

REKONSTRUKSI OBYEK TIGA DIMENSI DARI CITRA DUA DIMENSI MENGUNAKAN EPIPOLAR GEOMETRY

Rachmawati, Risanuri Hidayat, Sunu Wibirama
Jurusan Teknik Elektro, Fakultas Teknik
Universitas Gajah Mada
Jln. Grafika 2 Yogyakarta 55281 INDONESIA
E-mail : rachma_sy@yahoo.com

INTISARI

Teknik rekonstruksi objek tiga dimensi terbagi menjadi 2 kategori, yaitu metode aktif: akuisisi model 3D dapat dilakukan dengan laser scanner atau cahaya terstruktur. Metode pasif: rekonstruksi model 3D dengan sekumpulan gambar dari sebuah atau beberapa kamera. Object scanning terkadang membutuhkan peralatan yang mahal dan keahlian khusus untuk mengoperasikan. Sebagai pendekatan yang sederhana dan murah, rekonstruksi objek berdasarkan gambar menjadi lebih populer bagi para peneliti.

Paper ini menjelaskan rekonstruksi obyek tiga dimensi (3D) dari kumpulan gambar citra dua dimensi (2D) yang diambil menggunakan sebuah kamera dari posisi berbeda. Tahapan rekonstruksi dapat dibagi menjadi empat bagian: pertama, ekstraksi titik fitur, kedua menghitung fundamental matrik dari titik – titik yang berkoresponden, ketiga menghitung matrik kamera dan terakhir mendapatkan titik-titik 3D untuk semua pasangan titik citra. Tujuan dari penelitian ini adalah merekonstruksi sebuah obyek 3D menggunakan citra 2D yang diperoleh dengan sebuah kamera dari sudut pandang yang berbeda. Sebagai obyek penelitian adalah sebuah miniatur candi Borobudur. Rekonstruksi koordinat 3D objek candi Borobudur dilakukan dengan algoritma DLT, dan dari hasil rekonstruksi didapat error reproyeksi sebesar 8.864×10^{-06} piksel.

Kata kunci : Rekonstruksi 3D, citra, matriks fundamental, epipolar geometry

ABSTRACT

There are two classes of 3D object reconstruction : active method, 3D model acquisition can be performed by laser scanner or structured light. Passive methods, reconstruct 3D models by image sequences from a single camera or multiple cameras. Object scanning often demand expensive equipment and special skill to operate. As a simple and low cost approach, 3D reconstruction based images became more popular among the researches.

This paper address the problem of 3D reconstruction from 2D images. The reconstruction process can divided into four part: first, feature point extraction. Second estimate the fundamental matrix from point correspondences, third compute the camera matrix, and compute 3D point from image points. The goal of this research is to reconstruct an 3D object using 2D images that captures from different viewpoint with a camera. As objects of experiment used a miniatur of Borobudur temple. Experiment result show average reprojection error is $8,864 \times 10^{-06}$ pixels.

Keywords : 3D Reconstruction, image, fundamental matrix, epipolar geometry

PENDAHULUAN

Seiring berkembangnya teknologi di bidang Teknologi Informasi dan Komputer, teknologi tiga dimensi (3D) juga ikut berkembang pesat. Kebutuhan untuk memvisualisasikan objek 3D banyak digunakan dalam aplikasi animasi dan grafis, arsitektur, pendidikan dan pengenalan budaya, maupun *virtual reality*. Pemodelan 3D dari bangunan bersejarah pun sudah menjadi perhatian dalam beberapa dekade terakhir. Rekonstruksi 3D dapat merupakan dokumentasi yang akurat untuk rekonstruksi

atau restorasi jika bangunan tersebut hancur, dan bisa dijadikan sumber untuk pendidikan bagi pelajar sejarah dan budaya serta para peneliti. Banyak aplikasi membutuhkan keakuratan aspek geometri, detail dan kualitas tampilan yang tinggi.

Teknik rekonstruksi objek tiga dimensi terbagi menjadi 2 kategori, yaitu teknik aktif dan teknik pasif. Teknik aktif (*object scanning*) memerlukan kendali pada cahaya yang terstruktur. Beberapa peneliti menggunakan sebuah kamera dan sebuah proyektor atau viewer untuk menghasilkan cahaya terstruktur

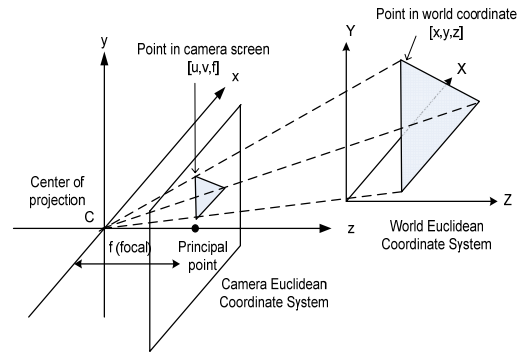
(Dipanda,2005). Peneliti lain menggunakan laser beam dan sebuah kamera video. Teknik pasif dilakukan dengan mengambil menggunakan dua atau lebih citra dari sebuah obyek dari berbagai macam posisi dengan kamera (Prakoonwit,2007). Teknik ini sering dikenal dengan adopsi *photogrammetry* atau *structure from motion*. *Object scanning* terkadang membutuhkan peralatan yang mahal dan keahlian khusus untuk mengoperasikan. Sebagai pendekatan yang sederhana dan murah, rekonstruksi objek berdasarkan gambar menjadi lebih populer bagi para peneliti (Zhang, 2010). Pada awalnya penelitian di bidang ini terfokus pada kamera terkalibrasi, namun pada perkembangan selanjutnya rekonstruksi 3D dapat juga dilakukan dengan kamera tidak terkalibrasi. Parameter kamera ditentukan dari korespondensi antar citra yang berbeda dalam satu urutan. Pada penelitian ini akan diimplementasikan teknik rekonstruksi 3D pasif menggunakan perangkat yang sederhana. Sebagai input adalah citra 2D dari objek yang diambil dengan menggunakan sebuah kamera dari sudut pandang yang berbeda, proses rekonstruksi dilakukan dengan menggunakan karakteristik geometri dari pasangan citra tersebut.

TINJAUAN PUSTAKA
Kamera Model

3D vision bertujuan untuk mendapatkan informasi 3D dari tampilan 2D, yang memiliki kesulitan dalam hal geometris dan radiometri.[Sonka,2008]. Masalah geometris berkaitan dengan citra tunggal yang tidak menyediakan informasi yang cukup tentang struktur 3D dan masalah radiometric berkaitan dengan kompleksitas proses dari pembentukan intensitas citra. Secara garis besar 3D vision membahas bagaimana manusia mempersepsikan objek 3D pada sistem pengindraannya. Persepsi ini sangat penting, terutama terkait dengan memperkirakan jarak antara objek 3D dengan manusia. Dengan memahami cara kerja penginderaan manusia, algoritma rekonstruksi visual objek 3D dari citra 2D dapat dikembangkan.

Sebagian besar citra 2D adalah hasil proyeksi adegan 3D (*3D scene*) yang dihasilkan oleh sistem kamera. Proses proyeksi adegan 3D ke citra 2D ini didapatkan dengan mengolah persamaan transformasi perspektif (*perspective transformation*). Model geometri kamera yang paling umum dan sederhana adalah model kamera *pinhole*

untuk menangani persamaan transformasi perspektif. Pinhole adalah sebuah bidang pencitraan dengan lubang kecil di tengah yang menahan semua sinar kecuali yang melewati lubang lensa di tengah.



Gambar 1. Transformasi Perspektif 3D ke 2D

Sebuah titik yang terletak di dalam ruang nyata 3D diekspresikan sebagai titik $[x \ y \ z \ 1]^T$ di sistem koordinat dunia. Titik $[x \ y \ z \ 1]^T$ yang direkam dengan sebuah kamera akan mengalami transformasi perspektif dari ruang 3D sistem koordinat dunia ke ruang 2D sistem koordinat kamera sebagaimana ditunjukkan Gambar 1. Titik $[u \ v \ w]^T$ pada sistem koordinat kamera diperoleh dengan persamaan koordinat homogen berikut: (Wibirama,2011)

$$\begin{bmatrix} u \\ v \\ w \end{bmatrix} = K [I_3 | 0_3] \begin{bmatrix} R & -T \\ 0_3^T & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \quad (1)$$

Atau

$$\begin{bmatrix} u \\ v \\ w \end{bmatrix} = M \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix},$$

dimana $M = [KR \ | \ -KRT]$ (2)

Matriks K dan $[I_3 \ | \ 0_3]$ pada persamaan (1) adalah parameter intrinsik yang mengandung variabel-variabel fisik pada kamera, seperti fokus dan *principal point*. Matriks K adalah matriks kamera kalibrasi yang didefinisikan sebagai berikut:

$$K = \begin{bmatrix} fa & 0 & u_0 \\ 0 & fb & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (3)$$

Dimana f adalah panjang fokus pada proyeksi kamera, a dan b adalah faktor konversi ukuran citra dari unit fisik (misal: cm atau mm) ke unit pixel pada - sumbu x dan y , (u_0, v_0) adalah *principal point* pada proyeksi kamera.

Matriks $[I_3 | O_3]$ didefinisikan sebagai :

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \text{Matriks} \begin{bmatrix} R & -T \\ 0_3^T & 1 \end{bmatrix} \text{ adalah}$$

parameter ekstrinsik

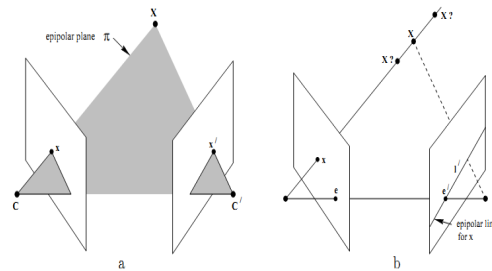
kamera yang mendefinisikan lokasi dan orientasi kamera terhadap sistem koordinat dunia. Matriks translasi T berisi tiga buah elemen translasi dari pusat sistem koordinat dunia terhadap sistem koordinat kamera. Matriks M adalah gabungan dari parameter intrinsik dan ekstrinsik kamera yang bias didapatkan dengan cara melakukan identifikasi titik-titik $[x \ y \ z \ 1]^T$ pada koordinat dunia dan identifikasi titik-titik $[u \ v \ w]^T$ yang akan menghasilkan persamaan :

$$\begin{bmatrix} x & y & z & 1 & 0 & 0 & 0 & 0 & -ux & -uy & -uz & 1 \\ 0 & 0 & 0 & 0 & x & y & z & 1 & -vx & -vy & -vz & 1 \end{bmatrix} \begin{bmatrix} m_{11} \\ m_{12} \\ \vdots \\ m_{14} \end{bmatrix} = 0 \quad (4)$$

Apabila terdapat n buah titik pada sistem koordinat dunia, matriks A akan berukuran $2n \times 12$. Untuk memperoleh X , digunakan metode *Singular Value Decomposition* (SVD) pada matriks A sehingga diperoleh $A = UDV^T$. Kolom terakhir pada V adalah solusi dari M .

Epipolar Geometry

Dalam rekonstruksi 3D *Epipolar geometry* merupakan teknik yang digunakan untuk mendapatkan persesuaian antara dua citra. Jika dua kamera melihat sebuah *scene*, maka terdapat geometri yang menghubungkan antara titik-titik 3D yang diamati dengan 2D proyeksinya (Hartley,2003). Jika X adalah sebuah titik 3D dicitrakan dalam dua sudut pandang, pada citra pertama x dan citra kedua x' . Titik citra x dan x' , titik ruang X , dan pusat kamera C dan C' adalah sebidang (*coplanar*), dan semua terletak pada bidang epipolar π seperti diperlihatkan pada Gambar 2a.



Gambar 2. Epipolar Geometry

Semua bidang yang berhimpit dengan *baseline* yang menghubungkan dua pusat kamera adalah bidang epipolar. Jika hanya diketahui sebuah titik x dalam citra pertama, maka bagaimana titik yang bersesuaian x' ditentukan dalam citra kedua. Bidang epipolar π ditentukan oleh *baseline* dan sinar yang memotong pusat kamera C yaitu x . Perpotongan antara π dan bidang citra kedua membentuk sebuah garis l' dimana x' seharusnya terletak. Garis l' adalah garis epipolar berhubungan dengan x , seperti diilustrasikan pada Gambar 2.b. Epipolar dapat digunakan agar memudahkan untuk melakukan pencocokkan untuk rekonstruksi citra, wilayah pencarian akan disempitkan pada garis epipolar.

Matriks Fundamental

Matriks fundamental F adalah sebuah matriks 3×3 yang merepresentasikan hubungan geometri tersebut. Algoritma 8-titik pertama dikembangkan oleh Longuet-Higgins. Merupakan metode yang paling sering digunakan untuk menghitung fundamental matriks dari delapan atau lebih titik-titik yang berkoresponden. Algoritma ini memiliki keuntungan sederhana dan mudah diimplementasikan. Jika $x = [x \ y \ 1]^T$ dan $x' = [x' \ y' \ 1]^T$, setiap titik terdapat dalam sebuah persamaan linier dalam F yang tidak diketahui. Persamaan ini dapat ditulis sebagai:

$$[x' \ y' \ 1] \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} = 0 \quad (5)$$

Solusi f diperoleh dengan menggunakan SVD dari $A : A = UDV^T$ dan f merupakan kolom terakhir dari V . Untuk mengurangi pengaruh noise, koordinat dari titik-titik yang berkoresponden dinormalisasi. Titik titik yang berkoresponden di translasi dan skala, sehingga centroidnya terdapat di tengah

koordinat dan jarak rata-rata titik dari origin sama dengan $\sqrt{2}$. Nilai fundamental matrik didapat dengan melakukan denormalisasi kembali dari hasil yang diperoleh.

Matrik essential

Matrik essential menyatakan relasi antara dua citra dengan menggunakan hubungan matrik rotasi dan vektor translasi. Melalui matrik essential diketahui posisi relatif antar kamera.

$$E = K^T F K \quad (6)$$

Dimana: E = Matrik essential

K = Matriks instrinsik kamera 1

K' = Matriks instrinsik kamera 2

F = Matriks fundamental

Dari perhitungan matrik essential dapat diperoleh posisi relatif antara dua kamera dengan skala yang tidak diketahui. Dengan matrik kamera

$M = [I \ 0]$ dan $M' = [R \ t]$, posisi relatif dinyatakan dengan matrik rotasi R dan vektor translasi t. SVD dari matrik essential dapat ditulis sebagai $E = U \text{diag}(1, 1, 0) V^T$, dengan U dan V dipilih $\det(U) > 0$ dan $\det(V) > 0$. Vektor translasi t sama dengan $+u_3$ atau

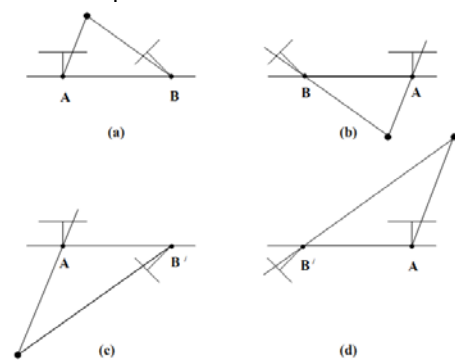
$-u_3$, dan $\|t\| = 1$. Dengan menggunakan matrik orthonormal W, matrik rotasi R adalah sama dengan $R_a = U W V^T$ atau $R_b = U W^T V^T$.

$$W = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (7)$$

Maka dari matrik essential yang diketahui dapat menghasilkan empat kemungkinan solusi untuk marik kamera kedua P', yaitu:

$$M' = [U W V^T | +u_3], \text{ atau } [U W V^T | -u_3], \\ \text{atau } [U W^T V^T | +u_3] \text{ atau } [U W^T V^T | -u_3].$$

Keempat solusi tersebut diilustrasikan pada Gambar (3), dapat dilihat bahwa hanya satu solusi yang terbaik yaitu (a) yang menghasilkan titik X yang direkonstruksi berada di depan kedua kamera



Gambar 3. Ilustrasi empat kemungkinan solusi

matrik kamera.

Ekstraksi Koordinat 3D

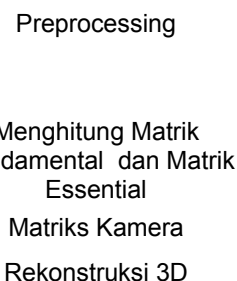
Direct Linear Transformation (DLT) adalah sebuah algoritma untuk memperkirakan koordinat 3D dari sebuah obyek yang diperoleh dengan menghitung data koordinat 2D pada beberapa gambar yang diambil pada berbagai posisi di sekeliling obyek tersebut. Proses kalibrasi dapat dilakukan untuk mendapatkan parameter instrinsik dan ekstrinsik kamera. Parameter untuk kamera 1 disimpan pada matriks M dan parameter untuk kamera 2 disimpan pada matriks M'. Jika koordinat (u', v') dan (u'', v'') adalah koordinat 2D yang diperoleh dari proses deteksi titik, maka untuk menghitung koordinat 3D obyek, persamaan homogen di bawah ini digunakan:

$$\begin{bmatrix} u' m_3^{i'T} - w' m_1^{i'T} \\ v' m_3^{i'T} - w' m_2^{i'T} \\ u'' m_3^{i'T} - w'' m_1^{i'T} \\ v'' m_3^{i'T} - w'' m_2^{i'T} \end{bmatrix} X = AX = 0 \quad (8)$$

Dimana m_i^T adalah baris ke-i pada matriks M dan X adalah matriks $[x \ y \ z \ 1]^T$. Hasil dari proses ini kemudian dimodelkan pada sebuah koordinat 3D.

PEMBAHASAN

Secara garis besar, diagram alur proses rekonstruksi 3D pada penelitian ini dapat dilihat pada Gambar 4

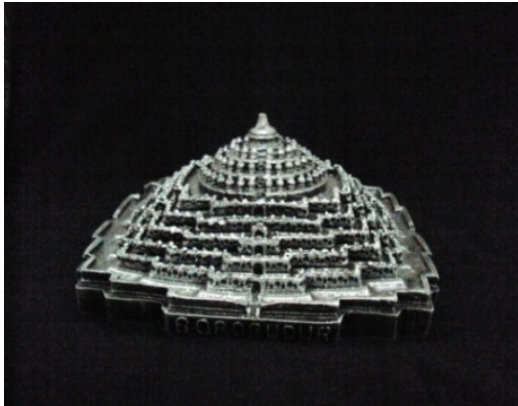


Gambar 4. Bagan alur penelitian

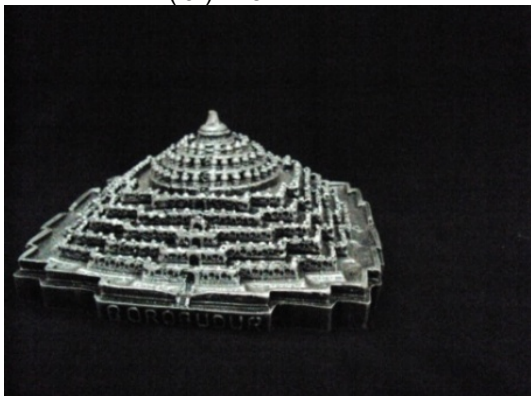
Pada tahap *preprocessing* dilakukan proses pengambilan citra objek menggunakan sebuah kamera yang sama dari dua posisi yang berbeda, sedangkan posisi objek tidak berubah, dan pada saat pengambilan citra tidak menggunakan zooming. Sebagai bahan penelitian digunakan dua buah citra dari miniatur Candi borobudur dengan tipe JPEG dan memiliki resolusi 3648 x 2736 piksel.

Gambar 5 menunjukkan hasil pengambilan citra dari dua posisi.

Selanjutnya dilakukan pemilihan titik-titik yang bersesuaian untuk setiap pasangan citra, yang akan digunakan untuk input pada proses rekonstruksi berikutnya. Langkah kedua perhitungan matriks fundamental dengan algoritma delapan titik, yang menyatakan korespondensi antara citra pada view pertama dengan citra pada view kedua.



(a) View 1



(b) View 2

Gambar 5. Citra Borobudur dari dua posisi

Langkah ketiga menghitung Matrik essential, karena menggunakan satu kamera yang sama, maka $K_1 = K_2$. Matriks proyeksi kamera diperoleh dari dekomposisi matrik essential. Hanya satu dari empat kemungkinan solusi tersebut yang terbaik, dilakukan pengujian dengan triangulasi untuk menentukan pasangan M dan M' yang menghasilkan titik 3D hasil rekonstruksi yang terletak di depan kedua kamera

Setelah diperoleh data koordinat titik-titik 2D x dan x' , serta matrik proyeksi kamera M dan M', maka dapat dilakukan rekonstruksi koordinat 3D menggunakan triangulasi *Direct Linear Transformation* (DLT). Merupakan

sebuah algoritma untuk memperkirakan koordinat 3D dari sebuah obyek dengan menghitung data koordinat 2D pada beberapa citra yang diambil dari berbagai posisi di sekeliling obyek tersebut.

ANALISA

Penelitian dilakukan menggunakan 2 citra dari objek miniatur candi yang diambil dari dua posisi berbeda menggunakan kamera digital Sony Cybershot DCS-W180 yang mempunyai focus 6.2 mm. Kedua citra harus mempunyai bagian obyek yang bersesuaian satu dengan yang lain. Selanjutnya pemilihan titik-titik yang akan direkonstruksi untuk masing-masing citra dilakukan secara manual, proses ini untuk menentukan titik yang mana dari citra pertama yang berhubungan dengan titik dari citra kedua.

Kemudian dilakukan perhitungan matrik fundamental, data yang diperlukan adalah posisi koordinat citra pada view pertama yang berpasangan dengan posisi koordinat pada view kedua. Citra pertama dan citra kedua dapat berkorespondensi dengan adanya matrik fundamental kedua citra tersebut. Dari hasil perhitungan diperoleh:

Matrik Fundamental:

$$F = \begin{bmatrix} -1.58e-009 & -2.0182e-007 & 0.0002754 \\ 1.8142e-007 & -1.9374e-008 & 0.0013088 \\ -0.00025091 & -0.0012877 & 0.056258 \end{bmatrix}$$

Pengujian hasil perhitungan matrik fundamental dilakukan dengan menghitung jarak antara koordinat titik terhadap garis epipolarnya merupakan nilai residual error. Untuk citra candi diperoleh residual error 2.69×10^{-07} piksel.

Selanjutnya dihitung parameter-parameter yang dibutuhkan untuk proses rekonstruksi seperti dijelaskan pada bagian sebelumnya, maka diperoleh:

Matrik Essential:

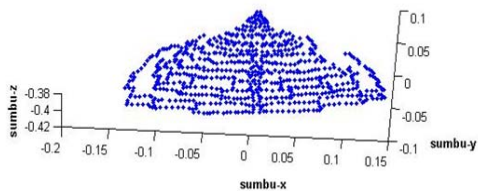
$$E = \begin{bmatrix} -0.021301 & -2.7208 & - & 0.013114 \\ 2.4458 & -0.26119 & 5.9232 & -0.020596 \\ -0.020596 & -6.177 & & 0.037376 \end{bmatrix}$$

Matrik Parameter Kamera:

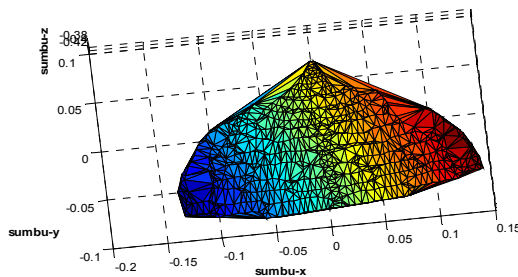
$$M = \begin{bmatrix} 3671.7 & 0 & 1824 & 0 \\ 0 & 3671.7 & 1368 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

$$M' = \begin{bmatrix} -0.99959 & 0.016622 & -0.023161 & 0.91507 \\ -0.016315 & -0.99978 & -0.013398 & 0.004571 \\ 0.023378 & 0.013014 & -0.99964 & -0.40326 \end{bmatrix}$$

Hasil rekonstruksi 3D obyek candi dapat dilihat pada Gambar 6 (a), dan kemudian dihubungkan menggunakan triangulasi Delaunay, seperti ditunjukkan Gambar 6(b). Pengujian hasil rekonstruksi dilakukan dengan mereproyeksikan kembali koordinat 3D yang diperoleh ke gambar 2D. Kemudian hasil reproyeksi tersebut dibandingkan dengan nilai koordinat (x,y) yang diperoleh di awal dari hasil inialisasi titik input yang dianggap sebagai hasil pengukuran. Selisih antara koordinat hasil pengukuran dan perhitungan merupakan error reproyeksi.



(a)



(b)

Gambar 6. Hasil Rekonstruksi 3D

KESIMPULAN

Dari hasil penelitian diperoleh beberapa kesimpulan yaitu:

1. Rekonstruksi objek 3D dari citra 2D dapat dilakukan dengan beberapa tahap yaitu : Ekstraksi koordinat citra, menghitung matrik fundamental, menghitung matrik kamera, dan rekonstruksi titik 3D untuk setiap pasangan titik-titik dari citra 2D.
2. Korespondensi antara koordinat citra dari view yang berbeda dapat ditentukan dengan matrik fundamental yang dihitung menggunakan algoritma delapan titik.
3. Dari hasil rekonstruksi didapat error reproyeksi sebesar 8.864×10^{-06} Piksel.

DAFTAR PUSTAKA

- Dipanda. A and S. Woo, "Towards a real-time 3D shape reconstruction using a structured light system," *Pattern Recognition*, vol. 38, pp. 1632-1650, 2005.
- Dianyong Zhang, Zhenjiang Miao, "Photorealistic 3D Volumetric Model Reconstruction By Voxel Coloring", *IAPRS vol.XXXVIII, Part 3B*, 2010
- Hartley. A and A. Zisserman, *Multiple View Geometry in Computer Vision*, Second Edition, Cambridge University Press, 2003.
- Prakoonwit.S and R. Benjamin, "3D surface point and wireframe reconstruction from multiview photographic images," *Image and Vision Computing*, vol. 25, pp. 1509-1518,2007.
- Park.J.S "Interactive 3D reconstruction from multiple images: A primitive-based approach," *Pattern Recognition Letters*, vol. 26, pp. 2558-2571, 2005.
- Sonka.M, V. Hlavac, and R. Boyle, "3D Vision," in *Image Processing, Analysis, and Machine Vision*, H. Gowans, Ed., 3rd ed. Toronto: Thomson, pp. 592-594, 2008.
- Wibirama.S "Fundamental Techniques for 3D Computer Vision," Department of Electrical Engineering Gadjah Mada University, Indonesia, 2011.