OPTIMAL CALIBRATION OF A PRISM-BASED VIDEOENDOSCOPIC SYSTEM FOR PRECISE 3D MEASUREMENTS

A.V. Gorevoy [1, 2, 3], *A.S. Machikhin* [1, 3]

[1] *Scientific and Technological Center of Unique Instrumentation, Russian Academy of Sciences, Moscow, Russia,*
[2] *Bauman Moscow State Technical University, Moscow, Russia,*
[3] *Moscow Power Engineering University, Moscow, Russia*

Abstract

Modern videoendoscopes are capable of performing precise three-dimensional (3D) measurements of hard-to-reach elements. An attachable prism-based stereo adapter allows one to register images from two different viewpoints using a single sensor and apply stereoscopic methods. The key condition for achieving high measurement accuracy is the optimal choice of a mathematical model for calibration and 3D reconstruction procedures. In this paper, the conventional pinhole camera models with polynomial distortion approximation were analyzed and compared to the ray tracing model based on the vector form of Snell's law. We, first, conducted a series of experiments using an industrial videoendoscope and utilized the criteria based on the measurement error of a segment length to evaluate the mathematical models considered. The experimental results confirmed a theoretical conclusion that the ray tracing model outperforms the pinhole models in a wide range of working distances. The results may be useful for the development of new stereoscopic measurement tools and algorithms for remote visual inspection in industrial and medical applications.

Introduction

Measurement technologies are now transforming videoendoscopes from pure visual inspection tools to precise equipment for characterization of hard-to-reach elements. With the measurement technologies available today, remote visual inspection is complementing or even replacing other non-destructive testing methods in the rocket and aircraft engineering, oil and gas industry and other fields [1, 2]. The mostly used approach to videoendoscope-based 3D measurements is the utilization of a miniature prism-based optics and matching two images of a surface from two different points [3]. This stereoscopic technique employs a calibration procedure and processing algorithms to compute a 3D coordinate for every matched pixel prior to the start of the measurement process resulting in a full 3D surface map of the inspected object [4 – 6]. The data processing pipeline of a 3D measurement endoscopic system with a prism-based stereo adapter is shown in Fig. 1.

Many researchers have proposed their understandings and implementation methods for prism-based single-lens stereovision systems in the last two decades. Lee and Kweon [7] considered an arbitrary point in 3D space as two virtual points after the refraction of a prism and derived 3D reconstruction technique for this kind of systems. Lim and Xiao [8] described that one image acquired by the system could be divided into two halves considered as captured by two virtual cameras. This concept converted the prism-based stereovision system to the conventional stereoscopic system using two cameras, which made possible to use conventional camera calibration, image processing and 3D reconstruction methods. Furthermore, this concept was extended from two-ocular to multiocular system [9]. This approach was also used to implement epipolar lines calculation method [10], to analyze the influence of prism's angle and position on a common field of view for the virtual cameras [11] and to estimate depth measurement error caused by errors of the prism and camera parameters estimation [12]. However, they made significant assumptions about the relative camera to prism position and analyzed refraction in 2D space only.
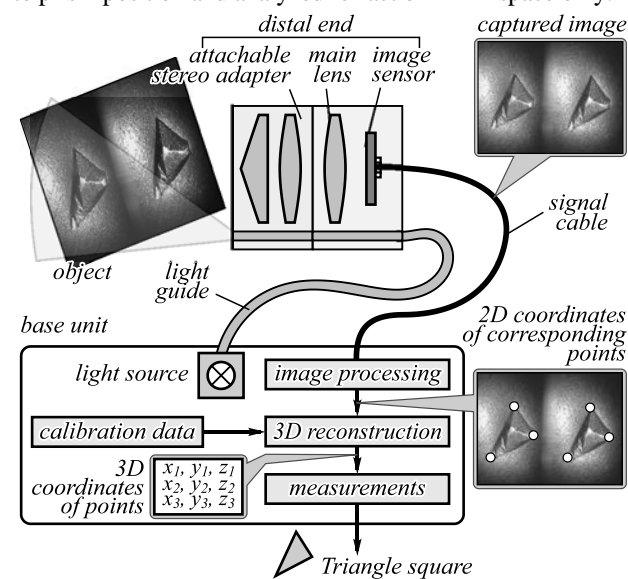

Fig. 1. The data processing pipeline of a 3D measurement endoscopy system

Cui et al. [13] pointed out the limitations of the 2D approach and proposed an accurate geometrical optics model of the prism-based stereovision system using the 3D vector form of Snell's law. It made possible to derive a transformation matrix which could express the relationship between an object point and its image as 3D affine transformation. This matrix could be composed with other transformation matrices describing rotation, translation and projection in traditional projective geometry to build epipolar lines and 3D reconstruction. They noted that this matrix is different for each image point and the prism-based stereovision system can not be represented with pure projective camera models with high precision. Wu et al. [14] used similar approach for the digital image correlation system. They implemented a modified virtual point model based on backward ray tracing from the image plane to the object space considering the refraction on prism faces and adapted it for bilateral telecentric lens later [15].

In all reviewed papers it was supposed that the camera parameters could be either estimated through a standard calibration procedure [16] or known from a technical documentation and that the position and the parameters of the prism were known exactly. In contrast, Cui et al. [17] proposed a calibration technique to estimate multiocular prism position after a separate camera calibration. They used transformation matrices from [13] and made several assumptions about prism to reduce the number of parameters.

In this paper, we introduce a classification and analysis of mathematical models for a prism-based single-lens stereovision system based on a considered model of an optical distortion induced by a prism. The first group [8 – 12] includes pure projective pinhole models ignoring prism distortion. Nevertheless, the concept of virtual camera allows to use any camera model with distortion to calibrate virtual cameras as independent ones. Following this way, Cui et al. [18] approximated scaling factor for every image point in [13] by radial distortion. Lim and Qian [19] made Taylor series analysis of the equations for ray refraction and offered the extended polynomial model to undistort acquired images. These models can be unified as the second group containing pinhole camera models with polynomial distortion. The third group [13 – 15, 17] consists of the accurate geometrical optics models derived from the 3D vector form of Snell's law. Finally, we may refuse global distortion model and use model-free local distortion correction as shown in [20]. Furthermore, the same approach can be used in 3D object space to correct the result of 3D measurements directly or to undistort images. Genovese et al. [21] used an additional complicated equipment and the second camera without a prism to measure deviations of 3D coordinates for thousands of control points in the working volume to obtain correction tables.

However, an endoscopic system with a prism-based stereo adapter has some peculiarities. First of all, the main lens has a very short focal distance and a wider field of view, the prism angles are steeper and the prism is placed close to the lens. These factors lead to significantly bigger image distortions than shown in the most of reviewed papers. The lack of techniques allowing to assess 3D measurement error and the

applicability of the proposed mathematical models for different camera and prism parameters makes it problematic to apply the results for large-scale prism-based systems to endoscopes. Moreover, the stereo adapter contains an additional lens. We consider the prism parameters and position to be unknown and to be estimated during a calibration procedure. Hence, we can not apply techniques based on known prism parameters as well as independent calibration of camera without a prism and estimation of prism parameters afterwards.

The patent [22] describing the stereo-measurement borescope included the description of calibration procedure using a single image of the calibration tool manufactured as a flat gridded base with a high step. It contains the distortion correction equations generated from the distorted positions of the grid points on the image, but didn't provide further details. Another patent [23] offered to apply polynomial distortion model to the main lens and to the prism with additional lens separately. Hence, we can refer it to the second group of our classification. Takata et al. [24] developed distortion correction method for images captured by endoscope with two wedge prisms using backward ray tracing similar to [14]. Unfortunately, none of these papers presented any results to estimate the uncertainty of 3D measurements which can be achieved for the endoscope with stereo adapter.

The results of computer simulation in our previous work [25] have shown that the pinhole camera model with polynomial distortion could not be equal to the ray tracing model based on the 3D vector form of Snell's law for the typical industrial endoscope with a prism-based stereoscopic adapter. The reason is that pinhole model doesn't consider a ray shift and non-homocentric beams after refraction. Additionally, the best criteria for industrial endoscope evaluation is based on the measurement uncertainties of length or area. These uncertainties vary significantly depending on position and direction inside the working space [25, 26] which makes a problem to choose an optimal mathematical model for these systems challenging.

This paper presents theoretical analysis and experimental evaluation of the mathematical models for the industrial endoscopes with the prism-based adapter considering the peculiarities of their optical systems and the specificity of the application. We introduce the basic description of three models and the calibration procedure; present the experimental verification of each model effectiveness and the comparison of the models.

Mathematical models

The general geometric model of the image formation may be written in the form $\mathbf{p}_i = P_i \circ E_i(\mathbf{l}_w)$, where $\mathbf{l}_w$ is the vector of 3D line coordinates in the world coordinate system (WCS) and $\mathbf{p}_i$ is the vector of 2D coordinates for the corresponding point on the image plane of the i-th camera, $i = 1...N$, N – number of cameras [4, 5, 26]. The vector of line coordinates $\mathbf{l}_w$ includes 3D coordinates of the origin point $\mathbf{c}_w$ and the direction vector $\mathbf{v}_w$: $\mathbf{l}_w = \left(\mathbf{c}_w^T, \mathbf{v}_w^T \right)^T$. We use E_i to denote Euclidean mapping

$\mathbf{x}_i = E_i(\mathbf{x}_w) = \mathbf{R}_i\mathbf{x}_w + \mathbf{t}_i$ translating 3D point $\mathbf{x}_w$ from the WCS to the coordinate system (CS) of i-th camera ($\mathbf{R}_i$ is the rotation matrix and $\mathbf{t}_i$ is the translation vector). Hence, $\mathbf{l}_i = E_i(\mathbf{l}_w) = (\mathbf{c}_i^T, \mathbf{v}_i^T)^T$ with $\mathbf{c}_i = \mathbf{R}_i\mathbf{c}_w + \mathbf{t}_i$, $\mathbf{v}_i = \mathbf{R}_i\mathbf{v}_w$. The mapping P_i determines the unique correspondence $\mathbf{p}_i = P_i(\mathbf{l}_i)$ between the ray $\mathbf{l}_i$ in the i-th camera object space and the point $\mathbf{p}_i$ in the image plane. We use «∘» notation here to define a composition of transformations, i.e. $P_i{\circ}E_i(\mathbf{l}_w) \equiv P_i(E_i(\mathbf{l}_w))$. The set of transformations E_i and P_i is parametrized by the vector $\mathbf{k}$ of intrinsic and extrinsic camera parameters, which components are to be determined during the system calibration. According to [8–12, 19], the prism-based stereoscopic imager can be considered as two virtual cameras and the same notation can be used.

Pinhole camera models. The pinhole camera model is widely used in computer vision and is well described [4, 5, 8, 16]. It assumes that all rays $\mathbf{l}_i$ for the i-th camera pass through the central point, i.e. $\mathbf{l}_i = ((0,0,0)^T, \mathbf{v}_i^T)^T$. Hence, it is possible to use one point $\mathbf{x}_i$ for each ray and write $\mathbf{x}_i$ instead of $\mathbf{l}_i$ in the equations for the pinhole models (see Fig. 2). Using the same approach as [5, 27], we consider the mapping P_i as the combination of simple transformations

$$\mathbf{p}_i = P_i \circ E_i(\mathbf{x}_w) = A_i \circ D_i \circ F \circ E_i(\mathbf{x}_w), \qquad (1)$$

where F is the central projection onto the unit plane $\mathbf{x}_i' = F(\mathbf{x}_i)$; D_i is the 2D distortion transformation $\mathbf{x}_i'' = D_i(\mathbf{x}_i')$ and A_i is the 2D affine transformation to the image coordinates in pixels $\mathbf{p}_i = A_i(\mathbf{x}_i'')$ [4, 5].
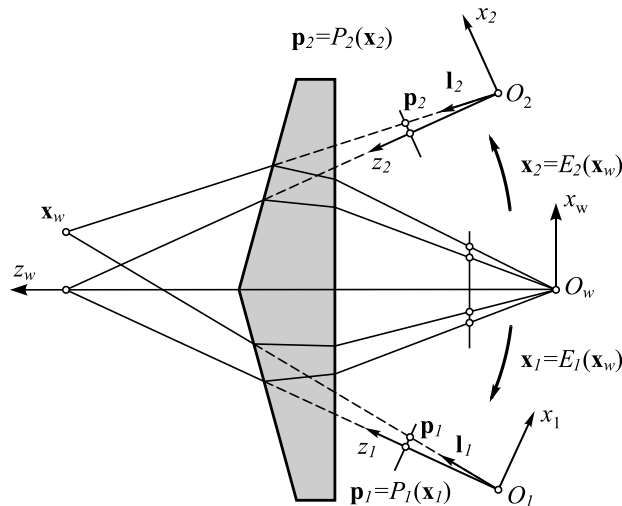

Fig. 2. A prism-based stereoscopic imager considered as two virtual pinhole cameras

One can find many possible approximations for the lens distortion, based on polynomial [16, 19, 23, 27, 28] or more complex [5] mathematical models. The widely used model (introduced by Brown [28]) represents distortion as the combination of radial (3rd, 5th and 7th orders) and tangential part to consider lens decentering as follows

$$\begin{pmatrix} x'' \\ y'' \end{pmatrix} = \left(1 + k_1 r^2 + k_2 r^4 + k_3 r^6\right)\begin{pmatrix} x_c' \\ y_c' \end{pmatrix} + \\ + \begin{pmatrix} \rho_1\left(r^2 + 2x_c'^2\right) + 2\rho_2 x_c' y_c' \\ 2\rho_1 x_c' y_c' + \rho_2\left(r^2 + 2y_c'^2\right) \end{pmatrix} + \begin{pmatrix} x_0' \\ y_0' \end{pmatrix}, \qquad (2)$$

where $r^2 = x_c'^2 + y_c'^2$, $x_c' = x' - x_0'$, $y_c' = y' - y_0'$ and (x_0', y_0') is the distortion center of the image, which is usually assumed to coincide with the center of projection; $k_1, k_2, k_3, \rho_1, \rho_2$ are the distortion parameters. We further refer to this model as "model 1". This model requires 10 parameters to describe each camera (5 parameters for A_i and 5 parameters for D_i) and 6 parameters for relative orientation of two cameras [4, 16]. As a result, vector $\mathbf{k}$ includes 26 parameters. We can extend the polynomial distortion model using the generic equation

$$\begin{pmatrix} x'' \\ y'' \end{pmatrix} = \sum_{m=0}^{B-1}\sum_{n=0}^{B-1}\begin{pmatrix} kx_{m,n} \\ ky_{m,n} \end{pmatrix} x_c'^m y_c'^n + \begin{pmatrix} x_0' \\ y_0' \end{pmatrix}, \qquad (3)$$

where coefficients $kx_{m,n}$ and $ky_{m,n}$ are the distortion parameters ($2B^2$ totally). We further refer to the extended polynomial model as "model 2". Generally, this model requires $2(5+2B^2)+6 = 16+4B^2$ parameters, but the number of the distortion coefficients can be reduced significantly based on theoretical analysis or computer simulation of the system typical parameters [19, 25] to simplify optimization procedures on the calibration stage.

In order to allow reconstruction of a ray in 3D object space for each image point, camera model should be invertible and should provide the back-projection transformation $\mathbf{l}_w = E_i^{-1} \circ P_i^{-1}(\mathbf{p}_i)$. Hence, each of the transformations in Eq. (1) should have a corresponding inverse one: F^{-1}, D^{-1} and A^{-1}. It is essential that polynomial models described by Eqs. (2) and (3) do not allow finding a closed-form solution for the inverse transformation D^{-1}, an iterative solver should be used in this case [5]. Thus, D^{-1} notation actually stands for this iterative procedure.

Ray tracing models. The alternative model using a ray tracing from the image plane to the object space is formulated similar to [14, 24]. We use $\mathbf{l}_{k,i} = S_{k,i}(\mathbf{l}_{k-1,i})$ notation to describe the refraction of each ray $\mathbf{l}_{k,i}$ on the k-th surface for the i-th image part. According to the vector form of Snell's law, we can derive transformation $S_{k,i}$ as

$$\mathbf{v}_{k,i} = b_{k,i}\mathbf{v}_{k-1,i} + \left(\sqrt{1 - b_{k,i}^2\left(1 - g_{k,i}^2\right)} - b_{k,i}g_{k,i}\right)\mathbf{s}_{k,i},$$

$$\mathbf{c}_{k,i} = \mathbf{c}_{k-1,i} - \frac{(\mathbf{c}_{k-1,i} - \mathbf{d}_{k,i})\cdot\mathbf{s}_{k,i}}{g_{k,i}}\mathbf{v}_{k-1,i}, \qquad (4)$$

$$b_{k,i} = \frac{n_{k-1,i}}{n_{k,i}}, g_{k,i} = \mathbf{v}_{k-1,i}\cdot\mathbf{s}_{k,i},$$

where $n_{k,i}$ is the refractive index between the k-th and the $(k+1)$-th surfaces, $\mathbf{s}_{k,i}$ is the normal to the k-th surface and $\mathbf{d}_{k,i}$ is the 3D coordinates of any point on the k-th surface (see Fig. 3). In order to find the ray coordinates $\mathbf{l}_{2,i}$ in the object space for each image pixel $\mathbf{p}_i$, we should

define ray coordinates $\mathbf{l}_{0,i}$. We consider the pinhole model with distortion for the main lens and use Eqs. (1) and (2) to find $\mathbf{l}_{0,i} = P_i^{-1}(\mathbf{p}_i)$. The inverse transformation P_i^{-1} depends only on the parameters of the image sensor and the main lens, it is the same for both image parts. Similar to Eq. (1), the complex inverse transformation for the ray tracing model can be represented in the form

$$\mathbf{l}_w = E(\mathbf{l}_{2,i}) = E^{-1} \circ S_{2,i} \circ S_{1,i} \circ F^{-1} \circ D^{-1} \circ A^{-1}(\mathbf{p}_i), \quad (5)$$

where Euclidean mapping E is the same for both channels due to our choice of the same CS for the left and the right parts of the prism, vectors $\mathbf{s}_{k,i}$ and $\mathbf{d}_{k,i}$ are determined in this CS.
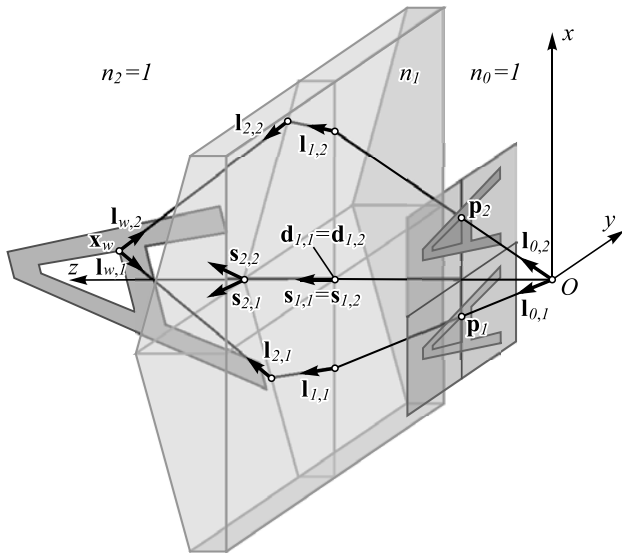

Fig. 3. Ray tracing through a prism

We assume that $\mathbf{s}_{k,i}$ is the unit vector and the point $\mathbf{d}_{k,i}$ is on the z-axis, so we can use 3 parameters for each prism surface. Accordingly, we have 9 parameters for 3 surfaces and one more parameter for the refractive index. Again, if we let E to be the unitary transformation, we need the vector $\mathbf{k}$ of 20 parameters to describe this model. In contrast to the pinhole model, the ray-tracing model cannot provide closed-form solution for the forward transformation $\mathbf{p}_i = P_i \circ E(\mathbf{l}_w)$ because it requires initially unknown direction vector of the line $\mathbf{l}_w$ from 3D point $\mathbf{x}_w$ (shown in Fig. 3). This problem is usually solved by the iterative technique called ray aiming or by look-up-table interpolation. We can formally write the forward transformation as $\mathbf{p}_i = P_i \circ E(\mathbf{x}_w)$ and extend it the same way as Eq. (5), if we consider that this notation stands for the iterative solver. We further refer to the ray tracing model as "model 3".

3D reconstruction problem. The reconstruction of 3D point coordinates $\mathbf{x}_w$ from N projections $\mathbf{p}_i$ may be considered as the ray intersection problem. Because of the uncertainties in coordinates $\mathbf{p}_i$, these rays are skew, and the triangulation algorithm T is used for the estimation of $\hat{\mathbf{x}}_w$ by minimizing a functional C, i.e.

$$\hat{\mathbf{x}}_w = T(\mathbf{p}, \mathbf{k}) = \underset{\hat{\mathbf{x}}_w}{\operatorname{argmin}}(C(\hat{\mathbf{x}}_w, \mathbf{p}, \mathbf{k})), \quad (6)$$

where $\mathbf{p} = (\mathbf{p}_1^T, \mathbf{p}_2^T, ..., \mathbf{p}_N^T)^T$. The selection of the appropriate cost functions is not trivial in the general case, because it depends on a priori data about the position of target point $\mathbf{x}_w$, its projections $\mathbf{p}_i$, the statistics of coordinate measurement errors and the properties of camera transformations P_i. If we assume that the deviation of the measured image coordinates $\mathbf{p}_i$ from their true values $\overline{\mathbf{p}}_i$ follows the Gaussian distribution with zero mean and covariation matrix Σ_{pi}, then we can use the Maximum Likelihood Estimator, which minimizes the Mahalanobis distance in the image plane [4, 29]

$$C(\hat{\mathbf{x}}_w, \mathbf{p}, \mathbf{k}) = \sum_{i=1}^{N} \left\| \mathbf{p}_i - P_i \circ E_i(\hat{\mathbf{x}}_w) \right\|_{\Sigma}^2 =$$
$$= \sum_{i=1}^{N} (\mathbf{p}_i - \hat{\mathbf{p}}_i)^T \Sigma_{pi}^{-1} (\mathbf{p}_i - \hat{\mathbf{p}}_i), \quad (7)$$

where $\hat{\mathbf{p}}_i = P_i \circ E_i(\hat{\mathbf{x}}_w)$ is the estimated position of the 3D point $\hat{\mathbf{x}}_w$ projection on the image plane of the i-th camera, and Σ_{pi}^{-1} is the inverse (rank-constrained generalized inverse) covariation matrix of the coordinate measurement error for $\mathbf{p}_i$. We assume that the measurements errors are independent for each image. The cost function (7) and its derivatives are calculated at each iteration, so it is sufficient to have a closed-form solution for the forward transformation. Hence, we can use this cost function for models 1 and 2. Alternatively, we can formulate the cost function based on the sum of distances from the 3D point $\hat{\mathbf{x}}_w$ to the rays $\mathbf{l}$, for example, we can use the mid-point of the common perpendicular for two skew rays [4]. This alternative cost function does not allow to achieve the theoretical bound for minimal variance (the Cramer-Rao lower bound) [29], but it is more suitable for the models that can faster calculate the inverse transformation than the forward one, such as model 3, though we can also use it for models 1 and 2.

Calibration problem. The aim of the calibration procedure is to determine the value of parameter vector $\mathbf{k}$ using different types of calibration targets (such as boards and corners [16] or steps [22]). We consider that we register images of the calibration target with M points in R positions; 3D coordinates of the points $\mathbf{x}_{tj}$ are known with a certain accuracy in the CS of the calibration target. Next, the image coordinates $\mathbf{p}_{i,j,k}$ are calculated for each point $j = 1...M$ and each position $k = 1...R$. The complete transformation for these points is $\mathbf{p}_{i,j,k} = P_i \circ E_i \circ E_k'(\mathbf{x}_{tj})$, where E_k' stands for the Euclidean mapping from the CS of the calibration target to the WCS. In addition to the parameter vector $\mathbf{k}$ for transformations P_i and E_i, we also need to find the vector $\mathbf{k}_t$ describing transformations E_k' for each k. If we introduce the composite vector $\mathbf{x}_t$ of all 3D points and the composite vector $\mathbf{p}$ of all projections, the calibration algorithm K may be written as

$$\hat{\mathbf{k}} = K(\mathbf{x}_t, \mathbf{p}) = \underset{\hat{\mathbf{k}}, \hat{\mathbf{k}}_t}{\operatorname{argmin}}(C(\mathbf{x}_t, \mathbf{p}, \hat{\mathbf{k}}, \hat{\mathbf{k}}_t)). \quad (8)$$

It is not necessary for each point at each position to be projected for each camera, some points or positions may be skipped. Following the same reasons as for the triangulation algorithm above, the most appropriate choice of the cost function C is the Mahalanobis distance in the image plane. The

equation for this cost function is similar to Eq. (7), but we should write the sum for all values i, j and k, i.e.

$$C\left(\mathbf{x}_t, \mathbf{p}, \hat{\mathbf{k}}, \hat{\mathbf{k}}_t\right) = \sum_{i=1}^{N}\sum_{j=1}^{M}\sum_{k=1}^{R}\left\|\mathbf{p}_{i,j,k} - \hat{P}_i \circ \hat{E}_i \circ \hat{E}'_k\left(\mathbf{x}_{tj}\right)\right\|_{\Sigma}^2 =$$

$$= \sum_{i,j,k}\left(\mathbf{p}_{i,j,k} - \hat{\mathbf{p}}_{i,j,k}\right)^T \Sigma_{\mathrm{p}\,i,j,k}^{-1}\left(\mathbf{p}_{i,j,k} - \hat{\mathbf{p}}_{i,j,k}\right). \quad (9)$$

Again, we can use this cost function for models 1 and 2 because they provide the closed-form solution for the forward transformation. For the model 3 we should derive the alternative cost function based on the sum of distances from 3D point $\mathbf{x}_t$ to rays $\hat{\mathbf{l}}_t$. If we use a flat calibration target, we can replace rays $\hat{\mathbf{l}}_{t\,i,j,k}$ by points $\hat{\mathbf{x}}_{t\,i,j,k}$ of ray intersections with XOY plane in the CS of the target. In that case, the cost function can be represented as follows

$$C\left(\mathbf{x}_t, \mathbf{p}, \hat{\mathbf{k}}, \hat{\mathbf{k}}_t\right) = \sum_{i,j,k}\left\|\mathbf{x}_{tj} - \hat{E}'^{-1}_k \circ \hat{E}_i^{-1} \circ \hat{P}_i^{-1}\left(\mathbf{p}_{i,j,k}\right)\right\|_{\Sigma}^2 =$$

$$= \sum_{i,j,k}\left(\mathbf{x}_{tj} - \hat{\mathbf{x}}_{t\,i,j,k}\right)^T \Sigma_{\mathrm{x}\,i,j,k}^{-1}\left(\mathbf{x}_{tj} - \hat{\mathbf{x}}_{t\,i,j,k}\right). \quad (10)$$

where $\Sigma_{\mathrm{x}\,i,j,k}^{-1}$ is the inverse covariation matrix of calibration target coordinate measurement error for $\mathbf{x}_{t,i,j,k}$ (see [29] for details about the estimation of this matrix). We can also perform calibration for models 1 and 2 with the cost function (10) which leads to slightly different results than calibration with (9).

Experimental results

Equipment. In order to evaluate and compare the considered mathematical models, we have conducted a series of experiments using industrial videoendoscope Mentor Visual iQ Videoprobe (GE Inspection Technologies) with forward-view stereo adapter. The probe has a diameter 6.1 mm and a 1/6" CCD image sensor with 440000 pixels. A prism-based stereo adapter provides two registration channels with the fields of view 55°/55°.

To acquire images for calibration and tests, we utilized two types of plane calibration targets: (1) with black circular dots on a white background and (2) with black and white chessboard pattern. Due to sufficient range of distances and magnifications, three samples of each calibration target type with 0.5 mm (small-sized), 1 mm (middle-sized) and 2 mm (large-sized) distance between markers have been manufactured. This distance corresponded to the distance between the centers of the dots and to the chessboard square size for two types of target respectively. Each target had a grid of 25×25 markers.

The distal end of the endoscope was fixed on a mechanical stand that allows adjustable movement along z-axis and rotation around y and z-axis. The axis of the probe was approximately coincident with the z-axis of the stand. Another mechanical stand was placed in front of the endoscope to mount calibration targets. The plane of calibration target was set approximately perpendicular to the z-axis of the stand.

Image acquisition and processing. We have conducted a few experiments using different types of calibration targets described above. At first, dot and chessboard targets have been used for independent experiments. Next, we have carried out an experiment replacing calibration target

in each position to obtain equal conditions for two types of targets. For each experiment we have captured two series of images: calibration sequence and test sequence.

The calibration sequence included 18 images captured at different positions of three calibration targets: small-sized, medium-sized and large-sized. Each of these targets has been placed approximately perpendicular to z-axis of the stand at the distance which allows to see at least eight markers in a horizontal row on both image halves. Then, the distal end was consequently rotated 30° around x and y axes to capture images at oblique angles. Finally, we have positioned the target perpendicular to z-axis at approximately 1.5× larger working distance than initial position. As a result, we have acquired 6 images for each size of the calibration target covering the total range of distances from 8 to 40 mm.

Test sequences were registered for medium-sized and large-sized calibration targets in the following way. Again, the calibration target has been placed approximately perpendicular to z-axis of the stand at reasonably close distance and oriented so that horizontal rows of markers were approximately horizontal on the image. We used translation stage to shift the distal end of the endoscope along z-axis and captured images with 1 mm step. Thus, we have obtained series of images for each calibration target. These series contained 16 images in the range from 12 to 27 mm for medium-size targets and 19 images in the range from 22 to 40 mm for large-sized ones. The range of distances was limited by the condition that the number of visible markers should not be too small and the condition of minimal marker scale suitable for robust image processing.

In order to calculate image coordinates $\mathbf{p}_{i,j,k}$ for each marker, we have implemented in MATLAB the image processing algorithms for detection of dot and chessboard pattern from [30]. All automatically processed images of calibration and test sequences have been checked manually afterwards to delete points in low-illuminated areas, around glare, etc. Captured images of two calibration targets with estimated marker positions are shown in Fig. 4. These images correspond to the calibration series.

Calibration. We have implemented the calibration algorithms for each earlier considered mathematical model using non-linear iterative solver from MATLAB for the constrained minimization problem formulated in Eq. (8). The constraints for the values of calibration parameters $\mathbf{k}$ are determined by the physical limitations of prism parameters: size (less than 5 mm), refraction index (from 1.4 to 1.8), etc. We used the cost function C based on distance in the image plane for models 1 (usual pinhole model) and 2 (pinhole model with extended polynomial distortion) according to Eq. (9).

The covariation matrix $\Sigma_{\mathrm{p}\,i,j,k}$ was assumed to be the identity matrix due to the lack of a priory information about the error distribution of the image coordinates $\mathbf{p}_{i,j,k}$. As mentioned above, model 3 (ray tracing model) does not provide closed-form solution for the forward transformation which makes this cost function inappropriate. Hence, we used another cost function based on the distance in the XOY plane of the CS of the calibration target as shown in Eq. (10).
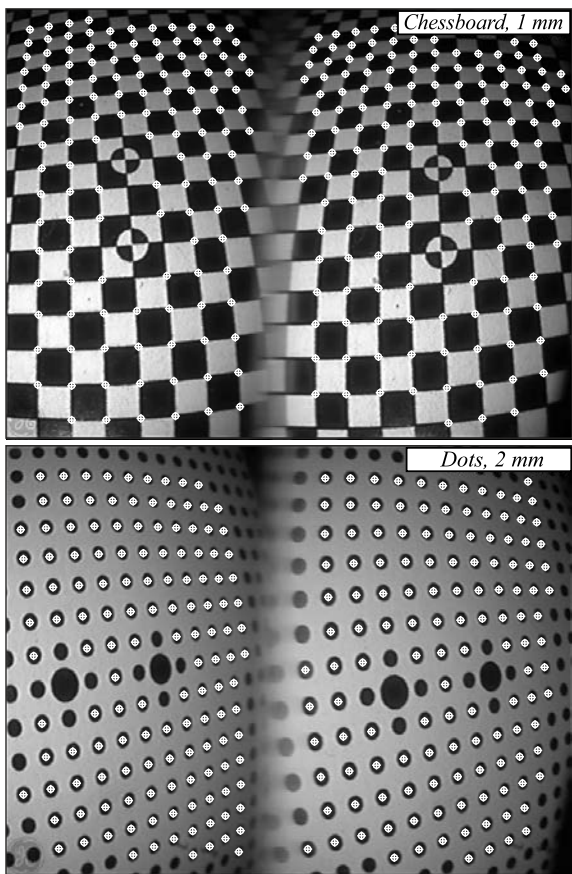
Fig. 4. Captured images with estimated marker positions for two types of calibration target

For simplicity, the covariation matrix $\Sigma_{x\,i,j,k}$ was also considered as the identity matrix, but this is sufficiently less reasonable assumption than the same one for $\Sigma_{p\,i,j,k}$ according to [4, 29]. In order to assess the impact of the calibration algorithm on the measurement accuracy we

have calibrated model 1 using the cost function (10). We further refer to this hybrid model as model 1' (usual pin-hole model calibrated in 3D space).

We have processed four calibration sequences and have calculated four vectors of parameters for each of the considered mathematical models. Two of these sequences have been captured during the first experiment. The first sequence included images of the dot calibration target and the second one included images of the chessboard target. These sequences were acquired independently and the positions of calibration target were different. The next two sequences have been captured replacing the calibration target in each position during the second experiment. Thus, the planes of the targets were approximately the same for each position. We have compared the calibration results for each model calculated with different targets to assess the impact of this factor. Our analysis has shown that the difference of parameters calculated for two sequences of the second experiment is close to the difference of parameters calculated for two series of the same calibration target.

The calibration results for the chessboard target from the first experiment are listed in Table 1. The affine transformation A_i is described by the focal length f_x, f_y and the image center c_x, c_y [4, 16], the skew parameter is considered to be zero. The parameters of distortion transformation D_i for model 1, 1' and 3 correspond to Eq. (2). Model 2 uses 17 parameters for the distortion transformation: $kx_{1,4}$, $kx_{0,4}$, $kx_{3,2}$, $kx_{1,2}$, $kx_{0,2}$, $kx_{1,1}$, $kx_{5,0}$, $kx_{3,0}$, $kx_{2,0}$, $ky_{0,5}$, $ky_{2,3}$, $ky_{0,3}$, $ky_{0,2}$, $ky_{4,1}$, $ky_{2,1}$, $ky_{1,1}$ and $ky_{2,0}$ according to Eq. (3). The prism parameters estimated for model 3 are as follows: Back face: $s_{1,1\,x}=0.014$, $s_{1,1\,y}=-0.004$, $d_{1,1\,z}=0.014$; Front face 1: $s_{2,1\,x}=-0.436$, $s_{2,1\,y}=-0.013$, $d_{2,1\,z}=3.401$; Front face 2: $s_{2,2\,x}=0.463$, $s_{2,2\,y}=-0.012$, $d_{2,2\,y}=3.429$ (all measured in mm) and the index of refraction $n_1=1.663$ (see Fig. 3 for notation).

Table 1. Calibration parameters

	Parameters	Units	Model 1	Model 1'	Model 2	Model 3
Camera 1	focal length f_x, f_y	pixel	594.73; 707.01	590.41; 703.16	616.81; 715.02	732.93; 749.92
	image center c_x, c_y	pixel	40.28; 314.74	47.66; 315.07	64.91; 317.15	392.2; 301.89
	distortion k_1, k_2	–	–0.55; 0.124	–0.566; 0.152	17 parameters for distortion	–0.567; 0.242
	distortion ρ_1, ρ_2	–	–0.005; 0.076	–0.007; 0.073		–
Camera 2	focal length f_x, f_y	pixel	582.28; 702.36	577.44; 699.29	598.13; 709.27	equal to the parameters of camera 1
	image center c_x, c_y	pixel	748.7; 307.26	744.02; 305.03	738.09; 311.71	
	distortion k_1, k_2	–	–0.51; 0.114	–0.533; 0.143	17 parameters for distortion	
	distortion ρ_1, ρ_2	–	–0.004; –0.069	–0.004; -0.067		
	Rotation ω_1, ω_2, ω_3	radians	–0.012; –0.398; –0.003	0.018; -0.383; -0.003	–0.009; –0.331; –0.008	–
	Translation t_1, t_2, t_3	mm	–1.264; –0.025; 0.298	–1.276; –0.058; –0.307	–1.276; –0.021; –0.257	–

Error analysis. The images of the test sequences have been used to calculate the 3D coordinates for each marker according to Eq. (6). The algorithm have been implemented in MATLAB similar to the calibration algorithm. The same way, we used the cost function (7) based on the distance in the image plane for models 1 and 2 with the identity covariation matrix. The cost function based on the sum of the distances from the 3D point to the back-projected rays was used for models 1' and 3. We have considered all covariation

matrices equal to the identity ones so that all coordinates, points and rays had equal weights in the cost functions.

In most applications of 3D measurement endoscopic systems, the calculated 3D point coordinates are used to measure the geometric parameters such as distances or areas. Hence, we should utilize the criteria based on the deviations of the length or the area to evaluate the mathematical models for these systems. We have processed 3D points calculated for each image of the test sequences to estimate the distance between neighbouring markers. Due to the chosen orientation of

the calibration targets these segments are approximately parallel to *x*- and *y*-axis of the stand. The orientation of the CS is the same as shown in Fig. 3. The true value of the distance is assumed to be 1 mm for the medium-sized target and 2 mm for the large-sized one. Next, we have processed 3D points for pairs of consequently captured images to estimate the distance between points corresponding to the same marker. The obtained segment is approximately parallel to *z*-axis and its true length is assumed to be 1 mm (equal to the shift step for the distal end of the endoscope). Additionally, the position of the

calibration target has been estimated for each image relying on the calculated 3D point coordinates.

We have estimated the distance measurement errors for all captured test sequences and all considered mathematical models using corresponding vectors of parameters from the calibration stage. Since the results for different types of calibration targets are almost coincident, we present the results for the chessboard target from the first experiment only. The absolute errors Δ*r* are shown in Fig. 5 for the medium-sized target.
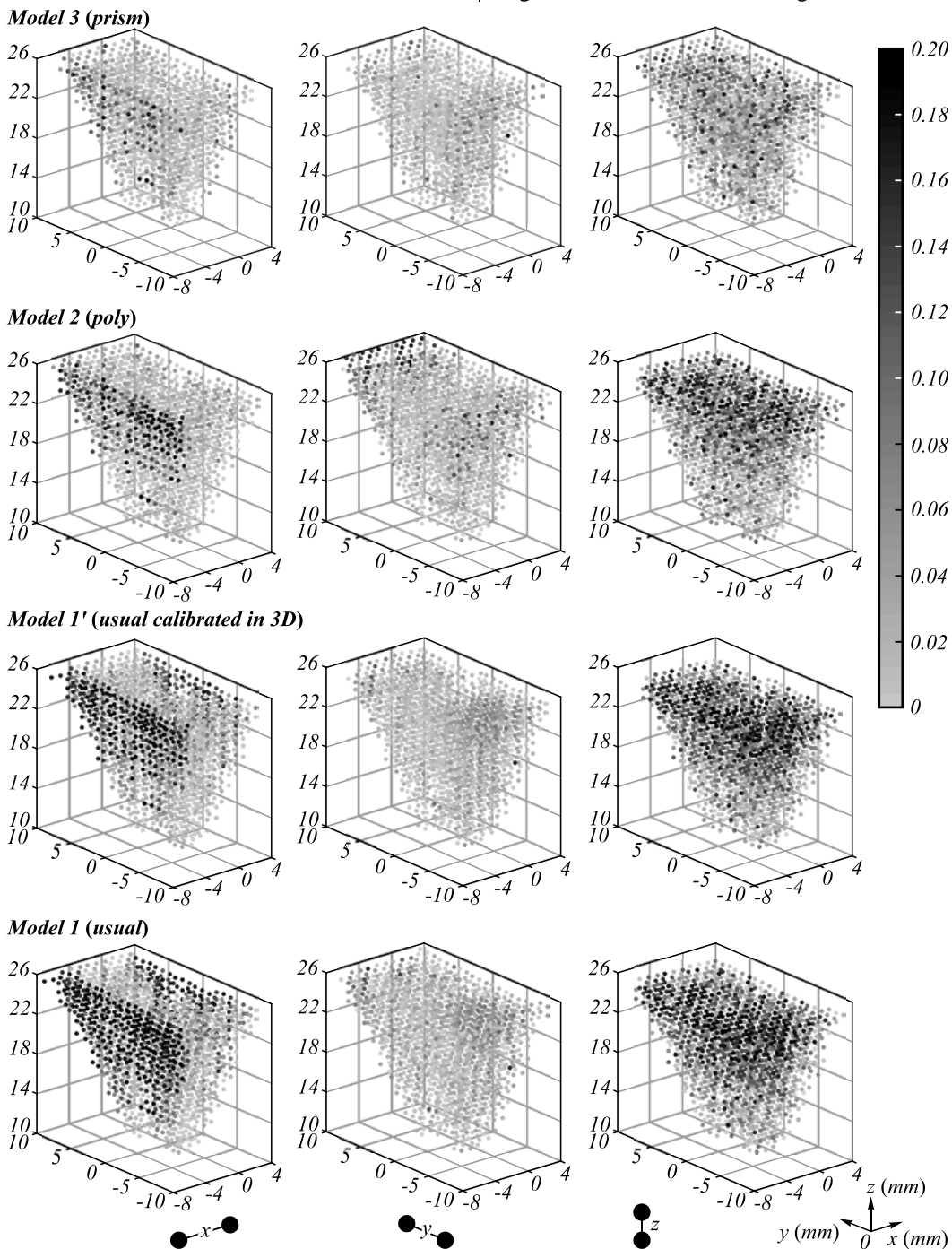

Fig. 5. Absolute error of distance measurements for medium-sized chessboard calibration target. The true value of each segment is 1 mm, all values are in mm. The value of error is indicated by dots' color according to the grayscale bar in the right side of the figure. Each row of plots represents one of four considered mathematical models, each column corresponds to segments parallel to x-, y and z-axis shown in the bottom part respectively

The error value is indicated by dots' color according to the grayscale bar in the right side of the figure. The dots indicating segments with error value larger than 0.2 mm are filled with the black color intended for 0.2 mm values. The coordinates of the dots correspond to the estimated positions of the calibration target for a visual clarity, the coordinates of the first point of segment are shown. Each row of the plots represents one of four considered mathematical models, each column corresponds to the segments parallel to x-, y- and z-axis respectively. The CSs for the plots are slightly different. The origin for model 3 is determined by the center of the internal pinhole camera as shown in Fig. 3. The middle point between two camera centers is chosen as the origin for models 1, 1' and 2 (see Fig. 2).

One can see that the error value varies significantly for different parts of the observed volume and the different orientation of the segment. Thus, the choice of the integral criterion for the quantitative evaluation representing the values distribution is not clear. According to common applications of 3D measurement endoscopic systems, we have divided the obtained data set into zones by z-coordinate and calculated mean and standard deviation of the segment length for every zone. The number of zones is equal to the number of images in the test sequence, hence, each captured image belongs to single zone. The results are shown in Fig. 6. The top row of the plots represents test sequence for large-sized calibration target, the true value of x- and y-segments is 2 mm and the true value of z-segments is 1 mm. The bottom row of plots represents test sequence for medium-sized target, the true value of all segments is 1 mm.

The results presented in Figs. 5 and 6 indicate that the error for x-segments grows significantly on the edges of the field of view for models 1, 1' and 2. In contrast, all models demonstrate relatively similar performance for y-segments. Furthermore, the error for x- and y-segments varies for each zone which leads to a large standard deviation due to a chosen evaluation method. In contrast, the standard deviation for z-segments is comparable for all models, but the usage of models 1, 1' and 2 leads to unacceptable bias which increases with distance. The correlation of spikes on the plots for the mean error of the length for z-segments reveals uncertainties in shifting of the distal end during the experiment. Model 1' demonstrates slightly better results than model 1 for x-segments, but the overall performance is relatively similar. Hence, the choice of the cost function has no great impact on the accepted evaluation of models. The main conclusion based on the experimental results is that model 3 has a superb performance in comparison to other models. Model 2 may be used for the working distances less than 20 mm but its usage is limited due to bias in measurements of length along z-axis at larger distances.
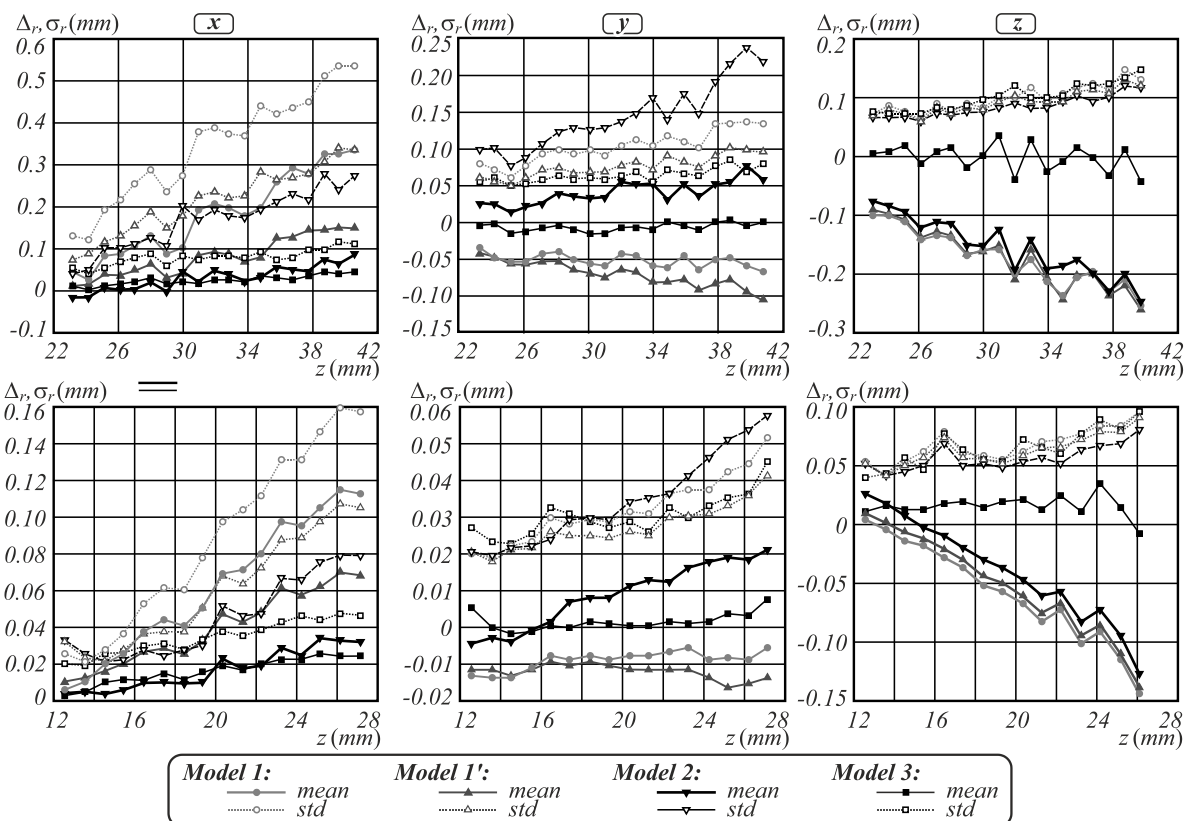

Fig. 6. Mean and standard deviation of distance: large-sized calibration target (top row) and medium-sized target (bottom row). Each column corresponds to segments parallel to x-, y- and z-axis respectively. The true value of x- and y-segments is 2 mm for the large-sized target and 1 mm for the medium-sized one. The true value of z-segments is 1 mm

Conclusion

In this paper, we first theoretically and experimentally analyzed and compared three different mathematical models of prism-based stereoscopic imagers. The experimental data confirmed the results previously obtained by computer simulation: any of considered pinhole models can not provide the required accuracy of the measurement over a wide range of working distances. The model based on ray tracing through the prism shows much better results. An increase in the number of degrees for distortion polynomial model does not help to improve the accuracy, apparently, because of the assumption that all rays pass through the single point. The error of the length segment measurement has a complex spatial distribution and depends strongly on the orientation of the segment. This fact should be taken into account for evaluation of the videoendoscopes and formulating the criteria for their measurement accuracy testing. The distribution can be used for the interactive user assistance by indicating the estimated measurement errors or giving recommendations on the location of the probe relative to the object. We have shown that the simple calibration method using a flat target allows to conduct the calibration of the endoscope with attached prism-based stereo adapter and reach the desired accuracy. The method does not require the results of other specific calibrations or the prism parameters. Type of the calibration target pattern selected for the calibration and test measurements has an insufficient effect on the result.

The obtained theoretical and experimental results may be useful for the improvement of the existing prism-based videoendoscopic imagers in terms of the measurement range and accuracy as well as for the development of the new small-size tools for remote visual inspection, especially for the important industrial tasks where high precision of defect characterization is crucial.

References

[1]   Mix PE. Introduction to nondestructive testing: A training guide. 2nd ed. New York: John Wiley & Sons; 2005. ISBN: 978-0-471-42029-3.

[2]   Zhu Y-K, Tian G-Y, Lu R-S, Zhang H. A review of optical NDT technologies. Sensors (Basel) 2011; 11(8): 7773-7798. DOI: 10.3390/s110807773.

[3]   Olympus Corp. Demonstration of stereo measurement. Source: ⟨http://www.olympus-ims.com/en/rvi-products/ip-lex-fx/demo-stereo-measurement/⟩.

[4]   Hartley RI, Zisserman A. Multiple view geometry in computer vision. 2nd ed. Cambridge: Cambridge University Press; 2004. ISBN: 978-0-521-54051-3.

[5]   Kannala J, Brandt SS. A generic camera model and calibration method for conventional, wide-angle, and fish-eye lenses. IEEE Trans Pattern Analysis and Machine Intelligence 2006; 28(8): 1335-1340. DOI: 10.1109/TPAMI.2006.153.

[6]   Zhang S, ed. Handbook of 3D machine vision: Optical metrology and imaging. Boca Raton, Fl: CRC Press; 2016. ISBN 978-1-138-19957-6.

[7]   Lee D, Kweon I. A novel stereo camera system by a bi-prism. IEEE Transactions on Robotics and Automation 2000; 16(5): 528-541. DOI: 10.1109/70.880803.

[8]   Xiao Y, Lim KB. Virtual stereovision system: New understanding on single-lens stereovision using a biprism. Journal of Electronic Imaging 2005; 14(4): 043020. DOI: 10.1117/1.2137654.

[9]   Xiao Y, Lim KB. A prism-based single-lens stereovision system: From trinocular to multi-ocular. Image and Vision Computing 2007; 25(11): 1725-1736. DOI: 10.1016/j.imavis.2007.01.002.

[10] Wang D, Lim KB, Kee WL. Geometrical approach for rectification of single-lens stereovision system with a triprism. Machine Vision and Applications 2013; 24(4): 821-833. DOI: 10.1007/s00138-012-0467-8.

[11] Kee WL, Lim KB, Tun ZL, Yading B. New understanding on the effects of angle and position of biprism on single-lens biprism stereovision system. J Electron Imaging 2014; 23(3): 033005. DOI: 10.1117/1.JEI.23.3.033005.

[12] Kee WL, Bai Y, Lim KB. Parameter error analysis of single-lens prism-based stereovision system. J Opt Soc Am A 2015; 32(3): 367-373. DOI: 10.1364/JOSAA.32.000367.

[13] Cui X, Lim KB, Guo Q, Wang D. Accurate geometrical optic model for single-lens stereovision system using a prism. J Opt Soc Am A 2012; 29(9): 1828-1837. DOI: 10.1364/JOSAA.29.001828.

[14] Wu L, Zhu J, Xie H. A modified virtual point model of the 3D dic technique using a single camera and a bi-prism. Measurement Science and Technology 2014; 25: 115008. DOI: 10.1088/0957-0233/25/11/115008.

[15] Wu L, Zhu J, Xie H. Single-lens 3D digital image correlatio system based on a bilateral telecentric lens and a bi-prism: validation and application. Appl Opt 2015; 54(26): 7842-7850. DOI: 10.1364/AO.54.007842.

[16] Zhang Z. Flexible camera calibration by viewing a plane from unknown orientations. Proc International Conference on Computer Vision 1999; 666-673.

[17] Cui X, Lim KB, Zhao Y, Kee WL. Single-lens stereovision system using a prism: position estimation of a multi-ocular prism. J Opt Soc Am A 2014; 31(5): 1074-1082. DOI: 10.1364/JOSAA.31.001074.

[18] Cui X, Zhao Y, Lim K, Wu T. Perspective projection model for prism-based stereovision. Opt Express 2015; 23(21): 27542-27557. DOI: 10.1364/OE.23.027542.

[19] Lim KB, Qian B. Biprism distortion modeling and calibration for a single-lens stereovision system. J Opt Soc Am A 2016; 33(11): 2213-2224. DOI: 10.1364/JOSAA.33.002213.

[20] Qian B, Lim KB. Image distortion correction for single-lens stereo vision system employing a biprism. Journal of Electronic Imaging 2016; 25(4): 043024. DOI: 10.1117/1.JEI.25.4.043024.

[21] Genovese K, Casaletto L, Rayas J, Flores V, Martinez A. Stereo-digital image correlation (DIC) measurements with a single camera using a biprism. Optics and Lasers in Engineering 2013; 51(3): 278-285. DOI: 10.1016/j.optlaseng.2012.10.001.

[22] Bendall C, Lia R, Salvati J, Chilek T. Stereo-measurement borescope with 3-D viewing. US Patent 7,564,626; published on Jule 21, 2009.

[23] Nakano S, Hori F, Ogawa K. Endoscope device and endoscopi image distortion correction method. US Patent 8,979,743; published of March 15, 2015.

[24] Takata Y, Torigoe T, Kobayashi E, Sakuma I. Distortion correction of wedge prism 3D endoscopic images. Proc 7th Asian-Pacific Conference on Medical and Biological Engineering (APCMBE 2008) 2008; 750-753. DOI: 10.1007/978-3-540-79039-6_185.

[25] Machikhin AS, Gorevoy AV. Calibration of miniature prism-based stereoscopic imagers for precise spatial meas-

urements. Proc SPIE 2015; 9917: 991707. DOI: 10.1117/12.2229837.

[26] Machikhin AS, Gorevoy AV, Perfilov AM. Accuracy evaluation for the non-contact area defect measurement at the complex shape-surfaces under videoendoscopic control [In Russian]. Scientific and Technical Journal of Information Technologies, Mechanics and Optics 2014; 14: 140-148.

[27] Wu C, Jaramaz B, Narasimhan S. A full geometric and photometric calibration method for oblique-viewing endoscope. Computer Aided Surgery: Official Journal of the International Society For Computer Aided Surgery 2010; 15(1-3): 19-31. DOI: 10.3109/10929081003718758.

[28] Brown DC. Decentering distortion of lenses. Photometric Engineering 1966; 32: 444-462.

[29] Kanatani K. Statistical optimization for geometric computation: Theory and practice. Mineola: Dover Publications; 2005.

[30] Mallon J, Whelan PF. Which pattern? Biasing aspects of planar calibration patterns and detection methods. Pattern Recogn Lett 2007; 28(8): 921-930. DOI: 10.1016/j.patrec.2006.12.008.

Authors' information

Alexey Vladimirovich Gorevoy (b. 1987) graduated from Bauman Moscow Technical University in 2010, Laser and Optic-Electronic Systems subdepartment. Currently he works as the researcher at the Scientific and Technological Center of Unique Instrumentation of RAS. Research interests include image processing, optical system design and computer graphics. E-mail: *gorevoy.a@gmail.com* .

Alexander Sergeevich Machikhin (b. 1984) graduated from Bauman Moscow Technical University in 2006, Development of Optic-Electronic Systems for Scientific Research subdepartment. PhD. Currently he works as the leading researcher at the Scientific and Technological Center of Unique Instrumentation of RAS. Research interests include machine vision, spectral imaging, acousto-optics and spectroscopy. E-mail: *aalexanderr@mail.ru* .

Р