2020 3D Media
Spatial Sound and Vision

D9.6 Industry white papers
and Abstracts of research papers

<table>
<thead>
<tr>
<th>Project ref. no.</th>
<th>ICT-FP7-215475</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project acronym</td>
<td>2020 3D Media</td>
</tr>
<tr>
<td>Start date of project (dur.)</td>
<td>1 March, 2008 (48 months)</td>
</tr>
<tr>
<td>Document due Date</td>
<td>11 February, 2011</td>
</tr>
<tr>
<td>Actual date of delivery</td>
<td>28 February, 2011</td>
</tr>
<tr>
<td>Leader of this deliverable</td>
<td>BM</td>
</tr>
<tr>
<td>Reply to</td>
<td><a href="mailto:Santi.fort@barcelonamedia.org">Santi.fort@barcelonamedia.org</a></td>
</tr>
<tr>
<td>Document status</td>
<td>For internal peer review</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Version</th>
<th>Date</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>11/2/2011</td>
<td>Version for internal peer review</td>
</tr>
<tr>
<td>1.1</td>
<td>02/03/2011</td>
<td>Reworked after peer review.- Version sent to Coordinator</td>
</tr>
<tr>
<td>1.2</td>
<td>03/03/2011</td>
<td>Final pdf. submitted</td>
</tr>
</tbody>
</table>
## Deliverable Identification Sheet

<table>
<thead>
<tr>
<th>Project ref. no.</th>
<th>ICT-FP7-215475</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project acronym</td>
<td>2020 3D Media</td>
</tr>
<tr>
<td>Project full title</td>
<td>2020 3D Media Spatial Sound and Vision</td>
</tr>
<tr>
<td>Document name</td>
<td>D9.6_ Industry white papers and Abstracts of research papers</td>
</tr>
<tr>
<td>Security (distribution level)</td>
<td>Public (PU)</td>
</tr>
<tr>
<td>Contractual date of delivery</td>
<td>February 28, 2011</td>
</tr>
<tr>
<td>Actual date of delivery</td>
<td>March 3, 2011</td>
</tr>
<tr>
<td>Deliverable number</td>
<td>D9.6</td>
</tr>
<tr>
<td>Deliverable name</td>
<td>Industry white papers and Abstracts of research papers</td>
</tr>
<tr>
<td>Type</td>
<td>Report</td>
</tr>
<tr>
<td>Status &amp; version</td>
<td>Final</td>
</tr>
<tr>
<td>Number of pages</td>
<td>16</td>
</tr>
<tr>
<td>WP / Task responsible</td>
<td>WP9, Santi Fort</td>
</tr>
<tr>
<td>Other contributors</td>
<td></td>
</tr>
<tr>
<td>Author(s)</td>
<td>Santi Fort, BM</td>
</tr>
<tr>
<td>EC Project Officer</td>
<td>Jorge Santos</td>
</tr>
<tr>
<td>Abstract</td>
<td>This report summarizes the professional and academic publications generated by the partners of the 2020 3D Media project and that are reflected in the resources area of the project website.</td>
</tr>
<tr>
<td>Keywords</td>
<td>Academic publication; paper, poster</td>
</tr>
<tr>
<td>Sent to peer reviewer</td>
<td>11/2/2011, Jordi Alonso, Mediapro</td>
</tr>
<tr>
<td>Peer review completed</td>
<td>16/02/2011</td>
</tr>
<tr>
<td>Circulated to partners</td>
<td>Via plone</td>
</tr>
<tr>
<td>Mgt. Board approval</td>
<td>To be approved at the next SB meeting</td>
</tr>
</tbody>
</table>
Table of contents

1. Public Executive Summary ........................................................................................................ 4
2. List of papers ............................................................................................................................. 5
1. Public Executive Summary

This report summarizes the professional and academic publications generated by the partners of the 2020 3D Media project. This collection, that comprises up to January 2011, consists in a group of 49 accepted conference and journal papers and posters abstracts on 3D media technologies and 5 white papers from consortium both research and industrial partners.

The documents described in this report, may be found at the Resources area of the project website (http://www.20203dmedia.eu) under the sections Public documents and presentations and Academic publications. In the website, all references to documents, which are sorted by date, include title, date, authorship, abstract and the downloadable version of the actual document when available for public dissemination or the link to the scientific publisher who has the copyright.
2. List of papers

"Being inside the image. Heightening the sense of presence in a video captured environment through artistic means: the case of CREW."

This paper explores the use of omni-directional video, a high impact immersive medium, by a performance group and multi-disciplinary team of artists and scientists.  
By Nele Wynants, Kurt Vanhoutte, Theatre, Film and Literature Studies, University of Antwerpen and Philippe Bekaert, Expertise Centre for Digital Media, University of Hasselt

"Cognitive immersion"

The goal of this paper is to present our research about the cognitive processes involved in 3D immersion and present our research plans about how this type of environment should be helpful for people with special needs, in particular people with cerebral paralysis.  
By Elena Parra and Raquel Navarro, BM

"Video Enhancement Using Reference Photographs"

Handheld digital video cameras have become increasingly popular and cheaper in recent years. Even still cameras offer additional functionality for shooting videos. Unfortunately, the small form factor of these devices limits the light sensitivity, and often the lens and sensor do not allow for satisfactory image quality. We wish to process a video into a more aesthetically pleasing version, by borrowing information from high quality reference photographs of the same scene. Since the process of taking a photograph is not time-critical, we can afford a longer exposure for reducing noise, and record more information to increase resolution.  
By Cosmin Ancuti, Tom Haber, Tom Mertens, Philippe Bekaert, Hasselt University

"Eunomia: Toward a Framework for Multi-touch Information Displays in Public Spaces"

Multi-touch interaction techniques are becoming more widespread because of new industrial initiatives to make this hardware available and affordable for the consumer market. To cope with the diversity in hardware setups and the lack of knowledge about developing generic multitouch applications, we created a framework, Eunomia, for abstracting the hardware from the software and to enable software developers to easily develop interactive applications taking advantage of multi-touch interaction.  
By Tom Cuypers, Jan Schneider, Johannes Taelman, Kris Luyten, Philippe Bekaert, Hasselt University

"IT leverage for media acquisition: New paradigms in the key area of digital cinematography and HD production workflows"

This paper proposes the demonstration of a comprehensive novel acquisition infrastructure based on standard 10G Ethernet interface technology. This infrastructure is suitable for digital cinematography as well as HD production workflows and comprises camera, mobile solid state recorder and monitor all connected through a 10G Ethernet based network. Optical Ethernet
technology enables bidirectional, high speed, long distance networking at affordable cost as conventional IT equipment - such as switches - are fully supported. Interface and file formats enable versatile metadata support and provide a back channel to the camera for control purposes. Advances in solid state storage capacity in combination with interface performance eliminate any need for costly and quality compromising compression.

By Brune, T. Kochale, A. Wittenburg, J.P, Technicolor

"Time-triggered Static Schedulable Dataflows for Multimedia Systems"

Software-based reactive multimedia computation systems are pervasive today in desktops but also in mobile and ultra-portable devices. Most such systems offer a callback-based architecture to incorporate specific stream processing. Article presented at the Electronic Imaging - Multimedia Networking and Computing conference, in January 2009, in San Jose, California

By Pau Arumí, BM and Xavier Amatriain, Telefonica Research

"3D-Audio with CLAM and Blender's Game Engine"

Blender can be used as a 3D scene modeler, editor, animator, renderer and Game Engine. This paper describes how it can be linked to a 3D sound platform working within the CLAM framework, making special emphasis on a specific application: the recently launched Yo Frankie! open content game for the Blender Game Engine. The game was hacked to interact with CLAM, implementing spatial scene descriptors transmitted over the Open Sound Control protocol, and allowing to experiment with many different spatialization and acoustic simulation algorithms. Further new applications of this Blender-CLAM integration are also discussed.

By Natanael Olaiz, Pau Arumí, Toni Mateos and David García, BM

"A Framework for Networked Interactive Surfaces"

The development of interactive surfaces has led to a number of applications as it allows for natural interaction and collaboration. The use of co-located collaboration on this kind of surfaces provides new possibilities. It is our belief, however, that an even higher degree of collaboration can be achieved by overcoming the boundaries of a single device or setup. Therefore, we extended our previously built multi-touch framework to realize collaborative multi-device setups. In order to assess the framework, a networked virtual world and three other applications were built to demonstrate its power.

By Tom Cuypers, Karel Frederix, Chris Raymaekers, Philippe Bekaert, Univ. of Hasselt.

"Estimating 3D camera motion for rendering audio in virtual scenes"

In the production phase of digital cinema content it is important to allow the director not only to preview the final rendered scene early in the shooting process, but also to prehear the 5.1 surround and HRTF-based binaural versions, thus enabling visual and auditory artistic decisions to be taken at the shooting stage. In many cases camera tracking data is not available for all cameras on the set (e.g. handheld ones) and thus the motion of the camera needs to be estimated. In this paper we describe an approach for estimation of 3D camera motion and its use for real-time audio rendering.

By Werner Bailer, JRS, and Pau Arumi, Toni Mateos, Adan Garriga, Jaume Durany, David García, BM.
"Summarizing Raw Video Material Using Hidden Markov Models"

Besides the reduction of redundancy the selection of representative segments is a core problem when summarizing collections of raw video material. We propose a novel approach for the selection of segments to be included in a video summary based on Hidden Markov Models (HMM), which are trained on an annotated subset of the content. The observations of the HMM are relevance judgments of content segments based on different visual features, the hidden states are the selection/non-selection of content segments. The HMM is designed to take all relevant scenes into account. We show that the approach generalizes well when trained on sufficiently diverse content.

By Werner Bailer and Georg Thallinger, JRS.

"Depth from sliding projections"

In this paper we present a novel method for 3D structure acquisition, based on structured light. Unlike classical structured light methods, in which a static projector illuminates a scene with time-varying illumination patterns, our technique makes use of a moving projector emitting a static striped illumination pattern. This projector is translated at a constant velocity, in the direction of the projector's horizontal axis. Illuminating the object in this manner allows us to perform a per pixel analysis, in which we decompose the recorded illumination sequence into a corresponding set of frequency components. The dominant frequency in this set can be directly converted into a corresponding depth value. This per pixel analysis allows us to preserve sharp edges in the depth image. Unlike classical structured light methods, the quality of our results is not limited by projector or camera resolution, but is solely dependent on the temporal sampling density of the captured image sequence. Additional benefits include a significant robustness against common problems encountered with structured light methods, such as occlusions, specular reflections, subsurface scattering, interreflections, and to a certain extent projector defocus.

By Tom Cuypers, Karel Frederix, Chris Raymaekers, Philippe Bekaert, Univ. of Hasselt.

"Capturing Multiple Illuminations using Time and Color Multiplexing"

Many vision and graphics problems such as relighting, structured light scanning and photometric stereo, need images of a scene under a number of different illumination conditions. It is typically assumed that the scene is static. To extend such methods to dynamic scenes, dense optical flow can be used to register adjacent frames. This registration becomes inaccurate if the frame rate is too low with respect to the degree of movement in the scenes. We present a general method that extends time multiplexing with color multiplexing in order to better handle dynamic scenes. Our method allows for packing more illumination information into a single frame, thereby reducing the number of required frames over which optical flow must be computed. Moreover, color-multiplexed frames lend themselves better to reliably computing optical flow. We show that our method produces better results compared to time- multiplexing alone. We demonstrate its application to relighting, structured light scanning and photometric stereo in dynamic scenes.

By Bert De Decker, Jan Kautz, Tom Mertens and Philippe Bekaert, Univ. of Hasselt.

"Methodology for studying the numerical speed of sound in finite differences"

The space and time discretization inherent to all FDTD schemes introduce non-physical dispersion errors, i.e. deviations of the speed of sound from the theoretical value predicted by the governing Euler differential equations. A general methodology for computing this dispersion
error via straightforward numerical simulations of the FDTD schemes is presented. The method
is shown to provide remarkable accuracies of the order of 1/1000 in a wide variety of two
dimensional finite difference schemes.
By C. Spa, UPF, T. Mateos, A. Garriga, BM

"Shadow Multiplexing for Real-Time Silhouette Extraction"

Automatic detection and tracking of feature points is an important part of many computer vision
methods. A widely used method is the KLT tracker proposed by Kanade, Lucas and Tomasi.
This paper reports work done on porting the KLT tracker to the GPU, using the CUDA
technology by NVIDIA. Describing the CUDA port of the point tracker has been accepted as full
paper at the Computer Graphics, Computer Vision and Mathematics (GraVisMa) workshop.
By Tom Cuypers, Yannick Francken, Johannes Taelman, Philippe Bekaert, Univ. of Hasselt

"Realtime KLT feature point tracking for High Definition video"

Automatic detection and tracking of feature points is an important part of many computer vision
methods. A widely used method is the KLT tracker proposed by Kanade, Lucas and Tomasi.
This paper reports work done on porting the KLT tracker to the GPU, using the CUDA
technology by NVIDIA. describing the CUDA port of the point tracker has been accepted as full
paper at the Computer Graphics, Computer Vision and Mathematics (GraVisMa) workshop.
By Hannes Fassold, Jakub Rosner, Peter Schallauer and Werner Bailer, JRS

"Relighting Objects from Image Collections"

We present an approach for recovering the reflectance of a static scene with known geometry
from a collection of images taken under distant, unknown illumination. In contrast to previous
work, we allow the illumination to vary between the images, which greatly increases the
applicability of the approach. Using an all-frequence relighting framework based on wavelets,
we are able to simultaneously estimate the per-image incident illumination and the persurface
point reflectance. The wavelet framework allows for incorporating various reflection models. We
demonstrate the quality of our results for synthetic test cases as well as for several datasets
captured under laboratory conditions. Combined with multi-view stereo reconstruction, we are
even able to recover the geometry and reflectance of a scene solely using images collected
from the Internet.
By Tom Haber, Christian Fuchs, Philippe Bekaert, H.P. Seidel. Univ. of Hasselt

"Gloss and Normal Map Acquisition of Mesostructures Using Gray Codes"

We propose a technique for gloss and normal map acquisition of fine-scale specular surface
details, or mesostructure. Our main goal is to provide an efficient, easily applicable, but
sufficiently accurate method to acquire mesostructures. We therefore introduce a setup
consisting of cheap and accessible components, including a regular monitor and a digital still
camera. We build on our previously proposed method [Francken et al. 2008], which acquires
normal maps by analyzing the reflection of binary patterns (Gray codes). These patterns are
successively refined in this process. The key idea in this paper, is that this refinement also
allows us to measure the shininess for each spatial location, resulting in a gloss map. Acquiring
spatially varying reflectance usually requires a complicated hardware setup, which measures
the BRDF at each spatial location. Our method is much simpler and cheaper. Even though we
assume a very simplified BRDF model, our technique is able to reproduce the mesostructure's
appearance faithfully.
By Yannick Francken, Tom Cuypers, Tom Mertens, Philippe Bekaert. Univ. of Hasselt

"Metadata for Creation and Distribution of Multi-view Video"

This paper reviews the metadata requirements for multi-view video content and analyze how well these requirements are covered in existing metadata standards, both in terms of the coverage of metadata elements and the capabilities to structurally describe multi-view video content. Workshop on Semantic Multimedia Database Technologies (SeMuDaTe).
By Bailer W., Höffernig M. JRS

"Formal Metadata Semantics for Interoperability in the Audiovisual Media Production Process"

Different metadata formats and standards are used in the stages of the process, containing metadata properties with similar and partly overlapping semantics, though not fully identical. We attempt to model the metadata properties used throughout the production process in a format independent way by creating an ontology that models these properties and the relations between them.
By Höffernig M., Bailer W, JRS

"How to Align Media Metadata Schemas? Design and Implementation of the Media Ontology"

Multimedia data is generated, shared, stored and distributed worldwide at an ever increasing rate. This huge amount of content comes with metadata represented in different formats which hardly interoperate although they partially overlap.
By Florian Stegmaier, Werner Bailer, Tobias Bürger, et al., JRS

"Fast GPU-based KLT feature point tracking using CUDA"

Automatic selection and tracking of feature points is a basic task for many CV algorithms (e.g. structure from motion, object tracking) Very popular due to good performance: KLT algorithm [Kanade, Lucas and Tomasi]
By Florian Stegmaier, Werner Bailer, Tobias Bürger, et al., JRS

"Comparing Fact Finding Tasks and User Survey for Evaluating a Video Browsing Tool"

In the multimedia information retrieval community evaluations following the Cranfield paradigm (as e.g. used in TRECVID) have been widely adopted. We have applied two TRECVID style fact finding approaches (retrieval and question answering tasks) and a user survey to the evaluation of a video browsing tool.
By Werner Bailer and Herwig Rehatschek, JRS

"A Video Browsing Tool for Content Management in Post-production"
We propose an interactive video browsing tool for supporting content management and selection in postproduction. The approach is based on a process model for multimedia content abstraction. A software framework based on this process model and desktop and Web-based client applications are presented. International Journal of Digital Multimedia Broadcasting, 2010.

By Werner Bailer, Wolfgang Weiss, Gert Kienast, Georg Thallinger, and Werner Haas, JRS

"Integrative Workflow-Embedded Support For Process Automation And Management"

This paper presents the on-going research performed in order to integrate process automation and process management support in the context of media production. This has been addressed on the basis of a holistic approach to software engineering applied to media production modeling to ensure design correctness, completeness and effectiveness. European and Mediterranean Conference on Information Systems 2009

By Badii A., Fuschi D., Zhu M., Univ. of Reading

"Confronto sperimentale tra microfono B-format Soundfield e sonda intensimetrica di pressione e velocità Microflown"

Il funzionamento dei microfoni B-Format si basa sull'elaborazione dei segnali ottenuti da una configurazione coincidente tetraedrica di capsule microfoniche direzionali, che vengono combinati linearmente per fornire in uscita quattro canali, indicati con W, X, Y e Z, proporzionali rispettivamente alla pressione sonora ed alle tre componenti cartesiane del vettore di velocità acustica. Questi segnali vengono equalizzati per compensare effetti di interferenza dovuti alla non perfetta coincidenza spaziale delle capsule. Le sonde intensimetriche di ultima generazione Microflown sono costituite da un microfono di pressione e tre trasduttori di velocità operanti in base al principio dell'anemometria a doppio filamento caldo, permettendo così una misura diretta e pressoché coincidente delle variabili acustiche e v. Per via del particolare principio di funzionamento i sensori di velocità sono caratterizzati da una curva di risposta in ampiezza e fase descrivibile con una serie di tre filtri passa-basso con diverse frequenze di taglio, che devono essere accuratamente compensati nella fase di acquisizione del segnale.

By G. Cengarle, T. Mateos and D. Bonsi., BM

"A Narrow Band Method for the Convex Formulation of Discrete Multilabel Problems"

We study a narrow band type algorithm to solve a discrete formulation of the convex relaxation of energy functionals with total variation regularization and non convex data terms. We prove that this algorithm converges to a local minimum of the original non linear optimization problem. We illustrate the algorithm with experiments for disparity computation in stereo and a multi-label segmentation problem and we check experimentally that the energy of the local minimum is very near to the energy of the global minimum obtained without the narrow band type method.

By A. Baeza, V. Caselles, P. Gargallo, N. Papadakis., BM


This paper presents the on-going research performed in order to migrate from process automation to process management support in the context of media production and more specifically 3D cinematographic immersive and interactive production.
"An improved method to determine the onset timings of reflections in an acoustic impulse response"

Determining the absolute onset time of reflections in an acoustic impulse response (IR) has applications for both subjective and physical acoustics problems. Although computationally simple, a first-order energetic analysis of the IR can lead to false-positive identification of reflections. This letter reports on a new method to determine reflection onset timings using a modified running local kurtosis analysis to identify regions in the IR where the distribution is non-normal. IRs from both real and virtual rooms are used to validate the new method and to find optimum analysis window sizes.

By J. Uscher, BM

"A new technology for the assisted mixing of sport events: application to live football broadcasting"

This paper presents a novel application for capturing the sound of the action during a football match by automatically mixing the signals of several microphones placed around the pitch and selecting only those microphones which are close to, or aiming at, the action. The sound engineer is presented with a user interface where he or she can define and move dynamically the point of interest on a screen representing the pitch, while the application controls the faders of the broadcast console. The technology has been applied in the context of a three-dimensional surround sound playback of a Spanish first-division match.

By G. Cengarle, T. Mateos, N. Olaiz, P. Arumi, BM

"Measuring impulse responses using speech and music"

Continuous measurement of room impulse responses (RIRs) in the presence of an audience has many applications for room acoustics: in-situ loudspeaker/room equalization; teleconferencing; and for architectural acoustic diagnostics. A continuous analysis of the RIR is often preferable to a single measurement, especially with non-stationary room characteristics such as from changing atmospheric or audience conditions. This paper discusses the use of adaptive filters updated according to the NLMS algorithm for fast, continuous in-situ RIR acquisition; particularly when the input signal is music or speech. We show that the dual-channel FFT (DCFFT) method has slower convergence and is less robust to colored signals such as music and speech. Data is presented comparing the NLMS and the DCFFT methods and we show that the adaptive filter approach provides RIRs with high accuracy and high robustness to background noise using music or speech signals.

By J. Uscher, BM

"Multi-label depth estimation for graph cuts stereo problems"

We describe here a method to compute the depth of a scene from a set of at least two images taken at known view-points. Our approach is based on an energy formulation of the 3D reconstruction problem which we minimize using a graph-cut approach that computes a local minimum whose energy is comparable with the energy of the absolute minimum. As usually done, we treat the input images symmetrically, match pixels using photo consistency, treat collusion and visibility problems and we consider a spatial regularization term which preserves discontinuities. The details of the graph construction as well as the proof of the correctness of
the method are given. Moreover we introduce a multi-label refinement algorithm in order to increase the number of depth labels without significantly increasing the computational complexity. Finally we compared our algorithm with the results available in the Middlebury database.

By N. Papadakis and V. Caselles, BM

"Compensation of the afterglow phenomenon in 2-D discrete-time simulation"

Due to high computational costs, the physics governed by the wave equation in 3D is often modeled via discrete-time numerical simulations conducted in 2D scenarios. Results are normally generalized to 3D rather straightforwardly, overlooking the fact that the propagation of a point-like impulse in 2D exhibits the so-called afterglow phenomenon, which consists on the fact that non-null field values are measured after the arrival of the first wavefront. This paper analyzes the consequences of this phenomenon, both theoretically and numerically, and presents a simple method to compensate for it. Verification of this method is presented using various discrete-time numerical schemes.

By T. Mateos, J. Escolano, C. Spa, A. Garriga, BM

"Determining a scene geometry from its impulse response"

This paper reports on the development of a system to characterize the geometric properties of a space from an acoustic impulse response measured within it. This can be thought of as the inverse problem to the common practice of obtaining impulse responses from either a real-world or a virtual space. Starting from a measured or synthesized impulse response in an original scene, the method described here defines a space of possible scenes and a distance measure on it, and discusses a strategy to select a scene from this space, which is perceptually as close as possible to the original one.

Potential applications of this novel method include audio forensics, re-equalization and re-mixing of music and audio.

By David García-Garzón; Daniel Arteaga; John Usher; Toni Mateos, BM

"Evaluating Detection of Video Segments"

In this paper we have implemented several evaluation measures found in literature and we apply them to real algorithm outputs and a simulated result data set. We then calculate the correlation between the results obtained with the different measures in order to investigate whether they can be compared or not. The results show that the correlation between the measures is some cases quite low, and some measures are especially sensitive to certain types of deviations from the ground truth.

By Werner Bailer, JRS

"2020 3D Media: New directions in Immersive Entertainment"

(Poster abstract) This research project is conducted by a consortium of European industrial and academic partners that include companies like: Technicolor, Digital Projection, DTS Europe, Doremi, Mediapro, Creative Wokers (CREW) and Datasat, and research centers: Barcelona Media, Joaneum research, University of Hasselt, University of Reading and Fraunhofer. It is aimed to research, develop and demonstrate novel forms of compelling entertainment experiences based on new technologies for the capture, production, networked distribution and display of three-dimensional sound and images.
"Spatial string matching for image classification"

This paper presents a spatial string matching method to incorporate spatial information into the bag-of-words model, which represents an image as an unordered distribution of local features. Spatial constraints among neighboring features are explored in order to achieve better discrimination power for image classification. The features from neighboring points are combined together and taken as a spatial string, and then our method matches the images according to the similarity of string pairs. The categorization problem can be formulated using KNN or SVM classifier based on the spatial string matching kernel. The proposed method is able to capture spatial dependencies across the neighboring features. Experiment results show promising performance for image classification tasks.

By Yunqiang Liu and Vicent Caselles, BM

"Stereoscopic image inpainting: distinct depth maps and images inpainting"

In this paper we propose an algorithm for in painting of stereo images. The issue is to reconstruct the holes in a pair of stereo image as if they were the projection of a 3D scene. Hence, the reconstruction of the missing information has to produce a consistent visual perception of depth. Thus, first step of the algorithm consists in the computation and in painting of disparity maps in the given holes. The second step of the algorithm is to fill-in missing regions using the complete disparity maps in a way that avoids the creation of 3D artifacts. We present some experiments on several pairs of stereo images.

By Alexandre Hervieu, Nicolas Papadakis, Aurélie Bugeau, Pau Gargallo, Vicent Caselles, BM

"A New Infrastructure for High Resolution 3D and Slow Motion Production-Sets Featuring the 'FlashPak II' Field-Recorder Demonstrator"

The current success of immersive 3D experiences in feature films and the resulting move to push such technology into TV call for the extension of today's standard digital HD or 2K/4K capture process towards stereo imaging or other means of depth representation. Recently, stereoscopic high resolution capture rigs are employing proprietary workflows and a non-unified infrastructure based on legacy technology taken from the 2D domain. As this sudden demand for at least doubled recording bandwidth and storage space are clearly stretching the practicability limits of this legacy infrastructure, the already overdue paradigm shift towards a modern, future proof capture and recording infrastructure is getting another push. This paper describes a comprehensive infrastructure approach based on new interface and field recording technologies for capture devices that targets the challenging requirements of today's and future 3D production setups. The new interface approach harmonizes and consolidates image data, depth information and general metadata to ease 3D production workflows. Furthermore, a handy field recorder solution based on solid state technology is introduced that is able to record up to seven uncompressed camera streams in parallel.

By T. Brune and J.P. Wittenburg, Technicolor

"CUDA Acceleration of Color Histogram Matching"

A common approach to histogram matching is done by means of the cumulative distribution functions (CDFs). First, we calculate their normalized histograms (hA and hB) and their respective cumulative histogram distribution functions(CDFA and CDFB). Then, a matching
between the CDFs are performed. Given the reference CDF, CDFA, for each gray level GI we find the corresponding gray level GJ in which CDFA(GI) = CDFB(GJ), if I and J correspond to different values then a replacement of a gray level in the image IB is performed in order to match their CDFs. Our approach considers the ideas of [Rolland et al. 2000] with a Nvidia 3D broadcast solution system using professional HD cameras.

By Antonio S. Montemayor, Raúl Cabido, Juan José Pantrigo, URJC, and Xavi Rodríguez, Sergi Sagàs, Mediapro

"Polyconvexification of the multi-label optical flow problem"

In this paper the problem of optical flow and occlusion mask estimation is raised. To that end, we consider a multi-label representation of the optical flow and we define an energy that models the problem. The convexification of the energy and its minimization with an iterative algorithm are studied. Our algorithm is implemented in GPU, since each pixel can be processed in parallel. From our experiments, the relation between the quality of the results obtained and computing time seems to be very promising.

By N. Papadakis, A. Baeza, P. Gargallo, BM and V. Caselles, UPF

"A Comprehensive Framework for Image Inpainting"

Inpainting is the art of modifying an image in a form that is not detectable by an ordinary observer. There are numerous and very different approaches to tackle the inpainting problem, though as explained in this paper, the most successful algorithms are based upon one or two of the following three basic techniques: copy-and-paste texture synthesis, geometric partial differential equations (PDEs), and coherence among neighboring pixels. We combine these three building blocks in a variational model, and provide a working algorithm for image inpainting trying to approximate the minimum of the proposed energy functional. Our experiments show that the combination of all three terms of the proposed energy works better than taking each term separately, and the results obtained are within the state-of-the-art.

By A. Bugeau, M. Bertalmio, V. Caselles, G. Sapiro, BM, UPF

"Tracking and Clustering Salient Features in Image Sequences"

Many applications in media production need information about moving objects in the scene, e.g. insertion of computer-generated objects, association of sound sources to these objects or visualization of object trajectories in broadcasting. We present a GPU accelerated approach for detecting and tracking salient features in image sequences and we propose an algorithm for clustering the obtained feature point trajectories in order to obtain a motion segmentation of the set of feature trajectories. Evaluation results for both the tracking and clustering steps are presented. Finally we discuss the application of the proposed approach for associating audio sources to objects to support audio rendering for virtual sets.

By Jakub Rosner, Silesian University and Werner Bailer, Hannes Fassold, Felix Lee, JRS

"Virtual Camera Synthesis for Soccer Game Replays"

In this paper, we present a set of tools developed during the creation of a platform that allows the automatic generation of virtual views in a live soccer game production. Observing the scene through a multi-camera system, a 3D approximation of the players is computed and used for the synthesis of virtual views. The system is suitable both for static scenes, to create bullet time effects, and for video applications, where the virtual camera moves as the game plays.

By N. Papadakis, A. Baeza, X. Armangué, I. Rius, A. Bugeau, O. D'Hondt, P. Gargallo, V. Caselles and S. Sagas, BM, Mediapro
"Unsupervised Motion Layer Segmentation by Random Sampling and Energy Minimization"

In this paper we introduce an unsupervised scheme for the segmentation of motion layers in video sequences. The number of layers is automatically determined by the method. Our approach first extracts the motion models thanks to a RANSAC-based random sampling algorithm improved by the use of geodesic distance information. Then those models are assigned to pixels in the color image by minimizing an energy functional thanks to graph-cut. Our energy takes into account motion residuals, color distributions, geodesic distance as well as temporal consistency of the layers. Moreover, we define a smoothness term that enforces a patch-wise spatial coherency on areas where optical flow is reliable and a pixel-wise coherency on occluded areas. The method leads to promising results on the tested sequences.

By Olivier D'Hondt and Vicent Caselles, BM

"A Comprehensive Infrastructure and Workflow for Acquisition of High-resolution Multi-view Content"

The current success of immersive 3D experiences in feature films and the trend towards 3D TV require advanced tools and workflows for high-quality capture of multi-view live scenes including depth information. These requirements are not fulfilled by today's capture workflows and infrastructures based on legacy technology from 2D capture solutions. We propose a comprehensive infrastructure for capture of multiview content supporting different methods to obtain depth information. The paper describes a device for field capture and tools for online and offline content analysis and browsing of the captured content that have been integrated into the proposed infrastructure.

By Werner Bailer, Christian Schober, JRS and Thomas Brune, Technicolor

"A Variational Framework for Exemplar-Based Image Inpainting"

In this paper we propose an algorithm for inpainting of stereo images. The issue is to reconstruct the holes in a pair of stereo image as if they were the projection of a 3D scene. Hence, the reconstruction of the missing information has to produce a consistent visual perception of depth. Thus, first step of the algorithm consists in the computation and inpainting of disparity maps in the given holes. The second step of the algorithm is to fill-in missing regions using the complete disparity maps in a way that avoids the creation of 3D artifacts. We present some experiments on several pairs of stereo images.

By Pablo Arias, Gabriele Facciolo, Guillermo Sapiro, Vicent Caselles, BM

"Stereoscopic Image Inpainting Using Scene Geometry"

In this paper we propose an algorithm for inpainting of stereo images. The issue is to reconstruct the holes in a pair of stereo image as if they were the projection of a 3D scene. Hence, the reconstruction of the missing information has to produce a consistent visual perception of depth. Thus, first step of the algorithm consists in the computation and inpainting of disparity maps in the given holes. The second step of the algorithm is to fill-in missing regions using the complete disparity maps in a way that avoids the creation of 3D artifacts. We present some experiments on several pairs of stereo images.

By Alexandre Hervieu, Nicolas Papadakis, Aurélie Bugeau, Pau Gargallo, Vicent Caselles, BM

"Video Browsing Using Object Trajectories"

Video browsing methods are complementary to search and retrieval approaches, as they allow
for exploration of unknown content sets. Objects and their motion convey important semantics of video content, which is relevant information for video browsing. We propose extending an existing video browsing tool in order to support clustering of objects with similar motion and visualization of the objects’ positions and trajectories. This requires the automatic extraction of moving objects and estimation of their trajectories, as well as the ability to group objects with similar trajectories. For the first issue we describe the application of a recently proposed motion trajectory clustering algorithm, for the second we use \( k \)-medoids clustering and the dynamic time warping distance. We present evaluation results of both steps on real world traffic sequences from the Hopkins155 data set. Finally we describe the description of analysis results using MPEG-7 and the integration into the video browsing tool.

By Felix Lee and Werner Bailer, JRS