D28.2 Technical Report # 2 on 3D Time-varying Scene Representation Technologies

Project Number: 511568
Project Acronym: 3DTV
Title: Integrated Three-Dimensional Television – Capture, Transmission and Display

Deliverable Nature: R
Number: D28.2
Contractual Date of Delivery: M29
Actual Date of Delivery: M30
Report Date: 16 February 2007
Task: WP8
Dissemination level: PU
Start Date of Project: 01 September 2004
Duration: 48 months
Organisation name of lead contractor for this deliverable: METU

Name of responsible: A. Aydın Alatan (alatan@eee.metu.edu.tr)

Editor: A. Aydın Alatan (alatan@eee.metu.edu.tr)

16 February 2007

3D Time-varying Scene Representation Technologies
TC1 WP8 Technical Report 2

EDITOR

A. Aydın Alatan (**METU**)

Contributing Partners to Technical Report:

Bilkent University (**Bilkent**)
Fraunhofer Gesellschaft zur Förderung der angewandten Forschung e.V. (**FhG-HHI**)
Institute of Media Technology, Technische Universitaet Ilmenau (**UIL**)
Informatics and Telematics Institute, Centre for Research and Technology Hellas (**ITI-CERTH**)
Koç University (**KU**)
Middle East Technical University (**METU**)
Momentum Bilgisayar Yazılım, Danışmanlık, Ticaret A.Ş. (**Momentum**)
University of West Bohemia in Plzen (**Plzen**)
University of Hannover (**UHANN**)
University of Tuebingen (**UNI-TUEBINGEN**)

REVIEWERS

Ugur Güdükbay (**Bilkent**)

Christian Weigel (**UIL**)

Xenophon Zaboulis (**ITI-CERTH**)

Tanju Erdem (**Momentum**)

Project Number: 511568
Project Acronym: 3DTV
Title: Integrated Three-Dimensional Television – Capture, Transmission and Display

TABLE OF CONTENTS

Executive Summary

This report summarizes the research efforts in EC IST 3DTV Network of Excellence (NoE) under Work Package 8 (WP8), entitled *3-D Scene Representation Technologies,* during Year-II of this project. The technical report consists of 20 novel research manuscripts, out of which 6 of them are obtained as a result of joint research activities between WP8 contributors. As well as these 20 publications, the technical report also includes an in-depth analysis over the results of the presented research activities, in terms of the applicability of the proposed technologies for the upcoming 3DTV systems. Noting that *3-D scene representation* can be simply defined as the description of the observed scenery in terms of geometric primitives, such as points, surfaces or volumes, the outline of the report follows a similar categorization. Special attention is also devoted to the 3-D representations of humans, as well as some specific physically-modeled objects, under separate sections. Finally, research results about the representations, which implicitly utilize 3-D information, are also examined under pseudo-3D representation category.

In this report, point-based scene representation technologies are examined along two different paths, one of which utilizes cameras from different angles of the scene to extract dense depth information at each *point* on the reference image. Such a dense representation is mostly utilized to remove the visual redundancy in different views of the scene to obtain better compression for multi-view content, rather than describing full 3-D geometry (Section 3.1.1). The other point-based representation approach, inputs 3-D points from active range-finding devices, such as laser scanners, and focuses on Bayesian modeling of this 3-D point-cloud for noise removal and hole-filling (Section 3.1.2). Both approaches give quite promising result in their respective application domains.

3-D surface representation research efforts of this report mostly focus on the generation of time-consistent (fixed-connectivity) wireframe representations, which are preferable for rate-efficiency (compression). In one of such research actions, the wireframe model is obtained through minimization of a cost function for active contours by considering the silhouettes of these models on the images (Section 3.2.1). In a different effort, a novel similarity measure between two mesh representations at different time-instants is also proposed by defining 4-D static tetrahedral mesh structures (Section 3.2.2). For obtaining time-consistent mesh representations between dynamic meshes at different time-instants, scene flow is utilized to propagate mesh-connectivity between different instants and re-meshing techniques are utilized to obtain time-consistent mesh representations (Section 3.2.4). A complete multi-camera dialogue system, utilizing such fixed-connectivity wireframe representations is also presented, which clearly signify the opportunities for achieving real-life applications of 3DTV systems by the help of such representations (Section 3.2.3).

Volumetric representations are mostly utilized to describe the outputs of scene extraction methodologies, which aim to determine 3-D information in the predefined volumetric primitives at the scene. In one of these contributions, a detailed comparison between such representations against a mesh-based description is presented in terms of visual and geometric reconstruction quality (Section 3.3.1). The compared volumetric approach is a novel plane-sweep based volumetric extraction and representation methodology and it is also applied to self-calibrated video (Section 3.3.2), as well as multi-view data (Section 3.3.4), all of which result with quite precise 3-D extraction results. The partitioning of 3-D space via planes is further extended into sphere-sweeping idea, in order to improve the precision of 3-D estimates, especially in non-planar surfaces (Section 3.3.5). Finally, utilization of tetrahedral

primitives, as volumetric representations, is examined by two different multi-resolution strategies to result with efficient representation for any complex descriptions (Section 3.3.6).

The research efforts on human face and body are mostly focused on bridging the gap between real observations and artificial visualizations. In these activities, the measurements, which are captured from the real scenes, are fed to the synthetic animations for obtaining more realistic temporal actions. In one of these approaches, the head movements and speech prosody of a human are both analyzed to jointly model these two different modalities via HMMs, so that the head movements of a typical synthetic animation could be made more realistic from the information inferred through the analysis of its speech content (Section 3.4.1). In another research action, the redundancy and jitter-noise in the measured data during motion capture from real life, is removed by fitting parametric curves to these measurements (Section 3.4.2).

The object-specific representation research the interaction between synthetically generated fluids and deformable objects are examined by unifying the approaches in Smoothed Particle Hydrodynamics (SPH) and mass-spring systems (Sections 3.5.1 and 3.5.2). Typical impressive results of this research are available on 3DTV webpage.

In the final group of representations, generation of the images of a scene from arbitrary viewing angles could be achieved by pseudo-3D representation techniques. By utilizing trifocal tensors, the location of a pixel in the third image for an arbitrarily selected view, could be obtained from the images of two other viewing angles (Section 3.6.2).

This report presents the promising results of research collaboration between contributors of WP8 on one of the fundamental technologies in 3DTV systems. These efforts will continue towards these directions in Year III, in order to determine preferable 3-D scene descriptions for 3DTV systems.

1. Introduction

State-of-the-art Survey Report on 3-D Scene Representation Technologies has clearly shown that there are different paths to pursue for the 3DTV systems for their representation technologies. The initial research activities for 3DTV NoE on these paths are presented in *Technical Report #1*. These initial efforts are further improved with some recent results, as well as some new research directions, which has been started from this second report.

In a typical 3DTV system, the pursued 3-D scene representation technologies should consider the following requirements:

- *Applicability* of the representation to the result (output) of the preferred 3-D scene extraction methodology
- *Generality* of the representation to make it possible to be utilized in any 3-D scene structure
- *Accuracy and perceptual quality* of the representation with respect to the real 3-D scene
    - Quality of the representation to yield perceptually pleasant and "real-looking" results
    - Level of Detail (*LoD*) scalability for obtaining results with various resolutions
- *Achievability of the transmission bandwidth* by the given representation
    - Efficiency of the utilized representation, resulting with better compression of the data
    - Progressive structure of the bit-stream for obtaining a version of represented structure, even if limited representation information is available
- *Editing and interaction availability* of the representation for some application scenarios
- *Compatibility* of the representation with the available 3-D display properties and specifications.
- *3-D perceptibility* of the representation on the 3-D display by the viewers.

Among the above requirements, two of them require further attention, while the first one puts a constraint for obtaining a subjectively pleasing (real-looking) outputs, as most of the research in computer graphics scientific community is focused on, whereas the second requirement forces the description to be transmittable, which will be strictly required in a 3DTV system during its operation.

All the research initiatives in this report try to fulfill the above requirements from different aspects. Obviously, for the synthetically generated content, subjective quality requirement becomes more dominant, since the major obstacle is to obtain realistic-looking outputs through the utilized representations. On the other hand, for multi-view video, which is expected to be the pioneering content-type of first-generation 3DTV systems, rate-distortion efficiency of the overall representation is more critical.

2. Analysis of the Results Reported in Publications

The technical report starts with two contributions for *point-based representation* technologies in Sections 3.1.1 and 3.1.2. The proposed dense depth representation in Section 3.1.1 jointly considers the extraction of 3-D scene information and compression of the resulting representation. In other words, the scene description is extracted in such a way that the representation is easy to compress. The proposed representation has a promising application in multi-view content-based 3DTV systems, by utilizing the 3-D information to remove the visual redundancy from multi-view content. The research effort is also an important step to bridge the gap between extraction and compression technologies in 3DTV systems.

In Section 3.1.2, Bayesian formulation of noise removal problem on a 3-D point set is given with some a priori models on the point clouds, such as smoothness between neighboring points. The experimental results indicate that it is possible to remove significant amount of errors (due to the inevitable noise in active sensing device), as well as completing some unavailable parts of the 3-D point data set by the help of such a formulation. As a future application, it could also be interesting to use the same formulation to remove outliers of the 3-D structural estimates, obtained for different 3-D scene extraction methods. In other words, the presented method is also applicable to remove outliers from 3-D structure estimates, which are obtained through visual data based passive (*shape-from-X*) approaches.

The contributions for *surface-based representations* are presented in Sections 3.2.1, 3.2.2, 3.2.3 and 3.2.4. All the surface-based representation research contributions in this report mainly focus on fixed-connectivity meshes, since this approach is quite promising due to the efficient compression of the resulting representation in time. In Section 3.2.1, the initial efforts for determining a time-consistent mesh is given, starting from a 3-D active closed surface, which fits on the silhouettes of the 3-D structure to obtain its final structure through iterations. The connectivity of this initially estimated mesh is to remain fixed in time in the planned joint future research, while the approach will update the position of the vertices based on the time-varying silhouette information.

The position update for a time-consistent mesh is also examined in another research effort in Section 3.2.4. In this research, 3-D scene flow of the vertices of an initial mesh is obtained by using tracked 2-D correspondences between frames. Hence, given a fixed-connectivity initial mesh, the vertices of this mesh are obtained by using conventional 2-D corner trackers for video. The research results on synthetic data are quite promising, whereas the experimental results on real data will follow in their future work.

Comparison between different meshes is crucial to asses the quality of any mesh at any time instant and a novel approach is proposed in Section 3.2.2 to fulfill this goal. In this approach, the compared 3D dynamic meshes are converted into a 4D static tetrahedral mesh. Then, the Hausdorff distance is adopted for this 4D static mesh to determine the distance between meshes. The proposed novel approach has not only applications in dynamic mesh comparison, but also in some other areas, as well, such as 3D search engines for shape-based retrieval and analysis.

Finally, a complete 3D dialogue system, based on the mesh representation is presented in Section 3.2.3. In this multi-camera setup with 20+4 cameras, 3-D information of the dynamic

objects are obtained from shape-from-silhouette method with foreground-background subtraction is achieved by chroma-keying. After mesh-generation, this representation is updated with the method in Section 3.2.2. This effort is important in the sense that it clearly shows the applicability of mesh-based representations for 3DTV applications.

The research efforts for *volumetric representations* follow two different paths, one of which describes a 3-D scene by defining small patches (planes) within each volumetric primitive. The position and orientation of these planes are precisely determined by using the texture uniqueness property between different images (views) of the scene (i.e.., correct position and orientation of the plane should yield good correlation between the consecutive locations in the image pair). Although, the algorithm not only determines the position, but also the orientation of the plane, only the position (i.e. *z-distance*) of these planes are being utilized in the resulting description for each volumetric element, which have intersections with the observed surface.

In Section 3.3.1, the representation performance of this textured-voxel approach is compared against a 3-D wire-frame-based approach in terms of backprojections of the reconstructed scene to the views and it is observed that the textured-voxel representation has a clear improvement over the wire-frame representation in terms of the backprojection error. The same representation is shown to be applicable for the cases, in which the internal calibration of the imaging system is not available. Self-calibration of the system makes it possible to obtain precise representation, even for uncalibrated imaging setups (Section 3.3.2). The extension of this representation for multi-view content is also achieved and tested via simulations in Section 3.3.4. The simulation results clearly indicate that the proposed method is also applicable to multi-view data. The aforementioned volumetric representation is also extended to non-planar patches in Section 3.3.5. Instead of plane sweeping, the observed scene is swept via spherical surfaces, which give fair resolution accuracy on different viewing angles. The simulation results enjoy an improvement in 3-D reconstruction accuracy, especially on the non-planar surfaces in the scene. The approach based on texture-uniqueness is a novel approach with a promising 3-D extraction and representation performance, whereas it is still immature to be considered in 3DTV systems.

The other pursued path in *volumetric representations* is about utilizing tetrahedral primitives for representing 3-D scenes. The approach in Section 3.3.6, applies a hierarchical description strategy, beginning from a coarse tetrahedral element, which is refined through within consecutive resolutions by finer primitives. The simulation results indicate an efficient, hierarchical, GPU-friendly algorithm, yielding high-quality simplified meshes.

The joint research activities related to the *representation of human face and body* are focused on generation of natural-looking movements of head and body, after capturing and modeling the temporal behavior of humans. The representation of a synthetically-generated human face might lack of realistic temporal movements (head gestures). In order to overcome this problem, in Section 3.4.1, a novel approach is proposed to relate temporal variations in pitch frequency and speech intensity to the typical head movements of a speaker via HMMs. Hence, it is possible to generate natural gestures of a synthetic head, from its speech content by utilizing the model in between.

In a different effort to efficiently capture human activities, the captured marker coordinates in key-frames are modeled parametrically by a Hermite curve to remove the redundancy in these marker positions (Section 3.4.2). According to the simulation results, such a representation

not only yields natural-looking body motions, but also becomes very efficient for storage and transmission purposes.

In a joint 3-D scene extraction and representation research effort, 3-D pose and motion of a human body is estimated from its foreground silhouettes in a mono-view, by matching these masks to that of synthetic 3-D human motion representations (Section 3.4.3). In other words, available 3-D body motion models are projected onto different image views for generating silhouettes, which are compared against the observed video human body segmentation masks.

Finally, a novel method to deploy a reference human face muscle model to an arbitrary face model is proposed in Section 3.4.4. In this way, the well-known muscle-based model of Keith Waters is transferred to any other model, while using only a set of feature points marked on both source and target face models. For 3DTV systems, the presented algorithm could be useful, since it allows quick integration and usage of new face models, even if such models do not have the integrated muscles.

Apart from generation of synthetic views of human beings, some other objects could also be generated in 3-D, which could be used as content for 3DTV systems. In Sections 3.5.1 and 3.5.2, a novel approach, which models the interaction between liquids and deformable objects, are presented. These interactions are modeled by using computational fluid dynamics and , applied onto particle systems. The stain creation is also synthesized as a result of the interaction between liquids and cloth objects. The visual results look quite realistic and promising for 3DTV systems.

In pseudo-3D representations, the main goal is usually set to render arbitrary views of a scene or real objects in virtual scenes by implicitly using 3-D information. As presented in Section 3.6.2, trifocal tensor relations could be utilized in order to determine a rendered view of an arbitrarily located virtual camera from two available views. In that formulation, the geometric relations between three cameras (two original and one virtual) are converted into an algebraic formulation, which gives the location of a particular pixel point on the virtual image, whose correspondences are available in the original images. The simulation results present a quite acceptable visual quality of the rendered views in this tensor-based method. Psuedo-3D methods are interesting and attractive for 3DTV systems in the sense that they could provide the required views for the autostereoscopic displays with no explicit utilization of 3-D scene information.

In the next part of the technical report, the abstracts of the aforementioned methods are presented in consecutive sections.

3. Abstracts of Publications for Year-II

3.1. Point representations

3.1.1. Multi-view Video Coding via Dense Depth Estimation

Authors: Burak Özkalaycı, O.Serdar Gedik, A. Aydın Alatan

Institution: METU

*Publication: "*Multi-view Video Coding via Dense Depth Estimation", *prepared in Year-II as an Internal Technical Report and submitted to 3DTV CONFERENCE 2007, The True Vision – Capture, Transmission and Display of 3D Video, Kos, Greece, May 2007*

A geometry-based multi-view video coding (MVC) method is proposed. In order to utilize the spatial redundancies between multiple views, the scene geometry is estimated as dense depth maps. The dense depth estimation problem is modeled by using a Markov random field (MRF) and solved via the belief propagation algorithm. Relying on these depth maps of the scene, novel view estimates of the intermediate views of the multi-view set is obtained with a 3D warping algorithm, which also performs hole-filling in the occlusion regions. The proposed MVC method, based on H.264 standard, encodes a number of reference views in a standard manner, whereas the residuals of the novel view predictions are encoded separately. The proposed MVC method is compared against simulcast coding of each view, yielding better rate-distortion performance, especially at lower bit-rates.

3.1.2. Post-processing of Scattered Point Data

Authors: P. Jenke, M. Wand, M. Bokeloh, A. Schilling, W. Straßer

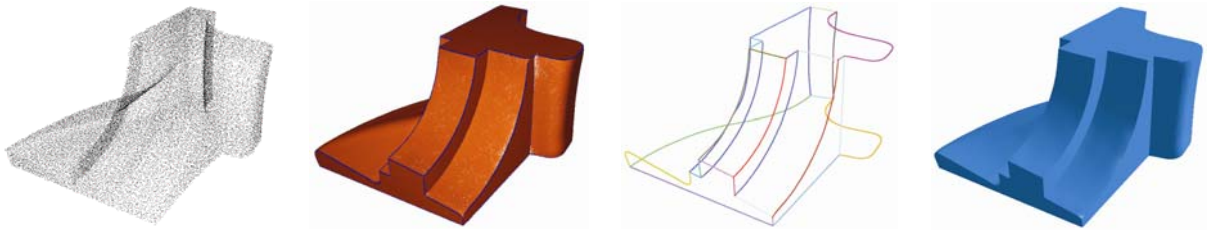
Institutions: Tuebingen, Stanford

*Publication: "*Bayesian Point Cloud Reconstruction", in *Computer Graphics Forum (Proceedings Eurographics '06)*, Volume 25, Number 3, 2006

Many different applications such as surgery planning, cultural heritage projects or material testing require acquisition of the surface a variety of objects. For this purpose, several scanning systems have been developed, including structured light scanners, time-of-flight (laser) systems or space carving approaches. Unlike the diversity of acquisition systems, a single representation, namely scattered point data, has become the ubiquitous primitive for the representation of such 3D surface data in an acquisition process. However, although the quality of the resulting point clouds has improved significantly, several common error sources can still be observed in the datasets. Therefore, it is required to post-process the result and eventually transform it into other representations. We have investigated two research directions in this process: surface reconstruction and hole-filling.

Surface reconstruction is a topic in the field of computer graphics which has been investigated for a long time by now. New acquisition and computation possibilities required for the development of new algorithms several times. We tried to create a statistics-based framework

which allows us to flexibly adjust the reconstruction system to application-specific purposes [1]. We specifically wanted the reconstruction system to handle sharp features in the data, a property which is often disregarded in other publication in this area. An approach that enables us to incorporate these requirements into a unified framework is Bayesian reasoning. Using this formulation, we can combine a generative measurement model with prior assumptions about the surface such as piecewise smoothness and sharp features (edges or corners). We were able to show that this framework can be used to successfully remove noise in data sets, can meaningfully fill in small holes in the data and can preserve and even enhance sharp features.



a) b) c) d)

Figure 1: Input to our algorithm is unstructured scattered point data (a). Explicit information about the features in the object (c) becomes available as part of the reconstruction (b). The result can be transformed into a triangle mesh representation (d).

For small holes, the Bayesian reconstruction approach can even be used to fill in the missing data. However, for larger holes, this is not sufficient and a more sophisticated system is required in order to reconstruct the missing information. An approach which has already proven its success in 2D texture synthesis is to use the local neighbourhood of missing data (pixel position in 2D) for finding similar patches which are then used for reconstruction. We extended this idea to the 3D domain by using a voxelization of the input data for representing the local neighbourhoods.

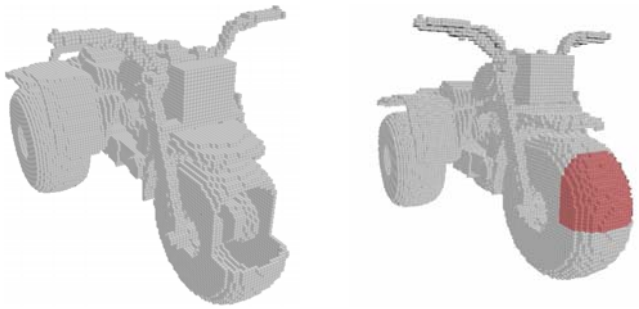


Figure 2: Hole-filling: points with similar local neighbourhoods are used to reconstruct the surface in areas with missing information (Images courtesy of Alexander Berner).

3.1.3. Summary, conclusion, and plans

The initial research of dense-depth representation for multi-view compression is promising. This approach will be extended to utilization of state-of-the-art multi-view coding algorithms that do not use 3-D information. In other words, the current representation will be tested by powerful compression algorithms, which only try to remove the visual redundancy between time and views, naively by 2-D block-matching. It is planned to continue working on dense depth representation technologies with emphasize on multi-view coding.

For Bayesian point-cloud modelling, the current simulation results for noise removal and hole-filling are quite remarkable. In future work, the results achieved until now for Bayesian point cloud modelling will be extend into several directions. The researchers would like to use self-similarity inherent in the data to improve the results of the current reconstruction process. It is the authors belief that this formulation will lead to superior results, since there will be no constrain on the surface, as strongly as currently exists (e.g. assuming piecewise smoothness everywhere). The authors also plan to investigate how well the developed algorithms can be extended to the reconstruction of dynamic data, namely animations.

3.2. Mesh representations

3.2.1. 3D Shape Recovery and Tracking from Multi-Camera Video Sequences via Surface Deformation

Authors: Sahillioglu, Y.; Yemez, Y.; Skala, V.

Institutions: Koç University and University of West Bohemia

*Publication: "*3D Shape Recovery and Tracking from Multi-Camera Video Sequences via Surface Deformation", *IEEE Signal Processing and Communications Applications:* 17-19. 2006

This paper addresses 3D reconstruction and modeling of time-varying real objects using multi-camera video. The work consists of two phases. In the first phase, the initial shape of the object is recovered from its silhouettes using a surface deformation model. The same deformation model is also employed in the second phase to track the recovered initial shape through the time-varying silhouette information by surface evolution. The surface deformation/evolution model allows us to construct a spatially and temporally smooth surface mesh representation having fixed connectivity. This eventually leads to an overall space-time representation that preserves the semantics of the underlying motion and that is much more efficient to process, to visualize, to store and to transmit.

3.2.2. A Spatio-temporal Metric for Dynamic Mesh Comparison

Authors: Libor Vasa, Vaclav Skala

Institutions: University of West Bohemia

*Publication: "*A Spatio-temporal Metric for Dynamic Mesh Comparison", Proc. of *AMDO* 2006:, pp. 29-37.

A new approach to comparison of dynamic meshes based on Hausdorff distance is presented along with examples of application of such metric. The technique presented is based on representation of a 3D dynamic mesh by a 4D static tetrahedral mesh. Issues concerning space-time relations, mesh consistency and distance computation are addressed, yielding a fully applicable algorithm. Necessary speedup techniques are also discussed in detail and many possible applications of the proposed metric are outlined.

3.2.3. Virtual 3D Person Models for Intuitive Dialog Systems

Authors: Patrick Klie, Torsten Büschenfeld, Jörn Ostermann

Institutions: Leibniz University of Hanover

*Publication: "*ViPiD - Virtual 3D Person Models for Intuitive Dialog Systems", *Proc. IEEE Workshop on Content Generation and Coding for 3D-Television, Eindhoven, Netherlands, June 2006.*

ViPiD is a complete framework for audio and 3D video capturing of one or several moving persons as well as the creation of 3D person models for intuitive dialog systems. Therefore we are setting up a multi-camera environment for 3D scene analysis, incorporating aspects such as 3D/4D reconstruction, motion estimation, virtual camera integration, coding of time variant 3D meshes and free viewpoint video. The entire framework serves as a basis for subsequent creation of dynamic mesh sequences with fixed connectivity.

3.2.4. Scene-Flow driven Creation of Time Consistent Dynamic Objects Using Mesh Parameterizations

Authors: Patrick Klie, Eugen Okon, Jörn Ostermann

Institutions: Leibniz University of Hanover

*Publication: "*Scene-Flow driven Creation of Time Consistent Dynamic Meshes Using Mesh Parameterizations", s*ubmitted to 3DTV CONFERENCE 2007, The True Vision – Capture,, Transmission and Display of 3D Video, Kos, Greece, May 2007*

This paper introduces a novel approach to create dynamic mesh sequences with fixed connectivity from synchronized multiple camera videos. The algorithm performs the following steps: Firstly, static reconstructions based on silhouette cone intersections and dense depth maps are created for every time step. Secondly, 3D scene flow is calculated from multiple optical flows and attached to the static mesh in time step n defining anchor points. Thirdly, mesh segmentation is performed based on these anchor points using Voronoi decompositions inducing patch correspondence. The resulting patches are re-meshed by means of mesh parameterizations, hence mesh connectivity is transferred patch-wise from one time step to the next one purveying motion vectors for every vertex of a given static mesh. Finally, motion vector trajectories are low pass filtered to give a realistic impression of the animation.

3.2.5. Summary, Conclusions and Plans

Time-varying mesh representations with connectivity as fixed as possible, but with changing vertex positions, would certainly provide enormous efficiency for storage, processing and visualization. There have been very few attempts to achieve such time-consistent representations, but these works are yet quite premature and can obtain time-consistent meshes only for very short time intervals. Koc University and University of West Bohemia have conducted joint research for developing tools to build such efficient time varying mesh representations. Koc University has focused on the problem of obtaining dynamic meshes whereas University of West Bohemia has mostly worked on simplification of dynamic meshes.

Koc University and University of West Bohemia have developed a snake-based surface deformation framework that can be used for tracking time-varying geometry, which is robust, efficient and easy to implement. An initial surface model, represented as a triangle mesh, is iteratively deformed towards the target boundary in a smooth manner under the guidance of external and internal forces, by restructuring the deformable mesh at each iteration by using local mesh transform operations. By appropriately defining the external forces, the deformation framework can be applied to any type of data that can be used to infer information about 3D geometry. We have applied the proposed deformation framework to the shape from silhouette problem. As also verified by the experimental results, the presented technique can recover any complex shape details (bifurcations, protrusions and cavities) that can be represented under the minimum edge length constraint. Thanks to the local mesh transform operations, the deformable mesh can easily adapt its geometry to the target shape, preserving its uniformity, topological robustness and optimal configuration without need for any re-parameterization.

Koc University and University of West Bohemia have yet used the developed deformation framework only for reconstruction of 3D static objects from their silhouettes. Since our deformation framework is based on Lagrangian approach, the connectivity information is not lost through deformation and hence the presented framework can also be employed for building efficient time varying surface representations. For the time-varying case, we however have only, though promising, very preliminary results. Our further research will include adaptation of the surface deformation technique to the time-varying case, first on synthetic data and then using multi-view video sequences acquired in 3DTV laboratory of Koc University.

Regarding the dynamic mesh simplification problem, we have generalized the idea of computing Hausdorff distance as a measure of difference of two static triangular meshes to the case of dynamic meshes. The generalization is quite straightforward using the dynamic mesh representation. First, both compared meshes (i.e. the original and the compressed or simplified version) are converted to the representation by a static tetrahedral mesh in 4d. Subsequently, both meshes can be uniformly sampled, from each point a closest point on the other mesh is found using a series of optimized point to tetrahedron distance test. The process is repeated for the backward distance, yielding a guess of distance between the dynamic meshes. We have also employed a spatial subdivision technique, which reduces the complexity of the problem from quadratic down to almost linear to the number of tetrahedra of each mesh. Although we do not yet have any implementation of a dynamic mesh simplification available, we have still tested the method on the available data. We have used it to compare two non-equal sequences of human jump, which produced a significantly lower distance measure than the case when a jump sequence has been compared to a walk sequence. Our implementation of the Hausdorff distance can also be used to map the found minimal distance to vertex colors, thus showing the distribution of the error, which may be useful when considering various simplification criteria.

Our further research on simplification will consider three possible scenarios: Frame by frame simplification, spatio-temporal simplification of a 4D representation of the mesh, and single connectivity simplification using a global simplification criterion. We suggest using our spatio-temporal Hausdorff distance evaluation tool to find the best possible way to simplify a dynamic mesh. One of the methods that will most likely be used in all of the three approaches is the quadric based simplification. We would also like to test this method using different approaches to constructing the aggregate quadrics. This intention is inspired by the various vertex normal estimation schemes, where weighting by simplex area usually isn’t the best

possibility. Another challenging problem is the extension of the volume preserving edge costs to the case of tetrahedral meshes, which could the possibly be used for simplification of dynamic meshes using the spatio-temporal method. The evaluation criteria will have to be updated to preserve the hyper-volume of the dynamic mesh. We hope that this extension will be possible in a manner similar to the extension of the quadric based metric.

UHANN is following an alternative approach for obtaining dynamic mesh sequences. These mesh sequences are created for the purpose of dynamic mesh coding as mesh connectivity is expensive to transfer for each frame. Further future plans address the problem of self-collision or collision of two or more foreground objects where a fixed connectivity cannot be uphold. For this reason, a set of local operations will be defined to handle local topology and connectivity changes in order to avoid retransmission of the entire connectivity. Another problem is that the above mentioned approach does not guarantee to give total control of the tangential drift. Further research will try to compensate this by deploying motion-based segmentation giving surface patches whose motion can be described by a three-dimensional affine transform. These new segmentations can be used as a post-processing step to define two-dimensional vector fields on the meshes for the purpose of "undoing" the tangential drift.

Time consistency will also be helpful for dynamic textures and dynamic surface light fields. If correspondence information of the vertices is established an appropriate motion estimation and compensation in texture space can enhance dynamic texture/light field compression.

3.3. Volume representations

3.3.1. Evaluation of 3D Reconstruction Using Multi-view Backprojection

Authors: K. Mueller, X. Zabulis, A. Smolic and T. Wiegand

Institutions: ITI-CERTH, FhG-HHI

Publication: "Evaluation of 3D Reconstruction Using Multi-view Backprojection"*, in the proceedings of ICOB, 2nd Workshop On Immersive Communication and Broadcast Systems*, 27-28 October 2005, Berlin, Germany.

This paper evaluates the final reconstruction quality of 3D objects from different reconstruction methods by comparing rendered views of a 3D model to the original views, initially taken from 2D cameras. The paper uses pixel-by-pixel error measures, like pixel-wise reconstruction error for non-textured objects and PSNR values for colored or textured objects. Concurrently, the limitations of such measures in connection with 3D reconstruction evaluation are highlighted and a reconstruction measurement based on differential values is investigated, where deviations from reference values are analyzed instead of absolute PSNR-values.

3.3.2. Efficient, Precise, and Accurate Utilization of the Uniqueness Constraint in Self-calibrated Stereo

Authors: Xenophon Zabulis, Uğur Topay and A. Aydın Alatan

Institutions: ITI-CERTH, METU

*Publication: "*Efficient, Precise, and Accurate Utilization of the Uniqueness Constraint in Self-calibrated Stereo"*,* in the *Proceedings of ICOB, 2nd Workshop On Immersive Communication And Broadcast Systems*, 27-28 October 2005, Berlin, Germany.

In this paper, the depth cue due to the assumption of texture uniqueness is reviewed. The spatial direction, over which a similarity measure is optimized, in order to establish a stereo correspondence, is considered and methods to increase the precision and accuracy of stereo reconstructions are presented. It is further presented that the proposed method is quite robust to projective distortions due to less accurate camera parameters, possibly obtained through self-calibration. An efficient implementation of the above methods is also offered, based on a scale-space treatment of the data. The above contributions are integrated in a generic and parallelizable implementation of the uniqueness constraint to observe speedup and increase in the fidelity of surface reconstruction.

3.3.3. Synchronous Image Acquisition based on Network Synchronization

Authors: Georgios Litos, Xenophon Zabulis and Georgios Triantafyllidis

Institutions: ITI-CERTH

*Publication: "*Synchronous Image Acquisition based on Network Synchronization", in the *Proceedings of IEEE Workshop on Three-Dimensional Cinematography (3DCINE'06),* June 22, New York City (in conjunction with CVPR)

In this paper, a software-based system for the synchronization of images captured by a low-cost camera framework is presented. It is most well suited for cases where special hardware cannot be utilized (e.g. remote or wireless applications) and/or when cost efficiency is critical. The proposed method utilizes messages to establish a consensus on the time of image acquisition and NTP-synchronization of computer clocks. It also provides with an error signal, in case of failure of the synchronization. The evaluation of the proposed algorithm using a precise LED array system (1ms accuracy) proves the effectiveness of this method.

3.3.4. Efficient, Precise, and Accurate Utilization of the Uniqueness Constraint in Multi-View Stereo

Authors: Xenophon Zabulis and Georgios Kordelas

Institutions: ITI-CERTH

*Publication: "*Efficient, Precise, and Accurate Utilization of the Uniqueness Constraint in Multi-View Stereo"*,* in *Proceedings of 3DPVT 2006, Third International Symposium on 3D Data Processing, Visualization and Transmission*, University of North Carolina, Chapel Hill, June 14-16, 2006.

In this paper, the depth cue due to the assumption of texture uniqueness is reviewed. The spatial direction, over which a similarity measure is optimized, in order to establish a stereo correspondence, is considered and methods to increase the precision and accuracy of stereo reconstructions are presented. An efficient implementation of the above methods is offered, based on optimizations that evaluate potential correspondences hierarchically, in the spatial and angular dimensions. Furthermore, the expansion of the above techniques in a multi-view framework where calibration errors cause the mis-registration of individually obtained reconstructions are considered, and a treatment of the data is proposed for the elimination of duplicate reconstructions of a single surface point. Finally, a processing step is proposed for the increase of reconstruction precision and post-processing of the final result. The above contributions are integrated in a generic and parallelizable implementation of the uniqueness constraint to observe speedup and increase in the fidelity of surface reconstruction.

3.3.5. Increasing the accuracy of the space-sweeping approach to stereo reconstruction, using spherical backprojection surfaces

Authors: Xenophon Zabulis, Georgios Kordelas, Karsten Mueller and Aljoscha Smolic

Institutions: ITI-CERTH, FhG-HHI

Publication: "Increasing the accuracy of the space-sweeping approach to stereo reconstruction, using spherical backprojection surfaces", in the *Proceedings of International Conference on Image Processing (ICIP) 2006*, Atlanta GA, 8-11 October 2006.

In this paper, interest is focused on the accurate and time-efficient stereo reconstruction, for

the purpose of generating 3D animated scenes from multiple synchronized videos. The plane-sweeping approach is reviewed as relevant to the goal of time-efficiency, since its execution can be optimized on a GPU. A method compatible for optimization on the GPU is proposed as a more accurate alternative to plane sweeping and to the derived visibility computation. The method is compared to plane sweeping as to its accuracy, by evaluating the backprojected 3D model against independent views and using n-fold cross validation to estimate the Peak Signal to Noise Ratio (PSNR). Finally, the method's output is casted integratable with multi camera stereo reconstruction frameworks.

3.3.6. Multi Resolution Tetrahedral Meshes

Authors: Ralf Sondershaus and Wolfgang Straßer

Institutions: University of Tuebingen

Publication: "Segment-Based Tetrahedral Meshing and Rendering", *will appear in the proceedings of ACM Graphite*, 2006.

We examined efficient data structures for multi-resolution representations of large unstructured tetrahedral meshes. Tetrahedral meshes are often used as finite element meshes in scientific simulations, like Computational Fluid Dynamics (CFD) or geophysical simulations. The simulations carry data along with the mesh which are usually scalar values like temperature or pressure, or vector values, like velocity, and can be attached to the vertices, edges, border faces, or to the tetrahedra themselves. The emerging need to visualize the simulation data has introduced tetrahedral meshes to volume visualization. But the number of tetrahedra within large finite element meshes doesn't allow for the full resolution mesh to be rendered at interactive frame rates. Both, direct volume rendering approaches like projected tetrahedra and indirect volume rendering approaches like extraction of isosurfaces need to be supported by multi-resolution representations of the mesh in order to be interactive. Thereby, the data structure enables the mesh to adapt to current viewing and classification parameters such that the approximation doesn't introduce rendering artefacts. The data structure must thereby adapt the mesh very fast and store it in a compact way. We designed and implemented two different multi-resolution representations which work on the basis of segments and are described shortly in the following sections.

Binary Segment Hierarchy: This representation uses a binary hierarchy of mesh segments which can be swapped to and from the core memory efficiently. The mesh is partitioned into several segments which are stored on disc and can be processed independently of each other. Every such segment represents a part of the tetrahedral mesh and is itself a small tetrahedral mesh.

The partition of the mesh is steered by an octree whose leaves are merged into segments. Given such a partition, the binary segment hierarchy is constructed iteratively. Every iteration merges two segments into a new segment which is simplified. The two segments are inserted as children of the new segment in the binary hierarchy. The borders between segments are changed only if they restrict the quality of the simplification too much. If the border between two segments is simplified, an additional dependency between both segments is added to the binary hierarchy.

Given the binary hierarchy and the stored additional dependencies, the mesh can be adapted at run time by replacing a segment with both its children (which refines the mesh) or by

replacing two segments with their common parent (which coarsens the mesh). Additionally, the adjacency information can be updated using a hash table of the boundary triangles. At any time, the mesh can be treated as one consistent mesh as well as a collection of independent segments.

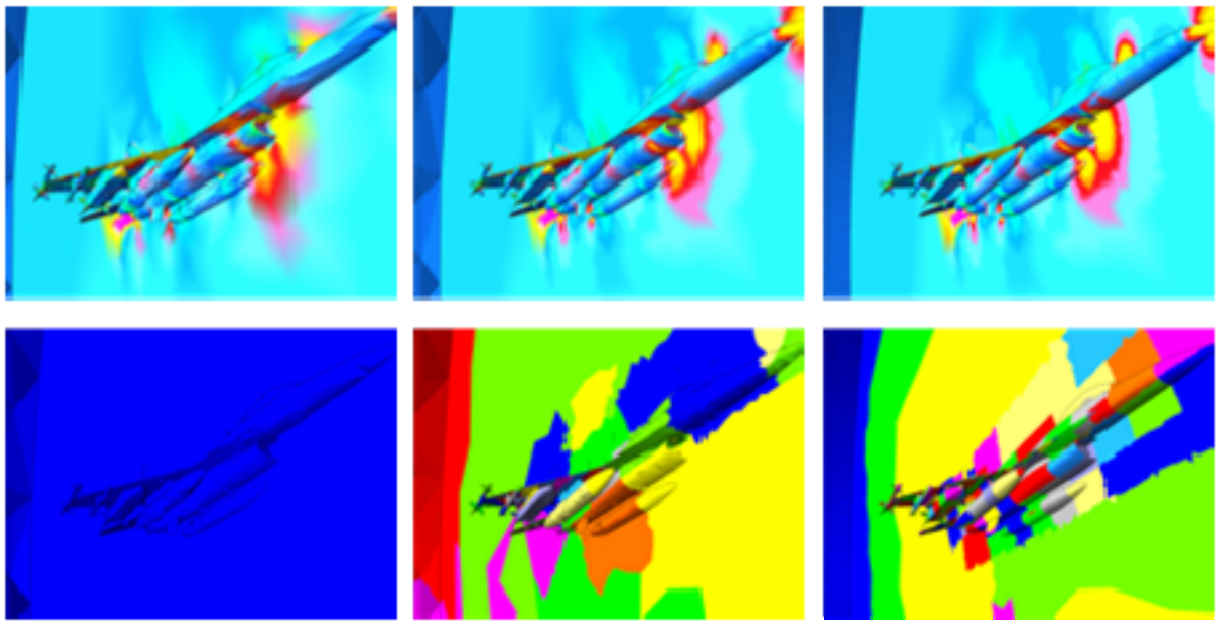


Figure 3. The F16-like model contains about 6 million tetrahedra. Only the border faces of the volume mesh are shown. The simplified model (left, 80,000 tetrahedra) is a coarse approximation of the scalar field. The adapted mesh (middle) contains about 400,000 tetrahedra and shows nearly no differences when compared to the full-resolution mesh (right).

N-ary Segment Hierarchy: In contrast to the binary hierarchy, the n-ary hierarchy stores dependencies between segments in a directed acyclic graph (DAG). The graph is constructed iteratively by a sequence of different partitions of the mesh. The segments of two consecutive partitions are cut against each other resulting in patches which can be thought of as small segments. A segment depends on another segment if the cut of both segments does not result in an empty patch. For triangular meshes, such a data structure is also called a multi triangulation.

Given a partition, its segments are simplified leaving the boundaries of the segments unchanged. So, the boundaries of segments are contained in both the current partition and the previous partition which enables that a segment can be replaced by the shared patches of its preceding segments (which refines the mesh) or by the shared patches of its consecutive segments (which coarsens the mesh).

Again, the adjacencies of the boundaries between segments can be adapted by a hash map that is addressed by the triangles of the boundaries. Depending on the size of the segments which is given by the user the mesh can be adapted very quickly. In order to transmit the meshes fast from the disc as well as over a network, the segments are stored in a compact format.

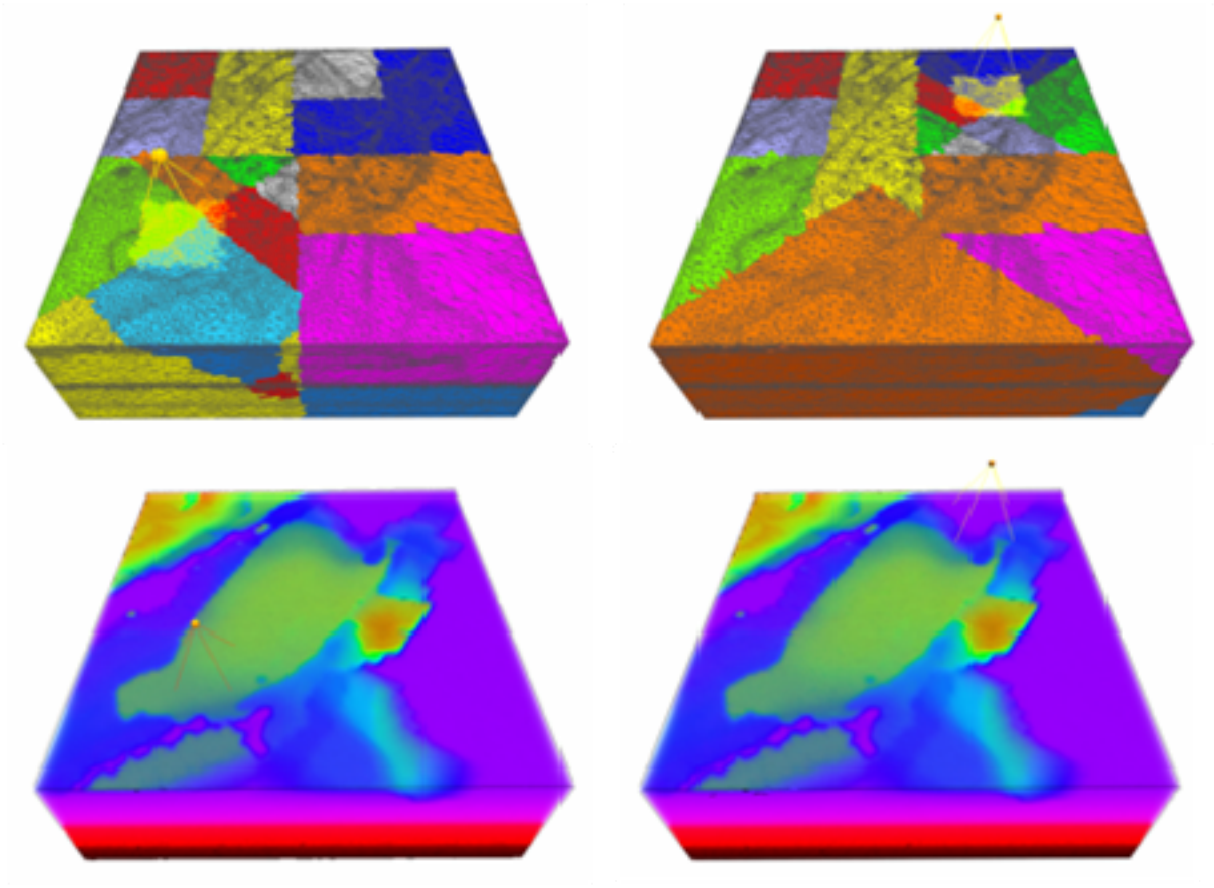


Figure 4. The Earthquake dataset contains more than 13 million tetrahedra. Our multi resolution model can be created in less than one hour and enables the mesh to adapt to viewing parameters very quickly by replacing segments with other segments. The mesh can be visualized in real time on a standard PC. The figures show above show the coloured segments together with the viewer position as well as the direct volume renderings. Both adaptations contain roughly 300,000 tetrahedra.

3.3.7. Summary, conclusion, and plans

Volumetric representations are very useful since they contain "neighborhoodness" information and, thus, facilitate the operation of a wide range of algorithms for the 3D reconstruction of surfaces, the post-processing of the reconstruction results and, last but not least, rendering of these reconstructions. Such methods are further reinforced by volumetric representations, because the latter utilize a regular matrix to represent surfaces, a data-structure that is native to most algorithmic implementations of stereo approaches. In addition, volumetric representations are useful for processing reconstruction results from multiple views because they facilitate the registration of view-dependent results in a common reference frame. Therefore, the main application of volumetric representations is found at data structures that store intermediate results of algorithmic methods. For the output of these methods, typically a more compact encoding of their content is utilized.

Progress was reported in the domain of volumetric representations by the enhancement of the precision, accuracy and efficiency of the computational implementation of one of the most powerful cues to depth-from-stereo approaches, namely the uniqueness constraint or its more recent variant photoconsistency. In addition, a volumetric rendering framework for large

unstructured tetrahedral meshes was proposed. This framework adapts at-run time to viewing and classification parameters and exploits multi-triangulation to increase the interactivity of direct volume rendering, thus permitting large meshes to be explored on standard PCs.

In the accompanying publications, the state-of-the-art in the utilization of the uniqueness cue for stereo reconstruction has been extended by volumetric methods that enhance reconstruction quality and increase algorithmic efficiency. The utilization of volumetric representations in this aspect was twofold. First, the consideration of texture matching in the 3D volume, instead of the 2D image space, was adopted to increase the accuracy and precision of the results. Second the topological matrix-representation offered by volumetric consideration of surfaces, has been exploited in the formulation of a coarse-to-fine approach that conserves memory and reduces computational complexity. Furthermore, the widely used "plane-sweeping" approach to stereo-reconstruction was reviewed as relevant to the goal of time-efficiency, since its execution can be optimized on a GPU. A method compatible for optimization on the GPU is proposed as a more accurate alternative to plane sweeping and to the derived visibility computation that is based on the utilization of spherical backprojection surfaces that projectively expand in volume.

In the field of rendering, progress was reported in multi-resolution volumetric representations of large unstructured tetrahedral meshes. In particular, large finite element meshes don't allow for the full resolution mesh to be rendered at interactive frame rates. Multi-resolution representations were employed to reduce the required memory capacity and, thus, increase the speed of interactive visualization of large meshes. Thereby, the proposed data structure enables the mesh to adapt to current viewing and classification parameters such that the approximation doesn't introduce rendering artefacts. The data structure must thereby adapt the mesh very fast and store it in a compact way. Two different multi-resolution representations which work on the basis of segments were designed and implemented.

Intention for continuation of collaboration between 3DTV partners in the domain was agreed as joint papers in the domain of volumetric representations have yielded fruitful results. ITI-CERTH is to continue (Section 3.3.2) collaborating with METU in the adoption of self-calibration methods of the latter in the reconstruction process and, thus, not only reduce the demand for manual calibration of cameras prior to 3D reconstruction but also correct for small calibration errors at run-time, e.g. for real-time reconstruction. ITI-CERTH is to expand the existing collaboration with FhG-HHI (Sections 3.3.1 and 3.3.5) in the adoption of more sophisticated evaluation techniques for reconstruction results as those are being developed at FhG-HHI, with a special interest in the evaluation of multi-view stereo reconstruction results. After the TR meeting of Florence, it became clear that ITI-CERTH and University of West Bohemia intend to collaborate in the optimal extraction of mesh representations from intermediate volumetric results, in order to increase the compatibility of the reconstruction software modules of 3DTV with those that will perform coding and rendering functionalities.

It is finally reported that during the conduct of experiments in multi-view stereo a need for software synchronization of cameras arose. In this context, a novel approach to such synchronization was proposed (presented in Section 3.3.3) by ITI-CERTH. Since, this work is more related to WP7, it was voluntary contributed there, although that ITI-CERTH has no participation in that WP.

3.4 Human Face and Body Specific Techniques

3.4.1 Automatic Head-Gesture Synthesis Using Speech Prosody

Authors: M. Emre Sargın, Engin Erzin, Çiğdem Eroğlu Erdem, Yücel Yemez, A. Murat Tekalp, A. Tanju Erdem, , Mehmet Özkan

Institutions: Momentum A. S. and Koc University

Publication: Internal report

A new method for automatic and realistic synthesis of head gestures of an avatar from speech prosody is presented. An audio-visual model is first built to predict a sequence of gesture patterns from the prosody pattern sequence computed for the input speech. Head gestures are represented by Euler angles associated with head rotations, and speech prosody by temporal variations in the pitch frequency and speech intensity. A two-stage analysis method is used to "learn" both elementary prosody and head gesture patterns for a particular speaker, as well as the correlations between these head gesture and prosody patterns from a training video sequence. In the first stage analysis, Hidden Markov Model (HMM), based on unsupervised temporal segmentation of head gesture and speech prosody features, is performed separately to determine elementary head gesture and speech prosody patterns, respectively. In the second stage, joint analysis of correlations between these elementary head gesture and prosody patterns is performed using Multi-Stream HMMs to determine an audio-visual mapping model. Objective and subjective evaluations indicate that the proposed synthesis-by-analysis scheme provides natural looking head gestures for the speaker with any input test speech, as well as in ``prosody transplant'' and ``gesture transplant'' scenarios.

3.4.2 Key Frame Reduction of Human Motion Capture Data

Authors: Onur Önder, Uğur Güdükbay, Bülent Özgüç, Tanju Erdem, Çiğdem E. Erdem,

Institutions: Bilkent University, Momentum A. S.

*Publication: "*Key Frame Reduction of Human Motion Capture Data", *WSCG'2006, Plzen.*

A method for combined filtering and key-frame reduction of motion capture data is proposed. Filtering of motion capture data is necessary to eliminate any jitter introduced by a motion capture system. Key-frame reduction also allows animators to easily edit motion data by representing animation curves with a significantly smaller number of key frames. The proposed technique achieves key frame reduction and jitter removal simultaneously, by fitting a Hermite curve to motion capture data using dynamic programming. Implementation details of the proposed filtering and key-frame reduction algorithm are provided.

3.4.3 Motion Capture from Single Video Sequence

Authors: İbrahim Demir, Yiğithan Dedeoğlu, Uğur Güdükbay

Institutions: Bilkent University

*Publication: "*Motion Capture from Single Video Sequence", *to be submitted as a journal paper*

In this work, a framework for the reconstruction of the 3D posture of a human body, from a sequence of single-view video frames, is proposed. This framework initiates by performing background estimation, based on a frame-by-frame image subtraction process, to extract the body's silhouette. Then, the body silhouettes are automatically labelled, utilizing a model-based approach. Finally, the 3D pose is reconstructed from the labelled silhouette, based on the assumption that the corresponding body is orthographically projected in the images. The proposed approach does not require camera calibration and assumes that (i) the camera is static, (ii) the input video has a static background and no significant perspective distortion and (iii) the performer is in upright position.

3.4.4 Algorithm for adaptation of a muscle model to different face models

Authors: S.Piekh, D. Zhdanov

Institutions: UHANN (Leibniz University of Hanover)

Publication: report to be submitted later

This paper describes a method for transferring the muscles used in the Waters muscle model from one reference 3D face model to another. The method can be applied in muscle-based facial animation and has been developed to replace or simplify the integration of muscles in different 3D face models. In addition, it facilitates a quick transfer of the animation from one model to another. The method can find application in object-based video coding where a range of generic 3D facial masks and Waters muscle model are used for the analysis and synthesis of facial mimic. Starting from a reference 3D face model with integrated muscles, the proposed algorithm transfers the muscle system to a target 3D face model. The method utilizes the feature points on 3D mesh for adaptation and, thus, the resolution of 3D face model is not relevant. The quality of adaptation can be controlled by number of selected feature points.

3.4.5 Summary, conclusion, plans

It has been shown that it is possible to model the correlation between speech prosody (pitch, intensity) and 3D head gestures using hidden Markov models. Experimental results

demonstrate that automatic head animation using speech as the only input is possible via the generated HMM model.

Implementation details of the proposed key-frame reduction algorithm are given and the preliminary results are encouraging. Further tests are necessary to verify the effectiveness of the algorithm.

Realistic animation of humans is an important part of 3D scenes. One way to achieve it is to use motion capture data. We proposed a framework for the animation of humans from a sequence of single view video frames. In the future, we will continue to investigate motion capture and pose reconstruction based on motion capture data from a single-view video sequence, which is very useful for constructing motion libraries, especially from public resources.

3.5 Object Specific Representations: Modeling, Rendering and Animation Techniques

3.5.1. Modeling Interaction of Fluid, Fabric and Rigid Objects

Authors: Serkan Bayraktar, Uğur Güdükbay, Bülent Özgüç

Institutions: Bilkent University

Publication: "Modeling Interaction of Fluid, Fabric, and Rigid Objects for Computer Graphics", in *Proc. of IEEE 14th Signal Processing and Communications Applications*, IEEE Computer Society, Antalya, Turkey, April 2006.

Simulating every day phenomena such as fluid, rigid objects, or cloth and their interaction has been a challenge for the computer graphics community for decades. In this article, techniques to model such interactions are explained briefly and some of the results of applying these techniques are presented.

3.5.2. A Unified Particle-Based Method for the Interaction of Fluids and Deformable Objects

Authors: Serkan Bayraktar, Uğur Güdükbay, Bülent Özgüç

Institutions: Bilkent University

Publication: "A Unified Particle-Based Method for the Interaction of Fluids and Deformable Objects", *submitted to Computers & Graphics.*

Simulating natural phenomena such as fluids, deformable objects, cloth, or fire has been a challenge for the computer graphics community. For most of these phenomena there exist models in computational physics and engineering. The field of Computational Fluid Dynamics (CFD) is a well-established research area with many applications in engineering and computational physics. CFD-based applications have been developed to simulate the fluid behavior for computer graphics.
In this paper, we propose a system based on CFD to animate the interaction of fluids and deformable objects. We employ some established methods, such as the Smoothed Particle Hydrodynamics (SPH) and mass-spring systems, and a unified particle-based representation for fluids, cloth, and deformable objects. This unified representation enables us to easily define the interactions between different types of objects. Moreover, by exploiting the data representation for the particle system, we are able to speed up the collision detection and object proximity tests.

3.5.3. Summary, conclusion, and plans

The interaction of fluids with deformable solid objects is an important area of research in Computer Graphics. It is also a very important part of various 3D Scenes; thus it should be realistically represented for 3DTV. We implemented a framework for the interaction of liquids with solid objects based on Computational Fluid Dynamics (CFD). The framework simulates the motion of the liquids and solids while they are interacting and renders them. We will continue our research in this area in two directions: (i) improvement of the simulation techniques for the interaction of fluids and solid objects, and (ii) improvements of the rendering techniques used for rendering the interaction of liquids with solids, such as stain creation.

3.6. Pseudo-3D representations

3.6.1. Framework for 3D video objects

Author: Christian Weigel

Institutions: UIL; Plzen

Publication: (UIL solo) Towards a 3D-TV System on the Basis of Image-Based Rendering Methods, *Proceedings 51. Internationales Wissenschaftliches Kolloquium (IWK), 2006 September 11-15, Ilmenau, Germany*

In the first technical report the establishment of collaboration between UIL and Plzen was reported. The aim was to merge the two software systems, namely ReVOGS and MVE 2, used at both institutions for a diversity of 3D related tasks. In first evaluation it came out, that, although the basic idea of the systems is quite common, it is almost impossible to merge them together to one system. The software architecture is too different. While ReVOGS is using a generic C++ approach, MVE is based on C# and running on the .NET environment. Furthermore MVE2 is heavily computer graphics oriented, the ReVOGS module do focus on computer vision related tasks. Altogether, it was decided not to follow the initial plan, since the effort to accomplish the task was in no reasonable relation to the benefit of it.

The development of ReVOGS was on of the main efforts of UIL within the reporting period. From a very rudimentary alpha state the software is know in quite usable shape in terms of scientific applications. A number of new functionalities were added in order to start with the tasks described in Section 0. Since it makes not really sense to publish the development of software in scientific papers, only one publication is related to this section. It gives an overview/vision of the whole 3D video object system with ReVOGS as basis.

Considering the development state of ReVOGS UIL decided to upload it to the 3DTV software repository in October 2006 for usage within the NoE. Starting with a binary distribution every institution interested in jointly developing modules with new functionality can contact UIL in order to get the sources.

3.6.2. Evaluation of different 3D video object synthesis methods

Author: Christian Weigel

Institutions: UIL, FhG-HHI

*Publication: (*UIL solo) Advanced 3D Video Object Synthesis Based on Trilinear Tensors*, Conference Proceedings of the 10th IEEE International Symposium on Consumer Electronics (ISCE). June 28 - July 01, 2006, St. Petersburg, Russia*

The joint work on this topic, unfortunately, was delayed. It was planned to have first results by the beginning of the second project period but due to several reasons the target could not be hit. On of the reason might be the underestimation of the efforts for software development

required in order to compare the methods. Anyway the target is still set and we are looking forward to accomplish this task in the near future.

Besides the software development, one of the preliminary works was the enhancement of UIL's view synthesis algorithms in a way that the viewpoint can be really chosen freely. Only this modification makes it possible to compare the two methods. The algorithm was changed form simple view morphing to a new warping technique. In addition to the originally proposed algorithm we enhanced it by some modification in the pre- and post-processing phase.

With the help of matching constraints obtained from epipolar geometry a dense disparity map can be estimated using a modification of Birchfield and Tomasi's pixel-to-pixel stereo algorithm. In a next step knowledge about the explicit scene geometry as well as the information acquired by several image correspondences can be used to synthesize novel views. The mapping method we employ is called trifocal transfer and based upon the usage of trilinear tensors introduced by Shashua and Avidan. This synthesis method is extended in order to work with sequences of images that have a masking alpha channel. Moreover, this transfer method avoids the issue of indirect view specification and allows the user to move freely in space. The previously mentioned pre-processing steps are combined efficiently with a trilinear warping method particularly with regard to obtain good quality even with wide baseline camera setups. Therefore hole-filling algorithms have been developed.

3.6.3. Scalable image-based video objects

Institutions: UIL

Publication: None in the current period. Section provided, since topic still considered!

As mentioned in the previous sections, the main focus within this subgroup was laying on the development of the software system as the basis for accomplishing the scientific tasks. Therefore, only some reserach was done in the area of scalable image based video object. One of the related topics was the evaluation of diffrent methods for dense depth evaluation. The idea behind this to drive the complexity of the 3D video object generation by employing different algorithms of DM estimation. A first collaboration was established with METU. Since this special topic is more related to Point representations please refer to this section for a more detailed description.

3.6.4. Summary, conclusion, and plans

We had have changed the focus within this subgroup during the second period. Due to the basic necessity of a test environment, it had been shifted to software development. Ideas to combine frameworks of institutions within the NoE, namely UIl & Plzen, could not have been realized as described in Section 3.6.1. However, the development of ReVOGS by UIL has made a comprehensive progress of which, hopefully, some institutions can benefit. The software was uploaded to the common internal software directory.

It has been planned that the collaboration with FhG-HHI about evaluation of synthesis methods will continue in the upcoming period. The research on scalable representation of 3D video object will come back to focus in the next period. Now, the basis for algorithm implementations and testing had been laid.

4. Conclusions and Future Directions

This technical report presents the outputs of 20 research activities, whilst 6 of these papers are result of joint efforts. It should be noted that Technical Report # 1 for WP8 also had 19 research outputs with 6 joint publications. In Year-II, based on updated DoW of 3DTV NoE, a number of high-priority joint research areas have been already determined and research efforts continue in all of these research topics.

The results of point-based representations are promising in the sense that multi-view coding technologies could exploit 3-D information of the scene to better remove the redundancies. The research direction in this topic is expected to focus on merger of the current multi-view coding standard algorithms by the proposed algorithm, so that the rate-distortion performance might improve, especially for the static scenes. On the other hand, Bayesian formulation of 3-D point-clouds could be useful in different applications, from active range sensing device outputs to sparse depth estimation algorithms. Both of these research efforts from METU and TUEB will continue in Year-III for obtaining better results.

Time-consistent and fixed-connectivity mesh representations are one of the most promising research directions for the maturity of the mesh-based techniques, as well as the efficiency of the representation in time due to the constant mesh topology. On the other hand, UHANN has also some impressive work on the similar research directions with an almost complete 3D extraction and representation system that performs in real-time. Future joint research efforts of BILKENT and PLZEN will be directed towards completion of the proposed dynamic mesh update algorithm. In Year-III, these research initiatives are expected to get mature to result with demonstrations.

Utilization of small planar patches to exploit the texture uniqueness is an elegant idea, leading to precise 3-D reconstruction results, while it is applicable to stereo, multi-view or even uncalibrated imaging systems. In Year-III, the same idea is expected to be applied jointly by ITI and METU to arbitrary-shaped pre-segmented regions for modeling 3-D structure of these segments via planes to obtain continuous planar fits to man-made structures.

The research on human face and body representation continues on two main paths. The joint modeling of head movements and speech prosody is a novel idea, jointly developed by KOC and MOMENTUM. The initial simulation results are quite promising and Year-III is devoted to the better modeling of these two human behaviors and to the performance of more tests on the proposed method. The other research path, which is key-frame reduction for motion extraction, is also a new approach to the efficient representation of human motion. BILKENT and MOMENTUM have reached to quite satisfactory results, in this research effort.

BILKENT has two solo contributions on the representation of the interactions between deformable objects and the liquids to particle systems. The resulting synthesized images are quite remarkable. The research activity is expected to continue in Year-III, especially for improvements in the rendering of the interaction fluids and deformable objects.

The research efforts by UIL in Pseudo3D representations are devoted to generation of arbitrary virtual views from given image pairs. The generated views look quite realistic with

no significant visual distortions. Further research is planned to be conducted on different depth estimation algorithms with some other partners.

5. Annex

5.1 Multi-view Video Coding via Dense Depth Estimation

5.2 Bayesian Point Cloud Reconstruction

5.3 3D Shape Recovery and Tracking from Multi-Camera Video Sequences via Surface Deformation

5.4 A spatio-temporal metric for dynamic mesh comparison

5.5 ViPiD - Virtual 3D Person Models for Intuitive Dialog Systems

5.6 A Framework For Scene-Flow Drıven Creatıon Of Tıme-Consıstent Dynamıc Objects Usıng Mesh Parametrızatıons

5.7 Evaluatıon of 3D Reconstruction Usıng Multıvıew Backprojectıon

5.8 Efficient, Precise, and Accurate Utilization of the Uniqueness Constraint in Self-Calibrated Stereo

5.9 Synchronous Image Acquisition based on Network Synchronization

5.10 Efficient, Precise, and Accurate Utilization of the Uniqueness Constraint in Multi-View Stereo

5.11 Increasıng The Accuracy Of The Space-Sweepıng Approach To Stereo Reconstructıon, Usıng Spherıcal Backprojectıon Surfaces.

5.12 Segment-Based Tetrahedral Meshing and Rendering

5.13 Automatic Head-Gesture Synthesis Using Speech Prosody

5.14 Key-Frame Reduction of Human Motion Capture Data

5.15 Motion Capture from Single Video Sequence

5.16 Algorithm for adaptation of a muscle model to different face models

5.17 Modeling Interaction of Fluid, Fabric, and Rigid Objects for Computer Graphics

5.18 A Unified Particle-Based Method for the Interaction of Fluids and Deformable Objects

5.19 Towards a 3D-TV System on the Basis of Image-Based Rendering Methods

5.20 Advanced 3D Video Object Synthesis Based on Trilinear Tensors