Metrological characterization of a low-cost portable stereo vision system for dynamic measurements

Fabio Brian Valli, Alberto Lavatelli, Emanuele Zappa
Department of Mechanical Engineering, Politecnico di Milano, Via La Masa, 1 – Milano - Italy

Email: alberto.lavatelli@polimi.it

Abstract. In recent years, camera and computer systems have been miniaturized considerably. The market now offers several low-cost products to assemble a small scale portable stereo vision system which can be completely autonomous (since it can acquire and process image). However, the metrological characteristics of such a portable stereo vision device should be assessed and compared with the state of the art of stereo vision system. In this work a portable stereo vision system composed by small board cameras and a single board computer is metrologically characterized and validated with respect to fixed non-portable systems. The analysis focused on the application of such a stereo rig to DIC and stereoscopic measurements performed with commercial software. For these tests two algorithms with different correlation techniques are evaluated in Matlab and in Halcon, moreover, a 3D DIC analysis is conducted on a H-shape specimen. To perform a metrological characterization, the points clouds obtained from the triangulations are confined to a region of the scene containing only a part of the specimen, after that they are fitted with a LOWESS fitting model. The uncertainty is assessed by computing the discrepancy between the points of the cloud and the ones of the fitting model. In some algorithms more than others the errors distribute in a strongly periodic manner with a period linked to integer values of disparity. Eventually the experimental analysis demonstrated that the accuracy of such a lightweight stereo system is similar to the accuracy of traditional not-portable ones.

1. Introduction

Contactless measurement techniques obtained from vision system have become a branch of relevant in the world of measures and the number of industrial applications that are adopting this practise is growing continuously.

Of particular interest in the recent years is the three-dimensional full field measurement performed with stereo system: by acquiring a couple of images from two cameras at the same time it is possible to obtain the three-dimensional coordinates of each point of the scene, (with just one camera it can’t be possible to obtain the depth datum).

Due to the rapid development of technology in the recent years, camera and computer systems have been miniaturized considerably and they have reduced their costs by much. Starting from these considerations it is possible to think to assemble a small scale portable stereo vision system which can be completely implemented to perform images acquisitions independently.
Since these products have the compactness as a main feature and not the accuracy, it is necessary to metrologically characterize and validate them with respect to fixed well-known non-portable systems. The basic idea of this work is to design a portable stereo vision system composed by small board cameras and a single board computer. Besides the design of the portable stereo rig, a depth work has been provided on the image processing part. The last aim of this thesis is to characterize the portable stereo system in terms of accuracy and reliability, to do this a state of art of the best stereo vision fixed system is achieved and used as a touchstone for the compact stereo system analysis.

The intention is to provide a robust stereo vision method based on different algorithms able to compete with the Digital Image Correlation (DIC) technique, considered today as the guideline technique in stereovision analysis. On the market several software able to perform a 3D scene reconstruction are available. In this work the analyses are conducted with Matlab, one of the most widespread engineering software, and Halcon, a tool completely focused on artificial vision.

2. Development of a portable stereo vision system

In the design of the portable stereo rig the weight and the dimensions are considered main characteristics in the selection of the components. Unfortunately, as it is in all technological fields, a reduction in size of any device results in an increase of the cost and in critical issue under the point of view of reliability and processing capacity. In the selection of the components it was tried to find a trade-off between the costs and the attributes of the devices. Thanks to the progressive technology shrink of the latest years, it is possible to find in the market cameras and computer of incredible small size able to guarantee reasonable performances.

2.1. Selection of camera system and lens setup

The cameras selected for the proposed stereo system are a pair of USB 3.0 monochrome CMOS sensor board cameras from The Imaging Source. The size of the whole sensors selected is extremely contained as it can be seen in Figure 1 (30 x 30 x 10 mm), the same for the weight that is about 15 g.

![Figure 1 Board camera with lens](image)

The monochrome choice is justified by a higher efficiency in the image processing steps: due to the lightweight data in respect to the colour images. For the same reason the resolution of about 1280x960 (1.2MP) is a good compromise between allowing a reasonable quality of the images and consenting a quite fast processing time during the analysis. The 3.0 USB choice is of particular interest for the high-speed data transfer rate (theoretical 5 Gb/s), and for the suitable possibility to perform trigger operations directly through it and without use additional cables allowing to save more weight in the resulting payload.

During the selection of the model, also the current consumption has to be taken into account, considering that the cameras have to be managed and supplied by a single board computer the current absorption has to be as limited as possible. For this model, the consumption is declared at 250mA @ 5 VDC that is an admissible value for our board.
The last parameter that has to be considered is the exposure of each pixel to the light that, for this application, has to happen at the same time and so a global shutter mode is necessary required at the expense of a rolling one.

The choice of the optics has been made by taking into account the working space and the cameras and lenses parameters like focal length, baseline between the cameras, rake angles etc. In this manner the lens’s focal length that enables to operate in a determinate space at a certain working distance is detected. The selection of the optics is made by considering a working distance of about 1.5 m, a baseline between the cameras of 200 mm and rank angles set to 10°. A simple 2-dimensional stereo system perspective is shown in Figure 2: in light-blue the field of view where it is possible to acquire stereo images.

![Figure 2 Stereo system top view scheme](image)

For the kind of application under investigation, a 6 mm focal length fish-eye lens is selected as the best candidate. With this configuration the point \( S_o \) has vertical coordinate equal to 161 mm that represents the closest points of the scene that can be subjected to triangulation, \( P_1 \) has coordinate (202, 486) mm, and symmetrically \( P_2 \) (-202, -486) mm. These lenses are characterized by an extremely low weight (6 g) and the whole assembly results in an overall weight of 21 g per camera.

### 2.2. Computer system and full setup configuration

A single-board lightweight computer is selected to manage the acquisitions of the two cameras. The idea to sort this kind of device is justified by several issues: the compact size and the limited weight make it suitable for our purpose, the low power consumption offers the possibility to supply it power with a low weight lithium battery and to provide energy to both the cameras.

Among several boards, the model chosen is LattePanda that shows several advantages in respect to the other competitors (i.e. Raspberry pi), in example it is able to run the full version of Windows 10 Home Edition, 3.0 USB ports, more efficient processor etc. Thanks to this, it is possible to develop a proper software for images acquisition, storage, triggering etc. on a common computer with Microsoft Windows installed as operating system and to run it on the LattePanda.

Moreover, through the Wi-Fi connection, one can provide on-shore commands for in-flight UAV acquisition so as to acquire images in the intended instants.

The single-board features are widely satisfactory for this kind of application:
- Intel Quad Core 1.8GHz processor, 500MHz GPU
- 2GB RAM
- 32GB onboard flash memory and 64GB MicroSD card memory
- USB 3.0 port suitable for the selected USB 3.0 cameras
• Wi-Fi and Bluetooth 4.0
• HDMI video output
• Power: 5V @ 2A

The dimensions are awfully small: 88 x 70 mm, and the weight is about 100 g.

Figure 3 LattePanda single board computer

A proper software is developed in LabVIEW environment with the aim of being installed on LattePanda to perform images acquisitions. With this software it is possible to select the image format and the number of frame per second directly from the LattePanda, moreover, for dynamic acquisitions, the number of shots to be acquired can be set to an arbitrary value. Even if this software has a basic working principle, it has to be considered as an early program that can be developed in future with some other features (i.e. Wi-Fi implementation for remote control).

The pair of cameras are managed by NI_I MAQdx and they are linked to the single board computer by means of a 3.0 USB bus. Before starting the acquisitions, the specific driver released by the Imaging Source is installed on LattePanda for both the cameras.

Figure 4 Final stereo system front view

3. Performance analysis and uncertainty characterization of stereoscopy algorithms for the development of a portable stereo vision system

To characterize the brand new portable stereo vision system, it is necessary to provide a state-of-art of the major stereoscopy algorithms proposed by two software available on the market: Matlab and Halcon. Matlab is a widespread software used in several field of engineering. Even though it is not properly designated to operate in the computer vision area, it proposes an effective toolbox that offers the possibility to perform triangulation between a pair of stereo images by exploiting the function named “disparity”.
Halcon is a comprehensive standard software for machine vision with an integrated developed environment that is used worldwide. In particular the software is specifically developed to work in the field of computer vision and it proposes different operators to compute the disparity between a pair of stereo images.

### 3.1. Uncertainty analysis methodology

To perform a reasonable uncertainty and performance comparison, the usage of the same couple of images is crucial during image processing analysis. This because any change in brightness, focus or bias can change drastically the results obtained during the investigations making the outcomes comparison not reliable. The specimen selected is a H-shaped thin sheet of steel characterized by a uniform black and white texture.

The first step is to obtain the 3D coordinates of every point of the scene by means of the triangulation procedure. Once this result is achieved for each software, the data are exported to a Matlab script developed to assess the goodness of the triangulation in order to have as a guideline the same evaluation criterion.

Starting from the set of coordinates of each point of the scene, the points cloud (that is a three-dimensional representation of each point in the space) is computed and, to evaluate the effectiveness of the measurements of the proposed model, it is fitted with a locally weighted linear regression scatterplot smoothing surface model: LOWESS.

The value of this algorithm is that it does not force every single point to belong to the surface, but it uses a local linear regression fit type that automatically approximates all the points in the most convenient way. The analysis is conducted to the only points belonging to a single arm of the specimen, in such a way none non-interest point will affect the results.

As an index of the accuracy of this method, root mean squared error (standard error or RMSE) is evaluated studying the discrepancy between the model and the points of the cloud. A sort of systematic error in the display of the points cloud has been detected with different intensity in the methods examined. After having studied the origin of this effect, it was concluded that the source comes from the evaluation of the disparity map, in particular, an evident dependency between the periodicity of the errors and the integer values of the computed disparity has been found.

To study the link between disparity values and the trend of the discrepancy errors, graphs are drawn which relate the trend of errors according to the value of disparity and a tendency line is tracked to study the trend.

To quantify the weight of this cyclical component, the autospectrum of the tendency line is evaluated. In Figure 5 a typical outcome is shown: starting from the signal, the peak (that it's present at disparity value of 1 pixel), is divided by the sum of all the other spectral components, this allows to quantify how much the component at 1 pixel weighing in respect to the rest of the signal’s value. The highest this value, the highest the periodic trend for integer values of disparity. In the further discussions this value will be called “periodicity index”.

![Figure 5 Periodicity index evaluation](image)
The stereoscopic techniques seem to work well for integer disparity values, while between them, the algorithms try to find an interpolation in the correlation procedure [1]. The origin of this outcome, that is more evident in some algorithms rather than others, has to be referred to as pixel-locking effect.

Finally, the last method selected to evaluate the goodness of each test is the Anderson-Darling goodness-of-fit hypothesis test. With this statistical tool it is possible to determine if the points obtained by the fitting procedure are coming from a normal distribution or not.

3.2. Algorithm comparison

A pair of algorithms are put under investigation for both Matlab and Halcon software. As regards to Matlab, its library offers a double choose matching algorithm to compute the disparity map: local block matching and semi-global block matching.

In the first case the disparity function implements the Sum of absolute Differences (SAD) correlation technique that computes the absolute differences of gray levels given a certain pixel at position \((r,c)\) in the left image \((L)\) and the pixel \((r,c-d)\) in the match right image \((R)\).

\[
SAD = \sum_{r,c} |L(r,c) - R(r,c - d) |
\]

While in the semi-global matching alternative, the algorithm is not limited in finding optimal disparities for small image regions (i.e. looking for the best matching on a single row) but it keeps care about disparity values of the neighbouring blocks and it favour the ones with similar disparity. This kind of operation is based on a multi-scan line optimization: beside taking the absolute difference of grey levels, the semi-global matching algorithm proposes a method based on mutual information originally studied to incorporate illumination changes between both input images [2].

In the same way, Halcon proposes two alternatives to perform triangulation: binocular_disparity and multi-scan line binocular_disparity functions. For the first case it is possible to choose the correlation technique among three different options: Sum of absolute differences, Summed Squared Differences or Normalized Cross Correlation.

Even if the NCC is the more time consuming, it is choosen as the best candidate for the analysis due to the higher reliability with respect to the other competitors.

\[
NCC(r,c,d) = \frac{\sum_{r'=-m}^{r+m} \sum_{c'=-n}^{c+n} (g_1(r',c') - \bar{g}_1(r,c))(g_2(r',c'+d) - \bar{g}_2(r,c))}{\sqrt{\sum_{r'=-m}^{r+m} \sum_{c'=-n}^{c+n} (g_1(r',c') - \bar{g}_1(r,c))^2(g_2(r',c'+d) - \bar{g}_2(r,c+d))^2}}
\]

With:
- \(r, c\) : row and column coordinates of the corresponding pixels of the two input images.
- \(g_1, g_2\) : gray values of the unprocessed input images.
- \(N = (2m + 1)(2n + 1)\) : size of correlation window
- \(\bar{g}(r,c) = \frac{1}{N} \sum_{r'=-m}^{r+m} \sum_{c'=-n}^{c+n} g(r',c')\) is the mean value within the correlation window of width 2m+1 and height 2n+1.

The multi-scan line disparity function, differently from the first, it computes the disparity between two rectified stereo images using multi-scanline optimization [3]. This technically allow to achieve better results in the low-texture regions of the images without blurring discontinuities, furthermore the previous algorithm returns more robust results if there is sufficient texture but fails where there is none [4].
In Figure 6 the disparity values-errors entity graphs of the 4 methods are reported in order to study the effects of pixel-locking issue.

As it can be seen the multi-scan line optimization algorithms suffers strongly the pixel-locking effect. This phenomenon results in the tendency for the measured location or displacement of a particle image to be biased towards integer values [5]. It is an intrinsic error imbedded in an image acquisition in particular when a sub-pixel optimization process is present in the triangulation algorithm [6].

In Table 1 a brief summary of the outcomes from each analysis are reported. It is possible to see that the best result in terms of fitting RMSE is achieved by the Halcon binocular disparity function with the matching correlation algorithm set on Normalized Cross Correlation.

With respect to Matlab the results are mildly worsened, in the case of the Local block matching the result obtained derives from a complex combination of the right parameters in the disparity function. The procedure to get there it is not as immediate as it was for the Halcon software, while for the Semi-Global matching algorithm there is no way to reduce the uncertainties obtained.

The ad-test gives positive results in the case of Matlab Local block matching and Halcon NCC, while it shows a negative feedback for both the multi-scan line algorithms.

With respect to the time complexity Matlab is the fastest software in computing the disparity, while Halcon is much slower, in some cases more than 10 times. As it obvious, to achieve greater results the complexity of the algorithm and the density of data increases and consequently also the computational time.

| Method                        | RMSE [mm] | Anderson-Darling test | Time complexity [s] | Periodicity index |
|-------------------------------|-----------|-----------------------|---------------------|------------------|
| Matlab, Semi-Global           | 0.529     | 1                     | 3.78                | 7.44             |
| Matlab, Local                 | 0.224     | 0                     | 0.68                | 0.17             |
| Halcon, binoc_disp, NCC       | 0.112     | 0                     | 33.46               | 0.19             |
| Halcon, binoc_disp_ms         | 0.511     | 1                     | 57.18               | 1.24             |

Table 1 Algorithms comparison
As already said, the multi-scan line algorithm suffers the pixel-locking effect, especially in the case of Matlab semi-global block matching the periodicity index value is resulted to be 7.4. In the figure the plot of the discrepancy errors on the fitting surface is proposed to appreciate the high periodicity in the trend.

![Figure 7 Error values distribution shown on the fitting surface](image)

As method of comparison, the same pair of images are processed with MatchID Digital Image Correlation and the resulting points cloud is fitted in the same way. The RMSE in this case gives 0.059 mm as value. Of course, a better result was expected with this technique that is considered as the rule model in computer vision for contactless measures. By the way, the outcomes achieved from Halcon NCC correlation technique are slightly worse but still comparable.

4. Case study: vibration measurement

To test the portable stereo vision system in complex situation under the computational point of view, a case of vibration motion of the same H-shaped specimen is put under analysis. Under the assumption of studying 6 periods of fluctuations, a set of 50 pair of stereo images is acquired at 96 fps with exposure time set to 0.0058 s for both the cameras during the perturbation.

The images are then processed in Matlab with the following procedure. Starting from the points cloud, a volume of space containing a portion of the specimen at the extremity of its arm is identified in each of the 50 clouds. As shown in Figure 8, the points of the clouds included in this volume (in light blue) represents a set of pixels of the vibrating arm from which it is possible to calculate the centre of gravity coordinates.

![Figure 8 Filtered points cloud in Matlab](image)

This set of coordinates represents the distance from the left camera optical centre to the selected region of the specimen frame by frame, and from that it is possible to extract the frame-distance graph as shown in figure.
As expected, the coordinates of the COG describe a fluctuating signal that represents the motion of vibration of the specimen’s arm.

As a comparison index, a vibration analysis of the same set of frames is conducted by performing 3D DIC with MatchID: a rectangular region of 39 squared blocks of 24 pixels in the extremity of the lower arm of the specimen is selected with similar size and in the same position with respect to the one from Matlab. The program automatically returns the out-of-plane displacement “W” of each block of the rectangular region with respect to the reference image automatically.

To perform a reliable comparison among the two methods, the data need to be reorganized as regards to conventions, and by estimate the vibration amplitude with respect to the static reference image. After that, the 2 outcomes from these different techniques are superimposed as shown in Figure 10.

In this comparison it is evident the excellent correspondence between the MatchID data (in blue dotted line) and the ones from Matlab (in light green line): the curves seem to be perfectly overlapped, the points in blue representing the MatchID data are plotted with a reduced size in order to appreciate the small discrepancy with respect to the green ones from Matlab.

---

**Figure 9** Optical centre-filtered cloud distance in each frame

**Figure 10** Vibration motion comparison
The RMSE, that represents the value of the mean discrepancy between the data extracted from the 3D Direct Image Correlation in MatchID and the ones from the stereo matching in Matlab is then evaluated:

$$RMSE = \sqrt[2]{\frac{\sum_{i=1}^{n}(x_i - \hat{x}_i)^2}{n}} = 0.070 \text{ mm}$$

By considering MatchID as the key model for these analysis, the result of 7 hundredths of millimeter of mean discrepancy between this and the proposed Matlab approach allows to validate this technique completely.

5. Conclusion
In this work, in order to lay the foundations for the portable stereo system, it was provided a detailed comparison in terms of uncertainties and computational time among the most widespread algorithm capable of performing a triangulation starting from a pair of stereo images.

It has been concluded that the best method to perform a 3D scene reconstruction by means of stereoscopy algorithms is by computing the binocular disparity function in Halcon with the correlation technique set on Normalized Cross Correlation. On the other side, since Matlab is a broader software that operates in a magnitude of fields including the stereo vision one, the processing procedure is easier and more immediate, but the results are characterized by a lower accuracy.

By the way, based on the kind of application to deal with and on the level of accuracy required, also the local block matching algorithm in Matlab can be considered as a reliable tool to perform a triangulation. After having set up the prototype and the acquisition software, the tool was tested in static and dynamic acquisitions. In both these applications the results to which it has come are to be considered definitely positive: the static outcomes are of the same order of magnitude of the ones achieved with the original stereo system. In the same way in the dynamic test, besides some little drawbacks related to the images acquisition in LattePanda, the final conclusion has to be considered successful.

Moreover, the novel technique to study the vibration motion by means of stereoscopic algorithms has proved to be comparable to the key model digital image correlation from MatchID for these specific tasks.

LattePanda has proved to be a device capable of performing this type of operations useful for rapid dynamic acquisitions of vibratory motions. Despite its small size, its process ability is sufficient to capture pairs of images from two different cameras simultaneously up to 120 fps. Obviously, it is not conceivable to use this prototype for long measurement operations or where an extreme accuracy is required in the results, by the way, the measurement accuracy guaranteed by this prototype is of interest for most applications.

After having established that the lightweight system is capable of performing reliable tests, the future work of interest would be to implement it on a UAV with a specific frame and to perform in-flight acquisition. The resulting weight of the final stereo system is about 300 g (excluded the lithium portable battery necessary for in-flight applications) and can be considered an acceptable payload for the majority of the work-drones.

The area of application in which this project can catch on are undoubtedly the civil works: the periodic checks on the large structures as bridges, dams, stadiums, etc. are expensive and extremely time consuming to be performed with the actual technology. The proposed method can lead to a radical reduction in time, money, risks and complexity during the operation by ensuring an analogous accuracy in the results.
References

[1] D. R. N. a. B. W. Dirk Michaelis1, “Peak-locking reduction for particle image,” *Measurement Science and Technology*, 2016.

[2] G. Egnal, “Mutual information as a stereo correspondence,” in *Technical Report MS-CIS-00-20*, Computer and Information Science, University of Pennsylvania, Philadelphia.

[3] MVTec, *HALCON 12.0*.

[4] V. A. D. D. Guide, Census Transform Algorithm Overview.

[5] J. Westerweel, “Effect of Sensor Geometry on the Performance of PIV Interrogation”.

[6] B. G. R. J. Hearst, “Quantification and adjustment of pixel-locking in particle image velocimetry,” vol. Exp Fluids, 2015.