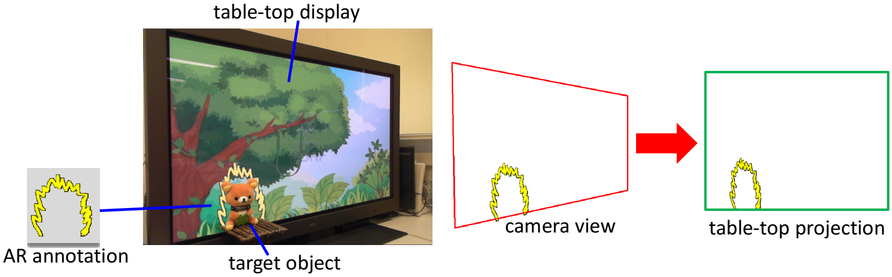


Fig. 11. Table-top display application. (left) table-top display and the target object. (right) conversion from camera view to table-top view.

4 Conclusion

We proposed a pen based tool based on image rectification method for non-frontal view of a planar object. This tool is designed to support online task of annotating planar objects. This tool enables on-line registration of AR annotations, marker-less tracking and annotation overlay. We have proposed four interactive image rectification methods to crop a fronto-parallel image. Method A (4 corner points based) is used for rectangle objects. Method B (line selection based) is used when user can observe at least two horizontal lines and two vertical lines in the target region. In method C, user can directly paint a stroke(s) on the target region. The stroke is used as a baseline to extract line segments. In method D, user can specify a bounding box as a region of interest. Three image rectification methods (B, C and D) are one of quadrangle estimation methods that are extended to general planar objects. In the proposed methods, it is assumed that target image has rich horizontal and vertical lines. In our experiments, we showed promising cropping results. And we showed marker-less tracking results using the reference image.

Our tool supports registering AR annotations. User can execute online creation of AR annotations with the target object. In particular, virtual graffiti can be directly painted on the preview image and then superposed on the live image immediately.

We would also like to examine our tool for various target objects. Currently, performance on the whole task depends on the tracking accuracy. There is probably room for improvement in marker-less tracking method. Future work includes consideration about AR annotation sharing with other users, which is typified by Stiktu [14].

References

1. Fue, P., Vincent, L.: Vision based 3D tracking and pose estimation for mixed reality. In: Emerging Technologies of Augmented Reality: Interfaces and Design, pp. 1–22 (2005)
2. Jain, A., Yu, B.: Automatic text location in images and video frames. In: Proc. 14th Intl. Conf. on Pattern Recognition, vol. 2, pp. 1497–1499 (1998)
3. Fragoso, V., et al.: TranslatAR: A Mobile Augmented Reality Translator. In: Proc. of IEEE Workshop on Applications of Computer Vision (WACV), pp. 497–502 (2011)
4. Clark, P., Mirmehdi, M.: Estimating the orientation and recovery of text planes in a single image. In: Proceedings of the 12th British Machine Vision Conference (2001)
5. Yin, X.-C., et al.: Robust Vanishing Point Detection for MobileCam-Based Documents. In: International Conference on Document Analysis and Recognition. IEEE (2011)
6. Lu, S., et al.: Perspective rectification of document images using fuzzy set and morphological operations. Image and Vision Computing 23(5), 541–553 (2005)
7. Bay, H., Ess, A., Tuytelaars, T., Gool, L.V.: Speeded-Up Robust Features (SURF). Computer Vision and Image Understanding Archive 110(3), 346–359 (2008)
8. Telea, A.: An image inpainting technique based on the fast marching method. Journal of Graphics Tools 9(1), 23–34 (2004)
9. Yonemoto, S.: An interactive image rectification method using quadrangle hypothesis. In: Proc. of International Conference on Image Analysis and Processing (ICIAP 2013), pp. 51–60 (2013)
10. Yonemoto, S.: A Reference Image Generation Method for Marker-less AR. In: Proc. of 17th Conference Information Visualisation, pp. 410–415 (2013)
11. Matas, J., Galambos, C., Kittler, J.V.: Robust Detection of Lines Using the Progressive Probabilistic Hough Transform. CVIU 78(1), 119–137 (2000)

12. Hartley, R., Zisserman, A.: Multiple view geometry in computer vision. Cambridge University Press, New York (2001)
13. Zhang, Z., He, L.-W.: Note-taking with a camera: whiteboard scanning and image enhancement. In: Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2004), vol. 3 (2004)
14. Stiktu, http://stiktu.com/
15. Yonemoto, S.: Seamless Annotation Display for Augmented Reality. In: Proc. of International Conference on Cyberworlds, p. 37 (2013)