Region-based tracking using sequences of relevance measures

Sandy Martedi
Keio University
3-14-1, Hiyoshi, Kohoku-ku,
Yokohama, Kanagawa,
223-8522, Japan
sandy@hvrl.icskeio.ac.jp
Bruce Thomas
University of South Australia
D1-07, Mawson Lakes SA
5095, Australia
Bruce.Thomas@unisa.edu.au
Hideo Saito
Keio University
saito@hvrl.ics.keio.ac.jp

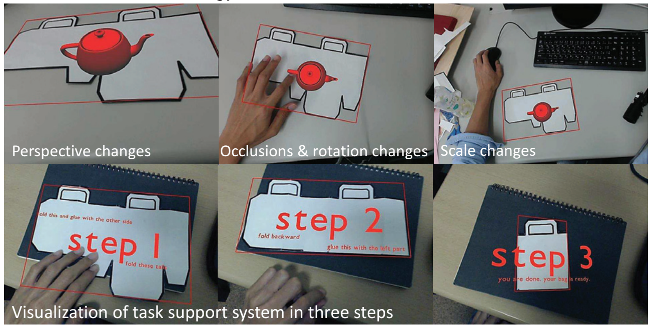


Figure 1: Region-based tracking in action (first row) and the visualization of task support system for making a paper craft object (second row).

IEEE International Symposium on Mixed and Augmented Reality 2013
Science and Technology Proceedings
1 - 4 October 2013, Adelaide, SA, Australia
978-1-4799-2868-2/13/$31.00 ©2013 IEEE

Abstract
We present the preliminary results of our proposal: a region-based detection and tracking method of arbitrary shapes. The method is designed to be robust against orientation and scale changes and also occlusions. In this work, we study the effectiveness of sequence of shape descriptors for matching purpose. We detect and track surfaces by matching the sequences of descriptor so called relevance measures with their correspondences in the database. First, we extract stable shapes as the detection target using Maximally Stable Extreme Region (MSER) method. The keypoints on the stable shapes are then extracted by simplifying the outline of the stable regions. The relevance measures that are composed by three keypoints are then computed and the sequences of them are composed as descriptors. During runtime, the sequences of relevance measures are extracted from the captured image and are matched with those in the database. When a particular region is matched with one in the database, the orientation of the region is then estimated and virtual annotations can be superimposed. We apply this approach in an interactive task support system that helps users for creating paper craft objects.

Author Keywords
Artificial, augmented, virtual realities

ACM Classification Keywords

H.5.1 [Multimedia Information Systems]: Artificial, augmented, and virtual realities.

Introduction

In design processes such as clothes making, designers usually draw particular shapes or write notes on the fabrics. Showing virtual annotations instead of actually writing the notes on the fabrics for helping designers to finish their work is interesting use case in augmented reality.

In order to show virtual information on fabrics, we need to register it. When the surface is registered, the camera pose then can be estimated and annotation can be superimposed. Conventionally, a rectangular fiducial marker [4] is used to register a planar surface. The marker must be visible in the surface for performing detection and tracking. Therefure, such marker is not suitable for system that requires physical change of the surface. On the other hand, random dot markers [6] can have irregular shapes. However, random dot markers and other texture-based registration methods are not fit for texture-less fabric. Therefore, we can only rely on other features such as edges or outlines of the shapes that are drawn on the fabric.

We propose a method that applies region detection (MSER) method for tracking shapes in a surface for realizing a task support system that uses some patterns printed or drawn on the fabric as illustrated in Figure 2.

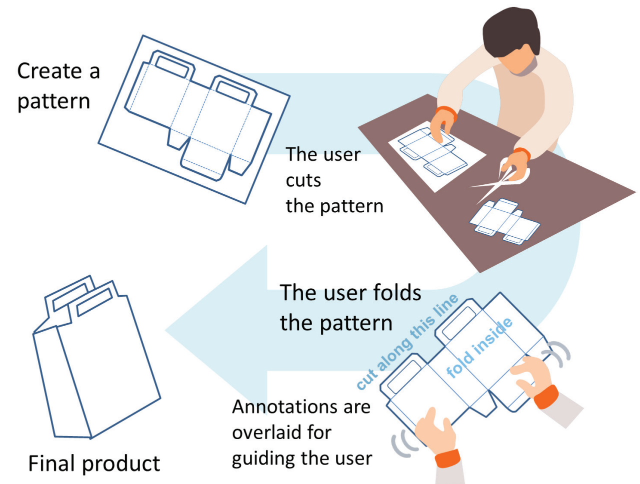


Figure 2: Scenario for making a paper craft object. The user cuts and folds a paper for making a paper bag as instructed in the virtual annotations.

Related Works

Shapes or regions registration problem has been explored in previous investigations. Bergig et al. have developed an application for augmented reality that recognize handwriting in real time [1]. Their method recognizes 2D drawings and displays its corresponding 3D shapes. Explorations on arbitrary shapes for planar registration is done by Hagbi et al. [3]. They used a classification method for searching region template in database. Similarly, Donoser et al. proposed a method for tracking regions using MSER [2]. We explore the similar method as proposed by Donoser et al [2] and add the robust local feature in the registration process. We simplify the matching process by keeping the keypoints from the polygon instead of using many MSER region templates.

Proposed Method

In a nutshell, we proposed a tracking method that consists of features extraction and shape registration using sequence of relevance measures as descriptors.

Features Extraction

The features extraction is initialized by applying MSER to the input image as illustrated in Figure 3. In order to detect a region, one MSER must be extracted. The border of the MSER is simplified using relevance measure that is computed using three consequent points that form two connected lines in the shape outline is defined as

$$r = \frac{\theta l_1 l_2}{l_1^2 + l_2^2} \qquad (1)$$

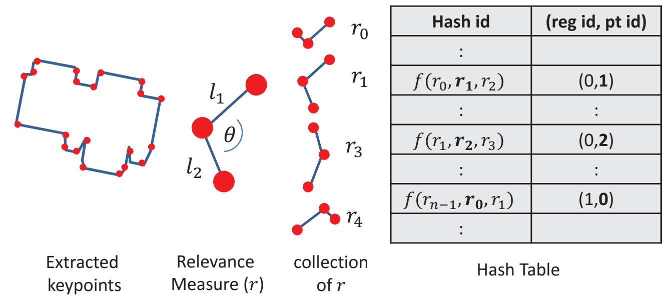
where l_1 and l_2 are the length of two connected segments (lines) and θ is the angle between two segments. The point that connects two segments is removed from the polygon if the relevance measure is smaller than threshold. This removal process is iterated until only the points with high relevance measure remains (the remaining points are called keypoints). The relevance measure are recomputed and then used to describe a particular keypoint for the shape registration.

The features extraction is done off-line for making the feature database (hash table) and shape database. It is also done on-line for registering the unknown shapes (see Figure 4).

Figure 3: A stable region is extracted using MSER and its outline polygon is simplified by filtering the keypoints with the high relevance measure.

Shapes Registration

During the shapes registration, the keypoints of unidentified region are extracted. The sequences of relevance measures are then computed and the corresponding tuple (region id, keypoint id) is looked up in the hash table (see Figure 4). Since the hash table is many-to-one relationship, in order to get the matched region and keypoint, the voting using a histogram of matched keypoints is performed. This process yields keypoints correspondences between a shape captured in the camera and a shape in the database.



Hash id	(reg id, pt id)
:	
$f(r_0, \boldsymbol{r_1}, r_2)$	(0,**1**)
:	:
$f(r_1, \boldsymbol{r_2}, r_3)$	(0,**2**)
:	:
$f(r_{n-1}, \boldsymbol{r_0}, r_1)$	(1,**0**)
:	

Figure 4: Sequences of relevance measures are stored in the database as the indices of the hash table. An index of the hash table refers to a tuple that consist of a region id (region id, for multiple regions detection) and keypoint id (pt id). $f(r_{n-1}, r_n, r_{n+1})$ is the hashing function for keypoint n that can be implemented as a string of (r_{n-1}, r_n, r_{n+1})

Technically, we choose three neighbouring relevance measure values to represent a keypoint of a shape. In this case, one keypoint is actually described by its four neighbours keypoints. We assume that we can also increase the number of relevance measure to four in order to represents one keypoints for example r_0, r_1, r_2, r_3 for keypoint with $id = 1$.

Pose Estimation

The camera pose is estimated using homography that is calculated using at least four keypoints correspondences as the result of the shape registration. The camera pose is then optimized using Levenberg-Marquardt [5] by minimizing the re-projection error that is the distance between the projected keypoints from the shape database and the extracted keypoints in the captured frames. The camera pose is also refined by considering the keypoints correspondence to the detected shape in previous frame. These two optimizations produce a stable camera pose.

Scenario of task support system

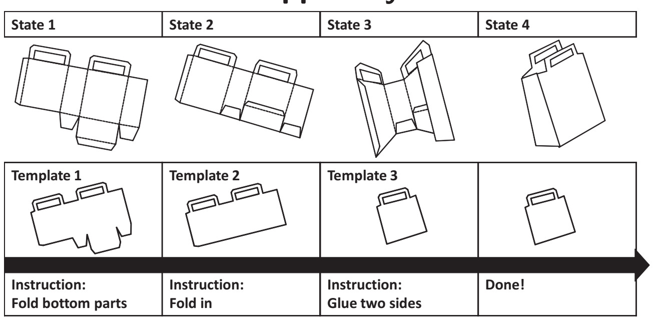


Figure 5: The flow of instructions seen by the user and the state for tracking individual region. Three templates are prepared beforehand in order to identify the current state of user's action. The user cuts the region of the paper craft and follows the instructions that are displayed.

We are developing a prototype of the task support system using the proposed region-based tracking. We use a piece of paper is used as the target surface. The user then cuts and folds the paper in order to make the final product by following steps illustrated in Figure 5. The information is superimposed virtually over the paper to guide the user as illustrated in Figure 1.

Evaluation

We show the accuracy of detection and tracking by calculating the re-projection error as the results of orientation and scale changes and also occlusions. We also evaluated our method by detecting multiple shapes.

Re-projection error

We capture the paper craft shape (template in the scenario) in 1184 image frames. The error are calculated by projecting the keypoints in the template (database) to the captured image using the computed homography. Then the average distance of the projected keypoints to the shape outline in the captured image is plotted as the error (See Figure 6).

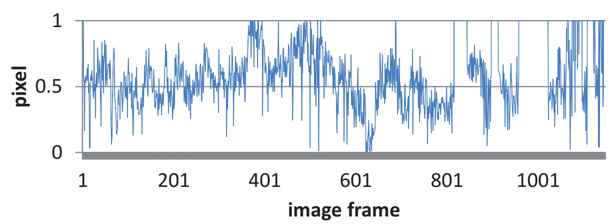


Figure 6: Re-projection error of tracking using paper craft shape. The error for successful detection and tracking is lower than 1 pixel. Error that higher than 1 pixel and missing values are because of failure detection, extreme orientation and occlusions.

Orientation and scale changes

For making a shape descriptor, we take into account the ratio of length of the edge and angle which are invariant to the scale and rotation changes. Therefore, the shape descriptor is effective to handle the scale changes as illustrated in Figure 7.

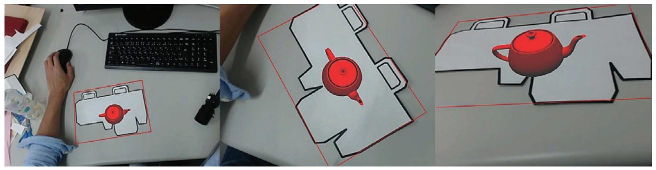


Figure 7: Scale changes.

Likewise, the angle between two edges does not change in every rotation, which make our method works robustly against rotation changes as illustrated in the Figure 8.

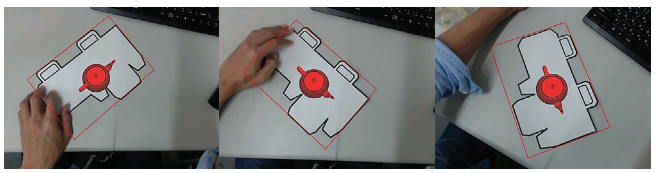


Figure 8: Rotation changes.

In handling the orientation changes, we update the descriptor database in every successful tracking for the next matching. Furthermore, we use also the previous frame in addition of the descriptor information from the database. Hence, our method could handle the perspective changes as illustrated in Figure 9.

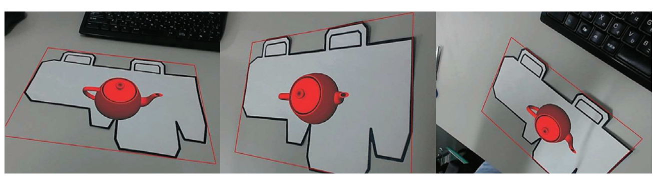


Figure 9: Perspective changes.

Occlusions
Our method maintains the sequence of local descriptors from the outline that describes a shape. Therefore, some missing descriptors due to occlusions can be tolerated as long as another part of the shape is visible as shown in Figure 10.

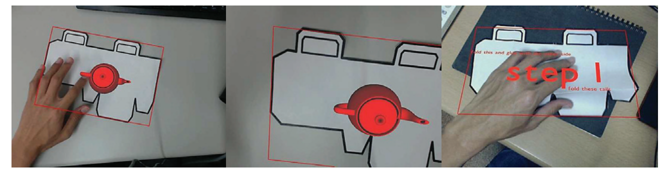


Figure 10: Occlusions.

Multiple shapes
We examine the effectiveness of the sequence of descriptors for detecting multiple shapes. We tested using three categories of 100 random shapes (based on angle and the length of segment) as listed in the Table 1. The test shows promising result because more than ten shapes can be detected.

Table 1: Multiple shapes detection result.

Shapes Categories	Angle ($^\circ$)	Segment (pixel)	Num. of Keypoints	Error (%)
	5-10	10-120	25	9
	15-30	10-120	21	25
	0-50	70-120	24	15

Current Limitations

The first limitation is that our method recognizes shapes with curves inaccurately because the extraction method produces inconsistent keypoints in every frame. We assume that by improving the shape extraction method, the accuracy can be improved. We are currently examining other extraction method to handle shapes that contain curves. The second limitation is in the implemented task support system. We use multiple templates for identifying tasks. The templates are ordered in the database based on the task order. The second template is the part of the first template due to the folding. Because our method use local descriptors, it becomes difficult to register the second template since the descriptors of second template is actually the part of the first template. Currently, after the first task (folding), the first template is set to be unavailable for matching. As a result, the flow of task is irreversible. Furthermore, the number of keypoints decreases due to the folding task that makes the matching often fails in the last task. In term of computation speed, since we rely on optimizations that involving all keypoints in the shape, for a shape that has 30 keypoints, the program runs in 23fps while for a shape that has 79 keypoints, the speed decreases drastically to 6fps. We are improving the current implementation so that the computation speed will not be affected by the number of keypoints.

Conclusion

We introduced a region-based tracking method for arbitrary shapes registration. Our proposal is utilizing the sequence of relevance measure that is extracted from arbitrary shape. We applied the MSER method for extracting the stable region in a surface. We then track the detected regions and estimate the pose of each region. We have shown that our method could handle the orientation and scale changes and also detect multiple shapes. We applied the tracking method for developing a task support system for helping users to make paper or fabric craft objects.

References

[1] Bergig, O., Hagbi, N., El-Sana, J., and Billinghurst, M. In-place 3d sketching for authoring and augmenting mechanical systems. In *Proc. ISMAR* (oct. 2009), 87 –94.

[2] Donoser, M., Kontschieder, P., and Bischof, H. Robust planar target tracking and pose estimation from a single concavity. In *Proc. ISMAR* (oct. 2011), 9 –15.

[3] Hagbi, N., Bergig, O., El-Sana, J., and Billinghurst, M. Shape recognition and pose estimation for mobile augmented reality. *Visualization and Computer Graphics, IEEE Transactions on 17*, 10 (oct. 2011), 1369 –1379.

[4] Kato, H., and Billinghurst, M. Marker tracking and HMD calibration for a video-based augmented reality conferencing system. In *Proc. IWAR* (1999), 85–94.

[5] Lourakis, M. levmar: Levenberg-marquardt nonlinear least squares algorithms in C/C++. [web page] `http://www.ics.forth.gr/~lourakis/levmar/`, Jul. 2004. [Accessed on 31 Jan. 2005.].

[6] Uchiyama, H., and Saito, H. Random dot markers. In *Proc. IEEE VR* (march 2011), 35 –38.