3D imaging and image processing – literature review

[1]*Michal VARGA (1st year),*
Supervisor: [2]*Ján JADLOVSKÝ*

[1,2]Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

[1]michal.varga@tuke.sk, [2]jan.jadlovsky@tuke.sk

***Abstract*—This review aims to provide an overview of the latest trends in 3D computer vision. Both in imaging and image processing, the research has advanced significantly over the past decade. This paper reviews the fundamentals and advanced techniques of 3D image acquisition and algorithms for storing, processing and understanding this information. This overview is focused on the application domain of autonomous mobile robotics and automated industry. Several most widely used passive and active optical range imaging techniques are reviewed with their strengths and weaknesses. This paper also describes some of the most common range image processing techniques and structures.**

***Keywords*—3D imaging, computer vision, laser scanning, mobile robotics, point cloud, quality control, stereo vision**

I. Introduction

Used in industry, autonomous robotics and other fields, 3D imaging provides very detailed and valuable information about the environment or examined objects. Especially in mobile robotics, 3D vision has become the area of interest of many researchers and numerous imaging and processing algorithms have been developed over the past decade. In modern industry, both 2D and 3D vision systems are the basis of automatic construction, product inspection and quality control.

The fundamental difference between 2D and 3D vision is the inclusion of the third coordinate – depth. This information can be acquired by various means ranging from stereoscopy (use of two specifically aligned cameras) to laser scanning of the environment, while each of these techniques have their own advantages and disadvantages. Downsides can often be eliminated by using multiple 3D imaging methods in conjunction.

In robotics, 3D representation of the environment allows for better navigation, obstacle avoidance and object classification. 3D scans of the surrounding objects enables the robot to interact with them more easily and in more sophisticated manner such as grabbing various objects and opening doors [1]. In industry and manufacturing, 3D vision systems provide unprecedented precision and flexibility in control, measurement and quality inspection. While these areas require different approaches to hardware construction and methods of 3D image acquisition, most of the image processing algorithms are used universally.

II. Methods of 3D image acquisition

The collection of techniques that are used to measure the distance to a set of points in a scene from a specific point is called range imaging. The result is a 2D *range image*, where the pixel values correspond to the distance. As with many other types of measurements, both active and passive sensors are available for this task.

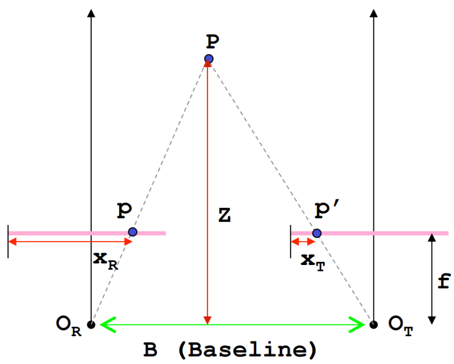

Fig. 1. Depth triangulation – based on the known intrinsic and extrinsic parameters (baseline B, focal length f, ...) and disparity ($x_R - x_T$), the depth Z is calculated [3].

A. Stereo triangulation

Perhaps the most well known range imaging technique is stereo vision. It is a passive method, since it only requires a pair of cameras and no energy is emitted. The depth is calculated by triangulation based on the *disparity* of each pixel – the difference between the x coordinate of two corresponding points, representing the same homologous point in the scene (Fig. 1). Based on intrinsic and extrinsic parameters of the setup (acquired by calibration), the physical coordinates can be calculated from the pixel position and disparity. While having many advantages (passive sensing, low cost and latency [2], small size, color information), stereo vision has its pitfalls. Solving the correspondence problem is difficult in homogeneous regions, repeating patterns or when there are specular reflection or transparent materials in the scene [3].

B. Laser triangulation

Utilization of lasers is very common in 3D scanning. One of the techniques using this device is laser triangulation. In this case, the laser shines on the scanned object, which is simultaneously recorded with a camera. The distance is triangulated using the known separation between the camera and the laser emitter as well as the angles to the illuminated point (Fig. 2).

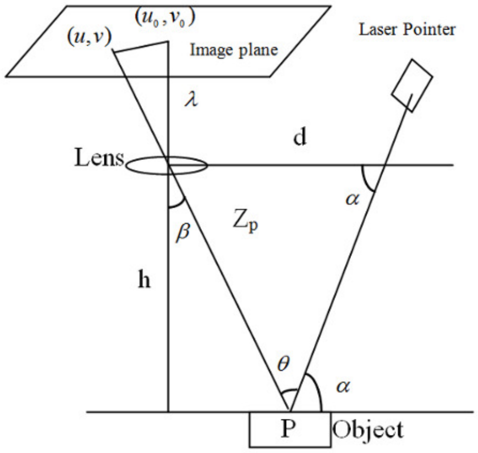

Fig. 2. Laser triangulation – distance to the object P is equal to the height of a triangle with two known angles (α, β) and one known side (d) [5].

Often a line laser is used instead of a single laser dot for faster acquisition process. Since energy is emitted towards the scene, all laser scanners are categorized as active sensors [4]. This technique offers extremely accurate depth measurements, however it struggles to correctly scan transparent and reflective surfaces, since those do not diffuse the laser well enough to be captured by the camera. Result of this technique yields no color or light intensity information.

C. LIDAR, Time-of-Flight

LIDAR (acronym for Light Detection And Ranging) is another laser-utilizing technique for 3D scanning of the environment (Fig. 3). Often used in geodesy, geography, and related fields, LIDAR has found its use in autonomous vehicles and robotics as well. LIDAR may use various techniques to infer the distance to the illuminated point. One of those is time-of-flight, where a precise timing mechanism measures the delay between emitting and receiving the laser pulse. Using this delay and the known speed of light in the environment, distance to the surface is calculated [6]. The whole scene is scanned point-by-point by slewing the laser beam around using a rotating mirror. Both 2D (planar) and 3D LIDARs are available. The main advantages of LIDAR are its hight precision and long range, however these devices are usually rather expensive due to their very precise hardware.

Time-of-Flight technique (Fig. 4) can also be used with fast gated CMOS cameras. In this case the whole scene is illuminated at once and the delay is measured per-pixel between the emission and reception of a single laser or LED pulse. The main advantage of this technique is its high acquisition speed reaching up to 160 frames per second.

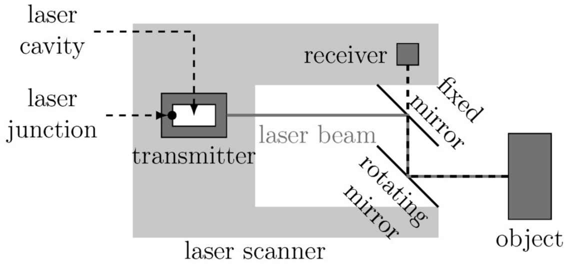

Fig. 3. Time-of-Flight LIDAR. Pulsed laser beam is reflected by a rotating mirror creating a fan-like scanning pattern. The pulse is then reflected from the scanned surface and returns to the receiver. The time of flight between transmission and reception is then used to estimate the distance to the surface [7].

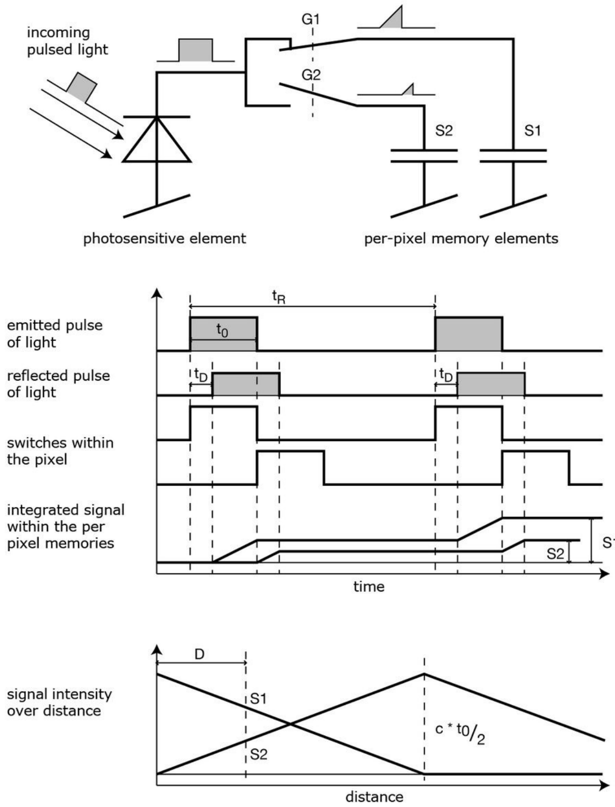

Fig. 4. Diagrams illustrating the principle of a time-of-flight camera with analog timing [8].

D. Structured Light 3D scanners

Structured-light 3D scanner is an active 3D scanning device which measures the depth by recording projected light patterns with a camera. The most widely used light pattern in this technique are parallel stripes generated by either projection or laser interference. Both the displacement and width of individual stripes, as well as their frequency and phase provide cues to calculate the depth. Several patterns (e.g. stripes of varying width) are usually used per scan (Fig. 5). Advantages of this technique are its low hardware cost and high scanning speed, however it requires a controlled environment and projector calibration.

A very popular example of a structured-light 3D scanner is an entertainment device called Kinect. Named *Light Coding*, this scanner is uses an infrared laser to project a pattern of dots which is then captured by a monochrome CMOS camera.

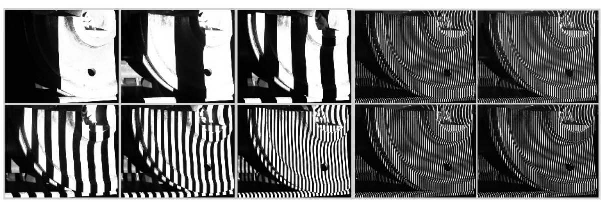

Fig. 5. Set of structured light patterns consisting of parallel stripes of varying width projected on a scanned object [9].

Due to its affordability and relatively good performance, it has found its use in many robotics projects [10].

E. Structure from motion

Structure from motion (SFM) is a range imaging technique which estimates 3D structures from sequences of 2D images captured at different points in the scene. These methods include both camera localization and 3D shape reconstruction. Traditional feature detectors, descriptors and matchers may be used to determine motion trajectories. Depth calculation algorithms similar to those used in stereo triangulation are then used to retrieve the 3D structure of the environment or the scanned object [11]. Some variants of SFM, called direct approaches, estimate geometric information without intermediate abstraction to features or corners [12]. Pitfalls similar to those in stereo triangulation are present here as well – reflective materials, transparency, homogeneous surfaces, etc. Occlusion are partially solved by having multiple points of view available for shape reconstruction.

A very related problem to SFM is SLAM (Simultaneous Localization And Mapping) which is often found in autonomous robotics. This is a problem of constructing and updating a map of environment by a mobile agent (robot) equipped with some kind of a sensing device. A pair of cameras or a LIDAR is often used, while solutions for single-camera (monocular) systems have been found too [12][13].

F. Combination of techniques

As mentioned, each of the techniques has its own advantages and disadvantages. To mitigate their downsides, solutions have been found to use multiple range imaging techniques simultaneously. One such combination is a laser scanner with a stereo camera, providing both precise depth measurements as well as color image [14]. When combining several sensors, thorough calibration has to be performed to align the acquired data. Algorithms exist that solve this problem even when the sensors' fields of view are not overlapping [15].

III. Range processing

Range images are often processed before applying any classification or scene-understanding algorithms. There are various forms of representation of 3D environments or objects, each of them having their use.

A. Point clouds

The most precise 3D scene representation and often the direct output of a depth scanner is a raw point cloud. This consists of a set of points in 3D space which may also bear color information and usually represent the external surface of an object (Fig. 6). Point clouds are often post-processed using some type of filtering algorithm to remove outliers introduced by sensor noise. One of the most commonly used filtering algorithm is RANSAC [17], or one of its later variants [18]. Filtering is also used to remove local noise – smoothing. Several scans – point clouds – of the scene can be aligned to create a complete representation of the environment.

Point clouds can be generated from range image using a simple conversion algorithm by providing the parameters from calibration. Many of the most frequently used algorithms for generating, manipulating and visualizing point clouds are implemented within the popular open-source library called PCL [16]. Raw point clouds can be difficult to process due to large amount of data they contain. To accelerate the manipulation, the sampled points can be organized in octrees [19].

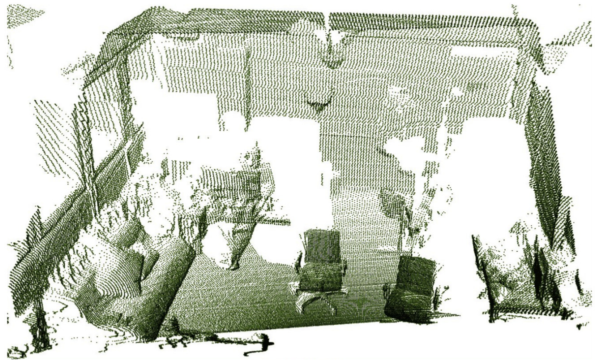

Fig. 6. Example of a point cloud representation of a room visualized using the PCL library [16].

B. Triangulation

Point clouds are not ideal for further modification and real-time visualization. For those purposes, it is often required to use a different geometric representation such as a polygonal model (mesh). Delaunay triangulation method is often used with 2D data (Fig. 7), while 3D input requires a more difficult approach called tetrahedralization or tetrahedrization [20]. This method can be very computationally expensive and modern techniques are able to accelerate this process by exploiting graphics hardware [21]. Surface normal estimation is often a part of the surface reconstruction process.

Geometry represented in this way can then be further optimized and compressed by traditional polygon mesh optimization algorithms to save space and accelerate further processing.

C. Object detection and pose estimation

Both in autonomous robotics and industry, the problem of detecting and classifying objects appears very frequently. In case the geometry of the object is known in advance (e.g. a CAD model is available), the task consists of finding this object in the scene and fitting it to the segmented data (estimating its position and rotation – 6DOF pose) [22]. Once this is achieved, the results can be used for various tasks, for example to navigate a mobile robot or detect and measure surface defects in industrial quality control [23].

Another type of problem is detection and general classification of objects in the environment such as people, cars,

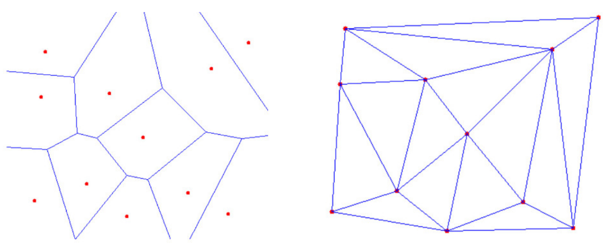

Fig. 7. Delaunay triangulation (right) corresponds to the dual graph of Voronoi diagram (left) [20].

furniture, switches, door handles, etc. This problem occurs mainly in autonomous robotics, where the robot has no prior information about the environment. Some of the tools used for solving this problem are Markov networks [24] and conditional random fields [25].

IV. Conclusion

The field of 3D imaging and processing is very wide and there are still numerous unsolved problems in both imaging (some techniques still struggle with transparent and reflective materials, occlusions, etc.) and range processing (scene understanding, semantic mapping, ...). As autonomous mobile robots are slowly becoming a part of the human society, we believe it is important to focus on the perception capabilities of these machines. Understanding its surroundings is crucial for safe and useful operation of these machines.

In our next work, we would like to focus on developing techniques and algorithms for scene understanding, object detection and classification with its main use in autonomous mobile robotics. We want to work on solving problems and improving existing solutions in the domain of segmentation and semantic labeling of point cloud data for the purpose of robot navigation and interaction with focus on both static and dynamic environments. Combining these methods with existing 3D imaging and SLAM algorithms will aid the development of fully autonomous robots with the ability to perform useful tasks in previously unknown environments. Since robotics has very strong presence in automated manufacturing, we believe our work will find its application in this field as well.

Acknowledgment

This work has been supported by the Research and Development Operational Program for project: University Science Park Technicom for innovative applications with knowledge technology support, ITMS code 26220220182, co-financed by the ERDF (80%) and by grant KEGA - 001TUKE-4/2015 (20%).

References

[1] M. Quigley, S. Batra, S. Gould, E. Klingbeil, Q. V. Le, A. Wellman, A. Y. Ng *et al.*, "High-accuracy 3d sensing for mobile manipulation: Improving object detection and door opening." in *ICRA*, 2009, pp. 2816–2822.

[2] B. Tippetts, D. J. Lee, K. Lillywhite, and J. Archibald, "Review of stereo vision algorithms and their suitability for resource-limited systems," *Journal of Real-Time Image Processing*, vol. 11, no. 1, pp. 5–25, 2016.

[3] S. Mattoccia. (2013, January) Stereo vision: Algorithms and applications. Department of Computer Science (DISI), University of Bologna. [Online]. Available: http://vision.deis.unibo.it/~smatt/Seminars/StereoVision.pdf

[4] E. D. Acosta, E. O. García, and E. J. Aponte, "Laser triangulation for shape acquisition in a 3d scanner plus scan," in *Electronics, Robotics and Automotive Mechanics Conference, 2006*, vol. 2. IEEE, 2006, pp. 14–19.

[5] W.-F. Xie, Z. Li, X.-W. Tu, and C. Perron, "Switching control of image-based visual servoing with laser pointer in robotic manufacturing systems," *Industrial Electronics, IEEE Transactions on*, vol. 56, no. 2, pp. 520–529, 2009.

[6] R. Horaud. Three-dimensional sensors, lecture 4: Time of flight cameras (pulse light modulation). INRIA Grenoble Rhone-Alpes, France. [Online]. Available: http://perception.inrialpes.fr/~Horaud/Courses/pdf/Horaud_3DS_4.pdf

[7] A. Alhashimi, D. Varagnolo, and T. Gustafsson, "Joint temperature-lasing mode compensation for time-of-flight lidar sensors," *Sensors*, vol. 15, no. 12, pp. 31 205–31 223, 2015.

[8] W. Commons. (2010) Diagrams illustrating the principle of a time-of-flight camera with analog timing. [Online]. Available: https://commons.wikimedia.org/wiki/File:TOF-camera-principle.jpg

[9] Z. Liu, X. Li, F. Li, X. Wei, and G. Zhang, "Fast and flexible movable vision measurement for the surface of a large-sized object," *Sensors*, vol. 15, no. 3, pp. 4643–4657, 2015.

[10] M. R. Andersen, T. Jensen, P. Lisouski, A. K. Mortensen, M. K. Hansen, T. Gregersen, and P. Ahrendt, "Kinect depth sensor evaluation for computer vision applications," *Technical Report Electronics and Computer Engineering*, vol. 1, no. 6, 2015.

[11] S. N. Sinha, D. Steedly, and R. Szeliski, "A multi-stage linear approach to structure from motion," in *Trends and Topics in Computer Vision*. Springer, 2010, pp. 267–281.

[12] J. Engel, T. Schöps, and D. Cremers, "Lsd-slam: Large-scale direct monocular slam," in *Computer Vision–ECCV 2014*. Springer, 2014, pp. 834–849.

[13] A. J. Davison, I. D. Reid, N. D. Molton, and O. Stasse, "Monoslam: Real-time single camera slam," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 29, no. 6, pp. 1052–1067, 2007.

[14] S. Budzan and J. Kasprzyk, "Fusion of 3d laser scanner and depth images for obstacle recognition in mobile applications," *Optics and Lasers in Engineering*, vol. 77, pp. 230–240, 2016.

[15] Y. Bok, D.-G. Choi, and I. S. Kweon, "Extrinsic calibration of a camera and a 2d laser without overlap," *Robotics and Autonomous Systems*, 2016.

[16] R. B. Rusu and S. Cousins, "3d is here: Point cloud library (pcl)," in *Robotics and Automation (ICRA), 2011 IEEE International Conference on*. IEEE, 2011, pp. 1–4.

[17] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[18] R. Schnabel, R. Wahl, and R. Klein, "Efficient ransac for point-cloud shape detection," in *Computer graphics forum*, vol. 26, no. 2. Wiley Online Library, 2007, pp. 214–226.

[19] R. Schnabel and R. Klein, "Octree-based point-cloud compression." in *SPBG*, 2006, pp. 111–120.

[20] R. Fabio *et al.*, "From point cloud to surface: the modeling and visualization problem," *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 34, no. 5, p. W10, 2003.

[21] L. P. Kobbelt and M. Botsch, "An interactive approach to point cloud triangulation," in *Computer Graphics Forum*, vol. 19, no. 3. Wiley Online Library, 2000, pp. 479–487.

[22] A. Aldoma, M. Vincze, N. Blodow, D. Gossow, S. Gedikli, R. B. Rusu, and G. Bradski, "Cad-model recognition and 6dof pose estimation using 3d cues," in *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*. IEEE, 2011, pp. 585–592.

[23] J. Xu, N. Xi, C. Zhang, Q. Shi, and J. Gregory, "Real-time 3d shape inspection system of automotive parts based on structured light pattern," *Optics & Laser Technology*, vol. 43, no. 1, pp. 1–8, 2011.

[24] D. Munoz, N. Vandapel, and M. Hebert, "Directional associative markov network for 3-d point cloud classification," 2008.

[25] R. B. Rusu, Z. C. Marton, N. Blodow, A. Holzbach, and M. Beetz, "Model-based and learned semantic object labeling in 3d point cloud maps of kitchen environments," in *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on*. IEEE, 2009, pp. 3601–3608.