Pattern Recognition Letters 68 (2015) 176–182

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Entire reflective object surface structure understanding based on reflection motion estimation☆

Qinglin Lu[a,*], Eric Fauvet[a], Anastasia Zakharova[b], Olivier Laligant[a]

[a] University of Burgundy, Le2i UMR 6306 CNRS. 12, Rue de la Fonderie, 71200, France
[b] INSA Rouen LMI EA3226. Avenue de l'Université, 76800, France

ARTICLE INFO

Article history:
Received 6 May 2015
Available online 13 October 2015

Keywords:
Entire reflective object
Sub-segmentation
Surface structure understanding
Reflection motion features
Elementary continuous surface

ABSTRACT

The presence of reflection on a surface has been a long-standing problem for object recognition since it brings negative effects on object's color, texture and structural information. Because of that, it is not a trivial task to recognize the surface structure affected by the reflection, especially when the object is entirely reflective. Most of the cases, reflection is considered as noise. In this paper, we propose a novel method for entire reflective object sub-segmentation by transforming the reflection motion into object surface label. To the best of our knowledge, the segmentation of entirely reflective surfaces has not been studied. The experimental results on specular and transparent objects show that the surface structures of the reflective objects can be revealed and the segmentation based on the surface structure outperforms the approaches in literature.

1. Introduction

The *object surface structure* (OSS) describes the geometric distribution of the elementary continuous surfaces of an object (the definition of elementary continuous surface is given in Section 3.2). It is a highly representative feature obtained by performing a sub-segmentation of the surface. The understanding of the OSS is considered as a building block for solving problems such as object recognition, detection, and classification. For non-reflective objects, the OSS can be easily recognized due to the object's contour, texture, and color. However, for the entire reflective objects, the reflective effects make the understanding of OSS extremely complicated. For instance, as shown in Fig. 1, Fig. 1a is the original image of an entire reflective object which consists of both specular and transparent surfaces; Fig. 1b is the ground-truth of the manual sub-segmentation according to the OSS. We can see that due to the reflection on the object, the boundaries are barely observable and the OSS is hard to recognize. Moreover, because of the transparent surface, undesired components inside the object are also visible. Thus, the sub-segmentation from Fig. 1a to b is not a trivial task. The objective of this paper is to sub-segment entirely reflective objects using the information provided by reflection.

In this paper, the reflection motion features are extracted in the image sequence as spatiotemporal information, then object is segmented by taking these features in order to understand the OSS. Both the camera and object are fixed, the light source is moving around the object in order to produce *reflection particles* (RP) on the object surface. The surface is supposed to be piecewise elementary continuous, i.e. it consists of several elementary continuous subsurfaces.

We assume that while the RP are moving on the object surface, their positions, directions, and velocities are extracted in each frame as reflection motion features. These features are matched in all the frames for tracking the RP in the whole sequence. The trajectories of RP are smooth along the subsurfaces. While they are passing through the boundary of two subsurfaces, irregular features (jumps) appear. Thus, we stop tracking when the trajectories are not smooth enough with respect to the previous frames. This guarantees that the trajectory of a moving RP stays on the same elementary continuous subsurface. Then, the surface is segmented by employing flood fill method [24] which takes the positions in the trajectory as seeds. As this process iteratively covers all the trajectories, different surfaces of the object could be respectively labeled.

Our primary contributions are: (1) we introduce an effective sub-segmentation method for the reflective surface structure understanding (on both specular and transparent surfaces). (2) Instead of removing reflection, we study the reflection motion and we consider it as additional information for sub-segmentation. (3) We use the reflection motion features as spatiotemporal coherence for video segmentation and fine-attributes for OSS understanding.

The rest of the paper is organized as follows. In Section 2, we give an overview of the related work. In Section 3, we present the reflection motion features extraction and RP matching and tracking, and also, we explain how sub-segmentation is performed in order to take

☆ This paper has been recommended for acceptance by Dr. Y. Liu.

* Corresponding author. Tel.: +33 650390506.
E-mail address: qinglin.lu@u-bourgogne.fr, luqinglin1024@gmail.com (Q. Lu).

http://dx.doi.org/10.1016/j.patrec.2015.09.006
0167-8655/

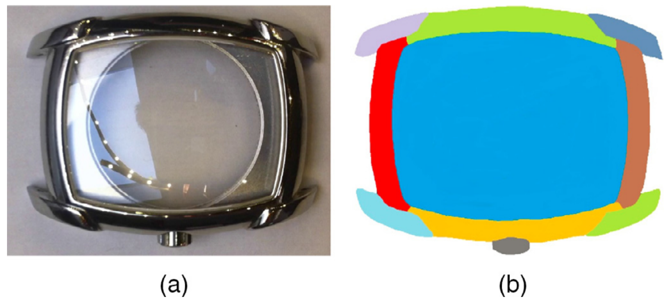

Fig. 1. Reflective object structure understanding. (a) Original image and (b) manually sub-segmented ground-truth image.

into account the reflection motion features. The results of our approach on multiple reflective objects and the comparison with other segmentation methods are shown in Section 4. Conclusion and future work directions are discussed in Section 5.

2. Related work

Dealing with reflection: Many works have been done in dealing with reflection in the image. The most common idea is to consider the reflection as noise, then try to remove or reduce it, such as the methods proposed in [9,16,22,23]. However, several attempts have been made to use information contained in reflections to extract object features. Savarese and Perona [18,19] propose an analysis of the relationship between a calibrated scene composed of lines through a point, and the geometry of a curved mirror surface on which the scene is reflected. This analysis is used to measure object surface profile. DelPozo and Savarese [7] use static specular flows features to detect specular surfaces on natural image. Barrois and Wohler [2] present a method which incorporates different channels of information, one of which is a polarization angle of light reflected from the object surface. It provides information on the rotation of an object relative to the camera.

Video object segmentation: Many methods have been proposed for video object segmentation. Most existing methods attempt to exploit the temporal and spatial coherence in the image sequance, in which pixels with similar appearance and spatiotemporal continuity are grouped together over a video volume [15,17,26]. There are also some works [12,21] that adapt graph-based image segmentation to video segmentation by building the graph in the spatiotemporal volume. Shi and Malik [20] use Nystrom normalized cuts, in which the Nystrom approximation is applied to solve the normalized cut problem for spatiotemporal grouping. Grundmann et al. [14] apply hierarchical graph-based approach in segmenting 3D RGBD point clouds by combing depth, color, and temporal information. Moreover, about scene segmentation using RGBD data, Bergamasco et al. [3] employ a game-theoretic clustering schema which benefits from the macropixels pairwise similarities to combine color and depth information.

Object sub-segmentation in detail: Approaches closest to ours investigate in extracting fine-gained attributes for object recognition [4,8,10,11,13]. Deng and Feifei [8] present an attribute-based framework for describing object in details which is generalized across object categories. Bourdev and Malik [4] use 3D data of human body which is annotated into different body parts to recognize the pose. Vedaldi et al. [25] propose a method for understanding objects in detail by studying the relation between part detection and attribute prediction. It diagnoses the performance of classifier that pool information from different parts of an object. However, the attributes used by these authors are no more accurate in presence of reflection, thus these methods are not robust in object segmentation in case of reflective surfaces.

The proposed approach extracts reflection motion features in the image sequence as spatiotemporal information, then sub-segment object by taking these features as fine-gained attributes in order to understand object surface structure. Comparing to other reflection dealing methods, we do not use any prior knowledge like calibrated camera or textured environment. Furthermore, to the best of our knowledge, the use of reflection motion features as spatiotemporal coherence for video segmentation and fine-attributes for object structure understanding has not been yet studied.

3. Methodology

Our goal is to transform the motion of reflections into useful information that can help to segment the different continuous surfaces of an object. The proposed pipeline is made up of three main tasks depicted in Fig. 2. First step is the RP motion feature extraction; followed by a RP tracking process; finally the sub-segmentation is conducted by taking the RP motion trajectories as labeling information.

3.1. Motion estimation of reflection

The motion of RP provides temporal information, thus in order to employ the RP moving information for object sub-segmentation, we firstly extract motion features of all the moving RP in the video.

3.1.1. Reflection motion features extraction

Since our object and camera are fixed, in the video, movements could only be produced by reflections due to the movement of the light source (Fig. 3). We use the motion history image [1,6] (*MHI*) to extract RP. The *MHI* $H_\tau(x, y, t)$ can be computed from an update function $\Psi_\tau(x, y, t)$:

$$H_\tau(x,y,t) = \begin{cases} \tau & if \quad \Psi_\tau(x,y,t) = 1 \\ max(0, H_\tau(x,y,t-1) - \delta) & if \quad \Psi_\tau(x,y,t) = 0 \end{cases} \qquad (1)$$

Precisely, if $\Psi_\tau(x, y, t) = 1$, then the pixel at position (x, y) in t-th frame has moved. The duration τ decides the temporal extent of the movement, and δ is the decay parameter. More details refer to [1,6]. This leads to a static scalar valued image where the more recently moving pixels are brighter. Then the moving direction can be efficiently calculated by convolution with separable Sobel filters in the X

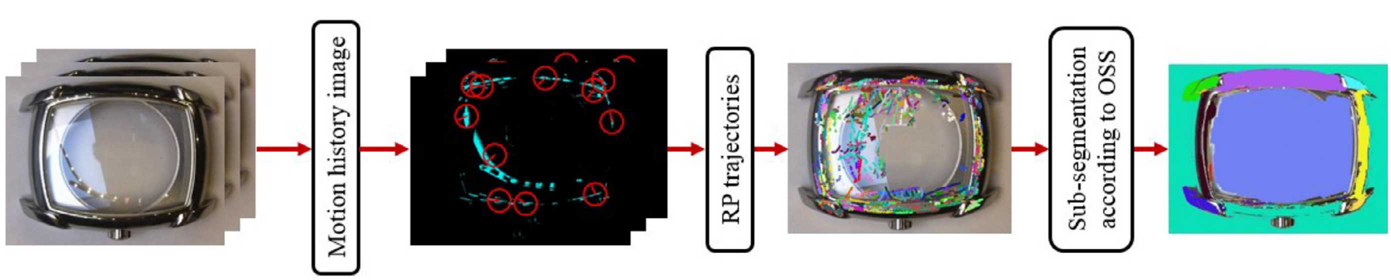

Fig. 2. Illustration of the proposed pipeline (see text for details).