2015 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)

Vision-based Registration for Augmented Reality-A short Survey

Akilan Thangarajah, Jonathan Wu
Electrical and Computer Engineering,
University of Windsor,
Windsor, Canada

Bakri Madon, Adib Kabir Chowdhury
School of Computing
University College of Technology Sarawak,
Sarawak, Malaysia.

***Abstract*— The purpose of this paper is to explore some existing techniques in vision-based registration for Augmented Reality (AR) and present them collectively. AR is a branch of computer vision which generally overlays Virtual Objects (VOs) on actual images of real-world scenes in order to provide additional information about the scene to the user. Due to its wide range of applications in the fields of medical, robotics and automotive, geographic and remote sensing, military and aerospace, it has gained high demand. In any AR system, registration is the key to make the augmented scene appearing natural. Registration process must avoid occlusion of VOs and objects in the real world and align the VOs precisely. Optics-based and video-based are two well-known industrial AR systems. Researchers show that even with a single camera model registration for an AR is plausible but, VOs may be registered in front of real-world objects. It is because the registration process lacks depth information of the scene. However, employing stereo vision system and utilizing available natural features in a real-world scene and set of arbitrary multiple planes one can improve accuracy of VO registration. Thus, an AR system becomes robust if it is devised with algorithms to extract and track natural features in real-time.**

***Keywords*— augmented reality, stereo-vision, 2D, 3D virtual object, object registration**

I. INTRODUCTION

Augmented Reality (AR) is a branch in computer vision which provides additional information about a scene in the real-world by overlying Virtual Objects (VOs) on actual images of scene. It has gain high demand in recent years due to its multivariate applications in the fields of medical, robotics and automotive, geographic and remote sensing, military and aerospace. Optics-based and video-based are the two well-utilized registration techniques in AR systems. In optics-based AR system, an optical combiner reflects a projected virtual scene to the user's display while it allows users to see the real-world scene directly. The conceptual diagram in Fig.1 describes this type of systems. On the other hand, video-based AR systems use video cameras and video composition arrangement to capture the real-world scenes and to combine the virtual object before the augmented scene is streamed to user. The conceptual diagram in Fig.2 describes this type of systems. In AR, registration plays a vital role in making the augmented scene appearing natural. Registration process is to align the VOs properly with the objects in the concerned real-world and solving the occlusion conflict between real objects and VOs.

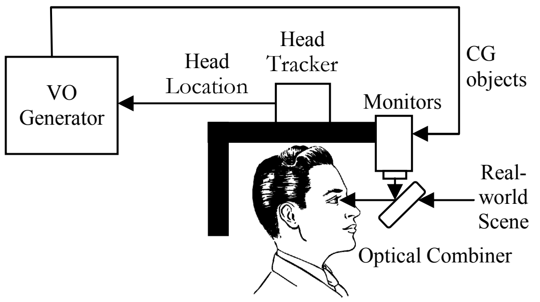

Fig.1. Optics-based AR [1].

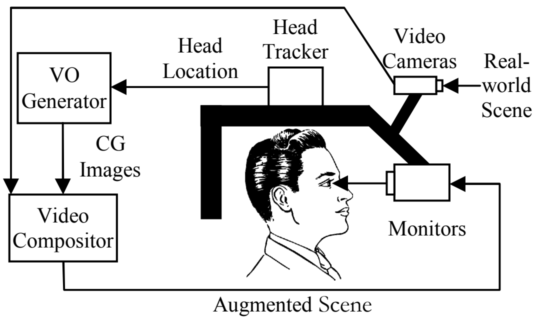

Fig.2. Video-based AR [1].

The registration process is classically done in a tripler-step: positioning, rendering, and merging. Positioning locates VOs at desired positions in the scene, while rendering creates 2D images of the VOs, then merging superimposes the rendered VOs on image of the real-world scene. First of all, AR registration process needs the exact coordinate of the location where the VOs to be placed. Getting this piece of information is not readily available since it requires metric 3D information of the scene, high accuracy measurement, and projection matrix of the real camera.

Thus, natural appearance of an augmented image is mainly based on how accurately the VOs are registered in the image of a scene with respect to the position and orientation of user's eye in real-time. There are many researches who explored different techniques to address this. Some methods use external sensing devices such as magnetic or ultrasonic sensors attached to the vision system to get the position and orientation of the users' eye.

The rest of this paper is organized as follows. Section II provides bench mark methods of 3D registration for augmented reality, collectively. Section III concludes the review and Section IV provides our future direcions.

II. REVIEW

This section summarizes the vision based registration fundamental methods presented in prominent literatures. In general the methods can be categorized as: registration with artificial markers, registration with natural markers, multi-camera model based registration, hybrid registration, and registration through blur estimation.

A. Registration with artificial markers

Okuma et al. [2] propose a vision-based registration algorithm along a prototype which requires known markers in the scene. The prototype captures images via a Sirius video capturing system and displays the augmented scene in a Sony Glasstron Head Mounted Display (HMD). At first the system takes four known points from the image of a scene and set four separate tracking windows around the points. Based on these points it calculates camera calibration matrix (M). If the system is able to track only three points out of the fours, it utilizes Finsterwalder's method to calculate M, as in

$$\mathrm{M} = \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (1)$$

Then world, camera, and screen coordinates of the four points are assumed to be $W_i = (x_{wi,}\ y_{wi,}\ 1)^t$; i = 1, 2, 3, 4, $C_i = (x_{ci,}\ y_{ci,}\ 1)^t$ and $S_i = (x_{si,}\ y_{si})^t$ respectively, which holds the matrix expression,

$$\begin{pmatrix} X_{ci} \\ Y_{ci} \\ Z_{ci} \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} & m_{14} \\ m_{21} & m_{22} & m_{24} \\ m_{31} & m_{32} & m_{34} \end{pmatrix} \begin{pmatrix} x_{wi} \\ y_{wi} \\ 1 \end{pmatrix}. \quad (2)$$

Similarly, the screen coordinate of the four points are assumed to be $S_i = (x_{si,}\ y_{si})^t = (x_{ci}/z_{ci},\ y_{ci}/z_{ci})^t$. Then a projection matrix is calculated based on the coordinates found earlier, sequentially the generated virtual objects are overlaid on the input image and the final augmented image displayed in HMD. Although, this approach projects the virtual object at a desired location and orientation, there is registration error in terms of rendering delay. It is reported that capturing, synthesizing, and displaying output takes one frame time for each.

Since, there is no depth information of the scene available in this system, the VOs are always overlaid the real-world objects/scene even if the real-world objects are in front of the VOs. So the authors believe that employing a stereo vision system to get the depth information would solve this issue.

B. Registration with natural markers

Using artificial markers subside robustness of the registration process. Therefore, a 3D registration without using artificial markers or external sensors to assist the registration process is proposed by Uematsu and Hideo [3]. The authors exploit naturally available planes, for instance, indoor floors and walls or outdoor surfaces of buildings i.e. a set of arbitrary multiple planes and their geometrical relationships to estimate the camera motion frame by frame. The authors assign separate 3D coordinate system for each plane by setting Z = 0. Then by using homography between 3D real-plane and the input image plane intrinsic and extrinsic parameters of the camera is calculated. Once these parameters are found the projection matrix can be calculated. The KLT-feature-tracker [4] is then employed to track the natural feature points of the 3D planes so that homographies and projection matrix can be calculated for each of the chosen arbitrary 3D planes. The authors find planner homography $\widehat{P}$ of the input images by deleting the third column of each projection matrices (i.e. Z = 0), as in

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \cong P \begin{pmatrix} X \\ Y \\ 0 \\ 1 \end{pmatrix} \cong \begin{bmatrix} p_{11} & p_{12} & p_{14} \\ p_{21} & p_{22} & p_{24} \\ p_{31} & p_{32} & p_{34} \end{bmatrix} \begin{pmatrix} X \\ Y \\ 1 \end{pmatrix} \cong \widehat{P} \begin{pmatrix} X \\ Y \\ 1 \end{pmatrix} \cong H \begin{pmatrix} X \\ Y \\ 1 \end{pmatrix}. \quad (3)$$

From the above calculations, the projection matrix is recalculated by using relationships, as in (4) and (5).

$$P = A[R \mid t] = A[r_1 r_2\ r_3\ t]. \quad (4)$$

$$\widehat{P} = A[r_1 r_2\ t] = H \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}. \quad (5)$$

By setting skew to 0 and aspect ratio to 1, center of the image, intrinsic parameters, and focal length are calculated, as in

$$\mathrm{A} = H \begin{bmatrix} f & 0 & c_x \\ 0 & f & c_y \\ 0 & 0 & 1 \end{bmatrix}, \quad (6)$$

$$\mathrm{A}^{-1}\mathrm{H} = [r_1 r_2\ t]. \quad (7)$$

Where, (C_x, C_y) is the center point and f is the focal length. Then the rotation matrix R is calculated while steepest descent method and homography are used, respectively, to improve accuracy and to optimize errors between initial point x_p and projected point x_h, as in

$$f^2 = \frac{(h_{11}\text{-}c_x h_{31})(h_{12} - c_x h_{32}) + (h_{21}\text{-}c_y h_{31})(h_{22}\text{-}c_x h_{32})}{\text{-}h_{31} h_{32}}, \quad (8)$$

$$\varepsilon = (x_h\text{-}x_p). \quad (9)$$

Since each projection matrix exclusively belongs to its corresponding plane, the accuracy will become lower if the VOs moves away from each plane. To overcome this issue the authors integrate the projection matrices to compute a matrix which uses information of multiple planes. The rich information: various poses and positions of the multiple planes allow a precise augmentation than using a single plane. The authors propose two methods for integrating the matrices: maximum likelihood estimation and merging with weights. The maximum likelihood estimation based integration is defined as in

$$\begin{pmatrix} X_{11} & Y_{11} & Z_{11} & 1 & 0 & 0 & 0 & 0 & -X_{11}x_{11} & -Y_{11}x_{11} & -Z_{11}x_{11} \\ 0 & 0 & 0 & 0 & 1 & X_{11} & Y_{11} & Z_{11} & -X_{11}y_{11} & -Y_{11}y_{11} & -Z_{11}y_{11} \\ & & & \vdots & & & & & & & \\ X_{\mathrm{nm}} & Y_{\mathrm{nm}} & Z_{\mathrm{nm}} & 1 & 0 & 0 & 0 & 0 & -X_{\mathrm{nm}}x_{\mathrm{nm}} & -Y_{\mathrm{nm}}x_{\mathrm{nm}} & -Z_{\mathrm{nm}}x_{\mathrm{nm}} \\ 0 & 0 & 0 & 0 & 1 & X_{\mathrm{nm}} & Y_{\mathrm{nm}} & Z_{\mathrm{nm}} & -X_{\mathrm{nm}}y_{\mathrm{nm}} & -Y_{\mathrm{nm}}y_{\mathrm{nm}} & -Z_{\mathrm{nm}}y_{\mathrm{nm}} \end{pmatrix} \begin{pmatrix} t_{11}^{PI} \\ t_{12}^{PI} \\ \vdots \\ t_{33}^{PI} \\ t_{34}^{PI} \end{pmatrix} = \begin{pmatrix} x_{11} \\ y_{11} \\ \vdots \\ x_{mn} \\ y_{mn} \end{pmatrix}. \quad (10)$$

Where, n is the number of planes exists, and m is set of corresponding points in every plane which is calculated by

$$T_k^{PI} = P_k\left(T_k^{WP}\right)^{-1}, \quad (11)$$

and merging with weights does the integration using (12).

$$T^{PI} = \frac{1}{n}\left[w_1, \ldots, w_n\right]\left[T_1^{PI}, \ldots, T_N^{PI}\right]^T. \quad (12)$$

Where, w_k is a weight parameter. Although their method uses un-calibrated camera, it produces adequate registration accuracy comparing to Simon et al. [5].

C. Multi-camera model based regisration

To achieve better result, William et al. [6] take advantage of using multiple CCD camera views and predefined markers. Their system acquire and track multiple objects starting from a wide variety of initial poses and extract feature points, position, and orientation (pose) of the scene for forming projection matrices followed by registration process. The authors use passive fiducial markings called "Concentric Contrasting Circle (CCC)" for landmark target points in the scene in a distinctive geometric pattern to simplify feature extraction and correspondence process. An example of such CCC placement is shown in Fig. 3 where the arrow overlays displayed in the HMD show the person how to pull off the PC cover. Once the image capturing is done, morphological operations is carried out to eliminate noise and to track the interested points which are invariant to changes in translation, scale, and rotation.

The authors place four CCC's in a flat rectangular pattern and fifth CCC on a side of the rectangular object to limit the rotation angle. Then, the authors use relatively simple camera model, the pinhole with an aspect ratio scaling factor and Hung-Yeh-Harwood pose estimation algorithm [7] to find the projection matrix. This algorithm is used to find the projection that generates the best relationship between the image features and their predicted locations. The main limitation of their system is that it requires at least five known points i.e. the CCC's. The authors found that if the target moves out of the field then orientation of the object cannot be calculated, thus, no registration process will be taken place. To overcome this issue, the authors propose to create more CCC's or to use internal sensors such as gyroscopes and accelerometers to get readily available head pose information.

Fig. 3 (a) Person Wearing the AR System, (b) Overlays shown in HMD [6].

Similar to the work of William et al., Kanbara et al. [8] and Yong et al. [9] also utilize the richness of stereovision system. Firstly, from a pair of images containing markers in a real-world scene, camera parameters are calculated by using stereo matching algorithm with epipolar constraint. Extraction of the known points is performed only in the first frame and tracked in the subsequent frames in order to reduce computation complexity. From the extraction and tracking information, screen coordinates of each marker is calculated. Secondly, the authors estimate depth of the scene and model matrix with the camera parameters. Finally, based on those calculations the authors superimpose desired VOs on the image of the scene. In the initial states, stereovision cameras are calibrated by using standard approach as described in [10]. To acquire 3D coordinates of the markers, the authors use the stereoscopic geometry shown in Fig. 4, where, the coordinates X, Y, and Z are defined, as in

$$X = \frac{B(x_l + x_r)}{2(x_l - x_r)}, \quad (13)$$

$$Y = \frac{B(y_l + y_r)}{2(x_l - x_r)}, \quad (14)$$

$$Z = \frac{fB}{x_l - x_r}. \quad (15)$$

Where, f and B are the known values of focal length and baseline respectively. Then to register CG objects (i.e. VOs) on an image of a real-world scene, the authors rely on $C=Mw$, which, holds the relationship between camera coordinate system – C, projection/ model-view matrix – M, and world coordinate system – w. The M, is a homogenous matrix augmented with rotation, R and translation, T. From the defined geometry in Fig. 5, T is taken as the camera coordinate of the marker m_1 then R is calculated, as in

$$\mathrm{R} = \left[x_n' \; y_n' \; z_n'\right]. \quad (16)$$

Where,

$$x_n = V_2 - V_1, \quad (17)$$

$$y_n = (V_3 - V_1) - \frac{x_n \cdot (V_3 - V_1)}{x_n \cdot x_n} x_n, \quad (18)$$

$$z_n = x_n \times y_n,\ x_n' = \frac{x_n}{\| x_n \|}, \tag{19}$$

$$y_n' = \frac{y_n}{\| y_n \|}, \tag{20}$$

$$z_n' = \frac{z_n}{\| z_n \|}. \tag{21}$$

By using the calculated matrix M, the real and virtual world can be aligned for perfection. In order to avoid occlusion i.e. conflict between real-world and virtual objects, the authors adopt bounding box registration method. Once position of VOs in world-coordinate system is determined, the M is used to transform its position in Camera-coordinate system. Then, a bounding box of the VO is projected onto the left image (refer Fig. 6) and same projection on the right image as well. Then, Sobel filter is utilized to detect edges in the region of depth estimation on the left and right images followed by sum of absolute differences (SAD) based stereo matching is performed with a window size of 5×5 pixels, while interpolation is used to determine the depth values at edges.

The authors report that the VOs are registered correctly without any occlusion conflicts in real-time. However, when they are not able to gather at least three markers, their system fails to estimate the model-view matrix, M. The authors believe that introducing: (i) more than three markers, (ii) an automatic system for feature detection and tracking of natural feature points would overcome the shortcoming. These kinds of techniques i.e. registration trough estimating model view matrix is used in Iterative Closest Point algorithm (ICP) [11 - 13] for 3D point clouds registration.

D. Hybrid registration model

A hybrid registration model is introduced to overcome the limitation of relying on predefined fiducial points by Kanbara et al. [14] as an extension of their work presented in [8]. The authors improve the AR system to track natural feature points along with the predefined fiducial markers. So even the initial markers go out of camera frame the system recovers projection matrix by tracking the natural features in the image of the scene and is able to achieve the registration of VOs. To enhance the feature tracking in subsequent images a prediction based hybrid method [15] is used while for detecting natural features Moravec's interest point operator [16] is employed. Hence, for estimating the projection matrix SSD based algorithm is used with the geometric description in Fig. 7. The *SSD* is defined, as in

$$SSD = \sum_i w_i (P_{rec,i} - M\widetilde{P}_{now,i})^2. \tag{22}$$

Where, w_i is a credibility parameter assign for the i^{th}feature, $P_{rec,i}$ and $\widetilde{P}_{now,i}$ are position of i^{th} feature recorded in the world-coordinate and camera-coordinate systems, respectively, in the current frame. By using this method the authors achieve a robust registration. Similarly, Meng et al. [17] also proposes a unified framework which uses multiple natural features. It represents the selected features in state space and tracks them to estimate camera pose, consequently, register a user selected VO in video.

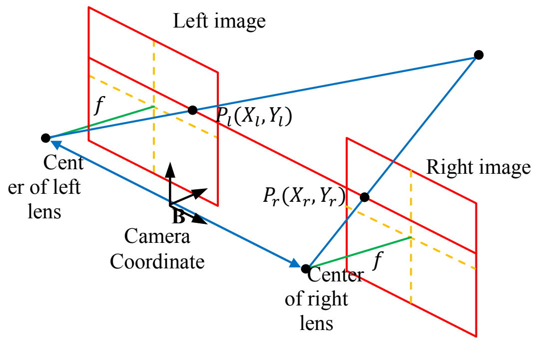

Fig. 4. Stereoscopic Geometry [8].

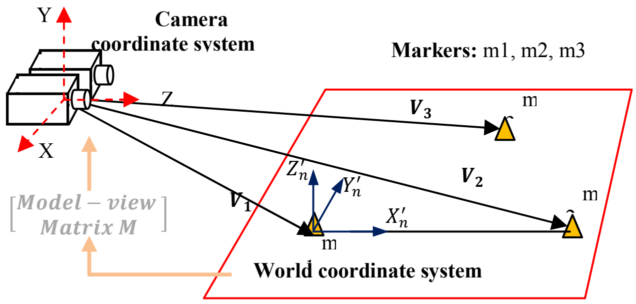

Fig. 5. Geometric Relationship Used in [8].

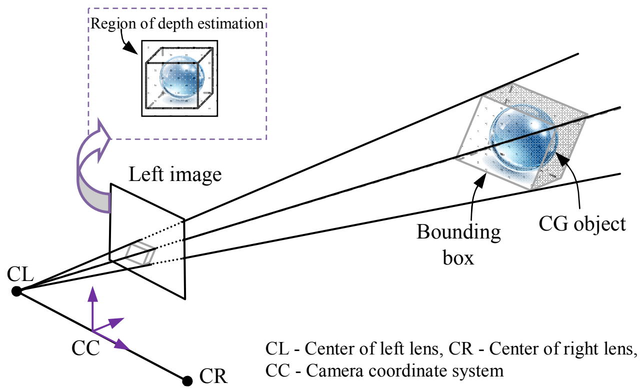

Fig.6. Occlusion Elimination Method [8].

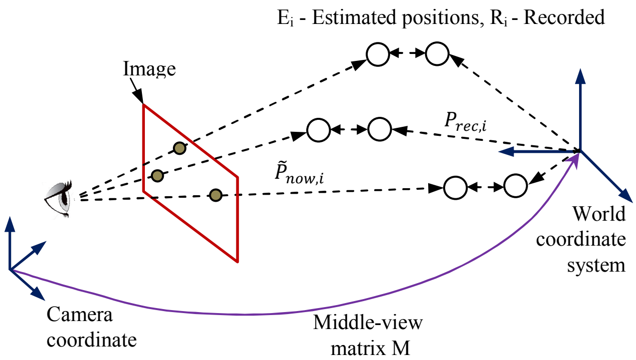

Fig.7. Registration Using Features [14].

E. Registration through blur effect estimation

All the above summarized works do not consider low quality input images; instead, require high quality and clearly detectable feature images. On the other hand, Okumura et al. [18] attempt to achieve precise geometric registration through estimation of blur effects of an image captured in degraded quality with fiducial markers in place (a sample maker is shown in Fig. 8). Through this method, the authors mainly calculate the posture and position of the camera since these are the fundamental information needed for VOs registration in AR. This approach follows as:

Step 1 – Markers of known color and shape are detected.

Step 2 – Size of blur is estimated in an edge region of the markers by using Point spread function (PSF) and the PSF parameters are calculated.

Step 3 – Based on the acquired PSF parameters and the marker shape, a template is generated. Then, SSD minimization is taken between the pixel-values of the template and captured images to estimate corner position of the markers.

Step 4 – The camera position and posture are calculated in marker coordinate system by minimizing a re-projection error of detected feature points. The employed PSF is given in (23).

$$P(x,y;r,l,\theta) = \begin{cases} \frac{1}{\Pi((r+l)^2+r^2)}; & \left(\frac{x'}{r+l}\right)^2 + \left(\frac{y'}{r}\right)^2 \leq 1 \\ 0 & ;\ otherwise \end{cases} \quad (23)$$

Where,

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}, \quad (24)$$

while (x, y), r, l, and θ are coordinate of the target pixel in the image, radius of defocusing blur, length of uniform motion, and the direction of motion respectively. The generated template from the estimated coordinates of the marker and PSF parameters are defined in (25).

$$P(x,y;r,l,\theta) = \sum_{s=-wt}^{w} \sum_{t=-w}^{w} M(x+s, +t).PSF(s,t;r,l,\theta). \quad (25)$$

Where, $M(x,y) = \begin{cases} \hat{\imath}_{black}; & a_j x + b_j y + c_j \geq 0\ (j=0 \ldots 3) \\ \hat{\imath}_{white}; & otherwise \end{cases}$, while (x, y) – a pixel coordinate in the image, w – window size = $r+l$, $\hat{\imath}_{black}$ and $\hat{\imath}_{white}$ are black and white part intensities of the marker respectively. Note that the expression $a_j x + b_j y + c_j \geq 0\ (j=0 \ldots 3)$ holds when (x, y) the point coordinate is inside the marker region. Hence, i and j are the indexes of corners and edges counted from top left marker in clockwise. Then, the SSD minimization function to estimate the corner position of the marker is defined as

$$E_{SSD,i}(\tilde{x}_i,\tilde{y}_i) = \sum_{s=-wt}^{w} \sum_{t=-w}^{w} \left[I(x_i+s, y_i+t) - (\tilde{x}_i,\tilde{y}_i)\right]^2. \quad (26)$$

If $E_{SSD,i}$ larger than the threshold th_{SSD}, the authors update shape of the marker with new a_j, b_j, and c_j from the estimated $(\tilde{x}_i,\tilde{y}_i)$.

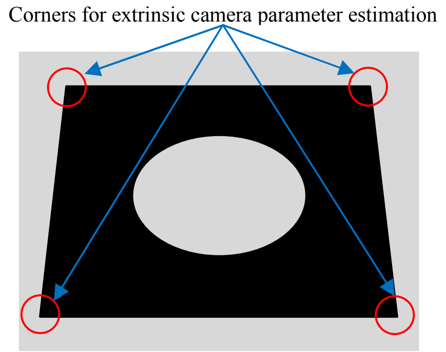

Fig. 8. An Example of the markers used in [18].

Their experiment results in comparison with Kato et al. [19] and Kanbara et al. [20] are tabulated in Table I and Table II. The results show that the method proposed by Okumura [18], improves the accuracy of the estimated corners and camera positions from hardly defocused and motion blurred images. However, the authors do not take computational cost, which is also an important consideration in AR. The Table I and Table II list the average and standard deviation errors of estimation. In table I the errors are in pixels while in table II are in millimeters.

TABLE I. CORNER COORDINATE ESTIMATION ERROR

Estimation Methods	**Corner 0**		**Corner 1**		**Corner 2**		**Corner 4**	
	Avg.	*Std.*	*Avg.*	*Std.*	*Avg.*	*Std.*	*Avg.*	*Std.*
[18]	0.59	0.31	0.45	0.20	0.72	0.28	0.56	0.25
[19]	1.43	2.34	0.71	2.48	2.22	3.47	1.91	2.87
[20]	2.08	0.45	1.03	0.38	2.97	0.35	2.70	0.41

TABLE II. DEPTH ESTIMATION ERROR.

Estimation Methods	**Exposure time**					
	15ms		**30ms**		**45ms**	
	Avg.	*Std.*	*Avg.*	*Std.*	*Avg.*	*Std.*
[18]	2.02	1.75	4.29	1.81	5.04	1.63
[19]	6.09	1.04	11.90	0.88	16.29	0.78
[20]	30.01	1.63	37.90	1.75	41.08	1.72

The advantages and disadvantages of the reviewed vision-based registration methods above are summarized in table III.

TABLE III. SUMMARY.

Method	**Advantage**	**Disadvantage**
[2]	Able to work even with minimum of three known points by using Finsterwalder's camera calibration matrix. Single camera model.	Requires known four points in the scene. Rendering delay. Lacking depth information causes overlaid VO.
[3]	Do not require artificial markers or external sensors.	Requires information of multiple planes with weightages, so high complexity.
[6], [8], [9]	Uses simple pinhole camera model. Precise since depth information is available.	Needs multiple cameras or stereo-vision and minimum of five markers for the best result. Camera calibration is required.
[14], [17]	Robust registration process.	Complex algorithm.
[18]	Able to work with blur and defocused images.	Computational cost.

III. CONCLUSION

In augmented reality, the registration plays a vital role. It depends on the accuracy of the camera parameters and projection matrix estimations. From this short review, it is found that AR is plausible with a single camera model but the VOs will be registered in front of an object in the real-world. It is because; such single camera models lack depth information of the scene. However, if an algorithm is designed to utilize images of stereo-vision and natural features available in real-world scene high accuracy of VO registration is achievable. An AR system will be more robust, if it is automated to extract and track the natural features in real-time rather than relying on predefined markers. Hence, a better vision-based registration for AR can be done through homography, blur effect estimation techniques such as PSF, and image enhancement techniques such as filtering and edge detection.

IV. FUTURE WORK

There are more literatures to be reviewed to account state of art technologies being used in industries for AR. Hence, new algorithms for vision-based registration for AR is to be implemented which will output perform the existing ones and can be applied in various applications ranging from auto mobile to medical. The results of the algorithms will be reported in the future conference proceedings and journals along with depth background information and mathematical models.

REFERENCES

[1] T. Ronald, "A survey of augmented reality," Presence, vol. 6, no. 4, pp. 355-385, August 1997.

[2] T. Okuma, K. Kiyoshi, T. Haruo and Naokaz, "An augmented reality system using a real-time vision based registration," Pattern Recognition Proceedings, Fourteenth International Conference , vol. 2, 1998.

[3] Y. Uematsu and S. Hideo, "Vision-based registration for augmented reality with integration of arbitrary multiple planes," International Conference on Image Analysis and Processing (ICIAP), pp. 155-162, 2005.

[4] S. Jianbo and T. Carlo, "Good features to track," IEEE Conference on Computer Vision and Pattern Recognition, pp. 593-600, 1994.

[5] G. Simon and M. Berger, "Reconstructing while registering: a novel apporach for markerless augmented reality.," Proc. of ISMAR, pp. 285-294, 2002.

[6] A. H. William, N. Khoi and L. Torsten, "Computer vision-based registration techniques for augmented reality," Proceedings of Intelligent Robots and Computer Vision XV, vol. 2904, pp. 538-548, 1996.

[7] Y. Hung, P. Yeh and D. Harwood, "Passive ranging to known planar point sets," IEEE International Conference on Robotics and Automation, St. Louis, Missouri, 1985.

[8] M. Kanbara, T. Okuma, H. Takemura and N. Yokoya, "A stereoscopic video see-through augmented reality system based on real-time vision-based registration," Virtual Reality, New Brunswick, NJ, 2000.

[9] S. Yong, S. Choi, J. Moon, J. Kim, and Y. W. Park "Real-time 3D registration of stereo-vision based range images using GPU," IEEE Conference on Applications of Computer Vision, Snowbird, 7-8 Dec., 2009.

[10] R. Tsai, "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses," IEEE Journal of Robotics and Automation, vol. 03, no. 4, p. 323–344, 1987.

[11] J. B. Li, F. Liu and Z. Song, "An accurate and practical means for the automatic registration of multiple 3D scanning data," 4th IEEE International Conference on Information Science and Technology (ICIST), pp. 619-622, 2014.

[12] J. Yang, H. Li and Y. Jia, "Go-ICP: solving 3D registration efficiently and globally Optimally," IEEE International Conference on Computer Vision, pp. 1457 - 1464, 2013.

[13] A. P. Bustos, T. Chin and D. Suter, "Fast rotation search with stereographic projections for 3D Registration," IEEE Conference on Computer Vision and Pattern Recognition, pp. 3930 - 3937, 2014.

[14] M. Kanbara, N. Yokoya and H. Takemura, "Registration for stereo vision-based augmented reality based on extendible tracking of markers and natural features," International Conference on Pattern Recognition, vol. 2, pp. 1045 - 1048, 2002.

[15] M. Kanbara, H. Fujii, H. Takemura and N. Yokoya, "A stereo vision-based augmented reality system with an inertial sensor," Munich, 2000.

[16] H. P. Moravec, "Visual mapping by a robot rover," International Joint Conference on Artificial Intelligence, pp.598-600, 1979.

[17] C. Meng-meng, C. Xiao-wu and L. Yi-li, "Integration of 3D registration methods based on multiple natural features," International Conference on Audio, Language and Image Processing, Shanghai, 23-25 Nov. 2010.

[18] B. Okumura, M. Kanbara and N. Yokoya, "Precise geometric registration by blur estimation for vision-based augmented reality," in International Symposium on Mixed and Augmented Reality, Nara, 13-16 Nov. 2007.

[19] B. Okumura, M. Kanbara and N. Yokoya, "Augmented reality based on estimation of defocusing and motion blurring from captured images," International Symposium on Mixed and Augmented Reality, Santa Barbard, CA, 22-25 Oct. 2006.

[20] H. Kato, M. Billinghurst, I. Poupyrev, K. Imamoto and K. Tachibana, "Virtual object manipulation on a table-top AR environment," International Symposium on Augmented Reality, Munich, Germany, 2000.