A Surface-Growing Approach to Multi-View Stereo Reconstruction

Martin Habbecke and Leif Kobbelt
Computer Graphics Group, RWTH Aachen University
http://www.graphics.rwth-aachen.de

Abstract

We present a new approach to reconstruct the shape of a 3D object or scene from a set of calibrated images. The central idea of our method is to combine the topological flexibility of a point-based geometry representation with the robust reconstruction properties of scene-aligned planar primitives. This can be achieved by approximating the shape with a set of surface elements (surfels) in the form of planar disks which are independently fitted such that their footprint in the input images matches. Instead of using an artificial energy functional to promote the smoothness of the recovered surface during fitting, we use the smoothness assumption only to initialize planar primitives and to check the feasibility of the fitting result. After an initial disk has been found, the recovered region is iteratively expanded by growing further disks in tangent direction. The expansion stops when a disk rotates by more than a given threshold during the fitting step. A global sampling strategy guarantees that eventually the whole surface is covered. Our technique does not depend on a shape prior or silhouette information for the initialization and it can automatically and simultaneously recover the geometry, topology, and visibility information which makes it superior to other state-of-the-art techniques. We demonstrate with several high-quality reconstruction examples that our algorithm performs highly robustly and is tolerant to a wide range of image capture modalities.

1. Introduction

In recent years the reconstruction of 3D geometry from images has been a highly active research field and several sophisticated techniques have been proposed. The main drawback of many existing techniques is, however, that they have rather strict requirements regarding the input data: Often a shape prior or exact image silhouettes are required, and the quality of the reconstructions quickly degrades when the scene is not sufficiently textured. We present a new reconstruction method based on several new ideas combined with a set of existing, well established techniques with the goal to loosen these restrictions. We demonstrate with multiple real-world reconstruction results – including an evaluation using the Middlebury datasets [18] – that our algorithm produces reconstructions of comparable quality to the current state-of-the-art techniques while being more versatile and requiring no input other than a set of calibrated images.

Our method utilizes a geometry representation based on oriented particles (*c.f.* [11, 21]). Specifically, we use a set of surface elements in the form of planar disks in object space each of which is defined by a center point, a normal vector, and a radius to approximate the unknown surface of a scene seen in a set of calibrated images. The proposed reconstruction algorithm consists of two alternating phases. In the first phase it computes seed disks by purely image-based homography matching (*i.e.* without the integration of camera calibration information) in an automatically selected pair of images. The seed disks are then corrected by a plane fitting algorithm that is able to take an arbitrary number of input images into account, this time utilizing the camera calibration. The second phase expands the information from the already recovered surface region employing a greedy growing strategy. New disks are spawned at the boundary of the known region until it cannot be smoothly expanded any further. The algorithm stops when all visible parts of the scene are covered by disks, resulting in a dense set of point samples with accurate normal information. For applications like rendering or new view synthesis, the disks are already sufficient. For cases where closed, manifold triangle meshes are required we can apply off-the-shelf triangulation techniques. Due to the point and normal information, the publicly available tools of Kazhdan [9, 10] are especially well suited in our case. We can hence concentrate on generating point and normal samples for the visible parts of the scene only and let the mesh generation tools care for a smooth and consistent interpolation for the invisible parts.

The disk-based surface representation leverages several important advantages, which in this combination cannot be found in any existing reconstruction method:

- The sampling density in object space and size of the disks' footprint can be varied arbitrarily and independently which cannot easily be achieved for classical

surface representations like polygon meshes or voxel grids. We vary both parameters automatically driven by the texturedness and resolution of the input images.

- Our algorithm does not need initialization information of any form. All geometry, topology and visibility information is completely derived from the calibrated but otherwise unmodified input images. This in particular means that no visual hull has to be computed. However, if image silhouettes are available, this additional information can easily be taken into account.
- We exploit the topological flexibility of point-sampled surfaces. This topology-free surface representation allows for a simple incremental procedure: All disks are fitted with visibility information derived from the already recovered surface. In case new information invalidates existing disks, these disks are erased and the affected part of the surface is treated as not yet recovered.

1.1. Comparison to Previous Work

In what follows we discuss several existing reconstruction techniques, many of which have been shown to produce results of extremely high quality, with an emphasis on the differences compared to our method. A more complete overview of related techniques can be found in the survey of Seitz *et al*. [18].

Existing reconstruction methods can, quite roughly, be classified into either surface oriented or volume oriented methods. Examples for surface-based methods are [6, 8, 24] that use polygon meshes to represent the surface of a scene and [16, 20] that use level-sets. The mesh-based approach has to deal with several difficulties: artifacts like self intersections or folded-over polygons either have to be prevented explicitly or have to be taken care of by artificial energy terms that are minimized simultaneously with the evolution of the mesh. Furthermore, the resolution of the polygon mesh has to be adjusted by tedious decimation, subdivision, and remeshing algorithms that again keep the mesh consistent. In contrast, our disk-based surface representation does not require any consistency constraints since it only keeps track of loose neighborhood relations instead of strict element connectivity. For the mesh-based methods the topology of the final mesh has to be either known a-priori ([6, 24]) or topology changes have to be detected and taken care of during the mesh evolution ([8]) which can be an error prone process. Level-set surface representations ([16, 20]) naturally support topology changes but controlling the correct topology is still a difficult issue. Due to the topology-free surface representation with planar disks we neither require initial knowledge of the topology nor need to take care of topology changes.

Many of the above methods require a good initial surface proxy obtained, *e.g*., from the visual hull. Since this can in general be quite far away from the true surface, most often extremely small image regions have to be used for correspondence computation (*c.f*. [6]). In the presence of image noise again artificial smoothness energies are required to prevent noisy surfaces. Furthermore, large distances between the initial and the true surface (*e.g*. in deep concavities) often lead to slow convergence of the deformation process. In our approach, the disks are always tangential to the true surface of the scene and are fitted in a perspectively correct way. We can hence allow for relatively large footprints in the images resulting in stable fitting. Smooth surfaces are achieved in a natural way by increasing the footprint of the disks. Since new disks are created close to the true surface, in most cases they have to be only slightly corrected by the fitting process. Hence our method is very efficient in terms of convergence speed.

The mesh-based reconstruction method of Furukawa and Ponce [4] uses affine matching of image regions between pairs of images to obtain approximate point positions in object space. Their approach also contains the idea of expanding the already recovered information. However, due to the affine matching, the expansion is limited to a grid of rectangular regions in image space. A similar approach is followed by Lhuillier and Quan in [13]. The information provided by sparse point matches in image pairs is expanded in image space to obtain a regular grid of quasi-dense correspondences. In our approach, we use perspectively correct matching of object space disks which, instead of being limited to a pair of images, gains robustness from taking an arbitrary number of images into account. Furthermore our technique performs a natural expansion of the recovered surface in object space. The method in [13] employs a level-set evolution approach to compute a surface based on the triangulated quasi-dense point correspondences, image silhouettes, and photo-consistency measures. In [4] the visual hull is deformed to fit the reconstructed points and in a second step the mesh is further deformed by a texture-driven force based on tangential surface patches. This is similar to our idea of fitting planar disks to the surface, but is strictly bound to the resolution of the mesh and generally suffers from the mesh consistency problems mentioned above.

Early work in the field of volumetric reconstruction include [19] and [12]. The main problem here is that for high resolutions of the voxel grid, the image footprints used for consistency determination become very small. This often results in noisy reconstructions in textureless regions or under the influence of images noise. To overcome this problem, more recent volumetric methods (*e.g*. [7, 22, 23]) extract a smooth surface from the noisy consistency volume using a global optimization technique like graph cuts. While this works very well for sufficiently textured objects, it generally does not solve the problem for objects without texture. For small voxel footprints large parts of the vol-

ume are tagged as consistent with the images such that even global optimization methods are often unable to extract the correct surface. Enlarging voxel footprints is difficult since it is unknown in general where exactly the true surface lies. In our method the sampling density in object space and especially the footprint size of the disks is not restricted by a volumetric grid: Due to the surface-aligned disks we are free to enlarge the footprints in textureless areas until a robust fitting is possible. Tran and Davis in [22] extend the work of Vogiatzis *et al*. [23] by adding silhouette constraints to avoid the surface shrinkage that volumetric graph cuts are affected with. This, however, again requires exact image segmentation and does not work in concave parts of the surface. In contrast, our method does, due to the localized disk fitting, not need to cope with global shrinkage issues and hence does not require any constraints based on, *e.g*., object silhouettes.

Among the work explicitly using planes for the reconstruction of a scene is [2] by Baker *et al*. They use a Levenberg-Marquardt approach to fit planes to a set of images. In [5] Habbecke and Kobbelt present a framework that exploits the specific properties of the plane fitting problem and hence is more efficient than a standard optimization approach. However, both systems are not fully automated but require manual plane initialization. Furthermore, the approach in [2] uses very few large planes with individual per pixel displacements and hence is more geared towards a classical stereo setup while we use many planes to densely sample the scene with arbitrarily distributed cameras. Rothganger *et al*. [17] use planes to approximate a scene with the goal of partial reconstruction and object recognition. Their system is based on affine matching and an affine camera model, while our method supports more realistic projective matching and projective cameras. Carceroni and Kutulakos [3] use planar elements as well to represent the surface of a scene. Their work is based on the space carving method [12] and places a planar element in each surface voxel. The correct plane parameters are computed by an algorithm that first exhaustively tests many different positions and orientations and then performs a non-linear optimization for a set of best candidates. Hence their method is again bound to a volumetric grid in terms of sampling density and element size, and furthermore is quite computationally expensive. Unfortunately, none of the above plane-based methods take part in the Middlebury evaluation [14], so a comparison to the current state-of-the-art is difficult.

2. Plane Fitting

Our reconstruction algorithm is based on the ability to fit planar disks to the surface of a scene. For this task we build on the method described in [5]. A brief summary of the key points of the plane fitting algorithm is given in the following for the sake of completeness.

The fitting of a scene plane $\mathbf{N}^T = (n_0, n_1, n_2, d) = (\mathbf{n}^T, d)$ is based on the plane-induced homography that, for two images I_1 and I_2 with projection matrices $\mathbf{P}_1 = (\mathbf{M}_1|\mathbf{m}_1)$ and $\mathbf{P}_2 = (\mathbf{M}_2|\mathbf{m}_2)$, is given by

$$\mathbf{H}_2(\mathbf{N}) = \left(d\mathbf{M}_2 - \mathbf{m}_2\mathbf{n}^T\right)\left(d\mathbf{M}_1 - \mathbf{m}_1\mathbf{n}^T\right)^{-1}. \quad (1)$$

It maps image points from the image I_1 to the plane $\mathbf{N}$ and then further to the image I_2. The key to simplified matching equations is the global coordinate transformation

$$\mathbf{B} = \left(\begin{array}{ccc|c} & \mathbf{M}_1^{-1} & & -\mathbf{M}_1^{-1}\mathbf{m}_1 \\ \hline 0 & 0 & 0 & 1 \end{array}\right) \in \mathbb{R}^{4\times 4} \quad (2)$$

that transforms the projection matrix $\mathbf{P}_1$ to $\mathbf{P}_1' = \mathbf{P}_1\mathbf{B} = (\mathbf{Id}_3|\mathbf{0})$. After the transformation the plane-induced homography (1) simplifies to

$$\mathbf{H}_2(\mathbf{N}') = \mathbf{H}_2(\mathbf{n}') = \mathbf{M}_2' - \mathbf{m}_2'\mathbf{n}'^T, \quad (3)$$

where $(\mathbf{M}_2'|\mathbf{m}_2')$ is the transformed projection matrix of I_2 and $\mathbf{N}'^T = (\mathbf{n}'^T, 1)$ is the transformed plane.

The method in [5] requires the specification of a reference image I_1, an arbitrary number of comparison images $I_2, \ldots, I_n$, an initial estimate of the plane parameters $\tilde{\mathbf{N}}$ and a set of pixels Ω in the reference image. It then computes the correct plane parameters $\mathbf{N}$ by iteratively minimizing the SSD objective function

$$E = \sum_{c=2}^{n} \sum_{\mathbf{p}\in\Omega} \left(I_1(\mathbf{p}) - I_c(\mathbf{H}(\mathbf{N})\mathbf{p})\right)^2. \quad (4)$$

Applied to our context of fitting disks, we find the required information as follows. For a given disk, we compute a reference and a set of comparison images using the procedure detailed in Section 3.4. The image region Ω is determined by projecting the disk into the reference image and rasterizing its footprint. How the actual size of the footprint is determined is discussed in Section 3.3. Finally, the initial plane parameters of a disk are computed in two different ways: In the case of a seed disk the plane parameters are initialized from a homography matched between a pair of images (*c.f*. Section 3.1), in the case of a disk that expands the already recovered surface the parameters are inherited from an existing disk (*c.f*. Section 3.2).

3. Surface Growing

The surface growing approach to multi-view stereo reconstruction we are proposing works by alternating two phases. These two phases, the generation of seed disks and the expansion of surface regions, will be discussed in the following two sections. After that we go into detail about how the sampling density in object space and the size of the disks' footprint is determined and how visibility information is computed.

3.1. Seed Disks

Seed disks are used to initiate the surface reconstruction process on not yet recovered parts of the scene. To find such a part we traverse the input images in a regular way and stop at the first image I_1 with a free pixel $\mathbf{p}$ (*i.e.* a pixel that is not covered when the currently recovered surface is projected into the image). More elaborate strategies to, *e.g.*, find the largest not yet recovered image region are possible but we found that the above method is completely sufficient. To generate a seed disk our algorithm then matches an unconstrained homography $\mathbf{H}$ with 8 degrees of freedom using the technique in [1]. This requires an image region Ω in I_1 and a second image I_2. As region Ω we use all pixels contained in a circle centered at the pixel $\mathbf{p}$. The radius of the circle is determined in the same way as the size of the disk footprints, *c.f.* Section 3.3. To be able to initialize the matching homography with the identity matrix, we choose the second image I_2 nearby to the first in terms of viewing direction and rotation orientation. Note that, although our algorithm does not require a segmentation of the input images in principle, the computation time necessary to reconstruct a desired object can be reduced greatly if a coarse separation of the object and the background is available. Then seed disks are constructed only for surface parts that the user is really interested in.

Once the 8 parameters of the homography have been computed, we need to generate the corresponding disk in object space by initializing the parameters $\mathbf{n}$ and d. We apply the transformation (2) with I_1 as reference to the projection matrices of I_1 and I_2 and then compute the transformed plane parameters $\mathbf{n}'$ such that the plane-induced homography (3) performs the same transformation on image points $\mathbf{x}$ as the matched homography $\mathbf{H}$:

$$\mathbf{y} := \mathbf{H}\mathbf{x} \stackrel{!}{=} (\mathbf{M} - \mathbf{m}\mathbf{n}'^T)\mathbf{x}, \tag{5}$$

with $(\mathbf{M}|\mathbf{m})$ being the transformed projection matrix of I_2. After explicitly writing out the de-homogenization

$$y_i = \frac{(\mathbf{M}_i^T - m_i\mathbf{n}'^T)\mathbf{x}}{(\mathbf{M}_3^T - m_3\mathbf{n}'^T)\mathbf{x}}, \quad i = 1, 2 \tag{6}$$

where $\mathbf{M}_i^T$ is the ith row of $\mathbf{M}$ and m_i the ith component of $\mathbf{m}$, we apply the often-used transformation

$$y_i \cdot (\mathbf{M}_3^T - m_3\mathbf{n}'^T)\mathbf{x} = (\mathbf{M}_i^T - m_i\mathbf{n}'^T)\mathbf{x}. \tag{7}$$

Now that the equation is linear in the elements of $\mathbf{n}'$ we construct a linear system for all correspondences $\mathbf{x} \leftrightarrow \mathbf{y}, \mathbf{x} \in \Omega$ and compute the least squares solution. After applying the inverse transformation $\mathbf{B}^{-1}$ to obtain the true plane parameters $(\mathbf{n}^T, d)$, the disk center is computed as the intersection of the plane with a ray from the camera center of I_1 through the pixel $\mathbf{p}$.

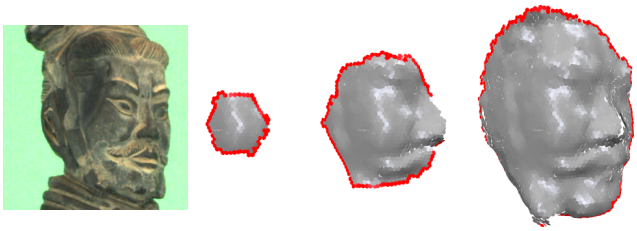

Figure 1. Surface growing on the head of a Chinese statue. The recovered region is expanded at active disks (red). A disk becomes inactive (grey) when its neighborhood is complete or it is not possible to smoothly expand the recovered region in this particular area. (The cutout of the image shown on the left has 256^2 pixels and that the size of the disks has been reduced for rendering.)

Clearly there is no guarantee that the homography matching does not get stuck in a local minimum and the seed disk hence does not lie on the scene surface. We employ two simple but effective mechanisms to detect such cases. Firstly, the Euclidean distance

$$||\mathbf{H}\mathbf{x} - (\mathbf{M} - \mathbf{m}\mathbf{n}'^T)\mathbf{x}|| \tag{8}$$

between points projected by the originally matched homography $\mathbf{H}$ and the homography induced by the recovered plane parameters has to be smaller than a threshold. This way homographies not representable as plane-induced homographies are easily detected. As second check we apply the projective plane fitting of Section 2 to the newly constructed seed disk with more than two images taken into account and compare the original to the fitted disk. If the disk moves or rotates by more than a threshold it is treated as outlier. As we will demonstrate in the result section, we can choose these thresholds once for all experiments and do not have to adapt them to each individual data set. In case object silhouettes are available, they can be used as additional check for outlying seed disks. If either of these checks fails the sampling procedure starts over with finding a new free image position.

3.2. Surface Expansion

The disks that are used to approximate the unknown surface of a scene can be classified as being either *active* or *inactive*, see Figure 1. Active disks usually lie on the boundary of the recovered surface region and are candidates for its further expansion. A disk becomes inactive either if its geodesic neighborhood on the surface is completely covered with disks or if it is not possible to smoothly expand the recovered region at this particular position. The recovered region is expanded as long as active disks are present. When all disks have become inactive a new region is initialized with a seed disk as single active disk (*c.f.* Section 3.1).

The main idea of the surface growing approach is to add new disks to complete the neighborhood of active disks, thus expanding the conquered region in tangential direction.

"Gaps" in the neighborhood of an active (parent) disk are detected using a simple criterion. The centers of all neighboring disks are projected into the supporting plane of the disk in question. Then the angles between the vectors from the disk center to the projections are computed and sorted. Each gap larger than 60 degrees is filled by adding a new disk at the corresponding position in the tangent plane of the parent disk. The initial plane parameters are inherited from the parent. After correcting the plane parameters with the fitting algorithm of Section 2 the final disk position is again computed as a ray-plane intersection with a ray from the reference camera center through the old disk center.

This strategy guarantees a dense coverage of the whole surface when the process finishes (notice that the threshold of 60 degrees corresponds to a dense circle packing). Moreover it satisfies the requirement of the plane matching algorithm for an initial estimate of the plane parameters: By initializing the normal of a newly added disk with the normal of its already fitted parent and by moving its center to a new position in the parent's tangent plane, the parameters are sufficiently close to allow for stable convergence of the plane matching algorithm.

There are, however, two cases in which the above approach is not able to smoothly expand the recovered region. Firstly, the initialization with the plane parameters of the parent disk might not be correct at corners or sharp edges, for example. Secondly, the visibility information might not be correct. As will be discussed in Section 3.4, the set of images used for the fitting of a disk is determined by the recovered surface information only. In early stages of the algorithm occluding geometry simply might not have been recovered yet, leading to a wrong choice of comparison images and consequently to badly fitted disks. Both cases are handled by limiting the movement and rotation a disk is allowed to undergo during fitting: If the assumption of a smooth surface is violated the corresponding disk is immediately discarded. If object silhouettes are available they can again be used as additional means to detect outlying disks.

3.3. Sampling Density and Footprint Size

The image region used to fit a particular disk is its footprint in its reference image. The position of the footprint is determined by the position of the disk in object space which, in turn, is determined by the expansion strategy. The only remaining degree of freedom is the size of the footprint controlled by the radius of the disk. We adjust the size of the footprint based on the following idea: For textured parts of an object the fitting process is generally stable. Hence the size of the disks can be relatively small to allow for geometric detail. However, in textureless parts the fitting needs to be stabilized by using larger image regions. Accordingly, the size of a disk is determined by the *intensity variance* of the footprint in the reference image. That is, starting from a minimal footprint size $f_{\min}$ the radius of a disk is increased until the variance of the footprint reaches a target threshold σ^2 (see Section 4 for details about parameter settings). Choosing the target variance above the image noise level allows us to effectively adapt the disks' size to the quality of the underlying images. This disk sizing strategy also integrates nicely with the fact that, under the assumption of Lambertian illumination, an image region with very similar intensities corresponds to a quasi-planar part of the scene that can extremely well be approximated by a large planar disk. Figure 2 shows a visualization of the distribution of disk sizes for the completely textureless Dino model from the Middlebury multi-view stereo evaluation [18]. Red color depicts surface areas with large disks due to little texture while green colored surface parts are covered with small disks.

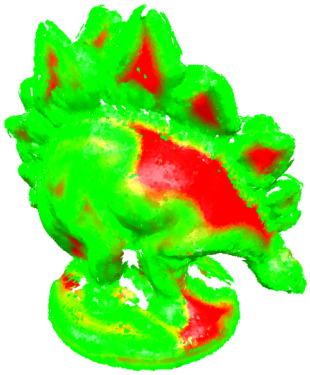

Figure 2. Variance-driven determination of disk size. Red colored parts of the surface are approximated by large disks due to little image texture, while green color depicts small disks.

The determination of the sampling density in space, *i.e.* the distribution of disk centers on the scene's surface, is part of the surface expansion. Specifically, the sampling density is controlled by the distance a new disk is moved away from the center of its parent disk. The most natural way of adjusting the sampling density would be based on the curvature of the surface. Unfortunately this is not feasible since the surface is unknown. We hence decided for a flexible solution that does not require knowledge about the absolute scene scale: New disks are moved by a user-specified fraction of the parent disk's minimal radius (corresponding to the footprint size $f_{\min}$). This way the sampling density is automatically adjusted to the resolution of the images. For images with large camera-to-object distance the disks are enlarged to project to a footprint of sufficiently many pixels. As a result, the sampling in object space becomes coarser which is a meaningful reaction to this situation. On the other hand, for close-up images of an object this approach increases the sampling density allowing for the reconstruction of fine geometric detail. Note that the sampling density depends on the minimal but not on the actual, variance adjusted disk radius. We do not use the latter solution since the variance in many cases does not change sufficiently smooth as it would be required for a regular sampling.

3.4. Visibility Determination

To perform reliable and correct disk fitting we need to be able to determine a set of images in which a particular disk

is visible. As stated earlier, the visibility information is derived from the recovered geometry only and hence no shape initialization is necessary. We apply an approach similar to the z-buffers of graphics systems: For each pixel of each input image we store the ID of a disk and a depth value. Initially all IDs are set to an invalid value and all depth values are set to infinity.

Whenever a new disk is created (either as seed disk or during the surface expansion) its footprint is determined in *all* images. In case the depth of the new disk is less than the depth stored in a particular image, the disk is classified as being visible in that image and the existing depth and ID are overwritten with the new information. From all images seeing a disk we choose the one with the most perpendicular view as reference. The comparison images are chosen from the remaining images that see the disk with a viewing angle of less than 45 degrees. To prevent a set of only small-baseline comparison cameras, images with a viewing angle of 30 degrees are preferred most.

The main point of our depth buffer approach is its incremental nature. Since the visibility of a disk is derived from the recovered surface only, it might not be completely correct (*i.e.*, one or more images might wrongly be classified as seeing a particular disk). Such a case is easily detected once the occluding surface part is conquered: When a new disk has a depth less than the stored depth value, it is checked if the disk corresponding to the stored ID has used the current image for fitting. If this is the case, this disk is erased. Furthermore, its neighbors are classified as active again such that the surface expansion algorithm is able to fill the hole in the surface eventually – this time with improved visibility information.

During the expansion of the recovered surface the algorithm requires access to the geodesic neighborhood of the active disks (a) to determine the uncovered surface areas, (b) to check the feasibility of a new disk and (c) to reactivate neighbors of erased disks. This information is also derived from the IDs stored in the reference image of the disk in question by looking up all IDs in an area with twice the diameter of the disk's footprint.

4. Implementation Details

The reconstruction system we are proposing has several parameters that need to be set. It turns out, however, that all of them have a natural interpretation and an easy to understand influence on the behavior of the algorithm. Hence no tedious parameter tuning is necessary.

To let the algorithm automatically adjust the size of the disks the user needs to specify a range $f_{\min}$, $f_{\max}$ defining the allowed footprint sizes in the reference image, measured in pixels. Furthermore, a target intensity variance σ^2 needs to be specified that guides the determination of the actual footprint size. Clearly, the size of the footprints on

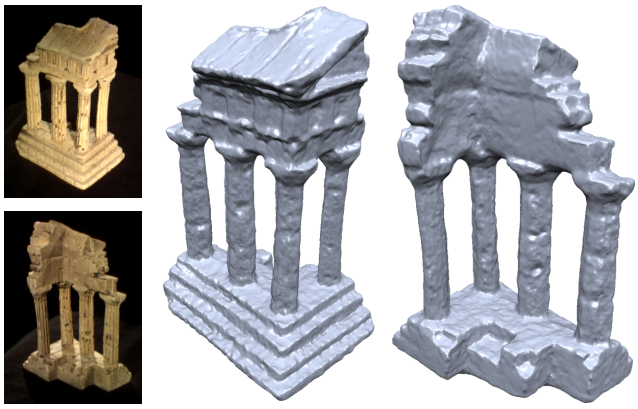

Figure 3. Reconstruction from the Middlebury Temple dataset. Although we fit planar disks to the surface of the temple, we are still able to reconstruct a lot of its structural detail. Middlebury evaluation result: 90% of our reconstruction is contained within 0.66mm of the ground truth, and 98% is contained within 1.25mm.

the one hand influences the smoothness of the result (larger footprints induce more smoothness) and on the other hand influences the running time (larger footprints mean longer fitting time). In our experiments we set $f_{\min} = 100$ (with a single exception for the Monkey model, see next section), $f_{\max} = 2000$ and $\sigma^2 = 36$ (for image intensities in $[0, 255]$) which lies sufficiently above the image noise level.

The sampling density δ is, as mentioned above, defined as a fraction of the disk radius corresponding to the minimal footprint size $f_{\min}$. The sampling density directly influences the level of detail the algorithm is able to reconstruct. It also influences the running time since denser sampling requires more computational effort. In the experiments we set δ to 20%.

The next two parameters regard the number of comparison images used during the disk fitting. Again we specify a range $c_{\min}$, $c_{\max}$. When enough images are available then the $c_{\max}$ best are used for fitting (*c.f.* Section 3.4). In case less then $c_{\min}$ cameras see a disk, the surface expansion is stopped at that particular position. These two parameters also directly influence the stability (more comparison images stabilize the fitting process) and the running time (more images require more time). In all the experiments we set the parameters to $c_{\min} = 4$ and $c_{\max} = 10$.

Finally, the algorithm requires the specification of a threshold regarding the admissible movement of child disks during the surface expansion. Given a parent disk, we compute the mean plane parameters over all disks in its two-ring neighborhood. Then the normal of a child disk is allowed to maximally deviate by an angle t_α from the mean normal and its center has to lie within a distance t_d to the mean plane. In our experiments an angle t_α of 15 degrees and a distance t_d of $1/4$ of the parent disk's radius have proven to work well. The same values are used to check the correctness of a seed disk.

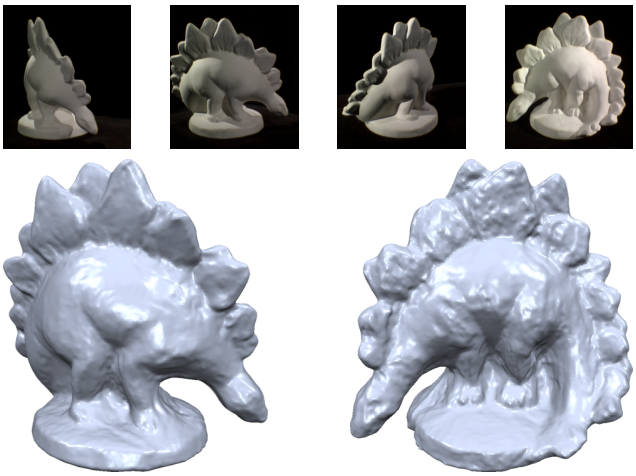

Figure 4. Reconstruction from the Middlebury Dino dataset. Due to the fitting of large disks our algorithm is able to compute a smooth reconstruction even in the absence of object texture. Middlebury evaluation result: 90% of our reconstruction is contained within only 0.43mm of the ground truth, and 99.7% is contained within 1.25mm.

Model	Images	Resolution	Regions	Disks	Time
Dino	363	640×480	91	105k	130 min
Temple	312	640×480	594	111k	91 min
Leo	86	1024×768	8	94k	29 min
Warrior	42	1024×768	51	264k	75 min
Monkey	46	720×576	117	51k	35 min

Table 1. Results of our surface-growing algorithm applied to several image sequences.

5. Results

In the following, several reconstructions from different image sets are shown. All experiments have been carried out on an Athlon64 system running at 2.2GHz with 4GB of RAM. Table 1 lists the number and resolution of the input images, the number of reconstructed surface regions, the total number of reconstructed disks, and the total computation time.

The first two reconstructions result from the Middlebury [18] Temple and Dino images. In both cases the full datasets with well over 300 images each were used. Figures 3 and 4 show the triangle meshes generated from the reconstructed points and normals using the method of [10].

The main difficulty of the Dino model is its textureless surface that results in almost constant intensity values in large image areas. Hence, many of the existing methods using small image regions for correspondence computation have trouble reconstructing the Dino (*c.f*. the Middlebury evaluation results [14]). Due to the adaptively sized disks, our approach computes a faithful reconstruction that ranks among the best methods at the time of writing.

The Temple model shows that our method is able to reconstruct objects of non-trivial topology without any a-priori knowledge about the shape and without the need to compute a visual hull that prescribes the topology. Because the resolution of the input images is only 640×480 pixels, fine geometric detail, *e.g*. on the columns, appears only a few pixels wide and hence cannot be captured correctly by the planar disks. On the other hand our result for the Temple can still compete with the state of the art (again *c.f*. [14]).

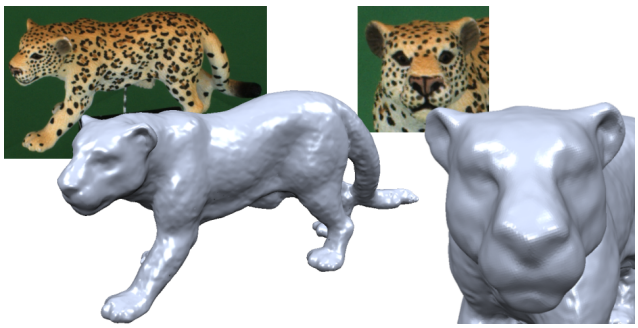

Figure 5. High-quality reconstruction of a toy leopard. Triangle mesh generated from the 94k reconstructed disks (left) and close-up of the leopard's head (right).

The reconstruction of the Leopard and Warrior shown in Figures 5 and 6 are examples of the high quality our algorithm is able to achieve. The images have been acquired with a turn-table setup and calibrated using a standard structure from motion approach. Particularly in the case of the Warrior our algorithm has been able to reproduce the geometric detail quite well due to the higher image resolution compared to the Middlebury data.

The last result in Figure 7 demonstrates the behavior of our algorithm for low-quality input data. Several images are affected by motion blur and limited depth of field and hence the calibration obtained via structure from motion is far from optimal. To improve the robustness of our algorithm, we increased the minimal footprint size to $f_{\min} = 300$. Moreover, since all viewpoints for this image sequence lie within a narrow cone, this example shows that our algorithm is even able to reconstruct relief-like structures where a shape initialization from object silhouettes is not possible.

6. Conclusion and Discussion

We have presented a new reconstruction method that is based on a robust plane fitting algorithm and exploits the flexibility of point-based surface representations. The main advantages of the presented approach are the ability to cope with textureless objects due to the surface-aligned, automatically sized disks and the fact that it does not depend on any a-priori shape, topology or visibility information. It is furthermore fully automatic and has been shown to be able to do high quality reconstructions in diverse image capture setups and competitive runtime.

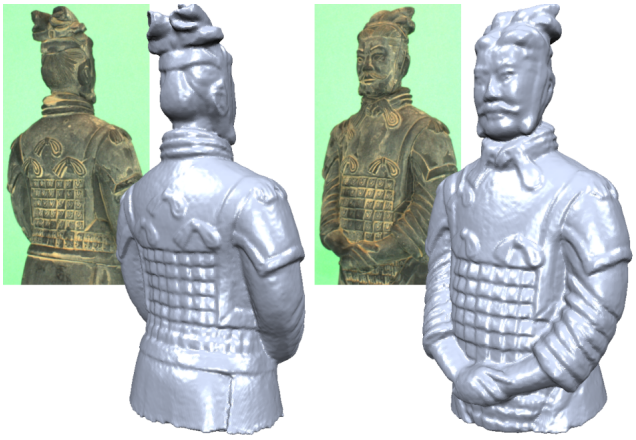

Figure 6. Chinese warrior statue reconstructed from 43 images. Notice that in this case more geometric detail is recovered than for the Temple due to the higher image resolution.

Due to the specific form of seed disk generation by homography matching, the current implementation is targeted at small baseline image sequences to work well. Future work includes the integration of a wide-baseline matching scheme (like, *e.g.*, [15] by Mikolajczyk and Schmid). In theory, due to the more efficient convergence properties of our surface expansion strategy, our method should be much faster than existing mesh-based approaches. In practice we lose some of the advantages due to the incremental visibility determination that requires the deletion and re-computation of already recovered surface regions. However, since the disks are fitted independently, it is possible to parallelize important parts of the algorithm in a shared memory environment to exploit modern multi-core CPUs. The examination of this issue is an interesting topic for future work.

References

[1] S. Baker and I. Matthews. Lucas-kanade 20 years on: A unifying framework. *IJCV*, 56(3):221–255, 2004.

[2] S. Baker, R. Szeliski, and P. Anandan. A layered approach to stereo reconstruction. In *CVPR*, pages 434–441, 1998.

[3] R. L. Carceroni and K. N. Kutulakos. Multi-view scene capture by surfel sampling. *IJCV*, 49(2):175–214, 2002.

[4] Y. Furukawa and J. Ponce. High-fidelity image-based modeling. Technical Report 2006-02, UIUC, May 2006.

[5] M. Habbecke and L. Kobbelt. Iterative multi-view plane fitting. In *Proc. of VMV*, pages 73–80, 2006.

[6] C. Hernández and F. Schmitt. Silhouette and stereo fusion for 3d object modeling. *Computer Vision and Image Understanding*, 96(3):367–392, 2004.

[7] A. Hornung and L. Kobbelt. Hierarchical volumetric multi-view stereo reconstruction of manifold surfaces based on dual graph embedding. In *Proc. of CVPR*, volume 1, pages 503–510, 2006.

[8] J. Isidoro and S. Sclaroff. Stochastic refinement of the visual hull to satisfy photometric and silhouette consistency constraints. In *Proc. of ICCV*, pages 1335–1342, 2003.

Figure 7. Reconstruction from low-quality images. Left: First and last image of the input sequence. Right: triangle mesh.

[9] M. Kazhdan. Reconstruction of solid models from oriented point sets. In *Proc. of SGP*, pages 73–82, 2005.

[10] M. Kazhdan, M. Bolitho, and H. Hoppe. Poisson surface reconstruction. In *Proc. of SGP*, pages 61–70, 2006.

[11] L. Kobbelt and M. Botsch. A survey of point-based techniques in computer graphics. *Computers & Graphics*, 28(6):801–814, 2004.

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

[13] M. Lhuillier and L. Quan. A quasi-dense approach to surface reconstruction from uncalibrated images. *IEEE Trans. PAMI*, 27(3):418–433, 2005.

[14] Middlebury multi-view stereo evaluation results. http://vision.middlebury.edu/mview/, Dec. 2006.

[15] K. Mikolajczyk and C. Schmid. Scale and affine invariant interest point detectors. *IJCV*, 60(1):63–86, 2004.

[16] J.-P. Pons, R. Keriven, and O. Faugeras. Modelling dynamic scenes by registering multi-view image sequences. In *Proc. of CVPR*, volume 2, pages 822–827, 2005.

[17] F. Rothganger, S. Lazebnik, C. Schmid, and J. Ponce. 3d object modeling and recognition using local affine-invariant image descriptors and multi-view spatial constraints. *IJCV*, 66(3):231–259, 2006.

[18] S. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *Proc. of CVPR*, volume 1, pages 519–526, 2006.

[19] S. M. Seitz and C. R. Dyer. Photorealistic scene reconstruction by voxel coloring. In *Proc of. CVPR*, pages 1067–1073, 1997.

[20] S. Soatto, A. J. Yezzi, and H. Jin. Tales of shape and radiance in multi-view stereo. In *ICCV*, pages 974–981, 2003.

[21] R. Szeliski and D. Tonnesen. Surface modeling with oriented particle systems. In *SIGGRAPH*, pages 185–194, 1992.

[22] S. Tran and L. Davis. 3d surface reconstruction using graph cuts with surface constraints. In *Proc. of ECCV*, volume 2, pages 219–231, 2006.

[23] G. Vogiatzis, P. Torr, and R. Cipolla. Multi-view stereo via volumetric graph-cuts. In *CVPR*, pages 391–398, 2005.

[24] L. Zhang and S. M. Seitz. Image-based multiresolution shape recovery by surface deformation. In *Proc. of SPIE*, pages 51–61, 2001.