D26.2 Technical Report # 2 on 3D Time-varying Scene Capture Technologies

Project Number: 511568

Project Acronym: 3DTV

Title: Integrated Three-Dimensional Television – Capture, Transmission and Display

Deliverable Nature: R
Number: D26.2
Contractual Date of Delivery: M29
Actual Date of Delivery: M30
Report Date: 7 March 2007
Task: WP7
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: Christian Theobalt (theobalt@mpi-inf.mpg.de)

Editor: Christian Theobalt (theobalt@mpi-inf.mpg.de)

7 March 2007

3D Time-varying Scene Capture Technologies
TC1 WP7 Technical Report 2

EDITOR

Christian Theobat (**MPG**)

Contributing Partners to Technical Report:

Bilkent University (**Bilkent**)
Bulgarian Academy of Sciences (**CLOSPI-BAS**)
Institute of Media Technology, Technische Universitaet Ilmenau (**UIL**)
Tampere University of Technology (**TUT**)
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**)
Max-Planck-Institut fuer Informatik (**MPG**)
University of Tuebingen (**UNI-TUEBINGEN**)
University of Hannover (**UHANN**)
Technische Universitaet Berlin (**TUB**)
University of Aberdeen (**UNIABDN**)
Yoğurt Bilgisayar Teknolojileri Tic. Ltd. Şti. (**Yogurt**)

REVIEWERS

Elena Stoykova (**CLOSPI-BAS**)

Nikos Grammalidis (**ITI-CERTH**)

Edilson de Aguiar (**MPG)**

Joern Ostermann (**UHANN**)

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

TABLE OF CONTENTS

Executive Summary 1

1. Introduction 2

2. Analysis of the Scientific Results.. 4

2.1. Multicamera 5

2.2. Single Camera 6

2.3. Human Face and Body 6

2.4. Holographic Camera Techniques 7

2.5. Pattern Projection 7

2.6. Motion Analysis 8

2.7. Registration 9

2.8. Object –Based Segmentation 9

3. Abstracts of Papers and Technical Reports.. 10

3.1. Multicamera 10

3.1.1. Graph Cut based Panoramic 3D Modeling and Ground Truth Comparison with a Mobile Platform - The Wägele.. 11

3.1.2. Multicamera Motion Capture for Articulated Body Model Animation 13

3.1.3. Relightable Free-Viewpoint Video of Human Actors.................................... 13

3.1.4. Enhanced Dynamic Reflectometry for Relightable Free-Viewpoint Video 14

3.1.5. Spatio-temporal Reflectance Sharing for Relightable 3D Video 15

3.1.6. A Framework for Natural Animation of Digitized Models............................. 15

3.1.7. Automatic Learning of Articulated Skeletons from 3D Marker Trajectories .. 16

3.1.8. Texture Replacement of Garments in Video 16

3.1.9. Weighted Minimal Hypersurface Reconstruction... 17

3.1.10. Synchronous Image Acquisition based on Network Synchronization 17

3.1.11. Conclusions and future plans 18

3.2. Single Camera 20

3.2.1. A Geometric Segmentation Approach for the 3D Reconstruction of Dynamic Scenes in 2D Video Sequences 20

3.2.2. Prioritized Sequential 3D Reconstruction in Video Sequences with Multiple Motions 21

3.2.3. A Modular Scheme for 2D/3D Conversion of TV Broadcast.......................... 21

3.2.4. Color High Dynamic Range (HDR) Imaging in Luminance- Chrominance Space 22

3.2.5. Conclusions and future plans 23

3.3. Human Face and Body 23

3.3.1. Automatic adaptation of face model to human face in video 24
3.3.2. Active Appearance Models for Facial Features Detection.............................. 24
3.3.3. 3D Face Recognition by point signatures and iso-contours. 25
3.3.4. Speech/Gesture Correlation Analysis for Prosody-Driven Head-Gesture Animation 25
3.3.5. 3D Face Motion Capture 26
3.3.6. Face Detection and Eye Localization in Video 26
3.3.7. Conclusions and future plans 27
3.4. Holographic Camera Techniques 29
3.4.1. Developments for underwater remote vision 29
3.4.2. Real time capture and reconstruction of digital holograms using reconfigurable hardware 30
3.4.3. Numerical Methods of Reconstruction employed in underwater Digital Holography (eHolocam)... 30
3.4.4. Conclusions and future plans 31
3.5. Pattern Projection 32
3.5.1. Real-time acquisition of depth and color images using structured light 32
3.5.2. Three-dimensional profilometry by symmetrical fringes projection technique34
3.5.3. Comparative study of fringes generation in two-Spacing phase-shifting profilometry.. 36
3.5.4. Phase retrieval techniques in coordinate measurement.................................... 36
3.5.5. Holographic and digital methods for recording and data processing in investigation of cultural heritage 37
3.5.6. Real-time phase-stepping pattern projection profilometry............................... 38
3.5.7. Conclusions and future plans 39
3.6. Motion Analysis and Tracking 40
3.6.1. Using Multi-target Tracking from Multiple Views for Football-related Applications 40
3.6.2. Traffic Monitoring Using Multiple Cameras, Homographies and Multi-Hypothesis Tracking 41
3.6.3. Model-based Head Motion Estimation with an Automatic Update of Texture Information.. 42
3.6.4. Conclusions and future plans 43
3.7. Registration 43
3.7.1. An Image Registration Technique Aimed at Super Resolution on Mobile Devices 43
3.7.2. Corner Validation based on Extracted Corner Properties 45
3.7.3. Conclusions and Future Work.. 46

3.8. Object-Based Segmentation ... 47
3.8.1. Motion-Based Initialization of Color Models ... 47
3.8.2. Conclusions and Future Plans ... 48
4. Conclusions ... 49
4.1. Multicamera ... 50
4.2. Single Camera ... 51
4.3. Human Face and Body ... 51
4.4. Holographic Camera Techniques ... 52
4.5. Pattern Projection ... 52
4.6. Motion Analysis and Tracking ... 53
4.7. Registration ... 53
4.8. Object-Based Segmentation ... 53
References ... 55
5. Annex ... 58

Executive Summary

This technical report summarizes the scientific results obtained within work package 7 of the 3DTV project, which performs “Joint research on 3D time-varying scene capture technology”. It is the second technical report published by the WP7 partners and summarizes the results obtained during the project period between month 17 and month 29.

The sixteen institutions participating in WP7 jointly investigate the theoretical and algorithmic foundations of an important sub-part of the 3D television production process, namely 3D scene recording and scene reconstruction. The technological concepts developed in WP7 will enable efficient and accurate reconstruction of 3D TV scene representations from captured real-world footage. The faithful reconstruction of such dynamic scene models is a prerequisite for high-quality 3D content display, and therefore the methods developed in this work package also lay the algorithmic foundations for many other work packages in the 3DTV NoE.

In order to systematically investigate the multitude of algorithmic challenges in time-varying scene capture, eight high-priority research sub-areas were defined. These sub-areas are *Multicamera Techniques, Single Camera Techniques, Human Face and Body Techniques, Holographic Camera Techniques, Pattern Projection Techniques, Motion Analysis and Tracking Algorithms, Registration Methods* and *Object Segmentation Approaches.*

In the reported period, the project partners have obtained significant new scientific results in many of the primary research areas. Exemplary result highlights were the development of novel model-based algorithms to create relightable 3D videos of human actors, methods for 2D video to 3D conversion, face and eye localization, multi-target tracking as well as research on color HDR imaging.

Overall, 36 papers were published in top tier journals, books and conference proceedings. This is a notable increase in the number of published papers in comparison to previous reporting period. An increase in the number of joint publications between multiple project partners could also be observed. The high scientific output also demonstrates the high research dynamics in the work package, as well as the good progress of individual research projects.

Representatives from the WP7 partners have met twice during the period, first in Saalbach, Austria, in February 2006, and again in Florence, Italy, in September 2006. Both meetings were very fruitful and spawned many novel research projects. Currently, there are 18 such projects going on.

The high quality of the scientific results summarized in this report shows that the work package is very successful both in terms of scientific impact, but also in terms of scientific exchange between multiple institutions. Overall, the partners have found solutions to some significant research problems that advance the field as a whole. The joint results are also made available to other project participants, as well as the scientific community via a joint software repository. In future, the WP7 partners plan to further intensify their cooperation, to initiate more joint research projects, and to increase the frequency of research visits.

1. Introduction

3D Television augments the traditional TV technology by showing the viewer not only sequences of 2D images but streams of three-dimensional scene representations. To the viewer at home this will mean a completely new media experience. He will perceive the displayed events in a more immersive way, and he may even get the chance to choose his own viewpoint to watch the displayed events. In the future, three-dimensional movies will become a standard and provide enhanced interactivity options, e.g. by allowing the user to navigate through the scenes.

The production pipeline for 2D television has developed into a mature and well-understood process over many years. Scenes are recorded with cameras from single view points, captured image streams are postprocessed, transferred to receivers, and displayed on planar screens. In contrast, the production process for 3D television requires a fundamental rethinking of the underlying technology. Scenes have to be recorded with multiple imaging devices that may be augmented with additional sensor technology to capture the three-dimensional nature of real-scenes. In addition, the data format used in 3D television is a lot more complex. Rather than normal video streams, time-varying computational models of the recorded scenes are required that comprise of descriptions of the scenes' shape, motion, and multi-view appearance. The reconstruction of these models from the multi-view sensor data is one of the major challenges that we face today. Finally, the captured scene descriptions have to be shown to the viewer in three-dimensions which requires completely new display technology.

Today, 3D Television is still in its early days. Many technological and computational problems in scene acquisition, scene reconstruction, and scene display are either unresolved or bring today's technology to its limits. Furthermore, the problems to be solved require expertise from many different areas in science and engineering, ranging from computer science, over physics, to electrical engineering. The European Union's network of Excellence on 3D television brings together a multi-disciplinary group of 21 leading European researcher centers that jointly works on the solution to the myriad of challenging technological problems. The project partners have split their research effort into seven major research work packages, each of which is focusing on one of the most intriguing technological problems that we face on the way towards 3D television.

The last twelve months has been an active and exciting period for work-package 7. The research projects carried out by partners, both joint and individual are continuing, and several more are being discussed. A direct result of the successful work is the texts and scientific papers presented in this report, six of them jointly authored by researchers from WP7 partners. During he reported period, the successful research in WP7 lead to 35 publications in major international conferences, journals as well as edited books and technical reports.

There are currently eight tasks identified as high priority research areas within Work Package 7. These are: *Multicamera, Single Camera, Human Face and Body, Holographic Camera Techniques, Pattern Projection, Motion Analysis and Tracking, Registration* and *Object Segmentation*. Most research by the partners in WP 7 is generally focused towards these tasks. However, naturally many projects will of course touch several areas. For instance, 3D scene capture technologies, such as Single- and Multi-camera methods and applications, will naturally use techniques from the Tracking task. Thus, even though this report organizes the abstracts of the contributed texts into sub-sections corresponding to the tasks, many of the contributions contain work in several WP7 areas.

This report is arranged as follows: In the next Section, an analysis of the presented material is given. In Section 3, abstracts of papers and technical reports produced by WP7 partners during the project period are presented. These are organized into subsections corresponding to the identified research tasks. There after, in Section 4, a brief conclusion summarizing the research is given. Section 5 contains a bibliography of references cited in the abstracts. Finally, attached to the document are reprints of the submitted papers and reports.

2. Analysis of the Scientific Results

In this section, we briefly analyze the scientific results presented in this report and describe their scientific impact. Given the variety of algorithmic challenges that one faces in time-varying scene capture, the individual tasks of WP7 naturally address many different topics. In a sense, this also mirrors the very many technologies being developed for 3D recording, reconstruction and vision. Analyzing the work, we can roughly say that the research performed under the name *Scene Capture* may be thought of as recording methods together with vision and applications. This, by itself, is already a notable contribution to the scientific community, because the work of WP7 is an excellent example for a research effort in which joint expertise from different disciplines, e.g. engineering, computer vision and computer graphics, enables novel applications.

In the reported period, the project partners developed a variety of important methods that advance the field and pave the trail for mature 3D video and 3D TV capturing and reconstruction technology.

On the recording side, novel ways for camera synchronization were developed which is an important prerequisite for high-quality input video data. WP7 also proposes a new hardware architecture for holographic recording. Finally, there are several contributions dealing with Pattern Projection techniques which push that area forward by analyzing and evaluating fringe generation, phase retrieval as well as color recording.

There has also been a lot of work in vision and applications, dealing with very versatile topics such as 3D reconstruction, feature detection and tracking of body models. In the 3D reconstruction field several papers in the Single Camera task deal with 2D to 3D conversion of video data, being an important technology to convert standard TV footage into 3D TV footage. However, there is also work in complete 3D reconstruction from multiple cameras, for example using Weighted Minimal Hypersurfaces. We also see a trend where not only a three dimensional model is reconstructed, but also lighting and material properties of the scene. This may be an important future step as this information could be used to render scenes on 3D-TV display systems under user-controlled environment conditions. Furthermore, compositing of 3D Video and 3DTV objects is facilitated.

Instead of a full reconstruction, model estimation and detection can be used for known objects in a scene. This report presents several approaches to tracking, motion estimation and detection. Many of the model-based results are focusing on the human face and body as they are natural candidates in many video sequences. Work dealing with body model animation and 3D face and speech recognition are representative for this area, but there is also work on tracking humans in multiple view video, as well as traffic monitoring which could prove to be important for large area scenery. In addition to the more application oriented contributions there is also work in basic vision and image registration. One paper is especially oriented towards image registration on mobile devices for instance. Finally, there have been some first steps taken towards dealing with High Dynamic Range recording, which probably will be a important part of future display systems enabling the reproduction of such higher fidelity video streams.

In the following subsections, the contributing partners analyze their results obtained in the different research areas and illustrate their impact on the scientific community.

2.1. Multicamera

For the Multicamera task, results for the whole graphics pipeline from acquisition to rendering have been obtained. One project (Section 3.1.10) developed a software-base method for synchronizing low-cost cameras. By developing a prototype system it was shown that accurate camera synchronization (<1 ms error) can be obtained with a software solution. Another project (Section 3.1.2) presents a newly installed multicamera acquisition system, which has resulted in successful 3D tracking and acquisition using colored markers. We are able to display the captured motion as 3D animation, in spite of some manipulations in marker points and some vibrations of limbs that occurred during the rapid movements. These are our preliminary results of and they play very significant role in integrating capture techniques leading to a low cost hardware/software combination.

In the next step of the pipeline, different 3D reconstruction techniques were used, such as graph cut-based omnidirectional stereo vision in Section 3.1.1. Automated 3D model acquisition techniques are essential for many emerging applications. 3DTV is one major area where 3D models are required as the user will be able to freely choose his view point in real-time. Stereo vision is one major approach to 3D modeling relying solely on the visual modality. An omnidirectional stereo vision approach based on graph cuts, which enables high quality results, is used. For this purpose, a mobile acquisition platform – the Wägele was designed. It can be transported easily and is capable of acquiring both indoor and outdoor scenes. Due to the fact that our localization technique of the measurement platform is robust and precise, we are in a position to use just one camera and sample the scene, which is much more inexpensive than a multi-camera approach yet yielding the same results for static scenes.

Work on weighted minimal hypersurfaces (Section 3.1.9), another reconstruction technique, now allows time-coherent capture of articulated bodies. As another result, it is the first method to capture arbitrary geometry of flowing water. Work on motion capture (Sections 3.1.3-3.1.7) has progressed towards photorealistic acquisition of surface reflectance properties. The fully-automatic system in Sect. 3.1.7 is able to learn articulated skeletons of arbitrary subjects from 3D marker trajectories. It does not use a priori information about the kinematics of the captured individual and can be applied to arbitrary subjects including humans and animals. The learned models are comparable to the ones obtained with commercial software in terms of accuracy and detail.

Up to now the focus in 3D video has been on reproducing the multi-view appearance of dynamic scenes under the lighting conditions that prevailed at the time of recording. However, in many applications, it may become necessary to combine captured 3D video footage with other free-viewpoint content. To serve this purpose, one can not simply merge the scene representations, since the lighting conditions will not match. We thus developed a method to capture relightable free-viewpoint videos of human actors that can not only be displayed from novel viewpoints, but also under arbitrary virtual illumination conditions. We have developed algorithms to reconstruct such relightable dynamic scene representations from only eight input video streams captured under calibrated lighting. Thus, our work shows that it is feasible to capture dynamic geometry, dynamic normal maps, and spatially-varying BRDFs, with a moderately sized acquisition setup. This is an important finding that proves the importance of model-based representations in 3D video as well as multi-camera techniques. The myriad of involved algorithmic problems were researched in two stages.

In stage I, described in “Relightable Free-Viewpoint Video of Human Actors”, the nuts and bolts of a dynamic reflectance estimation method that measures BRDFs and normal maps in dynamic scenes were developed. Visual quality was further improved by the work presented in stage II that is detailed in “Enhanced Dynamic Reflectometry for Relightable Free-Viewpoint Video”. There, new methods to solve many multi-view surface to texture registration problems, such as cloth shifting and geometry inaccuracies, were developed. Algorithms to automatically shape adapt a new improved single-skin body model were also invented. To our knowledge, this is the first approach in the literature that can capture relightable dynamic scene representations of such complexity completely passively. It lays the foundations to many interesting applications in 3D video display, but also editing and post-production since lighting can be changed in the aftermath.

2.2. Single Camera

The research on Single Camera techniques culminated in the reconstruction of 3D scenes from 2D broadcast video. This is a key achievement for conversion of 2D archive material into 3D, and it defines an alternative path for 3DTV broadcast. A multi-frame structure-from-motion (MFSfM) approach is preferred for the solution, due to the existence of many mature algorithms in the field. However, it is observed that there exist certain issues to be resolved for the application of such an approach to 2D/3D conversion of broadcast video problem.

The first of these issues is the fact that the default ordering of the frames is not amenable to processing, due to short baseline length. On the other hand, apparently, a longer baseline means less and unreliable correspondences. Thus, a metric taking into account both of these factors is devised.

The second issue is the handling of multiple motions present in a scene (i.e., dynamic scene), which is essential for the ability to process arbitrary broadcast content. The geometric constraint on the location of points is identified to be a powerful tool, and a segmentation algorithm employing the epipolar constraint is designed to tackle this problem.

One final issue was the recovery of internal calibration parameters from broadcast video, or self-calibration. The accuracy of internal calibration estimates considerably affects the quality of the metric reconstruction, which is important primarily for visualization purposes. The experiments on some simple self-calibration algorithms indicated that most methods suffer from robustness issues, and computations should be performed at projective frame unless necessary.

2.3. Human Face and Body

In Section 3.3, the Human Face and Body specific techniques are presented. A priori knowledge about 3D structure and motion of human face and body, and nature and limits of human motions are used in order to make the processing more efficient. In this section a wide spectrum of algorithms is presented which deal with the following research directions: a) 3D face recognition (publication presented in subsection 3.3.3), b) face and facial feature detection (publications presented in subsection 3.3.2 and 3.3.6), c) 3D face motion and head gesture analysis (publications presented in subsection 3.3.4 and 3.3.5), d) capturing/modeling 3D structure of human face (publication presented in subsection 3.3.1).

All these research directions are very important for 3DTV applications. On the one hand, the possibility to detect and trace of a person face and person's view point is necessary to render the correct view according to the observer position in 3D display systems. On the other hand the exact analysis of 3D face structure and motion is important for exact description and synthesis of the scenes. All handled topics complement each other. An overview of experimental results for each topic is given in subsection 3.3.7.

2.4. Holographic Camera Techniques

Digital holography as compared to traditional methods of holography is seen as the way forward in realizing practical, mass media 3D displays. However, there are still many technological advances needed to achieve this goal. The Holographic Camera Techniques group has been working towards this end. The successful demonstration of the use of SLM's for real-time reconstruction of remotely recorded objects is a significant step forward. Although the reconstructed images at this stage are low-resolution they indicate promise. The use of digital holography for underwater imaging of biological organisms has been growing in importance and some of the techniques being developed for recording and processing of images are likely to find their way into the 3DTV domain.

2.5. Pattern Projection

The information about spatial and color coordinates of the objects, as well as of their position in the scene is essential for their 3D presentation in pseudoscopic, autostereoscopic and holographic displays. Inherent property of optical methods for parallel acquisition of information stimulated development of a various optical techniques for non-contact three-dimensional measurement of surface profiles, displacements, strain/stress in industry, fracture mechanics, machine vision, biology and medicine.

Among existing techniques, the methods which rely on a functional relationship of the sought object data with deformation of the projected onto the object encoded light pattern occupy a special place as a full-field metrological means with non-complex set-ups and processing algorithms that are easy to implement in outdoor and industrial environment. To meet the capture requirements for the needs of the 3D dynamic display the pattern projection profilometric methods and systems must ensure accurate automated real-time full-field measurement of absolute 3D coordinates in a large dynamic range without loss of information due to shadowing and occlusion. The main goal of the different teams involved in this subtopic from the 3DTV consortium is to improve the dynamic range and accuracy of the measurement in dynamic mode on the basis of existing and newly developed pattern projection systems.

The real-time acquisition of depth and color images by color-encoded structured light pattern which is under development in ITI-CERTH allows combination of 2D and 3D image processing algorithms and provides a 2D color image in addition to the coordinate's data. To ensure system performance in uncontrolled environments and with arbitrary scenes, novel approaches for encoding and recognizing the projected light are used. The goal is to make the system practically independent of intrinsic object colors and to minimize the influence of the ambient light conditions. The profilometric systems which are under development in CLOSPI-BAS rely on a sinusoidal fringe projection and a phase-stepping algorithm. A two spacing projection method for three-dimensional measurement of the absolute coordinates of real objects is proposed without a reference plane that makes the method suitable for application directly at the object location. However, complexity of the investigated relief may require solving a problem with low/no information zones due to shadowing, differences in reflection and absorption of light which may seriously hamper successful surface restoration. To avoid influence of shadowing, double symmetrical illumination/observation is experimentally tested. The selection of large observation/illumination angle and spacing difference increases sensitivity and accuracy of the measurements. The suitable compromise should be made for achievement of better precision, as the partial shadow zones recovery could appear, leading to losses of information.

Double symmetrical observation method requires more complex technical implementation than the double symmetrical illumination. Comparison of different techniques for pattern generation as using of interferometer, DMD projector or a sinusoidal phase grating proves that both DMD and a grating can be used under certain limitations. The benefits of such solutions are easy implementation and no need of moving parts for fringe density and orientation control as well as minimization of the phase-shifting error. The main drawback is a higher frequency content. A newly developed system for real-time phase stepping pattern projection profilometry is designed which is based on the usage of four phase shifted sinusoidal phase gratings. The fact that the fringe spacing in the projected patterns does not depend on the wavelength permits to illuminate the object by spatially similar fringes at different wavelengths in order to record simultaneously the deformed fringe patterns by separate four synchronized CCD cameras. This technical solution overcomes the main drawback of the temporal phase-shifting profilometry in which patterns acquisition is made successively in time. The use of four different wavelengths for pattern projection in near IR offers possibility for simultaneous measurement of color and spatial coordinates of the objects in dynamic scenes. A PhD thesis has been presented and successfully defended which is completely devoted to the problem of shape and coordinate measurement, including comparative analysis of different methods of optical profilometry.

The obtained results and findings in the frame of this topic are essential and could be used in 3D time-varying scene representation technology (WP8), 3DTV coding techniques (WP9), data translation (WP10), signal processing (WP11) and 3DTV display techniques (WP12).

2.6. Motion Analysis

A football player tracking was developed by ITI-CERTH and TUB and was applied to football sequences with satisfactory results. A small amount of information (only a few floating point numbers per track) is used to represent the players' position, track ID and classification, so it is possible to transmit these data via low bit rate channels as desirable for some viewing applications (e.g. flash animations or mobile phones).

ITI-CERTH is also developing an intelligent system for tracking moving objects (vehicles, persons etc) focusing on different traffic monitoring applications, such as the traffic control of aircraft parking areas at airports and tunnels at highways. The main contributions are mainly related to data fusion techniques: Two methods was implemented, a grid based one that is faster and requires less bandwidth and one based on probability maps which is more consuming but shows promising results.

In the case of the head captured with high resolution, the 3D motion of the head can be estimated with a model-based motion estimation algorithm. The UHANN team improved the accuracy and robustness of out-of-plane rotation estimations by updating texture information.

This research is needed for many applications such as the accurate viewpoint detection of the anchor person in news.

2.7. Registration

In the work lead by TUT, speed and robustness problems of registration techniques were tackled. This work resulted in a feature-based image registration technique that is suitable for applications on portable devices. Experimental results showed that the technique is accurate enough for super-resolution implementation, fast enough using the advantage of feature-based registration and robust to the conditions to which portable devices may be exposed.

Corner validation study presented by METU resulted in an algorithm that validates corners regardless of their orientation, contrast and corner angle. Performance of the algorithm is very good for images under mild noise conditions. Corner angle, orientation and contrast estimates are by-products of the technique which can be utilized in applications of feature point matching, object recognition, pose estimation etc. As such the algorithm could be used in a number of 3D imaging applications.

2.8. Object –Based Segmentation

The contribution in this task is a motion-based initialization of a GGM color model presented by UIL. Due to a movement of researchers at the beginning of the period there is only one publication in this task, however segmentation is an important step for parts of the 3DTV pipeline and the work will continue.

3. Abstracts of Papers and Technical Reports

3.1. Multicamera

High quality 3D modeling of real world environments is a hot research topic, as many emerging 3DTV and 3D Video applications rely on the availability of high-quality dynamic scene models. This section describes algorithms that exploit the rich information contained in multi-view video footage to reconstruct high-fidelity scene models, as well as to perform advanced 3D Video Editing tasks.

In 3.1.1 a graph-cut-based approach to omnidirectional stereo vision is described. It is well-suited for reconstructing 3D scene models of very large scenes, such as office buildings, factory or safety critical environments like airports or power plants. Up to now, reconstruction of large environments for 3D Video scenarios is a very difficult problem that has been elegantly addressed in this work.

In 3.1.2 we present a multicamera acquisition system designed at Koç University, in collaboration with our partner Yogurt Tech. The system currently uses 8 cameras. The multi-camera system has been extensively tested on a variety of scenes, and a working database of sequences was captured. A preliminary color-marker based motion capture system for articulated human body model animation was developed and tested on the database. We presented our initial results at Siggraph 2006: International Conference and Exhibition on Computer Graphics and Interactive Techniques at Boston.

By means of passive optical motion capture real people can be authentically animated and photo-realistically textured. To import real-world characters into virtual environments, however, surface reflectance properties must also be known. We describe a video-based modeling approach that captures human shape and motion as well as reflectance characteristics from a handful of synchronized video recordings in 3.1.3.

In 3.1.5 we present an enhanced approach that jointly captures shape, motion and time-varying surface reflectance of people. Using only a handful of calibrated and synchronized video recordings, our algorithm allows us to photo-realistically visualize recorded people at interactive frame rates in changing lighting conditions and from arbitrary perspective.

Sect. 3.1.5 presents a robust way to estimate time-varying surface reflectance of human actors despite a bias in captured reflectance samples. This way, faithful relighting of dynamic scenes becomes possible although the space of reflectance samples is probed unevenly.

In this section, WP7 also presents a novel versatile, fast and simple framework to generate high-quality animations of scanned human characters from input motion data in 3.1.6. Our method is purely mesh-based and, in contrast to skeleton-based animation, requires only a minimum of manual interaction. Along this line of research, in 3.1.7 we present a novel fully-automatic approach for estimating an articulated skeleton of a moving subject and its motion from body marker trajectories that have been measured with an optical motion capture system.

In section 3.1.8 we present a video processing algorithm for texture replacement of moving garments in video recordings. We use a color-coded pattern which encodes texture coordinates within a local neighborhood in order to determine the geometric deformation of the texture. Our method enables exchanging fabric pattern designs of garments worn by actors as a video post-processing step.

Many problems in computer vision can be formulated as a minimization problem for an energy functional. If this functional is given as an integral of a scalar-valued weight function over an unknown hypersurface, then the sought-after minimal surface can be determined as a solution of the functional's Euler-Lagrange equation. Section 3.1.9 contains such a formalism for Weighted Minimal Hypersurface Reconstruction.

In 3.1.10, the software synchronization of cameras is considered and a software approach to the real-time synchronization of images captured by a low-cost camera framework is proposed. This contribution permits the low-cost assembly of a camera network, eases the implementation of multi-view experiments for stereo reconstruction and 3D vision. In addition, the proposed approach can be utilized for the synchronization of wireless sensors.

3.1.1. Graph Cut based Panoramic 3D Modeling and Ground Truth Comparison with a Mobile Platform - The Wägele

Authors: Sven Fleck, Andreas Schilling

Institutions: University of Tuebingen

Publication: Sven Fleck, Florian Busch, Peter Biber and Wolfgang Strasser, *Graph Cut based Panoramic 3D Modeling and Ground Truth Comparison with a Mobile Platform - The Wägele*, in: IAPR Canadian Conference on Computer and Robot Vision (CRV 2006), 2006

Efficient and comfortable acquisition of large 3D scenes is an important topic for many current and future applications in the field of robotics, factory and office visualization, 3DTV and cultural heritage. We present both an omnidirectional stereo vision approach for 3D modeling of static scenes based on graph cut techniques and also a new mobile 3D model acquisition platform where it is employed.

The platform comprises a panoramic camera and a 2D laser range scanner for self localization by scan matching as illustrated in Fig. 2. 3D models are acquired just by moving the platform around and recording images in regular intervals. This is equivalent to a multicamera approach for static scenes, however requires only one physical camera. Besides this stereo vision approach, we concurrently build 3D models using two supplementary laser range scanners. This enables the investigation of the stereo algorithm's quality by comparing it with the laser scanner based 3D model as ground truth. This offers a more objective point of view on the achieved 3D model quality which is important for 3DTV. The acquired 3D scenes shall serve as background models where live 3D actors can be embedded which offers a great link to 3DTV participants dealing with this topic.

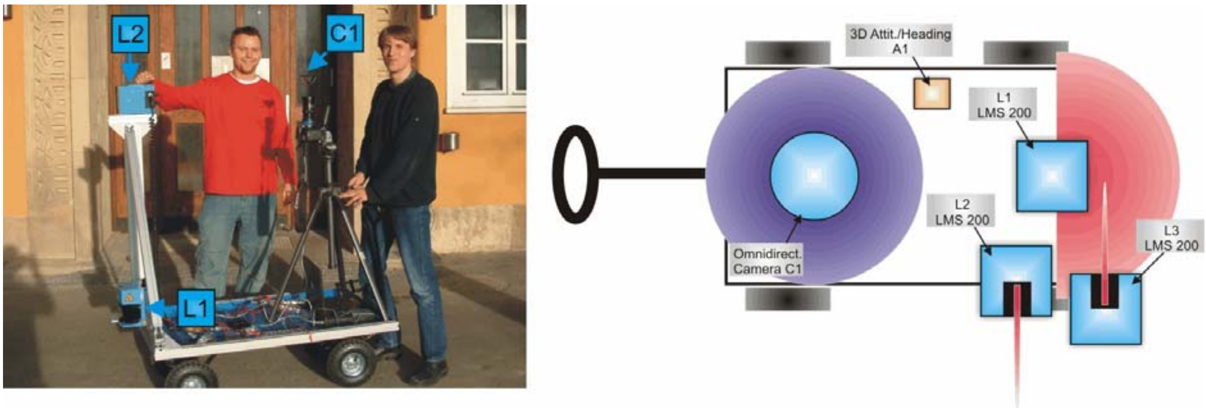


Fig. 1. *3D model acquisition platform – The Wägele.*

Fig. 2 illustrates renderings of the obtained 3D models from different viewpoints. The top row shows our graph cut based stereo vision approach, the center row shows the laser scanner based model considered as ground truth for comparison purposes and the bottom row depicts a mixed model, whereas the stereo model is tinted green and the laser model is tinted red.

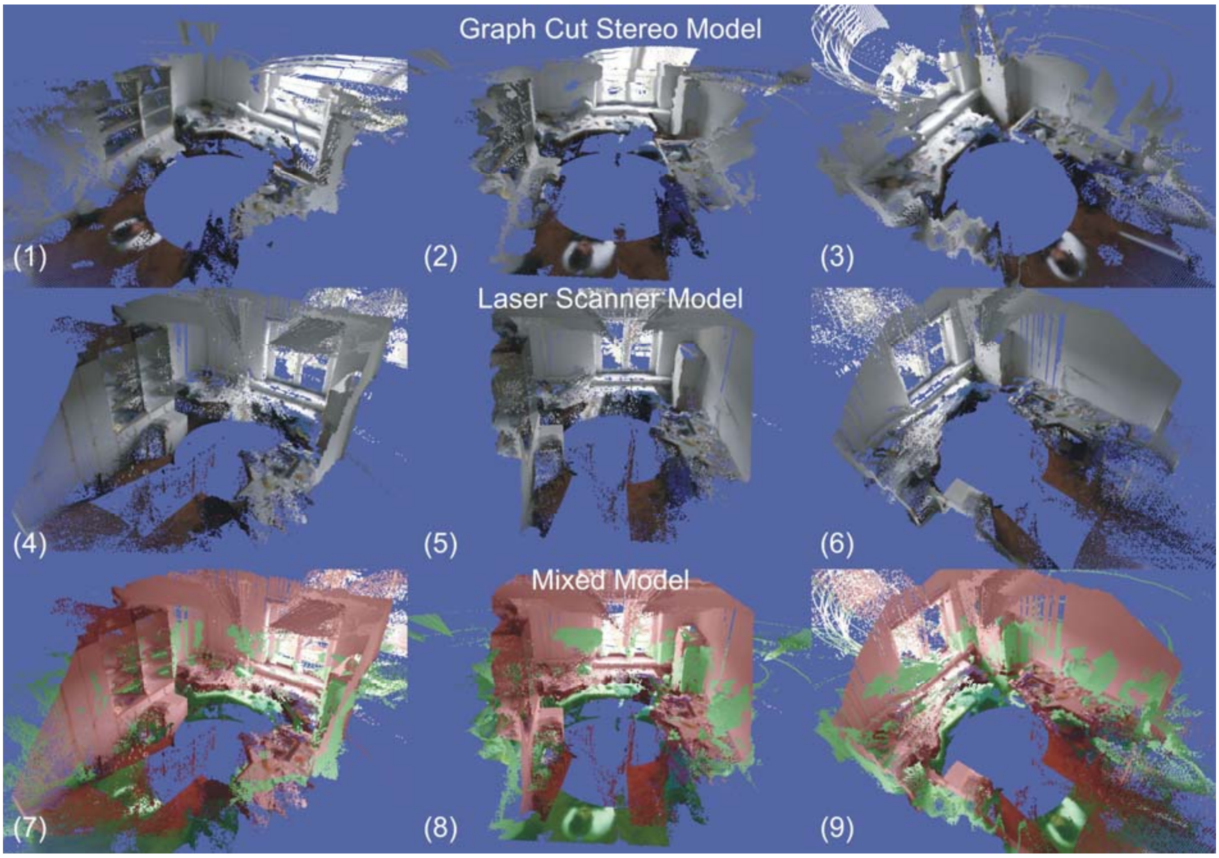


Fig. 2. *Rendering of resulting 3D point clouds from different viewpoints.*

A comparison of our graph cut (GC) method to well known winner-takes-all (WTA) stereo techniques is shown in Fig. 3, visual differences are obvious.

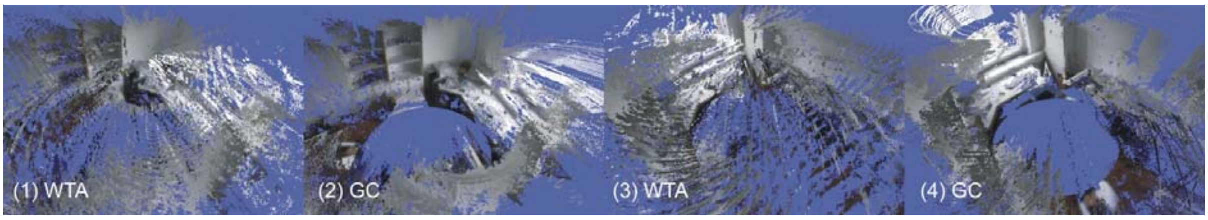


Fig. 3. *Results of the stereo processing with small (3x3) search window. (1, 3) Winner-takes-all stereo (WTA), (2 ,4) Our graph cut based method (GC).*

3.1.2. Multicamera Motion Capture for Articulated Body Model Animation

Authors: Ferda Ofli, Yücel Yemez, A. Murat Tekalp, Can J. Wetherilt, Cemil Türün

Institutions: Koç University, Yogurt Tech.

Publication: Internal report to be submitted later

We present an automated multicamera motion capture system for articulated human body model animation. The multiview video of a moving actor is acquired using 8 synchronized cameras. The proposed motion capture technique is based on 3D tracking of the markers attached to the person's body in the scene, using stereo color information without need for an explicit 3D model. The resulting set of 3D points is then used to animate a personalized skeleton-based 3D human body model. Preliminary results on body motion feature extraction, tracking and animation are provided.

3.1.3. Relightable Free-Viewpoint Video of Human Actors

Authors: C. Theobalt, N. Ahmed, Edilson de Aguiar, Gernot Ziegler, H. Lensch, M. Magnor, H.-P. Seidel

Institutions: MPI Informatik

Publication: technical report [2]

By means of passive optical motion capture real people can be authentically animated and photo-realistically textured. To import real-world characters into virtual environments, however, surface reflectance properties must also be known. We describe a video-based modeling approach that captures human shape and motion as well as reflectance characteristics from a handful of synchronized video recordings. The presented method is able to recover spatially varying surface reflectance properties of clothes from multi-view video footage. The resulting model description enables us to realistically reproduce the appearance of animated virtual actors under different lighting conditions, as well as to interchange surface attributes among different people, e.g. for virtual dressing. Our contribution can be used to create 3D renditions of real-world people under arbitrary novel lighting conditions on standard graphics hardware.

(a) (b)

Fig. 4. *Screenshots of rendered relightable free-viewpoint videos*

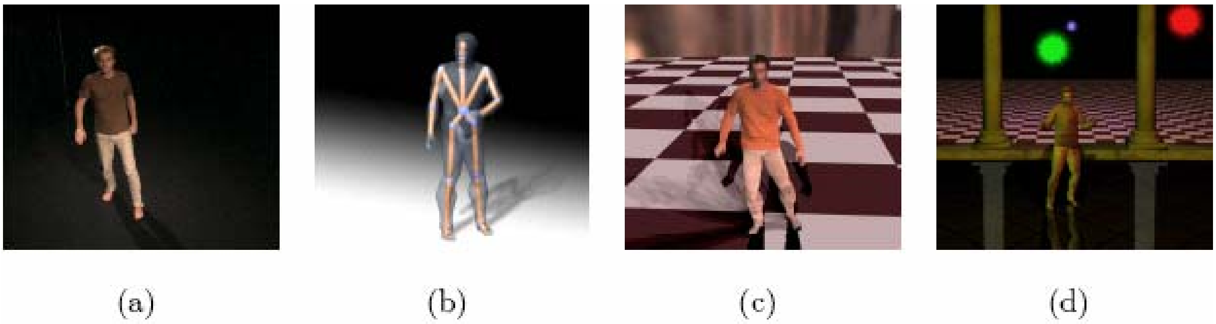


Fig. 5. *Our method jointly estimates shape, motion and dynamic surface reflectance of a human actor from input video (a). A dynamic scene description based on a template body model (b) enables us to photo-realistically render moving actors from arbitrary viewpoint and under both captured real-world (c) and artificial illumination (d).*

3.1.4. Enhanced Dynamic Reflectometry for Relightable Free-Viewpoint Video

Authors: Christian Theobalt, Naveed Ahmed, Hendrik Lensch, Marcus Magnor, Hans-Peter Seidel.

Institutions: Max Planck Institut Informatik

Publication: Seeing People in Different Light: Joint Shape Motion and Reflectance Capture, TVCG

Free-Viewpoint Video of Human Actors allows photo-realistic rendering of real-world people under novel viewing conditions. Dynamic Reflectometry extends the concept of free-view point video and in addition allows rendering under novel lighting conditions. In this work, we present an enhanced method for capturing human shape and motion as well as dynamic surface reflectance properties from a sparse set of input video streams. We augment our initial method for model-based relightable free-viewpoint video in several ways. Firstly, a single-skin mesh is introduced for the continuous appearance of the model. Moreover, an algorithm to detect and compensate lateral shifting of textiles in order to improve temporal texture registration is presented. Finally, a structured resampling approach is introduced which enables reliable estimation of spatially varying surface reflectance despite a static recording setup. The new algorithm ingredients along with the Relightable 3D Video framework enables us to realistically reproduce the appearance of animated virtual actors under different lighting conditions, as well as to interchange surface attributes among different people, e.g. for virtual dressing. Our contribution can be used to create 3D renditions of real-world people under arbitrary novel lighting conditions on standard graphics hardware. Fig. 4 shows two exemplary renderings of reconstructed human actors under captured real world illumination, (a), and spot light illumination as it is typical for computer games, (a).

3.1.5. Spatio-temporal Reflectance Sharing for Relightable 3D Video

Authors: Naveed Ahmed, Christian Theobalt, and Hans-Peter Seidel

Institutions: Max-Planck Institut für Informatik

Publication: To appear in the proceedings of MIRAGE 2007

In our previous work we have shown that by means of a model-based approach, relightable free-viewpoint videos of human actors can be reconstructed from only a handful of multi-view videostreams recorded under calibrated illumination. To achieve this purpose, we employ a marker-free motion capture approach to measure dynamic human scene geometry. Reflectance samples for each surface point are captured by exploiting the fact that, due to the person's motion, each surface location is, over time, exposed to the acquisition sensors under varying orientations. Although this is the first setup of its kind to measure surface reflectance from footage of arbitrary human performances, our approach may lead to a biased sampling of surface reflectance since each surface point is only seen under a limited number of half-vector directions. We thus propose in this paper a novel algorithm that reduces the bias in BRDF estimates of a single surface point by cleverly taking into account reflectance samples from other surface locations made of similar material. We demonstrate the improvements achieved with this spatio-temporal reflectance sharing approach both visually and quantitatively.

3.1.6. A Framework for Natural Animation of Digitized Models

Authors: Edilson de Aguiar, Rhaleb Zayer, Christian Theobalt, Marcus Magnor and Hans-Peter Seidel

Institutions: Max-Planck Institut für Informatik and TU Braunschweig

Publication: Technical Report at Max-Planck Institut für Informatik [3]

We present a novel versatile, fast and simple framework to generate high-quality animations of scanned human characters from input motion data. Our method is purely mesh-based and, in contrast to skeleton-based animation, requires only a minimum of manual interaction. The only manual step that is required to create moving virtual people is the placement of a sparse set of correspondences between triangles of an input mesh and triangles of the mesh to be animated. The proposed algorithm implicitly generates realistic body deformations, and can easily transfer motions between human subjects of completely different shape and proportions. Our approach handles many different types of input data, e.g. other animated meshes and motion capture data, in just the same way. Finally, and most importantly, it creates animations at interactive frame rates. We feature two working prototype systems that demonstrate that our method can generate lifelike character animations from both marker-based and marker-less optical motion capture data.

3.1.7. Automatic Learning of Articulated Skeletons from 3D Marker Trajectories

Authors: Edilson de Aguiar, Christian Theobalt and Hans-Peter Seidel

Institutions: Max-Planck Institut für Informatik

Publication: Proceedings of the 2nd International Symposium on Visual Computing [4]

We present a novel fully-automatic approach for estimating an articulated skeleton of a moving subject and its motion from body marker trajectories that have been measured with an optical motion capture system. Our method does not require a priori information about the shape and proportions of the tracked subject, can be applied to arbitrary motion sequences, and makes dedicated initialization poses unnecessary. To serve this purpose, our algorithm first identifies individual rigid bodies by means of a variant of spectral clustering. Thereafter, it determines joint positions at each time step of motion through numerical optimization, reconstructs the skeleton topology, and finally enforces fixed bone length constraints. Through experiments, we demonstrate the robustness and efficiency of our algorithm and show that it outperforms related methods from the literature in terms of accuracy and speed.

3.1.8. Texture Replacement of Garments in Video

Authors: Volker Scholz and Marcus Magnor

Institutions: MPI Informatik and TU Braunschweig

Publication: Rendering Techniques 2006 (Proc. of the 17th Eurographics Symposium on Rendering EGSR 2006) [6]

In this paper, we present a video processing algorithm for texture replacement of moving garments in video recordings. We use a color-coded pattern which encodes texture coordinates within a local neighborhood in order to determine the geometric deformation of the texture. A time-coherent texture interpolation is obtained by the use of 3D radial basis functions. Shading maps are determined with a surface reconstruction technique and applied to new textures which replace the color pattern in the video sequence. Our method enables exchanging fabric pattern designs of garments worn by actors as a video post-processing step.

(a) (b)

Fig. 6. *Input video frame (a) and retextured result (b).*

3.1.9. Weighted Minimal Hypersurface Reconstruction

Authors: Bastian Goldluecke, Ivo Ihrke, Christian Linz, Marcus Magnor

Institutions: Max-Planck-Institut fuer Informatik

Publication: accepted for publication PAMI

Many problems in computer vision can be formulated as a minimization problem for an energy functional. If this functional is given as an integral of a scalar-valued weight function over an unknown hypersurface, then the sought-after minimal surface can be determined as a solution of the functional's Euler-Lagrange equation. This paper deals with a general class of weight functions that may depend on surface point coordinates as well as surface orientation. We derive the Euler-Lagrange equation in arbitrary dimensional space without the need for any surface parameterization, generalizing existing proofs. Our work opens up the possibility to solve problems involving minimal hypersurfaces in dimension higher than three, which were previously impossible to solve in practice. We also introduce two applications of our new framework: we show how to reconstruct temporally coherent geometry from multiple video streams, and we use the same framework for the volumetric reconstruction of refractive and transparent natural phenomena, here bodies of flowing water.

3.1.10. Synchronous Image Acquisition based on Network Synchronization

Authors: Georgios Litos, Xenophon Zabulis and Georgios Triantafyllidis

Institutions: Informatics and Telematics Institute, Thessaloniki, Greece

Publication: 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 real-time 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 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.

3.1.11. Conclusions and future plans

In 3.1.1 we have presented a hassle free method and system for automatically generating realistic looking 3D models from omnidirectional images. All that is needed is a high quality panoramic camera and a mechanism to determine its external camera parameters. The introduced stereo method uses the combination of omnidirectional vision and graph cut methods, followed by several post-processing steps. We have designed a simple yet highly flexible sensor platform (The Wägele). It can be transported easily and is capable of acquiring both indoor and outdoor scenes. The visual quality of the resulting models is quite satisfying and far better than simple winner-takes-all stereo. It turned out that the results are relatively robust against parameter variations of the graph cut's energy function, so no scene specific adjustments have to be conducted. The main contribution of this research is an assessment of the quality of our graph cut based stereo method with respect to the needs of 3DTV. To measure the quality of our stereo method, additional laser range scanners are mounted on the Wägele to acquire ground truth depth information. The results of our stereo method have been compared to the point clouds of the laser scanners in 3D. The quality of our 3D models is quite convincing, both visually and with respect to ground truth. So although laser scanners have not become obsolete yet, the qualities of stereo methods seem to approach the laser scanner based models. However, the algorithm is relatively slow (about 30 minutes per image on a 3GHz machine). Additionally, we have to deal with vibrations of the platform in the future. Finally, the dynamic range of the camera is still limited, so using high dynamic range (HDR) imaging for stereo is future work we already started.

In 3.1.2, we have developed a marker-based motion capture system for articulated body animation. Based on preliminary experiments and obtained results, we have observed that the current system has two major problems that will be addressed in future work. The first problem is due to poor visibility of a marker and the second is due to abrupt motions that may occur in the scene. To handle these two problems, we currently allow users to intervene the process of tracking, if necessary, throughout an interface. Our future work will include introducing a skeleton model to the motion capture task in order to be able to make use of spatial relations of the markers on the body. The skeleton model will give us the opportunity to utilize the properties of kinematic chains in the body, as well. We will, furthermore, use this idea in our filtering process to change the state variables of Kalman filter from positions and velocities to angles and angular velocities. This will bring robustness and sophisticated error recovery functionality to our system. The obtained time-varying skeleton model will then be used to animate a 3D human body model.

In 3.1.3 we have developed a novel algorithm for realistic capturing of shape, motion and dynamic surface reflectance of human actors. Our commitment to an adaptable template body model enables us to achieve this goal using only eight synchronized video streams as input. We are currently improving our dynamic reflectance estimation procedure such that it can capitalize on spatial and temporal coherence in the reflectance data. Also, we work on methods to handle more general surface geometry. In particular, we plan to couple our method with a cloth tracking approach and work on novel tracking algorithms for high-quality laser-scanned surface geometry.

Our video-based modeling approach in 3.1.4 jointly captures shape, motion and surface reflectance of a person. From eight synchronized multi-video streams, we recover all information necessary to photo-realistically render a recorded person from arbitrary viewpoint and in arbitrary illumination. Despite a bias in the reflectance samples, we can robustly estimate dynamic reflectance parameters by employing a dynamic reflectance sharing approach, Sect. 3.1.5. The ability to perform convincing relighting enables us to implant real-world, animated people into virtual surroundings. The abstract description of people's appearance in terms of geometry, animation and surface reflectance further allows us to separate surface appearance from geometry. This way, we can interchange surface attributes among different people, e.g., for dressing one person with another person's clothes. Moreover, we employ a compact data format for our scene description that can be acquired with only a handful of imaging sensors. Joint shape, motion and reflectance capture cannot only be applied to humans but to any dynamic object whose motion can be described by a kinematic chain and for which a suitably parameterized geometry model is available. For BRDF parameter recovery, the proposed algorithm currently assumes that the subject is illuminated by a small number of spot lights. In the future, we would like to extend it to use more general lighting configurations, such as the light stage [7]. We would also like to extend our work in the direction of no a-priori geometry model and only use multi-view videos for all the estimations.

The novel animation framework in 3.1.5 simplifies the traditional, not so straightforward acquisition-to-animation pipeline. The proposed method is easy and intuitive to use and does not require any training. By means of the same efficient methodology the method solves the animation, the surface deformation and the motion retargeting problem. As a direction for future work, we would like to combine our technique with an approach to learn per-time-step surface deformations from input video footage. Also, we plan to merge this novel animation approach with the work on dynamic surface reflectance capture to generate even more natural and lifelike characters for 3D Video environments.

The fully-automatic system in 3.1.7 is able to learn an articulated skeleton model with constant bone lengths and its poses from 3D marker trajectories. It works with no a priori information about the kinematics of the captured individual and can be applied to arbitrary subjects including humans and animals. Through experimental evaluation we have shown that it performs better in terms of speed and accuracy than the most closely related methods from the literature. The learned models are comparable to the ones obtained with commercial software in terms of accuracy and detail. As future work, we plan to integrate our method with an automatic non-intrusive surface reconstruction approach in order to automatically learn complete virtual characters.

In 3.1.8 we have presented a system for automatic texture replacement of color-coded garments. Visually convincing replacement results are obtained by using 3D spatiotemporal RBF approximation. We also show that our shading maps can capture small details at cloth folds. Currently, the performance bottleneck in our implementation is the computation of the shading maps. Using faster solvers (e.g. multigrid) would improve our system's time and memory consumption.

As future work along the line of research in 3.1.9, we plan to include a global optimization of surface reflectance properties into the same unifying framework. Future work also includes research into the applicability of our method to the reconstruction of other refractive media. Next, we intend to develop a hierarchical representation of the underlying computational grid to achieve higher resolution reconstruction which allows resolving finer details.

Although providing the most precise synchronization, external triggering as proposed in 3.1.10 requires special cameras that can receive and recognize such input as well as additional hardware and wiring. This increases the camera's cost and casts hardware-synchronization weak in term of cost-efficiency. Moreover, in several types of applications independence from physical links (cable) is required due to the large distance between the cameras (e.g. ground traffic control, etc). This independence can also facilitate the synchronization of cameras mounted on wireless stations, suitable for mobile applications.

3.2. Single Camera

The studies on 3D reconstruction from monocular sequences currently aim to convert 2D video sequences into 3D, without any prior information on scene and imaging equipment. The solution of this problem is the corner stone of some important applications such as automatic conversion of visual media archives into 3D, and implementation of a 3D TV, utilizing the existing 2D TV network. To accomplish this goal, so far, three sub-problems received our attention:

- Decomposition of the multi-frame-multi-object structure from motion problem into single-object problems by segmentation of the features into spatially and geometrically consistent sets.
- Design of a sequential 3D reconstruction algorithm capable of identifying a pair processing order that both emphasizes reliability and rapid exploration of the structure for each reconstruction problem.
- Design of a proof-of-concept, automatic 2D/3D conversion algorithm for 3D reconstruction from broadcast 2D video, utilizing the tools described above, plus a self-calibration, and a dense reconstruction module.

Another focus of the research efforts is the construction of high dynamic range images. An ultimate goal of 3DTV is the acquisition and visualization of high quality 3D content of both indoor and outdoor scenes. This raises the problem of achieving high quality imaging of differently illuminated objects. High dynamic range (HDR) imaging aims at solving this problem by taking multiple images with different exposure and combining them into a single HDR image. Image modeling, image registration and edge detection are three fields that benefit from the results of this study.

The work on these topics is presented below.

3.2.1. A Geometric Segmentation Approach for the 3D Reconstruction of Dynamic Scenes in 2D Video Sequences

Authors: Sebastian Knorr, Evren Imre†, Aydin A. Alatan†, and Thomas Sikora

Institutions: Technical University of Berlin , † Middle East Technical University

Publication: EUSIPCO, Florence, Italy, 04.09.2006 - 08.09.2006

Structure from motion in static scenes is an extensively studied problem with some well established solutions [11]. However, these solutions are not capable of dealing with dynamic scenes with multiple moving objects, which are often encountered in practice. Hence, the intention of this study is to achieve both an accurate segmentation and reconstruction of the whole 3D scene including the dynamic elements.

In the literature, analysis of video sequences of dynamic scenes falls under the category of the multi-body MFSfM (multi-frame structure from motion) problem, which has the following definition for this special case:

Given a set of N features belonging to a background scene and K independently moving objects (IMOs), L views and the correspondence information, estimate the locations of the feature points in 3D world coordinates and the external calibration parameters.

Once the feature set is segmented into partitions corresponding to the background and the individual objects, the problem can be decomposed into several static MFSfM problems. In this work, we focused on the segmentation of multiple independently moving objects and on their reconstruction using a prioritized MFSfM approach.

3.2.2. Prioritized Sequential 3D Reconstruction in Video Sequences with Multiple Motions

Authors: Evren Imre, Sebastian Knorr †, Aydin A. Alatan, and Thomas Sikora†

Institutions: Middle East Technical University, † Technical University of Berlin

Publication: IEEE Int. Conf. on Image Processing (ICIP'06), Atlanta, GA, USA, 08.10.2006 - 11.10.2006

In this study, an algorithm is proposed to solve the multi-frame structure from motion (MFSfM) problem for monocular video sequences in dynamic scenes. The algorithm uses the epipolar criterion to segment the features belonging to the independently moving objects. Once the features are segmented, corresponding objects are reconstructed individually by using a sequential algorithm, which is also capable of prioritizing the frame pairs with respect to their reliability and information content, thus achieving a fast and accurate reconstruction through efficient processing of the available data A tracker is utilized to increase the baseline distance between views and to improve the F-matrix estimation, which is beneficial to both the segmentation and the 3D structure estimation processes. The experimental results demonstrate that our approach has the potential to effectively deal with the multi-body MFSfM problem in a generic video sequence.

3.2.3. A Modular Scheme for 2D/3D Conversion of TV Broadcast

Authors: Sebastian Knorr, Evren Imre†, Burak Özkalayci†, Aydin A. Alatan†, and Thomas Sikora

Institutions: Technical University of Berlin , † Middle East Technical University

Publication: 3rd International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT'06), Chapel Hill, USA, 14.06.2006 - 16.06.2006

3DTV technology is currently being investigated in many research labs worldwide [8]. In this context, conversion of existing 2D video material to 3D is of much interest. Many fundamental algorithms have been developed to reconstruct 3D scenes from an uncalibrated video sequence [9][10]. However, most of these approaches deal with the reconstruction of static scenes . When the scene is dynamic, i.e., contains independently moving objects (IMO), they usually fail because the triangulation techniques used for the reconstruction can only deal with one single relative motion.

In this work, a complete system for the 3D reconstruction of a scene from broadcast video is proposed. The input to the system is an uncalibrated 2D video sequence captured from typical TV broadcast, and the output is a dense 3D reconstruction of the scene observed in the sequence. In addition to some standard approaches, the system also delivers novel ideas for moving object segmentation and a new prioritized sequential algorithm for sparse reconstruction, respectively. An overview of the complete 2D/3D conversion scheme is illustrated in Figure 3.1.

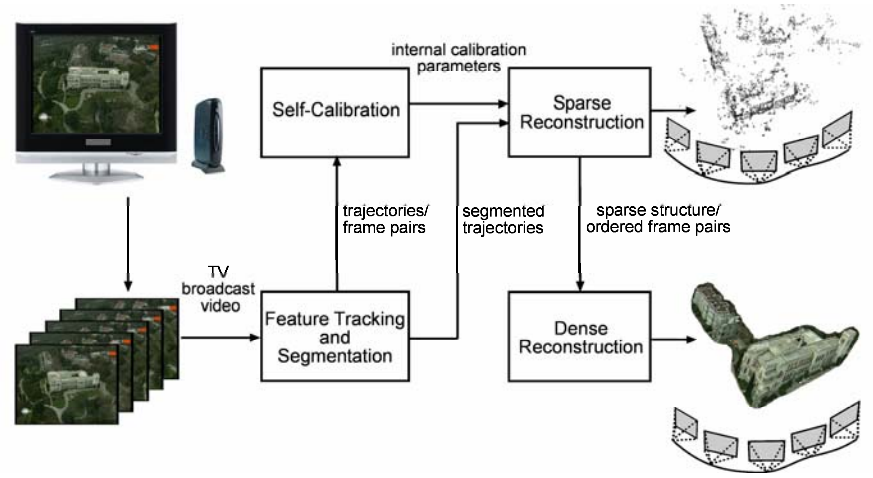


Figure 3.1: Block diagram of the proposed system

The proposed system is composed of four modules: Feature tracking and segmentation, self-calibration, camera pose estimation and sparse 3D reconstruction, and finally, dense reconstruction. Each module deals with a specific task and is the subject of a distinct field of research.

3.2.4. Color High Dynamic Range (HDR) Imaging in Luminance-Chrominance Space

Authors: Ossi Pirinen, Alessandro Foi, Atanas Gotchev

Institutions: Tampere University of Technology

Publication: 3rd International Workshop on Video Processing and Quality Metrics for Consumer Electronics VPQM-07, Scottsdale, Arizona, USA, Jan. 25-26, 2007.

Natural 3D scenes may have a very broad radiance dynamic range, which conventional cameras cannot capture without saturation. Therefore, the problem of recovering a high dynamic range (HDR) image from multiple exposures of a low dynamic range (LDR) imaging device has gotten increasing attention from the digital imaging community. Many image processing applications, such as edge detection, image registration, and image-based modeling benefit from HDR as it provides the true scene radiance instead of the distorted (non-linearly mapped) one normally generated by the LDR device. This is especially important for methods, which require obtaining true surface radiance in extremely different lighting conditions.

The methods currently employed for multiple exposure HDR capture date back to the first half of the 1990's and for gray-scale images they are essentially mature. As far as the issue of color is concerned, all known methods operate in RGB space. This however raises some problems, since when processing the three RGB channels independently one always ends up distorting the colors, thus creating a need for white-balancing and possibly more corrective procedures. From another side, much of the processing in the imaging devices as well as final storage of the images is done in a luminance-chrominance space. In our current research, we address the issue of color in HDR imaging and suggest moving the whole process from HDR composition to tone mapping from RGB to a luminance-chrominance space. We argue that processing in RGB introduces two redundant transformations from space to another, and may even cause unwanted artifacts. Moreover, processing in luminance-chrominance space saves us from dealing with three different layers.

3.2.5. Conclusions and future plans

Geometric segmentation successfully deals with the decomposition of the multiple-object MFSfM problems into several single-object ones. The reconstruction procedure presents a powerful and efficient information integration scheme, however a better understanding of the prioritization metric requires more work. Finally, 2D/3D conversion without any prior information is demonstrated to be possible, but computationally efficient determination of the key frames for self-calibration and prioritization remains a challenge.

Three topics are likely to be in the focus of the future research efforts in 2D/3D conversion. The first one is, obviously, the shortcomings of the current algorithm, explained in the previous paragraph. The second one is endowing the algorithm with the capability of interaction between tracking and reconstruction modules, to improve the reliability of the tracks by enforcing geometric constraints, and in turn to improve the reliability of the reconstruction. And finally, the capabilities of piecewise planar reconstruction as an alternative dense reconstruction representation is to be explored.

In the next research period, HDR imaging will be related with the problem of 3D model acquisition, where the aim is to provide the 3D background models where potential 3D actors can be embedded. A platform for 3D model acquisition has been developed by the 3DTV team of University of Tuebingen. In their system, the 3D modeling critically relies on the quality of the images taken with an omnidirectional camera having limited dynamic range. Current results will be extended toward automated acquisition of omnidirectional images in high dynamic range (HDR) to enable HDR based 3D model acquisition.

3.3. Human Face and Body

Capturing a 3D structure of human face and body is a very important research area since the human face and body are integral parts of almost any 3DTV application. The a priori knowledge about 3D structure and motion of human face and body, nature and limits of human motions can be used in order to make the processing more efficiently. Case specific algorithms can perform better than general purpose algorithms. For 3D display systems, detection and tracking of observer's eyes and observer's view point is necessary to render the correct view according to the observer position. The face and facial feature localization and tracking algorithms are also important for progress in the field of Human-Computer Interaction (HCI). In this section the human face and body specific techniques are presented. The section is organized as follows. In the subsection 3.3.1 an algorithm for adaptation of generic 3D face model to human face of speaking person in video is presented. In the subsection 3.3.2 an application of active appearance model for facial feature detection is studied. Subsection 3.3.3 describes 3D face recognition algorithm by point signatures and iso-contours. In the subsection 3.3.4 a new framework for joint analysis of head gesture and speech prosody patterns of a speaker towards automatic realistic synthesis of head gestures from speech prosody is presented. Subsection 3.3.5 devotes 3D face motion capture and in subsection 3.3.6 a method for face detection and eye localization is described. The summary of presented results is given in the subsection 3.3.7.

3.3.1. Automatic adaptation of face model to human face in video

Authors: S. Piekh, T.Bujalski

Institutions: UHANN (Leibniz University of Hannover)

Publication: report

In this contribution an algorithm for automatic adaptation of a generic 3D face model is presented. The algorithm adapts a given 3D face model to the individual face of the speaker in a video sequence using facial feature points detected in previous analysis steps. The algorithm contains two steps: global and local adaptation. The global adaptation uses relative positions of facial feature points to each other in order to find the rough position and scale of the face model in the scene. The local adaptation deforms the face model in order to exactly describe the face geometry of the speaking person. The advantage of the proposed adaptation algorithm is that after the adaptation the face model remains symmetric. Deviation between the calculated and detected facial feature points is used to evaluate the quality of the adaptation algorithm. The subjective quality of the adaptation is evaluated by synthesis of the face in different views and synthesis of different facial expressions using the adapted face model.

3.3.2. Active Appearance Models for Facial Features Detection

Authors: Xinghan Luo, Kang Liu

Institutions: TUT, UHANN (Leibniz University of Hannover)

Publication: report

Computer aided modeling still requires a great deal of expertise and manual control to achieve realistic animations. Humans are very sensitive to any abnormal lineaments, so that body and facial animation remains a challenging task until this day. Therefore a lot of research must still be carried out to achieve photo-realistic facial animations. Facial features are used to analyze the facial motion of humans. The motion of these features determines the animation. Hence, feature detection has to be performed accurately. We aim at developing algorithms detecting the mouth corners, lip contour, and jaw contour with sub pixel accuracy. More specifically, we rely on Active Appearance Models (AAM) to detect these features automatically. Building proper PCA spaces to describe adequately the shape and the texture of the detected features is of crucial importance. In the final report, the chosen approaches and comparative results will be given.

3.3.3. 3D Face Recognition by point signatures and iso-contours.

Authors: Iordanis Mpiperis, Sotiris Malassiotis and Michael G. Strintzis

Institutions: ITI-CERTH

Publication: internal report

This work addresses the problem of face recognition from range images. A novel technique based on matching of level contours is proposed and compared with a variant of the point signatures algorithm. Their efficiency is investigated on conditions of changes in expression and pose, and presence of glasses. Using a large database of range images, comparative experimental results are presented, showing that iso-contours outperform point signatures both in computational efficiency and in recognition rates. A conference publication is intended, while collaboration in the field with other 3DTV partners is sought.

3.3.4. Speech/Gesture Correlation Analysis for Prosody-Driven Head-Gesture Animation

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

Institutions: Koc University, Momentum A. S.

Publication: to be presented at IEEE Conf. On Acoustics, Speech and Signal Processing (ICASSP 2007)

State of the art visual speaker animation methods are capable of generating synchronized lip movements automatically from speech content; however, they lack automatic synthesis of speaker gestures from speech. Head and face gestures are usually added manually by artists, who is costly and often look unrealistic. Hence, learning the correlation between gesture and speech patterns of a speaker towards automatic realistic synthesis of speaker gestures from speech remains a challenging research problem.

We present a new framework for joint analysis of head gesture and speech prosody patterns of a speaker towards automatic realistic synthesis of head gestures from speech prosody. There are some open challenges involved in the joint analysis of head gestures and speech prosody: First, there does not exist a well-established set of elementary prosody and gesture patterns for gesture synthesis, unlike phonemes and visemes in speech articulation. Second, prosody and gesture patterns are speaker dependent, and may exhibit variations in time even for the same speaker. Third, synchronicity of gesture and prosody patterns may exhibit variations. We address these challenges by processing the head gesture and prosody features separately by a parallel HMM structure to learn and model the gestural and prosodic elements (elementary patterns), respectively, over training data for a particular speaker. We then employ a multi-stream parallel HMM structure to find the jointly recurring gesture-prosody patterns and the corresponding audio-to-visual mapping. The resulting audio-visual mapping model is then employed to synthesize natural head gestures from arbitrary input test speech given a head model for the speaker. 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.

3.3.5. 3D Face Motion Capture

Authors: Tanju Erdem, Çiğdem Eroğlu Erdem

Institutions: Momentum A.Ş.

Publication: internal report

A method for tracking the motion of a person's face for the purpose of animating a 3D face model of the same or another person is presented. The 3D face model carries both the geometry (shape) and the texture (color) characteristics of the person's face. The shape of the face model is represented via a 3D triangular mesh (geometry mesh), while the texture of the face model is represented via a 2D composite image (texture image). Both the global motion and the local motion of the person's face are tracked. Global motion of the face involves the rotation and the translation of the face in 3D. Local motion of the face involves the 3D motion of the lips, eyebrows, etc., caused by speech and facial expressions. The 2D positions of salient features of the person's face and/or markers placed on the person's face are automatically tracked in a time-sequence of 2D images of the face. Global and local motion of the face is separately calculated using the tracked 2D positions of the salient features or markers. Global motion is represented in a 2D image by rotation and position vectors while local motion is represented by an action vector that specifies the amount of facial actions such as smiling-mouth, raised-eyebrows, etc.

3.3.6. Face Detection and Eye Localization in Video

Authors: Mehmet Turkan, A. Enis Cetin

Institutions: Bilkent University

Publication: Face Detection is presented in EUSIPCO 2006; Eye Localization will be submitted to IbPRIA 2007

A human face detection and eye localization algorithm in images and video is presented. After determining all possible face candidate regions using color information in a given still image or video frame, single-stage 2D rectangular wavelet transform of each region is computed. In this way, wavelet domain sub-images are obtained. The low-high and high-low sub-images contain horizontal and vertical edges of the region, respectively. The high-high sub-image may contain almost all the edges, if the face candidate region is sharp enough. It is clear that the detail information within local facial areas, e.g., eyes, nose, and mouth, show noticeable discrimination ability for face detection problem of frontal view faces. We take advantage of this fact by summarizing these sub-images using their projections and obtain 1-D projection feature vectors corresponding to edge images of face or face-like regions. The advantage of the projections is that they can be easily normalized to a fixed size and this provides robustness against scale changes. A binary edge image is obtained by thresholding the absolute value of the high-high sub-image. The horizontal and vertical projections are simply computed by counting the pixels in any row and column in the binary edge image, respectively. Furthermore, filter-like projections are computed as additional feature vectors, similar to Viola and Jones approach. The final feature vector for a face candidate region is obtained by concatenating all the horizontal, vertical, and filter-like projections. These feature vectors are then classified using a support vector machine based classifier into face or non-face classes.

After detecting a human face by the face detector, the first step of the eye localization algorithm is to determine the approximate horizontal position of eyes using horizontal-cut edge image with its horizontal projection. Horizontal-cut edge image provides some robustness against the effects of edges of neck, mouth (teeth), and nose on the horizontal projections as the eyes are located in the upper part of a human face. The global maximum index in the horizontal projection after low-pass smoothing helps us to determine the approximate horizontal location of eyes. After determining the approximate horizontal position of eyes, the edge image is again cut vertically using a threshold in pixels, depending on the resolution of the input signal around the horizontal position. The second step is to compute both horizontal and vertical projections in both right-eye and left-eye vertical-cut regions separately. The eye candidate points are obtained by pairing up the peak value indices occurring in the associated projections for each eye. All the candidate points with a bounding rectangle depending on the resolution are then classified using a support vector machine based classifier. The feature vectors for each eye candidate region are calculated similar to the face detection algorithm by concatenating only the horizontal and vertical projections of the regions. The points that are classified as an eye are finally averaged in order to have exact positions of eyes. This procedure may be applied multiple times, if the first-pass is not satisfactory because of the sharpness of the face region, using another threshold value to obtain the binary edge image.

3.3.7. Conclusions and future plans

In this section the human face and body related techniques were presented. An improved method for adaptation of a generic 3D face model to human face of speaking person was described in subsection 3.3.1. The global and local adaptation of the face model is performed. The experimental results show a good quality of face model adaptation. The adapted face model remains symmetrical. The adapted face model can be used for future face recognition, facial expression analysis, facial animation or face motion analysis and tracking.

In subsection 3.3.2 is demonstrated the potential of the AAM-based facial feature detection. The basics of AAM method have been studied. An AAM model has been built using a training set and manually selected landmarks. The test of new video frame sequences with the subject (the same as in training) shows AAM method's fast, robust and accurate detection for facial features. In the future work, one will further investigate and modify the original AAM method, implement an automatic AAM-based detection system and adapt it to facial animation or other 3DTV applications. The detector should be also extended to work for a general set of faces/facial features.

In the work carried out by the ITI-CERTH in subsection 3.3.3 is demonstrated that iso-contours are capable to capture the discriminative information about the facial surface structure, which is suitable for recognition. ITI-CERTH is planning to extend their iso-contour work for facial surface structure to allow for images depicting facial expressions. The current algorithm does not handle this. Overcoming this problem is the target of the future work.

Momentum and Koc University presents a new two-stage joint head gesture and speech prosody analysis framework to drive automatic realistic synthesis of head gestures from speech prosody in subsection 3.3.4. The proposed two-stage analysis framework offers the following advantages: a) meaningful elementary gesture and prosody patterns are defined for a speaker in the first stage, b) a mapping between these elementary prosody and head gesture patterns is obtained with the unsupervised segmentation of joint gestureprosody label stream, c) the HMM-based analysis and synthesis yields flexibility in modeling structural and durational variations within gestural and prosodic patterns, d) automatic generation of the elementary gesture patterns produces natural looking prosodydriven head gesture synthesis. In the future, this method will be extended to estimate not only the head gesture but also the facial expressions from a given input speech.

Momentum presents also in subsection 3.3.5 a 3D face motion capture technique. Both the global motion and the local motion of the person's face are tracked. Global motion of the face involves the rotation and the translation of the face in 3D space. Local motion of the face involves the 3D motion of the lips, eyebrows, etc., caused by speech and facial expressions. The 2D positions of salient features of the person's face and/or markers placed on the person's face are automatically tracked in a time-sequence of 2D images of the face. Global and local motion of the face is separately calculated using the tracked 2D positions of the salient features or markers. Global motion is represented in a 2D image by rotation and position vectors while local motion is represented by an action vector that specifies the amount of different facial actions. The performance of the proposed algorithm will be demonstrated via experimental results.

In subsection 3.3.6 a human face detection and eye localization algorithm in images and video is presented. In this work, a wavelet domain signal and image processing algorithms for two specific applications is presented: (a) detection of human faces in images and video, and (b) localization of eyes in the detected face regions. The proposed human face detection algorithm is based on edge projections of face candidate regions in color images and video. The algorithm segments skin color regions out, extracts feature vectors in these regions and detects frontal upright faces using support vector machine based classifiers. A set of detailed experiments in both real-time and well-known datasets are carried out. The experimental results indicate that the proposed algorithm is better than any other wavelet domain methods including Haar wavelet dictionary, Symmlet-2, Symmlet-3 and Symmlet-5 with 150 atoms on the FERET Face Database. However, the current implementation is limited to the detection of frontal pose and upright orientation human faces. A possible and interesting extension would be expanding the proposed algorithm to include side-view faces as well.
An eye localization algorithm, which is based on the idea that eyes can be detected and localized from edges of a typical human face, is proposed assuming that a human face region in a given still image or video frame is already detected. This algorithm also works with edge projections of given face images. The performance of the developed system has been examined on two face databases, i.e., CVL and BioID, by comparing the estimated eye positions with the ground-truth values using a relative error measure. The localization results show that the algorithm is not affected by both illumination and scale changes since the BioID database contains images with a large variety of illumination conditions and face size. The algorithm can be applied to human-computer interaction applications, and be used as the initialization stage of eye trackers.

3.4. Holographic Camera Techniques

Digital holography holds the potential to overcome limitations of traditional film based holography which have prevented its wide spread use in many engineering applications, including 3DTV. Having said that, digital holography needs to address some technological issues before it can gain mainstream acceptance. Some of these issues include image resolution, data storage and retrieval speed, display of real and virtual images, and true color recording and reconstruction.

A digitally recorded hologram can be replayed by numerical reconstruction. An alternate method is to dynamically reconstruct the hologram using spatial light modulators (SLM). Both these techniques are investigated in this report. In subsection 3.4.1, experimental demonstration of real-time holographic reconstruction employing an SLM is presented. An object recorded underwater was reconstructed in real-time in a remote location. Subsection 3.4.2 deals with speed issues in the numerical reconstruction of digital holograms. A reconfigurable computer containing an FPGA for capture and reconstruction is being considered.

Underwater 3D recording would add value to ocean exploration. Design of any equipment and housings for this purpose would have to take into consideration of high pressure and subzero temperature conditions. Subsection 3.4.3 presents a compact holographic camera which was used in deep sea exploration. A description of the reconstruction algorithms used is also given. Laboratory reconstructed images of plankton species are included.

3.4.1. Developments for underwater remote vision

Authors: P. Benzie, H. Sun, N. Burns, J. Watson

Institutions: University of Aberdeen

Publication: 3DTV conference 2007, May 7 – 9, Kos Island, Greece

In recent years underwater digital holography has enabled high resolution optical data to be obtained with depths-of-field otherwise unobtainable by conventional non-holographic optical techniques. This method of remote three-dimensional visualization is of particular interest in identifying in-situ behavior of marine organisms without disturbing the natural habitat, position and relative location. Existing underwater holographic cameras have tended to rely on in-line recording geometries, and are thus reliant on recorded species being largely transparent, small (less than the sensor size) or distinguishable by edge information alone. With in-line geometry using a single sensor, objects can be located in three-dimensions with high resolution, but the reconstruction appears as a (high resolution) silhouette, thus longitudinal depth is difficult to retrieve without *a priori* knowledge of an objects refractive index and simple object geometries that act as a lens to the illuminating source. To combat this, multi-camera techniques or folding of the optical scene via mirrors may be applied, thus recording the in-line scene from different illumination directions. Yet generally this results in a few obtainable cross-sections of the recorded scene. Using this method, Katz *et al* improved the in-line geometry by simultaneously recording two in-line holographic images of objects from different angles thus enabling a volumetric 3D reconstruction of both lateral and longitudinal depth of objects. This is indeed effective for many marine species and as with all single camera in-line geometries without folding mirrors. Its primary disadvantage is that the surface information is lost by reliance on back illumination of the scene.

Numerical reconstruction methods for in-line holography are generally computationally slow to implement without the use of specialized hardware. Most algorithms rely on two-dimensional fast Fourier transforms (FFTs) of the total pixel count of recorded imaging area. For instance, if using the angular spectrum method it is necessary to implement two FFTs to reconstruct the object field. These algorithms are often implemented on commodity PCs using C-programming, however faster speeds may be obtained by field programmable gate arrays (FPGAs) implementation, or application specific integrated circuits (ASIC).

We propose here a method of remote reconstruction and visualization of real-time off-axis holograms using a SLM. This method has the advantage of recording surface information and optimizing the recording quality of holograms in situ without the necessity of numerical post-processing. Preliminary experimental research has demonstrated optical and numerical reconstructions at 25 Hz. Observed interlacing artefacts caused by the CCD camera used are presented and discussed. A brief discussion is given of the potential for using color-phase shifting holographic system for capturing underwater holograms. Optical and numerical reconstructions are presented using a spatial light modulator.

3.4.2. Real time capture and reconstruction of digital holograms using reconfigurable hardware

Authors: Richard Veitch, David Henry

Institutions: University of Aberdeen

Publication: Internal report

One of the many advantages of digital holography over more traditional methods is the possibility of recording holograms at frame rates of five frames per second or higher. However, recording at this speed is of limited use unless a reconstruction system can be developed which can keep up without forming a bottleneck. A proposed system to solve this problem is to implement capture and reconstruction on a reconfigurable computer containing an FPGA. Such a system would be able to read holograms from a video input into a frame buffer on the onboard RAM before performing a Pipelined Fresnel Transform on the FPGA to create a two dimensional reconstruction which could be displayed externally via a second frame buffer. A prototype reconstruction system has been developed which is capable of streaming reconstructions at 75 x 10^6 pixels per second from a static source hologram. Research suggests that this reconstruction rate will be limited by the availability of data when coupled with a video input system but should still be capable of transforms rates greater than 20 frames per second at a resolution of 1024 x 1024 pixels.

3.4.3. Numerical Methods of Reconstruction employed in underwater Digital Holography (eHolocam)

Authors: H. Sun, P. Benzie, N. Burns, D. C. Hendry, M. A. Player and J. Watson

Institutions: University of Aberdeen

Publication: Internal report

The ability to record, non-intrusively and non-destructively, high-resolution holograms of micro-objects in three dimensions (3D) in their natural environment gave marine scientists the opportunity to study the subsea environment in a way never before possible. Holography is seen to be particularly useful for analysis and precision measurement of the species and spatial distribution of marine plankton and particles within the subsea water column.

This publication reports the development of a compact digital holographic camera (eHolocam) and the techniques involved in generating and reconstructing electronic holograms (eholograms) of marine organisms and particles. Unlike recording a conventional hologram on photographic film, an ehologram is recorded on an electronic sensor and reconstructed numerically in a computer by simulating propagation of the optical field in a space. In comparison with other imaging techniques, an ehologram has several advantages such as three-dimensional spatial reconstruction, non-intrusive and non-destructive interrogation of the recording sampling volume, and the ability to record holographic videos. In a recent collaboration with marine biologists, the eHoloCam was deployed in the North Sea to record *in situ* holographic videos of plankton and other organisms present in the water column.

In digital holography, a hologram is recorded electronically on an imaging sensor such as a CCD (charge coupled device) or CMOS (complementary metal oxide semiconductor) array. A real image is then recreated by computer using a numerical reconstruction algorithm, such as the angular spectrum method or a variety of other algorithms. Holocameras based on this approach have usually deployed the in-line mode of recording because of the lower resolution requirements of electronic sensors compared with photographic films. A distinct advantage of electronic recording, apart from speed, convenience and freedom from wet chemical processing, is the ability to record holographic videos which not only allow 3D interrogation but introduce the fourth dimension, *viz*. time.

3.4.4. Conclusions and future plans

Summary of results

Optical and numerical reconstructions were carried out to demonstrate the principal of real-time optical reconstruction for underwater holography. Real-time reconstructions were achieved at 25 fps using an SLM to remotely view the virtual image of video streamed holograms. These holograms were recorded with a monochromatic laser, however, current efforts include the development of a RGB laser for capturing real-time color holograms.

Progress has been made in increasing the speed of numerical reconstruction process. A prototype of the pipelined Fresnel transform has been tested on static hologram of 512 x 512 pixels. Total transform time of approximately 7 milliseconds has been demonstrated.

A specialized underwater holographic camera has been built and successfully used to record images of marine organisms for studies of species populations and dynamics. A range of algorithms for numerical reconstruction of the holograms was developed and compared. In addition, the algorithms were subjected to pre- and post-processing methods for improving quality and noise levels of the holograms.

3.5. Pattern Projection

Pattern projection is the most appropriate technique for shape and coordinates measurement in real time of real objects in dynamic scenes. The information about spatial and color coordinates of the objects, as well as for their position in the scene is essential for their 3D presentation in pseudoscopic, autostereoscopic and holographic displays. The main goal of the different teams involved in this subtopic from the 3DTV consortium is to improve the dynamic range and accuracy of the measurement in dynamic mode on the basis of existing and newly developed pattern projection systems. The developed in ITI-CERTH (3.5.1) real-time acquisition of depth and color images by color-encoded structured light pattern allows combination of 2D and 3D image processing algorithms and provides a 2D color image in addition to the coordinate's data. Double symmetrical illumination/observation is proposed in 3.5.2 for elimination of the shadowing effects and loss of information. Different techniques for pattern generation are presented in 3.5.3 and 3.5.4. PhD thesis (3.5.5) is completely devoted to the problem of shape and coordinate measurement, including comparative analysis of different methods of optical profilometry. Newly developed system for real-time phase stepping pattern projection profilometry is presented in 3.5.6. The use of four different wavelengths for pattern projection in near IR offers possibility for simultaneous measurement of color and spatial coordinates of the objects in dynamic scenes. The obtained results and findings in the frame of this topic are essential and could be used in 3D time-varying scene representation technology (WP8), 3DTV coding techniques (WP9), data translation (WP10), signal processing (WP11) and 3DTV display techniques (WP12).

3.5.1. Real-time acquisition of depth and color images using structured light

Authors: F. Tsalakanidou and S. Malassiotis

Institution: ITI-CERTH

Publications: working draft report

In this work, a novel real-time 3D and color sensor for the mid-distance range (0.1-3m) based on color-encoded structured light pattern projection is presented. The basic principle lying behind this technique is the projection of an encoded light pattern on the scene and measurement of its deformation on the objects surfaces. The sensor is integrated using low-cost of-the-shelf components and allows the combination of 2D and 3D image processing algorithms, since it provides a 2D color image of the scene in addition to the range data. Its design is focused on enabling the system to operate reliably in real-world scenarios, i.e. in uncontrolled environments and with arbitrary scenes. To that end, novel approaches for encoding and recognizing the projected light are used, which make the system practically independent of intrinsic object colors and minimize the influence of the ambient light conditions.

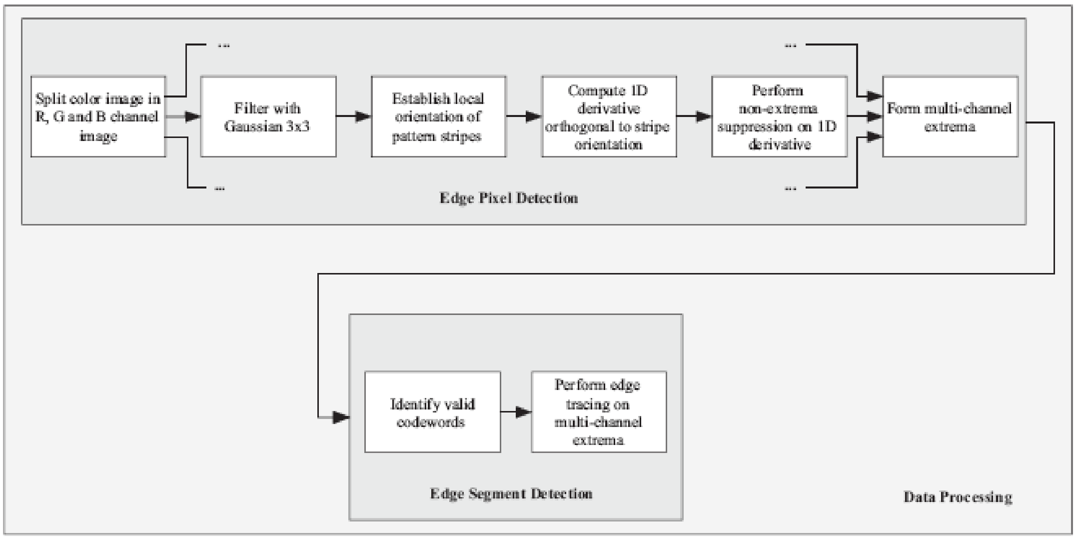


Fig. 1. Block diagram of the 3D acquisition system.

Fig. 2. Color coded pattern. (a) Clipping from the projection pattern and (b) image captured by the color camera with color pattern projected on the face.

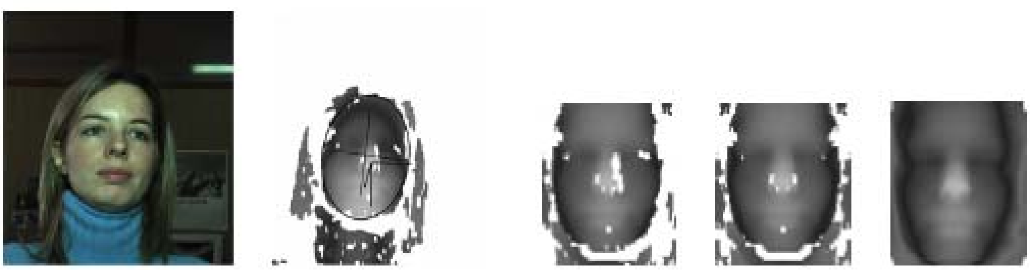


Fig. 3. Pose compensation example. (a) Original color image, (b) original depth image showing detected head blob and estimated local coordinate system fixed on the nose, (c) rectified depth image, (d) symmetry-based interpolation, (e) final linear interpolated depth image.

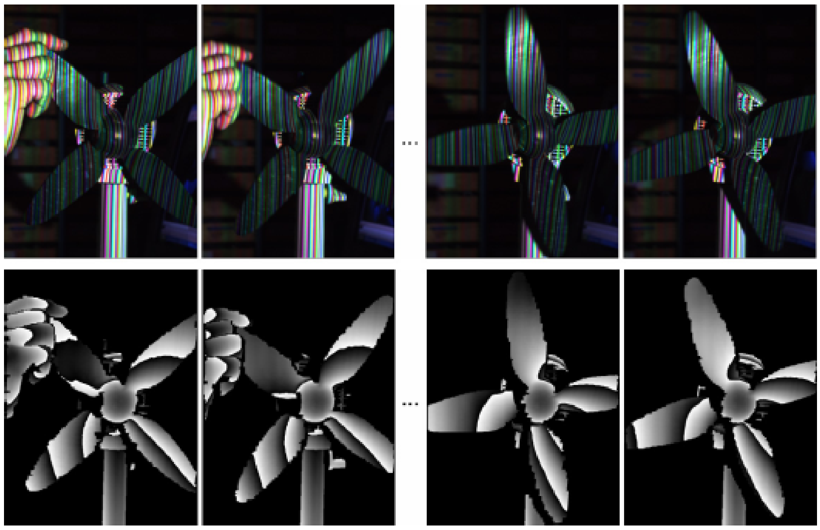


Fig. 4. Test sequence “rotating dark-green fan”

3.5.2. Three-dimensional profilometry by symmetrical fringes projection technique

Authors: Jana Harizanova, Ventseslav Sainov

Institution: CLOSPI-BAS

Publication: Optics and Lasers in Engineering 44, 1270-1282 (2006)

A precise interferometric method for three-dimensional surface measurements in a wide dynamic range is developed. The method is based on symmetrical two-spacing projection phase-stepping interferometry and allows absolute coordinates estimation of real objects in real time operation mode. Two approaches are proposed – one with double symmetrical illumination (DSI) and the other with double symmetrical observation (DSO), to reduce influence of shadowing at investigation of complex surfaces and scenes. The theoretical background description, experimental check as well as accuracy and sensitivity assessment of the method have been performed. The proposed technique is especially useful for remote, non-destructive in-situ measurements of real objects – for example historical objects and monuments.

The experimental results from surface measurements in laboratory conditions are obtained. Vertical interference patterns are generated by a compact portable interferometer, thus increasing speed and accuracy of the measurement. Influence of the external factors as vibration, air turbulence etc. could be neglected. The two spacing pattern projection interferometry allows absolute coordinate estimation of the investigated object without a reference plane that makes the method suitable for application directly at the object location. The selection of large observation/illumination angles and spacing difference increases sensitivity and accuracy of the measurements. The suitable compromise should be made for achievement of better precision, as the partial shadow zones recovery could appear, leading to losses of information. DSI method requires more complex technical implementation, while

DSO technique (Fig.1 and Fig.2) is more complicated from processing point of view. The choice of the suitable measurement approach depends on application field and type of the specimen. The developed approaches allow application in real time operation mode for measurement of moving objects and 3D presentation of real scenes.

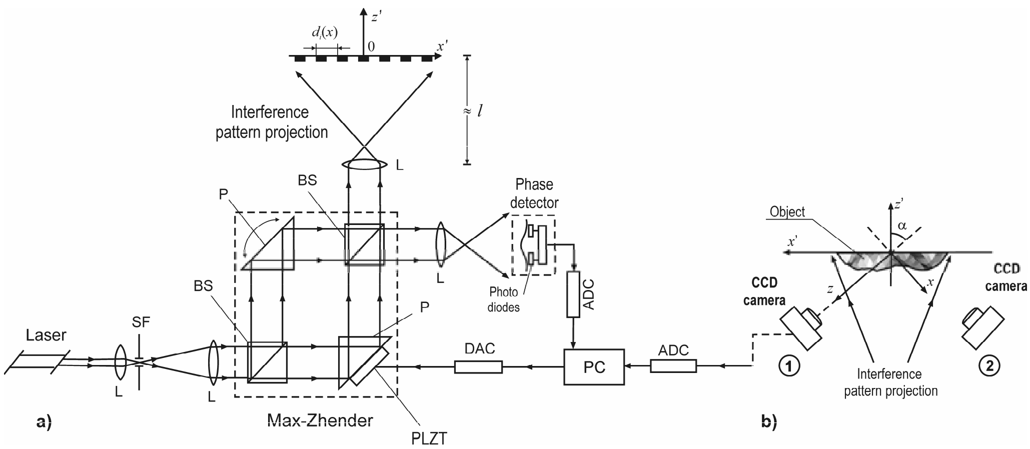
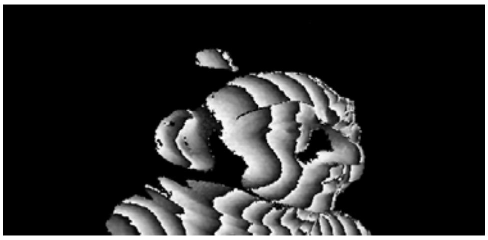
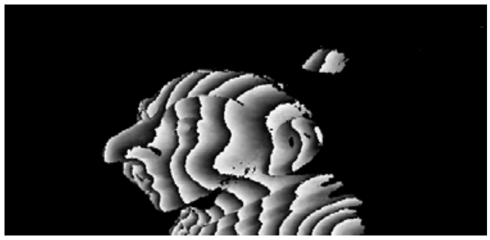


Fig.1. Optical arrangement of a double symmetrical observation pattern- projection system.

Left Right

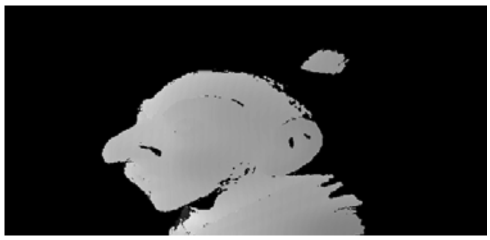



Fig.2. Shape measurement by fringe projection interferometry – double symmetrical observation; top – the wrapped phase maps, middle – the unwrapped phase maps, bottom – the reconstructed 3D image using both phase maps.

3.5.3. Comparative study of fringes generation in two-Spacing phase-shifting profilometry

Authors: Jana Harizanova, Angel Kolev

Institutions: CLOSPI-BAS

Publication: Proc. SPIE 6252, 625223-1-5 (2006); Eds. Y. Denisuyk, V. Sainov, E. Stoykova

A two-spacing phase-shifting technique for three-dimensional shape measurement based on projection of digitally synthesized fringes has been developed. The developed technique is compared to traditional laser phase-shifting pattern projection interferometry which is more vulnerable to influence of noise/vibration, as well as to possible errors in phase steps setting. The theoretical background and experimental results from surface measurements of test and real objects in laboratory conditions have been made. The results confirm applicability of the developed technique for surface profile measurement and show good agreement between both investigated methods. The proposed technique is especially useful for remote, non-destructive in-situ inspection of real objects. However, some limitations in focal depth and resolution of the developed technique restrict its applicability to specific object shapes.

3.5.4. Phase retrieval techniques in coordinate measurement

Authors: Jana Harizanova, Elena Stoykova, Ventseslav Sainov

Institutions: CLOSPI-BAS

Publication: Oral presentation at BPU'2006, Istanbul, Turkey; to be published in the Proc. of AIP

Recently different approaches for fringe generation have been proposed, based on DMD (Digital Micromirror Device) or LCD (Liquid Crystal Display), offering easy implementation and minimization of phase-shifting error. However some limitations occur at investigation of specimens with big relief depth due to ill-focusing in some areas along with non-sinusoidal profile of the fringes decreasing contrast of the projected pattern and accuracy. The application of diffraction gratings for fringe generation combines the benefits of classical interferometrical techniques and projector based methods such as compact setup, unlimited focal depth and coherent light source. Nevertheless the week points in such experimental solution are presence of higher diffractive orders and limited illumination efficiency.

In this work a precise pattern projection profilometry for three-dimensional shape measurements with different methods of fringe generation is presented. The application of phase-shifting algorithm along with two-spacing illumination allows for phase retrieval and relative and absolute coordinate estimation of the tested samples. The following experimental approaches for fringe generation are investigated: interferometric approach based on a classical Michelson interferometer, digital computation with a DMD projection and light modulation by a sinusoidal phase grating. The theoretical background, experimental results as well as comparison of the applied generation methods are analyzed. The obtained outcomes successfully display the applicability of this technique for surface profile measurement. The application of the proposed techniques for remote, non-destructive in-situ inspection of real objects from cultural heritage is discussed.

3.5.5. Holographic and digital methods for recording and data processing in investigation of cultural heritage

Author: Jana Harizanova

Supervisor: Ventseslav Sainov

Institutions: CLOSPI-BAS

Publication: Ph.D. Thesis; summary of the thesis (30 pages) is published as a book in Bulgarian with acknowledgment to 3DTV project

The accurate profile acquisition is of great importance for monitoring and non-destructive testing of cultural heritage. The need for its protection has significantly increased during the past century. Urban development, pollution, climate changes and natural disasters are some of the most important deterioration factors for monuments and artifacts, especially for those in outdoor conditions. Control of surface modification is one of the basic requirements for restoration and monitoring of monuments and artifacts. Diversity of locations, materials and reliefs of the objects however demands various techniques for high accuracy inspection, especially for large-scale specimens, that could not be measured in laboratory conditions. Surface profile measurements which utilize optical methods are widely applied due to their inherent advantages, such as remote interaction, high precision and automation.

In the PhD thesis following main tasks are completed:

- Development of methodology for precise absolute coordinate's measurements, appropriate for investigation of cultural heritage, based on original two-spacing projection phase-stepping interferometric approach. 3D measurements of real object surfaces are obtained without reference plane that makes method suitable for in-situ application.
- Development of methodology for minimizing influence of shadowing and information losses due to objects complicated relief, implementing two approaches: double symmetrical illumination double symmetrical observation of the surface under investigation that significantly increases accuracy of the measurement.
- Development of methodology for determining in plane and out of the plane displacements and deformations with increased sensitivity.
- Development of original software modules for relative/absolute coordinates determination and automated surface reconstruction as well as the multimedia database for 3D presentation of the investigated objects.

The PhD thesis comprises introduction, four chapters, appendix, conclusions and references. In introduction the state of the art of the subject is discusses and the main tasks of the thesis are formulated. Chapter 1 is dedicated to comparative analysis of main classical and contemporary methods for surface contouring and coordinates measurements. In Chapter 2 methodology of the measurements is described. The obtained results of absolute coordinates measurements based on two spacing projection interferometry and approaches for information losses reduction are presented in Chapter 3. Chapter 4 contains the outcomes from normal displacement measurements and displacements based on phase-stepping shear, projection and electronic-speckle interferometry. In the Annex the detailed description of specialized software modules and a structure of multimedia database for 3D visualization are provided.

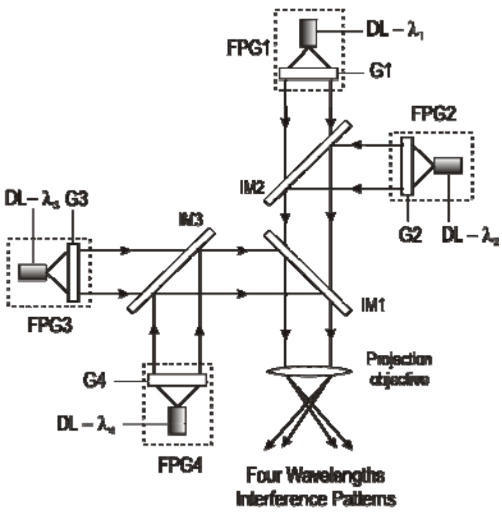
3.5.6. Real-time phase-stepping pattern projection profilometry

Author: Ventseslav Sainov, Elena Stoykova, Jana Harizanova

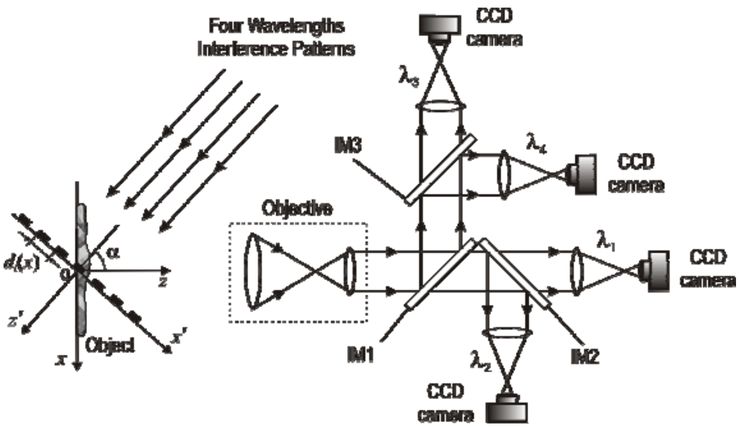
Institutions: CLOSPI-BAS

Publication: Proc. SPIE 6341, 63410P-1-6 (2006); oral presentation at Speckle'06, Nimes, France

This work proposes a single-shot FPP system based on simultaneous projection of four phase-shifted sinusoidal fringe patterns generated at four different wavelengths. In order to simplify the technical solution and to avoid the stringent requirement for stability in the case of interferometric fringe generation, we analyze the system realization by using of phase or holographic diffraction gratings. The results of the measurement of the relative and absolute coordinates of test objects for both types of gratings are presented.



a)



b)

Fig. 1 Optical arrangement of the four-wavelengths FPP system

a) fringe pattern generation module b) registration module

Optical arrangement of the four-wavelength FPP system is presented in Fig. 1. The fringe pattern generation module consists of 4 blocks (FPG1-4) corresponding to four different wavelengths ($\lambda_1 - \lambda_4$) as is shown in Fig 1 a), where the FPGs are Fringe Pattern Generators, DLs are 20 mW CW single mode diode lasers, G1 – G4 are phase or holographic gratings. The diode lasers emit in NIR spectral region at wavelengths: λ_1 = 785 nm, λ_2 = 808 nm, λ_2 = 830 nm and λ_4 = 850 nm. The fringe pattern generation module should ensure four sinusoidal patterns of equal intensity, contrast and spacing that are phase-shifted at π/2. To optimize the optical efficiency of wavelength mixing, the different interference mirrors (IM1 – 3) are used as follows: the mirror IM1 transmitting λ_1, λ_2 and reflecting λ_3, λ_4; the mirror IM2 transmitting λ_1 and reflecting λ_2 and the mirror IM3 transmitting λ_3 and reflecting λ_4. The object is illuminated by the four patterns simultaneously using a projection objective.

The registration module (Fig. 1 b) consists of four synchronized CCD cameras for capture of the deformed by the measured surface fringe patterns. The spectral separation of the individual patterns is provided by a second set of interference mirrors IM1 – 3. The precise positioning and adjustments of cameras and optical elements is required to avoid possible systematical errors and to ensure parallel recording of the phase-shifted patterns, modulated by the object surface.

Fig. 2 Experimental results for normal displacement measurement of a tilted plane with a phase grating at λ = 785 nm a) wrapped phase map; b) unwrapped phase map; c) 3D presentation

Fig. 3 Experimental results for relative coordinates measurement of a test object with a phase grating at λ = 785 nm a) wrapped phase map; b) unwrapped phase map; c) 3D presentation

3.5.7. Conclusions and future plans

The research efforts within the "Pattern projection" section for the reported period were oriented towards solving the problem for accurate retrieval of 3D absolute coordinates of objects with a complicated relief. Two systems are described based on color-encoded in visible structured light pattern projection and on simultaneous illumination of the investigated surface by four phase shifted at π/2 patterns generated with different wavelengths in the near IR spectral range and capture of the deformed patterns by four CCD cameras. The research included development of laboratory prototypes of the proposed system, creation of an appropriate software for 3D reconstruction of the objects, measurement of test and real object and initial analysis of accuracy, sensitivity and dynamic range of the measurement. Several key tasks connected with loss of information due to shadowing, low-level of illumination, discontinuities and singularities in the phase maps were thoroughly studied and some successful hardware and software solutions were proposed to avoid its deterioration effect. Different methods for generation of fringes were experimentally studied as interferometric approach based on a classical Michelson interferometer, digital computation with a DMD projection and light modulation by a sinusoidal phase grating. Both described measurement methods can be realized with low-cost technically available components but are still limited in their ability to measure large-scale scenes. Nevertheless, the obtained results are promising for solving the problem of color and 3D coordinates measurement in real time, required for creation of dynamic CGHs for the dynamic holographic display as well as for the pseudoscopic visualization. The proposed techniques could be used also for normal displacement measurement and for investigation of transient effects.

The future activities within this section should start with realization of real time operation mode. Electrically driven LC spatial light modulators will be used for generation of the fringes with sinusoidal profile and variable in broad limits spacing. In addition, absolute coordinate measurement will be performed by two-spacing phase stepping profilometry realized with a double symmetrical illumination. The further assessment of sensitivity, dynamic range and uncertainty of the measurement should be made on the basis of simulations. The most appropriate for the investigated objects noise filtration technique will be selected through the comparative analysis of different filtration algorithms, concerning productivity and information losses. As the sensitivity of the method varies in broad limits in comparison with the other interferometric techniques, the inspection of the objects will be realized in a wide dynamic range and working conditions. An appropriate software for data processing – denoising, filtration, unwrapping and fast data translation will be developed in connection to the tasks of the other work packages – WP8,9,10,11 and 12.

3.6. Motion Analysis and Tracking

Motion analysis generally detects objects in a scene, tracks them, and determines what types of movements are being performed. We are considering object motion which can be split into three stages: object detection, tracking and classification of its motion. The special human body part, the head, is handled separately due to the exceptional focus.

3.6.1. Using Multi-target Tracking from Multiple Views for Football-related Applications

Authors: Alexandra Koutsia, Nikos Grammalidis, Kosmas Dimitropoulos, Mustafa Karaman, Lutz Goldmann

Institutions: ITI-CERTH (Informatics and Telematics Institute – Centre of Research and Technology Hellas), TUB (Technical University of Berlin)

Publication: A paper is submitted to VISAPP

Soccer is probably the most popular sport activity in the world. Therefore, most of the research efforts have been devoted to analyze soccer games. Game analysis typically includes detection of many events such as corners, free-kicks, throw-ins, etc., and extraction of statistics. A detailed analysis can help coaches to design a specific training for each player or to evaluate different team formations and tactics.

In this work, information from multiple cameras placed over a football pitch is used to perform player recognition and tracking during a match. Detection of football players on each camera is done by the background subtraction method which uses Gaussian color model and temporal information. Multiple hypothesis tracking (MHT) is utilized to improve tracking results applied on the obtained 3d-positions of each player. With occlusion handling based on position information from multiple cameras, less tracking failures are obtained. The resulting data displayed on the ground plane can then be used in a series of applications, such as the creation of an automated system which provides football game analysis or football viewing using low bit rate channels (e.g. mobile phones, internet). Furthermore, game statistics can be obtained.

3.6.2. Traffic Monitoring Using Multiple Cameras, Homographies and Multi-Hypothesis Tracking

Authors: A. Koutsia, T. Semertzidis, K.Dimitropoulos, N. Grammalidis and A. Kantidakis

Institutions: ITI-CERTH (Informatics and Telematics Institute – Centre of Research and Technology Hellas), MARAC Electronics

Publication: A paper submitted to 3DTV-CON, May 2007, Kos, Greece

Traffic control and monitoring using video sensors has drawn increasing attention recently due to the significant advances in the field of computer vision. However, robust and accurate detection and tracking of moving objects still remains a difficult problem for the majority of computer vision applications. Especially in case of outdoor video surveillance systems, the visual tracking problem is particularly challenging due to illumination or background changes, occlusions problems etc. In this work the aim is to present a novel multi-camera video surveillance system, which supports functionalities such as detection, tracking and classification of objects moving within the supervised area.

The proposed system is based on a network of intelligent autonomous tracking units, which capture and process images from a network of pre-calibrated visual sensors. The fundamental goal is the development of a moving target tracking system which is fully scalable and parameterized and can be applied in a broad field of applications in the future. Two prototypes will be developed, each one focusing on a different aspect of traffic monitoring: traffic control of tunnels at highways and of the aircraft parking area (APRON) at airports. Each autonomous tracking unit of the proposed system can automatically deal with background changes (e.g. grass and trees moving in the wind) or lighting changes (e.g. day, night etc) using e.g. the mixture of Gaussians modeling of the luminance for each background pixel, although other background extraction algorithms can be easily integrated. The extraction of observations is performed by a subsequent step of moving blobs analysis, while the accurate position of targets in the scene is calculated by the available calibration information of each camera. An important step of the processing chain is the fusion of data extracted by the autonomous tracking units. Two data fusion techniques are considered in order to address occlusion problems and eliminate blobs merging / splitting errors due to the camera perspective: i) a grid-based fusion technique and ii) a foreground map fusion technique. Both techniques are tested using the same traffic monitoring sequence providing promising results.

3.6.3. Model-based Head Motion Estimation with an Automatic Update of Texture Information

Authors: Axel Weissenfeld, Kang Liu, Joern Ostermann

Institutions: Institut für Informationsverarbeitung, Leibniz Universität Hannover

Publication: internal report

A head tracking system estimates the rigid motion of the human face throughout an image sequence. Head tracking systems are important for many applications in computer vision like expression analysis, face identification, model-based coding and 3D facial animation systems. Head motion can be used to recognize simple gestures, like head shaking or nodding, or for capturing a person's focus of attention, providing a natural clue for human machine interfaces.

Existing approaches can be divided into motion-based and model-based systems. In the first approach, distinct facial features, such as eye corners or nostrils, are tracked throughout the image sequence. The displacements between corresponding feature points can be estimated using optical flow or block-based motion estimation methods. In this way, a 2D motion field is estimated in order to calculate the motion of the object model. The object model is only used to transform 2D motion vectors into object model motion vectors. One problem with methods based on this approach is the accumulation of motion estimation errors and thus a drift.

A model-based tracker stores texture information of the object and tries to adapt the object model's position to fit the new frame. Therefore, the motion estimation is dependent on the texture information of the initial and current frame and on the object model. Model-based motion estimation can be accomplished by optical flow or image registration in texture space. A recent algorithm combines optical flow information with image gradients. Many variations of motion estimation algorithms have been proposed in literature. Differences can be noticed in the boundary conditions, like the use of calibrated or uncalibrated image sequences or the used motion model.

In this report a gradient-based motion estimation algorithm is extended by a fully automatic update of texture information. The gradient-based algorithm estimates the motion of the face model by calculating the optical flow of the luminance values of feature points. A luminance value of a feature point can be considered as texture information. A careful selection and update of texture information of feature points is important for an accurate motion estimation. In literature, the update of texture points has been widely neglected. Mostly either feature points are updated after each frame or feature points are only once initially selected. Both approaches have disadvantages, a drift may occur in the first while the second approach is unable to estimate large out-of-plane rotations. Here a new method is presented, which automatically updates texture information if necessary, while allowing for an accurate motion estimation.

3.6.4. Conclusions and future plans

Tracking of football players on the pitch using multiple cameras was developed. Due to the limited object resolution small objects, such as the ball, are more difficult to track. Therefore, 3D ball tracking is left as a future goal to achieve more information about the game itself.

A novel multi-camera video surveillance system for traffic monitoring applications was described, focusing on the comparative evaluation of two data fusion techniques: grid-based and foreground probability map fusion. The grid-based technique is simpler, faster and requires less bandwidth, while the foreground map fusion technique was seen to robustly resolve occlusions. Future research will be focused on improved foreground detection techniques, robust to shadows and illumination changes. Furthermore, an extensive evaluation of both methods in terms of performance, speed and bandwidth requirements will be conducted.

We extended a model-based motion estimation algorithm for full head motion recovery by an automatic texture update. The proposed strategy significantly reduces the error as determined by a real sequence with respect to never or always updating texture information. Especially if large out-of-plane rotations occur, our algorithm is able to accurately track the pose of the human head.

After achieving of complete information from both systems of ITI-CERTH, methods for reconstructing and projecting in 3D should be considered to display the activities of objects in monitored fields as animated applications. Accurate information about head motion can yield better results for 3D animation which should be considered in the future.

3.7. Registration

Registration is the task of finding correspondences between images. Within the 3DTV NoE, registration is closely related to and used by many subpackages of WP7 such as multi-camera, single camera, pattern projection. It is expected that these results can be used by other partners of the NoE that use registration for their particular problem and will lead to further collaboration with the NoE.

3.7.1. An Image Registration Technique Aimed at Super Resolution on Mobile Devices

Authors: Mihail Georgiev, Ilian Todorov, Atanas Boev, Atanas Gotchev, Karen Egiazarian

Institutions: Tampere University of Technology

Publication: Submitted to EURAPSIP Journal of Embedded Systems

Super-resolution works by combining several slightly spatially-different low-resolution observations into a single high-resolution image [1]. It is essential that the spatial displacements of the low-resolution images are precisely known. The image registration task aims at finding the optimal transformation that would match two images. Together with the presence of rigid-body transformations, there are optical distortions introduced as well. Therefore, a good image registration algorithm for a mobile camera should handle noise, large displacements, local distortions, and to cope with lens distortions.

We have adopted the feature-based approach and develop a modified technique combining registration with lens-correction thus avoiding any initial resampling. Our technique is fast and robust and therefore well-suitable for further use for super-resolution on mobile devices. To demonstrate its usefulness for super-resolution applications, we have also implemented radial interpolation that combines the registered data on non-uniform grid and generates the desired high-resolution grid. Additionally, we implemented a super-resolution method, based on projections onto convex sets (POCS), [2]. To demonstrate feasibility of our technique we have built an application working on Texas Instruments Innovator Development Kit [3]. This platform is powered by OMAP1510 dual-core processor architecture incorporating fixed-point TMS320XC55 DSP CPU and ARM925 ARM CPU.

The proposed algorithm for registration is shown in Figure 1. It follows the three general steps of any feature-based image registration (FBIR), namely *feature selection*, *feature matching* and *transformation model estimation*. The algorithm handles rigid, affine and planar projective transformation.

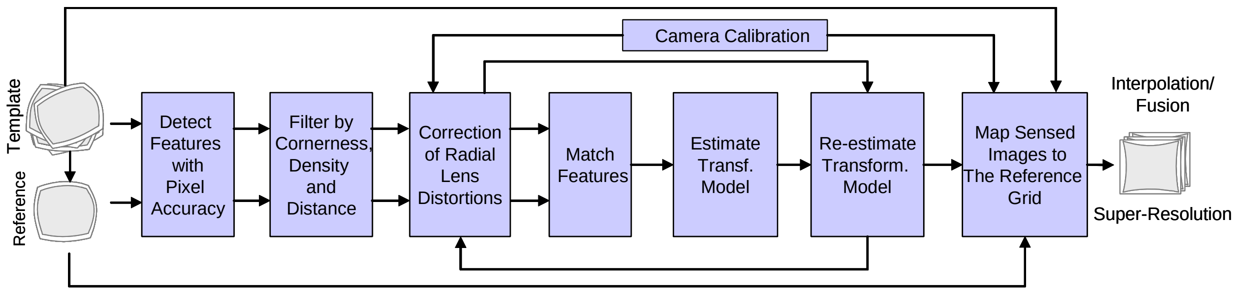


Figure 1: Block Diagram of FBIR algorithm

The algorithm was tested on real images with the following standard settings: number of filtered features – 50, threshold to accept the consensus set – 4, minimum of 15 accumulated correspondences to proceed re-estimation of transformation, total pixel displacement by rotation and translation – in the range of [0, 20] pixels. A example picture, taken with handheld camera is given in Figure 2. The super-resolved image is visually better and less noisy than the one enlarged by using bicubic interpolation. As the images have been registered using the plane containing maximum reliable features, objects out of this plane are slightly deformed and blurred. Photographed scene in Figure 3 was scanned at low resolution and various orientations. After registration, the images were fused using the POCS method.

(a) (b) (c)

Figure 2: Registration results with handheld camera images (a) test scene, (b) detail, doubled with bicubic interpolation (c) detail, after registration of five images and fusion with radially symmetric interpolation.

(a) (b) (c) (d)

Figure 3: Super-Resolution enhancement with POCS method of handheld camera images.
a) reference image with resolution doubled by bicubic interpolation, b) Fused image of 3 registered images, c) Zoomed region of reference image d) zoomed region of the enhanced image

3.7.2. Corner Validation based on Extracted Corner Properties

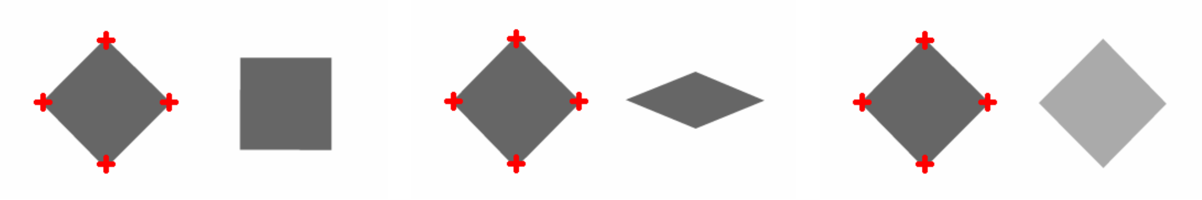
Authors: Yalin Bastanlar, Yasemin Yardimci

Institutions: Middle East Technical University

Publication: Submitted to Computer Vision and Image Understanding

We developed a method to validate and filter a large set of previously obtained corner points. We derived the necessary relationships between image derivatives and estimates of corner angle, orientation and contrast. Commonly used cornerness measures of the autocorrelation matrix estimates of image derivatives (such as Tomasi-Kanade [4] or Harris-Stephens [5]) are expressed in terms of these estimated corner properties. A candidate corner is validated if the cornerness score directly obtained from the image is sufficiently close to the cornerness score for an ideal corner with the estimated orientation, angle and contrast. We tested this algorithm on both real and synthetic images and observed that this procedure significantly improves the correct corner detection rates. Moreover, extracted corner properties can be used for tasks like feature point matching, object recognition and pose estimation. We tested the accuracy of our corner property estimates under various noise conditions.

It turns out that for image gradient based detectors two identical corners may not be selected for a given threshold if their orientations are different (Figure 1a). A similar problem is observed for objects with identical orientations but different corner angles (Figure 1b) and different contrasts (Figure 1c). Such changes are likely to occur under affine transformations and/or varying illumination conditions in different feature matching scenarios. It is important to design corner detectors that take corner angles, orientations and contrasts into account.



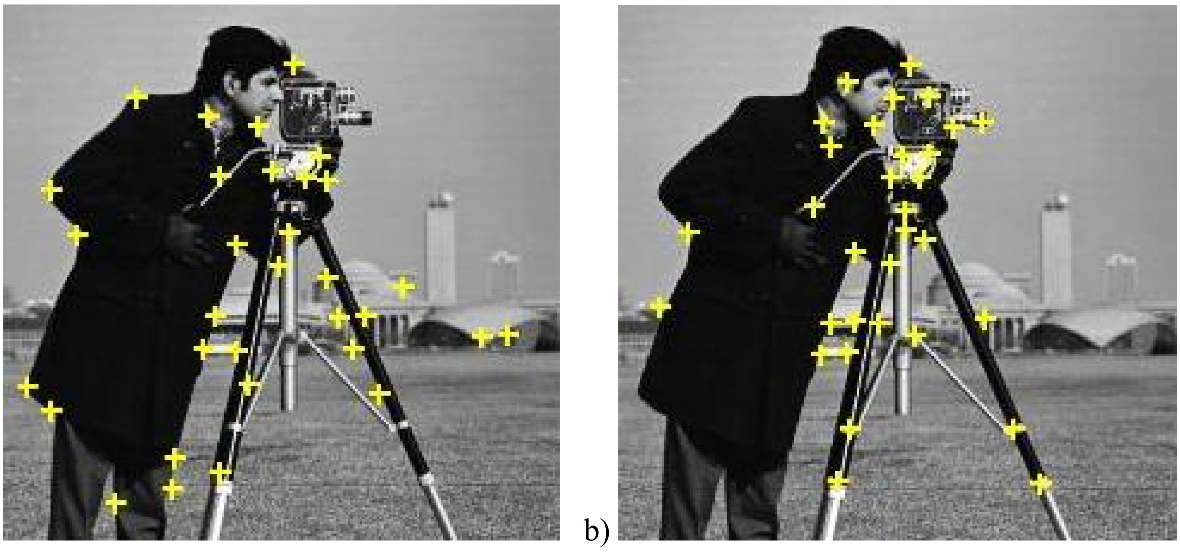
a) b) c)

Figure 1: For a given threshold, a) the corners of the square on the right square's corners are not detected due to a different orientation, b) the compressed square's corners are not detected due to different corner angles, c) the right square's corners are not detected due to a lower contrast.

When a corner point is given, we can extract the corner orientation (θ), contrast (Δ) and corner angle (α) in terms of image gradients that are either already computed for the cornerness measure or they can be computed easily from the derivative image.

We named our method as COVPEX (Corner Validation based on Corner Property Extraction). In COVPEX, we first estimate α, θ, Δ and compute the corresponding eigenvalue of a corner having these corner properties, $\hat{\lambda}_{\min}(\hat{\alpha},\hat{\theta},\hat{\Delta})$. Then, we compare this eigenvalue with the one calculated from the image ($\lambda_{\min}$). If they are close to each other, we conclude that the corner properties are successfully extracted and we validate that corner. If the Harris-Stephens method is preferred, we repeat this procedure for the cornerness measure $\hat{M}(\hat{\alpha},\hat{\theta},\hat{\Delta})$ and compare it with the one directly calculated from the image (M).

We tested our method on well-known 'cameraman' image. Initial corner set, obtained via Tomasi-Kanade (TK) using a low threshold, contains 113 points. In Figure 2a, 32 points validated by COVPEX are shown (feature window size is taken as 17x17 pixels). In Figure 2b, corner set obtained via TK is decreased to 32 points by increasing the TK threshold. It is observed that many well defined corners around the body of the cameraman that are validated by COVPEX (Fig. 2a) are missed by TK method (Fig. 2b). If smaller details on the camera are preferred a smaller window sizes can be used.



a) b)

Figure 2: Test on the 'cameraman' image, window size=17x17 pixels.
Initial corner set by TK is decreased to 32 points a) by COVPEX b) by increasing TK threshold

3.7.3. Conclusions and Future Work

In the contribution of TUT, a registration technique suitable for imaging applications on portable devices is developed. Experimental results have demonstrated that the approach is robust to conditions typical for hand-held camera use and accurate enough for super-resolution implementation. Presented implementation shows the advantage of feature-based image registration methods in terms of speed. The technique is quite general and can be used as a first step in low-complexity super-resolution imaging and also in related applications such as stereo matching, depth estimation, image mosaicking, video stabilization, etc.

A corner validation algorithm based on corner property extraction is presented in the work of METU. It validates corners regardless of their orientation, contrast and corner angle. Proposed algorithm depicted ideal behavior for noise-free images. It also provides better detection performance under mild noise conditions. Successful results are obtained on real image tests as well. Corners selected by the presented algorithm are not only high-eigenvalue regions but they also have 'clear' shapes resembling a corner. Moreover, extracted corner angle, orientation and contrast can be utilized in applications of feature point matching, object recognition, pose estimation etc. The potential use of this technique with omnidirectional images will be evaluated. Possible collaborations with the work of TUT and other NoE partners will also be investigated.

3.8. Object-Based Segmentation

Object segmentation is one of the main tasks for many kinds of image processing and computer vision applications. Especially for 3D object based registration and reconstruction algorithms a computation and segmentation of meaningful objects is essential.

The next section addresses the initialization step to color based segmentation tasks.

For color based segmentation the segmentation algorithm needs some kind of color model. In the case of the keyer developed at UIL a GMM (Gaussian mixture model) is used to describe the color distribution of foreground as well, as background. This color model has to be initialized. Doing this manually is a time consuming task. An automatic initialization step is shown to reduce the time needed to initialize the color model.

3.8.1. Motion-Based Initialization of Color Models

Authors: Michael Kirchner

Institutions: Technische Universität Ilmenau (UIL)

Publication: M. Kirchner, Motion Based Segmentation for Automatic Initialization of Color Based Keyers, internal technical report, IMT 2007

As an enhancement to the background color agnostic keying technique, developed at UIL, some methods for automatic initialization of the GMM color model were investigated. The first approach was a feature based object tracking method used to examine the moving parts of the image sequence. Since this method generates too few key points to calculate a usable segmentation, a second method was developed, that uses the optical flow and an intensity threshold to extract the moving parts. Due to the optical flows weakness on object borders, this method generates a foreground-background-bitmap that does not match the exact border of the objects. However results showed, that an exact border is not needed to decide whether a color component is part of the foreground or not.

Although not exactly motion based the third evaluated method is based on a kind of clean plate difference keying. To compare the quality of the latter two methods to the manual initialization method quality measurement methods were developed. The measurement results showed that the automatic methods principally can achieve a quality of the produced color models as high as with manual initialization. But in some more difficult sequences where one color component is located in the border area the optical flow based automatic methods failed.

3.8.2. Conclusions and Future Plans

While the principal functionality of the motion based initialization of a GMM color model has been shown, there is room for further improvements. One improvement could be the change from the expectation maximization algorithm to a parameterless approach as the mean shift algorithm. The next major step will be a motion based matte prediction to improve the quality of temporal constraints in the MRF-based keying process.

4. Conclusions

This report has presented the projects and research performed by the partners of work package 7 during project months 17 to 29. During this time, several important contributions to the area 3D scene capture have been made by 3DTV partners.

The tasks within this project are concerned with recording and processing of three dimensional scenes. The huge number of different approaches to tackle this problem and their variety, makes WP7 a most versatile and dynamic work package. As one part of the WP, recording oriented research areas are investigated, such as Multi- and Single-camera techniques, as well as Holographic and Pattern Projection techniques. In addition to that, contributions in processing and vision tasks such as Registration and Object-Based Segmentation were made, and also application-specific problems, such as special 3D TV techniques for Human Face and Body techniques, are investigated.

One of the reasons of the diversity of techniques in this report is that recording three dimensional scenes is a highly complex task, and there is yet no single method to capture a general scene. All methods have their own advantages and disadvantages. The work presented in WP7 provides researchers with a more profound understanding of the difficulties involved in time-varying scene capture. It also enables researchers to better decide which acquisition technology is suitable for which type of scene. As such, we anticipate that a lot of the work performed in WP7 will flow into the other technical work packages of the 3DTV Network of Excellence.

Number of partners	16
Reported planned or ongoing joint projects	18
Number of research visits related to WP7	12
Number of publications	36
Software uploads related to WP7	8
Data uploads by WP7 partners	7

Table 1: WP7 at a glance *- the reported number of joint projects and research visits reported is based on statistics collected in preparation for the TC1 meeting. Software and data uploads are calculated from downloads related to the WP7 tasks available at the 3DTV webpage by WP7 partners.*

The overall knowledge about time-varying scene capture gathered in our NoE is a unique pool of expertise that is shared by all 3DTV partners. All in all, WP7 is a very successful research effort. The collaborating partners have made significant contributions to top tier journals and conferences, Table 1. Overall, 36 research papers were published in the reporting period which is a notable increase over the 25 papers published in the last period. Seven of the papers presented here are joint publications by 3DTV partners, in contrast to only four in D26.1. As a whole the work package is making good progress with several unique projects. The partners have a dynamic cooperation with several research projects being reported as ongoing or discussed for the future. Also, a variety of research software prototypes were made available, and test data sets were assembled.

In future, we will make our knowledge available to the contributors to the other work packages. As part of this dissemination activity, the WP7 partners are currently working on a joint survey paper. In the following, the partners are summarizing their results presented in the different subsections and give an outlook to their future activities.

4.1. Multicamera

There are many remaining algorithmic difficulties which we intend to research in greater detail. MPG already started to develop a new dynamic reflectance sharing approach that produced better reflectance estimates despite a bias in the captured reflectance samples. By this means, the appearance of the videos under lighting conditions starkly different from the measurement setup will improve. A second major line of research will be the incorporation of more detailed geometry models in a free-viewpoint video framework. To this end, MPG currently researches ways to use laser-scanned static human geometry for marker-less motion capture, 3D video acquisition and display. The first step in this direction is already described in "A Framework for Natural Animation of Digitized Models".

In future we plan to proceed further in this direction and develop simple and fast deformation approaches that can be used to capture the motion of and animate complex geometry models without having to go through the classical skeleton-based animation pipeline. Having such more detailed models, the time-varying shape of clothing and subtle surface details will be more faithfully reproduced. MPG also plans to develop methods to reconstruct spatio-temporally consistent mesh models, i.e. models with consistent graph structure, directly from multi-view input video. Spatio-temporally consistent reconstruction is a very hard problem. However, with a proper solution, a larger range of models can be handled, the acquisition process will depend on less additional hardware, and the dynamic models can be encoded efficiently.

The animation framework presented in Sect. 3.1.5 simplifies the traditional, not so straightforward acquisition-to-animation pipeline. Therefore, we expect in the future that more researchers will start to use such techniques to generate realistic animations of humans and characters. Since a human performing can drive the animation of the virtual character efficiently, 3D Videos can also be easily generated. For future work, the system's efficiency and the quality of the resulting animations are expected to be improved, since faster and better methods for mesh deformation are developed. As future work for the fully-automatic system in Sect. 3.1.7, we expect that the integration of such system with an automatic non-intrusive surface reconstruction approach will allow us to automatically learn complete virtual characters, including its kinematics and skinning properties.

A future project involving minimal surfaces (Section 3.1.9) as a pre-requisite is the computation of a consistent surface parameterization. Currently the surfaces are defined implicitly as level sets of a function. Therefore correspondences on the surface over time are not easily established. The project is planned to solve this problem by reformulating known PDE-based optical flow methods on Cartesian grids to work on arbitrary manifolds. In the future it would be desirable to optimize shape and parameterization at the same time. Usually joint optimizations result in superior reconstruction quality if suitable initial guesses can be found.

4.2. Single Camera

The research efforts in Single Camera techniques have been focused on recovery of 3D structure from 2D broadcast data. This is demonstrated to be possible, however still there exist some issues to be resolved, mainly self-calibration, to improve the robustness of the system. The next step of the research aims to estimate a piecewise planar reconstruction from a point cloud, a natural extension to a sparse reconstruction algorithm. Apart from being a good model for man-made and most natural scenes, it also offers a very efficient scene representation, when compared to point-based and voxel-based representations. Image-consistent triangulation, in which a triangulation minimizing the errors on predicted images is sought, is identified as a proper approach to study this problem. However, a successful piecewise planar reconstruction algorithm requires the solution of the sub-problems, primarily, 3D mesh generation and optimization in triangulation domain.

4.3. Human Face and Body

In the next phase of the project *UHANN* plans to continue the work on facial feature analysis and facial motion estimation. The improvement of developed algorithms and combining them is a main research direction for the next months.

UHANN and *TUT* have investigated the AAM (Active Appearance Model) algorithm and evaluated the basic AAM approach . Several databases have been tested by AAM and the results are promising. In the next months *UHANN* and *TUT* will work on improvements of feature point detection. The facial features like mouth corners, lip contour, eye corners, and chin and jaw contour have to be located with subpixel accuracy. The detection of facial features will be carried out automatically so as to track the facial feature points in the sequences. The AAM approach will be improved and evaluated for the facial animation system. Regarding to the attribute of AAM, a general training database is also expected, with which different subjects can be detected.

The use of iso-contours to capturing the discriminative information of the facial surface structure suitable for recognition was investigated by *ITI-CERTH*. The comparisons were made with a simplified yet improved in computational complexity version of point signatures algorithm. As shown by experimental results, iso-contours outperform point signatures as the proposed algorithm is not sensitive to translation and planar rotation between the images compared. Still, the proposed algorithm is not robust enough for images depicting facial expressions and our future work within the next 12 months will try to overcome this problem.

Momentum and *Koc* have developed a method for automatically generating head motion parameters from an input speech, to be used for animating the head of an avatar. In the next 12 months this method will be further developed so that the facial expressions as well as the head motion can be estimated automatically from the given speech file. This is planned to be achieved via correlation analysis of facial expressions and speech prosody. After tracking the facial features of a talking person, an HMM based classification / correlation analysis technique will be employed for building a correlation model between the facial expressions and speech prosody. This model will then be used for automatic generation of facial expressions from speech.

Bilkent will improve the algorithm described in this report. The proposed human face detection and eye localization algorithms in addition with an object tracker will be combined and implemented for face recognition purposes in 3D environment using stereo video. The locations of eyes in each video would be matched in order to obtain also height information.

4.4. Holographic Camera Techniques

University of Aberdeen has contributed two projects. The first is a FPGA system used for real time capture and reconstruction of digital holograms. The current prototype is capable of streaming 75 x 10^7 pixels per second from a static source hologram.

The second contribution from *Uni. Aberdeen* considers using sub-pixel shifting to increase the resolution of hologram recording. This work suggests that hologram resolution can be improved using a shifting of the CCD targets and multiple exposures. These results are noteworthy, as resolution is of great importance in hologram recording.

In the next phase of this work, continued development of the numerical reconstruction algorithms is needed for faster data interrogation and connection with display aspects. We expect to pursue further work on remote underwater vision and develop digital underwater holography to the point where we can record large angle off-axis holograms. However, it should be noted that the work on underwater holography (digital or classical) is not covered by the 3DTV project and separately funded. To record color holograms of course still requires the development of a compact, efficient 3-colour laser and we hope to make significant progress on this when additional funds are available.

4.5. Pattern Projection

Two pattern projection systems are under design and testing for 3D coordinates measurement in real time within the 3DTV project. The first one relies on projection of patterns of color-encoded structured light whereas the second uses sinusoidal fringes generated by means of SLMs or phase gratings.

The research activity on Pattern Projection for the reported period includes development of laboratory prototypes of the proposed system, creation of appropriate software for 3D reconstruction of the objects, measurement of test and real object and initial analysis of accuracy, sensitivity and dynamic range of the measurement. Several key tasks connected with loss of information due to shadowing, low-level of illumination, influence of intrinsic object colors and the ambient light conditions, discontinuities and singularities in the phase maps have been thoroughly studied and some successful hardware and software solutions have been proposed to avoid its deterioration effect. Both systems are constructed using low-cost of-the-shelf components which is one of their main advantages.

The main efforts of the future research will be focused on enabling the systems to operate reliably in real-world scenarios, i.e. in uncontrolled environments and with arbitrary scenes. 2D color sensing simultaneously with the coordinate measurement is also envisaged. The sensitivity, dynamic range and uncertainty of the measurement should be further improved on the basis of simulations. The simulations will permit to choose the most appropriate noise filtration technique. Comparative analysis of different filtration algorithms, concerning productivity and information losses, is also obligatory. As the sensitivity of the methods varies in broad limits in comparison with the other interferometric techniques, the inspection of the objects will be realized in a wide dynamic range and working conditions. Appropriate software for data processing – denoising, filtration, unwrapping and fast data translation will be developed in connection to the tasks of the other work packages – WP8,9,10,11 and 12.

4.6. Motion Analysis and Tracking

The work of ITI-CERTH and TUB for football player tracking still needs a ball tracking unit to be able to give more complete information about the football game. Once this is achieved, methods for reconstructing and projecting the game in 3D should be considered. The same extension is suitable for ITI 's smart visual tracking application in order to have a full and all around view of the area under surveillance.

4.7. Registration

TUT team has completed a study on developing an image registration framework aiming super-resolution for mobile devices. They proposed a feature-based approach, consisting of feature extraction, feature filtering, feature matching, and transformation estimation. In the application, the transformation estimation is robust to local distortions, and is accurate enough to allow for a subsequent super-resolution on the registered images. Implementation on hardware is also realized.

METU has been working on an improved corner detection to obtain better feature points. The corner angles, orientations and contrasts of a candidate corner set are extracted first. Then, a candidate corner is validated if the cornerness score directly obtained from the image is sufficiently close to the cornerness score for an ideal corner with the estimated orientation, angle and contrast. Tests on real and synthetic images showed that the algorithm improves the correct corner detection rates. This technique can be applied to stereo matching which is used for mobile image registration by TUT. Its potential for feature point matching with omnidirectional images will also be evaluated.

4.8. Object-Based Segmentation

In the further research the initialization step will be improved with the change from the expectation maximization algorithm to a parameterless approach as the mean shift algorithm. Methods of improving the Markov Random Fields based matte creation with a motion based matte prediction will be implemented and evaluated.

WP7 partners are permanently evaluating their current collaboration, the quality of the scientific results, and the correctness of the current overall research direction. Although some significant results were made to the field as a whole, the partners have identified several issues which they plan to improve in the future.

With 36 papers, the overall scientific output is very high. However, WP7 partners plan to actively increase the number of joint research projects, and as a consequence, also increase the number publications co-authored by researchers from multiple institutions. Currently, seven papers are a result of joint projects.

As part of future strategic planning, it will also be decided if the subtasks *Object Segmentation* and *Registration* should be further pursued. During the last period, the research focus and manpower somewhat shifted to the other subtasks which shows that they are considered more important by the partners. This development should therefore also be reflected in the overall research agenda.

Finally, the partners also agreed to further intensify their patenting efforts and work on more efficient internal reporting standards that facilitate the supervision of the work package by the work package manager as well the discussion on future research projects.

Despite these issues, the work package seven is a very efficient and successful joint research effort that produced important scientific results and spawned a fruitful and constantly progressing cooperation between top European research institutions. The partners are confident that the high scientific output will continue and that many novel joint research projects will be initiated in the future to attack the remaining technological challenges in the field.

References

[1] C. Theobalt, N. Ahmed, H. Lensch, M. Magnor, H.-P. Seidel, *Seeing People in Different Light: Joint Shape Motion and Reflectance Capture*, Accepted for publication in IEEE Transactions on Visualization and Computer Graphics, 2006.

[2] C. Theobalt, N. Ahmed, Edilson de Aguiar, Gernot Ziegler, H. Lensch, M. Magnor, H.-P. Seidel, *Joint Motion and Reflectance Capture for Creating Relightable 3D Videos*, MPI Technical Report, MPI-I-2005-4-004, 2005

[3] E. de Aguiar, R. Zayer, C. Theobalt, M. Magnor and H.-P. Seidel, "A Framework for Natural Animation of Digitized Models," Technical Report MPI-I-2006-4-003, Max-Planck-Institut fur Informatik, 2006.

[4] E. de Aguiar, C. Theobalt, H.-P. Seidel, "Automatic Learning of Articulated Skeletons from 3D Marker Trajectories," *In Proc. of ISVC'06*, Lake Tahoe, USA, 2006, pp. 485-494

[5] G. Ziegler, M. Magnor, and H-P. Seidel, *GeoCast: Unifying Depth Video with Camera Meta-Data, Proc. 2nd Workshop on Immersive Communication and Broadcast Systems*, Heinrich-Hertz-Institute, Berlin, Germany, 2005.

[6] V. Scholz and M. Magnor, *Texture Replacement of Garments in Monocular Video Sequences*, Rendering Techniques 2006 (Proc. of the 17th Eurographics Symposium on Rendering EGSR 2006), Nicosia, Cyprus, pp. 305-312, June 2006.

[7] A. Wenger, A. Gardner, C. Tchou, J.Unger, T. Hawkins, and P. Debevec, "*Performance relighting and reflectance transformation with time-multiplexed illumination,*" in *ACM TOG (Proc. of SIGGRAPH'05),* vol. 24(3), pp. 756–764, 2005.

[8] O. Schreer, P. Kauff, T. Sikora (Eds.), "3D Videocommunication: Algorithms, Concepts and Real-time Systems in Human Centered Communication", John Wiley & Sons Ltd, Chichester, England, 2005

[9] M. Pollefeys, R. Koch, M. Vergauwen, L. Van Gool, "Automated reconstruction of 3D Scenes from Sequences of Images", ISPRS Journal of Photogrammetry and Remote Sensing (55) 4, pp. 251-267, 2000

[10] C. Tomasi, T. Kanade, ”Shape and Motion from Image Streams: A Factorization Method”, Journal of Computer Vision 9(2), pp. 137-154, 1992

[11] R. Hartley, A. Zisserman, Multiple view geometry, Cambridge University Press, UK, 2003

[12] S. C. Park, M. K. Park, “Super-Resolution Image Reconstruction: A Technical Overview”, IEEE Signal Processing Magazine, May 2003, Volume:20, Issue:3, pp. 21-36.

[13] A. M. Tekalp, M. K. Ozkan, and M. I. Sezan, "High-resolution image reconstruction from lower-resolution image sequences and space varying image restoration," ICASSP, San Francisco, CA., Mar. 1992, vol. III, pp. 169--172.

[14] http://focus.ti.com/general/docs/wtbu/wtbuproductcontent.tsp?templateId=6123&navigationId=12314&contentId=4666

[15] C. Tomasi, T. Kanade, Shape and Motion from Image Streams: a Factorization Method Part 3: Detection and Tracking of Point Features, Carnegie Mellon University Technical Report CMU-CS-91-132, April 1991.

[16] C.G. Harris, M. Stephens, A Combined Corner and Edge Detector, Proc. of Fourth Alley Vision Conference, Manchester, 1988, pp.182-192.

5. Annex

Section	Publication
3.1	Multicamera
3.1.1.	Graph Cut based Panoramic 3D Modeling and Ground Truth Comparison with a Mobile Platform - The Wägele
3.1.2.	Multicamera Motion Capture for Articulated Body Model Animation
3.1.3.	Joint Motion and Reflectance Capture for Creating Relightable 3D Videos
3.1.4.	Seeing People in Different Light: Joint Shape Motion and Reflectance Capture
3.1.5.	Spatio-temporal Reflectance Sharing for Relightable 3D Video
3.1.6.	A Framework for Natural Animation of Digitized Models
3.1.7.	Automatic Learning of Articulated Skeletons from 3D Marker Trajectories
3.1.8.	Texture Replacement of Garments in Video
3.1.9.	Weighted Minimal Hypersurface Reconstruction
3.1.10.	Synchronous Image Acquisition based on Network Synchronization
3.2.	Single Camera
3.2.1.	A Geometric Segmentation Approach for the 3D Reconstruction of Dynamic Scenes in 2D Video Sequences
3.2.2.	Prioritized Sequential 3D Reconstruction in Video Sequences with Multiple Motions
3.2.3.	A Modular Scheme for 2D/3D Conversion of TV Broadcast
3.2.4.	Color High Dynamic Range (HDR) Imaging in Luminance- Chrominance Space
3.3.	Human Face and Body
3.3.1.	Automatic adaptation of face model to human face in video
3.3.2.	Active Appearance Models for Facial Features Detection
3.3.3.	3D Face Recognition by point signatures and iso-contours.
3.3.4.	PROSODY-DRIVEN HEAD-GESTURE ANIMATION
3.3.5.	3D Face Motion Capture
3.3.6.	Face Detection and Eye Localization in Video
3.4.	Holographic Camera Techniques
3.4.1.	Developments for underwater remote vision
3.4.2.	Real Time Capture and Reconstruction of Digital Holograms using Reconfigurable Hardware
3.4.3	Numerical Methods of Reconstruction employed in underwater Digital Holography (eHolocam)
3.5.	Pattern Projection
3.5.1.	Real-time acquisition of depth and color images using structured light
3.5.2.	Three-dimensional profilometry by symmetrical fringes projection technique
3.5.3.	Comparative study of fringes generation in two-Spacing phase-shifting profilometry
3.5.4.	Phase retrieval techniques in coordinate measurement
3.5.5.	Holographic and digital methods for recording and data processing in investigation of cultural heritage
3.5.6.	Real-time phase-stepping pattern projection profilometry
3.6.	Motion Analysis and Tracking
3.6.1.	FOOTBALL PLAYER TRACKING FROM MULTIPLE VIEWS
3.6.2.	Model-based Head Motion Estimation with an Automatic Update of Texture Information
3.6.3	Traffic Monitoring Using Multiple Cameras, Homographies and Multi-Hypothesis Tracking
3.7.	Registration
3.7.1.	An Image Registration Technique Aimed at Super Resolution on Mobile Devices
3.7.2.	Corner Validation based on Extracted Corner Properties
3.8.	Object-Based Segmentation
3.8.1.	Motion Based Segmentation for Automatic Initialization of Color Based Keyers