The International Journal of Engineering and Science (IJES)
|| Volume || 13 || Issue || 12 || Pages || PP 53-64 || 2024 ||
ISSN (e): 2319 – 1813 ISSN (p): 2319 – 1805

Image Matching Approach Integrated To Image Evaluation Techniques To Generate Augmented Reality In The Open Area Environments

Fatih KUCUKUYSAL
[1] *Turkish Aerospace Industries, Ankara, Turkey*
Corresponding Author: Fatih KUCUKUYSAL

ABSTRACT
Positioning of a XR/VR glasses with high accuracy in the open area environments and augmentation of simulated objects refers to the problem to give the real environment feeling during the training in a real environment to the trainees. In this article, an image matching approach will be proposed for generating a high accuracy image matching algorithm to use in the simulation systems while maximizing interactivity in the training area using by locating the glasses user with high accuracy. Current methods use sensors and cameras of the glasses, to locate the glasses in the environment that can give erroneous results while the data generated by the sensors. This paper proposes a technique that uses image matching algorithms to locate glasses by using previously given images with a known position. HARRIS and SIFT algorithms will be compared in robustness, implementation side for usage in real time on glasses.

KEYWORDS; Augmented reality, HARRIS, SIFT, image matching, military, training, XR, VR

Date of Submission: 03-12-2024 Date of acceptance: 14-12-2024

I. INTRODUCTION

Real environment feeling during the training is a required capability to maximize the effectivity of the training. Especially, in military trainings, the concept of "train as you fight" is the main objective [1][2]. If augmented reality technology is used, a problem arises to locate the augmented objects at an accurate position in the real three-dimensional environment. However, existing methodologies use sensors and cameras of the glasses to locate the trainee and the glasses that the trainee uses, which are dependent on sensor error rates and image processing failures [3].

Currently, virtualization of the combat environment is carried out in flight simulators and to some extent through XR (Extended Reality) simulators by combining the real cockpit virtual environment with glasses [4][5]. However, these solutions are generally heavy, non-ergonomic, and based on completely recreating the image taken from the camera by blocking the field of view and giving it back to the glasses' screens [6].

In the existing products, since the reference information about the environment where the user is located is not included and used in the products, the processes of positioning the glasses and combining the images of both environments cannot be done realistically, especially in open areas [7]. Existing systems are limited in terms of interaction and do not provide full interaction. Therefore, they cannot reach the necessary prevalence on the user side, which has led us to design research that meets all these needs [8].

Here, **Our problem** is positioning errors in the use of virtual environment glasses in open areas with high accuracy.

A technique that can reduce positioning errors for glasses is image matching using invariant interest point detection. Among the most widely known techniques, SIFT (Scale-Invariant Feature Transform) and Harris Corner Detection are commonly compared for their robustness and accuracy [15][16]. SIFT offers superior performance in feature matching due to its invariance to scale, rotation, and illumination changes, while Harris is often preferred for its computational efficiency in real-time applications [15][16].

II. MATERIALS AND METHODS

In the following sections;

- Tested Methodologies,
- Working Principles of Tested Methodologies,

- Target Platform for Implementation of Design and
- Our Team and Reasons Motivate Us for This Approach will be given.

II.1. Tested Methodologies

In this design research, image matching approach for matching the images in the database and the images in the field of view of the glasses will be tested. Scale Invariant Interest Points are the searched points in the images and these Scale Invariant Interest Points are matched in between the images [15][16].

Two mostly known "Scale Invariant Interest Points" detection methods are examined. These methods are;

- "Scale Invariant Feature Transform (SIFT)" and
- "Harris Corner Detection (HARRIS)" used for interest point detection.

II.1.1. Working Principles of Tested Methodologies

HARRIS and SIFT algorithms are compared in robustness, implementation side for usage in real time on glasses.

For an exact positioning of the goggle in the real environment; in addition to positioning sensors image processing methodologies are used. This additional process guarantees the exact positioning probability quality increasing. Also, these processes guarantee the true augmentation of the additional images to be overlayed in the image at the exact position in the Glasses' camera Field of Regard.

For Feature matching, matching images in the database, finding a place in an image, Object Recognition; Detect feature point (key points) and Match these points in different images is implemented. For this purpose, below techniques are available:

- Harris corner
- Tomasi's "good features to track"
- SIFT: Scale Invariant Feature Transform
- SURF: Speeded Up Robust Feature
- GLOH: Gradient Location and Orientation Histogram ...

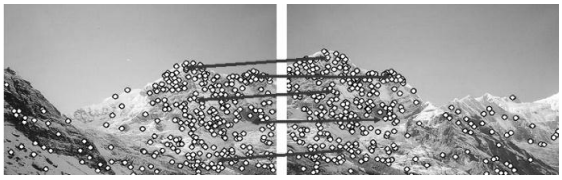

Fig. 1 Image Matching [27]

From these techniques, we have tested the HARRIS Corner and SIFT (Scale Invariant Feature Transform) for a better comparison

HARRIS uses shifted corners that produce some difference in the image that looks for large difference in shifted image. SIFT (Scale Invariant Feature Transform) is another technique used to detect Invariant local features that has an algorithm for finding points and representing their patches should produce similar results even when conditions vary. These "invariance" includes geometric invariance: translation, rotation, scale and photometric invariance: brightness, exposure… (7)

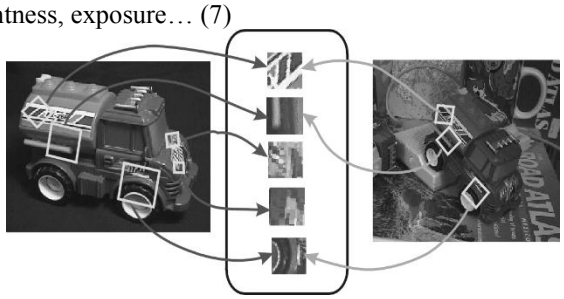

Fig. 2 Image Matching [27]

Suppose you're looking for corners. Key idea is finding a scale that gives local maximum of f. Here, f is a local maximum in both position and scale where f is Laplacian or difference between two Gaussian filtered images with different sigmas. (8)

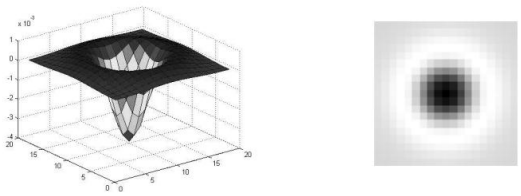

Fig. 3 Difference of Gaussians [27]

While HARRIS is invariant to translation and rotation, Scale is trickier and we need to detect features at many scales using a Gaussian pyramid but SIFT can find "the best scale" to represent each feature.

We can design an invariant feature descriptor that captures the information in a region around the detected feature point. The simplest descriptor: a square window of pixel. An example, invariance in Scale Invariant Feature Transform (9):

- We Take 16x16 square window around detected feature
- Compute edge orientation (angle of the gradient - 90°) for each pixel
- Throw out weak edges (threshold gradient magnitude)
- Create histogram of surviving edge orientations (9)

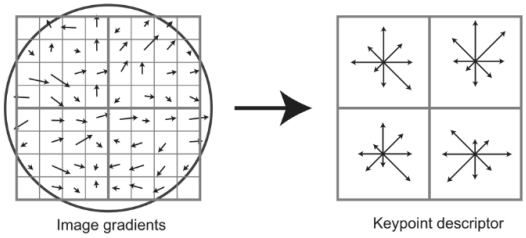

Fig. 4 Keypoint Descriptors [29]

A key point descriptor is created by first computing the gradient magnitude and orientation at each image sample point in a region around the key point location, as shown on the left. These are weighted by a Gaussian window, indicated by the overlaid circle. These samples are then accumulated into orientation histograms summarizing the contents over 4x4 subregions, as shown on the right, with the length of each arrow corresponding to the sum of the gradient magnitudes near that direction within the region. This figure shows a 2x2 descriptor array computed from an 8x8 set of samples, whereas the experiments in this paper use 4x4 descriptors computed from a 16x16 sample array. (9)

For a SIFT descriptor, basic idea: (10)

- Take 16x16 square window around detected feature
- Compute edge orientation (angle of the gradient - 90°) for each pixel
- Throw out weak edges (threshold gradient magnitude)
- Create histogram of surviving edge orientations
- Full version: (10)
- Divide the 16x16 window into a 4x4 grid of cells (2x2 case shown below)
- Compute an orientation histogram for each cell
- 16 cells * 8 orientations = 128 dimensional descriptor

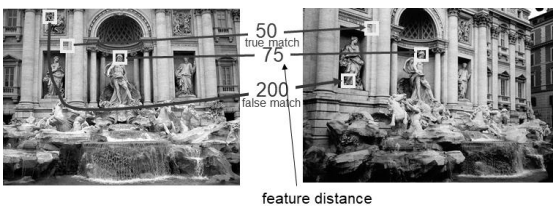

Fig. 5 True/false positives matching [28]

Properties of SIFT (11)

- Extraordinarily robust matching technique
- Can handle changes in viewpoint
  - Up to about 60 degree out of plane rotation
- Can handle significant changes in illumination
  - Sometimes even day vs. night (below)
- Fast and efficient—can run in real time

II.2. Tested Images Used to Test

Tested images are;

- 2 images of the same scene (800x600 pixels resolution in gray scale) with different scaling ratio and,
- 10 images of the same scene (with 800x600 pixels resolution in gray scale) with different rotation angles are used to test the methods in different scaling and rotation conditions [15][16].

II.3. Test Software and Formulas

MATLAB Program, is used to test the algorithms generated for the methods.

Used formulas in MATLAB for this research can be seen below:

$$E(x,y) = \sum_{(x_i,y_i)\in W} [I(x_i,y_i) - I(x_i+\Delta x, y_i+\Delta y)]^2 \quad \text{Eq. (1)}$$

$$\begin{aligned} E(\mathbf{x}) &= \sum_W [I_x\Delta x + I_y\Delta y]^2 \\ &= \sum_W [I_x^2\Delta^2 x + 2I_xI_y\Delta x\Delta y + I_y^2\Delta^2 y] \\ &= (\Delta\mathbf{x})^\top \mathbf{A}(\mathbf{x})\Delta\mathbf{x} \end{aligned} \quad \text{Eq. (2)}$$

$$\mathbf{A} = \begin{bmatrix} \sum_W I_x^2 & \sum_W I_xI_y \\ \sum_W I_xI_y & \sum_W I_y^2 \end{bmatrix} \quad \text{Eq. (3)}$$

$$\det \mathbf{A} - \alpha(\operatorname{tr}\mathbf{A})^2 = \lambda_1\lambda_2 - \alpha(\lambda_1+\lambda_2)^2 \quad \text{Eq. (4)}$$

Look at Euclidean distance between feature vectors to match.

$$d(\mathbf{u},\mathbf{v}) = \left(\sum_i (u_i - v_i)^2\right)^{1/2} \quad \text{Eq. (5)}$$

II.4. Target Platform for Implementation of Design

Since this study is a design research, a design concept including these Scale Invariant Interest Points detection methodologies is generated.

It is planned to develop a Virtual Environment Glasses Platform concept that will enable virtualization of real situations or virtual creation of unreal situations in many fields such as education or military purposes. The sensors on the glasses will be able to detect head movements and react accordingly. The glasses will include a camera on the front and will cover in real time.

Using Image matching approach and attached sensors on the hardware design of the glasses, will be expected to give place of the glasses in the three-dimensional space by global positioning and environment interaction independent of external devices

The images obtained via the camera in front of the glasses are evaluated with image processing algorithms using OpenCV library-based software, and accordingly, hand movements are converted into commands, and body and environment movements are converted into digital data.

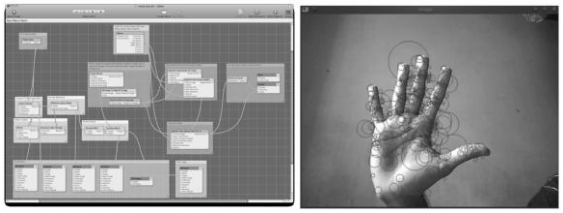

Fig. 6 OpenCV Based Image Evaluation Software [24], [31]

The environment images are subjected to image processing to precisely position the user of the glasses in the 3D environment and to position the images to be added to this environment with the Augmented Reality method exactly where they should be. The environment data in the database is compared with the data from the camera and the positioning and image superimposition processes are performed. An example of image placement in a land environment can be seen in the image below.

Fig. 7 Example of Positioning the Virtual Image in the Real Environment [25]

An example of image placement in an aerial environment can be seen in the image below.

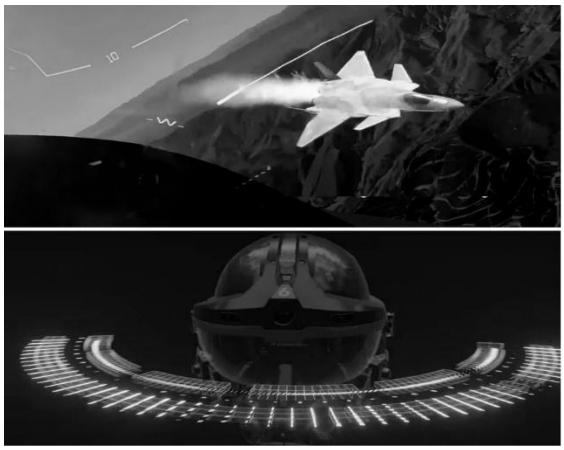

Fig. 8 Example of RED6 Company's Military Application with Helmet [26]

II.5. Our Team and Reasons Motivate Us for This Approach

For educational or military purposes; although there is a need to simulate the relevant environment in real terms, we felt the need to create a new concept because the existing systems are limited in terms of interaction. First of all, the process that led to the solution, which is the reason why existing systems cannot fully meet the need and we added different elements as capabilities;

- Global Positioning and Environment Interaction Independent of External Devices
- Holistic Solution in Battlefield Virtualization and Visualization
    - Providing Full Interactive Military Environment Integration
    - Military Units Environment Integration
    - Combat Operation Centers Environment Integration
    - Air/Land/Sea Platforms Environment Integration

We have been working with academic friends since 2013 in terms of concept creation and design development to develop this solution. Our team consists of 2 engineers working on augmented reality training systems and an academician having expertise on interactive education.

Also, we ranked in the top 10 in the defence category out of approximately 1500 projects in the Middle East Technical University New Ideas New Business Competition. New Ideas New Business (YFYI) is Turkey's first and largest technology-based entrepreneurship program organized by METU TECHNOPARK since 2005.

III. Experimental Results

Here are the experimental results obtained for feature matching. Image matching calculation performance in rotation case for, both HARRIS and SIFT are used for the same images.

We have tested the HARRIS and SIFT algorithms for HARRIS for 30, 60, 90, 120, 150, 180, 210, 240, 270, 300 degrees rotated 10 images. Below results are obtained:

Fig. 9 Harris (Left Images) vs SIFT (Right Images) Rotating

When looked at the images for matched points, values given in the below Table-1 are obtained.

Table 1, for comparing HARRIS and SIFT for Matched Points performance at different Rotation Angles with given below parameters

- Rotation Angle Approximately as Degrees
- HARRIS Matched Points as Quantity
- SIFT Matched Points as Quantity

Test No.	HARRIS versus SIFT Rotating		
	Rotation Angle Approximately (degrees)	**HARRIS Matched Points**	**SIFT Matched Points**
1	30	7	26
2	60	2	25
3	90	2	18
4	120	3	30
5	150	3	27
6	180	3	33
7	210	2	25
8	240	3	23
9	270	4	23
10	300	5	33

Table 1 (Harris vs SIFT Rotating – Matched Points versus Rotation Angle)

Also, we have tested the algorithms for the images taken with nearly 1/5 scaling. Image matching calculation performance in scaling case, both HARRIS and SIFT are used for the same images.

HARRIS algorithm gave the below matching result for the images taken with approximately 1/5 scaling:

Fig. 10 HARRIS Algorithm Scaling

SIFT algorithm gave the below matching result for the images taken with approximately 1/5 scaling:

Fig. 11 SIFT Algorithm Scaling

IV. Conclusions

Analysing the experimental results for compared image matching methodologies to precisely place augmented images demonstrates the importance of matching database images with real-time images captured by the glasses.

Test results for rotating case HARRIS gave 34 matched points totally while SIFT gave 263 matched points. Test results for scaling case HARRIS gave 5 matched points totally while SIFT gave 1 matched points.

These results gave us information about SIFT algorithm is better in feature matching in small view angle changes. Also having lots of corners in the images, HARRIS gave more matched points being efficient to detect corner points to be used for feature matching. But, it is not always possible to have lots of corners in an image, in open areas. Looking at the values for matched points quantity in our tests, and comparing lots of images with the algorithms, experimental studies shows us that the SIFT algorithm is better than HARRIS, when compared, as given with below capabilities:

- Easy to compute and match

- Accurate matching for small view angle changes
- Better in feature matching in small view angle changes
- Robust for illumination changes

The findings align with previous studies that indicate SIFT's robustness in small angle and illumination changes [15][16]. While HARRIS is computationally simpler and faster, SIFT's ability to adapt to varying scales and rotations has been widely recognized in image processing research [17][18].

Using an image matching approach to support the glasses' positioning data reduces sensor-generated errors. Improved positioning accuracy enhances the interactivity of training systems by ensuring augmented objects are precisely located within the real three-dimensional environment [9][15].

This increased interactivity accuracy also provides better experience for implementation areas;

- Battlefield virtualization and visualization
- Fully interactive military environment integration
- Combat operations centers environment integration
- Air/land/sea platforms environment integration

Existing systems restrict trainees to static images and limited areas. The proposed system overcomes these limitations by precisely determining the glasses' location and perspective using onboard sensors (compass, accelerometer, GPS/GNSS). Augmented images are seamlessly overlaid on the real environment, allowing soldiers to train in dynamic, large-scale environments [13][22].

Fig 12 Military Training in Limited Areas [30]

REFERENCE

[1]. [Azuma, R. T. (1997). A survey of augmented reality. Presence: Teleoperators and Virtual Environments, 6(4), 355-385. https://doi.org/10.1162/pres.1997.6.4.355

[2]. Billinghurst, M., Clark, A., & Lee, G. (2015). A survey of augmented reality. Foundations and Trends in Human-Computer Interaction, 8(2-3), 73-272. https://doi.org/10.1561/1100000049

[3]. Milgram, P., & Kishino, F. (1994). A taxonomy of mixed reality visual displays. IEICE Transactions on Information and Systems, E77-D(12), 1321-1329.

[4]. Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. Human Factors, 37(1), 32-64. https://doi.org/10.1518/001872095779049543

[5]. Sherman, W. R., & Craig, A. B. (2018). Understanding Virtual Reality: Interface, Application, and Design. Morgan Kaufmann.

[6]. Zhou, F., Duh, H. B. L., & Billinghurst, M. (2008). Trends in augmented reality tracking, interaction, and display: A review of ten years of ISMAR. Proceedings of the IEEE International Symposium on Mixed and Augmented Reality (ISMAR), 193-202. https://doi.org/10.1109/ISMAR.2008.4637362

[7]. Wang, X., & Dunston, P. S. (2013). Mixed reality technology applications in construction equipment operator training. Automation in Construction, 27, 111-122. https://doi.org/10.1016/j.autcon.2012.05.011

[8]. Cummings, M. L., & Guerlain, S. (2007). Developing operator capacity estimates for supervisory control of autonomous vehicles. Human Factors: The Journal of the Human Factors and Ergonomics Society, 49(1), 1-13. https://doi.org/10.1518/001872007779598060

[9]. Microsoft. (2020). HoloLens 2: Technology overview. Retrieved from https://www.microsoft.com/hololens2

[10]. Zhang, Y., & Yang, Y. (2021). Augmented reality applications in training and education. Journal of Advanced Reality Technology, 12(3), 45-58. https://doi.org/10.1016/j.jart.2021.03.002

[11]. Pico Interactive. (2022). Pico 4 Enterprise: Specifications and use cases. Retrieved from https://www.pico-interactive.com

[12]. HTC. (2022). Vive XR Elite: Transformable all-in-one VR headset. Retrieved from https://www.vive.com

[13]. Epson. (2021). Moverio BT-40 and BT-45CS: Augmented reality smart glasses. Retrieved from https://www.epson.com

[14]. VARJO. (2022). XR-4: High-performance extended reality headset. Retrieved from https://www.varjo.com

[15]. Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, 60(2), 91-110. https://doi.org/10.1023/B:VISI.0000029664.99615.94

[16]. Harris, C., & Stephens, M. (1988). A combined corner and edge detector. Proceedings of the Alvey Vision Conference, 23(1), 147-151. https://doi.org/10.5244/C.2.23

[17]. Leap Motion. (2019). Hand tracking for immersive technologies. Retrieved from https://www.leapmotion.com

[18]. Xu, W., Huang, Z., Li, B., & Zhang, J. (2020). Hand gesture recognition with depth sensor and neural networks for augmented reality applications. Journal of Visual Communication and Image Representation, 70, 102816. https://doi.org/10.1016/j.jvcir.2020.102816

[19]. Starke, S., & Kruse, S. (2021). A machine learning approach for hand gesture recognition in mixed reality. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 29, 987-998. https://doi.org/10.1109/TNSRE.2021.3070339

[20]. Lindeberg, T. (1998). Feature detection with automatic scale selection. International Journal of Computer Vision, 30(2), 79-116. https://doi.org/10.1023/A:1008045108935
[21]. Bay, H., Tuytelaars, T., & Van Gool, L. (2006). SURF: Speeded Up Robust Features. Proceedings of the European Conference on Computer Vision (ECCV), 404-417. https://doi.org/10.1007/11744023_32
[22]. Starke, S., & Kruse, S. (2021). A machine learning approach for hand gesture recognition in mixed reality. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 29, 987-998. https://doi.org/10.1109/TNSRE.2021.3070339
[23]. Zhang, Y., & Yang, Y. (2021). Augmented reality applications in training and education. Journal of Advanced Reality Technology, 12(3), 45-58. https://doi.org/10.1016/j.jart.2021.03.002
[24]. Image Source from website (Is This the Future of Augmented Reality Gaming? (kineme.net))
[25]. Image Source from website (SURF in OpenCV | Achu's TechBlog (roadtovr.com))
[26]. Image Source from website (AR Training in the Sky With Red 6 | Spotlight | Unreal Engine (youtube.com)
[27]. Image Source from website (s-ln.in/2013/04/18/hand-tracking-and-gesture-detection-opencv/)
[28]. Image Source from (Lecture 1: Images and image filtering (cornell.edu) CS5670: Computer Vision: Noah Snavely - page25
[29]. Image Source from (David G. Lowe, "Distinctive image features from scale-invariant keypoints," International Journal of Computer Vision, 60, 2 (2004), pp. 91-110.
[30]. Image Source from website (Virtual Reality: The Future of Military Training – Finabel)
[31]. Image Source from website (SURF in OpenCV | Achu's TechBlog (wordpress.com)