On Tablet 3D Structured Light Reconstruction and Registration

Matea Đonlić   Tomislav Petković

Tomislav Pribanić

University of Zagreb Faculty of Electrical Engineering and Computing

Unska 3, HR-10000 Zagreb, Croatia

{matea.donlic, tomislav.petkovic.jr, tomislav.pribanic}@fer.hr

Abstract

One of the very first tablets with a built-in DLP projector has recently appeared on the market while smartphones with a built-in projector have been available around for quite a while. Interestingly, 3D reconstruction solutions on mobile devices never considered exploiting a built-in projector for the implementation of a powerful active stereo concept, structured light (SL), whose main component is a camera-projector pair. In this work we demonstrate a 3D reconstruction framework implementing SL on a tablet. In addition, we propose a 3D registration method by taking the advantage in a novel way of two commonly available sensors on mobile devices, an accelerometer and a magnetometer. The proposed solution provides robust and accurate 3D reconstruction and 3D registration results.

1. Introduction

Various available 3D measurement devices use different technologies depending on their application field. Optical methods, and within those structured light (SL) strategies, are one of the most popular ones since they are offering very robust, non-invasive and highly accurate 3D measurements [4, 32]. In brief, SL is a representative of active stereo consisting of a projector and a camera. A projector illuminates an object of interest with a certain number of patterns which are subsequently captured by the camera. The patterns carry a particular code and the decoding of the code in the captured images enables 3D triangulation and reconstruction of object's surface. Unfortunately, almost all objects require multiple 3D scans using more than one spatial view to obtain the 3D reconstruction of the whole surface. These separate scans must be brought together into a common coordinate system in order to seamlessly join them. This is the task of 3D surface registration [40] which finds transformations to the common coordinate system. For rigid objects such transformations have six degrees of freedom: a three component translation vector and three rotation angles.

Typically a 3D registration on mobile devices is carried out either through the SLAM pipeline or combining outputs of magnetometer, accelerometer and gyroscope. The translation estimates coming from the inertial sensor is usually acquired by the double integration of accelerometer data and a rotation estimate is usually extracted from the gyroscopes by integrating angular speed during a dynamic movement. In practice this usually requires rather brisk (dynamic) movement of the smartphone in order to have reliable data for the integration [41]. Such manipulation with 3D device is not very convenient. Also, 3D reconstruction solutions which relies on the SLAM typically require processing a large number of captured images which is done in the cloud or on the GPU. Some of the first smartphone 3D solutions have been essentially based on some form of passive stereo [35], which unfortunately has a number of disadvantages compared to active stereo, particularly in the absence of sufficient object surface texture.

Obviously it would be preferable to have a reconstruction and registration method for a mobile device (tablet or smartphone) which requires fewer sensors for registration, does not require brisk movements, process only frames from very different views instead of processing all frames in-between and for 3D data computation itself uses one of the most accurate and robust principle—structured light.

We have noticed, relatively recently available on the market, Lenovo tablet with the built-in projector: Yoga Tab 3 Pro. As it will be shown, this tablet can be used as standalone 3D scanning mobile device (with no need for attaching additional equipment [28, 38]) since we choose a convenient setting where the Yoga tablet's camera and projector do have a common FOV. Additionally we point out the fact that nowadays tablets (and smartphones) have a number of additional sensors that can be readily used to solve 3D registration problem. This work builds on the work of [27] where the following was emphasized: typical 3D registration approaches explicitly search for all six unknown registration parameters whereas a tablet sensors, accelerometer and magnetometer, can be used to first initialize rotation,

allowing cleverly to focus the search for translation only. Thus, effectively reducing the search space from six to three unknowns. However, unlike [27] where a smartphone was used to solve the registration problem of a typical 3D SL scanner, this work takes it to a next level.

We demonstrate a full tablet based solution including both 3D SL reconstruction and 3D registration. First we show the effectiveness of one the most popular SL strategies, so called multiple phase shifting (MPS), on the Yoga Tab 3 Pro tablet having projector with the features still behind typical projectors. Next, we describe a 3D reconstruction approach which imitates IR speckle (random dots) pattern which is used in some of the most popular mobile 3D scanning devices such as Microsoft's Kinect v1 device, Occipital's Structure Sensor [38] or Google's Project Tango [30]. Considering that none of the aforementioned companies of nor anyone else openly published a complete and step by step solution how random dots pattern can be used for 3D reconstruction, we describe in full our proposal. Therefore, essentially showing that our proposal is applicable in the case of Kinect v1 approach too. Third, we propose a novel registration algorithm to register large data sets in a timely efficient manner. This efficiency is gained by uniquely transforming the typical search from a spatial domain, into frequency domain.

The remainder of this paper is structured as follows: Section 2 gives a brief overview of related work. Section 3 describes the proposed method in detail. Results and Discussion are presented in Section 4. We conclude in Section 5.

2. Related Work

2.1. 3D Registration

3D scanning system may be attached to various robot arms and/or turntables [20, 2]. Such hardware readily provides 3D registration parameters. Apart from the evident increase in the system cost, robotic manipulators are usually installed in a dedicated environment requiring the object to be brought in which is not always convenient or possible, e.g. large objects, excessive weight objects [16]). Therefore, purely software based 3D registration solutions have been proposed as well, roughly categorized as coarse and fine surface registration methods [33]. The former requires no (explicit) initial solution and, hopefully, eventually outputs an estimate that is good enough for some fine registration methods to start with. Concerning fine registration methods, iterative closest point (ICP) is nowadays considered a standard for general-purpose fine registration. Since the earliest ICP proposal [3], there have been many efficient variants studies [31], and they are successfully applied in specific applications [22]. Some ICP variants use uniform [44] or random [31] sampling of the data, or use multi-

resolution approach [11] to speed up registration. Others [7, 39] employ color information to reduce the search space in the ICP point correspondence process. However, for ICP to be successful it requires a rather good initial solution, commonly provided manually or by some coarse registration based on feature points [6].

The majority of published algorithms, except for a few PCA based exceptions [5], assume first a step of selecting a set of candidate solutions, followed by the step of the detection of the optimal candidate solution within a candidate set. Due to the intrinsic combinatorial complexity of the problem both steps are likely to be memory and time demanding. Consequently, a central issue in most approaches is time efficiency where many solutions use the appropriate data structures [45]. Efficiency is particularly challenging in the case of genetic-based algorithms which are in addition known to have problems such as defining clearly a criteria to stop iterating [34]. Approaches based on the shape descriptors tend to improve time efficiency measuring the local shapes of the sets around each point [21]. Some methods relax the problem by assuming only a certain type of environment (e.g. objects having straight edges and planar regions), at the same time inherently being restricted to a particular type of application [37]. On the other hand, there are methods which expect not only 3D point position data, but also the normal vectors of every 3D point [18]. Another generic way to increase the efficiency of any parallelized method is to make an additional implementation effort (e.g. require having NVIDIA CUDA enabled device) by providing GPU method implementation [24].

In terms of 3D registration the most similar work to ours is [27]. However, in that work no tablet or smartphone with a built-in projector was used. Only a common smartphone without a built-in projector was used and it was merely for 3D registration task whereas the 3D reconstruction was carried out using a standard 3D SL hardware components. Besides, [27] computes the translation from a set of candidates by evaluating each candidate using the nearest neighbor approach. On the other hand, we estimate the translation from a single search step by transforming the search in the frequency domain. Some methods [17, 47] use frequency domain to compute phase correlation and obtain both rotation and translation but with the drawback that the common region between contiguous 3D scans must be known.

2.2. Smartphone and Tablet 3D reconstruction

3D surface reconstruction methods applicable to smartphones and/or tablets may be categorized into passive and active methods.

Most common passive 3D reconstruction solutions require the user to go around the object carefully taking a relatively large number of images [1, 43] that are processed in a SLAM pipeline to extract a 3D shape. Image processing

(e.g. extracting image features and matching them between acquired images) is a central part of such approaches and is to a large extent typically done in the cloud requiring a network connection to upload acquired images. Also those solutions usually take around 10 or more minutes to produce a 3D model (e.g. when scanning a plastic mannequin head [13]). The complete on-smartphone solutions have also been presented [41], but they still require a fairly large number of images (features) to retrieve a coarse 3D shape. In addition, [41, 14] heavily rely on the use of other sensors such as gyroscope and accelerometer. Alternatively, the shape from silhouettes has also been proposed, still creating relatively coarse 3D models [10] and/or models with poor level of details [29].

Besides the passive approaches, some authors realized the possibilities to use a smartphone in the context of active stereo, more specifically using the photometric stereo. The smartphone screen is conveniently used as a light source, however, as noted by the authors themselves, a dark environment is required [46, 42]. Somewhat more robust solution, but at the expense of using an extra smartphone, is proposed in [48]. They used a pair of smartphones which collaborated as master and slave where the slave was illuminating the scene using the flash from appropriate viewing points while the master recorded images of the object.

Considering such approaches it become apparent that the lack of an appropriate light source on the smartphone is a substantial obstacle to implementing any form of active stereo. Project Tango by Google is perhaps one of the most well-known examples where in order to overcome that obstacle a custom made IR projector and IR camera are installed in a smartphone [30]. It appears that the principle used behind is presumably similar to Kinect v1 [12, 19], and can be shortly described as follows: a projector illuminates the scene with a random dots IR pattern; the acquired image is then matched against the reference image(s) taken in advance. Since a large part of processing is apparently stereo matching between a pair of images, rather than an explicit computation of some SL code, many do not consider the mentioned principle as type of SL approach. But even more importantly added IR projector may be used only for 3D depth sensing and nothing much beyond.

Some of the more recent work proposed the use of a laser line projector attached to a smartphone [36]. Although conceptually simple, such solution, similarly to all single line laser approaches, requires many images for 3D surface reconstruction. Besides, the authors in [36] impose a constraint that a marker has to be visible and tracked throughout the frames in order to 3D register all those separate laser line segments. Combining a pico-projector with almost any smartphone is always a possibility [26], but evidently not always a convenient solution. Apparently, the only smartphone standalone solution for 3D shape acquisition which
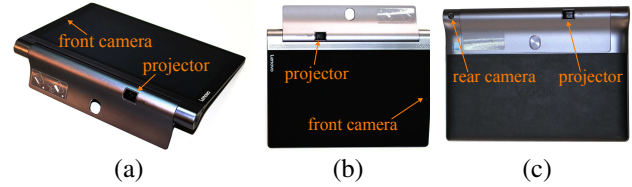

(a)   (b)   (c)

Figure 1. Lenovo Yoga Tab 3 Pro. Built-in projector's position: (a) for the projection of video content; (b) with the hinge completely open; (c) with the hinge completely closed – this position is used for 3D surface scanning.

implemented SL (active stereo) is [28] where a smartphone with a built in projector was used. However such approach requires the use of custom designed adapter to bring smartphone's built in projector and camera into the common field of view.

In the context of mentioned 3D scanning solution, we recall briefly some of the major advantages of our method compared to others:

- no need for Internet connections and requirement for processing taking place outside the smartphone [1];

- no requirement for IR projector [30], but instead we use a visible light projector that can be utilized for other purposes too;

- no extra light sources are needed [36];

- no need to have an extra dark environment [42];

- no need to use and process data from additional sensors (e.g. accelerometer, gyroscope) [41]; and

- no problems with a sparse reconstruction of small objects such as convex/concave surfaces [10].

3. Method Description

We first describe utilized hardware in Section 3.1 followed by the Section 3.2 describing our proposal of using a random dots pattern. Section 3.3 describes the used structured light method and Section 3.4 presents the proposed 3D registration method.

3.1. Hardware Components

Figure 1 shows the Lenovo Yoga Tab 3 Pro tablet which has a 13 MP rear camera, a 5 MP front camera, and a 50 lm DLP projector which is placed on a rotating hinge. The rotating hinge allows some flexibility in positioning the built-in projector with respect to both cameras. In this work we have chosen a repeatable positional setting of the closed hinge (Fig. 1c) where the built-in projector and the rear camera share a common field-of-view.

The camera-projector pair was calibrated using an adapted approach from [23].

3.2. 3D Reconstruction using Random Dots Pattern

We next describe our proposal how a Kinect v1 like random dots (speckle) pattern is processed in order to extract the depth.

Without loss of generality, assume a rectified camera-projector pair such that the epipolary rectified pair of projected random dots pattern and of acquired object image is available. Figure 2 shows the rectified sensors arrangement of our system, assuming that camera's coordinate system coincides with the world coordinate system. For the rectification, we have used an approach described in [8].

Consider a point A on the surface of an object illuminated using a random dots pattern. Let the coordinates of the point A be (x_{CA}, y_{CA}, z_{CA}) in the camera coordinate system, (x_{PA}, y_{PA}, z_{PA}) in the projector coordinate system, (u_{CA}, v_{CA}) in the camera image, and (u_{PA}, v_{PA}) in the projector image. The coordinates of the point A satisfy the following equations:

$$
\begin{aligned}
u_{CA} - u_0 &= \frac{f \cdot x_{CA}}{z_{CA}} \\
v_{CA} - v_0 &= \frac{f \cdot y_{CA}}{z_{CA}} \\
u_{PA} - u_0 &= \frac{f \cdot x_{PA}}{z_{PA}} = \frac{f \cdot (x_{CA} - b)}{z_{PA}} \\
v_{PA} - v_0 &= \frac{f \cdot y_{PA}}{z_{PA}}
\end{aligned}
\tag{1}
$$

In Eq. (1) f is an effective focal length, b is a baseline distance between optical centers $\mathbf{C}_C$ of the camera and $\mathbf{C}_P$ of the projector, and (u_0, v_0) is a principal point of camera and projector sensors within the rectified pair. Due to epipolar rectification $y_{CA} = y_{PA}$ and $z_{CA} = z_{PA}$. Also, the disparity d_A between camera and projector image coordinates simplifies to the difference between horizontal image coordinates only:

$$
d_A = u_{CA} - u_{PA} = \frac{f \cdot b}{z_{CA}}.
\tag{2}
$$

Eq. (2) is a well-known relation for a rectified sensor pair which enables recovery of spatial depth z_{CA} from the disparity d_A.

The disparity d_A is not directly measurable for a projector-camera pair; this is in direct contrast with a stereo camera pair for which the disparity d_A is directly measurable. The crux of the proposed 3D surface reconstruction method is that the disparity d_A between camera and projector image coordinates is never explicitly computed. Instead, consider another point B which lies in a reference plane as shown in Fig. 2 and which has the very same coordinates in the projector image as the point A (so $u_{PA} = u_{PB}$ and $v_{PA} = v_{PB}$). Then $v_{CA} = v_{CB}$ due to stereo rectification [8].

The disparity d_B for the point B is:

$$
d_B = u_{CB} - u_{PB} = \frac{f \cdot b}{z_{CB}}.
\tag{3}
$$

Difference of disparities of Eqs. (2) and (3) is another disparity D where all projector image coordinates conveniently cancel out,

$$
D = d_A - d_B = u_{CA} - u_{CB} = \frac{f \cdot b}{z_{CA}} - \frac{f \cdot b}{z_{CB}}.
\tag{4}
$$

The depth z_{CB} of the reference plane and the parameters f and b of the projector-camera pair are known; they are measured during the system calibration. The depth z_{CA} of the point A may then be recovered using

$$
z_{CA} = \frac{z_{CB}}{1 + z_{CB} \dfrac{D}{f \cdot b}},
\tag{5}
$$

provided the disparity D is measured from the image. The disparity D may be computed using any passive stereo matching [15] between the object image and the reference plane image. The object image is recorded during 3D scanning while the reference plane image is recorded once during the system calibration by projecting a random dots pattern on a white reference plane. Note that in the proposed approach the original random dots pattern image which is projected by the projector is not used for stereo matching.

Once depth z_{CA} is computed using Eq. (5) the other two coordinates x_{CA} and y_{CA} are found using the first two expressions from Eq. (1).

3.3. 3D Reconstruction using Structured Light

The basic principle of SL approach can be summarized as follows: a projector projects a certain number of images on the object of interest. The projected images have a particular structure, a code, which can be decoded on the acquired camera images. 3D position can be triangulated from the decoded SL code.

Among more than a dozen different SL patterns we have chosen one of the time multiplexing strategies, a well known phase shifting (PS) method [32]. PS consists of projecting a number ($N \geq 3$) of periodic sine pattern, shifted by some period amount. The patterns are sequentially projected with a projector on the object of interest, are recorded by the camera, and are then processed in order to compute a wrapped phase map. Due to the periodic nature of sine patterns, the wrapped phase map does not provide a unique code, rather it is said that the code is wrapped within $\langle -\pi, +\pi]$ interval. One way to unwrap the wrapped phase map and recover the SL code is to project additional PS patterns having a different number of periods compared to the first set. Such multiple phase shifting (MPS) procedure
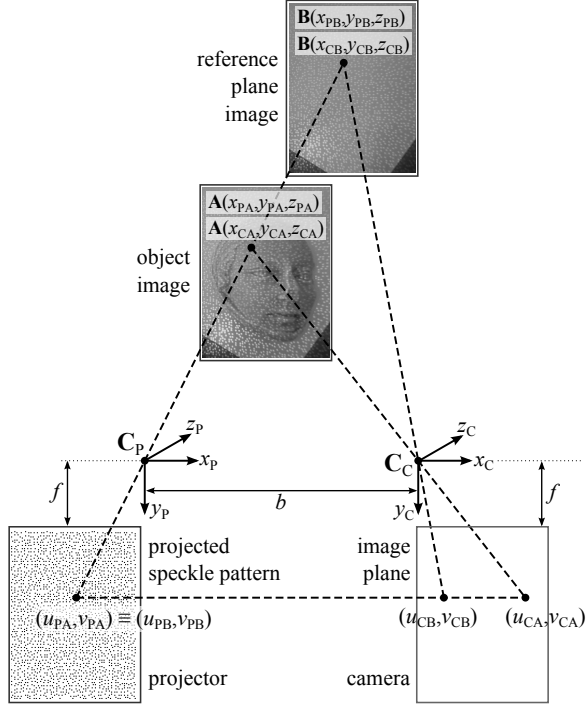
Figure 2. The principle of using a random dots pattern for 3D surface reconstruction. Points A and B are 3D points on the object and on the reference plane, respectively. Without loss of generality, points A and B correspond to the same point in the projector image space but they project to the different points in the camera image plane producing the disparity D. See text for more details.

provides two wrapped values ϕ_{w1} and ϕ_{w2}. Computing the unwrapped phase Φ_{uw} from the ϕ_{w1} and ϕ_{w2} and extracting the SL code can be done in number of different ways; we have followed the method described in [25].

3.4. 3D Registration

Assume two different point clouds $\mathcal{C}_1$ and $\mathcal{C}_2$ are acquired from the two different views of the same rigid body. Then, we propose finding the registration between two views by separately finding the rotational and the translational part of the transformation.

Rotation The rotation to align the second view to the first is found using the following steps:

Step 1. Given available accelerometer and magnetometer measurements we can compute the rotation matrix $\mathbf{R}_S$, describing tablet rotation (i.e. its Android's coordinate system $(X_S; Y_S; Z_S)$) w.r.t. the Earth coordinate system. For two different views we first compute $\mathbf{R}_{S1}$ and $\mathbf{R}_{S2}$ and then the relative orientation between two views as $\mathbf{R}_{S21} = \mathbf{R}_{S1}\mathbf{R}_{S2}^{-1}$.

Step 2. Similarly, we can define the rotation matrix $\mathbf{R}_C$ as the rotation of tablet 3D scanner (i.e. its camera coordinate system $(X_C; Y_C; Z_C)$) w.r.t. to some scanner calibra-

tion coordinate system. Therefore, for two different views we consider $\mathbf{R}_{C1}$ and $\mathbf{R}_{C2}$. The relative orientation between two views (rotational part of the searched registration from the second to the first view) is then given by $\mathbf{R}_{C21} = \mathbf{R}_{C1}\mathbf{R}_{C2}^{-1}$. Unfortunately, $\mathbf{R}_C$'s are not directly computable. However, throughout different views there is a fixed rotational relationship $\mathbf{R}_P$ between two coordinate systems $(X_S; Y_S; Z_S)$ and $(X_C; Y_C; Z_C)$. In turn, the rotation which aligns camera's (3D scanner) second view to the first view can be computed as:

$$\mathbf{R}_{C21} = \mathbf{R}_P \cdot \mathbf{R}_{S21} \cdot \mathbf{R}_P^{-1} \qquad (6)$$

Translation Translation to align the second view to the first may be found only after the rotation is known. To compute the translation we propose the following efficient algorithm:

Step 1. Voxelize each point cloud $\mathcal{C}_1$ and $\mathcal{C}_2$ so $\mathcal{C}_1$ is voxelized directly and so C_2 is voxelized using voxels rotated according to the rotation matrix $\mathbf{R}_{C21}$ given by Eq. (5). The voxelization is performed by constructing a regular three-dimensional grid of voxels; we have used $5\,\mathrm{mm} \times 5\,\mathrm{mm} \times 5\,\mathrm{mm}$ voxels. The voxel value is set to 1 if any point from the point cloud $\mathcal{C}$ falls inside the voxel and to 0 if no points fall inside the voxel. The voxelizations of point clouds $\mathcal{C}_1$ and $\mathcal{C}_2$ are binary multidimensional arrays of finite size; denote them by V_1 and V_2 respectively.

Step 2. The translation vector $\mathbf{t}_{21}$ which best aligns V_2 to V_1 is found by computing the cross-correlation C_{21} between V_2 and V_1; the position of the maximum of C_{21} is $\mathbf{t}_{21}$. We propose solving efficiently this problem in the Fourier transform domain by using the cross-correlation theorem which relates cross-correlation C_{21} and its Fourier transform $\hat{C}_{21}$:

$$C_{21}[x, y, z] = \sum_{a,b,c} V_1^*[a, b, c] \cdot V_2[a + x, b + y, c + z]$$

$$\hat{C}_{21}[\omega_x, \omega_y, \omega_z] = \hat{V}_1^*[\omega_x, \omega_y, \omega_z] \cdot \hat{V}_2[\omega_x, \omega_y, \omega_z] \qquad (7)$$

Let V_1 be of size $N_{1,x} \times N_{1,y} \times N_{1,z}$ and let V_2 be of size $N_{2,x} \times N_{2,y} \times N_{2,z}$. To find $\mathbf{t}_{21}$ both V_1 and V_2 are extended to $(N_{1,x} + N_{2,x} - 1) \times (N_{1,y} + N_{2,y} - 1) \times (N_{1,z} + N_{2,z} - 1)$ by zero-padding. Then the discrete Fourier transforms along each of three dimensions are computed yielding $\hat{V}_1$ and $\hat{V}_2$. Next the complex conjugate of $\hat{V}_1$ is element-wise multiplied by $\hat{V}_2$ yielding the Fourier transform $\hat{C}_{21}$. Applying the inverse discrete Fourier transform to $\hat{C}_{21}$ yields the cross-correlation C_{21}. The position of the maximum of C_{21} is the translation vector $\mathbf{t}_{21}$. The computation in the transform domain effectively reduces complexity from $O(N^4)$ to $O(N^3 \log N)$, where N is the maximum of $N_{1,x}$, $N_{1,y}$, $N_{1,z}$, $N_{2,x}$, $N_{2,y}$ and $N_{2,z}$.

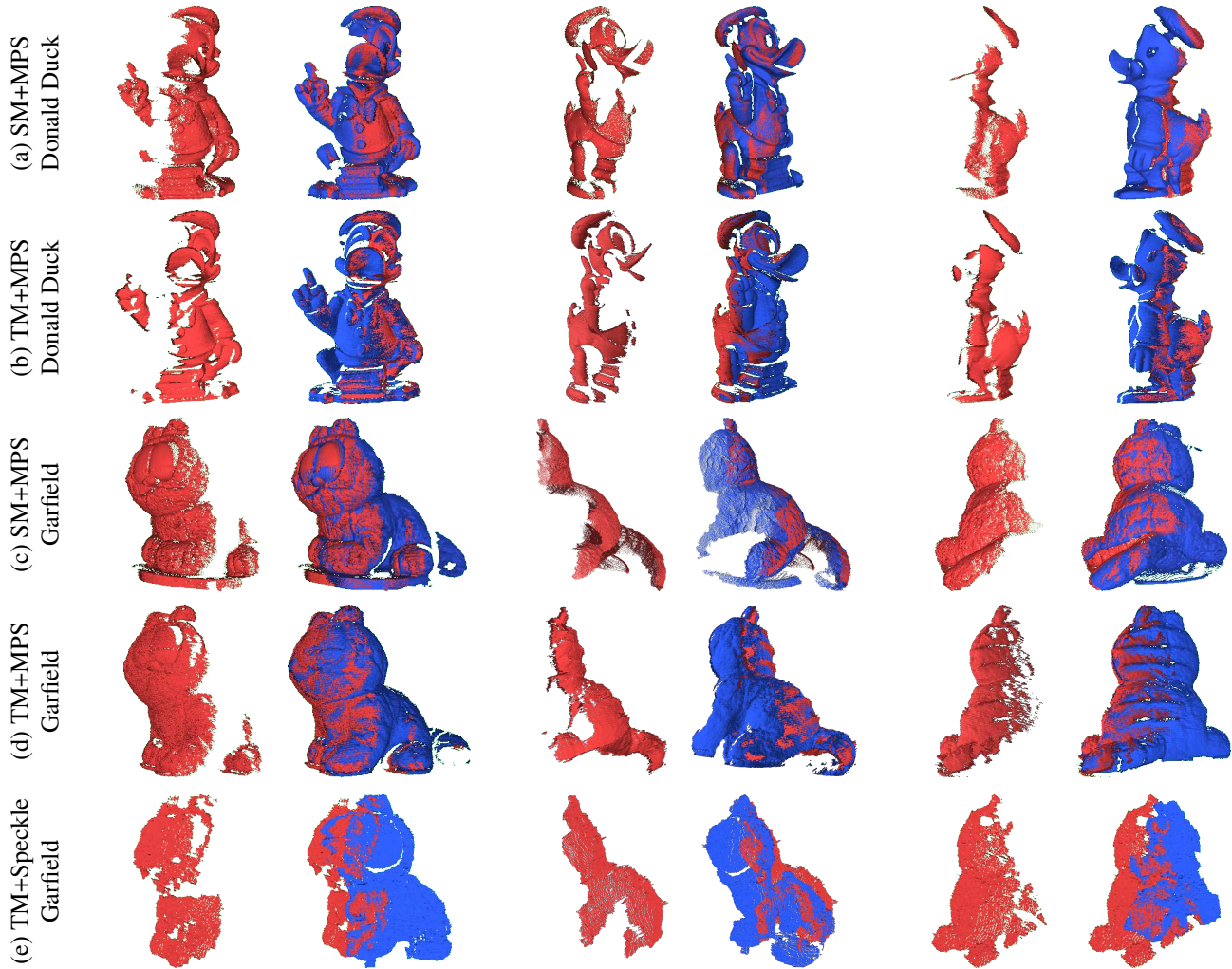Refinement The above computed 3D registration parameters are refined using the ICP.

Figure 3. Example registrations for view pairs. Row labels denote used reconstruction method and object. Odd columns show a single view, even column show two registered views.

4. Results and Discussion

Since our work is largely inspired by [27] we have also implemented that method. Let us recall that in [27] a standard 3D structured light scanner configuration is used for 3D scanning and a low end smartphone having accelerometer and magnetometer is used to implement 3D registration. That implementation will be hereafter referred to as SM method, in contrast to the proposed method using a tablet and which will be hereafter referred to as TM method. We have implemented the SM method using Acer X1260 projector, Point-Grey Dragon-Fly DR2-HICOL camera, and Samsung SM-A310F smartphone.

We compare both TM and SM methods to the ground truth (GT) registration data. The GT data is obtained by first manually selecting very carefully several correspondence points on each considered pair of views, which allowed computing an initial registration solution. Next we

have refined that solution by using ICP that is known to provide a very accurate 3D registration solution, given a good enough initial solution which was certainly satisfied in our case.

In our experiments we have 3D scanned and 3D registered models of Donald Duck and Garfield. Figure 3 shows representative point clouds before and after 3D registration. Each view consisted of ∼15000 points. It is evident that each registered view pair nicely adds new surface parts not present in individual members of a pair. On average the difference in at least one of the rotation angles between views is $57.65° \pm 6.77°$. It is worth noting that when we have applied ICP alone on such pairs of views, it mostly failed to register successfully.

For a qualitative evaluation we provide Table 1 showing rotation and translation errors of SM and TM against GT registration data. The errors are very small and can be attributed to the finite accuracy of 3D reconstruction data

Table 1. Comparison to the ground truth for MPS structured light reconstruction and final surface reconstruction from all view-pairs.

		Rotation – Absolute error [°]						Surface Reconstruction
		SM method			TM method			
		Pitch	*Roll*	*Yaw*	*Pitch*	*Roll*	*Yaw*	
Garfield	Pair 1	−0.49	−0.01	−1.85	0.18	−0.05	0.02	
	Pair 2	0.31	0.31	−0.36	0.17	0.31	0.53	
	Pair 3	−0.69	−0.22	1.29	1.03	0.21	−0.34	
	Pair 4	0.47	−0.59	−1.32	0.17	−2.02	0.08	
	Pair 5	0.16	0.73	1.44	−0.10	0.44	−0.04	
	Pair 6	0.03	−0.50	−0.81	−0.46	−0.04	0.52	
	Pair 7	−0.17	0.14	0.59	0.68	0.40	−0.17	
	Pair 8	0.20	0.17	0.22	−0.02	−0.11	0.21	
Donald Duck	Pair 1	0.23	−0.12	−0.04	0.22	−0.40	−0.01	
	Pair 2	0.07	0.07	−0.21	0.19	0.13	−0.29	
	Pair 3	0.31	−0.22	0.35	−0.15	−0.66	0.36	
	Pair 4	−0.39	−0.07	−0.42	0.30	−0.35	−0.07	
	Pair 5	0.11	0.05	0.37	0.02	0.19	0.15	
	Pair 6	0.24	0.00	0.28	−0.28	−0.21	0.22	
	Pair 7	0.03	0.03	−0.16	−0.08	0.62	−0.08	
	Pair 8	−0.03	−0.06	−0.05	−0.09	0.01	0.02	

		Translation – Absolute error [mm]						Textured Reconstruction
		SM method			TM method			
		T_x	T_y	T_y	T_x	T_y	T_z	
Garfield	Pair 1	0.16	0.93	1.11	0.17	−0.33	−0.12	
	Pair 2	0.49	0.36	−0.22	1.11	−0.65	−1.73	
	Pair 3	−0.99	−0.02	1.03	−2.62	−2.29	−0.91	
	Pair 4	0.89	0.10	0.71	−2.20	−0.41	4.27	
	Pair 5	−0.98	0.39	1.28	1.82	0.24	0.67	
	Pair 6	−0.12	−0.06	0.33	2.04	0.43	0.10	
	Pair 7	0.13	−0.15	0.27	0.54	−1.95	−1.29	
	Pair 8	−0.09	−0.01	0.16	0.11	0.14	−0.15	
Donald Duck	Pair 1	−0.10	0.95	−0.15	−0.75	−0.80	0.36	
	Pair 2	−0.30	−0.52	0.95	−0.24	−0.67	−0.18	
	Pair 3	1.65	3.06	−3.13	0.26	0.56	1.05	
	Pair 4	−3.44	−0.53	−0.46	−0.37	−1.01	0.59	
	Pair 5	1.73	−0.04	−1.98	1.09	0.26	−0.76	
	Pair 6	0.29	0.92	−1.06	0.39	0.78	0.29	
	Pair 7	−0.17	0.10	0.58	1.84	0.27	−1.12	
	Pair 8	−1.21	0.20	−0.02	0.65	0.28	−0.06	

and to the fact that due to a finite 3D reconstruction resolution many points simply do not have a perfect match in the correspondent view unless a certain interpolation on the 3D point cloud is performed. We recall that both methods consist of rotation estimation, translation estimation, and ICP refinement. The TM method is able to estimate translation in only matter of couple tenths of second and therefore is much faster then the SM method, i.e. for the order of magnitude (see Table 2). Finally, both methods require several more seconds for the ICP refinement.

We have used here MPS which is one of the most accurate SL algorithms. In addition we have contributed with our Kinect-like random dots implementation showing 3D registration results not as accurate which is due to a fact that initial 3D reconstruction is less accurate than the MPS approach (Table 3). The inherent feature of straightforward use of the random dots pattern is discretization of the depth, resulting in less accurate 3D reconstruction and con-

Table 2. Duration of translation estimation [s] for SM and TM methods.

		SM method	TM method
Garfield	Pair 1	1.271	0.133
	Pair 2	1.008	0.129
	Pair 3	1.371	0.135
	Pair 4	1.466	0.148
	Pair 5	1.167	0.128
	Pair 6	1.118	0.149
	Pair 7	1.125	0.126
	Pair 8	1.185	0.143
Donald Duck	Pair 1	3.952	0.251
	Pair 2	3.318	0.217
	Pair 3	4.075	0.274
	Pair 4	4.484	0.232
	Pair 5	4.503	0.228
	Pair 6	5.797	0.305
	Pair 7	5.621	0.229
	Pair 8	4.244	0.237

sequently less accurate 3D registration. Additionally, qualitative inspection of Figure 3 confirms the advantage of using MPS over the random dots pattern. Nevertheless we have demonstrated that the proposed registration method works in the case of random dots pattern too.

Our method estimates rotation from only two sensors in a straightforward fashion. Furthermore we point out, it is applicable not only in the case of tablets but in all those applications whenever accelerometer-magnetometer pair is available. Given the accuracy of estimated rotation data using smartphone or tablet sensors, the proposed method generally requires ICP as the final refinement step. However from the end-user's point of-view the introduction of ICP as the final registration step is completely transparent and the proposed method as a whole can be regarded to be the complete 3D registration solution.

Many 3D registration tablet (smartphone) solutions also require a dynamic tracking frame by frame while the pro-

posed TM method takes advantage of only two widely apart views and under much convenient static conditions. TM method complete relies on only accelerometer and magnetometer sensors. Nowadays, almost all smartphones and tablets are equipped with three axis accelerometer and three-axis magnetometer which we use to compute the local orientation. Other solutions tend to take advantage of additional sensors such as gyroscope and GPS. We point out that based on the data search from GSM Arena [9] only around one fifth of smartphones are equipped with gyroscope. Not surprisingly Lenovo tablet model we have used does not have a gyroscope. Nevertheless, Lenovo tablet used here is apparently one of the very first, and one of the still relatively few tablets with a built-in DLP projector, that rather recently have begun to appear on the market. In that sense, the proposed work is a nice example of going along with the newest technologies emerging on the market.

5. Conclusion

To the extent of our knowledge, this is the first demonstration of the 3D structured light scanning which combines 3D reconstruction and registration on a single device (tablet) to obtain a full 3D point cloud model. We have also shown how random dots (speckle) pattern can be modeled and used in 3D reconstruction from a single image. Besides the 3D reconstruction itself, we have proposed a novel 3D registration algorithm. Our algorithm cleverly takes advantage of the two commonly available sensors to estimate accurate rotation parameters. The remaining translation is found efficiently in the frequency domain, even for very large point clouds. As demonstrated with our quantitative and qualitative results, there is a high potential of the proposed approach for the 3D reconstruction and registration using tablet that goes beyond a simple gadget which is otherwise typical for many tablet (or smartphone) applications.

Acknowledgment

Table 3. Comparison to the ground truth for the TM method using random dots pattern.

		Rotation – Absolute error [°]			Translation – Absolute error [mm]		
		Pitch	*Roll*	*Yaw*	T_x	T_y	T_z
Garfield	Pair 1	−0.539	0.583	1.118	3.211	2.271	−2.551
	Pair 2	−0.116	2.319	1.769	4.841	0.885	−9.499
	Pair 3	−0.355	−0.046	1.371	−7.364	2.820	−1.447
	Pair 4	−3.123	2.558	0.096	4.054	6.292	−3.604
	Pair 5	−0.962	2.268	−0.570	5.366	4.415	−2.049
	Pair 6	−1.821	3.610	0.903	4.469	2.285	−5.854
	Pair 7	0.733	3.649	0.552	1.085	−2.513	−11.479
	Pair 8	−0.514	0.785	0.036	3.515	1.437	−0.350

References

[1] 123D Catch. http://www.123dapp.com/catch. [Online; Accessed: November 2016]. 2, 3

[2] Automated 3D Digitisation. http://www-old.igd.fraunhofer.de/en/Institut/Abteilungen/VHT/Projekte/Automated-3D-Digitisation. [Online; Accessed: July 2017]. 2

[3] P. J. Besl and N. D. McKay. A method for registration of 3-D shapes. *IEEE Trans. Pattern Anal. Mach. Intell.*, 14(2):239–256, 1992. 2

[4] F. Chen, G. M. Brown, and M. Song. Overview of three-dimensional shape measurement using optical methods. *Optical Engineering*, 39(1):10–22, 2000. 1

[5] D. H. Chung, I. D. Yun, and S. U. Lee. Registration of multiple-range views using the reverse-calibration technique. *Pattern Recognition*, 31(4):457–464, 1998. 2

[6] Y. Díez, F. Roure, X. Lladó, and J. Salvi. A qualitative review on 3D coarse registration methods. *ACM Comput. Surv.*, 47(3):45:1–45:36, Feb. 2015. 2

[7] S. Druon, M. J. Aldon, and A. Crosnier. Color constrained icp for registration of large unstructured 3d color data sets. In *2006 IEEE International Conference on Information Acquisition*, pages 249–255, Aug 2006. 2

[8] A. Fusiello, E. Trucco, and A. Verri. A compact algorithm for rectification of stereo pairs. *Machine Vision and Applications*, 12(1):16–22, 2000. 4

[9] GSM Arena. http://www.gsmarena.com/search.php3. [Online; Accessed: July 2017]. 8

[10] A. Hartl, L. Gruber, C. Arth, S. Hauswiesner, and D. Schmalstieg. Rapid reconstruction of small objects on mobile phones. In *CVPR 2011 WORKSHOPS*, pages 20–27, June 2011. 3

[11] T. Jost and H. Hugli. A multi-resolution icp with heuristic closest point search for fast and robust 3D registration of range images. In *3-D Digital Imaging and Modeling, 2003. 3DIM 2003. Proceedings. Fourth International Conference on*, pages 427–433, Oct 2003. 2

[12] K. Khoshelham and S. O. Elberink. Accuracy and resolution of kinect depth data for indoor mapping applications. *Sensors*, 12(2):1437, 2012. 3

[13] K. C. Koban, S. Leitsch, T. Holzbach, E. Volkmer, P. M. Metz, and R. E. Giunta. 3D-imaging and analysis for plastic surgery by smartphone and tablet: An alternative to professional systems? *Handchir Mikrochir Plast Chir*, 46(2):97–104, April 2014. 3

[14] K. Kolev, P. Tanskanen, P. Speciale, and M. Pollefeys. Turning mobile phones into 3D scanners. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 3946–3953, June 2014. 3

[15] N. Lazaros, G. C. Sirakoulis, and A. Gasteratos. Review of stereo vision algorithms: From software to hardware. *International Journal of Optomechatronics*, 2(4):435–462, 2008. 4

[16] M. Levoy, K. Pulli, B. Curless, S. Rusinkiewicz, D. Koller, L. Pereira, M. Ginzton, S. Anderson, J. Davis, J. Ginsberg, J. Shade, and D. Fulk. The digital Michelangelo project: 3D scanning of large statues. In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '00, pages 131–144, New York, NY, USA, 2000. ACM Press/Addison-Wesley Publishing Co. 2

[17] L. Lucchese, G. Doretto, and G. M. Cortelazzo. A frequency domain technique for range data registration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(11):1468–1484, Nov 2002. 2

[18] A. Makadia, A. Patterson, and K. Daniilidis. Fully automatic registration of 3D point clouds. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 1, pages 1297–1304, June 2006. 2

[19] M. Martinez and R. Stiefelhagen. Kinect unleashed: Getting control over high resolution depth maps. In *IAPR Conference on Machine Vision Applications*, pages 247–250, 2013. 3

[20] A. F. Martins, M. Bessant, L. Manukyan, and M. C. Milinkovitch. R^2OBBIE-3D, a fast robotic high-resolution system for quantitative phenotyping of surface geometry and colour-texture. *PLoS ONE*, 10(6):1–18, 2015. 2

[21] A. Mian, M. Bennamoun, and R. Owens. On the repeatability and quality of keypoints for local feature-based 3D object retrieval from cluttered scenes. *International Journal of Computer Vision*, 89(2):348–361, 2010. 2

[22] H. Mohammadzade and D. Hatzinakos. Iterative closest normal point for 3D face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(2):381–397, Feb 2013. 2

[23] D. Moreno and G. Taubin. Simple, accurate, and robust projector-camera calibration. In *2012 Second International Conference on 3D Imaging, Modeling, Processing, Visualization Transmission*, pages 464–471, Oct 2012. 3

[24] S.-Y. Park, S.-I. Choi, J. Kim, and J. S. Chae. Real-time 3D registration using GPU. *Machine Vision and Applications*, 22(5):837–850, 2011. 2

[25] T. Petković, T. Pribanić, and M. Đonlić. Temporal phase unwrapping using orthographic projection. *Optics and Lasers in Engineering*, 90:34–47, 2017. 5

[26] M. Piccirilli, G. Doretto, A. Ross, and D. Adjeroh. A mobile structured light system for 3D face acquisition. *IEEE Sensors Journal*, 16(7):1854–1855, April 2016. 3

[27] T. Pribanić, Y. Diez, F. Roure, and J. Salvi. An efficient surface registration using smartphone. *Machine Vision and Applications*, 27(4):559–576, 2016. 1, 2, 6

[28] T. Pribanić, T. Petković, M. Đonlić, V. Angladon, and S. Gasparini. 3D structured light scanner on the smartphone. In A. Campilho and F. Karray, editors, *Image Analysis and Recognition: 13th International Conference, ICIAR 2016, in Memory of Mohamed Kamel, Póvoa de Varzim, Portugal, July 13-15, 2016, Proceedings*, pages 443–450, Cham, 2016. Springer International Publishing. 1, 3

[29] V. A. Prisacariu, O. Khler, D. W. Murray, and I. D. Reid. Simultaneous 3D tracking and reconstruction on a mobile phone. In *Mixed and Augmented Reality (ISMAR), 2013 IEEE International Symposium on*, pages 89–98, Oct 2013. 3

[30] Project Tango. http://get.google.com/tango. [Online; Accessed: July 2017]. 2, 3

[31] S. Rusinkiewicz and M. Levoy. Efficient variants of the ICP algorithm. In *3-D Digital Imaging and Modeling, 2001. Proceedings. Third International Conference on*, pages 145–152, 2001. 2

[32] J. Salvi, S. Fernandez, T. Pribanić, and X. Llado. A state of the art in structured light patterns for surface profilometry. *Pattern Recognition*, 43(8):2666–2680, 2010. 1, 4

[33] J. Salvi, C. Matabosch, D. Fofi, and J. Forest. A review of recent range image registration methods with accuracy evaluation. *Image Vision Computing*, 25(5):578–596, 2007. 2

[34] J. Santamaría, O. Cordón, and S. Damas. A comparative study of state-of-the-art evolutionary image registration methods for 3D modeling. *Computer Vision and Image Understanding*, 115(9):1340–1354, 2011. 2

[35] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47(1):7–42, 2002. 1

[36] R. Slossberg, A. Wetzler, and R. Kimmel. Freehand laser scanning using mobile phone. In X. Xie, M. W. Jones, and G. K. L. Tam, editors, *Proceedings of the British Machine Vision Conference (BMVC)*, pages 88.1–88.10. BMVA Press, September 2015. 3

[37] I. Stamos and M. Leordeanu. Automated feature-based range registration of urban scenes of large scale. In *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, volume 2, pages II–555–Ii–561 vol.2, June 2003. 2

[38] Structure Sensor. http://structure.io. [Online; Accessed: July 2017]. 1, 2

[39] R. Y. Takimoto, M. de Sales Guerra Tsuzuki, R. Vogelaar, T. de Castro Martins, A. K. Sato, Y. Iwao, T. Gotoh, and S. Kagei. 3d reconstruction and multiple point cloud registration using a low precision rgb-d sensor. *Mechatronics*, 35:11 – 22, 2016. 2

[40] G. K. L. Tam, Z. Q. Cheng, Y. K. Lai, F. C. Langbein, Y. Liu, D. Marshall, R. R. Martin, X. F. Sun, and P. L. Rosin. Registration of 3D point clouds and meshes: A survey from rigid to nonrigid. *IEEE Transactions on Visualization and Computer Graphics*, 19(7):1199–1217, July 2013. 1

[41] P. Tanskanen, K. Kolev, L. Meier, F. Camposeco, O. Saurer, and M. Pollefeys. Live metric 3D reconstruction on mobile phones. In *2013 IEEE International Conference on Computer Vision*, pages 65–72, Dec 2013. 1, 3

[42] Trimensional. http://www.trimensional.com. [Online; Accessed: July 2017]. 3

[43] Trnio. http://www.trnio.com. [Online; Accessed: July 2017]. 2

[44] G. Turk and M. Levoy. Zippered polygon meshes from range images. In *Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '94, pages 311–318, New York, NY, USA, 1994. ACM. 2

[45] I. Wald and V. Havran. On building fast kd-trees for ray tracing, and on doing that in o(n log n). In *2006 IEEE Symposium on Interactive Ray Tracing*, pages 61–69, Sept 2006. 2

[46] C. Wang, M. Bao, and T. Shen. 3D model reconstruction algorithm and implementation based on the mobile device. *Journal of Theoretical & Applied Information Technology*, 46(1):255–262, Dec. 2012. 3

[47] C. Wang, X. Jing, and C. Zhao. Local upsampling fourier transform for accurate 2d/3d image registration. *Computers & Electrical Engineering*, 38(5):1346 – 1357, 2012. Special issue on Recent Advances in Security and Privacy in Distributed Communications and Image processing. 2

[48] J. H. Won, M. H. Lee, and I. K. Park. Active 3D shape acquisition using smartphones. In *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pages 29–34, June 2012. 3