An abstraction to 3D Reconstruction based on visual properties

Anonymous WACV submission

Paper ID ****

Abstract

3D reconstruction is an on-going research topic for decades, yet there is no easy way to apply the state-of-the-art techniques to everyday objects for various applications. This paper proposed a visual property based description that serves as an abstraction layer to find the best suited 3D reconstruction technique for an object with specific visual properties. We used reflection and geometry related visual properties to describe objects and select the best suited technique. We used off-the-shelf hardware as our capture system, chose 3 different reconstruction techniques and 6 everyday objects as the target objects to show the validity of our concept.

1. Introduction

3D reconstruction of real objects or scene has been an on-going research topic for decades. It has wide range of applications: the visual effects industry uses image-based dense 3D reconstruction for close range 3D scanning [3], digital mapping is another field, in which multi-view stereo (MVS) plays a crucila role, and Virtual Reality is in need of photorealistic 3D modelling of the world to provide a more realistic experience. Of which, creating photorealistic models of everyday objects for Fish-tank VR interests us the most since it provides more recreational and creative possibilities compared to immersive VR. Potential applications include but not limited to, exhibition of virtual replicas of cultural relics for the purpose of protection, 3D gaming which allows direct interaction with the virtual world such as 3D Minecraft, or a 3D design workspace that allows designers to directly view, interact or modify their design in 3D space.

Image based 3D reconstruction is an on-goting research direction that has significant achievements and is becoming increasingly robust, especially in the field of MVS, which estimates the most likely 3D shape that explains those images, under the assumptions of known materials, viewpoints, and lighting conditions. A typical MVS technique reconstructs the geometry of the object or scene from the camera parameters and a set of correspondences. Since camera calibration can be done either manually using a calibration pattern or automatically using SfM technique, solving for the 3D geometry of the scene is equivelent to solving the correpondence problem across the input images. However, dense binocular stereo correspondene is a challenging problem in its own right, it becomes even more complicated in the MVS scenario because 1) it's hard to generate possible corresponding candidates since we don't have the visibility prior 2) it's challenging to measure the similarity of a given candidate for surfaces with high specularity or uniform/periodic texture. Under some assumptions and constraints, for instance Lambertian reflection and textured surface, it's possible to achieve a satisfactory result. The early work is limited to images obtained from the lab setting, e.g. voxel colouring [31], space carving [20], optimization based techniques, then moved to outdoor small scale scenes [34], e.g. a building facade, a fountain, a statue, and scaled up to large scenes, e.g., Colosseum in Rome [1]. Thanks to the Middlebury dataset, more state-of-the-art algorithms are still being developed.

Another thread of research focuses on obtaining depth from a projector/emitter and camera/receiver pair. Controllable lighting is utilized to ease the task of correspondence searching. Laser scanner projects a single laser point onto the surface. If the position and orientation of the laser and camera are known, the 3D surface point is found as the intersection. In order to fasten the process of scanning one point at a time, more complicated patterns are utilized, such as laser strip, and spatially or temporally encoded structured light patterns. The problem is that it's highly sensitive to the surface properties, and not suitable to dynamic scenes. An alternative approach, time-of-flight, estimates the distance to a surface by using the time elapsed between the emitting and receiving of a signal. The drawback is that the depth resolution and accuracy remain below that of laser scanner and structured lighting.

Do these previous successes of 3D reconstruction techniques indicates that they're ready to be used out of box to various applications? Far from it, because 1) effective use of these methods requires extensive knowledge of how algorithms work and how their parameters affect the results, thus a basic understanding of the field or related area is a prerequisite, 2) due to the complexity of 3D reconstruction, most algorithms or techniques are only applicable to a limited group of objects, for instance, typical MVS only works for near-Lambertial, textured objects, visual hull works for textureless objects but fails on concave ones, active sensing fails when high specularity and interreflection exist, etc. All those previous efforts focused mainly on the algorithm side of the problem trying to improve the performance on some benchmark datasets instead of applying or embedding the techniques into real-world applications.

What's the main challenge to achieve this goal? It's not that we don't have a robust technique for a specific category of objects, but we don't have a suite of techniques that can work cohesively together to deal with objects with diverse and varying properties. Though the problem of 3D reconstruction remains challenging and deserve continuing research efforts, it's also important to come up with a framework to integrate as many technique as possible in a cohesive and complementary way to ease the task of building 3D models for everyday objects. This paper intends to bridge the gap between state-of-the-art algorithms and a robust, widely applicable 3D capturing system.

The contribution of the paper is a visual property based description applied as an abstraction to the 3D reconstruction problem, to choose the best suited technique for a specific object. We first investigate the visual properties that play a crutial role in reconstruction, then choose 3 categories of techniques for reconstruction and analyze how they are affected by each visual property, and propose a framework that chooses the best suited algorithm based on the visual properties of the object.

The advantages of this high level abstraction are: 1) Developers can focus on their applications main task, rather than the algorithms; 2) Advances in the state-of-the-art can be incorporated into existing systems without re-implementation; 3) Hardware acceleration of algorithms may be used transparently; 4) The limitations of a particular platform can be taken into account automatically e.g. mobile devices may require a set of low-power consuming algorithms; 5) Computer vision expertise can be more readily adopted by researchers in other disciplines and general developers. If any abstraction is used to access vision methods, hardware and software developers of the underlying mechanisms are free to continually optimise and add new algorithms [25, 24].

2. Related Work

Many attempts have been made to develop computer vision or image processing frameworks that support rapid development of vision applications. There are many open vision libraries that provide common vision functionality, such as OpenCV [6], VLFeat [35], XVL [18], etc. These libraries often provide utilities such as camera capture or image conversion as well as suites of algorithms, which has previously been shown to lessen the effectiveness on application [23]. All of these software frameworks and libraries provide vision components and algorithms without any context of how and when they should be applied, and so often require expert vision knowledge for effective use. In the case of image-based reconstruction, especially in the MVS field, we've witnessed many successful softwares: One of the most widely used open source softwares is PMVS developed by Furukawa [11], which is used not only by computer vision/graphics engineers, but also production companies like Industrial Light & Magic, and Google, etc. It's often used together with Bundler, which is a Structure from Motion software that estimate camera parameters from images developed by Noah Snavely [32], and Poisson Surface Reconstruction developed by Michael Misha Kazhdan, which is a surface mesh software that estimate the triangulated surface from oriented point cloud [19]. Some other notable open source softwares include VisualSfM [37], CMP-MVS [12], MVE [9], and openMVG [27]. However, effective use of those software requires a basic understanding of the relevant domain, including feature detection, matching, camera calibration, dense correspondence search, etc. This current situation motivates us to provide an description-based abstraction for non-vision experts to access the state-of-the-art techniques in their own applications, but first we need an general picture of the current development of 3D reconstruction field.

3D reconstruction has been a major research topic for decades, a wide range of techniques have been developed, which can be categorized into static and active techniques. For the active techniques, the position, direction or pattern of the light sources are specially controlled. For the static techniques, light is not controlled, and typically works with ambient light [26]. Another framework called spacetime stereo unifies the active and static approaches by incorporating infomation from both the spatial and temporal domains [7]. This following review is organized as the former way of categorization.

Passive techniques don't require special lighting, and works with reasonable ambient light, thus are not restricted by the environment. There are several visual cues that can be used for reconstruction: texture, defocus, contours, silhouette, and stereo correspondence. Among these, MVS, which is based on stereo correspondence, proves to be the most successful.

One of the early reviews of MVS by Dyer investigated volemetric scene reconstruction techniques from multiple viewpoints [8]. The object or scene is reconstructed using shape from silhouette or shape from photo-consistency techniques. The state-of-art algorithm were voxel colouring [31] and space carving [20]. Those techniques represent the scene as a volume of voxels, which are time and memory demanding, thus limited to small scale objects. Seitz [30] gave a thorough taxonomy for MVS techniques, introduced a benchmark dataset, and a evaluation methodology for the first time. The algorithms at that time used varied priors and initializations, had varied ways to represent the objects, and different reconstruction approach to maximize the photo-consistency measure, the state-of-the-art algorithms include PMVS proposed by Furukawa [11], and many optimisation based techniques, including level set, and graph cut based techniques. Later, as the improvement of digital camera and computational power, the MVS technique enters the era of high resolution, medium scale, outdoor scenes, Strecha proposed a dataset for medium size, outdoor scene [34] and Hiep proposed a new pipeline to deal with the problem that MVS techniques don't scale well for high resolution large scale objects. Their proposed pipeline achieved highly detailed reconstruction within reasonable amount of time [14]. In the meantime, another exciting direction is to take advantage the increasing volume of Internet photos to reconstruct a city-scale scene, which is enabled by the matured SfM technique developed by Snavely et al. [32, 33, 1]. We can clearly see the development of MVS from a lab-controlled, small scale environment to a medium scale, outside scene, to a large-scale city-scale reconstruction. The future direction of MVS would be to achieve much higher reconstruction accuracy and real-time performance.

Active depth sensing technique needs to project a detectable pattern onto the surface to help correspondence searching, and usually consists of projector/emitter and camera/receiver pair. The current techniques include laser point or strip scanner, structured light and time-of-flight techniques.

The laser scanners typically project a spot or a strip on the surface that can be easily detected by the camera. If both the position and orientation of these two devices are known, the position of the scene point can be found by triangulation. This technique is super slow since only one spot/strip on the surface is estimated at a time. The Digital Michelangelo Project led by Marc Levoy was done by using the laser scanner, and it took the team 2 years for planning, and another 7 months to scan the David statue [22].

Although it's possible to project a plane of light onto the surface, there should be a way to encode into the pattern to make distinction between each pixel. One way is to encode each row/column of pixel a different colour/intensity, but interference by the surface colour/intensity makes it challenging to identify the orignal pattern in the image. Alternatively, one can project several strip patterns in sequence so that each pixel has a unique sequence of binary code. Bouguet and Perona developed a simple and inexpensive method for extracting the 3D shape of objects [5]. It is based on weak structured lighting, and differs from the conventional structured lighting approaches in that it requires very little hardware besides the camera, a desk-lamp, a pencil and a checkerboard. The user moves a pencil in front of the light source casting a moving shadow on the object. The 3D shape of the object is extracted from the spatial and temporal location of the observed shadow. The well organized tutorial from Brown University gave a detailed and extensive introduction to building a structured light scanner from scratch [21]. IBM's Pieta Project was conducted using the structured light techniques [4].

The basic principle of time-of-flight technique is the measurement of the duration between the emission of a modulated signal and its return to the receiver. Depending on the types of signal sent out, there are radar (electromagnetic waves), sonar(acoustic waves), or optical radar (optical electromagnetic waves, including near-infrared). One disadvantage is that the surface shouldn't have interreflection, sub-surface scattering, or low albedo, as little light will be reflected.

Recently, several 3D depth sensing products have entered the market. However, the resolution is still limited and the accuracy of the depth is moderate even under the most ideal circumstances.

3. Description-based 3D Reconstruction

The primary goal of our contribution is to provide developers without vision expertiese with intuitive access to sophisticated 3D reconstruction techniques. We observed that many techniques or algorithms are developed under certain assumptions about object properties. For instance, MVS is for textured, Lambertian objects, visual hull is for convex objects with no particular reflectance constraints, and shape from shading is for objects with known reflection property. It's reasonable to argue that there exists a connection between the object's visual properties and the reconstruction techniques that is best suited for it, which serves as the basis for our description-based abstraction. We first investigate the visual properties used by current reconstruction techniques, then examing the strengths and weaknesses of the chosen techniques with respect to each property, then we establish a relation between the property and the reconstruction technique, which serves as the basis for the 3D reconstruction abstraction.

3.1. Abstract through visual properties

We employ a simple definition of the 3D reconstruction problem: estimate the depth/3D position of the object/scene points, directly or from a set of input images. The central idea of the abstraction is visual properties. Each property is associated with an expected quantity of how strong it is. Currently it is challenging to be highly specific with these, as the descriptors have not been evaluated to this level of detail. Therefore we employ a simple scale from None to Neutral to High to allow the user to indicate the level of property expected.

Why do we choose visual properties as the foundation of the description? Theoretically speaking, the visual properties of an object are influenced by the illumination, material and viewpoint. This is generally an ill-posed problem, but we can retrieve the geometry under certain assumption of illumination, material and viewpoint [10]. For instance, shape from shading estimate the surface orientation assuming distant illumination and viewpoint and a known reflectance property, most likely a Lambertian model for diffuse objects [29] and a double-delta model for specular ones [17]. Photometric stereo reduces the ambiguity by using more light sources with known directions [36]. Visual hull reconstructs a convex hull of the target object assuming the object is convex; MVS uses colour constancy to colourize or carve out certain voxels or texture correlation to help find stereo correspondences. Furthermore, there are other visual properties like defocus, texture distortion used in less successful techniques. In conclusion, various visual propertis the key for reconstruction.

How to describe the visual properties of objects? Both human visual and camera system perceive the world by receiving light after interacting with the object.One way to characterize the behaviour of the light is BXDF, which is BRDF for opaque object, BTDF for transparent object and BSSRDF for translucent object. However, these models generally have high dimensionality thus unsuitable as a description. We decided to use more perceptual visual properties that are familiar to human, but we're not looking for any visual properties, for instance, surface roughness is a common one, but it's of little help to 3D reconstruction.

We observed that reconstruction is always related to how the object is perceived. Objects are being seen by interacting with light in a particular and characteristic way [2]. The light pattern resulted from this interaction provides enormous visual cues to determine the shape and layout of our surroundings: the shading variation tells us the surface orientation while the occlusion informs us the relative depth. How is light structured by interacting with the objects? It is reflected when hitting on a perfect mirror, scattered on an opaque, rough surface, refracted when falls onto the still water surface, and transmitted through optic fibre. There is more to that, still water surface can mimick a mirror while a bowl of water a optical lens. The behaviour of light differs dramatically when the material and geometric properties of the object differ. Therefore the perception of the object is largely dependent on its *reflection*, *transmitance* and *geometry*, which are the visual properties we're interested in.

3.1.1 Reflectance

Lightness Surface lightness or albedo is the proportion of incident light that is reflected by the surface, and ranges from 'black' to 'white' in the grey scale axis. Color is a superset intensity, which takes account into the spectral composition of light. Both terms depend on illumination, surface normal, surface reflectance, and viewing direction.

Texture Texture is something that is easy to recognize, but hard to define. Here we only consider visual textures, which is the result of shape and reflection. Therefore, a surface with varying reflectance property can produces a textured surface, a flat surface with fixed reflectance property under different illuminations can also achieve textured effect. Even very weak texture can be a strong cue to object reconstruction as manifested by the Middlebury 'Dino' dataset.

Diffusion The mechanism for diffuse reflection is complex, it could result from a rough surface, or from subsurface reflection. Regardless the cause, the light is reflected in all directions, and a special case called Lambertian, the light is distributed evenly in all directions. This property is a widely used assumption in most MVS techniques to ensure the validity of photo-consistency.

Specularity The specular surface reflects light in almost a single direction when the microscopic surface irregularities is small compared to light wavelength, and no subsurface scattering present [28]. Unlike diffuse reflections, which we experience the lightness and color of an object, specular reflections carry information about the structure, intensity, and spectral content of the illumination field. In other word, specular reflections are simply images of the environment, or the illumination field, distorted by the geometry of the reflecting surface. A purely specular surface is a mirror.

Most materials cause diffuse reflection except for smooth metal, liquid, gas, and mirror, and all object can have a proportion of specular reflection if made smooth enough. Therefore, most objects have a mixture of specular and diffuse reflection. Since the reconstruction techniques requires sufficient visual information reflected, the surface generally cannot be too dark.

3.1.2 Transmittance

Translucency Light enters the translucent material from one point and exits at another. The light that enters the object is scattered in ways specific to the material and scale of the object. Therefore, it's difficult to acquire the surface using traditional methods due to the non-locality of light transport introduced by multiple scattering beneath the object surface [16].

Transparency Transparent objects Transparent material presents unique computational challenges for the visual system. The acquisition of complete surface descriptions of refractive objects with possibly inhomogeneous material properties is very complex. The image formation for such objects is non-trivial and to data no reconstruction approaches exist for the general problem [16].

Opacity An opaque object doesn't allow light to come through, which means it's neither translucent nor transparent.

Because majority of the acquisition approaches rely on observing light reflected off a surface, objects made of materials that exhibit significant effects of global light transport are difficult to handle. Generally speaking, there is no universal technique for transparent and translucent surface [16], thus we will leave out transparency and translucency out in the following discussion.

3.1.3 Geometry

The way in which the light is reflected is dependent on, among other factors, the microscopic shape characteristics. Microscopic irregularities can, to an extent, determine the amount of diffuse and specular reflection [28]. A smooth surface may reflect incident light in a single direction, while a rough surface may scatter the light in various directions. Mesoscopic or medium scale defines the visual coarseness of the surface, and macroscopic scale defines the 'shape' of an object. Since most of the reconstruction technique don't pose a strict constraint on the geometry of the object, we only consider the scale and convexity.

3.2. Selection of techniques

3D reconstruction techniques are always connected to the reflectance, transmissive, or geometric properties of the object. When applying 3D reconstruction techniques, it's important to meet the conditions that determine the validity of these techniques. This requires an understanding of the assumption made while developing these techniques. Most of these assumptions are related to the reflective properties and macroscopic geometry. We seek to answer those questions: 1) what assumptions are inherite to this technique, 2) under what conditions, does the technique work best or worst.

3.2.1 MVS

Reflection MVS uses stereo correspondence as the main cue for reconstruction, so there should be sufficient amount of texture to match image regions with similar or same appearance [10]. Most algorithms assume Lambertian reflectance, which assume even radiance in all directions. While Lambertian reflection is only valid for smooth matte surfaces, this assumption is generally robust enough for surfaces that contain sufficient amount of diffuse component. But it's still challenging to reconstruct surface with high specularity.

Textureless or repeated textured surfaces pose another great challenge to MVS algorithms since the photo-consistenc measure evaluates the correlation of texture patches. The 'Dino' dataset from the Middlebury benchmark demonstrates that the MVS algorithms can pick up intrinsic textures, mostly caused by shading or shadowing effects [10].

Geometry MVS algorithms generally evaluate the photo-consistency measure within a window instead of a single pixel. This makes it challenging to reconstruct objects with thin structures.

In conclusion, MVS algorithms work for surface with sufficient amount of diffuse reflection, textured, and structures that are not too thin.

3.2.2 Example based photometric stereo

Reflection Photometric stereo uses multiple light sources to eliminate the ambiguity faced by shape from shading technique for surfaces with known reflectance map. Reflectance map can be obtain from known BRDF or direct measurement, which is not readily accessible for most everyday object. Hertzmann et al. [13] introduced a practical appraoch called example based photometric stereo by introducing one or multiple reference objects. Under the assumption of distant illumination condition and orthographic projection, orientation consistency should be met, which states that scene points with the same surface orientation should have same color profile. This technique can deal with objects with varying albedo and surface reflectance properties. However, since it requires the variation of reflected light, it fails on dark surfaces. The estimated continuous normal map makes the integrated height map continuous, which leads to results comparable to that of laser scanner. The estimated BRDF can also be used in rendering to achieve better visualization.

Geometry The downside of photometric stereo is that it's fails in regions when there exist interreflection or self-reflection. Thus it often fails for objects with severe concavities. It's also an one-sided or 2.5D reconstruction and has no metric meaning since the viewpoint is fixed, which makes it impossible to retrieve the true depth of the object. This will bring up a serious issue when there exists occlusion in the scene since the relative depth will be inaccurate.

In conclusion, example based photometric stereo algorithm works for relatively bright object without severe interreflection and self-reflection.

3.2.3 Active sensing

Reflection Active sensing typically involves projecting either a single laser point or a light pattern onto the object surface to assist the process of correpondence searching.

Example of active sensing include laser strip scanner, various structured light techniques and time-of-flight scanner. It typically needs a projector/emmiter and camera/receiver pair. This techniques belong to the most accurate object acquisition approaches known today. However, most of them rely on a clearly detectable pattern of light being reflected off the object's surface, thus is more sensitive to surface material or geometric properties. These technique would fail when the surface exhibits significant effects of global light transport such as specularity, tranaparency, subsurface scattering, and low albedo. Therefore, it works best for opaque surface with high albedo, low specular reflection and no interreflection.

Geometry Nowadays, a commertially available 3D depth sensor have very limited resolution, and a minimal sensing distance even though real-time depth sensing is possible. It also fails for surface with severe concavity since self-reflection is present. Therefore, active sensing could be an alternative to MVS in cases where MVS fails or a coarse reconstruction is sufficient, but generally produces poor results on small-scale objects with self-reflection.

In conclusion, active sensing techniques works for relatively bright objects that don't exhibit significant effects of global light transport.

In this paper, we'll target opaque objects with a mixture of specular and diffuse reflection, and no constraints on surface texture. We won't discuss the extremly challenging cases where the surface has very low albedo or highly specular, or both. It's still extremely challenging to select the best suited algorithm based on a higher-level description of the visual properties: there may be many ranges under which it works well; it may perform best under certain optimal circumstances, but perform well enough under other conditions. Therefore, the proposed selection might not be perfect, but it's a solid first step towards an more sophisticated and robust abstraction. Based on the performance of the 3 techniques listed above, Figure 1 shows the flow chart for selecting the optimal technique based on visual properties.

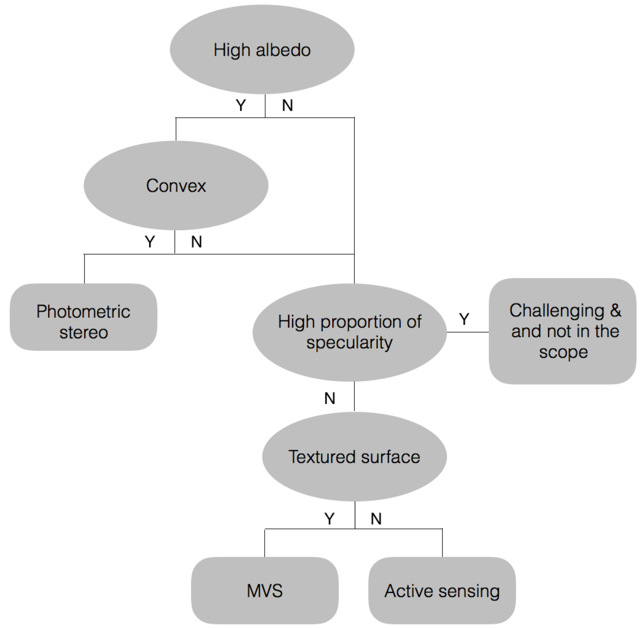

Figure 1: selection of reconstruction technique based on visual properties

4. Implementation details

We implemented our system mostly using Matlab, and used C++ to do computational heavy tasks.

MVS MVS technique consists of 2 steps: multi-camera calibration, multi-view stereo reconstruction. We used the state-of-the-art algorithms to achieve the best possible results for MVS in order to compare with those from other two techniques. We used SfM techniques to obtain the camera parameters when the target object is textured. If not, we place the object on top of a dot pattern, designed by Vogiatzis and Hernndez. We then run the CMVS/PMVS to get a dense reconstruction.

Example based PS We developed a modified version of the example based photometric stereo algorithm proposed by Hertzmann and Seitz [13]. Some modifications are made: the original paper uses a brute-force search for objects with spatially varying BRDF and took several hours to reconstruct. We employ a coarse-to-fine search strategy inspired by [15], and was able to complete the reconstruction around 10 minutes without parallelisation. In the future, we might develop a gradient descent approach to further improve the result. We observed several things to note when taking images: 1) choose a camera with larger focal length if possible to better approximate orthographic projection; 2) avoid grazing angle between the illumination and surface normal, 3) avoid interreflection between objects.

Active sensing techniques Since Kinect is a widely available depth sensor, we utilized Kinect Fusion as our active sensing technique for reconstruction. Even though Kinect has relative poor resolution, our intention is to demonstrate the conditions under which this type of techniques would work or fail, not to achieve laser scanner quality reconstructions.

5. Evaluation and Results

We select the following household objects as our targets for reconstruction, as shown in Table 1. The images are captured using a Nikon D70s camera with 70mm and 200mm lenses, the light source is a typical household lamp.

Photometric stereo techniques work for textureless surface with varying BRDF, as demonstrated by the first 2 ex-

Object	Reflection	Geometry	Best suited Technique
Cup	Bright, Glossy, Textureless	Nearly convex	PS
Bottle	Bright, Nearly diffuse, Textured with textureless area	Nearly convex	PS
Vase	Dark, Nearly diffuse, Textured	Convex	MVS
Jimu	Bright, Nearly diffuse, Textured	Non-convex	MVS
Dino	Bright, Diffuse, Textured with textureless area	Non-Convex	Active sensing
Statue	Bright, Diffuse, Textureless	Non-Convex	Active sensing

Table 1: The mapping from the visual properties to the reconstruction technique

amples. However, if fails when the object has sufficient amount of interreflection or self-reflection caused by convality or occlusion, as shown by example 4-6, and dark surface, shown by example 3.

MVS techniques work for nearly diffuse, textured surface, as shown by example 3, 4, even very intrinsic texture caused by shading. However, it fails in textureless surface, as shown in example 1, 2, 5, and 6.

Active sensing techniques works for objects lack of texture, but some have relatively low depth resolution, as demonstrated by example 5, 6, thus can be an alternative to MVS. But it fails in region with high specularity, thus causing a bump in example 3.

The results shown are consistent with our visual property based abstraction shown in Figure 1, which demonstrate the potential of visual properties as an abstration to hide the details of various 3D reconstruction techniques.

6. Conclusion

We have presented our formulation for a description-based abstraction for 3D reconstruction using visual properties. Our abstraction is designed to provide non-expert users with an interface to methods which does not expose algorithmic detail. We have demonstrated how the abstraction may be used to describe various visual properties of the target object and technique selection based on this description. We have chosen 3 methods used within 3D reconstruction fields and demonstrated how they map into our problem description, showing how it is possible for a sufficiently rich description to represent them and in turn present a useable interface to end-users.

It is our hope that other computer vision problems may be represented in this way. Representing vision problems through a description, and mapping the different problem conditions which impact algorithm performance, provides researchers and developers who are not experts at specific vision problems with access to sophisticated algorithms, without requiring direct knowledge of the field. We intend to continue developing this model to expanding the abstraction to cover new details and to provide a full proof-of-concept framework for researchers, developers and everyone else to try.

References

[1] S. Agarwal, N. Snavely, I. Simon, S. M. Seitz, and R. Szeliski. Building rome in a day. In *2009 IEEE 12th international conference on computer vision*, pages 72–79. IEEE, 2009.

[2] B. L. Anderson. Visual perception of materials and surfaces. *Current Biology*, 21(24):R978–R983, 2011.

[3] T. Beeler, F. Hahn, D. Bradley, B. Bickel, P. Beardsley, C. Gotsman, R. W. Sumner, and M. Gross. High-quality passive facial performance capture using anchor frames. In *ACM Transactions on Graphics (TOG)*, volume 30, page 75. ACM, 2011.

[4] F. Bernardini, H. Rushmeier, I. M. Martin, J. Mittleman, and G. Taubin. Building a digital model of michelangelo's florentine pieta. *IEEE Computer Graphics and Applications*, 22(1):59–67, 2002.

[5] J.-Y. Bouguet and P. Perona. 3d photography on your desk. In *Computer Vision, 1998. Sixth International Conference on*, pages 43–50. IEEE, 1998.

[6] G. Bradski and A. Kaehler. *Learning OpenCV: Computer vision with the OpenCV library*. " O'Reilly Media, Inc.", 2008.

[7] J. Davis, R. Ramamoorthi, and S. Rusinkiewicz. Spacetime stereo: A unifying framework for depth from triangulation. In *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, volume 2, pages II–359. IEEE, 2003.

[8] C. R. Dyer. Volumetric scene reconstruction from multiple views. In *Foundations of image understanding*, pages 469–489. Springer, 2001.

[9] S. Fuhrmann, F. Langguth, and M. Goesele. Mve-a multiview reconstruction environment. In *Proceedings of the Eurographics Workshop on Graphics and Cultural Heritage (GCH)*, volume 6, page 8, 2014.

[10] Y. Furukawa, C. Hernández, et al. *Multi-view stereo: A tutorial*. Citeseer, 2015.

[11] Y. Furukawa and J. Ponce. Accurate, dense, and robust multiview stereopsis. *IEEE transactions on pattern analysis and machine intelligence*, 32(8):1362–1376, 2010.

[12] J. Heller, M. Havlena, M. Jancosek, A. Torii, and T. Pajdla. 3d reconstruction from photographs by cmp sfm web service. In *Machine Vision Applications (MVA), 2015 14th IAPR International Conference on*, pages 30–34, May 2015.

(a) Original image (b) Results of MVS (c) Results of Photometric Stereo (d) Results of Kinect Fusion

Figure 2: The reconstruction results: each object is reconstructed using all 3 technique. For each object, the results of reconstruction are compared among different techniques to show the best-suited one.

[13] A. Hertzmann and S. M. Seitz. Example-based photometric stereo: Shape reconstruction with general, varying brdfs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8):1254–1264, 2005.

[14] V. H. Hiep, R. Keriven, P. Labatut, and J.-P. Pons. Towards high-resolution large-scale multi-view stereo. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 1430–1437. IEEE, 2009.

[15] Z. Hui and A. C. Sankaranarayanan. A dictionary-based approach for estimating shape and spatially-varying reflectance. In *Computational Photography (ICCP), 2015 IEEE International Conference on*, pages 1–9. IEEE, 2015.

[16] I. Ihrke, K. N. Kutulakos, H. Lensch, M. Magnor, and W. Heidrich. Transparent and specular object reconstruction.

In *Computer Graphics Forum*, volume 29, pages 2400–2426. Wiley Online Library, 2010.

[17] K. Ikeuchi. Determining surface orientations of specular surfaces by using the photometric stereo method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (6):661–669, 1981.

[18] P. V. Joe Mundy, Andrew Fitzgibbon et al. Vxl c++ libraries for computer vision research and implementation. `http://vxl.sourceforge.net`.

[19] M. Kazhdan, M. Bolitho, and H. Hoppe. Poisson surface reconstruction. In *Proceedings of the fourth Eurographics symposium on Geometry processing*, volume 7, 2006.

[20] K. N. Kutulakos and S. M. Seitz. A theory of shape by space carving. *International Journal of Computer Vision*, 38(3):199–218, 2000.

[21] D. Lanman and G. Taubin. Build your own 3d scanner: 3d photograhy for beginners. In *SIGGRAPH '09: ACM SIGGRAPH 2009 courses*, pages 1–87, New York, NY, USA, 2009. ACM.

[22] M. Levoy, K. Pulli, B. Curless, S. Rusinkiewicz, D. Koller, L. Pereira, M. Ginzton, S. Anderson, J. Davis, J. Ginsberg, et al. The digital michelangelo project: 3d scanning of large statues. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pages 131–144. ACM Press/Addison-Wesley Publishing Co., 2000.

[23] A. Makarenko, A. Brooks, and T. Kaupp. On the benefits of making robotic software frameworks thin. In *International Conference on Intelligent Robots and Systems*, volume 2, 2007.

[24] G. Miller and S. Fels. An abstraction for correspondence search using task-based controls. In *Asian Conference on Computer Vision*, pages 229–242. Springer, 2014.

[25] G. Miller, S. Oldridge, and S. Fels. Openvl: Abstracting vision tasks using a segment-based language model. In *Computer and Robot Vision (CRV), 2013 International Conference on*, pages 257–264. IEEE, 2013.

[26] T. Moons, L. Van Gool, and M. Vergauwen. *3D Reconstruction from multiple images, part 1: Principles*. Now Publishers Inc, 2009.

[27] P. Moulon, P. Monasse, R. Marlet, and Others. Openmvg. an open multiple view geometry library. `https://github.com/openMVG/openMVG`.

[28] S. K. Nayar, K. Ikeuchi, and T. Kanade. Surface reflection: physical and geometrical perspectives. Technical report, DTIC Document, 1989.

[29] A. P. Pentland. Local shading analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (2):170–187, 1984.

[30] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 1, pages 519–528. IEEE, 2006.

[31] S. M. Seitz and C. R. Dyer. Photorealistic scene reconstruction by voxel coloring. In *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*, pages 1067–1073. IEEE, 1997.

[32] N. Snavely, S. M. Seitz, and R. Szeliski. Photo tourism: exploring photo collections in 3d. In *ACM transactions on graphics (TOG)*, volume 25, pages 835–846. ACM, 2006.

[33] N. Snavely, S. M. Seitz, and R. Szeliski. Modeling the world from internet photo collections. *International Journal of Computer Vision*, 80(2):189–210, 2008.

[34] C. Strecha, W. von Hansen, L. Van Gool, P. Fua, and U. Thoennessen. On benchmarking camera calibration and multi-view stereo for high resolution imagery. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. Ieee, 2008.

[35] A. Vedaldi and B. Fulkerson. VLFeat: An open and portable library of computer vision algorithms. `http://www.vlfeat.org/`, 2008.

[36] R. J. Woodham. Photometric stereo: A reflectance map technique for determining surface orientation from image intensity. In *22nd Annual Technical Symposium*, pages 136–143. International Society for Optics and Photonics, 1979.

[37] C. Wu et al. Visualsfm: A visual structure from motion system. 2011.