[Supplementary Material]
GEOBIT: A Geodesic-Based Binary Descriptor Invariant to Non-Rigid Deformations for RGB-D Images

Erickson R. Nascimento[1], Guilherme Potje[1], Renato Martins[1,2], Felipe Cadar[1], Mario F. M. Campos[1], and Ruzena Bajcsy[3]

[1]Universidade Federal de Minas Gerais (UFMG) [2]INRIA [3]University of California Berkeley

{erickson,guipotje,renato.martins,cadar,mario}@dcc.ufmg.br, bajcsy@eecs.berkeley.edu

In this supplementary material, we provide auxiliary information about the implementation, collected data and the tracking application presented in our ICCV submitted paper. Please also have a look in our video demonstrating the non-rigid tracking application mentioned in the paper. We start presenting some implementation details in Section 1. Subsequently, we detailed our annotated RGB-D dataset of real and synthetic objects subjected to non-rigid deformations in Section 2. Finally, Section 3 presents more qualitative results of the tracking and recognition rate omitted in the paper due to space restrictions.

1. Implementation Details

Our method receives a triangular mesh as input. This mesh can be computed from any RGB-D image. We reconstructed the triangular mesh by triangulating neighboring pixels in image space, which is the fastest approach. Although this may produce triangles of irregular shapes, uneven sampling density and holes, this proved to be a good trade-off between accuracy and speed during our experiments. All RGB-D descriptors tested in the paper were fed with the same mesh for the sake of fairness. As pointed in the submitted manuscript, we applied a multi-resolution strategy with a Gaussian pyramid to both remove depth noise and accelerate the estimation of the geodesic isocurves. Table 1 shows the actual speed-up achieved by using this multi-resolution strategy, where we also report the mean time required for computing a single keypoint descriptor. We chose as the default parameter for the descriptor the pyramid maximum level $L2$, since it provides a good balance between speed-up and geometric detail.

We also provide the results obtained by testing different test patterns on the isocurves. Figure 1 shows the achieved recognition rates when testing different patterns, as explained in the paper. The default pattern chooses points in a uniform distribution considering the radius of the support region, while the Gaussian distribution uses a Gaussian sampling centered at the position of the keypoint. We chose the Gaussian with $1,024$ tests. Further doubling the size of descriptor provides small recognition rate gain in exchange of doubling its size and calculation time.

2. RGB-D Dataset with Non-Rigid Deformations

Dataset Annotation Tool. To annotate the ground-truth keypoint correspondences between the frames, we developed a tool that is also released to the community. The graphical interface of the tool is depicted in Fig. 2. Each image sequence

Table 1. Speed up achieved by the downscale pyramid sampling where L is the level of the pyramid per frame. Each level reduces the image resolution by half ($L0$ is the original resolution). Mean time represents total CPU usage by our descriptor extractor software on an Intel(R) Xeon(R) CPU E5-2650 v3 @ 2.30GHz machine to compute a single keypoint descriptor. Note that this time considers all steps, including loading the point cloud and saving results.

L	Resolution	Speed-up	Mean Time (s)
0	640×480	1.00	42.94
1	320×240	3.50	12.26
2	160×120	35.78	1.20
3	80×60	306.71	0.14

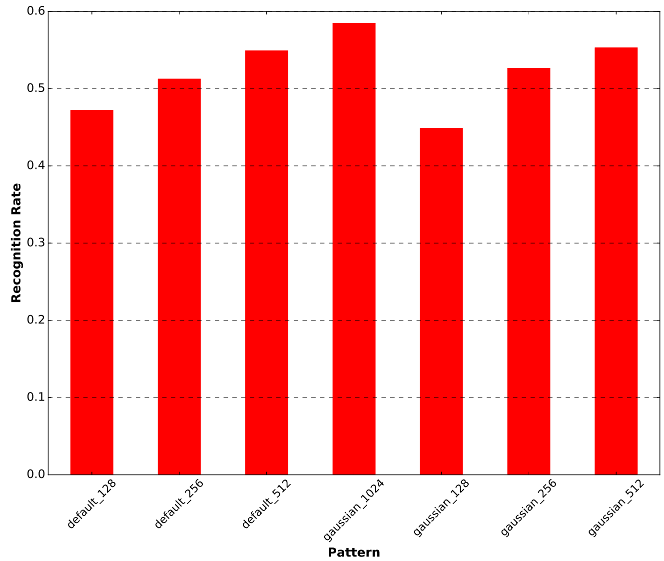


Figure 1. The average of the recognition rates for different distribution patterns. The Gaussian distribution with $1,024$ tests performed better than the other distributions.

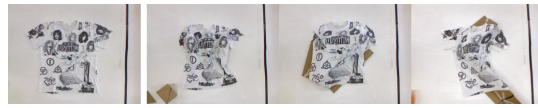
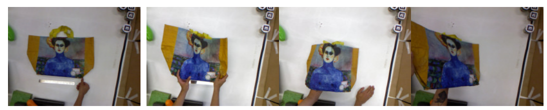
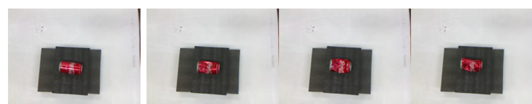
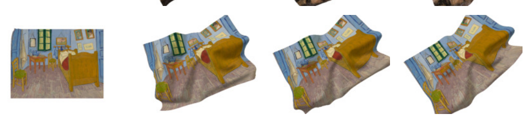
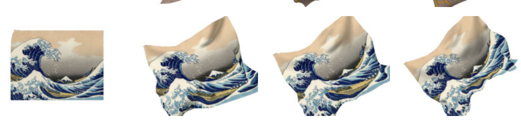
Figure 2. Graphical interface of the developed annotation tool. Magenta areas indicate that depth information is not available.

was annotated by at least three different individuals. If an annotated point differs from at least 8 pixels to other annotated points, it is considered a mismatch and automatically discarded. If none of the matches agree, they are summarily removed from the set of ground-truth correspondences. This way we lower the chances of generating false ground-truth matches. After validating the annotated points, we calculate the mean position for the remaining points having multiple annotations, to generate the final annotated keypoint position.

Real-world data. These image sequences are composed of 6 deformable objects and a total of 74 pairs of RGB-D images captured with a Kinect $^{\text{TM}}$. Naturally, non-linear illumination changes occurs when manipulating the surface of those objects. We manually annotated about 50 keypoints and the ground-truth correspondence for all datasets. Table 2 summarizes all details of our dataset and shows some examples of the frames therein.

Synthetic data. We applied physics' simulation of cloth to create arbitrary non-rigid isometric deformations with ground-truth correspondences. The texture is applied onto the mesh generated by the grid and rendered with diffuse illumination as the cloth moves (which causes non-linear illumination changes). The synthetic data is composed of 18 pairs of images comprising three different textures with arbitrary deformations, and rotations. Table 2 summarizes all details of our dataset and shows some examples of data.

Table 2. RGB-D deformable objects dataset.

Name	**#Pairs**	**#Keypoints**	**Material**	**Kind**	**Image samples**
Shirt 1	14	50	Fabric	Real	
Shirt 2	18	65	Fabric	Real	
Shirt 3	17	55	Fabric	Real	
Blanket	15	50	Fabric	Real	
Bag	4	51	Plastic	Real	
Can	5	31	Metal	Real	
Lascaux	18	94	-	Synthetic	
Vangogh	18	96	-	Synthetic	
Kanagawa	18	100	-	Synthetic	

We provide the dataset [1] and the source-code of the non-rigid deformation simulator engine used to generate the synthetic images.

3. Non-Rigid Tracking Application

Our tracking application uses the proposed descriptor and also its best competitor, namely DaLI [1], to track correspondences between frames, and warp the deformed object into its reference frame using the found matches. The warped image allows us to visually assess the descriptors performance, since for a perfect descriptor and assuming no occlusions induced by the deformation, the warped image must depict the original undeformed version of the object. Figure 3 shows the tracking results for three image sequences from our dataset.

[1] `https://www.verlab.dcc.ufmg.br/descriptors/iccv2019`

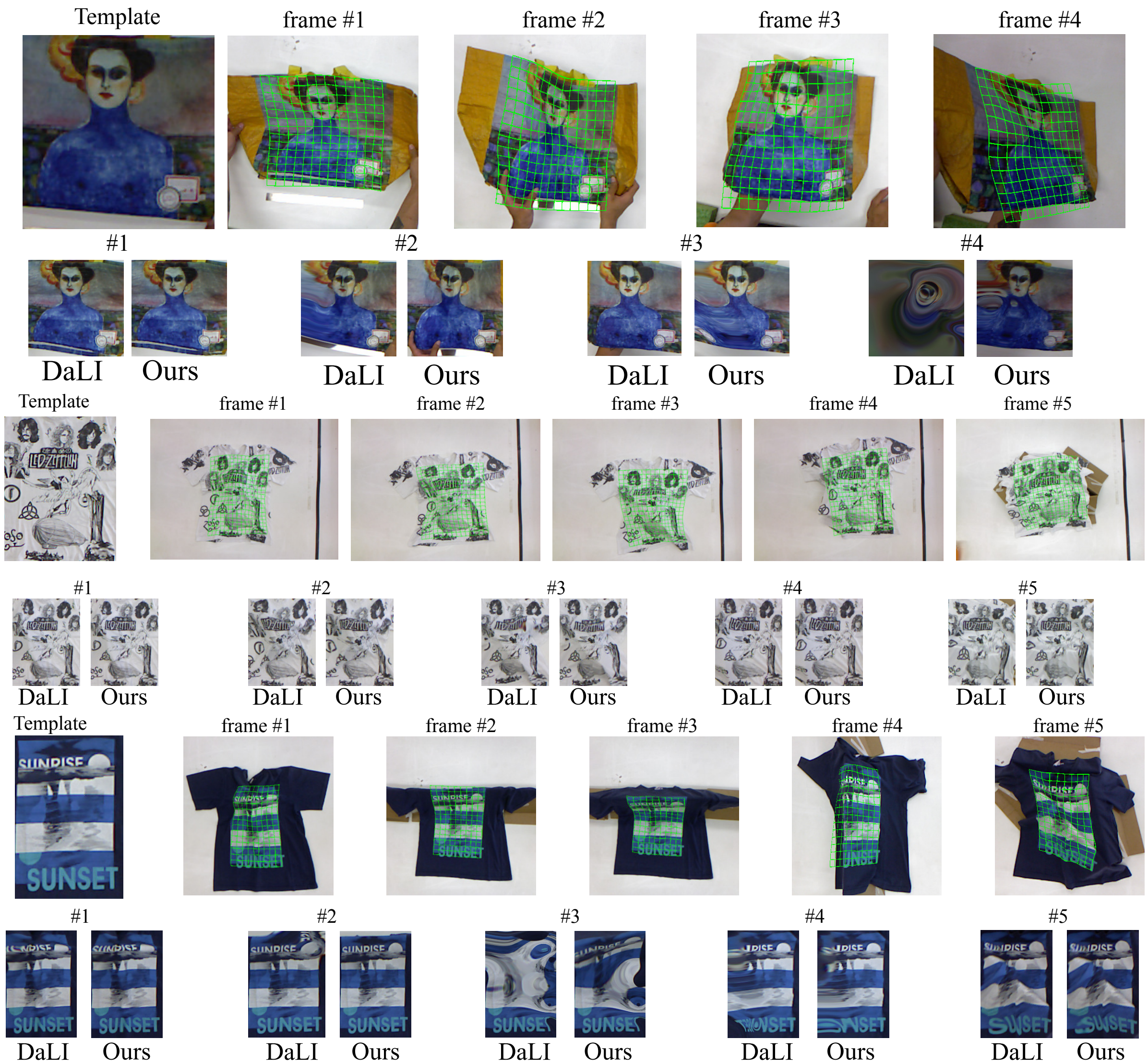


Figure 3. Results of tracking deformable surfaces using our descriptor and DaLI.

References

[1] F. Moreno-Noguer. Deformation and illumination invariant feature point descriptor. In *CVPR 2011*, pages 1593–1600, 2011.