Augmented Reality: Technology Merging Computer Vision and Image Processing by Experimental Techniques

Vaishnavi J. Deshpande, Jignyasa Sanghavi

Abstract*: Computer vision is an interdisciplinary field which forms the foundation for many evolving technologies. Augmented reality (AR) and virtual reality (VR) are the rapidly developing technologies used in many applications. Augmented Reality superimposes a computer generated image on the user's view of the real world. Virtual reality is an interactive computer-generated experience taking place within a simulated environment. With the research and progress, the technology has given many applications based on Augmented Reality to the real world. The technology has three main pillars: motion tracking, environmental understanding and light estimation. The research in this field is also influencing a new human-computer interaction. The objective of this paper is to understand the ongoing research in AR. This paper discusses all the technical details of the AR technology, the algorithms used to achieve various functionalities and the future scope of the technology.

***Index Terms*: Augmented Reality, motion tracking, SLAM, Computer Vision, image processing**

I. INTRODUCTION

Computer Vision is the field which has methods for acquiring, processing and analysing the digital images. These tasks are used in Augmented Reality. Augmented Reality is a technology which superimposes a computer generated image on the user's view of the real world. This composites the real and virtual world in device's space. It is the set of technologies that that superimposes the digital data to physical world. Computer vision algorithms have wide applications today. Nowadays, AR is the hottest topic for discussion. Augmented Reality technology was first introduced in 1968. This was invented by a Harvard professor and his student. The first innovation in this technology was featured with head mounteddisplay.

From here, many projects were conducted in augmented reality which would make user to experience computer graphics. Later on, with the use of image generated system, i.e., camera, display and GPS, synthetic vision system was used by NASA to improve navigation of the space crafts during the test flights.

The applications of AR are ranging from gaming to social media. The face filters used in instagram is one of the Case studies of augmented reality. Or the Pokemon Go which had crossed the number of users and became very popular in short time is also one of the applications of AR.

The field has avery wide range of applications as architecture, engineering applications and construction services.

There are three types of models in Augmented Reality:

1. Marker basedAR
2. MarkerlessAR
3. Location basedAR

In marker based augmented reality, contextual digital information and animations are tied to an augmented reality marker in the real world [5]. The question might arise, how computer vision is related to augmented reality. To answer this question, let us start with understanding the components of augmented reality system.

Components of Augmented Reality System:

- Userdevice
- Camera
- Sensors
- CPU

II. DETAILS OF SENSORMODALITIES

The perceptually enriched experiences offered by AR are because of computer vision which serve as additive to enhance user experience. The device needs sensors to simulate the data from real world. These sensors are required to get the location, to capture the real world data, i.e., the digital images, position of the device in the world space, capture or record audio, track the motion of the object, etc. The sensors can be more easily understood with the help of human capabilities:

- Ability to see the world:Visualisation
- Listen and speak: Audio Recorder & AudioListener
- Touch
- Ability of mind to recognise the object, location:Object recognition, Locationmapping

III. PROPOSED WORKSYSTEM

The following diagram is the simplest view of the AR Work flow. The components of the block diagram are explained in detail below. For any AR application, the first step is to detect the features and extract, i.e., refinement of the detected features.

The camera calibration technique relates to mapping of the 3D features on the image plane. In AR system, camera plays a vital role to capture the image frames. That is why, it is very essential to understand the camera features, intrinsic and extrinsic, orthographic and projective view systems, camera matrix calculations and so on.

Next comes the rendering part of the system. Rendering techniques are required to generate the image frames pixel by pixel. This involves graphical rendering of Augmented objects.

Revised Manuscript Received on December 22, 2018.

Vaishnavi J. Deshpande, M.Tech CSE Final Year Student, Shri Ramdeobaba college of Engineering & Management, Nagpur.
JignyasaSanghavi, Assistant Professor Department of Computer Science, Shri Ramdeobaba college of Engineering & Management, Nagpur.

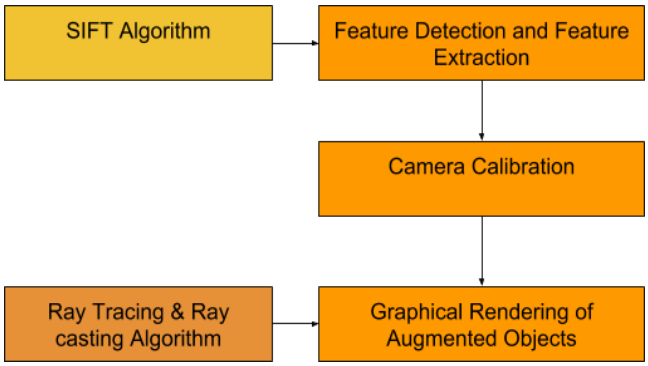

Fig.1 Working of AR System

IV. FEATURES

How the human brain recognizes particular object is because of the features of the object. For eg., chair has four legs and a typical structure which makes us understand that object with such a structure is a chair.

In the marker based AR model, we have one particular image which serves as a marker for the AR system. AR system calculates the histogram for the image marker.

Whenever the histogram of the real world data matches with that of the histogram of the image marker, the particular marker is detected and augmented object is placed on themarker.

Features (feature points) play an important role to find the solution for the tasks related to any augmented reality application. Feature detection and extraction are regarded as the preprocessing or the starting step of any computational task. Feature can be edge, corner, blob or ridge. There are various algorithms like Canny Edge Detector, Shi Tomasi algorithm to detect and extract features. In computer vision, feature is represented by feature descriptor.

V. CAMERACALIBRATION

Camera calibration is the process to match the camera parameters of the virtual camera to that of the physical camera. The goal of the camera calibration is to recover the mapping between the 3D space and the image plane, which can be separated into two sets of transformations [1].

The main application of this is the 3D scene reconstruction in AR which we will discuss in detail below. The camera matrix is obtained from the calibration of camera using intrinsic and extrinsic parameters.

	Extrinsic Parameters	Intrinsic Parameters
1.	Rigid transformation	Projective transformation
2.	3D world coordinate system-> 3D camera coordinate system	3D camera coordinate system-> 2D image coordinate system

Table: Difference in Extrinsic and Intrinsic Parameters

VI. SLAM

SLAM is an abbreviation for simultaneous localization andmapping. This algorithm simultaneously estimates the pose of robot and the map of the environment. SLAM is "chicken or egg" problem because of the following cycle:

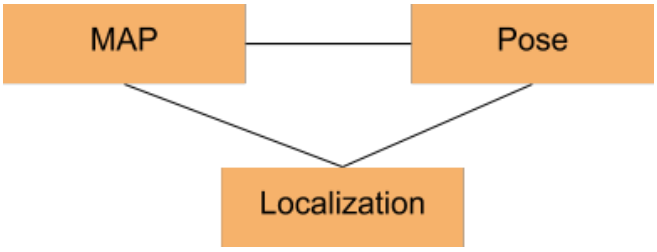

Fig 2. Working of SLAM

VII. IMAGEREGISTRATION

This algorithm is used in computer vision. As discussed previously, the real time data, i.e., moving images are captured with the help of camera. This uses different sensors like GPS, accelerometer, etc. In order to synchronize the data obtained from different measurements, we need image registration. In the past few years the global need for low computation, less time consuming, and good quality image mapping methods has caused an image registration technique alive in multiple application areas[2].

The basic steps of image registration algorithm are:

1. Feature detection andmatching
2. Feature mapping functiondesign
3. Imagetransformation
4. Resampling

VIII. VISUALODOMETRY

Environmental understanding is a very important aspect in augmented reality. The following would be very interesting questions in order to understand the environment:

- Environmental understanding deals with understanding the position of camera, and then analyzing the visual features of the scene. So the question arises, what sensor in the device is used for thispurpose.
- What algorithms are used to determine the visual features
- What are the typical modules required for estimating the 3D pose of thecamera

The "visual odometry" covers all the answers to the above questions. Inertial measurement unit(IMU) is a system in the device which is used for the calculation of odometry measurements.

The downside of IMUs is, however, that they measure relative poses only indirectly through rotational velocities and linear accelerations [3].The pose can be estimated with the help of accelerometer and gyroscope.

According to the efficiency and accuracy, SIFT and SURF(improvised version of SURF, i.e., OpenSURF) are the algorithms used when we talk of feature extraction. A feature is nothing but the point of interest.

Steps involved in this algorithm:

1. Feature detection andextraction
2. Feature matching across differentframes
3. Calculation of camera matrix(Cameracalibration)
4. Determine relative poses between the image frames

Assumptions:

When we talk of feature matching across different frames, this can be a simple process in case of monocular camera. But the process becomes complex or even fail, when stereopair scenario is taken into consideration.

IX. RAYTRACING

The algorithm deals with forming the image which is obtained by tracing the direction of ray to the object. The

image formed is more realistic. The realism of the image obtained by the camera depends on the rendering. In case of ray tracing, batch rendering is used to achieve realism. Ray tracing works on pixel by pixel.

Although the algorithm gives accurate results, it takes time when the scene involves large objects. Geometric rays are used to calculate the pixel intensity. Ray tracing renders 3D scenes to 2D images.

X. RAYCASTING

Computer graphics is the interdisciplinary field dealing with computer vision, image processing and augmented reality. Ray tracing and ray casting are the two algorithms in the computer graphics whose mechanism is completely opposite, but both are important in rendering the objects on the camera screen. Light has the basic characteristics like reflection, refraction and absorption. These characteristics of light are taken into the consideration because of the following reasons:

• The objects are visible on the camera because of light. It is necessary to understand the materialistic properties of objects. This is quite interesting as well as important because while rendering, we should get realism to the maximumestimate.

• The ray can be considered as the stream of photons. In ray casting, color of the light is computed. The shading kernel evaluates the color contribution of a given ray at the hit point [4]. Texture maps are used for rendering the 3D objects. The algorithm samples the radiance incident on the point that the rayhit.

XI. SEMANTIC SEGMENTATION

The idea behind semantic segmentation is to recognize and understand the image at the pixel level. This algorithm is somewhat the part of machine learning. For eg., consider the image containing table, chair, books and different objects. What semantic segmentation will do is the segregation of the different objects with the help ofcolors.

The task is quite challenging. In order to do semantic segmentation, it is necessary to understand the environment. The core of the task is to detect the features and classify the features based on the intensity differences in the pixel values.

Role of semantic segmentation in Augmented Reality:

When we fuse semantic segmentation with augmented reality, it is the advancement which the user can experience the robustness in technology. This is because we can integrate
deep learning and neural networks in th the augmented reality system after we do semantic segmentation. Semantic segmentation faces an inherent tension between semantics and location: global information resolves what while local information resolves where [6].

When the feature extraction is done, pixels are classified based on the features. Now, to classify the feature, we can use conditional random field.

In order to get the labeled and segmented structured data, conditional random field is used which is the probabilistic framework. Nowadays, fully convolutional networks with Recurrent Neural Networks areused in semantic segmentation. The network is trained back and forth on the maximum number of samples to get accurate classification ofdata.

Semantic segmentation has many applications in medical field. How we can integrate semantic segmentation in AR is for the applicationsofobject identification and classification, localization and mapping applications where we can provide additional information regarding the spatial location and so on. Semantic segmentation is approached with encoder and
decoder.

XII. APPLICATIONS

Nowadays, AR applications are boom in the market. In this
section, we discuss a few applications of augmented reality:

1. Learning applications for kids: Educational applications, Libraryapplication

This is the educational purpose application of AR. To teach the kids the basics like alphabets, numbers, poems and story telling activities, etc. we can use augmented reality which will be fun and interactive learning process.

2. Architectural applications: Urban planning and civil engineering

The customer would always love to know how the property-home, office, etc would look before actual construction. The AR application in which they will be able to view the 3D model of the house will be help full in this case. The Ikea like apps where one can view the furniture before purchasing is one of the applications of augmentedreality.

3. Robotics and Navigation: Pathfinding
4. Objects recognition andclassification
5. Plane detection andclassification
6. Measurement applications: Measure app,ARRuler app
7. Instagram face filters,snapchat
8. Gaming: PokemonGo
9. Instagram face filters,snapchat
10. Gaming: PokemonGo

XIII. FUTUREWORK

Augmented Reality can be enhanced more by integrating it with neural networks. The research in this direction has already started.

However, it has not gained momentum as it has expected. What next can be done is that convincing the user with more advanced UI experience which AR can provide.

Recently, there is a news that Google maps will provide navigation systems with the help of Augmented Reality technology.

XIV. RESULT

The following is the screenshot of the work done for marker
based AR.

Fig 3. Marker based AR Demo

When we focus the camera on the front cover of the book, the additional details regarding the book like information of the author, amazon reviews of the book, linkedin profile of the author and other related books are displayed on the screen as the augmented data.

XV. CONCLUSION

The main motto behind the launching of augmented reality technology is that visualisation technology can reduce the need for physical materials. Augmented reality uses computer vision, computer graphics and image processing domains which has increased the radius of research field to infinity. AR technology focuses from CPU computational power to 3D rendering pipeline to hardware acceleration.

The technology introduced in the last decade has become today the hottest technical topic for discussion. Many top companies like Google are investing in this technology.

REFERENCES

1. YueqiangZhang, Langming Zhou, Haibo Liu, and Yang Shang, "A Flexible Online Camera Calibration Using Line Segments", 1-16, Hindawi Publishing Corporation Journal of Sensors Volume2016
2. Sombir Singh Bisht, Bhumika Gupta, Parvez Rahi, "Image Registration Concept and Techniques: AReview", IJERA, Vol. 4, Issue 4, 30-35, April2014,
3. VladyslavUsenko, Jakob Engel, JorgStuckler, Daniel Cremers, "Direct Visual-Inertial , (Mapping on Demand) of German Research Foundation (DFG), 3-8, 2016
4. Timothy J. Purcell, Ian Buck, William R. Mark, Pat Hanrahan, "Ray Tracing on Programmable Graphics Hardware", SIGGRAPH '02 Proceedings of the 29thannual conference on Computer graphics and interactive techniques, 703-712, July2002
5. AhishAgrawal ,Gourav Acharya, Krishna Balasubramanian, Nehal Agrawal, Ratnesh Chaturvedi, "A Review on the use of Augmented Reality to Generate Safety Awareness and Enhance EmergencyResponse", 813-820, International Journal of Current Engineering and Technology, Vol. 6, 2016
6. Long, Shelhamer, Darrell, "Fully Convolutional Networks for Semantic Segmentation", CVPR, 1-10,2015

AUTHORS PROFILE

Vaishnavi Deshpande received bachelor of engineering degree in Computer science and engineering from Jawaharlal Darda Institute of Engineering & Technology, Yavatmal. Currently pursuing M.tech from Shri Ramdeobaba College Of Engineering And Management, Nagpur in computer science branch and research paper work focuses on Augmented Reality.

JignyasaSanghavi*is* Assistant Professor at Shri Ramdeobaba College of Engineering and Management, Nagpur, India. She received B.E. degree and M.Tech degree in Computer Science and Engineering from RashtrasantTukodoji Maharaj Nagpur University. She has teachingexperienceof8.5yearsinComputerScienceand Engineering department. She is member of ACM. Her research interest includes Image Processing, Parallel Computing, Databases and Algorithms. She has published 10 research papers in various conferences, workshops and InternationalJournals.