3D Visualization of Products for Online Shopping

Vishakha Patel
Dwarkadas J. Sanghvi College
of Engineering
Mumbai, Maharshtra, India

Gayatri Prabhu
Dwarkadas J. Sanghvi Coll
of Engineering
Mumbai, Maharshtra, India

Ekta Rita
Dwarkadas J. Sanghvi College
of Engineering
Mumbai, Maharshtra, India

Ruhina Karani
Asst. Prof.,
Dwarkadas J. Sanghvi College of Engineering
Mumbai,
Maharshtra, India

ABSTRACT

With the increasing demands on 3D applications and the easy capturing of 2D images nowadays, building 3D models from 2D images receives much attention in the past few years. 3D modeling is widely used in several fields- 3D graphics in computer games, software architecture models and 3D printing. 3D models represent a physical body using a collection of points in 3D space, connected by various geometric entities such as triangles, lines, curved surfaces, etc. 3D modeling is the process of developing a mathematical representation of any three-dimensional surface of an object. Today, 3D models are used in a wide variety of fields. The engineering community uses them as designs of new devices, vehicles and structures as well as a host of other uses. A variety of machine learning algorithms are being studied and implemented to find or estimate the depth information which is unavailable in conventional 2D image. We apply Computer Vision algorithm considering aspect of binocular disparity where we use 2 images of same scene captured from different viewpoints. Then we obtain depth of the object and further construct depth map. After mapping the points from depth maps of various images captured we apply correct texturing to obtain full 3D model of the object.capturing one eye's view, and depth information is computed using binocular disparity. Here we focus only on binocular and multi-ocular images as input. The two or more input images could be taken either by multiple fixed cameras located at different viewing angles or by a single camera with moving objects in the scenes.

A three-dimensional (3D) visualization enables consumers to interact with products and creates a sense of being in a simulated real world. As the consumer gets a real view of the products they tend to get attracted towards the product thus increasing the sale. It gives us an edge over other competitors as 3D visualization is different and it stands out from others. It also makes shopping more convenient and easy for the customers. In this paper we focus only on binocular and multi-ocular images as input. We study computer vision algorithm, binocular disparity, silhouette and visual hull.

In computer vision algorithm, SURF and ORB features descriptors are used to extract information from images. Binocular disparity uses 2 images of the same scene from different viewpoints. In silhouette the object is separated from the background and silhouette cones are formed. Intersection of silhouette cones is called visual hulls.

General Terms

2D images, Computer Vision, 3D reconstruction

Keywords

Computer Vision algorithm, Binocular disparity, Silhouette, Visual Hull

1. INTRODUCTION

3D Visualization of products enables the creation and manipulation of product images on a website. Using 3D may enhance a sense of telepresence, which leads consumers to get more information about the product and it adds a sense of fun to consumers' virtual experiences. Moreover, it proposes that both values might end up in enhancing users' approaches toward an online store. On online shopping sites, only snapshots of 2D items are displayed for users. Hence the users cannot estimate the size or dimension of overall product. Conversion of 2D image to 3D model will help users to visualize the product accurately. Several special cameras have been designed to generate 3D model directly. For example, a stereoscopic dual-camera makes use of a co-planar configuration of two separate, monoscopic cameras, each

2. LITERATURE SURVEY

2.1 Computer Vision Algorithm

As a scientific discipline, computer vision is concerned with the theory behind artificial systems that extract information from images. The image data can take many forms, such as video sequences, views from multiple cameras, or multi-dimensional data from a medical scanner. As a technological discipline, computer vision seeks to apply its theories and models to the construction of computer vision systems. In Computer Vision algorithm for 2D to 3D, the algorithm utilizes SURF and ORB feature descriptors to establish correspondence between 2D projective images of known camera positions. ORB (Oriented FAST and Rotated BRIEF) is a fast robust local feature detector, first presented by Ethan Rublee et al. in 2011, which is used in computer vision tasks like object recognition and 3D reconstruction. ORB uses an orientation compensation mechanism, making it rotation invariant. The epi-polar geometry gives the location of the feature point in 3D space for each correspondence pairs. After testing the virtual

models in OpenGL, we can apply the same reconstruction technique to real images taken in a noisy environment. After camera calibration and illumination adjustments, the computer vision algorithm generates fairly accurate results (plus noticeable noise) even with background clutter, imprecise body poses and movement.

Binocular Disparity

With two images of the same scene captured from slightly different viewpoints, the binocular disparity can be utilized to recover the depth of an object [1]. This is the main mechanism for depth perception. First, a set of corresponding points in the image pair are found. Then, by means of the triangulation method, the depth information can be retrieved with a high degree of accuracy (see Figure 1) when all the parameters of the stereo system are known. When only intrinsic camera parameters are available, the depth can be recovered correctly up to a scale factor. In the case when no camera parameters are known, the resulting depth is correct up to a projective transformation.
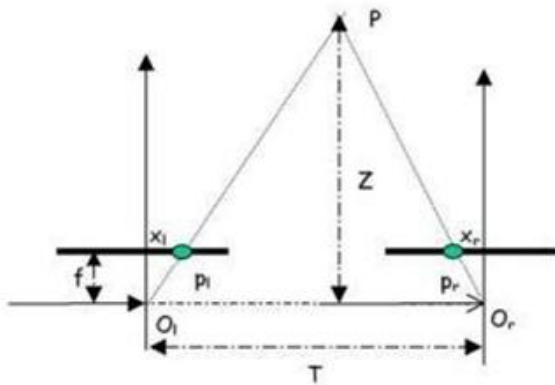

Fig 1: Binocular Disparity for depth estimation

Assume pl and pr are the projections of the 3D point P on the left image and right image; Ol and Or are the origin of camera coordinate systems of the left and right cameras. Based on the relationship between similar triangles (P, pl, pr) and (P, Oi, Or) shown in Figure 1, , the depth value Z of the point P can be obtained:

$Z = f*T/d$ [1]

where $d = x_i − x_r$, which measures the difference in retinal position between corresponding image points. The disparity value of a point is often interpreted as the inversed distances to the observed objects. Therefore, finding the disparity map is essential for the construction of the depth map.

2.2 Silhouette

A new sketch-based modeling approach is used in which models are interactively designed by drawing their 2D silhouettes from different views [2]. The core idea was to limit the input to 2D silhouettes, removing the need to explicitly create or position 3D elements. Here visual hull is used as a modeling tool where it is limited in complexity of models it can produce and it cannot generate concavities. The following were the steps to produce 3D model [2].

1. Taking input from user- a 3D model is assembled out of parts. Each part is specified by two or three silhouettes from front, side, or top views. As the user creates and refines the silhouettes of a part, the corresponding 3D model is automatically generated and displayed. A part's silhouette in one view imposes certain constraints on the part's silhouettes

2. in the other views. For example, the height of a part's silhouette in the side view should match the height of the part's silhouette in the front view.

3. Rotation: Once a part has been drawn into the model, it can be rotated in 3D relative to the rest of the model. For rotation silhouettes must be orthogonal to each other so that they are edited with respect to axes.

4. Applying Constructive Solid geometry to create final model- Models in this approach are composed out of parts defined by their silhouettes, which the user can combine using Boolean operations. A model before smoothing can therefore be represented by a CSG tree applied to silhouette cylinders. A silhouette cylinder is the orthographic equivalent of a silhouette cone: it is the infinite extrusion of a silhouette in the viewing direction.

2.3 Visual Hull using Silhouette

Multi view geometry based algorithms uses multiple silhouettes to reconstruct the real world objects and these algorithms involves three different steps.

1. To obtain the silhouettes. Szeliski [3] used a turntable when obtaining the silhouettes while Ladikos et al. [4] on the other hand use a fixed setup of 16 cameras mounted in a room. The Euclidian space containing the possible camera placements are referred to as viewing region and Laurentini [5] shows that different viewing regions may give different results. After an image have been obtained the object silhouette must be extracted, a process that can be done with an intensity threshold as mentioned in [3, 5], by using foreground/background subtraction as in [6, 7].

2. When the images are collected and silhouettes have been found the next step is to identify the camera parameters [8]. According to Szeliski these parameters include the cameras position relative the object, orientation and focal length.

3. The last step of multi view geometry based algorithms is the actual 3D reconstruction. Liang and Wong [4] classify the 3D reconstruction algorithms using 2D images as being part of one of two schools, the volumetric school and the differential school. Liang and Wong [4] describe the volumetric school as "treating objects as solid volumes" they also say that the differential school is defined by the fact that it "treats the object surface as an infinitesimally thin shell".
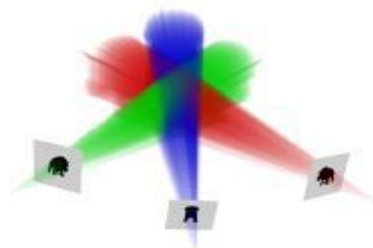

Fig 2: Geometric representation of object obtained after intersection of silhouette cones is called visualhull.

Visual Hull Reconstruction algorithm

The basic principle is to create a 3D representation of an object by its silhouettes within several images from different viewpoints. Each of these silhouettes by different camera views form in their projection a cone, called visual cone and an intersection of all these cones form a description of the real object's shape. By using this basic idea there are many advantages in using Shape-From-Silhouette techniques [9]. First of all the calculation of the silhouettes is easily to implement, when we assume an indoor environment with special conditions, like static light and static cameras. Without these assumptions it can become difficult to calculate an accurate silhouette out of the images, because of shadows or moving backgrounds. On the other hand are the implementations of the SFS-algorithm straight forward and especially compared to other techniques for shape estimations, like multi-baseline stereo far less complex. The result of the

SFS construction is an upper bound of the real object's shape in contrast to a lower bound, which is a big advantage for obstacle avoidance in the field of robotic or visibility analysis in navigation.

The main problem of the SFS-based algorithms is that they are not able to perform an accurate reconstruction of concave objects. An obvious question, which occurs in this context, is which parts of an object can be reconstructed by standard SFS techniques, or what are the limits of these approaches? To denote this difference Laurentini introduced the term of the Visual Hull in 1991. His formal definition of the Visual Hull is the following: "The visual hull $V H(S, R)$ of an object S relative to a viewing region R is a region of E3 such that, for each point $P \in V H(S, R)$ and each viewpoint $V \in R$, the half-line starting at V and passing through P contains at least a point of S."
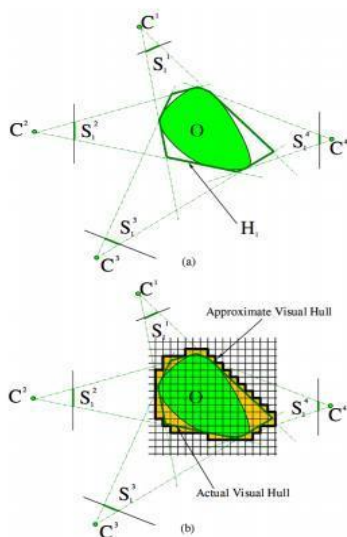

Fig 3: Approximate and actual Visual Hull

3. COMPARATIVE STUDY

	Computer Vision Hull using Binocular Disparity	Silhouette Visual using	
Number of Images	2-3	multiple	multiple
Feature Descriptors	SURF and ORB	Not Applicable	Not Applicable
Texturing Algorithm	Not Required	Required	Required
Accuracy	High	High only if object can be clearly separated from background	When more than 10 images are used accuracy is visible in smaller details
Colour Consistency	High	Medium	Medium
Images with cavities	Allowed	Not Allowed	Not Allowed

4. PROPOSED SYSTEM

The steps for the process are documented below.

1. Image Capturing: 2D images of a particular object that is to be converted to 3D model is captured from different angles. Two or more input images could be taken either by multiple fixed cameras located at different viewing angles or by a single camera with moving objects in the scenes using a rotating table. With input images from multiple views, multi-baseline stereo is used in the implementation to refine the surface estimates from focus. For each reference image, two previous reference images and two following reference images are selected for multi-baseline stereo analysis.

2. Feature Extraction: Image features at various levels of complexity are extracted from the image data. Typical examples of such features are lines, edges and ridges.

3. Depth Estimation: We view depth estimation as a small but crucial step towards the larger goal of image understanding, in that it will help in tasks such as understanding the spatial layout of a scene, finding walk able areas in a scene, detecting objects, etc. The 3D vision processes have as goal the reconstruction of this lost information, and, thus, the distances from each projected point to the image plane. On the one hand, some extra information has to be obtained, for absolute depth estimation. On the other hand, there are, geometrically, infinite points in the scene that are not projected and, then, must be, in some cases, interpolated. This is the case of occluded points.

4. Object Detection: An individual instance of an object is recognized. Examples include identification of a specific person's face or fingerprint. Feature based method is used which is as follows:

- A search is used to find feasible matches between object

features and image features.
The primary constraint is that a single position of the object
Must account for all of the feasible matches.
Methods that extract features from the objects to be
recognized and the images to be searched.

5.   3D Model Reconstruction: Given one or more images of
     an object, reconstruction aims at computing 3D

model of the object. In the simplest case the model can be a
set of 3D points. More sophisticated methods produce a
complete 3D surface model. The advent of 3D imaging not
requiring motion or scanning and related processing
algorithms is enabling rapid advances in this field. Grid-based
3D sensing can be used to acquire 3D images from multiple
angles.

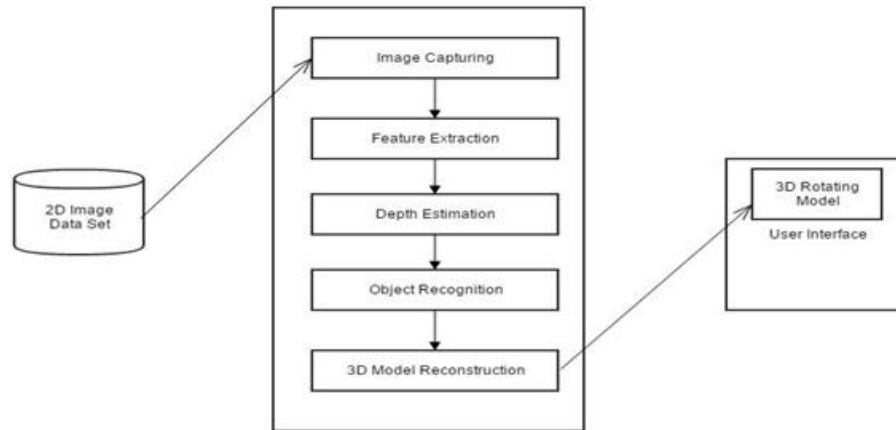Block diagram for the proposed system is given below in
Figure 4:

Fig 4: Model of Proposed System

5.   CONCLUSION AND FUTURE WORK

In today's world we see that the majority of the earth's
population prefers to go shopping online as it not only saves
your efforts but also time. But the most eminent drawback of
online shopping sites is that these sites are unable to provide
us an actual overall view of the object. This problem can be
resolved using 3-D modeling discussed in this paper, wherein
we take already available 2-D images of the object and
convert them into reliable 3-D models. A number of 2D to 3D
conversion algorithms are dedicated to recover the "structure"
or "shape" of objects in the images, which are understood to
mean the 3D coordinates of a small set of points in the scene.
In this paper we studied in detail about computer vision
algorithm, visual hull, binocular disparity and silhouette. We
plan to build a 3-D model with a collaboration of the same.
We also learnt that a new trend of the development of 2D to
3D conversion algorithms is to operate in association with
robust algorithms for image semantics analysis and to design
specialized conversion algorithm for each specific semantic
entity. The resulting 3-D model will not only be useful for
online shopping sites but also in numerous other fields like
building architecture, etc.

In future this can be taken to a next level by not only creating
3-D models of the object on the online shopping sites but also
of ourselves instantaneously at one go by uploading our 2-D
images and then trying the objects we intend to buy on the
created 3-D model. This would further make online shopping
convenient and attract more customers. Multiprocessing of
'the 2-D images taken as input' with multiple chips and
proper algorithm can also increase the speed of 3-D model
generation.

6.   ACKNOWLEDGMENTS

7.   REFERENCES

[1]   Assoc. Prof. Dr. Ir. E. A. Hendriks, Dr. Ir. P. A. Redert,
      "Converting 2D to 3D- A Survey Information and
      Communication Theory Group (ICT), Faculty of
      Electrical Engineering, Mathematics and Computer
      Science, Delft University of Technology, the
      Netherlands, 2005

[2]   Alec Rivers, Frédo Durand, and Takeo Igarashi. 2010.
      3D modeling with silhouettes. In ACM SIGGRAPH
      2010 papers (SIGGRAPH '10), Hugues Hoppe (Ed.).
      ACM, New York, NY,USA, , Article 109 , 8 pages

[3]   T. Matsuyama, X. Wu, T. Takai and T. Wada, "Real-
      Time Dynamic 3D Object Shape Reconstruction and
      High-Fidelity Texture Mapping for 3D Video", *Circuits
      and Systems for Video Technology, IEEE Transactions
      on*, vol. 14, no. 3, pp. 357-369, Mar. 2004.

[4]   C. Liang and K. K. Wong, "3D reconstruction using
      silhouettes from unordered viewpoints", *Image and
      Vision Computing*, vol. 28, no. 4, pp. 579-589, Apr.
      2010.

[5]   M. Li, M. Magnor and H. P. Seidel, "Improved
      Hardware-Accelerated Visual Hull Rendering", *Vision,
      Modeling, and Visualization 2003*, pp. 151-158, Nov.
      2003

[6]   M. Li, M. Magnor and H. P. Seidel, "Hardware-
      Accelerated Visual Hull Reconstruction and
      Rendering",*Graphics Interface 2003, Proceeding*, pp.
      65-71, 2003.

[7]   M. Li, M. Magnor and H.P. Seidel, "A Hybrid
      Hardware-Accelerated Algorithm for High Quality
      Rendering of

Visual Hulls", *Graphics Interface 2004, Proceedings*, pp. 41-48, 2004.

[8] Fuad Al-Amin, David Shuang Liu, Katherine Chen, YooHsiu Yeh, "Learning 3D Models", CS229 Final ProjectReport,Department of computer science and electrical engineering, Stanford University.

[9] M. Li, M. Magnor and H.P. Seidel, "A Hybrid Hardware-Accelerated Algorithm for High Quality

Rendering ofVisual Hulls", *Graphics Interface 2004, Proceedings*, pp. 41-48, 2004.

[10] Fuad Al-Amin, David Shuang Liu, Katherine Chen, YooHsiu Yeh, "Learning 3D Models", CS229 Final ProjectReport,Department of computer science and electrical engineering, Stanford University.