FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Uncalibrated Stereo Vision applied to Breast Cancer Treatment Aesthetic Assessment

João Miguel Trigo Soares

Master in Informatics and Computing Engineering

Supervisor: Jaime dos Santos Cardoso, Doctor

Second Supervisor: Hélder Filipe Pinto de Oliveira, Eng.

July 21, 2011

Uncalibrated Stereo Vision applied to Breast Cancer Treatment Aesthetic Assessment

João Miguel Trigo Soares

Master in Informatics and Computing Engineering

Approved in oral examination by the committee:

Chair: Luís Filipe Pinto de Almeida Teixeira (Doutor)

External Examiner: Hugo Pedro Martins Carriço Proença (Doutor)

Supervisor: Jaime dos Santos Cardoso (Doutor)

13th July, 2010

Abstract

Breast Cancer Conservative Treatment (BCCT) is the most well-established method of treatment used around the world for an attempted remission of breast cancer. These treatments have evolved so as to provide the patient with similar life expectancy as that of other, more radical methods. So, now the attention is turning towards the assessment of the aesthetical results of this type of treatments, to identify key variables that could be changed in order to achieve a better overall outcome.

This work tries to use state-of-the-art algorithms for stereo matching in an attempt to recreate 3D models from patient's breasts that were subject to breast cancer conservative treatment. Those models will provide an increase in the accuracy of measurements that influence the assessment of the Breast Cancer Conservative Treatment aesthetic result, improving the *Breast Cancer Conservative Treatment.cosmetic result* software developed in *INESC Porto*. A detailed description of the objectives is given, paired with mathematical basis that support the reconstruction.

Stereo reconstruction depends on a very specific and sequential workflow, whereas failure in a previous step will render the reconstruction impossible. The main problems are identified, such as good quality rectification and the uncertainness of stereo matching in situations of low or repetitive texture and high image ambiguity.

It is shown that the state-of-the-art methods selected to take part of this test solved all problems with some degree of success except the stereo matching problem, most certainly due to the high ambiguity of the images that this project deals with. As such, it is concluded that it is necessary to search or develop a better suited algorithm for this type of images.

Resumo

O tratamento conservador do cancro da mama está mundialmente estabelecido como uma das mais viáveis hipóteses no tratamento de episódios de cancro da mama. Este tipo de tratamento evoluiu de maneira a providenciar ao paciente taxas de sobrevivência similares aos outros métodos mais radicais. Assim, tem-se em atenção agora a avaliação dos resultados estéticos destes tratamentos, de forma a identificar variáveis que possam ser manipuladas de forma a obter melhores resultados estéticos globais.

Neste trabalho tenta-se aplicar um conjunto de métodos reconhecidos do estado da arte do campo da correspondência estereofotogramétrica numa tentativa de verificar se é possivel recriar os modelos 3D do peito de pacientes que foram submetidos a tratamento conservador do cancro da mama. Esse modelos poderão providenciar o aumento da fiabilidade das medidas que influenciam a avaliação do resultado estético do tratamento, melhorando o *software Breast Cancer Conservative Treatment.cosmetic result* que está a ser desenvolvido no *INESC Porto*. Todos os objectivos que se pretendem para esse software no momento, bem como a matemática que dá suporte à reconstrução são descritas no presente documento.

A reconstrução *stereo* depende de uma linha sequencial de algoritmos, sendo que uma falha em um dos passos leva a que a reconstrução se torne impossível. Assim, identificam-se também quais os maiores problemas que existem, como a dificuldade de se conseguir uma correcta rectificação e a incerteza associada à correspondência de *píxeis* quando a imagem é constituída por pouca textura ou esta é repetitiva e onde existe muita ambiguidade.

Demonstra-se que os algoritmos do estado da arte que foram escolhidos para fazer parte do teste resolveram quase todos os subproblemas com algum nível de sucesso, excepto o problema de correspondência, devido à elevada ambiguidade associada às imagens que este projecto usou. Assim, conclui-se que é preciso procurar ou desenvolver um algoritmo mais adequado a este tipo de imagens.

Acknowledgements

Many thanks are due to Professor Jaime dos Santos Cardoso, my advisor, for all his willingness and precise guidance. Also, for Engineer Hélder Filipe Pinto de Oliveira, my co-advisor, for his friendship, support and for being available every day. Also to INESC Porto, specially to Visual Computing and Machine Intelligence group, the place where I developed the whole thing and where I made some new and good friends. In the medical support, thanks to *Hospital da Trindade*, Dr. André Magalhães and Prof. Maria João Cardoso. Last, thanks to Prof. Jorge Alves Silva and Prof. Rui Rodrigues for their very significant initial remarks when I started.

To my friends, those that bother to tolerate me despite all my widely recognised social problems and notable communication difficulties, thank you for 17 wonderful years. I won't write any name here, for you are too many and I won't risk leaving anyone aside. Special thanks for Bruno Mendes and Magui Pinto for being my support in the first days of this harsh journey.

Thank you, Jacinto Galvão.

My daily mates, Inês Coímbra Morgado and Paulo Silva: working together, having lunch and every other many things we have done will be specially missed and remembered. A word for all B320 people too.

My dear musicians, students, friends of music, Chico, Luís, Luís, Cristiana, Tixa, Miguel, etc, we have 12 memorable years together. Let's have another 12. Among these, two have a special place in my life. Thank you, Igor, for your discipline and for keeping me down to earth, and Duarte for doing quite the opposite, your madness somewhat refreshing. Also, pizza nights were excellent, and so were those cocktail nights I miss so much.

An enormous gratitude flag is to be raised to Ana Sara Morais for three long and difficult years, where so much happened and so much was wasted. It was hard but we got through it. Thank you.

Dear Micael Pinho, well...I'd rather not get started with this fellow. He knows better...

To my very special and strange family, mother, father and grandpa, thank you for having put so much love, dedication and effort in my bringing up, and you did it so well that You made me the man I became. Thank you for having given me everything I needed and even what I sometimes didn't dare asking for. I love you all deeply.

Carece aqui uma tradução do colocado acima:
À minha estranha mas muito especial família, mãe, pai e avô, obrigado por dedicarem toda a vossa vida a educar-me, fazer-me o homem que hoje sou, sofrerem por mim e por me darem tudo o que preciso, o que não preciso e mesmo o que seria completamente desnecessário. Amo-vos.

Now, last, but not at all least, thank you Di, for `System.OutOfMemoryException`.

“**a wild thesis appears...*”

The Internet

Contents

List of Figures

List of Tables

Abbreviations

BCCT	Breast Cancer Conservative Treatment
BCCT.core	Breast Cancer Conservative Treatment cosmetic results Software
INESC Porto	*Instituto Nacional de Engenharia e Sistemas de Computadores – Porto*
SNR	Signal-to-noise ratio
3D	Three-dimensional, usually referred to a three-dimensional model
IR	Infra-red
SVM	Support Vector Machine, a popular artificial intelligence algorithm for unsupervised classification
ELAS	*Efficient LArge-scale Stereo*, a stereo matching algorithm introduced by Geiger et al. [22].
PCA	Principal Component Analysis
SAD	Sum of Absolute Differences

ABBREVIATIONS

Chapter 1

Introduction

Computer vision is the area of computer science dedicated to study how can machines can mimic the human perception and interpretation of light and colors around themselves. Humans possess unique capabilities regarding vision, being able to recognize objects with enormous accuracy, read, remember places, even full cities, reconstruct and even have some tri-dimensional perception either inside one's mind or by drawing, sculpting, or others, and can go further and understand emotions or hidden meanings, patterns, clues. Nowadays, some of these aspects can be imitated by computers, for example, Optical Character Recognition (OCR), aerial 3D model building, motion capture, etc. Science has even improved those capabilities, like adding readings outside the visible spectrum and at many optical magnification levels. That permitted improvements like fingerprint recognition or medical analysis [47].

This work intends to apply computer vision state-of-the-art methods to assess the results of Breast Cancer Conservative Treatment (BCCT). These treatments are widely established as standard for they are as successful as more radical methods and result in improved aesthetic outcome [17, 50]. Subjects that undergo this type of surgery live, in 80% of the cases, 10 years further, so the aesthetic aspect is a factor they must cope with for a long time. For that reason, stable and robust evaluation is needed to identify and adapt key variables that would ease an improvement of that outcome, and that search has become a priority [13]. For now, this evaluation is almost always performed in a subjective way, with a panel of experts [24]. Some groups have tried and proposed various attempts to introduce objective measures, but those have continuously been found to be insufficient and are, at the moment, used to complement and/or correlate the subjective evaluation [1, 8, 10].

1.1 Context

The work from [8, 10] present the software *Breast Cancer Conservative Treatment.cosmetic result (BCCT.core)* that implements a set of measures for objective assessment. It is located in the position explained in the previous section, where it tries to correlate previous classifications performed by experts, in order to advance towards a semi-automated and complete evaluation tool. Through the use of semi-automatic feature extraction and machine learning algorithms, the overall assessment is predicted in a scale commonly used for the subjective evaluation, thus, directly related to the state of the art in the area [24].

For now, *BCCT.core* is capable of processing only the patient's frontal photographs based on a semi-automatic process (it requires the user to mark some feature points before attempting the final identification of all fiducial points and mammary contour). Lateral pictures are being studied in parallel to this work and may be introduced in a later development [41].

This project was undertaken for the group developing the software described. It aims to search for early developments on new measures that, when proven to be robust, can be added to the software.

1.2 Motivation

It is noted that *BCCT.core* can be improved by adding more measures. Particularly, it is intended to add dimensionality to the measures in order to drop the limitation of only measuring based on what can be seen in a frontal photograph. The capability of manipulating and measuring over 3D readings from the breasts can improve the accuracy and objectivity of the tool.

3D capabilities are recognised as having high clinical potential. However, the current techniques face two major problems: the high cost of required equipment and the need for specialized operators to work with them. Current techniques are based on specially designed cameras and hardware, mainly resorting to many lenses on the same camera or to laser scanners. Due to these special needs, 3D applications are considered pricey and are not commonly implemented, thus, the benefit of 3D modelling is not availed. Some of those methods are described in Chapter 4. As so, it is needed to search for new methodologies that ought to be cheaper and easier to use. The practitioners themselves must be able to use the tool without a hassle. It is the group's desire to develop a prototype of that methodology and compare it with those more expensive approaches commonly used.

If this 3D modelling proves to be robust enough, the software can be further extended to adopt surgical simulation. Those types of simulations are notable desired features, that would provide the means for anticipation of possible surgical options and outcomes, allowing better education of patients towards a more informed choice and the improvement of surgical techniques and skills.

1.3 Objectives

This project aims to research the state-of-the-art algorithms for stereo matching and reconstruction based on uncalibrated views from free-moving cameras when applied to female breasts, after breast cancer conservative treatments. The goal is to find out if the existing, most common and recent algorithms have some applicability to the type of scenes that *BCCT.core* deals with, and if 3D reconstruction can be performed directly from this methods.

This project's global objectives are summarised as such:

- Develop a robust high-level workflow for stereophotogrammetry reconstruction that ought to be based on current methods; it must possess enough parametrisation to accommodate the necessary changes those methods must undergo because of the type of images that are used.
- Test the highest number of possible methods, algorithms, and workflow changes, with emphasis on those that, for their characteristics and proven results, have higher probability of success.
- Verify what type of reconstruction is possible to get from these methods and possible applications to those.
- Establish a set of contributions and guidelines for future development in this direction, highlighting the solutions that are considered to have higher probabilities of improving the software.

1.4 Contributions

This work introduces a new approach to 3D body parts reconstruction that aims to be easy to use and affordable for general medical staff to use. The sate-of-the-art methods are glued together and the results are presented, so they can be used as orientation to future upgrades aiming the development of this tool.

Later on, some future improvements and guidance, based on the obtained results and on all the preliminary and regular study involved, are described, so it can help further developments.

Apart from this, some articles are being planned and developed, one of which is almost finished, to be submitted to *ORBS 2011*, with the working title *3D Model for Aesthetic*

Objective Evaluation after Breast Cancer Surgery using Infrared laser Projector. It's a new development, somehow apart from the approach being described here, based on active lighting methods (this work is concerned only with passive lighting methods).

1.5 Structure

Along with this chapter, this document has the following structure: Chapter 2 gives the fundamental knowledge about breast cancer and the software tool involved. Chapter 3 states the mathematical basis involved in this work and the general constraints involved. Chapter 4 describes the some related works that were found, studied and used as part of the proposed test solution. In Chapter 5 is shown the structure and steps of the proposed solution and the algorithms that were involved and then the results of applying the tests that were planned, followed by some evaluation. Chapter 6 wraps everything up along with the final statement about our hypothesis. Last, Chapter 7 provides some orientation as for the future of this project.

Chapter 2

Breast Cancer

Breast cancer is a worryingly common disease affecting mostly women, reaching such an enormous scale as to be considered a public health problem. It is known that one out of ten women will develop breast cancer at some point in her life.

> "Every day thirteen new breast cancer cases are detected and, in Portugal, four to five women die every day" - Mário Bernardo, *Liga Portuguesa Contra o Cancro*[1], in *O Público* 21st February, 2006. Free translation.

Many of the recognized risk factors are linked with oestrogens. Risk of contraction rises in early menarche (first menstrual cycle), late menopause and by obesity in post menopausal women [28] (see Figure 2.1). The cancer's incidence rises rapidly with age during the woman's reproductive phase and then increases slowly after 50 years of age. Childbearing, breastfeeding, oral contraceptives and hormonal therapy are some factors that might change the probability of contracting cancer. The exposure to radiation that occurs in mammographies has little, if any, effect on cancer incidence risk. Any possible effect is proven to be overcome by the demonstrated benefits of earlier detection of breast cancer [32].

Even though, 90% of breast cancer cases might be curable if detected in a relatively early stage and treated accordingly. Due to its incidence, prevalence, exposure and frequency, and also because of the enormous impact the disease inflicts in the feminine body, breast cancer is one of the most publicized diseases, fact that influenced the widespread of preventing and monitoring actions. For example, one may reckon that were deployed nation-wide screening programs in developed countries; also, family health care practice is now sensitive to the problem, taking into account adequate risk factors, undertaking

[1] The Portuguese Cancer League (*Liga Portuguesa Contra o Cancro* - LPCC) is a private non-profit non-governmental organization, declared by law of public utility (relevant social aim), founded in April 1941. http://www.ligacontracancro.pt/

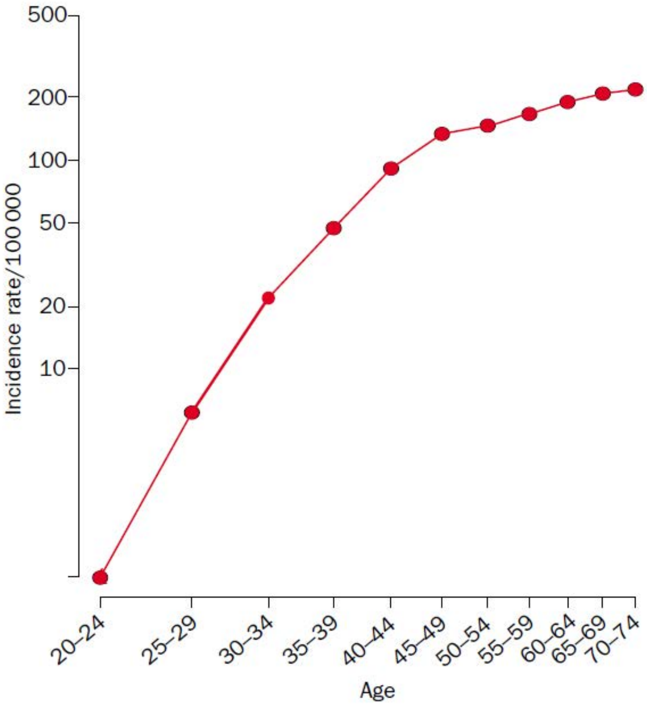


Figure 2.1: Age-incidence curve of breast cancer; log-log plot (from data for England and Wales 1983–87). Taken from [28].

the necessary actions or forwarding the patient to appropriate locations to ease a relatively in-time early diagnosis. All these factors combined may contribute to a successful treatment.

2.1 Beast Cancer Fundamentals

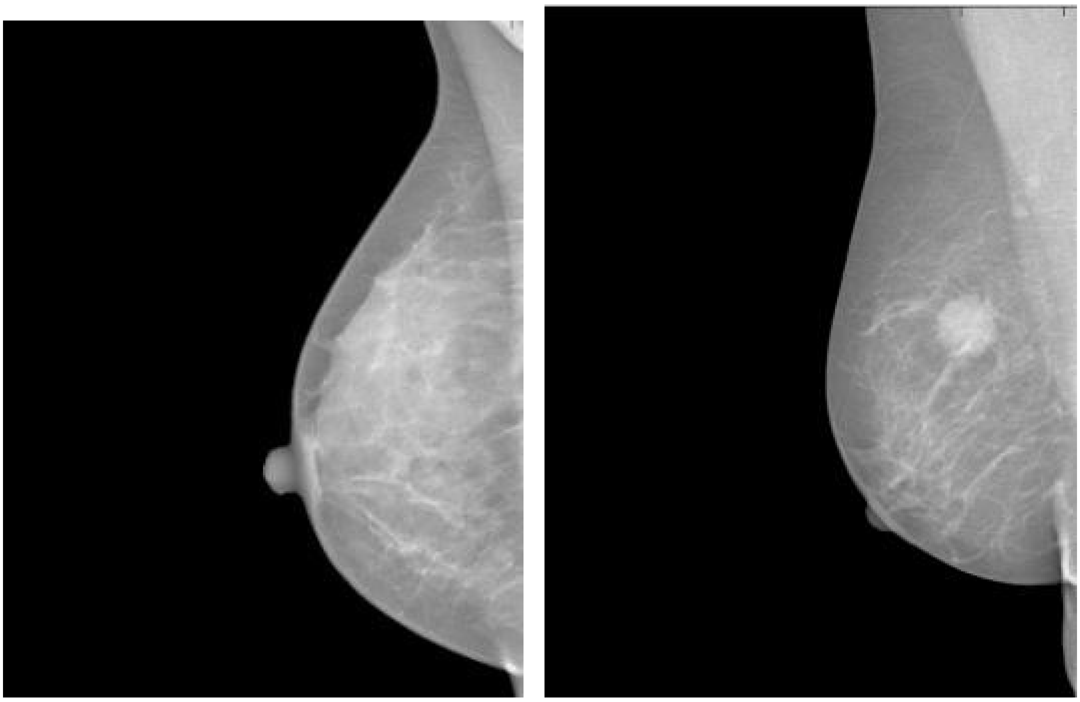
The origin of breast cancer, as well as all the other cancers, is yet unknown. Nevertheless, Kumar and Cotran [32] reckon three sets of influences that might be relevant:

- Genetic changes, as some genes have been proved to influence abnormal cell function and reproduction; this might come from family heritage or from sporadic genetic changes.
- Hormonal influences, as hormonal imbalance has a very important and recognised role, namely, risk of cancer rises with the exposure to high levels of oestrogen.
- Environmental variables, suggested by different cancer incidence depending on geographical location; also, high exposure to radiation or exogenous oestrogen may induce this and other types of cancer.

General risk factors are summarised in Table 2.1.

Factor	**Relative Risk**
Well-Established Influences	
Geographic factors	Varies in different areas
Age	Increases after age 30yr
Family history	
First-degree	relative with breast cancer 1.2-3.0
Premenopausal	3.1
Premenopausal and bilateral	8.5-9.0
Postmenopausal	1.5
Postmenopausal and bilateral	4.0-5.4
Menstrual history	
Age at menarche	<12yr 1.3
Age at menopause	>55yr 1.5-2.0
Pregnancy	
First live birth from ages 25 to 29yr	1.5
First live birth after age 30yr	1.9
First live birth after age 35yr	2.0-3.0
Nulliparous	3.0
Benign breast disease	
Proliferative disease without atypia	1.6
Proliferative disease with atypical hyperplasia	>2.0
Lobular carcinoma in situ	6.9-12.0
Less Well-Established Influences	
Exogenous estrogens	
Oral contraceptives	
Obesity	
High-fat diet	
Alcohol consumption	
Cigarette smoking	

Table 2.1: Breast cancer risk factors [32]



(a) Mammogram without any mass or abnormality

(b) Mammogram showing an abnormal mass

Figure 2.2: Typical mammogram screening images, with different diagnostics [40].

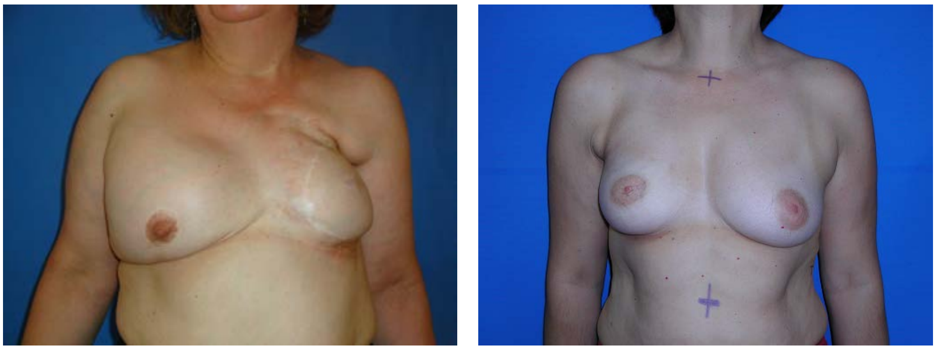
Commonly, breast cancer originates in the inner lining of milk ducts or near lobules that are linked and provide milk for those ducts. A cancer that has its origin in mammary ducts is known as ductal carcinoma, while one originating from lobules is known as lobular carcinoma. The treatment is dependent on size, rate of growth and other tumour characteristics and may include drugs, radiation, surgery and immunotherapy, but also of patient characteristics and personal opinion about possible surgical options [18]. Early detection is normally performed through mammographic screening, an X-Ray to both of the breasts, in order to try and observe any abnormal mass (see Figure 2.2). If such mass is observed, it is biopsied to verify if it is cancer, as it can be some simpler thing such as a cyst.

Male carcinoma is, on the other hand, very rare. It occurs with a frequency ratio of 1 to 125 female cases, normally in advanced age. It has a relatively fast infiltration rate, due to scarce amount of breast substance in the male. Unfortunately, nearly half of the cases spread beyond the initial location, to regional nodes or even beyond, before the cancer is detected [32].

2.2 Tumour removal

For many years, the accepted tumour removal treatment was a radical mastectomy, which consisted on the complete removal of the breast tissue and necessary surrounding lymph nodes. Then, the techniques evolved so as to reduce the amount of tissue needed to be removed.

The quadrantectomy was the first milestone, and was proven successful in the late 1970s. A median follow-up of 20 years survival rate found that the conservative approach and the more evasive and radical one were, in fact, equivalent [35]. Later studies found that Breast Cancer Conservative Treatment (BCCT) through the employment of the technique called lumpectomy, together with proper irradiation, has the same survival rates than those obtained by mastectomy alone. In the conservative treatment, a small portion of the breast containing the tumour and some additional tissue is extracted, along with a few lymph nodes. After that surgery, the patient may require additional radiotherapy. This technique has been verified to provide similar survival rates than those obtained with the radical mastectomy, although with a better cosmetic result [17, 50]. Comparison of the expected result from a mastectomy and a lumpectomy plus irradiation can be found in Figure 2.3.



(a) Result from a mastectomy.

(b) Result from a lumpectomy plus irradiation.

Figure 2.3: Two examples of possible results after the very different surgeries.

2.3 Cosmetic Assessment

The cosmetic assessment in conservative treatment is, as one can imagine, of utter importance for both the patient and the practitioner. As for the patient, the quality of life is at stake, sideways with the psychological effect this kind of surgical procedure inflicts on the subject itself, companion, family or friends. The patient may incur into distress due to a degraded effect the surgery might have in self-image and, consequently, self-esteem. As such, the practitioner should take into serious consideration the evaluation of the surgical aesthetic outcomes. This assessment should provide the means for the identification of variables that interfere with the cosmetic outcome so as to refine the surgical techniques.

The absence of a standard method for measuring the aesthetic outcome has been considered an obstacle in the assessment and evaluation of the techniques applied. Until recently most used methods were based on a subjective evaluation, made by one or more observers by visual inspection [24]. This type of evaluation presents the following problems:

- Exemption is not always possible, as the evaluating professionals were often involved in the treatment;
- The evaluation reproducibility is not usually high, as the same case can be evaluated differently and even the agreement between observers is only low or moderate;
- The potential invasion of the patients' privacy.

This precipitated the introduction of objective methods.

Many methods were attempted, based on measurements taken directly from the patient or from patient photographs, that are essentially based on asymmetries between treated and non-treated breasts [1, 13]. As of today, those objective methods are not yet well developed and accepted, and are often complemented with the established subjective evaluation, as on a search for correlation between both measures. In *INESC Porto*, *Institute for Systems and Computer Engineering of Porto*, a software is being developed, presented in [10] and [8], that attempts to create an independent and objective assessment tool, the *Breast Cancer Conservative Treatment.cosmetic result*.

BCCT.core classifies the aesthetic outcome of BCCT into excellent, good, fair, and poor classes. To achieve the classification, first, a concise representation of a BCCT image is obtained based on the aspects mentioned before. These measurements are preceded by the semi-automatic localization of fiducial points (nipple complex, breast contour and jugular notch of sternum) on the digital photographs [9, 11]; measures are then supported on these fiducial points. After this, all the measures are automatically converted onto an overall objective classification of the aesthetical outcome, using the SVM classifier, trained to predict the overall aesthetical classification on the aforementioned scale of four classes. The software was created based on a database of 120 patients, with frontal 4M pixel images, with or without auxiliary reconstructive surgery and, at least, one year over all treatments. Twenty-four practitioners were selected, located in 13 different countries, and asked to classify the pictures. From this point on, a Support Vector Machine classificator was used to try and correlate some asymmetric measures with the panel result.

Chapter 3

Background

This chapter aims to describe briefly the basic math evolved in the recovery of depth information from images. The reader is conveniently directed to appropriate sources for thorough information and formal proof.

3.1 Camera model

We start by defining our camera model and by defining some of the basic math related to image acquisition. The information of this section is based on Bradski and Kaehler [6] and Szeliski [47]. It is common in the literature to use the pinhole camera model, as it simplifies the math evolved and provides an intuitive view of how the image plane is formed. Imagine, for this model, that the projection is based on light being projected in a plane, entering from a small hole, so small only a ray can enter for each point of the scene (Figure 3.1). This model can be rearranged so that the math comes out easier and so that the image plane is not inverted in relation to the scene (Figure 3.2).

The projection that occurs on a camera when transforming 3D coordinates to 2D plane coordinates is a true 3D *perspective*, as can be found in [47]. So, any point Q will be transformed into an image point q (in inhomogeneous coordinates) as follows:

$$q = \mathscr{P}(Q) = \begin{bmatrix} X/Z \\ Y/Z \\ Z/Z \end{bmatrix} \tag{3.1}$$

However, this transformation is based on an ideal pinhole camera, which is extremely difficult, if not impossible, to create. Once the 3D points are transformed through the pinhole, it is still needed to transform those coordinates into sensor space coordinates (camera coordinates). A camera can be simplified as a sensor and a lens, where the lens gather the maximum amount of light it can and redirects that light into the sensor. An example of that process can be found in Figure 3.3. So, a point $q_s = [x_s, y_s]^T$ is formed

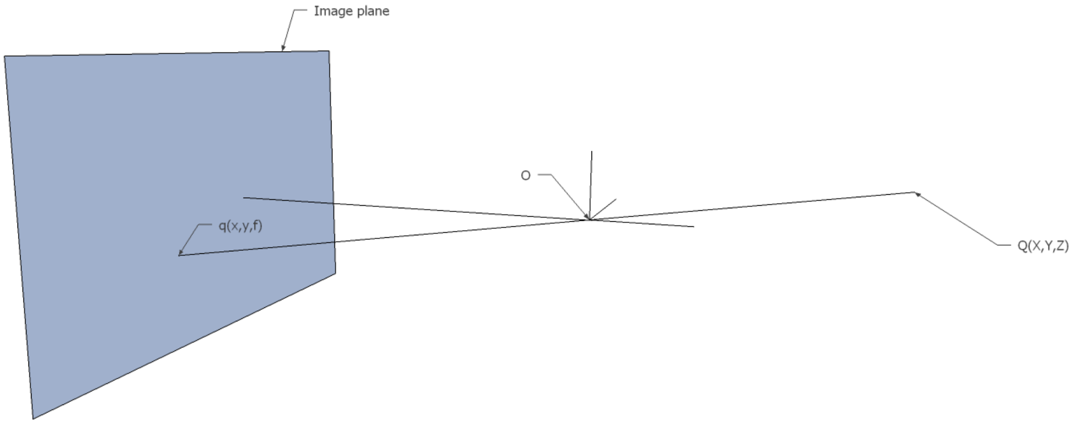


Figure 3.1: Pinhole camera model, the Z axis shall be considered the optical axis, f stands for the focal length, O is the origin of coordinates, Q is a point in space and q is the projection of that point in the image plane. Please note the change in coordinates from Q to q.

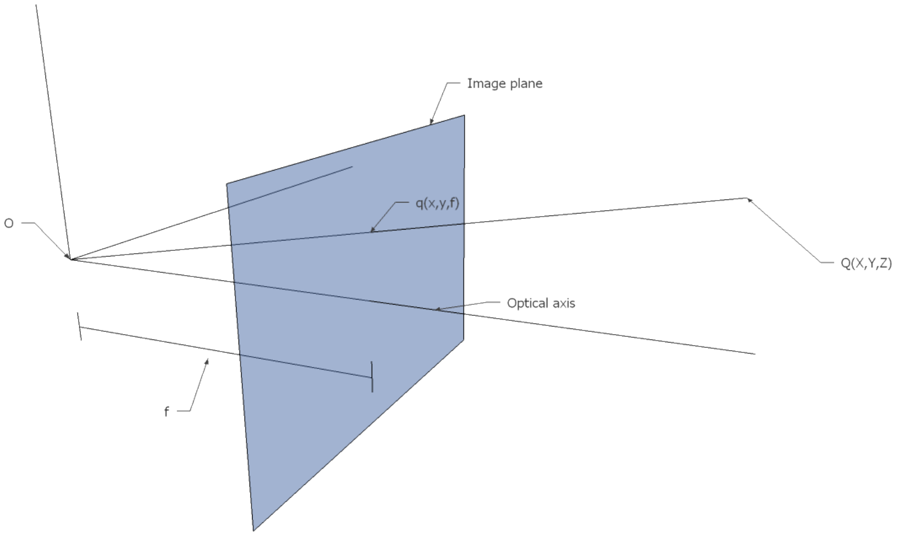


Figure 3.2: Pinhole camera model with front facing panel, for simplification.

from a scene point $Q = [X,Y,Z]^T$ by some simple relations:

$$x_s = f\frac{X}{Z} + c_x \qquad\qquad y_s = f\frac{Y}{Z} + c_y \tag{3.2}$$

Here, x_s and y_s are image coordinates, f is the focal length and c_x and c_y are the coordinates for the center of the image plane . These two additional parameters result from the possible displacement there might exist between the sensor and the optical axis. Perfect aligning the sensor and the optical center of the lens would require a level of precision difficult to acquire, especially in low-end cameras. There is also the possibility that the sensor is not perpendicular enough to the optical axis as to create a skew, s, between the sensor axis. So, as for now, we can build a the camera calibration matrix as such:

$$\mathbf{K} = \begin{bmatrix} f & s & c_x \\ 0 & af & c_y \\ 0 & 0 & 1 \end{bmatrix} \tag{3.3}$$

where a is the aspect ratio of the image. In practice, for many applications, it can be assumed that $a = 1$ and $s = 0$, resulting in the final calibration matrix:

$$\mathbf{K} = \begin{bmatrix} f & 0 & c_x \\ 0 & f & c_y \\ 0 & 0 & 1 \end{bmatrix} \tag{3.4}$$

This equation allows for correcting the ray coordinates for each point in space in the image plane, through the following relation:

$$\hat{x} = \mathbf{K}^{-1}x \tag{3.5}$$

Now that the form of the calibration matrix is agreed, it can be put together with the camera extrinsics, the camera orientation in space (the camera position regarding the space coordinates). That position can be parametrized by a rotation matrix and a translation vector, respectively, **R** and **t**. As such, one can define the camera matrix:

$$\mathbf{P} = \mathbf{K}\begin{bmatrix}\mathbf{R}|\mathbf{t}\end{bmatrix} \tag{3.6}$$

It is preferable to use an invertible $4x4$ matrix, defined as such:

$$\tilde{P} = \begin{bmatrix} \mathbf{K} & \mathbf{0} \\ \mathbf{0}^T & 1 \end{bmatrix} \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix} \tag{3.7}$$

With this matrix, a point in space can directly be transformed into screen coordinates through the equation

$$\bar{q} \sim \tilde{P}\bar{Q} \tag{3.8}$$

where $\bar{Q} = (x, y, z, 1)$ and $\bar{q} = (x_s, y_s, 1, d)$, d is the point's disparity and $\sim$ denotes equality up to a scale.

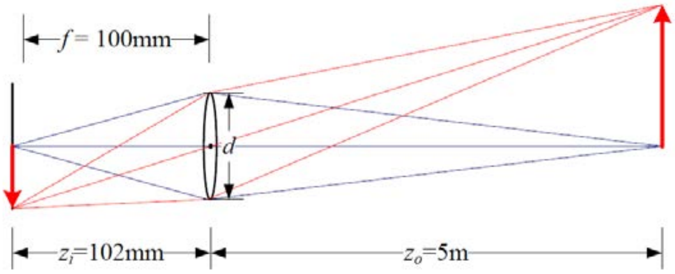


Figure 3.3: Simplified camera and lens example; note the possible similarities between this example and the pinhole model in Figure 3.1. Taken from [47].

3.2 Binocular disparity and stereo vision

Part of the problem that is addressed in this project is that of binocular disparity, namely the different displacement objects undergo when seen through different viewpoints. The human brain uses this process to identify distances to objects. The same process is used in computer vision, where similar features are matched and their disparity is calculated. Then, with some computation and provided that there exist some knowledge between the cameras (or that knowledge is inferred, as it will be seen in Section 3.3.2) it is possible to reconstruct the scene up to a metric reconstruction, where the lengths and distances of that reconstruction have a direct relation to real world measures. If such knowledge is unknown, every reconstruction is relative to a trivial projective transformation, thus the measures of that projection might have no direct relation to real world measures.

Based on Hartley and Zisserman [25], Pollefeys [43], Szeliski [47], it is possible to build a mathematical model in order to relate two views (see Figure 3.4). Although we do not know the exact position of point p, we know that it must be located somewhere in the ray cast between the camera centres and that point, that pass through the image plane in each one of the points x and x_1. After some geometrical relations and based on the camera matrix discussed in the previous section, it is possible to define that there is a matrix $\mathbf{E}$, 3x3 of rank-2, that relates every corresponding point as such:

$$\hat{x}_1^T \mathbf{E} \hat{x} = 0 \tag{3.9}$$

where $\mathbf{E}$ is defined as:

$$\mathbf{E} = [\mathbf{t}]_\times \mathbf{R} \tag{3.10}$$

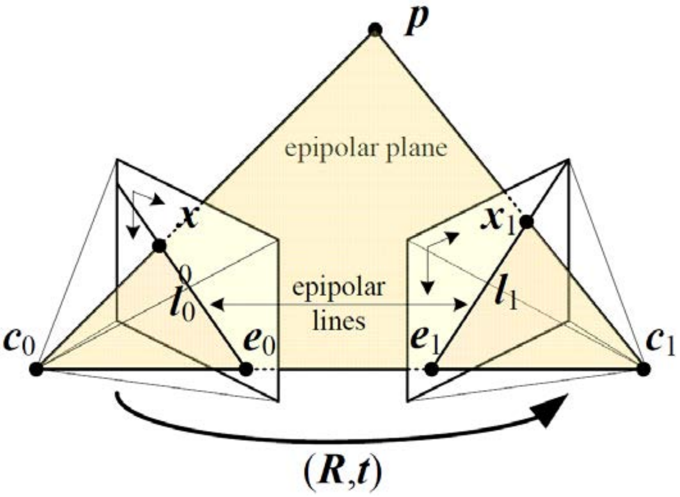


Figure 3.4: Geometric model for two-view geometry based on the pinhole camera model. The projection of point p is, respectively, x and x_1 in the left and right view, c_0 and c_1 are the camera centers, and e_0 and e_1 the epipoles, finally l_0 and l_1 are the epipolar lines that pass through x and x_1, respectively, and are contained in the epipolar plane for point p. Note that all epipolar lines and planes go through the epipoles. Also note that, between the two cameras, it is assumed to exist a rotation and translation associated, $(\mathbf{R},\mathbf{t})$. Taken from [47].

This is, of course, if both **R** and **t** are known. In this project's case of study, those are not known, as it deals with free-moving hand-held cameras. Also, as it is an uncalibrated environment, there is no easy access to camera parameters and the cameras' calibration matrix. In such situation, it is possible to define a new matrix **F**, the fundamental matrix, that still respects the epipolar constraint, as follows:

$$\hat{x}_1^T \mathbf{E} \hat{x} = x_1^T \mathbf{K}_r^{-T} \mathbf{E} \mathbf{K}_l^{-1} x = x_1^T \mathbf{F} x = 0 \qquad (3.11)$$

The matrix defined in (3.11) [25, 43] has many applications, and it is of utter importance, as we will see through this chapter. **F** is normally found using, at least, 8 matches in both images, solving a system of equations using Singular Value Decomposition[1]. It can, anyway, be estimated using only 7 correspondences solving for a system of non-linear equations, as this matrix has only seven degrees of freedom. If more than eight points are available, then it is possible to minimise the effect the noise can introduce while estimating **F**. If using automatic established correspondences, then there might exist some outliers (wrongly-matched points) that must be addressed. Those techniques will be discussed later.

The process of stereo matching consists on finding correspondences for the maximum number of pixels in each image. Assuming a low baseline (the displacement, or disparity, between the two views is much smaller than the distance to the objects on the scene), it is possible to assume that most of the pixels from either images will match, in other words,

[1] Simple and intuitive explanation in http://www.cs.unc.edu/~marc/tutorial/node54.html

almost all scene points are visible in both views. It is possible to ease the correspondence problem and the computation of correspondent distances by pre-aligning the views, a process called *rectification*, so that corresponding epipolar lines are horizontally aligned. That way, the stereo matching problem is reduced to a one-dimension search and the math is reduced to what is summarised in Figure 3.5. From there it is possible to note that the problem, as it is posed here, assumes that the focal length f is the same for both views. That is common, as even with autofocus, in low baseline problems that value tends to be the same [6], even though the state-of-the-art methods do try to transform that (and other) intrinsic parameter, assuming it is not shared by both the views, so that our model holds. By simple triangle similarity, it is possible to verify that:

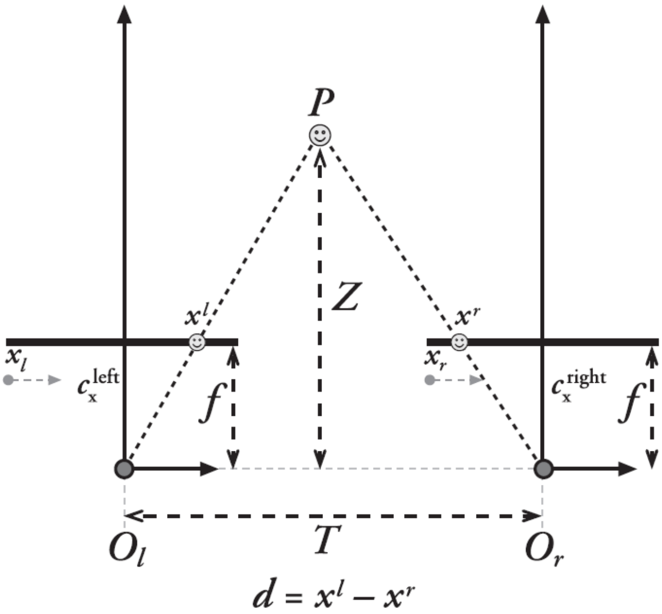


Figure 3.5: Disparity search summary assuming a low baseline ($T << Z$). The search space for correspondences is limited to the x-dimension and the disparity value is very simple to compute. Taken from [6].

$$\frac{T-(x_l-x_r)}{Z-f} = \frac{T}{Z} \Rightarrow Z = \frac{fT}{x_l-x_r} \Rightarrow Z = \frac{fT}{d} \tag{3.12}$$

3.3 Rectification

From the previous section it is noted that rectification is a very important and necessary step in the reconstruction process. This step is highly based on the epipolar geometry already discussed, where the main goal is to find a way to alight horizontally corresponding epipolar lines.

3.3.1 Based on known objects

If a known object is placed on the scene, visible to both views, and with identifiable and matchable features, then it is possible to recover the pose of the two cameras from those measures. In this special case, it is possible to find both the translation vector and the rotation matrix for the cameras concerning that object, and (3.10) is applicable. In this case, the rotation necessary to align the epipolar lines is known, and by Bouguet's algorithm, the rotation is split between the right and left cameras. That way, it is possible to minimize the reprojection distortion and maximize the common viewing area [6].

3.3.2 Uncalibrated rectification

The relation between the fundamental matrix **F** and the rectification process is that of finding a pair of homografies that will transform the images so that they become rectified [6, 43, 47]. In the rectified case, the fundamental matrix has the form [21]:

$$\mathbf{F} = [u_1]_\times = [(1,0,0)]_\times = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \tag{3.13}$$

where $[x]_\times$ means the cross-product of the vector x. In order to rectify a pair of images it is necessary to search for a pair of homographies so that the epipolar constraint 3.10 of a rectified pair is verified. If H_r and H_l are such homographies, then it is defined as such:

$$(H_r x_r)^T [u_1]_\times (H_l x_l) = 0 \tag{3.14}$$

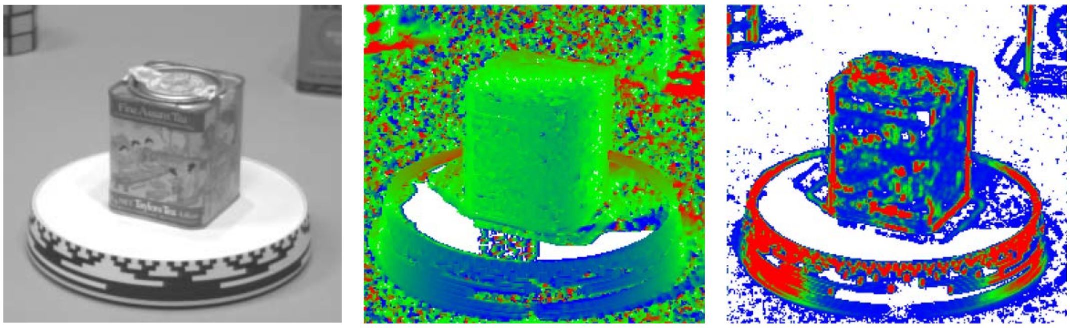
The way those homographies are searched is algorithm-dependant, and will be discussed later in the appropriate section.

3.4 Stereo matching

The stereo matching process is almost always difficult and prone to errors, mainly because it is ill-posed [51]: due to the possible colour repeatability, its solution might not be unique. There will be multiple situations where it won't be possible to define, with confidence, the correct correspondence between pixels. The colour values in some regions might be equal or very similar in some areas, thus preventing a confident matching. The classical algorithm for finding pixel correspondences is based on simple correlation, also called "brute force" algorithm. It can be simply described as such: for each pixel in a "base" image, calculate a weight function in a window around that pixel; then, for each pixel in the same row of the "target" view, calculate the weight of the window around it. The "target" pixel that has a value closer to that of the "base" pixel is, potentially, the best match. It is not difficult to see that this algorithm can greatly suffer from problems in low or repeated textured scenes, where there would be similar valued pixels and where an unique match is not possible [47]. Common window evaluation functions are, for example:

$$\begin{array}{l} \text{Normalized correlation: } \frac{\sum_x \sum_y L(x,y)R(x,y)}{\sqrt{\left(\sum_x \sum_y L(x,y)\right)^2 \left(\sum_x \sum_y R(x,y)\right)^2}} \\ \text{Sum of squared differences: } \sum_x \sum_y \left(L(x,y) - R(x,y)\right)^2 \\ \text{Sum of absolute differences: } \sum_x \sum_y |L(x,y) - R(x,y)| \end{array} \quad (3.15)$$

where L and R are, respectively, the left and right images.



(a) Original image.

(b) Estimated depth map (blue is closer).

(c) Estimated confidence (red is higher).

Figure 3.6: Uncertainty in stereo matching, as can be seen from the last image. Taken from [47].

3.5 General uncalibrated reconstruction workflow

So, wrapping everything up, the basic workflow can be seen as the following steps sequence:

1. Search for feature points;
2. Try and match the highest number of those feature points in a robust way;
3. Use the matched features to find the fundamental matrix F with the lowest possible error (with the lowest inconsistency within the largest possible set of inliers);
4. Use that matrix and the inlier matches to compute rectification homographies, and then rectify both views;
5. Perform the largest stereo matching possible, with the highest confidence available for each match;

6. If camera parameters were found somewhere in the fundamental matrix estimation or rectification algorithm, use them to try and return a quasi-euclidean reconstruction, else return the projective reconstruction.

Failure in any previous step (either due to lack of features, or to failure in matching enough of them) will invalidate the remainder of the reconstruction, so each step is critical in the process. This workflow is visually drawn in Figure 3.7.

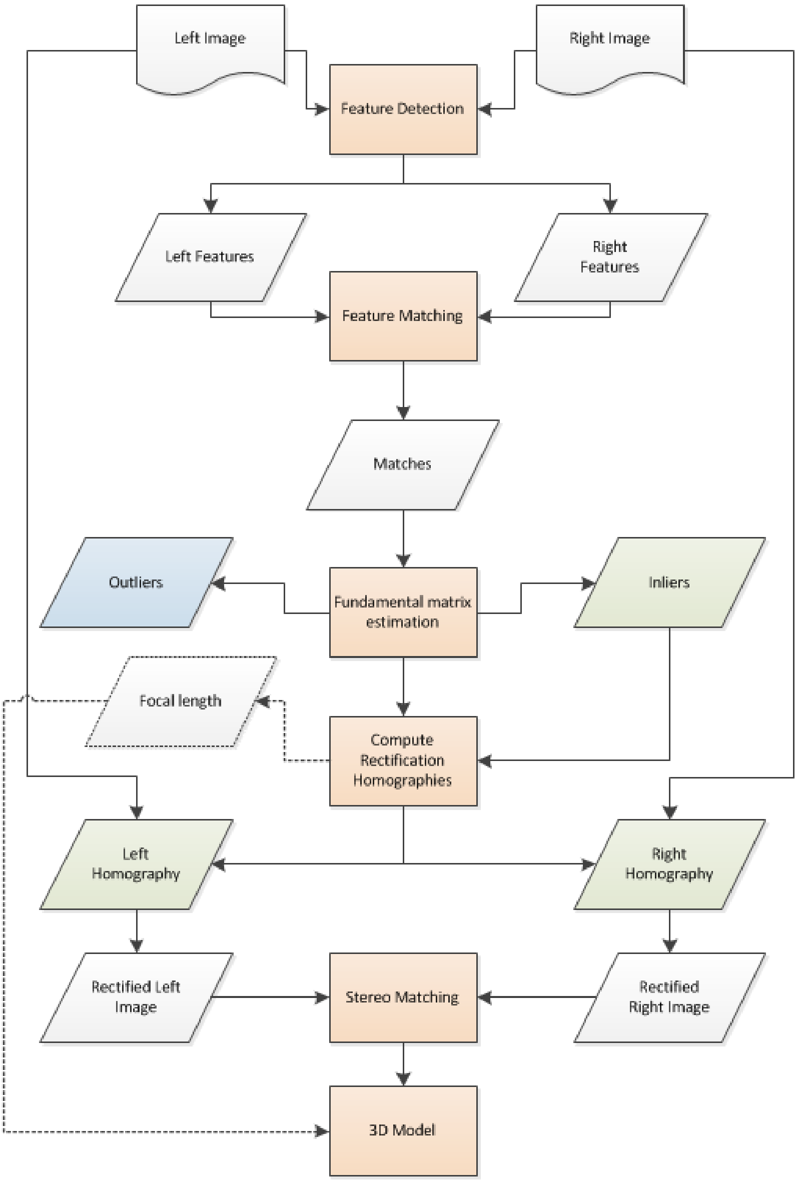


Figure 3.7: Stereo reconstruction generic workflow.

Chapter 4

Literature Review

The 3D reconstruction problem has been studied for several decades [57]. There is thorough work developed in the literature about this problem, with wide applications, ranging from world comprehension for robotics, aerial building reconstruction, to virtual tours for on-line tourism.

This section aims to present a brief summary of the most interesting contributions found in the literature for solving the workflow presented in Section 3.5: feature detection and matching, rectification and stereo matching. Because of the large amount of algorithms and applications available, it is necessary to select, *a priori*, the candidate applications that might prove to be success. For that reason, it is convenient to present Figure 4.1 to enhance the reading of this chapter with better understanding of what properties and structure this project's scene deals with.

4.1 Feature detection

One of the best contributions which can be found in the literature as far as explanation and comparison of feature detection and matching algorithms are concerned is that of Tuytelaars and Mikolajczyk [48]. There are also detailed explanations in the previously referred books, namely Bradski and Kaehler [6], Szeliski [47]. A feature can be defined as an image point or pattern that is relevant, different from its neighbour, and for the purpose of this project, especially interesting if visible in both views and mathematically matchable. Those features can be isolated points (pixels), but might also be edges or even other portions (e.g. blobs). Matching normally occurs either through measures around a window or through the computation of specially designed descriptors. Window matching is normally not invariant to transformations, so descriptors are created trying to create that invariance.

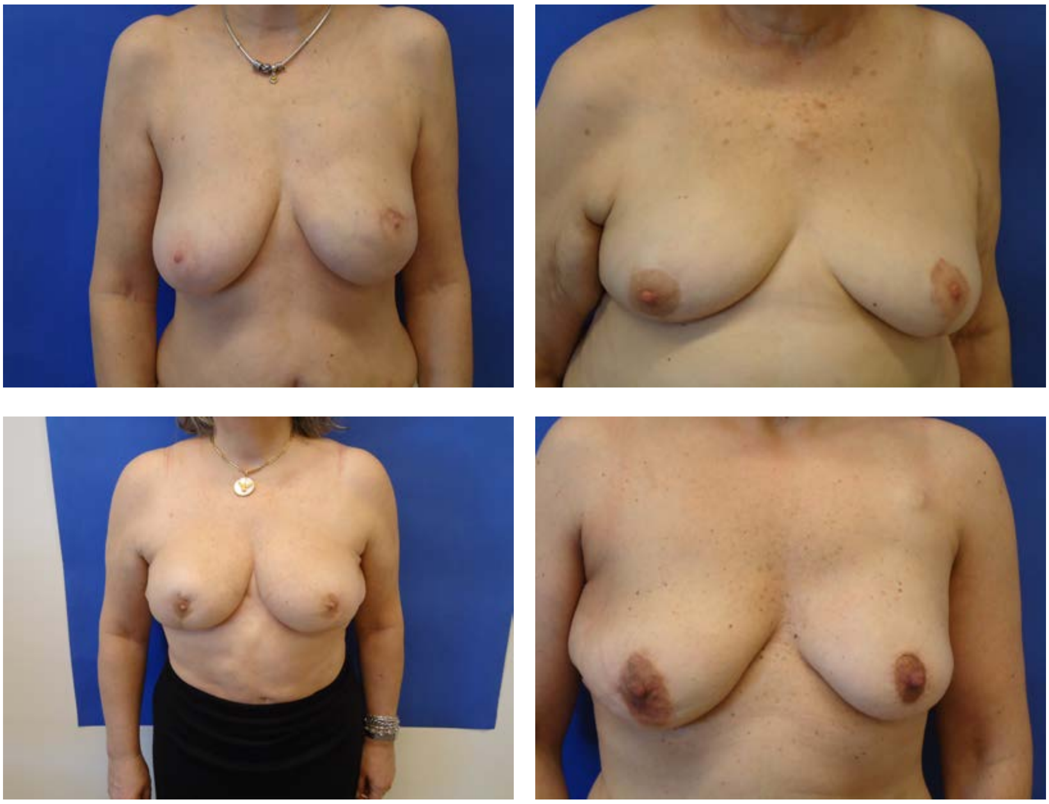


Figure 4.1: Part of the test images used in this project. Note the texture repeatability, the commonly unreasonable background and the lack of common distinctive features, like corners.

Notice that this project is dealing with free-moving cameras. As such, it is not known, *a priori*, what type of transformations the scene might incur from one view to another. Thus, the ideal type of features must be invariant to the highest number of transformations possible. Good features have associated a set of commonly desired properties [48]. For this project, all but one property is necessary: *Efficiency*, for it is not desired to create a time-critical application, detection can take some time to be performed. So, the features to be looked for must have these properties:

- *Repeatability*, once most of the features detected in one view should be visible and detected in the other;
- *Distinctiveness*, as features should be informative enough to allow an accurate and robust matching between all features present in an image;
- *Locality and Accuracy*, in Section 3.1 it was learned that the geometrical model would require local points to be matched, and so the detected features should have an accurate location in the image plane;
- *Quantity*, even though a small minimum number of correspondences are necessary, a larger number of features would reduce noise-induced errors.

4.1.1 Corner detectors

Corners, in a planar imaging sense, are points of high curvature. They might not correspond to actual corners in 3D objects. The most classical algorithm for corner detection is that of Harris and Stephens [23], commonly called **Harris corner detector**. It is based on finding large changes in the derivatives signal of the image plane. A large change in both directions is a strong indication there might be a corner in that location. This algorithm was then updated to a scale-invariant version and then to an affine-invariant version [48].

SUSAN [46] (*Smallest Univalue Segment Assimilating Nucleus*) is another commonly used algorithm. The principle is simple: to define a circular patch around each pixel of the image, being the center called the *nucleus*; each pixel in that radius is assigned as having similar or different intensity values to that of the *nucleus*. So, a corner is defined as a *nucleus* where the circular neighbour has less than a defined threshold of similar pixels. Commonly, that value is somewhat near 25%.

Observing the images presented at the beginning of this chapter, a good performance from the corner detectors is not expected. There are, visually, not many points of high curvature salient in the scenes that this project deals with.

4.1.2 Blob detectors

Although not so well defined as corners, blobs constitute a feature where corners are not available. A blob might be imagined as a pixel or region of pixels that are "darker" or "brighter" than the surrounding neighbours.

One possible algorithm to find blobs that are invariant to some transformations (namely, affine) is the *Hessian-Laplace detector*. This algorithm is based on a matrix of second order derivatives of the image intensities, created around points detected by the determinant of Hessian [48, 55]. Features are detected when the trace of such matrix attains a local maxima.

Another possibility is through the use the Difference of Gaussian detector type. This algorithm requires only a convolution on the images in a constant matter, depending on a Gaussian smoothing parameter. The image scale-space is developed with Gaussian filters at different scales. The features are detected searching for extremal values by comparison with the neighbours of the pixels at the various scales and octaves [34] (Figure 4.2).

Regions might also be searched through *Maximum Stable Extremal Regions* [36]. The objective is to search for regions of distinguishable stable and invariant shape. To detect those regions, the image is threshold at all possible gray levels. While that threshold is changing, the area of connected components (pixels) is monitored. Regions where that area changes below a defined threshold are defined as maximally stable.

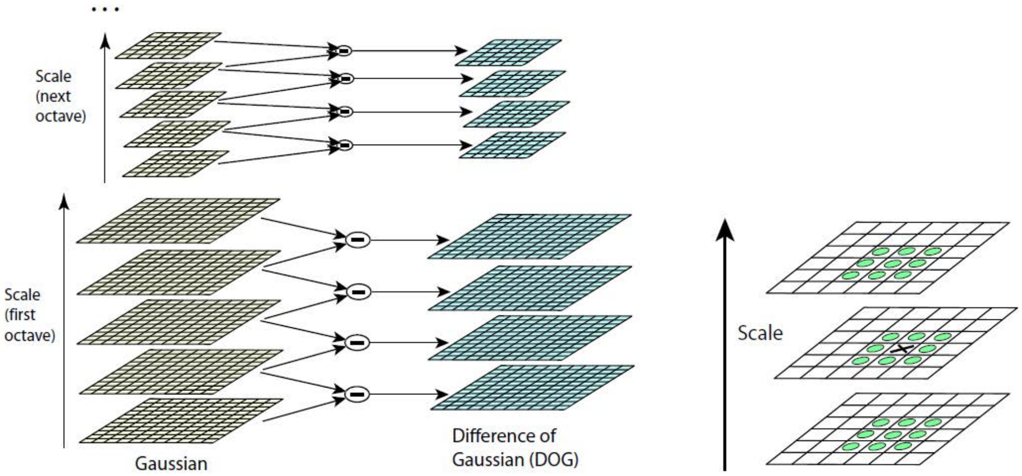


Figure 4.2: Difference of Gaussian space-scale pyramid visual description (left) with extremal pixel finding by comparison to the neighbours in each level (right). Taken from [34].

4.1.3 Edge detectors

Edges are expected to greatly vary from one view to another, as a rotation and translation almost certainly would hide and/or transform the detected edges, invalidating the matching tasks. For that reason, the edge detectors were, concerning the time span of this project, left behind.

4.2 Feature Descriptors

After finding all features, the next task is to try and match them with features in the other view. The most robust way to do that is to define and compute a descriptor for each feature found, and then search for correspondences on those descriptors. It is common for the features and the image to, at least, suffer some translation and rotation. It is also possible that there is some type of affine transformation involved. As such, it is of great importance to use robust feature descriptors in this project.

Brown et al. [7] defined the *Multi-Scale Oriented Patches*, a descriptor based on normalized intensity patches. It defines a 5x5 patch around the feature location, and the normalization process transforms the intensities so that their average is zero and the variance is one. It performs well whereas small image transforms occur.

Lowe [34] developed a robust descriptor along with the *SIFT* algorithm. Consider a window of 16x16 around the feature location, using the level of the Gaussian filter in witch the feature was detected. The descriptor is then based on computing the gradient in each of the pixels contained in that window, downgraded by a Gaussian fall-off function (to reduce the effect of those pixels far from the feature center). Then, in each quadrant of that window, it develops a orientation histogram by adding the value of each pixel to one of the eight orientation beans using trilinear interpolation in a $2 \times 2 \times 2$ histogram (4 eight-bean histogram for each quadrant). That forms a 128-Dimension vector that is then normalized to unit-size, clipped to 0.2 and then re-normalized.

Ke and Sukthankar [27], inspired by the *SIFT* descriptors, proposed a more simple approach. Around each feature, compute the x and y derivatives around a patch of 39×39 pixels. That amounts for a 3042-dimensional vector that is then downsampled to a 36-dimensional one by Principal Component Analysis (PCA) [47, Section 14.2.1 and Appendix A.1.2].

Last, Mikolajczyk and Schmid [37], again by considering the *SIFT* descriptors, proposed to change those a little. So, instead of using the usual square-defined descriptor, their algorithm uses log-polar coordinates (still, centred in the feature point) to define the histogram structure. As such, the bin space are of radius 6, 11 and 15, with eight bins per coordinate, except for the center coordinate, for a grand total of 17 bin-coordinates, each with 16 orientation bins. The 272-dimension resulting vector is then mapped to a 128-dimension one by PCA trained on a large database.

After detecting the features and obtaining the descriptors, it is still needed to define a matching strategy. The most commonly used is simple Euclidean distance between all "target" descriptors, provided that a certain distance threshold is attained [47]. However, nearest-neighbour methods, for example, are suitable if efficiency is crucial.

4.3 Rectification

There is extensive work presented in the literature about uncalibrated rectification. The possibility of reconstructing scenes from sets of frames or pictures is much appreciated and is extensively studied. Here only recent and recognised algorithms will be presented, among many others that can be found in the literature.

As it can be seen in Section 3.3.2, the process of planar rectification is to find a pair of homografies that will align the epipolar lines. So, it is possible to establish these assumptions:

1. All epipolar lines are parallel to the x axis;

2. All image features and points have the same corresponding y coordinate.

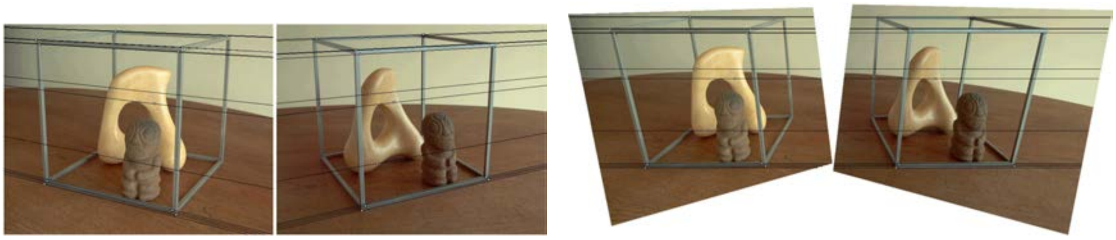
The algorithm from Loop and Zhang [33] is based on finding a pair of homographies that will maintain the epipolar constraint. The goal is to find those homographies by establishing a set of simpler transforms, computing each component to achieve a desired effect and satisfy some conditions. So, the homography **H** is decomposed in two homographies, so that:

$$\boldsymbol{H} = \boldsymbol{H_a}\boldsymbol{H_p} \tag{4.1}$$

being H_a an affine transformation and H_p a projective one. Then, H_a is decomposed as

$$\boldsymbol{H_a} = \boldsymbol{H_s}\boldsymbol{H_r} \tag{4.2}$$

where $\boldsymbol{H_s}$ is a shearing transformation and $\boldsymbol{H_r}$ a similarity. All these transformations are computed through minimization with criteria specific for each objective. This algorithm's result can be found on Figure 4.3.



(a) Original unrectified pair with some epipolar lines

(b) Result of the rectification process, with epipolar lines aligned

Figure 4.3: Results from Loop and Zhang [33] rectification algorithm.

Hartley and Zisserman [25] created a method that is based on the process of relocating the epipoles in both images. So, the algorithm attempts to transform the epipoles so that their location is set at infinity (with the last coordinate as 0, in homogeneous coordinates). Starting from a set of matches $x_i \leftrightarrow x'_i$, which can never be less than seven (though more than that is preferred). Compute the fundamental matrix $\mathbf{F}$ from those matches and find the epipoles e and e' so that $e'\mathbf{F} = 0$ and $\mathbf{F}e = 0$. Then, compute the projective transformation **H'** that maps the epipole e' to infinity, $(1,0,0)^T$. With that, find the matching projective transformation H that minimizes the least-squares distance

$$\sum_i d(\mathbf{H}x_i, \mathbf{H'}x'_i). \tag{4.3}$$

An example of the result from this algorithm can be found in Figure 4.4.

Fusiello and Irsara [21] developed recently a new rectification algorithm that attempts to get close to that of the euclidean rectification. This method attempts to use some notions of autocalibration for trying and estimating the camera parameters (that are assumed to be equal for both views, as it is the case for the images used – based on the EXIF tags). From a set of correspondences $m_l^j \leftrightarrow m_r^j$ it defines a Sampson error function for each correspondence

$$E_j^2 = \frac{({m_r^j}^T \boldsymbol{F} m_l^j)^2}{(\boldsymbol{F} m_l^j)_1^2 + (\boldsymbol{F} m_l^j)_2^2 + ({m_r^j}^T \boldsymbol{F})_1^2 + ({m_r^j}^T \boldsymbol{F})_2^2} \tag{4.4}$$

where $(\bullet)_i$ is the i^{th} component of the normalized vector. The homographies are then

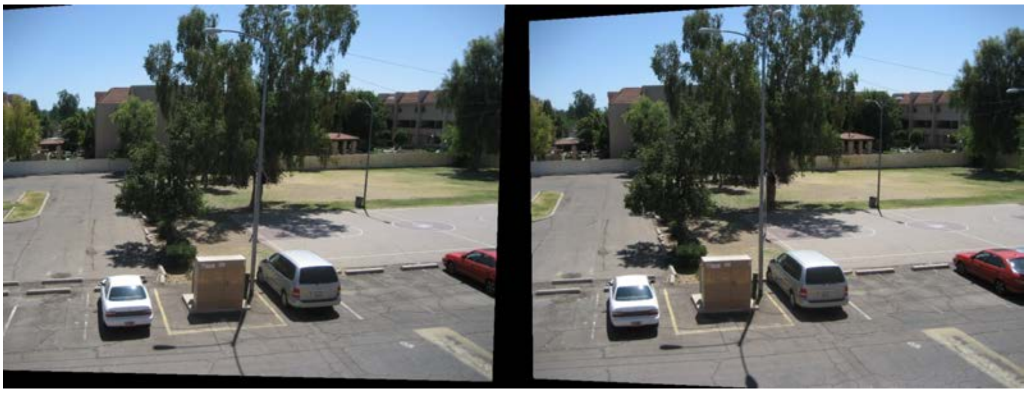


Figure 4.4: Result from Hartley and Zisserman [25] for a generic real image. Taken from [56].

obtained by creating a system of non-linear equations to every j correspondence ($E_j = 0$). The solution is obtained through least-squares using the Levenberg-Marquardt algorighm, and through parametrisation so that

$$\boldsymbol{H_i} = \boldsymbol{K}_{ni}\boldsymbol{R_i}\boldsymbol{K}_{oi}^{-1} \tag{4.5}$$

where i is l or r, for each one of the views. It is shown that the rotation of one of the views around the x-axis is irrelevant, and thus, discarded. Some rectification examples can be found in Figure 4.5. This algorithm was then improved by Monasse and Salgado [39] and

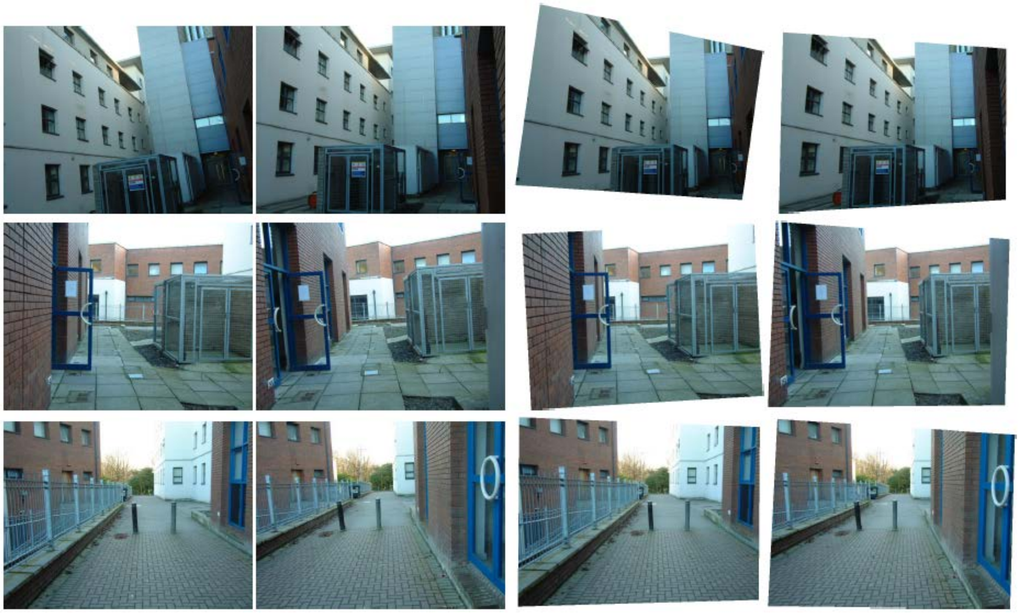


Figure 4.5: Examples from Fusiello and Irsara [21]. The original views can be seen on the left, while the rectified ones stand on the right.

is showing better results. The biggest changes are:

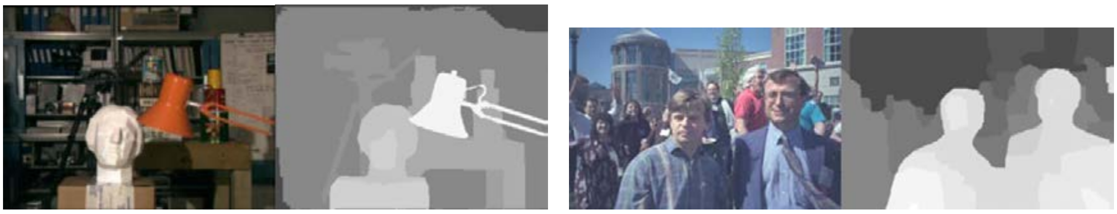
- Improved Jacobian matrix calculations;

- Some changes on the terms and geometric interpretations;
- Actual use of the Levenberg-Marquardt minimization;
- Use of the ORSA algorithm[38] instead of the common RANSAC to find and eliminate outliers from the matched points.

4.4 Stereo Matching

When faced with a pair of stereo images, it could be difficult to establish a dense correspondence for each pixel in both images. This section aims to present some interesting algorithms that are considered top-notch to solve that problem in a correct and efficient way. Nevertheless, the reader should note that efficiency it is not a concern, at this stage, and so there will be no performance analysis in spatio-temporal or complexity terms.

Kolmogorov and Zabih [30, 31] present a well known and recognised algorithm for stereo matching. Their approach is based on energy minimisation via graph cuts, and it is presented as a fast and accurate matching method. They give some degree of interest to the problem of occlusion detection. Their work treats both images symmetrically, handle visibility properly, and imposes spatial smoothness while preserving discontinuities. The problem of energy minimisation is proven to be NP-hard, so graph cuts are necessary to be possible to compute a local minimum of the energy function. Their experimental results are very promising and testify the performance of the algorithm. Those results can be viewed in Figure 4.6.



(a) Result for the Tsukuba dataset

(b) Results on the Dayton sequence

Figure 4.6: Results from Kolmogorov and Zabih taken from [31]. A lighter color means that the points are closer to the camera.

Radim Šára contributed with various algorithms for efficient stereo matching. His work was based in trying to solve the problems of ambiguity and occlusion. Believing that studying the ambiguity related to the reconstruction would extensively help the reconstruction problem, he presented a paper [51] describing an algorithm that is intended to find the largest unambiguous portion of matched points, identifying some variables that can be manipulated to identify what is or is not ambiguous. As a result, he defined a new *stability property* that is a condition a set of matches must satisfy to be considered unambiguous at a given confidence level. The experimental results can be seen in Figure 4.7.

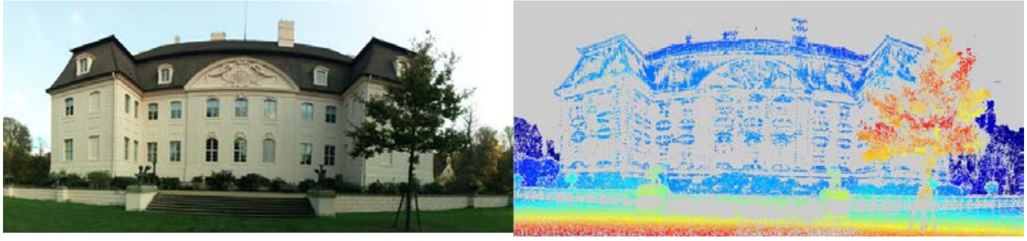


Figure 4.7: Results from the algorithms presented in [51].

This author presented later, in [52], a new set of algorithms, based on his previous work. It aim was to build a robust framework for correspondence matching in computer vision. This framework is featured with the ability of adding some well-structured prior knowledge. In this work some considerations and definitions about visibility, occlusions and ambiguity are presented, also pointing the most common tasks attributed to solve the stereo problems. The paper also shows a large degree of interest in detailing that the prior knowledge or model severely affects the solution found, thus explaining that a contradicting prior model will result in artefacts and wrong disparity calculations. The framework is based on the notion of digraph kernels, that are considered excellent classification tools and have widespread applications and uses [5].

Concerned with computational time while producing accurate results, Čech and Šára [49] present a two-step stereo matching method. This algorithm visits only a small fraction of the disparity space, finding a semi-dense disparity map. From a set of seeds previously selected (using helper algorithms like [23]) the matching is performed growing the correspondence from those seeds on, dropping, under a theoretically well-grounded rule, the uniqueness constraint until the result can no longer be improved. Then, a robust global optimality task is performed, selecting from the competing patters those that are unambiguous. It is shown that the algorithm is very resistant to repetitive patterns and other ambiguities. The most interesting result is that, contrary to initial belief, there is no need for high quality seeds, as the algorithm seems to be able to perform matching in very complex scenes from a set of pseudo-random initial points. Their experimental results are very promising and can be viewed in the Figure 4.8.

Hirschmuller presented some work [26] on a stereo matching algorithm with a semi-global approach. Through the use of a specially designed cost function, the color variations that pixels from the same 3D point may exhibit in the two images are compensated. The matching is performed through a global cost function, that propagates the matching for a determined pixel in all directions trying to reach a global optimisation value. Some post-processing steps are also considered, that try to improve the solution found. Tests

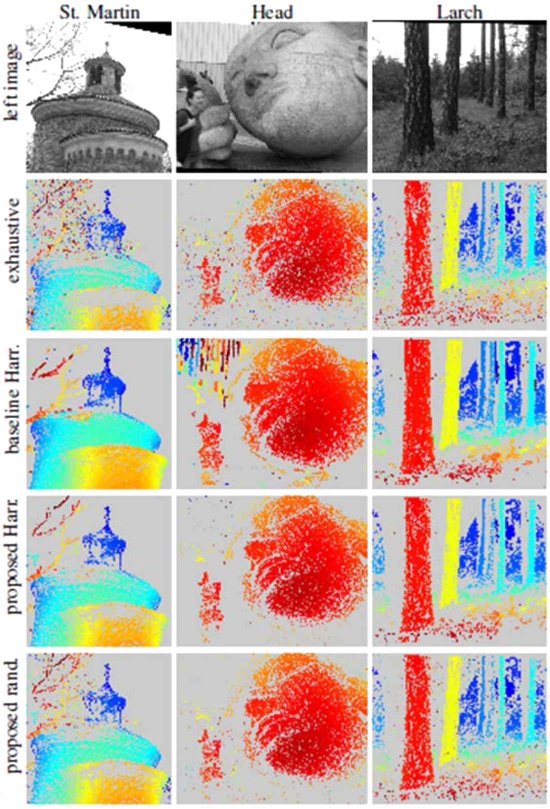


Figure 4.8: Experimental results from the algorithm presented in [49] for some combinations of parameters and different images. Notice that, by "proposed" it is meant the proposed algorithm, "Harris" is the feature detector presented in [23] and "exhaustive" is the application of a classical exhaustive search algorithm (as presented in Chapter 3.4).

and comparisons found that this algorithm is well ranked among the commonly existing algorithms, and the author claims that it can perform better if some subpixel accuracy is computed in the resulting disparity map. This algorithm is implemented with optimization in the *OpenCV*[1] computer graphics library.

Alagoz [2] created two stereo matching algorithms based on classical global energy minimisation functions. The approach was, however, to try and add a smoothing function so as to eliminate unreliable disparity estimations. In the first algorithm, it is attempted to smooth the energy function by applying an averaging filter a large number of times, assigning, for each pixel, a new minimum disparity value, if found in each iteration. In the second algorithm, the approach is based on region growing along image rows, defining root points where regions begin. A root point is defined as a point that does not belong to any other region, and its disparity is computed through the energy function previously defined by the author. The region is then grown from that point on while the energy function stays equal or less the value for the root point. Otherwise, a new root point is

[1] The *OpenCV* library can be found in http://opencv.willowgarage.com/

selected and the region starts growing from there.

Bhatti and Nahavandi [4] published another approach to the stereo matching problem. Their claim is that few work in the area was attempted using wavelets/multiwavelets multiresolution analysis. As such, their work tries to develop a stable algorithm that uses translation invariant multiwavelet transforms to establish correspondences on the stereo pair. In the paper, they also studied some base changes and the effects it could make on the performance of the correspondence estimation. Their experimental results on some of the classical datasets can be found in Figure 4.9.

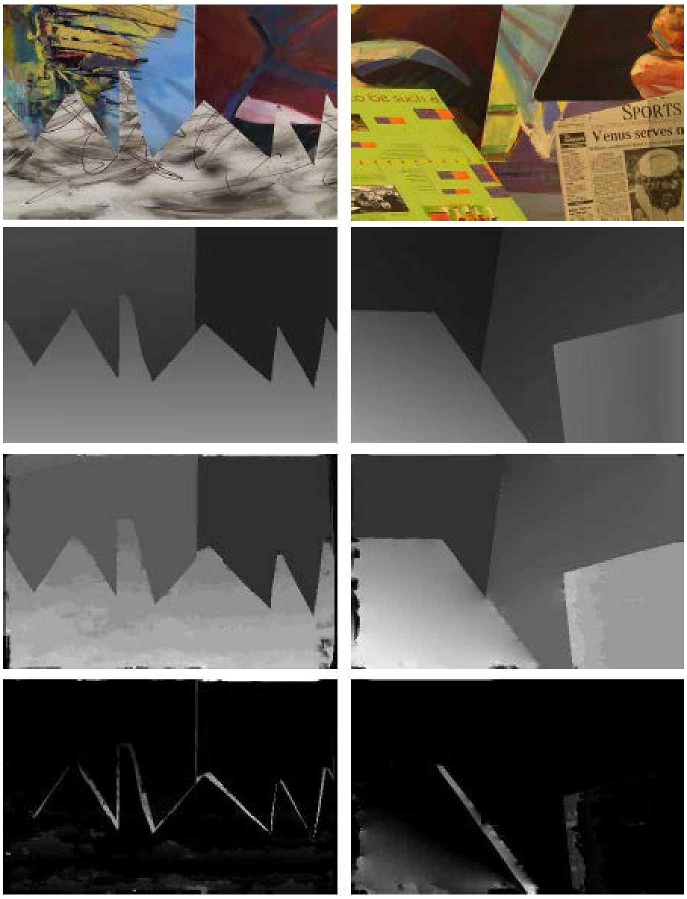


Figure 4.9: Results for Bahatti et al. algorithm for the classical Sawfoot and Venus stereo datasets.

Wang and Road [54] built a stereo matching algorithm based on competitive and cooperative region optimisation. It is an energy minimisation approach. So, first, through segmentation, it defines a set of image regions, establishing initial correspondences with simple window-based methods. Those disparities are then refined through a plane fitting process that uses voting for the assignments. Finally, the disparities of all planes are refined by an inter-regional cooperative optimization procedure based on global energy minimisation. Up till this moment, it's the 3$^{\text{rd}}$ best performing algorithm on the *Middlebury Stereo Evaluation* ranking (see next paragraph). Their experimental results can be seen in the Figure 4.10.

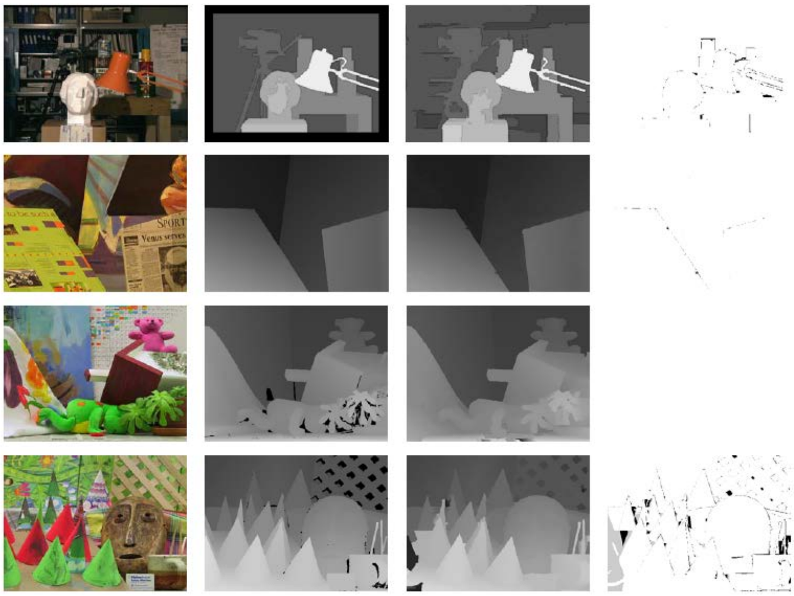


Figure 4.10: Experimental results from Wang and Road [54] for the Tsukuba, Venus, Teddy and Cones databases.

Klaus, Sormann and Karner presented in [29] another matching algorithm. Their approach is based on color segmentation on the reference image followed by an adapting matching score, developed for maximizing an unambiguous solution. After identifying areas of similar color, the matching is performed on a local window of every point (in this paper, a 3x3 window). The idea is that disparity is smooth in areas with the same color, while large disparity variations should occur in discontinuities. After that, the disparity attribution is based on deriving a set of planes that should be adequate to represent the scene structure. Finally, each segment found in the first phase is assigned to a specific plane, through an optimisation algorithm based on energy minimization. It is one of the top ranked algorithms found in *Middlebury Stereo Evaluation* ranking[2], right after a best algorithm, submitted anonymously, and for that reason excluded from this work. The experimental results can be found in Figure 4.11.

Geiger et al. [22] have very recently introduced the algorithm they named as *Efficient Large-scale Stereo* (ELAS) focusing on the problem of creating one algorithm that would be computationally less expensive while achieving comparable results. Their approach is based on Bayesian probability computation. Having in mind that most pixels in a pair of images might be ambiguous and repetitive, their method relies on a set of "support points", defined as points of strong, unambiguously matched features. Then, the assumption is that, along that ambiguous portions, the disparities should vary smoothly. After

[2] The ranking can be found in http://vision.middlebury.edu/stereo/eval/

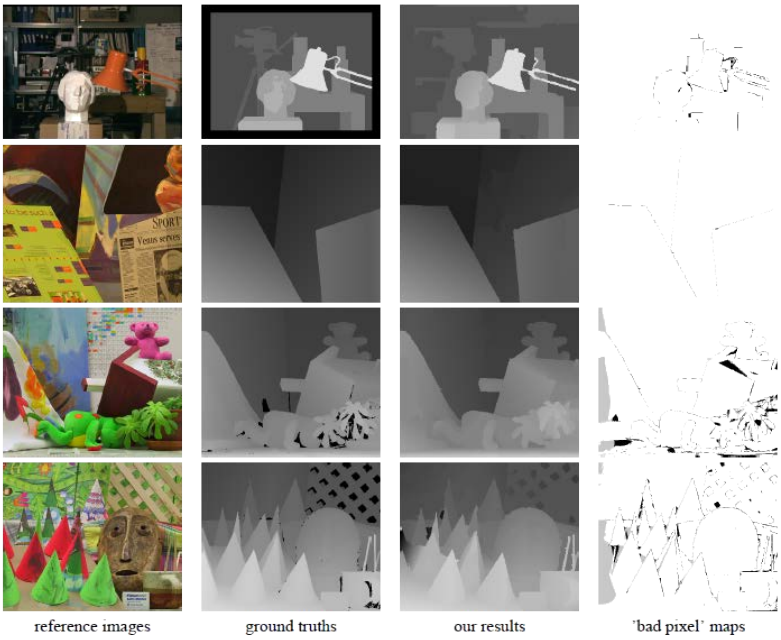


Figure 4.11: Experimental results from Klaus et al. for the Tsukuba, Venus, Teddy and Cones databases.

computing the disparity of that "support points', they create a 2D mesh via Delaunay triangulation. Then, it's up to find a piecewise linear function that is able to approximate that mesh to the real disparities. Results for the classical evaluation datasets are shown in Figure 4.12.

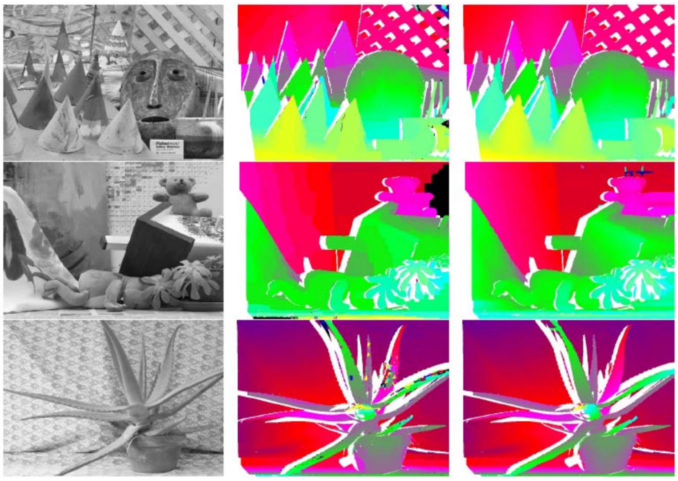


Figure 4.12: Classical result set for the algorithm by Geiger et al. [22].

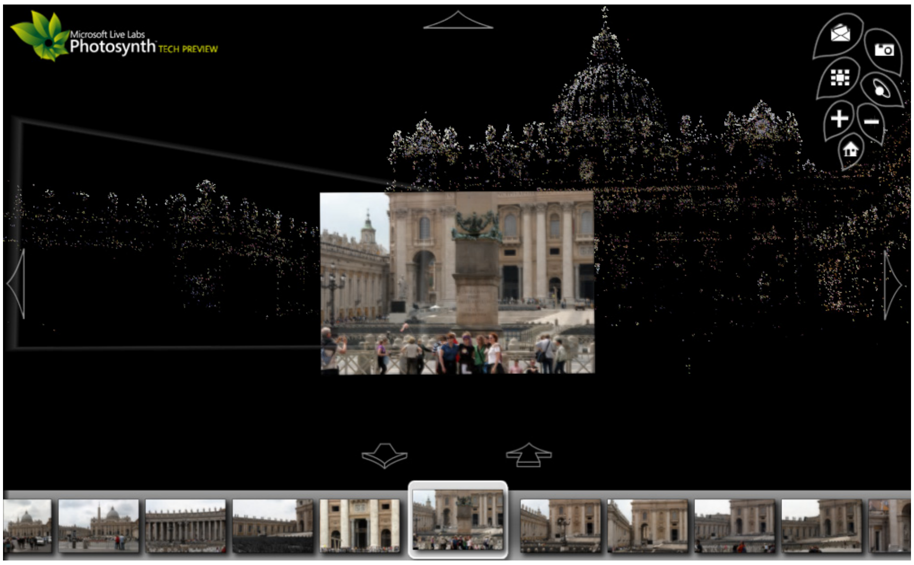


Figure 4.13: Microsoft Photosynth$^{\text{TM}}$screenshot showing the *Piazza San Pietro, Vatican City* 3D cloud point and a picture used to created it.

4.5 Software Examples

There are some commercial examples that use stereo reconstruction. The simplicity in use and low cost of data has motivated many commercial implementations and there exists a plenitude of software available.

The first example is that of **Microsoft** ® **Photosynth**$^{\text{TM}}$ (screenshot in Figure 4.13). Through the use of pictures taken from any camera from any location, users are (freely) invited to create their own reconstructions. The product's website features many famous places that were reconstructed through the use of multiple contributions from many users. One famous and recent use of **Photosynth**$^{\text{TM}}$ can be seen in the *CNN* special *President Inauguration* ceremony[3].

Another example is the **DI3Dcapture**$^{\text{TM}}$ **3D Capture Software** from *Dimensional Imaging*. According to their marketing description, their software is able to calculate a dense range (disparity) map as well as creating 3D models from some uncalibrated views. The software automatically produces 3D meshes that are rendered using modern GPU units and is able to merge multiple consecutive meshes in order to create a larger and more complete one. More details are available in http://www.di3d.com.

[3] http://edition.cnn.com/SPECIALS/2009/44.president/.inauguration/themoment/, accessed 9th February 2011.

Chapter 5

Experimental Work

5.1 Methodology

This project is about studying the possible approaches and the state-of-the-art methods about uncalibrated stereopsis and reconstruction. This section describes the steps which took part of the solution attempt. The results and some discussion will take place in the next section.

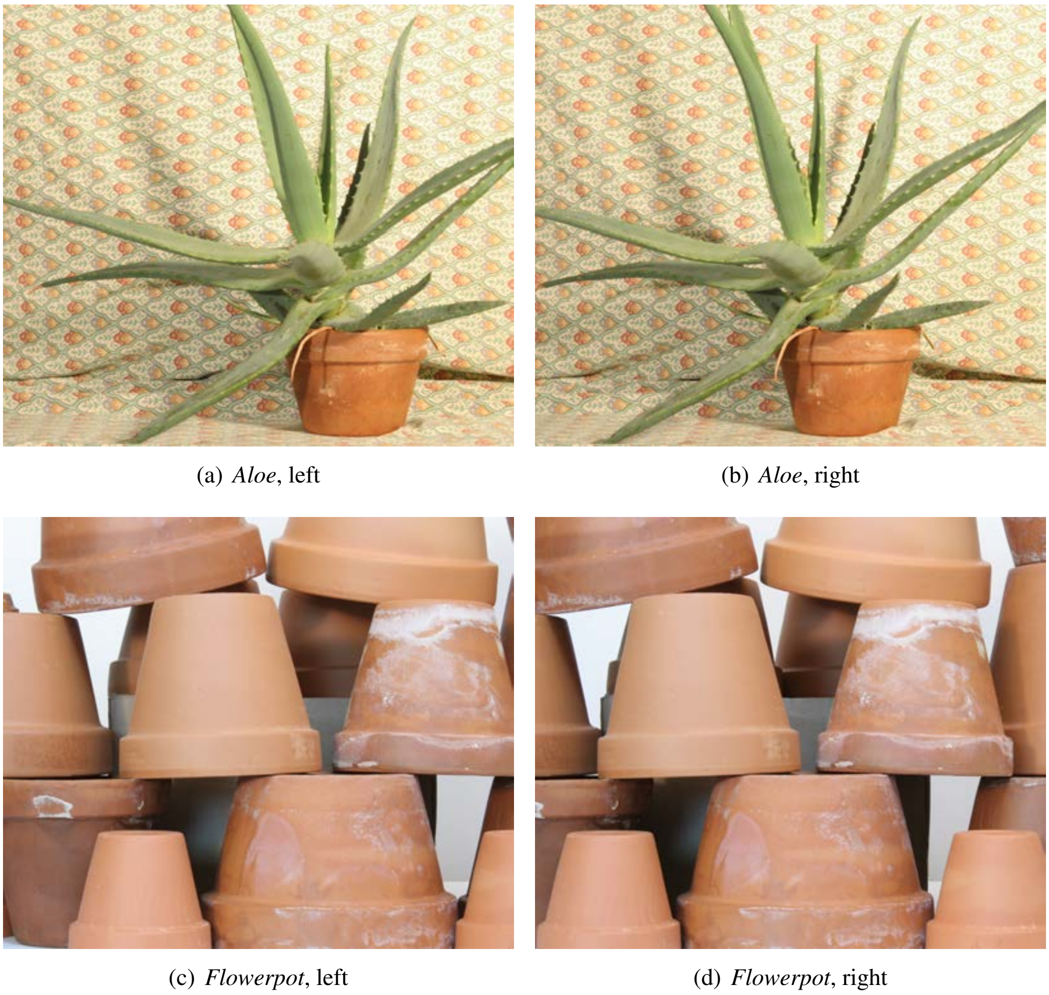
5.1.1 Test images

After some analysis on the state-of-the-art stereo matching descriptions, it is noticeable that they are developed and tried on some, well known and somewhat featured images. The testing datasets evolved too[1] and it was decided that, before attempting any algorithm, they should be tested against a "standard" image that had some similarity with this project's images: presence of zones of high ambiguity due to low texture or its repeatability . That would allow for some overview of how would the algorithms behave on those smaller and almost perfectly rectified images, before introducing the bigger, not so well rectified views. As such, the *Aloe* and *Flowerpot* images from the 2006 *Middlebury* datasets were chosen to verify if the algorithms performed well against texture repeatability and large portions of ambiguity. The selected pairs are in Figure 5.1.

Later on, there was a brief test on two very ambiguous and virtual images. The nature of those images would pose a very hard to solve test on all known algorithms and would allow a strong starting point for this study. Those images can be seen in Figure 5.2.

The next set of images is based on a dummy (original: 5.3(a)) that was modified to incorporate some volume change between the two breasts. Although poorly executed, it was attempted to mimic the natural skin ambiguity by using almost always the same color,

[1] Some datasets can be found in `http://vision.middlebury.edu/stereo/data/` and the more recent have been used in the best "scoring" algorithm.



(a) *Aloe*, left

(b) *Aloe*, right

(c) *Flowerpot*, left

(d) *Flowerpot*, right

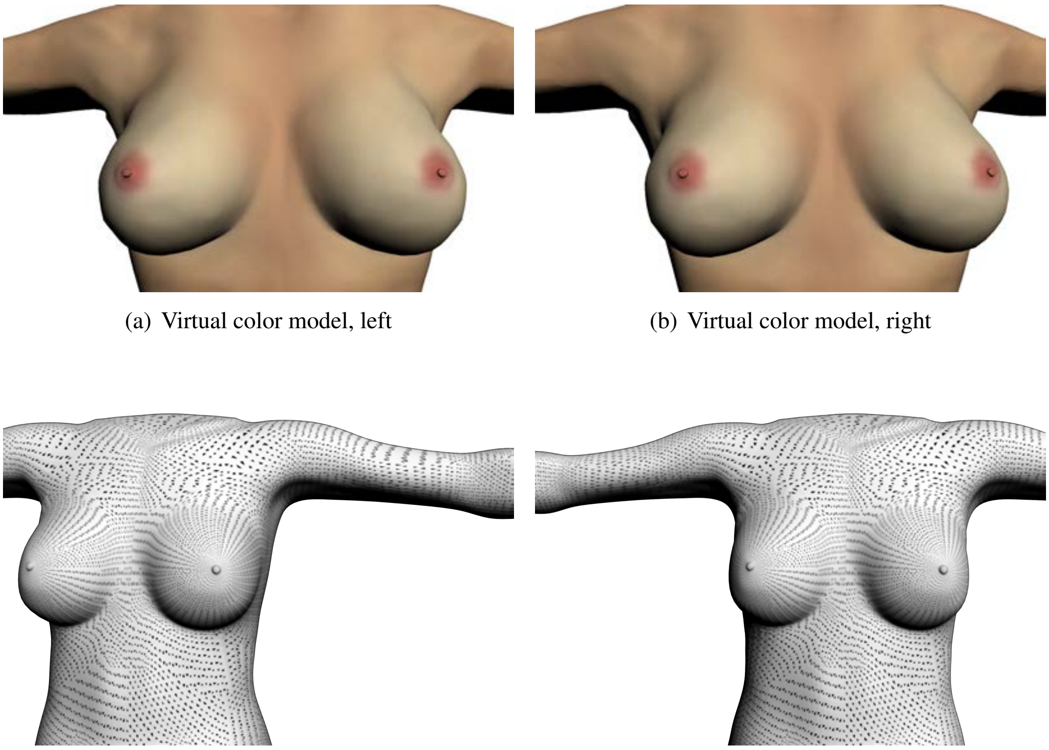
Figure 5.1: Initial test images, from the 2006 dataset in *Middlebury* website.

although with some points of color change that would copy the effect of skin lesions, moles, freckles or blemishes. That model and the selected views are in Figure 5.3.

5.1.2 Feature point detection

It is easily perceived that there are not many distinctive features in the images of this project. The most common and known algorithms were tested for the number and quality of features. The objective is to find the detector that is able to find the biggest number of quality features that can then be matched. The tested algorithms were:

- Harris corner detector [23]
- SIFT [34]
- SURF [3]
- MSER [36]



(a) Virtual color model, left

(b) Virtual color model, right

(c) Virtual blank model, left

(d) Virtual blank model, right

Figure 5.2: Virtual models for large ambiguity tests.

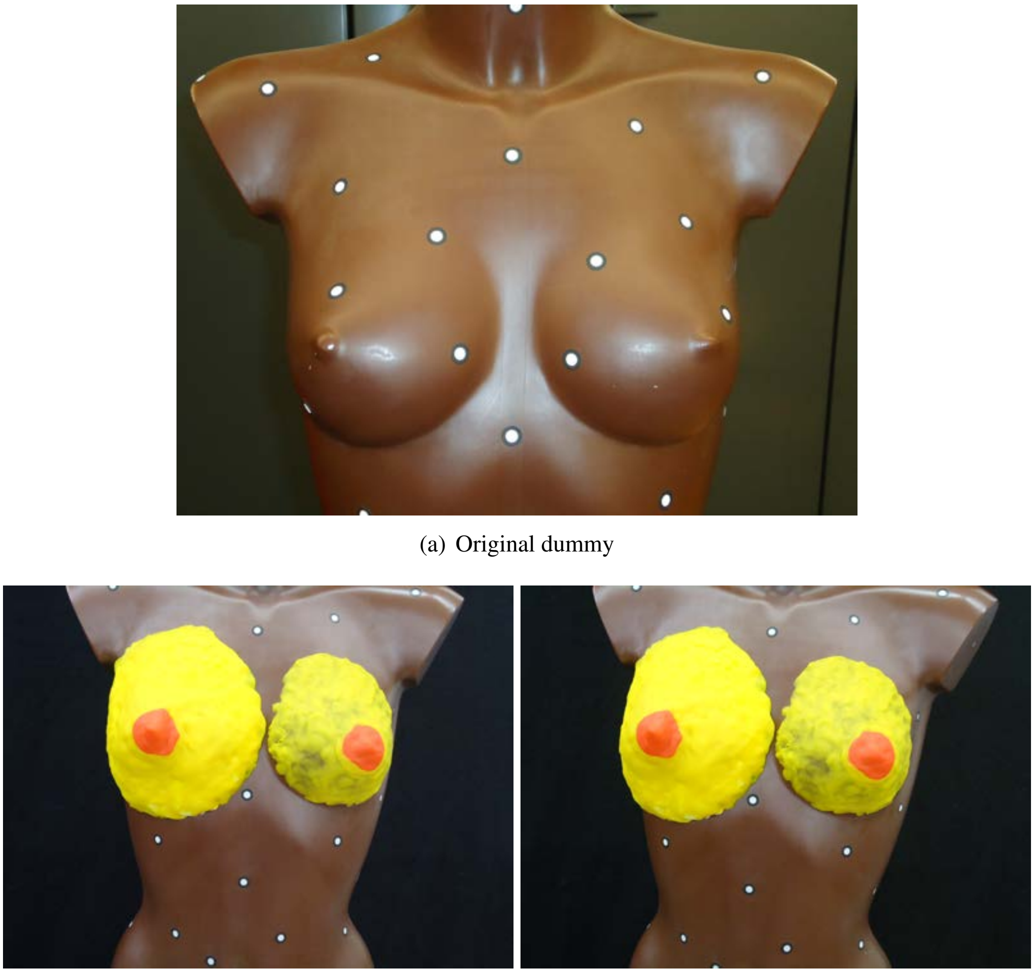
some others were proposed but either an implementation was not found, or the tests failed to run in the used configuration. Some examples:

- Ferns [42]
- STAR [48]
- HoG lines and points [14]
- Dias et al. [16]

5.1.3 Image rectification

Hartley and Zisserman [25] method was computed and attempted using the tools present in *Matlab*®[2]. The Fusiello and Irsara [21] method is tested with its own *Matlab*®toolbox, and so is that method's improvement by Monasse and Salgado [39].

[2] http://www.mathworks.com/products/computer-vision/demos.html?file=/products/demos/shipping/vision/videorectification.html



(a) Original dummy

(b) Modified dummy, left view

(c) Modified dummy, right view

Figure 5.3: Dummy model for more real tests, with induced volume difference in the breasts.

5.1.4 Disparity map generation

The disparity tests were performed using the author's code, either through executables or through their *Matlab*®wrappers. Not many implementations were found, whereas the most common code available is that of very simple matching presented in Section 3.4. Appart from those, the tested algorithms were:

- Shawn Lankton implementation of a selective mode filter algorithm[3].
- Hirschmüller [26]
- Čech and Šára [49]
- Alagoz [2]

[3] It was not possible to find a publication on this code. It can, however, be found in `http://www.shawnlankton.com/2008/04/stereo-vision-update-with-new-code/`

- Klaus et al. [29]
- Geiger et al. [22]

All but the first one from this list are described in Section 4.4. About the first one, the algorithm processes the images by first computing disparity values comparing shifted versions of images; then a 2D mode filter tries and replaces low confidence calculations with information from its neighbours.

5.2 Results

This chapter presents the results to the tests planned in Chapter 5. A figurative testbed was developed and there are now conditions to make all tests for any image that is required. The parameters used are those that were shown (by the authors) to have the best results. Later, some parameter variation was attempted in some algorithms, to see what changes those parameters would bring to the resulting map. It was not possible to test for all parameters on all algorithms due to the time and resources each test consumes.

5.2.1 Feature point detection and matching

The testbed was run on all images and the results are present in Table 5.1. Please note that the extra column about the *SIFT* algorithm is based on matching against *SIFT* descriptors using squared Euclidean distance, and is present because, according to Lowe [34], some matches should be rejected when considered too ambiguous.

		Harris	SIFT		SURF	MSER	Total
		Corners	Features	Matched	Features	Regions	
Aloe	l	26	10383	5727	8606	11306	72648
	r	27	10447		8510	11024	
Flowerpot	l	8	4605	2127	1055	591	
	r	10	4611		1093	597	
Color model	l	75	3751	1748	202	9	
	r	92	3726		191	10	
Artificial model	l	500	8810	4219	9803	2364	
	r	500	8759		9409	2352	
Dummy	l	7	9832	1072	960	197	
	r	9	9663		971	205	
Total (avg. unmatched)		627	37294	—	20400	14328	
% of total		0.8	51.3		28.1	19.7	

Table 5.1: Results for the testbed and some evaluation. Note that, because there is no defined matching refusal criteria for the remainder algorithms, it was decided to compare the global number of features detected, as a larger number raises the probability of correct matches.

The *SIFT* algorithm appears to be largely superior to the remainder, and therefore it was selected for the rest of the tests. The feature representation (as close as possible to that returned by the algorithm) is shown in Figures 5.4 and 5.5. It is possible that a combination of all four algorithms would increase the accuracy, but that depends on defining an algorithm that would:

1. Eliminate duplicate or too near features;
2. Scale those features according to some quality level, where only the high quality features would be used.

5.2.2 Rectification

The rectification was performed on the colored artificial 3D model and on the marked dummy, because those get close to human body representations. The other test images were discarded for these tests. The pairs used for rectification were obtained through *SIFT* both for detection and matching, in this configuration:

- Dummy: 1072 matched pairs;
- Artificial coloured model: 1748 matched pairs.

Hartley and Zisserman [25] rectification algorithm was parametrized to use different fundamental matrix generation methods. Those that got tested were the *RANASC*, *Least Trimmed Squares* and *Least Median of Squares* [25]. All methods considered the entire set of 1072 pairs as inliers for every pair of images. The selected parameters were:

- Confidence: 99.99%
- Distance: 0.01 Sampson
- Number of trials: 10000

This method achieved good results, but more recent algorithms performed better.

Fusiello and Irsara [21] achieved the best results of the test, with the lowest possible error. However, it is also the slower to run and sometimes it might not end successfully, due to the random characteristics of the Non-linear Least Squares method, that initializes the focal length with a random value and then iteratively tries to improve that value. The initialization of this value, taken from the image's *metadata*, would possibly improve the result. It is not part of the initial algorithm to search and eliminate outlier matches, so it was done by fitting an initial fundamental matrix through *RANSAC* at a required minimum distance to the unrectified epipolar lines for each pair set to a very restrictive 0.00001 pixel.

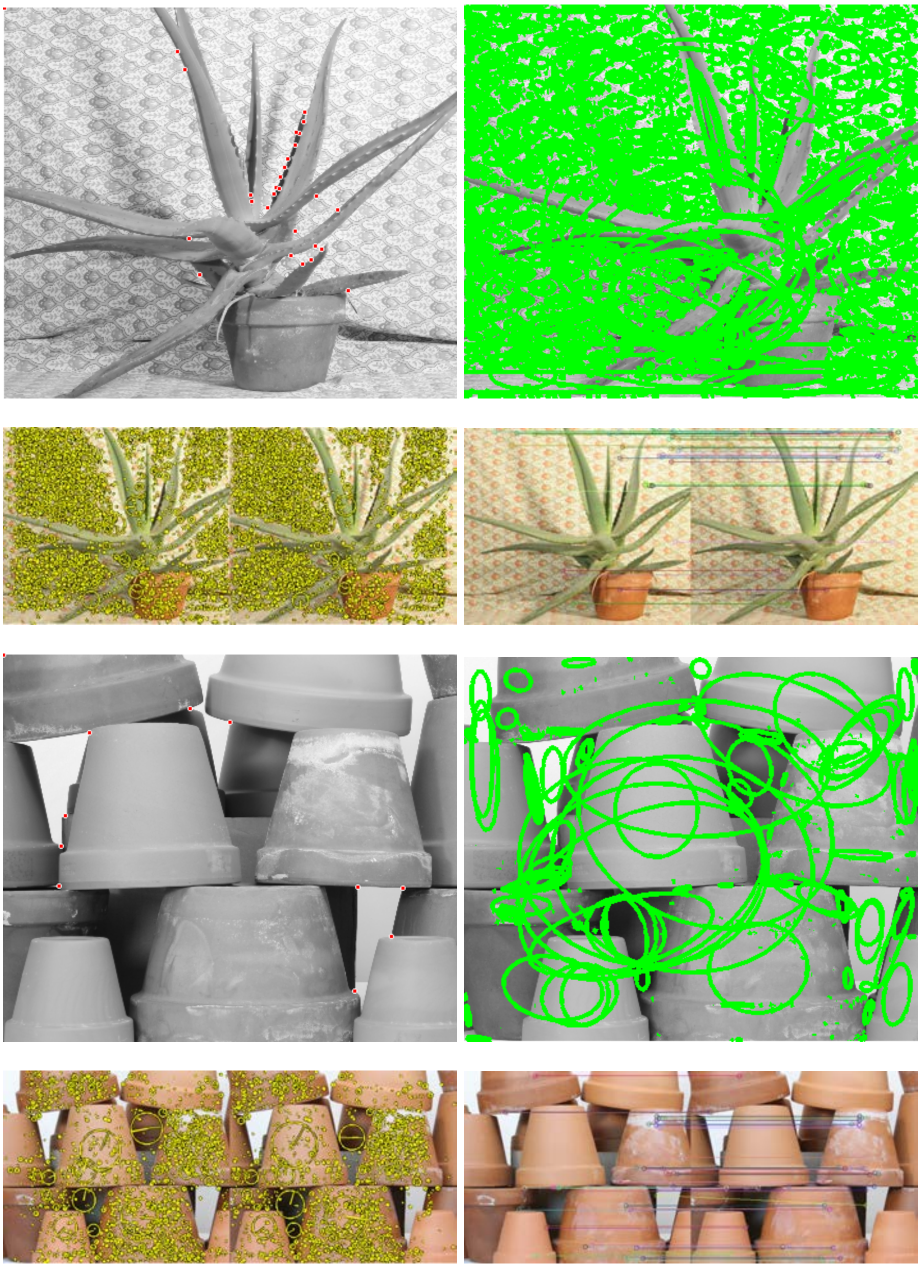


Figure 5.4: Feature detection performed in the *Aloe* and *Flowerpot*, algorithms in this order: *Harris*, *MSER*, *SIFT* and *SURF*.

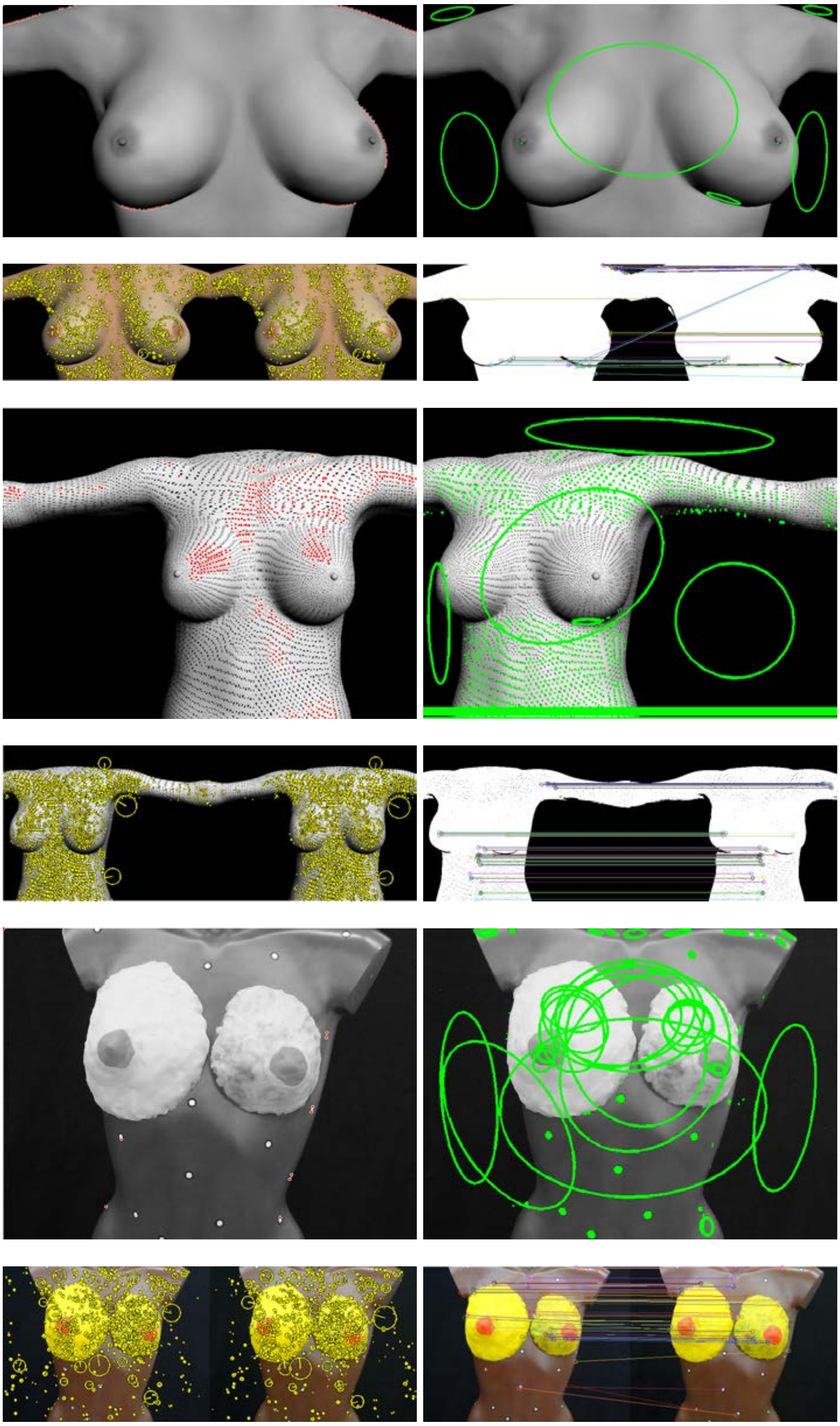


Figure 5.5: Feature detection performed in the three artificial models used, algorithms in this order: *Harris*, *MSER*, *SIFT* and *SURF*.

			Color Model	Dummy
		# Features	1755	1069
Hartley	RANSAC	Error	0.15292	1.75616
		Inliers	1755	1069
	LMedS	Error	0.180815	0.688782
		Inliers	1755	1069
	LTS	Error	0.173151	0.775647
		Inliers	1755	1069
Fusiello	Error		0.019211	0.024518
	Inliers		595	586
	Focal length		1006.5	2872.5
Monasse	Error		0.581935	0.044068
	Inliers		595	751
	Focal length		2824.61	2691.91

Table 5.2: Rectification results and comparison. **Error** is established through Equation 5.1 and is in pixel.

Monasse and Salgado [39] method achieved, in this scenario, highest rectification errors though the author claims the algorithm performs best. It might be due to a very restrictive set of inliers computed for the previous algorithm or it could be related to the scene or the quality of features found.

The results are summarised inf Table 5.2. The rectified pairs are visually very similar, so only the result with the minimum error is presented in Figure 5.6. The presented error was obtained by applying the Sampson error suggested by [20] and defined as:

$$e_{rec} = \frac{1}{N}\sqrt{\sum_j E_S^j} \text{ where } E_s^j = \frac{(\boldsymbol{m}_r^j)\boldsymbol{F}\boldsymbol{m}_l^j)^2}{||[\boldsymbol{u}_3]_\times \boldsymbol{F}\boldsymbol{m}_l^j||^2 + ||\boldsymbol{m}_r^{jT}\boldsymbol{F}\boldsymbol{u}_3]_\times||^2} \tag{5.1}$$

and can be seen as the root mean squared distance of each feature point to its horizontal epipolar line.

5.2.3 Stereo matching

The same images used for studying the rectification algorithms will now be used to try and generate the most dense disparity maps possible. The rectification algorithm used was that with the lowest error, according to Table 5.6. There is no ground truth available and there is no clue about minimum or maximum disparity, so it is always assumed to range between 0 and a large number, chosen to be 255.

The results for all the algorithms running over the coloured 3D model with the default parameters can be viewed in Figure 5.7. It is evident that the large ambiguity the image presents is terribly handled by all algorithms, and the resulting disparity maps are visually wrong, sparse and unintelligible.

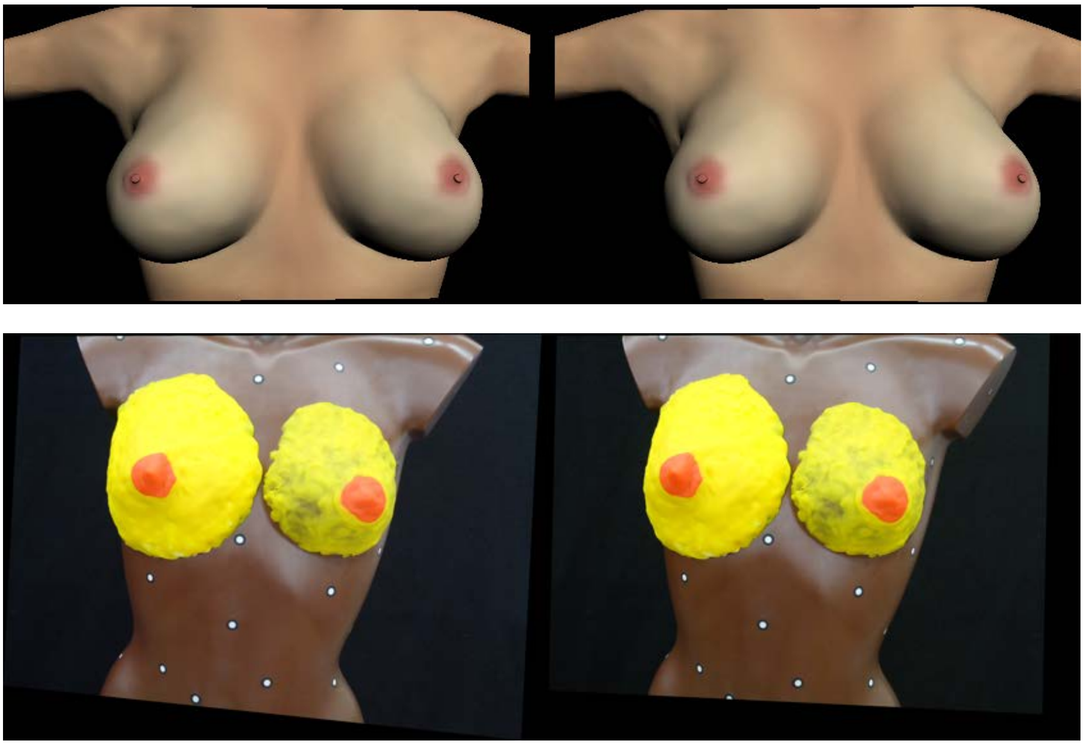


Figure 5.6: Best results from rectification algorithms, both by Fusiello and Irsara [21].

Unfortunately, for the Dummy images the maps did not improve as much. The large texture and color ambiguity in the surface of the Dummy can be attributed to these weak results. In any case there is a disparity map good enough for 3D reconstruction and in most of them there are no distinguishable differences between parts of the left and the right breast, although originally they were at different depth levels. The disparity maps can be viewed in Figure 5.8. Note that result 5.8(h) was rotated for unknown reasons by Geiger et al. [22] own code.

5.2.4 Real scenario

A small dataset containing some photographs taken from medical personnel was available for this project. One of the first immediately spotted difficulties was for the medic to retain the “low baseline” constraint, which did not occur in most of the shots. Taking into account that it was specifically requested and reasoned, it denotes that enforcing this constraint within medical staff might be a concern.

In this section it was decided to present two cases, namely the best and the worse for the available dataset. They were named **Patient A** and **Patient B** (see Figure 5.9). The rectification results are summarised in Table 5.3 and can be seen in Figure 5.10. Both patients were tested for all the disparity algorithms that were tested in the previous section.

	Patient A	Patient B
Features	1377	329
Inliers	755	168
Focal length	4505.1	5185.1
Error	0.025873	0.334152

Table 5.3: Rectification values for the Patient photographs using *SIFT* for features and Fusiello and Irsara [21] for rectification.

For **Patient A**, (Figure 5.11) the results appear to be better than the testing ones, in some cases. On the other hand, for the remainder, they are catastrophically bad. There is still no disparity map dense and correct enough for a decent reconstruction and some algorithms tend to continuously give close disparity values to both breasts and, in some cases, even to the chest.

The images from Patient B (Figure 5.12) are more difficult to process than the rest, for lack of features and because the images are a little blurred. The results are pretty terrible, and there are no distinguishable depth differences between the chest and the breasts. No disparity map is usable for an attempted reconstruction.

It is feasible to conclude that, for this kind of images, it is not possible to use the most common state-of-the-art algorithms. It is needed to search for a more suitable algorithm, one that does a wider, more globally oriented search that would better fit the ambiguity of these images. Nonetheless, a global algorithm would escalate in terms of complexity, so it might be a trouble accounting for the target of this software.

5.2.5 Parameter variation

Some parameter variation was attempted for some algorithms to verify what changes that would bring to the generated map. Those variations were done on the two most common parameters: window size and maximum disparity value. Other parameters were available for Klaus et al. [29] and Alagoz [2] solutions, but those proved to have little if no effect on the final result. This section presents some of the results obtained, found to be the most relevant and consistent among various runs.

The first algorithm that was varied was Sum of Absolute Differences. For the coloured 3D dummy, the results are presented in Table 5.4 where the results for the dummy are presented in Table 5.5. A larger window appears to have a detrimental effect on the definition of edges and color changes, while increasing the maximum number of disparity variation appears to introduce more noise to the final result. Similar interpretation is in order for Patient A (Table 5.6) and Patient B (Table 5.7) images, added that the noise, in those cases, is more evident.

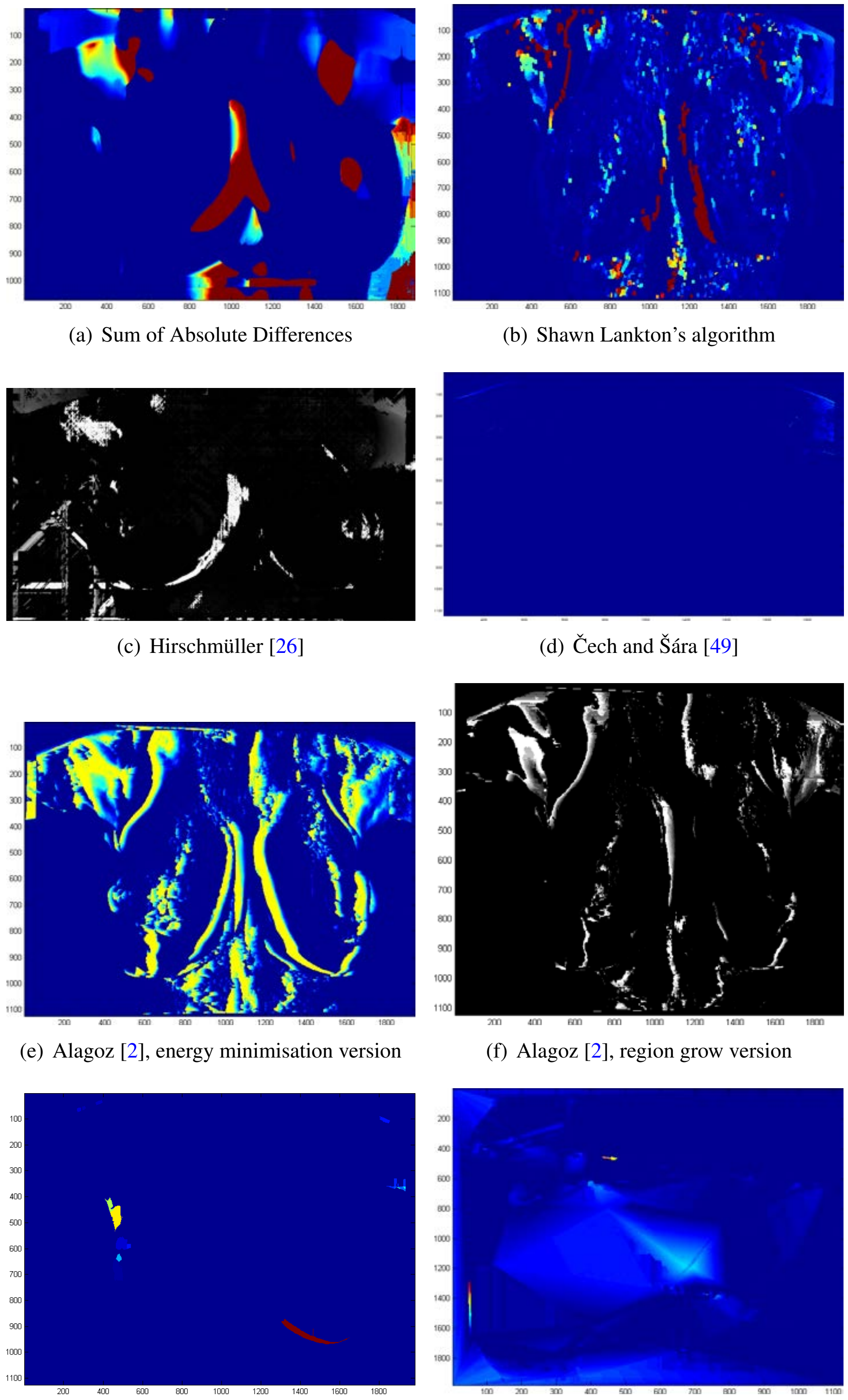
For Shawn Lankton's algorithm, the results for the 3D model and the Dummy were unintelligible and, in some cases, the algorithm wouldn't run on the computer framework available. For these reasons, no result will be presented for this algorithm. Somewhat similar, Alagoz [2] was taking too much memory and time to run, and it became infeasible to vary the parameters in a significant way.

Čech and Šára [49] solution had only the option to adapt a different window size to the matching. The other parameters were already parameterised to optimality, i.e. the disparity values were set to be possibly infinite (capped to the maximum and minimum representable integer) and the maximum number of disparity candidates for each pixel. The results (Table 5.8), however, are unintelligible from each other, forcing the conclusion that varying the window size has no effect on this algorithm.

Geiger et al. [22] on the other hand had the option to vary the maximum disparity value but no control over the matching window. Some other parameters were available, but after some runs it was found that there was no significant variation in the final result. The results for those variations are found in Table 5.9. Increasing the maximum disparity value has a direct effect on the resulting map, increasing its definition and improving the final result. However, the map is not intelligible enough.

5.2.6 Discussion

As early suggested and, somewhat, predicted, the state-of-the-art algorithms could not successfully reason the correct correspondences for a too large portion of the pairs, both virtual and real. From the literature, that can be attributed to the fact that it is very difficult to attribute a confident matching in many of the areas of the images used in this project. That difficulty can be verified with a simple zoom on similar areas in both images. The difficulty of reasoning is notable, and even a human, with visual capabilities that overcome the actual computer possibilities in many areas, is incapable of reasoning about the correct correspondences.



(a) Sum of Absolute Differences

(b) Shawn Lankton's algorithm

(c) Hirschmüller [26]

(d) Čech and Šára [49]

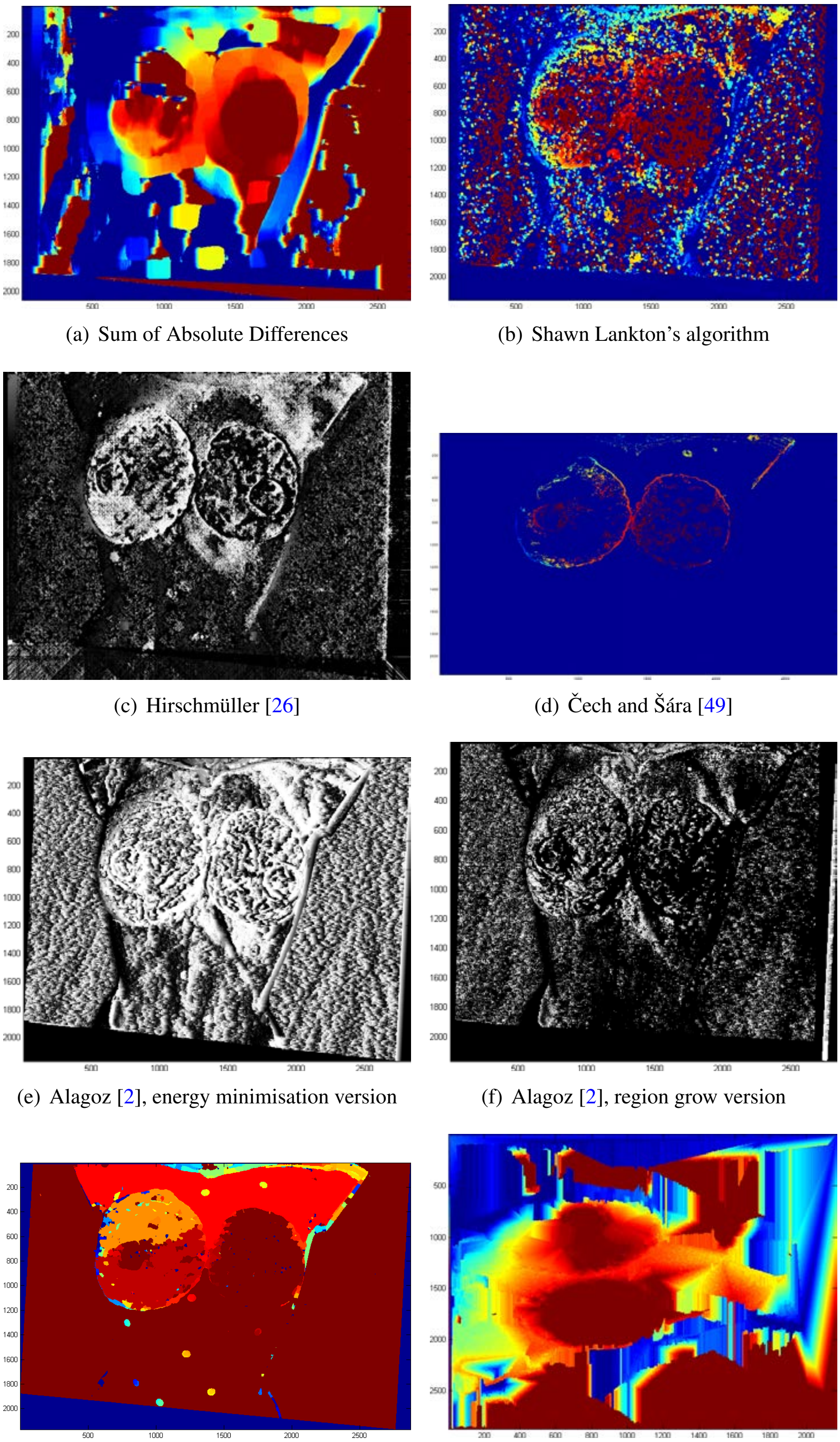
(e) Alagoz [2], energy minimisation version

(f) Alagoz [2], region grow version

(g) Klaus et al. [29]

(h) Geiger et al. [22]

Figure 5.7: Results for stereo matching algorithms running over the 3D coloured model.



(a) Sum of Absolute Differences

(b) Shawn Lankton's algorithm

(c) Hirschmüller [26]

(d) Čech and Šára [49]

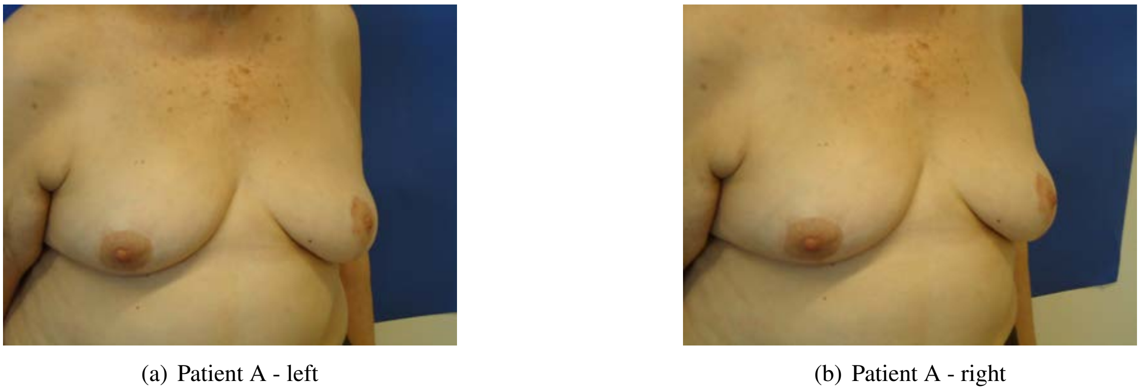
(e) Alagoz [2], energy minimisation version

(f) Alagoz [2], region grow version

(g) Klaus et al. [29]

(h) Geiger et al. [22]

Figure 5.8: Results for stereo matching algorithms running over the Dummy model.



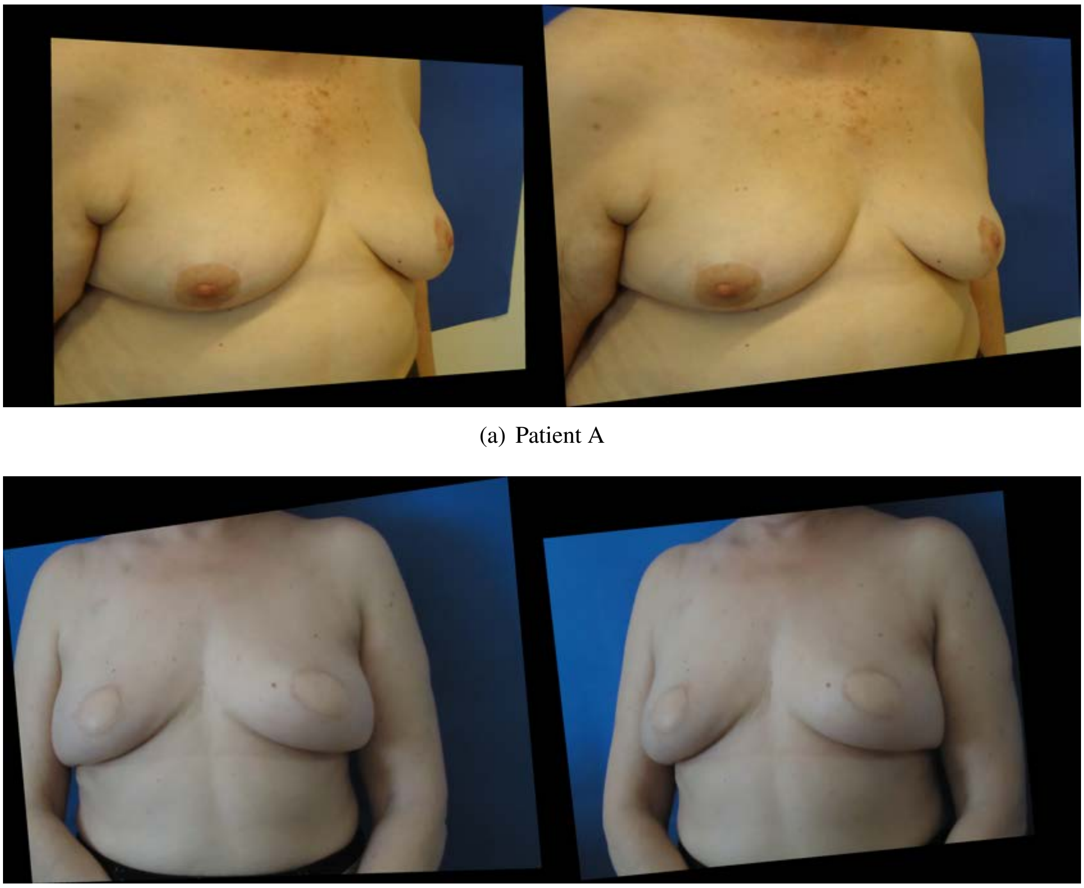
(a) Patient A - left

(b) Patient A - right

(c) Patient B - left

(d) Patient B - right

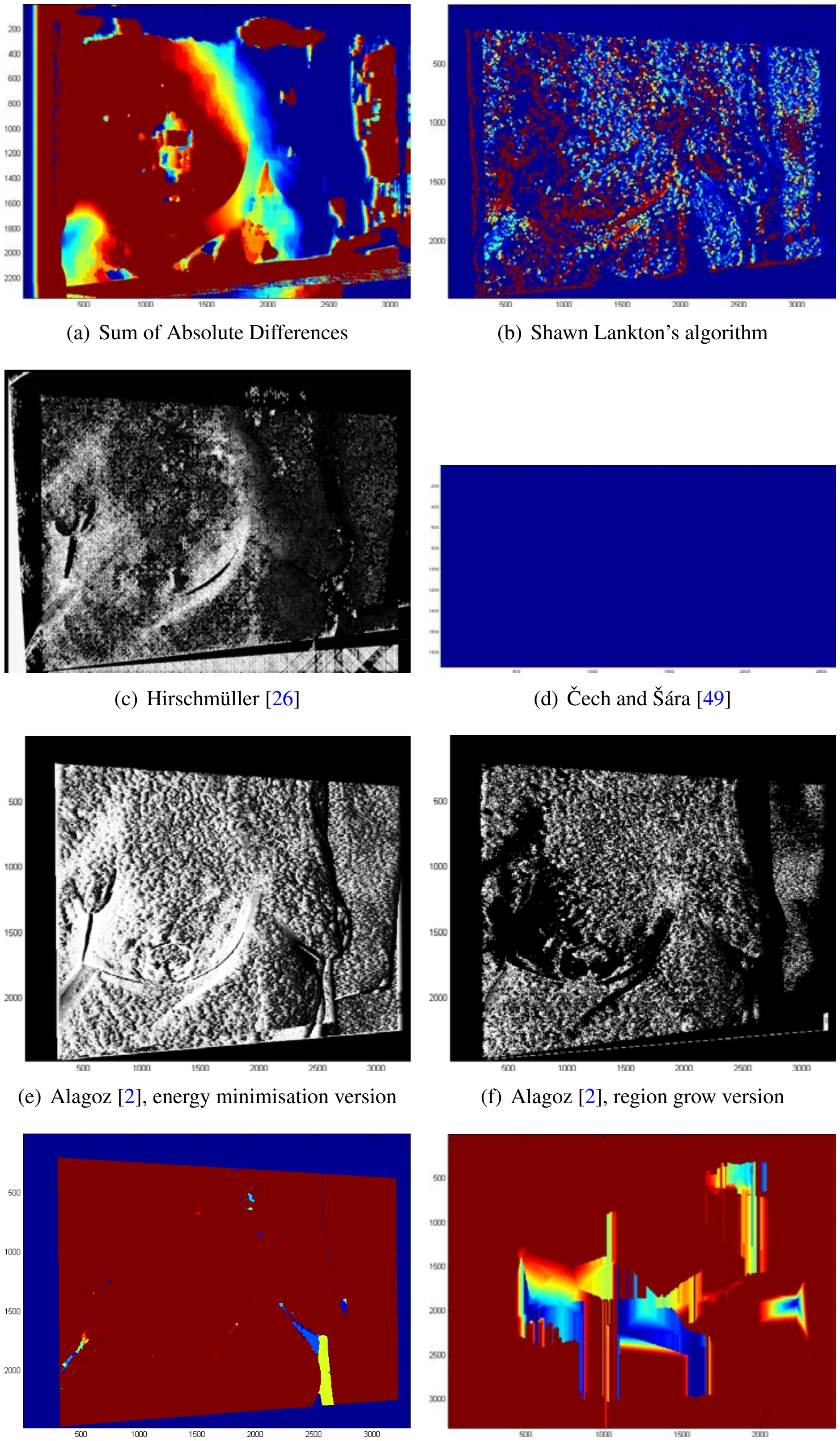
Figure 5.9: Original patient pairs, prior to processing.



(a) Patient A

(b) Patient B

Figure 5.10: Rectification results for the patients' pairs.



(a) Sum of Absolute Differences

(b) Shawn Lankton's algorithm

(c) Hirschmüller [26]

(d) Čech and Šára [49]

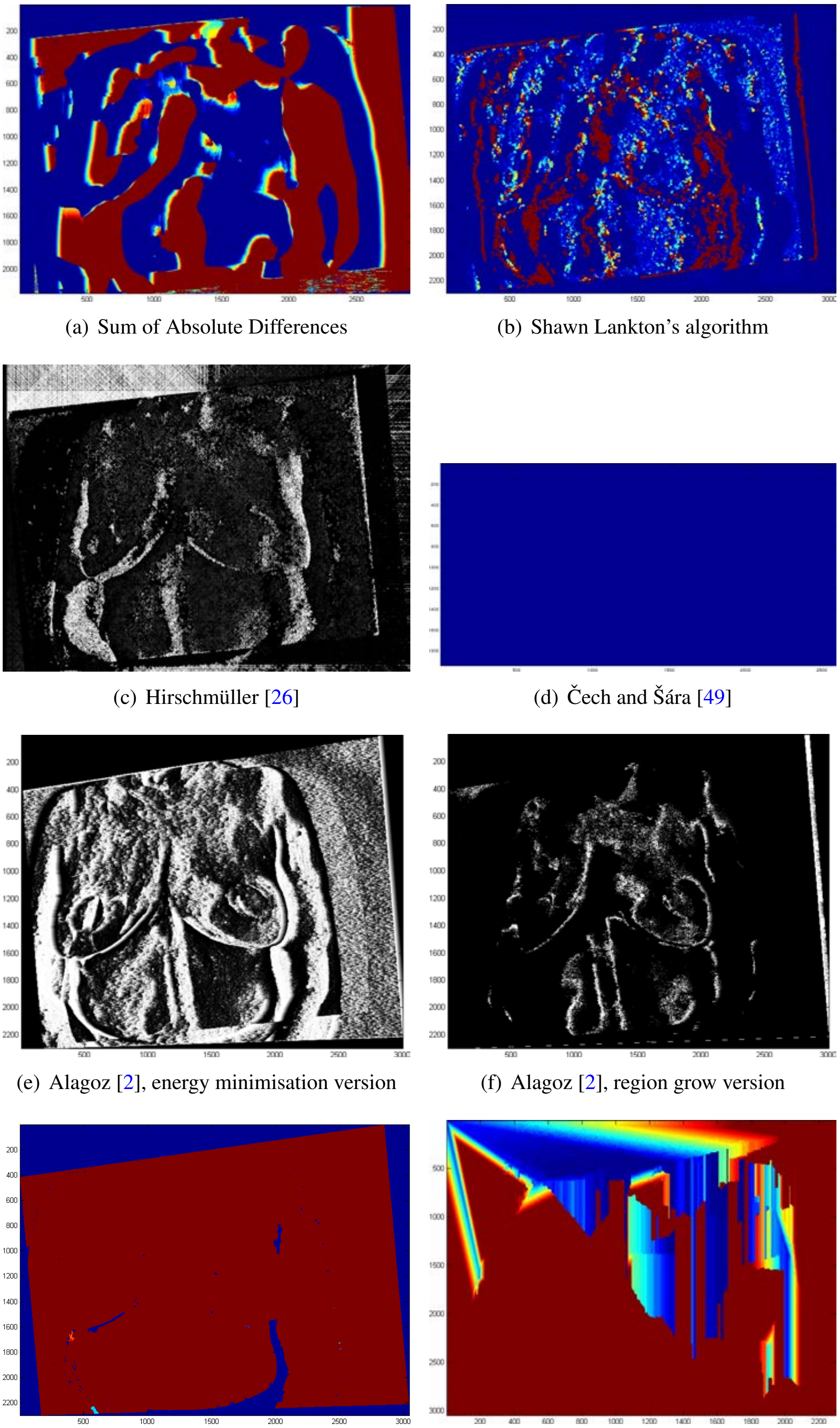
(e) Alagoz [2], energy minimisation version

(f) Alagoz [2], region grow version

(g) Klaus et al. [29]

(h) Geiger et al. [22]

Figure 5.11: Results for stereo matching algorithms running over Patient A photographs.



(a) Sum of Absolute Differences

(b) Shawn Lankton's algorithm

(c) Hirschmüller [26]

(d) Čech and Šára [49]

(e) Alagoz [2], energy minimisation version

(f) Alagoz [2], region grow version

(g) Klaus et al. [29]

(h) Geiger et al. [22]

Figure 5.12: Results for stereo matching algorithms running over Patient B photographs.

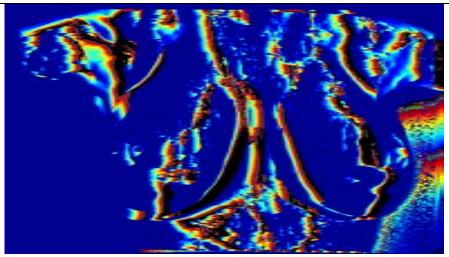
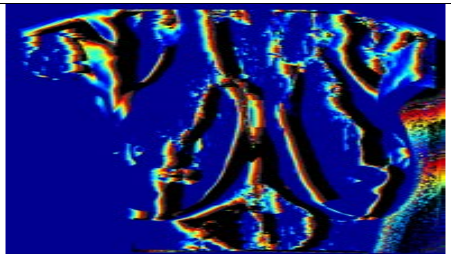
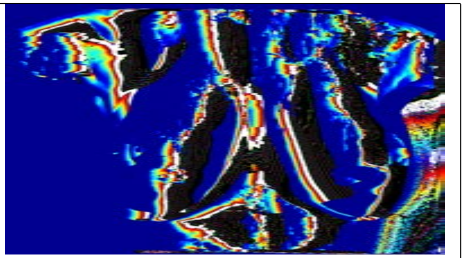
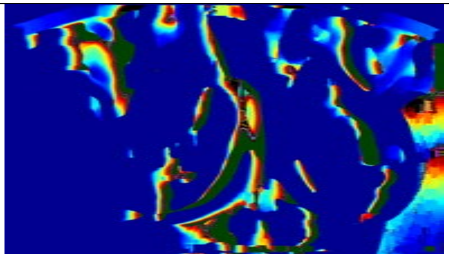
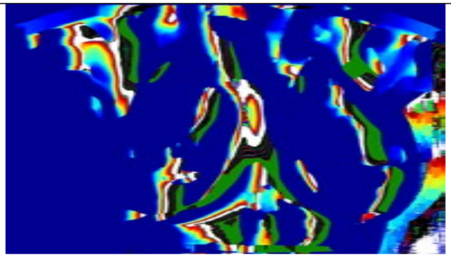
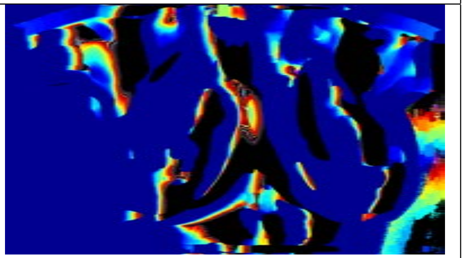
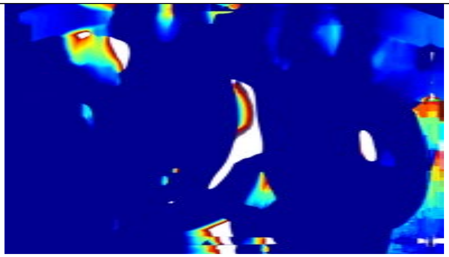
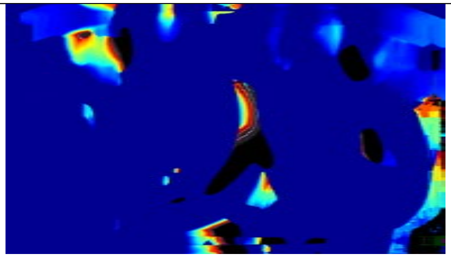
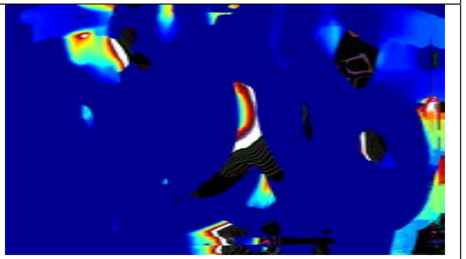
	80	160	255
9			
49			
121			

Table 5.4: Parameter variation for SAD algorithm, for the coloured 3D model. Column: window size; Row: Maximum Disparity.

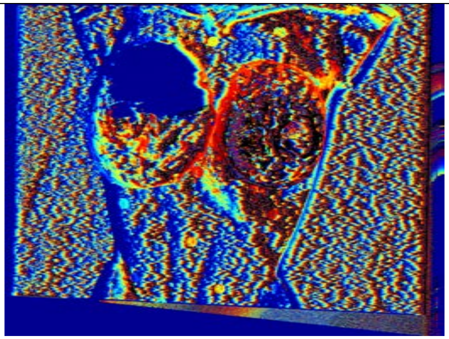
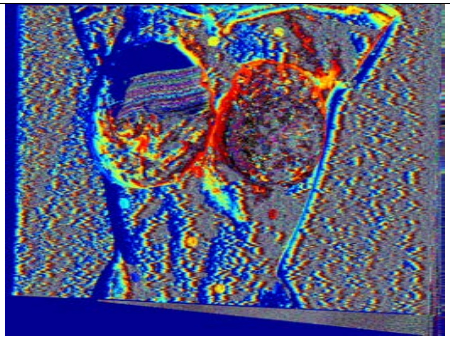
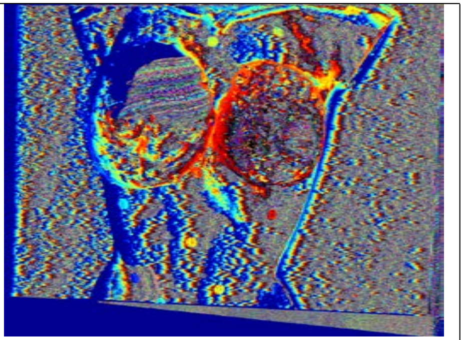
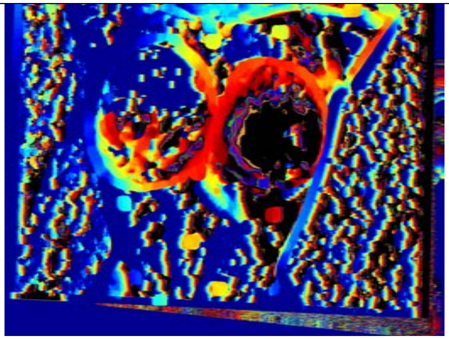
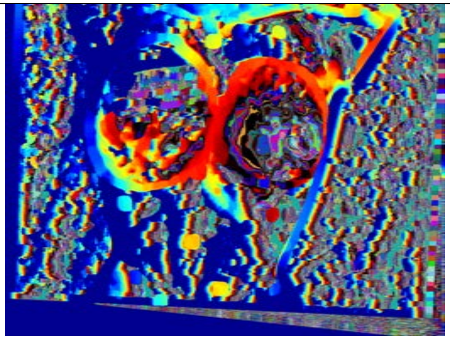
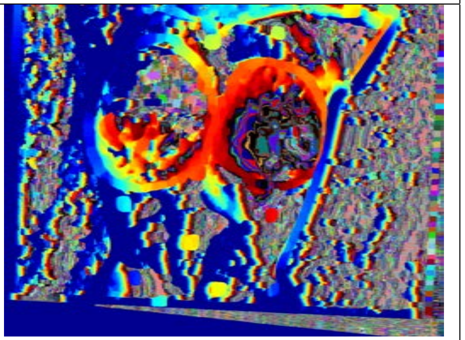
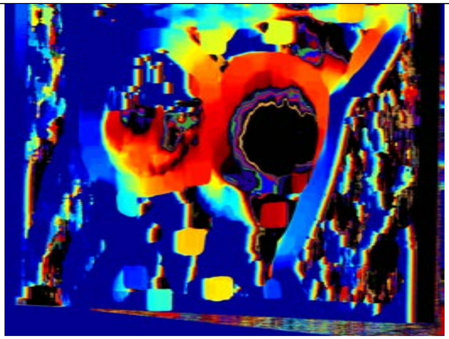
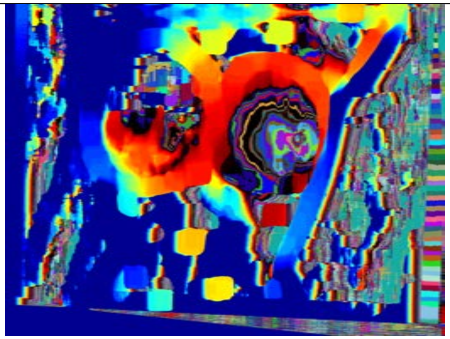
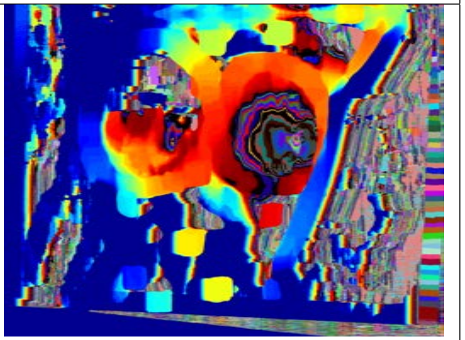
	80	160	255
9			
49			
121			

Table 5.5: Parameter variation for SAD algorithm, for the Dummy model. Column: window size; Row: Maximum Disparity.

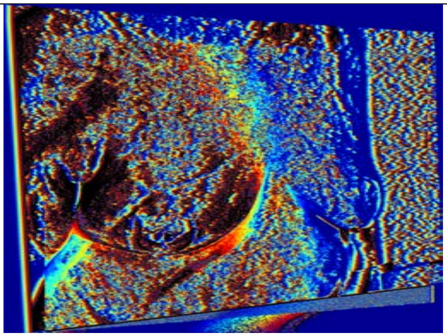
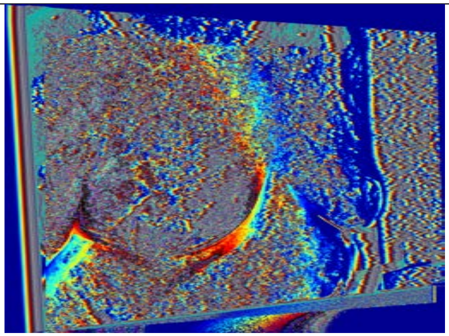
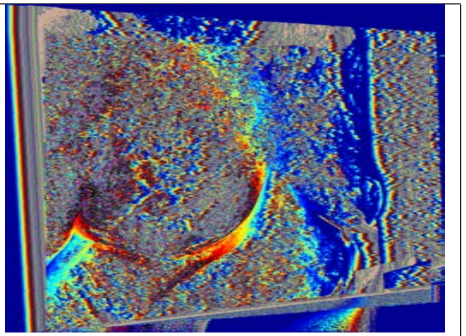
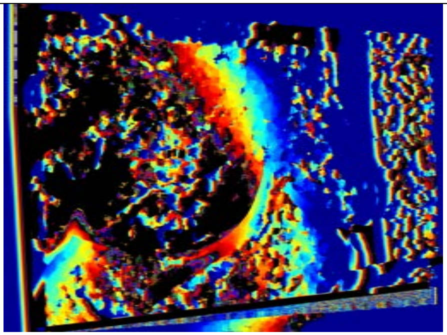
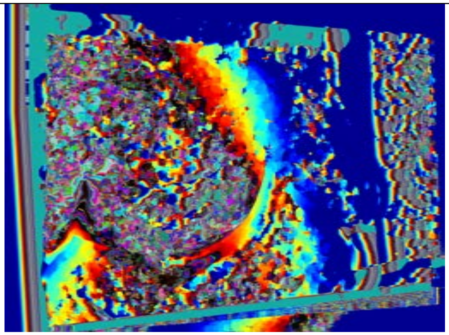
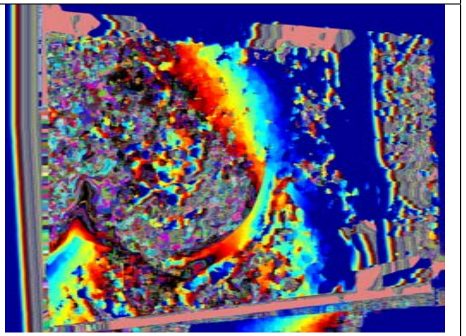
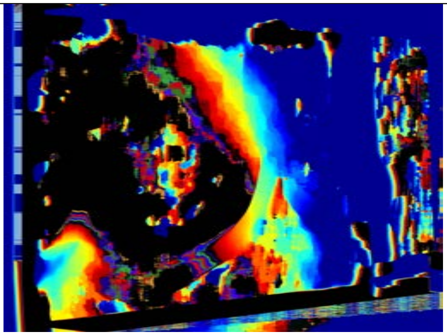
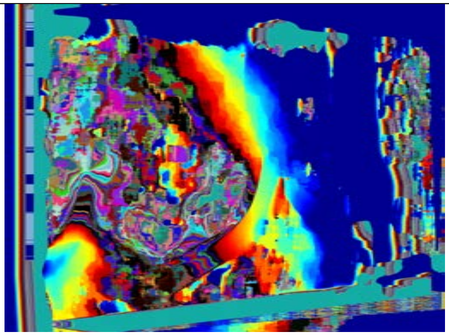
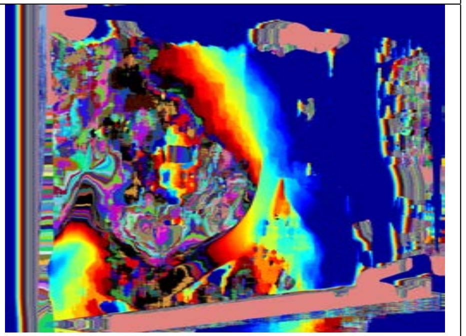
	80	160	255
9			
49			
121			

Table 5.6: Parameter variation for SAD algorithm, running for Patient A images. Column: window size; Row: Maximum Disparity.

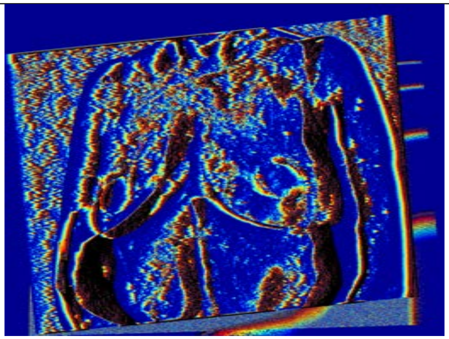
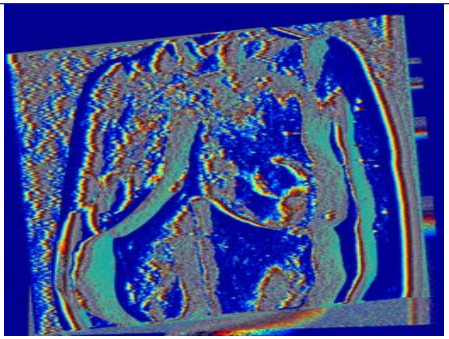
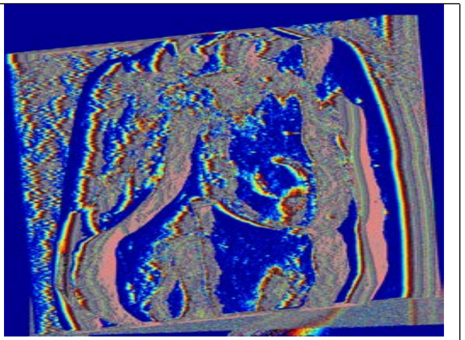
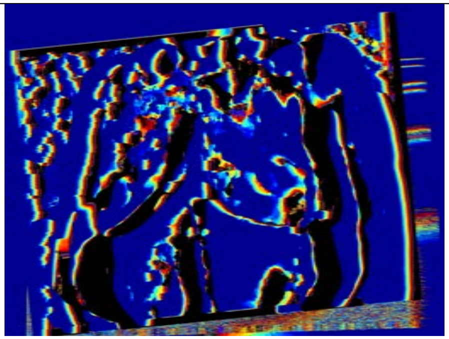
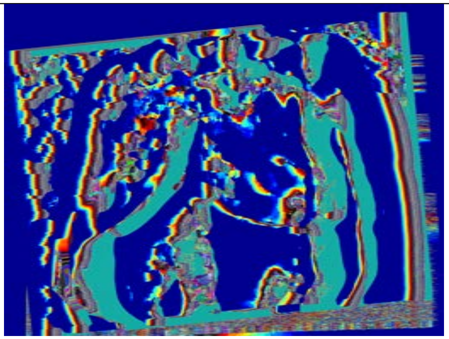
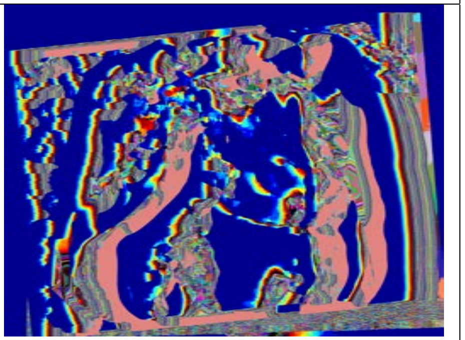
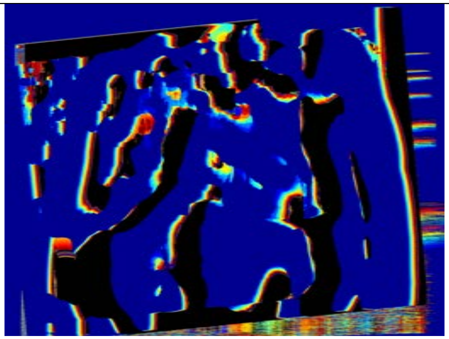
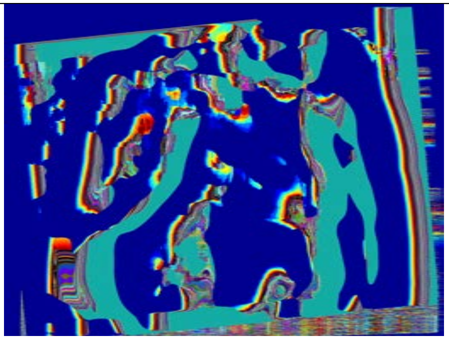
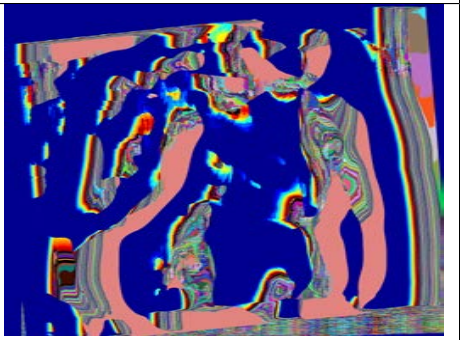
	80	160	255
9			
49			
121			

Table 5.7: Parameter variation for SAD algorithm, running for Patient B images. Column: window size; Row: Maximum Disparity.

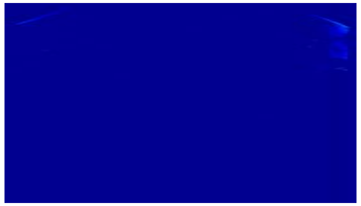
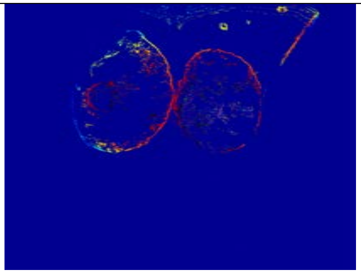
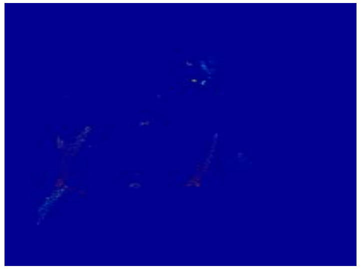
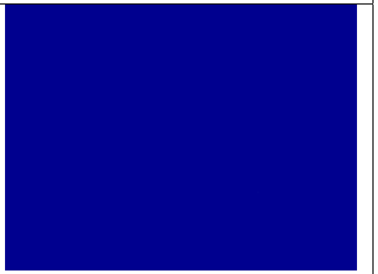
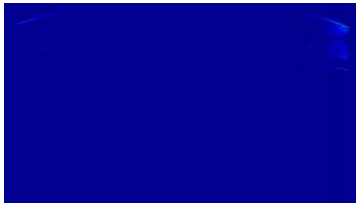
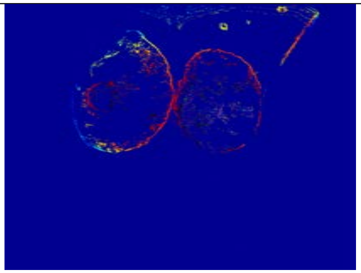
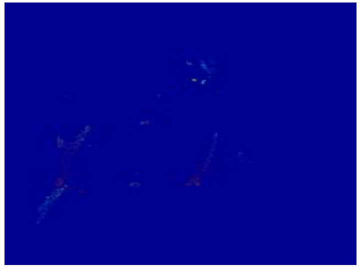
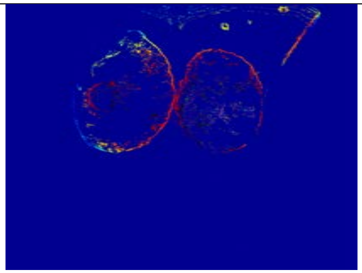
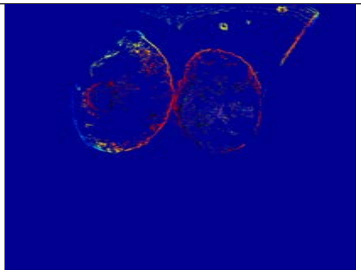
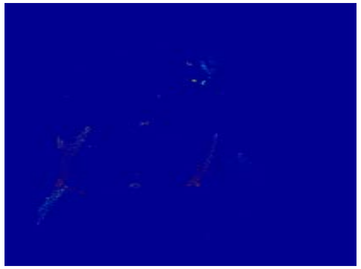
	3D model	Dummy	Patient A	PatientB
9				
49				
121				

Table 5.8: Window size variation for the four pairs of images under test, for Čech and Šára [49].

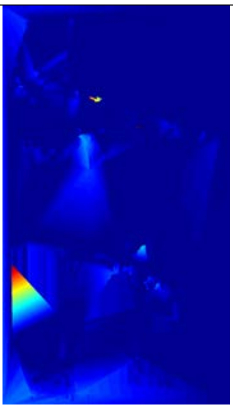
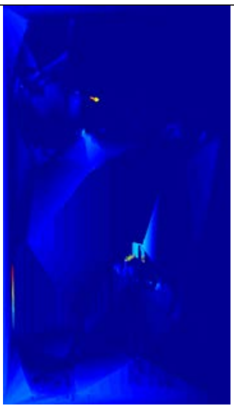
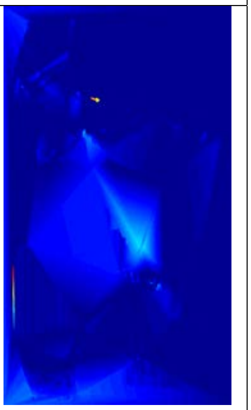
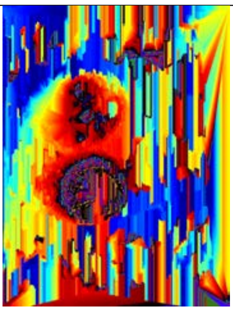
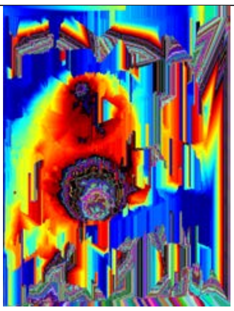
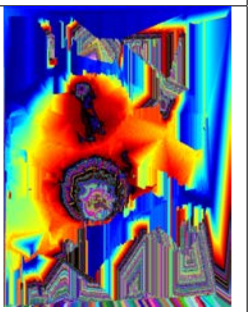
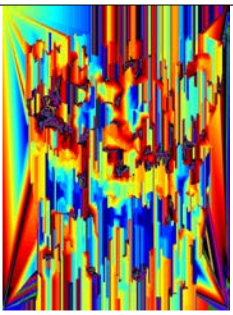
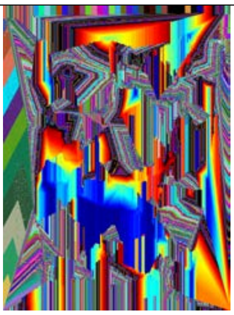
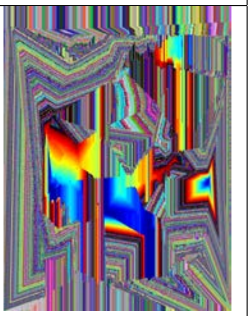
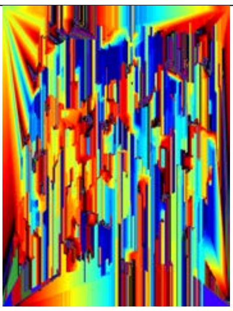
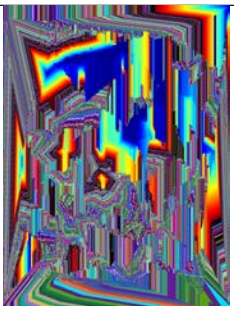
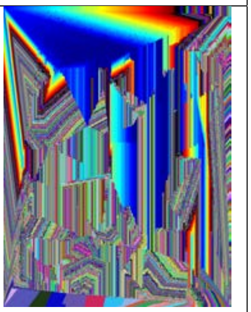
	80	160	255
3D model			
Dummy			
Patient A			
PatientB			

Table 5.9: Maximum disparity value variation for the four pairs of images under test, for Geiger et al. [22].

Chapter 6

Conclusions

This work attempted to create a conclusive set of tests for the 3D reconstruction based on stereo disparity search. After having defined the common mathematics associated, it was possible to establish some basis for the validity of reconstruction and the general problems were posed.

The literature related to this problem is thorough and wide, so a selection of the most recognised, proven and broadly used methods was chosen to take part on this test. The parameters used are those that were shown (by the authors) to have achieved the best results. The results are summarised in the previous chapter.

It is therefore fair to conclude that the current state-of-the-art algorithms for stereo matching are not capable of reliably attributing a dense estimation of the disparity map from the type of images this project deals with. The estimated maps returned from these algorithms are generally very sparse. Visual inspection is enough then to tell that the maps are generally also wrong, whereas areas of the chest sometimes have close disparity from those of the breast. This can be attributed mainly to the difficulties of reasoning to what pixel on the “target” image does one pixel on the “base” pixel belong, or is correlated. Where repetitive texture is present, the reasoning difficulty is very high, and too many options are available for a correct match.

As such, it is required to search for more suitable algorithms that would make more globally oriented decisions. Those algorithms could be helped by a parametric model that would incorporate some restrictions to the possible shape of the breast, reducing the complexity associated to the stereo matching process. Multi-view and/or video incorporation could allow for ambiguity reduction, using methods like optical flow. The biggest adversary to global solution is the elevated complexity associated, and it can be a difficult matter accounting for the possible deployment on normal PCs.

For the evaluation alone it is not needed to have a very high quality disparity map. Some measures can be introduced into the classificator and it might be sufficient for achieving better results. So, a sparse disparity map with just the right features would suffice. The right features would be those enough to add volume loss perception in some mathematical model. A semi-automated solution could improve the rectification process. By asking the user to enter some close-matched points, and then attempting to use the feature selection algorithms to search for features around those points, it could be possible to obtain better quality features in a reasonable number.

There is much further work needed and some upgrades are necessary, like multi-view adaptations, plane fitting algorithms that would allow for adapting a sparser disparity map, and some attention is needed to the problem of difficulties in maintaining the low baseline restriction. There is a large set of methods in these subjects that were not part of this work because of time restrictions.

Chapter 7

Future work

This chapter describes some possible improvements that, although planned and with recognisable possibility, could not be part of the present work, for various reasons. They are to be understood as simple suggestions, once the applicability has not yet been tested.

7.1 Better feature detection and matching

The process of rectification is extensively influential by the correct correspondences being fed to those algorithms. As such, feature detection and matching is, inherently, the biggest concern prior to rectification. There were some algorithms that were not tested, and some more might be better suitable to the scenes that are being dealt with. Some suggestions are present in Section 5.1.2, but there might be more, better suited, or that could be adapted.

Contrary to initial belief, the number of features is not important, as the fundamental matrix needs only 8 points to be correctly formulated. It is much more important the quality of those features. Of course, it is suggestible to have more than 8 points, but too many of mixed quality could produce as bad results, as the good quality model could be lost in so many possibilities. For example, if *RANSAC* is used, then the number of possible solutions for N points is $\binom{N}{8}$. For one of the tests here presented, that would amount to: $\binom{586}{8} \approx 3.2872 \times 10^{17}$ possibilities, what is computationally impossible to test.

7.2 POSIT with a cube

An idea came along while reading [6], of using the *robotics-famous* algorithm *Pose from Orthography and Scaling with Iteration* to establish the camera's full calibration (intrinsics and extrinsics) and, thus, both allowing for a seamlessly rectification, but also to cope with lens distortion and, the most desired feature, a metric reconstruction. The idea is basically to place a small known, recognisable and detectable object with enough distinguishable features. That could be, for example, a cube textured with the classical chessboard images, or a calibration rig like those used by [25]. It should be small enough to be placed beside the patient, and big enough to be usable. Higher definition images would, in principle, reduce the minimum required size for the cube.

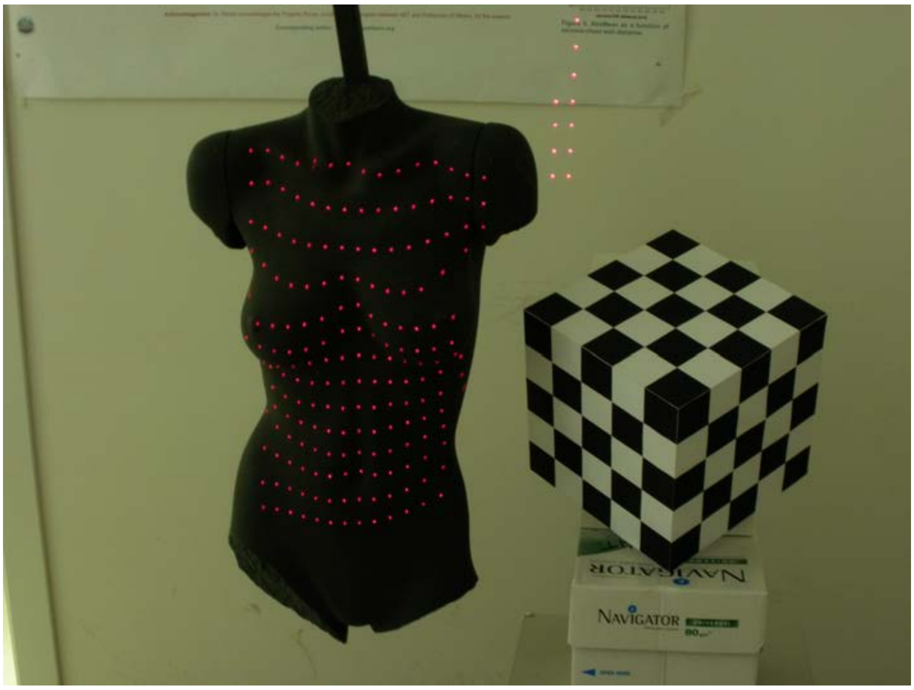


Figure 7.1: Calibration cube near a dummy, for POSIT algorithm. Kindly provided by *Politecnico di Milano*

7.3 Dynamic Time warping

It is possible to use the extension to the original *Dynamic time warping* algorithm to be held at two dimensions. And so, one existing suggestion was to try and use this algorithm to create a very sparse but possibly useful disparity line (or plane) for cosmetic assessment. The idea is to ask for the user to select some feature points manually. Then, using *Dynamic time warping*, try and build a disparity calculation in a line formed by those points, that would be assumed to cross the whole breast through the nipple, defining a disparity measure for a “string” containing the breast. Those measures would be added to those already used in *Breast Cancer Conservative Treatment.cosmetic result* and could improve the cosmetic assessment. Later on, this algorithm could constitute a basis for a new disparity algorithm, adapted to this type of images.

7.4 3D Reconstruction from Multi-view and other techniques

Multiple contributions are present on the literature, both in the form of papers and books. Not much attention was given to the 3D reconstruction from multi-view and other techniques, but there was still some preliminary study on the existing possibilities. Multiple material has been found, and it is depicted in the next paragraphs very briefly.

Ziegler et al. in [58] presents a method for 3D reconstruction based on color and volume consistency using user feature points identified prior to the algorithms application.

Checchi et al describes in [12] a method based on Videogrammetry for reconstructing volcanic surfaces. It deals with automatic camera calibration and 3D measurements from multiple views. Their claim is that reconstruction accuracy is about $10^{-}4$ m for a distance between the object and the camera of 0.5 m, it is potentially several orders of magnitude higher for surfaces of finer texture and uses higher precision sensors.

Wang et al. published in [53] a reconstruction algorithm based on auto-calibration and geometry reconstruction that is based and shared from Zisserman et al. and others. It is based on feature identification and straight lines marking for homographic parameters retrieval. Nevertheless, it requires some user input to identify lines in the images.

Sinha et al. presents in his paper [45] a novel approach for reconstruction that is based on photo consistency (like stereo-matching) and silhouette consistency. Their algorithms are based on global graph-cut optimisations.

Fraundorfer et al. depicts in [19] a method for homograph estimation from a set of identified planes in the images. It grows the reconstruction from feature points that act like seeds.

d'Angelo et al. presents in their paper [15] a combination of photometric methods together with real-aperture measures (real depth sensors) and some notions of Lambertian models (surface illumination calculations).

Salzmann et al. published in [44] a method for coping with noisy video data that take advantage of deploying deformable 3D surface models. The shape of a triangular mesh can be parametrized, so as to create a representative set of potential shapes.

Zhu et al. [57] describes a 3D reconstruction method from uncalibrated cameras without prior knowledge from the structure of the scene. Both intrinsic parameters and feature points can be recovered from the images through the use of a global cost function. The reconstruction generally does not require any further non-linear optimization for general applications.

7.5 Optical flow from video

Video can be considered as a collection of images. Nowadays, with the dissemination of video compression and advanced video codecs, it is not exactly as so. However, it is still possible to use optical flow algorithms based on video that would greatly add the number of views used. For video, there are different algorithms for reconstruction, as can be found in [6, 25, 47] and is a successfully area of studies. One notable implementation is that of *Insight3D*[1].

7.6 Point Cloud Library

When an enough dense mesh is found, this library[2] will allow extensive manipulation and visualisation of that mesh and the resulting reconstruction. It was, once again, created for robotic applications, but it is believed to work in this project. It is computationally improved for speed and reliability, because it is meant to run on limited devices installed on robots. It allows useful operations like point cloud concatenations, filtering, down-sampling, directly applying statistical analysis to detect outliers, estimation of surface normals for surface smoothing, 3D feature points, easing object segmentation, and many other capabilities[3].

[1] As can be seen in `http://insight3d.sourceforge.net/`

[2] Point Cloud Library found in `http://pointclouds.org/`.

[3] Please refer to `http://pointclouds.org/documentation/tutorials/`

References

[1] S. K. Al-Ghazal, R. W. Blamey, J. Stewart, and A. L. Morgan. The cosmetic outcome in early breast cancer treated with breast conservation. *European journal of surgical oncology: the journal of the European Society of Surgical Oncology and the British Association of Surgical Oncology*, 25(6):566–70, December 1999.

[2] B. B. Alagoz. Obtaining depth maps from color images by region based stereo matching algorithms. *OncuBilim Algorithm And Systems Labs*, 08(04):1–13, 2008.

[3] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool. Surf: Speeded up robust features. *Computer Vision and Image Understanding (CVIU)*, 110(3):346–359, 2008.

[4] A. Bhatti and S. Nahavandi. Wavelets/multiwavelets in stereo correspondence estimation: a comparative study. In *Digital Image Computing: Techniques and Application*, volume 2009, pages 309–316, 2009.

[5] K. M. Borgwardt. *Graph Kernels*. PhD thesis, Ludwig-Maximilians University, Germany, 2007.

[6] G. R. Bradski and A. Kaehler. *Learning OpenCV*. O'Reilly, 2008.

[7] M. Brown, R. Szeliski, and S. Winder. Multi-Image Matching Using Multi-Scale Oriented Patches. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 510–517, 2005.

[8] J. S. Cardoso and M. J. Cardoso. Towards an intelligent medical system for the aesthetic evaluation of breast cancer conservative treatment. *Artificial Intelligence in Medicine*, 40(2):115–126, June 2007.

[9] J. S. Cardoso and L. F. Teixeira. Automatic breast contour detection in digital photographs. In *International Conference on Health Informatics (HEALTHINF)*, pages 91–98, 2008.

[10] J. S. Cardoso, J. F. Pinto Da Costa, and M. J. Cardoso. SVMs applied to objective aesthetic evaluation conservative breast cancer treatment. *IEEE International Joint Conference on Neural Networks*, pages 2481–2486, 2005.

[11] J. S. Cardoso, R. Sousa, and L. F. Teixeira. Breast contour detection with stable paths. In *Biomedical Engineering Systems and Technologies*, volume 25, pages 439–452, 2009.

[12] E. Cecchi. N-view reconstruction: a new method for morphological modelling and deformation measurement in volcanology. *Journal of Volcanology and Geothermal Research*, 123(1-2):181–201, April 2003.

[13] D. R. H. Christie, J. A. O'Brien, T. Kron Christie, S. A. Ferguson, C. S. Hamilton, and J. W. Denham. A comparison of methods of cosmetic assessment in breast conservation treatment. *The Breast*, 5(5):358–367, October 1996.

[14] N. Dalal and B. Triggs. Histograms of Oriented Gradients for Human Detection. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 886–893, 2005.

[15] P. D'Angelo and C. Wohler. Image-based 3D surface reconstruction by combination of photometric, geometric, and real-aperture methods. *ISPRS Journal of Photogrammetry and Remote Sensing*, 63(3):297–321, May 2008.

[16] J. M. S. Dias, R. Bastos, J. Correia, and R. Vicente. Semi-Automatic 3D Reconstruction of Urban Areas Using Epipolar Geometry and Template Matching. *Computer-Aided Civil and Infrastructure Engineering*, 21(7):466–485, 2006.

[17] B. Fisher, S. Anderson, J. Bryant, R. G. Margolese, M. Deutsch, E. R. Fisher, J. H. Jeong, and N. Wolmark. Twenty-year follow-up of a randomized trial comparing total mastectomy, lumpectomy, and lumpectomy plus irradiation for the treatment of invasive breast cancer. *New England Journal of Medicine*, 347(16):1233–1241, 2002.

[18] A. Florescu, E. Amir, N. Bouganim, and M. Clemons. Immune therapy for breast cancer in 2010-hype or hope? *Current Oncology*, 18(1):e9–e18, January 2011.

[19] F. Fraundorfer, K. Schindler, and H. Bischof. Piecewise planar scene reconstruction from sparse correspondences. *Image and Vision Computing*, 24(4):395–406, April 2006.

[20] A. Fusiello and L. Irsara. Quasi-euclidean uncalibrated epipolar rectification. In *19th International Conference on Pattern Recognition*, pages 1–4, December 2008.

[21] A. Fusiello and L. Irsara. Quasi-Euclidean epipolar rectification of uncalibrated images. *Machine Vision and Applications*, 22(4):663–670, May 2011.

[22] A. Geiger, M. Roser, and R. Urtasun. Efficient large-scale stereo matching. In *Asian Conference on Computer Vision*, pages 25–38. Springer, 2010.

[23] C. Harris and M. Stephens. A combined corner and edge detector. In *In Proceedings of The Fourth Alvey Vision Conference*, pages 147–151, 1988.

[24] J. R. Harris, M. B. Levene, G. Svensson, and S. Hellman. Analysis of cosmetic results following primary radiation therapy for stages I and II carcinoma of the breast. *International Journal of Radiation Oncology, Biology, Physics*, 5(2):257–261, 1979.

[25] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Cambridge University Press, 2008.

[26] H. Hirschmüller. Stereo processing by semiglobal matching and mutual information. *IEEE transactions on pattern analysis and machine intelligence*, 30(2):328–41, February 2008.

[27] Y. Ke and R. Sukthankar. PCA-SIFT: A more distinctive representation for local image descriptors. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2:506–513, 2004.

[28] T. J. Key, P. K. Verkasalo, and E. Banks. Epidemiology of breast cancer. *The lancet oncology*, 2(3):133–140, 2001.

[29] A. Klaus, M. Sormann, and K. Karner. Segment-Based Stereo Matching Using Belief Propagation and a Self-Adapting Dissimilarity Measure. *18th International Conference on Pattern Recognition (ICPR'06)*, pages 15–18, 2006.

[30] V. Kolmogorov and R. Zabih. Computing visual correspondence with occlusions using graph cuts. In *In International Conference on Computer Vision*, pages 508–515. IEEE Computer Society, 2001.

[31] V. Kolmogorov and R. Zabih. Multi-camera Scene Reconstruction via Graph Cuts. In *Computer Vision — ECCV 2002*, volume 2352, pages 8–40. Springer, 2002.

[32] V. Kumar and R. S. Cotran. *Robbins Basic Pathology*. Saunders Philadelphia, PA, 2003.

[33] C. Loop and Z. Zhang. Computing rectifying homographies for stereo vision. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 125–131, 1999.

[34] D. G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2):91–110, November 2004.

[35] A. Luini, G. Gatti, V. Galimberti, S. Zurrida, M. Intra, O. Gentilini, G. Paganelli, G. Viale, R. Orecchia, P. Veronesi, and U. Veronesi. Conservative treatment of breast cancer: its evolution. *Breast cancer research and treatment*, 94(3):195–8, December 2005.

[36] J. Matas. Robust wide-baseline stereo from maximally stable extremal regions. *Image and Vision Computing*, 22(10):761–767, September 2004.

[37] K. Mikolajczyk and C. Schmid. Performance evaluation of local descriptors. *IEEE transactions on pattern analysis and machine intelligence*, 27(10):1615–30, October 2005.

[38] L. Moisan and B. Stival. A Probabilistic Criterion to Detect Rigid Point Matches Between Two Images and Estimate the Fundamental Matrix. *International Journal of Computer Vision*, 57(3):201–218, May 2004.

[39] P. Monasse and A. Salgado. Quasi-Euclidean Epipolar Rectification (acessed 20/07/2011). 2011. URL http://www.ipol.im/pub/algo/m_quasi_euclidean_epipolar_rectification/.

[40] I. Moreira, I. Amaral, I. Domingues, A. Cardoso, M. J. Cardoso, and J. S. Cardoso. INbreast: Towards a Full Field Digital Mammographic Database.

[41] H. P. Oliveira, A. Magalhães, M. J. Cardoso, and J. S. Cardoso. Improving the bcct. core model with lateral information. In *Proceedings of the 10th IEEE International Conference on Information Technology and Applications in Biomedicine*, pages 1–4, 2010.

[42] M. Özuysal, M. Calonder, V. Lepetit, and P. Fua. Fast keypoint recognition using random ferns. *IEEE transactions on pattern analysis and machine intelligence*, 32 (3):448–461, March 2009.

[43] M. Pollefeys. Visual 3D Modelling from Images (acessed 20/07/2011). 2002. URL http://www.cs.unc.edu/~marc/tutorial/.

[44] M. Salzmann, J. Pilet, S. Ilic, and P. Fua. Surface deformation models for nonrigid 3D shape recovery. *IEEE transactions on pattern analysis and machine intelligence*, 29(8):1481–7, August 2007.

[45] S. N. Sinha and M. Pollefeys. Multi-view reconstruction using photo-consistency and exact silhouette constraints: a maximum-flow formulation. In *20th IEEE International Conference on Computer Vision (ICCV'05)*, pages 349–356, 2005.

[46] S. M. Smith and J. M. Brady. SUSAN—A new approach to low level image processing. *International journal of computer vision*, 23(1):45–78, 1997.

[47] R. Szeliski. *Computer vision: algorithms and applications*. Springer-Verlag New York Inc, 2010.

[48] T. Tuytelaars and K. Mikolajczyk. Local Invariant Feature Detectors: A Survey. *Foundations and Trends in Computer Graphics and Vision*, 3(3):177–280, 2007.

[49] J. Čech and R. Šára. Efficient sampling of disparity space for fast and accurate matching. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2007.

[50] U. Veronesi, N. Cascinelli, L. Mariani, M. Greco, R. Saccozzi, A. Luini, M. Aguilar, and E. Marubini. Twenty-year follow-up of a randomized study comparing breast-conserving surgery with radical mastectomy for early breast cancer. *The New England journal of medicine*, 347(16):1227–32, October 2002.

[51] R. Šára. Finding the Largest Unambiguous Component of Stereo Matching. In *Computer Vision — ECCV 2002*, volume 2352, pages 900–914, 2002.

[52] R. Šára. Robust correspondence recognition for computer vision. In *Compstat 2006-Proceedings in Computational Statistics*, pages 119–131, 2006.

[53] G. Wang, H. Tsui, and Z. Hu. Reconstruction of structured scenes from two uncalibrated images. *Pattern Recognition Letters*, 26(2):207–220, January 2005.

[54] Z. Wang and J. Road. A region based stereo matching algorithm using cooperative optimization. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, June 2008.

[55] W. L. Zhao and C. W. Ngo. lip-vireo Manual. Technical report.

[56] J. Zhou and B. Li. Image rectification for stereoscopic visualization. *J. Opt. Soc. Am. A*, 25(11):2721–2733, November 2008.

[57] Q. Zhu, H. Wang, and W. Tian. A practical new approach to 3D scene recovery. *Signal Processing*, 89(11):2152–2158, November 2009.

[58] R. Ziegler, W. Matusik, H. Pfister, and L. McMillan. 3D reconstruction using labeled image regions. In *Proceedings of the Eurographics Symposium on Geometry Processing*, pages 1–12, 2003.

REFERENCES