


Ground Control Station Berbasis Web menggunakan ROS dan Flask dengan Multimodal LLM untuk Analisis Citra Udara

Maulana Al Iqbal Widodo¹, Dedi Gunawan²

¹Universitas Muhammadiyah Surakarta, Jl. A. Yani, Mendungan, Pabelan, Kec. Kartasura, Kabupaten Sukoharjo, Jawa Tengah, Indonesia

² Universitas Muhammadiyah Surakarta, Jl. A. Yani, Mendungan, Pabelan, Kec. Kartasura, Kabupaten Sukoharjo, Jawa Tengah, Indonesia

 Email korespondensi: L200210136@student.ums.ac.id

Abstrak. Perkembangan teknologi drone dan kecerdasan buatan (AI) membuka peluang besar dalam pemanfaatan teknologi di berbagai sektor termasuk pengumpulan dan analisis data citra udara. Penelitian ini mengusulkan sistem monitoring wahana dan analisis citra berbasis *Multimodal Large Language Model* (LLM) yang terintegrasi dengan wahana menggunakan *firmware ArduPilot* dan MAVROS sebagai jembatan komunikasi data wahana ke sistem *Robot Operating System* (ROS). Sistem dikembangkan dengan mengusung dua sub sistem utama yaitu monitoring wahana melalui *Ground Control Station* (GCS) berbasis web yang dibangun dengan *framework Flask*, dan sistem analisis citra yang menerima gambar dari *Raspberry Pi* dan modul kamera yang dihubungkan ke GCS melalui jaringan Wi-Fi dalam bentuk *video streaming*. Berikutnya GCS dapat menangkap gambar yang nantinya akan dianalisis oleh Multimodal AI (*Llama 4 Scout*) untuk menghasilkan deskripsi tekstual berdasarkan citra udara. Pengujian sistem dilakukan dalam skenario simulasi menggunakan *Gazebo Simulator* dan uji lapangan dengan wahana drone fisik. Diharapkan dari penelitian ini menunjukkan potensi pengembangan sistem drone cerdas berbasis AI yang nantinya dapat dimanfaatkan dalam berbagai aplikasi seperti pengawasan, pemetaan, dan lain-lain.

Kata kunci: *Aerial Image Analysis; LLM; Multimodal AI; MAVROS; ROS*

PENDAHULUAN

Pesatnya perkembangan teknologi UAV (*Unmanned Aerial Vehicle*) terutama drone telah marak dalam beberapa tahun terakhir, penggunaannya yang fleksibel di berbagai bidang seperti pemetaan wilayah, pengawasan, pertanian, penelitian riset geologi,



aplikasi militer seperti pemantauan dan pengintaian, hingga operasi penyelamatan [1]. Salah satu aspek penting dalam penggunaan drone adalah bagaimana sistem dapat mengontrol dan memantau pergerakan drone secara efektif terutama dalam misi otonomus. Solusi yang efektif untuk memenuhi kebutuhan tersebut adalah *Ground Control Station (GCS)*, yaitu sebuah sistem pusat kendali wahana UAV yang mengintegrasikan antarmuka visual, komunikasi data dan pengendalian misi wahana dalam suatu platform terpadu.

Sistem *Ground Control Station (GCS)* berbasis web dapat menjadi solusi dalam membuat sistem yang mampu mengontrol dan memantau pergerakan drone yang juga karena berbasis web dapat memberikan fleksibilitas tanpa memerlukan perangkat lunak tambahan. Ditambah dengan integrasi *Robot Operating System (ROS)* dan MAVROS, sistem pengendali wahana drone dapat dilakukan melalui komunikasi berbasis protokol MAVLink [2]. ROS adalah *framework* standar *open-source* pada industri robotika yang memungkinkan integrasi serta manipulasi sensor, komponen aktuator, navigasi robot melalui arsitektur pesan terdistribusi [3]. Sehingga penggunaan ROS dapat memberikan keleluasaan dalam menambahkan sensor-sensor yang berguna dalam navigasi wahana UAV untuk misi tertentu seperti analisis Citra Udara [4].

Namun, dalam penggunaannya salah satu tantangan utama dalam analisis data visual secara *real-time* menggunakan *video stream* yang nantinya dianalisis penuh dengan AI memerlukan *bandwidth* yang besar dan tentu saja daya komputasi tinggi. Oleh karena itu, pendekatan yang lebih efisien untuk dilakukan yaitu dengan mengambil gambar secara selektif dari citra yang dikirimkan wahana drone dan dianalisis dengan *Multimodal Large Language Model (Multimodal LLM)* seperti *Llama 4 Scout*. Kemampuan Multimodal LLM yang mampu memahami konteks gambar dengan teks tambahan tentu sangat fleksibel dalam berbagai skenario misi wahana UAV [5]. Penggunaan metode tradisional untuk deteksi objek seperti dengan YOLO atau CNN mungkin akan mengalami kesulitan dalam mendeteksi objek dengan orientasi yang berubah-ubah karena metode tersebut



menggunakan anotasi dengan kotak horizontal untuk melakukan pelabelan objek, hal ini dapat berpotensi menyebabkan menurunnya akurasi deteksi objek [6].

Penelitian ini penting dilakukan dalam rangka mengintegrasikan *Next-generation* AI dalam bidang robotika, terutama meningkatkan aplikasi wahana drone di berbagai sektor. Penelitian sebelumnya [7] telah mengembangkan wahana drone untuk misi bantuan bencana, seperti pencarian korban bencana, identifikasi rintangan pada akses penyelamatan, dan pengiriman logistik bantuan untuk tim penyelamat. Selanjutnya penerapan *Visual Language Models* (VLM) kini menjadi penting dalam meningkatkan penggunaan kecerdasan buatan pada wahana drone. VLM dapat memberikan deskripsi konten visual dengan tata bahasa yang baik dan mudah dipahami dengan mengidentifikasi pola pada konten gambar atau kumpulan gambar [8]. Tujuan dari penelitian ini untuk mengembangkan pendekatan yang lebih revolusioner terhadap perkembangan kecerdasan buatan dan robotika dengan menerapkan Multimodal AI. Aplikasinya yang tak terbatas pada deteksi objek, tetapi juga mampu memberikan analisis yang lebih deskriptif dan kontekstual pada suatu gambar, seperti memahami kondisi lingkungan, menganalisis situasi secara lebih mendalam sehingga drone dapat memberikan wawasan yang berharga untuk berbagai tugas di berbagai skenario nyata [9].

METODE

Bagian ini menjelaskan komponen utama dan kerangka kerja sistem GCS dengan implementasi Multimodal AI termasuk integrasi perangkat keras dan perangkat lunak yang mendukung operasional wahana drone. Selain itu penelitian ini juga mencakup pengujian sistem di dua lingkungan berbeda yaitu simulasi untuk validasi konsep, serta pengujian langsung di dunia nyata guna mengevaluasi sistem dalam kondisi sebenarnya. Pengujian dalam simulasi memungkinkan identifikasi potensi masalah tanpa risiko terhadap perangkat keras wahana drone sehingga dapat memastikan bahwa sistem GCS berbasis Multimodal AI dapat berfungsi dengan baik, aman, dan efisien sebelum diterapkan pada pengujian dunia nyata [10].

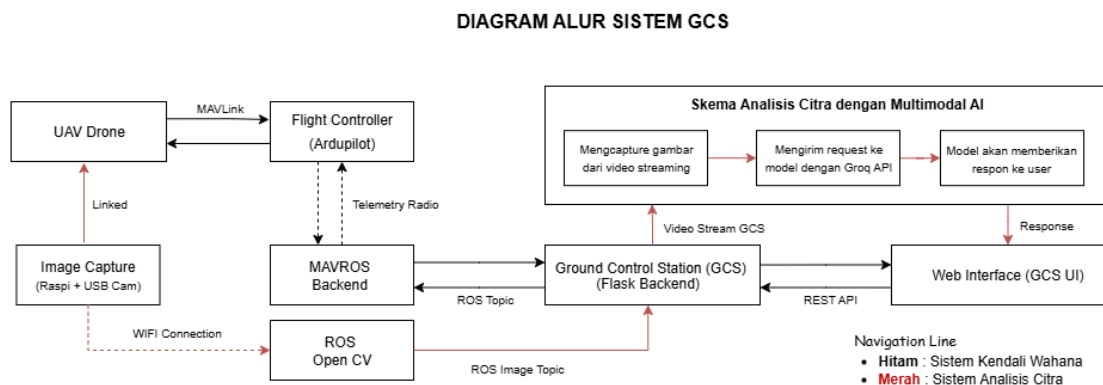


Persiapan enviroment simulasi

Untuk melakukan visualisasi dan simulasi terhadap wahana drone dapat dilakukan dengan *software gazebo simulator*. Gazebo merupakan suatu perangkat lunak yang memungkinkan simulasi robotika berbasis fisika yang dapat mengintegrasikan ROS sehingga dapat menyediakan lingkungan simulasi yang realistis dengan dukungan model fisik, sensor, dan aktuator sehingga memungkinkan peneliti menguji algoritma serta integrasi sistem tanpa menggunakan perangkat keras [11]. Dengan integrasi ROS, gazebo dapat memungkinkan komunikasi langsung antara simulasi dan sistem kontrol drone sesuai dengan kondisi yang dibutuhkan dalam penelitian.

Perancangan GCS dengan Implementasi Multimodal AI

Perancangan GCS bertujuan untuk menyediakan antarmuka kontrol dan monitoring bagi wahana drone yang dikembangkan. GCS dibangun menggunakan *framework* Flask sebagai *backend* untuk menghubungkan data wahana dengan tampilan antarmuka web melalui MAVROS, serta HTML CSS dan JavaScript untuk *frontend*. Penggunaan MAVROS sebagai penghubung antara data wahana melalui pesan-pesan *mavlink* ke ROS dengan menerima data telemetri secara *real-time*.



Gambar 1. Diagram Kerja Sistem GCS.

Berdasarkan diagram diatas, penelitian ini mengusung dua sistem utama yang bekerja secara terintegrasi dalam satu kesatuan arsitektur sistem GCS yaitu kendali wahana dan



analisis citra. Sistem kendali wahana dibangun menggunakan *firmware ArduPilot* yang kemudian dikendalikan melalui protokol *MAVLink* yang dikirim serta diterima melalui Modul Telemetri yang terpasang pada wahana (*Air*) dan GCS (*Ground*) [12]. Data telemetri seperti posisi, orientasi, dan status wahana dikirim secara dua arah sehingga memungkinkan kendali serta pemantauan langsung melalui antarmuka web yang dibangun menggunakan *framework* Flask. Data telemetri diintegrasikan ke dalam ROS menggunakan package MAVROS yang memungkinkan pesan-pesan MAVLink diubah menjadi ROS *Topic* yang selanjutnya dapat diproses dan dikontrol melalui sistem GCS berbasis web sehingga memungkinkan fleksibilitas tinggi dalam pengembangan sistem otonom dan monitoring wahana [13].

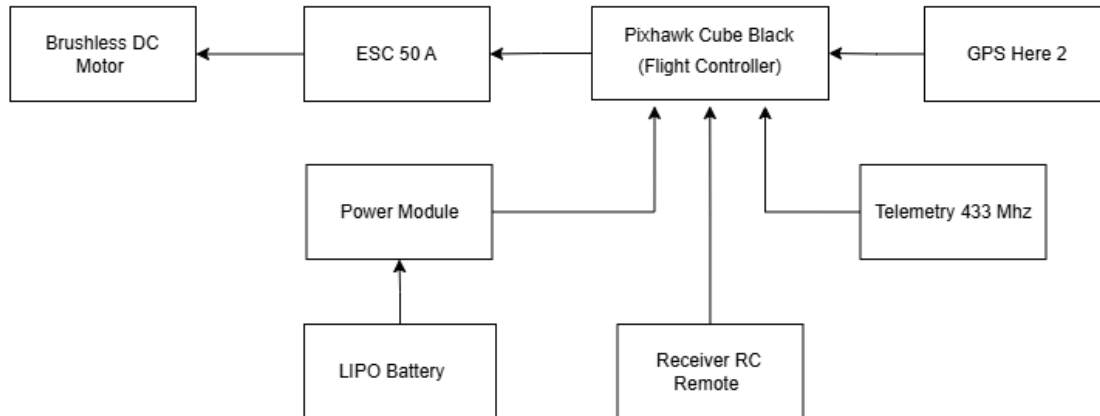
Sementara itu, sistem analisis citra dirancang terpisah dari sistem kendali drone karena perbedaan protokol pengiriman data yang dimana sistem analisis citra mentransmisikan video citra udara melalui jaringan Wi-Fi menggunakan *Raspberry Pi* dengan kamera USB dari wahana ke GCS. Hasil *streaming* pada sistem yang terpasang di wahana dapat ditangkap melalui *web interface* pengguna dan mengirim *request* ke model *Llama 4 Scout* melalui *Groq API*. AI akan memberikan respon berupa deskripsi terhadap citra yang dikirim dan kemudian ditampilkan kembali pada web GCS sebagai umpan balik bagi operator. Hal ini memungkinkan deteksi kondisi visual secara cerdas dan efisien daripada deteksi objek menggunakan metode konvensional.

Perancangan Wahana Drone

Rancangan wahana drone dalam penelitian ini menggunakan *ArduPilot* sebagai *firmware* utama untuk sistem navigasi dan kendali penerbangan yang nantinya dikombinasikan dengan GCS.



DIAGRAM HARDWARE DRONE



Gambar 2. Diagram Hardware Wahana.

Pada gambar diagram diatas menjelaskan komponen apa saja yang diperlukan untuk merancang wahana, perancangan hardware wahana merujuk pada arsitektur sistem *quadcopter open source* yang tersebar luas di literatur ilmiah maupun komunitas seperti PX4 dan ArduPilot. *Flight controller Pixhawk Cube Black* yang menjadi sistem inti kendali utama, terdiri dari berbagai komponen sensorik seperti kompas, barometer untuk mengukur ketinggian, serta IMU yang mengukur percepatan dan orientasi dari wahana. Komponen pendukung lainnya seperti *GPS Here 2* sebagai sensor navigasi dalam memberikan informasi lokasi wahana, selanjutnya modul telemetri sebagai komunikasi data dua arah antara wahana di udara dan GCS. Sinyal radio diterima oleh *receiver* dari *remote control* pilot lalu *receiver* akan meneruskan *input* dari *remote* ke *flight controller* yang nantinya data *input* sinyal radio diubah menjadi *output* sinyal PWM (*Pulse with Modulation*) untuk mengatur kecepatan *Brushless Motor* melalui modul ESC (*Electronics Speed Controller*) [14].

Pengujian Sistem

Tahap akhir dalam penelitian ini adalah pengujian keseluruhan sistem, termasuk dalam lingkungan simulasi. Pengujian dilakukan untuk memastikan bahwa wahana drone dapat beroperasi dengan stabil, mengirimkan data secara akurat, serta mampu



menganalisis gambar citra udara dengan Multimodal AI. Pada tahap uji coba lingkungan terkendali, dilakukan menggunakan ROS dan *Gazebo Simulator* untuk memastikan drone dapat mengeksekusi perintah dengan baik serta mampu mengirimkan citra dari udara ke GCS. Setelah pengujian sistem dalam simulasi berhasil maka dilakukan pengujian sistem secara langsung menggunakan wahana drone yang sudah dirancang sebelumnya dengan tetap memperhatikan keamanan dalam penelitian.

Tujuan dari pengujian pada sistem nyata ini diharapkan dapat menguji respon Multimodal AI terhadap citra udara asli dengan berbagai variabel seperti pencahayaan dan objek dinamis serta memberikan wawasan terhadap batasan sistem seperti ketahanan baterai, jarak sinyal efektif dan peforma wahana untuk analisis citra udara.

HASIL

Hasil dari penelitian ini adalah sebuah sistem yang terdiri dari empat komponen utama: (1) lingkungan simulasi sistem berbasis ROS dengan *Gazebo Simulator*, (2) GCS berbasis *Flask* dan ROS, (3) sistem analisis citra udara berbasis Multimodal AI, serta (4) wahana drone berbasis *AutoPilot*. Seluruh komponen berhasil diintegrasikan baik dalam lingkungan simulasi menggunakan *Gazebo* dan pengujian langsung wahana fisik.

Konfigurasi dan Gambaran Umum Lingkungan Simulasi

Pengembangan sistem yang dirancang memerlukan simulasi yang dapat meniru skenario pemantauan dari udara secara realistis dalam berbagai kondisi. Lingkungan simulasi dikonfigurasi dan dijalankan pada sistem operasi *Ubuntu 20.04 LTS*, menggunakan *framework ROS Noetic* yang merupakan distribusi ROS spesifik untuk sistem operasi *Ubuntu 20.04* dan *MAVROS*. Terdapat tiga skenario lingkungan simulasi yang dirancang secara khusus yaitu :



1. Lingkungan Perkotaan (*Urban Environment*)

Skenario ini meniru susana perkotaan dengan beberapa gedung, jalan raya, perumahan dan kendaraan.



Gambar 3. Gambar skenario simulasi Lingkungan Kota.

Pada gambar diatas objek manusia diletakan di trotoar atau dekat kendaraan. Kondisi tersebut dirancang untuk mewakili situasi analisis citra udara pada wilayah padat penduduk, dimana sistem diharapkan dapat mendeteksi memberikan deskripsi tentang situasi lingkungan pada kompleksitas visual dan struktur vertikal.

2. Lingkungan Pedesaan (*Village Environment*)

Desain simulasi ini mencerminkan area pedesaan dengan perumahan rendah dan ladang terbuka.

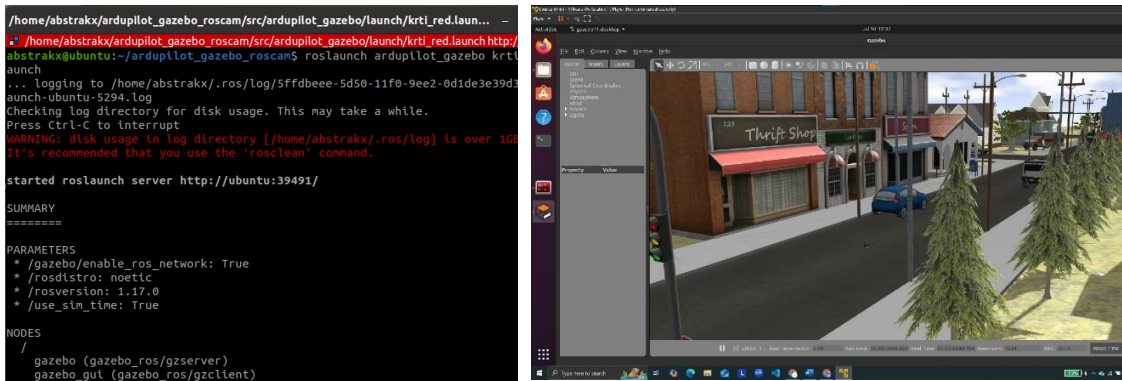


Gambar 4. Gambar skenario simulasi Lingkungan Desa.

Lingkungan pedesaan pada gambar diatas terdiri dari beberapa model yang cukup kompleks seperti jalan, perumahan, dan vegetasi yang berada di tengah desa. Kondisi ini dipakai untuk mengevaluasi peforma sistem di area minim gangguan visual dan area yang lebih lapang namun tetap menantang dari sisi variasi jenis objeknya.

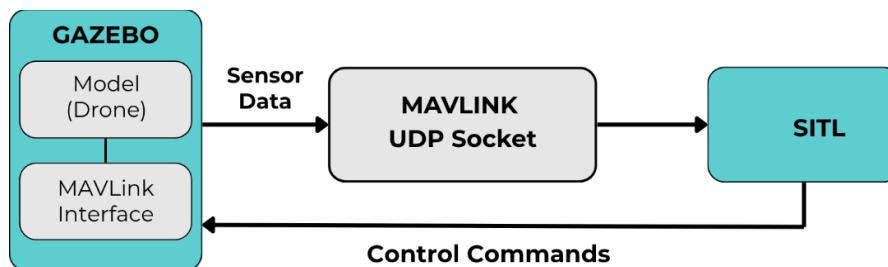
Simulasi dijalankan dengan memuat file *environment* dalam format *.launch* pada gazebo dieksekusi melalui terminal linux sebagai berikut :





Gambar 5. Menjalankan simulasi dengan terminal linux dan tampilan UI gazebo.

Setelah dijalankan *node gazebo* akan memuat simulasi termasuk model drone dan objek lingkungan. Namun, proses ini hanya mencakup visualisasi dan *spawn model*, perlu beberapa terminal tambahan untuk kendali wahana drone secara virtual menggunakan *Ardupilot Software In The Loop (SITL)*. Agar model drone dari gazebo dan SITL dapat berkomunikasi, digunakan *plugin* bernama *gazebo_mavlink_interface* yang mengirimkan data sensor dari lingkungan simulasi seperti orientasi drone, posisi GPS, kecepatan dan akselerasi serta informasi sensor dikemas menjadi pesan MAVLink dan dikirim melalui *UDP socket*.



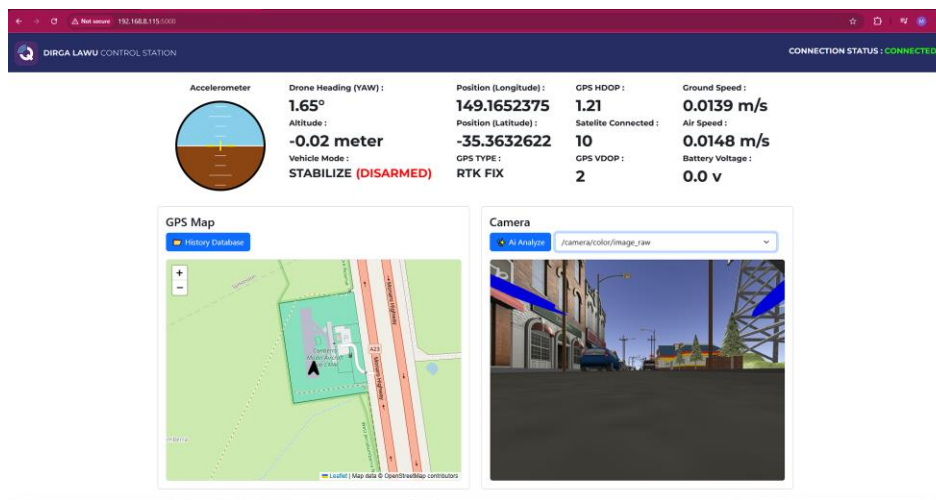
Gambar 6. Diagram Koneksi Gazebo Model dan SITL.

Pada gambar diatas, SITL memproses data untuk menghasilkan perintah kontrol seperti kecepatan rotasi motor atau manuver tertentu, dan kemudian dikembalikan kembali ke *plugin* dan diterapkan pada model drone sehingga simulasi dapat berjalan secara *real-time* dan realistis seolah drone fisik sedang beroperasi di lingkungan nyata.



Integrasi GCS Berbasis ROS dan Flask

Secara umum, ROS digunakan sebagai *middleware* komunikasi antar komponen pada sistem wahana. *Package* MAVROS digunakan sebagai penghubung antara protokol MAVLink yang digunakan dalam *flight controller* dengan sistem ROS. Sehingga dengan MAVROS data wahana seperti posisi global, orientasi, kecepatan, status sistem, hingga topik navigasi dapat dipublikasi (*publish*) dan disubskripsi (*subscribe*) oleh *node-node* ROS, yang kemudian dapat diakses oleh GCS.



Gambar 7. Tampilan antarmuka GCS.

Sementara itu, *Flask* digunakan untuk membangun antarmuka web yang ringan untuk menampilkan data telemetri secara *real-time*, seperti posisi koordinat, ketinggian, status *arming* dan perintah lainnya [15]. Komunikasi antara *Flask* dan ROS menggunakan integrasi HTTP *endpoint* yang memungkinkan interaksi dua arah antara *backend* ROS dan *frontend* *Flask*. GCS terdiri dari tiga komponen utama yaitu :

1. Drone Data Information

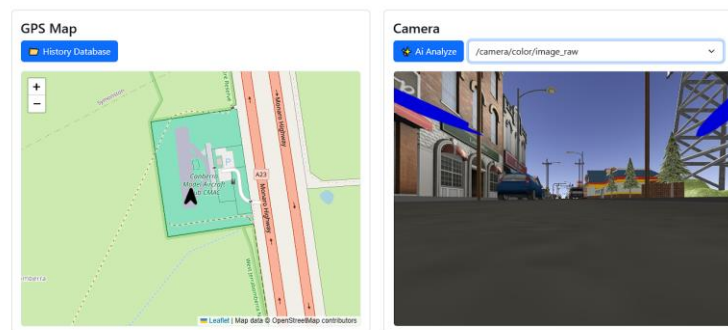


Gambar 8. Tampilan komponen informasi data wahana.



Berdasarkan gambar diatas, komponen ini akan menampilkan data-data yang *publish* oleh package MAVROS. Informasi pada komponen ini penting untuk diketahui operator GCS untuk navigasi dan kontrol. Informasi seperti status wahana *arming* (wahana siap lepas landas) dan status lainnya bertujuan untuk meningkatkan kesadaran situasional operator terhadap status fisik dan lingkungan UAV sehingga operator dapat mengambil keputusan secara cepat, akurat dan aman selama misi berlangsung

2. Map dan Video Stream Section



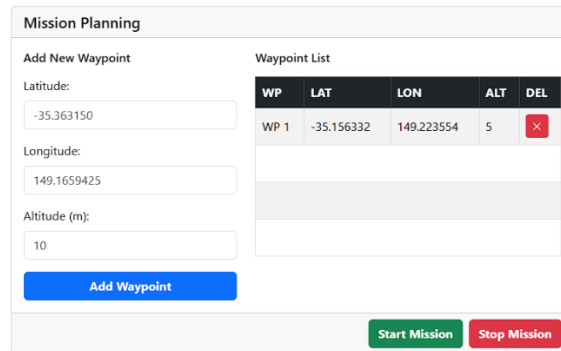
Gambar 9. Tampilan komponen Map dan Video Stream GCS.

Selanjutnya adalah komponen *map* dan *video stream* yang dapat memberikan informasi spasial dan visual secara *real-time*. Komponen visualisasi peta mampu menampilkan posisi dan lintasan wahana berdasarkan data GPS. Sementara itu *video stream* dari kamera memberikan informasi kontekstual terhadap lingkungan wahana yang penting untuk observasi misi maupun analisis citra udara dengan Multimodal AI. Kedua komponen tersebut memperkuat kemampuan monitoring dan pengambilan keputusan terhadap misi wahana.

3. Mission Control Dashboard

Komponen *Mission Control Dashboard* berfungsi sebagai pusat pengendalian wahana secara otonomus. Dashboard ini menjadi titik interaksi langsung antara operator dan sistem drone.

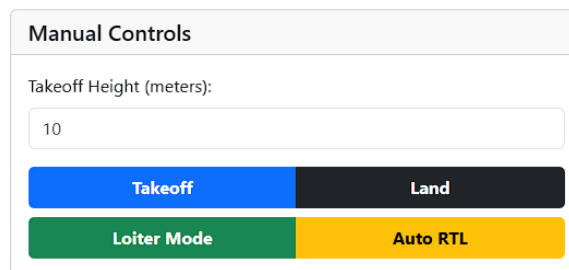




The screenshot shows a 'Mission Planning' window. On the left, there is a form to 'Add New Waypoint' with input fields for Latitude (-35.363150), Longitude (149.1659425), and Altitude (m) (10). A blue 'Add Waypoint' button is below. On the right, a 'Waypoint List' table contains one entry: WP 1 with LAT -35.156332, LON 149.223554, ALT 5, and a red 'X' delete button. At the bottom right, there are green 'Start Mission' and red 'Stop Mission' buttons.

Gambar 10. Tampilan sub-komponen Mission Planning.

Pada gambar diatas sub-komponen *Mission Planning* berfungsi sebagai perencanaan penerbangan wahana secara otonom berdasarkan koordinat GPS yang diinputkan pada *form field* termasuk data *latitude*, *longitude* dan ketinggian wahana pada *waypoint* tersebut yang kemudian ditampilkan pada tabel. *Button start Mission* dan *stop Mission* berfungsi untuk memulai dan menghentikan proses penerbangan otonomus.

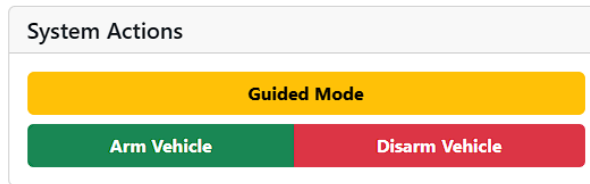


The screenshot shows a 'Manual Controls' window. It features an input field for 'Takeoff Height (meters)' with the value 10. Below the input field are four buttons: 'Takeoff' (blue), 'Land' (black), 'Loiter Mode' (green), and 'Auto RTL' (yellow).

Gambar 11. Tampilan sub-komponen Manual Controls.

Selanjutnya sub-komponen *Manual Control* mengatur kontrol wahana untuk *takeoff*, *landing*, mengatur mode wahana seperti RTL (*Return To Land*) dan LOITER. *Input field* ketinggian akan menginisiasi berapa ketinggian wahana ketika *takeoff* dalam satuan meter. Mode RTL berfungsi untuk mengembalikan wahana ke posisi semula berdasarkan *home point* ketika wahana menghadapi kendala dalam misi. Mode LOITER berfungsi agar wahana dapat diam di tempat sesuai titik koordinat penerbangannya saat itu, memungkinkan pilot dapat mengambil alih wahana secara manual melalui *remote control*.





Gambar 12. Tampilan sub-komponen System Actions.

Sub-komponen *system action* berfungsi untuk melakukan *preflight test* (pengujian sebelum penerbangan). Tombol *armed* berfungsi untuk menandakan bahwa wahana siap terbang serta semua sistem kendali penerbangan seperti GPS, *gyroscope*, dan *accelerometer* berfungsi. Sedangkan *disarmed* untuk memastikan jika komponen penggerak wahana non-aktif sehingga mencegah baling-baling berputar secara tidak sengaja saat melakukan pengaturan. Mode GUIDED merupakan kunci utama dalam penerbangan otonomus, mode ini dapat membuat wahana menerima *input* dari sumber *eksternal* selain dari *remote control*.

Implementasi AI Multimodal

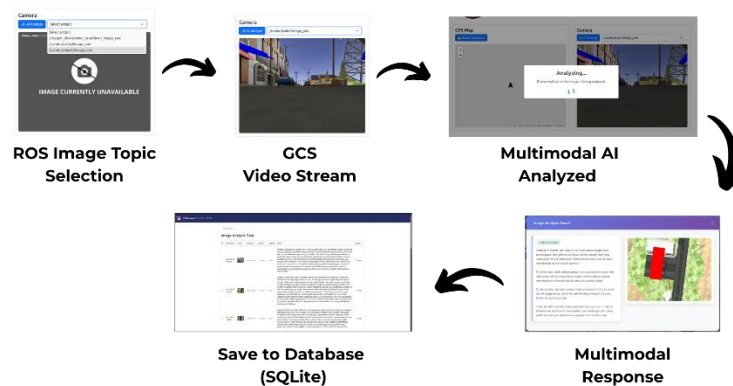
Salah satu komponen penting yang diintegrasikan dalam sistem GCS adalah Multimodal AI. Teknologi ini memungkinkan sistem mampu menerima data visual gambar dan memahami serta memberikan respon berbasis teks terhadap informasi dalam gambar tersebut. Kemampuan LLM modern yang telah dilatih terhadap data lintas-modal dapat memberikan analisis semantik terhadap citra termasuk citra udara.

Proses Pengambilan dan Pengiriman Citra Udara

Citra ditangkap melalui modul kamera yang terpasang pada wahana menggunakan kamera nirkabel yang terhubung ke jaringan Wi-Fi, seperti modul kamera yang ditransmisikan dengan *raspberry pi* atau melalui *plugin* kamera pada gazebo. Pada aplikasi nyata kamera tidak terhubung langsung ke *flight controller*, melainkan beroperasi secara independen melalui jaringan lokal. Integrasi dengan sistem ROS, *stream* video akan ditangkap dan dikonversi menjadi ROS *image topic*



menggunakan *node* ROS seperti *image_transport*. Pendekatan ini memungkinkan citra dari kamera dapat dipublikasikan sebagai topik ROS (misalnya */camera/image_raw*), sehingga dapat digunakan oleh *node* lain dalam ekosistem ROS.



Gambar 13. Diagram penggunaan analisis citra udara GCS.

Pada gambar diagram diatas operator perlu memilih topik kamera yang akan dilakukan analisis citra udara, hal ini memungkinkan wahana untuk memiliki lebih dari satu kamera yang berfungsi menangkap konteks lingkungan dari sudut pandang yang berbeda seperti bagian depan wahana atau bawah wahana. Topik dari kamera tersebut akan ditampilkan oleh GCS sebagai *video stream*. Operator dapat memulai analisis citra udara melalui *button capture* yang akan mengirimkan gambar ke AI melalui *API Groq* dalam format *Base64* melalui metode *POST*. Selanjutnya Multimodal AI akan mengembalikan paragraf naratif deskriptif tentang hasil citra udara sebagai respon dari gambar dan *prompt* yang kita kirimkan. *Respons* tersebut akan disimpan dalam *database web flask* menggunakan *SQLite*.

Desain Prompt untuk Analisis Citra Udara

Agar Multimodal AI dapat memberikan hasil yang terarah, perlu disusun sebuah *prompt* khusus yang menjelaskan konteks misi. *Prompt* yang digunakan dalam penelitian ini adalah sebagai berikut :



“Gambar ini diambil dari udara oleh drone dalam rangka misi pemantauan. Identifikasi apakah terdapat manusia dan kendaraan (seperti mobil, motor, truk) di dalam gambar. Berikan deskripsi dalam bentuk paragraf naratif yang menjelaskan lokasi dan aktivitas objek secara alami dan informatif. Jika terdapat objek menarik lain (bangunan, vegetasi, aktivitas tidak biasa), sebutkan secara singkat”.

Prompt diatas dirancang untuk menghasilkan keluaran berupa paragraf deskriptif yang menjelaskan situasi terhadap citra udara yang kita tangkap. Kelebihan menggunakan Multimodal AI memungkinkan penggunaan sistem secara fleksibel saat menjalankan misi tertentu karena operator dapat menyesuaikan *prompt* terhadap kebutuhan misi yang dijalankan.

Hasil Perancangan Wahana Drone

Wahana yang digunakan dalam penelitian ini dirancang agar dapat terintegrasi dengan sistem kontrol berbasis ROS. Wahana menggunakan *frame quadcopter* berukuran 450 mm yang terbuat dari material plastik agar bobot wahana tetap ringan.

Wahana Drone yang Telah Dirakit

Wahana yang dirakit telah sesuai dengan desain *hardware* pada gambar 2. Komponen seperti *flight controller*, modul GPS, telemetri dan komponen elektronik pendukung lainnya telah terpasang sesuai skema.



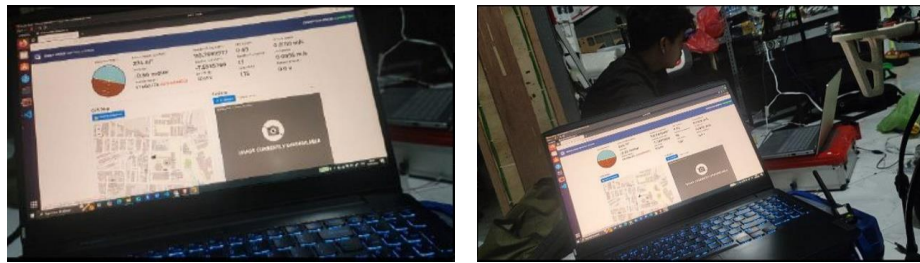
Gambar 14. Wahana drone yang telah dirakit.



Pada gambar diatas wahana dirakit dengan sistem propulsi empat motor *brushless Holybro 2216* dengan nilai KV 920 dan dikombinasikan dengan ESC *HobbyWing Skywalker 50A* untuk mengatur kecepatan putaran motor. *Propeller* yang digunakan berbahan plastik berukuran 10x45 inci yang cukup untuk drone dapat *takeoff* dan bermanuver. Tenaga listrik disuplai oleh baterai LiHV 4s 6500mAh yang mampu memberikan keseimbangan antara kapasitas dan bobot wahana.

Koneksi Wahana dengan GCS

Komunikasi antara wahana dan GCS, menggunakan modul *telemetry Holybro SiK V2* dengan frekuensi 433 MHz yang terhubung ke GCS melalui protokol MAVROS.



Gambar 15. Koneksi antara wahana dan GCS.

Gambar diatas memperlihatkan hasil koneksi antara wahana dan GCS. Data sensor dari *flight controller* seperti *accelerometer*, *gyroscope*, *barometer*, dan *magnetometer*. Lalu informasi yang dikirim mencakup informasi posisi GPS, mode penerbangan, kecepatan, dan informasi kesehatan GPS melalui modul telemetri secara *real time*.

PEMBAHASAN

Pengujian sistem dilakukan untuk mengevaluasi performa dan fungsionalitas dari GCS yang sudah dirancang sebelumnya. Pengujian terhadap sistem GCS dilakukan dengan pendekatan *Black Box Testing* sehingga evaluasi fungsionalitas sistem dilakukan tanpa membedah kode sumber atau struktur internal sistem [16]. Hasil pengujian sistem dapat



dilihat pada tabel 1 yang menunjukkan sistem GCS dapat memberikan hasil sesuai harapan meskipun tidak menutup kemungkinan terjadi bug pada sistem yang sudah dirancang.

Tabel 1. Black Box Testing Sistem GCS

Nama Fitur	Detail <i>Test Case</i>	Hasil yang Diharapkan	Hasil yang Terjadi	Status
Main Dashboard	Operator dapat menghubungkan sistem ke wahana	Wahana terkoneksi dengan sistem GCS	Wahana berhasil terkoneksi dengan sistem GCS	<i>Passed</i>
	Operator dapat menghetikan koneksi sistem ke wahana	Sistem GCS putus koneksi dengan wahana	Sistem GCS berhasil disconnected dari wahana	<i>Passed</i>
	Operator GCS dapat melihat infromasi data wahana setelah terkoneksi	Informasi pada section Drone Data Infromation dapat diterima oleh GCS	Data wahana berhasil dikirimkan dan ditampilkan oleh GCS	<i>Passed</i>
	Operator dapat melihat posisi wahana secara realtime pada maps	Posisi wahana dapat ditampilkan secara realtime pada bagian maps	Posisi wahana berhasil ditampilkan secara realtime berdasarkan koordinat wahana pada maps	<i>Passed</i>
	Operator dapat mengganti video stream pada GCS berdasarkan total kamera pada wahana	Dapat mengganti topic image pada video stream GCS	Operator berhasil mengganti topic image pada video stream GCS	<i>Passed</i>
	Operator dapat melakukan analisis citra udara dengan menekan tombol AI Analyze	Multimodal AI akan melakukan analisis citra udara ketika button ditekan	Multimodal AI berhasil memberikan respon deskriptif terhadap citra udara yang diupload ke AI	<i>Passed</i>
Drone Control	Operator dapat menginputkan koordinat dan ketinggian waypoint wahana	Dapat menginputkan koordinat untuk misi penerbangan otonomus	Berhasil menginputkan koordinat dan waypoint tampil di tabel.	<i>Passed</i>



Operator dapat memulai dan menghentikan misi penerbangan otonomus berdasarkan waypoint	Dapat memulai dan menghentikan misi otonomus berdasarkan waypoint yang diberikan	Berhasil memulai dan menghentikan misi otonomus wahana	<i>Passed</i>
Operator dapat menginputkan ketinggian untuk wahana takeoff	Dapat menginputkan nilai untuk menerbangkan wahana ke ketinggian yang diinputkan	Berhasil menginputkan nilai untuk ketinggian takeoff wahana	<i>Passed</i>
Operator dapat memulai takeoff wahana berdasarkan input ketinggian	Dapat memulai takeoff wahana berdasarkan input ketinggian lewat button takeoff	Wahana berhasil takeoff ke ketinggian tersebut berdasarkan input sebelumnya	<i>Passed</i>
Operator dapat mengganti mode wahana menjadi mode RTL	Dapat mengganti mode wahana menjadi RTL ketika button RTL ditekan	Wahana berhasil berubah ke mode RTL setelah button ditekan	<i>Passed</i>
Operator dapat mengubah mode wahana menjadi mode LOITER	Dapat mengganti mode wahana menjadi LOITER ketika button LOITER ditekan	Wahana berhasil berubah ke mode LOITER setelah button ditekan	<i>Passed</i>
Operator dapat mengubah wahana menjadi mode GUIDED	Dapat mengganti mode wahana menjadi GUIDED ketika button GUIDED ditekan	Wahana berhasil berubah ke mode GUIDED setelah button ditekan	<i>Passed</i>
Operator dapat arm atau disarm wahana untuk preflight test	Dapat melakukan arm atau disarm ketika button ditekan	Berhasil arming dan disarmed wahana setelah button ditekan	<i>Passed</i>




Validasi Hasil Deteksi Multimodal AI

Untuk memastikan efektivitas dan keakuratan sistem Multimodal AI dalam menganalisis citra udara, dilakukan proses validasi oleh peneliti terhadap hasil deteksi kondisi aktual yang terekam dalam gambar. Validasi ini dilakukan baik pada

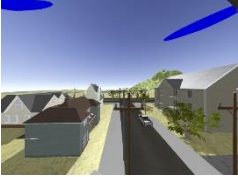


lingkungan simulasi maupun kondisi lapangan dengan membandingkan hasil yang dihasilkan oleh AI dengan objek aktual.

Tabel 2. Hasil Analisis Citra Udara Pada Environment Simulasi


Gambar	Nama Gambar	Deskripsi Multimodal AI
	IMG_01.jpg	<p>Gambar ini tampaknya merupakan hasil rendering 3D dari sebuah kota kecil yang tenang. Dari perspektif udara, terlihat sebuah jalan lurus yang dikelilingi oleh bangunan-bangunan dengan gaya arsitektur klasik. Di sisi kiri jalan, terdapat sebuah bangunan besar berwarna merah bata dengan jendela-jendela melengkung yang menambah kesan klasik pada suasana. Beberapa mobil terlihat sedang bergerak di sepanjang jalan, salah satunya berwarna biru muda yang sedang menuju ke arah bawah gambar. Di sisi kanan jalan, terlihat beberapa pohon kecil yang ditanam rapi, serta sebuah menara struktur logam yang mungkin merupakan bagian dari infrastruktur kota. Di kejauhan, sebuah bangunan dengan atap biru, garis-garis berwarna merah dan kuning, serta lapangan hijau dengan bukit kecil menambah keindahan pemandangan. Tidak ada manusia yang terlihat dalam gambar ini. Adapun objek menarik lainnya adalah menara logam dan bangunan bergaya klasik yang mendominasi pemandangan.</p>
	IMG_02.jpg	<p>Gambar ini diambil dari udara dan menunjukkan pemandangan sebuah jalan dengan sebuah mobil berwarna biru yang sedang bergerak di atasnya. Mobil tersebut tampak sedang melaju di jalan yang terlihat relatif kosong. Di sekitar jalan, terdapat beberapa bangunan, termasuk sebuah toko yang ditandai dengan tulisan "Thrift Shop" di fasadnya. Bangunan lainnya terlihat memiliki dinding berwarna hijau dengan jendela-jendela dan sebuah pintu garasi. Tidak ada manusia yang terlihat dalam gambar ini. Adapun kendaraan yang terlihat hanya mobil biru tunggal yang sedang bergerak. Objek menarik lainnya termasuk vegetasi yang tampak di sekitar bangunan, memberikan kesan adanya lingkungan yang cukup rindang dan asri. Bangunan-bangunan tersebut tampak memiliki arsitektur yang sederhana namun fungsional. Tidak ada aktivitas tidak biasa yang terlihat dalam gambar ini.</p>
	IMG_03.jpg	<p>Dari gambar yang diambil dari udara oleh drone, tampak sebuah truk berada di atas jalan yang terlihat berada di antara bangunan-bangunan. Truk tersebut berwarna abu-abu dengan garis merah biru di sampingnya. Di sekitar truk, tidak terlihat adanya manusia yang jelas, namun ada satu titik kuning kecil yang mungkin bisa diartikan sebagai manusia atau objek kecil di dekat</p>



	<p>truk. Di sebelah kiri truk, ada sebuah bangunan dengan dinding berwarna krem dan beberapa jendela. Di depan bangunan tersebut, ada sebuah pohon yang cukup besar dengan dedaunan hijau. Aktivitas yang terlihat hanyalah truk yang mungkin sedang berhenti atau melintas di jalan tersebut. Tidak ada aktivitas tidak biasa yang terlihat dari gambar ini.</p>
 <p>IMG_04.jpg</p>	<p>Gambar ini merupakan pemandangan udara dari sebuah lingkungan perumahan yang diambil oleh drone. Dalam gambar tersebut, terlihat beberapa rumah dengan berbagai warna dan gaya arsitektur yang tersebar di sepanjang jalan. Di tengah jalan, terdapat sebuah truk yang sedang melaju. Tidak ada manusia yang terlihat di gambar ini. Selain itu, terdapat beberapa pohon dan vegetasi yang tumbuh di sekitar rumah-rumah tersebut, menambah kesan alami pada lingkungan ini.</p>

Tabel 3. Hasil Analisis Citra Udara Pada Environment Lapangan

Gambar	Nama Gambar	Deskripsi Multimodal AI
	<p>IMG_05.jpg</p>	<p>Gambar ini diambil dari udara oleh drone, menunjukkan sebuah jalan setapak yang terbuat dari batu bata dengan sebuah mobil hitam yang diparkir di sepanjang dinding putih yang terlihat sudah mulai kusam. Mobil tersebut menghadap ke arah pojok kiri bawah gambar. Di bagian atas gambar, terlihat sebagian dari bangunan dengan dinding berwarna ungu. Bangunan tersebut memiliki sebuah pintu kayu yang terlihat cukup kokoh. Selain itu, ada juga sebagian kendaraan lain yang terparkir di depannya, tetapi tidak terlalu jelas. Tidak ada manusia yang terlihat dalam gambar ini. Adapun aktivitas yang tampak hanyalah mobil yang sedang terparkir. Objek menarik lainnya yang dapat dilihat adalah vegetasi yang mungkin ada di sekitar, tetapi tidak terlihat jelas dalam gambar ini. Bangunan dengan pintu kayu yang unik juga menjadi poin menarik. Secara keseluruhan, gambar ini menunjukkan suasana yang tenang dan tidak ada aktivitas yang mencolok.</p>
	<p>IMG_06.jpg</p>	<p>Gambar ini diambil dari udara, menunjukkan sebuah area parkir yang terletak di depan sebuah bangunan. Lokasi ini tampak seperti area parkir umum atau parkir khusus untuk sebuah gedung. Di dalam gambar, terdapat dua kendaraan yang terlihat, yaitu sebuah mobil berwarna hitam yang terparkir di pojok kanan dan sebuah mobil berwarna abu-abu yang juga terparkir di bagian tengah belakang. Tidak ada manusia yang terlihat jelas melakukan aktivitas di dalam gambar ini. Bangunan di bagian belakang tampak memiliki dua lantai dengan atap berwarna</p>

	<p>merah. Beberapa pohon juga terlihat di area parkir ini, termasuk sebuah pohon besar di tengah gambar yang tampak baru dipangkas.</p>
 <p>IMG_07.jpg</p>	<p>Gambar ini diambil dari udara oleh drone, menunjukkan pemandangan sebuah area yang mungkin sedang dalam proses pembangunan atau renovasi. Lokasinya tampak seperti sebuah kawasan yang terdiri dari beberapa bangunan, salah satunya adalah bangunan dua lantai dengan atap merah di sisi kiri gambar. Bangunan ini terlihat memiliki beberapa motor yang terparkir di depannya, tetapi tidak ada manusia yang terlihat jelas di sekitar area tersebut. Di bagian tengah gambar, terdapat sebuah jalan yang cukup lebar dengan beberapa barikade biru dan putih di depannya, yang mungkin digunakan untuk mengatur lalu lintas atau akses ke area tersebut. Sebuah mobil putih terlihat berada di tengah jalan, menghadap ke arah bangunan yang lebih besar dan belum selesai di bagian belakang. Bangunan yang belum selesai ini memiliki sebuah crane besar yang sedang digunakan untuk proses pembangunannya. Tidak ada aktivitas manusia yang terlihat jelas dalam gambar ini, namun terdapat beberapa pohon dan vegetasi di sekitar bangunan yang menambah kesan alami pada area tersebut. Secara keseluruhan, gambar ini menunjukkan suasana yang relatif tenang dan proses pembangunan yang sedang berlangsung.</p>

Tabel 4. Hasil Pengujian Multimodal AI

Skenario	Nama Gambar	Objek Aktual	Objek Terdeteksi AI	Akurasi	Keterangan
Environment Simulasi	IMG_01.jpg	Beberapa Mobil, Bangunan gaya klasik, Tower, Jalan, Bukit kecil, Pepohonan	Beberapa Mobil, Bangunan gaya klasik, Menara logam, Jalan, Bukit kecil, Pepohonan	92%	Multimodal AI mampu mengetahui objek serta kondisi lingkungan dengan sangat baik. Bahkan AI mampu mengetahui environment tersebut adalah simulasi
	IMG_02.jpg	Mobil biru, Toko dengan tulisan thrift shop, Bangunan, Pepohonan	Mobil biru, Toko dengan tulisan thrift shop, Bangunan, Pepohonan	100%	Multimodal AI mampu mengetahui objek serta kondisi lingkungan dengan sangat baik.



	Bangunan, Ambulans, Manusia, Pohon	Bangunan, Truk, Manusia, Pohon		Konteks gambar dari citra udara tentu menjadi tantangan. Ambiguitas objek cukup terlihat dimana AI mendeskripsikan ambulans sebagai truk dan manusia. Namun AI berhasil menjelaskan konteks dengan sangat baik
	IMG_03.j pg		87%	
	Beberapa Rumah, Pepohonan, Truk	Beberapa Rumah, Pepohonan, Truk		AI mampu menjelaskan dengan sangat baik lingkungan pada citra udara yang ditangkap wahana.
	IMG_04.j pg		100%	
	Mobil hitam, Rumah warna ungu dengan pintu kayu, Mobil putih, vegetasi	Mobil hitam, Bangunan warna ungu dengan pintu kayu, Mobil putih, vegetasi		AI sangat mampu menjelaskan situasi lingkungan pada gambar dengan sangat baik terdapat salah penafsiran terhadap bangunan dan rumah akan tetapi model sudah sangat baik menafsirkan citra udara yang diterima.
	IMG_05.j pg		100%	
Environment Lapangan	Gedung, Mobil hitam, Mobil abu abu, pepohonan, manusia	Gedung, Mobil hitam, Mobil abu abu, pepohonan,		AI tidak dapat mengenali objek manusia pada kondisi lingkungan yang diakibatkan oleh kualitas gambar yang semakin buruk seiring jauhnya Raspberry Pi pada wahana dari GCS
	IMG_06.j pg		80%	



IMG_07.jpg	Bangunan 2 lantai, Motor, Crane, Mobil putih, Barikade, Vegetasi	Bangunan 2 lantai, Motor, Crane, Mobil putih, Barikade, Vegetasi	100%	Ai dapat mengenali setiap objek dengan sangat rinci termasuk suasana lingkungan pada citra udara tersebut.
------------	---	---	------	--

Berdasarkan hasil pengujian, Tabel 2 dan Tabel 3 menyajikan deskripsi naratif terhadap citra udara yang dianalisis oleh sistem Multimodal AI. Deskripsi ini memberikan wawasan terhadap konteks visual di lapangan. Sebagai contoh, pada IMG_03.jpg sistem tidak mampu mengenali ambulans pada *image* tersebut akan tetapi sistem mengenali objek tersebut sebagai *truck* walaupun memang masih dalam kategori kendaraan. Hal ini disebabkan oleh sudut pandang objek dari atas bisa tampak mirip sehingga mempengaruhi segmentasi dalam deskripsi yang dihasilkan Multimodal AI.

Hasil tersebut menjelaskan bahwa Multimodal AI masih memiliki keterbatasan dalam menganalisis citra udara terutama pada lingkungan nyata dimana hasil deskripsi yang dihasilkan sangat bergantung pada kualitas jaringan WI-FI dari GCS ke wahana, semakin jauh akan meningkatkan *delay* pada penerimaan citra dan vibrasi yang dihasilkan wahana ketika terbang juga dapat mempengaruhi hasil analisis citra nantinya. Selanjutnya pada Tabel 4, memberikan keterangan perbandingan hasil pengamatan oleh Multimodal AI dan objek aktual menjelaskan tentang performa model secara ringkas dan padat terhadap hasil analisis yang diberikan pada Tabel 2 dan Tabel 3.

KESIMPULAN

Secara keseluruhan, hasil pengujian menunjukkan sistem GCS berbasis Multimodal AI dapat memberikan performa yang baik. Deskripsi naratif citra udara yang dihasilkan oleh AI dapat memberikan pemahaman terhadap konteks visual dari citra udara yang diambil. Meskipun terdapat beberapa batasan terutama pada citra dengan sudut pandang tampak atas, koneksi WI-FI dari *video stream* citra udara ke GCS dan *noise* yang diakibatkan getaran dari wahana dapat mempengaruhi hasil analisis citra udara oleh Multimodal AI.



Dengan demikian, sistem yang dikembangkan menunjukkan potensi yang kuat untuk diaplikasikan dalam berbagai misi berbasis UAV, pemanfaatannya yang cukup luas membuka peluang bagaimana AI dan UAV dapat menjadi suatu integrasi yang memudahkan manusia seperti pada misi pencarian korban bencana, pemetaan, inspeksi, dan lainnya. Penelitian ini juga diharapkan dapat menjadi awal dalam menjembatani terbatasnya dataset dalam pemanfaatan kecerdasan buatan dan UAV sehingga pemanfaatan UAV dapat lebih fleksibel dimasa mendatang.

UCAPAN TERIMA KASIH

Penulis menyampaikan terima kasih kepada semua pihak yang telah memberikan dukungan teknis dan kontribusi dalam penyusunan artikel ini, baik dari kalangan individu maupun instansi yang terkait.

DAFTAR PUSTAKA

- [1] Alfiandy, Suprayogi, and F. Nurwulan, "Iot (Internet of Things) Navigasi Drone Berdasarkan Waypoint," *e-Proceeding Eng.*, vol. 8, no. 2, pp. 1–8, 2021.
- [2] F. Rodrigues, C. Cardeira, J. M. F. Calado, and R. Melício, "Controlo 2016," vol. 402, pp. 691–699, 2017, doi: 10.1007/978-3-319-43671-5.
- [3] A. Ahmad and S. T. B. Handaga, "Deteksi Dini Kegagalan Catu Daya di Ruang Server menggunakan Teknik Pengolahan Citra dengan Raspberry pi dan Robot Operating System (ROS)," 2022, [Online]. Available: http://eprints.ums.ac.id/id/eprint/100498%0Ahttp://eprints.ums.ac.id/100498/12/NASKAH_PUBLIKASI.pdf
- [4] U. Fadlilah *et al.*, "Design and Build A Logistics Delivery Drone to the Patients ' Quarantine Room Using ROS Technology Without GPS : An Initial Study," no. 3, pp. 1–12, 2023.
- [5] Y. Liu, Z. Zhou, J. Liu, L. Chen, S. Member, and J. Wang, "Multi-Agent Formation Control Using Large Language Models," 2024.
- [6] H. I. Ashqar, A. Jaber, T. Alhadidi, and M. Elhenawy, "Advancing Object Detection in Transportation with Multimodal Large Language Models (MLLMs): A Comprehensive Review and Empirical Testing Advancing Object Detection in Transportation with Multimodal Large Language Models (MLLMs): A Comprehensive Review," *Computation*, vol. 13, no. 6, p. 133, 2025.
- [7] S. Speth *et al.*, "Deep learning with RGB and thermal images onboard a drone for monitoring operations," *J. F. Robot.*, vol. 39, no. 6, pp. 840–868, 2022, doi:



10.1002/rob.22082.

- [8] Z. Wang, J. Yu, A. W. Yu, Z. Dai, Y. Tsvetkov, and Y. Cao, "Simvlm: Simple Visual Language Model Pre-Training With Weak Supervision," *ICLR 2022 - 10th Int. Conf. Learn. Represent.*, no. VI, pp. 1–17, 2022.
- [9] A. Ahmed, M. Farhan, H. Eesaar, K. T. Chong, and H. Tayara, "From Detection to Action: A Multimodal AI Framework for Traffic Incident Response," *Drones*, vol. 8, no. 12, pp. 1–19, 2024, doi: 10.3390/drones8120741.
- [10] Z. Chen, J. Yan, B. Ma, K. Shi, Q. Yu, and W. Yuan, "A Survey on Open-Source Simulation Platforms for Multi-Copter UAV Swarms," *Robotics*, vol. 12, no. 2, pp. 1–24, 2023, doi: 10.3390/robotics12020053.
- [11] C. Bernardeschi, A. Fagiolini, M. Palmieri, G. Scrima, and F. Sofia, *ROS/Gazebo Based Simulation of Co-operative UAVs*, vol. 11472 LNCS. Springer International Publishing, 2019. doi: 10.1007/978-3-030-14984-0_24.
- [12] T. Indriyanto, A. R. Rizki, M. L. Hariyadin, M. F. Akbar, and A. A. A. Syafi, "Centralized swarming UAV using ROS for collaborative missions," *AIP Conf. Proc.*, vol. 2226, no. April, 2020, doi: 10.1063/5.0002616.
- [13] J. Gustave, J. Chahal, and A. Belbachir, "Functional architecture using ROS for autonomous UAVs," *ICINCO 2020 - Proc. 17th Int. Conf. Informatics Control. Autom. Robot.*, no. Icinco, pp. 506–512, 2020, doi: 10.5220/0009888305060512.
- [14] A. Haniffullah, I. Abdi Bangsa, and D. A. Stefanie, "Implementasi Quadcopter Pembawa Arm Robot Pembersih Kaca Jendela Gedung Bertingkat," *Edu Elektr. J.*, vol. 9, no. 2, pp. 36–42, 2020.
- [15] A. Ibrahim and A. Fatmawati, "Sistem Penghitung Orang Menggunakan Yolov8 Dan Flask (Studi Kasus Di Laboratorium Informatika Universitas Muhammadiyah Surakarta)," 2024, [Online]. Available: https://eprints.ums.ac.id/126135/1/NASKAH_PUBLIKASI.pdf
- [16] A. Verma, A. Khatana, and S. Chaudhary, "A Comparative Study of Black Box Testing and White Box Testing," *Int. J. Comput. Sci. Eng.*, vol. 5, no. 12, pp. 301–304, 2017, doi: 10.26438/ijcse/v5i12.301304.

