Noname manuscript No.
(will be inserted by the editor)

Large-Scale Data for Multiple-View Stereopsis

Henrik Aanæs · Rasmus Ramsbøl Jensen · George Vogiatzis · Engin Tola · Anders Bjorholm Dahl

Received: date / Accepted: date

Abstract The seminal multiple-view stereo benchmark evaluations from Middlebury and by Strecha et al. have played a major role in propelling the development of multi-view stereopsis (MVS) methodology. The somewhat small size and variability of these data sets, however, limit their scope and the conclusions that can be derived from them. To facilitate further development within MVS, we here present a new and varied data set consisting of 80 scenes, seen from 49 or 64 accurate camera positions. This is accompanied by accurate structured light scans for reference and evaluation. In addition all images are taken under seven different lighting conditions. As a benchmark and to validate the use of our data set for obtaining reasonable and statistically significant findings about MVS, we have applied the three state-of-the-art MVS algorithms by Campbell et al., Furukawa et al., and Tola et al. to the data set. To do this we have extended the evaluation protocol from the Middlebury evaluation, necessitated by the more complex geometry of some of our scenes. The data set and accompanying evaluation framework are made freely available online.

Based on this evaluation, we are able to observe several characteristics of state-of-the-art MVS, e.g. that there is a tradeoff between the quality of the reconstructed 3D points (accuracy) and how much of an object's surface is captured (completeness). Also, several issues that we hypothesized would challenge MVS, such as specularities and changing lighting conditions did not pose serious problems. Our study finds that the two most pressing issues for MVS are those of meshing (forming 3D points into closed triangulated surfaces) and lack of texture.

H. Aanæs, R.R. Jensen and A.B. Dahl
Technical University of Denmark, Lyngby, Denmark
E-mail: {aanes, raje, abda}@dtu.dk

G. Vogiatzis
Aston University, Birmingham, England
E-mail: g.vogiatzis@aston.ac.uk

E. Tola
Aurvis, Ankara, Turkey
E-mail: tola@aurvis.com

1 Introduction

Stereopsis from both two and multiple views (MVS) is one of the central problems in computer vision. Stereopsis allows easy capture of the environment such that appealing 3D models can be made. This has many applications in entertainment, augmented reality, robotics, as well as industrial inspection and aerial cartography. During the last decade, the advances in MVS have been driven by benchmark MVS data sets. Central benchmark data sets are the Middlebury Multi-View Stereo data set [32] and the building data set by Strecha et al. [34]. Although these data sets have been tremendously useful, they also have their limitations due to their relatively small sizes – Middlebury contains two scenes and Stretcha et al. contains six. To continue the important advancement of MVS, the basis for empirical development comparison and evaluation has to advance, along with the methodology.

In order to further advance the development of MVS algorithms, we have compiled a large data set, consisting of 80 different scenes, and we present this here. This data set is almost an order of magnitude larger than the current state of the art. We show that it is large enough to detect the effects of central aspects of MVS algorithms in a statistically significant manner, the latter being central for scientifically solid advances in MVS.

Fig. 1 Subset of point clouds in our reference data set. The images show point reconstructions of scenes with variability in geometry, reflectance, and texture. These images are grouped in our analysis into categories like groceries and vegetables.

The full data set is free and available for download at http://roboimagedata.compute.dtu.dk/.

Examples of point clouds from the proposed data set are shown in Fig. 1, and as outlined in Section 3, the composition of the data set is such that it spans much of the scene variation central to MVS, such as varying degrees of specularity, geometric complexity, texture and light variation. The data set was compiled using a 6-axis industrial robot, with the evaluation reference achieved via a structured light scanner. We have chosen the term *reference data* instead of ground truth, to emphasize that these are also physical measurements.

The added geometric complexity in the scenes of the proposed data set required further development of the otherwise well thought through protocol of the Middlebury evaluation [32], with a more direct handling of the occluded regions. This extension is another contribution of this paper.

To demonstrate the usability of the proposed data set, as well as to gain insight into the abilities of the state-of-the-art MVS, we have applied the MVS algorithms of Tola et al. [35], Furukawa and Ponce [12] and Campbell et al. [7] to the data set (referred to as *Tol*, *Fur* and *Cam*). This also provides a benchmark for others to compare their algorithms against. The results of this empirical evaluation are given in Section 5, where we investigate the effects of specular surfaces, light variations and converting estimated 3D point clouds into dense triangulated surfaces, i.e. meshing. A previous and more limited version of this study appeared in [17].

2 Related Work

The first work that attempted to benchmark MVS algorithms was Seitz et al. [32], in which the performance of six algorithms was measured across two different scenes. The authors subsequently invited submissions of reconstruction results from dozens of different algorithms, and these were publicly ranked against each other. The somewhat artificial, low-resolution setup of Middlebury [32] was subsequently improved in the evaluation effort by Strecha et al. [34] that consisted of high-resolution images of outdoor scenes. Both [32] and [34] made an invaluable contribution to the advancement of MVS technologies by providing a solid platform on which improvement to existing state-of-the-art can be measured and recorded.

Our work contributes to the evaluation of MVS, albeit with a different focus. In [32, 34], the evaluators' basic question was, "which MVS algorithm works best for *this scene*?" In our work we ask the question "what scene types work best for this MVS algorithm and what scene features make MVS reconstruction fail?" Posing the question this way facilitates more detailed understanding of current state-of-the-art MVS and several future research challenges for it. The evaluations of [32, 34] consider a small number of 3D scenes that are thought to be representative of real-world application domains for MVS. In practice, they chose well-textured diffuse-reflectance 3D objects on which MVS algorithms tend to perform quite well. They then applied several algorithms in order to create a performance-ranking for each scene. Our approach is to consider the wide range of 3D scenes one might encounter in real applications, and then consider how particular types of MVS algorithms perform on each type of scene. This approach sheds light on the performance of MVS technology as a whole and its overall suitability for particular applications.

Most successful MVS algorithms can be divided into two main categories: point-cloud-based methods (e.g. [7, 12, 13, 15, 35, 37]) and volume-based methods (e.g. [14, 21, 23]). Volume-based methods aggregate photo-consistency data in a 3D volume and compute a 3D surface within that volume using surface optimisation. On the other hand, point-cloud-based methods convert photo-consistency data into a 3D point-cloud, which is then converted into a 3D triangulated surface using standard meshing techniques such as Poisson reconstruction [18], graph cuts [37] or signed distance functions [28]. In this work we focus on point-cloud-based

methods because we can easily isolate the point-cloud stage from the surface extraction stage and all the filtering and regularisation this entails.

Within point-cloud-based methods we can distinguish two different paradigms: Feature expansion *Fur*[12] and depth-map fusion [7, 13, 15, 35, 37]. Under the feature expansion paradigm the algorithm starts from a set of 3D features in the scene, which then expand into nearby 3D points while outliers are filtered using occlusion reasoning. Depth-map fusion works by computing independent depth maps for each image using neighboring images. These depth maps are then merged into a single point cloud. We chose *Fur*[12], *Cam*[7], and *Tol*[35] as representative algorithms from the feature expansion and depth-map fusion families. It must be stressed again that our aim is not to directly compare the three methods or the three families of algorithms. Rather, by running these methods on a large selection of data sets we highlight the effect on performance of different types of 3D scenes.

Perhaps closer in spirit to the present work are some previous attempts at investigating in detail different aspects of MVS performance. In [20] there is a theoretical analysis of the impact of scene geometry on feature-expansion MVS methods. A serious evaluation of MVS algorithms based on depth-map fusion is presented in [16]. Our work can be seen as an empirical analysis of both families of MVS algorithms.

A recent trend in MVS research has been to automate all aspects of the MVS pipeline, including viewpoint selection and image capture. For example, in [3, 11] MVS is applied to photographs of famous landmarks, harvested from online photo-collections. Similarly, the authors of [38] propose using MVS with sequences of images obtained by a remote controlled model helicopter for the purposes of automatic 3D mapping. These examples highlight a detailed understanding of the performance of MVS algorithms under different conditions, which is the purpose of the proposed data set.

The problem of evaluating 3D reconstruction is of course not unique to MVS technologies. In [5] the authors describe a detailed study of several laser-based scanners for large-scale, architectural scenes. The large-scale evaluation of time-of-flight systems is the focus of [27]. That work carefully collects a number of design principles that must be adhered to by a ground-truth data set designed to evaluate time-of-flight systems. Several different scanners are tested in [4] with RMS errors reported on a single 3D scene. In [24], a portable test rig is created and scanned by several technologies. The emphasis here is on automation and ease of use. The same theme is followed in [25] where a benchmark for evaluating different types of 3D scanners is presented. In that work a variety of technologies based on several methods like laser triangulation, structured light and time-of-flight are tested against a single, portable object that exhibits multiple different types of reflectance and relatively simple geometry (plane and hemisphere). Apart from the usual reporting of RMS and completeness measures, an evaluation into the effects of specular reflection is also presented.

3 Data

High-performing MVS algorithms are expected to precisely recover 3D surface geometry of natural scenes. Under natural imaging conditions many factors may vary, which makes 3D reconstruction a challenging task. Factors include camera pose, scene variation, including the non-static nature of many scenes, scene illumination, etc. Our aim with the proposed data set is to evaluate MVS performance in relation to such key performance-influencing factors, and to be able to distinguish the effect of the individual factors. In order to obtain this we have constructed a highly controlled setup for data acquisition, where we have chosen to systematically vary the camera position, scene, and illumination. In total we have 80 scenes, with the same 49 or 64 camera positions depicted under varying lighting conditions. This allows a detailed statistical analysis of MVS performance. The image resolution is 1200×1600 pixels in 8-bit RGB color, with practically all the scene being in the depth of field (due to long exposure and small aperture).

A fair argument against this approach is that it does not capture *all* aspects of unconstrained hand-held photography, such as motion blur, typical user behavior, natural sunlight, etc. But a rigorous and systematic evaluation of MVS requires some of the frivolity to be removed, e.g. in order to capture reference surface information, and we believe that the presented data set does capture most of the relevant issues.

3.1 Scene Choice

Apart from making the data set large enough to capture a large variability of scene types and to allow for statistically significant analysis of relevant aspects, we have also strived to span some of the relevant issues relating to MVS, as exemplified by the images shown in Fig. 3. Firstly, we included subsets of scene type clusters into the data set to enable within-class analysis on more refined details. Specifically, we included

– 16 scenes of model houses, c.f. Fig. 3-a.

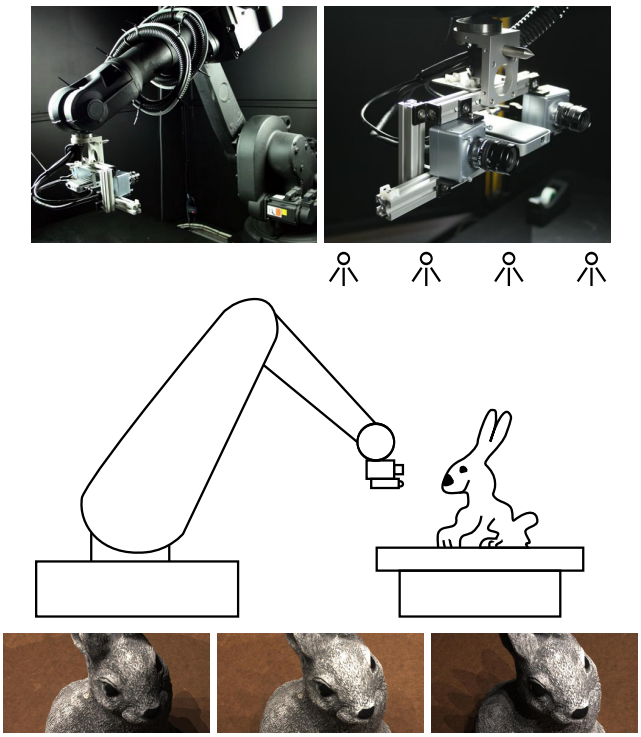

Fig. 2 Top shows photos of the industrial robot mounted with the two cameras and the projector. Both cameras are used for structured-light reconstruction, but the input views for the datasets are only collected by one camera. In the middle is a schematic illustration of the setup, consisting of the industrial robot, LEDs in the ceiling, and the scene placed on a table. The bottom shows three different illuminations of the scene.

- 7 scenes of building materials with diffuse reflectance including wood and concrete, c.f. Fig. 3-b.
- 11 scenes of groceries, c.f. Fig. 3-c.
- 6 scenes of fruit and vegetables, c.f. Fig. 3-d.
- 7 scenes of stuffed animals, c.f. Fig. 3-e.

In addition to this, we have composed the scenes such that they span geometric variation, e.g. Fig. 3-f and 3-g, specular reflections, e.g. Fig. 3-b, 3-f, 3-i, 3-j and 3-k, as well as variation in the degree of texture. For example large parts of the grocery scenes are without texture Fig. 3-l. We thus captured most of the variability of scene types that we hypothesize are of importance for MVS performance. A deliberate omission, however, was very thin structures, which we did not include as we were not sure that the structured light would give reference data of sufficient quality.

3.2 Image Positioning

Our data acquisition was done in a controlled environment similar to [32]. In our, setup we mounted a camera and a structured light scanner on a 6-axis industrial robot, providing a precise and flexible camera pose, c.f. Fig. 2. In order to vary the illumination we acquired images using 16 individually controlled light emitting diodes (LEDs) placed above the scene, see Fig. 4 and 5 and Tab. 1. This setup has previously been used in [1, 19] to produce different data sets but in a similar manner.

The robot provided very precise camera positioning due to its very high position repeatability. By coding the robot with a set of predefined positions calibrated photogrammetrically using a fixed checkerboard pattern, we acquired images from the same positions for the 80 scenes in our data set. By using the industrial robot arm we obtained a flexible design space for our experiments, which we used to let the robot move to camera positions on concentric spheres – something that would not be possible with a static setup.

The 80 scenes contained different number of camera positions. 59 scenes contained 49 camera positions and 21 scenes contained 64 camera positions. The camera positions of the smaller sets were placed on one sphere with a radius of 50 cm, i.e. around 35 cm from the scene surfaces. The larger sets contained an additional 15 positions on a concentric sphere with a radius of 65 cm at a distance around 50 cm from the scene centers as shown in Fig. 6. The inner/main sphere allowed each scene point to be observed from many different angles. The outer sphere was included to allow investigations into the effect of scale changes.

3.3 Reference Scan

The reference points, obtained from the structured light scans, are based on binary gray code, which is recommended as being one of the most precise structured light methods [29, 30, 31]. The scans are, however, not complete. The main cause is that only the front of the objects were covered, and there are areas seen by the cameras that have not been covered. This occurred because of object self-occlusion and small holes where the structured light images were severely underexposed. Despite these minor incompleteness issues, the scans are very dense, each containing 13.4 million points on average.

Note, only the scene objects were used in the evaluation. This was done by removing the part of the reconstruction containing the supporting table, simply by discarding points below a manually placed plane.

3.4 Accuracy

Our experiments were dependent on the accuracy of the structured light scans, and we therefore measured the

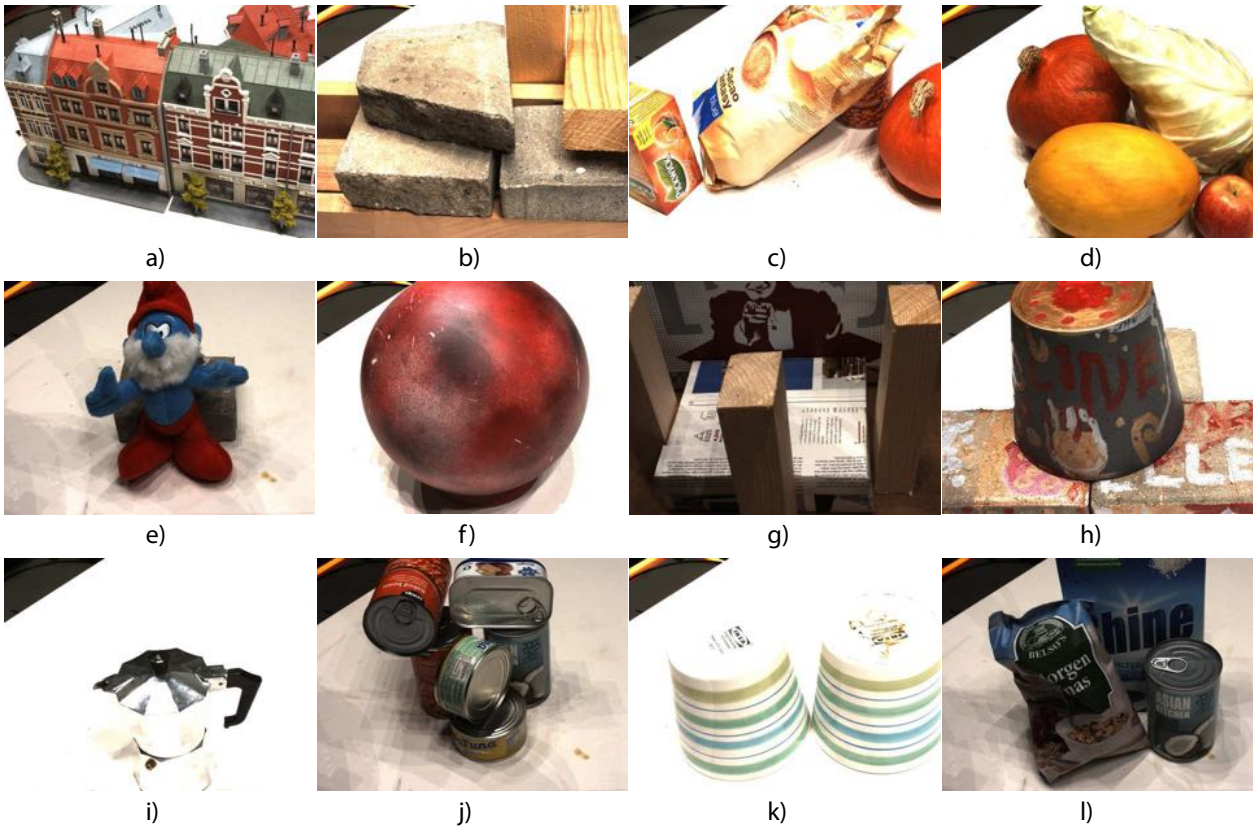

Fig. 3 Examples images from our data set including examples of the five different scene categories in a) to e) and the rest illustrating the variability in geometric complexity, specularity, and texturedness.

scan precision using an object with known geometry. We chose a bowling ball, because it is a spherical object of suitable size with a simple and known geometry. A reference scan was obtained from each camera position, and all the scans were combined to make up the total reference data for each scene. For each scan we estimated the centre position and the radius of the sphere form the surface points using linear least squares. This also enabled us to estimate the deviation of the individual points from the sphere's surface. We obtained a standard deviation of 0.17 mm on the centre position estimates, and an average standard deviation on the surface points of 0.14 mm, which corresponds roughly to 0.6 pixels. Positioning repeatability of the robot turned out to be very high. Over the two months of the data acquisition period, we performed 10 complete calibrations, and the average standard deviation of the camera positions was 0.0552 mm. The reprojection error here was 0.067 pixels.

3.5 Varying Illumination

In some situations, e.g. online photo collections, the scene illumination varies significantly. In order to obtain 3D reconstructions from such data, MVS algorithms must be able to handle large variation in illumination. To enable evaluation under changing lighting conditions, we chose to vary the scene illumination. This variation was achieved using 16 LEDs placed in the ceiling, as illustrated in Fig. 5. In each camera position, seven different illuminations were obtained by strobing the LEDs in groups as illustrated in Fig. 4. This resulted in images with varying degrees of directional illumination and one with diffuse illumination[1]. Note that we denote the lighting of pattern 4 in Fig. 4 as diffuse, even though it is only an emulation, with all 16 LEDs turned on.

4 Evaluation Protocol

To evaluate MVS stereo algorithms based on our data set, an evaluation protocol is required. This protocol takes a structured light point cloud and an MVS reconstruction, and returns the mean and the median point-wise reconstruction error, quantifying how well the latter fits the former. The protocol is an integral

[1] In a few of the extreme positions, the robot shaded a few of the LEDs.

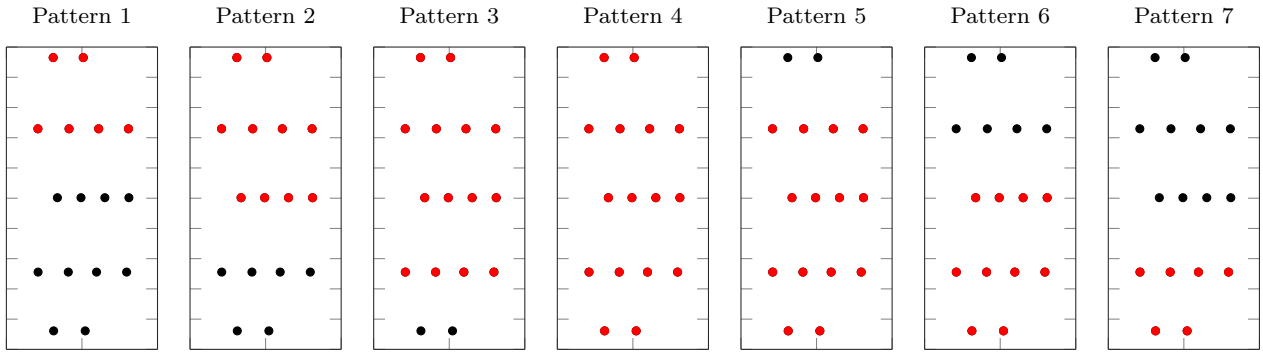

Fig. 4 LED illumination pattern. LEDs that are turned on are marked in red.

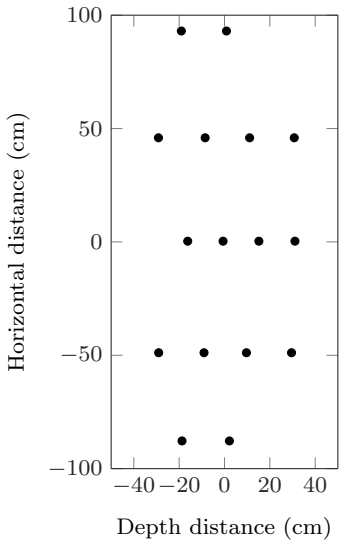

Fig. 5 Overview of how the LEDs are placed above the scene.

LED #	θ	ϕ	LED #	θ	ϕ
1	269.5°	56.2°	9	332.8°	89.7°
2	281.6°	55.7°	10	358.9°	83.3°
3	236.2°	68.3°	11	121.2°	67.7°
4	256.4°	71.2°	12	101.2°	70.3°
5	280.5°	71.4°	13	79.5°	70.3°
6	302.4°	68.6°	14	59.3°	67.8°
7	180.6°	77.4°	15	91.4°	57.7°
8	181.2°	83.8°	16	78.0°	57.1°

Table 1 Azimuth (ϕ) and elevation (θ) angles in degrees for all LEDs numbered according to Fig. 5 (top left to bottom right). The centre of the coordinate system is the surface of the table where the scenes are placed.

part of the experimental design, and its details are presented in this section. Here, we take as a starting point the protocol from the Middlebury MVS evaluation [32], which we modify, among other things to account for the higher geometric complexity of our data.

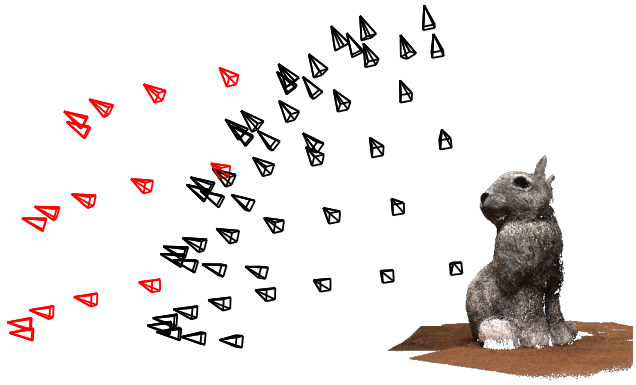

Fig. 6 Camera positions on a 50 cm sphere (black) and a 65 cm sphere (red).

4.1 Quantifying Distances Between Point Clouds

As mentioned above, we use a modified version of the protocol in [32]. As in [32] we also use accuracy and completeness as evaluation measures, where;

- **Accuracy** is measured as the distance from the MVS reconstruction to the structured light reference, encapsulating the quality of the reconstructed MVS points.
- **Completeness** is measured as the distance from the reference to the MVS reconstruction, encapsulating how much of the surface is captured by the MVS reconstruction.

Both measures are needed for a fair comparison. If only accuracy were reported, it would favor MVS algorithms that only include estimated points of high certainty, e.g. high-textured surface parts. On the other hand, if only completeness were reported it would favor MVS algorithms that include everything, regardless of point quality.

These distances are measured by comparing structured light and MVS-reconstructed 3D point clouds. More specifically, we measure the distance from every point in one point cloud to the closest point in the other

point cloud and then we record statistics about the distribution of these. We chose to characterize these empirical probability distribution functions (PDFs) by their mean and median, after removing observations with distances above 20 mm. The latter was done so that a few large outliers would not dominate the result.This reduction or projection of the PDFs is slightly different than [32]. They report a high fractile where we report the mean and median. This change is done in accordance with standard statistical practice where mean and median are the typical fist projections of a PDF to be reported, [36]. The motivation for including the median is because it is a standard robust measure, and allow us to gain better insight into the effect of 'small outliers' (not removed by our 20 mm threshold).

4.2 Missing Data and Observability

When using structured light scanning, it is common to have holes in the 3D surface model, as was the case in [32] as well as in our data. The essential property of the reference data (ground truth), for this type of MVS evaluation, is that it segments 3D space into where there is a surface and where there is not. In relation to MVS evaluation, an implication of the surface holes is that some of the reference surface has not been observed.

In [32], this issue is addressed by closing the holes in the reference model via a hole-filling algorithm, in effect by using interpolation. When evaluating accuracy, i.e. the distances from points on the MVS reconstruction to the structured light scan, an MVS point is discarded from the evaluation if its closest point is a result of such interpolation. An interpretation of this is that the Voronoi regions of the hole-filled parts of the reference data are the parts of 3D space classified as non-observable. To avoid point misclassification, the hole-filled surface must be close to the true surface, which requires the holes to be small or the geometry to be simple. Therefore, the surface scans must either be almost complete or simple in geometry. This is hard to obtain with complex-shaped objects with a large degree of self occlusion. We have strived after large variation in our data set, including geometric complexity, which implies that a hole-filling approach will not be applicable for our data, see e.g. the scenes in Fig. 3-d and Fig. 3-g.

To address the issue of observability, we instead explicitly computed an observability mask, which provides information about the visible parts of the scene with reference data. This was done by representing the relevant part of 3D space by a voxel grid (of voxel size 1 mm^3), and initializing all parts as being not observed. Then, for *every* structured light point, we computed the ray to the camera recording that point and all voxels along that ray were set as observed. This ray was extended 10 mm behind the 3D point, allowing reconstructions in this range to be evaluated. The described algorithm produced a binary 3D observability mask representing where the 3D surface could be observed by camera sensors. The mask could then be used to restrict the evaluation of MVS algorithms, by ignoring accuracy or completeness of masked points. Apart from handling holes in the structured light scan, this observability mask also handles the fact that our data set only has objects scanned from one side[2].

4.3 Sampling Reconstructions and Meshing

As mentioned, our structured light reconstruction was merged from a number of structured light scans. A side effect of this is that the sampling density is uneven, for example with prominent parts being visible from more angles resulting in higher sampling density. Many state-of-the-art MVS algorithms, including the ones evaluated here, have a similar trait of uneven point sampling, e.g. because they at some stage are a merger of two-view stereo.

A side effect of this uneven sampling is, that in comparing point clouds point to point, the quality of the higher sampled surface areas are weighted up. This results in unduly biasing the evaluation towards prominent points, and towards high-textured areas, where stereo algorithms are more likely to give a response. We found an area integral more appropriate implying the need for a uniform sampling on the surfaces.

To address this issue, we reduced the sample density of the MVS and structured light point-clouds. This was done by considering the points of a given point cloud in random order, and only keeping a point if there were no previously considered points within a distance of 0.2 mm. The 0.2 mm threshold was chosen, since this is a conservative estimate of the accuracy of our structured light scans.

The effect of this was to randomly down-sample areas of density higher than 0.2 mm down to 0.2 mm, while leaving other areas unchanged. Lower sampled regions were not upsampled, firstly because considerably lower density implied less reliably estimated regions, and also because there is no clear way of how to upsample without getting into the hole-filling issues mentioned above. The latter would have biased the result towards some heuristic prior imposed by us.

[2] In the online data set, 360° scans of some models are included by combining four scans. In these cases we only included one data set into the evaluation, in order to avoid biasing the data set unnecessarily.

The choice of sub-sampling influences the structured light reference data because points are removed, which will give a bias towards larger error measures. In order to quantify the effect of sub-sampling, we ran our evaluation protocol with the structured light scans as data, but down-sampled in another random order. Averaging over all scenes, in the same manner as in Section 5.1 and Fig. 8, the results were a difference of 0.0631 mm for the mean and 0.0301 mm for the median, which are significantly smaller than most differences in the performance measures. Despite this difference, the sub-sampling is unlikely to influence the relation between the performances measured for different MVS methods, since the choice of removing a point influences the performance measure as a point-wise stochastic process. Therefore, all points in a given MVS reconstruction are equally likely to be affected by the sub-sampling, and since we have very large reference point sets, it is highly unlikely to influence the performance measure.

An alternative would have been to fit a surface to the structured light points, as done in [32], for example. Fitting a surface would, however, imply using interpolation and thus a surface prior. Such a prior can be seen as a bias, and cannot be averaged out.

In addition to evaluating the MVS point reconstructions we also evaluated meshed versions of the point clouds, forming triangulated surfaces. The triangulated surfaces were evaluated by converting them to point clouds by first uniformly sampling each triangle of the triangulated surface and then reducing it to a minimum 0.2 mm sampling density using the same method as mentioned above. This method gave very similar evaluation protocols for the point and triangulated surface reconstructions.

4.4 Protocol Outline

The MATLAB code for evaluating MVS reconstructions via the data and protocol is available together with the data online. In short, the proposed protocol can be outlined as follows: given an MVS reconstruction and structured light scan, both as point clouds, in the same frame of reference:

1. Reduce the sampling density of both point clouds as described in Section 4.3.
2. For every point in the structured light scan compute the distance to the closest point in the MVS reconstruction. This gives the completeness distribution.
3. For every point in the MVS reconstruction, if it is in the observability mask c.f. Section 4.2, compute the distance to the closest point in the structured light scan. This gives the accuracy distribution.
4. For each of the PDFs in items two and three, remove outliers and compute the mean and median.

If the MVS reconstruction is a triangulated surface and not a point cloud, convert the triangulated surface into a point cloud by uniform sampling as mentioned in Section 4.3.

Acknowledging, that the proposed protocol involves parameters set by our best, albeit subjective, judgement, we performed a sensitivity analysis on these turn-button parameters. Specifically, we investigated the 0.2 mm sampling distance threshold, the 20 mm outlier rejection threshold and the 10 mm ray-extension threshold, by rerunning our experiments with each of these parameters changed by plus and minus ten percent. The effects hereof were so minor, with mean effects of approximately a hundredth of a millimeter, that we confidently conclude that the evaluations are very insensitive to these parameters.

5 Empirical Investigations

A natural part of proposing our data set and protocol aimed at MVS is to apply state-of-the-art MVS algorithms to it. The purpose of doing so is threefold: firstly to validate that the proposal is useful for its intended purpose, secondly to set a benchmark on which others can compare their algorithm, and thirdly to gain insight into the state-of-the-art of MVS, i.e. what are the current issues and challenges?

To do these experiments we chose to apply the MVS methods of Campbell et al. [7], Furukawa and Ponce [12], and Tola et al. [35]. These methods represent the state of the art within MVS well – c.f. Section 2 – and provide a baseline on the proposed data set, and as such serve our purpose. The three methods provide point clouds that were meshed, i.e. creating a dense triangulated surface, via the Poisson surface-reconstruction algorithm [18]. As such, both the 3D point reconstructions, as well as the triangulated surface aggregates were tested. Poisson surface reconstruction was chosen, because it is one of the most popular methods.

For all MVS methods, we used original implementations, *without* optimizing the parameters for better performance, because this would take them away from their original form. However, we made one alteration, in relation to the meshing, where all three methods use the Poisson reconstruction [18]. Here we standardized the parameter settings using depth 11, and trimmed such that areas with depths less than 8 were removed. We judged this would give a fairer comparison[3].

[3] This standardization of the Poisson reconstruction parameters was done after the preliminary version of this work

We present two experiments concerning (i) general evaluation of all scenes using full illumination, and (ii) evaluation of changing illumination for 10 scenes. In the general evaluation experiments we also report the results of scene categories for a selection of scenes. An overview of our experiments is given in Tab. 2.

Experiment	# scenes	Varying illumination
(i) General	80	No
Categories	–	No
- model houses	16	No
- groceries	7	No
- vegetables	11	No
- building material	6	No
- stuffed animals	7	No
(ii) Illumination	10	Yes

Table 2 Overview of experiments. Note that category experiment is a subset of the general experiment.

An important point in including 80 scenes in our dataset is to allow for thorough statistical analysis of the performance, because effects that may accidentally occur in one scene are averaged out by repetition. In addition the large number of scenes allows for investigating different factors affecting the performance. We apply the standard statistical way of analyzing such data, namely an analysis of variance (ANOVA) [2]. An ANOVA computes the effects and cross effects of the different factors of our experiment as well as the statistical strength or significants. The factors included in our analysis include:

- Overall mean performance μ.
- Algorithm a_i ($i \in \{\text{Tol, Fur, Cam}\}$).
- Scenes s_j ($j \in \{1, \ldots, 80\}$).
- Meshing m_k ($k \in \{\text{Used, Not used}\}$).
- Illumination l_n ($n \in \{\text{Full, Varying direction}\}$)

Four performance measures of the MVS algorithms are considered, which are the mean and median values of the completeness and accuracy scores. Both one-way and two-way interactions are considered, and two way interactions are e.g. denoted as_{ij} for the cross effect of algorithm i and scene j – note that a variable is estimated for each combination. The model for the general experiment becomes

$$y_{ijk} = \mu + a_i + s_j + m_k + as_{ij} + am_{ik} + sm_{jk} + \epsilon_{ijk} \ ,$$

and the model for varying illumination is

$$y_{ijn} = \mu + a_i + s_j + l_n + as_{ij} + al_{in} + sl_{jn} + \epsilon_{ijn} \ ,$$

where y is the performance measure (either mean or median of completeness or accuracy) and ϵ is the residual error. Results for these models are shown and discussed in the following.

5.1 General Evaluation with Full Illumination – (i)

Even though the proposed data set includes many possibilities for investigation, e.g. varying light and scene types, the natural first experiment to perform is to apply the MVS algorithms to all the diffuse (As mentioned in Section 3, this is only an emulation of diffuse light) lighted images. Sample reconstructions from this experiment are seen in Fig. 12, and a summary of the overall performance is shown in Fig. 8. Here the results are retrieved both as raw point clouds as well as triangulated surfaces computed from these. As explained in Section 4 we evaluate the performance based on accuracy and completeness (to aid others in the use of our data set, all results on a point to point basis is found on the homepage associated with the data set).

Fig. 8 clearly shows that there is a tradeoff between completeness and accuracy with *Tol*[35] being the most accurate and *Cam*[7] being the most complete. This finding is confirmed by looking at the individual reconstructions, where this tradeoff manifests itself in a choice between the obtained detail at the expense of more errors, most notably outliers. So, this study does not show one of the three methods to be superior compared to the other. Furthermore, the method of *Tol*[35] was developed for much higher resolution images than the ones used here, which in turn translates into a high accuracy and low completeness on these images. Results of the analysis are shown in Tab. 3.

	Accuracy		Completeness	
Method	Mean	Median	Mean	Median
MVS algorithm, $\mu + a_i$				
Tol	0.408	0.224	1.040	0.424
Fur	0.952	0.427	0.772	0.418
Cam	1.082	0.530	0.551	0.250
Meshing, $\mu + m_k$				
Meshing used	0.562	0.335	0.829	0.359
Meshing not used	1.066	0.452	0.746	0.370
Cross effects – algorithm and meshing, $\mu + am_{ik}$				
Tol – no mesh	0.327	0.205	1.106	0.466
Fur – no mesh	0.605	0.321	0.842	0.431
Cam – no mesh	0.753	0.480	0.540	0.179
Tol – mesh used	0.488	0.244	0.974	0.382
Fur – mesh used	1.299	0.534	0.702	0.405
Cam – mesh used	1.411	0.579	0.562	0.322

Table 3 Overall performance of the MVS with the average of the main effects of the reconstruction algorithms and the use of meshing, as provided by the ANOVA. The unit is in *mm* and *all* entries are significant on at $p < 0.001$ level.

[17], which is why there is a slight discrepancy between the result of this paper and the preliminary version.

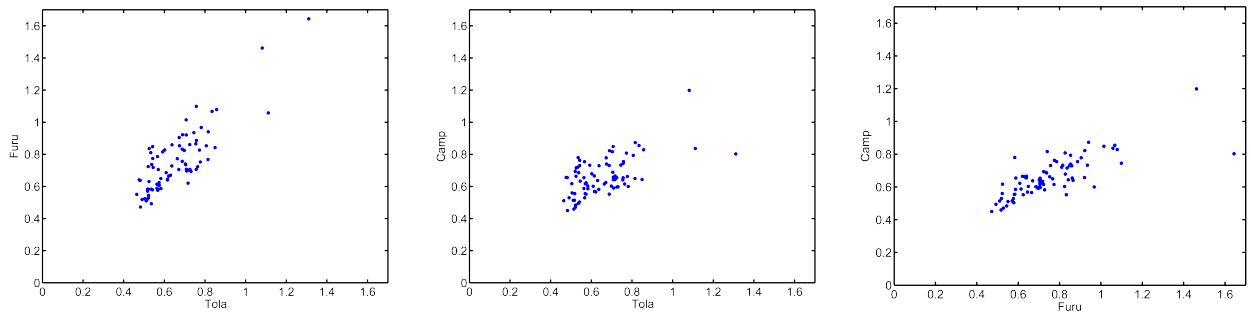

Fig. 7 Pairwise plots of the combined performance score for each of the three tested point reconstruction methods. This combined score is the sum of the median accuracy and the median completeness. Here it is seen that a) there is a high correlation between the performance of the different methods,although this is least obvious comparing the methods of Tola and Campbell, which are also the most different with regard to completeness and accuracy tradeoff, and b) there is no tendency of clustering.

Several things are seen from this ANOVA: firstly that the data set is large enough to give statistically significant results on the aspects we are interested in. This is an effect highly related to the number of observations, in this case scenes. As such, this is a strong validation of our data set compared to state of the art. Secondly, the tradeoff between accuracy and completeness is also confirmed by this ANOVA, as seen by the significance in the difference between average performance of the algorithms where the algorithm with highest accuracy has lowest completeness and vice versa.

Fig. 9 shows a selection of scenes categorized according to their surface reflectance properties. This categorization shows that categories such as (model) houses and diffuse square building materials are well suited for MVS, whereas less traditional objects, such as texture-poor and specular objects found in a grocery store, are more challenging. Although this is not surprising, we still believe it is interesting that generally held hypothesis can be validated in a more rigorous manner.

We also observed that the different methods were approximately equally challenged by the same scenes, i.e. if one algorithm is challenged by a given scene, the other algorithms are likely to be too. To exemplify this in a straight forward manner, we summed the median accuracy and completeness for the point reconstructions. This gave a single scalar value for each algorithm and scene, making the presentation easier. The results of this are presented in Fig. 7, where it is seen that there is a clear linear trend, which is also observed from the associated cross-correlation matrix, given by

$$\rho = \begin{bmatrix} 1.0000 & 0.8333 & 0.6011 \\ 0.8333 & 1.0000 & 0.7764 \\ 0.6011 & 0.7764 & 1.0000 \end{bmatrix}$$

where the ordering of the methods is 'Tol', 'Fur', 'Cam'. It is also seen from Fig. 7 that there is no apparent clustering of the results.

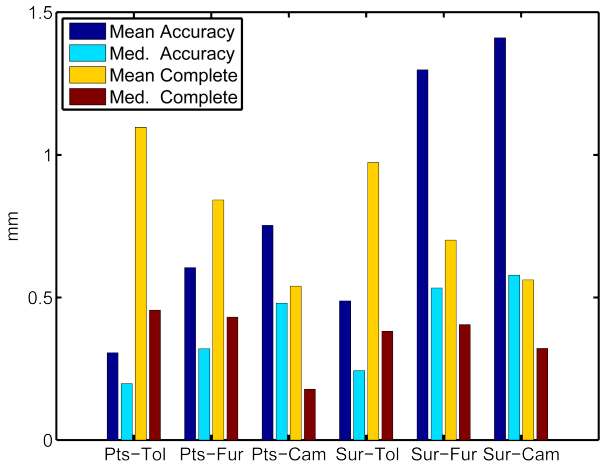

Fig. 8 Performance over all 80 scenes of accuracy and completeness of reconstructed points (Pts) and triangulated surfaces (Sur). The error is measured both as mean and median. Tol is Tola et al. [35], Fur is Furukawa and Ponce [12], and Cam is Campbell et al. [7].

With the vast data set and evaluation presented here we have observed some general trends for the investigated MVS methods. Firstly, we found that the largest source of poor performance is by far the lack of texture, as seen in Fig. 10. In many cases the meshing closes holes which compensates for this lack of texture. The success of this, however, depends on the noise and the complexity of the surface. The box sequence shown in Fig. 10, for example, is improved by meshing where the surface meshing fills holes that closely follow the reference surface points. For more complicated geometries, the meshing does not, however, improve performance, but will often corrupt finer details.

More surprisingly, we found that many other factors, which we expected to seriously corrupt the results, were not as problematic. As an example, the geometric complexity of the scenes did not influence the results

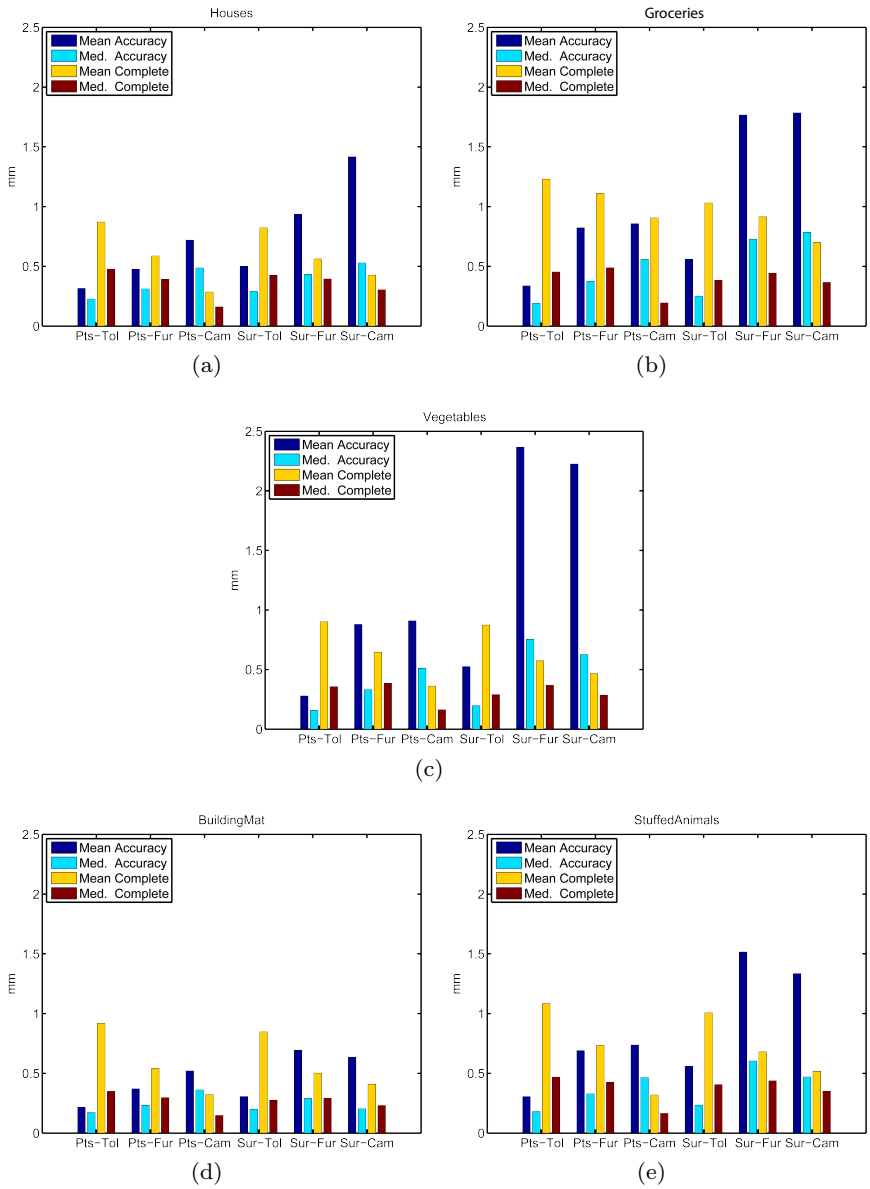

Fig. 9 Performance for different scene types. (a) is model houses, (b) is groceries, (c) is vegetables, (d) is building material, and (e) is stuffed animals.

to the extent we had expected. This was especially true for the point reconstructions. Similarly, specular surfaces and a change of lighting did not influence the reconstructions as negatively as expected, as described in Section 5.2.

5.2 Evaluation with Varying Illumination Direction (ii)

As mentioned in Section 3, an aspect we were particularly interested in was lighting conditions and surface reflectance. Our working hypothesis was that this would be one of the major challenges for MVS, mainly since lighting change was the most corruptive factor found in a previous study on point features [1] – performed in a similar experimental setting. As mentioned above, this hypothesis was disproven. The lack of a highly degrading effect from specular surfaces is illustrated in Fig. 10, where a textureless metal espresso-can has been reconstructed.

The images of the proposed data set have been taken in seven different lighting conditions, ranging from directional to nearly diffuse. This allows us to emulate the type of changing lighting conditions arising from taking images of an object at different times of day, and subsequently attempting an MVS reconstruction. To il-

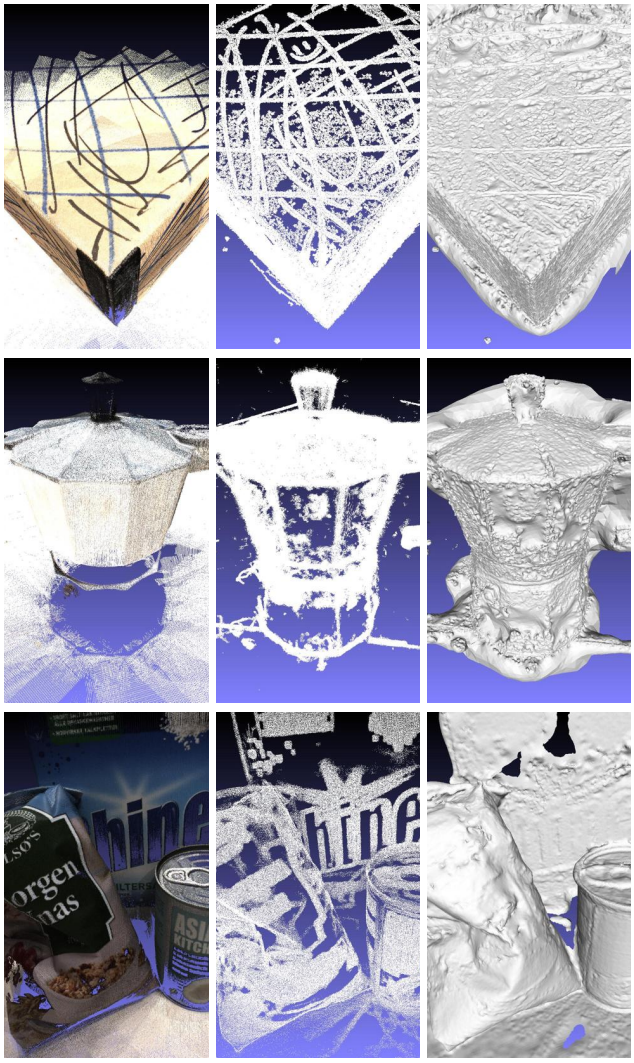

Fig. 10 The top row shows an example of an object with missing texture resulting in reconstructions with holes. The simple geometry of the box did however recover the holes well. From left to right: the reference data points, the reconstructed points by *Fur*[12], and the surface-reconstruction of these points [18]. The middle row shows an expresso pot with almost mirroring surfaces. From left to right: the reference data points, the point reconstruction *Cam*[7] and the surface reconstruction of these points [18]. The bottom row shows a scene with both specularities and lack of texture. From left to right: the reference data points, the point and surface reconstructions of *Tol*[35].

lustrate this feature of our data set, and investigate the effect of lighting on MVS, we chose ten scenes from our data set, on which we made the following experiment;

1. For each of the (49 or 64) camera positions we at random drew an image corresponding to one of the seven lighting conditions.
2. Based on this 'new' data set, we computed new MVS reconstructions and compared them to the ones made with only full illumination.

An example result from this experiment is shown in Fig. 11, where it is heavily indicated that the effect of varying lighting conditions is very limited. This is also the conclusion from a visual inspection of the reconstructions.

To quantify effects of light variation, we also applied an ANOVA to this experiment, and the results are shown in Tab. 4. All one-way and two-way effects are significant for the completeness, meaning that their means are significantly different. The conclusion is that the scans become slightly less complete when the light varies both measured as a mean and as a median error. For the accuracy, the differences are mainly insignificant, however, so randomly varying the light does not affect the accuracy of the scans.

	Accuracy		Completeness	
Method	Mean	Median	Mean	Median
MVS algorithm, $\mu + a_i$				
Tol	0.288 †	0.186 †	1.109 †	0.454 †
Fur	0.681 †	0.311 †	0.729 †	0.409 †
Cam	0.760 †	0.473 †	0.514 †	0.170 †
Light, $\mu + l_n$				
Full light	0.576	0.322 ‡	0.735 †	0.334 †
Light varied	0.576	0.325 ‡	0.832 †	0.354 †
Cross effects – algorithm and light, $\mu + al_{in}$				
Tol – full light	0.300	0.188	0.996 †	0.432 †
Fur – full light	0.676	0.308	0.715 †	0.403 †
Cam – full light	0.751	0.469	0.494 †	0.167 †
Tol – light varied	0.276	0.184	1.221 †	0.475 †
Fur – light varied	0.685	0.314	0.743 †	0.414 †
Cam – light varied	0.768	0.476	0.534 †	0.173 †

Table 4 Light experiment performance of the MVS with the average of the main effects of the reconstruction algorithms and full vs. varying illumination. Significance levels are † $p < 0.001$ and ‡ $p < 0.05$. No mark indicates no significance.

Our hypothesis related to this lack of effect from light and specularities is that; the tested MVS methods in essence propagate the results from image pair matching and even if some or most of such image pairs are corrupted, if just a few are OK, this will mostly result in a good 3D reconstruction. This explains the good performance in the changing light experiment, in that there is almost always two close images with similar lighting. The few cases where this is not the case can explain the slight degradation in completeness.

Thus, the robust workings of the MVS algorithms are able to pick out the good estimates. Lastly, it should be noted that specularities have a high visual effect, but only in limited directions [8]. This implies that only images in one direction can be effected by highlights per point light source.

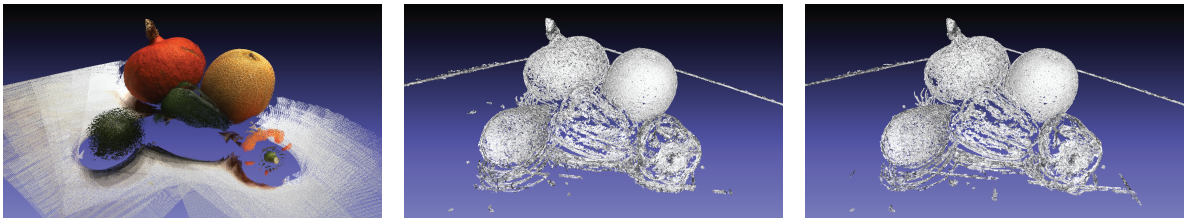

Fig. 11 An example of the effect of lighting variation. Left: our colored structured light reconstruction. Middle: the reconstruction of *Fur*[12] with full lighting. Right: the reconstruction of *Fur*[12] with *varying* lighting direction. The effect of varying the light seems negligible.

5.3 Points vs. Surfaces

The state-of-the-art in MVS has, to a great degree, converged to an approach where a 3D point cloud model of a scene is first reconstructed and then it is transformed into a triangulated surface [10, 13, 18, 22, 26]. The triangulation, or meshing, for these methods is commonly in a form of iso-surface extraction – the most popular of which is Poission reconstruction [18]. This is also the case for the three state-of-the-art methods presented here. We evaluate both the 3D point reconstructions and the triangulated surface aggregates in order to investigate the properties of the meshing, but also because there is debate as to which is correct to report. Additionally, as most meshing methods use the point clouds as input, it is important to evaluate the success of these clouds independently from the meshing stage.

Here, we should note that there are many variations as to the way a triangulated surface is computed from the point-cloud data. Some of the best performing ways are iso-surface extraction methods [9, 10, 18, 33] and graph-cut-based methods [6, 15, 22, 26]. In this work, we choose to use the Poisson reconstruction method [18], firstly because it is used with the three methods evaluated and presented here, secondly because its code is open source and hence easy to use, and finally because there are more readily reported results in the literature with this method and thus it is easier to correlate with our results. We do not expect our following conclusions about surface models to vary greatly with different methods but it will nevertheless improve our understanding of the state of the art for this stage as more methods are evaluated through our datasets and evaluation protocol. This is one of the reasons why we made these available to the community.

As seen in Fig. 8, the point reconstructions in general perform best, which expresses a very clear trend looking at the individual reconstructions. As a general observation, the cases where the meshed results are best are as the box in Fig. 10, where there are large texture-poor regions for which no points are estimated *and* the geometry is simple enough for the implicit smoothing prior of the meshing to smooth noise and fill holes. Typically this applies to flat or spherical surfaces.

Examples of surface meshing are shown in Fig. 12, which illustrates how fine surface details are preserved by the method of *Cam*[7], where many surface points are reconstructed, whereas many of these details are smoothed away in *Tol*[35]. Complex geometry as seen in the middle front part of the house images are, however, severely corrupted by the surface meshing, however. This is one of the scenes where the meshing performed worst relative to the 3D point reconstructions. Firstly, it is seen that the meshing has problems with finer details. Such fine details are *inconsistent* with the implicit smoothing prior of the meshing algorithm. Secondly it is seen that more fine details are captured in *Cam*[7], but also more gross errors. This relates back to the accuracy/completeness trade-off discussed above, in that more complete 3D point data gives more data to constrain the meshing. On the other hand the meshing process is relatively sensitive to outliers, which are increased by poorer accuracy. Sometimes these outliers also seem to result in large surface portions being hallucinated.

Overall, our investigation shows that the three state-of-the-art surface reconstruction algorithms investigated here have high precision in reconstructing surface points. Depending on the number of generated points, a more or less detailed set of surface points can be obtained. Even small features, like a small antenna of a thickness of around 1 mm on a model house, were covered by precisely reconstructed surface points. Extending these surface points to a triangulated surface, however, is not easily done and many of these fine details are often lost. This is not surprising, because it can be hard to distinguish points on small surface details from groups of falsely detected points. Meshing the surfaces is, however, an important task for applying MVS in many of its intended uses in e.g. entertainment, robotics, industrial inspection or aerial cartography. We see this as a great challenge and hope that the provided data set can aid in this development as well as many other investigations within MVS or other computer vision problems.

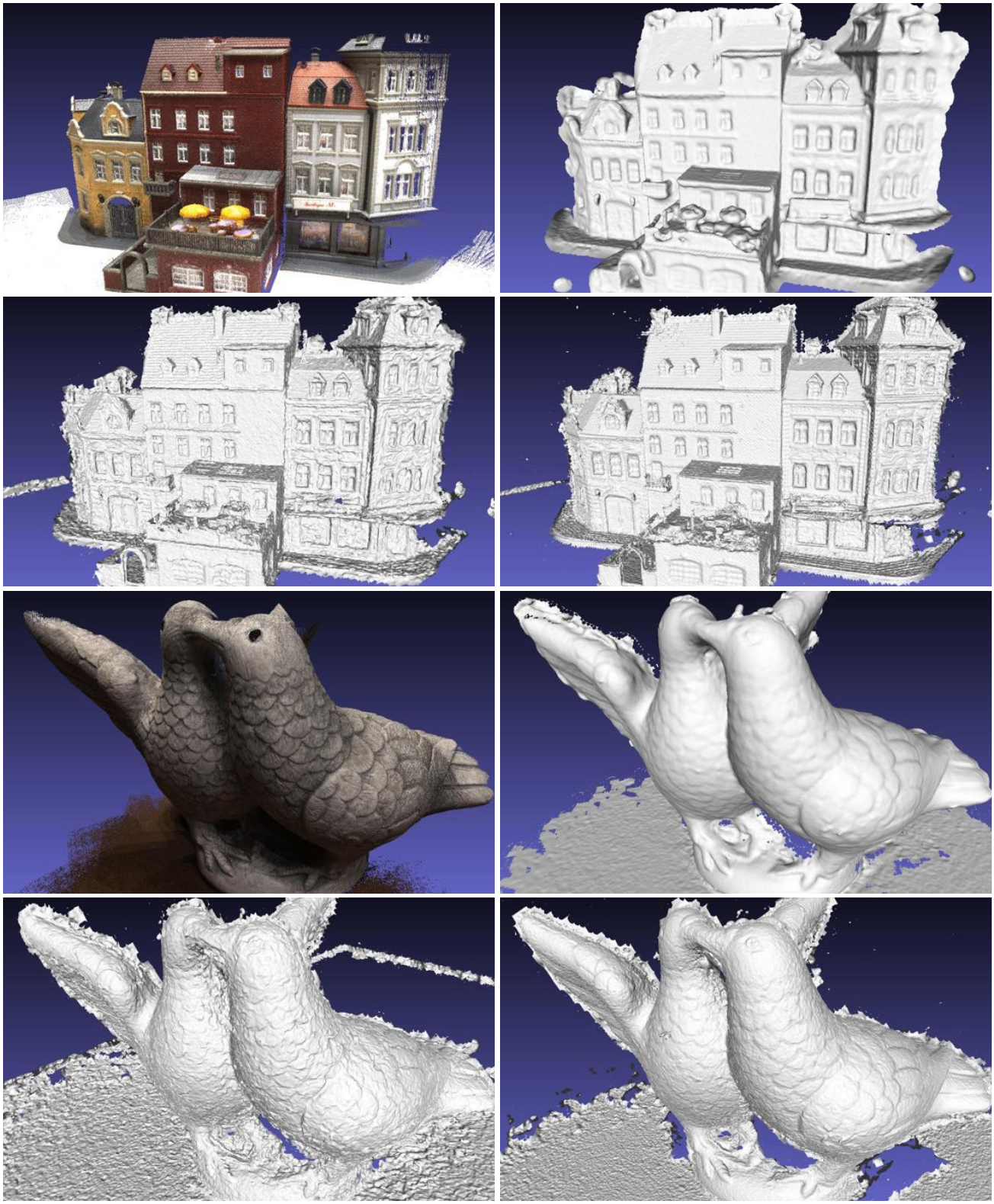

Fig. 12 Reference points (upper left) and triangulated surfaces of buildings where details are corrupted by the smoothing introduced by surface meshing. Upper right is *Tol*[35], lower left is *Fur*[12], and lower right is *Cam*[7]. The statuette of doves is reconstructed following the same order. As with the buildings, a slight corruption of detail is the result of surface reconstruction. In both scenes the artifacts around the edges are results of the surface reconstruction step and are not present in the point reconstruction.

6 Discussion & Conclusion

We have presented a dataset and accompanying evaluation protocol aimed at MVS. This data set captures many of the central issues of MVS, such as varying degrees of specularity, texturedness and geometric complexity. In addition to this, the images are taken under seven different lighting conditions, which allows for an investigation into the effects of light change. It is demonstrated that the data set is large enough to reach statistically significant conclusions on central aspects of MVS, which we see as a main contribution. We have made all relevant data of this dataset available for free download[4].

The three state-of-the-art MVS methods by Campbell et al. [7], Furukawa and Ponce [12], and Tola et al. [35] have been applied to the dataset, thus giving a benchmark for others to compare against, validating that reasonable results can be achieved from our dataset, and lastly illustrating some of the challenges of modern-day MVS.

As for the latter, our investigations showed several things. Firstly, we observed a tradeoff between accuracy and completeness in the three methods, such that the method by *Tol*[35] has highest accuracy but lowest completeness whereas *Cam*[7] obtained the highest completeness but lowest accuracy. This trade-off can be caused by the extent of discrimination towards reconstructed points in the respective methods. High discrimination gives good accuracy but less completeness, whereas the opposite is seen with less discrimination.

Secondly, many of the issues that are typically very disruptive for two-view stereo, such as changing lighting conditions and specular surfaces, surprisingly showed not to be a main issue for MVS. Our hypothesis is that all the employed methods use robust aggregates of two view stereo, implying that if just a few image pairs are good for every part of the surface, then the result will in general not degenerate. The lack of texture, however, still seems to be a main challenge.

The three applied MVS methods were both evaluated in relation to estimated surface points and triangulated surfaces. We observed that surface meshing has a smoothing effect, which is beneficial for simple geometries, because it tends to fill out holes. In general, the effect of meshing does not, however, improve the performance, because small details are generally corrupted. This demonstrates the need to improve meshing algorithms in relation to MVS.

As future work, we aim to get an even better understanding of how surface properties influence the MVS quality. To do this we are contemplating a data set with single 'atomic' surface properties, e.g. a single wood slab, as a supplement to the more varied scenes of the presented data set. This would hopefully allow us to better model the relationship between surface properties and MVS reconstruction quality, by better isolating the effects. In regard to this, the study presented here has given us valuable insights into what properties such atomic surfaces should span.

[4] http://roboimagedata.compute.dtu.dk/

References

1. Aanæs, H., Dahl, A., Steenstrup Pedersen, K.: Interesting interest points. IJCV **97**, 18–35 (2012)
2. Anderson, T.: An Introduction to Multivariate Statistical Analysis. Wiley & Sons (1984)
3. Bailer, C., Finckh, M., Lensch, H.P.A.: Scale robust multi view stereo. In: ECCV, pp. 398–411. Springer-Verlag (2012)
4. Beraldin, J.A., Gaiani, M.: Evaluating the Performance of Close Range 3D Active Vision Systems for Industrial Design Applications. In: SPIE: Videometrics IX, vol. 5665 (2005)
5. Boehler, W., Vicent, B., Marbs, A.: Investigating laser scanner accuracy. In: CIPA (2003)
6. Boykov, Y., Kolmogorov, V.: Computing geodesics and minimal surfaces via graph cuts. In: ICCV, pp. 26–33 (2003)
7. Campbell, N.D., Vogiatzis, G., Hernández, C., Cipolla, R.: Using multiple hypotheses to improve depth-maps for multi-view stereo. In: ECCV, pp. 766–779 (2008)
8. Cook, R.L., Torrance, K.E.: A reflectance model for computer graphics. SIGGRAPH Comput. Graph. **15**(3), 307–316 (1981)
9. Curless, B., Levoy, M.: A volumetric method for building complex models from range images. In: Conference on Computer Graphics and Interactive Techniques, pp. 303–312. ACM (1996)
10. Fuhrmann, S., Goesele, M.: Floating scale surface reconstruction. ACM Transactions on Graphics (TOG) **33**(4), 46 (2014)
11. Furukawa, Y., Curless, B., Seitz, S.M., Szeliski, R.: Towards internet-scale multi-view stereo. In: CVPR, pp. 1434–1441 (2010)
12. Furukawa, Y., Ponce, J.: Accurate, dense, and robust multiview stereopsis. PAMI **32**(8), 1362–1376 (2010)
13. Goesele, M., Curless, B., Seitz, S.M.: Multi-view stereo revisited. In: CVPR, pp. 2402–2409 (2006)
14. Hernández, C., Vogiatzis, G., Cipolla, R.: Probabilistic visibility for multi-view stereo. In: CVPR, pp. 1–8 (2007)

15. Hiep, V.H., Keriven, R., Labatut, P., Pons, J.P.: Towards high-resolution large-scale multi-view stereo. In: PAMI, pp. 1430–1437 (2009)
16. Hu, X., Mordohai, P.: Evaluation of stereo confidence indoors and outdoors. In: CVPR, pp. 1466–1473 (2010)
17. Jensen, R., Dahl, A., Vogiatzis, G., Tola, E., Aanaes, H.: Large scale multi-view stereopsis evaluation. In: CVPR, pp. 406–413 (2014)
18. Kazhdan, M., Bolitho, M., Hoppe, H.: Poisson surface reconstruction. In: Eurographics symposium on Geometry processing, pp. 61–70 (2006)
19. Kim, S., Kim, S., Dahl, A., Conradsen, K., Jensen, R., Aanæ s, H.: Multiple view stereo by reflectance modeling. 2011 International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission (2012)
20. Klowsky, R., Kuijper, A., Goesele, M.: Modulation transfer function of patch-based stereo systems. In: CVPR, pp. 1386–1393 (2012)
21. Kolev, K., Brox, T., Cremers, D.: Fast joint estimation of silhouettes and dense 3D geometry from multiple images. PAMI **34**(3), 493–505 (2012)
22. Labatut, P., Pons, J.P., Keriven, R.: Robust and efficient surface reconstruction from range data. In: Computer Graphics Forum, vol. 28, pp. 2275–2290. Wiley Online Library (2009)
23. Liu, S., Cooper, D.B.: A complete statistical inverse ray tracing approach to multi-view stereo. In: CVPR, pp. 913–920 (2011)
24. Luhmann, T.: Comparison and verification of optical 3-d surface measurement systems. In: Int. Archives Photogram., Remote Sensing Spatial Inf. Sci., vol. 37 (2008)
25. Møller, B., Balslev, I., Krüger, N.: An automatic Evaluation Procedure for 3D Scanners in Robotics Applications. IEEE Sensors Journal **13**(2), 870–878 (2013)
26. Mücke, P., Klowsky, R., Goesele, M.: Surface reconstruction from multi-resolution sample points. In: VMV, pp. 105–112. Citeseer (2011)
27. Nair, R., Meister, S., Lambers, M., Balda, M., Hofmann, H., Kolb, A., Kondermann, D., Jähne, B.: Ground truth for evaluating time of flight imaging. In: Time-of-Flight and Depth Imaging, vol. 8200, pp. 52–74. Springer (2013)
28. Newcombe, R.A., Lovegrove, S.J., Davison, A.J.: Dtam: Dense tracking and mapping in real-time. In: ICCV, pp. 2320–2327 (2011)
29. Salvi, J., Fernandez, S., Pribanic, T., Llado, X.: A state of the art in structured light patterns for surface profilometry. Pattern recognition **43**(8), 2666–2680 (2010)
30. Salvi, J., Pages, J., Batlle, J.: Pattern codification strategies in structured light systems. Pattern Recognition **37**(4), 827–849 (2004)
31. Scharstein, D., Szeliski, R.: High-accuracy stereo depth maps using structured light. In: CVPR, vol. 1, pp. I–195 (2003)
32. Seitz, S.M., Curless, B., Diebel, J., Scharstein, D., Szeliski, R.: A comparison and evaluation of multi-view stereo reconstruction algorithms. In: CVPR, vol. 1, pp. 519–528 (2006)
33. Shalom, S., Shamir, A., Zhang, H., Cohen-Or, D.: Cone carving for surface reconstruction. In: ACM Transactions on Graphics (TOG), vol. 29, p. 150. ACM (2010)
34. Strecha, C., von Hansen, W., Van Gool, L., Fua, P., Thoennessen, U.: On benchmarking camera calibration and multi-view stereo for high resolution imagery. In: CVPR, pp. 1–8 (2008)
35. Tola, E., Strecha, C., Fua, P.: Efficient large-scale multi-view stereo for ultra high-resolution image sets. Machine Vision and Applications **23**(5), 903–920 (2012)
36. Tukey, J.W.: Exploratory data analysis, vol. 1. Pearson - Addison-Wesley (1977)
37. Vogiatzis, G., Hernández, C., Torr, P.H.S., Cipolla, R.: Multiview stereo via volumetric graph-cuts and occlusion robust photo-consistency. PAMI **29**(12), 2241–2246 (2007)
38. Wendel, A., Maurer, M., Graber, G., Pock, T., Bischof, H.: Dense reconstruction on-the-fly. In: CVPR, pp. 1450–1457 (2012)