



OSIRIS

Original **S**ystems for Image **R**endition via Innovative **S**creens

Funded by the European IST Program (FP6)

<http://www.osiris-project.eu/>

3D Multi View capture and rendering of natural scenes

White Paper

THOMSON HOLOGRAFIKA

Authors : Valter DRAZIC, Zoltan Megyesi

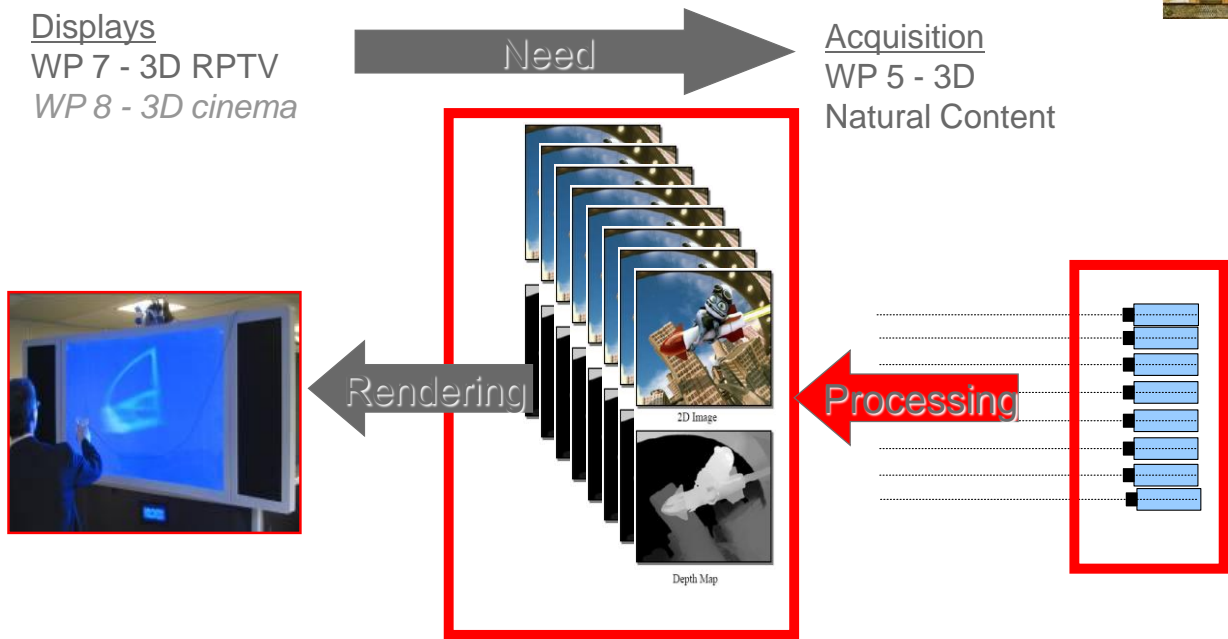
Date: 2009 - June 23rd

Ref Plenoptic 1 vs 3

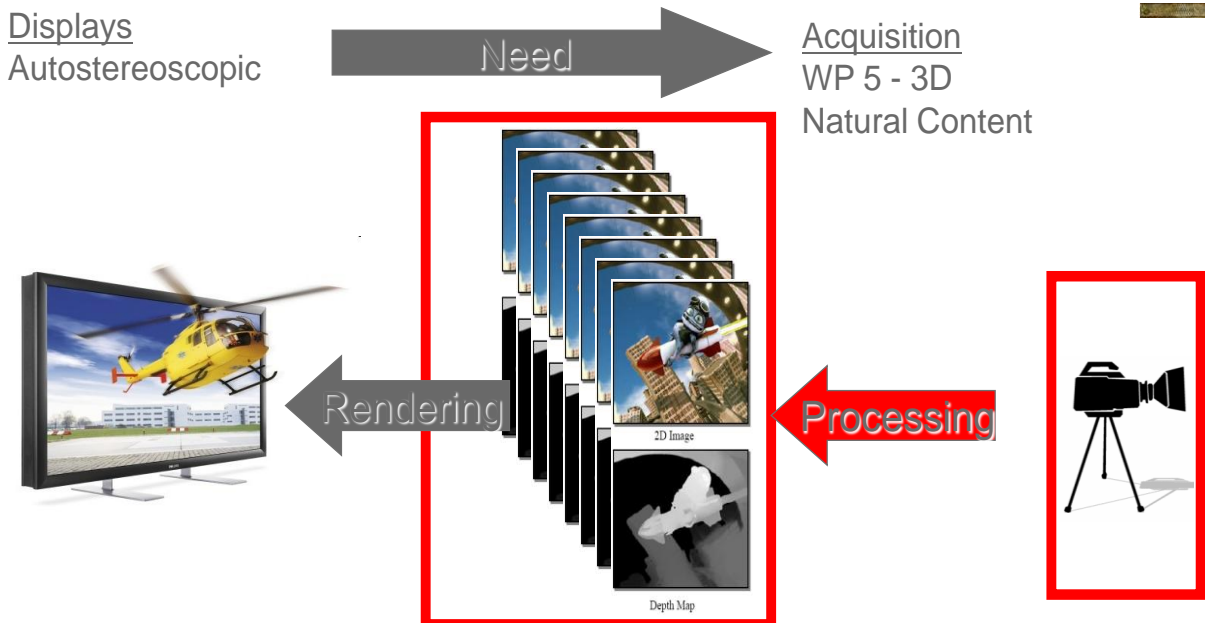
Revision Approved/1.0

1. Introduction

The goal of the work package “Natural 3D content” is to develop a capturing mean for providing video images of the real world to display systems developed within WP7 and WP8 which are very wide field angle capable displays.



It is also aimed to develop capturing means for different displays, like auto-stereoscopic displays which are narrow field angle display systems.



Main idea: simplify the complexity and amount of image processing for the depth map

The first schematic is bound to a linear array of cameras, less than 10 however, distributed on a large baseline. On the other hand, in the second approach, we would like to develop a single lens camera capable of capturing multiple views.

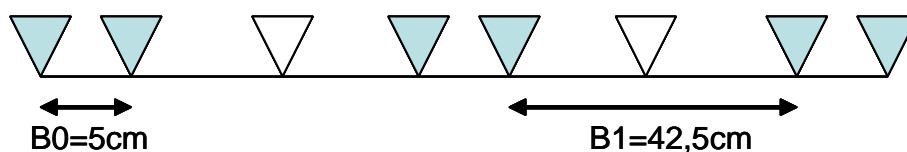
2. Acquisition

2.1 Multiple camera acquisition

The system consists in a camera rig of 8 or 9 cameras. As the scene that needs to be caught must be a very wide angle one (70 degrees), shot not too far away, that means that the total dimension of the rig will be somewhere between 1 and 2 meters. The following drawing shows a rig of 1 meter. If the cameras are equally distributed, the distance from one to the next results in very big disparities what make the calculation of the depth maps inaccurate. We will get a better quality for the depth map if the pair of cameras that participate in the depth estimation are close to each other (blue triangles). The rendering used to display all the signals makes it necessary to have a video in between (white triangles).

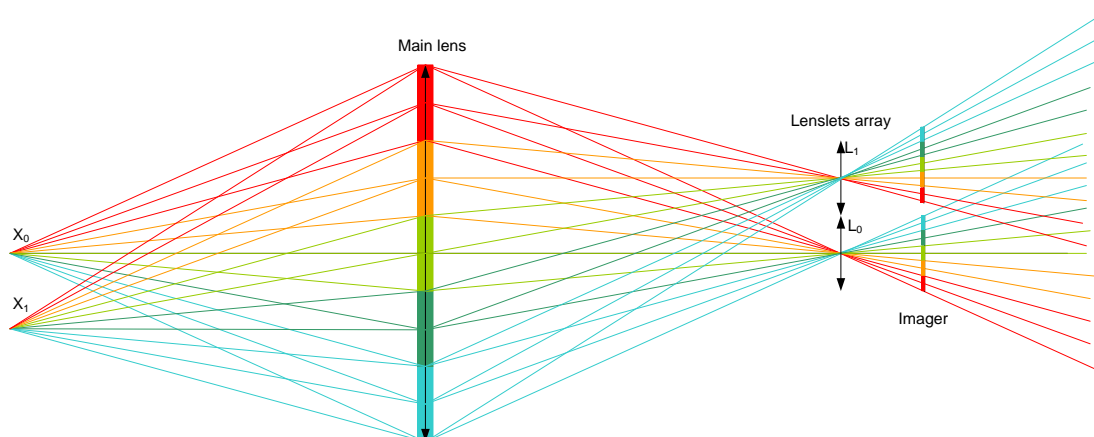
The main challenge here is the precise positioning of all the cameras, the calibration, rectification and above all the calculations of the depth maps.

The schematics here may not be the final one, discussions are ongoing to define the best configuration. The blue triangles denote the *depth estimator* cameras, the white triangles denote the *image provider* cameras.



Still pictures are taken according to different configurations and a final one shall be approved by WP7 & 8.

2.2 Single lens multi-view acquisition



The principle used for acquiring multiple views with a single sensor is the plenoptic recording. Compared to a classical camera system, there is a lenslet array in front of the imager. The system is depicted in the previous figure where different colors are used to emphasize the new art of signal acquisition in a plenoptic camera.

In the example above, there are 5 pixels for each lenslet.

Each group of 5 pixels sees one specific location in the object space. The group of pixels from the lenslet L_1 sees the position X_1 whereas the group L_0 sees the position X_0 .

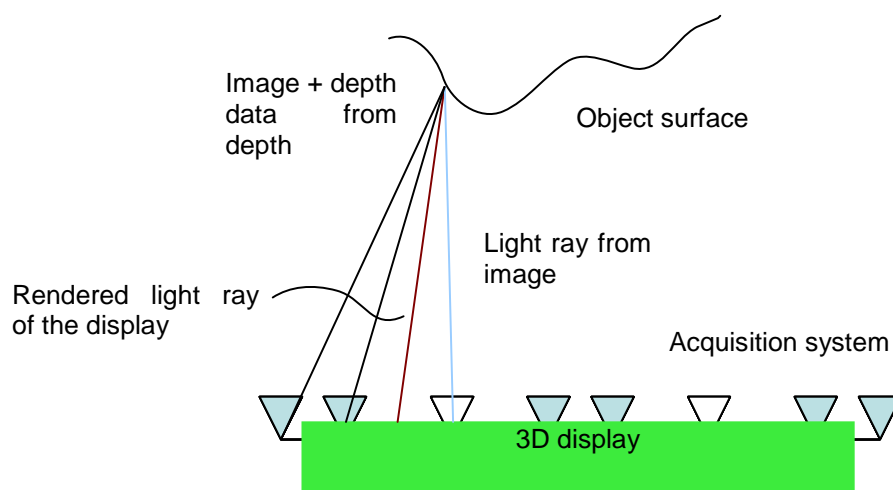
Within a group of 5 pixel, each one of them sees a different patch of rays through a small aperture on the pupil of the main lens \Rightarrow Within a group of 5 pixels, each one of them sees a specific direction.

The plenoptic camera stores the rays coming out of the scene in a 2D array of 2D arrays. The 2D array of the lenslets registers the position of the objects in the scene, and the 2D array of the pixels below each lenslets stores the parallax under which a point from the scene is seen from different viewpoints of the main lens pupil.

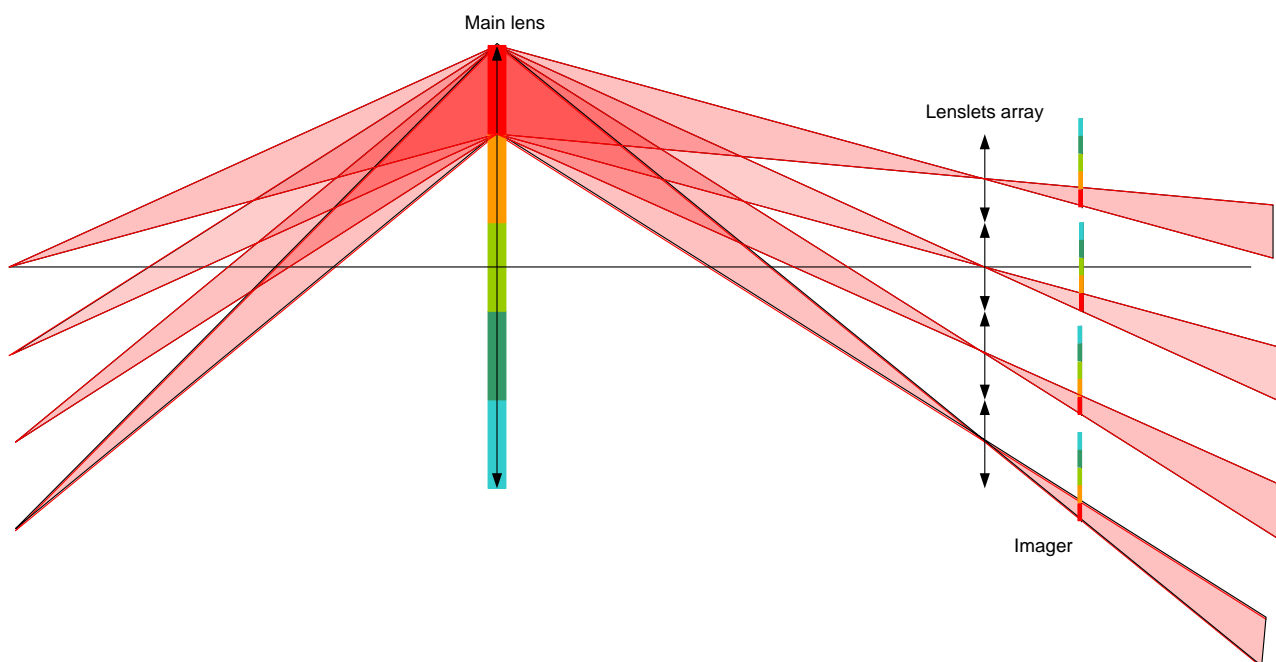
3. View rendering

3.1 Rendering of the multi-camera content

The purpose of the acquisition system is to provide input a for the continuous parallax 3D displays developed in WP7 & 8. Therefore the images captured by the cameras need to be converted into a dense light field defined by the displays. This task is performed on the side of the display using Image Based Rendering techniques enhanced by the depth data obtained from the short baseline camera pairs (blue triangles). Ideally, the rendering of the dense light field would require multiple camera images, however, our purpose is to create a novel flexible and compact acquisition system. To address this problem, we use all image information from the acquisition system (including the images from both depth estimator and image cameras) and the missing information will be replaced using the depth maps of the depth estimator cameras. The depth maps are also necessary to speed up the rendering process and facilitate real-time rendering.

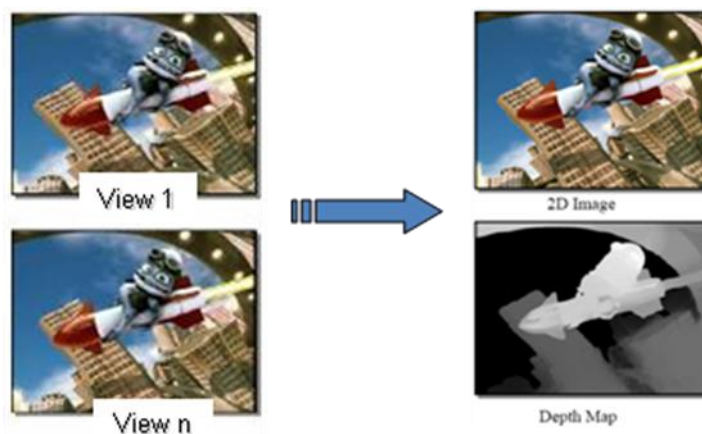


3.2 Rendering of the single lens camera content



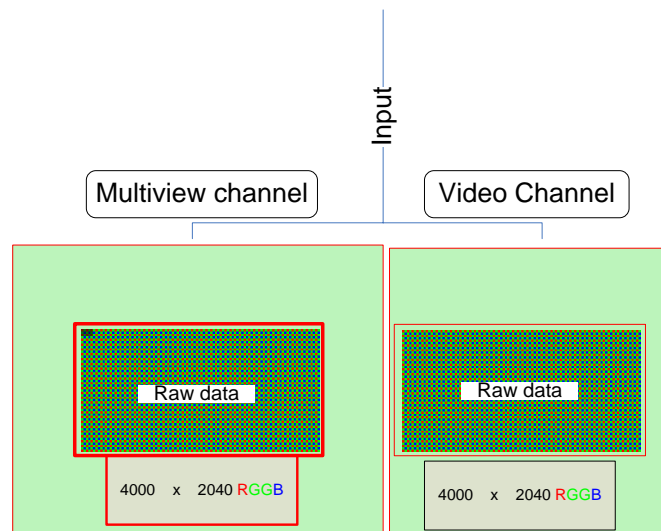
The rendering of the multiple views is just a spatial de-multiplexing process. In the figure just above, it can be seen that if the red pixels are put side by side, one can reconstruct the object as seen through the red aperture of the main lens. The same can be done for the other pixels and we'll end up with five views of the main field as seen from five different parallaxes.

The plenoptic camera can be used in two ways. We can either feed the different views directly to an auto-stereoscopic display, or we can transmit the central view and a depth map calculated thanks to the n views.

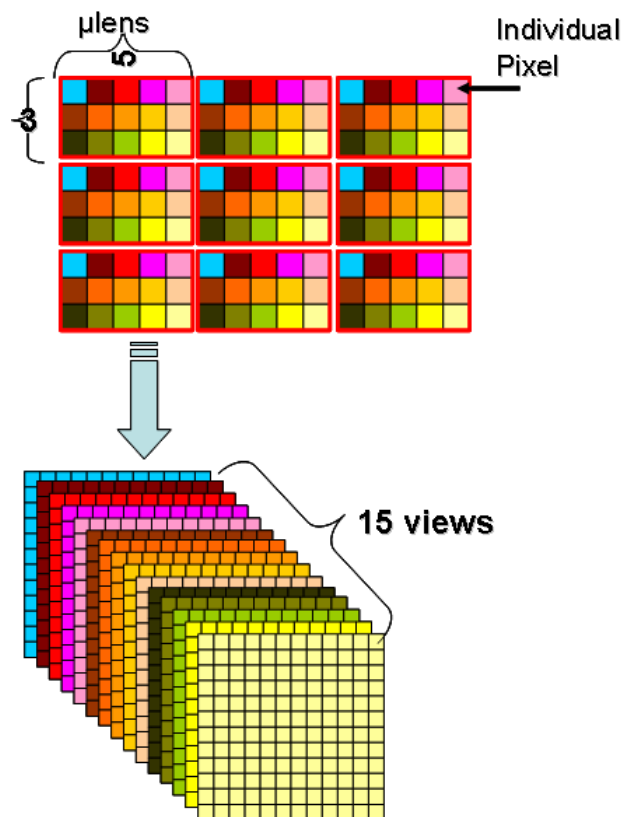


But going back to the very general and high level view of demultiplexing with a sensor equipped by a color filter array (CFA), known as a Bayer pattern, we can summarize the flow with some synthetical charts.

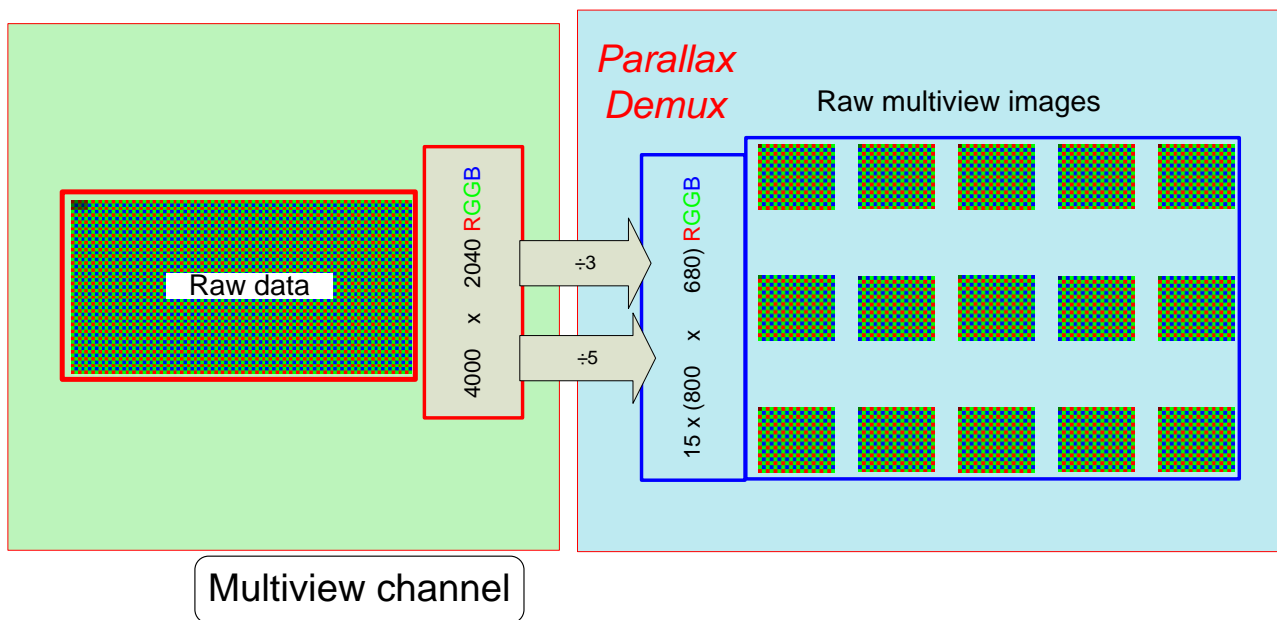
Let us assume for explanation purpose that we have a sensor with 4000×2040 pixels with a CFA. Let us also assume that we have a 2D lenslet array, each lenslet having the dimension of 5 horizontal pixels and 3 vertical ones.



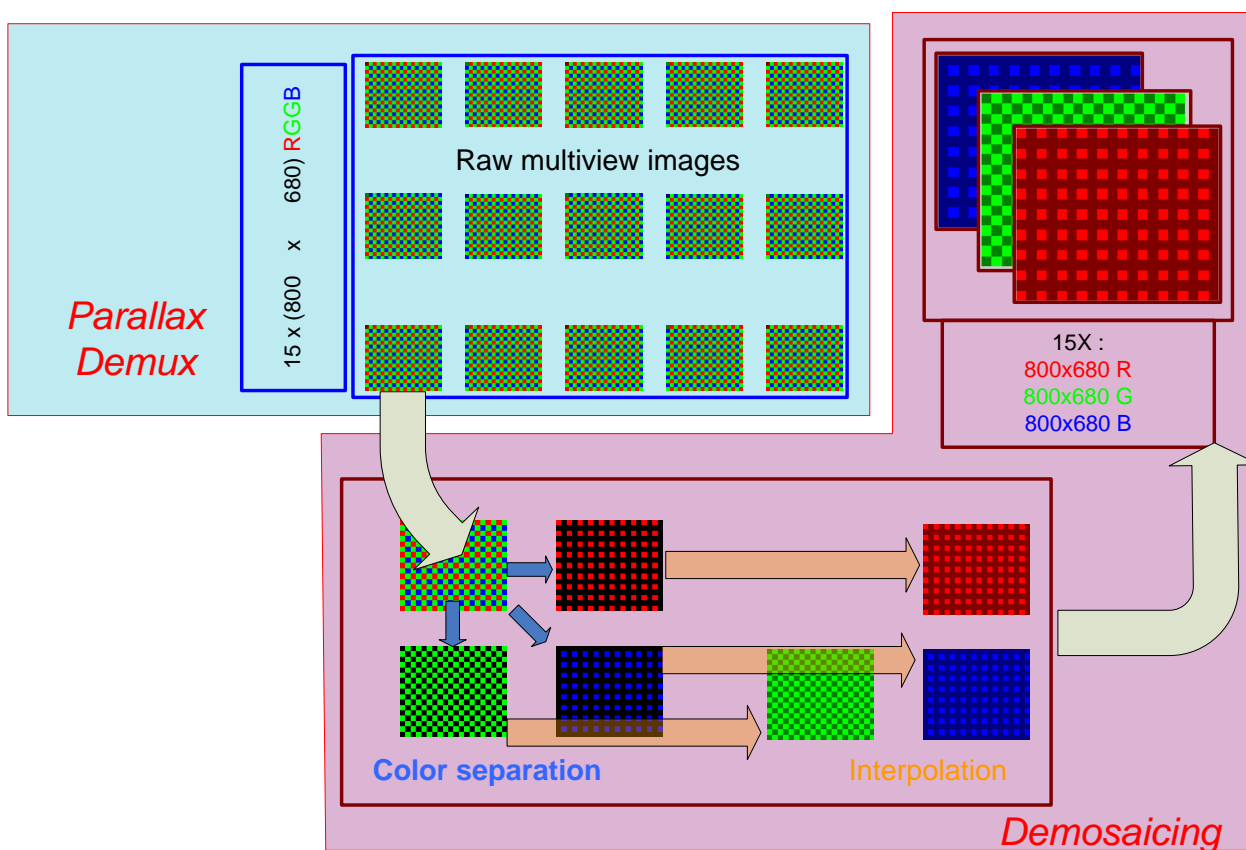
We first de-multiplex the 15 different views by grouping pixels together.



As the sensor is provided with a CFA, the 15 different views will be RAW data (not yet demosaiced), according to:



At this stage we have 15 RAW views, each one having a resolution of 800x680 pixels. The next step is the demosaicing, which consists in the separation of each view in 3 color planes (true color file), and also the interpolation of the missing pixels.



4. Display of the content

4.1 Continuous parallax 3D display

Parallax in 3D displaying is used to indicate the different viewer positions to which the display is projecting images containing displacement and distortion that apparently belongs to a valid 3D scene. As an example, simple systems with stereo glasses provide images with binocular parallax. A continuous parallax display provides valid 3D views in the entire viewing area, regardless of position and eye distance. An example of continuous parallax 3D display is the Hologivio™ display system.

The motivation behind these displays is holographic technology. Holograms can reconstruct the complete wavefront of the light field by storing and reproducing not only the intensity but also the phase of the lightwaves. The interference image created by the holographic photography is reproduced by applying a reference wave (a light source) to the photographic plate. The reference wave is diffracted through the plate, thus a virtual image is created for the viewer. Moving holograms can also be created by using optically addressed spatial light modulators, however usable sizes, and viewing angles are far from practical applicability.

In the Hologivio™ display technology each point (voxel) of the holographic screen emits light beams of different colour and intensity to the various directions (exactly how a point of a window does), in a controlled manner. The light beams are generated through a specially arranged light modulation system and the holographic screen makes the necessary optical transformation to compose these beams into a perfect 3D view. The light beams cross each other in front of the screen or they propagate as if they were emitted from a common point behind the screen. With a proper software control of the light beams viewer or viewers see objects behind the screen or floating in the air in front of the screen. These systems can be upgraded to large scale (wall-size holographic screens), resolution, brightness, etc. is not limited by principle.

The main advantage of this approach is that, similarly to the pure holographic displays, it is able to provide all the depth cues and it is truly multi-user within a reasonably large field of view. This is a high-end solution compared to other technologies and fulfills all the requirements of real 3D displaying simultaneously. It creates all light beams that are present in a natural 3D view, that is the reason why one sees the same as in reality. This is qualitatively very different from other contemporary multi view technology that forces users into approximately fixed positions, because of the abrupt view-image changes that appear at the crossing of discrete viewing zones. By contrast, the holographic displays provides continuous horizontal parallax with approximately 0.8° angular resolution for a full ~50° field of view, free of any transitions between view images. Reconstruction of the perfect 3D view makes this technology a favorable choice for 3D displays.

4.2 Multi-view display

The single lens multiview content acquired by the plenoptic camera is very well suited to the multiview auto-stereoscopic displays where its short baseline is naturally adapted to the small disparities involved by this display technology. The simplest and immediate way to display the acquired content is to build a camera that captures as many views as needed by the display. This makes however the camera display specific. Another approach is to use a depth map calculated from the multiview content to interpolate intermediate views and adapt their numbers to the type of auto-stereoscopic display. Discussion

5. Discussion

5.1 Wide field of view acquisition and display systems

5.1.1 Advantages

The major advantage of using wide field of view in both acquisition and displaying is the fact that more information can be provided to the viewers in the same time.

Multiple viewers watching a wide field of view display can observe different details, while the viewers can change the focus of their interests by natural movement without changing the content or the rendering viewpoint. This indeed have many advantages in situations where multiple viewers wish to share the information from the same scene, or when it is important to display information to a large angle at any given time (like in museums, advertising situations, but also in research where 3D data analysis is important).

Naturally, wide FOV displaying require wide FOV content. But this is not the only reason why wide FOV acquisition is important. Capturing more geometric information of the same scene in the same time can be an advantage regardless of the displaying technology. It allows inspecting the scene from multiple viewpoints even in the case of dynamic scenes, and it makes changing the viewpoint during displaying possible. This can be especially important in the case of natural scenes, where any moment can be unrepeatable. These acquisition systems are useful to *Interactive* or *Free Viewpoint Television* or simply when it is important to store more complete 3D data.

5.1.2 Drawbacks

Unfortunately there are several drawbacks related to wide field of view acquisition and displaying, that make this task challenging. Firstly, the multiplied size of the data induce problems in either capturing, storing, transferring or rendering the content. For consideration, the raw data for full continuous parallax displaying would require up to 50 times the information that is needed for a simple stereo displaying solution.

Fortunately, exploiting the existence of 3D information in the captured data (in the form of depth) can reduce this need to controllable quantities. The second problem is the increased importance of synchrony and calibration accuracy on both acquisition and displaying sides. Any lag or lack of precision can lead to geometric inconsistency and 3D errors that reduce overall viewer satisfaction. The large distance and the number of components to synchronize and calibrate make this task difficult. On the good side, the large amount of redundant information that is captured can be used for alignment and thus can help in solving these problems reliably.

5.2 Single lens acquisition and multi-view display systems

5.2.1 Advantages

There is just a single camera required, hence this means that no calibration is necessary, no rectifications between views which are just translated, there is no skewness. Hence the depth should be easily calculated since the correspondence problem is naturally minimized, and the image processing algorithms shall be simpler and faster. A real time depth map could be at hand. One camera is much simple to handle than multiple ones. We can get parallaxes in both directions, and it's a passive depth measuring system that does not need additional means to project interferences patterns or pulses of IR light onto the objects.

5.2.2 Drawbacks

As the baseline for estimating the depth depends on the size of the main lens (triangulation), its depth discrimination accuracy is lower than for binocular stereo systems. One goal within the work package is to study how good or how bad is the depth discrimination of a plenoptic camera.

Another drawback is that the spatial resolution of the image is divided by the total number of recorded views. (position and parallax captured on same sensor used traditionally for just position sensing), so the resolution of each view is much lower than that of a camera used in a stereo pair.

To address this issue, some new schemes have to be worked out.

6. Conclusion

The twofold axis of research in this work package are addressing very innovative and new applications. The multiple camera rig shall provide for the first time some natural content for continuous viewpoint displays like the Holovizio™. The plenoptic camera on the other hand is meant as a natural 3D content provider for auto-stereoscopic displays, but also as a very simple apparatus in its handling that could provide images and calculate depth map with on board chip.