MATEC Web of Conferences **32**, 06004 (2015)
DOI: 10.1051/matecconf/20153206004

A multi-view stereo based 3D hand-held scanning system using visual-inertial navigation and structured light

Shirazi Muhammad Ayaz[1], Khan Danish[1], Joon-Young Bang [1], Soo-In Park[1], YoungJun Roh [3] and Min Young Kim[1,2,a]

[1]*School of Electronics Engineering, IT College, Kyungpook National University, 1370 Sankyuk-dong, Buk-gu, Daegu, 702-701 Korea*
[2]*Research Center for Neurosurgical Robotic System, Kyungpook National University, 1370 Sankyuk-dong, Buk-gu, Daegu,702-701 Korea*
[3]*Production Engineering Research Institute(PRI), LG Electronics, LG-ro 222, Jinwi-myeon, Pyeongtaek 451-713, Korea*

Abstract. This paper describes the implementation of a 3D handheld scanning system based on visual inertial pose estimation and structured light technique.3D scanning system is composed of stereo camera, inertial navigation system (INS) and illumination projector to collect high resolution data for close range applications. The proposed algorithm for visual pose estimation is either based on feature matching or using accurate target object. The integration of INS enables the scanning system to provide the fast and reliable pose estimation supporting visual pose estimates. Block matching algorithm was used to render two view 3D reconstruction. For multiview 3D approach, rough registration and final alignment of point clouds using iterative closest point algorithm further improves the scanning accuracy. The proposed system is potentially advantageous for the generation of 3D models in bio-medical applications.

1 Introduction

Three dimensional measurements is well known in computer vision due to its applications in several areas such as medical and scientific imaging, industrial inspection and recognition, reverse engineering and 3D map building etc. 3D sensing systems may be generally categorized into contact and non-contact techniques. The contact measurement lacks due to its characteristic of touching object, slow performance and high cost of employing mechanically calibrated passive arms [1,5].The non-contact measurement is further divided into two kinds of optical techniques for 3D reconstruction i.e., active and passive techniques [2]. Passive technique constitutes the scene imaged by digital cameras from two or more viewpoints and poses correspondence problem due to the absence of strong texture on the surface of 3D object [3].To cope with this problem, active technique based on structured light was employed to create artificial texture on the surface of 3D objects [4]. 3D sensing systems based on structured light may be categorized into two types, camera-projector system and stereo camera with non-calibrated projector [5, 6].The comparison of passive stereo, camera projector system and stereo camera with non-calibrated projector was described in [6].Structured light techniques can also be classified into sequential (multiple-shot) or single shot techniques and single shot technique is commonly employed for the moving 3D object with stringent constraint on the acquisition time[7].

Due to self-occlusion, object size and limited field of view, 3D modelling system may not render the 3D model in single measurement step and multiple views are needed to merge data in to the 3D model and the multiview 3D approaches are based on estimation of the rotation and translation parameters to register multiple views in real time [8]. Several researchers employed robotic manipulators, passive arms, turntables and electromagnetic devices to accomplish 3D handheld scanning, but these devices not only restrict the user's mobility and need accurate external hand-eye calibration but these external positioning systems are also considered to be the largest and most expensive part of 3D sensing systems [9].Despite various advantages of digital camera, the geometric and perspective geometry issues entangles the geometric information obtained from cameras thus making it hard to get real time pose estimations solely from image sensors and user may overcome these issues using inertial measurement unit (IMU) or inertial navigation system (INS) which is a better solution to digital camera in term of measuring rate and temporal precision [8].The purpose of INS is to estimate the relative pose and position of the system between different viewpoints and accomplish the multiview registration using these parameters [10,11].For visual inertial navigation, both the visual and inertial pose may be fused either in time or stochastically and one sensor may

[a] Corresponding author: mykim@ee.knu.ac.kr

support pose estimation within another sensor's estimation process [8].
In this research, we propose a handheld 3D sensing system based on stereo camera, IMU and projector for close range applications. This research is organised as follows: Section 2 gives a brief account of proposed 3D sensing system and section 3 describes the visual and inertial navigation approach. Two view and multiview 3D reconstruction are accounted in sections 4 and 5 respectively. We discuss the experimental results in section 6 while section 7 includes the conclusion of this research and also reveal the directions of future work.

2 Proposed 3D Sensing System

Handheld 3D scanning system is composed of stereo camera and projector connected to the inertial navigation system (INS).These devices are connected to PC via several communication interfaces. The stereo camera with wide baseline was employed for 3D scanning system.
We extended the system described in [6] for 3D handheld scanning approach based on multiview stereo. The system proposed in [11] used a high dynamic range (HDR) image sensor besides stereo camera for visual pose estimation while we do not employ an extra camera for parameter estimation in our setup.
The stereo camera utilizes a pair of Basler Ace2500 14/gm monochrome cameras connected with A100P Zeus pocket projector and mySen-C INS device. The proposed system is depicted in figure 1 while the specifications of camera, projector and INS are shown in table 1.

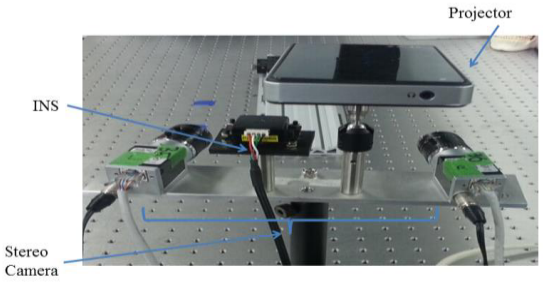


Figure 1. Proposed 3D Handheld Scanning system

Table 1. Specifications of camera, projector and INS

Item	Specifications
Basler Ace2500 14/gm	Sensor size : 2592x1944. Optical size : 1/2.5" Pixel size : 2.2 -2.2 µm Max. Frame rate : 14.6 fps
A100P Zeus	Resolution:854x480 Projection distance:20- 300cm Brightness:100 ANSI lumens
mySen-C INS	Static accuracy: ≤1 deg Dynamic accuracy:4 deg RMS Update rate: 100 Hz

3 Visual and Inertial Navigation

One of the characteristics of our system is the estimation of visual and inertial poses from one camera and INS device respectively. In our approach, the pose of one sensor may support the pose estimation process in other sensor [8].

INS was employed in our system to improve the overall pose estimation process. For inertial navigation, we need the coordinate matching between the camera data and INS data. For this purpose, coordinate axis correction is required. This process uses acceleration sensor, angular velocity sensor and geomagnetic sensor. The equation governing this algorithm is as follows:

$$R_{cam1}^{cam2} R_{cam}^{ins} = R_{cam}^{ins} R_{ins1}^{ins2}$$

R_{cam1}^{cam2} =camera rotation matrix between position 1 and position 2

R_{cam}^{ins} =Rotation matrix between camera and INS

R_{ins1}^{ins2} =INS rotation matrix between position 1 and position 2.

We have used two approaches for visual pose estimation. One approach is based on features matching using SURF (speeded up robust features) [12] while other approach utilizes chessboard target. When there are sufficient 2D features or texture on 3D object, we use features matching based approach. When texture or enough features are absent, we employ chessboard target for accurate pose estimation. Both the approaches are based on homography estimation [13] between the points of different viewpoint images of single camera and relative orientation and translation parameters were estimated using Homography decomposition provided that the intrinsic parameters of one camera are known by pre-calibration. We carried out pre-calibration of single camera using Zhang's method [14].The block diagram of visual pose estimation is shown in figure 2.

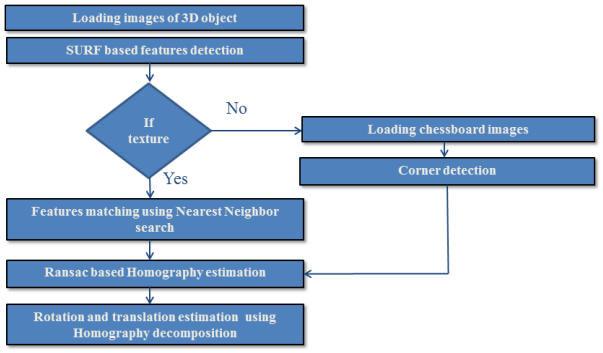


Figure 2. Visual Pose estimation for handheld 3D scanning system

4 Two View 3D Reconstruction

The 3D reconstruction of a single view of stereo camera is based on single shot pattern projection on 3D object. Our multiview stereo based approach including two view 3D reconstructions scans the static 3D object and due to stringent constraint on acquisition time [7], we have used the M-array based single shot pattern.

4.1. M-Array pattern design

Random arrays of dimensions a x b in which a sub-array of p x q is unique, is called M array or perfect map. User may construct the M-array pattern theoretically following equation $ab=2^{pq}$, but practically we do not consider the zero sub matrix and get a total of $ab=2^{pq}$ -1 unique sub-arrays of p x q size [5].In this research, we contributed by employing the binary coded M-array pattern proposed in [15] for our handheld 3D scanning system while this pattern was previously used for single camera projector system. This pattern has just one symbol of white square, no connected ones and unique sub-matrix in the pattern. The less no of symbols and connectivity constraint simplifies the pattern segmentation in decoding stage while the unique window property tends to simplify the correspondence problem [15].In order to increase the resolution; we used pixel replication method to replace zeros or ones in pattern with n x n grid of same symbol. The pattern of 200x200 resolution having 9x9 unique window with pixel replication using n=2 are shown in figure 3.

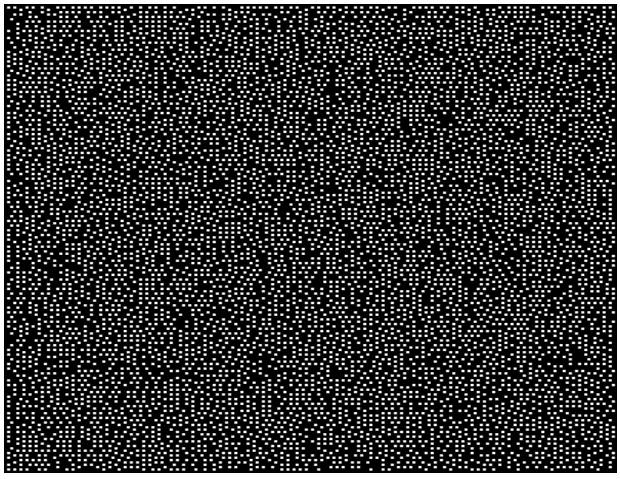


Figure 3. (200x200) pattern generated with (9x9) unique window and n=2.

4.2 Block Matching Algorithm

Since our proposed system is based on stereo camera and non-calibrated projector [5], conventional decoding methods for M-Array are not required in this research and matching was performed on stereo images with projected pattern, not between the ideal pattern and single camera image as in camera projector system.

The two view 3D reconstruction is based on fast and effective block matching algorithm [13] applied on pattern projected images of stereo camera. This algorithm matches the rectified stereo images using sum of absolute difference (SAD) window and estimates depth for every pixel in highly textured scenes.

After finding the matched image features in both images, 2D points in homogenous coordinate may be projected to the three dimensional coordinate using the following equation [13]:

$$Q\begin{bmatrix} x \\ y \\ d \\ 1 \end{bmatrix} = \begin{bmatrix} X \\ Y \\ Z \\ W \end{bmatrix}$$

$$Q = \begin{bmatrix} 1 & 0 & 0 & -c_x \\ 0 & 1 & 0 & -c_y \\ 0 & 0 & 0 & f \\ 0 & 0 & -1/T_x & (c_x - c_x')/T_x \end{bmatrix}$$

Q= reprojection matrix
(c_x , c_y)=coordinates of Principle point of left camera
T_x=translation along x-axis of left camera
f =focal length of left camera
c_x'=x coordinate of principle point for right camera
(x,y)=image coordinates
(X/W,Y/W,Z/W)=3D coordinates
d= disparity
W=scale factor

5 Multiview 3D Reconstruction

Multiview 3D reconstruction is based on two stages such as rough registration of different view point clouds and their refinement based on iterative closest point (ICP) algorithm. The rough registration uses the visual and inertial pose estimated in section 3.The ICP algorithm then aligns the roughly registered point clouds with the reference view point cloud.

5.1 Rough Registration

The rough registration is based on the transformation of point clouds of different views into coordinate of reference view point cloud via visual and inertial pose estimation described in section 3. The equation governing the rough registration is as follows:

$$X_i^{ref} = R_i X_i + T_i$$

X_i=3D point cloud of ith view
R_i =relative rotation between ith view and reference view point cloud
T_i= relative translation between ith view and reference view point cloud
X_i^{ref}=ith point cloud transformed into reference view

5.2 ICP based point cloud alignment

In order to refine the roughly registered point clouds, ICP algorithm [16] was applied in this research. This algorithm takes the point clouds which are in approximate registration with the reference view. The ICP algorithm consists of the following main steps.

1. Nearest neighbour search based Delaunay triangulation and convex hull estimation for 3D points matching between ith view and reference view point cloud
2. Finding the relative orientation and rotation parameters from matched 3D points
3. Transformation of ith point cloud into the coordinate of reference view point cloud using estimated parameters
4. Repeating steps 1, 2 and 3 till ICP converges to acceptable solution

Suppose we consider two point clouds 'A' and 'B' whose matched points are A_1 to A_n and B_1 to B_n respectively. Let C_a and C_b are the centroids in A and B respectively. We consider 3x3 covariance matrix 'M' decomposed into 'U' and 'V' matrices using singular value decomposition (SVD) while the rotation and translation parameters are denoted by 'R' and 'T'. The following equations govern the estimation of R and T parameters.

$$M = \sum_{i=1}^{n} (A_i - C_a) * (B_i - C_b)^T$$

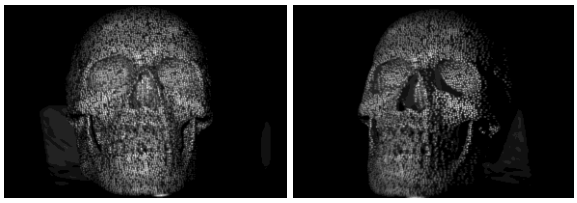
$$R = V * U^T$$

$$T = C_b - R * C_a$$

6 Experiments and Results

For demonstration of our handheld scanning system, we have employed artificial skull object for biomedical applications. The object was placed at 30 cm from the hardware setup and M-array pattern was projected on it. The stereo camera images with the projection of M-Array pattern are shown in figure 4.Two view 3D reconstruction result using block matching algorithm is shown in figure 5 when the reference view and three other view point clouds are visualized at the same time in Geomagic Verify Viewer software.

For multiview 3D reconstruction approach, we have used reference view point cloud and other three point clouds of skull. The results of refinement using ICP is shown in figure 6 when the reference view and three view point clouds are visualized. After refinement using ICP, the shape of skull is visible in figure 6 as compared to the result in figure 5.The result in figure 6 shows the good alignment of other view with the reference view and shape of skull was also improved.



(a) **(b)**

Figure 4. The images of left and right camera with the projection of M-Array pattern on skull phantom

For 3D handheld scanning, acquisition time is very critical. Table 2 shows the processing time of different algorithms used in this research.

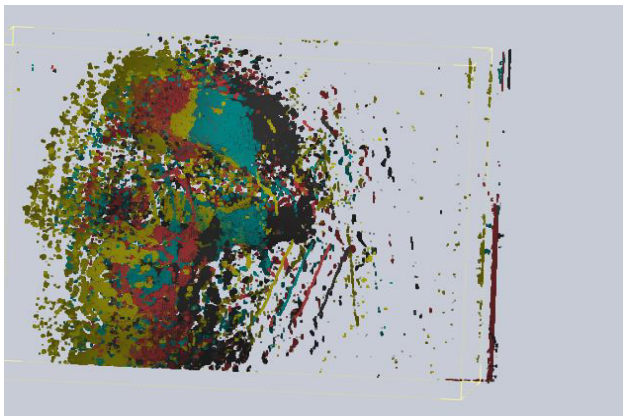


Figure 5. The visualization of reference view and three different view point clouds before ICP based refinement, point clouds shown in different colors

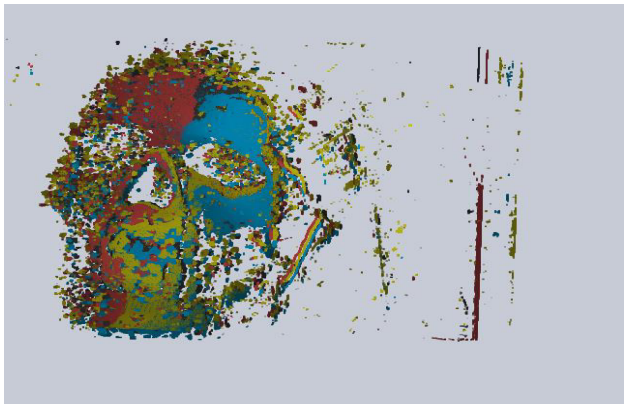


Figure 6. The visualization of reference view and three different view point clouds after registration using ICP, point clouds shown in different colors

Table 2. Processing time for different steps

S.No.	Steps	Processing time(s)
1.	Features based Visual pose estimation	2.7
2.	Target based Visual pose estimation	5.5
3.	Two view 3D reconstruction	10
4.	Rough Registration	8
5.	ICP algorithm	9
	Overall time	35.2

7 Conclusion

In this paper, we have implemented a 3D handheld scanning system based on visual inertial navigation and structured light. The system consists of stereo camera, IMU and projector to get the point clouds of different viewpoints. The system employed features and chessboard target based visual pose estimation algorithm. M-array pattern was employed for solving the correspondence problem of stereo camera. Block matching algorithm was employed for two view 3D reconstruction. For multiview 3D approach, rough registration and final alignment of point clouds using ICP further improves the accuracy of 3D scanning. The proposed system is feasible for 3D scanning and finds application in biomedical imaging.

The proposed system yields 700-900 k data points per single scan with the mean error of 0.056 mm in the dimension of 3D object. The visual inertial navigation approach results mean error of 0.85 mm and 0.29 deg in the pose and position respectively. Our system has better accuracy in estmating the visual inertial pose and position as compared to 3D modeller [8].The system proposed in [8] operates at longer range of 30cm to 2m and generates 3D model in less scanning time as compared to our system.

We can reduce the scanning time using feature-based stereo vision without processing whole images.One possibility to reduce scanning time is to assign separated computing threads to different algorithms to meet the application requirement via parallel processing.We will also intend to improve feature matching algorithm so as to avoid the use of IMU in this research.

References

1. Li, Yadong, and Peihua Gu. "Free-form surface inspection techniques state of the art review." *Computer-Aided Design* 36.13 (2004): 1395-1417.
2. Bruno, F., et al. "Experimentation of structured light and stereo vision for underwater 3D reconstruction." *ISPRS Journal of Photogrammetry and Remote Sensing* 66.4 (2011): 508-518.
3. Shi, Chun-qin, and Li-yan Zhang. "A 3D Shape Measurement System Based on Random Pattern Projection." *Frontier of Computer Science and Technology (FCST), 2010 Fifth International Conference on*. IEEE, 2010.
4. Schaffer, Martin, Marcus Grosse, and Richard Kowarschik. "High-speed pattern projection for three-dimensional shape measurement using laser speckles."*Applied optics* 49.18 (2010): 3622-3629.
5. Salvi, Joaquim, et al. "A state of the art in structured light patterns for surface profilometry." *Pattern recognition* 43.8 (2010): 2666-2680.
6. Kim, Min Young, Shirazi Muhammad Ayaz, Jaechan Park, and YoungJun Roh. "Adaptive 3D sensing system based on variable magnification using stereo vision and structured light." *Optics and Lasers in Engineering* 55 (2014): 113-127.
7. Geng, Jason. "Structured-light 3D surface imaging: a tutorial." *Advances in Optics and Photonics* 3.2 (2011): 128-160.
8. Strobl, Klaus H., et al. "The self-referenced DLR 3D-modeler." *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on*. IEEE, 2009.
9. Munkelt, Christoph, et al. "Handheld 3D Scanning with Automatic Multi-View Registration Based on Optical and Inertial Pose Estimation." *Fringe 2013*. Springer Berlin Heidelberg, 2014. 809-814.
10. Byczkowski, Tomasz, and Jochen Lang. "A stereo-based system with inertial navigation for outdoor 3D scanning." *Computer and Robot Vision, 2009. CRV'09. Canadian Conference on*. IEEE, 2009.
11. Kleiner, Bernhard, et al. "Handheld 3-D Scanning with Automatic Multi-View Registration Based on Visual-Inertial Navigation." *International Journal of Optomechatronics* 8.4 (2014): 313-325.
12. Bay, Herbert, et al. "Speeded-up robust features (SURF)." *Computer vision and image understanding* 110.3 (2008): 346-359.
13. Bradski, Gary, and Adrian Kaehler. *Learning OpenCV: Computer vision with the OpenCV library*. " O'Reilly Media, Inc.", 2008.
14. Zhang, Zhengyou. "A flexible new technique for camera calibration." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 22.11 (2000): 1330-1334.
15. Wijenayake, Udaya, and Soon-Yong Park. "Dual pseudorandom array technique for error correction and hole filling of color structured-light three-dimensional scanning." *Optical Engineering* 54.4 (2015): 043109-043109.
16. Mian, Ajmal S., Mohammed Bennamoun, and Robyn Owens. "Three-dimensional model-based object recognition and segmentation in cluttered scenes." *Pattern Analysis and Machine Intelligence, IEEE Transactions on*28.10 (2006): 1584-1601.