Depth Map Improvement by Combining Passive and Active Scanning Methods

Ondrej KALLER, Libor BOLECEK, Ladislav POLAK, Tomas KRATOCHVIL

Dept. of Radio Electronics, Brno University of Technology, CZ-61600 Brno, Czech Republic

xkalle00@stud.feec.vutbr.cz, kratot@feec.vutbr.cz

Manuscript received November 29, 2015

Abstract. *The paper presents a new method of more precise estimation of the depth map in 3D videos. The novelty of the proposed approach lies in sophisticated combination of partial results obtained by selected existing passive and active 3D scanning methods. The aim of the combination is to overcome drawbacks of individual methods and this way to improve the accessible precision of the final depth map. The active method used is incoherent profilometry scanning which fails on surface discontinuities. As a passive method, a stereo pair matching is used. This method is currently the most widely applied method of depth map estimation in the field of 3D capturing and is available in various implementations. Unfortunately, it fails if there is a lack of identifiable corresponding points in the scanned scene. The paper provides a specific way of combining these methods to improve the accuracy and usability. The proposed innovative technique exploits the advantages of both approaches. Specifically, the more accurate depth profiles of individual discontinuous objects obtained from the active method, and information about mean depths of the objects from the stereo pair are combined. Two implementations of the passive method have been tested for combination with active scanning: matching from stereo pair, and SIFT.*

The paper includes a brief description of the active and passive methods used and a thorough explanation of their combination. As an example, the proposed method is tested on a simple scene whose nature enables straight assessment of the achieved accuracy. The choice of a suitable implementation of the passive component is also shown and discussed. The obtained results of individual existing methods used and of the proposed combined method are given and compared. To demonstrate the contribution of the proposed combined method, also a comparison with the results obtained with a commercial solution is presented with significantly good results.

Keywords

Profilometry, stereoscopical capturing, depth map, phase unwrapping, stereo pair, 3D images, map estimation, shadow detection

1. Introduction

3D video capturing can be realized by various camera systems working on many physical principles. We can observe two paths of development, namely active and passive 3D capturing systems. Active capturing systems utilize the projection of the measurement pattern on a scanned scene which could be in visible light spectrum [1], [2], near infrared field [3] or projected by a focused laser spot [4]. In the mostly used passive system, the depth of the pixel is determined from its disparity in a stereo pair. The depth can be also estimated from a multicamera facility [5], depth field camera [6] or from monoscopic camera auto-focusing parameters [7].

Regardless the variety of capture systems principles, most of them have a similar output format (2D + a depth map). Based on this output format, it is possible to render more views by Depth Image Based Rendering (DIBR). These views are then compressed by Multiview Video Coding (MVC) and used also in television broadcasting [8].

Some part of current 3D video shooting systems is based on combinations of more depth acquisition methods, such as Time-of-Flight (TOF) IR camera with a stereo pair, where a depth image is rectified to the color camera images [2]. In some other approaches, the stereo pair is fructified by combining advanced and conventional methods of image segmentation [9], or more monocular cues are combined to estimate the depth map [10]. Another combination is a profilometry scanning system with two cameras [1] which is a very similar system design as proposed in this paper, but with completely different data processing. The approach proposed in this article was first mentioned in our previous contribution [11] where two possible modifications were outlined.

The aim of this paper is to introduce a new specific capture system for depth map estimation in 3D TV. The idea is based on a combination of the active scanning with a passive method in which depth information is estimated from a stereo pair. A precise 3D model of the scene providing the true depth map was created to demonstrate good accuracy of the proposed system. The relevance of

our method is also demonstrated by comparing of the results with a professional (commercial) 3D active system.

The rest of the paper is organized as follows. Section 2 contains a brief description of active and passive 3D capturing methods of interest. Section 3 deals with the definition of the theoretical depth accuracy. The proposed algorithm for the depth information synthesis is described in Sec. 4. Analysis of practical implementations of the proposed method and evaluation of obtained depth maps are presented in Sec. 5 and Sec. 6, respectively. Finally, Section 7 concludes the paper.

2. Current Methods of Depth Map Generation

As mentioned above, there is a huge variety of depth map estimation methods. Before dealing with the proposed combined method, a brief description of existing methods of interest is given in this section, altogether with available technical information about the commercial KinectTM system.

2.1 Depth Map Estimation from a Stereo Pair

Most of today's 3D captures systems use a passive method for depth map estimation, based on a stereo pair analysis.

There are mainly two types of this method. Firstly, classic and older approaches are referred to as area-based methods [4]. In most cases, well-matched camera parameters are assumed, namely focal length, depth of field and resolution. The description of epipolar geometrical parameters is epitomized in a fundamental matrix to be found. Then rectification [3] is performed, meaning transformation of input stereo pair images is carried out so that epipolar lines of output images correspond with the same image rows. After that the corresponding points are sought for just along these lines. The basic algorithms for Disparity Space Imaging (DSI) are Sum of Squared Differences (SSD), Sum of Absolute Differences (SAD) and Normalized Cross Correlation (NCC) [4].

The second category consists of feature-based methods. They can find corresponding points within the whole images of the stereo pair. The Scale-Invariant Feature Transform (SIFT) or Speeded Up Robust Feature (SURF) are algorithms which assign a descriptor of each characteristic pixel. Correspondent points are then found at the base of this description [4].

Problems occur when objects of the scene have a large monochromatic surface, and thus characteristic points cannot be identified in order to find the correspondences. A similar situation can happen when the surface of the photographed object has a fine periodical structure in the horizontal direction. Although the algorithms for depth estimation are “best effort”, meaning they choose the most probable variant of the depth map, an inaccuracy or error cannot be detected or reduced.

2.2 Profilometry Scanning

Profilometry is a very common method for accurate surface topography measurement. It can use coherent light, but in macroscopic scanning systems, incoherent methods (such as Fourier’s profilometry, phase-shifting profilometry or moiré topography) are usually used.

In this work, the phase shifting profilometry is used because it is very easy to implement [12], [13]. It should be noted that in the case of the profilometry ideal functionality, it is not important which implementation is applied for our purposes.

Incoherent methods are based on triangulation of a measured system. On its way from the source to the detector, reflection of a particular ray from the measured surface takes single valued information about the depth. The intensity of each ray (pixel) is modulated by a sample of the sine pattern which is projected to the scene. The pattern is phase-shifted in time. This basic principle also yields an advantage which is utilized in the proposed method. In case of a continuous surface, profilometry provides continuous information about the depth, meaning the depth value for each visible pixel [14].

2.3 Professional Active Scanning System

To show practical usability of the designed system, described in Sec. 4, it is useful to put its parameters into context with a commercial solution. For comparison, the commercial Kinect device with depth sensors by PrimeSense was used. Unfortunately, producers have not published details, but experimentally obtained parameters can be found in the report [15].

The sensor combines two methods of active scanning: a structural light analysis and a depth from focus. The first one is a triangulation method as well as profilometry is. Nevertheless, the classical profilometry approach codes information by a specific pattern to identify the position of each projected scanning point. Kinect structures infrared light to speckles of points which are randomly spread. The information about correspondence between projected and observed light spots has to be added another way.

Depth from focus is one of classical monocular cues of depth which rates blurring of an image, projected beside the focus plane [16]. The producer PrimeSense uses astigmatic lenses for structure light projection. This solution is based on the changing of geometrical parameters of projected spots along the depth dimension. The combination of mentioned methods joins high accuracy of structural light scanning in a continuous surface of scanned objects with a robust approach to the detection of their mutual position.

3. Theoretical Achievable Accuracy of the Described Methods

In this section, first, the term of the depth map accuracy is defined. Then the interval of depth values, to which the true depth value of a particular pixel belongs, is mentioned. In this section, attention is focused on the accuracy of the depth estimation of individual pixels disregarding specific depth profile of the scene as a whole. An example of particular depth maps obtained by various methods is discussed in the following section.

3.1 Depth Obtained from a Stereo Pair

In the following explanation, the passive method with full-pixel accuracy is assumed. In modern algorithms using n-sub-pixels accuracy [17], final depth error intervals could be reduced n-times.

Figure 1 shows the corresponding pixel P which is viewed by the left and right camera. Both cameras have finite horizontal resolution h_r. Transformation of pixel´s width to the y-depth object (plane) is d_p. It can be seen that the real corresponding point, which is sampled as pixel P, could lie in the Δy interval. The formula for the width of the pixel d_p at particular depth y is obvious:

$$d_p(y) = 2\frac{y}{h_r}\tan\frac{\alpha}{2}. \quad (1)$$

Inner parameters of the cameras and their parallel optical axes are assumed to match perfectly.

Geometrical parameters such as cameras' stereo base d, horizontal viewing angle α and depth value y are defined (see Fig. 1).

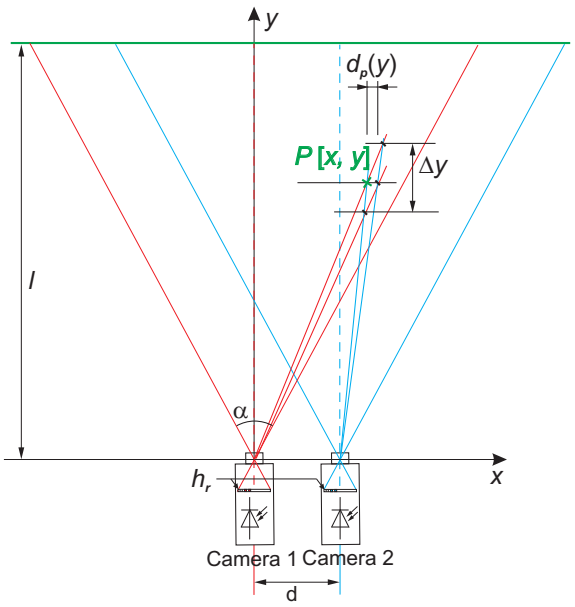

Fig. 1. The maximum theoretical accuracy of the depth value calculated from a stereo pair in relation with cameras' parameters and configuration.

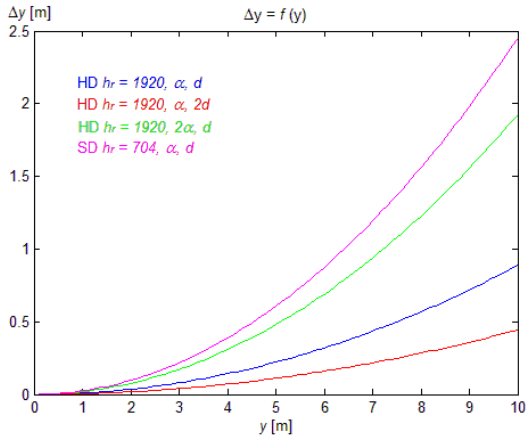

Fig. 2. Maximum theoretical accuracy of the depth estimation from a stereo pair in the case of variable cameras' horizontal resolution, viewing angle and stereo base ($\alpha = 30°$, $d = 6.3\times10^{-2}$ m).

Then the depth uncertainty Δy could be calculated as follows:

$$\Delta y(y) = 2\frac{d\, d_p y}{d^2 - d_p^2}. \quad (2)$$

The practical graphical interpretation of the previous equations is presented in Fig. 2. It demonstrates how the course of the function $\Delta y = f(y)$ depends on the mentioned parameters. Particular values in our examples are $h_r = 1920$ pix (704 pix), $\alpha = 30°$, $d = 63$ mm.

3.2 Depth from Profilometry

Profilometry scanning with sine phase shifted sets of patterns is quite a simple method which has an essential disadvantage, embedded ambiguity of depth, compared to alternative pattern. Figure 3 demonstrates the mentioned problem. Black lines illustrate light rays for a particular phase of the projected pattern. From the camera point of view it is not possible to distinguish from which of the green planes the light has been reflected. Examples of four planes are depicted in Fig. 3. Their distance corresponds to the period of the projected pattern. In other words, periodicity of the projected pattern results in the depth ambiguity: the same depth information is assigned to objects at a particular distance y and also $(y - \Delta l)$.

The function (3) maps phase shift $\Delta\phi$ between the measurement pattern projected on the reference plane and the pattern projected on the observed surface, to the distance h between the reference plane and the observed surface

$$h = \frac{\Delta\phi}{\frac{\Delta\phi}{l} - \frac{2\pi\tan\beta}{p}}. \quad (3)$$

For the phase shift $\Delta\phi$ equal to $2k\pi$, $k \in \mathbb{N}$, the following formula expresses the dependency of the depth ambiguity

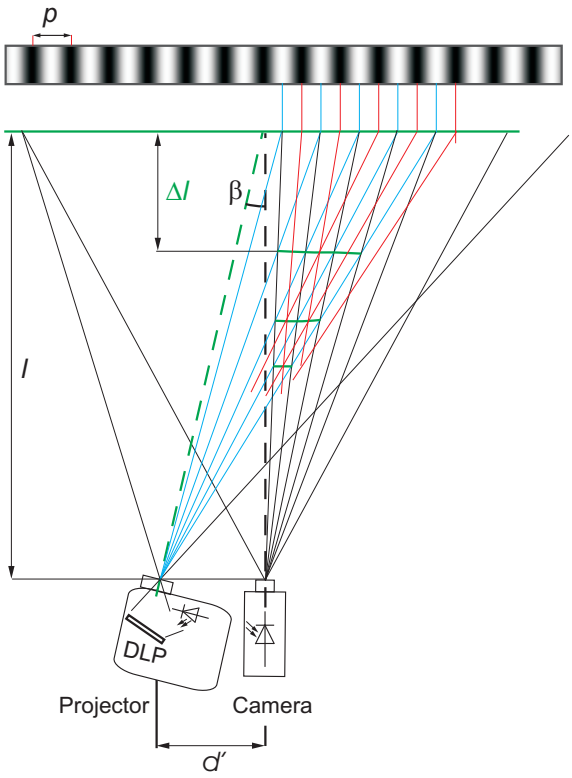

Fig. 3. The ambiguity of the depth representation caused by the periodical repetition of phase-coded depth information.

interval Δl on the parameters of the profilometry capture system, i.e. the period of the measurement pattern p, the distance of the camera and the projector focal points d', and the distance between the camera and the reference plane l:

$$\Delta l = \mathrm{f}\left(d', p, l\right) = \frac{pl}{\left|p - d'\right|}. \tag{4}$$

Figure 4 shows this dependence for $l = 2$ m and $p = 1\times10^{-2}$ m.

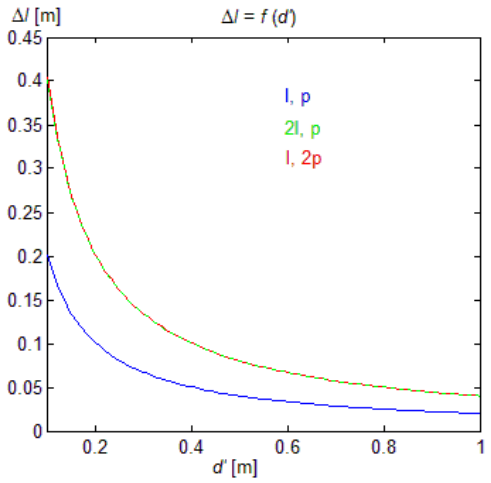

Fig. 4. The dependency of the depth ambiguity interval Δl on parameters of the profilometry capture system ($l = 2$ m, $p = 1\times10^{-2}$ m).

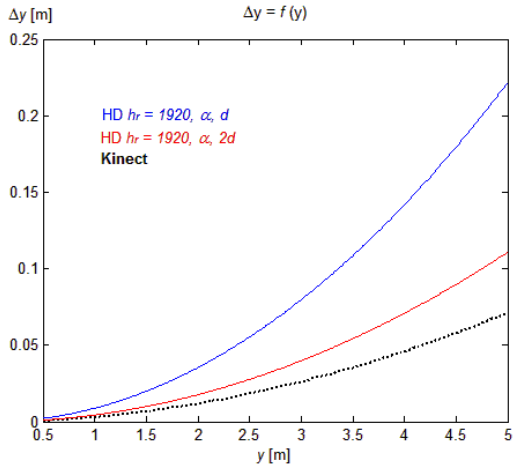

Fig. 5. The theoretical accuracy of Kinect compared with the depth from a stereo pair ($\alpha = 30°$, $d = 6.3\times10^{-2}$ m).

3.3 Professional Solution Application

The analysis of Kinect parameters is not explicit because the depth range, linearity and resolution could be influenced by specific software variant, even if the same hardware is used. The hardware specifications [15], [18] define the sensor's nominal depth range from 0.8 m to 3.5 m. However, from the practical application, it is obvious that the sensor works from 0.5 m up to 15 m in specific conditions [18]. The same problem is with the depth resolution which is declared as 1 cm in a 2 m distance.

The depth quantization step q has been found in [15]:

$$q\left(y\right) = 2.73\times10^{-3}\ y^2 + 7.4\times10^{-4}\ y - 5.8\times10^{-4}. \tag{5}$$

The quantization step is a parameter which is comparable with the depth uncertainty Δy, in the case of the depth from a stereo pair. The theoretical accuracies of the mentioned methods are compared in Fig. 5.

4. The Proposed Procedure: Combination of Two Methods

This section describes the implementation of the proposed combined system. The basic idea is based on combination of the two above mentioned methods. The static scene is captured by using of each of them and the information is combined to improve the relevance of the final depth map. All the objects are assumed to be illuminated both by the ambient light and the measurement pattern.

A flowchart of the proposed procedure is given in Fig. 6. Phase unwrapping is the most difficult step in active scanning method. The output from the block "Calculation of wrapped phase" is the phase structure within the range $-\pi$ to π, in which wraps (rapid phase shifting by π) occur. We adopt the method Unwrapping via Graph [14] for unwrapping. However, the algorithm failed because a rapid change of phase occurs in the shadow region too often.

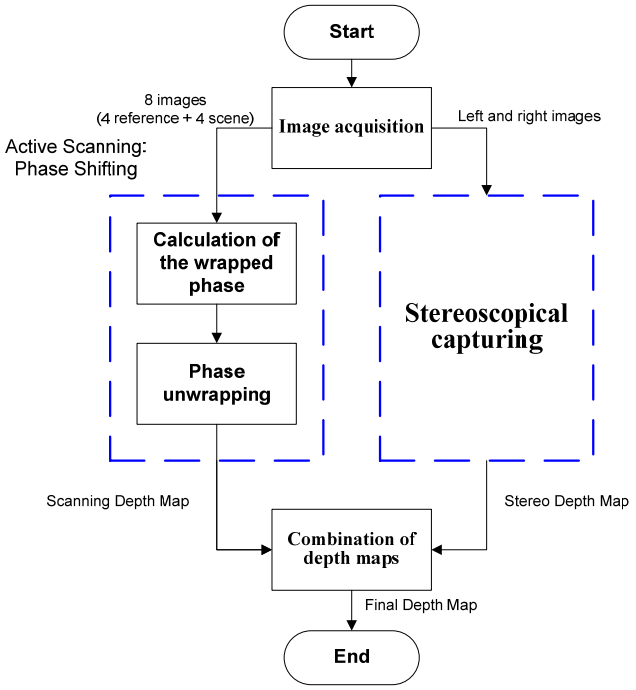

Fig. 6. The flowchart of the proposed procedure: Active scanning profilometry with depth from stereo pair.

Therefore, first, shadows must be detected and their influence eliminated.

4.1 Shadow Detection

For the shadow detection a formerly presented algorithm is used [19]. Its flow chart is shown in Fig. 7. The input "Stereo depth map" has information about topography obtained from the stereo pair and 2D image of the captured scene.

In $L \times a \times b$ space the background of the scene is thresholded and Suspicions for shadow (S_S) are found. The shadows are than excluded in areas of objects. Suspicions for objects (S_O) are detected from the smoothed depth map.

In the next step, data from both images (S_O, S_S) are combined. The basic assumption says that a pixel cannot be simultaneously included to the foreground and to the shadow because no of the objects is hidden in a shadow. In accordance with this assumption, the assignment of each pixel to shadow region is confirmed as expressed by the following pseudo-code:

```
If (S_S = = 1 & S_O = = 0)
    Shadow = 1;
Else
    Shadow = 0;
End
```

In the final step, small disturbing artifacts are removed by morphological operations and the MATLAB function *bwreaopen*. In the resultant shadow map of the scene, pixels belonging to shadow regions are labeled by logic 1 values.

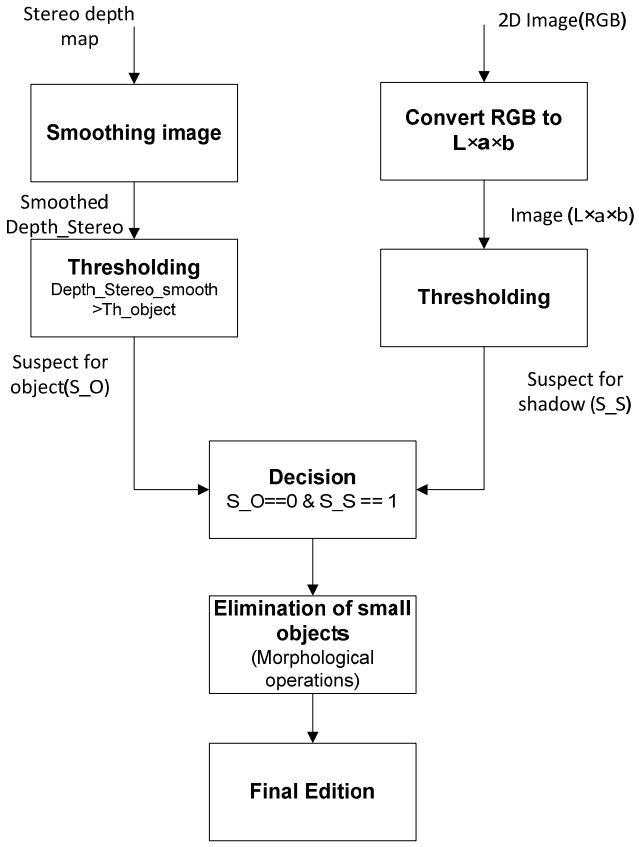

Fig. 7. Flowchart of the proposed algorithm for shadow detection [19].

4.2 Combination of Depth Maps

The main part of the proposed procedure consists in the combining of the two obtained depth maps. Inputs to this algorithm are the depth map achieved by the stereo method, the depth map obtained by the phase shifting profilometry, the shadow map and the original image of the scene.

The process of the combination is based on the properties of each depth map. The stereo depth map provides good information about mutual positions of objects, but the profile of each object is inaccurate. On the contrary, the profilometrical depth map has a precise profile of each object but does not provide the relationship among the positions of the objects. Therefore, it is needed to obtain the profile of each object from the profilometrical depth map and to transform it to the range given by the stereo map.

Firstly, individual objects in the image must be found. For this purpose, the shadow map and the profilometrical depth map will be used. This step is based on the assumption that an object belongs to the foreground, hence its values of the depth map will be high. Concurrently, objects are assumed not stay in the shadow. In consequence, we use the following condition:

$$(Shadow.map == 0) \ \& \ (stereo.depth.map > threshold).$$

The pixel which satisfies this condition belongs to the object and its value in the new matrix **Object** is logic 1.

In the following step, objects are classified. The registration of an image means that for each object, linking pixels are defined. As a result, the matrix **Class_objects** (1920 × 1080) is obtained whose elements are integers $i = 1,2,\ldots, n$ defining the assignment of each pixel to one of n registered objects. In the next step, the range of the depth of each object is found. All the pixels belonging to the object are sorted according to their depth. Subsequently, the upper and lower threshold (th_{low}, th_{up}) are determined as values corresponding to 95 and 5 percent of the depth of the object. This way, the range of depth of each object in the stereo depth map is obtained. This range is used as the range of object's depth in the final depth map *DM*. The minimum and maximum depths of each object in the profilometrical depth map are also found (*min, max*). Thus, each of n different objects is characterized by parameters (th_{low}, th_{up}, *max, min*). Then, the profilometrical depth map DM_{prof} is transformed separately for each object as follows:

$$DM_i = \left(th_{\text{up}\,i} - th_{\text{low}\,i}\right) \cdot \frac{DM_{\text{prof}\,i} - min_i}{max_i - min_i}, \quad i = 1,2,\ldots,n. \quad (6)$$

5. Implementations and Verification of the Idea

To verify the depth map accuracy improvement, a laboratory setup was prepared. Positions of three simple geometrical objects (two cylinders made of paper, one sphere made of white glass), of the cameras, and of the projector in the static scene are obvious from Fig. 8.

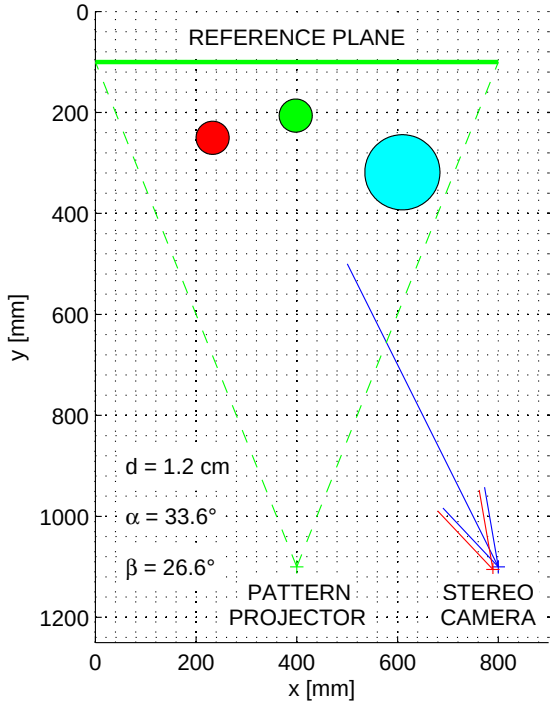

Fig. 8. The ground plan of the experimental scene.

The photo of the scanning equipment is taken in from a perspective of 3D scene (see Fig. 9). Starting from the right, a projector for sinusoidal pattern projection, a stereoscopic camera with a reduced stereo base, an active camera Kinect and finally, a PC to record and process the captured signals can be seen.

One of possible principles of combining the two scanning methods is plotted in Fig. 10. The DLP data projector, which projects a measurement pattern by unpolarized light, is complemented with a linear polarizing filter. This filter is oriented vertically. Besides the projector, the scene is illuminated by another source of light (a spotlight). The second polarization filter with horizontal polarization is added to the left objective of the stereo camera.

To actively scan and to record a stereo pair simultaneously, the measurement pattern can be projected by the projector and captured by the right camera, in which the light intensity of this pattern is added to the background intensity of the spotlight. The left camera then captures just

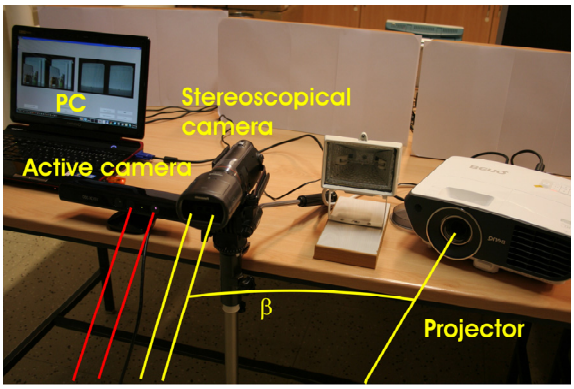

Fig. 9. The laboratory workspace for depth map estimation.

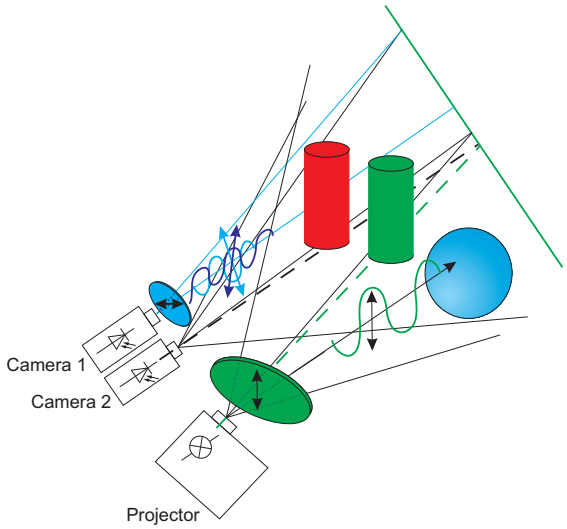

Fig. 10. The scheme of possible setup for depth map estimation by combining of stereo pair matching and profilometry.

the pattern. During profilometry scanning, by using a signal processing, it is possible to separate the measurement pattern from ambient light. This filtered image forms the second image of the stereo pair. This system of measurement pattern separation has been tested and works fairly for metal objects or with metalized surfaces. However, most of dielectric surfaces do not retain polarization of the reflected light.

A wider practical application of such a system could be expected with near-infrared light (NIR) projection. The NIR projector, nowadays quite available, produces the measurement pattern. Its reflection from scanned objects with added ambient visible light is captured by the right camera. The left camera has an IR filter installed to be insensitive to the measurement pattern. The main motivation for the described methods of measurement pattern separation is movement in the scene. For static scenes, time multiplex can be sufficient for separation of the measurement pattern and the image itself. In such a way, the results presented below have been collected.

5.1 The True Depth Map

Comparison of the results of the proposed combined method and the results of individual sub-methods is described in the following. For rating the efficiency of the new method, the true (exact) depth map is needed.

For this purpose, an experimental scene has been designed and its accurate 3D model (with potential deviation less than 0.05%) has been prepared in MATLAB. Based on known intrinsic and extrinsic parameters of the real camera, perspective projection on Camera 1 sensor plane has been computed (Fig. 11 a). The true depth map (Fig. 12 a) has been also calculated from the precise 3D model, as the distance of the modeled object surface to the virtual camera's focal plane. The achievable accuracy of the 3D model and the true depth map derived from it is the main reason why quite a simple scene has been chosen for this experiment.

5.2 Metrics for Depth Map Error Estimation

In general, the depth map is a function which maps the pixels of the image into a 3D surface (generally discontinuous):

$$DM_{\mathrm{A}} = \mathrm{f}\left(x,y\right) = \mathrm{f}\left(\vec{\mathbf{X}}\right), \vec{\mathbf{X}} \in \mathbf{R} \tag{7}$$

where **R** is the space in the coordinate system of the original image, where the original image and also the depth map are placed. Output values of the function are depth values for a particular camera setup. These values should be in units of length, expressing the distance from the camera focus plane orthogonally to the mapped point on the object surface. This particular depth map is referred to as absolute values (DM_{A}). For later processing (e.g. compression, etc.) and TV broadcasting, it is not important to preserve the information about absolute depth and scale. All of the following realizations with arbitrary real coefficients a, b can be considered as true depth maps:

$$DM_{\mathrm{A}} = a \cdot DM_{\mathrm{R}} + b = a \cdot \mathrm{f}\left(\vec{\mathbf{X}}\right) + b,\ \vec{\mathbf{X}} \in \mathbf{R},\ a,b \in \mathbb{R}, \tag{8}$$

where DM_{R} is depth map in relative scale.

The proposed method has inbuilt segmentation to n blocks with continuous surface (objects, see Sec. 4.2) and background $\mathbf{R}_{\mathrm{R}}$.

$$\begin{aligned} &\mathbf{R}_1 \cup \mathbf{R}_2 \cup \ldots \cup \mathbf{R}_n \cup \mathbf{R}_{\mathrm{R}} = \mathbf{R} \\ &\mathbf{R}_j \cap \mathbf{R}_k,\ j \neq k. \end{aligned} \tag{9}$$

Profilometry scanning provides a set of depth maps of each object surface DM_i while coefficients a_i, b_i are obtained from information provided by the conventional depth map estimator (from stereo pairs)

$$\begin{aligned} &DM_1 = a_1 \cdot \mathrm{f}_1\left(\vec{\mathbf{X}}\right) + b_1,\ \vec{\mathbf{X}_1} \in \mathbf{R}_1 \\ &\ldots \\ &DM_n = a_n \cdot \mathrm{f}_n\left(\vec{\mathbf{X}}\right) + b_n,\ \vec{\mathbf{X}} \in \mathbf{R}_n \\ &a_i, b_i \in \mathbb{R},\ i = 1,2,\ldots,n. \end{aligned} \tag{10}$$

The resulting map combines information from two methods and their inaccuracies influence the final values. The described combination of the methods assumes the condition $\mathrm{f}_1 = \mathrm{f}_2 = \ldots = \mathrm{f}_n = \mathrm{f}_{\mathrm{true}}$, where functions $\mathrm{f}_1 \ldots \mathrm{f}_n$ are just windowed parts (for sets $\mathbf{R}_1,\ldots,\mathbf{R}_n$) of the true depth map mapping function $\mathrm{f}_{\mathrm{true}}$. The error of this assumption is caused by the error from profilometry scanning. The second source of the error is a premise that the stereo pair matching provides true information about minimum and maximum depth value for each object even if it does not have enough information about the surface. As shown further, this claim is not true, because both sets of coefficients (a_i, b_i, $i = 1,2,\ldots,n$) are set at the base of inexact prerequisite. Both errors are multiplied, which is one of the disadvantages of the proposed combination of methods, (Sec. 4).

We have used two objective methods for depth map evaluation. An objective method means that the influence of incorrectness in the depth map on the stereo/multi-view Quality of Experience (QoE) was not determined. In the first method, the mean values of depth for each segment $\mathbf{R}_i$ in the evaluated depth map are compared with the true depth map. In the second method, the minimum mean square error (*MMSE*) between the evaluated depth map (DM_{E}) and the true one (DM_{T}) is found as follows:

$$\begin{aligned} &MSE = \frac{1}{\mathbf{R}_{\mathrm{E}} \cap \mathbf{R}_{\mathrm{T}}} \sum_{\vec{\mathbf{X}} \in \mathbf{R}_{\mathrm{E}} \cap \mathbf{R}_{\mathrm{T}}} \left(DM_{\mathrm{E}} - DM_{\mathrm{T}}\right)^2, \\ &MSE = \mathrm{E}\left\{\left(a_n \cdot \mathrm{f}_{\mathrm{E}}\left(\vec{\mathbf{X}}\right) + b_n - \mathrm{f}_{\mathrm{T}}\left(\vec{\mathbf{X}}\right)\right)^2\right\}, \\ &MMSE = MSE\left\{DM_{\mathrm{E}} - DM_{\mathrm{T}}\right\} \Big|\ a,b. \end{aligned} \tag{11}$$

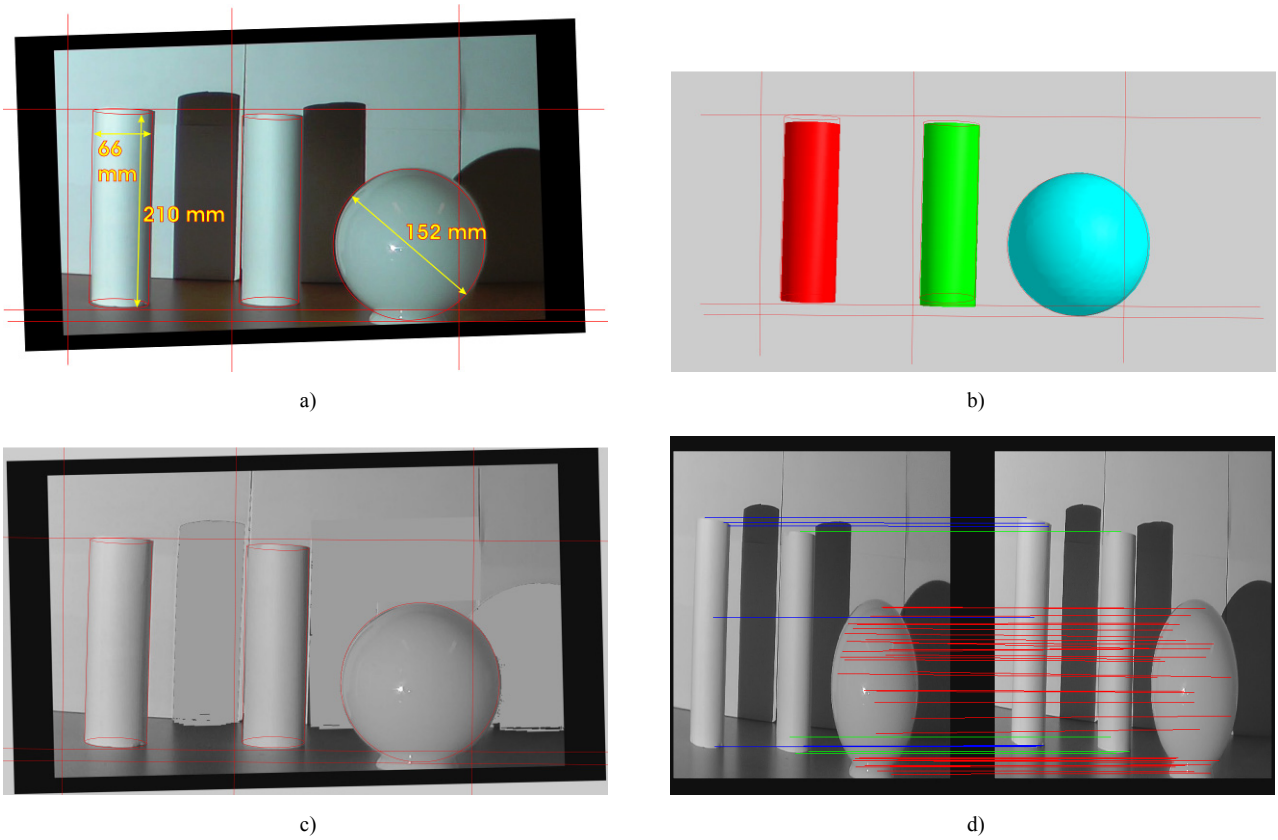

Fig. 11. The image of the scene: a) captured by the left half of the stereo camera, b) captured from the model in MATLAB, c) after removal of the shadow, d) with highlighted corresponding points in the stereo pair.

The differences (*MSE*) between DM_E and DM_T on sets of object surface points (excluding background) are calculated. Symbol |.| means cardinality of a (countable) set. Coefficients a, b represent multiplicative, respectively additive parameters from both depth maps. The *MMSE* is then the global minimum of the function $MSE(a, b)$.

5.3 Alternative Finding of Mean Depth Value

The proposed system for depth map generation, as described above, is very sensitive to the accuracy of each object's extreme depth map values.

The first improvement which can suppress this drawback is the usage of 5% and 95% quantiles of depth values distribution for (a_i, b_i, $i = 1,2,\ldots,n$) calculation (10), instead of negative, respectively positive peak value. This approach filters extreme values which can occur due to noise on edges or by inaccurate object segmentation.

If the camera and the projector are focused to infinity (see Fig. 3), the multiplicative factors of the depth map's segments can be assumed to be the same for all sets $\mathbf{R}_i$ in profilometry scanning, i.e. $a_1 = a_2 = \ldots = a_n$. Then there is no need to search for multiplicative factors and only additive factors b_i need to be found from mean depth values. In this case, errors surely increase with decreasing focal lengths and also with differences in b_i.

The experiments have shown that better data on mean depth are needed than those provided by the conventional implementation of depth from stereo pair matching (by SW Triaxes Stereo Tracker [20], [21], Fig. 12 c). That's why horizontal parallax of corresponding points has been used to estimate mean values of depths. Scale-Invariant Feature Transform (SIFT) is the known method which provides 128-dimension features for specific image points. These features are invariant or "almost" invariant to many image geometrical transformations and they are also useable for finding corresponding points in a stereo pair.

In this work, the implementation from the free MATLAB toolbox, described in [20] was used. The corresponding points of both halves of the stereo pair, laying on the object's surface and simultaneously having high probability of correspondence, are shown in Fig. 11 d).

6. Comparison of Various Methods for Depth Map Generation

Examples of the resulting depth maps can be seen in Fig. 12. As mentioned above, the first map (Fig. 12 a) is the true depth map which has been computed as the perspective projection of a 3D model (Sec. 5.1).

Figure 12 b) presents the depth map provided by the

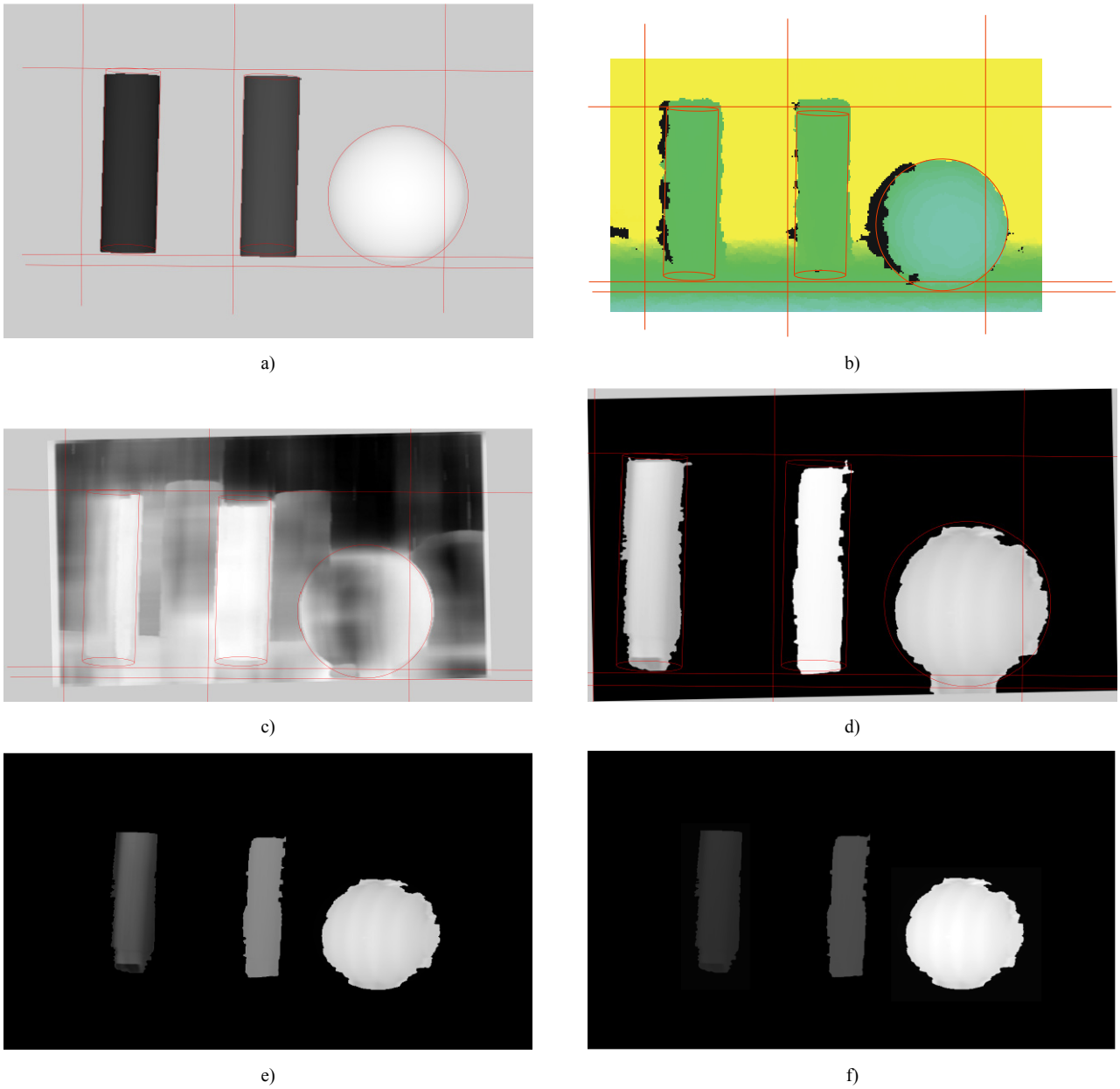

Fig. 12. The depth map of the scene: a) the true depth map computed from MATLAB model, b) captured by Kinect, c) computed from a stereo pair matching only, d) obtained by combining stereo pair matching with profilometry scanning, e) obtained by combining SIFT with profilometry scanning, f) obtained from ideal mean depth and profilometry.

professional device Kinect. This depth sensor maps a 16-bit dynamical range of depth to three 8-bit color components. For further processing, only 3 parts of a dynamic range are used wherein surfaces of objects lay.

The depth map from the stereo pair matching (provided by SW Triaxes Stereo Tracker [20], [21]) is shown in Fig. 12 c). The figure illustrates shortcomings of this sub-method with estimating depth caused by problems with finding correspondences. The obtained values are acceptable around edges, but the algorithm obviously fails in almost all monochromatic areas. Unfortunately, this failure is not cured by the combination with profilometry scanning and the error of the passive method manifests also in the final depth map. The result affected by these dynamic range errors can be observed in Fig. 12 d).

Relative mean values of depth for 3 objects in the depth map	
True depth map	1.0000 : 1.5388 : 5.0874
Kinect	1.0852 : 1.0000 : 1.2329
From stereo pair matching	1.0000 : 1.1188 : 1.0201
SIFT algorithm	1.0000 : 1.9836 : 5.7730

Tab. 1. The relative mean values of depth for 3 scanned objects (red cylinder $\mathbf{R}_1$, green cylinder $\mathbf{R}_2$, cyan sphere $\mathbf{R}_3$).

MMSE		
Stereo pair matching only	0.1554	Fig. 12 c)
Kinect	0.1430	Fig. 12 b)
Ideal mean depth + profilometry	0.0032	Fig. 12 f)
Stereo pair matching (Triaxes) + profilometry	0.1740	Fig. 12 d)
SIFT + profilometry	0.0294	Fig. 12 e)

Tab. 2. Minimum mean square error (MMSE) of the estimated maps relative to the true depth map. The last two rows demonstrate the influence of two different implementations of the passive method giving the mean depth.

As depicted in Fig. 12 e), much better depth map is obtained if profilometry scanning is combined with parallaxes from SIFT. The corresponding points have been chosen from SIFT significant points at the base of three parameters. Firstly, the pairs with minimal Euclidean distance of their SIFT feature values have been chosen as corresponding points. Secondly, the corresponding points have been chosen according to the fact that they have to belong to the same set $\mathbf{R}_i$, and thirdly, according to the fact that straight lines for all corresponding points' pairs should be parallel (in the case of rectified images they are parallel and exactly horizontal).

Figure 12 f) is presented just for comparison. The depth map of the scene is obtained from the profilometry scanning combined with accurate information about the objects' mean depths (ideal coefficients b_i applied).

6.1 Final Score

In this subsection, the benefits of the proposed system are demonstrated by comparing its results with those of individual methods. Furthermore, we also show competitiveness with the commercial depth sensor Kinect.

Table 1 compares the ratio among mean depths of three scanned objects $\mathbf{R}_{1\div3}$ (colored by red, green and cyan in the 3D model image, shown in Fig. 11 a,b). The biggest deviation from the true depth map can be observed if the commercial implementation of stereo pair matching provided by SW Triaxes Stereo Tracker [20], [21] is applied. The algorithm is not suitable even to order objects correctly. The Kinect device also has a problem with this basic task. This is due to the intentional setting for scanning within a very small part of its dynamic range. However, it has to be mentioned that there is less error compared to the map from a stereo pair. It is predictable that the proposed combination of depth from stereo pairs with profilometry suffers from the same problem. The results from the parallax of corresponding points provided by SIFT with subpixel accuracy are presented in the last row of Tab. 1. These results are the best estimation of mean depth of objects from the tested primary methods.

Table 2 sums up the *MMSE* values of the particular depth maps relative to the true one. The first and second rows are calculated from the maps resulting from the stereo pair matching and Kinect. The last three rows represent errors in the case of depth map combinations. Ideal mapping of the depth maps of segments obtained from profilometry scanning to the true depth dynamical range is performed and described by the error value in the third row. This value also determines the minimum achievable error of our setting of the profilometry scanning system. The fourth line of Tab. 2 gives the results obtained from the original version of the proposed method (Sec. 4) combining the commercial implementation of stereo pair matching provided by SW Triaxes Stereo Tracker [20], [21] with profilomery scanning. As explained above, this combination suffers from vague inputs resulting from stereo pair matching. The last value in the fifth row of Tab. 2 refers to an alternative source of mean depth value obtained by SIFT (see Sec. 5.3). This result is obviously the best and demonstrates the contribution of the proposed combination of individual sub-methods, introduced in this paper.

7. Conclusion and Future Work

This paper, in detail, describes the combination of two depth map constructing methods and compares this combination with the results of the commercial depth sensor Kinect. Our method has been tested in a laboratory environment to prove better results than partial methods and the competitiveness with a contemporary depth camera. From various scanned scenes, a simple one has been chosen to demonstrate the obtained results, to compare them mutually and also with exactly defined real data. A significant improvement has been achieved by the proposed combination of the profilometry scanning with the stereo pair matching with SIFT.

In our future works, we would like to modify system parameters for instances with movement within the scene and a moving camera. For dynamic scenes, the time multiplex of the depth scanning method should be replaced by a different mentioned multiplexing method. Near infrared projection of a measurement pattern seems to be promising. It also solves the problem with ambient light conditions and shifts the proposed combined method from the laboratory to the practical usage. It could work sufficiently almost in the whole dynamic range of used cameras.

Nevertheless, in practice, a price of a device would definitely be an important aspect. So, to avoid utilization of the NIR projector, the system with dichroic filters in visible light range could be tested to separate measurement patterns from ambient light. Maybe, also time-multiplexed scanning methods could be further used even in scenes with moving objects or cameras, if the scanning rate is increased sufficiently. Anyway, further analyses, computations and testing are planned to refine the proposed com-

bined method of depth map estimation, to adapt it to moving scenes, to judge its feasibility, its advantages and drawbacks under various conditions, and last but not least to take into account possible economical aspects of its practical applicability.

Acknowledgments

This work was supported by the Ministry of Education, Youth and Sports of the Czech Republic (MEYS) under the national projects no. LD12005 and LD15020, and by the BUT project no. FEKT-S-14-2177. The research described in this paper was financed by the MEYS in the frame of the National Sustainability Program under grant LO1401. For research, infrastructure of the SIX Center was used.

References

[1] HERAKLEOUS, K., POULLIS, C. 3D UNDERWORLD-SLS: *An Open-Source Structured-Light Scanning System for Rapid Geometry Acquisition*. Immersive and Creative Technologies Lab, Cyprus University of Technology, June 26, 2014, ICT-TR-2014-01

[2] LEE, CH., SONG, H., CHOI, B., HO, Y-S. 3D scene capturing using stereoscopic cameras and a time-of-flight camera. *IEEE Transaction on Consumer Electronics*, 2011, vol. 57, no. 3, p. 1370–1376. DOI: 10.1109/TCE.2011.6018896

[3] JAVIDI, B., OKANO, F. *Three-Dimensional Imaging, Visualisation and Display*. 1st ed. New York: Springer, 2009. ISBN: 978-0-387-79334-4

[4] OZAKTAS, H., ONURALL, L. *Three-Dimensional Television: Capture, Transmission, Display (Signals and Communication Technology*). 1st ed. New York: Springer, 2007. ISBN: 978-3-540-72531-2

[5] FEHRMAN, B., MCGOUGH, J. Depth mapping using a low-cost camera array. In *2014 IEEE Southwest Symposium on Image Analysis and Interpretation* (SSIAI). San Diego (CA, USA), 2014, p. 101–104. DOI: 10.1109/SSIAI.2014.6806039

[6] ZELLER, N., QUINT, F., STILLA, U. Establishing a probabilistic depth map from focused plenoptic cameras. In *International Conference on 3D Vision (3DV 2015)*. Lyon (France), 2015, p. 91–99. DOI: 10.1109/3DV.2015.18

[7] GALABOV, M. 3D Capturing with monoscopic camera, *Radioengeneering,* 2014, vol. 23, no. 4, p. 1208–1212.

[8] ETSI TS 101 547-3 V1.1.1: *Digital Video Broadcasting (DVB); Plano-stereoscopic 3DTV; Part 3: HDTV Service Compatible Plano-stereoscopic 3DTV*. [Online] Cited 2016-05-16. Available at: http://www.etsi.org/

[9] KAMENCAY, P., BREZNAN, M., JARINA, R., LUKAC, P., ZACHARIASOVA, M. Improved depth map estimation from stereo images based on hybrid method. *Radioengeneering,* 2012, vol. 21, no. 1, p. 70–78.

[10] AABED, M., TEMEL, D., ALREGIB, G. Depth map estimation in DIBR stereoscopic 3d videos using a combination of monocular cues. In *Conference Record of the Forty Sixth Asilomar Conference on Signals, Systems and Computers (ASILOMAR).* Pacific Grove (CA, USA), 2012, p. 729–733. DOI: 10.1109/ACSSC.2012.6489108

[11] KALLER, O., BOLECEK, L., KRATOCHVIL, T. Profilometry scanning for correction of 3D images depth map estimation. In 2011 *Proceedings of the 53rd Symposium ELMAR*. Zadar (Croatia), 2011, p. 119–122.

[12] PAGÈS, J., SALVI, J., GARCÍA, R., MATABOSCH, C. Overview of coded light projection techniques for automatic 3D profiling. In *Proceedings of IEEE International Conference on Robotics and Automation ICRA '03*. Girona (Spain), 2003.

[13] SU, X., CHEN, W. Fourier transform profilometry: a review. *Optics and Lasers in Engineering*, 2001, vol. 35, no. 5, p. 263 to 284. DOI: 10.1016/S0143-8166(01)00023-9

[14] MOORE, A. J., MENDOZA-SANTOYO, F. Phase demodulation in space domain without a fringe carrier. *Optics and Laser in Engineering*, 1995, vol. 23, no. 5, p. 319–330. DOI: 10.1016/0143-8166(95)00037-O

[15] ANDERSEN, M. R., JENSEN, T., LISOUSKI, P., MORTENSEN, A. K., HANSEN, M. K. Kinect depth sensor evaluation for computer vision applications. *Technical Report ECE-TR-6*, Department of Engineering – Electrical and Computer Engineering, Aarhus University, Denmark

[16] FREEDMAN, B. *Distance-Varying Illumination and Imaging Techniques for Depth Mapping. United States Patent Application* Publication US2010/0290698 A1. [Online] Cited 2010-11-18. Available at: http://www.patentlens.net/imageserver/getimage/US_2010_0290698_A1.pdf?id=23222535&page=all

[17] LIN, H.Y., HUANG, P.K., LIN, T.Y., et al. Stereo matching architecture for 3D pose/gesture recognition and distance-measuring application. In *2013 International Conference on 3D Imaging (IC3D)*. Liege (Belgium), 2013, 6 p. DOI: 10.1109/IC3D.2013.6732095

[18] SMISEK, J., JANCOSEK, M., PAJDA, T. 3D with Kinect. In *Proceedings of IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*. Barcelona (Spain), 2011, p. 1154 to 1160. DOI: 10.1109/ICCVW.2011.6130380

[19] BOLECEK, L., RICNY, V. MATLAB detection of shadow in image of profilometry. In *Proceedings of Technical Computing Prague (TCP)*. Prague (Czech Republic), 2011, p. 22–30.

[20] VEDALDI, A., FULKERSON, B. *VLFeat: An Open and Portable Library of Computer Vision Algorithms*. [Online] Cited 2016-05-16. Available at: http://www.vlfeat.org/

[21] TRIAXES, *Stereo Tracker Version 7 User Guide*. Triaxes Lab LLC, Tomsk (Russia). [Online] Cited 2016-05-16. Available at: http://www.triaxes.com/

[22] JAN, J. *Medical Image Processing, Reconstruction and Restoration - Concepts and Methods*. (Signal Processing and Comm.) Boca Raton (FL, USA): CRC Press, Taylor and Francis Group, 2006. ISBN: 0-8247-5849- 8

About the Authors …

Ondrej KALLER was born in Frydek-Mistek, Czech Republic in 1986. He received his master degree from the Faculty of Electrical Engineering and Communication, Brno University of Technology (BUT), in 2010. Currently he is a PhD. student at the Department of Radio Electronic (DREL), BUT. His field of interest includes digital television broadcasting systems. He is focused on 3D video capturing, transmission, interpretation and evaluation.

Libor BOLECEK was born in Sternberk, Czech Republic, in April 1985. He graduated from the Faculty of Electrical Engineering and Communication (FEEC), Brno University of Technology (BUT), in 2010. The field of his interest includes image processing, quality evaluation and photo-stereometric systems.

Ladislav POLAK was born in Sturovo, Slovakia in 1984. He received the M.Sc. degree in 2009 and the Ph.D. degree in 2013, both in Electronics and Communication from the Brno University of Technology (BUT), Czech Republic. Currently he is an assistant professor at the Department of Radio Electronic (DREL), BUT. His research interests are Digital Video Broadcasting (DVB) standards, wireless communication systems, signal processing, video image quality evaluation and design of subjective video quality methodologies. He has been an IEEE member since 2010.

Tomas KRATOCHVIL was born in Brno, Czech Republic, in 1976. He received the M.Sc. degree in 1999, Ph.D. degree in 2006 and Assoc. Prof. position in 2009, all in Electronics and Communications from the Brno University of Technology. He is currently an associated professor at the Department of Radio Electronics, Brno University of Technology. His research interests include digital television and audio broadcasting, its standardization and video and multimedia transmission including video image quality evaluation. He has been an IEEE member since 2001.