ISSN 2356-4407

IC-ITechS

www.STIKI.ac.id

menuju 30th STIKI
Bersama memajukan ICT Indonesia

PROCEEDING
IC - ITECHS 2014

The 1st International Conference on Information Technology and Security

Malang, November 27, 2014
Published by:
Lembaga Penelitian dan Pengabdian pada Masyarakat
Sekolah Tinggi Informatika dan Komputer Indonesia

PROCEEDING

The 1st International Conference on Information Technology and Security (IC-ITechs)

November 27, 2014

Editors & Reviewers:

Tri Y. Evelina, SE, MM Daniel
Rudiaman, S.T, M.Kom Jozua
F. Palandi, M.Kom

Layout Editor:

Eka Widya Sari

LEMBAGA PENELITIAN & PENGABDIAN KEPADA MASYARAKAT
Sekolah Tinggi Informatika & Komputer Indonesia (STIKI) – Malang
Website: itechs.stiki.ac.id E-mail: itechs@stiki.ac.id

PROCEEDING

The 1st International Conference on Information Technology and Security (IC-ITechs) November 27, 2014

ISSN 2356 - 4407

viii + 276 hlm; 21 X 29,7 cm

Reviewers & Editors:

Tri Y. Evelina, SE, MM
Daniel Rudiaman, S.T, M.Kom
Jozua F. Palandi, M.Kom

Layout Editor:

Eka Widya Sari

Published by:

LEMBAGA PENELITIAN & PENGABDIAN KEPADA MASYARAKAT

Sekolah Tinggi Informatika & Komputer Indonesia (STIKI) – Malang

Jl. Raya Tidar 100 Malang 65146, Tel. +62-341 560823, Fax. +62-341 562525

Website: itechs.stiki.ac.id E-mail: itechs@stiki.ac.id

GREETINGS

Head of Committee IC-Itechs

For all delegation participants and invited guest, welcome to International Conference on Information Technology and Security (IC-Itechs) 2014 in Malang, Indonesia.

This conference is part of the framework of ICT development and security system that became one of the activities in STIKI and STTAR. this forum resulted in some references on the application of ICT. This activity is related to the movement of ICT development for Indonesia.

IC-Itechs aims to be a forum for communication between researchers, activists, system developers, industrial players and all communications ICT Indonesia and abroad.

The forum is expected to continue to be held continuously and periodically, so we hope this conference give real contribution and direct impact for ICT development.

Finally, we would like to say thanks for all participant and event organizer who involved in the held of the IC-Itechs 2014. We hope all participant and keynote speakers got benefit from this conference.

LIST OF CONTENT

3D Interaction in Augmented Reality Environment with Reprojection Improvement on Active and Passive Stereo

Eko Budi Cahyono[1], Ilyas Nuryasin[2], Aminudin[3]
[1,2,3]Department of Informatics, University of Muhammadiyah Malang,
[1]ekobudi@umm.ac.id, [2]ilyas@umm.ac.id, [3]aminudin@umm.ac.id

Abstract

This research proposes the three dimensional (3D) interaction on the augmented reality (AR) environment based on interposition. Interposition is captured from the intersection of reconstruction points. That points should be increased to generate accurate interposition. The optimal triangulation is used to increase the performance for reconstructing 3D. This method is applied to Logitech camera (passive stereo) and Kinect camera (active stereo) to analyze the result of reconstruction. Based on the calibration testing, optimal triangulation gives a significant deference accuration on both of Logitech and Kinect cameras. On the passive stereo, accurate interaction is generated by combining subpixel detection and the optimal triangulation algorithm, while on the active stereo, the accuration of depth image can be improved by just applying the optimal triangulation algorithm.

Keywords: *3D interaction, augmented reality, optimal triangulation, active stereo, passive stereo*

1. INTRODUCTION

The 3D hand gesture interaction is the most relevant approach when interacts on the augmented reality environment. This approach is natural, simple, cheap, fast, accurate, and immersive (Lee et al, 2008). The interaction is performed through the triangulation stereo interface, either the active (structured light) or passive stereo (multiple view geometry). This research aims to:

a. Developing an interaction model between hand gesture and AR environment.
b. Improving the accuration of reprojection point through an optimal triangulation.

Two dimensional (2D) interaction that performed in the desktop doesn't meet the special requirement when interacts in 3D, such in AR environment. This requirement is the freedom of movement (gesture) from one point to another across fields. By using triangulation stereo (Luong and Faugeras, 1996; Hartley, 1997; Hartley and Zisserman, 2003), reconstruction of 3D point is performed within AR environment. The Interposition of reconstruction above is the basis of interaction.

This research proposes the model of 3D interaction based on interposition. The improvement of reprojection point is performed before reconstructed to be 3D point using optimal triangulation, either for active stereo (Smisek et al, 2011) or passive stereo (Luong and Faugeras, 1996; Hartley, 1997) and thus resulting the more precise reconstruction point as the element of the interposition generator on a loosely scale.

2. RESEARCH METHOD

The research is conducted in a planned and systematic method. The approach that used in this study is an experimental analysis. The method that used by the previous researchers is analyzed and applied to meet the research goals. The selected method is a simple, fast and accurate. Technically, the method is divided to be 3 sections: calibration, hand gesture detection and interaction.

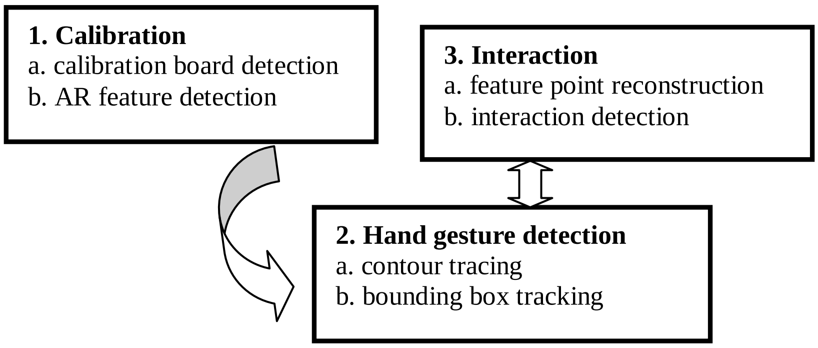


Figure 1. Methodology of research

2.1 CALIBRATION MODEL

Calibration relates to the determination of intrinsic and extrinsic parameters on the camera. In this study, calibration is performed indirectly through fundamental matrix (Hartley, 1997). The calibration model consists of 3 sub-models: calibration board detection, AR feature detection, and AR feature reconstruction.

2.1.1 CALIBRATION BOARD DETECTION

Calibration board that selected is a checkerboard with 8x6 vertices. A vertex is formed from 2 vertical lines. Vertex is a the midpoint of four squares. The pattern that formed inside is a square with black or white side.

Eight points corresponded that obtained from calibration board detection is location of vertex pixel in the image coordinate at index 0, 6, 12, 18, 24, 30, 36, 42 for each stereo image, either left or right image. Pixel location of the vertex is selected by consideration of equitable distribution of vertex on the calibration board.

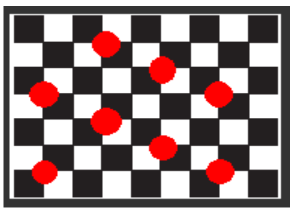
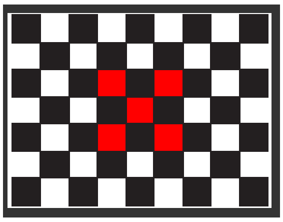


Figure 2. Eight points corresponded

2.1.2 AR FEATURE DETECTION

The AR environment is represented by AR feature as many as five points taken from midpoint of square (AR object). Square model is assumed to be on the planar and each side is perpendicular. So, that mid point is calculated by following formula:



$$\begin{aligned} midpoint_x &= x_1 + dx_1 \\ midpoint_y &= y_1 + dy_1 \end{aligned} \quad (1)$$

$$\begin{aligned} dx_1 &= (x_2 - x_1) / 2 \\ dy_1 &= (y_3 - y_1) / 2 \end{aligned}$$

Figure 3. Model of AR environment

2.2 MODEL OF HAND GESTURE DETECTION

Hand gesture is modeled by a red marker that attached on the tip of the right index finger. That marker is called as hand feature. Hand feature has coordinate system on the real coordinate that represented as P_f (x_f , y_f , z_f). This model consists 2 parts: contour tracing and bounding box tracking.

2.2.1 CONTOUR TRACING

Contour is an ordered list of edge pixels or connected component. Tracing the contour can be performed by using contour tracing method (Chang et al, 2004).

Contour tracing method is a method for detecting outer contour and may be inner contour**,** recognizing and labeling the inside of each component. This method is based on the fact that a component is completely determined by the contour.

Labeling is performed in a single visit from the top to the bottom and from the left to the right for each line. However, the contour point is possibly visited more than once to a specific number. In addition, it is not only for labeling to a connected component, but also extracting its contour. So, it results a series of sequence contour point (Chang et al, 2004).

2.2.2 BOUNDING BOX TRACKING

Tracking is an activity to detect any change of object position from one state to another. Bounding box is an area that bounds the points are in the selected area. On the 2D image, that area is a rectangular contour.

Bounding box tracking (Senior, 2002) is selected because the hand gesture detection results a connected component labeling that has been recognized its area, so it will be more effective if performed on that area.

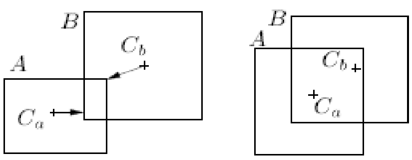


Figure 4. Bounding box tracking

Based on figure 4, there are 2 bounding boxes that represent 2 components, object A (AR environment) and B (hand gesture). The distance between 2 bounding boxes is the shortest distance from the midpoint C_a to the nearest point on the bounding box B or the shortest distance from the midpoint C_b to the nearest point on the bounding box A (left image). If the midpoint is in the another bounding box, the distance of two bounding boxes is equal to 0 (right image).

2.3 INTERACTION MODEL

Interaction between hand gesture and AR environment is by pointing a specific location by right index finger on the calibration board as an AR symbol. This model consists of two parts: feature point reconstruction and intersection detection / interaction.

2.3.1 FEATURE POINT RECONSTRUCTION

Feature point reconstruction is a series of tasks for mapping the real and virtual points on the computer. Basis of this reconstruction is the geometry of stereo vision. On the passive stereo, a real point is taken from two different viewpoints with two cameras and each camera takes one viewpoint. On active stereo, a real point is taken from two different viewpoints with one camera and projector.

A set of stereo image results correspondence $x_i \leftrightarrow x_i^{'}$. This correspondence starts from X_i on the real coordinate that projected to x_i and $x_i^{'}$ on the camera coordinate. On this state, reconstruction performs to reprojects x_i and $x_i^{'}$ to be X_i on the real coordinate. Triangulation is directly related with epipolar geometry, a projection geometry for stereo vision. The core of epipolar geometry are fundamental and camera matrix (Luong and Faugeras, 1996; Hartley, 1997). Fundamental matrix is algebraic representation of epipolar geometry. Geometrically, fundamental matrix is derived by two steps: point mapping on a board and constructing epipolar line. First step, X is mapped to some *X'* points along epipolar line $l^{'}$. One of X' point is potentially corresponded with X point. Second step, epipolar line $l^{'}$ defined as a connector $x^{'}$ to epipolar $e^{'}$.

Referring to figure 5, the light that is passing through the center of first camera cut the image plane at x point and also cut the ∏ plane on the X point. In the same way, all x_i points on the left image plane will correspond with x_i' on the right image plane. This is because it uses equivalence projection and then called as homography $H_{\prod}$. In the mathematical notation $l^{'} = [e^{'}]_x H_{\prod} x = Fx$. While in general notation fundamental $F = [e^{'}]_x H_{\prod}$, where $H_{\prod}$ is a function for mapping the set of point from one image to another through ∏.

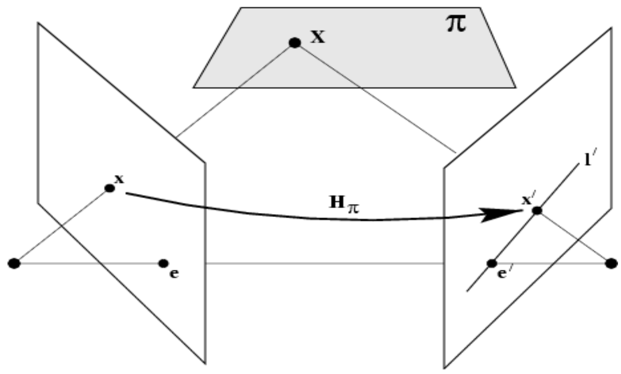


Figure 5. Epipolar geometry

Camera matrix is a matrix that describes the mapping of pinhole camera from real point to image point and written by notation x=PX. X is the real point in homogeneous coordinate (vector-4), whether X is image point in the pinhole camera with homogeneous coordinate (vector-3). Camera matrix can be obtained from fundamental matrix, because the fundamental matrix is representation of algebraic epipolar geometry. It also contains information related to the camera matrix. On the stereo vision, camera matrix can be expressed by two matrices, $P = [I \mid 0]$ and $P^{'} = [[e^{'}]_x F \mid e^{'}]$.

When two cameras or camera-projector capture the real point from two different position, it establishes a geometry connection between real point and its projection point on the image plane. That connection generates a correspondence of points on the image plane.

If the camera matrix is P and P', it means x and x' meeting the epipolar requirement $x'^{T}Fx = 0$, or x' is on the epipolar line Fx. It also means that two projection back lights from x and x' are on the single epipolar plane, as seen at figure 5. Because two lights are on the single plane, they met at some points. Two back lights started from the center of camera, each of them passes through the epipolar, then both of them met at reconstruction point X. This method is called as triangulation. Triangulation is applied with camera matrix P and P' as equation below:

$$X=\Gamma(x, x', P, P')=H^{-1}\,\Gamma(x, x', P H^{-1}, P' H^{-1}) \tag{2}$$

An optimal triangulation is obtained by recalculating corresponding point, so the geometrical error can be reduced. Hartley has analyzed this issue by proposing an algorithm called optimal triangulation algorithm (Hartley and Zisserman, 2003).

The procedure to apply this algorithm:

a. Determining corresponding point, (x1, y1) and (x2, y2)
b. Defining the translation matrix
c. Calculating matrix $F \leftarrow (T2^{T}*F)*T1$
d. Calculating the epipole by processing the singular value decomposition (SVD)
e. Defining the rotation matrix
f. Calculating matrix $F(R2*F)*R1^{T}$.
g. Determining $f1=E1_{2,0}$ $f2=E2_{2,0}$ $a=F_{1,1}$ $b=F_{1,2}$ $c=F_{2,1}$ $d=F_{2,2.}$
h. Calculating six roots of the equation $g(t) = k6*t^{6} + k5*t^{5} + k4*t^{4} + k3*t^{3} + k2*t^{2} + k1*t + k0$
i. Getting the value of the smallest (t) root of the equation
   $s(t) = t^{2}/(1+f1^{2}*t^{2})+(c*t+d)^{2}/(a*t+b)^{2}+f2^{2}*(c*t+d)^{2}$
j. Finding the new corresponding point

The optimal triangulation has a transformation method that is similar with the direct linear transformation / DLT (Aziz and Karara, 1971). On the paired stereo image obtained that $x = PX$, $x' = P'X$ and that measurement is combined to be a linear equation of X or AX=0, where X is the real coordinate system.

Homogeneous scale factor is removed by cross multiplication to get three equations, where two of them linear-independent. For example, $x \times (PX)=0$ results:

$$\begin{aligned} x(p^{3T}X) - (p^{1T}X) &= 0 \\ y(p^{3T}X) - (p^{2T}X) &= 0 \\ x(p^{2T}X) - y(p^{1T}X) &= 0 \end{aligned} \tag{3}$$

where p^{iT} is i^{th} row of matrix P. This equation is linear as a part of matrix X. Then, the equation AX=0 can be arranged by A.

$$
\begin{matrix} xp^{3T} - p^{1T} \\ yp^{3T} - p^{1T} \\ x'p^{3T} - p'^{1T} \\ y'p^{3T} - p'^{2T} \end{matrix} \qquad (4)
$$

Equation (4) results four linear equations. They can be solved by using DLT, or generating the correspondence between singular vector unit and the smallest singular solution of matrix A.

2.3.2 INTERSECTION DETECTION

Intersection between AR feature reconstruction point and hand feature reconstruction can be detected by using axis-aligned bounding boxes / AABB (Christer Ericson, 2005). One type of AABB is center-radius. Center-radius is selected because both of AR and hand feature come from the midpoint of bounding box. AR feature comes from midpoint of the calibration board, while hand feature comes from the midpoint of bounding box the hand gesture component.

With this approach, the procedure of intersection detection and interaction as follows:
a. Detecting and reconstruction AR feature then getting pont $P_{b.}$
b. Detecting and reconstructing hand gesture then getting point $P_{f.}$
c. Intersection occurs if $|x_b-x_f|<r_x$ && $|y_b-y_f|<r_y$ && $|z_b-z_f|<r_z$
d. If the step c the intersection occurs, the interaction is activated.
e. Repeating steps b-c-d till the last frame.

3. RESULT AND DISCUSSION

Based on the issues on chapter 2, defined the variables that affect the result as follows:

a. Amount of camera: (two types) two cameras Logitech and a set of Kinect cameras.
b. Amount of feature point: calibration (eight), AR (five), hand gesture (one).
c. Camera pose (three random configurations).
d. The accuration of feature point detection (pixel, subpixel).
e. The accuration of reconstruction (DLT, optimal triangulation).

This chapter is divided to three parts: epipolar geometry, optimal triangulation accuration dan interaction model.

3.1 EPIPOLAR GEOMETRY

To calculate the fundamental matrix is by knowing the corresponding point set in a pair of stereo image. The stability of fundamental matrix can be measured by using epipolar requirement $x'^T F x = 0$, then derived to be epipolar difference equation (ede). If be getting closer to zero, the fundamental matrix is more stable.

$$ede = d(x'^T F x ,0)^2 \qquad (5)$$

The result of comparison between (ede) and the variables are follows:

Table 1. Test result of fundamental matrix

Logitech

pose	DLT		optimal triangulation	
	pixel	subpixel	pixel	subpixel
1	1,88E-04	1,42E-08	2,66E-11	1,85E-13
2	3,30E-07	3,38E-06	1,55E-13	2,56E-13
3	7,01E-06	7,76E-08	1,79E-15	1,54E-13
mean	6,50E-05	1,16E-06	8,92E-12	1,99E-13

Kinect

pose	DLT		optimal triangulation	
	pixel	subpixel	pixel	subpixel
1	2,14E-04	7,63E-04	5,49E-12	7,58E-13
2	2,48E-06	1,17E-05	9,96E-15	6,12E-15
3	1,28E-05	1,20E-05	5,56E-14	2,38E-13
mean	7,63E-05	2,62E-04	1,85E-12	3,34E-13

Based on table 1, it is known that fundamental matrix has a high stability when applied by optimal triangulation, either Logitech or Kinect camera.

On the Logitech camera, fundamental matrix is more stable if the calibration point set is detected by using subpixel detection, while in Kinect not always. This is because IR camera on the Kinect sensor affected by structured light that are characteristic of active triangulation. This makes subpixel detection is not affect the fundamental matrix stability.

3.2 OPTIMAL TRIANGULATION ACCURATION

AR environment is represented by AR feature point. Five AR feature points are taken from calibration board. They are registered once after epipolar geometry construction is formed. Registration consists of detection and AR feature point reconstruction. Registration depends on the existing epipolar geometry construction, so it doesn't need a recalibration. Once the epipolar geometry construction formed, fundamental and camera matrix are resulted. Fundamental matrix is used to calculate optimal triangulation, while camera matrix is used to calculate DLT. Camera matrix can be calculated from the fundamental matrix.

Table 2. Mean of reprojection errors from reconstruction point

Logitech

	DLT	optimal triangulation
pixel	6.72E+00	1.31E-10
subpixel	7.28E+00	1.04E-10
mean	7.00E+00	1.17E-10

Kinect

	DLT	optimal triangulation
pixel	1.20E+00	2.29E-10
subpixel	6.76E-01	3.70E-10
mean	9.37E-01	4.11E-10

Based on tabel 2, the optimal triangulation significantly affects the accuration of reconstruction. It occurs on both Logitech and Kinect cameras. Thus, the optimal triangulation is strongly affected by the stability of fundamental matrix.

3.3 INTERACTION MODEL

Interaction model between AR environment and hand gesture is by pointing a specific location by the right index finger on the calibration board as a AR interaction symbol. AR environment is represented by AR feature point, while hand gesture is represented by hand gesture feature point. Five AR feature points are taken calibration board, while an hand gesture feature point from a colored marker that attached to the tip of right index finger.

Both of the real points are represented by AR feature reconstruction point and hand gesture feature to detect the interposition. If interposition occurs, the interaction will be activated.

Interaction response between hand gesture feature and AR environment is more influenced by feature point detection accuration than reconstruction accuration. This can be evidenced by Table 3 that the feature point detection accuration is smaller by 3% compared to the reconstruction accuration by 97%.

On the Kinect camera, the accuration of depth image can be improve by applying the optimal triangulation to get the new corresponding point before reconstructed by DLT. This depth image is similar with reconstruction image. Hopefully, the mapping of reconstruction point is more accurate. It means, the mapping of real point to the pixel point or vice versa is closer to the reality for the more loosely distance between Kinect camera and the real object.

Table 3. Interaction response rate

pose	feature point detection accuration		reconstruction accuration	
	pixel	subpixel	DLT	optimal triangulation
1	006	002	248	242
2	002	008	070	066
3	000	000	008	008
subtotal	008	010	326	316
total	018		642	
percentage	03		97	

4. CONCLUSION

The conclusion of this study:

a. The fundamental matrix has a high stability when applied the optimal triangulation. It occurs on both Logitech and Kinect cameras. On the Logitech camera, the fundamental matrix is more stable when the set of calibration point is detected using subpixel detection, while Kinect camera is not always.

b. The optimal triangulation significantly affects the reconstruction accuration. It occurs on both Logitech and Kinect camera. On the Logitech camera, the most accurate 3D interaction is obtained from the combination of subpixel detection and optimal triangulation, while on the Kinect camera, the accuration from the depth image can be improved by just applied the optimal triangulation.

c. This research evidences that the model of 3D based on interposition can be applied for AR environment. The improvement of reprojection point is performed before reconstructed to be 3D point using optimal triangulation, either active or passive stereo and thus resulting the more precise reconstruction point as the element of the interposition generator on a loosely scale.

REFERENCES

[1] Abdel-Aziz, Y.I. and Karara, H.M. (1971), "Direct Linear Transformation from Comparator Coordinates into Object Space Coordinates in Close-Range Photogrammetry", *Proceedings of the ASP Symposium on Close-Range Photogrammetry*; page 1-18.

[2] Chang, F., Chen, C.J, and Lu, C.J. (2004), “A Linear Time Component Labeling Algorithm Using Contour Tracing Technique”, *Computer Vision and Image Understanding;* 93 (2); page 206–220.

[3] Ericson, C. (2005), *Real-Time Collision Detection*, Morgan Kaufmann Publishers.

[4] Hartley, R.I. (1997), “In Defense of the Eight Point Algorithms”, *IEEE Transactions on Pattern Analysis dan Machine Intelligence;* 19 (6); 580-593.

[5] Hartley, R.I, and Zisserman, A (2003), *Multiple View Geometry in Computer Vision*, 2rd edition, Cambridge University Press, United Kingdom.

[6] Lee, M., Green, R. and Billinghurst, M. (2008), “3D Natural Hand Interaction for AR Applications, Image and Vision Computing”, *23rd International Conference,* New Zealand; page 1-6.

[7] Luong, Q.T. and Faugeras, O.D. (1996), "The Fundamental Matrix: Theory Algorithms, and Stability Analysis", *The International Journal of Computer Vision;* 17 (1); 43-75.

[8] Senior, A. (2002), "Tracking People with Probabilistic Appearance Models", *Third IEEE International Workshop on Performance Evaluation of Tracking and Surveillance*, page 48-55.

[9] Smisek, Jan and Jancosek, Michal and Pajdla, Tomas. (2011), “3D With Kinect”, *IEEE International Converence on Computer Vision Workshop.*