

3D DIGITIZATION OF TRANSPARENT AND GLASS SURFACES: STATE OF THE ART AND ANALYSIS OF SOME METHODS

Ali Karami ^{1,2*}, Roberto Battisti ¹, Fabio Menna ¹, Fabio Remondino ¹

¹ 3D Optical Metrology (3DOM) unit, Bruno Kessler Foundation (FBK), Trento, Italy
Web: <http://3dom.fbk.eu> – Email: <akarami><rbattisti><fmenna><remondino>@fbk.eu

² Dep. of Information Engineering and Computer Science, University of Trento, Italy

Commission 2, WG II/7

KEY WORDS: Photogrammetry, Shape from X, 3D reconstruction, Transparent objects, Machine learning, Inspections

ABSTRACT

In the field of industrial metrology, there is a rising need for 3D information at a very high resolution for micro-measurements and quality control of transparent objects such as glass bottles (beer, wine, cola, cosmetics, etc.). However, such objects are particularly challenging for optical-based 3D reconstruction methods and systems such as photogrammetry, photometric stereo, structured light scanning, laser scanning, typically resulting in poor metrological performances. Indeed, these methods require the surface of the object to diffusely reflect the incoming light, which is not the case with the glass material where refraction and absorption phenomena do not permit their use. Over the years, various methods have been investigated and developed to avoid the coating (or powdering) treatment often used to make transparent objects opaque and diffusely reflecting. Most of the approaches require either some a priori knowledge of the transparent object or assumptions about how light interacts with the surface. This paper provides a general overview of state-of-the-art 3D digitization methods for optically non-cooperative surfaces featuring absorption, scattering, and refraction. The paper reviews research works summarizing them into four categories including shape-from-X, direct ray measurements, hybrid, and learning-based approaches. Moreover, we provided some 3D results to better highlight the advantages and disadvantages of each method in practice when dealing with transparent objects.

1. INTRODUCTION

In the field of industrial metrology, there is a rising need for 3D information at a very high resolution for quality inspection (i.e., shape and textures) and 3D monitoring over time of manufactured products. 3D object reconstruction is generally performed using either active (range-based) or passive (image-based) methods (Blais, 2004; Remondino and El-Hakim, 2006; Remondino et al., 2013; Ahmadabadian et al., 2019; Karami et al., 2021). However, these methods are not directly applicable to the 3D reconstruction of transparent objects such as glass bottles (beverages, cosmetics, oil, etc.). This is mainly because glass does not diffusely reflect the incoming light and, also, for passive methods, do not have a texture of their own needed for image matching tasks. Instead, because of refraction and specular reflections, their appearance depends on the object's shape, surrounding background, and lighting conditions with light traveling through the surface, distorting or changing the path of

the light in the process. This makes standard techniques inappropriate, causing large errors, and most often failures in the process of 3D reconstruction (Figure 1). One of the traditional approaches generally used in industrial sectors to deal with such objects is to spray a thin layer of powder onto the object's surface (Figure 1) to make its surface opaque and diffusely reflecting (Palousek et al., 2015; Lin et al., 2017; Pereira et al., 2019). This supplementary treatment, on the other hand, is challenging, time-consuming, and may not always be feasible in real-time 3D inspection of industrial components (Pereira et al., 2019; Karami et al., 2022). Moreover, the added layer could increase the overall volume of the object and may negatively affect the final accuracy depending on the powder thickness and its homogeneity (Palousek et al., 2015; Pereira et al., 2019).

Over the years, various approaches have been investigated and developed to avoid the coating treatment. Most of them require either some a priori knowledge of the transparent object's shape or assumptions about how light interacts with the surface.

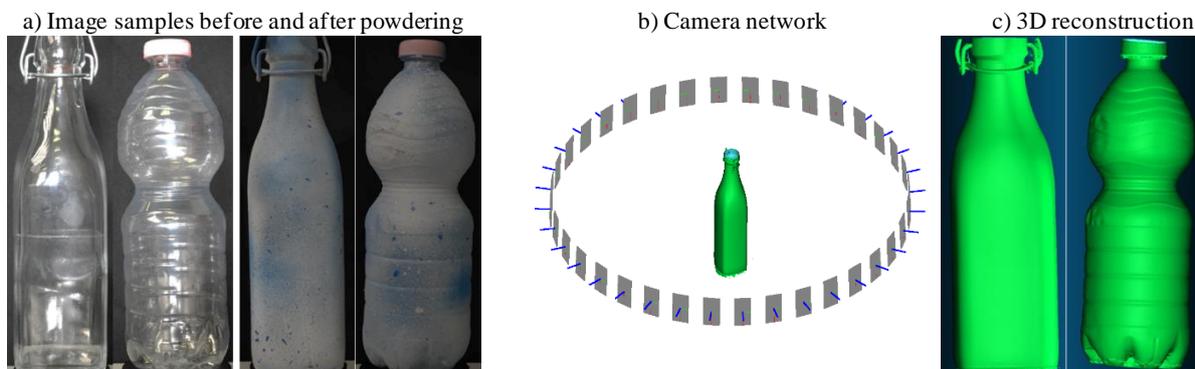


Figure 1. 3D reconstruction results for a glass bottle (400x80x80 mm) and a plastic bottle (300x70x70 mm) (a) using photogrammetry and 36 images (ground sample distance – GSD \approx 38 μ m) acquired with a turntable (b). Without powdering, no results are achieved. Only after powdering a successful 3D result can be obtained (c).

* Corresponding author

A comprehensive review was provided in (Ihrke et al., 2010), which presents a taxonomy of nine object types to describe light transport, ranging from ideally diffuse to a more complicated surface with absorption, scattering, and refraction. This work was further completed by Mériaudeau et al. (2012), in which all techniques were categorized based on the physical interactions in conjunction with developed techniques like transmission, reflection, and emission. Stolz et al. (2016) proposed a short overview of only polarimetric imaging-based methods for 3D measurements of transparent objects.

1.1 Aim of the work

This paper aims to provide an updated and general overview of 3D reconstruction approaches for optically non-cooperative surfaces, in particular transparent ones (Figure 2). It reviews the most related research works, summarizing them into different categories, including shape-from-X (distortion, silhouette, reflection, polarization, heating), direct ray measurements, hybrid, and learning-based approaches. The main benefits and drawbacks, as well as their applications in industrial sectors, are also discussed and assessed. Moreover, for some techniques, we provide a comparative or visual evaluation to show, with practical examples, the advantages and disadvantages of each method when dealing with transparent objects.

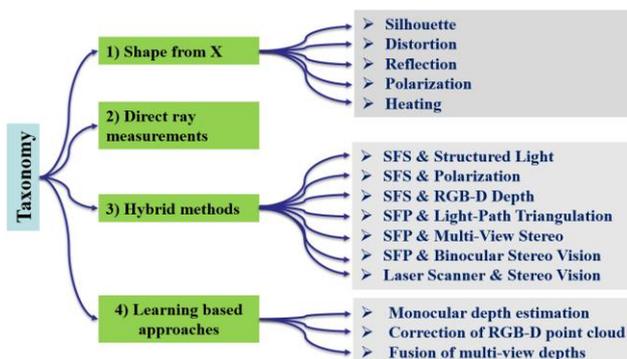


Figure 2. General taxonomy of 3D digitization of transparent and glass surfaces. SFS and SFP are the abbreviation for Shape from Silhouette and Shape from Polarization, respectively.

2. STATE OF THE ART

In this Section, we are summarizing many research works related to the 3D measurement of transparent surfaces into four different categories: shape-from-X, direct ray measurements, hybrid, and learning-based approaches.

2.1 Shape from X

Several approaches known as Shape from X techniques have been developed for extracting shape information from 2D images, where X could be distortion, Silhouette, reflection, polarization, heating, and so on.

Shape from distortion, also known as **Deflectometry**, is one of the earliest methods specifically developed for transparent objects. This technique recovers the 3D shape of an object by analyzing the distortion of a known pattern placed behind or near the surface. This approach has been investigated for long to reconstruct either mirror-like surfaces (Tarini et al., 2005), liquids (Murase, 1990; Jähne et al., 1994), or solid refractive surfaces (Ben-Ezra and Nayar, 2003; Wetzstein et al., 2011; Tanaka et al., 2016; Kim et al., 2017). The 3D reconstruction of refractive surfaces is more complex than the corresponding specular, or textureless surfaces because the ray path depends on

the refractive index in addition to the dependence on the surface normal (Wu et al., 2018; Lyu et al., 2020). These approaches are also limited to the recovery of a single refractive surface or the reconstruction of parametric surface with simple geometry and therefore are not generalizable if not with approximation to a wider range of object categories (Wu et al., 2018; Lyu et al., 2020).

Shape from Silhouette (SFS) is a well-known 3D reconstruction method applied to a wider range of object categories. This method reconstructs the 3D shape of an object using a sequence of images taken from different views, where the silhouette of the object is the sole relevant feature of the image. Depending on the geometric projection of the imaging system (e.g.: telecentric, central perspective) the silhouette of the object at each station (image) can be seen as the base of a prismatic /conic volume in three-dimensional space. The silhouette itself represents the locus of tangent points on the straight line departing from the perspective center of the camera (for a central perspective). By intersecting the pyramidal volumes, which is also known as Space Carving, a 3D reconstruction of an object can be generated. This method was first presented by Baumgart in 1974. Since then, various versions of the SFS have been proposed. For example, (Martin and Aggarwal, 1983; Kim and Aggarwal, 1986) used volumetric descriptions to represent the reconstructed shape. Following this, some works (Potmesil, 1987; Ahuja and Veenstra, 1989) used an octree data structure to speed up the 3D reconstruction process. Szeliski (1993) built a non-invasive 3D digitizer using a turntable and a single camera with SFS as the reconstruction method.

As shown in Figure 3, SFS can recover the 3D shape of an object regardless of the object's property and shape as long as the region of the object in each image is distinguishable from the background (Figure 3-red and yellow boxes demonstrate how a non-distinguishable background affects the 3D model). On the other hand, as photogrammetric 3D results directly depend on the object's surface property, the method completely failed to reconstruct the object's 3D shape before powdering the surface due to the refraction of the light (Figure 1).

However, the accuracy of SFS is directly depending on the silhouette boundary binarization, which can be done using automated or user-defined global thresholding of an image. In many cases, it might be difficult to determine the optimum threshold for distinguishing transparent objects from the background. As a result, the silhouette of an object may be reduced or increased, making the resulting 3D model smaller or larger than the real size of the object or making it noisier. To evaluate this, 3D results achieved with SFS were geometrically compared against photogrammetric data where its surface was coated, and the generated 3D model was of relatively higher accuracy than those of SFS.

The quantitative analysis in Figure 3 shows that the accuracy of the generated 3D data using the SFS approach before powdering was 0.6mm, while after coating the surface, it decreased slightly to 0.51mm. This is due to the fact that, without powdering, refraction makes it more difficult to identify the object's silhouette on the captured images and distinguish it from the background.

Moreover, another primary issue with SFS is that concavities on an object's surface remain unseen, finding it unsuitable for reconstructing the inside of a hole or concave areas (Figure 3-red and brown boxes). To deal with this issue, Zuo et al. (2015) incorporate internal occluding contours into traditional SFS methods to recover the concavities on an object's surface. Wu et al. (2018) and Lyu et al. (2020) started with an initial 3D shape reconstruction generated from traditional SFS, and then gradually optimizes the model.

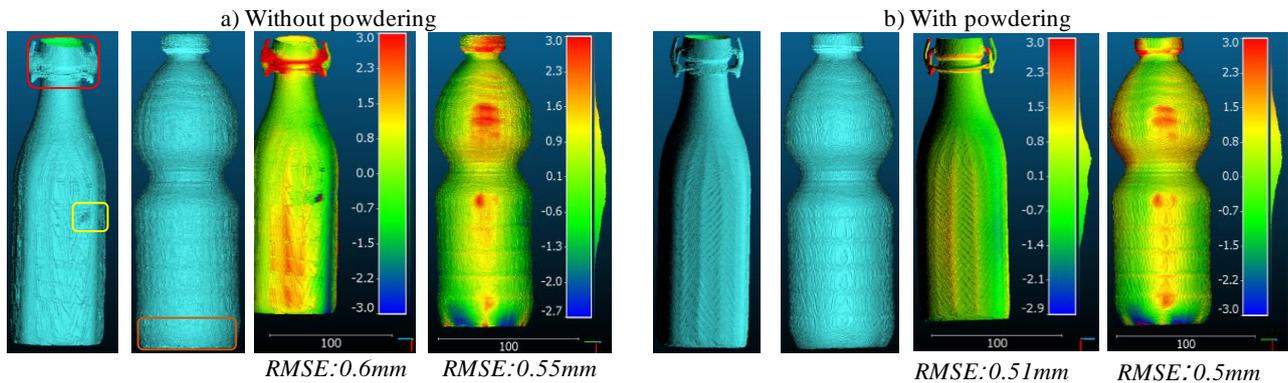


Figure 3. Quantitative analyses for the SFS-based 3D results on the two transparent objects of Figure 1. Using the 3D data achieved with photogrammetry on the powdered objects (Figure 1), the RMSE of a cloud-to-cloud comparison is computed.

Shape from reflection/refraction is also another approach introduced for the first time by Morris and Kutulakos (2007) to recover the 3D shape of the transparent object. This approach usually describes the behavior of rays as they pass through a refractive object by controlling the background behind the refractive object (Morris and Kutulakos, 2007; Yeung et al. 2011; Yeung et al. 2015; Han et al., 2015; Han et al., 2021). However, the data collection of this method may result complex and ineffective, and it is necessary to manually rotate a spotlight around the hemisphere to illuminate the object and a reference sphere from various angles. Following a similar idea, Yeung et al. (2011, 2015) used a more convenient data collection method to obtain the specular reflection information on the surface of a transparent object and applies the graph cut theory to recover and optimize the normal vectors, consequently the depth map. Although the results are insufficiently precise for industrial inspection, they are promising for 3D computer graphics animation.

Iwabuchi et al. (2011) also presented a similar method based on inverse ray-tracing. This method uses multiple sensors placed around a transparent object with simple geometry and can recover the shape and refraction index of the object. Chari et al. (2013) proposed a method that combines both geometric and radiometric information to do reconstruction. The position and direction for each light-path were recovered and combined with light radiance at the beginning and end of each light-path. More recently, Han et al. (2015, 2021) employed a single camera that was set in place with a refractive object in front of a checkerboard background. The approach required two images with the background pattern placed in two different known locations. However, the approach required a change in refractive index, necessitating immersion of the object in water, which is a significant disadvantage for industrial purposes.

Shape from Polarization (SFP) Miyazaki et al., (2002, 2003, 2004), Miyazaki and Ikeuchi (2005), Huynh et al. (2010), Cui et al. (2017), Sun et al. (2020) recover the 3D shape of an object from polarization information of the reflected light. The basic principle is that after capturing the polarization information such as the intensity, degree of polarization, polarization phase angle, the surface normal can be recovered by analyzing the relationship between the surface normal and the polarization image formation model. This method has been applied on different object types with various reflection properties such as dielectrics (Huynh et al., 2010), black (Miyazaki et al., 2016), metal (Morel et al., 2006), translucent (Chen et al., 2007) and transparent (Miyazaki and Ikeuchi, 2005; Huynh et al., 2010; Cui et al., 2017) objects. This method is also quite robust and stable to different lighting conditions such as indoors, outdoors, or under patterned illumination as long as incident light is unpolarized (Durou et al., 2020). These methods calculate surface normals, which must

afterward be converted into a height map. However, the results are highly vulnerable to noise since they depend solely on the weak shape cue supplied by polarization and do not ensure integrability (Durou et al., 2020). The ambiguity in polarization analysis is also one of the main issues for this approach. To resolve the azimuth and zenith angle ambiguity, for example, Miyazaki et al. (2002) used the polarization degree in the far-infrared wavelength for estimating the surface orientation instead of the visible wavelength. Morel et al. (2006) recommended using active lighting. Stolz et al. (2012) proposed a multispectral method for determining the optimal zenith angle, and Garcia et al. (2015) used circularly polarized light. More recently, ambiguities in this approach are adjusted by combining with other approaches in which rough geometric information is provided such as Multi-View Stereo (Miyazaki et al., 2004; Zhu and Smith, 2019) binocular stereo vision (Tian et al., 2022), light-path triangulation (Xu and Qiao, 2016; Xu et al., 2017) and etc. (Durou et al., 2020).

Shape from heating is another technique for 3D reconstruction of transparent objects (Eren et al., 2009) that, unlike the previously described approaches, ignores the refractive properties of the object. Laser range scanning of transparent objects is possible using an IR laser rather than visible light since long-wave and thermal infrared spectrum is not refracted by glass. This technique is based on the principle of infrared thermal imaging, in which the infrared source heats up the object, and then the IR-sensitive sensor detects and records the geometric surface information of the object. Aubret et al. (2013) also demonstrated a very similar approach for high specular objects utilizing high-power lasers. Since these approaches utilized single laser spots as activating light sources, their measurement areas and acquisition speed are restricted owing to the time required for scanning. There are additional limitations in spatial resolution and precision because of the size of the laser dots. To overcome these restrictions, Wiedenmann et al. (2015) developed a demonstrator system based on a CO₂ laser with a single thermal camera and phase-shifting projection technique of sinusoidal heat patterns. Brahm et al. (2016) developed a stereo-vision configuration consisting of two uncooled long-wave infrared (LWIR) cameras to detect the emitted heat radiation from an object induced by a pattern projection unit generated by a CO₂ laser. More recently, Landmann et al. (2019) demonstrated real-time 3D thermographs with a 30-frames per second frame rate (fps). This technique is well suited to applications where the geometry or temperature distribution of the objects is rapidly changing. Landmann et al. (2021) developed a simplified and robust projection approach based on a focused single thermal fringe that can rapidly scan across the object's surface. Higher intensities were obtained using such focused single thermal fringe compared to multi-fringe

projection, which increased acquisition speed while improving measurement accuracy.

2.2 Direct ray measurement

Direct ray measurement techniques, which detect light rays directly, have for long been utilized for refractive surface 3D reconstruction. Kutulakos et al. (2008) published foundational work on measuring the geometry of refractive objects using light-ray correspondences. By mapping the light rays which reach and depart from the object, the geometry of transparent objects characterized by depths and surface normal can be determined. As shown in Figure 4, The projection of a point is defined by the 3D path(s) that light would take to reach the camera, given an arbitrary 3D point p , a known viewpoint c , and a known image plane. As expressed by Kutulakos et al. (2008), refractive surface reconstruction problems are expressed as N-K-M triangulation, where N represents view-points required for reconstruction, K represents refractive surface points on a piecewise linear path, and M represents the number of calibrated reference points along the ray exiting the refractive object.

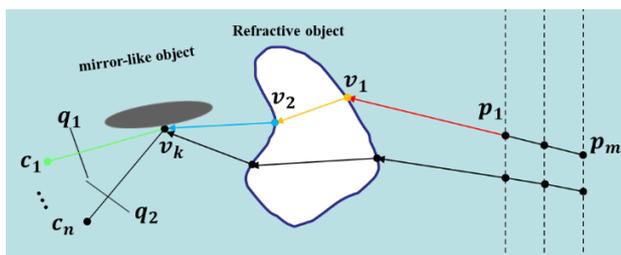


Figure 4. The geometry of N-K-M triangulation expressed by Kutulakos et al. (2008). To reach point q on the image plane, the light path from p crosses three surfaces, including refractive and mirror-like ones, passing from three vertices, v_1 , v_2 , and v_3 , which form four segments. The objective of light-path triangulation is to estimate the normals and coordinates of the vertices using the known coordinates of c , q and p .

However, methods based on light-path triangulation are known to have collinearity ambiguities as the 3D surface point can be located anywhere along the optical ray that passes through the pixel. To remove the ambiguity, Tsai et al. (2015) assumed that the light rays are refracted twice. They recovered the geometry of a transparent object with only one monocular image using a monitor controlling the background image without even needing to immerse the object in the water.

Some researchers (Morris and Kutulakos, 2011; Ding et al., 2011) employed stereo/multiple cameras to record the refractive surface, relying on a cross-view normal consistency constraint: the normals computed using the pixel-point correspondences obtained from multiple viewpoints must be consistent. Alternatively, some studies have been conducted (Ye et al., 2012; Wetzstein et al., 2011; Tsai, 2020; Tsai et al., 2021) to estimate ray-ray correspondences utilizing specific devices such as Bokode (Ye et al., 2012) and light field probes (Wetzstein et al., 2011; Tsai et al., 2021) by capturing the incident rays released from the background and the exiting rays traveling to the camera. Although 3D results appear to be highly promising, the high cost of such devices is an important downside of these approaches. In addition, one of the main common shortcomings of the aforementioned approaches is that they provide only dependable normals but noisy depths. To provide the boundary condition for the integration of normal, they need to presume a planar surface near the boundary (Ye et al., 2012; Ding et al., 2011) or approximate the border using noisy depths (Morris and

Kutulakos, 2011; Wetzstein et al., 2011). To address the restrictions mentioned above, Qian et al. (2016, 2017) propose a position-normal consistency based on a global optimization method to restore depth maps of the surface from front and back. Similarly, Kim et al. (2017) proposed a method based on optimizing the object's shape and refractive index to minimize the disparity between observed and simulated transmission/refraction rays traveling through an object. It cannot, however, be applied to any non-symmetric objects. Following that, Wu et al. (2018) expanded this technique and provided the non-intrusive method to reconstruct the whole geometry of a transparent object; nevertheless, the results are always over-smoothed due to their independent optimization and multi-view fusion of recovered point clouds. Lately, Lyu et al. (2020) expanded this work by optimizing directly the surface mesh generated from the SFS method using differentiable rendering algorithms. However, these approaches rely on feature correspondence across several views to discover similar features for triangulation, requiring more assumptions and constraints making it insufficient for actual industrial applications that must struggle with a wide range of circumstances or environments.

2.3 Hybrid methods

This group of methods includes combinations of different approaches. The primary goal of combining two techniques is to overcome the constraints of one method by leveraging the strengths of the other, allowing complete and precise 3D reconstruction of optically non-cooperative objects to be generated. For instance, SFS is considered a more suitable and practical approach to reconstruct the 3D shape of transparent objects regardless of object's property and shape. However, the concavities on an object's surface remain unseen. Therefore, some works (Kampel et al., 2002; Tosovic, 2002) have been conducted to correct the problem of SFS by combining a structured light method.

Some researchers have also tried to merge the range sensor and silhouette information to provide more reliable sensor data on transparent objects. For instance, Chiu et al. (2011) described a method to improve Microsoft Kinect depth maps by employing a cross-modal stereo path derived from disparity matching between the Kinect's included IR and RGB sensors. Narayan et al. (2015) merged the silhouette information and depth images on the 2D image domain, which can improve 3D reconstruction for concave and transparent objects with interactive segmentation. Ji et al. (2017) also combined silhouette information and depth from an RGB-D sensor to retrieve the missing surface of transparent objects. First, they seek the 3D region from multiple views that includes the transparent object using incorrect depth led by transparent materials. The 3D shape was then retrieved inside these noisy areas using SFS technology.

Another solution developed to deal with transparent surfaces is to combine SFP with other approaches such as light-path triangulation (Xu and Qiao, 2016; Xu et al., 2017), conventional raytracing (Miyazaki et al., 2007), Multi-View Stereo (Miyazaki et al., 2004; Zhu and Smith, 2019), binocular stereo vision (Tian et al., 2022). For instance, Miyazaki et al. (2007) developed a polarization raytracing approach, which combines traditional raytracing (calculates the path of light rays) with SFP (calculates the polarization state of the light). Starting with an initial shape of the transparent object, by modifying the shape, the difference between the input polarization data and the rendered polarization data obtained by polarization raytracing was minimized.

More recently, He et al., (2022) developed a pipeline based on the fusion of the laser tracking frame to frame (LTFtF) method and stereo vision to distinguish and extract the reflected laser

lines on the front surface from several laser reflection candidates caused by the refraction of the transparent objects.

2.4 Learning-based methods

Recently, many researchers have used (machine or deep) learning-based approaches to solve the problem of measuring 3D transparent objects. These approaches could be categorized into three groups as follows.

2.4.1 Multi-view 3D reconstruction

Li et al. (2020) suggested a physically-based network for generating the 3D geometry of transparent objects using multiple images acquired from different viewpoints while also taking into account light transport patterns. More similar to Lyu et al. (2020), this method (Li et al., 2020) optimizes surface normals corresponding to a back-projected ray from both sides of the object using an in-network differentiable rendering layer, given the visual hull construction as an initial 3D reconstruction. Despite the fact that their method is less restrictive than previous ones (Wu et al., 2018; Lyu et al., 2020) that utilized multi-view images, it still requires the environment map and the object's refraction index. It is also difficult to be used in real-time applications because of the time-consuming optimization procedure. Furthermore, these data-driven algorithms rely on training synthetic images since getting a significant quantity of real image training data is difficult (Lyu et al., 2020).

2.4.2 Depth completion (from partial RGB-D depths)

These approaches use different learning-based methods to fill in missing depths (where transparent objects are) acquired with an RGB-D sensor (Figure 5). Sajjan et al. (2020) presented a deep learning approach (named ClearGrasp) for predicting the 3D geometry of transparent objects partly surveyed with an RGB-D sensor. Deep networks are used to identify masks, occlusion borders, and surface normals given RGB images, and then the initial depth is optimized using the network predictions. The optimization, however, needs transparent objects having interaction boundaries with non-transparent objects. Otherwise, the depth of the transparent region remains unpredictable. Figure 5 shows an example of a depth completion using the method of Sajjan et al. (2020): the missing parts of the scene (where both transparent objects are located) are predicted and the new point cloud is more complete.

Zhu et al. (2021) proposed another learning-based technique which uses a local implicit neural representation built on ray-voxel pairs that can generalize to unseen objects and fill in missing depth on given noisy depth maps.

2.4.3 Monocular shape prediction

This group of approaches requires only a single image as input in order to predict the 3D shape of transparent objects. Stets et al. (2019) proposed a deep convolutional neural network (CNN) method for determining depths and normals of a transparent object using a single image obtained under an arbitrary environment map. More recently, Eppel et al. (2022) presented a method for predicting 3D points of transparent objects straight from an image taken from unknown source using an advanced neural net that is independent of camera parameters. In this method, each pixel in the predicted map is assigned with the X, Y, Z coordinates of a point rather than the distances to that point. To train the net, 50k transparent container images containing 13k different objects, 500 different environments, and 1450 material textures were utilized. A total of 104 real-world transparent images of various containers with depth maps were also utilized. Instead of using absolute XYZ coordinates to calculate the training loss, the distance between pairs of points inside the 3D

model was utilized, making the loss function translation invariant.

Unlike previous methods, this approach does not require camera parameters and can work with images from unknown cameras. The method was designed for specific manipulation applications of transparent chemical bins but with specific re-training operations, it could be generalized to other objects. Figure 6 shows some results obtained using the method presented in Eppel et al. (2022). It can be seen that the predicted 3D shape is only an approximate 3D shape with also anisotropic scaling issues remaining unsolved.

3. CONCLUSIONS

This paper presented a general overview of 3D digitization methods for non-cooperative surfaces featuring absorption, scattering, and refraction. The paper reviewed the most relevant research works, summarizing them into four categories, including shape-from-X, direct ray measurements, hybrid, and learning-based approaches. Shape from silhouette has become a popular 3D reconstruction method for static objects. Besides, the accuracy of SFS is directly dependent on the binarization of acquired images, making it difficult to determine the optimum threshold for distinguishing transparent objects from the background.

Shape from reflection may also be used to generate a precise 3D reconstruction of transparent objects with complex, inhomogeneous interiors when additional limitations are taken into account. However, data collection is more complex making it difficult and unsuitable for many real-time applications.

Approaches based on the shape from heating appear to be very promising for 3D reconstruction of optically non-cooperative objects compared to other approaches. However, their resolution is limited since the incident illumination's wavelength is significantly longer than visible light. In addition, high-resolution IR cameras are quite expensive.

SFP is a quite accurate method and its recent integration with other technologies like light-path triangulation, traditional raytracing, and Multi-View Stereo indicates some further possibilities.

Most light path-based approaches for refractive object shape reconstruction rely on feature correspondence across several views to discover similar features for triangulation. To simplify and make the problem more affordable, the majority of the techniques use assumptions and constraints. As a result, the application window of their approach becomes too limited, making them unstable and untrustworthy for actual industrial applications that must struggle with a wide range of circumstances or environments.

More recent learning-based approaches learn directly from real or synthetic training data and do not require assumptions or constraints such as controlled data acquisition, darkroom environments, or other limitations. Nevertheless, these approaches are still significantly less accurate than traditional methods, making them unsuitable for industrial applications where accuracy, reliability, and traceability of 3D measurements are mandatory. In our experience, learning-based approaches can deliver a rough 3D shape of transparent objects which can only be utilized for low accuracy applications. Nevertheless, learning-based methods have demonstrated encouraging results with a cost-effective and re-trainable approach.

However, training such algorithms need huge datasets annotated for the specific case. And there is still a gap between real-world and synthetic images, making it difficult to generalize to real-world input datasets.

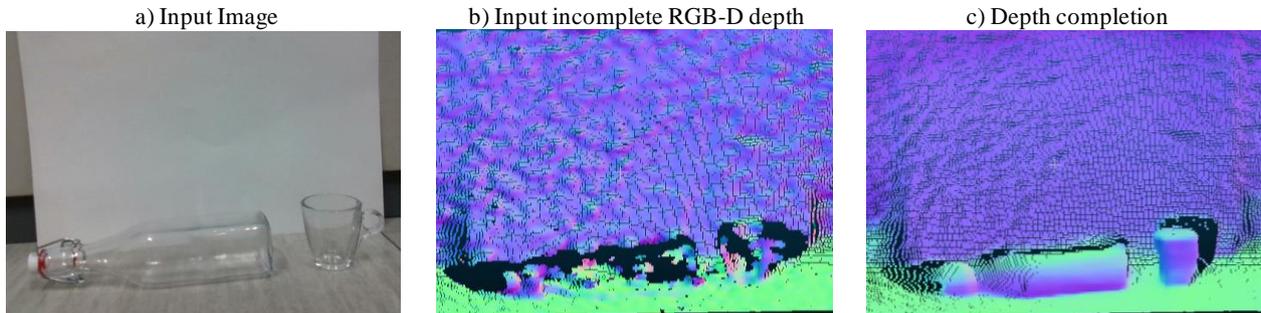


Figure 5. Depth completion using a learning-based method (Sajjan et al., 2020): a depth map of a scene with two transparent objects (glass bottle and tea cup) placed at the scene. Given an RGB input image (a) with an uncompleted depth map (b), the missing areas from input depth were predicted (c).

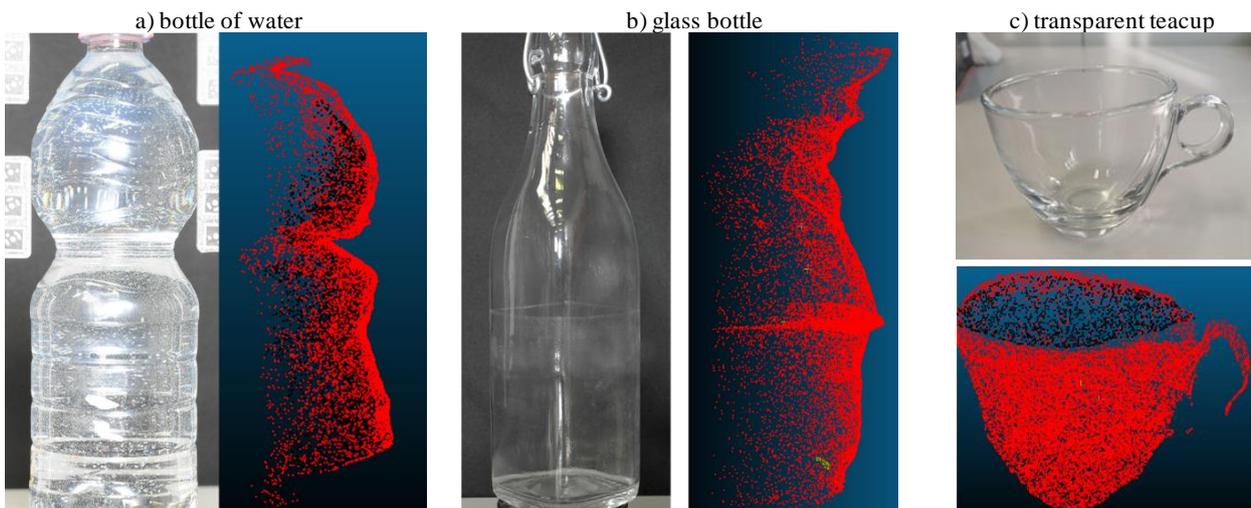


Figure 6. Learning-based 3D reconstruction of three transparent objects (bottle of water, glass bottle, and teacup) from a single image based on Eppel et al. (2022).

ACKNOWLEDGEMENTS

We are grateful to Simone Rigon (3DOM-FBK) for his valuable technical contribution in the experiments related to depth completion and shape prediction.

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.
- Ahuja, N. and Veenstra, J., 1989. Generating octrees from object silhouettes in orthographic views. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(2), pp.137-149.
- Aubreton, O., Bajard, A., Verney, B. and Truchetet, F., 2013. Infrared system for 3D scanning of metallic surfaces. *Machine Vision and Applications*, 24(7), pp.1513-1524.
- Baumgart, B.G., 1974. Geometric Modeling for Computer Vision. *PhD Thesis, Stanford University*.
- Ben-Ezra, M. and Nayar, S.K., 2003. What does motion reveal about transparency?. In *Computer Vision, IEEE International Conference on (Vol. 3, pp. 1025-1025)*. IEEE Computer Society.
- Blais, F., 2004. Review of 20 years of range sensor development. *Journal of Electronic Imaging*, 13(1), pp.231-243.
- 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. In *Dimensional Optical Metrology and Inspection for Practical Applications V (Vol. 9868, p. 98680C)*. International Society for Optics and Photonics.
- Chari, V. and Sturm, P., 2013. A theory of refractive photo-light-path triangulation. *Proc. CVPR*, pp. 1438-1445.
- Chen, T., Lensch, H.P., Fuchs, C. and Seidel, H.P., 2007. Polarization and phase-shifting for 3D scanning of translucent objects. *Proc. CVPR*, pp. 1-8.
- Chiu, W.W.C., Blanke, U. and Fritz, M., 2011. Improving the Kinect by Cross-Modal Stereo. In *BMVC (Vol. 1, No. 2, p. 3)*.
- Cui, Z., Gu, J., Shi, B., Tan, P. and Kautz, J., 2017. Polarimetric multi-view stereo. *Proc. CVPR*, pp. 1558-1567.
- Ding, Y., Li, F., Ji, Y. and Yu, J., 2011. Dynamic fluid surface acquisition using a camera array. *Proc. ICCV*, pp. 2478-2485.
- Durou, 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.
- Garcia, N.M., De Erasquin, I., Edmiston, C. and Gruev, V., 2015. Surface normal reconstruction using circularly polarized light. *Optics Express*, 23(11), pp.14391-14406.
- Han, K., Wong, K.Y.K. and Liu, M., 2015. A fixed viewpoint approach for dense reconstruction of transparent objects. *Proc. CVPR*, pp. 4001-4008.
- Han, K., Wong, K.Y.K. and Liu, M., 2021. Dense Reconstruction of Transparent Objects by Altering Incident Light Paths Through Refraction. *arXiv e-prints*, pp.arXiv-2105.
- He, K., Sui, C., Huang, T., Dai, R., Lyu, C. and Liu, Y.H., 2022. 3D Surface reconstruction of transparent objects using laser scanning with LTF method. *Optics and Lasers in Engineering*, 148, p.106774.
- 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.
- Iwabuchi, Y., Li, L., Baba, M. and Ohtani, K., 2011. 3D shape measurement of a transparent object with unknown refractive index by inverse ray tracing method. In *SICE Annual Conference 2011*, pp. 16-21.
- Jähne, B., Klinke, J. and Waas, S., 1994. Imaging of short ocean wind waves: a critical theoretical review. *JOSA A*, 11(8), pp.2197-2209.
- Ji, Y., Xia, Q. and Zhang, Z., 2017. Fusing depth and silhouette for scanning transparent object with RGB-D sensor. *International Journal of Optics*.
- Kampel, M., Tosovic, S. and Sablatnig, R., 2002. Octree-based fusion of shape from silhouette and shape from structured light. In *Proceedings. First International Symposium on 3D Data Processing Visualization and Transmission*, pp. 754-757.
- 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., Menna, F., Remondino, F. and Varshosaz, M., 2022. Exploiting light directionality for image-based 3d reconstruction of non-collaborative surfaces. *The Photogrammetric Record*.
- Kim, J., Reshetouski, I. and Ghosh, A., 2017. Acquiring axially-symmetric transparent objects using single-view transmission imaging. *Proc. CVPR*, pp. 3559-3567.
- Kim, Y. and Aggarwal, J., 1986. Rectangular parallelepiped coding: A volumetric representation of three-dimensional objects. *IEEE Journal on Robotics and Automation*, 2(3), pp.127-134.
- Kutulakos, K.N. and Steger, E., 2008. A theory of refractive and specular 3D shape by light-path triangulation. *International Journal of Computer Vision*, 76(1), pp.13-29.
- Landmann, M., Heist, S., Kühmstedt, P. and Notni, G., 2019. 3D shape from thermal patterns: investigation of projection parameters in simulation and experiment. In *Optical Measurement Systems for Industrial Inspection XI (Vol. 11056, p. 1105615)*. International Society for Optics and Photonics.
- 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.
- Lin, H., Gao, J., Zhang, G., Chen, X., He, Y. and Liu, Y., 2017. Review and comparison of high-dynamic range three-dimensional shape measurement techniques. *Journal of Sensors*, 9576850.
- 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.
- Martin, W.N. and Aggarwal, J.K., 1983. Volumetric descriptions of objects from multiple views. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. (2), pp.150-158.
- Mériaudeau, F., Rantoso, R., Fofi, D. and Stolz, C., 2012. Review and comparison of non-conventional imaging systems for three-dimensional digitization of transparent objects. *Journal of Electronic Imaging*, 21(2), p.021105.
- Miyazaki, D., Saito, M., Sato, Y. and Ikeuchi, K., 2002. Determining surface orientations of transparent objects based on polarization degrees in visible and infrared wavelengths. *JOSA A*, 19(4), pp.687-694.
- Miyazaki, D., Kagesawa, M. and Ikeuchi, K., 2003. Polarization-based transparent surface modeling from two views. *Proc. ICCV*, Vol. 3, pp. 1381-1381.
- Miyazaki, D., Kagesawa, M. and Ikeuchi, K., 2004. Transparent surface modeling from a pair of polarization images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(1), pp.73-82.
- Miyazaki, D. and Ikeuchi, K., 2005. Inverse polarization raytracing: estimating surface shapes of transparent objects. *Proc. CVPR*, Vol. 2, pp. 910-917.
- Miyazaki, D. and Ikeuchi, K., 2007. Shape estimation of transparent objects by using inverse polarization ray tracing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(11), pp. 2018-2030.
- Miyazaki, D., Shigetomi, T., Baba, M., Furukawa, R., Hiura, S. and Asada, N., 2016. Surface normal estimation of black specular objects from multiview polarization images. *Optical Engineering*, 56(4), p.041303.
- Murase, H., 1992. Surface shape reconstruction of an undulating transparent object. *Proc. ICCV*.
- Morel, O., Stolz, C., Meriaudeau, F. and Gorria, P., 2006. Active lighting applied to three-dimensional reconstruction of specular metallic surfaces by polarization imaging. *Applied Optics*, 45(17), pp.4062-4068.
- Morris, N.J. and Kutulakos, K.N., 2007. Reconstructing the surface of inhomogeneous transparent scenes by scatter-trace photography. *Proc. ICCV*, pp. 1-8.

- Morris, N.J. and Kutulakos, K.N., 2011. Dynamic refraction stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(8), pp.1518-1531.
- Narayan, K.S., Sha, J., Singh, A. and Abbeel, P., 2015. Range sensor and silhouette fusion for high-quality 3D Scanning. *Proc. IEEE ICRA*, pp. 3617-3624.
- 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.
- Potmesil, M., 1987. Generating octree models of 3D objects from their silhouettes in a sequence of images. *Computer Vision, Graphics, and Image Processing*, 40(1), pp.1-29.
- Qian, Y., Gong, M. and Yang, Y.H., 2016. 3d reconstruction of transparent objects with position-normal consistency. *Proc. CVPR*, pp. 4369-4377.
- Qian, Y., Gong, M. and Yang, Y.H., 2017. Stereo-based 3D reconstruction of dynamic fluid surfaces by global optimization. *Proc. CVPR*, pp. 1269-1278.
- Remondino, F. and El-Hakim, S., 2006. Image-based 3D modelling: a review. *The Photogrammetric Record*, 21(115), pp.269-291.
- 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.
- Stolz, C., Ferraton, M. and Meriaudeau, F., 2012. Shape from polarization: a method for solving zenithal angle ambiguity. *Optics Letters*, 37(20), pp.4218-4220.
- Stolz, C., Kechiche, A.Z. and Aubreton, O., 2016. Short review of polarimetric imaging based method for 3D measurements. *In Optics, Photonics and Digital Technologies for Imaging Applications IV*, Vol. 9896, p. 98960P.
- 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.
- Szeliski, R., 1993. Rapid octree construction from image sequences. *CVGIP: Image Understanding*, 58(1), pp.23-32.
- 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.
- Tarini, M., Lensch, H.P., Goesele, M. and Seidel, H.P., 2005. 3D acquisition of mirroring objects using striped patterns. *Graphical Models*, 67(4), pp.233-259.
- Tsai, D.Y.P., 2020. Light-field features for robotic vision in the presence of refractive objects, *Doctoral Dissertation, Queensland University of Technology*.
- 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.
- Tsai, D., Corke, P., Peynot, T. and Dansereau, D.G., 2021. Refractive Light-Field Features for Curved Transparent Objects in Structure from Motion. *IEEE Robotics and Automation Letters*, 6(4), pp. 6923-6930.
- Tian, X., Liu, R., Wang, Z. and Ma, J., 2022. High quality 3D reconstruction based on fusion of polarization imaging and binocular stereo vision. *Information Fusion*, 77, pp.19-28.
- Tosovic, S., 2002. Adaptive 3D modeling of objects by combining shape from silhouette and shape from structured light. *Technical Report, PRIP, TU Wien, Austria*.
- 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*.
- Xu, X. and Qiao, Y., 2016. 3D reconstruction of transparent objects by direct polarized light measurements. *Proc. IEEE PIC*, pp. 306-310.
- Xu, X., Qiao, Y. and Qiu, B., 2017. Reconstructing the surface of transparent objects by polarized light measurements. *Optics Express*, 25(21), pp.26296-26309.
- Ye, J., Ji, Y., Li, F. and Yu, J., 2012. Angular domain reconstruction of dynamic 3d fluid surfaces. *Proc. CVPR*, pp. 310-317.
- Yeung, S.K., Wu, T.P., Tang, C.K., Chan, T.F. and Osher, S., 2011. Adequate reconstruction of transparent objects on a shoestring budget. *Proc. IEEE CVPR*, pp. 2513-2520.
- Yeung, S.K., Wu, T.P., Tang, C.K., Chan, T.F. and Osher, S.J., 2015. Normal Estimation of a Transparent Object Using a Video. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 37(04), pp.890-897.
- Zhu, D. and Smith, W.A., 2019. Depth from a polarisation+RGB stereo pair. *Proc. CVPR*, pp. 7586-7595.
- 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. IEEE CVPR*, pp. 4649-4658.
- Zuo, X., Du, C., Wang, S., Zheng, J. and Yang, R., 2015. Interactive visual hull refinement for specular and transparent object surface reconstruction. *Proc. IEEE CVPR*, pp. 2237-2245.