UNIVERSITAS GADJAH MADA
Persepsi Multi-Modal Berbasis Deep Learning dengan Representasi Birdâ€™s-Eye View sebagai Dasar
Kendali Autonomous Excavator
Muhammad Aditya Rahmansyah Baskoro, Ahmad Ataka Awwalur Rizqi, S.T., Ph.D.; Dr.Eng. Ir. Igi Ardiyanto, S.T.,M.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

DAFTAR ISI

HALAMAN PENGESAHAN ii
PERNYATAAN BEBAS PLAGIASI iii
HALAMAN PERSEMBAHAN iv
KATA PENGANTAR v
DAFTAR ISI vi
DAFTAR TABEL xi
DAFTAR GAMBAR xiii
DAFTAR SINGKATAN xvi
INTISARI xvii
ABSTRACT xviii
BAB I Pendahuluan 1
1.1 Latar Belakang 1
1.2 Rumusan Masalah 2
1.3 Tujuan Penelitian 3
1.4 Batasan Masalah 3
1.5 Manfaat Penelitian 3
1.6 Sistematika Penulisan 4
BAB II Tinjauan Pustaka dan Dasar Teori 5
2.1 Tinjauan Pustaka 5
2.1.1 Otomasi Ekskavator dan Manipulasi Alat Berat 5
2.1.2 Persepsi Robotik pada Lingkungan Konstruksi 6
2.1.3 Persepsi 3D Multi-Modal pada Kendaraan Otonom 7
2.1.4 Persepsi yang Berorientasi pada Tugas 8
2.1.5 Analisis Kesenjangan dan Posisi Penelitian 8
2.2 Dasar Teori 9
2.2.1 Persepsi Robotik 9
2.2.1.1 Keadaan, observasi, dan estimasi 9
2.2.1.2 Ketidakpastian pada pengukuran sensor 10
2.2.1.3 Kebutuhan keadaan untuk manipulasi 10
2.2.2 Kinematika Robot Manipulator 11
2.2.2.1 Rantai kinematik dan kinematika maju 11
2.2.2.2 Kinematika balik dan keterjangkauan 12
2.2.2.3 Error target dan titik ujung 12
2.2.3 Representasi Data Sensor 13
2.2.3.1 Citra RGB 13
2.2.3.2 Peta kedalaman 13

UNIVERSITAS GADJAH MADA
Persepsi Multi-Modal Berbasis Deep Learning dengan Representasi Birdâ€™s-Eye View sebagai Dasar Kendali Autonomous Excavator
Muhammad Aditya Rahmansyah Baskoro, Ahmad Ataka Awwalur Rizqi, S.T., Ph.D.; Dr.Eng. Ir. Igi Ardiyanto, S.T.,M.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

2.2.3.3 *Point cloud* 13
2.2.4 Geometri Sensor dan Kalibrasi 14
2.2.4.1 Kerangka koordinat 14
2.2.4.2 Model proyeksi kamera 14
2.2.4.3 Model pengukuran LiDAR 15
2.2.4.4 Kalibrasi multi-sensor 15
2.2.5 Observabilitas Target 16
2.2.5.1 Visibilitas dan observabilitas 16
2.2.5.2 Dukungan sensor terhadap pusat objek 17
2.2.6 Representasi Spasial 18
2.2.6.1 Representasi perspektif 18
2.2.6.2 Representasi titik 3D 18
2.2.6.3 Grid metrik 19
2.2.6.4 *Bird's-Eye View* 19
2.2.7 Pembelajaran Representasi Sensor 19
2.2.7.1 Fitur visual berbasis CNN 19
2.2.7.2 Encoding *point cloud*: titik, voxel, dan pilar 20
2.2.7.3 Pembelajaran kedalaman dan proyeksi fitur 20
2.2.8 Fusi Sensor Multi-Modal 21
2.2.8.1 Tujuan fusi sensor 21
2.2.8.2 Tingkat fusi: data, fitur, dan keputusan 21
2.2.8.3 Fusi adaptif dan batasnya 22
2.2.9 Deteksi Objek 3D 23
2.2.9.1 Kelas dan lokalisasi objek 23
2.2.9.2 Deteksi berbasis pusat objek 23
2.2.9.3 *Heatmap*, offset, dan asumsi pusat 24
2.2.9.4 Target dan non-target mirip target 24
2.2.10 Evaluasi Persepsi 25
2.2.10.1 Presisi, *recall*, dan AP 25
2.2.10.2 Error translasi dan distribusi error 25
2.2.10.3 Metrik berbasis toleransi tugas 26
2.2.11 Simulasi dan *Domain Randomization* 26
2.3 Analisis Perbandingan Metode 27
2.3.1 Alternatif Representasi 27
2.3.2 Alternatif Modalitas 27
2.3.3 Alternatif Kepala Deteksi 27
2.3.4 Alternatif Penghubung Persepsi dan Kendali 27
2.3.5 Ringkasan Pilihan Metode 28
BAB III Metode Penelitian 30

Persepsi Multi-Modal Berbasis Deep Learning dengan Representasi Birdâ€™s-Eye View sebagai Dasar
Kendali Autonomous Excavator
Muhammad Aditya Rahmansyah Baskoro, Ahmad Ataka Awwalur Rizqi, S.T., Ph.D.; Dr.Eng. Ir. Igi Ardiyanto, S.T.,M.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

3.1 Rancangan Penelitian 30
3.2 Definisi Tugas dan Kebutuhan Informasi 31
3.3 Lingkungan Simulasi dan Rancangan Sensor 33
3.3.1 Platform Simulasi 33
3.3.2 Model Ekskavator 33
3.3.3 Konfigurasi Sensor 34
3.3.4 Generasi Lingkungan dan *Domain Randomization* 34
3.4 Rancangan Dataset dan Akuisisi Data 37
3.4.1 *Pipeline* Koleksi Data 37
3.4.2 Praproses Sensor dan Transformasi Koordinat 38
3.4.3 Format Data, Anotasi, dan Metadata Visibilitas 38
3.4.4 Pembagian Dataset dan Peran Split 39
3.4.5 Karakteristik Dataset Gabungan 40
3.5 Rancangan Arsitektur Persepsi 41
3.5.1 Dari Data Sensor Mentah ke Representasi Bersama 41
3.5.2 Ruang BEV dan Kontrak Keluaran 42
3.5.3 Cabang Kamera: EfficientNet-B0 dan LSS 43
3.5.3.1 Dimensi Tensor Cabang Kamera 43
3.5.4 Cabang LiDAR: PointPillars 44
3.5.4.1 Dimensi Tensor Cabang LiDAR 44
3.5.5 Fusi BEV dengan Gerbang Adaptif 45
3.5.5.1 Dimensi Tensor Fusi 45
3.5.6 Kepala Deteksi Berbasis CenterPoint 46
3.5.6.1 Dimensi Tensor Kepala Deteksi 46
3.5.7 Varian Model dan Ablasi 46
3.6 Strategi Pelatihan 47
3.6.1 Fungsi *Loss* 49
3.7 Protokol Evaluasi Persepsi 49
3.7.1 Pencocokan Prediksi dan *Ground Truth* 49
3.7.2 Metrik Deteksi 3D 50
3.7.3 Metrik Berbasis Toleransi Tugas 50
3.7.4 Analisis Terstratifikasi 51
3.7.5 Pembanding Geometri Mentah 51
3.7.6 Evaluasi Non-Target Mirip Target dan Ketahanan Sensor 51
3.8 Protokol Evaluasi terhadap Kendali 52
3.8.1 Kinematika Balik Analitik 52
3.8.1.1 Model kinematik dan kalibrasi 52
3.8.1.2 Penyelesaian kinematika balik 53
3.8.1.3 Target replay dan batas penggunaan 54

Persepsi Multi-Modal Berbasis Deep Learning dengan Representasi Birdâ€™s-Eye View sebagai Dasar
Kendali Autonomous Excavator
Muhammad Aditya Rahmansyah Baskoro, Ahmad Ataka Awwalur Rizqi, S.T., Ph.D.; Dr.Eng. Ir. Igi Ardiyanto, S.T.,M.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

3.8.2 *Replay* Scene Berbasis Keluaran Persepsi 55
BAB IV Hasil dan Pembahasan 57
4.1 Konteks Evaluasi dan Data Uji 57
4.2 Hasil Utama Deteksi dan Lokalisasi 59
4.3 Peran Pembelajaran dalam Estimasi Pusat 59
4.4 Kontribusi Fusi Kamera-LiDAR 60
4.5 Ketergantungan Sensor dan Batas Operasi 61
4.6 Sintesis Hasil Persepsi 63
BAB V Evaluasi Persepsi terhadap Kendali 65
5.1 Pertanyaan Evaluasi dan Kontrak Uji 65
5.2 Hasil Utama Pemilihan Pasangan Target Berbasis Skor 65
5.3 Mengapa Fusi Menambah Pasangan yang Layak 66
5.4 Contoh Eksekusi Gazebo Berbasis Keluaran Persepsi 67
5.5 Makna Operasional dan Batas Klaim 69
BAB VI Kesimpulan dan Saran 70
6.1 Kesimpulan 70
6.2 Saran 71
DAFTAR PUSTAKA 73
LAMPIRAN L-1
L.1 Struktur Data Episodik dan Format Anotasi L-1
L.1.1 Struktur Direktori Episode L-1
L.1.2 Format Anotasi 3D L-1
L.1.3 Kalibrasi dan Sinkronisasi Sensor L-2
L.1.4 Format *Visibility Sidecar* L-2
L.2 Probe Diagnostik Observabilitas L-3
L.2.1 Baseline Input Kosong L-3
L.2.2 Oracle Decode dan Pemisahan Tahap Kesalahan L-4
L.2.3 Pemulihan Sinyal Setelah Konfigurasi Mast L-5
L.2.4 Oracle Depth pada Objek LiDAR-Blind L-5
L.3 Detail Implementasi Kendali L-6
L.3.1 Kalibrasi Kinematika Gazebo L-6
L.3.2 Cakupan Geometris Uji Utama L-6
L.3.3 Hasil Pendukung Replay Berbasis Persepsi L-7
L.3.4 Batas Model Kontak Gazebo L-8
L.4 Tabel Hasil Lengkap L-9
L.4.1 Hasil Held-Out dengan All-Point AP L-9
L.4.2 Bootstrap Berpasangan Hasil Utama Bab 4 ..... L-9
L.4.3 Hasil Best-F1 dari Evaluator 11 Titik L-9
L.4.4 Ablasi Arsitektur L-10

UNIVERSITAS
GADJAH MADA
**Persepsi Multi-Modal Berbasis Deep Learning dengan Representasi Birdâ€™s-Eye View sebagai Dasar
Kendali Autonomous Excavator**
Muhammad Aditya Rahmansyah Baskoro, Ahmad Ataka Awwalur Rizqi, S.T., Ph.D.; Dr.Eng. Ir. Igi Ardiyanto, S.T.,M.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

L.4.5 Dust Sweep L-10
L.4.6 Recall Berdasarkan Dukungan LiDAR L-10
L.4.7 Hard-Tail dan Oracle Depth L-11
L.5 Cuplikan Kode Implementasi L-12
L.5.1 Transformasi LiDAR Mast L-12
L.5.2 Gated Fusion L-12
L.5.3 Decode CenterPoint L-13
L.5.4 Penulisan Visibility Sidecar L-14

Persepsi Multi-Modal Berbasis Deep Learning dengan Representasi Birdâ€™s-Eye View sebagai Dasar
Kendali Autonomous Excavator
Muhammad Aditya Rahmansyah Baskoro, Ahmad Ataka Awwalur Rizqi, S.T., Ph.D.; Dr.Eng. Ir. Igi Ardiyanto, S.T.,M.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

DAFTAR TABEL

Tabel 2.1 Posisi penelitian terhadap penelitian terkait 9
Tabel 2.2 Ringkasan alasan pemilihan metode 28

Tabel 3.1 Spesifikasi sensor pada konfigurasi mast 35
Tabel 3.2 Pembagian dataset dan peran evaluasi 39
Tabel 3.3 Karakteristik utama dataset gabungan 40
Tabel 3.4 Distribusi jumlah titik LiDAR di dalam kotak *ground truth* 41
Tabel 3.5 Dimensi tensor cabang kamera, dari citra mentah hingga fitur BEV . 44
Tabel 3.6 Dimensi tensor cabang LiDAR, dari *point cloud* mentah hingga fitur BEV 44
Tabel 3.7 Dimensi tensor pada tahap fusi gerbang adaptif 45
Tabel 3.8 Dimensi tensor pada kepala deteksi CenterPoint yang disederhanakan 46
Tabel 3.9 Varian model dan tujuan ablasinya 47
Tabel 3.10 Konfigurasi pelatihan model utama 48
Tabel 3.11 Istilah evaluasi kompatibilitas IK berbasis persepsi 55

Tabel 4.1 Cakupan LiDAR pada split *test_ungated* 57
Tabel 4.2 Hasil utama pada split *test_ungated* 59
Tabel 4.3 Geometri mentah dan model BEV 60
Tabel 4.4 Prediksi pada skenario non-target mirip target (245 target dan 170 *distractor*) 60
Tabel 4.5 Performa model fusi pada perubahan masukan sensor 61
Tabel 4.6 Performa model fusi berdasarkan jumlah titik LiDAR pada objek ... 62

Tabel 5.1 Hasil utama pasangan target terpilih pada held-out *test_ungated* 66

Tabel L.1 Elemen utama pada berkas anotasi. L-2
Tabel L.2 Elemen utama pada metadata sinkronisasi sensor. L-3
Tabel L.3 Elemen utama pada berkas *visibility sidecar*. L-4
Tabel L.4 Probe input kosong pada model era A12 sebagai pembanding diagnostik. L-4
Tabel L.5 Probe decode untuk memisahkan kapasitas representasi dan pemilihan pusat. L-4
Tabel L.6 Ringkasan probe pemulihan sinyal LiDAR. L-5
Tabel L.7 Batas oracle-depth untuk bin tanpa dukungan LiDAR. L-5
Tabel L.8 Ringkasan kalibrasi IK pada Gazebo. L-6
Tabel L.9 Cakupan kinematik scene pada *test_ungated*. L-7
Tabel L.10 Konsistensi hasil replay pada set uji pendukung. L-7
Tabel L.11 Statistik berpasangan hasil utama *score-only* pada 74 scene. L-8
Tabel L.12 Batas suplai dan error target terpilih pada *test_ungated*. L-8
Tabel L.13 Batas interpretasi eksperimen kendali fisika. L-8
Tabel L.14 All-point AP pada split held-out, match radius 1.0 m. L-9
Tabel L.15 Interval bootstrap berpasangan untuk perbandingan utama Bab 4. ... L-9
Tabel L.16 Hasil best-F1 dari evaluator 11 titik, match radius 1.0 m. L-10
Tabel L.17 Ablasi arsitektur pada `test_ungated`, all-point AP. L-10

UNIVERSITAS
GADJAH MADA
**Persepsi Multi-Modal Berbasis Deep Learning dengan Representasi Birdâ€™s-Eye View sebagai Dasar
Kendali Autonomous Excavator**
Muhammad Aditya Rahmansyah Baskoro, Ahmad Ataka Awwalur Rizqi, S.T., Ph.D.; Dr.Eng. Ir. Igi Ardiyanto, S.T.,M.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

Tabel L.18 Dust sweep pada `test_ungated`; AP dihitung dengan evaluator 11 titik. L-11
Tabel L.19 Bootstrap CI untuk gain fusi pada dust sweep `test_ungated`. L-11
Tabel L.20 Recall berdasarkan dukungan LiDAR pada split ungated, $t = 0.30$. L-11
Tabel L.21 Split hard-tail untuk bin tanpa titik LiDAR. L-11
Tabel L.22 LSS depth probe pada objek LiDAR-blind. L-12

Persepsi Multi-Modal Berbasis Deep Learning dengan Representasi Birdâ€™s-Eye View sebagai Dasar
Kendali Autonomous Excavator
Muhammad Aditya Rahmansyah Baskoro, Ahmad Ataka Awwalur Rizqi, S.T., Ph.D.; Dr.Eng. Ir. Igi Ardiyanto, S.T.,M.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

DAFTAR GAMBAR

Gambar 2.1 Skema rantai kinematik ekskavator. Rumah putar mengatur arah horizontal lengan, sedangkan boom, stick, dan bucket membentuk rantai sendi yang menentukan posisi titik ujung bucket terhadap kerangka dasar mesin. 6
Gambar 2.2 Perbedaan antara visibilitas visual dan observabilitas metrik. Objek dapat terlihat jelas pada citra kamera, tetapi pusat 3D-nya belum tentu teramati apabila tidak ada bukti LiDAR di dalam kotak objek. Bukti empiris atas konsep ini ditunjukkan secara terukur pada Gambar 4.3..... 17
Gambar 2.3 Perbandingan tampilan perspektif kamera dan representasi *Bird's-Eye View*. Representasi perspektif mengikuti bidang citra, sedangkan BEV memetakan objek ke bidang atas yang berhubungan langsung dengan koordinat metrik..... 18
Gambar 2.4 *Pipeline* PointPillars. Titik LiDAR dikelompokkan ke dalam pilar vertikal pada grid XY, diencode menjadi fitur pilar, lalu disebarkan kembali sebagai *pseudo-image* BEV..... 20
Gambar 2.5 *Pipeline* Lift-Splat-Shoot. Tahap *lift* membentuk fitur frustum dari fitur citra dan distribusi kedalaman. Tahap *splat* memproyeksikan fitur tersebut ke grid BEV. Tahap *shoot* menggunakan BEV untuk tugas *downstream*. 21
Gambar 3.1 Diagram arsitektur sistem persepsi multi-modal. Enam citra kamera diproses oleh EfficientNet-B0 dan LSS untuk membentuk fitur BEV kamera. *Point cloud* LiDAR diproses oleh PointPillars menjadi fitur BEV LiDAR. Kedua fitur BEV digabungkan dengan *gated fusion*, kemudian kepala deteksi berbasis CenterPoint menghasilkan pusat objek 3D. 31
Gambar 3.2 Fase tugas *pick-and-place* dan kebutuhan informasi persepsi. Persepsi dibutuhkan untuk menyediakan pusat *rubble* pada fase *reach* dan pusat *bin* pada fase *transport*..... 32
Gambar 3.3 Tata letak sensor dilihat dari atas. Enam kamera memberikan cakupan visual 360°, sedangkan LiDAR mast pada $z = 3{,}25\,\mathrm{m}$ memberi cakupan geometri 360° pada bidang horizontal. 35
Gambar 3.4 Contoh variasi lingkungan dari beberapa episode. Variasi utama meliputi jumlah objek, posisi target, dan kondisi oklusi pada ruang kerja..... 36
Gambar 3.5 Alur pengumpulan data. Setiap episode melewati tahap generasi lingkungan, stabilisasi fisika 3 detik, perekaman sensor, penyimpanan anotasi, dan pembersihan lingkungan. 37
Gambar 3.6 Diagram rinci arsitektur persepsi. Cabang kamera membentuk fitur BEV melalui EfficientNet-B0 dan LSS, cabang LiDAR membentuk fitur BEV melalui PointPillars, kemudian kedua fitur digabungkan dengan *gated fusion* sebelum didekode oleh kepala CenterPoint yang disederhanakan untuk prediksi pusat objek. 42

UNIVERSITAS
GADJAH MADA
**Persepsi Multi-Modal Berbasis Deep Learning dengan Representasi Birdâ€™s-Eye View sebagai Dasar
Kendali Autonomous Excavator**
Muhammad Aditya Rahmansyah Baskoro, Ahmad Ataka Awwalur Rizqi, S.T., Ph.D.; Dr.Eng. Ir. Igi Ardiyanto, S.T.,M.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

Gambar 4.1 Contoh frame uji dari dataset mast. Baris citra menunjukkan enam kamera egosentris, sedangkan panel BEV menunjukkan titik LiDAR, kotak *ground truth*, dan prediksi model fusi pada ambang operasi. 58
Gambar 4.2 Perubahan performa model fusi ketika salah satu modalitas dihilangkan atau didegradasi. R@0.25 perlu dibaca bersama P90 karena error bersyarat yang rendah dapat berasal dari subset prediksi yang lebih mudah. 62
Gambar 4.3 R@0.25 menurut dukungan LiDAR 63
Gambar 5.1 Kelayakan target peletakan terhadap toleransi 67
Gambar 5.2 Observasi dan target terpilih pada contoh Gazebo 68
Gambar 5.3 Urutan eksekusi Gazebo berbasis prediksi..... 68

Persepsi Multi-Modal Berbasis Deep Learning dengan Representasi Birdâ€™s-Eye View sebagai Dasar
Kendali Autonomous Excavator
Muhammad Aditya Rahmansyah Baskoro, Ahmad Ataka Awwalur Rizqi, S.T., Ph.D.; Dr.Eng. Ir. Igi Ardiyanto, S.T.,M.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

DAFTAR SINGKATAN

Simbol

$\mathbf{x}$	:	Keadaan (*state*) objek target
$\hat{\mathbf{x}}$	:	Estimasi keadaan objek
$\mathbf{z}$	:	Vektor observasi sensor
$h(\cdot)$	:	Fungsi pengamatan (*observation function*)
$\mathbf{v}$	:	Gangguan (*noise*) pengukuran
(c_x, c_y, c_z)	:	Koordinat pusat objek 3D
k_i	:	Kelas objek ke-i
d_{thresh}	:	Ambang toleransi tugas (0,25 m)
$\mathbf{R}, \mathbf{t}$	:	Matriks rotasi dan vektor translasi
$\mathbf{K}$	:	Matriks intrinsik kamera
(u, v)	:	Koordinat piksel citra
λ	:	Faktor skala kedalaman
$\mathbf{H}$	:	Jacobian pengukuran $\partial h/\partial \mathbf{x}$
$\mathcal{I}(\mathbf{x})$	:	Matriks informasi Fisher

Singkatan

UNIVERSITAS
GADJAH MADA
**Persepsi Multi-Modal Berbasis Deep Learning dengan Representasi Birdâ€™s-Eye View sebagai Dasar
Kendali Autonomous Excavator**
Muhammad Aditya Rahmansyah Baskoro, Ahmad Ataka Awwalur Rizqi, S.T., Ph.D.; Dr.Eng. Ir. Igi Ardiyanto, S.T.,M.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

AP	:	*Average Precision*
ATE	:	*Average Translation Error*
BEV	:	*Bird's-Eye View*
CNN	:	*Convolutional Neural Network*
DOF	:	*Degree of Freedom*
FK	:	*Forward Kinematics*
FOV	:	*Field of View*
FP	:	*False Positive*
GPU	:	*Graphics Processing Unit*
GT	:	*Ground Truth*
IK	:	*Inverse Kinematics*
LiDAR	:	*Light Detection and Ranging*
LSS	:	*Lift-Splat-Shoot*
mAP	:	*mean Average Precision*
mATE	:	*mean Average Translation Error*
NMS	:	*Non-Maximum Suppression*
PID	:	*Proportional-Integral-Derivative*
RGB	:	*Red Green Blue*
RL	:	*Reinforcement Learning*
RMS	:	*Root Mean Square*
ROS	:	*Robot Operating System*
TP	:	*True Positive*
URDF	:	*Unified Robot Description Format*