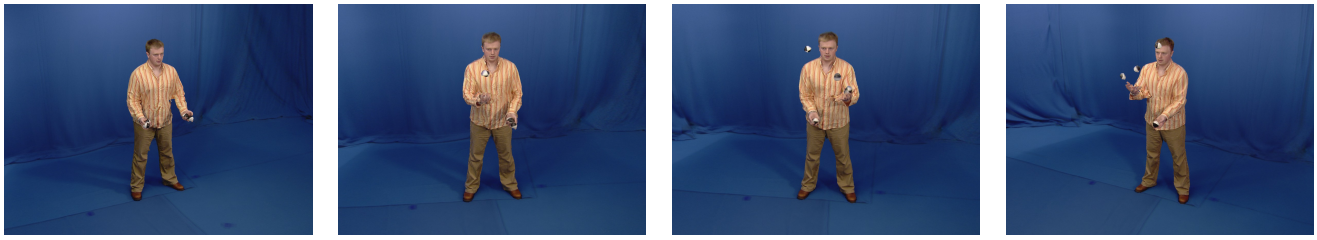
SAFE HULLS

Gregor Miller and Adrian Hilton

Centre for Vision, Speech and Signal Processing,
University of Surrey, UK
{Gregor.Miller, A.Hilton}@surrey.ac.uk



Novel views synthesised using safe hulls and only rendering definite foreground areas

Abstract

The visual hull is widely used as a proxy for novel view synthesis in computer vision. This paper introduces the *safe hull*, the first visual hull reconstruction technique to produce a surface containing only foreground parts. A theoretical basis underlies this novel approach which, unlike any previous work, can also identify phantom volumes attached to real objects. Using an image-based method, the visual hull is constructed with respect to each real view and used to identify *safe zones* in the original silhouettes. The safe zones define volumes known to only contain surface corresponding to a real object. The zones are used in a second reconstruction step to produce a surface without phantom volumes. Results demonstrate the effectiveness of this method for improving surface shape and scene realism, and its advantages over heuristic techniques.

Keywords: Surface reconstruction, free-viewpoint video, visual hull

1 Introduction

This paper presents a novel contribution to the visual hull literature, a method which overcomes inaccuracies of the visual hull by reducing the number of phantom volumes. Consequently the visual quality of novel views rendered using the safe hull as a proxy surface is improved by eliminating visual artefacts. This is achieved without increasing the number of cameras or using heuristic methods.

The visual hull is defined as the maximum volume consistent with the observed views' silhouettes and is widely used in the graphics and vision communities for a range of applications. Many free-viewpoint video techniques rely on it either as an initialisation for further improvement or as the final model[12, 14]. It is also used in applications as diverse as crowd surveillance, 3D modelling of objects and medical imaging.

However, the surface produced from a visual hull reconstruction comes with two major problems. Silhouettes are unable to represent concavities of objects and so neither can the visual hull (e.g. it could not reconstruct the inside surface of a coffee mug). Research has concentrated particularly on this problem, especially in free-viewpoint video, where colour matching and model fitting are used to recover concavity shape.

The second problem, which this paper addresses, is phantom volumes: surfaces produced in the reconstruction that do not represent objects in the scene. They are a product of multiple or non-convex objects, illustrated in Figure 1(a), and are consistent with the original silhouettes. The perceived realism in a synthesised view can be negatively affected by the odd shapes they form, as shown in Figure 5(c). These often have to be removed by hand, or through heuristic methods which may incorrectly remove surface belonging to a foreground object.

Previous approaches have attempted to remove phantom volumes by adding more cameras[2], however this is not a guaranteed solution. A common issue for reconstruction of people is extra limbs, such as a 'tail' (Figure 4(c)), which appear as connected phantom volumes (surfaces which do not represent a foreground object but are connected to a surface which does). Removal of this volume requires a camera positioned to look directly between the legs at all times, which is impossible for a dynamic subject.

Although additional cameras can reduce the size and number of phantom volumes, studios generally do not have many cameras due to time and financial constraints, therefore research into free-viewpoint video is often targeted toward a minimal number of well-placed cameras. This highlights the importance of a solution to phantom volumes, since it increases the quality of the results without requiring additional cameras.

The research presented in this paper illustrates how to identify volumes in three dimensions which are part of the foreground

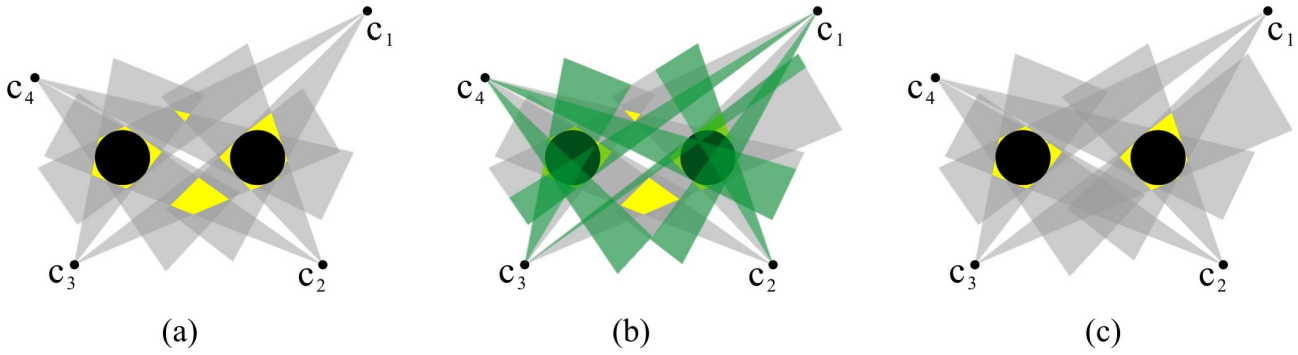


Figure 1: Phantom volumes are caused by multiple objects in the scene, shown in (a). The black circles represent scene objects, the gray areas represent silhouette cones from the cameras and the yellow shapes represent the result of visual hull reconstruction. The green areas in (b) represent the safe zones defined by the cameras, and the safe hull reconstruction is shown in (c), with phantom volumes removed.

i.e. safe zones which definitely do not contain phantom volumes, and to reclassify the remaining occupied space as unsafe. The unsafe space can be removed completely, therefore guaranteeing removal of all phantom volumes (including those connected to the subject), it can be processed further, for example using colour constraints to identify foreground, or it can be rendered differently.

Previous work on visual hull and phantom volume removal is described in the next section. The safe hulls algorithm is introduced in Section 3, followed by results in Section 4 which demonstrate the effectiveness of the solution. Finally, the technique is discussed in Section 5.

2 Previous Work

Various algorithms for constructing the visual hull have been presented since it was introduced by Laurentini[7], the most common of which is the volumetric approach. A volumetric grid where each element is tested against $\mathcal{S}$ is a simple and robust way to generate an approximate surface[11]. Franco et al. [5] presented a technique to recover the exact representation of the visual hull corresponding to a polyhedral approximation of the silhouette contour. Brand et al. [3] describe a method of applying differential geometry to obtain a close estimate to the exact visual hull surface from silhouette contours. Matusik et al. presented image-based visual hulls[8], an approximate view-dependent visual hull to efficiently render novel views without explicit reconstruction. Exact view-dependent visual hull (VDVH)[9] evaluates the surface in the image domain to produce an accurate depth map representing the visual hull, based on the original contours of $\mathcal{S}$. All of these visual hull approaches construct the entire surface without identifying areas of definite foreground or those with possible phantom volumes.

There are many free-viewpoint video techniques that employ visual hull for novel view synthesis. Vedula et al. introduced scene flow, based on volumetric visual and photo hull[12]. Wuermlin et al. used a variant of image-based visual hulls for free-viewpoint video using point samples instead of mesh with texture[14]. Miller et al. construct an image plus depth representation for every original view using visual hull as an initialisation for a global stereo optimisation[10]. Since these techniques all use visual hull as a proxy surface, they are all vulnerable to phantom volumes.

Other approaches to free-viewpoint video do not use visual hull and so do not suffer from phantom volumes, but are more constrained. Carranza et al. used a model-based approach with silhouette initialisation[4], which requires prior knowledge of the captured subject and so reconstruction of an arbitrary scene (such as the juggling example) is not possible. The novel view system presented by Zitnick et al.[16] simultaneously estimates image segmentation and stereo correspondence to produce video quality virtual views, but is restricted to a narrow baseline camera setup (8 cameras over 30°). Goesele et al.[6] present a multi-view stereo reconstruction system that produces high quality surfaces with a large number of narrow baseline views, which is prohibitive for dynamic scenes. Adopting the visual hull as a basis allows for arbitrary dynamic scenes to be reconstructed from a relatively small number of widely spaced cameras.

The problem of removing phantom volumes from a visual hull reconstruction has largely been ignored in previous research. Adding more cameras can reduce the problem but artefacts still occur. The visual hull has been applied to crowd surveillance [15], with temporal filtering and heuristic methods based on size used to remove phantom volumes. These approaches can be unreliable, for example in juggling (Figure 6) the balls could be removed by a threshold on size, and temporal filtering would not work on a connected phantom volume (such as the tail in Figure 4).

The following section presents a theoretical basis for reliably extracting definite foreground surface from a visual hull reconstruction. Unlike previous approaches, this guarantees removal of all phantom volumes, including those connected to real volumes.

3 Safe Hulls

This section presents a novel fully automatic method for reliably detecting real volumes in a visual hull reconstruction, using a theoretical basis to construct a *safe hull*. The construction of the safe hull is accomplished via a two-pass algorithm, where the full visual hull is constructed and analysed to supply information about the original images. The information is used to define *safe zones* in the original images: regions known not to back-project from the camera centre to phantom volumes. A second construction takes place, similar to visual hull but incorporating the safe zones so that all phantom volumes are excluded. The final result is a scene partitioned into definite foreground, definite background and a middle ground which may contain both.

The theory of visual hull is briefly explained and a short overview presented of the chosen algorithm for surface construction, the exact view-dependent visual hull[9]. This is followed by the theory for partitioning the surface into definite foreground and volumes containing phantoms. The main algorithm for safe hull construction is then described.

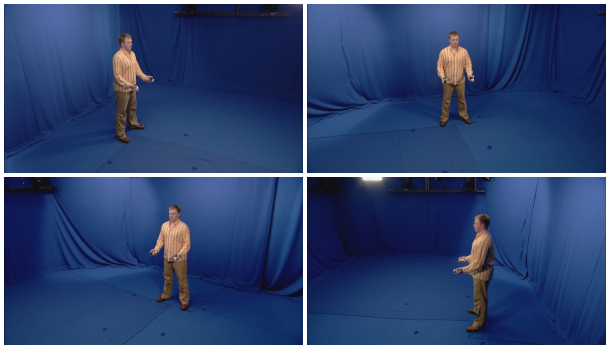
3.1 Visual Hull

The *visual hull* is widely used as a proxy for novel view synthesis in free-viewpoint video. The first step of visual hull construction is foreground extraction from the set of captured images $\mathcal{I} = \{\mathcal{I}_n : n = 1, \ldots, N\}$ to produce the set of silhouette images $\mathcal{S} = \{\mathcal{S}_n : n = 1, \ldots, N\}$, where N is the number of calibrated views. The *silhouette cone* for the n^{th} view is produced by casting rays from the camera centre $\boldsymbol{c}_n$ through the occupied pixels in the silhouette S_n. The visual hull is the three dimensional shape formed by the intersection of all views' silhouette cones[7].

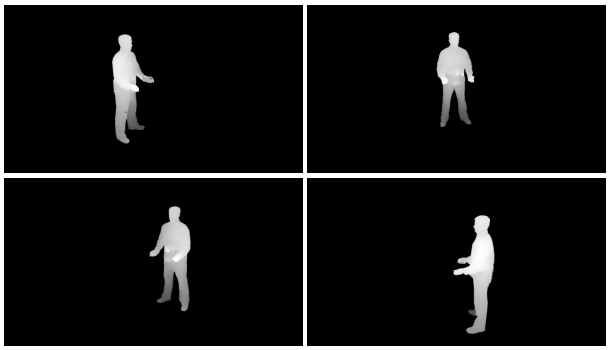
Phantom volumes are the result of a visual hull reconstruction of multiple objects or a non-convex object (e.g. a person) in a scene, as shown in Figure 1(a). They are volumes which are consistent with all silhouettes but do not represent a scene object.

3.2 View-Dependent Visual Hull

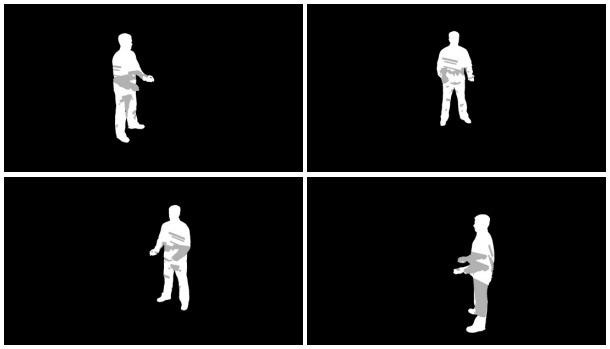
The VDVH is an image-based approach which constructs the exact view-dependent visual hull with respect to a specific



(a) original images



(b) VDVH depth maps



(c) safe zones (white)

Figure 2: Stages of the safe hull construction process. Starting with the original images in (a), the VDVH is constructed and a depth image for each view produced, shown in (b). The number of intervals in a depth image pixel determines whether it belongs to a safe zone (white) or an unsafe zone(grey), illustrated in (c). The safe hull is constructed using these zones to produce the surface shown in Figure 5(b). The surface on the left is the original visual hull. Notice the safe hull has removed the growth under the arm, the majority of excess surface in front of the body, and improved the smoothness of the surface of the legs.

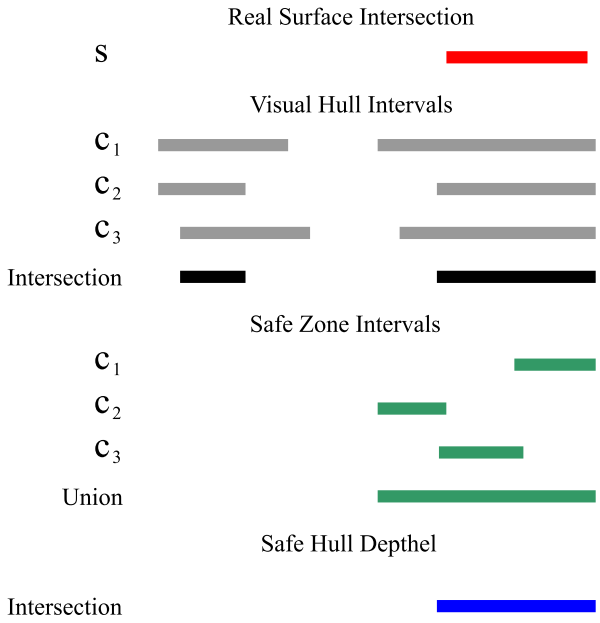


Figure 3: The grey lines represent the intervals from three cameras projected onto a virtual ray, and the visual hull represented below them as their intersection. The green lines represent the safe zone intervals from these cameras and below them their union to define which volumes are definitely not phantom. The blue line shows the result of an intersection of the visual hull depthel with the safe zone depthel: the safe hull depthel. Notice that the object to the left has been removed, and may have been a phantom.

viewpoint. The algorithm for VDVH has been extended to produce the entire exact visual hull for a specific viewpoint. The result is a multi-layer depth image composed of depthels:

Definition 1 *A* depthel *is a single pixel in a multi-layer depth image representing every interval where the ray through that pixel from the camera centre is inside the visual hull.*

Depthels have a dynamic number of intervals, produced during visual hull construction to represent the entire surface. Using a method which constructs a global representation to compute the multi-layer depth images would require us to find the intersection of rays from each camera with the surface, which involves multiple resampling steps. The VDVH was chosen to form the basis of this technique because it efficiently produces the exact visual hull with no additional quantisation, and the multi-layer depth image it produces directly represents the intervals of the visual hull with respect to a particular view. This is required for construction of the safe hull.

3.3 Foreground Detection

The algorithm relies upon the ability to detect regions in an image which definitely do not contribute to a phantom volume and are therefore part of the foreground. The following results demonstrate how this can be accomplished. This first result is the basis of the method:

Theorem 1 *Given the set of pixels* $Q = \{q : q \in \mathcal{I}_n\}$ *which lie on the projection of a phantom volume in image* $\mathcal{I}_n$ *and the set of multi-layer depth images* $\mathcal{V} = \{V_n : n = 1, \ldots, N\}$ *produced using VDVH, every depthel in the set* $D_n = \{d(q) : d(q) \in \mathcal{V}_n, q \in Q\}$ *has more than one interval.*

Proof Define the set of pixels $P = \{p : p \in \mathcal{I}_n, \mathcal{S}_n(p)$ is occupied$\}$, and the set of rays $R = \{\boldsymbol{r}(p) : p \in P\}$ through P from the camera centre $\boldsymbol{c}_n$, each ray $\boldsymbol{r} \in R$ has at least one interval which lies inside the real object described by $\mathcal{S}_n$. Phantom volumes are consistent with all views' silhouettes (by definition of visual hull), therefore they exist inside the silhouette cones for real objects. Now define the set of pixels $Q = \{q : q \in \mathcal{I}_n, q \subseteq p,$ q corresponds to a phantom volume$\}$. Since each depthel $d \in D_n$ already has at least one interval for the real object, the phantom volume intersected by $\boldsymbol{r}(q) \in R, q \in Q$ introduces at least one more. Therefore d must have a minimum of two intervals.

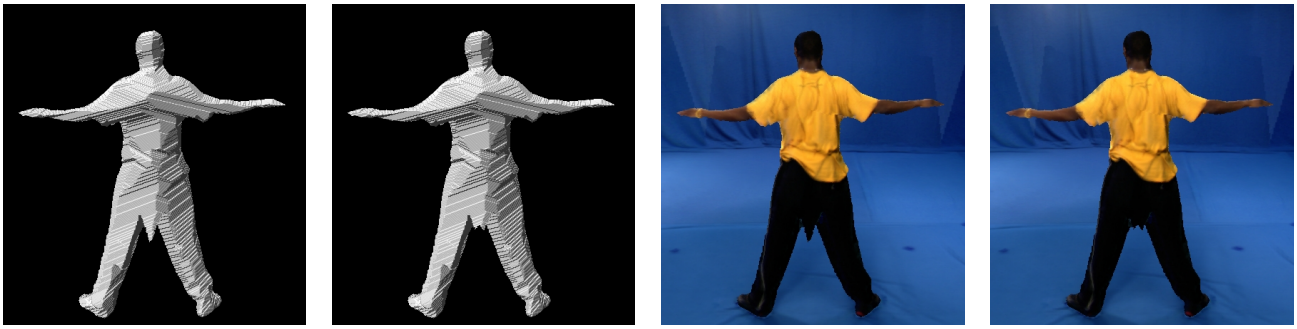
Theorem 1 does not work in the reverse: regions of the image with multiple intervals are not necessarily phantom volumes, they could for example be an arm occluding the body. However, we can use it to deduce the following result:

Corollary 1 *Depthels which have only one interval represent a real volume and do not contain phantom volumes.*

Proof Theorem 1 states that depthels containing phantom volumes must have more than one interval, so it follows that depthels with one interval describe a real volume.

This allows us to partition each image in $\mathcal{I}$ into three regions: we can mark regions of $\mathcal{I}_n$ with more than one interval in $\mathcal{V}_n$ as 'unsafe zones', regions with only one interval as 'safe zones', and the rest remains as background. This leads to the important result which allows us to remove phantom volumes: *any point in the visual hull whose projection lies inside a safe zone of a single image does not contain a phantom volume*.

This is important because it shows that for any point in the volume, only *one* view with this point's projection in the safe zone is required for it to be considered part of a real volume,



(a) Visual hull surface (b) Safe hull surface (c) Visual hull rendered (d) Safe hull rendered

Figure 4: This example illustrates the most common situation for phantom volumes to appear when capturing humans. This is a connected phantom volume, and often appears between the legs or under the shoulders. The quality of a synthesised view (c) is dramatically decreased when a subject spontaneously grows a 'tail', and once removed the image quality is improved (d).

as shown in Figure 1. There is no need for all views to agree, which is exactly the opposite concept to the visual hull.

Since all views have their own safe zones, the union of visual hull volumes corresponding to each forms the safe hull and completely eliminates phantom volumes. The volumes in the visual hull excluded from the safe hull contain all phantom volumes and parts of the object which did not project to safe zones. These can be examined manually for phantom volume removal, or further processed, for example using colour consistency as a constraint.

The algorithm for safe hull construction is as follows:

1 Construct the visual hull

2 Find safe zones in the original images

   a Find intersections of rays from occupied pixels in original views with the visual hull surface

   b Partition occupied pixels in the silhouette into safe and unsafe zones (mark pixels with one interval as safe)

3 Construct safe hull

   For a given point in the visual hull volume, accept it if it lies in a safe zone in at least one camera. Otherwise reject it.

3.4 Safe Zones

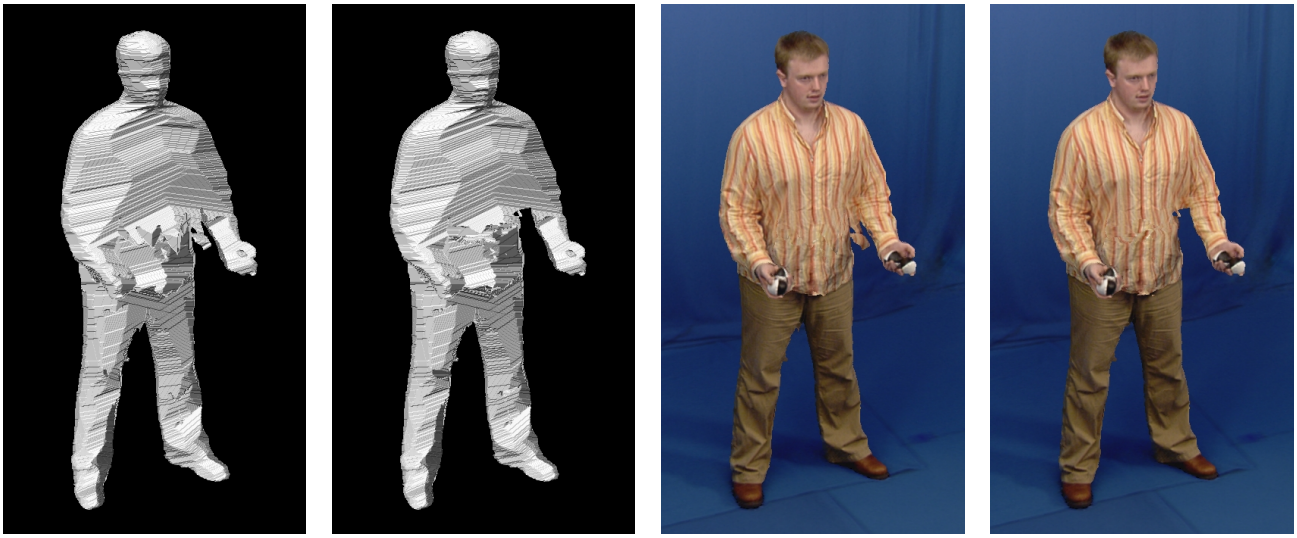
The first step is to construct the VDVH with respect to each real viewpoint, using that view's silhouette as a mask to make construction more efficient. The result is a multi-layer depth image containing the set of intervals inside the visual hull surface, which immediately gives us the required form for partitioning the image into safe and unsafe zones. A safe zone is made up of the pixels in the VDVH whose rays contain a single visual hull interval. The depth images which result from VDVH construction are shown in Figure 2(b). Pixels with depthels of only one interval are marked as safe, and every other occupied pixel marked as unsafe. Figure 2(c) shows the safe zones as white areas and the unsafe zones as grey areas.

The safe hull cannot be constructed by removing the unsafe zones from the silhouette, since these regions may correspond to a volume that has been declared safe by another camera. Instead the unsafe zones are used to determine the validity of points in the volume, or in the case of the VDVH, to select the correct interval.

3.5 Safe Hulls

The VDVH is constructed by casting rays out through the pixels of the virtual image, projecting them onto the real view's images and finding the intervals where the projected rays are inside the silhouette. The intervals are projected onto the original rays from the real view, and the mathematical intersection of intervals on each ray provides the depthel of the visual hull for that pixel (shown in the top section of Figure 3).

Safe hulls are constructed in a similar way to visual hull, with an additional selection process. When finding the intersections of a projected ray with a silhouette, we can also find the intersections with the safe zone in that image. The safe zone intervals are projected onto the original ray as well as the silhouette intervals. As for the visual hull, the mathematical intersection of the silhouette intervals gives the depthel of the visual hull. The depthel for the safe hull is provided by computing the mathematical intersection of the visual hull depthel with the union of all cameras' safe zones intervals (illustrated in Figure 3). Only visual hull intervals which are completely outside the safe zones are removed; those that are



(a) Visual hull surface (b) Safe hull surface (c) Visual hull rendered (d) Safe hull rendered

Figure 5: Taken from a juggling sequence, (b) shows the surface of the object after the phantom volumes from (a) have been removed. The rendered views of these surfaces are shown in (c) and (d). The quality of the synthesised view is severely affected by the presence of a phantom volume between the arm and body in (c). As a result of safe hull construction, (d) is much more realistic.

partly inside a safe zone are considered safe, because if part of the interval is a real object, the remainder must also be a real object.

The equivalent process in a volumetric construction is to test that a voxel is consistent across all silhouettes and that it appears in at least one safe zone, and should therefore be accepted.

Figure 5(a) displays a visual hull reconstruction of a person, with phantom volumes in front of the body, between the body and the arm, and around the inside of the legs. Figure 5(b) shows a safe hull reconstruction with these shape artefacts removed.

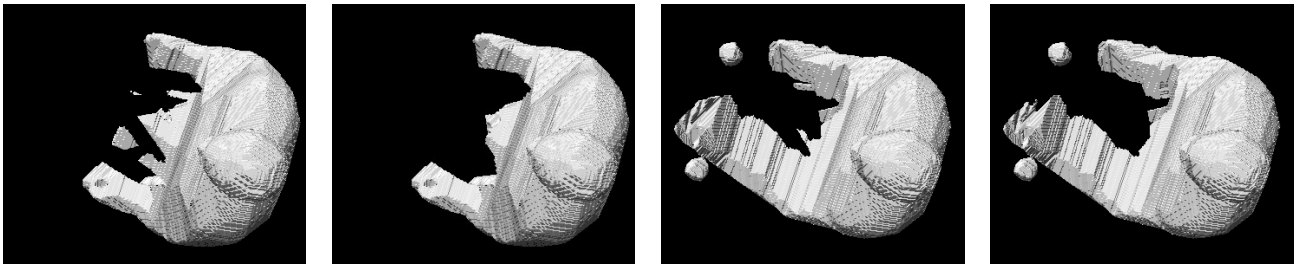
4 Results

This section presents results which demonstrate the effectiveness of safe hull construction, and how it enhances the realism of a virtual scene.

Two separate studio setups were used for capturing multiple synchronized video sequences for testing: Setup 1 captured at 25Hz SD resolution (720×576) progressive scan from eight equally spaced cameras in a complete circle of radius $6m$. Setup 2 captured video at 25Hz HD resolution (1920×1080) progressive scan from eight equally spaced cameras in an arc spanning $180°$, radius $4m$; Intrinsic and extrinsic camera parameters were estimated in both cases using the public domain calibration toolbox [1]. A third setup for static subjects used a Fuji s6500fd digital camera recording images at a resolution of 2048×1536 and calibrated using the GML Calibration Toolbox[13]. Tests were performed on an AMD 3100+ Sempron with 1GB RAM and results rendered using OpenGL on an nVidia 6600 graphics card.

Figure 4 shows the most common problem with multiple view video capture. This frame is from a sequence captured in Setup 1, and illustrates the problem of connected phantom volumes. These appear generally at the meeting point of two objects, and form a cone shape. Safe hull reconstruction removes these since they do not appear in any safe zone, and the generated result is of a higher visual quality. The slight stump in Figure 4(d) remaining in the safe hull is part of a safe zone; this volume cannot be removed and is considered part of the foreground. Under refinement this volume could be improved, whereas the removed part could only be improved by removing it. The pre-computation of visual hull for each real viewpoint required 2 seconds, and every safe hull thereafter took 6 seconds (virtual view size of 720×576).

The images shown in Figure 5 are produced from sequences captured in Setup 2. The visual hull produced a surface with phantom volumes in front of the person, between the body and the arm, and around the legs. Figure 5(b) shows the safe hull with these shape artefacts removed. Figure 6 shows a top-down view demonstrating the removal of the phantom volumes. The synthesised novel view is more realistic after safe hull construction. Pre-computation took 6 seconds per camera, and 20 seconds for safe hull construction (virtual view size



(a) Visual hull surface (b) Safe hull surface (c) Visual hull surface (d) Safe hull surface

Figure 6: Surfaces viewed from above: image (b) demonstrates removal of entire phantom volumes from (a); image (d) shows the safe hull reconstruction of (c), with the juggling balls intact - a heuristic solution based on size would have removed them. This would also be more difficult to produce using a model-based method.

of 1280×960).

The static setup was used to capture the images in Figure 7. This illustrates a worst-case scenario where there are very few original images (three in this case) of a subject with multiple surfaces and each view has an occlusion. The visual hull reconstruction is shown in Figure 7(a) shows the outcome of multiple occlusions: a large phantom volume in the centre of the subject, and some smaller phantoms at the top edges. Figure 7(b) shows the improvement the safe hull reconstruction has made, where only definite foreground remains and the phantoms have been removed. Some small parts of the foreground surface were also removed in the process, since the original views did not provide a comprehensive coverage of safe zones. Pre-computation took 8 seconds per camera, and 22 seconds for safe hull construction (virtual view size of 1280×960, silhouette area much larger for this capture).

The safe hull technique works well for subjects such as humans as shown in Figures 4, 5 and 6, and also for more complicated objects such as that in Figure 7. The results images demonstrate the higher quality of the rendered views using safe hull construction rather than visual hull.

5 Conclusions

This paper has introduced the first known constraint which allows phantom volume removal from visual hull reconstructions. This improves reconstruction accuracy and the overall quality of novel view synthesis. The approach presented here uses information from the visual hull and is reliable since it does not use heuristics or require additional cameras to remove shape artefacts.

The surface produced by the safe hull reconstruction is limited by the number of safe zones in the original images. If there are too few safe zones due to many occlusions and not enough viewpoints then parts of the real surface may be removed. However for many setups, especially those involving people, the safe hull produces good results with a small number of cameras.

For future work further processing of the surface not marked as definite foreground will be investigated, by applying feature and colour constraints. We also intend to try safe hull reconstruction on objects with multiple inter-occlusions causing phantom volumes, such as a tree, to find out the extent to which the safe hull can reduce the number of phantom volumes from visual hull.

References

[1] J.-Y. Bouguet. Camera calibration toolbox for matlab: www.vision.caltech.edu/bouguetj/calib-doc. Technical report, MRL-INTEL, 2003.

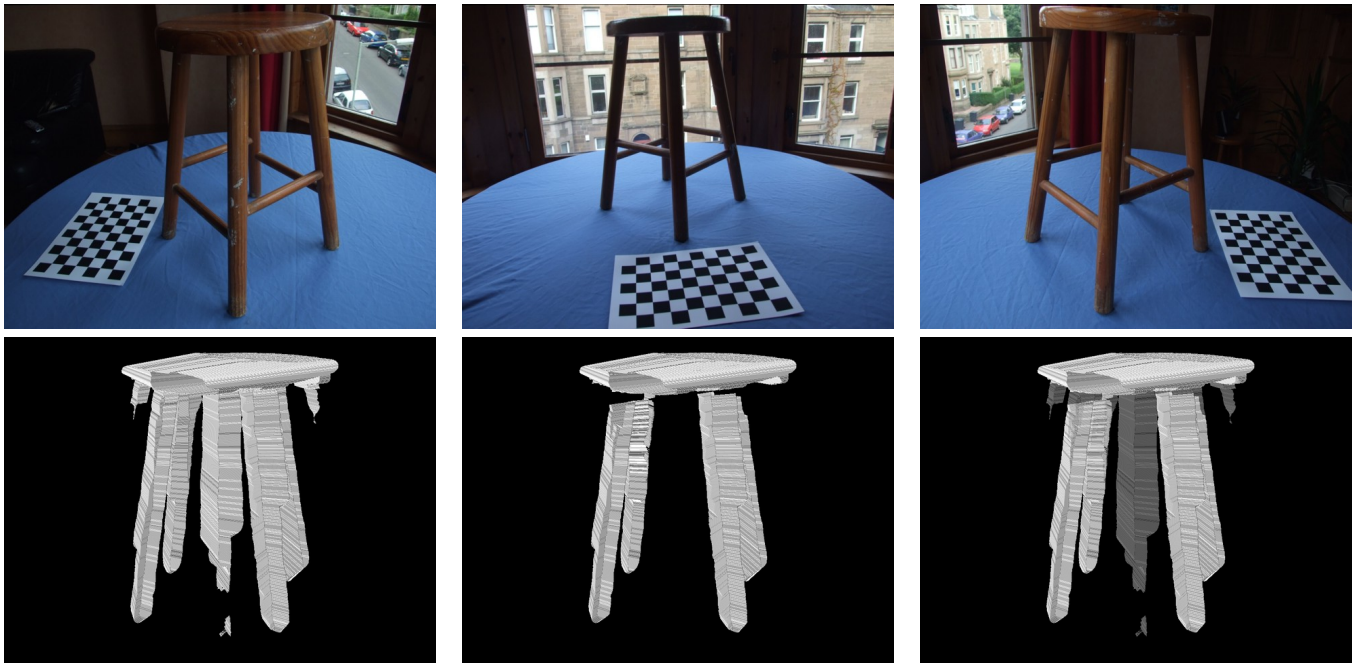
[2] E. Boyer and J. S. Franco. A hybrid approach for computing visual hulls of complex objects. In *Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume I*, pages 695–701. IEEE Computer Society, 2003.

[3] M. Brand, K. Kang, and D. B. Cooper. An algebraic solution to visual hull. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2004.

[4] J. Carranza, C. Theobalt, M. Magnor, and H.-P. Seidel. Free-viewpoint video of human actors. *Proceedings ACM SIGGRAPH*, 22(3):569–577, 2003.

[5] J.-S. Franco and E. Boyer. Exact polyhedral visual hulls. In *Fourteenth British Machine Vision Conference (BMVC)*, pages 329–338, September 2003. Norwich, UK.

[6] M. Goesele, S. M. Seitz, and B. Curless. Multi-view stereo revisited. In *Proceedings of CVPR*, June 2006.



(a) Visual hull surface

(b) Definite foreground surface

(c) Combined surface

Figure 7: The top row shows all original images used for the capture and the bottom row shows a virtual view of the surface. This capture illustrates a worst-case scenario with occlusion causing a large phantom volume to appear, shown in bottom row (a). The result in (b) shows the definite foreground areas, with the phantoms removed and some small sections where no safe zone existed. The final image in (c) shows the definite foreground with the rest of the surface rendered with transparency for comparison.

[7] A. Laurentini. The visual hull concept for silhouette-based image understanding. *IEEE Trans. Pattern Anal. Mach. Intell.*, 16(2):150–162, 1994.

[8] W. Matusik, C. Buehler, R. Raskar, S. J. Gortler, and L. McMillan. Image-based visual hulls. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pages 369–374. ACM Press/Addison-Wesley Publishing Co., 2000.

[9] G. Miller and A. Hilton. Exact view-dependent visual hulls. In *Proc. 18th International Conference on Pattern Recognition*. IEEE Computer Society, August 2006.

[10] G. Miller, J. Starck, and A. Hilton. Projective surface refinement for free-viewpoint video. In *Proc. 3rd European Conference on Visual Media Production*. IET, November 2006.

[11] G. Slabaugh, B. Culbertson, T. Malzbender, and R. Schafer. A survey of methods for volumetric scene reconstruction from photographs. In *Proc. of the Joint IEEE TCVG and Eurographics Workshop*. Springer Computer Science, 2001.

[12] S. Vedula, S. Baker, and T. Kanade. Spatio-temporal view interpolation. In *Proceedings of the 13th ACM Eurographics Workshop on Rendering*. ACM, June 2002.

[13] V.Vezhnevets and A.Velizhev. Gml c++ camera calibration toolbox: http://research.graphicon.ru/calibration/gml-c++-camera-calibration-toolbox.html. Technical report, 2005.

[14] S. Wuermlin, E. Lamboray, and M. Gross. 3d video fragments: Dynamic point samples for real-time free-viewpoint video. *Computers and Graphics, Special Issue on Coding, Compression and Streaming Techniques for 3D and Multimedia Data*, 28(1):3–14, 2004.

[15] D. B. Yang, H. H. Gonzalez-Banos, and L. J. Guibas. Counting people in crowds with a real-time network of simple image sensors. In *Proceedings of the Ninth IEEE International Conference on Computer Vision*, page 122, Washington, DC, USA, 2003. IEEE, IEEE Computer Society.

[16] C. Zitnick, S. Kang, M. Uyttendaele, S. Winder, and R. Szeliski. High-quality video view interpolation using a layered representation. In *SIGGRAPH*, pages 600—608, 2004.