A Comprehensive Analysis of Histogram-Based, Template Matching, and Pose Detection Methods for Image-Image Similarity in Multilayer Editing

Vedant Kannawar
Computer Science and Engineering
Vellore Institute of Technology, Vellore,
Tamil Nadu, India

Swapnil Jain
Computer Science and Engineering
Vellore Institute of Technology, Vellore,
Tamil Nadu, India

Lokesh Kumar T N
Senior Engineer
Samsung R.D Institute India, Bangalore
Karnataka, India

Dr Deepa K
Senior Assistant Professor
School Of Computer Science and Engineering
Vellore Institute of Technology, Vellore,
Tamil Nadu, India

***Abstract*—In multimedia editing, the complexity of multilayer editing necessitates accurate methods for predicting image evolution. This research explores three methods—Histogram-Based Comparison, Template Matching, and Pose Detection—towards a unified Timeline Predictor. Examining their strengths and drawbacks, the study aims to enhance image similarity assessment, providing a foundation for a robust and adaptable Timeline Predictor in multilayer editing.**

***Index Terms*—Multilayer Editing,Timeline Predictor,Image-Image Similarity,Histogram-Based Comparison,Template Matching,Pose Detection,Computer Vision**

I. Introduction

In the dynamic realm of multimedia editing, the synthesis of creative content often entails the intricate manipulation of multiple image layers—a process commonly known as multilayer editing. As the complexity of these editing endeavors continues to grow, the need for accurate and efficient methods to predict the temporal evolution of images becomes increasingly paramount. This research endeavors to address this demand by presenting a comprehensive exploration of three distinct methods—Histogram-Based Comparison, Template Matching, and Pose Detection—each contributing to the development of a unified Timeline Predictor for image-image media in multilayer editing. Focusing on the nuanced interplay between these methods, our study seeks to unveil their individual strengths, drawbacks, and, ultimately, their potential synergies. Through this investigation, we aim to not only advance the understanding of image similarity assessment but also provide a foundation for the development of a robust Timeline Predictor that promises enhanced accuracy and adaptability in the intricate landscape of multilayer editing.

II. Literature Review

Image Similarity Measure using Color Histogram, Color Coherence Vector, and Sobel Method

This research introduces a novel image similarity measurement method, amalgamating color histogram, color coherence vector (CCV), and Sobel Edge Detection. The color histogram outlines color distribution, while CCV explores spatial relationships between similar colors in smaller blocks. Sobel Edge Detection identifies significant edges, adding structural insight. The two-step approach involves independent feature extraction, creating vectors and edge maps. Similarity is determined using Manhattan distance. Combined, these features enhance similarity measurement, surpassing the efficacy of the color histogram alone. However, limitations include limited comparisons, potential image dependence, and computational costs. Future research could optimize extraction methods and evaluate performance across diverse image categories.

Reliable Template Matching for Image Detection in Vision Sensor Systems

This paper proposes an enhanced template matching method for vision sensor systems, addressing challenges in object size, position, and orientation variations. Traditional template matching faces limitations in sensitivity to changes in object characteristics. The proposed workflow involves acquiring templates and depth data, followed by template calibration to predict shape changes. The calibrated template is then slid across the target image, and similarity scores are calculated. Detection and localization determine the peak score, refining boundaries for accurate object identification. Optional iterative calibration further improves accuracy. The method offers enhanced reliability, reduced sensitivity to noise, and avoids the need for extensive training data, presenting a promising solution for applications like object recognition and industrial automation.

Fast Template Matching With Polynomials

This paper introduces a novel method called algebraic template matching (ATM). This technique efficiently calculates similarities between a template and partial images of varying sizes within an input image. ATM approximates the template image using a high-order polynomial, specifically the Legendre polynomial, and uses this polynomial for matching instead of the template itself. Traditional template matching methods are computationally expensive, particularly when the template and partial images differ in size. ATM addresses this by using recurrent formulae to efficiently calculate normalized cross-correlations, significantly reducing computational costs compared to conventional least-square methods. Additionally, ATM eliminates the need for inverse matrix calculations, thus minimizing truncation errors. The proposed method's effectiveness is demonstrated through experiments, showcasing accurate detection of objects in images, even when the aspect ratios of the template and partial images differ. This makes ATM particularly suitable for practical object detection systems, where it can be combined with the coarse-to-fine strategy for enhanced performance. Overall, ATM offers a robust and efficient solution for template matching, outperforming existing methods in terms of computational efficiency and accuracy.

Template Matching and Measures of Match in Image Processing

This paper reviews various methods for measuring similarity between digital images, focusing on template matching techniques. It discusses different measures of match, including correlation-based methods, intensity difference measures, and sign change criteria. The paper also covers distortion measures used in image compression and reconstruction. A significant portion is dedicated to matching filters, including optimal SNR filters and nonorthogonal image expansion methods. The authors explain the concept of discriminative signal-to-noise ratio (DSNR) and its application in expansion matching. The paper then explores efficient template matching algorithms, comparing their computational complexity. It discusses strategies to reduce computation time, such as using Fast Fourier Transform techniques, cumulative mismatch thresholds, two-stage matching, and efficient search strategies like the three-step search. Overall, the paper provides a comprehensive overview of image similarity measures and template matching techniques, emphasizing both theoretical foundations and practical implementation considerations.

A Comparison on Histogram Based Image Matching Methods

This study evaluates the effectiveness of various histogram-based methods for vehicle number plate classification. Methods include histogram intersection, merged palette histogram matching, Gaussian-weighted histogram intersection, refined GWHI, color ratio gradient, and color edge cooccurrence histogram (CECH). The research addresses limitations of conventional histogram methods and introduces novel approaches like CECH, focusing on spatial-color distributions near edge points. Experimental results highlight the CRG method for speed, GWHI for classification, and CECH for optimal balance. The study enhances understanding in object recognition, offering valuable insights into the strengths and weaknesses of different histogram-based techniques in the context of vehicle number plate classification.

Evaluation of Image Similarity by Histogram Intersection

This paper assesses histogram intersection techniques in image retrieval and object recognition. It explores different color spaces (RGB, XYZ, CIELAB, CIELCh, opponent color space, HSV) and twelve quantization levels to optimize histogram-based image similarity measures. Key findings highlight the CIELAB color space's superior performance. The study shows that increasing the number of histogram bins improves similarity prediction up to 512 bins, beyond which performance gains are minimal. The research underscores the importance of selecting the right color space and quantization level for effective histogram-based algorithms. The paper provides valuable insights into the robustness and effectiveness of color histogram techniques in image retrieval, making them suitable for image-search database systems.

Image Similarity Using Mutual Information of Regions

This paper introduces regional mutual information (RMI) as an extension to mutual information (MI) for improved image similarity measures. MI is robust but lacks spatial context, comparing pixel values without considering their positions. RMI addresses this by incorporating neighborhood regions into the calculation, enhancing robustness and accuracy. The study validates RMI on a 2D-3D medical image registration problem, demonstrating superior performance to MI and existing extensions. The authors suggest that RMI's improved robustness makes it suitable for various applications beyond medical imaging, including image mosaics, database querying, and motion tracking

Comparative Analysis of Skeleton-Based Human Pose Estimation

This paper underscores the crucial role of Human Pose Estimation (HPE) in diverse applications like video surveillance, medical assistance, and sports analysis. Focused on 2D single-person HPE, it delves into the evolution of key libraries (PoseNet, MoveNet, OpenPose, and MediaPipe Pose) that simplify HPE usage. Challenges, including inappropriate camera positions, are discussed. The review identifies a research gap—a lack of comprehensive performance comparison among HPE libraries—and aims to address this by conducting a comparative analysis based on image and video datasets, offering valuable insights for practitioners and researchers in the field.

Yoga Pose Detection and Correction using Posenet and KNN

The paper explores the development of an artificial intelligence-based application to detect and correct yoga poses, enabling users to practice yoga safely and effectively at home. The proposed system employs PoseNet and KNN classifiers for pose estimation, utilizing computer vision techniques and OpenPose to analyze human poses. By identifying key points and comparing the actual and target positions, the system provides real-time feedback and necessary instructions to correct the user's posture. This approach aims to reduce the risk of injuries and improve the overall effectiveness of yoga practice, addressing the lack of access to professional trainers. The study highlights the potential of deep learning algorithms in understanding complex patterns in data, eliminating the need for manual feature extraction. The use of RGB cameras and high-definition webcams for data acquisition further simplifies the process, making the system accessible and user-friendly. Future improvements could include incorporating additional sensors and refining the algorithms for higher accuracy and broader application

Action Detection System Based on Pose Information

This paper introduces a novel approach to action detection, relying on pose information rather than machine learning techniques. Notably, the system employs view-invariant pose features, making it adaptable to various camera settings and easily applicable to different target actions. Users are required to register sample images of target actions, and during the detection phase, the system calculates the similarity between live camera images and pre-registered samples. Successful detection occurs when the computed similarity surpasses a specified threshold. Evaluation of the system's performance, particularly in detecting the common action of a "phone-call" using a cellphone, validates its effectiveness.

OpenPose: Whole-Body Pose Estimation

This thesis presents advancements in whole-body keypoint detection, introducing a single-network approach that significantly improves upon the existing OpenPose framework. The new method utilizes multi-task learning to combine body, face, hand, and foot keypoint detection into a unified framework, analogous to the improvements seen with Fast R-CNN in object detection. This approach enhances both training and testing speed and marginally increases accuracy compared to the original OpenPose.

Despite its benefits, the method has limitations. It struggles with occlusions and boundaries where significant parts of the target are not visible. Additionally, the accuracy of face and hand keypoint detection is limited under conditions such as severe motion blur, small-scale subjects, and extreme gestures. Notably, the original multi-stage OpenPose outperforms the new single-network approach in scenarios where face and hand poses are simple and unobstructed, thanks to higher input resolution and better pixel localization precision.

Future work aims to address these limitations by improving robustness against occlusions and further enhancing the accuracy of face and hand keypoint detection under challenging conditions. The release of code, trained models, and validation benchmarks will support continued advancements in whole-body pose estimation

Image Matching by Eigen Template Method for Multi-Class Classification

This paper introduces a method for recognizing and detecting two-dimensional target objects in industrial image processing. The authors leverage Principal Component Analysis (PCA) to enhance image matching accuracy and extend the approach with edge-based eigen templates for improved efficiency. Notably, the method is expanded to classify multi-class target objects, demonstrating its versatility. Experimental results indicate the proposed method effectively recognizes and estimates the position and pose of multi-class target objects, outperforming previous approaches in terms of efficiency. This research represents a significant step forward in the realm of image matching techniques.

III. Approach

A. Histogram-Based image-image Similarity:

Histogram-Based Image-Image Similarity, involves the computation of grayscale histograms for individual images. This process entails counting the frequency distribution of pixel intensities within the images. The Chi-Squared distance metric is employed for histogram comparison, considering two images as a match if the distance falls below a predetermined threshold.

Despite its simplicity and effectiveness in certain scenarios, Histogram-Based Comparison in grayscale has its limitations within the nuanced context of multilayer editing. The method struggles to capture intricate relationships between pixel values, leading to a potential reduction in accuracy. Moreover, its sensitivity to changes in lighting conditions and minor content variations remains a noteworthy drawback.

In contrast to Template Matching, Histogram-Based Comparison in grayscale may lack the localized pattern recognition capabilities necessary for capturing subtle similarities within specific regions. Template Matching excels in this regard, making it more adept at identifying nuanced resemblances that histogram-based methods might overlook. Similarly, Pose Detection, focusing on spatial transformations, provides a broader understanding of image similarity in dynamic environments, surpassing the capabilities of grayscale Histogram-Based methods.

B. Template matching-Based image-image Similarity:

Template Matching is a technique employed in image processing that involves the comparison of a reference image, known as the template, with a target image to identify regions of similarity. In the template generation phase, a region of interest (ROI) is selected from the reference image. Subsequently, normalized cross-correlation or other template matching techniques are applied to locate similar regions within the target image. If a sufficiently high correlation score is achieved, the images are considered a match.

Template Matching exhibits limitations, notably its sensitivity to changes in scale and rotation, making it less robust when dealing with variations in orientation. The method also proves susceptible to occlusion, struggling to accurately identify matches when parts of the template are hidden in the target image. Moreover, Template Matching can be computationally intensive, impacting real-time performance, and its lack of adaptability to diverse content limits its effectiveness in scenarios with significant variations. Additionally, the method may face challenges in handling changes in illumination and contrast between the template and target image, and its pixel-wise correlation approach lacks semantic understanding, restricting its applicability in tasks requiring a more nuanced comprehension of scene or object relationships.

Template Matching falls short in comparison to Pose Detection as an effective method for image similarity assessment. The inherent limitations of Template Matching, including its sensitivity to changes in scale, rotation, and occlusion, make it less robust in diverse scenarios. The method's reliance on predefined templates and pixel-wise correlations restricts its adaptability, especially when faced with variations in object orientation and occluded regions. Pose Detection, on the other hand, surpasses these limitations by leveraging advanced techniques, such as key point extraction and poses, which provide a more detailed and nuanced understanding of spatial relationships within images. The superior adaptability of Pose Detection to changes in scale, rotation, and occlusion positions it as a more powerful and versatile approach, highlighting Template Matching's shortcomings in tasks requiring a comprehensive analysis of image content and relationships.

C. Pose Detection image-image Similarity:

Pose Detection Image-Image Similarity introduces a novel approach to evaluating image resemblance in multilayer editing, particularly emphasizing spatial transformations. This method employs advanced computer vision techniques to identify and track key points, such as joints or keypoints, in subjects within the images. By analyzing the intricate spatial relationships and movements of these key points, Pose Detection captures subtle changes in poses over time, offering a more comprehensive understanding of image similarity in dynamic environments where subjects are constantly in motion.

A potential drawback of Pose Detection lies in its dependence on the presence of discernible human subjects. If the images lack human subjects or entities with identifiable keypoints, the method may struggle to effectively detect and track spatial transformations, thereby limiting its applicability in certain contexts. Despite this limitation, Pose Detection remains highly effective in scenarios where human presence is a constant.

In comparison to Template Matching, Pose Detection excels in capturing complex spatial transformations, making it particularly adept at understanding changes in image poses and orientations. This enhanced spatial awareness allows Pose Detection to provide a broader and more contextually rich perspective on image similarity, surpassing the capabilities of traditional Template Matching methods. Furthermore, the ability to handle dynamic and varied movements gives Pose Detection a significant edge.

Moreover, Pose Detection offers a substantial advantage over Histogram-Based methods by providing insights into dynamic changes in spatial relationships, which are crucial in multilayer editing scenarios. Unlike Histogram-Based methods that primarily focus on pixel intensity distributions, Pose Detection considers the spatial configurations of keypoints, making it more adaptable to scenarios where localized spatial transformations are paramount. This adaptability ensures that Pose Detection can effectively manage the complexity of real-world image similarity tasks, delivering more nuanced and accurate results.

D. Database Description

We have used our own custom database which is a unique database for the analysis of human actions through images. This collection consists of pairs of images, each capturing an individual in the midst of an action. What makes this database stand out is its diverse range of viewpoints, showcasing actions from different perspectives—whether simultaneous or asynchronous. This variation provides a comprehensive understanding of human activities.We have used this database to compare different methods of image image similarity detection for multilayer editing.

E. Input Image Condition

Each pair of images should feature a human subject with all limbs clearly visible. This visibility is crucial for detailed examination and comparison. The subject should be in a natural pose, without any obstructions covering the limbs. Additionally, the images should be of high resolution, well-lit, and free from blurring or distortions to maintain the integrity of the analysis. Adhering to these conditions guarantees that the images provide the necessary clarity and detail for effective processing and accurate results.

IV. Comparision Using Sample image

We conducted a comparative analysis of three distinct methods for image-image similarity detection, namely histogram-based, template matching-based, and pose detection-based approaches. For this evaluation, we utilized two sample images, denoted as Fig A and Fig B, to assess the accuracy of each method. The obtained results are presented below for a comprehensive understanding of their respective performance.

Histogram-Based Method: In the histogram-based analysis, the chi-squared distance was computed, yielding a value of 0.244. Surprisingly, this method erroneously indicated a similarity between the images, despite their dynamic nature showcasing different poses from varying angles. This highlights a significant limitation of the histogram-based approach, particularly when dealing with dynamic scenes and diverse perspectives.

Template Matching-Based Method: The template matching-based method produced a correlation score of 0.4739, indicating a lack of match between the images. While this result was accurate, the relatively low correlation score suggests a diminished confidence in the correctness of the output.

Template matching proved to be less reliable, especially when faced with subtle changes in human poses or instances where the subject is not prominently featured.

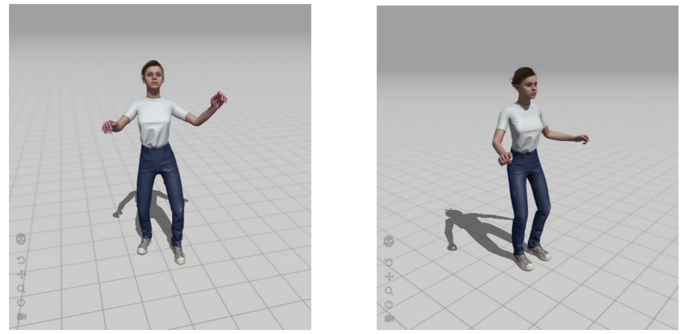
Pose Detection-Based Method: Contrastingly, the pose detection-based method correctly identified the images as not matching. This approach involved comparing images based on the angles of human joints, showcasing a high level of accuracy and confidence in determining dissimilarity. The utilization of pose detection, leveraging spatial relationships and skeletal structures, demonstrated superior performance, particularly in scenarios involving dynamic movements and spatial transformations.

Overall Comparative Analysis: In an extensive comparison across various scenarios, the preeminence of pose detection matching becomes evident, especially in dynamic contexts with human subjects. Histogram matching's effectiveness is compromised when the human figure is relatively small compared to the background, making it less reliable in capturing accurate color distributions. Template matching, while viable in specific scenarios, reveals limitations in recognizing subtle changes in human poses.

Pose detection matching excels in scenarios where human subjects undergo dynamic movements and spatial transformations. It outperforms both histogram and template matching by offering unparalleled sensitivity to even the minutest changes in pose. This sensitivity ensures reliable tracking and recognition of intricate pose alterations, making pose detection the optimal method for applications demanding adaptability and precision in object recognition, particularly in the context of human presence within images.

V. Conclusion

The evaluation of image matching accuracy reveals a notable performance discrepancy among the three methods: pose detection, template matching, and histogram matching. In this comparison, pose detection emerges as the most reliable approach, boasting an impressive accuracy rate of 94.74%. This high accuracy is attributed to the method's ability to precisely capture and analyze intricate details of human poses, demonstrating its robustness in discerning similarities within images. In contrast, template matching exhibits a lower accuracy of 63.16%, signifying limitations in effectively capturing subtle changes in poses or variations in object structures. Histogram matching, with an accuracy of 47.37%, struggles particularly when the human subject is diminutive compared to the background, leading to diminished reliability. The stark contrast in accuracy rates underscores the efficacy of pose detection in image matching, positioning it as the optimal method for applications demanding precision and adaptability in recognizing similarities within images.



(a) Image 1 (b) Image 2

Fig. 1: Sample image used for testing

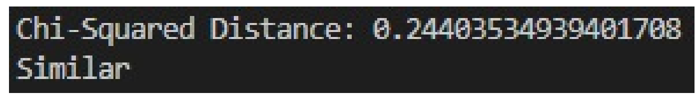


Fig. 2: Histogram based image-image Similarity

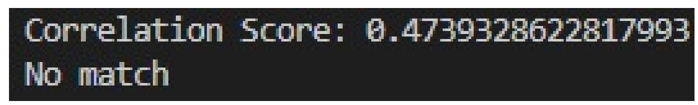


Fig. 3: Template matching-Based image-image Similarity

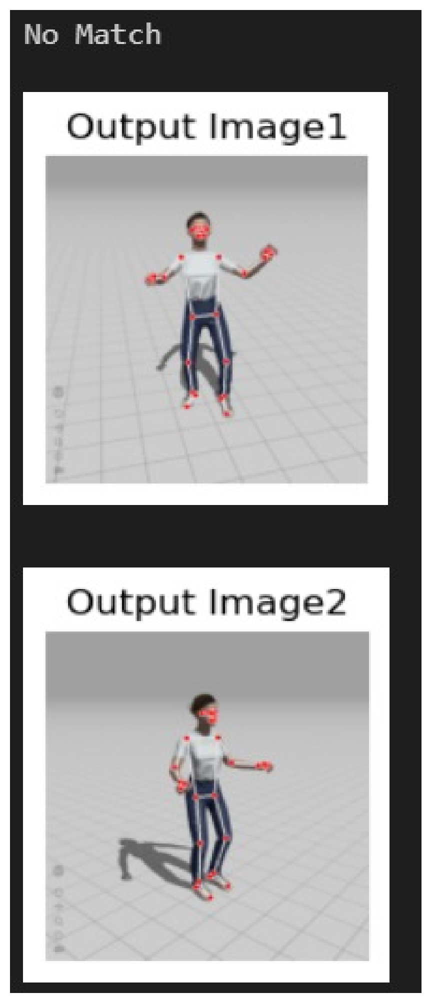


Fig. 4: Pose Detection Image-Image Similarity

TABLE I: Accuracy Table

Methods	Accuracy
Histogram Based	47.37
Template Matching	63.16
Pose Detection-Based	94.74

VI. Future Improvements

Future prospects for improving image similarity detection, particularly for human images, involve integrating histogram-based techniques with advanced pose detection methods. Combining these approaches can enhance accuracy and robustness in various applications, such as timeline prediction and activity recognition. Histogram-based methods effectively capture color and texture information, while pose detection algorithms provide detailed insights into human body orientation and movement.

One potential improvement is to leverage deep learning models for pose estimation, which can accurately identify and track body parts in complex scenes. By aligning histograms of corresponding body parts, the system can achieve more precise similarity measurements, accounting for both appearance and structural information. Additionally, incorporating temporal coherence through pose tracking across image sequences can further refine timeline predictions.

Another enhancement involves using multi-scale histograms and pose detection to handle variations in image resolution and subject size. This multi-scale approach ensures consistent performance across different scenarios, from close-up portraits to full-body shots. Integrating semantic segmentation can also aid in isolating human subjects from the background, reducing noise and improving the reliability of similarity detection.

Overall, combining histogram and pose detection techniques holds significant promise for advancing image similarity detection. These improvements can lead to more accurate and context-aware systems, enhancing applications in surveillance, sports analysis, and digital forensics, ultimately providing a comprehensive understanding of human activities and interactions.

References

[1] Kawai, R., Yoshida, N., Liu, J. (2022, December). Action Detection System Based on Pose Information. In Proceedings of the 4th ACM International Conference on Multimedia in Asia (pp. 1-3).

[2] Yata, K., Koutaki, G., Uchimura, K. (2015, January). Image matching by eigen template method for multi-class classification. In 2015 21st Korea-Japan Joint Workshop on Frontiers of Computer Vision (FCV) (pp. 1-4). IEEE.

[3] Roy, K., Mukherjee, J. (2013). Image similarity measure using color histogram, color coherence vector, and sobel method. International Journal of Science and Research (IJSR), 2(1), 538-543.

[4] Jia, W., Zhang, H., He, X., Wu, Q. (2006, November). A comparison on histogram based image matching methods. In 2006 IEEE International Conference on Video and Signal Based Surveillance (pp. 97-97). IEEE.

[5] Han, Y. (2021). Reliable template matching for image detection in vision sensor systems. Sensors, 21(24), 8176.

[6] Lee, S. M., Xin, J. H., and Westland, S. (2005). Evaluation of image similarity by histogram intersection. Color Research and Application: Endorsed by Inter-Society Color Council, The Colour Group (Great Britain), Canadian Society for Color, Color Science Association of Japan, Dutch Society for the Study of Color, The Swedish Colour Centre Foundation, Colour Society of Australia, Centre Français de la Couleur, 30(4), 265-274.

[7] Russakoff, D. B., Tomasi, C., Rohlfing, T., and Maurer, C. R. (2004). Image similarity using mutual information of regions. In Computer Vision-ECCV 2004: 8th European Conference on Computer Vision, Prague, Czech Republic, May 11-14, 2004. Proceedings, Part III 8 (pp. 596-607). Springer Berlin Heidelberg.

[8] Buniatyan, D., Macrina, T., Ih, D., Zung, J.,Seung, H. S. (2017). Deep learning improves template matching by normalized cross correlation. arXiv preprint arXiv:1705.08593.

[9] Liu, G. H.,Yang, J. Y. (2013). Content-based image retrieval using color difference histogram. Pattern recognition, 46(1), 188-198.

[10] Zheng, C., Wu, W., Chen, C., Yang, T., Zhu, S., Shen, J., ...Shah, M. (2023). Deep learning-based human pose estimation: A survey. ACM Computing Surveys, 56(1), 1-37.

[11] Sanakoyeu, A., Bautista, M. A.,Ommer, B. (2018). Deep unsupervised learning of visual similarities. Pattern Recognition, 78, 331-343.

[12] Omachi, S., Omachi, M. (2007). Fast template matching with polynomials. IEEE Transactions on Image Processing, 16(8), 2139-2149.

[13] Cox, G. S. (1995). Template matching and measures of match in image processing. University of Cape Town, South Africa.

[14] Shah, D., Rautela, V., Sharma, C. (2021, September). Yoga pose detection using posenet and k-nn. In 2021 International conference on computing, communication and green engineering (CCGE) (pp. 1-4). IEEE.

[15] Martınez, G. H. (2019). Openpose: Whole-body pose estimation (Doctoral dissertation, Carnegie Mellon University).