Available online at www.sciencedirect.com
ScienceDirect
Computerized Medical Imaging and Graphics 31 (2007) 625–637
Computerized Medical Imaging and Graphics
www.elsevier.com/locate/compmedimag

A binocular machine vision system for three-dimensional surface measurement of small objects☆

Dimitris Gorpas[a,*], Kostas Politopoulos[b,1], Dido Yova[b,2]

[a] National Technical University of Athens, School of Electrical and Computer Engineering, Laboratory of Biomedical Optics and Applied Biophysics, 9 Iroon Polutexneiou Street, Zografou Campus, 157 73 Zografou, Greece
[b] National Technical University of Athens, School of Electrical and Computer Engineering, Laboratory of Biomedical Optics and Applied Biophysics, 9 Iroon Polutexneiou Street, Zografou Campus, 157 73 Zografou, Greece

Received 6 November 2006; received in revised form 11 July 2007; accepted 16 July 2007

Abstract

Rendering three-dimensional information of a scene from optical measurements is very important for a wide variety of applications. However, computer vision advancements have not yet achieved the accurate three-dimensional reconstruction of objects smaller than 1 cm diameter. This paper describes the development of a novel volumetric method for small objects, using a binocular machine vision system. The achieved precision is high, providing a standard deviation of 0.04 mm. The robustness, of the system, issues from the lab prototype imaging system with the crucial *z*-axis movement without the need of further calibration and the fully automated volumetric algorithms.

Keywords: Photogrammetry; Calibration; Structured light; Conjugate points; Small object imaging; 3D skin tumour imaging

1. Introduction

Traditionally, photogrammetry has been defined as the process of deriving metrical information about an object through measurements made on photographs of the object [1]. Modern photogrammetry covers a considerably wider domain. Imagery of all types, both passive, such as photography, and active (i.e. providing its own energy source), such as fluorescence imaging, is used. The fundamental task of photogrammetry is to rigorously establish the geometric relationship between the image and the object, as it existed at the time of the imaging event. Once this relationship is correctly recovered, one can then derive information about the object strictly from its imagery.

Several types of large-scale problems – such as long-distance measurement of a landscape via aerial, terrestrial or even satellite imagery [2,3] and industrial quality control of relatively large work-pieces [4,5] – are well solved, although there are still no fully automatic commercial systems available [3]. On the other hand, close-range photogrammetry is not yet well encountered, especially when objects with quite small, less than 1 cm diameter, dimensions are involved.

In regards to biomedical technology, there is increasing interest in methods for acquiring three-dimensional (3D) models of the human face (e.g. pre- and post-operative planning for facial surgery, creating human avatars to populate virtual environments, facial encoding, etc.), many of which obtain acceptably small 3D reconstruction errors [6,7]. However, some efforts to reconstruct and measure human hand have just recently been established, while there is still to overcome major problems faced in skin lesion 3D measurements.

This paper presents a novel volumetric method for gauging applications of objects smaller than 1 cm diameter. The need of accurate surface rendering and measurement is present in numerous modern life activities, such as industrial quality control measurements, archeology, and biology. All these fields involve 3D reconstruction and measurement of small objects with high accuracy levels. The hardware premises are described providing images with great resolution, under certain image acquisition protocol, since the input information is among the most important aspects for all problems. The use of telecentric lenses is introduced in a binocular biomedical imaging applica-

☆ A binocular gauging system for small objects.

* Corresponding author. Tel.: +30 2107722293; fax: +30 2107723894.
E-mail addresses: dgorpas@mail.ntua.gr (D. Gorpas), policent@central.ntua.gr (K. Politopoulos), didoy@central.ntua.gr (D. Yova).
[1] Tel.: +30 2107723048; fax: +30 2107723894.
[2] Tel.: +30 2107722283; fax: +30 2107723894.

0895-6111/$ – see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compmedimag.2007.07.005

tion. The calibration algorithm is based on the photogrammetric method [8]; the inspection of a 3D calibration object, for telecentric lenses, is analyzed. A novel image-processing algorithm is presented for detecting and labeling the structured light pattern features. The segmentation outcomes of this algorithm have been proven extremely accurate, even for features barely visible in the original stereo pair, due to low intensity levels. Finally a custom volume calculation method is introduced, based on finite pyramidal frustums. The developed system was tested and evaluated in 3D reconstruction and measurement of non-melanoma skin cancer (NMSC) animal model tumours, during successive photodynamic therapy (PDT) sessions.

The measurements during this work were preformed in vivo on SKH-HR1 albino hairless mice. The animal model treatment and all measurements were carried out according to the strict guidelines, established by the European Council Directive 86/909/EC and the Greek Committee for animal models.

2. System configuration

The configuration of the vision system was designed taking into account the type of cameras, the objects to be measured, the illumination, the targets and the flexibility of the system. Two 1/1.8 in. progressive scan digital color CCD cameras, with IEEE-1394a interface, were used (PointGrey Scorpion SCOR-20SO). With the specific type of cameras the acquisition of the images was automatically synchronized at the hardware level via the firewire clock pulses, since both cameras were connected on the same bus. The image size was 1600 pixels × 1200 pixels, with pixel size of 4.4 μm × 4.4 μm. Those cameras do not require external power supply, since the firewire port powers them.

Perspective error, also known as parallax, is part of the everyday experience in gauging distance; closer objects appear larger than objects that have the same size but are further away. While this is useful for estimating the distance of objects with known sizes, is most troublesome in measurement applications involving objects with depth, or objects moving relatively to the lens. Telecentric lenses are designed to minimize perspective error. They optically correct for perspective, and objects retain the same perceived size, independent of their location within a depth of field and field of view, defined by the lenses. Furthermore, with the telecentric lenses the acquired image is distortion free, since all distortion factors are optically compensated. Considering all these premises, two telecentric gauging lenses were selected, with magnification 0.29× and depth of field 8 mm, (Melles Griot Invarigon-R Trade Mark Telecentric Gauging Lenses) and connected to the cameras. Magnification of 0.29× provides the possibility to inspect object of maximum dimensions 14.7 mm × 19.8 mm.

One of the most important problems in machine vision systems with multiple views is the well-known correspondence problem [9]. In order to increase the accuracy of the matching between different stereo pair images, the system included a custom structured light projector. Illumination from a non-ultraviolet blue light source (dominant wavelength at 440 nm [10]) was delivered to a high quality printed film through a fiber bundle. Then, via a lenses system the film pattern was projected on the tumour-target. This way the number of successful matches increased sufficiently. The structured light pattern, Fig. 1, was consisted of a line matrix, with size 10 mm × 10 mm, while the diameter of the shapes, inside the line matrix, was 0.5 mm. Two types of shapes, circles and squares, were used to code the structured light pattern, thus the correspondence between each stereo pair was defined with increased accuracy.

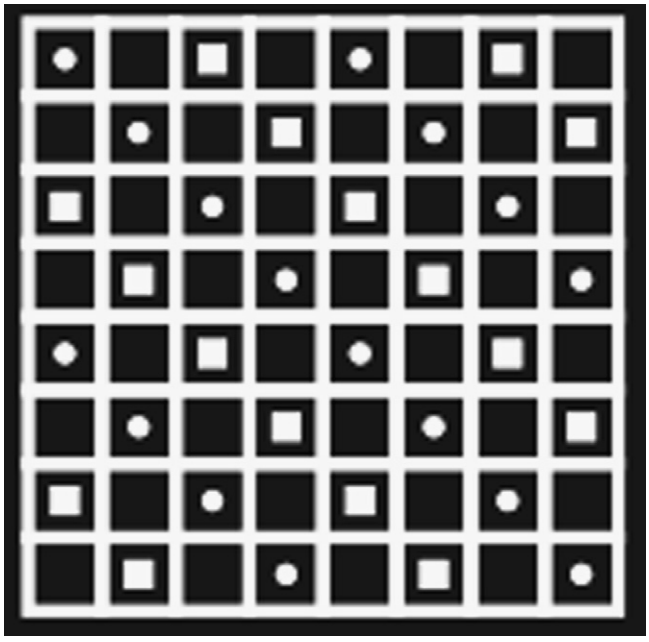

Fig. 1. The structured light pattern.

The object used for calibrating the system was a prism containing an equally spaced grid of circular tie-points (38 × 38) on the two perpendicular planes, Fig. 2. Each circle had 0.3 mm diameter, while one of them had 0.5 mm diameter, in order to correlate the acquired calibration stereo pair and to assign the same world coordinates for both cameras. The number of tie-points was chosen to be of that order, so that to cover the max-

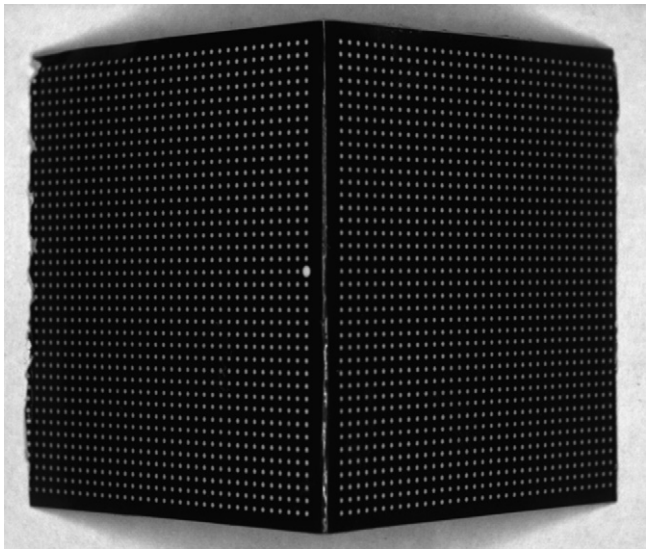

Fig. 2. The calibration object.

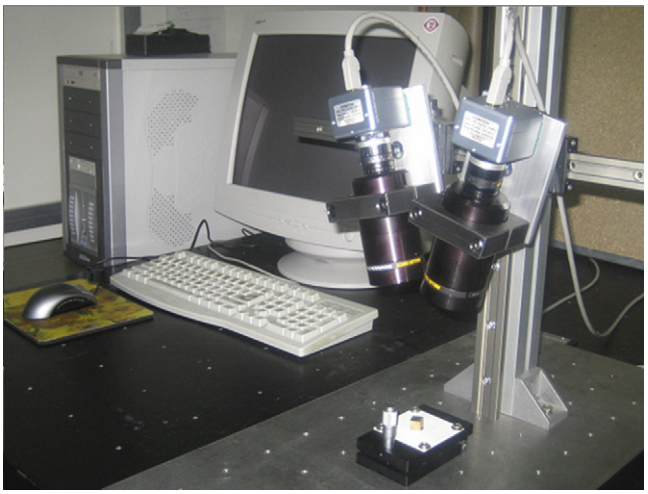

Fig. 3. The binocular machine vision system.

imum image space for more accurate calibration. Furthermore, the color of the tie-points was white, on a black background, minimizing calibration error factors due to reflectance.

The binocular machine vision system is presented in Fig. 3. It is obvious that both cameras can move independently along a horizontal axis and alter their view angle, so to achieve the highest levels of accuracy. In addition, both cameras can move vertically and by using the telecentric lenses there would be no need for further calibration after that movement. Two white light sources were also included to the system for the acquisition of colored stereo pairs.

3. Camera calibration

The cameras of the binocular machine vision system, although designed for high accuracy purposes, were non metric. That made the calibration of the system decisive for the accuracy of the gauging application.

Camera calibration in the context of 3D optical measurements is the process of determining the internal camera geometric and optical characteristics (intrinsic parameters) and/or the 3D position and orientation of the camera frame relative to a certain world coordinate system (extrinsic parameters) [11]. Extrinsic parameters are needed to transform object coordinates to a camera centered coordinate system. In many cases the overall performance of a machine vision system strongly depends on the accuracy of the camera calibration procedure [1,9,12–16].

Due to the purpose of this research, it is obvious that there was no need for real-time reconstruction and inspection of the target, thus the calibration could be accomplished beforehand using a calibration target. Moreover, what is important is measurement accuracy and image processing rapidity rather, than simple real-time target tracking. Under these premises, there was no need for adopting self-calibration of virtual targeting techniques [17–19], since photogrammetric calibration [11,14] can provide by far more accurate camera parameters than the other techniques [9].

3.1. Camera model

The most important premise for an accurate calibration is the correct mathematical expression of the camera model. By using telecentric lenses the pinhole model of the wide angle cameras [9,11,14], was not applicable. Telecentric lenses perform orthographic projection, thus the projection of an arbitrary point to the image plane in metrical units $(\tilde{u}_i, \tilde{v}_i)$ is expressed as:

$$\begin{bmatrix} \tilde{u}_i \\ \tilde{v}_i \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_i \\ y_i \\ z_i \\ 1 \end{bmatrix} \tag{1}$$

where the coordinates (x_i, y_i, z_i) is the transformation of the real world coordinates (X_i, Y_i, Z_i) of an arbitrary point to the camera centered system. This transformation is preformed through the expression:

$$\begin{bmatrix} x_i \\ y_i \\ z_i \\ 1 \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{bmatrix} \tag{2}$$

where $R = [r_{ij}]$ is the rotation matrix and $T = [\, t_x \quad t_z \quad t_z \,]^{\mathrm{T}}$ is the translation matrix. These two matrices define the extrinsic parameters of the camera frame, which are the 3D position in a global coordinate system.

The corresponding image coordinates (u'_i, v'_i) in pixels are obtained from the projection $(\tilde{u}_i, \tilde{v}_i)$ of the Eq. (1) by applying the following transformation:

$$\begin{bmatrix} u'_i \\ v'_i \\ 1 \end{bmatrix} = \begin{bmatrix} a_u & s & 0 \\ 0 & a_v & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \tilde{u}_i \\ \tilde{v}_i \\ 1 \end{bmatrix} \tag{3}$$

where the coefficients a_u, a_v and s are scale factors for transforming the image plane metrical units to image pixels and represent the camera intrinsic parameters. From Eq. (3) is obvious that for telecentric lenses there is not a principal point (u_0, v_0), as there is not defined a projection center (parallel projection).

Combining Eqs. (1)–(3), the full orthographic camera model is formed and expressed by the equation:

$$\begin{bmatrix} u'_i \\ v'_i \\ 1 \end{bmatrix} = \begin{bmatrix} a_u & s & 0 \\ 0 & a_v & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \times \cdot \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{bmatrix} \tag{4}$$

Although it has been already mentioned that telecentric lenses optically compensate for distortions, in fact there should be taken into account at least the most common types of distortion, to

avoid any possible lens construction errors, which would assign to the camera pinhole functionality.

The most widespread distortion factor is the radial lens distortion, which causes the actual image point to be displaced radially in the image plane. On the other hand, centers of curvature of lens surfaces are not always strictly collinear. This introduces another common distortion type, decentering distortion, which has both a radial and a tangential component. The overall distortion equation is expressed as follows:

$$\begin{bmatrix} \delta u_i \\ \delta v_i \end{bmatrix} = \begin{bmatrix} \tilde{u}_i \cdot r_i^2 & \tilde{u}_i \cdot r_i^4 & 2 \cdot \tilde{u}_i \cdot \tilde{v}_i & r_i^2 + 2 \cdot \tilde{u}_i^2 \\ \tilde{v}_i \cdot r_i^2 & \tilde{v}_i \cdot r_i^4 & r_i^2 + 2 \cdot \tilde{v}_i^2 & 2 \cdot \tilde{u}_i \cdot \tilde{v}_i \end{bmatrix} \cdot \begin{bmatrix} k_1 \\ k_2 \\ p_1 \\ p_2 \end{bmatrix} \tag{5}$$

where (k_1, k_2) are the radial distortion coefficients (typically, one or two coefficients are enough to compensate for this distortion type), (p_1, p_2) are the coefficients for tangential distortion and $r_i = \sqrt{\tilde{u}_i^2 + \tilde{v}_i^2}$ is the radial displacement [20,21].

Taking into consideration the Eq. (5) to the camera model, the set of intrinsic parameters (a_u, a_v, s) is augmented with the four distortion coefficients (k_1, k_2, p_1, p_2). These parameters are also known as physical camera parameters, since they bear a certain physical meaning. Generally, the objective of a camera calibration procedure is to determine optimal values for these parameters based on image observations of a known 3D target.

It is obvious, from Eq. (1) that the camera frame z-axis position does not affect the projection of the point (X_i, Y_i, Z_i) to the image plane, thus providing the ability of z-axis system movement without requiring further calibration. This fact is of great importance to applications where each target may bear different z-axis coordinate and increases the robustness of the described system, compared to any system with wide angle lenses.

3.2. Calibration algorithm

The adopted photogrammetric calibration algorithm is consisted of three steps. During the first step, initial values for the camera model parameters in Eq. (4) are estimated via direct linear transformation (DLT). At the second step, initial values of the distortion parameters in Eq. (5) are approximated using the results of the first step and applying again the DLT method. Finally, all the parameters values are refined, implementing an iterative optimization algorithm and using the outcomes of the two previous steps as initial values [9].

The DLT method is based on the orthographic camera model of Eq. (4), and it ignores the nonlinear radial and tangential distortion components. The first calibration step consists of two parts; the solution of the linear transformation from the object coordinates (X_i, Y_i, Z_i) to image coordinates (u'_i, v'_i) and the decomposition of the outcomes to obtain the camera parameters.

Eq. (4) can be transformed to:

$$w_i \cdot \begin{bmatrix} u'_i \\ v'_i \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} & h_{14} \\ h_{21} & h_{22} & h_{23} & h_{24} \\ h_{31} & h_{32} & h_{33} & h_{34} \end{bmatrix} \cdot \begin{bmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{bmatrix} \tag{6}$$

The parameters of the DLT matrix are solved by eliminating the arbitrary weight coefficient. Solving Eq. (6) and rearranging its components, two matrices arise:

$$L = \begin{bmatrix} X_1 & Y_1 & Z_1 & 1 & 0 & 0 & 0 & 0 & -X_1 \cdot u'_1 & -Y_1 \cdot u'_1 & -Z_1 \cdot u'_1 & -u'_1 \\ 0 & 0 & 0 & 0 & X_1 & Y_1 & Z_1 & 1 & -X_1 \cdot v'_1 & -Y_1 \cdot v'_1 & -Z_1 \cdot v'_1 & -v'_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ X_i & Y_i & Z_i & 1 & 0 & 0 & 0 & 0 & -X_i \cdot u'_i & -Y_i \cdot u'_i & -Z_i \cdot u'_i & -u'_i \\ 0 & 0 & 0 & 0 & X_i & Y_i & Z_i & 1 & -X_i \cdot v'_i & -Y_i \cdot v'_i & -Z_i \cdot v'_i & -v'_i \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ X_N & Y_N & Z_N & 1 & 0 & 0 & 0 & 0 & -X_N \cdot u'_N & -Y_N \cdot u'_N & -Z_N \cdot u'_N & -u'_N \\ 0 & 0 & 0 & 0 & X_N & Y_N & Z_N & 1 & -X_N \cdot v'_N & -Y_N \cdot v'_N & -Z_N \cdot v'_N & -v'_N \end{bmatrix}$$

and

$$h = \begin{bmatrix} h_{11} & h_{12} & h_{13} & h_{14} & h_{21} & h_{22} & h_{23} & h_{24} & h_{31} & h_{32} & h_{33} & h_{34} \end{bmatrix}^{\mathrm{T}}$$

and the following matrix equation for N control points is attained:

$$L \cdot h = 0 \tag{7}$$

By replacing the correct image points (u'_i, v'_i) with the observed values U_i, V_i the parameters $[h_{ij}]$ are estimated in a linear least squares fashion. In Eq. (7) is obvious the existence of the trivial solution $h = 0$. There are quite a few suggestions to overcome this problem in the case of the wide angle lens [22,23]. However, a closer inspection of the Eq. (4) reveals that $h_{31} = h_{32} = h_{33} = 0$ and $h_{34} = 1$, thus the use of telecentric lenses provides unique solution to the matrix Eq. (7) and different than the trivial one. It is obvious, though, that matrix L, since it contains metrical and pixel quantities, is badly scaled, so a proper normalization of the data is needed for the solution to be accurate enough.

The second part of the linear parameter estimation is the extraction of camera physical parameters from the solution h. A modified RQ decomposition has been applied to the DLT matrix. Furthermore, RQ correction has also been applied, in case scale factors a_u and a_v are calculated negatively signed [24].

The initial values of the distortion coefficients, at the second step of the calibration algorithm, are estimated in the same manner Eq. (4) was solved. Eq. (5) obviously is linear to the distortion coefficients, thus using the previously estimated parameters this equation is transformed to the:

$$\begin{bmatrix} \tilde{u}_i - U_i \\ \tilde{v}_i - V_i \end{bmatrix} = \begin{bmatrix} \tilde{u}_i \cdot r_i^2 & \tilde{u}_i \cdot r_i^4 & 2 \cdot \tilde{u}_i \cdot \tilde{v}_i & r_i^2 + 2 \cdot \tilde{u}_i^2 \\ \tilde{v}_i \cdot r_i^2 & \tilde{v}_i \cdot r_i^4 & r_i^2 + 2 \cdot \tilde{v}_i^2 & 2 \cdot \tilde{u}_i \cdot \tilde{v}_i \end{bmatrix} \times \cdot \begin{bmatrix} k_1 \\ k_2 \\ p_1 \\ p_2 \end{bmatrix} \quad (8)$$

which is solved for (k_1, k_2, p_1, p_2) in a linear least squares approximation.

Since no iterations are required, direct methods are computationally fast. By using telecentric lenses the solutions of the linear methods provide very good precision, fastening the nonlinear optimization for camera parameters.

With real cameras the image observations are always contaminated by noise. There are various error components incorporated in the measurement process. If the systematic parts of the measurement error are compensated for, it is convenient to assume that the error is white Gaussian noise. Then, the best estimate for the camera parameters can be obtained by minimizing the residual between the model and N bservations (U_i, V_i). In the case of Gaussian noise, the objective function is expressed as a sum of squared residuals:

$$F = \sum_{i=1}^{N} (U_i - u_i)^2 + \sum_{i=1}^{N} (V_i - v_i)^2 \quad (9)$$

The least squares estimation technique was used to minimize Eq. (9). Due to the nonlinear nature of the camera model, simultaneous estimation of the parameters involves application of an iterative algorithm. For this problem the Levenberg–Marquardt [25] optimization method provides the fastest and more accurate convergence. However, without proper initial parameters values the optimization may stick in a local minimum and thereby cause the calibration to fail. This problem has been avoided by using the parameters from the DLT method as the initial values for the optimization.

4. Image acquisition

Besides the accurate stereo cameras calibration, another crucial factor for the accuracy of the binocular machine vision system is the instantly synchronized image acquisition. In the case of a static scene the time difference between capturing the two images would not affect the gauging accuracy of the system, but in the case of animal models, motion due to breath rate is more than observable. This fact led to the need of highly synchronized stereo pair image acquisition.

The used cameras, as mentioned previously, had the advantage of synchronization at hardware level, when two or more cameras are connected on the same firewire bus. This was implemented via the firewire clock pulses which instantly activated all the cameras that are connected on this bus.

An algorithm that activates the cameras in video mode and via the clock pulses stores a predefined number of stereo pair images was developed. It was useful to acquire more than one stereo pair, since animal model breath rate may lead to blurred image capturing due to motion, a fact that would decrease the accuracy of the system. Acquisition of ten stereo images provided at least two to three motion free pairs. The implementation of the image acquisition algorithm was based on the current cameras function library provided by the constructor.

5. Three-dimensional reconstruction

One of the major problems in multiple vision systems is the well known "matching" problem, which is the determination of conjugate points between stereo pair images. This problem issues from the inherited difficulty in determining, automatically, features from a series of images, without changing any algorithm parameter (i.e. threshold values, intensity limits).

Successful 3D reconstruction from photogrammetric stereo pairs is strongly depended on the existence of conjugate points. Those points are pixels bearing exactly the same information in both images. The more the conjugate points, the higher is the accuracy of a stereo vision system. Therefore, two are the premises for accurate 3D reconstruction, as far as the stereo pair images are concerned. Firstly is the existence of a great number of conjugate points, which will lead to more detailed representation, and secondly is the detection only of the true conjugate points, minimizing registration errors.

5.1. Image preprocessing

Instead of processing the acquired RGB images, only the blue channel was used, decreasing the computational cost significantly, while all the details were preserved. The choice of the blue channel was based on the fact that the system was using a blue light source for the structured light. An example of a blue channel stereo pair is shown in Fig. 4.

However, each image if the stereo pair had different intensity levels, due to the different view angles of the cameras. In order both images to have equal grey scale levels, intensity adjustment would be the first thing to perform, otherwise image processing algorithms that apply to one image would fail to the other.

The intensity adjustment was performed by mapping the intensity values in each image to new values, such that values between a low threshold and a high threshold mapped to values between 0 (black) and 1 (white). Values below low and above high thresholds were clipped; that is, values below the low threshold mapped to 0 and those above the high threshold mapped to 1. For the specific protocol of image acquisition, the appropriate mapping was the one that was weighted toward higher (brighter) output values. This way, all the structured light elements were included in the adjusted images. The new stereo pair retained all the details as the input images, but in addition they had very similar intensity levels, Fig. 5.

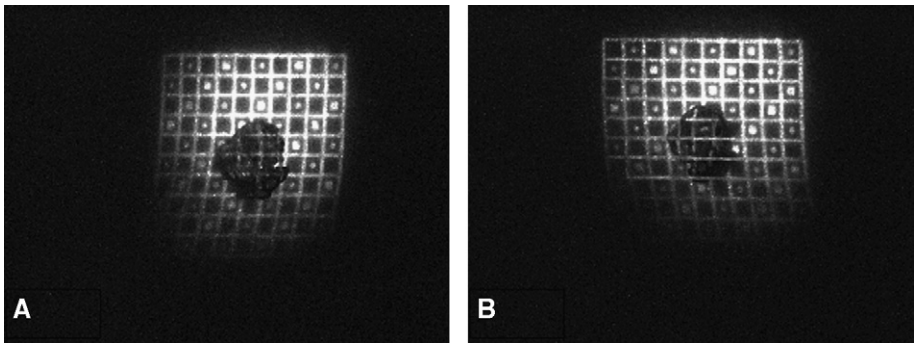

Fig. 4. Grey level (blue channel) stereo pair images. (A) Image acquired from the left camera. (B) Image acquired from the right camera.

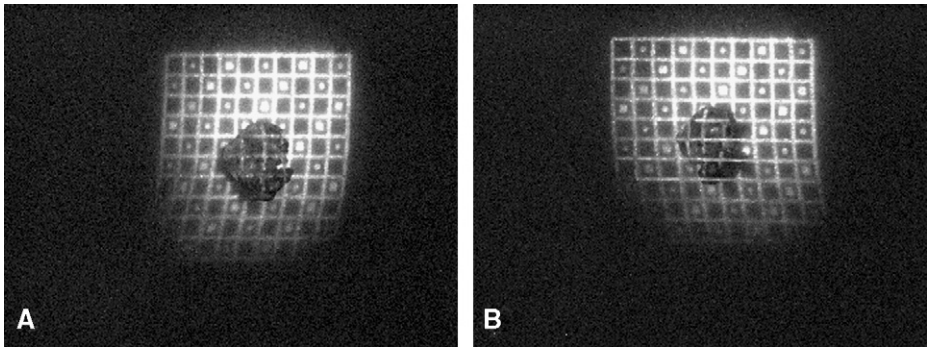

Fig. 5. Stereo pair images after intensity adjustment. (A) Image acquired from the left camera. (B) Image acquired from the right camera.

An expected side-effect of the intensity adjustment is the amplification of the inherited noise, something that is more than obvious in Fig. 5. The two images were restored by implementing the Wiener filtering [26–28]. A Wiener filter seeks an estimate $\hat{f}$ that minimizes the statistical error function expressed by the following equation:

$$e^2 = E\{(f - \hat{f})^2\} \tag{10}$$

where E is the expected value operator and f the undegraded image. This filter in its complete form requires the knowledge of a good deal about the signal and the properties of the noise that infects it. By considering the inherited image noise to be Gaussian the resulted restored stereo pair images are shown in Fig. 6.

The difference between images in Figs. 5 and 6 is rather obvious. The stereo pair in Fig. 6 was restored and it retained all the details in comparison with the original image. Trying to extract the structured light pattern, segmenting the stereo pair, would result in leaving image features undetected. The reason is that the intensity of some features is quite smaller than the intensity due to diffuse reflectance on the animal model skin tumour.

However, the described algorithm overcame this difficulty via morphological image processing. More specific, opening filter can produce a reasonable estimate of the background across the image, as long as the selected structuring element is large enough not to fit entirely within any of the pattern elements. Subtracting the resulted image from the restored, a new one with reasonable even background arise (top-hat transformation). Repeating the same procedure, but instead of opening applying

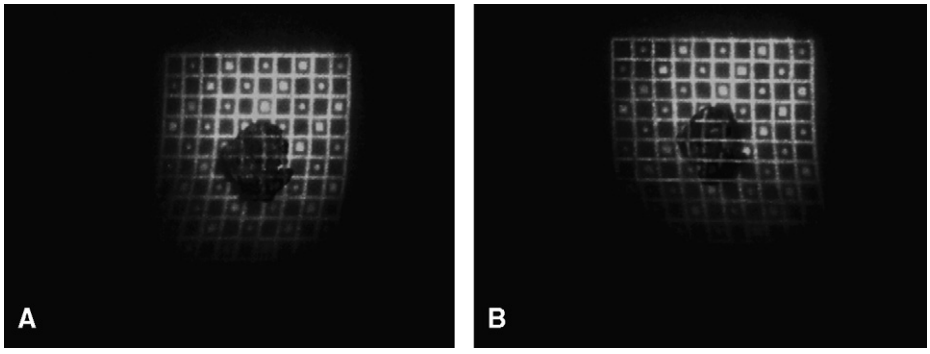

Fig. 6. Restored grey level stereo pair images. (A) Image acquired from the left camera. (B) Image acquired from the right camera.

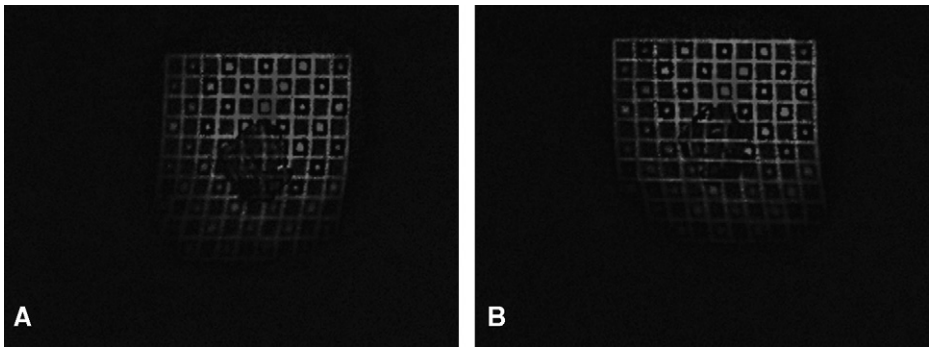

Fig. 7. Stereo pair images with contrast enhancement. (A) Image acquired from the left camera. (B) Image acquired from the right camera.

the closing filter (bottom-hat transformation), and multiplying the outcomes, a contrast enhancement was succeeded, as shown in Fig. 7.

5.2. Image segmentation

After the preprocessing procedure the stereo pair images were ready for the detection of the pattern elements. Simple thresholding would lead either to the exclusion of some elements or to the region growth of others. On the other hand, application of any edge detection or segmentation algorithm would lead to falsely detected edges or over-segmentation. The reason was that the pattern elements did not have sharp edges or relevant intensities. In order to avoid such problems, this algorithm performed a combination of marker-controlled watershed transformation [29,30] and thresholding. As a result even the lowest intensity features were detected. The resulted segmented images are presented in Fig. 8.

5.3. Stereo matching

When the entire pattern shapes had been detected on the two images, the correspondences between shapes were estimated by using feature properties, such as eccentricity. A very useful tool for this was also the codification of the structured light pattern, that is the sequence of the pattern shapes. After segmentation all the pattern shapes were labeled uniquely, providing instant matching between the two images.

Since the correspondences between features were known, the determination of the conjugate points was achieved. The determination was based on object properties (relative rotation, distance from centroid, etc.) and maxima correlation values, but instead of searching the whole image or applying epipolar line constrains, the search was constrained only to the corresponding features. This fact made the algorithm significantly faster and accurate than area matching algorithms.

5.4. 3D coordinates calculation

According to Eq. (2) an arbitrary real world point (X_i, Y_i, Z_i) was transformed to the left camera coordinates system as:

$$\begin{bmatrix} x_i^L \\ y_i^L \\ z_i^L \end{bmatrix} = R_L \cdot \begin{bmatrix} X_i \\ Y_i \\ Z_i \end{bmatrix} + T_L \tag{11}$$

where the indicator L represents the left camera and $[R_L|T_L]$ are the rotation and translation matrices of that camera.

On the other hand, the same real world point is transformed to the right camera coordinates system, using again Eq. (2), as:

$$\begin{bmatrix} x_i^R \\ y_i^R \\ z_i^R \end{bmatrix} = R_R \cdot \begin{bmatrix} X_i \\ Y_i \\ Z_i \end{bmatrix} + T_R \tag{12}$$

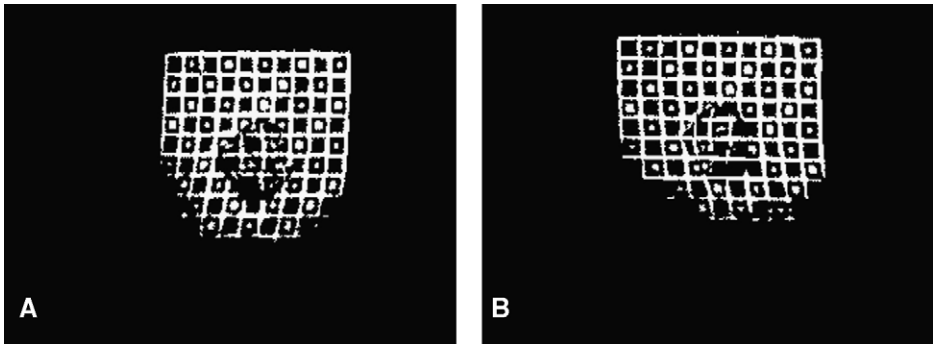

Fig. 8. The segmented stereo pair. (A) Image acquired from the left camera. (B) Image acquired from the right camera.

Combining Eqs. (11) and (12) and solving for the real-world point (X_i, Y_i, Z_i), the following equations resulted:

$$\begin{bmatrix} x_i^L \\ y_i^L \\ z_i^L \end{bmatrix} = R_L \cdot R_R^{-1} \cdot \begin{bmatrix} x_i^R \\ y_i^R \\ z_i^R \end{bmatrix} - R_L \cdot R_R^{-1} \cdot T_R + T_L \Rightarrow \begin{bmatrix} x_i^L \\ y_i^L \\ z_i^L \end{bmatrix}$$

$$= R \cdot \begin{bmatrix} x_i^R \\ y_i^R \\ z_i^R \end{bmatrix} + T \tag{13}$$

where $[R|T]$ are the rotation and translation matrices that transform the left camera coordinates system to the right camera coordinates system.

From the calibration procedure all rotation and translation matrices are known, thus the unknowns for the system in Eq. (13) are the coordinates (x_i^L, y_i^L, z_i^L) and (x_i^R, y_i^R, z_i^R). Having more unknowns than equations makes the system unsolvable.

However, it is also known that the plane coordinates (x_i^L, y_i^L) are related to the left image coordinates in pixels with the equation:

$$\begin{bmatrix} u_i^L \\ v_i^L \end{bmatrix} = \begin{bmatrix} a_u^L & s^L \\ 0 & a_v^L \end{bmatrix} \cdot \begin{bmatrix} x_i^L \\ y_i^L \end{bmatrix} + \begin{bmatrix} \delta u^L \\ \delta v^L \end{bmatrix} \tag{14}$$

according to Eqs. (1), (3) and (5). From this equation the coordinates (x_i^L, y_i^L) for the point (X_i, Y_i, Z_i) are calculated, since all other parameters are known either from calibration procedure (intrinsic parameters) or from image information (detected points). With the same technique the coordinates (x_i^R, y_i^R) are also estimated.

Now Eq. (13) can be solved for the coordinates z_i^L and z_i^R, thus providing the 3D coordinates of the real-world point (X_i, Y_i, Z_i), which was detected to both images.

5.5. Interpolation and geometrical characteristics

One disadvantage of the stereo vision is that the 3D coordinates are known only for those points recognized as conjugate. For the volume calculation, information about the whole surface is essential, thus an interpolation scheme is used. Delaunay triangulation [1] provided the most accurate 3D interpolation, since it converges towards the real 3D surface with the increase of the points with known 3D coordinates. This technique is a pure geometrical method and its result for the stereo pair of Figs. 4–8 is presented in Fig. 9.

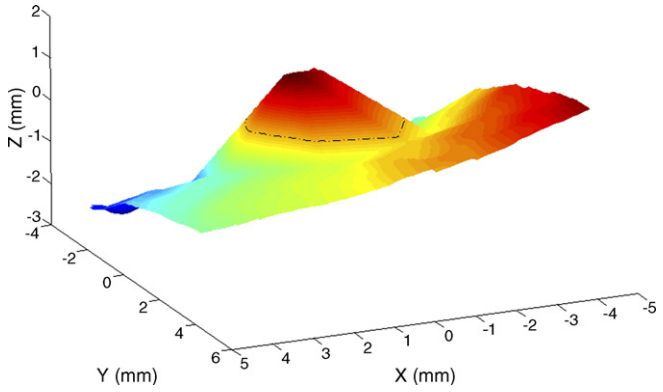

Fig. 9. 3D reconstruction of skin tumour.

Table 1
Geometrical characteristics calculated by the binocular machine vision system

Feature	Description
Volume	The volume of the reconstructed object
Height	The height of the reconstructed object
Area	The area of the reconstructed object base region
Perimeter	The perimeter of the reconstructed object base region
Equivalent diameter	The diameter of a circle having the same area as the base region of the reconstructed object
Major axis	The length of the major axis of the ellipse that has the same normalized second central moments as the base region of the reconstructed object
Minor axis	The length of the minor axis of the ellipse that has the same normalized second central moments as the base region of the reconstructed object

However, trying to calculate volume directly from the triangulation outcomes will provide significant error, due to the surrounding area curvature. To overcome this issue, the contours of the reconstructed shape are used. Examining the distribution of the contours of the reconstructed area, arouses that only the tumour contours were closed, while the contours of the surrounding area were not.

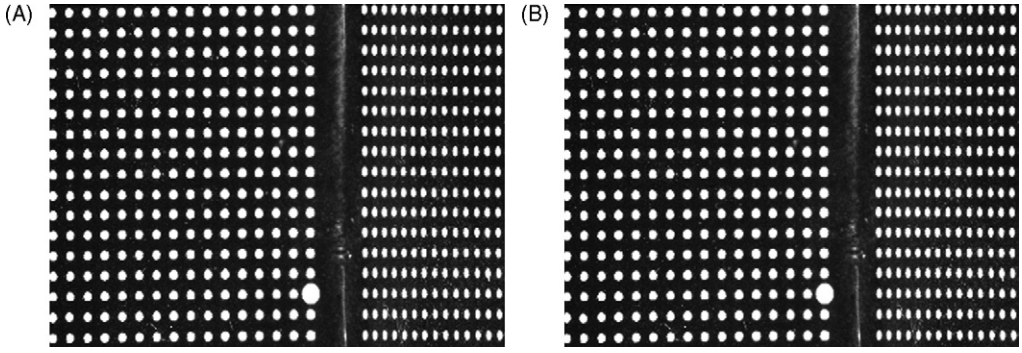

Fig. 10. The calibration object stereo pair. (A) Image acquired from the left camera. (B) Image acquired from the right camera.

In case that a great number of contours are considered, two sequential contours shape a very narrow pyramidal frustum. The finite volume of that frustum can be calculated by the following formula:

$$dv = \frac{1}{3}\, dh(A_1 + A_2 + \sqrt{A_1 \cdot A_2}) \quad (15)$$

where dh is the height of the frustum and A_i are the two base areas. Adding the finite volumes of all the frustums, the volume of the tumour is calculated.

The geometrical characteristics that are finally calculated are summarized in Table 1.

6. Experimental results

6.1. Camera calibration results

Before the application of the calibration algorithm, it is obvious that a stereo pair of the calibration object should be captured. In Fig. 10 this stereo pair is presented.

The coordinates of the calibration object in the real-world coordinates system are portrayed in Fig. 11.

By detecting the tie-point with the larger diameter, correspondence between the real world and the acquired images will be possible, since the real position of that point is known. The detection is performed by segmenting the input images and using the geometrical properties of the image features. After this procedure, all the tie-points are labeled with their corresponding real-world coordinates and the calibration algorithm is applied. In Figs. 12 and 13, the results of the calibration are shown for both cameras, after performing back-projection to the detected tie-points and using the calibration outcomes.

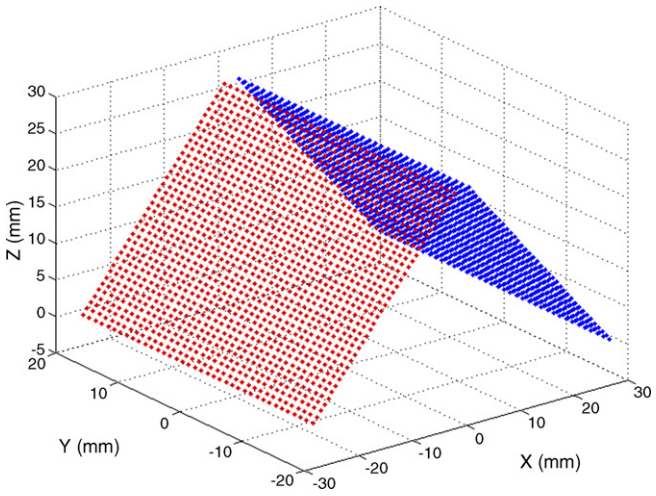

Fig. 11. Coordinates of the calibration object in the real world.

The accuracy of the calibration algorithm is approximately 0.04 mm, and this quantity is limited only by the construction of the object.

6.2. Application in tumour recording after PDT treatments

Photodynamic therapy experiments were performed in non-melanoma skin cancer (NMSC) animal model tumours. Three PDT sessions were preformed in each animal, once every seven days. The final evaluation of the tumour diminishing was performed fourteen days after the third PDT treatment. In Figs. 14–18, the 3D reconstruction of NMSC tumours after each PDT treatment for therapy and drug evaluation are presented.

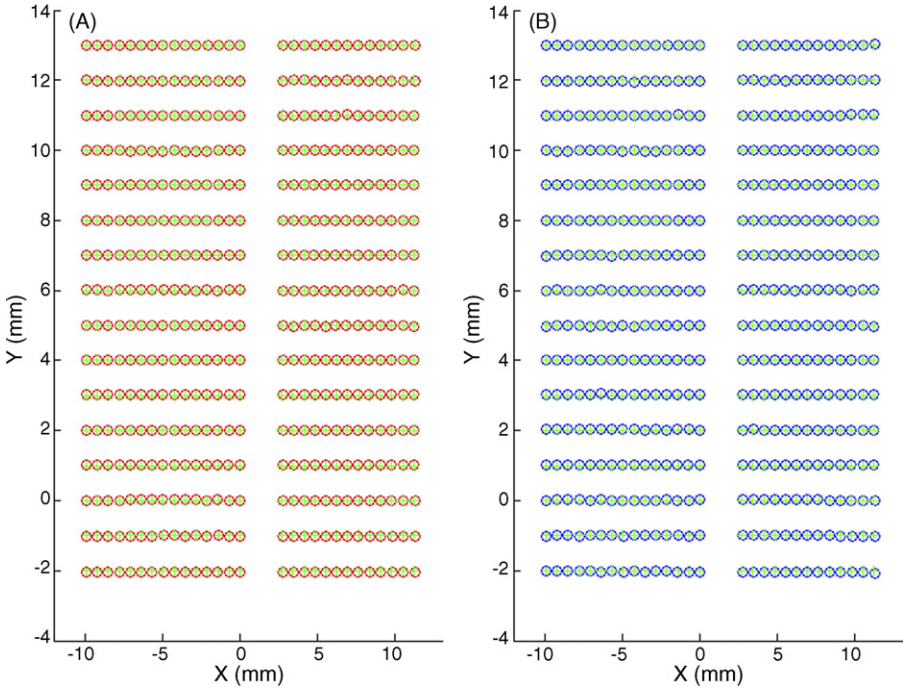

Fig. 12. The coordinates of the real world (+) and the calculated by back-projection (o) for XY plane. (A) For the left camera. (B) For the right camera.

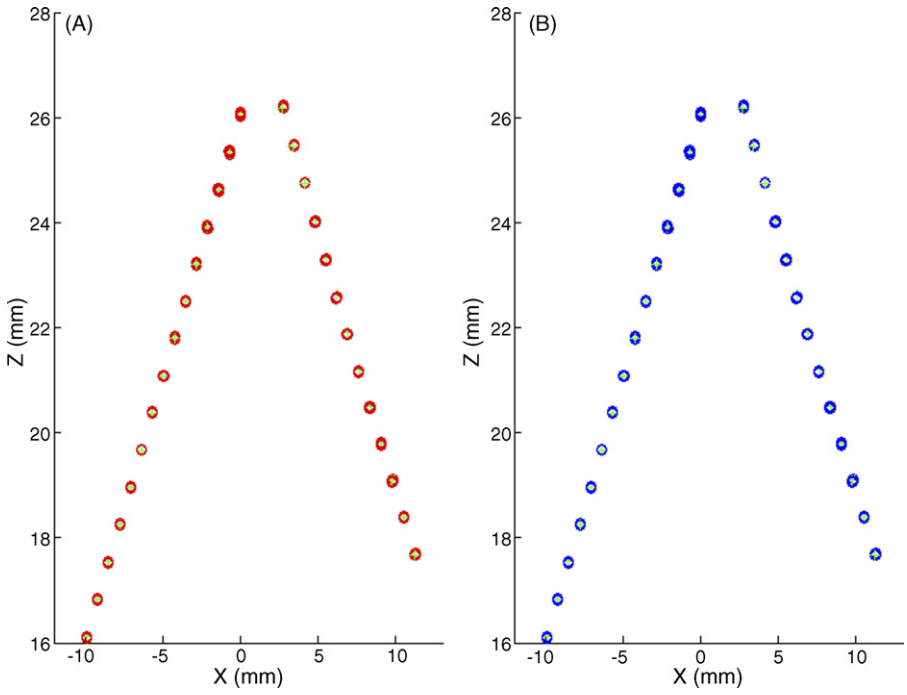

Fig. 13. The coordinates of the real world (+) and the calculated by back-projection (o) for XZ plane. (A) For the left camera. (B) For the right camera.

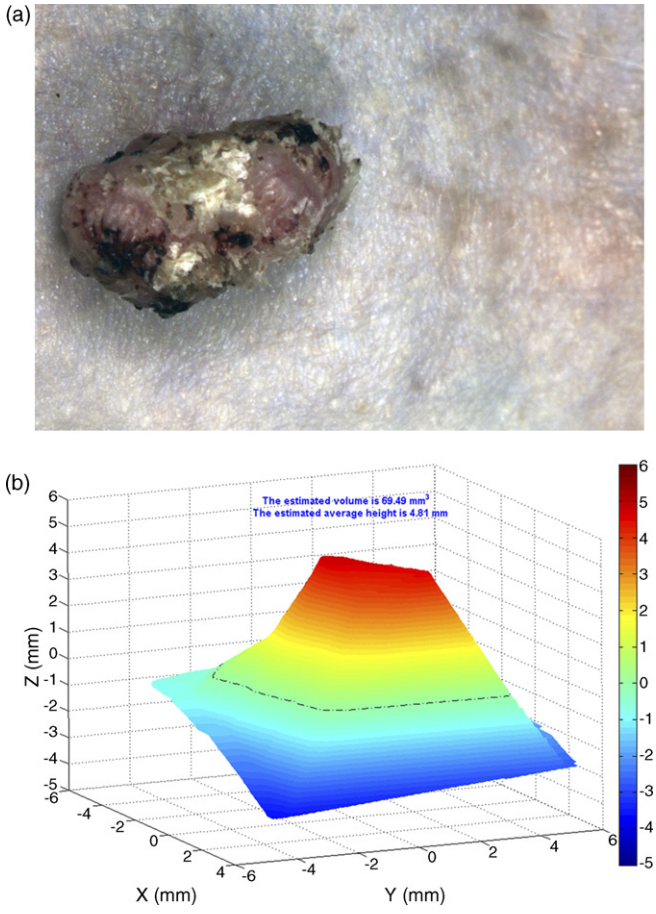

Fig. 14. Before first PDT.

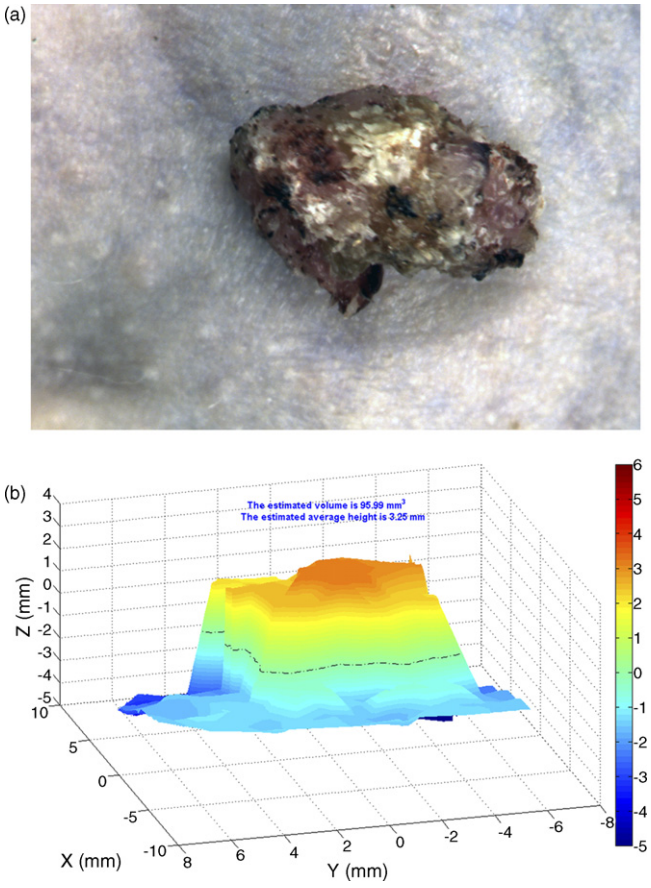

Fig. 15. Seven days after first PDT.

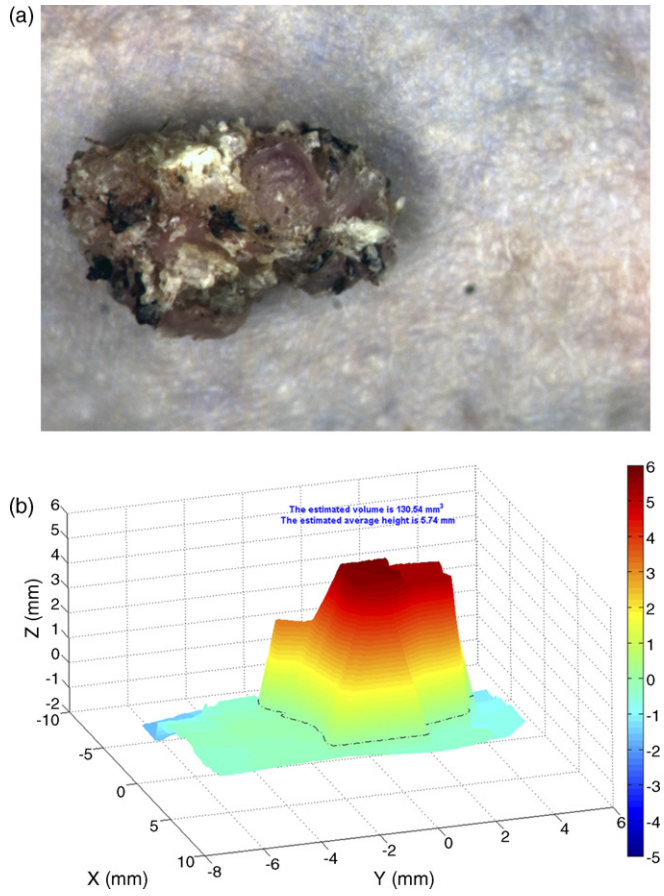

Fig. 16. Seven days after second PDT.

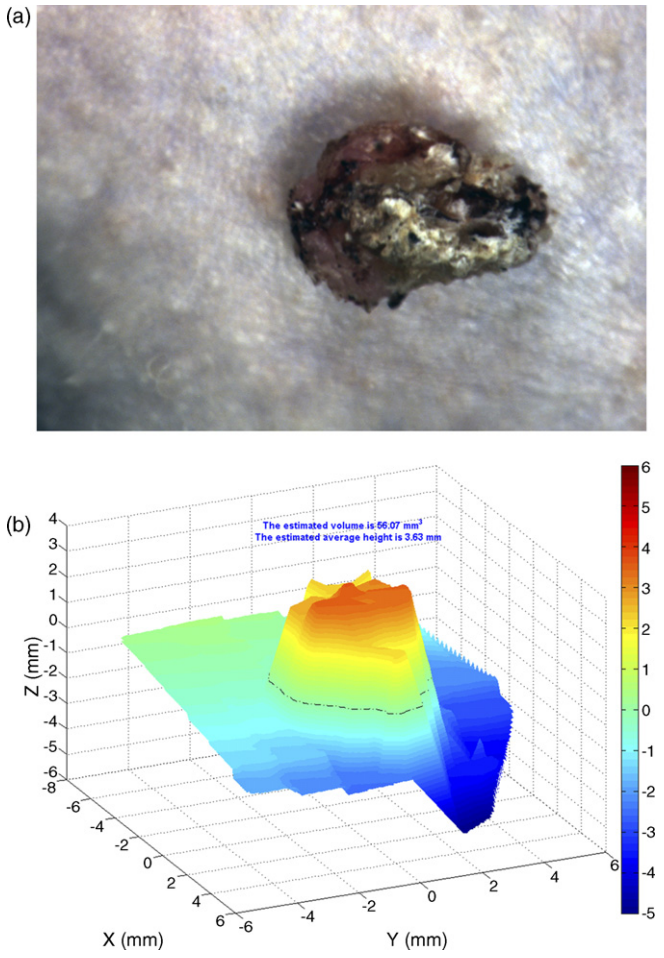

Fig. 17. Seven days after third PDT.

7. Discussion

The configuration of the vision system was designed taking into account the type of cameras, the objects to be measured, the illumination, the flexibility and the cost. The use of telecentric lenses compensated for perspective error, while provided the very important z-axis movement of the system without the need of further calibration.

The feature detection algorithm succeeds to recognise structured light pattern features almost invisible due to very low intensity levels and it fully compensates noise due to skin and tumour diffuse reflections. The normalization of intensity levels prevents the algorithm from failing among different stereo pairs, leading to fully automated operation. This is very important, especially for non computer vision experts. The matching algorithm is an object based algorithm, providing better matches in discontinuous scenes, such as NMSC tumours, than area matching algorithms, since it matches objects visible from both cameras [1]. Another important characteristic of the developed method is also the automated recognition of the tumour from the non planar background, by using contours of the 3D reconstruction.

The robustness of this system issues from the lab prototype gauging imaging system with the crucial z-axis movement and the fully automated volumetric algorithms.

A novel volumetric method for accurate small objects 3D reconstruction and measurement was presented in this paper. This method can be a very useful tool for industrial quality control applications. Modern machineries require the best size accuracy of their particles for optimal functionality and the binocular system can provide the level of dimension accuracy during quality controls [31].

Another field where the binocular machine vision system would provide valuable assistance is Archeology, where volume and surface geometry measurements of small items are frequently essential [8]. In biology there is also the need to observe and study the growth rates of various small species, like insects. Due to the absolute synchronization of image acquisition, the developed system can measure the growth rates of those fast-moving specimens and provide all the necessary geometrical features.

In addition to these applications, this system is the first CCD-based monitoring system providing 3D tumour imaging of very small skin tumours and monitoring their vanishing between PDT treatments. However, the developed system could have remarkable impact in dermatology in other applications as well. Burn depth assessment is one of the great and unsolved problems. Burn wound is still assessed via visual inspection in most cases, while there are some efforts to assess the grade by infrared imag-

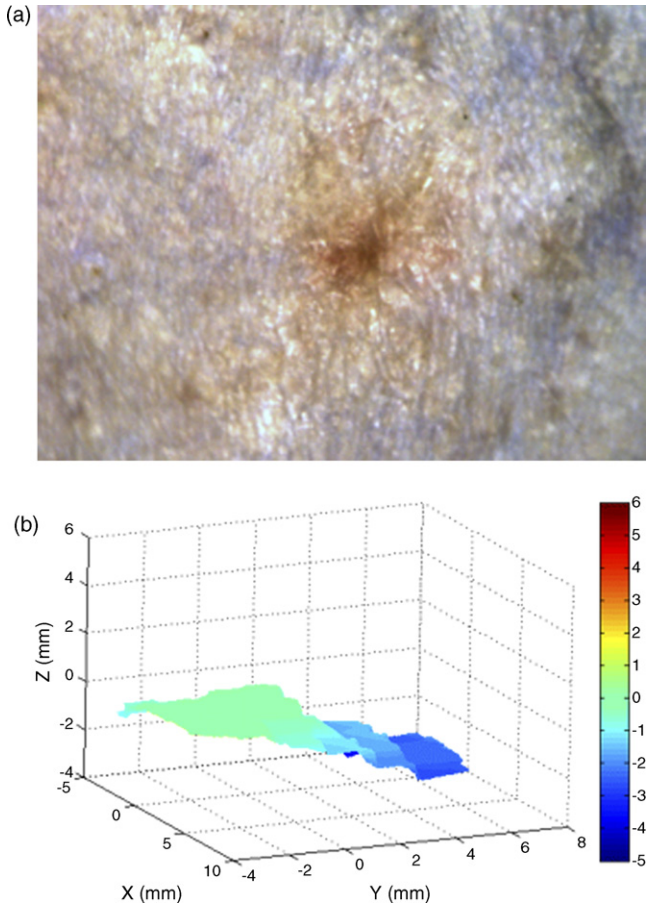

Fig. 18. Fourteen days after third PDT.

ing [32]. However, there is intense scientific interest in the depth estimation via optical imaging [33]. This system can push those efforts a bit forward and provide valuable information towards the burn depth estimation. Another biomedical scientific field that the system can be used is the pre- and post-clinical surgery planning. It can provide accurate volumetric measurements of small skin lesions and assist medical doctors plan a successful operation.

8. Summary

Rendering three-dimensional information of a scene from optical measurements is very important for a wide variety of applications. Although there are still no fully automatic commercial systems available, the problem of 3D reconstruction of relatively large objects has been solved. Despite the recent advances in computer vision, close-range photogrammetry of very small objects has not yet been encountered.

The hardware configuration was strictly based on the nature of the problem, thus emerged into highest efficiency during image acquisition. Telecentric lenses provided the z-axis movement, so the system can achieve optimal focus, without further calibration, free from perspective error and distortion factors, this being a major advantage over the commonly used wide-angle lenses. In addition stereo pair image acquisition was absolutely synchronized at hardware level, minimizing reconstruction errors of fast moving objects, as in the case of breath rating of tumour bearing animal models.

Fully automated algorithms with increased accuracy on the 3D reconstruction have been developed. More specific, a preprocessing procedure, which included intensity adjustment, noise reduction and contrast enhancement algorithms, was applied to each one of the stereo pair images. This procedure, apart from highlighting the structured light features, it also ensured the automated function of the algorithms. At the second image-processing phase, the features of the structured light pattern were detected via a combination of modified watershed algorithm and thresholding, and they were uniquely labeled. That led to the successful confrontation of the well-known "matching" problem. The use and the successful detection of a custom coded structured light pattern provided a large number of conjugate points, which made the three-dimensional reconstruction possible for any gauging application involving objects with diameter less than 1 cm, even if inherited object features, such as edges or color variations, were absent. Furthermore, a custom mathematical algorithm, based on surface contours and eccentricity alignment, has been applied in order to extract the object under inspection from the surrounding area, after the 3D reconstruction. This extraction is vital in gauging applications, where the possible curvature of the surrounding surface may be included in the geometrical characteristics calculation, decreasing the accuracy of the system.

The developed system is a very useful tool for industrial quality control measurements, where accuracy and inspection are very important. Most of the modern machineries consist of a large number of very small parts and the need for automated three-dimensional reconstruction algorithms is still urgent to control the successful construction of those parts. Another field where the Binocular Machine Vision System would provide valuable assistance is Archeology, where volume and surface geometry measurements of small items are essential. In biology there is also the need to observe and study the growth rates of various small species, like insects. Due to the absolute synchronization of image acquisition the developed system can measure the growth rates of those fast moving specimens and provide all the necessary geometrical features. In addition to these applications, this system is the first CCD-based monitoring system providing 3D tumour imaging of very small skin tumours and monitoring their vanishing between photodynamic treatments.

Acknowledgements

This research was supported by the EU Integrated Project "Wide Wavelength light for public Welfare: High-Brightness Laser Diode Systems for Health, Telecom and Environment Use", IST – 511722-2003.

References

[1] E.M. Mikhail, J.S. Bethel, J.C. McGlone, Introduction to Modern Photogrammetry, John Wiley and Sons, Inc., 2001.
[2] Kucukkaya AG. Photogrammetry and remote sensing in archeology. J Quant Spectrosc Radiat Transfer 2004;88:83–8.

[3] Brenner C. Building reconstruction from images and laser scanning. Int J Appl Earth Obs Geoinfo 2005;6:187–98.
[4] Bösemann W. Advances in photogrammetric measurement solutions. J Comput Ind 2005;56:886–93.
[5] Zhang G, Wei Z, Li X. 3D double-vision inspection based on structured light. ASME J Manuf Sci Eng 2003;125:617–23.
[6] Huq S, Abidi B, Goshtasby A, Abidi MA. Stereo matching with energy-minimizing snake grid for 3D face modeling, biometric technology for human identification. Proc SPIE 2004;5404:339–50.
[7] Ansari AN, Abdel-Mottaleb M. Automatic facial feature extraction and 3D face modeling using two orthogonal views with application to 3D face recognition. Pattern Recogn 2005;38:2549–63.
[8] G. Medioni, S.B. Kang, Emerging Topics in Computer Vision, Pearson Prentice Hall, New Jersey, USA, 2005.
[9] O. Faugeras, Three-dimensional Computer Vision: A Geometric Viewpoint, The MIT Press, Fourth Printing, 2001.
[10] Elli A. Understanding the Color of Human Skin. In: SPIE Proceedings, Conference on Human Vision and Electronic Imaging. 2001. p. 243–51.
[11] Tsai RY. A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses. IEEE J Robot Automat 1987;RA-3:323–44.
[12] Horn BKP. Closed-form solution of absolute orientation using unit quaternions. J Opt Soc Am A 1987;4:629–42.
[13] Horn BKP. Relative orientation revisited. J Opt Soc Am A 1991;8:1630–8.
[14] Horn BKP. Relative orientation revisited. J Opt Soc Am 1991;A8:1630–8.
[15] Horn BKP, Hilden HM, Negahdaripour S. Closed-form solution of absolute orientation using orthonormal matrices. J Opt Soc 1988;5:1127–35.
[16] Micheals RJ, Boult TE. On the Robustness of Absolute Orientation. In: Proceeding of the International Association for Science and Technology Development (IASTED) Conference on Robotics and Automation. 2000.
[17] Dornaika F, Chung R. An algebraic approach to camera self-calibration. Comput Vis Image Und 2001;83:195–215.
[18] X. Chen, J. Davis, P. Slusallek, Wide area camera calibration using virtual calibration object, in: Proceedings of the IEEE Comp. Soc. Conf. on Computer Vision and Pattern Recognition Proc., CVPR '00, 2000, pp. 2520–2527.
[19] Ihrke I, Ahrenberg L, Magnor M. External camera calibration for synchronized multi-video systems. J WSCG 2004;12:537–44.
[20] Fryer JG, Clarke TA, Chen J. Lens distortion for simple C-mount lenses. Int Arch Photogrammetry Remote Sens 1994;30:97–101.
[21] Ahmed MT, Farag AA. Differential methods for nonmetric calibration of camera lens distortion. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR'01, 2. 2001. p. 477–82.
[22] Abdel-Aziz YI, Karara HM. Direct linear transformation into object space coordinates in close-range photogrammetry. In: Proceediong of the Symposium on Close-Range Photogrammetry. 1971. p. 1–18.
[23] O.D. Faugeras, G. Toscani, Camera Calibration for 3D Computer Vision, in: Proceedings of the International Workshop on Industrial Applications of Machine Vision and Machine Intelligence, 1987, pp. 240–247.
[24] P.G. Ciarlet, Introduction to Numerical Linear Algebra and Optimization, Cambridge University Press, Cambridge, 1995.
[25] More JJ. The Levenberg–Marquardt algorithm: implementation and theory. Numerical Analysis, Lecture Notes in Mathematics, vol. 630. 1977. p. 105–16.
[26] J.R. Parker, Algorithms for Image Processing and Computer Vision, Wiley Computer Publishing, 1997, John Wiley and Sons, Inc., New York.
[27] N. Wiener, Cybernetics, II Ed., MIT Press, 1961.
[28] Murli A, D'Amore L, De Simmone V. The Wiener Filter and Regularization Methods for Image Restoration Problems. In: Proceedings of the 10th International Conference on Image Analysis and Processing. 1999. p. 394–9.
[29] Beucher S. The watershed transformation applied to image segmentation. Scan Microsc Int 1992;Suppl. 6:299–314.
[30] Meyer F, Maragos P. Multiscale morphological segmentations based on watershed, flooding, and eikonal PDE. In: Proceedings of the International Conference on Scale-Space Theories in Computer Vision (SCALE-SPACE'99, Lecture Notes in Computer Science, vol. 1682). 1999. p. 351–62.
[31] Aguilar JJ, Lope M, Torres F, Blesa A. Development of a stereo vision system for non-contact railway concrete sleepers measurement based in holographic optical elements. Measurement 2005;38:154–65.
[32] Hargroder AG, Davidson JE, Luther DG, Head JF. Infrared imaging of burn wounds to determine burn depth. SPIE Proc 1999;3698:103–8.
[33] Devan L, Bhat S, Aylward S, Spence RJ. Modalities for the assessment of burn wound depth. J Burns Wounds 2006;5:7–15.

Dimitris Gorpas received his BEng degree in Electrical and Computer Engineering from the National Technological University of Athens (NTUA), Greece, in 2003 and since then he is performing his doctoral studies in Biomedical Optics. His research interests include biomedical optics, computer vision and image analysis mainly in the field of biomedical applications.

Kostas Politopoulos, PhD, is a Lecturer of Biomedical Optics and Applied Biophysics, at the School of Electrical and Computers Engineering. His research interests include design and development of analog digital and optical systems, development of computer programs for data processing of biomedical data and signals and development of laser optical systems for biomedical purposes.

Dido Yova, PhD, is a Professor of Biomedical Optics and Applied Biophysics, Director of Biomedical Optics Lab and Chairman of Applied Electromagnetism, ElectroOptics & Electronic Materials Department, at the School of Electrical and Computers Engineering of NTUA. She has also been a Harvard researcher, from 1991 to 1995 engaged in research at the Wellman Laboratory of Photomedicine and Photobiology. Her main research fields for more than 14 years have been in the fields of Fluorescence Tissue Diagnostics, Photodynamic Therapy, Fluorescence Imaging microscopy, Molecular imaging and recently in Computer vision. She has been involved in many projects and has been an expert evaluator in FP6 Calls.