



UNIVERSITÀ DELLA CALABRIA
Dipartimento di Meccanica

DOTTORATO DI RICERCA IN INGEGNERIA MECCANICA
CICLO XXV (2009-2012)

SCUOLA DI DOTTORATO "PITAGORA" IN SCIENZE INGEGNERISTICHE

SSD: ING-IND/15 - DISEGNO E METODI DELL'INGEGNERIA INDUSTRIALE

Innovative methodologies
for multi-view 3D reconstruction
of Cultural Heritage

A DISSERTATION SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE
DOCTORAL RESEARCH DEGREE IN MECHANICAL ENGINEERING

Doctoral Research Director
Prof. Sergio Rizzuti

Supervisor
Prof. Fabio Bruno

Candidate
Ing. Alessandro Gallo

a.a. 2011/2012



UNIVERSITÀ DELLA CALABRIA
Dipartimento di Meccanica

DOTTORATO DI RICERCA IN INGEGNERIA MECCANICA

CICLO XXV (2009-2012)

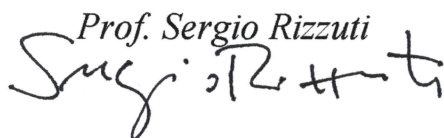
SCUOLA DI DOTTORATO "PITAGORA" IN SCIENZE INGEGNERISTICHE

SSD: ING-IND/15 - DISEGNO E METODI DELL'INGEGNERIA INDUSTRIALE


Innovative methodologies
for multi-view 3D reconstruction
of Cultural Heritage

A DISSERTATION SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE
DOCTORAL RESEARCH DEGREE IN MECHANICAL ENGINEERING

Doctoral Research Director

Prof. Sergio Rizzuti


Supervisor

Prof. Fabio Bruno


Candidate

Ing. Alessandro Gallo



a.a. 2011/2012

A Silvia

Abstract

This dissertation focuses on the use of multi-view 3D reconstruction techniques in the field of cultural heritage. To name just a few applications, a digital 3D acquisition can be used for documentation purposes in the event of destruction or damage of an artefact, or for the creation of museums and virtual tourism, education, structural studies, restoration, etc... All these applications require high precision and accuracy to reproduce the details, but there are other important characteristics such as low cost, ease of use, the level of knowledge needed to operate the systems, which have also to be taken into account. At the present time, the interest is growing around the use of images for the digital documentation of cultural heritage, because it is possible to obtain a 3D model by the means of common photographic equipment. In this work, we have investigated multi-view 3D reconstruction techniques in two specific fields that have not been treated in literature: the 3D reconstruction of small objects (from few mm to few cm) and the survey of submerged archaeological finds.

As for the 3D reconstruction of small objects, a new methodology based on multi-view and image fusion techniques has been developed. The used approach solves the problems related to the use of macro lenses in photogrammetry, such as the very small depth of field and the loss of quality due to diffraction. Since image matching algorithms cannot work on blurred areas, each image of the sequence is obtained by merging pictures acquired at different focus planes. The methodology has been applied on different case studies, and the results have shown that it is possible to reconstruct small complex objects with a resolution of 20 microns and an accuracy of 10 microns.

For which concerns the underwater imaging, a preliminary comparative study between active and passive techniques in turbid water has been conducted. The experimental setup consists in a 3D scanner designed for underwater survey, composed by two cameras and a projector. An analysis on the influence of the colour channel has been conducted, showing how it is possible to obtain a cleaner reconstruction by using the green channel only. The results have shown a denser point cloud when using the passive technique, characterized by missing areas since the technique is more sensible to

turbidity. By contrast, the reconstruction conducted with the active technique have shown more stable results as the turbidity increases, but a greater noise.

A multi-view passive technique has been experimented for the survey of a submerged structure located at a depth of 5 meters, on a seabed characterized by poor visibility conditions and the presence of marine flora and fauna.

We performed an analysis of the performances of a multi-view technique commonly used in air in the first instance, highlighting the limits of the current techniques in underwater environment. In such conditions, in fact, it has not been possible to obtain a complete reconstruction of the scene. The second stage of the process was the testing of image enhancement algorithms in order to improve matching performances in poor visibility conditions. In particular, a variational analysis of the factors that influence the quality of the 3D reconstruction, such as the image resolution and the colour channel, has been performed. For this purpose, the data related to the parameters of interest, such as the number of features extracted or the number of oriented cameras, have been evaluated. The statistical analysis has allowed to find the best combination of factors for a complete and accurate 3D reconstruction of the submerged scenario.

Sommario

La presente tesi tratta l'impiego di tecniche di ricostruzione 3D multi-vista nell'ambito dei beni culturali. Un'acquisizione digitale 3D trova impiego nella documentazione in caso di distruzione o danneggiamenti di un manufatto, nella creazione di musei e turismo virtuale, educazione, analisi strutturale, pianificazione di interventi di restauro, ecc... Queste applicazioni richiedono un'alta precisione e accuratezza per riprodurre tutti i dettagli dell'oggetto, ma altri importanti requisiti di cui si deve tenere conto sono rappresentati dal costo, semplicità di utilizzo e dal grado di specializzazione richiesto. Attualmente, vi è un crescente interesse per l'utilizzo di immagini digitali per la documentazione del patrimonio culturale, in quanto rendono possibile la generazione di contenuti 3D mediante l'utilizzo di comuni apparecchiature fotografiche. Nel presente lavoro sono state studiate tecniche di ricostruzione 3D multi-vista in due ambiti specifici che attualmente non sono stati sufficientemente trattati in letteratura: la ricostruzione 3D mediante tecniche passive di piccoli oggetti (da pochi mm a pochi cm) ed il rilievo di reperti archeologici sommersi.

Nell'ambito della ricostruzione di piccoli oggetti, è stata sviluppata una nuova metodologia basata su tecniche multi-vista e di image fusion. L'approccio utilizzato risolve i problemi che si riscontrano nell'utilizzo di ottiche macro in fotogrammetria, quali la ridotta profondità di campo e la perdita di nitidezza dovuta alla diffrazione. Dal momento che gli algoritmi di image matching non lavorano al meglio se sono presenti aree sfocate, ogni immagine della sequenza è il risultato della fusione di più immagini riprese a diverse profondità di campo. La metodologia è stata applicata su differenti casi studio, e i risultati ottenuti hanno dimostrato che è possibile ricostruire oggetti complessi con una risoluzione di 20 micron e un'accuratezza di circa 10 micron.

Nell'ambito dell'imaging subacqueo, è stato condotto un confronto preliminare tra tecniche attive e passive in condizioni di acqua torbida. Il setup sperimentale si basa su di uno scanner 3D progettato per il rilievo subacqueo composto da due fotocamere ed un proiettore. È stata condotta un'analisi sull'influenza del canale colore, che ha mostrato come sia possibile ottenere ricostruzioni più pulite utilizzando il solo canale verde. I risultati mostrano una maggior densità della nuvola di punti ricostruita mediante

la tecnica passiva, che è caratterizzata però da aree mancanti data la maggior sensibilità alla torbidità. Di contro, le ricostruzioni effettuate con la tecnica attiva presentano una maggior stabilità all'aumentare della torbidità ma un rumore maggiore.

Una tecnica passiva multi-vista è stata sperimentata per il rilievo di una struttura sommersa rinvenuta ad una profondità di 5 metri su un fondale caratterizzato da condizioni di scarsa visibilità e presenza di flora e fauna acquatica. Una prima fase ha previsto l'analisi delle prestazioni di una tecnica multi-vista largamente utilizzata in aria, evidenziando i limiti delle attuali tecniche in ambito subacqueo. In tali condizioni, infatti, non è stato possibile ottenere una ricostruzione completa della scena. Una seconda fase ha previsto la sperimentazione di algoritmi di image enhancement al fine di migliorare le prestazioni degli algoritmi di matching in condizioni di scarsa visibilità. In particolare è stata utilizzata un'analisi variazionale dei fattori che hanno una certa influenza sulla qualità della ricostruzione 3D, quali ad esempio la risoluzione delle immagini ed il canale colore. A tal fine sono stati valutati i dati relativi ad opportuni parametri di interesse, quali il numero di features estratte o il numero di camere orientate. L'analisi statistica ha consentito di trovare la combinazione di fattori ottimale per una completa ed accurata ricostruzione 3D dello scenario sommerso.

Acknowledgments

The research presented in this work would not have been possible without the invaluable guide and support of the following people:

my supervisor Prof. Fabio Bruno for the constant assistance and encouragement during my work;

the Prof. Maurizio Muzzupappa and the other members of my research team;

the Prof. Sergio Rizzuti, Doctoral Research Director;

my colleagues with whom I constantly shared knowledge and experience;

the head of the Centre of Machine Perception (CMP) in Prague, Prof. Vaclav Hlavac and the Prof. Radim Šára for the interesting study experience during the doctoral stage;

the technicians of Laboratory of the Department of Mechanics at the University of Calabria;

and finally Silvia and my family for the constant moral support.

List of Figures

| | |
|---|----|
| Figure 1.1. Image formation pipeline: representation of various sources of noise and the typical digital post-processing steps. | 7 |
| Figure 1.2. Bayer pattern (a) and demosaicing procedure (b). | 11 |
| Figure 1.3. The Foveon sensor. | 12 |
| Figure 1.4. Pinhole camera model. | 13 |
| Figure 1.5. For every pixel m , the corresponding pixel in the second image m' must lie somewhere along a line l' . This property is referred to as the <i>epipolar constraint</i> . See the text for details. | 15 |
| Figure 1.6. Survey of the historical centre of Cosenza (Italy): aerial view of the reconstructed area including camera positions (top) and 3D pointy clouds related to relevant points of interest. | 22 |
| Figure 1.7. 3D point cloud of the Certosa of Serra San Bruno (Italy), destroyed in 1783 by a terrible earthquake. | 23 |
| Figure 1.8. Triangulated surface (a) and textured 3D model (b) of the B. Telesio Statue located in Cosenza (Italy). | 23 |
| Figure 1.9. Triangulated surface (a) and textured 3D model (b) of a wooden statue of Virgin Mary. | 24 |
| Figure 1.10. Triangulated surface (a) and textured 3D model (b) of a bronze archaeological find. | 25 |
| Figure 1.11. Triangulated surface (a) and textured 3D model (b) of a bas-relief with dimensions of 50 x 50 cm. | 25 |
| Figure 2.1. Size of the Airy disk at increasing f-numbers relative to the pixel dimension of a Nikon D5000 digital reflex camera. | 33 |
| Figure 2.2. Pipeline of the implemented methodology for the acquisition and 3D reconstruction of small sized objects. | 35 |
| Figure 2.3. Example of three images acquired at different focus distances and the staked image resulting from the fusion of 22 slices. | 37 |
| Figure 2.4. Effect of diffraction due to the use of a very small aperture. The image resulting from the fusion of 22 slices taken at different focus distance with the aperture set to f-11, is compared to the one taken at f-64, highlighting the loss of sharpness. | 38 |
| Figure 2.5. Experimental setup. A Nikon D5000 reflex camera equipped with a Tamron 90 mm macro lens and the rotational stage, fixed on an optical table. Some led lamps ensure proper lighting. | 40 |
| Figure 2.6. White support plate used to scale the final model and to help the image pre-processing (masking and white balance). | 41 |
| Figure 2.7. Two certified gauge blocks are placed on a rectified surface and mounted on the rotational stage in order to verify the accuracy. | 42 |
| Figure 2.8. Sample image enhanced by using the Wallis filter (right) compared to the original uncorrected image (left). | 42 |

| | |
|---|----|
| Figure 2.9. System resolution and accuracy evaluation: the reconstructed 3D model (a) has been compared to the relative CAD model (b)..... | 43 |
| Figure 2.10. First case study, a bronze zoomorphic fibula with dimensions (LxWxH) of 37x18x16 mm. . | 45 |
| Figure 2.11. Second case study. Four specimens of encrustations, (a) T2, (b) T9 (c) T10 and (d) T16, extracted from a marble statue of a bearded Triton. The support plate is a square of 25 x 25 mm. . | 46 |
| Figure 2.12. Third case study. A calcareous sample (left) and a marble sample (right). The support plate is a square of 25x25mm..... | 47 |
| Figure 2.13. First case study: one of the input image (a), the 3D point cloud (b), the reconstructed polygonal mesh (c) and the final textured 3D model (d)..... | 49 |
| Figure 2.14. First case study. Detail of the tail and wings..... | 49 |
| Figure 2.15. Second case study, T2 specimen: one of the input image (a), the 3D point cloud (b), the reconstructed polygonal mesh (c) and the final textured 3D model (d)..... | 50 |
| Figure 2.16. Second case study, T9 specimen: one of the input image (a), the 3D point cloud (b), the reconstructed polygonal mesh (c) and the final textured 3D model (d)..... | 51 |
| Figure 2.17. Second case study, T10 specimen: one of the input image (a), the 3D point cloud (b), the reconstructed polygonal mesh (c) and the final textured 3D model (d)..... | 52 |
| Figure 2.18. Second case study, T16 specimen: one of the input image (a), the 3D point cloud (b), the reconstructed polygonal mesh (c) and the final textured 3D model (d)..... | 53 |
| Figure 2.19. Detail of pitting on the T9 specimen (a) and gastropod in the T16 specimen. The holes due to pitting measure about 0.4 mm while the gastropod is about 4 mm long..... | 54 |
| Figure 2.20. Third case study, M1 specimen: one of the input image (a), the 3D point cloud (b), the reconstructed polygonal mesh (c) and the final textured 3D model (d)..... | 55 |
| Figure 2.21. Third case study, M2 specimen: one of the input image (a), the 3D point cloud (b), the reconstructed polygonal mesh (c) and the final textured 3D model (d)..... | 56 |
| Figure 2.22. Third case study, details of serpulids on M1 (a) and M2 (b) specimen. The mean diameter of the serpulidae on specimen M2 measures about 1 mm, while from the traces on specimen M1 the diameter of the serpulids range from 0.1 to 0.5 mm. | 57 |
| Figure 2.23. Comparison between the 3D reconstructions obtained by our approach for the specimens T2 (a), T9 (c) and T16 (d) and the one obtained using the multi-view stereo technique on a sequence of images taken at an aperture value of f-64 (b), (d) and (f). The diffraction effect causes a loss of detail that result in a smoother surface. | 59 |
| Figure 2.24. Comparison between the 3D reconstruction obtained by our approach for the specimen M1 (a) and M2 (c) and the one obtained using a sequence of images taken at the maximum f-number (b) and (d). The diffraction effect causes a loss of detail that result in a smoother surface. | 59 |
| Figure 2.25. Comparison between the 3D reconstruction obtained by our approach for the specimens T2 (a) and T16 (d) and the one obtained using a depth from focus algorithm on a stack of 14 images taken at f-11 (b) and (e) and on a stack of 70 images taken at the smallest aperture of f-5.6 (c) and | |

| | |
|--|-----|
| (f). The depth resolution relative to the depth from focus algorithm is equal to half DoF, 0.5 mm and 0.25 mm respectively, against a resolution of 20 micrometer of our approach. | 61 |
| Figure 2.26. Comparison between the 3D reconstruction obtained by our approach for the specimens M1 (a) and the one obtained using a depth from focus algorithm on a stack of 86 images taken at f-4.5 (b). The depth resolution relative to the depth from focus algorithm is equal 0.25 mm, against a resolution of 20 micrometer of our approach..... | 61 |
| Figure 3.1. On the left, triangulation of a stereo configuration composed by two cameras and a projector. The points m_L and m_R in left and right images respectively, are the projection on the image planes I_L and I_R of the same 3D point w of the object. On the right, examples of binary patterns and code shifting. | 69 |
| Figure 3.2. Optical setup of the underwater 3D acquisition system: underwater calibration (left) and acquisition (right). | 71 |
| Figure 3.3. Optical setup of the underwater 3D system. | 72 |
| Figure 3.4. Images of the two objects with a projected gray-code pattern, acquired at different turbidity levels. | 73 |
| Figure 3.5. Complete distortion model (radial plus tangential distortions) in air and clear water. The cross and the circle indicate the image centre and the principal point respectively. The contours represent the optical distortions values (in pixel)..... | 75 |
| Figure 3.6. Distribution of the re-projection error (in pixel) in air and clear water (RGB images)..... | 75 |
| Figure 3.7. Point cloud cleaning procedure. Outlier point manual selection a), selected outlier points (b), cleaned point cloud after outlier deletion (c) and final point cloud after noise filtering (d). | 79 |
| Figure 3.8. Point clouds obtained in different environment condition for both active and passive technique. | 81 |
| Figure 3.9. Acquired images and unclean point clouds at a light turbidity level (T1) in each colour channel: the blue points are due to scattering effect, highly reduced in green channel. | 81 |
| Figure 4.1. Geographical localization of underwater archaeological site of Baia (Italy)..... | 89 |
| Figure 4.2. Calibration panel used to manually correct the white balance..... | 91 |
| Figure 4.3. Image acquisition using a standard aerial photography layout (a) and using oblique photographs (b)..... | 92 |
| Figure 4.4 Results of the camera orientation process on original uncorrected pictures: due to the presence of a low contrasted sandy seabed, the whole scene has been reconstructed in two separated blocks relative to the North part (up, 384 pictures) and to the South part (down, 116 pictures)..... | 94 |
| Figure 4.5. Reconstructed point cloud of the North area..... | 95 |
| Figure 4.6. Reconstructed point cloud of the South area..... | 95 |
| Figure 4.7. Reconstructed surface and textured model of the North area. | 96 |
| Figure 4.8. Reconstructed surface and textured model of the South area. | 97 |
| Figure 5.1. A sample image enhanced with the ACE method (b), HIST method (c) and PCA method (d) compared to the original image (a)..... | 103 |

| | |
|--|-----|
| Figure 5.2. Pipeline of the implemented methodology for the reconstruction of submerged scenes. | 104 |
| Figure 5.3. Sample images related to image set 1 (a), image set 2 (b), image set 3 (c), image set 4 (d), image set 5 (e), image set 6 (f) and image set 7 (g)..... | 108 |
| Figure 5.4. Example of an anova table related to the parameter “mean extracted features”..... | 112 |
| Figure 5.5. Graphs obtained with the <i>multicompare</i> function of the factors image enhancement method (a) and image pyramid level (b) related to the parameter “Mean number of extracted features”..... | 113 |
| Figure 5.6. Example of the histogram representing the data related to the mean number of extracted features..... | 115 |
| Figure 5.7. Anova table relative to the parameter “mean extracted features”..... | 116 |
| Figure 5.8. Mean number of extracted features: detailed comparison related to all the configurations. ... | 117 |
| Figure 5.9. Mean values and confidence intervals related to the parameter <i>mean number of extracted features</i> | 118 |
| Figure 5.10. Anova table related to the parameter “mean extracted features” measured on images belonging to the second image set..... | 119 |
| Figure 5.11. Mean values and confidence intervals related to the parameter <i>mean number of extracted features</i> of the second image set..... | 119 |
| Figure 5.12. Anova table relative to the parameter “ <i>percentage of matched features</i> ”..... | 120 |
| Figure 5.13. Mean values and confidence intervals related to the parameter <i>percentage of matched features</i> | 121 |
| Figure 5.14. Percentage of matched features: detailed comparison related to all the configurations. | 122 |
| Figure 5.15. Anova table related to the parameter “ <i>percentage of oriented cameras</i> ”..... | 123 |
| Figure 5.16. Mean values and confidence intervals related to the parameter <i>percentage of oriented cameras</i> | 124 |
| Figure 5.17. Anova table related to the parameter “percentage of oriented cameras” computed from the data related to the RGB images only, which include the single component extracted by PCA correction. | 125 |
| Figure 5.18. Mean values and confidence intervals related to the parameter <i>percentage of oriented cameras</i> , computed from the data related to only the RGB images which include the single component from PCA correction. | 125 |
| Figure 5.19. Percentage of oriented cameras: detailed comparison related to all the configurations. | 126 |
| Figure 5.20. Anova table related to the parameter “ <i>bundle adjustment mean reprojection error</i> ”..... | 127 |
| Figure 5.21. Mean values and confidence intervals related to the parameter <i>bundle adjustment mean reprojection error</i> | 128 |
| Figure 5.22. Bundle adjustment mean reprojection error: detailed comparison related to all the configurations..... | 129 |
| Figure 5.23. Mean values and confidence intervals related to the parameter <i>mean number of extracted features</i> : the analysis has been performed on images acquired using custom white balance correction. | 131 |

| | |
|---|-----|
| Figure 5.24. Performances comparison of the image enhancement methods for each measured parameter. The performances of the HIST, ACE and PCA image enhancement algorithms, are compared to a reference value related to the original images (OR)..... | 132 |
| Figure 5.25. Performances comparison of the image resolution (image pyramid) for each measured parameter. The performances of the first pyramid (High res – images resized to 50%), the second pyramid (Med res – images resized to 25%) and the third pyramid (Low res – images resized to 6.25%), are compared taking as reference the value related to the low resolution images. | 132 |
| Figure 5.26. Performances comparison of the colour channels for each measured parameter. The performances of the R, G and B components, are compared taking as reference the value related to the RGB images..... | 133 |
| Figure 5.27. Results of the camera orientation process (enhanced pictures with HIST method): sparse point cloud and 533 oriented pictures..... | 133 |
| Figure 5.28. Reconstructed dense point cloud. | 134 |
| Figure 5.29. Reconstructed surface. | 134 |
| Figure 5.30. Final textured 3D model - a..... | 137 |
| Figure 5.31. Final textured 3D model - b. | 138 |

List of Tables

| | |
|--|-----|
| Table 2.1. The three samples prepared arranging in pair three certified gauge blocks..... | 41 |
| Table 2.2. mean values of the extracted SIFT features for the datasets relative to the tree specimens, reconstructed using the original images and the one enhanced by using the Wallis filter. | 43 |
| Table 2.3. System accuracy evaluation. The experimental results show that the system is capable to resolve up to a 20 μm step with an accuracy of about 10 microns..... | 44 |
| Table 2.4. Results of the 3D reconstruction relative to the three case studies..... | 57 |
| Table 2.5. Time required to complete the whole 3D reconstruction pipeline. Most of the time is needed for the image acquisition phase. | 58 |
| Table 3.1. Measurements of the turbidity conditions. | 70 |
| Table 3.2. Intrinsic parameters (mean values for left and right camera) in air and water (RGB images)... | 74 |
| Table 3.3. Accuracy evaluation on a planar and cylindrical sample in air and water..... | 78 |
| Table 3.4. Acquired 3D points per 100 pixel (N_p for passive and N_a for active stereo) and percentage values of deleted 3D points ($N_{pc}\%$ for passive $N_{ac}\%$ for active stereo) for Amphora and Aeolus... | 79 |
| Table 3.5. Percentage values of deleted 3D points for active stereo technique in each colour channel, red ($N_{ac}\%$ -R), green ($N_{ac}\%$ -G) and blue ($N_{ac}\%$ -B), compared respect the values computed from RGB images ($N_{ac}\%$ -RGB). The table is related to the object Amphora..... | 82 |
| Table 3.6. Geometrical error (mean μ and standard deviation σ) calculated for active technique in the green channel: the point clouds are compared respect to a reference (Ref.)..... | 82 |
| Table 3.7. Geometrical error (mean μ and standard deviation σ) calculated for passive technique in the green channel: the point clouds are compared respect to a reference (Ref.)..... | 83 |
| Table 5.1. Influence factors and related levels used to perform the analysis of variance. | 106 |

Contents

| | |
|--|-----|
| Abstract..... | ii |
| Sommario..... | iv |
| Acknowledgments | vi |
| List of Figures..... | vii |
| List of Tables..... | xii |
| Introduction | 1 |
| Thesis objectives..... | 4 |
| Structure of the thesis | 5 |
| 1. Image formation and 3D reconstruction from multiple views | 6 |
| 1.1. Image formation..... | 6 |
| 1.1.1. The digital camera | 6 |
| 1.1.2. Colour filter arrays and demosaicing..... | 10 |
| 1.2. Geometry of image formation and multi-view Geometry..... | 12 |
| 1.2.1. Pinhole Camera Model | 12 |
| 1.2.2. Epipolar Geometry | 15 |
| 1.2.3. Projective Reconstruction and Self-Calibration..... | 16 |
| 1.3. Multi-view 3D Reconstruction..... | 16 |
| 1.3.1. Camera calibration..... | 17 |
| 1.3.2. Structure from motion | 18 |
| 1.3.3. Multi view stereo | 19 |
| 1.4. Experimentation of a multi-view stereo algorithm in real case studies | 21 |
| 1.4.1. Large scale scenes | 21 |
| 1.4.2. Monuments and statues | 23 |
| 2. Multi-view 3D reconstruction of small sized objects..... | 26 |
| 2.1. Introduction..... | 26 |
| 2.2. Related works..... | 28 |
| 2.3. Theoretical background | 29 |

| | | |
|--------|--|----|
| 2.3.1. | Macro photography..... | 29 |
| 2.3.2. | Image fusion techniques for extended Depth of Field..... | 30 |
| 2.3.3. | Depth from focus..... | 31 |
| 2.3.4. | Multi-view photogrammetric reconstruction..... | 31 |
| 2.3.5. | Problems in macro photogrammetry..... | 32 |
| 2.4. | The proposed methodology for small object reconstruction..... | 34 |
| 2.4.1. | Image acquisition..... | 34 |
| 2.4.2. | Determination of shooting parameters..... | 34 |
| 2.4.3. | Image stack acquisition..... | 36 |
| 2.4.4. | Image fusion and pre-processing..... | 36 |
| 2.4.5. | Multi-view 3D reconstruction..... | 37 |
| 2.4.6. | Triangulation..... | 39 |
| 2.4.7. | Surface refinement..... | 39 |
| 2.4.8. | Scaling and texturing..... | 39 |
| 2.5. | Experimental setup..... | 40 |
| 2.6. | Accuracy evaluation..... | 41 |
| 2.7. | Experimentation..... | 44 |
| 2.7.1. | Description of the case studies..... | 44 |
| 2.7.2. | Image acquisition..... | 47 |
| 2.7.3. | 3D reconstruction..... | 48 |
| 2.7.4. | Experimental results..... | 48 |
| 2.8. | Comparison with depth from focus technique..... | 60 |
| 2.9. | Summary and discussion..... | 62 |
| 3. | Comparative analysis between active and passive stereo techniques for underwater 3d modelling of close-range objects..... | 64 |
| 3.1. | Introduction..... | 64 |
| 3.2. | Related works..... | 65 |
| 3.2.1. | Optical imaging systems..... | 65 |
| 3.2.2. | Camera calibration in underwater environment..... | 66 |
| 3.3. | Active and passive stereo techniques..... | 67 |
| 3.3.1. | 3D stereo reconstruction..... | 67 |
| 3.3.2. | Calibration..... | 69 |

| | | |
|--------|---|-----|
| 3.4. | Experimentation..... | 69 |
| 3.4.1. | Underwater setup..... | 71 |
| 3.4.2. | Calibration..... | 74 |
| 3.4.3. | 3D reconstruction - Passive stereo..... | 76 |
| 3.4.4. | 3D reconstruction - Active stereo..... | 77 |
| 3.4.5. | Accuracy evaluation..... | 77 |
| 3.5. | Results analysis..... | 77 |
| 3.5.1. | Point cloud editing..... | 78 |
| 3.5.2. | 3D points density..... | 78 |
| 3.5.3. | Multi-channel analysis..... | 80 |
| 3.5.4. | Geometrical error..... | 82 |
| 3.6. | Discussion..... | 83 |
| 3.7. | Concluding Remarks..... | 84 |
| 4. | Experimentation of multi-view 3D reconstruction techniques in underwater environment..... | 85 |
| 4.1. | Introduction..... | 85 |
| 4.2. | Related works..... | 86 |
| 4.3. | Archaeological context..... | 88 |
| 4.4. | Experimentation..... | 89 |
| 4.4.1. | Experimental setup..... | 90 |
| 4.4.2. | Image acquisition..... | 90 |
| 4.4.3. | Multi-view 3D reconstruction..... | 93 |
| 4.5. | Conclusions..... | 98 |
| 5. | Underwater image pre-processing algorithms for the improvement of multi-view 3D reconstruction..... | 99 |
| 5.1. | Introduction..... | 99 |
| 5.2. | State of the art of underwater image pre-processing algorithms..... | 101 |
| 5.3. | Colour enhancement of underwater images..... | 102 |
| 5.4. | Description of the adopted methodology..... | 104 |
| 5.4.1. | Design of the experimental campaign..... | 104 |
| 5.4.2. | Datasets generation..... | 110 |
| 5.5. | Statistical analysis..... | 110 |

| | |
|--|-----|
| 5.5.1. Anova analysis..... | 111 |
| 5.5.2. Mean extracted features..... | 114 |
| 5.5.3. Percentage of matched features | 119 |
| 5.5.4. Percentage of oriented cameras | 123 |
| 5.5.5. Bundle adjustment mean reprojection error | 127 |
| 5.5.6. Effects of white balance correction | 128 |
| 5.5.7. Discussion..... | 130 |
| 5.6. Results..... | 133 |
| 5.7. Conclusions..... | 136 |
| Conclusions | 139 |
| Bibliography | 141 |

Introduction

3D modelling is defined as the process that starts with the acquisition of metric data and ends with the creation of a 3D virtual model with which is possible to interact on a computer. Usually the term is used to denote the transition from a point cloud to a continuous surface, which is actually only one of the steps of a more complex process.

The modelling of 3D objects and scenes is nowadays a topic of great interest not only in industry, robotics, navigation and body scanning, but also in the field of Cultural Heritage. A digital 3D acquisition can be used for documentation in the event of destruction or damage of an artifact, the creation of museums and virtual tourism, education, structural studies, restoration, etc... All these applications require high precision and accuracy to reproduce the details, but there are other important characteristics such as low cost, ease of handling, the level of automation in the process, which have also to be taken into account.

Digital models are nowadays employed everywhere, widely diffused through the Internet and handled on low-cost computers and smartphones. Even if the creation of a simple 3D model seems an easy task, a considerable effort is still required to get a precise model of a complex object, painted with a high resolution texture.

Generally speaking, the three-dimensional object measurement techniques can be classified into *contact methods*, such as CMM (coordinate measuring machines), and *non-contact methods*, such as laser scanning and photogrammetry.

There are different approaches that can be used for the generation of a 3D model using non-contact systems, which can be grouped in active and passive techniques in relation to the type of sensor employed.

Regarding both active and passive techniques, the methods used to build a 3D model from digital images can be classified in:

- 1) *Image-based rendering*. This method creates new views of a 3D scene directly from the images (Shum & Kang 2000), and for this purpose it is necessary to know the camera positions. Alternatively it is possible to compute the camera motion from a sequence of very small baseline images.

- 2) *Image-based modelling*. The image based methods use passive sensors (still or video cameras) to acquire images from which 3D information can be extracted. This method is widely used for city modelling (Grün 2000) and architectural object reconstruction (El-Hakim 2002), and may be implemented with simple and low-cost hardware. Although the best results can be obtained when user interaction is involved, the image-based modelling methods retrieve the 3D coordinates of a pixel by solving the correspondence problem or using shape-form-x approaches, such as shape from stereo, shape from shading and shape from silhouette methods.
- 3) *Range-based modelling*. This method requires specialized and expensive hardware, commonly known as 3D scanners, but directly provides the 3D data of the object in the form of a point cloud. Usually this system does not provide colour information, but recent laser scanners integrate a colour camera mounted in a fixed and known position relative to the scanner reference system. However, in order to obtain the best results in terms of surface texture, it is necessary to use separated still cameras (Beraldin *et al.* 2002; Guidi *et al.* 2003). 3D scanners can be used on different scales, from tens of meters (e.g.: buildings, monuments, etc.) to few centimetres (e.g.: coins, pottery, etc.), and often require multiple scans that will be aligned to cover the whole object surface.
- 4) *Combination of both image-based and range-based modelling*. In some cases, such as the reconstruction of large architectural places, the use of just one of the two methods is not sufficient to describe the geometrical information of the object, so the use of a technique integrating range-based and image-based methods is necessary (El-Hakim & Beraldin 1995).

After the measurements, a polygonal surface has been created in order to build a realistic representation of the object. Then, the model can be mapped with a high quality texture to allow a photo-realistic visualization, necessary for immersive visualization and interaction.

This thesis will focus on *image-based modelling*, with special attention on passive imaging methods used to acquire 3D information from multiple views in the context of Cultural Heritage.

At the present time, there is an increasing interest in the use of images for the digital documentation of cultural heritage (Zubrow 2006). Their high flexibility, reduced cost and low level of knowledge required to inexperienced users, make the use of images the best choice for the 3D survey of individual artefacts or complex archaeological sites, although the usability of the raw data produced is non-straightforward and some post-processing becomes necessary.

Several new tools that can be used to obtain three-dimensional information from unorganized image collections are now available for the public and used for survey. Both COTS and open source solutions based on dense stereo matching are available, giving the possibility to generate 3D content without the need of high-cost hardware, just by using a simple still camera.

The commercial software PhotoScan (Agisoft LCC. 2010) has demonstrated to be an effective tool for 3D content development from overlapping images (Verhoeven 2011), making the reconstruction very straightforward, since the process is almost automatic and consist in a simple three-step procedure (image alignment, 3D model creation, texturing).

The use of open source 3D modelling solutions is becoming increasingly popular in the documentation of Cultural Heritage. The workflow based on Bundler (Snavely *et al.* 2007) for camera network orientation and PMVS2 (Furukawa & Ponce 2007) for dense reconstruction is one of the most widely used. These tools do not require taking care during acquisition, since the process can be successfully applied on unordered image collections or on pictures taken from the Internet (Agarwal *et al.* 2009). Moreover, the image orientation stage is carried out without user interaction and there is no need to use markers, which were usually employed in the standard photogrammetric survey. However, the “freedom” of these tools may affect the quality and reliability of the survey by reducing the possibility to use their output as a metric model, making them fit for visualization purposes only. With a more careful planning of the acquisition phase and with a little effort in the post-processing stage, these methods can be used for accurate 3D survey.

In the field of Cultural Heritage these 3D modelling solutions have been successfully applied in different works. In (Dellepiane *et al.* 2012) a methodology based on the use

of open source dense stereo matching systems has been presented. It has been applied on the monitoring of the archaeological excavation at the archaeological site of Uppàkra. In (Barazzetti *et al.* 2011) low cost software has been used to reconstruct the “G1” temple in MySon, Vietnam, which was listed as a UNESCO World Cultural Heritage site, demonstrating the effectiveness and accuracy of the methodology.

Image-based 3D modelling techniques represent a valid alternative to more specialized and expensive equipment and many research areas and applications still have to be explored in depth.

The research presented in this thesis focuses on two specific areas that, at the present, have not yet been adequately addressed: the 3D reconstruction of small objects and the survey of submerged structures in turbid underwater environment. Specific methodologies and tools have been developed to increase the performances of multi-view 3D reconstruction algorithms in these two challenging fields.

Thesis objectives

The main objective of this thesis is the study and the application of passive multi-view stereo techniques for Cultural Heritage purposes, aiming to satisfy the following needs:

- Developing a methodology and an experimental setup for the acquisitions of small sized objects;
- Testing the performances of the combination of a passive multi-view stereo technique and an image fusion method for the 3D reconstruction of small sized objects;
- Comparing two whole-field (active and passive) techniques for 3D underwater acquisition;
- Understanding the performances of image enhancement algorithms for acquisitions in underwater environment;
- Experimentation of a multi-view stereo technique for the 3D reconstruction of submerged structures in high turbidity conditions.

Structure of the thesis

Chapter 1 presents the background about the image formation in digital cameras and 3D reconstruction from multiple views.

Chapter 2 presents a new methodology for the 3D reconstruction of small sized objects (from few millimetres to few centimetres) using a multi-view stereo algorithm on a sequence of images acquired with a macro lens. Each image of the sequence is obtained with an image fusion technique that merges a stack of pictures acquired at different focus planes to extend the small depth of field at high magnification.

A comparison between an active stereo technique and a passive stereo technique in underwater environment is discussed in Chapter 3. The experimental tests on medium-sized objects have been conducted in a water tank to control the turbidity.

Chapter 4 describes the survey conducted at the underwater archaeological site of Baiae (Naples, Italy), characterized by a poor visibility due to turbidity. This condition negatively affects the 3D reconstruction, preventing to create a complete model of the site.

Chapter 5 presents the methodology developed for the 3D reconstructions in underwater environment. This is based on the use of image enhancement methods to improve the performance of a multi-view stereo technique. The statistical approach used to test the effects of different colour enhancement methods on the performance of image matching algorithms has been described. In particular the DOE (Design of Experiments) has been used to find the best combination of factors for a complete and accurate reconstruction of the submerged structure described in Chapter 4, demonstrating the effectiveness of image pre-processing.

The final Conclusions chapter summarizes the considerations about the results of multi-view imaging techniques applied in unfavourable conditions.

Image formation and 3D reconstruction from multiple views

The real world three-dimensional geometries are projected in two-dimensional information in an image, which express this data by discrete colour or intensity values. 3D reconstruction algorithms can recover 3D information from a sequence of 2D overlapping images.

1.1. Image formation

A digital image is a matrix of elements, the pixels (picture element), which contain the radiometric information that can be expressed by a continuous function. The radiometric content can be represented by black and white values, gray levels or RGB (Red, Green and Blue) values.

Regardless of the method of acquisition of a digital image, must be considered that an image of a natural scene is not an entity expressible with a closed analytical expression and therefore it is necessary to search for a discrete function representing it.

The digitization process converts the continuous representation in a discrete representation, by sampling the pixels and quantizing the radiometric values. This is the process that occurs directly in a digital camera.

The imaging sensor collects the information carried by electromagnetic waves and measures the amount of incident light, which is subsequently converted in a proportional tension transformed by an Analog to Digital converter (A/D) in a digital number.

1.1.1. The digital camera

The light starting from one or more light sources which reflects off one or more surfaces and passes through the camera's lenses, reaches finally the imaging sensor. The photons arriving at this sensor are converted into the digital (R, G, B) values forming a digital

image. The processing stages that occur in modern digital cameras are depicted in fig. 1.1 which is based on camera models developed by (Healey & Kondepudy 1994; Tsing *et al.* 2001; Liu *et al.* 2008).

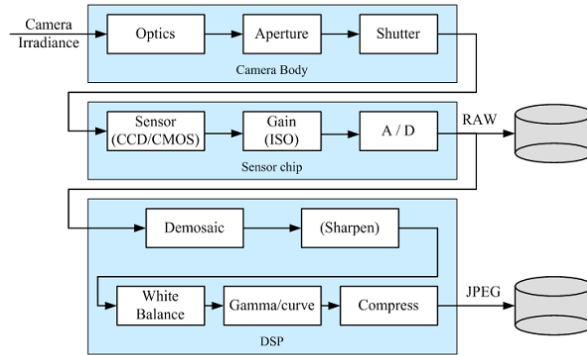


Figure 1.1. Image formation pipeline: representation of various sources of noise and the typical digital post-processing steps.

Light arriving on an imaging sensor is usually collected by an active sensing area, integrated for the exposure time (expressed as the shutter speed in a fraction of a second, e.g., 1/60, 1/200, etc.), and then passed to a set of amplifiers.

The two main types of sensor used in digital still and video cameras are *charge-coupled device* (CCD) and *complementary metal oxide on silicon* (CMOS). Each of them is composed by an array of elements, photodiodes, which are able to convert the light intensity in electric charge. What differentiates the two types is the mode with which the electric charge is converted into voltage values and its transfer from the chip to the camera

In a CCD, photons are accumulated during the exposure time in each active well. Then, the charges are transferred from well to well until they are deposited at the amplifiers, which amplify the signal transferring it to an ADC (analog-to-digital converter). In older CCD sensors the blooming effect may occurs when charges are transferred from one over-exposed pixel into adjacent ones, but anti-blooming technology in the most newer CCDs avoid this problem.

In CMOS, the photons falling on the sensor directly affect the conductivity (or gain) of a photodetector, which can be selectively managed to control exposure duration, and locally amplified before being transferred using a multiplexing scheme.

The main factors affecting the performance of a digital image sensor are represented by:

- shutter speed
- sampling pitch,
- fill factor
- sensor size
- analog gain
- sensor noise
- resolution
- ADC quality

Shutter speed

The shutter speed (namely the exposure time) controls the amount of light falling on the sensor and, hence, determines the correct exposure of the image. Given the focal length there is a minimum value of the shutter speed that have to be used to avoid blur due to the photographer vibrations. A practice rule indicates a safety time equal to $1/f$ (focal length) but newer vibration reduction system allow increasing this value.

This rule is only valid for movement due to the operator. For a dynamic object, the shutter speed also determines the amount of motion blur in the resulting picture.

Sampling pitch

The sampling pitch is the physical spacing between adjacent photodetector on the sensor. A smaller sampling pitch gives a higher sampling density providing a higher resolution (in terms of pixels) for a given active area on the imaging sensor. However, a smaller pitch implies a smaller area which makes the sensor less sensitive to light and more sensible to noise.

Fill factor

The fill factor indicates the size of the light sensitive photodiode relative to the surface of the pixel. Because of the extra electronics required around each pixel the fill factor

must be balanced with the space required by this electronics. To overcome this limitation, often an array of microlenses is placed on top of the sensor. Higher fill factors are usually preferable, as they result in more light capture and less aliasing.

Sensor size

Digital SLR (Single-lens reflex) cameras make use of sensor with dimensions closer to the traditional size of a 35mm film frame, while video and point-and-shoot cameras traditionally use small chip areas. Common sensors used in reflex cameras are the full frame sensors (36 x 24 mm) and APS-C sensors (23.6 x 15.7 mm for Nikon cameras). If the sensor does not fill the 35mm full frame format, a multiplier effect has to be considered on the lens focal length. For example, an APS-C sensor gives a multiplier factor of 1.6. A larger sensor size is preferable, since each cell can be more photo-sensitive but are more expensive to produce.

Analog gain

Usually the signal is boosted by an amplifier before the transfer to the ADC. In video cameras, the gain was controlled by an automatic gain control (AGC) logic, which would adjust these values to obtain a good overall exposure. Digital still cameras allow additional control over this gain through the ISO sensitivity setting, expressed in ISO standard units (ISO 100, ISO 200, etc ...). The ISO setting is the third degrees of freedom after exposure and shutter control when the camera is used in manual control mode. In theory, a higher gain allows the camera to perform better under low light conditions but higher ISO settings amplify the sensor thermal noise.

Sensor noise

The noise added to the images comes from various sources: fixed pattern noise, dark current noise, shot noise, amplifier noise and quantization noise (Healey & Kondepudy, 1994; Tsing *et al.* 2001). The final amount of noise in the image depends on all of these quantities, as well as the incoming light, the exposure time, and the sensor gain.

ADC resolution

The final step in the analog processing chain is the analog to digital conversion (ADC). The two quantities of interest are the resolution of this process (how many bits it yields) and its noise level (how many of these bits are useful in practice). The radiometric resolution determines how finely a system can represent or distinguish differences of intensity, so with higher values better fine differences of intensity or reflectivity can be represented. In practice, the effective radiometric resolution is typically limited by the noise level, rather than by the number of bits of representation

Digital post-processing

After the conversion of the light falling on the imaging sensor into digital bits, a series of digital signal processing (DSP) operations has been performed to enhance the image before compression and storage. These include colour filter array (CFA) demosaicing, white balance setting, and gamma correction to increase the perceived dynamic range of the signal.

1.1.2. Colour filter arrays and demosaicing

The sensible elements of an image sensor are monochrome photosensors. In order to acquire a RGB colour image it is necessary to place in front of the sensor a Colour Filter Array (CFA) which decompose the incident light into the three channels Red, Green and Blue. The Bayer CFA is the most common filter used in digital cameras (fig. 1.2-a); the Bayer pattern (Bayer, 1976) arrange the RGB colour filter on a square grid of photosensors, and is composed by a 50% of green pixels, 25% of red pixel and 25% of blue pixels (RGBG or RGGB pattern).

The principle of operation of the filter is based on the fact that only a specific band of wavelengths of light let through the sensor, which corresponds to a well-determined colour. Thus only one colour is measured for each pixel of the sensor and the camera has to estimate the two missing colours at each pixel by performing an operation that is called demosaicing (fig 1.2-b). The estimate of the missing colours in each pixel takes

place on the basis of neighbouring pixels, using an interpolation that can be of type nearest neighbour, linear or bilinear.

The demosaicing reduces the contrast of the resulting image and causes a misregistration of the colour channel (colour aliasing) which must be taken into account when using the images for metric purpose.

Colour aliasing is caused by spatial phase differences among the colour channels (Guttosh 2002) because the photosensors relative to each RGB component are placed close. To reduce colour aliasing (at the expense of image sharpness), an optical filter (Anti-Aliasing filter) or an optical low-pass filter (OLPF), is usually employed. Two of these filters (one in the horizontal direction, the other in the vertical) are typically placed in the optical path.

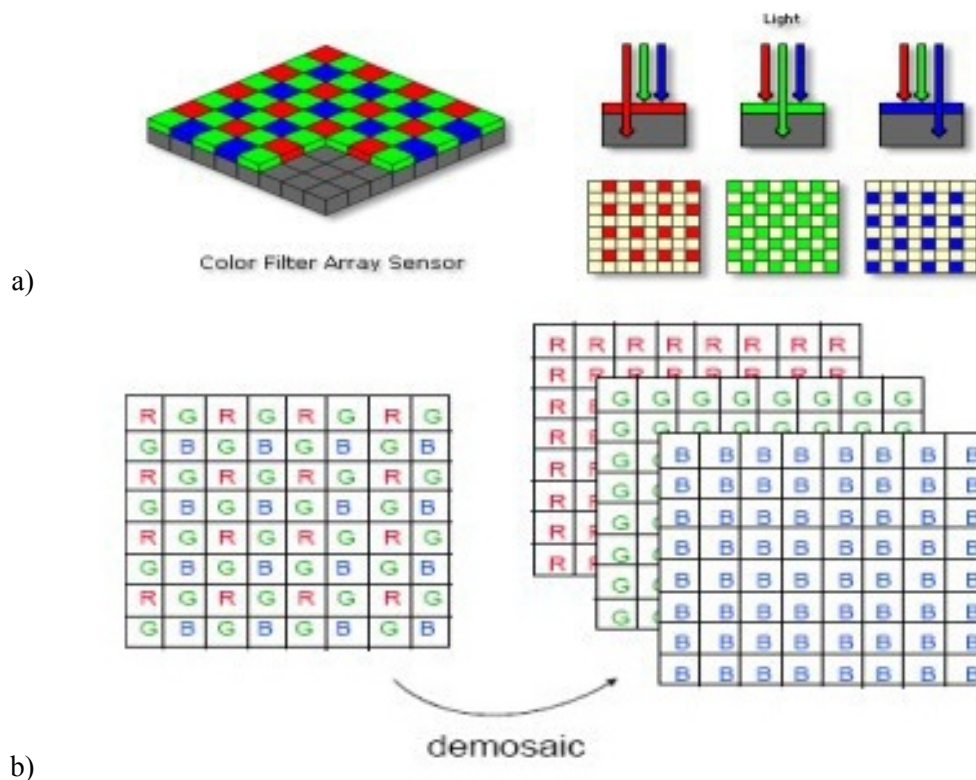


Figure 1.2. Bayer pattern (a) and demosaicing procedure (b).

Another technology is based on a different type of sensor, the Foveon (fig. 1.3), that is a CMOS sensor in which a stack of photodiodes is present at each pixel location. Each photodiodes in the stack filters different wavelength of light, so the final image results

from the combination of this kind of filtering, which does not require a demosaicing stage.

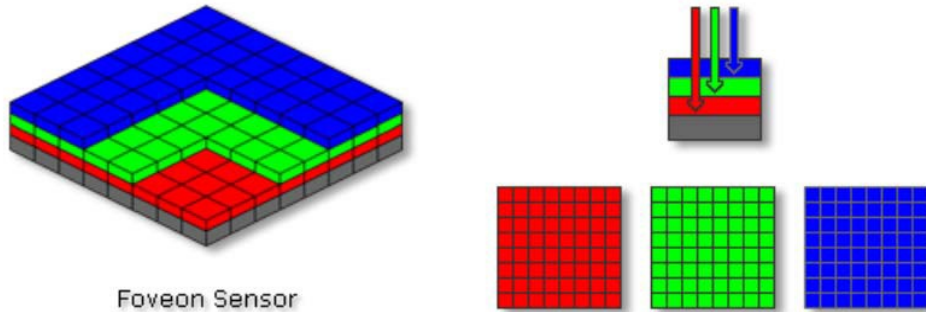


Figure 1.3. The Foveon sensor.

1.2. Geometry of image formation and multi-view Geometry

This section provides a brief review which covers the theory of camera models, multi-view geometry, and image similarity metrics.

1.2.1. Pinhole Camera Model

The pin-hole or perspective camera model (shown in Figure 1.4) is commonly used to explain the geometry of image formation. Given a fixed centre of projection C (the pin-hole or the *camera centre*) and an image plane, every 3D point $M(X, Y, Z)$ other than the *camera centre* itself, maps to a 2D point $m(x, y)$ on the image plane, which is assumed to be at a distance f from C .

They are related as follows:

$$x = f \cdot \frac{X}{Z}, \quad y = f \cdot \frac{Y}{Z}$$

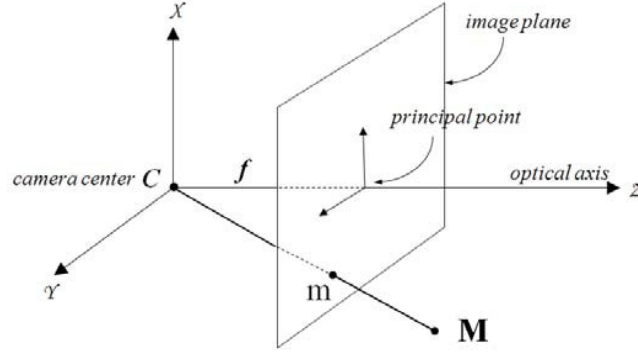


Figure 1.4. Pinhole camera model.

Using homogeneous coordinates for 2D and 3D points, this can be written in matrix form:

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (1.1)$$

The line through C, perpendicular to the image plane is called optical axis and it meets the image plane at the principal point. Often, 3D points are represented in a different world coordinate system. The coordinates of the 3D point in the camera coordinate system M_c , can be obtained from the world coordinates M_w , as follows:

$$\mathbf{M}_c = \mathbf{R}\mathbf{M}_w + \mathbf{t}$$

Here \mathbf{R} represents a 3×3 rotation matrix and \mathbf{t} represents a translation vector for 3D points. This can be written in matrix form.

$$\mathbf{M}_c = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix} \mathbf{M}_w \quad i.e. \quad \mathbf{M}_c = \mathbf{T}_w \mathbf{M}_w \quad (1.2)$$

The image coordinate system is typically not centred at the principal point and the scaling along each image axes can vary. So the coordinates of the 2D image point undergoes a similarity transformation, represented by \mathbf{T}_c . Substituting these into the perspective projection equation, we obtain:

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \mathbf{T}_c \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \mathbf{T}_w \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (1.3)$$

Or simply,

$$m = \mathbf{P}M$$

where \mathbf{P} is a 3×4 non-singular matrix called the *camera projection matrix*. This matrix \mathbf{P} can be decomposed as shown in Eq. 1.4, where \mathbf{K} is the matrix containing the *camera intrinsics parameters*, while \mathbf{R} and \mathbf{t} together represents the *camera extrinsics parameters*, namely the relative pose of the camera in the 3D world.

$$\mathbf{P} = \mathbf{K}[\mathbf{R}|\mathbf{t}] \quad \text{where} \quad \mathbf{K} = \begin{bmatrix} f_x & s & p_x \\ 0 & f_y & p_y \\ 0 & 0 & 1 \end{bmatrix} \quad (1.4)$$

The intrinsics \mathbf{K} can be parameterized by f_x, f_y, s, p_x and p_y (Eq. 1.4), where f_x and f_y are the focal length f measured in pixels in the x and y directions respectively, s is the skew and (p_x, p_y) is the principal point in the image.

Thus, in the general case, a Euclidean perspective camera can modelled in matrix form with five intrinsic and six extrinsic parameters (three for the rotation and three for the translation) which defines the transformation from the world coordinate system, to the coordinate system in the image.

Real cameras deviate from the pin-hole model due to various optical effects, amongst which, the most pronounced is the effect of radial distortion that affect above all wide-angle lenses, which manifests itself as a visible curvature in the projection of straight lines. Unless this distortion is taken into account, it becomes impossible to create highly accurate reconstructions. For example, panoramas created through image stitching constructed without taking radial distortion into account will often show evidence of blurring due to the misregistration of corresponding features before blending.

Radial distortion is often corrected by warping the image with a non-linear transformation. Thus, the undistorted image coordinates (\tilde{x}, \tilde{y}) can be obtained from the distorted image coordinates (x, y) as follows:

$$\tilde{x} = x_c + (x - x_c)L(r)$$

$$\tilde{y} = y_c + (y - y_c)L(r)$$

$$L(r) = (1 + k_1 r + k_2 r^2 + \dots)$$

Here k_1, k_2 etc. are the coefficients of radial distortion, (x_c, y_c) is the centre of radial distortion in the image and r is the radial distance from the centre of distortion.

1.2.2. Epipolar Geometry

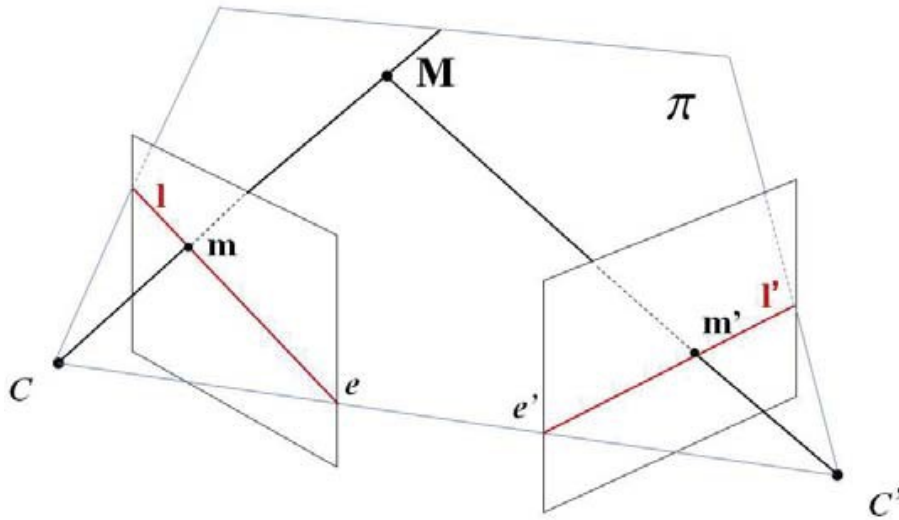


Figure 1.5. For every pixel m , the corresponding pixel in the second image m' must lie somewhere along a line l' . This property is referred to as the *epipolar constraint*. See the text for details.

The epipolar geometry captures the geometric relation between two images of the same scene. When a 3D point M projects to pixels m and m' on the two images m and m' are said to be in correspondence. For every point in the first image, the corresponding point in the second image is constrained to lie along a specific line called the *epipolar line*. Every plane such as π that contains the *baseline* (the line joining the two camera centres), must intersect the two image planes in corresponding epipolar lines, such as l and l' , respectively. All epipolar lines within an image intersect at a special point called the *epipole*. Algebraically,

$$m'^T \mathbf{F} m = 0 \quad (1.8)$$

where \mathbf{F} is called the *fundamental matrix* and has rank two. Points m and m' can be transferred to the corresponding epipolar lines in the other image, using the following relations.

$$\mathbf{l} = \mathbf{F}^T m' \quad \mathbf{l}' = \mathbf{F} m$$

The epipoles are also the left and right null-vector of the fundamental matrix:

$$\mathbf{F} \mathbf{e} = 0 \quad \mathbf{F}^T \mathbf{e}' = 0 \quad (1.9)$$

Since F is a 3×3 matrix unique up to scale, it can be linearly computed from 8 pair of corresponding points in the two views using Equation 1.8, which is often called the epipolar constraint. This is known as the *8-point algorithm*. However, when the rank constraint is enforced, F can be computed from 7 pairs of correspondences using the non-linear *7-point algorithm*. Refer to (Hartley & Zisserman 2005) for the details. Any pair of cameras denoted by camera matrices P and P' , results in a unique fundamental matrix. Given a fundamental matrix F , the camera pairs are determined up to a projective ambiguity.

1.2.3. Projective Reconstruction and Self-Calibration

Without special knowledge about the contents of the scene, it is impossible to recover the position, orientation, and scale of a 3D scene reconstructed from images. When the camera intrinsics are unknown, there is a higher degree of ambiguity in the reconstruction – it can only be determined up to a projective transformation of 3 space. By recovering the whole projective structure starting from only point correspondences in multiple views, one is able to compute a *projective reconstruction* of the cameras and the scene. Note that this can be done without any knowledge of the camera intrinsics, and makes it possible to reconstruct 3D scenes from uncalibrated sequences.

This projective cameras and scene differs from the actual cameras and scene (often referred to as a Euclidean or *metric reconstruction*) by a projective transformation. There exists classical techniques to transform a projective reconstruction to a metric one by computing this unknown homography – this is called *auto-calibration* or *self-calibration* (Fraser 1997). Please refer to (Hartley & Zisserman 2005) and (Pollefeys *et al.* 2004) for more details.

1.3. Multi-view 3D Reconstruction

The techniques that use images to reconstruct the 3D shape of an object can be divided into two categories: *active* and *passive* techniques. While in *active* methods some kind of lighting is employed to recover 3D information, *passive* methods recover all the information only from the object's texture.

Active techniques such as laser scanning (Levoy *et al.* 2000), structured or unstructured projected light methods (Zhang *et al.* 2003), and active depth from defocus (Ghita *et al.* 2005) typically require expensive hardware or custom equipment. These techniques produce good results in controlled scenes, but there are some issues when using them in real environment.

The passive techniques are based on different shape-from-x approaches: shape from texture (Kender 1981), shape from silhouettes (Hirano *et al.* 2009), shape from shading (Horn & Brooks 1989), shape from focus (Nayar & Nakagawa 1990), shape from stereo, etc. These are signals that the human vision uses to perceive the 3D shape of an object. When some information about the scene is known, it is possible to recover coarse shape and fairly inaccurate 3D information from a single image relative to an unknown scene (Criminisi *et al.* 2000; Hoiem *et al.* 2005; Saxena *et al.* 2008).

It is possible to recover the 3D structure using geometric techniques such as multi-view triangulation by computing dense pixel to pixel correspondences between multiple calibrated images. The 3D reconstruction problem is quite challenging due to image noise, specular surfaces, bad illumination, moving objects, texture-less areas and occlusion. Often the reconstruction conduct to a computational expensive optimization problem, involving different mathematical formulations based on different criteria (Faugeras & Keriven 1998; Boykov & Kolmogorov 2003 and 2004).

These methods can be classified into *local* and *global* methods. While global methods usually ensure higher accuracy their computational complexity is usually much higher than the local methods (Scharstein & Szeliski 2002), and cannot be used for real time applications.

1.3.1. Camera calibration

In traditional camera calibration procedures, images of a calibration target represented by an object with a known geometry, are first acquired. Then, the correspondences between 3D points on the target and their imaged pixels found in the images are recovered. The next step is represented by the solving of the *camera-resectioning* problem which involves the estimation of the intrinsic and extrinsic parameters of the camera by minimizing the reprojection error of the 3D points on the calibration object.

The camera calibration technique proposed by Tsai (Tsai 1992) requires a three dimensional calibration object with known 3D coordinates, such as two or three orthogonal planes. This type of object was difficult to build complicating the overall.

A more flexible method has been proposed by (Zhang 2000) and implemented in a toolbox by (Bouguet 2012). It is based on a planar calibration grid that can be freely moved jointly with the camera. The calibration object is represented by a simple checkerboard pattern that can be fixed on a planar board.

This method can be very tedious when accurate calibration of for large multi-camera systems have to be performed; in fact with the checkerboard can only be seen by a small group of cameras at one time, so only a small number of cameras can be calibrated at the same time. Then it is necessary to merge the results for all calibration session in order to calibrate the full camera network.

A new method for multi-camera calibration, based on the use of a single-point calibration object represented by a bright led that is moved around the scene, has been proposed by (Svoboda *et al* 2005). This led is viewed by all the cameras and can be used to calibrate a large volume.

Although all these methods can produce accurate results, they require an offline pre-calibration stage that is often impractical in some application. Furthermore in the last decade significant progress has been made in automatic feature detection and feature matching across images. Robust structure from motion methods, that allow the recovery of 3D structure from uncalibrated image sequences, have been developed.

Such structure from motion algorithms can be used with videos (Pollefeys *et al.* 2004), or with large unordered image collections (Brown & Lowe 2005; Snavely *et al.* 2006).

1.3.2. Structure from motion

Most of the structure from motion techniques used for sequence of unknowns cameras start by estimating the fundamental matrix if the correspondence is performed in two views or the trifocal tensor in the case of three view correspondences. The trifocal tensor and the fundamental matrix present the same role in the three-view and two views case respectively (Hartley & Zisserman 2005).

Most of the approaches used for large-scale structure from motion computes in an incremental way the reconstruction of the cameras and the scene, though in the state of the art are present various approaches that perform the reconstruction simultaneously (Sturm & Triggs 1996; Triggs 1996).

The SFM process developed by (Pollefeys *et al.* 2004) start from an initial reconstruction recovered from two views, then other cameras are added making possible to reconstruct the scene incrementally. A projective bundle adjustment (Triggs *et al.* 1999) refine all the camera parameters and the 3D point previously computed, by minimizing the reprojection error.

The obtained reconstruction is determined up to an arbitrary projective transformation. So it is necessary a method that can lead to a metric reconstruction determined up to an arbitrary Euclidean transformation and a scale factor. This can be done though the imposition of some constraints on the intrinsic camera parameters (self-calibration) followed by a *Euclidean bundle adjustment* to determine the optimal camera parameters.

1.3.3. Multi view stereo

In multi-view stereo algorithms (Seitz *et al.* 2006) the object's colour or texture is used to compute dense correspondence between pixels in different calibrated views, then the depth of a 3D point in the scene can be recovered by triangulating corresponding pixels. Image matching algorithms often fail to find robust correspondences because of ambiguities due to texture-less regions, problems due to occlusions or non-Lambertian properties of the surface.

Stereo matching algorithms represent the object's shape using a disparity maps, which represent the distance d (disparity) between a pixel $p_1(i, j)$ and its correspondent $p_2(i+d, j)$ along an epipolar line. This is a 2D search problem because the images are previously rectified in order to ensure that corresponding pixels lie on the same scanline (epipolar line).

In the multi-view case, the correspondence between matched pixels is represented by a *depth-map* that is the depth from a particular viewpoint. The disparity map computation can be viewed as a pixel labelling problem, which can be solved by

energy minimization methods such as graph cuts (Boykov & Kolmogorov 2004) and loopy belief propagation (Yedidia *et al.* 2000).

Similarity Measures

The (*dense*) *correspondence problem* in computer vision involves finding, for every pixel in one image, the corresponding pixel in the other image. As individual pixel values are subject to the presence of noise and alone are not distinctive enough the similarity is often computed using a support window around a pixel. Typically is a $n \times n$ square window centred in the current pixel, but some methods make use of adaptive windows (Tombari *et al.* 2008; Kanade & Okutomi 1994). Different windows-based similarity functions based on difference and correlation measures are used for this task. One of the most common pixel-based matching method, is the Sum of Absolute Differences (SAD) (Kanade 1994) given by the expression:

$$\sum_{(u,v) \in W} |I_1(x_1 + u, y_1 + v) - I_2(x_2 + u, y_2 + v)| \quad (1.11)$$

Another common metric is the Sum of Squared Differences (SSD) (Hannah 1974) given by the expression:

$$\sum_{(u,v) \in W} (I_1(x_1 + u, y_1 + v) - I_2(x_2 + u, y_2 + v))^2 \quad (1.12)$$

The Normalized Cross Correlation (NCC) (Hannah 1974; Bolles *et al.* 1993; Evangelidis & Psarakis 2008) is given by the expression:

$$\frac{\sum_{(u,v) \in W} I_1(x_1 + u, y_1 + v) \cdot I_2(x_2 + u, y_2 + v)}{\sqrt{\sum_{(u,v) \in W} I_1^2(x_1 + u, y_1 + v) \cdot \sum_{(u,v) \in W} I_2^2(x_2 + u, y_2 + v)}} \quad (1.13)$$

The Zero-mean Normalized Cross Correlation (ZNCC) (Faugeras *et al.* 1993) is given by the expression:

$$\frac{\sum_{(u,v) \in W} (I_1(x_1 + u, y_1 + v) - \bar{I}_1) \cdot (I_2(x_2 + u, y_2 + v) - \bar{I}_2)}{\sqrt{\sum_{(u,v) \in W} (I_1(x_1 + u, y_1 + v) - \bar{I}_1)^2 \cdot \sum_{(u,v) \in W} (I_2(x_2 + u, y_2 + v) - \bar{I}_2)^2}} \quad (1.14)$$

SAD and SSD produce a value of zero for identical support windows, but are not normalized because the similarity measure depends on the window extension.

Although it is the most expensive in terms of computational time, contrary to SAD, SSD and NCC, the ZNCC tolerates linear intensity changes providing better robustness and should be used when brightness differences are present in the images. When a relevant noise is present in the images, similar *texture-less* areas provide high similarity scores using SAD and SSD, when the ZNCC instead produce a low similarity score because it tries to correlate two random signals. For more details about similarity measures in the context of stereo matching, please refer to (Hirschmuller & Scharstein 2007).

1.4. Experimentation of a multi-view stereo algorithm in real case studies

A multi-view stereo technique makes use of a simple camera to reconstruct the desired object, and can be applied at different scales by simply changing the optical setup. In this section some examples of 3D reconstruction performed on different objects (monuments, buildings, statues) will be presented. A procedure based on open source solution consisting in Bundler (Snavely *et al.* 2007) for structure from motion and PMVS2 (Furukawa & Ponce 2007) for dense reconstruction has been adopted.

1.4.1. Large scale scenes

Using a careful planning of the photographic session, and through the analysis of lighting conditions and environmental factors it is possible to reconstruct very complex scenes such as, for example, entire towns.

In fig. 1.6 the 3D reconstruction of the historical centre of Cosenza (Italy) is shown. About 1000 pictures have been acquired for a total number of about 20 million points belonging to the facades of the buildings across the main street and around the three main squares.

The example in fig. 1.7 represents the survey conducted on the Certosa of Serra San Bruno (Calabria - Italy). From a dataset containing about 150 pictures, it has been possible to reconstruct about 3 million points which are sufficient for the analysis of the degradation. In these applications the most important issues are related to occlusions

and inaccessible areas (particularly in old historical centres), changes of lighting conditions and moving objects (pedestrians, cars, etc...).

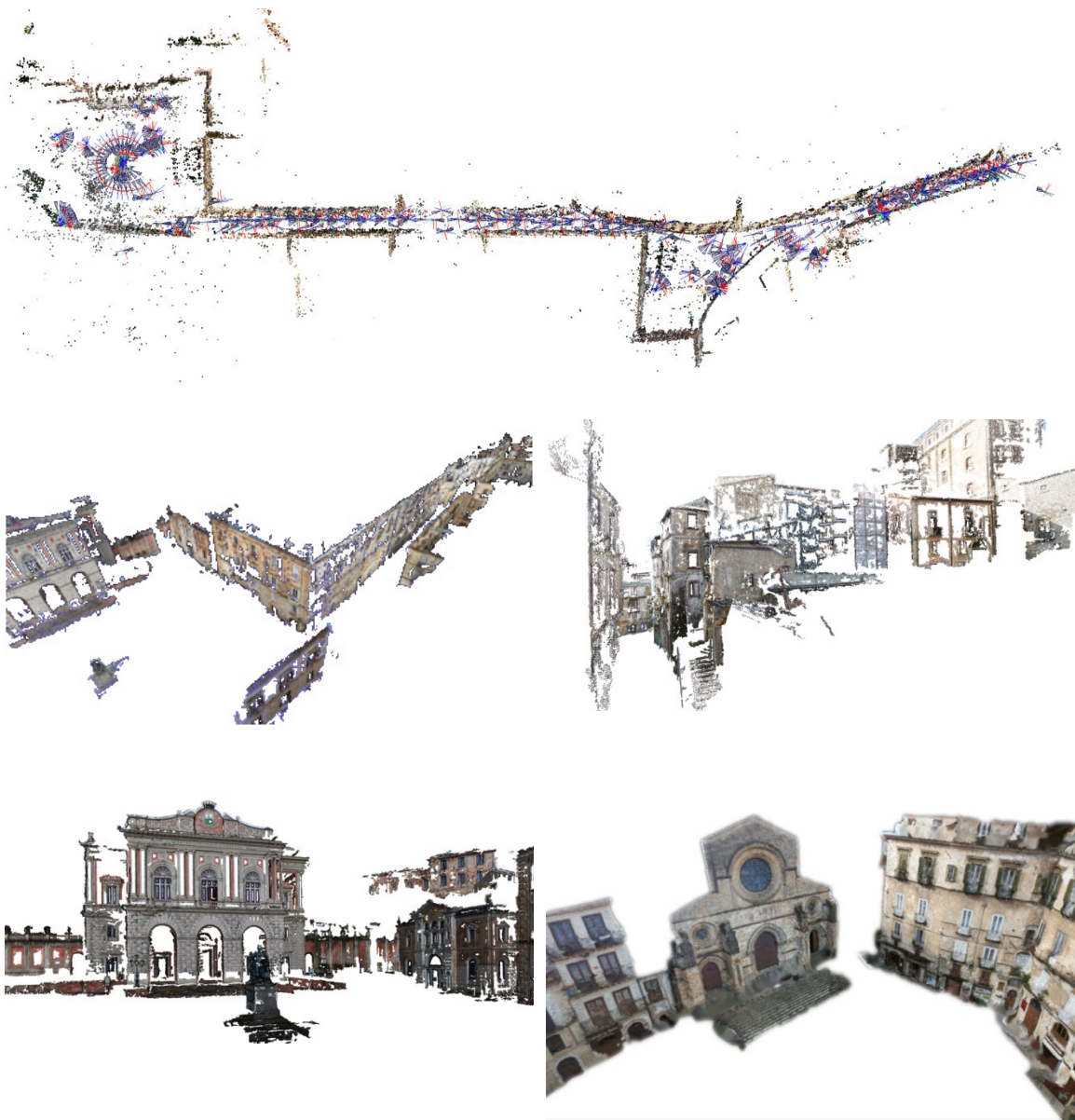


Figure 1.6. Survey of the historical centre of Cosenza (Italy): aerial view of the reconstructed area including camera positions (top) and 3D pointy clouds related to relevant points of interest.



Figure 1.7. 3D point cloud of the Certosa of Serra San Bruno (Italy), destroyed in 1783 by a terrible earthquake.

1.4.2. Monuments and statues

In fig. 1.8 there is an example of the reconstruction conducted in the B. Telesio statue in Cosenza (Italy). From a dataset of about 300 pictures, acquired using different lenses (wide angle and telephoto lenses), has been possible to reconstruct about 1.5 million points and about 3 million triangles. From these data a textured model usable for interactive visualization has been created.

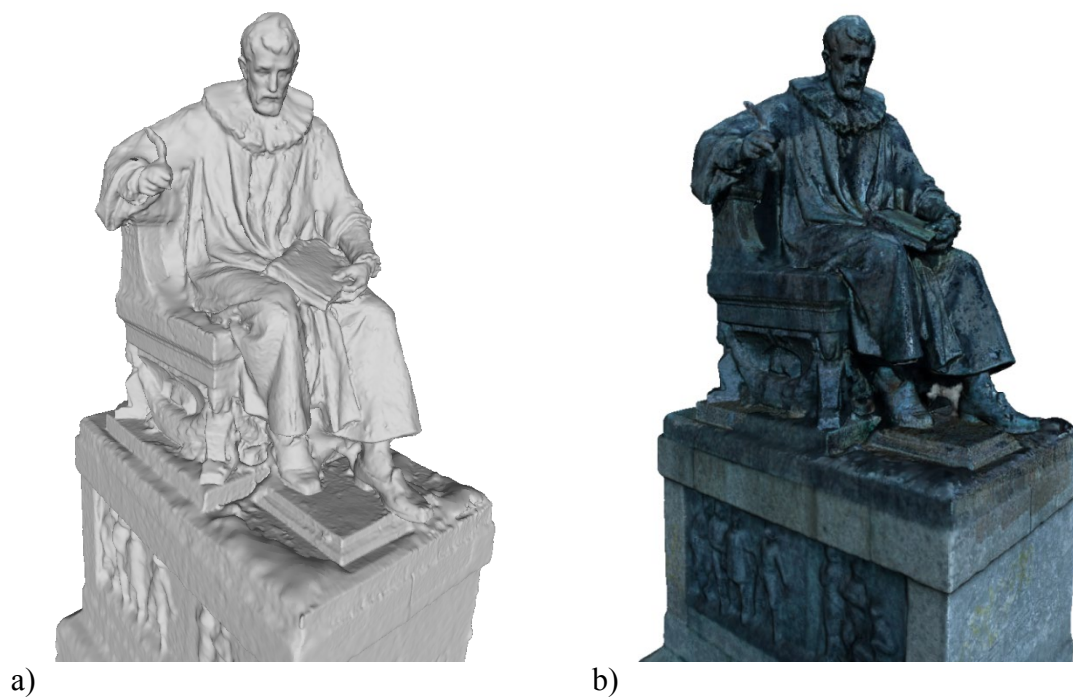


Figure 1.8. Triangulated surface (a) and textured 3D model (b) of the B. Telesio Statue located in Cosenza (Italy).

The example in fig. 1.9 represent a wooden statue of the Virgin Mary conserved in Marsico Nuovo (PZ - Italy) reconstructed from 130 images acquired inside a church in very bad lighting conditions. In order to ensure a high quality pictures, some artificial lights and diffusive panels has been used. The reconstructed model consists in 1.3 million points and about 2.5 million triangles.



Figure 1.9. Triangulated surface (a) and textured 3D model (b) of a wooden statue of Virgin Mary.

1.4.3. Small archaeological finds

Using a proper optical setup (controlled lighting, proper lenses, etc...) it is possible to reconstruct small objects, with dimension starting from 10-15 cm. The 3D reconstruction of very small objects (dimensions less than 5 cm) involves the use of an higher magnification ratio which causes the problems that will described in chapter 2.

Fig. 1.10 shows the reconstruction of a bronze archaeological find at the National Museum of Alta Val D'Agri (PZ - Italy), characterized by a very dark texture and a complex surface. About 2 million points have been measured from 50 images.

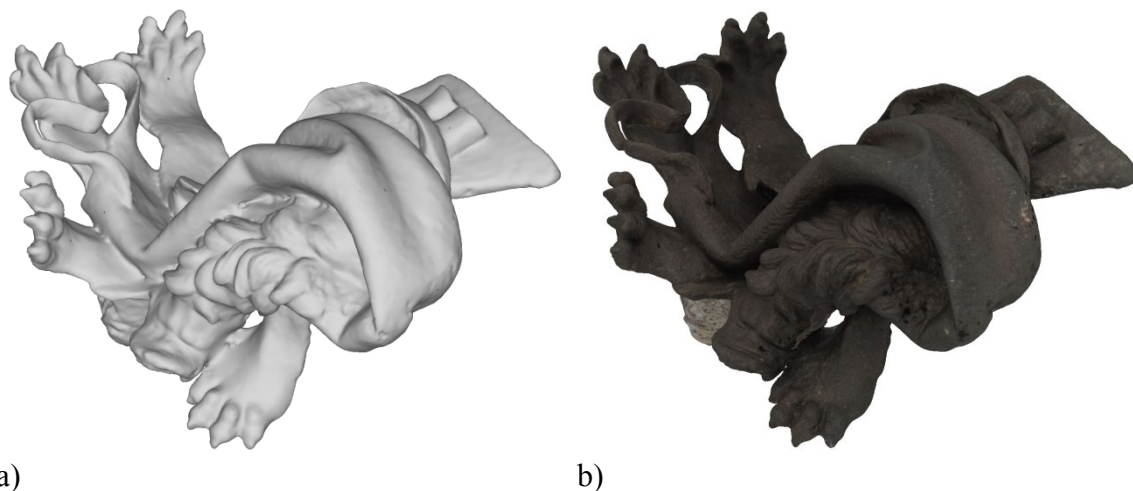


Figure 1.10. Triangulated surface (a) and textured 3D model (b) of a bronze archaeological find.

In fig. 1.11 the reconstruction of a bas relief of the XIV century made of alabaster is presented. Despite the challenging surface characteristics (no texture, reflections, etc...) it has been possible to reconstruct about 3 million points for a frame of 50 x 50 cm from 36 pictures.

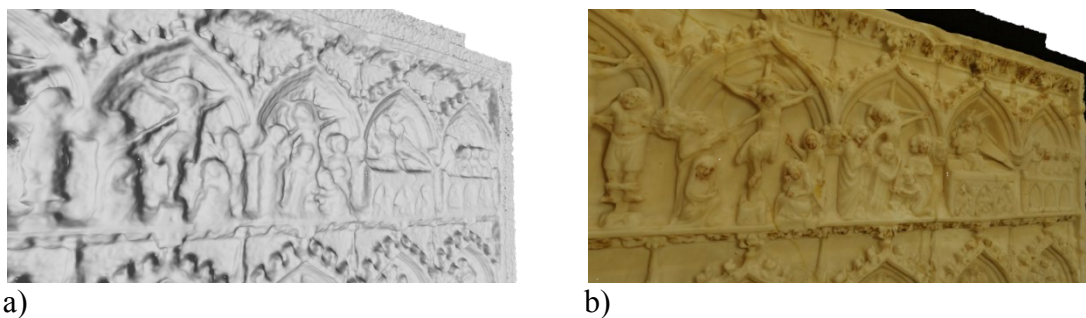


Figure 1.11. Triangulated surface (a) and textured 3D model (b) of a bas-relief with dimensions of 50 x 50 cm.

Multi-view 3D reconstruction of small sized objects

A new methodology for the 3D reconstruction of small sized objects based on a multi-view passive stereo technique applied on a sequence of macro images is described. In order to overcome the problems related to the use of macro lenses, such as the very small depth of field and the loss of sharpness due to diffraction, each image of the sequence is obtained by merging a stack of images acquired at different focus planes by the means of an image fusion algorithm. The results related to objects with dimensions ranging from few millimetres to few centimetres are shown.

2.1. Introduction

In the field of Cultural Heritage there are many applications of the 3D reconstruction techniques for both diagnostic and fruition purposes. Many interesting case studies are related to small sized objects such as jewellery, coins and miniatures, which in most cases have been treated just to allow their virtual presentation.

Recently, there is also an increasing number of applications related to material analysis and non-destructive investigations. This approach is very important for the diagnostic studies carried out for the preliminary restoration stages. In the last few decades the diagnostic analysis of decay processes on small sized objects is increasingly important in the restoration and maintenance of ancient archaeological artefacts. It requires the examination of the materials used and a careful characterization of decay products, in order to evaluate degenerative processes.

In a previous work scanning electron microscopy (SEM) has been used for the analysis of morphological features and degradation forms present on the archaeological artefacts (Crisci *et al.* 2010). In this case the samples were analysed in conditions of low vacuum (10–5 mbar pressure) and were coated with a thin, highly conductive film (graphite) to obtain a good image. For this purpose the 3D reconstruction is useful for different

analytical steps: identification of degradation forms and determination of a qualitative distribution map, choice and mapping of sampling points. Unfortunately the use of standard 3D scanners, that are optimized for medium sized object (potteries, statues, etc.), does not allow the researchers to rely on precision that could be sufficient for an in-depth analysis of the object shape.

Commonly the analysis of small samples are carried out on an image captured using optical microscopes or SEMs. The images present a reference scale to perform the measurements of the characterizing elements, and in some cases a more rigorous approach is required. Since this process works on a two dimensional image, if the element to be measured is not placed on a plane that is orthogonal to the optical axis of the camera, the measurements will be affected by errors.

The adoption of a methodology based on interactive measurements performed on a three-dimensional model of the sample being studied will not only provide a more rigorous approach to the survey, but it would allow to build models suitable for being used in virtual museums or duplications.

In this chapter will be described the developed methodology for the 3D reconstruction of small objects based on a multi-view stereo technique applied on a sequence of images acquired by a digital camera equipped with a macro lens. Our contribution concerns the developing and the experimentation of a set of solutions designed to overcome some problems in the context of the 3D survey of small objects using passive optical techniques in the field of cultural heritage. The problems related to the use of macro lenses, such as the very small Depth of Field (hereinafter, DoF) and the loss of sharpness at high f-number due to diffraction, has been addressed by taking a series of pictures at different focus planes with an aperture below the limit of diffraction. These images are subsequently merged by an image fusion algorithm in order to obtain a picture with an extended DoF suitable for 3D reconstruction.

The validity of the reconstruction procedure has been evaluated on real objects. In particular the methodology has been tested on three case studies, whose complexity makes it possible to evaluate the performance of the used algorithms across a wide range of issues related to 3D reconstruction, such as surface properties and occlusions.

Our approach is directed only to the morphological study of small samples, for which the pushed magnification that can be obtained by SEM is not necessary.

The chapter is organized as follows. Section 2.2 presents a short review of the previous works concerning the 3D reconstruction of small sized objects, while in section 2.3 a short theoretical background will be presented. Section 2.4 describes the methodology used for the 3D reconstruction, while in section 2.5 and 2.6 the experimental setup and the accuracy evaluation procedure will be presented. The experimentation is described in Section 2.7 while in section 2.8 a comparison with a depth from focus technique is reported. In section 2.9 the results are discussed.

2.2. Related works

In the state of the art, there are many techniques successfully applied in the field of 3D reconstruction of small objects, such as laser scanners, TOF (Time Of Flight) cameras, X-Ray Tomography, microscopy. These techniques commonly involve the use of delicate, expensive and not much versatile equipment: the use of still images represent a valid and economic alternative, although an intervention of an operator is required for 3D model refinement (Remondino & El-Hakim 2006). In the field of cultural heritage, there are several methods and technologies for the 3D digitization, which can be classified according to object size and shape complexity (Pavlidis *et al.* 2007).

The images taken from two cameras mounted on an optical stereo microscope (Schreier *et al.* 2004) or on a scanning electron microscope (Cornille *et al.* 2003) can be used to extract 3D information. This approach does not require the use of a precision calibration target for camera parameters estimation. The microscope can be accurately calibrated with a common bundle-adjustment technique using a simple speckle pattern, and subsequently the 3D point cloud of the micro object can be reconstructed by means of stereo matching algorithms (see chapter 1 for more details). A recent approach (Atsushi *et al.* 2011) features the use of the SFS (Shape from Silhouette) technique on a sequence of images acquired through a digital microscope. This system is capable to reconstruct a polygonal model of a micro object with dimensions ranging from several hundred microns to several millimetres, with an accuracy of about the 0.5% of the bounding box. Alternatively, a single camera equipped with a macro lens mounted on a motorized

micrometric translational bar (Tiano *et al.* 2009) can be suitable for 3D measurements and dense reconstructions of small objects, by applying stereo matching algorithms on a triplet of images. The main target of this system is represented by surface analysis and high precision measurements of planar surfaces characterized by a low shape complexity and a thickness that does not exceed the macro lens Depth of Field (hereinafter, DoF). In the work of (Bitelli & Girardi 2010) the 3D reconstruction of small objects using photogrammetric techniques has been compared to the one obtained with a laser scanner, for objects from 5 to 10 centimetres. The issues related to the uses of macro lenses, such as the very small DoF and camera calibration anomalies, have been highlighted showing how this influences the quality of the final model.

2.3. Theoretical background

In this section a brief description of important concepts on which is based the experimentation is reported.

2.3.1. Macro photography

Macro-photography is a field of photography characterized by a very high level of detail. The magnification represents the ratio between the size of the photographed object on the imaging sensor, and its physical dimension. There are different techniques and equipment to take macro pictures. The most common tool for macro-photography is represented by macro lenses, designed for close-up work with magnification up to 1:1. The use of extension tubes or bellows allows to increase the distance between the lens and the sensor, resulting in a closer focus distance, a higher magnification and a darker image (given the same aperture). Another option is to mount a close-up lens in front of the objective, thus adding dioptres to the optical power of the lens, reducing the minimum focus distance and increasing the magnification. The quality is heavily reduced, because the auxiliary lens may cause chromatic aberration and loss of sharpness. For reproduction ratio up to 10-20x, it is possible to mount a reversed standard lens (commonly a 50 mm fixed lens) directly on the camera body or on the front of a normally mounted long focal lens (used like an extension tube, but preserving

all the automatisms such as exposure meter and focus control), with very high quality results.

2.3.2. Image fusion techniques for extended Depth of Field

Macro images, especially at high magnification, show a very small DoF. To overcome this limitation, it is possible to acquire a series of optical sections (slices) of the object continuously adjusting the focus plane. In the literature many image processing methods related to image fusion have been proposed (Pieper & Korpel 1983), with the aim to combine all the in-focus information of each slice into a composite focused image. These algorithms can be classified into:

- Pixel variance algorithm
- Edge detection algorithm
- Contrast measurements algorithms
- Multi-resolution based approaches.

Pixel based fusion is modelled on two approaches: point based and neighbourhood based. In the point based approach, pixels with the same coordinates are compared in all the images of a stack using a minimum or maximum rule to select the best in-focus pixel. In the second case, the algorithm is based on the assumption that the larger intensity variations occur in a region surrounding the pixel. Thus the pixel along the z axis with the biggest variance to its surrounding pixels will be selected for the fused image. This algorithm is computationally simple, but sensitive to brightness variations which are reduced by normalising the variance across the image first (Eltoukhy & Kavusi 2003; Aggarwal & Ahuja 2000).

Edge detection is based on the concept that focused areas have a high intensity and therefore a large luminance variation between close pixels, once the image has been appropriately filtered. A number of filters are available to do this, including the Sobel edge detection operators and the Laplacian filter.

In contrast measurements algorithms, as for edge detection methods, the contrast is measured as the difference in luminance values between neighbouring pixels. Various measures based on this approach have been proposed and have the advantage of being

very fast to be computed, since they involve only simple subtractions of luminance values.

The multi-resolution approach (Liu *et al.* 2001) is based on the assumption that focused areas contain many details and therefore high frequency components. In first instance, input images are decomposed using a multi-resolution technique, and then processed using the Discrete Wavelet Transformation, mathematically very similar to the Fast Fourier Transform (FFT) but with better properties (Forster *et al.* 2004; Valdecasas *et al.* 2001; Bradley & Bamdford 2004);

2.3.3. Depth from focus

Given a set of images taken by adjusting the focus plane, it is possible to recover depth information (Nayar & Nakagawa 1990; Pentand 1982). Assuming that the lens focal length and the subject distance is available, the lens distance can be solved using the lens law. Changes in the focus settings results in image blur, and by minimizing this blur, it is possible to estimate the depth through the thin lens equation (Jacobson *et al.* 2000). Several techniques have been proposed to quantify the image blur (Subbarao *et al.* 1993), based on variance of image intensity (Yeo *et al.* 1993), image gradients, squared Laplacian (Sperling 1970) and Tenengrad operator (Tenenbaum 1971). In the interesting work of (Schechner & Kiryati 2000), depth form focus (and defocus) techniques are compared versus stereo techniques, showing that the apparent differences reside in the size of the physical setups.

2.3.4. Multi-view photogrammetric reconstruction

To reconstruct all the details of a complex object overcoming occlusion, it is necessary to merge together different point clouds, acquired from different point of views. A solution is represented by the multi-view stereo techniques, through which it is possible to reconstruct a scene from a sequence of images, by solving the structure from motion (SFM) problem (see chapter 1 for more details). These techniques stars from the identification in each image of a sparse set of interest points (features), robustly detected by means of algorithms based on edge detection (Harris & Stephens 1988) or with

descriptor such as the SIFT operator (Lowe 2004), and the subsequently tracking of these features along the image sequence. Since the multi-view techniques make use of a common camera, it is possible to reconstruct scenes of any size by simply changing the optical setup. For this reason, it is possible to reconstruct small objects through the use of macro lenses. This practice, unfortunately, which presents some issues that will be discussed in the next section.

2.3.5. Problems in macro photogrammetry

Macro lenses are characterized by a very high level of detail, which would make them ideal for photogrammetric applications. Unfortunately, there are some photogrammetric problems related to the use of macro lenses in terms of camera calibration and DoF.

Working with macro lenses, whose behaviour can be assumed similar to long focal length lenses (Girardi 2011), leads to problems related to the analytical aspects of camera parameters retrieving (in particular for internal and external orientation parameters). As the focal length increase, the FOW (Field of View) get smaller, making the perspective projection model (with convergent optical rays) similar to an orthogonal projection model (with parallel rays) (Stamatopoulos *et al.* 2010). The Brown eight-parameter model presents some anomalies with long focal length lenses - in particular, the correlation between camera parameters and interior and exterior orientation parameters increase as focal length increase (Fraser & Al-Ajlouni 2006). For a macro lens, a 4-parameters subset of the Brown's model is sufficient to describe the optical properties and, in particular, the third order coefficient K_1 is sufficient to describe the radial distortion profile (Wiley & Wong 1995; Yanagi & Chikatsu 2010).

The biggest limitation of macro lenses is represented by the very small DoF, that may be within a half millimetre for a unitary magnification, and may reach values lesser than 100 micrometers with higher magnifications. This means that only a very small portion of the image appears sufficiently sharp and can be usable for 3D reconstruction. In particular there are problems to recover camera parameters and network orientation because image matching algorithms cannot work on blurred areas.

A way to extend the DoF is to use a higher f-number (smaller lens aperture) but, unfortunately, a smaller aperture leads to a longer exposure time, with the risk of image

blur due to camera shake or object movement. Furthermore, the optical resolution is highly reduced due to the diffraction effect of the light that has to pass through a small aperture.

For an ideal circular aperture, the 2D diffraction pattern is a disk surrounded by dark and light rings that is called Airy disk. Its width is used to define the theoretical maximum resolution for an optical system, and the diameter of the first dark ring is defined by (Ficher *et al.* 2008):

$$D = 2.44 \cdot \lambda \cdot N \quad (2.1)$$

in which λ represent the wavelength of the diffracted light and N the f-number being used. When the Airy disk's diameter becomes large in relation to the pixel size (or to the maximum tolerable *Circle of Confusion* - CoC), the effect of diffraction cannot be negligible (Fig. 2.1) or, alternatively, if the distance between two Airy disks become closer than their radius, they are no longer resolvable (Rayleigh criterion) (Born & Wolf 1999). A digital sensor utilizes a Bayer array to capture one primary colour at each pixel location, interpolating these colours to produce the final full colour image (see chapter 1). Considering the presence of an optical low-pass filter in front of the sensor used to suppress colour aliasing (see chapter 1), it is possible to assume that the Airy disk can have a diameter of about three pixels before the diffraction comes to limit the resolution. Nevertheless, diffraction starts to have a significant impact prior to reaching this diameter.

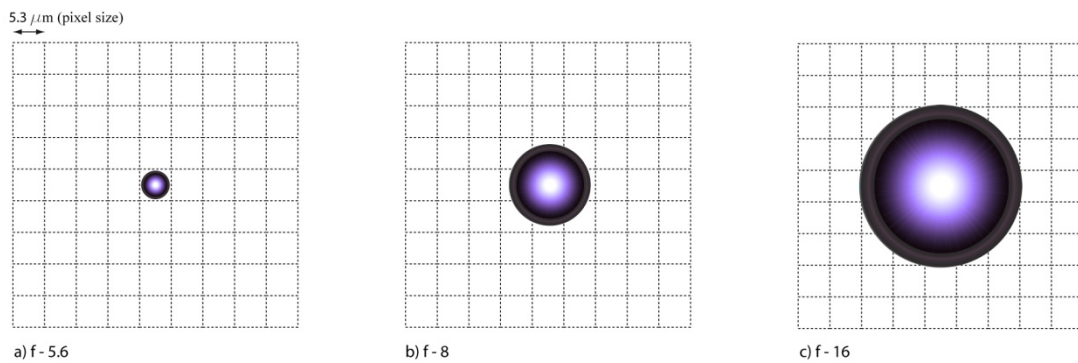


Figure 2.1. Size of the Airy disk at increasing f-numbers relative to the pixel dimension of a Nikon D5000 digital reflex camera.

2.4. The proposed methodology for small object reconstruction

The proposed methodology is made up of several steps with different degrees of automation, schematically represented in Fig. 2.2. The acquisition phase consists in capturing a set of images at different focus planes (image stack) for each camera pose. Subsequently each image stack has been processed by an image fusion algorithm to achieve a single image with an extended DoF (stacked image). Prior to start the automatic 3D reconstruction pipeline, each image has been manually pre-processed for background removal. This allows to speed up the reconstruction process, also increasing the quality of the final model.

In the next sections the proposed methodology will be presented in a general way, suggesting methods and software solutions to obtain high quality 3D models of small objects.

2.4.1. Image acquisition

The reproduction ratio had to be chosen in relation to object dimensions and desired resolution. The object should be placed on a stable surface, preferably on a rotating support, while the camera can be fixed on a solid tripod. It is recommended to place the subject in a soft box with diffusive light sources, in order to avoid strong shadows and reflections that could negatively affect the 3D reconstruction. The minimum working distance depends on the focal length and had to be chosen in order to provide enough space between the object and the front lens. To prevent vibration, the camera must be controlled by a PC and exposure delay mode must be used. The exposure time is related to the scene lighting and must be kept fixed during the whole acquisition process.

2.4.2. Determination of shooting parameters

The diffraction limit aperture is chosen considering that the Airy disk diameter evaluated through equation (2.1), has to be smaller than the CoC, that can be evaluated with good approximation by the “Zeiss formula” (Zeiss 1997), by which the CoC is estimated as $d/1500$, where d is the diagonal measure of the sensor. Regarding the

wavelength of light, it is possible to assume a value in the middle of the visible spectrum (about 550 nm).

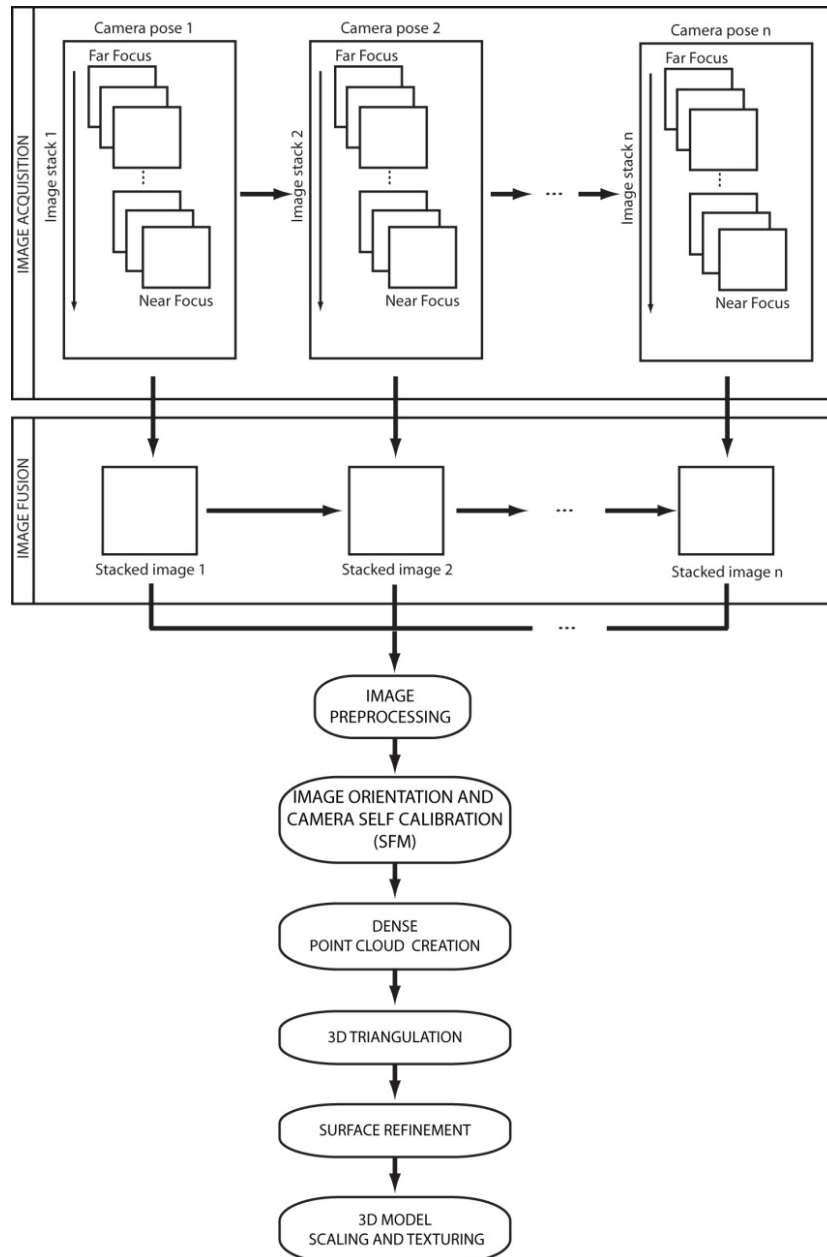


Figure 2.2. Pipeline of the implemented methodology for the acquisition and 3D reconstruction of small sized objects.

2.4.3. Image stack acquisition

Macro pictures are characterized by a very small DoF that can measure fractions of a millimetre at high magnification. The result is a picture in which only few areas are perfectly sharp. A plausible solution is to acquire a number of images at different focus planes, which are subsequently fused together. The number of pictures needed depends on the object size and DoF.

The DoF for an asymmetrical lens in terms of magnification can be expressed by:

$$DoF_{m,P} = \frac{2 \cdot f \cdot \left(\frac{1}{m} + \frac{1}{P}\right)}{\frac{f \cdot m}{N \cdot c} - \frac{N \cdot c}{f \cdot m}} \quad (2.2)$$

in which f is the focal length, m is the magnification, N the f-number, c the CoC and P the pupillary magnification. The latter is evaluated as the ratio between the exit pupil d and the entrance pupil D and is equal to unity for symmetrical lens. For macro lenses, the lens asymmetry leads to a pupillary magnification P different from unity, increasing the extension of the DoF with $P < 1$. The manufacturers do not provide this technical feature, so it should be measured experimentally.

The number of pictures needed to cover the distance between the near focus and far focus planes, can be choose reminding that it is necessary to have at least a 50% overlap between each slice, so is sufficient to shift the focus plane by a distance equal to half of the DoF.

Different commercial solutions are available to automate the image stack acquisition, such as a motorized linear slider to shift the camera or a camera control software to remotely drive the autofocus motor.

2.4.4. Image fusion and pre-processing

Each acquired image presents only a small portion that is in focus; therefore it is necessary to merge together the stack of pictures at each camera pose. The result is an image with an extended DoF (Fig. 2.3).

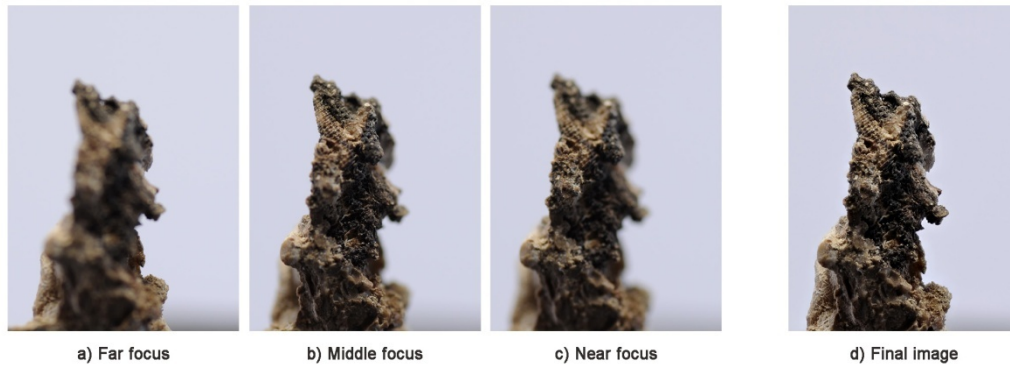


Figure 2.3. Example of three images acquired at different focus distances and the stacked image resulting from the fusion of 22 slices.

A simpler way consists in a single shot taken at the maximum f-number, in order to increase the DoF. The image shows a noticeable loss of sharpness due to diffraction (Fig. 2.4) and, moreover, the longer exposure time needed increase the risk of blur due to vibrations.

There are various commercial and free focus stacking software developed around different image fusion algorithms, such as Helicon Focus (HeliconSoft 2012), CombineZP (CombineZP 2012), Zerene Stacker (Zerene Systems 2012), that can be used to automate the image acquisition. In order to obtain high quality textures, it is necessary to pre-process the resulting images with white balance, levels, gamma, brightness and contrast adjustment.

2.4.5. Multi-view 3D reconstruction

To reconstruct the whole object with a multi-view stereo technique it is necessary to take a sequence of overlapping pictures. In the case of small objects, the better way is to maintain fixed the camera placing the object in a rotating table. It is important to point out that using a rotating table, it is necessary to create a mask to remove the background from the images in order to avoid matching problems due to features that remains fixed respect to the camera.

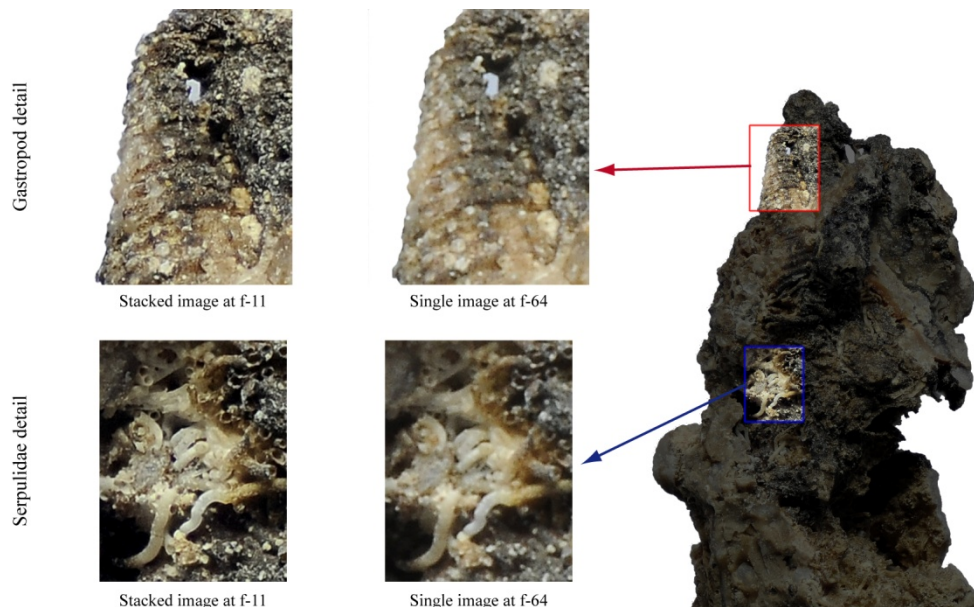


Figure 2.4. Effect of diffraction due to the use of a very small aperture. The image resulting from the fusion of 22 slices taken at different focus distance with the aperture set to f-11, is compared to the one taken at f-64, highlighting the loss of sharpness.

The number of camera poses varies according to the shape complexity and object dimension. The photographs must be taken at least every 10-15 degrees around the object and ideally a 50-75% overlap between images should be maintained during acquisition.

The image orientation is not an automatic process in commercial software packages and commonly is helped by using coded targets. Usually these targets are printed on paper and applied on the object surface. Clearly, the printing and positioning of a target suitable for small object reconstruction is rather difficult. In particular, the target can reach dimension of few millimetres, so the paper microstructure cause a bleeding effect that alters the circular shape of the printed dot that cannot be recognized. For this reasons it is necessary to identify the corresponding points from the natural texture using a feature extraction algorithm, such as SIFT (Lowe 2004), Harris (Harris & Stephens 1988), or SURF (Bay *et al.* 2008).

There are several commercial packages and open source research software that allow the creation of a dense point cloud from a sequence of images used with success for the 3D reconstruction of cultural heritage (Barazzetti *et al.* 2011). Moreover, are available several online 3D reconstruction services such as My3dscanner (My3DScanner 2012),

CMP sfm WEB service (CMP sfm 2012), ARC 3D (ARC 3D 2012), that provide a 3D point cloud ready to be edited.

2.4.6. Triangulation

The dense point cloud can be read by many CAD software in order to perform a surface triangulation to provide an optimal surface description.

Since the point cloud results from images that can be affected by noise, it may include some outlier points due to false matches or reconstructed areas that do not belong to the object. Prior to triangulate the point cloud, it is necessary to remove these noisy points by manual selection, using the appropriate tools included in all mesh processing software.

2.4.7. Surface refinement

Low texture areas, occluded regions, reflections and noise results in holes and artifacts in the reconstructed surface. Therefore, an additional manual refining step is necessary to correct all the surface defects. Nevertheless, some mesh processing software integrates efficient algorithms for automatic hole filling and noise reduction that can speed up this process.

2.4.8. Scaling and texturing

Since the 3D reconstruction is carried out up to a scale factor, it is necessary to scale the model to their real dimensions. An object with known dimensions included in the scene, such as scale bar or gauge blocks, can be used as a reference measure. In addition, it is possible to use two easily identifiable points with known distance.

The last step is represented by the texture mapping. The simplest method provides to map the colour of each 3D point, with good results only if a very dense point cloud is available. Since the viewpoints are known, the better way is to project the acquired images directly on the 3D surface.

2.5. Experimental setup

The experimental setup consists in a Nikon D5000 reflex camera equipped with a Tamron 90mm macro lens, designed for magnification up to 1:1 with a minimum working distance of 29 centimetres. In order to avoid vibrations that can cause blur, the setup is fixed on an optical table (Fig. 2.5) and the camera is remotely controlled by a PC.

The object to be acquired is placed on a white support plate fixed on a micrometric rotational stage, which provides a set of black dot targets, printed with high accuracy through serigraphy, to scale the final model in conjunction with the use of gauge blocks (fig 2.6). Furthermore, the white surface helps the image pre-processing, in particular the image masking and white balance correction. In order to ensure adequate lighting, a series of led lights and diffusive panels are placed around the object.

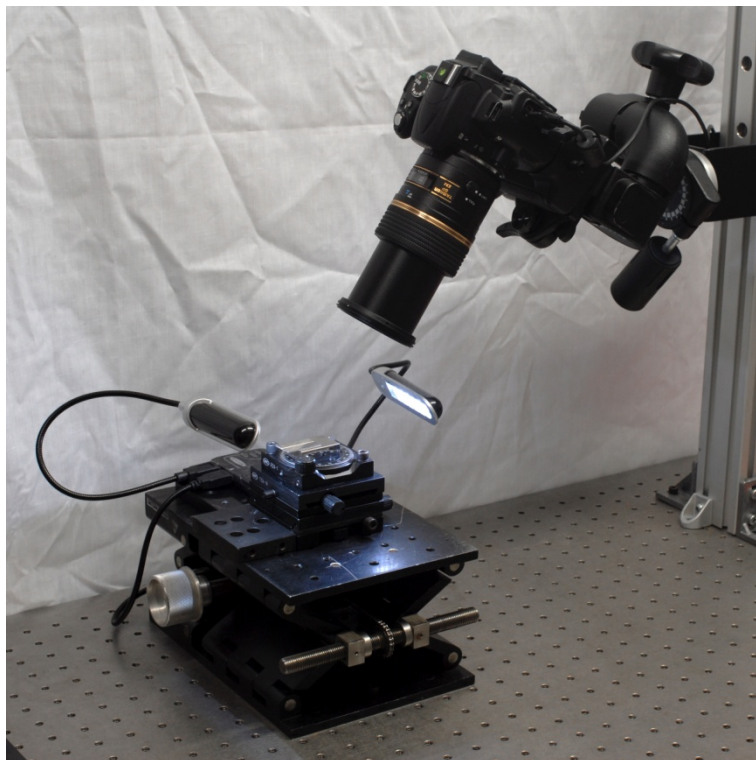


Figure 2.5. Experimental setup. A Nikon D5000 reflex camera equipped with a Tamron 90 mm macro lens and the rotational stage, fixed on an optical table. Some led lamps ensure proper lighting.

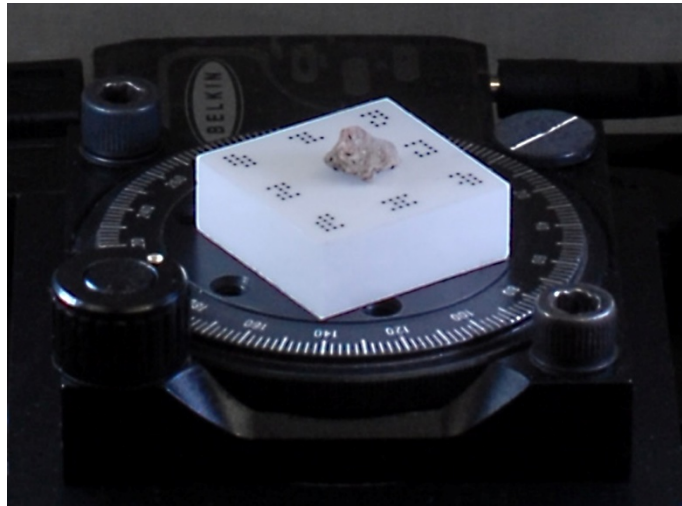


Figure 2.6. White support plate used to scale the final model and to help the image pre-processing (masking and white balance).

2.6. Accuracy evaluation

The effectiveness of the proposed methodology has been evaluated by performing a series of experimental tests on different sets of gauge blocks. The measurements have been conducted using certified gauge blocks of 1.20, 1.22, 1.25 and 1.30 mm, manufactured with dimensional and shape errors that are negligible compared to the ones we want to evaluate. The gauge blocks are placed on a rectified reference block and combined in pairs (table 2.1 and Fig. 2.7), in order to prepare three different specimens forming a step of 100 μm 50 μm and 20 μm . The three samples have been placed on the rotational stage and reconstructed through our approach. For each pair, a CAD model has been created with the nominal dimensions of each sample, and used as reference for accuracy evaluation.

| Gauge block (mm) | 1.20 | 1.22 | 1.25 | 1.30 |
|------------------|------|-----------------------|-----------------------|------------------------|
| 1.20 | - | 20 μm step | 50 μm step | 100 μm step |

Table 2.1. The three samples prepared arranging in pair three certified gauge blocks.



Figure 2.7. Two certified gauge blocks are placed on a rectified surface and mounted on the rotational stage in order to verify the accuracy.

The image orientation stage has been performed by using a feature extraction algorithm to extract a set of point of interest from the images that are subsequently matched along the image sequence. The matching process is sensitive to noise, reflection and lack of texture, so it is very difficult to extract and match a set of robust features. In fact in all the experiments, although it has been extracted a sufficient number of points, has not been possible to orient the images because the particular surface caused ambiguities.

In order to improve the matching, each image has been pre-processed by the means of the Wallis Filter (Wallis 1976). The enhanced images showed better contrast and more details (fig. 2.8), improving the image matching algorithm and allowing to find an higher number of robust features (table 2.2) making possible to orient the whole dataset.

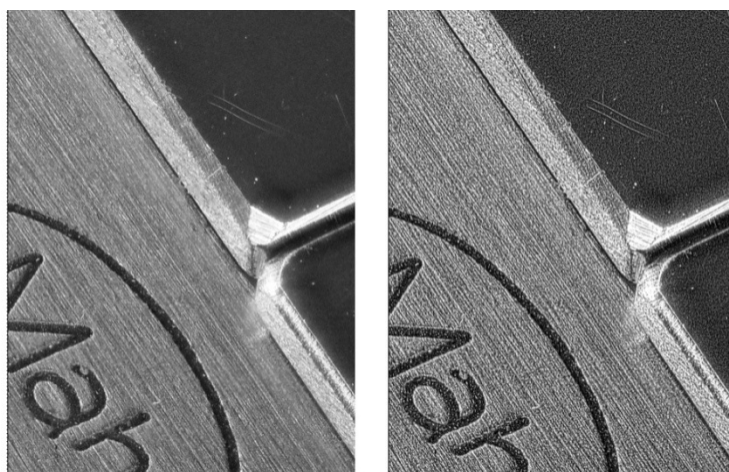


Figure 2.8. Sample image enhanced by using the Wallis filter (right) compared to the original uncorrected image (left).

| | Extracted SIFT features (mean values) | |
|--|---------------------------------------|-------------------------------|
| | <i>Original images</i> | <i>Wallis filtered images</i> |
| 20 μm specimen | 8165 | 14759 |
| 50 μm specimen | 7443 | 14581 |
| 100 μm specimen | 7868 | 14179 |

Table 2.2. mean values of the extracted SIFT features for the datasets relative to the tree specimens, reconstructed using the original images and the one enhanced by using the Wallis filter.

The reconstructed surface has been compared with the respective CAD model using the top surface of the 1.20 mm gauge block as reference plane (fig. 2.9). The distance between the reconstructed surface and the CAD model has been measured using the surface inspection tool provided with Rapidform (Rapidform 2012). The accuracy is estimated in terms of mean distance and standard deviation, and the measured values are summarized in Table 2.3. The results show that the system is capable of a maximum depth resolution of 0.020 mm, with an accuracy of about 10 microns, values comparable to the technical specifications of structured light scanners used for dental application, such as DentalScanner by Scansystems (Scansystems 2012).

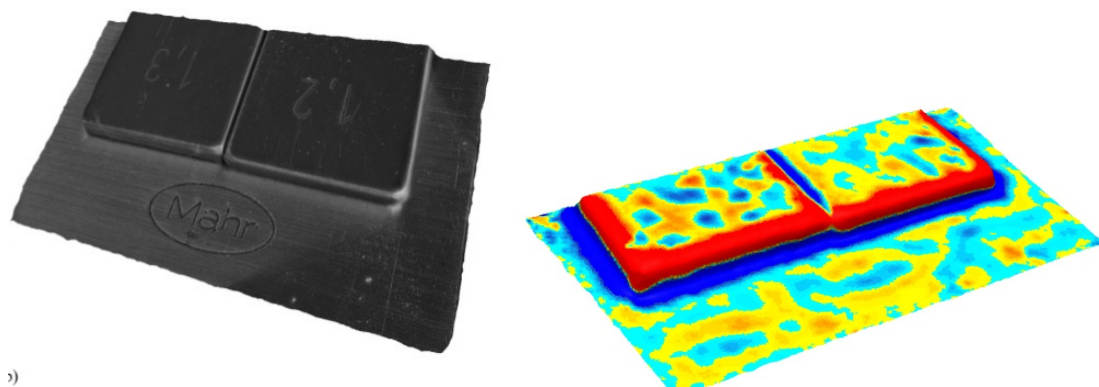


Figure 2.9. System resolution and accuracy evaluation: the reconstructed 3D model (a) has been compared to the relative CAD model (b).

| Parameter | 100 μm step | 50 μm step | 20 μm step |
|--------------------------------|--|---|---|
| <i>Mean (mm)</i> | 0.011 | 0.009 | 0.009 |
| <i>Standard deviation (mm)</i> | 0.009 | 0.006 | 0.008 |

Table 2.3. System accuracy evaluation. The experimental results show that the system is capable to resolve up to a 20 μm step with an accuracy of about 10 microns.

2.7. Experimentation

The experimentation has led to encouraging results in terms of depth resolution and accuracy. In order to validate the reconstruction procedure on real objects, the methodology has been tested on three case studies, whose complexity makes it possible to evaluate the performance of the used algorithms across a wide range of issues related to 3D reconstruction. The following description is referred to a specific experimental setup but our approach provides a large degree of customization and can be applied with any setup for macro photography.

2.7.1. Description of the case studies

The experimentation involved three case studies, with the aim to study benefits and drawbacks of the proposed methodology with different types of objects. Since the methodology can be applied to objects very different in shape, complexity, material, surface characteristics and size, the case studies have been chosen in order to represent a wide range of issues. For this reason, we chose an ancient jewel with enamel details and a series of small encrustation specimens.

The first case study concerns the reconstruction of a bronze zoomorphic fibula, depicting a bird with spread wings, with traces of enamel deposits on the wings and on the tail (Fig. 2.10). The piece, stored at the National Archaeological Museum of “Alta Val d’Agri” situated in Grumentun Nova (PZ, Italy), is dated between the sixth and seventh century AD and belongs to a typology that was widely used for female costumes in the Dark Ages, especially in southern Italy. The object size is enclosed in a bounding box of 37x18x16 mm, and has been acquired at 1:2 magnification.



Figure 2.10. Fist case study, a bronze zoomorphic fibula with dimensions (LxWxH) of 37x18x16 mm.

The second case study includes four specimens of encrustations extracted from a marble statue of a bearded Triton (Fig. 2.11), most probably dating back to the Roman era, which was discovered in the waters of the Blue Grotto of Capri in the late Sixties. The first specimen (classified as T2) presents, on almost the entire surface, deposits of bryozoa and coatings are also evident due to encrusting red algae. The T9 presents a surface covered by bryozoa deposits and traces of sponge, while the specimen T10 shows traces of calcareous red algae. This sample is made of a small piece of marble, which is highly degraded due to micro perforations. Calcareous tubes produced by serpulidae are also visible. The specimen T16 presents brownish encrustations and traces of serpulidae and calcareous sponges and a clearly visible structure of a gastropod is located at one end. The maximum dimensions of the four specimens range from 9 mm (T2) to 19.6 mm (T16), and have been acquired at 1:1 magnification.

The third case study is represented by two samples (Fig. 2.12) extracted from a pavement located at the underwater archaeological site park of Baiae (Naples, Italy), precisely at a depth of 5 meters in the area belonging to the Villa Protiro. The first specimen (M1) is a calcareous sample, having the following measurements: length 21.5 mm, width 22 mm, and depth 17 mm. The degradation forms recognized on the surface of the sample are: calcareous deposits and pitting, sponges, encrusting algae, serpulids. Moreover, bryozoa were also detected. In particular, the images obtained by our optical setup have highlighted the qualitative distribution of degradation agents on the stone material (an area colonized for about 50% by sponges, 35% by encrusting algae, 10%

by serpulids and 5% by bryozoa). The second sample (M2) is a piece of marble, having the following measurements: length 15 mm, width 11 mm, and depth 10.6 mm. Also in this case the typologies of damage identified are: calcareous deposits and pitting, sponges and serpulids. In particular, acquired data showed that the surface of the sample is colonized by sponges for about 70% and for 30% by serpulids.

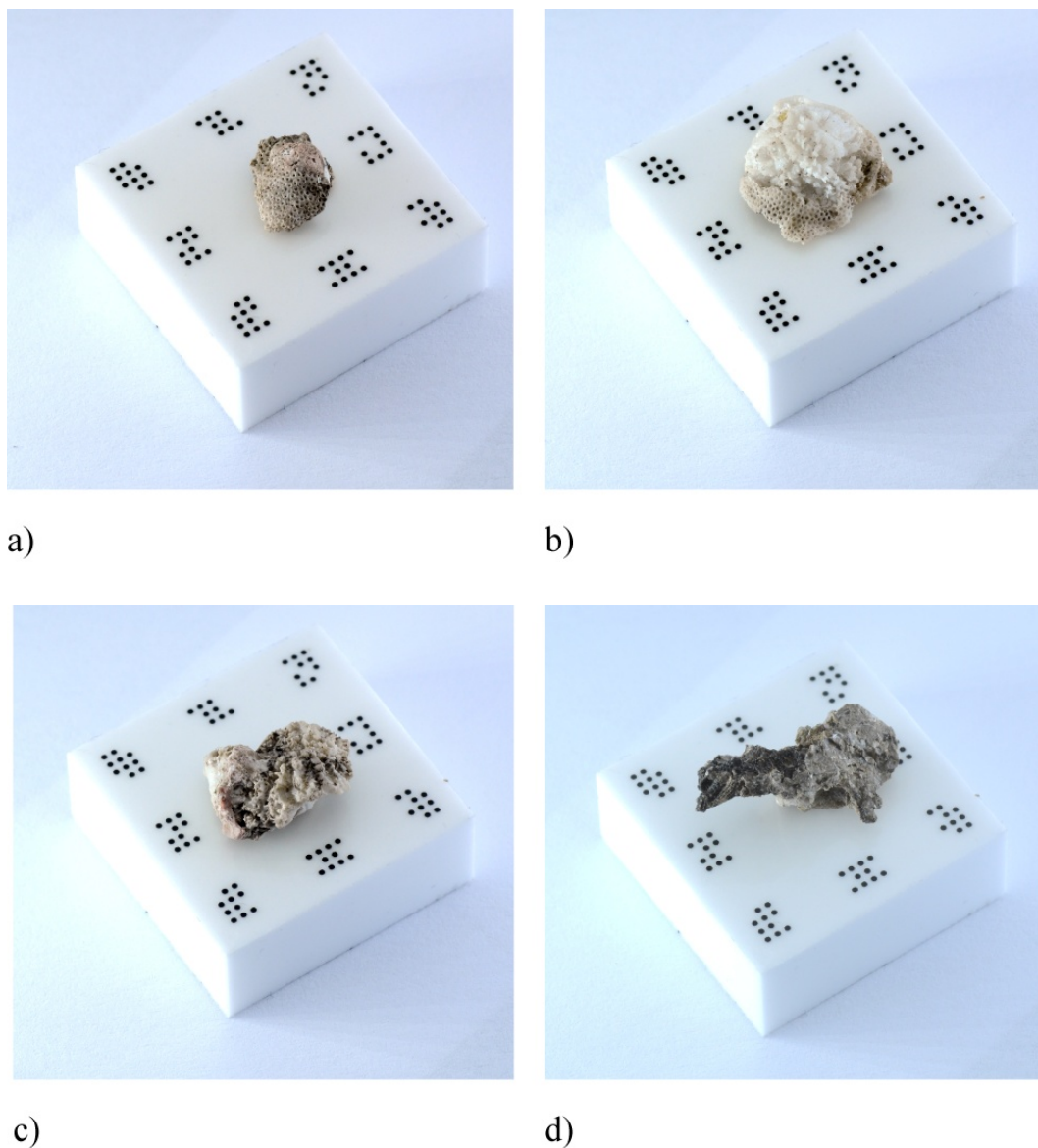


Figure 2.11. Second case study. Four specimens of encrustations, (a) T2, (b) T9 (c) T10 and (d) T16, extracted from a marble statue of a bearded Triton. The support plate is a square of 25 x 25 mm.

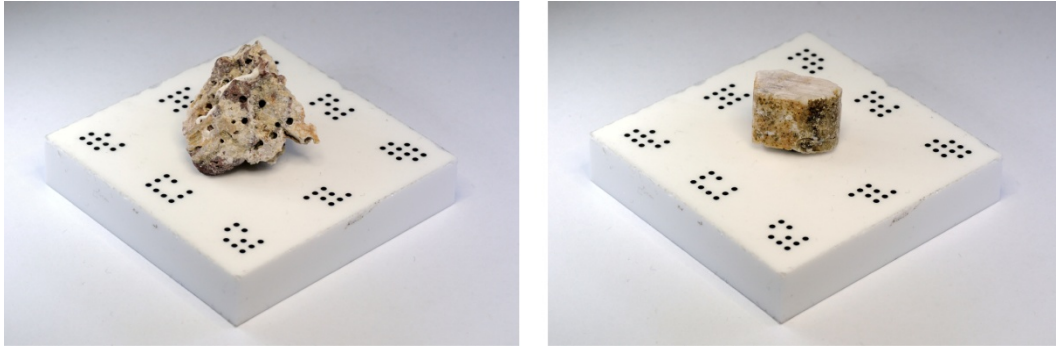


Figure 2.12. Third case study. A calcareous sample (left) and a marble sample (right). The support plate is a square of 25x25mm.

2.7.2. Image acquisition

The shooting parameter has been set following the guidelines previously presented in section 2.4. The size of CoC measures 0.019 mm and has been measured a pupil magnification of 0.63 at the minimum focus distance. Using a conservative value of 700 nanometres for the wavelength in equation (2.1), a diffraction limit aperture of f-11 has been assessed.

To remotely control the shooting parameters the camera control software included in Helicon Focus (Heliconsoft, 2012) has been used. The number of pictures in each image stack has been chosen accordingly to the object dimensions. Assuming a DoF evaluated using equation (2.2) of about 1 mm and a focusing step of 50 μm for the lens, it is sufficient to take a picture every 0.5 mm in order to ensure a 50% overlap.

Different software solution and image fusion algorithms have been tested. Both the algorithms included in Helicon Focus did not lead to encouraging results, especially in resolving the finer details. For this reason has been used the pyramid based image fusion algorithm embedded in Zerene Stacker (Zerene Systems 2012), which showed better results.

For each stacked image has been created a mask with the help of the white background, in order to help the reconstruction process by avoiding false matches and speed-up the 3D point cloud computation.

2.7.3. 3D reconstruction

The image orientation has been carried out through the self-calibration bundle-adjustment procedure implemented in Bundler (Snavely *et al.* 2007), using the features extracted and matched by the SIFT operator. Then, a dense 3D point cloud has been computed using the multi-view stereo software PMVS2 (Furukawa & Ponce 2007). The results of the SFM process and dense reconstruction are summarized in Table 2.4 and 2.5 respectively.

The point cloud has been edited with mesh processing software to remove the outliers points and build a triangular mesh. Then, the 3D model has been scaled using the reference gauge blocks and the black dots printed on the support plate and completed with high quality textures extracted by projecting the original high resolution images.

2.7.4. Experimental results

The results show a high level of detail of the reconstructed surface but, as it was expected, are present some non-reconstructed areas due to the surface properties and occlusions.

The fibula has been reconstructed with about 3.5 million of points and 7 million of triangles (fig. 2.13). The high quality resulting model allows to analyse all the details of the fibula, such as the painted areas on tail and wings (fig. 2.14).

All the four specimens belonging to the second case study have been successfully reconstructed (see figures 2.15, 2.16, 2.17, 2.18) and are present only some small unreconstructed areas due to texture-less regions, especially in the reconstruction of T2 (fig. 2.15) and T9 (fig. 2.16) specimens in which the characteristics of the marble material are more influent. Despite this issues, has been possible to make measurements of very small details, such as hole due to pitting with a diameter of about 0.4 mm, and features of about 0.2 mm on a gastropod with a length of 4 mm (fig. 2.19).

High resolution 3D model are also obtained for the specimens belonging to the third case study (fig. 2.20 and fig. 2.21). In particular has been possible to identify and effectively measure the effects of pitting and traces of serpulids (fig. 2.22).

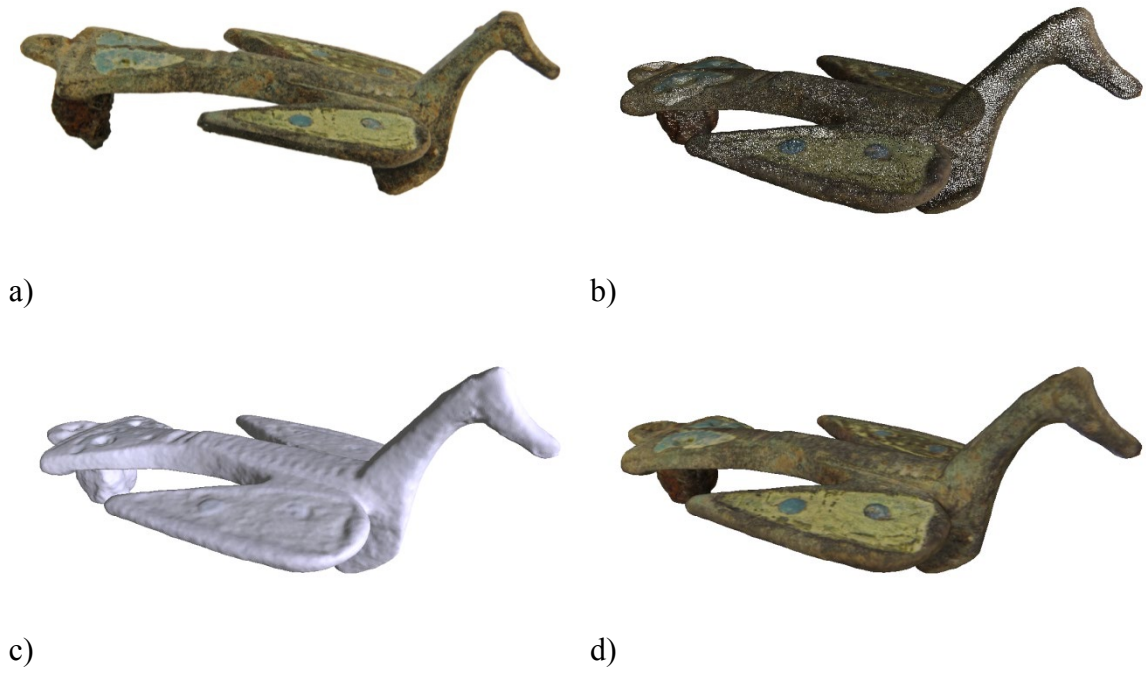


Figure 2.13. First case study: one of the input image (a), the 3D point cloud (b), the reconstructed polygonal mesh (c) and the final textured 3D model (d).

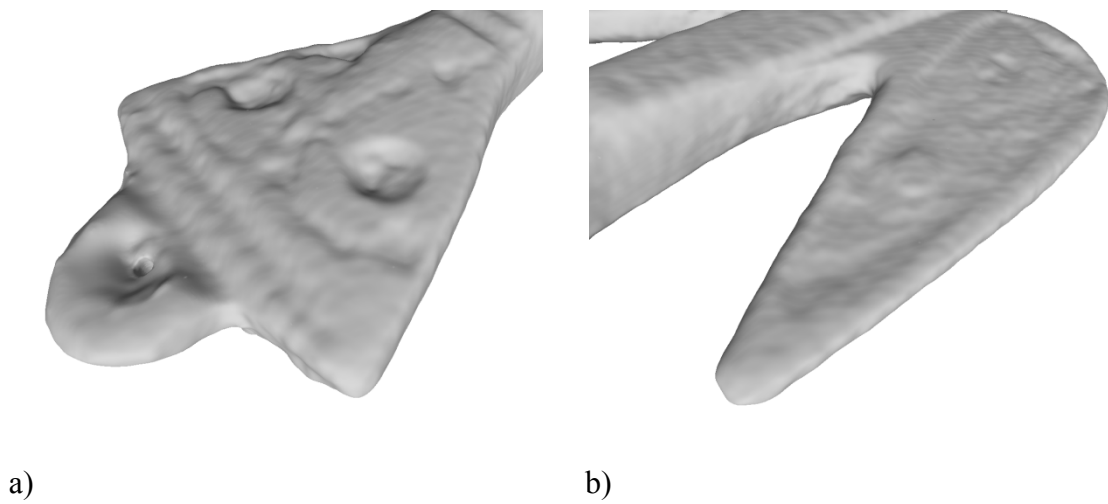


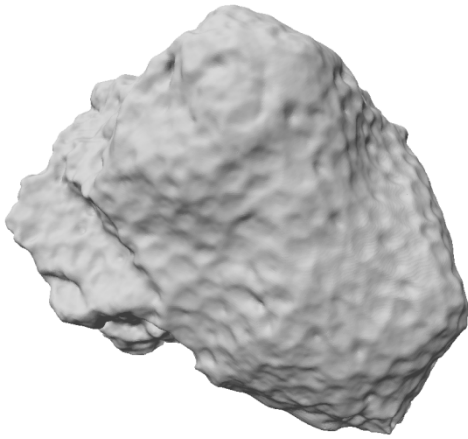
Figure 2.14. First case study. Detail of the tail and wings.



a)



b)



c)

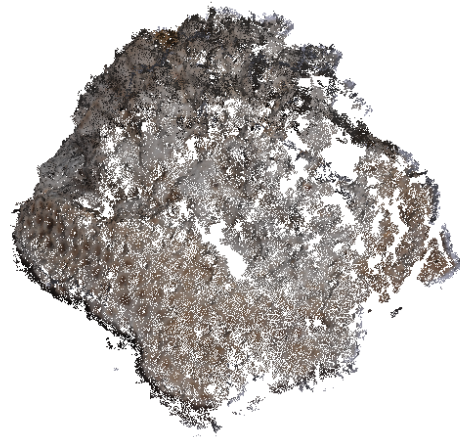


d)

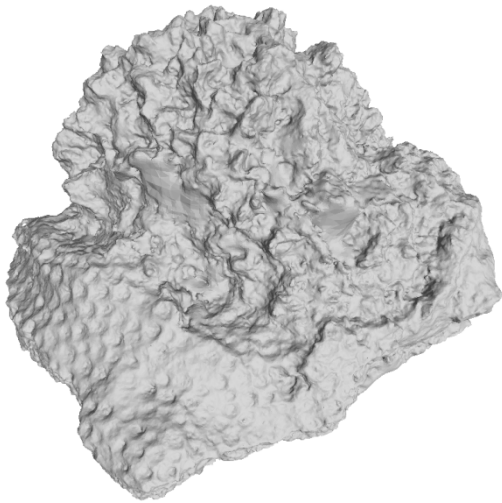
Figure 2.15. Second case study, T2 specimen: one of the input image (a), the 3D point cloud (b), the reconstructed polygonal mesh (c) and the final textured 3D model (d).



a)



b)



c)

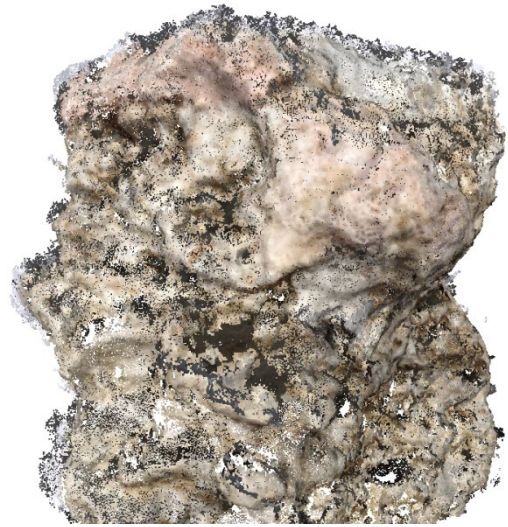


d)

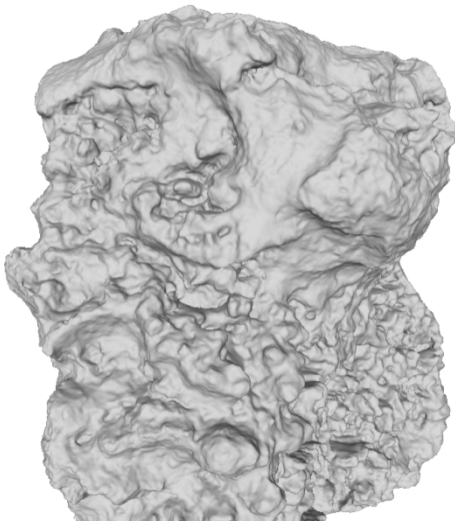
Figure 2.16. Second case study, T9 specimen: one of the input image (a), the 3D point cloud (b), the reconstructed polygonal mesh (c) and the final textured 3D model (d).



a)



b)



c)



d)

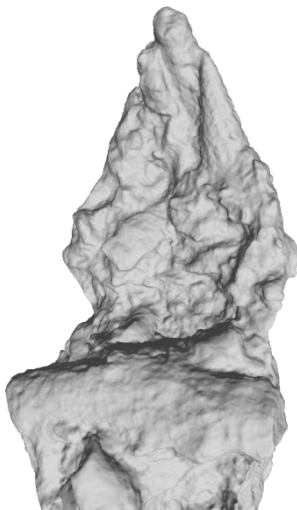
Figure 2.17. Second case study, T10 specimen: one of the input image (a), the 3D point cloud (b), the reconstructed polygonal mesh (c) and the final textured 3D model (d).



a)



b)

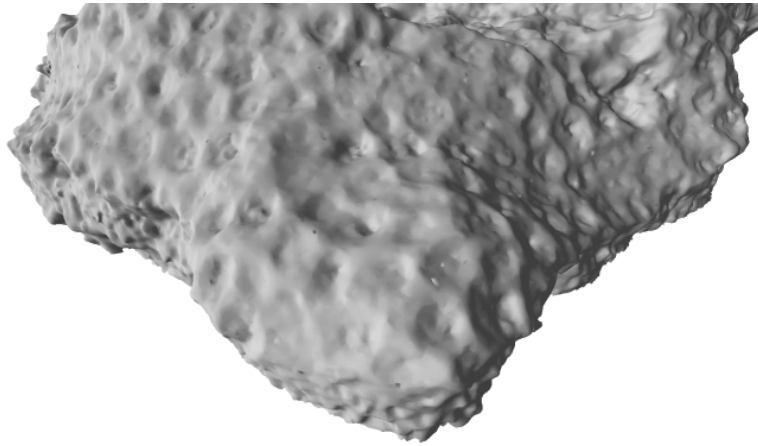


c)

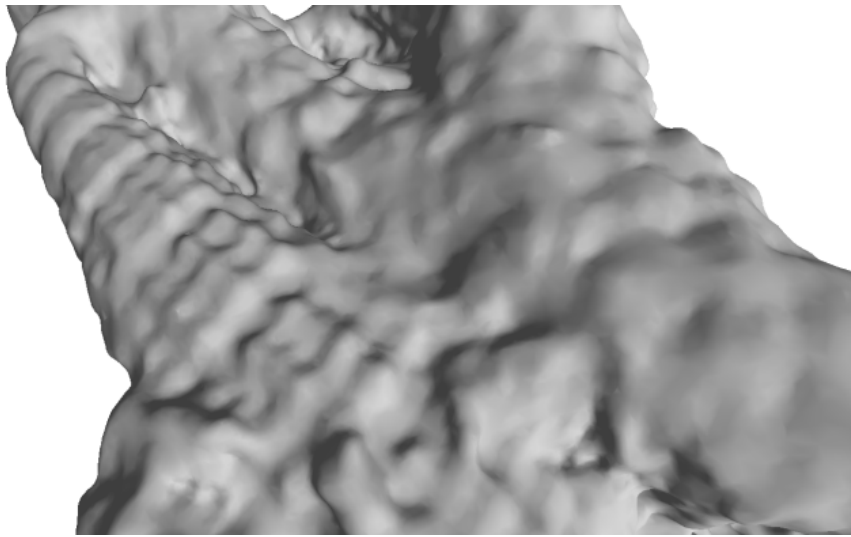


d)

Figure 2.18. Second case study, T16 specimen: one of the input image (a), the 3D point cloud (b), the reconstructed polygonal mesh (c) and the final textured 3D model (d).



a)



b)

Figure 2.19. Detail of pitting on the T9 specimen (a) and gastropod in the T16 specimen. The holes due to pitting measure about 0.4 mm while the gastropod is about 4 mm long.

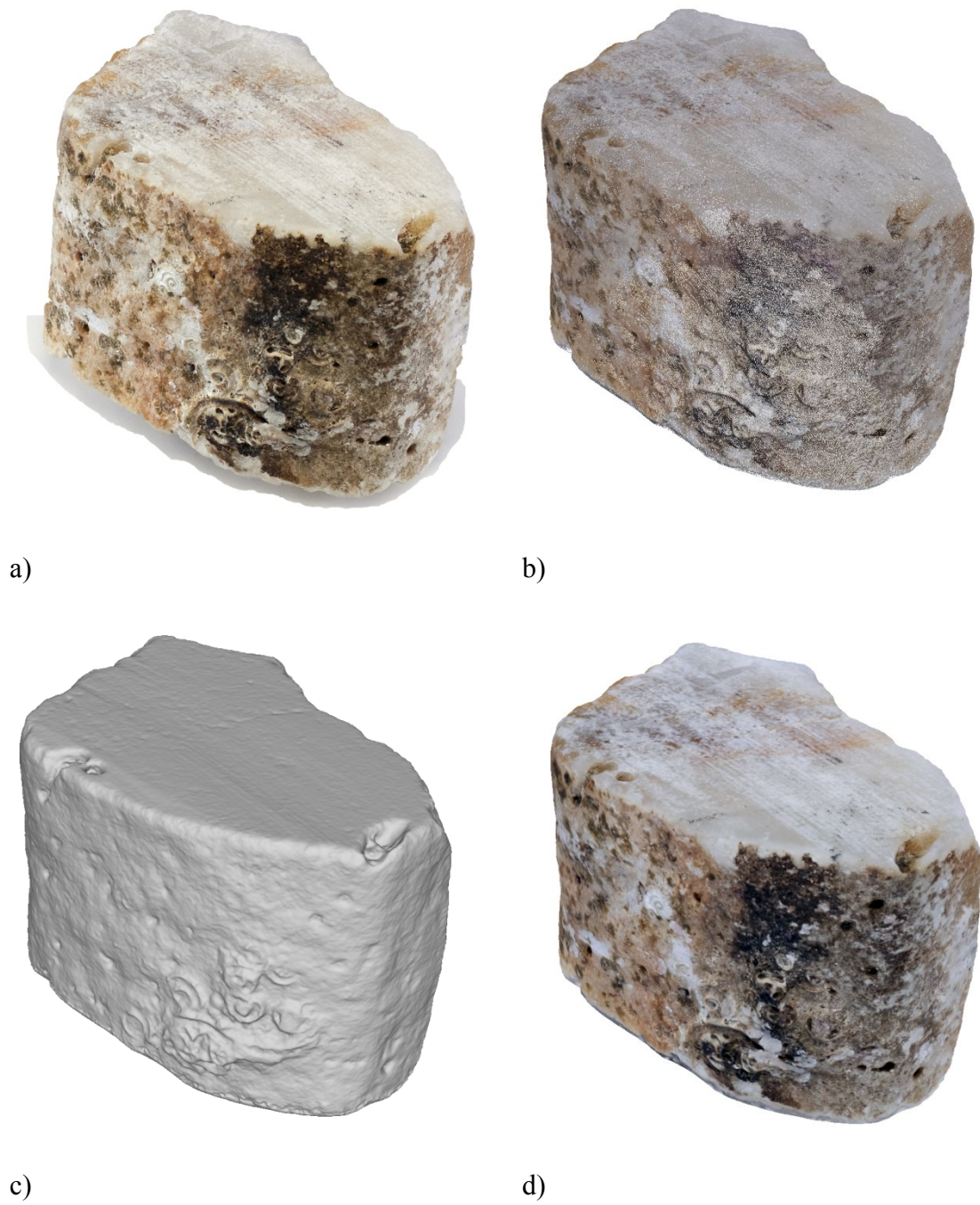


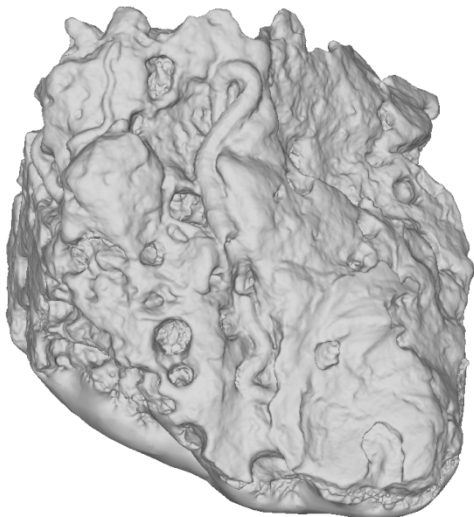
Figure 2.20. Third case study, M1 specimen: one of the input image (a), the 3D point cloud (b), the reconstructed polygonal mesh (c) and the final textured 3D model (d).



a)



b)

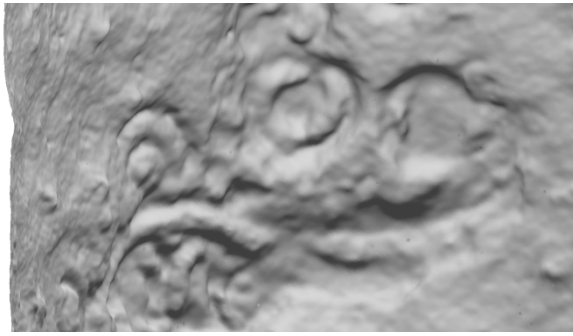


c)

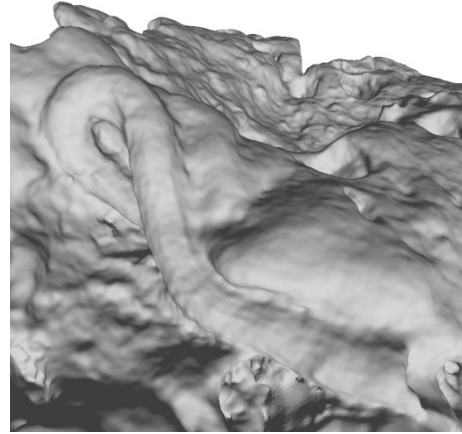


d)

Figure 2.21. Third case study, M2 specimen: one of the input image (a), the 3D point cloud (b), the reconstructed polygonal mesh (c) and the final textured 3D model (d).



a)



b)

Figure 2.22. Third case study, details of serpulids on M1 (a) and M2 (b) specimen. The mean diameter of the serpulidae on specimen M2 measures about 1 mm, while from the traces on specimen M1 the diameter of the serpulids range from 0.1 to 0.5 mm.

| | First case study | Second case study | | | | Third case study | |
|-------------------------------------|------------------|-------------------|-------------|-------------|-------------|------------------|----------|
| | Fibula | T2 | T9 | T10 | T16 | M1 | M2 |
| <i>Dimensions (L x w x h, mm)</i> | 37 x 18 x 15 | 9 x 7 x 5 | 15 x 13 x 7 | 13 x 10 x 7 | 20 x 11 x 6 | 21.5x22x17 | 15x11x11 |
| <i>Magnification</i> | 1:2 | 1:1 | 1:1 | 1:1 | 1:1 | 1:2 | 1:1.5 |
| <i>Number of camera poses</i> | 51 | 30 | 26 | 24 | 29 | 36 | 36 |
| <i>Number of slices</i> | 18 | 12 | 13 | 14 | 22 | 16 | 18 |
| <i>Reconstructed points[x 1000]</i> | 3500 | 735 | 500 | 990 | 1500 | 1770 | 1450 |

Table 2.4. Results of the 3D reconstruction relative to the three case studies.

The total reconstruction time depends mainly on object dimensions and surface complexity and in a second instance on computational resources. A complex object requires more camera poses to cover the whole surface, while larger dimensions need more slices to explore the entire DoF.

The acquisition time depends mainly on the speed of the autofocus motor and the data transfer rate. In addition, the process is only partially automated, because we use a non

motorized rotational stage and because human intervention is needed to start the shooting for each angular position.

The 3D reconstruction time depends mainly on hardware resources. All the tests were conducted on a machine equipped with an Intel core i5 CPU, 8 Gb Ram, Ati Radeon 5770 GPU and Windows 7 64 bit. The results (Table 2.5) show that computational time ranges from 30 to 70 minutes, depending on the number of images and object coverage on the imaging sensor.

The total time is affected by some manual operations, especially during the acquisition and pre-processing time. However, the time spent to pre-process each image for mask creation has a positive impact on the 3D reconstruction. In fact the computational time can be reduced up to 75%, making the manual cleaning operation on the resulting point cloud unnecessary.

| | First case study | Second case study | | | | Third case study | |
|--|------------------|-------------------|-----|-----|-----|------------------|-----|
| | Fibula | T2 | T9 | T10 | T16 | M1 | M2 |
| <i>Image acquisition (minutes)</i> | 130 | 51 | 48 | 47 | 90 | 80 | 90 |
| <i>Image fusion</i> | 32 | 13 | 12 | 12 | 22 | 20 | 20 |
| <i>Image Pre-processing (minutes)</i> | 17 | 10 | 8 | 8 | 10 | 10 | 10 |
| <i>3D reconstruction (minutes)</i> | 60 | 35 | 32 | 30 | 34 | 65 | 70 |
| <i>Scaling and texturing (minutes)</i> | 10 | 10 | 10 | 10 | 10 | 10 | 10 |
| <i>Total time (minutes)</i> | 249 | 119 | 110 | 107 | 166 | 185 | 200 |

Table 2.5. Time required to complete the whole 3D reconstruction pipeline. Most of the time is needed for the image acquisition phase.

Increasing the DoF by closing the aperture is a solution that can reduce significantly the whole acquisition time. In (Fig. 2.23) and (fig 2.24) is presented a comparison between the 3D reconstruction obtained by our approach and the ones resulting by using the maximum f-number. The object's shape has been reconstructed in both cases, but the loss of detail due to diffraction is evident.

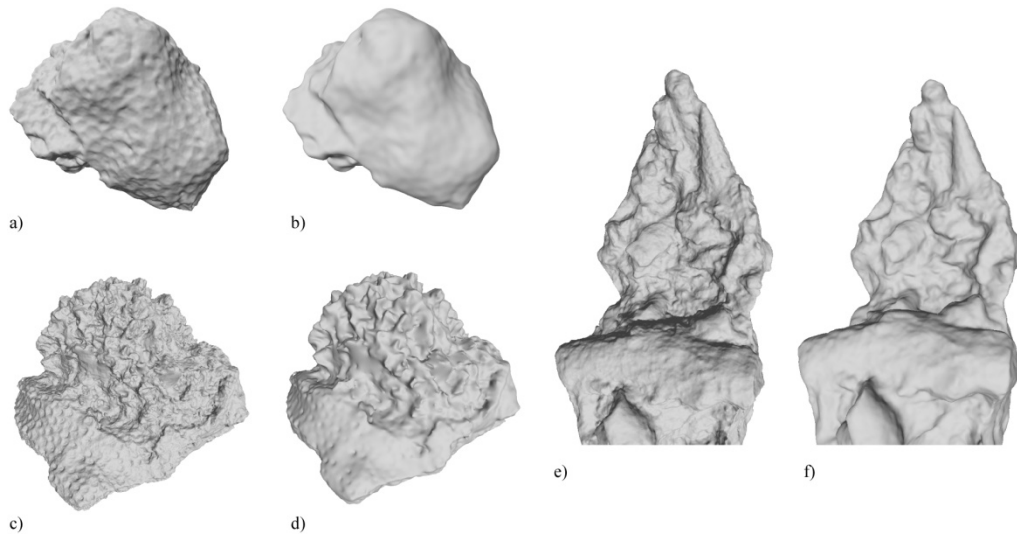


Figure 2.23. Comparison between the 3D reconstructions obtained by our approach for the specimens T2 (a), T9 (c) and T16 (d) and the one obtained using the multi-view stereo technique on a sequence of images taken at an aperture value of f-64 (b), (d) and (f). The diffraction effect causes a loss of detail that result in a smoother surface.

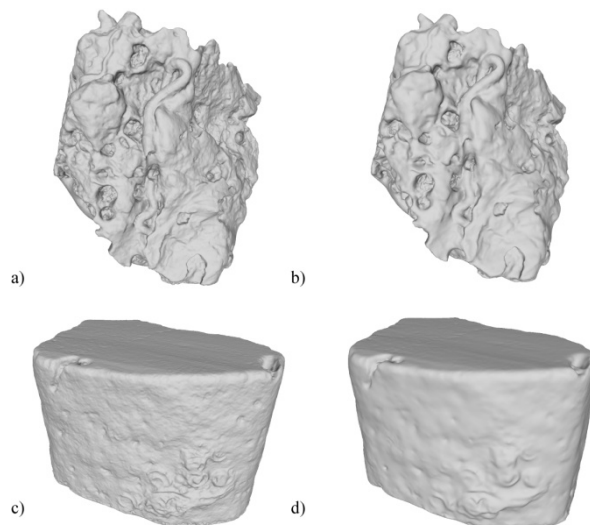


Figure 2.24. Comparison between the 3D reconstruction obtained by our approach for the specimen M1 (a) and M2 (c) and the one obtained using a sequence of images taken at the maximum f-number (b) and (d). The diffraction effect causes a loss of detail that result in a smoother surface.

2.8. Comparison with depth from focus technique

The proposed methodology requires a high number of pictures in order to produce a complete 3D model. In order to compare our approach with an optical technique that can share the same optical setup, we chose the depth-from-focus algorithm implemented in Helicon Focus.

In this section will be presented a comparison between the models obtained using our approach and the one obtained with the depth-from-focus technique relative to an important selection from the second (fig 2.25) and third (fig 2.26) case studies.

Above all, depth from focus has the advantage of requiring only one stack of images to compute a 3D point cloud. The differences lay mainly on depth resolution and object completeness. The model obtained using depth-from-focus is a 2.5D height map, composed by a set of slices which. Furthermore, the model is acceptable only from a viewpoint, while with the multi-view technique it is possible to recover the whole object. The depth resolution is strongly related to mechanical and optical constraints, in particular the minimum DoF and on the size of each focusing step.

In fig. 2.25 there is a comparison between the models obtained using our approach and the ones obtained with the depth-from-focus technique relative to the specimens T2 and T16 of the second case study, while in fig. 2.26 is reported the comparison relative to the specimen M1 of the third case study.

The 3D reconstructions showed in fig. 2.25 have been obtained with an aperture set to f-11 (Fig. 2.25 b-e) and to f-5.6 (Fig. 2.25 c-f). The former has been obtained from a stack of 14 images taken with a focus step of 0.5 mm, the latter from 70 images with a focus step of 0.25 mm. Both focus steps are equal to the DoF at the respective apertures and to the measured depth resolution

In the comparison in fig. 2.26 is presented the 3D reconstruction obtained with a stack of 86 images taken at f-4.5. Also in this case the thickness of each slice by which the model is composed is equal to 0.25 mm.

These values are much higher than those measured on the 3D models obtained applying our methodology based on multi view stereo.

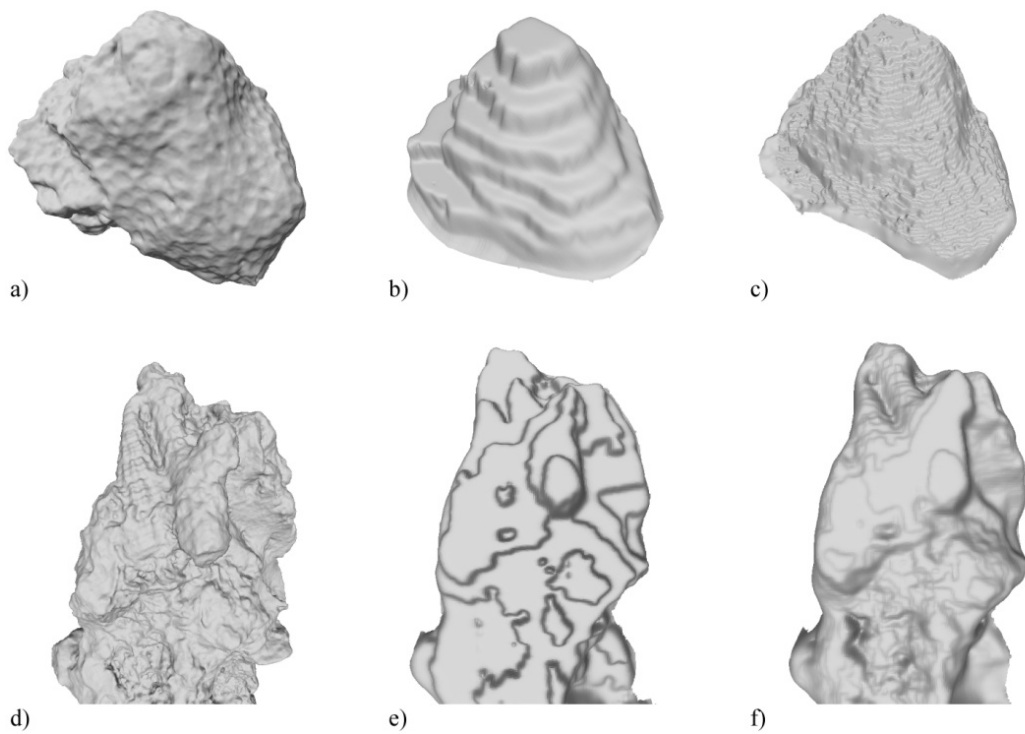


Figure 2.25. Comparison between the 3D reconstruction obtained by our approach for the specimens T2 (a) and T16 (d) and the one obtained using a depth from focus algorithm on a stack of 14 images taken at f-11 (b) and (e) and on a stack of 70 images taken at the smallest aperture of f-5.6 (c) and (f). The depth resolution relative to the depth from focus algorithm is equal to half DoF, 0.5 mm and 0.25 mm respectively, against a resolution of 20 micrometer of our approach.

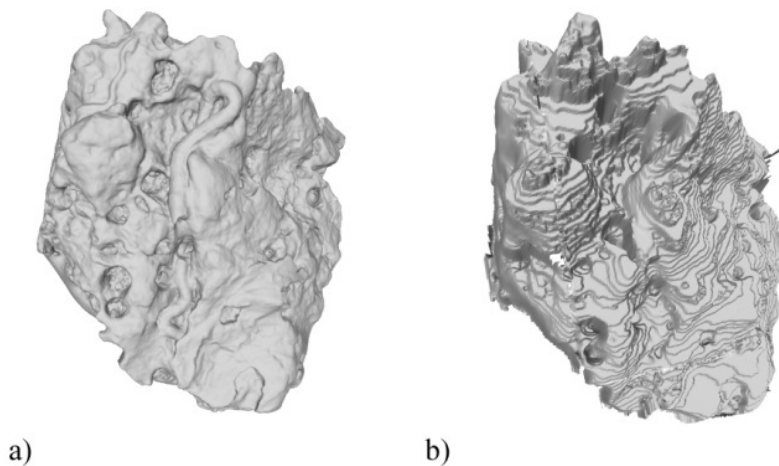


Figure 2.26. Comparison between the 3D reconstruction obtained by our approach for the specimens M1 (a) and the one obtained using a depth from focus algorithm on a stack of 86 images taken at f-4.5 (b). The depth resolution relative to the depth from focus algorithm is equal 0.25 mm, against a resolution of 20 micrometer of our approach.

2.9. Summary and discussion

In this chapter has been presented a new methodology to achieve high quality 3D reconstruction of small objects using off-the-shelf photographic equipment and with the joint use of image fusion algorithms and multi-view stereo techniques. The research has been designed to solve some problems related to the use of passive 3D reconstruction optical techniques, such as the very small depth of field at high magnification and the loss of contrast due to diffraction.

The whole 3D reconstruction pipeline can be performed with the use of open source and low cost software, with very high automation and customization possibilities.

The problem of verifying the system accuracy has been addressed in our laboratory tests. Since is not possible to use commercial equipment (such as laser scanners) capable to acquire dense point clouds of objects with a maximum dimension of few centimetres, the accuracy has been evaluated through the 3D reconstruction of a set of gauge blocks with a manufacture accuracy smaller than the one we want to measure. We have evaluated a system resolution of 20 microns and an accuracy of about 10 microns. These values are similar to the ones achieved with structured light systems used for dental application.

Our methodology has been applied on three case studies. In particular the 3D modelling of a small artefacts and the 3D reconstruction of small stone samples deteriorated by marine environment characterized by a very complicated surface has been addressed. The first case study was the 3D reconstruction of a zoomorphic fibula, depicting a bird with spread wings, located at the National Archaeological Museum of “Alta Val d’Agri” situated in Grumentum Nova (PZ, Italy) the second one was the challenging 3D reconstruction of four specimens of encrustations extracted from a marble statue of a bearded Triton. The third case study is relative to the 3D reconstruction of two small stone sample extracted from a pavement located at the underwater archaeological site of Baiae (Naples, Italy).

A great attention has been devoted to the development of a high quality texture capable of keep all the surface information. For this purpose, shooting parameters and lighting conditions have been chosen to limit the problems arising from the use of macro lens at

high magnification, in particular the very shallow DoF and the loss of sharpness due to diffraction.

The results show a very high quality 3D model which makes it suitable for virtual representation, virtual restoration, duplication or interactive measuring. The reconstructed surfaces are characterized by an accuracy that is significantly higher than the one that can be obtained with other optical techniques used for small objects reconstruction, as we demonstrate in a comparison with the depth-from-focus technique. The downside is represented by the time needed for the whole process. For the most complex object the whole process lasted about 250 minutes, of which about two hours were spent only for the acquisition, due to the manual intervention of the operator and mechanical and transfer rates limits. The acquisition time can be reduced to few minutes using the smallest aperture, but with an evident loss of quality.

The next step of our research will focus on improving the process automation to increase the system productivity (by integrating the camera control software with a motorized rotational stage) and on field testing of our methodology on other types of small challenging artefacts, such as jewels.

Comparative analysis between active and passive stereo techniques for underwater 3d modelling of close-range objects

Two 3D optical techniques (active and passive) based on stereo vision has been tested and compared. A 3D system has been designed, made-up of a digital projector and two still cameras lodged in waterproof housings. Experimental tests were conducted in laboratory, dissolving different concentrations of clay in a water tank in order to simulate the typical visibility conditions of a seafloor. The RGB images acquired in turbidity have been analysed separately for each channel to improve the reconstruction quality. The performances of the two techniques are described and discussed

3.1. Introduction

In the field of underwater imaging, optical systems are widely used in applications like survey of submerged finds or monitoring of the marine flora. For these purposes movie and still cameras installed on robots or handled by scuba divers are employed to recover both texture and 3D shape. 3D imaging in underwater environment is a challenging issue, because of the hostile conditions, the poor visibility and the surface state that can be corroded and covered by marine flora and fauna. So, this limits the use of such systems at the close-range measurements. Using stereo systems 3D models are obtained from a pair of images taken from two views illuminating the scene with artificial or ambient light (passive approach). While a different stereo approach uses a projection of gray-code patterns (active approach) that serve to codify the scene for automatic stereo matching. This latter technique typically used in air, has been experimented in (Bruno *et al.* 2011) where acceptable results in close-range 3D acquisition are obtained.

In this chapter, this active technique is compared with a passive technique until now employed only in air applications. The goal of this research is to test the effectiveness of

stereo algorithms already consolidated in air applications, regardless the optical corrections in the image formation due to scattering media. Thus, the experimentations have been conducted in laboratory in a water tank to control the lighting conditions and the turbidity, in order to ensure the repeatability of measurements and to test their reliability. These preliminary tests aim to recover useful information that can be adopted for the design of a new 3D system based on stereo vision for real applications. For an even comparison the experimental setup does not change during the acquisition process for both the techniques. Moreover, we have taken into account the optical limitations due to the turbidity considering a multichannel image analysis to improve the quality of 3D models.

In the next section *related works* we provide a brief review of 3D optical techniques usually employed in underwater environment, then in the section *active and passive stereo techniques* we detail the 3D stereo reconstruction based on active and passive approaches. Subsequently, in the section *experimentation*, the setup and the experimental tests conducted in a water tank will be described, in the section *results analysis* the results are discussed and comments about the comparison between the two techniques are reported in section *discussion*. The *conclusions and remarks* are reported in the last section.

3.2. Related works

In this section a quick review of underwater 3D optical techniques is presented, paying a particular attention to stereo systems based on the correspondence problem that can be solved with active and passive approaches.

3.2.1. Optical imaging systems

Optical imaging techniques used in underwater environment for 3D data acquisition are based on multiple acquisitions of the scene taken by different viewpoints, with movie or still camera. In these techniques, artificial lights (lamps and spotlights, for example) are used just to illuminate the scene if needed, and are not exploited in the

triangulation of the 3D points, which is in turn based on the knowledge of similar points in the image sequence, found through stereo matching algorithms.

Stereo systems use two digital calibrated cameras to capture the scene; they may be installed on underwater robots in applications like seabed mapping (Allais *et al.* 2007), in situ estimation of the length of marine organisms (Harvey & Shortis 1995). The use of three synchronized cameras (Menci 2012) represents a good compromise between accuracy and device size for underwater survey conducted by a scuba diver. Movie or still cameras, lodged in water-proof housings, are used for example in seabed mapping (Leone *et al.* 2006), survey of archaeological sites (Drap *et al.* 2007). To reconstruct a scene from a sequence of overlapping images acquired by a single moving camera, structure-from-motion techniques are widely used. They involve the automatic extraction of points of interest (a sparse set of features, such as corners), the tracking of this sparse set of features across the image sequence, and the estimation of their 3D positions using multiple views (Saito *et al.* 1995).

3.2.2. Camera calibration in underwater environment

A calibration procedure is a necessary requirement for the extraction of 3D metric measurements from a set of digital images. In the literature, a lot of algorithms and methods for camera calibration have been developed, generally based on perspective or projective camera models. The calibration of an underwater optical device must consider the effect of refraction at the air-acrylic and acrylic-water interfaces, which are present when a camera is mounted in its housing (Lavest *et al.* 2000). The refraction can be considered through the explicit modelling of the optical paths using ray tracing (Li *et al.* 1997) or with the implicit incorporation of refraction effect into camera calibration parameters (Schewe *et al.* 1996a). In the first case, the cameras are first calibrated in air, and then calibrated in water to derive the geometry of the refractive interfaces. The reason for this two stage process is partly due to analogy between the fact that the principal component of both refractive effect and image distortion is radial (Fryer & Fraser 1986). Although the assumption on the refractive indices for the acrylic dome port and the internal air path is reliable, regarding the water changes in temperature, pressure, salinity can cause small change of the refractive index of the media, which

cannot be eliminated. The alternative implicit approach incorporates refractive effects of the optical components and refractive interfaces into camera calibration parameters.

3.3. Active and passive stereo techniques

In this section, are described methods and algorithms on which the two stereo techniques used in the experimentation are based.

3.3.1. 3D stereo reconstruction

Given two images acquired from slightly different viewpoints, a stereo matching algorithm tries to identify the corresponding points (solving the so-called correspondence problem) in both the images related to the same scene point (Fig. 3.1). Knowing these correspondences and the camera geometry, the 3D world coordinates can be reconstructed by triangulation. Finding a pair of correspondent points is not so trivial. Since the scene is acquired from different points of view, it is possible to find false correspondences due to occlusion, radiometric distortion and perspective distortion. The correspondence problem can be solved by using stereo matching algorithms (see chapter 1 for more details), but a different approach involves the projection of a sequence of structured-light patterns in order to codify the scene to be acquired. In this way, the matching of the points is performed automatically because each point object is uniquely determined by a retrieved code. In literature, gray-code and fringe projection technique are used to assign a binary code (Barone & Razionale 2004) or to detect the phase map, respectively.

Passive stereo

In this experimentation a Multi-View Stereo (MVS) algorithm has been used to solve the correspondence problem on a calibrated image pair (Furukawa & Ponce 2007). This algorithm typically employed for 3D reconstruction by multiple views, in the laboratory tests has been used in a stereo configuration for the comparison with the stereo active technique already tested in (Bruno *et al.* 2011). The number of 3D points

calculated is strictly related to image resolution, image quality (contrast, focus, etc.) and object surface properties, i.e. a low textured object can produce false correspondence and a lesser number of points.

Active stereo

Whole-field structured-light techniques are based on the projection of black and white light patterns. Different patterns as gray-code bands, sinusoidal fringes, grids, etc. are projected on the surface and the scene is acquired by a camera placed in a known position to calculate the triangulation (Salvi *et al.* 2004). The active stereo technique used in this experimentation is based on the codification of a black/white pattern set projected on the object by a digital projector. In particular, the gray-code technique generates a more effective matching between correspondent points in stereo pairs (Barone & Razionale 2004). The object is illuminated by a set of n temporally encoded patterns of black/white bands, with a progressively halved width, so that n images are captured by each camera. A binary code (n bit) is assigned to each point of the image, and the values 0 and 1 are associated to intensity levels, i.e. 0=black and 1=white (see Fig. 3.1). This procedure allows to codify $2n-1$ lines defined as crossing zones between white and black bands. Moreover, coded patterns with a bandwidth of 4 pixels, shifted in steps of 1 pixel for a total of 4 pattern positions, are used to exploit the minimum resolution of the projector. By projecting both horizontally and vertically striped coded patterns, a double code is assigned to the intersection points through horizontal and vertical lines. This procedure allows to codify automatically each point of the object surface. In this work a set of 8 vertical and 8 horizontal patterns (8-bit code) is used for gray-code. Other 4 + 4 patterns are projected for vertical and horizontal code shifting, respectively, with a bandwidth of four pixels. A projector resolution of 800x600 points allow to codify 799 lines x 599 lines = 478.601 points. The projector is only used to establish the correspondences and is not involved in the triangulation, so it is not required to calibrate its optics. In contrast to traditional passive approaches, this technique does not rely on images with consistent textures, because each point on the object surface is precisely identified thanks to a double binary code.

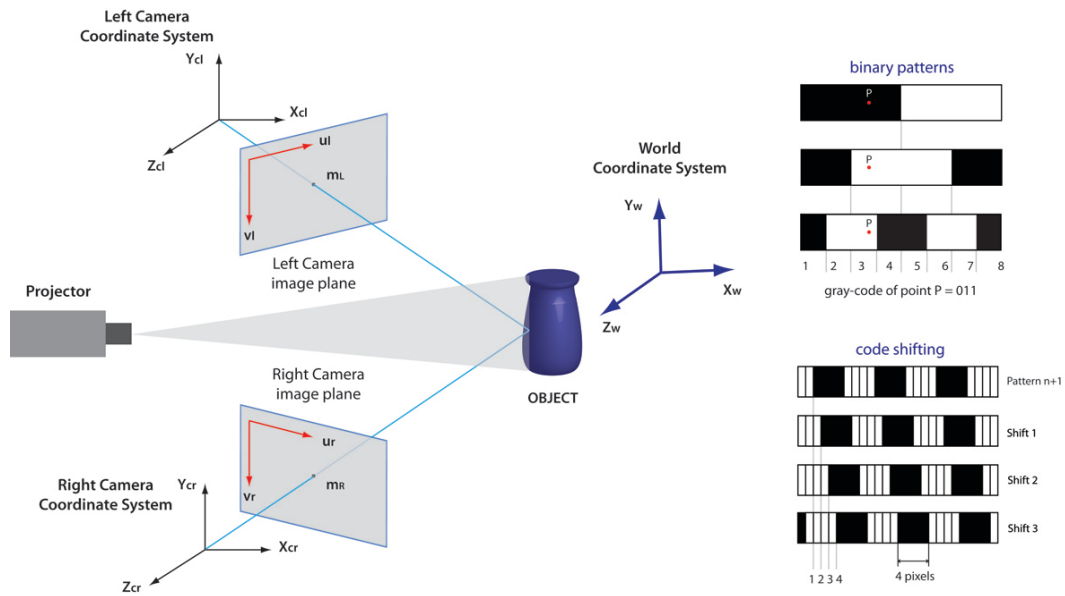


Figure 3.1. On the left, triangulation of a stereo configuration composed by two cameras and a projector. The points m_L and m_R in left and right images respectively, are the projection on the image planes I_L and I_R of the same 3D point w of the object. On the right, examples of binary patterns and code shifting.

3.3.2. Calibration

A calibration procedure is needed to compute the intrinsic parameters of each camera (focal length, coordinates of the principal points, radial and tangential distortions, pixel size) and the extrinsic parameters (translation and rotation with respect to a world coordinate system) of camera and stereo system. These parameters are computed using the well-known Camera Calibration Toolbox for Matlab (Bouguet 2012). The intrinsic and extrinsic parameters of each camera are obtained by correlating the coordinates of known points located on a calibration sample (i.e. checkerboard) with the corresponding coordinates on the image plane. The next step is to compute the extrinsic parameters of the system (stereo calibration), relating each camera frame to a unique world coordinate system. So, a relationship between the world coordinate system and the left and right camera coordinate systems is found.

3.4. Experimentation

Underwater imaging is seriously compromised by scattering and absorption in the medium, which decrease image contrast and attenuate light intensity with consequent

loss of details and colour alteration. The experimental tests have been conducted in laboratory using a water tank in order to control the lighting conditions, to ensure the repeatability of measurements and to test their reliability. In order to worsen the transparency of water (turbidity), simulating a real marine environment, different quantities of clay (with particle size of about few micrometers) have been dissolved in the water tank (2 x 3 m). In this way at each turbidity level, increases the loss of visibility and decreases the ratio of light reflected by the objects. A calibration of a volume of about 400 x 300 x 200 mm is performed at a distance of 1 meter and two ceramic objects have been acquired. As lighting source we have used only the projector, and we have measured the illuminance with a luxmeter at the working distance in front of the system, as reported in Table 3.1 at the different turbidity levels. The laboratory tests aim to analyze the effects of optical distortions, refraction and turbidity in the 3D reconstructions of close-range objects for both the techniques. Moreover, the variation of absorption of colour spectrum with the depth (at around 20-25m only blue and green wavelengths remain) has been neglected because an artificial light source (projector) has been used. In order to improve the quality of 3D reconstructions a spectral discrimination approach is used analysing the data obtained by elaborating the images in the three colours channels separately.

| <i>Turbidity level</i> | <i>Clay dissolved (mg/l)</i> | <i>Illuminance (lux)</i> |
|------------------------|----------------------------------|------------------------------|
| Clear Water | 0 | 96 7 |
| Turbidity level 1 | 5 | 80 4 |
| Turbidity level 2 | 10 | 70 5 |
| Turbidity level 3 | 15 | 44 6 |
| Turbidity level 4 | 20 | 21 5 |

Table 3.1. Measurements of the turbidity conditions.

3.4.1. Underwater setup

Two cameras Nikon D200, with a CCD (Charge Coupled Device) sensor size of 23.6x15.8 mm and resolution of 3872x2592 pixels equipped with an AF-Nikkor 35mm lens, are used to acquire the scene. The projector used in our experimentation is a Mitsubishi PK20, characterized by a very small size (123x48x97 mm), an acceptable resolution of 800x600 pixels with a brightness of 25 lumens and a low power consumption that allow to use a battery pack as power source. The three optical devices are protected in acrylic waterproof housings and fixed on an aluminium support; the optical setup is depicted in Fig.3.2. A dome port is mounted on the camera housing that reduces the refractive effects at the air-water interface (Edge 2006).

The system size is 950 x 530 x 560 mm, with a weight of about 17 kg in air. The two cameras are mounted on a common bar, in order to have a 620 mm baseline (the distance between the camera's optical centres) and a 36 degrees angle between the optical axes. The projector is fixed on an independent bar in a lower position, to take into account the differences between the Field-Of-VIEWS (FOVs) of the cameras and projector (Fig. 3.3). The latter generates an oblique off-axis projection, while the former presents a symmetric FOV with respect to the optical axis. We chose this setup in order to allow the acquisition of an object of 30-40 cm at a working distance of 1 m, maintaining a reduced scattering volume due to the overlap of the FOVs of the imaging devices. The system can be used also for different working distances, changing the position of cameras (baseline and orientation) on the aluminium frame as shown in Fig. 3.3.

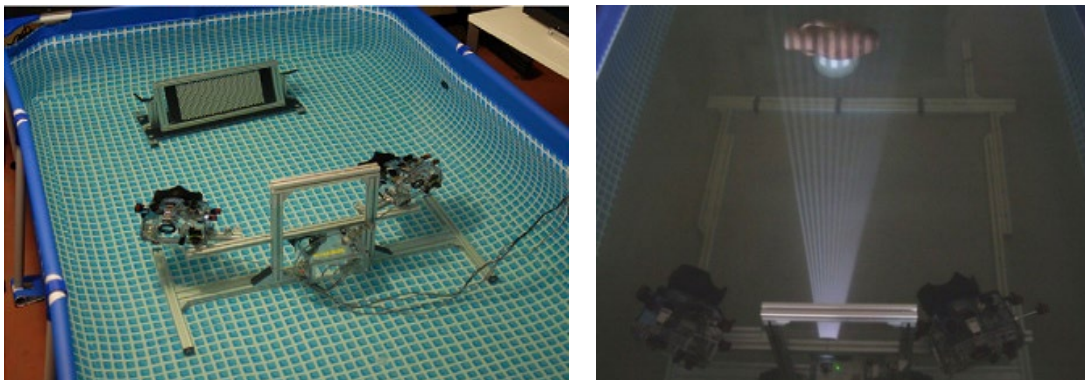


Figure 3.2. Optical setup of the underwater 3D acquisition system: underwater calibration (left) and acquisition (right).

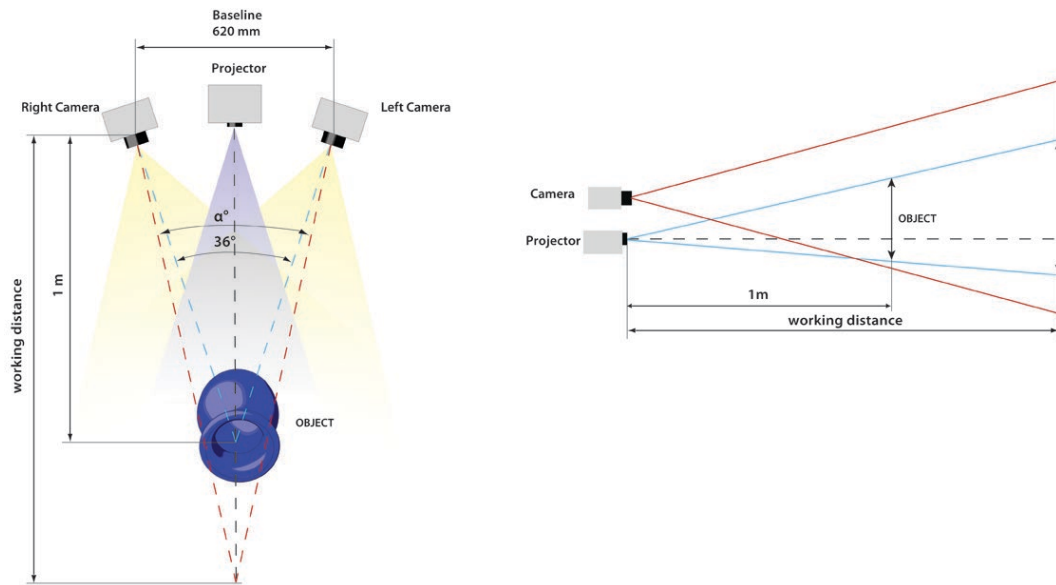


Figure 3.3. Optical setup of the underwater 3D system.

The two cameras are remotely controlled by a PC through a USB interface connection and the projection of the image sequence is controlled with an appropriate cable in order to optimize the acquisition time, manage and setting camera parameters and analyse in real time the captured images. The optical configuration of the system does not change for the active and passive techniques, thus we can compare more effectively the results of 3D acquisitions obtained with the same camera positions, optical devices, calibration parameters and working distance. In particular, the same calibration procedure is performed in water for both techniques, computing the intrinsic and extrinsic parameters without changing the cameras configuration. The use of a common setup can ensure a higher reconstruction accuracy avoiding systematic errors due to multiple calibration procedures. So we can perform an “even” comparison between the two techniques.

An amphora and a mask (Aeolus) have been acquired in air and water in this experimentation (Fig. 3.4). For the passive stereo technique, only an image pair has been acquired projecting a white pattern, while 50 images are necessary for the active stereo technique. The images have been taken in RAW format to ensure high quality and to extract each colour channel directly from RAW camera data. A 35 mm lens has been used in order to enclose in the camera FOV the whole object at a working distance

of about 1 m. During the acquisitions the aperture of camera is fixed to f/11 that ensures a depth of field of about 100 mm, and the exposure time varies from 0.8 sec for the more luminous patterns until 10 sec for dark ones, while for passive technique has been projected a white pattern.

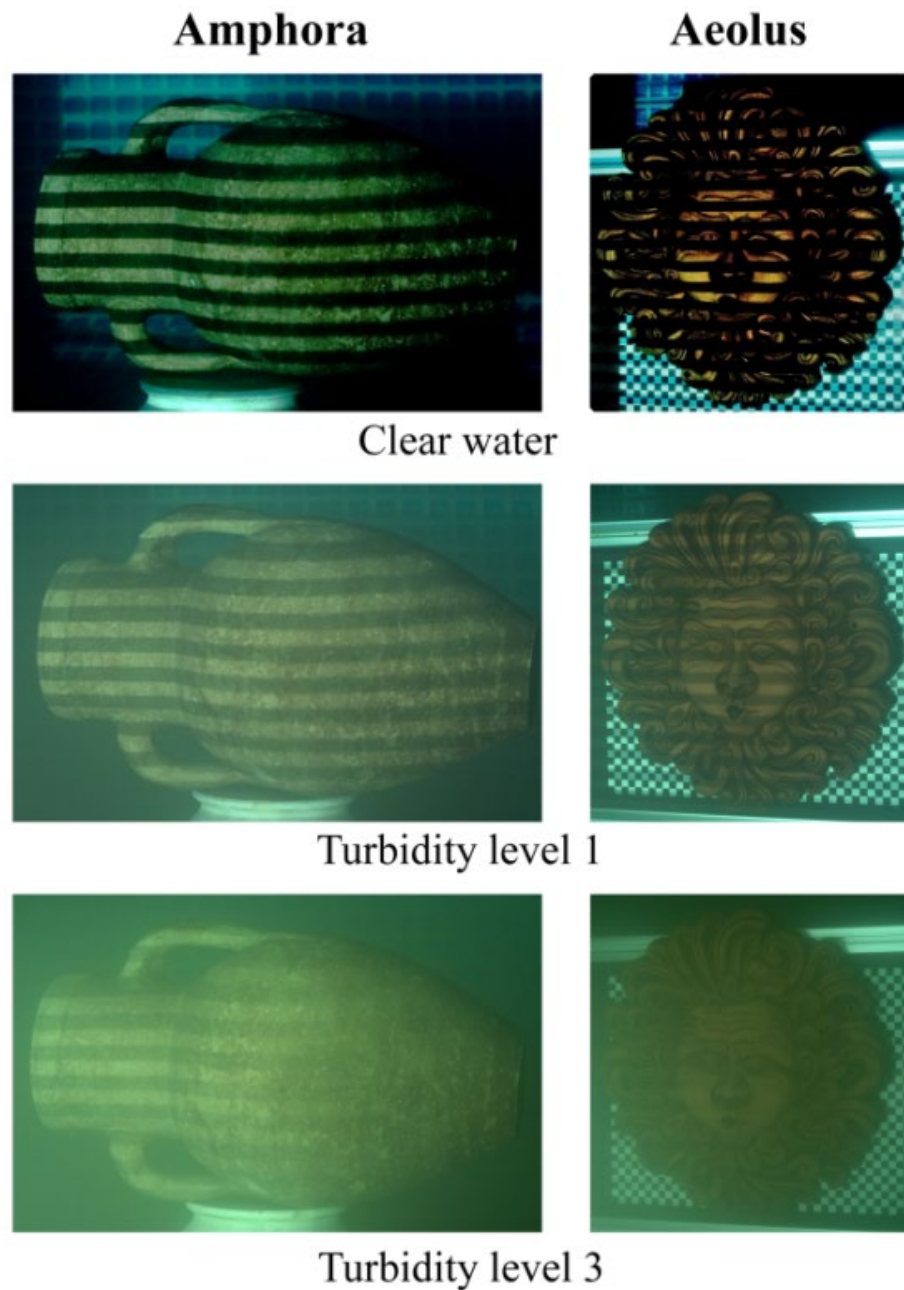


Figure 3.4. Images of the two objects with a projected gray-code pattern, acquired at different turbidity levels.

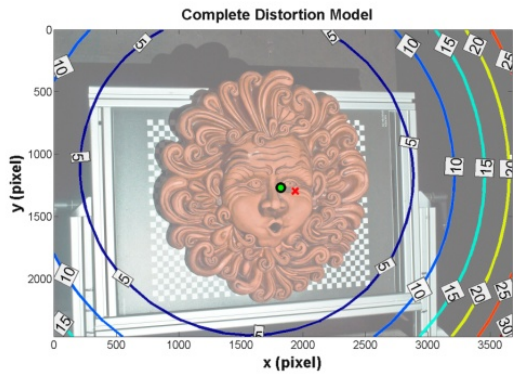
3.4.2. Calibration

Calibration is performed in air and in water tank, in order to compare the results of intrinsic and extrinsic parameters obtained in both conditions and to evaluate its impact on 3D reconstructions. The calibration is carried out with a planar black and white checkerboard pattern (Fig. 3.2) composed by 32 x 42 squares with an 8mm side. This calibration sample have been placed in several poses and acquired by the two cameras in RAW RGB format in order to extract each channel (Red, Green, and Blue) and perform a single colour calibration.

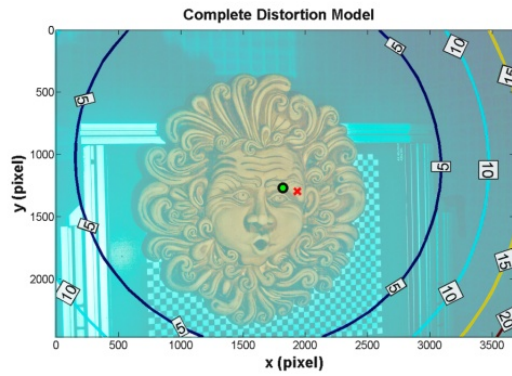
The system has been first calibrated using the RAW camera files, extracting demosaicked grey scale images. The comparison between calibration parameters in air with and without the dome port (Table 3.2) shows that the acrylic interface affects only the principal point coordinates that differ by 1 mm and 0.5 mm in x and y direction respectively, while focal length variation is equal to just 0.37 mm. This is due to a non-perfect alignment between the optical axes of camera lens and dome port, and to asymmetric components like the dome port shape, for example. The comparison between the values in air and in water shows a difference on the focal length due to the refractive effect of 4.87 mm that correspond to an increase of about 13%. This variation of focal length implies a decrease of the FOV proportional to the medium index because the CCD size is constant. Finally, the analysis of the principal point coordinates shows non evident changes. The complete distortion model defined as the sum of radial and tangential components for each camera in Fig.3.5, shows the displacement of the principal point respect to the image centre in x and y direction. The maps of optical distortions show that, in the image centre where the object to be captured is present, deviations are very small (maximum deviation at object border of about 3 pixels).

| <i>Parameter (mm)</i> | <i>Air</i> | <i>Water</i> |
|-----------------------|------------|--------------|
| Focal length | 37·19 | 42·43 |
| Principal point px | 12·23 | 11·04 |
| Principal point py | 8·22 | 7·42 |

Table 3.2. Intrinsic parameters (mean values for left and right camera) in air and water (RGB images).

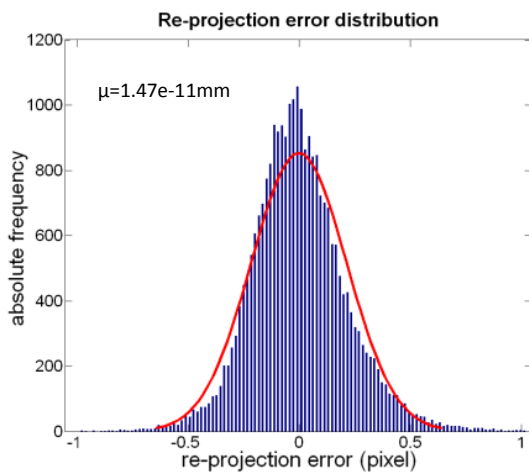


(a) Air.

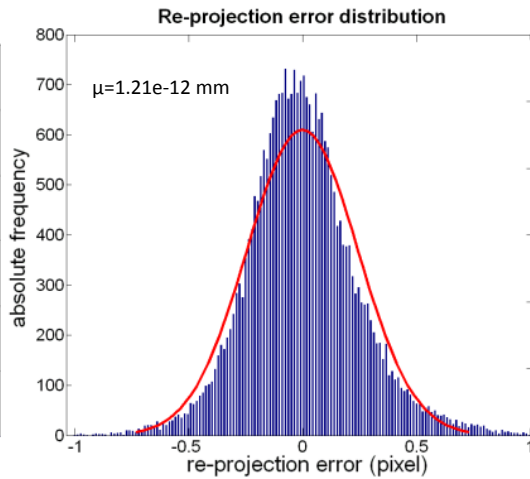


(b) Water.

Figure 3.5. Complete distortion model (radial plus tangential distortions) in air and clear water. The cross and the circle indicate the image centre and the principal point respectively. The contours represent the optical distortions values (in pixel).



(a) Air.



(b) Water.

Figure 3.6. Distribution of the re-projection error (in pixel) in air and clear water (RGB images).

To evaluate the effectiveness of the underwater calibration, we considered the re-projection error for each corner point of the calibration checkerboard, defined as the distance between the points of the actual image, and those computed from the back projection of the ideal checkerboard. The error dispersions are Gaussian-distributed and the values of mean and standard deviation obtained in water, computed on both x and y directions, are comparable with the ones obtained in air (see Fig.3.6). This comparison

allows us to retain that the assumed linear model of calibration in air can be used also in water, without take into account the refraction effects.

Finally, in order to take into account the effect of chromatic aberration we have estimated the calibration parameters in the red, green and blue colour channels, separately. This optical phenomenon can cause blur in the images due to a degraded focus quality (longitudinal chromatic aberration) and a misregistration of the colour channels (transversal chromatic aberration) that are difficult to reduce (Cronk *et al.* 2006). Despite this issue, the results have not shown significant differences in the values obtained by performing the camera calibration in water for each individual colour channel.

3.4.3. 3D reconstruction - Passive stereo

The dense stereo matching algorithm PVMS is used for the 3D reconstruction with the passive stereo technique. The algorithm inputs are an undistorted image pair and the 3×4 camera projection matrix, and the output is a coloured 3D point cloud. The camera projection matrix is computed from the same calibration parameter (intrinsic parameter matrix, rotation matrix and translation vector between left and right camera) of the stereo rig used for 3D reconstruction with the active technique. To correct the distortion of the image pair, a tool provided with the Camera Calibration Toolbox has been used (Bouquet 2012). The PMVS parameters used to finely tune the 3D reconstructions are the size of the correlation window and the level in the internal image pyramid that is used for the computation.

In this experimentation a fixed correlation window with a size of 7×7 pixels has been used, while the image resolution (image pyramid level) has been chosen accordingly to the turbidity level. A series of preliminary tests have been conducted in order to properly set up the algorithm parameters. In particular, for each turbidity level the pyramid level has been progressively reduced from a value of 3 (input image 8 times smaller) to 1 (halved images, 4 times less pixels) maintaining the correlation window fixed to a 7×7 pixels. Furthermore, the image pyramid level is fixed to 1, increasing the correlation window size from 7×7 pixels to 21×21 pixels. The results shown that as the turbidity increase, a wider correlation window is needed to overcome the loss of

contrast due to the scattering effect. Otherwise, the same results can be obtained by using a fixed correlation window on a smaller image (with a higher pyramid level). A larger correlation window means higher computation time, hence, in order to preserve the reconstruction quality and save time, has been used a small window of 7 x 7 pixels with a pyramid level value equal to 3 for the highest turbidity level, a value of 2 (images four time smaller) for the medium turbidity level and a value of 1 otherwise.

3.4.4. 3D reconstruction - Active stereo

The different image sets of the two objects have been processed to obtain the point clouds. In the active case, the software for 3D reconstruction is provided by Scansystems (Scansystems 2012), so from 50 image pairs and the calibration data has been obtained the 3D point cloud of the scene.

3.4.5. Accuracy evaluation

To evaluate the accuracy of the 3D reconstruction in air and clear water for both two techniques using the optical configuration of Fig. 3.3, a series of experiments based on a statistical approach have been planned. Two rectified samples, a plane (297 x 420 mm) and a cylinder (height 400 mm, radius 150 mm), have been acquired in air. These two samples allow to quantify and compare the displacements of the acquired 3D points respect to two known surfaces considered as reference. So, the accuracy has been evaluated measuring the discards of 3D data respect to the best fit surface by the means of Rapidform® software. In Table 3.3 are reported the higher mean values of discards for the two objects and the relative standard deviation. By the comparison we can conclude that the two techniques show a very similar behaviour in air and water.

3.5. Results analysis

The point clouds computed through the active and passive techniques must be further elaborated to obtain the final 3D models, because many outlier points are present. Rapidform® software has been used to edit the point clouds and also it

provides a series of tools to perform measurements and compare 3D data. In this analysis, we aim to quantify the performances of both techniques computing the density of 3D points, the portion of surface reconstructed, the percentage of outliers and the geometrical error of the obtained 3D model.

| | <i>Air</i> | | <i>Water</i> | |
|----------------|------------|---------------|--------------|---------------|
| | μ (mm) | σ (mm) | μ (mm) | σ (mm) |
| Active stereo | 0.20 | 0.10 | 0.50 | 0.72 |
| Passive stereo | 0.24 | 0.35 | 0.63 | 0.88 |

Table 3.3. Accuracy evaluation on a planar and cylindrical sample in air and water.

3.5.1. Point cloud editing

The first step consists in the manual deletion of outlier points that do not belong to the object surface (Fig 3.7-a, - b, -c). This is the most time consuming process, because is highly dependent upon the surface complexity of the object being processed. Whereas for a simple object such an amphora are required only few steps with the lasso selection tool, for a more complex surface is necessary to intervene manually, in a more precise way. The second cleaning step is represented by the application of the noise filtering implemented in the software (Fig. 3.7-d), that deletes automatically the noisy points close to the object surface.

3.5.2. 3D points density

The density of 3D points acquired is computed considering the ratio between the number of 3D points obtained after the editing procedure and the number of pixels that covers the object in the image. So, the number of 3D points per 100 pixels (N_p , N_a for passive and active, respectively) are reported in Table 3.4 for both techniques.

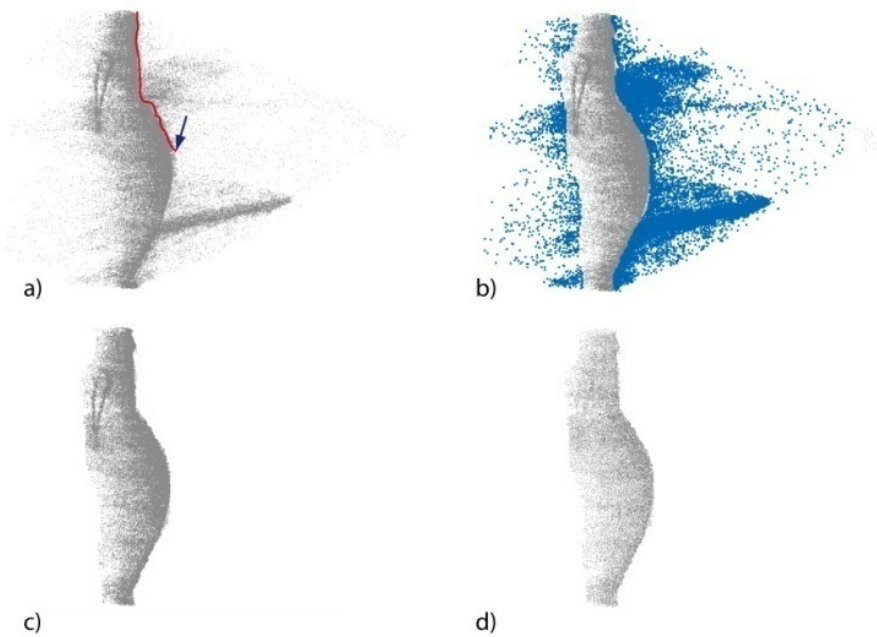


Figure 3.7. Point cloud cleaning procedure. Outlier point manual selection a), selected outlier points (b), cleaned point cloud after outlier deletion (c) and final point cloud after noise filtering (d).

| <i>Condition</i> | <i>Amphora</i> | | | | <i>Aeolus</i> | | | |
|------------------|----------------------|------------------------|----------------------|------------------------|----------------------|------------------------|----------------------|------------------------|
| | <i>N_p</i> | <i>N_{pc}%</i> | <i>N_a</i> | <i>N_{ac}%</i> | <i>N_p</i> | <i>N_{pc}%</i> | <i>N_a</i> | <i>N_{ac}%</i> |
| Air | 13.0 | 1.93 | 3.4 | 4.52 | 11.9 | 0.98 | 3.2 | 3.17 |
| CW | 14.3 | 2.79 | 3.9 | 17.50 | 13.8 | 0.94 | 3.5 | 22.61 |
| T1 | 14.2 | 4.23 | 4.1 | 23.96 | 3.8 | 3.53 | 3.2 | 24.93 |
| T2 | 2.9 | 6.35 | 3.3 | 34.58 | 2.8 | 9.93 | 2.9 | 20.50 |
| T3 | 3.8 | 5.07 | 3.5 | 30.63 | 2.5 | 3.21 | 1.7 | 42.81 |
| T4 | 0.2 | 20.42 | 2.2 | 47.99 | - | - | - | - |

Table 3.4. Acquired 3D points per 100 pixel (N_p for passive and N_a for active stereo) and percentage values of deleted 3D points ($N_{pc}\%$ for passive $N_{ac}\%$ for active stereo) for Amphora and Aeolus.

By comparing N_p and N_a in air, we found that the active technique has a lower point density than a factor 3, because it depends on the projector resolution, 800 x 600 in our case, while in the passive case depends on the number of pixels of the object in the image. However, from medium turbidity T2 this difference is reduced because the passive technique suffers more the lighting conditions of the scene. In fact, as we can see in Table 3.4, for passive stereo the increase of turbidity causes a drastic decrease of

acquired points, while for active stereo the reduction of points is less noticeable. By analysing the acquired point clouds (Fig. 3.8) we can point out that the active technique is able to reconstruct the entire surface of the object in each condition, while in the passive 3D reconstructions the point density is variable on the surface and any portions are not reconstructed.

Moreover, it is important to analyse the percentage of points deleted during the preliminary editing procedure. The percentage values of deleted points ($N_{pc\%}$ and $N_{ac\%}$) reported in Table 3.4 show that the active stereo technique is more affected by scattering even in low turbidity condition. Therefore, as mentioned before a time-consuming process is needed to delete manually 3D points before using the noise filter that works only on points very close to the object's surface. The passive case requires a lower noise reduction, but the scattering effect causes a loss of image contrast that compromises the matching of correspondent points and may generate holes in the reconstructed point clouds.

3.5.3. Multi-channel analysis

To improve the quality of the 3D models we have studied 3D reconstruction in each colour channel. In fact, contrast of acquired images can be very different in the three colour channels depending on the illumination and the object texture. So it is useful to understand if the channel separation of light performs a spectral discrimination of component noise. The acquired RGB images used in the 3D reconstruction have been separated in the Red, Green and Blue channels and then elaborated. In the passive case significant differences have not been noted, while in active case substantial differences in the point density have been registered. As we can see in Fig.3.9 the outlier points are highly reduced using the Green channel. The percentage of deleted points at turbidity level T1 has been reduced by about 66%, hence limiting manual operation of cleaning. Table 3.5 summarizes the results for the object Amphora at each turbidity level, showing that the improvement ranges from about 40% (turbidity level T4) to about 70% (turbidity level T2).

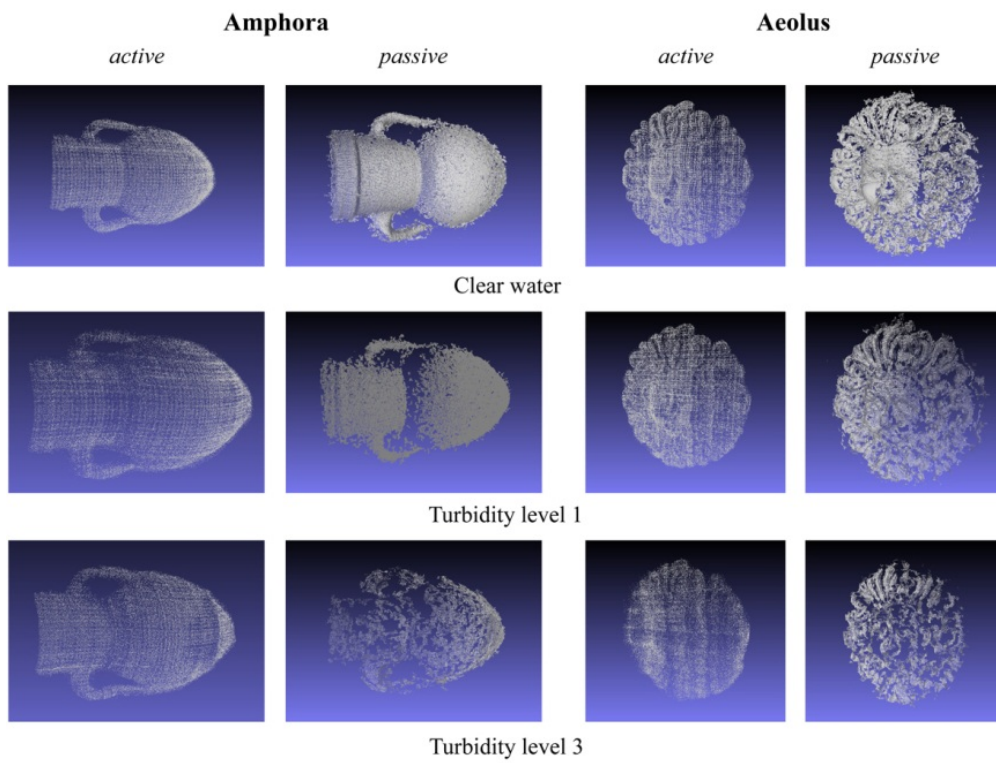


Figure 3.8. Point clouds obtained in different environment condition for both active and passive technique.

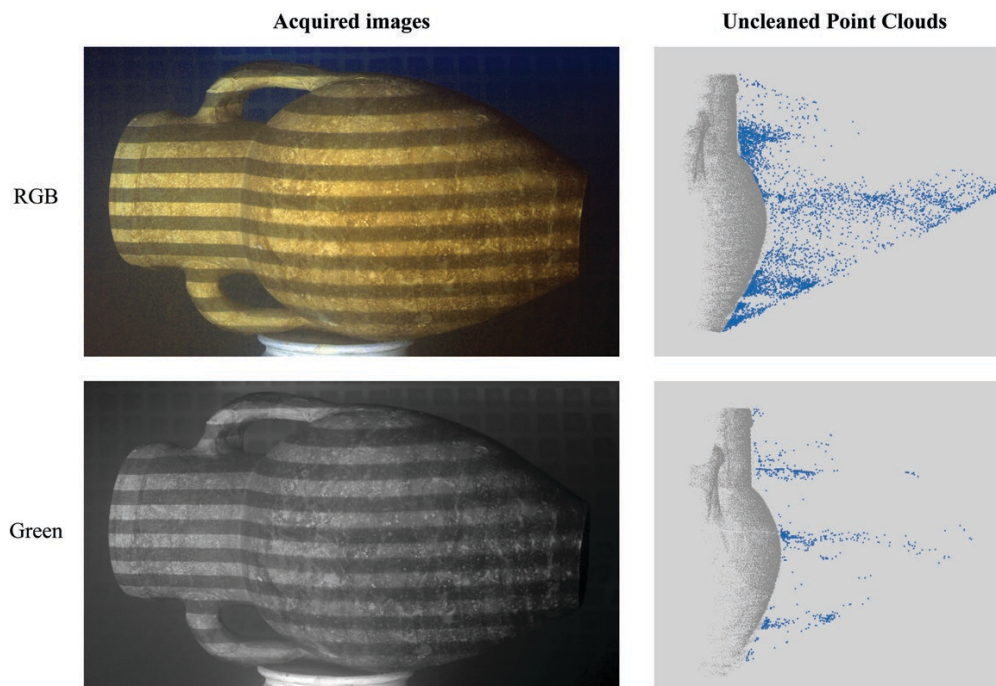


Figure 3.9. Acquired images and unclean point clouds at a light turbidity level (T1) in each colour channel: the blue points are due to scattering effect, highly reduced in green channel.

| <i>Condition</i> | <i>Na</i> | <i>Na</i> | <i>Na</i> | <i>Na</i> | <i>Nac %</i> | <i>Nac</i> | <i>Nac</i> | <i>Nac</i> |
|------------------|------------|-----------|-----------|-----------|--------------|------------|------------|------------|
| | <i>RGB</i> | <i>R</i> | <i>G</i> | <i>B</i> | <i>RGB</i> | <i>%-R</i> | <i>%-G</i> | <i>%-B</i> |
| Air | 3.4 | 3.3 | 3.3 | 2.8 | 4.52 | 3.49 | 2.67 | 17.02 |
| CW | 3.9 | 3.8 | 4.4 | 4.0 | 17.50 | 12.33 | 6.45 | 11.72 |
| T1 | 4.1 | 4.4 | 5.2 | 3.9 | 23.96 | 11.33 | 8.09 | 16.09 |
| T2 | 3.3 | 3.0 | 2.4 | 2.0 | 34.58 | 15.84 | 9.26 | 16.11 |
| T3 | 3.5 | 3.6 | 2.3 | 1.5 | 30.63 | 23.55 | 16.56 | 36.75 |
| T4 | 2.2 | 2.1 | 1.2 | 0.5 | 47.99 | 29.03 | 28.10 | 58.56 |

Table 3.5. Percentage values of deleted 3D points for active stereo technique in each colour channel, red (Nac%-R), green (Nac%-G) and blue (Nac%-B), compared respect the values computed from RGB images (Nac%-RGB). The table is related to the object Amphora.

3.5.4. Geometrical error

Finally, using the previous results obtained by the multi-channel analysis, we have evaluated on the Green channel, the geometrical errors for each technique separately, comparing the point clouds reconstructed in water with the one obtained in air. The geometrical error of the acquired 3D point clouds is expressed in terms of mean value and standard deviation of the discards of points (Tables 3.6 and 3.7). There is a slight increase in standard deviation in presence of low and medium turbidity, and with the same environmental conditions, the discriminant value is the material properties: the Aeolus mask shows a higher standard deviation due to its dark texture.

| <i>Condition</i> | <i>Amphora</i> | | <i>Aeolus</i> | |
|------------------|----------------|------------------|---------------|------------------|
| | μ (mm) | σ (mm) | μ (mm) | σ (mm) |
| CW | 0.25 | 0.35 | 0.75 | 0.81 |
| T1 | 0.44 | 0.23 | 1.61 | 1.26 |
| T2 | 0.47 | 0.23 | 1.98 | 1.56 |
| T3 | 0.75 | 0.66 | 2.32 | 1.98 |
| T4 | 0.86 | 0.88 | 2.83 | 2.81 |

Table 3.6. Geometrical error (mean μ and standard deviation σ) calculated for active technique in the green channel: the point clouds are compared respect to a reference (Ref.).

| <i>Condition</i> | <i>Amphora</i> | | <i>Aeolus</i> | |
|------------------|----------------|------------------|---------------|------------------|
| | μ (mm) | σ (mm) | μ (mm) | σ (mm) |
| CW | 0.22 | 0.55 | 0.70 | 1.06 |
| T1 | 0.52 | 0.65 | 1.53 | 0.90 |
| T2 | 0.69 | 0.78 | 2.08 | 1.49 |
| T3 | 0.85 | 0.87 | 2.44 | 2.00 |
| T4 | 2.23 | 1.86 | - | - |

Table 3.7. Geometrical error (mean μ and standard deviation σ) calculated for passive technique in the green channel: the point clouds are compared respect to a reference (Ref.).

3.6. Discussion

In the light of our experimentation, it is apparent that the active technique is not suitable at the moment for real applications because of the large number of patterns projected and the long acquisition time. So we think its application with a stable platform with any improvements regarding the reduction of number of patterns and the powerful of light source. However, a series of possible solutions to encourage the application in real cases can be proposed as follows:

- using fringe projection technique to reduce the number of patterns from 50 to 1;
- projecting colored patterns to increase the image contrast of acquired pattern;
- projecting a fringe pattern using a grating fixed on a flash lamp handled by a scuba diver;
- employing a more powerful light source.

On the contrary, the passive technique is ready to be tested in real cases, installing the stereo system on an underwater robot or handled by a scuba diver, simply synchronizing the shoot between cameras. Moreover, it offers the opportunity to be employed in stereo or in multi-view configuration using just one camera and taking a sequence of overlapped pictures. For the reasons explicated above we have decide to apply the passive technique in the acquisition of a real case and postpone the application of the active technique to future works.

3.7. Concluding Remarks

In this chapter active and passive stereo techniques has been compared using a compact and relatively low-cost 3D optical system. Experimentation has been conducted in a water tank with different values of clay concentration to simulate the bad visibility conditions that can occur in a marine environment. A system for 3D underwater acquisition in stereo configuration has been developed, composed by two cameras and a projector lodged in the waterproof housing. For an even comparison the experimental setup does not change during the acquisition process for both the techniques. The results of calibration show that a linear model can be employed in 3D underwater acquisition, without taking into account the refractive index of media with an acceptable accuracy.

The point clouds obtained with the active technique gives more stable results due to the use of coded patterns to solve the correspondence problem, but they are affected by a noticeable noise level due to scattering even in low turbidity, thus requiring more manual operations to clean the point cloud. To reduce this problem has been investigated the use of a single colour channel for the 3D reconstruction, finding that acquiring underwater images in the Green channel only can give good improvements in term of noise reduction and surface accuracy. The passive technique gives a cleaner and denser point clouds in clear water and at low turbidity level, but the results are strongly related to the surface texture of the acquired object. In fact, in low textured areas, as the turbidity increases, the loss of contrast due to turbidity puts in evidence holes and missing areas. Another difference concerns the acquisition time: the active technique requires the projection of 50 patterns while passive stereo needs only an image pair.

The results of this experimentation demonstrate that two stereo algorithms widely employed in air applications for 3D reconstruction of close-range objects can be used for in water applications without carry any modifications, or take into account optical model of refraction. Some solution to apply the active technique in real cases has been suggested, reducing notably the number of patterns by the means of fringe projection technique.

Future developments of this Research will address the feasibility of extending the applicability of the technique in active marine application improving the 3D acquisition integrating both passive and active techniques.

Experimentation of multi-view 3D reconstruction techniques in underwater environment

3D imaging techniques based on multi-view capture offer a valid and effective tool in underwater applications because they require capturing the scene by different viewpoints using just one still or movie camera. A technique for 3D reconstructions from multiple images has been tested on the submerged structures in the underwater archaeological site of Baiae (Italy). This case study is particularly critical because of the high level of turbidity, the heavy presence of marine flora and the changing of lighting conditions typical of shallow water.

4.1. Introduction

Many underwater applications require to gather 3D data of flora, fauna or submerged structures, for different tasks like monitoring, analysis, dissemination or inspection. In particular, in the Cultural Heritage field, the reconstruction of submerged finds or entire archaeological sites is of great interest, otherwise they could be unexplored and unknown to researchers and lovers because located in inaccessible and hostile environments. Moreover, according to the guidelines of UNESCO that suggest the preservation *in situ* of underwater heritages (Unesco 2001), techniques and related tools for 3D reconstructions are widely employed by underwater archaeology in the last decade. Among the 3D techniques suitable for underwater applications, the photogrammetry represents a valid solution to reconstruct 3D scene from a set of images taken from different view-points. The acquisition devices (still or movie camera with the proper water-proof housings) are very affordable and easy to use if compared to dedicated devices like LIDAR, Multi-beam, etc. and they can be installed on underwater robots or handled by scuba divers.

In this context, in this chapter will be proposed a photogrammetric technique based on multiple captures to reconstruct the 3D model of submerged structures in the shallow

archaeological site of Baiae (Italy) that is particularly interesting because covers a submerged area of many hectares and presents a wide range of different architectural structures, with few decorations still preserved. This site is the subject of study of the COMAS project (Comas 2012) that aims to develop new methodologies and tools for conservation and restoration of submerged archaeological artifacts. Thus, the project includes a preliminary 3D acquisition in order to document the conditions of architectural structures and to organize the subsequent operations of restoration and conservation. In fact, the structures are covered by aquatic flora and fouling organisms that root and deteriorate the materials. The lighting conditions of the submarine environment in which the site is immersed are characterized by high level of turbidity due the suspended particles and caustic effects that are present typically in shallow water (maximum depth about 14 meters).

As preliminary test the experimentation involves the acquisition of a portion of the site, focusing the attention on a room of the so-called "Villa Protiro". This case study allowed to test algorithms for 3D reconstruction from a sequence of overlapped images, widely used in air applications, in a very critical condition due to the presence of strong turbidity and various disturbance of the scene like caustic effects, algae and fishes.

The chapter is organized as follows. Section 4.2 presents the related works about underwater 3D reconstructions. In Section 4.3 the archaeological site of Baiae is briefly described. Section 4.4 is dedicated to the 3D reconstruction and finally conclusions are presented in Section 4.5.

4.2. Related works

In underwater archaeology the acoustic and optical technologies are used for mapping, survey and 3D reconstruction of finds, artifacts and submerged sites (Singh *et al.* 2010). In particular, among the optical techniques available for underwater applications the photogrammetry has taken an important role (Telem & Filin 2010) because its easy applicability that requires just one or more digital cameras, also in stereo configuration, so such acquisition device can be installed on submarine robots or handled by scuba. Moreover, the algorithms for 3D reconstruction by photogrammetry are widely

consolidated in air applications, and sometimes with some modifications employed in water.

Underwater photogrammetric approach has been used to survey the archaeological site of Pianosa, a Hellenistic shipwreck, for example (Drap *et al.* 2002; Georgopoulos *et al.* 2011). A series of control points are dispersed in the scene and acquired by a camera handled by a diver in order to establish the correspondence among a sequence of overlapped images. In case of deep water like for the survey of Grand Ribaud Etruscan Wreck (Drap *et al.* 2002), a diver cannot stay long time and the mapping of find under those conditions would be near to impossible. Thus, a light digital photogrammetry method is employed by using a non-metric digital camera, mounted in a waterproof housing attached to a bar on a submarine robot. A preliminary procedure of calibration must be performed to calculate the optical parameters of the camera and to consider the effect of refraction at the air-acrylic and acrylic-water interfaces, which are present when a camera is mounted in its housing. The refraction can be considered through the explicit modelling of the optical paths using ray tracing (Li *et al.* 1997) or with the implicit incorporation of refraction effect into camera calibration parameters (Schewe *et al.* 1996b).

In (Meline *et al.* 2010) using an uncalibrated single camera a method has been proposed to calibrate this camera integrating variation of the water optical refractive index to the distortion model, for 3D underwater reconstruction of archaeological objects. An interesting application of underwater 3D reconstructions for the Mazotos wreck by photogrammetry is proposed in (Skarlatosa *et al.* 2012). Using a digital camera the capture images have been elaborated by using open source algorithms (PMVS and Blundler) widely used in air for multi-view 3D reconstruction that have demonstrate their effectiveness in underwater application compared with a structured-light technique (Bianco *et al.* 2011), thus without take into account the optical refraction model. The site of Mazotos lies in deep water and the turbidity is very low, so factors of disturbance like turbidity, variation light condition, and presence of marine flora have not been considered in the experimentation.

4.3. Archaeological context

The underwater archaeological site of Baiae is located few kilometres north of Naples (Italy) (Fig.4.1) and belongs to the imposing volcanic complex of the Phlegrean Fields. Baiae was a famous seaside town much prized in Antiquity for its temperate climate, beautiful setting and the properties of its mineral waters which have been exploited since the second century BC. It was the most popular resort of Roman Aristocracy and the Imperial Family up to the end of the third century AD, when the bradyséisme caused the submersion of the city. Now the remains of villas lie underwater along the seaside to a distance of 400-500 meters. Next to luxurious maritime villas and imperial buildings it is possible to find more modest residences, private *thermae*, the so-called *tabernae* and all those structures that characterize the Roman age cities.

In the past years the ISCR with the project “Restoring Underwater” has restored some sectors of ancient villas and structures situated in the Park of Baiae (Pietraggi & Mancinelli 2004; Pietraggi & Davidde 2007): the Villa con ingresso a Protiro, the Villa dei Pisoni, the Via Herculanea and the Edificio con cortile porticato at Portus Iulius. In particular in the Villa con ingresso a Protiro in the 2003 and in 2009 were restored a room paved with a white mosaic and a thermal room (*calidarium*) were this 3D project has been realized. The Villa con ingresso a protiro is located at 5 meters depth; the rooms which composed the Villa extend for 40 meters on the road flanked by *thermae*, *tabernae* and other villas; however, its real size must have been larger. The threshold – framed in stucco pilasters – is delimited by two red-plastered masonry benches. The name ‘*prothyrum*’ comes from the presence of two stuccoed columns shafts – no longer existing – that were placed on two short parting walls built in front of the threshold. The villa’s floors and walls are largely decorated with marbles; almost all the surveyed rooms have marble-covered walls, and mosaic floors mostly made of white tesserae (e.g.: room no restored by ISCR in 2003). A white and black mosaic patterned with hexagons, rosette diamonds and plants – 2nd century ad – decorated a room on the North-East of the entry hall; this entry hall had a sheets wainscot in red marble of Tenario with ‘Lunense’ marble mouldings. The villa underwent a second construction phase represented by the building of a wide apsidal room south of the hall. The room had two floor levels, built from large white marble sheets which were used as covers for

the walls as well. The villa is also composed of other service rooms like an area identified as a kitchen in the north-western section, and a rectangular courtyard. A garden was located behind these areas, it contained the aforementioned Alkamenes Afrodite in kepois statue. (*l'Afrodite dei giardini*) recovered in 1993 and now displayed in the garden of Campi Flegrei Archaeological Museum.

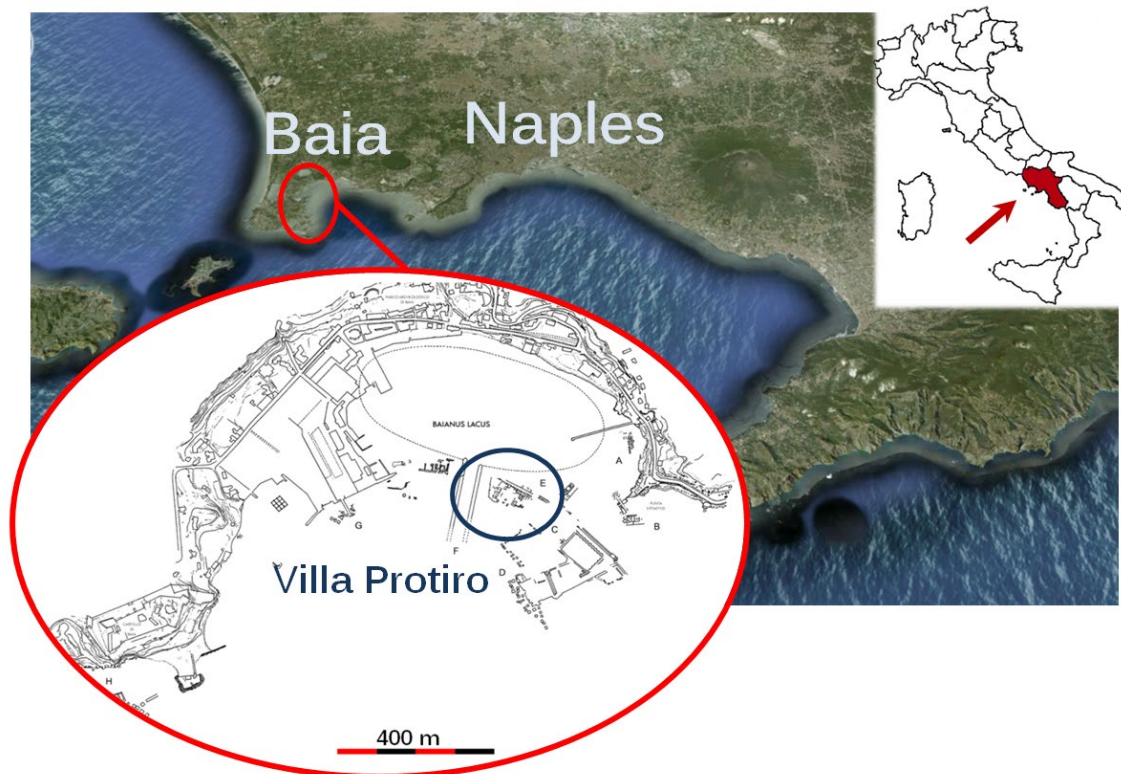


Figure 4.1. Geographical localization of underwater archaeological site of Baia (Italy).

4.4. Experimentation

As mentioned in the previous sections, the archaeological site of Baiae presents several peculiarities compared to other underwater sites. In fact, it well know that the Baiae environment conditions, in terms of water turbidity and presence of flora, are very critical and it is not possible to be sure about the possibility to use a multi-view photogrammetric technique for obtaining a sufficiently detailed model of the structures present in the site.

For this reason a first feasibility study has been performed to test a 3D reconstruction technique available in literature to acquire a portion of the site elected as a case-study. Thus, the main goal of this study was to verify if a well-known 3D reconstruction technique based on multi-view stereo may be employed in the critical conditions of the Baiae site.

The room selected as a case-study is a portion of “Villa Protiro” sized 5 x 8 meters. The room is delimited by a wall with an elevation ranging around 0.5 meters. There is one aperture on the west side and it presents several blocks of bricks spread over the floor. This particular room has been selected because it presents various materials (bricks, mortar, tile floors, etc.) and, at the moment of the acquisition, it was affected by a strong colonization of algae and other kind of bio-fouling, so constituting a very hard test case for the 3D acquisition techniques.

4.4.1. Experimental setup

The equipment used to acquire the underwater pictures consists in a Nikon D7000 reflex camera with a 20 mm fixed lens. The camera is mounted in an underwater case manufactured by Ikelite equipped with a spherical housing. In addition two underwater strobes have been used to acquire pictures of small portion of the site that, for the presence of shadows, are in very low lighting conditions. A second compact camera, precisely a Canon S100, mounted in a proprietary underwater housing, has been used by a second diver to record a series of full-HD videos in order to have redundancy of data reducing the risk of uncovered areas.

4.4.2. Image acquisition

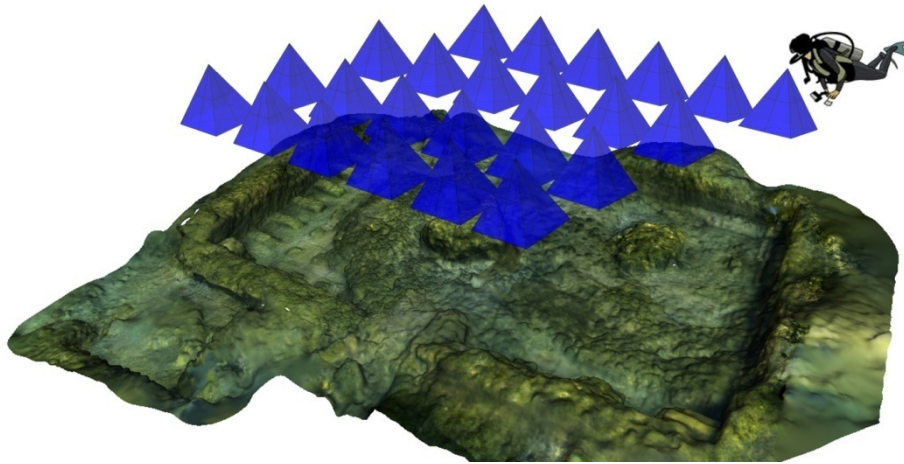
The acquisition has been carried out in two different dives, related to the North part and to the South part of the submerged area. A third dive has been planned in order to complete the missed areas. The manual mode has been used during the acquisition, setting values between 1/250 and 1/320 of a second for the exposure time and an aperture between f-5.6 and f-8 to ensure a sufficient depth of field. The automatic ISO sensitivity function has been used in order to ensure that exposure adapts to variable

lighting conditions, taking advantage of the recent performance of the D7000 imaging sensor, capable of producing images with no noise even at high ISO. In order to test the effect on the 3D reconstruction, a set of images has been acquired using the pre-measured value obtained acquiring an underwater white balance panel *Lastolite* (Fig. 4.2).

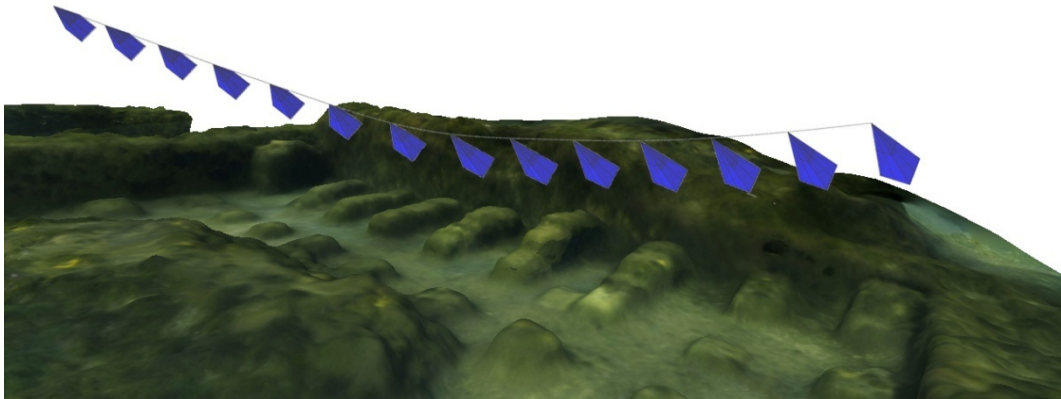


Figure 4.2. Calibration panel used to manually correct the white balance.

Following a standard aerial photography layout, the diver swims at a certain height/depth above the site (about 2 meter in this particular experimentation) taking overlapping pictures ensuring about a 75% of common coverage between two consecutive photos, following a straight line with constant speed in North-South direction (Fig. 4.3-a). Several overlapping stripes have been acquired covering the whole area. Another set of stripes have been acquired in East-West direction. The occluded areas like walls and all the features not visible in vertical pictures, have been acquired a series of oblique photographs (Fig.4.3-b). In total about 700 images have been gathered.



a)



b)

Figure 4.3. Image acquisition using a standard aerial photography layout (a) and using oblique photographs (b).

4.4.3. Multi-view 3D reconstruction

The 3D reconstruction pipeline starts from the orientation of the 722 pictures composing the whole dataset by using Bundler (Snavely *et al.* 2007). The first phase uses the SIFT (Scale Invariant Feature Transform) operator (Lowe 2004) to extract a set of descriptors across the collection of images that are subsequently matched. The output of Bundler consists in the intrinsic and extrinsic parameters for each camera pose and in a sparse reconstruction of the scene.

The full set of images has not been successfully oriented, and it has been necessary to divide the dataset into two groups, for the North and South part, that have been reconstructed separately (Fig.4.4). In particular 384 images have been oriented for the North dataset and 116 for the South dataset.

This is mainly caused by the sandy seabed present in the central part of the room, which makes difficult the extraction and matching of feature due to the low contrast. Furthermore, the lack of overlapping areas in the reconstructed surface has made it impossible to align the two blocks.

Subsequently for each block, the Bundler's outputs and the undistorted images have been processed with PMVS2 (Patch-based Multi-View Stereo) algorithm (Furukawa & Ponce 2007) to create a dense point cloud. This algorithm estimates the surface orientation while enforcing the local photometric consistency, which is important to obtain accurate models for low textured objects or with images affected by blur due to turbidity in underwater environment. Furthermore PMVS2 automatically reject moving objects, such as fishes and algae.

Two 3D point clouds have been generated, for the North (fig. 4.5) and South (fig. 4.6) area respectively. The North area has been reconstructed with about 3 million points and 7 million triangles, while about 1.5 million points and 3 million of triangles have been reconstructed for the South area.

Since there is no common area between the two blocks, it is impossible to register the point cloud by picking pairs of corresponding points, making not possible to generate a complete model of the site. Moreover, the use of uncorrected pictures produces a low-quality texture making the model unusable for interactive visualization purposes.

Reconstructed surfaces and textured models are reported in fig. 4.7 and fig. 4.8.

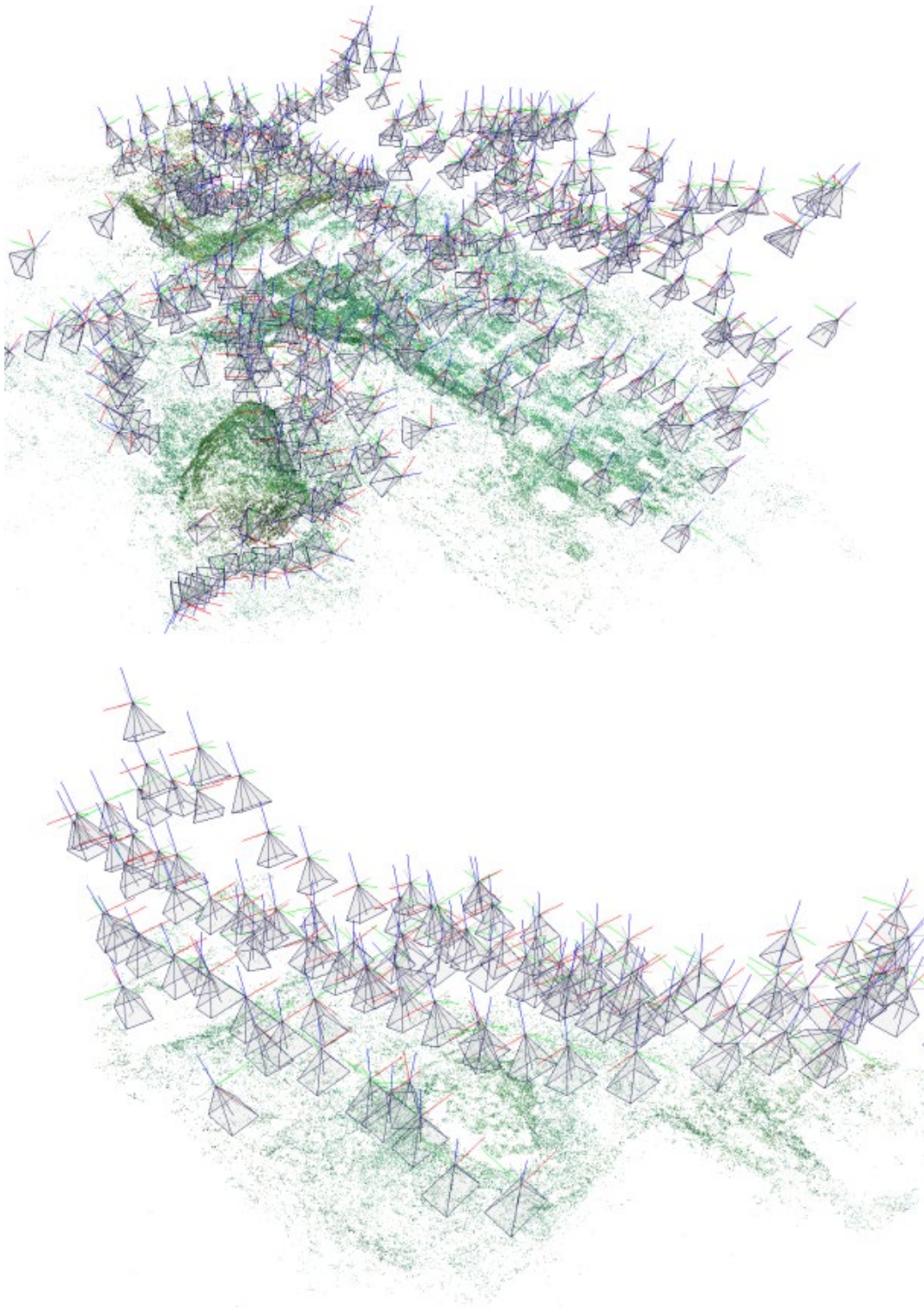


Figure 4.4 Results of the camera orientation process on original uncorrected pictures: due to the presence of a low contrasted sandy seabed, the whole scene has been reconstructed in two separated blocks relative to the North part (up, 384 pictures) and to the South part (down, 116 pictures).

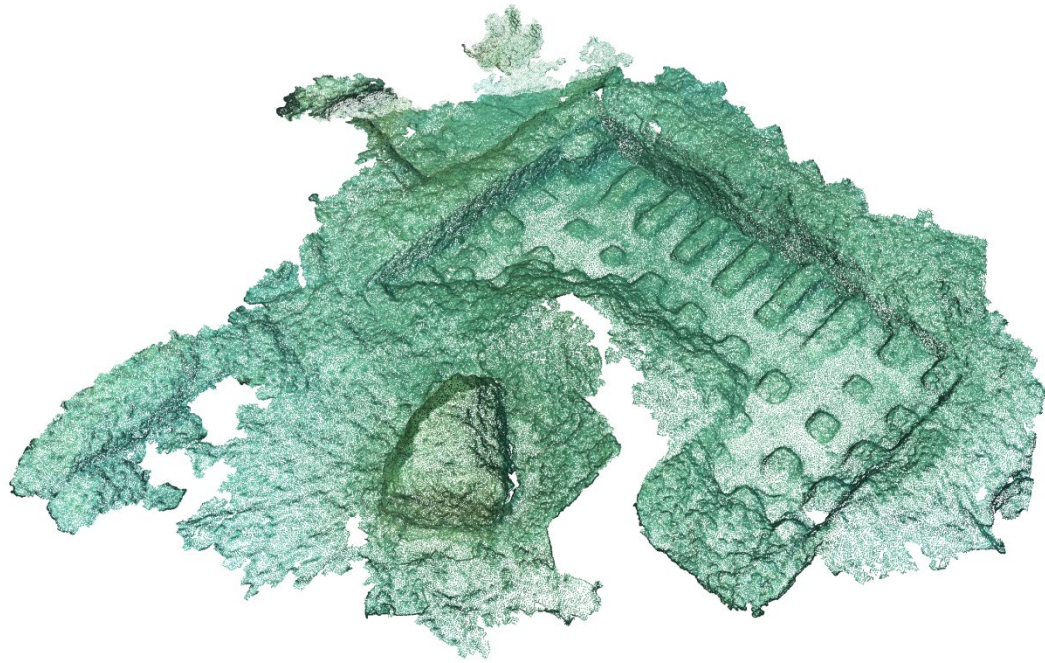


Figure 4.5. Reconstructed point cloud of the North area.

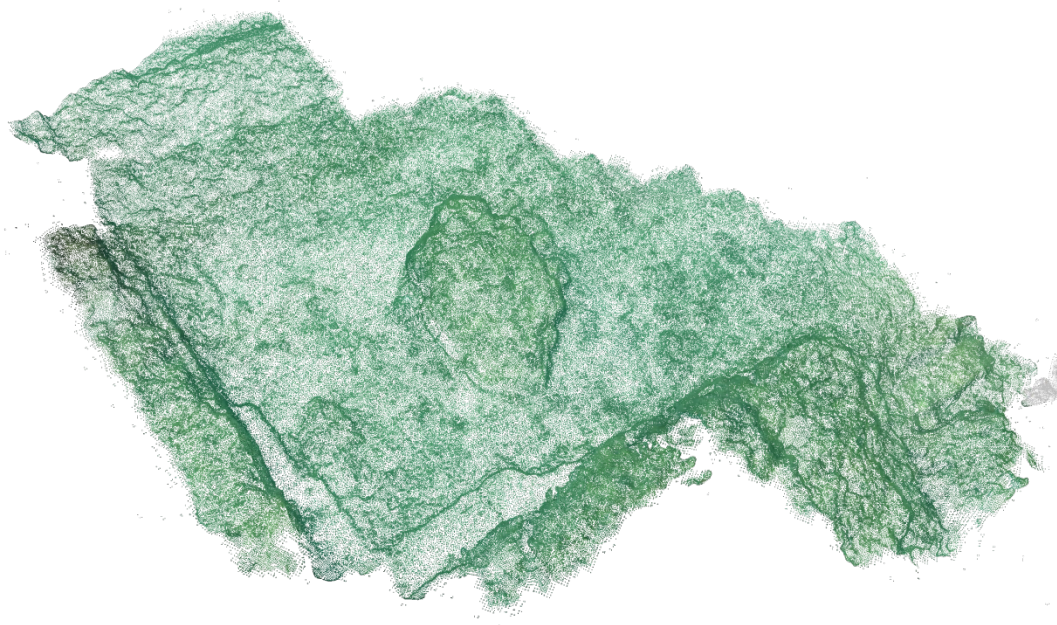


Figure 4.6. Reconstructed point cloud of the South area.

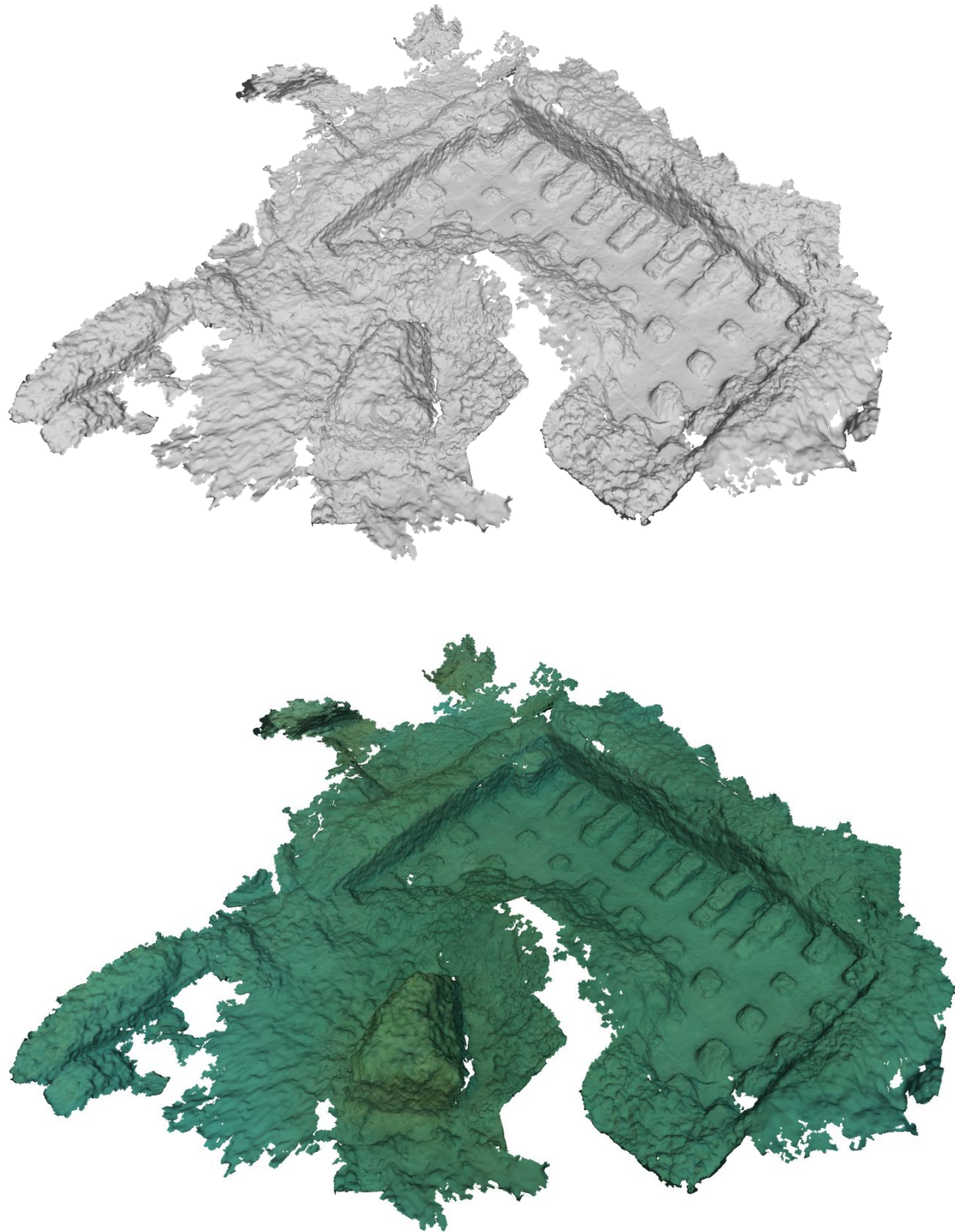


Figure 4.7. Reconstructed surface and textured model of the North area.

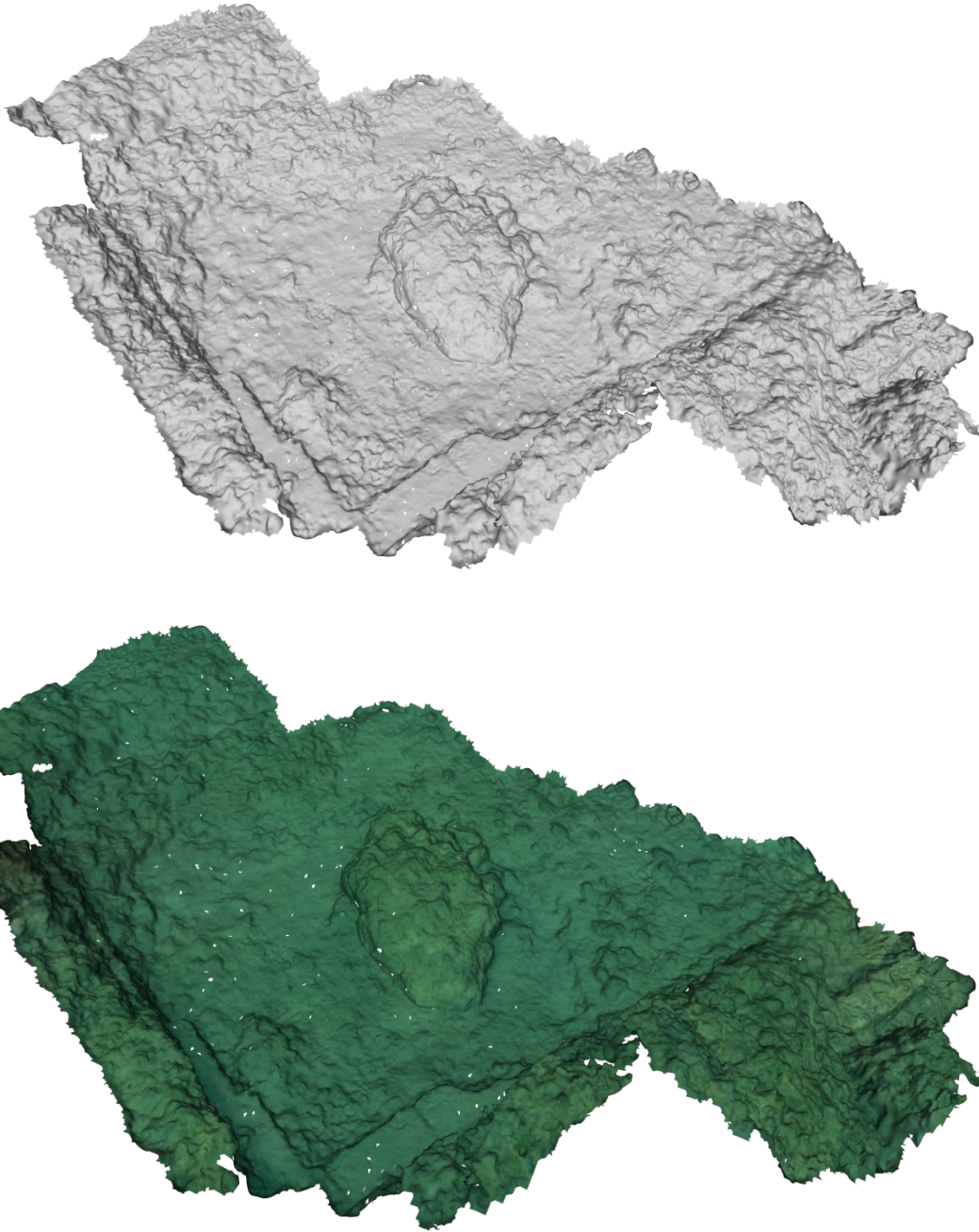


Figure 4.8. Reconstructed surface and textured model of the South area.

4.5. Conclusions

In this chapter has been presented a preliminary experimentation of 3D reconstructions in the underwater archaeological site of Baiae (Italy), useful for the creation of a virtual scenario of a site portion for fruition and analysis purposes. A limited area of the site (Villa Protiro room) has been acquired and used to experiment the performances of a multi-view 3D reconstruction algorithm commonly used in air.

The lighting conditions of this site are characterized by high level of turbidity, heavy presence of marine flora and changing of lighting conditions typical of shallow water, which makes the site chosen for the experimentation very challenging.

The results have been shown how the environment affects the 3D reconstruction. In particular has not been possible to orient the whole dataset and build a complete model of the submerged site. The loss of contrast due to turbidity worsen the performances of image matching algorithms, making it necessary to develop specific tools to increase the quality of the 3D reconstruction.

Underwater image pre-processing algorithms for the improvement of multi-view 3D reconstruction

The results described in the previous chapter of the thesis have demonstrated how the effects of underwater environment, such as bad visibility and moving marine organisms, affect the quality of the 3D reconstruction. A specific methodology must be used to improve the results of the 3D reconstruction. In this chapter the DOE (Design of Experiments) has been used to find the factors that influence the reconstruction accuracy the most. A statistical approach has been used to test the impacts of underwater colour enhancing algorithms, demonstrating how a specific pre-processing stage can improve considerably the quality of the 3D reconstruction.

5.1. Introduction

In the previous chapter the preliminary survey conducted on the thermal room in Villa Protiro at the underwater archaeological site of Baiae (Naples, Italy) has been presented. The results shown that the marine environment negatively affects the quality of the reconstruction. In particular, the original images did not allow to make a complete model of the area. Image-based acquisition suffers the critical environment conditions that depend on several factors like depth, flora, weather and geographical position, that reduce the visibility and alter the colours. These problems limit its application to close range applications and force the operators to adopt algorithms in order to improve the image quality (Schettini & Corchs 2000).

In the experimentation described in this chapter, two existing colour enhancement methods, ACE (Automatic Colour Enhancement) and PCA (Principal Component Analysis), have been compared with a new method based on Histogram stretching and manual retouching. These algorithms have been tested on the images acquired in the real environment of Baiae.

A new methodology for the 3D reconstruction in underwater environment based on the enhancement of underwater pictures will be proposed herein. The enhancement of underwater images is necessary not only to improve the quality of the 3D reconstruction, but also to create a more realistic texture, which is an essential part of an high quality 3D model. Unfortunately, often the images enhanced for texturing purposes cannot be used for the creation of an accurate 3D model.

A texture showing the “true” appearance of the scene can be obtained using photo editing tools already present in the state of the art, but for the generation of a high quality model a more detailed and rigorous analysis is required.

For this purpose, an experimental campaign has been planned using the DOE (Design of Experiment) (Montgomery 2009), in order to identify the factors that influence the reconstruction accuracy the most. In particular the camera orientation and calibration phase have been analysed, using a SFM algorithm that works on the points of interest evaluated by a feature extraction and matching algorithm. The results of this step are related to the image quality, for this reason the effect of image pre-processing has been investigated through a statistical approach.

The first step is the measurement of the parameters of interest, such as the number of extracted and matched features and the number of oriented cameras. The subsequent step is the selection of the influencing factors that cause a variation in the parameters, such as, for example, the methodology of colour enhancement used to improve the underwater images and the colour channel. The collected data have been processed with an analysis of variance technique (*anova*) and classical statistical tools.

The purpose of this analysis is to evaluate, on a statistical basis, if a factor or a combination of factors have an influence on a parameter with a certain confidence level, in order to find the best combination of factors which will be used to create the 3D reconstruction of the submerged site.

The chapter is organized as follows. Section 5.2 presents the related works about underwater image pre-processing methods. Section 5.3 deals with the image pre-processing stage. Section 5.4 describes the adopted methodology and in section 5.5 the parameters used in the statistical analysis are described. Section 5.6 presents the 3D

reconstruction obtained using enhanced images and finally conclusions are presented in Section 5.7.

5.2. State of the art of underwater image pre-processing algorithms

As it is well known, underwater photography is not an easy task, because the environmental conditions lead to several problems related to visibility, refraction and changing of lighting conditions. Underwater pictures generally suffer of light absorption, which causes some defects mostly on the red channel (that is the first component of the light spectrum that is absorbed), and this effect is already noticeable at few meters of depth. The pre-processing of underwater images can be conducted with two different approaches: image restoration techniques or image enhancement methods (Mahiddine *et al.* 2012; Schettini & Corchs 2000). Image restoration techniques need some known environment parameters such as scattering and attenuation coefficients, while image enhancement methods do not require an *a priori* knowledge of the environment.

The ACE (Automatic Colour Equalization) algorithm (Rizzi & Gatta 2004) is inspired by human visual system, which is able to adapt to highly variable lighting conditions, and extract visual information from the underwater environment (Chambah *et al.* 2004). The algorithm combines the Patch White algorithm with the Gray World algorithm, taking into account the spatial distribution of colour information. In the first part of the ACE method the chromatic data, the pixels are processed and adjusted according to the information contained in the image. Subsequently the colours in the output image are restored and enhanced (Petit 2010).

(Bazeille *et al.* 2006) proposed an automatic enhancement algorithm which does not require any correction parameter. Each source of errors is corrected sequentially. The first step removes the moiré effect, then a homomorphic or frequency filter is applied to make the brightness uniform and to enhance the contrast. Regarding the acquisition noise, a wavelet denoising filter followed by an anisotropic filtering has been applied. Finally, a dynamic expansion is applied to increase contrast followed by colour equalization. The process is performed on the component Y in YCbCr space to optimize the computation time. It is important to point out that the use of a homomorphic filter

affects the geometry, which results in errors on the reconstructed scene. (Iqbal *et al.* 2007) have used slide stretching algorithms both on RGB and HIS colour to enhance underwater images. First, a contrast stretching on RGB colour space has been performed, and then the result has been converted to HSI colour space where a saturation and intensity stretching is performed.

The aim of underwater colour correction is not only useful for obtaining better quality images, but also to improve the performance of feature extraction algorithms in terms of feature point's detection. In (Kalia *et al.* 2011) the effects of different image pre-processing methods on the performance of the SURF detector (Bay *et al.* 2008) have been investigated, and a new method - IACE (Image Adaptive Contrast Enhancement) has been proposed.

5.3. Colour enhancement of underwater images

In this experimentation, the original untouched images have been compared with the ones enhanced by using the ACE model by (Rizzi & Gatta 2004) and with an approach based on Principal Component Analysis (PCA) (Tonazzini *et al.* 2004). The ACE and PCA method have been implemented in Matlab and applied on the full dataset of RGB images.

For this particular study, a new enhancement methodology (hereinafter, HIST), which is not present in the state of the art, has been developed. This method is based on histogram stretching and on a manual colour retouching procedure performed on a sample image of the dataset, and it has been implemented using batch actions in Adobe Photoshop® to rescue the maximum amount of information from a set of defective and noisy pictures.

For this purpose, a representative image of the underwater scenario has been chosen as a sample in order to perform retouching and equalization, by applying a four-step procedure:

- preliminary histogram stretching to improve the contrast;
- mixing of the colour channels to balance the missing information on the red one;

- creation of a set of adjustment layers on the sample image, ranging from saturation enhancement for some missing hues, contrast masks, colour balancing and equalizing;
- export of the statistics of the sample image on the other ones, with different mixing percentages.

In figure 5.1, there is a comparison between an original uncorrected image and the ones enhanced with ACE, HIST and PCA methods.

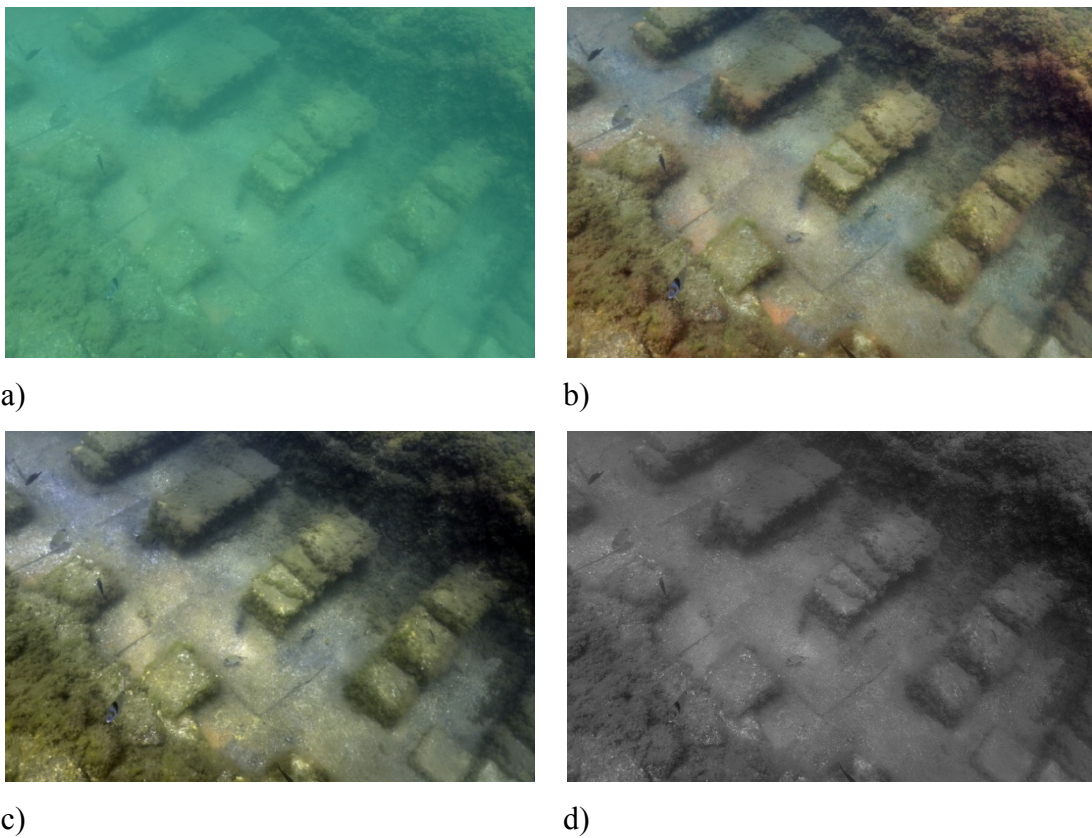


Figure 5.1. A sample image enhanced with the ACE method (b), HIST method (c) and PCA method (d) compared to the original image (a).

5.4. Description of the adopted methodology

The proposed methodology provides separate workflows for the 3D reconstruction and the texture creation steps (fig. 5.2), since often the images edited for texturing purposes, cannot be used for the creation of an accurate 3D model.

The texturing step can be addressed with common image editing tools, in order to generate a set of images that reproduce the correct colours and are characterized by a uniform brightness and colour balance.

The 3D reconstruction step is carried out using a multi-view 3D reconstruction technique with a set of underwater images, enhanced in order to improve the performances of the image matching algorithm. There are several methods in the state of the art that can be used for underwater image enhancement, and multiple parameters which have a relevant impact on the 3D reconstruction.

In order to identify how these factors affect the accuracy of the reconstructed model, it is necessary to perform a rigorous analysis.

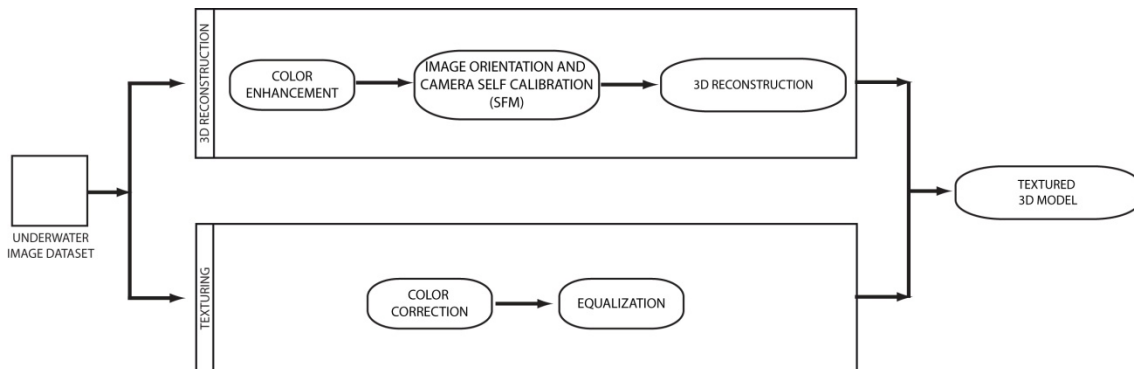


Figure 5.2. Pipeline of the implemented methodology for the reconstruction of submerged scenes.

5.4.1. Design of the experimental campaign

The dataset described in Chapter 4 (consisting in 722 images relative to a submerged area of about 40 square meters located at a depth of 5 meters) has been used to test how a colour enhancement algorithm can improve the reconstruction quality.

The experimental campaign has been planned using the DOE (Design of Experiment) (Montgomery 2009), in order to identify the most influential factors that affect the results of the 3D reconstruction. A particular attention has been paid about how the image enhancement methods affect the camera orientation and the self-calibration bundle adjustment process. The measured data have been analysed with statistical tools, in order to evaluate if a factor (or a combination of factors) has an influence on a parameter with a certain confidence level. From this analysis it is possible to evaluate the best combination of factors (such as the best colour channel and the best enhancement method) which should be used to make a highly accurate and dense 3D reconstruction by using a multi-view stereo technique.

The preliminary phase involved the identification of the factors that could have an influence on the quality of a 3D reconstruction in underwater environment performed with a multi-view technique. The following factors have been identified according to the experience:

- Visibility conditions
- Lighting
- Underwater flora and fauna
- Colour channel
- Image resolution
- Focal length
- Focus
- Distance from the subject
- Type of shooting (oblique photographs or aerial layout)
- Object type
- Method used to enhance the underwater images

Some of these factors cannot be controlled on the field, such as the presence of marine organisms in motion and the level of turbidity of the water, for instance. Other factors may also be considered as a direct consequence of the others. For example the setting of focus depends on the distance from the subject. Furthermore, it is almost impossible for a diver to keep a fixed distance from the object being photographed. With regard to the focal length, it can be set according to the extension of the area of acquisition and to the

working distance defined according to the level of turbidity and depth. In shallow water, for example, a wide angle lens is necessary to ensure a good coverage. For these reasons, we used a subset of the identified factors (Table 5.1).

| Influence factor | Level | Symbol |
|---------------------------|--------------------|---------------|
| Colour enhancement method | OR, HIST, ACE, PCA | EN |
| Image pyramid level | 1, 2, 3 | PYR |
| Colour channel | RGB, R, G, B | CH |
| Image set | SET 1,2,...,SET 7 | SET |

Table 5.1. Influence factors and related levels used to perform the analysis of variance.

The first factor is the colour enhancement algorithm used to improve underwater images (hereinafter, EN). As shown in (Mahiddine *et al.* 2012), the performance of a feature extraction and matching algorithm can vary significantly according to the method used to correct the images. In this experimentation, the original images were compared with the ones corrected by ACE (Automatic Colour Enhancement) method and with a specially developed enhancement methodology, based on the histogram stretch and on a manual correction of a sample picture (hereinafter, HIST). In addition, an analysis has been performed using the images enhanced through a method based on the PCA (Principal Component Analysis). Since this correction gives a single component, the comparison was made between each RGB channel and the corrected image with PCA.

The second factor is represented by the image resolution (hereinafter, PYR). An image pyramid has been created, so the levels of this factor correspond to the pyramid levels: at the first level (PYR = 1) the images are halved (4 times less pixels), at the second level (PYR = 2) the images are 4 time smaller (16 times less pixels) and at the third pyramid level (PYR = 3) the images contain 64 time less pixels.

The third factor is represented by the colour channel used for the 3D reconstruction. The levels of this factor are the three components R (Red), G (Green) and B (Blue) and the composite RGB image.

Seven subsets containing a different number of images belonging to different regions of the scanned area have been extracted from the complete dataset, made of 722 pictures.

In particular, the first (21 images, an example in fig. 5.3-a) and the second (20 images, and example in fig. 5.3-b) set are related to the northern part of the area, in which the remains of the supporting structures of the floor of the thermal room are clearly visible. The photos have been taken mostly from the top, following a standard aerial layout; the difference between the two sets consists in the fact that the second set has been acquired after that cleaning operations have been conducted on the floor.

The third set (fig. 5.3-c) includes 9 pictures taken from the top, covering a small part on the north-west wall. The images are characterized by high overlap and good visibility due to the reduced distance of the camera from the submerged structures.

The fourth and fifth sets (fig. 5.3-d and fig. 5.3-e) were shot with a Canon s100, using custom white balance. The two sets, which are composed by 27 and 28 pictures respectively, are related to the north-west of the room, and cover the outside masonry structures of the outer walls.

The sixth set (fig. 5.3-f) includes 70 oblique photographs taken with a Nikon d7000 covering the whole south area. The last set (fig. 5.3-g), is a sequence of 21 oblique overlapping pictures taken travelling along the west wall from north to south. The difference between the two sets consists in the working distance, which is higher for the set number 6 because it was necessary to acquire a larger area.

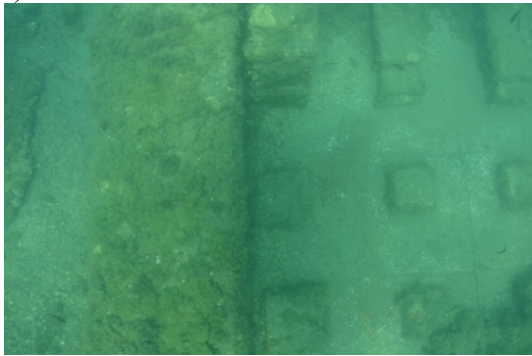
In order to perform a quantitative analysis of the influence of the shot type on the processing results, the type of image set (hereinafter, SET) was included as an influencing factor in the variational analysis. In particular, two different experiments were carried out. The first one included sets 1,2,3,6 and 7 (images taken with the Nikon d7000) and the second one included sets 4 and 5 (images taken with a Canon s100 using the custom white balance setting for underwater photography)



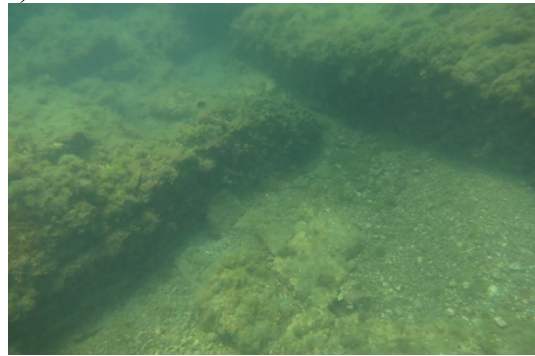
a)



b)



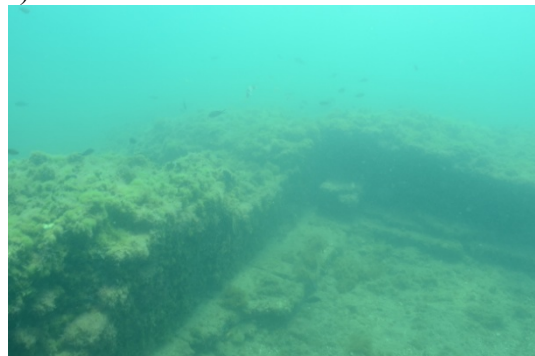
c)



d)



e)



f)



g)

Figure 5.3. Sample images related to image set 1 (a), image set 2 (b), image set 3 (c), image set 4 (d), image set 5 (e), image set 6 (f) and image set 7 (g).

The mean number of extracted features from each image and the percentage of matched feature within each set were important parameters to be measured. Furthermore, in order to evaluate the goodness of the features to be used in the computation of both the scene structure and the camera calibration parameters, the percentage of oriented cameras in each set and the mean bundle adjustment reprojection error at the end of the SFM process have been taken into account.

The first parameter, which stands for the number of extracted features, has been evaluated by using the SIFT (Scale Invariant Feature Transform) (Lowe 2004) operator in order to extract a set of descriptors across the collection of images that are subsequently matched. Since each set contains a different number of images, the data used in the statistical analysis is represented by the mean number of extracted features, evaluated as:

$$\text{mean number of features}_{EN,PYR,CH,SET} = \frac{\text{number of features}_{EN,PYR,CH,SET}}{\text{number of images}_{SET}}$$

In which the *number of features* stands for the total number of features extracted for each configuration, and *number of images* is the number of images included in each set at each configuration.

The other parameters have been evaluated using Bundler (Snavely *et al.* 2007), a structure-from-motion system that can handle unordered image collections. The system works on a set of images, image features (computes using SIFT), and image matches (computed by matching SIFT features), and produces the set of oriented cameras (and camera calibration parameters) and (sparse) scene geometry as output. The percentage of oriented cameras, the percentage of matched points and the mean bundle adjustment reprojection error have been extracted from the Bundler output,. In particular the percentage of matched points has been evaluated as:

$$\% \text{ of matched features}_{EN,PYR,CH,SET} = \frac{\text{matched features}_{EN,PYR,CH,SET}}{\text{number of features}_{EN,PYR,CH,SET}}$$

Where the *matched features* represents the number of 3D points in the sparse scene reconstruction resulting at the end of the bundle adjustment process, and the *number of features* represent the total number of features extracted for each configuration.

The percentage of oriented cameras has been evaluated as:

$$\% \text{ of oriented cameras}_{EN,PYR,CH,SET} = \frac{\text{oriented cameras}_{EN,PYR,CH,SET}}{\text{number of images}_{SET}}$$

Where *oriented cameras* is the number of oriented images for each configuration, and *number of images* is the number of images included in each set at each configuration.

All the measured values have been collected in a table, which represents the set of data used for the following phase of processing and interpretation of results

5.4.2. Datasets generation

196 images belonging to the seven image sets have been extracted from the whole dataset and pre-processed, to perform the statistical analysis. First, the three colour enhancement methods (ACE, PCA, HIST) have been applied on the original full resolution images using batch actions in Photoshop (for HIST method) and Matlab script (for ACE and PCA method). A Matlab script has been prepared to generate the collection of datasets.

The red, green and blue colour components have been subsequently extracted only from images enhanced with ACE and HIST methods, while no action has been taken on images enhanced with PCA method, because this produces a single channel output. Then, all the images have been resized according to the pyramid levels.

The script prepares a folder for each combination of factors (for example: method = ACE, pyramid = 2, Channel = Red, image set = Set 6), putting into it the corresponding pre-processed images.

A total number of 315 datasets corresponding to each possible configuration of factors have been prepared, and for each of them the feature extraction and matching processes have been performed and Bundler has been executed.

5.5. Statistical analysis

The data related to the measured parameters have been processed by using analysis of variance techniques in Matlab environment. The results have allowed to determine which are, in statistical terms, the factors or the combination of factors that effectively cause a variation in the chosen parameters. In particular, two different analysis have

been performed. The first one includes the image sets 1,2,3,6 and 7, acquired using a Nikon D7000 reflex camera with the automatic white balance setting. The second one includes the image sets 4 and 5, acquired using a Canon S100 with a custom white balance setting.

In the next sections, the results relative to the analysis conducted on image set 1, 2, 3, 6 and 7 will be reported, then these results will be compared to the ones obtained analysing the image sets 4 and 5 in section 5.5.6.

5.5.1. Anova analysis

An analysis of variance has been conducted using the Matlab Statistical toolbox, in particular the function *anovan* has been used to perform a multiway (n-way) analysis of variance (*anova*) for testing the effects of multiple factors on the mean of the values of the current parameter. The function generates a table with six columns (fig 5.4):

- The first shows the source of the variability (factors and combination of factors).
- The second shows the sum of squares (*SS*) due to each source.
- The third shows the degrees of freedom (*df*) associated with each source.
- The fourth shows the mean squares (*MS*), which is the ratio SS/df .
- The fifth shows the values relative to the Fisher statistics (*F statistics, F*), which are the ratios of the mean squares.
- The sixth shows the *p-values* for the F statistics, which is computed from the cumulative distribution function (*cdf*) of F, and is inversely proportional to it.

If any *p-value* is near to zero, there maybe doubts on the associated null hypothesis. For example, a sufficiently small *p-value* for the factor EN (Enhancement algorithm) suggests that at least one EN-sample mean is significantly different from the other EN-sample means; that is, there is a main effect due to factor EN. It is necessary to choose a bound for the *p-value* to determine whether a result is statistically significant or not. The *p-value* represents the complementary of the confidence level. A value α of this parameter means that the factor (or the combination of factors) is influent with a confidence level of $(1 - \alpha) \cdot 100$. It is common to declare a result significant if the *p value* is less than 0.05.

| Source | Sum Sq. | d. f. | Mean Sq. | F | Prob>F |
|------------|--------------|-------|--------------|---------|--------|
| EN | 1.47532e+009 | 2 | 737662186.6 | 1065.41 | 0 |
| PYR | 2.01772e+009 | 2 | 1008858245.9 | 1457.1 | 0 |
| CH | 9.92779e+007 | 3 | 33092640.6 | 47.8 | 0 |
| SET | 2.56979e+008 | 4 | 64244818.8 | 92.79 | 0 |
| EN*PYR | 9.27072e+008 | 4 | 231767966.4 | 334.74 | 0 |
| EN*CH | 1.24859e+008 | 6 | 20809914.1 | 30.06 | 0 |
| EN*SET | 1.15487e+008 | 8 | 14435888 | 20.85 | 0 |
| PYR*CH | 1.25265e+008 | 6 | 20877515.3 | 30.15 | 0 |
| PYR*SET | 3.70207e+008 | 8 | 46275819.3 | 66.84 | 0 |
| CH*SET | 3.53563e+007 | 12 | 2946360.1 | 4.26 | 0.0001 |
| EN*PYR*CH | 1.40029e+008 | 12 | 11669113.1 | 16.85 | 0 |
| EN*PYR*SET | 1.48124e+008 | 16 | 9257723.1 | 13.37 | 0 |
| EN*CH*SET | 3.06397e+007 | 24 | 1276653.8 | 1.84 | 0.0356 |
| PYR*CH*SET | 4.28374e+007 | 24 | 1784889.7 | 2.58 | 0.0026 |
| Error | 3.3234e+007 | 48 | 692374.8 | | |
| Total | 5.94241e+009 | 179 | | | |

Figure 5.4. Example of an anova table related to the parameter “mean extracted features”.

The table in fig. 5.4, related to the parameter “*mean extracted features*”, may lead to deduce the following conclusions:

- All the factors, EN (*Image enhancement method*), PYR (*Image pyramid level*), CH (*Colour channel*) and SET (*Image set*), are influent with a confidence level of 95%;
- All the combination of factors are influent with a confidence level of 95%;
- The most influent factors for the parameter “*mean extracted features*” are in order: Image pyramid level, image enhancement method, image set and colour channel;
- The second level combination “EN*PYR” is more influent than the colour channel and image set;
- The second level combination “PYR*SET” is more influent than the colour channel.

The *anova* table for the other parameters shows that the factor “*Image enhancement algorithm*” (EN) ranks second in terms of influence for which regards the number of extracted features, but it is the most influent factor for the percentage of matched features and the percentage of oriented cameras. The results for each parameter will be discussed in detail in the next section. In order to obtain additional information about the statistical variability of the data, the function *multicompare* in the Matlab statistical toolbox has been used. The output is a graph (an example in fig 5.5) in which the

sample mean of each parameter (circles) grouped by the level of the current factor, and the relative 95% confidence interval (lines passing through the circles) are depicted.

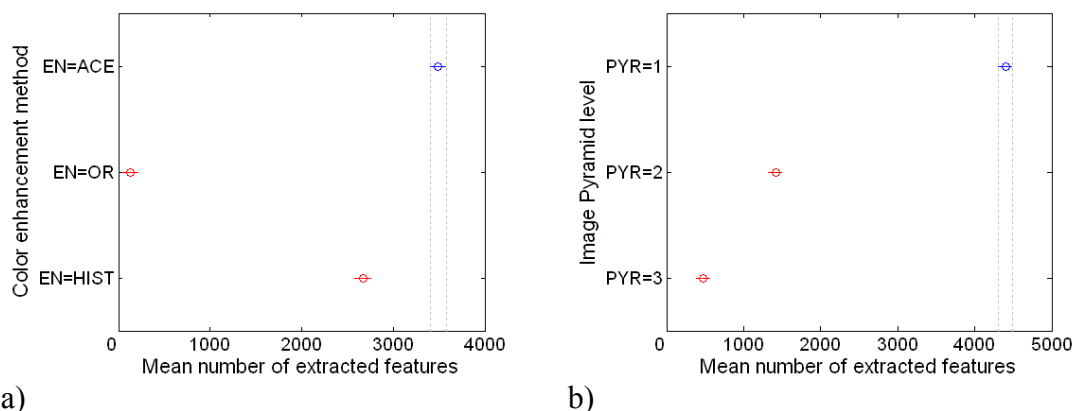


Figure 5.5. Graphs obtained with the *multicompare* function of the factors image enhancement method (a) and image pyramid level (b) related to the parameter “Mean number of extracted features”.

These graphs have been used to study in detail the influence of each factor, confirming the results obtained with the *anova* tables. In the example shown fig 5.5, it is possible to notice that there is no overlap between the confidence intervals, which means that the factors are all influential with a confidence level of 95%. In addition, the confidence intervals are quite narrow, indicating a low uncertainty on the population means.

In order to gather more detailed information about the influence of a single factor on a parameter, it is possible to compare the results for the levels of a particular factor with all the other ones, that is, the results for each parameter have been compared for all the combinations. For this purpose a series of bar graph such as in fig. 5.6 has been prepared.

In this graph, for example, the measured values of the mean extracted features for all the levels of the considered factors are reported, grouped by the image set. It's clear that the behaviour is similar between all the datasets, and only for SET_3 a higher number of features have been extracted. This is due to the higher amount of information of aerial-type layout photographs and to the reduced distance from the scene in this particular case. It is possible to notice that each enhancement algorithm presents a different behaviour considering the colour channels. In particular the red channel of the ACE method shows a higher amount of features than the HIST method, in which the best channel is the blue one. The data related to the original images show a very low number

of features, such that it cannot be used with success (detailed results will be shown in the next paragraph).

The statistical analysis has shown how each factor affects the accuracy and the completeness of the SFM procedure, resulting in important information that may improve the performance of the acquisition system in underwater environment. The validity of the choice of the influencing factors, which are all statistically relevant, has also been verified. The relevant results obtained for each parameter of interest will be reported in detail below.

5.5.2. Mean extracted features

The conducted variational analysis has shown that the most influential factor on the parameter “*mean number of extracted features*” is the image resolution (identified by the image pyramid level), followed by the image enhancement algorithm being used, while the factors colour channel and image set are less influential, as shown in the anova table in fig 5.7 and from the histogram in fig 5.8.

The results in fig. 5.7 do not include the images enhanced with PCA method, since only one component is available. A separate anova analysis has been conducted comparing PCA images to the other channels related to ACE and HIST methods. The histogram shows the values obtained using the PCA correction.

The colour channel appears to be less influential, as enhancement algorithms operate a mixing between the various colour channels in a different way. The histogram in Fig 5.8 shows that by using the ACE method, the red channel does a much better performance in terms of number of extracted features, while using the HIST method the blue appears to be the best channel, even if the difference is small.

The number of extracted points strongly depends on image resolution. In fig. 5.8 it is possible to see that for all combinations a greater number of features are obtained for images at higher resolution.

In particular, it is possible to see how there is no overlap between the confidence intervals, so all factors have an influence with a confidence level of 95%. Furthermore it is possible to detect how globally the ACE method and the first pyramid level leads to a higher number of extracted features.

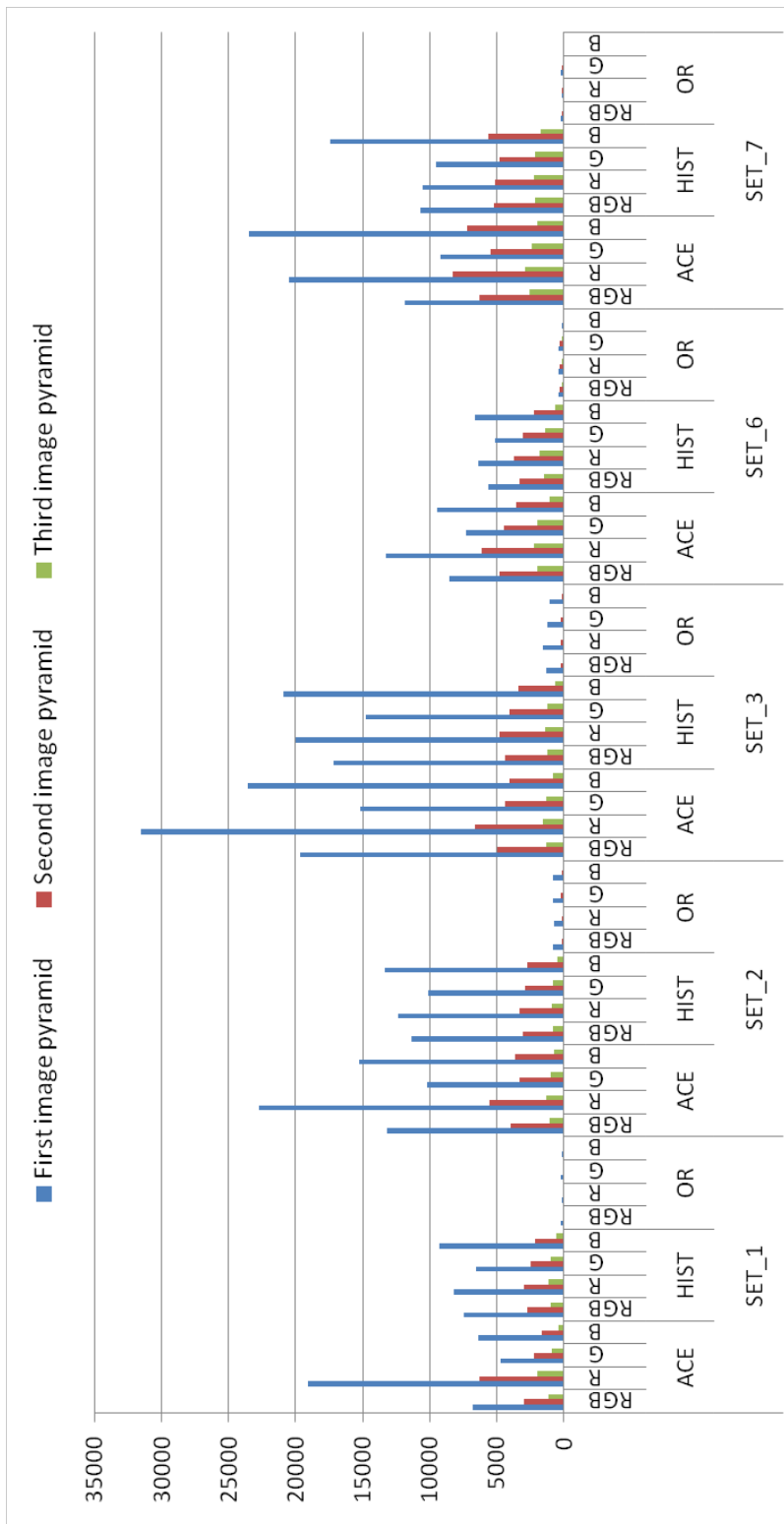


Figure 5.6. Example of the histogram representing the data related to the mean number of extracted features.

| Source | Sum Sq. | d. f. | Mean Sq. | F | Prob>F |
|------------|--------------|-------|--------------|---------|--------|
| EN | 1.47532e+009 | 2 | 737662186.6 | 1065.41 | 0 |
| PYR | 2.01772e+009 | 2 | 1008858245.9 | 1457.1 | 0 |
| CH | 9.92779e+007 | 3 | 33092640.6 | 47.8 | 0 |
| SET | 2.56979e+008 | 4 | 64244818.8 | 92.79 | 0 |
| EN*PYR | 9.27072e+008 | 4 | 231767966.4 | 334.74 | 0 |
| EN*CH | 1.24859e+008 | 6 | 20809914.1 | 30.06 | 0 |
| EN*SET | 1.15487e+008 | 8 | 14435888 | 20.85 | 0 |
| PYR*CH | 1.25265e+008 | 6 | 20877515.3 | 30.15 | 0 |
| PYR*SET | 3.70207e+008 | 8 | 46275819.3 | 66.84 | 0 |
| CH*SET | 3.53563e+007 | 12 | 2946360.1 | 4.26 | 0.0001 |
| EN*PYR*CH | 1.40029e+008 | 12 | 11669113.1 | 16.85 | 0 |
| EN*PYR*SET | 1.48124e+008 | 16 | 9257723.1 | 13.37 | 0 |
| EN*CH*SET | 3.06397e+007 | 24 | 1276653.8 | 1.84 | 0.0356 |
| PYR*CH*SET | 4.28374e+007 | 24 | 1784889.7 | 2.58 | 0.0026 |
| Error | 3.3234e+007 | 48 | 692374.8 | | |
| Total | 5.94241e+009 | 179 | | | |

Figure 5.7. Anova table relative to the parameter “mean extracted features”.

It is also evident that the red channel allows to extract a greater number of points while the blue and green channels and the RGB composite image lead to almost the same results.

It is important to analyse the factor “*image set*”. The analysis of the graphs in Fig. 5.9 shows a difference of behaviour within the various sets. In particular, the highest number of feature is extracted from the images belonging to the set number three, which is characterized by pictures taken using a standard aerial photography layout with a reduced distance to the subject (less scattering volume and more object coverage and pictures overlap).

The analysis of the graphs obtained by the Matlab’s *multicompare* function (fig. 5.9) confirm what has just been assumed.

The first two sets are related to the same area, with the difference that the second set has been acquired after removing the sand that cover the tiled floor of the thermal room (see section 5.4.1), making the point detection more reliable.

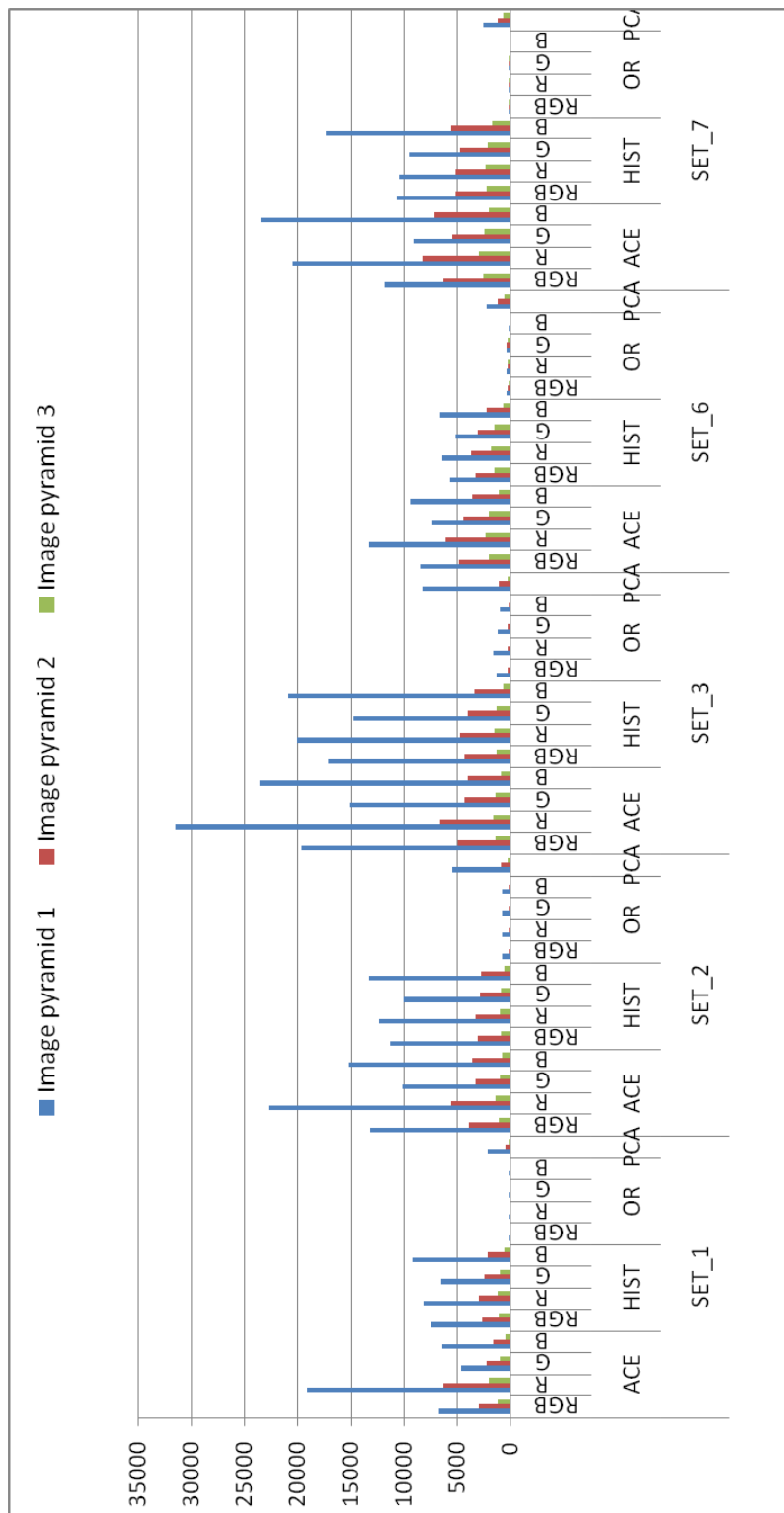


Figure 5.8. Mean number of extracted features: detailed comparison related to all the configurations.

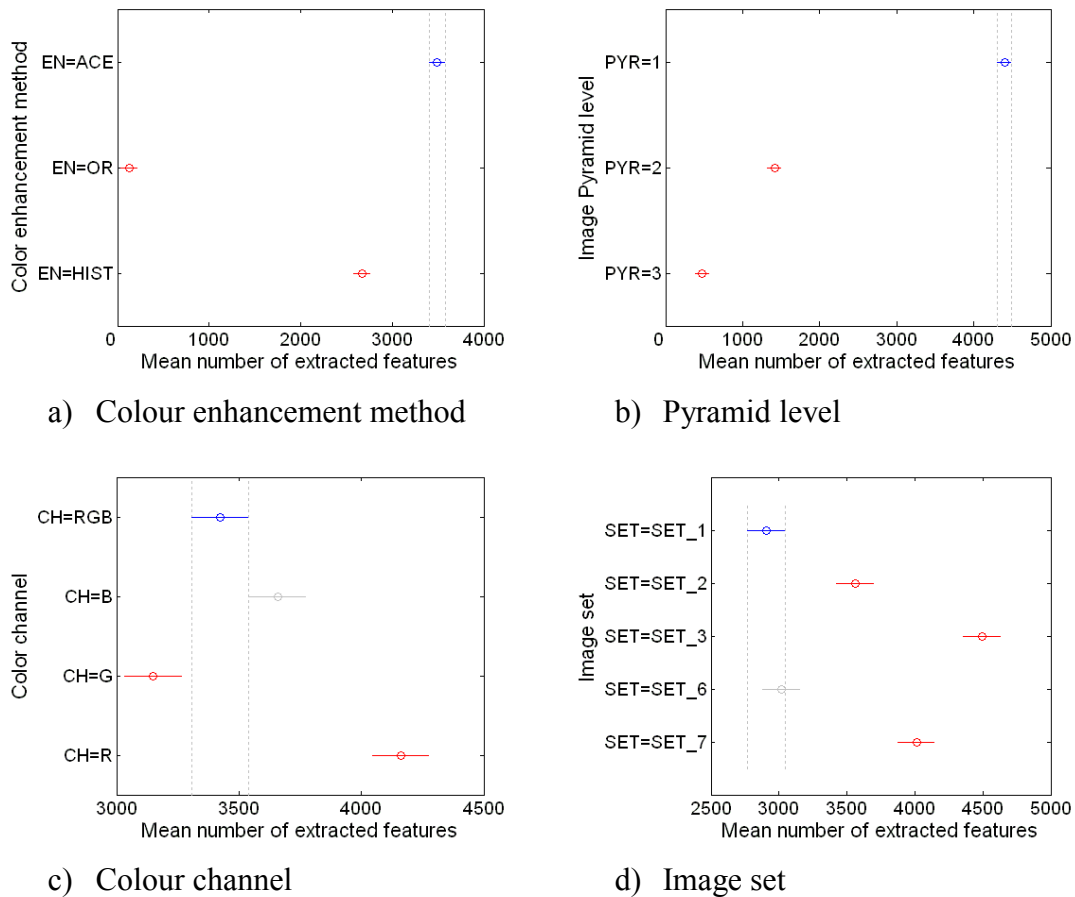


Figure 5.9. Mean values and confidence intervals related to the parameter *mean number of extracted features*.

The sets number 6 and 7 are characterized by oblique pictures taken at different distances. The set 6 shows a lower number of features, because the pictures have been taken from a greater distance and the presence of the blue background is more evident.

A further analysis has been carried out by analysing the anova tables separately for each set of images (example reported in fig. 5.10 and fig 5.11 related to the second image set). The data demonstrate what has been assumed before, that is, a similar trend of the data for all sets in which the most influential parameter is represented by the image resolution followed by the enhancement algorithm.

The colour channel appears the less influential, with a similar behaviour for the blue and green channels and the RGB composite image, with respect to a red channel that features a greater number of points.

| Source | Sum Sq. | d. f. | Mean Sq. | F | Prob>F |
|--------|--------------|-------|--------------|-------|--------|
| EN | 2.749e+008 | 2 | 1.3745e+008 | 56.85 | 0 |
| PYR | 5.11107e+008 | 2 | 2.55553e+008 | 105.7 | 0 |
| CH | 1.86649e+007 | 3 | 6.22165e+006 | 2.57 | 0.1027 |
| EN*PYR | 2.21239e+008 | 4 | 5.53096e+007 | 22.88 | 0 |
| EN*CH | 2.48532e+007 | 6 | 4.1422e+006 | 1.71 | 0.2011 |
| PYR*CH | 2.2426e+007 | 6 | 3.73767e+006 | 1.55 | 0.2449 |
| Error | 2.90116e+007 | 12 | 2.41763e+006 | | |
| Total | 1.1022e+009 | 35 | | | |

Figure 5.10. Anova table related to the parameter “mean extracted features” measured on images belonging to the second image set.

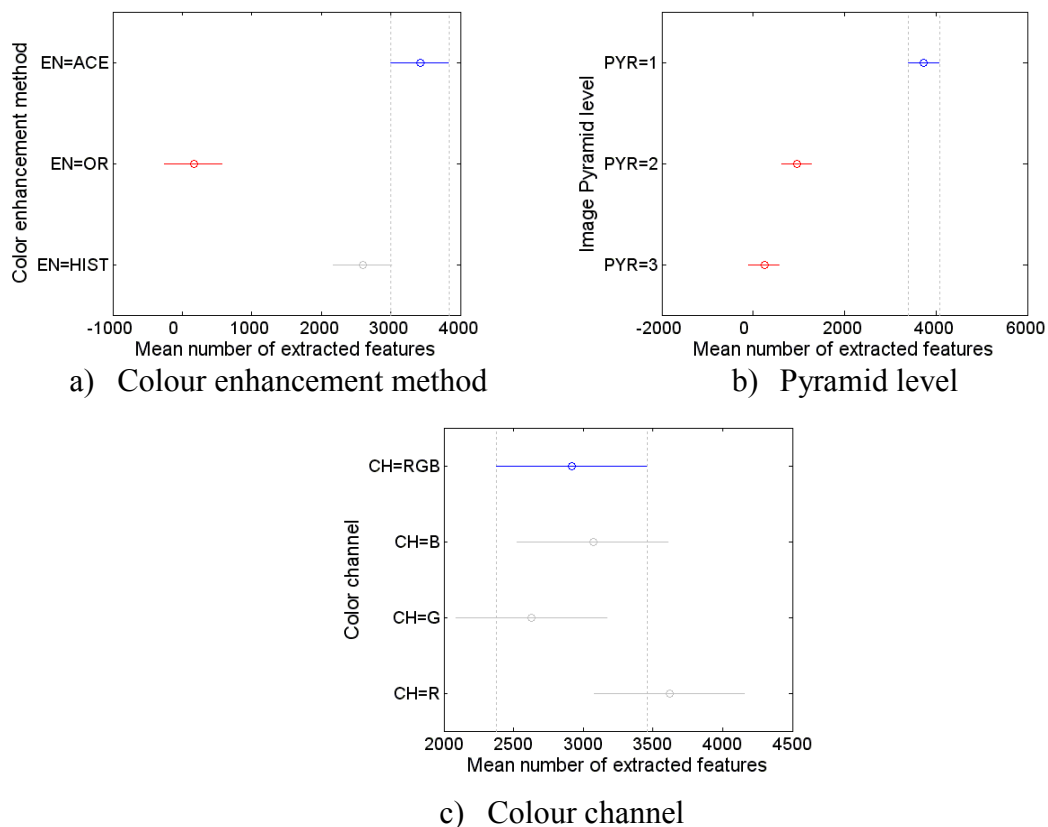


Figure 5.11. Mean values and confidence intervals related to the parameter *mean number of extracted features* of the second image set.

5.5.3. Percentage of matched features

The most influent factor on the percentage of matched features is the image enhancement algorithm being used (fig. 5.12). In particular, by analysing the confidence

intervals (fig. 5.13), it is possible to notice that the HIST method allows to match a higher number of features. The second factor in terms of influence is represented by the colour channel. The graph in fig. 5.13 shows that the RGB images and the green channel allowed to match the maximum number of features.

| Source | Sum Sq. | d. f. | Mean Sq. | F | Prob>F |
|------------|---------|-------|----------|-------|--------|
| EN | 52.568 | 2 | 26.2839 | 21.07 | 0 |
| PYR | 14.801 | 2 | 7.4007 | 5.93 | 0.005 |
| CH | 54.429 | 3 | 18.1429 | 14.54 | 0 |
| SET | 38.733 | 4 | 9.6833 | 7.76 | 0.0001 |
| EN*PYR | 38.797 | 4 | 9.6991 | 7.77 | 0.0001 |
| EN*CH | 33.912 | 6 | 5.652 | 4.53 | 0.001 |
| EN*SET | 124.594 | 8 | 15.5742 | 12.48 | 0 |
| PYR*CH | 11.387 | 6 | 1.8979 | 1.52 | 0.1915 |
| PYR*SET | 41.895 | 8 | 5.2368 | 4.2 | 0.0007 |
| CH*SET | 59.91 | 12 | 4.9925 | 4 | 0.0003 |
| EN*PYR*CH | 15.473 | 12 | 1.2895 | 1.03 | 0.4351 |
| EN*PYR*SET | 144.093 | 16 | 9.0058 | 7.22 | 0 |
| EN*CH*SET | 30.938 | 24 | 1.2891 | 1.03 | 0.4478 |
| PYR*CH*SET | 32.851 | 24 | 1.3688 | 1.1 | 0.3818 |
| Error | 59.889 | 48 | 1.2477 | | |
| Total | 754.271 | 179 | | | |

Figure 5.12. Anova table relative to the parameter “percentage of matched features”.

The analysis of the histogram in fig. 5.14 is not so immediate, because the behaviour is not regular as for the mean number of features histogram. The first information that can be extracted are the performances of the original images, especially in the image set number two and three. These values are related to a small amount of extracted features (see histogram in fig.5.8) and to the two “easiest” datasets, characterized by highly textured objects (the cleaned pavement in set 2) and a better visibility (the camera is near the structures in set 3).

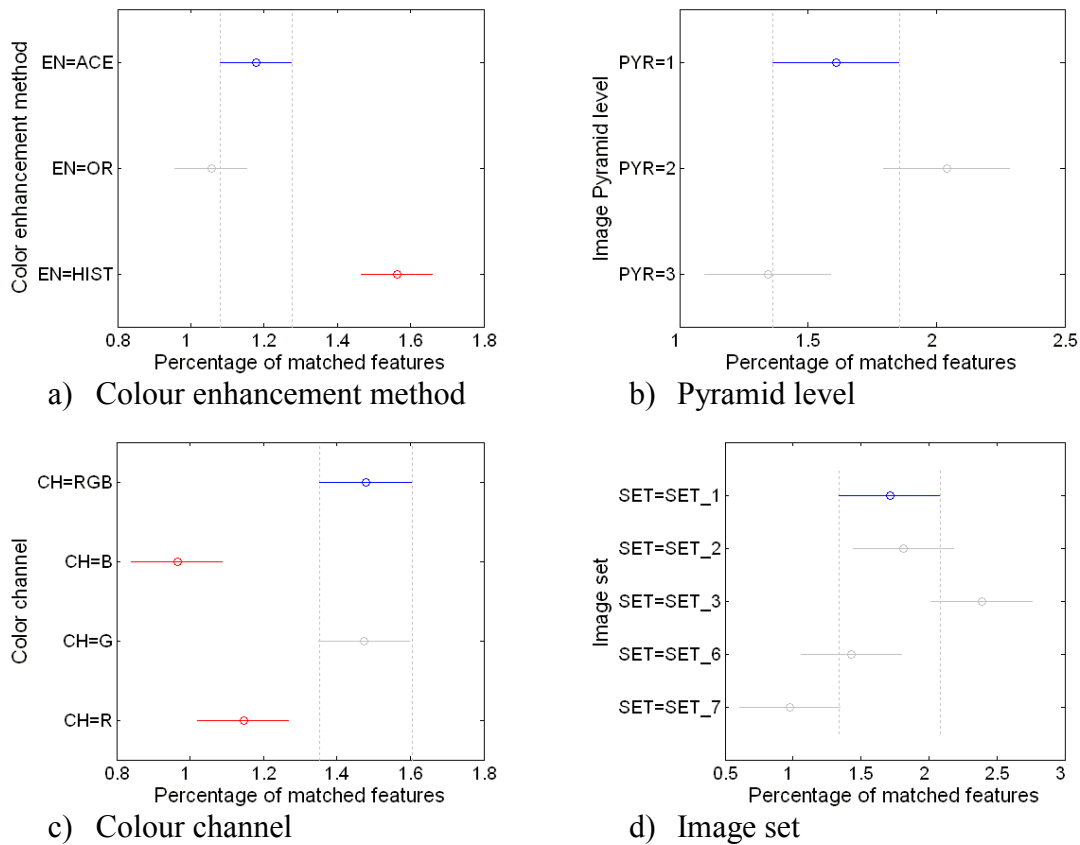


Figure 5.13. Mean values and confidence intervals related to the parameter percentage of matched features.

The results related to the factor “*image pyramid level*” reflects the low influence deduced from the anova table (fig. 5.12), but it is possible to point out that resized images leads to an higher percentage of matched points in some cases. For example, the histogram stretching method applied on the second set is able to match a sufficient number of features only on the third pyramid (lower resolution). In set 7, for all the three enhancement algorithms used in the experimentation, it is possible to notice that resized images lead to a higher performance of the feature matching algorithm. One of the reasons for this behaviour resides in the fact that by reducing the resolution of the images, it is possible to find more robust features, as they are extracted from the more evident details, whereas the noise of the marine environment (such as suspended particles) is attenuated by the reduced resolution.

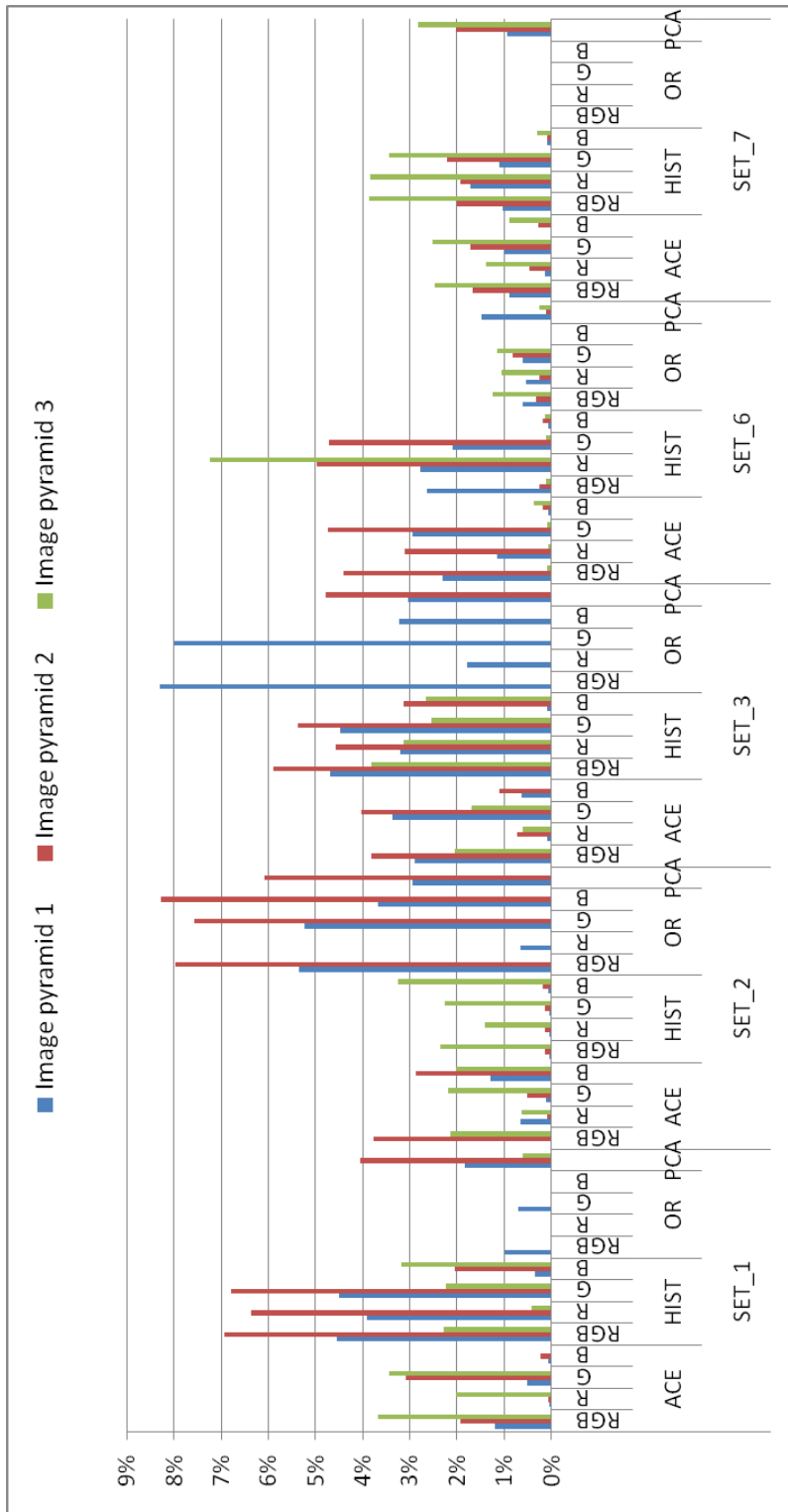


Figure 5.14. Percentage of matched features: detailed comparison related to all the configurations.

5.5.4. Percentage of oriented cameras

From the analysis of the anova table related to the parameter “percentage of oriented cameras” (fig. 5.15), the image enhancement algorithm ends up to be the most influent factor. In fact, from the graphs in fig 5.16, the performances of both ACE and HIST methods are clearly better.

In order to compare the results of the single channel resulting from the PCA correction, an additional series of anova tables has been prepared for each channel. Fig. 5.17 contains the anova table computed from the data relative to only the RGB images, which include the single component from PCA correction and the confidence intervals for the factor “*Image enhancement method*” (fig. 5.18). The most influent factor also in this case is the enhancement method, but it is apparent that the PCA correction shows similar results to ACE and HIST methods.

| Source | Sum Sq. | d. f. | Mean Sq. | F | Prob>F |
|------------|----------|-------|----------|-------|--------|
| EN | 39910 | 2 | 19955.01 | 91.08 | 0 |
| PYR | 9031.1 | 2 | 4515.54 | 20.61 | 0 |
| CH | 8146.6 | 3 | 2715.54 | 12.39 | 0 |
| SET | 16335.1 | 4 | 4083.78 | 18.64 | 0 |
| EN*PYR | 2901.5 | 4 | 725.37 | 3.31 | 0.0179 |
| EN*CH | 7186 | 6 | 1197.67 | 5.47 | 0.0002 |
| EN*SET | 16560.2 | 8 | 2070.03 | 9.45 | 0 |
| PYR*CH | 849.3 | 6 | 141.54 | 0.65 | 0.6929 |
| PYR*SET | 7519 | 8 | 939.88 | 4.29 | 0.0006 |
| CH*SET | 10638.4 | 12 | 886.54 | 4.05 | 0.0002 |
| EN*PYR*CH | 2180.9 | 12 | 181.74 | 0.83 | 0.6201 |
| EN*PYR*SET | 17863.8 | 16 | 1116.49 | 5.1 | 0 |
| EN*CH*SET | 6031.3 | 24 | 251.3 | 1.15 | 0.3345 |
| PYR*CH*SET | 4824 | 24 | 201 | 0.92 | 0.5798 |
| Error | 10516 | 48 | 219.08 | | |
| Total | 160493.2 | 179 | | | |

Figure 5.15. Anova table related to the parameter “percentage of oriented cameras”.

This is clearly noticeable also in the histogram in fig. 5.19. It is also possible to notice that the original images (OR) allowed to orient a large number of cameras in sets 2 and 3. The reasons are the same explained in the previous discussion about the percentage of

matched features: the set 2 and 3 represent the “easiest” datasets. On the contrary, no cameras have been oriented in set 7 using original images. Comparing this histogram and the graph in fig 5.16, there is a certain stability with respect to the results for the first and second levels of the image pyramid.

This means that if the maximum number of oriented camera is needed, a lower resolution should be used. This may also save some computational time.

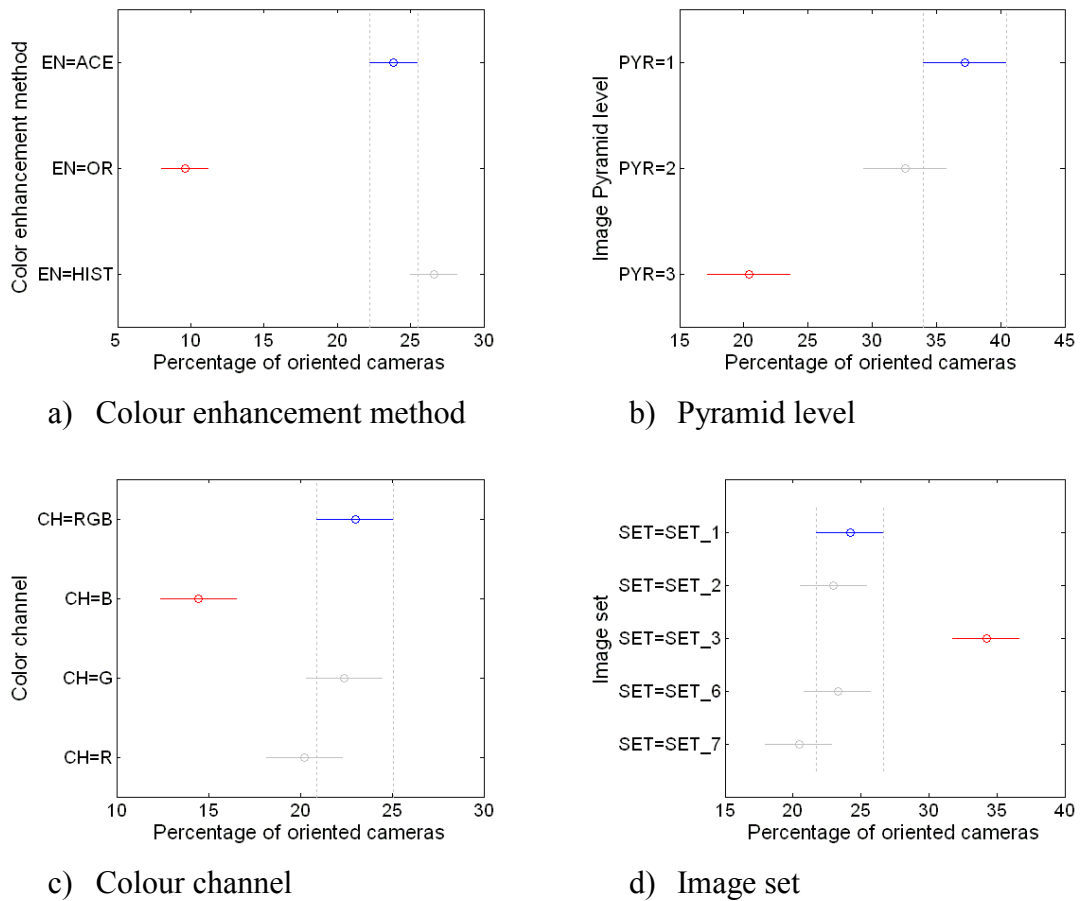


Figure 5.16. Mean values and confidence intervals related to the parameter percentage of oriented cameras.

| Source | Sum Sq. | d. f. | Mean Sq. | F | Prob>F |
|---------|---------|-------|----------|------|--------|
| EN | 13018.4 | 3 | 4339.47 | 8.93 | 0.0004 |
| PYR | 7592.1 | 2 | 3796.05 | 7.81 | 0.0024 |
| SET | 9105.7 | 4 | 2276.43 | 4.69 | 0.0062 |
| EN*PYR | 3214.4 | 6 | 535.73 | 1.1 | 0.3896 |
| EN*SET | 8766.5 | 12 | 730.55 | 1.5 | 0.1905 |
| PYR*SET | 4640.6 | 8 | 580.07 | 1.19 | 0.3434 |
| Error | 11660 | 24 | 485.83 | | |
| Total | 57997.7 | 59 | | | |

Figure 5.17. Anova table related to the parameter “percentage of oriented cameras” computed from the data related to the RGB images only, which include the single component extracted by PCA correction.

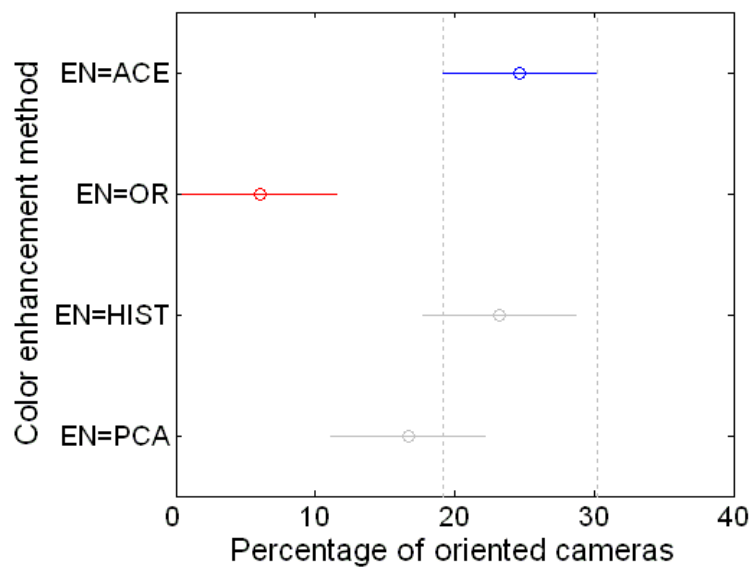


Figure 5.18. Mean values and confidence intervals related to the parameter *percentage of oriented cameras*, computed from the data related to only the RGB images which include the single component from PCA correction.

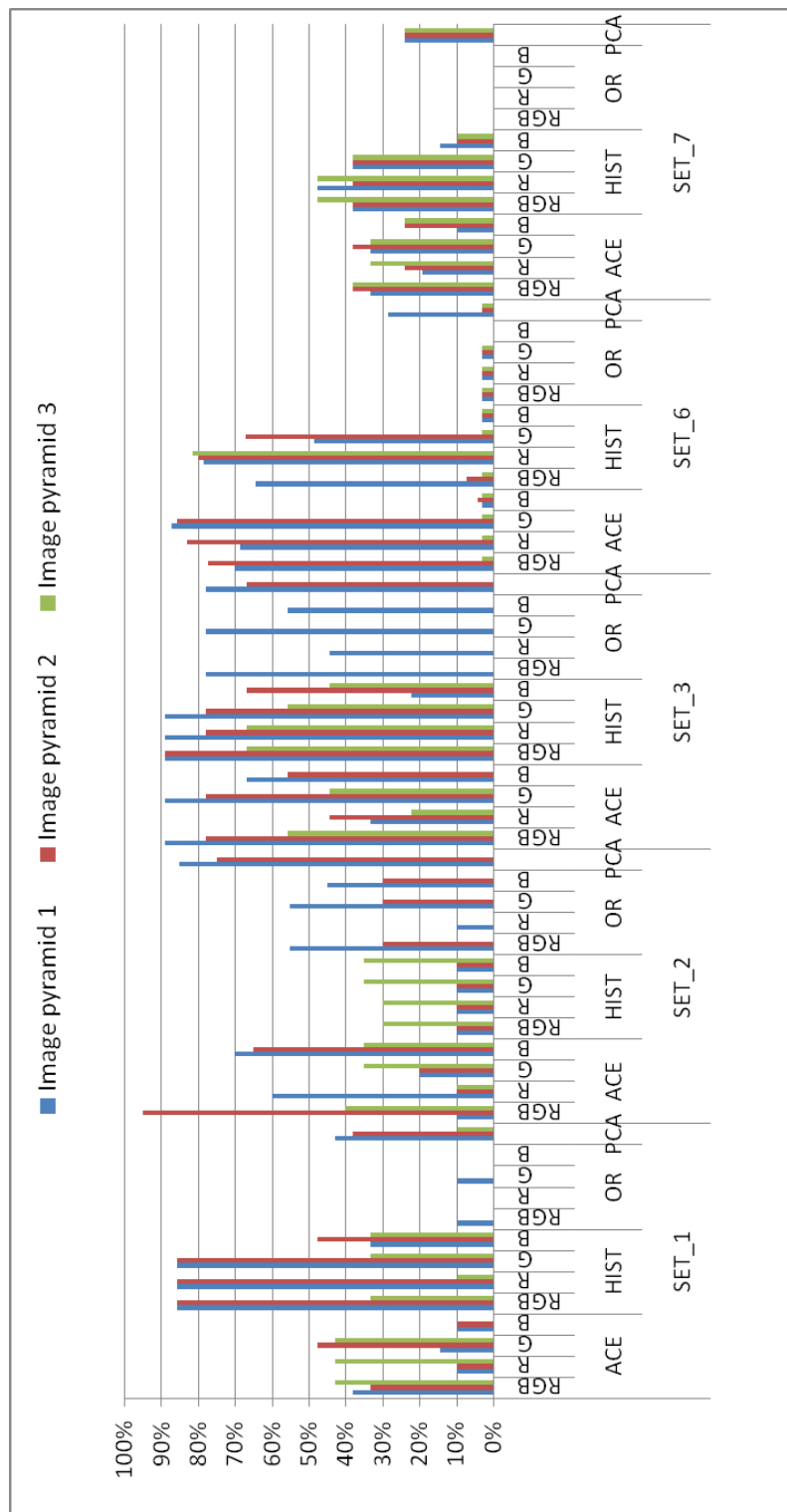


Figure 5.19. Percentage of oriented cameras: detailed comparison related to all the configurations.

5.5.5. Bundle adjustment mean reprojection error

The last parameter is the mean reprojection error measured at the end of the bundle adjustment procedure performed with Bundler. The most influent factor is the image resolution (pyramid level, PYR), as shown in the anova table in fig.5.20.

| Source | Sum Sq. | d. f. | Mean Sq. | F | Prob>F |
|--------|---------|-------|----------|-------|--------|
| EN | 0.20628 | 2 | 0.10314 | 9.17 | 0.0002 |
| PYR | 0.31086 | 2 | 0.15543 | 13.82 | 0 |
| CH | 0.00592 | 3 | 0.00197 | 0.18 | 0.9129 |
| SET | 0.27504 | 4 | 0.06876 | 6.11 | 0.0002 |
| Error | 1.44007 | 128 | 0.01125 | | |
| Total | 2.1073 | 139 | | | |

Figure 5.20. Anova table related to the parameter “*bundle adjustment mean reprojection error*”.

Higher errors are measured for all the combinations on the reconstructions performed using the images resized to 50% (first pyramid level). This means that a lower resolution results in a higher accuracy and subsequently in a lower computational time. Analysing the confidence intervals in fig 5.21 and also the data reported in the histogram in fig. 5.22, it is possible to notice that a higher reprojection error on the two last image sets has been measured. This error is due to the oblique photographs which compose the two sets. The presence of the blue background in almost all the pictures reduces the accuracy of the bundle adjustment process. Comparing the results obtained by the different enhancement methods, there are no significant differences. These results confirm the (Skarlatosa 2012) results. Regarding the influence of the colour channel, almost the same values have been measured in the entire configuration for all components. In fact, if we analyse the anova table in fig. 5.20, the factor colour channel (CH) does not show to be influent on the mean reprojection error.

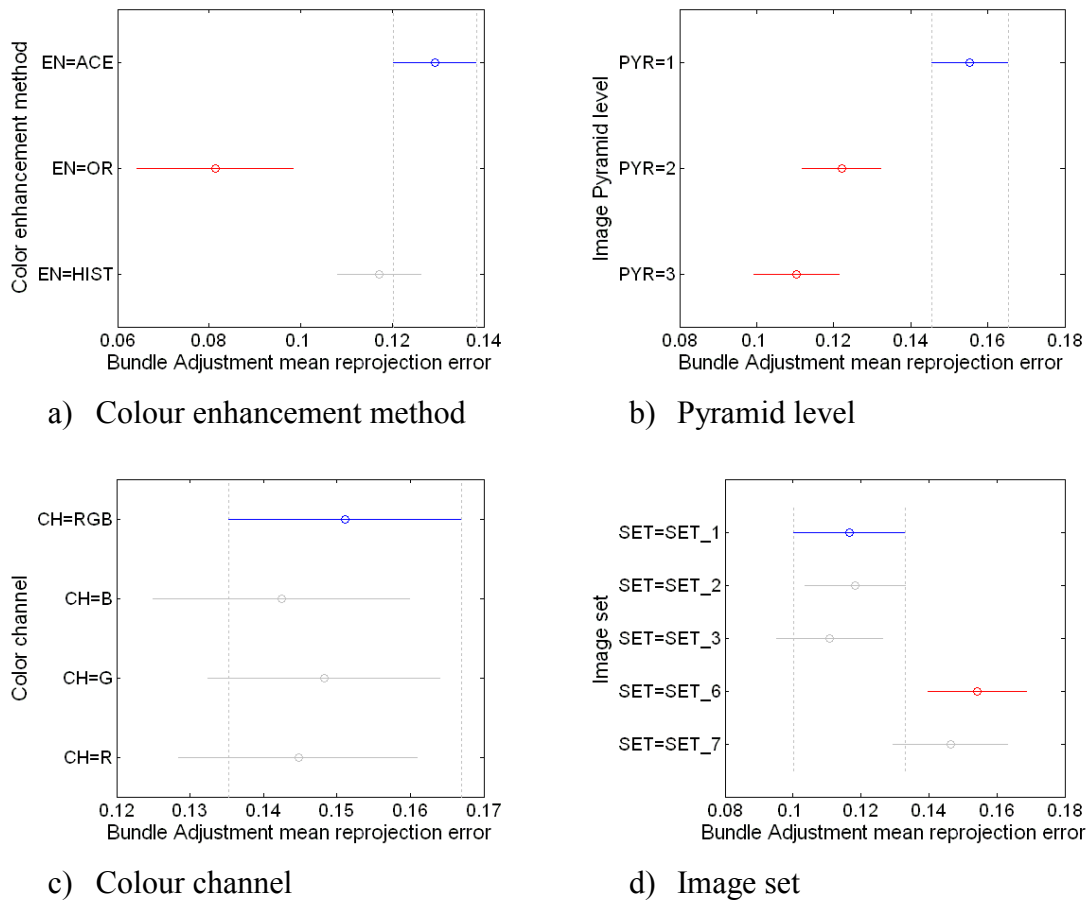


Figure 5.21. Mean values and confidence intervals related to the parameter *bundle adjustment mean reprojection error*.

5.5.6. Effects of white balance correction

A separate analysis has been performed using the images belong to the datasets 4 and 5, acquired balancing the colours with the pre-measured white balance setting. The correct value has been measured using an underwater panel by Lastolite.

The analysis of the anova table and the confidence intervals has been shown similar results to images captured without white balance correction, except for the parameter “*mean number of extracted features*”. In particular the graphs in fig. 5.23 show an about 100% increment of the number of extracted features and, in particular, the good results of uncorrected (OR) images.

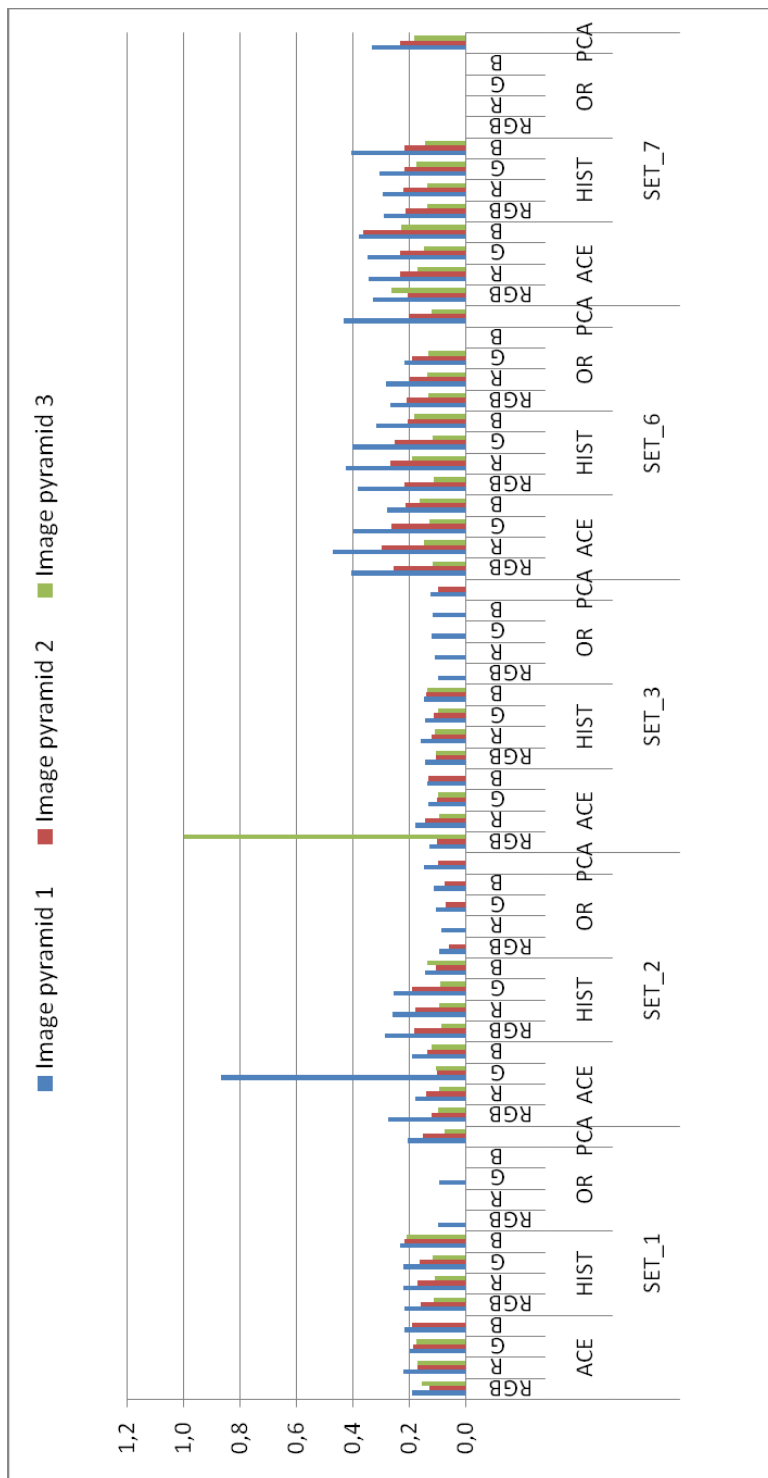


Figure 5.22. Bundle adjustment mean reprojection error: detailed comparison related to all the configurations.

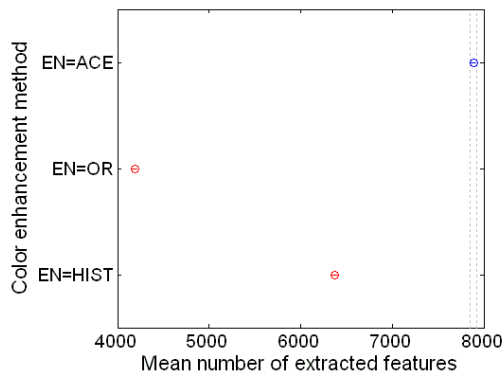
There are no relevant differences between the data related to the percentage of matched features and the percentage of oriented cameras, measured using custom and auto-white balance settings. In addition, less stable results have been obtained. The white balance correction has been performed at the beginning of the survey and changing of lighting conditions can invalidate the measured values producing images with a colour balance that is not uniform across the image sequence.

5.5.7. Discussion

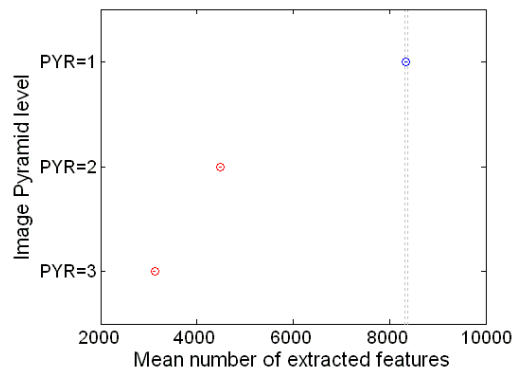
The statistical analysis allowed to choose the best combination of factors which should be used to perform the 3D reconstruction of the site. The graphs in fig. 5.24, 5.25 and 5.26 summarize the results related to the mean number of extracted features, the percentage of the matched features and the percentage of oriented cameras respectively. The accuracy of the SFM procedure (namely the mean reprojection error), is mainly related to the camera network orientation. In fact sets 1, 2 and 3 are characterized by convergent images with a high overlap, forming a more robust network. Moreover, a lower resolution leads to lower errors, but also to a higher percentage of matched features

Regarding the colour enhancement algorithms, both ACE and HIST methods showed the best results (fig. 5.24). In particular the HIST and ACE methods increase considerably the performance of image matching, with a percentage of matched features much higher than the original images. The HIST method produces better results also in terms of percentage of oriented cameras and matched features, increasing the performance by a factor of about 150% and 50% respectively.

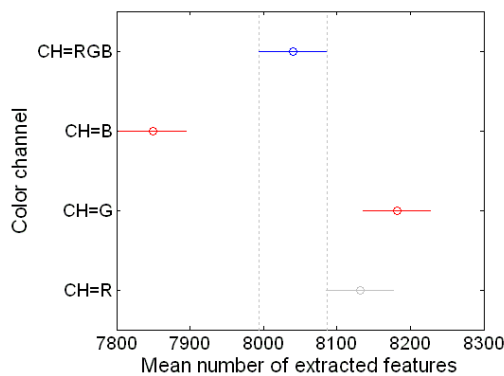
The results in terms of number of oriented cameras between the first and second image pyramid are quite similar (fig. 5.25), so the best choice is to use the lower resolution in order to save computational time.



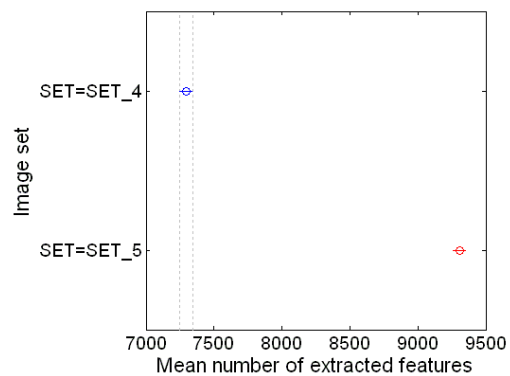
a) Colour enhancement method



b) Pyramid level



c) Colour channel

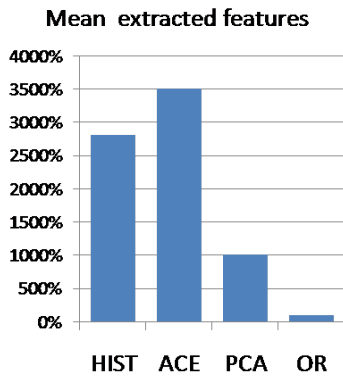


d) Image set

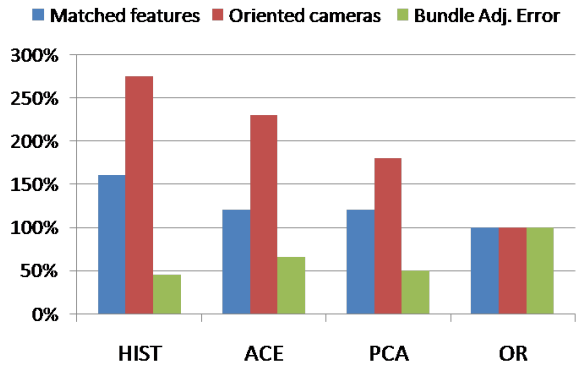
Figure 5.23. Mean values and confidence intervals related to the parameter *mean number of extracted features*: the analysis has been performed on images acquired using custom white balance correction.

Regarding the colour channel, the data have demonstrated the better performances of the RGB images and green channel (fig. 5.26), which outperforms the other channels results in terms of matched features and oriented images. The red channel shows a 20% increment in terms of extracted features, but the results have demonstrated that a higher number of features does not leads to a high percentage of matched features and oriented cameras.

In conclusion, the best combination is represented by RGB images resized to 25% (second pyramid level) and enhanced with the HIST method.

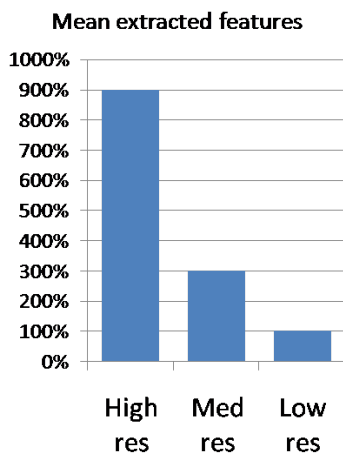


a)

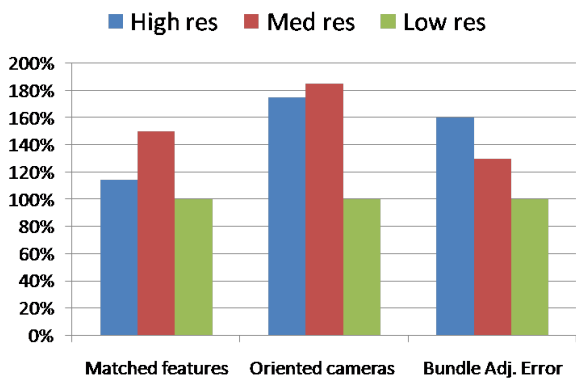


b)

Figure 5.24. Performances comparison of the image enhancement methods for each measured parameter. The performances of the HIST, ACE and PCA image enhancement algorithms, are compared to a reference value related to the original images (OR).



a)



b)

Figure 5.25. Performances comparison of the image resolution (image pyramid) for each measured parameter. The performances of the first pyramid (High res – images resized to 50%), the second pyramid (Med res – images resized to 25%) and the third pyramid (Low res – images resized to 6.25%), are compared taking as reference the value related to the low resolution images.

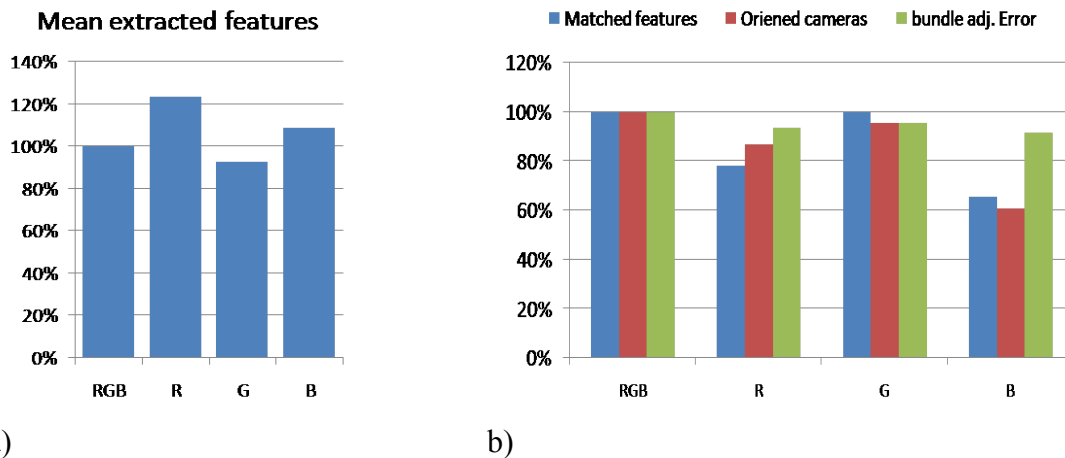


Figure 5.26. Performances comparison of the colour channels for each measured parameter. The performances of the R, G and B components, are compared taking as reference the value related to the RGB images.

5.6. Results

The 722 enhanced images have been processed with Bundler, and a subset of 533 images related to the whole area have been aligned (fig. 5.27), allowing to generate a complete 3D point cloud without registering different meshes.

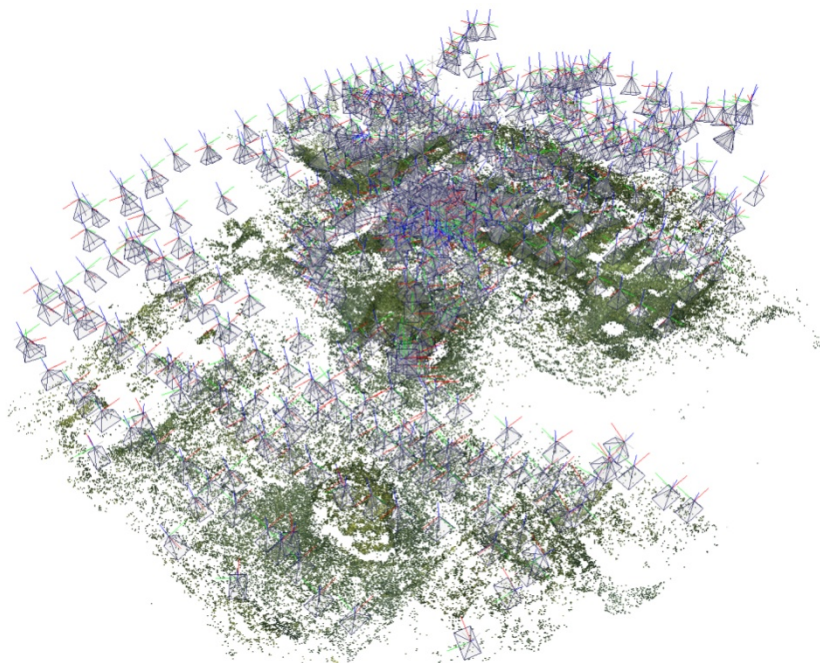


Figure 5.27. Results of the camera orientation process (enhanced pictures with HIST method): sparse point cloud and 533 oriented pictures.

The colour correction improves considerably the matching results. In fact, the full set of original images have not been successfully oriented, thus making it necessary to divide the dataset into two separate groups (North and South part) where the images have been oriented separately (see Chapter 4). This is mainly caused by the sandy seabed present in the central part of the thermal room, which makes the extraction and matching of feature difficult, because of the low contrast.

The Bundler's outputs and the undistorted images have been processed with PMVS2 (Furukawa & Ponce 2007) to create a dense point cloud of the whole site (about 10 millions of points) with RGB information for each 3D point (Fig. 5.28).

The point cloud has been elaborated by using Meshlab tools. The first operation was a manual selection and deletion of unwanted areas. Then, a watertight surface with about 25 millions of triangles (Fig.5.29) has been obtained through a Poisson Surface Reconstruction algorithm.

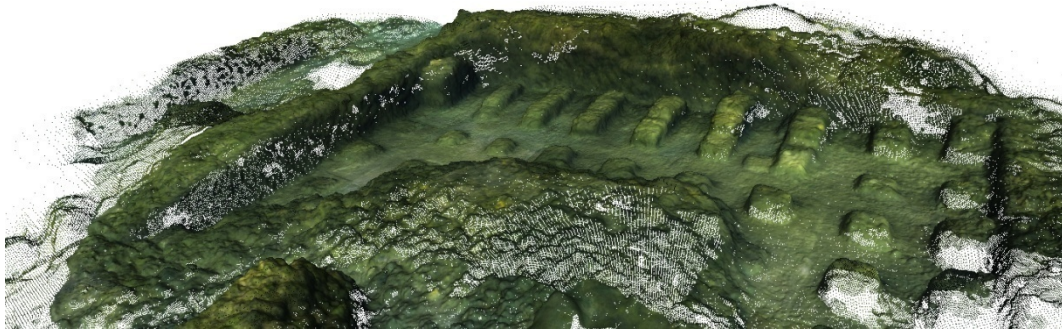


Figure 5.28. Reconstructed dense point cloud.

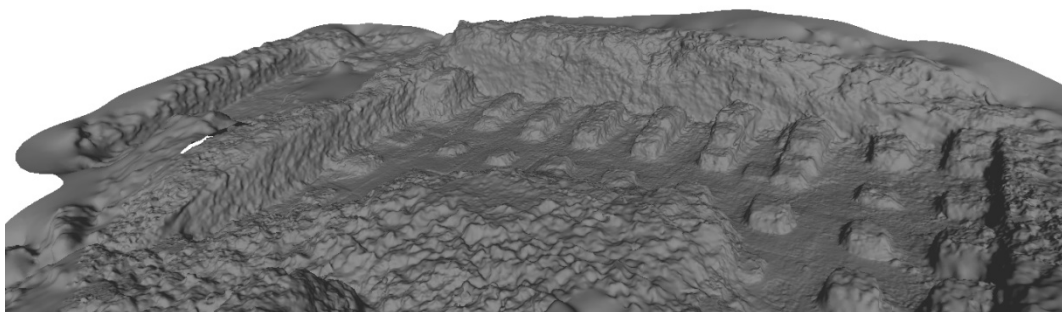


Figure 5.29. Reconstructed surface.

The resulting surface has been subsequently decimated in a mesh of 6.5 million of triangles and 3 million of points, in order to be handled more effectively without losing details. Since the camera orientation procedure has been carried out with an unknown scale factor, it is necessary to scale the model by selecting two points with a known distance. In this experimentation a scale bar has been placed in the scene and reconstructed in order to evaluate the scale factor.

The last step consists in the application of the texture on the 3D surface. Colour information can be extracted directly from the coloured point cloud, but this method does not allow the creation of a high quality texture, because its resolution depends on the point cloud density. Moreover, since the enhancement is often performed to favour the feature extraction process by increasing the contrast without taking into account the fidelity of the colours (usually a single component or greyscale images are used), the colour information stored into the pixels cannot be used.

Since the camera positions are known, the method chosen for texture mapping consists into the projection and blending of the high resolution images directly on the 3D surface. In particular, an image subset has been selected because the averaging among neighbourhood values during the blending on the images works better if a small overlapping area is present (blur effects are reduced).

This subset of images has been extracted from the ones enhanced with the HIST method, which gave the best results also in terms of texture quality. This is due principally to the manual retouching step performed on a sample image and then exported to the whole dataset.

The result of this procedure is a texture with a resolution comparable to the original images and characterized by a high quality level (fig.5.30 and fig. 5.31).

5.7. Conclusions

This work introduced a methodology for the 3D reconstruction of submerged structures. A colour enhancement algorithm and a multi-view stereo technique have been used to obtain an accurate model of an underwater scene in very challenging visibility conditions.

The DOE (Design of the Experiments) has been used to determine the most influential factors on the quality of the 3D reconstruction. In particular three different colour enhancement methods, ACE (Automatic Colour Enhancement), PCA (Principal Component Analysis) and HIST (the developed method based on Histogram stretching and manual retouching), have been tested on images acquired in the real environment of Baiae. The performances of this enhancement methodology have been evaluated using the analysis of variance (*anova*) including the effects of other factors such as image resolution and colour channel.

The results allowed to choose the best combination of factors in terms of SFM bundle adjustment mean reprojection error, extracted features, oriented cameras and matched features. Using this information and enhancing the images, it has been possible to reconstruct the whole site. Moreover, the corrected images allowed to produce a model mapped with a high resolution texture.

Even if these techniques have been used in other works related to underwater archaeology, this experimentation represents a significant case-study to verify their robustness in presence of strong turbidity and bad environmental conditions.

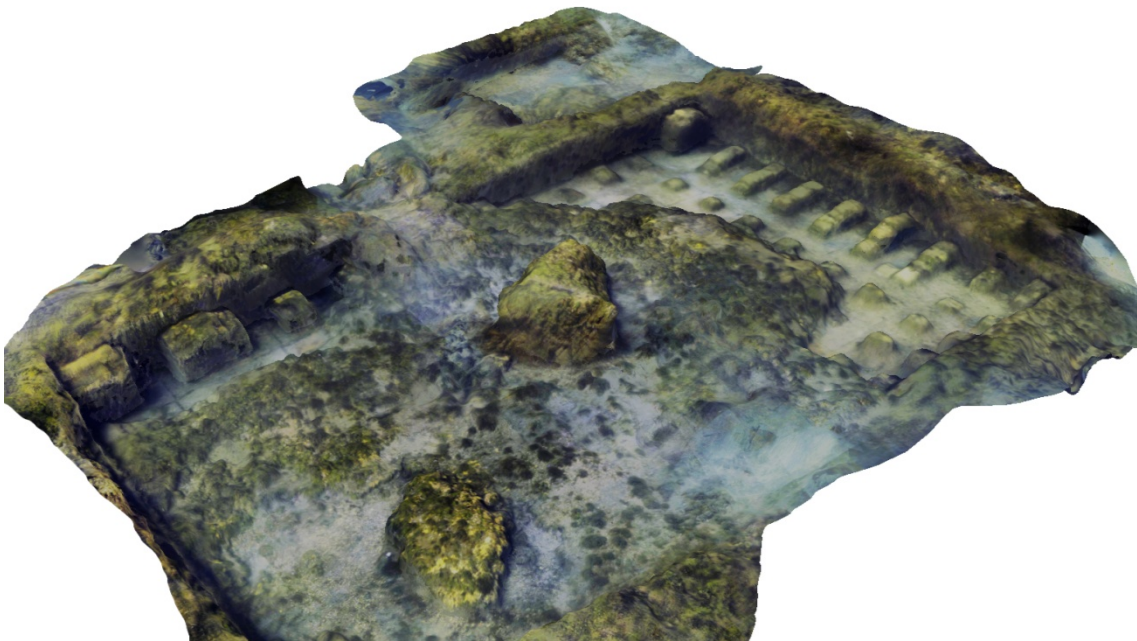
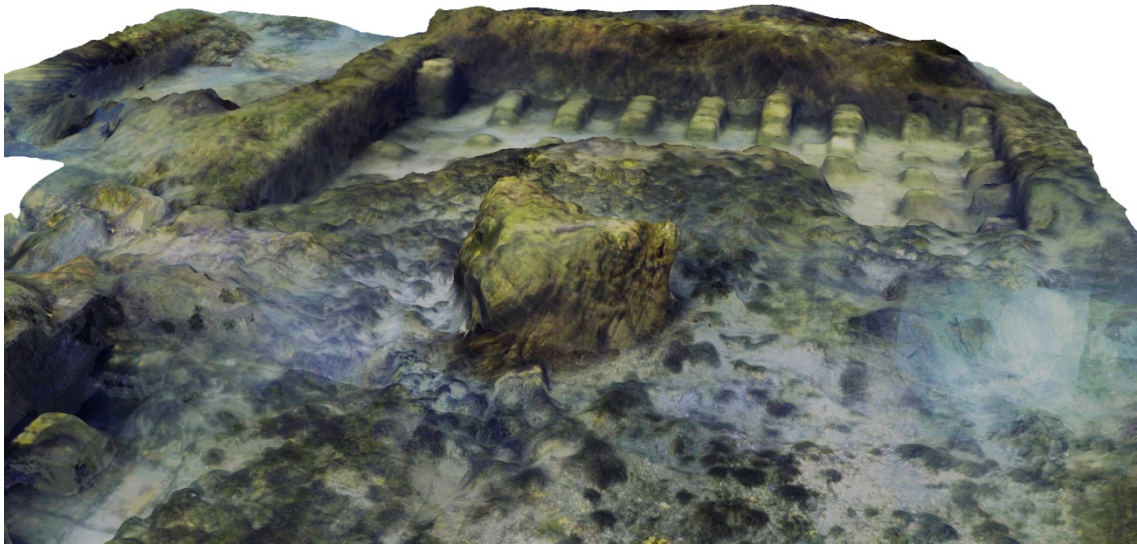


Figure 5.30. Final textured 3D model - a.

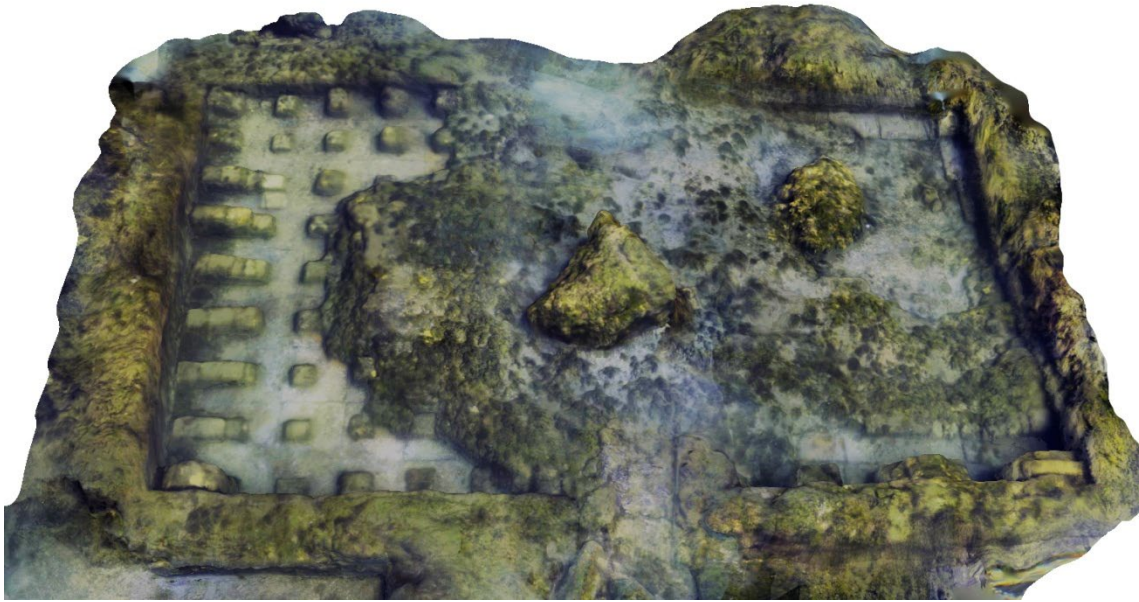
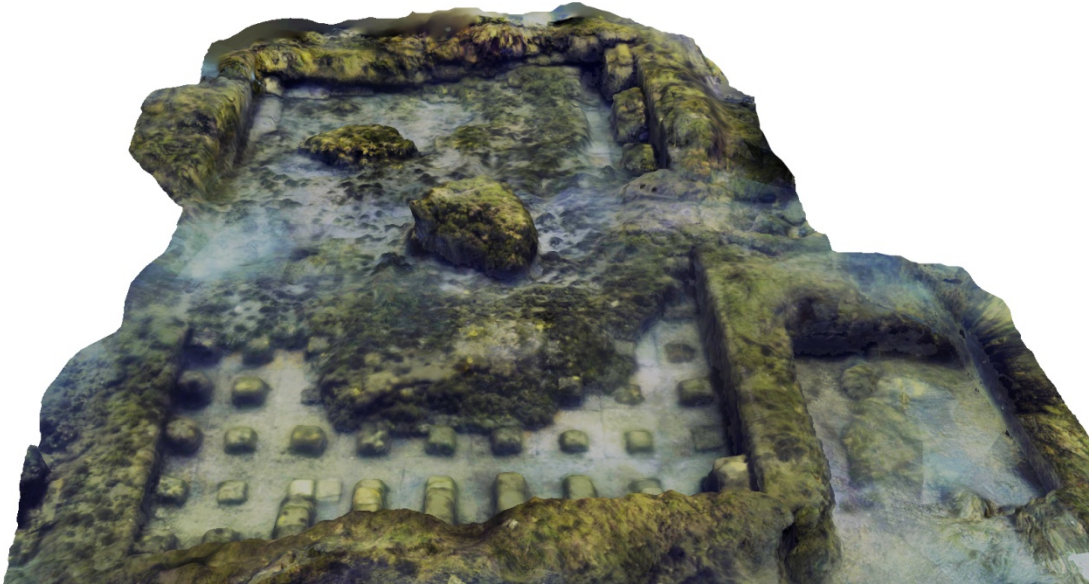


Figure 5.31. Final textured 3D model - b.

Conclusions

This thesis has investigated the 3D imaging techniques based on the use of digital images applied in context of Cultural Heritage. A multi-view passive technique commonly used in air has been experimented in two specific areas: the 3D reconstruction of small sized objects and the survey of underwater archaeological finds.

The 3D reconstruction approach used for small object reconstruction is based on the use of common cameras and macro lenses. The problems related to the use of macro lenses in photogrammetry are the small Depth of Field and the loss of sharpness due to diffraction at small apertures. Image matching algorithms cannot work on blurred areas, and they can't be used to reconstruct the object. The developed methodology overcomes these limitations by using a set of images acquired at different focus distances, which are merged by the means of an image fusion algorithm in order to have the whole object in focus. Unfortunately this method increases the number of required images and the acquisition time, but the methodology gives space to high automation possibilities.

Concerning the underwater 3D imaging, a comparison between an active and a passive stereo technique has been carried out. The results demonstrate that two stereo algorithms widely employed in air applications for 3D reconstruction of close-range objects can be used for underwater applications without implementing any modification, or taking into account optical models of refraction. The analysis has been conducted on each colour channel separately, showing more stable results for the active technique that uses coded patterns. Moreover the noise is higher if compared to passive technique, which suffers the loss of contrast due to turbidity, resulting in holes and missed areas.

A passive multi-view stereo technique has been tested on the field at the underwater archaeological site of Baiae (Naples, Italy), characterized by very challenging environment conditions (fishes, turbidity, algae). In order to improve the performance of the adopted stereo matching algorithms, different image enhancement techniques have been experimented. A statistical approach has been used to determine the most influential factors, allowing to collect relevant data to improve the quality of the 3D reconstruction related to the following experimental campaigns

The work presented in this thesis demonstrates that a passive multi-view technique can be successfully applied in challenging applications such as the reconstruction of small objects or the survey of submerged archaeological sites.

Bibliography

(Agarwal *et al.* 2009) Sameer Agarwal, Noah Snavely, Ian Simon, Steven M. Seitz and Richard Szeliski. Building Rome in a Day. International Conference on Computer Vision, 2009, Kyoto, Japan.

(Aggarwal & Ahuja 2000) M. Aggarwal, N. Ahuja. On Generating Seamless Mosaics with Large Depth of Field. ICPR (2000) 1588-1591.

(AgiSoft LLC. 2010) AgiSoft PhotoScan. <http://www.agisoft.ru/products/photoscan/> (accessed on September 2012).

(Allais *et al.* 2007) Allais, A.G., Brandou, V., Dentrecolas, S., Gilliotte, J.P. and Perrier, M., 2007. Iris - a vision system to reconstruct natural deep-sea scenes in 3D. *Seventeenth International Offshore and Polar Engineering Conference*, Lisbon - Portugal: 111–118.

(ARC 3D 2012) ARC 3D Webservice, <http://homes.esat.kuleuven.be/~visit3d/websevice/v2/>, (accessed on September 2012).

(Atsushi *et al.* 2011) K. Atsushi, H. Sueyasu, Y. Funayama, T. Maekawa. System for reconstruction of three-dimensional micro objects from multiple photographic images. *Computer-Aided Design* 43(8) (2011) 1045-1055

(Barazzetti *et al.* 2011) Luigi Barazzetti, Luigia Binda, Marco Scaioni, Paolo Taranto, Photogrammetric survey of complex geometries with low-cost software: Application to the ‘G1’ temple in Myson, Vietnam, *Journal of Cultural Heritage*, Volume 12, Issue 3, July–September 2011, Pages 253-262, ISSN 1296-2074.

(Barone & Razionale 2004) Barone, S. and Razionale, A.V., 2004. A reverse engineering methodology to capture complex shapes. *XVI Congreso Internacional de Ingeniería Gráfica*. Saragoza – Spain : 1–10.

(Bay *et al.* 2008) Herbert Bay, Andreas Ess, Tinne Tuytelaars, & Luc Van Gool. (2008). Speeded-Up Robust Features (SURF). *Comput. Vis. Image Underst.*, Vol.:110, pp.346-359, isbn/issn:1077-3142.

(Bayer 1976) Bayer, B. E. (1976). Color imaging array. US Patent No. 3,971,065.

(Bazeille *et al.* 2006) Stéphane Bazeille, Isabelle Quidu, Luc Jaulin, & Jean-Phillipe Malkasse. (2006, 16 - 19 Octobre 2006). Automatic Underwater Image Pre-Processing. Paper presented at the CMM'06 - CARACTERISATION DU MILIEU MARIN

(Beraldin *et al.* 2002) Beraldin, J.-A., Picard, M., El-Hakim, S. F., Godin, G., Latouche, C., Valzano, V. and Bandiera, A., 2002. Exploring a Byzantine crypt through a high-resolution texture mapped 3D model: combining range data and photogrammetry. *Proceedings of the CIPA WG6 International Workshop on Scanning for Cultural Heritage Recording, Corfu, Greece, 1st to 2nd September 2002.* 159 pages: 65–72.

(Bianco *et al.* 2011) Bianco G., Gallo A., Bruno F., Muzzupappa M.: A comparison between active and passive techniques for underwater 3d applications. In *3D-ARCH 2011 (Trento, Italy, 2-4 March 2011)*.

(Bitelli & Girardi 2010) G. Bitelli, F. Girardi. Problematrice nel rilievo e modellazione tridimensionale di oggetti di piccole dimensioni nel campo dei Beni Culturali. *Atti 14a Conferenza ASITA, Brescia, Italy, 9-12 nov. 2010*, pp 285-290.

(Bolles *et al.* 1993) Bolles, R. C., Baker, H. H., and Hannah, M. J. (1993). The JISCT stereo evaluation. In *Image Understanding Workshop*, pp. 263–274.

(Born & Wolf 1999) M. Born, E. Wolf. *Principles of Optics*. Cambridge: Cambridge University Press, 1999, p. 461.

(Bouguet 2012) Camera Calibration Toolbox for Matlab, Bouguet, Y-J. <http://www.vision.caltech.edu/bouguetj/>. (accessed on September 2012).

(Boykov & Kolmogorov 2003) Yuri Boykov and Vladimir Kolmogorov. Computing geodesics and minimal surfaces via graph cuts. In *ICCV '03: Proceedings of the Ninth IEEE International Conference on Computer Vision*, page 26, Washington, DC, USA, 2003. IEEE Computer Society.

(Boykov & Kolmogorov 2004) Yuri Boykov and Vladimir Kolmogorov. An experimental comparison of min-cut/maxflow algorithms for energy minimization in vision. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(9):1124–1137, 2004.

(Bradley & Bamford 2004) A. P. Bradley, P. C. Bamford. A one-pass extended depth of field algorithm based on the over-complete discrete wavelet transform. in *Proc. Image and Vision Computing*, 2004, pp. 279–284.

(Brown & Lowe 2005) M. Brown and D. G. Lowe. Unsupervised 3d object recognition and reconstruction in unordered datasets. In *3DIM '05: Proceedings of the Fifth International Conference on 3-D Digital Imaging and Modeling*, pages 56–63, Washington, DC, USA, 2005. IEEE Computer Society.

(Bruno *et al.* 2011) Bruno, F., Bianco, G., Muzzupappa, M., Barone, S. and Razonale, A. V., 2011. Experimentation of structured light and stereo vision for underwater 3D reconstruction. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(4):508-518.

(Chambah *et al.* 2004) Majed Chambah, Dahbia Semani, Arnaud Renouf, Pierre Courtellemont, and Alessandro Rizzi. (2004). Underwater Color Constancy: Enhancement of Automatic Live Fish Recognition. Paper presented at the 16th Annual symposium on electronic imaging, United States.

(CMP sfm 2012) CMP SfM Web Service, Center for Machine Perception, FEE, CTU Prague, <http://ptak.felk.cvut.cz/sfmservice/>, accessed September 2012.

(Comas 2012) COMAS, <http://comas.unical.it>, (accessed on September 2012).

(CombineZM 2012) CombineZM, Alan Hadley
<http://hadleyweb.pwp.blueyonder.co.uk/CZP/News.htm>, (accessed on September 2012).

(Cornille *et al.* 2003) N. Cornille, D. Garcia, M.A. Sutton, S.R. McNeill, J.J. Orteu. Automated 3-D reconstruction using a scanning electron microscope. In: Proceedings of SEM 2003 annual conference on experimental and applied mechanics, Charlotte, North Carolina, USA, 2003.

(Criminisi *et al.* 2000) Antonio Criminisi, Ian D. Reid, and Andrew Zisserman. Single view metrology. *International Journal of Computer Vision*, 40(2):123–148, 2000.

(Crisci *et al.* 2010) G. Crisci, M. La Russa, M. Macchione, M. Malagodi, A. Palermo and S. Ruffolo, Study of archaeological underwater finds: deterioration and conservation, *Applied Physics A* (2010) 100: 855–863.

(Cronk *et al.* 2006) Cronk, S., Fraser, C.S. and Hanley, H.B. 2006. Automatic metric calibration of colour digital cameras. *The Photogrammetric Record*, 21(116): 355-372.

(Dellepiane *et al.* 2012) Matteo Dellepiane, Nicolò Dell’Unto, Marco Callieri, Stefan Lindgren, Roberto Scopigno, Archeological excavation monitoring using dense stereo matching techniques, *Journal of Cultural Heritage*, Available online 17 July 2012, ISSN 1296-2074.

(Drap *et al.* 2002) Drap P., Bruno E., Long L., Durand A., Grussenmeyer P.: Underwater photogrammetry and xml based documentation system: The case of the ‘Grand Ribaud F’ Etruscan wreck. In ISPRS COMMISSION V SYMPOSIUM Close- Range Imaging, Long-Range Vision (Corfu, Greece, 2-6 September 2002).

(Drap *et al.* 2007) Drap, P., Scaradozzi, D., Gambogi, P., Long, L. and Gauche, F., 2007. Photogrammetry for virtual exploration of underwater archeological sites. *XXI*

International Scientific Committee for Documentation of Cultural Heritage (CIPA) Symposium, Athens, Greece.

(Edge 2006) Edge, M., 2006. Chapter 5 in *Dome port theory*. Elsevier Ltd. page 314.

(El-Hakim & Beraldin 1995) El-Hakim, S. F. and Beraldin, J. A., 1995. Configuration design for sensor integration. Proceedings of SPIE Videometrics IV, Philadelphia, Pennsylvania, 22nd to 26th October 1995. Vol. 2598, 390 pages: 274–285.

(El-Hakim 2002) El-Hakim, S. F., 2002. Semi-automatic 3D reconstruction of occluded and unmarked surfaces from widely separated views. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 34(5): 143–148.

(Eltoukhy & Kavusi 2003) H. A. Eltoukhy, S. Kavusi. A Computationally Efficient Algorithm for Multi-Focus Image Reconstruction. Department of Electrical Engineering, Stanford University (2003).

(Evangelidis & Psarakis 2008) Evangelidis, G. D. and Psarakis, E. Z. (2008). Parametric image alignment using enhanced correlation coefficient maximization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(10):1858–1865.

(Faugeras & Keriven 1998) Olivier D. Faugeras and Renaud Keriven. Complete dense stereovision using level set methods. In *ECCV '98: Proceedings of the 5th European Conference on Computer Vision-Volume I*, pages 379–393, London, UK, 1998. Springer-Verlag.

(Faugeras *et al* 1993) Faugeras, O, Hotz, B., Mathieu, H., Vuille, T., Zhang, Z., Fua, P., Theron, E., Moll, L., Berry, G., Vuillemin, J., Bertin, P., Proy, C., 1993. Real time correlation-based stereo: Algorithm, implementations and applications, INRIA, Tech. Rep. RR-2013, 1993.

(Ficher *et al.* 2008) R. Ficher, B.T. Galeb, P.R: Yoder. Optical System Design, second edition SPIE Press- McGraw-Hill, 2008.

(Forster *et al.* 2004) B. Forster, D. Van de Ville, J. Berent, D. Sage, M. Unser. Complex wavelets for extended Depth of Field: A new Method for the Fusion of Multichannel Microscopy Images. *Microscopy Research and Technique*.(2004) 65 33-42.

(Fraser & Al-Ajlouni 2006) C. S. Fraser, S. Al-Ajlouni. Zoom-dependent camera calibration in digital close-range photogrammetry. *Photogrammetric Engineering & Remote Sensing*, 72(2006) 9 1017-1026.

(Fraser 1997) C. Fraser. Digital camera self-calibration. *ISPRS Journal of Photogrammetry and Remote Sensing*, 52(1997) 4 149-159.

(Fryer & Fraser 1986) Fryer, J. G. and Fraser, C. S., 1986. On the calibration of underwater cameras. *Photogrammetric Record*, 12(67): 73-85.

(Furukawa & Ponce 2007) Y. Furukawa, J. Ponce. Accurate, Dense, and Robust Multi-View Stereopsis. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2007.

(Georgopoulos *et al.* 2011) Georgopoulos A., Diamanti E., F.Vlachaki: Geometric documentation of underwater archaeological sites. In XXIIIrd International CIPA Symposium (Prague, Czech Republic, 12-16 September 2011).

(Ghita *et al.* 2005) Ovidiu Ghita, Paul F. Whelan, and John Mallon. Computational approach for depth from defocus. *Journal of Electronic Imaging*, 14(2), 2005.

(Girardi 2011) F. Girardi. Survey and 3D modeling for small sized objects. PhD thesis, University of Bologna, Italy, 2011.

(Grün 2000) Grün, A., 2000. Semi-automated approaches to site recording and modelling. *International Archives of Photogrammetry and Remote Sensing*, 33(5/1): 309–318.

(Guidi *et al.* 2003) Guidi, G., Beraldin, J.-A., Ciofi, S. and Atzeni, C., 2003. Fusion of range camera and photogrammetry : systematic procedure for improving 3-D models metric accuracy. *IEEE Transactions on Systems, Man, and Cybernetics*, 33(4): 667–676.

(Guttosh 2002) R. J. Guttosh, “Investigation of color aliasing of high spatial frequencies and edges for bayer-pattern sensors and foveon x3 direct image sensor,” tech. rep., Foveon, San Antonio, USA, 2002.

(Hannah 1974) Hannah, M. J. (1974). *Computer Matching of Areas in Stereo Images*. Ph.D. thesis, Stanford University.

(Harris & Stephens 1988) C. Harris, M. Stephens. A combined corner and edge detector. *Proceedings of the 4th Alvey Vision Conference*, 1988, pp 147–151.

(Hartley & Zisserman 2005) Richard Hartley and Andrew Zisserman. *Multiple View Geometry in Computer Vision*, volume 23. Cambridge University Press, New York, NY, USA, 2005.

(Harvey & Shortis 1995) Harvey, E. and Shortis, M., 1995. A system for stereo-video measurement of sub-tidal organisms. *Marine Technology Society Journal*, 29: 10–22.

(Healey & Kondepudy 1994) Healey, G. E. and Kondepudy, R. (1994). Radiometric CCD camera calibration and noise estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(3):267–276.

(Heliconsoft 2012) Helicon Focus, Heliconsoft Ltd., <http://www.heliconsoft.com>, (accessed on September 2012).

(Hirano *et al.* 2009) Hirano D, Funayama Y, Maekawa T. 3D shape reconstruction from 2D images. *Computer-Aided Design and Applications* 2009;6(5):701–10.

(Hirschmuller & Scharstein 2007) Heiko Hirschmuller and Daniel Scharstein. Evaluation of cost functions for stereo matching. In *CVPR*. IEEE Computer Society, 2007.

(Hoiem *et al.* 2005) Derek Hoiem, Alexei A. Efros, and Martial Hebert. Automatic photo pop-up. In *ACM SIGGRAPH*, August 2005.

(Horn & Brooks 1989) Horn, B.K.P. and Brooks, M.J. (Eds.), 1989. Shape from Shading. MIT Press, Cambridge, Massachusetts. 577 pages.

(Iqbal *et al.* 2007) Kashif Iqbal, Rosalina Abdul Salam, Azam Osman, & Abdullah Zawawi Talib. (2007). Underwater Image Enhancement Using an Integrated Colour Model. *IAENG International Journal of Computer Science*, Vol.:34.

(Jacobson *et al.* 2000) R. E. Jacobson, S. F. Ray, G. G. Attridge, N. R. Axford. The Manual of Photography: Photographic and Digital Imaging, 9th ed. Focal Press (2000).

(Kalia *et al.* 2011) Robin Kalia, Keun-Dong Lee, Samir B.V.R., Sung-Kwan Je, & Weon- Geun Oh. (2011, 9-11 Feb.). An analysis of the effect of different image pre-processing techniques on the performance of SURF: Speeded Up Robust Feature. Paper presented at the 17th Korea- Japan Joint Workshop on Frontiers of Computer Vision (FCV). pp.1-6.

(Kanade & Okutomi 1994) Kanade, T. and Okutomi, M. (1994). A stereo matching algorithm with an adaptive window: Theory and experiment. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(9):920–932.

(Kanade 1994) Kanade, T. (1994). Development of a video-rate stereo machine. In *Image Understanding Workshop*, pp. 549–557, Monterey.

(Kender 1981) Kender, J. R., 1981. Shape from Texture. Carnegie Mellon University, AAT 8114629. 222 pages.

(Lavest *et al.* 2000) Lavest, J. M., Rives, G., and Lapresté, J. T., 2000. Underwater Camera Calibration. *Lecture Notes in Computer Science*, 1843: 654-668.

(Leone *et al.* 2006) Leone, A., Diraco, G. and Distanto, C., 2006. Stereoscopic system for 3D seabed mosaic reconstruction. *International Conference on Image Processing, ICIP*: II541–II544.

(Levoy *et al.* 2000) Marc Levoy, Kari Pulli, Brian Curless, Szymon Rusinkiewicz, David Koller, Lucas Pereira, Matt Ginzton, Sean Anderson, James Davis, Jeremy Ginsberg, Jonathan Shade, and Duane Fulk. The digital michelangelo project: 3D scanning of large statues. In *Proceedings of ACM SIGGRAPH 2000*, pages 131–144, July 2000.

(Li *et al.* 1997) Li, R., Li, H., Zou, W., Smith, R. G. and Curran, T. A., 1997. Quantitative photogrammetric analysis of digital underwater video imagery. *IEEE Journal of Oceanic Engineering*, 22(2): 364-375.

(Liu *et al.* 2001) Z. Liu, K. Tsukada, K. Hanasaki, Y.K. Ho, Y.P. Dai. Image fusion by using steerable pyramid. *Pattern Recognit. Lett.*, 22(2001) 929–939.

(Liu *et al.* 2008) Liu, C., Szeliski, R., Kang, S. B., Zitnick, C. L., and Freeman, W. T. (2008). Automatic estimation and removal of noise from a single image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(2):299–314.

(Lowe 2004) D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2004) 2 91-110.

(Mahiddine *et al.* 2012) Mahiddine A., Seinturier J., Peloso D., Boï J.-M., Drap P., Merad D.: Underwater image pre-processing for automated photogrammetry in high turbidity water. In *VSMM2012* (Milano, Italy, 2012).

(Meline *et al.* 2010) Meline A., Triboulet J., Jouvencel B.: A camcorder for 3d underwater reconstruction of archeological objects. In *OCEANS' 10* (Seattle, USA, 2010), pp. 1- 9.

(Menci 2012) Menci software, <http://www.menci.com> (accessed on September 2012).

(Montgomery 2009) Montgomery, Douglas C. (2009). Design and Analysis of Experiments (7th Edition). John Wiley & Sons.

(My3DScanner 2012) My3DScanner, <http://my3dscanner.com/>, (accessed on September 2012).

(Nayar & Nakagawa 1990) S. K. Nayar, Y. Nakagawa. Shape from focus: an effective approach for rough surfaces. In Proceedings of IEEE International Conference on Robotics and Automation, Cincinnati, OH. 1990 pp 218-225.

(Pavlidis *et al.* 2007) G. Pavlidis, A. Koutsoudis, F. Arnaoutoglou, V. Tsioukas, C. Chamzas. Methods for 3D digitization of Cultural Heritage. Journal of Cultural Heritage 8 (2007) 93-98.

(Pentland 1982) A. P. Pentland. Depth of Scene from Depth of Field. Proceedings of the Image Understanding Workshop, 1982 pp. 253-259.

(Petit 2010) Frédéric Petit. (2010). Traitement et analyse d'images couleur sous-marines : modèles physiques et représentation quaternionique. Doctorat, Sciences et Ingénierie pour l'Information, Poitier.

(Pieper & Korpel 1983) R. J. Pieper, A. Korpel. Image processing for extended depth of field. Appl. Opt. 22(1983) 10 1449–1453.

(Pietraggi & Davidde 2007) Petriaggi R., Davidde B.: Restauration subaquatique: le bilan de cinq années de travaux expérimentaux de l'institut central pour la restauration. In IVèmes Rencontres Internationales Monaco et la Méditerranée (Monaco, France, 22-24 Mars 2007), pp. 105–116. 3

(Pietraggi & Mancinelli 2004) Petriaggi R., Mancinelli R.: An experimental conservation treatment on the mosaic floor and perimeter walls of room n.1 of the

so-called 'villa con ingresso a protiro' in the underwater archaeological park of Baia. *Archaeologia Maritima Mediterranea* 1 (2004)

(Pollefeys *et al.* 2004) Marc Pollefeys, Luc Van Gool, Maarten Vergauwen, Frank Verbiest, Kurt Cornelis, Jan Tops, and Reinhard Koch. Visual modeling with a hand-held camera. *Int. J. Comput. Vision*, 59(3):207–232, 2004.

(Rapidform 2012) Rapidform, INUS Technology Inc., <http://www.rapidform.com>, (accessed on September 2012).

(Remondino & El-Hakim 2006) F. Remondino, S. El-Hakim. Image-based 3D modeling: A review. *Photogrammetric Record*. 21(2006) 115 269–291.

(Rizzi & Gatta 2004) Alessandro Rizzi, & Carlo Gatta. (2004). From Retinex to Automatic Color Equalization: issues in developing a new algorithm for unsupervised color equalization. *Journal of Electronic Imaging*, Vol.:13, pp.75-84.

(Sablatnig & Menard 1997) Sablatnig, R. and Menard, C., 1997. 3D reconstruction of archaeological pottery using profile primitives. *Proceedings of the International Workshop on Synthetic–Natural Hybrid Coding and Three-Dimensional Imaging*, Rhodes, Greece, 5th to 9th September 1997. 290 pages: 93–96.

(Saito *et al.* 1995) Saito, H., Kawamura, H. and Nakajima, M., 1995. 3D shape measurement of underwater objects using motion stereo. *IECON Proceedings (Industrial Electronics Conference)*: 1231–1235.

(Salvi *et al.* 2004) Salvi, J., Pages, J. and Batlle, J., 2004. Pattern codification strategies in structured light systems. *Pattern Recognition*, 37: 827–849.

(Saxena *et al.* 2008) Ashutosh Saxena, Sung H. Chung, and Andrew Y. Ng. 3-d depth reconstruction from a single still image. *International Journal Computer Vision*, 76(1):53–69, January 2008.

(Scansystems 2012) DentalScanner, Scansystems, <http://www.scansystems.it>, (accessed on September 2012).

(Scharstein & Szeliski 2002) D. Scharstein, R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47(2002) 1/2/3 7-42.

(Schechner & Kiryati 2000) Y. Y. Schechner, N. Kiryati. Depth from Defocus vs. Stereo: How Different Really are They? *International journal of Computer Vision*, 39(2000) 2 141-162.

(Schettini & Corchs 2000) Schettini R., Corchs S.: Imaging for underwater archaeology. *American Journal of Field Archaeology* 27, 3 (2000), 319–328.

(Schewe *et al.* 1996a) Schewe H., Moncreiff E., Gruendig L.: Quantitative photogrammetric analysis of digital underwater video imagery. *International Archives of Photogrammetry and Remote Sensing* 31, 5 (1996), 524–529

(Schewe *et al.* 1996b) Schewe, H., Moncreiff, E. and Gruendig, L., 1996. Improvement of fish farm pen design using computational structural modelling and large-scale underwater photogrammetry. *International Archives of Photogrammetry and Remote Sensing*, 31(5): 524-529.

(Schreier *et al.* 2004) H. W. Schreier, D. Garcia, M.A. Sutton. Advances in light microscope stereo vision *Exp. Mech.* 44(2004) 278-88.

(Seitz *et al.* 2006) Steven M. Seitz, Brian Curless, James Diebel, Daniel Scharstein, and Richard Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *CVPR '06: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 519–528, Washington, DC, USA, 2006. IEEE Computer Society.

(Shum & Kang 2000) Shum, H.-Y. and Kang, S. B., 2000. A review of image-based rendering techniques. *Proceedings of IEEE/SPIE Visual Communications and Image Processing (VCIP)*, Perth, Australia, 21st June 2000. SPIE Vol. 4067, 1680 pages: 2–13.

(Singh *et al.* 2010) Singh H., Adams J., Foley B. P., Mindell D.: Underwater image processing: State of the art of restoration and image enhancement methods. *EURASIP Journal on Advances in Signal Processing* 2010 (2010), 1–14. 2

(Skarlatosa *et al.* 2012) Skarlatosa D., Demestihab S., Kiparissia S.: An “open” method for 3d modelling and mapping in underwater archaeological sites. *International Journal of Heritage in the Digital Era* 1, 1 (2012), 2–24.

(Snavely *et al.* 2006) Noah Snavely, Steven M. Seitz, and Richard Szeliski. Photo tourism: Exploring photo collections in 3d. In *SIGGRAPH Conference Proceedings*, pages 835–846, New York, NY, USA, 2006. ACM Press.

(Snavely *et al.* 2007) N. Snavely, S.M. Seitz, R. Szeliski. Modeling the World from Internet Photo Collections. *International Journal of Computer Vision* (2007).

(Sperling 1970) G. Sperling. Binocular vision: A physical and a neural theory. *Amer. J. Psychol.* 83(1970) 461-534.

(Stamatopoulos *et al.* 2010) C. Stamatopoulos, C. Fraser, S. Cronk. On the self-calibration of long focal length lenses. *ISPRS, XXXVIII, Part 5 Commission V Symposium*, Newcastle upon Tyne, UK, 2010.

(Sturm & Triggs 1996) P. Sturm and B. Triggs. A factorization based algorithm for multi-image projective structure and motion. In *ECCV*, pages 709–20, Cambridge, England, apr 1996. Springer-Verlag.

(Subbarao *et al.* 1993) M. Subbarao, T. Chio, A. Nikzad. Focus techniques. *Optical engineering*, (1993) 2824-2836.

(Svoboda *et al.* 2005) Tomáš Svoboda, Daniel Martinec, and Tomáš Pajdla. A convenient multi-camera selfcalibration for virtual environments. *PRESENCE: Teleoperators and Virtual Environments*, 14(4):407–422, August 2005.

(Telem & Filin 2010) Telem G., Filin S.: Photogrammetric modeling of underwater environments. *ISPRS journal of photogrammetry and remote sensing* 65, 5 (2010), 433–444.

(Tenenbaum 1971) J. M. Tenenbaum. Accomodation in computer vision. PhD dissertation, Standford University, 1971.

(Tiano *et al.* 2009) P. Tiano, P. Salvo, F. Ceccaroni. The microphotogrammetry as a portable diagnostic tool for monumental surface monitoring. In *DGaOProceedings*, 2009.

(Tombari *et al.* 2008) F. Tombari, S. Mattoccia, L. Di Stefano, E. Addimanda, “Classification and evaluation of cost aggregation methods for stereo correspondence”, *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR 2008)*, 2008.

(Tonazzini *et al.* 2004) Anna Tonazzini, Emanuele Salerno, Matteo Mochi and Luigi Bedini. Blind Source Separation Techniques for Detecting Hidden Texts and Textures in Document Images. *IMAGE ANALYSIS AND RECOGNITION Lecture Notes in Computer Science*, 2004, Volume 3212/2004, 241-248.

(Triggs 1996) B. Triggs. Factorization methods for projective structure and motion. In *CVPR*, San Francisco, June 1996.

(Triggs *et al.* 1999) B. Triggs, P. F. McLauchlan, R. Hartley, A. Titzgibbon. Bundle adjustment – a modern synthesis. *Proceedings of the International Workshop on Vision Algorithms: Theory and Practice*, Corfu, September 1999.

(Tsai 1992) Roger Y. Tsai. A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. *Radiometry*, pages 221–244, 1992.

(Tsin *et al.* 2001) Tsin, Y., Ramesh, V., and Kanade, T. (2001). Statistical calibration of CCD imaging process. In *Eighth International Conference on Computer Vision (ICCV 2001)*, pp. 480–487, Vancouver, Canada.

(Unesco 2011) <http://www.unesco.org/new/en/culture/themes/underwater-cultural-heritage/>. (accessed on September 2012).

(Valdecasas *et al.* 2001) A. G. Valdecasas, D. Marshall, J. M. Becerra, J. J. Terrero. On the extended depth of focus algorithms for bright field microscopy. *Micron* 32(2001) 6 559-569.

(Verhoeven 2011) Geert Verhoeven. Taking computer vision aloft – archaeological three-dimensional reconstructions from aerial photographs with PhotoScan. *Archaeological Prospection* Volume 18, Issue 1, pages 67–73, January/March 2011.

(Wall 1984) Wahl, F. M., 1984. A coded light approach for 3-dimensional (3D) vision. IBM Research Report, RZ 1452, 8 pages.

(Wallis 1976) Wallis, K.F. (1976) Seasonal adjustment and relations between variables *Journal of the American Statistical Association*, 69(345) pp. 18-31.

(Wiley & Wong 1995) A. G. Wiley, K.W. Wong. Geometric calibration of zoom lenses for computer vision metrology. *Photogrammetric Engineering & Remote Sensing* 61(1995) 1 69-74.

(Yanagi & Chikatsu 2010) H. Yanagi, H. Chikatsu. Performance evaluations of macro lenses for digital documentation of small objects. *SPIE Videometrics, Range Imaging, and Applications X Proceedings Paper*, San Diego, CA, USA, 2010.

(Yedidia *et al.* 2000) Jonathan S. Yedidia, William T. Freeman, and Yair Weiss. Generalized belief propagation. In *NIPS*, pages 689–695, 2000.

(Yeo *et al.* 1993) T. T. E. Yeo, S. H. Ong, Jayasooriah, R. Sinniah. Autofocusing for tissue microscopy. *Image and Vision Computing*, 11(1993) 10 629-639.

(Zeiss 1997) C. Zeiss. "Depth of Field – An Insider's Look". *Camera Lens News* #1, 1997.

(Zerene Systems 2012) Zerene Stacker, Zerene Systems LLC, <http://zerenesystems.com>, (accessed on September 2012).

(Zhang 2000) Zhengyou Zhang. A flexible new technique for camera calibration. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(11):1330–1334, 2000.

(Zhang *et al.* 2003) Li Zhang, Brian Curless, and Steven M. Seitz. Spacetime stereo: Shape recovery for dynamic scenes. *cvpr*, 02:367, 2003

(Zubrow 2006) E. Zubrow, Digital Archaeology A historical context, in Digital Archaeology bridging method and theory, edited by Thomas Evans and Patrick Daly, 2006 pp. 10–31.