POLITECNICO
MILANO 1863

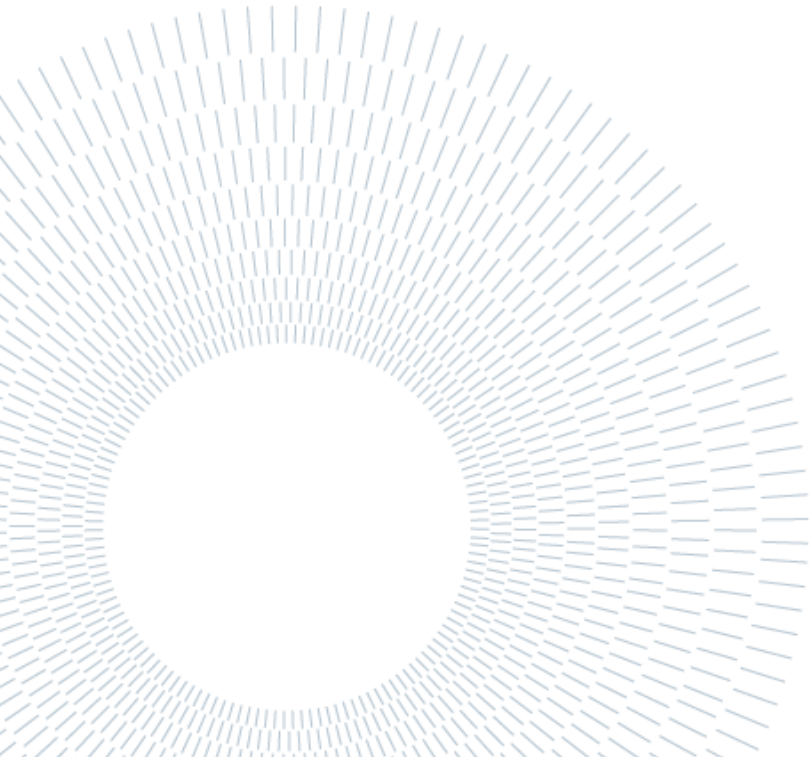
SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

RGB-D Digital Image Correlation technique for crack assessment through homography-based movement compensation on drones

TESI DI LAUREA MAGISTRALE IN
MECHANICAL ENGINEERING
INGEGNERIA MECCANICA

Authors: **Carlo Folcio, Luigi Rovelli**

Students ID: 969758, 968654
Advisor: Emanuele Zappa
Co-advisor: Daniele Marchisotti, Simone Paganoni
Academic Year: 2021-22



POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

EXECUTIVE SUMMARY OF THE THESIS

RGB-D Digital Image Correlation technique for crack assessment through homography-based movement compensation on drones

TESI MAGISTRALE IN MECHANICAL ENGINEERING – INGEGNERIA MECCANICA

AUTHORS: CARLO FOLCIO, LUIGI ROVELLI

ADVISOR: EMANUELE ZAPPA

COADVISORS: DANIELE MARCHISOTTI, SIMONE PAGANONI

ACADEMIC YEAR: 2021-2022

1. Introduction

The thesis addresses an RGB-D digital image correlation [1] approach for crack measurements with sensors mounted on a moving reference (ex. drone). The aim of the thesis is to develop an algorithm capable of compensating the movements of the drone through homographies [2], thus providing accurate displacement results. The RGB-D sensor is composed by an RGB or greyscale camera (Flir camera) and by a depth sensor (Time of Flight – Blaze 101).

The capabilities of the approach are firstly tested with a simulator and then in a real laboratory application, showing promising results.

2. Algorithm description

On the acquired in motion images, 2D DIC is performed. With the depth information coming from the Time of Flight (ToF) sensor, it is possible to build the centre subset point clouds (x_{3D}, y_{3D}, z_{3D}) for each frame, thanks to the pin-hole model (Eq.2.1):

$$\begin{cases} x_{3D} = (x_{pix} - c_x) * \dfrac{z_{3D}}{f_x} \\ y_{3D} = (y_{pix} - c_y) * \dfrac{z_{3D}}{f_y} \end{cases} \qquad 2.1$$

Where:

- x_{pix} and y_{pix} are the coordinates of each centre subset coming from DIC software (respectively for x-coordinates and y-coordinates);
- f_x, f_y, c_x, c_y are the intrinsics parameters of the camera (f: focal length (in pixels), c: optical centre (in pixels));
- z_{3D} is the z coordinate of each centre subset coming from the depth sensor.

The movement compensation is performed on an approximately fixed part of the measurand, if present, or on the entire set of points to remove the

average displacement and leave the deformations. It is always done between a defined frame (reference frame) and all the others, to report all the data to the same reference.

Three different solutions for the camera movement compensation are evaluated:

- homography estimation with calculation based on the realignement of 3D point clouds of the centre subsets (approach recalled 'H 3D');

The use of homographies limits the application to approximately plane surfaces of the object of interest. This choice is made to average the effects of noisy depth to the best fitting plane, since ToF sensors have a poor accuracy, of the order of magnitude of millimeters [3].

The calculation of the homography is done according to Eq.2.2:

$$H = K * \left(R - t * \frac{n}{d}\right) * inv(K) \qquad 2.2$$

where the parameters are:

-K is the intrinsic parameters matrix in the form presented here below:

$$K = \begin{bmatrix} f_x & 0 & 0 \\ 0 & f_y & 0 \\ c_x & c_y & 1 \end{bmatrix}$$

-R and t are the rotation matrix and the translation vector which bring the reference frame XYZ coincident with the successive frame X'Y'Z;

-n is the normal of the plane evaluated from the reference frame XYZ;

-d is the known term of the plane ax + by + cz + d = 0 evaluated from the reference frame XYZ;

The inverse of this homography is applied to 2D homogenous coordinates of the centre subsets of the successive frames to report data to the reference frame.

- homography estimation based on the realignment of 2D set of points of the centre subsets (approach recalled 'H 2D');

The estimation is based on the MSAC algorithm, that, with a minimization procedure, calculates the induced best-fitting homography between reference and moved views. This allows to minimize eventual issues related to an imperfection in the depth estimation, because it is performed on 2D homogenous coordinates of the centre subset points. At the same time, since the algorithm is based on a projective trasformation, some deformations effects can be confounded as projective effects by the minimization algorithm, thus causing their undesired compensation.

- rototranslation estimation from point clouds of the centre subsets (approach recalled 'PC 3D');

The direct application of the rototranslation to the point clouds allows to work with objects that aren't strictly planar and provide information also on the third coordinate (z), but the noise effects of the ToF can induce inaccurate results.

Knowing the centre subset point clouds in the space also allows to apply a final transformation to report these points perpendicular to the optical axis of the camera. In this way, the misalignment of the object with respect to the axis of the camera is corrected.

After the above mentioned compensations, actual displacements and deformations are retrieved.

The outputs of the algorithm in case of the presence of a fixed part and a moving one for the x displacement is represented in Figure 2.1. The same output is obtainable for y (and z in case of 'PC 3D').

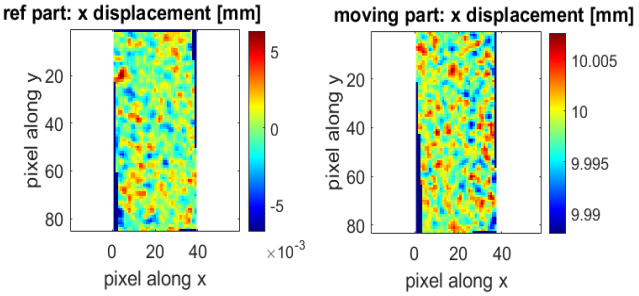


Figure 2.1: final output of the algorithm for reference and moving part

3. Algorithm testing methodology

In this paragraph the validation of the algorithm with the simulator and the experimental test are presented.

3.1. Numerical simulator validation

The algorithm is tested in a developed simulation environment to assess the ideal capabilities of the approach. The simulator generates greyscale images, acquired with a virtual moving camera, and corresponding depth maps. The speckle image (an example is shown in Figure 3.1) is divided in a fixed part, used for the movement compensation, and a moving part on which the displacement is tested. The moving part position is imposed in the simulator. The displacement between fixed and moving part simulate the crack opening.

The algorithm is tested with 7 different reference images (where the moving part displacement is set to 0) with diverse rototranslation values to validate the realignment perpendicular to the optical axis of the camera. Moreover, a set of 40 images with controlled crack size (moving part displacement) and defined rototranslation are considered to study the camera movement compensation.

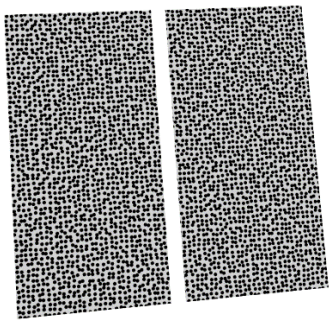


Figure 3.1: greyscale image obtained through the simulator

The results of the simulation tests are satisfying. The obtained uncertainty with all the 3 approaches for the camera movement compensation is, in average, below 2-3 hundredths of pixels, comparable with the intrinsic uncertainty of the standard DIC procedure [4]. This validates the proposed algorithm.

Moreover, the effects of the increase of the angles of rotations of the camera and of the increase of the crack size are assessed. An approximately linear trend for the average displacement value is present when the pitch and yaw angles increase but contained in a range comparable with the uncertainty of the measurement. Also, the standard deviation increases when the rotation is higher but not of a significant amount.

3.2. Experimental validation of the technique

After the simulator validation, an experiment is conducted to evaluate the approach in a real scenario as well as to understand the accuracy reachable with the sensors considered.
The laboratory experiment consists in the dynamical analysis of the crack displacement of a xps panel (shown in Figure 3.2). The pre-cracked panel, on which a speckle is painted, is cyclically loaded with a three point bending system. The RGB-D system is handheld to simulate the movements of the drone. Its results are compared to the data coming from a 3D DIC system that is considered as the ground truth.

Figure 3.2: xps pre-cracked panel with painted speckle

The 'PC 3D' approach is discarded due to the fact that the accuracy of the Blaze is ±5 mm and because the panel is approximately flat.
The 'H 2D' and 'H 3D' approaches are tested against the DIC 3D. In correspondence of the crack two zones, one on the left and one on the right side of the crack are considered. The relative displacement between right and left is calculated

in the time. The derived time histories describing the crack opening are compared here below in Figure 3.3

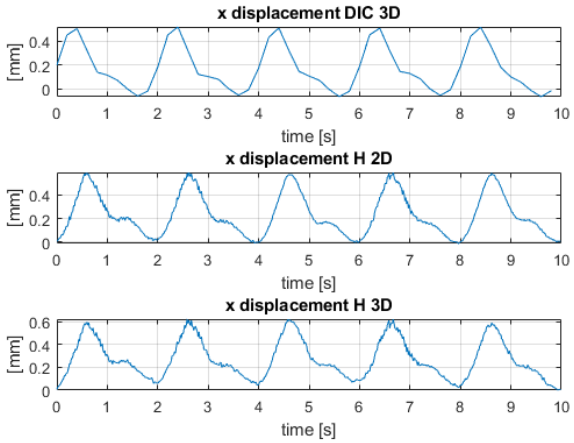


Figure 3.3: time histories retrieved with DIC 3D, H 2D and H 3D methods.

The perfect synchronization between the 3D DIC system and the RGB-D one was not possible. To compare more precisely the results, the normalized cross correlation between the two approaches with the 3D DIC time histories is performed. The cross correlation peak retrieval allows to realign the time histories.

Visually, the superimposition of the time histories is satisfying. The RMS error is calculated (Eq. 3.1) on the realigned graphs (Figure 3.4) to compare the two approaches (results shown in Table 3.1).

$$RMS_{error} = \sqrt{\frac{\sum_i (x(H)_i - x(3D\ DIC)_i)^2}{n-1}} \qquad 3.1$$

	H 3D	H 2D
RMS error	0.038 mm	0.032 mm

Table 3.1: RMS error for H 3D and H 2D

'H 2D' seems to perform slightly better than 'H 3D' in terms of Root Mean Square value. This is not so significant to define the best technique since the calculation is valid only locally, in correspondence of the crack. Moreover, the values are very small and similar, comparable with the uncertainty of the measurement approach.

Considering all the displacement field of the panel, it is possible to see that qualitatively (according to Figure 3.5) the 'H 3D' approach behaves far better than the 'H 2D' one. The homography calculation on 2D points loses information in different parts of the panel. The deformations effects are confounded as projective movements of the camera by the minimization algorithm. The application of the derived homography causes their undesired compensation.

Concluding that, in the experimental scenario, with the addition of non perfectly planar object and of the deformations of the panel, the 'H 3D' approach results more robust compared to the 'H 2D' one. The calculation of the homography passing through 3D point clouds data allows to

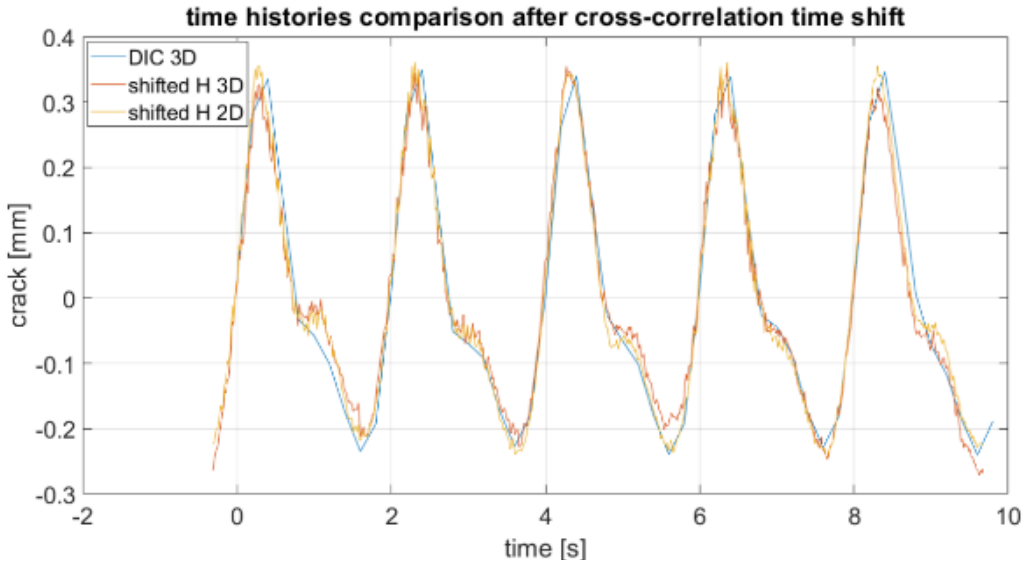


Figure 3.4: graph with the realignement of the time histories for the three different methods (DIC 3D, H 2D and H 3D)

avoid the 2D problem of confounding deformations and shape effects as projective effects by the 2D minimization algorithm. At the same time, the actual degree of uncertainty associated to ToF sensors does not permit to work directly with rototranslation applied to point clouds of centre subset points ('PC 3D'). The use of homography with the plane assumption minimizes the effects of ToF noise, providing accurate results.

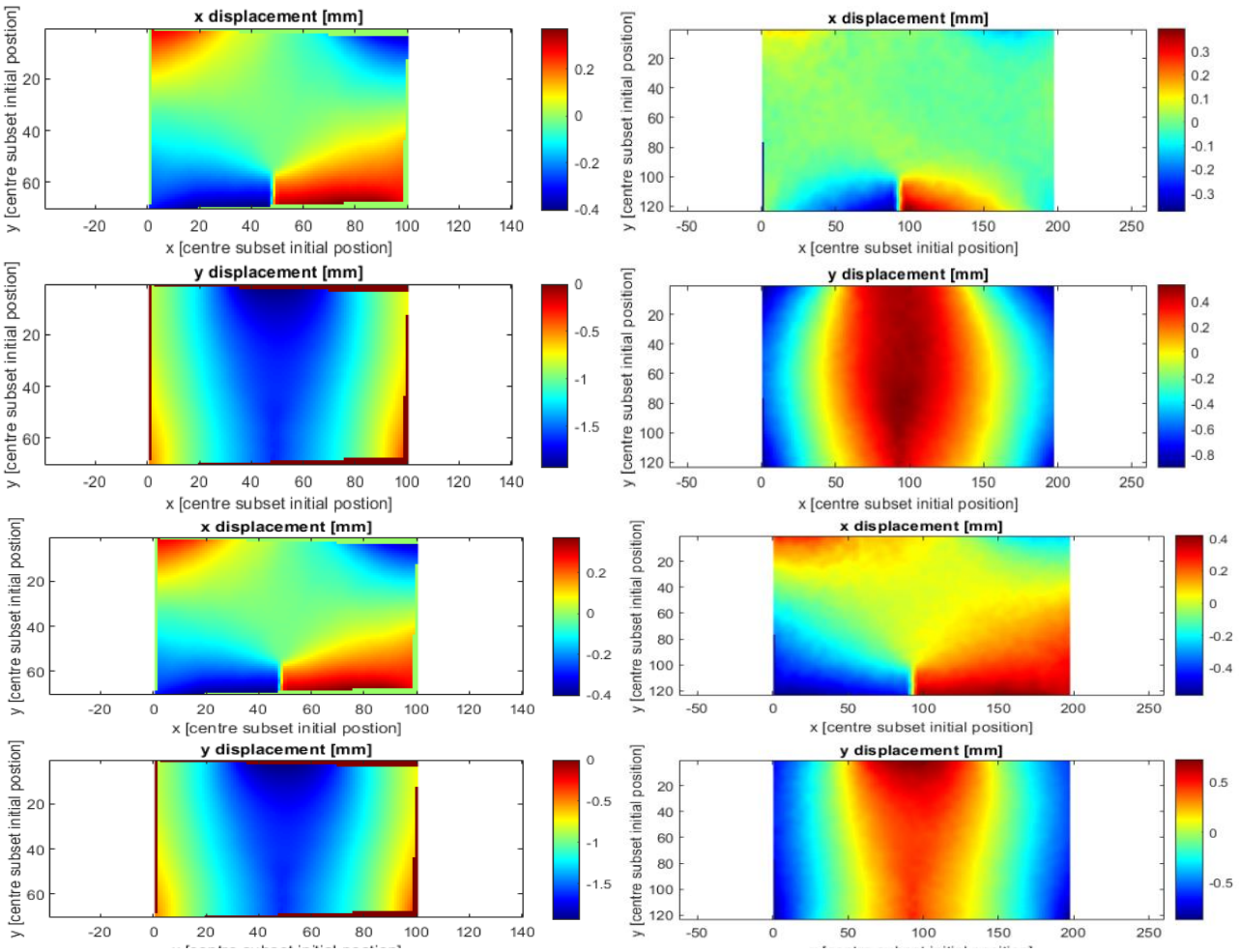


Figure 3.5: Final output with data from DIC 3D on the left part and from H 2D and H 3D respectively on the upper right (2 plots, one for x and one for y) and bottom right part (2 plots, one for x and one for y)

Conclusions

RGB-D digital image correlation with sensors mounted on a drone is an innovative technique that allows crack assessments on critical places (for example bridge decks), not easily reachable by men. The RGB-D sensor is suitable for drone transportation compared to a 3D DIC system, as well as much less expensive.

The implemented algorithm demonstrated to effectively compensate the drone movements, providing accurate information in the simulation environment.

In the laboratory experiment both the compensation with homographies from 3D rototranslation and from 2D centre subset points reconstructed accurately the local time history of the crack. The RMS error was contained below 0.04 mm against the ground truth 3D DIC in the two cases.

Instead, considering the entire deformation field of the panel, the approach based on the realignment with homographies calculated starting from the rototranslation of 3D centre subset point clouds was better. The calculation of the homographies on the 2D data brought a problem to the surface. In a real application the effects of deformations and not perfectly planar objects can be confounded as projective effects by the minimization algorithm, thus causing their undesired compensation. Working with a depth sensor and with point clouds allow to avoid this mistake.

4. References

[1] J. M. Franco, J. M. Caicedo, J. Marulanda, M. Sutton, and P. Thomson, "RGB-D-DIC technique for low-cost 3D displacement fields measurements", *Eng Struct*, vol. 197, Oct. 2019, doi: 10.1016/j.engstruct.2019.109457.

[2] G. Chen, Q. Liang, W. Zhong, X. Gao, and F. Cui, "Homography-based measurement of bridge vibration using UAV and DIC method", *Measurement (Lond)*, vol. 170, Jan. 2021, doi: 10.1016/j.measurement.2020.108683.

[3] D. Marchisotti and E. Zappa, "Uncertainty mitigation in drone-based 3D scanning of defects in concrete structures", 2022 *IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, 2022, doi: 10.1109/I2MTC48687.2022.9806652.

[4] D. Reagan, A. Sabato, and C. Niezrecki, "Feasibility of using digital image correlation for unmanned aerial vehicle structural health monitoring of bridges", *Struct Health Monit*, vol. 17, no. 5, pp. 1056–1072, Sep. 2018, doi: 10.1177/1475921717735326.

Abstract

The thesis presents an RGB-D-DIC measurement approach with sensors mounted on a moving reference (ex. UAVs) to perform crack assessments in critical areas. The developed algorithm compensates the effects of the reference movement by realigning centre subset coordinates on a defined approximately fixed part of the measurand. This allows to retrieve the real displacement field not corrupted by the drone movement. The movement compensation is done on centre subset point clouds through rototranslations or on centre subset homogenous coordinates via homographies. The coupling of a greyscale camera (Flir) with the Time of Flight (ToF) depth sensor (Blaze) also gives information about the relative position of the centre subset point clouds and of the RGB-D sensor, thus allowing a final correction to report data perpendicular to the optical axis of the camera to provide accurate results. The measurement approach is firstly validated through a simulator that works with ideal greyscale and depth images, giving satisfying results both with point clouds and homographies realignments. It is then tested on real laboratory application where a three point bending system cyclically loads a pre-cracked panel. The RGB-D sensor is handheld to simulate drone movements. In this case the movement compensation is performed only via homographies due to the poor accuracy of the ToF sensor, mediating its noisy effects on a fitted plane (homographies can only work with plane objects). The time history of the crack displacement is retrieved based on digital image correlation corrected data. A qualitative and quantitative comparison of the results is performed against a DIC 3D system considered as the ground truth model.

Key-words: DIC, RGB-D, drone, homography, point cloud, crack, bridge.

Abstract in italiano

Questa tesi propone un approccio di misura RGB-D-DIC con sensori montati su un riferimento in movimento (es. UAV) per eseguire valutazioni di cricche in aree critiche non facilmente raggiungibili dall'uomo. L'algoritmo sviluppato compensa gli effetti del movimento del riferimento riallineando le coordinate dei centri subset su una parte approssimativamente fissa del misurando. Ciò consente di recuperare il campo di spostamento reale non corrotto dal movimento del drone. La compensazione del movimento viene effettuata sulle nuvole di punti dei centri subset mediante rototraslazioni o sulle coordinate omogenee dei centri subset mediante omografie. L'accoppiamento di una telecamera in scala di grigi (Flir) con il sensore di profondità a tempo di volo (ToF - Blaze) fornisce anche informazioni sulla posizione relativa delle nuvole di punti dei centri subset e del sensore RGB-D, consentendo così una correzione finale per riportare i dati perpendicolari all'asse ottico della camera e fornire risultati accurati. L'approccio di misura viene prima validato attraverso un simulatore che lavora con immagini ideali in scala di grigi e di profondità, fornendo risultati soddisfacenti sia con riallineamenti basati su nuvole di punti 3D che con compensazioni omografiche. L'algoritmo viene poi testato su un'applicazione reale di laboratorio con un sistema di flessione a tre punti che carica ciclicamente un pannello pre-criccato. Il sensore RGB-D è sorretto a mano per simulare i movimenti del drone. In questo caso la compensazione del movimento viene eseguita solo tramite omografie, a causa della scarsa accuratezza del sensore ToF, in modo da mediare il rumore su un piano interpolato (le omografie possono lavorare solo con oggetti piani). La storia temporale dell'apertura di cricca viene calcolata tramite i dati compensati provenienti dalla digital image correlation. Viene eseguito un confronto qualitativo e quantitativo dei risultati con un sistema DIC 3D preso come riferimento di misura.

Key-words: DIC, RGB-D, drone, omografia, nuvola di punti, cricca, ponte.

Contents

Abstract ..... **i**

Abstract in italiano ..... **ii**

Contents ..... **iii**

Introduction **1**

1 State of the art **5**

2 Proposed approach **9**

2.1. Digital Image Correlation 10

2.2. Acquired data manipulation 12

2.3. Compensation of camera movement 14

2.3.1. 'H 3D': homography from 3D point clouds 14

2.3.2. 'H 2D': homography from 2D greyscale images 16

2.3.3. 'PC 3D': Rototranslation from point cloud 17

2.4. Realignement perpendicular to the optical axis of the camera 17

2.5. Data representation 20

3 Numerical simulator development and validation **25**

3.1. Simulator structure and implementation 25

3.1.1. Parameters definition 26

3.1.2. Speckle and greyscale images generation 28

3.1.3. Simulator camera calibration procedure 30

3.1.4. Depth maps generation 32

3.2. Design of numerical tests 36

3.2.1. Reference images 36

3.2.2. Fixed crack dimension and moving camera images 38

3.2.3. Fixed camera and increasingly crack dimension images 40

3.3. Numerical tests results 41

3.3.1. Fixed crack dimension and moving camera results 41
3.3.2. Fixed camera and increasingly crack dimension results 51
4 Experimental validation of the technique 53
4.1. Experimental setup description 54
4.1.1. Three point bending system 54
4.1.2. Ground truth 3D DIC reference system 56
4.2. Measurement system of the technique 58
4.2.1. RGB-D sensor assembly 58
4.2.2. RGB-D sensor calibration procedure 61
4.3. Practical implementation of the algorithm 63
4.3.1. Further developments for real world application 64
4.3.2. ROI definition for movement compensation in real application . 67
4.4. Experimental results 71
4.4.1. Qualitative analysis 72
4.4.2. Quantitative analysis 74
5 Conclusion and future developments 83
Bibliography 85

Introduction

In this thesis an approach to perform crack measurements with the coupling of a greyscale camera and a depth sensor mounted on a moving reference (for example a drone) is proposed. The combination of the two sensors with the digital image correlation allows to compensate the movement of the reference, thus obtaining accurate measurement results. The proposed algorithm is validated through a simulator, working with ideal greyscale and depth images. A laboratory experiment is conducted to evaluate the capabilities on a real application. The realignment to the reference position is, in a real scenario, done with homographies to reduce the effect of the depth sensor variability. The use of homographies limits the crack measurement procedure to planar surfaces.

2D digital image correlation allows to get displacement and strain information on a plane surface orthogonal to the optical axis of the camera. It is based on the comparison between a reference digital image with other images representing the deformed state in time. The analysis is carried out defining subsets of the reference image and calculating the position of each subset in the following ones, calculation done with a numerical iterative method based on correlation algorithm. The displacement of the subsets is represented as a displacement field of the object of interest, from which strains can be calculated. To retrieve each subset in the images a proper speckle random pattern on the measurand is needed. Nowadays the accuracy of the technique reaches the sub-pixel level thanks to interpolation algorithms. The 2D DIC measurement presents a lot of positive aspects: it is contactless, retrieves non uniform full field deformations with great resolution and it is relatively easy and cheap compared to other methods. In fact, strain gauges focus only on pointwise measurements while other contactless optical techniques as the interferometric ones result more complicated and expensive. 2D DIC also suffers some disadvantages, mainly regarding the fact that the measurement depends heavily on the quality of the imaging system and that the object must be planar and kept perpendicular to the optical axis of the camera during the entire procedure. Also, the camera must remain still because its movement corrupts the results.

To solve the out of plane movement problem, to work with non planar target and to correctly estimate the displacement and deformation fields in 3D (thus including the out of plane direction), 3D DIC was invented. 3D DIC is performed with 2 cameras pointing the measurand object from different positions. From each camera perspective the displacement map is calculated, then results are matched thanks to stereo calibration data and triangulation procedure. In this way it is possible to know the entire shape of the object (undeformed shape) and all the 3D displacements and strains. 3D DIC is more complicated in the setup. It needs more space for the placement of the cameras and a trigger system for simultaneous acquisition. Even in this case cameras must not be moved (with respect to each other, rigid movement is instead allow) not only during acquisition but even after the stereo calibration.

To perform digital image correlation in the 3D space other solutions are emerging. The coupling of a RGB or greyscale camera with a depth sensor allows to calculate displacement with 2D DIC and add the information about the depth with, for example, a Time of Flight (ToF) sensor, thus taking out 3D data with a much portable and small system. The 2 sensors can be placed as near as possible contrary to the cameras in 3D DIC. The ToF sensor calculates depth images based on the round-trip time of an artificial emitted light coming from the sensor itself. This technique is not free from problems. First of all, the simultaneous acquisition between the 2 sensors is not always guaranteed or, to the knowledge of the authors, not developed in a standardized way up to now. RGB and ToF cameras usually work with different framerate. Interpolation between depth images can reduce the uncertainty related to this aspect, that still remains not negligible. Moreover, ToF sensors provide nowadays an accuracy on the order of magnitude of millimeters. Depending on the application, results could be much worse compared with 3D DIC, for example in cases in which the analysis is performed on very small objects or the desired uncertainty is less than 1 millimeter. At the other extreme, ToF sensors have a working field limited, usually around 10 meters, so far away measurements are not possible.

DIC measurements in critical areas such as bridges, skyscrapers, high temperature or corrosive environments can't be directly supervised by men. The placing of the cameras and of the various instrumentation needed cannot be done directly by a person. An emerging idea nowadays involves the use of UAV systems with the instrumentation mounted on it. The reference is not fixed so the movements of the drone must be either neglected or compensated. The payload of drones do not allow heavy cameras mounting. In some studies ([1], [2]) 3D stereo cameras placed on the

drones are proposed, but the entire system results cumbersome due to the needed angle between cameras.

In this thesis an approach with RGB-D-DIC mounted on UAVs is proposed. The RGB-D system is much more suitable for drone transportation due to reduced dimensions and weight. Moreover, the cost of the RGB-D sensor is much less than the stereo system. The movements of the drone are corrected by considering an approximately fixed part of the framed object. Point clouds coming from depth images of successive frames are realigned to the reference one either with a rototranslation or with a homography calculated on the fixed part. The RGB-D sensor also allows to report the coordinates perpendicular to the optical axis of the camera. What remains is the actual displacement map. The proposed algorithm is firstly validated with a simulator in which ideal acquisition of greyscale and depth images are made and later also with an experimental test in a real world application.

The focus of the algorithm for the experimental part is the crack detection on bridge decks reachable only by a drone. Due to the poor accuracy of the ToF sensor the analysis is concentrated on approximately plane surfaces. In this way it is possible to conduct the realignment with homographies that mediate the effects of uncertainties related to the depth detection.

In the final part of the thesis an experiment is conducted on a panel of plane shape to the naked eye. On the panel a proper speckle is applied considering the distance at which the RGB-D sensor will work. A pre-crack is performed and then the panel is loaded with a hydraulic piston with an imposed cycle. The RGB-D sensor is handheld during acquisition to simulate the drone movement and the results are compared to a fixed stereo DIC system that acts as a reference ground truth. Results are presented in the last chapter of the thesis.

1 State of the art

The accuracy of DIC both for 2D and 3D reaches nowadays 1/100 of the dimension of the pixel [1]. The retrieval of each subset is performed with cross-correlation (CC) or sum of squared differences (SSD) algorithms, where the normalized and mediated ones are more robust in terms of noise rejection and change in lighting conditions [3], [4]. The search for the subset is performed with at least first order shape functions so that even deformation can be evaluated. The subpixel accuracy above mentioned is guaranteed thanks to registration algorithms such as the coarse fine search, the peak finding, the spatial gradient and genetic or FEM algorithms but the more robust approach is the Newton-Raphson (NR) one [3]– [5]. Errors can be introduced due to incorrect speckle, image distorsion, noise and correlation algorithm problems. The subset size must be large enough to encapsulate at least 3 speckle dots. The speckle pattern must respect these characteristics [3], [6]:

- High contrast;
- Randomness;
- Isotropy;
- Stability.

For 2D DIC uncertainties related to out of plane deformations or not perfectly planar object can be problematic while in 3D DIC stereo calibration issues can arise. As described in [3] other main problems that are still present for DIC are:

- Speckle pattern fabrication in harsh and difficult environments or for very small/large objects;
- adaptive selection of optimal subset size and shape function for subset matching;
- fully automatic, robust and fast estimation of initial guess;
- adaptive selection of optimal strain window size and shape function for strain estimation.

DIC can be a tool in diverse fields. Generally, it allows to measure the displacement or deformation of a structural element subjected to loading. [7] proposed a DIC

approach to study the fatigue behaviour of aluminum specimen to get out the sinusoidal imposed cycle, the crack growth and the stress intensity factor with satisfying results. DIC is also used at a microscopic level. [8]–[11] analyzed the capability of an image acquiring system coupled with a metallurgical microscope concluding that results are comparable with macroscopic DIC till magnification levels of 1000-2000 pixels/mm when out of plane displacements are kept within 3 micrometers and lens aberattions are limited to the minimum.

In the biomechanical field DIC applications can be found at different dimensional scales, on a wide range of biological specimens (soft and hard tissue) and for a variety of tests [12]. All this is possible with no loading effect of the measuring system due to the use of optical cameras. DIC is also applicable with regular metals, polymeric and composite materials [12]–[15]

To perform crack assessment on bridges different solutions are present. On site visual inspection is now giving place to unmanned evaluation due to an easier access to the desired place and the elimination of risks for the operators. UAVs with cameras allows the detection of cracks with different image processing techniques[16]–[21]. Other solutions are emerging like trained neural networks [22], [23] and deep learning algorithms [24] as well as methods based on principal component analysis (PCA) [25].

The use of UAV systems to perform civil structures assessment started years ago and it is now becoming more popular. In 2008 an autonomous UAV for visual inspection on bridges and pipelines was proposed as a starting point without providing numerical information [26]. A 3D UAV DIC system is proposed by [1] with 2 ccd cameras mounted with a 25° separation angle and a working distance of approximately 1.75m to reach a field of view of 1.4m x 1m. The movement of the drone does not induce a significant error to the measurement according to this study, except for the pitch rotational angle. When the pitch angle reaches +/-5° the error doubles. The experimental results indicate that the error when the measurement is done with the UAV compared to the fixed DIC system is increased, but remains of the same order of magnitude. The UAV DIC system is capable of detecting cracks not visible to the naked eye, till the dimension of $1x10^{-5}$.

In [27] a UAV DIC system to detect cracks on a beam is analyzed showing promising results as a preliminary assessment tool at inaccessible locations. [2] shows a UAV stereo-DIC system to study a pre-stressed concrete beam subjected to four point bending in laboratory conditions pointing out that for stereo dic on drone the payload of the UAV is a big disadvantage. [28] proposed an approach with DIC system mounted on drone to study the health of wind turbine blades showing high

correlation with strain data coming from strain-gauges. It is remarked that full-field deformation or strain with no interference between the blade and the measuring system is only possible in this way.

The approach proposed in this thesis consists in the use of an RGB-D system for DIC mounted on drone.

[29] studied the 3D reconstruction made with ToF sensor, coming to the conclusion that the systematic error is about 0.5-2 mm with a dispersion of raw data of 2-4.5mm when the target is at 2 meters of distance. The study points out that ToF sensors are compact and light, suitable for drone transportation. Moreover, compared to stereo system, ToF sensors are more robust under critical light conditions since they work with infrared frequencies. Between the sensors analyzed it is important to remark that the Basler Blaze 101 present a larger dispersion of data due to its longer wavelength (940nm). At the same time, it is more robust in case of sunlight disturbances. Basler Blaze is the ToF sensor chosen for the experimental part.

RGB-D DIC needs synchronous acquisition between the two cameras. Then, with data coming from stereo calibration depth and color or greyscale images must be realigned. After the realignment, 2D DIC is performed and thanks to the pin-hole model it is possible to get out the 3D displacement field [30]

In case of planar surface, the best fit of the plane performed on the area of interest reduces the variability and allows to work with homographies on the homogenous coordinates.

The UAV 2D DIC with the realignment of different frames with homographies is presented in [31]. It shows satisfying results comparing it with a fixed camera. The realignment of successive frames is made with homographies estimated on an approximately fixed part of the structure. The comparison is also done on FRFs from the UAV data and the fixed camera. The relative error of natural frequency is only 0.65%.

2 Proposed approach

In this chapter a complete overview of the proposed approach for RGB-D-DIC performed on UAVs is given.

The idea behind the algorithm starts from [30] and [31] works. In [30] the procedure to realign data coming from a RGB sensor and a depth sensor is described for RGB-D-DIC with fixed camera position. In [31] 2D DIC is performed on images coming from a drone with the compensation of movements calculated on a limited number of fixed points through homography.

The here proposed technique works with RGB-D sensor and it exploits the use of 2D-DIC method that coupled together with the depth information allows the compensation of the movement of the camera either by homography (both from images (2D) and depth maps (3D)) or by point cloud (approach used in a completely ideal situation, see Chapter 3, but not in the real application because of the poor accuracy of the depth sensor, see Chapter 4). This compensation, contrary to what is presented in [31], is calculated with the least square method on 3D centre subset points, coming from DIC and the depth sensor, of an approximately fixed part of the structure, thus providing a more robust estimation of the homography (or of the rototranslation in case of point cloud compensation). Moreover, working with point clouds 3D data coming from the RGB-D sensor allows to apply a final transformation to report centre subset points coordinates perpendicular to the optical axis of the camera (while in [31] it is assumed that the first frame is approximately perpendicular to the optical axis).

A detailed explanation of the calculation method and of the code implementation is presented in this chapter, starting from the acquired images and depth maps and finishing with the data representation.

More in details the RGB-D sensor for the acquisitions is composed by a Flir monochrome camera coupled with a Time of Flight (ToF) sensor (Blaze) for the depth maps (more information at Paragraph 4.2.1). This compact solution could be easily mounted on a drone thus allowing the acquisition of images even in difficult

and not accessible places to humans (the overall mass is approximately below 800 grams).

To perform the crack measurements with the moving sensor, it is firstly necessary to estimate correctly the rototranslation and the camera movement; in order to do that a reference part of the measurand object should be defined. This part should theoretically remain approximately fixed in the space and it shouldn't be subjected to any change in shape, dimensions, position (actually, as it'll possible to notice in Paragraph 4.3.2 this assumption could be relaxed in a real application, obtaining anyway good results). This is necessary to correctly compensate the effect of the movement of the drone and to solve the multiple-views scene problem (using only 2D-DIC and the ToF sensor). Moreover, it's important to define one reference frame among the multiple acquisition, typically the first one. In fact, multiple frames of the same object will be acquired to evaluate how the crack behaves in time.

The first step of the procedure is the 2D DIC. For the fixed part, knowing the position of the centre subsets for each frame, coupled with the depth information, allows to compensate the camera movements. For the moving part, i.e. the part on which the crack assessment is performed, the same correction related to the camera movements is applied and what remains is the actual displacement describing the crack opening.

2.1. Digital Image Correlation

The very first step after the acquisition carried out with RGB-D sensor, consists in performing the DIC on the set of greyscale images acquired without compensating for radial and tangential distorsion (the compensation is carried out afterwards in MATLAB to avoid uncertainties related to resampling of the image).

With the assumption that the left half of the plane is fixed and free from deformations, two ROIs are selected as shown in Figure 2.1 which is an example of a reference image used in the analysis.

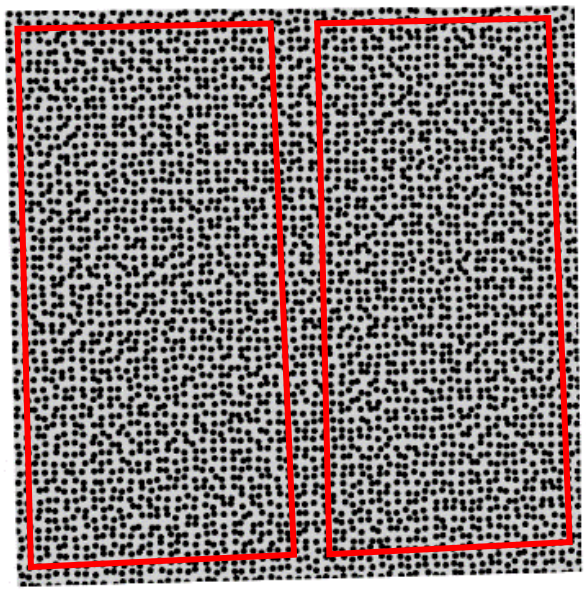


Figure 2.1: example of reference image with two ROIs put in evidence (left one for the fixed part and right one for the moving)

The choice of the subset and the step size is done according to [32] so that each subset contains at least three speckles.

DIC is performed on a set of cracked images, as in Figure 2.2, with the assumption that the movement is associated only to the right half ROI. Working with sensors mounted on a drone, even the left part will show displacements in the DIC analysis. These data must be used to compensate the undesired effects of the drone movements.

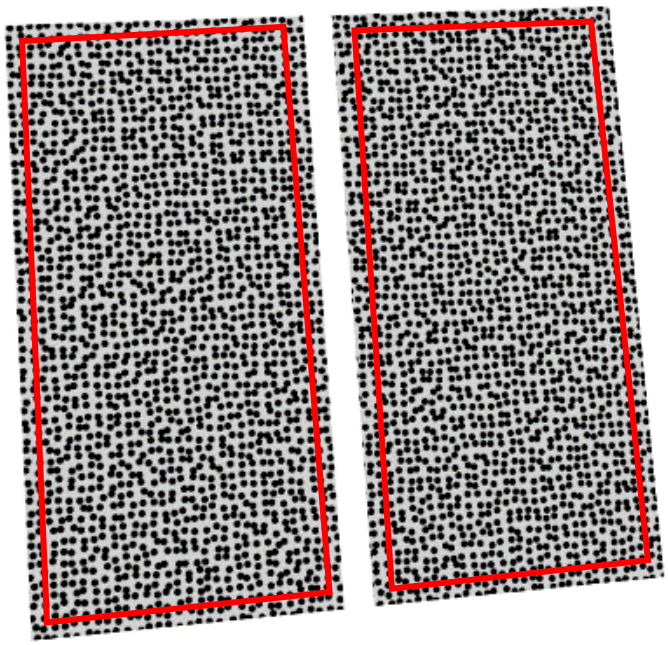


Figure 2.2: example of cracked image with the two ROIs put in evidence

2.2. Acquired data manipulation

Once the DIC is completed (as briefly explained in the previous paragraph), all the data are arranged in different folders for pure convenience, in particular one with the photos obtained from the FLIR sensor (without radial and tangential distorsion compensation), one with alla the data coming from the DIC and a last one regarding the depth maps files (actually a conversion from depth map files to plane models is carried out to improve the quality of the final results, but this will be explained better later in Paragraphs 3.1.4 and 4.3.1, which show some differences between them).

It's important also to load the stereo calibration file with the data coming from Caltech stereo camera calibration called "Calib_Results_stereo.mat" (the calibration procedures will be explained better later in Paragraphs 3.1.3 and 4.2.2).

Depth and DIC data are loaded on MATLAB and stored. The indexes of the coordinates are increased by 1 pixel since the first pixel in the VIC software is (0,0) while in MATLAB convention indexes start from 1. Then, the compensation of only radial distorsion is carried out.

2D DIC data with the corresponding centre subset coordinates (x_{pix} and y_{pix}) coupled with the depth plane model (imposed in the simulator and retrieved by point cloud data coming from ToF sensor in real application, see Paragraph 4.3 for more information on real application) allow to obtain the depth information associated to each centre subset. For the sake of clarity, here below the line of code used to get the depth information are reported (better explanation at Eq.2.1):

$$\begin{aligned} g_x &= (x_{pix} - c_x)/f_x \\ g_y &= (y_{pix} - c_y)/f_y \\ z_{3D} &= -\frac{d_{ref}}{\left[g_x \; g_y \; 1\right] * n_{ref}} \end{aligned} \qquad 2.1$$

Where:

- x_{pix} and y_{pix} are the coordinates of each centre subset coming from DIC software (respectively for x-coordinates and y-coordinates);
- f_x, f_y, c_x, c_y are the intrinsics parameters of the camera (as shown also in Eq. 2.4);
- d_{ref} is the coefficient of the estimated plane d;
- n_{ref} is the vector normal to the plane in the form $[a \; b \; c]'$;
- z_{3D} is the z coordinate of each centre subset.

Then with the pin-hole camera model (see Eq.2.2) x and y coordinates in the real world (x_{3D} , y_{3D}) are retrieved. Finally the point cloud associated to the centre subset is built (x_{3D} , y_{3D} , z_{3D}).

$$\begin{cases} x_{3D} = \left(x_{pix} - c_x\right) * \dfrac{z_{3D}}{f_x} \\ y_{3D} = \left(y_{pix} - c_y\right) * \dfrac{z_{3D}}{f_y} \end{cases} \qquad 2.2$$

2.3. Compensation of camera movement

The purpose of this paragraph is to give a detailed explanation of how the rototranslation of the camera is compensated.

Different solutions are analysed:

- Homography from 3D point clouds: homography estimation with calculation based on the realignement of 3D point clouds of the centre subsets (this approach is recalled later in the paper as 'H 3D');
- Homography from 2D greyscale images: homography estimation based on the realignment of 2D set of points of the centre subsets (this approach is recalled later in the paper as 'H 2D');
- Rototranslation from point cloud: rototranslation estimation from point clouds of the centre subsets (this approach is recalled later in the paper as 'PC 3D');

2.3.1. 'H 3D': homography from 3D point clouds

The situation taken into account is summarised in Figure 2.3, where π is the framed plane, m and m' respectively as the reference sensor plane and moving sensor plane.

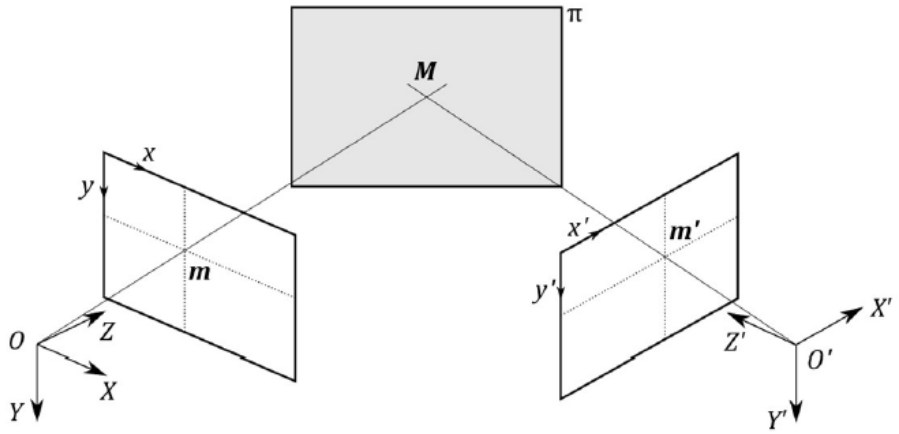


Figure 2.3: Summary of the considered situation: multiple views scene problem

Since the centre subset point cloud of the reference fixed part is known for each frame, the rototranslation between the first one and the others is calculated with the MATLAB function 'estimateGeometricTransform3D' with 'rigid' option as input, since we are dealing with a rigid body.

This function is based on the M-estimator sample consensus (MSAC) algorithm, which is an improved variant of the RANdom SAmple Consensus (RANSAC) algorithm. RANSAC is an iterative method to estimate parameters that deals with data containing outliers in a proper way. It is not deterministic cause it is based on the repeated random sub-sampling.[33]–[36]

Moreover, making the assumption that the acquired object is approximately plane, it is possible to demonstrate that the induced homography between m and m' is calculated as shown in Eq.2.3:

$$H = K * \left(R - t * \frac{n}{d}\right) * inv(K) \qquad 2.3$$

where the parameters are:

- K is the intrinsic parameters matrix K in the form presented in Eq. 2.4:

$$K = \begin{bmatrix} f_x & 0 & 0 \\ 0 & f_y & 0 \\ c_x & c_y & 1 \end{bmatrix} \qquad 2.4$$

- R and t are the rotation matrix and the translation vector which bring the reference frame XYZ coincident with the successive frame X'Y'Z'

- n is the normal of the plane evaluated from the reference frame XYZ
- d is the known term of the plane ax + by + cz + d = 0 evaluated from the reference frame XYZ, that's why the minus sign is used (while plus sign should be used if the actual distance between camera and origin would have been considered).

  The way in which the homography is evaluated allows to estimate parameters n and d once and for all, however in the application of H the inverse matrix should be used according to Eq.2.5.

$$m = inv(H) * m' \qquad 2.5$$

The reason of the inverse is related to the fact that the successive frame should be brought on the reference one. This equation is applied obviously on both the fixed part centre subsets data and the moving ones.

The main problem of this approach is the correct estimation of the parameters used to define H, problem that could be avoided introducing the homography based on greyscale images

2.3.2. 'H 2D': homography from 2D greyscale images

As already mentioned, alternatively, the homography H can be estimated with a function based on MSAC and RANSAC algorithm [35], [36].

In MATLAB there's the already built in function 'estimateGeometricTransform2D' with 'projective' as option.

In this case it works directly with the homogenous coordinates of the centre subsets and with a minimization procedure calculates the induced best-fitting homography for the realignment of the centre subsets of the fixed part. This allows to minimize eventual issues related to an imperfection in the depth estimation.

This doesn't mean that this approach is free from other issues and problems. In particular since the algorithm used to estimate H is based on a projective trasformation, some deformations effects can be confounded as projective effects by the minimization algorithm, thus causing their undesired compensation.

It is important to underline that even though the technique proposed in this paragraph and the one already seen in Paragraph 2.3.1 are based on the same algorithm (MCAS), the results in terms of homography will be different since in the former case the estimation is projective in the 2D world without the depth information, while, for the latter, the homography is calculated on the rigid rototranslation of the 3D point cloud of the centre subsets based on the depth image.

2.3.3. 'PC 3D': Rototranslation from point cloud

The last approach doesn't require the use of a homography, but it's based entirely on the point clouds of the centre subsets.

Once calculated the rototranslation on the basis of the fixed part of the object (as seen also in Paragraph 2.3.1), the realignement is done directly on the 3D point cloud by applying the calculated rototranslation to the point cloud. Since there's no passage to homogenous coordinates, 3D coordinates are used taking into account also the depth.

The goodness of the algorithm can be evaluated not only on the alignment of x and y, but also on z. This approach should give the same results as the 'H 3D' one in a completely ideal case, while could present poor results with respect to the 'H 3D' and 'H 2D' when dealing with a sensor with poor resolution in terms of depth (in this thesis, as already mentioned, it'll be exploited only for the ideal case and not for the real one); anyway it could be used also with surfaces and objects that aren't planar and it provides displcament information also on the third dimension.

2.4. Realignement perpendicular to the optical axis of the camera

Here the explanation about the last step before the data representation discussion is presented. This step consists in the calculation of matrix R0, the rotation matrix responsible for bringing the current frame perpendicular to the sensor's one.

To respect this constraint the z-axis of the panel must be parallel with the z-axis of the sensor, while x and y-axes could be chosen arbitrarly.

To obtain this final result the use of the principal component analysis is recommended. This analysis allows to define the principal direction of the considered point cloud (in this case a plane), thus making it possible to define the three principal axis.

The calculation is made automatically using eigenvalues and eigenvectors of the following matrix (the same analysis could be performed with the already implemented MATLAB function "pca"):

$$\begin{bmatrix} x_1 - \mu_x & \dots & x_n - \mu_x \\ y_1 - \mu_y & \dots & y_n - \mu_y \\ z_1 - \mu_z & \dots & z_n - \mu_z \end{bmatrix} * \begin{bmatrix} x_1 - \mu_x & y_1 - \mu_y & z_1 - \mu_z \\ \dots & \dots & \dots \\ x_n - \mu_x & y_n - \mu_y & z_n - \mu_z \end{bmatrix}$$

The critical aspect regards the coupling between axes (x,y,z) and their corresponding eigenvector.

Knowing that to each eigenvalue an eigenvector is associated, the shape of the selected region of interest (ROI) helps making the coupling possible.

In particular, the code automatically detects if the ROI is a vertical or a horizontal rectangle (by quickly analysing the size of the DIC output matrix). In the first case the eigenvectors are rearranged in order to obtain:

z: smallest eigenvalue (so the corresponding eigenvector will be placed in the last column of R0)

y: highest eigenvalue (so the corresponding eigenvector will be placed in the middle column of R0)

x: remaining eigenvalue (so the corresponding eigenvector will be placed in the first column of R0)

In the second case:

z: smallest eigenvalue (so the corresponding eigenvector will be placed in the last column of R0)

x: highest eigenvalue (so the corresponding eigenvector will be placed in the first column of R0)

y: remaining eigenvalue (so the corresponding eigenvector will be placed in the second column of R0)

Moreover, to respect the convenction (right-hand rule), all the terms in the diagonal must be positive and the determinant of R0 must be equal to one.

Then matrix R0 is used according to Eq.2.6:

$$H0 = K * R0 * inv(K)$$
$$m_{camera} = H0 * m \qquad 2.6$$

Anyway, to improve the quality of the results in terms of readability, x and y-axes shouldn't be defined with pca, but it's better to define them manually depending on the crack present.

In fact, in real situation the crack isn't necessarily generated along one of the axis (x or y) and it isn't typically a straight line.

It is decided to leave the choice of the other two axes (x and y) to the user.

As it can be seen from Figure 2.4 the reference image (not correctly aligned with the camera axis) appears to the user in MATLAB and he's asked to select two points which belong to the desired x-axis of the framed plane (as an example the two red ones).

Knowing the position of these two points the code can retrieve the x-axis and with a rotation of 90° also the y-axis.

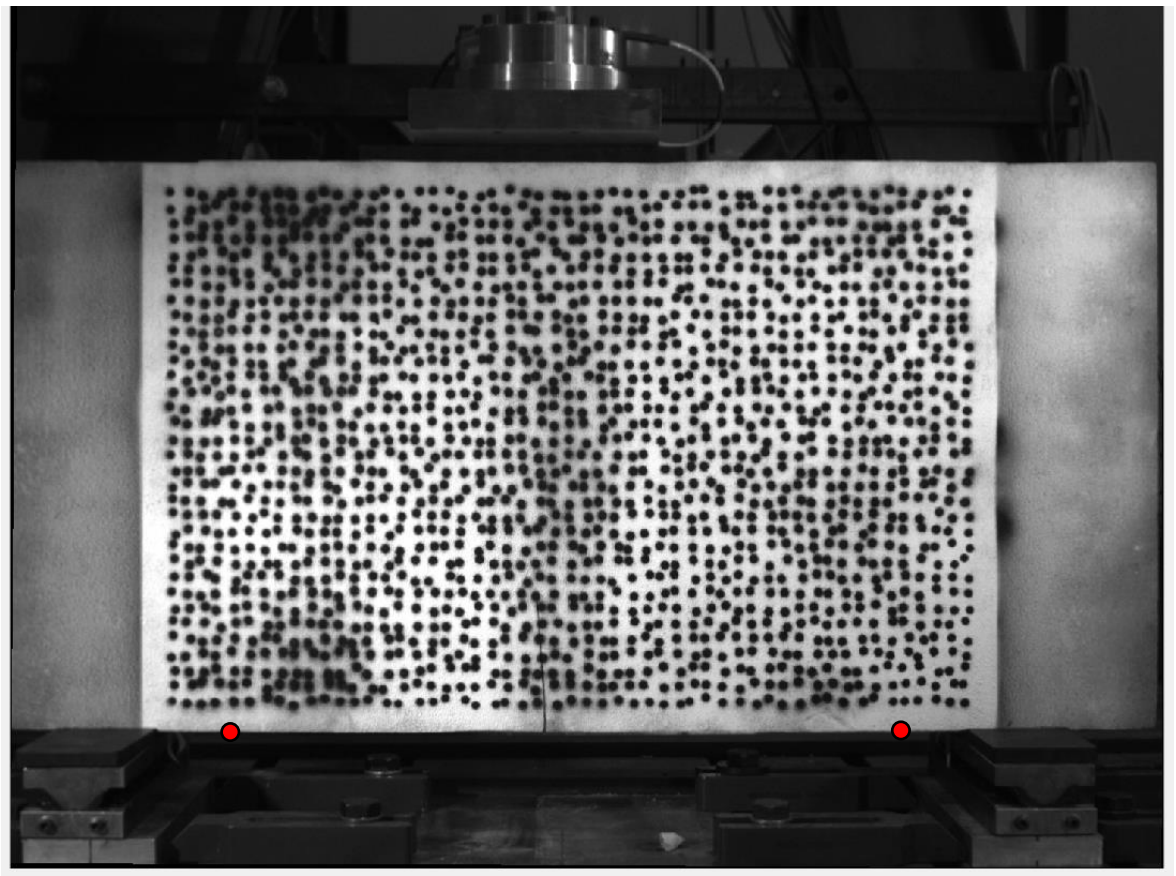


Figure 2.4: pictures of the cracked specimen that appears in MATLAB so that the user can choose x and y axis as he prefers

This is of paramount importance for the data analysis part, since a clear knowledge of which are the axes allows to read the final graphs without difficulties and misunderstandings.

To conclude, after the application of camera movements correction either by homography or point cloud rototranslation and the realignment perpendicular to the optical axis of the camera, the actual displacement of the moving part is correctly retrieved, while ideally the reference part should show null displacement.

2.5. Data representation

In this paragraph a brief discussion about the way in which outputs are generated is presented.

After the realignement between the reference frame and the successive one and the global realignement perpendicular to the camera are performed, the data are stored in a matrix $n_s x 3$ where n_s represents the number of subsets in which the acquired image is divided.

Being these coordinates homogeneous (for the 'H 3D' and 'H 2D' cases), all the three columns are normalized with respect to the third one in order to obtain it full of ones.

Then the difference between the corresponding points is performed.

Once the difference is computed the coordinates are converted from pixel to millimeters multiplying x and y coordinates respectively for $conv_x$ and $conv_y$ (according to the conversion parameters shown in Eq.2.7).

$$conv_x = -\frac{d}{f_x}$$
$$conv_y = -\frac{d}{f_y} \qquad 2.7$$

This procedure is carried out both for the fixed part of the object (reference) and for the moving one.

The results are then plotted as images where to each centre subset a color with different intensity is associated (ranging from the minimum value obtained to the maximum one).

An example of the results is shown in Figure 2.5.

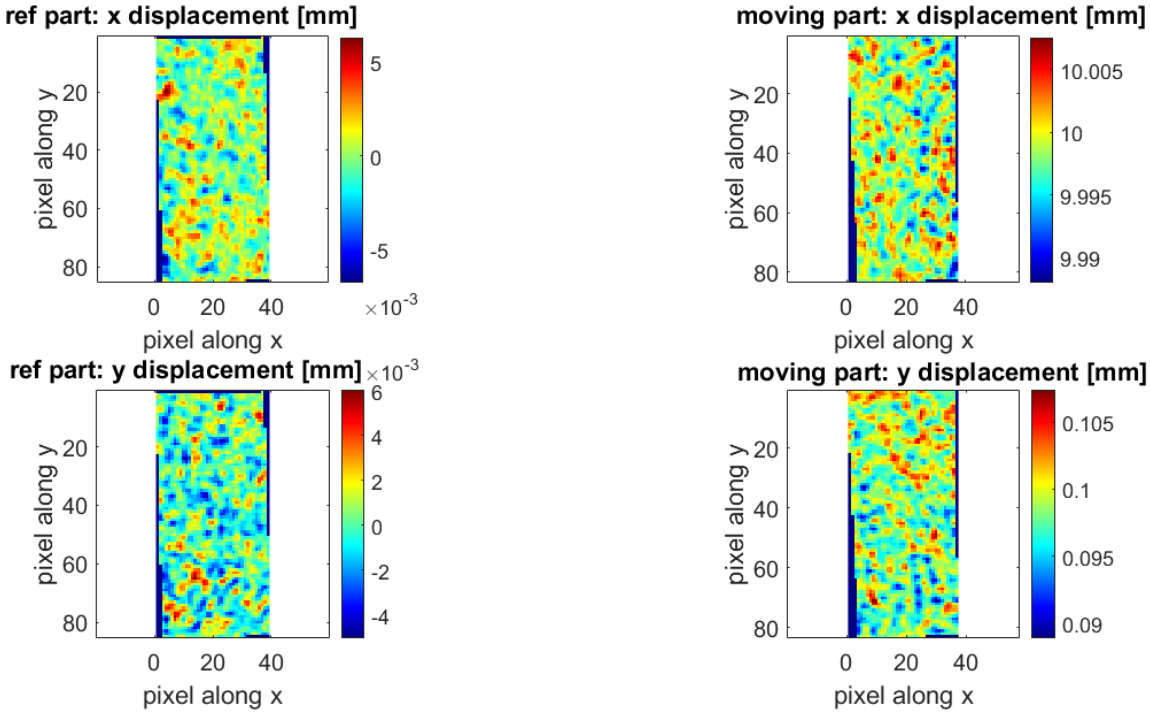


Figure 2.5: final output of the algorithm. Starting from the upper left corner the displacement of the fixed part along the x axis is shown, while on the right one there's the displacement on the moving one. The same is presented also for the bottom part, but all is referred to y.

Moreover, it has to be noted that with the approach previously discussed in Paragraph 2.3.3 ('PC 3D') it's possible to retrieve information related to the z displacement too. This is limited to the simulator (ideal case) since the quality of the sensor doesn't allow the proper use of this technique in a real world application. The results are shown in Figure 2.6.

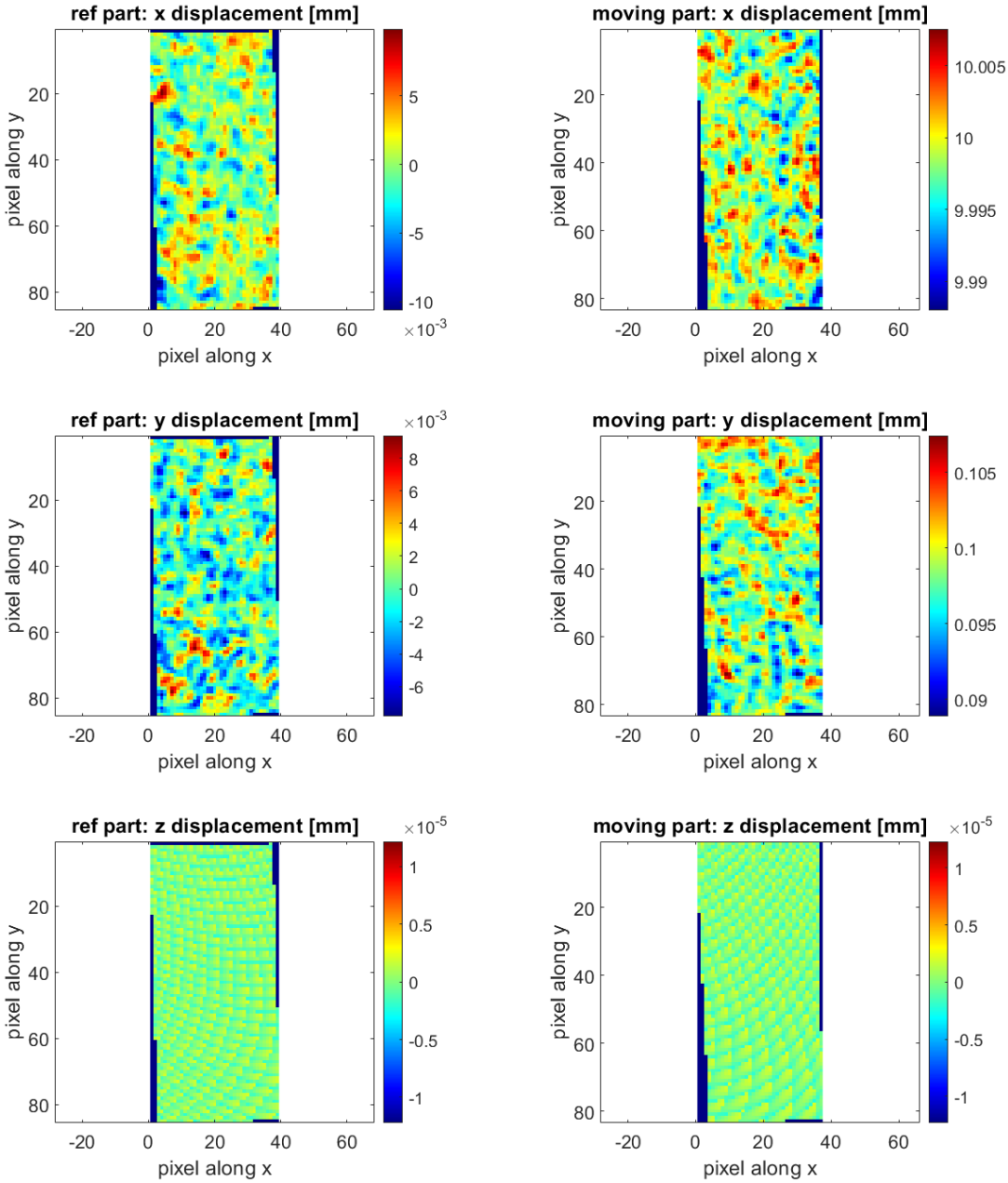


Figure 2.6: the plots are equivalent to the ones previously shown in Figure 2.5 with the addition of the last two plots that are related to the z displacement for the reference part (left) and for the moving part (right)

It's possible to see that the plots related to x and y displacements are identical to the one previously shown in Figure 2.5 (since they're coming from an ideal case). This wouldn't be true in the case of a real world application.

Due to the fact that the crack won't probably be perfectly aligned to the x-axis or to the y one, the square root of the sum of the two components squared is carried out.

This allows the proper estimation of the maximum crack displacement.

Afterwards the mean and the standard deviation of these data are evaluated by means of the already implemented MATLAB functions "mean" and "std".

The so evaluated standard deviation is related to the mean value of the data, but it's not useful to correctly estimate the dispersion of the data around the crack value imposed in the simulator. In order to do that, the Root Mean Square error is introduced as shown in Eq.2.8, where crack is the crack value imposed and x_i are the displacement of the centre subsets after the quadrature.

$$RMS_{error} = \sqrt{\frac{\sum_i (x_i - crack)^2}{n-1}} \qquad 2.8$$

3 Numerical simulator development and validation

In this chapter the analysis of virtual images in a simulation environment with ideal and known crack size and direction is presented. The position and orientation of the cracked plane in the space is controlled and defined with reference to the virtual camera. Controlling most of all the variables that will be faced in a real experiment allows to verify the algorithm and to define the ideal capabilities of the approach described in this thesis.

In an on-field experiment scenario many issues may arise such as possible random noise associated with the camera, problems related to the compensation of the images, unsatisfying resolution of the depth sensor, incorrect color/greyscale image-depth alignment, wrong detection of the plane in the space or of the 3D model.

These eventualities could hide issues strictly related to the algorithm and not to instrumentation or external factors ones.

Therefore, to validate the algorithm and to eliminate all the possible factors and causes of errors or imperfections, an analytical simulator is considered. The simulator is capable of generating ideal greyscale images with defined speckle patterns and crack (similar to those that will be later acquired directly on field) and corresponding depth images obtained by meaning of the analytical plane equation.

3.1. Simulator structure and implementation

The following paragraphs show the overall structure of the simulator, analysing step by step the defined parameters, the generation of the greyscale images and of the depth maps and the related procedures.

3.1.1. Parameters definition

The definition of the relative camera-plane position, orientation and the dimension of the plane and of the crack inside the simulator are here reported.

It is important to highlight the fact that in the various acquisitions it is the camera that changes position and, thus rototranslates, while the plane remains perfectly still, except for the imposed crack opening.

The camera starts to acquire in the origin of the reference frame, with its own axis Xg Yg and Zg coincident with the axis of the reference frame X Y Z and facing YZ-plane as evidenced in Figure 3.1.

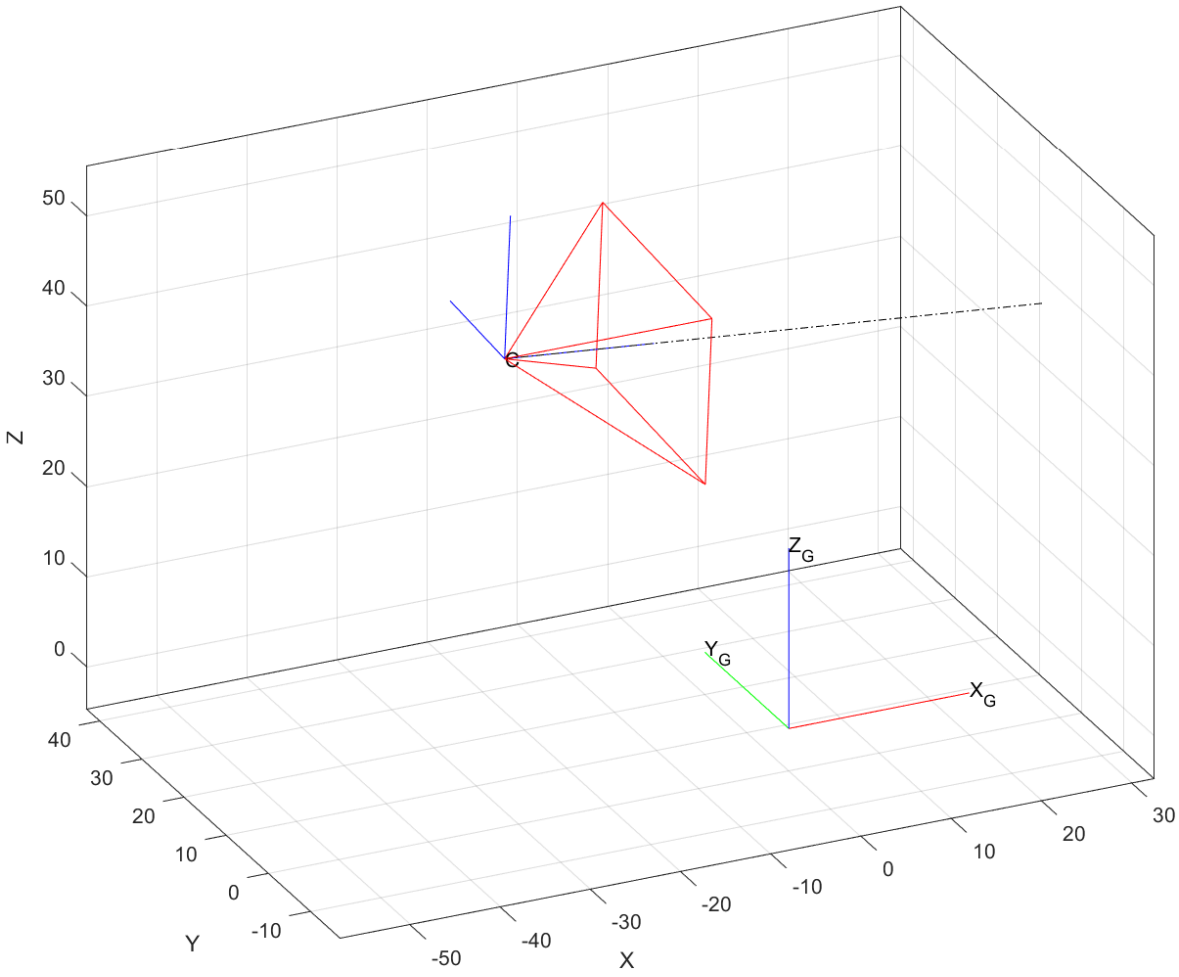


Figure 3.1: starting position of the camera in the reference frame (the red lines were placed only to highlight the portion of the space which will be acquired).

The plane is splitted in two smaller rectangles (1000x500mm) as shown in Figure 3.2 to allow the simulation of the crack opening. The center of gravity of each rectangle is placed in the following points:

$$C_1 = [1000, 250, 0]$$
$$C_2 = [1000, -250 - crack, 0]$$

Where the coordinates are expressed in millimeters in the usual order [x, y, z] and with crack equal to the crack opening desired in millimeters.

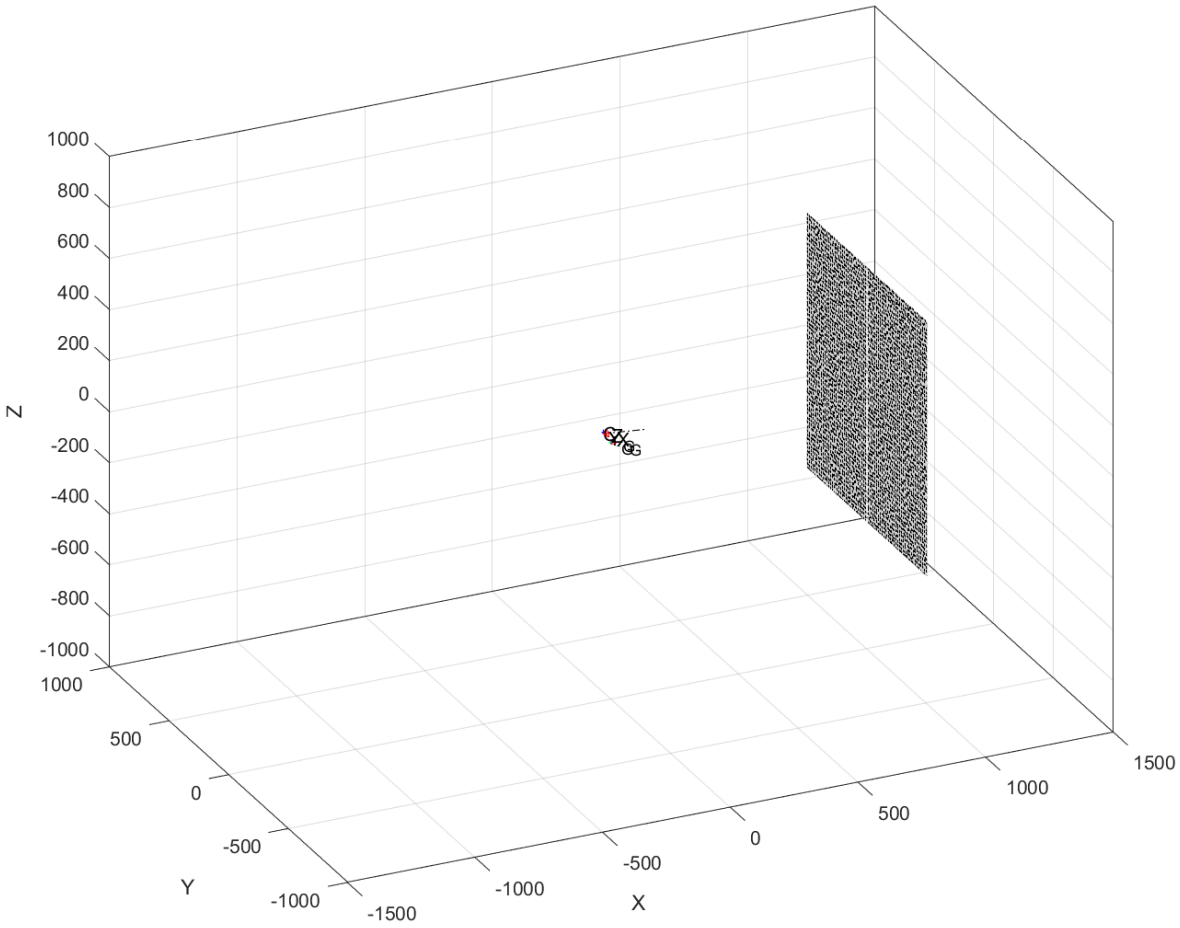


Figure 3.2: in this figure the whole scene composed by camera and plane with crack is shown. Here the plane already presents speckle patterns, but a better description will be given in the next section (Paragraph 3.1.2)

To describe the rototranslation of the camera the following six parameters are used:

$$orientation = [yaw\ pitch\ roll] * \frac{\pi}{180°}$$
$$position = [t_X\ t_y\ t_z]$$

Where the angles yaw, pitch, roll follow the left-hand rule, are expressed in degrees and converted in radians by means of the multiplication with the factor $\pi/180°$. The position vector is expressed in millimeters.

While the final camera position in the space will be $[t_X\ t_y\ t_z]$, the exact orientation is stricly influenced by the order of rotation (in this case yaw, pitch, roll) due to the fact that the set of angles is converted in a rotation matrix.

3.1.2. Speckle and greyscale images generation

To obtain the ideal greyscale images it is necessary to generate the proper speckle and superimpose it to the cracked plane of the simulator.

To obtain a proper speckle pattern reccommendation from [3], [6] are taken into account:

- randomness: non-periodic and non-repetitive pattern to facilitate full field displacement mapping;
- isotropy: no directionality in the pattern;
- high contrast: varying greyscale intensities and relatively large intensity gradients;
- dimension: speckle granules with a size of 3–5 pixels or slightly greater.

The generation is done with the software proposed by [37] with the final output reported in Figure 3.3, thus allowing to satisfy randomness and isotropy constraints. Among them, the pattern variability is set to 49% and the in-centre spacing is set to 65 pixel (with a speckle diameter equal to 50 pixel).

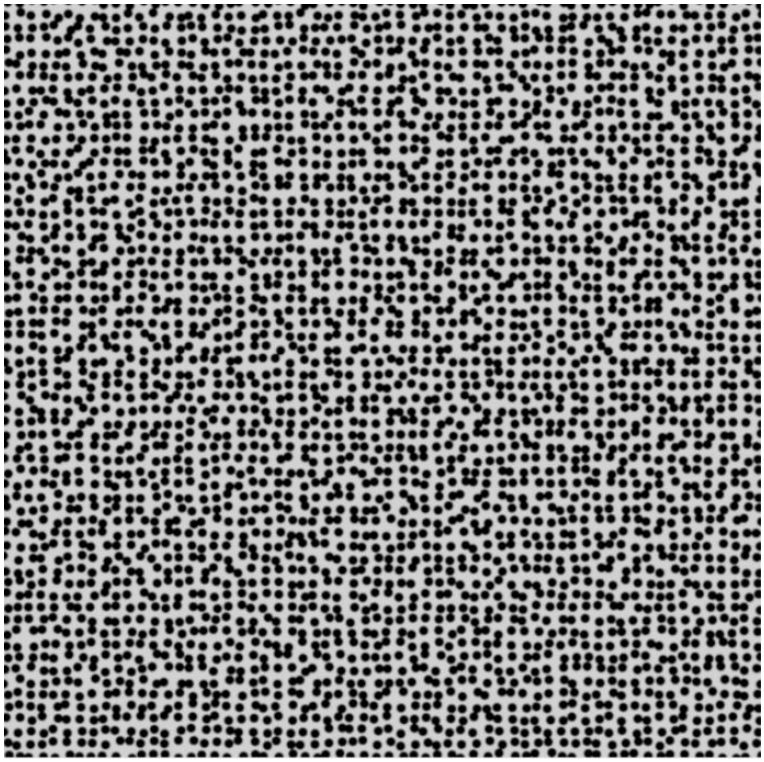


Figure 3.3: speckle pattern

The camera's parameters are the following ones:

$$Sensor\ size = [1.6\ \ 0.9]\ mm$$

$$Pixel\ size = \frac{1.25\ mm}{10\ px}$$

$$Focal\ length = 0.8\ mm$$

The pixel size is divided by 10 because of a binning procedure that will be carried out afterward. This procedure is fundamental in order to guarantee the high contrast.

It is appropriate to perform DIC on images that are greyscale and not on binary images.

The final speckle is presented in Figure 3.4. The size of the speckle could be reduced but the choice is conservative to guarantee no problems with the DIC search algorithm (wrong or imperfect centre subset detection).

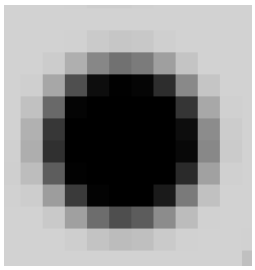


Figure 3.4: The final speckle respects all the constraints.

The final output in terms of image is instead shown in Figure 3.5.

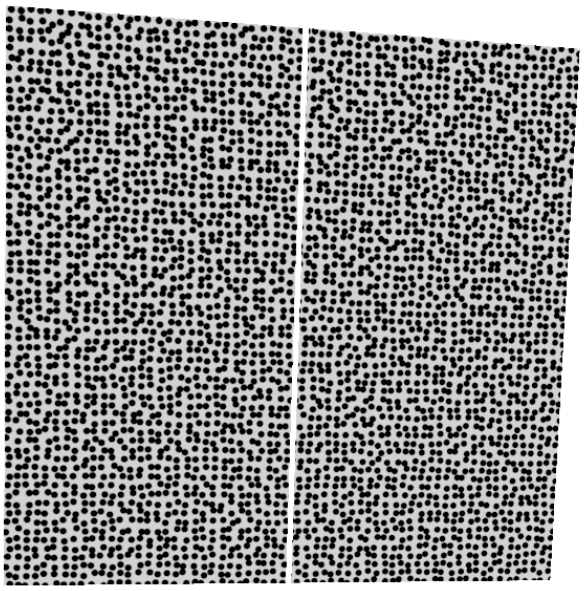


Figure 3.5: greyscale image with proper speckle applied

3.1.3. Simulator camera calibration procedure

The calibration procedure is necessary to know precisely the intrinsic parameters of the virtual camera, specifically the focal length and the optical centre position.

The calibration procedure is carried out using an algorithm that can automatically generate a series of images portraying a checkerboard taken from different angles in order to cover the entire field of view. The checkerboard presents 15 x 20 squares (rows times columns) with edges of 50mm for each square. A binning procedure is carried out too as in the greyscale image generation. 28 images for calibration are considered.

Below in Figure 3.6 it is represented a calibration image as example.

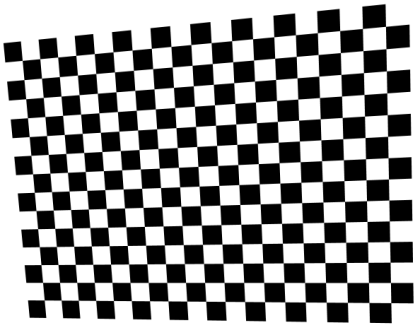


Figure 3.6: example of the acquired chessboard

The calibration procedure is performed with the 'Matlab Single Camera Calibrator' app where the detected points are accurately identified (see Figure 3.7).

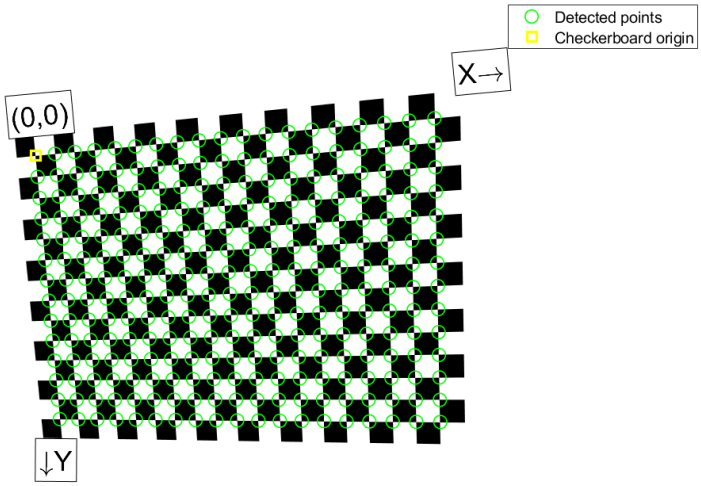


Figure 3.7: example of the acquired chessboard with the detected points in green.

The final output of the calibration procedure is the intrinsic parameters matrix K reported in Eq.3.1 using the same form previously seen in Eq. 2.4

$$\begin{matrix} f_x = 637.24 & 0 & 0 \\ 0 & f_y = 640.08 & 0 \\ c_x = 640.50 & c_y = 360.48 & 1 \end{matrix} \qquad 3.1$$

Note that all the values are written with pixels as measurement units and, for compactness, they're reported here with only two decimal digits, while they are used with all the decimal numbers in the code.

3.1.4. Depth maps generation

In this paragraph the discussion on how to retrieve depth information is presented; particular attention is given on the relationship between the analytical plane (used in the end with the real world application and the RGB-D sensor) and the depth image (used only in the simulator and useful in the case of sensors that give as output only the depth image and not the point cloud).

Before taking into account the depth image generation, it is appropriate to briefly introduce the analytical theory behind the code.

The idea behind the following steps is to create an image (depth image) starting from the knowledge of the 3D plane analytical equation. In each pixel there will be the x coordinate (which in our reference system for the depth maps generation is the depth coordinate) of the point portrayed.

Since the depth images generation is strictly coupled with the color images generation, the same consideration presented up to now are valid; in particular, it is appropriate to consider as starting point the camera in the origin of the reference frame and the plane perpendicular to the x-axis at a distance of 1000mm from the center.

By considering the straight line which passes through the reference frame origin and through a point $P = (p_x, p_y, p_z)$ where P is the pixels' center, it is possible to write in parametric form (Eq.3.2):

$$\begin{cases} x = p_x * t \\ y = p_y * t \\ z = p_z * t \end{cases} \tag{3.2}$$

That become in cartesian form (Eq. 3.3):

$$\begin{cases} y - \frac{p_y}{p_x} * x = 0 \\ z - \frac{p_z}{p_x} * x = 0 \end{cases} \tag{3.3}$$

The intersection with the analytic plane model is reported in Eq.3.4:

$$\begin{cases} y - \frac{p_y}{p_x} * x = 0 \\ z - \frac{p_z}{p_x} * x = 0 \\ a * x + b * y + c * z = -d \end{cases} \tag{3.4}$$

Multiplying for -b the first one and for -c the second one Eq. 3.5 is obtained:

$$\begin{cases} -b * y + b * \frac{p_y}{p_x} * x = 0 \\ -c * z + c * \frac{p_z}{p_x} * x = 0 \\ a * x + b * y + c * z = -d \end{cases} \tag{3.5}$$

By summing up all the equations we take out x, which is the coordinate of interest (Eq.3.6):

$$x = \frac{-d}{a + b * \frac{p_y}{p_x} + c * \frac{p_z}{p_x}} \qquad 3.6$$

It has to be noted that in the end Eq.3.6 is the same of Eq.2.1, which is enough for the correct implementation of the algorithm for the real world application.

While in the simulator the depth image is also retrieved by imposing as x_{pix}, y_{pix}, c_x and c_y the correct data of a pixel grid 1280x720 with the intrinsics matrix shown in Paragraph 3.1.3

The final results in terms of output is reported below in Figure 3.8 where the greyscale image has been already presented (Figure 3.5) and it can be viewed that the depth map is correspondent to the greyscale one.

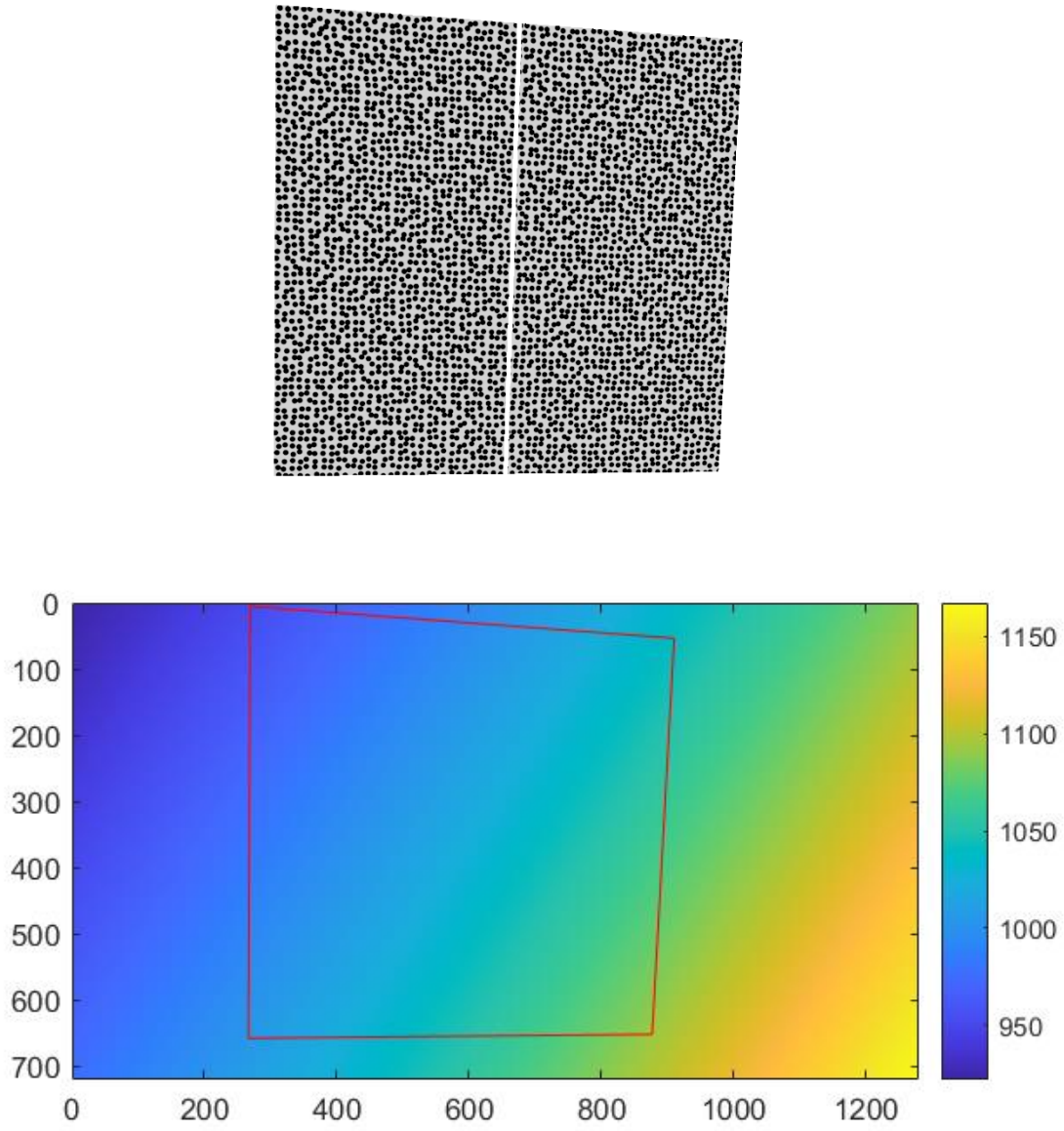


Figure 3.8: on the upper side the color image, on the bottom one the depth image

3.2. Design of numerical tests

In this paragraph a complete overview of the design of the numerical tests is presented.

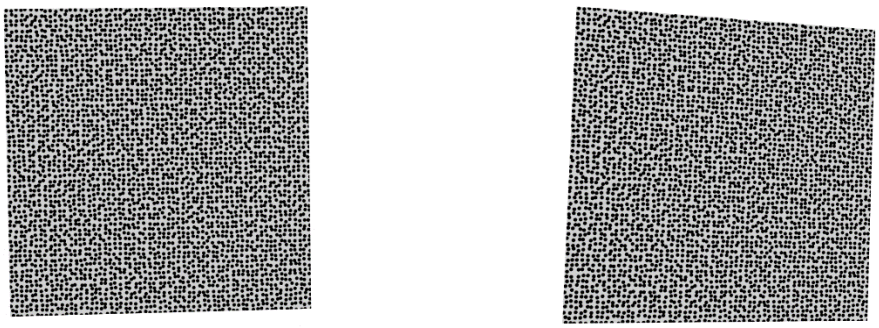
In particular two tests are performed where the images are taken respectively:

- with different camera angles
- with different crack dimension

Moreover, a set of different reference images is used to be more confident with the final results.

3.2.1. Reference images

To be more confident with the final results 7 reference images (no crack opening) are considered as shown in Figure 3.9 with parameters reported in the Table 3.1. It is possible to notice that each image is acquired from a different position in order to be sure that the procedure regarding the realignement perpendicular to the optical axis is free from errors.



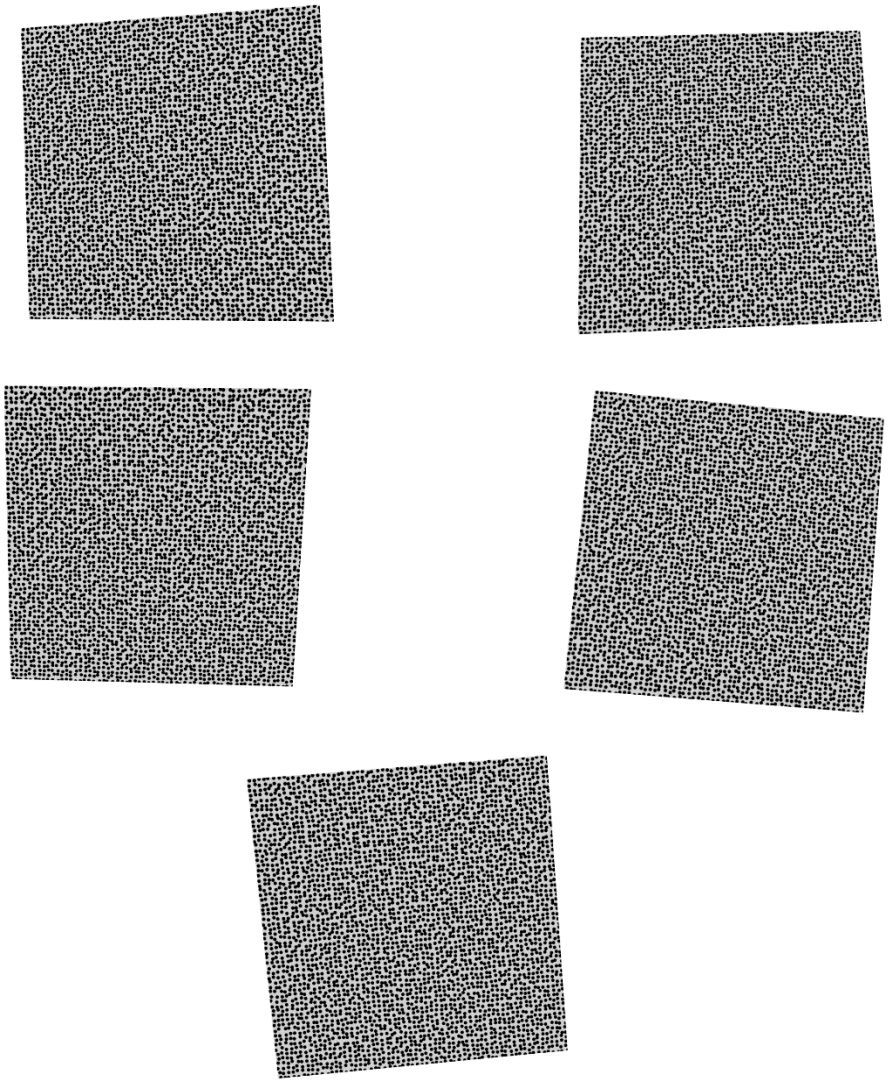


Figure 3.9: reference images used during the tests. Going from the upper left corner to the bottom right one the image numbers are 0000, 0001, 0002, 0003, 0004, 0005 with the last one at the bottom which is 0006.

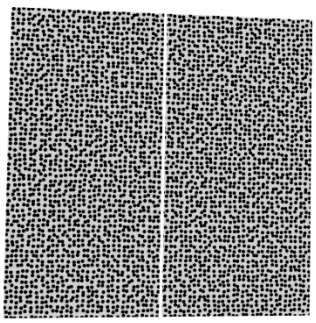
Image number	Rotation angles (degrees)	Vector of position (mm)
0000	[1 -0.5 -1]	[3 1 -2];
0001	[5 1 2]	[-5 10 2]
0002	[-5 1 -2]	[-5 -12 -4]

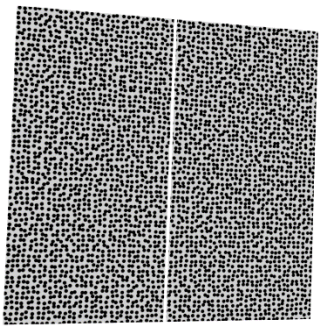
0003	[1 5 -2]	[-35 -5 -55]
0004	[-1 -5 1]	[-22 +5 +78]
0005	[+1 2 5]	[-20 4 -15]
0006	[+1 -2 -5]	[-20 4 +12]

Table 3.1: reference images (without crack) data

3.2.2. Fixed crack dimension and moving camera images

A set of 30 images is generated with the following characteristics: crack of 10mm with rotation angles that vary one at a time from -15° to +15° with a fixed step of 3° (avoiding the 0°), keeping the other two random in a range between -2° and +2° (visual explanation in Figure 3.10). The vector of position $[t_x\ t_y\ t_z]$ is changed to fit all the plane in the field of view of the camera. To better understand the strategy used, the first 10 images (varying the yaw angle) are reported in Table 3.2.; the overall set of images is created in order to be confident that the approach doesn't show any problems for all the possible camera angles taken into account (in particular higher camera angles could be critical for the proper use of the algorighm).





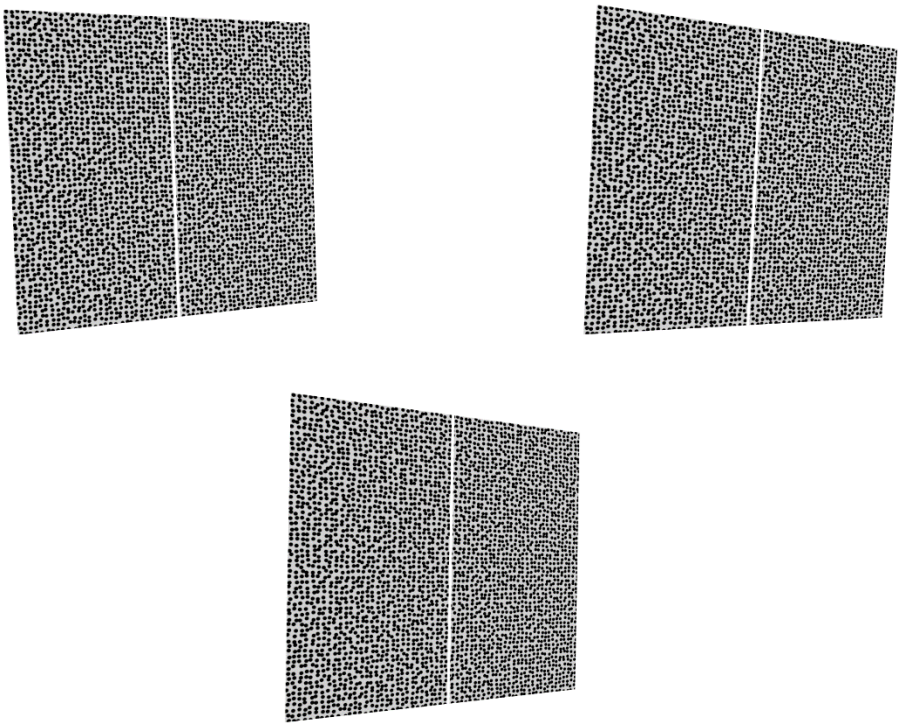


Figure 3.10: images with gradually increasing yaw angle. Going from the upper left corner to the bottom right one the image numbers are 0007 (with 3° yaw angle), 0008, 0009, 00010 with the last one at the bottom which is 0011 (with 15° yaw angle).

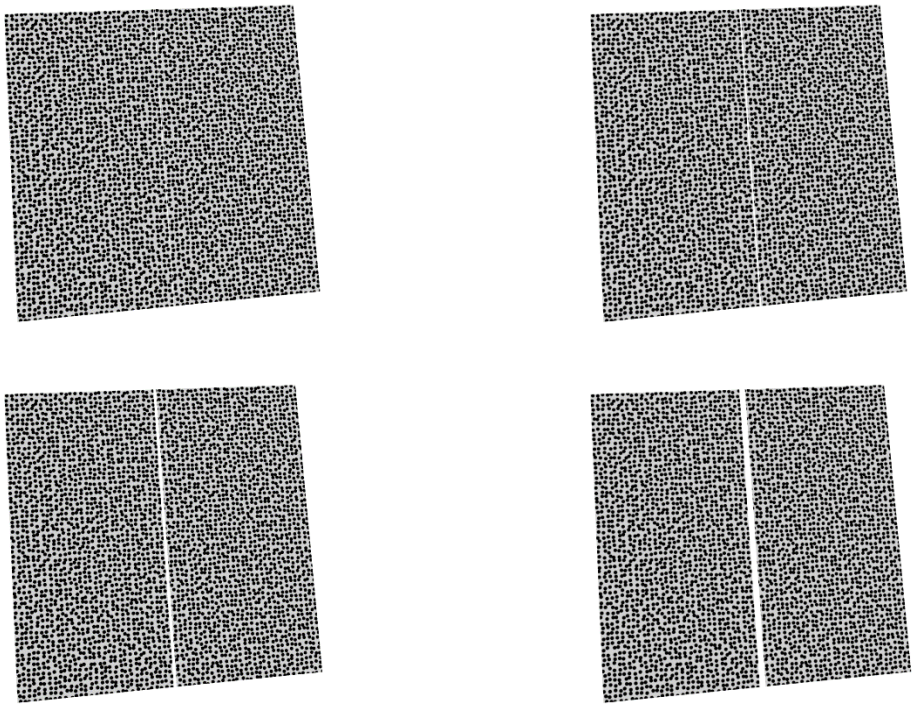
Image number	Rotation angles (degrees)	Vector of position (mm)
0007	[3 1 1]	[-4 10 6]
0008	[6 1 2]	[-8 20 6]
0009	[9 -1 -2]	[-15 65 -2]
0010	[12 -1 +3]	[-25 105 42]
0011	[15 0 +1]	[-32 202 22]
0012	[-3 1 -1]	[-4 -15 6]
0013	[-6 2 -1]	[-10 -29 -9]
0014	[-9 0 -2]	[-13 -85 -1]

0015	[-12 +1 -2]	[-25 -139 -2]
0016	[-15 1 -2]	[-32 -222 +2]

Table 3.2: data of the images with camera movement

3.2.3. Fixed camera and increasingly crack dimension images

For the second test, 10 images splitted in two sets of five are generated where each set presents fixed imposed camera angles and vector t, but increasingly crack dimension, respectively of 2, 5, 10, 20, 50mm (the first set is shown in Figure 3.11). This particular choice is performed in order to see if the errors in the final results are somehow influenced by the crack dimension.



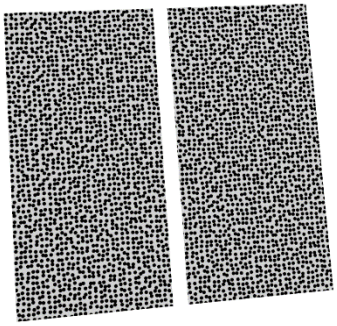


Figure 3.11: first set of five images with increasingly crack dimension (the camera orientation is fixed in the space during all the acquisition).

3.3. Numerical tests results

In this paragraph the 3 variants of the algorithm ('H 3D', 'H 2dD', 'PC 3D' as discussed in Paragraph 0) are analysed for all the images. As described before, two tests are carried out, one with fixed crack dimension and moving camera and one with increasingly carck dimension and fixed camera. Results are reported in the following paragraphs here below.

3.3.1. Fixed crack dimension and moving camera results

The analysis starts with the 30 images where the crack size is fixed to 10 mm while the camera is changing its position in the space in order to test the goodness of the approach for several camera angles. Moreover, different reference images are considered to validate the goodness of the realignment to the optical axis. The considered case is a static analysis, since each frame has always the same crack dimension.

Firstly the reference image 0 is considered. The final measurement result, shown in Table 3.3, consisting in the average of the average values coming from each image processed, is:

H 3D	H 2D	PC 3D
9.9996mm +/- 0.0025 mm	9.9978 +/- 0.0013 mm	9.9996mm +/- 0.0025 mm

Table 3.3: final measurements results for the three proposed approach

With the uncertainty expressed as the standard deviation of the averages of each image. It's possible to see that 'H 3D' and 'PC 3D' in a perfectly ideal case produce the exact same output, while 'H 2D' seems to perform slightly better in terms of data dispersion.

The accuracy is around 1-2 thousandths of millimeters for this static analysis.

Considering all the 30 images, in Table 3.4 maximum and minimum values are reported for average, standard deviation and RMS error:

	H 3D		H 2D		PC 3D	
	min	max	min	max	min	max
Average [mm]	9.9944	10.0052	9.9952	10.0003	9.9944	10.0052
Std dev [mm]	0.0024	0.0058	0.0024	0.0044	0.0024	0.0058
RMS error [mm]	0.0026	0.0079	0.0026	0.0058	0.0026	0.0079

Table 3.4: minimum and maximum values of the 30 images

The RMS error is defined as in Paragraph 2.5.

For the first reference image the range of the average is of 1 hundredth of millimeters for 'H 3D' and 'PC 3D', while it is halved with 'H 2D'. The standard deviation and RMS error for each cracked image processed are always below 1

hundredth of mm. The RMS error is bigger than the standard deviation since it is calculated on the actual known crack size (10 mm).

The same results (presented from Table 3.5 to Table 3.16) are calculated for each reference image to be sure that the procedure regarding the realignement perpendicular to the optical axis is free from errors:

- Ref 1:

H 3D	H 2D	PC 3D
9.9996mm +/- 0.0027 mm	9.9969 +/- 0.0013 mm	9.9996mm +/- 0.0027 mm

Table 3.5: final measurements results

Even with this reference image the retrieval of the average crack value is more accurate in terms of central value with 'H 3D' and 'PC 3D', while 'H 2D' shows less variability.

	H 3D		H 2D		PC 3D	
	min	max	min	max	min	max
Average [mm]	9.9938	10.0067	9.9939	10.0004	9.9938	10.0067
Std dev [mm]	0.0031	0.0068	0.0031	0.0049	0.0031	0.0068
RMS error [mm]	0.0033	0.0092	0.0036	0.0076	0.0033	0.0092

Table 3.6: minimum and maximum values of the 30 images

The maximum range for the average is below 1.5 thousandths of millimeters for 'H 3D' and 'PC 3D' and even lower with 'H 2D'. Standard deviation and RMS error are contained below 1 hundredth of mm.

- Ref 2:

H 3D	H 2D	PC 3D
10.0008 mm +/- 0.0036 mm	10.0018 +/- 0.0151 mm	10.0008 mm +/- 0.0036 mm

Table 3.7: final measurements results

With this reference image 'H 3D' and 'PC 3D' work well while the 'H 2D' approach shows some criticality with the uncertainty that reaches 1.5 hundredths of millimeters.

	H 3D		H 2D		PC 3D	
	min	max	min	max	min	max
Average [mm]	9.9961	10.0119	9.9960	10.0814	9.9961	10.0119
Std dev [mm]	0.0028	0.0059	0.0029	0.0378	0.0028	0.0059
RMS error [mm]	0.0027	0.0122	0.0029	0.0845	0.0027	0.0122

Table 3.8: minimum and maximum values of the 30 images

The maximum range for the average is around 1.5 hundredths of mm with 'H 3D' and 'PC 3D' while is higher, slightly below 1 tenth of mm for 'H 2D'. Standard deviation and RMS error are higher than the last two reference images, in particular for 'H 2D'.

The loss of accuracy with 'H 2D' approach could be symptomatic of an estimation for the movement correction not as robust as the one performed with 3D data. 'H 3D' based on the homography calculated with the rigid rototranslation values on centre subset point clouds is more repeatable than the projective estimation on 2D centre subset points calculated with the MSAC non deterministic algorithm.

- Ref 3:

H 3D	H 2D	PC 3D
10.0010 +/- 0.0029 mm	9.9979 +/- 0.0012 mm	10.0010 +/- 0.0029 mm

Table 3.9: final measurements results

	H 3D		H 2D		PC 3D	
	min	max	min	max	min	max
Average [mm]	9.9949	10.0072	9.9955	10.0003	9.9949	10.0072
Std dev [mm]	0.0029	0.0071	0.0030	0.0049	0.0029	0.0071
RMS error [mm]	0.0028	0.0079	0.0032	0.0031	0.0028	0.0079

Table 3.10: minimum and maximum values of the 30 images

The results for average, standard deviation and RMS error are good and comparable with the first two reference images cases.

- Ref 4:

H 3D	H 2D	PC 3D
9.9989 +/- 0.0026 mm	9.9999 +/- 0.0013 mm	9.9989 +/- 0.0026 mm

Table 3.11: final measurements results

This case presents very accurate results for 'H 2D' both in terms of central value and dispersion. 'H 3D' and 'PC 3D' provide satisfying results.

	H 3D		H 2D		PC 3D	
	min	max	min	max	min	max
Average [mm]	9.9948	10.0069	9.9972	10.003	9.9948	10.0069
Std dev [mm]	0.0027	0.0073	0.0026	0.0047	0.0027	0.0073
RMS error [mm]	0.0031	0.0085	0.0025	0.0046	0.0031	0.0085

Table 3.12: minimum and maximum values of the 30 images

It is remarkable here the contained range for the average for 'H 2D' approach as well as the standard deviation and RMS error.

Comparing this case (Ref 4) with Ref 2 it is possible to notice the variability in the behaviour of 'H 2D' approach. While in Ref 2 it reaches unsatisfying results, here it provides the best ones. This aspect comes from the different approach in the estimation of the homography. Even if data dispersion is tendentially lower with 'H 2D', it doesn't mean that this behaviour is repeatable every time, as anticipated in Ref 2 case. The results of this static simulation analysis show that 'H 3D' is more repeatable in all the different considered cases.

- Ref 5:

H 3D	H 2D	PC 3D
10.0002 +/- 0.0027 mm	9.9979 +/- 0.0013 mm	10.0002 +/- 0.0027 mm

Table 3.13: final measurements results

	H 3D		H 2D		PC 3D	
	min	max	min	max	min	max
Average [mm]	9.9945	10.0062	9.9955	10.0009	9.9945	10.0062
Std dev [mm]	0.0027	0.0064	0.0026	0.0045	0.0027	0.0064
RMS error [mm]	0.0026	0.0070	0.0027	0.0057	0.0026	0.0070

Table 3.14: minimum and maximum values of the 30 images

- Ref 6:

H 3D	H 2D	PC 3D
9.9991 +/- 0.0025 mm	9.9967 +/- 0.0013 mm	9.9991 +/- 0.0025 mm

Table 3.15: final measurements results

	H 3D		H 2D		PC 3D	
	min	max	min	max	min	max
Average [mm]	9.9940	10.0046	9.9944	9.9993	9.9940	10.0046
Std dev [mm]	0.0026	0.0063	0.0027	0.0045	0.0026	0.0063
RMS error [mm]	0.0026	0.0078	0.0030	0.0064	0.0026	0.0078

Table 3.16: minimum and maximum values of the 30 images

Results from Ref 5 and Ref 6 images follow the average behaviour of the other reference images cases. The central value is retrieved better with 'H 3D' and 'PC 3D' while 'H 2D' shows less data dispersion within the considered case.

To resume, the results from the simulated environment are satisfying in average terms, since the maximum range (difference between minimum and maximum average values) is always either below or around 1.5/100 of mm (with exception of Ref 2 evaluated with 'H 2D' method, suggesting that the calculation passing through 3D data is more robust compared to the one on 2D).

About the dispersion evaluated with the three different approaches 'H 3D, 'H 2D' and 'PC 3D', it's possible to notice that the final results are very similar and limited below the 1/100 of mm.

With the DIC it is expected to reach an accuracy of the order of 1/100 of pixel [1]. In all the images processed the peak to peak difference remains limited to 2 or 3 hundredth of mm (as it is possible to notice also in the colorbars of Figure 3.12). Considering the scale factor (Eq.3.7):

$$\mathrm{conv_x} = -\frac{\mathrm{d}}{\mathrm{f_x}} = 1.57\,\frac{mm}{pix} \qquad 3.7$$

It is possible to conclude that the remaining minimal uncertainty is caused by the DIC procedure and not by algorithm mistakes.

It is remarkable to underline that usually small trends appear at higher rotation angles sometimes for the reference and sometimes for the moving part, not around the roll axis, where there is only rotation around the axis coming out from the camera. This is shown in the following set of figures which are called Figure 3.12.

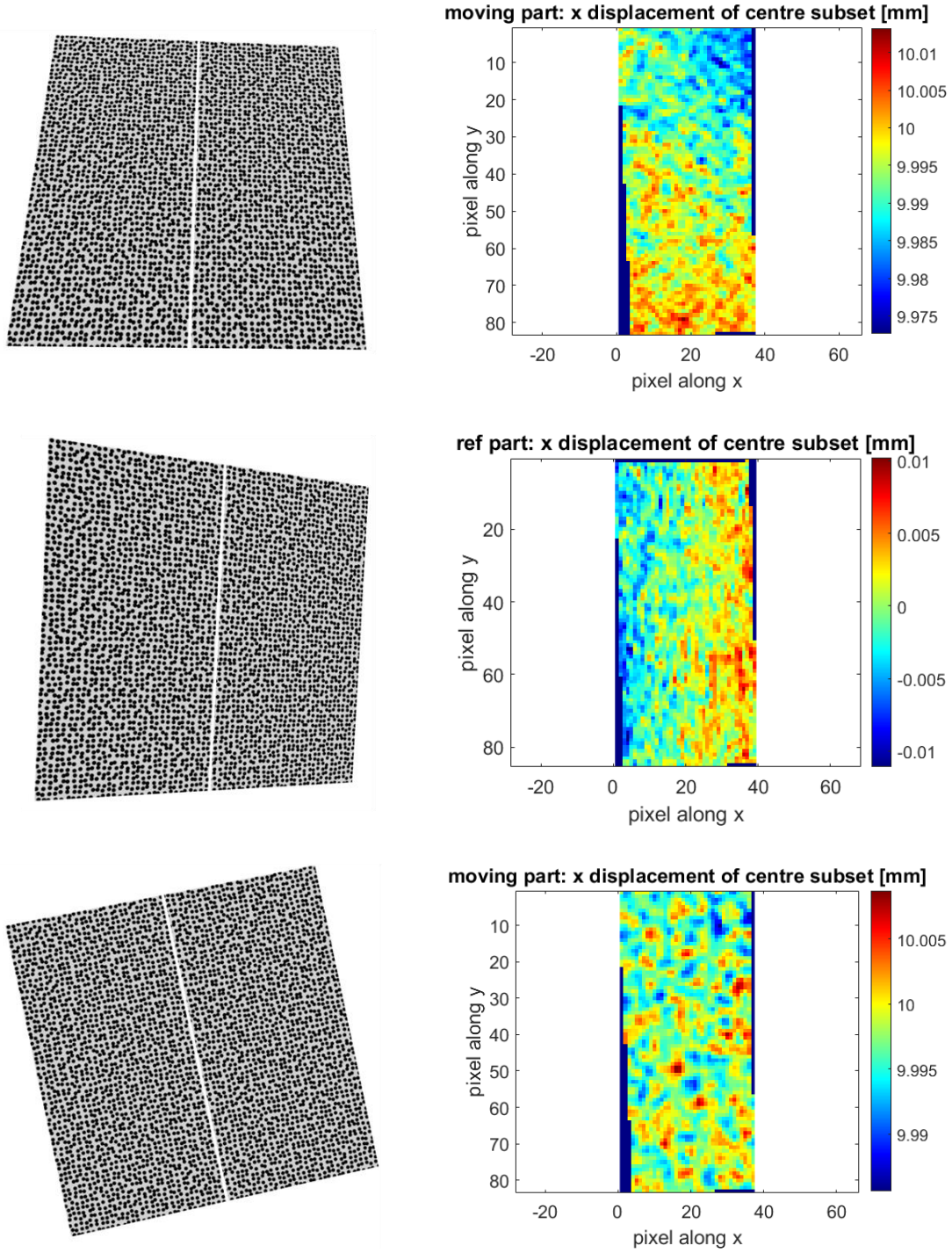


Figure 3.12: trends appear at high rotation angles both for pitch and yaw angles and both for moving and reference part (with the exception of the x-axis rotation)

A more detailed analysis on the results regarding the influence of the angle values of rotation is here presented, as it can be seen in Figure 3.13.

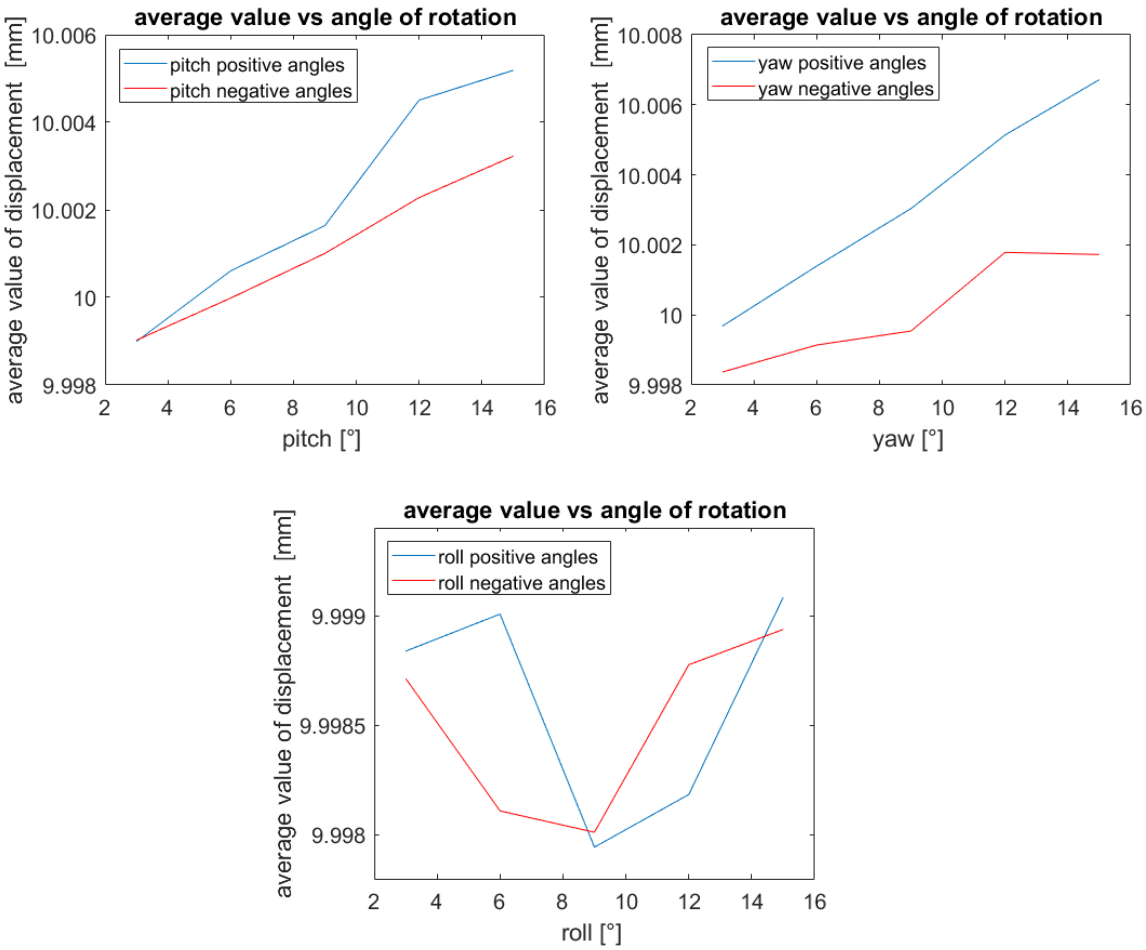


Figure 3.13: the average value of displacement of the ficticious crack shows a trend when the angle increases for pitch and also yaw, while for roll the behaviour is random.

With pitch and yaw, when the angle increases, also the average value of displacement increases with a relationship that seems more or less linear. With roll no trend is visible, data are distributed randomly around the mean. This behaviour is probably caused by the perspective distorsion that comes with pitch and yaw rotations and not with the roll one. Nevertheless, the trend is contained below 1 hundredth of millimeters, so it is comparable with the uncertainty of the digital image correlation.

For the values of standard deviation and RMS error the trend is always increasing (as it can be seen from Figure 3.14) in all the cases considered with the increase of the angles. Usually the RMS error is above the standard deviation since it is calculated on the actual value of the crack and not around the average.

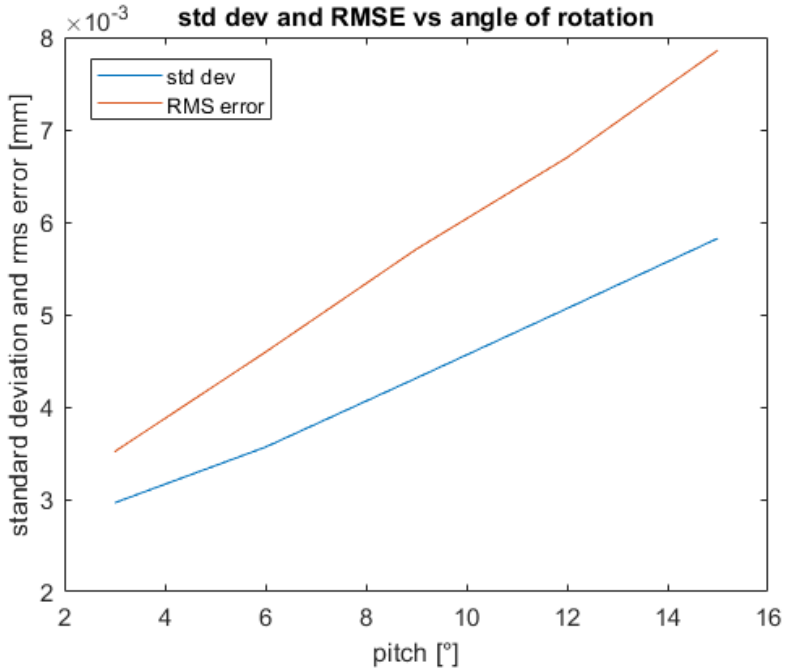


Figure 3.14: plot of RMS error against standard deviation for different values of pitch angle

These graphs suggest that, if possible, the angle of rotation of the drone must remain as low as possible, even though the increase of percentage error (compared to the actual value) is very small. It could be due to the imperfections in the image generation and not to the algorithm itself, so it's not as significant as it visually appears in the graphs.

3.3.2. Fixed camera and increasingly crack dimension results

The analysis here proposed is focused on two set of five images with fixed imposed camera angles and vector t, but increasingly crack dimension, respectively of 2, 5,

10, 20, 50mm (as already pointed out in Paragraph 3.2). This particular choice is performed in order to see if the errors in the final results are somehow influenced by the crack dimension.

In Figure 3.15 the results of the analysis for both the sets are presented, with particular attention on the absolute difference between the measured value and the actual one (2, 5, 10, 20, 50 mm respectively).

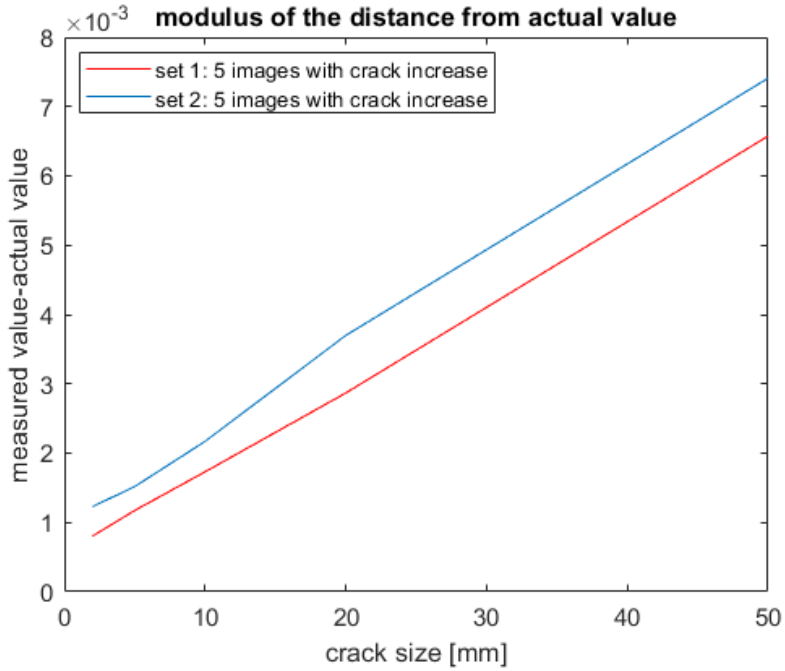


Figure 3.15: difference between measured value and actual value of the crack for increasing value of the crack size

In the case in which only the crack size is increasing, (while the orientation is fixed) the difference between the measured value and the actual value has a strong linear trend. This is probably related to a non perfect calculation of the scale factor, that causes the linear magnification of the error when considering higher crack values.

To conclude, the results from the simulated environment are satisfying in average terms. As already said, the remaining uncertainty is comparable with the DIC uncertainty. This validates the proposed algorithm.

4 Experimental validation of the technique

After eliminating all possible external causes of error not closely related to the algorithm and establishing that it works in ideal situations (Chapter 3) it is decided to test it in the experimental field, specifically in a laboratory environment to further minimize any criticality.

Before moving on to tests involving the use of the sensor installed on the drone (with all the various problems that can arise in such conditions), it is appropriate to test (through basic experimental tests) the chosen sensor that might not even have turned out to be the most suitable for this type of work.

The experiment consists in the analysis with the chosen sensor of the crack generated on a panel with the algorithm presented in (Chapter 2). The RGB-D sensor is held with hands to simulate the movements of the drone. The panel is loaded with a hydraulic press controlled with a PID (Figure 4.1).

In order to have an assurance of the quality of the results obtained, the measurements are compared with those obtained from a stereo system (Paragraph 4.1.2). The analysis is performed using the DIC 3D software.

A more detailed explanation of the sensor used for the various acquisitions is also given. Moreover, the entire experimental setup used to conduct the test in the laboratory environment is highlighted.

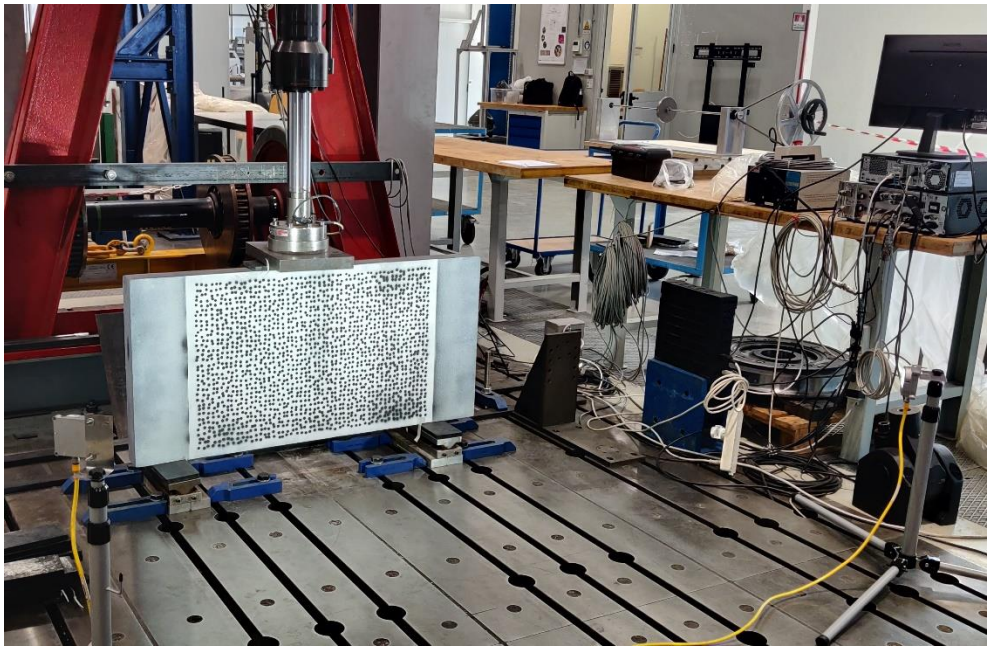


Figure 4.1: picture of the overall setup with piston and panel highlighted

4.1. Experimental setup description

In this paragraph the overall experimental setup used is presented.

In particular, the description of the specimen and of the loading machine is proposed, as well as an overview of the stereo system used as ground truth.

4.1.1. Three point bending system

The specimen is a xps panel has dimensions of 120cm x 60 cm. The first step is the staining of the specimen of white with spray paint. With the software proposed by [37] a proper speckle is then generated considering the working distance of the RGB-D sensor (approximately 2 meters).

A stencil, that works as a mask, is built, with holes where the black dots of the speckle are present. The stencil is superimposed to the panel, fixed at it, and sprayed with dark matt black, thus obtaining the black dots, as shown in Figure 4.2.

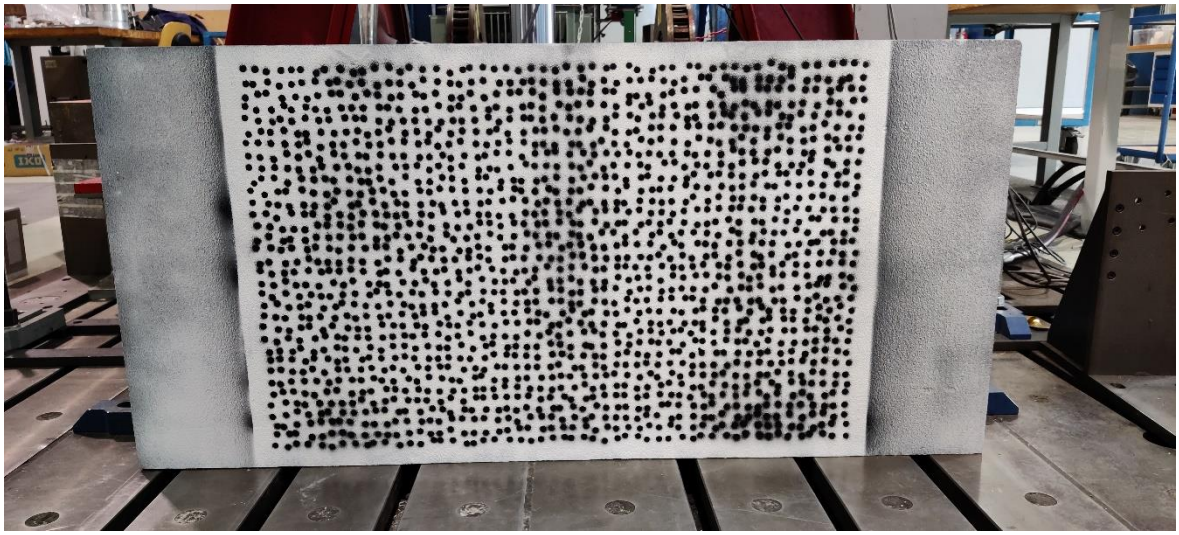


Figure 4.2: picture of the panel with the speckle applied

To avoid a shock fracture of the panel when loaded with the hydraulic press, that would not allow a proper analysis, a crack starter point is made below the piece as shown in Figure 4.3.

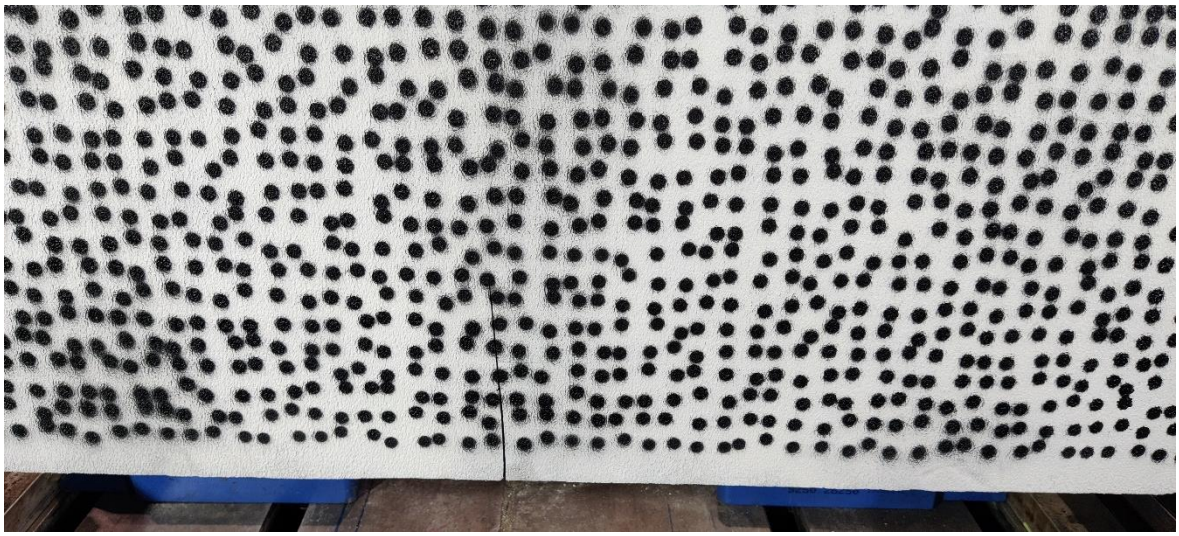


Figure 4.3: crack starter point below the piece put in evidence

Then, the pre-cracked panel is loaded with the hydraulic piston to analyze the behaviour of the crack.

The movement of the piston can be either controlled with fixed steps displacement or with cycles.

It is decided to work with cycles. In this way the focus is on the dynamic response of the panel. A sinusoidal cycle is imposed as input to the hydraulic piston. With the fixed parameters of the PID, the hydraulic piston couldn't follow precisely the imposed cycle, nevertheless the shape of the signal is accepted for cyclic dynamical loading.

The frequency of this "sinusoidal" wave is 0.5 Hz with a peak to peak distance of approximately 3mm.

4.1.2. Ground truth 3D DIC reference system

To compare the results of the RGB-D sensor with a precise reference, a stereo system with 2 cameras to perform the DIC 3D is set up as shown in Figure 4.4.

Figure 4.4: picture of the stereo system used

More information about the cameras used are shown here:

Prosilica GX 3300	
Sensor	Kodak KAI-08050
Type	CCD Progressive
Resolution (pixel)	3296x2472
Max frame rate (fps)	17
Focal length (mm)	50
Camera lens	Zeiss Makro-Planar T* f2/50 ZF.2

Table 4.1: Stereo system camera main information

The stereo system with the use of DIC 3D easily allows the reconstruction of the 3D deformation fields based on the triangulation of deformation fields of the two 2D images taken from the cameras.

The stereo system is calibrated with a set of images once the desired position, focus and diaphragm aperture is fixed. For the calibration procedure a different software (DIC 3D from Correlated Solutions) is used and, accordingly, also the calibration panel should be different (see Figure 4.5).

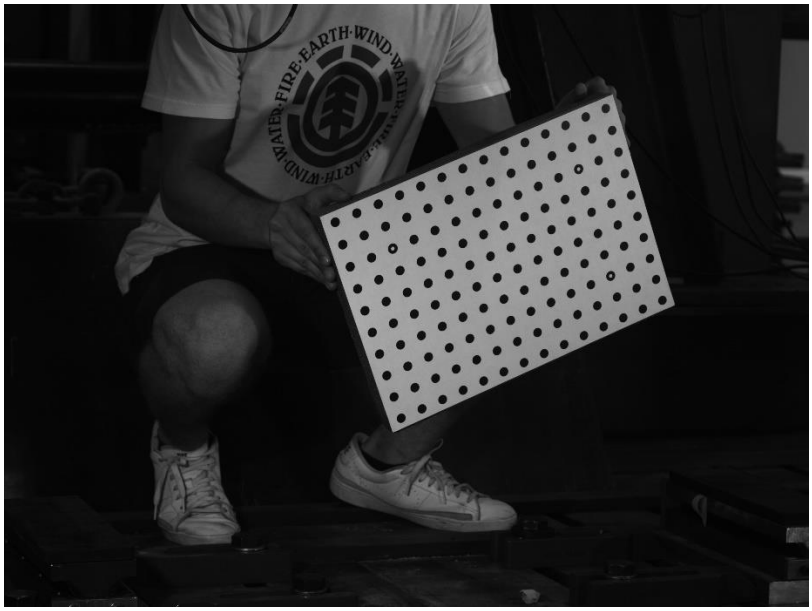


Figure 4.5: calibration panel used for stereo system

The 2 cameras are working with a simultaneous acquisition thanks to a triggering system. The acquisition rate is 5 fps.

The stereoscopyc system allows to get information about x-y-z diplacements, strains and also about the shape of the panel. This is also important to understand if the plane hypothesis of the panel is acceptable or not.

4.2. Measurement system of the technique

In this paragraph a detailed explanation regarding the sensor fusion and the calibration procedure related to the chosen RGB-D system used is proposed.

4.2.1. RGB-D sensor assembly

The overall sensor is composed by a Flir monochrome camera coulped with a ToF sensor (Blaze) (Figure 4.6).

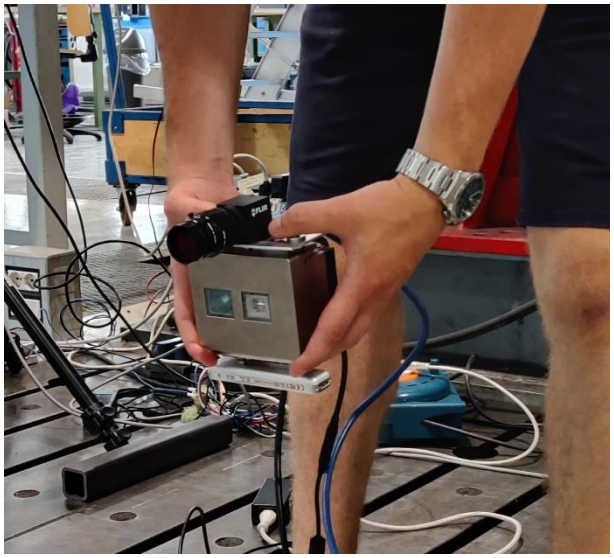


Figure 4.6: RGB-D sensor handheld to simulate the movement of the drone

More information about the overall sensor are reported respectively in Table 4.2 and Table 4.3:

FLIR Monochrome Camera	
Model number	BFS-U3-16S2M-CS
Type of sensor	Progressive Scan CMOS
Resolution (pixel)	1440x1080
Pixel size (μm)	3.45 x 3.45
Type of shutter	Global
Exposure time	4μs- 30s

Table 4.2: Flir monochrome camera main parameters

Blaze – 101 ToF	
Sensor Type	Sony IMX556 Area Scan CMOS (Time-of-Flight)
Measuring Method	Time-of-Flight
Resolution (pixel)	640x480
Range (m)	Short: 0.3–1.5 Long: 0.3–10
Accuracy (typical)	±5 mm (0.5–5.5 m)
Frame Rate	20 fps (Default) 30 fps (FastMode)

Table 4.3: Blaze – 101 ToF main parameters

Moreover, for both sensors, it is important to note that the acquisition is done in motion to simulate the real experimental case on drone.

To do this, the sensor isn't fixed to the ground, (as is the case with the stereo system) but is held in hand.

This can lead to significant blurring problems.

To try to avoid this risk, it is necessary to change the exposure time and to consider all the possible parameters carefully. In particular:

- exposure time: it is set equal to 2ms for the FLIR (with a gain of 10) and 0.7 ms for the BLAZE. This could lead to the problem of dark acquisitions in an indoor environment, that's why LED lamps are added in the experimental environment pointing the panel;
- acquisition frame rate: it is set equal to 60 Hz for the FLIR and 20 Hz for the blaze. This will lead for sure to the problem of not correct synchronization between the depth acquisitions and the color ones. Having such a high frame

rate for the Flir allows to choose the closest depth acquisition (in terms of time) related to the greyscale one, without introducing interpolation between the images (better explanation at Paragraph 4.3.1);
- depth range: for the Blaze a depth range between 1 and 3 meter is defined.

Other issues that must be taken into account regard respectively the scene complexity and the presence of material with different refractive index.

For the first issue, when light gets reflected multiple times before being received by the camera's ToF sensor, the delay and displaced location of the reflected light negatively impacts depth calculations [38]. For instance, if the surface of the target is very complex, with multiple concave and convex corners in a variety of angles, or if the target is placed in an environment with various angled surfaces, the optical paths of the ToF emitted light combine and interfere with one another.

While for the second issue it has to be noted that transparent materials, such as plastic or glass bottles, let too much light pass through their bodies and so they make poor targets for ToF cameras because they lack sufficient reflectivity [38].

All these aspects are minimised in the experimental application because the panel is approximately planar; the only remaining criticality is the presence of the black speckle on a white background, since black colour is perceived as farther away respect to its real position. The estimation of the plane must be done with a larger set of points from the point cloud to have a more robust result.

4.2.2. RGB-D sensor calibration procedure

As already discussed for the virtual camera (Chapter 3) to be sure that the parameters used in the algorithm are correct a calibration procedure is required.

The calibration is carried out by means of the Camera Calibration Toolbox for MATLAB.

For this procedure a checkerboard (Figure 4.7) is generated which presents 9 times 12 squares (rows times columns) with edges of 30 mm each.

The distance at which the acquisitions are taken is approximately the same of the one used during the experimental acquisition to avoid problems of blurring.

It's important to point out that since there are two sensors (one for color acquisition and one for depth) the calibration must be performed twice and then a stereo calibration should be carried out too, in order to retrieve the rototranslation between them.

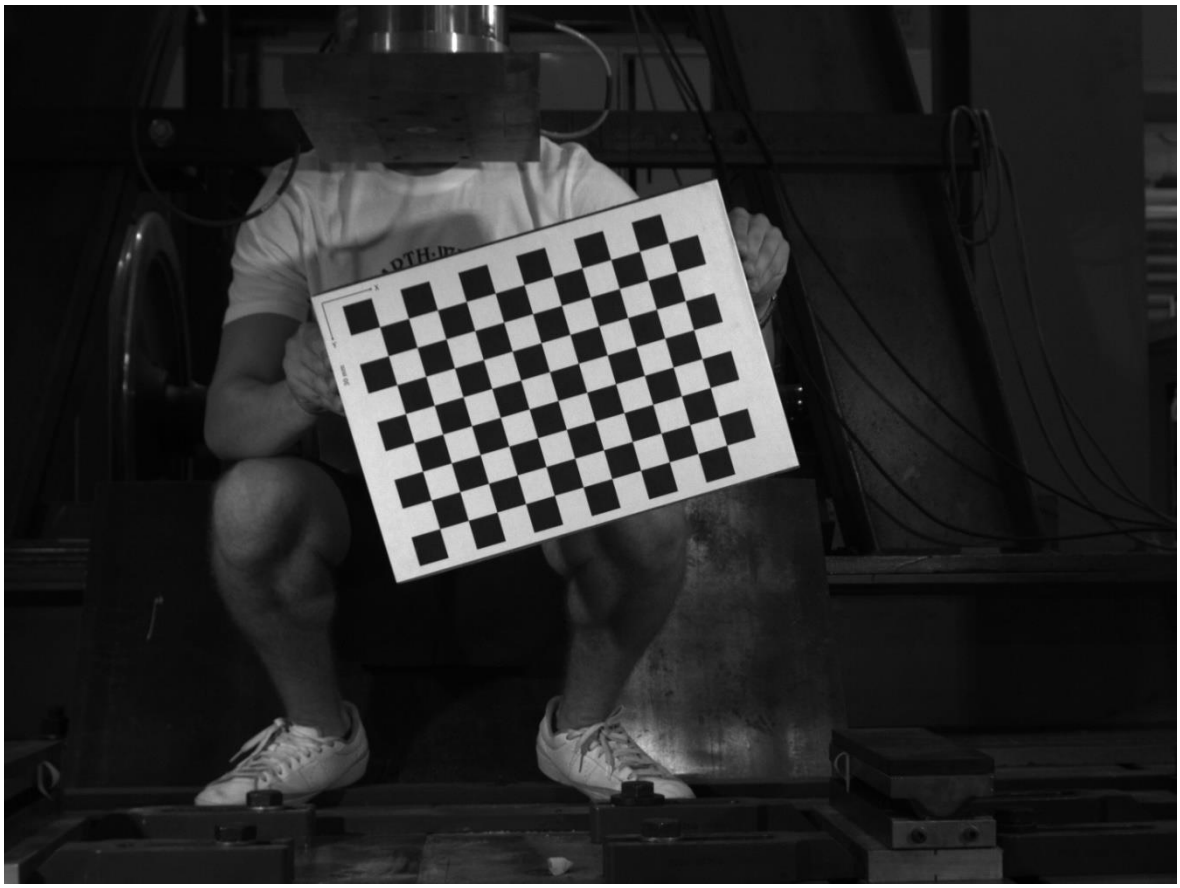


Figure 4.7: checkerboard used during the calibration procedure for the RGB-D sensor

In addition, it is decided to perform the calibration considering:

- a possible radial distortion in the color camera (only for the first two orders and putting the tangential one to zero);
- no distorsion for the ToF sensor (as indicated in the datasheet);
- fixed center for both the sensor placed in the exact center considering half of the resolution.

The final results are shown in the following Table 4.4:

Color camera intrinsic parameters matrix (pixels)	$\begin{matrix} 2458.5 & 0 & 0 \\ 0 & 2465.9 & 0 \\ 719.5 & 539.5 & 1 \end{matrix}$
Depth camera intrinsic parameters matrix (pixels)	$\begin{matrix} 521.3 & 0 & 0 \\ 0 & 522.2 & 0 \\ 319.5 & 239.5 & 1 \end{matrix}$
Radial distorsion coefficient color camera	$2^{nd} order$: 0.0084 $4^{th} order$: 0.0843

Table 4.4: sensors' intrinsic parameters matrix with the addittion of radial distorsion coefficient for the color camera

Anyway, the most critical aspect regards the realignement of the two sensors.

In fact, a shift between the color and the depth images is for sure present and it should be compensated by applying the approriate rototranslation.

The overall explanation of this procedure is shown in Paragraph 4.3.1

However, here below in Table 4.5 are already shown both the rotation matrix and the translation vector that allow the correct realignement of the two sensor.

Rotation matrix	$\begin{matrix} 0.9996 & -0.0036 & 0.0271 \\ 0.0044 & 0.9995 & -0.0309 \\ -0.0270 & 0.0310 & 0.9992 \end{matrix}$
Translation vector	$\begin{matrix} -25.86 \\ -66.70 \\ 61.09 \end{matrix}$

Table 4.5: rototranslation parameters that allow the correct coupling between ToF and Color sensor

4.3. Practical implementation of the algorithm

In this paragraph the implementation of the algorithm is explained, taking into account all the further improvements done, which are essential for dealing with a real world situation and with the sensor chosen for the test.

4.3.1. Further developments for real world application

As already explained in Paragraph 4.2.2, the calibration procedure is carried out for both the monochrome camera and the ToF sensor. Moreover, a shift between the greyscale and the depth images is for sure present and it should be compensated by applying the approriate rototranslation. This chapter shows in detail this procedure and how to manipulate the acquired data coming from the sensor.

The first step consists in the coupling of the greyscale image to the depth image. Since the acquisition rate for the two cameras is not the same, the images are not taken simultaneously. Knowing the timestamps of each acquisition both for the greyscale and the depth cameras, it is possible, with a MATLAB script, to associate to each greyscale image, on which the DIC will be performed, the nearest depth image in terms of timing. Due to the high frame rate of the Flir sensor (60 Hz) no interpolation between depth images is done.

Once the depth image is chosen (.bin format), it is converted in .ply format. Then, the obtained point cloud is rototranslated with the values presented in Table 4.5 to adopt the same reference system for the Flir and the Blaze.

To perform the fit of the plane, the region of the point cloud representing the panel must be detected. To do this, the idea is to bring the point cloud to the Flir sensor plane with the pin-hole model, following (Eq. 4.1):

$$\varphi m = K[R \mid t]M \qquad 4.1$$

Where:

- $M = [X, Y, Z, 1]^T$ is the 3D point in homogenous coordinates;
- $m = [x, y, 1]^T$ is the homogenous representation of the point reported to the sensor plane;
- φ: scaling factor;
- K: intrinsic matrix;
- $[R \mid t]$: rototranslation values (Table 4.5)

The points reported to the sensor can be saved or discarded based on the selection of the panel region on the greyscale image (Figure 4.8).

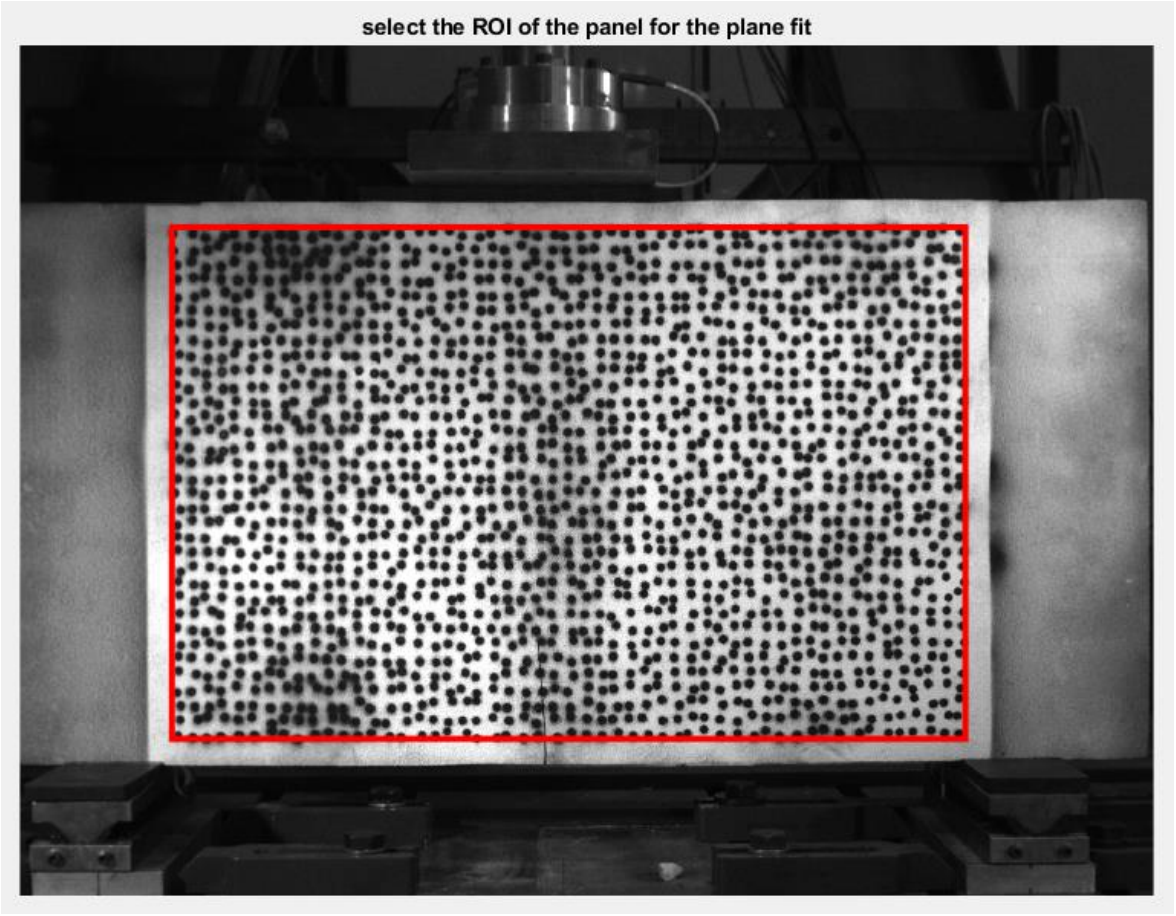


Figure 4.8: manual selection of the ROI for the plane fit

The biunivocal correspondence between the points on the sensor and the actual 3D points allows to retrieve the 3D model of the panel only (Figure 4.9). On these 3D points the fit of the plane is done and the analytical equation is taken out.

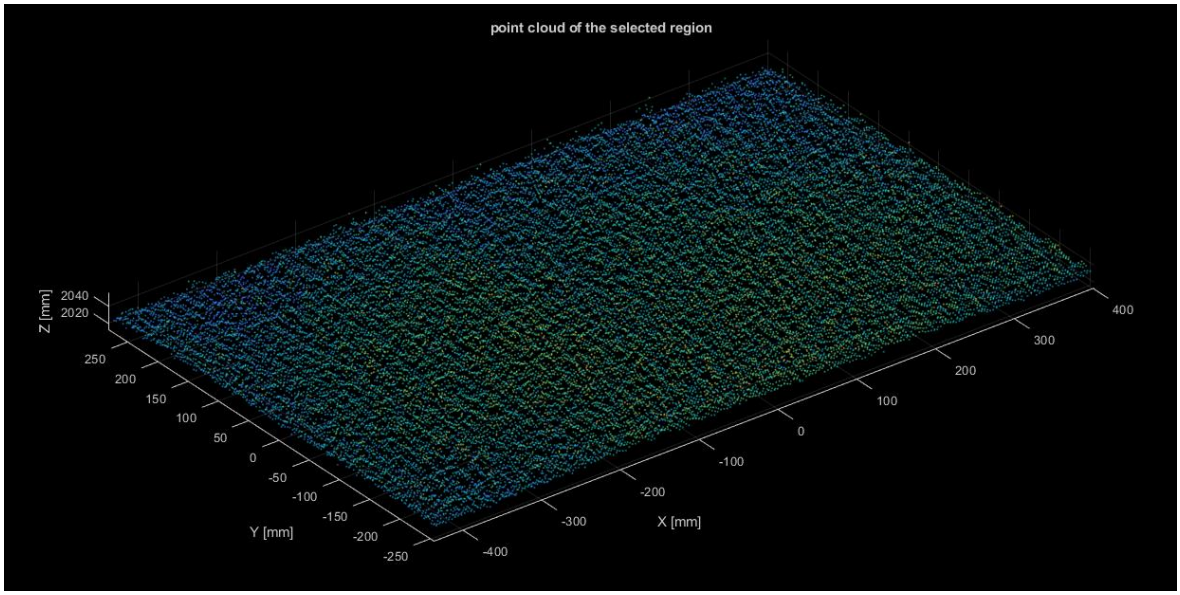


Figure 4.9: point cloud of the region selected

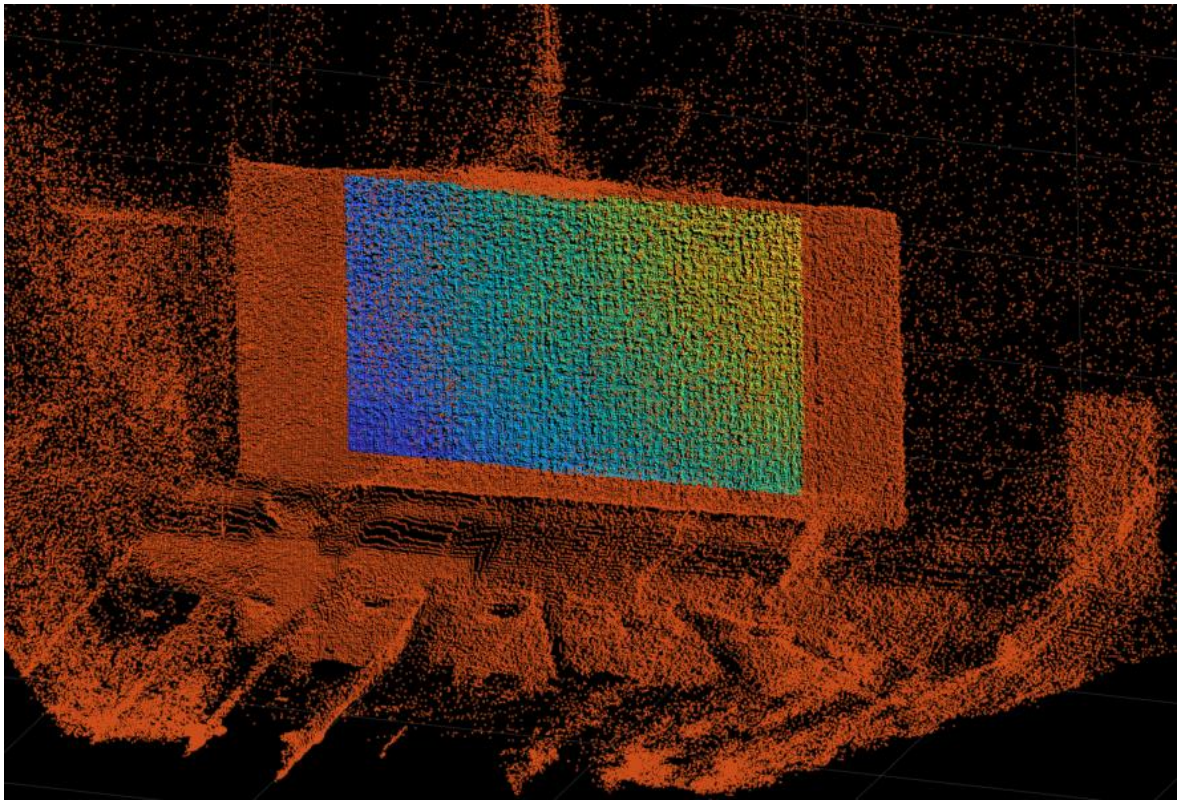


Figure 4.10: point cloud selected part on the entire point cloud in red

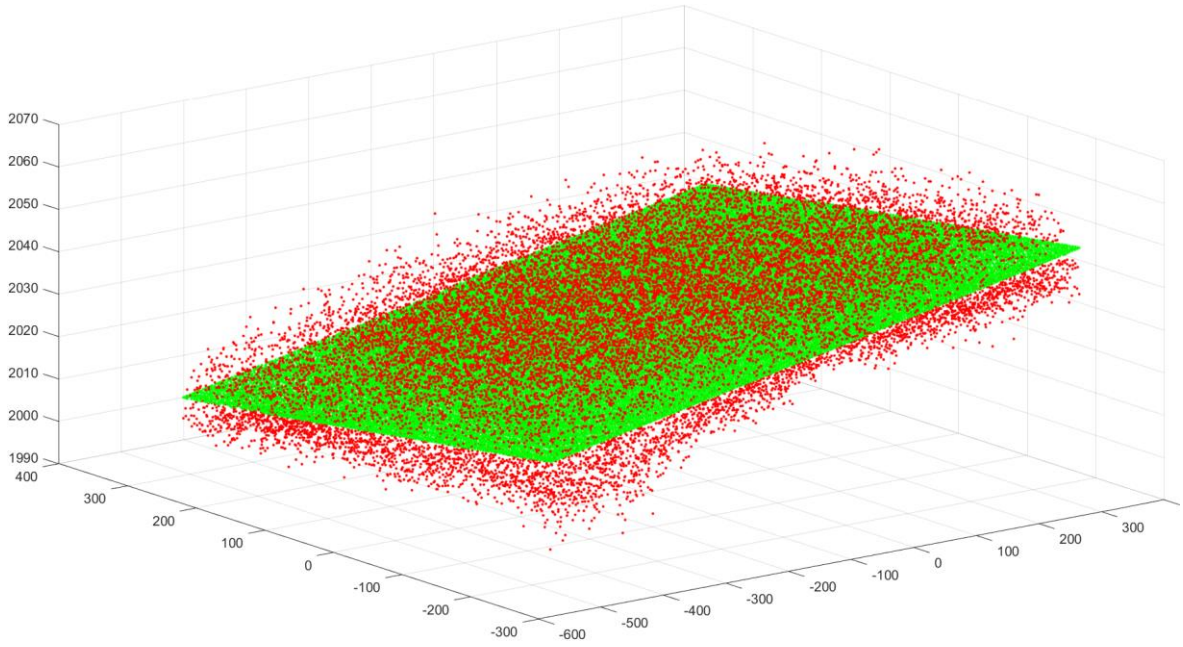


Figure 4.11: comparison between the plane model (green) and the selected point cloud (red)

As it is visible in Figure 4.11, the depth sensor presents a significant amount of noise that can add uncertainty to the measurement.

To give an idea the value of average RMS error on a set of 10 images on the z differences between the point cloud and the plane retrieved is 4.97 mm.

4.3.2. ROI definition for movement compensation in real application

In this paragraph considerations about the choice of the region of interest (ROI) are present. The ROIs considered are the ones for compensation of the movement of the drone and not the ones used for the final output (final crack opening).

This is important with experimental acquisitions because several issues to be taken into account arise.

Among them, it is worth highlighting:

- there is not necessarily a fixed reference zone that is not subjected to deformation;

- there is the possibility that out of plane displacement is present, thus affecting the quality of the results;
- the transformation could be not rigid;

To overcome these problems the following ROIs with their own pros and cons are chosen.

4.3.2.1. Small ROI close to the crack

As shown in Figure 4.12 a ROI in a small region close to the crack is selected.

This seems the most intuitive solution since that region should remain approximately undeformed however this isn't true for this particular case, since a gradient in the deformation is present, (for better understanding see Paragraph 4.4.2).

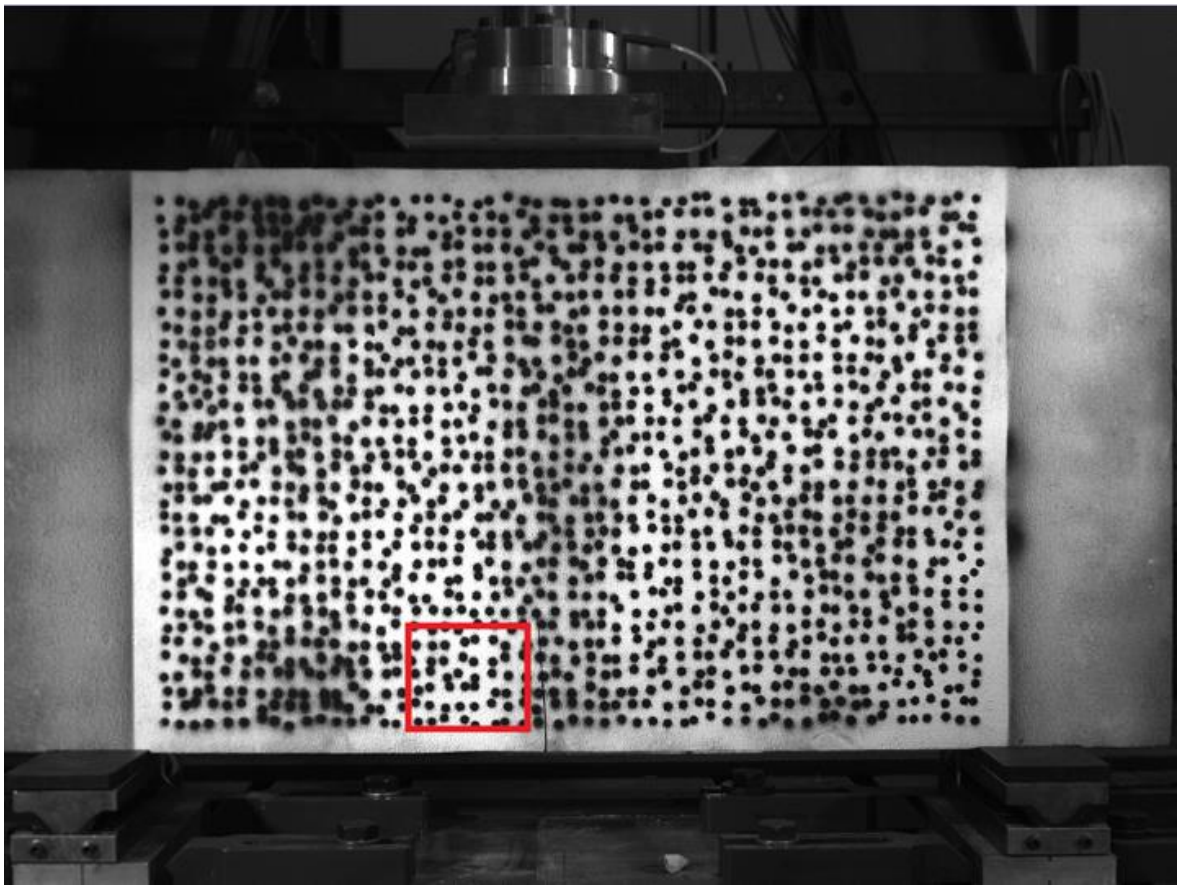


Figure 4.12: ROI put in evidence in red

At the end, this solution is rejected due to the fact that the ROI contains too few points (even with a smaller step choice in the DIC analysis). This brings to a non optimal estimation of the plane model and of the rototranslation or homography.

4.3.2.2. ROI as half of the panel

Similar to what is done with the simulator, a ROI on half of the specimen is selected (in this case the left one as shown in Figure 4.13). Anyway, different problems arise, in particular this region is for sure not subjected to a rigid deformation and referring all the information of the moving part to this reference makes the graphs difficult to be interpreted.

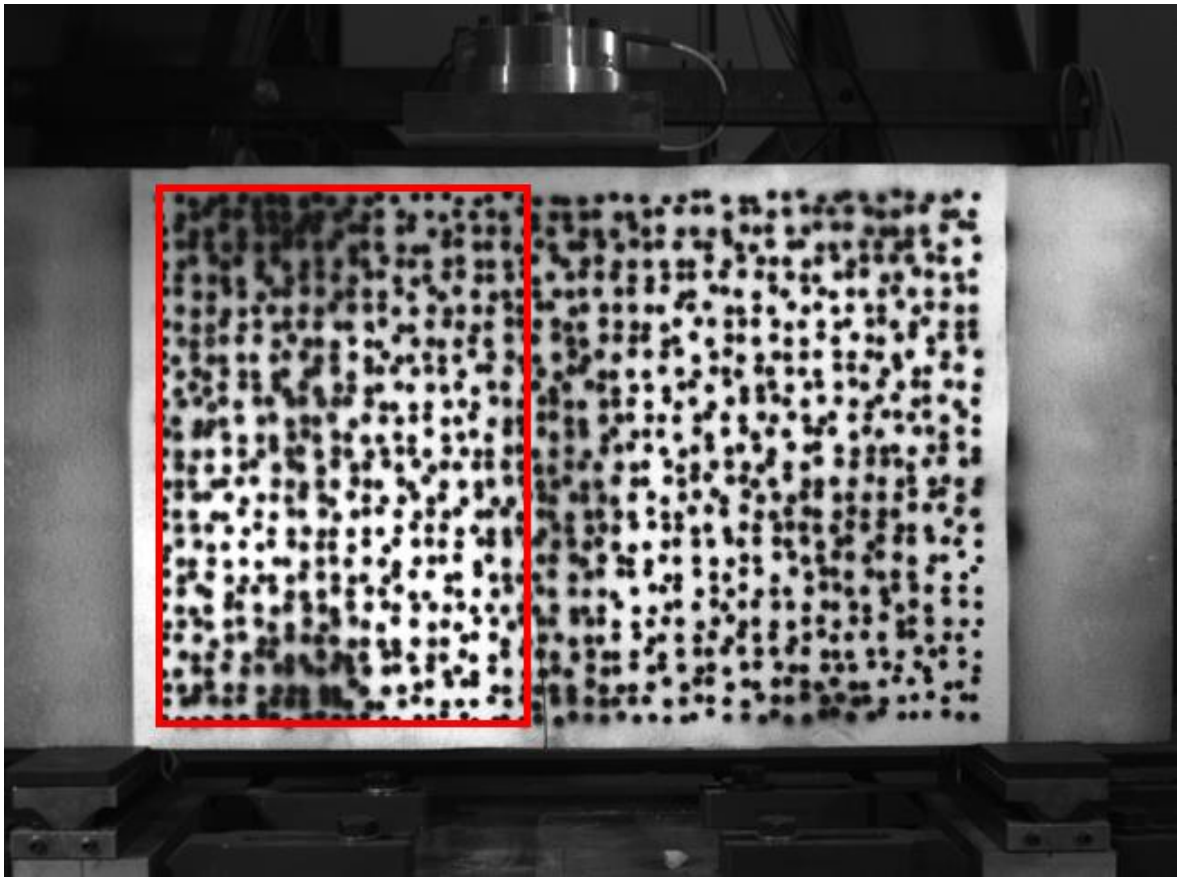


Figure 4.13: ROI put in evidence in red

4.3.2.3. ROI on all the panel

As already shown in Figure 4.8 the selection of the total specimen as ROI seems to be the best solution.

As a matter of fact, it seems reasonable that if the ROI is taken as a symmetric region, the effects of the left and right part of the panel are averaged out (different orientation of the panel between left and right, asymmetric deformation if only the left part is considered), leaving only the result of the deformation around the average.

This choice also makes the graphs easy to be read.

4.4. Experimental results

The analysis of dynamic loading process and the related results with an imposed cycle with the hydraulic piston are here discussed.

As already mentioned in Paragraph 4.1.1 a sinusoidal cycle is imposed as input to the hydraulic piston, however with the fixed parameters of the PID, the hydraulic piston couldn't follow precisely the imposed cycle. The frequency of this "sinusoidal" wave is 0.5Hz with a peak to peak distance of approximately 3 mm.

The approach considered in the experiment for the realignment of data from different frames is the homography one, due to the fact that the panel has an approximate plane shape. In this way the effects of imperfect or noisy depth are minimised.

The calculation of the homographies, as presented in Paragraph 0, can be done with an estimation from 2D data or by retrieving the rototranslation values from 3D data (Eq. 2.3). Typically the estimation coming from 2D data, since it's based on a minimization procedure, gives accurate results, especially in case of rigid movements. In this experiment a remarkable deformation of the panel complicates the analysis. Moreover, the plane assumption necessary to use homographies is not precisely respected considering the information coming from DIC 3D, as visible from a qualitative point of view in

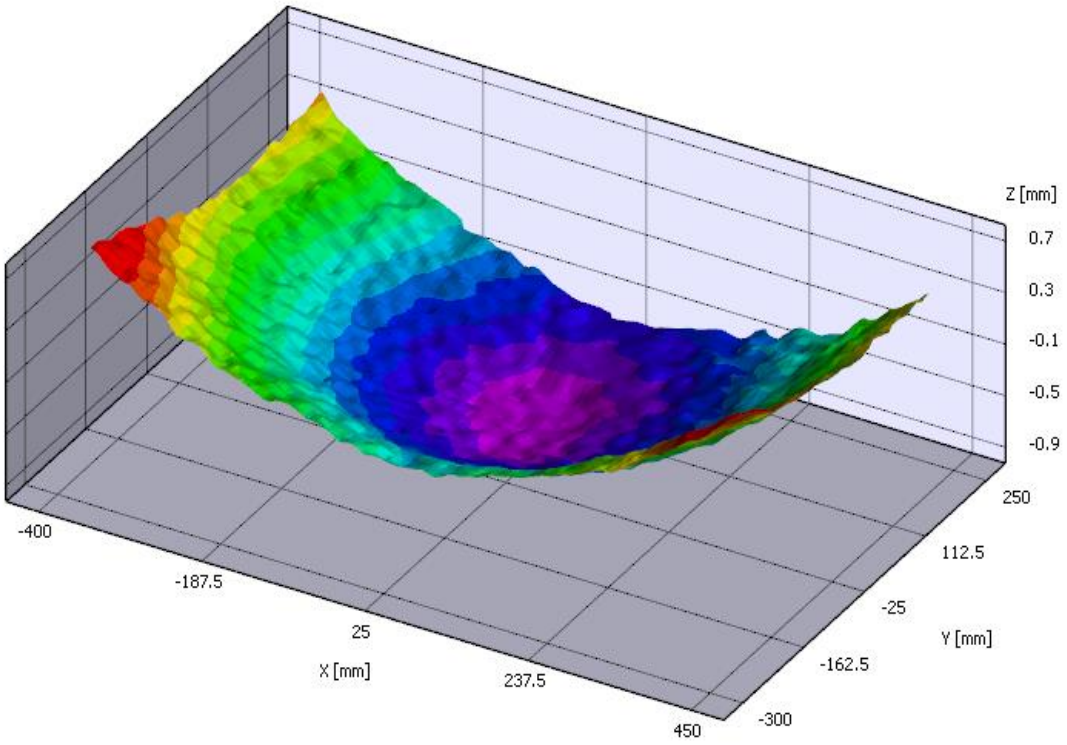


Figure 4.14: 3D deformation plot obtained through DIC 3D

4.4.1. Qualitative analysis

Initially, to validate the proposed technique, a qualitative analysis of the obtained graphs for 'H 2D' and 'H 3D' is carried out, comparing them with the ones obtained from DIC 3D (respectively in Figure 4.15 and Figure 4.16). In particular, in order to obtain a more precise comparison, a reference picture with close crack is chosen, but the sensor and the stereo camera aren't synchronized and they acquire with different frame rate, thus making it impossible to obtain ideal conditions.

As it is clearly visible the results obtained from 'H 2D' analysis (Figure 4.15) seems good in terms of crack reconstruction, in fact the peak to peak values in both x and y directions are similar; however, the 'H 2D' case loses some information especially for the x component in the upper part.

From a qualitative point of view, comparing the results with the 3D DIC taken as the reference, the 'H 3D' approach seems better than the 'H 2D' one. The explanation could be that some deformations effects, combined with a non perfectly plane panel, can be confounded as projective effects by the minimization algorithm of the 'H 2D', thus causing their undesired compensation, while the 'H 3D' approach is able to reconstruct with a better accuracy the deformation field.

Anyway the poor quality in terms of accuracy of the sensor (±5 mm for the Blaze) could strongly affect the quality of the final output in terms of crack reconstruction.

It must be noted that the convention of the software VIC 2D and VIC 3D are opposite for the y coordinate.

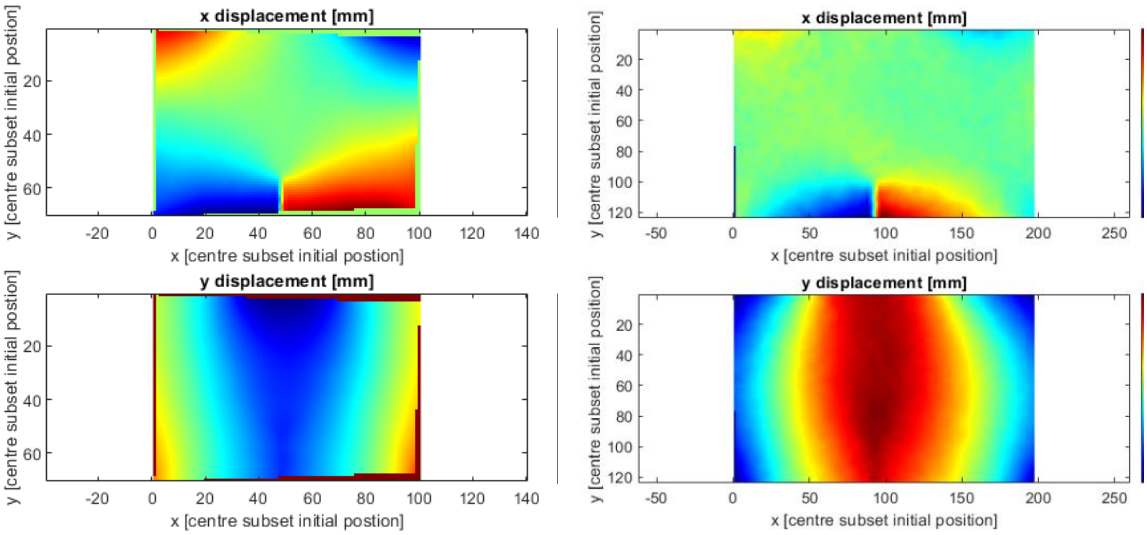


Figure 4.15: Data from DIC 3D on the left and from 'H 2D' on the right

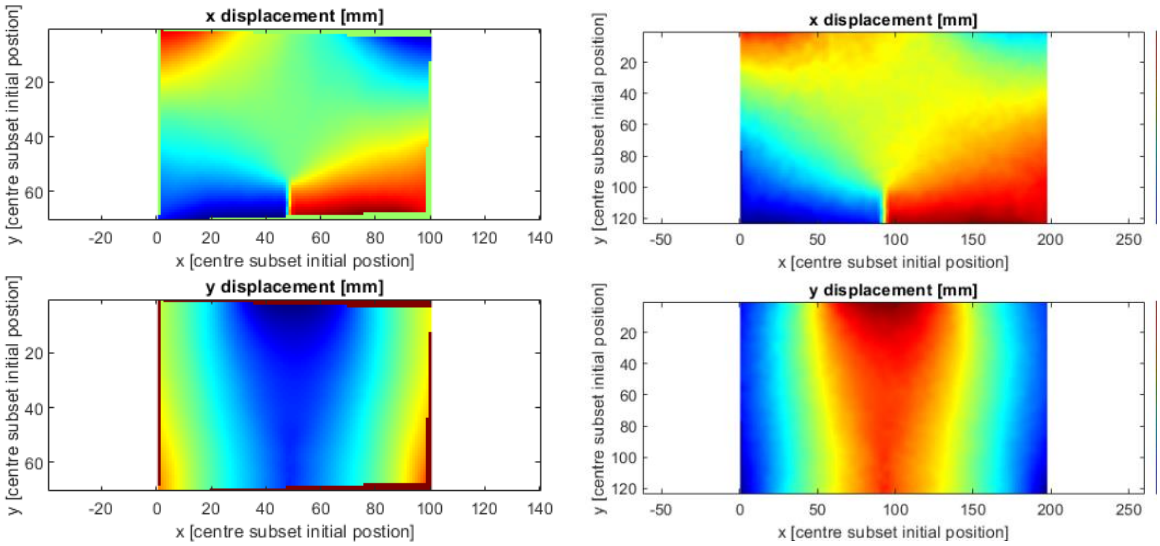


Figure 4.16: Data from DIC 3D on the left and from 'H 3D' on the right

4.4.2. Quantitative analysis

After the first qualitative analysis made by graphs comparison, two zones (one on the left and one on the right of the crack as shown in Figure 4.17) are chosen and the time history of the difference between their displacement is computed.

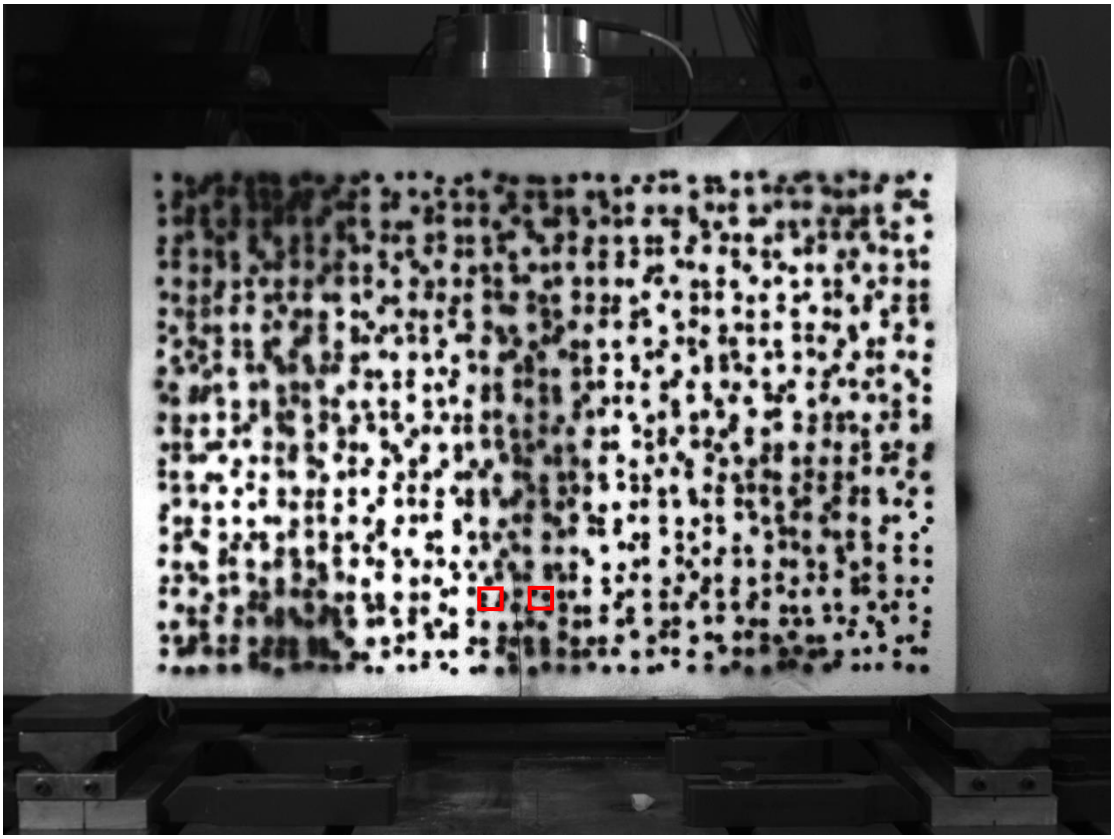


Figure 4.17: rectangles putting in evidence the chosen zone for the time history

The idea is to allow the user to select two regions (symmetric w.r.t. the crack itself) and to analyze the overall displacement in x-y direction between them. The difference between right and left rectangle average displacement is carried out and then plotted in time.

First of all, the user has the possibility to select two different rectangles in whichever position of the panel and with arbitrary size chosen by him (to make the analysis more accurate averaging in the choice of the rectangles is performed).

As it can be seen in Figure 4.18, even if the user chooses two rectangles that seem more or less identical and symmetric, the difference in terms of spatial derivative inside them (Figure 4.19) could strongly affect the results.

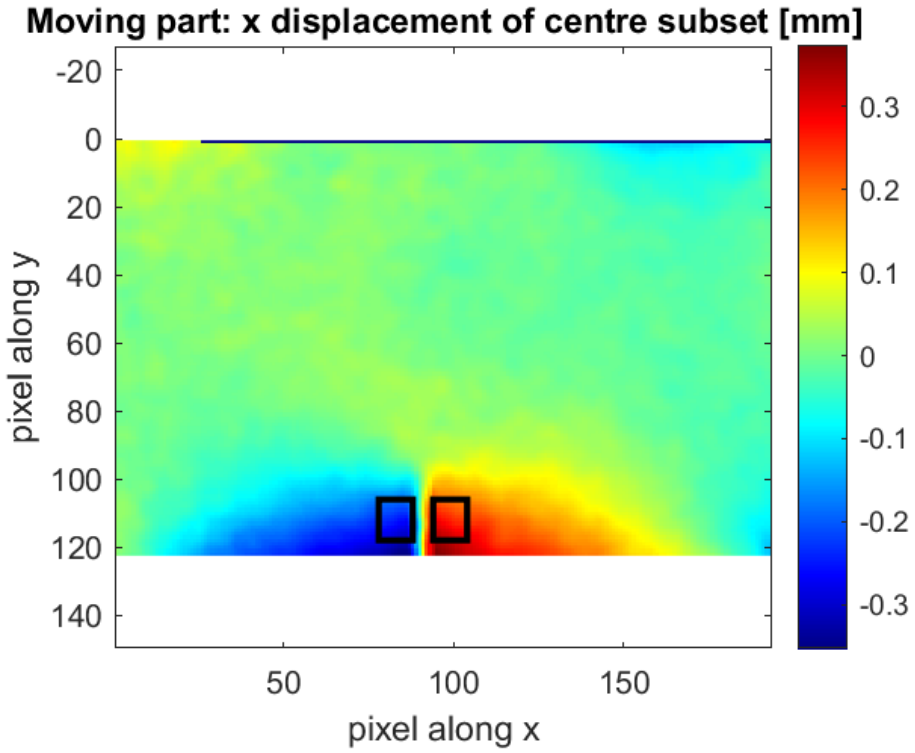


Figure 4.18: picture of x displacement with the two rectangles chosen by the user highlighted

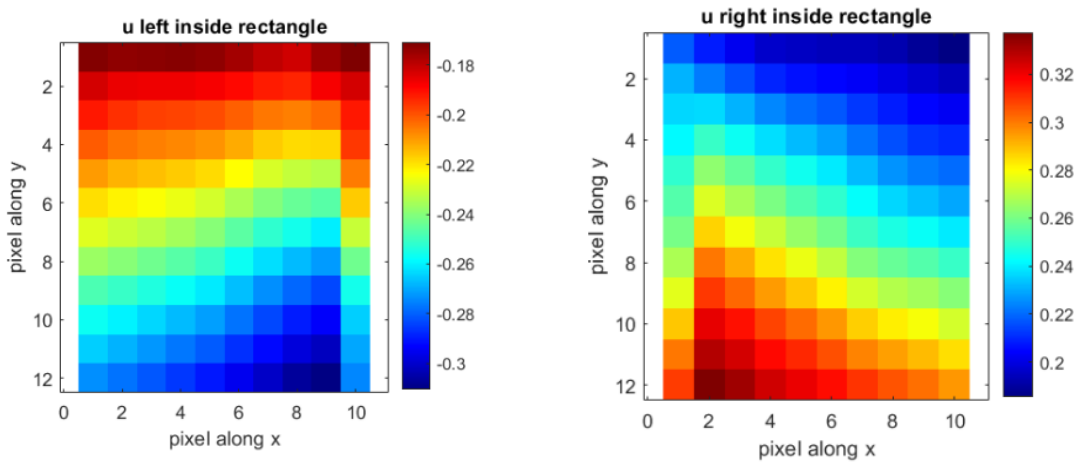


Figure 4.19: zoom on previously highlighted rectangles of Figure 4.18Figure 4.18

The problem of leaving total autonomy to the user in the decision is that, since the displacement in the panel is not negligible, even small misalignements in the regions selection could dramatically affect the results.

That's why another possibility is implemented to evaluate better the time history.

The idea is to make the user select one point on the crack and the dimensions (in pixels) of the rectangles that he wants to analyze.

The code automatically define two rectangles symmetric w.r.t. the crack and of the same dimensions thus improving the quality of the results.

The obtained results (after performing the average for all the centers subset of the two rectangles) are presented in Figure 4.20 (remembering that absolute displacement is referred to the difference between right and left rectangle).

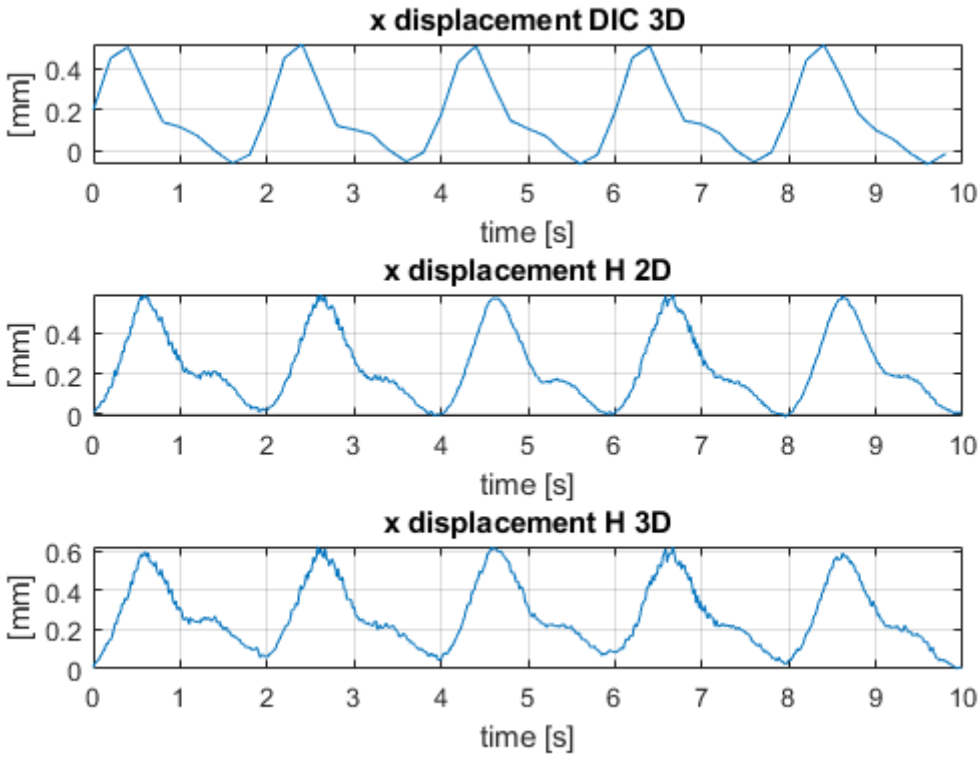


Figure 4.20: time histories comparison between the results obtained from stereo-system, homography from images and homography with depth information.

As it is clearly visible the three time histories are very similar, even if they start from slightly different points (that's due to a non optimal synchronization) and the

results obtained with the stereo-system seem less noisy (because of the lower sampling frequency and at the same time because of the higher accuracy).

Further analyses are then carried out in order to guarantee a more robust validation of the results.

From a simple Fourier analysis, the following results are obtained (Figure 4.21 and Figure 4.22).

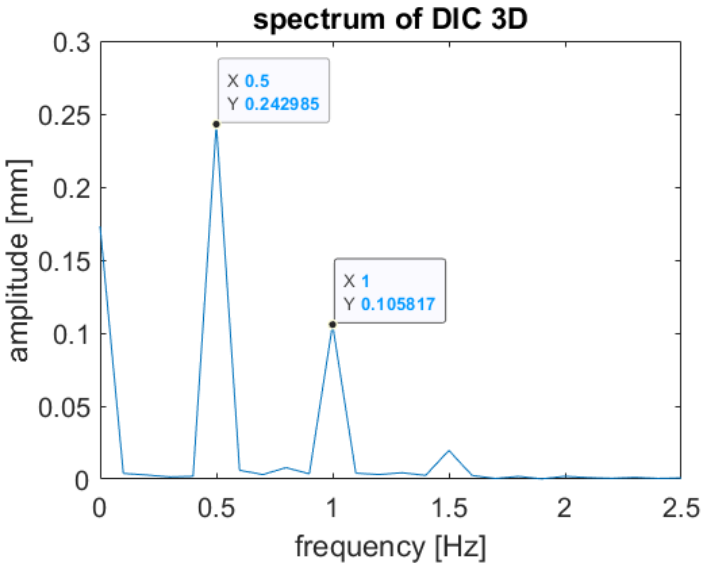


Figure 4.21: Spectrum obtained from DIC 3D (stereo-system)

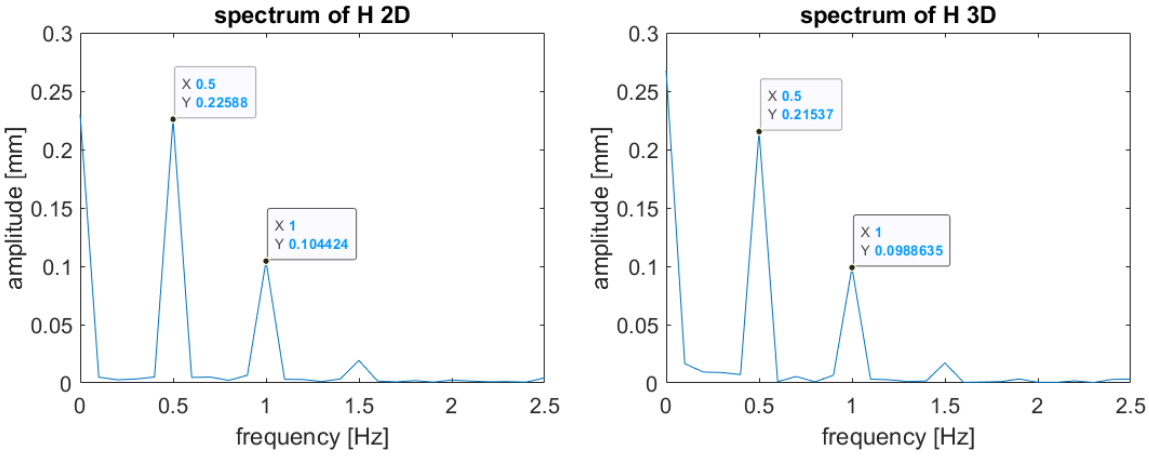


Figure 4.22: Spectrum obtained with homography from images ('H 2D') on the left and from depth ('H 3D') on the right

It's possible to notice that the frequency retrieved is exactly the same for all the three cases, while some minor differences in the amplitude are present (probably because of the non equal starting point as already pointed out due to synchronization and because of the presence of noise).

To evaluate more precisely the differences between results of the DIC 3D and of the two proposed approaches, cross correlation is performed. To correctly calculate it, the 'H 3D' and 'H 2D' time histories must be downsampled to the acquisition rate of the DIC 3D cameras. To reach a desired resolution for the time shifts of the cross correlation, interpolating functions for the 60 Hz time histories are used. To resume the procedure, time history from DIC 3D is kept fixed while the time history of either 'H 3D' or 'H 2D' (cross correlation is performed with DIC 3D vs 'H 3D' time histories and then DIC 3D vs 'H 2D' ones) is shifted of an imposed time shift, interpolated to obtain a fitting function and sampled at 5 Hz (the sampling frequency of stereo cameras). Cross correlation is calculated at every loop for the increasingly time shift. The graphs of the cross correlation are presented in Figure 4.23. After a second order fitting (Figure 4.24) the peak is retrieved, thus allowing the realignment of the time histories of the proposed approaches with the time history of DIC 3D, as visible in Figure 4.25 and Figure 4.26.

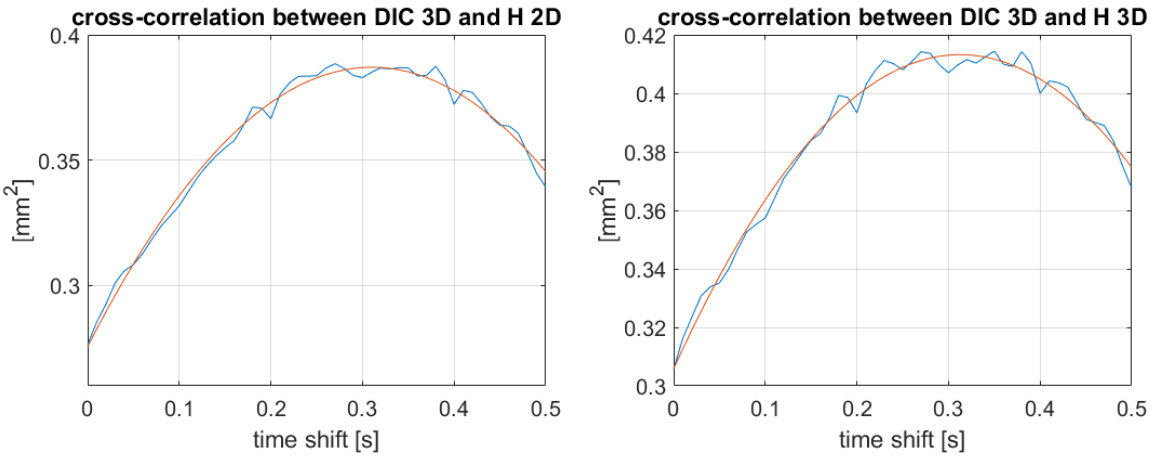


Figure 4.23: cross-correlation between DIC 3D and 'H 2D' on the left and cross correlation between DIC 3D and 'H 3D' on the right. In red the second order fitting function to retrieve the peak

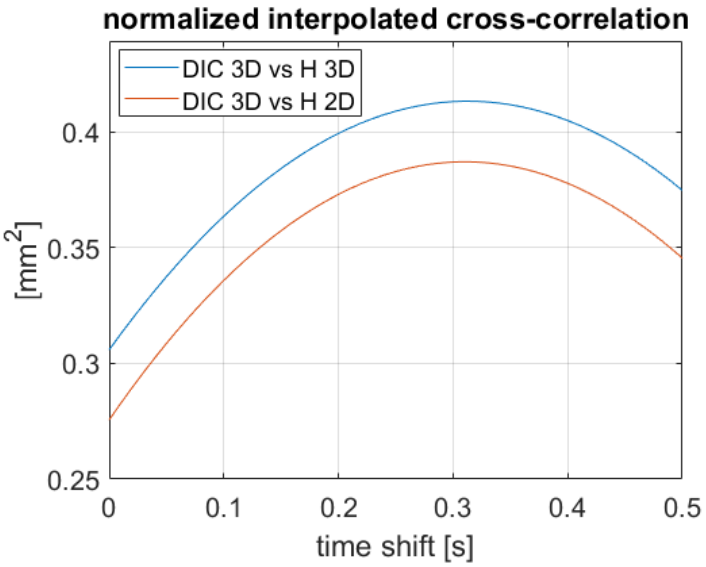


Figure 4.24: second order interpolation of cross-correlation function

The cross correlation peak is in correspondence of the same value of time shift (around 0.3 seconds) both for 'H 2D' and 'H 3D'. The time resolution chosen for the cross correlation is 0.01 s.

It is now possible to realign the time histories imposing the time shift calculated in correspondence of the cross correlation peak.

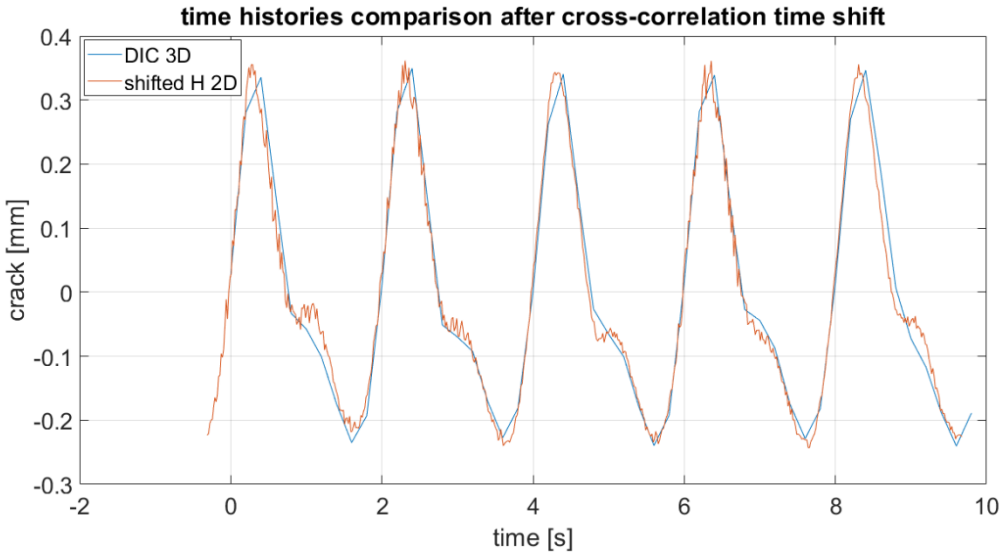


Figure 4.25: time histories comparison (H 2D vs DIC 3D) after cross correlation time shift

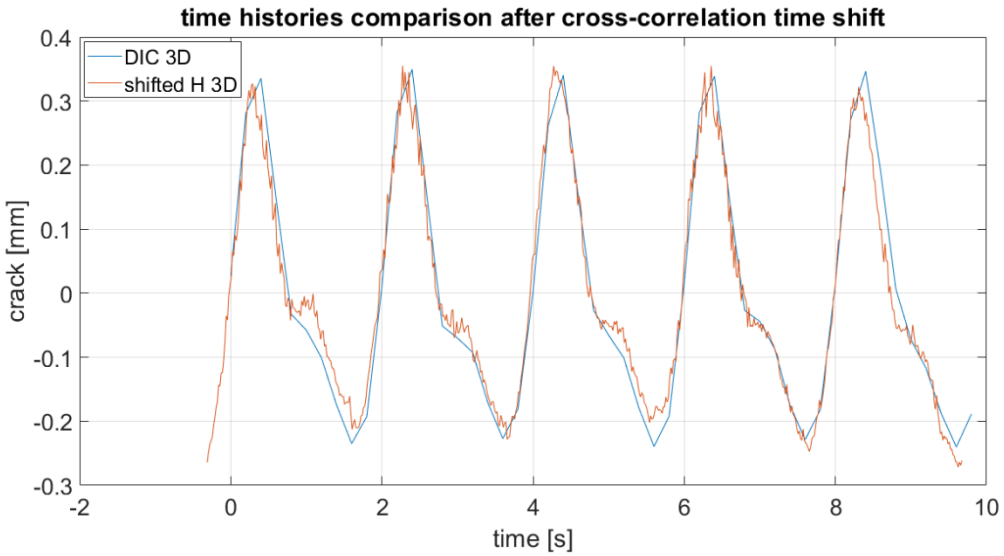


Figure 4.26: time histories comparison (H 3D vs DIC 3D) after cross correlation time shift

Here below in Figure 4.27 the overall comparison is shown.

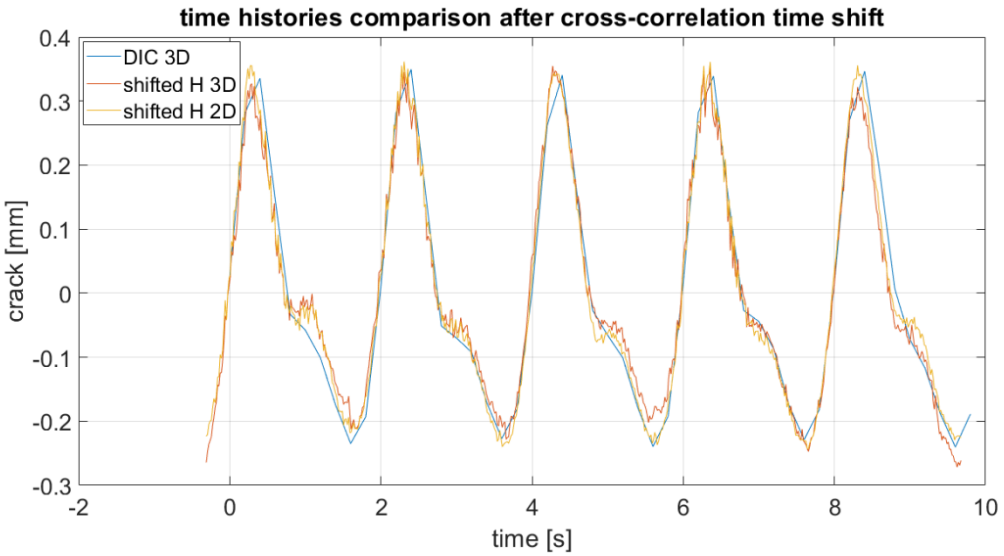


Figure 4.27: time histories comparison after the realignment of the 2 approaches

Visually the superimposition of the three time histories is great. Only some limited zones, where more noise is present for 'H 2D' and 'H 3D', seem far from the ground truth 3D DIC data. From this qualitative point of view, the risk of using 'H 3D' with the poor accuracy of the ToF sensor does not stand. This is for sure the result of

using rototranslation matrices not directly on the point clouds ('PC 3D' approach), but to calculate the homographies, that mediates the noisy effects of depth.

With the resampled shifted signal it is possible to calculate the Root Mean Square error between the points of the DIC 3D time history and the points of the two approaches proposed.

Two different interpolation functions are considered, in particular 'csaps' and 'interpft', but the results in terms of RMS error (Root Mean Square error) are the same. Results are presented in Table 4.6.

	H 3D	H 2D
RMS error [mm]	0.038	0.032

Table 4.6: RMS error comparison between the two approach

'H 2D' seems to perform slightly better than 'H 3D' in terms of Root Mean Square value. This is not so significant to define the best technique since the calculation is valid only locally, in correspondence of the crack. The loss of information in the upper part on the right and on the left due to deformation effects confounded as projective transformation is not considered within this index. Moreover, the values are very small and similar, comparable with the uncertainty of the measurement approach.

Concluding that, in the experimental scenario, with the addition of non perfectly planar object and of the deformations of the panel, the 'H 3D' approach results more robust compared to the 'H 2D' one. The calculation of the homography passing through 3D point clouds data allows to avoid the 2D problem of confounding deformations and shape effects as projective effects by the 2D minimization algorithm. At the same time, the actual degree of uncertainty associated to ToF sensors does not permit to work directly with rototranslation applied to point clouds of centre subset points ('PC 3D'). The use of homography with the plane assumption minimizes the effects of ToF noise, providing accurate results.

5 Conclusion and future developments

RGB-D digital image correlation with sensors mounted on a drone is an innovative technique that allows crack assessments on critical places (for example bridge decks), not easily reachable by men. The combination of a color or greyscale camera with a depth sensor is suitable for drone transportation compared to a 3D DIC system, as well as much less expensive. The implemented algorithm effectively compensates the drone movements, providing accurate information.

The entire technique has been tested and validated with a simulator on ideal reference and cracked images with both different orientations and crack sizes, coupled with ideal depth maps. In this simulated environment the accuracy reaches 2–3 hundredths of pixels, comparable with the intrinsic DIC uncertainty.

Moving on to a laboratory experiment, the real capabilities of the technique have been tested. A pre cracked panel with a painted speckle has been loaded by a three point bending system with an imposed cycle. The considered sensor, composed by a Flir camera and a Blaze sensor, was tested against the 3D stereo DIC system considered as the ground truth. The crack behaviour in time was evaluated. The movement compensation was tested both with homographies calculated with a minimization procedure on the 2D centre subset points and with homographies derived from 3D point clouds. In terms of local crack reconstruction both the approaches gave satisfying results with a Root Mean Square error contained under 0.04 millimeters, comparing the time histories with the one of the DIC 3D on 5 entire loading cycles. Instead, considering the entire deformation field of the panel the approach based on the realignment with homographies, calculated starting from the rototranslation of 3D centre subset point clouds was better. The calculation of the homographies on the 2D data brought a problem to the surface. In a real application the effects of deformations and not perfectly planar objects can be confounded as projective effects by the minimization algorithm, thus causing their undesired compensation. Working with a depth sensor and with point clouds allow to avoid this mistake.

As for future developments, the algorithm must be tested in a laboratory experiment with a larger data set to increase the confidence of the technique. Then, it must be evaluated on field with the RGB-D sensor mounted on the drone to perform crack assessments. With developments on the accuracy of depth sensors, the technique could be used directly with point clouds, without passing through the homography to mediate the effects of noise, for the camera movement compensation. This would also allow to work on objects with not strictly planar shape, as well as to provide displacement results for all the 3 dimensions (x-y-z), as seen with the simulator.

Bibliography

[1] D. Reagan, A. Sabato, and C. Niezrecki, "Feasibility of using digital image correlation for unmanned aerial vehicle structural health monitoring of bridges", *Struct Health Monit*, vol. 17, no. 5, pp. 1056–1072, Sep. 2018, doi: 10.1177/1475921717735326.

[2] M. Kalaitzakis, N. Vitzilaios, D. C. Rizos, and M. A. Sutton, "Drone-Based StereoDIC: System Development, Experimental Validation and Infrastructure Application", *Exp Mech*, vol. 61, no. 6, pp. 981–996, Jul. 2021, doi: 10.1007/s11340-021-00710-z.

[3] B. Pan, "Digital image correlation for surface deformation measurement: Historical developments, recent advances and future goals", *Measurement Science and Technology*, vol. 29, no. 8. Institute of Physics Publishing, Jun. 28, 2018. doi: 10.1088/1361-6501/aac55b.

[4] B. Pan, K. Qian, H. Xie, and A. Asundi, "Two-dimensional digital image correlation for in-plane displacement and strain measurement: A review", *Meas Sci Technol*, vol. 20, no. 6, 2009, doi: 10.1088/0957-0233/20/6/062001.

[5] H. Bruck, S. McNeill, M. Sutton, and W. Peters III, "Digital Image Correlation Using Newton-Raphson Method of Partial Differential Correction", *Exp Mech*, vol. 29, pp. 261–267, 1989, doi: 10.1007/BF02321405.

[6] Y. L. Dong and B. Pan, "A Review of Speckle Pattern Fabrication and Assessment for Digital Image Correlation", *Exp Mech*, vol. 57, no. 8, pp. 1161–1181, Oct. 2017, doi: 10.1007/s11340-017-0283-1.

[7] S. Vanlanduit, J. Vanherzeele, R. Longo, and P. Guillaume, "A digital image correlation method for fatigue test experiments", *Opt Lasers Eng*, vol. 47, no. 3–4, pp. 371–378, Mar. 2009, doi: 10.1016/j.optlaseng.2008.03.016.

[8] Z. Sun, J. S. Lyons, and S. R. Mcneill, "Measuring Microscopic Deformations with Digital Image Correlation", *Opt Lasers Eng*, vol. 27, no. 4, pp. 409–428, 1997, doi: 10.1016/S0143-8166(96)00041-3.

[9] H. W. Schreier, D. Garcia, and M. A. Sutton, "Advances in light microscope stereo vision", *Exp Mech*, vol. 44, no. 3, pp. 278–288, Jun. 2004, doi: 10.1177/0014485104041546.

[10] T. A. Berfield, J. K. Patel, R. G. Shimmin, P. v. Braun, J. Lambros, and N. R. Sottos, "Micro-and nanoscale deformation measurement of surface and internal planes via digital image correlation", *Exp Mech*, vol. 47, no. 1, pp. 51–62, Feb. 2007, doi: 10.1007/s11340-006-0531-2.

[11] W. G. Knauss, I. Chasiotis, and Y. Huang, "Mechanical measurements at the micron and nanometer scales", [Online]. Available: www.elsevier.com/locate/mechmat, 2003.

[12] M. Palanca, G. Tozzi, and L. Cristofolini, "The use of digital image correlation in the biomechanical area: A review", *International Biomechanics*, vol. 3, no. 1. Taylor and Francis Ltd., pp. 1–21, 2016. doi: 10.1080/23335432.2015.1117395.

[13] Y. Wang and A. M. Cuiti, "Full-field measurements of heterogeneous deformation patterns on polymeric foams using digital image correlation", [Online]. Available: www.elsevier.com/locate/ijsolstr, 2002.

[14] J.-N. Périe, S. Calloch, C. Cluzel, and F. Hild, "Analysis of a Multiaxial Test on a C/C Composite by Using Digital Image Correlation and a Damage Model", *Exp Mech*, vol. 42, pp. 318–328, 2002, doi: 10.1007/BF02410989.

[15] D. Zhang and D. D. Arola, "Applications of digital image correlation to biological tissues", *J Biomed Opt*, vol. 9, no. 4, p. 691, 2004, doi: 10.1117/1.1753270.

[16] B. Lei, N. Wang, P. Xu, and G. Song, "New Crack Detection Method for Bridge Inspection Using UAV Incorporating Image Processing", *J Aerosp Eng*, vol. 31, no. 5, Sep. 2018, doi: 10.1061/(asce)as.1943-5525.0000879.

[17] H. Yu, W. Yang, H. Zhang, and W. He, "A UAV-BASED CRACK INSPECTION SYSTEM FOR CONCRETE BRIDGE MONITORING", *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2017, doi: 10.1109/IGARSS.2017.8127704.

[18] C. M. Yeum and S. J. Dyke, "Vision-Based Automated Crack Detection for Bridge Inspection", *Computer-Aided Civil and Infrastructure Engineering*, vol. 30, no. 10, pp. 759–770, Oct. 2015, doi: 10.1111/mice.12141.

[19] Y. F. Liu, X. Nie, J. S. Fan, and X. G. Liu, "Image-based crack assessment of bridge piers using unmanned aerial vehicles and three-dimensional scene

reconstruction", *Computer-Aided Civil and Infrastructure Engineering*, vol. 35, no. 5, pp. 511–529, May 2020, doi: 10.1111/mice.12501.

[20] A. Ellenberg, A. Kontsos, F. Moon, and I. Bartoli, "Bridge related damage quantification using unmanned aerial vehicle imagery", *Struct Control Health Monit*, vol. 23, no. 9, pp. 1168–1179, Sep. 2016, doi: 10.1002/stc.1831.

[21] N. Metni and T. Hamel, "A UAV for bridge inspection: Visual servoing control law with orientation limits", *Autom Constr*, vol. 17, no. 1, pp. 3–10, Nov. 2007, doi: 10.1016/j.autcon.2006.12.010.

[22] H. Xu, X. Su, Y. Wang, H. Cai, K. Cui, and X. Chen, "Automatic bridge crack detection using a convolutional neural network", *Applied Sciences (Switzerland)*, vol. 9, no. 14, Jul. 2019, doi: 10.3390/app9142867.

[23] R. S. Adhikari, O. Moselhi, and A. Bagchi, "Image-based retrieval of concrete crack properties for bridge inspection", *Autom Constr*, vol. 39, pp. 180–194, Apr. 2014, doi: 10.1016/j.autcon.2013.06.011.

[24] I. H. Kim, H. Jeon, S. C. Baek, W. H. Hong, and H. J. Jung, "Application of crack identification techniques for an aging concrete bridge inspection using an unmanned aerial vehicle", *Sensors (Switzerland)*, vol. 18, no. 6, Jun. 2018, doi: 10.3390/s18061881.

[25] I. Abdel-Qader, S. Pashaie-Rad, O. Abudayyeh, and S. Yehia, "PCA-Based algorithm for unsupervised bridge crack detection", *Advances in Engineering Software*, vol. 37, no. 12, pp. 771–778, 2006, doi: 10.1016/j.advengsoft.2006.06.002.

[26] S. 'Rathinam and R. 'Sengupta, "Vision-based monitoring of locally linear structures using an unmanned aerial vehicle", *Infrastruct Syst*, 2008, doi: 10.1061/(ASCE)1076-0342(2008)14:1(52).

[27] K. Kumarapu, S. Mesapam, V. R. Keesara, A. K. Shukla, N. V. S. K. Manapragada, and B. Javed, "RCC Structural Deformation and Damage Quantification Using Unmanned Aerial Vehicle Image Correlation Technique", *Applied Sciences*, vol. 12, no. 13, p. 6574, Jun. 2022, doi: 10.3390/app12136574.

[28] A. Khadka, A. Afshar, M. Zadeh, and J. Baqersad, "Strain monitoring of wind turbines using a semi-autonomous drone", *Wind Engineering*, vol. 46, no. 1, pp. 296–307, Feb. 2022, doi: 10.1177/0309524X211027814.

[29] D. Marchisotti and E. Zappa, "Uncertainty mitigation in drone-based 3D scanning of defects in concrete structures", 2022 *IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, 2022, doi: 10.1109/I2MTC48687.2022.9806652.

[30] J. M. Franco, J. M. Caicedo, J. Marulanda, M. Sutton, and P. Thomson, "RGB-D-DIC technique for low-cost 3D displacement fields measurements", *Eng Struct*, vol. 197, Oct. 2019, doi: 10.1016/j.engstruct.2019.109457.

[31] G. Chen, Q. Liang, W. Zhong, X. Gao, and F. Cui, "Homography-based measurement of bridge vibration using UAV and DIC method", *Measurement (Lond)*, vol. 170, Jan. 2021, doi: 10.1016/j.measurement.2020.108683.

[32] website on "digitalimagecorrelation.org", visited on Mar. 2022.

[33] J. D. Foley, M. A. Fischler, and R. C. Bolles, "Graphics and Image Processing Random Sample Consensus: A Paradigm for Model Fitting with Apphcatlons to Image Analysis and Automated Cartography", *Communications of the ACM*, vol. 24, no.6, 1981.

[34] R. Schnabel, R. Wahl, and R. Klein, "Efficient RANSAC for Point-Cloud Shape Detection", *Computer Graphics forum*, vol. 26, no. 2, 2007.

[35] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*, 2004.

[36] P. H. S. Torr and A. Zisserman, "MLESAC: A new robust estimator with application to estimating image geometry", *Computer Vision and Image Understanding*, vol. 78, no. 1, pp. 138–156, 2000, doi: 10.1006/cviu.1999.0832.

[37] A. Lavatelli, "Uncertainty analysis and improvement of vision based measurement techniques in vibration testing", [Online]. Available: https://re.public.polimi.it/retrieve/handle/11311/1001560/624768/11311-1001560%20Zappa.pdf , 2019.

[38] J. Chang, "TIME-OF-FLIGHT FORGES AHEAD", *Quality, suppl. VISION & SENSORS*, [Online]. Available: https://www.qualitymag.com/articles/95956-time-of-flight-forges-ahead, Mar. 2020.