

ORIENTATION OF IMAGES WITH LOW CONTRAST TEXTURES AND TRANSPARENT OBJECTS

L. Morelli ^{a,b}, A. Karami ^{a,c}, F. Menna ^a, F. Remondino ^a

^a3D Optical Metrology (3DOM) unit, Bruno Kessler Foundation (FBK), Trento, Italy
Web: <http://3dom.fbk.eu> – Email: <lmorelli><akarami><remondino><fmenna>@fbk.eu

^bDept. of Civil, Environmental and Mechanical Engineering (DICAM), University of Trento, Italy

^cDept. of Information Engineering and Computer Science (DISI), University of Trento, Italy

Commission II

KEY WORDS: Transparent objects, Normalized cross-correlation, Image orientation, Image matching, Tie points.

Abstract:

Objects with non-collaborative or transparent surfaces pose challenges to image orientation procedures and are an open research task in photogrammetry and computer vision. In this paper, we analyse the critical issues that cause image orientation failures and propose two approaches that leverage the low-contrast textures present on object surfaces to accurately orient an image block. Both approaches privilege tie point detection on low-contrast textures, discarding specular reflections and static tie points. In the first approach local descriptors are extracted in those regions where roughness and micro-structures are better highlighted, applying the normalized cross-correlation (NCC) on the gradient map of the images to fully exploit the geometrical content of the patches. The second approach builds on the first method modifying the classic RootSIFT pipeline and obtaining a faster and more reliable approach. Different transparent objects with different surface characteristics are tested to evaluate the efficiency and reliability of the proposed pipelines for image orientation and successive dense 3D reconstruction.

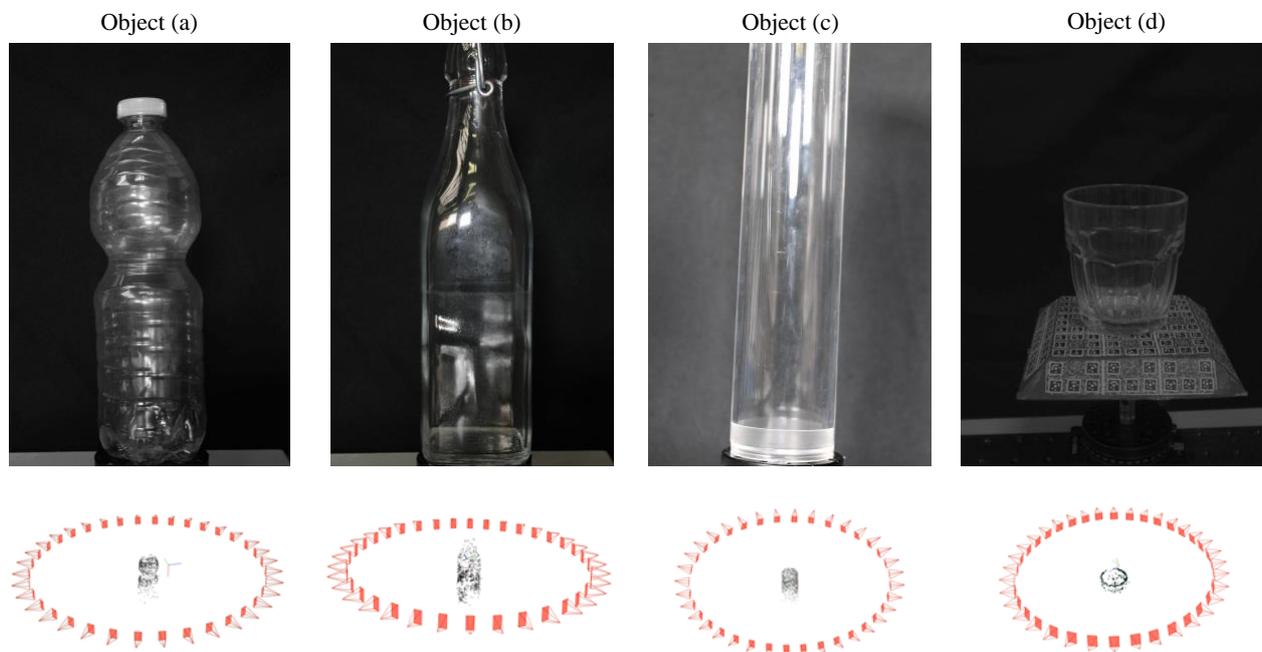


Figure 1: Example of transparent objects used in this investigation (upper row) and visual representations of recovered camera network and sparse 3D reconstruction (lower row).

1. INTRODUCTION

Estimating the position and orientation of 3D objects is one of the fundamental problems in photogrammetry and computer vision and requires repeatable, reliable, and well distributed tie points (Remondino et al., 2021; Bellavia et al., 2022). Among the different fields of application, photogrammetry is widely used for industrial measurements and inspections, although several challenges exist when dealing with non-collaborative surfaces (Figure 1), such as textureless (Ahmadabadian et al., 2017, 2019),

reflective (Karami et al., 2021, 2022b), and refractive objects (Wu et al., 2018; Liu et al., 2020; Karami et al., 2022a). In transparent objects, the ability to diffusely reflect light is very limited, and in addition, they are almost textureless. Due to refraction and mirror-like reflections, a part of the surface recorded textures of such objects is not invariant to the camera viewpoint being also dependent on the object's shape, surrounding environment, and illumination conditions. Therefore, standard procedures like those implemented in SfM applications become ineffective, leading to significant mistakes

and, most frequently, failures in the image matching and orientation process (Hosseinaveh et al., 2015; Wu et al., 2018; Karami et al., 2022c).

To deal with 3D reconstruction of transparent objects, different approaches have been proposed. For instance, shape from distortion is a well-known approach for 3D shape reconstruction of transparent objects. This method analyzes the distortion of a known pattern positioned beneath or near the surface (Ben-Ezra and Nayar, 2003; Wetzstein et al., 2011; Tanaka et al., 2016; Kim et al., 2017). These techniques, however, are limited to the recovery of a single refractive surface or the reconstruction of objects with simple geometry, making them difficult to use for other complex objects (Wu et al., 2018; Lyu et al., 2020; Karami et al., 2022b). Shape from Silhouette (SFS) is a well-known 3D reconstruction approach applied to various object categories, including opaque, translucent, and transparent objects, as long as the region of the object in each image can be separated from the background (Baumgart, 1974; Wu et al., 2018; Lyu et al., 2020). However, concavities on an object's surface is the main difficulty with SFS making it impossible to reconstruct the inside of a hole or concave region. Another technique for determining an object's 3D reconstruction is called Shape from Polarization (SFP) (Miyazaki et al., 2004; Huynh et al., 2010; Cui et al., 2017; Sun et al., 2020) which uses the polarization information of the reflected light. One of the major problems with this strategy is the uncertainty in polarization analysis (Durou et al., 2020; Karami et al., 2022b). 3D reconstruction based on the shape from heating disregard the object's refractive characteristics making it suitable for a wide range of objects (Eren et al., 2009; Brahm et al., 2016; Landmann et al., 2021). However, the application of this approach is limited due to the high-cost IR camera and low acquisition speed. Methods based on the direct ray measurement (Kutulakos et al., 2008; Tsai et al., 2015; Qian et al., 2016; Kim et al., 2017; Lyu et al., 2020) have also been used for 3D reconstruction of transparent objects. However, these methods not only struggle with collinearity ambiguity but also require additional assumptions and constraints, making them inadequate for practical industrial applications with a wide range of situations (Ihrke et al., 2010; Karami et al., 2022b).

Various learning-based works (Stets et al., 2019; Sajjan et al., 2020; Zhu et al., 2021; Eppel et al., 2022) have been recently proposed. Stets et al. (2019), for example, introduced a deep convolutional neural network (CNN) technique using only a single image taken under an arbitrary environment map to generate depth of a transparent object. More similarly, Eppel et al. (2022) proposed an advanced learning-based method to predict 3D points of transparent objects from an image captured using an unknown camera. Instead, Sajjan et al. (2020) and Zhu et al. (2021) proposed different learning-based techniques for filling in missing depth (where transparent objects are) from a scene. Li et al. (2020) proposed an advanced physically-based network for constructing the 3D geometry of transparent objects using various multi-view images. Learning-based approaches, in contrast to previous methods, learn from actual or synthetic training data and do not require assumptions or constraints such as controlled data collection (Karami et al., 2022b). However, these methods are still significantly less precise than conventional methods and need large datasets annotation of specific object types, making them unsuitable for industrial applications requiring 3D measurement accuracy, dependability, and traceability (Karami et al., 2022b).

One of the practical surface treatments that is used for 3D inspection of manufactured parts is to coat the surface of such objects by spraying a thin layer of white or colored powder.

Although, this treatment makes the surface opaque and diffusely reflective (Palousek et al., 2015; Pereira et al., 2019; Karami et al., 2021), it is still difficult, resource-intensive, and impractical for real-time 3D inspection of industrial materials (Pereira et al., 2019; Karami et al., 2022b). Moreover, this surface treatment is well suited only for active systems such as laser scanners and fringe projection because of the textureless appearance.

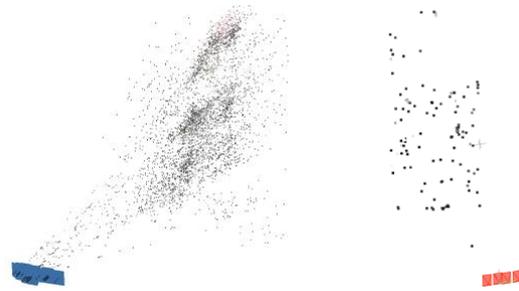


Figure 2: Partial and incorrect image orientation in Metashape¹ and COLMAP².

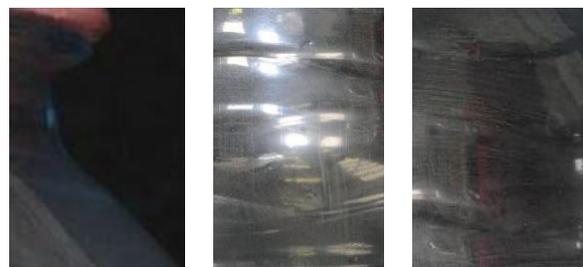


Figure 3: Silhouette regions (left) and specular reflections (centre) appear almost constant in different images. Low contrast textures areas (right).

Contrary to the previous approaches, we present two methods that only utilise the low contrast textures present on transparent objects to estimate the external orientation of the cameras. During our investigations, we analysed the results of COLMAP (Schonberger and Frahm, 2016) and Agisoft Metashape¹, which failed at the image matching stage, thus no camera parameters could be recovered (Figure 2). The reason is that SIFT-like descriptors prioritize image patches with high contrast, e.g. textures along the silhouette of the transparent object (Figure 3a), or reflections coming from surrounding objects and light sources (Figure 3b), and discard all the low-contrast textures.

If the camera position and the surroundings remain constant and only the object rotates, the reflected textures remain quite steady from one image to the next if the object is a solid of revolution (or the surface of the surveyed object changes slowly). In this situation, the descriptor finds matches in the same position across all images, assuming that the object stayed consistent. Conversely, if the object is not a solid of revolution, a patch that is first located on the silhouette in a successive image will be located inside the object with a completely different appearance. In both cases the matching step of SfM generally fails. In addition, regular descriptors that prioritize high contrast textures ignore low contrast textures on extremely high resolution images that exhibit surface flaws, dust, or other particles (Figure 3c), even if they can be utilized for image matching.

¹ <https://www.agisoft.com/>

² <https://colmap.github.io/>

1.1 Paper aim and contribution

In this paper, we leverage the low contrast textures present on the surface of transparent objects for image matching and image orientation using the classical structure-from-motion (SfM) pipeline with high resolution RGB images as input. The working pipeline proposed and afterwards described, starts from the previous analysis and relies on the following observations:

1. A dark background helps to reduce the reflections from surrounding objects and absorb the refracted light that passes through the object.
2. Many transparent surfaces when imaged at high resolution reveal some feature points and textures that are generally rejected from common descriptors and can be used for image matching.
3. Reflections and shadow must be eliminated since they do not rigidly rotate from one image to the other.
4. The radiometric appearance of low-contrast textures can vary dramatically from one image to the next, hence the employed descriptor must be robust to radiometric variations.
5. A simple but effective approach to describe and match patches is the normalized cross-correlation (NCC), which can be used for both image orientation and dense reconstruction. NCC is the starting point of our analysis, but other descriptors could also be considered, therefore we present also a second approach based on RootSIFT (Arandjelović and Zisserman, 2012).
6. To take full advantage of the geometrical content of the patches, NCC must be run on the gradient map of the grayscale image since applying NCC on grayscale images of transparent objects is not robust enough.

2. METHODOLOGY

The steps of the two proposed pipeline including the detection of candidate key points, the matching with NCC and RootSIFT, and the image orientation in COLMAP with self-calibration are afterwards presented. The detection and description parts of the Cross-Correlation pipeline rely on the python-opencv (Bradski, 2000) and PIL (Clark, 2015) libraries.

2.1 Cross-Correlation pipeline

The detection is strongly related to the description which relies on the similarity of patches extracted on the gradient map. Therefore, the best candidate tie points should be those surrounded by discriminative regions, where there is enough gradient. Furthermore, we need to eliminate reflection regions since they usually do not move solidly with the object rotating on the turn table. Based on these assumptions, Figure 4 depicts the essential steps of the proposed detection pipeline with an example.

1. A sequence of images are used as input (Figure 1). High resolution images must be preferred to highlight surface roughness and details on the transparent surfaces.
- 2-3. Backgrounds are removed with Removal.AI (<https://removal.ai/>), a deep-learning tool for background removal, and converted to masks.
4. Masks are applied to process only image areas containing the transparent object, and especially discarding the coded targets for metric evaluations. The purpose of the paper is in fact to orient the images with only the texture of the object.
- 5-6. The gradient of the image is estimated to highlight and emphasise the geometrical content of the textures, then a gaussian kernel with a large standard deviation is applied to mask specular reflections and keep only those parts where roughness and microstructures are highlighted better.

7-8-9. A density map on cells of 33x33 pixels (the same size we use for NCC) is estimated and then keypoints are extracted and ranked by their density. This step can be also substituted with a Harris detector. In the first qualitative tests of Figure 1 both the "density" method and the Harris detector were successfully used. In the quantitative tests (Table1-2, and Figure 7) only the Harris detector was used, in the future, we will also test the "density" method.

10. Finally, non-maxima suppression is applied to a neighbourhood of thirty-three pixels.

The description and matching step leverage the knowledge of the approximate epipolar lines (almost horizontal because of the acquisition network) to search for the best match of a candidate 33x33 pixel patch cropped around each keypoint. Each patch is converted in its gradient map (Figure 5a) and compared with NCC within a rectangular searching window extracted along the a-priori known epipolar lines (Figure 5b and c). For NCC we used the formula from Zhao et al. (2006) without the window normalization (Equation 1), since our datasets do not present scale changes.

$$S_{m1,m2} = \sum_{u=-w}^w \sum_{v=-w}^w \left[[A_{uv} - \underline{A}] \cdot [B_{uv} - \underline{B}] \right] / (\sigma(A) \sigma(B)) \quad (1)$$

Where S is the score, $m1$ the reference patch cropped around the keypoint and $m2$ is one of all the possible patches inside the rectangular searching window. The patch size is $(2w + 1) \cdot (2w + 1)$, u and v are the local coordinate system with $(0, 0)$ located in the center of the patch, A_{uv} and B_{uv} are the gradient intensity in position (u, v) . \underline{A} and \underline{B} are the average intensity value of each patch, $\sigma(A)$ and $\sigma(B)$ are the standard deviations. The output is a score map (Figure 5d) and the best match is extracted in the maximum. For computational constraints the matches of each image were calculated for only two images forward and two backward for the qualitative datasets (Figure 1), and one forward and one backward for the quantitative datasets (Figure 7). The patches are limited to a 33x33 pixel size. The method has several hyper-parameters that have been chosen to balance computation times and final accuracy.

The raw matches are imported to be geometrical verified in COLMAP with RANSAC. All matches, both inliers and outliers are extracted along approximate epipolar lines, therefore the RANSAC maximum error threshold was set to 1 pixel to be very restrictive and be able to extract more correct matches. Because of the simple, non-redundant acquisition network, a weak initialization might lead the solution to converge to a local minimum rather than the global maxima.

2.2 First experiments and RootSIFT-based pipeline

The first experiments aimed to understand why Metashape and COLMAP generally fail (Figure 2) to orient transparent objects with low contrast textures, leading to the consideration of section 1.1 and to the proposal of the Cross-Correlation pipeline that is mainly based on the elimination of "static" tie points between image pairs related to reflection effects, and the extraction of tie points on low-contrast textures that are normally discarded by traditional local features. With the proposed pipeline we oriented the image blocks of the objects shown in Figure 1, where the acquisition network and the sparse reconstruction are qualitatively correct.

From these first results, it was possible to adapt the standard COLMAP pipeline based on the RootSIFT descriptor for these types of datasets. The second approach is based on the following

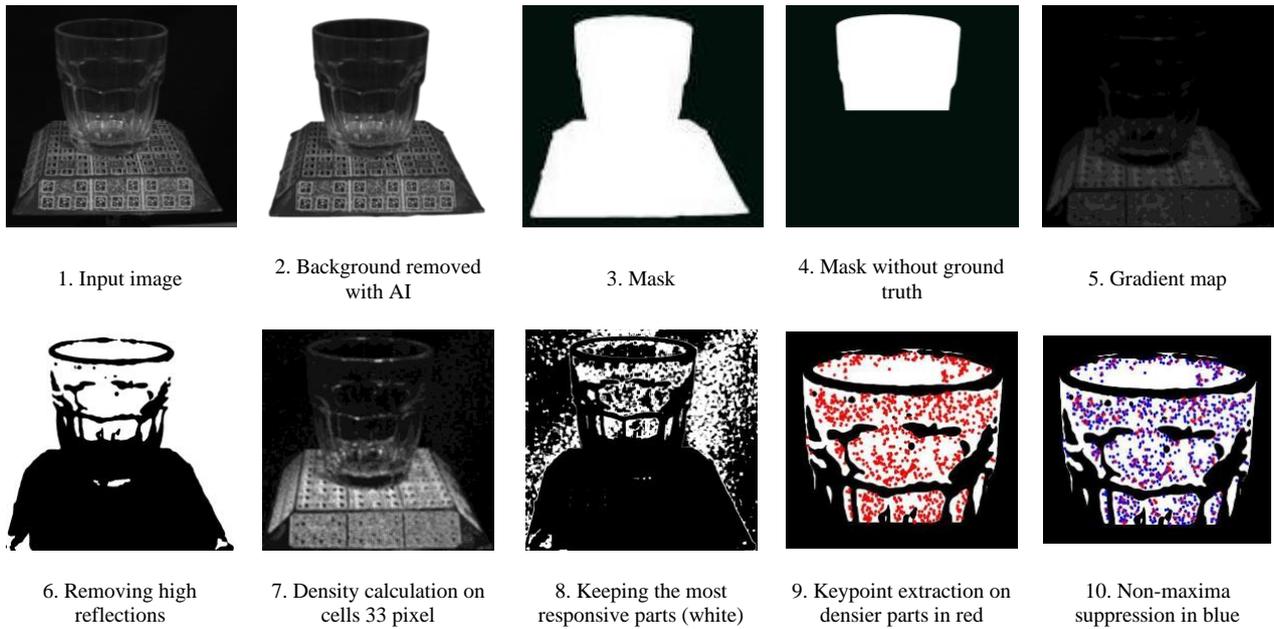


Figure 4: Proposed detection pipeline for the cross-correlation pipeline. Steps 7-8-9 can be substituted with the Harris detector.

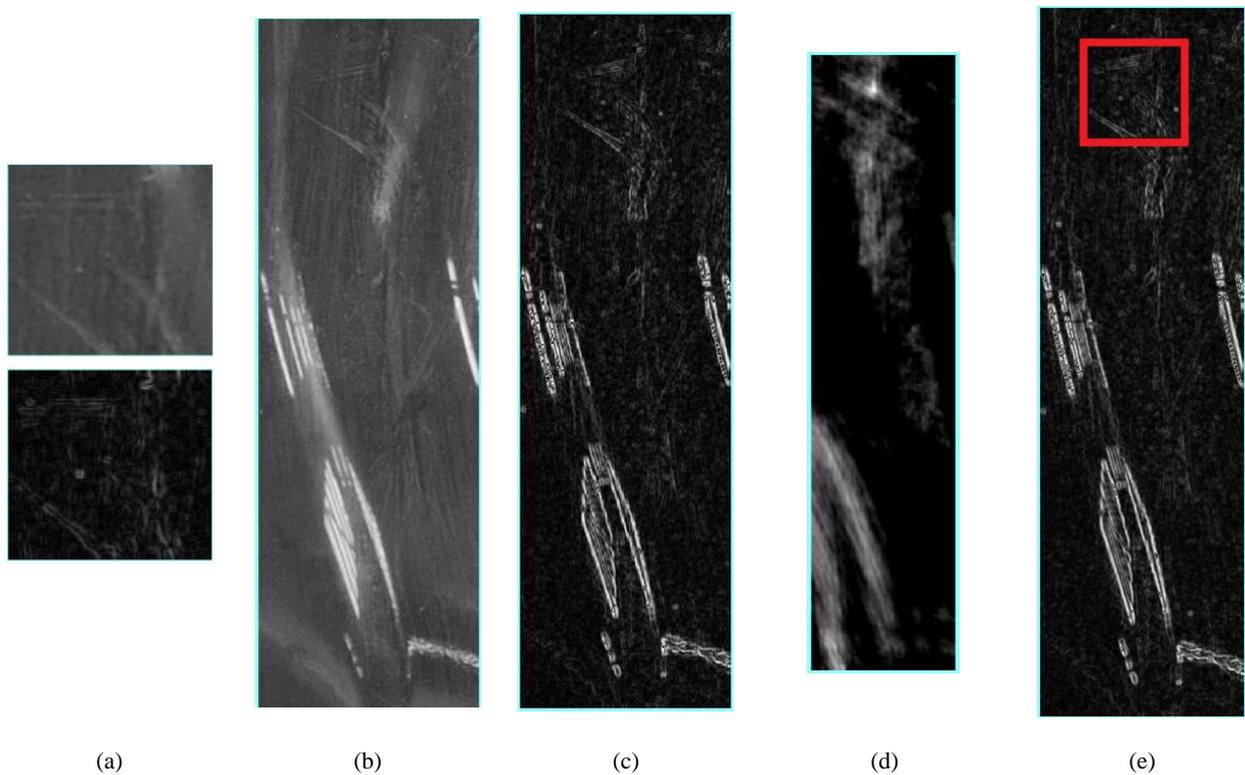


Figure 5: Description and matching with NCC: (a) patches extracted around each candidate keypoint, (b) searching window on the previous or following two images, (c) gradient of the searching window, (d) score map after NCC, (e) visualisation of the patch around the maximum score.

changes to the standard COLMAP pipeline and will later be referred to as RootSIFT-based pipeline:

1. The extraction of 8000 interest points with RootSIFT upright, using the COLMAP implementation, decreasing the contrast peak threshold from 0.066 to 0.026 to detect local features also in the low contrast areas. In addition, using a not

rotation invariant descriptor (the “upright” version) increases the number of absolute correct matches and the inlier ratio.

2. The usage of the near neighbour strategy (NN) instead of the near neighbourhood ratio (NNR) to increase the number of matches.

3. The usage of sequential matching instead of the brute-force approach. In fact, the brute-force matching without NNR

can lead to consider as good matching images that do not have overlap, causing a partial failure in the orientation. Sequential matching works in our case study with only a strip of circularly acquired images, but with other cases a rough knowledge of the camera motion is enough to avoid failures.

4. The elimination of static tie points checking that for each consecutive image pair the apparent flux, like the one of Figure 6, which refers to manually taken tie points, is larger than a certain threshold, e.g., 50 px.

5. Finally the usage of RANSAC as global geometric constraint to filter outlier matches.

The RootSIFT-based pipeline contains several hyper-parameters: the optimal value for the contrast threshold in the detection step, the threshold error for RANSAC, RootSIFT upright vs rotation invariant RootSIFT, the near neighbourhood ratio (NNR) vs the near neighbourhood (NN) approach, intersection vs union strategy for brute-force matching (Jin et al., 2021), RGB images vs the gradient map as input for the descriptor. To set these hyper-parameters, we selected a pair of images from the Plastic Bottle dataset and we determined the parameters combination that generates the higher number of correct matches with a high inlier ratio. To estimate the number of correct matches we employed a reference fundamental matrix calculated from a set of manually selected tie points.

To further filter out the incorrect matches that lie along the epipolar lines, we also calculated the approximate flux from the manually taken tie points (yellow lines in Figure 6), and we interpolated the data to estimate the apparent flux for all the area covered by the bottle. This approximate ground-truth flux has been used to filter out the outliers along the epipolar lines, not filtered by the fundamental matrix.

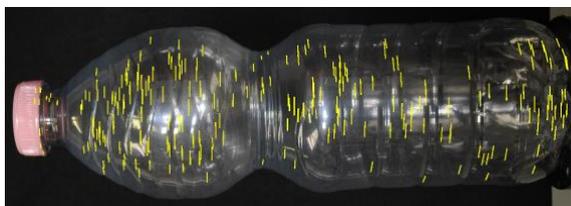


Figure 6: ground-truth flux calculated from manually taken tie points.

Table 1 reported the results for different sets of hyper-parameters in terms of absolute correct matches and inlier-ratio. The parameters to be evaluated are reported in bold,

while the best parameters are highlighted in grey, which are: 0.0026 for the contrast (peak) threshold, NN and not NNR with the intersection strategy for the brute-force matching, RootSIFT upright as descriptor, and RGB images used as input for the descriptor.

3. EXPERIMENTS AND EVALUATION

3.1 Image orientation evaluation

To evaluate the accuracy of the two proposed pipelines, we tested the plastic bottle, the glass bottle and the tea cup shown in Figure 1. For each object, a set of images was captured from 36 different stations under constant lighting, directed to enhance the appearance of the low-contrast texture, and a set of photogrammetric coded targets were printed and mounted on a rotating table with known relative distances. These targets were then employed to provide a metric assessment using different criteria such as mean reprojection error (MRE), and the mean error and standard deviation. The results are reported in Table 2.

For the tea cup dataset (object d), both pipelines achieve an MRE of about 0.4 pixels. RootSIFT obtains a Std of 0.052 mm with Mean Absolute Error (MAE) of 0.048 mm, while the cross-correlation pipeline obtains an Std of 0.079 mm with MAE of 0.083 mm, twice larger than RootSIFT. For the plastic bottle dataset with GSD of 0.05 mm/px, the RootSIFT-based pipeline has a Std of 0.012 mm, MAE of 0.011 mm and MRE of 0.581 px. The cross-correlation pipeline achieves a similar MRE of 0.532 px, while a significantly worse Std and MAE of 0.068 mm and 0.106, respectively. This is because it did not orient four images so that in this case we do not have a loop-closure with a decrease in the accuracy. Failure to close the network may also be due to the lower number of local features used in the cross-correlation method compared to RootSIFT, which has been kept small for computational reasons. For the glass bottle dataset (object b) the cross-correlation approach reaches a better result for all criteria compared to RootSIFT approach.

The visual representation of the generated sparse 3D reconstruction and an example image for each of the three transparent objects are presented in Figure 7.

input	HYPER-PARAMETERS				RANSAC error	MATCHES			RESULTS	
	RootSIFT	NNR	NNR strategy	gradient peak		matches after NN	static filtering	after RANSAC	correct matches	inlier ratio
RGB	upright	1.00	intersection	0.0026	4 px	5469	228	165	124	0.75
RGB	upright	1.00	intersection	0.0026	2 px	5469	228	155	122	0.79
RGB	upright	1.00	intersection	0.0026	1 px	5469	228	132	112	0.85
RGB	upright	1.00	intersection	0.0036	1 px	4565	205	117	97	0.83
RGB	upright	1.00	intersection	0.0026	1 px	5469	228	132	112	0.85
RGB	upright	1.00	intersection	0.0016	1 px	6764	249	143	118	0.83
RGB	upright	1.00	intersection	0.0066	1 px	2673	118	66	50	0.76
RGB	upright	1.00	intersection	0.0026	1 px	5469	228	132	112	0.85
RGB	upright	0.90	intersection	0.0026	1 px	1563	109	89	76	0.85
RGB	upright	0.80	intersection	0.0026	1 px	768	56	51	43	0.84
RGB	upright	1.00	intersection	0.0026	1 px	5469	228	132	112	0.85
RGB	upright	1.00	union	0.0026	1 px	25430	698	217	157	0.72
RGB	upright	1.00	intersection	0.0026	1 px	5469	228	132	112	0.85
RGB	no rotation	1.00	intersection	0.0026	1 px	5673	158	97	79	0.81
RGB	upright	1.00	intersection	0.0026	1 px	5469	228	132	112	0.85
gradient	upright	1.00	intersection	0.0026	1 px	4087	201	86	65	0.76

Table 1: Fine-tuning of the hyper-parameters of the RootSIFT-based pipeline.

DATASET	ROOTSIFT-BASED APPROACH				CROSS CORRELATION			
	Camera Focal length Pixel size Average distance	GSD [mm/px]	Mean Absolute Error [mm]	Std [mm]	Mean Reprojection Error [px]	Mean Absolute Error [mm]	Std [mm]	Mean Reprojection Error [px]
PLASTIC BOTTLE (object a)	NIKON D750 28 mm 5.98 μ m 240 mm	0.05	0.011	0.012	0.581	0.106	0.068	0.532
GLASS BOTTLE (object b)	NIKON D750 28 mm 5.98 μ m 240 mm	0.05	0.054	0.044	0.648	0.016	0.011	0.291
TEA CUP (object d)	NIKON D3X 60 mm 5.98 μ m 300 mm	0.03	0.048	0.052	0.436	0.083	0.079	0.423
Average		0.043	0.037	0.036	0.555	0.068	0.052	0.41

Table 2: Accuracy evaluation for two proposed pipelines using different criteria of Mean Reprojection Error (MRE) in pixel and the residuals in mm (mean absolute error and standard deviation).

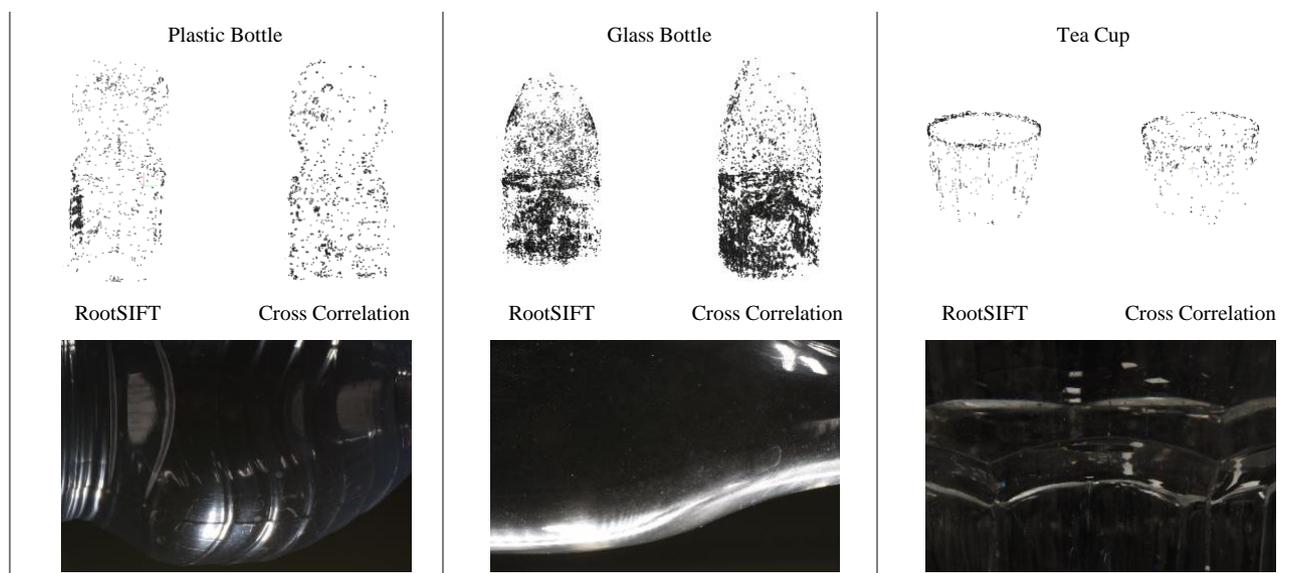


Figure 7: Sparse reconstruction and image detail for the three objects used in the accuracy evaluation.

3.2 Dense cloud evaluation

After orienting the images, the dense cloud was generated using different pipelines, including OpenMVS (Moulon et al., 2013), COLMAP (Schonberger and Frahm, 2016), and Shape from Silhouette (SFS) to propose different possible solutions for 3D dense reconstruction. To generate reference data for each object, their surface was covered with a thin layer of random colored powder to i) make it diffusely reflecting and remove refraction and ii) provide texture on the surface. After surface treatment, an additional photogrammetric 3D reconstruction was employed to generate a dense 3D reconstruction (considered as reference data).

To evaluate the accuracy potential of the suggested method in low-frequency domain, the 3D results achieved with each pipeline were geometrically compared against reference data (photogrammetry). To this end, the obtained 3D point clouds were registered to the reference data using an Iterative Closest Point (ICP) algorithm. The RMS of the shortest distance between the homologous points on the reconstructed and reference models was then calculated and compared. The

results of the point-to-point comparison for both objects (a and b) are presented in Figure 8.

The quantitative analysis shown in Figure 8 demonstrate that the SFS can recover the 3D shape of an object regardless of whether it is textureless or transparent, as long as the region of the object in each image is distinguishable from the background. However, SFS failed to reconstruct concavities or holes on an object's surface making it difficult for geometrically complex objects. It is quite possible that the silhouette of an object would be trimmed or expanded, resulting in a 3D model that is smaller or bigger than the actual size of the object.

On the other hand, the 3D results using OpenMVS and COLMAP (Figure 8) directly depend on the high-resolution texture of the object surface, hence it completely failed or generate noisy point cloud in the area where microstructure and roughness are not highlighted well (see red boxes).

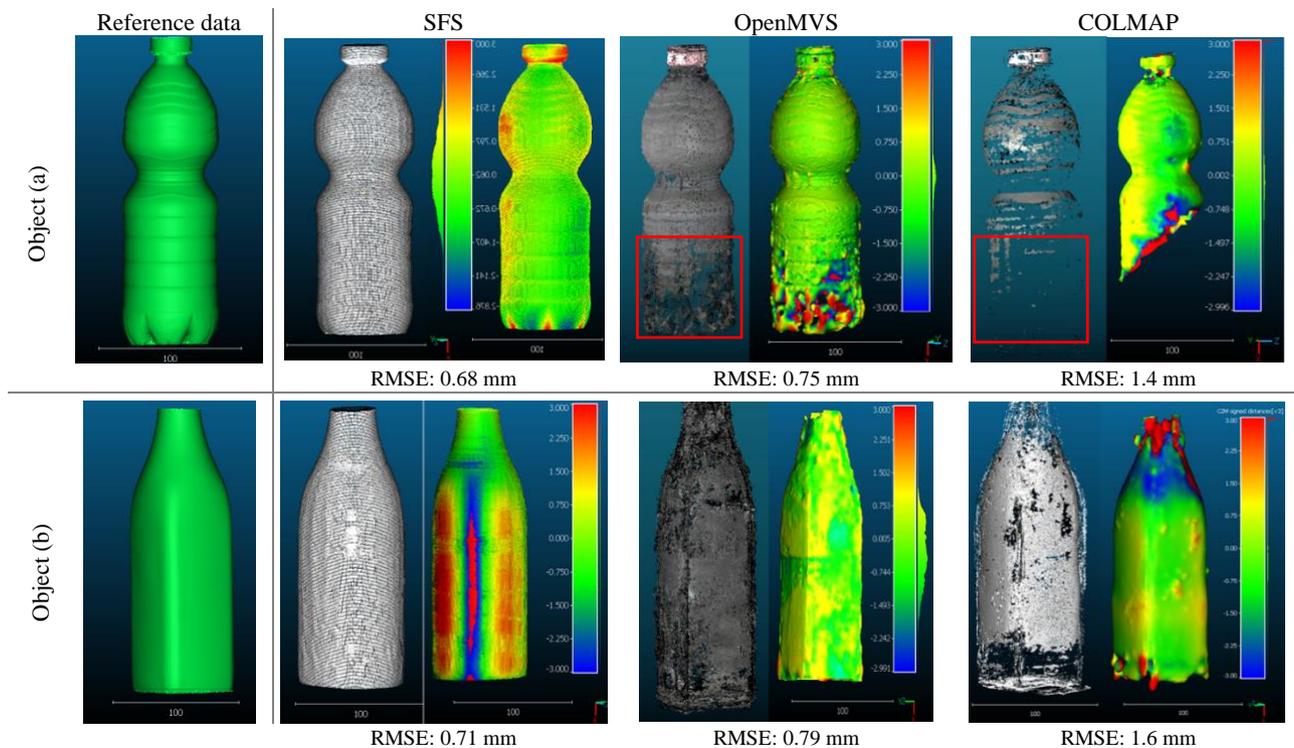


Figure 8. Results of the point-to-point comparison for objects *a* and *b* using three different dense reconstruction pipelines.

4. CONCLUSIONS AND FUTURE WORK

In this paper, two approaches were proposed based on leveraging low-contrast textures present on the surface to precisely orient the images. The first method extracted local descriptors in areas where roughness and microstructures were more prominent, the normalized cross-correlation (NCC) was then applied to the gradient map of the images to advantage of the geometrical information of the patches.

The second approach, however, modifies the classic RootSIFT pipeline to achieve faster and more reliable results. To assess the accuracy of the suggested pipelines for image orientation and dense reconstruction, three transparent objects with varying surface properties were investigated. Different criteria including mean reprojection error (MRE), the mean point-to-point error and standard deviation were used to evaluate the performance of both cross-correlation and RootSIFT descriptors. From the estimated errors, it can be concluded that RootSIFT descriptor performed slightly better than cross-correlation. After image orientation, the dense cloud was generated using various pipelines, including OpenMVS COLMAP and Shape from Silhouette (SFS), to present many viable options for 3D dense reconstruction.

As future work, we will investigate on deep-learning descriptors instead of cross-correlation and RootSIFT. Applying on more complex datasets with more than one strip or multi-camera acquisitions. Furthermore, taking our estimated poses as input, it would be intriguing to apply NERF-based techniques to generate 3D reconstructions of transparent objects.

REFERENCES

Ahmadabadian, A.H., Karami, A. and Yazdan, R., 2019. An automatic 3D reconstruction system for texture-less objects. *Robotics and Autonomous Systems*, 117, pp. 29-39.

Ahmadabadian, A.H., Yazdan, R., Karami, A., Moradi, M. and Ghorbani, F., 2017. Clustering and selecting vantage images in a low-cost system for 3D reconstruction of texture-less objects. *Measurement*, 99, pp.185-191.

Arandjelović, R. and Zisserman, A., 2012, June. Three things everyone should know to improve object retrieval. *Proc. IEEE CVPR*, pp. 2911-2918.

Baumgart, B.G., 1974. Geometric Modeling for Computer Vision. *PhD Thesis, Stanford University*.

Bellavia, F., Colombo, C., Morelli, L. and Remondino, F., 2022. Challenges in image matching for cultural heritage: an overview and perspective. *Proc. FAPER2022, Springer LNCS*.

Ben-Ezra, M. and Nayar, S.K., 2003. What does motion reveal about transparency? In *Proc. ICCV*, Vol. 3, pp. 1025-1025.

Bradski, G., 2000. The openCV library. *Dr. Dobb's Journal: Software Tools for the Professional Programmer*, 25(11), pp. 120-123.

Brahm, A., Rößler, C., Dietrich, P., Heist, S., Kühmstedt, P. and Notni, G., 2016. Non-destructive 3D shape measurement of transparent and black objects with thermal fringes. *SPIE Proc. 9868*.

Clark, A., 2015. Pillow (pil fork) documentation. *readthedocs*. Available at: <https://pillow.readthedocs.io/en/stable/reference/index.html>.

Cui, Z., Gu, J., Shi, B., Tan, P. and Kautz, J., 2017. Polarimetric multi-view stereo. *Proc. CVPR*, pp. 1558-1567.

Drouot, J.D., Falcone, M., Quéau, Y. and Tozza, S. eds., 2020. Advances in Photometric 3D-Reconstruction. *Springer International Publishing*.

Eppel, S., Xu, H., Wang, Y.R. and Aspuru-Guzik, A., 2021. Predicting 3D shapes, masks, and properties of materials,

- liquids, and objects inside transparent containers, using the TransProteus CGI dataset. *arXiv preprint arXiv:2109.07577*.
- Eren, G., Aubreton, O., Meriaudeau, F., Secades, L.S., Fofi, D., Naskali, A.T., Truchetet, F. and Ercil, A., 2009. Scanning from heating: 3D shape estimation of transparent objects from local surface heating. *Optics Express*, 17(14), pp.11457-11468.
- Hosseininaveh, A., Yazdan, R., Karami, A., Moradi, M. and Ghorbani, F., 2015. A low-cost and portable system for 3D reconstruction of texture-less objects. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 40(1), p.327.
- Huynh, C.P., Robles-Kelly, A. and Hancock, E., 2010. Shape and refractive index recovery from single-view polarisation images. Proc. *CVPR*, pp. 1229-1236.
- Ihrke, I., Kutulakos, K.N., Lensch, H.P., Magnor, M. and Heidrich, W., 2010. Transparent and specular object reconstruction. In *Computer Graphics Forum*, Vol. 29, No. 8, pp. 2400-2426. Oxford, UK: Blackwell Publishing Ltd.
- Jin, Y., Mishkin, D., Mishchuk, A., Matas, J., Fua, P., Yi, K.M. and Trulls, E., 2021. Image matching across wide baselines: From paper to practice. *International Journal of Computer Vision*, 129(2), pp.517-547.
- Karami, A., Menna, F. and Remondino, F., 2021. Investigating 3d Reconstruction of non-collaborative surfaces through photogrammetry and photometric stereo. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43, pp. 519-526.
- Karami, A., Battisti, R., Menna, F. and Remondino, F., 2022a. 3D digitization of transparent and glass surfaces: state of the art and analysis of some methods. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43, pp. 695-702.
- Karami, A., Menna, F., Remondino, F. and Varshosaz, M., 2022b. Exploiting Light Directionality for Image-Based 3D Reconstruction of Non-Collaborative Surfaces. *The Photogrammetric Record*, 37(177), pp. 111-138.
- Karami, A., Menna, F. and Remondino, F., 2022c. Combining Photogrammetry and Photometric Stereo to Achieve Precise and Complete 3D Reconstruction. *Sensors*, 22(21), p.8172.
- Kim, J., Reshetouski, I. and Ghosh, A., 2017. Acquiring axially-symmetric transparent objects using single-view transmission imaging. Proc. *CVPR*, pp. 3559-3567.
- Kutulakos, K.N. and Steger, E., 2008. A theory of refractive and specular 3D shape by light-path triangulation. *Int. Journal of Computer Vision*, 76(1), pp.13-29.
- Landmann, M., Speck, H., Schmieder, J.T., Heist, S. and Notni, G., 2021. Improvement of Thermal Fringe Projection for Fast and Accurate 3D Shape Measurement of Transparent Objects. *of Materials*, p.99.
- Li, Z., Yeh, Y.Y. and Chandraker, M., 2020. Through the looking glass: neural 3D reconstruction of transparent shapes. Proc. *CVPR*, pp. 1262-1271.
- Liu, X., Jonschkowski, R., Angelova, A. and Konolige, K., 2020. Keypose: Multi-view 3d labeling and keypoint estimation for transparent objects. *CVPR*, pp. 11602-11610.
- Lyu, J., Wu, B., Lischinski, D., Cohen-Or, D. and Huang, H., 2020. Differentiable refraction-tracing for mesh reconstruction of transparent objects. *ACM Transactions on Graphics*, 39(6), pp.1-13.
- Miyazaki, D., Kagesawa, M. and Ikeuchi, K., 2004. Transparent surface modeling from a pair of polarization images. *IEEE Transactions on PAMI*, 26(1), pp.73-82.
- Moulon, P., Monasse, P. and Marlet, R., 2013. Global fusion of relative motions for robust, accurate and scalable structure from motion. Proc. *ICCV*. pp. 3248-3255.
- Palousek, D., Omasta, M., Koutny, D., Bednar, J., Koutecky, T. and Dokoupil, F., 2015. Effect of matte coating on 3D optical measurement accuracy. *Optical Materials*, 40, pp. 1-9.
- Pereira, J.R.M., de Lima e Silva Penz, I. and da Silva, F.P., 2019. Effects of different coating materials on three-dimensional optical scanning accuracy. *Advances in Mechanical Engineering*, 11(4), p.1687814019842416.
- Qian, Y., Gong, M. and Yang, Y.H., 2016. 3d reconstruction of transparent objects with position-normal consistency. Proc. *CVPR*, pp. 4369-4377.
- Remondino, F., Menna, F. and Morelli, L., 2021. Evaluating hand-crafted and learning-based features for photogrammetric applications. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43, pp. 549-556.
- Sajjan, S., Moore, M., Pan, M., Nagaraja, G., Lee, J., Zeng, A. and Song, S., 2020. Clear grasp: 3d shape estimation of transparent objects for manipulation. Proc. *IEEE ICRA*, pp. 3634-3642.
- Schonberger, J.L. and Frahm, J.M., 2016. Structure-from-motion revisited. Proc. *CVPR*, pp. 4104-4113.
- Stets, J., Li, Z., Frisvad, J.R. and Chandraker, M., 2019. Single-shot analysis of refractive shape using convolutional neural networks. Proc. *IEEE WACV*, pp. 995-1003.
- Sun, Z., Qiao, Y., Jiang, Z., Xu, X., Zhou, J. and Gong, X., 2020. An Accurate Fourier-Based Method for Three-Dimensional Reconstruction of Transparent Surfaces in the Shape-From-Polarization Method. *IEEE Access*, 8, pp.42097-42110.
- Tanaka, K., Mukaigawa, Y., Kubo, H., Matsushita, Y. and Yagi, Y., 2016. Recovering transparent shape from time-of-flight distortion. Proc. *CVPR*, pp. 4387-4395.
- Tsai, C.Y., Veeraraghavan, A. and Sankaranarayanan, A.C., 2015. What does a single light-ray reveal about a transparent object? Proc. *IEEE ICIP*, pp. 606-610.
- Wetzstein, G., Roodnick, D., Heidrich, W. and Raskar, R., 2011. Refractive shape from light field distortion. Proc. *ICCV*, pp. 1180-1186.
- Wu, B., Zhou, Y., Qian, Y., Gong, M. and Huang, H., 2018. Full 3D reconstruction of transparent objects. *arXiv preprint arXiv:1805.03482*.
- Zhao, F., Huang, Q. and Gao, W., 2006, May. Image matching by normalized cross-correlation. In *2006 IEEE international conference on acoustics speech and signal processing proceedings*, Vol. 2, pp. II-II.
- Zhu, L., Mousavian, A., Xiang, Y., Mazhar, H., van Eenbergen, J., Debnath, S. and Fox, D., 2021. RGB-D Local Implicit Function for Depth Completion of Transparent Objects. Proc. *CVPR*, pp. 4649-465.