Vision-based Tracking Technology for Augmented Reality: A Survey

Fadhil Noer Afif, Ahmad Hoirul Basori, and Nadzari Saari

Abstract—This paper reviewed some development of tracking techniques used for Augmented Reality. It exposes importance of tracking for building a well AR system. This also presents general principles and some methods that has been used by researchers for developing their techniques from earlier development until recently.

Index Terms—reality, tracking, tracking evaluation

1 INTRODUCTION

HUMAN has an ability to expand their intelligence in a process called imagination. By imagining in their mind, people can mentally create and manipulate things that cannot be described by hearing, scent or other senses. Children frequently use their imagination to create worlds and objects that exists only in their mind, such as when they do storytelling or role-playing. In a learning process, imagination will help improving human understanding of a particular knowledge. This concept of creating and manipulating imaginary objects is realized with a technology called Augmented Reality.

Augmented Reality (AR) attempts to combine computer-generated objects to a real-world space. The technology works by superimposing virtual objects onto a scene of real world captured by a video camera. User can see these objects with a display devices, resulting as if the virtual and real objects are lying together on the same space. AR can display information that cannot be directly perceived by sight or hearing, thus extends human senses. [1] describes AR as a system that combines real and virtual objects, positions these objects seamlessly with each other and interactive in real time. Survey about Augmented Reality principles and concepts can be found on the same paper [1] and the complement in [2].

By augmenting virtual objects onto real world to display information, AR can be used to improve user performance of real-world activities. As such, there are many fields that can be explored with AR. In the medical domain, AR can be used to guide the doctor by displaying virtual imagery of patient's inner organs that cannot be seen with bare eyes. AR can also applied for virtually displaying information of objects in a museum, displaying interactive content on advertising, or guiding. Any other domains can also get benefits from AR such as military, engineering or entertainment.

- *Fadhil Noer Afif is with Faculty of Computing, Universiti Teknologi Malaysia, Skudai, Malaysia. E-mail: fadhilnoerafif@gmail.com*
- *Ahmad Hoirul Basori is with Faculty of Computing, Universiti Teknologi Malaysia, Skudai, Malaysia. E-mail: ahmadhoirul@utm.my*
- *Nadzari Saari is with Faculty of Computing, Universiti Teknologi Malaysia, Skudai, Malaysia. E-mail:*

Augmented Reality requires all of augmented objects to appear together seamlessly in with the real world. When the user moves around and changes the view, ideally these virtual objects remain in their designated place, as if these objects are completely overlaid in space. This condition demands computer to be able to track changes of surrounding environment. Thus, tracking techniques, in purpose for registering virtual world into real world, is essential for developing a decent Augmented Reality system [3–5].

Attempts for registering real-world space with virtual imageries and objects have been done in many researches. A recent comprehensive review by [6] exposed trends of research areas in Augmented Reality. According to this survey, tracking techniques are one of the fundamental topics for AR system development, and is the most popular sub-fields to be explored.

This paper reviews some development of techniques that has been proposed by researchers, mainly focusing on vision-based techniques. The methods are organized by their general principles. Strengths and weaknesses for these principles also will be explored. Current trends, as well as future possibilities will be discussed and concluded by the end of this survey.

2 PROCEDURE FOR PAPER SUBMISSION

An early tracking technology utilizes many devices such as transducers or sensors to help the tracking process. These instruments, such as ultrasonic sensors, infra-reds, GPS, gyroscopes and so on give informations to the processing unit over-time by measuring coordinates and orientation of real environment. Despite their capability to provide sufficient information for object registration, these devices are limited in terms of specifications, expensive, and practically not portable. See survey by [7] for some methods of tracking by using them, namely sensor-based techniques.

Different approach is taken by vision-based tracking, which unlike sensor-based, exploits functionalities of camera's vision. Vision-based tracking uses image processing techniques to develop method for calculate position of the camera. Researches so far categorized vision based tracking into two; namely marker-based and markerless tracking techniques.

Marker-based tracking are techniques that tells the computer to track predefined fiducial markers that placed in arbitrary locations. One of the earlier technique was proposed by [8] for object identification and registration using matrix fiducials. In order to obtain reliable inputs, sequence of images captured by camera were binarized onto black-white pixels by using adaptive thresholding method. By analyzing connected component and fitting of code frame returned by matrix, information of specific markers were then gained and projected. This information were then used to estimate the camera position, thus track the markers relative to the camera. The proposed technique showed ability to identify and track 2^{16} different objects identified by matrix barcodes.

A line-based tracking technique by [4] proposed a method to recognize and track a specific marker that has linear features. A dark rectangular landmark that contains rows of red dots was used as a marker. Recognition of the linear side (and corners) was done by extracting contour samples and evaluating them statistically with the predefined criterion. Each of square markers has a different formation of red dots. By comparing these formations to the database, each marker can be distinguished and augmented uniquely. The algorithm has been evaluated in two different machine and accomplished performance of about 23 Hz.

Another marker-based tracking technique is the well-known ARToolKit by [9] that proposed a square-shaped fiducial markers for tracking, together with new AR interface. ARToolKit utilized square markers with known size and pattern to be regarded as "Virtual monitors", so that camera coordinates are projected onto them. By thresholding input and template matching, the markers were detected and then augmented using the interface provided. As reported in the result, ARToolKit were reliable to provide accurate, robust and real-time Augmented Reality experience. However, tracking errors occur when fiducial markers were placed too far. The method also recognize linearly with number of markers, thus performance will drop when numerous markers are being used.

In addition to rectangular shape, other types of markers were also utilized by some researchers, such as a ring fiducials tracking techniques developed by [3]. The method attemped to recognize concentric multicolor fiducials with different sizes allowing scalability of tracked markers. Scanline segmentation along with segment and line finding was used to detect ellipse projection of captured circular markers, and the multi-ring shape of the marker increases the robustness of this technique. The method managed to track predefined fiducials with minimum width of five pixels.

In research done by [10], a circular bar-coded fiducials were used as the markers for the AR tracking based on homomophics, focused more on tracking many fiducials in wide area environments. Bar-coded fiducal design allowed wide choice for tracking different locations. Shapes of fiducial used is shown in Fig. 1.

Fig. 1. Circular bar-coded markers. Taken from [10]

Marker-based tracking techniques "drives" computers to track specific, predefined fiducials on scenes, thus is beneficial in terms of reducing computational costs. See [11] for comparison between several marker-based techniques. However, dependency of markers reduces seamless interaction, because tracking can only be done when they are visible and close enough. Also, some environments such as large buildings, workspaces are not fully appropriate with use of markers. Such limitations become drawbacks of these techniques.

Hence, markerless tracking techniques offer solution for aforementioned problem. The technique aims to search for natural elements on the real-world, and registers onto them, therefore more appropriate to definition of Augmented Reality stated by Azuma. One such technique was done by [12] for tracking edge and texture of object captured by camera. This method attempts to extract both interest points and contour points and utilized them as complementary source to feed information to camera. By combining information, the methods were able to track and augment objects in a corridor.

Some other researchers utilized known model in order to improve robustness and stability of tracking, namely model-based markerless tracking. These models are trained into the AR system. With feature-matching techniques, models captured on the screen frame will be recognized and virtual objects can be superimposed into them. A research by [13] proposed one such method for augmenting virtual object for e-commerce applications using two-dimensional images as the model. Divided into two steps, the method firstly extracted image features by applying scale invariant feature extraction to extract robust image features. Then, in order to recognize and track this feature on video sequence captured by the camera, the author utilized chaotic neural network as feature-matcher. According to the experimental result, the method was able to track and recognize pretrained image features with match probability of 90.3%.

In 2010, [14] proposed a method for tracking and augmenting printed photos. This technique trained the system to recognize any pretrained photos based on tracks of features by using bi-layer clustering process. During the training process, features of each image were collected and organized to form a visual vocabulary which would be used as basis for tracking. The method were able to augment hundreds of pretrained pictures simultaneously in real-time, and also detect photos with low illumination and partial occlusion.

In addition to planar images, point cloud models of a structure can also be used as models. A two-stage algorithm by [15] constructed a point cloud database of desired target by using subtrack optimization algorithm. Captured video sequences were then compared with these database for tracking camera pose and augmenting the virtual content. Point cloud models were also used to reconstruct three dimensional scene by using Kinect sensor [16].

Although tracking performance can be increased by giving knowledge about scene by training, model-based methods are limited to track well-informed scene only. When these models are not abundant or not available at all, significant drop in quality is likely to occur [6]. Feature-based tracking is another approach performed by adding unknown scene to computer's knowledge as more objects appear in scene. Parallel Tracking and Mapping (PTAM) by [17] proposed a scheme to develop model and track previously unknown scene in small environment such as on top of a desk. In order to approach such target, PTAM separates its tracking and mapping into two parallel threads, so that the process can run simultaneously. In mapping process, the method applied bundle adjusment technique to add information based on corner features of captured scene. When the camera sees new objects, more features will be added, expanding current map as it goes by. Tracking threads then will calculate camera's pose and location based on information obtained from the map. Results showed that PTAM were able to track an example scene of 4997 point features, with promising processing speed of about 20 ms. Other techniques also used feature-based tracking such as [18–24]

3 Conclusion

In this paper some tracking technologies for Augmented Reality system has been explored. In order to integrate virtual object into real-world seamlessly, Augmented Reality system must own ability to recognize and track its desired environment. Challenges arised because of complexity of each environment. Sensor-based techniques are potential to provide accurate real-world measurement, but has limitations in terms of affordability and usability. On the other hand, vision-based techniques are able to overcome such limitations because they only uses camera as the device. Early development of these principles utilize fiducials as their target for tracking, able to provide fast and robust augmentation. However, the techniques rely heavily on markers, thus more recent techniques called markerless tracking tries to overcome this limitation with attempt to track natural features on the environment. In spite of breakthroughs gained in markerless tracking techniques, there are some issues encountered. As mentioned before, model-based techniques require to train models or maps beforehand, so they are less applicable to track unknown environments. Some techniques to track natural edges are dependent to scenes at a particular moment, that it will fail if edges are lost caused by e.g. motion blur. While these can be handled by adding features to maps and improving them like what is done by [17], the techniques are computationally expensive. Furthermore, they also suffer from situation when scene is lacking of abundant edges and textures. Such condition causes low tracking quality because not enough information can be added to the maps. However, markerless techniques are still challenging and very potential to be explored, in purpose to develop a seamless Augmented Reality experience.

Acknowledgment

This research is supported by the Ministry of Science and Technology (MOSTI) and collaboration with Research Management Center (RMC), Universiti Teknologi Malaysia (UTM). This paper is financially supported by E-Science Grant Vot. No:R.J130000.7911.4S026.

References

[1] R. T. Azuma, "A Survey of Augmented Reality," *Presence : Teleoperators and Virtual Environments*, vol. 6, no. 4, pp. 355–385, 1997.

[2] R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier, and B. MacIntyre, "Recent Advances in Augmented Reality," *Computer Graphics And Applications*, vol. 21, no. 6, pp. 34–47, 2001.

[3] Y. Cho and U. Neumann, "Multiring Fiducial Systems for Scalable Fiducal-Tracking Augmented Reality," *Presence*, vol. 10, no. 6, pp. 599–612, 2001.

[4] D. Stricker, G. Klinker, and D. Reiners, "A Fast and Robust Line-based Optical Tracker for Augmented Reality Applications," in *Proceedings of the 1st International Workshop on Augmented Reality (IWAR 1998)*, 1998, pp. 31–46.

[5] W. A. Hoff, K. Nguyen, and T. Lyon, "Computer vision-based registration techniques for augmented reality," in *Proceedings of Intelligent Robots and Computer Vision XV,* 1996, vol. 2904, pp. 538–548.

[6] F. Zhou, H. B.-L. Duh, and M. Billinghurst, "Trends in Augmented Reality Tracking, Interaction and Display: A Review of Ten Years of ISMAR," in *Proceedings of the 7th International Symposium of Mixed and Augmented Reality (ISMAR 2008)* : ACM & IEEE, 2008, pp. 193–202.

[7] J. P. Rolland, Y. Baillot, and A. A. Goon, "A Survey of Tracking Technology for Virtual Environments," Fundamentals of Wearable Computers and Augmented Reality, vol. 1, pp. 67–112, 2001.

[8] J. Rekimoto, "Matrix: A Realtime Object Identication and Registration Method for Augmented Reality," in *Proceedings of Asia Pacific Computer Human Interaction 1998 (APCHI'98),* 1998, pp. 63–68.

[9] H. Kato and M. Billinghurst, "Marker tracking and HMD calibration for a video-based augmented reality conferencing system," in *Proceedings 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR'99)*, 1999, pp. 85–94.

[10] L. Naimark and E. Foxlin, "Circular data matrix fiducial system and robust image processing for a wearable vision-inertial self-tracker," in *Proceedings of the 1st International Symposium on Mixed and Augmented Reality*, 2002, pp. 27–36.

[11] X. Zhang, S. Fronz, and N. Nassir, "Visual marker detection and decoding in AR systems: A comparative study," in *Proceedings of the 1st International Symposium on Mixed and Augmented Reality (ISMAR 02)*, 2002, pp. 97–106.

[12] L. Vacchetti, V. Lepetit, and P. Fua, "Combining Edge and Texture Information for Real-Time Accurate 3D Camera Tracking," in *Proceedings of the 3rd International Symposium of Mixed and Augmented Reality (ISMAR 04)*, 2004, no. Ismar, pp. 48–57.

[13] X. Li and D. Chen, "Augmented reality in e-commerce with markerless tracking," in *2010 2nd IEEE International Conference on Information Management and Engineering,* 2010, pp. 609–613.

[14] J. Pilet and H. Saito, "Virtually augmenting hundreds of real pictures: An approach based on learning, retrieval, and tracking," in *Virtual Reality Conference*, 2010, pp. 71–78.

[15] J. Mooser, S. You, U. Neumann, and Q. Wang, "Applying robust structure from motion to markerless augmented reality," in *2009 Workshop on Applications of Computer Vision (WACV)*, 2009, pp. 1–8.

[16] R. A. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A. J. Davison, P. Kohli, J. Shotton, S. Hodges, and A. Fitzgibbon, "KinectFusion : Real-Time Dense Surface Mapping and Tracking ∗ ," in *Proceedings of the International Symposium of Mixed and Augmented Reality 2011 (ISMAR 2011) : ACM & IEEE*, 2011, pp. 127–136.

[17] G. Klein and D. Murray, "Parallel Tracking and Mapping for Small AR Workspaces," in *Proceedings of the 6th International Symposium on Mixed and Augmented Reality (ISMAR 2007) : IEEE & ACM*, 2007, pp. 225–234.

[18] S.-H. Huang, Y.-I. Yang, and C.-H. Chu, "Human-centric design personalization of 3D glasses frame in markerless augmented reality," *Advanced Engineering Informatics*, vol. 26, no. 1, pp. 35–45, Jan. 2012.

[19] F. Huang, Y. Zhou, Y. Yu, Z. Wang, and S. Du, "Piano AR: A Markerless Augmented Reality Based Piano Teaching System," in *2011 Third International Conference on Intelligent Human-Machine Systems and Cybernetics*, 2011, pp. 47–52.

[20] G. R. Rodrıguez-Canosa, S. Thomas, J. del Cerro, A. Barrientos, and B. MacDonald, "A Real-Time Method to Detect and Track Moving Objects (DATMO) from Unmanned Aerial Vehicles (UAVs) Using a Single Camera," *Remote Sensing,* vol. 4, pp. 1090–1111, 2012.

[21] K. Kim, V. Lepetit, and W. Woo, "Scalable real-time planar targets tracking for digilog books," *The Visual Computer*, vol. 26, no. 6–8, pp. 1145–1154, Apr. 2010.

[22] T. Lee and T. Höllerer, "Multithreaded hybrid feature tracking for markerless augmented reality.," *IEEE transactions on visualization and computer graphics*, vol. 15, no. 3, pp. 355–68, 2009.

[23] J. Platonov and M. Langer, "Automatic contour model creation out of polygonal CAD models for markerless Augmented Reality," in *2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality,* 2007, pp. 1–4.

[24] Y. Uematsu and H. Saito, "Multiple planes based registration using 3D Projective Space for Augmented Reality," *Image and Vision Computing*, vol. 27, no. 10, pp. 1484–1496, Sep. 2009.

Fadhil Noer Afif was born in Bandung, Indonesia in 1990. He received the B.Sc in Physics from University of Padjadjaran in 2011. Currently he is studying M.Sc in Computer Science in Universiti Teknologi Malaysia, Johor Bahru, Johor. His research interests including Computational Physiscs, Augmented Reality and Mixed Reality.

Ahmad Hoirul Basori was born in Wirowongso Village, jember City, in 1982. He received the B.Sc(Software Engineering), in 2004 and the Ph.D. degree in Computer Graphics from Universiti Teknologi Malaysia, Johor Bahru, Johor, in 2011.From 2004 to present, he was a lecturer with the Department of Informatics, Faculty of Information Teknologi, Institut Teknologi Sepuluh Nopember Surabaya, Indonesia. Since 2011, he has been appointed as Assistant Professor with the Department of Computer Graphics and Multimedia, Universiti Teknologi Malaysia. He is the author of Editorial board of some international journal, more than 32 articles, and also a member of professional membership IEEE, ACM SIGGRAPH, IAENG and Senior Member of IACSIT. His research interests include Computer Graphics, Facial Animation, Cloth Simulation, Medical Visualization, Haptic Interaction and Man Machine Interaction.