



Universitat de Girona

ONE-SHOT PATTERN PROJECTION FOR DENSE AND ACCURATE 3D RECONSTRUCTION IN STRUCTURED LIGHT

Sergio FERNÁNDEZ NAVARRO

Dipòsit legal: GI. 1232-2012

<http://hdl.handle.net/10803/83621>



One-shot pattern projection for dense and accurate 3D reconstruction in structured light està subjecte a una llicència de [Reconeixement 3.0 No adaptada de Creative Commons](https://creativecommons.org/licenses/by/3.0/)

One-shot pattern projection for dense and accurate 3D acquisition in structured light



Sergio Fernandez
Supervisor: Prof. Joaquim Salvi

VICOROB

University of Girona

A thesis submitted for the degree of
Philosophiæ Doctor (PhD) in Technology

2012

Abstract

This thesis focuses on the problem of 3D acquisition using coded structured light (CSL). CSL aims to retrieve the 3D information of an object shape by means of a camera or set of cameras and a coded active device that projects a pattern onto the scene. This pattern imposes the illusion of texture onto an object, increasing the number of correspondences even in presence of textureless surfaces. Afterwards, the camera(s) images the scene, and 3D acquisition is pursued using the same triangulation principle used in stereovision. The vast amount of literature on 3D acquisition using structured light gives an idea of the relevance of the topic. Nowadays, an active research is being done in CSL techniques that are able to work in moving scenarios. This implies the use of few or just one projected pattern, from which 3D information must be extracted. In this thesis, a review of the main CSL approaches is presented. The main features of the proposed CSL algorithms present in the literature are studied. We propose a first approach for one-shot dense acquisition using Wavelet Transform (WT) analysis and color multiplexing of different fringe patterns. This algorithm performs well for smooth surfaces, but fails under presence of discontinuities as the slopes are not detected optimally by the WT algorithm. Therefore, a deep study of the two most used frequency-based techniques is performed, and a new proposal for automatic selection of the window width using Windowed Fourier Transform (WFT) is made. Using this analysis, we implemented a new technique for one-shot dense acquisition. The technique is based on adaptive WFT and DeBruijn coding. The experimental results show that the proposed method obtains accuracy levels comparable to DeBruijn algorithm, but providing absolute dense acquisition. Finally, the last part of the thesis focuses on the problem of registration, as many applications need to register more than one single scan into a big 3D acquisition of a large shape. With this we finish the work of this thesis. The thesis concludes with an analysis of the pros and cons of the technique. The proposed algorithm sets a new trend in CSL as it merges the density of WFT frequency coding with the accuracy of DeBruijn spatial coding, which had been separated approaches until now.

Keywords: Pattern Projection, Structured light, Windowed Fourier Transform, 3D Measuring Devices, Active Stereo, Computer Vision

Resumen

La presente tesis doctoral estudia el problema de la reconstrucción 3D utilizando luz estructurada codificada (CSL). El objetivo de CSL es extraer la información 3D de la superficie de un objeto por medio de una cámara (o cámaras) y de un dispositivo emisor de luz (usualmente un proyector digital) que proyecta un patrón sobre la superficie. De este modo, se imprime textura artificial sobre el objeto, incrementando el número de correspondencias incluso para superficies carentes de textura. Tras capturar la imagen se realiza la reconstrucción 3D, utilizando los mismos principios que en stereovisión pasiva. Actualmente se está trabajando intensamente en las técnicas de CSL aplicables a entornos dinámicos. Ésto implica el uso de muy pocos o incluso un único patrón proyectado. Esta tesis realiza primero una revisión de las principales técnicas de CSL presentes en la literatura, indicando y comparando las características comunes de los diferentes algoritmos de CSL. Posteriormente se propone un nuevo y único patrón de proyección CSL. Dicho patrón utiliza el análisis mediante la Transformada Wavelet (WT), así como la multiplexión en color para unir diferentes componentes sinusoidales en una única proyección. El algoritmo obtiene buenos resultados para superficies suaves. Sin embargo, se detectan errores en la reconstrucción de discontinuidades, debido principalmente a tipo de análisis frecuencial utilizado. Esto nos lleva a realizar un estudio detallado de los diferentes algoritmos de análisis presentes en el dominio frecuencial. Resultado de este trabajo es la propuesta de un nuevo algoritmo de Windowed Fourier Transform (WFT) donde la selección del ancho de la ventana de análisis se calcula óptimamente de manera automática. Utilizando este algoritmo para la fase de análisis, se desarrolla un nuevo y único patrón basado en sinusoidales coloreadas siguiendo una secuencia DeBruijn. Los resultados experimentales muestran unos niveles de precisión comparables con otras técnicas DeBruijn, con la ventaja de que se obtiene una reconstrucción densa usando un único patrón. La tesis concluye realizando un análisis de las ventajas y desventajas de patrón diseñado. Dicho patrón establece un nuevo hilo en CSL al unir dos campos separados hasta hoy, como son la densidad de reconstrucción para entornos móviles (propio de las técnicas frecuenciales), y la precisión que proporcionan las técnicas basadas en secuencias DeBruijn.

Acknowledgements

This thesis was made possible thanks to the contribution of many persons. I would like to thank Munther (my supervisor in Liverpool) and Andrea and Umberto (my supervisors in Verona), for the guidance and support they provided me during my stages. Also, I want to thank the reviewers who helped me to improve this thesis with precise and meaningful suggestions. Pero por supuesto quiero agradecer a Quim, mi supervisor, por ser un soporte no solo en el plano profesional sino tambien en el personal. Apretando más cuando era posible y siendo comprensivo cuando lo necesitaba, ha conseguido que esta tesis saliera adelante, sin prisa pero sin pausa. En el plano personal, quiero agradecer a mis amigos del laboratorio de Visió y en general a VICOROB, por haber aceptado a este 'sevillanu' como compañero de trabajo (y de fiestas, por que no decirlo). Son muchos los ratos que hemos pasado codo con codo, y me quedan grandes amigos de esta etapa. Solo por eso ya vale la pena hacer una tesis. Non voglio dimenticare nemmeno a gli amici di Verona, loro sono i responsabili di farmi sentire un italiano in piu durante questi quattro mesi che sono stato la. Y por ultimo (last but not least) mi familia, que no los enumero aqui pero que son, con distancia y sin ella, mi mejor ayuda. A ellos dedico esta tesis.

Contents

List of Figures	ix
List of Tables	xv
1 Introduction	1
1.1 The importance of 3D acquisition in Computer Vision	1
1.2 Motivation and objectives	3
1.3 Context	6
1.4 Structure of the thesis	8
2 State of the art on Coded Structured Light	11
2.1 Overview of structured light techniques	11
2.2 Classification of coding strategies for SL	13
2.3 Sparse reconstruction methods	16
2.3.1 Spatial multiplexing	16
2.3.1.1 DeBruijn based techniques	16
2.3.1.2 Non-formal coding	18
2.3.1.3 M-arrays	19
2.3.2 Time multiplexing	20
2.3.2.1 Temporal binary codes	21
2.3.2.2 Temporal n-ary codes	21
2.3.2.3 Temporal hybrid codes	22
2.4 Dense reconstruction methods	23
2.4.1 Time multiplexing	23
2.4.1.1 Discrete shifting methods	23
2.4.1.2 Continuous phase shifting methods	23

CONTENTS

2.4.2	Frequency multiplexing	26
2.4.2.1	Fourier Transform	26
2.4.2.2	Window Fourier Transform	30
2.4.2.3	Wavelet Transform	32
2.4.2.4	The problem of phase unwrapping	33
2.4.2.5	Alternatives to sinusoidal grating	34
2.4.3	Spatial multiplexing (grading methods)	35
2.5	Experimental results	35
2.5.1	Qualitative results	37
2.5.2	Quantitative results	38
2.6	Conclusions	42
3	First approach to one-shot dense reconstruction	45
3.1	Introduction	45
3.2	System proposal	47
3.2.1	Pattern coding	49
3.2.2	Pattern decoding	50
3.2.3	Extraction of the wrapped phase	50
3.2.4	The cost function	52
3.2.5	Combination of the wrapped phased	53
3.2.6	3D or depth extraction	55
3.3	Implementation and results	56
3.3.1	Simulated results	56
3.3.2	Experimental results	59
3.4	Conclusions	60
4	Automatic window selection in Frequency Transform techniques	65
4.1	Introduction	65
4.2	Comparative study between WT and WFT	66
4.2.1	Adaption of the wavelet signals to the use in WFT	69
4.3	A new proposal for the automatic selection of the window size	70
4.3.1	Preprocessing the image	71
4.3.2	Setting the average period and the standard deviation	71
4.3.3	Setting the range of frequencies and the window	73

4.3.4	Computing WFT	73
4.3.5	Phase unwrapping	73
4.4	Testing the algorithm	74
4.4.1	Simulated results	74
4.4.2	Experimental results	76
4.5	Conclusions	79
5	A proposal on one-shot absolute pattern for dense reconstruction	83
5.1	Introduction	83
5.2	A new proposal for one-shot dense reconstruction	84
5.2.1	Pattern creation	86
5.2.2	Geometric and color calibration	86
5.2.3	Pattern projection and recovery	87
5.2.4	DeBruijn analysis	88
5.2.5	Dynamic programming	89
5.2.6	Windowed Fourier Transform analysis	89
5.2.7	Combination of DeBruijn and wrapped phase patterns	90
5.2.8	Triangulation	91
5.2.9	Filtering	91
5.2.10	Meshing	93
5.3	Results	93
5.3.1	Simulation results	93
5.3.2	Empirical results	94
5.3.3	Reconstructing color surfaces	95
5.4	Conclusions	96
6	Registration of single-views from Structured Light	111
6.1	Introduction	111
6.2	Brief overview of global registration algorithms	112
6.3	A novel pipeline for global registration	114
6.3.1	Keypoint detection	116
6.3.2	Keypoint description	116
6.3.3	Matching	118
6.3.3.1	Selection of the keypoint candidates for matching	118

CONTENTS

6.3.3.2	Geometrical constraints	119
6.3.3.3	Compute the rigid motion	119
6.3.3.4	Calculate the residual	120
6.4	Results	121
6.4.1	Results using a closed-form synthetic object	121
6.4.2	Results of registration of SL partial views	124
6.5	Conclusions	124
7	Conclusions	127
7.1	Discussion	127
7.2	Contributions	131
7.3	Publications	132
7.3.1	Journals	132
7.3.2	Conferences	132
7.3.3	Book chapters	133
7.4	Future work	134
A	Geometric camera-projector calibration	135
A.1	Classical calibration systems	135
A.2	A novel proposal for projector calibration	136
A.2.1	Camera calibration	138
A.2.2	Projection of a checkerboard onto a planar surface	138
A.2.3	Extract 3D coordinates of the marked points	139
A.2.4	Compute homography	139
A.2.5	Compute 3D points of the projected pattern	140
A.2.6	Calibrate the projector	140
A.3	Testing the proposed calibration	141
B	Color calibration and RGB channel alignment	147
B.1	Color calibration	147
B.2	RGB channel alignment	149
C	Application of the Remainder Theorem to the use in Multiple Phase Shifting pattern projection	151

D Dynamic programming	153
D.1 Classical dynamic programming	153
D.2 Multi-pass dynamic programming	153
References	157

CONTENTS

List of Figures

1.1	Comparison between human stereo-vision and a classical passive stereo-vision device.	2
1.2	Example of some applications of coded structured light.	5
2.1	General idea of a coded structured light system [1].	13
2.2	Pattern proposed by Pages et al. [2] (RGB pattern and luminance channel).	18
2.3	Code generation direction followed by Morano et al. [3].	20
2.4	Spatial temporal algorithm proposed by Ishii et al. [4].	22
2.5	Composite pattern formed by the multiplexation of modulated phase shifting profilometry (PMP) patterns using the algorithm of Guan et al. [5].	26
2.6	Pair of projected sinusoidal patterns, having two different frequencies (k is the number of periods).	27
2.7	Results of Monks et al. [6], Posdamer et al. [7] and Guhring [8], respectively.	40
2.8	Results of Pribanic et al. [9], Su et al. [10], and Carrihill and Hummel [11], respectively.	41
3.1	Surface slope producing a decoding error in traditional fringe pattern methods	46
3.2	Diagram of the proposed SL algorithm using WT and the remainder theorem.	48
3.3	Combination of Red, Green and Blue channels to create the color multiplexed pattern.	49
3.4	Pattern decoding (image processing).	51

LIST OF FIGURES

3.5	On the top, simulated noised fringe pattern. On the left column, example of the WT and ridge (in blue) for a specific position in the input image (top), and the corresponding extracted phase (bottom). On the right column, similar results using the cost function.	54
3.6	Simulated surface	57
3.7	Imaged patterns of Fig. 3.6 with added noise values of 5%, 10% and 20% of the maximum dynamic range, respectively.	57
3.8	Reconstructed surface of Fig. 3.6 with a noise of 5%, and corresponding error map re-scaled from [0%, 3.48%] to [0, 255].	58
3.9	Projected pattern on the object of Fig.3.1 and reconstructed surface. . .	59
3.10	Original and reconstructed sheets of paper.	60
3.11	Sheets of paper of Fig. 3.10: captured image (top-left), enhanced image (top-right), one color channel (bottom-left) and its wrapped phase (bottom-right).	61
3.12	Original and reconstructed surface of a ceramic face (the 3D shape is inverted to appreciate the details of nose and eyes).	61
3.13	Ceramic face of Fig. 3.12: captured image (top-left), enhanced image (top-right), one color channel (bottom-left) and its wrapped phase (bottom-right).	62
4.1	Diagram of Morlet (<i>a</i>), Paul (<i>b</i>) Shannon (<i>c</i>) and Spline (<i>d</i>) wavelets, traditionally used in fringe pattern analysis.	69
4.2	Diagram of the proposed algorithm showing the required steps.	72
4.3	Flow chart of the unwrapping algorithm proposed by Herraes et al. [12].	75
4.4	Peak function used as the input image, containing phase variation at different speeds, and added noise	76
4.5	Simulation results of the wrapped phase (from $-\pi$ to π): on the left column the four adapted mother wavelets (Morlet, Paul, Shannon, Spline). On the right column, their corresponding results in WT.	77
4.6	Recovered unwrapped phase, in radians, for the simulated input fringe pattern of Fig. 4.4. Morlet WFT has been used.	78
4.7	Input image and reconstructed unwrapped phase map in radians (using Paul wavelet for WFT).	79

4.8	Plastic bended sheet: on the left column, experimental results for the four adapted mother wavelets (Morlet, Paul, Shannon, Spline). On the right column, their corresponding results in WT.	80
4.9	Input image and reconstructed unwrapped phase map in radians (using Paul wavelet for WFT).	81
4.10	Radiotherapy RANDO phantom: on the left column, experimental results for the wrapped phase (from $-\pi$ to π) of the four adapted mother wavelets (Morlet,Paul,Shannon,Spline). On the right column, their corresponding results in WT.	82
5.1	Diagram of the proposed algorithm	85
5.2	Proposed pattern: HSV representation of the DeBruijn sequence (top-left), fringe pattern (top-right) and the resulting RGB pattern for $m = 64$ (bottom).	87
5.3	Behaviour of the 1st and 2nd derivative on a synthetic sinusoidal signal.	88
5.4	Visual representation of a Morlet signal with $n = 3$ periods.	90
5.5	On the left, detail of the wrapped phase and a crest maxims (in red), and its corresponding slits line position (in green). On the right, the wrapped section before and after correction, and the correction interpolation error.	91
5.6	Ray plane intersection diagram	92
5.7	Proposed setup: a portable frame with the camera and the pico-projector attached on it.	98
5.8	Peaks signal and recovered pattern for the proposed algorithm and noise of $std = 0.1$	99
5.9	Normalized error on reconstructed depth positions, for different values of noise.	99
5.10	3D retrieval of a bended piece of paper sheet. On the top, original and color rectified image. On the middle, RGB color channels before DeBruijn and WFT analysis. On the bottom, extracted DeBruijn color slits and WFT wrapped phase.	100
5.11	3D cloud of points corresponding to the bended piece of paper sheet. 802768 points were reconstructed.	101

LIST OF FIGURES

5.12	3D retrieval of a manufactured piece of white cork. On the top, original and color rectified image. On the middle, RGB color channels before DeBruijn and WFT analysis. On the bottom, extracted DeBruijn color slits and WFT wrapped phase.	102
5.13	3D cloud of points corresponding to the manufactured white cork. 620496 points were reconstructed.	103
5.14	3D retrieval of a ceramic figure of 'hello kitty'. On the top, original and color rectified image. On the middle, RGB color channels before DeBruijn and WFT analysis. On the bottom, extracted DeBruijn color slits and WFT wrapped phase.	104
5.15	3D cloud of points corresponding to the figure of 'hello kitty'. 666900 points were reconstructed.	105
5.16	3D retrieval of a ceramic sculpture of a horse. On the top, original and color rectified image. On the middle, RGB color channels before DeBruijn and WFT analysis. On the bottom, extracted DeBruijn color slits and WFT wrapped phase.	106
5.17	3D cloud of points corresponding to the ceramic sculpture of a horse. 722127 points were reconstructed.	107
5.18	3D retrieval of a Macbeth checker plane. On the top, original and color rectified image. On the middle, RGB and HSV channels before DeBruijn and WFT analysis. On the bottom, extracted DeBruijn color slits and WFT wrapped phase.	108
5.19	3D cloud of points corresponding to the Macbeth checker plane.	109
6.1	Diagram of the proposed registration algorithm.	115
6.2	Creation of the 2D histograms from the local planes, $b_s \times b_o$ and bins (from [13]).	118
6.3	2D representation of triplets in view 1 and view 2. As can be seen, the relative distances within the keypoints of both images are not related for the triplet on the top of second image, but will allow a match for the triplet on the bottom.	119

6.4	Results of the proposed detection + description. Top: input image view. Middle: similarity and binarized similarity matrices. Bottom: potential keypoint candidates for matching.	122
6.5	Registration results for the object <i>bunny</i>	123
6.6	Registration results for the two partial views of a ceramic sculpture of a horse. On the top, partial views and selected keypoints (red points). On the bottom, results of registration. The green circle marks the first view, while the red circle marks the second view.	125
A.1	Flow chart of the different steps in the calibration process.	137
A.2	Checkerboard images for camera calibration.	138
A.3	Planar surface (left image), and two different positions of the projected checkerboard (middle and right images).	139
A.4	Rays coming from the camera and going to the grid corners of the projected pattern.	140
A.5	Corners detection on the projected grid pattern.	141
A.6	Projector-camera setup used to test the proposed algorithm.	142
A.7	Projector-camera calibration results for the given setup (extrinsic parameters).	144
A.8	Reprojection error of the projector calibration algorithm.	145
B.1	Received color intensities for projected increasing values of Red, Green and Blue, respectively.	148
B.2	Projector-camera model for color calibration proposed by Caspi et al. [14].	148
D.1	On the left, example of the violation of the monotonicity assumption. On the right, the resulting match grid.	155

LIST OF FIGURES

List of Tables

2.1	Proposed classification embracing every group of CSL.	15
2.2	Selected methods, with their main attributes.	36
2.3	Quantitative results. The headings are: author's name of the technique; average deviation of the reconstructing error; standard deviation of the reconstructing error; number of 3D points reconstructed; number of projected patterns.	38
3.1	Error rates for the given input noise going from noises of 5% to 20% of the data dynamic range.	58
3.2	Quantitative results reconstructing a flat plane. The headings are: author's name of the technique; average deviation of the reconstructing error; standard deviation of the reconstructing error; number of 3D points reconstructed.	59
4.1	Relative mean error of the recovered phase	76
5.1	Quantitative results. The headings are: author's name of the technique; average deviation of the reconstructing error; standard deviation of the reconstructing error; number of 3D points reconstructed; number of projected patterns.	94
A.1	Calibration results for the intrinsic parameters.	143
A.2	Calibration results for the extrinsic parameters.	143

LIST OF ACRONYMS

List of Acronyms

In alphabetical order.

- 2D: Two dimensional range data.
- 3D: Three dimensional range data.
- AC: Alternating frequencies (high frequencies in Computer Vision analysis).
- AQUAVISION: 3D computer vision algorithms for underwater cartography and aquaculture.
- CCD: Charge-coupled device.
- CIS: Correlation Image Sensor.
- CPU: Central Processing Unit.
- CSL: Coded Structured Light.
- CV: Computer Vision.
- CWT: Continuous Wavelet Transform.
- DC: Direct Current (low frequencies in Computer Vision analysis).
- DLP: Digital Light Projector.
- DP: Dynamic Programming.
- FT: Fourier Transform.
- GFT: Gabor Fourier Transform.
- HOG: Histogram of Gaussians.
- HSI (HSV): Hue, Intensity and Illuminance (Value) channels.
- ICP: Iterative Closest Points.
- LCD: Liquid Crystal Display.

LIST OF ACRONYMS

- LS: Least Squares.
- MWFT: Multiscale Windowed Fourier Transform.
- MUMAP: multi-modal maps.
- PANDORA: Persistent Autonomy through Learning, Adaptation, Observation Re-planning.
- PCA: Principal Components Analysis.
- PMP: Phase Measurement Profilometry.
- PS: Phase Shifting.
- RAM: Random Access Memory.
- RGB: Red, Green and Blue channels.
- ROI: Region Of Interest.
- SL: Structured Light.
- SPD: Spatial Phase Detection.
- TRIDENT: Marine Robots and Dextrous Manipulation for Enagling Autonomous Multipurpose Intervention Missions.
- UDG: University of Girona.
- VICOROB: Computer Vision and Robotics Group.
- WFF: Windowed Fourier Filtering.
- WFR: Windowed Fourier Ridge.
- WFT: Window Fourier Transform.
- WT: Wavelet Transform.

1

Introduction

This chapter provides an introduction to the problem of three dimensional acquisition in Computer Vision. The importance of 3D acquisition and principal approaches are outlined in section 1.1, while section 1.2 presents the motivations behind this work. The context in which this work has been carried out is presented in section 1.3. Finally, section 1.4 outlines the structure of this thesis.

1.1 The importance of 3D acquisition in Computer Vision

This thesis is focused on the problem of 3D acquisition (i.e. 3D reconstruction) using Structured Light (SL). 3D reconstruction constitutes a valuable ability for any machine that needs to understand the environment where is being involved or where any action is required from its part. There are different ways to perform a 3D reconstruction of the environment. Doing an analogy with the human body (the most developed and complex machine) two different senses are associated to 3D reconstruction: the sense of touch and the sense of vision. Using the sense of touch, 3D shape of an object can be recovered by 'touching'; i.e., by passing a haptic sensor through the surface to scan. Analogously, a group of techniques use the same principle in 3D scene reconstruction. They are the so called *contact techniques*. Nowadays, one can obtain accurate reconstruction using any available contact technique. However, there are many applications where the direct contact with the object to scan is not feasible or recommended. Vision is the other sense used by humans for reconstruction. Vision is the most important and developed human sense, which allows us to perceive color, texture, shape and depth. Analogously, computer vision is a subfield of artificial intelligence that investigates how to make computer algorithms which are able to perceive and understand the world

1. INTRODUCTION

through images. Roughly, computer vision tries to emulate the visual perception of the human being from the first stage of light detection till the complex task of understanding what is being perceived. In this case, light is detected by cameras, while the high level tasks of image understanding are processed by using computer algorithms. Following with the analogy with humans, the ability of perceiving depth is based on the binocular stereopsis formed by the eyes (see Fig. 1.1). The slightly different position of the eyes on the head provokes that an object appears in different horizontal positions in each image provided by each eye. This difference on relative positions, known as disparity, gives a cue about the object depth. Computer vision tries to copy human stereopsis by using two cameras as if they were two eyes in what is called passive stereovision [15]. An alternative consists in using a single camera and moving it to different known positions for perceiving the scene from multiple points of view. This approach is known as structure from motion [16], [17]. Furthermore, disparity variations on a sequence of stereo images can be used for rigid motion estimation [18], [19].



Figure 1.1: Comparison between human stereo-vision and a classical passive stereo-vision device.

Stereovision is one of the most important topics in computer vision since it allows the three dimensional position of an object point to be obtained from its projective points on the image planes [20]. The setup is usually formed by electromagnetic sensor devices, usually working in the visual range (normal cameras). The ambient light or the projected signal gets reflected by the scene, typically the surface of an object. This reflection is imaged by the camera or set of cameras. Afterwards, a complete pipeline for filtering, detection and matching of points is pursued to provide a 3D model of the object shape. The most difficult problem in stereovision is the determination of homologous points in two images, i.e. determining which pair of projective points represent the same three dimensional object point. This problem is known as the correspondence problem, which is the main limitation of stereovision since once it is solved the rest has been already formalized [20]. Even if a set of geometrical constraints, known as the epipolar geometry [21], is able to simplify the correspondence problem, it is not a defini-

tive solution, mainly because density is directly related to the texture of the object. For example, the correspondence problem cannot be solved when observing non-textured objects, when points only appear in one of the images due to a surface occlusion, when points are multiply matched between the images or under adverse lighting conditions. Any of these situations complicates the process of finding correspondences in presence of textureless surfaces [22]. Therefore, stereovision is rather limited to reconstruct dense 3D surfaces, due to the problem of finding correspondences [23]. Methods based on structured light came to cope with this issue, creating correspondences and giving specific codewords to every unitary position on the image. In this approach one of the cameras is substituted by an active device (nowadays a Digital Light Projector), which projects a structured light pattern onto the scene. The projected pattern imposes the illusion of texture onto an object, increasing the number of correspondences [24]. Therefore, surface reconstruction is possible when looking for differences between projected and recorded patterns. The former structured light techniques were based on projecting simple primitives like a single dot or a single line of light, usually provided by lasers. The advantage of projecting such structured light primitives is that the correspondence problem of the illuminated points on the images is directly solved. Nevertheless, the number of correspondences per image is very small. In order to increase the number of correspondences, structured light patterns like arrays of dots, stripes, grids or concentric circles were introduced. However, with this solution the identification of different pattern regions on the images becomes ambiguous so that the correspondence problem is not directly solved. This fact provoked the emergence of coded structured light [23]. In this case, the projected patterns are coded so that each element of the pattern can be unambiguously identified on the images. This thesis contributes in the field of coded structured light for the dense and accurate retrieval of moving surfaces.

1.2 Motivation and objectives

There are many applications where stereo-vision plays an important role, not only in the fields strictly related to computer vision but also in many other industrial environments. Some of them are (see Fig. 1.2):

- Range sensing: as a way to measure distances in an environment where a specific machine needs to move through.
- Industrial inspection of manufactured parts: this constitutes a key factor in quality check of any industrial component. A 3D reconstruction of the piece reveals

1. INTRODUCTION

necessary for many products where the accuracy of the 3D shape represents a crucial factor.

- Reverse engineering: digitization of complex, free-form surfaces in order to reproduce exactly a previous created object.
- Object recognition: as a first step for recognition, it is necessary to have a reliable 3D model of the object. In this case the accuracy of the reconstruction can determine the success on a posterior recognition step.
- 3D map building: with the new trends in augmented reality, many applications require a 3D map of a set scenario. This is the case of 3D reconstruction of large surfaces (i.e. applications in terrestrial or submarine mapping). This is possible only if a previous 3D retrieval has been pursued.
- Biometrics: the specific shape and volume of a person can be an interesting for a variety of applications where a 3D database would facilitate a posterior recognition of a person among others.
- 3D surgery: in the last years, 3D surgery has revealed as a perfect aid tool for doctors. Having a 3D reconstruction of some internal parts of the object, while a operation is being pursued, facilitates the recognition of the regions where a work must be done. Roughly speaking, it is like looking at the body from inside.
- Clothing design: as a contribution to fashion, there is also a lot of interest in the design of clothes adapted, or better fitted, to the human body.

As mentioned before, in passive stereo-vision systems the density of the 3D reconstruction is directly related to the texture of the object. This complicates the process of finding correspondences in presence of textureless surfaces [22], and reduces its range of applications only to rich textured surfaces where a good number of matches between the captured images can be done. The aim of coded structured light is to robustly obtain a large set of correspondences per image independently of the appearance of the object being illuminated and the ambient lighting conditions. The different SL techniques available in the literature differ in three main aspects. First, in the way in which every point in the pattern is identified, i.e. what kind of codeword is used and whether it encodes a single axis or the two axes of an image, and how many projected patterns are required. In reality, it is only necessary to encode a single axis, since a 3D point can be obtained by intersecting two lines (i.e. when both pattern axis are coded)

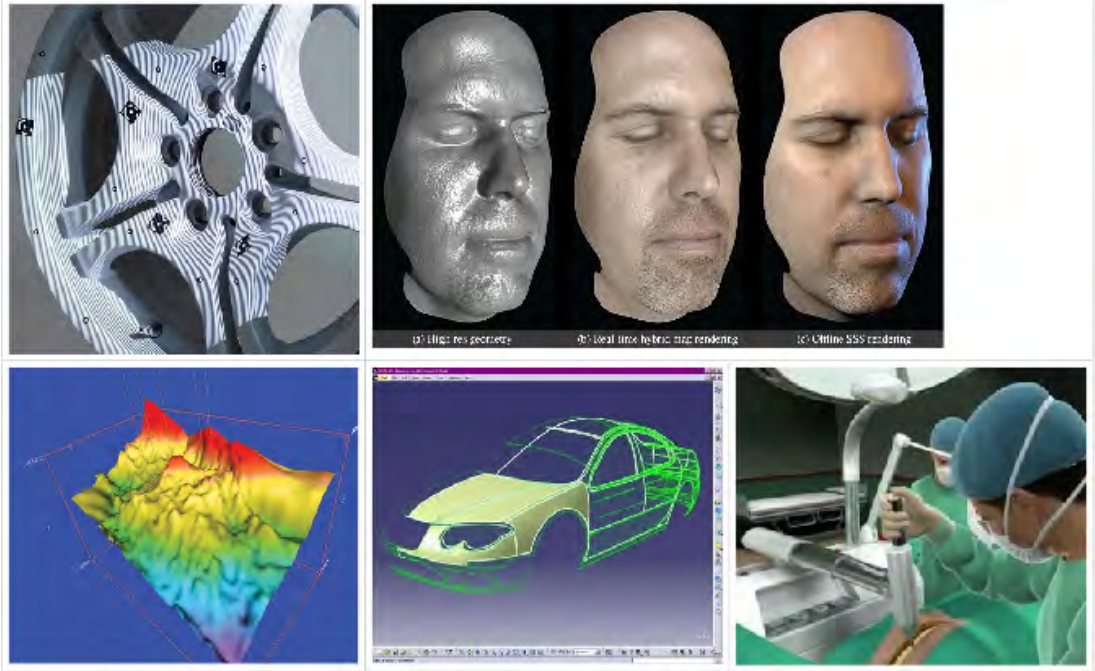


Figure 1.2: Example of some applications of coded structured light.

or intersecting one line (the one which contains a pixel of the camera image) with a plane (i.e. when a single pattern axis is coded). Second, the analysis algorithm and the domain in which the feature extraction and the matching step is pursued. Both spatial and frequency analysis can be carried out, separately or combined at some point of the processing step. Finally, the more important aspect to consider in a SL pattern is the output that it provides, and the sparsity or density of the obtained 3D reconstruction.

Regarding the number of projected patterns required to identify a point, SL approaches can be categorized depending on whether they are intended to work in static or in dynamic scenarios. The first SL approaches that can be found in the literature consist in the projection of a set of patterns that are successively projected onto the surface to measure. The codeword for a given pixel is usually formed by the sequence of illumination values for that pixel across the projected patterns. Therefore, the codification is called temporal because the bits of the codewords are multiplexed in time. This kind of patterns can achieve high accuracy in the measurements. However, as a main drawback, they are not able to work in moving scenarios. This represents a problem in many applications where the object to scan does not remain still for much time, or even is moving across the scanning scenario. The ability to measure moving surfaces (up to the acquisition time required by the camera) is only achieved by

1. INTRODUCTION

one-shot patterns. Plenty of proposals for one-shot projection has been done during decades. For instance, classical spatial multiplexing techniques like DeBruijn and M-arrays-based patterns perform one-shot 3D reconstruction with good accuracy results ([6], [25], [6], [26], [3], [27], [28]). However, they produce sparse (feature wise) reconstructions, due to their digital profile that imposes the same codeword for a set of points in the recovered image. Another group of techniques is grouped in the fringe profilometry approaches. These techniques make use of some frequency analysis to extract the phase deviation of a fringe recovered pattern, with respect to the one that was previously projected. With a unique projection, it is possible to extract the phase deviation and from that the depth of the object. However, inaccuracies can occur at surface discontinuities due to the non-absolute (periodic) coding intrinsic to the method [29]. There exist some techniques that obtain density and absolute coding by using one-shot spatial grading [11] [30], but both achieve a rather low accuracy [29]. Therefore, the problem that must be addressed is to design a SL pattern able to:

- Image the scene using only one-shot projection: this enables its use for moving scenarios, and could be potentially used for real-time applications.
- Dense 3D reconstruction: this is considered an asset in many of the previously mentioned applications. The more 3D points are available in the reconstruction, the better the detection of crucial points (being this good or bad) can be achieved.
- Absolute coding and accuracy: the designed algorithm must be able to reconstruct uniquely any point in the imaged scenarios. That is, uncertainties in the reconstruction must not appear. Moreover, the accuracy must be at least similar to the best accuracy results obtained nowadays in active reconstruction.

These are the challenges overcome in the SL algorithm proposed in this thesis.

1.3 Context

This thesis was carried out in the VICOROB laboratory at the University of Girona. The research areas of the group are underwater robotics and vision, mobile robotics, 3D perception and medical imaging. The research activities are currently supported by several national projects and an European projects. The work developed in this thesis has been partially funded by the following Spanish projects:

- CICYT Project AQUAVISION Vision Systems for computer cartography and underwater aquaculture (Ref DPI2007-66796-C03-02), funded by the Spanish Ministry of Education and Science.
- FP7-ICT-2011-7: PANDORA Persistent Autonomy through Learning, Adaptation, Observation and Re-planning (Ref 288273) funded by the European Commission and the project
- CICYT Project RAIMON Autonomous Underwater Robot for Marine Fish Farms Inspection and Monitoring (Ref CTM2011-29691-C02-02), funded by the Spanish Ministry of Science and Innovation.

Within the 3D perception group, 3D reconstruction using coded structured light has produced many contributions to VICOROB. Always under the supervision of prof. Joaquim Salvi, several researchers and PhD students have done previous contributions. This is the case of the works of David Fofi (*Navigation d'un Vehicule Intelligent l'aide d'un Capteur de Vision en Lumiere Structure et Code*) Jordi Pages (*Assisted visual servoing by means of structured light*) and Radu Orghidan (Catadioptric Stereo based on Structured Light Projection). Nevertheless, 3D reconstruction is a generic step that can find applications in a variety of contexts. For example, among the interests of the VICOROB group it is possible to find also projects related to the construction of multi-modal maps (MuMAP) and to the development of autonomous underwater vehicles (robots) for multi-purpose intervention missions (TRIDENT and 7PMSTREP). Both tasks require a system able to create 3D images of the explored area in order to increase the knowledge of the environment where the robot is moving across.

Moreover, this thesis has been developed in two other places as part of research stages. One is the General Engineering Research Institute (GERI) of the John Moores University (Liverpool, U.K.). This group is specialized in fringe pattern projection and decoding, among other disciplines. Therefore, part of my research was done in this center in order to learn about the 3D reconstruction using this kind of coded structured light patterns. The second place I visited was the department of informatics of the University of Verona (Italy). They work in stereo-vision applications, and have a spin-off, 3Dflow, working on 3D reconstruction of large scenarios using stereo-vision. Therefore, my stage at this center was addressed to learn the existing techniques for registration of 3D range data, and merging single views into a complete 3D reconstruction.

1.4 Structure of the thesis

The material presented in this thesis is structured as follows.

- Chapter 2 presents the state of the art of structured light. First, a new classification of the different techniques is proposed. Afterwards, the study of the different groups regarding the proposed classification is pursued, focusing on the advancements presented in the last years. Afterwards, the results of implementing some of the most relevant techniques are showed, comparing their pros and cons. Finally, the advantages and disadvantages of the most relevant techniques are discussed, pointing out the new potential fields of research.
- Chapter 3 presents a first approach of one-shot dense reconstruction algorithm. First, the theory behind the proposed technique is explained, focusing on the frequency analysis and the color multiplexing use to create the pattern. Afterwards, details on the implementation and results are shown. The pros and cons are discussed, revealing some problems in the reconstruction of surfaces having discontinuities. An explanation of this problem is found in the way the frequency analysis is pursued. Therefore, a more deep study of the different solutions existing for frequency analysis is pointed out as the next research step.
- Chapter 4 performs a comparative study of the two frequency-based analysis most used in SL, the Wavelet Transform (WT) and the Windowed Fourier Transform (WFT). Afterwards, a new proposal for frequency analysis based on a modification of the traditional WFT algorithm is presented. Some test are done showing the better performance of this phase retrieval algorithm against the classical WT and WFT.
- Chapter 5 presents a one-shot algorithm for dense absolute reconstruction of 3D shape. The algorithm is based on the combination of DeBruijn and WFT techniques, using the procedure explained in chapter 4. Details of the implementation are explained, and both simulated and experimental results of reconstructions are presented. The performance in terms of accuracy is comparable to that obtained for DeBruijn patterns, while dense reconstruction is achieved. A discussion comparing the proposed method with the ones existing in the literature is pursued.
- Finally, chapter 6 faces the problem of registration of single views. The aim is to perform a 3D acquisition of coarse objects by registering more than one single views. This can be used for retrieval of big manufactured products, large

regions of the human body, or similar. Registration of several views provides 3D information of the object. This can be used as the final outcome of the method, or as a previous 3D localization for further reconstruction of more small and detailed regions.

- To complete the thesis, chapter 6 gives a summary of the presented work. Conclusions are drawn and future directions are discussed. The list of publications is presented at the end of this chapter.

In order to clarify some steps of the proposed SL approaches, a total of four appendices are presented at the end of the thesis. Appendix A explains the camera-projector calibration technique, required to know the intrinsic parameters of the camera and the projector, as well as the relative position between them. This is used to perform the triangulation between the projected and the recovered patterns. Afterwards, appendix B explains the color calibration and RGB channel alignment, required steps when projecting a color pattern. Appendix C explains the principles of the Remainder theorem used in the first approach of SL pattern. Finally, appendix D shows the principles of dynamic programming applied to DeBruijn pattern analysis.

1. INTRODUCTION

2

State of the art on Coded Structured Light

Projecting structured light patterns onto the scene in order to extract the 3D shape is a common solution used in computer vision, and it is considered one of the most reliable approaches among all 3D reconstruction techniques. Having a calibrated projector-camera pair, a light pattern is projected onto the scene and imaged by the camera. Correspondences between projected and recovered patterns are found and used to extract 3D surface information. The main advantage among other approaches is that the projected visual features are easily distinguished by the camera. This chapter presents an up-to-date review and a new classification of the existing techniques. Some of these techniques have been implemented and compared, obtaining both qualitative and quantitative results. The advantages and drawbacks of the different techniques and their potentials are discussed.

2.1 Overview of structured light techniques

The term Structured Light (SL) is used to refer to a vision system taking profit of an active light source which projects a light pattern onto the environment. In computer vision, SL is used as an active stereovision system to obtain 3D reconstruction by triangulation, in the same way as is done for stereo-vision. In this approach an active device (typically a projector) is used to project a SL pattern onto the scene. The projected pattern imposes the illusion of texture onto an object, increasing the number of correspondences [24]. This solves the main problem experimented in passive stereovision, which is a sparse reconstruction since density is directly related to the texture

2. STATE OF THE ART ON CODED STRUCTURED LIGHT

of the object, thus complicating the process of finding correspondences in presence of textureless surfaces [22], [23]. The first shape acquisition systems based on structured light were laser scanners [31]. These devices are typically based on scanning the object with a laser plane and detecting the projected line in the camera image for triangulating all the illuminated points. The advantage of these scanners is the large resolution and accuracy obtained leading to high quality 3D surface reconstruction. The main drawback is that they are limited to static objects and that a large number of images must be acquired. Furthermore, in order to scan the object either the laser plane must be rotated, or both the laser and the camera or the object must be moved at each iteration. In the latter case, the displacement must be known so that free-moving objects cannot be reconstructed. All these problems appear because in each acquired image only few points can be triangulated, i.e the points belonging to the laser stripe. This limitation can be minimized by projecting more complex patterns like a laser grid [32]. However, a new problem arises: since the grid has a unique color the identification of every grid region on the image becomes ambiguous. Coded structured light is aimed to solve these limitations [1]. In Coded SL the active device is typically a Digital Light Projector (DLP), and is modeled as the inverse of a camera (see Fig. 2.1). Therefore the calibration step is a similar procedure to the one used in stereo vision [33]. Using this techniques, surface reconstruction is possible when looking for differences between projected and recorded images, which are called patterns since they present a globally structured appearance. The simplest pattern is a black image with an illuminated pixel. In this case, only one point can be reconstructed by triangulation by using the pixel coordinates of the illuminated point in the pattern and the corresponding coordinates in the camera image. Note that this case is equivalent to use a camera and a laser pointer. In general, all the patterns available with laser technology can be reproduced with a DLP device.

In this chapter, an exhaustive analysis of the different coding strategies used in active structured light is done, focusing on the advancements presented in the last years. A new classification regarding the strategy used to create the pattern is proposed, comparing some common characteristics between them. Feasibility and accuracy are analyzed, giving both qualitative and quantitative results for the implemented techniques. This chapter is structured as follows: section 2.2 presents a classification of the different techniques. Discrete pattern based codification is studied in section 2.3, while section 2.4 deals with the continuous ones. The results of implementing some of the most relevant techniques are showed in section 2.5, comparing their pros and cons. Finally, section 2.6 concludes with a discussion of the surveyed methods, pointing out

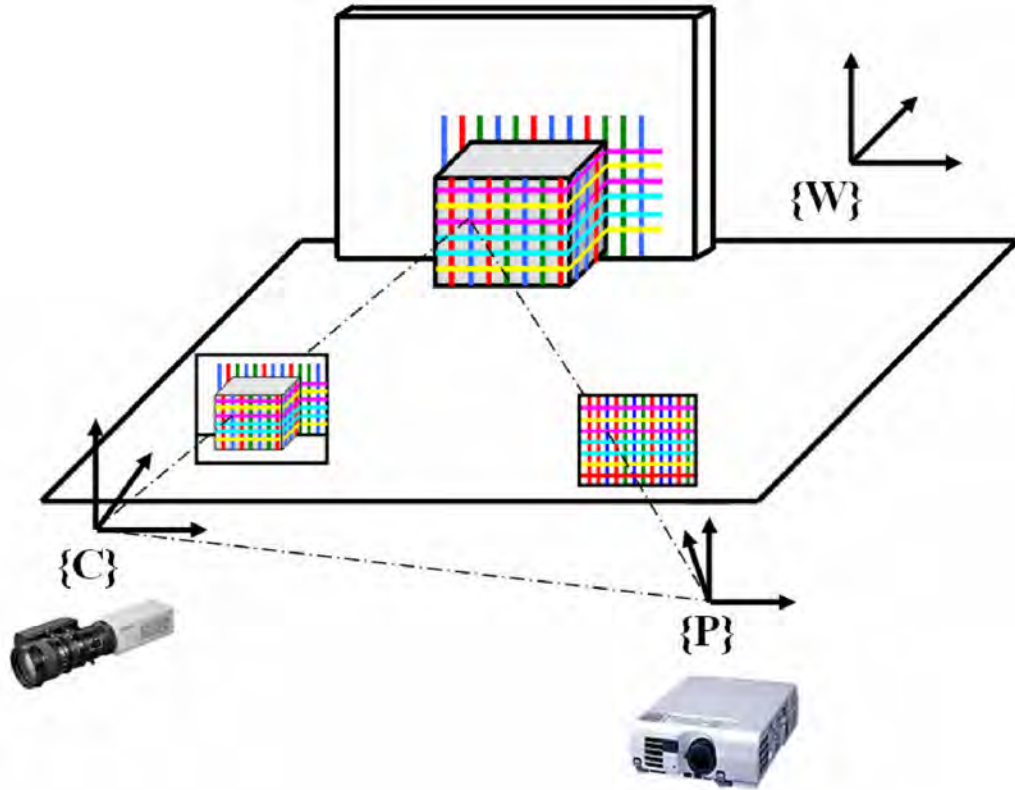


Figure 2.1: General idea of a coded structured light system [1].

advantages and disadvantages of the most relevant ones. In addition, general guidelines for choosing the most suitable technique, given the specifications of an application, are proposed.

2.2 Classification of coding strategies for SL

Coded structured light (CSL) is based on the projection of one pattern or a sequence of patterns that uniquely determine the codeword of a projecting pixel (or feature) within a non periodic region. CSL has produced many works during the last decades and some recopilatory works can be found in the literature. This is the case of the surveys presented by Batlle et al. [23] and Salvi et al. [1], that analyzed the different coded structured light techniques existing in temporal and spatial multiplexing domains from 1998 until 2004, respectively. Regarding frequency multiplexing, Su [34] reviewed the Fourier Transform (FT) techniques proposed until 2001. However, there is not any

2. STATE OF THE ART ON CODED STRUCTURED LIGHT

previous work comparing the three approaches together. Therefore, a classification extracting and analyzing attributes common in all the approaches is missing. This is overcome in the present survey, which also incorporates the most recent contributions done in CSL in the last years.

Table 2.1 shows a new classification of the existing pattern projection techniques. The main distinction has been done regarding the sparse or dense 3D reconstruction achieved. Patterns providing sparse reconstruction present a digital profile having the same value for the region represented by the same codeword. The size of this region largely determines the density of the reconstructed object. On the other hand, dense reconstruction is achieved by projecting either a sequence of digital patterns superposed over time to obtain full pixel coverage, or with a smooth profile pattern where every pixel has a unique codeword within the non-periodicity region. Both approaches achieve dense reconstruction. A posterior sub-classification is done regarding spatial, time and frequency multiplexing. Columns on the right indicate the value of some intrinsic attributes common to all the patterns. These attributes are:

- *Number of projected patterns*: determines whether the method is valid or not for measuring moving objects.
- *Number of cameras*: the method uses stereovision (2 or more cameras) coupled to a non calibrated pattern used only to get texture on the surface pattern, or a unique camera coupled to a calibrated projector.
- *Axis codification*: the pattern is coded along one or two axes.
- *Pixel depth*: refers to the color and luminance level of the projected pattern (B, G and C stands for Binary, Grayscale and Color respectively).
- *Coding strategy*: refers to the periodicity of the set of patterns projected on the surface (A stands for Absolute and P stands for Periodic).
- *Sub-pixel accuracy*: determines whether the features are found considering sub-pixel precision, thus providing better reconstruction results (Yes or No).
- *Color*: determines whether the technique can cope with colored objects (Yes or No).

2.2 Classification of coding strategies for SL

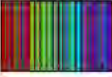
SPARSE	SPATIAL Multiplexing	De Bruijn	Boyer 1987	1	1	1	C	A	Y	N
			Savit 1998	1	1	1	C	A	Y	Y
DENSE	TIME Multiplexing	Non formal	Murata 1992	1	1	1	C	A	Y	N
			Page 2004	1	1	1	C	A	Y	N
DENSE	FREQUENCY Multiplexing	Single coding frequency	Foster 2007	1	1	1	C	A	Y	N
			Fuchsler 2008	1	1	1	C	A	Y	N
DENSE	SPATIAL multiplexing	Grading	Torino 2008	1	2	1	C	A	N	Y
			Miyasawa 1993	1	1	2	B	A	N	Y
DENSE	TIME Multiplexing	Shifting codes	Kanowski 2009	1	1	2	C	A	N	Y
			Li 1995	1	1	2	G	A	N	Y
DENSE	TIME Multiplexing	MultiplePhase Shifting(MPS)	Konacks 2006	1	1	2	C	P	Y	Y
			Griffin 1992	1	1	2	C	A	Y	Y
DENSE	TIME Multiplexing	Binary codes	Morano 1995	1	1	2	C	A	Y	Y
			Page 2006	1	1	2	C	A	Y	N
DENSE	TIME Multiplexing	N-ary codes	Adinar 2007	1	1	2	B	A	N	Y
			Podemot 1982	-2	1	1	B	A	N	Y
DENSE	TIME Multiplexing	Single Phase Shifting (SPS)	John 2007	-2	1	1	B	A	N	N
			Sun 2006	-2	2	1	B	A	Y	Y
DENSE	TIME Multiplexing	MultiplePhase Shifting(MPS)	Caspi 1988	-2	1	1	C	A	N	N
			Zhang 2002	-2	1	1	C	A	Y	N
DENSE	TIME Multiplexing	MultiplePhase Shifting(MPS)	Simon 2000	-2	1	1	G	A	Y	Y
			Galberg 2001	-2	1	1	G	A	Y	Y
DENSE	TIME Multiplexing	Single Phase Shifting (SPS)	Srinivasan 1985	-2	1	1	G	P	Y	Y
			Ueno 2004	-2	1	1	G	P	Y	Y
DENSE	TIME Multiplexing	MultiplePhase Shifting(MPS)	West 1991	1	1	1	C	P	Y	N
			Grav 2004	1	1	1	G	P	Y	Y
DENSE	TIME Multiplexing	MultiplePhase Shifting(MPS)	Grady 1991	-2	1	1	G	A	Y	Y
			Hyman 2000	-2	1	1	G	A	Y	Y
DENSE	FREQUENCY Multiplexing	Single coding frequency	Takeda 1984	1	1	1	G	P	Y	Y
			Erbilfi 2009	1	1	1	G	P	Y	Y
DENSE	FREQUENCY Multiplexing	Single coding frequency	Su 1990	2	1	1	G	P	Y	Y
			Ho 2000	2	2	1	G	P	Y	Y
DENSE	FREQUENCY Multiplexing	Single coding frequency	Chen 2007	1	1	1	C	P	Y	N
			Vae 2006	1	1	1	G	P	Y	Y
DENSE	FREQUENCY Multiplexing	Single coding frequency	Chen 2005	2	1	1	G	P	Y	Y
			Bercyman 2008	1	1	1	G	P	Y	Y
DENSE	FREQUENCY Multiplexing	Single coding frequency	Gdenaw 2006	1	1	1	G	P	Y	Y
			Zhang 2003	1	1	1	G	P	Y	Y
DENSE	FREQUENCY Multiplexing	Single coding frequency	Liu 1995	2	1	1	G	P	Y	Y
			Huang 2005	-2	1	1	G	P	Y	Y
DENSE	FREQUENCY Multiplexing	Single coding frequency	Ji 2007	1	1	1	G	P	Y	Y
			Wu 2006	1	1	1	G	P	Y	Y
DENSE	FREQUENCY Multiplexing	Single coding frequency	Pirmaschi 2003	1	1	1	C	A	Y	N
			Carroll 1985	1	1	1	G	A	Y	N
DENSE	FREQUENCY Multiplexing	Single coding frequency	Tajima 1990	1	1	1	C	A	Y	N
				Shots	Cameras	Axis	Pixel depth	Coiling strategy	Subpixel acc.	Color

Table 2.1: Proposed classification embracing every group of CSL.

2.3 Sparse reconstruction methods

In sparse reconstruction methods the pattern presents a digital profile; that is, a region of the pattern constituted by more than one pixel is represented by the same codeword. Two techniques, named spatial multiplexing and temporal multiplexing, can be employed to image the scene. Spatial multiplexing techniques code the pattern using the surrounding of a given feature, while temporal multiplexing creates the codeword by the successive projection of patterns onto the object. In addition, some methods combine spatial and temporal information to take advantage of both techniques.

2.3.1 Spatial multiplexing

Spatial multiplexing groups all techniques where the codeword of a specific position is extracted from surrounding points. Intensity or color variations are used to create the codeword. Three different coding strategies can be distinguished within this group: DeBruijn patterns, non-formal coding and M-arrays.

2.3.1.1 DeBruijn based techniques

DeBruijn sequences are a set of pseudo random values having specific properties between them. A k -ary DeBruijn sequence of order n is a circular sequence $d_0, d_1, \dots, d_{n^k-1}$ (length n^k) containing each substring of length k exactly once (window property of k). DeBruijn sequences can be constructed by taking a Hamiltonian or Eulerian path of a n -dimensional DeBruijn graph (see [35] for more details). This algorithm allows us to create univocal stripe sequences in the pattern, being able to extract the position by looking at the color of the stripes placed in the same window. Several proposals can be found using DeBruijn sequences, with both striped and multi-slit patterns. First proposals of DeBruijn-based striped patterns are found in the method developed by Boyer and Kak [36]. In this approach, RGB space was used to code the sequence of stripes. Being c_i^k the color of the stripe i of the sub-pattern k , the distance between two sub-patterns k and l is given by eq. (2.1):

$$d = \sum_{i=1}^N \delta_i \quad (2.1)$$

where

$$\delta_i = \begin{cases} 0 & \text{if } c_i^k = c_i^l \\ 1 & \text{otherwise} \end{cases} \quad (2.2)$$

The pattern proposed by Boyer and Kak [36] contains more than 300 stripes colored by three different colors. Color detection was done with a stripe indexing algorithm preceded by a Hamming filtering. However, no color calibration was pursued to suppress the effect of different albedo (that is, the diffuse reflectivity or reflecting power of the surface), leading to some errors due to leakage from blue to green channel.

A different approach was followed by Monks et al. [6], where a multi-slit-based DeBruijn sequence was projected. A total of 6 colors were used to color the slits, separated by black gaps. The slit colors were chosen so that every subsequence of three colors appeared only once. Colors were chosen in the Hue channel (HSI space), despite the projection was performed in RGB and transformed back to HSI once the image was captured by the camera. Full saturation and full intensity were chosen in the SI channels. A previous color calibration step was performed by the authors in order to determine the transfer function of the optical system. Once the system was calibrated, captured colors were corrected before applying fringe detection. A minimum cost matching algorithm was used in the decoding step in order to find the most probable matching between projected and recovered patterns, considering that some slits might be imaged partly occluded or bad segmented [1].

To simplify the peak detection process, Salvi et al. [24] created a grid of horizontal and vertical colored slits. Every crossings of the two slits were extracted by simple peak intensity detection. Hue channel was again used (in HSI space) to encode the colors. Three colors were assigned for the horizontal lines and other three for the vertical lines, using a DeBruijn sequence of order 3. The decoding step is done back in HSI space, showing negligible errors scanning planar surfaces under scene light control. However, some problems were encountered due to the sensitivity of the Hue channel under different albedo of the illuminated object. Some years later, Pages et al. [2] [25] proposed an alternative approach to traditional striped or multi-slit-based pattern. They combined a striped pattern in the Hue channel with a multi-slit pattern in the Intensity channel (see Fig. 2.2), which defined dark and bright areas within the same color stripe. Therefore, the high resolution of classical striped patterns and the accuracy of multi-slit patterns were combined. The half illuminated stripes were colored according to a DeBruijn sequence for a sub-pattern of n stripes, while bright slits were colored equally within the same sub-pattern. In the experiments, a 128 striped pattern having 4 colors and a window property of 3 encoded stripes was applied. Using this codification, their approach doubled the resolution of traditional DeBruijn stripe-based techniques.

2. STATE OF THE ART ON CODED STRUCTURED LIGHT

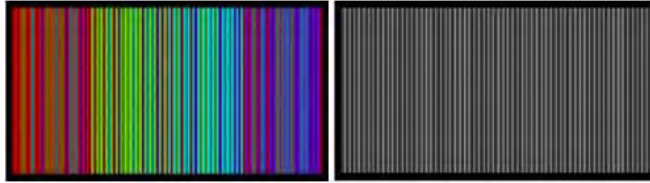


Figure 2.2: Pattern proposed by Pages et al. [2] (RGB pattern and luminance channel).

2.3.1.2 Non-formal coding

Non-formal coding is constituted by all the techniques having non-orthodox codification, in the sense that the pattern is designed to fulfill some particular requirements. Both one-axis and two-axes encoding are suitable for these methods. One-axis coding methods are based on striped or multi-slit patterns. This is the case of Forster [37] and Fechteler and Eisert [38] proposals, which created color-based patterns in which two adjacent colors must differ in at least two color channels in the receptor device (red, green and blue). This condition is not usually accomplished in DeBruijn sequences. Forster used a striped pattern, while Fechteler and Eisert employed a multi-slit pattern. In Fechteler and Eisert a parabola was fitted in every RGB channel (or combination of channels for non-pure RGB colors, option selected by Forster). Optionally, surface color was acquired by projecting an extra white pattern. Tehrani [39] applied the idea of color slits to reconstruct images taken from two camera views, using 10 hue values to create the slit pattern (the difference between colors was maximal for adjacent slits).

There are also some proposals based on two-axes encoding. For instance, Maruyama and Abe [40] proposed a pattern of randomly cut black slits on a white background. In this approach, coding information was held in the length of the slits and their position within the pattern. Every recorded segment had its own length, which can be similar for several segments. The codeword corresponding to a segment was determined by its own length and the lengths of its 6 adjacent segments. The main drawback of this method is that the length of segments is affected by the projector-object and object-camera distances, as well as by the camera optics, therefore reducing the reliability of the system. Another solution based on stripe lengths has been recently developed by Kawasaki et al. [22], who established a pattern of horizontal and vertical lines. In this work, the uniqueness of a specific location has been coded in the spacing between horizontal lines (in blue), whereas vertical lines (in red) have been equally spaced. A peak detection algorithm was applied to locate the crossing points (dots) in the recovered image, and a posterior comparison with distances to neighboring dots determined their positions in the projected pattern. Ito and Ishii [41] did not use stripes or slits for

coding, but a set of square cells (like a checkerboard), having one out of three possible intensity values. Every node (intersection between four cells of the checkerboard) was associated with the intensity values of the forming cells. In order to differentiate nodes having the same subcode, epipolar constraints between the camera and the projector were employed. The idea of using epipolar constraints was also applied in the work presented by Koninckx and Van Gool [42]. They proposed an adaptive system where green diagonal lines (named coding lines) were superimposed to a grid of vertical black lines (named base pattern). If a coding line was not coincident with an epipolar line, intersections created with the base pattern would all have laid on different epipolar lines on the camera image. This determines a unique point in the projected pattern, being able to perform the matching and the triangulation. A greater inclination of diagonal lines gave a higher density of the reconstruction, but a lower noise resistance. Therefore, the density of reconstruction could be chosen depending of how noisy the environment was, giving an adaptive robustness versus accuracy.

2.3.1.3 M-arrays

First presented by Etzion [43], M-arrays (perfect maps) are random arrays of dimensions $r \times v$ in which a sub-matrix of dimensions $n \times m$ appears only once in the whole pattern. Perfect maps are constructed theoretically having dimensions $rv = 2^{nm}$, but for real applications the zero submatrix is not considered. Thus, giving a total of $rv = 2^{nm} - 1$ unique submatrices in the pattern and a window property of $n \times m$. M-arrays represent in a two-dimensional space what DeBruijn patterns are in a one-dimensional space (see [43] and [44] for more details). Choosing an appropriate window property will determine the robustness of the pattern against pattern occlusions and object shadows for a given application. Morita et al. [45] proposed a two projection-based technique where an encoded matrix of black dots on a white background was projected, while in the second projection some black dots were removed according to a binary-encoded M-array. There are different approaches to represent non binary M-arrays, which are classified regarding the approach used to code the M-array: colored dots (color-based) or geometric features like circles and stripes (feature-based). For instance, Griffin et al. [46] generated an array of 18×66 features using an alphabet of four words 1, 2, 3, 4 comparing color and feature-based approaches. As the second approach is not color dependent, better results were obtained in presence of colored objects. Morano et al. [3] used a brute force (non-DeBruijn-based) algorithm to generate the pattern. An iterative algorithm adding one new code word and checking it against all the previous ones was

2. STATE OF THE ART ON CODED STRUCTURED LIGHT

performed. If all the distances between values were at least equal to the specified minimum Hamming distance, the new word was accepted and the next iteration was followed, until the pattern was created. The directions in which the pattern was created are indicated in Fig. 2.3.

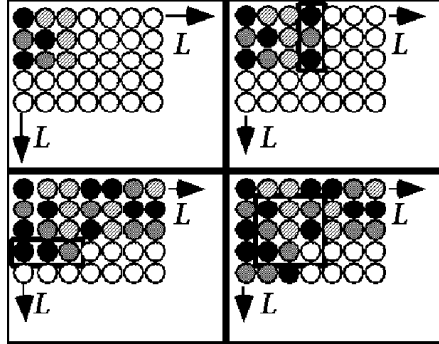


Figure 2.3: Code generation direction followed by Morano et al. [3].

This algorithm was posteriorly used by Pages et al. [27] to design a 20×20 M-array-based pattern with an alphabet of three symbols and a window property 3×3 . A color approach was used for the dots codification, using Red, Green and Blue in order to separate them in the camera sensor. The decoding algorithm analyzed the four neighbors of every dot. Once this was done, a comparison between all possible combinations of 8 neighbors was performed, in order to unequivocally locate the recorded dot in the projected pattern and perform the triangulation. A different approach has been followed by Albitar et al. [47], who used a 3×3 window property and three different symbols (black circle, circumference and stripe) to represent the codeword. As no color codification was employed, this solution presented robustness against colored objects. In the detection step, orientation of the projected pattern was extracted from the direction of the projected stripes. Once this is done, location of the symbols in the projected pattern was accomplished. Albitar et al. employed this method to create a 3D scan for medical imaging purposes (scanning of parts of the body), stating that this one-shot technique was robust against occlusions (up to a certain limit) and suitable for moving scenarios.

2.3.2 Time multiplexing

Time multiplexing methods are based on the codeword created by the successive projection of patterns onto the object surface. Therefore, the codeword associated to a position on the image is not completely formed until all patterns have been projected.

Therefore, they are used only for static scenarios. Usually the first projected pattern corresponds to the most significant bit, following a coarse-to-fine paradigm. Accuracy directly depends on the number of projections, as every pattern introduces finer resolution on the image. In addition, codeword basis tend to be small, providing higher resistance against noise. There are several approaches in sparse time multiplexing, which are exposed herebelow.

2.3.2.1 Temporal binary codes

These codes were first proposed by Posdamer and Altschuler [7] in 1982. A sequence of patterns having black and white stripes was projected onto the object. The number of stripes increased by two in every pattern, following a coarse-to-fine strategy. Therefore, the length of the codeword was given by 2^m bits, where m was the total number of projected patterns. An edge detection algorithm was employed to localize the transition between two consecutive stripes (black/white or viceversa). Moreover, Hamming distance between the codeword of two adjacent points could be maximized to reduce errors in the detection step, as was proposed by Minou et al. [48].

2.3.2.2 Temporal n-ary codes

Based on the use of n-ary codes, Caspi et al. [14] proposed a color based pattern where n^m stripes were coded in RGB space. The parameters to set were the number of colors to be used (N), the number of patterns to be projected (M) and the noise immunity factor alpha (α). For the calibration step, Caspi et al. proposed a reflectivity model given by eq. (2.3):

$$\underbrace{\begin{bmatrix} R \\ G \\ B \end{bmatrix}}_{\vec{c}} = \underbrace{\begin{bmatrix} a_{rr} & a_{rg} & a_{rb} \\ a_{gr} & a_{gg} & a_{gb} \\ a_{br} & a_{bg} & a_{bb} \end{bmatrix}}_A \underbrace{\begin{bmatrix} k_r & 0 & 0 \\ 0 & k_g & 0 \\ 0 & 0 & k_b \end{bmatrix}}_K \underbrace{\vec{P} \begin{Bmatrix} r \\ g \\ b \end{Bmatrix}}_{\vec{c}} + \underbrace{\begin{bmatrix} R_0 \\ G_0 \\ B_0 \end{bmatrix}}_{\vec{c}_0} \quad (2.3)$$

where \vec{c} is the projected instruction for a given color, \vec{P} is the non-linear transformation from projected instruction to the projected intensities for every RGB channel, A is the projector-camera coupling matrix, K the reflectance matrix (constant reflectance in every RGB channel is assumed) and C_0 is the reading of the camera under ambient light.

2. STATE OF THE ART ON CODED STRUCTURED LIGHT

2.3.2.3 Temporal hybrid codes

In order to reduce the number of projections, Ishii et al. [4] proposed a system where temporal and spatial coding were combined. The level of spatial or temporal dependence was given by the speed and accuracy requirements. For a given pixel $p(x, y)$ at time t of the projected pattern, the value was determined by using eq. (2.4):

$$I(x, y, t) = G(\lfloor \frac{x}{m} + t \rfloor \pmod{n}, y) \quad (2.4)$$

where

$$G(k, y) = G(\lfloor \frac{2^k y}{I_y} + \frac{1}{2} \rfloor \pmod{2}) \quad (2.5)$$

being G a binary image obtained from a camera at time t , n the space code size, m the light pattern width in the x direction, and I_y the image size in the y direction. There were n selectable code values for a pixel at time t , depending on the importance of temporal encoding or spatial encoding. As shown in Figure 2.4, combination of temporal and spatial information can be done from total temporal encoding (represented by $p = 1$) to total spatial encoding (given by $p = 8$). The parameter p is called the space coding weightier, as it provides an idea of how temporal or spatial the codification is.

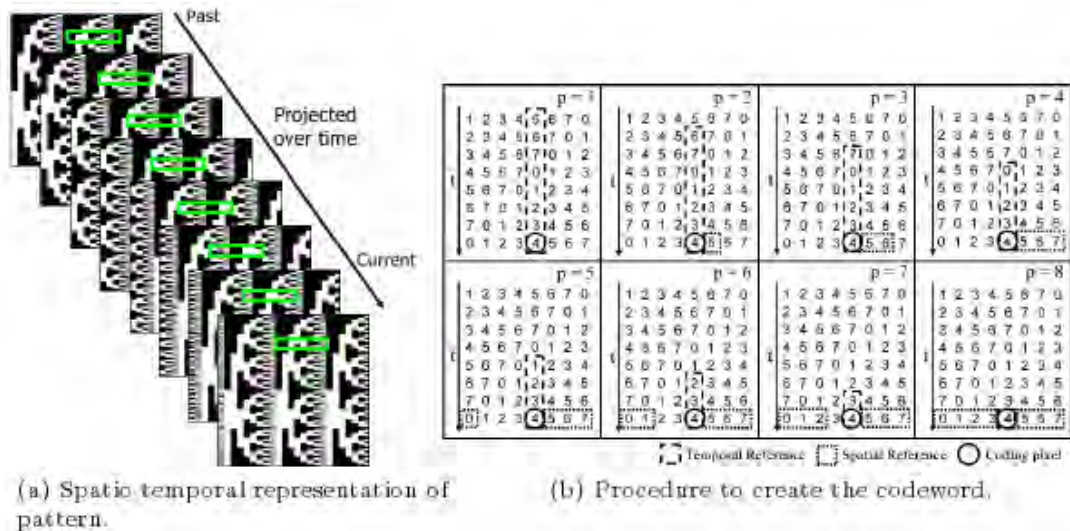


Figure 2.4: Spatial temporal algorithm proposed by Ishii et al. [4].

2.4 Dense reconstruction methods

This group of techniques provide 3D reconstruction of all the pixels captured by the image device. It is constituted by discrete or continuous shifting patterns, frequency patterns and spatial grading, showing continuous variations on intensity or color throughout one or two axes. Among these methods, the use of periodic and absolute patterns can be found. Periodic patterns are used in time multiplexing shifting methods and in frequency multiplexing. Besides, absolute patterns are based on spatial grading.

2.4.1 Time multiplexing

The same concept of time multiplexing in sparse reconstruction techniques is applied also for dense reconstruction approaches. Dense time multiplexing is represented by shifting techniques, both with discrete and continuous patterns.

2.4.1.1 Discrete shifting methods

There are some discrete implementations that use the shifting of patterns to obtain dense reconstructions. This is the case of Sansoni et al. [49], Guhring [8] and Zhang et al. [26]. The proposals of Sansoni et al. and Guhring projected a set of black and white striped patterns (like in binary codes). Afterwards, the work of Sansoni et al. projected 4 shifted versions of the last pattern, while Guhring proposal projected shifted versions of a slit-based pattern covering every pixel on the image. Binary patterns provided an absolute location of the information given by shifted patterns, avoiding ambiguities in the decoding step. Using a different strategy Zhang employed color to project DeBruijn sequences, being smoothed and shifted versions of the same pattern. The smoothing process provided subpixel accuracy to this method. In order to avoid errors due to occlusions and discontinuities, multi-pass dynamic programming (a variance of the dynamic programming proposed by Chen et al. [50]) was employed to match observed to projected patterns. An explanation of multi-pass dynamic programming can be found in Appendix D.

2.4.1.2 Continuous phase shifting methods

When projecting a sinusoidal grating onto a surface, every point along a line parallel to the coding axis can be characterized by a unique phase value. Any non-flat 3D shape will cause a deformation in the recorded pattern with respect to the projected one, which is recorded as a phase deviation. This phase deviation provides information about the

2. STATE OF THE ART ON CODED STRUCTURED LIGHT

illuminated shape. Matching the recovered image with the projected pattern, the object shape is recovered. The pattern must be shifted and projected several times in order to extract the phase deviation (this is not the case of frequency multiplexing approaches). Due to the grayscale nature of the projected patterns, they present advantages like resistance to ambient light and resistance to reflection variation. Depending on the number of frequencies used to create the pattern, we can distinguish between single and multiple Phase Shifting (PS) methods.

Single phase shifting

These techniques use only one frequency to create the sequence of patterns. In order to recover phase deviation, the pattern is projected several times, every projection shifted from the previous projection by a factor of $\frac{2\pi}{N}$, being N the total number of projections, as shown in eq. (2.6) (super-index P indicates the projected pattern):

$$I_n^p(y^p) = A^p + B^p \cos(2\pi f_\phi y^p - 2\pi n/N) \quad (2.6)$$

where A^p and B^p are the projection constants and (x^p, y^p) are the projection coordinates, $n = 0, 1, \dots, N$. The received intensity values from the object surface, once the set of patterns is projected is:

$$I_n(x, y) = \alpha(x, y) [A + B \cos(2\pi f_\phi y^p + \phi(x, y) - 2\pi n/N)] \quad (2.7)$$

As can be observed from eq. (2.7), it suffers of intensity and phase deviation, being necessary to cancel the effect of different albedo ($\alpha(x, y)$) to correctly extract the phase. This is shown in eq. (2.8):

$$\phi(x, y) = \arctan \left[\frac{\sum_{n=1}^N I_n(x, y) \sin(2\pi n/N)}{\sum_{n=1}^N I_n(x, y) \cos(2\pi n/N)} \right] \quad (2.8)$$

From a minimum of three projected shifted patterns is possible to create a relative phase map and to reconstruct the phase deviation caused by the object shape. However, the arctan function returns values between the range $(-\pi, \pi]$ and therefore a phase unwrapping procedure is necessary to work with a non-ambiguous phase value out of the wrapped phase. This is the reason why these patterns provide effective dense reconstruction only under the restriction of smoothed surfaces.

Phase shifting methods has been used in a variety of applications during the last years. For instance, Ono et al. [51] created the so-called correlation image sensor

(CIS), a device which generates temporal correlations between light intensity and three external reference signals on each pixel using phase shifting and a space-temporal unwrapping. Some approaches using phase shifting have also been developed from the work proposed by Srinivasan et al. [52].

One of the drawbacks of phase shifting methods is the necessity to project several patterns in time, which is more than the theoretic minimum of three patterns considered for real conditions. A solution to reduce the total time required in the projection step is to multiplex the patterns either in color space or in frequency. Following this idea, Wust and Capson [53] proposed a method which projected three overlapping sinusoidal patterns shifted 90 degrees and coded in red, green and blue. Therefore, in this way the camera recorded phase deviation of every pattern in a different color channel and a normal phase extraction algorithm like the one shown in eq. (2.9) could be used:

$$\Phi(x, y) = \arctan\left(\frac{I_r - I_g}{I_g - I_b}\right) \quad (2.9)$$

where $\Phi(x, y)$ is the phase of a given pixel, and I_r , I_g and I_b are the red, green and blue intensities, respectively.

A different approach was proposed by Guan et al. [5], where the patterns were combined in frequency using the orthogonal dimension, as shown in Fig. 2.5. Basically, a traditional band pass filtering was performed to the recorded pattern, as it is theoretically done in communications for frequency multiplexing. This step filters noise without suppressing the information hold in the surroundings of the carriers. In particular, [5] used a maximally-flat magnitude Butterworth filter. Once this step was done, a normal phase extraction was performed over the obtained patterns. This method provided higher signal to noise ratio than color multiplexing approaches and it was not dependent on the surface color. However, some errors arose in presence of different albedo and abrupt shape variations.

Multiple phase shifting (MPS)

The use of more than one frequency in phase shifting comes to cope with the uncertainty created in the extracted wrapped phase. As stated in the remainder theorem [54], an absolute phase map can be computed from two different relative phase maps having frequencies that are relative prime numbers. This principle was used by Gushov and Solodkin [55] for interferometry, where an interferometer able to deal with vibrations or relief parameters was constructed. More recently, Pribanic et al. [9] presented a multiple-phase shifting-based technique where only two patterns were used to create

2. STATE OF THE ART ON CODED STRUCTURED LIGHT

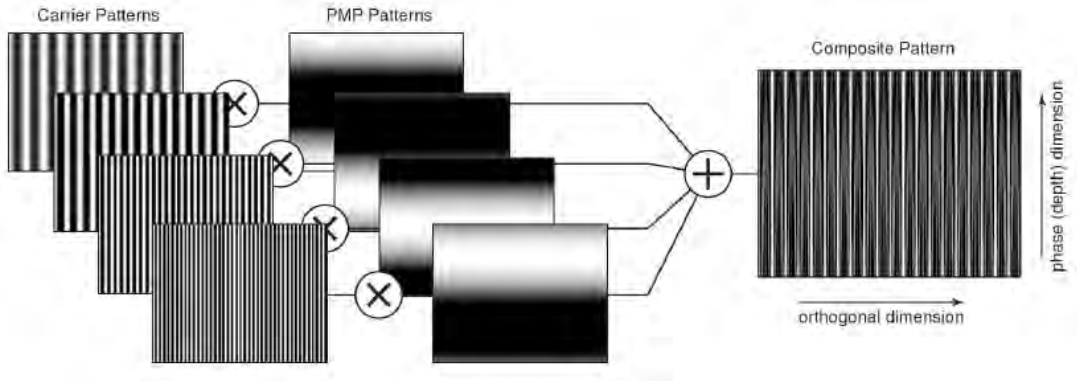


Figure 2.5: Composite pattern formed by the multiplexation of modulated phase shifting profilometry (PMP) patterns using the algorithm of Guan et al. [5].

the relative phase maps. Two sinusoidal patterns were shifted and projected in time, in order to recover phase deviation (see Fig. 2.6). From these sets of images was possible to obtain two relative phase maps, using normal phase shifting decoding algorithms (as shown in eq. (2.8)). Having this, the absolute phase map was recovered (an explanation about the use of the Remainder Theorem in Multiple Phase Shifting pattern projection can be found in Appendix C). This map can be directly compared to the ideal phase-shifting map, providing correspondences for the triangulation step. The algorithm was tested for different pairs of frequencies over a flat surface. Finally the reconstruction of a footprint and a face were pursued, providing small 3D reconstruction errors.

2.4.2 Frequency multiplexing

Frequency multiplexing methods group all the techniques where phase decoding is performed in the frequency domain rather than in the spatial domain. There are different approaches depending on the frequency analysis performed to the image. Fourier Transform has been traditionally used to extract the depth information from the information provided by the recovered phase. However, other techniques like Spatial Phase Detection, Windowed Fourier Transform, Wavelet Transform are also employed.

2.4.2.1 Fourier Transform

Fourier Transform (FT) was introduced to solve the necessity of having a phase-shifting-based method for moving scenarios. FT was first proposed by Takeda and Mutoh [56], who extracted depth from one single projected pattern. A sinusoidal grating was projected onto the object, and the reflected deformed pattern was recorded. The

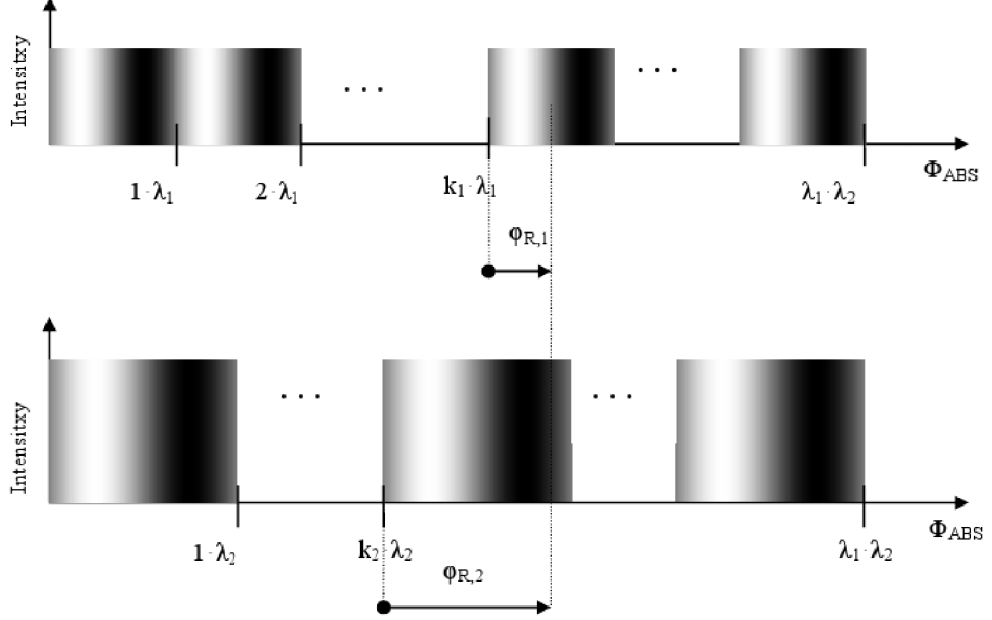


Figure 2.6: Pair of projected sinusoidal patterns, having two different frequencies (k is the number of periods).

projected signal for a sinusoidal grating was represented in eq. (2.10):

$$I_n^p(y^p) = A^p + B^p \cos(2\pi f_\phi y^p) \quad (2.10)$$

Once reflected onto the object, phase component was modified by the shape of the object, thus giving an intensity value expressed in eq. (2.11):

$$I(x, y) = \alpha(x, y)[A + B \cos(2\pi f_\phi y^p + \phi(x, y))] \quad (2.11)$$

Phase component must be isolated to extract shape information. This was achieved performing a frequency filtering in the Fourier domain. The background component was suppressed and a translation in frequency was done to bring the carrier component (which holds the phase information) to zero frequency axes. Applying the sequence of equations showed below, the phase can be extracted from the signal. First, the input signal was rewritten as shown in eq. (2.12):

$$I(x, y) = a(x, y) + c(x, y)e^{2\pi i f_\phi y^p} + c^*(x, y)e^{-2\pi i f_\phi y^p} \quad (2.12)$$

2. STATE OF THE ART ON CODED STRUCTURED LIGHT

where

$$c(x, y) = \frac{1}{2}b(x, y)e^{i\phi(x, y)} \quad (2.13)$$

and $c^*(x, y)$ is the complex value of constant $c(x, y)$. Finally, the phase component was extracted from the imaginary part of eq. (2.14):

$$\log[c(x, y)] = \log\left[\left(\frac{1}{2}\right)b(x, y)\right] + i\phi \quad (2.14)$$

The obtained phase component ranges from $(-\pi, \pi]$, being necessary to apply an unwrapping algorithm in order to obtain a continuous phase related to the object. Once the phase was unwrapped, the relative depth information was extracted using eq. (2.15):

$$h(x, y) = L \cdot \frac{\Delta\phi(x, y)}{(\Delta\phi(x, y) - 2\pi f_0 d)} \quad (2.15)$$

where L is the distance to the reference plane and d is the distance between the camera and the projector devices. FT has been widely used in industrial applications. For instance, Cobelli et al. [57] used FT for global measurement of water waves. In their work, two sources of noise were considered in the filtering step. The first one was related with illumination inhomogeneities of background variations over the field of view, which remains present as an additive variation. The second one was due to the local surface reflectivity. As this reflection varies much slower than the sinusoidal modulation impinged on the surface, it can also be treated as background noise. Thus, both sources of noise can be suppressed using the background component filtering procedure proposed by Takeda et al. [56]. Due to the periodic nature of the projected pattern, this method was constrained by the maximum reconstructible slope given by eq. (2.16):

$$\left|\frac{\partial h(x, y)}{\partial x}\right|_{MAX} < \frac{L}{3d} \quad (2.16)$$

In order to increase this slope limitation, Su et al. [10] proposed the so-called π -phase shifting FT. Two sinusoidal patterns were projected using this method, being the second one a half-period shifted version of the first one. This solution multiplies by three the detectable range in depth slope. This principle was used by Hu and He [58] to scan moving objects having uniform velocity (like in an assembly line). In their work two scan line cameras were used, and one single pattern was projected. The distance between the two cameras corresponded to half the period of the grating. As the velocity of the object was known, matching two scanning of the same point at different instants

of time could be done. This procedure avoids the projecting of two patterns, and takes advantage of the uniform motion present in assembly lines.

There are some proposals that combine both π -phase shifting FT patterns in one single projected pattern using color or frequency multiplexing. For instance, Chen et al. [59] used color space to project a bi-color sinusoidal fringe pattern consisting in the sum of π -phase shifting FT patterns, represented by blue and green patterns. Another approach was considered by Yue et al. [60]. In this work the same principle used by Guan for phase shifting was developed for FT. Appropriate carrier frequencies were chosen regarding the characteristics of the projector and camera used, assuming that the Nyquist sampling theorem was satisfied. These frequencies were kept away from zero frequency as much as possible. When analyzing the results, standard deviation error is slight lower than for normal FT, while accuracy remains unaltered.

In the case of scanning coarse objects where discontinuities and speckle-like structures can appear, two dimensional FT filtering must be used [34], as it permits better separation of the desired information from noise. This is due to the fact that noise is normally 2D distributed in a fringe pattern, having a spectra scattered in a 2D frequency domain. For instance, Hung and more recently Lin and Su [61] proposed a method for 2D FT scanning where the filtering step, aimed to prevent from frequency spreading, was performed using a 2D Hanning window. However, some other filters having similar characteristics can also be used. This is the case of Chen et al. [62], who applied a Gaussian filter. 2D FT filtering has been used by Berryman et al. [63] to create a low cost automated system to measure the three dimensional shape of the human back, obtaining an accuracy of $\pm 1mm$.

Spatial Phase Detection (SPD) constitutes an alternative to FT. This method was initially proposed by Toyooka and Iwaasa [64]. The analysis of the received signal (eq.(2.17)) is done using the sine and cosine functions, as can be observed in eq.(2.18), eq.(2.21):

$$I(x, y) = \alpha(x, y) [A + B\cos(2\pi fy^p + \phi(x, y))] \quad (2.17)$$

$$I_c(x, y) = \alpha(x, y) [A + B\cos(2\pi fy^p + \phi(x, y))] \cdot \cos(2\pi fy^p) \quad (2.18)$$

$$= \alpha(x, y) \cdot A\cos(2\pi fy^p) + \frac{1}{2}\alpha(x, y) \cdot B\cos(4\pi fy^p) + \frac{1}{2}\alpha(x, y) \cdot B\cos(\phi(x, y)) \quad (2.19)$$

$$(2.20)$$

2. STATE OF THE ART ON CODED STRUCTURED LIGHT

$$I_s(x, y) = \alpha(x, y) [A + B \cos(2\pi f y^p + \phi(x, y))] \cdot \sin(2\pi f y^p) \quad (2.21)$$

$$= \alpha(x, y) \cdot A \sin(2\pi f y^p) + \frac{1}{2} \alpha(x, y) \cdot B \sin(4\pi f y^p - \frac{1}{2} \alpha(x, y) \cdot B \sin(\phi(x, y))) \quad (2.22)$$

$$(2.23)$$

Now $\phi(x, y)$ varies slower than any term containing f and so only the last term in each new function is a low-frequency term. This part of the function can then be extracted by low-pass filtering. Regarding the Euler's formula for the sine and cosine functions and the principles of Fourier Transform applied on sinusoidal functions [65], this step provides similar results than obtaining the real and the imaginary components of the Fourier Transform applied to the incoming signal. Therefore, the last step is to extract the phase component from these components, which is obtained by applying the arctangent function (eq.(2.24)):

$$\phi(x, y) = \arctan \left[\frac{r(x, y) * I_s(x, y)}{r(x, y) * I_c(x, y)} \right] \quad (2.24)$$

where $r(x, y)$ represents a low-pass filter, and $*$ denotes convolution. It is important to note that Toyooka and Iwaasa use integration to extract the phase terms, whereas other authors using related spatial domain methods apply different low-pass filters [66]. As in FT, this method suffers from leakage distortion when working with fringe patterns, as no local analysis is performed to avoid spreading errors due to discontinuities and different albedo.

2.4.2.2 Window Fourier Transform

The task of suppressing the zero component and avoiding the frequency overlapping between background and data (the leakage distortion problem) has also been studied using other frequency-based approaches.

This is the case of the **Windowed Fourier Transform** (WFT), which splits the signal into segments before the analysis in frequency domain is performed. The received signal is filtered applying the WFT analysis transform shown in eq.(2.25, 2.27)

$$Sf(u, v, \xi, \eta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \cdot g(x - u, y - v) \cdot \exp(-j\xi x - j\eta y) dx dy \quad (2.25)$$

being (x, y) , (ξ, η) the translation and frequency coordinates respectively, and $g(x, y)$ the windowing function. When $g(x, y)$ is a Gaussian window, the WFT is called a Gabor transform; that is:

$$g(x, y) = \frac{1}{\sqrt{\pi\sigma_x\sigma_y}} \cdot \exp\left(-\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2}\right) \quad (2.26)$$

where σ_x and σ_y are the standard deviations of the Gaussian function in x and y , respectively. Eq.(2.25) provides the 4-D coefficients $Sf(u, v, \xi, \eta)$ corresponding to the 2D input image. The windowing permits the WFT to provide frequency information of a limited region around each pixel. The Gaussian window is often chosen as it provides the smallest Heisenberg box [67]. Once the 4D coefficients are computed, the phase can be extracted. There are two main techniques for phase extraction in WFT: Windowed Fourier Filtering (WFF) and Windowed Fourier Ridge (WFR). In WFF the 4D coefficients are first filtered, suppressing the small coefficients (in terms of its amplitude) that correspond to noise effects. The inverse WFT is then applied to obtain a smooth image:

$$f(\bar{x}, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\eta_1}^{\eta_h} \int_{-\xi_1}^{\xi_h} Sf(u, v, \xi, \eta) \cdot g_{u,v,\xi,\eta}(x, y) d\xi d\eta du dv \quad (2.27)$$

where:

$$Sf(u, v, \xi, \eta) = \begin{cases} Sf(u, v, \xi, \eta) & \text{if } |Sf(u, v, \xi, \eta)| > \text{threshold} \\ 0 & \text{if } |Sf(u, v, \xi, \eta)| < \text{threshold} \end{cases} \quad (2.28)$$

The estimated frequencies $\omega_x(x, y)$ and $\omega_y(x, y)$ and corresponding phase distribution is obtained from the angle given by the filtered WFF, as explained in [67]. In WFR, however, the estimated frequencies are extracted from the maximum of the spectrum amplitude, as shown in eq.(2.29).

$$[\omega_x(u, v), \omega_y(u, v)] = \operatorname{argmax}_{\xi,\eta} |Sf(u, v, \xi, \eta)| \quad (2.29)$$

The phase can be directly obtained from the angle of the spectrum for those frequency values selected by the WFR (phase from ridges), or integrating the frequencies (phase by integration). An optional cost function can be applied to find the optimal angle values of a pixel taking into account the angle values of the neighboring pixels. Phase from ridges represents a better solution than phase from integration (despite

2. STATE OF THE ART ON CODED STRUCTURED LIGHT

some phase correction may need to be applied [67]), as in phase from integration errors are accumulated and lead to large phase deviations. Using WFT, Chen et al. [59] proposed the use of Windowed Fourier transform (Gabor transform) to eliminate the zero spectrum. However, as was demonstrated by Gdeisat et al. [68], Chen's technique was not able to eliminate the zero spectrum neither in fringe patterns that have large bandwidths, nor in cases where the existence of large levels of speckle noise corrupts the fringe patterns. This is mainly caused by an erroneous selection of the width and shape of the window for the Fourier analysis. The window size must be small enough to reduce the errors introduced by boundaries, holes and background illumination, at the same time it must be big enough to hold some periods and hence allow the detection of the main frequency to perform an optimal filtering. However, in applications where the frequency varies considerably during the analysis (in space or in time) this trade-off is difficult to achieve and noise arises due to a wrong frequency detection.

2.4.2.3 Wavelet Transform

Wavelet Transform (WT) was proposed to solve the aforementioned trade-off. In WT the window size increases when the frequency to analyze decreases, and vice-versa. This allows to remove the background illumination and prevent the propagation of errors produced during the analysis, which remain confined in the corrupted regions alone [68]. Additionally the leakage effects are reduced, avoiding having large errors at the edges of the extracted phase maps. The Continuous Wavelet Transform (CWT) is a sub-family of WT that perform the transformation in the continuous domain. Moreover, it is common to use CWT with complex wavelets for the analysis of the fringe patterns [69]. The 1D-CWT algorithm analyses the fringe pattern on a row by row basis, whereas the 2D-CWT algorithm is an extension of the analysis to the two dimensional space. In 2D analysis a 4D transform is obtained from WT (the daughter wavelets are obtained by translation, dilation and rotation of the previously selected mother wavelet). Once this is performed, phase extraction is pursued using the phase from ridges or the phase by integration algorithms, also named phase estimation and frequency estimation (similarly to WFT). As in WFT, it has been proven that the phase from ridges provides better results than the phase from integration, due to the accumulative effect in the phase from integration algorithm [69]. The work done by Gdeisat et al. [68] applied a two dimensional wavelet function to the recovered image, based on phase from ridges extraction. Rotation and scale were considered jointly with x and y coordinates resulting in a four dimensional wavelet transform. To

apply the transformation, the mother wavelet $\psi(x, y)$ must satisfy the admissibility condition. Under this condition Gdeisat used a Differential of Gaussian as the mother wavelet, while Zhang [70] employed a 2D complex Morlet wavelet. Four sub-images were created at one iteration of the wavelet decomposition algorithm, corresponding to the low and high frequencies in both axes. Phase component was extracted from the ridge information present in the corresponding high frequency sub-image. The task of choosing appropriate values for rotation and scale parameters determined the results of filtering and phase extraction. Related to this, a novel method for choosing the adaptive level of discrete wavelet decomposition has been proposed by Zhang et al. [70]. They have achieved higher accuracy in the principal frequency estimation and low frequency energy suppression against traditional zero suppression algorithms used in FT. However, some problems arise related to the relationship between the window size and the frequency of the fringes. In WT the window size increases when the horizontal or vertical fringe frequencies decrease. This can be a troublesome for the analysis of some fringe patterns where the carrier frequency is extremely low or high, as was pointed out by Kemaio et al. [71]. Moreover, in computational applications a dyadic net is used to generate the set of wavelet functions. That is, the size of the wavelet is modified by the factor 2^j . This can lead to some problems in applications like fringe pattern analysis, where the change in the spatial fringe frequencies throughout the image is not high enough to produce a relative variance of 2^j in the size of the optimal wavelet.

2.4.2.4 The problem of phase unwrapping

Phase unwrapping represents a crucial step in frequency multiplexing techniques. In absence of noise, if all phase variation between neighboring pixels is less than π , the phase unwrapping procedure can be reduced to add the corresponding multiple of 2π when a discontinuity appears. Unfortunately, noise, local shadows, under-sampling, fringe discontinuities and irregular surface brightness make the unwrapping procedure much more difficult to solve. Plenty of approaches have been presented ([72],[73],[68]). For instance, phase unwrapping based on modulation follows an iterative algorithm, starting from the pixel with higher intensity value and comparing it to the pixels inside a 3×3 surrounding square region. The comparison step is done one by one, queuing the affected pixels from maximum to minimum intensity. This method can also be applied when dealing with moving objects, substituting the searching area to a $3 \times 3 \times 3$ voxel. Besides, Wu and Peng [72] presented a phase unwrapping algorithm based on region

2. STATE OF THE ART ON CODED STRUCTURED LIGHT

growing. The phase was unwrapped from the smoothest area to the surroundings, according to a linear estimation. In order to decrease the error, a quality map was used to guide the unwrapping. The map can be defined in different ways as far as it provides quality information. For instance, second-order partial derivative can be used to determine the pixels to unwrap; that is, those pixels having this value lower than a specified threshold. Statistical methods can also be used considering the variance within a mask, for every pixel. Finally, Gorthi and Lolla [74] projected an extra color-coded pattern, which can be unequivocally identified once the image was captured, thus giving a rough information about the required phase to add or subtract in the unwrapping step. A further explanation of different unwrapping methods used in profilometry can be found in [75].

2.4.2.5 Alternatives to sinusoidal grating

Not all frequency transform methods use sinusoidal fringes for the projected pattern. As Huang et al. [76] stated, structured light techniques based on sinusoidal phase-shifting methods have the advantage of pixel level resolution, large dynamic range and few errors due to defocussing. However, the arctangent computation make them relatively slow. As an alternative, they used three 120 degrees phase-shifted trapezoidal fringe patterns. The phase deviation was extracted from the so-called intensity-ratio image, shown in eq. (2.30).

$$r(x, y) = \frac{I_{med}(x, y) - I_{min}(x, y)}{I_{max}(x, y) - I_{min}(x, y)} \quad (2.30)$$

where $I_{min}(x, y)$, $I_{med}(x, y)$ and $I_{max}(x, y)$ are the minimum, median, and maximum intensities of the three patterns for the image point (x,y). Image defocus does not cause major errors when using sinusoidal pattern, as it is still sinusoidal when the image is defocused. However, errors caused by blurring have to be taken into account when dealing with trapezoidal patterns. Modeling these errors as a Gaussian filtering, Huang et al. experiments yielded defocussing errors not bigger than 0.6%. More recently, another approach using triangular patterns has been proposed by Jia et al. [77]. This approach used only two triangular patterns shifted half the period, making it more feasible to be implemented in real time applications. Ronchi grating has also been used in pattern projection as an alternative to sinusoidal grating. This is the case of Lin and Su [61], who proposed an algorithm where only one pattern was needed. Phase

information was obtained taking the imaginary part of eq. (2.31):

$$\Delta\Phi(x, y) = \log[\hat{I}(x, y)\hat{I}_0^*(x, y)] \quad (2.31)$$

where $\hat{I}(x, y)$ and $\hat{I}_0(x, y)$ are the recorded illumination from the setup and the reference plane, respectively. A Ronchi grating was also used by Spagnolo et al. [78] in real applications, in order to recover 3D reconstructions of artwork surfaces.

2.4.3 Spatial multiplexing (grading methods)

Grading methods refer to all techniques containing the entire codeword for a given position only in its pixel value. Therefore, the resolution can be as high as the pixel resolution of the projector device is. However, these methods suffer from high sensitivity to noise and low sensitivity to surface changes, due to the short distances between the codeword of adjacent pixels. This is the reason why some authors use these methods introducing temporal redundancy, projecting the same pattern several times. As a drawback, note that restriction to static scenarios is imposed when projecting more than one pattern. There are two main techniques based on grading methods: grayscale-based patterns and color-based patterns. Regarding grayscale based methods, Carrihill and Hummel [11] proposed a linear grayscale wedge spread going from white to black, along the vertical axis. The authors achieved a mean error of $1cm$, due to the high sensitivity to noise and non-linearity of the projector device. In color-based patterns, the pixel is coded using color instead of grayscale values. As a drawback, color calibration is required. Tajima and Iwakawa [30] presented a rainbow pattern codified in the vertical axis. In order to project this spectrum, a nematic liquid crystal was used to diffract white light. Two images were projected to suppress the effect of colored surfaces.

2.5 Experimental results

In order to test the effectiveness of the different strategies proposed in the literature a set of 6 representative techniques of table 2.1 have been implemented and compared. These methods are presented in table 2.2:

2. STATE OF THE ART ON CODED STRUCTURED LIGHT

Table 2.2: Selected methods, with their main attributes.

Group		Method	Characteristics
Sparse coding	Spatial m.	Monks et al. [6]	De Bruijn slits pattern. 6 Hue colors (1 pattern)
Sparse coding	Time m.	Posdamer et al. [7]	Stripes patterns. 7 bits Gray code (24 patterns)
Dense coding	Time m. (PS)	Guhring [8]	Time multiplexing + shifting (16 patterns)
Dense coding	Time m. (PS)	Pribanic et al. [9]	Multiple Phase Shifting (18 patterns)
Dense coding	Frequency m.	Su et al. [10]	Sinusoidal pattern, π -shifting (2 patterns)
Dense coding	Spatial m.	Carr. Hummel [11]	Grading grayscale pattern (1 pattern)

Two sparse coding and four dense coding techniques have been chosen and implemented. It is important to mention that all the methods presented here have been implemented directly from the corresponding papers (original code was not available), and the parameters have been set in order to obtain optimal reconstruction results. Among sparse coding spatial multiplexing, one axis coding was chosen as it presents an easier decoding algorithm than two axes coding. Among them, Monks et al. [6] technique presents a color slits pattern based technique that provides bigger vocabulary than grayscale approaches as well as easier detection and matching than stripes patterns techniques. Among sparse coding time multiplexing, Posdamer algorithm [7] was selected for being a well known effective technique in time multiplexing. Among dense coding time multiplexing, shifting codes are proposed by Sansoni et al. [49] and Guhring [8]. Between them, Guhring method was selected because it uses slits shifting, easier to segment than the fringes shifting used by Sansoni et al. Moreover, the technique presented by Pribanic et al. [9] was selected for being the latest time multiplexing technique using multiple phase shifting. In continuous frequency multiplexing, π -phase shifting FTP method proposed by Su et al. [10] provides higher resistance to slopes than the traditional FTP of Takeda and Mutoh [56], without the necessity to perform Wavelet filtering or having to deal with blurring associated to non-sinusoidal patterns. Chen et al. [59] and Yue et al. [60] use the same π -phase shifting FTP multiplexing the patterns into one single projection. However, the main idea remains unaltered, and therefore the simpler solution proposed by Su et al. is still a good representative to evaluate the performance of these techniques. Finally, the greyscale spatial grading proposed by Carrhill and Hummel [11] was chosen against the rainbow pattern implemented by Tajima and Iwakawa [30] which employs a nematic liquid crystal.

The setup used for the tests was composed of an LCD video projector (Epson EMP-400W) with a resolution of 1024×768 pixels, a camera (Sony 3CCD) and a frame grabber (Matrox Meteor-II) digitizing images at 768×576 pixels with 3×8 bits

per pixel (RGB). Both camera and video projector were calibrated using the projector camera calibration method explained in appendix A. The baseline between camera and projector was about $1m$. The results and time estimates were computed using a standard Intel Core2 Duo CPU at 3.00GHz and 4GB RAM memory. The algorithms were programmed and ran in Matlab 7.3.

2.5.1 Qualitative results

The reconstruction of a real object permits to analyze the performance of the programmed techniques in terms of accuracy and noise sensitivity. The reconstructed object used to perform the qualitative analysis of the results is a ceramic figure placed at a distance of about $80cm$ to the camera. In order to show the results, both 3D cloud of points and surfaces are used. The surface has been generated performing a 2D Delaunay triangulation over (x,y) coordinates. It is important to mention that no smoothing step has been applied to the 3D points or surfaces. Therefore the results are compared without a post-processing step.

As can be observed in Fig. 2.7, and Fig. 2.8, the best results are obtained with time multiplexing shifting approaches (the case of Guhring [8] and Pribanic et al. [9]). These techniques obtain the best accuracy results providing also dense reconstruction. Furthermore, both algorithms perform well in presence of surface slopes, as can be observed in some details of the reconstructed object (see for instance the ears of the horse). However, the number of projections necessary to reconstruct the object is more than one, which make them unable to cope with moving scenarios. This is also the case of the original time multiplexing algorithm proposed by Posdamer et al. [7], which has also been implemented in order to compare it to the other techniques. Despite the fact that the resolution obtained is higher than the other sparse coding techniques, it suffers some noise in the recovered cloud of points. This is mainly caused by nonlinearities of the camera, which produces some leakage from white to black fringes that can lead to some errors in the position of the recovered edges. Among one-shot techniques, DeBruijn based coding present the best results in terms of accuracy. This is the case of Monks algorithm [6], which employs DeBruijn color coding to obtain a dynamic sparse reconstruction. It is important to mention that its resolution is doubled in the approach of Pages et al. [2], which employed a more complex codification algorithm. In spite of using 6 color values in the Hue channel, previous color calibration is required. This can be simply done by comparing projected and recovered colors and working under light controlled conditions (i.e. dark scenario), or performing a full color calibration step

2. STATE OF THE ART ON CODED STRUCTURED LIGHT

using a colorimeter, in order to obtain a model like the one shown in eq. (2.3). Another approach was proposed by Su et al. [10], which employs frequency multiplexing (the π -phase shifting). This provides also one-shot dense reconstruction. Recent proposals combined the two patterns into one single projection ([60], [59]), following the same initial ideas proposed by Su et al. However, high frequencies are lost in the filtering step, causing the loss of some information in the surface details. Moreover, traditional frequency multiplexing approaches can work only on smooth surfaces having slopes not exceeding three times the value given in eq. (2.16). It is important to mention that the method chosen for phase unwrapping employs a qualitative map to determine the region where the unwrapping should start. Finally, the grading technique proposed by Carrihill and Hummel [11] resulted highly sensitive to noise and low sensitive to changes in depth, caused by the low range existing between adjacent pixels.

2.5.2 Quantitative results

Quantitative results have been analyzed reconstructing a white plane at a distance of about 80cm to the camera. Principle Component Analysis (PCA) was applied to obtain the equation of the 3D plane for every technique and for every reconstruction. This technique is used to span the 3D cloud of points onto a 2D plane defined by the two eigenvectors corresponding to the two largest eigenvalues. The results of the experiment are shown in table A.1. Observe that the algorithm of Su et al. [10] is conceived to measure deviation of smooth surfaces with respect to the reference plane, therefore a plane is not conceived to be reconstructed by depth deviation.

Table 2.3: Quantitative results. The headings are: author’s name of the technique; average deviation of the reconstructing error; standard deviation of the reconstructing error; number of 3D points reconstructed; number of projected patterns.

Technique	Average (mm)	Stdev (mm)	3D Points	Patterns	Time (s)
Monks et al.	1.31	1.19	13899	1	45.29
Posdamer et al.	1.56	1.40	25387	14	32.18
Guhring	1.52	1.33	315273	24	158.22
Pribanic et al.	1.12	0.78	255572	18	165.65
Su et al.	—	—	—	1	—
Carr.and Hummel	11.9	5.02	202714	1	150.57

As can be observed, among the techniques obtaining sparse reconstruction, De-Bruijn one-shot projection algorithm developed by Monks et al. [6] presents the best results in terms of average error and standard deviation, against traditional time multiplexing represented by Posdamer et al. [7]. Dense reconstruction techniques can be

divided into one-shot and multiple pattern projection techniques. Among one shot techniques, the technique proposed by Carrihill and Hummel [11] obtains the poorest results due to the low variance existing between adjacent pixels in the projected pattern. In contrast, Fourier analysis represented by the proposed technique presents lower error rate thanks to the frequency filtering process that is performed in the analysis. Among multiple pattern projection techniques the method developed by Pribanic et al. [9] gives the best results in terms of sensitivity to noise, as can be extracted from the values of average error and standard deviation. Regarding the computing time it can be observed that methods obtaining dense reconstructions (the case of Guhring, Pribanic et al., Su et al., and Carrihill and Hummel) need to compute more 3D points, requiring higher computational time. Among methods providing sparse reconstruction the color calibration step makes Monks et al. algorithm slower than Posdamer et al., despite it preserves the same order of magnitude. Finally, the computational cost of the algorithm proposed by Su et al. was tested using the dataset of previous section (qualitative results). The computational time was 108.52s, the same order of magnitude achieved in the other algorithms providing dense reconstructions.

2. STATE OF THE ART ON CODED STRUCTURED LIGHT

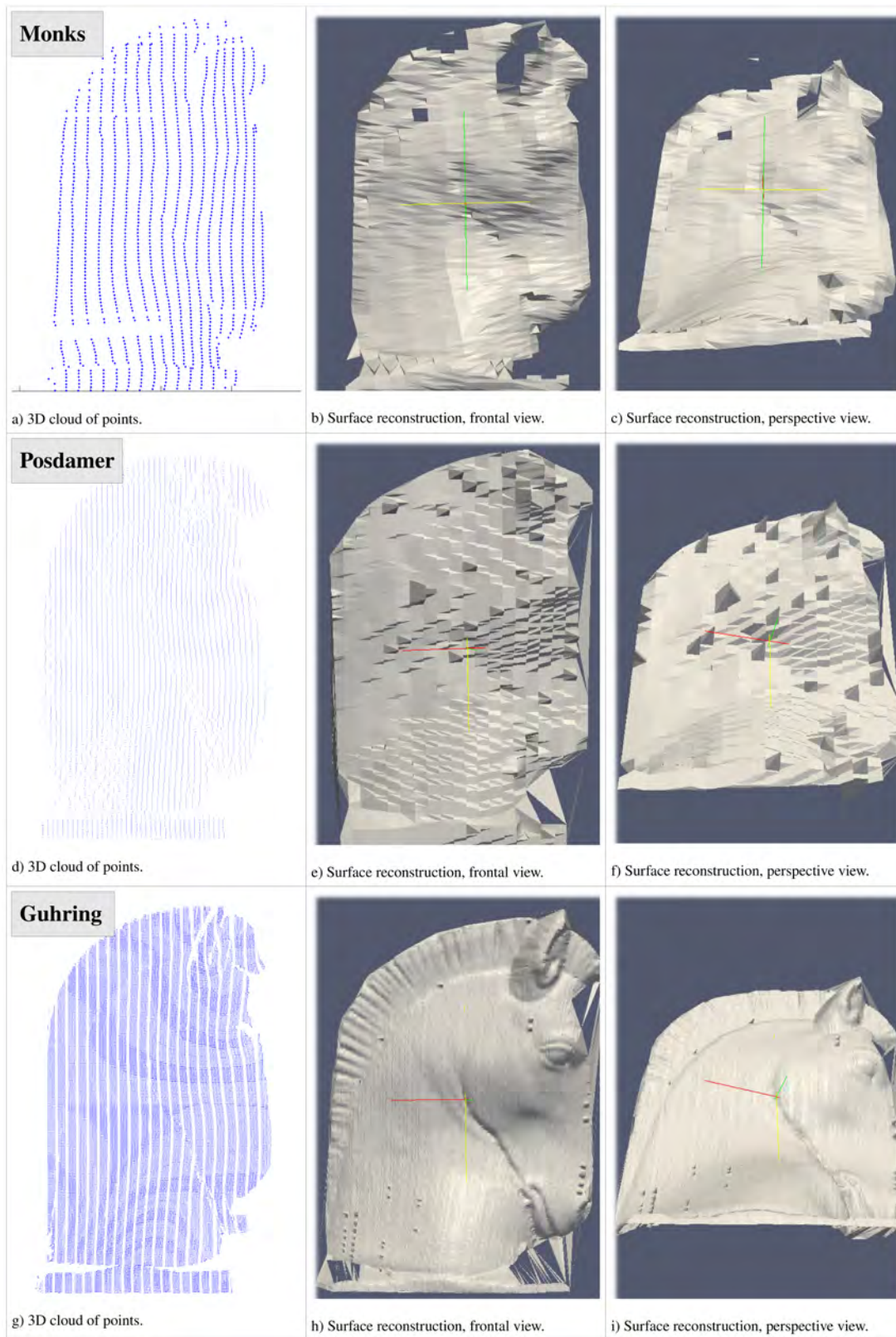


Figure 2.7: Results of Monks et al. [6], Posdamer et al. [7] and Guhring [8], respectively.

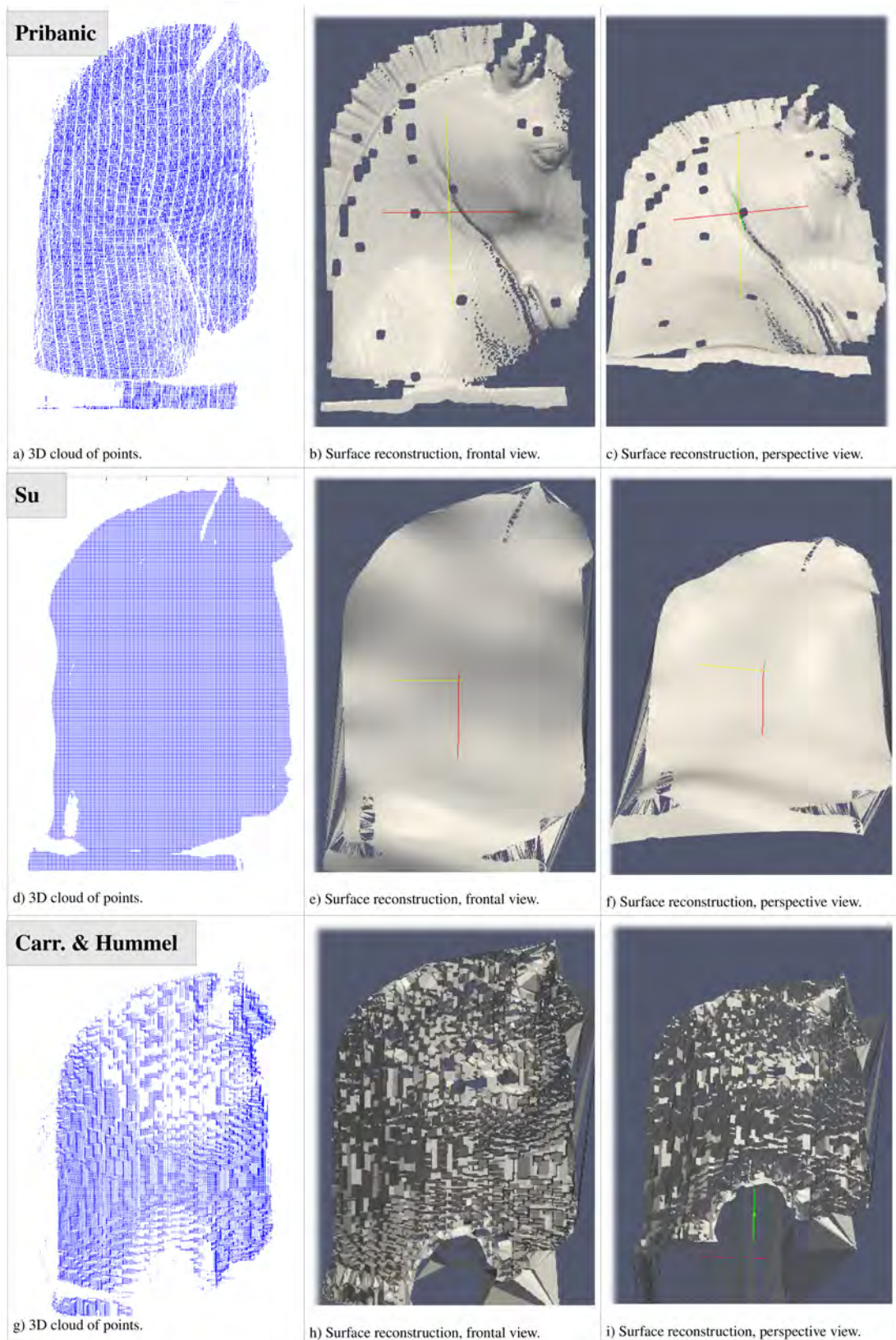


Figure 2.8: Results of Pribanic et al. [9], Su et al. [10], and Carrihill and Hummel [11], respectively.

2.6 Conclusions

In this chapter a state of the art of the methods existing in coded structured light has been pursued. A new classification of the different CSL techniques has been proposed embracing and updating the spatial, temporal and frequency multiplexing strategies existing in the literature. Common attributes to all the techniques have been analyzed and compared. Moreover, an update of the contributions done during the last years has been performed. Two main groups have been distinguished depending on the sparse or dense nature of the reconstruction.

Sparse reconstruction coding is created using stripes or slits in a unique axis codification, or geometric features (circles, checkerboard) or multi-slit for two axes codification. The final resolution depends on the number of encoded features present in the pattern. Among sparse coding techniques, spatial multiplexing and temporal multiplexing are distinguished. In the former, the codeword is determined by the pixel value in the pattern and the values of its surrounding pixels. In the later, the codeword is created by the sequence of patterns projected onto the surface. Spatial multiplexing needs only one or few patterns to create the codeword, being usually suitable for moving scenarios. However, they present lower spatial resolution than time multiplexing techniques, as all the information must be condensed in less projections. Among them, techniques based on De Bruijn codes, M-arrays and non-formal codification can be distinguished. De Bruijn codes create a striped or multi-slit based color pattern where the position on the image is determined by the color of a stripe (or slit) and the color of surrounding fringes. They present a trade-off between the resolution of the system and the window property (related with the separation between similar colors). The matching process must take the repeatability of the sequence into account, in order to avoid errors caused by occlusions. Multi-slit patterns present higher accuracy than striped patterns, at expense of lower resolution. Algorithms using slit-based patterns were proposed by Monks et al. [6] and Salvi et al. [24] in one and two axes coding, respectively. Besides, Pages et al. [2] proposed a combination of striped pattern in Hue channel with a multi-slit pattern in Intensity channel (in the HSI space), obtaining the advantages of both approaches. Other techniques create the color pattern having different values in at least two channels for adjacent stripes, in order to increase the resistance to errors ([37] and [38]). Regarding M-arrays (perfect maps), some efficient techniques have been proposed [3], [27], [47]. In spite of being difficult to generate, M-arrays take advantage of coding both axes to include higher degree of redundancy. There are two ways to represent M-arrays in a pattern: using color dots or using an alphabet of symbols. As

stated in [46] and [2], features based implementations are more robust against colored objects and can be segmented more easily. The coded points can be easily found in grid techniques using tracking algorithms, as they are placed in the intersections of edges. Besides, time multiplexing was the first paradigm of coded structured light used to obtain 3D data from an unknown surface. Having an easy implementation, time multiplexing methods achieve higher spatial resolution and accuracy than spatial multiplexing (in general terms), with the constraint of having to project several patterns, not being able for moving scenarios. Binary codes and n-ary codes have been proposed within this group. N-ary codes have a greater alphabet, so the number of patterns to project is reduced compared to binary codes. Moreover, the distance between pixels having different codewords decreases. However, this does not imply a higher sensitivity to noise for the alphabet size used in practice, as can be observed in the work of Caspi et al. [14]. Hybrid techniques combining spatial and temporal information have also been proposed [4]. Shifting approaches are grouped within time multiplexing, as more than one pattern are necessary to create the codeword. In sparse coding, a discrete pattern is projected and shifted over time, until all the object is covered. These are the cases of Zhang et al. [26] Sansoni et al. [49] and Guhring [8], who projected and shifted a De Bruijn smoothed pattern, a stripe-based pattern, and a slit-based pattern respectively, obtaining good accuracy results and dense reconstructions.

Dense reconstruction coding strategies achieve density by creating a pattern that changes the value between adjacent pixels. The depth of a given point on the image is determined by the deviation of its gray or color value with respect to the projected pattern. Among them, phase shifting techniques use the same principle mentioned above, but considering the information hold in the phase. Using Ronchi or sinusoidal patterns, phase deviation provides information of the surface shape when compared with the ideal phase map. These methods show good resistance to ambient light and to reflection variation, due to the greyscale nature of the projected patterns. For time multiplexing techniques is necessary to project at least three shifted patterns in order to suppress the effect of the albedo and to recover the phase deviation, in the so-called phase shifting approaches [52], [51]. Combinations of shifted patterns into one single composite pattern have been proposed to achieve real time ([53], [5]). However, periodicity of the pattern imposes the assumption of smooth surfaces, as the presence of slopes would yield to some reconstruction errors. This problem is overcome in Multiple Phase Shifting approaches, which create an absolute phase map from two relative phase maps, according to the ideas proposed by [54]. Pribanic et al. [9] used this principle to create dense reconstruction of surfaces having slopes. For frequency multiplexing

2. STATE OF THE ART ON CODED STRUCTURED LIGHT

techniques, phase decoding is performed in the frequency domain rather than spatial domain. Fourier methods have been traditionally used. From the first proposal of Takeda and Mutoh [56], an evolution using two projected patterns instead of one was proposed by Su et al. [10] in order to suppress the effect of background illumination in the phase extraction. This principle was preserved in Yue [60] and Chen et al. [59] approaches, who combined the two patterns in one single projection multiplexing them in the orthogonal axis or in different color channels, respectively. Other approaches used trapezoidal grating instead of sinusoidal grating [76], stating that the post processing becomes faster as there is no need to compute the arctangent function. However, errors arise due to defocussing, which does not affect the sinusoidal patterns. Finally grading methods, which belong to spatial coding methods, project one single pattern where the codeword is hold only in every projected point for that pixel. They achieve high spatial resolution with few projected patterns [11], [30]. However, sensitivity to errors is extremely high in these methods, as the received pattern can be affected by the resolution of the projector and the sensor device, or the reflectivity of the surface. The experimental results show that the best results are obtained by the time phase shifting techniques([9]). They obtain dense reconstruction and good accuracy results. However, they are only able to work in static scenarios. Among one-shot techniques, DeBruijn based algorithms ([6], [2]) achieve the lowest deviation error in quantitative results. However, only sparse reconstruction is achieved. Dense reconstruction in one-shot techniques is achieved by the frequency-based analysis methods ([56], [60], [59]). They obtain good results if smooth surfaces are scanned, but fail under presence of big slopes and discontinuities.

Summarizing the main contributions done in structured light in the last years, it is important to mention that most of the works have been concerned into dense reconstruction by means of frequency multiplexing approaches, trying to increase the robustness in the decoding step and the resistance to slopes under the constraint of moving scenarios ([60], [59], [68], [70]). Time multiplexing in phase shifting has arise also to overcome the problem of slopes in the objects [9]. However, they are no longer able to work under moving scenarios. Furthermore, hybrid approaches have also experienced a big growth ([37], [38], [22], [27], [4]), due to their ability to merge different characteristics of previous method into one pattern or set of patterns. It would be interesting to develop a SL technique that provides dense reconstruction for one-shot projection. Using the principles of frequency multiplexing, an hybrid approach would eventually avoid the complexity and errors caused by the classical phase unwrapping algorithms required to recover the phase deviation.

3

First approach to one-shot dense reconstruction

In this chapter we present a first approach to one-shot dense reconstruction based on frequency fringe analysis. To achieve this, most of the works present in the literature are based on the projection of a single one-shot fringe pattern. Depth is computed using frequency analysis, extracting the phase deviation of the imaged pattern with respect to the projected pattern. However, the algorithms employed to unwrap the recovered phase are computationally slow and can fail in the presence of depth discontinuities and occlusions. The proposed approach employs color multiplexing and wavelet analysis to create the pattern, combined with a novel phase unwrapping algorithm. This enables the acquisition of dense 3D cloud of points and absolute coding. The advantages and disadvantages associated to the use of WT for frequency analysis are discussed at the end of the chapter.

3.1 Introduction

As discussed in chapter 2, most of the works presented ultimately concern the achievement of dense reconstruction for moving scenarios. However, the ability to work in real time conditions regardless object motion (up to the acquisition time required by the camera) is only achieved by one-shot projection techniques. Moreover, absolute coding represents a must for most of the applications mentioned above. Two main fields of research are opened, regarding the use of frequency analysis approaches or the combination of well known spatial multiplexing techniques. Different techniques using De Bruijn codes and M-arrays have been developed [24], [37], [3], [25], obtaining

3. FIRST APPROACH TO ONE-SHOT DENSE RECONSTRUCTION

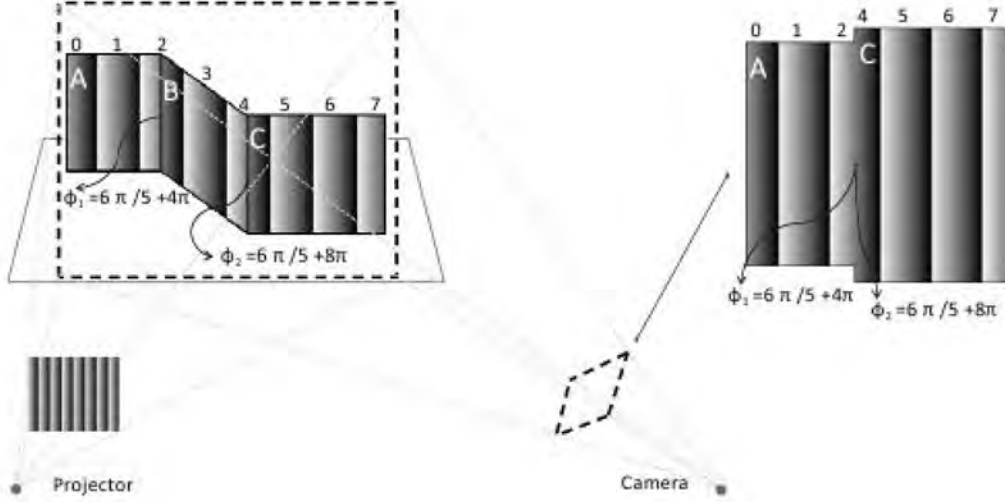


Figure 3.1: Surface slope producing a decoding error in traditional fringe pattern methods

a sparse acquisition with absolute coding and good accuracy results. Regarding the techniques using one-shot projections to obtain dense acquisition and absolute coding, grading techniques like the ones proposed by Carrihill and Hummel [11] and Tajima and Iwakawa [30] were proposed time ago. However, as stated by Salvi et al. [29], both techniques suffer of small signal to noise ratio and low accuracy. The other big group is constituted by the techniques based on frequency analysis (the so-called fringe profilometry methods). Although they achieve one-shot dense reconstructions, most of the algorithms require an unwrapping step to correctly unwrap the phase and extract the object depth. Noise, local shadows, under-sampling, fringe discontinuities and irregular surface brightness make the unwrapping procedure much more difficult than simply adding the corresponding multiple of 2π when a discontinuity appears [29]. Moreover, the phase unwrapping algorithm that can be found in the literature usually require a high computational cost and can fail into errors when the surface present a slope like the one shown in Fig. 3.1. In this case, the periodicity of the projected pattern make this overlapping impossible to detect regarding the recovered image. Therefore, absolute coding is not feasible using a unique frequency fringe pattern.

In order to obtain better performance, more information must be embedded in the one-shot pattern leading to a multiplexing of patterns in frequency or color space. The algorithm proposed in this chapter utilizes the ideas of dense acquisition combined with a novel method for color multiplexing and phase unwrapping in order to obtain a

one-shot dense acquisition having absolute coding.

The chapter is structured as follows. Section 3.2 presents the idea of absolute coding unwrapping. Also, the design of the technique is described. The experimental results with both simulated and real data are presented in section 3.3. Finally, section 3.4 concludes with a discussion of the proposed method, where the main advantages and disadvantages of the proposal are pointed out.

3.2 System proposal

The novel method we propose to overcome this problem is based on the remainder theorem [54] (an explanation about the use of the Remainder Theorem in Multiple Phase Shifting pattern projection can be found in Appendix C). A diagram of showing the different steps of the proposed algorithm can be seen in Fig. 3.2.

Using the formulas provided by the remainder theorem, an absolute phase map can be computed from two different relative phase maps having frequencies that are relative prime numbers between them. Having two relative phase maps with different frequencies and their corresponding phase values ϕ_1, ϕ_2 , the absolute phase value is given by eq. (3.1):

$$\Phi_{ABS1,INT} = \sum_{i=1}^N \phi_{Ri,INT} e_i \text{mod}(\lambda_1 \lambda_2) \quad (3.1)$$

being λ_i the period wavelengths and e_i a number which divided by λ_i yields a remainder 1, and 0 otherwise. A solution to $\Phi_{ABS1,INT}$ can be obtained from eq. 3.1, providing an absolute phase map from a minimum of two relative phases. Another advantage of this technique relies on its simplicity and non dependence on the neighboring pixels, as the phase value is computed directly from a linear combination of the two relative phase map values for the given pixel. The proposed model employs one-shot color multiplexing, wavelet analysis and absolute coding. As mentioned in chapter 2, there are different frequency analysis techniques. Wavelet analysis was chosen among FT, SPD and WFT because it provides frequency localization adaptive to the carrier frequency used. Therefore the patch size is set inversely proportional to the frequency to analyze, minimizing the frequency overlapping. Color multiplexing is used to combine different fringe patterns in a unique one-shot color projection.

3. FIRST APPROACH TO ONE-SHOT DENSE RECONSTRUCTION

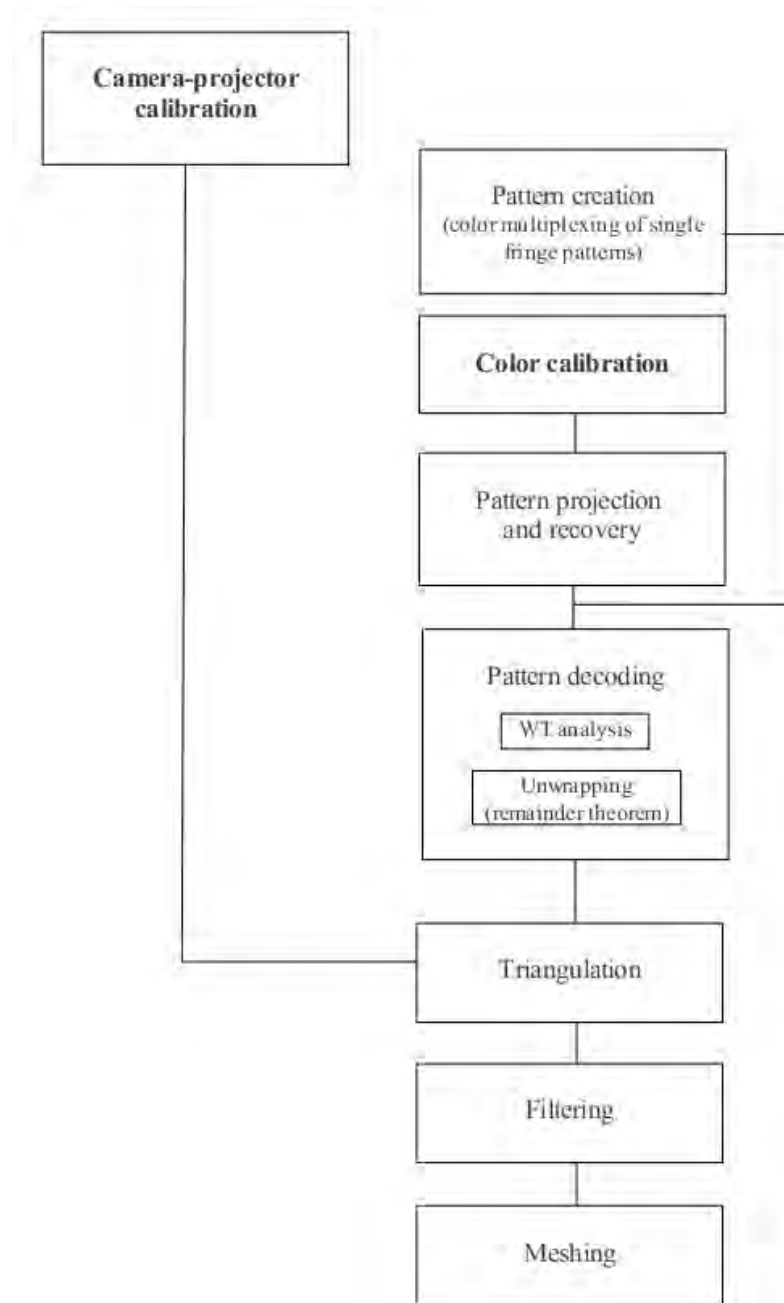


Figure 3.2: Diagram of the proposed SL algorithm using WT and the remainder theorem.

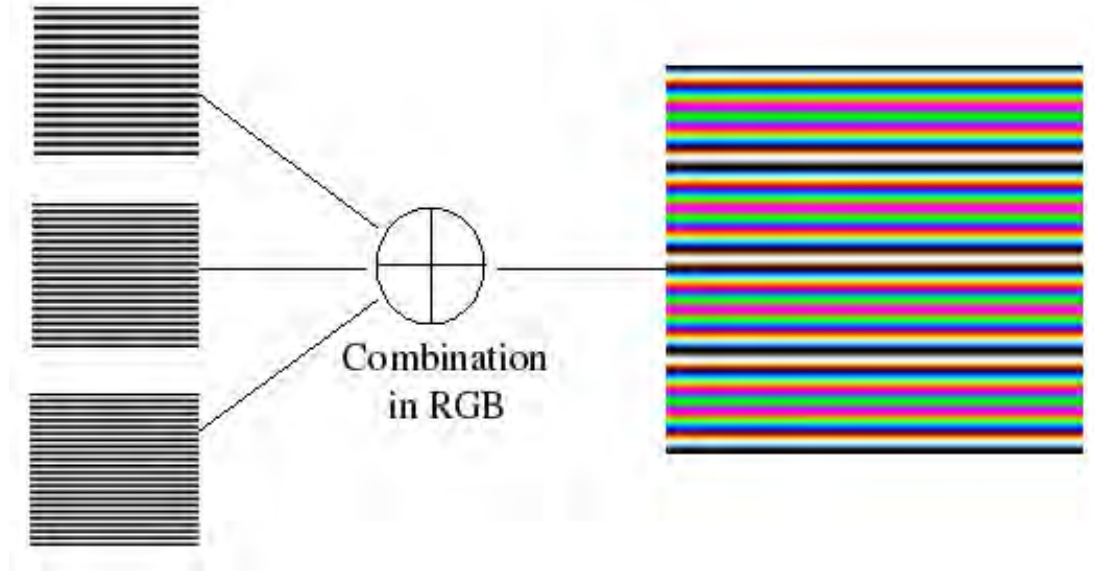


Figure 3.3: Combination of Red, Green and Blue channels to create the color multiplexed pattern.

3.2.1 Pattern coding

The idea of the proposed algorithm is to multiplex three different fringe patterns in the color space (Fig. 3.3), taking advantage of the Red, Green and Blue separated channels of the projector and camera devices. Compared to the frequency multiplexing approach, this technique permits the use of all the frequency bandwidth in every channel, avoiding errors caused by interferences of the harmonic frequencies. The fringe patterns are created having frequencies that are relative prime numbers between them. The sinusoidal patterns use one axis coding, as in other WT approaches. The projected pattern is represented by eq. (3.2), where A_i^p , B_i^p and f_i^p represent the low-frequency (DC) and high-frequency (AC) modulus and the frequency values for every channel of the projected pattern (r , g and b):

$$\begin{aligned}
 I^p(y^p) = & A_r^p + B_r^p \cos(2\pi f_r y^p) + \\
 & A_g^p + B_g^p \cos(2\pi f_g y^p) + \\
 & A_b^p + B_b^p \cos(2\pi f_b y^p)
 \end{aligned} \tag{3.2}$$

3. FIRST APPROACH TO ONE-SHOT DENSE RECONSTRUCTION

3.2.2 Pattern decoding

Once projected onto the object surface and imaged by the camera, the received pattern (Fig.3.4) can be represented as:

$$\begin{aligned}
 I_n(x, y) = & \\
 & \delta(x, y) \cdot (\alpha(x, y) \cdot (A_r^p + B_r^p \cos(2\pi f_r y^p + \phi(x, y))) + \\
 & \beta(x, y) \cdot (A_g^p + B_g^p \cos(2\pi f_g y^p + \phi(x, y))) + \\
 & \gamma(x, y) \cdot (A_b^p + B_b^p \cos(2\pi f_b y^p + \phi(x, y)))) \quad (3.3)
 \end{aligned}$$

where $\delta(x, y)$ represent the different albedo and $\alpha(x, y)$, $\beta(x, y)$ and $\gamma(x, y)$ the effect of crosstalk between the different color channels. The first task is to split the three color channels obtained from the camera and perform a color enhancement to reduce the effect of albedo and crosstalk in every color channel. To cope with this, the color calibration explained in appendix B is applied to the input image. The matrix of eq.(B.1) represents the whole system (projector-camera) and aims to subtract the effect of crosstalk between color channels. This matrix is applied to every color channel (R, G and B) of the recovered image, obtaining the corresponding fringe pattern.

3.2.3 Extraction of the wrapped phase

The applied color calibration provides the color filtered red, green and blue channels. However, as it approximates the system as a linear transformation between projected and received images, some errors will persist due to non-linearities. This error, jointly with the different albedo and noise, must be filtered by the wavelet analysis algorithm. The wavelet analysis employs a 2D Continuous Wavelet Transformation (2D-CWT) using a Morlet mother wavelet (eq.(3.2.3)).

$$\begin{aligned}
 WT_1(a, \theta, b_1, b_2) = & \frac{1}{a} \iint \{A(x, y) + \frac{B(x, y)}{2} [e^{-i\phi(x, y)} + e^{i\phi(x, y)}]\} \cdot \\
 & \exp(2\pi i f_c \frac{(x-b_1)(\cos(\theta)+\sin(\theta))+(y-b_2)(\cos(\theta)-\sin(\theta))}{a}) \cdot \exp(-\frac{(x-b_1)^2+(y-b_2)^2}{a^2 f_b}) dx dy
 \end{aligned}$$

where a, θ, b_1, b_2 are the scale, rotation and translation parameters, respectively. As stated in [69], Morlet wavelet is optimal in case we deal with signal having a low Signal-To-Noise ratio, which is the case when working in real conditions. Furthermore, 2D wavelet analysis performs better than 1D wavelet analysis. The output of the 2D wavelet analysis is a 5D matrix of dimensions *height · width · scales · translations ·*

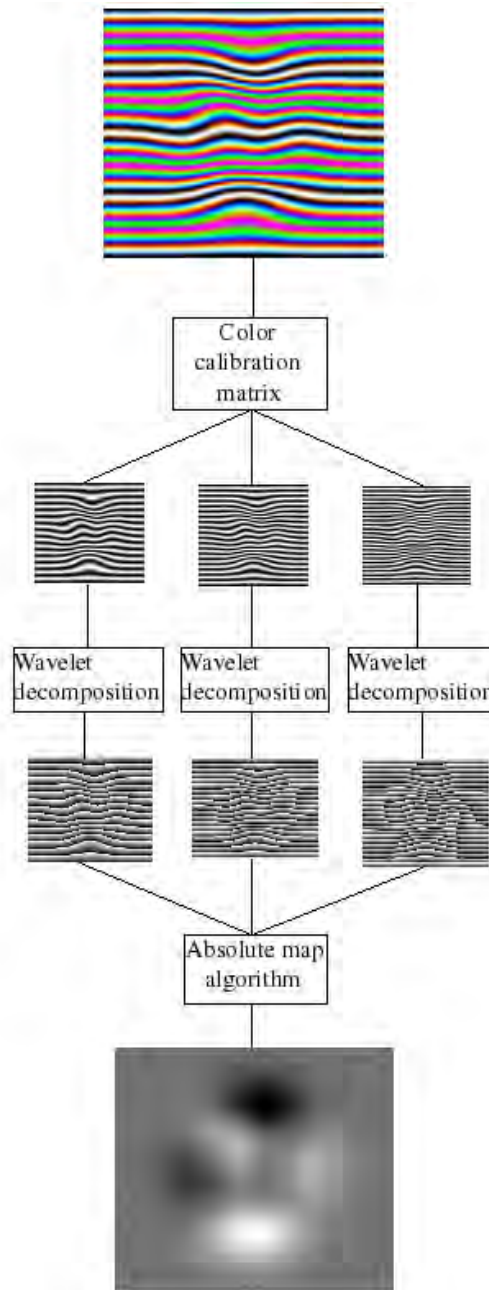


Figure 3.4: Pattern decoding (image processing).

3. FIRST APPROACH TO ONE-SHOT DENSE RECONSTRUCTION

orientations.

After the computation of the 2D-CWT, a phase from ridges (WFR) algorithm combined with cost function is employed to extract the optimal value of a, θ, b_1, b_2 for every position at every color channel. As mentioned in chapter 2, the WFR computes the maximum value of the modulus for every pixel. For a particular value of translation across both x and y axes, the transform will reach its maximum value when the dilated and rotated 2D wavelet and the fringe pattern are locally most similar. That is, the modulus of the transform has a maximum value when the 2D wavelet frequency is very close to the fringe frequency and the rotation of the 2D wavelet is very close to the direction of the fringes. This produces a ridge in the transform. The phase value corresponding to the position of the ridge is selected.

3.2.4 The cost function

After applying the WT without the cost function, the WFR algorithm would select for every pixel the daughter wavelet having the maximum of the modulus (within all scales, translations and orientations available) and its corresponding angle for that position. This would produce errors caused by local inaccuracies of the estimated pixel phase. The cost function is introduced to ensure continuity and avoid errors due to local errors that can be identified and corrected regarding its neighbors. The cost function works along the y axis of the camera (though any other direction could be selected according to the direction of fringes in the projected pattern), selecting the combination of daughter wavelet that best performs in terms of modulus maxims and continuity. The cost function is presented in eq.(3.4):

$$Cost = \sum_{b=2}^W \{-|S(\phi(b), b)|^2 + |\phi(b) - \phi(b-1)|^2\} \quad (3.4)$$

where $\phi(b)$ represents any value of the scaling parameter, b is the shifting parameter in the coding axis (y axis in our case), $|S(\phi(b), b)|$ is the modulus value of $S(\cdot)$ at both $\phi(b)$ and b , and W is the total width of the fringe pattern, in pixels. For every column, the algorithm works at follows: Eq.(3.4) calculates the cost of the step variation. As this algorithm searches through many local maxims, thus many different potential paths will occur. However, an optimal path should be selected and considered to be the true ridge of the transform. If we assume that the optimal path goes through the point (p, j) , i.e., $(j) = p$, then the cost can be reformatted as in eq.(3.5):

$$\begin{aligned}
 Cost = & \sum_{b=2}^{j-1} \{-|S[\phi(b), b]|^2 + |\phi(b) - \phi(b-1)|^2\} + \\
 & (-|S[p, j]|^2 + |p - \phi(j-1)|^2) + \\
 & \sum_{b=j+1}^W \{-|S[\phi(b), b]|^2 + |\phi(b) - \phi(b-1)|^2\}
 \end{aligned} \tag{3.5}$$

From eq.(3.5), we can conclude that at a given point (p, j) , the optimization of the complete path can be divided into two parts: the optimization for the path $(b = 1)$ to $(b = j)$ and the optimization for the path $(b = j)$ to $(b = W)$. Some simulated results of the use of the cost function are shown in Fig. 3.5. As can be seen, errors in the WFR due to local inaccuracies in the WT result are suppressed by the cost function algorithm.

```

for orientation = 1 to N do
    compute 2D daughter wavelet map (height · width · scales · translations)
end for
for column = 1 to width do
    for orientation = 1 to N do
        Compute best path in terms of:
        Maxima of the modulus, contributes decreasing the cost.
        change in scale, contributes increasing the cost.
    end for
    Select path with minimum cost
end for
Select orientation with minimum cost

```

This algorithm reduces the errors in presence of local holes in the received fringe pattern. Once the appropriate daughter wavelets have been selected, their corresponding angle values are computed and the wrapped phase map is extracted.

3.2.5 Combination of the wrapped phased

The next step is to unwrap the phase in order to extract the phase deviation. This is done applying the unwrapping algorithm of absolute coding. A minimum of two patterns are required to this end, but the proposed method utilizes the three color channels

3. FIRST APPROACH TO ONE-SHOT DENSE RECONSTRUCTION

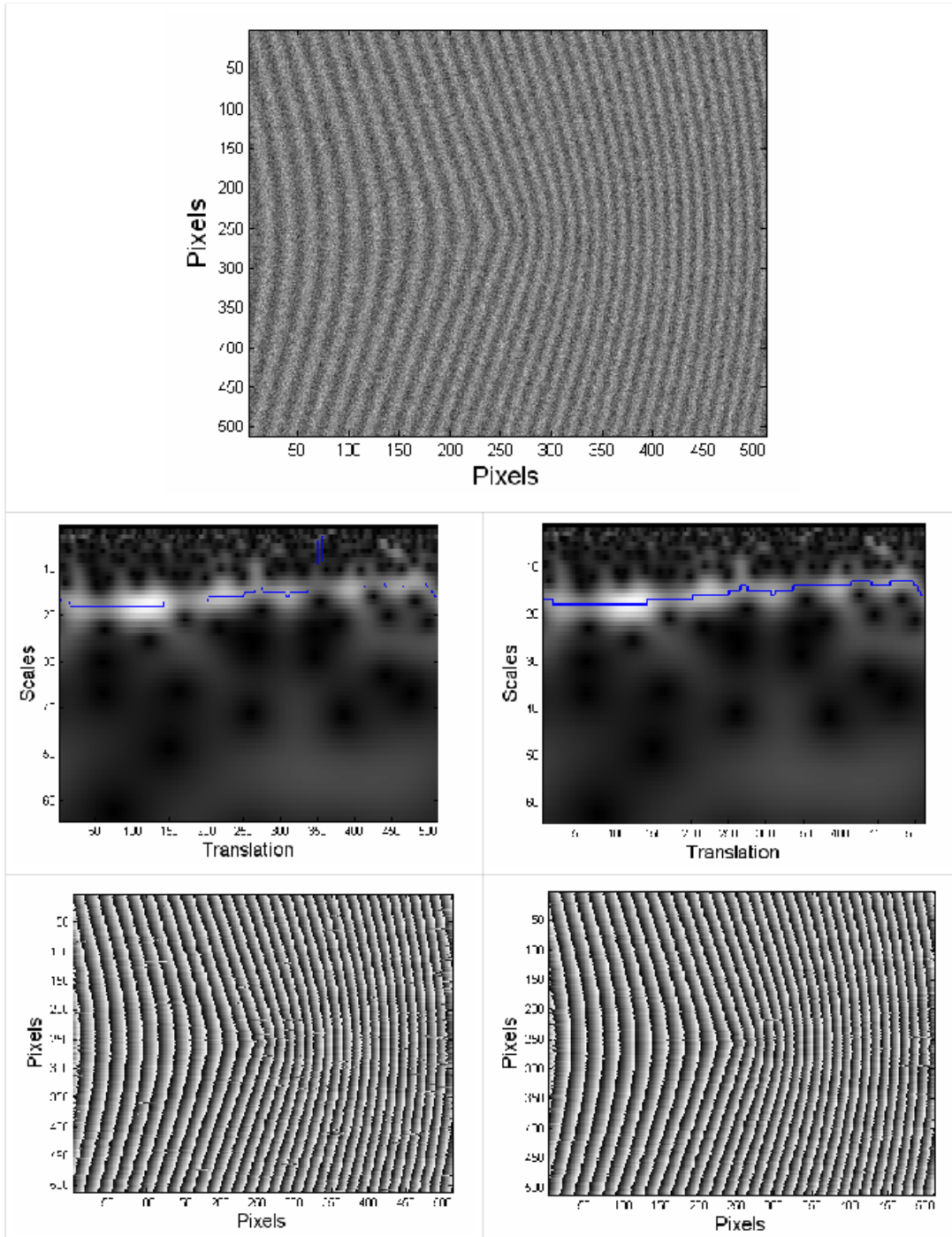


Figure 3.5: On the top, simulated noised fringe pattern. On the left column, example of the WT and ridge (in blue) for a specific position in the input image (top), and the corresponding extracted phase (bottom). On the right column, similar results using the cost function.

(red, green and blue) to create the absolute phase map. This allows us to increase the redundancy and hence reduce the errors that may propagate to the absolute phase map. The algorithm combines every two of the color channels to create an absolute phase map. That is, a total of $\binom{n}{2}$ combinations are created, being $n = 3$ the number of channels. An absolute phase map is computed for every one of the combinations following the idea of relative phase numbers. Posteriorly, an optimization algorithm is pursued to extract the optimal phase map that minimizes the error. This process is done in order to reduce the noise created by the non linearities of the projector-camera pair in color or intensity representation, which propagates from the wrapped phases to the absolute phase map. The optimization algorithm computed the best absolute map in terms of minimization of the laplacian map, and works as follows:

```

Compute the laplacian matrix of every absolute phase matrix
InitialLaplacianMatrix = min(sum(laplacian matrices))
FinalAbsoluteMatrix = Abs. matrix corresponding to InitialLaplacianMatrix
Average = mean(InitialLaplacianMatrix)
for row = 1 to height do
  for column = 1 to width do
    if Initial-laplacian-matrix(row,column) > Average then
      FinalAbsoluteMatrix(row,column) = argmin(InitialLaplacianMatrix), within
      all absolute phase matrices pixels
      Recompute InitialLaplacianMatrix
      Recompute Average
    end if
  end for
end for

```

This algorithm suppresses the error provided by isolated pixels.

3.2.6 3D or depth extraction

There are two ways to extract the 3D cloud of points from the information provided by the deviated phase map. The first way is to create a correspondences map between the projected and the recovered phases. This gives us, for every pixel in the recovered phase map, its corresponding interpolated position in the projected pattern. Every pair of correspondences is then triangulated using the information provided by the geometric calibration matrix, giving us the 3D points in the world coordinates system. Another

3. FIRST APPROACH TO ONE-SHOT DENSE RECONSTRUCTION

approach, profusely used in fringe profilometry, uses an approximated reconstruction of the depth (depth reconstruction). The formula was first proposed by Takeda and Mutoh [56]:

$$h(x, y) = L \cdot \frac{\Delta\phi(x, y)}{(\Delta\phi(x, y) - 2\pi f_0 d)} \quad (3.6)$$

where L is the distance to the reference plane and d is the distance between the camera and the projector devices. This was the approach used in this chapter, as it is enough to show us the performance of the method in a fast way under presence of noise and discontinuities.

3.3 Implementation and results

The proposed algorithm has been implemented and tested in both simulated and real environments and compared to other one-shot techniques. The setup used for the tests was composed of an LCD video projector (Epson EMP-400W) with a resolution of 1024×768 pixels, a camera (Sony 3CCD) and a frame grabber (Matrox Meteor-II) digitizing images at 768×576 pixels with 3×8 bits per pixel (RGB). The baseline between camera and projector was about $0.5m$. The algorithm run on an Intel Core2 Duo CPU at 3.00GHz. The selected frequencies for the three fringe pattern were $p_1 = 15$, $p_2 = 19$ and $p_3 = 23$ periods, providing a good resolution in details, while preserving the sinusoidal shape once captured by the camera.

3.3.1 Simulated results

The proposed algorithm was tested using simulated data. The *peaks* function available in Matlab (shown in Fig. 3.6) has become a benchmark for fringe pattern analysis, as stated in [29]. Moreover, the simulated object shape was obtained for different values of noise. The error introduced is a Gaussian zero mean random noise in the range 5%, 10% and 20% of the total dynamic range of the input image. The resulting patterns used as input images are shown in Fig. 3.7.

The reconstructed object shape obtained using the input image of noise 5% of the dynamic range is shown in Fig. 3.8. As can be observed, the algorithm reconstructs the simulated object at the same time the noise existing in the input image is reduced. A scaled map of the error is also presented in Fig. 3.8, where the error has been re-scaled

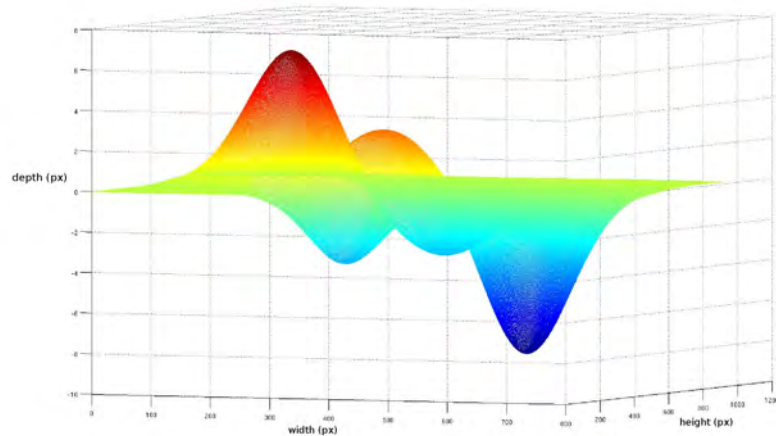


Figure 3.6: Simulated surface

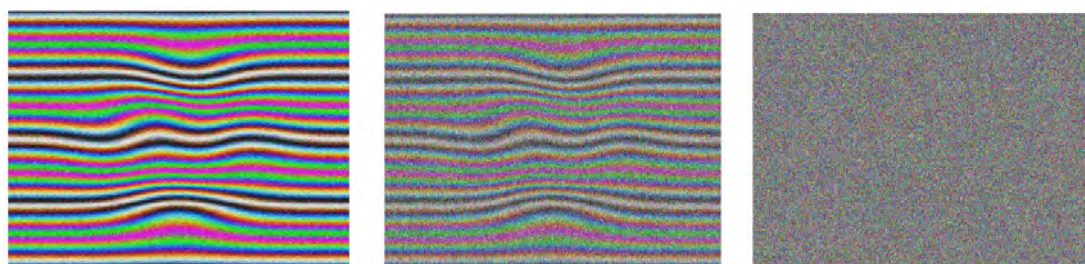


Figure 3.7: Imaged patterns of Fig. 3.6 with added noise values of 5%, 10% and 20% of the maximum dynamic range, respectively.

3. FIRST APPROACH TO ONE-SHOT DENSE RECONSTRUCTION

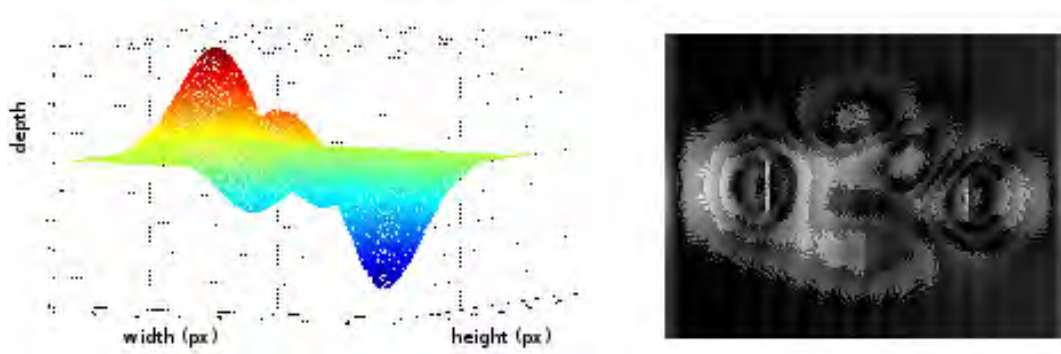


Figure 3.8: Reconstructed surface of Fig. 3.6 with a noise of 5%, and corresponding error map re-scaled from $[0\%, 3.48\%]$ to $[0, 255]$.

from $[0\%, 3.48\%]$ to $[0, 255]$. The error is uniformly distributed, with some peaks in the regions of major inclinations of the *peaks* function. This is due to the fact that a faster change in the phase frequency is more likely to suffer detection errors when noise appears. The algorithm was also tested with the other noised input images, and the results of average error are presented in table 3.1. The error is highly reduced for values lower than a noise of 5% of the dynamic range. Besides, noise values higher than 20% of dynamic range makes the decoding impossible. It is important to note that noise introduced depends on the object depth. For the object depths analyzed in real applications the noise remains under 5% of the data dynamic range, and hence the method is able to filter the noise in the reconstructed shape, as will be observed in the experimental results.

Table 3.1: Error rates for the given input noise going from noises of 5% to 20% of the data dynamic range.

Noise percentage (%)	Error rate (%)
5	0.57
10	19,8
20	—

Finally, the algorithm was tested in the acquisition of the stair step of Fig.3.1. The performance of the absolute coding used in the proposed algorithm is shown in Fig.3.9. The slope has been detected despite it was not visually perceptible. However, some error arose in the vicinity of the discontinuity, due to the erroneous phase estimation in the surroundings of the slope. This problem is posteriorly analyzed.

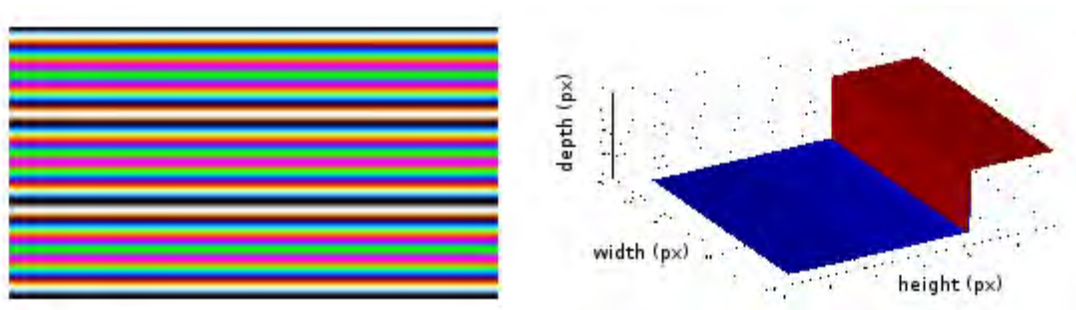


Figure 3.9: Projected pattern on the object of Fig.3.1 and reconstructed surface.

3.3.2 Experimental results

The proposed technique was tested under real conditions. First, a flat plane was reconstructed to perform a quantitative comparison with other one-shot absolute coding techniques existing in the literature. The results of reconstructing a 576×478 pixels region are shown in table A.2. The first method tested was the algorithm proposed by Carrihill and Hummel [11], which provides also dense acquisition by projecting a greyscale grading pattern. The second technique is the DeBruijn-based pattern proposed by Pages et al. [2], which employs a DeBruijn pattern and it is considered one of the most accurate techniques in sparse one-shot absolute coding patterns. The proposed method filters possible effects of noise on the image, and the flat plane is reconstructed with absolutely no error. This is due to the Morlet mother wavelet used to extract the phase (which is optimal in case of having low signal to noise ratio) and to the cost function, which corrects any erroneous estimated phase. It is important to mention, however, that the major problems of the proposed method arise in presence of discontinuities.

Table 3.2: Quantitative results reconstructing a flat plane. The headings are: author’s name of the technique; average deviation of the reconstructing error; standard deviation of the reconstructing error; number of 3D points reconstructed.

Technique	Average (mm)	Stdev (mm)	3D Points
Carr.and Hummel	11.9	5.02	202714
Pages et al.	1.31	1.19	13899
Proposed method	0.22	0.12	275328 (full)

In order to test the algorithm with real non-flat objects, a smooth volume done with sheets of paper having different orientations (Fig. 3.10), and a ceramic face (Fig.3.12) were reconstructed. These two objects attempt to cover the usual requirements of the

3. FIRST APPROACH TO ONE-SHOT DENSE RECONSTRUCTION

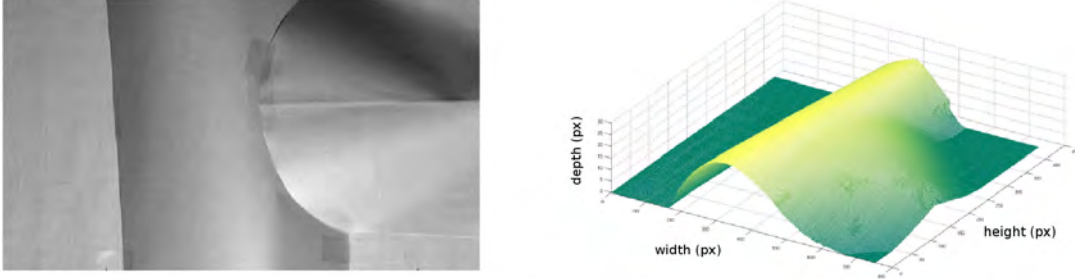


Figure 3.10: Original and reconstructed sheets of paper.

methods regarding its applicability to 3D dense acquisition. It is important to note that these objects have been chosen having lambertian white surface, in order fit into a linear color calibration. The decoding steps are presented in Fig. 3.11 and Fig.3.3.2.

The sheets of paper are reconstructed without error, not depending on the orientation and shape of the object. For the ceramic face, the main volume has been detected and details like the nose and eyes (identified as shadows) are represented in the final surface. The acquisition presents, however, some errors in the discontinuities (see Fig.3.12), the same kind of error that revealed in the simulated results. This is caused by the wavelet-based phase extraction algorithm and was reported in the work of Abid [69] as the major drawback of the wavelet technique employed in this algorithm, which is up to some point common in all frequency based analysis. An erroneous phase estimation at the vicinity of the discontinuity propagates to the absolute phase map, yielding to big errors and holes in the reconstruction.

3.4 Conclusions

Continuous coding strategies achieve dense acquisition by creating a pattern that changes the value between adjacent pixels. The depth of a given point on the image is determined by the deviation of its grey value with respect to the projected pattern. Among them, combinations of shifted patterns into one single composite pattern have been proposed to achieve dense reconstruction with a unique projection ([53],[5]). However, periodicity of the pattern imposes the assumption of smooth surfaces, as the presence of slopes would yield to some acquisition errors. This problem is overcome in Multiple Phase Shifting approaches, which create an absolute phase map from two relative phase maps, according to the ideas proposed by [54]. In this chapter, we have presented a one-shot dense acquisition technique using an absolute coding phase unwrapping algorithm. In order to combine three patterns in one single shot, the algo-

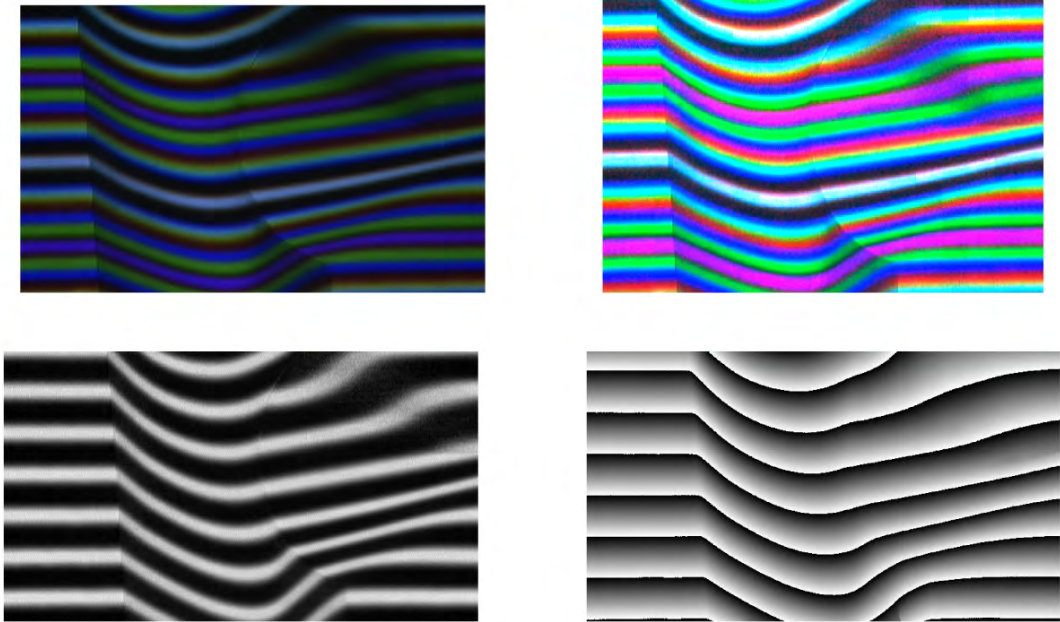


Figure 3.11: Sheets of paper of Fig. 3.10: captured image (top-left), enhanced image (top-right), one color channel (bottom-left) and its wrapped phase (bottom-right).

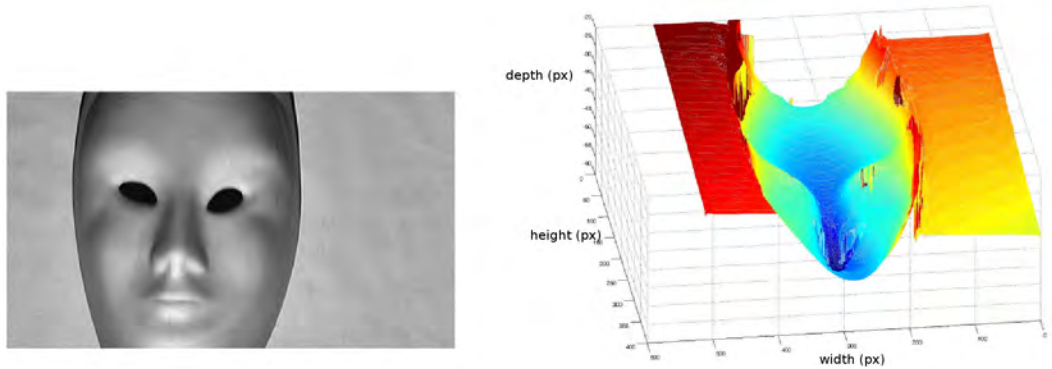


Figure 3.12: Original and reconstructed surface of a ceramic face (the 3D shape is inverted to appreciate the details of nose and eyes).

3. FIRST APPROACH TO ONE-SHOT DENSE RECONSTRUCTION

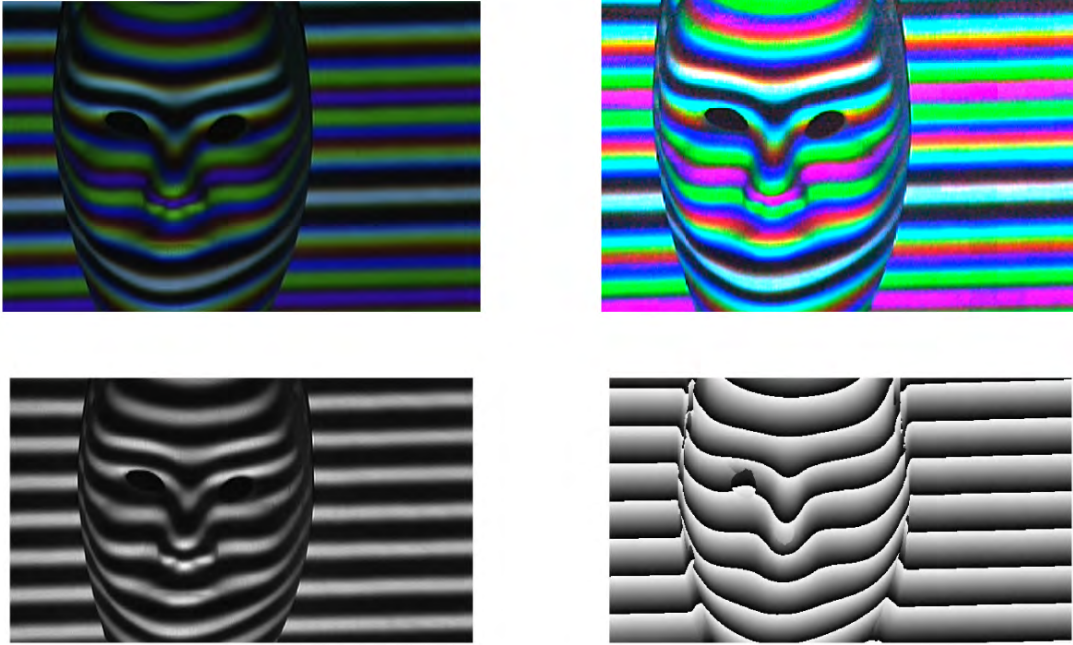


Figure 3.13: Ceramic face of Fig. 3.12: captured image (top-left), enhanced image (top-right), one color channel (bottom-left) and its wrapped phase (bottom-right).

rithm performs a multiplexing in color space of the three different fringe patterns. The phase of these patterns is extracted using wavelet decomposition combined with a cost function algorithm. An absolute coding based algorithm merges the individual relative phase maps to create the absolute phase map and extract the depth deviation. Finally, a consistence mapping is pursued in order to avoid errors created by a bad detection of slopes and error propagation between pixels during the unwrapping step. The algorithm has been tested in both simulated and real data, and a quantitative comparison with other one-shot absolute coding acquisition techniques had been pursued, showing the effectiveness of the proposed technique in terms of noise reduction. Moreover, the simulations performed optimally using the standard benchmark without and with added noise. Finally, a set of real experiments was pursued: a smooth volume done with sheets of paper having different orientations, and a ceramic face. The 3D depth maps present good reconstruction results, showing that the method works optimally under real conditions. However, some problems were encountered with surfaces having discontinuities. This is partially due to the effect of the dyadic net intrinsic to the wavelet implementation, which does not consider little frequency variations when setting the window size. Therefore, could be beneficial to implement an algorithm for the

adaptive selection of the best analysis mother signal, in terms of frequency and shape, depending on the frequency components present in a given patch of the image. Having this, a proper frequency analysis can be pursued, minimizing the problems presented in the discontinuities.

3. FIRST APPROACH TO ONE-SHOT DENSE RECONSTRUCTION

4

Automatic window selection in Frequency Transform techniques

As shown in previous chapter, a proper selection of the frequency analysis used for fringe decoding reveals crucial for an optimal extraction of the phase deviation. Many errors can arise due an incorrect detection of the carrier frequency in a fringe pattern region, causing noised phase map extraction which originates 3D reconstruction errors. This problem is more accused under presence of slopes, as was observed in Fig.3.12 of previous proposal. The aim of this chapter is first to perform a comparative study of the most suitable frequency based techniques for fringe pattern analysis. Afterwards, a new frequency analysis proposal is done to accurately recover the phase deviation even under presence of slopes. This proposal is based on the automatic selection of the window width in WFT.

4.1 Introduction

As expose in chapter 2, there are five different techniques used traditionally for phase extraction: Phase Measurement Profilometry (PMP), Spatial Phase Detection (SPD), Fourier Transform (FT), Windowed Fourier Transform (WFT) and Wavelet Transform (WT). Among them, only those based on frequency analysis (FT, WFT and WT) project one single shot and thus are able to work with moving objects. Regarding these frequency-based techniques, the main differences among them are related to the section of the imaged pattern that is considered in the frequency analysis. FT performs a global analysis, which is appropriate for stationary signals with poor spatial localization. However, this is not the case in CSL, which is by nature limited in space and thus non-

4. AUTOMATIC WINDOW SELECTION IN FREQUENCY TRANSFORM TECHNIQUES

stationary. This fact led to the use of the other two frequency based transforms (WFT and WT), which analyze local information in the imaged pattern. WFT and WT are constituted by two main steps: windowing the imaged pattern in local patches and computing the transform at every local patch. The crucial point in these techniques relies on the necessity of selecting an optimal window size, which constitutes a trade-off between resolution in space and resolution in frequency. Whereas WFT uses a fixed window size for all frequency components, WT changes the window size inversely proportional to the carrier frequency by a factor of 2^j (dyadic net). This chapter proposes a new algorithm for the automatic selection of the window size, based on a modification of the traditional WFT. Moreover, as it is not possible to find in the literature a study of windowing signals assuring good results for fringe pattern analysis using WFT, the adaption of four well known mother wavelet (Morlet, Paul, Shannon and Spline) is applied to WFT in order to compare the proposed algorithm for WFT with the WT.

The chapter is structured as follows: Section 4.2 performs an analytical comparison of WFT versus WT techniques. Section 4.3 proposes a novel algorithm for the automatic selection of the window. Section 4.4 shows quantitative and qualitative results obtained with both simulated and real data for the four different mother wavelets. Finally, Section 4.5 states the conclusions, pointing out the suitability of the four mother wavelets employed in the tests and their performance in comparison to the traditional WT.

4.2 Comparative study between WT and WFT

In order to analyze the pros and cons of WFT and WT techniques applied to fringe pattern analysis, a theoretical and a practical comparison is required. The main difference between both techniques is the way the window size is set, depending on whether they have a fixed or a variable value. Regarding WFT, its definition is given by eq.(4.1):

$$Sf(u, v, \xi, \eta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \cdot g(x - u, y - v) \cdot \exp(-j\xi x - j\eta y) dx dy \quad (4.1)$$

being (x, y) , (ξ, η) the translation and frequency coordinates respectively, and $g(x, y)$ the windowing function. When $g(x, y)$ is a Gaussian window, the WFT is called a

4.2 Comparative study between WT and WFT

Gabor transform; that is:

$$g(x, y) = \frac{1}{\sqrt{\pi\sigma_x\sigma_y}} \cdot \exp\left(-\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2}\right) \quad (4.2)$$

where σ_x and σ_y are the standard deviations of the Gaussian function in x and y , respectively. Regarding WT, there are two main steps in the process. First part finds the optimal values for dilation and rotation to use in the wavelet function. Expressing the wavelet function by means of sinus and cosines, the Continuous Wavelet Transformation (CWT) of the received signal is given by eq. (4.3) (sine are represented by s and cosine by c):

$$WT_f(a, \theta, b_1, b_2) = \frac{1}{a} \iint I(x, y) \cdot \psi\left[\frac{(x - b_1c(\theta)) - (y - b_2s(\theta))}{a}, \frac{(x - b_1s(\theta)) + (y - b_2c(\theta))}{a}\right] dx dy \quad (4.3)$$

For instance for the 2D complex Morlet wavelet of eq. (4.4):

$$\psi(x, y) = \frac{1}{\sqrt{\pi f_b}} \cdot e^{2i\pi f_c(x, y)} \cdot e^{-\frac{x^2 + y^2}{f_b}} \quad (4.4)$$

the Wavelet Transform results:

$$\begin{aligned} WT_1(a, \theta, b_1, b_2) &= \frac{1}{a} \iint \{A(x, y) + \frac{B(x, y)}{2} [e^{-i\phi(x, y)} + e^{i\phi(x, y)}]\} \cdot \\ &\exp\left(2i\pi f_c \frac{(x - b_1)(\cos(\theta) + \sin(\theta)) + (y - b_2)(\cos(\theta) - \sin(\theta))}{a}\right) \cdot \exp\left(-\frac{(x - b_1)^2 + (y - b_2)^2}{a^2 f_b}\right) dx dy = \\ &W_1(a, \theta, b_1, b_2) + W_2(a, \theta, b_1, b_2) + W_3(a, \theta, b_1, b_2) \end{aligned}$$

If we express the phase in terms of its Taylor series, considering negligible any term higher than first derivative, the dilation factor and the rotation factor can be both computed through $\frac{|WT_1|^2}{a^2}$, that is:

$$a = 2\pi f_c \sqrt{\frac{2}{\phi_x'^2 + \phi_y'^2}} \quad (4.5)$$

$$\sin(2\theta) = \frac{\phi_x'^2 - \phi_y'^2}{\phi_x'^2 + \phi_y'^2} \quad (4.6)$$

From this, a discretization of a in order to work in discrete domain has to be done, choosing k so that $a = 2^k$ (dyadic net). K value is obtained from 4.7, being y the

4. AUTOMATIC WINDOW SELECTION IN FREQUENCY TRANSFORM TECHNIQUES

codification axis.

$$k(y) = 0.5 + \log_2(f_c/f_0) - \log_2\left(\frac{\phi'_x}{2\pi f_0}\right) \quad (4.7)$$

Once k has been calculated, 2D DWT is applied to the input image, obtaining the low-pass and high-pass decomposition for the corresponding frequency level. This is done sequentially for all the frequency levels until a frequency decomposition among the dyadic net is obtained.

As stated in [79], WT performs better with signals having a wide range of frequencies with shorter correlation times for the higher frequencies than for the lower frequencies. This is the case in natural scenes, where low-frequency components usually last for longer durations than high-frequency components. However, in fringe patterns their periodicity and spatial extension does not depend on the selected frequency. Nevertheless, they mostly present spatial-harmonic components around the selected frequency. This is the reason why, despite many authors claim the goodness of WT [69], [68], there are some recent works that state the best suitability of WFT [80], [67]. Another point to consider is the resistance to noise. It has been demonstrated [80] that for noiseless fringe patterns the frequency components can be accurately recovered in either small or large windows, regardless the frequency value. However, under presence of higher noise on the imaged fringe pattern, an optimal selection of the window size reveals crucial for filtering the noise while preserving the main frequency components. Under these circumstances, the fixed window size of WFT performs better than the variable window size of WT. This is mainly due to the dyadic net used in practical applications of WT. This net changes geometrically (by two) the window size for adjacent levels of dilation, being excessive for some applications where the main frequency stands close to a fixed value (like in fringe pattern analysis).

Another point to consider is the importance of selecting a window having good localization in both frequency and space, in order to perform an optimal analysis of the fringe pattern. In WT, the mother wavelet signals usually used in fringe pattern analysis are, among others, the Morlet wavelet, the Paul wavelet, the Shannon wavelet and Spline wavelet [81], [82], [83], [84]. All of them use a low-pass envelope signal modulating a frequency sinusoidal signal, thus presenting good localization in time and frequency. In WFT, the Gabor transform has been traditionally used in fringe pattern analysis, as it provides the smallest Heisenberg box [85], [86]. However, is not possible to find in the literature a study of windowing signals assuring good results for fringe pattern analysis using WFT. As this work has been already done in WT [69], it is

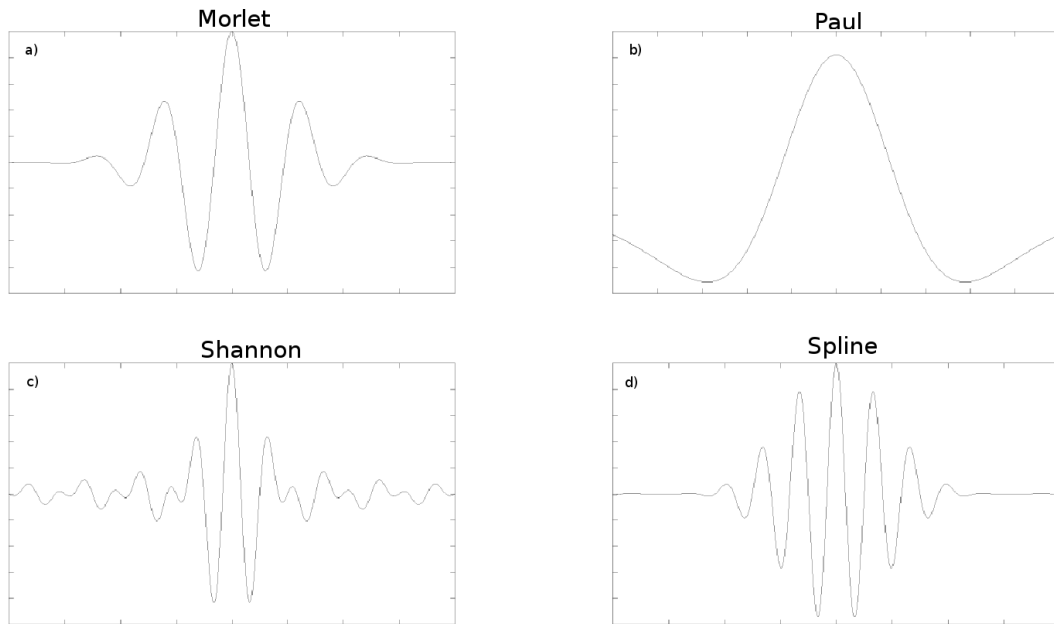


Figure 4.1: Diagram of Morlet (*a*), Paul (*b*) Shannon (*c*) and Spline (*d*) wavelets, traditionally used in fringe pattern analysis.

recommended to adapt those optimal mother wavelets to WFT. This would permit a fair comparison between both techniques.

4.2.1 Adaption of the wavelet signals to the use in WFT

The suitability of some mother wavelets for fringe pattern analysis in WT has been outlined in [68], [69], [71]. The best situation is given when the signal presents good localization in both space and frequency; that is, presenting some zeros at infinite (low pass shape). This is achieved by some mother wavelets like the Morlet, the Paul, the Shannon and the Spline wavelet (Fig. 4.1 and eqs. 4.8, 4.9, 4.10, and 4.11). Among them, the Paul wavelet has the best time localization capability, but at the same time it has the worst frequency localization [81]. This makes the Paul mother wavelet the more suitable for demodulating fringe patterns that exhibit high signal to noise ratio and rapid phase variations. Besides, the Morlet wavelet presents a Gaussian shape and thus has better localization in the frequency domain than the Paul wavelet. Therefore, it is more suitable for demodulating fringe patterns with slow phase variations and low signal to noise ratios.

4. AUTOMATIC WINDOW SELECTION IN FREQUENCY TRANSFORM TECHNIQUES

$$\Psi_{Morlet}(x) = \frac{1}{(f_b^2 \pi)^{1/4}} \exp(2\pi i f_c x) \cdot \exp\left(\frac{-x^2}{2f_b^2}\right) \quad (4.8)$$

$$\Psi_{Paul}(x) = \frac{2^n n! (1 - ix)^{(n+1)}}{2\pi \sqrt{\frac{(2n)!}{2}}} \quad (4.9)$$

$$\Psi_{Shannon}(x) = \sqrt{f_b} \exp(2\pi i f_c x) (\text{sinc}(f_b x)) \quad (4.10)$$

$$\Psi_{b-spline}(x) = \sqrt{f_b} \exp(2\pi i f_c x) \left[\text{sinc}\left(\frac{f_b x}{m}\right)\right]^m \quad (4.11)$$

where n is the order of the Paul mother wavelet, f_c is the mother wavelet central frequency, f_b is the variance of the window and m is an integer value that determines the Spline wavelet. The selected mother wavelets have been adapted to the use in the multiresolution WFT algorithm proposed in these lines. It must be mentioned that all of them contain a modulated sinusoidal frequency in its definition. Making a comparison with WFT (Eq.(2.25)), this would correspond to the exponential modulating frequency employed. Therefore, the window of WFT is equivalent to the shape of the selected mother wavelet. The introduction of a sinusoidal frequency becomes necessary when it is not implicitly contained in the wavelet definition. Another point to take into account is the normalization of the adapted wavelet signals, as a change in the window size must be compensated by an increment of the modulus of the signal, to preserve the value of energy provided in WFT. Finally, it must be considered the ability to adapt the size of the wave envelope relative to the wave period for many mother wavelets (Morlet, Shannon, Spline). In wavelet analysis, this parameter is used to create a set of complex mother wavelets within the same wavelet family. In WFT this is equivalent to just changing the size of the window, as the preset frequency does not change with this size.

4.3 A new proposal for the automatic selection of the window size

Given the signal in the form $f(x) = a(x, y) + b(x, y) \cdot \cos[\phi(x, y)]$, the accuracy of the retrieved phase is directly linked to the size of the window, the signal envelope and its behavior in frequency domain. Hence, an algorithm to set the optimal window for any WFT signal and for any fringe pattern reveals to be necessary. Regarding this point,

4.3 A new proposal for the automatic selection of the window size

recently Li and Yang [87] proposed a two-step algorithm to determine locally, among a set of patches, the most likely window size for WFT. First, the instantaneous frequencies on x and y direction of the modulated fringe pattern are determined by two-dimensional Gabor wavelet transform (2D-GWT) [88] and, then the local stationary lengths are obtained. Furthermore the so-called Two-dimensional Multiscale Windowed Fourier Transform (2D-MWFT) was applied. This algorithm applied local two-dimensional Gaussian windows, and is performed for each section of the modulated fringe pattern to achieve multiresolution analysis and phase demodulation. Despite the computational cost associated to the two frequency transformations required in this technique, quite good results are obtained as can be observed in [87]. In our work a new proposal for the automatic setting of the window size is done. The proposed algorithm is executed in only one WFT step, taking into account the uncertainty of the received image.

The proposed algorithm is depicted in Fig. 4.2, and described in the following section, emphasizing the steps that permit the automatic detection of the window size.

4.3.1 Preprocessing the image

The preprocessing step consists in a salt and pepper filtering and a histogram equalization. This reduces the noise present in the captured image and enhances the image contrast for a latter frequency component extraction. Finally, a DC filter is applied to extract the DC component of the image. This step delivers an enhanced image where the fringes are perceived more clearly.

4.3.2 Setting the average period and the standard deviation

This step represents the main idea of the automatic selection of the window. The algorithm extracts an approximated value of the number of periods existing in every line along the coding axis, of the image. To do so, a local maximum extraction is performed for the both maximum and the minimum values in every line along the coding axis. The algorithm avoids false positive by suppressing those local maximum that are not followed by a local minimum. Once the number of periods is extracted for every image column, an average of the global period, the corresponding frequency and its variance are computed. This variance represents the uncertainty in the estimated frequency, and is crucial to perform a global analysis of the image. Regarding this point a discussion about whether the selection of global or local variance for patches on the image is required. In principle, a local selection seems to be more appropriate as it can distinguish frequencies of different patches. However, it requires more computation

4. AUTOMATIC WINDOW SELECTION IN FREQUENCY TRANSFORM TECHNIQUES

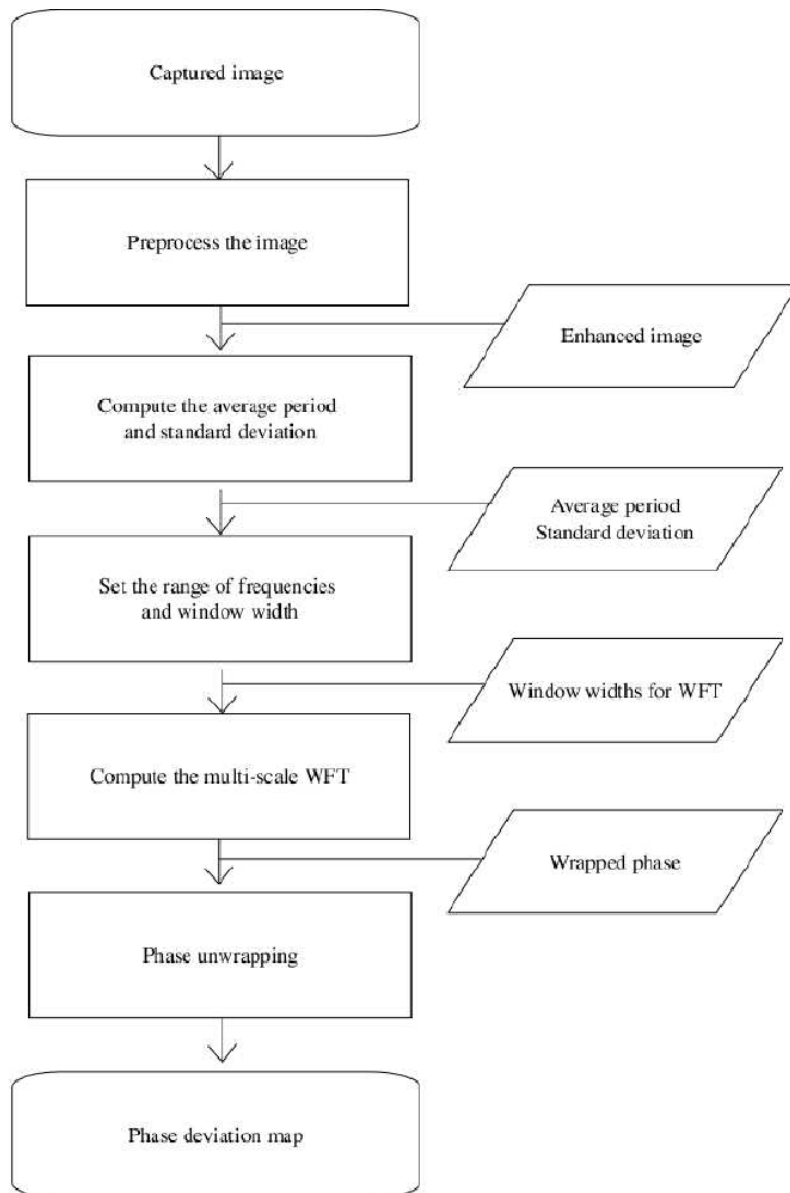


Figure 4.2: Diagram of the proposed algorithm showing the required steps.

4.3 A new proposal for the automatic selection of the window size

as the algorithm must be applied in every patch. Moreover, a global WFT gives a better idea of the carrier frequency, and the global variance is computed accordingly. Therefore it is easier to detect those frequencies corresponding to outliers.

4.3.3 Setting the range of frequencies and the window

The selection of the appropriate range of frequencies is done according to the variance and the average values of the period. For instance, considering the range $[f_m - 3 \cdot std(f), f_m + 3 \cdot std(f)]$ in both x and y axes, the 95% of detected frequencies are analyzed, according to the Central Limit Theorem [89]. The frequencies outbounding this range are considered outliers. In practice, this range can be reduced to $[f_m - 2 \cdot std(f), f_m + 2 \cdot std(f)]$ (90% of the frequencies are represented) without a significant lose in accuracy. Another variable to consider is the window size related to the number of periods of the sinusoidal signal. In contrast to the mother wavelets in WT, WFT does not require the number of periods to be linked to the sinusoidal oscillation of the signal. In WT the number of periods determines a mother wavelet within the same wavelet family, and usually goes from one up to three or four periods, allowing to hold information about the frequency without losing local information. In WFT though, the number of periods can be directly set from the definition of the signal. In our algorithm it has been tested from one up to three periods, determining the optimal value by the ridge extraction algorithm (WFR).

4.3.4 Computing WFT

Once all the parameters are defined, the set of signals having different sinusoidal frequencies and windows are convolved with the enhanced image. As result, a 4D matrix is obtained (having dimensions of x and y axes, window size and frequency). The WFR algorithm is then applied to compute the most likely values of window (wx, wy) , and the corresponding phase value, delivering the wrapped phase in the interval $[-\pi, \pi]$. Finally, the cost function algorithm presented in section 3.2.4 is applied here in order to ensure continuity and avoid errors due to local errors that can be identified and corrected regarding its neighbors.

4.3.5 Phase unwrapping

In order to obtain the unwrapped phase and compute the phase difference with the projected pattern, a phase unwrapping algorithm must be applied. To this end, we use the algorithm of Herraiez et al. [12]. A flow chart of the unwrapping steps is shown in

4. AUTOMATIC WINDOW SELECTION IN FREQUENCY TRANSFORM TECHNIQUES

Fig. 4.3. The algorithm performs a phase unwrapping based on sorting by reliability following a non-continuous path. As stated in [12], the algorithm is suitable for fringe analysis as it minimizes the effects of the noise present in the wrapped phase and provides robust results under presence of slopes (although errors can still arise in some specific discontinuities due to the periodicity of the pattern).

4.4 Testing the algorithm

In order to test the performance of the proposed analysis, it is necessary to observe the output; that is, the wrapped phase, obtained from a set of different fringe patterns. To this end, two kind of test have been done, under simulated and under real conditions. The setup used for the real tests was the same used in chapter 3. All the algorithms were programmed on MATLAB and executed on a standard Intel Core2 Duo CPU at 3.00GHz.

4.4.1 Simulated results

The proposed algorithm was tested using simulated data. The peak function available in Matlab has been considered in the test, since it has become a benchmark for fringe pattern analysis, as stated in [69] (Fig. 4.4). The peak function presents different levels of deformations, therefore it is optimal for analyzing the performance of the adapted signals under different 3D shapes. Some blurring was performed to the image to simulate the noise introduced by the projector-camera pair (Fig. 4.4). In addition, the peak function was also analyzed using WT (specifically the 2D-CWT) in order to compare the performance of both techniques. The results of the wrapped recovered phase are shown in Fig. 4.5.

As can be observed in Fig. 4.5, the best performance of WFT was obtained using the adapted Morlet wavelet and the adapted Paul wavelet. Both wavelets present a clear free of artifacts wrapped phase image. Moreover, the boundaries have been recovered optimally. Worse results are obtained with the adapted spline wavelet (with $m = 2$), as the change in its envelope signal does not suit optimally with the fringe pattern sinusoidal shape. Phase errors are also present across all the image for the analysis using the adapted shannon wavelet, which actually is the spline wavelet with $m = 1$. Looking at the wrapped phase obtained using the wavelet transform, we observe how the errors presented in WFT also appear in WT. This enforces the idea that Morlet and Paul wavelets are more suitable for fringe pattern analysis than Shannon and Spline wavelets. Moreover, we can appreciate some errors on the image corner for the case of

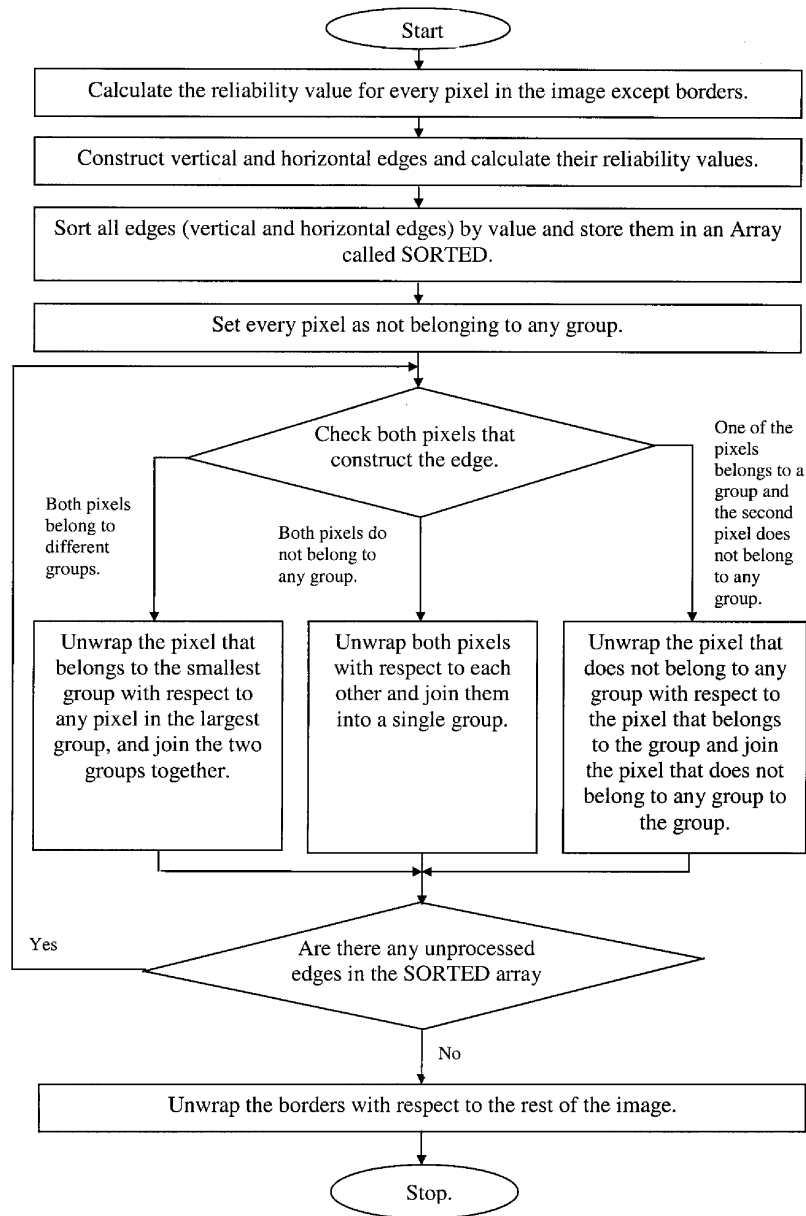


Figure 4.3: Flow chart of the unwrapping algorithm proposed by Herraez et al. [12].

4. AUTOMATIC WINDOW SELECTION IN FREQUENCY TRANSFORM TECHNIQUES

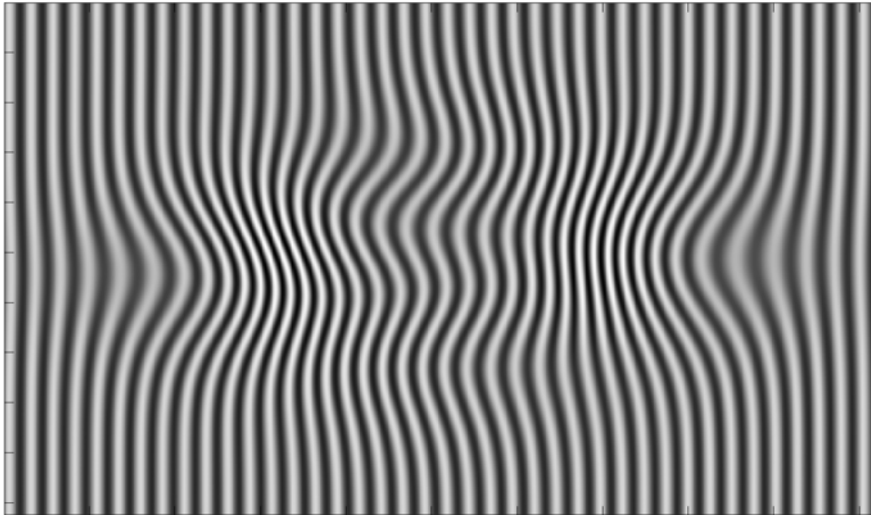


Figure 4.4: Peak function used as the input image, containing phase variation at different speeds, and added noise

the Shannon and the Spline wavelet analysis, not present with WFT phase extraction algorithm proposed in this work. Table 4.1 provides some quantitative results of the previous recovered phases compared to the input phase, where the relative mean error represents the mean of the sum of the absolute error between the computed wrapped phase map and the input phase map. Finally, as example of the reconstruction results, the unwrapped phase map of the Morlet WFT is shown in Fig. 4.6.

Table 4.1: Relative mean error of the recovered phase

Error (%)	Morlet	Paul	Shannon	Spline
WFT	0.073	0.058	0.156	0.169
WT	0.115	0.067	0.162	0.174

4.4.2 Experimental results

This section analyses the performance of the proposed algorithm reconstructing two different objects. The first is a plastic lambertian sheet having a smooth surface with an irregular deformation in the middle of the object. The second object is a rubberised elastic polymer, representing the face of a radiotherapy RANDO phantom. The proposed technique is applied and the corresponding wrapped phase maps are compared. The unwrapping algorithm developed by Herraes et al. [90] is employed to extract the unwrapped phase map of the objects. The input images and their corresponding un-

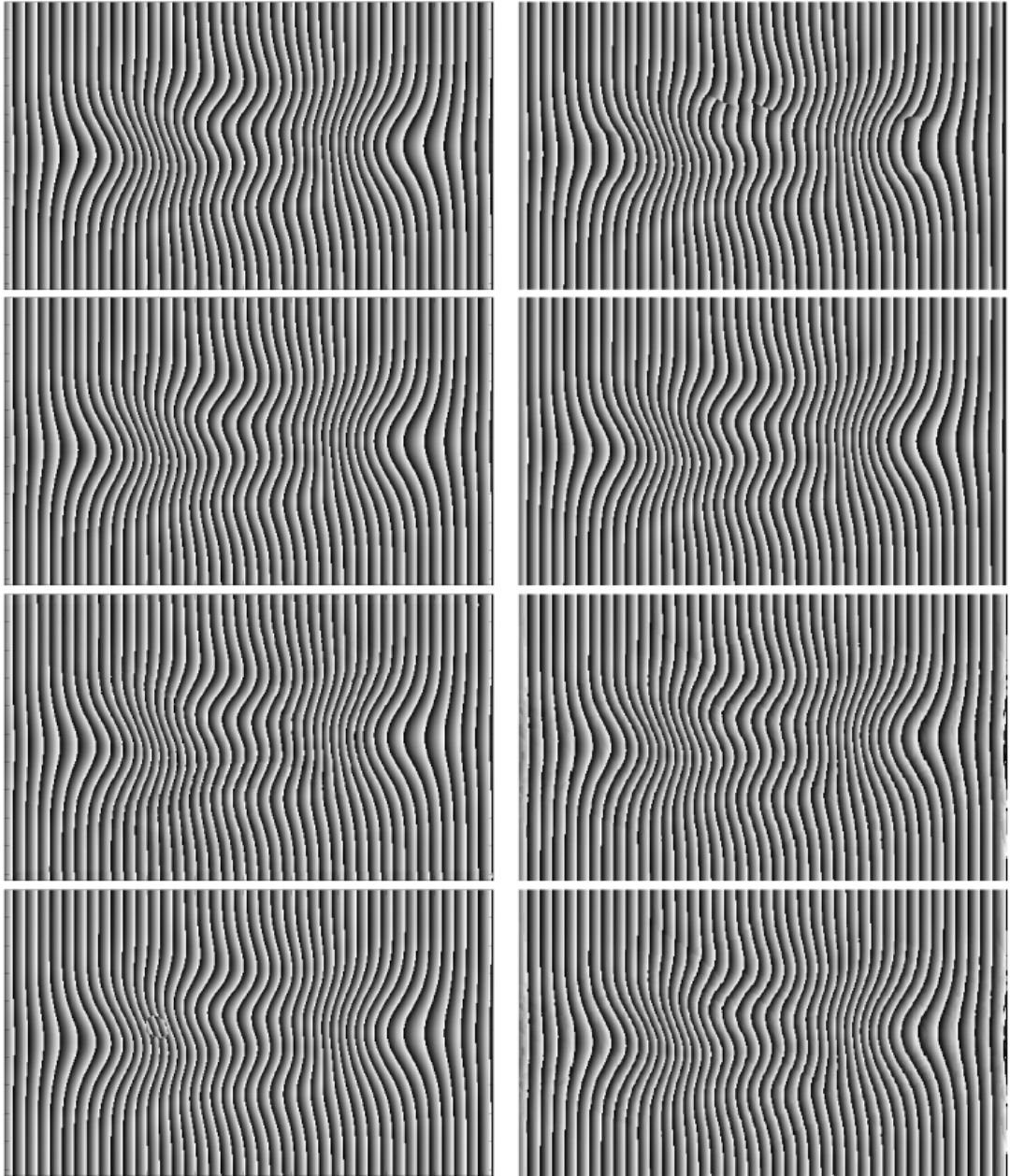


Figure 4.5: Simulation results of the wrapped phase (from $-\pi$ to π): on the left column the four adapted mother wavelets (Morlet, Paul, Shannon, Spline). On the right column, their corresponding results in WT.

4. AUTOMATIC WINDOW SELECTION IN FREQUENCY TRANSFORM TECHNIQUES

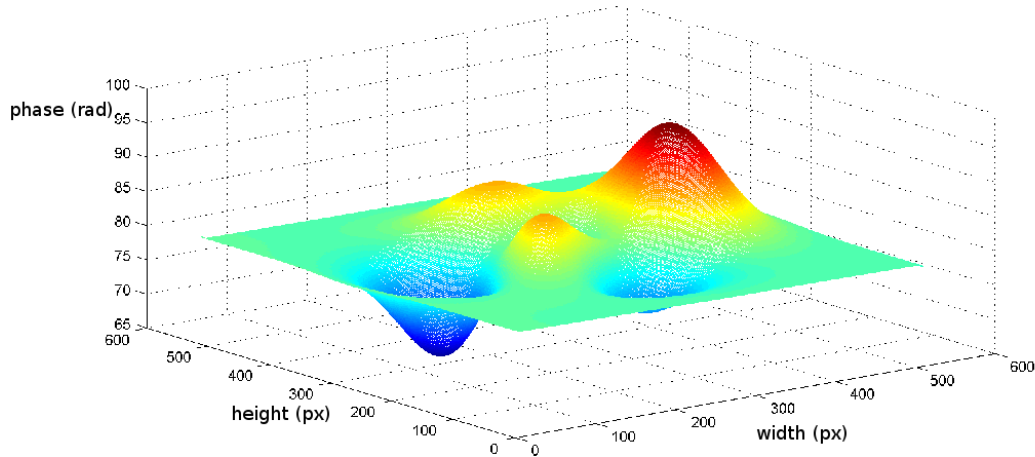


Figure 4.6: Recovered unwrapped phase, in radians, for the simulated input fringe pattern of Fig. 4.4. Morlet WFT has been used.

wrapped phase maps are shown in Fig. 4.7 and Fig. 4.9, respectively. The wrapped phase maps corresponding to either the output of WFT or WT (specifically the 2D-CWT) are analyzed in detail hereafter.

Plastic sheet

The Morlet and the Paul wavelets suffer from larger error than the respective signals in WFT, as can be observed in Fig. 4.8. The errors are located in the region having larger variation in depth. The wavelet analysis introduces some doubled frequency components due to the effect of the dyadic net in the phase estimation. This does not happen in WFT, where the discrepancy to the correct phase is lower. The same errors are present in the Shannon WT mode. The Shannon WFT, however, performs optimally for this image. Increasing the value of m up to $m = 2$, though (in fact the Spline function with $m = 2$) errors arise in some region of the captured image. Therefore, Morlet, Paul and Shannon WFT techniques perform optimally in this case.

Radiotherapy RANDO phantom

Phase estimation has been pursued for the four wavelet signals in both WT and WFT. The results are shown in Fig. 4.10. Best results are obtained for the Morlet signal, as in the previous images. Among them, WFT performs better than WT in presence of slopes, as can be noise near the nose and the eyes.

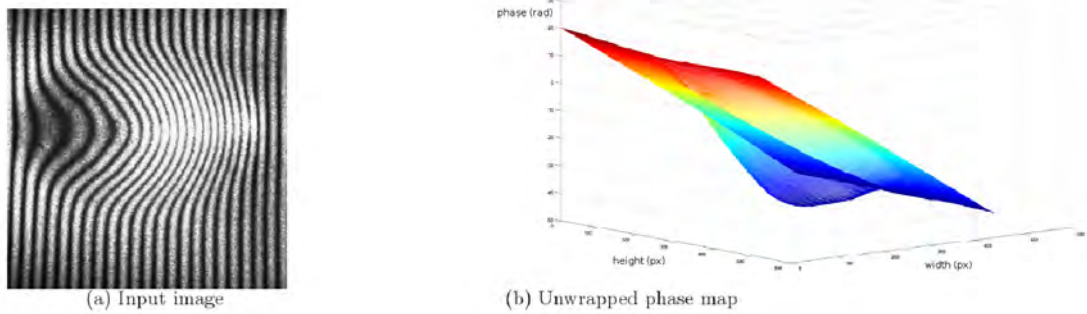


Figure 4.7: Input image and reconstructed unwrapped phase map in radians (using Paul wavelet for WFT).

4.5 Conclusions

In this chapter, we faced the problem of fringe pattern analysis in the frequency domain. Both WFT and WT are currently used in fringe profilometry, and pros and cons are found in both cases [69]. First, we performed a comparative study of both techniques. The results show how WFT with automatic window size detection performs better compared to the traditional WT technique. This is mainly due to the nature of the dyadic net used in WT, which scales the window by a factor of 2^j between adjacent size values. Afterwards we proposed an automatic window width selection algorithm for WFT, which performs a fine tune of the window size between a set of values around the optimal size corresponding to the global frequency average. Some simulated and real results have been obtained, proving that the Morlet and the Paul wavelets show better performance than the Shannon and the Spline wavelets, thanks to the greater similarity of its signal shape to the sinusoidal nature of the fringe pattern. Summarizing, the fine tuning provided by the automatic window selection algorithm assures an optimal selection of the local frequency. This method can be used in any fringe-based pattern analysis, minimizing the error in the recovered phase deviation map. Therefore, an optimal approach would be to use this frequency analysis in an hybrid pattern that avoids any phase unwrapping step by means of color or frequency multiplexing.

4. AUTOMATIC WINDOW SELECTION IN FREQUENCY TRANSFORM TECHNIQUES

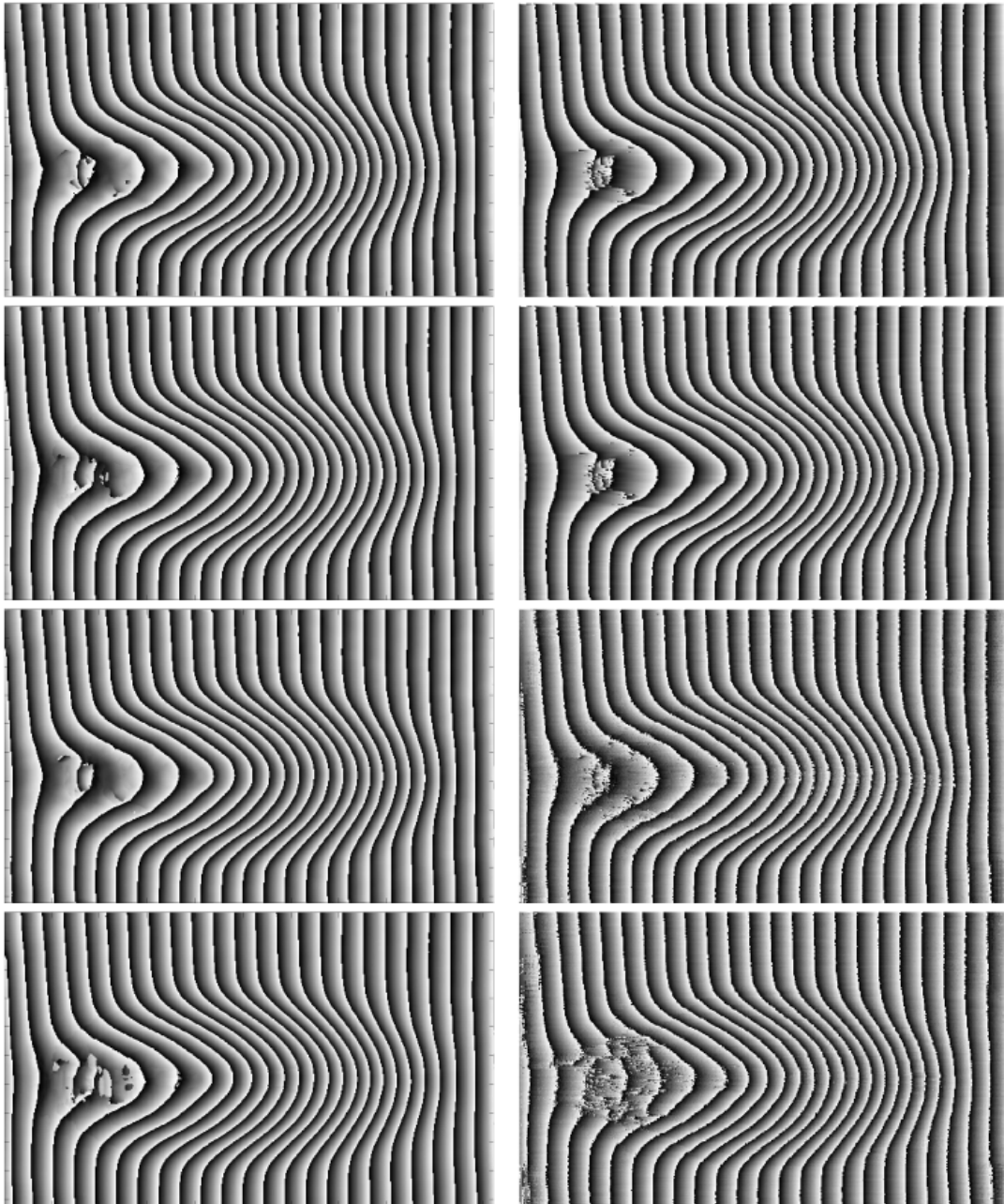


Figure 4.8: Plastic bended sheet: on the left column, experimental results for the four adapted mother wavelets (Morlet, Paul, Shannon, Spline). On the right column, their corresponding results in WT.

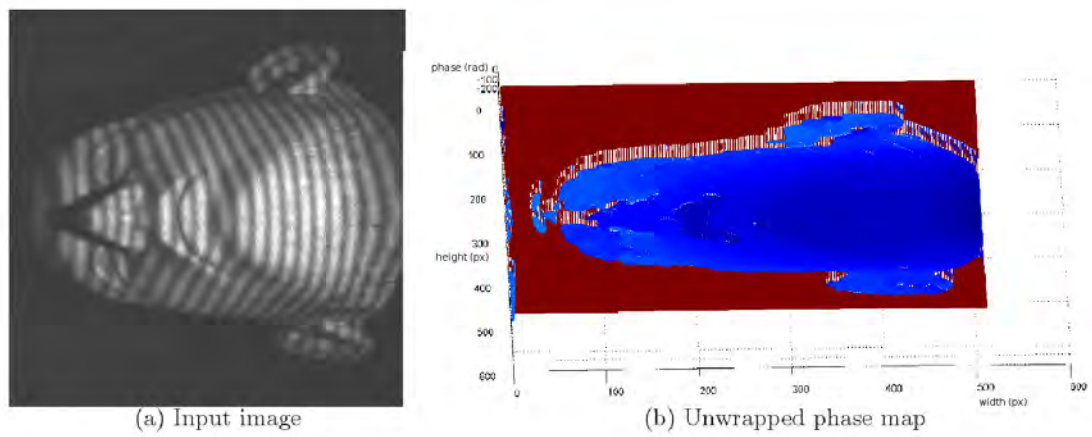


Figure 4.9: Input image and reconstructed unwrapped phase map in radians (using Paul wavelet for WFT).

4. AUTOMATIC WINDOW SELECTION IN FREQUENCY TRANSFORM TECHNIQUES

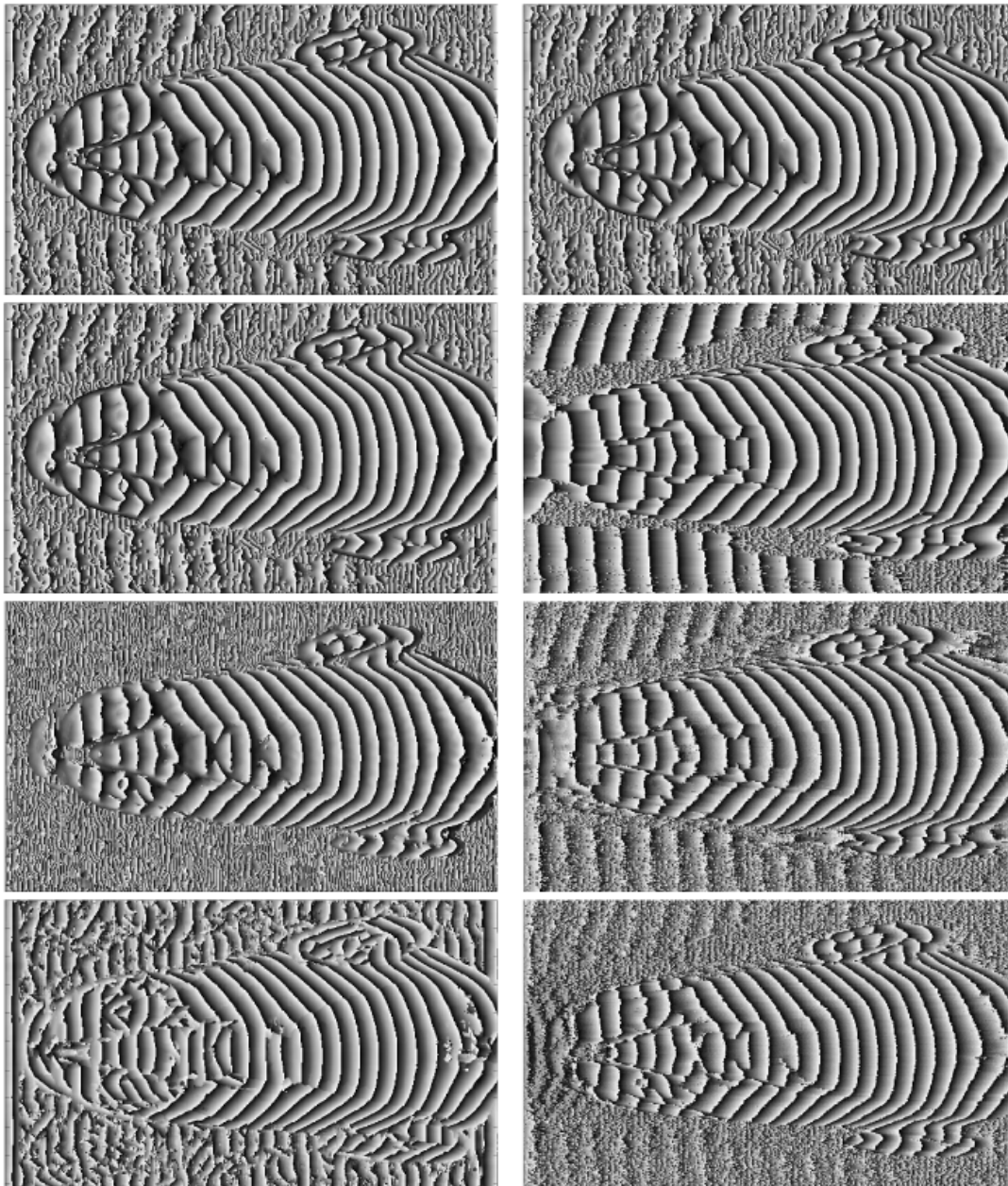


Figure 4.10: Radiotherapy RANDO phantom: on the left column, experimental results for the wrapped phase (from $-\pi$ to π) of the four adapted mother wavelets (Morlet, Paul, Shannon, Spline). On the right column, their corresponding results in WT.

5

A proposal on one-shot absolute pattern for dense reconstruction

In this chapter we propose a new algorithm for absolute phase unwrapping based on the use of DeBruijn coding. Combining this approach with the modified WFT algorithm proposed in chapter 4, the new proposal obtains a dense, absolute, accurate and computationally fast 3D reconstruction using a one-shot pattern. This is achieved thanks to the good accuracy results provided by DeBruijn coding, and the density of reconstruction provided by WFT. The algorithm is presented and compared with other existing techniques. The results presented at the end of the chapter show that the pattern obtains levels of accuracy comparable to classical DeBruijn patterns, but with dense reconstruction results. Finally, the advantages and disadvantages of the proposed technique are pointed out.

5.1 Introduction

As was pointed out in previous chapters, an important effort has been done in order to generate patterns able to obtain correspondences with a unique projection. Moreover, the need to obtain dense reconstructions and absolute coding is an asset in any solution proposed. In chapter 4 we proposed a new WFT-based analysis to recover the wrapped phase map of an input fringe pattern. The main advantage of the proposed analysis is that the behavior under discontinuities is optimal, in the sense that errors caused by a wrong selection of the carrier frequency are minimized, even under regions with big changes in frequency like discontinuities. The next step would be to unwrap the phase in order to obtain the depth and hence the 3D reconstruction of the scene. Many

5. A PROPOSAL ON ONE-SHOT ABSOLUTE PATTERN FOR DENSE RECONSTRUCTION

different phase unwrapping algorithms can be found in the literature ([72],[73],[68]). However, despite being complex and computationally slow they can fail in the presence of depth discontinuities and occlusions, performing only periodic coding (non-absolute). Furthermore, the problem of absolute coding still remains unsolved. In this chapter we propose a new algorithm for absolute phase unwrapping based on the use of DeBruijn color coding. It is seen from chapter 2 that DeBruijn coding provides the best performance in terms of accuracy among sparse reconstruction techniques. The idea is to combine these benefits with dense fringe-based patterns using the Windowed Fourier Transform (WFT) analysis proposed, with the goal of obtaining from a unique image an absolute, accurate and computationally fast 3D reconstruction.

The chapter is structured as follows: section 5.2 presents the design of the new technique especially focused on the absolute coding unwrapping. Experimental results with both simulated and real data are presented in section 5.3, including a comparison with other existing SL techniques. Finally, section 5.4 concludes with a discussion of the proposed method, analyzing its advantages and disadvantages compared to literature.

5.2 A new proposal for one-shot dense reconstruction

The proposed technique combines the benefits of DeBruijn coding in dense fringe-based patterns using Windowed Fourier Transform (WFT) analysis. Therefore, it is possible to obtain the accuracy provided by classical DeBruijn stripe-based patterns and the density of fringe-based patterns. A general scheme of the algorithm is shown in Fig 5.1 and explained in the following lines.

5.2 A new proposal for one-shot dense reconstruction

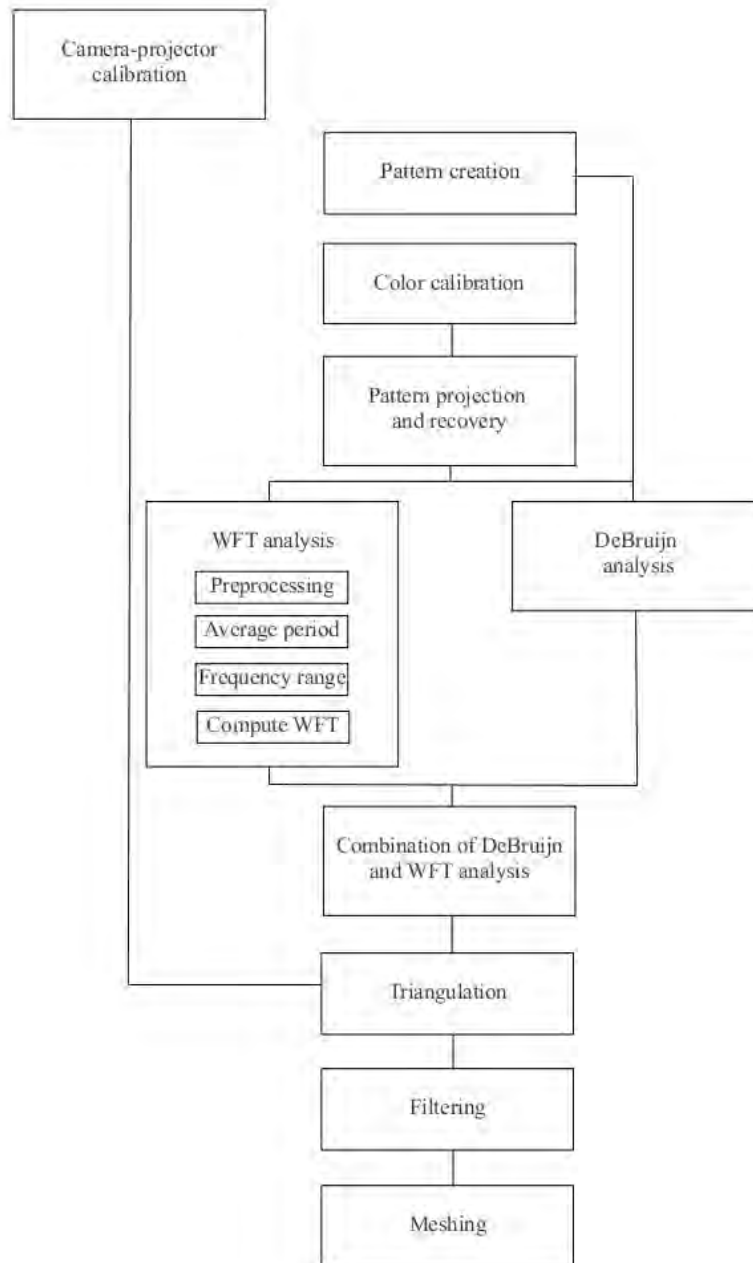


Figure 5.1: Diagram of the proposed algorithm

5. A PROPOSAL ON ONE-SHOT ABSOLUTE PATTERN FOR DENSE RECONSTRUCTION

5.2.1 Pattern creation

The proposed pattern consists in a colored sinusoidal fringe pattern, where the color of the different fringes follows a DeBruijn sequence. DeBruijn sequences are a set of pseudo random values having specific properties between them. A k -ary DeBruijn sequence of order n is a circular sequence $d_0, d_1, \dots, d_{n^k-1}$ (length n^k) containing each substring of length k exactly once (window property of k). DeBruijn sequences can be constructed directly from the Hamiltonian or Eulerian path of a n -dimensional DeBruijn graph (see [35] for more details). In our approach we set $n = 3$ as we work only with red, green and blue colors. The number of fringes contained in the pattern is determined by two constraining parameters: the pixel resolution of the camera and that of the projector device. In any case, it is selected to be a factor number of the total height of the pattern, in order to have an entire number of periods. Regarding the devices used in our experiments, we set the pattern to have 64 fringes which is the maximum resolution that the camera can capture without a significant loss in the shape of the sinusoidal signal. Therefore, $n^k \geq 64$, so we set the window property to $k = 4$. An algorithm performing the sequence generation provides us an arbitrary DeBruijn circular sequence d_0, d_1, \dots, d_{80} . The pattern, of size $m \times n$, is generated in the HSV space. This is done to minimize the effect of crosstalk in the fringe pattern (where every pixel has a different value), as it would happen using RGB and one color channel instead. For every column $j = 1..m$ of the V channel, the sinusoidal signal is represented as in eq. 5.1:

$$I(i, j) = 0.5 + 0.5 \cdot \cos(2\pi f i) \quad (5.1)$$

where $i = 1..n$ and the discrete frequency $f = 64/n$.

The H channel maps a value of the previously computed DeBruijn sequence to every period of the V channel. The S channel is set to 1 for all the pixels to obtain the maxims of the saturation value. Finally, the created HSV matrix is transformed into RGB values. The resulting pattern is shown in Fig. 5.2.

5.2.2 Geometric and color calibration

Using the calibration steps proposed in appendix A, the intrinsic and extrinsic transformation matrices of the projector-camera system are determined. This gives us the relative position between both devices, as well as their optical parameters. Moreover, color calibration matrix proposed in appendix B is applied to the incoming images in order to reduce the effects of crosstalk and attenuation of some color channels among others. Finally, RGB color channel alignment must be applied to the incoming image

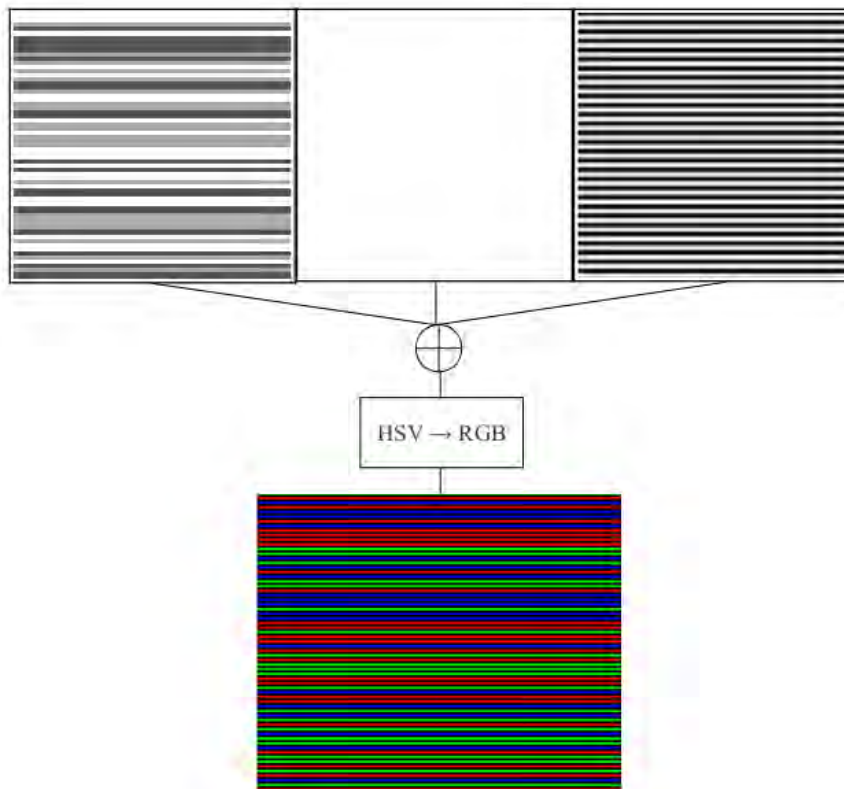


Figure 5.2: Proposed pattern: HSV representation of the DeBruijn sequence (top-left), fringe pattern (top-right) and the resulting RGB pattern for $m = 64$ (bottom).

in order to subtract the small misalignment present between the three color channels of the DLP and the camera.

5.2.3 Pattern projection and recovery

The designed pattern is projected by the active device. Once reflected onto the object the pattern is recovered by the camera. The camera calibration matrix is applied to the RGB image, obtaining the corrected color values. The corrected RGB image is transformed to the HSV space. Afterwards, a Region Of Interest (ROI) is selected regarding the information held in the V plane. To this end, a closure morphological operation followed by a binarization is computed. In the binarization step, the pixels exceeding the value given by the Otsu thresholding algorithm are set as white, and the rest as zero. The positions corresponding to the white pixels constitute the ROI, whereas the other positions will not be considered for the pattern analysis. The ROI is applied to both the corrected RGB image and the V matrix. The masked RGB image

5. A PROPOSAL ON ONE-SHOT ABSOLUTE PATTERN FOR DENSE RECONSTRUCTION

is the input of the DeBruijn detection algorithm, whereas the masked V matrix is used in the Windowed Fourier Transform Analysis.

5.2.4 DeBruijn analysis

The aim of this step is to extract the color associated to every projected colored fringe. We followed the same approach used in slit-based pattern, as the recovered fringes present a Gaussian-like shape similar to that present in slit-based patterns ([23], [29]). Therefore a maxims localization algorithm is applied, searching local maxims of every color channel on the current scan-column. Then, it calculates the sub-pixel center of mass of the region taking into account only those pixels for which its normalized intensity is higher than a certain threshold (set to 0.7 during the experiments). This is done using the first and second derivative of every column, whose zero-crossing and maxims provides us with an accurate subpixel position (see Fig 5.3).

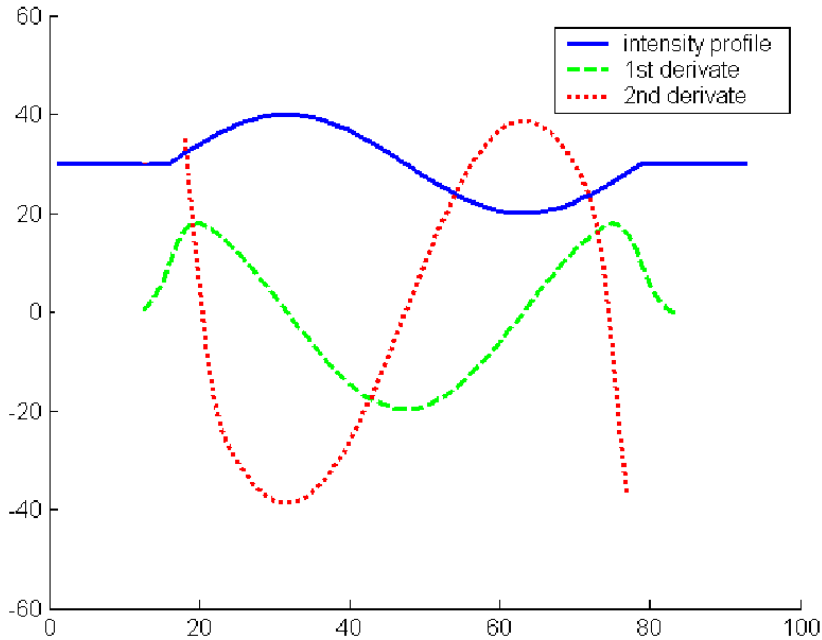


Figure 5.3: Behaviour of the 1st and 2nd derivative on a synthetic sinusoidal signal.

The implemented algorithm takes into account the total of $n = 64$ periods present in the pattern, and the consecutive maxims-minims distribution. Furthermore, a global threshold suppressing peaks lower than the 70% of maximum peak is applied. These steps prevent from false peaks detection.

5.2.5 Dynamic programming

Ideally, the matching of correspondences between the projected sequence and the perceived one is straightforward. However, usually the whole sequence of projected stripes is not visible on the image scan-line or some of them are incorrectly labeled or disorders may occur. Therefore, the use of dynamic programming becomes compulsory. Dynamic programming works as a scoring function, measuring the similarity between a certain element of the projected sequence and an element of the perceived one, and scoring to the maximum correspondence (an explanation of dynamic programming and the employed multi-pass dynamic programming can be found in appendix D). In our case the dynamic programming algorithm set the correspondence between the recovered sequence of color stripes and the corresponding section of color stripes in the projected DeBruijn sequence. This procedure is done minimizing the errors due to noise and occlusions.

5.2.6 Windowed Fourier Transform analysis

WFT has been chosen for frequency fringe analysis, as it avoids leakage distortion and a more precise window width selection than in WT. First, a salt and pepper filtering and a histogram equalization is applied to the V channel. Afterwards, an adapted Morlet wavelet is chosen for WFT analysis. Regarding the work of Fernandez et al. [91], this provides good frequency and spatial localization at the same time. Morlet signal definition is shown in eq. 5.2:

$$\Psi_{Morlet}(x) = \frac{1}{(f_c^2 \pi)^{1/4}} \exp(2\pi i f_c x) \cdot \exp\left(\frac{-x^2}{2f_b^2}\right) \quad (5.2)$$

where f_c is the mother wavelet central frequency and f_b is the window size.

The average and standard deviation of the fringe period is estimated counting the number of periods existing in every column along the coding axis (using the same algorithm employed to find the local maxims of the DeBruijn sequence). Average period (pm) and standard deviation (std) are extracted from the single periods corresponding to each column. The std represents the uncertainty in the estimated frequency, and is crucial to perform a global analysis of the image. The average frequency for the $n \times m$ pattern is computed as $f_m = n/pm$. The frequencies analyzed are in the range $[f_m - 3 \cdot std, f_m + 3 \cdot std]$ in both x and y axes, where f_m is the average frequency. Using this range the 99% of detected frequencies are analyzed. In practice, this range can be reduced to $[f_m - 2 \cdot std, f_m + 2 \cdot std]$ (95% of the frequencies are represented)

5. A PROPOSAL ON ONE-SHOT ABSOLUTE PATTERN FOR DENSE RECONSTRUCTION

without a significant loss in accuracy. Another variable to consider is the window size related to the number of periods of the sinusoidal mother signal. In contrast to the mother wavelets in WT, WFT does not require the number of periods to be linked to the sinusoidal oscillation of the signal. In our algorithm it has been used from one up to three periods. The optimal value is selected applying the ridge extraction algorithm (WFR), followed by the cost function algorithm. This step compute the most likely values of window (w_x, w_y) and the corresponding phase value. Afterwards, the wrapped phase in the interval $[-\pi, \pi]$ is obtained.

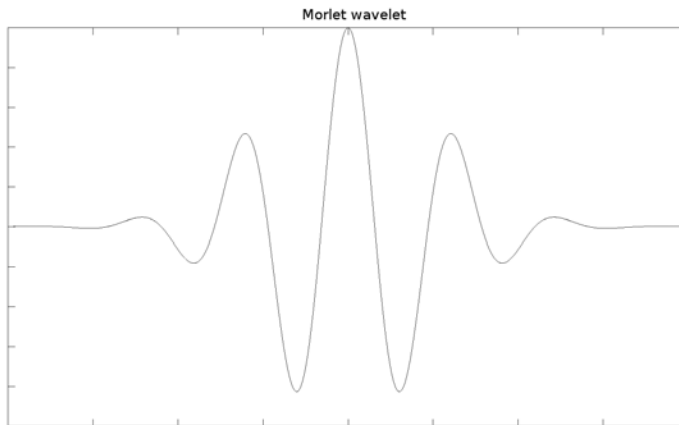


Figure 5.4: Visual representation of a Morlet signal with $n = 3$ periods.

5.2.7 Combination of DeBruijn and wrapped phase patterns

The next step is to merge the information obtained from WFT and DeBruijn. The wrapped phase is merged with the extracted colored lines. Due to the 2D nature of WFT (which may include some frequencies of adjacent positions in the Fourier Transform), the phase value of an specific position may have some deviation. This effect is corrected shrinking or expanding the wrapped phase accordingly to the DeBruijn correspondences for the maxims. A non-linear 4th order regression line is used to this end, matching the maxims of the wrapped phase map with the position of the colored lines in the DeBruijn map. This process is done for every column on the image, obtaining corrected wrapped phase map. This is shown in Fig. 5.5. Finally, the correspondence map provided by the DeBruijn lines is expanded using the wrapped phase map. The phase values between two adjacent lines go in the range $(-\pi, \pi)$. Therefore a direct correlation is set between these values and the position of the projected and the recovered color intensities. A

5.2 A new proposal for one-shot dense reconstruction

full (dense) correspondence map is obtained.

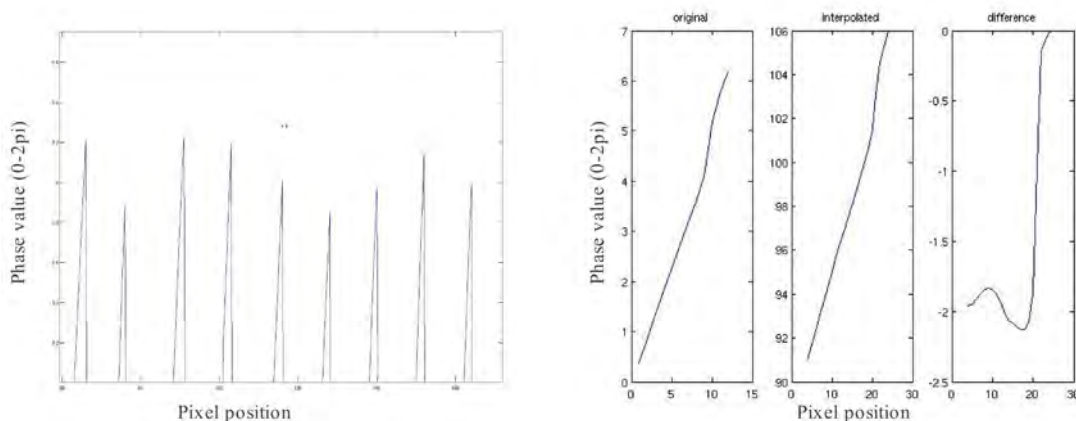


Figure 5.5: On the left, detail of the wrapped phase and a crest maxima (in red), and its corresponding slits line position (in green). On the right, the wrapped section before and after correction, and the correction interpolation error.

5.2.8 Triangulation

Every pair of (x, y) projector-camera coordinates given by the matching step are inputs in the triangulation module, which also makes use of the extrinsic and intrinsic parameters provided by the calibration module. The output is a cloud of points in (x, y, z) representing the shape of the reconstructed object. It is important to mention that the triangulation is done following a ray-to-plane intersection approach (Fig. 5.6), as one of the coordinates of the projected pattern remains unknown.

5.2.9 Filtering

A posterior filtering step reveals necessary due to some erroneous matchings that originate outliers in the 3D cloud of points. Two different filtering steps are applied.

3D statistical filtering: in the 3D space, the outliers are characterized by their extremely different 3D coordinates regarding the surrounding points. Therefore, pixels having 3D coordinates different than the 95% of the coordinates of all the points are considered for suppression. This is done in two steps for all the points in the 3D cloud. First the distance to the centroid of the cloud is computed, for every pixel. Afterwards, those pixel having a distance to the centroid greater than two times the standard deviation of the cloud of points are considered as outliers.

5. A PROPOSAL ON ONE-SHOT ABSOLUTE PATTERN FOR DENSE RECONSTRUCTION

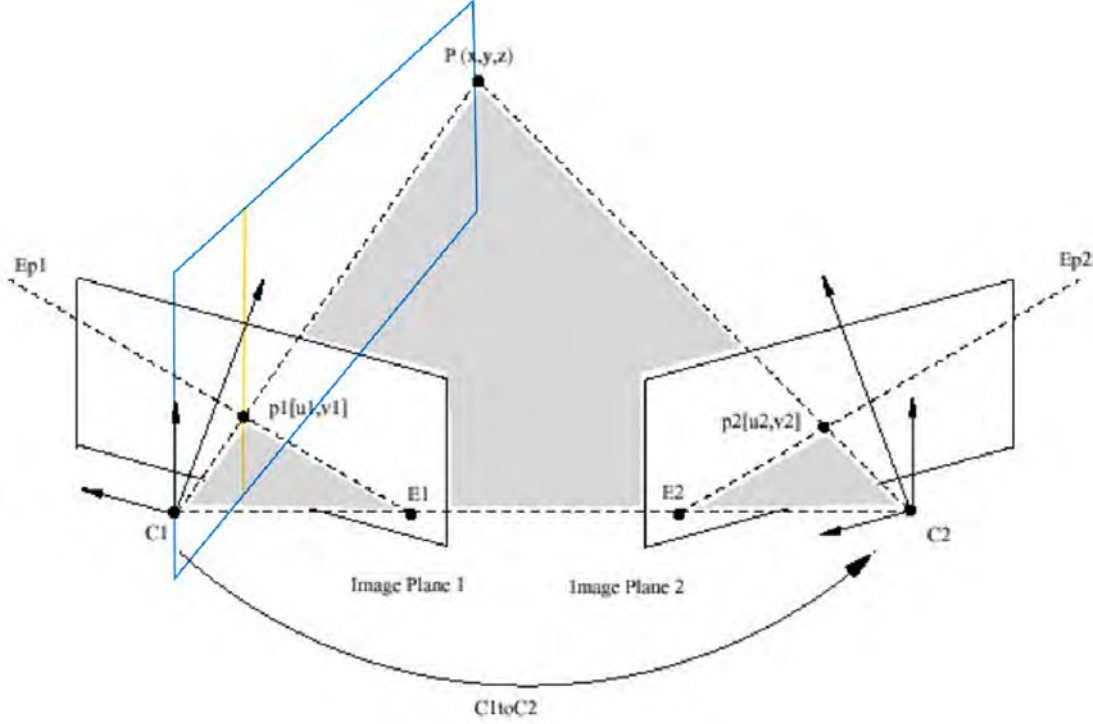


Figure 5.6: Ray plane intersection diagram

Bilateral filtering: Still, there can be some misaligned points after applying the statistical filtering. In this case it would be profitous to apply some anisotropic filtering that filters the data while preserving the slopes. To this end, an extension to 3D data of the 2D bilateral filter proposed by Tomati and Manduchi [92] was implemented. The bilateral filter is a non-recursive anisotropic filter whose aim is to smooth the cloud of points (up to a given value) while preserving the discontinuities, by means of a nonlinear combination of nearby point values. The proposed 3D bilateral filtering is described in eq. 5.3 and eq. 5.4.

$$G(x, y) = \exp(-((x - x_c)^2 + (y - y_c)^2)/(2 * \sigma_1^2)) \quad (5.3)$$

$$H(z) = \exp(-(z - z_c)^2/(2 * \sigma_2^2)) \quad (5.4)$$

where x, y, z are the 3D coordinates of a given point, $G(x, y)$ is the distance mask, $H(z)$ is the height mask, and both σ_1 and σ_2 are values to set empirically. The algorithm works as follows: given a set of points X, Y, Z around the selected 3D point, the cor-

responding masks $G(x, y)$ and $H(z)$ are computed, providing the filtered results. This modified the height of those isolated pixels having a 3D coordinates much different than their vicinity, while preserving the slopes.

5.2.10 Meshing

Finally, an optional meshing step can be applied to obtain a surface from the 3D cloud of points. To do so, the 2D bidimensional Delaunay meshing should be applied on the 3D coordinates with respect to the camera, in order to avoid duplicities in the depth value as this can not occur from the camera point of view.

5.3 Results

The proposed algorithm was implemented and tested in both simulated and real data. Moreover, a comparison with other representative SL algorithms was pursued. To this end we used the techniques presented in chapter 2, as they correspond to the main groups existing in SL, not only in dense but also sparse reconstruction. The new setup used for the experiments is composed by a compact handy projector-camera system. The DLP is a video pico-projector 3M MPro 150 with a resolution of 1024×768 pixels. The camera is a Canon EOS 50D set to a resolution of 1568×2352 pixels with 3×8 bits per pixel (RGB). The lens used is a Canon EF of $24mm$. The baseline between camera and projector was about $40cm$. The setup is shown in Fig. 5.7.

Experiments were computed in a desktop computer, Intel Core2 Duo CPU at 3.00GHz and 4GB RAM memory. The algorithms were programmed and ran in Matlab 7.3. It is important to mention the methods used for comparison were re-programmed from the corresponding papers, since at the best of our knowledge source codes were not available.

5.3.1 Simulation results

The *peaks* function available in Matlab (shown in Fig. 5.8) was chosen for 3D reconstruction. Moreover, an analysis against noise was pursued. The error introduced in the recovered (incoming) image was a Gaussian zero mean random noise with standard deviation of 0.05%, 0.1%, 0.15% and 0.2%. The results are shown in Fig. 5.9, where a comparison with other one-shot techniques selected in [29] was done.

As can be observed, the worst results are obtained by the algorithm of Carrhill and Hummel. The algorithm of Su et al. performs optimally for low noise values thanks

5. A PROPOSAL ON ONE-SHOT ABSOLUTE PATTERN FOR DENSE RECONSTRUCTION

to the nature of the 2D frequency analysis, which smooths the incoming data. This occurs also for the proposed pattern. However, for noisy images having $std > 0.1$ the 1D unwrapping step introduces discontinuities in the recovered phase, leading to errors in the pixel position. Besides, Monks algorithm suffers the low amount of reconstructed points, which penalizes the errors produced in the slits position. Finally, the proposed algorithm provides a much denser reconstruction with 2D fourier analysis and no need to perform any phase unwrapping. This gets reflected in the results, performing the best among the three tested techniques.

5.3.2 Empirical results

Quantitative results were analyzed reconstructing a white plane at a distance of about $80cm$ in front of the camera. Principal Component Analysis (PCA) was applied to obtain the equation of the 3D plane for every technique and for every reconstruction. PCA is used to span the 3D cloud of points onto a 2D plane defined by the two eigenvectors corresponding to the two largest eigenvalues. The results of the experiment are shown in table 5.1.

Table 5.1: Quantitative results. The headings are: author’s name of the technique; average deviation of the reconstructing error; standard deviation of the reconstructing error; number of 3D points reconstructed; number of projected patterns.

Technique	Average (mm)	Stdev (mm)	3D Points	Patterns	Time (s)
Monks et al.	1.31	1.19	13899	1	45.29
Posdamer et al.	1.56	1.40	25387	14	32.18
Guhring	1.52	1.33	315273	24	158.22
Pribanic et al.	1.12	0.78	255572	18	165.65
Carr.and Hummel	11.9	5.02	202714	1	150.57
Proposed technique	1.18	1.44	357200	1	160.75

Note that the algorithm of Su et al. [10] is conceived to measure deviation of smooth surfaces with respect to the reference plane, therefore a plane is not conceived to be reconstructed by depth deviation. As can be observed, the proposed technique obtains one of the best accuracy results in terms of average and standard deviation of the error, only overcome by the method of Pribanic et al. [9], which requires a total of 18 projected patterns. Among the one-shot projection techniques, the proposed technique obtains the best accuracy results jointly with another DeBruijn based technique, the sparse reconstruction algorithm proposed by Monks et al. [6]. Regarding the computing time it can be observed that methods obtaining dense reconstructions (the case of Guhring, Pribanic et al., Su et al., and Carrihill and Hummel and the proposed algorithm)

need to compute more 3D points, requiring higher computational time. However, our proposal does not need to compute many images, nor any unwrapping algorithm is required. This makes our technique faster in terms of computational time. Among methods providing sparse reconstruction the color calibration step makes Monks et al. algorithm slower than Posdamer et al. (also affects the proposed technique) despite it preserves the same order of magnitude. Still, real time response is achievable working with the appropriate programming language and firmware.

Finally, **Qualitative results** were pursued reconstructing several 3D objects. The lambertian objects were placed at a distance of about 80cm to the camera.

Results of 3D reconstruction of several objects are shown in Fig. 5.10, Fig. 5.11, Fig. 5.12, Fig. 5.13, Fig. 5.14, Fig. 5.15, Fig. 5.16, and Fig. 5.17. The first reconstruction corresponds to a bended piece of paper sheet. The second reconstruction is a piece of manufactured white cork, used to protect an electronic device inside a box. Third and fourth reconstructions are a ceramic figure of a 'hello kitty' and the sculpture of a horse, respectively. For every object, the first row corresponds to the input and the color calibrated images. The second rows are the extracted color channels, after color rectification. The third row corresponds to the results of DeBruijn analysis (the slits image) and WFT computation after merging with the DeBruijn images; that is, the tuned wrapped phase image. On the bottom, the corresponding triangulated 3D cloud of points is shown. As can be observed the objects are reconstructed optimally. Only the 'hello kitty' present some points missing, as the filtering suppressed 3D outliers associated to the low illuminated and blurred regions present in the recovered image.

5.3.3 Reconstructing color surfaces

Color surfaces represent a challenge for DeBruijn based SL patterns. In the proposed approach the information related to spatial decoding is held in the sequences of colors of the fringes. When projecting onto a color surface, the color of the fringes get distorted by the original color of the object. Therefore, the identification of the color sequence by the DeBruijn decoding algorithm may present some errors. This phenomena has been tested for a real colored object. The aim was to reconstruct a planar surface having different colors, in different positions not related with the orientation of the fringes. A Macbeth colorchecker was employed for this. This is a color calibration target consisting in a cardboard-framed arrangement of 24 squares of painted samples. Originally proposed by McCamy et al. [93], the charts color patches have spectral reflectance intended to mimic those of natural objects such as human skin, foliage,

5. A PROPOSAL ON ONE-SHOT ABSOLUTE PATTERN FOR DENSE RECONSTRUCTION

and flowers, to have consistent color appearance under a variety of lighting conditions, especially as detected by typical color photographic film, and to be stable over time. The Macbeth checker was scanned using the proposed SL pattern. The results of scanning and 3D reconstruction are shown in Fig. 5.18 and Fig. 5.19:

As can be observed, the Hue and the Illumination channels gets corrupted and color of the fringes is not well retrieved. Therefore, the reconstruction present big errors all around the plane, as the dynamic programming algorithm failed when minimizing the distance between the projected and the recovered color fringes (see slits plane and fringe plane). This effect occurs mainly in dark regions, as the reflected illumination is not high enough to detect the fringe colors in the DeBruijn algorithm. Fortunately, these errors are filtering by the statistical and the bilateral filtering. As result, the 3D reconstruction present big holes in the regions having low illumination rate, but the structure of the flat plane is preserved.

5.4 Conclusions

Continuous coding strategies achieve dense acquisition by creating a pattern that changes the value between adjacent pixels. The depth of a given point on the image is determined by the deviation of its grey value with respect to the projected pattern. Moreover, one-shot projection reveals necessary in order to work in moving scenarios. The combination of dense reconstruction and one-shot projection has been overcome by the fringe pattern approaches. However, they fail under presence of big slopes and occlusions, due to the periodicity of the fringe pattern. Therefore, the difficulty falls on the achievement of absolute coding for one-shot dense projection patterns. This chapter proposes a new technique for one-shot dense 3D surface reconstruction, which combines the accuracy of DeBruijn spatial multiplexing with the density of frequency multiplexing in fringe projection. The proposal was implemented and compared both quantitatively and qualitatively with some representative techniques of Structured Light. Simulation results and empirical quantitative results showed the good performance of the proposed technique in terms of resistance to noise and accuracy of a reconstructed plane. Among one-shot techniques, our proposed method achieves the best results in terms of accuracy, comparable with other DeBruijn-based spatial coding. Moreover, dense reconstruction and absolute coding is assured with the proposed technique. Besides, other frequency multiplexing methods provide dense reconstruction for moving scenarios, but present high sensitivity to non-linearities of the camera reducing the accuracy and sensitivity to details in the surface, and can fail under presence of big

slopes. Among all the compared approaches, our method was only overcome by the time multiplexing shifting approach proposed by Pribanic et al. [9], which is only valid for static scenarios. Finally, 3D retrieval of some real objects was pursued to show the effectiveness of the algorithm in terms of density and perceived quality of the one-shot reconstruction.

5. A PROPOSAL ON ONE-SHOT ABSOLUTE PATTERN FOR DENSE RECONSTRUCTION



Figure 5.7: Proposed setup: a portable frame with the camera and the pico-projector attached on it.

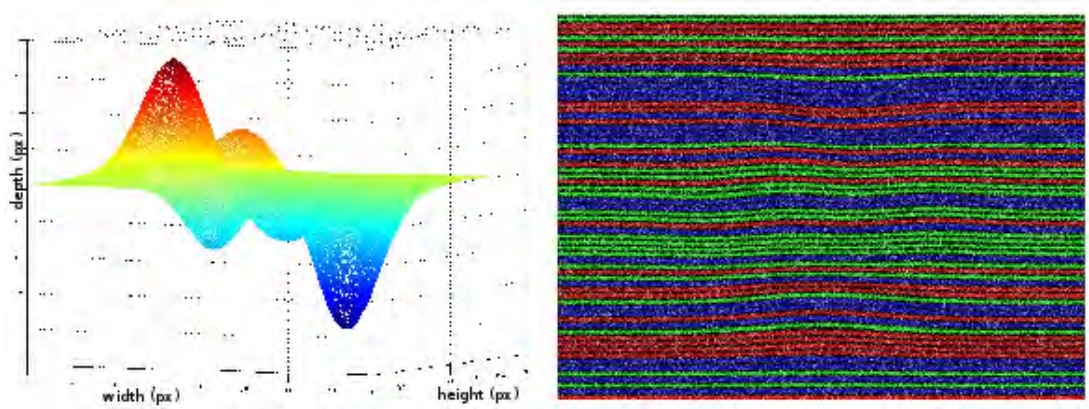


Figure 5.8: Peaks signal and recovered pattern for the proposed algorithm and noise of $std = 0.1$.

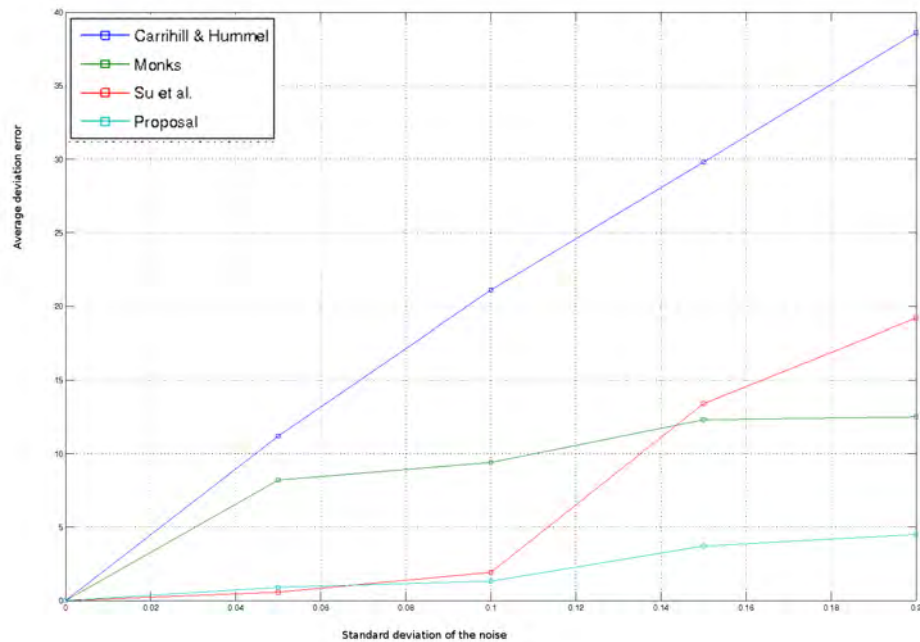


Figure 5.9: Normalized error on reconstructed depth positions, for different values of noise.

5. A PROPOSAL ON ONE-SHOT ABSOLUTE PATTERN FOR DENSE RECONSTRUCTION

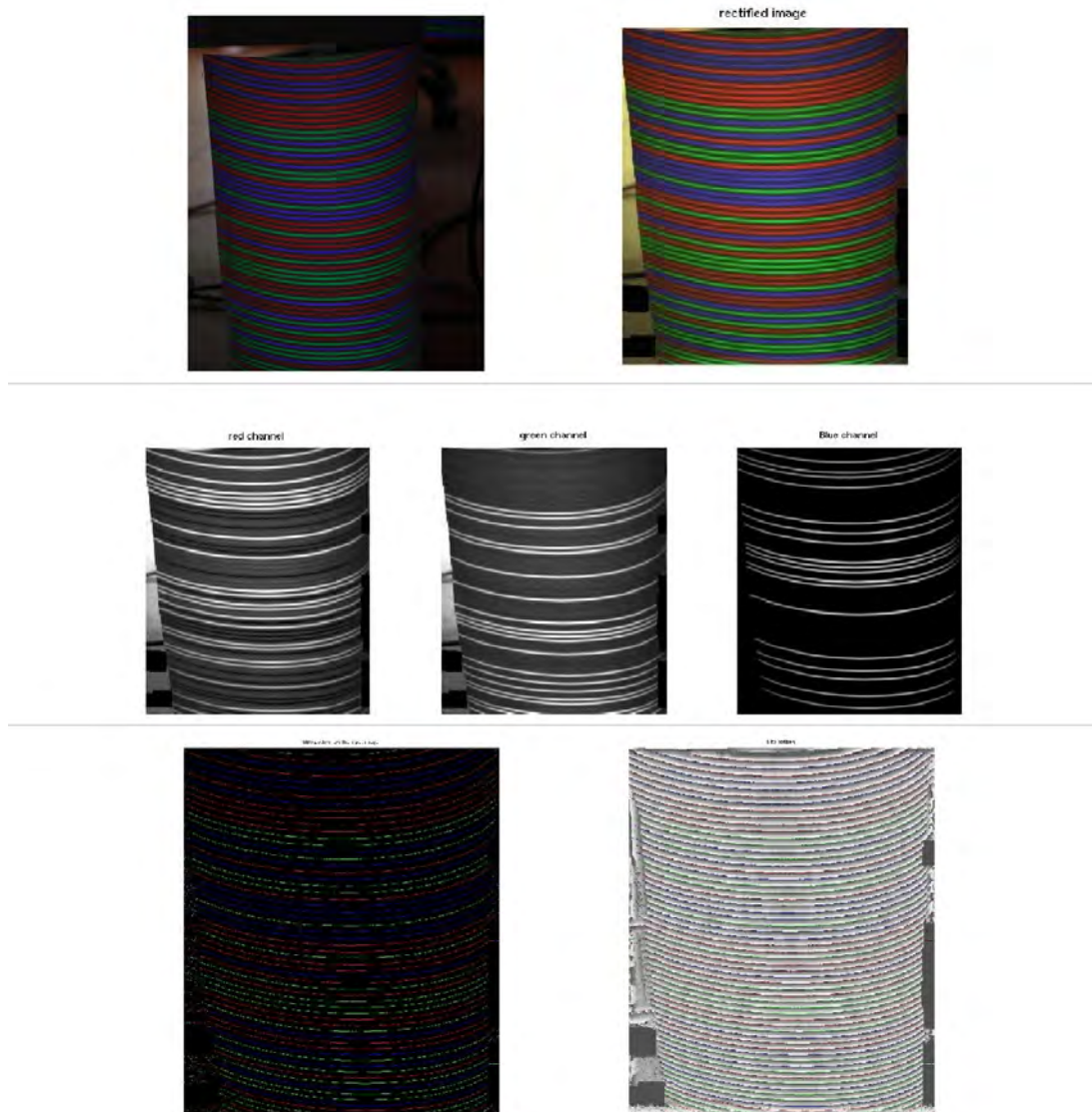


Figure 5.10: 3D retrieval of a bended piece of paper sheet. On the top, original and color rectified image. On the middle, RGB color channels before DeBruijn and WFT analysis. On the bottom, extracted DeBruijn color slits and WFT wrapped phase.

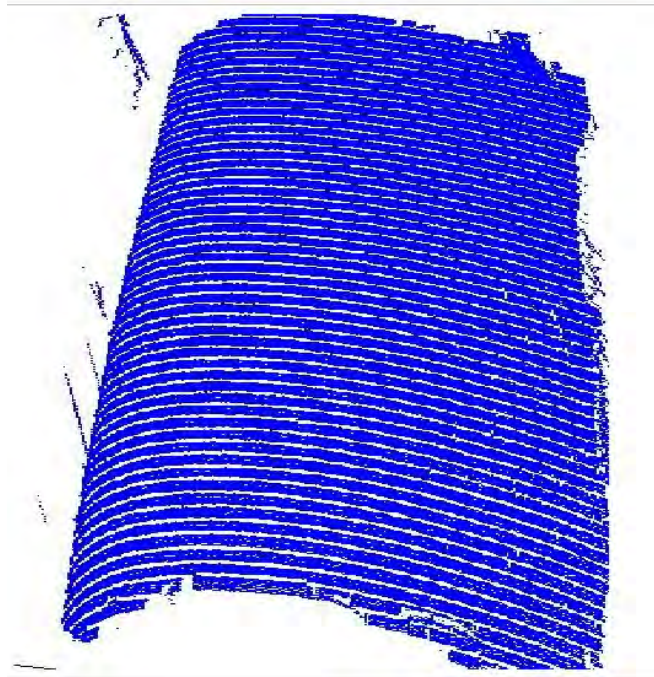


Figure 5.11: 3D cloud of points corresponding to the bended piece of paper sheet. 802768 points were reconstructed.

5. A PROPOSAL ON ONE-SHOT ABSOLUTE PATTERN FOR DENSE RECONSTRUCTION

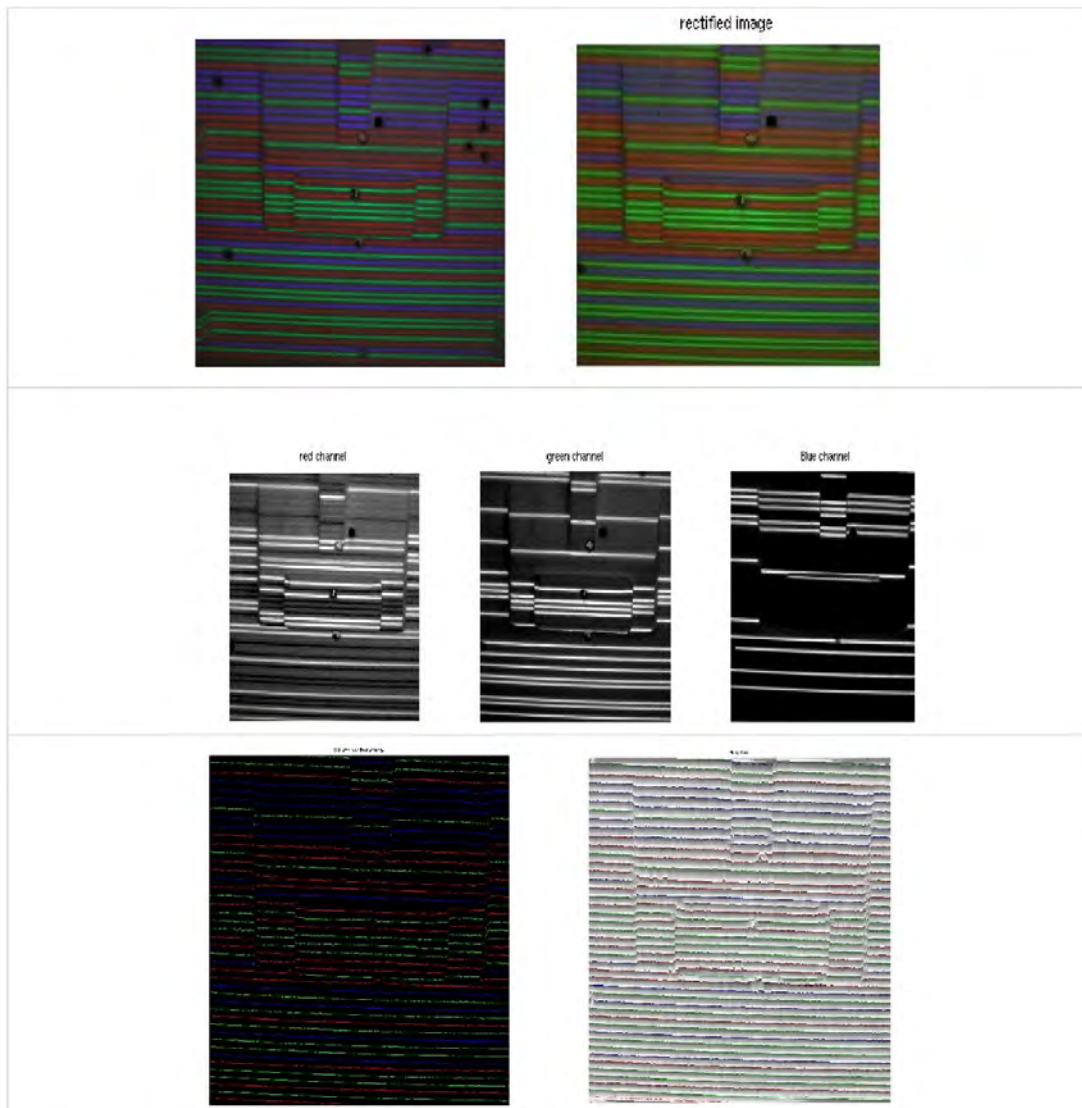


Figure 5.12: 3D retrieval of a manufactured piece of white cork. On the top, original and color rectified image. On the middle, RGB color channels before DeBruijn and WFT analysis. On the bottom, extracted DeBruijn color slits and WFT wrapped phase.

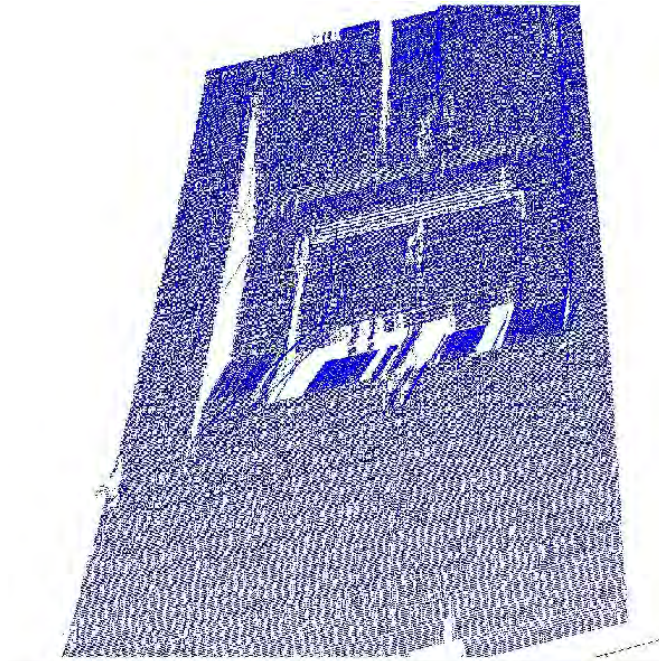


Figure 5.13: 3D cloud of points corresponding to the manufactured white cork. 620496 points were reconstructed.

5. A PROPOSAL ON ONE-SHOT ABSOLUTE PATTERN FOR DENSE RECONSTRUCTION

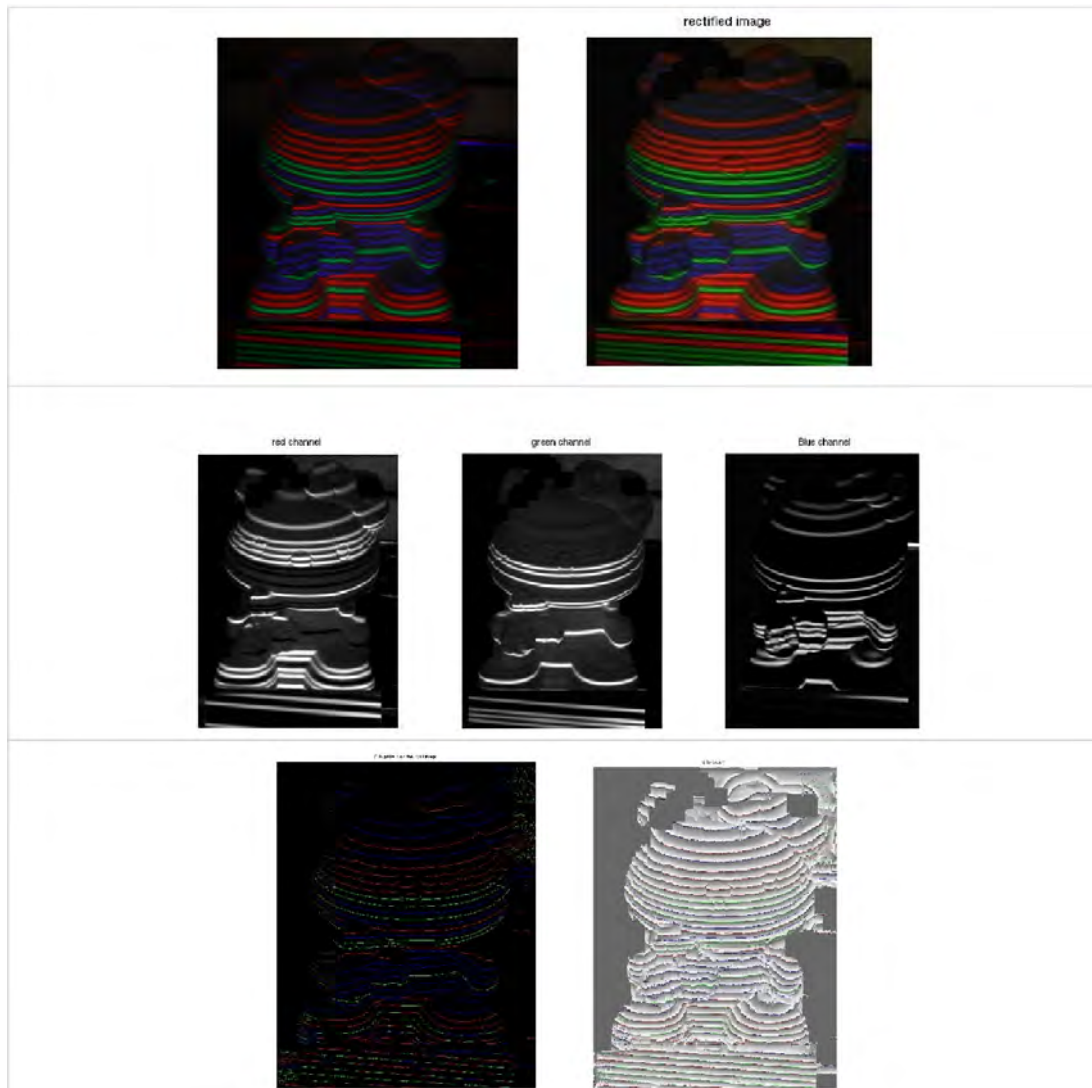


Figure 5.14: 3D retrieval of a ceramic figure of 'hello kitty'. On the top, original and color rectified image. On the middle, RGB color channels before DeBruijn and WFT analysis. On the bottom, extracted DeBruijn color slits and WFT wrapped phase.

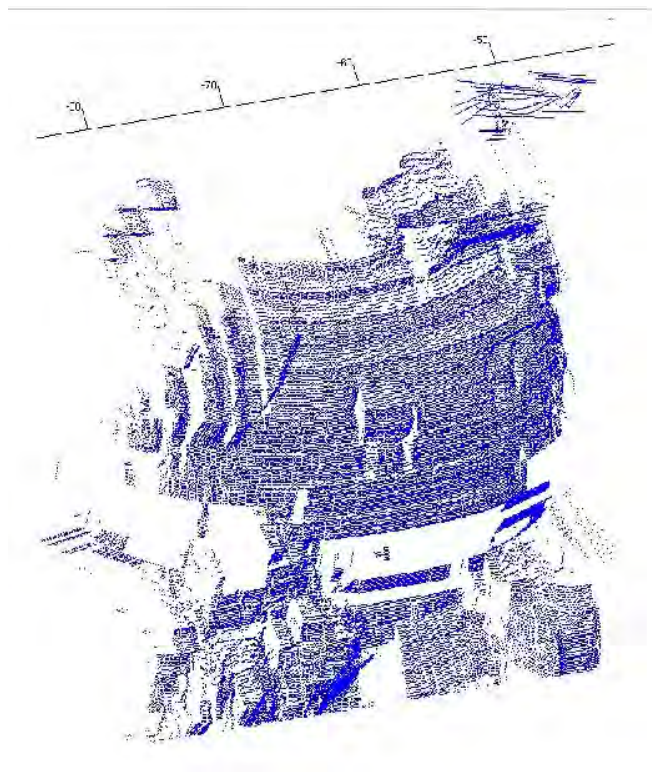


Figure 5.15: 3D cloud of points corresponding to the figure of 'hello kitty'. 666900 points were reconstructed.

5. A PROPOSAL ON ONE-SHOT ABSOLUTE PATTERN FOR DENSE RECONSTRUCTION

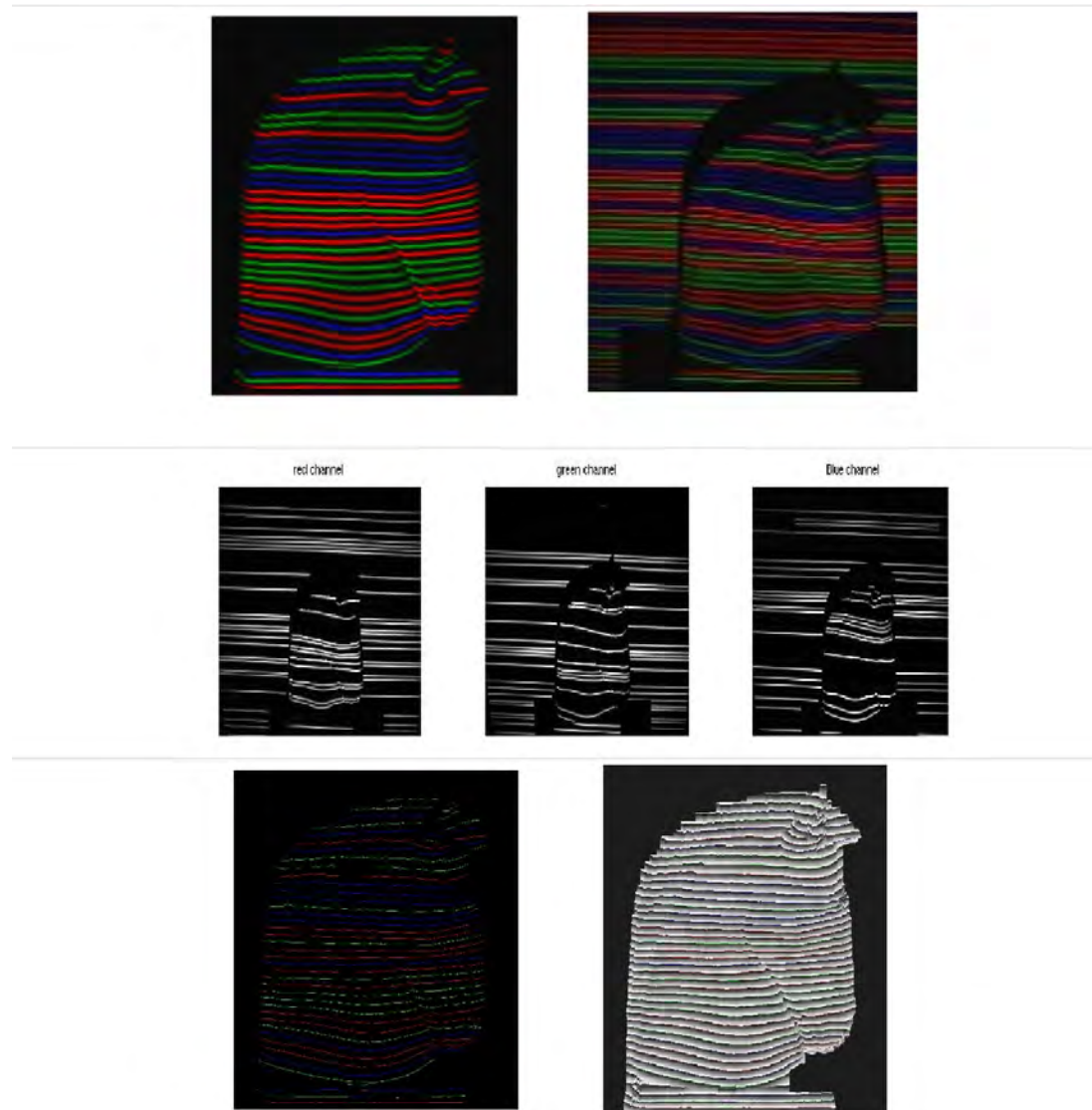


Figure 5.16: 3D retrieval of a ceramic sculpture of a horse. On the top, original and color rectified image. On the middle, RGB color channels before DeBruijn and WFT analysis. On the bottom, extracted DeBruijn color slits and WFT wrapped phase.

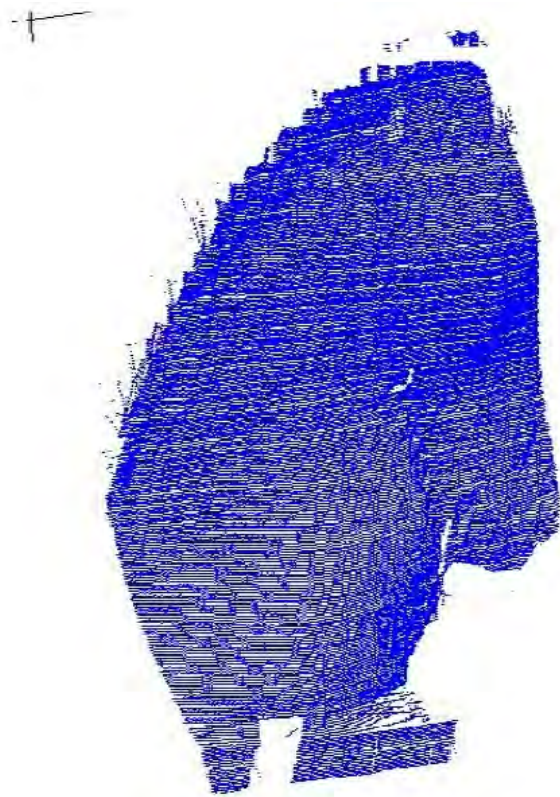


Figure 5.17: 3D cloud of points corresponding to the ceramic sculpture of a horse. 722127 points were reconstructed.

5. A PROPOSAL ON ONE-SHOT ABSOLUTE PATTERN FOR DENSE RECONSTRUCTION

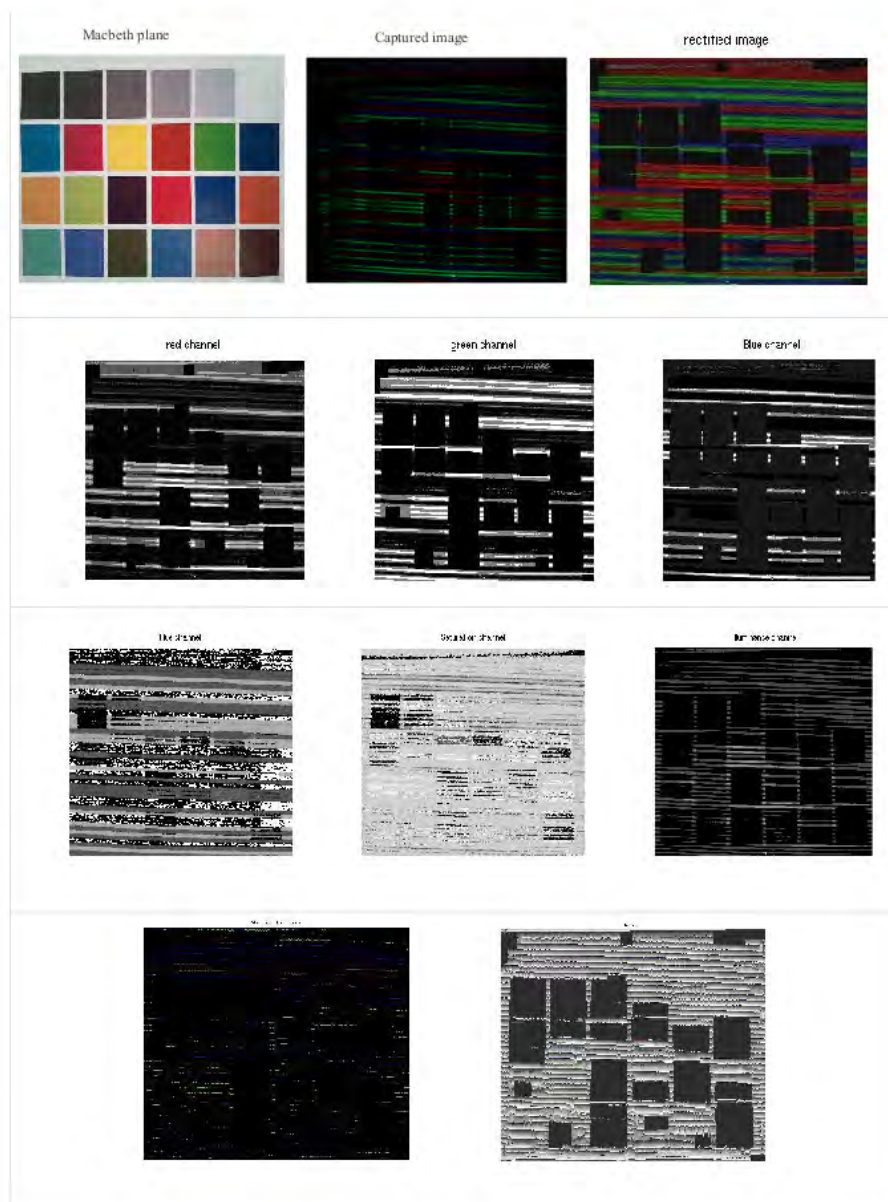


Figure 5.18: 3D retrieval of a Macbeth checker plane. On the top, original and color rectified image. On the middle, RGB and HSV channels before DeBruijn and WFT analysis. On the bottom, extracted DeBruijn color slits and WFT wrapped phase..

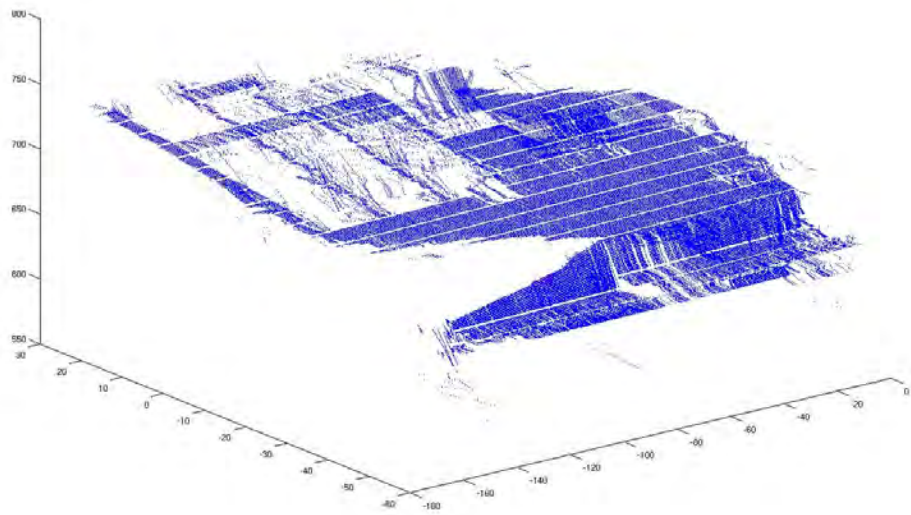


Figure 5.19: 3D cloud of points corresponding to the Macbeth checker plane.

5. A PROPOSAL ON ONE-SHOT ABSOLUTE PATTERN FOR DENSE RECONSTRUCTION

6

Registration of single-views from Structured Light

In this chapter we face the problem of registering; that is, the alignment of several 3D reconstructions in a common framework delivering a more complete 3D cloud of points of a given object. After studying the state of the art in 3D registration, we have implemented an optimal solution for structured light. Both global and local 3D registering are considered in the algorithm, in order to obtain the best results in terms of accuracy. Finally, tests are performed using real data and combining 3D views from single image captures.

6.1 Introduction

Registration of 3D images acquired by passive stereo-vision or structured light is an active field of research [94], [95], [96], [97], [98]). In particular, when two or more than two views are involved and initial pose estimations are unknown, three main inter-related sub-problems need to be solved: the overlapping regions between each view, their relative positions, and the absolute pose of the views. Regarding the pairwise registration (which in fact is a simplification of the multi-view case), the different approaches proposed in the literature can be broadly classified into two categories, depending on whether an initial information is required (fine or local registration) or not (coarse or global registration). Typically, the fine registration process consists in iterating the following two stages: firstly, the correspondence between every point from the two cloud of points should be found. These correspondences are used to define the residual error of the registration. Secondly, the best set of parameters that minimizes

6. REGISTRATION OF SINGLE-VIEWS FROM STRUCTURED LIGHT

the sum of these residuals should be found. These two stages are iteratively applied until convergence is reached. The Iterative Closest Point (ICP) [94] is one of the most widely used registration techniques using this two-stage scheme. Since then, several variations and improvements have been proposed in order to increase the efficiency and robustness of the method ([99], [100], [101], [102]). The main problem of ICP and other fine registration algorithms is that they can fail in a local minima in the relative pose estimation. To avoid this, a good approximation of the relative position must be given at the beginning of the iterative minimization algorithm. This is achieved using global registration, which attempts to find the rigid transform that optimally positions one set of data to another, until all partial shapes are registered. This problem is particularly hard when no information is available about the initial position of the model and data shapes, the inputs contain noise, and the shapes overlap only over parts of their extent (and the overlaps may not be known in advance). The solution adopted consist in extracting some feature keypoints of every 3D cloud of points, and perform the global registration using a matching algorithm based on minimization of the total Euclidean distance. This distance is computed as the accumulative distances from every pair of matched keypoints descriptors. Global registration provides a more robust result than fine registration, at expense of having more error in the estimation of the relative position of each view. Some of the global registration approaches existing in the literature were analyzed in this work. However, the nature of the 3D data provided by SL need of a specific registration pipeline. We propose an algorithm for global registration that is being developed by Umberto Castellani, Roberto Toldo, Maurizio Galassi and the aid of Sergio Fernandez during the stay at the University of Verona. This algorithm aims to be optimal for registration of noisy partial views like the ones acquired using SL. The reason is that a main focus is put on the selection and description of reliable 3D points, in order to avoid the effect of outliers, common in SL 3D results. The chapter is structured as follows: first, section 6.2 provides an overview of the global registration algorithms. Afterwards, the proposed registration pipeline is explained in section 6.3. Registration results of this technique are presented in section 6.4, and compared with those obtained with the 4PCS. Finally, section 6.5 discusses the implemented algorithm, pointing out the most important steps regarding its use in SL.

6.2 Brief overview of global registration algorithms

Global registration aims to align optimally two partial views of the same 3D shape, placed at arbitrary initial positions. Therefore, some rigid transformation must be

6.2 Brief overview of global registration algorithms

applied between the two views. This process is part of most 3D shape acquisition pipelines, where self occlusions and scanner limitations usually require the acquisition of multiple partial scans that overlap. To build a complete model, the partial scans need to be brought into a common coordinate system. This is usually done by pairwise registration, as the multiview registration can be obtained iteratively from pairwise registrations. The registration is particularly hard when the inputs contain noise and outliers, and the shapes overlap only over parts of their extent (and the overlaps may not be known in advance). As mentioned before, fine registration algorithms work iteratively computing a one-to-one rigid transformation for every pair of keypoints until a local minima (hopefully the absolute minima) is obtained. Contrary to this, global registration algorithms use the so-called voting methods, searching for the small number of parameters needed to specify the optimal motion. In order to compute this motion, distances between correspondences in different views are minimized. The most common correspondences are points, curves and surfaces. Global registration techniques can be classified on shape features or matching methods, which can be used jointly or separately. The first group searches for characteristics of points, using usually neighborhood information, in order to search for correspondences. Examples of this group are Point Signature [103] or Spin Image [104]. Matching methods are based on the process of matching points from both surfaces, using Ransac [105] or Genetic Algorithm [106]. A complete classification of techniques can be found in the work of Salvi et al. [107]. Regarding the process of matching, some proposals are described. The most common is based on the calculation of the residual. For each triplet of 3D locations or features in each view, the joint rigid transformation between the triplets is computed using an optimization algorithm. Afterwards, a residual is recorded, which tell us about the 'goodness' of the proposed transformation. This process is done for every combination of triplets in the two views. Finally, the entry with the best residual gives the optimal aligning transform. Several proposals can be found using this idea ([108],[109],[110]). Another variant of this scheme, the alignment method proposed by Huttenlocher and Ullman [111], counts for each transform proposed by two triplets of points how many points of the data are brought by the transform close to a point in the model. The transform which brings the most data points within a threshold of a point in the model is chosen as the optimal aligning transform. Voting methods provide the optimal alignment between the data and model shapes, and are independent of the initial pose of the input shapes. More recently, a method based of residual was proposed by Aiger et al. [112]. The so-called 4-Points Congruent Sets uses a total of four points, instead of three to compute the rigid transformation. The

6. REGISTRATION OF SINGLE-VIEWS FROM STRUCTURED LIGHT

method extracts all coplanar 4-points sets from a 3D point set that are approximately congruent, under rigid transformation, to a given set of coplanar 4-points. In order to align two point sets P and Q in arbitrary initial positions, a pair of triplets, one from P and one from Q, is enough to uniquely define a rigid transformation. For a base from P, the algorithm randomly selects a 3-points base from Q. Naively, there are $O(n^3)$ such candidate triplets from Q, where n is the number of points in Q. However, they use a set of 4 coplanar points from P as base B to find all subsets of 4-points from Q that are approximately congruent to B, in the sense that the two 4-points sets can be aligned, up to some allowed tolerance, using rigid transformation. This makes the problem easier than when working with only 3 points. This extraction procedure runs in roughly $O(n^2 + k)$ time, where n is the number of candidate points and k is the number of reported 4-points sets. Furthermore, the method reduces the number of trials required to establish a reliable registration between the underlying surfaces in the presence of noise, without any assumptions about starting alignment.

However, none of the global registering algorithms presented in these lines has been specifically designed for a SL set of data. A classical 3D cloud of points from SL present a high level of outliers and holes (an usual problem when dealing with discontinuities or occluded regions). Therefore, a robust pipeline able to select only the interest points and to describe its local region optimally reveals necessary. This would increase the robustness of the registration against errors in the recovered 3D points. Following this idea, a novel pipeline for global registration in SL is proposed hereafter. The algorithm is split in three main steps: detection of relevant keypoint, optimal description of these 3D points, and matching.

6.3 A novel pipeline for global registration

This section presents a novel method for feature-based global registration that combines a multi-scale based 3D saliency points detector, a mesh histogram of Gaussian (meshHOG) based feature descriptor, and a three to three keypoint based matching algorithm. The multi-scale saliency points detection algorithm, jointly with the polar representation used for the mesh HOG, provide robustness to the global registering, avoiding errors due to false similar regions in the object shape. This is an important issue when dealing with 3D points acquired using SL, as borders or discontinuities usually present a high level of noise. A diagram of the proposed pipeline including the three main steps (detection, description and matching) is shown in Fig. 6.1.

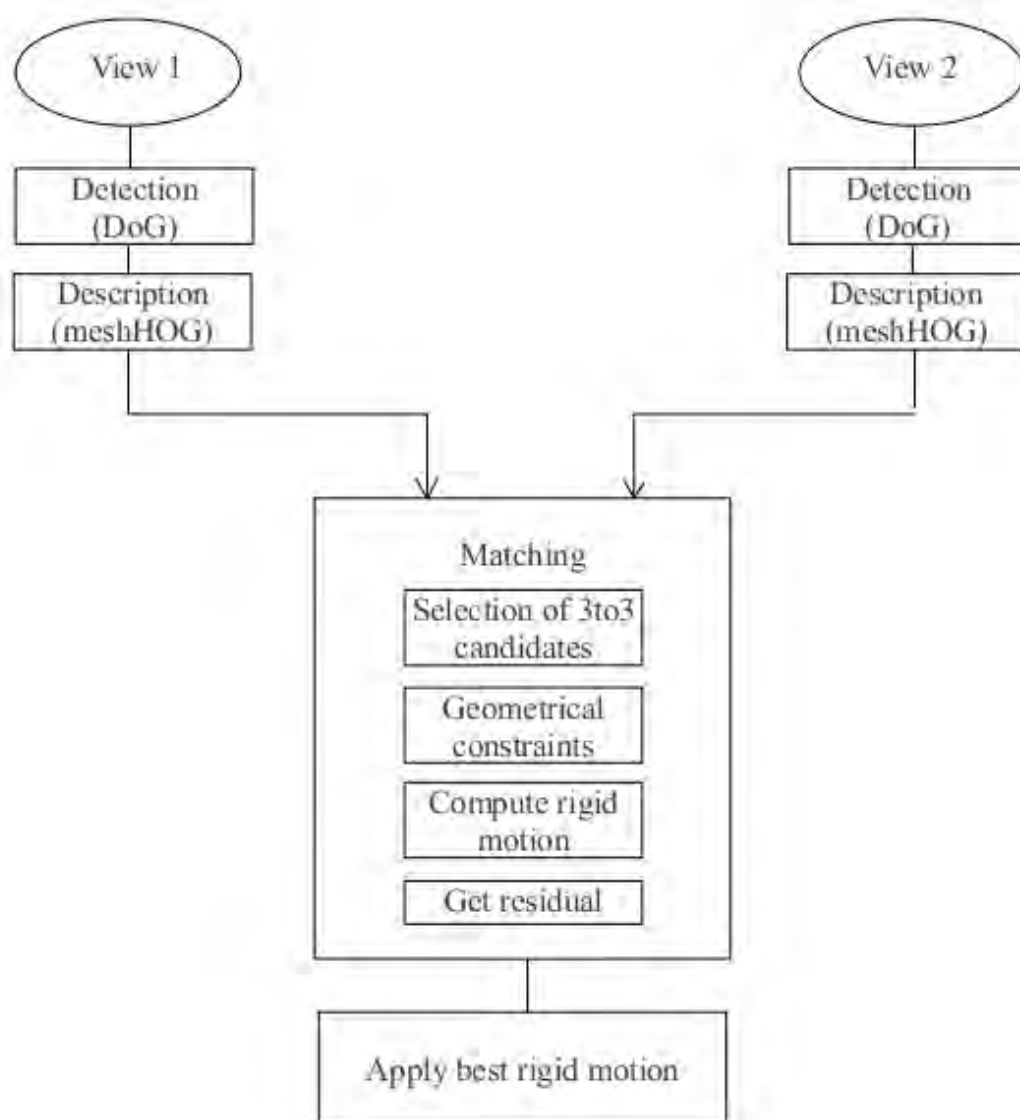


Figure 6.1: Diagram of the proposed registration algorithm.

6. REGISTRATION OF SINGLE-VIEWS FROM STRUCTURED LIGHT

6.3.1 Keypoint detection

The objective of keypoint detection is to extract some distinctive feature points from the 3D shape. These feature points are named saliency points, and will be used for the description of the surface. To this end, multi-scale representation, 3D saliency measure definition and keypoints selection is pursued in this step. The first step consist on computing the multi-scale representation of the surface. Given the definition of the Difference-of-Gaussian (eq.(6.1)):

$$F_i^d = g(v, \sigma_i) - g(v, 2\sigma_i) \quad (6.1)$$

where v is the analyzed vertex and σ_i is the value of the standard deviation associated to scale i , the multi-scale representation is obtained by applying a N multidimensional Gaussian filters $F_i^d, i = 1, \dots, N$, up to a distance equal to 2.5σ . Six scales of filtering have been fixed, corresponding to standard deviation of values $\sigma_i \in 1\gamma, 2\gamma, 3\gamma, 4\gamma, 5\gamma, 6\gamma$ where γ amounts to 0.1% of the main diagonal of the surface. F_i^d can be taken as a saliency feature after reducing it to the scalar quantity projected to the normal $n(v)$ of the vertex v . This value, named the scale map M_i^d , is obtained as in eq.(6.2). Finally, the saliency map is computed by simply adding the contribution of each scale map to vertex v . A saliency map is computed for every vertex in the surface. The last step is to determine which vertex are relevant enough; that is, to extract the saliency points. The saliency points are obtained as maxims of the saliency map, considering those values higher than the 30% of the global maximum.

$$M_i^d(v) = ||(n(v) \cdot (g(v, \sigma_i) - g(v, k\sigma_i)))|| \quad (6.2)$$

6.3.2 Keypoint description

The next step is focused on building an appropriate keypoint descriptor. The first, classical approach implemented was the spin-image. Firstly proposed by Johnson and Hebert [104], this feature descriptor is rotation invariant, therefore can be used for global registering. Given an oriented point $(v, n(v))$, a spin map S_0 is defined as the function that projects 3D points x to the local coordinate system defined by (v, n) (eq. 6.3):

$$S_v(x) \rightarrow (\alpha, \beta) = (\sqrt{(|x - v|^2 - (n \cdot (x - v))^2)}, n \cdot (x - v)) \quad (6.3)$$

6.3 A novel pipeline for global registration

A spin-image for point v is generated by applying the spin-map to all the points x of a region around the surface and then accumulating the results in a discretized (α, β) space. If surfaces are uniformly sampled, the spin-images of two corresponding points on different instances of the same object will be similar. Therefore, the corresponding Euclidean distance will be small, being this pair of points good candidates for matching. However, spin images present ambiguity in the position of the evaluated point x with respect to v (α takes only positive values). Moreover, geometrical or texture weighting is not able when using this technique. These are the reasons why a proposal for a keypoint descriptor based on the mesh Histogram of Gaussian (meshHOG) was implemented afterwards. Originally proposed by Zaharescu et al. [13] in 2009, the mesh Histogram of Gaussian is based on an histogram gradient which is computed for a specific region around the considered 3D point. The first step is to define the neighborhood region for computing the meshHOG. To this end, we consider the n -ring around the considered pixel v_i , defined as the group of vertex placed at a distance of N points to v_i . Then, we accumulate the N -ring points until we reach the 1% of the total surface. Afterwards, we compute the Discrete Gradient for the selected points. The discrete gradient $\Delta_S f(v_i)$ of f at $v_i \in S$ is defined from eq. and eq.:

$$\Delta_S f(v_i) = \sum_{v_j \in \text{erg}(v_i, 1)} (\omega_{ij} \cdot D_{\vec{e}_{ij}} f(v_i)) \vec{u}_{ij} \quad (6.4)$$

where $D_{\vec{e}_{ij}} f(v_i)$ is the directional derivative of f at v_i $\forall i \neq j$. ω_{ij} weights the contribution of $D_{\vec{e}_{ij}} f(v_i)$ and \vec{u}_{ij} is the normalized projected direction of $v_i v_j$ in the tangent plane at v_i . The weights ω_{ij} are chosen in order to balance the contributions of the local directional derivatives with respect to their associated directions in the tangent plane. The next step is to compute the local coordinate system. The first direction of the local coordinates is given by the normal n_{v_i} of the point. The second direction, defined as a_{v_i} , is provided by the direction associated to the dominant bin in a polar histogram of $b = 36$ bins. This polar histogram is computed from the projection of the considered 3D points magnitudes $|\Delta_S f(v_i)|$ onto the plane defined by n_{v_i} . Finally, the third direction is given by $n_{v_i} \times a_{v_i}$. Once the local coordinate system has been defined, the HOG descriptor is created from the concatenation of the three different 2D histogram values, where each 2D histogram is computed from the projection of 3D points magnitudes $|\Delta_S f(v_i)|$ onto one of the corresponding planes defined by the local coordinates system. For every plane, a 2D histogram with a total of $b_s \times b_o$ bins is provided, being $b_s = 4$ a previous sub-region division made on the plane (see Fig. 6.2). A modification of the standard approach has been done considering the weights ω_{ij} as

6. REGISTRATION OF SINGLE-VIEWS FROM STRUCTURED LIGHT

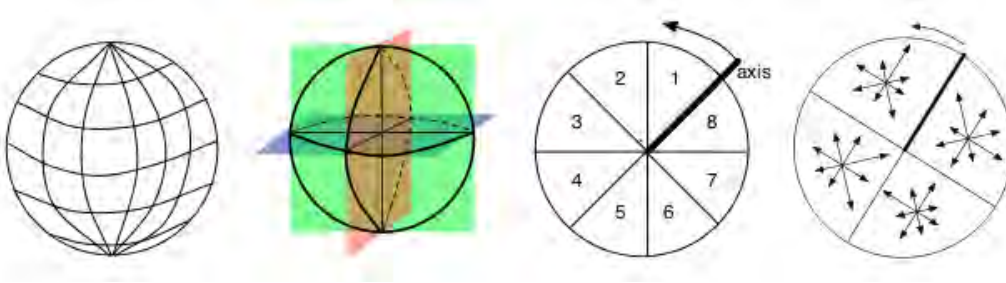


Figure 6.2: Creation of the 2D histograms from the local planes, $b_s \times b_o$ and bins (from [13]).

the saliency values provided by the saliency map, for every 3D point under consideration. This gives us a weighting strictly based on the geometrical properties of the mesh. Other approaches consider features linked to color or texture, but is out of the scope of 3D registering for SL. Moreover, the contribution of the 3D points magnitudes to the total count is weighted by their distance to the point n_{v_i} . The contribution is accumulated in the 2D histogram following a bilinear interpolation between neighboring bins, in order to reduce the aliasing and boundary effects.

6.3.3 Matching

Having the descriptors corresponding to two different views of a single image, the objective is to find matching keypoints with some overlap. We follow the approach of Brown and Lowe [113] for 2D image mosaicking.

6.3.3.1 Selection of the keypoint candidates for matching

First, we compute the correlation matrix between the set of keypoint descriptors from the two views. Direct Euclidean distance between each pair of descriptor is computed (zero-shifting in the correlation). A keypoint similarity matrix is obtained as result. Afterwards, we binarize this matrix selecting only the six greatest matches for every row; that is, each descriptor of the first image is candidate to its $L = 6$ nearest neighbors of the second image, in the feature space. For multiview matching, every view would be matched with the m views that have the greatest values in the 2D histogram, therefore a m -dimensional similarity matrix would be used. Afterwards, we select for every combination of three keypoints (triplet) from the first image every possible combination of three keypoints (triplet) from the second image. This is done regarding the binarized similarity matrix, and taking into account that every keypoint can be matched only

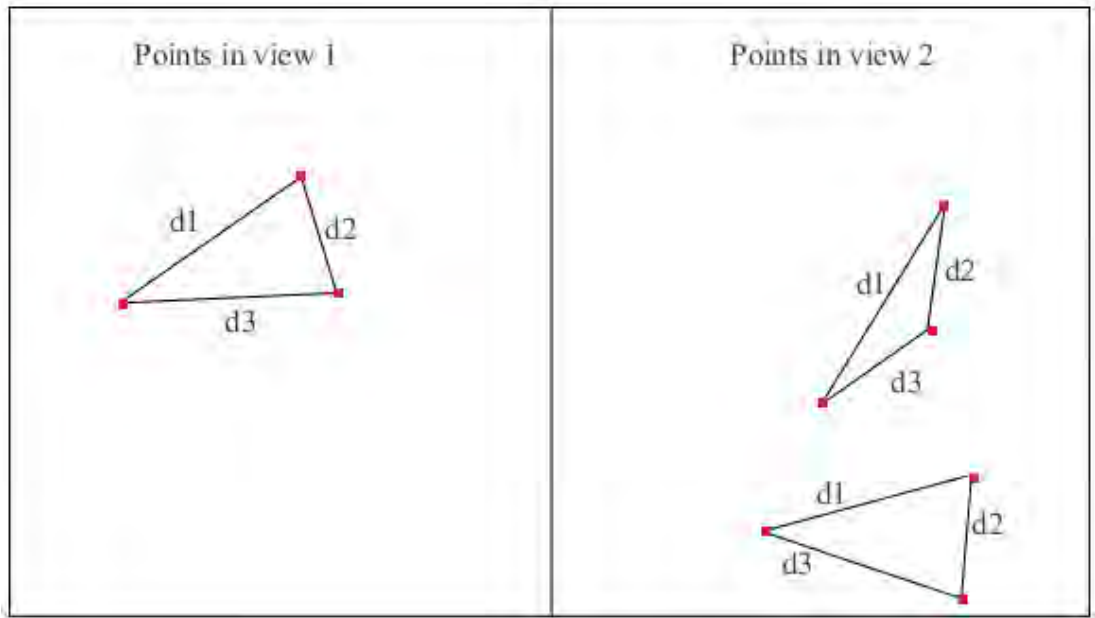


Figure 6.3: 2D representation of triplets in view 1 and view 2. As can be seen, the relative distances within the keypoints of both images are not related for the triplet on the top of second image, but will allow a match for the triplet on the bottom.

once. This triplet to triplet set is called *sextet*.

6.3.3.2 Geometrical constraints

There are two geometrical constraints that must be applied to the sextets. The first one refers to the relative distances within the two triplets. That is, the distances between the three keypoint positions of the first image must be similar to those for the second image. Otherwise the triplets do not have similar spatial distribution, so they do not correspond to the same region. The second constraint relevance of a given triplet. It has been proven that a triplet gives relevant information if at least one of its vertex is farther from the other two vertex in more than the average vertex distance of the mesh. Therefore, this constraint is also applied when selecting the triplets. A graphical representation of the constraints can be observed in Fig. 6.3.

6.3.3.3 Compute the rigid motion

Once the sextet is determined and the constraints have been applied, the rigid motion between the two views is computed. We are trying to find the rotation matrix R and

6. REGISTRATION OF SINGLE-VIEWS FROM STRUCTURED LIGHT

the translation vector t that minimizes the eq.(6.5).

$$\sum_{i=1}^3 \omega_i \cdot \|R \cdot v_i^a + t - v_i^b\| \quad (6.5)$$

where ω_i is a weighting vector (ones in our case), and v_i^a and v_j^b , $i, j = 1, 2, 3$ correspond to the triplet of the first and the second view, respectively. The first step is to compute the weighted centroids of the triplet (eq.(6.6) and eq.(6.7)).

$$v_c^a = \frac{\sum_{i=1}^3 \omega_i \cdot v_i^a}{\sum_{i=1}^3 \omega_i} \quad (6.6)$$

$$v_c^b = \frac{\sum_{i=1}^3 \omega_i \cdot v_i^b}{\sum_{i=1}^3 \omega_i} \quad (6.7)$$

From this, the centered vectors are computed as $x_i := v_i^a - v_c^a$, $y_i := v_i^b - v_c^b$. We create the $d \times d$ covariance matrix $S = X \cdot W \cdot Y^T$, where X and Y are the $d \times n$ matrices that have x_i and y_i as their columns, respectively, and $W = \text{diag}(\omega_1, \omega_2, \dots, \omega_n)$.

The final step is to apply Singular Value Decomposition (SVD) to S , such that $S := U \cdot \Sigma \cdot V^T$. The rotation matrix R and translation vector t are obtained from eq.(6.8) and eq.(6.9), respectively.

$$R = V \cdot [\cdot \cdot \cdot \det(VU^T)] \cdot U^T \quad (6.8)$$

$$t = v_c^b - R \cdot v_c^a \quad (6.9)$$

The details about the mathematical background can be found in the work of Sorkine [114]. Finally, the 3D cloud of points corresponding to the first view are transformed by R and t

6.3.3.4 Calculate the residual

It is necessary to compute a measure of the goodness of the proposed rigid transformation, in order to choose the sextet providing the best transformation between the two views. This is computed regarding the Euclidean distance between the transformed cloud of points, and minimizing this parameter. To this end, the computation of the residual reveals necessary. Technically speaking, the residual computes partial distances between pair of 3D points from the two views. For any 3D point in the transformed first view, we look for the closest point in the second view, and compute its Euclidean

distance. This is done for every pair of 3D points not farther than a pre-set threshold. The residual is taken as the sum of the partial distances. Therefore, the best rigid transformation is selected as the one minimizing the residual and maximizing the percentage of point matchings between the two views. It is important to note that, due to the big amount of 3D points of the surfaces, usually some data reduction is applied. In this approach only the 10% of the total number of points in the first view are considered to compute the residual. These points are selected randomly, therefore they are expected to represent the whole surface, thus providing similar residual at the same time the computational cost is decreased. Finally, a variation of the ICP fine registration called Levenberg-Marquardt-ICP (Fitzgibbon [115]) is applied in order to refine the registration provided by the proposed global registration.

6.4 Results

The proposed registration algorithm has been tested for a set of images corresponding to a given database of closed 3D objects, as well as for reconstructed 3D views from SL.

6.4.1 Results using a closed-form synthetic object

First, the proposed detector and descriptor have been tested on a closed 3D object, the Stanford *Bunny* model (available at <http://graphics.stanford.edu/data/3Dscanrep/>), for which 24 partial views are provided. The results are shown in Fig. 6.4. On the top are represented all the keypoints of views 1 and 2. On the middle, the corresponding $n \times m$ similarity matrix A and its binarized version with only 6 candidates for every row. On the bottom, the considered 3D keypoint (on the left) and their corresponding $L = 6$ nearest neighbors in the feature space of view 2 (on the right). As can be observed, the selected keypoint in image 1 is considered as a potential matching in image 2, provided the preliminary adjacency matrix. Afterwards, the matching algorithm is applied to compute the optimal rigid transform that aligns the views. The results of registration are shown in Fig. 6.5.

As can be observed, the global registration provides optimal alignment of the two views, without visual matching errors. Regarding the quantitative results, the residual obtained in the previous registration is $res = 45.6434$. This value corresponds to the sum of distances between matched pairs of points, after applying the rigid transformation. The percentage of matched points from the first view is $perc = 0.9589$, for a total

6. REGISTRATION OF SINGLE-VIEWS FROM STRUCTURED LIGHT

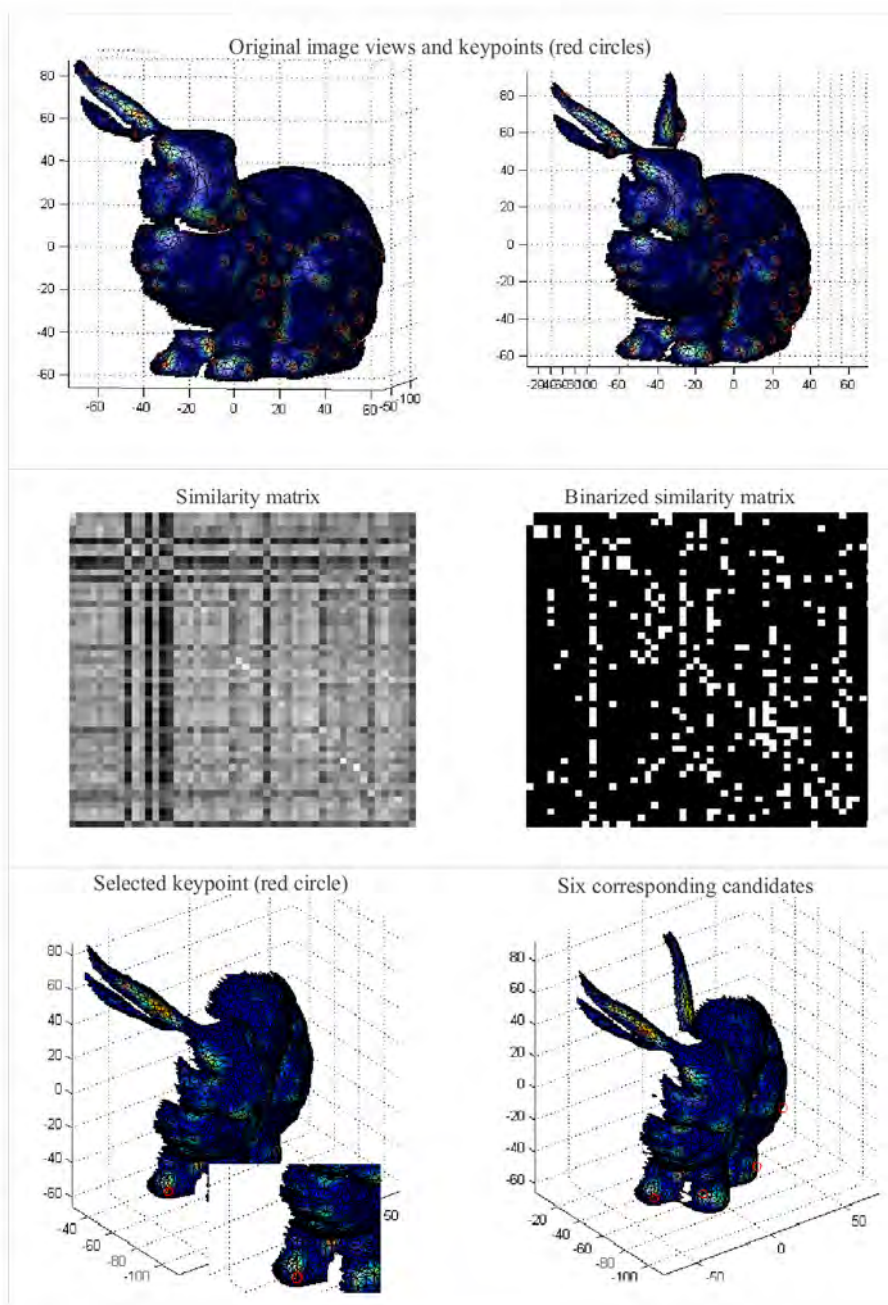


Figure 6.4: Results of the proposed detection + description. Top: input image view. Middle: similarity and binarized similarity matrices. Bottom: potential keypoint candidates for matching.

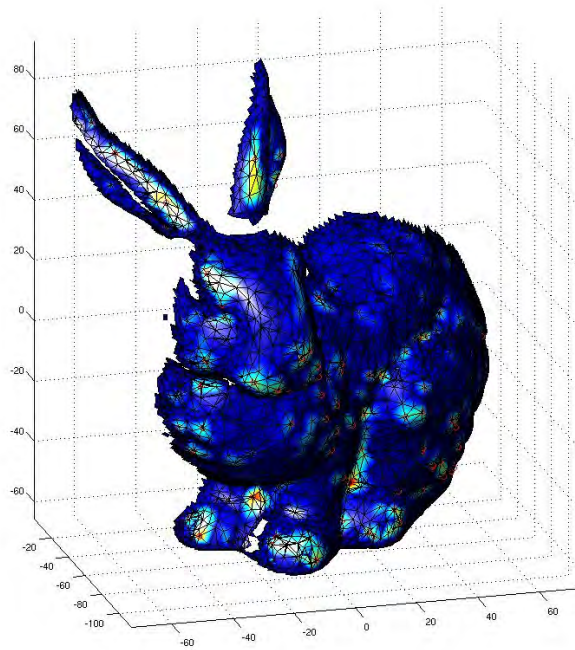


Figure 6.5: Registration results for the object *bunny*.

6. REGISTRATION OF SINGLE-VIEWS FROM STRUCTURED LIGHT

of $n_v = 2920$ points in the first view. Therefore, the average distance between two matched points after registration is (eq.(6.10)):

$$dist_{av} = \frac{res}{perc \cdot n_v} \quad (6.10)$$

This gives a value of $dist_{av} = 0.0163$. Considering that the average Euclidean distance between neighboring points in the mesh is $d = 1.7464$, we can say that the rigid registration is optimal, being the error lower than the 1% of the distances in the mesh.

6.4.2 Results of registration of SL partial views

Finally, the complete registration pipeline was applied to a set of partial views extracted from the previous SL acquisition technique. A ceramic sculpture of a horse was reconstructed from two non-centered views taken at a distance of around 80cm from the camera. Registration results can be observed in Fig. 6.6. We can see that the registration did not work properly. This is mainly caused by the effect of borders of the reconstructed shape, rather than the outliers present within the 3D cloud of points. Therefore, the next step will be to include some constraint in the detection step that avoid the selection of these positions as keypoints. This would allow the description of only inner positions, permitting a comparable representation of the two views for matching.

6.5 Conclusions

In this chapter we face the problem of 3D registration of partial views from SL. For the use in SL, where the initial relative position of the different views is unknown, a global registering prior to perform any fine registration algorithm reveals necessary. Regarding the global registration techniques, we first studied the state of the art, focusing on the so-called 4PCS algorithm developed by Aiger et al. [112]. However, in order to increase the robustness of the registration against errors in the recovered 3D points (an usual problem in SL when dealing with discontinuities or occluded regions), we proposed a new pipeline for global registration. Three main steps are pursued: detection, description and matching. Regarding the detection step, a multi-scale DoG was applied to the surface, in order to obtain a saliency measure for every 3D point. Afterwards, we choose the points having a greater value of saliency as the keypoints for that surface. Afterwards, a descriptor is computed to every keypoint. We implemented a mesh HOG descriptor with adapted weighting. This weight is adapted to

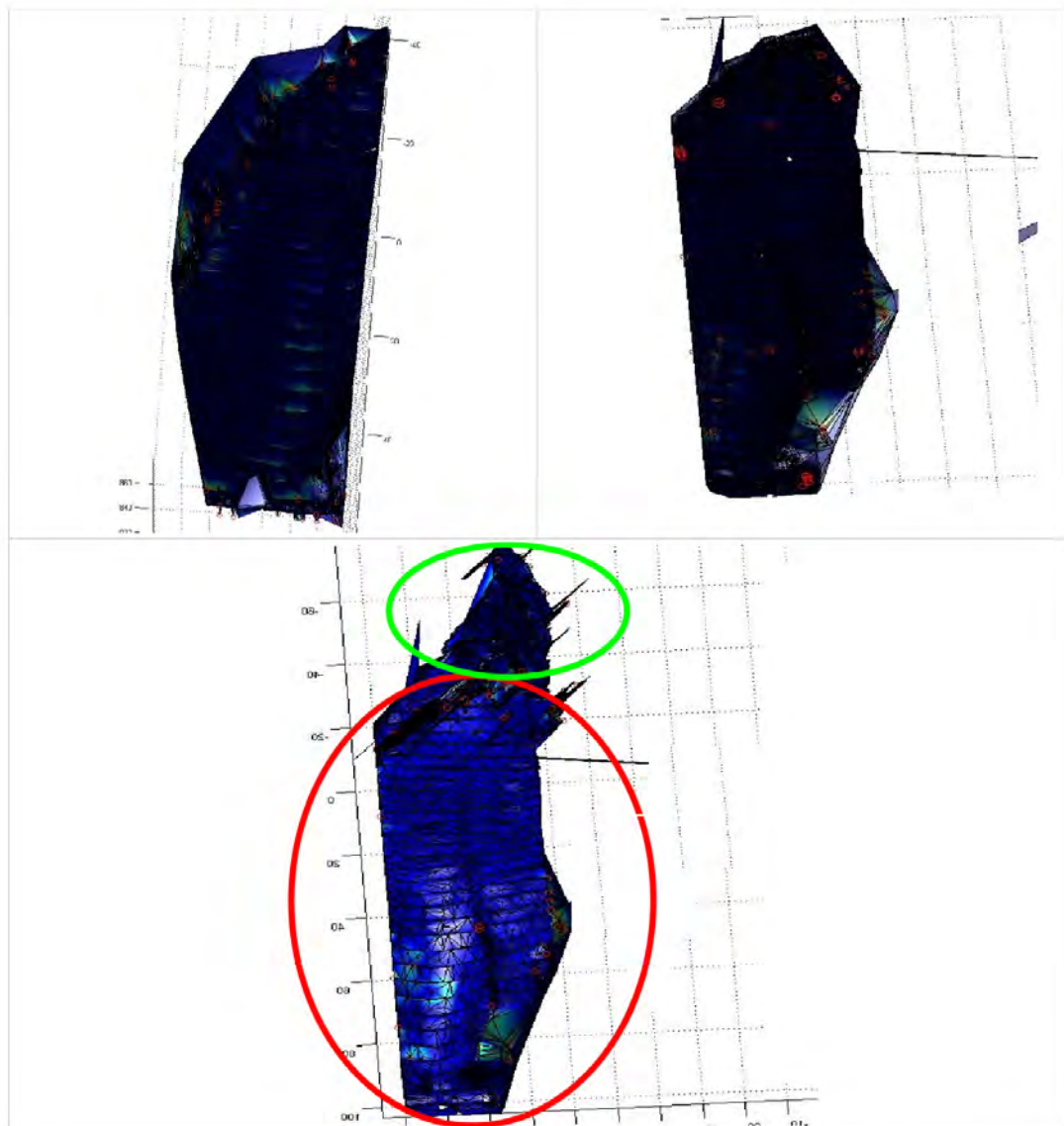


Figure 6.6: Registration results for the two partial views of a ceramic sculpture of a horse. On the top, partial views and selected keypoints (red points). On the bottom, results of registration. The green circle marks the first view, while the red circle marks the second view.

6. REGISTRATION OF SINGLE-VIEWS FROM STRUCTURED LIGHT

the saliency of the single points corresponding to the selected sub-region around every point. The weighted meshHOG provides information about the structure of the mesh, and is a good descriptor of the keypoint. Finally, a matching procedure based on Euclidean distance was applied. This algorithm calculates the optimal rigid motion between every pair of triplets from the two views. The best transformation is chosen regarding the residual value, given as the sum of Euclidean distances between every pair of matched 3D points. The proposed algorithm was first tested on a closed object (a *bunny* figure), showing the potential keypoints for matching, the optimal performance of the proposed descriptor and the final global registration. The complete pipeline was tested also under real 3D reconstruction. A ceramic sculpture of a horse was used for registering. However, due to the selection of keypoints at the corner of the partial views, the registration failed. This is an issue to amend as future work.

7

Conclusions

In this final chapter a summary of the thesis and its main contributions are presented. Future directions connected to this work are also pointed out. The chapter ends with a summary of publications and remarks related to this thesis.

7.1 Discussion

In this thesis the problem of dense 3D reconstruction using one-shot structured light has been discussed. First, **chapter 1** introduces the concepts of stereo-vision and structured light, and its importance in computer vision. Moreover, a statement with the objectives of the thesis is presented, as well as the context where it has been developed. In **chapter 2**, an up-to-date review and a new classification of the different techniques existing in structured light have been proposed. The classification was done regarding the sparse or dense 3D reconstruction of the imaged scene. Among them, a sub-classification regarding the spatial, frequency or time multiplexing strategy was done. A selection and implementation of representative techniques of every group has been done, and qualitative and quantitative comparisons have been performed extracting advantages and drawbacks of each technique. The results show the good performance of shifting approaches, which obtain dense reconstruction and excellent accuracy. However, they are only valid for static scenarios. Among one-shot techniques able to work in moving scenarios, frequency multiplexing methods achieve dense reconstruction. However, they present high sensitivity to non linearities of the camera reducing the accuracy and sensitivity to details in the surface and can fail under presence of big slopes. Spatial multiplexing can also work in moving scenarios. Among them, De Bruijn codes provides the best results in terms of accuracy, at the expense of having

7. CONCLUSIONS

sparse reconstruction.

Chapter 3 proposes a first approach of 3D dense reconstruction using one-shot projection. The principle is based on frequency analysis and color multiplexing of three different frequency channels. As was extracted from Chapter 2, in frequency based techniques the periodicity of the pattern imposes the assumption of smooth surfaces, as the presence of slopes would yield to some acquisition errors. This problem is overcome in Multiple Phase Shifting approaches, which create an absolute phase map from two relative phase maps, according to the ideas proposed by [54]. The proposed algorithm performs a multiplexing in color space of the three different fringe patterns. The phase of these patterns is extracted using wavelet decomposition combined with a cost function algorithm. An absolute coding based algorithm merges the individual relative phase maps to create the absolute phase map and extract the depth deviation. Finally, a consistence mapping is pursued in order to avoid errors created by a bad detection of slopes and error propagation between pixels during the unwrapping step. The simulations performed optimally using the standard benchmark without and with added noise. However, some problems were found in the experimental results. The algorithm fails under presence of slopes, due to the effect of the dyadic net intrinsic to the wavelet implementation, which does not consider little frequency variations when setting the window size. To solve this problem, an algorithm for the adaptive selection of the best analysis mother signal, in terms of frequency and shape, is proposed in **chapter 4**. First, a comparative study of WT and WFT was done. The results show how WFT with automatic window size detection performs better compared to the traditional WT technique. This is mainly due to the nature of the dyadic net used in WT, which scales the window by a factor of 2^j between adjacent size values. Afterwards, an automatic window width selection algorithm for WFT was proposed, which performs a fine tune of the window size between a set of values around the optimal size corresponding to the global frequency average. Some simulated and real results have been obtained, proving that the Morlet and the Paul wavelets show better performance than the Shannon and the Spline wavelets, thanks to the greater similarity of its signal shape to the sinusoidal nature of the fringe pattern. A proposal of a novel one-shot dense reconstruction algorithm is presented in **chapter 5**. The idea is to combine the accuracy of DeBruijn spatial multiplexing with the density of frequency multiplexing in fringe projection. The proposal was implemented and compared both quantitatively and qualitatively with some representative techniques of Structured Light. Simulation results and empirical quantitative results showed the good performance of the proposed technique in terms of resistance to noise and accuracy of a reconstructed plane. The proposed method

was only overcome by the time multiplexing shifting approach proposed by Pribanic et al. [9], which is only valid for static scenarios. Among one-shot techniques, our proposed method achieves the best results in terms of accuracy, comparable with other DeBruijn-based spatial coding. Moreover, dense reconstruction and absolute coding is assured with the proposed technique. Besides, other frequency multiplexing methods provide dense reconstruction for moving scenarios, but present high sensitivity to non-linearities of the camera reducing the accuracy and sensitivity to details in the surface, and can fail under presence of big slopes. Finally, 3D reconstruction of some real objects was pursued to show the effectiveness of the algorithm in terms of density and perceived quality of the one-shot reconstruction.

Most of the works presented in SL during last years have been concerned into frequency multiplexing approaches, trying to increase the robustness in the decoding step and the resistance to slopes under the constraint of moving scenarios ([59], [68]). Under this scenario, the proposal made in this work of merging DeBruijn and frequency-based one-shot patterns achieves a dense reconstruction with the robustness in the decoding step provided by frequency analysis, jointly with the accuracy given by spatial DeBruijn-based patterns. This combination gives us a one-shot absolute dense pattern with the highest accuracy achievable for moving scenarios.

The last chapter was about registration of 3D reconstruction obtained from single captures. This is a complementary step in SL, which becomes necessary when reconstructing large surfaces where the required level of detail imposes a careful reconstruction of every region. First, the differences between global registering and fine registering is explained, and the pros and cons are pointed out. Afterwards, an available software for registration called 4PCS has been tested, showing that registration is possible with the output provided by the proposed algorithm. Moreover, the chapter explains a new registration algorithm that is being developed currently. This is a global registration pipeline, where the matching is done using some saliency map and matching some feature descriptors. Some work has been done in the development and comparison of different feature descriptor algorithm, in order to select the most suitable for registration in SL.

Finally, some appendix are provide to clarify specific parts of the proposed algorithms. Appendix A presents a new algorithm for geometric calibration of the projector-camera system, based on a plane-based calibration structured light projector model. The proposed method makes use of the Bouguet's camera calibration toolbox which implements Zhang's calibration [116]. This technique has been also used to implement the projector calibration, as in the plane structured light model the projector is re-

7. CONCLUSIONS

garded as the inverse of a camera. Linear and non-linear distortion is considered for the calibration of both devices. Although we used the planar target and the nonlinear projector model, the idea is also suitable for 3-D reference objects and other projector models. Some simulations and real calibration experiments were pursued, proving that the proposed technique has similar precision for the projector calibration to the one obtained for camera calibration using DeBouguet algorithm. Afterwards, appendix B presents a method for color calibration and another for RGB channel misalignment calibration. Both techniques are required for the colored pattern projection that are presented afterwards as proposed 3D reconstruction algorithm. Appendix C explains the principle of the Remainder theorem used for the unwrapping algorithm employed in the first SL pattern proposal. Finally, appendix D gives some mathematical explanation of the dynamic programming algorithm employed in DeBruijn pattern decoding.

7.2 Contributions

The main contributions of this thesis are:

- A new state of the art and a new classification of the SL approaches present in the literature. A main classification is done regarding the sparse or dense 3D retrieval obtained. Afterwards, a low level classification is done regarding the spatial, time or frequency multiplexing of the technique.
- A one-shot dense reconstruction algorithm based on fringe pattern and the principle of the remainder theorem. This method obtains optimal results for 3D smooth surfaces.
- A study of the best frequency-based analysis for fringe pattern.
- A new algorithm for the automatic selection of the window size in WFT.
- A one-shot pattern projection technique for 3D dense reconstruction. The algorithm uses WFT and DeBruijn analysis to extract the 3D information from a colored fringe pattern. The obtained accuracy is similar to traditional sparse DeBruijn approaches.
- A new pipeline for 3D alignment of partial views obtained from SL.

7. CONCLUSIONS

7.3 Publications

The work developed in this thesis led to the following publications.

7.3.1 Journals

1. J. Salvi, S. Fernandez, T. Pribanic, "A state of the art in structured light patterns for surface profilometry", *Pattern Recognition*, 34, pp 2666-2680, 2010.
2. S. Fernandez, M. Gdeisat, J. Salvi, D. Burton, "Automatic window size selection in windowed Fourier Transform for 3D reconstruction using adapted mother wavelets profilometry", *Optics Communication*, 284(12), pp 2797-2807, 2011.
3. S. Fernandez and J. Salvi. One-shot absolute pattern for dense reconstruction using DeBruijn coding and WFT. Submitted to *Image and Vision Computing*.
4. S. Fernandez, D. Fofi, J. Salvi and J. Battle. Projector-camera calibration using a planar-based model. Submitted to *International Journal of Pattern Recognition and Artificial Intelligence*.

7.3.2 Conferences

1. S. Fernandez, J. Salvi and T. Pribanic. Absolute Phase Mapping for One-shot Dense Pattern Projection. PROCAMS'2010, IEEE Workshop on Projector-Camera Systems, in conjunction with IEEE International Conference on Computer Vision and Pattern Recognition, Article number 5543483, Pages 64-71, San Francisco (USA) June 18, 2010.
2. S. Fernandez, J. Forest and J. Salvi. Active stereo-matching for one-shot dense reconstruction. International Conference on Pattern Recognition Applications and Methods, Faro (Portugal) 6th/8th February 2012.
3. S. Fernandez, J. Salvi. Planar-based Camera-Projector Calibration. IEEE 7th International Symposium on Image and Signal Processing and Analysis (ISPA 2011), Dubrovnik (Croatia), September 4-6, 2011.
4. S. Fernandez, J. Salvi. A novel structured light method for one-shot dense reconstruction. IEEE International Conference on Image Processing (ICIP 2012). Coronado Spring, Florida (USA), September 30th - October 3th 2012.

7.3.3 Book chapters

1. S. Fernandez and J. Salvi, 3D reconstruction strategies in Structured Light, in Handbook of 3D machine vision: Optical metrology and imaging, compiled by Song Zhang. In press.

These publications, as well as some code, are available at the website:
<http://www.sergiofn.tk>

7. CONCLUSIONS

7.4 Future work

There are three main trends that must be studied as future work steps.

- The first step is to continue working on the registration problem. As was exposed in this thesis, many problems arise on the determination of the best algorithm for registration of 3D cloud of points. A first step refers to the selection of the best saliency points. As was observed in the experimental results, some effort must be put to avoid the selection of border positions as keypoints, as they do not represent the shape of the 3D object. Furthermore, the selection is done based on a DoG which is applied at N multidimensional filtering dimensions. The selection of the six scales of filtering is fixed, corresponding to standard deviation of values $\sigma_i \in 1\gamma, 2\gamma, 3\gamma, 4\gamma, 5\gamma, 6\gamma$ where γ amounts to 0.1% of the main diagonal. It would be interesting to perform an adaptive selection of these scales, which may be tuned to the shape of the surface. Moreover, the matching algorithm is nowadays working under Euclidean distance. It would be interesting to try other distance measure algorithm and compare them to see which one represents better the dissimilarity between surfaces. Furthermore, the optimal number of candidates for the triplets could be also studied.
- Regarding the 3D reconstruction and its performance under different scenarios, an important field to work on refers to the reconstruction of colored surfaces. This represents the main problem for the use of color patterns in structured light. As was shown in the experimental results of Chapter 5, the color of the surface get mixed with the colors used in the pattern. The Hue and the Illumination channels gets corrupted and color of the fringes is not well retrieved. This causes that the reconstruction present big holes on some regions of the plane, where the Illumination channel or the Hue channel are corrupted. Therefore, an interesting trend for research would be to develop a preprocessing algorithm able to minimize the effects of colored surface in the Hue and the Illumination channels, or retrieve more information from the damaged pattern.
- Another future step is to optimize the processing time and code the proposed algorithm into C++ or any other language that could suit on board computers of any 3D retrieval system. This would accelerate the post-processing step required until now, thus enabling to employ the proposed algorithm in real-time applications, which represent an interesting trend within the reconstruction of moving surfaces.

Appendix A

Geometric camera-projector calibration

A.1 Classical calibration systems

Camera projector calibration is constituted by two different steps: the camera calibration and the projector calibration. Camera calibration has been widely studied in the literature [116], [117], [118], [119]. Some contributions can also be found for projector calibration, which are usually grouped regarding three different parameters. The first one is the calibration object, which can be in 2D or 3D. 2D reference-object-based calibration includes methods having a planar pattern where a known image is projected by the projector and imaged by the camera. Instead, 3D reference-object-based calibration includes one or more than one fixed or mobile planes [120], [121], [122], or 3D calibration targets with known geometry [123]. The second parameter is the estimation technique, which can be Least Squares (LS) in both 2D and 3D space, and bundle adjustment. 2D LS estimation works with the projector linear method, therefore it is not possible to model non-linear distortion [124], [125]. A solution to this problem is constituted by the 3D LS algorithms, which work in the 3D space [123], [126], [127] and estimates both linear and non-linear parameters. However, 2D LS is easier to implement and faster than the 3D LS estimation techniques. Finally, bundle adjust apply a post-processing to a 2D linear estimation technique in order to minimize the cost function associated to the reprojection error. The last parameter refers to the projector model, and three models are found. In the line model the projector is described as a laser spot. Therefore, six parameters are considered (three for the center and three for the direction). The light-stripe model uses a plane to describe the projector. There-

fore the center coordinate and the plane direction are described. Finally, in the plane structured light technique the projector is regarded as the inverse of a camera, having the same parameters than the camera model. This is the model that is employed more often, as all the theory used for camera calibration can be adapted conveniently.

A.2 A novel proposal for projector calibration

The proposed projector calibration algorithm is based on the plane structured light model, where the projector is regarded as the inverse of a pinhole camera. However, an important constraint must be considered. In contrast with the calibration of a camera, there is a problem associated with the nature of the projector that complicates the calibration: the 3D points corresponding to the 2D projected pattern are not imaged by the projector (as it occurs in the camera calibration). Therefore, finding the correspondences between the 2D projected pattern and the 3D points implies the use of a calibrated camera to find the 3D position of the projected pattern. Using the geometry provided by the camera calibration it is possible to perform a projector calibration based on the plane structured light model. Following this idea, the extrinsic parameters of the projector are calculating placing the world coordinates at the camera center, therefore the computation of the transformation matrix is straightforward. That is, being the camera intrinsic parameters defined as in eq.(A.1), eq.(A.2), and being K_c , R_c , K_p and R_p the corresponding intrinsic and extrinsic parameters of the camera and the projector, respectively, we have:

$$R_c = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{A.1}$$

Therefore

$$P_c = K_c \cdot R_c = K_c \tag{A.2}$$

The transformation matrix P from the camera to the projector becomes (eq.(A.3)):

$$P = P_p \cdot P_c = K_p \cdot R_p \cdot inv(K_c) \tag{A.3}$$

The analytical inversion of radial and tangential distortion does not exist, as is not a linear distortion. However, an approximation can be done using an iterative method. In every iteration, the undistortion algorithm estimates the undistorted normalized point

A.2 A novel proposal for projector calibration

\tilde{m}_n as in (eq.(A.4)):

$$\tilde{m}_n = \tilde{m}_d - f_n(m_d, \delta) \quad (\text{A.4})$$

$$\tilde{m}_d = \tilde{m}_n$$

The previous iteration is run until the update of \tilde{m}_n becomes lower than a set threshold. A flow chart showing the different steps required for calibrating the projector can be seen in Fig. A.1. The key part of the proposed algorithm is the extraction of the 2D to 3D correspondences of the projector, from which the calibration can be pursued.

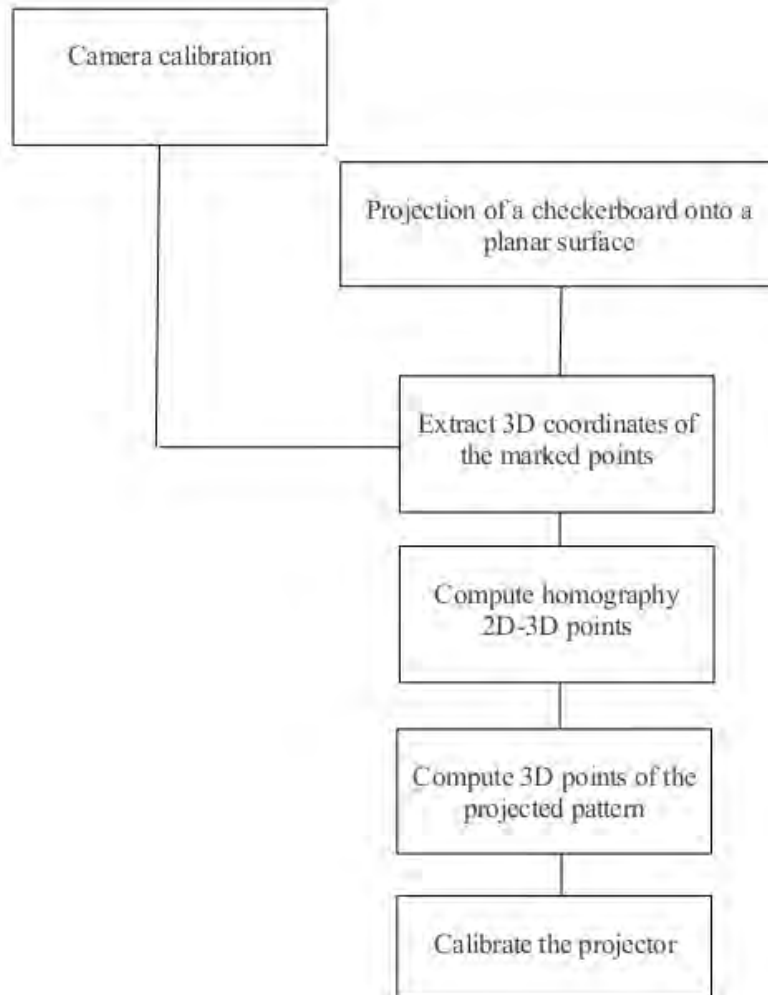


Figure A.1: Flow chart of the different steps in the calibration process.

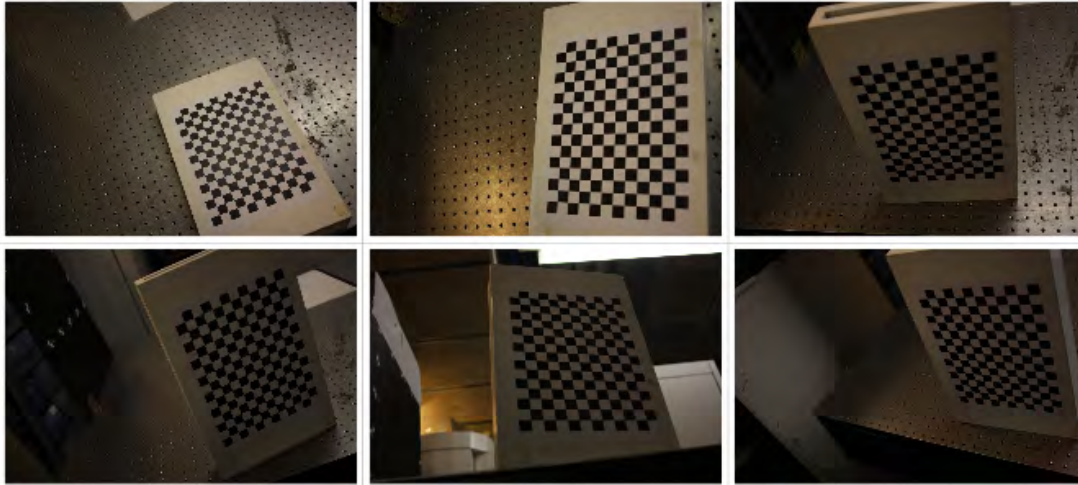


Figure A.2: Checkerboard images for camera calibration.

A.2.1 Camera calibration

The camera calibration algorithm follows the procedure proposed by Zhang [128] (an implementation of this method was done by Jean-Yves Bouguet [116] and is available for Matlab). Using this technique, a flat black and white checkerboard is placed at different positions on the 3D scene and imaged by the camera (see Fig. A.2). A minimum of two different images are necessary to extract the extrinsic parameters of the camera with respect to a chosen plane (the first imaged checkerboard). The intrinsic parameters need more orientations of the checkerboard to be computed (a total of $n = 14$ different poses are imaged in the experiments). The algorithm uses the extracted corner points of the checkerboard pattern to compute a projective transformation between the image points for the n different images, up to a scale factor. Afterwards, the camera intrinsic and extrinsic parameters are recovered using a closed-form solution, while the sixth-order radial distortion terms are recovered within a linear least-squares solution. A final nonlinear minimization of the reprojection error, solved using a Levenberg-Marquardt method, refines all the recovered parameters. At this stage, it is important to say that only the intrinsic parameters will be used, as we are only interested in the extrinsic parameters of the projector-camera pair.

A.2.2 Projection of a checkerboard onto a planar surface

The next step is the projection of a checkerboard using a Digital Light Projector. This pattern is projected onto a planar surface, placed at different positions on the 3D scene



Figure A.3: Planar surface (left image), and two different positions of the projected checkerboard (middle and right images).

and imaged by the camera. The border of the planar surface contains some marked points at relative fixed distances one from another, as can be seen in Fig. A.3.

A total of $m = 8 \sim 16$ different positions are imaged, in order to have enough variability in the 3D space.

A.2.3 Extract 3D coordinates of the marked points

In this step the 3D coordinates of some marked points are extracted for every image, so as to obtain the 3D coordinates of the planar surface. From the 2D positions of these points on the image scene, we compute the 3D rays passing through the camera center and crossing the 3D points (camera calibration is employed to this end). As we know the real 3D distances between the marked points, it is possible to extract their 3D coordinates from the 3D rays. It is important to note that the marked points must be spread all around the planar surface, in order to minimize the effect of radial and tangential distortion of the camera (which might have some error in the estimated values). Sub-pixel accuracy is considered for the detection of the 2D positions on the image. In practice, a total of $l = 8$ marked points equally distributed all around the planar surface give enough information to compute the homography.

A.2.4 Compute homography

Having the 2D to 3D correspondences of the marked positions in the planar surface, for every image (with respect to the camera frame), the next step is to compute the homography matrix that passes from 2D to 3D coordinates. A proper algorithm considering both linear and non-linear distortion of the camera is applied. It is important to mention that the 2D points must be normalized first; that is, must be expressed in millimeters instead of pixel coordinates.

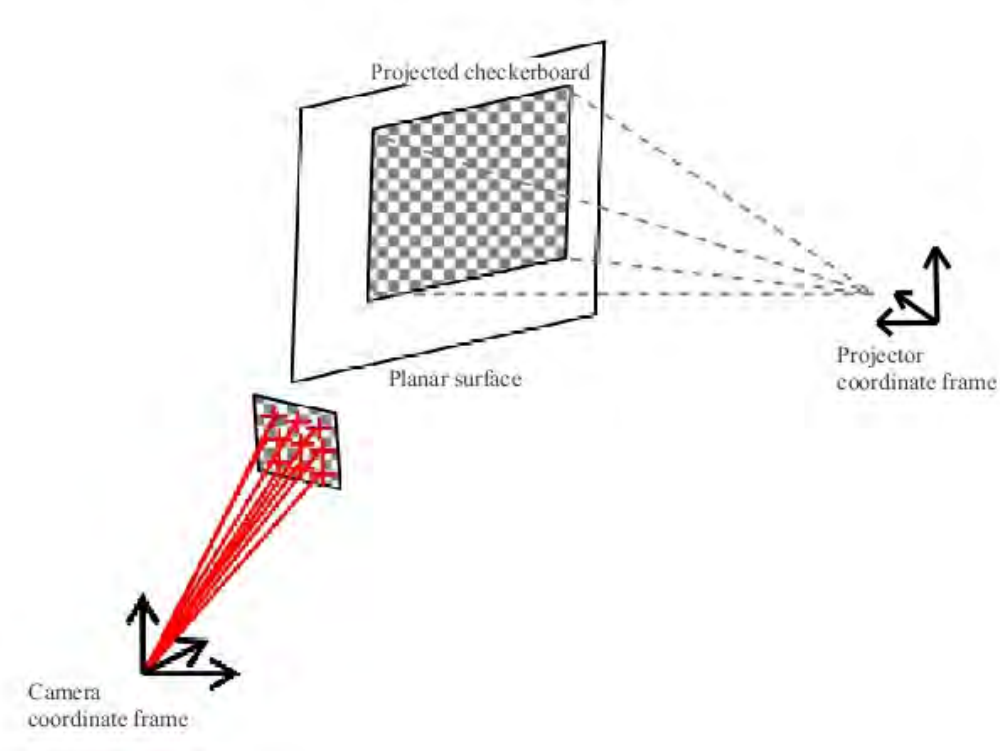


Figure A.4: Rays coming from the camera and going to the grid corners of the projected pattern.

A.2.5 Compute 3D points of the projected pattern

Next step is the extraction of the 2D corner coordinates of the projected checkerboard (with respect to the camera frame). A grid corner extraction algorithm with subpixel accuracy is applied. Afterwards, we use the homography to calculate the corresponding 3D coordinates. This is expressed in eq.(A.5) and Figs. A.4 and A.5.

$$\begin{bmatrix} sX \\ sY \\ sZ \end{bmatrix} = H \cdot \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (\text{A.5})$$

This is done for all the corners on the image and for different images so as to obtain a big number of non-coplanar 3D points for the calibration.

A.2.6 Calibrate the projector

The last step is to apply Zhangs method to the 2D to 3D correspondences for all the images, obtaining the optimized intrinsic and extrinsic parameters for the projector-

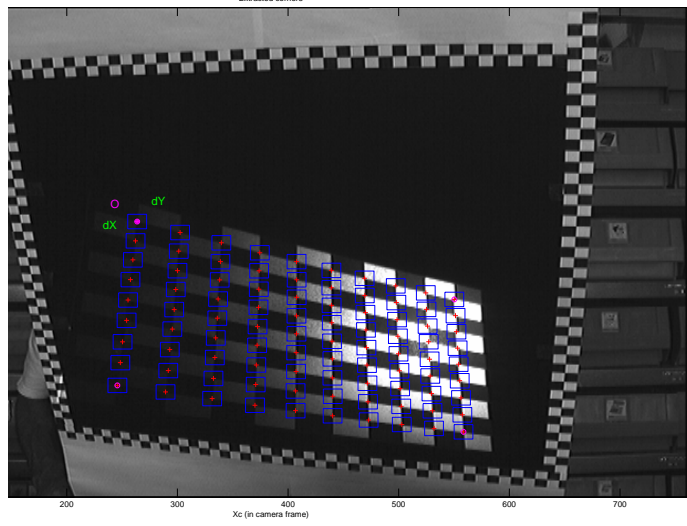


Figure A.5: Corners detection on the projected grid pattern.

camera pair. It is important to note that, as the projector works inversely to the camera, its radial and tangential distortion parameters must be obtained inversely to those for the camera calibration. Therefore, the undistortion algorithm is applied as in eq.(A.6):

$$\begin{aligned}
 m_n &= m_d + f_d(m_d, \delta), \\
 f_d(m_d, \delta) &= (k_1 r_d^2 + k_2 r_d^4) m_d + \left| \frac{2p_1 u_d v_d + p_2 (r_d^2 + 2u_d^2)}{p_1 (r_d^2 + 2v_d^2) + 2p_2 u_d v_d} \right| \\
 r_d^2 &= u_d^2 + v_d^2.
 \end{aligned} \tag{A.6}$$

A.3 Testing the proposed calibration

The proposed algorithm was tested under real conditions. The setup used for the tests was composed of a DLP video projector (Epson EMP-400W) with a resolution of 1024×768 pixels, and a digital camera (Sony 3CCD) digitizing images at 768×576 pixels. The baseline between camera and projector was about 0.5 m. The setup can be observed in Fig. A.6. The calibration pattern is composed of a painted checkerboard for camera calibration (Fig. A.2) placed on a 40×25 cm ceramic planar surface, and a blank metallic planar surface of dimensions 60×100 cm for the projection of the checkerboard (Fig. A.3). Both surfaces have a flatness deviation error lower than 0.5 mm. The

A. GEOMETRIC CAMERA-PROJECTOR CALIBRATION



Figure A.6: Projector-camera setup used to test the proposed algorithm.

implemented toolbox can be downloaded at http://eia.udg.edu/~sergiofn/proj_calib.zip. As example, the algorithm was run with 14 images for camera calibration and 14 images for projector calibration. The calibration results for the intrinsic and the extrinsic parameters are shown in table. A.1 and table. A.2, respectively. Furthermore, extrinsic geometry results can be observed in Fig. A.7.

As can be observed, the focal length presents an uncertainty error three order of magnitude lower than the computed value, which is comparable to the accuracy of the camera calibration. The non-linear distortion parameters present relatively higher error uncertainty than the focal length, but still lower than the provided values. Regarding the extrinsic parameters, uncertainty is much lower than the computed values for both rotation and translation. Moreover, a visual congruence can be observed between Fig. A.6 and Fig. A.7. Another way to analyze the error is to use the reprojection error functions available in the Matlab code for camera calibration. This tool computes the reprojection error of the extracted checkerboard points, for every image. Using these functions with the data of projector-camera calibration, we obtained the following error map shown in Fig. A.8. The method showed a maximum pixel error of around 2 pixels over the calibration error of the camera, which lead to a total reprojection error on the projector-camera setup of around 3 pixels. This is a reasonable value to work with

A.3 Testing the proposed calibration

Table A.1: Calibration results for the intrinsic parameters.

Parameter	Value (pixel)	Uncertainty (pixel)
α_x	906.75281	4.46296
α_y	917.71655	6.53165
u_0	524.99369	0.00000
v_0	825.16654	0.00000
ω	0.00000	0.00000
k_1	0.00671	0.01926
k_2	0.00135	0.01354
p_1	-0.02263	0.00481
p_2	0.00743	0.00101

Table A.2: Calibration results for the extrinsic parameters.

Parameter	Value (mm)	Uncertainty (mm)
$r_{1,1}$	1.0000	0.0000
$r_{1,2}$	-0.0003	0.0007
$r_{1,3}$	-0.0037	0.0015
$r_{2,1}$	0.0014	0.0007
$r_{2,2}$	0.9545	0.0000
$r_{2,3}$	0.2983	0.0026
$r_{3,1}$	0.0034	0.0015
$r_{3,2}$	-0.2983	0.0026
$r_{3,3}$	0.9545	0.0000
$t_{1,1}$	46.7205	1.2171
$t_{1,2}$	-281.1255	2.0332
$t_{1,3}$	-566.1478	2.1457

A. GEOMETRIC CAMERA-PROJECTOR CALIBRATION

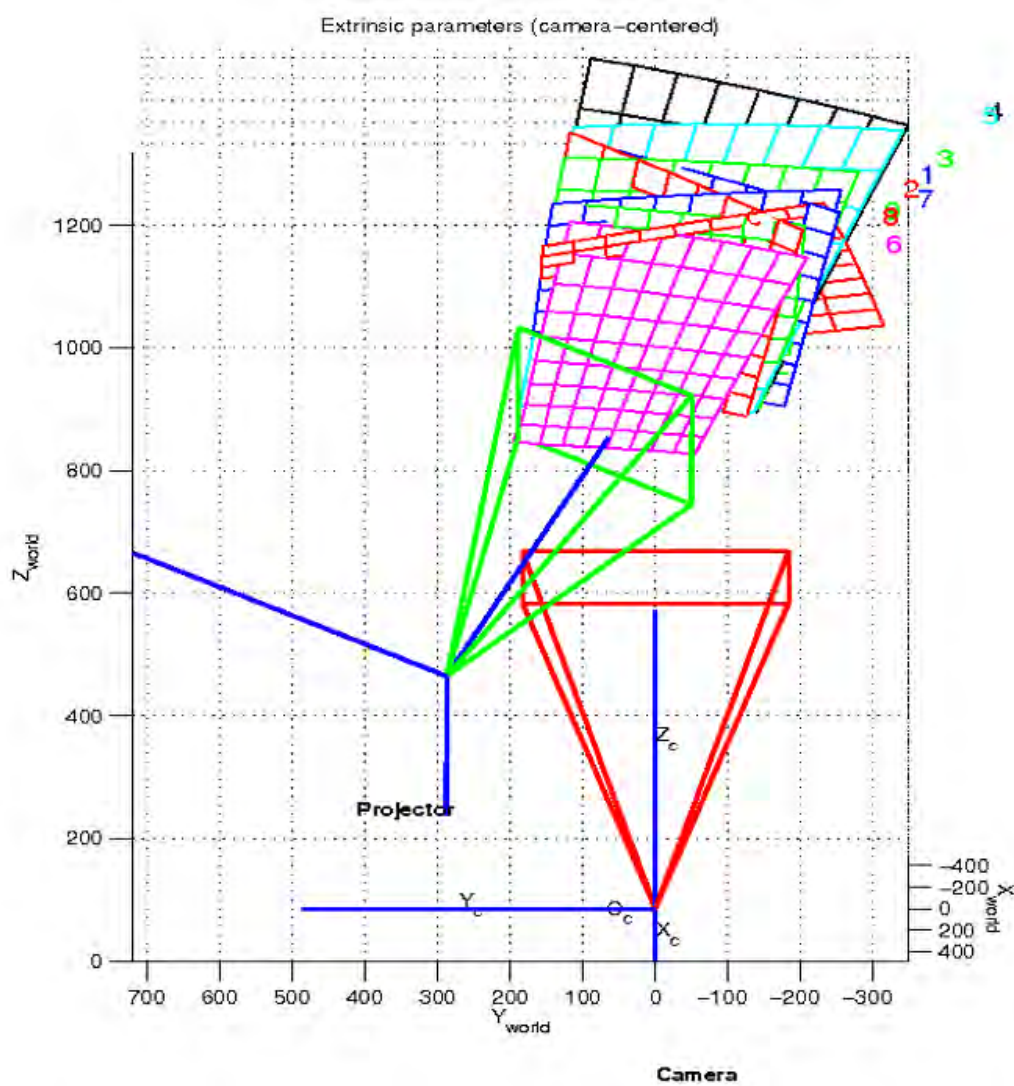


Figure A.7: Projector-camera calibration results for the given setup (extrinsic parameters).

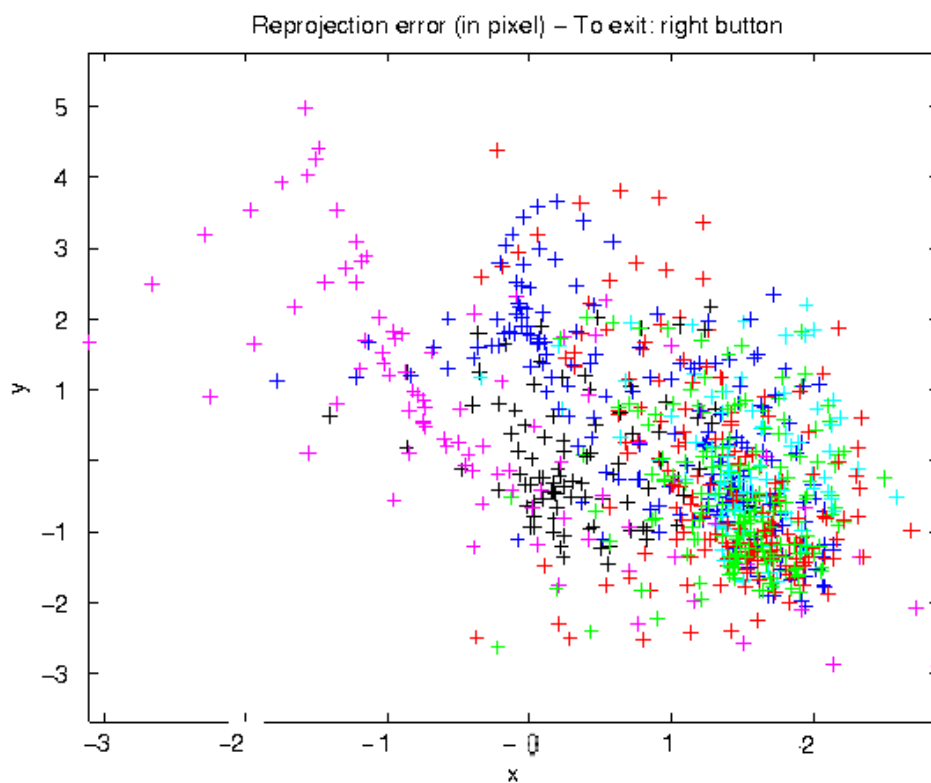


Figure A.8: Reprojection error of the projector calibration algorithm.

structured light systems.

A. GEOMETRIC CAMERA-PROJECTOR CALIBRATION

Appendix B

Color calibration and RGB channel alignment

This appendix sets the principles of color calibration and RGB channel alignment, necessary when working with color patterns.

B.1 Color calibration

Ideally, any discrete RGB instruction c with the same level of intensity i should produce the projection of light with the same intensity I . Similarly, a perfect camera should be able to digitize any incident light of wavelength λ and a certain intensity I to a discrete RGB triplet C with intensity i . In real conditions, however, the mapping from the RGB projection instruction c to the imaged RGB triplet C is a strongly non-linear process.

Fig. B.1 shows the actual projected and captured color values for red, green and blue, respectively, for a given pixel on the image. As can be observed, crosstalk effects and loss of intensity are present in the recovered pixel intensities. Experimentally it can be observed that usually the strongest crosstalk appears when projecting green, since it is not only detected by the green channel of the camera but also by the red one. In order to minimize the color crosstalk, a color calibration becomes necessary. Several radiometric models of a structured light system composed of a DLP and a color camera can be found in the literature ([6], [2], [14]). The most exhaustive color calibration technique was proposed by Caspi et al. [14], who developed a precise color calibration algorithm based on linearizing the projector-camera matrix and the surface reflectance matrix, specific for every scene point projected into a camera pixel. The

B. COLOR CALIBRATION AND RGB CHANNEL ALIGNMENT

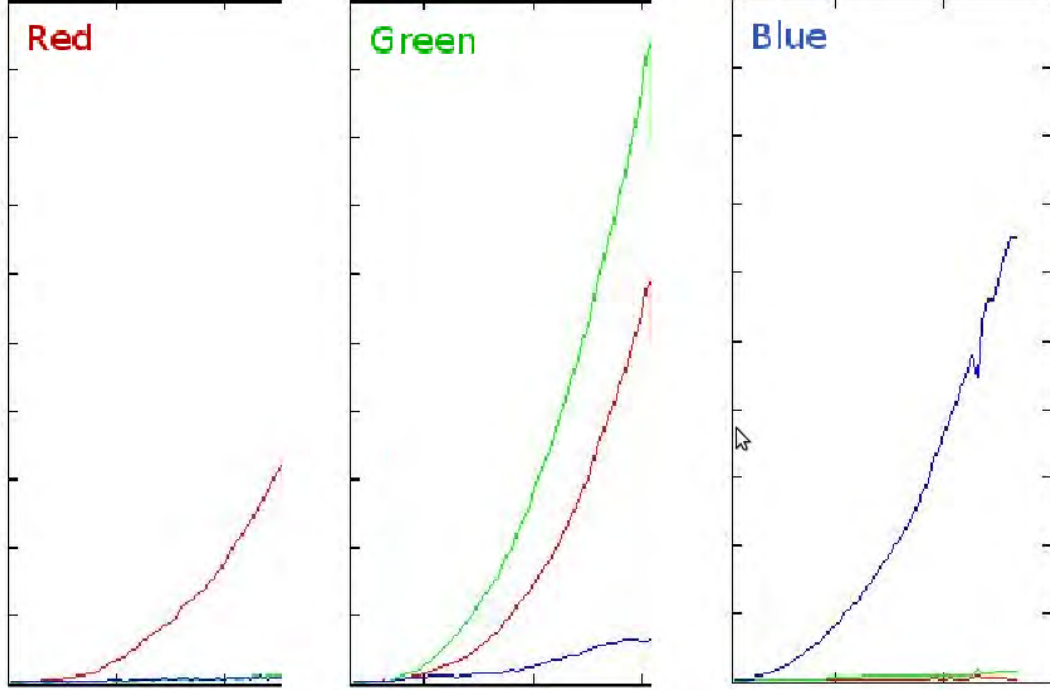


Figure B.1: Received color intensities for projected increasing values of Red, Green and Blue, respectively.

model presented by Caspi et al. is described in eq. B.2:

$$\underbrace{\begin{pmatrix} R \\ G \\ B \end{pmatrix}}_C = \underbrace{\begin{pmatrix} a_{rr} & a_{rg} & a_{rb} \\ a_{gr} & a_{gg} & a_{gb} \\ a_{br} & a_{bg} & a_{bb} \end{pmatrix}}_A \mathbf{KP} \underbrace{\begin{Bmatrix} r \\ g \\ b \end{Bmatrix}}_c + \underbrace{\begin{pmatrix} R_0 \\ G_0 \\ B_0 \end{pmatrix}}_{C_0}$$

Figure B.2: Projector-camera model for color calibration proposed by Caspi et al. [14].

where c is the RGB projection instruction sent to the projector and C the corresponding RGB triplet digitized by the camera. The consign c is actually modified by the non-linear behavior of the projector which actually projects a color denoted by P . K is a 3×3 matrix modeling the albedo of the illuminated object, and A is the color crosstalk matrix, while C_0 is the RGB tripled digitized by the camera when there is only ambient lighting. Therefore, A expresses how the RGB channels of the camera are affected by the RGB channels of the projector.

A simpler version of this method has been performed in our work, considering that

only three basic colors (red, green and blue) are projected and recovered by the color camera. The proposed algorithm uses least squares to linearize the combination matrix corresponding to the projector-camera pair and the surface reflectance matrices, in terms of response to color intensity, for each pixel in the received image and each color channel (red, green and blue). For every pixel and every color channel, the projected intensity is increased linearly and the corresponding captured color is stored. The stored color intensities must be compensated by the transformation matrix. A linear regression is computed that yields a matrix estimation of the projected color values for every received value. Having the set of three received color values R, G, B the estimated projected values R_0, G_0, B_0 are given by eq.(B.1). It is important to note that this calibration has been done under the assumption that all objects have a reflectance similar to the flat white lambertian plane used for calibration.

$$\begin{bmatrix} R_0 \\ G_0 \\ B_0 \end{bmatrix} = \begin{bmatrix} a_{rr} & a_{rg} & a_{rb} \\ a_{gr} & a_{gg} & a_{gb} \\ a_{br} & a_{bg} & a_{bb} \end{bmatrix} \begin{Bmatrix} R \\ G \\ B \end{Bmatrix} \quad (\text{B.1})$$

B.2 RGB channel alignment

Ideally, color cameras should perceive an intensity peak of white light at the same image coordinates in the three RGB channels. In practice there is an offset between the subpixel location of the intensity peaks in every RGB channel. This phenomenon is known as RGB channel misalignment. It is caused by spatial misalignments in the different CCD cells perceiving the red, green and blue light respectively. Although the order of these misalignments is usually below or around one pixel, it can produce higher order errors in 3D reconstruction. Furthermore, also DLP projectors suffer from this misalignment. Some authors propose to reduce the camera RGB channel misalignment by viewing an object providing reference points (like a checkerboard) and locating such points in the three channels separately. Afterwards, an homography can be calculated relating the position of the points in the red channel with respect to the ones in the green channel, and another homography doing the same between the points in the blue and the green channel. These homographies are then used to reduce the misalignment on the images [Zhang et al., 2002]. Nevertheless, this method totally ignores the RGB misalignment in the LCD projector.

We propose to minimize the RGB misalignment observed in the camera images taking into account both the camera and the projector at the same time, in the same

B. COLOR CALIBRATION AND RGB CHANNEL ALIGNMENT

way we did for color calibration. Our algorithm projects a checkerboard black and white pattern, which is imaged by the camera. For every color channel, the corners are detected with subpixel accuracy using Harris corner detector [129]. Afterwards, the relative position between every point in the red channel and their two corresponding points in the blue and in the green channels is extracted. We have observed that the relative positions of the channels coincide in all the checkerboard positions and that the relative offsets are very similar. That is why we finally store two unique offsets between the central channel and the other two. In our experimental setup we have found that on the images the central channel is the blue while the green and red channels have a little displacement to the left and to the right, respectively. These offsets have been named H_g^b and H_r^b , respectively. In order to reduce the global misalignment observed in an image it is necessary to apply the offset H_g^b to the green channel and the offset H_r^b to the red one and then combine the transformed channels with the original blue channel in order to obtain the rectified image. Note that the intensity of every transformed pixel in the new channels must be interpolated from the neighboring pixels in the corresponding source channel since the offsets are at sub-pixel precision.

Appendix C

Application of the Remainder Theorem to the use in Multiple Phase Shifting pattern projection

This information is extracted from the work of Pribanic et al. [9].

Two integer numbers Φ_{ABS} and ϕ_R are in congruence if they give the same remainder when they are divided by a given number λ . Hence, Φ_{ABS} and ϕ_R are said to be congruent modulo λ , as depicted in eq. C.1, eq. C.2:

$$\Phi_{ABS} \equiv \phi_R \pmod{\lambda} \tag{C.1}$$

$$\begin{aligned} \Phi_{ABS} &\equiv \phi_{R1} \pmod{\lambda_1} \\ \Phi_{ABS} &\equiv \phi_{R2} \pmod{\lambda_2} \\ &\vdots \\ \Phi_{ABS} &\equiv \phi_{Rk} \pmod{\lambda_k} \end{aligned} \tag{C.2}$$

where Φ_{ABS} and ϕ_{Ri} are integers and λ_i are positive integers but also relative primes. A solution to eq. C.1 is provided by the famous Chinese remainder theorem [54] depicted in:

$$\Phi_{ABS} = \sum_{i=1}^k \phi_{Ri} \cdot e_i \pmod{\lambda_1 \cdot \lambda_2 \cdot \dots \cdot \lambda_k} \tag{C.3}$$

C. APPLICATION OF THE REMAINDER THEOREM TO THE USE IN MULTIPLE PHASE SHIFTING PATTERN PROJECTION

where the coefficients e_i can be computed as follows:

$$\begin{aligned} e_i &\equiv 1 \pmod{\lambda_i} \\ e_i &\equiv 0 \pmod{\lambda_j} \quad i \neq j \end{aligned} \tag{C.4}$$

Hence, some e_i is a number which divided by the corresponding λ_i yields the remainder 1.

Given two arbitrary fringe pattern where their are relative prime numbers between the, the following condition holds:

$$\Phi_{ABS} = k_1 \cdot \lambda_1 + \phi_{R1} = k_2 \cdot \lambda_2 + \phi_{R2} \tag{C.5}$$

where k_1 and k_2 are the number of periods typically unknown in practice, but necessary to reach some corresponding Φ_{ABS} unwrapped value given some known ϕ_{R1} and ϕ_{R2} . Note that both Φ_{ABS} and ϕ_{R1} divided by λ_1 give the same remainder, ϕ_{R1} . Similarly Φ_{ABS} and ϕ_{R2} divided by λ_2 give the same remainder, ϕ_{R2} . Hence, we can set up a system of congruence equations C.2 and find a solution for C.3. Until now and for simplicity we have assumed integer values for the wrapped phases ϕ_{R1} and ϕ_{R2} . Of course, in practice, ϕ_{R1} and ϕ_{R2} are real numbers. Usually, the integer part is considered as an initial guess for computing Φ_{ABS} in eq. C.3. Then, ideally, the fractional parts ϕ_{R1}^{FRAC} and ϕ_{R2}^{FRAC} of both wrapped phases should be equal. However, in order to minimize noise influence, in practice, the mean values of both fractional parts are added to the initially computed Φ_{ABS} in order to find the ultimate solution for the unwrapped phase. The well-known problem in the original G-S method was that ϕ_{R1} and ϕ_{R2} were rounded and then used in eq. C.3.

Appendix D

Dynamic programming

The information explained in this Annex can be check in the work of Zhang et al. [26].

D.1 Classical dynamic programming

Let $G_{j;i}$ be the sub-grid defined by $[0; j]x[0; i]$, and $\phi_{j;i}^*$ be the optimal path in $G_{j;i}$. Three possible solutions for $\phi_{j;i}^*$ can appear. The first one consists of vertex $(j; i)$ and the optimal path $\phi_{j-1;i-1}^*$ in $G_{j-1;i-1}$. The second one is entirely in the sub-grid $G_{j-1;i}$, and the third one is in the subgrid $G_{j;i-1}$. In the latter two cases $\phi_{j;i}^* = \phi_{j-1;i}^*$ or $\phi_{j;i}^* = \phi_{j;i-1}^*$, respectively. Consequently, $\sigma(\phi_{j;i}^*)$ may be recursively computed as in eq. D.1):

$$\left\{ \begin{array}{l} 0, \text{ if } j = 0 \text{ or } i = 0 \\ \left\{ \begin{array}{l} \sigma(\phi_{j-1;i-1}^*) + \text{score}(q_j, e_i) \\ \sigma(\phi_{j-1;i}^*) \\ \sigma(\phi_{j;i-1}^*) \end{array} \right\} \\ \text{otherwise} \end{array} \right\} \quad (\text{D.1})$$

The cost of the optimal solution ϕ^* is given by $\sigma(\phi_{N;M}^*)$, where $\phi_{N;M}^*$ is computed by tracking back through the cost matrix computed from eq. D.1.

D.2 Multi-pass dynamic programming

A fundamental limitation of matching algorithms based on dynamic programming (DP) is the assumption of monotonicity, which is violated in the presence of occlusions. Due to the occlusion, the order of projected transitions and detected edges is not the same,

D. DYNAMIC PROGRAMMING

```
procedure MultiPassDP(grid)
  set path to be empty;
  while(path1 := DP(grid) is not empty)
    path := path ∪ path1;
    remove columns and rows in path1 from grid;
  return path;
end MultiPassDP
```

resulting in a non-monotonic path in the grid. The dynamic programming algorithm therefore fails to find the optimal path; instead, it will identify the optimal monotonic solution. While this solution could potentially be quite different than the optimal path, in practice we have seen that it corresponds to a monotonic component of the optimal solution. In the case of Fig. D.1, DP identifies the sub-path consisting of $(A; B; C; D; E)$. The rest of the optimal solution, sub-path $(F; G)$, is itself monotonic and can be identified by applying dynamic programming on the sub-grid obtained by removing columns $(1; 2; 4; 5; 6; 9)$, and rows $(1; 2; 5; 6; 7; 8)$ from the original grid. The same procedure may be repeated until all rows and columns are exhausted. This procedure, which we call MultiPassDP, is summarized as follows:

MultiPass dynamic programming computes the monotonic components of the optimal path in multiple passes, enabling solution of correspondence problems with occlusions that are not possible with traditional dynamic programming. Instead of exhausting the positive monotonic components, *path1*, in the grid, the number dynamic programming passes can also be specified by a user, based on prior knowledge of how many layers of structure the scene contains.

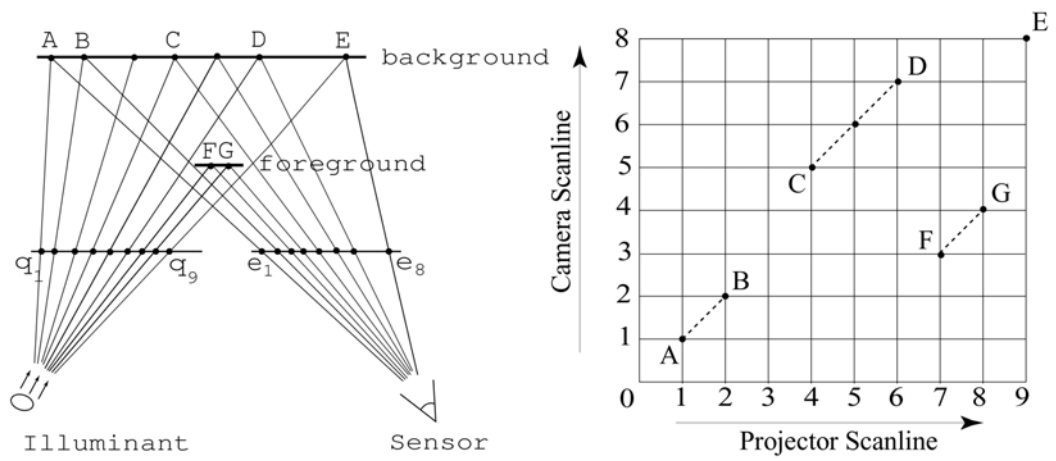


Figure D.1: On the left, example of the violation of the monotonicity assumption. On the right, the resulting match grid.

D. DYNAMIC PROGRAMMING

References

- [1] J. SALVI, J. PAGES, AND J. BATLLE. **Pattern codification strategies in structured light systems.** *Pattern Recognition*, **37**(4):827–849, 2004. [ix](#), [12](#), [13](#), [17](#)
- [2] J. PAGES, J. SALVI, AND J. FOREST. **A new optimised de bruijn coding strategy for structured light patterns.** In *17th International Conference on Pattern Recognition, ICPR 2004*, **4**, pages 284–287, 2004. [ix](#), [17](#), [18](#), [37](#), [42](#), [43](#), [44](#), [59](#), [147](#)
- [3] RA MORANO, C. OZTURK, R. CONN, S. DUBIN, S. ZIETZ, AND J. NISSANO. **Structured light using pseudorandom codes.** *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **20**(3):322–327, 1998. [ix](#), [6](#), [19](#), [20](#), [42](#), [45](#)
- [4] I. ISHII, K. YAMAMOTO, K. DOI, AND T. TSUJI. **High-speed 3D image acquisition using coded structured light projection.** In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 925–930, 2007. [ix](#), [22](#), [43](#), [44](#)
- [5] C. GUAN, L. HASSEBROOK, AND D. LAU. **Composite structured light pattern for three-dimensional video.** *Optics Express*, **11**(5):406–417, 2003. [ix](#), [25](#), [26](#), [43](#), [60](#)
- [6] TP. MONKS, JN. CARTER, AND CH. SHADLE. **Colour-encoded structured light for digitisation of real-time 3D data.** In *IEE 4th International Conference on Image Processing*, pages 327–30, 1992. [ix](#), [6](#), [17](#), [36](#), [37](#), [38](#), [40](#), [42](#), [44](#), [94](#), [147](#)
- [7] JL POSDAMER AND MD ALTSCHULER. **Surface measurement by space-encoded projected beam systems.** *Computer Graphics and Image Processing*, **18**(1):1–17, 1982. [ix](#), [21](#), [36](#), [37](#), [38](#), [40](#)

REFERENCES

- [8] J. GÜHRING. **Dense 3-D surface acquisition by structured light using off-the-shelf components.** *Videometrics and Optical Methods for 3D Shape Measurement*, **4309**:220–231, 2001. [ix](#), [23](#), [36](#), [37](#), [40](#), [43](#)
- [9] T. PRIBANIC, H. DAPO, AND J. SALVI. **Efficient and Low-Cost 3D Structured Light System Based on a Modified Number-Theoretic Approach.** *EURASIP Journal on Advances in Signal Processing, Volume 2010, Article ID 474389, 11 pages*, 2009. [ix](#), [25](#), [36](#), [37](#), [39](#), [41](#), [43](#), [44](#), [94](#), [97](#), [129](#), [151](#)
- [10] J. LI, X. SU, AND L. GUO. **Improved Fourier transform profilometry for the automatic measurement of three-dimensional object shapes (Journal Paper).** *Optical Engineering*, **29**(12):1439–1444, 1990. [ix](#), [28](#), [36](#), [38](#), [41](#), [44](#), [94](#)
- [11] B. CARRIHILL AND R. HUMMEL. **Experiments with the intensity ratio depth sensor.** *Computer Vision, Graphics, and Image Processing*, **32**(3):337–358, 1985. [ix](#), [6](#), [35](#), [36](#), [38](#), [39](#), [41](#), [44](#), [46](#), [59](#)
- [12] M.A. HERRÁEZ, D.R. BURTON, M.J. LALOR, AND M.A. GDEISAT. **Fast two-dimensional phase-unwrapping algorithm based on sorting by reliability following a noncontinuous path.** *Applied Optics*, **41**(35):7437–7444, 2002. [x](#), [73](#), [74](#), [75](#)
- [13] A. ZAHARESCU, E. BOYER, K. VARANASI, AND R. HORAUD. **Surface feature detection and description with applications to mesh matching.** In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 373–380. IEEE, 2009. [xii](#), [117](#), [118](#)
- [14] D. CASPI, N. KIRYATI, AND J. SHAMIR. **Range imaging with adaptive color structured light.** *IEEE Transactions on Pattern analysis and machine intelligence*, **20**(5):470–480, 1998. [xiii](#), [21](#), [43](#), [147](#), [148](#)
- [15] R. HARTLEY AND A. ZISSERMAN. *Multiple view geometry in computer vision*, **2**. Cambridge Univ Press, 2000. [2](#)
- [16] T.S. HUANG AND A.N. NETRAVALI. **Motion and structure from feature correspondences: A review.** *Proceedings of the IEEE*, **82**(2):252–268, 1994. [2](#)
- [17] L. ZAPPELLA. **Motion segmentation from feature trajectories.** *Master's thesis, University of Girona, Spain*, 2008. [2](#)

-
- [18] X. ARMANGUÉ, H. ARAÚJO, AND J. SALVI. **A review on egomotion by means of differential epipolar geometry applied to the movement of a mobile robot.** *Pattern Recognition*, **36**(12):2927–2944, 2003. [2](#)
- [19] N. DE LA BLANCA, J. FUERTES, AND M. LUCENA. **3D Rigid Facial Motion Estimation from Disparity Maps.** *Progress in Pattern Recognition, Speech and Image Analysis*, pages 54–61, 2003. [2](#)
- [20] O. FAUGERAS. *Three-dimensional computer vision: a geometric viewpoint.* the MIT Press, 1993. [2](#)
- [21] Z. ZHANG. **Determining the epipolar geometry and its uncertainty: A review.** *International Journal of Computer Vision*, **27**(2):161–195, 1998. [2](#)
- [22] H. KAWASAKI, R. FURUKAWA, R. SAGAWA, AND Y. YAGI. **Dynamic scene shape reconstruction using a single structured light pattern.** In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, pages 1–8, 2008. [3](#), [4](#), [12](#), [18](#), [44](#)
- [23] J. BATLLE, E. MOUADDIB, AND J. SALVI. **Recent progress in coded structured light as a technique to solve the correspondence problem: a survey.** *Pattern Recognition*, **31**(7):963–982, 1998. [3](#), [12](#), [13](#), [88](#)
- [24] J. SALVI, J. BATLLE, AND E. MOUADDIB. **A robust-coded pattern projection for dynamic 3D scene measurement.** *Pattern Recognition Letters*, **19**(11):1055–1065, 1998. [3](#), [11](#), [17](#), [42](#), [45](#)
- [25] J. PAGES, J. SALVI, C. COLLEWET, AND J. FOREST. **Optimised De Bruijn patterns for one-shot shape acquisition.** *Image Vision and Computing*, **23**:707–720, 2005. [6](#), [17](#), [45](#)
- [26] L. ZHANG, B. CURLESS, AND SM SEITZ. **Rapid shape acquisition using color structured light and multi-pass dynamic programming.** In *3D Data Processing Visualization and Transmission*,, pages 24–36, 2002. [6](#), [23](#), [43](#), [153](#)
- [27] J. PAGES, C. COLLEWET, F. CHAUMETTE, J. SALVI, S. GIRONA, AND F. RENNES. **An approach to visual servoing based on coded light.** In *IEEE Int. Conference on Robotics and Automation, ICRA*, **6**, pages 4118–4123, 2006. [6](#), [20](#), [42](#), [44](#)

REFERENCES

- [28] C. DOIGNON AND D. KNITTEL. **A structured light vision system for out-of-plane vibration frequencies location of a moving web.** *Machine Vision and Applications*, **16**(5):289–297, 2005. [6](#)
- [29] J. SALVI, S. FERNANDEZ, T. PRIBANIC, AND X. LLADO. **A state of the art in structured light patterns for surface profilometry.** *Pattern recognition*, **43**(8):2666–2680, 2010. [6](#), [46](#), [56](#), [88](#), [93](#)
- [30] J. TAJIMA AND M. IWAKAWA. **3-D data acquisition by rainbow range finder.** In *Pattern Recognition, 1990. Proceedings., 10th International Conference on*, **1**, pages 309–313, 1990. [6](#), [35](#), [36](#), [44](#), [46](#)
- [31] J. FOREST AND J. SALVI. **An overview of laser slit 3d digitasers.** *IEEE/RSJ Int.Conference on Robots and Systems*, 2002. [12](#)
- [32] JJ LE MOIGNE AND A.M. WAXMAN. **Structured light patterns for robot mobility.** *Robotics and Automation, IEEE Journal of*, **4**(5):541–548, 1988. [12](#)
- [33] S. ZHANG AND P.S. HUANG. **Novel method for structured light system calibration.** *Optical Engineering*, **45**(1):1–8, 2006. [12](#)
- [34] X. SU AND W. CHEN. **Fourier transform profilometry: a review.** *Optics and Lasers in Engineering*, **35**(5):263–284, 2001. [13](#), [29](#)
- [35] H. FREDRICKSEN. **A survey of full length nonlinear shift register cycle algorithms.** *SIAM Review*, pages 195–221, 1982. [16](#), [86](#)
- [36] KL. BOYER AND AC. KAK. **Color-encoded structured light for rapid active ranging.** *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **9**(1):14–28, 1987. [16](#), [17](#)
- [37] F. FORSTER. **A High-Resolution and High Accuracy Real-Time 3D Sensor Based on Structured Light.** In *Proc.3th International Symposium on 3D Data Processing, Visualization, and Transmission*, pages 208–215, 2006. [18](#), [42](#), [44](#), [45](#)
- [38] P. FECHTELER AND P. EISERT. **Adaptive color classification for structured light systems.** In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pages 1–7, 2008. [18](#), [42](#), [44](#)

-
- [39] M.A. TEHRANI, A. SAGHAEIAN, AND O.R. MOHAJERANI. **A New Approach to 3D Modeling Using Structured Light Pattern.** In *Information and Communication Technologies: From Theory to Applications, 2008. ICTTA 2008.*, pages 1–5, 2008. [18](#)
- [40] M. MARUYAMA AND S. ABE. **Range sensing by projecting multiple slits with random cuts.** *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **15**(6):647–651, 1993. [18](#)
- [41] M. ITO AND A. ISHII. **A three-level checkerboard pattern (TCP) projection method for curved surface measurement.** *Pattern Recognition*, **28**(1):27–40, 1995. [18](#)
- [42] TP KONINCKX AND L. VAN GOOL. **Real-time range acquisition by adaptive structured light.** *IEEE transactions on pattern analysis and machine intelligence*, **28**(3):432–445, 2006. [19](#)
- [43] T. ETZION. **Constructions for perfect maps and pseudorandom arrays.** *IEEE Transactions on information theory*, **34**(5 Part 1):1308–1316, 1988. [19](#)
- [44] FJ MACWILLIAMS AND NJA SLOANE. **Pseudo-random sequences and arrays.** *Proceedings of the IEEE*, **64**(12):1715–1729, 1976. [19](#)
- [45] H. MORITA, K. YAJIMA, AND S. SAKATA. **Reconstruction of surfaces of 3-d objects by m-array pattern projection method.** In *Computer Vision., Second International Conference on*, pages 468–473, 1988. [19](#)
- [46] P.M. GRIFFIN, L.S. NARASIMHAN, AND S.R. YEE. **Generation of uniquely encoded light patterns for range data acquisition.** *Pattern Recognition*, **25**(6):609–616, 1992. [19](#), [43](#)
- [47] C. ALBITAR, P. GRAEBLING, AND C. DOIGNON. **Design of a monochromatic pattern for a robust structured light coding.** In *IEEE Int. Conf. Image Process. ICIP*, **6**, pages 529–532, 2007. [20](#), [42](#)
- [48] M. MINOU, T. KANADE, AND T. SAKAI. **A method of time-coded parallel planes of light for depth measurement.** *Trans. IECE Japan*, **64**(8):521–528, 1981. [21](#)
- [49] G. SANSONI, M. CAROCCI, AND R. RODELLA. **Calibration and performance evaluation of a 3-D imaging sensorbased on the projection of structured**

REFERENCES

- light**. *IEEE Transactions on instrumentation and measurement*, **49**(3):628–636, 2000. [23](#), [36](#), [43](#)
- [50] C.S. CHEN, Y.P. HUNG, C.C. CHIANG, AND J.L. WU. **Range data acquisition using color structured lighting and stereo vision**. *Image and Vision Computing*, **15**(6):445–456, 1997. [23](#)
- [51] KURIHARA T. ONO N., SHIMIZU T. AND ANDO S. **Real-time 3-D imager based on spatio-temporal phase unwrapping**. **21**, pages 437–440, 2004. [24](#), [43](#)
- [52] V. SRINIVASAN, HC LIU, AND M. HALIOUS. **Automated phase-measuring profilometry: a phase mapping approach**. *Applied Optics*, **24**:185–188, 1985. [25](#), [43](#)
- [53] C. WUST AND D.W. CAPSON. **Surface profile measurement using color fringe projection**. *Machine Vision and App.*, **4**(3):193–203, 1991. [25](#), [43](#), [60](#)
- [54] P. RIBENBOIM. *Algebraic numbers*. R. Courant, L. Bers, J.J. Stoker. John Wiley and Sons. New York, 1972. [25](#), [43](#), [47](#), [60](#), [128](#), [151](#)
- [55] VI GUSHOV AND Y.N. SOLODKIN. **Automatic processing of fringe patterns in integer interferometers**. *Opt. Lasers Eng*, **14**(4-5):311–324, 1991. [25](#)
- [56] M.T.K. TAKEDA M, MUTOH. **Fourier transform profilometry for the automatic measurement of 3-D object shapes**. *Appl. Opt*, **22**:3977–3982, 1983. [26](#), [28](#), [36](#), [44](#), [56](#)
- [57] P.J. COBELLI, A. MAUREL, V. PAGNEUX, AND P. PETITJEANS. **Global measurement of water waves by Fourier transform profilometry**. *Experiments in Fluids*, **46**(6):1037–1047, 2009. [28](#)
- [58] E. HU AND Y. HE. **Surface profile measurement of moving objects by using an improved π phase-shifting Fourier transform profilometry**. *Optics and Lasers in Engineering*, **47**(1):57–61, 2009. [28](#)
- [59] W. CHEN, P. BU, S. ZHENG, AND X. SU. **Study on Fourier transforms profilometry based on bi-color projecting**. *Optics and Laser Technology*, **39**(4):821–827, 2007. [29](#), [32](#), [36](#), [38](#), [44](#), [129](#)

-
- [60] H.M. YUE, X.Y. SU, AND Y.Z. LIU. **Fourier transform profilometry based on composite structured light pattern.** *Optics and Laser Technology*, **39**(6):1170–1175, 2007. [29](#), [36](#), [38](#), [44](#)
- [61] J.F. LIN AND X. SU. **Two-dimensional Fourier transform profilometry for the automatic measurement of three-dimensional object shapes.** *Optical Engineering*, **34**:3297–3297, 1995. [29](#), [34](#)
- [62] W. CHEN, X. SU, Y. CAO, Q. ZHANG, AND L. XIANG. **Method for eliminating zero spectrum in Fourier transform profilometry.** *Optics and Lasers in Engineering*, **43**(11):1267–1276, 2005. [29](#)
- [63] F. BERRYMAN, P. PYNSENT, J. FAIRBANK, AND S. DISNEY. **A new system for measuring three-dimensional back shape in scoliosis.** *European Spine Journal*, **17**(5):663–672, 2008. [29](#)
- [64] IWAASA Y. TOYOOKA S. **Automatic prolometry of 3-D diffuse objects by spatial phase detection.** *Applied Optics*, **25**(10):1630–1633, 1986. [29](#)
- [65] J.G. PROAKIS, D.G. MANOLAKIS, DG MANOLAKIS, AND JG PROAKIS. *Digital signal processing: principles, algorithms, and applications*, **3**. Prentice Hall New Jersey, 1996. [30](#)
- [66] F. BERRYMAN, P. PYNSENT, AND J. CUBILLO. **A theoretical comparison of three fringe analysis methods for determining the three-dimensional shape of an object in the presence of noise.** *Optics and lasers in Engineering*, **39**(1):35–50, 2003. [30](#)
- [67] Q. KEMAO. **Two-dimensional windowed Fourier transform for fringe pattern analysis: principles, applications and implementations.** *Optics and Lasers in Engineering*, **45**(2):304–317, 2007. [31](#), [32](#), [68](#)
- [68] M.A. GDEISAT, D.R. BURTON, AND M.J. LALOR. **Eliminating the zero spectrum in Fourier transform profilometry using a two-dimensional continuous wavelet transform.** *Optics Communications*, **266**(2):482–489, 2006. [32](#), [33](#), [44](#), [68](#), [69](#), [84](#), [129](#)
- [69] ABDULBASIT ZAID AHMED ABID. *Fringe Pattern Analysis using Wavelet Transforms*. Phd thesis, general engineering research institute (geri). liverpool john moores university, Liverpool, UK, 2008. [32](#), [50](#), [60](#), [68](#), [69](#), [74](#), [79](#)

REFERENCES

- [70] Q. ZHANG, W. CHEN, AND Y. TANG. **Method of choosing the adaptive level of discrete wavelet decomposition to eliminate zero component.** *Optics Communications*, **282**(5):778–785, 2008. [33](#), [44](#)
- [71] Q. KEMAO. **Windowed Fourier transform for fringe pattern analysis.** *Applied optics*, **43**(17):3472–3473, 2004. [33](#), [69](#)
- [72] L.S. WU AND Q. PENG. **Research and development of fringe projection-based methods in 3D shape reconstruction.** *Journal of Zhejiang University-Science A*, **7**(6):1026–1036, 2006. [33](#), [84](#)
- [73] A. BALDI, F. BERTOLINO, AND F. GINESU. **On the performance of some unwrapping algorithms.** *Optics and Lasers in Engineering*, **37**(4):313–330, 2002. [33](#), [84](#)
- [74] S.S. GORTHI AND K.R. LOLLA. **A new approach for simple and rapid shape measurement of objects with surface discontinuities.** In *Proceedings. SPIE*, **5856**, pages 184–194, 2005. [34](#)
- [75] TR JUDGE AND PJ BRYANSTON-CROSS. **A review of phase unwrapping techniques in fringe analysis.** *Optics and Lasers in Engineering*, **21**(4):199–240, 1994. [34](#)
- [76] P.S. HUANG, S. ZHANG, F.P. CHIANG, ET AL. **Trapezoidal phase-shifting method for three-dimensional shape measurement.** *Optical Engineering*, **44**:142–152, 2005. [34](#), [44](#)
- [77] P. JIA, J. KOFMAN, AND C. ENGLISH. **Two-step triangular-pattern phase-shifting method for three-dimensional object-shape measurement.** *Optical Engineering*, **46**:083201, 2007. [34](#)
- [78] G.S. SPAGNOLO, G. GUATTARI, C. SAPIA, D. AMBROSINI, D. PAOLETTI, AND G. ACCARDO. **Contouring of artwork surface by fringe projection and FFT analysis.** *Optics and Lasers in Engineering*, **33**(2):141–156, 2000. [35](#)
- [79] AF. LAINE. **Wavelets in temporal and spatial processing of biomedical images.** *Annual Review on Biomedical Engineering*, (2):511–550, 2000. [68](#)
- [80] L. HUANG, Q. KEMAO, B. PAN, AND A.K. ASUNDI. **Comparison of Fourier transform, windowed Fourier transform, and wavelet transform methods for phase extraction from a single fringe pattern in fringe pro-**

- jection profilometry.** *Optics and Lasers in Engineering*, **48**(2):141–148, 2010. [68](#)
- [81] M.A. GDEISAT, A. ABID, D.R. BURTON, M.J. LALOR, F. LILLEY, C. MOORE, AND M. QUDEISAT. **Spatial and temporal carrier fringe pattern demodulation using the one-dimensional continuous wavelet transform: recent progress, challenges, and suggested developments.** *Optics and Lasers in Engineering*, **47**(12):1348–1361, 2009. [68](#), [69](#)
- [82] A. DURSUN, Z. SARAÇ, H. SARAÇ TOPKARA, S. OZDER, AND F. NECATI ECEVIT. **Phase recovery from interference fringes by using S-transform.** *Measurement*, **41**(4):403–411, 2008. [68](#)
- [83] M. AFIFI, A. FASSI-FIHRI, M. MARJANE, K. NASSIM, M. SIDKI, AND S. RACHAFI. **Paul wavelet-based algorithm for optical phase distribution evaluation.** *Optics Communications*, **211**(1-6):47–51, 2002. [68](#)
- [84] P. TOMASSINI, A. GIULIETTI, L.A. GIZZI, M. GALIMBERTI, D. GIULIETTI, M. BORGHESI, AND O. WILLI. **Analyzing laser plasma interferograms with a continuous wavelet transform ridge extraction technique: the method.** *Applied optics*, **40**(35):6561–6568, 2001. [68](#)
- [85] STEPHANE MALLAT. *A Wavelet Tour of Signal Processing*. AP Professional, London, 1997. [68](#)
- [86] S. QIAN. *Introduction to time-frequency and wavelet transforms*. Prentice Hall PTR, 2002. [68](#)
- [87] H. LI AND C. YANG. **Two-dimensional multiscale windowed Fourier transform based on two-dimensional wavelet transform for fringe pattern demodulation.** *Optics & Laser Technology*, **43**(1):72–81, 2011. [71](#)
- [88] J. WENG, J. ZHONG, AND C. HU. **Phase reconstruction of digital holography with the peak of the two-dimensional Gabor wavelet transform.** *Applied optics*, **48**(18):3308–3316, 2009. [71](#)
- [89] J.A. RICE. *Mathematical statistics and data analysis*. Duxbury press Belmont, CA, 1995. [73](#)
- [90] M.A. HERRÁEZ, J.G. BOTICARIO, M.J. LALOR, AND D.R. BURTON. **Agglomerative clustering-based approach for two-dimensional phase unwrapping.** *Applied optics*, **44**(7):1129–1140, 2005. [76](#)

REFERENCES

- [91] SERGIO FERNANDEZ, MUNTHER A. GDEISAT, JOAQUIM SALVI, AND DAVID BURTON. **Automatic window size selection in Windowed Fourier Transform for 3D reconstruction using adapted mother wavelets.** *Optics Communications*, **284**(12):2797–2807, 2011. [89](#)
- [92] C. TOMASI AND R. MANDUCHI. **Bilateral filtering for gray and color images.** In *Computer Vision, 1998. Sixth International Conference on*, pages 839–846. IEEE, 1998. [92](#)
- [93] C.S. McCAMY, H. MARCUS, AND JG DAVIDSON. **A color- rendition chart.** *Journal of Applied Photographic Engineering*, **2**(3):95–99, 1976. [95](#)
- [94] P.J. BESL AND N.D. MCKAY. **A method for registration of 3-D shapes.** *IEEE Transactions on pattern analysis and machine intelligence*, **14**(2):239–256, 1992. [111](#), [112](#)
- [95] C. YANG AND G. MEDIONI. **Object modelling by registration of multiple range images.** *Image and vision computing*, **10**(3):145–155, 1992. [111](#)
- [96] A.R. SPECHT, A.D. SAPPA, AND M. DEVY. **Edge registration versus triangular mesh registration, a comparative study.** *Signal Processing: Image Communication*, **20**(9):853–868, 2005. [111](#)
- [97] Y. TSIN AND T. KANADE. **A correlation-based approach to robust point set registration.** *Computer Vision-ECCV 2004*, pages 558–569, 2004. [111](#)
- [98] F. BOUGHORBEL, M. MERCIMEK, A. KOSCHAN, AND M. ABIDI. **A new method for the registration of three-dimensional point-sets: The Gaussian Fields framework.** *Image and Vision Computing*, **28**(1):124–137, 2010. [111](#)
- [99] M. GREENSPAN AND G. GODIN. **A nearest neighbor method for efficient ICP.** In *3-D Digital Imaging and Modeling, 2001. Proceedings. Third International Conference on*, pages 161–168. IEEE, 2001. [112](#)
- [100] G.C. SHARP, S.W. LEE, AND D.K. WEHE. **ICP registration using invariant features.** *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, **24**(1):90–102, 2002. [112](#)
- [101] T. ZINSSER, J. SCHMIDT, AND H. NIEMANN. **A refined ICP algorithm for robust 3-D correspondence estimation.** In *Image Processing, 2003. ICIP*

-
2003. *Proceedings. 2003 International Conference on*, **2**, pages II-695. IEEE, 2003. [112](#)
- [102] J. HO, A. PETER, A. RANGARAJAN, AND M.H. YANG. **An algebraic approach to affine registration of point sets**. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 1335–1340. IEEE, 2009. [112](#)
- [103] C.S. CHUA AND R. JARVIS. **Point signatures: A new representation for 3d object recognition**. *International Journal of Computer Vision*, **25**(1):63–85, 1997. [113](#)
- [104] A.E. JOHNSON AND M. HEBERT. **Using spin images for efficient object recognition in cluttered 3D scenes**. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, **21**(5):433–449, 1999. [113](#), [116](#)
- [105] J. FELDMAR AND N. AYACHE. **Rigid, affine and locally affine registration of free-form surfaces**. *International journal of computer vision*, **18**(2):99–119, 1996. [113](#)
- [106] K. BRUNNSTROM AND AJ STODDART. **Genetic algorithms for free-form surface matching**. In *Pattern Recognition, 1996., Proceedings of the 13th International Conference on*, **4**, pages 689–693. IEEE, 1996. [113](#)
- [107] J. SALVI, C. MATABOSCH, D. FOFI, AND J. FOREST. **A review of recent range image registration methods with accuracy evaluation**. *Image and Vision Computing*, **25**(5):578–596, 2007. [113](#)
- [108] G. STOCKMAN. **Object recognition and localization via pose clustering**. *Computer Vision, Graphics, and Image Processing*, **40**(3):361–387, 1987. [113](#)
- [109] Y.C. HECKER AND R.M. BOLLE. **On geometric hashing and the generalized hough transform**. *Systems, Man and Cybernetics, IEEE Transactions on*, **24**(9):1328–1338, 1994. [113](#)
- [110] H.J. WOLFSON AND I. RIGOUTSOS. **Geometric hashing: An overview**. *Computational Science & Engineering, IEEE*, **4**(4):10–21, 1997. [113](#)
- [111] D.P. HUTTENLOCHER AND S. ULLMAN. **Recognizing solid objects by alignment with an image**. *International Journal of Computer Vision*, **5**(2):195–212, 1990. [113](#)

REFERENCES

- [112] D. AIGER, N. J. MITRA, AND D. COHEN-OR. **4-points Congruent Sets for Robust Surface Registration.** *ACM Transactions on Graphics*, **27**(3):#85, 1–10, 2008. [113](#), [124](#)
- [113] M. BROWN AND D.G. LOWE. **Recognising panoramas.** In *Proceedings of the Ninth IEEE International Conference on Computer Vision*, **2**, page 5, 2003. [118](#)
- [114] O. SORKINE. **Least-Squares Rigid Motion Using SVD.** *Technical notes*, 2009. [120](#)
- [115] A.W. FITZGIBBON. **Robust registration of 2D and 3D point sets.** *Image and Vision Computing*, **21**(13-14):1145–1153, 2003. [121](#)
- [116] J.Y. BOUGUET. **Camera calibration toolbox for matlab**, 2004. [129](#), [135](#), [138](#)
- [117] J. SALVI, X. ARMANGUÉ, AND J. BATLLE. **A comparative review of camera calibrating methods with accuracy evaluation.** *Pattern recognition*, **35**(7):1617–1635, 2002. [135](#)
- [118] Z. ZHANG. **A flexible new technique for camera calibration.** *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, **22**(11):1330–1334, 2000. [135](#)
- [119] TA CLARKE AND JG FRYER. **The development of camera calibration methods and models.** *Photogrammetric Record*, **16**(91):51–66, 1998. [135](#)
- [120] F.S. MARZANI, Y. VOISIN, L.F.C.L.Y. VOON, AND A. DIOU. **Calibration of a three-dimensional reconstruction system using a structured light source.** *Optical Engineering*, **41**:484, 2002. [135](#)
- [121] I.D. REID. **Projective calibration of a laser-stripe range finder.** *Image and Vision Computing*, **14**(9):659–666, 1996. [135](#)
- [122] C. CHEN AND A. KAK. **Modeling and calibration of a structured light scanner for 3-D robot vision.** In *Robotics and Automation. Proceedings. 1987 IEEE International Conference on*, **4**, pages 807–815. IEEE, 1987. [135](#)
- [123] DQ HUYNH, RA OWENS, AND PE HARTMANN. **Calibrating a structured light stripe system: a novel approach.** *International Journal of computer vision*, **33**(1):73–86, 1999. [135](#)

- [124] Z. SONG AND R. CHUNG. **Use of LCD panel for calibrating structured-light-based range sensing system.** *Instrumentation and Measurement, IEEE Transactions on*, **57**(11):2623–2630, 2008. [135](#)
- [125] S. ZHANG AND P.S. HUANG. **Novel method for structured light system calibration.** *Optical Engineering*, **45**:083601, 2006. [135](#)
- [126] F. ZHOU AND G. ZHANG. **Complete calibration of a structured light stripe vision sensor through planar target of unknown orientations.** *Image and Vision Computing*, **23**(1):59–67, 2005. [135](#)
- [127] K. YAMAUCHI, H. SAITO, AND Y. SATO. **Calibration of a structured light system by observing planar object from unknown viewpoints.** In *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, pages 1–4. IEEE. [135](#)
- [128] Z. ZHANG. **A flexible new technique for camera calibration.** *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **22**(11):1330–1334, 2000. [138](#)
- [129] C. HARRIS AND M. STEPHENS. **A combined corner and edge detector.** In *Alvey vision conference*, **15**, page 50. Manchester, UK, 1988. [150](#)