Systems and Methods for Modeling Three-Dimensional Objects From Two-Dimensional Images

1-29-2013

Mubarak Shah  
University of Central Florida

Saad Khan  
University of Central Florida

Pingkun Yan  
University of Central Florida

Find similar works at: http://stars.library.ucf.edu/patents

Recommended Citation

http://stars.library.ucf.edu/patents/575

This Patent is brought to you for free and open access by the Technology Transfer at STARS. It has been accepted for inclusion in UCF Patents by an authorized administrator of STARS. For more information, please contact lee.dotson@ucf.edu.
SYSTEMS AND METHODS FOR MODELING THREE-DIMENSIONAL OBJECTS FROM TWO-DIMENSIONAL IMAGES

Inventors: Saad M. Khan, Hamilton, NJ (US);
Pingkun Yan, Peekskill, NY (US);
Mubarak Shah, Oviedo, FL (US)

Assignee: University of Central Florida Research Foundation, Inc., Orlando, FL (US)

Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 1050 days.

Appl. No.: 12/366,241
Filed: Feb. 5, 2009

Prior Publication Data

Related U.S. Application Data
Provisional application No. 61/026,661, filed on Feb. 6, 2008.

Int. Cl.
G06K 9/36 (2006.01)
G06T 15/00 (2011.01)

U.S. Cl. ...................... 382/154; 382/285; 345/419

Field of Classification Search ...................... 382/131,
382/242, 238, 203, 154, 285; 345/419, 163;
600/425; 359/9, 22, 23; 375/E7.081, E7.228,
375/E13.014, E13.019, E13.023, E13.025,
375/E13.033, E13.04, E13.044, E13.059,
375/E13.067, E13.071, E13.072, E13.073,
375/E13.038, 240.14; 707/999.002

See application file for complete search history.

ABSTRACT

In one embodiment, a system and method for modeling a three-dimensional object includes capturing two-dimensional images of the object from multiple different viewpoints to obtain multiple views of the object, estimating slices of the object that lie in parallel planes that cut through the object, and computing a surface of the object from the estimated slices.

28 Claims, 8 Drawing Sheets
FIG. 1
FIG. 3
CAPTURE MULTIPLE IMAGES OF OBJECT WITHIN A SCENE FROM MULTIPLE DIFFERENT VIEWPOINTS TO OBTAIN MULTIPLE VIEWS OF THE OBJECT

IDENTIFY FOREGROUND SILHOUETTES OF THE OBJECT IN EACH VIEW

DESIGNATE ONE OF THE VIEWS AS THE REFERENCE VIEW

WARP THE OTHER IMAGES TO THE REFERENCE IMAGE

FUSE THE WARPED SILHOUETTES OF EACH VIEW TO OBTAIN A CROSS-SECTIONAL SLICE OF A VISUAL HULL OF THE OBJECT THAT LIES IN THE REFERENCE PLANE

ESTIMATE FURTHER CROSS-SECTIONAL SLICES OF THE VISUAL HULL PARALLEL TO THE FIRST SLICE

FIG. 4A
STACK THE SLICES ON TOP OF EACH OTHER ALONG THE UP DIRECTION

COMPUTE OBJECT SURFACE FROM THE SLICE DATA

RENDER THE SURFACE USING A GRAPHICS ENGINE

OBJECT COLOR DESIRED?

PERFORM COLOR MAPPING

END

FIG. 4B
FIG. 9

FIG. 10
SYSTEMS AND METHODS FOR MODELING THREE-DIMENSIONAL OBJECTS FROM TWO-DIMENSIONAL IMAGES

CROSS-REFERENCE TO RELATED APPLICATION

This application claims priority to U.S. provisional application entitled "A Homographic Framework for the Fusion of Multi-View Silhouettes" having Ser. No. 61/026,561, filed Feb. 6, 2008, which is entirely incorporated herein by reference.

NOTICE OF GOVERNMENT-SPONSORED RESEARCH

This invention was made with Government support under Contract/Grant No.: NBCHCOB0105, awarded by the U.S. Government VACE program. The Government has certain rights in the claimed inventions.

BACKGROUND

Three-dimensional modeling is useful in various applications, including object localization, object recognition, and motion capture. There are a variety of methods that are currently used to model three-dimensional objects. One such method is a visual hull-based method in which silhouette information from two-dimensional images taken from multiple views in three-dimensional space are fused. Although such methods are viable, they typically require camera calibration, which is cumbersome. It would be desirable to be able to model three-dimensional objects from two-dimensional images without having to perform such calibration.

BRIEF DESCRIPTION OF THE FIGURES

The present disclosure may be better understood with reference to the following figures. Matching reference numerals designate corresponding parts throughout the figures, which are not necessarily drawn to scale.

FIG. 1 is a diagram that schematically illustrates a homography induced by a planar surface between two views of an object.

FIG. 2A is a diagram that schematically illustrates warping silhouettes of an object to a reference view using a planar homography induced by a first plane.

FIG. 2B is a diagram that schematically illustrates warping silhouettes of an object to a reference view using a planar homography induced by a second plane parallel to the first plane of FIG. 2A.

FIG. 3 is a diagram that illustrates the geometrical relationship of the homography of an image plane to two parallel scene planes.

FIGS. 4A and 4B together comprise a flow diagram that illustrates an embodiment of a method for three-dimensional modeling.

FIG. 5 comprises four images of an object captured from different viewpoints.

FIG. 6 is a view of four estimated object slices superimposed onto the outline of the object of FIG. 5.

FIG. 7 comprises two views of an object structure comprised of segmented object slices obtained for the object of FIG. 5 after rendering with a point rendering algorithm.

FIG. 8 comprises four zoomed-out views of the object structure of FIG. 7.

FIG. 9 is a schematic view of a system for three-dimensional modeling.

FIG. 10 is a block diagram of an example architecture for a computer system illustrated in FIG. 9.

DETAILED DESCRIPTION

Introduction

Disclosed herein are systems and methods for three-dimensional modeling using a purely image-based approach to fusing foreground silhouette information from multiple two-dimensional views. Unlike prior solutions, three-dimensional constructs such as camera calibration are unnecessary.

As described in greater detail below, visual hull intersection is performed in the image plane using planar homographies and foreground likelihood information from a set of arbitrary views of an object. A two-dimensional grid of object occupancy likelihoods is generated representing a cross-sectional slice of the object. Subsequent slices of the object are obtained by extending the process to planes parallel to a reference plane in a direction along the body of the object. Occupancy grids are then stacked on top of each other, creating a three-dimensional data structure that encapsulates the object shape and location. Finally, the object structure is segmented out by minimizing an energy function over the surface of the object in a level sets formulation.

The problem of determining a slice of an object can be stated as finding the region on a hypothetical plane that is occupied by the object. Through homographic warping of silhouette information from multiple views to a reference view, visual hull intersection on a plane can be achieved. If foreground information is available in each view, the process delivers a two-dimensional grid of space occupancies: a representation of a slice of the object cut out by the plane. Starting with homographies between views due to a reference plane in the scene (e.g., a ground plane), homographies of successively parallel planes can be obtained in the framework of plane-to-plane homographies using the vanishing point of the reference direction (the direction not parallel to the reference plane). With those homographies, an arbitrary number of occupancy grids/slices along the body of the object can be obtained, each being a discrete sampling of three-dimensional space of object occupancies.

Discussion of the Modeling Approach

Planar homographies will first be discussed in relation to FIG. 1, which illustrates a scene occupied by a cylinder 10. Assume x=(x, y, 1) denotes the image location (in homogeneous coordinates) of a three-dimensional scene point in a first view, View 1, and x'=(x', y', 1) are the coordinates of the scene point in a second view, View 2. Let H_1 denote the homography between the two views with respect to a scene plane \( \pi \). Warping View 1 onto View 2 using \( \text{Hom}_{\pi} \), the point x is transformed to \( x' \), where \( x'=H_1 x \). For scene points on plane \( \pi \), \( x_w-x' \), while for scene points off plane \( \pi \), \( x_w-x' \). The misalignment \( x_w-x' \) is called the plane parallax. Geometrically speaking, warping x from the View 1 to View 2 using homography H, amounts to projecting a ray from the camera center of View 1 through the scene point at location x and extending it until it intersects the plane \( H_1 \) at a point often referred to as the "piercing point" of x with respect to plane \( H_1 \). The ray is then projected from the piercing point onto the second view. The point in the image plane of the second view that the ray intersects is \( x_w \). In effect, \( x_w \) is the location at which the image of the piercing point is formed in the second camera. This phenomenon is illustrated by piercing points 12 and 14 in FIG. 1. As can be seen from that figure, scene points
on the plane $\pi$ have no plane parallax, while those off the plane have considerable plane parallax. Using these concepts, cross-sectional slices of the visual hull cut out by arbitrary planes in the scene can be obtained by the homographic fusion of multiple silhouettes onto a reference view.

FIG. 2A illustrates a scene containing a cylinder 16 that is viewed from several angles. The view $I_i$ is chosen as the reference view, although any of the views could be selected. By warping view $I_i$ to the reference view $I_1$, using homography $H_{1,i}$ induced by scene plane $\pi$, every foreground pixel in $I_i$ is projected to its piercing point on $\pi$. This process can be visualized as the foreground object casting a shadow on $\pi$ (an analogy if the cameras are replaced by point light sources), as depicted by the shadow regions 18 in FIG. 2A. The shadow is then projected onto the reference view $I_1$ to complete the operation of the homographic warping.

Computing the shadow is equivalent to determining the region on plane $\pi$ that falls inside the visual hull of the object image in $I_i$. The fusion of these shadows projected from various views therefore amounts to performing visual hull intersection on plane $\pi$, depicted by the region 20 in FIG. 2A. This process is performed implicitly, when all of the views are warped onto the reference view $I_1$ and fused to obtain the region 20 in the reference view $I_1$. Without loss of generality, the image plane of reference view $I_1$ after homographic fusion of foreground data can be viewed as a projectively transformed planar slice of the object (strictly speaking a perspective with only 6 degrees of freedom).

Instead of using binary foreground maps, a more statistical approach can be pursued and the background can be modeled in each view to obtain foreground likelihood maps, thereby using cameras as statistical occupancy sensors (foreground interpreted as occupancy in space). In the case of non-stationary cameras, object detection is achieved in a plane+parallax framework assigning high foreground likelihood where there is high motion parallax. A reason to adopt a soft approach is the act of thresholding preventing any premature decisions on pixel labeling; an approach that has proven to be very useful in visual hull methods due to their susceptibility to segmentation and calibration errors. Assume that $I_i$ is the foreground likelihood map (each pixel value is likelihood of being foreground) in view $i$. Consider a reference plane $\pi$ in the scene inducing homographies $H_{i,\pi}$ from view $i$ to view $j$. Warping $I_i$'s to a reference view $I_1$, the warped foreground likelihood maps are: $I_j=[H_{1,i}]I_i$. Visual hull intersection on $\pi$ (AND-fusion of the shadow regions) is achieved by multiplying these warped foreground likelihood maps:

$$\theta_{ref} = \prod_{i=1}^{n} I_i,$$

[Equation 1]

where $\theta_{ref}$ is the projectively transformed grid of object occupancy likelihoods. Notably, a more elaborate fusion model can be used at the expense of simplicity. For instance, a sensor fusion strategy that explicitly models pixel visibility, sensor reliability, or scene radiance can be transparently incorporated without affecting the underlying approach of fusing at slices in the image plane rather than in three-dimensional space.

Each value in $\theta_{ref}$ identifies the likelihood of a grid location being inside the body of the object, indeed representing a slice of the object cut out by plane $\pi$. It should be noted that the choice of reference view is irrelevant, as the slices obtained on all image planes and the scene plane $\pi$ are projectively equivalent. This computation can be performed at an arbitrary number of planes in the scene, each giving a new slice of the object. Naturally, this does not apply to planes that do not pass through the object’s body, since visual hull intersection on these planes will be empty. FIG. 2B demonstrates a second plane $\phi$ and a second slice 22 of the cylinder 16.

Starting with a reference plane in the scene (typically the ground plane), visual hull intersection can be performed on successively parallel planes in the up direction along the body of the object. The probabilistic occupancy grids $\theta_j$ obtained in this fashion can be thresholded to obtain object slices, but this creates the problem of finding the optimum threshold at each slice level. Moreover, the slices have a strong dependency on each other as they are parts of the same object(s), and should as such be treated as a whole. This dependency can be modeled by stacking the slices, creating a three-dimensional data structure $\Theta=[0_1; 0_2; \ldots; 0_n]$. $\Theta$ is not an entity in the three-dimensional world or a collection of voxels. It is simply put, a logical arrangement of planar slices representing discrete samplings of the continuous occupancy space. Object structure is then segmented out from $\Theta$, i.e., simultaneously segmented out from all the slices as a smooth surface that divides the space into the object and background. Discussed in the following paragraphs is an image-based approach using the homography of a reference plane in a scene to compute homographies induced between views by planes parallel to the reference plane.

Consider a coordinate system XYZ in space. Let the origin of the coordinate frame lie on the reference plane, with the X and Y axes spanning the plane. The Z axis is the reference direction, which is thus any direction not parallel to the plane. The image coordinate system is the usual xy affine image frame, and a point X in space is projected to the image point $x$ via a 3x4 projection matrix $M$ as:

$$x=MV=x(X, Y, Z, W),$$

[Equation 2]

where $x$ and $X$ are homogenous vectors in the form: $x=(x, y, w)^T$, $X=(X, Y, Z, W)^T$, and $"\cdot"$ means equality up to scale. The projection matrix $M$ can be parameterized as:

$$M=[v_{xy}, v_{yz}, v_{zw}, v_{zw}].$$

[Equation 3]

where $v_{xy}$, $v_{yz}$, and $v_{zw}$ are the vanishing points for X, Y, and Z directions respectively and $v_1$ is the vanishing line of the reference plane normalized.

Suppose the world coordinate system is translated from the plane $\pi$ onto the plane $\pi'$ along the reference direction (Z) by z units as shown in FIG. 3, then a new projection matrix $M'$ can be parameterized as:

$$M'=[v_{xy}, v_{yz}, v_{zw}, v_{zw}].$$

[Equation 4]

where $\alpha$ is a scale factor. Columns 1, 2, and 4 of the projection matrices are the three columns of the respective plane to image homographies. Therefore, the plane-to-image homographies can be extracted from the projection matrices, ignoring the third column, to give:

$$H_{\phi}=[v_{xy}, v_{yz}, v_{zw}, v_{zw}].$$

[Equation 5]

In general:

$$H_{\phi}=[H_{\phi}][v_{xy}, v_{yz}, v_{zw}, v_{zw}].$$

[Equation 6]

where $H_{\phi}$ is the homography of the reference plane, $\gamma$ is a scalar multiple encapsulating $\alpha$ and $z$, $[0]$ is a 3x2 matrix of zeros, and $v_{ref}$ is the vanishing point of the reference direction.

Using this result it can be shown that if the homography $H_{\gamma}$ is induced by a reference scene plane $\pi$ between views $i$ and $j$, then the homography $H_{\phi}$ induced by a plane $\phi$ parallel to $\pi$ in the direction of reference is given by:
As described above, slices computed along the body of the object can be stacked to create a three dimensional data structure \( \Theta \) that encapsulates the object structure. To segment out the object, a parameterized surface \( S(q) : [0, 1] \rightarrow RK^3 \) can be evolved that divides \( \Theta \) between the object and the background. This is achieved by formulating the problem in a variational framework, where the solution is a minimizer of a global cost functional that combines a smoothness prior on slice contours and a data fitness score. The function can be defined as:

\[
E(S) = \int_{S} \left| \nabla \Theta \left( S(q) \right) \right|^2 dq + \int_{S} \left| \frac{\partial S(q)}{\partial q} \right|^2 dq,
\]

where \( \nabla \Theta \) denotes gradient of \( \Theta \), and \( g \) denotes a strictly decreasing function: \( g(x) = 1/(1+x^2) \). The first term at the right side of Equation 8 represents external energy. Its role is to attract the surface towards the object boundary in \( \Theta \). The second term, called the internal energy, computes the area of the surface. Given the same volume, a smoother surface will have a smaller area. Therefore, this term controls the smoothness of the surface to be determined. When the overall energy is minimized, the object boundary will be approached by a smooth surface.

Minimizing Equation 8 is equivalent to computing geodesic in a Riemannian space:

\[
E(S) = \int g \left| \nabla \Theta \left( S(q) \right) \right|^2 dq + \int g \left| \frac{\partial S(q)}{\partial q} \right|^2 dq.
\]

With the Euler-Lagrange equation deduced, this objective function can be minimized by using the gradient descent method by an iteration time \( t \) as

\[
\frac{\partial \Theta}{\partial t} = g |\nabla \Theta | \nabla \Psi + \nabla g |\nabla \Theta | \nabla \Psi.
\]

Starting with an initial estimate for \( S \) and iteratively updating the level set function using Equation 11 leads to a segmentation of the object.
consecutive aligned views. Any pixel that has a mismatch from one view to the other (i.e., the color value difference is greater than a threshold) is marked as a pixel pertaining to the object. The alignment between such views can be determined, by finding the transformation, i.e., planar homography, between the views. In some embodiments, the homography can be determined between any two views by first identifying features of the ground plane using an appropriate algorithm or program, such as scale invariant feature transform (SIFT) algorithm or program. Once the features have been identified, the features can be matched across the views and the homographies can be determined in the manner described in the foregoing. By way of example, at least four features are identified to align any two views. In some embodiments, a suitable algorithm or program, such as a random sample consensus (RANSAC) algorithm or program, can be used to ensure that the identified features are in fact contained within the ground plane.

Once the foreground silhouettes have been identified in each view, visual hull intersection can be performed on multiple planes that cut through the object. Through that process, multiple slices of the object can be estimated, and those slices can then be used to compute a surface that approximates the outer surface of the object. As described above, the visual hull intersection begins with a reference plane. For purposes of this discussion, it is assumed that the reference plane is the ground plane.

With reference to block 34 of FIG. 4A, one of the obtained views is designated as the reference view. Next, each of the other views is warped to the reference view relative to the reference plane, as indicated in block 36. That is, the various views are transformed by obtaining the planar homography between each view and the reference view that is induced by the reference plane (i.e., the ground plane). Notably, those homographies can be obtained by determining the homographies between consecutive views and concatenating each of those homographies to produce the homography between each of the views and the reference view. Such a process may be considered preferable given that it may reduce error that could otherwise occur when homographies are determined between views that are spaced far apart from each other.

After each of the views, and their silhouettes, has been transformed (i.e., warped to the reference view using the planar homography), the warped silhouettes of each view are fused together to obtain a cross-sectional slice of a visual hull of the object that lies in the reference plane, as indicated in block 38. That is, a first slice of the object (i.e., a portion of the object that is occluded from view) that is present at the ground plane is estimated.

As described above, this process can be replicated to obtain further slices of the object that lie in planes parallel to the reference plane. Given that those other planes are imaginary, and therefore comprise no identifiable features, the transformation used to obtain the first slice cannot be performed to obtain the other slices. However, because the homographies induced by the reference plane and the location of the vanishing point in the up direction are known, the homographies induced by any plane parallel to the reference plane can be estimated. Therefore, each of the views can be warped to the reference view relative to new planes, and the warped silhouettes that result can be fused together to estimate further cross-sectional slices of the visual hull, as indicated in block 40.

As described above, the homographies can be estimated using Equation 7. In that equation, γ is a scalar multiple that specifies the locations of other planes along the up direction. Notably, the value for γ can be selected by determining the range for γ that spans the object. This is achieved by incrementing γ in Equation 7 until a point is reached at which there is no shadow overlap, indicating that the current plane is above the top of the object. Once the range has been determined, the value for γ at that point can be divided by the total number of planes that are desired to determine the appropriate value of γ to use. For example, if γ is 10 at the top of the object and 100 planes are desired, γ can be set to 0.1 to obtain the homographies induced by the various planes.

At this point in the process, multiple slices of the object have been estimated. FIG. 6 illustrates four example slices, slices 70-76, which have been superimposed over the outline of the object 60 from FIG. 5. As with the number of views, the greater the number of slices, the more accurate the results that can be obtained. In the experiment performed with the object 60, 200 slices were estimated.

Once the slices have been estimated, their precise boundaries are still unknown and, therefore, the precise boundaries of the object are likewise unknown. One way in which the boundaries of the slices could be determined is to establish thresholds for each of the slices to separate image data considered part of the object from image data considered part of the background. In the current embodiment, however, the various slices are first stacked on top of each other along the up direction, as indicated in block 42 of FIG. 4B to generate a three-dimensional “box” (i.e., the data structure 8) that encloses the object and the background. At that point, a surface can be computed that divides the three-dimensional box into the object and the background to segment out the object surface. In other words, an object surface can be computed from the slice data, as indicated in block 44 of FIG. 4B.

As described above, the surface can be computed by minimizing Equation 8, which comprises a first term that identifies portions of the data that have high gradient (thereby identifying the boundary of the object) and the second term identifies the surface area of the object surface. By minimizing both terms, the surface is optimized as a surface that moves toward the object boundary and has as small a surface area as possible. In other words, the surface is optimized to be the tightest surface that divides the three-dimensional surface of the object from the background.

After the object surface has been computed, the three-dimensional locations of points on the surface are known and, as indicated in block 46, the surface can be rendered using a graphics engine. FIG. 7 illustrates an object structure 80, comprised of parallel segmented slices 82, that results when such rendering is performed. In FIG. 7, the left image is a close up of a side of the object 60 and the right image is a view of the object from its bottom looking upward.

At this point, a three-dimensional model of the object has been produced, which can be used for various purposes, including object localization, object recognition, and motion capture. It can then be determined whether the colors of the object are desired, as indicated in decision block 48 of FIG. 4B. If not, flow for the process is terminated. If so, however, the process continues to block 50 at which color mapping is performed. In some embodiments, color mapping can be achieved by identifying the color values for the slices from the outer edges of the slices, which correspond to the outer surface of the object. A visibility check can be performed to determine which of the pixels of the slices pertain to the outer edges. Specifically, pixels within discrete regions of the slices can be “moved” along the direction of the vanishing point to determine if the pixels move toward or away from the center of the slice. The same process is performed for the pixels across multiple views and, if the pixels consistently move toward the center of the slice, they can be assumed to com-
A method for modeling a three-dimensional object, the method comprising:
1. capturing two-dimensional images of the object from multiple different viewpoints to obtain multiple views of the object;
2. estimating slices of the object that lie in parallel planes that cut through the object; and
3. computing a three-dimensional data structure that represents the object and its background from the estimated slices by stacking the estimated slices on top of each other along an up direction of the object.

We claim:
1. A method for modeling a three-dimensional object, the method comprising:
   1. capturing two-dimensional images of the object from multiple different viewpoints to obtain multiple views of the object;
   2. estimating slices of the object that lie in parallel planes that cut through the object; and
   3. computing a three-dimensional data structure that represents the object and its background from the estimated slices by stacking the estimated slices on top of each other along an up direction of the object.

2. The method of claim 1, wherein estimating slices comprises identifying foreground silhouettes of the object in each view.

3. The method of claim 2, wherein estimating slices further comprises designating one of the views as the reference view.

4. The method of claim 3, wherein estimating slices further comprises warping each view other than the reference view to the reference view to obtain warped silhouettes.

5. The method of claim 4, wherein warping comprises determining the planar homography between each other view and the reference view that is induced by a ground plane upon which the object is supported.

We claim:
5. A method for modeling a three-dimensional object, the method comprising:
   1. capturing two-dimensional images of the object from multiple different viewpoints to obtain multiple views of the object;
   2. estimating slices of the object that lie in parallel planes that cut through the object; and
   3. computing a three-dimensional data structure that represents the object and its background from the estimated slices by stacking the estimated slices on top of each other along an up direction of the object.

6. The method of claim 5, wherein estimating slices comprises identifying foreground silhouettes of the object in each view.

7. The method of claim 6, wherein estimating slices further comprises estimating planar homographies induced by additional planes parallel to the ground plane using the planar homographies induced by the ground plane and the location of the vanishing point in the up direction of the object.

8. The method of claim 7, wherein estimating slices further comprises, as to each additional plane, warping each view to the reference view relative to the additional plane to obtain warped silhouettes and fusing the warped silhouettes to obtain a further cross-sectional slice of the visual hull of the object.

We claim:
8. A method for modeling a three-dimensional object, the method comprising:
   1. capturing two-dimensional images of the object from multiple different viewpoints to obtain multiple views of the object;
   2. estimating slices of the object that lie in parallel planes that cut through the object; and
   3. computing a three-dimensional data structure that represents the object and its background from the estimated slices by stacking the estimated slices on top of each other along an up direction of the object.

9. The method of claim 8, wherein computing a three-dimensional data structure comprises dividing the object from the background in the three-dimensional data structure to segment out the surface of the object.

We claim:
9. A method for modeling a three-dimensional object, the method comprising:
   1. capturing two-dimensional images of the object from multiple different viewpoints to obtain multiple views of the object;
   2. estimating slices of the object that lie in parallel planes that cut through the object; and
   3. computing a three-dimensional data structure that represents the object and its background from the estimated slices by stacking the estimated slices on top of each other along an up direction of the object.

10. The method of claim 9, wherein dividing the object from the background comprises minimizing an energy function over the object surface.

We claim:
10. A method for modeling a three-dimensional object, the method comprising:
   1. capturing two-dimensional images of the object from multiple different viewpoints to obtain multiple views of the object;
   2. estimating slices of the object that lie in parallel planes that cut through the object; and
   3. computing a three-dimensional data structure that represents the object and its background from the estimated slices by stacking the estimated slices on top of each other along an up direction of the object.

11. The method of claim 10, further comprising rendering the object surface.

We claim:
11. A method for modeling a three-dimensional object, the method comprising:
   1. capturing two-dimensional images of the object from multiple different viewpoints to obtain multiple views of the object;
   2. estimating slices of the object that lie in parallel planes that cut through the object; and
   3. computing a three-dimensional data structure that represents the object and its background from the estimated slices by stacking the estimated slices on top of each other along an up direction of the object.

12. The method of claim 11, further comprising color mapping the rendered object surface.

We claim:
12. A method for modeling a three-dimensional object, the method comprising:
   1. capturing two-dimensional images of the object from multiple different viewpoints to obtain multiple views of the object;
   2. estimating slices of the object that lie in parallel planes that cut through the object; and
   3. computing a three-dimensional data structure that represents the object and its background from the estimated slices by stacking the estimated slices on top of each other along an up direction of the object.
estimating further cross-sectional slices of the visual hull that lie in additional planes above and parallel to the ground plane; stacking the cross-sectional slices on top of each other along an up direction of the object to create a three-dimensional data structure that represents the object and its background; and computing an object surface from the data structure by minimizing an energy function for the data structure.

14. The method of claim 13, wherein warping comprises determining the planar homography between each view and the reference view that is induced by the ground plane.

15. The method of claim 14, wherein estimating further cross-sectional slices comprises estimating planar homographies induced by the additional planes using the planar homographies induced by the ground plane and the location of the vanishing point in the up direction of the object.

16. The method of claim 15, wherein estimating further cross-sectional slices further comprises, as to each additional plane, warping each view to the reference view relative to the additional plane to obtain warped silhouettes and fusing the warped silhouettes to obtain a further cross-sectional slice of the visual hull of the object.

17. The method of claim 14, wherein minimizing an energy function comprises minimizing an energy function that comprises a gradient term and a surface area term.

18. The method of claim 14, further comprising rendering the object surface.

19. The method of claim 14, further comprising color mapping the rendered object surface.

20. A non-transitory computer-readable medium comprising:

logic configured to compute a three-dimensional data structure that represents the object and its background from the estimated slices by stacking the estimated slices on top of each other along an up direction of the object.

21. The non-transitory computer-readable medium of claim 20, further comprising logic configured to warp views to a reference view to obtain warped silhouettes related to the ground plane.

22. The non-transitory computer-readable medium of claim 21, wherein the logic configured to warp is configured to determine planar homographies between the reference view and the other views that are induced by the ground plane.

23. The non-transitory computer-readable medium of claim 22, further comprising logic configured to fuse the warped silhouettes to obtain a cross-sectional slice of a visual hull of the object that lies in the ground plane.

24. The non-transitory computer-readable medium of claim 23, wherein the logic configured to estimate further slices is configured to estimate planar homographies induced by the additional planes using planar homographies induced by the ground plane and the location of the vanishing point in the up direction of the object.

25. The non-transitory computer-readable medium of claim 20, wherein the logic configured to compute a surface is configured to divide the object from the background in the three-dimensional data structure to segment out the surface of the object.

26. The non-transitory computer-readable medium of claim 25, wherein the logic configured to compute a surface is further configured to minimize an energy function over the object surface.

27. The non-transitory computer-readable medium of claim 20, further comprising logic configured to render the object surface.

28. The non-transitory computer-readable medium of claim 20, further comprising logic configured to color map the rendered object surface.