Journal of Applied Informatics and Computing (JAIC)
Vol.9, No.3, June 2025, pp. 667~673
e-ISSN: 2548-6861

Augmented Reality Development for Creating Interactive Experiences in Tourism Places

Caraka Aji Pranata [1]*, Muhammad Fairul Filza [2]*, Miquel Jan Tanudidjaja [3]*
* Teknologi Informasi, Universitas AMIKOM Yogyakarta
caraka@amikom.ac.id [1], fairul.f@amikom.ac.id [2], miqueltan@students.amikom.ac.id [3]

Article Info

Article history:

Received 2025-02-21
Revised 2025-03-30
Accepted 2025-04-09

Keyword:

Augmented Reality,
Research and Development,
ADDIE,
OpenCV,
Unity Engine.

ABSTRACT

Technology, hardware and software are developing rapidly as time goes by. One of the results of this rapid technological development is Augmented Reality (AR) technology. Current tools in augmented reality (AR) were offered by Vuforia Software Development Kit (SDK). However, with the rapid development of Unity Engine, now we can integrated OpenCV in Unity Engine. This integration presents ideas for the development of AR technology. This study uses an approach that uses research and development (R&D) methodology. The model that will be used is the utilization of the ADDIE development framework. We encountered three main problems, problems faced are "How many stages of implementing AR with mediapipe in Unity Engine?", “How people satisfaction while using this product?” and " What is the maximum distance and time required to conduct the interaction process?". From the research conducted, we found that there are four important AR development stages that must be carried using Unity Engine and OpenCV. Next for RQ2 we got a level of approval from users of 86.53% or strongly agree with what we are doing. Then for the last RQ, we got the results for optimal hand detection distance is 3 meters, and the speed with the fastest value is 0.098s.

I. Introduction

Technology, hardware and software are developing rapidly as time goes by. One of the results of this rapid technological development is Augmented Reality (AR) technology. As we know, Augmented Reality is a technology transformation that enriches our perception of the real world by putting digital elements into our physical environment. Thus, enhancement can take various forms, including visual components such as images, animations, and 3D models, seamlessly integrated into our surroundings. As an example, when using AR applications, users might see virtual objects that appear to be real-world objects, creating an interactive experience that can draw the line between the digital and physical realms [1].

In addition to visual enhancements, AR can also combine with audio effects, providing users with soundscapes that complement the visual information. This audio segmentation can include anything from background music to specific sound effects that respond to user interactions or environmental cues, further immersing users in the augmented experience. In textual information, another critical aspect of AR, allowing for displaying relevant data, instructions, or details that can enhance understanding and engagement. For example, when pointing a device at marker, users might see pop-up text that provides some facts, information, or even user-generated content related to that markers [2].

Overall, AR creates a more interactive and informative experience, enabling users to engage with their environment in novel ways. By blending the digital and physical worlds, AR has applications across various fields, including education, gaming, retail, health, and tourism, making it a tool for enhancing learning, entertainment, and everyday activities [3]. AR is used in various fields as a tool in applications, software, and hardware and is also used to predict advances in future technology [4]. Research shows that current AR development uses Unity Engine software [5].

Current tools in augmented reality (AR) were offered by Vuforia Software Development Kit (SDK) [6]. Vuforia uses a fast corner detection method that allows the identification of markers in the image, which then allows visualization of three-dimensional objects [7], [8]. Vuforia's SDK supports various devices and platforms, allowing developers to build applications that can run on smartphones or desktop software. Image recognition and tracking capabilities enable the identification of various objects and surfaces in the real world, which can be enhanced with digital content [9]. Vuforia provides several features that can be used in AR implementation, including Image Targets, Multi Targets, Cylinder Targets, and Barcode Scanners [10]. From research data conducted by Nasirudin in 2024, the author found that 45% of the published research uses Unity and Vuforia [11]. However, some of the features provided require additional tools for interacting. The extra tools can be paper with markers, marker types or controllers such as mice or joysticks. The author wants to provide the latest developments generally used in AR development research. The development will be done by interacting with virtual objects using only the palm of your hand and not requiring additional hardware.

With the rapid development of Unity Engine, now we can integrate OpenCV in Unity Engine. This integration presents ideas for the development of AR technology. Applying triggers to various objects and other things, such as body part training, can be done as triggers. As an example, is hand recognition training. OpenCV allows the implementation of training processes for hand detection [9], hand tracking [12], and hand gesture recognition [13].

Return to AR, this technology can be implemented in the fields of education [14], health [15][16] and also tourism [17]. Tourist attractions are identical to places with natural scenery, entertainment, and other activities that can make visitors relax and have fun [17]. For example, one of the tourist attractions in the Special Region of Yogyakarta Province is Gembira Loka Zoo (GLZ). GLZ is a tourist attraction with one mission: to become a creative, interesting, educational, and environmentally friendly recreation area [18]. GLZ has various public facilities that visitors can enjoy. These facilities include Zoovenir Outlets, Mammal Educational Presentations, Aves Educational Presentations, Playgrounds, and Animal Interactions. Visitors can take photos or selfies with animals at the animal interaction facility accompanied by a supervisor. In addition to getting a pleasant experience by interacting more closely with animals, the facility certainly also has certain risks. The risk is in the form of negative treatment by visitors, which impacts animal habits. Visitors who have negative interactions, such as feeding animals without permission from officers, shouting at animals, or hitting animals when they are going to take photos with [19].

Based on the prior discussion, it has been identified that a significant issue arises in preventing acts of violence towards animals when visitors attempt to capture selfies with them. When visitors prioritize their experiences, they may engage in harmful or distressing behaviours to the animals involved. For instance, visitors might invade animals' personal space, provoke them for a more dramatic photo, or even handle them inappropriately, all to capture the perfect selfie. Such actions can lead to stress, injury, or even trauma for the animals, damage their natural behaviours.

With Unity, AR and openCV technology, we are trying to develop a system that can be used in the development of AR to the next stage that involves its users without having to use markers but simply by opening their hands and being able to interact with virtual objects that have been provided. This initial achievement is expected to trigger virtual objects and bring up animals that have also been prepared for selfies.

II. Method

As explained in the previous chapter, the purpose of the research conducted is to develop AR technology using Unity Engine as software, openCV, and hand recognition. This project aims to get the optimal distance from the camera to detect and then scan the hand with hand recognition. Then, from the detection, the user can interact with the virtual object directly using only the camera and without any additional hardware.

This study uses an approach that uses research and development (R&D) methodology. The model that will be used is the utilization of the ADDIE development framework, a well-established instructional design model consisting of five main stages: analysis, design, development, implementation, and evaluation [20]. The flow of this method can be seen in Fig 1.

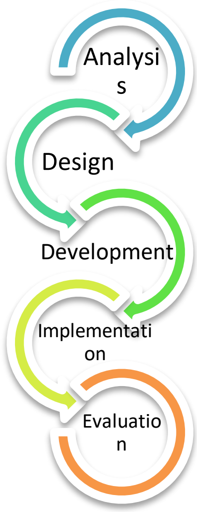

Fig 1. Reseach and Development ADDIE Models Flow

Brief overview of Fig. 1 about R&D ADDIE Models, each stage carried out is as follows:

- Analysis: In the analysis stage, the research identifies the problem. This basic step ensures that the next stage of the same problem is not found again.
- Design: Design stage outlines the needs and also the content to be provided. This stage includes determining the expected distance, the type of animal to be used, and the calculation method

- Development: The development stage is the stage where research begins to be developed. In addition to content creation, multimedia element development, and layout arrangement, we also try to dissect how video from the camera can appear on the screen in real-time. Testing and feedback are often carried out during this stage to refine the content before entering the next stage.
- Implementation: After all the materials needed are complete at the development stage, the implementation stage involves closed testing and evaluation of content delivery to the test target. There will be 10 test targets carried out. This stage includes taking data on the camera's viewing distance and hand recognition of the trigger. At this stage we also will calculate the optimal distance and its effectiveness in trigger detection. The final result of the evaluation stage is visitor satisfaction data.
- Evaluation: After completing the implementation phase, we proceed next step of research to the evaluation stage. The evaluation was carried out by conducting a trial to visitors and asking them to provide feedback with 3 questions. The first question concerns animal resemblance. We think this is necessary to ask visitors because in the future we are trying to reduce the negative impact of visitors on animals in the zoo which can make animals stressed. But we also don't want to make visitors have a bad experience because the three-dimensional animal models prepared don't resemble the original. The next question is on the safety of taking pictures and photo experience. In addition, from the problems previously explained about user safety in taking photos and also the photo experience. We consider that these two points also have an influence in the future to make this project a success.

In the research conducted, there are three main results that will be the final results after we start the various phases of this project. The expected results include: determining the number of phases in application development, detecting distance and time and analysing the level of user satisfaction in the user interaction process. Determining the number of phases in application development aims to create a clear framework for the stages needed to develop the application effectively. This will provide future benefits by considering the complex development process. So that in the future by documenting the steps that have been taken, it is hoped that it can facilitate complex development. Then for the next results, the evaluation of user interaction with virtual objects against distance and time. The variables obtained can be used as a reference to get more accurate camera placement and more clearly about how the camera can detect users. We assume that distance testing is very important before it can be used directly by users because it can anticipate errors that will occur. The last result is the collection and analysis of data on user satisfaction which is very necessary to be able to get feedback on the application being developed. From the three questions asked, users will choose five answer options, the answers are strongly agree, agree, neutral, disagree, and strongly disagree. From the research conducted by Victor [21] there are two stages, the first stage is the calculation of the Likert scale by categorizing the results according to the points applied. The points applied to the highest value are 5 for the answer strongly agree and 1 for strongly disagree. Next there is the interval calculation stage, namely by adding the results of the Likert score then dividing it by the maximum value of the Likert and multiplying it by 100 percent.

III. Results

In the early stages of the method used, we encountered three main problems, problems faced are:

- RQ 1: How many stages of implementing AR with mediapipe in Unity Engine?
- RQ 2: How people satisfaction while using this product?
- RQ 3: What is the maximum distance and time required to conduct the interaction process?

After several initial factors to consider, on second step was to conduct an in-depth literature study on AR technology. We did this so that the prototype we prepared did not misleading from the basic understanding of AR. After the initial stage is complete, the next step is to prepare the prototype concept to be developed. The consideration in our initial development is about hand recognition and continued with hand tracking. We use a dataset from a mediapipe landmark model for this research [12], [13], [22].

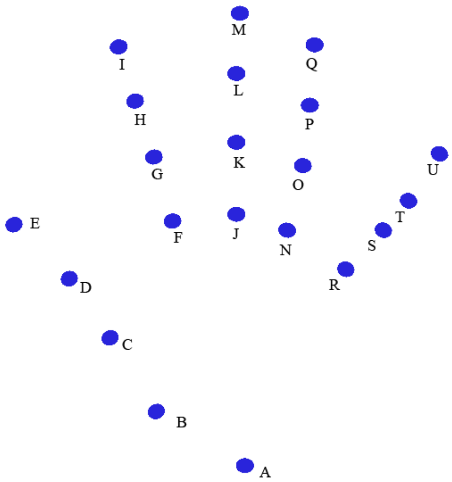

Fig 2. Mediapipe Landmark

Some terms used in determining key points in Fig. 1 are Carpometacarpal (CMC), Interphalangeal (PIP), Metacarpophalangeal (MCP), and Interphalangeal (DIP). Several points are used as key points, and the descriptions can be seen in Table 1.

TABLE 1
MEDIAPIPE KEYPOINT

Hand Point			
Point	**Joint Name**	**Point**	**Joint Name**
A	Wrist	L	Middle Finger DIP
B	Thumb CMC	M	Middle Finger TIP

C	Thumb MCP	N	Ring Finger MCP
D	Thumb PIP	O	Ring Finger DIP
E	Thumb TIP	P	Ring Finger PIP
F	Index Finger MCP	Q	Ring Finger TIP
G	Index Finger PIP	R	Pinky MCP
H	Index Finger DIP	S	Pinky PIP
I	Index Finger TIP	T	Pinky DIP
J	Middle Finger MCP	U	Pinky TIP
K	Middle Finger PIP		

After understanding the main points of the mediapipe, we started to provide the optimal target audience position. In addition to the position, we also considered the resolution size provided by the camera. From the considerations made, we used a webcam from Logitech with the C270 HD series. The selection of this use is based on a camera that can display a resolution of 16: 9 1280 x 720. Then, for the distance, we took the test at a distance of 1 - 4 meters and camera was placed at a height of 2 meters.

After completing the design phase, the next step is the development phase. We must go through several stages in the development of Unity Engine software: camera to material, hand tracking and pose estimation with mediapipe, hand tracking to 3D object, and virtual object interaction.

A. Camera to Material

The process of converting camera input into material within the Unity Engine begins at the beginning of the development process. The process of converting camera input into materials in Unity Engine starts early in the development process. We use the CoreModule and ImgProcModule from openCV in our implementation. CoreModule and ImgProcModule in OpenCV is a vital library component, offering a suite of functions and data structures tailored for image processing from computer vision. It supports mathematical operations on matrices, including addition, subtraction, multiplication, and division, and image manipulation tasks such as resizing, cropping, and flipping. The module also features basic filtering techniques like Gaussian and median blur, geometric transformations for rotation, scaling, translation, and shape drawing functions. Additionally, it facilitates colour space such as Grayscale, RGB, BGR, RGBA and BGRA.

B. Hand Tracking and Hand Pose Estimation

After some process of converting camera input into material in Unity Engine, the next step is to add a hand tracking point into the material. At this stage, we use the model explained in the previous chapter (Fig 2 and Table 1). The flow in using the hand tracking module is presented in Fig. 3.

Fig 3. Hand Tracking Process

A brief explanation of Fig 3: first, the image from the camera is converted into a desired format. After that, the MediaPipePalmDetector module is used to identify the presence of palms in the image. Next, the MediaPipeHandPoseEstimator is used to analyze and determine the pose of the detected hand. Finally, the identified hands are visualized on the original image as material.

C. Hand Tracking to 3D Game Object

Once the key points on the hand are identified, such as fingertips, joints, and other important points, these points are used to build a skeletal model that accurately reflects the user's hand movements in real time. Skeleton model allows tracking of each precise representation of hand movements in real-time.

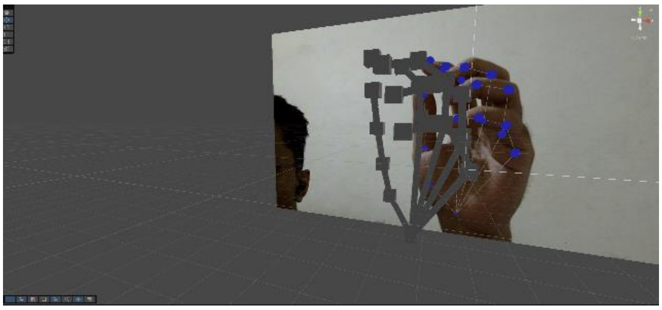

Fig 4. Hand Tracking to 3D Game Object

In short, from the hand tracking and hand pose estimation sections, this phase will visualize hand movements through a three-dimensional skeletal model and allow real-time adjustments to improve user interaction. The intended interaction is interaction with virtual objects that do not exist in the real world.

D. Virtual Object Interaction

From the previous stage, we experimented with virtual object interaction. From the experiments carried out, the prepared triggers can be detected. This further strengthens the previously stated theory that by using openCV, we can develop more interactive AR.

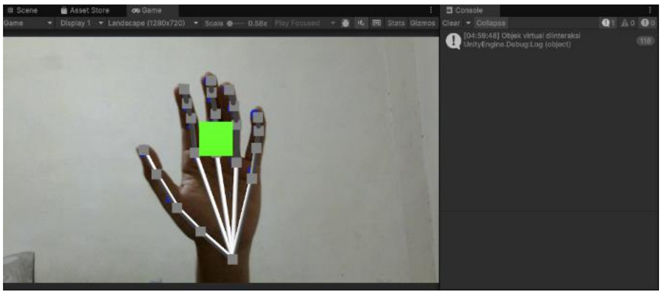

Fig 5. Virtual Object Interaction

Adding a simple script, rigid body, and collider to the virtual object that will be interacted with in the development stage is successfully carried out. The final step of the development stage is to include the virtual objects that have been prepared.

The existing assets are then imported into the project. The Virtual Animal used as a sample is a peacock. From the initial observation results, we saw that peacocks often experience violence from visitors. Starting from screams from visitors, flash lights from cellphone cameras, and beatings of peacocks. The appearance of the project to be tested can be seen in Fig 6.

Fig 6. Final Prototype

After completion of the merger, we entering the next stage. We will collect data for optimal distance and also detection time of hand recognition. Of the ten people we sampled, the test data can be seen in Table 2.

TABLE 2
RANGE AND TIME DATA SAMPLE

Sample	Range and Time			
	1 Meters	**2 Meters**	**3 Meters**	**4 Meters**
1	0,099	0,5	1,6	-
2	0,198	0,59	0,19	-
3	0,696	0,7	0,11	-
4	0,292	0,3	0,9	-
5	0,3	0,9	1,803	10,57
6	0,301	0,19	1,7	8,795
7	0,604	1,3	0,208	-
8	0,39	0,123	0,199	9.2
9	0,098	0,38	1,39	-
10	0,2	0,79	0,5	12,604

From the data that we can see in Table 2, at a distance of 4 meters, there are 6 people whose hands cannot be detected. In addition, the detection time obtained by the other four people was quite high. From these results, we judged that the distance of 4 meters for hand detection conditions was not optimal, so we concluded that 3 meters was the optimal distance for hand detection by cameras.

In the final stage of evaluation, we concluded that the fastest time was 0.098 seconds, and the optimum speed was 3 meters. As a visualization, we display the results of the data processing in Figure 6.

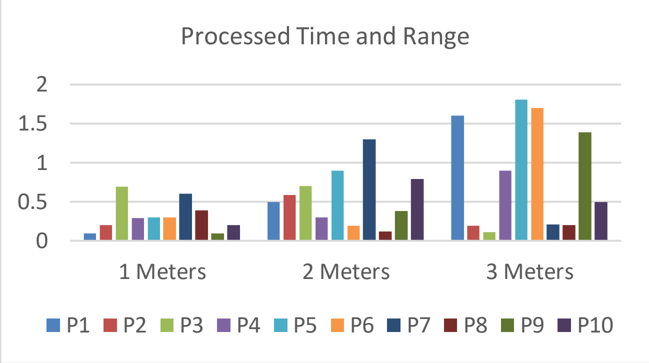

Fig 7. Processed Time and Range

E. Evaluation

From the results of the previous stages, it becomes a reference for us to adjust the camera position in order to get the optimal position for the user. The evaluation will be carried out on 50 people, with 5 optional answers and in order to protect the privacy of our users, we do not collect data after taking pictures. The users we choose are people who like peacocks. Next, the data we get will be calculated using the Likert scale formula $L = T \times Pn$.

From the formula written, L explains Likert, while T explains the total number of respondents who chose, then Pn is for the number of choices from the Likert score.

The results for the first question (Q1), 22 person answered strongly agree, 14 person answered agree, 9 people answered neutral, and 5 person answered disagree. Next one is second question (Q2), 25 person answered strongly agree, 11 person answered agree, and 14 person answered neutral. The third question (Q3), 35 person answered strongly agree and 15 person answered agree. After that, we using Likert formula written earlier to calculate score and obtained 203 point on Q1, 211 point on Q2, and 235 on Q3.

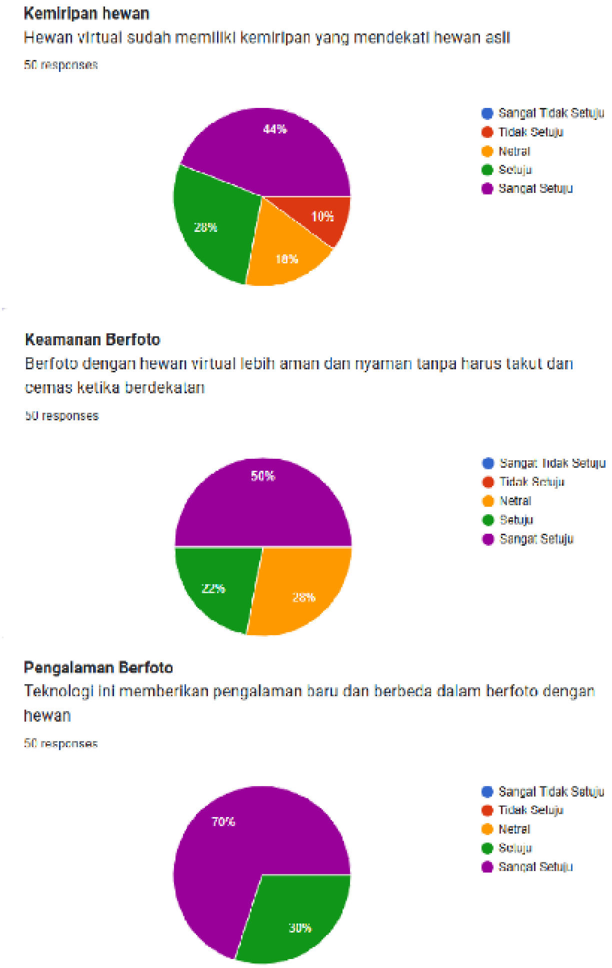

Fig 8. Respondent to Q1, Q2 and Q3

After getting the score, next we will look for the percentage of the score obtained. We use the Likert percentage formula $P = \frac{\sum R}{N} \times 100\%$ and we got the calculation was $P = \frac{649}{750} \times 100\%$ and the results of user satisfaction obtained from the calculation were 86.53% or strongly satisfied.

IV. Conclusion And Discussion

From the research, the answers to the questions in the Research Question (RQ) have been fully answered. From the research conducted, we found that there are four important AR development stages that must be carried using Unity Engine and OpenCV. These steps namely: camera to material, hand tracking and pose estimation with mediapipe, hand tracking to 3D objects, and virtual object interaction. Next for RQ2 we got a level of approval from users of 86.53% or strongly agree with what we are doing. Then for the last RQ, we got the results for optimal hand detection distance is 3 meters, and the speed with the fastest value is 0.098s.

As a discussion material, further research is expected to conduct testing on the light used. We think there is a possibility that the light conditions can affect the test results. We hope this research will continue and evolve over time, ultimately leading to significant advances in augmented reality development. We envision a future where the findings and innovations from this research will drive rapid progress, enabling developers and researchers to create more sophisticated and immersive augmented reality experiences. By fostering a collaborative environment and encouraging continued exploration, we hope to contribute to a dynamic landscape where augmented reality technology can develop quickly and effectively, meeting the growing demands of users and the industry.

Acknowledgments

This study was supported by Kemendikbud Ristek Dikti Republic of Indonesia No. 107/E5/PG.02.00.PL/2024, LLDIKTI V No. 0609.20/LL5-INT/AL.04/2024 and Universitas Amikom Yogyakarta No. 015/KONTRAK-LPPM/AMIKOM/VI/2024.

References

[1] K. Khan, “Advancements and Challenges in 360 Augmented Reality Video Streaming: A Comprehensive Review,” *International Journal of Computing*, vol. 13, no. 1, pp. 1–20, 2024.

[2] V. Nisi, S. James, P. Bala, A. Del Bue, and N. J. Nunes, “Inclusive Digital Storytelling: Artificial Intelligence and Augmented Reality to Re-centre Stories from the Margins,” in *Interactive Storytelling: 16th International Conference on Interactive Digital Storytelling, ICIDS 2023, Kobe, Japan, November 11–15, 2023, Proceedings, Part I*, Berlin, Heidelberg: Springer-Verlag, 2023, pp. 117–137. doi: 10.1007/978-3-031-47655-6_8.

[3] D. C. Villagran-Vizcarra, D. Luviano-Cruz, L. A. Pérez-Dom\'\inguez, L. C. Méndez-González, and F. Garcia-Luna, “Applications analyses, challenges and development of augmented reality in education, industry, marketing, medicine, and entertainment,” *Applied Sciences*, vol. 13, no. 5, p. 2766, 2023.

[4] H. Altınpulluk, “Current Trends In Augmented Reality And Forecasts About The Future,” Dec. 2017, pp. 3649–3655. doi: 10.21125/iceri.2017.0986.

[5] Caraka Aji Pranata, M Fairul Filza, and Akbar, “Systematic Literature Review dengan Mengidentifikasi Software serta Metode Pengembangan Augmented Reality,” *Indonesian Journal of Computer Science*, vol. 13, no. 1, Feb. 2024, doi: 10.33022/IJCS.V13I1.3742.

[6] J. Simon, “Augmented Reality Application Development using Unity and Vuforia,” *Interdisciplinary Description of Complex Systems*, vol. 21, no. 1, pp. 69–77, 2023, doi: 10.7906/INDECS.21.1.6.

[7] C. A. Pranata, E. Utami, E. T. Luthfi, and U. Amikom Yogyakarta, “Marker Based Augmented Reality Pada Buku Poa Dengan Metode Fast Corner Detection,” *EXPLORE*, vol. 11, no. 2, pp. 58–64, Jul. 2021, doi: 10.35200/EXPLORE.V11I2.461.

[8] Mhd. A.-G. Sinaga and M. Alda, “Penerapan Algoritma Fast Corner Dalam Perancangan Media Pembelajaran Awan Menggunakan Augmented Reality Berbasis Android,” *Journal Of Science And Social Research*, vol. 7, no. 2, pp. 463–472, May 2024, doi: 10.54314/JSSR.V7I2.1860.

[9] P. Shukla and P. Das, “Enhancing Human-Computer Interaction: Hand Detection for Air Writing Utilizing NumPy and OpenCV,” *Proceedings - International Conference on Technological Advancements in Computational Sciences, ICTACS 2023*, pp. 517–521, 2023, doi: 10.1109/ICTACS59847.2023.10390179.

[10] "Vuforia Engine Overview | Vuforia Library." Accessed: Dec. 26, 2024. [Online]. Available: https://developer.vuforia.com/library/getting-started/vuforia-features

[11] M. A. Nasirudin, M. F. M. Fudzee, N. Senan, C. S. C. Dalim, D. Witarsyah, and A. Erianda, "Systematic Literature Review on Augmented Reality with Persuasive System Design: Application and Design in Education and Learning," *JOIV: International Journal on Informatics Visualization*, vol. 8, no. 2, pp. 862–873, 2024.

[12] T. H. Chandhan, K. K. R, N. Raj, N. N. K. Reddy, and M. Z. A N, "Air Canvas: Hand Tracking Using OpenCV and MediaPipe," *SSRN Electronic Journal*, Aug. 2023, doi: 10.2139/SSRN.4527784.

[13] C. Luna-Jiménez, M. Gil-Martín, R. Kleinlein, R. San-Segundo, and F. Fernández-Martínez, "Interpreting Sign Language Recognition using Transformers and MediaPipe Landmarks," *ACM International Conference Proceeding Series*, pp. 373–377, Oct. 2023, doi: 10.1145/3577190.3614143.

[14] F. Achmad, L. Septiningrum, Y. Prambudia, A. A. Rumanti, and I. I. Wiratmadja, "Tourism Industry Development Model through Support System Facilities, Stakeholder Participation, and Environmental Dynamism: Dealing with Post-Pandemic Covid-19 Approach," *International Journal of Innovation, Management and Technology*, vol. 15, no. 1, 2024.

[15] K. I. Putra, P. J. L. Dawa, Y. D. Burgos, and F. I. Maulana, "Implementation of Augmented Reality in Study for Human Anatomy," *Procedia Comput Sci*, vol. 227, pp. 709–717, 2023, doi: https://doi.org/10.1016/j.procs.2023.10.575.

[16] C. Sukotjo, D. E. Bertucci, J. Y. Patel, J. C.-C. Yuan, and M. Santoso, "Evaluating augmented reality e-typodont to improve a patient's dental implant health literacy," *J Prosthet Dent*, 2023.

[17] S. Jiang, B. Molye, R. Yung, L. Tao, and N. Scott, "Augmented Reality and the Enhancement of Memorable Tourism Experiences at Heritage Sites," *Current Issues in Tourism*, Dec. 2022, doi: 10.1080/13683500.2022.2026303.

[18] "Tentang Kami | Gembira Loka Zoo." Accessed: Dec. 19, 2024. [Online]. Available: https://gembiralokazoo.com/about-us

[19] C. Collins, S. McKeown, and R. O'Riordan, "A comprehensive investigation of negative visitor behaviour in the zoo setting and captive animals' behavioural response," *Heliyon*, vol. 9, no. 6, p. e16879, Jun. 2023, doi: 10.1016/J.HELIYON.2023.E16879.

[20] H. Abuhassna, S. Alnawajha, F. Awae, M. A. Bin Mohamed Adnan, and B. I. Edwards, "Synthesizing technology integration within the Addie model for instructional design: A comprehensive systematic literature review," *Journal of Autonomous Intelligence*, vol. 7, no. 5, p. 1546, May 2024, doi: 10.32629/JAI.V7I5.1546.

[21] V. H. Pranatawijaya, W. Widiatry, R. Priskila, and P. B. A. A. Putra, "Penerapan Skala Likert dan Skala Dikotomi Pada Kuesioner Online," *Jurnal Sains dan Informatika*, vol. 5, no. 2, pp. 128–137, Dec. 2019, doi: 10.34128/jsi.v5i2.185.

[22] "Hand Gesture Recognition using OpenCV | IEEE Conference Publication | IEEE Xplore." Accessed: Dec. 19, 2024. [Online]. Available: https://ieeexplore.ieee.org/document/10112341