



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.

© 2015 Elsevier B.V. All rights reserved.

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 seg-

mented 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).

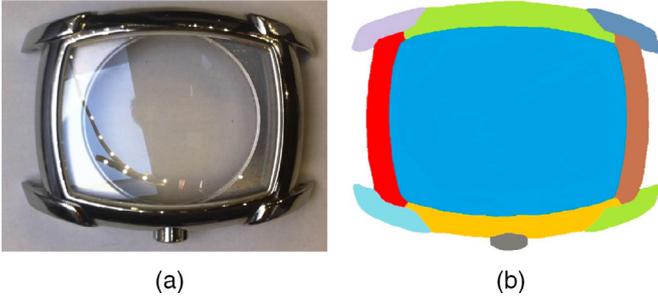


Fig. 1. Reflective object structure understanding. (a) Original image and (b) manually sub-segmented ground-truth image.

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 sequence, 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 $\Psi_\tau(x, y, t)$:

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

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

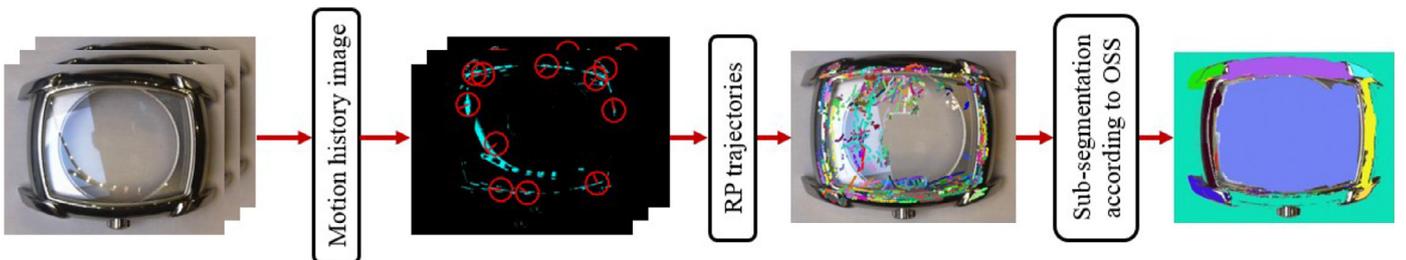


Fig. 2. Illustration of the proposed pipeline (see text for details).

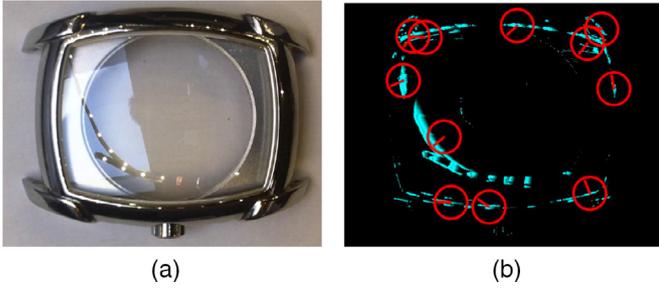


Fig. 3. (a) Original frame; and (b) motion history image of current frame. White pixels represent moving reflection particles. Red clocks represent moving directions of correspondent reflection particles. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

and Y directions yielding the spatial derivatives: $F_x(x, y)$ and $F_y(x, y)$, respectively. The gradient orientation (\varnothing) of the pixel is:

$$\varnothing = \arctan \frac{F_y(x, y)}{F_x(x, y)}. \quad (2)$$

Note that these gradient vectors will point orthogonally to moving object boundaries at each step in the *MHI*. It gives us a normal optical flow representation. After that, a downward stepping flood fill [24] is used to label motion regions connected to the current *MHI*. In this method, one puts a connected RP to be a family of neighbor pixels having similar motion direction. From the frame at time t , we extract the n moving RP (later denoted by C_i^t , $i \in [1 : n]$) as 8-connected pixels of the similar motion. From each C_i^t , a motion feature vector $f(C_i^t) = \{d_i^t, p_i^t\}$ is extracted, where d_i^t and p_i^t present the direction and the position of the RP, respectively. d_i^t is obtained by taking the average direction of all the pixels in C_i^t while p_i^t is the center of a bounding box that contains C_i^t . The motion features are used in the following section to match and track each RP in the image sequences.

3.1.2. Reflection particles matching

As the motion features $f(C_i^t)$ are extracted independently from each frame, the matching should be adapted to link the temporal information and to filter the impossible match. The matching of a reference particle feature $f(C_i^t)$ and a candidate particle feature $f(C_j^{t+\Delta t})$ needs to satisfy two following constraints:

$$err_p(i, j) = \sqrt{(p_i^t.x - p_j^{t+\Delta t}.x)^2 + (p_i^t.y - p_j^{t+\Delta t}.y)^2} < \delta, \quad (3)$$

$$err_d(i, j) = (d_i^t - d_j^{t+\Delta t})^2 < \alpha. \quad (4)$$

We set $\alpha = 20$ and $\delta = 10$ based on the experiment results. Eq. (3) shows the condition of particle position, where $err_p(i, j)$ is the position difference between C_i^t and $C_j^{t+\Delta t}$. Eq. (4) gives the condition of moving direction, where $err_d(i, j)$ is the direction difference between C_i^t and $C_j^{t+\Delta t}$. From a pair of matched features, a velocity feature $v_j^{t+\Delta t}$ is computed by:

$$v_i^{t+\Delta t} = \frac{\sqrt{(p_i^t.x - p_j^{t+\Delta t}.x)^2 + (p_i^t.y - p_j^{t+\Delta t}.y)^2}}{\Delta t}. \quad (5)$$

Then the updated motion feature of the reference particle is $f(C_i^{t+\Delta t}) = \{d_i^{t+\Delta t}, p_i^{t+\Delta t}, v_i^{t+\Delta t}\}$. The matching algorithm is illustrated as follows.

If no candidate particle features can be matched to reference particle feature, $f(C_i^{t+\Delta t})$ will be updated using the previous reference feature $f(C_i^t)$. On the other side, if there exist several candidate particle features which could be matched to $f(C_i^t)$, the C_i^t is computed as $\text{argmin}\{err_d(i, j)\}$.

Algorithm 1 Reflection particles matching.

Input: $f(C_i^t) = \{d_i^t, p_i^t\}$, $f(C_j^{t+\Delta t}) = \{d_j^{t+\Delta t}, p_j^{t+\Delta t}\}$.

Output: $f(C_i^{t+\Delta t})$.

do matching $f(C_i^t)$ and $f(C_j^{t+\Delta t})$ with Eqs. (3) and (4)

if matching is true **then**

1. compute $v_i^{t+\Delta t}$ by using Eq. (5)
2. update $f(C_i^t)$ to $f(C_i^{t+\Delta t})$
3. return $f(C_i^{t+\Delta t})$

else

1. $f(C_i^{t+\Delta t}) = f(C_i^t)$
2. return $f(C_i^{t+\Delta t})$



Fig. 4. Reflection moving trajectories. (a) 15 longest trajectories and (b) all the trajectories. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

3.1.3. Reflection particles tracking

The tracking of RP suffers from several problems: the high frequency of appearance and disappearance of the RP, the shape evolution of the RP, as well as multiple reference RP need to be tracked in the same time. Our tracker is composed by an iterative matching computation. The tracker is initialized for each detection, the state of a reference RP (C_i^t) is presented as $S(C_i^t) = \{p_i^t, d_i^t, v_i^t\}$. The state transition density is defined as follows:

$$p_i^t = p_i^{t-1} + v_i^{t-1} \times 1, \quad v_i^t = v_i^{t-\Delta t}. \quad (6)$$

The sampling processes a predictive circle window with the radius of δ and the center at the position predicted by Eq. (6). It is due to the RP motion features have already been extracted in each frame. Instead of sampling candidate RP (note as cc_j^t with its feature $f(cc_j^t) = \{dc_j^t, pc_j^t, vc_j^t\}$) with a weight which costs computational extremely expensive, a predictive sampling window is employed. Then each reference RP and candidate RP pair in the predictive window is scored by the difference of the moving direction:

$$err_d^c(C_i^t, cc_j^t) = (d_i^t - dc_j^t)^2, \quad (7)$$

and the $\text{argmin}\{err_d^c(C_i^t, cc_j^t)\}$ is computed to find the best match. Here we also present a threshold parameter β to break current reference RP tracking when the RP moving direction hugely changes. In our experiments, the value of β is set to 30. This tracking phase guarantees to keep all the associated RP on the same surface.

During tracking RP in frames, positions of all tracking results are saved as the moving trajectory. The trajectory of C_i is denoted as $T(C_i) = \{p_i^1, p_i^2, \dots, p_i^t\}$. One trajectory is considered as one label for a continuous surface on the object. As the RP could go through one surface in different directions, we save trajectories respectively for each direction. In this case, it ensures that one trajectory labels only one surface. On the other hand, some trajectories label the same surface. In Fig. 4, one color presents one trajectory of moving reflection;

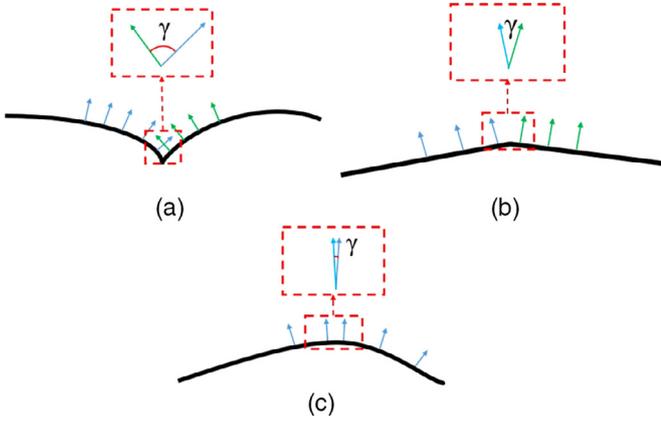


Fig. 5. (a, b) Discontinuous surfaces and (c) elementary continuous surface.

Fig. 4a contains 10 longest trajectories, Fig. 4b contains all the trajectories.

3.2. Continuous surfaces segmentation

In order to solve the problem of multi-labeled surfaces, an iterative surfaces segmentation is computed on the object based on RP moving trajectories. For convenience, we introduce a notation of an elementary continuous surface. It is defined according to the variation of γ of the object surface, γ being the difference between two neighbor normals. Then at each point of the surface, if the corresponding γ is below the threshold parameter ψ , the surface is considered as an elementary continuous surface, otherwise it is not. In fact, ψ denotes the limit of a surface normal variance which is not visible in the image. We put the distance between two neighbor points on the object surface equal to 1 mm. After experiments on various objects, ψ is set to be equal 2.2 degrees. As can be seen, Fig. 5a and b shows both discontinuous surfaces since their variations of γ are beyond the threshold parameter ψ , while Fig. 5c is an elementary continuous surface since γ is small.

Segmentation of the elementary continuous surfaces is to describe the surface structure of the object. As some trajectories are labeling the same surface, an iterative flood fill function is applied to merge the segmentation results of different trajectories on the same surface. The seeds which need to be flood filled are systematic sampled positions with a skip of 5 in the trajectory. Since the surface is elementary continuous, a trajectory can cover the surface regions with different brightness levels, the flood fill produces only one surface and the reflection does not produce additional subsegments. The flood fill method which we used during the segmentation is the same for the reflection particle detection. The pixel value $I(x, y)$ is considered to belong to the labeling domain if:

$$I(x', y') - d_l < I(x, y) < I(x', y') + d_h, \quad (8)$$

where d_l and d_h stand for maximum lower/upper brightness difference between the current observed pixel and one of its neighbors belonging to the surface, respectively. Algorithm of the segmentation process is illustrated as follows.

Since the trajectories do not have the same length and they may contain numerous positions, we order trajectories by increasing lengths and then systematic sample the positions by a skip of 5. Finally the flood fill is performed by starting from the sampled seeds in shorter trajectories to the sampled seeds in longer trajectories. In case of an elementary continuous surface, the segments containing shorter trajectories could be merged into other segments if there exists a suitable longer trajectory which covers all the segments. In this case, segments containing the seeds of shorter trajectories are relabeled according to the labeling of seeds of longer trajectory. As the

Algorithm 2 Segmentation process.

1. Trajectory sampling
 - (a) Sort trajectories by size in increasing order
 - (b) Systematic sampling of each trajectory with a skip of 5
 2. Segmentation
 - (a) Update filling color to the color of $T(C_i)$
 - (b) Flood fill all $p_i^t \in T(C_i)$ with current filling color
 3. Morphology component regrouping
 - (a) Update current filling color to the color of $T(C_j)$
 - (b) Regroup and fill all the components passed by $T(C_j)$ with current filling color ($i < j$)
 4. Final processing
 - (a) Fill holes which are surrounded by segmented regions with the surrounding color
-

reflection on the surface is highly variable, the segmentation phase might not cover the whole surfaces. In consequence, the final processing fills the holes which are surrounded by segmented regions with the surrounding color.

4. Results and evaluation

The experiments are conducted in using the camera with the resolution of 5 megapixels. An LED grow light is used to produce reflections on the object. Note that the light source is consisted by multiple light dots and it can be any shape, here we use a circle one. For the outdoor experiments, two projectors are used. The number of acquired frames depends on the complexity of the object surfaces and the number of light sources. In order to keep a reasonable number of acquired images, our LED grow light contains 30 light spots. The ground-truth images are manually labeled according to the 3D models of the objects which are obtained by a non-contact 3D digitizer VI-910.

4.1. Implementation and processing time

The method is implemented in Matlab and executed on a laptop with Inter(R) Core(TM)i7-360QM CPU@ 2.30 GHz. As previously mentioned, the processing time varies according to the complexity of the object surfaces and the number of light sources. The frame size in our experiment is 1280×720 . Regarding indoor experiments, as the light source is a LED grow light that contains 30 light spots, it yields a processing time of 0.8–0.95 s per image for the segmentation. About the outdoor experiments, as two projectors are used to produce the reflection, the tracking of reflections are much less complicated and less trajectories are saved. The processing time is 0.2–0.3 s per image for the segmentation. For both indoor and outdoor experiments, the objects which have less than 5 surfaces, the video length can be controlled in 5 s. On the other hand, for the objects which have about 10 surfaces, the video needs to be of 7–8 s.

4.2. Qualitative results

As the considered objects are reflective and/or transparent, the images contain many high-variability regions. Three of the comparison segmentation methods are graph based method [12,15]. They are based on k nearest neighbors, adjacent, and hierarchical graph, respectively. The graph-based methods are chosen since they have the ability to preserve detail in low-variability image regions while ignoring detail in high-variability regions. The fourth comparison method is EM segmentation [5]. It is a pixel clustering method in a joint feature space. It segments the image with the information from different aspects (color–texture–position). Over 20 objects have been processed,

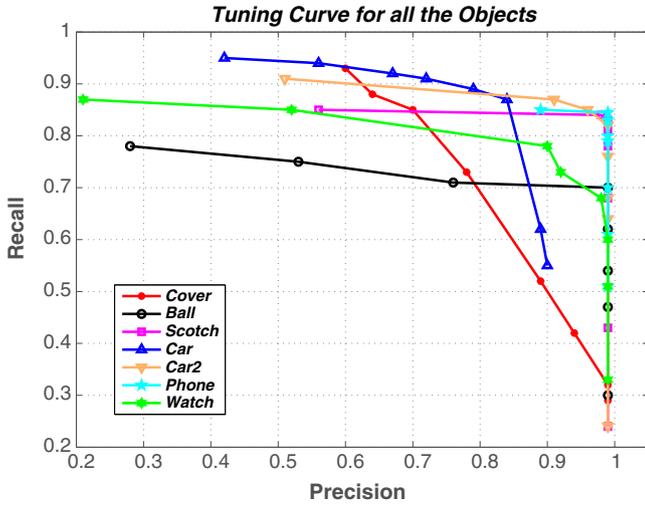


Fig. 6. ROC curve for the objects. All the curves were generated in using d_l from 1.5 to 5.5, d_h from 6.5 to 9.5. Each point corresponds to one combination of d_l and d_h . Objects with more subsurfaces have smoother curves.

7 of them are shown in Fig. 7. Due to the similarity of the three graph-based results and the lack of space, only KNN graph-based results are illustrated in Fig. 7. The objects Cover, Ball and Car2 have completely specular surfaces, the third object Scotch is transparent, and the other three objects contain both specular and transparent surfaces. The experiments for two cars are carried out outdoors. From the results, we can see that graph-based methods work reasonable in segmenting the object, but about the sub-segmentation of the object surfaces, it does not work meaningfully. EM segmentation preserves very well the contour of the objects but also the contour of the reflection that yields the poor sub-segmentation performance. Conspicuously, the results obtained by our method are more accurate. In consequence of a high sub-segmentation performance, the OSS is well presented.

4.3. Quantitative evaluation of our segmentation results

The purpose of our object surfaces segmentation is to understand the structure of the reflective objects. Therefore, to evaluate our proposed method, we manually labeled all the elementary continuous surfaces of the object to generate the ground-truth image as reference. Then we verify the segmentation performance with a pixel-wise evaluation.

4.3.1. Evaluation in details

To evaluate our proposed method in details, we calculate true positives (TP), false positives (FP), false negatives (FN), precision and recall for each surface, which are computed as follows:

$$TP = \frac{NTP}{PG}, FP = \frac{NFP}{PD}, \quad (9)$$

$$precision = \frac{NTP}{NTP + NFP}, \quad (10)$$

$$recall = \frac{NTP}{NTP + NFN}, \quad (11)$$

where NTP , NFP , NFN stand for the number of the true positive pixels, false positive pixels and false negative pixels, respectively; PD , PG , ND , NG stand for number of positives detected, number of positives in ground-truth mask, number of negatives detected and number of negatives in ground-truth. After computing precision and recall for each surface, a weighted combination of evaluations on each surface

is proposed to verify the entire performance for a whole object. The total pixel number N of the ground-truth object is computed as:

$$N = \sum_{i=1}^n PG(i), \quad (12)$$

where n is the number of surfaces. Then a weight w_i is defined by the percentage of the pixel number of current surface on that of the whole object, where i is surface index.

$$w_i = \frac{PD(i)}{N}. \quad (13)$$

With the weights of each surface, the precision ($precision_o$) and recall ($recall_o$) of the object can be computed as follows:

$$precision_o = \sum_{i=1}^n precision_i \times w_i; \quad (14)$$

$$recall_o = \sum_{i=1}^n recall_i \times w_i; \quad (15)$$

Then, we generate the receiver operating characteristic (ROC curves) for objects in the experiment by varying the parameters d_l and d_h of the flood fill method. We use 5 different values for $d_l \in [1.5, 2.5, 3.5, 4.5, 5.5]$ and 3 different values for $d_h \in [6.5, 7.5, 8.5]$. From the ROC curves, we can see that for Scotch, Ball and Phone, the precision values keep very high at the beginning and suddenly go down during the raising of recall values (Fig. 6). This is due to the fact that these objects all have two surfaces. Within the change of parameters of flood fill method, the labeling color of one surface overfills the other surface. Then the sudden overfilling makes precision value suddenly drop down. For the other objects, as they have approximately ten surfaces, the curves are more smooth. For all the indoor experiments (except one of the car), the precision values reach 0.99 and recall values are more than 78. For the outdoor experiments on the cars, under a natural environment without controlling illumination condition except our light source, the precision values reach 0.99 and the recall values are more than 0.88. These results illustrate the robustness of our segmentation method in OSS understanding under different experiment conditions and of various objects.

4.3.2. Comparison with other works

To our best knowledge, no segmentation method is designed for dealing with reflective objects; thus it is not a trivial task to compare with other methods. Among existing methods, graph-based and region based segmentation methods are most likely to treat the case of reflective surfaces (Fig. 7). Three graph based (KNN, adjacent, and hierarchical) methods and one region based (EM) method are chosen for the comparison. We would like to point out that only the proposed method and hierarchical graph-based method [15] take advantage of temporal information while the other two methods use static data. We did not compare with contour based methods since in this case, reflections would produce false true negative contours which lead to a poor segmentation. To evaluate the segmentation performance, we employ f -score as criterion which is a harmonic mean of precision and recall. It is computed as:

$$f\text{-score} = 2 \times \frac{precision \times recall}{precision + recall}. \quad (16)$$

Therefore, we choose f -score as the criterion of segmentation performance evaluation in order to compare our proposed method with the state-of-the-art approaches. In Table 1, we compare our proposed method to 4 well known segmentation methods. We can see that the f -score of object 'Cover' is 0.76, which is much lower than for the other objects computed by all the methods. This is due to the fact that the surfaces of this object are concave, moving reflection vanish extremely quick even though the surfaces are smooth and moving trajectories are split into smaller trajectories. On the other hand,

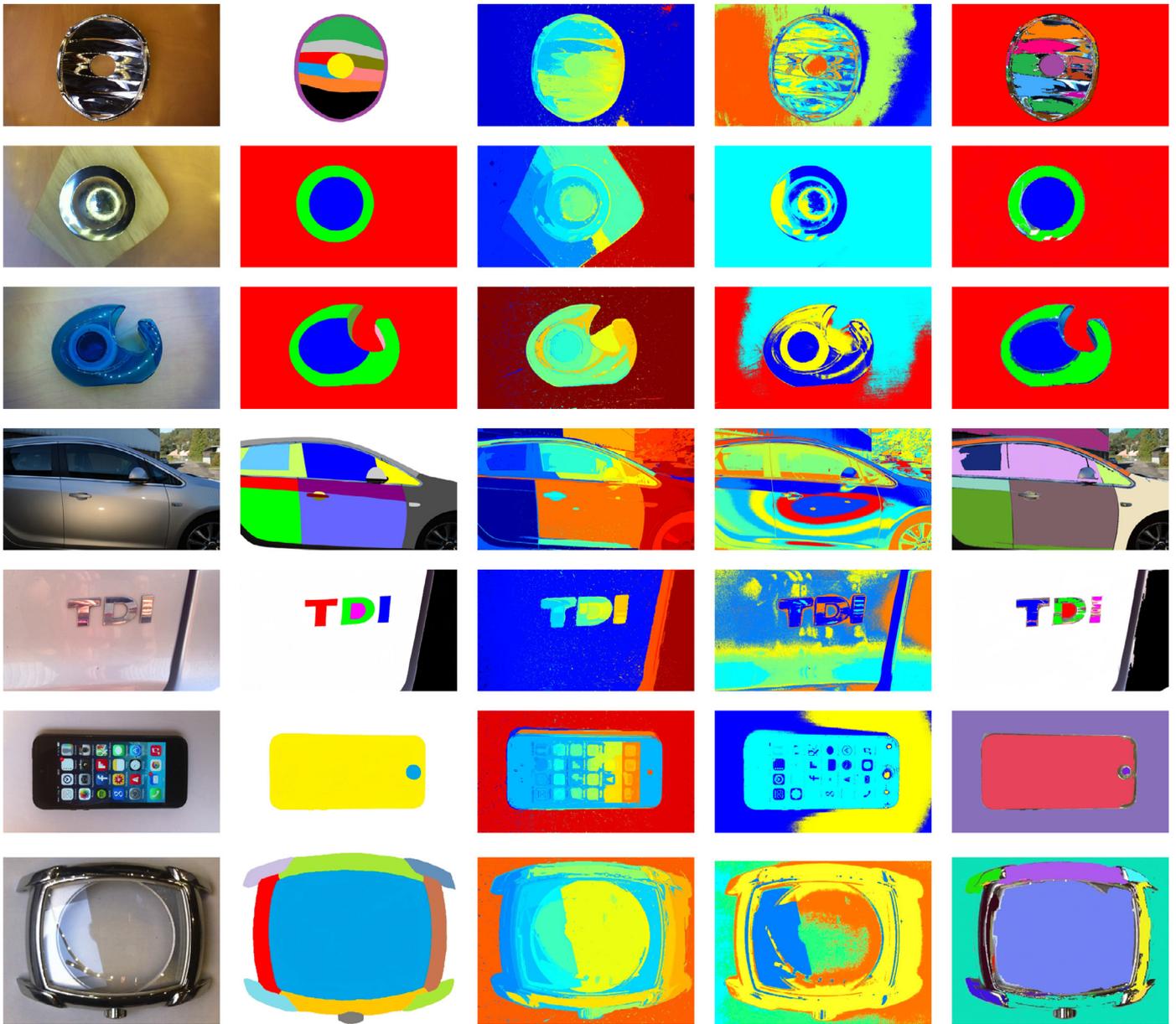


Fig. 7. First column: original images. Second column: ground-truth segmentation. Third column: k nearest neighborhood graph-based segmentation [12]. Forth column: EM segmentation [5]. Last column: Segmentation by our proposed method based on reflection motion estimation. (better see in color). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Best f -score of the objects.

f -score	Cover	Ball	Scotch	Car	Car2	Phone	Watch
KNN graph [12]	0.56	0.38	0.48	0.73	0.83	0.51	0.74
Ajdacent graph [12]	0.48	0.34	0.54	0.66	0.79	0.48	0.75
EM [5]	0.17	0.41	0.46	0.54	0.27	0.79	0.47
Hierarchical graph [15]	0.46	0.32	0.39	0.72	0.81	0.43	0.44
Our method	0.76	0.84	0.89	0.86	0.93	0.91	0.84

f -score of ‘Ball’ is also only 0.84 because of the presence of high intensity variations in small regions. In the experiment of object ‘Phone’, despite the fact that intensity variations are important on the whole object, it is not the case for small sub-regions. Thus, the final processing of our method can fill the holes and yields the value of f -score to 0.91. As for the two outdoor experiments, both provide meaningful results. The f -score of Car2 reaches 0.93 which means high rate in both precision and recall. We would like to emphasize that, in

dealing with reflective and transparent objects, our method outperforms significantly (at least 9% higher) the state-of-the-art methods.

5. Conclusion and perspectives

We have presented a segmentation method based on reflection motion features in order to deal with reflective and transparent objects. Due to a simple constraints (object and camera are fixed), our

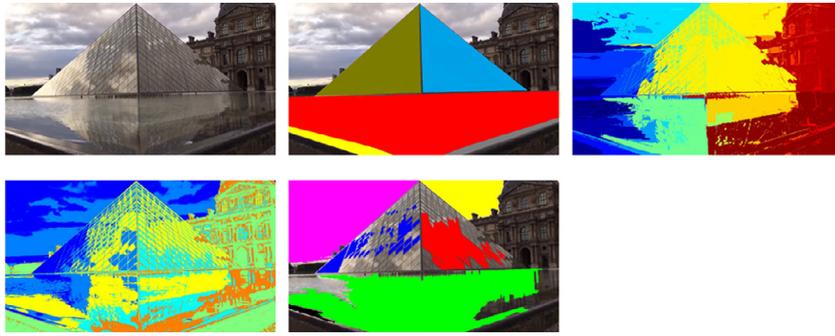


Fig. 8. Future work: object segmentation by employing fully natural light source.

method can be widely used in the industry for object recognition and retrieving. More importantly, instead of removing and reducing reflections, taking its advantage is pioneering work in a new direction. The results show that the reflection motion features can be used as a robust signature for labeling continuous surfaces on reflective and transparent objects. In comparison with conventional segmentation approaches, our method can overcome the difficulties produced in case of reflective and transparent objects and leads to higher performances in terms of accuracy and robustness. This efficiency has been proved through multiple experiments over various objects and under different type of illumination conditions (indoor and outdoor). This series of test highlight the advantage given by our approach against the state-of-the-art methods.

Regarding future work, we intend to use nature illumination source for the object segmentation. An example is shown in Fig. 8, where no man-made light source is used. It is a time lapse video of 10 s which requires 4 h of image acquisition. The reflection motion of the cloud is used to perform a surface segmentation. However, the faces of the pyramid do not satisfy our constraint of the elementary continuous surface, thus they are not detected as entire surfaces. One of possible directions of future research will be to adapt our method to such surfaces. We are also interested in exploring the evolution of reflection shape.

Acknowledgments

This work is supported by the company ALITHEON and French National Research and Technology Association (ANRT under contract CIFRE : 0888/2012). We also thank anonymous reviewers for valuable comments which helped us to improve the paper. We are extremely grateful to members of laboratory Le2i for day to day help and collaboration.

References

- [1] M.A. Ahad, J.K. Tan, H. Kim, S. Ishikawa, Motion history image: its variants and applications, *Mach. Vis. Appl.* 23 (2012) 255–281.
- [2] B. Barrois, C. Wohler, 3D pose estimation based on multiple monocular cues, in: *Proceedings of the BenCOS Workshop, 2007*, pp. 1–7.
- [3] F. Bergamasco, A. Albarelli, A. Torsello, M. Favaro, P. Zanuttigh, Pairwise similarities for scene segmentation combining color and depth data, in: *Proceedings of the International Conference on Pattern Recognition, ICPR, 2012*.
- [4] L. Bourdev, J. Malik, Body part detectors trained using 3D human pose annotations, in: *Proceedings of the International Conference on Computer Vision, ICCV, 2009*.
- [5] C. Carson, S. Belongie, H. Greenspan, J. Malik, Blobworld: image segmentation using expectation-maximization and its application to image querying, *IEEE Trans. Pattern Anal. Mach. Intell.* 24 (1999) 1026–1038.
- [6] J. Davis, Hierarchical motion history images for recognizing human motion, in: *Proceedings of the IEEE Workshop DREV, 2001*.
- [7] A. Delpozzi, S. Savarese, Detecting specular surfaces on natural images, in: *Proceedings of the Conference on Computer Vision and Pattern Recognition, CVPR, 2007*.
- [8] J. Deng, L. Feifei, Fine-grained crowdsourcing for fine-grained recognition, in: *Proceedings of the Conference on Computer Vision and Pattern Recognition, CVPR, 2013*.
- [9] M. D'Zmura, P. Lennie, Mechanism of color constancy, *J. Opt. Soc. Am.* 3 (1986) 1162–1172.
- [10] A. Farhadi, I. Endres, D. Hoiem, D. Forsyth, Describing objects by their attributes, in: *Proceedings of the Conference on Computer Vision and Pattern Recognition, CVPR, 2009*.
- [11] P. Felzenszwalb, R. Girshick, D. McAllester, D. Ramanan, Object detection with discriminatively trained part based models, *IEEE Trans. Pattern Anal. Mach. Intell.* 32 (9) (2010) 1627–1645.
- [12] P. Felzenszwalb, D. Huttenlocher, Efficient graph-based image segmentation, *Int. J. Comput. Vis.* 59 (2) (2004) 167–181.
- [13] V. Ferrari, A. Zisserman, Learning visual attributes, *Adv. Neural Inf. Process. Syst.* (2007).
- [14] M. Grundmann, V. Kwatra, M. Han, I. Essa, Efficient hierarchical graph-based segmentation of RBGD videos, in: *Proceedings of the Conference on Computer Vision and Pattern Recognition, CVPR, 2010*.
- [15] M. Grundmann, V. Kwatra, M. Han, I. Essa, Efficient hierarchical graph-based video segmentation, in: *Proceedings of the Conference on Computer Vision and Pattern Recognition, CVPR, 2010*.
- [16] X. Guo, X. Chao, Y. Ma, Robust separation of reflection from multiple images, in: *Proceedings of the Conference on Computer Vision and Pattern Recognition, CVPR, 2014*.
- [17] S. Paris, F. Durand, A topological approach to hierarchical segmentation using mean shift, in: *Proceedings of the Conference on Computer Vision and Pattern Recognition, CVPR, 2007*.
- [18] S. Savarese, P. Perona, Local analysis for 3D reconstruction of specular surfaces, in: *Proceedings of the Conference on Computer Vision and Pattern Recognition, CVPR, 2001*.
- [19] S. Savarese, P. Perona, Local analysis for 3D reconstruction of specular surfaces part II, in: *Proceedings of the Conference on Computer Vision and Pattern Recognition, CVPR, 2002*.
- [20] J. Shi, J. Malik, Normalized cuts and image segmentation, *IEEE Trans. Pattern Anal. Mach. Intell.* 22 (2000) 888–905.
- [21] K.J.F. de Souza, A. de Albuquerque Arajo, Z.K. do Patrocio Jr., S.J.F. Guimarães, Graph-based hierarchical video segmentation based on a simple dissimilarity measure, *Pattern Recognit. Lett.* 47 (2014) 85–92.
- [22] R. Tan, K. Ikeuchi, Reflection components decomposition of textured surfaces using linear basis functions, in: *Proceedings of the Conference on Computer Vision and Pattern Recognition, CVPR, 2005*.
- [23] R. Tan, K. Ikeuchi, Separating reflection components of textured surfaces using a single image, *IEEE Trans. Pattern Anal. Mach. Intell.* 25 (2005a) 178–193.
- [24] A. Treuenfels, An efficient flood visit algorithm, *C/C++ Users J.* 12 (8) (1994) 39–62.
- [25] A. Vedaldi, S. Mahendran, S. Tsogkas, S. Maji, B. Girshick, J. Kannala, E. Rahtu, I. Kokkinos, M.B. Blaschko, D. Weiss, B. Taskar, K. Simonyan, N. Saphra, S. Mohamed, Understanding objects in detail with fine-gained attributes, in: *Proceedings of the Conference on Computer Vision and Pattern Recognition, CVPR, 2014*.
- [26] C. Xu, J. Corso, Evaluation of super-voxel methods for early video processing, in: *Proceedings of the Conference on Computer Vision and Pattern Recognition, CVPR, 2012*.