4 차원 라이트필드 영상을 위한 특징 기술자

조나단 사무엘, 윌리엄, 박인규

인하대학교, 정보통신공학과

{jlumentut@inha.edu, 22112195@inha.edu, pik@inha.ac.kr}

Research area in feature descriptor has been done widely along with feature detector to increase feature matching performance. Viewpoint invariant is one of the issues in feature matching. To obtain better matching result in scenes with different viewpoint, robust viewpoint invariant descriptor is needed. In this framework, we introduced method to obtain viewpoint invariant descriptors by utilizing light field images. Initial features are extracted from the central image of sampled light field images. Then, corresponding features in neighboring images are tracked using KLT tracker. Their features descriptors are then computed and used to describe the initial features. Experimental results show the proposed descriptors obtained from light field images give better matching result compared to normal descriptors.

1. Introduction

It is already more than a decade that features and descriptors are created and widely used in many image processing and computer vision algorithms. Feature descriptors are made to be invariant of illumination, rotation, scale, and viewpoint change. Among them, viewpoint difference has been an interesting topic, yet arduous, for generating robust descriptor. Feature descriptor such as MSER [1] and ASIFT [2] algorithms have been proposed to increase feature's robustness on viewpoint change. MSER is not scale invariant because it does not simulate the blur in image with depth distance of camera. ASIFT [2] handles the transformation dilemma in MSER algorithm by applying manual projective transformation to achieve smooth deformation. Their method of simulating camera axis orientation with fixed parameters, such as rotation angle and tilt angle, produces significant number of inliers compared to normal matching. Using this idea to gain similar objective, our method relies on multiview light field images to substitute the projective simulation part.

We start by capture light field image using Lytro Illum camera [3]. The raw light field image captured from the Lytro camera is adjusted and extracted using light field image extractor toolbox provided by Dansereau *et al*. [4]. There are total 11 × 11 light field images extracted from a raw data. From total 11 × 11 light field images, $n \times n$ images are sampled. There is a central image in each sampled light field images. Initial features are detected on this image using SIFT detector. These initial features are then used in KLT tracker [5] to track their correspondences in the neighboring sampled images. The reason not to extract features individually in all sampled images is to keep the total feature numbers to be constant. Descriptors are then computed from those detected features.

2. Proposed Method

By substituting the manual orientation and tilting simulation from previous work by [2], we use the sampled light field images to extract descriptors from different viewpoint. As shown in Fig. 1, the proposed method starts by extracting initial features from central image that is located in [$(n+1)/2$, $(n+1)/2$] index from $n \times n$ sampled light field images. After initial features are extracted, their correspondences from neighboring images are tracked using KLT tracker. The descriptors of each detected key point are calculated using SIFT descriptor algorithm. Finally, exhaustive matching with complete combination is done to all detected features. More detail explanation of our proposed method is described in the following subsections:

2.1 Image Sampling

Image sampling is done by choosing 3 × 3 or 5 × 5 images from 121 light images. Central image is located precisely at the center of $n \times n$ light field images. Note that the further location of one neighbor image to its center, the bigger viewpoint difference it has. In our experiment, the maximum n value is limited to 5 since the time complexity for exhaustive matching is $O(n^4)$.

2.2 Feature and Descriptor Extraction

After light field images sampling is done, initial features are extracted using SIFT detection algorithm. The initial features are detected on central image. Then, the corresponding features from all $n \times n - 1$ neighboring images are detected using KLT tracker. This approach keeps the number of detected key points to be constant in each sampled light field image.

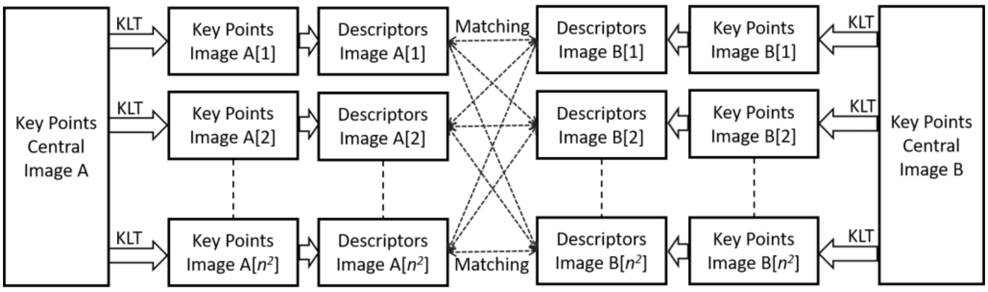
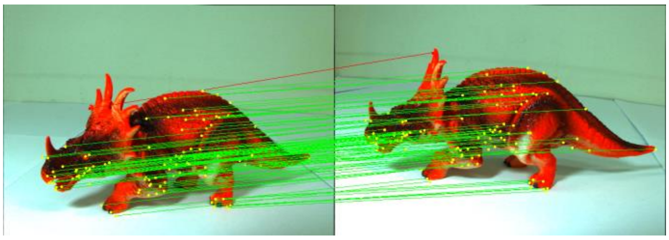
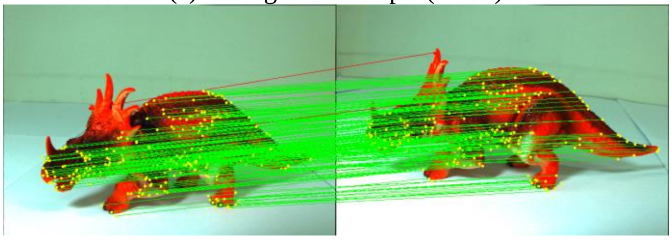


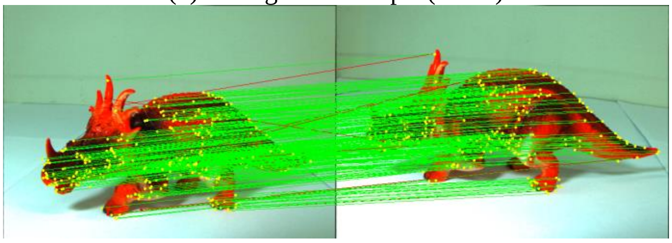
Fig. 1. Light field descriptor and matching framework



(a) Orange triceratops (105/1)



(b) Orange triceratops (271/2)



(c) Orange triceratops (396/8)

Fig. 2. Match result from sampled dataset: (a) Normal SIFT, (b) LF-SIFT 3 × 3 sampling, and (c) LF-SIFT 5 × 5 sampling

Table I. Inlier rate

	Normal-SIFT	LF-SIFT (3 × 3)	LF-SIFT (5 × 5)
Average	0.87	0.85	**0.88**

For computing local descriptor, SIFT descriptor algorithm is applied in every detected feature from $n \times n$ light field images. The patch size of each detected feature is 16 × 16. Each subset of descriptor represents the same feature from single view point (i). Note that each subset descriptors i from total n^2 multiview descriptors are not concatenated.

2.3 Descriptor Matching

After collecting all features descriptors from central image and its neighbors, the next step is to do exhaustive matching to obtain correspondences between 2 compared data.

3. Experimental Result

This experiment is run on an Intel i7-6700 @ 3.4 GHz computer with 16 GB RAM. As depicted in Fig. 2, comparison is done between Normal SIFT and our proposed method, Light Field SIFT (LF-SIFT). Descriptors of Normal SIFT are extracted only from the central image while descriptors of LF-SIFT are extracted from central image and its neighbors. In this experiment, we also included the result of using 5 × 5 sampling along with 3 × 3 sampling. Inliers and outliers are checked manually during matching. Performance evaluation of proposed method can be seen in Table I.

4. Conclusion

In this paper, we proposed a framework to extract descriptors that are invariant to viewpoint. The idea of collecting features from sampled light field images led to produce more inlier result than using one viewpoint image (normal) in matching process. For future works, we will use more variation such as illumination, rotation, and scale in the datasets.

Acknowledgement

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. NRF-2013R1A2A2A01069181).

References

[1] M. Donoser and H. Bischof, “Efficient Maximally Stable Extremal Region (MSER) Tracking,” *Proc. IEEE CVPR*, pp. 553-560, June 2006.

[2] J-M. Morel and G. Yu, “ASIFT: A new framework for fully affine invariant image comparison,” *SIAM Journal on Imaging Sciences*, vol. 2, pp. 438-469, April 2009.

[3] The Lytro camera, https://www.lytro.com.

[4] D. G. Dansereau, O. Pizarro, and S. B. Williams, “Decoding, calibration and rectification for lenselet-based plenoptic cameras,” *Proc. IEEE CVPR*, pp. 1027-1034, June 2013.

[5] B. Lucas and T. Kanade, “An iterative image registration technique with an application to stereo vision,” *Proc. IJCAI*, vol. 2, pp. 674-679, August 1981.