IMPART: BIG MEDIA DATA PROCESSING AND ANALYSIS FOR FILM PRODUCTION

Josep Blat[1], Alun Evans[1], Javi Agenjo[1], Hansung Kim[2], Evren Imre[2], Adrian Hilton[2], Anastasios Tefas[3], Nikos Nikolaidis[3], Ioannis Pitas[3], Lukas Polok[4], Pavel Smrz[4], Pavel Zemcik[4]

[1]Universitat Pompeu Fabra, Spain; [2]University of Surrey, UK;
[3]Aristotle University of Thessaloniki, Greece; [4]Brno University of Technology, Czech Republic
josep.blat@upf.edu, a.hilton@surrey.ac.uk, pitas@aiia.csd.auth.gr, zemcik@fit.vutbr.cz

ABSTRACT

A typical high-end film production generates several terabytes of data per day, either as footage from multiple cameras or as background information regarding the set (laser scans, spherical captures, etc). The EU project IMPART (impart.upf.edu) has been researching solutions that improve the integration and understanding of the quality of the multiple data sources to support creative decisions on-set or near it, and an enhanced post-production as well. The main results covered in this paper are: a public multisource production dataset made available for research purposes, monitoring and quality assurance of multicamera set-ups, multisource registration, anthropocentric visual analysis for semantic content annotation, acceleration of 3D reconstruction, and integrated 2D-3D web visualization tools.

Index Terms— Multi-modal data processing, big media data analysis, web 3D visualization

1. INTRODUCTION

The amount of data captured onset for film production is vastly increasing, and several Terabytes are generated per day for a typical high-end film. This paper deals with data coming from varied capture devices such as LIDAR scanners, spherical cameras, still cameras, HD video cameras, 2.7K/4K cameras and RGBD cameras (Fig. 1 shows some of the sources). While data storage is getting cheaper, all of this data needs to be sorted, indexed and processed, which requires an immense amount of manual task. In fact, the high volume of data generation during a shot prevents the immediate assessment of whether footage is fit for purpose. The current solution, "if in doubt, re-shoot", leads to potentially redundant data, costing extra time and money. On the other hand, even more data is generated during post-production, as the raw input is usually processed multiple times, in order to obtain the output desired by the director.

This process needs to be streamlined, and understanding the quality of data, and its content are key aspects, to invert the trend of producing and storing even more data, towards keeping the suitable data instead. IMPART has been researching into these issues and the paper outlines several of the contributions, which we indicate very briefly next:

a) A multisource dataset which is representative of production data, captured with a wide variety of devices and in different environments, which has been made public for research purposes.

b) Tools to monitor setups, and to enhance quality assurance in the context of multisource data acquisition and processing.

c) Automatic registration of multisource data into a common coordinate system (a "unified 3D space", visually represented in Fig. 1) for efficient data management, and improved post-production.

d) High-level visual information analysis for human-centered semantic metadata extraction and description, to support fast big visual data ingestion, search and retrieval for post-production and archival use.

e) Re-formulation of 3D reconstruction from still images, leading to much faster processing, while providing at the same time enhanced quality assessment

f) Integrated interactive web visualization of 2D and 3D source and processed large data and metadata, for increased efficiency in quality assessment, creative on-set decisions and post-production planning.

Most of the research advances presented in this paper have been already integrated in tools used by industrial partners of the project, and have been tested on film related material, some of it from actual productions, some of it part of the research dataset mentioned previously.

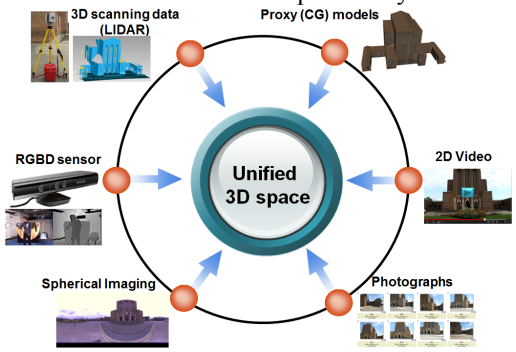

Fig. 1. Multisource data processing and managing.

2. DATA ACQUISITION AND REGISTRATION

2.1. Public Multisource Production Datasets

To support researches in multisource data processing, we provide a big production database (around 10TB, uncompressed) acquired in various indoor and outdoor environments. Various capture devices such as LIDAR scanners, spherical cameras, still cameras, HD video cameras, 2.7K/4K cameras and RGBD cameras were used. The material of other public datasets made available for research has been captured by a single device in a constrained environment.

The dataset with detailed notes on the capture configuration and data formats is available at: http://cvssp.org/impart/. The figures shown in the paper correspond to material coming from this dataset.

2.2. Set-up Monitoring and Quality Assurance

The set-up monitoring and quality assurance toolset developed within IMPART enables on-set detection of capture problems such as poor coverage, calibration invalidation and synchronization loss.

Poor coverage occurs when the camera configuration is out-of-focus, poorly framed or lacks the necessary detail. This problem is detected by placing a "sphere cloud", a synthetic lattice with fixed-size volume elements attached to each vertex, and analyzing the statistics of this projection over the set-up. This offers a prediction of the quality of the coverage, before the footage is captured [1].

Calibration invalidation occurs when the pose or focal length of some elements of the set-up is accidentally altered after calibration. It is detected by comparing the relative calibration predicted from the existing calibration hypothesis to that estimated from the image pairs. This approach successfully identifies the cameras with invalid or unreliable calibration. The true camera parameters can be recovered on set or by repairing the calibration in post-production.

Synchronization loss and frame drops can occur due to various hardware problems. The IMPART synchronization tool builds upon [2], whose key observation is that a feature on a dynamic scene element satisfies the epipolar constraint, only if the associated image pair depicts the same time instant. This leads to an image similarity metric, with which the indices of the corresponding frames are identified. Synchronization parameters and frame-drop events are estimated by fitting a 2D broken-line to the index correspondences for each pair, and fusing the pairwise estimates. IMPART improves [2] for faster and more robust operation, so that the algorithm can meet the demands of on-set operations under unfavorable conditions.

2.3. Multisource Data Registration

A unified 3D space where 2D and 3D data are registered is introduced for efficient multisource data management as illustrated in Fig. 1. 3D data from active sensors is directly registered to the reference coordinates through 3D feature detection and matching. 2D footage is registered via 3D reconstruction such as stereo matching or structure-from-motion techniques.

A robust multi-domain 3D feature description method and hybrid RANSAC matching algorithms were developed and integrated into the Double Negative's Jigsaw software [3] which offers a flexible user interface for organizing massive collections of data files.

Multisource data can be automatically registered with the proposed registration pipeline. It received very positive feedback "the quality achievable in a short amount of time and with minimal user input allowed for a much improved throughput" from artists at film industry.

3. SEMANTIC CONTENT ANALYSIS

Semantic content annotation through anthropocentric visual analysis can highly contribute to film production, by enabling novel functionalities for fast annotation, browsing and preview of footage, as well as retrieval of the most relevant streams of the entire available footage. To this end, Aristotle University of Thessaloniki (AUTH) has worked on semantic temporal video segmentation beyond the standard shot-cut detection, exploiting motion information. The video segment is the elementary asset in which we can apply video segment clustering for summarization and batch processing, face detection/ recognition/clustering, person activity recognition and saliency detection. A multi-view frame with the automatically extracted metadata (i.e., temporal segmentation timeline, semantic view in yellow frame, activity and person recognized) is depicted in Fig.3.

Fig. 3. Semantic metadata of multi-view video.

In more detail, the objective of temporal video segmentation is to produce video segments depicting single human activities, thus enabling fast activity-based content summarization, key-frame extraction and browsing of takes/dailies based on the activities performed. Such a segmentation scheme can greatly contribute to activity

clustering, allowing the determination of similar scenes, groups of activities, etc. and the application of the same series of treatments (i.e., filters, enhancements) to similar video segments.

Fast face clustering and recognition of actors needs face detection and/or tracking as first step. In a multi camera set up it is important to have distributed and multi-core, multi-threaded implementations exploiting also the existing algorithmic methodologies for fast computations. Thus, in Brno University of Technology (BUT) a fast distributed clustering framework has been proposed [4] that can be exploited in many different large scale learning problems where fast similarity matrix construction is needed.

In terms of performance, the accuracy of face and activity recognition algorithms is very important for the usability and acceptability of the developed tools by the industry. To this end, we have improved the state-of-the-art on facial image analysis in many ways. The most recent development was the exploitation of the face symmetry in the learning tasks for improving the recognition accuracy [5]. Production-related automatic recognition of specific human activities is very important in background scene construction (e.g., crowd replication). Thus, we have proposed several neural network based methods for human activity recognition. One of the proposed approaches combines the linear discriminant analysis objective with fast randomized neural network learning [6]. Finally, as video summarization is crucial for handling big media data, we have proposed a saliency-based content selection and semantic summarization method [7]. All the proposed approaches have been tested in various film related datasets with promising results.

4. ACCELERATION AND QUALITY ASSESSMENT OF 3D RECONSTRUCTION

Tools for 3D reconstruction from stills [8] are very popular in film production, especially for special effects support. Although there is quite a variety of implementations, they are usually too slow to be used on-set, and there is no indication of quality of the reconstruction that could be relied upon. The University of Surrey has developed an efficient Bundle Adjustment (BA) solver, which can process production-scale data in reasonable time, and as a bonus it also provides reconstruction quality indication (see an example in Fig. 4), which can be displayed to the user and more data can be captured to improve the reconstruction, if needed.

While formulating BA as an estimation problem using a graphical model leads to flexibility, representing it as a sparse *block* matrix yields significant computational benefits in comparison with conventional sparse matrices. Block methods [9] have been known in the literature for some time, but were generally not used. We implement sparse *block* matrix BLAS, which exploits the inherent block structure of the BA problem. This yields important performance advantages, enabling loop unrolling and vectorization in the BLAS kernels. We further show that the same optimization can be done for a GPU implementation.

In order to provide quality assessment of the final reconstruction, marginal covariances are calculated. The marginal covariance matrix is a square matrix with the same dimension as the sum of the optimized variables, it equals the inverse of the system matrix and is fully dense. For practical size problems, such a matrix would not fit into memory of most today's systems, therefore an algebraic manipulation is employed to calculate only sparse parts of the inverse. We further propose an efficient algorithm to update this inverse incrementally, yielding two orders of magnitude speedup, compared to other state of the art graphical solvers.

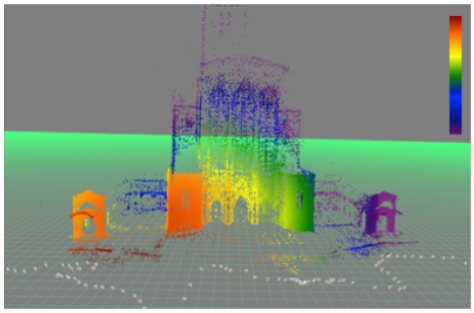

Fig. 4. Quality assessment of the 3D reconstruction (levels of quality are color coded).

The developed tools were evaluated on production data from the *Fast and Furious 6* movie. In the original production, calculating the 3D reconstruction took 8 hours on a workstation. With the optimized implementation, it is possible to do the same in a single hour, on a laptop, including the quality estimation.

Currently, the solution is limited to reconstruction from stills. Spherical cameras are another interesting modality, and it would be interesting to formulate 3D reconstruction from both stills and spherical capture.

5. INTEGRATED WEB VISUALIZATION TOOLS

The advent of cloud technology has revolutionized modern digital workflows, with remote and collaborative workflows and web-based tools now becoming common in the workplace. In the digital production world, interactive web tools allow easier onset sharing and facilitate creative decisions. With the advent of WebGL, hardware accelerated 3D graphics applications can now present multimodel 3D data, such as that created by this project, via a web-based interface. Universitat Pompeu Fabra has developed an integrated 2D-3D interactive web visualization of multi-source (processed) data, namely laser scans, 3D reconstructed from 2D, spherons, video, metadata coming from other Partners (see an example on Fig. 5).

Our work extends and goes beyond the dual-mode user interface paradigm of Jankowski and Decker [10], which

proposed a tighter integration between traditional web hypermedia and 3D web graphics. We use the multimodal data of the project to create a hybrid visualization, which mixes visualization of 2D images and video with registered 3D data, allowing users to obtain an overview of all the recorded data in a single application.

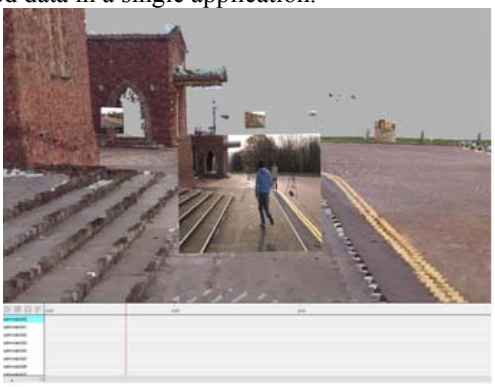

Fig. 5. Multimodal web-based visualisation.

One of the major challenges of web-based 3D applications is related to the transmission of data from the server to the remote client. Our work is based on solutions for the progressive download of very large point-cloud data [11] generated within the project, which are comparable in efficiency to the state-of-the art proposed for the more widely studied progressive download of meshes (see [12], for instance).

While the performance and quality are more limited than what could be obtained through desktop based solutions, our future work plans to taking advantage of browser access to accelerometers, cameras, etc. which permits mixed reality applications and potentially new methods of viewing multimodal data.

7. DISCUSSION AND PERSPECTIVES

In this paper we have presented several approaches to improve the processing – and subsequent managing - of the large multisource data generated onset in typical high-end productions.

One of the key issues is to understand the quality of the data, for instance, the fitness-for-purpose of the data in our case. We presented a range of tools, from monitoring the quality of the set-ups to capture the data, e.g., in terms of coverage, through visualizing in a more integrated way both 2D and 3D raw and processed data, to speeding up the 3D reconstruction while offering as a bonus an assessment of its quality. As a measure of their applicability, these tools have been integrated in the software used by one of the industrial partners of the project (Double Negative) to organize and process the vast data captured.

From another point of view, the results advance the state-of-the art in several aspects, e.g., in the processing of multisource data, in integrated visualization of (very large) 2D and 3D, in orders of magnitude speed up of graphical solvers, and could be applied in other big data areas, besides film production and postproduction.

Some future research work, such as applying the 3D reconstruction methods used in still images to spherical ones, or moving the integrated visualizations towards augmented reality tools to support onset creative decisions has been mentioned.

ACKNOWLEDGEMENTS: This work was supported by the European Commission, FP7 IMPART project (grant agreement No 316564).

8. REFERENCES

[1] E.Imre and A. Hilton, “Coverage Evaluation of Camera Networks for Facilitating Big Data Management in Film Production,” accepted to *ICIP 2015*.

[2] E. Imre, J.-Y. Guillemaut and A. Hilton, “Through-the-Lens Multi-Camera Synchronisation and Frame-Drop Detection for 3D Reconstruction,” in *3DIMPVT 2012*. 3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT), 2012 Second International Conference on , pp.395,402, 13-15 Oct. 2012

[3] S Pabst, H.Kim, L.Polok, V.Ila, T.Waine, A.Hilton, J.Clifford. “Jigsaw - Multi-Modal Big Data Management in Digital Film Production”. Accepted at SIGGRAPH 2015.

[4] N. Tsapanos, A. Tefas, N. Nikolaidis and I. Pitas, “A distributed framework for trimmed Kernel k-Means clustering”, *Pattern recognition*, 2015.

[5] K. Papachristou, A. Tefas and I. Pitas, "Symmetric Subspace Learning for Image Analysis", *IEEE Trans. on Image Processing*, 2014.

[6] A. Iosifidis, A. Tefas and I. Pitas, "Minimum Class Variance Extreme Learning Machine for Human Action Recognition", *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 23, no. 11, 1968-1979, 2013.

[7] V. Mygdalis, A. Iosifidis, A. Tefas and I. Pitas, "Exploiting Subclass Information in One-class Support Vector Machine for Video Summarization", in *ICASSP 2015*, Brisbane, Australia, 19-24 April, 2015.

[8] Noah Snavely, Steven M. Seitz, Richard Szeliski. “Modeling the World from Internet Photo Collections”. *International Journal of Computer Vision*, Vol 80(2) Pages 189-210, November 2008

[9] Ng, Esmond G., and Barry W. Peyton. "Block sparse Cholesky algorithms on advanced uniprocessor computers." *SIAM Journal on Scientific Computing* 14, 5: 1034-1056. 1993

[10] Jankowski, J., Decker, S.. On the design of a dual-mode user interface for accessing 3d content on the world wide web. *International Journal of Human-Computer Studies 71*, 7, 838–857. 2013

[11] A.Evans, J.Agenjo, J.Blat Web-based Visualisation of On-set Pointcloud Data. 11th European Conference on

Visual Media Production (CVMP2014), London, England (November 2014)

[12] G. Lavoué, L. Chevalier, and F. Dupont. Streaming Compressed 3D Data on the Web using JavaScript and WebGL. In *ACM International Conference on 3D Web Technology (Web3D)*, San Sebastian, Spain, pages 19–27, 2013. □