TÉCNICO LISBOA

2D Image Rendering for 3D Holoscopic Content using Disparity-Assisted Patch Blending

João Filipe Oliveira Lino

Thesis to obtain the Master of Science Degree in

Communications Networks Engineering

Examination Committee

Chairperson: Prof. Paulo Jorge Pires Ferreira

Supervisor: Prof. Fernando Manuel Bernardo Pereira

Co-Supervisor: Prof. Paulo Jorge Lourenço Nunes

Members of the Committee: Prof. Tomás Gomes da Silva Serpa Brandão

October 2013

Acknowledgements

I would first like to thank my supervisors Prof. Fernando Pereira and Prof. Paulo Nunes for their guidance and total availability during the various stages of this work. Without their knowledge, patience and critic, this work would not be possible.

I would also like to thank:

All my ex-colleagues at the Image Group at Instituto de Telecomunicações for their support throughout the various stages of this work. A especial thanks to Caroline Conti, Matteo Naccari, Catarina Brites, João Ascenso, Tomás Brandão and Paula Queluz for their support and knowledge.

My grand-parents, Maria Carolina Oliveira and José Vicente Oliveira, for always believing in me and supporting me.

My friends, for always giving me perspective and sharing their knowledge.

And finally, all brave men and women that dared to seek for answers to questions they did not fully understand.

Abstract

Holoscopy is a 3D technology that targets solving some of the limitations of current 3D technology, such as the need to wear special glasses to get the depth perception and the visual discomfort caused inherent accommodation issues. The display technologies available today at home are not yet compatible with 3D holoscopic image formats, for this reason methods that convert 3D holoscopic formats into 2D and current generation 3D image formats are required. There is, however, no fully automated, all-in-Focus and high performance 2D extraction method available in the literature. Moreover, the development of this type of conversion technology requires appropriate testing and assessment. Besides subjective testing with real people, there is nowadays no other reliable way to assess the performance of this type of conversion technology.

In this context, and after making an extensive review of 3D holoscopic capture and display technologies, this Thesis: a) proposes a novel, fully automated, 2D image extraction method for 3D holoscopic images and assesses its performance; and b) identifies potential No-Reference Image Quality Assessment metrics able to rate this type of 2D extractions and assesses their correlation with the human perception of quality.

The proposed 2D extraction method – *Disparity-Assisted Patch Blending* – outperforms, in "normal" conditions, all the available alternative methods. The NIQE metric seems promising as a potential candidate to reliably rate 2D extractions, although more research work should be done on BRISQUE as well to make it even more reliable.

Keywords

3D Holoscopic, Disparity-Assisted Patch Blending, Micro-image patch size computation, Disparity estimation, No-Reference Image Quality Assessment

Table of Contents

List of Tables

List of Figures

Glossary

1D	One-Dimensional
2D	Two-Dimensional
3D	Three-Dimensional
3DAV	3D Audio-Visual
ACR	Absolute Category Rating
AFX	Animation Framework Extension
AoV	Angle-of-View
AQI	Anisotropic Quality Index
AVC	Advanced Video Coding
AVe	Angle-of-View extraction
BLIINDS	BLind Image Integrity Notator using DCT-Statistics
BRISQUE	Blind/Referenceless Image Spatial QUality Evaluator
CBIQ	Code Book Image Quality
CfP	Call for Proposals
CGS	Coarse-Grain Scalable
CPU	Central Processing Unit
DA	Distortion Aware
DAPBe	Disparity-Assisted Patch Blending extraction
DBe	Depth Blending extraction
DIBR	Depth-Image-Based Rendering
DIIVINE	Distortion Identification-based Image Verity and INtegrity Evaluation
DMe	Disparity Map extraction
DU	Distortion Unaware
FR	Full Reference
GLSL	OpenGL Shading Language
Gm	Global model
GPU	Graphical Processing Unit
HEVC	High Efficiency Video Coding
IEC	International Electrotechnical Commission
IQA	Image Quality Assessment
ISO	International Organisation for Standardisation
ITU-T	International Telecommunications Union - Telecommunications
JMVM	Joint Multiview Video Model
JVT	Joint Video Team
LBIQ	Learning Blind Image Quality
LGm	Local plus Global model
MAC	Multiple Auxiliary Components
MLA	Micro-Lens Array
MPEG	Moving Picture Experts Group
MSE	Mean Square Error
MVC	Multiview Video Coding
MVD	Multiview Video plus Depth

MVP	Multiview Video Profile
NCC	Normalized Cross-Correlation
NIQE	Natural Image Quality Evaluator
NR	No-Reference
NSS	Natural Scene Statistics
OA	Opinion Aware
OU	Opinion Unaware
PoV	Point-of-View
PSNR	Peak Signal to Noise Ratio
QoE	Quality of Experience
RD	Rate-Distortion
RDC	Rate-Distortion-Complexity
RMSE	Root Mean Squared Error
SNR	Signal to Noise Ratio
SSIM	Structural SIMilarity
SSPBe	Single-Sized Patch Blending extraction
SSPe	Single-Sized Patch extraction
SVC	Scalable Video Coding
TV	TeleVision
VCEG	Video Coding Experts Group
VSBe	View Selective-Blending extraction

1 Introduction

The purpose of this first chapter is to present the context and motivation of this Thesis, as well as the proposed objectives and the Thesis report structure.

1.1 Context and Motivation

Providing a more immersive multimedia experience to the home-users has been a major trend in the scientific and industrial community with increased intensity more recently. The way people communicate, collaborate, socialize and entertain has fundamentally changed in recent years because of this. For 3D video services to become practical and sustainable, adequate data formats for representing and delivering 3D video content considering different constraints are needed. Moreover, it is essential that factors minimizing the consumer quality of experience be avoided, such as viewing discomfort or fatigue and the need for wearing special gear.

Holoscopy is a technology that solves the issues of previous and current 3D technology. The display technologies that exist today in homes are not yet compatible with 3D holoscopic image formats. To solve the issue of compatibility there is the need for methods that convert from 3D holoscopic formats into 2D or current generation 3D image formats.

The currently known conversion methods are either well defined and produce 2D images with artefacts, or partially defined promising to produce high quality 2D images. There is however no fully automated, well defined, 2D extraction methods described in literature.

An issue that arises from the development of this type of conversion technology is the testing of a proposed solution. Short of submitting it to tests with real people, there is no other documented way to know if a 2D extracted image has good or bad subjective quality. Part of the reason is that the image

quality assessment metrics documented where never tested with 2D holoscopic extractions. The other part of the reason is because most documented and proven ways of assessing perceived quality in images require an original 2D image to compare the 2D extraction to. The problem is that in this conversion scenario there never is an original 2D image because the original is a 3D holoscopic image.

1.2 Objectives

Based on the context and motivations, this Thesis has the following objectives:

- Review 3D holoscopic capture and display technologies;
- Proposing a novel, fully automated, 2D image extractions method for 3D holoscopic images;
- Identify potential No-Reference Image Quality Assessment metrics able to rate 2D extractions;
- Assess the performance of the proposed 2D extraction method;
- Assess the correlation between the identified No-Reference Image Quality Assessment metrics and the human perception of quality.

1.3 Thesis Outline

In this first chapter, the main subject and objectives of this Thesis are introduced.

The second chapter covers relevant concepts and also both the capture and display evolution of 3D holoscopic technologies throughout the years. However, a special emphasis is given to the 3D holoscopic technology, which is the theme of this Thesis.

In the third chapter, the existing methods developed to convert 3D holoscopic content into 2D content - 2D extraction methods - are reviewed.

In the forth chapter, a novel fully automated, 2D image extractions method for 3D holoscopic images is proposed. This novel method is methodically described and extensively detailed.

In the fifth chapter, the test methodology to assess the quality of the 2D extracted images is covered. First the available resources to test the proposed 2D extraction method are presented. Secondly the identified no-reference image quality assessment methods are presented. Finally, the test procedures are described.

In the sixth chapter, the results of applying the test methodology are analyzed. Two analyses are preformed: one to assess the performance of the proposed 2D extraction method; the other to assess the correlation between the identified no-reference image quality assessment metrics and the human perception of quality. Conclusions are drawn in relation to what is the 2D extraction method available that produces the best perceived quality and what is the no-reference image quality assessment metric that correlates best with human perception of quality.

The seventh and final chapter of this Thesis contains the conclusion and the proposed future work the author has identified.

2
Holoscopic Imaging: Reviewing Concepts and Systems

This chapter reviews the main concepts and systems related to holoscopic imaging, which is the core technology of this Thesis. A review of the main holoscopic imaging capture systems available follows. Finally, to close the chapter, a brief review of the current display technology relevant for 3D holoscopic imaging is presented.

2.1 Basic Concepts and Definitions

In this section, the basic concepts and subjects involved in *Holoscopic Imaging* are reviewed and defined. In order of appearance, this section will cover the concepts of *Light Field*, *Radiance*, *Focal Length, Depth of Field, 4D Light Field, Stereoscopy,* and finally *Holoscopic images*.

Light Field

Michael Faraday first defined a light field, in an 1846 lecture titled “Thoughts on Ray Vibrations”, as a function describing the evolution of all the rays of light passing through every point in space, in any angular direction, for any wavelength, throughout time [1]. Describing a light ray in this manner, with full detail, requires three coordinates to describe its spatial positioning in space (*x*, *y* and *z* in Figure

2.1), two coordinates for the angular information of the rays (θ and φ in Figure 2.1), one dimension to account for wavelength and one dimension for time to account for continuous propagation. A light field function is, therefore, a 7D function.

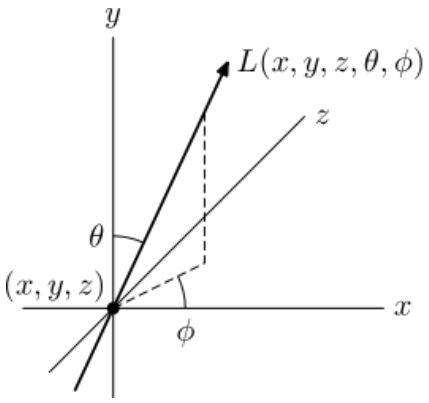


Figure 2.1 - Parameterization of a light ray in a 3D space by position (x,y,z) and direction (ɵ,ϕ).

In 1991, Adelson [2] made the time and wavelength dimensions to be constants in the light field function (static scene and a single wavelength) and called this 5D function the *plenoptic function*. In a simple manner, this function represents all the rays of light passing through every point in space, in any possible angular direction, in a single time instant and wavelength, this means a snapshot of the light field. Traditional imaging, commonly known as *photography*, attempts to capture an even simpler version of the light field function, in this case with only 2 spatial dimensions plus the wavelength dimension for colour differentiation.

Radiance

Photons are the elementary particles responsible for carrying the electromagnetic field. The phenomenon of propagation of these particles is called electromagnetic propagation made through electromagnetic waves. Visible light has a wavelength in the range of about 380 nm (nanometres) to about 740 nm, this means between the invisible ultraviolet, with shorter wavelengths, and the invisible infrared, with longer wavelengths. This type of radiation is commonly referred to as *light rays*.

The radiance represents the amount of electromagnetic energy that passes through or is emitted from a surface and falls within a given solid angle in a specific direction. Photographic cameras use two-dimensional (2D) sensors to detect light rays, measuring the amount of radiance hitting its surface, as defined by Equation (1).

$$L = \frac{d^2\Phi}{dAd\Omega\cos\theta} \approx \frac{\Phi}{A\Omega\cos\theta} \qquad (1)$$

In this equation, *L* is the measured radiance coming from direction θ, *d* stands for the differential operator, Φ is the total radiant flux or power emitted, θ is the angle between the surface normal and the specified direction, *A* is the area of the surface and *Ω* is the solid angle associated to the observation or measurement. The approximation in (1) only holds for small *A* and *Ω*, where *cos* θ is approximately constant [3].

Focal length

The power of a lens is the degree to which it converges or diverges light. The focal length of a lens is the distance from the lens to the point in space where all light rays parallel to the optical axis of the

lens hitting the lens converge to. The Lens-Maker's equation (2) relates the power of a lens with its focal length and its physical attributes [4]:

$$P = \frac{1}{f} = (n-1)\left[\frac{1}{R_1} - \frac{1}{R_2} + \frac{(n-1)d}{n\,R_1 R_2}\right] \qquad (2)$$

Where *P* stands for the power of the lens, *f* stands for the focal length of the lens, *n* is the refractive index of the lens material, R_1 is the radius of curvature of the lens surface closest to the light source, R_2 is the radius of curvature of the lens surface farthest from the light source and *d* is the thickness of the lens (corresponding to the distance along the lens axis between the two surface vertices). The diameter of a lens is also called pitch. A graphical representation of the relevant situation is shown in Figure 2.2.

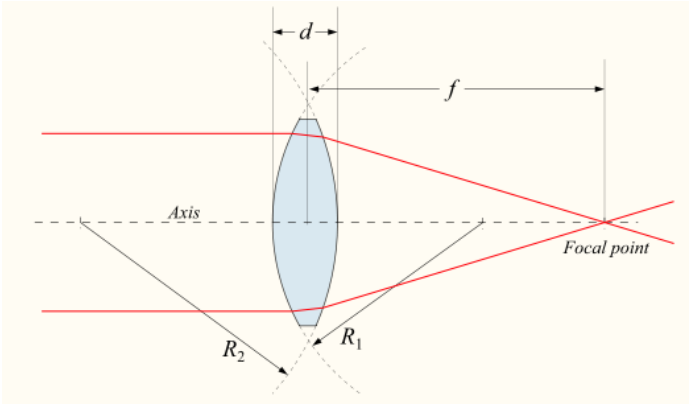


Figure 2.2 – Example geometry of a lens [4]**.**

This Thesis will often refer ahead to micro-lenses. A *micro-lens* is a very small *lens* with very small thickness d, in the order of the microns. Plugging in a very small d into Equation (2) if becomes Equation (3), a.k.a. the *thin lens equation*:

$$P = \frac{1}{f} = (n-1)\left[\frac{1}{R_1} - \frac{1}{R_2}\right] \qquad (3)$$

Approximating *n* to the ideal value of 0, meaning the lens is approximately non reflective, results in the Lens Equation (4), as demonstrated in [4]. This equation relates S_1, the distance of an object playing the role of a light source to the optical centre of the lens, with S_2, the distance of the projection, also known as *real image*, from that object on the other side of the lens to the optical centre of the lens, and *f*, the focal length of the lens [4]. A graphical representation of the relevant situation is included in Figure 2.3.

$$\frac{1}{S_1} + \frac{1}{S_2} = \frac{1}{f} \qquad (4)$$

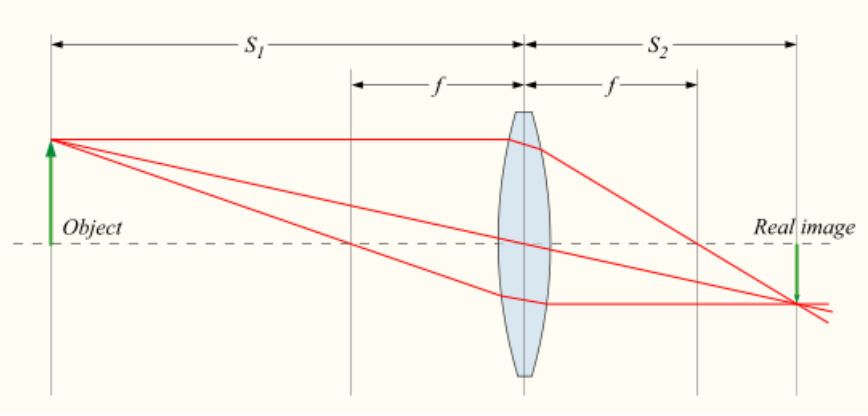


Figure 2.3 - Light source projection by a lens [4]**.**

Depth of Field

Aperture is the area of the opening letting in the light in a camera body. Depending on the aperture of a photographic camera, pictures may be captured with different sharpness at the various distances of the captured scene. Figure 2.4 a) and b) represent two image capturing scenarios: for both scenarios, the same type of lens is used, guaranteeing identical optical characteristics, i.e., focal length. The difference between the two scenarios is the aperture. Figure 2.4 a) shows a capturing case with a large aperture with the consequence of letting into the camera light rays from a larger *Angle-of-View* (AoV) while Figure 2.4 b) shows the opposite case.

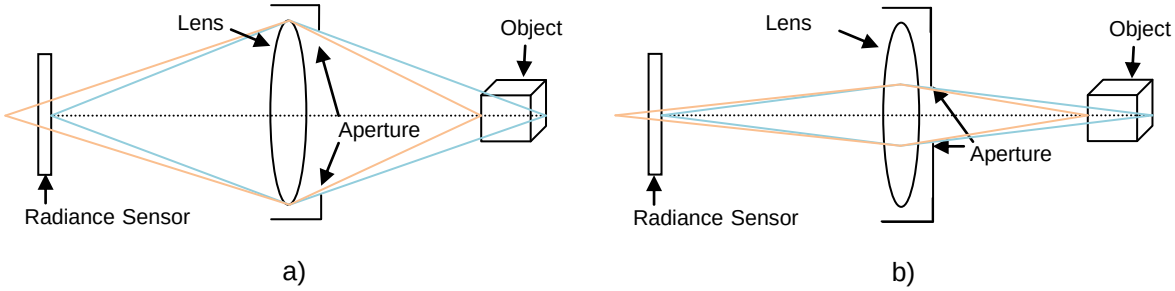


Figure 2.4 – Two image capture scenarios with different apertures: left) high aperture resulting in a small depth of field; right) low aperture resulting in a large depth of field.

Light rays coming from the focal length are represented by blue light rays in both scenarios while the orange lines represent the light rays coming from a different distance than the focal length. Considering the implications of Eq. (4) in both of these scenarios, the closer the light rays come from the point in space corresponding to the focal length, the smaller is the spatial range on the radiance sensor that will pick up the projection of light rays coming from that point in space. In summary, light rays coming from a single point in space scatter across the radiance sensor differently, notably proportionally to their distance to the lens focal length, with a growth rate determined by the size of the aperture (according to Eq. (4)).

The *depth of field* is then the range of distances in a scene where objects can be considered sharp enough in the captured image. There is, however, an undesirable consequence in the relation between depth of field and aperture. As the aperture becomes smaller, less light rays get into the camera, which translates into a reduction of the overall captured radiance, thus resulting in darker images. These concepts are illustrated in Figure 2.5, where only objects at distances close to the focal length, within the depth of field, are sharp.

Figure 2.5 - Depth of field examples [5]**.**

4D Light Field

In some scenarios, for instance capturing the light field inside a room, it is unpractical with current technology to capture, and subsequently store, radiance from all three spatial dimensions (x, y z) because an extremely large amount of sensors and storage devices would be required. However, if that task could be fulfilled, it would be very relevant to know how much information would be enough to match human sensorial capabilities.

Capturing enough information to be able to reconstruct the actual light field, as defined by Michael Faraday, may not always be required. In fact, humans are not sensorially equipped to handle such high amount of information. Biologically, humans are only able to process, at a time, the rays of light passing through two points in space, the focal points inside the eyes. Due to the human eye physiology and the way the eyes are disposed in the human head, notably horizontally separated by approximately 60 millimetres, humans can only capture light rays coming from about 180 degrees horizontally, 135 degrees vertically (see Figure 2.6), around the direction where the eyes are converging and with good detail only at much shorter angular ranges.

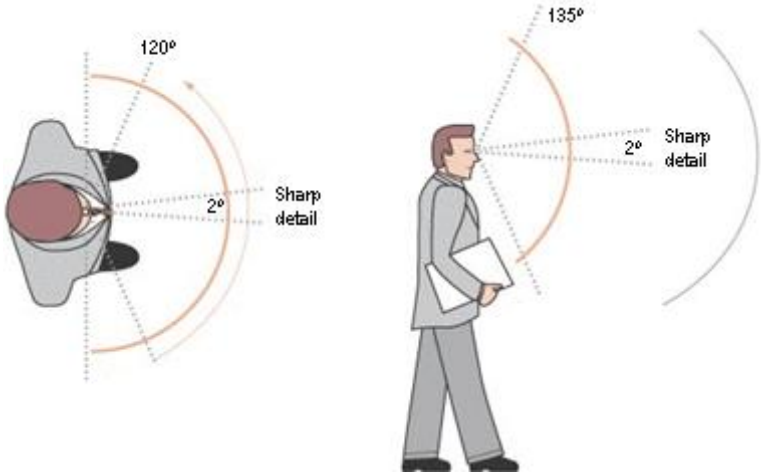


Figure 2.6 - Human field of vision [6]**.**

To describe the position of the human eyes in 3D space, knowing that they are 60 millimetres apart in the x-axis, only two other coordinates are required: one in the y-axis to describe the distance between them and another in the z-axis to describe height. Representing the angular directions of the rays hitting each eye in a fixed plane can be done with the angular coordinate φ (left image in Figure 2.6) and θ (right image in Figure 2.6). This angular coordinate system allows for the expression of human perception inside the 120 degrees of horizontal amplitude and 135 degrees of vertical amplitude. In summary, a representation with four dimensions (y, z, φ, θ) is required to express, and ultimately match, the amount of information the human visual system is designed to handle, the so-called *4D light field* function.

Stereoscopy

Stereoscopy is a technique for creating the illusion of depth in visual content by means of delivering two different perspectives of the visual scene to the two eyes. By combining the information of each perspective, the brain extrapolates the depth from the disparity of the objects present in both perspectives. This concept has been around for a long time, notably since it was described by Sir

Charles Wheatstone in 1838 [7], [8][1]. The principle known as *stereopsis* states that an object presents itself with a different perspective to each human eye. Observing a cubic object by closing one eye at a time, with the object at a certain angle and at close distance from the observer, the observer can perceive a face with one eye, Figure 2.7 a), that is hidden from the other eye, but visible when switching the eyes, Figure 2.7 b) [7].

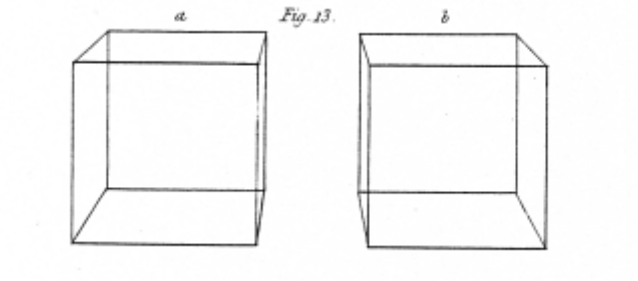


Figure 2.7 - Perception by each eye of a cubic object at a close distance to the eyes [7]**.**

Stereoscopy has been widely disseminated in recent years and has been adopted in several fields namely space exploration, movie industry, gaming industry, etc., to increase the Quality of Experience (QoE) in the process of consuming visual content, with remarkable success and public acceptance.

Holoscopic Image

A *holoscopic image* is a collection of 2D representations of the scene where each of these 2D representations corresponds to a different Point-of-View (PoV) and has a predetermined resolution. The common denomination for each of these 2D representations is *micro-image*, which is conceptually similar to a traditional 2D image. Figure 2.8 presents five micro-images, each of them corresponding to one specific but neighbouring PoV; this will be called in the following a *continuous set of micro-images*. This continuity is apparent in the object's location in each of the micro-images.

Figure 2.8 - Five contiguous micro-images captured from a scene.

By spatially grouping the several micro-images, according to the corresponding PoV, a simplified 2D representation of a 4D light field is obtained; in Figure 2.8, the micro-images are grouped in such manner. Figure 2.9 presents a full holoscopic image, highlighting the section corresponding to Figure 2.8.

[1] In his experiments to understand the phenomena of depth perception in human vision, Leonardo da Vinci used a sphere with which it is harder to observe the difference in perception of each eye, compared to a cubic object for instance.

Figure 2.9 - Example of a holoscopic image [9]**.**

This type of 2D light field representation, of a 4D light field, is commonly known as *holoscopic imaging*, *integral imaging, light field imaging,* or *plenoptic imaging*. In this Thesis, the denomination *holoscopic imaging* will be adopted.

2.2 Holoscopic Imaging Capture

This section will start by reviewing the basics on the traditional *2D light field* capture to provide insight to the initial light field capture methodologies. Next *4D light field* or holoscopic imaging capture is covered, introducing the basics of this technology. With all the basics covered, the rest of the section will review various systems to capture light fields as holoscopic images, starting with the basic conceptual camera setup, moving on to other relevant available designs including the Plenoptic 2.0 camera, the Lytro camera, the Raytrix camera and, finally, the 3D VIVANT camera(s).

2.2.1 Basics on Traditional 2D Light Field Capture

Radiance sensors in traditional 2D cameras, common photographic machines, are fixed in a flat 2D plane, as can be seen in Figure 2.10, and they allow light capture in only two spatial dimensions. A system of lenses, including the relay and main lens, is placed in front of the 2D sensor array (also radiance sensor) to refract onto the radiance sensor of the camera the light allowed into the camera, creating a projection of the scene on the radiance sensor. The vertical and horizontal angles at which light is allowed into the camera, the AoV, are determined by the *aperture* of the lens system, which essentially controls the size of the camera's entrance for the light rays.

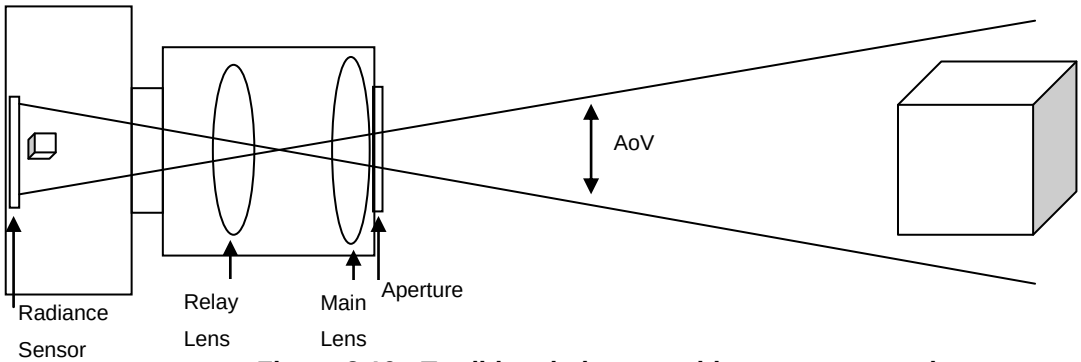


Figure 2.10 - Traditional photographic camera scenario.

The time dimension in this scenario is simplified to become a constant, but time may come to play an important role in video capture technology. This means that, for traditional imaging, the third spatial coordinate is fixed, both angular coordinates are also fixed and time is simplified to become a

constant, thus remaining only two spatial coordinates composing a *xy* plane where the samples are placed. The result is a 2D function which can be directly represented in a piece of paper or a projection screen, commonly known as a *picture*.

2.2.2 Basics on Holoscopic Imaging Capture

The 4D light function described above was called *photic field* in 1953 by Parry Moon [10], *4D light field* by Levoy in 1996 [11], *Lumigraph* by Gortler in 1996 [12], *4D radiance function* by Georgiev in 2008 [13] and, in practice, it is also possible to call it *light field* and *plenoptic function* since it is a simplification of an actual *light field or plenoptic function*.

In 1908, Lippmann suggested the means to capture this simplified 4D light field, using width, height, vertical and horizontal angular directions as dimensions, by projecting the rays of light travelling to a point in space in an organized manner, to which he called *Integral Photography* [14]. The idea is to refract, with very small lenses, which are commonly referred to as *micro-lens*, the rays of light coming from several angular directions, commonly referred to as Angle-of-View (AoV), that would otherwise all converge into a single point in space, commonly referred to as a Point-of-View. To illustrate how each micro-image refracts the rays of light, corresponding to each captured PoV, so each ray can be captured as part of an AoV range refer to Figure 2.11. Figure 2.11 a) shows an example of light rays converging to a single point in space while in Figure 2.11 b) a radiance sensor is placed at the point of convergence to capture the radiance, much like what happens in a 2D light field capture; finally, in Figure 2.11 c), the same light rays and sensor are present but a micro-lens is added to refract the light rays, causing them to hit the sensor in a different place.

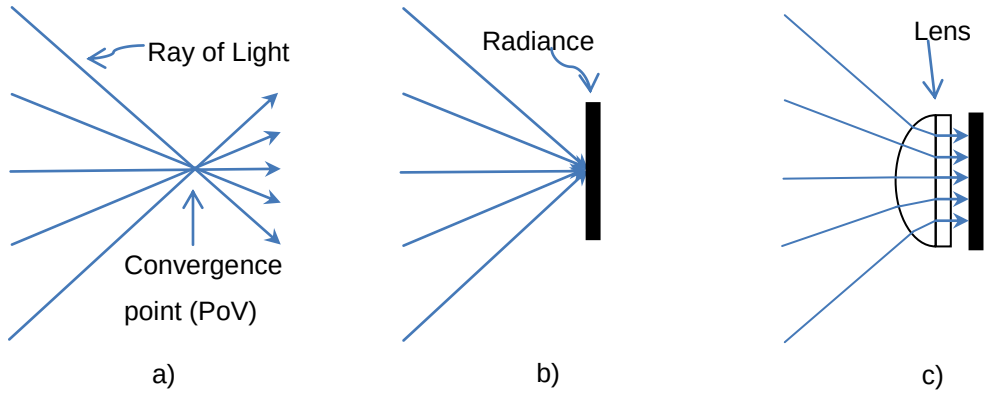


Figure 2.11 – PoV angular segmentation: a) Light rays converging at a point in space; b) The same light rays with a radiance sensor at the point of convergence; c) The same light rays being refracted by a micro-lens before hitting the sensor.

In 2D light field capture, only one lens is used, meaning that only one PoV is ever captured. By using multiple lenses, lined side-by-side, several discrete contiguous PoV can be captured simultaneously. The Lippmann's light field capture apparatus [14] consists on several of these devices, disposed in a 2D array, to allow capturing several micro-images for a given scene in a certain time instant. A graphical representation of this apparatus is presented in Figure 2.12, using the same components present in Figure 2.11 c).

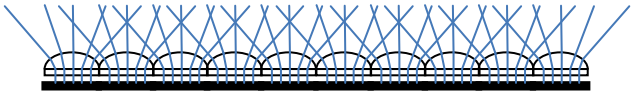


Figure 2.12 – Array of micro-lenses placed over a radiance sensor.

In practice, the capture can be achieved with greater performance when combined with current 2D image capture technology, placing the micro-lenses array (MLA) over a single radiance sensor, plus the elements already present in a traditional 2D photographic camera - a movable relay lens, an aperture element and a movable main lens (as depicted in Figure 2.13) to add focus and depth of field control. The subject of camera configuration will be extensively covered in the following section.

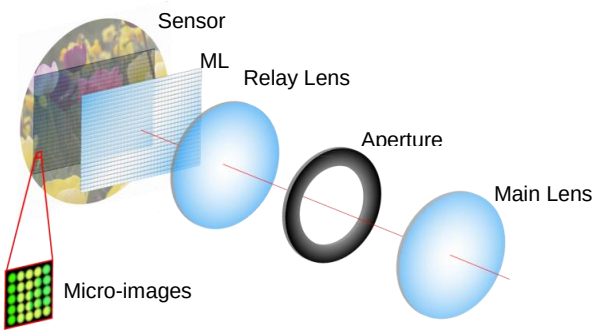


Figure 2.13 - Example of a light field capturing apparatus.

Naturally, there are limitations with the practical application of this technology that must be addressed. In theory, each PoV is composed of all the parallel light rays that enter the holoscopic camera's body. In practice, however, detecting all possible PoVs requires each radiance sensor to be the size of a photon and as fast as light to only detect one single light ray at a time which is not viable with current technology. In practice, depending on the resolution and *exposure*[2] of the radiance sensor, each captured radiance value is, in fact, an integral of light rays coming from a continuous range of points of view (see the left image of Figure 2.14). Moreover, the radiance captured for a specific angular direction is also a mathematical integral of a range of light rays coming from a range of angular directions (see the right image in Figure 2.14).

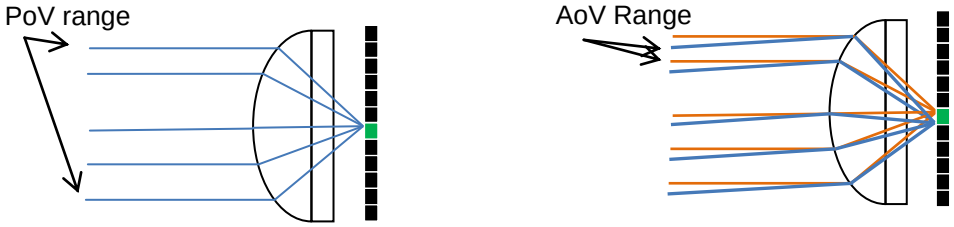


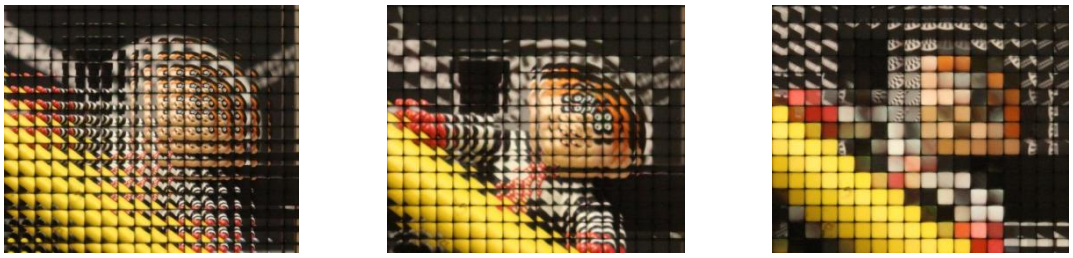
Figure 2.14 – Radiance sample AoV and PoV analysis: Left) Light rays from several different points of view being captured as the same point of view; Right) Light rays from distinct angular directions captured as the same angular direction.

To improve the precision of the captured light field, the size of each micro-lens must be as small as possible, to capture thinner PoV ranges, and the resolution of the radiance sensor as high as possible, to capture thinner angular ranges. Meeting these requirements will approach the real capture to the maximum theoretical performance by ultimately capturing each light ray with a single radiance sensor, knowing precisely its AoV.

Light rays coming from objects outside the depth of field will scatter widely across the image sensor, making the final image increasingly blurred. This effect can be found in all 2D photographic cameras and is most unpleasant when the depth of field is not centred at the region of interest chosen by the photographer. This is also the case with holoscopic imaging.

[2] The time the radiance sensor is detecting light rays.

Other effects, not present in 2D image cameras but present in holoscopic cameras, are the spatial and angular redundancy effects. If an object is far from the camera, the light rays emanating from it, and hitting the sensors, are close to perpendicular. For objects far enough away, the captured light rays will likely hit nearly all micro-lenses, resulting in representations of that object in nearly all micro-images. The object will register with high spatial sampling[3], meaning that the same region of the object will be sampled in multiple micro-images (see Figure 2.15 a)), but with low angular sampling[4], meaning that all representations of the object will have a wide AoV. However, if the object is very close to the camera (see Figure 2.15 c)), the light rays emanating from it and reaching each micro-lens will have a much higher angular sampling, so high that they may become disjoint among neighbour micro-images, and a much lower spatial sampling; Figure 2.15 b) shows the representation of an object with an average angular and spatial sampling.



a) b) c)

Figure 2.15 – Holoscopic image: a) Light field of an object placed far away from the camera; b) Light field of an object placed close to the camera; c) Light field of an object placed very near to the camera.

To better illustrate this effect, consider the experiment of an observer standing still while looking at an object far away, 5 meters for instance. Moving his head to get a different perspective will not yield any apparent result as the object will look identical. If the observer then moves closer, e.g. to about 30 cm from the object, moving his head will yield different perspectives (i.e. PoVs), with the object looking different as the observer moves his head. This effect is analogous with holoscopic capture and related to the concepts of spatial and angular sampling.

If objects are far away, they will have poor angular resolution, resulting in light fields with very low angular resolutions because all micro-images where the distant objects appear have approximately the same spatial representation of it. On the opposite extreme, if objects are very close, they will have excessively high angular resolution; in extreme cases, spatial regions of the object may not be detected – much like observing a large wall with the observers head touching its surface, resulting in omissions. To capture rich angular and spatial light field information, the object should be placed at a position where spatial samples slightly overlap each other, as in Figure 2.15 b). The variation of these parameters can also be controlled by placing an optical zoom element in the camera, as will be discussed in the next section.

2.2.3 Basic Holoscopic Camera

This camera solution is based on a theoretical model of what an holoscopic camera should look like, as envisioned by its inventor, G. Lippmann [14]. Significant differences exist among all presented

[3] Number of radiance samples corresponding to a point in space in a holoscopic image.
[4] Angular amplitude of the radiance samples corresponding to a point in space in the holoscopic image.

cameras in the following section and the original one by Lippmann. The presented cameras use digital sensors for capturing radiance as a substitute for the film in Lippmann's camera.

Basic Camera Elements

This type of camera can be built by simply placing a MLA over a radiance sensor. The basic camera elements are graphically depicted in Figure 2.16 and listed below:

Micro-lens array: A sheet composed of micro-lenses disposed in a square or hexagonal grid formation. All presented cameras have spherical shaped micro-lenses; however, there have been experiments done with cylindrical shaped micro-lenses.

Pickup device: Radiance sensor, sensitive to light rays.

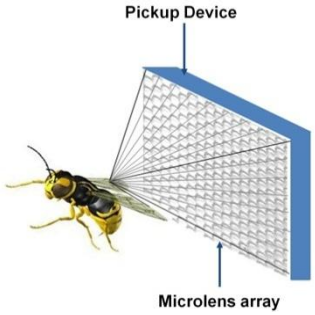


Figure 2.16 - Basic design of a light field camera [15]**.**

Main Features

In theory, this setup would work well provided that the sensor and MLA were exactly the same size to have a perfect mapping of the light rays hitting the sensor. The sensor would have to be infinitely sensitive, i.e. with an extremely high resolution, to detect single photons and the micro-lenses, in the MLA, would have to be small enough for the transitions between PoV to be unnoticeable by humans. However, the current technology does not allow for these ideal technical specifications.

Radiance sensors found in consumer electronics and also for professional use, like shooting a Hollywood blockbuster, are rather small, reaching today about 3 cm across, to match analogue film, with a maximum resolution of approximately 6 MPixels in professional video recording cameras and approximately 25 MPixels in professional photographic cameras. Micro-lens sizes can be as small as 90 microns but have to take into account the resolution of the radiance sensor placed behind it to guarantee good angular sampling.

- **Strengths**: Able to capture richer light field representations over traditional 2D cameras.
- **Weaknesses**: No ability to modify the placement of the depth of field, which has to be adjusted by manually moving the camera in 3D space to place the centre of the depth of field range at the region of interest; no ability to modify the aperture, being left with a very narrow (paper thin) static depth of field.

2.2.4 Plenoptic 2.0 Holoscopic Camera

This camera design, proposed by A. Lumsdaine and T. Georgiev in 2009 [16], introduces changes to the previous camera design to achieve increased performance in focus and depth of field control.

Basic Camera Elements

This type of camera is an improvement on the *Basic Holoscopic Camera*, thus including all its elements. A scheme of this camera, with the additional elements, can be found in Figure 2.17. The additional elements in this design are:

- **Main lens**: This lens has the function to manipulate the light field, following the discussion regarding the lens maker's equation (2), to allow for AoV control for all micro-images.
- **Aperture element:** This element determines both the depth of field and the shape of each micro-image; the aperture is static in this design.

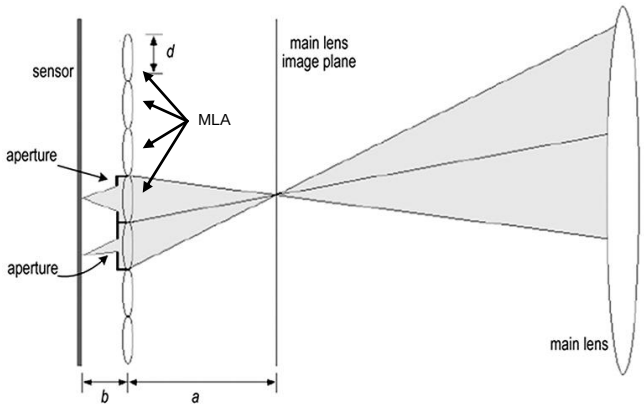


Figure 2.17 – Basic design of the Plenoptic 2.0 camera [17]**.**

Main Features

In this camera, the aperture is fixed, composed of light barriers placed behind each lens of the MLA, between the MLA and the sensor. The main lens is movable to allow variations of *a*, according to the lens Eq. (4), while *b* is constant and related to the focal length of the micro-lenses. The main lens placement allows for depth of field control as well as some AoV control.

- **Strengths**: Compared to the previous camera design, this camera can adjust its focal distance in relation to the MLA by adjusting the distance of the lens in relation to the MLA, thus shifting its depth of field in 3D space without the photographer having to adjust the camera position in relation to the scene; the aperture element allows for a larger depth of field than the previous camera model.
- **Weaknesses**: The aperture element of this camera does not allow for variations, resulting in a fixed depth of field and micro image ratio[5].

2.2.5 Lytro Holoscopic Camera

There are also companies that mass produce cameras with holoscopic technology for commercial purposes, namely Lytro and Raytrix, targeting the consumer electronics and the industrial inspection markets, respectively.

Lytro, Inc. was founded by Ren Ng in 2006, a light-field photography researcher at Stanford University. This was the first company to enter the 2D photography consumer electronics market with a plenoptic camera. The output of the camera goes into a piece of software to be processed. With the software, the user can chose the object to be in focus in the final 2D photograph output by simply using a

[5] Relationship between the spatial size, horizontal and vertical, of the micro-image.

regular 2D monitor. A preview is also available on the back of the camera body. More information on the company and its founder can be found at their website [18].

Basic Camera Elements

The public specification of the only camera model currently being produced by Lytro is presented in Figure 2.18; currently, its price is $399.00 at Amazon.

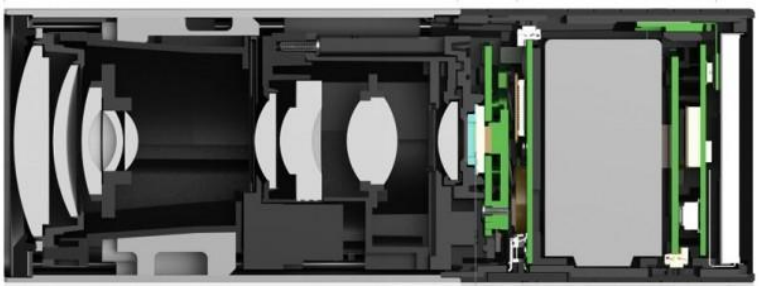


Figure 2.18 – Basic design of Lytro camera [18]**.**

The camera design is essentially the same as the Plenoptic 2.0 camera above. However, this camera is capable of displaying a 2D scene representation at acquisition time. This camera has these additional elements, in relation to the previously presented one, such as:

- **8x Optical Zoom**: System of lenses to filter AoV and manipulate the distance between PoV;
- **System of relay lenses:** Lenses system designed to compensate for multiple optical aberrations that may occur in optical systems; they may be a group of lenses (this case) or a single lens, depending on the aberrations introduced by the rest of the optical system;
- **Light Field Engine 1.0:** Module in charge of transforming the captured light field into a 2D reconstruction of the scene, based on some user input provided using a touch screen;
- **LCD display/Touch screen** – Small screen at the back of the camera receiving information from the *Light Field Engine 1.0* to display 2D reconstructions of the scene. The screen is also touch sensitive to allow the user to select the area of the scene he wants to bring into focus.

Main Features

The main characteristics of this camera solution are the ability to zoom into a specific area of the scene, thanks to the 8x Optical Zoom, giving the photographer more control over the light field capture process and the ability to convert that information into a 2D picture at capture time that the user can immediately see.

- **Strengths**: On-the-fly representation and display of the light field information into a human friendly format; optical zoom allowing for increased control of the captured scene without having to move the camera.
- **Weaknesses**: Depth of field range is still constant; no video capability.

2.2.6 Raytrix Holoscopic Camera

This is the second commercial light field camera available on the market. Raytrix, the company behind these cameras, was founded by Lennart Wietzke and Christian Perwaß in 2007 [19]. The main application of this camera technology is quality assessment for the industrial inspection of parts manufactured by assembly lines. The price is not publicly available, only upon request.

Basic Camera Elements

The public camera description is different from the Lytro's camera. Because it focuses on a different market, the Raytrix camera does not have some features such as the 2D LCD preview display, while it adds other features such as a Mix Focused MLA. The LCD preview display above is not needed for the application scenarios of these cameras; the same happens with the module creating the images to be displayed in the LCD. All the other camera elements are present. In summary, the design elements in this camera are:

- **Main lens:** Specifications unknown.
- **Aperture element:** Aperture control is unknown in this camera.
- **System of relay lenses**: Specifications unknown.
- **Mix Focused MLA**: Micro-lenses composing the MLA have different optical properties to capture PoVs with different focal distances. This enables the camera to gather multiple samples of a point in space with different sharpness, artificially extending the depth of field.
- **Dual-GigE [6]Ethernet**: Output interface to connect the camera to an inspection terminal where the holoscopic images are treated with custom software built for a specific application.

The left image in Figure 2.19 shows an example of an MLA composed of the mix focused micro-lenses used by this camera, while the right image in Figure 2.19 shows the corresponding group of captured micro-images.

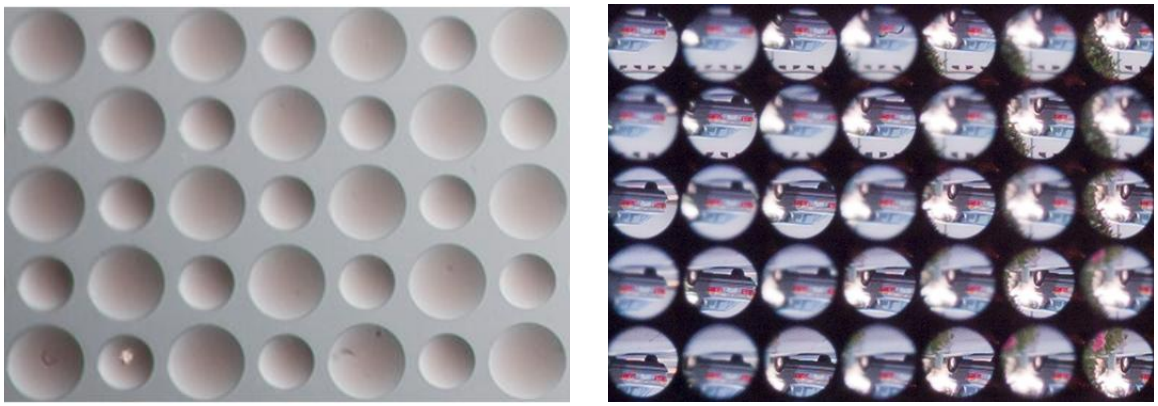


Figure 2.19 – Left) MLA with micro-lens of different focal lengths, which are coupled with the diameter of each micro-lens; Right) Array of micro-images captured with the MLA on the left; notice the differently focused micro-images, interleaved accordingly.

Main Features

Because some of the Raytrix camera's specifications are trade secret, this description cannot go into much detail. The main feature of this camera is the new type of MLA with several groups of lenses, each with identical characteristics. Each group of lenses may have a different number of lenses, all evenly distributed across the MLA, depending on their characteristics. The characteristics of each group of micro-lenses essentially allow varying the focal distance of the micro-lenses. The even distribution allows capturing radiance for points of view at even intervals inside each group. This is required so that every depth of field supported by each micro-lens group, and ultimately by the camera, shows good sampling properties for the corresponding focused light rays.

[6] 2 Gigabit Ethernet Over two physically separated 1 Gigabit Ethernet cables.

- **Strengths**: Enables good sampling at all distances with the new type of MLA, resulting in sharpness at all planes of a captured scene, commonly referred to as *all-in-focus*; it has video capabilities, but no public specification is available.
- **Weaknesses**: Very niche market oriented with the consequence that it does not include some interesting features like the Lytro's LCD preview display.

2.2.7 3D Vivant Holoscopic Cameras

In September 2013, two cameras were under development in the 3D VIVANT European project, one aiming at photography and another aiming at video capture. The 3D VIVANT European project aims to develop and implement the technology for a complete 3D holoscopic video system, all the way from capture to visualization [20]. Instituto de Telecomunicações is a partner of this project, with the main task of developing a codec to efficiently represent this novel imaging format. Figure 2.20 shows a picture of the video camera prototype under development by ARRI, a company based in Munique. The prototype is based on the Arri Alexa camera [21]. The prototype aiming at photography is being developed by Brunel University in London. Both prototypes use the same optical elements and essentially vary the camera body. The camera body includes only the sensor and the electronics responsible capturing and processing radiance sensor data. Since the photographic model resembles the Plenoptic 2.0 Holoscopic Camera, this description will focus on the 3D Holoscopic ARRI Alexa video camera, put forward by the 3D VIVANT project.

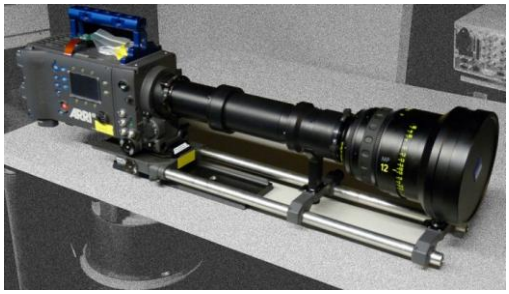


Figure 2.20 - ARRI Alexa camera fitted with the ARRI holoscopic capture tube [15]**.**

Basic Camera Elements

The 3D VIVANT video camera is similar to the Lytro camera in terms of optics, but the 3D VIVANT camera allows for video capture. Because it focuses on the professional moving picture market, it does not have some of Lytro features; however, it adds other features such as video recording capability. In this case, neither the LCD nor the image processing modules to provide images to the LCD are present. All the optical camera elements are present, i.e. the zoom element, the main lens and relay lens. The design elements introduced by this camera are:

- **Professional video capture**: Body of the camera is the ARRI Alexa camera, capable of capturing light field video sequences. The current camera is then a professional 2D video camera, with no modification apart from what is inside the acquisition tube (see Figure 2.21);
- **Changeable aperture piece**: Module placed in front of the acquisition tube to control the aperture of the camera.

A diagram of the camera's acquisition tube design can be found in Figure 2.21.

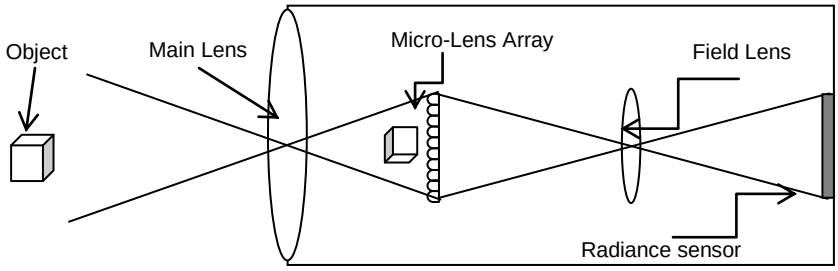


Figure 2.21 – Basic design of the 3D Vivant camera [15]**.**

Main Features

This camera is the first holoscopic camera to be built for the purpose of light field video capture. The recording formats are provided by the ARRI Alexa camera, which specification is available in [21]. For now, this camera is a prototype and the specifications for holoscopic video capture are being investigated. The holoscopic technology is also being studied for post-production effects such as refocusing and depth of field manipulation.

- **Strengths**: Capability to record professional video sequences in holoscopic raw video format;
- **Weaknesses**: Unstable prototype designs because the camera is still under development.

2.3 Holoscopic Imaging Displaying

This section will go over 2D display technology for a quick reference and review after the available 3D and holoscopic display technology, including previous generation and next generation visualization devices developed to take full advantage of holoscopic image and video content.

2.3.1 2D Displaying

2D displaying was the first display process to appear and it is to date the most widely used. Although the perception of depth in this type of displays is limited, there is still some depth perception as there are many depth perception cues that work with mono-view content. This type of display is labelled as mono-view displays as only a single perspective of the scene is delivered to (both eyes of) the viewer.
As holoscopic images are multi-viewpoint content, 3D holoscopic content cannot be fully exploited when displayed through 2D display systems as these systems are simply not able to appropriately interpret the structure of the light field information in holoscopic content (even in raw format). Thus, holoscopic image content must be pre-processed, prior to display, to extract data to be effectively displayed in 2D displays.

2.3.2 Stereoscopic Displaying

Stereoscopic displays are those able to simultaneously present to the observers two different PoV of the same scene, normally having in mind the two eyes. There are a multitude of ways by which a system can deliver two different views of a scene, one to each eye, thus exploiting the stereopsis effect, with different strengths and weaknesses. The most common solutions are:

- **Wavelength-Multiplexed Display:** This type of stereoscopic displays uses colour difference, as explained ahead, to present both view perspectives in one single image. The concept of colour difference behind this technology was first referred by W. Rollmann as *anaglyph images* [22] and consists in mixing the images for the right and left eye perspectives together as a single image,

with the help of colour filters. The reds are filtered out of the image corresponding to the left eye and the blues or greens are filtered out of the image corresponding to the right eye. The resulting image is viewed through a pair of anaglyphic glasses, where each lens filters out the colours used for the representation of the other perspective. Trying to make sense of these colour incomplete images, the brain interprets them as two different perspectives and fills in the missing colours. An example of these anaglyphic glasses can be found in Figure 2.22.

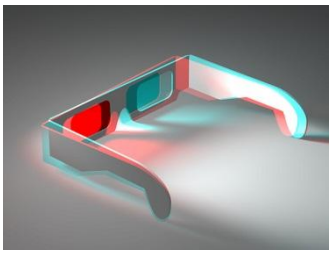


Figure 2.22 - Anaglyph image of 3D glasses used for wavelength-multiplexed displays [23]**.**

- **Time-Multiplexed Display**: This technology was patented in 1924 by Laurens Hammond [24]. The basic idea is that the pictures for each of the two view points, corresponding to each eye, are projected alternatively in the viewing screen. Since these images are slightly dissimilar, if a viewer were to look at the screen with the naked eye, the images would appear *jumped*[7] and not clear. To properly view these pictures, the viewer must wear some active shutter glasses which cover only the right eye while the left image is being projected on the screen and vice-versa, so that only one eye views the screen at a time. An example of this type of glasses can be found in Figure 2.23.

Figure 2.23 - Active shutter glasses used for time-multiplexed displays [25]**.**

- **Polarization-Multiplexed Display**: This type of stereoscopic display was invented by Sir David Brewster in 1879 [26] based on the idea of differently polarizing the light waves targeting each of the eyes. In this case, the two images in the stereoscopic pair are projected superimposed onto the same screen through individual projectors. The light beams from each projector are passed through corresponding orthogonal polarizing filters with polarization at 45 and 135 degrees. The viewer then uses a passive pair of glasses where each lens has a polarization filter oriented according to each of the projectors. In this way, each filter in the viewer's glasses only passes the light from one projector, creating the stereoscopic effect. A graphical representation of this process is presented in Figure 2.24.

[7] Portions of the image appear to repeat in two distinct areas of the image, depending on the depth of the objects in those areas.

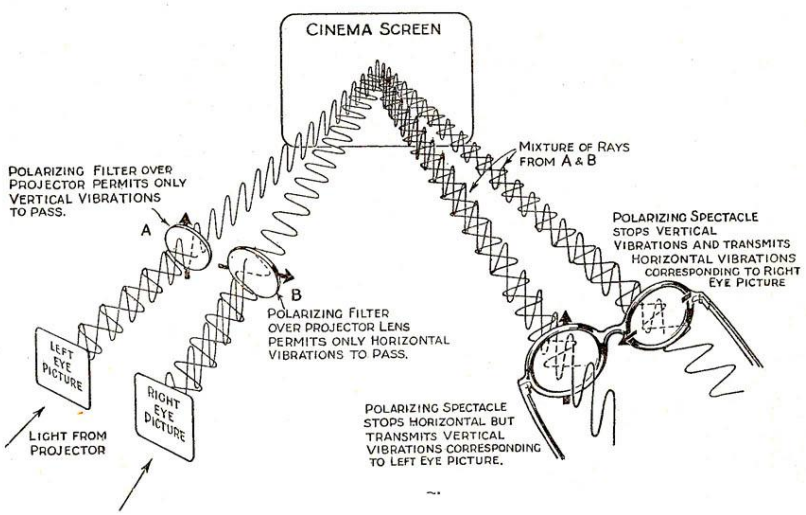


Figure 2.24 – Representation of the polarization-multiplexed display scenario [27]**.**

As in the regular mono 2D displays, also stereoscopic displays cannot directly deal with holoscopic images. Although 2D displays are mono-viewpoint and stereoscopic displays are stereo enabled, meaning they deliver two, instead of one, viewpoints of a scene, this is still not enough for holoscopic imaging. In this context, the original holoscopic content still has to be processed to appropriately extract two viewpoints, separated by approximately 60 mm horizontally and 0 mm of vertical misalignment, one viewpoint for each eye, in order an effective 3D experience is provided to the viewers using stereo displays.

2.3.3 Multiview Auto-stereoscopic Displaying

Multiview auto-stereoscopy displaying is conceptually similar to two-views stereoscopy with the main differences that a larger number of view-points may be displayed and no glasses are required. The user has a more natural experience with these technologies because they rely on the natural human action of moving the head to get another perspective, the so-called motion parallax. The displays either have built in devices to assert the user PoV or simply have fixed PoVs to help delivering the views to the correct eye. The most common types of display in this class are:

- **Specular Display**: Here two or more images are projected onto a curved screen so that the viewer perceives only two images, one with each eye, through reflection. In Figure 2.25, there is an example of a commercially available specular display system, the *Z-Dome* [28].

Figure 2.25 - Z-Dome specular display [28]**.**

- **Parallax Stereogram Display**: In this case, a multitude of points of view is displayed at the same time by the screen. To prevent light from non-relevant PoV to reach the viewer eyes, some barriers are placed in front of the screen, commonly referred to as *parallax-barriers*. When the

viewer moves his/her head, the relative position of the parallax barriers change and thus another point of view is seen by the viewer.

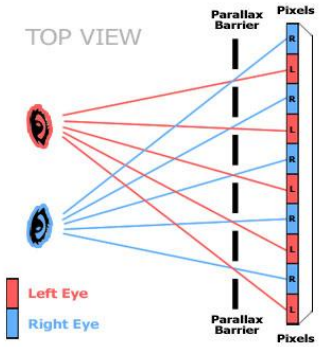


Figure 2.26 – Parallax barrier display scenario [29]**.**

- **Head Tracking Display**: This solution is coupled with a head tracking system for the display to dynamically decide what view to display to each viewer eye. The display screen may adopt any of the stereoscopic or multiview auto-stereoscopic display methods with the main difference here is that the image being delivered to each eye is real-time determined by the position of the viewer's head. Figure 2.27 shows an example of this type of display, using a parallax barrier display method.

Figure 2.27 - LG's D2500N-PN head tracking 3D display [30]**.**

Although these displays are more powerful, the same situation regarding the direct compatibility between holoscopic content and this type of displays happens. While they deliver a richer representation of a scene by showing more than one stereoscopic perspective, still holoscopic content enables many more perspectives, with less spacing between them, which these technologies cannot deliver (more granular horizontal parallax and also vertical parallax). In this context, some conversion processing is still required for effective visualization in this type of displays.

2.3.4 Holoscopic Displaying

3D holoscopic displays should be capable of recreating and representing a true 3D optical model with multiple PoV. This type of displays can handle the format and all the information present in a holoscopic image/video. In this technology, the perception of depth requires no glasses or any other gear whatsoever to aid the illusion. It produces a true sensation of depth by using natural light field reconstruction, projecting light rays in the direction they were travelling at capture time, resulting in an accurate reconstruction of the light at the moment of capture with natural horizontal and vertical parallaxes.

Currently, the company Holografika [31], also a partner in the 3D VIVANT project, started by Tibor Balogh in 1989, is developing the so-called *HoloVisio* displays, capable of naturally displaying 3D holoscopic content. The current product line features several display systems; one display system for

cinema (see Figure 2.28 a)), a professional application monitor (see Figure 2.28 b)) and, finally, one system for personal monitor use (see Figure 2.28 c)).

a) b) c)

Figure 2.28 - HoloVizio product line: a) HoloVizio C80 3D cinema system; b) HoloVizio 721RC high end model; c) HoloVizio 240P low end model [31]**.**

Both monitor-style and HoloVizio large-scale systems introduce a fundamentally new comprehensive approach to 3D displaying. Since HoloVizio is not a stereoscopic or multi-view system, it lacks most of the backlogs and drawbacks currently associated with 3D displays. HoloVizio technology is based on holographic geometrical principles with special focus on reconstructing the key elements of spatial vision. The pixels, or rather *voxels*[8]*,* of the holographic screen emit light beams of different intensity and colour to the various directions. A light-emitting surface composed of these voxels will act as a digital window or hologram and will be able to show 3D scenes undoubtedly being 3D [31]. A graphical representation of the basic difference between previously presented display technology and HoloVizio display technology is presented in Figure 2.29 where light rays in different directions have different radiance intensities.

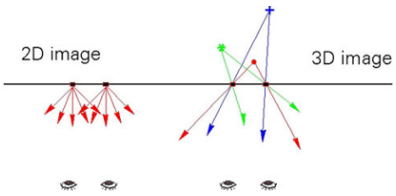


Figure 2.29 - Principle of the HoloVizio 3D display technology [31]**.**

In summary, each voxel of the display is able to emit light beams for the different colour components with different intensity in the various directions.

- **Strengths** - No need to use glasses; multi user support; continuous 3D effect across the PoV range; viewer with the ability to focus on different points of the image.

Weaknesses - Expensive (values upon request) because of its experimental status. Despite the various capture methods available, essentially all of them capture light fields. Because light field information does not have a natural process of representation yet available, as stated before, in cases other than the HoloVizio displays, the holoscopic data typically has to be processed before it can be displayed, for proper visualization. Several methods to extract views from holoscopic images have been developed over the years; a few of them will be reviewed in the next chapter.

[8] A *voxel* is a volumetric pixel associated to an intensity value in a grid in the 3D space; this is equivalent to a pixel which represents a value on a 2D space grid.

3
Extracting Views from Holoscopic Imaging: a Review

After formulating the problem of extracting views from holoscopic imaging targeting the currently available displays, this chapter will review the main solutions available in the literature to extract 2D views from holoscopic images, organized in two main classes, notably texture and depth based.

3.1 Problem Definition

Extracting 2D views from holoscopic images is a necessary process to appropriately display the captured light field information in 2D displays and in the various types of 3D (stereoscopic and auto-stereoscopic) displays nowadays available (and already reviewed in Chapter 2). In principle, displaying a full 3D holoscopic image [32] may be achieved by applying the inverse of the capturing process, this means by replacing the radiance sensor in a holoscopic camera by a flat panel display projecting the captured holoscopic image (see Figure 3.1). However, displaying part of the captured light field information in other types of displays requires processing the holoscopic image to extract the information in the appropriate format, notably: i) a single 2D image; ii) a stereo pair; and iii) multiple 2D views.

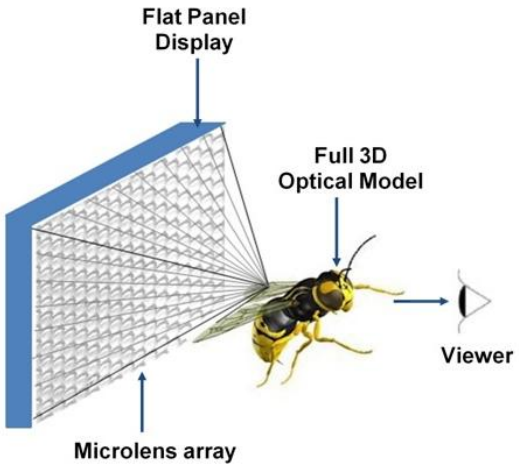


Figure 3.1 - Basic design of a light field display [15]**.**

Although the information required to reconstruct a 2D image of a scene is intrinsically included in the captured light field, the appropriate information must be selected and extracted to emulate the process of taking a 2D picture from a specific light field, in this case the captured light field. As explained in the previous chapter, in traditional imaging, and holoscopic imaging alike, the radiance values captured represent a mathematical integral of light rays. To better understand the extraction problem addressed in this chapter, it is time to go deeper into this issue and better characterize the structure of the captured light field data, in order to have an idea how the extraction/reconstruction methods may work. In traditional imaging, the angular range of light rays allowed into the camera by the main lens is spread out through the radiance values captured by the radiance sensor, meaning that no angular coordinate ever repeats in the camera sensor. This phenomenon is illustrated by the ray trace diagram in Figure 3.2 a) where the traditional imaging case is represented. In holoscopic imaging, the same angular range repeats throughout all portions of the radiance sensor behind each micro-lens. This means that each micro-image consumes the total range of angles allowed into the camera by the main lens. Each of these combinations of angle coordinates is repeated as many times as the number of micro-lenses in the camera; Figure 3.2 b) presents this case.

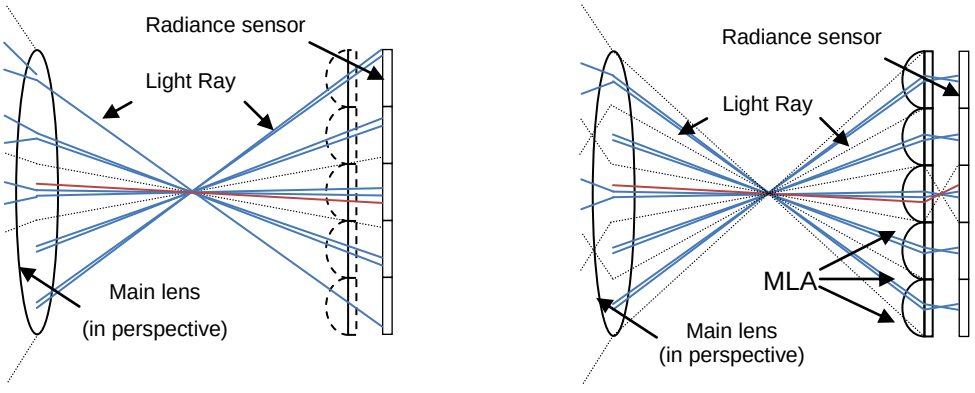


Figure 3.2 - Light field capture: a) traditional imaging capture; b) holoscopic imaging capture. The Red, Blue and Gray ray traces in case a) and b) are the same cases in both images, only in b) there is a MLA in place. All rays come through the Main lens in the direction of the Radiance sensor. The Red ray traces represent rays that in case b) correspond to the most central PoV. The Gray ray traces represent the limit angle allowed into the camera through the Main lens, at its edge. The Blue ray traces represent, in case b), samples corresponding to far edge PoVs.

Because of these angular properties, it is appropriate to say that each micro-image corresponds to a single traditional image. Although the angle ranges are the same for all micro-images, and equal to the total angular range the main lens allows, the PoV varies slightly among all micro-images (see Figure 2.12).

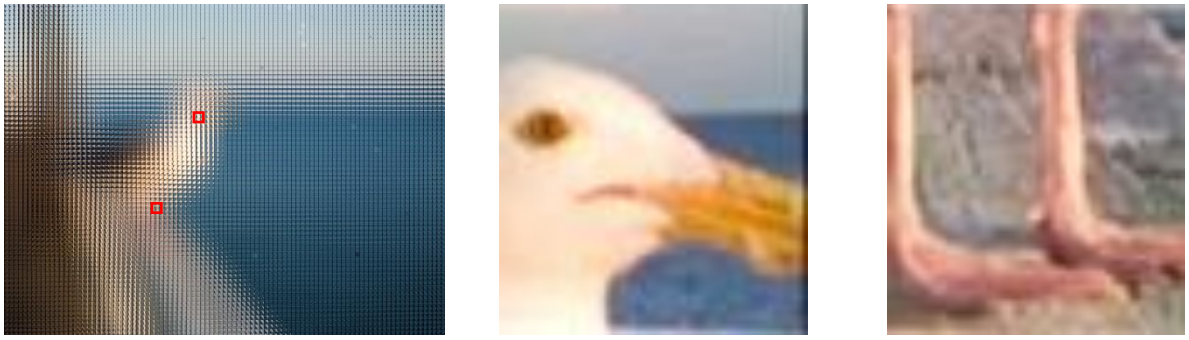
However, in practice, there are large differences between micro-images and traditional 2D images that come about for technological reasons, notably the radiance sensor resolution. In practice, the resolution of a micro-image is much smaller than the resolution of a traditional 2D image because the same radiance sensor technology used to capture a traditional 2D image is used to capture the multiple points of view present in the holoscopic image. This results in a factorization of the micro-image resolution, proportional to the number of micro-images in a holoscopic image. To illustrate this point, imagine that both cases in Figure 3.2 have a square radiance sensing area behind each micro-lens with 50x50 pixels. Imagining the MLA in Figure 3.2 b) 20x20 micro-lenses and the resolution of images coming from Figure 3.2 a) setup is 1000x1000, then the resolution of each of the 400 micro-images in Figure 3.2 b) setup would be 50x50 pixels.

This relation between different acquisition technologies is an important issue as it implies that to extract a 2D representation of a scene from a light field captured as a holoscopic image it would be enough to simply use one of the micro-images. However, there are several issues and problems with this simple scene representation using a single micro-image:

- Each micro-lens, plus the sensor behind it, acts like a *low resolution* traditional 2D camera with a high depth of field (see Section 2.2.1);
- The quality of the scene representation would decrease if the micro-images are upsampled to compensate for their low resolution;
- If each micro-lens is looked as a low resolution traditional 2D camera, each micro-image corresponds to a specific point of view of the scene. Because each micro-image corresponds to a small AoV range of the main lens, a single micro-image only captures a small fraction of the total scene, and thus it can easily be considered a very incomplete scene representation;

In conclusion, the choice of data extracted from the captured holoscopic image to reconstruct a 2D image to represent the scene must be carefully made to obtain a representation as faithful as possible of the full scene.

The second to last issue above is illustrated by the holoscopic image in Figure 3.3 a) from which two micro-images were selected. The upper selection corresponds to the micro-image in Figure 3.3 b) and the lower selection corresponds to the micro-image in Figure 3.3 c). It becomes obvious that a straightforward extraction approach like choosing a single micro-image to represent the whole scene is a very limitative solution.



a) b) c)

Figure 3.3 - Single micro-image representation: a) full holoscopic image; b) marked upper micro-image; c) marked lower micro-image.

A 2D reconstruction, at least one attempting to represent the whole scene, must be done in such a way that the whole scene is represented as faithfully as possible. There are several methods already in the literature with this target, each with its strengths and weaknesses, notably: i) the ability to sharpen objects in one image plane and blur objects in other image planes, apparently refocusing an image; ii) the ability to sharpen several image planes at once, resembling the process of regulating a cameras aperture; and iii) the ability to represent the scene as seen from a different point of view. These extraction methods will be presented in the next sections, organized depending on the main type of information they rely on, notably texture or depth. In a simple way, a texture based extraction is a form of texture processing based on the geometry of the camera, properties of the lenses and distance of the camera to the objects in the scene while depth based extraction extracts from texture some form of depth information to guide more complex geometry oriented extraction processes.

3.2 Texture Based 2D Image Extraction Solutions

The first type of 2D image extraction solutions will be presented in this section; these are extraction solutions based on the geometry of the camera, properties of the lenses and distance of the camera to the objects in the scene.

3.2.1 Angle of View Based 2D Image Extraction (AVe)

The first 2d extraction algorithm presented cannot be attributed to a single author or organization as it simply corresponds to the most obvious and simplest method that may be used for the extraction of a 2D image from a 3D holoscopic image.

Objectives and Basic Approach

The main objective of this method is to extract a 2D image from a 3D holoscopic image by simply grouping into one single image the pixel values corresponding to a specific AoV from all micro-images to reconstruct a 2D representation of the original scene. Consequently, the PoV of the 2D images reconstructed by this method corresponds to the centre of the chosen AoV pixels, which considerably limits the number of possible outputs.

As mentioned before, the total angular range allowed by the main lens is distributed among the radiance values in each micro-image, as illustrated in Figure 3.2 b). By further analysing Figure 3.2 b), it becomes clear that, for the MLA cases where all the micro-images have the same focal distance, as

all micro-images have the same size, the total angular range is equally distributed among the radiance values in every micro-image. Consequently, each radiance value in a micro-image corresponds to a subdivision of the total angular range depending on its spatial position inside the micro-image. For instance, the radiance value of the upper left corner of a micro-image corresponds to the same angular range in the upper left corner of all micro-images in the holoscopic image. This is true for all the other radiance values as well. This behaviour justifies the extraction method described in this section which reconstructs a 2D image by simply collecting the pixel values from all micro-images with the same angle of view.

Architecture and Walkthrough

The architecture of this method is presented in Figure 3.4.

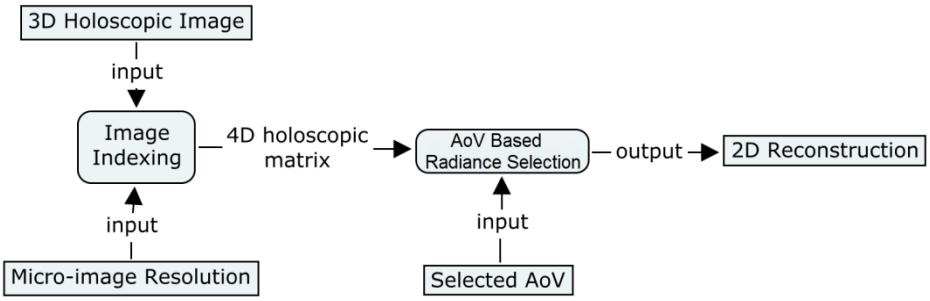


Figure 3.4 – Angle of View based extraction architecture.

From the architecture, it is possible to derive the processing walkthrough as:

1. **Image Indexing** – This module provides an easy indexing of the data in the original holoscopic image to facilitate the access to each radiance value; the inputs and outputs are:
    a. Input 3D Holoscopic Image – A 2D representation of the original light field organized in micro-images;
    b. Input Micro-image resolution – Horizontal and vertical size of each micro-image in pixels;
    c. Output 4D holoscopic matrix – This matrix organizes the data according to four dimensions: the first two dimensions define the PoV, in other words they index the micro-images in the full holoscopic image while the second two dimensions define the AoV, and in other words they index the radiance values inside each micro-image.
2. **AoV Based Radiance Selection** – The 2D image is defined by selecting the radiance values corresponding to the selected AoV, this means by selecting one radiance value from each micro-image, all in from the same position inside each micro-image; the inputs and outputs are:
    a. Input 4D holoscopic matrix – Output of the previous module;
    b. Input Selected PoV – Selected AoV (2D position) which allows defining the relevant radiance positions within each micro-image;
    c. Output 2D matrix – Radiance values matrix composed by the values corresponding to each *Selected AoV* position of each micro-image, thus composing the final 2D image.

Main Tools

The most important tool in the angle of view based extraction method presented in this section is the AoV based radiance selection tool as this is where the gathering of the data for the 2D image reconstruction from the holoscopic imaging happens.

AoV Based Radiance Selection

This module performs the key task in this method as it is here that the 4D light field representation is transformed into a 2D light field. This simple extraction algorithm relies on the angular symmetry of the radiance values in every micro-image to reconstruct the 2D images. The 2D image reconstruction provided by this algorithm is the result of assembling the radiance values for the same angular range in all micro-images into one single 2D image. The 2D image reconstructions produced by this algorithm are commonly referred as “views”. In fact, there is a finite number of different “views”, each one with a different PoV that can be extracted with this algorithm because the number of radiance values, i.e. angular ranges, in a micro-image is also finite. Figure 3.5 illustrates how the 4D holoscopic matrix, a rich representation of the original holoscopic image, is indexed to allow for the assembly of “views”. Each pixel in each micro-image (associated to a different colour in Figure 3.5) corresponds to a radiance value with a specific AoV. Each of the full coloured filled squares are “views” assembled by putting together each of the individual coloured squares from each micro-image in the holoscopic image. This assembling process is done respecting the spatial relation between the micro-images.

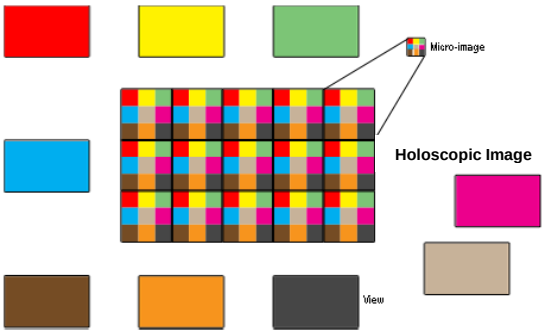


Figure 3.5 – Angle of View based extraction: the differently coloured rectangles correspond to different extract 2D views.

Further analysis of Figure 3.5 reveals that the resolution of each view becomes a fraction of the holoscopic image resolution. In the example, each micro-image has 3x3 radiance values and the holoscopic image has 5x3 micro-images. This allows for nine possible resulting “views” as output of this method, each with 5x3 radiance values (this also means pixels), compared to the 15x9 radiance values of the original holoscopic image.

Performance Assessment

A clear advantage of the angle of view based extraction algorithm is its simplicity as there is basically no data processing, besides matrix indexing. Thus, this is by far the extraction solution with the lowest implementation complexity.

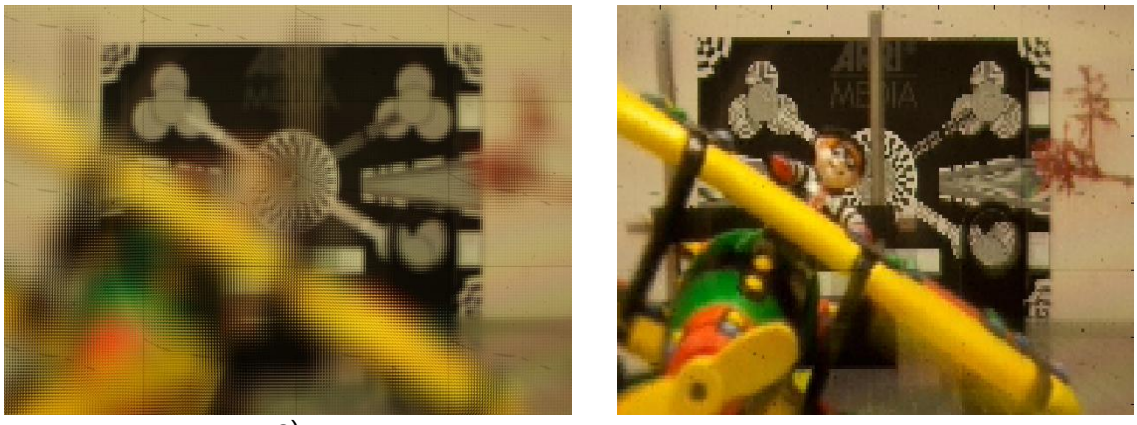
Regarding its disadvantages, this algorithm presents no means to address, counteract, compensate or control the following issues:

1. **Depth effects on micro-images** – As stated in the previous section and easily accessible in Figure 2.15 b), depending on the distance from an object to the camera, the AoV ranges for which there is a light field representation varies. Because the micro-images AoV range may overlap very little (see Figure 2.15 b), or not at all (see Figure 2.15 c), objects over the background may appear incomplete in the reconstruction or simply disappear, although there may be information of the missing region in other places of the light field. Figure 3.6 shows a 2D image extracted from the holoscopic image presented in Figure 2.15. The right ear of the doll seems to be missing in the

extracted 2D image although it is visible in the holoscopic images. In the case where the AoV ranges largely or completely overlap, foreground objects may have repeated sections, causing visual distortions.

2. **Stereo limitations** - The number of points of view allowed by this extraction method is equal to the resolution of each micro-image. Consequently, the ability to extract multiple pairs of PoV, ideally separated by 60 mm, for stereo displays is limited.
3. **MLA and micro-image ratios mismatch** – This method overlooks the relation between the MLA and the micro-image sizes. For instance, if each micro-image is 50x50 pixels and the MLA is 33x23 micro-lenses, the horizontal and vertical AoV size corresponding to each pixel does not match. In this method, because one pixel from each micro image is used in the reconstruction, all pixels are treated as if they correspond to the same size of AoV, which is not always true as demonstrated in the example. This raises a complex issue because either one direction is over-sampled or the other is under-sampled, to match the correct ratio of the output 2D scene representation. Consequently, the image ratio of the output "view" becomes incorrect if the MLA ratio[9] and the micro-image ratio do not match, resulting in a contraction or a stretch effect, depending if the ratios differ more vertically or horizontally, respectively.

A possible input for this 2D image reconstruction algorithm is presented in Figure 3.6 a) and the corresponding output for the most central PoV possible is presented in Figure 3.6 b).



a) b)

Figure 3.6 - Extraction example: a) original holoscopic image; b) extracted 2D view corresponding to the most central PoV possible (scaled to the holoscopic image size).

The image in Figure 3.6 a) is a 5576x3744 holoscopic image with 192x129 micro-images, each with a resolution of 29x29 (841) pixels, corresponding to a 1:1 ratio. The second image, Figure 3.6 b), corresponds to one of the 841 views that can be extracted with this algorithm, in this case the most central PoV. Notice that, although the micro-image ratio is 1:1, the extracted view does not have a 1:1 ratio, derived from the fact that the number of horizontal (192) and vertical (129) micro-lenses used in the MLA do not match. Although it is difficult to notice the three previously described effects in Figure 3.6 b) without using for comparison an actual 2D representation of the scene captured by the traditional 2D capture method, the scene is in fact horizontally stretched, the background and foreground sharpness is diminished and the extracted view is the most central possible, not the most central in absolute.

[9] Relationship between the number of micro-lenses, horizontal and vertical, in an MLA.

3.2.2 View Selective-Blending 2D Image Extraction (VSBe)

This method was developed in the context of the European 3D Vivant Project as a means to reconstruct 2D image viewpoints from holoscopic images, and after to perform object segmentation in holoscopic images, as stated in the project report [33]. The description of the method presented in this Thesis is also based on the report available in [34].

Objectives and Basic Approach

This algorithm aims to reconstruct 2D high resolution images from holoscopic images incorporating the extraction method presented in the previous section, the AVe algorithm. With this purpose in mind, a number of selected views are initially extracted based on the algorithm described in Section 3.2.1, with these views having adjacent angular ranges. These views are then blended to generate an output 2D image reconstruction. The PoV of the 2D image reconstruction corresponds to the centre of the PoVs of the views used as an input to this algorithm, this means the output PoV is the central of the views PoVs. Because of this, new PoVs can now be represented with this method that could not be addressed by the previously presented one. The bottom line here is that all extraction methods have a limited number of PoVs they can extract. By processing the information in a more efficient way than the previous method, this method is more granular and so can deliver reconstructed 2D images representing PoVs between those possible with the previous method. These new PoVs will be referred in the following as *Sub-PoVs*.

The views are combined by stacking them one on top of the other. When the views align perfectly, it can be said that they haven't drifted from each other; otherwise, if they don't stack perfectly, it can be said they have drifted. More precisely, the *drift* accounts for the distance between adjacent views, views with the lowest distance from the PoV allowed by the previous method, in the stack of views. A feature of this type of views combination is that they can be combined with different drifts from each other. The range of drifts that can be applied in the view stack can be understood as the range of depths the method is able to focus the scene. The focus can be varied within a discrete range of values. This focusing interpretation will be more extensively covered ahead.

Architecture and Walkthrough

The architecture of this method is presented in Figure 3.7.

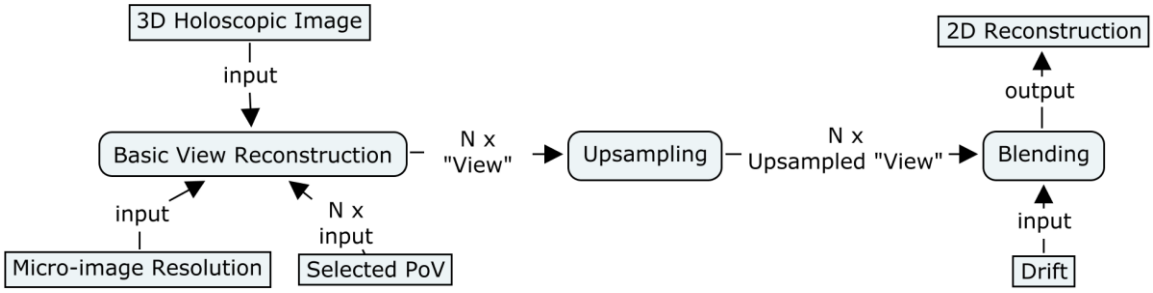


Figure 3.7 - View selective-blending 2D image extraction architecture.

The main processing modules in this method are:

1. **Basic View Reconstruction** – First, an unspecified number of adjacent views, typically 5 to 7, starting from a selected PoV (see Figure 3.8), are extracted from the holoscopic image using the algorithm described in Section 3.2.1; the inputs and outputs are:

a. Input 3D Holoscopic Image – A 2D representation of the original light field;
b. Input Micro-image Resolution – Horizontal and vertical resolution of each micro-image, this means the number of pixels in each direction;
c. Input Selected PoV – 2D position within a micro-image pinpointing a radiance value in each micro-image, one for each reconstructed view. An unspecified number of these values are used as input here. The number can be between two and the micro-image full resolution;
d. Output N Views – Multiple 2D views are extracted, i.e. 2D light field representations of the original scene.

2. **Upsampling** - Each view is upsampled with a bicubic interpolation algorithm to generate additional radiance values, thus increasing the 2D extracted images resolution. This should allow the views to be blended with a more granular drift interval in the next module; the inputs and outputs are:
    a. Input N Views – Output of the previous module;
    b. Output N Upsampled Views – N upsampled 2D light field representations of the original scene.
3. **Blending** – The upsampled views are layered on top of each other. To control the depth, the images can be overlaid with a specific drift between them. This drift is an integer offset value, meaning that the offsets allowed must be multiples of the pixel dimensions. This ensures that pixels will stay precisely on top of each other. The average value of the stacked pixel is calculated for every position, resulting in a high resolution 2D image output.
    a. Input N Views – Output of the previous module;
    b. Input Drift – Value controlling the overlay between the views (further explanation of this value can be found ahead).
    c. Output 2D Reconstruction – High resolution 2D image corresponding to the light field representation of the original scene for a specific sub-PoV.

Main Tools

The main tools of this method are the Upsampling, as it is where the views 'gain' resolution, and the Blending module, as it is where the individual PoVs from each view are combined to create a central estimated PoV, thus defining the output.

Upsampling

In the literature [33] and [34], only a few examples regarding this tool are described, and not the actual processing details. Figure 3.8 shows an example with the set of views that can be selected by this method to generate the 2D images corresponding to a left, a right and a central point of view, respectively. Using as reference the view (0,0) at the top left corner, the views selected for the extraction of the central point of view are views (1, 1), (1, 2), (2, 1) and (2, 2). These are the most central and continuous PoVs, therefore the best suited to extract the most central PoV.

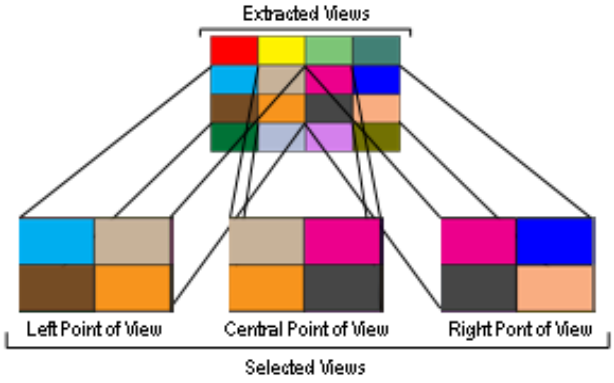


Figure 3.8 - Selecting views for the view selective-blending 2D image extraction method.

To reconstruct points of view other than the central PoV, the algorithm must choose groups of views centred further away from the central PoV. The algorithm then interpolates sub-pixel values for each view (this means non-existing pixels between the existing pixels), artificially increasing the resolution of each of the selected views through bicubic interpolation. The upscaling factor is proportional to the number of views used in the reconstruction, meaning that if the views have a resolution of 8x8 pixels and 9 are chosen, 3 horizontal and 3 vertical, the views are scaled horizontally and vertically by a factor of 3, making each one 24x24, and the spatial resolution equal to the sum of the selected views resolutions. In Figure 3.8, the scaling factor is 2 for the central view reconstruction, meaning that the size of each upsampled view is doubled in each direction regarding the original, without correcting the ratio deformation that may be introduced in the reconstruction of the views.

Blending

This blending operation is essentially a two stages process where the upsampled views generated by the previous module are merged together:

1. First, all upsampled views are overlaid on top of each other;
2. Second, the position $position(x,y)$ of the views in relation to each other is offset, horizontally and vertically, according to a $offset$ drift parameter in relation to their original position $viewNumber(x,y)$. The drifted position of each radiance value is calculated with Equation (5):

$$position(x,y) = offset * viewNumber(x,y) \qquad (5)$$

3. Third, for each overlaid pixel position, the average value among all the upsampled views is calculated.

The blending process is depicted in Figure 3.9. In the first stage, the upsampled views are side by side. In the second stage, the upsampled views have been overlaid precisely on top of each other, but not averaged together just yet. Finally, on the third stage, the upsampled views are drifted to their position according to the drift parameter and the values occupying the same positions are averaged, completing the blending process.

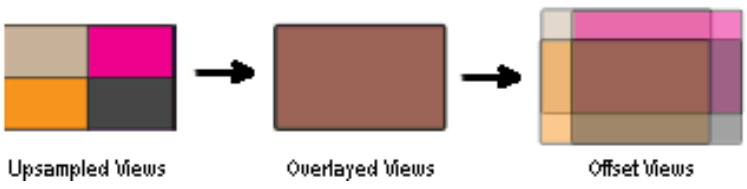


Figure 3.9 - Overlay and Average

As mentioned before, there is an input parameter, the *drift*, to artificially adjust the focus at which to render the final reconstructed 2D image. When the drift variable is set to 0, the upsampled views are perfectly stacked on top of each other, this means perfectly aligned. When the drift is set to an integer positive or negative value, the images will no longer align perfectly, but they will rather drift from each other, with a distance equal to the absolute value of the drift variable, in the direction set by the sign. The reason why this process works as focus control is that, when the angular ranges are being averaged together, the prevailing AoV is the angular range contributing with the most radiance energy for the average operation.

Performance Assessment

One advantage of this method regarding the previous one is the possibility to represent more PoVs. Another big advantage is the possibility to create 2D reconstructions with higher resolution than the previous method, thanks to the upsampling tool.

A drawback of this method is the complexity inherent to the bicubic interpolation, as it is more complex than bilinear and nearest-neighbour interpolations. In terms of output, although the bicubic interpolation produces smoother results than other methods, it also has some inevitable interpolation artefacts. Another drawback, still relating to the bicubic interpolation, is that the interpolation operation will modify the original data, meaning that the original radiance values are not guaranteed to be preserved, thus introducing error in the original radiance data [35]. Another drawback regards the rigidity of the method, resulting in its inability of bringing some depths into focus. As the overlaid pixels need to match perfectly in the blending module, the drift parameter controlling the focus can only be an integer value. By design, fractional numbers are not allowed by the adopted blending algorithm. Consequently, there can be depths for which this extraction method cannot provide focused images.

In comparison with the previous extraction method, this method counteracts the “depth effects on micro-images” effect with the drift mechanism by attempting to place the objects on their correct spatial position, according to its depth. It also partially addresses the “stereo limitations” with the upsampling tool by increasing the total number of possible outputs. Finally, the “MLA and micro-image ratios mismatch” still remains unaddressed; resulting in disproportional output images, with the source of the disproportionality being the views generated by the included AVe method which does not guaranty scene proportion.

3.2.3 Single-Sized Patch Based 2D Image Extraction (SSPe)

This method developed by Todor Georgiev, in 2010, reconstructs 2D radiance representations of the original scene based on the captured radiance information adopting a patch based approach [36]. This solution is based on essentially the same principles as the previous solution while adopting a different and well defined approach towards reaching high quality scene reconstructions.

Objectives and Basic Approach

This extraction method enables the accurate selection of the PoV by modelling the reconstruction considering the geometry of the scene capture with the holoscopic camera. In essence, this algorithm is a sophisticated version of the algorithm presented in Section 3.2.2, modifying some of the modules to better replicate the process of taking a 2D picture of a real scene light field, in this case with a

"virtual 2D camera" pointed at the captured light field. This means that this extraction method mimics a 2D camera behaviour when it is presented with a light field.

Architecture and Walkthrough

The architecture of this extraction solution is presented in Figure 3.10.

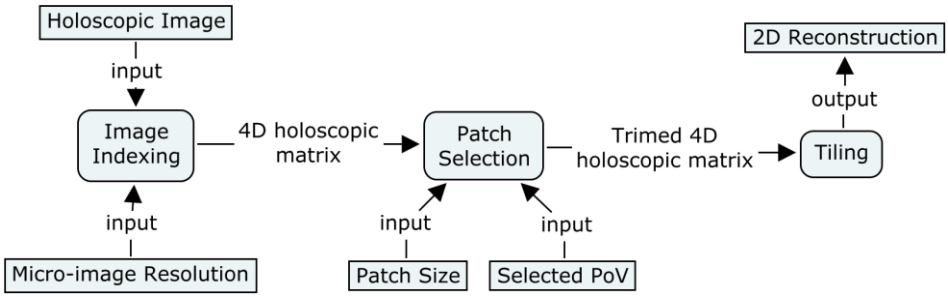


Figure 3.10 - Single-sized patch based 2D image extraction.

The main processing modules of this method are:

1. **Image Indexing** – This module is the same as for the extraction solution in the previous section.
2. **Patch Selection** – From each micro-image, a square patch of radiance values is extracted based on an input parameter. Because of the continuity of the radiance values inside each patch, patches are associated to horizontal and vertical angular ranges. Varying the angular ranges, i.e. grabbing patches not centred with the micro-image, will result in a perspective, i.e. PoV, shift. This process corresponds to moving a real 2D camera, horizontally or vertically, depending on the shift direction. For each micro-image, radiance values outside the chosen patch are removed from the 4D holoscopic matrix; the inputs and outputs are:
    a. Input 4D holoscopic matrix – Output of the previous module;
    b. Input Patch Size – Parameter defining the focal distance of the reconstruction by controlling the size of the radiance values patch coming from each micro-image;
    c. Input Selected PoV – Parameter defining the PoV for the reconstruction by controlling the positioning within each micro-image of the window defining the patch to be used;
    d. Output trimmed 4D holoscopic matrix – Corresponds to the input 4D holoscopic matrix with values outside the selected patches removed;
3. **Tiling** – The selected patches from the original light field are arranged side-by-side, as if they were tiles, to reconstruct the final image;
    a. Input Trimmed 4D holoscopic matrix – Output of the previous module;
    b. Output 2D Reconstruction - 2D light field representations of the original scene for a sub-PoV.

Main Tools

The main tool of this method is the Patch Selection because this is the main innovation regarding the previously presented methods and also because it corresponds to a good abstraction of the 2D acquisition process.

Patch Selection

As explained in previous sections, the aperture directly influences the depth of field at the expense of resolution. This is achieved by allowing into the camera narrower and narrower sets of AoV in order to

reduce the blurring effect caused by light scattering from AoVs farther away from the central one. It has also been established that aperture influences the total range of AoV captured. With this in mind, the two variables controlling the aperture and the position of the "virtual camera" are presented:

1. **Patch Position** – To control the "virtual camera's" position, this method uses a two-dimensional parameter to place the patch extraction window, horizontally and vertically, in relation to the micro-images, giving the method the ability to adjust the PoV of the extracted 2D images.
2. **Patch Size** - To control the "virtual camera's" aperture, this method uses a parameter with the sole function of cropping each micro-image, virtually varying the size of all micro-images. This process corresponds to the aperture control in an actual 2D camera, by cutting out AoV ranges and gathering data on narrower AoV, virtually modifying the depth of field. In summary, varying the patch size will artificially focus the reconstructed image by virtually modifying the depth of field.

To illustrate this tool, consider Figure 3.11 a), b) and c) where various patch sizes are used to assemble a full image representation of the letter "A". In the three cases, three different patch sizes, *M,* are used to assemble a full representation of the captured letter "A", by three micro-images; what varies between the three cases is only the patch size.

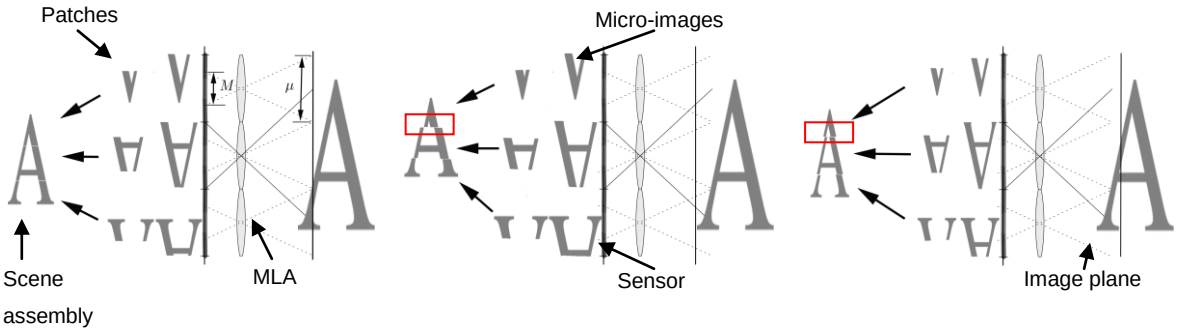


Figure 3.11 - Single-sized patch based 2D image extraction examples: a) proper scene assembly; b) *M* is too small, resulting in artefacts; c) *M* is too large, resulting in artefacts [36].

In Figure 3.11 a), there is an example of the virtual image capture geometry, when a proper scene assembly is achieved for a particular object. In that case, the chosen patch size, *M,* enables a sharper representation of the letter "A", than in b) and c). This is noticeable because the letter "A" in a) very closely resembles the original letter "A". In Figure 3.11 b), there is an example of scene assembly when *M* is too small for the object in the scene being rendered, resulting in the artefacts marked by the red rectangle. In Figure 3.11 c), there is an example of scene assembly when *M* is too large for the object in the scene being rendered, resulting in the artefacts marked again by the red rectangle.

This process of patch selection is best described by analysing the relation between patch size and object size presented in Equation (6). The derivation of Equation (6) from Equation (4) is done in [16].

$$\mu = M * \frac{a}{b} \qquad (6)$$

In this equation, *M* represents the size of the object region being projected on the sensor, through the micro-lens with size *μ, a* is the distance between each micro-lens and the image plane, i.e. the light source, inside the holoscopic camera, and *b* is the distance between each micro-lens and the radiance sensor, equal to the micro-lens focal length.

Performance Assessment

An advantage of this method is its simplicity, resulting in a very light to implement algorithm compared to the previously presented one. Another advantage is that it does not modify the data of the original light field, making the process fully reversible.

A disadvantage of this method is that it produces very noticeable artefacts for objects at depths not artificially brought into focus, as can be seen in Figure 3.12, in the marked region of the background. Another disadvantage is the low resolution when the patch size becomes too small.

Figure 3.12 - Out of focus rendering effect: the patch is too large for the background, resulting in obvious artefacts.

3.2.4 Single-Sized Patch Blending Based 2D Image Extraction (SSPBe)

This 2D image extraction method was developed by Todor Georgiev and Andrew Lumsdaine and is fully described in [36] [35] [37].

Objectives and Basic Approach

This 2D image extraction method is an optimization of the method presented in the previous section with the same basic objectives but providing increased output resolution. In this optimization, the *M* sized patches are still tiled together side-by-side, like in the previous method, but instead of cutting them to fit, the tiles are blended together.

This method eliminates less original information, which ultimately translates into higher output resolution because, unlike the previously described methods, the blending means interleaving pixels, not averaging them. As a consequence, the patch resolution does not need to be adjusted to guarantee the same horizontal and vertical resolutions. In fact, since this method intends to use more information outside the patch sizes, this is more an advantage than a problem because more information from the original light field is being used to assemble a 2D extraction of the captured light field.

Architecture and Walkthrough

The architecture of this method is presented in Figure 3.13.

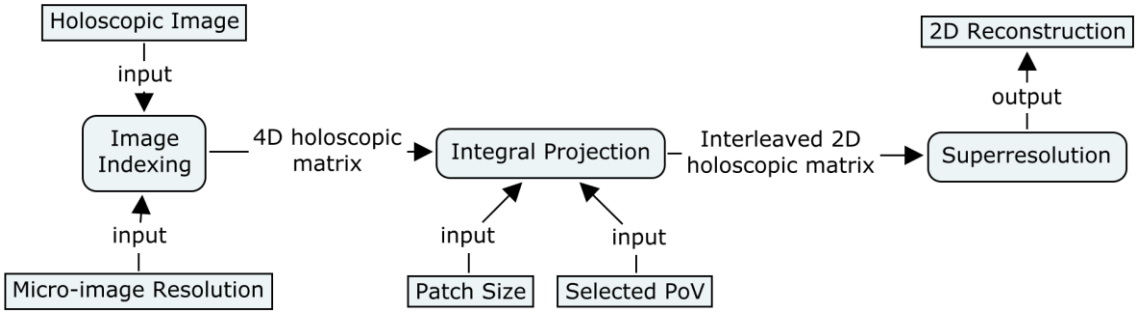


Figure 3.13 - Single-sized patch blending based 2D image extraction architecture.

The main processing modules of this method are:

1. **Image Indexing** – This module is the same as for the extraction solution in the previous section.
2. **Integral Projection** – Each micro-image is projected onto a 2D projection plane which can be abstracted as a plane where the final extracted image will be assembled. The *Patch Size* input determines the size of a square block of pixels in the centre of each micro-image that will correspond to the original size of a micro-image on the projection plane, i.e. scaling up the micro-images in the projection plane. The *Selected PoV* input parameter determines the horizontal and vertical drift, in the 2D projection plane, of each micro-image projection. If we look at the centre of non-drifted projections, we find samples of a central AoV range. This can be interpreted as extracting a central PoV from a central AoV. If there is a drift, the entire projection is drifted; in this case, some samples fall out of the projection plane and also some regions become empty. This issue will be resolved in the next module. For all intents and purposes, this drift equates to moving a real 2D camera horizontally or vertically, depending on the drift direction, while still pointing it to a certain point in the scene. This process, in a way, produces the same result as the previous SSPe *Patch Selection* module because patches from all the micro-images are being placed side by side, according to the organization of the micro-images in the holoscopic image. However, now their margins are not discarded, and they are rather stored in the adjacent micro-image projections as sub-pixel (pixels between the real pixel positions) values, interleaving the micro-images in the 4D holoscopic matrix. The inputs and outputs are:
    a. Input 4D holoscopic matrix – Output of the previous module;
    b. Input Patch Size – Parameter defining the focal distance of the reconstruction by controlling the size of the radiance values patch coming from each micro-image;
    c. Input Selected PoV – Parameter defining the PoV of the reconstruction by controlling the position of the window where the patch will be centred;
    d. Output Interleaved 2D holoscopic matrix – Corresponds to the input 4D holoscopic matrix where each micro-image is repositioned to interleave the radiance values according to the patch size parameter and shifted to reflect the selected PoV parameter in the resulting 2D image.

3. **Superresolution** – In the previously presented method, the patch margins are discarded and thus the resolution regarding the original holoscopic image is dramatically decreased, especially for small patch sizes. However, in this method, the margins are used as sub-pixel values. By increasing the number of pixels in the extraction method, the resolution goes up, hence the name of *Superresolution*. This method allows non-integer patch sizes for finer depth-focus, leading to uneven pixel spacing that is addressed by using a Gaussian function to estimate sub-pixel values. Details on the Gaussian function are presented on the next section. For this module, the inputs and outputs are:
    a. Input Interleaved 2D holoscopic matrix – Output of the previous module;
    b. Output 2D Reconstruction - 2D light field representation of the original scene for a sub-PoV.

Main Tools

The main tools of this method are the integral projection and the Superresolution creation tools.

Integral Projection

The basic idea behind the Integral Projection tool detailed in [37] can be explained by analyzing Figure 3.14 and Figure 3.15. Figure 3.14 highlights overlapping projections of neighbouring micro-images as examples of projection and blending. When the rendering image line comes closer or moves farther to the plenoptic image, the radiance value density decreases or increases, respectively.

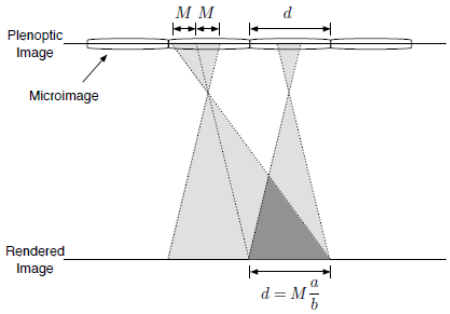


Figure 3.14 - Rendering with blending, through Integral Projection [36]**.**

Another way to explain the way micro-images are blended is depicted in Figure 3.15, which for simplicity assumes point pixels (both in position and angle). The process of projection and subsequent blending is accomplished by the following steps:

1. All micro-images are placed side by side, with an angle, in relation to a projection plane.
2. All micro-images are then spaced with such a distance that, if all pixels within the chosen patch size (rounded to integer) of each micro-image were to be projected vertically to the image plane, the image projected would correspond to the one in the SSPe method;
3. All micro-images are moved up or down, in relation to the projection plane, according to the PoV parameter, to shift the perspective;
4. Finally, the micro-images are projected onto the plane with an angle (not vertically), representing a 2D reconstruction, i.e. an interleaved 2D holoscopic matrix. The angle of projection is directly related to the geometry of the capturing process, allowing for the projections to overlap precisely at the image plane to be focused; in other words, the radiance values are interleaved to spatially match the AoV, intersecting at a certain depth.

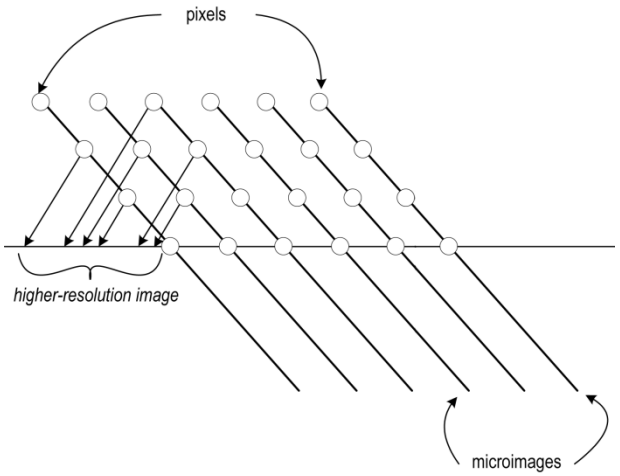


Figure 3.15 - Projections can produce images with higher resolution than the original micro-images [37]**.**

Because micro-images capture radiance from overlapping points in space, the desired angle of projection in a reconstruction can produce higher pixel densities, much higher than the density in the previously presented algorithm.

Superresolution

The output 2D reconstructed image will correspond to the central region of the Integral Projection output because it is where the radiance value density is highest [37]. Without this, the image would have under sampled regions on the borders.

In the cases where the angle of projection causes the pixel projections to not be equally spaced, a Gaussian function is used to normalize the output by weighting the pixels into the nearest pixel location, where the next perfect interleave will be. A perfect interleave is to be understood as when pixels are evenly spaced in the blending process. An illustration of this weighting process can be found in Figure 3.16. The process is divided in the following steps:

1. A series of Gaussian functions are applied to the micro-image, each centred at the closest missing sub-pixel position;
2. The extrapolated pixel values are calculated by summing all the contributions from each pixel.

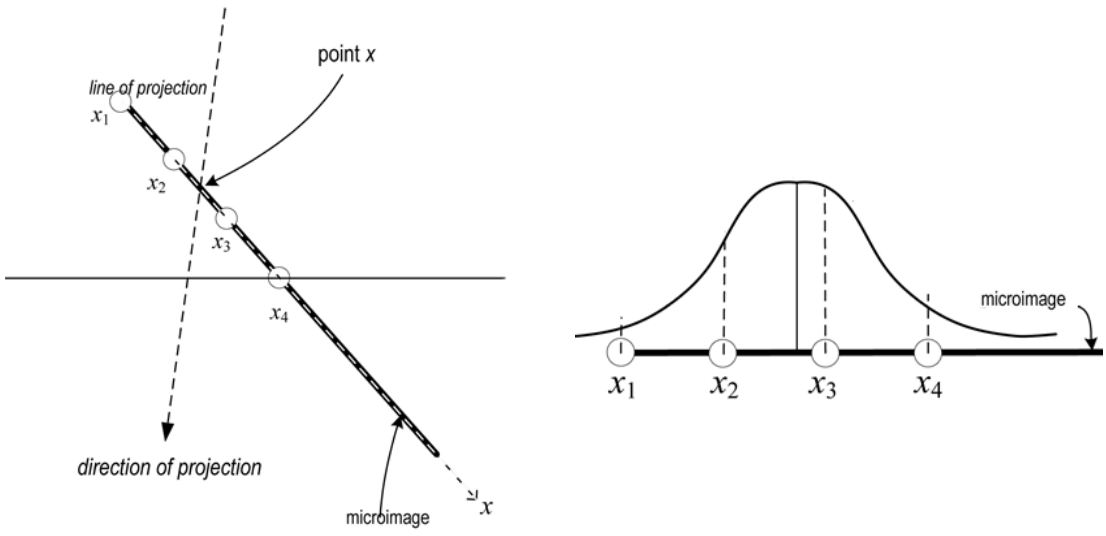


Figure 3.16 - Normalization of the interleaving process: left) projection plane and pixels of a micro-image that do not match with the position of the output pixel; right) weighting Gaussian function used to calculate the pixel value applied to the same micro-image [37]**.**

Performance Assessment

Figure 3.17 presents two 2D extracted image examples for performance assessment. In the left image, the background is in focus and the foreground is out of focus and slightly blurred; however, on the right image, the foreground is in focus and the background is out of focus and blurred. The blurring effect is a desired effect of the blending process resulting from grouping radiance values as if an actual picture was being taken, i.e. mixing together AoV that do not match for a given depth, just like what happens in a 2D traditional photographic machine.

An advantage of this algorithm is that depths not targeted to be in focus will become blurred instead of displaying sharp artefacts. A disadvantage is that the region brought into focus has to be chosen prior to the extraction process.

Figure 3.17 - Rendering with the single-sized patch blending based 2D image extraction algorithm: left) image rendered with a smaller patch size (7 pixels); right) image rendered with a larger patch size (10 pixels) [36]**.**

3.3 Depth Based 2D Image Extraction Solutions

The methods presented in this section address the 2D image extraction/reconstruction problem using not only texture data but also (estimated) depth or disparity data (assuming this data is not originally available).

3.3.1 Disparity Map Based 2D Image Extraction (DMe)

This 2D image extraction method was developed by Todor Georgiev and Andrew Lumsdaine and is fully described in [36].

Objectives and Basic Approach

This 2D reconstruction algorithm is an optimization of the method already presented in Section 3.2.3, now using also depth data. The objective of this method is to reconstruct an artefact-free, full-focused, 2D image from a holoscopic image with a patch size determined by the disparity map data. The disparity information extraction tool, applied to the holoscopic image, attempts to determine the patch size at which the information inside each micro image better matches its neighbour patches. This result in a multitude of patch sizes, with each set of identical patch sizes corresponding to objects at a specific depth in the scene. Since micro-lenses behave as pinhole cameras, the sharp representation of the objects already exists in the micro-images, and thus the different patch sizes simply do a better

job of piecing the information together. As a consequence of using multi-sized patches, the 2D final image reconstruction forces the magnification of some of the patches, so that they all match in size.

Architecture and Walkthrough

The architecture of this method is presented in Figure 3.18.

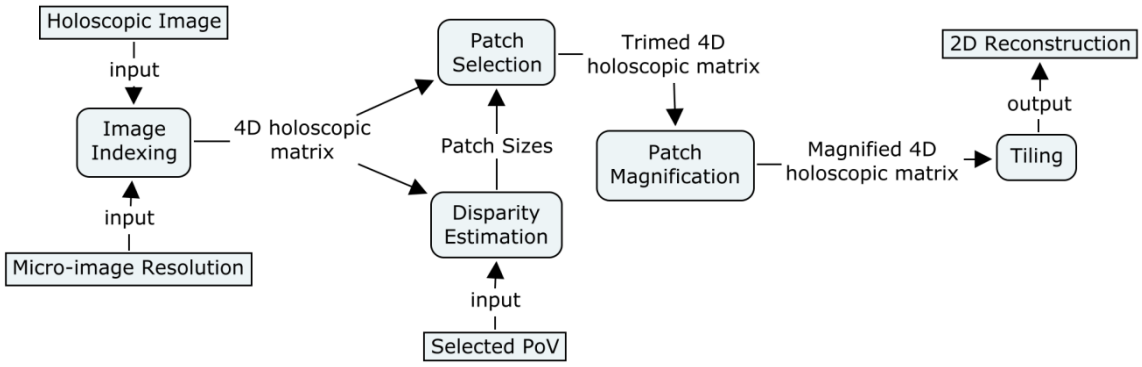


Figure 3.18 - Disparity map based 2D image extraction architecture.

The main processing modules of this method are:

1. **Image Indexing** – This module is the same as for the extraction solution in the previous section.
2. **Disparity Estimation** – The disparity between neighbouring micro-images is estimated to assert the best patch size to be used to reconstruct the final 2D image; the inputs and outputs are:
    a. Input Selected PoV – This parameter determines the position of the window where the disparity estimation algorithm will be looking for adequate patch sizes inside the micro-images;
    b. Output Patch Sizes – A patch size map describing the patch sizes that should be used for each micro-image in the reconstruction process, created from the estimated disparity map.
3. **Patch Selection** – This module is the same as for the 2D extraction solution presented in Section 3.2.3 with the difference that, instead of a single patch size for all micro-images, there is now a specific patch size for each micro-image; the inputs and outputs are:
    a. Input 4D holoscopic matrix – Output of the Image Indexing module above;
    b. Input Patch Sizes – Output of the Disparity Estimation module above;
    c. Output Trimmed 4D holoscopic matrix – Input 4D matrix with values outside the patches removed; notice that the patches do not have all the same size.
4. **Patch Magnification** – All the patches are interpolated to match the largest patch size selected in the previous module; the inputs and outputs are:
    a. Input trimmed 4D holoscopic matrix – Output of the previous Patch Selection module;
    b. Output magnified 4D holoscopic matrix – Input 4D holoscopic matrix with all patches upsampled to have the same size.
5. **Tiling** – The selected patches are arranged side-by-side, as if they were tiles, to reconstruct the final image; the inputs and outputs are:
    a. Input magnified 4D holoscopic matrix – Output of the previous module;

b. Output 2D image reconstruction - 2D light field representation of the original scene for a specific sub-PoV.

Main Tools

The main tool of this method is the Disparity Estimation module which enables selecting the local patch sizes. Although the patch magnification is also important, there is not much detail available in the literature.

Disparity Estimation

This section briefly describes the algorithm proposed in [36] to determine the patch size for each micro-image. Because micro-images inside a holoscopic image behave as different 2D PoV image captures of the scene, there are no significant vertical variations on horizontal neighbouring micro-images within a holoscopic image. This is also true for horizontal variations on micro-image columns. Although radiance values are not precisely the same as each one represents light coming from the same region of space from a different angle, they are very similar, almost the same, because the difference in angle is not very large in adjacent micro-images. This feature is exploited by the disparity estimation algorithm which basic principle is to find similar radiance patches in neighbouring micro-images, notably at very specific ranges supported by the horizontal and vertical variation constraints, to be able to pinpoint the same spatial regions in adjacent micro-images. This information is then used to extrapolate the disparity of objects representations among adjacent micro-images. Note that these disparity values can, and are, used as patch sizes because they essentially describe a distance from the centre of each micro-image, horizontally and vertically, within which there is a representation of a common spatial region in adjacent micro-images.

The algorithm proceeds with the following steps:

1. A *m*x*m* patch is selected at the centre of each micro-image (for the central PoV). The value of *m* is not defined, neither is described in any way in literature;
2. The best cross correlation between the *m*x*m* patch above and patches in all possible horizontal positions, at the same vertical position, on the micro-image to the left, referred to as K_x, is calculated;
3. The best cross correlation between the *m*x*m* patch above and patches in all possible vertical positions, at the same horizontal position, on the micro-image below, referred to as K_y, is calculated;
4. The K_x and K_y correlations are averaged to obtain the final value of K, which will correspond to the disparity value of the micro-image.

Figure 3.19 shows an illustration of the patch matching process.

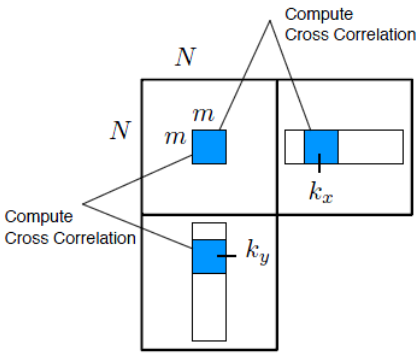


Figure 3.19 - Disparity estimation algorithm [36]**.**

This tool will generate a so-called *coarse grain depth map* [36], with each value representing a patch size.

Performance Assessment

An advantage of this method is presented in Figure 3.20 which shows an all-in-focus 2D reconstruction of the scene created using this method. Considering what has been said about holoscopic image geometry, in principle, this algorithm has the advantage of fully respecting the original depth of field and the focal position characterizing the original light field.

A drawback of this algorithm is that the reconstruction requires a dynamic upsampling of the patches to make them spatially matching, resulting in the localized introduction of errors depending on the ratio between each patch size and the maximum patch size.

Figure 3.20 – Disparity map based reconstructed image [36]**.**

Figure 3.21 shows a representation of the disparity estimation data used to generate the reconstructed image in Figure 3.20.

Figure 3.21 - Estimated disparity where the lighter regions correspond to the foreground (and thus larger patch sizes) [36]**.**

3.3.2 Depth Blending Based 2D Image Extraction (DBe)

This second depth-based 2D image extraction method has been proposed by Todor Georgiev and Andrew Lumsdaine and a full description is available in [38].

Objectives and Basic Approach

This extraction method is a combination of the two previously presented extraction methods, notably the blending and disparity map based methods, presented in Sections 3.2.4 and 3.3.1. The objective of this method is again to reconstruct an artefact-free, full-focused, 2D image from a holoscopic image using the depth estimated from the available radiance values using a depth map estimation tool. The depth information estimation tool, applied to the holoscopic image, attempts to determine the depth of each pixel in order to be able to place it at the correct location as determined by its depth, depending on the desired PoV. Reconstructing the 2D final image, respecting the depth of each pixel, may cause interleaving problems that are solved by the included Demosaicing module. This Demosaicing module aims to blend and interpolate the samples at colour component level.

Architecture and Walkthrough

The architecture of this method is presented in Figure 3.22.

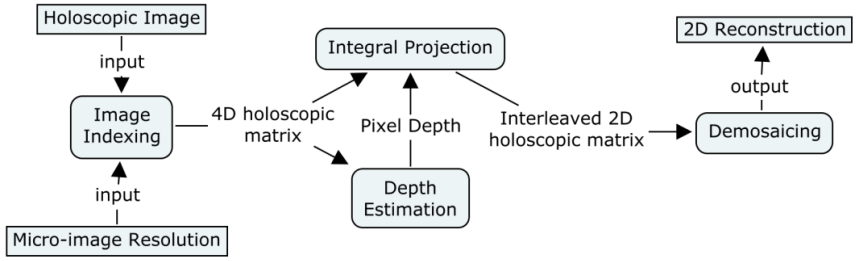


Figure 3.22 – Depth blending based 2D image extraction architecture.

The main processing modules of this method are:

1. **Image Indexing** – This module is the same as for the extraction solution in the previous section.
2. **Depth Estimation** – A multiview depth estimation algorithm based on *Graph Cuts* [39] is used to estimate the depth of each pixel from the captured radiance; the inputs and outputs are:
    a. Input 4D holoscopic matrix – Output of the previous module;
    b. Output Pixel Depth – A depth map describing the depth of the pixels in the captured radiance.
3. **Integral Projection** – This module is the same as for the 2D extraction solution in the SSPBe method.
4. **Demosaicing** – A similar process to the Superresolution module described in Section 3.2.4 is performed. However, this method includes an optimization step to minimize the noise created by Superresolution, called *mosaicing,* when it is applied alone. Instead of using luminance pixel values, the radiance information is processed by colour components (red, green and blue) separately; after, the various components are joined together in the reconstructed image; the inputs and outputs are:
    a. Input interleaved 2D holoscopic matrix – Output of the previous module;
    b. Output 2D Reconstruction - 2D light field representation of the original scene for a specific sub-PoV.

Main Tools

The main tool of this method is the Depth Estimation module, which should enable precise radiance information positioning based on depth.

Depth Estimation

In this section, the algorithm used to determine the depth of each pixel in the holoscopic image is described [38]. As each micro-image captures part of the scene from a different view point, multi-view depth estimation is feasible to retrieve the depth information from the entire captured radiance. This depth estimation algorithm iteratively computes the depth of each pixel based on Graph Cuts [39]. using the following steps:

Algorithm 1 Calculate depth of each pixel d_i

```
Require: Captured radiance r
  for j = 1 → M do
    for i = 1 → N do
      if p_i has been assigned with any depth value then
        Continue;
      end if
      Compute data term e_d of pixel p_i;
      if e_d > k_j then
        {Data term is bigger than current threshold}
        Continue;
      else
        Add p_i to graph
      end if
    end for
    Perform graph cuts;
    Assign depths to pixels in the graph;
  end for
```

Figure 3.23 - Depth estimation algorithm for holoscopic imaging [40]**.**

The algorithm is performed as follows:

1. Given a certain depth, the algorithm utilizes the variance of the corresponding pixels among micro-images as the data term, e_d.
2. One common problem with depth estimation is that, even though the correct depth is assigned to one pixel, the data term could still be large due to occlusions. This problem is severe because many views are involved in the computation. To resolve this issue, a gradually increasing confidence threshold, k_j, is adopted in each iteration, so that the pixels with lower depth will converge first and the pixels with larger depth may still generate data terms larger than the threshold, due to occlusion.
3. The algorithm covers all the *N* pixels, p_i , in each *M* micro-image, performing graph cuts for each micro-image.
4. Once the depth of one pixel is decided, it will not be involved in the computation anymore. Thus, the pixels with larger depth will avoid impacting the lower depth pixels and will produce a low data term if assigned the correct depth in the later iterations.

Performance Assessment

This method is relatively recent and not much information is available to accurately assess its performance, although it is interesting enough to be mentioned because it includes tools to remove artefacts that typically come about in the reconstruction process. An advantage of this method is the fact that it produces less extraction artefacts due to the independent processing of the colour components. Another advantage is that, according to the conclusions presented in the literature [38], the results are “much prettier” than the previous methods.

A first drawback of this method is that no mention to PoV selection is made in the literature. Another drawback is that a depth map is only good for a single PoV so extending it to multiple PoVs will require multiple Depth Map calculations.

4
The Proposed Disparity-Assisted Patch Blending 2D Extraction Algorithm

The proposed Disparity-Assisted Patch Bending 2D extraction (DAPBe) algorithm is presented in this chapter. The basic idea of this new algorithm is to extract 2D images from holoscopic images without the need to:

- Specify to the algorithm a particular plane of the scene that will be in focus in the extraction method. This algorithm generates images with the same depth of field of each micro-lens, typically a large depth of field, generating All-in-Focus 2D images;
- Manually improve the depth estimations calculated from the holoscopic image to improve the perceived quality of the final result. This algorithm is a fully automated solution to generate All-in-Focus 2D extractions from a holoscopic image.

First, the overall architecture is presented and the high-level processes behind the proposed method are briefly described. After, a detailed presentation of each of the modules involved in the extraction of All-in-Focus 2D images provided by this method is made.

4.1 Architecture and Walkthrough

The overall architecture of the image extraction method proposed in this chapter is presented in Figure 4.1.

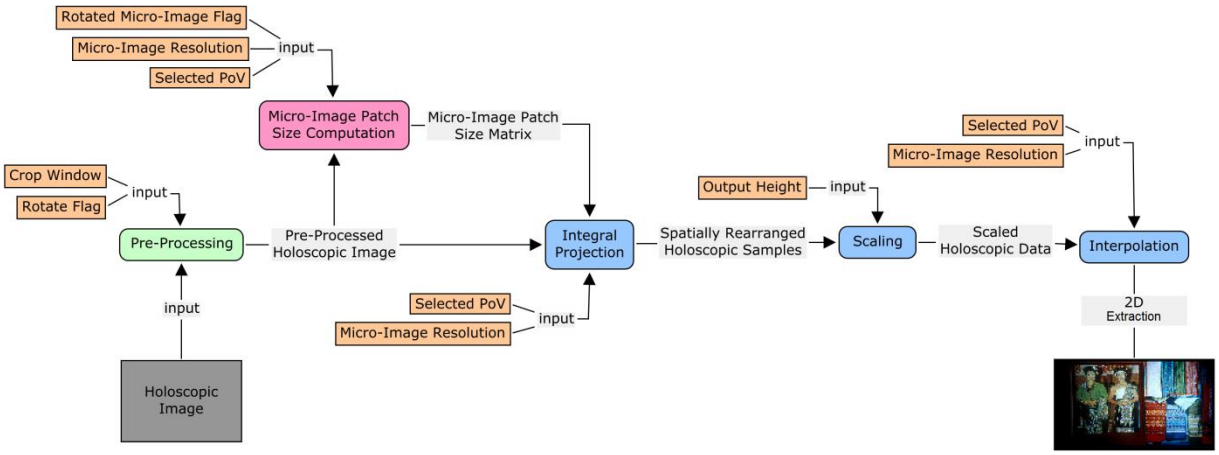


Figure 4.1 - Disparity-Assisted Patch Blending 2D image extraction architecture. Orange boxes represent inputs, while blue and pink boxes represent extraction modules with the pink boxes specifically related to disparity estimation.

The walkthrough of the proposed image extraction method is presented below. For each module, a brief description of the problem addressed as well as its major objectives and output(s) are also provided.

1. **Pre-Processing** – The problem this module tackles, e.g. micro-images rotation and misalignment, occurs due to the characteristics of some optical setups used in the acquisition process. For example:
    - Holoscopic images may exhibit a rotation of 180 degrees meaning that, depending on the relative position of the main lens focal plane relatively to the micro-lens array (see Chapter 2), the micro-images may appear inverted in the sensor plane;
    - Some micro-images may be incomplete at the borders of the holoscopic image due to some misalignments between the micro-lens array and the capturing sensor.

    Thus, the objective of this module is to align some of the most popular types of holoscopic data to be ready for processing with the proposed image extraction architecture since the proposed architecture expects complete non-rotated micro-images as input.

    The micro-images rotation problem is resolved by mirroring the samples over vertical and horizontal central lines. To ensure that only complete micro-images are considered in the following processing modules, a crop operation of the holoscopic image is performed around the image borders. The cropping is done manually according to the *crop window* input parameter. This value is one of the specifications of a holoscopic camera. It typically varies among different models and because there is no automatic means of detecting it, in this algorithm, it needs to be provided.

    The inputs and outputs of this module are the following:

    a. Input Holoscopic Image – Original holoscopic image;
    b. Input Cropping Window – Set of 4 values, each representing the distance in samples from the up, down, left and right limits of the holoscopic image, defining the cropping window;
    c. Input Rotate Flag – Binary flag indicating if the holoscopic image needs to be rotated or not;

d. Output Pre-Processed Holoscopic Image – The cropped holoscopic image containing only complete micro-images rotated in order to have the scene oriented upwards.

2. **Micro-Image Patch Size Computation** – The problem this module tackles is the definition of micro-image patch sizes to guide the extraction process. Patch sizes are directly related to the extraction plane depth and directly related to the disparity of the central region of each micro-image regarding neighbouring micro-images. This disparity manifests horizontally in the left and right adjacent micro-images and vertically in the up and down adjacent micro-images.

   The objective of this module is to guide the Integral Projection module in the process of spatially arranging the holoscopic image samples according to their relative scene depth.

   To resolve the problem, the pre-processed image undergoes, first a sampling step - **Micro-Image Sampling** - at micro-image level, where central square portions, with different sizes, are gathered from each micro-image, followed by a disparity estimation step - **Disparity Estimation** - where a search process tries to find the best match in neighbouring micro-images for the central square portions sampled from each micro-image.

   Micro-Image Sampling: The sampling is performed on the centre of each micro-image, where there is a representation of a *square portion A* of the scene, and on regions of the neighbours where it is very likely to find representations of that same region *A*; this happens when the selected PoV input points to the centre of the micro-images. If the selected PoV input points to another position (e.g., when the final objective is to extract a stereo pair, two different PoVs are defined, one for the left view and another for the right view) of the micro-images, then the sampling is performed in relation to that position and region representations in neighbouring micro-images.

   Disparity Estimation: After the square portions have been sampled for each micro-image, a search process is performed to determine the disparity of the region representations within each micro-image, as it relates to their depth in the captured scene.

   Using a Normalized Cross-Correlation criterion, for each micro-image, several square portions of the neighbours (the search region) are compared to the sampled *square portion A*, essentially searching for the *square portion A* in neighbouring micro-images. This disparity data if further processed to obtain a single value of disparity for each micro-image; thus disparity value expresses the spatial disparity, in samples, of the region represented in the centre of a micro-image and its neighbouring micro-images. The process of searching and subsequent processing of the data resulting from the search process, which results in the patch size value for each micro-image, will be described in more detail in the following section.

   The inputs and outputs of this module are the following:

   a. Input Pre-Processed Holoscopic Image – Output of the previous module which will be used as the source data for the sampling process;
   b. Input Micro-Image Resolution – Two values, height and width, each is representing the vertical and horizontal dimensions in samples of the micro-images. This input is used to delimit the micro-images when gathering samples for the search process;
   c. Input Selected PoV – A two dimensional coordinate pinpointing a position inside the micro-image and determining the region of the micro-images where the disparity estimation is performed; the selected PoV defines the centre of this region;

d. Input Rotated Micro-Image Flag – Binary flag responsible for informing the module if the image is rotated or not.
e. Output Micro-Image Patch Size Matrix – A patch size matrix defining the square patch size to be used for each micro-image in the extraction process.

3. **Integral Projection** – The problem this module tackles is the spatial arrangement of the 4D holoscopic samples (see Section 2.1) into a simplified 2D image representation. Originally, the holoscopic samples have 4D spatial information, i.e., (y, z, φ, θ). In this module, the 4D spatial information is transformed into a 2D spatial representation by selecting the appropriate information as the output of this module.

   The objective of this module is to transform the 4D holoscopic samples into a "friendly" representation to the HVS and the display at hand, i.e., a 2D representation.

   To resolve the aforementioned problem, each micro-image is projected onto a 2D projection plane which can be abstracted as the plane where the 2D extracted image will be assembled from the available samples. The *Micro-Image Patch Size Matrix* input determines the size of a square region of samples centred on the Selected PoV of each micro-image that will correspond to the original size of a micro-image, on the projection plane, i.e. scaling up the micro-images through the projection process. The *Selected PoV* input determines the horizontal and vertical displacements of the projections, in the 2D projection plane, of each micro-image projection. The amount of displacement depends on the amount of scaling that was done in the projection, i.e., each micro-image is displaced proportionally to the distance of its projected samples; otherwise, all micro-image projections would be displaced by the same amount. This displacement can be seen as moving a real 2D camera horizontally or vertically, depending on the displacement direction, while still pointing it to a certain position in the scene. Because of the way the micro-images are scaled, they no longer occupy neatly defined positions in a matrix. Therefore, alternative means of storage need to be considered, which is covered in the next section.

   The inputs and outputs of this module are the following:

   a. Input Pre-Processed Holoscopic Image – A holoscopic image containing only complete micro-images, with the top of the scene oriented upwards; due to the structure of this input data, it can feed this module with holoscopic samples and also their position in the holoscopic image;
   b. Input Micro-Image Patch Size Matrix – A patch size matrix defining the patch size to be used for each micro-image in the projection process.
   c. Input Micro-Image Resolution – Two values, height and width, representing the micro-images dimensions; this input is used to delimit micro-images when asserting to which micro-image a sample belongs to;
   d. Input Selected PoV – A two dimensional coordinate pinpointing a sample inside the micro-images. This sample results from the exposure to light from a particular angle (determined by camera optics). All these samples together (from all micro-images), makeup a scene representation from a particular PoV (as covered in the previous chapter.
   e. Output Spatially Rearranged Holoscopic Samples – Holoscopic samples and corresponding 2D coordinates on the 2D projection plane that forms a 2D representation of the captured scene, for the selected PoV. Because the samples may no longer have integer positions, they are no longer in matrix form but rather in an array form where each

position has a sample, its new position in 2D space, and also the position of the sample inside the source micro-image, displaced by the *Selected PoV*.

4. **Scaling** – The problem this module tackles is to output a 2D image with a given spatial resolution out of the Rearranged Holoscopic Samples array produced by the previous module.

   The objective is to provide flexibility in the output resolution, while preserving scene proportions by respecting the ratio between horizontal and vertical sizes.

   To resolve the problem, the spatial information of the samples is scaled proportionally to the ratio of the *Pre-Processed Holoscopic Image*. Due to the ratio constraint, the final dimension of the output has to be set in relation to the width or height. Here, the height was chosen to facilitate compatibility with line resolution of a given display system; then the width is calculated by multiplying the ratio of the *Pre-Processed Holoscopic Image* with the *Output Height*.

   The inputs and outputs of this module are the following:

   a. Input Spatially Rearranged Holoscopic Samples – The output of the previous module.
   b. Input Output Height – The height of the output 2D reconstructed image; the width is determined by multiplying the ratio of the Pre-Processed Holoscopic image with this input value;
   c. Output Scaled Holoscopic Data – The scaled version of the *Spatially Rearranged Holoscopic Samples* input.

5. **Interpolation** - The problem this module tackles is to compute the 2D image samples for the positions of the output image matrix, the *2D Extraction*. In a "traditional" 2D image, image samples are arranged in a matrix fashion (regularly spaced). Although the samples composing the *Scaled Holoscopic Data* input are arranged in a 2D space, they are not regularly spaced, which means that some positions in the output image matrix are empty.

   The objective of addressing this issue is to create a full 2D image representation of the scene, i.e., a representation of the scene where samples are uniformly distributed in a matrix arrangement. To resolve the problem, an empty image is created to be populated by processing the information in the *Scaled Holoscopic Data* input. To that end, the input data is laid on top of the empty image to determine which micro-images lay on top of which empty sample positions. For each micro-image lying on top of a sample position, a sample is interpolated, from each micro-image, to that sample position. Further details on the interpolation method used in this algorithm will be covered in the next section. After all interpolated values are calculated, they are blended in a weighted average operation. The weight of each sample is determined through a weighting function following a Gaussian distribution. The samples closer to the displaced (by the Selected PoV input) centre of the corresponding micro-image have higher value than values far from the displaced centre of the corresponding micro-image. This is because, as covered in Chapter 2, the further from the displaced centre samples are, the less relevant to the selected PoV they are. Further details of this process will be covered in the next section.

   The inputs and outputs of this module are the following:

   a. Input Scaled Holoscopic Data – The output of the previous module, this means an incomplete 2D representation of the scene, where some sample positions are empty;
   b. Output 2D Extraction – A complete 2D image representing the scene from a specific PoV with a specific spatial resolution.

In the next section, the details on the modules above will be presented.

4.2 Detailed Module Descriptions

In this section, the modules in the proposed 2D extraction algorithm will be described in greater detail, with a sub-section reserved for each module. Higher complexity modules, however, may require additional sub-division.

4.2.1 Pre-Processing Module

This module consists of two independent sub-modules, applied sequentially to the data in the input holoscopic image.

4.2.1.1 Cropping the Image

The first sub-module of the *Pre-Processing* module has the target to crop the input holoscopic image. The objective of this operation is to get rid of all the samples belonging to incomplete micro-images, resulting in a holoscopic image where only complete micro-images remain.

The inputs and outputs of this sub-module are the following:

a. Input Holoscopic Image – Original Holoscopic Image is composed of samples, $I(x,y)$, with $0 \leq x \leq M'-1$ and $0 \leq y \leq N'-1$, where the x axis represents the horizontal direction of the image and the y axis represents the vertical direction of the image;

b. Input Crop Window – Set of 4 values, d_y^U, d_y^D, d_x^L, d_x^R, defining the distance in samples from the up, down, left and right limits of the holoscopic image, defining the cropping window;

c. Output Cropped Holoscopic Image – Holoscopic image, $I_{crop}(x,y)$, containing only complete micro-images.

Given an input holoscopic image, $I\ (x,y)$, with horizontal and vertical dimensions, respectively, *M*' and *N*", the cropped image, $I_{crop}(x,y)$, is a subset of $I\ (x,y)$ as defined in Equation (7)

$$I_{crop}(x,y) = \{I(x,y) : x_{min} \leq x \leq x_{max} \wedge y_{min} \leq x \leq y_{max}\} \qquad (7)$$

where $\left(x_{min} = d_x^L, y_{min} = d_y^U,\right)$ and $\left(x_{max} = M' - 1 - d_x^R, y_{max} = N' - 1 - d_y^D\right)$ define the limits of the cropping window.

The processing in this operation consists in creating a window with the size of the holoscopic image resolution and subsequently eliminating the values outside the *Cropping Window*. All samples outside the window are discarded, the remaining compose the *Cropped Holoscopic Image* output.

4.2.1.2 Rotating the image

The second pre-processing sub-module rotates each micro-image of the cropped holoscopic image, if needed. The objective of this operation is to give the holoscopic image upright horizontal and vertical orientation, prior to the extraction process. This is done by rotating the image, thereby placing left, right, up and down regions of the scene in the left, right, up and down regions of the holoscopic image.

The inputs and outputs of this sub-module are the following:

a. Input Cropped Holoscopic Image – Holoscopic image composed of complete micro-images, $I_{crop}(x,y)$[10];

[10] From this point onwards, image coordinates are referred relatively to the cropped image unless otherwise stated.

b. Input Rotate Flag – Binary flag indicating if the holoscopic image needs to be rotated by 180º or not;
c. Output Pre-Processed Holoscopic Image – Cropped holoscopic image appropriately rotated to have the scene oriented upwards.

Given an input holoscopic image, $I_{crop}(x, y)$, with horizontal and vertical dimensions, respectively M and N in sample units, the rotated image, $I_{rot}(x, y)$, is defined in function of $\boldsymbol{I_{crop}}$ as described in Equation (8)

$$\boldsymbol{I_{rot}(x, y) = I_{crop}(M - 1 - x,\ N - 1 - y)} \qquad (8)$$

where $0 \leq x \leq M_m - 1$ and $0 \leq y \leq N_m - 1$.

This operation consists in mirroring the samples present in the holoscopic image, both horizontally and vertically, in relation to the centre of the holoscopic image, achieving a 180º rotation.

4.2.2 Micro-Image Patch Size Computation Module

This module consists of three independent sub-modules, applied sequentially to the input data, with the purpose of computing each micro-image patch size.

4.2.2.1 Computing Scene Elements Similarity

The objective of computing the scene elements displacement is to gather disparity data to feed the following modules. This data characterizes the displacement of the regions, from a micro-image point of view, relating to their depth (closer regions have higher displacements than farther regions). Therefore, this module essentially gathers the data needed for the following modules to determine the depth of the central region in each micro-image.

The input and output of this operation are:

a. Input Pre-Processed Holoscopic Image – Output of the previous module to be used as a source of data for the sampling process;
b. Input Micro-Image Resolution – Two values representing the micro-image width and height dimensions, respectively, M_m and N_m, to be used to delimit micro-images and gather samples for the search process that happens in this module;
c. Input Selected PoV – A two dimensional coordinate, (x_{pov}, y_{pov}), pinpointing a position inside all micro-images, where $-\frac{M_m}{2} \leq x_{pov} \leq \frac{M_m}{2}$ and $-\frac{M_m}{2} \leq y_{pov} \leq \frac{M_m}{2}$. This value determines the region of the micro-images where spatial redundancy is sought. The value $(0,0)$ references the centre of the micro-images;
d. Input Rotated Micro-Image Flag – Binary flag responsible for indicating if micro-images are rotated 180º or not.
e. Output *Similarity Map* – Structure containing the results of a series of similarity tests performed among micro-images. The information is encapsulated in the forward described $Sim(k, l, r, q)$ function, $0 \leq k \leq M_{mla}$, $0 \leq l \leq N_{mla}$, $0 \leq r \leq 15$ and $0 \leq q \leq \frac{M_m}{2} - p$, where $M_{mla} = M/M_m$ and $N_{mla} = N/N_m$. The input parameters are, respectively, a horizontal and vertical position in the MLA in micro-lens units, a r sequential number of a particular disparity test conducted in the (k, l) micro-image and a sequential number q referencing a region of

the neighbour micro-image $(k \mp 1, l \mp 1)$ analysed in test r. For each micro-image, this structure stores r series of Normalized Cross-Correlation[11] (NCC) values as a function of disparity q.

For each micro-image, $I_m^R(i,j)$, called *reference micro-image* (see yellow region in Figure 4.2), a spatial redundancy search operation is conducted in several adjacent micro-images, called *search micro-images*, $I_m^{S_\alpha}(i,j)$. This is performed with the intent of finding regions of samples that closely match each other among neighbouring micro-images.

The spatial redundancy searches are conducted in the entire image with each search proceeding as follows:

- Each search operation is composed of 16 individual spatial redundancy lookups[12] r, where $r \in \mathbb{N}: 0 \leq r \leq 15$;
- The 16 lookups are conducted in 4 (marked pink in Figure 4.2) different *search micro-image*s $I_m^{S_\alpha}(i,j)$, where $\alpha \in \{up, down, left, right\}$;
- In each *search micro-image*, 4 lookups are performed inside a *Patch Search Regions* (see purple, yellow, green and blue overlapping regions inside the 4 pink regions in Figure 4.2);
- In each of the 4 lookups, the *Search Template Patch* is slid across the same colour coded *Patch Search Region*, from the centre to an edge, while computing for each position the similarities between both.

An illustration of the search process described above is presented in Figure 4.2.

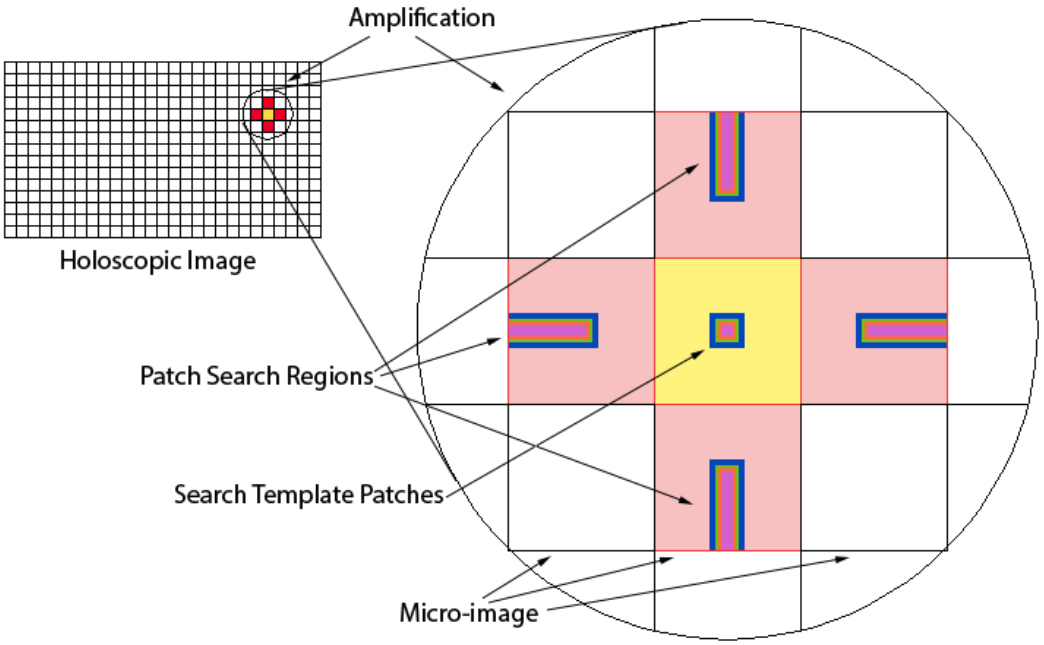


Figure 4.2 - Sampling with a 4-way disparity search pattern

The *Search Template Patches*, $P^T(x,y)$, must be placed at the centre of $I_m^R(i,j)$, offset by x_{pov} and y_{pov}, and limited by $x,y \in \mathbb{N}: 0 \leq x,y \leq p$, where $p \in \{7, 11, 15, 23\}$. The number of patches $P^T(x,y)$ used as *Search Template Patches* has been determined experimentally and can be varied. Increasing

[11] Forward described in Equation (9).
[12] Operations that methodically look for similarity between two different micro-images, in specific regions. This concept is explored further in this step.

the number of $P^T(x,y)$ increases the output quality and the output time of the algorithm; the difference in processing time for the increase in quality does not seem to compensate.

Decreasing the number of $P^T(x,y)$ leads to a dramatic decrease of the output quality and slight decrease of output time of the algorithm; the quality drop is considerable enough to justify an increase in processing time by adding another patch $P^T(x,y)$.

The number of luminance samples used for the *Search Template Patches* has been determined experimentally and can be varied. The values chosen for p represent the range where consistent results appear; lower values tend to result in unreliable (very high variance) data while high values tend to result in inconclusive (very low variance) data.

Only adjacent micro-images are used as *search micro-images* as going further would greatly increase the complexity and present gains only in cases where high spatial redundancies exist, i.e. adjacent micro-images represent nearly the same region of a scene. Highly redundant adjacent micro-images are engineered out in the capture process to ensure higher resolution output in 2D extractions [36].

The *Patch Search Regions* $R^{S_{up,down}}(x,y)$ are limited by $0 \leq x \leq p,\ 0 \leq y \leq \frac{M_m+p}{2} - 1 - \overline{y_{pov}}$ and $x,y,p \in \mathbb{N}$. The *Patch Search Regions* $R^{S_{left,right}}(x,y)$ are limited by $0 \leq x \leq \frac{M_m+p}{2} - 1 - \overline{x_{pov}}$, $0 \leq y \leq p$ and $x,y,p \in \mathbb{N}$. They are rectangular, starting from the centre of the $I_m^{S_x}(i,j)$, offset by x_{pov}, y_{pov}, and extending to one of its edges:

- When Rotated Micro-Image Flag is active, the $R^{S_x}(x,y)$ extends from the centre of $I_m^{S_x}(i,j)$ until its farthest edge from $I_m^R(i,j)$;
- When Rotated Micro-Image Flag is active, the $R^{S_x}(x,y)$ extends from the centre of $I_m^{S_x}(i,j)$ until its closest edge from $I_m^R(i,j)$.

When a *Search Template Patch* $P^T(x,y)$ is being slid across a same colour coded *Patch Search Region* $R^{S_x}(x,y)$, it can only occupy one of q given positions inside, where $q \in \mathbb{N}: 0 \leq q \leq \frac{M_m}{2} - p - \max(\overline{x_{pov}}, \overline{y_{pov}})$. In each q position, $P^T(x,y)$ is overlapping a same sized patch $P^S(x,y)$ from the *Patch Search Region* $R^{S_x}(x,y)$. To reference all the same sized patches $P^T(x,y)$ inside the *Patch Search Region* $R^{S_x}(x,y)$, they are abstracted as $P^S(q,x,y)$, where q specifies the position of $P^T(x,y)$.

Note that the search is only done horizontally for $I_m^{S_{left,right}}(i,j)$ and only vertically for $I_m^{S_{up,down}}(i,j)$ as these are the directions from which disparity manifests in each of those cases.

For each of the 16 individual spatial redundancy lookups, similarity values are calculated for q. Because the similarities are calculated for several positions $P^S(q,x,y)$ inside $R^{S_x}(x,y)$ to find a match, the position q where that match is found relates to the disparity of $P^T(x,y)$ in any $I_m^{S_x}(i,j)$. To assure a direct connection between q and the disparity, for all lookups, the central position of $R^{S_x}(x,y)$ must correspond to zero disparity, $q = 0$, and the limit edge position corresponds to the maximum value of disparity, $q = \frac{M_m}{2} - p - \max(\overline{x_{pov}}, \overline{y_{pov}})$.

The similarities between $P^T(x,y)$ and a $P^S(q,x,y)$ are measured using the NCC metric, shown in Equation (9), which uses the mean, shown in Equation (10) and standard deviation, shown in Equation (11).

$$\boldsymbol{NCC(P^T, P^S) = \frac{1}{n}\sum_{x,y} \frac{(P^T(x,y) - \overline{P^T})(P^S(x,y) - \overline{P^S})}{\sigma(P^T)\sigma(P^S)}, \forall q \in \mathbb{N}: 0 \leq q \leq \frac{M_m}{2} - p} \qquad (9)$$

$$\boldsymbol{\overline{X} = E[X] = \sum_{i=0}^{n-1} \frac{x_i}{n}} \qquad (10)$$

$$\boldsymbol{\sigma(X) = \sqrt{Var[X]} = \sqrt{\frac{\sum_{i=0}^{n-1}(x_i - \bar{x})^2}{n-1}}} \qquad (11)$$

The NCC was picked for this application for the following reasons: i) although the NCC is a computationally complex algorithm to assess similarities between images, it makes a good trade-off with performance in this application, as determined by testing; and ii) the normalized version of the cross-correlation compensates for different lighting and exposure conditions, in image processing, a key functional element in this application where lighting conditions change from one micro-image to the other.

The result of calculating $NCC(P^T, P^S)$ for all q positions, in all r lookups, can be obtained through the Similarity function $Sim(r,q)$, presented in Equation (12). $P^T(r)$ represents a $P^T(r)$ for a particular lookup r and $P^S(r,q)$ represents a $P^S(x,y)$ for a particular disparity q inside a lookup r.

$$\boldsymbol{Sim(r,q) = NCC\left(P^T(r), P^S(r,q)\right) = \frac{1}{n}\sum_{x,y}\frac{(P^T(r,x,y) - \overline{P^T(r)})(P^S(r,q,x,y) - \overline{P^S(r,q)})}{\sigma(P^T(r))\sigma(P^S(r,q))}} \qquad (12)$$

The $Sim(r,q)$ function, for a given r, becomes a simple series of NCC values as a function of the disparity q. Within this framework of similarities search, this series is considered to be inconsistent when it starts as a decreasing function and/or it ends as an increasing function:

1. The first case indicates that the centres of all micro-images are highly correlated, which makes no sense in holoscopic images as all micro-images would essentially be identical, which is impossible because each micro-lens cannot capture the scene from the same PoV of their neighbours.
2. The second case indicates that the centre of the micro-images highly correlate with an erred area, i.e. the damaged edges of neighbouring micro-images, more than with values not so much affected by errors.

The output <u>*Similarity Map*</u> is the result of applying the $Sim(r,q)$ function for every micro-image in the pre-processed holoscopic image, becoming $Sim(k,l,r,q)$ where $0 \leq k \leq M_{mla}$, $0 \leq l \leq N_{mla}$, $0 \leq r \leq 15$ and $0 \leq q \leq \frac{M_m}{2} - p$, with $M_{mla} = M/M_m$ and $N_{mla} = N/N_m$.

4.2.2.2 Calculating Disparity

The objective of this operation is to gain enough knowledge on the scene geometry, in the form of disparity values, to be able to guide the extraction method. As will be seen ahead, patch size values are the critical data to be used in the 2D extraction.

The input and output of this operation are:

a. Input <u>*Similarity Map*</u> – A structure containing the results of a series of similarity tests performed among all micro-images present in the pre-processed holoscopic image. This information is encapsulated in the previously described $Sim(k,l,r,q)$ function;
b. Output <u>Spatial Similarity Estimator Map</u> – A series of disparity estimators, one for each micro-image, encapsulated in a forward described $\hat{\theta}(k,l,q)$ function, where (k,l) pinpoints a micro-image in the pre-processed holoscopic image, q is a particular value of disparity and the function assumes values with the same characteristics of the input $Sim(k,l,r,q)$ function. Generally speaking, each disparity estimator is the result of combining several NCC functions into one general NCC function;

c. Output Disparity Matrix – A forward described matrix $Disp(k,l)$ containing estimated disparity values, one for each micro-image, where (k,l) pinpoint a micro-image in the pre-processed holoscopic image and the matrix contains a disparity value. The disparity values are calculated using the disparity estimator proposed in this module.

The processing performed in this operation consists in building, as much as possible, efficient[13] and unbiased[14] disparity estimators to produce unavailable disparity values. Following, these estimators are used to calculate a disparity map.

The efficiency of an estimator is measured with the formula presented in Equation (13), with the help of the Fisher information formula presented in Equation (14). This estimator measures if the actual parameter θ varies as the estimator $\hat{\theta}$ varies.

$$\boldsymbol{e}\left(\widehat{\boldsymbol{\theta}}\right) = \frac{\mathbf{1}}{\boldsymbol{\mathcal{L}(\theta)}} / \boldsymbol{\sigma(\hat{\theta})} \qquad (13)$$

$$\boldsymbol{\mathcal{L}(\theta)} = \boldsymbol{E}\left[\left(\frac{\partial}{\partial \boldsymbol{\theta}}\log \boldsymbol{f}(\mathbf{X};\boldsymbol{\theta})\right)^{2}\middle|\boldsymbol{\theta}\right] = \int\left(\frac{\partial}{\partial \boldsymbol{\theta}}\log \boldsymbol{f}(\mathbf{x};\boldsymbol{\theta})\right)^{2}\boldsymbol{f}(\boldsymbol{x};\boldsymbol{\theta})\,\boldsymbol{dx} \qquad (14)$$

The Fisher information measures the entropy of the given parameter θ; in this case, θ is the disparity of the central portion of a micro-image, off-set for a given PoV, in relation to its neighbour micro-images. Calculating $\mathcal{L}(\theta)$ requires knowledge of the actual disparity θ or an estimate (which is available) of it. It also requires the probability distribution function $f(\mathrm{X};\theta)$ of θ, which is typically studied or assumed. Because this algorithm aims to perform disparity estimation for all cases, nothing can be assumed about $f(\mathrm{X};\theta)$ without compromising in some way the performance of $\hat{\theta}$. In other words, maximum entropy is always assumed to virtually maximize the efficiency of the estimator.

The bias of an estimator $\hat{\theta}$ is calculated by checking its alignment with the actual parameter θ or its measurements. In other words, an estimator $\hat{\theta}$ is unbiased if its expected value $E[\hat{\theta}]$, or its mean value, matches the expected value $\bar{\theta}$, or the actual mean value, of the actual parameter θ (see Equation (15)).

$$\boldsymbol{Bias}\left[\widehat{\boldsymbol{\theta}}\right] = \boldsymbol{E}\left[\widehat{\boldsymbol{\theta}}\right] - \overline{\boldsymbol{\theta}} \qquad (15)$$

The bias may be calculated for the estimators being built. Moreover, the estimators are built backwards to ensure $Bias[\hat{\theta}] \approx 0$. Estimators are built by averaging values of $Sim(k,l,r,q)$, which are the closest measure of disparity available, while assuring $E[\hat{\theta}] \approx \bar{\theta}$.

The estimators for each micro-image are built through Equation (16). This equation, for a particular micro-image (k,l), returns an estimated NCC function value for a particular q disparity common to all adjacent micro-images. This virtual NCC function value is calculated by averaging all r NCC function values for the same spatial position q in the $Sim(k,l,n,q)$ similarity map. The function $\hat{\theta}(k,l,q)$ essentially merges all r NCC functions into one single/virtual spatial similarity function referred as the Spatial Similarity Estimator Map.

$$\widehat{\boldsymbol{\theta}}(\boldsymbol{k},\boldsymbol{l},\boldsymbol{q}) = \frac{\sum_{r=0}^{15}\boldsymbol{Sim}(\boldsymbol{k},\boldsymbol{l},\boldsymbol{r},\boldsymbol{q})}{\mathbf{16}} \qquad (16)$$

[13] Efficient Estimator is an estimator that estimates the quantity of interest in some "best possible" manner.

[14] Bias of an Estimator is the difference between this estimator's expected value and the true value of the parameter being estimated.

This solution is based on the theory stating that the average of a unknown sampled parameter delivers the best estimation performance when nothing can be said about a real data distribution of that unknown sampled parameter [41]. This theory was also experimentally tested for this case in particular by comparing the performance with median built estimators which resulted in higher error.
The $\hat{\theta}(k,l,q)$ function is then used to calculate a Disparity Matrix $Disp(k,l)$ as described in Equation (17).

$$\boldsymbol{Disp(k,l) = pos\left(\widehat{\boldsymbol{\theta}}(\boldsymbol{k},\boldsymbol{l},\boldsymbol{q}),\ \max\left(\widehat{\boldsymbol{\theta}}(\boldsymbol{k},\boldsymbol{l},\boldsymbol{q})\right)\right) = D_{k,l}} \qquad (17)$$

$Disp(k,l)$ returns the position q of the maximum value of $\hat{\theta}(k,l,q)$ for a particular micro-image (k,l).

4.2.2.3 Optimizing the Patch Sizes

The objective of this operation is the elimination of visual artefacts, caused by a big deviation between real disparity values and the disparity values estimated in the previous step, by choosing patch sizes that have both some statistical support and minimize artefacts. To obtain patch size PS values from disparity D values, no calculations are required as $PS_{k,l} = D_{k,l}$.
The input and output of this operation are:

a. Input Spatial Similarity Estimator Map – A series of disparity estimators, one for each micro-image, encapsulated in the previously described $\hat{\theta}(k,l,q)$ function. Generally speaking, each disparity estimator is the result of combining several NCC functions into one general/virtual NCC function;
b. Input Disparity Matrix – The previously described matrix $Disp(k,l)$ containing a disparity value for each micro-image.
c. Output Micro-Image Patch Size Matrix – A patch size matrix, $Patch(k,l)$, where (k,l) pinpoints a micro-image in the pre-processed holoscopic image and the matrix includes the patch size values to be used for each micro-image in the remainder of the extraction process.

The processing done in this operation consists in detecting potential estimation errors, i.e. outliers, in the estimated disparity values, $D_{k,l}$. When an adjustment is required, the "magnitude" of the estimation error determines the nature of the adjustment. After this stage of statistical optimization, the disparity values are converted into patch size values through a direct conversion.
To detect the necessity for adjustments, two *statistical models*[15] characterizing the relation between each pair of estimated disparity values $D_{k',l'}$ in $Disp(k,l)$ are created:

i. **Global model (*Gm*)** – This model is created based on global statistics of the input $Disp(k,l)$. By finding the global mean $\overline{D_{k,l}}$: $k \in [0,\ M_{mla}] \wedge l \in [0,\ N_{mla}]$ and global standard deviation $\sigma(D_{k,l})$: $k \in [0,\ M_{mla}] \wedge l \in [0,\ N_{mla}]$ of all disparity values of $D_{k,l}$, this model states, $D_{k',l'} \in Gm = \left[\overline{D_{k,l}} - \sigma(D_{k,l}),\ \overline{D_{k,l}} + \sigma(D_{k,l})\right] \wedge k \in [0,\ M_{mla}] \wedge l \in [0,\ N_{mla}]$, meaning the estimated disparity values have to be between the global disparity mean, plus or minus, the global disparity standard deviation.
ii. **Local plus Global model (*LGm*) (local model with global statistics adjustment)** – This model is created based on both the Gm's global statistics and the local statistics of a particular

[15] A *statistical model* is the formal mathematical description of the relationship between variables in statistics.

$D_{k',l'}$. By finding the local mean of the 8 adjacent disparity values $\overline{D_{k'',l''}}: k'' \in \{k'-1,\ k'+1,\} \wedge l'' \in \{l'-1,\ l'+1,\}$ and local standard deviation of the 8 adjacent disparity values $\sigma\left(D_{k'',l''}\right), k'' \in \{k'-1,\ k'+1,\} \wedge l'' \in \{l'-1,\ l'+1,\}$, this model states, $PS_{k',l'} \in LGm = Gm \cup \left[\overline{D_{k'',l''}} - \sigma\left(D_{k'',l''}\right),\ \overline{D_{k'',l''}} + \sigma(D_{k'',l''})\right] \wedge k \in [0,\ M_{mla}] \wedge l \in [0,\ N_{mla}] \wedge k'' \in \{k'-1,\ k'+1,\} \wedge l'' \in \{l'-1,\ l'+1,\}$, meaning disparity values have to be within an interval defined by the global disparity mean, plus or minus the global disparity standard deviation or they have to be within the local disparity mean, plus or minus the local disparity standard deviation.

The estimated disparity values $D_{k',l'}$ that may require adjustment (the potential outliers) are those that fulfil the condition $D_{k,l} \notin Gm\ \forall\ k \in [0,\ M_{mla}] \wedge l \in [0,\ N_{mla}]$, meaning they are outside the interval of the Gm model.

The adjustment for each disparity value in $Disp(k,l)$ is performed with the help of the corresponding estimator provided by $\hat{\theta}(k,l,q)$. Previously, the disparity value was the q value of the maximum $\hat{\theta}(k,l,q)$ value. Now, all values are considered, except for the maximum, until a $q \in LGm$ is found and used as a suitable replacement for the "potential outlier" disparity value $D_{k',l'} = q \in LGm$. With this goal in mind, local maxima of $\hat{\theta}(k,l,q)$ are considered first, after, the remaining values, from the highest to the lowest are considered. If no $q \in LGm$ is found, $q = \overline{D_{k'',l''}}: k'' \in \{k'-1,\ k'+1,\} \wedge l'' \in \{l'-1,\ l'+1,\}$.

The output of this sub-module, and ultimately the output of this module, is the modified disparity matrix $Patch(k,l)$, presented in Equation (18),

$$\boldsymbol{Patch(k,l)} = \begin{cases} \boldsymbol{Disp(k,l)}, & \boldsymbol{,Disp(k,l) \in Gm} \\ \boldsymbol{pos}\left(\boldsymbol{\hat{\theta}(k,l,q)},\ \max\left(\boldsymbol{\hat{\theta}(k,l,q) \cap LGm}\right)\right), & \boldsymbol{,Disp(k,l) \notin Gm \wedge \hat{\theta}(k,l,q) \cap LGm \neq \emptyset} \\ \boldsymbol{\overline{D_{k'',l''}}}, & \boldsymbol{,Disp(k,l) \notin Gm \wedge \hat{\theta}(k,l,q) \cap LGm = \emptyset} \end{cases} \quad (18)$$

where well behaved (within Gm) disparity values are returned as disparity values, as described in the first branch of Equation (18), "outlier disparity values" are replaced with more statistically relevant ones, as described in the second branch of Equation (18); "outlier disparity values" that have no statistical relevant substitutes are exchanged by their local disparity mean, as described in the third branch of Equation (18). This last fall back solution aims to minimize the appearance of visual artefacts after the extraction process.

4.2.3 Integral Projection Module

This module consists in projecting the image samples onto a new frame of reference. It has the same resolution of the Pre-Processed Holoscopic Image, $I(x,y)$. This assures the spatial redundant areas of the micro-images are projected on top of each other, with no regard for a uniform distribution of samples on the new frame of reference, resulting in a 2D representation of the scene.

4.2.3.1 Scaling the Micro-Images

The objective of this operation is to space the samples within each micro-image according to the depth of the spatial region represented in each micro-image. This ensures that the representations of all

regions, at every depth, in the region of the micro-image chosen to represent the selected PoV, will have the same spatial dimensions in all micro-images.

The input and output of this operation are:

a. Input Pre-Processed Holoscopic Image – This input data feeds this module holoscopic samples and their position in the holoscopic image;
b. Input Micro-Image Resolution – Two values, height and width, each representing a dimension's size of the micro-images; this input is used to delimit micro-images when asserting to which micro-image a sample belongs;
c. Input Micro-Image Patch Size Matrix – Patch size matrix, with the same dimensions of the cropped holoscopic image in micro-image units, where each value within the matrix is used as scaling factor for a corresponding micro-image;
d. Output Scaled Micro-Image Samples – Holoscopic samples correctly spatially arranged to form a 2D representation of the captured scene, for a central PoV. Because the samples may no longer have integer positions, they are no longer in matrix form but they are rather in array form, where each position has a sample, its new position in 2D space and also the position of the sample inside the source micro-image.

The processing performed in this operation consists in scaling the micro-images. In the *Micro-Image Patch Size Matrix* input, there is a patch size value $Patch(k', l') = M_{k',l'}$ for each micro-image $I_m(i, j)$. Each micro-image with size (M_m, N_m) is scaled from the centre outwards depending on the corresponding patch size. The new size of a micro-image (M'_m, N'_m) can be calculated according to Equation (19).

$$(\boldsymbol{M'_m}, \boldsymbol{N'_m}) = \frac{\boldsymbol{M_{k',l'}}^2}{(\boldsymbol{M_m}, \boldsymbol{N_m})} \qquad (19)$$

The position of each sample is proportionally scaled and placed in the *Scaled Micro-Image Samples* output. Because the samples may no longer have integer positions they are no longer in matrix form; they are in array form where each position has a sample, its position in 2D space and the position of the sample inside the source micro-image.

4.2.3.2 Selecting the PoV

The objective of this sub-module is to displace the samples according to the selected PoV, proportionally to the patch size of the micro-image each sample belongs, in order to represent the scene from the requested PoV.

The input and output of this sub-module are:

a. Input Scaled Micro-Image Samples – Output of the previous operation;
b. Input Micro-Image Resolution – Two values, height and width, each representing a dimension's size of the micro-image, to be used to move the samples according to their scale factor;
c. Input Selected PoV – A two dimensional coordinate pinpointing a position inside the micro-images. This value pinpoints the position of the most relevant AoV range in the holoscopic image for the selected PoV. This input is used to move samples according to their scale factor;

d. Input Micro-Image Patch Size Matrix – Patch size matrix used as scaling factor for the micro-images;
e. Output Spatially Rearranged Holoscopic Samples – Holoscopic samples correctly spatially arranged to form a 2D representation of the captured scene, for the selected PoV. The samples don't have integer; each position has a sample, its new position in 2D space and also the drifted position of the sample inside source the micro-image.

The processing performed in this operation consists in moving the micro-images according to the *Selected PoV* input using the scale each micro-image was scaled to. For centred PoVs, the micro-images are not moved, while non-centred PoVs require the drifting of the scaled micro-images. The new position $p(x,y)'$ of a sample depends on the source micro-image resolution M, the size of the patch μ attributed to that micro-image and the selected PoV $\theta(x,y)$. The relation is expressed in Equation (20).

$$\boldsymbol{p(x,y)' = \frac{M}{\mu} * \theta(x,y)} \qquad (20)$$

Since not all patch sizes have the same size, the micro-images are differently moved. The new position of the samples is updated in the *Spatially Rearranged Holoscopic Samples* output, as well as the position of each sample inside each micro-image, which is directly drifted by the Selected PoV vector.

4.2.4 Scaling Module

This module consists of two independent sub-modules, applied sequentially to the input data, with the purpose of scaling the 2D arranged samples.

4.2.4.1 Computing the Final Resolution

The objective of this operation is to compute the final resolution while maintaining the scene proportions.

The input and output of this operation are:

a. Input Output Height – The height, N_{2D}, of the output 2D reconstructed image. The width, M_{2D}, is determined by multiplying the aspect ratio, AR, of the Pre-Processed Holoscopic image with this input value, i.e., $M_{2D} = N_{2D} \times AR$, with $AR = M/N$,;
b. Input Resolution of the Pre-Processed Holoscopic Image – Width and height of the pre-processed holoscopic image;
c. Output Final Resolution – Width and height of the 2D extracted image.

The processing performed in this operation consists in calculating the width w' of the final 2D extraction based on the ratio between the dimensions w/h of the Pre-Processed Holoscopic image and the output height h' given as input. This is calculated according to Equation (21).

$$\boldsymbol{w' = w/h * h'} \qquad (21)$$

It is worth noticing that, according to Equation (21), this algorithm cannot reconstruct a scene in an arbitrary resolution, rather it must be a multiple of the *Pre-Processed Holoscopic* Image resolution, which is determined by the resolution of the original *Holoscopic Image* minus the *Cropping Window*.

The final resolution is then tied directly to the original resolution of the original holoscopic image, with this module only able to determine one component of the resolution and calculating the other. Either width or height could have been chosen to be an input; however, the height was chosen to facilitate compatibility with the line resolution of a given display system.

4.2.4.2 Scaling the Samples to Match the Final Resolution

The objective of this operation is to scale the samples, up or down, to match the final resolution.
The input and output of this operation are:

a. Input Final Resolution – Width and height of the 2D extraction;
b. Input Resolution of the Pre-Processed Holoscopic Image – Width and height of the pre-processed holoscopic image;
c. Input Spatially Rearranged Holoscopic Samples – Output of a previous module with the holoscopic samples that will be have their spatial positioning scaled proportionally to the Output Height value;
d. Output Scaled Holoscopic Data – Scaled version of the *Spatially Rearranged Holoscopic Samples* input, $I(x, y)$.

The processing performed in this operation consists in scaling the spatial information attributed to the samples in relation to a referential. Because patch sizes calculated in the *Micro-Image Patch Size Computation* module can vary for extractions with a different *Selected PoV*, given the same holoscopic image as input, there is no assurance that the limits of the *Spatially Rearranged Holoscopic Samples* input are the same for every PoV. For this reason, a standard, common reference parameter to all extractions based on a holoscopic image given as input, must be used, in this case the resolution of the Pre-Processed Holoscopic Image.
All samples are scaled proportionally in relation to the Pre-Processed Holoscopic Image resolution. The new position of each sample $p(x, y)'$ is calculated according to Equation (22), which takes into account the original position $p(x, y)$ of a sample, the Pre-Processed Holoscopic Image resolution (w, h) and the new resolution calculated in the previous sub-module $(w',\ h')$.

$$p(x, y)' = \frac{(w', h') * p(x, y)}{(w, h)} \qquad (22)$$

4.2.5 Interpolation Module

This module consists of three independent sub-modules, applied sequentially to the input data, with the purpose of converting the 2D representation of the scene into a “traditional” 2D image.

4.2.5.1 Sampling with Bi-Linear Interpolation

The objective of this operation is interpolating samples, within micro-images, for the unoccupied sample positions in the final 2D extraction. The samples interpolated are for each micro-image that will contribute for a particular final sample.
The input and output of this operation are:

a. Input Scaled Holoscopic Data – Output of a previous module. A 2D representation of the scene, where each sample has the original position of the sample and the new position of the samples;

b. Output 2D Extraction Framework – 2D matrix where each sample position has associated one or more interpolated samples from relevant micro-images. Each interpolated sample has associated to it the same information the samples from the input structure this means a spatial position, a location displaced by the PoV inside the source micro-image and the interpolated sample.

The processing performed in this operation begins by creating a matrix, with the dimension equal to the calculated output resolution. Because the micro-images where scaled, for each matrix position one or more scaled micro-image may cover a specific matrix location. For each scaled micro-image, a sample is created to occupy the matrix position, through bilinear interpolation.

Figure 4.3 illustrates the bilinear interpolation used to compute a sample at a particular position. The original four samples are in the positions (14, 20) with value 91, (15, 20) with value 210, (14, 21) with value 162 and (15, 21) with value 95. These samples will be referenced throughout this sub-module description as s_1, s_2, s_3, and, s_4, respectively. The new sample, calculated through bilinear interpolation for position (14,5; 20,2), has the value 146,1. The auxiliary samples (14,5;20) with value 150,5 and (14,5;21) with value 128,5 are calculated in an intermediary step. These two samples will be referenced throughout this sub-module description as s', s_a and s_b.

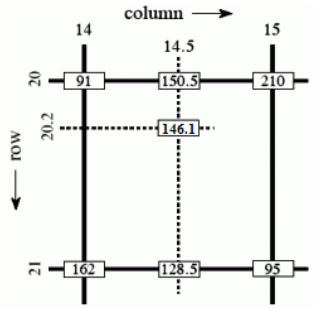


Figure 4.3 - Bilinear interpolation example

First, the original samples are interpolated horizontally, creating two auxiliary samples, as expressed in Equation (23) and Equation (24):

$$\boldsymbol{s_a} = \frac{s_{2,x} - s_{a,x}}{s_{2,x} - s_{1,x}} * \boldsymbol{s_1} + \frac{s_{a,x} - s_{1,x}}{s_{2,x} - s_{1,x}} * \boldsymbol{s_2} \qquad (23)$$

$$\boldsymbol{s_b} = \frac{s_{4,x} - s_{b,x}}{s_{4,x} - s_{3,x}} * \boldsymbol{s_3} + \frac{s_{b,x} - s_{3,x}}{s_{4,x} - s_{3,x}} * \boldsymbol{s_4} \qquad (24)$$

The results from Equation (23) and Equation (24) are then interpolated to create the new sample through Equation (25):

$$\boldsymbol{s'} = \frac{s_{b,y} - s'_y}{s_{b,y} - s_{a,y}} * \boldsymbol{s_a} + \frac{s'_y - s_{a,y}}{s_{b,y} - s_{a,y}} * \boldsymbol{s_b} \qquad (25)$$

4.2.5.2 Computing Sampling Weights

The objective of this operation is to weight each interpolated sample according to its relevance for the final 2D extracted image.

The input and output of this operation is:

a. Input 2D Extraction Framework – Output of previous sub-module. A 2D matrix where each sample position has one or more interpolated samples from relevant micro-images.

b. Input Micro-Image Resolution – Height and width, each representing a dimension of the micro-images, used to calculate the centre of the micro-images;
c. Output Weighted 2D Extraction Framework – The same data structure in the input plus a weight value associated to each sample.

After the creation of each interpolated sample, a weight value is attributed to it using a two dimensional Gaussian distribution, as defined by Equation (26). This weight expresses the impact an interpolated sample from a particular micro-image should have in the final sample it corresponds to.

$$\boldsymbol{f(x,y) = Ae^{-(\frac{(x-(x_0+x_{pov}))^2}{2\sigma_x^2}+\frac{(y-(y_0+y_{pov}))^2}{2\sigma_y^2})}} \qquad (26)$$

For each micro-image, the Gaussian distribution has a spread (σ_x, σ_y) equal to the corresponding micro-image patch size. For centred PoVs, each sample is weighted according to its distance from the centre of its micro-image (x_0, y_0), the centre being the position with highest weight. For non-centred PoVs, the centre of each micro-image is considered to be the result of adding the location of the micro-image centre with the *Selected PoV* input vector (x_{pov}, y_{pov}), corresponding to the position with highest weight. The process is done with a Gaussian function because theoretically it appropriately models the importance of the samples in the final extraction [36].

4.2.5.3 Computing the Weighted Average

The objective of this operation is to blend the interpolated samples into the samples that will compose the final 2D extracted image output.

The input and output of this operation are:

a. Input Weighted 2D Extraction Framework – Same as above;
b. Output 2D Extraction – A 2D image representing the scene from a specific PoV, with a specific resolution, were the samples are uniformly disposed in matrix form.

After the interpolated samples are created and weighted, for each matrix position, it remains blending them together. By using the weights attributed to the samples, a weighted average operation is performed to blend the samples for each matrix position, creating a single final blended sample, b', that will describe the scene in that position. This calculation is presented in Equation (27) where n interpolated samples are weighed according to the collective weight, w_t, of the samples s, which is calculated through Equation (28), and its own weight w_i.

$$\boldsymbol{b' = \frac{\sum_{i=0}^{n}(s_i * w_i)}{w_t}} \qquad (27)$$

$$\boldsymbol{w_t = \sum_{i=0}^{n} w_i} \qquad (28)$$

The blended samples are placed in the output 2D extraction matrix.

5
Proposed Test Methodologies

This chapter intends to describe the methodologies proposed for the battery of subjective and objective image quality assessment tests to be performed on a selection of 2D d images, created through the View Selective-Blending (VSBe), Single-Sized Patch Based (SSPe), Single-Sized Patch Blending Based (SSPBe) and Disparity-Assisted Patch Bending (DAPBe) extraction methods; these methods are the best performing from those presented in previous chapters. The motivations to perform the forward described tests are: i) to determine the viability of using currently available objective Image Quality Assessment (IQA) methods to evaluate the quality of 2D image extractions created from 3D holoscopic images; ii) to analyse the subjective quality perception performance of the 2D extraction methods referenced in this Thesis, as well as of the DAPBe method proposed in this Thesis.

5.1 Test Resources

This first section will define the test conditions to be used, notably the holoscopic resources as well as the parameters used for the 2D extractions methods to be assessed. First, the resources are presented and categorized into meaningful data sets, each representing different holoscopic acquisition conditions. Next, baseline configurations for all extraction methods used to generate 2D images (from a 3D holoscopic image) are defined as well as the data sets associated to them. Finally, because of the impracticality of using all the available content, the available test material content is sampled into data sets, each representing a relevant characteristic.

5.1.1 Holoscopic Test Resources

The tests will be performed on a selection of samples from a database of holoscopic images and holoscopic video frames. The samples are divided in categories, each containing resources from a single capture method. Because 3D holoscopic images are not a common resource, each category has between 4 to 7 entries, depending on the availability of samples for each category. Each entry corresponds to a particular scene. For reference in this document, these categories are colour coded, each colour corresponding to the following capture method:

1) Image Resources

a) Plenoptic 2.0 holoscopic camera [9] – The content in this image category has a high pixel resolution and a low MLA resolution. This content is presented first in Table 5.1 and can be found in [42].

b) 3D VIVANT Canon holoscopic camera (version 2) [20] – The content in this image category has a low pixel resolution and a low MLA resolution. This content is presented second in Table 5.2 and can be found in [43].

c) 3D VIVANT Canon holoscopic camera (version 1) [20] – The content in this image category has a high pixel resolution and a high MLA resolution. This content presented third in Table 5.3 and can be found in [44].

2) Video Resources

a) 3D VIVANT Arri Alexa holoscopic video camera (version 1) [20] – The content in this video category has a medium pixel resolution and a low MLA resolution. This content presented forth in Table 5.4and can be found in [45].

b) 3D VIVANT Arri Alexa holoscopic video camera (version 2) [20] – The content in this video category has a high pixel resolution and a low MLA resolution. This content is presented fifth in Table 5.5and can be found in [46].

The images and video frames chosen from the database are presented in Table 5.6.

Table 5.6 – Resources selected from the Test 3D holoscopic database

Resource		Holoscopic Image Resolution			MLA Resolution		
Name	Type	Width	Height	Count	Width	Height	Count
Fountain	Image	7240	5433	39334920	96	72	6912
Fredo	Image	7240	5433	39334920	96	72	6912
Jeff	Image	7240	5433	39334920	96	72	6912
Laura	Image	7240	5433	39334920	96	72	6912
Seagull	Image	7240	5433	39334920	96	72	6912
Sergio	Image	7240	5433	39334920	96	72	6912
Zhengyun1	Image	7240	5433	39334920	96	72	6912
4Foot 2	Image	5616	3744	21026304	70	46	3220
Airplanes 2	Image	5616	3744	21026304	70	46	3220
Cars 2	Image	5616	3744	21026304	70	46	3220
Dino 2	Image	5616	3744	21026304	70	46	3220
Fish 2	Image	5616	3744	21026304	70	46	3220
Humans 3	Image	5616	3744	21026304	70	46	3220
DynamicWithoutLight	Image	5556	3704	20579424	188	125	23500
Composition1	Image	5616	3744	21026304	190	127	24130
Composition2	Image	5616	3744	21026304	190	127	24130

Composition3	Image	5616	3744	21026304	190	127	24130
Plane129	Video	1920	1080	2073600	68	38	2584
Robot287	Video	1920	1080	2073600	69	39	2691
Robot292	Video	1920	1080	2073600	69	39	2691
Robot1029	Video	1920	1080	2073600	69	39	2691
AntennaTelaio_tot375	Video	2880	1620	4665600	76	42	3192
Demichelis_backmotion180	Video	2880	1620	4665600	76	42	3192
Demichelis_cut250	Video	2880	1620	4665600	76	42	3192
Demichelis_dolly160	Video	2880	1620	4665600	76	42	3192
RegistratoreLorentz_fix375	Video	2880	1620	4665600	76	42	3192
ValvoleRadio_dettpan300	Video	2880	1620	4665600	76	42	3192

In Table 5.6, the 1st and 2nd columns contain the name and type of content, respectively. The name is simply a common denominator for the resource while the type indicates if the resource is a holoscopic image or a frame from a holoscopic video. The 3rd, 4th and 5th columns contain the width, height and total number of pixels in the holoscopic content while the 6th, 7th and 8th columns contain the width, height and total number of micro-images in the holoscopic content.

A graphical analysis of the content used as test resources in terms of image and MLA resolutions, based on the information presented on Table 5.6, can be found in Figure 5.1. Each entry of the legend of Figure 5.1 aggregates each of the categories presented at the beginning of this sub-section, in order of appearance.

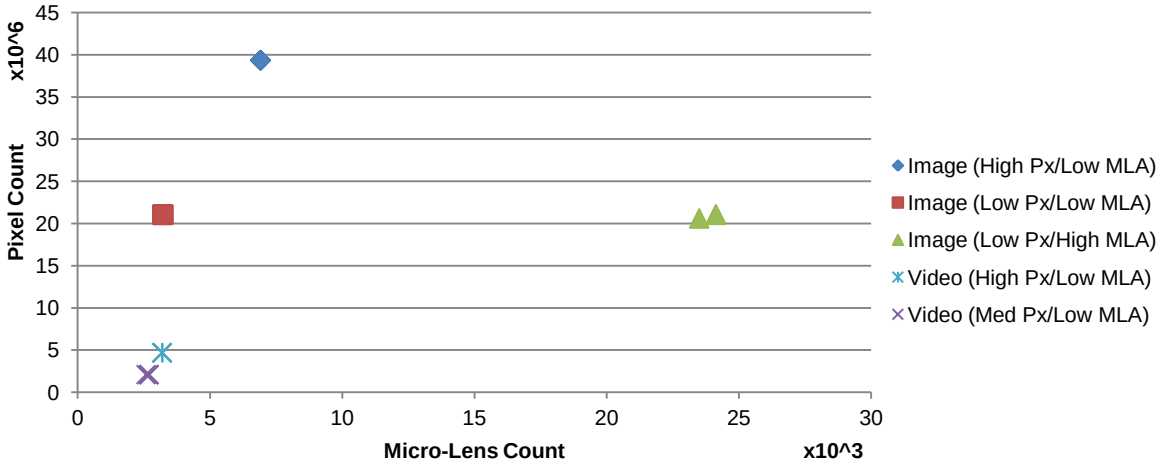


Figure 5.1 - Graphical analysis of the test resources

The last two sets seem very similar, although they are not. The last one instead of square micro-lenses, like all the others, has circular micro-lenses. This results in a lower usable pixel count than the second to last. This will enable the comparison between circular and square micro-lens holoscopic images.

5.1.2 2D Extractions Test Conditions

Comparing the 2D extraction methods previously presented requires having at hand implementations of the methods in order to produce comparable results. To that end, the various extraction methods were implemented by the author of this Thesis, using instructions from the original authors, to replicate the described 2D extraction methods as accurately as possible. The implementations where tuned to

meet resolution, angle of view, compression and consistency requirements, as described below, to allow extracting images to be comparable.

All methods where implemented in C++ for the Central Processing Unit (CPU) as well as for the Graphics Processing Unit (GPU), making use of the GPU's immensely superior graphics processing power which allows for fast processing. However, the GPU interaction also required the use of the high-level shading language OpenGL Shading Language (GLSL)[47].

A set of resolution, angle of view, compression and consistency requirements were established in order the quality of the extracted images could be maximized. Each of these requirements assures the following:

i. Resolution – Since the 2D extraction methods described produce 2D images at various resolutions, this test requirement was adopted to assure that all methods reconstruct images with the same resolution. However, as previously mentioned, some methods may not be able to produce images with the selected resolution. In those cases, the 2D extractions were scaled to 1080 pixels in height, keeping the original picture aspect ratio. To minimize the effect this interpolation has on the quality assessment results, all methods use the same interpolation algorithm, in this case the OpenGL Bi-Linear interpolation implementation on the GPU [48]. After interpolation, the 2D extractions match in height but the image ratios are not guaranteed to match. Maintaining the extracted images height, the image ratios are forced to match the ratio of the original holoscopic image through modification of the images width. By multiplying the ratio of the original holoscopic image by the common height value, the target width value is calculated. If the extraction exceeds the target width, the image is cut in the left and right sides equally to match the target width. If, however, the image width is shorter than the target width, two black bars, with the same size, are added to the image on the left and right edges, to fill the image and match the target width. This assures that all the created images have the same resolution.
ii. Angle of Vision - To some extent, all methods support extractions from multiple AoVs. As this aspect will not be under testing, the AoV should be fixed to the same setting for all methods. In this case, all methods were configured to reconstruct the central AoV of the scene.
iii. Compression – To avoid artefacts possibly caused by a lossy encoding process, no compression was performed on the 2D extractions after the extraction itself. This is assured by reading the 2D extracted images directly from the GPU memory bank into a BMP file on disk.
iv. Consistency – The extraction methods should produce consistent results, meaning that it produces the same results every time it is run with an identical set of inputs. No method requiring manual interaction can be considered as a consequence of this requirement.

Based on these conditions, and other reasons that will be covered later, some methods were excluded from testing. A case by case analysis is presented:

i. The Depth Blending Based 2D Image Extraction method was excluded from testing because currently there is no implementation available and its description in literature is extremely vague;
ii. The Disparity Map Based 2D Image Extraction method was excluded from testing because it requires human interaction to produce disparity map results. This constraint violates the consistency requirement above.

iii. The Angle of View Based 2D Image Extraction method was also excluded from testing because it produces very low resolution 2D extractions prior to interpolation. Any quality assessment preformed on these images would essentially reflect the quality of the interpolation method and less the quality of the extraction method.

The four remaining 2D extraction methods where configured to deliver as comparable results as possible. All the assessed texture based extraction methods, notably the View Selective-Blending (VSBe), Single-Sized Patch (SSPe) and Single-Sized Patch Blending (SSPBe) based methods, require focusing at a single image plane in each extraction. Thus, between 1 to 5 different focused extractions where created with each extraction method, depending on the viability of focusing the extractions on multiple planes of a scene. As the Disparity-Assisted Patch Bending (DAPBe) method proposed in this Thesis produces All-In-Focus extractions, only one extraction for each holoscopic image was created with this method.

In Table 5.7, the number of images extracted with different focus with each method for each holoscopic image resource is presented. The final test material is a set of 294 images as indicated in Table 5.7.

Table 5.7 - Number of images extracted with each 2D extraction method for the testing phase.

Resource		Number of Extracted Images				
Name	**Type**	**AVe**	**VSBe**	**SSPe**	**SSPBe**	**DAPBe**
Fountain	Image	1	4	4	4	1
Fredo	Image	1	4	4	4	1
Jeff	Image	1	2	2	2	1
Laura	Image	1	3	3	3	1
Seagull	Image	1	3	3	3	1
Sergio	Image	1	4	4	4	1
Zhengyun1	Image	1	4	4	4	1
4Foot 2	Image	1	3	3	3	1
Airplanes 2	Image	1	3	3	3	1
Cars 2	Image	1	3	3	3	1
Dino 2	Image	1	3	3	3	1
Fish 2	Image	1	3	3	3	1
Humans 3	Image	1	3	3	3	1
DynamicWithoutLight	Image	1	4	4	4	1
Composition1	Image	1	5	5	5	1
Composition2	Image	1	5	5	5	1
Composition3	Image	1	4	4	4	1
Plane129	Video	1	1	4	4	1
Robot287	Video	1	3	4	4	1
Robot292	Video	1	3	4	4	1
Robot1029	Video	1	3	4	4	1
AntennaTelaio_tot375	Video	1	1	1	1	1
Demichelis_backmotion180	Video	1	1	1	1	1
Demichelis_cut250	Video	1	1	1	1	1
Demichelis_dolly160	Video	1	1	1	1	1
RegistratoreLorentz_fix375	Video	1	1	1	1	1
ValvoleRadio_dettpan300	Video	1	1	1	1	1

Total	**294**	27	76	82	82	27	

5.1.3 Sampling Test Resources

The details on the tests to be performed on the 2D extracted resources will be covered on the next sections but the reasons for sampling the set of 294 images will be motivated and presented in this section. There will be two types of tests performed on these 2D extracted images:

1. Objective tests - the extracted 2D images are rated according to some objective, mathematical metrics. Processing 294 images by a computer, as complex a task as that may be, it is a rather practical and feasible one.
2. Subjective tests - the extracted 2D images are rated by human test subjects according to their perception of quality. In this case, processing 294 images may prove unpractical.

Motivated by the impracticality of performing such long subjective tests, the decision for resorting to statistical methodology was made. The idea is to lower the number of subjective tests (decreasing the number of images submitted to subjective testing) while still assuring that the final assessment results are close to what they would be if all 294 images where submitted to subjective testing. Statistically speaking, one can argue that if tests are executed on a sample of the 2D extractions dataset, the results of that subset will reflect closely the results for the entire set, if that sample is big enough. To ensure that this statistical truth holds, the sample must appropriately represent the population of 294 images. To this end, an unbiased subset of the 294 images was chosen using a uniformly distributed random generator through the following process:

1. Each holoscopic image receives a number. The numbers are unique only inside a holoscopic image category.
2. For each category, a random number generator picks a number from the interval of numbers attributed to that category.
3. The 2D extractions corresponding to the holoscopic image that has that number are picked for the testing phase.

As a result, five sets of 2D extractions were picked, totalling a much more reasonable number of 52 test images to be used during testing. The 52 images originate from holoscopic content as described in Table 5.8.

Table 5.8 - Sampling of the holoscopic database

Resource		Holoscopic Image Resolution			MLA Resolution			Possible Planes of Focus	
Name	**Type**	**Width**	**Height**	**Count**	**Width**	**Height**	**Count**	**Enumerated**	**Count**
Fountain	Image	7240	5433	39334920	96	72	6912	Background, Trees, Water Splash, Fountain, All	13
Dino 2	Image	5616	3744	21026304	70	46	3220	Figure Back, Figure Middle, Figure Front, All	9
Composition1	Image	5616	3744	21026304	190	127	24130	Background, Plane, Hood, Dino, Spiderman, All	16
Plane129	Video	1920	1080	2073600	68	38	2584	Plane, Support, Doll, Wing, Propeller, All	10
Demichelis_cut250	Video	2880	1620	4665600	76	42	3192	Demichelis, All	4

Now that the test content has been selected, the following section will describe the two test phases, objective and subjective.

5.2 Objective Evaluation Methodology

Automated IQA through objective evaluation is used not only in research but also in industry to guide image quality sensitive processes. To this end, it is important to have IQA metrics that are in line with human perception and rate images in a similar way as a human would. To date, no work has been done towards finding a suitable objective IQA metric that fits the task of evaluating 2D extractions from holoscopic images and rates them according to human quality perception. This section will cover this issue and define the test methodology to hopefully shine light on the issue.

The various 2D extractions methods output different images for the same scene. Since all methods attempt to extract 2D images from holoscopic content, but deliver different output given the same input, it is important to measure how good a method actually is in extracting 2D images.

Generally, IQA relies on a reference or “ground truth” image to provide a quantitative measure of comparison between an original and a ‘copy’ [49]. Because of this characteristic, methods that use a reference for the quality assessment are in a category known as Full Reference (FR) metrics. Popular quality metrics in this category include the *Peak Signal Noise Ratio* (PSNR), the *Root Mean Squared Error* (RMSE) [50] and *Structural SIMilarity* (SSIM) [51]. However, a reference image is not always available. When this is the case, other IQA methods exist to obtain the quality assessment: these methods are in a category known as No-Reference (NR) metrics.

Holoscopic images are not fit to use as a reference for any Full Reference metric that is available because 4D holoscopic images are very different from the 2D images that come out of the extraction methods. Thus, the only option to measure quality, resorting to objective IQA metrics, is to adopt NR IQA metrics.

Some NR IQA algorithms are based on models that can learn, through a training process, to predict human judgments of image quality from collections of human-rated ‘distorted’ images [52], [53], [54], [55], [56], [57]. Although they are in line with human perception, they are necessarily limited, since they can only assess quality degradations arising from the distortion types that they have been trained on. These algorithms are known as ‘opinion-aware’ (OA) because they have been trained on human rated distorted images and associated subjective opinion scores.

There are also NR IQA algorithms that do not base their score on human judgement. These algorithms are known as ‘opinion unaware’ (OU) and attempt to perform IQA based on distortion analysis metrics without human input. The OU algorithms in turn fall into two sub-categories depending on what type of distortion data is assumed or used in a training phase. Thus, an algorithm is classified as ‘distortion aware’ (DA) by design or by training on (and hence tuned to) specific distortion models to guide the QA process; algorithms classified as ‘distortion unaware’ (DU) rely only on exposure to naturalistic images or image models to guide the QA process.

Since no IQA metric has been tested or designed for the purpose of doing IQA on 2D extraction methods from holoscopic images, metrics from all categories of NR IQA metrics where considered as a possible source of objective IQA. The following list summarizes the motivations behind the choice of objective metrics, organized by the existing NR IQA categories, to be considered as 2D extraction IQA metric:

i. **Opinion Aware Distortion Aware** – Although many methods exist within this category, none is suited for this application because, in essence, this category represents 'built to purpose' NR IQA. Nearly all the available metrics target encoded content, which present very specific distortions that are not necessarily in line with the distortions found in 2D extractions. For this reason, no metric was considered appropriate and thus chosen from this category.

ii. **Opinion Aware Distortion Unaware** – This category may be a viable option for IQA within the holoscopic context because it does not rely on any specific distortion. Thus, algorithms like *Distortion Identification-based Image Verity and INtegrity Evaluation* (DIIVINE) [53], *Code Book Image Quality* (CBIQ) [55], *Learning Blind Image Quality* (LBIQ) [56], *BLind Image Integrity Notator using DCT-Statistics* (BLIINDS) [54] and *Blind/Referenceless Image Spatial QUality Evaluator* (BRISQUE) [52] are viable OA IQA solutions. As demonstrated in [54], from all these methods, BRISQUE is the one presenting the best overall performance [52] in comparison to other IQA metrics; for this reason, it was chosen for this Thesis.

iii. **Opinion Unaware Distortion Aware** – There is a single algorithm in this category that seemed relevant. The Anisotropic Quality Index (AQI) detects distortions associated with image integrity by calculating image anisotropy, and rates the content accordingly. Since the measured distortion relates to integrity, it may prove a good metric for NR IQA of 2D extractions. In testing for non-application specific IQA [58] [59], this method shows good performance against PSNR, SSIM and RSME; for this reason, it was chosen for this Thesis.

iv. **Opinion Unaware Distortion Unaware** – As far as NR goes, this is the category corresponding to the truly blind IQA algorithms. Of the few algorithms in this category, the *Natural Image Quality Evaluator* (NIQE) is the one worth mentioning at this time because it seems to hold up in testing [60] against other NR IQA metrics like BRISQUE.

From each category, an IQA metric was picked for testing in an attempt to find the best performing one for the 2D extraction problem at hand. A more in depth analysis of the AQI, NIQE and BRISQUE metrics will be done in the next sections.

5.2.1 Anisotropic Quality Index (AQI)

In the context of image processing, anisotropy is the property of being directionally dependent, meaning it relates to the directional properties of the image samples. This IQA method is based on the assumption that some degradation processes, such as blur or noise, introduce a substantial change in the scene's directional information. Anisotropy, as a directionally-dependent quality of images, decreases as more degradation of this kind (blur or noise) is added to the image.

To obtain an AQI measurement for an image, the following apply:

1. For each sample, a measure of spatial-frequency is calculated for several directions based on the neighbour samples, using the Wigner distribution [61].
2. For each direction, in each sample, an entropy function is constructed and used as a measure of directional change.
3. These directional entropy functions are averaged together for the corresponding samples, resulting in one average entropy function for each sample.
4. The standard deviation of these functions is calculated to express the AQI score of an image.

Because, during the generation of test images, the 2D extraction methods were instructed to recreate scenes to the same specifications (all extraction methods are given the same problem to solve), the rating attributed by this metric could be significant in a global sense. Therefore, the AQI ratings are comparable for 2D extractions from the same holoscopic image as well as for 2D extractions from different holoscopic images.

The maximum and minimum values attributed by this metric are not specified by the authors [58][59][62][52]. For this reason, the maximum and minimum possible ratings will correspond to the maximum and minimum ratings observed with the test resources.

During testing, for each of the test resources, an AQI score is calculated. However, to obtain a meaningful and comparable IQA rating based on the AQI metric, ratings have to be normalized. This normalization is performed using the maximum and minimum ratings observed for the entire testing set, as the limits of the AQI index.

5.2.2 Natural Image Quality Evaluator (NIQE)

The Natural Image Quality Evaluator is a completely blind image quality analyser that only makes use of measurable deviations from statistical regularities observed in natural images, without training on human-rated distorted images and without any exposure to distorted images. The basic principle behind this metric is to rate content based only on the deviation of a calculated natural scene statistic (NSS) model; the NSS statistic model is built out of the natural images used to train the method.

The NIQE rating process follows these steps:

1. The original image is partition into several square patches.
    a. In [60], the size of the patches was experimentally determined to be between 32x32 and 160x160 pixels. The adopted value should be as close to the middle of the interval as possible.
    b. The whole image has to be partitioned, without leaving any sample out.
    c. The partitions can go beyond the margins of the image, but this has to be minimized because it interferes with the results.
2. The sharpness of each patch is estimated by calculating the average of the local standard deviations within each patch. The range of values used to calculate the local standard deviations is not specified in the literature.
3. Patches with overall low sharpness are discarded from the rating process to minimize the influence of blurriness caused by out of focus regions in an image; this is done because, even in natural images, , there may be blurry regions due to the depth of field used to capture the image.
4. Based on the principle described in [63] which states that natural images possess certain statistical consistencies that hold within any random sized image patch, a NSS statistical model is built for each patch, relating 18 different features of each patch. This is also done for a scaled down version, to half size, of each patch. This results in a total of 36 different features for each patch.
5. Based on the NSS models built for each patch, the 36 different features expressing the naturalness of the region of interest are used to calculate the mean and covariance of each of these 36 features.

6. Using the mean and covariance calculated in the previous step, the distance from the 36 computed features to the 36 features computed in the training phase is determined. The distance between the two models, expressed by the 36 features distance, is the NIQE score.

The 36 features computed in the training phase are obtained through the following process:

1. 36 features are computed for each training natural image, as it is done in the rating process from step 1 through step 4;
2. To each of the 36 features calculated for each training image, a top rating (best quality rating possible) is attributed to each of those values for each feature;
3. 36 pairs of mean and covariance, one for each of the 36 features, are calculated to represent the best rating that an image can have.

The NIQE index, $D(\nu_1, \nu_2, \sum_1, \sum_2)$, is calculated using the mean vectors, with the 36 mean values, ν_1, ν_2 and the covariance vectors $\sum_1$, $\sum_2$, each holding 36 covariance values, as described in Equation (29).

$$\boldsymbol{D(\nu_1, \nu_2, \sum_1, \sum_2) = \sqrt{((\nu_1 - \nu_2)^T (\frac{\sum_1 + \sum_2}{2})(\nu_1 - \nu_2))}} \quad (29)$$

The rating attributed by this metric is significant in a global sense, meaning that it can be used to compare 2D extractions from the same resource as well as 2D extractions from different resources. The maximum and minimum values attributed by this metric are not specified by the authors [60]. For this reason, the maximum and minimum possible ratings will correspond to the maximum and minimum ratings observed in the test resources.

During testing, for each of the test resources, a NIQE score is calculated. To get a meaningful and comparable IQA rating based on the NIQE metric, ratings have to be normalized. The normalization will be done using the maximum and minimum ratings observed for the entire set as the limits of the NIQE index.

5.2.3 Blind/Referenceless Image Spatial QUality Evaluator (BRISQUE)

Like NIQE, the Blind/Referenceless Image Spatial QUality Evaluator is a NSS-based distortion-generic NR IQA model. The main difference between these two metrics is that while NIQE is trained only on pristine natural images, BRISQUE is trained on all sorts of images, including pristine and also distorted images, with each accompanied with a human-sourced rating.

Although BRISQUE is trained on distorted images, it does not compute distortion specific features such as ringing, blur or blocking, but instead uses scene statistics, weighed by the human IQA, to quantify possible losses of ‘naturalness’ in the image due to the presence of distortions [62]. In other words, BRISQUE rates images according to distortions it can detect by attributing a rating to an image that presents particular statistical behaviours. The particular statistical behaviours that BRISQUE looks for are determined during training, where, for each given training image, a distortion specific statistical model is created for each test image. Each statistical model then gets a human sourced quality rating associated to it. During the evaluation stage (rating images), BRISQUE has to create a statistical model for the image being rated, compare it to the statistical models to find the closest match, and attribute the corresponding rating.

More specifically, the BRISQUE steps to rate images are as follows:

1. Based on the principle described in [63], stating that natural images possess certain statistical consistencies that apply within any random sized image patch, a NSS statistical model is built for the entire image, relating 18 different features for the image. This is also done for a scaled down version, to half size, of the image. The result is a total of 36 different features for each image.
2. From the NSS model built for the image, the 36 different features that relate to the naturalness of the region of interest are used to calculate the mean and covariance of each of the 36 features.
3. Using the mean and covariance of each of the 36 features, the 36 calculated features are framed into 36 corresponding “ranges” (each feature can vary inside a range, calculated in the training phase, as explained ahead). A rating is attributed to each feature depending on its position inside the range. The distance between the two models, compared through the 36 features, is the BRISQUE score.

The feature ranges are computed in the training phase through the following process:

1. 36 features are computed for each training image like in step 1 above;
2. The human sourced rating attributed to a particular image is associated to each of the 36 features calculated for each training image;
3. Using regression to combine the multiple rated sets of each of the 36 features, a range of possible values emerges for each of the 36 features. With the regression model, quality values are attributed to different positions of the ranges. The result is 36 different ranges, each corresponding to each feature, able to judge image quality according to each of the 36 features.

As can be observed by the BRISQUE description, it shares mechanisms with NIQE; for this reason, it also shares some of its properties. Like NIQE, the ratings are significant in a global sense; as the maximum and minimum values attributed by this metric are not specified by the authors [60], the normalization will be done using the maximum and minimum ratings for the entire set as the limits of the BRISQUE index.

5.3 Subjective Evaluation

After the content has been rated automatically by the IQA methods best suiting the 2D image creation scenario at hand, the results must be compared to some subjective scores used as reference. This will allow benchmarking the obtained objective quality assessment results with human perception. This phase of testing attempts to construct that reference by means of tests performed with human subjects who rate the same images used for the objective tests.

The test procedure adopted for the subjective assessment will follow the recommendations of ITU-T P.910 [64]. Testing is performed with at the least 30 different test subjects for the results to bear statistical significance.

The subjective rating will be measured with one of the most popular subjective assessment methodologies, the Absolute Category Rating (ACR) method, which has been standardized for images and video in ITU-T P.910 [64]. In ACR, several test subjects rate images under controlled conditions using a discrete 1-5 scale.

After the subjective tests are completed and the results can be analyzed, it is possible to cross match the results of the subjective evaluation tests with the results of the objective evaluation thus understanding how good are the selected objective metrics to assess quality in the holoscopic scenario under study.

5.3.1 Explaining the Test Procedure

Because the recommendations in ITU-T P.910 [64] are the industry standard for this type of test procedure, they were adopted for the scenario under study. Each of the 32 subjective evaluation tests will begin with a complete explanation, directed at the test subject, of what the test procedure will consist of, and what is expected from each subject. To this end, the following text will be explained to the test subjects and made available to them:

"The purpose of this test session is to gather data on the subjective quality of a set of randomly chosen images. The full test session will consist of:

1. *Visual acuity phase Two different tests to assess the test subjects visual acuity and determine if his/her visual system functions properly;*
2. *Training phase - A training phase to familiarize the test subjects with the type of images that will be presented to them during the actual quality assessment tests stage;*
3. *Test phase - The actual quality assessment test stage where the subject rates 5 randomly ordered sets of 10 images each.*

This test will proceed in the following way:

1. *The test subjects are placed at 3 meters from the screen; the Snellen chart (see* Figure 5.2*) appears and the test subjects read the characters, top to bottom, left to right, first only with the left eye open, then only with the right eye open and, finally, with both eyes open.*

Figure 5.2 - Snellen Eye Chart

2. *The test subjects are placed at 75 cm from the screen; 17 Ishihara plates (see* Figure 5.3*) will appear one after the other, spaced by 3 seconds; the test subjects say aloud what is the content of the plate within the 3 seconds the plates are on screen.*

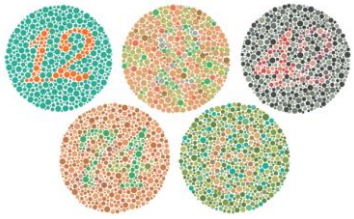


Figure 5.3 - Sample Ishihara Plates

3. *The test subjects are placed at 1.8 meters from the screen and will be given a sheet with Absolute Category Rating (ACR) scales (see* Figure 5.4*) where the test subjects should express their opinion in relation to the images that will appear on screen. To do this, the user must mark with an 'X' the box next to the quality appreciation that he/she thinks best suits a particular image. 12 images will appear on screen; the test subject rates one by one the images. The images appear for 10 seconds and then disappear for 10 seconds. During the 10 seconds when no picture is no screen, the test subject has to rate the previous on screen image.*

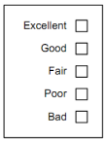


Figure 5.4 - ACR scale used for the test subjects to assess the image quality

4. *The test subjects maintain the same position and a clean sheet with ACR scales are provided to him/her. Instead of 12 images, now 52 images are presented for the test subjects to rate. The same 10 seconds for visualization and 10 seconds to vote apply.*
5. *The test is over.*

To express their opinion regarding image quality, the subjects will use a scale like the one presented in Figure 5.4*.*

The ACR scale is freely marked with the personal opinion of the test subjects by marking the option that best fits the perceived image quality with an 'X'. Only one box should be marked. If there is a mistake, the test subjects draw a circle around the answer they mean to nullify. The subject should also avoid only using the top and bottom scores of the scale."

After this phase, the test subjects are considered informed of the test procedure and the real tests begin.

5.3.2 Visual Acuity Phase

Two visual acuity tests will be administered to the test subjects in this phase. The purpose of these tests is to evaluate each test subject's visual system. This is done to ensure that the test subjects are able to see correctly, validating the results of each subjective evaluation test.

The first of the visual acuity tests is the Snellen Eye Chart test. Snellen Eye Chart is a chart with 11 lines of letters, with large letters in the top line, which gradually scale down until the bottom line, where they are very small. An example of the Snellen Eye Chart is presented in Figure 5.2. The test subject reads the letters left to right, from top to bottom, until he/she reads at least one letter incorrectly from a line. After that, the test proceeds.

Since all test subjects are adult and know how to read, a standard Snellen Eye Chart will be used. The Snellen fractions, 20/20, 20/30, etc., are measures of sharpness of sight and can be found on the chart. They relate to the ability to identify small letters with high contrast at a specified distance. They give no information about seeing larger objects and objects with poor contrast; it also does not inform as to whether or not meaning is obtained from visual input, how much effort is needed to see clearly or singly, and whether or not vision is less efficient when using both eyes as opposed to each eye

individually. In short, this visual acuity test measures only the smallest detail humans can see; it does not represent the quality of vision in general.

When checking visual acuity, one eye is covered at a time and the vision of each eye is recorded separately, as well as for both eyes together. In the Snellen fraction 20/20, the first number represents the test distance, 20 feet (roughly 6 meters). The Snellen chart will be displayed in a computer monitor with half size, so the distance of the subject to the monitor will be 3 meters. The second number represents the distance at which the average eye can see the letters on a certain line of the eye chart. For the purpose of this test, and following the recommendations of [64], if a test subject reads the first 8 lines correctly, it is considered to have at least an average vision in terms of detail, this means a 20/20 vision, and the test subject passes the test.

The second test is the Ishihara Colour Test that evaluates colour perception for red-green colour deficiencies. The test consists on a number of coloured plates, called Ishihara plates, being sowed to the test subject, in a particular order, from the distance of 75cm that he/she must read and state each plate's content within 3 seconds. Each plate contains a circle of dots appearing randomized in colour and size. Some examples of the plates used can be found in Figure 5.5, Figure 5.6 and Figure 5.7.

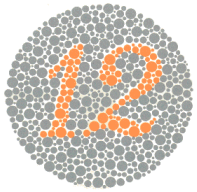


Figure 5.5 - Plate nº1 of the Ishihara test.

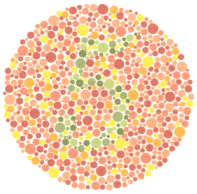


Figure 5.6 - Plate nº10 of the Ishihara test.

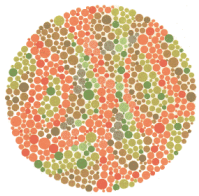


Figure 5.7 - Plate nº15 of the Ishihara test.

Within each pattern there are dots forming a number or shape clearly visible to those with normal colour vision, and invisible, or difficult to see, to those with a red-green colour vision defect, or the other way around. The full test consists of 38 plates, but the existence of a deficiency is usually clear after a few plates.

There is also the smaller test consisting only of 24 plates which is the one that will be administered. Since the purpose of administering this test is to detect and eliminate colour blind test subjects, rather than thoroughly diagnose colour blindness, only the first 17 plates of the 24 plates are needed. Every test subject that is not colour blind should be able to correctly determine the content of all 17 plates, with a maximum of 3 seconds per plate to assess its content.

If the subject is able to pass these two tests, the next phase of the subjective test is administered.

5.3.3 Training Phase

In the training phase of each subjective evaluation test, the subjects should get familiarized with the type of images that will be presented in the actual test phase and how the rating will proceed.

From the 294 images extracted for the test stages, 52 have been chosen for testing and will not be used in this phase. The images selected for the training phase are selected from the remaining 241 2D extracted images by picking 3 extracted images from each of the 4 extraction methods selected, amounting to a total of 12 images for training. Within each method, the 3 images are randomly selected by a computer. The separation from the main test set, as well as the inclusion of all types of

extraction methods in equal proportion, will allow the test subjects to be familiar with 2D extractions generated by all the extraction methods.

A sheet with ACR scales is provided to the test subjects and the training phase begins. The images are rated sequentially and, for each image that must be rated, the ACR method considers 2 stages:

1. A visualization stage, lasting up to 10 seconds, in which the test subject is presented with an image. The test subjects cannot rate the images during this stage;
2. A voting stage, lasting no more than 10 seconds. This stage starts with the removal of the image from the screen and showing the test subject a grey screen. After the grey screen appears, the test subjects can rate the images in a sheet containing ACR scales.

Following the recommendations in [64], the test images are preformed in a well lit room.. For visualization of images, the tests are performed using a Samsung SyncMaster 2343 NW monitor. For FullHD resolution, the manufacturer recommends a viewing distance between 0.9 and 2.7 meters. Since the resolution of all the test content is FullHD, the test subjects are placed in the middle of the range, in this case at 1.8 meters from the screen.

During this training phase test, the 12 images will be presented in random order. The test subjects are informed about the ACR method, the length of the test and that the training session only serves the purpose of getting the test subjects familiar with the type of images and the rating system.

5.3.4 Test Phase

The test phase of each subjective evaluation session is when the test subjects will actually rate the sample of 52 2D extracted images. The rating will be measured with the same ACR method used for the training phase (see description in Section 5.3.3). Following the recommendations in [64], the test images will be presented as described in Section 5.3.3. During the test phase, 52 images will be presented in a computer generated random order. The test subjects are reminded about the ACR method, the length of the test and that the ratings provided in this test session, in the form of ACR tables, count as IQA results. In the next section, the results of the described test sessions will be presented and analyzed.

6 Performance Assessment: Scores and Analysis

This chapter presents the scores from both the subjective and objective quality evaluation tests described in the previous chapter. After a comprehensive presentation of the relevant data, a thorough analysis of the scores is performed trying to understand which the best 2D extraction methods are and which objective quality metrics better correlate with the subjective quality assessments.

6.1 Comparing the 2D Extraction Methods Performance: Scores and Analysis

This section aims to present and analyse the results of the various tests performed according to the methodologies presented in Chapter 5, starting with the subjective test scores and afterwards the objective test scores. After the presentation of the scores for the two types of tests, they are analysed individually, using the numerical data collected as well as through visual comparisons.

6.1.1 Using Subjective Scores

This subsection presents the scores of the subjective tests performed according to the methodologies presented in Chapter 5. The presentation of the data will be done in three stages, each characterized by a particular level, or group of levels, of subjective quality. This choice was made to group together three relevant types of 2D extracted images, for the purpose of understanding what 2D extraction methods can produce 2D extractions with: i) acceptable quality, ii) poor quality and iii) bad quality. This

grouping will facilitate the exclusion of 2D extraction methods that show a below average performance in general.

The following groups of MOS scores correspond to the ACR scale (see Chapter 5.3) scores and are used here for analysis purposes:

Average and Above - The scores of the extractions that obtained a MOS between *Excellent* and *Fair,* including *Good*, i.e., 4.0 and 1.50 scores in the ACR scale, as described in Chapter 5. The subjective test scores for this group can be found in Table 6.1 and Figure 6.1;

Poor - The scores of the extractions that obtained a *Poor* MOS score, i.e. scored between 1,50 and 0,50 in the ACR scale, as described in Chapter 5. The subjective test scores for this group can be found in Table 6.2 and Figure 6.2;

Bad - The scores of the extractions that obtained a *Bad* MOS score, i.e. scored between 0,50 and 0,00 in the ACR scale, as described in Chapter 5. The subjective test scores for this group can be found in Table 6.3 and Figure 6.3.

In Table 6.1, the 1st column indicates the original holoscopic images used (see Appendix 1 trough 5), the 2nd column indicates the 2D extraction method (see Chapters 3 and 4 for the methods description and Chapter 5 for the list of those that can be tested), the 3rd column indicates the region of the extracted images in focus (if any), the 4th column contains the MOS rating obtained from the 32 individual subjective tests, as described in Chapter 5, and the 5th column contains the MOS deviation error, ε, calculated according to Equation (30).

$$\varepsilon = \frac{(1-\alpha)\sigma}{\sqrt{n}} \qquad (30)$$

The error ε is used to calculate the confidence interval $CI_{95\%}$ (see Equation (31)) for the MOS rating, with a 95% α confidence, a MOS standard deviation σ, using the 32 individual subjective test sessions n.

$$CI_{95\%} = [MOS - \varepsilon, MOS + \varepsilon] \qquad (31)$$

Table 6.1 is ordered by MOS score, from the highest to the lowest, with the highest scored extractions at the top and the lowest scored at the bottom. To separate the single *Excellent*, at the top, the *Good*, in the middle, and the *Fair*, at the bottom there are two lines between them.

Table 6.1 - MOS scores ("Excellent", "Good" and "Fair" scores) for 2D extractions

Image	Method	Focus	MOS	ε
Fountain	DAPBe	All	3,84	0,09
Fountain	SSPBe	Water Splash	3,00	0,00
Fountain	SSPBe	Background	2,91	0,07
Fountain	SSPBe	Trees	2,75	0,18
Fountain	SSPe	Trees	2,69	0,24
Fountain	SSPe	Water Splash	2,41	0,12
Fountain	SSPBe	Fountain	2,41	0,12
Dino2	SSPBe	Fig. Back	2,31	0,11
Fountain	SSPe	Background	2,22	0,31
Dino2	DAPBe	All	2,13	0,04
Plane129	SSPBe	Doll	2,06	0,06
Dino2	SSPBe	Fig. Middle	2,00	0,00
Plane129	SSPBe	Support	2,00	0,00
Fountain	SSPe	Fountain	1,97	0,04
Plane129	DAPBe	All	1,84	0,09
Composition1	SSPBe	Dino	1,72	0,11
Composition1	SSPBe	Plane	1,69	0,12

Composition1	SSPBe	Hood	1,69	0,12

Figure 6.1 represents the data present in Table 6.1 in graphical form.

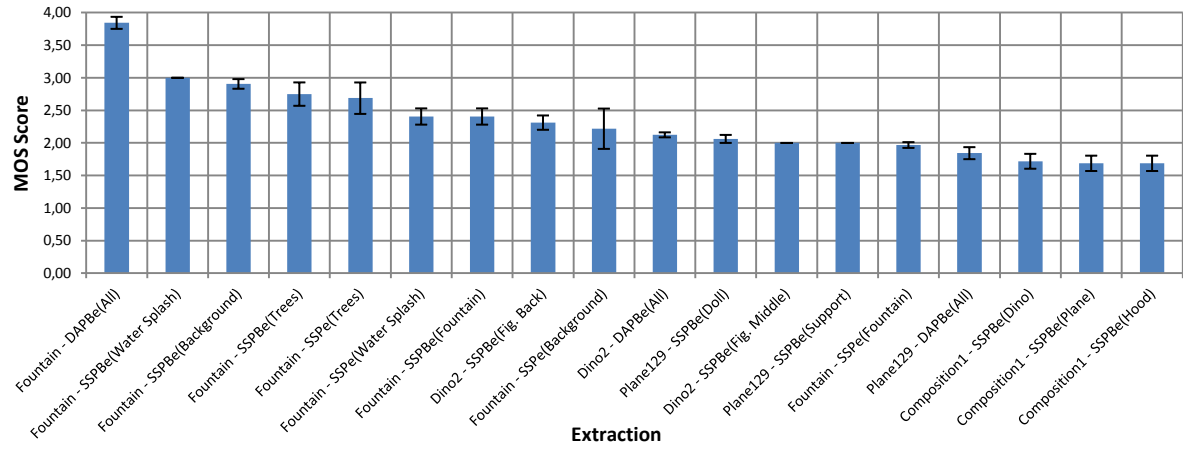


Figure 6.1 - MOS scores ("Excellent", "Good" and "Fair" scores) for 2D extractions

As in Table 6.1, in Table 6.2 the images are ordered with the highest scored extractions at the top and the lowest scored at the bottom.

Table 6.2 - MOS scores ("Poor" scores) for 2D extractions

Image	Method	Focus	MOS	ε
Demichelis_cut	SSPBe	Demichelis	1,22	0,22
Composition1	DAPBe	All	1,19	0,10
Demichelis_cut	DAPBe	All	1,16	0,09
Dino2	SSPe	Fig. Front-Centre	1,00	0,00
Composition1	VSBe	Hood	0,97	0,04
Composition1	SSPBe	Spiderman	0,97	0,04
Plane129	SSPBe	Propeller	0,97	0,04
Composition1	SSPe	Plane	0,94	0,06
Composition1	SSPe	Dino	0,94	0,06
Dino2	SSPe	Fig. Back	0,94	0,03
Dino2	SSPe	Fig. Middle	0,94	0,03
Fountain	VSBe	Water Splash	0,94	0,06
Plane129	SSPBe	Wing	0,94	0,06
Dino2	SSPBe	Fig. Front-Centre	0,91	0,03
Composition1	SSPBe	Background	0,66	0,12

Figure 6.2 represents the data present in Table 6.2 in graphical form.

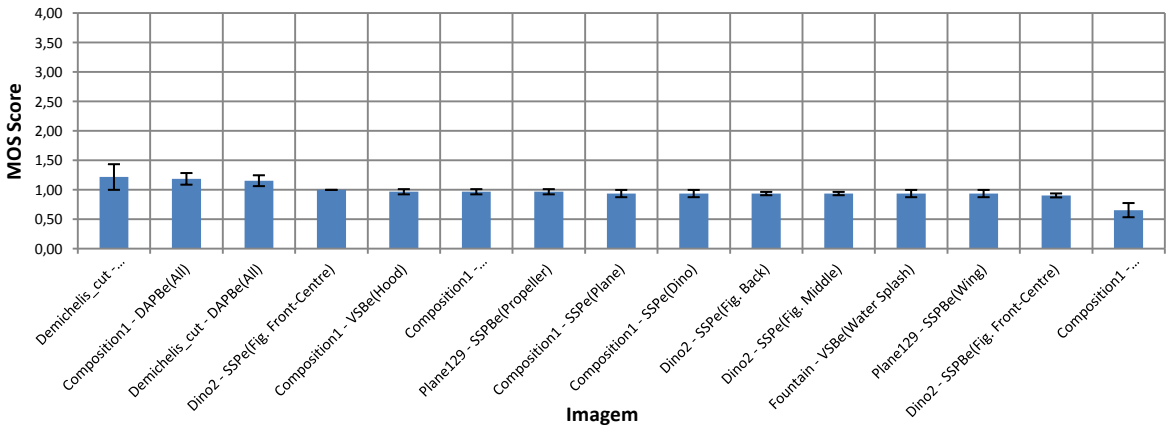


Figure 6.2 - MOS scores ("Poor" scores) for 2D extractions

As in Table 6.1 and Table 6.2, in Table 6.3 the images are ordered with the highest scored extractions at the top and the lowest scored at the bottom.

Table 6.3 - MOS scores ("Bad" scores) for 2D extractions

Image	Method	Focus	MOS	ε
Fountain	VSBe	Trees	0,38	0,12
Plane129	SSPe	Support	0,38	0,12
Plane129	SSPe	Wing	0,38	0,12
Composition1	SSPe	Hood	0,34	0,12
Composition1	SSPe	Spiderman	0,34	0,12
Plane129	SSPe	Doll	0,34	0,12
Composition1	VSBe	Plane	0,31	0,12
Dino2	VSBe	Fig. Middle	0,06	0,03
Fountain	VSBe	Fountain	0,06	0,06
Composition1	VSBe	Background	0,03	0,04
Composition1	VSBe	Dino	0,03	0,04
Composition1	VSBe	Spiderman	0,03	0,04
Composition1	SSPe	Background	0,03	0,04
Demichelis_cut	VSBe	Demichelis	0,03	0,04
Demichelis_cut	SSPe	Demichelis	0,03	0,04
Dino2	VSBe	Fig. Back	0,03	0,04
Plane129	SSPe	Propeller	0,03	0,04
Dino2	VSBe	Fig. Front	0,00	0,00
Fountain	VSBe	Background	0,00	0,00
Plane129	VSBe	Plane	0,00	0,00

Figure 6.3 represents the data present in Table 6.3 in graphical form.

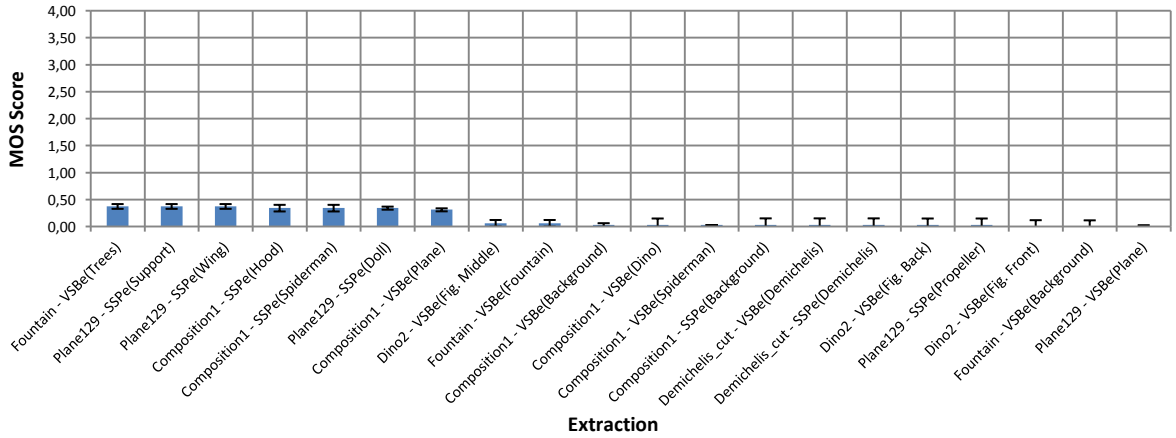


Figure 6.3 - MOS scores ("Bad" scores) for 2D extractions

Analysis I – Excluding Methods with "Below Average" Ratings

The purpose of this analysis is to identify 2D extraction methods that consistently perform below average. Based on the presented subjective scores (see Table 6.1, Table 6.2 and Table 6.3),
Table 6.4 accounts for the percentage of the extracted images, for each method, that are rated average and above (*Excellent*, *Good* and *Fair*) and below average (*Poor* and *Bad*).

Table 6.4 - Percentage of MOS ratings attributed to each 2D extraction method

Method	Average and Above				Below Average		
	Excellent	Good	Fair	Total	Poor	Bad	Total
SSPBe	0%	18%	47%	65%	35%	0%	35%

DAPBe	20%	0%	40%	60%	40%	0%	40%
SSPe	0%	6%	18%	24%	29%	47%	76%
VSBe	0%	0%	0%	0%	14%	86%	100%

According to Table 6.4 there are two methods that perform below average, the VSBe and the SSPe:

100% of the VSBe method's extractions are rated below average, indicating that 2D images extracted by this method show, typically, "Bad" subjective quality. This was expected because all images extracted with this method show a noticeable blur effect (as stated in Chapter 2), which can be seen in Figure 6.4a;

76% of the SSPe method's extractions are rated below average, indicating that 2D images extracted by this method show, typically, also "Bad" subjective quality. This is mainly due to the noticeable artefacts that, typically, appear in the image planes not in focus. This aggravates as the distance between the plane of focus and the planes not in focus increases, as stated in Chapter 2 and can be seen in Figure 6.4b.

(a) (b)

Figure 6.4 - Examples of "Below Average" 2D extractions: (a) Fountain, VSBe method, focused on the trees, MOS rated 0,38; (b) Dino2, SSPe method, focused on the dinosaur at the right, MOS rated 0,03

Analysis II – Comparing Methods with Average and Above Ratings

The purpose of this analysis is to characterize the performance of 2D extraction methods that perform mostly average and above average. According to Table 6.4 there are two methods that meet this condition:

65% of the SSPBe method's extractions are rated average and above;

60% of the DAPBe method's extractions are rated average and above.

Comparing the totals, the SSPBe method has a small advantage. However, the DAPBe method is the only one with *Excellent* ratings, while the SSPBe only goes as high as *Good*, giving the DAPBe method an advantage here.

Analysis III – Visual Artefacts in the Top Rated 2D Extractions

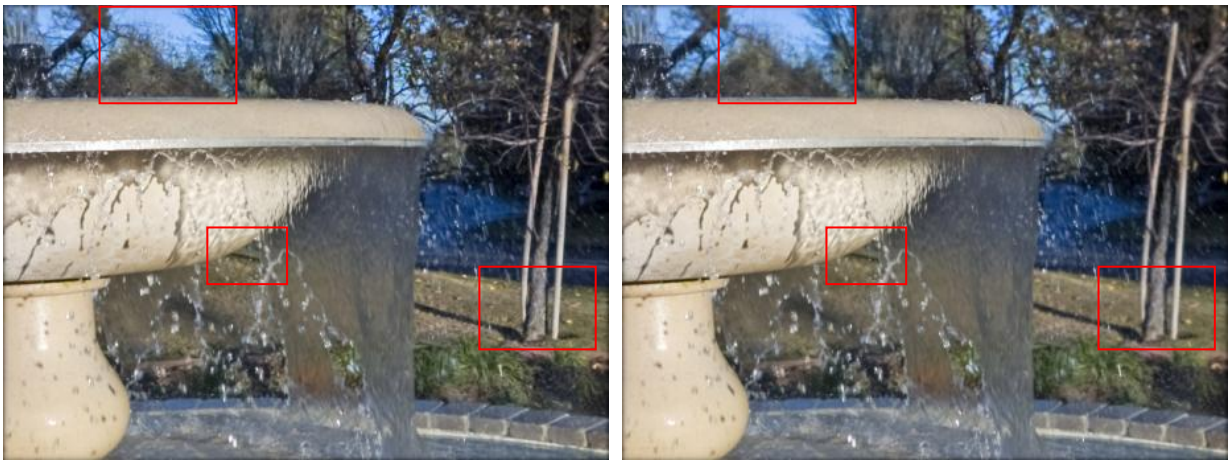
The purpose of this analysis is to identify visual artefacts in the top rated 2D Extractions. Since the DAPBe and the SSPBe methods are the only two with average and above ratings, this analysis is going to focus on them, particularly in the regions where visual artefacts exist in their extractions. The

extractions used for the comparison are from the *Fountain* holoscopic image (see Appendix 1) to facilitate comparison and to replicate the conditions in both methods as much as possible. Figure 6.5 shows the top rated 2D extractions for the DAPBe and SSPBe methods. The red squares were used to mark the three regions; one where the SSPBe method is completely out of focus, one where the SSPBe extraction is slightly in focus and one where both extractions are in focus. For each chosen region, the following observations were made:

The marked region at the top left – This region is shown amplified in Figure 6.6. As stated in Chapter 2, the process used by the SSPBe method to extract 2D images results in artefacts on regions no brought into focus. The big advancement in this method is the ability to disguise these artefacts with a blurring effect, hiding them, to achieve better perceived quality. In this region, the SSPBe method shows in fact blurring covering artefacts, although not perfectly, while the DAPBe method is sharp. Please note that the region of focus for the SSPBe extraction in Figure 6.6 is the fountain, meaning that this is the region farthest from the focus region where the most artefacts exist.

The marked region at the bottom right – This region is shown amplified in Figure 6.7. As in the previous region, in the SSPBe extraction this is for a region not brought into focus. This region however is closer to the region in focus, the water splash of the fountain. The artefacts are less noticeable in this region because the artefacts weren't as serious when the blurring was applied to cover them.

The marked region at the centre - This region is shown amplified in Figure 6.8. For both the SSPBe and DAPBe method, this region is in focus except for the ground region. No artefacts exist in this fountain and water splash region, in fact, not only they look the same, they were built the same way, only through different processes.



(a) (b)

Figure 6.5 - 2D Images extracted from the Fountain holoscopic test image using: (a) DAPBe method (all-in-focus), MOS rated 3,84; (b) SSPBe method (focused on the water splash), MOS rated 3,00

(a) (b)

Figure 6.6 - A section with trees above the fountain, from the extractions of the Fountain resource; the (a) at the left is from the DAPBe method and is all-in-focus, showing no artefacts; the (b) on the right is from the SSPBe method and is focused on the water splash, showing artefacts poorly repaired with a burring effect

(a) (b)

Figure 6.7 - A section from the tree at the right side of the fountain, from the extractions of the Fountain resource; the (a) at the left is from the DAPBe method and is all-in-focus, showing no artefacts; the (b) on the right is from the SSPBe method and is focused on the water splash, showing artefacts repaired with a burring effect

(a) (b)

Figure 6.8 - A section from the fountain, from the extractions of the Fountain resource; the (a) at the left is from the DAPBe method and is all-in-focus; the (b) on the right is from the SSPBe method and is focused on the water splash

Through these observations immerge the facts that the DAPBe method, with less information about scene topography, namely a focus region to base the extraction on, is able to:

- Outperform the SSPBe method in regions not in focus because, in the SSPBe method, artefacts become increasingly noticeable in regions distant from the focus region and the blurring effect is incapable to properly conceal them;
- Perform as good as the SSPBe method in regions that are in focus.

Analysis IV – Visual Artefacts in the Average Rated 2D Extractions

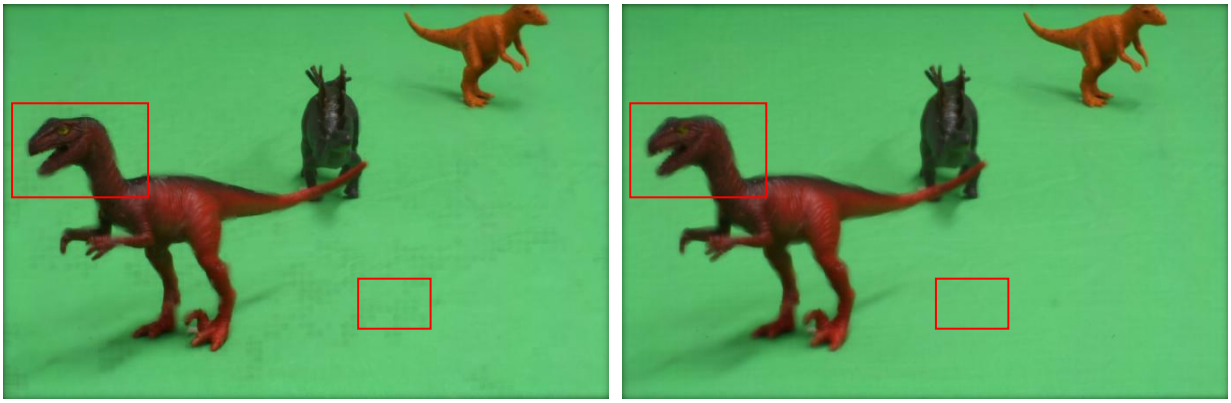
The purpose of this analysis is to identify visual artefacts in the average rated 2D Extractions (see the selection at the bottom of Table 6.1, rated between 2,50 and 1,50). Since the DAPBe and the SSPBe methods are the only two with average and above ratings, this analysis, as the previous one, is going to focus on them, particularly in the regions where visual artefacts exist in their extractions. The extractions used for the comparison are from the same holoscopic image, Dino2, (see Appendix 2) to facilitate comparison and to replicate the conditions in both methods as much as possible. Figure 6.9a shows an average rated 2D extraction of the Dino2 image done by the DAPBe method and Figure 6.9b shows an average rated 2D extraction of the Dino2 image done by the SSPBe methods. The red squares where used to mark two regions; one where the SSPBe method is out of focus and both the DAPBe and SSPBe methods generated artefacts in that region and another region where the DAPBe method generates artefacts and the SSPBe method does not. For each chosen region, the following observations where made:

The marked region at the left – This region is shown amplified in Figure 6.10. In this region, the SSPBe method (see Figure 6.10b) also shows blurring covering artefacts, while the DAPBe method shows noticeable artefacts and blurring in some areas (see Figure 6.10a). There are two independent issues here:

- The presence of these artefacts in the body and near the edges of the dinosaur. For the SSPBe method (see Figure 6.10b), this is partly because this is not a region brought into focus. The other part of the issue is directly linked with the fact that the Dino2 holoscopic resource does not have geometrically identical micro-images throughout the holoscopic image. As a result, all patches contain wrong samples of the extractions, even in regions brought into focus, creating the artefacts that neither extraction methods have a specific compensating mechanism for. This issue however may be compensated for with pre-processing. Although, the DAPBe method (see Figure 6.10a) hides it with the blurring, better in focused regions than unfocused regions. To prove this point, removing the blurring mechanism from the SSPBe method should show a misalignment of the patches in a region brought into focus. Since the SSPe method is essentially the SSPBe method without the blurring mechanism, Figure 6.12 shows this case. The misalignment is present there.
- The blurring effect in the DAPBe method. This blurring mechanism, as covered in Chapter 4, serves the purpose of smoothing out transitions between patches that have different disparity values. So, if there is blurring, the algorithm wrongly calculated, through the disparity

calculation mechanism, that there was a disparity transition there. An example of this can be seen in Figure 6.10a. The reason for the error in disparity calculation is that the DAPBe method assumes the micro-image geometry is perfect when calculating the disparity, resulting in disparity calculations based on multiple micro-images instead of one (see Section 4.2.2.1) when this is not the case. As a consequence, the outliers are not removed from the right positions (see Section 4.2.2.3), resulting in wrong disparity values. The visual consequences vary, in regions where there are disparity transitions this effect is very noticeable and in regions where there aren't many disparity transitions, the effect can vary depending on the amount of detail.

The marked region at the right – This region is shown amplified in Figure 6.11. In this side by side comparison, Figure 6.11a, in contrast to Figure 6.11b, presents some noticeable artefacts. These artefacts are caused by a limitation of the NCC method (see Section 4.2.2.1) that manifests when a scene is highly uniform. A consequence of this is a drop in the NCC efficiency when calculating similarities among neighbour micro-images. The similarities are so many (nearly identical), that they are confused with error when the NCC series are averaged to become the Similarity function (see Section 4.2.2.1). Although the statistical analysis performed during the DAPBe method (see Section 4.2.2.3) should absorb this, when a region is highly regular this issue arises because all similarity functions in those regions have bad disparity values. With little to none good values to base the statistical optimization, the errors cannot be concealed and result in the presence of these artefacts.



(a) (b)

Figure 6.9- 2D Images extracted from the Dino2 holoscopic test image using: (a) DAPBe method (all-in-focus), MOS rated 2,31; (b) SSPBe method (focused on the orange dinosaur at the back (top right corner)), MOS rated 2,13

(a) (b)

Figure 6.10 - A section from the dinosaur on the left, from the Dino2 resource; (a) image extracted with the DAPBe method, all-in-focus, showing artefacts at the edges of objects and on the objects; (b) image extracted with the DAPBe method, focused on the dinosaur in the back, showing artefacts covered with a blurring effect

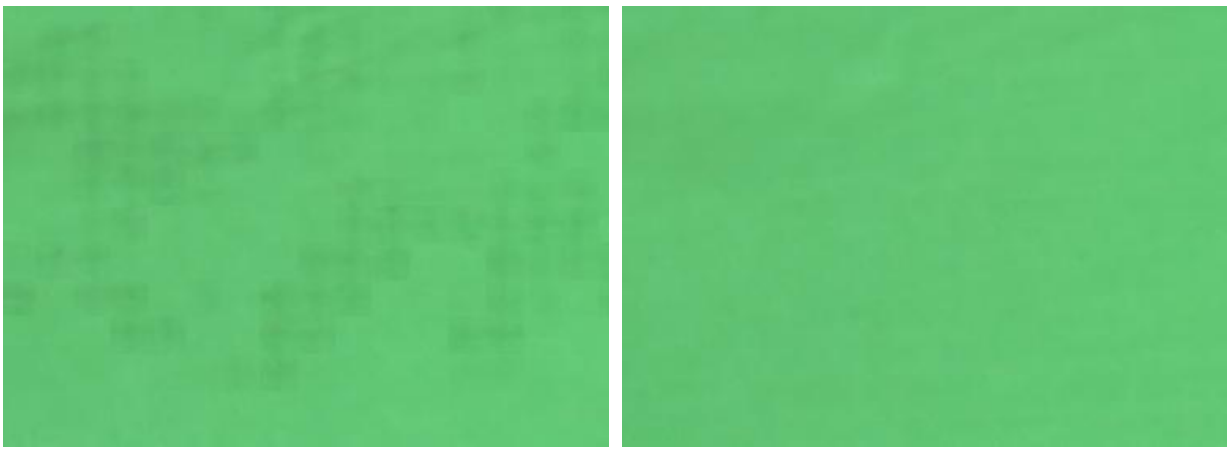


Figure 6.11 - A section of the green screen, from the Dino2 resource; (a) image extracted with the DAPBe method, all-in-focus, showing artefacts in smooth regions of the image; (b) image extracted with the DAPBe method, focused on the dinosaur in the back, showing no artefacts in smooth regions of the image

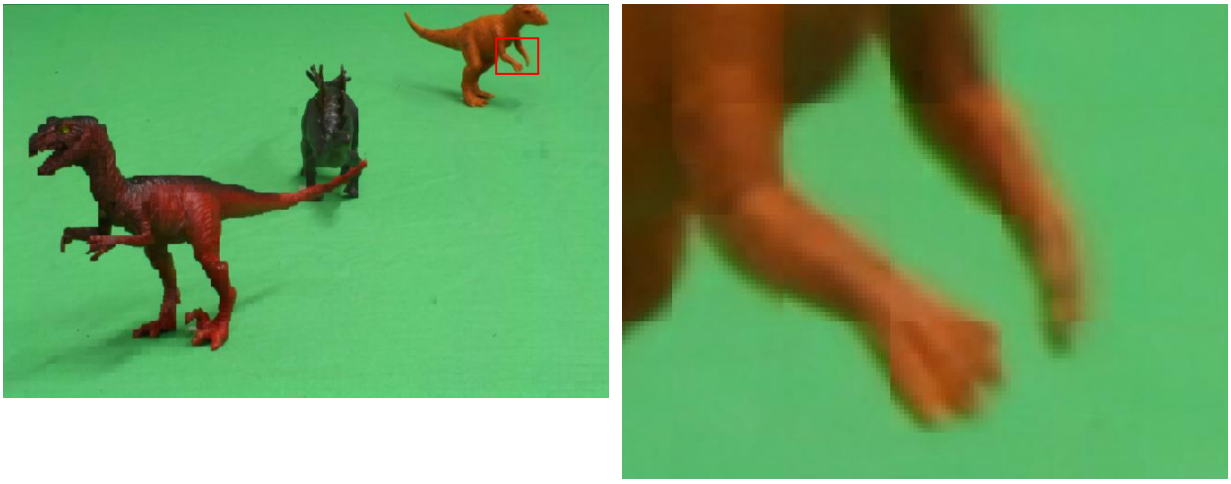


Figure 6.12 - Extractions performed on the Dino2 resource, focuses on the orange dinosaur on the top right region, by the SSPe method; (a) is the extraction with a portion of the in focus region marked; (b) the amplification of the square region marked in (a), showing poor adaptations between patches, resulting in artefacts

Through these observations immerge the facts that

The SSPBe method deals better with slight differences in micro-image geometry than the DAPBe method, through its blurring mechanism;

The DAPBe method does not deal as well with smooth regions in comparison with the SSPBe method because the NCC (see Chapter 4) that scans the holoscopic image for similarities, which in turn will provide a basis for the DAPBe method to make decisions on how to build the 2D reconstruction, finds a lot of similarities in smooth regions that both are the same and are not the same region. These errors reflect in the reconstruction process, resulting in the artefacts present in Figure 6.1.

In conclusion

The VSBe and the SSPe typically perform below average and for this reason can't be considered the best methods.

The DAPBe method is the only tested method able to produce *Excellent* quality 2D extractions, notably an extraction method that constructs a 2D representation of a scene with the same large depth of field of each micro-lens, typically called *All-In-Focus*. For its performance and algorithm of extraction, this author considers the DAPBe method the best in analysis. The DAPBe method is the one that extracts 2D image representation of a scene more faithfully to the original scene and therefore the best choice.

However, there are still subjects in the DAPBe method that need to be corrected for it to always perform better than the SSPBe method, namely:

The capacity to compensate for elliptical micro-images or not perfectly square micro-images;

The capacity to compensate for errors in smooth regions of a scene.

Figure 6.13 through Figure 6.17 (originals at [65])present the 2D extractions of the available content captured by the Plenoptic 2.0 holoscopic camera (see chapter 2). These All-in-Focus extractions serve as examples of the capability of the DAPBe method, by applying it to holoscopic images that have micro-images with a uniform structure.

Figure 6.13 - Fredo resource, reconstructed by the DAPBe method

Figure 6.14 - Jeff resource, reconstructed by the DAPBe method

Figure 6.15 - Laura resource, reconstructed by the DAPBe method

Figure 6.16 - Seagull resource, reconstructed by the DAPBe method

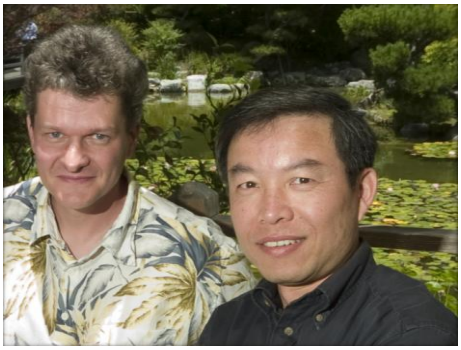


Figure 6.17 - Sergio resource, reconstructed by the DAPBe method

Figure 6.18 - Zhengyun1 resource, reconstructed by the DAPBe method

6.1.2 Using Objectives Scores

This subsection presents the objective scores of the extractions. After that the scores are analyzed to find the best extraction method according to the objective IQA metrics. The scores can be found in Table 6.5, which contains all scores for the NIQE, AQI and BRISQUE objective metrics. The best MOS scored extraction, with the respective objective metric score, is marked bold for each objective metric. The images are ordered with the best scored extractions at the top and the worst scored at the bottom.

Table 6.5 - Objective test scores for the 2D extractions submitted for testing

Extraction	NIQE	Extraction	AQI (10^6)	Extraction	BRISQUE
Composition1 - SSPe(Spiderman)	12,94	Composition1 - SSPe(Spiderman)	2300,76	Fountain - SSPe(Fountain)	24,99
Composition1 - SSPe(Dino)	12,97	Composition1 - SSPe(Background)	2208,85	Composition1 - SSPe(Spiderman)	25,21
Composition1 - SSPBe(Spiderman)	13,29	Fountain - SSPe(Fountain)	2106,49	Composition1 - SSPe(Dino)	26,67
Composition1 - SSPe(Hood)	13,42	Composition1 - SSPe(Dino)	2008,74	Dino2 - SSPe(Fig. Front-Centre)	33,16
Composition1 - SSPBe(Dino)	13,60	Composition1 - SSPe(Plane)	1962,79	Fountain - SSPBe(Fountain)	35,32
Composition1 - DAPBe(All)	13,72	Plane129 - SSPe(Propeller)	1944,38	Dino2 - SSPe(Fig. Middle)	35,39
Composition1 - SSPe(Plane)	13,93	Plane129 - SSPe(Support)	1918,79	Dino2 - SSPe(Fig. Back)	36,01
Composition1 - SSPBe(Hood)	14,14	Composition1 - SSPe(Hood)	1870,88	Composition1 - SSPBe(Spiderman)	36,51
Dino2 - SSPe(Fig. Front-Centre)	14,29	Plane129 - SSPe(Wing)	1845,16	Fountain - SSPe(Water Splash)	37,69
Composition1 - SSPe(Background)	14,46	Plane129 - SSPe(Doll)	1842,62	Composition1 - SSPe(Hood)	38,09
Dino2 - SSPe(Fig. Middle)	14,56	Fountain - SSPe(Water Splash)	1760,48	Composition1 - DAPBe(All)	39,64
Composition1 - SSPBe(Plane)	14,61	Fountain - SSPe(Trees)	1525,43	Dino2 - SSPBe(Fig. Front-Centre)	42,83
Plane129 - SSPBe(Doll)	14,68	Fountain - SSPe(Background)	1447,72	Fountain - SSPe(Trees)	44,28
Demichelis_cut - DAPBe(All)	14,68	**Fountain - DAPBe(All)**	**1268,07**	Composition1 - SSPBe(Dino)	46,47
Dino2 - DAPBe(All)	14,68	Demichelis_cut - SSPe(Demichelis)	1223,24	Fountain - SSPBe(Water Splash)	47,71
Dino2 - SSPBe(Fig. Front-Centre)	14,72	Composition1 - SSPBe(Spiderman)	1192,76	Dino2 - SSPBe(Fig. Middle)	47,89
Plane129 - SSPe(Propeller)	14,73	Composition1 - DAPBe(All)	1171,83	Dino2 - SSPBe(Fig. Back)	48,11
Plane129 - SSPBe(Wing)	14,75	Fountain - SSPBe(Trees)	1079,61	Composition1 - SSPe(Plane)	49,53
Dino2 - SSPBe(Fig. Middle)	14,78	Fountain - SSPBe(Water Splash)	1016,51	**Fountain - DAPBe(All)**	**49,83**
Dino2 - SSPe(Fig. Back)	14,79	Fountain - SSPBe(Background)	1011,77	Fountain - SSPe(Background)	50,06
Composition1 - SSPBe(Background)	14,81	Composition1 - SSPBe(Dino)	921,45	Dino2 - DAPBe(All)	51,86
Plane129 - SSPBe(Propeller)	14,84	Fountain - SSPBe(Fountain)	899,25	Fountain - SSPBe(Trees)	53,15
Plane129 - SSPe(Wing)	14,86	Composition1 - SSPBe(Hood)	565,78	Fountain - SSPBe(Background)	53,86
Dino2 - SSPBe(Fig. Back)	14,89	Plane129 - SSPBe(Propeller)	366,956	Demichelis_cut - SSPe(Demichelis)	54,62
Demichelis_cut - SSPBe(Demichelis)	14,91	Composition1 - SSPBe(Plane)	363,13	Composition1 - SSPe(Background)	55,79
Plane129 - SSPBe(Support)	14,91	Plane129 - SSPBe(Wing)	356,739	Plane129 - SSPe(Propeller)	56,06
Plane129 - SSPe(Doll)	14,94	Dino2 - SSPe(Fig. Back)	351,47	Composition1 - SSPBe(Hood)	57,59
Composition1 - VSBe(Plane)	15,10	Plane129 - SSPBe(Doll)	319,034	Demichelis_cut - SSPBe(Demichelis)	58,31
Composition1 - VSBe(Background)	15,17	Dino2 - SSPe(Fig. Front-Centre)	300,49	Plane129 - SSPe(Wing)	59,02
Composition1 - VSBe(Hood)	15,25	Dino2 - SSPe(Fig. Middle)	298,51	Demichelis_cut - DAPBe(All)	60,29
Plane129 - SSPe(Support)	15,26	Plane129 - DAPBe(All)	256,443	Plane129 - SSPe(Doll)	60,85
Plane129 - DAPBe(All)	15,28	Plane129 - SSPBe(Support)	254,094	Plane129 - SSPe(Support)	61,88
Fountain - SSPBe(Background)	15,40	Composition1 - SSPBe(Background)	238,54	Plane129 - SSPBe(Propeller)	69,39
Composition1 - VSBe(Dino)	15,59	Composition1 - VSBe(Background)	215,91	Plane129 - SSPBe(Wing)	70,27
Fountain - SSPBe(Trees)	15,70	Composition1 - VSBe(Plane)	214,89	Plane129 - SSPBe(Doll)	70,72
Demichelis_cut - SSPe(Demichelis)	15,83	Plane129 - VSBe(Plane)	156,68	Composition1 - SSPBe(Plane)	71,01
Fountain - SSPe(Water Splash)	15,90	Dino2 - SSPBe(Fig. Front-Centre)	139,43	Plane129 - SSPBe(Support)	73,41
Fountain - SSPe(Fountain)	15,94	Dino2 - DAPBe(All)	114,36	Composition1 - SSPBe(Background)	75,42
Fountain - DAPBe(All)	**16,00**	Composition1 - VSBe(Hood)	108,42	Composition1 - VSBe(Background)	76,59
Dino2 - VSBe(Fig. Middle)	16,06	Dino2 - SSPBe(Fig. Middle)	103,11	Plane129 - DAPBe(All)	77,06
Composition1 - VSBe(Spiderman)	16,06	Demichelis_cut - SSPBe(Demichelis)	88,16	Fountain - VSBe(Fountain)	77,74
Plane129 - VSBe(Plane)	16,13	Demichelis_cut - DAPBe(All)	80,14	Composition1 - VSBe(Plane)	77,87
Dino2 - VSBe(Fig. Back)	16,16	Dino2 - SSPBe(Fig. Back)	73,31	Fountain - VSBe(Water Splash)	79,22
Fountain - VSBe(Water Splash)	16,23	Fountain - VSBe(Trees)	60,96	Fountain - VSBe(Trees)	81,28
Fountain - SSPBe(Water Splash)	16,23	Fountain - VSBe(Background)	55,17	Composition1 - VSBe(Hood)	81,83
Fountain - VSBe(Trees)	16,25	Fountain - VSBe(Water Splash)	45,14	Fountain - VSBe(Background)	82,80
Fountain - VSBe(Fountain)	16,33	Fountain - VSBe(Fountain)	30,22	Composition1 - VSBe(Dino)	87,88
Fountain - SSPBe(Fountain)	16,39	Composition1 - VSBe(Dino)	28,91	Demichelis_cut - VSBe(Demichelis)	88,32
Fountain - VSBe(Background)	16,39	Dino2 - VSBe(Fig. Back)	6,85	Composition1 - VSBe(Spiderman)	92,09
Demichelis_cut - VSBe(Demichelis)	16,58	Dino2 - VSBe(Fig. Middle)	5,48	Dino2 - VSBe(Fig. Back)	92,26
Dino2 - VSBe(Fig. Front)	16,81	Composition1 - VSBe(Spiderman)	4,29	Dino2 - VSBe(Fig. Middle)	93,52
Fountain - SSPe(Trees)	16,84	Demichelis_cut - VSBe(Demichelis)	2,13	Plane129 - VSBe(Plane)	94,82
Fountain - SSPe(Background)	17,52	Dino2 - VSBe(Fig. Front)	1,40	Dino2 - VSBe(Fig. Front)	102,48

Analysis I – Direct Comparison with Subjective Scores

This analysis consists of comparing the three rating methods, AQI, NIQE, and BRISQUE (see Section 5.2.1) directly with the obtained MOS scores (see Section 5.3). This will be done by direct comparison to check if the best extraction rated subjectively is also rated as the best extraction method by any of these three metrics.

Looking at Table 6.5, where the highest MOS rated 2D extraction is marked bold, it becomes apparent that neither of the three metrics agree with the MOS in terms of the best method. In fact, all of them indicate an extraction performed by the SSPe method as the best. This method, as analyzed in Section 6.1.1, typical performs below average in terms of perceived quality.

According to the obtained scores, the NIQE and AQI methods agree that the extraction performed by the SSPe method on the Composition1 resource, with focus on the Spiderman doll, (see Figure 6.19) is the extraction with the best quality. A close up of the plane featured in the scene can be found in Figure 6.20 for consideration.

Figure 6.19 – 2D extracted image Rated as the best extraction by the NIQE and AQI method

Figure 6.20 - A close up of the plane in the extraction rated as the best extraction by the NIQE and AQI methods

Also in line with the obtained scores, the BRISQUE method indicates the extraction performed by the SSPe method on the Fountain resource, with focus on the fountain, (see Figure 6.21) is the 2D extraction with the best quality. A close up of the fountain and tree line featured in the scene can be found in Figure 6.22 for consideration.

Figure 6.21 - Rated as the best extraction by the BRISQUE method

Figure 6.22 - A close up of the tree line at the right, in the extraction rated as the best extraction, by the BRISQUE method

Because the objective metrics do not seem to have a direct correlation with the MOS ratings presented in Section 6.1.1, without further analysis, the only conclusion that can be drawn based only

on the data from the objective metric is that, according to the objective metrics the SSPe method is the best performer.

6.1.3 Comparing Subjective and Objective Scores: Conclusions

Comparing the scores presented in the previous Section 6.1.1 and 6.1.2, a big discrepancy is evident between subjective and objective results. Not only the best MOS rated extraction don't match with the best objective rated extractions, the respective worst rated extractions don't match either. Moreover, the underperformer SSPe method seems to be the best according to all three objective metrics.
Using the visual examples provided in the previous two sub-sections, with a simple visual inspection supported by the MOS ratings, it becomes apparent that the objective ratings are not adjusted to human perceived quality.

6.2 Comparing the Objective Metrics Performance: Scores and Analysis

Based on the result obtained for the objective test scores and the apparent lack of connections between them and the subjective test scores further analysis is required to assess how much the objective and subjective test scores correlate with each other.
Ideally, the analysis that would reveal the most information would be to calculate the correlation between the extractions of a single holoscopic image, extracted by a single 2D extraction method, with the MOS scores. This grouping of data would eliminate more variables, however there isn't enough material to perform a statistically relevant analysis of that kind because the groups would have between 4 and 1 image each.
Since one of the objectives of this thesis is to find an objective metric that correlates best with human perception, this type of analysis is going to be performed on 4 groups of images, presented in Table 6.6 through Table 6.9, each containing all images extracted by a single extraction method.

Table 6.6 – Normalized test scores for subjective and objective tests performed on DAPBe extracted images

Image	MOS Score	AQI Score	NIQE Score	BRISQUE Score
Demichelis_cut - DAPBe - All	0,289	0,034	0,619	0,495
Composition16 - DAPBe - All	0,297	0,509	0,829	0,790
Plane138 - DAPBe - All	0,461	0,111	0,488	0,254
Dino11 - DAPBe - All	0,531	0,049	0,619	0,615
Fountain - DAPBe - All	0,961	0,551	0,331	0,644

Table 6.7 - Normalized test scores for subjective and objective tests performed on SSPBe extracted images

Image	MOS Score	AQI	NIQE	BRISQUE
Composition11 - SSPBe - Background	0,164	0,103	0,591	0,278
Dino10 - SSPBe - Figure Front-Centre	0,227	0,060	0,610	0,745
Plane136 - SSPBe - Wing	0,234	0,154	0,605	0,352
Composition15 - SSPBe - Spiderman	0,242	0,518	0,924	0,835
Plane137 - SSPBe - Propeller	0,242	0,159	0,585	0,364
Demichelis_cut - SSPBe - Demichelis	0,305	0,037	0,570	0,523

Composition13 - SSPBe - Hood	0,422	0,245	0,737	0,533
Composition12 - SSPBe - Plane	0,422	0,157	0,635	0,341
Composition14 - SSPBe - Dino	0,430	0,400	0,856	0,692
Dino9 - SSPBe - Figure Middle	0,500	0,044	0,598	0,672
Plane134 - SSPBe - Support	0,500	0,110	0,570	0,307
Plane135 - SSPBe - Doll	0,516	0,138	0,620	0,345
Dino8 - SSPBe - Figure Back	0,578	0,031	0,574	0,669
Fountain - SSPBe - Fountain	0,602	0,390	0,247	0,852
Fountain - SSPBe - Trees	0,688	0,469	0,397	0,597
Fountain - SSPBe - Background	0,727	0,439	0,463	0,587
Fountain - SSPBe - Water Splash	0,750	0,441	0,280	0,675

Table 6.8 - Normalized test scores for subjective and objective tests performed on SSPe extracted images

Image	MOS Score	AQI	NIQE	BRISQUE
Fountain - SSPe - Trees	0,672	0,663	0,148	0,724
Fountain - SSPe - Water Splash	0,602	0,765	0,352	0,818
Fountain - SSPe - Background	0,555	0,629	0,000	0,641
Fountain - SSPe - Fountain	0,492	0,915	0,343	1,000
Dino7 - SSPe - Figure Front-Centre	0,250	0,130	0,705	0,883
Composition9 - SSPe - Dino	0,234	0,873	0,994	0,976
Composition7 - SSPe - Plane	0,234	0,853	0,783	0,649
Dino5 - SSPe - Figure Back	0,234	0,152	0,596	0,842
Dino6 - SSPe - Figure Middle	0,234	0,129	0,645	0,851
Plane130 - SSPe - Support	0,094	0,834	0,492	0,472
Plane132 - SSPe - Wing	0,094	0,802	0,579	0,513
Composition10 - SSPe - Spiderman	0,086	1,000	1,000	0,997
Composition8 - SSPe - Hood	0,086	0,813	0,895	0,812
Plane131 - SSPe - Doll	0,086	0,801	0,563	0,486
Composition6 - SSPe - Background	0,008	0,960	0,667	0,559
Plane133 - SSPe - Propeller	0,008	0,845	0,609	0,555
Demichelis_cut - SSPe - Demichelis	0,008	0,531	0,368	0,576

Table 6.9 - Normalized test scores for subjective and objective tests performed on VSBe extracted images

Image	MOS Score	AQI	NIQE	BRISQUE
Composition3 - VSBe - Hood	0,242	0,046	0,494	0,186
Fountain - VSBe - Water Splash	0,234	0,019	0,282	0,223
Fountain - VSBe - Trees	0,094	0,026	0,277	0,194
Composition2 - VSBe - Plane	0,078	0,093	0,527	0,243
Fountain - VSBe - Fountain	0,016	0,012	0,258	0,245
Dino3 - VSBe - Figure Middle	0,016	0,001	0,318	0,019
Composition1 - VSBe - Background	0,008	0,093	0,512	0,261
Composition4 - VSBe - Dino	0,008	0,012	0,420	0,099
Demichelis_cut - VSBe - Demichelis	0,008	0,000	0,205	0,093
Composition5 - VSBe - Spiderman	0,008	0,001	0,318	0,039
Dino2 - VSBe - Figure Back	0,008	0,002	0,297	0,037
Fountain - VSBe - Background	0,000	0,023	0,245	0,172
Plane129 - VSBe - Plane	0,000	0,067	0,302	0,000

Dino4 - VSBe - Figure Front	0,000	0,000	0,154	0,000

Analysis I – Correlation of Objective Metrics with 2D Extraction Metrics

The purpose of this analysis is to assert how linearly dependent the objective test scores are, in relation to each metric, based on the subjective MOS. The first step is to group the data for each 2D extraction method (see Figure 6.23 though Figure 6.26).

Figure 6.23 groups the data for the DAPBe method, Figure 6.24 for the SSPBe method, Figure 6.25 for the SSPe method and Figure 6.26 for the VSBe method. The values of the ratings are normalized and sorted by the subjective score, from lowest to highest. AQI, BRISQUE and NIQE are all normalized according to the highest and lowest value observed in testing. The MOS scores are normalized with the maximum and minimum value of the ACR scale, 0 and 4 (see Chapter 5).

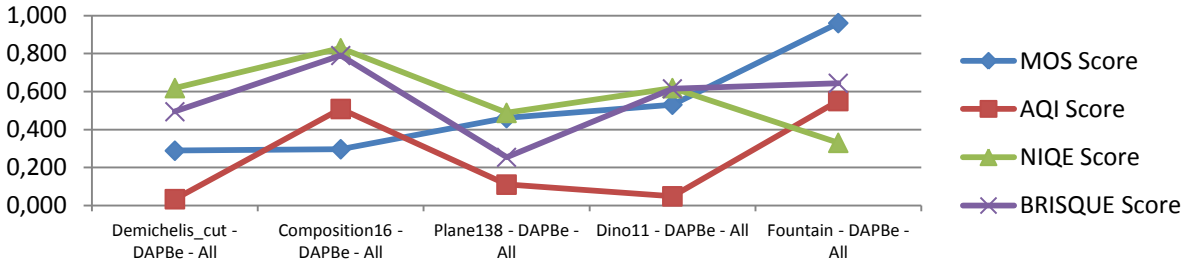


Figure 6.23 – Scores for the DAPBe method

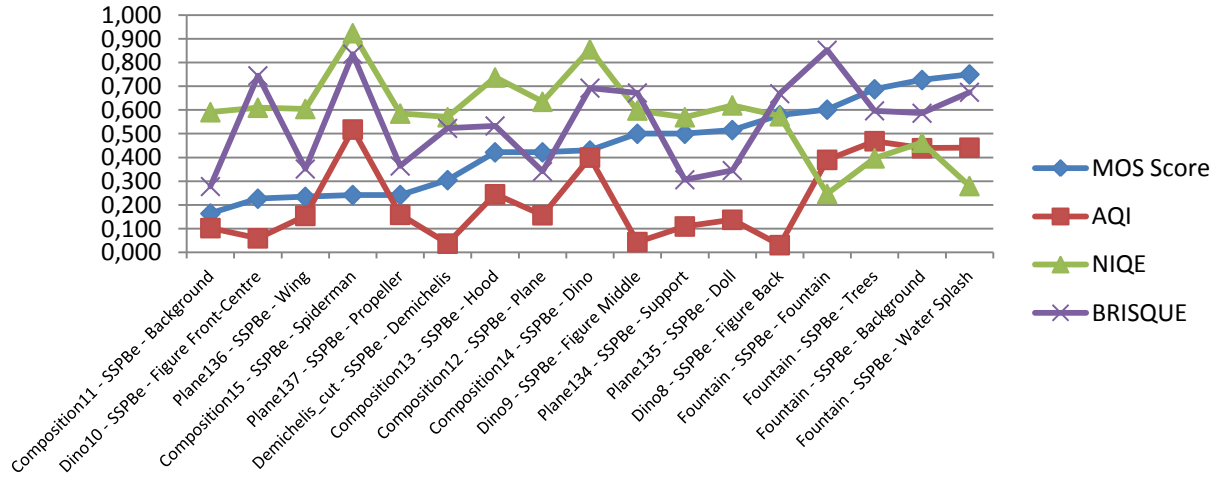


Figure 6.24 – Scores for the SSPBe method

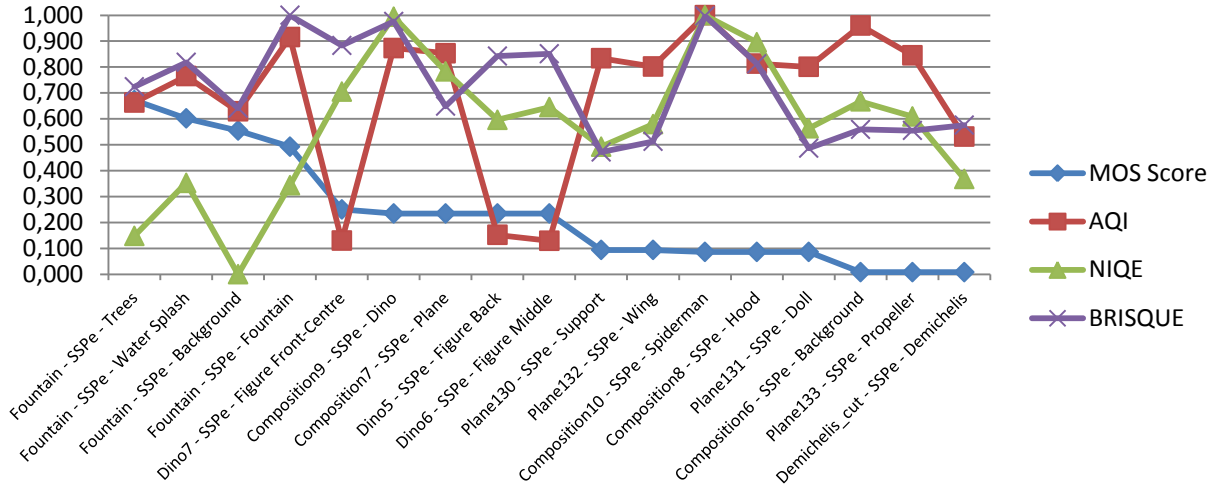


Figure 6.25 – Scores for the SSPe method

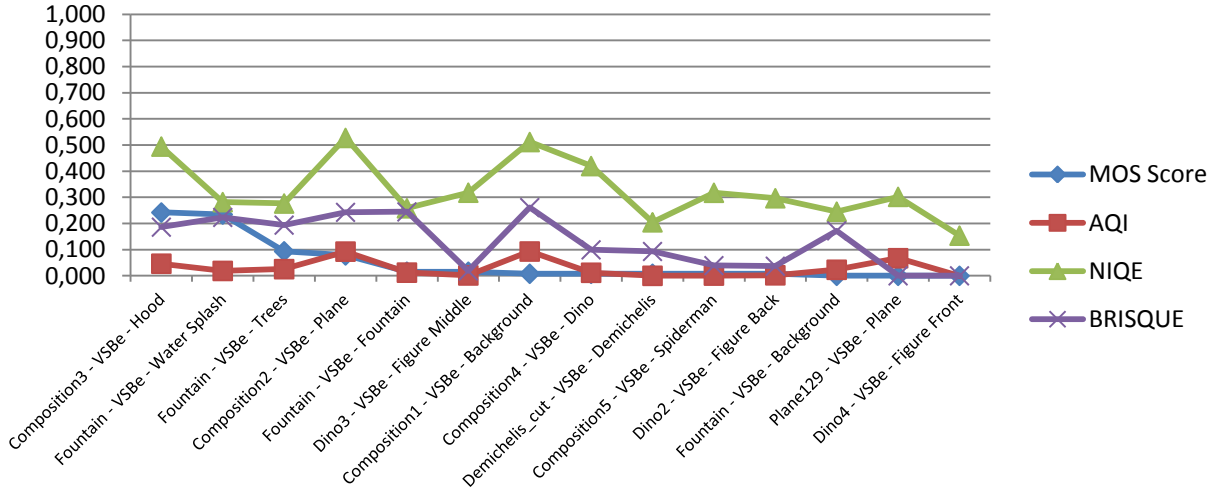


Figure 6.26 – Scores for the VSBe method

The Pearson product-moment correlation coefficient [66] is a measurement of the linear correlation between two parameters. It can assume values between -1 and 1, where 0 indicates no linear correlation between the two parameters, 1 indicates perfect linear correlation between the two parameters and -1 indicates perfect inverted correlation.

To calculate a Pearson product-moment correlation coefficient $\rho_{X,Y}$ the covariance between the two parameters $cov(X,Y)$ is calculated and divided by the product of the means of both parameters $\sigma_X \sigma_Y$. The formula is presented on Equation (32).

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} \qquad (32)$$

The Spearman's rank correlation coefficient [67] is also a measurement of linear correlation between two parameters. It is calculated using the same formula used for the Pearson product-moment correlation coefficient (see Equation (32)) and assumes the same values with the same meanings. The difference is that instead of plugging in the values from a parameter X and Y, the values are sorted in a proportional order and the values of X and Y are substituted by the value 1 for the first position, the value 2 for the second position and so on, for both parameters. In the case where a

parameter has duplicate, triplicates, and so on, values, consecutive ranks that should be attributed to those repeated values are averaged and the average attributed to the repeated values.

Based on the values presented on Table 6.6 to Table 6.9 comes the second step of this analysis, calculating the correlation. The Spearman's Rank Correlation coefficients and the Pearson Product-Moment Correlation Coefficients were calculated between the objective and the subjective ranks and scores respectively to determine if a hidden linear dependence exists.

Both these linear correlation metrics, Spearman and Pearson, are calculated, the X parameter corresponding to the subjective MOS values and Y to one of the three objective IQA metrics. The scores obtained are presented in Table 6.10.

Table 6.10 – Spearman and Pearson correlation between the MOS and the objective test scores, for each extraction metric

Method	IQA	Spearman	Pearson
DAPBe	AQI	0,60	0,46
	NIQE	0,80	0,84
	BRISQUE	-0,20	-0,09
SSPBe	AQI	0,34	0,44
	NIQE	0,59	0,61
	BRISQUE	-0,27	-0,29
SSPe	AQI	-0,34	-0,13
	NIQE	0,43	0,62
	BRISQUE	-0,45	-0,29
VSBe	AQI	0,28	0,16
	NIQE	-0,37	0,36
	BRISQUE	-0,56	-0,45

A good result for the Pearson or Spearman correlations is typically above 0,9 or below -0,9, as can be noted by the results presented for all objective metrics considered, in [52], [53], [54], [55], [56], [58], [59] and [60]. Based on the scores presented in Table 6.10 that mark is never reached. In this regard there are grounds to conclude that all metrics have poor correlation with the MOS.

There is however enough data to choose the one that best correlates to the MOS ratings:

- For the DAPBe and SSPBe method - The NIQE rating is the best according to both Spearman and Pearson correlation;
- For the SSPe method – The BRISQUE rating is the best according to the Spearman correlation and the NIQE rating is the best according to the Pearson correlation;
- For the VSBe method - The BRISQUE rating is the best according to both Spearman and Pearson correlation;

Based on the fact that the NIQE ratings correlate best with MOS ratings within each group of single extraction method images, the NIQE metric is chosen as the Objective IQA metric that best correlates with the human perception of quality.

.

7
Conclusions and Future Work

3D holoscopic technology has arrived at the consumer market and soon will hit the professional market as a consequence of ambitious projects like the 3D VIVANT Project. With the appearance of a company (and its products) called Lytro, everybody can now buy a 3D holoscopic camera and produce his own holoscopic content. However, because the display technology for 3D holoscopic images is not yet at an affordable price for the common buyer, Lytro had to limit its products to produce 2D images only. However, this should change in the future. For this change to happen, 3D holoscopy has not only to deliver a better experience than the previous 2D and 3D technologies, but it will also have to find some degree of compatibility with them. The currently available conversion methods to extract 2D images from 3D holoscopic images, solving part of the compatibility problem, still have performance issues. In this regard, there still is significant room for improvement.

Related to this issue, there is also the need to find quality metrics able to assess the perceivable quality of 2D extracted images, avoiding performing time consuming subjective quality tests with real people to score 2D extracted images. This kind of tool would play an important role on both 3D holoscopic technology development and deployment.

7.1 Conclusions

In this context, this work has made the following contributions:

- **A novel, fully automated method for the extraction of 2D images from 3D holoscopic images** – The 2D extraction method proposed in this thesis – *the Disparity-Assisted Patch Blending 2D extraction method* – is capable of both estimating the scene relative depth and, following that estimate, to extract a 2D representation of the scene based only on the specifications of the capturing camera and the 3D holoscopic capture itself. The process requires no human intervention during the reconstructions process and no input regarding any of the scenes depth properties. Moreover, the proposed 2D extraction solution outperforms, in "normal" conditions, all the available alternative methods; however, in "bad" conditions, it generates unwanted artefacts that compromise its performance making it similar to the other methods (in some cases, it may even be beaten by the SSPBe extraction method);
- **Potential No-Reference Image Quality Assessment objective metrics able to reliably rate 2D extractions from 3D holoscopic images** – While none of the no-reference objective image quality assessment metrics, the AQI, NIQE and BRISQUE, has been identified as a relevant potential candidate to reliably rate 2D extractions from 3D holoscopic images, NIQE seems to be promising. However, it is important to point out that both NIQE and BRISQUE need to be trained and are dependent on human opinion, although not in the same way. It may also be concluded that, with the originally trained model provided by its authors without considering 2D extractions, these NR IQA metrics can't reliably rate 2D extractions from 3D holoscopic images.

7.2 Future Work

Upon implementation, experimentation and review of the test results, a series of enhancements have been identified for the technologies proposed and identified in this Thesis. To cover these enhancements, this section is divided in the two themes discussed: NR IQA metrics and 2D Extraction methods.

7.2.1 2D Extraction Methods

Further development of the proposed DAPBe method can follow different paths. There are, however, advances that should improve the number of scenarios in which it performs better than the other available extraction methods:

- **Detection of the holoscopic images structure** – Currently, the DAPBe method assumes perfectly square micro-images as input. Moreover, the size of the squares has to be provided manually. An automatic mechanism to detect the shape of each micro-image will undoubtedly improve the quality of the 2D reconstruction in the cases where the SSPBe method is better than the DAPBe method;
- **Improved focusing** – The difficult job of reconstructing a scene as represented in the original 3D holoscopic image is done by the DAPBe method. There are, however, features like the blurring effect available in other methods (the SSPBe method for instance), that may be interesting to have in the DAPBe method too. In this case, the burring would not be used to conceal artefacts but only to simulate aperture adjustment, changing the depth of field;

- **Variable PoV testing** - Although a mechanism to vary the PoV was defined, it was not properly tested; thus researching adequate test cases to consolidate this mechanism is still required;
- **Improving the Maximum-Likelihood estimator** – The estimation method employed in this Thesis delivers very good performance. Although the Maximum-Likelihood estimator was tested, a mixed estimation framework might provide greater performance using this estimation technique;
- **Reducing the extraction method complexity** – As it is, the proposed 2D extraction algorithm requires intensive processing of the holoscopic images to assess the disparities between micro-images. Experiments with other disparity detection methods could bring the complexity of this method down; further fine tuning of the current process could also prove fruitful;
- **Application to holoscopic video content** – All tests have been conducted using still images or video frames. Testing the DAPBe and SSPBe methods with video content by generating 2D extracted video sequences is the next big step for these algorithms.

7.2.2 No-Reference Image Quality Assessment Metrics

It has been concluded that the reviewed NR IQA metrics have in fact some degree of sensitivity to the distortions present in 2D extracted images. However, although they are sensitive to them, they fail to properly assess the impact of those distortions in the perceived quality. In this regard, the following subjects were identified as possible future work:

- **Training NR IQA metrics with 2D extracted images** – The NIQE metric is image structure oriented, requiring training in the structure of subjectively highly rated image cases. Thus, it should be productive to input very highly (subjectively) rated 2D extracted images ("perfect images") in the training phase, and compare its behaviour with the previous training model. BRISQUE is also image structure oriented and requires a training phase but it accepts all type of images, highly and poorly rated ones. Training BRISQUE on 2D extractions only may also prove productive;
- **Modifying Distortion Unaware NR IQA metric for 2D extracted images** - Both BRISQUE and NIQE are distortion unaware. In that regard, it may produce some interesting results to modify BRISQUE and NIQE to be sensitive to distortions identified in 2D extracted images;
- **Proposing a Distortion Aware NR IQA metric for 2D extracted images** - Although AQI is distortion oriented, it only looks at anisotropy to rate images. A new metric entirely, or a hybrid metric incorporating the AQI method, may prove productive to properly assess perceivable image quality in 2D extracted images.

Bibliography

[1] M. Faraday, “Thoughts on Ray Vibrations,” *The London, Edinburgh and Dublin Philosophical Magazine and Journal of Science*, vol. 28, pp. 345–350, 1846.

[2] E. H. Adelson and J. R. Bergen, “The Plenoptic Function and the Elements of Early Vision,” *Computation Models of Visual Processing*, pp. 3–20, 1991.

[3] C. DeCusatis, “APPENDIX I The SI system and SI units for Radiometry and photometry,” in *Handbook of Applied Photometry*, American Inst. of Physics, 1997, p. 484.

[4] J. E. Greivenkamp, “Physical optics,” *Journal of the Optical Society of America A*, vol. 28, no. 2, pp. 2–44, 2011.

[5] S. McHugh, “Cambridge In Colour,” 2011. [Online]. Available: http://www.cambridgeincolour.com/tutorials/depth-of-field.htm.

[6] Hamster, “Where’s the moolah?,” 2012. [Online]. Available: http://moolahdb.wordpress.com/2012/06/12/parallel-processing-with-a-human-brain/. [Accessed: 13-Feb-2013].

[7] C. Wheatstone, “Contributions to the Physiology of Vision. Part the First. On Some Remarkable, and Hitherto Unobserved, Phenomena of Binocular Vision,” *Philosophical Transactions of the Royal Society of London*, vol. 128, no. 1838, pp. 371–394, 1838.

[8] L. da Vinci, “Trattato della pittura,” *Nella Stamperia de Romanis*, pp. 43 – 511, 1817.

[9] T. Georgiev, “Todor Georgiev Personal Website,” 2012. [Online]. Available: http://tgeorgiev.net/. [Accessed: 27-Nov-2012].

[10] P. Moon and D. Eberle Spencer, “Theory of the photic field,” *Journal of the Franklin Institute*, vol. 255, no. 1, pp. 33–50, 1953.

[11] M. Levoy and P. Hanrahan, “Light Field Rendering,” in *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques SIGGRAPH 96*, 1996, pp. 31–42.

[12] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen, “The lumigraph,” in *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques SIGGRAPH 96*, 1996, vol. 30, no. Annual Conference Series, pp. 43–54.

[13] A. Lumsdaine and T. Georgiev, “Full Resolution Lightfield Rendering,” in *SIGGRAPH*, 2008, no. January, pp. 1–12.

[14] G. Lippmann, “Epreuves Reversibles Donnant la Sensation du Relief,” *Journal de Physique Théorique et Appliquée*, vol. 7, no. 1, pp. 821–825, 1908.

[15] A. Aggoun, E. Tsekleves, D. Zarpalas, A. Dimou, P. Daras, P. Nunes, and L. D. Soares, “Immersive 3D Holoscopic Video System,” *IEEE Multimedia*, 2012.

[16] A. Lumsdaine and T. Georgiev, “The Focused Plenoptic Camera,” 2009.

[17] T. Georgiev, “New results on the Plenoptic 2.0 camera,” in *2009 Conference Record of the Forty-Third Asilomar Conference on Signals, Systems and Computers*, 2009, pp. 1243–1247.

[18] Lytro, “Lytro Website,” 2012. [Online]. Available: https://www.lytro.com/. [Accessed: 26-Nov-2012].

[19] Raytrix, “Raytrix Website,” 2012. [Online]. Available: http://www.raytrix.de/. [Accessed: 26-Nov-2012].

[20] 3DVivant, “3D Vivant Website,” 2012. [Online]. Available: http://www.3dvivant.eu/. [Accessed: 27-Nov-2012].

[21] Arri, “Arri Alexa Website,” *12/12/12*, 2012. [Online]. Available: http://www.arri.com/camera/digital_cameras/cameras/camera_details.html?no_cache=1&product=9&cHash=a8f59e1416. [Accessed: 11-Dec-2012].

[22] W. Rollmann, “Zwei neue stereoskopische Methoden,” *Zeitschrift fur Naturwiss*, vol. 166, no. 9, pp. 186–187, 1853.

[23] M. 3D, “Mundo 3D Blog,” 2012. [Online]. Available: http://3dtudo3d.blogspot.pt/. [Accessed: 27-Nov-2012].

[24] L. Hammond, “Televiev,” U.S. Patent 1,435,5201922.

[25] Anandteck, “Anandteck Website,” 2012. [Online]. Available: http://www.anandtech.com/show/4364/computex-2011-nvidia-announces-wired-3d-vision-glasses. [Accessed: 27-Nov-2012].

[26] S. D. Brewster, “On the construction of kaleidoscopes wich combine the colours and forms produced by polarized light,” *The Kaleidoscope, Its History, Theory, and Construction*, 1879.

[27] R. Zone, “The 3D Zone: Its Past & Its Future,” *Creative Cow Magazine*, 2012. [Online]. Available: http://magazine.creativecow.net/article/a-creative-cow-magazine-extra-the-3d-zone. [Accessed: 27-Nov-2012].

[28] C. Payatagool, “Z-Dome - 3D Immersive Display,” 2008. [Online]. Available: http://www.telepresenceoptions.com/2008/02/gdc_2008_eurotouch_zdome/. [Accessed: 27-Nov-2012].

[29] Aiptek, “Portable 3D display features,” 2012. [Online]. Available: http://www.2fidelity.com/Display/3D/features/. [Accessed: 27-Nov-2012].

[30] LG., “LG. D2500N-PN Product Webpage,” 2012. [Online]. Available: http://www.lg.com/jp/monitor/lg-D2500N-PN-cinema-3d. [Accessed: 27-Nov-2012].

[31] Holografika, “Holografica Website,” 2012. [Online]. Available: http://www.holografika.com/. [Accessed: 27-Nov-2012].

[32] A. Aggoun, “3D Holoscopic video content capture, manipulation and display technologies,” in *2010 9th Euro-American Workshop on Information Optics*, 2010, pp. 1–3.

[33] A. Amar, Z. Dimitris, C. Paulo, and P. Peter, “3D VIVANT – Deliverable 4.1 - Accurate Depth Computation and Object Segmentation,” 2011.

[34] A. Amar and O. Fatah, “Depth Mapping of Integral Images Through Viewpoint Image Extraction.” 2012.

[35] T. Georgiev and A. Lumsdaine, “Superresolution with Plenoptic Camera 2 . 0,” *Camera*, no. April, pp. 1–9, 2009.

[36] “Focused Plenoptic Camera and Rendering Focused Plenoptic Camera and Rendering Todor Georgiev 1,” *Camera*, pp. 1–28, 2010.

[37] T. Georgiev, G. Chunev, and A. Lumsdaine, “Superresolution with the focused plenoptic camera,” in *SPIE 7873*, 2011, vol. 7873, no. April, p. 78730X–78730X–13.

[38] A. Lumsdaine and T. Georgieg, “Color Demosaicing in Plenoptic Cameras,” pp. 3–10, 2012.

[39] V. Vineet and P. J. Narayanan, *CUDA cuts: Fast graph cuts on the GPU*, vol. 0, no. July. Ieee, 2008, pp. 1–8.

[40] A. Lumsdaine, “An Analysis of Color Demosaicing in Plenoptic Cameras,” *Camera*.

[41] S. M. Ross, *Introduction to probability and statistics for engineers and scientists*. Academic Press, 2004, p. 624.

[42] T. Georgiev, “Plenoptic 2.0 Holoscopic Camera Resources,” 2013. [Online]. Available: http://bordalo.img.lx.it.pt/apendix1.rar. [Accessed: 15-Oct-2013].

[43] 3DVivant, “3D VIVANT Canon Holoscopic Camera (version 1) Resources,” 2013. [Online]. Available: http://bordalo.img.lx.it.pt/apendix2.rar. [Accessed: 15-Oct-2013].

[44] 3DVivant, “3D VIVANT Canon Holoscopic Camera (version 2) Resources,” 2013. [Online]. Available: http://bordalo.img.lx.it.pt/apendix3.rar. [Accessed: 15-Oct-2013].

[45] 3DVivant, “3D VIVANT Arri Alexa Holoscopic Video Camera (version 1) Resources,” 2013. [Online]. Available: http://bordalo.img.lx.it.pt/apendix4.rar. [Accessed: 15-Oct-2013].

[46] 3DVivant, “3D VIVANT Arri Alexa Holoscopic Video Camera (version 2) Resources,” 2013. [Online]. Available: http://bordalo.img.lx.it.pt/apendix5.rar. [Accessed: 15-Oct-2013].

[47] OpenGL, “OpenGL Website.” [Online]. Available: http://www.opengl.org/. [Accessed: 26-May-2013].

[48] V. July, M. Segal, and C. Frazier, “The OpenGL Graphics System : A Specification,” 2006.

[49] H. R. Sheikh, A. C. Bovik, and G. De Veciana, *An information fidelity criterion for image quality assessment using natural scene statistics.*, vol. 14, no. 12. Ieee, 2005, pp. 2117–2128.

[50] B. J. S. Armstrong and F. Collopy, “Error Measures For Generalizing About Forecasting Methods: Empirical Comparisons By J. Scott Armstrong and Fred Collopy Reprinted with permission form,” *International Journal of Forecasting*, vol. 8, no. 1, pp. 69–80, 1992.

[51] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: from error visibility to structural similarity.,” *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[52] A. Mittal, A. K. Moorthy, and A. C. Bovik, “No-reference image quality assessment in the spatial domain.,” *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 21, no. 12, pp. 4695–708, Dec. 2012.

[53] A. K. Moorthy and A. C. Bovik, *Blind Image Quality Assessment: From Natural Scene Statistics to Perceptual Quality*, vol. 20, no. 12. IEEE, 2011, pp. 3350–3364.

[54] M. Saad, A. C. Bovik, and C. Charrier, “Blind Image Quality Assessment: A Natural Scene Statistics Approach in the DCT Domain.,” *IEEE Transactions on Image Processing*, vol. 21, no. 8, pp. 3339–3352, 2012.

[55] P. Ye and D. Doermann, “No-reference image quality assessment using visual codebooks.,” *IEEE Transactions on Image Processing*, vol. 21, no. 7, pp. 3129–38, 2012.

[56] H. Tang and N. Joshi, “Learning a Blind Measure of Perceptual Image Quality,” *Test*, vol. 1, pp. 305–312, 2011.

[57] T. Brandão and M. P. Queluz, *No-Reference Quality Assessment of H.264/AVC Encoded Video*, vol. 20, no. 11. 2010, pp. 1437–1447.

[58] S. Gabarda and G. Cristóbal, “Blind image quality assessment through anisotropy.,” *Journal of the Optical Society of America A*, vol. 24, no. 12, pp. B42–B51, 2007.

[59] S. Gabarda and G. Cristóbal, “Image quality assessment through a logarithmic anisotropic measure,” *SPIE Photonics Europe*, vol. 7000, no. 34, p. 70000J–70000J–11, Apr. 2008.

[60] A. Mittal, R. Soundararajan, and A. C. Bovik, “Making a ‘Completely Blind’ Image Quality Analyzer,” *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 209–212, Mar. 2013.

[61] E. Wigner, “On the Quantum Correction For Thermodynamic Equilibrium,” *Physical Review*, vol. 40, no. 5, pp. 749–759, 1932.

[62] A. Mittal, A. K. Moorthy, and A. C. Bovik, “Referenceless Image Spatial Quality Evaluation Engine,” in *45th Asilomar Conference on Signals, Systems and Computers*, 2011.

[63] D. Ruderman, “The statistics of natural images,” *Network Computation in Neural Systems*, vol. 5, no. 4, pp. 517–548, 1994.

[64] T. Installations and L. Line, “Subjective video quality assessment methods for multimedia applications,” *Networks*, vol. 910, no. P.910, p. 42, 1999.

[65] J. Lino, “Extractions of the Plenoptic 2.0 Holoscopic Camera Resources with the DAPBe Method,” 2013. [Online]. Available: http://bordalo.img.lx.it.pt/apendix6.rar. [Accessed: 15-Oct-2013].

[66] J. L. Rodgers and W. A. Nicewander, “Thirteen Ways to Look at the Correlation Coefficient,” *American Statistician*, vol. 42, no. 1, pp. 59–66, 1988.

[67] C. Spearman, “The proof and measurement of association between two things,” *The American Journal of Psychology*, vol. 15, no. 1, pp. 72–101, 1904.