STUDY OF 3D RECONSTRUCTION FROM MULTI-VIEW OR MULTI-POINT IMAGE PROCESSING AND IT'S APPLICATIONS

Author's Name: Ashadul Haque

Affiliation: Assistant Professor, Department of Computer Science, Vivekananda Mission Mahavidyalaya, West Bengal, India

E-Mail ID: ashadulhaque81@gmail.com

DOI No. – **08.2020-25662434**

Abstract

An image can be represented using 2d array of pixels and each pixel can have value between 0-255. An image can be gray or color based on number of plane and value of red, green and blue plane. A 3D image represented more realistic view than 2D images as it has more information or data than 2d images. A 3D image can be constructed using multiple camera and reconstruct an image after fusion of two or more images taken from different camera. In this paper, I have discussed about 3D image, different applications of 3D images and related works done on 3D image like dynamic object tracking, stereo image, re-identification, image mosaicking, hazard identification, depth estimation, 3D reconstruction etc. In this paper I have discussed limitations and compared different methods.

Keywords: *3D Reconstruction, Multi-View, Multi-Point, Image, Application, Processing*

INTRODUCTION

A simple image is two dimensional (2 -D) array of pixel where an 8 bit gray pixel represent the pixel intensity values between 0-255. An image can be presented more real which have three-dimensional (3-D) view. It is possible to present an image or video more realistic using two or more camera placing at different location placing on same object or place it slightly different view point. This allows production of multi-view sequence of video[1]. There are many researches going on 3-D image processing and multi-view display. For example, 3DTV project or DISTIMA aimed at developing a system for capturing, encoding, transmitting and presenting digital stereoscopic sequences [1]. This project leads to another project called PANOROMA which try to enhance visual information in telecommunication using multi-view concepts. Multi-view display are more realistic and natural compare to auto-stereoscopic display technologies such as in multi-view 3-D display user does not require any special glass. In multi-view system not require any head tracking to provide motion parallax [2].

Multi-view image processing is also known as data fusion or data integration from multiple images or multiple feature sets. 3D image reconstruction from 2D images has been a long standing problem in computer vision and it has many applications such as virtual reality, 3D visualization, image mosaicing, depth estimation and rendering, 3D augmentation, 3D rendering etc.[3]. There are mainly four types of image fusion can be categorized, which are multi-modal fusion, multi-temporal fusion, multi-focus fusion, multi-view fusion [4].Images are taken from different viewpoint and merge these images based on feature set is called multi-view fusion. Image are taken same scene but in different times is called multi-temporal fusion. Multi-focus fusion fuses different images have common scene but focuses on different parts [4]. Multi-modal fusions are like image have taken from different devices which may have different configuration.

RELATED WORK

Dynamic Object Tracking

A simple picture capture by only one camera have less details than same picture capture by multiple camera and reconstruct it. An image capture from multiple cameras hasmany applications in human identification, 3D image reconstruction, entertainment, sport,activity recognition etc. An static object can capture from multi point camera in different location set up as ring of camera across the fixed point object.But in case of dynamic object like sport, it is very difficult to predict early the position of the object's direction, therefore it is necessary to adjust dynamically the position of the object in multi-view or multipoint system. Dynamic object can be track in real time system by adjusting camera's tilt,focusing point, and zooming pan to acquiresynchronous multi-view video[5].

Stereo

One of the classic research problems is to reconstruction of three dimensional (3-D) object or shape from two or more image in computer vision which is called stereo. Applications of Stereo are many like robot navigation, object recognition, realistic scene virtualization. Stereo matching is the process to recover a 3-D object from pair of image. It is observe that using two or more image dramatically improved the quality of the reconstruction[6]. As more image increase the occluded portion or portion of an object which is partially visible in an image may be visible in other image. So occlusion problem may be resolve using two or more image. Some technique to resolve occlusion problem are firstly take combination of shift able window and match dynamically with selected subset of neighbor image. Secondly occluded pixel is label within global energy minimization framework and match with truly visible framework.

Video understanding

Video understanding algorithm have been developing to automatically detect a moving object, vehicle or human etc and track them using network active sensor[7]. They present the object to the controller who controls using graphical user interface of their three dimensional location using geospatial site model. This help to automatically collect and disseminate real time information to improve situational awareness of security provider and decision maker[7]. Background maintenance is one of field of video surveillance system. Three level component systems are there for background maintenance which are pixel level component, region level component and frame level component. In pixel level model, preliminary differentiate between backgrounds with foreground. Adapt changes in background but avoid some common problem like moved objects, time of day etc. All pixel level processing happens within same pixel and ignore the change in other pixel. In region level, it considers relationship or changes between inter pixel. It helps to identify the classification in pixel level. In frame level address the light switch problems [8]. It checks sharp change in large part in the image and changes in background. In frame level each image is subtracted with previous image, if difference between this two are higher than some threshold value then it will mark as foreground. Other parts areconsidered as background.

Re-identification

One of research work in multi-camera surveillance system is person re-identification. In one camera takes short video and store interest-point based on the short video. It matches with interest-point which is generated from other camera video sequence and match with store feature-point. The system generally takes logarithmic dependence with number of stored person models. Re-identification algorithms to identify some person have to be robust even in challenging situation like cause by different view-point and orientation,pose, lighting or changes in cloth appearance. Simple category of re-identification is using biometric technique like face detection or retina detection. Second group of re-identification are signature based color histograms, panoramic model from multi-view or texture characteristics [9]. Re-identification method can be details in five step:

a) **Model building:** A signature is built and store the feature point of different people based on the taken video sequences.
b) **Query building**: In this step exactly same way as models target persons is built on several evenly time-spaced images.Number of image for query building is less than the model building.
c) **Descriptor comparison**: Sum of Absolute is used for measuring the similarity between two interest point descriptor.
d) **Robust fast matching**: Using Camellia function a robust and very fast matching is done in KD tree contains all models.
e) **Identification**: A vote is added for each model contains a close enough descriptor, finally identification is made based on highest voted model [9].

Multi view concept applies on many fields but it has some disadvantages like increase in computational complexity of the system and visual information.

Image mosaicing

Image mosaicingis a method to combine two or more image into larger image. Image are combined based on two method, first method is to match the pixel intensity of an image with other image. Other method is feature based method where feature of an image are taken and combine the two images based on the similarity of the feature. Image mosaicing consist of five phases which are feature point extraction, Image registration, Homograph computation, Warping and Blending [10].In first phase feature of the two images are detected. In image registration, geometric alliance is generated based on common reference point to compare, transform or analysis two images. In image warping phase is to correct the distorted image and Blending is the technique to modify the boundary between images to obtain a smooth transition between two images.

Hazard identification

To travel autonomously, mobile robots have to identify different objects in different situations to not get into hazards. AniketMurarkaet.al. presented real-time stereo based mapping algorithm in [11] for identify various hazards. To travel in urban area, mobile robot has to identify and deals with different kind of potential hazards like drop-off or incline surface. Robot has to calculate the slope of the incline surface unless it will slip. There are many kinds of potential hazards like Static obstacle are wall or furniture where dynamic obstacles are people, door etc. Drop-offs obstacle are sidewalk curbs, downward stairs and inclines obstacle are wheelchair ramp, slope sidewalks etc. In [11], presents an annotated 2-D grid map called local safety map based on algorithm identifies safe and unsafe region in the robot's surrounding. The stereo image process in four steps which are given below:

a) First stereo image computed and convert it into depth map, depth reading are transform into map coordinate [11].
b) New depth reading update the 3-D model containing 3-D grid and 3-D point cloud using an occupancy grid algorithm.
c) Planes are fit to potentially traversable ground segments in the 3-D model [11]. Using linear least square algorithm segment the 3-D grid and fitting plane to points corresponding segment.

d) Finally, from safety map segment and plane are analysis for safety.

Depth Estimation and Depth based image rendering

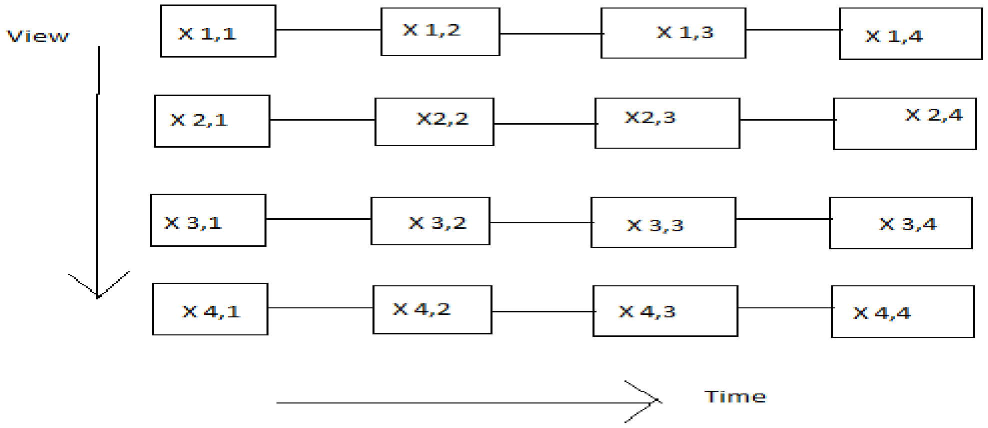
We can convert 2-D image into 3-D image by generating another view which involve two steps, one is depth estimation and other depth base image rendering. In depth estimation step we convert a 2-D image into a depth map or depth image which store depth information in 8-bit gray value of each pixel. In depth map, 0 or a black pixel means farthest element or object and 255 or white pixel means nearest object. Multi-view video consume more space or need more bandwidth compare to 2-D video. Due to huge space needed to communicate in 3-D video, we need more efficient compression technique. In multi-view an image is capture from different position, so we can get common object or similarity of the multi-view. This similarity can be classified into two types, one is inter-view similarities between images of the adjacent camera and other is temporal similarities which are temporally similarities between successive images of the same video [12]. Markus Flierl et. al. present in [12] natural presentation of multi-view video classify into Matrix of Picture (MOP). In MOP each row present temporal successive image of a camera or image of one view where each column represent picture taken from different view or different camera at same time. An object is nothing but interest point or entity in an image. Object can be airplane in air, car in road or fish in aquarium. Object can be presented various way, they are-

a) Points: Objects can be represented as set of points which are centroid.

b) Primitive geometric shape: Object can be represented by rectangle, ellipse or other primitive geometric shape.

c) Object Silhouette and Contour: Contour means edge or boundary of an object where Silhouette means region inside contour which is suitable for non rigid object.

d) Skeletal model.



[Fig 1] Matrix of Picture (MOP)

3D Reconstruction and Inter-frame prediction

Stephan et. al present in [13] 3D video fragment , a dynamic point sample framework which can be insert, delete or update. Each point consists of set of attributes likes colours, geo-metric location or position, surface normal vector etc. In point based 3D image representation is more efficient for dynamically updating, compression, progressive streaming and sampling due to lack of local connectivity. Stephan et. al acquire a 3D image in [13]. 3D point sample are generated from set of active camera and set of supporting camera are used to improve the 3D reconstruction. Differential operator is generated from inter-frame prediction in an image which helps to update point sample attributes including position or colour dynamically.Different types of operator are used on image fragment like INSERT, DELETE or UPDATE. INSERT operator is used to insert a new fragment to the view of the input camera. DELETE operator is used to delete fragment from view of the input camera where as UPDATE operator used change appearance or geometric attribute of the frame from prior appearance.

Alternative of LIDAR

C. Strechaet. al. discuss in [14] whether image based 3D modelling technique can replace Light Detection and Ranging (LIDAR) technique for outdoor 3D data acquisition. Camera calibration (internal and external) and dense multi-view stereo image are the two issue address in [14]. Several technique are available like laser measurement (LIDAR), active stereo, passive stereo or NMR imaging to measure an object in 3D. Active stereo can measure 3D coordinate of an object in real-time in laboratories or controlled indoor environment but not outdoor environment. Another approach to measure uncontrolled outdoor environment is LIDAR technique. LIDAR is method which is used to measure distance to a target by illuminating the target using pulse laser light and measuring the reflect pulse using a sensor. LIDAR use ultraviolet, visible or near infrared light to measure object like non-metallic object, terrain, rock, rain, chemical compound etc. Data use and generate using LIDAR are digital. So LIDAR system can be use in GIS or satellite. LIDAR systems are able to produce directly 3D point cloud based with accuracy of less than 1 cm but time consuming process to acquire the data and costly [14]. An alternate to LIDAR system is multi view image reconstruction method as it is a low cost alternative.

RELATED ARTICLE AND THEIR LIMITATION

Sl. No	Work Description	Limitation
1	Xiaoduan FENG et. al present a new method to acquire not only multi-view image [15] but also in multi-luminance that work effectively in to image shadows, noise and high light.	This method not very effective in more complicated environment.
2	Mustafa Oral et. al. describe four types image fusion method and compare their performance [16]. In these four types, Principal Component Analysis and Unique Color are two spatial fusion methods where Discrete Cosine Transform and Discrete Wavelet Transform are the two frequency fusion methods.	Out of four fusion method, Unique Color method is simple and easy to implement then DCT method [14] but other method performed unsatisfactory. Unique Color method have some disadvantage also as sharp edge in blurred part have more Colors than focus part [16].
3	Jianguo Li et. al. presented a depth map merging based multi view stereo reconstruction using a novel two stage bundle optimization algorithm [17]. They able to produce high quality point cloud and remove outlier though generated depth maps do not have sub pixel level precision or erroneous [17].	In paper [17], not focus on adaptive DAISY radius during bundle optimization and many-core/GPU based implementation.
4	Jian Yang et. al. present an algorithm for face model registration [18] which combine active structured light and multi point cloud, analyzed difference between this two model and also compare with standard ICP algorithm and normal vector algorithm.	There is not mentioned about how effective to enhance the real-time

5	Yilong Liu et. al. present an algorithm to construct a multi-view stereo image to realize faster reconstruction procedure [19]. This algorithm not used visual hull prior as input reference and also it is a robust reconstruction method.	This model not worked properly in some extreme texture-less situation. This model also not worked well on uniform color images and smooth images.

CONCLUSION AND FUTURE WORK

In this paper I have discussed about 3D image, different applications of 3d images like Depth estimation,dynamic object tracking, hazard identification, stereo image, re-identification, image mosaicing, 3D reconstruction etc. Author also compared different methods and their limitations.

Our future work is to analyse an object in multi point or multi-view system and try to give more realistic view of an object using two or more camera placing at different locations focusing on same the object. And also we try to develop a more efficient system for capturing, encoding, transmitting and presenting three dimensional (3 -D) images or video in multi-view or multi-point systems.

REFERENCES

1. Yongtae Kim, Jiyoung Kim, KwanghoonSohn, "Fast Disparity and Motion Estimation for Multi-view Video Coding", IEEE Transaction on Consumer Electronics, Vol. 53, No. 2, MAY 2007.
2. Matthias Zwicker, Anthony Vetro, Sehoon Yea, WojciechMatusik, HanspeterPfister and Fredo Durand, " Resampling, Antialiasing, and Compression in Multiview 3-D Displays", IEEE Signal Processing Magazine, November 2007, 1053-5888/07.
3. Sai Bi, Zexiang Xu, KalyanSunkavalli, David Kriegman, Ravi Ramamoorthi,"Deep 3D Capture: Geometry and Reflectance from Sparse Multi-View Images", CVF Conference on Computer Vision and Pattern Recognition (CVPR).
4. Mustafa Oral, Sultan SevgiTurgut,"A Comparative Study for Image Fusion", 2018 Innovations in Intelligent Systems and Applications Conferences, IEEE.
5. Robert T. Collins, OmeadAmidi. and Takeo Kanade, "An active camera system for Acquiring multi -view video", Robotics Institute, Carnegie Mellon University, IEEE ICIP 2002, 0-7803-7622-6/02/.
6. Sing Bing Kang, Richard Szeliski, Jinxiang Chai, "Handling Occlusions in Multi-view Stereo", IEEE 2001,0-7695-1272-0/01.
7. Robert T. Collins, Alans J. Lipton, Hironobu Fujiyoshi and Takeo Kanade, "Algorithm for Cooperative Multisensor surveillance", Proceedings of the IEEE, VOL. 89, NO. 10, October 2001.
8. Kentaro Toyama, John Krumm, Barry Brumitt and Brian Meyers," Wallflower: Principles and Practice of Background Maintenance", Microsoft Research, Redmond, WA 98052.
9. Omar Hamdoun, Fabien Moutarde, Bogdan Stanciulescu and Bruno Steux, "Preson re-identification in Multi-camera system by Signature based on Interest point descriptors collected on short video sequences", Robotics laboratory (CAOR), Mines ParisTech,France, 978-1-4244-2665-2/08/ 2008 IEEE.
10. ChaudhuryKajalbenMohanbhai, Mrs. HetalBhaidasna, "A survey On Image Mosaicing Using Feature Based Approach", IJEDR 2017,Volume -5.
11. AniketMurarka, Benjamin Kuipers, " A Stereo Vision Based Mapping Algorithm for Detecting Inclines, Drop-off, and Obstacles for Safe Local Navigation", Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on 10-15 Oct. 2009.
12. Markus Flierl and Bernd Girod, "Multiview video Compression, Exploiting inter-image

similarities", IEEE Signal Processing Magazine, November 2007,1053-5888/07/$25.00

13. Stephan Wurmlin, Edouard Lamboray, Markus Gross, "3D video fragments: dynamic point samples for real-time free-viewpoint video", Computers & Graphics 28 (2004) 3-14, Elsevier.
14. C. Strecha, W. Von Hansen, L. Van Gool, P. Fua, U. Thoennessen, " On Benchmarking Camera Calibration and Multi-view Stereo for High Resolution Imagery", 978-1-4244-2243-2/08, 2008 IEEE.
15. Xiaoduan FENG, Yebin Liu, Qionghai DAI," MULTI-VIEW STEREO USING MULTI-LUMINANCE IMAGES", 3DTV-CON 2009,IEEE.
16. M. Oral, Sultan Sevgi Turgut, "A Comparative Study for Image Fusion", 2018 Innovation in Intelligent Systems and Applications Conference (ASYU), 2018, IEEE.
17. Jianguo Li, Eric Li, Yurong Chen, Lin Xu, Yimin Zhang, "Bundled Depth-Map Merging for Multi-View Stereo", Computer Society Conference on Computer Vision and Pattern Recognition, IEEE, 13-18 June 2010.
18. Jian Yang, Hui Chen," The 3D Reconstruction of Face Model with Active Structured Light and Stereo Vision Fusion", 3rd International Conference on Computer and Communications, 2017, IEEE.
19. Yilong Liu, Yuanyuan Jiang, YebinLiu,"A VISUAL HULL FREE ALGORITHM FOR FAST AND ROBUST MULTI-VIEW STEREO",International Conference on 3D Imaging, 2011, IEEE.