Calibration and disparity maps for a depth camera based on a four-lens device

Cécile Riou, Bruno Colicchio, Jean-Philippe Lauffenburger, Olivier Haeberle, Christophe Cudel

Calibration and disparity maps for a depth camera based on a four-lens device

Cécile Riou,[a] Bruno Colicchio,[a] Jean Philippe Lauffenburger,[a] Olivier Haeberlé,[a] Christophe Cudel [a]
[a] Université de Haute Alsace, laboratoire MIPS (EA2332), 61 rue Albert Camus, Mulhouse, France, 68093

Abstract. We propose a model of depth camera based on a four-lens device. This device is used for validating alternate approaches for calibrating multi-view cameras and also for computing disparity or depth images. Calibration method arises from previous works, where principles of variable homography were extended for 3-D measurement. In this paper, calibration is performed between two contiguous views obtained on the same image sensor. This approach leads to propose a new approach for simplifying calibration, by using the properties of the variable homography. The second part of this paper addresses new principles for obtaining disparity images without any matching. A fast algorithm using a contour propagation algorithm is proposed without requiring structured or random pattern projection. These principles are proposed in a framework of quality control by vision, for inspection in natural illumination. By preserving scene photometry, some other standard controls, as for example calipers, shape recognition, or barcode reading, can be done conjointly with 3-D measurements. Approaches presented in this paper are evaluated. Firstly, we show that rapid calibration is relevant for devices mounted with multi lenses. Secondly, synthetic and real experimentations validate our method for computing depth images.

Keywords: Multi-view cameras, depth cameras, homography, disparity map, camera calibration

Address all correspondence to: Christophe Cudel, University of Haute Alsace, MIPS, 61 rue Albert Camus, Mulhouse, France, 68093; Tel: +33 389-337-661; Fax: +33 389-337-605; E-mail: christophe.cudel@uha.fr

1 Introduction

Multi-view imaging is a large domain, where numerous approaches can take place for computing the depths in a scene. Stereovision[1], multi-camera array[2], light-field imaging[3,4] or coded aperture imaging[5] are the most well-known technics for capturing several point of views of a scene. Using a metric calibration on these systems allows for addressing recently growing applications in a research or industrial context. In multi-view imaging, the depth estimation problem is then related to the disparity between same pixels projected on the different coordinates in the different views. This task usually relies on the identification of the similarity between the different views, allowing for computing displacements of identified points and their corresponding 3-D positions. The similarities identification is usually a time consuming task and is referenced as matching

methods based on local or global methods[6]. Local methods compare two regions of interest by measuring similarity with well-known criteria such as block matching or feature approaches[7]. These approaches are efficient on images containing a large percentage of textured area, but fail for uniform regions, because these areas contain too few information for the matching process. Occluded regions, due to the difference of point of views, are another cause of occurring errors in the matching process. Global approaches can overcome some of these issues, by estimating a disparity map that minimizes energy criteria on the whole image. Graph cuts[8], belief propagation[9] or dynamic programming[10] methods are the most well-known global approaches. Computational complexity of these methods is often high, and they are not able to properly solve the cases where images contain large uniform surfaces. For robotics or industrial control, light-patterns are projected on whole objects and background of the scene, for helping the matching process. In this case, patterns modify the aspect of homogeneous objects, which seem to be textured. These active techniques are efficient for computing disparity map, but images are then unusable for any other controls such as edge detections, intensity measurements, barcode reading or OCR.

In order to deduce a depth from disparity, a metric calibration of the acquisition system is needed. Usually, calibration consists in determining intrinsic parameters for the acquisition device used for metric measurements, and permits to compute distances between scene objects and the camera device. An interesting survey of the reference methods is proposed by Zhang[11], where approaches are divided in three categories following the reference object used for calibration. Highest accuracy can theoretically be obtained by using a 3-D object for calibrating, but in practice, calibration with a 2-D apparatus seems to be the best choice in most situations, because of its ease of use and good accuracy. In this case, calibration is performed by taking

images of the 2-D pattern plane under different orientations by moving either the plane or the camera. A dozen of acquisitions is then a recommended minimum. Calibration of multi-camera devices has been studied by Vaish[2]. This work compares a non-metric approach, using plane and parallax, with metric calibration for computing synthetic aperture images. If the operating mode for calibrating is similar to our approach developed in this paper, the proposed affine model leads to generate images focused at different relative or uncalibrated depths.

We propose in this work a multi-lens device (Fig. 1) and a post-processing algorithm using contours in order to avoid computing a disparity map of homogeneous objects by using projected light patterns. A fast and flexible calibration step is also proposed. The calibration of this device could be qualified as an indirect method, because intrinsic parameters are not determined individually as is usually done, but in an overall manner using a reference homography. One of the main interests is to simplify the calibration step, which is then easily achievable by an operator for industrial applications. This method is an improvement of previous works[12], in which we have extended the principles of "variable homography", defined by Zhang and Greenspan[13] for measuring the height of emergent fibers on glass and non-wovens fabrics. This method has been defined for working with fabric samples progressing on a conveyor belt. Triggered acquisition of two successive images was needed to perform 3D measurements. In this work, we have retained advantages of variable homography for measurements along the optical axis, but we have reduced the acquisitions to a unique one, by developing a device made of four lenses placed in front of a single image sensor. We have also adapted the variable homography formulation for this device, and we give a new formulation to calculate the depth. This method is presented in the first part of the paper and can be applied to any disparity maps calculated from various methods. In the second part, we propose also a new framework for computing disparity

map adapted to our device, which then becomes a depth camera. The disparity map computation is an alternate to matching methods, exploring a contour based solution, where no correspondence matching is needed, drastically reducing the computation time. In 3-D industrial control by vision, most of stereo-vision systems project structured or random patterns on objects for providing texture on uniform surfaces. This method is efficient for measuring disparities between the multiple views, but in this case, only the depth images can be used for image analysis because the scene photometry is strongly modified by the projected pattern. On the contrary, we propose an algorithm preserving gray image levels. The camera could then be used to both measure object height and perform any other classic vision control, such as optical characters or pattern recognition. However, this method is not fully universal, and we have only studied cases for measuring flat objects, in a context of industrial control.

This paper presents firstly the geometry of our device and describes the extension of variable homography for 3-D measurements. This approach leads to propose a new calibration scheme, for which only a reference homography is needed, instead of the intrinsic parameters, as usually. Secondly, we have proposed an alternate manner of computing a depth map. Preliminary and promising results on synthetic and real images are presented in the last part.

Fig. 1 (a) Exploded view of the camera with its 4 mini-lenses (b) Assembled camera (c) Corresponding sub-images (North-West, North-East, South-West and outh-East).

2 Variable homography and depth measurements from disparity

The concept of variable homography has been defined by Zhang and Greenspan[13] for parallax compensation in image mosaicking without metric calibration. This concept has then been extended by Xu et al.[12], for 3-D measurements by using two successive acquisitions, of fabrics scrolling on a conveyor belt.

This part presents how the variable homography can be used for modeling our multi-view device, described on the figure 1. Figure 2 gives a schematic representation of our system made out of mini-lenses and a single image plane. This figure is a 2D section representing only two neighboring projections, given the sub-images noted I_i and I_{i+1} (we will use $i=1$ for next expressions). With four projections on the image sensor, each sub-image has three possible neighbors. If index i represents the *SE* view (South-East), index $i+1$ can be any view of *NE* (North-East), *SW* (South-West) or NW (North-West). This scheme simplifies our system to a parallel stereovision system, where triangulation could be used to determine the depth of point P when intrinsic and distortion parameters are known. Ideal situation is encountered when these

parameters are identical for each camera, which is however never the case in practice. Precise calibration is then needed to determine numerical values of optical centers and focal distances. These parameters are necessary for the geometric re-projection, for computing depth by intersecting the two rays of each associated left and right image pixel. For example, on fig. 2, f_1, f_2, O_1, and O_2 must be known for recovering the depth of a point P from the rays $(O_1\ p_{1B})$ and $(O_2\ p_{2B})$.

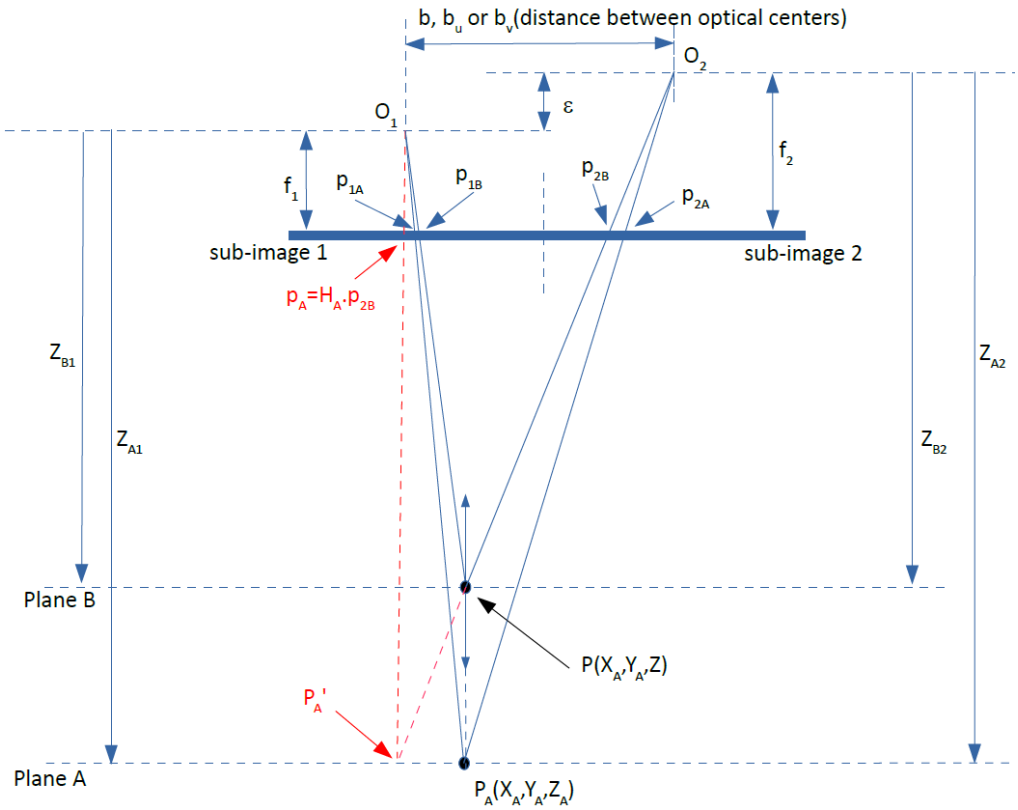

Fig. 2 Geometric cross-section between two adjacent sub-images. O_1, O_2: optical centers of sub-images 1 and 2; b: distance between O_1, O_2; f_1, f_2: lenses focal distances; ε difference between f_1, f_2. Z_{Ai}, Z_{Bi}: distances between O_i and planes A and B. p_{1A} and p_{2A}: P_A projections on both sub-images. p_{1B} and p_{2B} : P projections on both sub-images. p_A: projection on sub-image 1 of the virtual point P'_A.

An alternative approach is to take benefit of variable homography for modeling this device.

2.1 Variable homography definition for a pair of views

Variable homography formulation introduces a matrix K giving some interesting properties between both homography H_A and H_B defined for the A and B parallel planes and verifying $p_{1A}=H_A.p_{2A}$ and $p_{1B}=H_B.p_{2B}$:

$$H_B(k) = K_i.H_A.K_{i+1}^{-1} \quad \text{where} \quad K_n = \begin{pmatrix} k_n & 0 & (1-k_n)u_{0n} \\ 0 & k_n & (1-k_n)v_{0n} \\ 0 & 0 & 1 \end{pmatrix}, \tag{1}$$

with $n=i$ or $i+1$ and where u_{0n} and v_{0n} stand for the coordinates of the optical centers O_n projected on the image sensor. The main parameter of K is the distance ratio k_n defined by $k_n=Z_{An}/Z_{Bn}$, giving the possibilities to compute H_B from H_A. By sweeping k around realistic values and by using a matching comparison method, it would be possible to find reliable correspondences where $k=argmax(Matching(p_{1B}(k), p_{2B}))$. This is the idea presented by Zhang and Greenspan[13] in their paper to compensate parallax in image mosaicking. We propose to extend this method, by developing the expressions of H_A and H_B, in order to simplify the re-projection expression to compute depth and easily include in our model all intrinsic parameters described on fig. 2.

2.2 H_A and H_B expressions

Plane A is considered as the reference plane, and is used for calibrating the device. As our device uses 4 mini-lenses, it's interesting to study the case where these lenses don't have exactly the same focal distances. H_A is then determined by assuming that homography is a combination of an intrinsic parameter and translation matrices, and is given in Ref. [1] as:

$$H_A = \begin{pmatrix} \frac{f_1}{f_2}\frac{Z_{A2}}{Z_{A1}} & 0 & u_{01} + \frac{b_u f_1}{Z_{A1}} - \frac{f_1}{f_2}\frac{Z_{A2}}{Z_{A1}}u_{02} \\ 0 & \frac{f_1}{f_2}\frac{Z_{A2}}{Z_{A1}} & v_{01} + \frac{b_v f_1}{Z_{A1}} - \frac{f_1}{f_2}\frac{Z_{A2}}{Z_{A1}}v_{02} \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \alpha & 0 & u_{01} + \frac{b_u f_1}{Z_{A1}} - \alpha.u_{02} \\ 0 & \alpha & v_{01} + \frac{b_v f_1}{Z_{A1}} - \alpha.v_{02} \\ 0 & 0 & 1 \end{pmatrix}, \quad (2)$$

where b_u and b_v are horizontal and vertical distances between optical centers, respectively. By considering $Z_{A2} = Z_{A1} + \varepsilon \approx Z_{A1}$ at the working distances, H_A has been simplified in eq. 2 by introducing the focal distance ratio $\alpha = f_1/f_2$. Expressions (1) and (2) lead to give a simplified expression for H_B, representing the geometric transformation for any point P located in a virtual plane B, parallel to the reference plane A:

$$H_B = \begin{pmatrix} \alpha & 0 & u_{01} + k_1 \frac{b_u f_1}{Z_{A1}} - \alpha.u_{02} \\ 0 & \alpha & v_{01} + k_1 \frac{b_v f_1}{Z_{A1}} - \alpha.v_{02} \\ 0 & 0 & 1 \end{pmatrix}. \quad (3)$$

2.3 P point depth computation from a pair of views

From H_A and H_B homographies, we propose to compute the depth of any point P, by the definition of disparity d given by $d = H_B.p_{2B} - H_A.p_{2B}$. By considering that $p_{1B} = H_B.p_{2B}$ and with expressions (2) and (3), d can be written as:

$$d = \underbrace{H_B.p_{2B}}_{p_{1B}} - \underbrace{H_A.p_{2B}}_{p_A} = dist(p_{1B}, p_A) = \left(\left(\underbrace{(k_1 - 1)\overbrace{\frac{b_u f_1}{Z_{A1}}}^{C_u}}_{d_u} \right)^2 + \left(\underbrace{(k_1 - 1)\overbrace{\frac{b_v f_1}{Z_{A1}}}^{C_v}}_{d_v} \right)^2 \right)^{1/2}. \quad (4)$$

Assuming that $k_1 = Z_{A1}/Z_{B1}$ and $b = \sqrt{b_u^2 + b_v^2}$, relations between depth Z_{B1} and disparity d are finally given by:

$$Z_{B1} = Z_{A1}\frac{\sqrt{C_u^2 + C_v^2}}{d.Z_{A1} + \sqrt{C_u^2 + C_v^2}} = \frac{bf_1}{d + \dfrac{bf_1}{Z_{A1}}} \qquad \text{or} \qquad d = \frac{bf_1}{Z_{B1}} - \frac{bf_1}{Z_{A1}} \ . \tag{5}$$

This definition of disparity is slightly different of the one usually given. In the next part, we study the differences and advantages to use this definition for 3-D measurements, especially in a 3-D industrial vision context.

2.4 Comparison with triangulation method

When two cameras are separated by lateral translation with no rotation, triangulation method leads to obtain the depth relative to p_{1B} and p_{2B} coordinates. From figure 2, depth Z_{B1} can be expressed as:

$$Z_{B1} = \frac{b + \dfrac{\varepsilon}{f_2}(O_2 - p_{2B})}{\dfrac{(p_{2B} - O_2)}{f_2} - \dfrac{(p_{1B} - O_1)}{f_1}} \tag{6}$$

This relation is usually simplified assuming $f_1=f_2$, and $O_1=O_2$ in the coordinate systems of each sub-image:

$$Z_{B1} = \frac{bf_1}{d} \qquad \text{with} \qquad d = p_{2B} - p_{1B} \ . \tag{7}$$

By comparing this expression with the one obtained with variable homography, there is an evident similarity, and variable homography translates the disparity d by a term of $(-bf_1/Z_{A1})$. This is observed on fig. 3, comparing expressions (7) and (5). Curves are plotted according to the real dimensions of our device: pixels size of *10 µm*, optical centers spaced by *9 mm* and $f_1 = f_2 = 8$ *mm*. Triangulation curve is plotted in red, and the ones obtained from equation (5) are plotted in blue for several values of the reference distance Z_{A1}. In function of Z_{A1}, blue curves are thus shifted to the left. In fact, this translation offers some interesting properties. When the range of

measurement is predicable, it would be then possible to work in the same range of disparity whatever the working depth, by choosing correctly Z_{A1}.

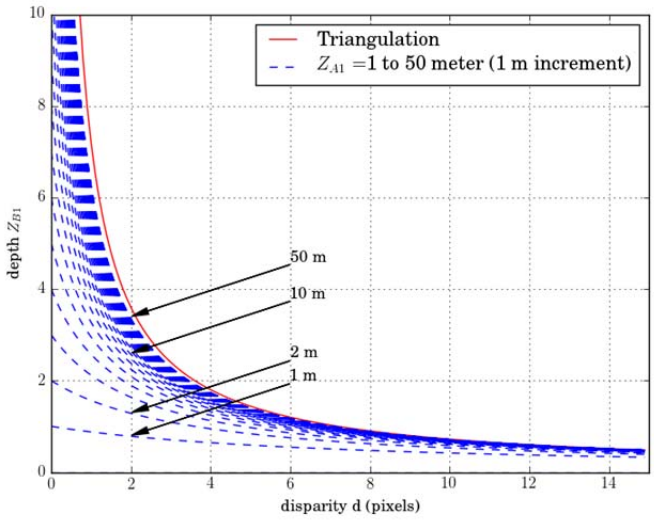

Fig. 3 Depths Z_{B1} as a function of disparity d. Triangulation and variable homography comparison.

It is quite easy to demonstrate that the depth error ∂Z_{B1} stays identical for both approaches, meaning this translation does not influence the accuracy of measurement:

$$\frac{\partial d}{\partial Z_{B1}} = -\frac{bf_1}{Z_{B1}^2} \quad \rightarrow \quad \partial Z_{B1} = -\frac{Z_{B1}^2}{bf_1}\partial d \tag{8}$$

There are some advantages to use variable homography for working within the same range of disparity. The same device and same algorithm can be used for a wide working distance range, from macroscopic applications to higher scales. For example, if Z_{A1} equal *7 cm*, depth Z_{B1} is ranged between *7* and *5.8 cm* for an interval of *d* of *[0,20] pixels*. For 3-D industrial vision inspection, when Z_{A1} is fixed at *90 cm*, the corresponding measured depths are ranged from *90 cm* to *26 cm*, while keeping the same *d* interval. We can note that, for these different scales of measurements, we have used the same device. Switching between the different scales only

requires adjusting the lenses focal distances. This last point is important, because in this case, accurate measurement would need to re-determine all intrinsic parameters using expression (6) when triangulation method is used. Next part explains how variable homography leads to an easier calibration step.

2.5 Simplified calibration for multi-view cameras

In practice, lenses are never perfectly aligned in multi-lenses cameras (lenses are not sorted), and lenses focal distances are also never strictly equal[14]. This is especially true for our prototype and this creates some undesirable distortions affecting depth measurements. In practice, imperfections are rarely taken into account and literature usually gives expression (7) as reference method to compute depth instead of the complete expression given by eq. (6). By considering a multi-view device, the complete determination of these intrinsic parameters could be an uncomfortable task, requiring multiple acquisitions[15], but nevertheless mandatory for providing accurate measurements.

A great advantage of variable homography is to include in matrix H_A all intrinsic parameters between the views of 2 sub-images. By projecting measurement points from sub-image 2 to sub-image 1 with H_A, the number of required parameters for 3-D measurement is then reduced, and they are easy to determine via a simple calibration process. Sole parameters to recover during the calibration step are: Z_{A1}, H_A, and the constant bf_1. The easy and robust proposed calibration method is realized by using a reference chessboard pattern for detecting corners and by performing two successive acquisitions. The first acquisition is done at reference depth Z_{A1}, and the second one at depth $Z_{B1}=Z_{A1}+d_z$. Corners of first acquisition are used to calculate all combinations of existing H_A homographies between sub-images. For each acquisition, distances in pixels between the first and last corners detected on the chessboard are computed. We note d_A

the distance for the dataset of first acquisition and d_B the distance for dataset of the second one. Z_{A1} is then obtained by triangulation and is formulated as:

$$Z_{A1} = \frac{d_z . d_B}{d_B - d_A} \quad . \tag{9}$$

Finally, Z_{A1}, H_A, d_z and distances d, computed for corresponding corners p_{1B} and p_{2B} during the second acquisition, lead to find best value for bf_1. In conclusion, interest of this calibration lies in its simplicity when calibrating, because it is easier to calibrate via two acquisitions than to determine the intrinsic parameters with several snap images, as usually done with well-known standard methods[15].

This calibration process for determining the variable homography parameters has been established by considering two views. This principle is easily extensible for a n-view camera, i.e our device. Calibration is then performed for each pair of working views, without additional acquisition.

2.6 Improved calibration with lens distortion correction

Our multi-view camera prototype uses low-cost lenses, presenting a distortion coefficient close to 1.5% (value is issued from lens datasheet). To improve measurements accuracy, we propose to take into account the distortion in our calibration scheme. By using standard approaches, such as corrections by using radial distortion models, the precise positions of the optic centers of each sub-image can be determined. As these intrinsic parameters are combined in H_A, they are not immediately available at this step with our calibration process. We propose therefore an alternate approach. We consider that distortion effects are comparable to a non-uniform image magnification on image plane. This approach leads to adjust the computation of Z_{B1} as a function of the spatial location of correspondences in sub-images. This can be done by using the proportional relationship existing between magnification, focal distance and parameter bf_1. The

correction is performed by using the Γ operator, which transforms a vector representing a point P of coordinates *(u,v,1)* in the lifted coordinates[16] $\widetilde{P}$:

$$P(u,v,1)^T \rightarrow \widetilde{P} = \Gamma(P,P) = (u^2, v^2, u.v, u, v, 1)^T \quad . \tag{10}$$

By using p_{B1} as reference location, bf_1 is then computed by the quadratic functions defined by:

$$bf_1 = \begin{pmatrix} a & b & c & d & e & f \end{pmatrix}.\widetilde{p}_{B2} \quad . \tag{11}$$

Corners collected previously form a dataset and permit to compute the polynomial coefficients by using a least mean square method.

2.7 Measurement accuracy

The measurements obtained in this part have been obtained with our device presenting the following characteristics: CCD size: 15x15 mm^2; CCD resolution: 2048x2048 pixels; mini-lenses diameters: 8 mm; mini-lenses focal distance: 7.5 mm; distance between mini-lenses: 9 mm; sub-images resolution: 550x550 pixels. We present experimental measurements for three ranges of depth, in order to estimate the accuracy of our device in term of depth measurement. Here, only measurements performed on SW and SE sub-images are presented for this experimentation. As shown by expression (8), depth error is a function of two factors: geometric parameters estimation and disparity measurement.

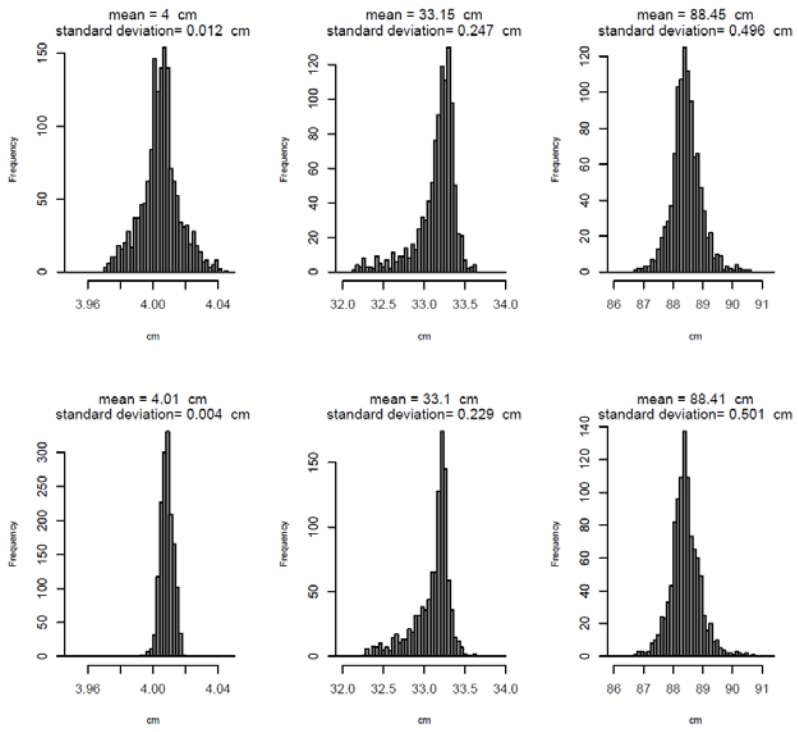

Fig. 4 Depth measurement distributions for three ranges of measurements: 4 cm, 33 cm and 88.5 cm.

Top: results of simplified calibration method. Bottom: results of improved calibration method.

At this point, we are interested in measuring the quality of the parameters estimated during calibration. For minimizing the role of disparity measurement in our estimation, we have used a pattern chessboard as depth target and corners detection is performed with the corresponding function of the OpenCV library. For each depth level, numerous acquisitions are performed, leading to about one thousand measurements. We have set the camera-chessboards distances at 4 cm, 33 cm and 88.5 cm, and the measured depth corresponds to Z_{B1} (distance separating the chessboard and the optical center). Figure 4 gives distributions of depth measured for each corners. By working at a short distance, captured images present important optical distortions. In this case, the second method offers a significant improvement, providing a standard deviation lower than 1%. For evaluating depth error, percentage is not significant here, because the

uncertainty of measurements is directly linked to the depth range (see eq. 8), meaning depth error increases systematically with depth.

This is the reason why our evaluation is presented by comparing theoretically predicted error with the measured standard deviation. In this case, the main difficulty concerns evaluation of the disparity variation ∂d , which is a difficult to assess parameter. Even if we use a subpixel corner detector, ∂d heavily depends on the pattern chessboard scale in sub-images and thus of the working distance. To provide best estimations, we have bounded its values between two realistic limit values, fixed here at *0.1* and *0.5* pixel. These values are then used for the result comparisons presented in table 1. This table shows that standard deviations observed during our measurements are comprised within the range of estimated depth error. This confirms that the proposed calibration is well-adapted to our measurement approach by variable homography. These values of depth error allow for keeping a good uncertainty of measurements for applications in the field of metrology by vision.

Table 1 Comparison between theoretical and experimental depth measurement error. (bf_1=7.2 pixels is obtained during the calibration stage). SC: simplified calibration; IC: Improved calibration

Depth Z_{B1} (cm)	**Predicted depth error** $\partial Z = (z^2 / bf_1).\partial d$	**Range of estimated theoretical depth error**		**Observed experimental standard deviation**
		$\partial d_{\min} = 0.1$ pixel	$\partial d_{\max} = 0.5$ pixel	**SC / IC**
4	$(0.22\ 10^{-3}).\partial d$ =0.05 mm	0.022 mm	0.111 mm	0.12 mm / 0.04 mm
33	$(15.1\ 10^{-3}).\partial d$ =3.78 mm	1.51 mm	7.56 mm	2.47 mm / 2.29 mm
88.5	$(107\ 10^{-3}).\partial d$ =2.69 cm	10.7 mm	53.8 mm	49.6 mm / 50.1 mm

2.8 Sub-images rectifications

Previous measurements were performed on coordinates of detected corners. To transform our device into a depth camera, disparity has to be computed on all pixels of the sub-image. As we have defined disparity as $d=p_{1B}-H_A.p_{2B}$, a simple pre-processing steps can be performed, for

rectifying each sub-image with homographies H_A. In this case a reference sub-image must be chosen. Next examples consider *SE* as reference sub-image, as illustrated on fig. 5, but any other sub-image could be selected instead. By rectifying sub-images, we ensure that correspondences between images are well aligned along the epipolar line, as well-known in stereovision. By organizing lens positions in a square, epipolar lines are then horizontal, vertical or diagonal following positions of sub-images used for computing disparity. Multi-views approach is then interesting because in the reference sub-image, epipolar lines calculated from other views could be combined to improve the matching for establishing correspondences and thus enhanced the final disparity map.

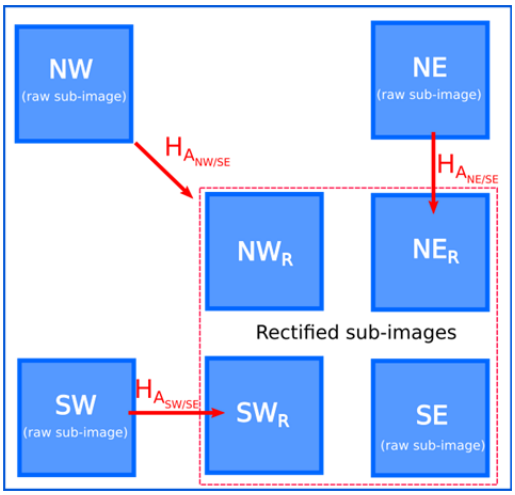

Fig. 5 Sub-images rectification: pre-processing before computing disparity maps. By choosing SE as reference sub-image, disparity maps between (SE, SW_R), (SE, NE_R) and (SE, SE_R) can be computed.

3 Computation of disparity with the four mini-lenses device

With the set of rectified images, we propose to compute the corresponding disparity image (or disparity map), in order to use our device as a depth camera. The main idea is to subtract distances map between sub-images in order to highlight disparity values. This approach is to be

used in a normal lightning environment with uniform objects. The next section shows that the proposed method doesn't need complex calculations, which is an advantage for real-time applications. However, this method is not universal, and can, up to now, be used for flat objects only.

3.1 Principle

The four sub-images delivered by our device can be combined for computing up to three disparity images. We have validated the principle by using only two associations with three sub-images: disparity (SE, SW_R) and disparity (SE, NE_R). The reference sub-image is then *SE*, as illustrated on fig. 5. Our goal is to fill the content of any objects by the disparity existing between comparable contours of two sub-images. To do so, we propose an algorithm composed of four main steps and operating between two sub-images:

- computation of Distance Images DI^{xx};
- computation of pseudo-Subtraction Images pSI^{xx};
- application of a fusion process;
- computation of a disparity map by decoding the image given by the previous step.

As up to three disparity maps can be computed from the reference sub-image, a second and final step merges these maps in order to eliminate some artifacts. The depth image is then directly obtained, thanks to the previous calibration, and using expression (5).

- **Distance images: DI^{xx}**

The first step generates distance images (*DI*) following predetermined directions. When the working sub-images are horizontals, then DI^{LR} (Distance Image Left to Right) and DI^{RL} (Distance Image Right to Left) are generated. Similarly, Top/Bottom distance images are

generated when vertical sub-images. Diagonal extension is also possible. Distance Images are constituted of values indicating the distances between pixels to the previous contour encountered following the scanning direction. A flag × is used when no contour has been yet encountered by scanning. Figure 6 illustrates examples to compute DI^{LR} and DI^{RL} images from *SE* and SW_R views. On this example, a rectangular object is represented by its contour pixels, highlighted in red. A disparity *d* (or parallax d_u) of *4* pixels is simulated and the green points P_1 and P_2 represent a point having same coordinates on both original sub-images. The main property used for our method is based on the fact that the difference $DI^{LR}(SE) - DI^{LR}(SW_R)$ or $DI^{RL}(SE) - DI^{RL}(SW_R)$ for any pixel inside the object is a function of the disparity value *d*.

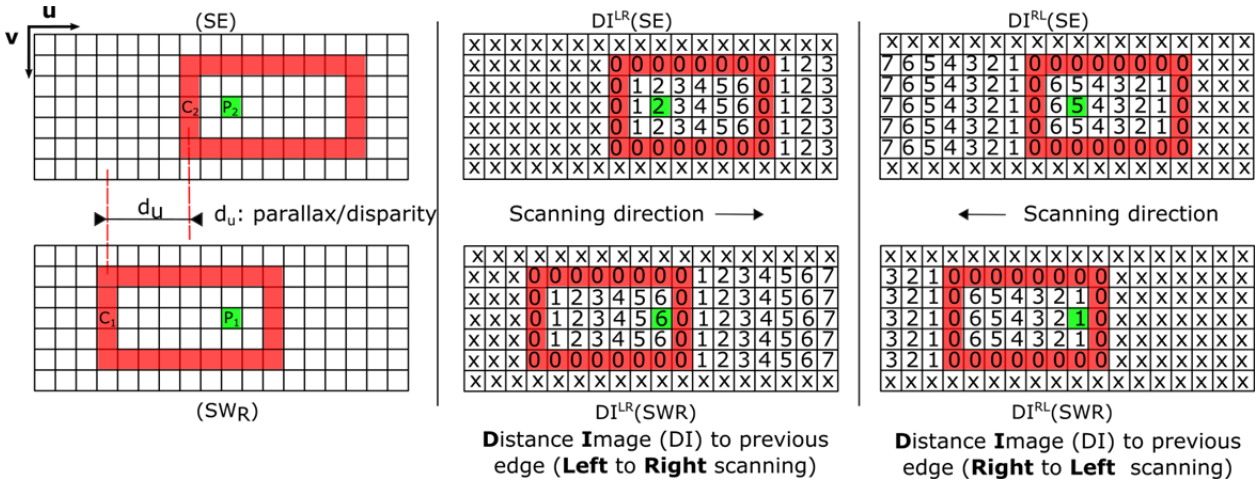

Fig. 6 Example of computation of distance image DI^{LR} and DI^{RL}

- **Pseudo-Subtraction Image: *pSI*^{xx}**

We have defined two distinct operators for exploiting the property that subtraction between *DI* images can give the disparity. Furthermore, this property is not true on the whole image, especially when the flag × is encountered. We have established that the best result, where object is filled by the disparity value, is obtained by using two kinds of subtraction called here *pseudo-*

subtraction, where the result value can be the flag $\times$, the first operand or the true difference. These operators are defined with arguments a and b, where a and b are pixel values of the reference image and the second input image, respectively. Definitions of these operators are for DI^{LR} images:

$$pSI^{LR}(a,b) \rightarrow \begin{cases} (a-b) \ if \ a,b \in \mathrm{N} \\ \times otherwise \end{cases}, \tag{12}$$

and for DI^{RL} images:

$$pSI^{RL}(a,b) \rightarrow \begin{cases} (a-b) \ if \ a,b \in \mathrm{N} \\ a \ if \ b = \times \\ \times otherwise \end{cases}. \tag{13}$$

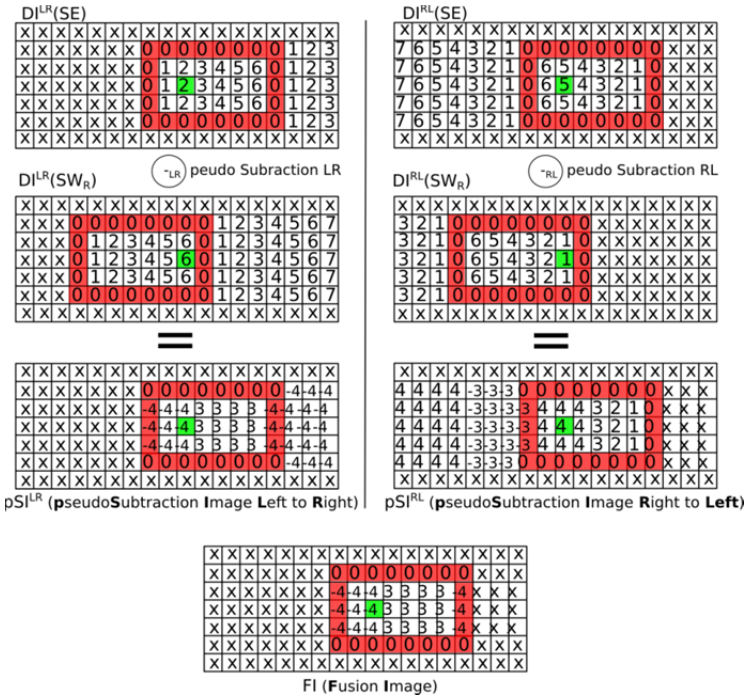

Fig. 7 Examples of pseudo-subtractions pSI^{LR}, pSI^{RL} and FI for an object measuring l=8 width and for a d=4 disparity.

- **Fusion image**

The fusion uses pSI^{LR} and pSI^{RL} images for replacing on pSI^{LR} image, the values appearing outside object by flag $\times$. This operation is simply defined by:

$$FI(a,b) \rightarrow \begin{cases} a & if\ a,b \in \mathrm{N} \\ \times & otherwise \end{cases}. \tag{15}$$

- **Decoded fusion image**

The goal of this final step is to decode the content of fusion image in order to identify each object by its disparity, position and dimension. Thanks to the previous steps, especially the distance propagation and pseudo-subtraction ones, a complete description of objects contained in original sub-images, can be extracted from the pixel values of the fusion image. These values are constituted of combinations of objects positions, sizes and disparities. For N separated objects in each view, we have established by simulations, that each pixel value in fusion image, can be decoded. We note and define:

- d_i, l_i and p_i, respectively, the disparity, length and first position of object i;
- A_i=-d_i; B_i=l_i-d_i-1; C_i=p_i-d_i-p_{i-1}-l_{i-1}+1;
- $Seq_1 = |A_1 \ldots A_1 B_1 \ldots B_1 A_1 \ldots A_1|$;
- $Seq_i = |C_i \ldots C_i A_i \ldots A_i B_i \ldots B_i A_i \ldots A_i|$.

Each line of the fusion image is then constituted of the following typical sequences arrangement: $[\times \ldots \times; Seq_1; Seq_2; \ldots Seq_i; \ldots Seq_N; \times \ldots \times]$. The number of occurrences of terms A_i, B_i and C_i inside each sequence depends on the disparity, length and position of objects in the sub-images. This information is not used for the decoding. For the example of figure 7, with d_1=4 and l_1=8, where

image fusion is constituted of the arrangement $[\times..\times;-4;-4;-4;3;3;3;3;-4;\times...\times]$, one can recognize Seq_1.

Finally, the role of the decoding stage is to use this property for producing the disparity map. Depth images are then established with expression (5), linking depth and disparity.

3.2 Merging disparity maps

The algorithm presented above computes disparity map from two sub-images. A complete overview, resuming step connections, is shown on fig. 8. This scheme is designed work with any couple of sub-images. As this approach is only based on the contours detected in each sub-images, the quality of initial contours is a primary initial pre-preprocessing step. For the moment, no matching process is performed between sub-images and this is an advantage of this approach. Even if matching could be useful for ensuring that each contour point in sub-image *i* would have a corresponding point in sub-image *i+1*, we have decided to skip it, for increasing the speed of our method. But, with no matching process, some contours in sub-image *i* may have no correspondence in sub-image *i+1*. *DI* images are then affected and some artifacts in disparity images can appear. Following scanning direction, they are characterized by horizontal, diagonal or vertical trails. A solution for attenuating these artifacts has been found by exploiting the multi-view capability of our device. As from a reference image (see fig. 5) three disparity maps can be computed, we propose to merge them for smoothing artifacts and retaining the most plausible disparity value for each pixel.

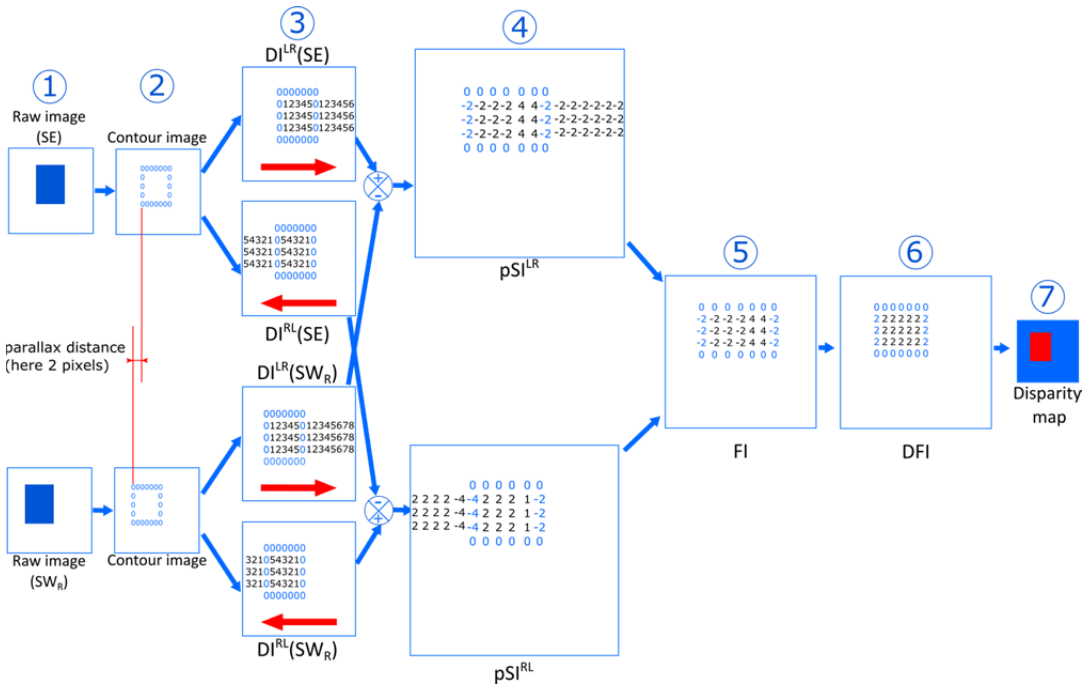

Fig. 8 Steps used for computing disparity map from SE and SW_R sub-images. 1: sub-image acquisition; 2: binary contour extraction; 3: Distance images computation according scanning direction; 4: pseudo-subtraction images computation: 5; fusion step; 6: image fusion decoding; 7: final disparity map

3.3 Results

Figure 9 presents some results on synthetic scenes, showing objects located at several depths. The first case shows a basic configuration, where objects are located far away from each other. A second case shows superimposed objects. For these two cases, no final merging is necessary and only one disparity map is sufficient to generate the final depth image. For the last example, we consider a situation where all contours pixels have not correspondences in sub-images, as can happen in experimental conditions. The map image is then computed with the merging process using here disparity images (SE, SW_R) and (SE, NE_R) as inputs. Final depth image still contains residual errors, but stays very close to the expect result and is suitable for most of 3-D inspection processes in industrial vision.

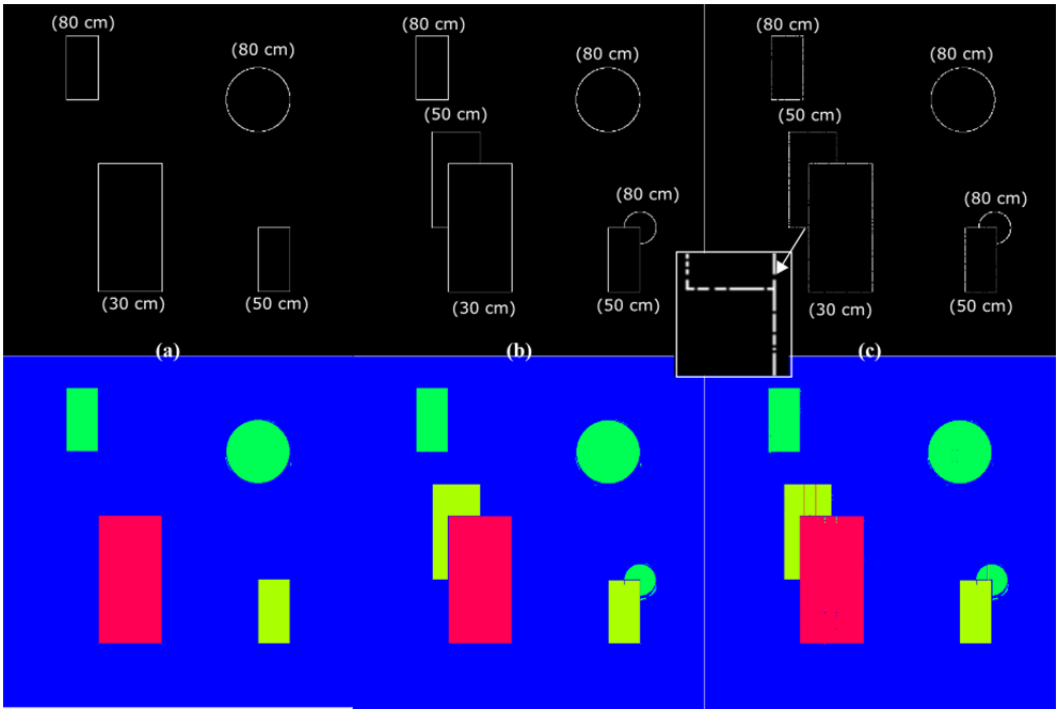

Fig. 9 Synthetic images (only sub-image SE is shown) containing objects at different depths and corresponding depth maps. (a): 4 objects without superposition are placed in the scene; (b): some superposed objects are placed; (c): contours in sub-image are randomly degraded.

Instead of presenting experimental results in the form of depth images, we prefer defining the height images, which facilitates comparison with reference objects: *height image* = Z_{A1}*-depth image*. Differences are just a translation of measurement values by the reference distance Z_{A1}. Figure 10 shows some typical height images computed on real scenes. We explore situations with objects having large uniform areas in natural lightning, for which standard matching methods often fail when no structured illumination is used. To do so, we have adapted the calibration distance Z_{A1} to the objects scale, as explained in paragraph 2.4. Heights measured for the four reference objects are then well recovered. One can however see some artifacts, which are still visible on the final depth map. The main characteristic of the proposed method is its ability to compute disparity map without requiring any matching. With a single image sensor snapping simultaneously four views, the captured images present similar brightness and contrast.

It could be expected that edge detection provides same contours for the four views. In practice, due to the image sensor sampling, it is not always the case and some no-exact correspondences between detected contours are sometime encountered. Figure 11 presents an example of this situation, when no-exact correspondences between detected contours corrupt the depth map. Other artifacts could also be caused by the difference of point of views, resulting in occultations or hidden edges between the sub-images. With our method, this kind of artifact is significantly smoothed by the merging process and is not the most frequently seen during our experiments.

Our approach for computing depth map is therefore validated by these synthetic and experimental results. This work should provide some new perspectives for establishing disparity maps without matching procedure, which is the main interest of the approach. The existing artifacts encountered in some cases should disappear or be attenuated by improving and adapting an edge detection method for multi lenses device.

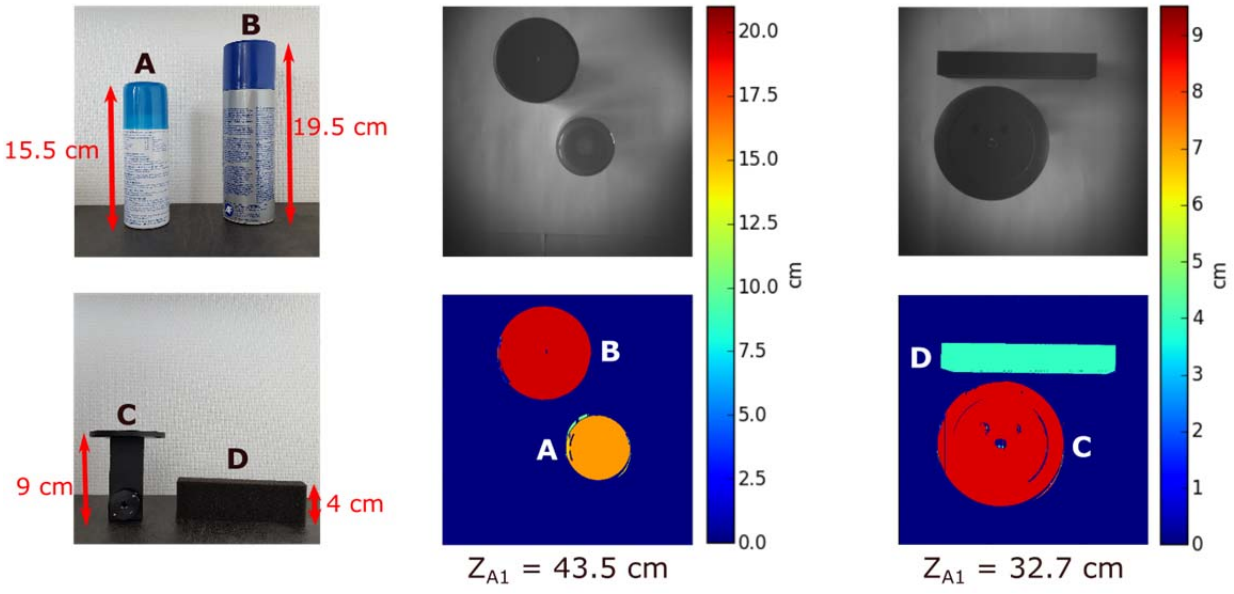

Fig. 10 Experimental results images (only sub-image SE is shown) with four objects. Measured heights are: A: 15.6 cm; B: 19.4 cm; C: 8.7 cm and D: 3.9 cm.

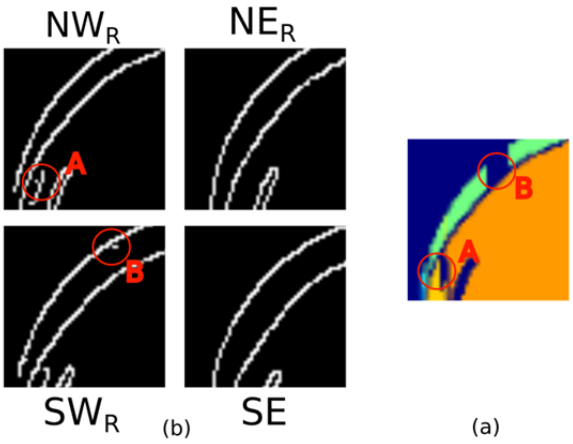

Fig. 11 Zooms on no-similar contours detection and corresponding artifacts on disparity maps.

(a) 4 contours sub-images; (b) artifact areas where the contours are not identical.

4 Conclusions and perspectives

We have described a multi-view device considered as a multi-view camera and used as a depth camera. This device can work for detecting objects to a depth of up to two meters and could find interests for industrial vision. Main contributions consist firstly in an alternate approach for modeling and calibrating this device with variable homography, and secondly in an original and new approach for computing disparity map or depth images.

Variable homography is an alternate approach for modeling multi-view systems. Main interest is to propose a rapid calibration procedure requiring only two acquisitions, without having first to determine intrinsic parameters. A comparison with the well-known triangulation method is performed, and we have demonstrated that the measurement uncertainties are then comparable, but calibration is easier in our case, especially if intrinsic parameters are not considered as identical for each sub-device. When optical distortions are significant, as for measurements performed at macroscopic scale, we have also proposed an improved calibration method, significantly reducing the measurement uncertainty. Calibration stays identical for normal and

improved method and always requires only two acquisitions. The principles developed with the variable homography method lead to use a reference homography called H_A, and measurement are based on a disparity d defined as $d = H_B.p_{2B} - H_A.p_{2B}$.

In the second part of this paper, a new approach has been presented for computing disparity map, according to the definitions given previously. The major interest of this approach is to propose a new framework requiring no matching process defeated by uniform or poorly textured areas. This approach is well adapted for measuring height of objects placed under a camera for industrial vision control. By using contours, this method is efficient on flat and uniform objects in natural lightning. Our solution can be considered as an interpolating method of the disparity or depth information in uniform areas, giving a realistic rendering of simple objects. As the scene photometry is preserved, unlike measurements performed with active illumination, some other standard controls as for example calipers, shape recognition or barcode reading, can be done conjointly with 3-D measurement.

This work has validated the interest of our approach, and some interesting perspectives are envisaged. First, we project to extend the calibration for other models of multi-views cameras, such as the focused plenoptic[17] ones. A study of the best compromise between resolution, depth accuracy measurement, working distance, pixel size and number of mini or micro lenses should be performed. Secondly, we would like to improve the original method proposed for computing disparity map. In the future, we will focus on edge detection, in order to reduce artifacts and also extend our approach for natural and complex scenes.

References

1. R. Hartley, *Multiple view geometry in computer vision*, 2nd ed, Cambridge University Press, Cambridge, UK ; New York (2003).
2. V. Vaish et al., "Using plane + parallax for calibrating dense camera arrays," in In Proc. CVPR, pp. 2–9 (2004).
3. C. Birklbauer and O. Bimber, "Panorama light-field imaging: Panorama light-field imaging," Comput. Graph. Forum **33**(2), 43–52 (2014) [doi:10.1111/cgf.12289].
4. E. H. Adelson and J. Y. A. Wang, "Single lens stereo with a plenoptic camera," IEEE Trans. Pattern Anal. Mach. Intell. **14**(2), 99–106 (1992) [doi:10.1109/34.121783].
5. A. Levin et al., "Image and depth from a conventional camera with a coded aperture," 2007, 70, ACM Press [doi:10.1145/1275808.1276464].
6. D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," in International Journal of Computer Vision, pp. 7–42 (2002).
7. R. Szeliski, *Computer Vision*, Springer London, London (2011).
8. V. Kolmogorov and R. Zabih, "Computing visual correspondence with occlusions using graph cuts," 2001, 508–515, IEEE Comput. Soc [doi:10.1109/ICCV.2001.937668].
9. Y.-C. Tseng, N. Chang, and T.-S. Chang, "Low Memory Cost Block-Based Belief Propagation for Stereo Correspondence," July 2007, 1415–1418, IEEE [doi:10.1109/ICME.2007.4284925].
10. O. Veksler, "Stereo Correspondence by Dynamic Programming on a Tree," 2005, 384–390, IEEE [doi:10.1109/CVPR.2005.334].
11. S. B. Kang and G. Medioni, Eds., *Emerging topics in computer version*, Prentice Hall PTR, Upper Saddle River, N.J (2005).
12. J. Xu et al., "Original method to compute epipoles using variable homography: application to measure emergent fibers on textile fabrics," J. Electron. Imaging **21**(2), 1–12 (2012) [doi:10.1117/1.JEI.21.2.021103].
13. S. Zhang and M. Greenspan, "Variable Homography Compensation of Parallax Along Mosaic Seams," in Image Analysis and Recognition **4633**, M. Kamel and A. Campilho, Eds., pp. 271–284, Springer Berlin Heidelberg, Berlin, Heidelberg (2007).
14. D. G. Dansereau, O. Pizarro, and S. B. Williams, "Decoding, Calibration and Rectification for Lenselet-Based Plenoptic Cameras," June 2013, 1027–1034, IEEE [doi:10.1109/CVPR.2013.137].
15. Z. Zhang, "A flexible new technique for camera calibration," IEEE Trans. Pattern Anal. Mach. Intell. **22**(11), 1330–1334 (2000) [doi:10.1109/34.888718].
16. J. P. Barreto and K. Daniilidis, "Epipolar Geometry of Central Projection Systems Using Veronese Maps," 2006, 1258–1265, IEEE [doi:10.1109/CVPR.2006.111].
17. C. Perwass and L. Wietzke, "Single lens 3D-camera with extended depth-of-field," 9 February 2012, 829108 [doi:10.1117/12.909882].

Cecile Riou is Ph.D. student at MIPS laboratory, University of Haute-Alsace, Mulhouse, France, since October 2014. She received MS degree in Optics, Image and Vision from University Jean Monnet, Saint Etienne, France and engineering degree from Telecom Saint-Etienne, Saint Etienne, France, in 2013. Her current research interests include 3-D vision, depth measurements, light-field cameras and optical device.

Caption List

Fig. 1 (a) Exploded view of the camera with its 4 mini-lenses (b) Assembled camera (c) Corresponding sub-images (North-West, North-East, South-West and outh-East).

Fig. 2 Geometric cross-section between two adjacent sub-images. O_1, O_2: optical centers of sub-images 1 and 2; b: distance between O_1, O_2; f_1, f_2: lenses focal distances; ε difference between f_1, f_2. Z_{Ai}, Z_{Bi}: distances between O_i and planes A and B. p_{1A} and p_{2A}: P_A projections on both sub-images. p_{1B} and p_{2B} : P projections on both sub-images.

pA: projection on sub-image 1 of the virtual point P'_A.

Fig. 3 Depths Z_{B1} function of disparity d. Triangulation and variable homography comparison.

Fig. 4 Depth measurement distributions for three ranges of measurements: 4 cm, 33 cm and 88.5 cm. Top: results of simplified calibration method. Bottom: results of improved calibration method.

Table 1 Comparison between theoretical and experimental depth measurement error. (bf_1=7.2 pixels is obtained during the calibration stage). SC: normal calibration; IC: Improved calibration

Fig. 5 Sub-images rectification: pre-processing before computing disparity maps. By choosing SE as reference sub-image, disparity maps between (SE, SW_R), (SE, NE_R) and (SE, SE_R) can be computed.

Fig. 6 Example of computation of distance image DI^{LR} and DI^{RL}

Fig. 7 Examples of pseudo-subtractions pSI^{LR}, pSI^{RL} and FI for an object measuring l=8 width and for a d=4 disparity**.**

Fig. 8 Steps used for computing disparity map from SE and SWR sub-images. 1: sub-image acquisition; 2: binary contour extraction; 3: Distance images computation according scanning direction; 4: pseudo Subtraction images computation: 5; fusion step; 6: image fusion decoding; 7: final disparity map

Fig. 9 Synthetic images (only sub-image SE is shown) containing objects at different depths and corresponding depth maps. (a): 4 objects without superposition are placed in the scene; (b): some superposed objects are placed; (c): contours in sub-image are randomly degraded.

Fig. 10 Experimental results images (only sub-image SE is shown) with four objects. Heights measured are: A: 15.6 cm; B: 19.4 cm; C: 8.7 cm and D: 3.9 cm.

Fig. 11 Zooms on no-similar contours detection and corresponding artifacts on disparity maps. (a) 4 contours sub-images; (b) artifact areas where the contours are not identical.