Chapter 6

An Overview of Three-Dimensional Videos: 3D Content Creation, 3D Representation and Visualization

Lourena Rocha and Luiz Gonçalves

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/50177

1. Introduction

The upcoming of digital video has caused a technological revolution that has changed audiovisual communication in several ways. The digital format, in its essence, is appropriate to computational processing. As a consequence, it has a huge impact in the cinema and television industries. Nowadays, with advances experimented in Internet and wireless networking, digital video has been consolidated as a new and important media. For example, the Skype application relies in this kind of media in order to allow partners that are distant far away communicate to each other.

Current generation of digital video brings revolutionary aspects as the incorporation of new data types in the media. Depth information is certainly one data type that is typically natural, inserted in digital videos in order to provide more realism. That is, the insertion of depth agrees with human perceptual system and also makes easier the scene analysis using computers, mainly if the goal is to extract high-level information. In this way, three-dimensional video (or simply 3D video) comes up, used to reproduce images in movement with the third dimension sensation or to recreate a dynamic scene visualization with other viewpoints besides the one that the movie has been filmed. 3D videos that allow the scene visualization from new viewpoints can be constructed using an image or model based approach. These type of 3D videos are known as free-viewpoint video, or so-called FVV, and 3D videos providing depth perception are so-called 3DV or stereoscopic videos.

So, in the scope of this text, the main characteristic of a 3D video is that it captures the dynamics and movement of the scene during the filming, offering to the user the possibility to change the point of view during the exhibition, beyond supplying the three-dimensional model of visualized objects. Automatic construction of three-dimensional photo-realistic models of a scene is important in applications such as interactive visualization of environment

or objects that are remotely located, for example. One could provide a modification of a real scene for virtual reality tasks. Other applications of 3D video are in Archeology, Oceanography, Historic and Cultural Sites, Arts, Education and Entertainment.

In general, an end-to-end 3D video system pipeline consists of the following stages: capture system setup, 3D reconstruction, 3D representation, coding, transmission, decoding, rendenring and 3D display. They can be classified in four main blocks: *3D Content Creation* (capture and 3D recontruction stages), *3D Representation*, *Delivery* (coding, transmission and decoding stages) and *Visualization* (rendering and 3D display stages).

In this text, we provide an extensive literature review on 3D Content Creation, 3D Representation and Visualization blocks of the 3D video pipeline. The Delivery block regarding coding, transmission and decoding techniques is not in the scope of this text. It is mainly intended for applications involving some network channels, such as, internet applications and 3D TV.

3D videos are one of the most active research topics and other reviews have already been proposed [63, 66, 74].

The chapter is organized as follows. Section 2 explains the pipeline of 3D videos from capture to display. As part of the 3D Content Creation block, we discuss acquisition systems and 3D reconstructions techniques in Section 3. Section 4 presents the most popular 3D representations formats in the context of 3DV and FVV. The Visualization block, with rendering and 3D display stages, is discussed in Section 5. Finally, Section 6 concludes the chapter.

2. Pipeline of 3D video systems

3D videos are now a huge success due to the release of Avatar film in 2010. Besides its use in cinemas, applications that require some sort of 3D video transmission, such as internet and 3D TV is also receiving attention. 3D TV, for example, is a reality and the first 3D commercial channels are available.

For such sort of applications, an end-to-end 3D video system is subdivided into four main blocks: *3D Content Creation*, *3D Representation*, *Delivery* and *Visualization* (see Fig. 1).

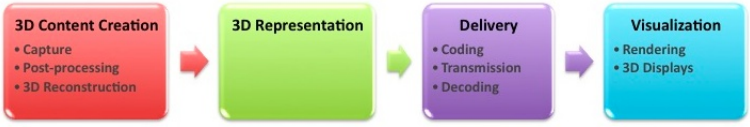

Figure 1. Pipeline of an end-to-end 3D video system.

The 3D Content Creation block (Fig. 2) is responsible for providing the data used to create the 3D video. The process starts at the Capture stage (Subsec. 3.1) with the choice of equipments that will be used to capture the scene and process data. Examples of devices for scene capture are 3D scanners, time-of-flight (TOF) sensors and digital cameras. The latter is the most widely used for capturing dynamic scenes, sometimes combined with other sensors. Other necessary equipments are computers, disks, grabber cards, etc. Projectors are also used in some systems

to improve the quality of captured data. The number of cameras in a setting varies and it depends on the application, as well as, its costs. For example, in literature we can find systems with more than 50 cameras [28] and also systems composed by only one camera and one projector [81].

After capture stage the data is sent to post-processing where low-level algorithms are applied to correct and improve data accuracy. For example, algorithms for color correction, correction of lens distortion and keystone distortion, camera calibration, features extraction and tracking, image rectification and alignment are within this stage. For explanations on these algorithms, we refer the reader to any Computer Vision book, such as the one in [75].

The processed data is sent to the 3D Reconstruction stage. The 3D reconstruction problem refers to the recovering of scene geometry, i.e., the 3D coordinates of objects that compose the scene. This stage is responsible for creating the data that will be used within the 3D video representation. Common techniques performed for geometry recovery are structure-from-stereo, shape-from-silhouette, structure-from-motion, shape-from-focus and defocus, as well as, shape-from-shading. In Subsection 3.2 we will discuss structure from stereo, structure from motion and shape from silhoettes techniques in the context of 3DV and FVV. Structure-from-stereo methods are the most popular in 3D videos literature and have been investigated by the MPEG group for standardization. Another research line on 3D reconstruction fuses data obtained from digital cameras and ToF sensors [29, 89].

A review of dynamic scenes capture can be found in [72].

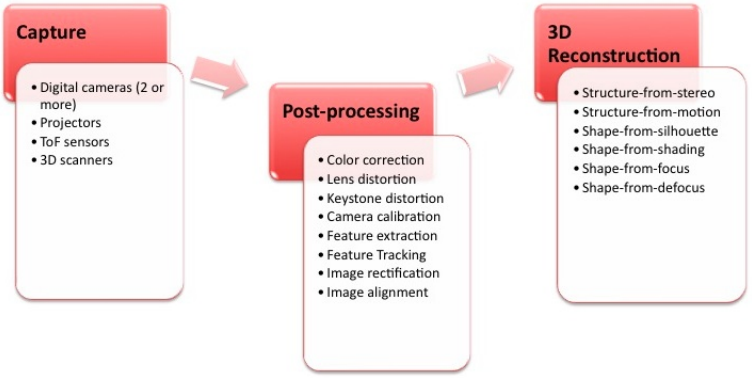

Figure 2. 3D Content Creation block. Sensors capture the scene and the acquired data is processed by low-level algorithms in Post-Processing stage. 3D reconstruction method is applied in order to create data that will be used by the 3D representation.

At the 3D Representation stage (Section 4) a format is chosen to store data from the 3D Content Creation block. There are a variety of 3D representation schemes in literature [64]. Its choice depends on the target application and capture devices. They can be classified in image-based (Subsec. 4.1), geometry-based (Subsec. 4.2) and a representation based on depth maps (Subsec. 4.3), which combines image and geometry aspects [63]. Geometry-based formats represent data as we know from Computer Graphics. They offer a full navigation of the scene or object, but it has realistic rendering issues due to errors in reconstruction step.

On the other hand, image-based formats avoid the explicit 3D reconstruction of the scene and provides a more realistic visualization. Depth-maps formats are more suitable for 3DV and FVV coding and has been investigated for standardization by the MPEG group.

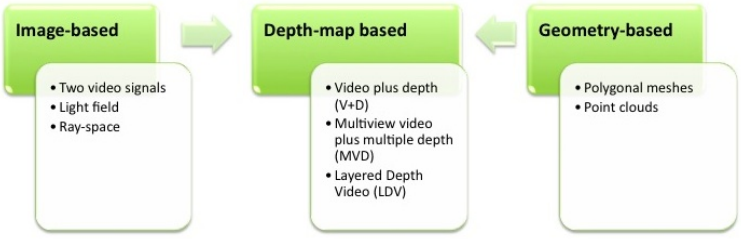

Figure 3. Categorization of 3D Representation formats.

The Delivey block is responsible for 3D video coding, transmission and decoding. Usually, it is necessary in applications with some type of network, such as Internet and 3D TV. Moreover, coding and decoding of 3D videos are important for development of storage media, e.g., Blu-ray discs. These are not in the scope of this text. We refer the reader interested in coding of 3D videos to the works in [64, 65, 80]. Readers interested in transmission and also storage of 3D videos are refered to references [22, 57]. A discussion about technologies to deliver 3D content to mobile devices can be found in [20].

The last building block of a 3D video system is the most important to the end user, because it deals with Visualization of the 3D content. It comprises Rendering stage and 3D Displays (Fig. 4). The Rendering stage 5.1 is responsible for employing algorithms to render the data stored at the representation format. The main focus is the view synthesis methods. They are necessary for free view point functionality and autoestereoscopic displays. More than others stages, this one is in charge of providing a realistic view of 3D dynamic scenes. Of course, its performance depends on several factors, such as the accuracy of the reconstructed data and data loss during transmission. In a 3D TV scenario it also depends on the receiver processing capability.

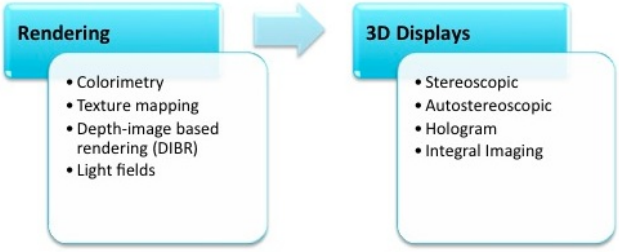

Figure 4. Visualization

3D Displays (Subsec. 5.2) are responsible for depth perception of stereoscopic videos. Also, for free-viewpoint videos they have to be able to provide means of interaction with the visualized content. 3D displays technologies are in constant development since 3D media became more accessible to home user. Specialists in consumer electronics predict that in 2015

more than 30% of all high-definition panels at home will be equipped with 3D capabilities. Stereoscopic videos technologies are mature and a huge success in cinemas, but there is room for improvement, specially regarding 3D displays. Stereoscopic displays are the most popular 3D display in the market, but in order to provide depth perception they require the use of uncomfortable glasses. To overcome this limitation, researches on autostereoscopic displays are under development. Autostereoscopic displays allow depth perception and FVV with no requirement of eyewear. Other types of 3D displays are holography and integral imaging. We refer readers interested in advances in holography and integral imaging to references [51] and [7], respectively.

3. 3D content creation

3.1. Capture

There are a variety of technologies for digitally acquiring the geometry of a 3D object. The choice of the acquisition setup strongly depends on the application, and of course, its costs. Digital cameras, 3D laser scanners and time-of-flight (TOF) sensors are the most popular devices for geometry and color acquisition.

An important laser scanner system has been presented in [55]. It utilizes a laser triangulation scanner and a high-resolution color camera to scan the 5m tall Davi, a sculpture of Michelangelo. Structured light scanners settings are composed by a projector and one or more cameras [3, 73]. In these systems a pattern is projected onto the object surface in order to improve the quality of the captured 3D object coordinates. In reference [9] the authors propose the scanning of 3D objects using a ToF camera. All systems cited above capture 3D information of static scenes. Figure 5 shows the simple acquisition setup utilized to capture the geometry of Parthenon sculptures [73].

Figure 5. Simple structured light scanner consisting of a digital camera, a projector and a tripod used in [73] and, on the right, a sculpture model obtained after 14 scans.

For dynamic scenes the most used devices are digital cameras. Systems with one or two cameras can be found in literature. For example, in reference [52] scene structure and motion

are retrieved using a hand-held camera and a real-time 3D system with a high-definition camera and a projector is presented in [81]. However, most settings utilize several digital cameras as in [26], for example. The concept of *3D video bricks* was introduced in [82]. One 3D video brick is composed by a projector, two black-and-white cameras and a high-definition color camera. The complete setting comprises multiple 3D video bricks.

Cameras can be arranged in a parallel or convergent setup (see Figure 6). One of the pioneering projects in this area is presented in reference [26]. The *3D Dome Studio* uses 51 cameras mounted on a 5m diameter dome and applies stereo techniques to reconstruct the shape of a moving object. The same techniques have been used in a circular setup with more than 30 cameras to shoot a football game. All cameras are sinchronized and pointing to the same target from different angles. The set of captured multi-view images are processed and a 3D model is reconstructed. In reference [6], 7 cameras where placed in a convergent fixed setup pointing to the center of the scene. Cameras are synchronized and calibrated. The main goal is the reconstruction and rendering of human bodys from any viewpoint and estimate its motion parameters. Another curved setting can be found in [40] where 12 cameras where placed at the ceiling, around the scene.

An example of parallel setup can be seen in [58]. It uses six consumer quality Fire-Wire video cameras aligned in two rows. Cameras where partitioned in stereo pairs, and every stereo pair is connected to one PC for stereo processing. The 3D system in [76] captures dynamic events with several cameras displaced in sequence and generates novel views with interpolation methods. Another example of parallel camera arrangement can be seen in [43].

Figure 6. Example of convergent setup with 51 cameras proposed in [26] (left) and a parallel one with 16 cameras in [43] (right).

All studio settings shown above use controled illumination to facilitate reconstruction processes. As a consequence, studio setups rigorously restricts the type of observed scene. In [8] authors use auto-exposure and gain changes compensation in order to capture outdoor scenes which has a large variation in illumination. The setup is portable and can be hold in a backpack or vehicle mount. It consists of a GPS, an inertial sensor and an omnidirectional camera, with six cameras within (see Fig. 7).

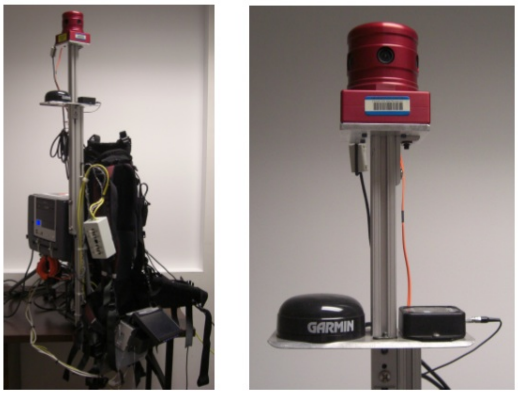

Figure 7. On the left the backpack mount. On the right the sensors head with GPS, inertial sensor and the omniderctional camera. Figures taken form [8]

Recently a new system configuration has been investigated. These 3D systems employs sensor fusion combining depth sensors and digital cameras [29, 90]. Their main goal is obtain more accurate depth maps by combining stereo methods and data acquired by depth sensors.

Commercial solutions for ease 3D acquisition are available. They are called stereo- or 3D cameras. Figure 8 shows the stereo camera Bumblebee XB3 from Point Grey Research and the full-HD professional 3D Panasonic AG-3DA1. Both cameras are available at Natalnet Laboratory. Bumblebee XB3 has a 3-sensor multi-baseline with variable resolutions and come with softwares for stereo processing. The Panasonic AG-3DA1 has integrated twin-lens and records and process synchronized left and right streams. The recorded channels are stored on memory cards in AVCHD format.

Figure 8. Bumblebee XB3 from Point Grey Research (left) and full-HD professional 3D Panasonic AG-3DA1 (right).

3.2. 3D reconstruction

After images are captured and pre-processed they are sent to the reconstruction stage. The 3D reconstruction problem refers to the recovering of scene geometry, i.e., the 3D coordinates of objects that compose the scene. This stage is responsible for creating the data that will be used within the 3D video representation.

3D video systems in literature differ on the employed reconstruction methods. Examples of such methods are shape from focus, shape from shading, structure from motion, shape from silhouette and structure from stereo. We refer the reader to any Computer Vision book [75] for a broad discussion about existing reconstruction methods. However, structure-from-stereo techniques have shown be more suitable for 3DV and FVV [68].

Here we will review some works of structure-from-stereo, structure-from-motion and shape-from-silhouette techniques within the context of 3D videos.

3.2.1. Structure from stereo

The most popular method of 3D reconstruction is stereo [78]. It is based in the principle of stereo vision (or stereopsis) which copes with the human visual system [38]. Because of the position of our eyes, our brain receives two views of a same scene from two slightly different viewpoints at the same horizontal level. Our brains fuse these two images and measure the disparity in order to estimate depth [38]. Computationally, stereo process has three main steps: selection of a particular location of the surface in one image (*feature extraction*); the selected location must be identified in the other image (*matching or correspondence problem*); the disparity in two correspondent locations must be computed (*reconstruction*) [38]. The process used to obtain 3D point coordinates from a set of known corresponding image locations is called *triangulation* [75, 78]. Overviews about the problem of recovering 3D structures from stereo can be found in literature [5, 13].

Over the years many efforts have been made by academics to compute stereo efficiently for static and dynamic events. The literature stereo is very extensive. In [56] an important work surveying and evaluating binocular stereo algorithms has been presented. The authors have categorized dense binocular stereo according to: matching cost computation, cost aggregation, disparity computation and disparity refinement.

Multiview stereo

For free-viewpoint video development it is mandatory the acquisition of images from many different viewpoints (see Fig. 9). Thus, the problem of reconstructing 3D scenes from more than 2 frames arises, the so-called multi-view stereo reconstruction problem [23].

Many algorithms to compute multi-view stereo has been developed [72]. A taxonomy for multi-view stereo methods has been proposed [59], similar to the one presented in [56] for binocular stereo methods evaluation. The multi-view algorithms are classified and evaluated according to six categories: scene representation, photoconsistency measure, visibility model, shape prior, reconstruction algorithm, initialization requirements. According to this taxonomy the reconstruction algorithms can be classified in four mais classes [59]:

- Cost computation on a 3D volume - for example, voxel coloring methods [60];
- Minimization of a cost function - for example, space carving methods [31];
- Computation of depth-maps;
- Extraction and matching of feature points.

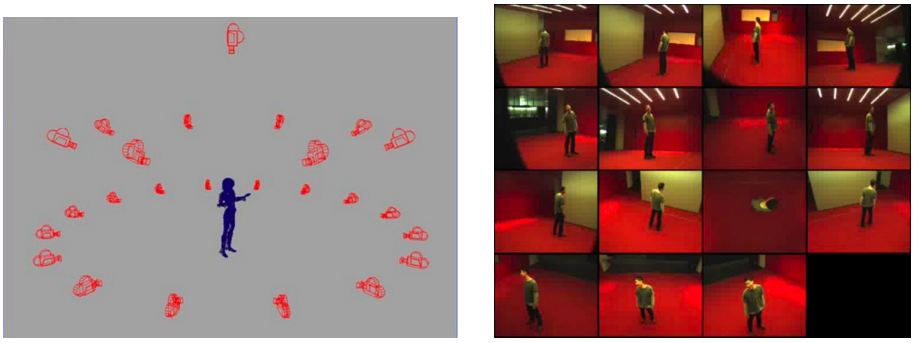

Figure 9. Example of multi-camera setup (left) and images of a same scene captured from many different viewpoints (right). Figures taken from [63]

In [18] the authors propose a new algorithm to implement multi-view stereo reconstruction by employing a pipeline other than Feature Extraction, Matching and Reconstruction as traditional stereo methods. It starts with a sparse set of matched points that are expanded to a more dense set and filtered using visibility constraints. This process results in a patch-based representation of the surface which is transformed into a mesh-based representation.

Multiview stereo algorithms have been applied to obtain 3D objects geometry from photos [36]. Also, many 3D video systems based on multiview stereo algorithms have been proposed [44, 82, 84, 92]. In the context of FVV one of the pioneering works can be seen in [26]. The authors use the multi-baseline stereo algorithm of [50] to obtain depth maps that are edited to remove innacuracys. It reconstructs fore- and back-ground objects. The system in [27] is also based on the same algorithm.

An recent overview of coding algorithms to stereo and multiview video can be found in [80].

Active stereo

The most difficult part in stereo computation is the matching or correspondence problem [38]. Active stereo methos try to overcome this limitation by emitting and projecting some sort of waves onto the surface. In structured light approaches a controlled illumination pattern is projected. This methodology has been applied to obtain 3D models of cultural artifacts, such as statutes [3, 34, 73].

Many 3DV and FVV systems benefits from this idea [26, 81–84, 92]. In [81], for example, a real-time 3D system is presented. It utilizes only one camera and one projector. They must be synchronized to guarantee that the projected pattern it will be projected at the same time the camera captures it. Camera and projector have to be calibrated, as well. The projector projects slides with a sequence of colored stripes and consecutive stripes may not have the same color(see Fig. 10). Experiments where made with static and also reasonably fast movements scenes. The system needs improvements on the quality of reconstructed scenes but it is a promising approach towards real-time 3D video system. Unlike the previous setting, the multi-view stereo system in [82] projects a binary vertical stripes pattern with randomly varying stripes width (see Fig. 10).

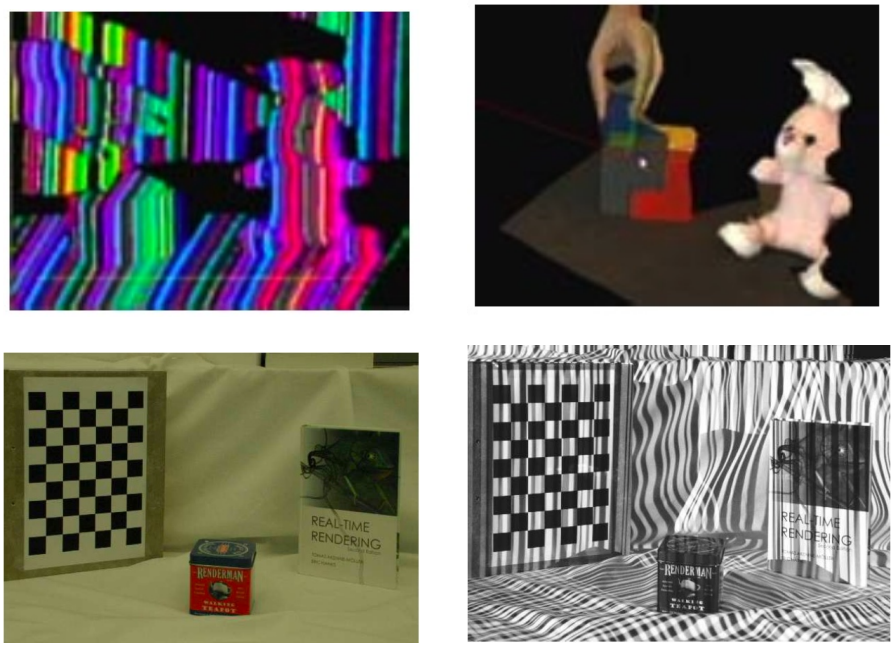

Figure 10. Upper row: scene illuminated with colored stripes (left) and the reconstructed scene (right) [81]. Lower row: color image (left) and same image with structured light illumination (right) [82].

Spacetime stereo

Methods to compute depth via triangulation have been widely investigated by the computer vision community. Stereo, laser scanning and time- or color- structured light are the most popular. Usually they are classified as active or passive methods. In [11] a new classification of the 3D reconstruction methods based on triangulation is proposed. Instead of passive or active approaches, the methods would be classified according to the domain where corresponding features are located. Techniques such as laser scanning and passive stereo identify features only in spacial domain. Methods such as time structured light use features only in temporal domain. The spacetime stereo approach looks for features in both spatial and temporal domains (see Fig 11). This new methodology has been applied for dynamic scenes reconstruction [10, 48].

In parallel, other research groups where also interested in spatio-temporal benefits. In reference [88] the authors have employed spacetime approach in three different cases. For static scenes they have used structured light to obtain high-quality depth maps and where observed improvements over traditional stereo methods. They have tested the spacetime theory in quasi-static objects such as waterfalls and it proved to be more efficient. For dynamic scenes under natural lighting conditions it behaved like traditional stereo. The approach presented in [88] have been used to develop a scalable 3D video system [83].

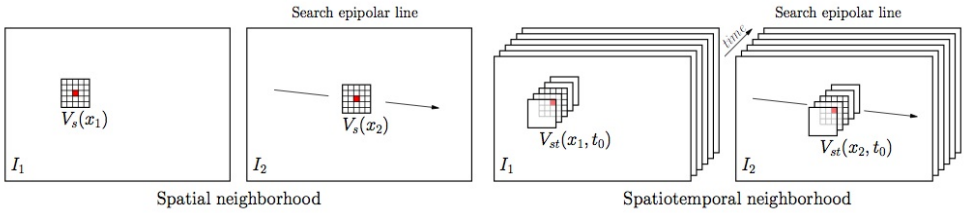

Figure 11. Spacetime principle. The search for correspondences in traditional stereo is done in spatial domain only (left). In spacetime stereo search for correspondences is done in spatial and temporal neighborhoods (right). Figure taken from [48].

In reference [62] the spacetime approach was used to improve the video resolution of dynamic scenes. The super-resolution is obtained simultaneasly in space and time and makes the system capable of recovering dynamic events that happens faster than video frame-rate.

3.2.2. Structure from motion

In Computer Vision, the problem of recovering the Structure From Motion (SFM) [75] refers to the process of finding the three dimensional structure of an object by analyzing its motion over time. We perceive a lot of information from the three dimensional structure of the environment by moving around. The same happens when the objects perform some movement in the scene.

The SFM problem is similar to stereo vision. In both approaches, the image correspondences and the 3D coordinates of the object must be computed. But in SFM, in order to find correspondences between images, features such as corners must be tracked from an image to another. The trajectories of these features are used to reconstruct the 3D object and the camera motion. Because of features tracking, SFM is especially effective with video sequences.

Most SFM techniques reconstructs scenes with rigid objects, but in [4, 77] the authors deal with scenes with non-rigid objects, such as animals and humans. A limitation of SFM is that the pixels correspondences can only be calculated accurately for salient features.

In [91] the authors use structure from motion to reconstruct statics scenes from a sequence of uncalibrated images. For such, a hand-held camera is used. They required restrict camera motion, specially camera rotations. No prior information is required besides the images themselves. One limitation is that it strongly depends on image texture because it is a feature based approach.

The reconstruction of 3D scenes captured by a hand-held camera was the main goal of other works[52, 54], as well. Structure from motion techniques were used to reconstruct citys architecture [53]. The authors try to fuse the data obtained by SFM approach and GPS measurements.

3.2.3. Shape from silhouette

Many algorithms of 3D reconstruction are based on object's silhouettes. This class of techniques are known as Shape-from-Silhouette [75]. The important concept of Visual

Hull of an object S was introduced in [32] to identify which parts of S are important to silhouette-based approaches. A formal definition is:

> " The visual hull $VH(S,R)$ of an object S relative to a viewing region R is a region of E^3 such that, for each point $P \in VH(S,R)$ and each viewpoint $V \in R$, the half-line starting at V and passing trough P contains at least a point of S." [32]

For each viewpoint V, the lines starting at V and passing trough P form a silhouette cone. The volume generated by intersecting all silhouette cones from all viewpoints V is the visual hull 12. Volume carving [31] is the approach commonly used for such. Since volumetric techniques are traditionally slow, an image-based visual hull (IBVH) [42] have been developed to overcome this limitation. It is real-time and like all image-based rendering technique it provides a realistic rendering of the scene. It is pertinent to observe that silhouettes approaches suffer from one important limitation: they are not able to distinguish concave surface regions. Thus, the reconstruction of concave objects is not guaranteed with silhouette approaches only. Efforts to overcome this problem have been made [17], as well.

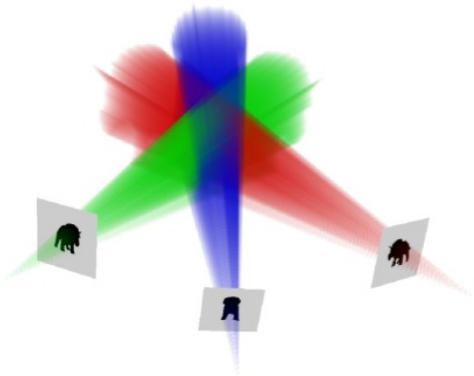

Figure 12. Intersection of silhouette cones. The result is the visual hull volume. Figure taken from [42]

In the context of 3DV and FVV silhouettes approaches have been widely used to recover the 3D object surface. The systems in [39, 40, 45, 67] employ the same volumetric strategy: the visual hull volume is computed, then it is divided in voxels. For each frame and viewing position all voxels are marked as occupied (object portion) or empty (background portion). After this process the remaining voxels contain the object and form a voxel-based representation of it. Finally, the marching cubes algorithm transforms the voxels model into a triangle mesh, which represents the object surface.

3D video systems using variants of IBVH have been already proposed [21, 85, 86]. Reference [37] presents a complete 3DV and FVV system combining visual hull, surface texture, image features and inertia constraints to perform a high quality reconstruction of dynamic scenes.

4. 3D video representation

Various representation schemes for 3D videos can be found in literature [64]. Usually its choice depends on the target application. But for some authors [63] it determines completely the 3D video system design.

3D scene representation formats can be classified in image- and geometry-based formats and also a hybrid representation based on depth maps (Subsec. 4.3), which combines image and geometry aspects [63].

Geometry-based modeling (Subsec. 4.2) represents data as we know from Computer Graphics. In order to use this format the 3D scene has to be reconstructed and the geometry stored in a well know format such as, polygonal meshes or point clouds. They offer a full navigation of the scene or object, but it has realistic rendering issues due to errors in reconstruction step. On the other hand, image-based formats (Subsec. 4.1) avoid the explicit 3D reconstruction of the scene and provides a more realistic visualization. But there is a critical trade-off between realistic rendering and size of stored data.

4.1. Image-based representation

The popular format of a three-dimensional video is a stereoscopic video composed by two video signals, one for each eye. It is the image-based format used by movie theaters and current 3D TV for home entertainment. Due to its simple format it can be encoded using existing video codecs, by performing spatial or temporal interleaving. For spatial interleaving the images for the right and left eye are resized and packed into a single frame. They can be arranged in side-by-side or top-bottom. In a temporal interleaving the right and left images are shown in alternate times.

For FVV systems exist Light fields [19, 33] and Ray-space [76] representations. Both representations do not perform any geometric reconstruction, avoiding the artifacts generated by this process. Thus, they lead to a more realistic rendering of the scenes. However, the realistic rendering is paid by the cost of the huge amount of necessary data. They need to store and transmit a set of views that are, at the receiver side, interpolated in order to generate novel views. If only a few views are transmitted the rendering quality is poor.

4.2. Geometry-based representations

4.2.1. Polygonal meshes

Polygonal meshes [16] are the most popular 3D scene representation in many industries such as architecture and entertainment. Due to realism requirements in computer graphics and the development of 3D scanning technologies, polygonal meshes representing 3D surfaces contain millions of polygons. On one hand they can represent satisfactorily almost any geometric detail of the surface. On the other hand these meshes are complex and computationally expensive to be stored, transmitted and rendered. To overcome these limitations, many techniques to compress and simplify complex meshes have been developed leading to progressive approaches [24], even for time-varying meshes [30].

Important projects that build 3D polygonal mesh models from scanner systems or photos have been proposed [3, 53]. Many developed 3DV and FVV systems are based on polygonal mesh representation [6, 37, 71]. In [39–41, 45, 67] a triangular mesh is obtained from a voxel representation via marching cubes algorithm, after silhouette-based reconstruction. In reference [37] instead of marching cubes algorithm the authors perform multi-level partition of unity implicits (MPLU) [49]. Reference [6] uses a prior body model consisting of 16 closed triangle meshes. Researchers in [39, 41] present a deformable three-dimensional mesh model which allows the recovery of the 3D shape and 3D motion. The shape is represented by the triangular mesh, while the movement by vertices translations. Deformations occur inter- and intra-frames, with photometric and smoothness constraints, for example. Figure 13 shows a result obtained after intra-frame deformation.

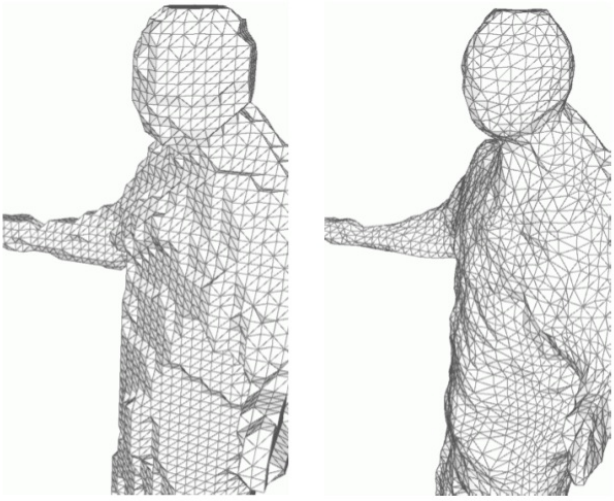

Figure 13. (Left) Mesh obtained from a voxel representation via marching cubes algorithm, after silhouette-based reconstruction. (Right) Mesh smoothed after intra-frame deformation. [39]

4.2.2. Point-based representation

In point-based schemes the geometry is represented by a set of points sampled from the surfaces in the scene [35]. Neither topological nor connectivity informations are explicitly stored. Points offer advantages over other representations because they are the simplest geometric primitive.

Progressive approaches have also been applied to point-based representations [34, 87]. The need arises in applications which deal with a huge amount of data and/or make some sort of data transmission, such as internet or broadcast. In [87] the 3D objects geometry and texture are encoded in terms of surface particles associated to an octree [16]. The encoding is done in an appropriate order which allows the surface be reconstructed progressively. The same idea has been employed to reconstruct and render the Davi statue [34]. In the last one a hierarchy

of spheres have been used instead of an octree and the resulting representation have been rendered using splatting techniques [55].

3D video systems claiming high-quality rendering of point-based representations are available in literature [81–83]. In [82] each point of the representation is associated with its color, avoiding the use of textures. Also, each point is modeled by a Gaussian ellipsoid generated by three vectors, with origin in its center. This is a probabilistic model representing the positional uncertainty of each point.

The authors in [86] propose a framework for recording 3D videos. The prototype have been tested to capture and reproduce dynamic scenes with one human in movement. They utilize a time-varying three-dimensional hierarchical point-based data structure to store the 3D video. One such data structure is constructed per frame. Then, two different splatting techniques are employed for rendering a continuous surface of the 3D object.

In [21] a point-based variant of image-based visual hull [42] is used in the design of an immersive environment for virtual design and collaboration. Authors of [85] propose a real-time free-viewpoint system based on the concept of 3D video fragments. 3D video fragments are point samples of a 3D object surface with some attributes, e.g., position, surface normal and color. It uses an inter-frame prediction scheme to dynamically update those attributes in order to avoid recompute the full 3D representation for each frame.

Comercials 3D video systems based on point representations are already available in the market. For example, Libero Vision Company [25] offers products for creating realistic virtual views for arbitrary viewpoints of sports .

4.3. Depth maps-based representation

Depth map [78] is a special case of digital image. In a depth map each pixel represents the distance from the sensor to a visible point at the scene. Thus, it reproduces the 3D scene structure and can be interpreted as a surface sampling .

Nowadays, representations based on depth maps are the most popular and promising representation for 3DV and FVV. This is due to the fact that some representations based on depth maps are able to perform at the same time 3DV coding - where the left and right images are encoded - and FVV coding - where view synthesis can be performed. Explanation on depth-image based representations and a recent review of 3D video representations using depth-maps are available in literature [1, 46].

In order to build a reliable 3D model a dense depth map must be established, that is, a depth estimate corresponding to each pixel in the intensity images. The pionnering works in [26, 27] computes dense depth maps from all available views. A scene description is created using the depth map aligned with the intensity image for each recorded angle. However, they convert each depth map into a triangle mesh and employs texture mapping for rendering the scene. This representation reproduces only free-viewpoint video and it is not capable of rendering stereoscopic videos. The work developed in [92] utilizes a layered representation - intensity image and associated depth map - for view interpolation. They also convert the depth maps into a triangle mesh to benefit from programmable GPUs.

A 3D representation that combines conventional 2D video stream with synchronized depth informations have been proposed in [12] during the ATTEST project [15]. It is called video plus depth (V+D) and it allows the rendering of two virtual views corresponding to a stereo pair. This format have been standartized by the MPEG group and it is known as MPEG-C Part 3 [68].

The video plus depth format has been extended to the multiview video plus depth (MVD) [69]. With this format only a subset of M images and its associated depth maps are transmitted to a display of N views. The remaining views are interpolated via image-based warping.

Another available format is layered depth video (LDV) [47]. It is based on the concept of layered-depth image (LDI) [61]. An LDV is composed by a 2D video (color image), the associated depth maps and other layers, for example, an occlusion layer or residual layers of depth and color. This representation is more compact than MVD. However, due to redundancy MVD format provides a more realistic rendering. Both formats are under investigation at MPEG group [68].

5. 3D video visualization

5.1. Rendering

Most developed 3D video systems aim to provide realistic visualization. The rendering technique employed strongly depends on the 3D representation used to model the 3D scene.

Popular approaches are texture mapping and colorimetry for surface-based representations, light fields [33] and depth-image based rendering (DIBR) [14]. Examples of FVV systems employing these rendering techniques can be found in [26, 43, 81]. Systems based on point-cloud representations usually apply splatting techniques [82, 86]. For 3DV rendering, video plus depth (V+D) representations achieve depth perception by performing DIBR techniques of the second video.

One important task of the rendering stage is to generate virtual views. This is important not only for FVV systems, but also for autostereoscopic displays. The general idea behind virtual view synthesis is to project the image into the 3D space and then project it again at a chosen virtual camera at the desired position. Inherent problems with this processing are occlusions and object boundaries areas. An occluded region in a natural view could be visible from a virtual view position, leading to holes at the novel view. Object boundaries areas are difficult to handle because they have back- and foreground colors. Also depth estimation of such areas are unreliable. Both situations lead to artifacts after projection into novel views.

The view interpolation schemes of MVD and LDV representations presented in [69] and [47], respectively, are good strategies to overcome these limitations. MVD identify unreliable regions by extracting a main and two boundaries layers - one for background boundaries and another for foreground boundaries (Fig. 14). Following layers extraction, they are projected into the 3D space and the virtual view position is interpolated from the original view positions trough spherical linear interpolation. After that, all layers in 3D space are projected separately in proper order and the results are merged. Finally, the artifacts naturally introduced by image-based 3D warping are detected and corrected.

LDV approach also identify unreliable regions and extract a main layer but, unlike MVD, it extracts only one boundary layer combining either back- and foreground boundaries (see Fig.14 for comparison). In LDV representations only a central view and associated residual layers are transmitted, leading to some color difference in novel views. Thus, MDV performs better than LDV regarding rendering aspects, but the latter is a more compact representation.

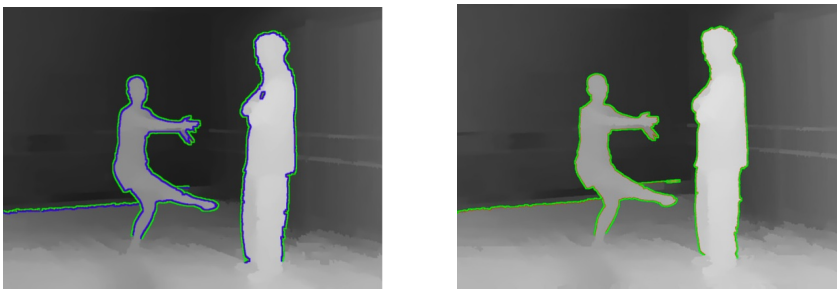

Figure 14. (Left) Layers of MVD: main layer in gray, foreground boundary layer in blue and background boundary layer in gree. (Right) Layers of LDV: main layer in gray and one boundary layer combining either back- and foreground boundaries

5.2. 3D displays

Mechanisms offering the perception of depth is a reality. 3D cinemas is experimenting huge success and 3D TV for home entertainment is now a reality. The popularization of 3D TV is due to advances in the whole 3D video pipeline, specially in 3D displays.

Examples of 3D displays are stereoscopic and autostereoscopic displays [79], holograms [51] and integral imaging [7]. Here we will briefly present the most intended for home entertainment: stereoscopic and autostereoscopic displays.

Stereoscopic displays are the most popular type of 3D display. It projects two multiplexed images at the screen. Both images show the same scene captured from two slightly different angles. A viewer needs to wear special glasses that separates the multiplexed image into two images - one for the left eye and one for the right eye. In particular, the glasses make each viewer's eye view only one of the two images. Schemes of images multiplexing rely on color, polarization or time multiplexing. Thus, the separation is possible because each image uses a different color (e.g., red and cyan), polarization or are projected in alternate frame sequencing. In each case, anaglyph, polarized or shutter glasses are required to send each image to the correspondent eye, respectively. The major drawback of this approach is that the viewer must to wear glasses for depth perception.

Autostereoscopic displays offer depth perception without the requirement of using any device such as special glasses or user-mounted devices. The main limitations of this technology are the cost and number of users able to perceive depth at the same time. Autostereoscopic displays are based on viewing areas the user should remain making one image to be visible

to the right eye and another to the left. It could be a two-view or multi-view display. In the first case only one stereo pair is displayed allowing 3DV capabilities. In the second, multiple stereo pairs are displayed and allows 3DV and FVV functionalities. Here, FVV is in the sense that when the observer moves in front of the display, he/she can perceive a natural motion parallax impression. Technologies employed in two-view autostereoscopic displays are parallax barrier and lenticular sheets. In the multi-view case, the performed methods are multiview parallax barrier, time multiplexing combined with parallax barrier and lenticular arrays combined with pixelated emissive displays.

An excellent discussion of underlying mechanisms of 3D displays is presented in [70]. We refer reader to references [2, 79] for reviews on 3D displays.

6. Conclusion

This chapter provides an overview of 3D videos production pipeline. We have concentrated in systems with no interest in 3D data coding and transmission. 3D video is a broad research area and here we outlined its main issues and advances briefly. An extensive list of publications is provided below for readers interested in more details.

3D media is already in our everyday lives and for this reason many leading researches are under development. Regarding capture devices, 3D cameras are already in the market, even for professional use. Still they are expensive. Although there are not many options for home users, they are becoming cheaper with development of new technologies.

Along with the quality of produced 3D content and advances in 3D displays, standardization plays an important role in 3D videos success. For such, MPEG group works on standardization of depth-maps based representations, which have shown be more suitable in this context. In parallel, the development of multiview autostereoscopic displays intend to make them the next generation of TV sets.

Author details

Lourena Rocha
Faculty of Federal University of Rio Grande do Norte, Departament of Applied and Exact Sciences, Caicó-RN, Brazil
PhD student at Federal University of Rio Grande do Norte, Departament of Computing Engeneering and Automation, Natalnet Laboratory, Natal-RN, Brazil

Luiz Gonçalves
Faculty of Federal University of Rio Grande do Norte, Departament of Computing Engeneering and Automation, Natalnet Laboratory, Natal-RN, Brazil

7. References

[1] Bayakovski, Y., Levkovich-Maslyuk, L., Ignatenko, A., Konushin, A., Timasov, D., Zhirkov, A., Han, M. & Park, I. K. [2002]. Depth image-based representations for static and animated 3d objects, *Image Processing. 2002. Proceedings. 2002 International Conference on*, Vol. 3, pp. III–25 – III–28 vol.3.

[2] Benzie, P. W., Watson, J., Surman, P., Rakkolainen, I., Hopf, K., Urey, H., Sainov, V. & von Kopylow, C. [2007]. A survey of 3dtv displays: Techniques and technologies, *IEEE Trans. Circuits Syst. Video Techn.* 17(11): 1647–1658.

[3] Bernardini, F., Rushmeier, H., Martin, I. M., Mittleman, J. & Taubin, G. [2002]. Building a digital model of michelangelo's florentine pieta, *Computer Graphics and Applications, IEEE* 22(1): 59–67.

[4] Brand, M. [2001]. Morphable 3d models from video, *IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 456–463.

[5] Brown, M. Z., Burschka, D. & Hager, G. D. [2003]. Advances in computational stereo, *IEEE Trans. Pattern Anal. Mach. Intell.* 25(8): 993–1008.
URL: *http://dx.doi.org/10.1109/TPAMI.2003.1217603*

[6] Carranza, J., Theobalt, C., Magnor, M. A. & Seidel, H.-P. [2003]. Free-viewpoint video of human actors.

[7] Cho, M., Daneshpanah, M., Moon, I. & Javidi, B. [2011]. Three-dimensional optical sensing and visualization using integral imaging, *Proceedings of the IEEE* 99(4): 556 –575.

[8] Clipp, B., Raguram, R., Frahm, J.-M., Welch, G. & Pollefeys, M. [n.d.]. A mobile 3d city reconstruction system, *Proceedings of IEEE Virtual Reality Workshop on Cityscape*.

[9] Cui, Y., Schuon, S., Chan, D., Thrun, S. & Theobalt, C. [2010]. 3d shape scanning with a time-of-flight camera, *The Twenty-Third IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2010, San Francisco, CA, USA, 13-18 June 2010*, IEEE, pp. 1173–1180.

[10] Davis, J., Nehab, D., Ramamoorthi, R. & Rusinkiewicz, S. [2005]. Spacetime stereo: A unifying framework for depth from triangulation, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27: 296–302.

[11] Davis, J., Ramamoorthi, R. & Rusinkiewicz, S. [2003]. Spacetime stereo: A unifying framework for depth from triangulation, *CVPR*, Vol. II, pp. 359–366.

[12] de Beeck, M. O. & Redert, A. [2001]. Three dimensional video for the home., *Proc. EUROIMAGE International Conference on Augmented, Virtual Environments and Three-Dimensional Imaging (ICAV3D'01), Mykonos, Greece*, pp. 188–191.

[13] Dhond, U. R. & Aggarwal, J. K. [1989]. Structure from stereo-a review, *Ieee Transactions On Systems Man And Cybernetics* 19(6): 1489–1510.
URL: *http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=44067*

[14] Fehn, C. [2004]. Depth-image-based rendering (dibr), compression, and transmission for a new approach on 3d-tv, *Proceedings of SPIE* 5291(2): 93–104.
URL: *http://link.aip.org/link/?PSI/5291/93/1&Agg=doi*

[15] Fehn, C., Kauff, P., de Beeck, M. O., Ernst, F., I Jssel-Steijn, W., Pollefeys, M., Gool, L. V., Ofek, E. & Sexton, I. [2002]. An evolutionary and optimised approach on 3d-tv, *Proceedings ofInternationalBroadcastConference*, pp. 357–365.

[16] Foley, J. D., van Dam, A., Feiner, S. K. & Hughes, J. F. [1995]. *Computer Graphics:Principles and Practice in C*, 2 edn, Addison-Wesley Publishing Company.

[17] Furukawa, Y. & Ponce, J. [2009]. Carved visual hulls for image-based modeling, *Int. J. Comput. Vision* 81(1): 53–67.
URL: *http://dx.doi.org/10.1007/s11263-008-0134-8*

[18] Furukawa, Y. & Ponce, J. [2010]. Accurate, dense, and robust multiview stereopsis, *IEEE Trans. Pattern Anal. Mach. Intell.* 32(8): 1362–1376.
URL: *http://dx.doi.org/10.1109/TPAMI.2009.161*

[19] Gortler, S. J., Grzeszczuk, R., Szeliski, R. & Cohen, M. F. [1996]. The lumigraph, *SIGGRAPH '96: Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, ACM, New York, NY, USA, pp. 43–54.

[20] Gotchev, A., Akar, G., Capin, T., Strohmeier, D. & Boev, A. [2011]. Three-dimensional media for mobile devices, *Proceedings of the IEEE* 99(4): 708 –741.

[21] Gross, M., Würmlin, S., Naef, M., Lamboray, E., Spagno, C., Kunz, A., Koller-Meier, E., Svoboda, T., Van Gool, L., Lang, S., Strehlke, K., Moere, A. V. & Staadt, O. [2003]. blue-c: a spatially immersive display and 3d video portal for telepresence, *SIGGRAPH '03: ACM SIGGRAPH 2003 Papers*, ACM, New York, NY, USA, pp. 819–827.

[22] Gü andrler, C., Gö andrkemli, B., Saygili, G. & Tekalp, A. [2011]. Flexible transport of 3-d video over networks, *Proceedings of the IEEE* 99(4): 694 –707.

[23] Hartley, R. I. & Zisserman, A. [2004]. *Multiple View Geometry in Computer Vision*, second edn, Cambridge University Press, ISBN: 0521540518.

[24] Hoppe, H. [1996]. Progressive meshes, *SIGGRAPH '96: Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, ACM, New York, NY, USA, pp. 99–108.

[25] *http://www.liberovision.com/* [n.d.].
URL: *http://www.liberovision.com/*

[26] Kanade, T., Narayanan, P. & Rander, P. [1995]. Virtualized reality: Concepts and early results, *IEEE Workshop on the Representation of Visual Scenes* .

[27] Kanade, T., Rander, P. & Narayanan, P. J. [1997]. Virtualized reality: constructing virtual worlds from real scenes, *Ieee Multimedia* 4(1): 34–47.
URL: *http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=580394*

[28] Kanade, T., Rander, P., Vedula, S. & Saito, H. [1999]. Virtualized reality: Digitizing a 3d time-varying event as is and in real time, *in* H. T. Yuichi Ohta (ed.), *Mixed Reality, Merging Real and Virtual Worlds*, Springer-Verlag, pp. 41–57.

[29] Kim, Y. M., Theobalt, C., Diebel, J., Kosecka, J., Micusik, B. & Thrun, S. [n.d.]. Multi-view image and tof sensor fusion for dense 3d reconstruction, *3DIM 2009*.

[30] Kircher, S. & Garland, M. [2005]. Progressive multiresolution meshes for deforming surfaces, *Proceedings of the 2005 ACM SIGGRAPH/Eurographics symposium on Computer animation*, ACM, New York, NY, USA, pp. 191–200.

[31] Kutulakos, K. N. & Seitz, S. M. [2000]. A theory of shape by space carving, *Int. J. Comput. Vision* 38(3): 199–218.

[32] Laurentini, A. [1994]. The visual hull concept for silhouette-based image understanding.

[33] Levoy, M. & Hanrahan, P. [1996]. Light field rendering, *in* ACM (ed.), *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, New York, NY, USA, pp. 31–42.

[34] Levoy, M., Pulli, K., Curless, B., Rusinkiewicz, S., Koller, D., Pereira, L., Ginzton, M., Anderson, S., Davis, J., Ginsberg, J., Shade, J. & Fulk, D. [2000]. The digital michelangelo project: 3D scanning of large statues, *Proceedings of ACM SIGGRAPH 2000*, pp. 131–144.

[35] Levoy, M. & Whitted, T. [1985]. The use of points as a display primitive, *Technical report*, University of North Carolina, Chapel Hill.

[36] Li, Y., Shum, H.-Y., Tang, C.-K. & Szeliski, R. [2004]. Stereo reconstruction from multiperspective panoramas, *IEEE Trans. Pattern Anal. Mach. Intell.* 26(1): 45–62.
URL: *http://dx.doi.org/10.1109/TPAMI.2004.1261078*

[37] Liu, Y., Dai, Q. & Xu, W. [2009]. A wide base line multiple camera system for high performance 3d video and free viewpoint video.

[38] Marr, D. & Poggio, T. [1979]. A computational theory of human stereo vision, *Proceedings of the Royal Society of London. Series B, Biological Sciences* 204(1156): 301–328.

[39] Matsuyama, T. [2004]. Exploitation of 3d video technologies, *Proceedings of the International Conference on Informatics Research for Development of Knowledge Society Infrastructure*, ICKS '04, IEEE Computer Society, Washington, DC, USA, pp. 7–14.
URL: *http://dx.doi.org/10.1109/ICKS.2004.10*

[40] Matsuyama, T. & Takai, T. [2002]. Generation, visualization, and editing of 3d video, *3DPVT02*, pp. 234–245.

[41] Matsuyama, T., Wu, X., Takai, T. & Nobuhara, S. [2004]. Real-time 3d shape reconstruction, dynamic 3d mesh deformation, and high fidelity visualization for 3d video, *Computer Vision and Image Understanding* 96(3): 393–434.

[42] Matusik, W., Buehler, C., Raskar, R., McMillan, L. & Gortler, S. [2000]. Image-based visual hulls, *SIGGRAPH 2000*.

[43] Matusik, W. & Pfister, H. [2004]. 3d tv: a scalable system for real-time acquisition, transmission, and autostereoscopic display of dynamic scenes, *SIGGRAPH '04: ACM SIGGRAPH 2004 Papers*, ACM, New York, NY, USA, pp. 814–824.

[44] Min, D. B., Kim, D., Yun, S. & Sohn, K. [2009]. 2d/3d freeview video generation for 3dtv system, *Sig. Proc.: Image Comm.* 24(1-2): 31–48.

[45] Moezzi, S., Tai, L.-C. & Gerard, P. [1997]. Virtual view generation for 3d digital video, *IEEE MultiMedia* 4(1): 18–26.

[46] Mü andller, K., Merkle, P. & Wiegand, T. [2011]. 3-d video representation using depth maps, *Proceedings of the IEEE* 99(4): 643 –656.

[47] Muller, K., Smolic, A., Dix, K., Kauff, P. & Wiegand, T. [2008]. Reliability-based generation and view synthesis in layered depth video, *Multimedia Signal Processing, 2008 IEEE 10th Workshop on*, pp. 34 –39.

[48] Nehab, D. [2007]. *Advances in 3D Shape Acquisition*, PhD thesis, Princeton University.

[49] Ohtake, Y., Belyaev, A., Alexa, M., Turk, G., Seidel, H.-P. & Saarbrücken, M. [2003]. Multi-level partition of unity implicits, *ACM Transactions on Graphics* 22: 463–470.

[50] Okutomi, M. & Kanade, T. [1993]. A multiple-baseline stereo, *IEEE Trans. Pattern Anal. Mach. Intell.* 15(4): 353–363.
URL: *http://dx.doi.org/10.1109/34.206955*

[51] Onural, L., Yaraş and, F. & Kang, H. [2011]. Digital holographic three-dimensional video displays, *Proceedings of the IEEE* 99(4): 576 –589.

[52] Pollefeys, M., Gool, L. V., Vergauwen, M., Verbiest, F., Cornelis, K., Tops, J. & Koch, R. [2004]. Visual modeling with a hand-held camera.

[53] Pollefeys, M., Nistér, D., Frahm, J.-M., Akbarzadeh, A., Mordohai, P., Clipp, B., Engels, C., Gallup, D., Kim, S.-J., Merrell, P., Salmi, C., Sinha, S., Talton, B., Wang, L., Yang, Q., Stewénius, H., Yang, R., Welch, G. & Towles, H. [2008]. Detailed real-time urban 3d reconstruction from video.

[54] Pollefeys, M., Vergauwen, M., Cornelis, K., Tops, J., Verbiest, F. & Van Gool, L. [2001]. Structure and motion from image sequences, *PROC. CONF. ON OPTICAL 3-D MEASUREMENT TECHNIQUES* pp. 251–258.

[55] Rusinkiewicz, S. & Levoy, M. [2000]. Qsplat: A multoresolution point rendering system for large meshes, *Proc. of ACM SIGGRAPH*.

[56] Scharstein, D. & Szeliski, R. [2002]. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms.

[57] Schierl, T. & Narasimhan, S. [2011]. Transport and storage systems for 3-d video using mpeg-2 systems, rtp, and iso file format, *Proceedings of the IEEE* 99(4): 671 –683.

[58] Schirmacher, H., Li, M. & peter Seidel, H. [2001]. On-the-fly processing of generalized lumigraphs, *EUROGRAPHICS 2001*, pp. 165–173.

[59] Seitz, S. M., Curless, B., Diebel, J., Scharstein, D. & Szeliski, R. [2006]. A comparison and evaluation of multi-view stereo reconstruction algorithms, *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 1*, CVPR '06, IEEE Computer Society, Washington, DC, USA, pp. 519–528.
URL: *http://dx.doi.org/10.1109/CVPR.2006.19*

[60] Seitz, S. M. & Dyer, C. R. [1997]. Photorealistic scene reconstruction by voxel coloring, *CVPR '97: Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition (CVPR '97)*, IEEE Computer Society, Washington, DC, USA, p. 1067.

[61] Shade, J., Gortler, S. J., wei He, L. & Szeliski, R. [1998]. Layered depth images, *SIGGRAPH '98*.

[62] Shechtman, E., Caspi, Y. & Irani, M. [2002]. Increasing space-time resolution in video, *ECCV '02: Proceedings of the 7th European Conference on Computer Vision-Part I*, Springer-Verlag, London, UK, pp. 753–768.

[63] Smolic, A. [2010]. 3d video and free viewpoint video—from capture to display, *Pattern Recognition* 44(9): 1958–1968.
URL: *http://linkinghub.elsevier.com/retrieve/pii/S0031320310004450*

[64] Smolic, A., Alatan, A., Yemez, Y., Gudubkay, U., Zabulis, X., Mueller, K., Erdem, C. E. & Weigel, C. [2007]. Scene representation technologies for 3dtv - a survey.

[65] Smolic, A. & Kauff, P. [2005]. Interactive 3-d video representation and coding technologies, *Proceedings of IEEE*, number 1, pp. 98–110.

[66] Smolic, A., Kauff, P., Knorr, S., Hornung, A., Kunter, M., Mü andller, M. & Lang, M. [2011]. Three-dimensional video postproduction and processing, *Proceedings of the IEEE* 99(4): 607 –625.

[67] Smolić, A., Mueller, K., Merkle, P., Rein, T., Kautzner, M., Eisert, P. & Wieg, T. [2004]. Representation, coding, and rendering of 3d video objects with mpeg-4 and h.264/avc.

[68] Smolic, A., Mueller, K., Merkle, P. & Vetro, A. [2009]. Development of a new mpeg standard for advanced 3d video applications, *Image and Signal Processing and Analysis, 2009. ISPA 2009. Proceedings of 6th International Symposium on*, pp. 400 –407.

[69] Smolic, A., Muller, K., Dix, K., Merkle, P., Kauff, P. & Wiegand, T. [2008]. Intermediate view interpolation based on multiview video plus depth for advanced 3d video systems, *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*, pp. 2448 –2451.

[70] Son, J., Javidi, B. & Kwack, K. [2006]. Methods for displaying three-dimensional images, 94(3): 502–523.

[71] Starck, J., Maki, A., Nobuhara, S., Hilton, A. & Matsuyama, T. [2009]. The multiple-camera 3-d production studio, *IEEE Trans. Cir. and Sys. for Video Technol.* 19(6): 856–869.
URL: *http://dx.doi.org/10.1109/TCSVT.2009.2017406*

[72] Stoykova, E., Alatan, A., Benzie, P., Grammalidis, N., Malassiotis, S., Ostermann, J., Piekh, S., Sainov, V., Theobalt, C., Thevar, T. & Zabulis, X. [2007]. 3d time-varying scene capture technologies - a survey, *IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on Multi-view Video Coding and 3DTV* 17(11): 1568–1586.

[73] Stumpfel, J., Tchou, C., Yun, N., Martinez, P., Hawkins, T., Jones, A., Emerson, B. & Debevec, P. [2003]. Digital reunification of the parthenon and its sculptures, *4th International Symposium on Virtual Reality, Archeology and Intelligent Cultural Heritage*, Brighton, UK.

[74] Su, G.-M., Lai, Y.-C., Kwasinski, A. & Wang, H. [2011]. 3d video communications: Challenges and opportunities, *Int. J. Commun. Syst.* 24(10): 1261–1281.
URL: *http://dx.doi.org/10.1002/dac.1190*

[75] Szeliski, R. [2010]. *Computer Vision : Algorithms and Applications*, Vol. 5 of *Texts in Computer Science*, Springer-Verlag New York Inc.

[76] Tanimoto, M. [2006]. Overview of free viewpoint television, *Signal Processing Image Communication* 21(6): 454–461.
URL: *http://linkinghub.elsevier.com/retrieve/pii/S0923596506000166*

[77] Torresani, L., Yang, D. B., Alexander, E. J. & Bregler, C. [2001]. Tracking and modeling non-rigid objects with rank constraints, *IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 493–500.

[78] Trucco, E. & Verri, A. [1998]. *Introductory techniques for 3-D computer vision*, Prentice Hall.

[79] Urey, H., Chellappan, K., Erden, E. & Surman, P. [2011]. State of the art in stereoscopic and autostereoscopic displays, *Proceedings of the IEEE* 99(4): 540 –555.

[80] Vetro, A., Wiegand, T. & Sullivan, G. [2011]. Overview of the stereo and multiview video coding extensions of the h.264/mpeg-4 avc standard, *Proceedings of the IEEE* 99(4): 626 –642.

[81] Vieira, M., Sa, A., Velho, L. & Carvalho, P. C. [2005]. A camera-projector system for real-time 3d video, *Proceeedings of PROCAMS*.

[82] Waschbüsch, M., Würmlin, S., Cotting, D. & Gross, M. [2007]. Point-sampled 3d video of real-world scenes, *Image Commun.* 22(2): 203–216.

[83] Waschbusch, M., Würmlin, S., Cotting, D., Sadlo, F. & Gross, M. [2005]. Scalable 3d video of dynamic scenes.

[84] Waschbüsch, M., Würmlin, S. & Gross, M. H. [2007]. 3d video billboard clouds, *Comput. Graph. Forum* 26(3): 561–569.

[85] Würmlin, S., Lamboray, E. & Gross, M. H. [2004]. 3d video fragments: dynamic point samples for real-time free-viewpoint video, *Computers & Graphics* 28(1): 3–14.

[86] Wurmlin, S., Lamboray, E., Staadt, O. G. & Gross, M. H. [2002]. 3d video recorder, *Proceedings of Pacific Graphics*.

[87] Yemez, Y. & Schmitt, F. [1999]. Progressive multilevel meshes from octree particles, *Proc. of Second International Conference on 3D Imaging and Modeling*, IEEE Computer Society, Los Alamitos, CA, USA.

[88] Zhang, L., Curless, B. & Seitz, S. M. [2003]. Spacetime stereo: Shape recovery for dynamic scenes, *Proc. Computer Vision and Pattern Recognition Conf. (CVPR)*.

[89] Zhu, J., Wang, L., Yang, R. & Davis, J. [2008]. Fusion of time-of-flight depth and stereo for high accuracy depth maps, *CVPR*.

[90] Zhu, J., Wang, L., Yang, R., Davis, J. & Pan, Z. [2011]. Reliability fusion of time-of-flight depth and stereo geometry for high quality depth maps, *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 33(7): 1400 –1414.

[91] Zisserman, A., Fitzgibbon, A. & Cross, G. [1999]. Vhs to vrml: 3d graphical models from video sequences, *Proceedings of the IEEE International Conference on Multimedia Computing and Systems - Volume 2*, ICMCS '99, IEEE Computer Society, Washington, DC, USA, pp. 9051–.
URL: *http://dx.doi.org/10.1109/MMCS.1999.779119*

[92] Zitnick, C. L., Kang, S. B., Uyttendaele, M., Winder, S. & Szeliski, R. [2004]. High-quality view interpolation using a layered representation, *SIGGRAPH '04*, Vol. 23.