3D Models from the Black Box: Investigating the Current State of Image-Based Modeling

Hoang Minh Nguyen
The University of Auckland,
New Zealand
hngu039@aucklanduni.ac.nz

Burkhard Wünsche
The University of Auckland,
New Zealand
burkhard@cs.auckland.ac.nz

Patrice Delmas
The University of Auckland,
New Zealand
p.delmas@cs.auckland.ac.nz

Christof Lutteroth
The University of Auckland,
New Zealand
lutteroth@cs.auckland.ac.nz

ABSTRACT

3D models have become an essential part of many applications ranging from computer games to architectural design, virtual heritage, and visual impact studies. Traditionally, 3D model creation is done using modelling systems such as Maya or Blender. However, these systems have a steep learning curve and require a considerable amount of training to use. Thus, there is a critical need for tools which allow non-expert users to easily and efficiently create complex 3D scenes. To answer to that demand, a number of commercial image-based modelling packages have been introduced recently. Such software offers a very intuitive means to create 3D models from a sequence of images. However, the algorithms employed by these systems are usually kept secret, which makes it difficult to compare them algorithmically and identify common underlying concepts. This paper evaluates the most promising 3D reconstruction software packages with regard to efficiency, accuracy, limitations, constraints and compares them with a system developed by us in order to give an insight into their performance. To achieve that, we first describe our own 3D reconstruction system as a reference in order to make deductions about common concepts and differences. Then, we use a set of benchmark datasets to evaluate all considered systems, and gage their limitations with regard to the number of input images they need and the image resolution. Our evaluation shows that as the number of input images decreases, the geometry of models created using correspondence-based approaches contains more holes. However, the structure and geometry still reflect the original model. In contrast, silhouette-based methods produce coarse and distorted geometry as the number of input images decreases. Models obtained using silhouette-based methods from few input images are often unrecognizable.

Keywords

image-based modeling, correspondence-based reconstruction, silhouette-based reconstruction

1 INTRODUCTION

Creating 3D models of a scene has long been an important task in computer graphics. While conventional geometry-based modeling approaches enable the construction of highly realistic and complex 3D models via interaction with 3D meshes, they have a steep learning curve and require a considerable amount of training to use. These restrictions render them unsuitable for non-expert users. The recent advancement in hardware specialized in 3D model reconstruction has made it possible for non-professionals to reconstruct 3D scenes. However, tools such as laser scanners and structured lighting systems are often costly, have a limited range and resolution, are not very portable and flexible to use. Additionally, they have constraints with respect to material properties and environmental conditions such as string sunlight.

Digital cameras overcome many of these limitations and their ubiquitous use and integration into computing devices such as smart phones is making them an increasingly attractive proposition for 3D scene reconstruction. Recovering 3D structure from photographic images is an efficient and intuitive way to create 3D digital models of objects. Compared with conventional geometry-based modeling and hardware-heavy approaches, the image-based modeling method can be employed to extract original texture and illumination directly from images for visual 3D modeling, without the need for complicated processes, such as geometry modeling, shading and ray tracing. The techniques are

usually less accurate, but offer very intuitive and low-cost methods for reconstructing 3D scenes and models.

The past few years have seen significant progress toward automatic creation of 3D models. There are now a number of software packages that offer the ability to acquire 3D models from a set of images without any a priori information about the scene to be reconstructed. Once supplied with the input images, these systems automatically process and produce a 3D model. As the algorithms used by these systems to reconstruct 3D models are usually kept secret, it is difficult to identify fundamental limitations and commonalities, and hence to compare them on an algorithmic level. The objective of this paper is to give an insight into the performance of the currently best-performing systems by evaluating them using benchmark datasets. We use the following methodology to provide some insight into the current state of image-based modeling: we include a reference system that allows us to make grounded assumptions on algorithmic differences, and we systematically vary the number of input images and their resolution to identify the current limitations.

After a description of related work on image-based modeling, a brief overview of our reference system is presented. In the second part, a set of benchmark datasets is used to stress test these systems under various conditions.

2 RELATED WORK

Ideally, image-based modeling algorithms can be categorized depending on the visual cues employed to perform reconstruction, i.e. silhouettes, texture, transparency, defocus, shading or correspondence. Traditionally, the most well-known and successful visual cues have been shading, silhouettes, and correspondence [HVC08]. Silhouettes and correspondence offer the highest degree of robustness due to their invariance to illumination changes. The shading cue requires more control over the illumination environment, but can produce excellent results [HVC08]. However, the requirement for strict constraints over lighting conditions renders shape from shading impractical for general applications.

The shape from silhouette class of algorithms is very efficient and has been proved to be stable with regard to object surface properties (color, texture and material). It is, however, very limited in the object geometries it can handle [FLB06, MBR$^+$00, NWDL11]. The earliest attempt of using silhouettes for 3D shape reconstruction was made by Baumgart in 1974. In his pioneering work [Bau74], Baumgart exploited silhouette information from four input images to compute the 3D shapes of a baby doll and a toy horse. Following Baumgart's work, many different variations of the shape from silhouette paradigm have been proposed

Grauman et al. [GSD03] presented a Bayesian approach to account and compensate for errors introduced as the result of false segmentation. The approach has been shown to produce excellent error-compensated models from erroneous silhouette information. The disadvantage of this method is that it requires prior knowledge about the objects to be reconstructed and large ground-truth training data. This makes them impractical for general applications.

Cheung et al. [CA84, mCBK05a, mCBK05b] proposed a method that aligns multiple silhouette images of a non-rigidly moving object over time in an attempt to improve the quality of the constructed visual hull. Their method showed a significant improvement in reconstruction quality over previous methods.

Amongst the vast body of literature available on image-based modeling techniques, recent work on multiple view reconstruction has become a growing area of interest with many different techniques achieving a high degree of accuracy. These techniques are based mainly on correspondence cues and focus on producing models that resemble the original 3D scene from a sequence of calibrated or uncalibrated images. The concept underpinning these techniques is the extraction and combination of information from several overlapping images taken from distinct locations at different instants to deduce the relations between those images. When relations between images are properly established, the 3D structure of the observed scene can be inferred.

One of the most famous and successful reconstruction systems is the Façade system, which was proposed by Debevec et al. [DTM96]. The Façade system was designed to model and render simple architectural scenes by combining a hybrid geometric and image-based approach. The system requires only a few images and some known geometric parameters. It was used to reconstruct compelling fly-throughs of the Berkeley campus and was employed for the MIT City Scanning Project, which captured thousands of calibrated images from an instrumented rig to compute a 3D model of the MIT campus. While the resulting 3D models are often impressive, the system requires considerable time and effort from the user to decompose the scene into prismatic blocks and manually select features and their correspondence in different views, followed by the estimation of the pose of these primitives. Consequently, the system is impractical for reconstructing large scenes.

More recently, Xiao et al. [XFT$^+$08] developed a semi-automatic image-based approach to reconstruct 3D façade models of high visual quality from a sequence of street view images. Their method employed a systematic and automatic decomposition scheme of façades for both analysis and reconstruction. The decomposition is achieved by a recursive subdivision that partitions the whole façades into small segments,

while still preserving the architectural structure. Users are required to provide feedback on façade partition. This method demonstrated excellent results.

Brown et al. [BL05] presented an image-based modeling system that aims to recover camera parameters, pose estimates and sparse 3D scene geometry from a sequence of images. Snavely et al. [SSS06] introduced the Photo Tourism (Photosynth) system which is based on the work of Brown, with some significant modifications to improve scalability and robustness. Agarwala et al. [AAC+06] proposed another related technique for composing panoramas of roughly planar scenes. Although these approaches address the same SfM concepts as we do, their aim is not to reconstruct and visualize 3D scenes and models from images, but only to allow easy navigation between images in three dimensions.

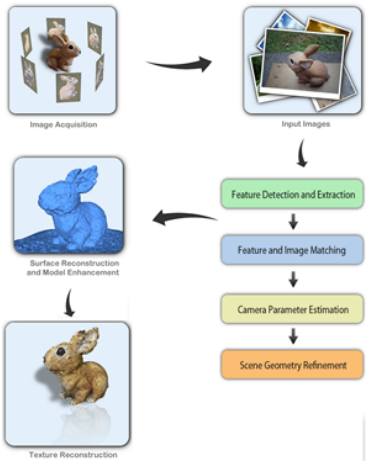

Figure 1: Stages of the reconstruction process.

3 DESIGN OF 3D RECONSTRUCTION ALGORITHMS

Ideally, 3D reconstruction algorithms can be categorized depending on the visual cues employed to perform reconstruction, i.e. silhouettes, texture, transparency, defocus, shading or correspondence. The best-known and most successful visual cues have been shading, silhouettes, and correspondence [HVC08, Qua10]. Silhouettes and correspondence offer the highest degree of robustness due to their invariance to illumination changes. The shading cue requires more control over the illumination environment, but can produce excellent results.

In this section, we review and analyze the two most popular reconstruction techniques: silhouettes and correspondence based methods. In order to be able to reconstruct a 3D scene without any a prior knowledge, the intrinsic and extrinsic parameters of the camera being used must first be estimated. This process is further divided into three sub-steps: feature extraction, feature matching, and camera parameter estimation. 3D scene geometry can then be recovered by either back projecting and interpolating 3D points (correspondence-based), or using silhouette information (silhouette-based). Figure 1 depicts several stages of the reconstruction process.

3.1 Feature Detection and Extraction

The objectives of this step are to identify features of interest in each image and to match the features across views. The accuracy of the entire reconstruction process relies on the features of the scene that can be identified, extracted and automatically matched. Consequently, occlusions, illumination variation, limited locations for the image acquisition and reflective surfaces are problematic. However, recent invariant feature detector, such as SIFT [BL05], have proved to be fairly robust under large image variations. Feature points extracted by SIFT are highly distinctive and invariant to different transformations and changes in illumination, as well as having a high information content [BL05, HL07].

The SIFT operator works by first identifying potential points of interest. This is achieved by isolating points located at the extrema of the Difference-of-Gaussian (DoG) function in scale space. The location and scale of each key point is then computed and key points are selected based on measures of stability. Unstable extremum points (key points with low contrast or edge response features along an edge) are rejected as they are too sensitive to noise for accurate localization. Each detected key point is then assigned one or more consistent canonical orientations based on local image gradients. The key point descriptor is then described relative to this canonical orientation, thereby achieving invariance to rotation. Finally, using local image gradient information, a descriptor is produced for each key point [SSS06]. Figure 2 shows an example of detected key points from the Queen Victoria statue dataset (Auckland, New Zealand) before and after localization.

Figure 2: Left: Candidate key points detected from the first stage. Middle: After discarding low contrast key points. Right: After discarding key points located on edges (key-points near the edges and corners of images are now removed).

3.2 Feature Matching

Once features have been identified and extracted from all the images, they are matched. This is known as the correspondence problem. Given a feature in an image I_1, what is the corresponding feature (the projection of the same 3D point) in the other image I_2? This can be solved by defining a distance function that compares the two feature descriptors. All the detected features in I_2 are tested and the one with the minimum distance is selected [CA84]. The Euclidean distance is employed to measure the similarity between two key points A and B.

A small distance indicates that the two key points are close and thus similar. However, a small distance does not necessarily mean that the points represent the same feature. For instance, a scene can contain many similar features such as corners of windows in an large building. It merely indicates that the two features have the highest resemblance of all processed features. In order to accurately match a key point in the candidate image, we identify the closest and second closet key points in the reference image using a nearest neighbor search strategy. If the ratio of them is below a given threshold, the key point and the closest matched key point are accepted as correspondences, otherwise that match is rejected [Low04, Low99].

Since multiple images may view the same point in the world, each image is matched to the nearest neighbors. During this process, image pairs whose number of corresponding features is below a certain threshold are removed. In our experiment, the threshold value of 20 seems to produce the best results.

As the matching procedure is subject to errors and mismatches, many of our matches are spurious. It is possible to eliminate many spurious matches by enforcing a geometric consistency. This is predicated on the fact that, assuming a stationary scene, not all corresponding features between two images are physically resizable, regardless of what the actual shape of the scene is. This geometric constraint is known as the epipolar constraint. The epipolar constraint requires that a pair of corresponding features, $(x_1, y_1) \rightarrow (x_2, y_2)$ between two images satisfies the equation:

$$\begin{bmatrix} x_2 & y_2 & 1 \end{bmatrix} \mathrm{F} \begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} = 0 \qquad (1)$$

Where F denotes the *Fundamental matrix*, which defines a bilinear constraint between the coordinates of corresponding image points. Thus, for a given image pair, only matching features that agree with the epipolar constraint are admissible. All other matches are rejected.

3.3 Camera Parameter Estimation

Given a set of matching images, the goal of this stage is to recover the geometry of the scene and the motion information of the camera (camera parameters) simultaneously. The motion information includes the extrinsic (position, orientation) and intrinsic parameters of the camera for the captured images. This is accomplished using the Structure from Motion (SfM) technique [SSS06, Sze, COM+11].

The reconstruction process begins by estimating parameters for an initial pair. The selection of the initial image pair to be reconstructed is highly critical. If the reconstruction of this initial pair gets stuck in undesirable local minima, the optimization is unlikely to ever converge. To avoid such cases, the initial pair must be selected carefully. The chosen images should have a large number of correspondences, but also have a relatively large baseline (the distance between camera optical centers). This is to ensure that the location of the 3D observed point is well-conditioned, so that the initial two-frame reconstruction can be robustly estimated.

The estimation of the extrinsic parameters for this initial pair is as follows [SSS06]: First, the Essential matrix is approximated using the five-point algorithm. Next, the projection matrix can be retrieved by decomposing the Essential matrix. Feature tracks visible in the two images are then triangulated, giving an initial set of 3D points. Once the structure of the scene and the motion information have been estimated for the first pair, they are further refined using Bundle Adjustment. Bundle Adjustment refines a visual reconstruction to produce the optimal 3D structure and motion information. This last step is critical for the accuracy of the reconstruction, as concentration of pairwise homographies would accumulate errors and disregard constraints between images. The recovered geometry parameters should be consistent. That is, the reprojection error, which is defined by the distance between the projections of each key point and its observations, is minimized (Figure 3).

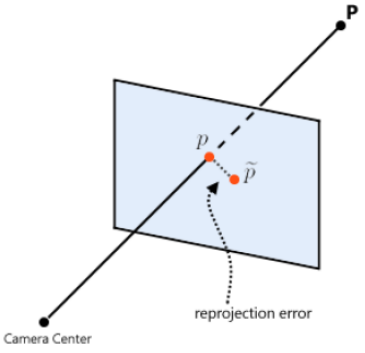

Figure 3: The reprojection error is the distance between the projected image point p and the observed image point $\tilde{p}$.

Subsequent images are added to the optimization one at a time, with the best matching image being added at each step. Best matching images are those that share the largest number of tracks whose 3D locations have already been estimated. Each new added image is initialized with the same orientation, and focal length as the image that it matches best. This has proved to work very well even though images have different rotation and scale. Next, Bundle Adjustment is applied to refine the solution. This procedure is repeated, one image at a time, and terminates when no more images can reliably be added. The reliability test is determined based on the number of correspondences. A camera is only added when it shares a sufficient number of correspondences. In our system, we use a threshold value of 25, which was empirically selected from our experiments. Figure 4 demonstrates several stages of the SfM algorithm.

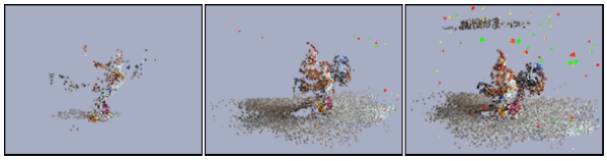

Figure 4: Several SfM stages of the reconstruction of our Rooster dataset. Left: the initial two-frame reconstruction. Middle: an intermediate stage after 20 images have been added. Right: the final construction with 42 images.

3.4 Correspondence- and Silhouette-based Reconstruction

Once the camera parameters have been successfully estimated, the 3D scene geometry can be computed. For correspondence-based approaches, the final steps involved triangulating correspondences to obtain a sparse set of 3D points. This initial sparse set of points then undergoes a refinement process which often involves denoising and resampling. Surfaces are then applied onto the point clouds to produce the final 3D model [ASG, SSS06, HL07, NWDL11, FP09].

For silhouette-based approaches, the 3D model is produced by exploiting silhouette information to create intersected visual cones, which are subsequently used to derive the 3D representation of an object. The construction of these cones requires the coordinates of a set of silhouette contour vertices for a given camera, and the coordinates of the camera's optical center as input. Rays extrapolated from the camera's optical center through the contour image points and beyond define a silhouette cone for that view, which is guaranteed to contain the original object. The intersection of silhouette cones from different viewpoints defines a polyhedral visual hull as an approximation of the object [HVC08, Lam02, FB10, LW10].

4 EVALUATION

In this section, we present an evaluation of four 3D reconstruction systems. These include two systems that seem to be correspondence-based (our own correspondence-based system and *Hyper3D*) and two reconstruction systems that seem to be silhouette-based (*Agisoft* and *123D Catch*). These systems were chosen because they are among the best with regard to the quality of reconstruction. Our goal is to evaluate their efficiency, accuracy, constraints and limitations in order to provide insight into the current state of image-based reconstruction in general.

In order to evaluate a system we used a repository of 40 objects. After an initial tests using different objects we selected one object which reflected the capabilities of all tested algorithms and used it to investigate the effect of image acquisition parameters. For this selected object, we created eight different datasets. Some of the input images are shown in Figure 5. The datasets vary in the number of input images they contain and the resolution of the input images.

Figure 5: Four input images from four distinct viewpoints.

All images are taken with a *Logitec Webcam* in an indoor environment. Some input images are intentionally captured with other unrelated moving objects, to test how different algorithms handle unrelated moving object in the scene. The original model has a bumpy surface with a reasonable number of distinctive features.

4.1 Effect of the Number of Input Images

The objective of this test is to determine the effect of the number of input images on reconstruction quality. We evaluated 4 datasets of the bird model with a constant resolution of 1600 $\times$ 1200 and varying numbers of input images: 30, 20, 12 and 7. Figure 6 shows the reconstruction results for all the evaluated systems. The perspectives in which the reconstructed models are shown in the table were chosen so that reconstruction deficiencies are most visible.

30 Images For this dataset, our system, *123D Catch*, and *Agisoft* produce a qualitatively good model. There

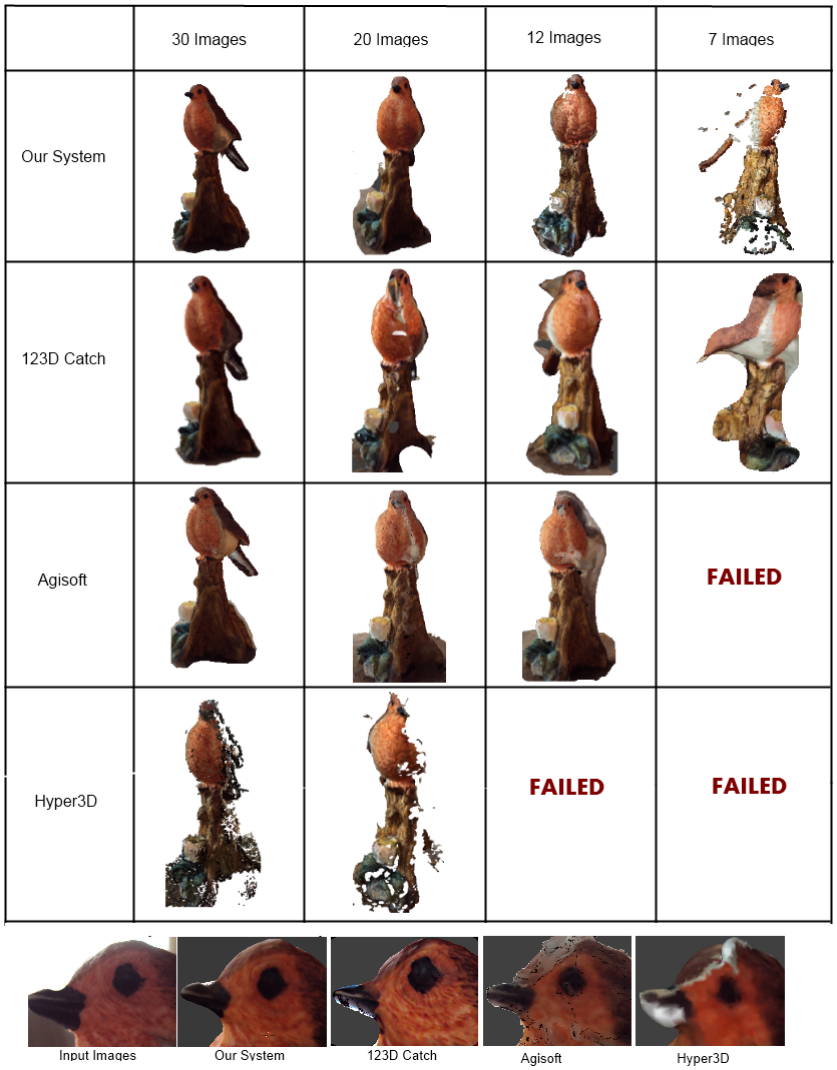

Figure 6: Reconstruction results of the first test suit. Last row: Close-up screen shots of the reconstructed models from 30 images.

is a slight deformation (on the side) in *123D Catch*'s model, but overall the resulting reconstruction is visual-pleasing. Our reconstruction has a more blurry texture than that generated using *123D Catch*. This is because we generate the texture by performing color interpolation between 3D points while they use a projection-based texture. However, in terms of geometry, our reconstruction bears the highest resemblance to the original model. *Hyper3D* was only able to construct a partial model.

20 Images *Agisoft*'s system produced a reasonably good model, although there is a slight disruption in the geometry in the chest of the bird model. *123D Catch*'s model is not as good. There is a large chunk of the background glued to the model. This is probably caused during the background subtraction process in which the object was not properly segmented. The model produced from our system has the best geometry, however the texture has become even more blurry. This is understandable as there are fewer distinctive features in the input image sequence leading to fewer 3D points generated. *Hyper3D*, again, was only able to produce a partial reconstruction.

12 Images For the third dataset, *Hyper3D* was unable to produce any result. *Agisoft* model's geometry was disfigured. There is a large bit missing on the side of the model. *123D Catch* has reasonably good geometry, although similar to the previous case there is a bit of the background attached to model (Figure 6). For our model, there is a small missing region at the top of the model. Apart from that, the geometry still retains the highest resemblance to the original model.

7 Images For this dataset, the reconstruction results from our system and *123D Catch* are shown in Figure

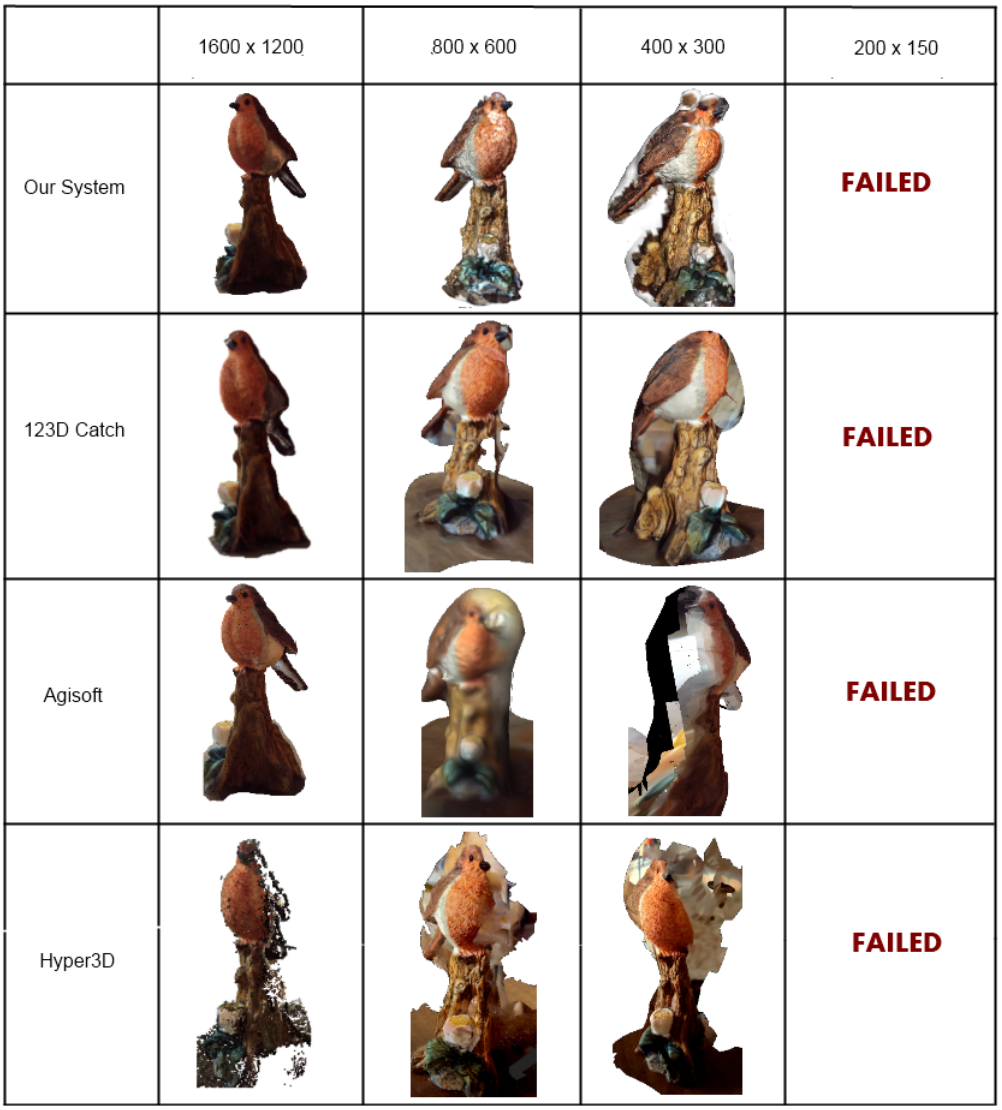

Figure 7: Reconstruction results of the second test suit.

6. *Agisoft* and *Hyper3D* system were unable to produce any result. Both *123D Catch* and our system were only able to construct a partial model. Although the geometry of *123D Catch* model seems more complete, it is almost unrecognisable and does not share much in common with the original model. In contrast, the model created using our system still has some resemblance to the original model. This test indicates that our system and *123D Catch* are amongst the most robust with regard to limited number of input images.

The result of this test clearly differentiate correspondence-based from silhouette-based approaches for a decreasing number of input images. Correspondence-based approaches, although producing models with good geometry, tend to have more missing geometry when the number of input images decreases. Additionally, textures generated from this approach are often blurry as the result of interpolation. Silhouette-based approaches do not create holes, but they show coarse and distorted geometry for small numbers of images. The resulting models become more refined with an increasing number of input images. To be able to construct a reasonable quality model of a small sized object, today's systems would need at least 20 input images from a standard consumer-level camera. Adding more than 30 images does not improve the reconstruction quality significantly. Models reconstructed from 12 images or less are usually unsatisfactory.

4.2 Effect of the Input Image Resolution

We aim to evaluate the performance of these systems with regard to image resolution. For this test suit, we reduce the resolution of the input images. There

	Our System	123D Catch	Agisoft	Hyper3D
Speed	Medium	Fast	Slow	Medium
Geometry	Good	Good	Good	Average
Texture	Average	Good	Good	Good
#Images	Small	Small	Medium	Medium
Resolution	Small	Small	High	High
Constraints	None	None	None	None
Min #images	12	12	12	20
Min resolution	400×300	400×300	800×600	1600×1200

Table 1: Summary of the four system's performance.

are four datasets in this test suit, each contains 30 input images with resolution of 1600 × 1200, 800 × 600, 400 × 300, and 200 × 150 respectively. The objective is to stress the systems further to determine how well each system performs in the case of low resolution input images. Figure 7 illustrates the resulting reconstructions from all the systems.

1600 × 1200 Due to the large resolution of the input images, most resulting models are well reconstructed. *Hyper3D* produces the worst model, which has a large missing region.

800 × 600 For this dataset, *123D Catch*'s system yields the most qualitatively accurate model which has its texture properly recreated, although there remain many missing regions in the final model. Our model has the most complete and well-reconstructed geometry of all resulting models. Our texture, however, is noisy. Models from *Agisoft* and *Hyper3D* are mostly disfigured. One half of the reconstructed models has completely vanished. This is mostly due to the fact that their systems are not able to register views when there only a limited number of features in each input images. In the case of *Agisoft*, the texture appears very blurry.

400 × 300 For this test, *Agisoft*'s model is completely unrecognizable. There is a large bit missing from the reconstructed model. Our system, *123D Catch* and *Hyper3D* were able to produce some outputs. Although the resulting reconstructions are only partial. This is also the result of insufficient number of distinctive features, which leads to failure to establish global image correspondence. Models from our system and *Hyper3D* are reasonably reconstructed. The resulting models still reflect the structure of the original object. In the case of *123D Catch*, the resulting model bears almost no resemblance to the original

200 × 150 In this test, all the systems were unable to register images due to insufficient overlap between image features.

The results show that silhouette-based approaches seem to be less robust for low resolutions. This is probably because silhouette-based methods are naturally deterministic and do not account for errors that might be present in views. The errors are typically caused by inaccurately estimated camera parameters. For silhouette-based systems, resolution significantly below 1600 × 1200 does not seem to yield satisfactory models anymore. Correspondence-based approaches are slightly more robust with regard to image resolution.

Models reconstructed with low-resolution images using correspondence-based approaches often have noisy surfaces, but appear more complete. However, for correspondence-based approaches, a resolution below 800 x 600 does not seem to produce satisfactory models anymore.

5 CONCLUSION

We described the overall design of image-based reconstruction algorithms, and evaluated a number of 3D reconstruction systems. The evaluation shows that there are general differences between the different algorithms, particularly between correspondence-based and silhouette-based algorithms.

Correspondence-based algorithms produce good details for larger numbers of input images ($\geq$20), but tend to produce missing geometry (holes) as the number of input images decreases. The textures they generate are often blurry because of interpolation. They are fairly robust with regard to image resolution and still produce models with fairly complete geometry for low resolutions, although the surfaces become noisy.

Silhouette-based approaches do not create holes, but they show coarse and distorted geometry for small numbers of images. They tend to produce better textures because they backproject the original images as the silhouettes are constructed. However, they tend to be less robust for low-resolution input images, as they are more sensitive to camera parameter estimation errors. A summary about various aspects of the four systems is shown in Table 1.

To gain a deeper understanding into today's reconstruction algorithms, it is necessary to investigate the effect of other parameters such as illumination, distortion, occlusion and object types.

6 REFERENCES

[AAC+06] Aseem Agarwala, Maneesh Agrawala, Michael Cohen, David Salesin, and Richard Szeliski. Photographing long scenes with multi-viewpoint panoramas. *In ACM Transactions on Graphics*, 25(3):853–861, 2006.

[ASG] Pierre Alliez, Laurent Saboret, and Gael Guennebaud. Surface reconstruction from point sets. Available at `http://www.cgal.org/Manual/3.5/doc_html/cgal_manual/Surface_reconstruction_points_3/Chapter_main.html`. Last accessed on May 22nd 2011.

[Bau74] Bruce Guenther Baumgart. *Geometric modeling for computer vision*. Doctoral Dissertation, Stanford University, 1974.

[BL05] Matthew Brown and David Lowe. Unsupervised 3D object recognition and reconstruction in unordered datasets. *In International Conference on 3D Digital Imaging and Modelling*, pages 56–63, 2005.

[CA84] Chang Ho Chien and J. K Aggarwal. A volume surface octree representation. *In Seventh International Conference on Pattern Recognition, Montreal, Canada*, pages 817–820, 1984.

[COM+11] Wei Cheng, Wei Tsang Ooi, Sebastien Mondet, Romulus Grigoras, and Geraldine Morin. Modeling progressive mesh streaming: Does data dependency matter. *ACM Transaction on Multimedia Computing*, pages 1–24, 2011.

[DTM96] Paul E Debevec, Camillo J Taylor, and Jitendra Malik. Modeling and rendering architecture from photographs: A hybrid geometry and image-based approach. *In ACM Transactions on Graphics*, pages 11–20, 1996.

[FB10] Jean Sebastien Franco and Edmond Boyer. Efficient polyhedralmodeling from silhouettes. *IEEE Transaction on Pattern Analysis and Machine Intelligences*, 31:853–861, 2010.

[FLB06] Jean-Sebastien Franco, Marc Lapierre, and Edmond Boyer. Visual shapes of silhouette sets. *In 3D Data Processing, Visualization and Transmission*, pages 397–404, 2006.

[FP09] Yasutaka Furukawa and Jean Ponce. Accurate, dense, and robust multi-view stereopsis. *In IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(8):1362–1376, 2009.

[GSD03] Kristen Grauman, Gregory Shakhnarovich, and Trevor Darrell. A bayesian approach to image-based visual hull reconstruction. *In IEEE International Conference on Computer Vision and Pattern Recognition*, 1:187–194, 2003.

[HL07] Shungang Hua and Ting Liu. Realistic 3D reconstruction from two uncalibrated views. *In International Journal of Computer Science and Network Security*, 7:178–183, 2007.

[HVC08] Carlos Hernandez, George Vogiatzis, and Roberto Cipolla. Multi-view photometric stereo. *In IEEE Transaction on Pattern Recognition and Machine Intelligence*, 30:548–554, 2008.

[Lam02] Bruce Lamond. *An Investigation into the Recovery of Three-Dimensional Structure from Two-Dimensional Images*. Master Thesis, School of Computer Science, University of Edinburgh, 2002.

[Low99] David G Lowe. Object recognition from local scale-invariant features. *In International Conference on Computer Vision*, 2:1150–1157, 1999.

[Low04] David G Lowe. Distinctive image features from scale-invariant keypoints. *In International Journal of Computer Vision*, 60:91–110, 2004.

[LW10] Chen Liang and Kwan Yee Wong. 3d reconstruction using silhouettes from unordered viewpoints. *Image and Vision Computing*, 28(4):579–589, 2010.

[MBR+00] Wojciech Matusik, Chris Buehler, Ramesh Raskar, Steven Gortler, and Leonard McMillan. Image-based visual hulls. *In Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pages 369–374, 2000.

[mCBK05a] Kon man Cheung, Simon Baker, and Takeo Kanade. Shape-from-silhouette across time part 1: Theory and algorithms. *In International Journal of Computer Vision*, 62(1):221–247, 2005.

[mCBK05b] Kon man Cheung, Simon Baker, and

Takeo Kanade. Shape-from-silhouette across time part 2: Applications to human modeling and markerless motion tracking. *In International Journal of Computer Vision*, 63(1):225–245, 2005.

[NWDL11] Hoang Minh Nguyen, Burkhard Wunsche, Patrice Delmas, and Christof Lutteroth. Realistic 3d scene reconstruction from unconstrained and uncalibrated images. *In Proceedings of GRAPP 2011, Algarve, Portugal*, 31:67–75, 2011.

[Qua10] Long Quan. *Image-Based Modeling*. Springer Press, 2010.

[SSS06] Noah Snavely, Steven Seitz, and Richard Szeliski. Photo tourism: Exploring photo collections in 3D. *In ACM Transactions on Graphics*, 25(3):835–846, 2006.

[Sze] Richard Szeliski. Image alignment and stitching. A tutorial in Computer Graphics and Vision, 2006. Available at `http://research.microsoft.com/apps/pubs/default.aspx?id=70092`. Last accessed on May 21st 2011.

[XFT+08] Jianxiong Xiao, Tian Fang, Ping Tan, Peng Zhao, Eyal Ofek, and Long Quan. Image-based façade modeling. *In ACM Transactions on Graphics*, 27(5):26–34, 2008.