Camera Pose Estimation by Vision-inertial Sensor Fusion: An Application to Augmented Reality Books

Juan Li[1,], Hamid Aghajan[2,3], José R. Casar[1] and Wilfried Philips[2]*

[1] *School of Telecommunication, Technical University of Madrid, Spain*
[2] *TELIN-IPI-iMinds, Ghent University, Sint-Pietersnieuwstraat 41, Ghent, Belgium*
[3] *Ambient Intelligence Research (AIR) Lab, Stanford University, CA, USA*

[*] *li.juan@grpss.ssr.upm.es*

Abstract

As the booming of mobile technologies, handheld Augmented Reality draws increasing attention. One crucial process is the position and orientation estimation of the handheld device relative to a reference coordinate system in order to display the virtual content accurately upon the real world. An emerging application is to integrate multimedia elements (e.g., text, images, sound, video) in physical real books to enhance the learning process or provide amusement and interaction. Square markers are popularly used due to their ease-to-use and high accuracy. However, these markers cause visual discomfort. This paper introduces a novel pose tracking approach for mobile devices by combining a tricolored strip, an internal camera and embedded inertial sensors. Furthermore, the application to books is studied. The proposed approach is less intrusive than square markers due to its reduced size and provides accurate six-degree-of-freedom pose estimation in real time.

Introduction

Augmented Reality (AR) is a view of the physical, real-world whose elements are augmented by virtual artificial information (e.g., text, images, sound, video). With the aid of advanced AR technologies, users can better percept, interact with or manipulate the surrounding real world [1]. As the booming of mobile technologies, mobile devices are equipped with multiple sensors, touchscreen display, memory card and powerful computational capabilities. Together with the popularity and portability, mobile platforms have become favorable for designing AR applications on. Handheld AR has become increasingly attractive in recent years and has been applied in various domains, such as gaming, tourism, smart space services and education [2, 3].

One specific emerging application field in education is AR books. According to a study by Marshall et al. [4], readers still love the physical printed books, even though the electronic books are getting popular recently. This fact invoked a research study of applying AR technologies to books to digitally enhance and augment traditional real books. As an example, Billinghust et al. proposed the "MagicBook" in 2001 [5], allowing users to see animated 3D content. An observational study implemented by Dünser [6] found that, AR books are attractive to users and can enhance the learning process thanks to the augmented multimedia content, such as videos, animations.

One crucial challenge in AR is the registration of virtual content with the real world, that is, the determination of the position and orientation of the display device relative to the real world, as it decides if the virtual content is displayed in the proper position. Fiducial markers are popularly used in the literature due to their ease-to-use and high accuracy. Besides, they provide adequate patterns. ARToolKit [7] is an representative example of fiducial markers. These markers are typically made up of a thick black rectangle or circle shaped border and an embedded pattern for identification. They are designed to have high contrast to be easily detected. However, this causes their sensibility to lighting conditions. Additionally, the size of the markers and the viewing angle will also affect the accuracy. One obvious drawback of these marker-based approaches applied to books is that markers have to be big enough to provide accuracy results, thus they take up a certain amount of space in the page. This largely influences the visual effect or aesthetic. In order to get rid of fiducial markers, some research on markerless approaches has been made. SIFT [8] and SURF [9] are popularly used for natural feature tracking. However, they have high computational cost. Therefore, it still remains a challenge for real-time performance on mobile devices. Yang et al. proposed a hybrid visual tracking approach which merges the merits of fiducial marker-based tracking and markerless tracking [10]. They use a tiny marker for page recognition and a random forest for pose estimation. However, no accuracy assessment is provided.

In this context, the contribution described in this paper is the proposal and validation of a six-degree-of-freedom camera pose estimation approach by fusing vision and inertial data. A linear colored strip is employed as fiducial marker. Unlike the traditional square or circle shaped markers, this linear marker can be printed in the page border which causes no interference with the content and reduces the visual discomfort. The marker is designed to provide two reference points which are detected with an easy and fast algorithm. The marker detection results are then fused with the acceleration measurements from the embedded three-axis accelerometer to determine the position and orientation of a handheld device. In the proposed approach, only gravitational accelerations are extracted and employed to the pose calculation. As no integration is involved, this approach avoids the severe drift problem that most inertial sensor-based approaches have. In other words, our approach is drift-free. The integration of accelerometers contributes to inclination estimation and markers' 3D position reconstruction. Thus, the workload of the computer vision task is

Figure 1. *The proposed marker composed of three colors (yellow, magenta and cyan), providing two reference points P_1 and P_2.*

lightened. The experimental results show that our proposed system provides high accuracy and fulfills the real-time requirement. Page determination can be achieved by some already existed approaches, such as QR codes [11], number recognition using optical character recognition [12] or the tiny marker proposed in [10]. This paper is focused on the novel approach to handheld camera pose estimation and we will not go further into detail about the page determination.

Marker Design and Detection

The proposed marker is composed of three colors, providing two reference points, indicated as **P$_1$** and **P$_2$**, as shown in Fig. 1. In this paper we choose yellow, magenta and cyan for implementation, but it can be any combination of three different colors. According to the experimental results, the ideal width is 2 mm ∼ 10 mm and the length can vary from several centimeters to tens of centimeters, which is a suitable size for printing on the page border without any interference with the content.

The objective of the accompanying detection algorithm is to locate the positions of these two reference points in the image, which are denoted by (u_1, v_1) and (u_2, v_2) respectively. The algorithm is designed taking into account both the color and the shape of the marker. It is fast, accurate and robust to environment noises. The reference point detection algorithm is described as follows.

Firstly, the RGB image is converted to Hue-Saturation-Value (HSV) color space. Then, color segmentation is carried out to obtain regions of interest (ROIs) for the desired colors. Note that the three colored rectangles in the marker are successive. Each of the ROIs is morphologically dilated with a same structuring element. Afterwards, a logic 'AND' operation is applied to the dilated images to get the overlapped regions. The centroids of the overlapped regions are considered as the 2D locations of the reference points **P$_1$** and **P$_2$** in the image. Considering the noises (similar colors) from the page content or the environment, there may be other detected overlapped regions. To exactly locate the marker, the algorithm calculates the percentage of 'magenta' (central color) pixels in the segment between each pair of reference point candidates. This percentage is compared with an upper threshold of $T = 90\%$, which has been experimentally validated as a suitable trade-off between the detection rate and imperfect color perception. Apart from the marker, it is rare to find regions composed by the three selected colors in the image. Based on this idea, the algorithm simply chooses the longest candidate pair as the final result. Algorithm 1 summarizes previous processing.

Pose Estimation Approach

In the previous section, the 2D positions of the two reference points in the image are located by a straightforward algorithm. This section gives mathematical derivations of 6-DoF pose from the marker detection results and the accelerometer measurements.

Algorithm 1 Marker Detection

Input: Captured image I

Output: Reference points' positions in the image (u_1, v_1) and (u_2, v_2)

1: Capture the image I
2: Convert the image from RGB color space to HSV color space
3: Filter the image using the thresholds for each color and get 3 binary image, I_c(cyan), I_m(*magenta*) and I_y(yellow)
4: Morphologically dilate I_c, I_m and I_y separately using a disk kernel with a radius of 3 pixels
5: Do logic 'AND' operation and get $I_m \& I_c = I_{mc}$, $I_m \& I_y = I_{my}$
6: Find contours of I_{mc} and I_{my}, then save centroids of each contour as candidates
7: Check the pixels between each two distinct candidates. If they are mostly magenta, that is if magenta pixels/all checked pixels $> T$, save the pair as one pair candidate
8: Among all the pair candidates, choose the longest pair as the final result

---

Firstly, two involved coordinate systems are defined, as schematically depicted in Fig. 2.

- World coordinate system: This is the global reference system used for describing the position and the orientation (pose) of the camera; its origin is located in the reference point **P$_1$**; its X-axis is aligned with the marker's longitudinal direction; its Y-axis is perpendicular to the X-axis and aligned with the page border pointing up; its Z-axis is perpendicular to the X-axis and Y-axis, following the right-hand rule.

- Camera coordinate system: The origin is located in the camera optical center. The Z-axis is along the optical axis, and therefore X-Y axes are parallel to the image plane.

We denote the orientation of the camera by a rotation matrix R, which represents a rotation from the world coordinate system to the camera coordinate system. According to the Euler's rotation theorem, any rotation can be given as a composition of basic rotations about three axes. These three basic rotations are given by Euler angles, which are defined as pitch (ψ, rotation about the x axis), roll (θ, rotation about the y axis) and yaw (θ, rotation about the z axis), as shown in Fig. 2b. In principle, there are six equally valid orderings of these three angles. However, there are only two orderings that are feasible to determine the device's inclination from pure accelerometer measurements, as explained in [13]. In this paper, the rotation sequence yaw-roll-pitch is selected to express the rotation matrix. Therefore, the rotation matrix R results:

$$R = R_x(\psi)R_y(\theta)R_z(\phi) =$$
$$\begin{bmatrix} \cos\theta\cos\phi & \cos\theta\sin\phi & -\sin\theta \\ \sin\psi\sin\theta\cos\phi - \cos\psi\sin\phi & \sin\psi\sin\theta\sin\phi + \cos\psi\cos\phi & \sin\psi\cos\theta \\ \cos\psi\sin\theta\cos\phi + \sin\psi\sin\phi & \cos\psi\sin\theta\sin\phi - \sin\psi\cos\phi & \cos\psi\cos\theta \end{bmatrix}. \quad (1)$$

A. Inclination sensing using accelerometers

Assume that the book is laid on a horizontal surface, i.e., the Z-axis of the world coordinate system is pointed upward to the sky. The accelerometer embedded in the device measures the accelerations on three axes due to the forces applied to the device, including the force of gravity and external forces. In this paper,
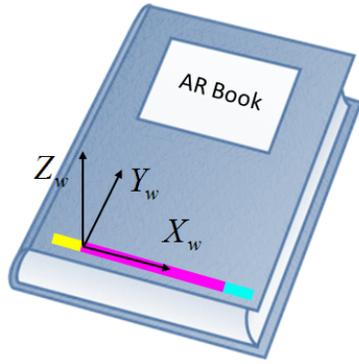
(a)
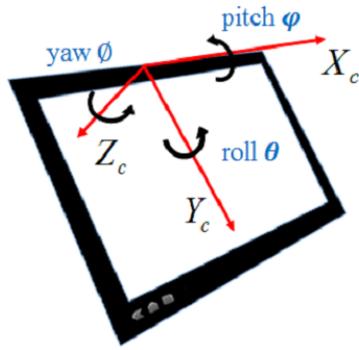
(b)

Figure 2. *Coordinate systems. (a) World coordinate system. (b) Camera coordinate system.*

only the gravitational acceleration is isolated and used for pose estimation. The isolation can be achieved by applying a low-pass filter. In case of an Android platform, a synthetic gravity sensor is available since API 9 was released.

The gravitational acceleration measurement is a three dimensional vector indicating the direction and magnitude of gravity, denoted by $\mathbf{a}_g = [g_x, g_y, g_z]^T$. Note that $\mathbf{a}_g$ is expressed in the local accelerometer coordinate system. In the world coordinate system, the gravitational acceleration is express as $[0, 0, g]^T$, where g is the value of earth gravitational acceleration. These two vectors are related through the rotation matrix R as

$$\mathbf{a}_g = \begin{bmatrix} g_x \\ g_y \\ g_z \end{bmatrix} = R \begin{bmatrix} 0 \\ 0 \\ -g \end{bmatrix} = \begin{bmatrix} g\sin\theta \\ -g\sin\psi\cos\theta \\ -g\cos\psi\cos\theta \end{bmatrix}. \quad (2)$$

Then, pitch and roll angles can be deduced from the measurements of gravitational accelerations as

$$\psi = \arctan\frac{g_y}{g_z}, \quad (3)$$

$$\theta = \arcsin\frac{g_x}{g}. \quad (4)$$

B. Pose estimation approach by sensor fusion

The camera pose can be described as a composition of a 3×3 rotation matrix R and a 3×1 translational vector $\mathbf{t}$ (the translation

from the origin in the world coordinate system to the camera coordinate system). The objective is to solve R and $\mathbf{t}$ at each time instant.

Let us start with the detected reference points from the images. A pinhole camera model is adopted and the mapping of 3D points in world coordinates to 2D points in the image plane can be mathematically described as

$$\lambda \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K(R\mathbf{P}+\mathbf{t}) = \begin{bmatrix} f_x & 0 & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \left(\begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} \right), \quad (5)$$

where (u, v) are the coordinates of a 2D point; K is a matrix containing camera intrinsic parameters: the focus length f_x, f_y and the principal point (u_o, v_o); λ is a scale factor.

Firstly, we apply Eq. 5 to the reference point $\mathbf{P_1}$, whose coordinates in the world coordinate system is $(0,0,0)^T$ and whose coordinates in the camera frame coordinate system is (u_1, v_1) and get

$$\lambda_1 \begin{bmatrix} u_1 \\ v_1 \\ 1 \end{bmatrix} = K(R \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} + \mathbf{t}) = K\mathbf{t} = \begin{bmatrix} f_x t_x + u_o t_z \\ f_y t_y + v_o t_z \\ t_z \end{bmatrix}. \quad (6)$$

Eq. 6 can be rewritten as two separate equations by eliminating the variable λ_1:

$$u_1 = (f_x t_x + u_o t_z)/t_z, \quad (7)$$

$$v_1 = (f_y t_y + v_o t_z)/t_z. \quad (8)$$

Then, the term $K\mathbf{t}$ can be substituted as

$$K\mathbf{t} = t_z[u_1, v_1, 1]^T. \quad (9)$$

The reference point $\mathbf{P_2}$ is expressed in the world coordinate system as $(D,0,0)^T$, where D is a known value indicating the distance between two reference points. Its 2D position in the image is indicated as (u_2, v_2). We apply Eq. 5 to the reference point $\mathbf{P_2}$ and get

$$\lambda_2 \begin{bmatrix} u_2 \\ v_2 \\ 1 \end{bmatrix} = K(R \begin{bmatrix} D \\ 0 \\ 0 \end{bmatrix} + \mathbf{t}) = KR \begin{bmatrix} D \\ 0 \\ 0 \end{bmatrix} + K\mathbf{t}. \quad (10)$$

By substituting Eq. 9 into Eq. 10, we can obtain

$$\lambda_2 \begin{bmatrix} u_2 \\ v_2 \\ 1 \end{bmatrix} = \begin{bmatrix} Df_x r_{11} + Du_o r_{31} + u_1 t_z \\ Df_y r_{21} + Dv_o r_{31} + v_1 t_z \\ Dr_{31} + t_z \end{bmatrix}. \quad (11)$$

Eq. 10 can also be rewritten as two separate equations by eliminating the variable λ_2:

$$u_2 = (Df_x r_{11} + Du_o r_{31} + u_1 t_z)/(Dr_{31} + t_z), \quad (12)$$

$$v_2 = (Df_y r_{21} + Dv_o r_{31} + v_1 t_z)/(Dr_{31} + t_z). \quad (13)$$

Please remind that two rotation angles from the rotation matrix R are resolved by accelerometer measurements in Eq. 3 and Eq. 4. Therefore, in the system of equations composed by Eq. 12 and Eq. 13, there are only two unknown variables, i.e., t_z and the rotation angle θ. Thus, t_z and θ can be solved from this determined system.

As a consequence, the elements t_x and t_y of the translation vector can be calculated from Eq. 7 and Eq. 8. Until here, we have resolved all the parameters of the six-degree-of-freedom pose. The whole algorithm architecture is shown in Fig. 3.
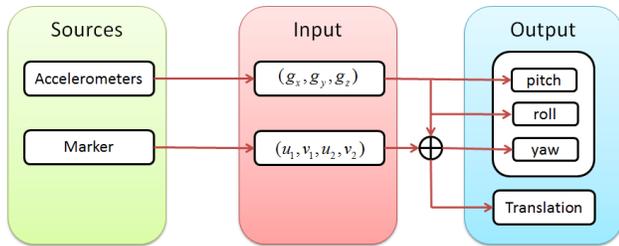
Figure 3. *The architecture of the proposed pose estimation approach by sensor fusion*

Figure 4. *An example of a 3D virtual model popping up out of a page.*

Augmentation

Texts and planar images may be abstract for learners to understand or imagine in spatial domain. However, with the aid of AR technologies to superimpose supplementary information rendered in a multimedia format, learning or reading can become easier and more attractive than using a physic book. For example, AR can enable readers to visualize a world globe upon a 2D world map or text. Also, it can enhance the learning process of chemistry, physiology, astronomy, etc. in an immersive and interactive way.

A variety of multimedia elements can be integrated in a physical book:

- 2D content: texts, images, illustrations, hyperlinks, videos and animations. For example, an animation of how to write alphabets is helpful for preschool children.

- 3D content: 3D models, animations. In Fig. 4, we give an example of a 3D tiger model popping up out of a text related to tigers. Readers can also rotate the book or the display device to have a 360-degree view of the model.

- Interaction: users are allowed to interact with the physical world through manipulating the virtual content. For example, users can move and place an interactive character to accomplish a game.

- Sound: music, audio guide.

Experiments

In this section, the accuracy and the real-time performance of the propose approach are assessed.

A. Prototype set up

A tablet PC Nexus 10 running on the Android operating system is chosen as the AR device. The marker is designed as 19 cm$\times$ 0.3 cm (yellow: 2.5 cm, magenta: 14 cm and cyan: 2.5 cm), which is a suitable size for a A4 page. The image resolution is set at 640 $\times$ 480 pixels.

B. Testing procedure

The pose estimation accuracy is estimated dynamically in terms of projection errors in pixels, i.e., the displacement of the augmented object from the real object in the image. We benchmark our results to the ground truth generated by OpenCV library [14], an open source, widely-used and cross-platform computer vision library. OpenCV finds the position of internal corners of a chessboard using the function *findChessboardConers()*.

To set this benchmark up, a chessboard composed of 4×3 grids of 20 mm $\times$ 20 mm has been chosen as the physical augmented object (forming 12 corners). The chessboard is placed in a known position in the global coordinate system, as shown in Fig. 5b. The detected positions of these 12 corners by OpenCV are considered as ground truth. In our proposed system, firstly, the position and the orientation of the tablet camera relative to the stripe marker is estimated by fusing vision and inertial data. Based on the estimated pose, the 12 corners with known world coordinates are projected to the image. Then, the root-mean-square (RMS) of the displacements (in pixels) of the calculated positions of the corners from ground truth are considered as the average projection error in the current frame. The whole procedures can be summarized as follows:

(1) Calibrate the device camera's intrinsic parameters.

(2) Estimate the position and orientation of the camera relative to the world coordinate system using our proposed approach.

(3) Project chessboard corners with known 3D positions to the image using the estimated camera pose.

(4) Detect the positions of the chessboard corners in the image using the function *findChessboardConers()* from OpenCV. The results are used as ground truth.

(5) Calculate the displacement error of the projected 2D points from the detected points.

In Fig. 5, two captured examples of the running system are presented. Based on the estimated camera pose, the origin and X-Y-Z axes of the world coordinate system are projected in the image, as shown in Fig. 5a. The comparison of the proposed system to the OpenCV chessboard corner detection is shown in Fig. 5b. As we can see, the projected chessboard corners using our approach are in close correspondence to the detected chessboard corners.
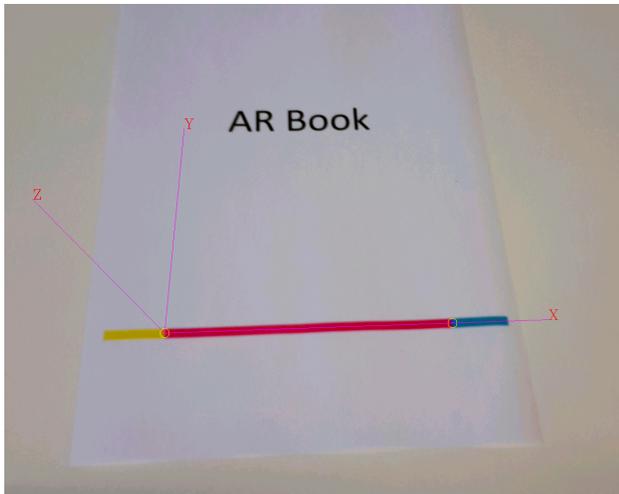
(a)
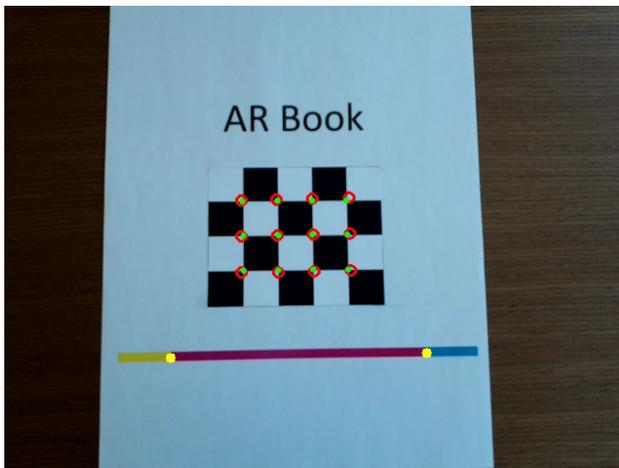
(b)

Figure 5. *(a) The world coordinate system is projected in the image based on the estimated camera pose. (b) Accuracy assessment using the corner detection results from OpenCV as ground-truth. Red circles are the calculated corners positions and the green circles are the corner detection results from OpenCV. The yellow circles are detected reference points positions in the image.*

The RMS errors for 400 frames with varying camera motions and varying distances to the marker are illustrated in Fig. 6. Then, we calculated the overall RMS error of these 400 frames and obtained an average projection error of 4.8 pixels. This error is insignificant for most of the applications.

C. Execution time

Regarding the execution time, an evaluation was carried out by computing the average time needed for frame capturing, image processing, pose calculation and image visualization. The OpenCV chessboard detection method (in particular, the function *findChessboardCorners()*) has two working modes: a fast check mode and an adaptive thresholding mode. The fast check mode runs a fast check on the image to look for chessboard corners and shortcuts the call if none is found. It drastically speeds up when no chessboard is observed. Adaptive thresholding mode is ro-
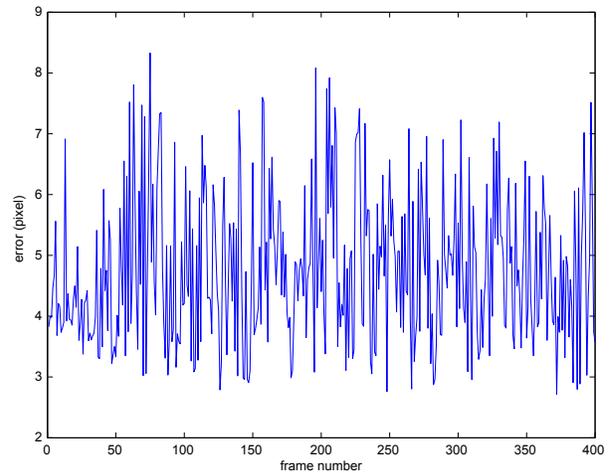

Figure 6. *RMS errors for 400 frames with varying camera motion and varying distance.*

bust to the illumination changes. However, it is time-consuming. We compared the execution time of the proposed system with the OpenCV method both in the fast check mode and the adaptive thresholding mode.

During the experiment, we implemented the three methods on the same device and calculated the average processing time for 1200 frames. Both the proposed system and the reference system in fast check mode have achieved 11 frames/s, while the reference system in adaptive thresholding mode has achieved 2.5 frames/s.

D. Some notes on error sources

When analyzing the error sources of the experiments, the following aspects are considered:

- Imperfect camera calibration: Even though the camera calibration can estimate the intrinsic parameters of the camera, it is not perfect. Also, all cameras are inevitable to have some lens distortion.

- Reference point detection errors: These errors are due to image acquisition and the reference point detection algorithm, such as image noise, sensor spatial quantization.

- Gravitational acceleration measurement errors: All accelerometers suffer from a variety of error sources, such as temperature, mechanical misalignments, stochastic variation (random noises) [15].

Conclusions and Future Work

The paper presents a novel hybrid approach to handheld camera pose estimation by combining marker detection and inertial sensors. An application to AR books is described and analyzed. The experimental results show that the system is easy-to-use, fast and accurate, fulfilling the requirements of AR applications.

Although in this paper the page determination is not implemented, it will be included in the future work. Furthermore, an AR book with augmented content and interaction will be completed and related issues, such as UI design, content rendering, will be studied.

Acknowledgments

References

[1] Azuma, Ronald T. "A survey of augmented reality." Presence 6.4 (1997): 355-385.

[2] Li, Juan, et al. "A new 360-degree immersive game controller." Proceedings of the 9th International Conference on Distributed Smart Camera. ACM, 2015.

[3] Van Krevelen, D. W. F., and R. Poelman. "A survey of augmented reality technologies, applications and limitations." International Journal of Virtual Reality 9.2 (2010): 1.

[4] Marshall, Catherine C. "Reading and interactivity in the digital library: Creating an experience that transcends paper." Proceedings of CLIR/Kanazawa Institute of Technology Roundtable. Vol. 5. No. 4. 2005.

[5] Billinghurst, Mark, Hirokazu Kato, and Ivan Poupyrev. "The magicbook-moving seamlessly between reality and virtuality." Computer Graphics and Applications, IEEE 21.3 (2001): 6-8.

[6] Dnser, Andreas, and Eva Hornecker. "An observational study of children interacting with an augmented story book." Technologies for E-Learning and Digital Entertainment. Springer Berlin Heidelberg, 2007. 305-315.

[7] Kato, Hirokazu, and Mark Billinghurst. "Marker tracking and hmd calibration for a video-based augmented reality conferencing system." Augmented Reality, 1999.(IWAR'99) Proceedings. 2nd IEEE and ACM International Workshop on. IEEE, 1999.

[8] Lowe, David G. "Distinctive image features from scale-invariant keypoints." International journal of computer vision 60.2 (2004): 91-110.

[9] Bay, Herbert, Tinne Tuytelaars, and Luc Van Gool. "Surf: Speeded up robust features." Computer visionECCV 2006. Springer Berlin Heidelberg, 2006. 404-417.

[10] Yang, Hyun S., et al. "Hybrid visual tracking for augmented books." Entertainment Computing-ICEC 2008. Springer Berlin Heidelberg, 2009,pg. 161-166.

[11] Kan, Tai-Wei, Chin-Hung Teng, and Wen-Shou Chou. "Applying QR code in augmented reality applications." Proceedings of the 8th International Conference on Virtual Reality Continuum and its Applications in Industry. ACM, 2009.

[12] Qadri, Muhammad Tahir, and Muhammad Asif. "Automatic number plate recognition system for vehicle identification using optical character recognition." Education Technology and Computer, 2009. ICETC'09. International Conference on. IEEE, 2009.

[13] Tuck, Kimberly. "Tilt sensing using linear accelerometers." Freescale Semiconductor Application Note AN3107, 2007

[14] OpenCV, http://www.opencv.org, Accessed:2015-11-29

[15] Park, Minha, and Yang Gao. "Error and performance analysis of MEMS-based inertial sensors with a low-cost GPS receiver." Sensors 8.4 : 2240-2261, 2008.

Author Biography

Juan Li received her BS in electrical engineering from Beihang University in 2010. Since September 2011, she has been doing her Ph.D in School of Telecommunication in Technical University of Madrid. She worked in Ghent University as a visiting scholar from January to August, 2015. Her work has focused on object pose estimation, augmented reality, interaction in smart spaces, indoor localization and image processing.

Hamid Aghajan is with the Department of Telecommunication and Informatics (TELIN) in Gent University, and has also been director of the Ambient Intelligence Research (AIR) Lab at Stanford University. He received his BS degree in 1989 from Sharif University of Technology, Tehran, Iran, and his MS and PhD degrees from Stanford University in 1991 and 1995, respectively, all in Electrical Engineering. Focus of research in his group is on methods and applications of Ambient Intelligence with an emphasis on behaviour modelling based on activity monitoring with multi-camera networks.

José R. Casar is a full-time professor in Universidad Politécnica de Madrid and is heading the Data Processing and Smart Spaces Group. He graduated in Telecommunication Engineering in 1981 and gained a Ph.D. degree in 1983 from Department of Signals, Systems and Radio communications of Universidad Politécnica de Madrid (UPM). His current interests include the Internet of things, advanced interaction methods, user experience design and smart spaces. His group is running an Experience Laboratory on the Spaces of the Future.

Wilfried Philips is a full-time professor in Ghent University and is heading the research group "Image Processing and Interpretation", which is also part of the Flemish ICT research institute iMinds. In 1989, he received the Diploma degree in electrical engineering and in 1993 the Ph.D degree in applied sciences, both from Ghent University, Belgium. Some of the recent research activities in the group include image and video restoration and analysis, image and video quality assessment and image analysis and computer vision.