Light Field Distortion Feature for Transparent Object Recognition

Kazuki Maeno, Hajime Nagahara, Atsushi Shimada, Rin-ichiro Taniguchi
Graduate School of Information Science and Electrical Engineering, Kyushu University
744, Motooka, Nishi-ku, Fukuoka, 819-0395 Japan

Abstract

Current object-recognition algorithms use local features, such as scale-invariant feature transform (SIFT) and speeded-up robust features (SURF), for visually learning to recognize objects. These approaches though cannot apply to transparent objects made of glass or plastic, as such objects take on the visual features of background objects, and the appearance of such objects dramatically varies with changes in scene background. Indeed, in transmitting light, transparent objects have the unique characteristic of distorting the background by refraction. In this paper, we use a single-shot light field image as an input and model the distortion of the light field caused by the refractive property of a transparent object. We propose a new feature, called the light field distortion (LFD) feature, for identifying a transparent object. The proposal incorporates this LFD feature into the bag-of-features approach for recognizing transparent objects. We evaluated its performance in laboratory and real settings.

1. Introduction

Visual object recognition is important to robotics and computer vision applications. Statistical learning methods such as bag-of-features (BoF) are currently a very active area of research for image annotation and object recognition. These methods commonly use local features, such as scale-invariant feature transform (SIFT) and speeded-up robust features (SURF), for visual object recognition. However, these features and learning algorithms cannot apply to transparent objects. Our daily environments, kitchen, living room, and office, are filled with many transparent objects, such as glasses, bottles, bowls, jars, vases, and windows, to name a few. Depending on backgrounds, whether stationary or moving, the appearance of a transparent object drastically changes, as such objects do not have features entirely of their own, but rather transmitted background images. Using the standard approaches requires modeling local features not of the transparent object but the background. In this paper, we propose a novel object feature, called the light field distortion (LFD) feature. We discuss how the LFD feature models and visually recognizes transparent objects, which to date have been ignored as exceptions in applications of visual object recognition or annotation.

Transparent objects are made of refractive materials, such as glass or plastics, and distort rays emanating from the scene background. Different objects produce different distortions, each carrying intrinsic characteristics of the transparent object, namely the refractive index of material and the object's shape, both of which influence the distortion. The LFD feature is calculated from a light field image, a single shot captured by a light field camera, and describes this distortion of the field caused by the refraction in transparent object. Compared with conventional cameras, which capture 2D photos from a single perspective, light field cameras obtain richer 4D images that include multiple 2D viewpoints (position resolution) as well as standard 2D image coordinates (angular resolutions). The LFD feature models the distortion from differences in corresponding points between viewpoints including the 4D light field, whereas common features, such as gradients or edges, model the appearance. This is an entirely original concept for feature description with the advantage that LFD is less affected by background changes, as it uses patterns of ray distortions caused by the object, not patterns from the object's appearance.

The light field camera was originally proposed for image-based rendering for use in the graphics community, and has been used for a variety of different visualization applications, such as generating free-view images, 3D graphics, and digital refocusing. Early light field cameras, e.g. Stanford multi-camera array [1], were huge and quite expensive systems. However, the latest light field cameras consisting of a micro-lens array between the sensor and main lens are becoming inexpensive and compact [2, 3], some of these are available in the product market [4, 5, 6]. Hence, we believe that the light field camera is becoming a popular input device in computer vision applications.

The contribution of this paper is 1) in tackling a difficult computer vision problem, transparent object recognition with a single-shot image, 2) in proposing a new feature

1063-6919/13 $26.00 © 2013 IEEE
DOI 10.1109/CVPR.2013.359

for transparent object recognition, called the LFD feature, and 3) in applying a light field image to recognition application. We implemented our method based on the BoF approach and performed laboratory and real experiments, using seven objects and different backgrounds, to assess the effectiveness of the LFD features. Our result shows that the LFD feature classified transparent objects without explicit physics-based refraction analysis and refraction models.

2. Related Work

In recent years, the BoF-based approach has been attracting much attention in object recognition research. Local features such as SIFT are widely used owing to their invariance to scaling, rotation and illumination [7, 8, 9]. Local features are divided into several clusters and a representative feature in each cluster is assigned by vector quantization. Objects in the same category are expected to have similar frequency within this representative feature. This approach implicitly assumes that the majority of local features are extracted from an object's surface rather than the background. Therefore, if local features are drawn from a more dominant background than an object's surface, existing learning and recognition methods perform poorly. A transparent object yields less information about its appearance. Its actual appearance depends largely on the visible background as viewed through the object. In consequence, extracting scene-independent local features from a transparent object area is difficult.

Thus, these approaches find local transparent structure by applying a latent factor model before quantizing into a visual word representation [10]. Although such approaches recognize a transparent object without any knowledge of background scenes at test time, the learning step requires many training images in which the transparent object is captured under various environments.

In other directions, there has been much research on measuring refraction responses in transparent objects using cameras to obtain physical parameters, such as surface curvature or refractive index. It is well known that refraction polarizes light. Miyazaki et al. measured light intensities from transparent objects through polarizing filters [11, 12]. Schlieren photography [13, 14] has also been used for fluid, gas flows, and shock wave analysis. This method visualizes the refraction response in a scene as a gray-scale or color image by using special optics, although it requires high-quality optics and precise alignment. Hence, its applicability is restricted to laboratory environments, and not for common practical use. Wetzstein et al. [15] proposed light-field background-oriented Schlieren photography that obtains Schlieren photos using a common hand-held camera and a special-purpose optical sheet. Although this technique recovers the transparent surface [16], it also has restricted practical use as the special sheet is always required as a background object.

Figure 1. Background distortion from different objects

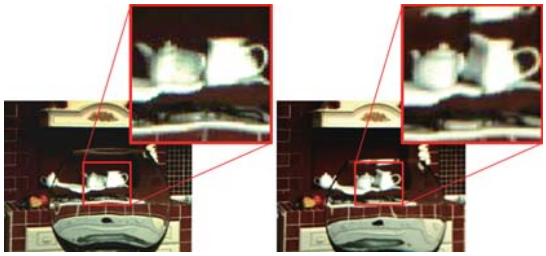


Figure 2. Background distortion from changing viewpoints

Approaches, similar to our own, obtain shape from optical flow caused by refraction. In particular, Ben-Ezra et al. [17] proposed a model-based method to recover shape and pose from video taken with known camera motions. Similarly, Aagrwal et al. [18] recovered shape from video acquired while the background behind the object moves. Morris et al. [19] used two calibrated cameras to estimate the refractive indices over time-varying liquid surfaces from distortions of known grid patterns at the bottom of a tank. In contrast to these approaches, the novelty of our work is to apply refraction to transparent object recognition, realized from a single shot image, using a light field camera as an input device. Unlike previous methods, there are no constraints on background texture, camera motion or known parameters.

3. Light Field Distortion Feature

By refraction, a transparent object deforms the background scene. Different objects produce different images of the same scene (Figure 1), because refraction by objects is affected by shape and refractive index. Using the background distortion caused by refraction is our means to recognize transparent objects. In fact, we modeled the background distortion to the appearance difference from different perspectives (Figure 2). In theory, the modeled distortion itself is independent of background texture, although the background determines image appearance, the distortion for corresponding points from different viewpoints is maintained. Therefore, our proposal is to model the object's refraction as a distortion of multiple viewpoints cap-

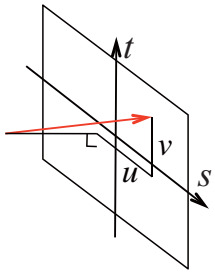
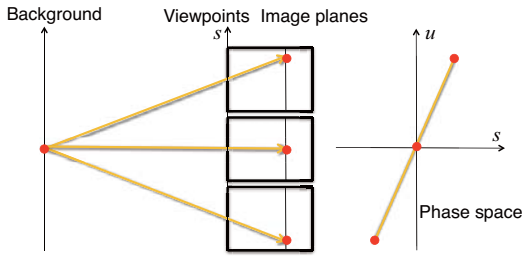
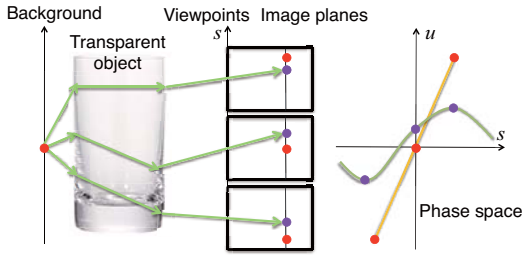


Figure 3. Definition of light field representation $L(s,t,u,v)$.



(a) Undistorted light field



(b) Distorted light field

Figure 4. Light field propagation

tured by the light field camera. In this section, we define the LFD feature and outline its use in transparent object recognition. The light field is a function that describes the amount of light emitting in every direction from every point in a scene. Conventional cameras only record that part of the light field passing through a single viewpoint of a 2D image. In contrast, a light field camera obtains a 4D light field image which has multiple viewpoints. There are various representations of the light field. Here, we use the 4D-ray representation of the light field $L(s,t,u,v)$ determined by the intersection of a plane (s,t) and a slant of ray (u,v) (see Figure 3). Figure 4(a) illustrates the functioning of a camera array and shows the relation between light field and phase space representations. Figure 4(a) shows only a 2D slice of the light field and phase space for ease in understanding.

Figure 4(a) depicts a scene where there is no object between background and camera; i.e., light propagates in free-space with no refraction, reflection, scattering, or absorption. As illustrated, if rays emitted from a point in the background are straight, the observed light field has constant disparities over the images for the different viewpoints. The rays from the same point are distributed on a line in the su-phase space (Figure 4(a)), and the slope of the line depends upon the distance between camera and background. In fact, these rays are distributed on a hyperplane in $stuv$-space because the actual light field and phase space is a 4D function. In contrast, if a transparent object intervenes between background and camera, the ray distribution deviates from the line or the hyperplane (Figure 4(b)). This LFD is caused by refraction occurring within the object, which is characterized by the material (refractive index) and shape of the transparent object. We call this the LDF feature that is to be used as a feature in transparent object recognition.

Here, we denote an arbitrary point in the image taken from the center of the viewpoints $(0,0)$ as $p_{0,0} = (u,v)$ and the corresponding point in the image taken from another viewpoint (s,t) as $p_{s,t} = (u',v')$. Actually, we denote the i-th feature point $(i = 0,\ldots,N-1)$ and the corresponding points as $p_{0,0}(i)$ and $p_{s,t}(i)$ respectively, with the image having N feature points. To make the LFD feature independent of the position of the point (u,v), we use relative differences defined by the following expression,

$$\Delta p_{s,t}(i) = p_{s,t}(i) - p_{0,0}(i), (s,t) \notin (0,0). \quad (1)$$

Finally, the i-th LFD feature is defined as the set of relative differences,

$$LFD(i) = \{\Delta p_{s,t}(i) \mid -m \leq s \leq m, -n \leq t \leq n\}, \quad (2)$$

where $2m+1$ and $2n+1$ are the numbers of viewpoints.

4. Transparent object recognition

In this section, we describe an algorithm of our transparent object recognition. Figure 5 shows the overview of the algorithm.

We used a commercial light field camera, Pro Fusion 25(ViewPlus Inc.). This camera system has 25 VGA resolution (640×480 pixels) cameras and can simultaneously capture images from 25 viewpoints (5 horizontal viewpoints × 5 vertical viewpoints). We transformed the 25 captured images to a rectified light field image (s,t,u,v) as shown in Figure 6 by Xu's calibration method [21].

In the LFD feature acquisition, we obtain correspondences between the image of the center view and those of the other viewpoints. A disparity $\Delta p_{s,t}(i)$ can be calculated from $p_{0,0}(i)$ and its corresponding point $p_{s,t}(i)$. An LFD feature is composed of these disparities, which are represented by equation 2. Figure 7-top shows examples of the correspondences between the center of three views. We describes the 2D disparity vectors to color representations, in which hue and saturation indicate direction and length of the vector respectively as shown in Figure 7-bottom, and

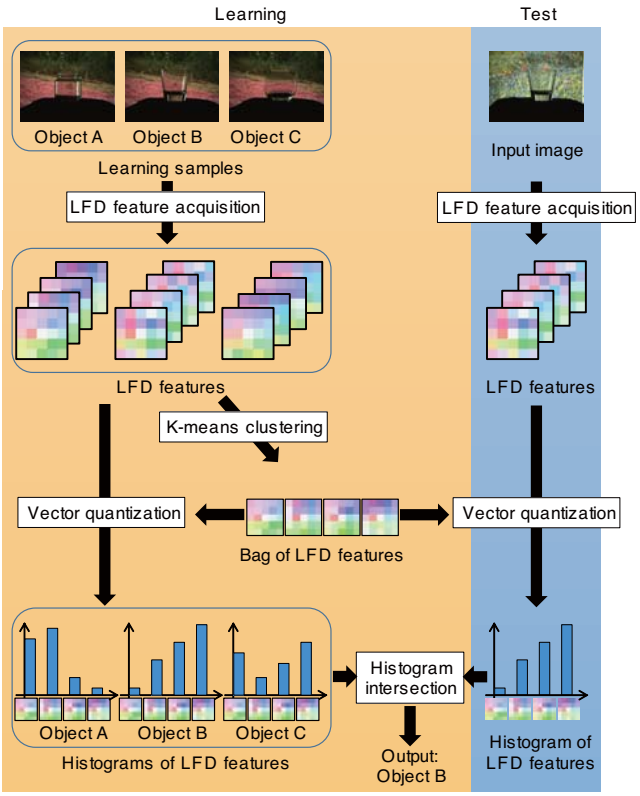


Figure 5. Overview of algorithm

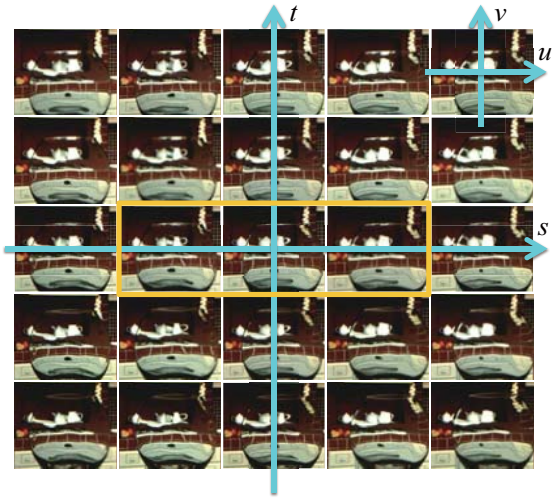


Figure 6. Light field image

we uses the colors for indicating LFD patterns in this paper. We estimated the disparities between the center and the other 24 viewpoints by the optical flow method proposed by Farneback et al. [20]. The LFD feature is obtained by a 48-dimensional vector which describes the disparities between the center view and the other views. The LFD features are extracted at pixel by pixel in an image. The LFD features

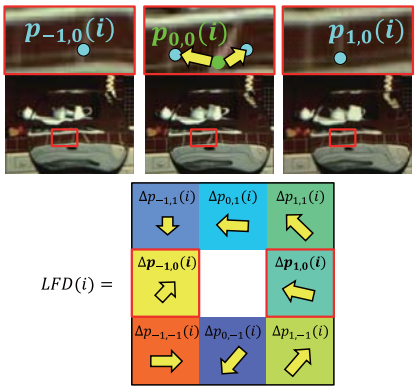


Figure 7. LFD feature and corresponding points. This is an enlargement of the central images of figure 6. The LFD is also an example of 3×3 case; these images are actually taken from a 25-viewpoint light field image. Hue and saturation of the color represent direction and length of the $\Delta p_{s,t}$ vectors.

coming from the transparent object has a larger distortion than these from background, since the disparities contain refraction effect and deviate from hyperplane assumed as Lambertian reflection in phase space as described in Figure 4. We filtered out the background LFD features by thresholding of the deviation. As a result, we can obtain a set of N' LFD features from the single light field image.

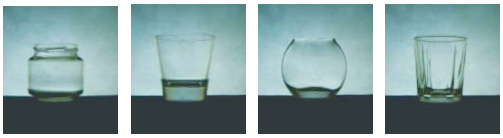
Learning and recognition processes are performed by a typical BoF approach. We use the LFD features as visual words. In the training phase, the LFD features are quantized by k-means clustering for obtaining visual words. We represent classes of transparent objects as patterns of histograms of the visual words. In the testing phase, we extracted LFD features from input image as a similar manner, and calculate the similarities of the distances by histogram matching for classification. Finally, we determine the class of the object as a minimum distance of the matching.

5. Experiment

5.1. Assumption

We evaluated our proposed method by classification of transparent objects in a laboratory setting and real environments under following assumptions;

- There is one transparent object as a recognition target in a scene.
- The target object appears in all of the viewpoints of the light field camera.
- Relative positions and poses of the camera and target object are almost same between a training and testings.
- Background is reasonably far away from the object.
- Background scenes have sufficient texture.



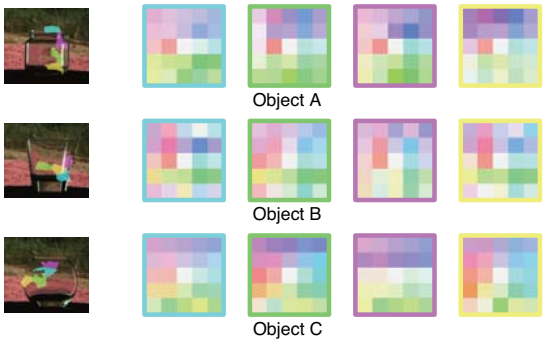
(a) Object A (b) Object B (c) Object C (d) Object D

(e) Object E (f) Object F (g) Object G

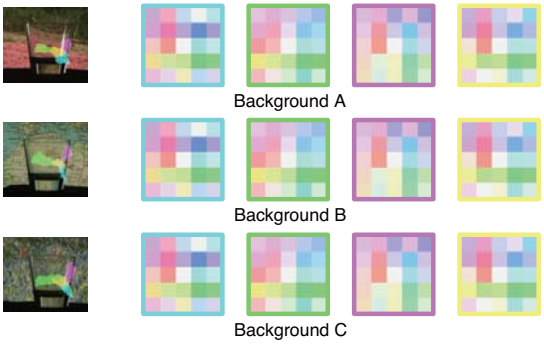
Figure 8. Target transparent objects

(a) Background A (b) Background B (c) Background C

Figure 9. Three of the five background patterns



(a) Different objects with same background



(b) Same object with different backgrounds

Figure 10. Examples of primal LFD features by color representation. Each row shows the different object or different background. The 1st column shows the objects and regions of the pixels where the primal LFD come form. The 2-5 columns indicate frequent LFDs describing the objects. The colors of the bounding box of the LFDs are corresponding to that of the regions in the 1st column. Hue and saturation of the LFDs represent direction and length of the Δp vectors on 5×5 viewpoints as similar to Figure 7.

We performed some experiments in a laboratory and real setting to evaluate robustness and limitations of our proposed method.

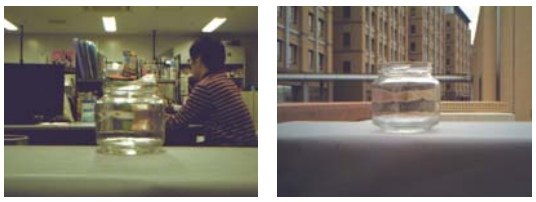
5.2. Experimental Results

We performed some recognition experiments in a laboratory setting. We used five printed images for a backdrop of a scene, three of them are depicted as examples in Figure 9. We used as a reference position for LFD learning a setting where camera position was 40 cm in front and background was 150 cm behind the object position. The optimal numbers of clusters K for the BoF approach was determined based on the preliminary experiments and set $K = 500$ for the LFD feature. Our task is classifying seven various shapes of the objects (Figure 8) into the seven classes under the various background textures. We calculated avarage recognition ratio amaung the 7 objects using leave-one-out cross validation for scenes, one scene is used for training and the other for testing.

Figure 10 shows difference of the 4 of 500 ($K = 500$) frequent visual words as primal LFD features described by color representation. Figure 10(a) shows the frequent LFD features obtained by different objects with the same background. The patterns of the LFD features are different for the different objects, despite these objects looking similar visually and being placed in front of the same background. Also the LFDs came from the different regions of the objects. It means that each object was uniquely modeled by the LFDs. Figure 10(a) shows our methods utilized not only silhouette LFD features but also LFD features inside region of the object.

In contrast, Figure 10(b) shows the LFDs from the same object in front of the different backgrounds. It shows that these LFD patterns are the similar and coming from similar regions of the object, although the visual appearance so different among the background differences. We confirmed that LFD feature is irrespective of the background difference, since the feature models not intensity pattern but geometrical distortion cased by object refraction. As a result, the proposed method achieved 80% of average classification accuracy over the 7 object classes in front of 5 different backgrounds, although it realized transparent object recognition from a single-shot image.

We also performed real experiments in indoor and outdoor settings (Figure 11). Objects were placed 40 cm from



(a) Office (b) Outside

Figure 11. Examples of the scenes for real experiments

Table 1. Recognition ratios for real experiment

	racognition ratio
Proposed LFD feature	0.714
Standard SIFT	0.119

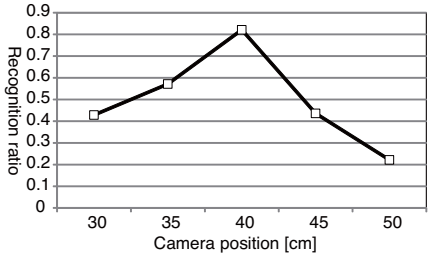


Figure 12. Recognition ratios for camera position changes

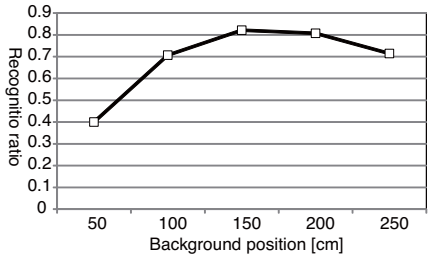


Figure 13. Recognition ratios for background position changes

the camera against real backgrounds of structures at various depths, i.e., distances sufficiently far (more than 1m) from the objects. Table 1 shows recognition ratios by leave-one-out cross validations for three different scenes. We also used a similar recognition method in using the SIFT feature. Numbers of clusters K for the SIFT approach was set $K = 100$. Using SIFT, these recognition ratios are fairly low. This result shows that local features are unsuitable for transparent object recognition. The proposed method using the LFD feature achieved an average 70% accuracy (Table 1), while standard SIFT was not working at all.

5.3. Limitation Analysis

We evaluated and discuss limitations of our proposed method in this section. We used the same conditions to the experiments in a laboratory setting as described in Sec. 5.2. We changed camera and background positions, object poses and lighting conditions for the evaluation. Figures

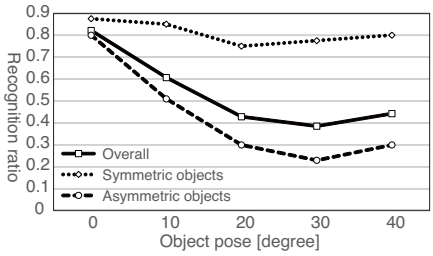


Figure 14. Recognition ratios for object pose changes

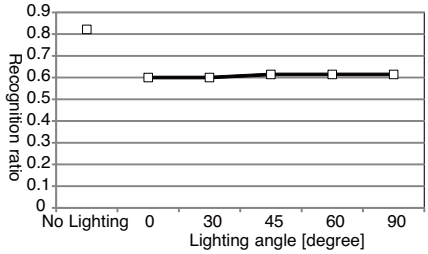


Figure 15. Recognition ratios for illumination changes

12-15 show decreases of the recognition accuracies by these changes from the reference setting on the learning step.

We moved the camera over a range ± 10 cm from the reference position 40 cm. Figure 12 shows that the recognition ratios are decreased when the displacement form the reference position is increased, because the LFD feature are changed from the learned pattern related to the distance between the camera and object. We would consider the margins for object deviation to be about 7.5 cm if we accept 20% decrease in the recognition ratio.

We also moved the background position over the range of 50 cm to 250 cm from the object, while the reference position of the background is 150 cm. Figure 13 shows the recognition ratio decreased when the background displaced from the reference position. The decrease was moderate in compared with the changes of the object position as shown in figure 12. The ratio is not so changed when the background is away from the object, while it is steeply decreased when the background position is approaching to the object. This is because that the LFD difference cased by the depth disparity is nonlinearly occurred, near position is larger and far is small. The background position did not affect much about the recognition ratio and we can apply this method to more realistic no planer scene background, if we can assume that the background objects of the scene are places reasonably far positions, e.g. more than 100 cm.

We rotated objects up to 40 degree about their central axes for evaluating the effect of the pose change. Figure 14 shows the results splitting to a symmetric group (object A, B) and an asymmetric group (object C-G) of the objects, as well as overall ratio. As we expected, the ratio of the sym-

(a) Uniform background

(b) False in uniform pattern

Figure 16. Falsely detected feature points

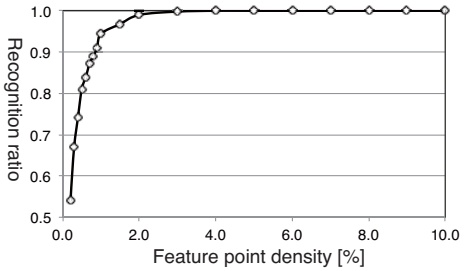


Figure 17. Recognition ratio vs. density of feature points

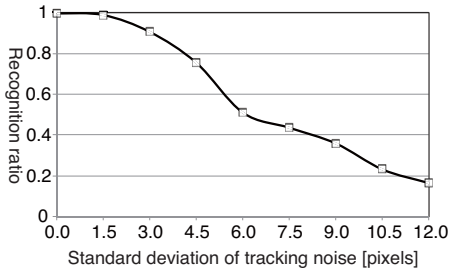


Figure 18. Recognition ratio vs. tracking noise.

metric group was invariant to poses, since the shape and its LFDs would be not changed. The ratio of asymmetric group was decreased gradually and the limitation on object pose variation is within 10 degree if we accept a 20% degradation in recognition ratios.

We also evaluated effects of illumination change for recognition. We placed an additional point light source to the global illumination that was used in the all of the experiments. We changed the direction of the light source from above (0 degree) to the side (90 degree) with respect to the target object. There were inter-reflections and specular reflections from the light source and these effects were changed as we moved the light source. Figure 15 shows the recognition ratios across the lighting directions. The left most label indicate the recognition ratio without lighting which is same condition of learning setting. This figure shows that the internal and specular from the light source contaminated the LFDs and decreases averagely 20% of the recognition. It is not related to the directions of the settings.

5.4. Analysis for Texture Density

The background patterns used in the experiment have complex textures (see Figure 9) from which correspondence detection can be easily performed. Meanwhile, LFD features were not appropriately extracted in certain background scenes (Figure 16). Because textural information is minimal, correspondences between cameras were difficult to find. In Figure 16(a), the LFD features were extracted from only the edges of the transparent object, with no LFD feature taken interior to the object. For another background (Figure 16(b)), LFD features were wrongly extracted exterior to the transparent object (see the top-left part of the figure). Therefore, the performance is affected by the accuracy in correspondence detection.

We evaluated how many LFD feature point is needed to accurate recognition in simulation. First, to obtain ideal feature points, a dot pattern was displayed as a background to the transparent object for easy to detect the correspondence of the LFDs. A total of N feature points were captured; note that the number of N corresponds to a whole number of pixels. Second, a percentage $d\%$ of LFD features were randomly selected, then leave-one-out cross-validation was performed to acquire the recognition accuracy. This procedure was repeated 100 times if d was less than 1%, otherwise, just ten times. The recognition accuracy curve is plotted in Figure 17. This figure shows that our approach requires at least 3% of the LFD features to obtain almost 100% recognition accuracy. In terms of practical uses, extracting LFD features for at least 3% of the image size is not such a difficult problem. Therefore, our proposed LFD feature is considered effective in transparent object recognition.

We also evaluated mistracking for estimating the LFD vectors. We used the same simulated features above and randomly selected 3% of the features. We added zero-mean Gaussian noise with different standard deviations to the LFD features as a simulate tracking noise. The recognition ratios across different standard deviation of noise (Figure 18) show that ratios was decreased when error levels was increased. We confirmed that less than 5.0 pixels of the error is required if we desired 70% recognition levels.

6. Conclusion

This paper proposed a novel feature termed the LFD feature, which models refraction in objects as distortions between multiple views captured by a light field camera. In addition, we designed a learning and recognition method with this LFD feature and performed recognition experiments. We compared this approach based on our LFD feature with the SIFT-based method. Our method using the LFD feature achieved on average 70% accuracy with seven objects against different backgrounds as assessed by leave-

one-out cross-validation in real environments, hence verifying the effectiveness of our LFD feature. We also evaluated the robustness and limitation of the proposed method under various conditions such as: camera and background positions, object poses, and lighting conditions. Furthermore, we conducted simulation experiments that evaluated texture densities of backgrounds and tracking error in detecting LFD feature. In conclusion, we have been successful in: 1) producing a transparent object recognition approach based on a single-shot image, 2) employing a novel feature, namely refraction, for transparent object recognition, and 3) introducing the light field camera array in learning and recognition applications.

In our practical experiments (in Sec.5.2), many LFD features (more than 3% of the image size) were extracted. Nevertheless, the recognition accuracy did not exceed 70% and it is not so high in light of applications. One reason for the result is false correspondence detection, that tracking errors decrease recognition accuracy as indicated from simulation results. Robust correspondence matching has to be considered in future work.

Acknoledgements

The part of this work was supported by Konica Minolta Science and Technology Foundation and Grant-in-Aid for Scientific Research on Innovative Areas "Shitsukan" (No. 23135524) from MEXT, Japan.

References

[1] B. Wilburn, M. Smulski, K. Lee, and M. A. Horowitz, "The light field video camera", *SPIE Electronic Imaging: Media Processors*, vol. 4674, 29-36, 2002.

[2] R. Ng, M. Levoy, M. Brdif, G. Duval, M. Horowitz, P. Hanrahan, "Light Field Photography with a Hand-Held Plenoptic Camera", *Stanford University Computer Science Tech Report CSTR 2005-02*.

[3] A. Lumsdaine and T. Georgiev, "Full Resolution Lightfield Rendering", *Tech. rep., Adobe Systems*, January 2008.

[4] www.viewplus.co.jp/product/camera/profusion25.html

[5] www.raytrix.de/

[6] www.lytro.com/

[7] J. Sivic and A. Zisserman, "Video Google: A Text Retrieval Approach to Object Matching in Videos", *Int. Conf. Computer Vision*, 2003.

[8] G. Csurka, C. Bray, C. Dance, and L. Fan, "Visual Categorization with Bags of Keypoints", *Workshop on Statistical Learning in Computer Vision, European Conference on Computer Vision*, pp. 59-74, 2004

[9] D. Nister and H. Stewenius, "Scalable Recognition with a Vocabulary Tree", *IEEE Conf. Computer Vision and Pattern Recognition*, 2006.

[10] M. Fritz, M. Black, G. Bradski, S. Karayev, and T. Darrell, "An Additive Latent Feature Model for Transparent Object Recognition", *Advances in Neural Information Processing Systems*, 2009

[11] D. Miyazaki, M. Kagesawa, and K. Ikeuchi, "Transparent Surface Modeling from a Pair of Polarization Images", *IEEE Trans. Pattern Analysis Machine Intelligence*, Vol. 26, No. 1, pp. 73-82, 2004.

[12] D. Miyazaki and K. Ikeuchi, "Inverse Polarization Ray Tracing: Estimating Surface Shapes of Transparent Objects", *IEEE Conf. Computer Vision and Pattern Recognition*, pp. 910-917, 2005.

[13] G. S. Settles, "Schlieren and Shadowgraph Techniques", *Cambridge University Press*, 2001.

[14] G. S. Settles, "Important Developments in Schlieren and Shadowgraph Visualization during the Last Decade", *Int. Sympo. Flow Visualization*, paper no. 267, June 2010.

[15] G. Wetzstein and R. Raskar, W. Heidrich, "Hand-Held Schlieren Photography with Light Field Probes", *Int. Conf. Computational Photography*, 2011.

[16] G. Wetzstein, D. Roodnick, W. Heidrich, R. Raskar, "Refractive Shape from Light Field Distortion", *Int. Conf. Computer Vision*, 2011.

[17] M. Ben-Ezra and S. K. Nayar, "What Does Motion Reveal About Transparency?", *Int. Conf. Computer Vision*, pp. 1025-1032, 2003.

[18] S. Agarwal, S. P. Mallick, D. Kriegman, and S. Belongie, "On Refractive Optical Flow", *Europen Conference on Computer Vision*, pp. 483-494, 2004.

[19] N. J. W. Morris and K. N. Kutulakos, "Dynamic Refraction Stereo", *Int. Conf. Computer Vision*, pp. 1573-1580, 2005

[20] G. Farneback, "Two-Frame Motion Estimation Based on Polynomial Expansion", *Scandinavian Conference on Image Analysis*, 2003.

[21] Y. Xu, K. Maeno, H. Nagahara and R. Taniguchi, "Mobile Camera Array Calibration for Light Field Acquisition", *Int. Conf. Quality Control by Artificial Vision*, 2013.