3D Image Sensors, an Overview

MARIUS OTESTEANU, VASILE GUI,
Department of Communications,
Faculty of Electronics and Telecommunications,
"Politehnica" University of Timisoara,
Bd. Vasile Parvan No. 2,
ROMANIA

marius.otesteanu@etc.upt.ro, vasile.gui@etc.upt.ro

Abstract: - We briefly introduce the current state in the most widely used 3D image sensors in computer vision today. Passive and active sensing methods are discussed, with highlight on difficult computer vision problems that can be avoided by using active methods. Active optoelectronic imaging methods like time of flight, interferometry and triangulation are presented. The 3D image processing applications included in this issue are reviewed and some conclusions drown.

1 Introduction.

Undoubtedly, vision is our the most important source of information, facilitating intelligent interaction with a 3D world. Efforts to provide machines with the same kind of vision were constantly made throughout several decades of work in image processing [1]-[5]. Now machine vision, computer vision, image analysis and multimedia are very active fields of modern engineering. Currently, 3D image processing applications include biomedical image analysis, robot guidance, autonomous navigation, car safety, human computer interface, surveillance, remote sensing, oceanography, industrial inspection, microscopy, astronomy, virtual reality, science, etc. This amazing diversity of applications is based on a wide range of image sensors. Image information usually commes in the form of a radiation, captured and quantified by some kind of sensor. Most often than not, the radiation is in the visible part of the electromagnetic radiation spectrum. However, radiology uses X rays, thermal imaging is based on infrared radiation, echography and sonography use accoustic vibrations.

Depending on the particular type of sensor used, 3D image data presents different levels of resolution, accuracy and noise. As a result, useful information extraction from sensor data needs to consider both sensor charateristics and the type if information desired.

The term "3D" deserves a short comment. In the signal processing teminology, a 3D signal may be any physical quantity that depends on 3 free variables. In a strict sense of the definition, video sequences can be considered as 3D signals, depending on two spatial variables and one temporal variable. A 3D video sequence, like the one obtained in one of the papers included in this issue, is a sequence of 3D images, but could be actually treated as a 4D signal, with three space variables and one time variable. Conversely, range images, containing only depth information at each image pixel are sometimes refered to as 2.5D images. True 3D images are volume images generated for example by computer tomagraphy (CT) scans. 3D images provided by modern CCD cameras provide depth image in addition to reflected light intensity, so intensity can be viewed as a signal depending on three spatial variables.

Optoelectronical 3D imaging has a long history. Methods to obtain range image information can be classified as active or passive ones. Passive methods rely only on radiation received from the environment, while active methods project a structure of coherent and/or modulated light on the target in order to obtain shape information.

Most of the previous work in computer vision focused on passive methods. Classic examples of 3D computer vision tasks include photo sculptures from silhouete images, shape from

shading, photogrammetry, focus/defocus or confocal microscopy and stereo vision. Much work has been done to solve the last one [6]. Stereo vision exploits the principle of triangulation to derive depth information from left camera generated image and right camera generated image disparity. Depth computation is based on point correspondences, which unfortunately are unknown and therefore have to be estimated from local image similarity. The problem is far from trivial, especially on flat image areas. Moreover, subpixel accuracy is needed in order to derive useful results, since small errors in feature localization result in high errors in depth estimation.

Passive 3D imaging methods work well on structured surfaces, with well defined edges, lines, corners and many uniquely identifiable "landmark" points. They do not require special purpose hardware development and have been used successfully in several applications.

Active 3D optoelectronical imaging methods project a structure of regularly spaced lines of light on the sensed object and yield a grid of known depths. Among the most widely used active optical methods are time of flight, interferometry and triangulation.

A time of flight camera [7],[8],[9] detects the delay of the wave reflected by the target with respect to the emited vawe. The delay is proportional to the distance to target and inversely proportional to the speed of wave. This principle was first used with ultrasonic waves and is still widely used in echography. Ultrasonic waves travel through any environment, be it solid, liquid or gas. The wave is reflected at any discontinuity of a surface, making possible the detection and measurement of its position. More recently, laser based time of flight cameras have been built using the same principle. Compared to triangulation, time of flight based sensors do not suffer from shading (or occlusion) problems, caused by lateral illumination in the first case.

Despite the simplicity of the idea, the electronics used to measure time delays needs to be able to resolve accurately time delays of the order of magnitude of 10^{-15} seconds (that is femto seconds), a formidable performance at date. Low reflected energy levels in conjunction with photodetector noise make the task nearly impossible. Hence, for high precision, the method is best applied in long range 3D measurements of large objects, such as monuments, buildings or in elevation map building from aerial LIDAR imaging.

However, recent developments opened the way to applications in machine vision. CMOS 3D time of flight single chip cameras provide in real time, 30 frames per second sequences of depth and intensity images. This is a significant brake through in machine vision, bypassing many difficult issues like motion estimation, pose estimation, object segmentation or tracking. Shadows, reflections, illumination changes, complex texture patterns, camouflage are no more crucial difficulties with the availability of depth information. Time of flight cameras can work with pulsed light or continuously modulated light, facilitating delay measurements. Continuous (amplitude or frequency) modulation makes posible depth measurement from the modulator wave phase shift. Since phase shifts can be measured more precisely than absolute time shifts, better resolution can be obtained by this method. The image sensor amplifier at each pixel is made phase sensitive, being synchronised with light emission. As a result, phase conveys depth information. Infrared sources are prefered in order to make the projected light structure invisible for humans.

Currently, single chip CMOS time of flight 3D cameras for computer vision have relatively low resolutions, because of the relatively short distances to the target. For high resolution measurements, interferometric methods [10] can be used. In the interferometric 3D imaging method, a regular grid of light is projected on the surface of the imaged object and the reflected light is mixed with the reference pattern. The resulting Moiré patterns are used to derive local depths changes. They work well on regular surfaces, but may encounter phase discrimination problems in the presence of depths discontinuities. Moiré methods are closely related to holographic interferometry and may be also considered as a special form of triangulation. In the classical optical triangulation method, two systems of horizontal and vertical deflection mirrors are used to scan the imaged object with a laser beam [7],[11]. The angle of the reflected light depends on the depth of the reflecting point and changes the incidence point on a linear image sensor array. A complete aquisition of the shape of an object can be obtained by rotating the object. In the active optical triangulation method, the scene is illuminated from one direction by a coherent source of light and viewed from another direction. The depth information can be computed based on the illumination angle, the viewing angle and the base line distance between the illuminator and the sensor.

2 Scanning this issue

The papers included in this issue cover several applications of 3D image sensors. No one is concerned with sensor design. We briefly discuss them below, with the aim of giving the reader the guidence needed to select the papers containing his or her topics of interest.

The paper of **Takahashi** describes a system that generates in real time a sequence of 3D colour and depth images covering the entire shape of the imaged object. To this end, multiple range images obtained simultaneously from four different directions, are fused together quickly and continuously. Range information is infered using the principle of stereo vision. Each camera contains three colour CCD units. Disparity information is extracted from those three images, by stereo matching. Fast data communication between sensors, local stations and a central processor is essential for real time operation. The "point cloud" data structure used to encode the 3D shape information is shown to be efficient from both the data communication point of view as well as for solving connectivity problems. In contrast with mesh representations, the point cloud data structure does not have the connectivity information and therefore an important data reduction is obtained. Annoying shape artefacts may result from improper stitching of adjacent images generated by different sensors, given calibration and resolution limits of the stereo vision based range sensors. An algorithm to remove quickly and efficiently invalid points, using a cylindrical projection method is described by the authors. Details of the characteristics and size of the equipment are included in the presentation. The algorithm and the whole system work in real time.

An application of a CMOS time of flight 3D camera in computer vision is described in the paper of **Jivet**. The 3D CMOS time of flight camera used has a resolution of 176×132 pixels and works with infrared light, amplitude modulated with 50 MHz frequency rectangular pulses. Reportedly, the maximum unambiguous range is around 3 m with a resolution on the order of one cm. The authors describe a 3D image segmentation method, capable of extracting large objects in a scene. Potential applications are robot vision, collision avoiding and blind people assistance. The proposed segmentation works in two steps. The first step is very fast and consists of a multiple thresholding in the global depth image histogram, based on peak detection. Significant peaks correspond to large objects exposing flat surfaces normal to the viewing direction. However, oblique angled surfaces like the floor and lateral walls generate locally uniform depth histograms in symmetric regions, making the peak detection based thresholding to fail to detect them. Therefore, a second step is used to segment such surfaces, based on local depth histograms of horizontal and vertical image strips. Effective object segmentation of a scene is demonstrated by experiments with an office image. Real time scene perception is claimed by the authors. Higher level semantic scene representation is described. The overall performance of the system is evaluated in terms of computational efficiency and semantic relevance in envisaged applications.

A second paper dealing with 3D image segmentation, signed by **Lascu**, is devoted to 3D echocardiographic imagery. Second generation 3D echocardiography systems provide high quality data, but its use is still in infancy. New applications such as image guided interventions and therapy emerged. Old image processing methods in echocardiography need to be re-assessed, given the higher quality of sensor data. The authors exploit prior information available about intracardiac ultrasound images obtained in vivo, using pathology information as a reference for the ground truth. Their method is essentially a seeded region growing segmentation, carried out in a multifeature data space. Namely, four texture features are used along with the grey level data extracted from the 3D image. Region growing is approached with the basic morphological watershed tool, enriched with elaborated metrics, like geographic priority privilege, geographic similarity and equal opportunity competence criteria, used to evaluate region similarity. Starting from a manually selected region of interest, real time segmentation is obtained, on a Labview programming platform. Experimental results demonstrated the effectiveness and reliability of the proposed segmentation. The shape and spatial distribution of segmented infarction and ischemic regions obtained in experiments matched closely the pathology images.

The paper of **Boehnke** deals with 3D object handling in the framework of the classical problem of robotic bin picking, consisting in the extraction of a desired object from a set of objects jammed in a single box. Since bean picking requires very accurate object surface registration and object segmentation, 3D scanning laser sensors are considered. The author actually developed a 3D sensor simulator, based on principles from virtual reality. The approach is supported well by existing programming libraries, like directX and OpenGL.

These libraries were designed and distributed for application developers. The available software development tools not only provide powerful functions, but also fully exploit acceleration facilities from the PC graphic cards hardware. Objects are modeled in this work using 3D polygonal meshes. A virtual scene with a virtual 3D sensor can be created with the development tool designed and described in the paper and the results are compared with real 3D laser sensor data. The system is used to study and optimize an object registration process, based on a hierarchical version of the iterative closest point algorithm, in conjunction with progressive mesh data representations. Since the system is modular and abstract, it can be used for any application for object localization, robot control, grasp point definition and collision avoiding strategy development. New objects can be easily added, as well as new sensors. Sensor noise effects can be incorporated into the model, resulting in realistic and reliable simulations. The paper of Boehnke also includes an excellent presentation of some non contact 3D image sensors used in non-contact industrial measurements.

A second paper dealing with industrial robots and 3D sensors is signed by **Roebrock**. The paper approaches the problem of a sensor-based tool positioning system for an industrial robot. Sensors are treated as abstract data sources, therefore different kinds of sensors can be considered at the same time in order to fulfill a required task. The system uses a closed-loop control in conjunction with a real time robot interface. Very high repeat accuracies, around 0.15 mm were obtained in a stable way by the author. Good robustness and excellent maintainability are additional assets of the proposed system. Future developments of the system are also discussed in the paper.

3 Conclusion

A brief introduction of the most widely used 3D optoelectronical sensors in computer vision was provided, with highlights on the benefits of active methods. The introduction was meant to set the right context for the subsequent discussion of the work on 3D sensors reported this issue. The included papers cover distinct and representative areas from the very broad range of now days' 3D sensor applications.

References

[1] Azriel Rosenfeld and Avinash Kak (1982). *Digital Picture Processing*. Academic Press. ISBN 0-12-597301-2.

[2] David Marr (1982). *Vision*. W. H. Freeman and Company. ISBN 0-7167-1284-9.

[3] Milan Sonka, Vaclav Hlavac and Roger Boyle (1999). *Image Processing, Analysis, and Machine Vision*. PWS Publishing. ISBN 0-534-95393-X.

[4] Bernd Jähne (2002). *Digital Image Processing*. Springer. ISBN 3-540-67754-2.

[5] Nikos Paragios, Yunmei Chen and Olivier Faugeras (2005). *Handbook of Mathematical Models in Computer Vision*. Springer. ISBN 0-387-26371-3.

[6] Richard Hartley and Andrew Zisserman (2003). *Multiple View Geometry in computer vision*. Cambridge University Press. ISBN 0-521-54051-8.

[7] Vanam Upendranath, *Smart CMOS sensor for 3D measurement,* PhD thesis, Informatica e Telecomunicazioni, DIT - University of Trento

[8] R. Lange and P. Seitz. *Solid-State Time-of-Flight Range Camera,* IEEE Journal of Quantum Electronics, vol. 37, no. 3, pp. 390-397, 2001.

[9] L. Viarani, D. Stoppa, L. Gonzo, M. Gottardi, and A. Simoni, *A CMOS Smart Pixel for Active 3-D Vision Applications*, IEEE Sensors Journal, vol. 4, no. 1, pp. 145-152, 2004.

[10] P. Besl, *Advances in Machine Vision*, Springer-Verlag, chapter 1 - Active optical range imaging sensors, pp. 1–63, 1989.

[11] J.-A. Beraldin, F. Blais, L. Cournoyer, G. Godin, and M. Rioux, *Active 3D Sensing,* published in Modelli E Metodi per lo studio e la conservazione dell'architettura storica, University: Scuola Normale Superiore, Pisa, 10: 22-46;2000. NRC 44159. April 2000.