Computer Vision and Sensor fusion for efficient Hybrid Tracking in Augmented Reality Systems

Deepti Prit Kaur
School of Electronics and Electrical Engineering
Chitkara University
Rajpura, Punjab
deeptiprit.kaur@chitkara.edu.in

Dr. Archana Mantri
Chitkara University Research and Innovation Network
Chitkara University
Rajpura, Punjab
archana.mantri@chitkara.edu.in

***Abstract*— Augmented Reality is one of the most fundamental enabling technologies of this era which modifies our perception to such an extent that we are able to see, hear and feel the ordinary everyday objects in a new and enriched way. This technology is an amalgamation of Real and Virtual Worlds which aims to enhance the physical information by exactly super-imposing the computer generated content on it. The technology can prove helpful to surgeons and doctors to visualize and train a surgery, can instruct a mechanic for repair of an unknown piece of equipment, can serve as a source of entertainment, can help soldiers for spotting enemy snipers, and can be used as an efficient teaching aid for education. A typical Augmented Reality (AR) system requires the integration of hardware and software with factors such as Tracking, Registration, Interaction and Display playing an important role in the system design. The focus of this paper is specifically towards surveying the available Tracking Techniques for various Augmented Reality applications. Based on the existing methods of sensor based, vision based and hybrid tracking, a new method for efficient tracking is proposed.**

***Index Terms*— Augmented Reality, Sensor Based Tracking, Vision Based Tracking, Hybrid Tracking, Mobile AR, Augmented Reality for Education.**

I. INTRODUCTION

Augmented Reality finds many application areas, education being the newest of all. According to L. Johnson, A. Levine, R. Smith and S. Stone [3], AR has very strong potential to provide both learning experiences and exploration of information in the real world. During last few years, researchers have been developing theories and applications for adoption of AR into academic settings to enhance the education and training efficiency of students and trainers. In an AR System, the human perception is refined by overlaying virtual information on the physical objects, thereby enabling the human users of this system to sense the physical world around them augmented with computer generated virtual information. The user interaction with the real world happens in a natural way and during same time, a computer is used for generating the information related to the real world. The user then interacts with computer-generated virtual objects in the real time. The virtual objects represent the information which is not directly detectable by the user's own senses. Thus, a combined view of real object and a computer-generated virtual object is generated for the user in a typical AR system. The virtual object augments the real environment with additional information but preserving the user's sense of being in the real world by merging the virtual data with a view of the real scene to create the augmented display. According to R. Azuma [2], augmented reality has following characteristics:

- Combines the real and virtual objects in the real environment thereby allowing the user to see virtual and real information in a combined view simultaneously.
- Interactive in real time. As it runs at interactive frame rates, the information can be superimposed in real-time allowing user interaction.
- Registered in 3D by rendering the virtual content which is placed with respect to the real.

Simplification of the reality-virtuality continuum (RVC) by P. Milgram and F. Kishino [4] is represented in Figure 1, which has real environment and virtual environment at the two opposite extremes with mixed reality (MR) as the middle region. Mixed reality includes Augmented Reality and Augmented Virtuality.

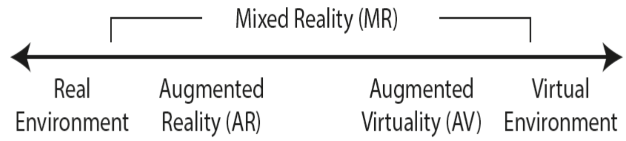

Figure 1: The Reality-Virtuality Continuum

Two basic technologies are used to accomplish the real and virtual combining for an AR application. One uses head-mounted display (HMD) equipment which has a wearable computer, and the other makes use of computer-generated graphic images overlaid on the video captured by the camera [5]. According to D.W.F. van Krevelen and R. Poelman [21], the development of a typical Augmented Reality System involves integration of hardware and software, and consists of each one of tracking, sensing, display and interaction methods assembled together. Before an AR system can display virtual content into real world information, the system must sense the environment and track the relative movement of the viewer

978-1-4673-6747-9/15/$31.00 ©2015 IEEE

(preferably with six degrees of freedom - 6DOF). This movement calculation consists of three variables for position and three angles for orientation. Thus, for developing an augmented reality system, the key challenges are tracking and registration, which deals between the computer-generated and real world objects. Proper tracking is one of the main issues for any augmented reality application [1]. So remainder of this paper will detail various techniques available for Tracking in an Augmented Reality System and their performance analysis.

II. How to "Augment" the Reality?

Let us imagine a Technology, with which we are able to see, hear and feel the ordinary everyday objects more than their actual existence. This answer to this imagination is Augmented Reality, in which, human perception is enhanced by overlaying virtual information on the physical objects so that user is able to view or feel the modified real world information. According to Wendy E. Mackay [22], there are three basic strategies for augmenting the reality, in which an electronic system is merged into the physical world, instead of completely replacing it.

A. UserAugmentation

In this approach, user wears a device to obtain information about the physical objects or carries it. This technique is useful in Medicines, Education or Defense services, where AR devices such as Google Glass, Data Gloves or VR Helmets augment the reality for user.

B. ObjectAugmentation

When some input / output or computational device is embedded on the physical object, the object becomes intelligent enough to enhance the information for user. This technique is useful in Education and Office environments where electronic paper can be used for augmenting the reality. Global Positioning System (GPS) also falls under this category which is nowadays used for 'Ubiquitous Computing'.

C. Augmentation of the surrounding environment of User and the Physical Object

The information is captured or collected from the environment which is surrounding the user and the physical object to be augmented. For example, Entertainment Industry makes use of Video Cameras. Barcode scanners also falls under this category which may be used to turn 2D pictures into 3D, still images into Dynamic Video sequences etc., without modifying the physical objects with which the user is interacting.

III. Trends in Augmented Reality Tracking

Tracking the objects in an augmented reality system is significant for establishing the position and orientation of the object of interest in the real scene. In an AR System, when a user in Real World moves his/her position, the virtual objects must remain aligned with the position and orientation of real objects. The geometrical relationship between the viewpoint or position of the user, the display, and the environment needs to be known such that the combinations of real and virtual worlds are aligned accurately and properly. The process of finding this relationship is called Tracking. There is variety of Tracking Techniques to recover the position and/or orientation of users, displays and objects in the AR systems. The need for user tracking depends on many factors, such as how and where the virtual imagery is created and the display's placement relative to the user. The Tracking Techniques can be categorized as Sensor-based, Vision-based, and Hybrid tracking techniques and are applicable to both Indoor and Outdoor environments.

A. Sensor Tracking

For tracking the orientation of a camera pose, magnetic, optical, acoustic, inertial, and/or mechanical sensors can be used. Every sensor has its own advantages and disadvantages [1], [4]. Table 1 presents the analysis of various sensor techniques used for tracking in Augmented Reality applications.

TABLE 1: ANALYSIS OF SENSING TECHNIQUES

Sensing Technique	*Advantages*	*Disadvantages*
Magnetic Sensor; sensitive to Electro-magnetic waves	• high update rate • no occlusion problem (obstruction to the view) • light weighted	• less accurate (as the magnetic field from any nearby metal can disturb their performance)
Optical Sensors; sensitive to Light	• accurate • have high update rate • have much better resolution	• highly affected by optical noise • suffer from the occlusion problem.
Acoustic Sensors; sensitive to Temperature/ Humidity/ Pressure	• light-weighted • small in size • suffer no distortion	• suffer from occlusion • prone to ultrasonic noise • are less accurate in calculation of camera pose
Inertial Sensors; sensitive to Friction.	• very accurate • do not require any source or preparation in the environment	• they suffer from drift errors

Researchers are now exploring the fusion of different types of sensors so that robust tracking with effective sensor hand-over between different tracking technologies may be achieved. Combination of fixed optical tracking sensor is done with a head mounted optical see through device and a CCD camera in [6]. It results in more accurate pose estimation of head-to-object using a head mounted display. Dynamic fusion of Gyroscope is made in a technology given by D. Pustka and G. Klinker [7], with three different mobile and stationary sensors for Ubiquitous tracking setup. It is based on the concept of spatial relationship graphs (SRG) and patterns and increases the robustness and accuracy of the tracking technique. It also enables new wide-area concepts. A sensor fusion approach is presented by L. Porzi [8], which exploits the visual cues and inertial sensors of a Smartphone. It has robustness towards the appearance variations and is useful for local feature tracking.

B. Vision Tracking

Computer Vision methods are based on image processing as basic principle of operation. These are used to calculate the camera pose relative to the real objects. In fact, vision-based tracking is the most active area of research in AR [1]. The purpose of camera tracking is to continuously update the extrinsic parameters of a camera [5] which is used to determine the relationship between the camera view (which appears on the image plane) and actual scene (where it is located in 3-dimensional world). Camera tracking accurately tracks the motion of the camera in a known 3-dimensional environment and dynamically estimates the location of the camera. Vision Tracking is either Marker based or Marker-less [1], [20]. In Marker based tracking, fiducial (reference) markers are used to track the orientation of camera pose. The orientation of the marker is determined using Image Processing Algorithms and then the recognized marker region is used for estimating the position and orientation of the camera relative to the marker tagged object [5]. The fiducial design allows to stock different codes and enables uninterrupted tracking throughout a large building at very reasonable cost. Rather than using artificial markers, camera pose can also be determined from naturally occurring features, such as points, edges, textures or lines. In Fact, vision-based tracking research is based on marker-less approach nowadays. With a fixed camera, which captures the moving object in a scene, the term "outside-in tracking" is used because the object is observed from outside. On the contrary, if a camera is movable such as in a handheld device, and captures any fixed/moving object, the term "inside-out tracking" is used [17]. Table 2 presents the analysis of various vision techniques used for tracking in Augmented Reality applications.

TABLE 2: ANALYSIS OF VISION TECHNIQUES

Vision Technique	*Advantages*	*Disadvantages*
Marker based; Fiducial markers / Visual tags are used	• accurate and robust • Low jitter or drift problems • allows to stock different codes and enables uninterrupted tracking throughout a large building at very reasonable cost.	• Not suitable for large scale navigation, as information is augmented only when marker is in sight. • Also they result in visual clutter.
Marker-less; Natural feature extraction from real environment	• Accurate and robust • Dynamic data update is present for error correction • Low occlusion or drift problems	• High cost • More complexity • Execution time is very high, resulting in low update rate

W. Guan, S. You and U. Neumann [10] presented a recognition-based, marker-less user tracking and augmented reality system. This System makes use of Global Positioning System (GPS) for Outdoor Tracking and Patch-Retrieval Technique of Image Processing for precisely finding the specified location in a Large Area so that augmented information about the building (such as building names and locations) can be overlaid for a user who captures the image of a building facade. Although GPS is not able to provide information about camera poses, it is still needed so that the task of searching long ranges in image database is reduced. The patch-retrieval method is employed for efficient computations and real-time camera pose recovery, which increases the robustness of the tracking method towards occlusions and dynamics in the scenes. Also, the system simulates the feature of "soft handoff" to avoid frequent swaps in memory resource. A full image-based approach is presented by Y. Liu and X. Granier in [11] which works for the videos captured with moving cameras and provides online tracking of outdoor illumination variations. But the outdoor illumination depends on the weather, therefore; the lighting condition may change for different frames. Spatial and temporal coherence of illumination are then exploited to estimate the relative intensities of skylight and sunlight by using an optimization process. The estimations thus obtained are visually coherent along the video sequences. A People Tracking method is presented by A. L. Christian Waechter, D. Pustka, G. J. Klinker in [12], which is applicable to Augmented Reality for indoor Applications. Tracking is under research for various surveillance applications or human behavior estimation, but in this paper, Ubiquitous People Tracking is employed.

C. Hybrid Tracking

Image Processing and Computer vision alone cannot provide a robust tracking solution for some of the AR applications. According to Azuma [2], additional sensors and GPS besides the video cameras can aid registration and provide complete solution for Outdoor Augmented Reality. Therefore, many hybrid methods have been developed by researchers, which combine several vision and sensor technologies. Hybrid tracking techniques present the most promising way to deal with various challenges encountered in indoor and outdoor mobile augmented reality environments [19]. Figure 2 shows a typical hybrid tracking system for real-time object detection.

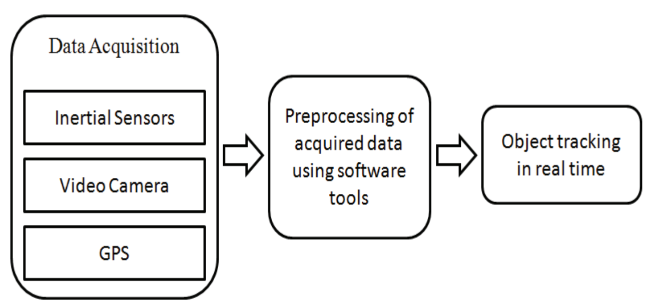

Figure 2: Real-time object tracking

G. Schall, D. Wagner, G. Reitmayr, E. Taichmann, M. Wieser, D. Schmalstieg and B. Hofmann-Wellenhof [9] present a multi-sensor fusion system, which is used in the Global Pose Estimation for Outdoor Augmented Reality Applications. J. Ventura, G. Reitmayr gave a technique in [13], which is combination of a global localization method and a key-frame based monocular SLAM system. The SLAM (Simultaneous Localization And Mapping) system works on a camera-equipped mobile client and is able to provide continuous, relative 6DoF pose estimation. It also provides key frame images with computed camera locations. The global registration correction is returned to the mobile client and localization result is updated whenever a key frame is added. This system provides 6DoF tracking and mapping for a mobile device and overcomes the localization difficulties with a narrow field-of-view mobile-phone camera. This system performs equally well for Indoor and Outdoor AR Applications. A novel marker-less approach for camera tracking is described by T. Lee, T. Hollerer in [14] which is used for augmented reality (AR) on unprepared table-top environments. Here, real-time system architecture combines two types of feature tracking. To achieve real-time performance, various operations such as capturing of video frame, tracking features (using optical flow), detecting distinctive invariant features, and rendering an output frame are processed in a synchronized multi-threaded manner. A Marker-less Hybrid Tracking method is represented by J. P. Lima, V. Teichrieb, J. Kelner and Robert W. Lindeman in [15], which is based on natural feature tracking of 3-Dimensional objects and works for the development of various standalone augmented reality applications for handheld devices. K. Kumar, A. Varghese, P. K. Reddy, Narendra N., P. Swamy, M. G. Chandra, P. Balamurlidhar [16] presents a hybrid approach for improved tracking in an AR system. This method uses Inertial Measurement Unit (IMU) fused with Vision based tracking and provides accurate results for tracking an object movement with or without occlusion. L. Chai, W. A Hoff and T. Vincent [18], presents a hybrid method using two vision cameras and head mounted inertial sensors. The technique thus presented is suitable for tracking in wide/small area unprepared environments. Table 3 provides analysis of various sensor, vision and hybrid tracking methods for Augmented Reality Applications presented mostly during last 5 years.

TABLE 3: ANALYSIS OF TRACKING METHODS FOR AR APPLICATIONS

Tracking Technique	***Objective***	***Findings from Conclusion***
Sensor Based Methods [6], [7], [8]	To analyse the accuracy of Head-to-Object position and orientation using head mounted displays.	• Based on fusion of sensor data from fixed and head mounted sensors. • Fixed optical tracking sensor combined with head mounted optical see through device and CCD camera. • More accurate pose estimation with sensor fusion.
	Automatic selection of a suitable fusion algorithm for Ubiquitous Tracking	• Based on Spatial Relationship Graphs (SRGs) and Patterns • Gyroscope fusion with three other sensors in dynamically changing tracking situations • Increased Robustness
	Sensor fusion approach to exploit visual cues and the smartphone’s inertial sensors	• Robust to appearance variations and local features tracking • Computationally Complex • Applicable in Outdoor Environment for Mobile AR
Vision Based Methods [10], [11], [12]	Tracking in Large Scale Areas using Global Positioning System and Image Processing	• Robust tracking to occlusions and dynamics in the scenes • Soft handoff to avoid frequent swaps in memory resource • Outdoor Tracking for precisely finding the specified location in a Large Area
	To provides online tracking of outdoor illumination variations	• Works for the videos captured with moving cameras • Full image-based approach for tracking • Visually coherent • Outdoor Tracking
	Deploying stationary infrastructural hardware in indoor environments for people tracking	• Fusion of anonymous position information and personal orientation information • Ubiquitous Augmented Reality • Indoor Navigation

Hybrid Methods [9], [13], [14], [15], [16], [18]	Fusion of Tracking Technologies for registration of three dimensional overlays on the real environment	• Increased Accuracy, Robustness and range for coverage area of tracking • Multi-sensor integration of inertial and vision tracking technologies • Applicable to Indoor and Outdoor Environments
	To overcome the problem of localization with a narrow field-of-view mobile-phone camera	• Keyframe-based monocular SLAM system and a global localization • Provides real-time tracking on a mobile device
	Hybrid tracking method using the multi-threaded framework for real-time AR	• Marker-less camera tracking for unprepared tabletop environments • robust against drift errors • Swift recovery from tracking failures Frame-to-frame tracking by computing optical flow
	Standalone augmented reality for handheld devices	• Natural feature tracking of fully 3-Dimensional objects • Edge-based tracking for Pocket PC • Moderate tracking accuracy • Marker-less algorithms
	Fusion of IMU and Vision approaches to improve the object tracking	• Markerless AR application • Combines the fast rotational response from IMU and accurate pose estimation of Vision techniqes. • Robust and appropriate for handling occlusion failures
	To track the position, orientation and motion of user head alongwith the 3-D structure of the scene simultaneously	• Fusion of data from head mounted cameras and inertial sensors • Suitable for unprepared environments • Uses two cameras for vision tracking

IV. CONCLUSION

For any augmented reality application, the most important requirement is to track a camera pose accurately including the position and orientation of the object of interest. Due to the refined methodology, improvements are taking place in Augmenting the Reality over last 5 years. It is clear from the study that vision based tracking methods are robust, accurate and have low jitter and drift problems. These methods are analogous to closed loop systems in which dynamic data update is present for error correction. But such systems are slow as data recovery process is time consuming. Moreover, marker-based vision methods are not suitable for large scale navigation, as information is augmented only when marker is in sight. Sensor based methods are fast and robust and can be used for motion prediction when changes occur rapidly. But these techniques suffer from drift problems due to noise accumulation. The combination of various sensors is possible with vision methods to make hybrid sensor tracking systems. Each tracking system has its limitations, and hybrid systems provide a relatively good solution by combining the advantageous aspects of different technologies. Although, these hybrid methods increase the complexity and the cost of a system yet these can readily improve the efficiency of an Augmented Reality System.

V. PROPOSED WORK

As a result of analyzing the available methods of tracking, their applicability and advantages for particular areas, it is proposed to develop a hybrid system for tracking which uses combination of sensor and vision methods. For future work, the emphasis will be given on the development of a fully mobile tracking Augmented Reality system using Hybrid sensing, which is certainly the right concept for wide area coverage in unprepared environment. This system may prove useful for AR applications in education and Training. The combination of inertial tracking and vision based inside-out tracking is especially suited for this application.

REFERENCES

[1] F. Zhou, Henry B. Duh, M. Billinghurst, “Trends in Augmented Reality Tracking, Interaction and Display: A Review of Ten Years of ISMAR”, International Symposium on Mixed and Augmented Reality (IEEE, ISMAR – 2008), pp 193-202.

[2] R. Azuma, “A Survey of Augmented Reality”, In Presence: Teleoperators and Virtual Environments-6, August 1997, pp 355-385.

[3] L. Johnson, A. Levine, R. Smith, S. Stone, “Simple Augmented Reality”, The 2010 Horizon Report, 2010, pp 21-24.

[4] P. Milgram, F. Kishino, "A Taxonomy of Mixed Reality Visual Displays,” IEICE Transactions on Information Systems, Vol. E77D, No.12, 1994.

[5] P. Liu, N. D. Georganas, P. Boulanger, “Designing Real-Time Vision Based Augmented Reality Environments for 3D

Collaborative Applications", IEEE Canadian Conference on Electrical and Computer Engineering, 2002, pp 1-6.

[6] W. Hoff, T. Vincent, "Analysis of Head Pose Accuracy in Augmented Reality", IEEE Transactions on Visualization and Computer Graphics, Vol. 6, No. 4, 2000, pp 319-334.

[7] D. Pustka, G. Klinker, "Dynamic gyroscope fusion in Ubiquitous Tracking environments", IEEE International Symposium on Mixed and Augmented Reality (ISMAR) - 2008, pp 13-20.

[8] L. Porzi, "PhD Forum: Sensor Fusion for Outdoors Augmented Reality on Android", ICDSC '14, November 04 - 07 2014, Venezia Mestre, Italy.

[9] G. Schall, D. Wagner, G. Reitmayr, E. Taichmann, M. Wieser, D. Schmalstieg and B. Hofmann-Wellenhof, "Global Pose Estimation using Multi-Sensor Fusion for Outdoor Augmented Reality", IEEE International Symposium on Mixed and Augmented Reality, 2009, pp 153-162.

[10] W. Guan, S. You, U. Neumann, "GPS-aided recognition-based user tracking system with augmented reality in extreme large-scale areas", MMSys '11 Proceedings of the second annual ACM conference on Multimedia systems, pp 1-10.

[11] Y. Liu and X. Granier, "Online Tracking of Outdoor Lighting Variations for Augmented Reality with Moving Cameras", IEEE Transactions on Visualization and Computer Graphics, (Volume:18 , Issue: 4), pp 573-580.

[12] A. L. Christian Waechter, D. Pustka, G. J. Klinker, "Vision based people tracking for ubiquitous Augmented Reality applications", International Symposium on Mixed and Augmented Reality, 2009, pp 221 – 222.

[13] J. Ventura, G. Reitmayr, "Global localization from monocular SLAM on a mobile phone", IEEE Transactions on Visualization and Computer Graphics 2014.Vol. 20, No. 4, pp 531-539.

[14] T. Lee, T. Hollerer, "Multithreaded Hybrid Feature Tracking for Marker-less Augmented Reality", IEEE Transactions on Visualization and Computer Graphics 2009, .Vol. 15, No. 3, pp 355-368.

[15] J. P. Lima, V. Teichrieb, J. Kelner, Robert W. Lindeman, "Standalone Edge-Based Markerless Tracking of Fully 3-Dimensional Objects for Handheld Augmented Reality", VRST 2009, pp 139-142.

[16] K. Kumar, A. Varghese, P. K. Reddy, Narendra N., P. Swamy, M. G. Chandra, P. Balamurlidhar, "An improved tracking using IMU and Vision fusion for Mobile Augmented Reality applications", The International Journal of Multimedia & Its Applications (IJMA) Vol.6, No.5, October 2014, pp 13-29.

[17] H. Uchiyama, E. Marchand, "Object Detection and Pose Tracking for Augmented Reality: Recent Approaches", FCV 2012, pp 1-8.

[18] L. Chai, W. A Hoff, T. Vincent, "3-D Motion and Structure Estimation Using Inertial Sensors and Computer Vision for Augmented Reality", submitted to Presence: Teleoperators and Virtual Environments.

[19] D. P. Kaur, A. Mantri, "Comparative Analysis of indoor and outdoor Tracking Techniques for Augmented Reality Applications", published in the proceedings of 4th International Conference on Wireless Networks and Embedded Systems (WECON), 2015.

[20] D. Wagner, T. Langlotz, D. Schmalstieg, "Robust and unobtrusive marker tracking on mobile phones", IEEE International Symposium on Mixed and Augmented Reality (ISMAR) 2008, pp 121-124.

[21] D.W.F. van Krevelen, R. Poelman, "A Survey of Augmented Reality Technologies, Applications and Limitations", The International Journal of Virtual Reality, 2010, Vol. 9, No. 2, pp 1-20.

[22] Wendy E. Mackay, "Augmented Reality: Linking real and virtual worlds, A new paradigm for interacting with computers", in proceedings of AVI'98, ACM Conference on Advanced Visual Interfaces.

[23] J. Lobo, J. Dias, "Vision and Inertial sensor cooperation using Gravity as a Vertical Reference", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 25, No. 12, December 2003, pp 1597-1608.