A Comparative Study of Various Key-point Detector-Descriptor Algorithms for Augmented Reality Applications

[1]Dr. Jayshree R. Pansare, [2]Kanad Naleshwarkar, [3]Nitish Prajapati, [4]Aditya Muley
Department of Computer Engineering, Modern Education Society's College of Engineering, Pune
[1]jayshree.pansare@mescoepune.org, [2] kanad.naleshwarkar@gmail.com, [3] prajapati.nitish.np@gmail.com, [4] adityamuley07@gmail.com

***Abstract*— This study aims to compare various feature detector-descriptor algorithms. The algorithms compared are ORB, SIFT, SURF, BRISK, KAZE and AKAZE. The methodology used in this analysis segregates fetching of the frames (from video input) from the actual processing so that time consuming I/O operation does not affect time taken by each algorithm to process the input. This analysis shows that ORB is faster than remaining algorithms with frame processing rate of 23.9 Frames per Second (FPS), while SIFT is more accurate than others for feature detection and description. The results of this analysis can then be used for Augmented Reality applications implementing one of these algorithms.**

Index Terms— ***Augmented Reality, Feature Detection, Feature Description, ORB, SIFT, SURF.***

I. INTRODUCTION

Augmented Reality is one of the leading technologies in the industry. AR has a lot of potential in industries such as advertising, gaming, production, etc. The term AR refers to the enhancement of physical world with the help of virtual objects. It fuses digital data into the real world to produce a new perception altogether. AR enables users to visualize objects which are physically absent or inaccessible. It eliminates the need for the object to be present at desired location.

What lies at the core of this technology is identifying the surface or object onto which the augmentation is rendered, this (surface or object) is called as a target. Target identification can be performed using marker-based or marker-less approach of AR. In the marker-based approach, the target is a bar code or a QR code. A QR code makes it easier to detect features due to the presence of edges.

In the marker-less approach, the target is an object. The challenge here is detecting features. Features can be specific structures such as edges, points, etc. In this approach, features are first detected and then described in a form which can be used to compare these features with the features of what is there in the live feed of the camera. If these two set of features match, the virtual object is rendered.

There are various algorithms for feature detection and description. This analysis aims to compare these algorithms for an AR application which renders a note onto a surface previously registered by the user. In order to maintain the consistency, we have used a 20 sec video input consisting of 621 frames, for each of the algorithm.

II. Related Work

Scale Invariant Feature Transform (SIFT) [1] keypoint detector and descriptor has been successful for over a decade now for various applications such as Object Recognition, Image Stitching, Vision Mapping and many more. To detect corners/features in different scaled perspectives, SIFT uses scale-space filtering(blurring). The Laplacian of Gaussian(LoG) is used with varying values of scaling parameter which acts as blob detector. The LoG in crude sense, refers to the first derivative of an image. Blob detection is used with different values of scaling parameter 'σ'. But, due to cost complexity of LoG, the Difference of Gaussian(DoG) is used which is an estimation of LoG. The local maxima and minima are then found out by comparing current space's neighboring eight pixels with previous scale space's nine and next scale space's nine pixels.

Speeded Up Robust Features (SURF) [2], which also relies on Gaussian scale space analysis, presents a novel scale and rotation invariant interest point detector and descriptor. SURF is developed for providing both a detector and a descriptor, which are faster to compute without affecting the performance.

Binary Robust Invariant Scalable Keypoints (BRISK) [3] offers detector-descriptor pair aimed at being faster with comparable matching performances with SIFT and SURF. BRISK has two steps:

- Scale-space keypoint detection
- Keypoint description

Once generated, these keypoints can be matched efficiently because of the binary nature of the descriptors.

Even after considering its speed, SIFT is computationally expensive considering its usage in real-time applications. In ORB [4], the authors have proposed an efficient alternative to SIFT having similar performance in terms of matching, and is better suited for real-time applications. The proposed method is developed on well-known FAST keypoint detector and BRIEF keypoint descriptor; hence the name ORB (Oriented FAST and Rotated BRIEF).

The authors also address several limitations of these techniques, the most important of them being the absence of Rotational Invariance in BRIEF.

The important contributions of this paper are:

- Inclusion of orientation element to FAST
- Better computation of oriented BRIEF features

The Gaussian blurring in SIFT blurs details and noise to the same extent, affecting natural boundaries of objects which reduces localization accuracy and distinctiveness. In a newer approach called as KAZE [5], this issue was addressed. In this algorithm 2D features are detected and described in a non-linear scale space using non-linear diffusion filtering as opposed to linear diffusion in previous approaches. The non-linear scale space is built using Additive Operator Splitting (AOS) techniques. Even though this approach is helpful, the algorithm amounts to increase in computational cost due to the complexity involved in building a non-linear scale space. In continuation of this approach, Accelerated-KAZE (AKAZE) [6] exploits benefits of non-linear scale space. In addition, this algorithm speeds-up detection using Fast Explicit Diffusion (FED) and introduces a Modified-Local Difference Binary (M-LDB) descriptor which is highly efficient as claimed by the authors.

A comprehensive study [7] of the detectors and descriptors was proven to be helpful to get the gist of their performances. This study gives a detailed comparison of feature detector and feature descriptor methods and their combinations. The authors have experimented with various detector-descriptor combinations, a total of 23 to be precise. The parameters considered were: accuracy, time, angle difference between keypoints, number of correct matches and distance between correctly matched keypoints. The important conclusions are:

(a) The SIFT-SURF combination (SIFT as keypoint detector and SURF as descriptor) was found to be the most accurate one with 98.41% accuracy

(b) The ORB-BRIEF combination (ORB as keypoint detector and BRIEF as keypoint descriptor) was found to be the computationally fastest one.

Another study of detection-description algorithms [8] introduces a new approach for evaluation of various keypoint detectors. Instead of the accuracy parameter computed with respect to keypoints detected in a single reference image, the authors have proposed to evaluate the keypoints' performance by computing repeatability parameter across a set of images obtained by rotating a single image from 0 to 180˚ with the interval of 1˚

III. Methodology

This experiment was conducted using OpenCV. All the algorithms for detection and description are available. Of all, SIFT and SURF are patented and are available for educational use in specific versions of OpenCV.

The video used for this comparative analysis is first converted to a stream of frames. These frames are later fetched for processing individually. We have used the concept of multithreading in order to make the pipeline faster. One thread is responsible for fetching frames from the video and another thread processes the frames. Fetching frames in parallel saves time.

IV. Use of Threads to Buffer Frames

We are using multithreading and queue data structure to improve the processing rate, FPS. The read() method of cv2.VideoCapture blocks I/O operations, affecting the processing speed. The solution is to create a separate thread to handle the fetching of the frames. This thread fetches frames and adds them in a queue data structure. This queue acts as a buffer from which frames are dequeued.

The reason behind using queue data structure is that it is thread safe in python.

Next, the key-points and descriptor of reference image are calculated. The descriptors of frame and reference image are matched using KNN matcher and good matches are extracted based on 75% neighborhood distance.

These good matches are used to calculate the perspective transformation matrix by finding the homography.

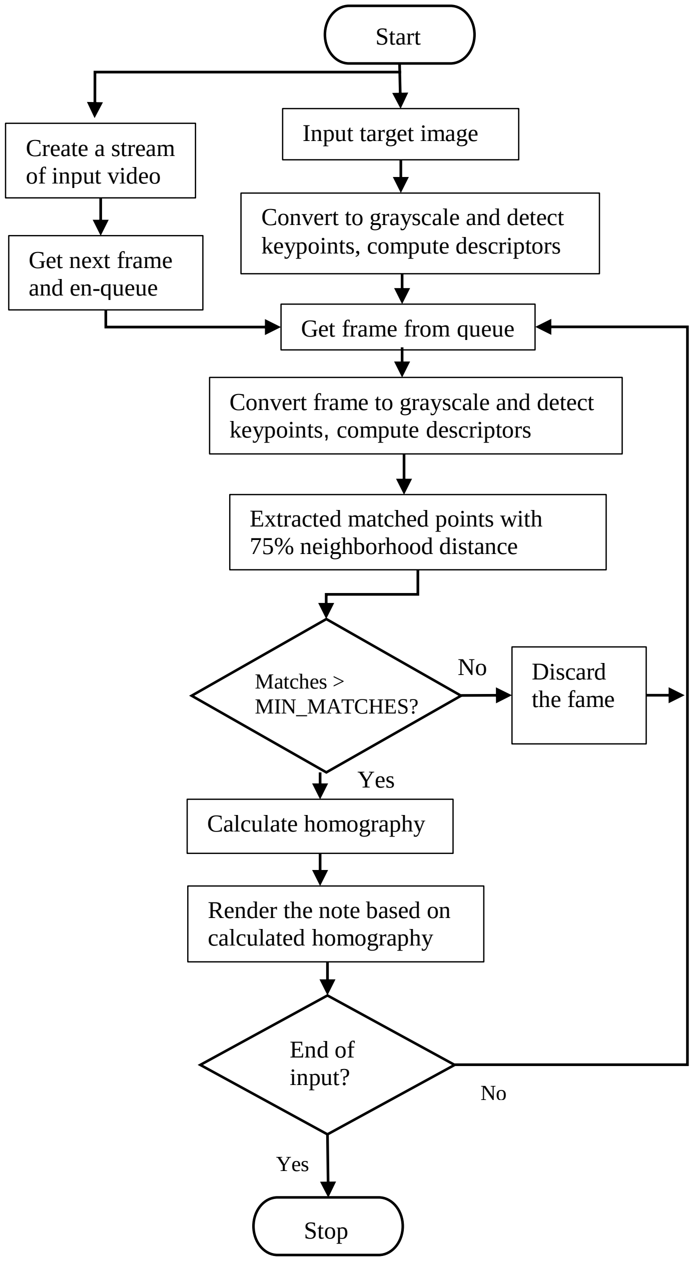

Figure 1: Flowchart of the process

V. RESULTS

Following results were observed for this analysis which used a 20 sec long video containing total of 621 frames.

Elapsed Time:

Algorithm	ORB	SIFT	SURF	BRISK	KAZE	AKAZE
Elapsed Time	25.8	47.8	55.24	58.79	102.9	28.9

Table 1

Frames per Second (FPS):

Algorithm	ORB	SIFT	SURF	BRISK	KAZE	AKAZE
FPS	23.9	12.9	11.24	10.56	6.03	21.49

Table 2

Average Number of Matches:

Algorithm	ORB	SIFT	SURF	BRISK	KAZE	AKAZE
Avg. Matches	37.5	214	178.5	79.79	96.68	23.12

Table 3

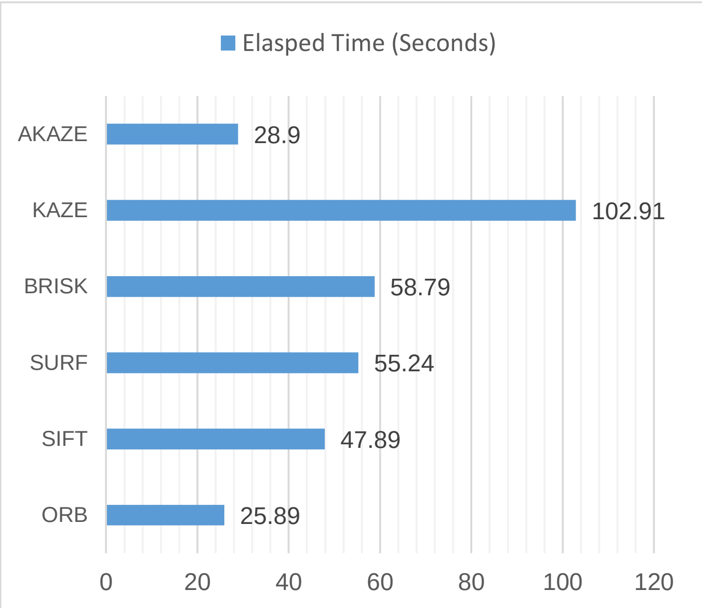

Figure 2: Comparison of Elapsed Time

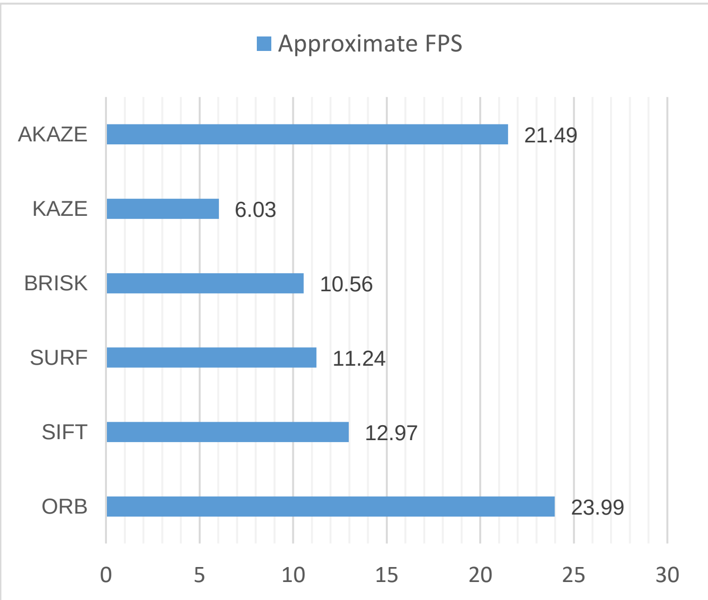

Figure 3: Comparison of FPS

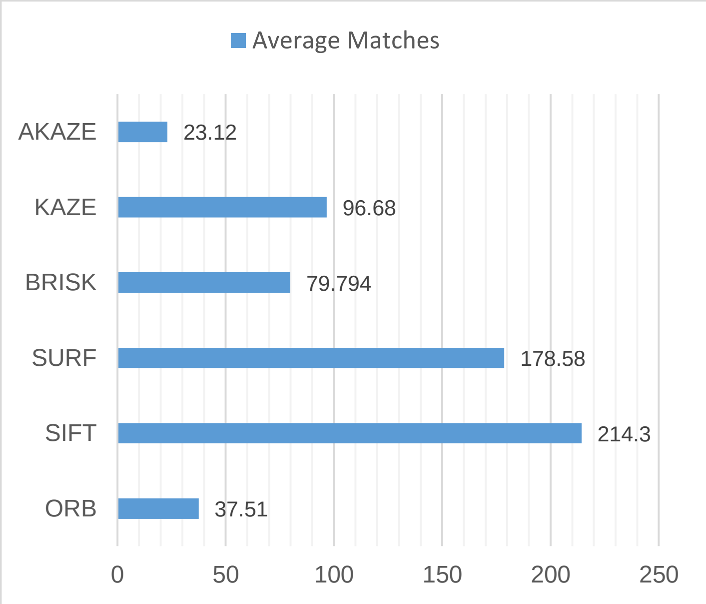

Figure 4: Average Number of Matches

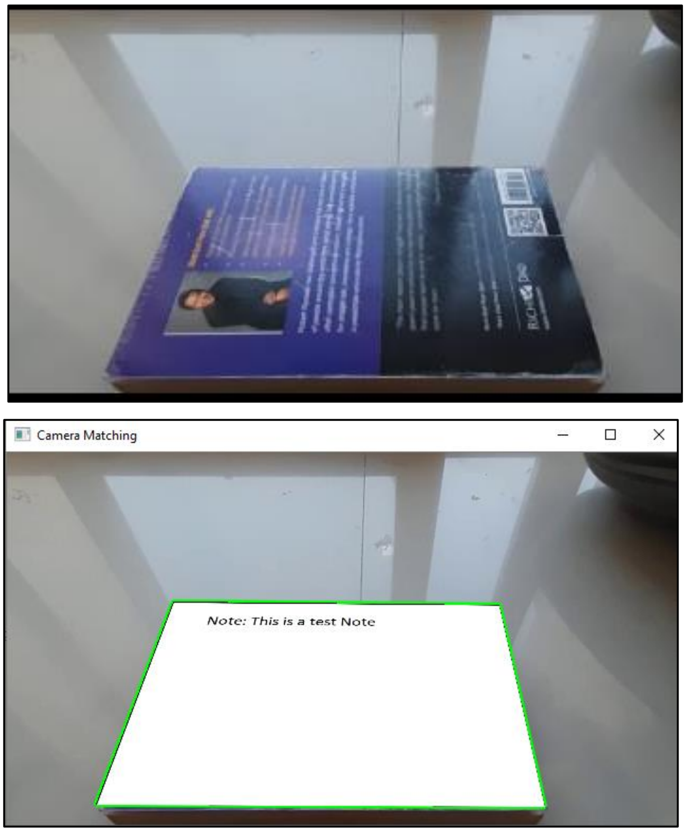

Figure 5A: Rendering of a note after calculating homography

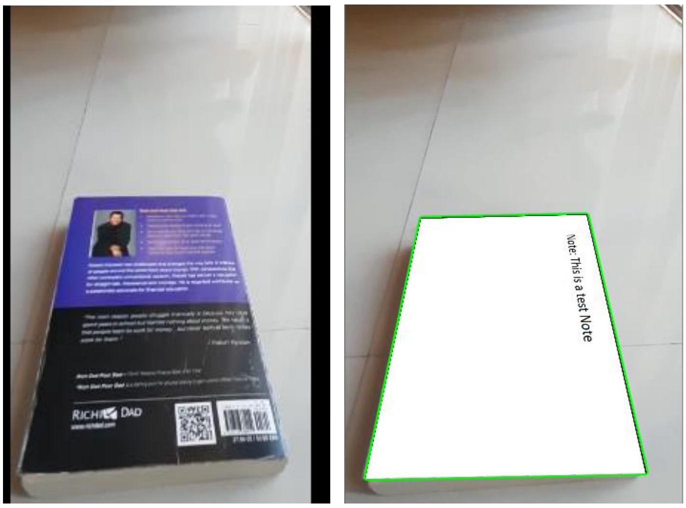

Figure 5B: Rendering of a note after calculating homography

Conclusion

Based on the three parameters considered for the analysis, we can conclude that ORB is the fastest algorithm for feature detection-description. It is suitable for real-time applications. However, SIFT was still proven to be the most accurate considering the number of matches calculated.

These observations can be visually cross-checked with the help of the augmented note. For algorithms with poor accuracy, the augmented note tends to flicker more. Additionally for slower algorithms, the video is processed at comparatively slower speed. This study can be helpful for developers implementing augmented reality application which does not use off-the-shelf solutions.

VI. REFERENCES

[1] D. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer Vision*, 2004.

[2] T. T. a. L. V. G. Herbert Bay, "SURF: Speeded Up Robust Features," *Computer Vision and Image Understanding.*

[3] M. C. a. R. Y. S. Stefan Leutenegger, "BRISK: Binary Robust Invariant Scalable Keypoints," *IEEE International Conference on Computer Vision, Barcelona.*

[4] V. R. K. K. G. B. Ethan Rublee, "ORB: an efficient alternative to SIFT or SURF," *International Conference on Computer Vision, Barcelona, 2011, pp. 2564-2571.*, p. 8, 2011.

[5] A. B. a. A. J. D. Pablo F. Alcantarillay, "KAZE features," in *European Conference on Computer Vision*, Berlin, 2012.

[6] J. N. a. A. B. Pablo F. Alcantarilla, "Fast explicit diffusion for accelerated features in nonlinear scale spaces," in *British Machine Vision Conference*, Bristol, 2013.

[7] P. B. Ertugrul Bayraktar, "Analysis of feature detector and descriptor combinations with a localization experiment for various performance metrics," *Turkish Journal of Electrical Engineering & Computer Sciences*, p. 11, 2017.

[8] P. S. P. S. Karol Matusiak, "Unbiased evaluation of keypoint detectors with respect to rotation invariance," *IET Computer Vision*, 2017.

AUTHORS PROFILE

Dr. Jayshree R. Pansare
Associate Professor
B.E., DAC (C-DAC), M.Tech, P.hD.
Teaching Experience-20 years
Research Experience-11 years
Publications: Research Articles-60+, Patents-02, Copyrights-09
Research Work**:** In Computer vision, Pattern Recognition, Hand Gesture Recognition for Devnagari Sign Language and American Sign Language.
Membership: IETE, CSI, ISTE
Remarkable Achievements:

1. Selected for **Best Citizen of India Gold Medal Award** by Government of India and award will be issued on 15th August 2019 by Ministry of Tamilnadu, India.
2. Received **Best Project Award at ACM-WIR 2016 Symposium on 21st March 2016.**
3. Received **Best Paper Presentation at International Conference on Image, Vision and Computing 2016 (ICIVC 2016) held in University of Portsmouth, UK on 3.08.2016 to 5.08.2016.**
4. Received "**Young Scientist Reasearch Trvel Grant**" for presenting and publishing research at WS4 2017 Springer Conference held in London, United Kingdom (UK). (15th Feb. to 16th Feb. 2017)

Kanad Naleshwarkar Final year student in Dept. of Computer Engineering, Modern Education Society's College of Engineering, Pune.
ACM ICPC 2017 and 2019 Asia-Amritapuri Double-Site Regional Qualifier

Nitish Prajapati Final year student in Dept. of Computer Engineering, Modern Education Society's College of Engineering, Pune.
ACM ICPC 2017 and 2019 Asia-Amritapuri Double-Site Regional Qualifier.
Certification for CodeChef Data Structures & Algorithm Programme (CCDSAP)

Aditya Muley Final year student in Dept. of Computer Engineering, Modern Education Society's College of Engineering, Pune.