Design of a new visual data representation for interactive 3D systems

Pere Farrero i Roger

Project Proposal and Work Plan

REVISION HISTORY AND APPROVAL RECORD

Revision	Date	Purpose
0	19/06/2013	Document creation
1	19/06/2013	Document revision

DOCUMENT DISTRIBUTION LIST

Name	E-mail
Pere Farrero i Roger	perefarrero@gmail.com
Pascal Frossard (EPFL)	pascal.frossard@epfl.ch
Ferran Marquès Acosta (UPC)	ferran.marques@upc.edu
Thomas Maugey (EPFL)	thomas.maugey@upc.edu

WRITTEN BY:		REVIEWED AND APPROVED BY:	
Date	19/06/2013	Date	19/06/2013
Name	Pere Farrero i Roger	Name	Thomas Maugey
Position	Project Author	Position	Project Supervisor

0. CONTENTS

In this document it is possible to see all the formulation of the project as well as his fundamental steps of implementation and study.

There is, also, the formulation of the problem with a clear scheme of what the project it is going to be.

1. PROJECT OVERVIEW AND GOALS

The project is carried out at Laboratoire de Tratament du Signaux 4 (LTS4) of École Polytechique Fédérale de Lausanne leadered by professor Pascal Frossard and supervised also by Thomas Maugey.

The purpose of this project is to design a new system of representation for 3D multiview data. Research in 3D video compression has to deal with very different aspects: scene capture, data representation, compression, transmission, rendering and display. All these issues are very different but cannot be handled independently. Therefore, one of the major questions in the design of 3D video transmission system is related to the data representation, which has to be flexible enough for effective transmission and interactive rendering of 3D information. Nowadays several types of representation exist in the literature. Whereas they are very effective in specific scenario they remain too limited and non adaptable when the transmission conditions become different. In this project, we will study a new representation for 3D video transmission and validate its use in simple specific scenarios of interactive multi-view video streaming.

The project main goals are:

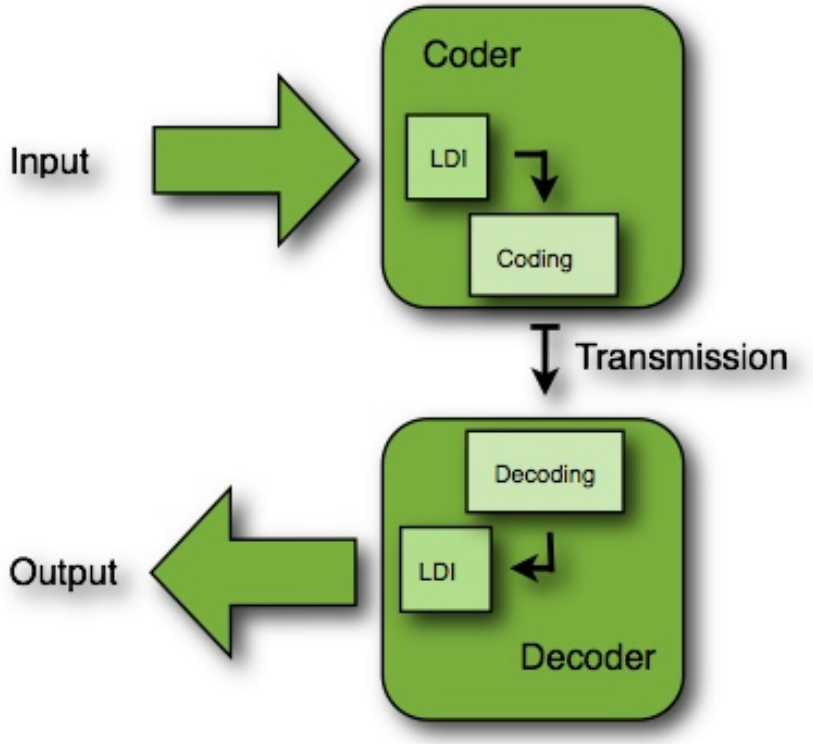
1.- Reduce the sent amount of information required to sent.

2.- To build an adaptive system to multi-resolution viewing problems due to distances and positions of the user.

The idea of the project is to design a representation system for multi-view 3D data improving some of the current systems' features. For that, we have set some targets. One is to code the minimum of data possible in order to compress the amount of information to transmit. Other target is to allow a multi-resolution system taking into account different Regions of Interest (ROI). We do not want a static-resolution system, but a variable one. There are some parts of the image that have more important details than some others and, for that, it is important the variability of the resolution for a specific place. The idea is that if we are looking inside a museum, for example, it is interesting to add more resolution to the part of the picture that contains a painting that to a flat white wall. It is for that reason that the system proposed includes a layered resolution stream with which it can be improved the resolution in relation of a region of interest. That resolution variability has also to allow not only movements in one or the other sides, but movements backward and forward as well.

The main purpose of the project is the following. We start start having a 3D Mesh and the position of the cameras that we will call C = {c1, ... , ck} being k the total number of cameras. The

main target is to get a rendering of that scene (which actually is the same set of images) again with the minimum distortion possible and, also, transmitting the lowest bitrate possible.



Being xi the image that the particular i camera gets (and are the same that the projection of the 3D mesh on the camera position ci),

$$X = [x1, \dots , xn]$$

and ^xi the rendered image (the output of the system) for that position,

$$\hat{X} = [\hat{x}1, \dots , \hat{x}n]$$

the main purpose of the project is to minimize the Distortion between those two sets of images

$$\text{argmin}\{D(X, \hat{X})\}$$

being D a criteria of distortion and tanking in account the bitrate to transmit.

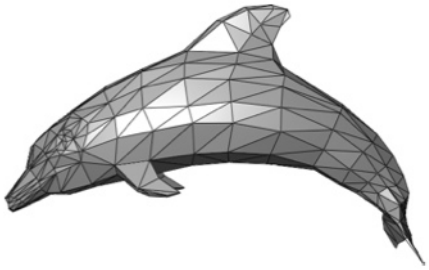
The minimum distortion between the input and output pictures will be when it is represented the same image in the rendering output that in the input. However, the bitrate is not going to be high. For that, there will be a trade off with the bitrate produced that will be mandatory to take in account.

So, the project consists in an input of a 3D mesh from which has to be builded the representation of the 3D scene. After that, when the 3D image has been transformed in a set of layers that represent one image with all the depths, it has to be coded in some way for getting a good representation bitstream. Then, it has to be defined the way of sending the packets of the coded image in order to represent all the image with the definitions improvements.

Finally, it is needed the inverse system that decodes the image and represents right the navigation over the scene.

From 3D Mesh to LDI

The starting point is a 3D Mesh, what could be understood as an extension of 2D images onto three-dimensions by adding a depth dimension. Usually, those three dimensions are represented by a set of polygons with different shapes that models the figure. From that representation system, we want to reach Layered Depth Image representation. It is very interesting because it codes only the part of the scene that is useful for the current Point of View. If there is an object that we can not see from the current viewer position or near it, there is no need to code it and to do not benefit all the available resources.



That representation change is done by analyzing the mesh from a main perspective and then, with a defined viewer segment (which represents all the available points of view for the user), all the possible viewable hidden objects are computed and added to the back layers.

From LDI to the multi-resolution friendly bitstream

To code the bitstream from LDI, we study the Wavelets techniques and the Discrete Wavelet Transform. This system is used in the standard JPEG-2000 and gives known good results.

It codes the input in a set of packets that allows different resolution scalability. That is done by packing first the basic layers and then enhancements of that basic image by the other resolution layers.

For providing multi-resolution to different segments of the image independently, JPEG-2000 uses Embedded Block Coding with Optimal Truncation (EBCOT). That is based on coding the image in independent NxN blocks in order to be able to give different resolutions to any of this blocks if needed. The current project will be based on that in order to allow the variable resolution system.

As the system doesn't know which will be the user's point-of-view, the ROI has to be defined for all the layers. In order to define well the resolution for each part, it is important to take a look on the depth of each point. That is because of perspectives and movement. For example, if the user move a step forward to the scene, two objects in different distances will increase their size for the user, but not in the same measure. The increasing of the pixels' size of the nearer object will be higher that the size of the backwards' object pixel. Then, we will need more resolution scalability in the nearer objects than in the more distant ones.

Going back: From the bitstream to an image

At the moment we have all the data coded and transmitted, it is needed to undo all the process and to reach again an image. For that, we have to anti-transformate the DWT blocks. The transmission system will code only those required blocks of enhancement layers that that are included on the region of interest. Hence, it is needed to have a good control of where do each block belong.

It is important to remind the problem that not-knowing the Point of View, the ROI could be in all the layers, but then not be shown due to the position of the user.

Rendering: From LDI to the final image

It's time to know the position of the user in order to compute the final image that it is going to be shown in screen.We have the complete image with all the resolution coefficients. What is needed to do now is to compute what appears in the image and what not and, then, to project it. All the projections are computed for each pixel and, then, the image is redistributed. The holes are filled with the projection of the back layers, also, iteratively, until having the image complete.

Of course, all of that is done for multi-resolution images. In the end, it is the same but projecting a different amount of pixels per block, for example. Once it is done, the image is recovered and able to be shown.

2. PROJECT BACKGROUND

The origin of the project is a previous project that a student made as his master thesis. This project was a Layered Depth Image coder. This images had the need of being compressed, hence this project was born.

The system's implementation requires, then, several different frameworks. The first kind are the image processing libraries and frameworks such as OpenCV, OpenGL, OpenJPEG... And the other main framework needed is the previous LDI coder, which requires to be modified in order to be adapted to the current system.

The project was initially a very open topic what gave lot of different possibilities for working. However, with some meetings with the supervisors, the project context was narrowed and finally well defined.

3. PROJECT REQUIREMENTS AND SPECIFICATIONS

Project requirements:

- Knowledge of image and video compression
- Multi-view video Representation
- Matlab (for auxiliar tools) and C++ (for the main implementation)

4. WORK PLAN

4.1. Work Breakdown Structure

The main structure of the project meanwhile its execution is: State of the art, background documentation and design, coding and implementation and writing and documenting the project.

4.2. Work Packages, Tasks and Milestones

Project: 3D Reperesentation	WP ref: 1	
Major constituent: State of the art and design		
Short description: To describe how the state of the art of the problem is as well as investigate other background informations related to the problem. Then, using all of this information, to design the system.	Planned start date: 01/03/2013 Planned end date: 01/04/2013	
	Start event: 01/03/2013 End event: 01/04/2013	

Project: 3D Reperesentation	WP ref: 2	
Major constituent: Implementation		
Short description: To write all the needed code in order to execute the described and designed previously.	Planned start date: 01/04/2013 Planned end date: 20/06/2013	
	Start event: 01/04/2013 End event: 20/06/2013	

Project: 3D Reperesentation	WP ref: 3	
Major constituent: Thesis writing and public defense		
Short description: To move from the author's head and implementation procedure all the knowledge into a paper written thesis. Then, to present an oral defense of this project in front of the tribunal who is going to evaluate it and interested people.	Planned start date: 20/06/2013 Planned end date: 01/07/2013	
	Start event: 20/06/2013 End event: 01/07/2013	

5. GENERIC SKILLS

#	Generic Skill	Assessed
1	Innovation and entrepreneurship	X
2	Societal and environmental context	
3	Communication in a foreign language	X
4	Oral and written communication	X
5	Teamwork	X
6	Survey of information resources	X
7	Autonomous learning	X
8	Ability to identify, formulate and solve engineering problems	X
9	Ability to Conceive, Design, Implement and Operate complex systems in the ICT context	X
10	Experimental behaviour and ability to manage instruments	X