Foundations of a New Interaction Paradigm for Immersive 3D Multimedia

I. Galloso, F.P. Luque-Oostrom, L. Piovano, D. Garrido, E. Sánchez, C. Feijóo
Center for Applied ICTs – Technical University of Madrid
UPM-CEDINT

NEM 2012 (Istanbul), October 18th, 2012

Index

1. Introduction
2. ImmersiveTV Project
3. UPM-CeDInt contribution: research for immersive and interactive 3D Multimedia
4. II-3DM Prototype at the I-Space Infrastructure at UPM-CeDInt VR Lab
5. Results
6. Conclusions and future works

1. Introduction (I)

INTERACTIVITY

Communication:

- Human-to-human vs. human-to-machine (e.g., computer, cars, devices);
- It involves psychology, computer and information science, industrial design, …;
- Bi-directional flow of information (action-reaction principle);
- Communication channel: interfaces to play an active role over the environment

IMMERSION

Perception of the environment:

- Basically, a mental state connecting the self-awareness and the surroundings;
- It involves psychology, computer science, technology, medicine (nervous system, brain physiology, …);
- Full involvement of all the senses;
- Envelopment, inclusion and interaction

USER EXPERIENCE

Evaluation of quality of experience (QoE):

- User-centered approach;
- Active exploration of the environment by exploiting interactive communication and facilities;
- Non-linearity of the multimedia content progress;
- Commercial opportunities and new services exploitation when technological and social impact are meeting

1. Introduction (II)

IMMERSIVE ENVIRONMENT

CAVE™:

- A theater made up by a variable number of large projection screens (walls) and arranged in several configurations + high resolution projectors;
- 3D stereoscopic content seen through specific glasses;
- User movements tracking through infrared cameras;
- Stereo speakers;
- Remote controller devices;

INTERACTIVE MULTIMEDIA

Content production and enrichment:

- Stereo video streaming;
- Panoramic display;
- Insertion of synthetic objects onto the original streaming;
- HCI policies;

ADDED VALUES

Towards the future:

- Exploiting the higher level of stereo displaying techniques;
- Bringing CAVEs environment beyond usual VR approach;
- Inserting Augmented and Mixed Reality components and paradigms;
- Evaluation of user's reactions in order to properly tailor further applications

2. ImmersiveTV project

“ImmersiveTV: An approximation to immersive media”. AVANZA 2010, MICINN

Relevant steps across the entire value chain:

1. Analysis of requirements;
2. Content creation and post-production (3D multimedia and interactive computer-generated objects);
3. Manipulation, codification and transmission (hybrid approach: broadcast + internet);
4. Reception, visualization and interaction;
5. Experimental pilots: Assessment of the QoE in immersive and interactive 3D Multimedia applications

3. CeDInt VR Lab Goals (I)

COMPUTER VISION

Synthetic vs. traditional audio-visual content:

- Well-known 3D displacements vs. estimation of depth layers;
- Well-known 3D objects geometry vs. estimation of shapes and models;
- Multi-view allowed vs. fixed point of view;
- 3D audio-visual content comes at a very expensive cost;
- Much more traditional stereo stuff;

INTERACTIVE MULTIMEDIA

Main functionalities:

- Generation of immersive, panoramic video contents to be visualized in a multi-screen projection system;
- Merge external 3D elements to create hybrid content;
- Visualization in an immersive space with real-time interaction (e.g., zoom, selection of a point of view, 3D models manipulation, …)

USER EXPERIENCE

User's assessment and feedback on:

- Natural interaction (e.g., easiness in its use, learning overhead, …);
- Quality of visual experience (e.g., 3D rendering and effects)
- Possible technical limitations (e.g., controller device, interaction limits, …);
- Possible contraindications (e.g., headache or sickness, drawbacks after continuous use, …)

3. CeDInt VR Lab Goals (II)

COMPUTER VISION

Dense two-frame stereo matching algorithm:

- Video acquisition from a generic couple of commercial cameras;
- Video frame rectification (for each pair of contiguous views);
- Dense two-frame stereo matching techniques for depth map estimation from uncalibrated stereo video sequences;
- Segmentation, borders detection and other image processing techniques to enhance depth maps resulting from stereo matching;
- Creation of the panoramic content (both video and depth-maps);

INTERACTIVE MULTIMEDIA

Insertion and interaction algorithms:

- Definition of the user interaction paradigm (e.g. objects to be inserted, interaction devices, scenario features, …);
- Real-time computation of occlusions;
- Real-time interaction;
- Synchronization between interaction and video streaming features;

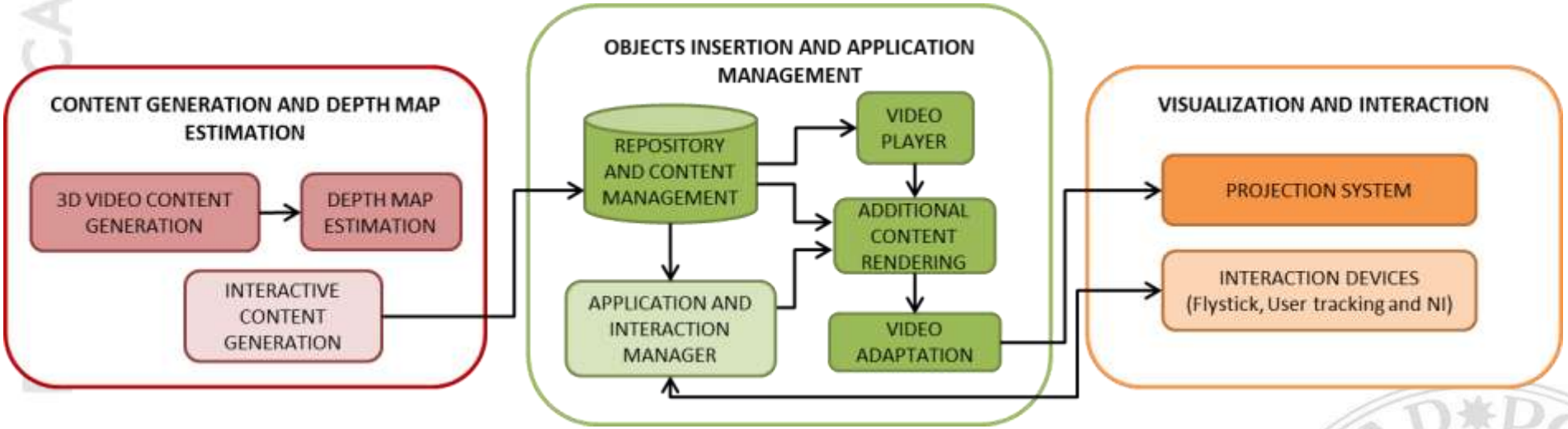
USER EXPERIENCE

Evaluation of quality of experience (QoE):

- Overall quality of the interactive and immersive multimedia experience: subjective and objective assessment
- New Measures: Physiological signals, gestural behavior, interactions and questionnaires
- Influence of media form (focus on immersion and interaction) in the emotional response, attention and perceived presence

COMING SOON

3. Prototype architecture



3. Panoramic stereo videos

TECHNICAL FEATURES

- Elongated field of view (greater than human sight , that is ≥ 160º and ≥ 75º)
- Composition of multiple video patches (e.g., by stitching) by exploiting slightly overlapping areas;
- Stereo acquisition rig (2D cameras) with (possibly) known geometry (displacement and angle of convergence) and technical camera parameters (focal length, calibration setups, …)
- Stereo geometry to be algorithmically reconstructed

PROS

- Fitting naturally into a 5-sided CAVE™
- Real sense of immersion
- Huge field of depth
- Based on real content;
- Simple extension of TV- or cinema-like application approach;

CONS

- Geometrical distortion of original videos
- Depth estimation is usually not so accurate for furthest objects;
- Possible different illumination conditions across views:
  - Visible discontinuities in adjacent frames;
  - Difficulties in matching the same objects across views;
- Noise (e.g., from cameras, unfocused details, undistinguishable patterns, …)

3. Depth estimation (I)

PROBLEM STATEMENT

Stereo matching algorithm:

Given (at least) a pair of images representing the same scene taken from different points of views, describe how far they are with respect to the observer.

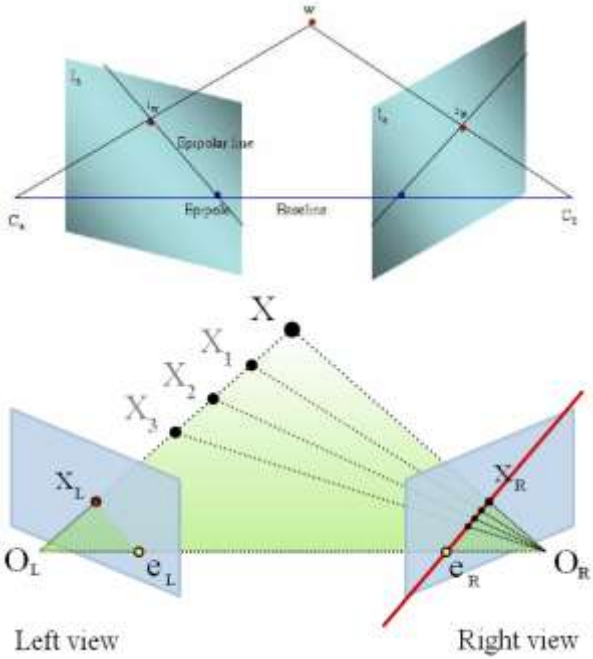
Analogies with human sight:

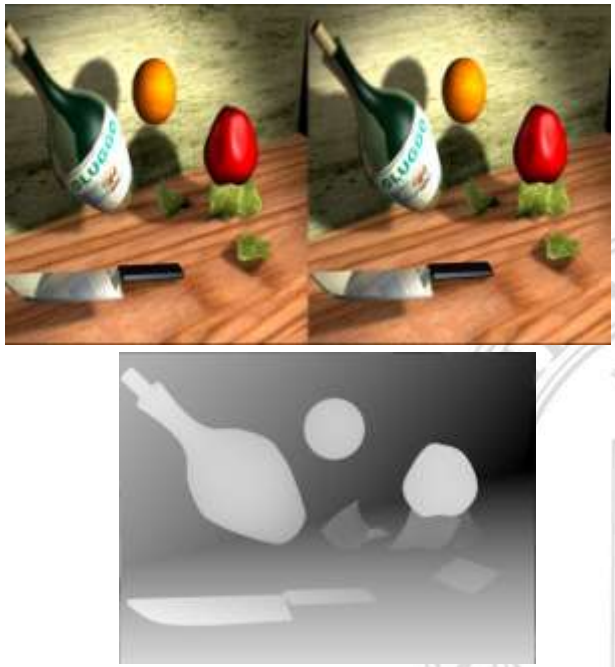
Stereopsis (binocular vision): recovering depth levels as the inverse function of relative positions of object in the views;

Triangulation;

Analogies with computer graphics:

Depth maps similar to Z-buffer techniques





3. Depth estimation (II)

PROBLEM STATEMENT

Stereo matching algorithm:

Given (at least) a pair of images representing the same scene taken from different point of views, describe how far they are with respect to the observer

3D USEFULNESS

Accurate 3D reconstruction required for:

Correct object insertion (in terms of visual feedback and perspective coherence)

Correct object interaction (in terms of handling occlusions and possible geometrical transformations)

ISSUES

Stereo matching is difficult because of:

Illumination conditions and light effects (e.g., darkness; reflections, reverberations, haloes), especially for outdoor scenes;

Lack of textures;

Occlusions across views;

Noise (e.g., blur and lack of contrast);

Small disparity range vs. depth of view (further objects tend to fail in matching correctly)

MATCHING FEATURES

Single points (e.g. corners):

Lines;

Segments;

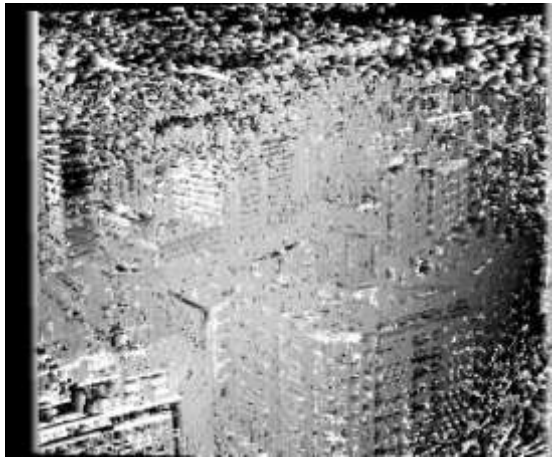
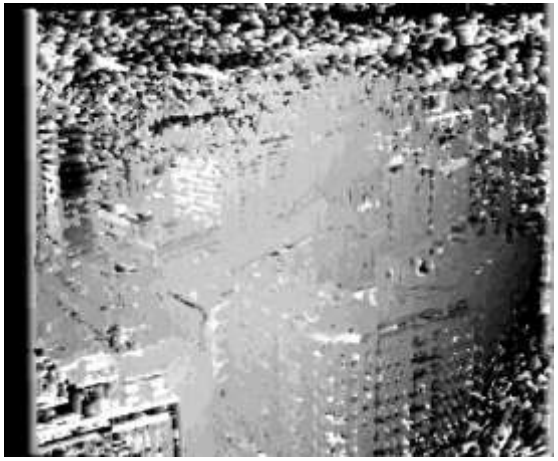
Color vs. shapes descriptors;

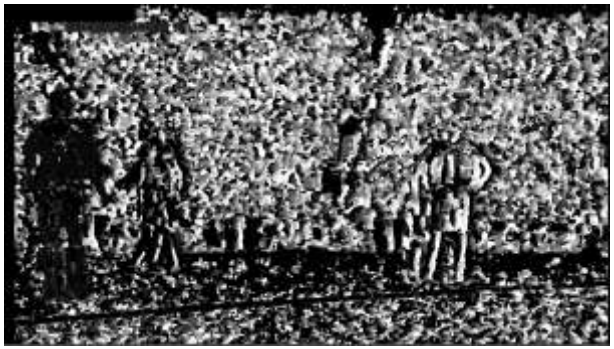
3. Depth estimation (III)

PROBLEM STATEMENT

Stereo matching algorithm:

Given (at least) a pair of images representing the same scene taken from different point of views, describe how far they are with respect to the observer




3. Depth estimation (IV)

PROBLEM STATEMENT

Stereo matching algorithm:

Given (at least) a pair of images representing the same scene taken from different point of views, describe how far they are with respect to the observer

DEPTH MAPS LIMITS

Depth maps suffers of:

- noise;
- occlusions;
- incomplete 3D information (objects are not fully determined because of the point of view)

2D + ½ & 3D

Improving 3D reconstruction:

Integration with external information sources:

- City landscapes → building models, street maps, complex topology;
- Sport / cultural events → sport / cultural facilities
- …

Nowadays, such kind of models are easy to be retrieved / imported / modified.

Modeling (parts of a) scene can be done once for all and independently by how the video content is shown into the video.

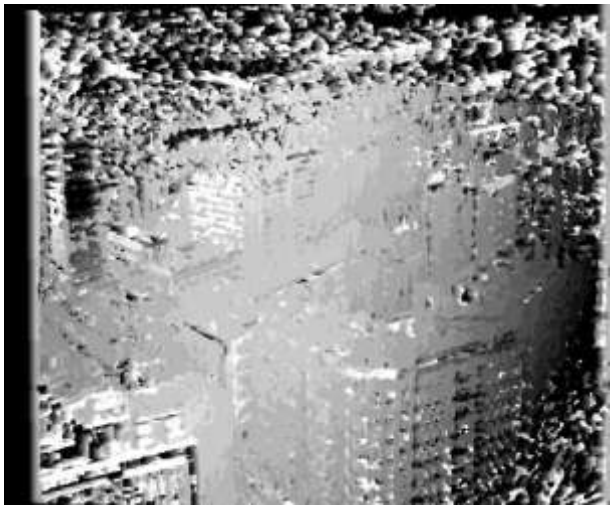
3D objects can help improving stereo matching reconstruction to avoid mismatching and inconsistencies, filling holes and unmatched areas, define objects more clearly, …

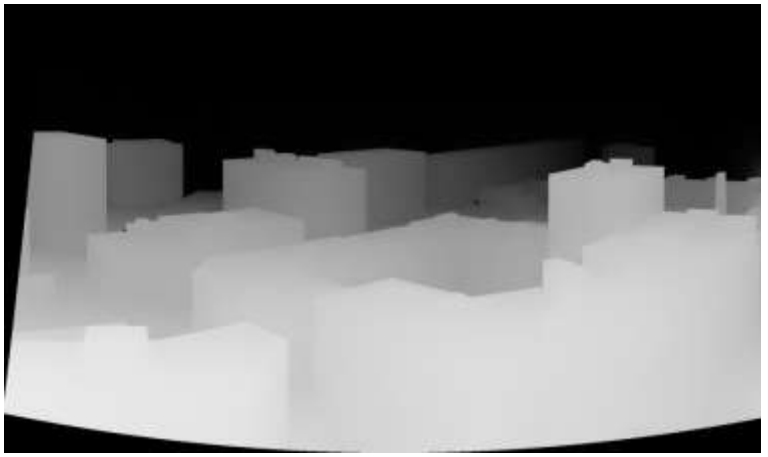
3. Depth estimation (V)

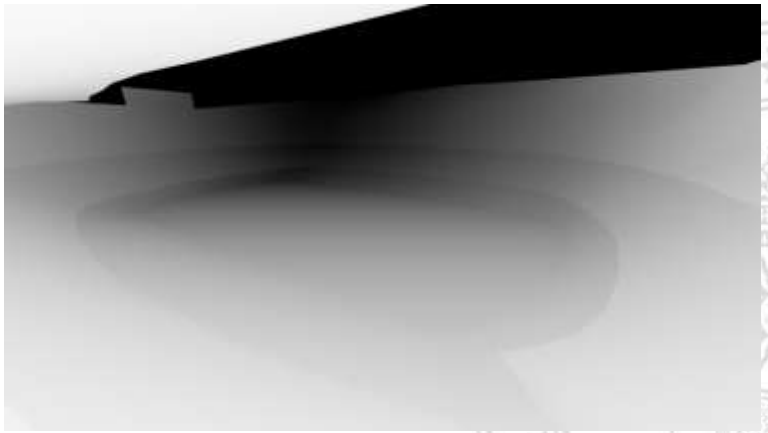
PROBLEM STATEMENT

Stereo matching algorithm:

Given (at least) a pair of images representing the same scene taken from different point of views, describe how far they are with respect to the observer







3. Object insertion

PROBLEM STATEMENT

3D models insertion while a video is reproducing:

Enriching the original video content by allowing users to add external models into the video stream. The insertion is made through suitable remote controller devices such as Flystick® and/or joysticks.

Inserted object can be moved inside the frame (e.g. scaling and roto-translation transformations) and in accordance to the depth range of the scene itself.

Occlusions and collisions should be taken into account for more realistic visual effect.

IDEAS

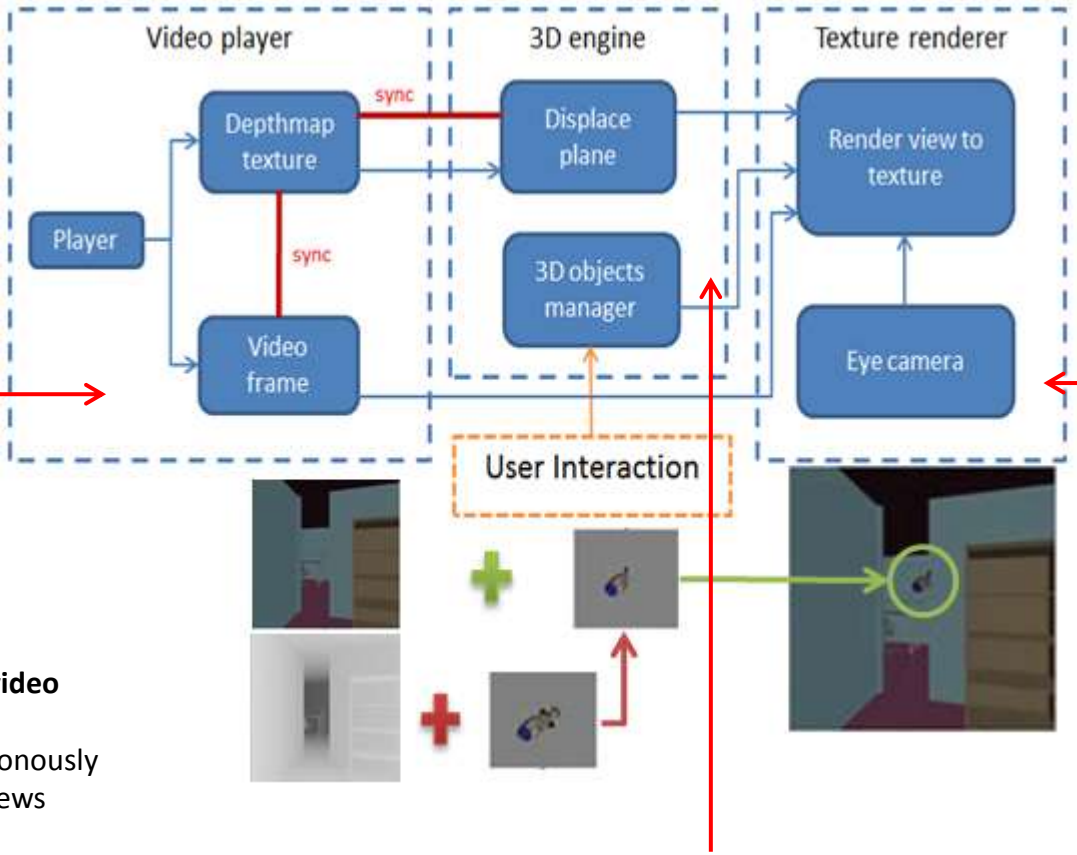
Generate a depth displacement model of the 2D scene and compute object occlusions attending to the Z-buffer information from the scene and the 3D model;

Insertion can be performed for only one video frame (e.g., the left input). The stereoscopy could be then reconstructed through the depth map previously computed;

Video synchronization between left and right eye streams to be considered carefully;

To avoid unwanted performance dropdowns, the whole process has to be completed within the time imposed by the video frame rate (e.g., 25 fps → time processing of ca. 40 ms)

3. Object insertion (II)



Inputs:
- **Stereo video;**
- **Stereo depth-map video**

For each frame, synchronously extract left and right views and depth texture

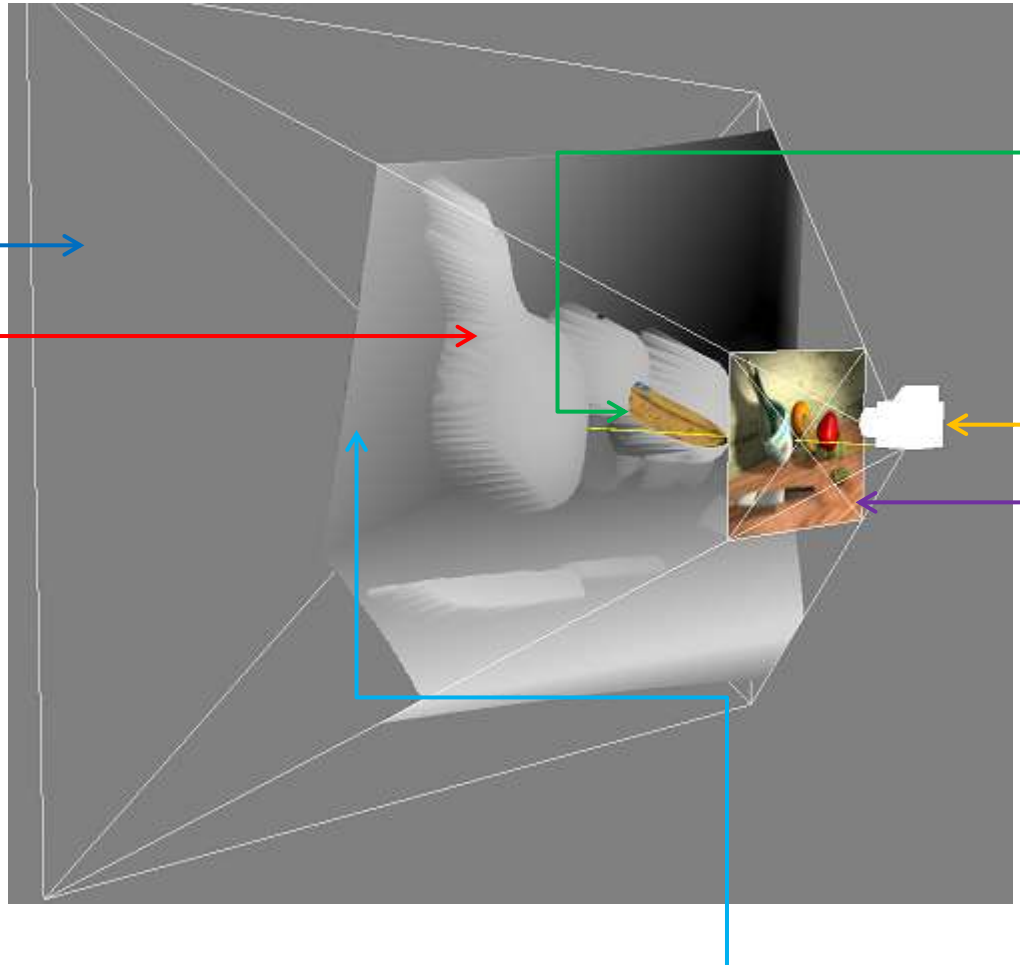
Inputs:
- **3D extra content**

It handles objects to be inserted according to user interaction. Occlusions are handled in real time.

Inputs:
- **Original frames;**
- **3D object**

A new frame, replacing the original one, is generated by merging original and new content. A virtual eye camera (same extrinsic parameters as the original one) takes a picture of the new scene and uses it as the new frame to be displayed.

3. Object insertion (III)



The 3D model insertion will not affect the original depth-map but it is used to compute occlusions and required transformation for its proper representation in the stream

Displace plane: set at the furthest distance computed by the depth-map. The scene is reconstructed by moving its vertices towards the camera according to the value of its corresponding pixel position in the depth map texture. The obtained mesh is used to compute occlusions.

3. Content adaptation and visualization

3D SPRITE

- Once the new frame have been produced, it has to be visualized into the I-Space environment;
- Stereoscopic visualization → 3D sprite having as a stereo material the corresponding frames pair;
- Synchronization with stereo glasses allows the texture to be correctly displayed according to each eye perspective;

3D OBJECTS RENDERING

- 3D models are rendered in the virtual space between the video 3D sprites and the user;
- No need to compute objects occlusions;
- Object visualization allows multiple user point of views (of course, only for objects, not for video content);
- Explicitly exploitation of CAVE capabilities;

SYNCHRONIZATION

- 3D sprite for each view → multiple sprites to be rendered for panoramic views
- Each screen has a different rendering and object insertion process
- Trade-off between CPU load (occlusions and frame creation) number of sprites into the frustum at any given time;
- It could affect the normal flow of video streams, turning into criticality in user 3D experience (e.g., flickering, loosing 3D perception, …)
- Maximum difference in time (computed on video stream) before experiencing asynchrony depends on video and user sensibility
- As the threshold is exceeded, the reproduction speed of the corresponding stream is reduced by 10% until it becomes synchronized again

3. Content adaptation and visualization

I2Tech- CAMPUS MONTEGANCEDO
Universidad Politécnica de Madrid

CeDInt POLITÉCNICA

3D SPRITE

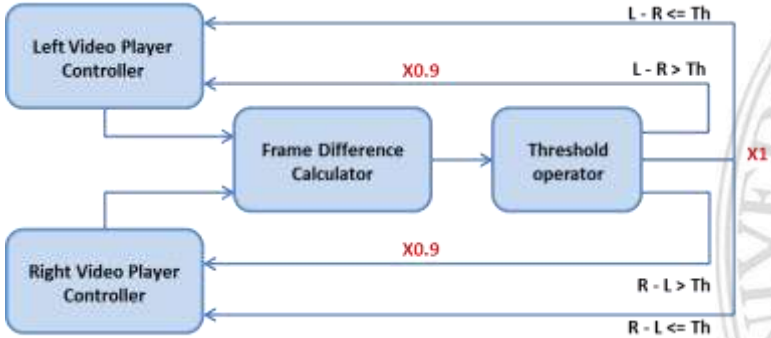
- Once the new frame have been produced, it has to be visualized into the I-Space environment;
- Stereoscopic visualization → 3D sprite having as a stereo material the corresponding frames pair;
- Synchronization with stereo glasses allows the texture to be correctly displayed according to each eye perspective;

3D OBJECTS RENDERING

- 3D models are rendered in the virtual space between the video 3D sprites and the user;
- No need to compute objects occlusions;
- Object visualization allows multiple user point of views (of course, only for objects, not for video content);
- Explicitly exploitation of CAVE capabilities;

SYNCHRONIZATION

- Maximum difference in time (computed on video stream) before experiencing asynchrony depends on video and user sensibility
- As the threshold is exceeded, the reproduction speed of the corresponding stream is reduced by 10% until it becomes synchronized again



4. CeDInt VR Lab: Infrastructure

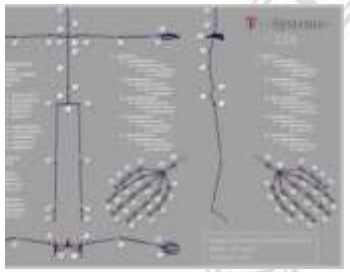
- Panoramic video / multiview (5)
- Stereo and HD
- Interactive content:
  - Pre-existing video with a combined field of horizon ≥ 180º
  - Synthetic objects (e.g. 3d models, animations, …)

Content repository

Frontal projector and screen

Processing, rendering and immersive visualization

Lateral projector and screen



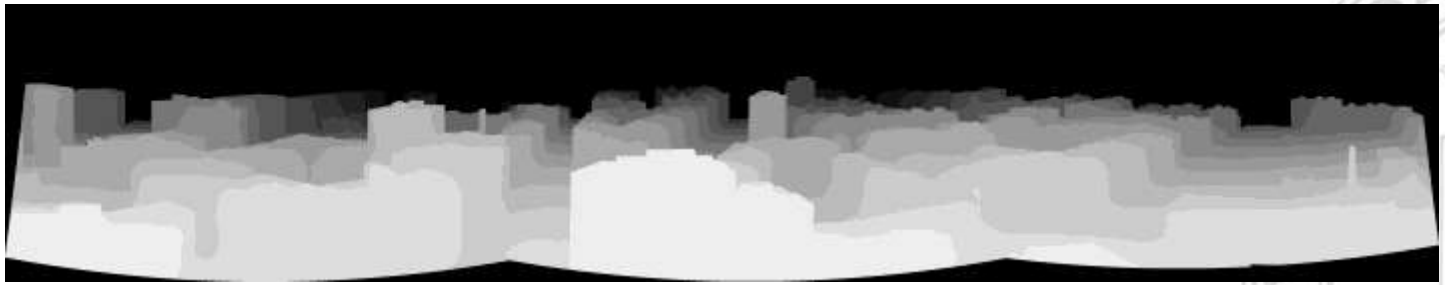
Motion capture: tracking and natural interaction

5. Results – Barcelona panoramic video (I)

5. Results – Barcelona panoramic video (II)




Stitching and depth-map (ground-truth) generation

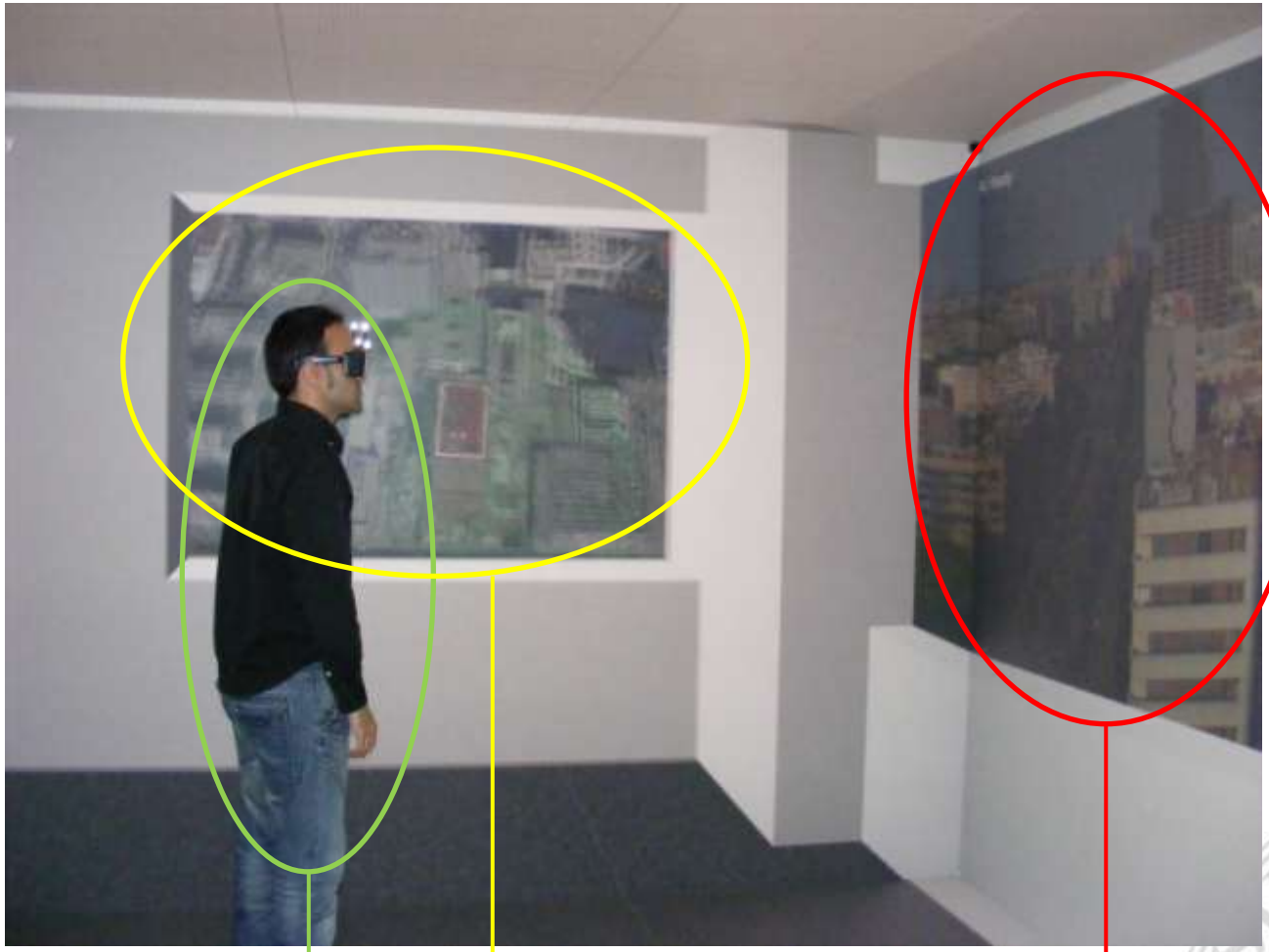



5. Results – Barcelona panoramic video (III)

Inserting a building model into the video.

Highlighting a specific model from the video content (*left*); Reproducing the stereo stream into an I-Space *(right)*

5. Results – Barcelona panoramic video (IV)



User tracking;

Synthetic environment: interior of a room; windows with a sight on the city;

Stereo video reproduction;

Bird-sight view of the city

5. Results – Barcelona panoramic video (VI)

Synthetic environment: interior of a room; windows with a sight on the city;

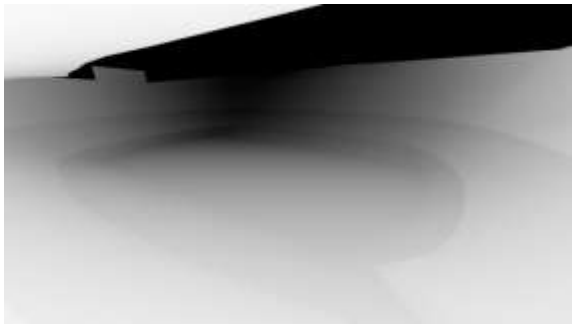
Stereo video reproduction;

Synthetic view of a city;

5. Results – Football match video

ORIGINAL FRAMES + DEPTH MAP

3D BUILDING (DEPTH MAP + MODEL)



3D SAMPLES

6. Conclusions

FUTURE APPLICATIONS

New opportunities in entertainment industry (in terms of services provided and exploitation of presented content);

New opportunities for advertising and marketing strategies;

New opportunities for educational tools and gaming;

CONCLUSIONS

ImmersiveTV prototype:

- Multimedia production content mixing stereoscopic videos and 3D object insertion (augmented-reality-like approach);
- Specifically tailored for highly immersive environment;
- Stress on user-centered applications exploiting interactivity;
- Stress on using real rather than synthetic video sequences;
- Preliminary results presented;

FUTURE WORKS

Computer vision:

- GPU implementation;
- Enhancing the creation of more reliable depth maps;

Interactive multimedia:

- Enhanced interaction modes and tools (e.g. gesture/motion recognition – based);
- Application over mobile platforms;

Evaluation of quality of experience (QoE):

- Evaluation of the user experience according to subjective and objective assessment