MJSAT

Malaysian Journal of Science and Advanced Technology

Journal homepage: https://mjsat.com.my/

Enhanced CAMSHIFT with Perceptual Grouping, Weighted Histogram, Selective Adaptation and Kalman Filtering: Improving Stability and Accuracy in Face Detection and Tracking

Alex Kok Bin See[*1] and Ji Xing[2]

[1] *School of Engineering, Ngee Ann Polytechnic, Singapore*
[2] *School of Electrical Engineering and Computer Science,University of Newcastle, Australia*

KEYWORDS

Enhanced CAMSHIFT
Perceptual grouping
Kalman filtering
Predict centroid position
Selective adaptation

ARTICLE HISTORY

Received 12 June 2024
Received in revised form 12 July 2024
Accepted 15 July 2024
Available online 16 July 2024

ABSTRACT

This study addresses the weakness of the traditional CAMSHIFT (Continuous Adaptative MeanShift) algorithm. This paper reports the development of an enhanced CAMSHIFT model for theoretical face detection and tracking. The model integrates advanced techniques including Perceptual Grouping, Weighted histogram distribution, Selective Adaptation and with Kalman Filtering to increase the face detection and tracking strategies. Results reveal increased performance in scenarios like hand occlusions, varying illumination, disturbance from multiple faces. The normalized log-likelihood index serves as consistent indicator for face tracking analysis. This new model with Kalman filtering can predict the face's centroid position and increased tracking stability. This new model has achieved low Mean Absolute Percentage Error (MAPE) both predicted ($\hat{X}$) and ($\hat{Y}$) at 9.32 % and 9.70 % respectively. Low RMSE values of X and Y coordinates reported as 8.3 pixel and 8.8 pixel suggest that the Kalman Filtering predicted values are reliable and accurate. It strongly indicates that on average, the forecast/predicted by Kalman Filtering algorithm deviates from the actual values by low margin and this model is effective in predicting and tracking the face target. Further observation suggests that the x-errors and y-errors have both positive and negative values, suggesting no systematic bias in over- or under-prediction in this developed Kalman Filtering model. This model is a significant advancement in face detection methods, promising improved adaptability and tracking.

1. INTRODUCTION

Over the past three years at the time of this writing, face detection and tracking algorithms have experienced remarkable advancements in the scientific and research field. ChatGPT was introduced in 2022 by OpenAI. It marked a significant advancement in natural language processing and AI capabilities, building upon the foundation of previous GPT models. Deep learning, particularly through Generative AI and advanced models like GPT-4o, has been instrumental in these advanced developments. AI with sophisticated image processing capabilities has introduced new techniques for face detection and tracking, improving the performance and accuracy of these systems. Innovations such as deeper network architectures and transfer learning have further enhanced these advancements. Emphasizing real-time efficiency, these technologies are now more adept for applications such as video surveillance, facial recognition systems, and augmented reality. Additionally, notable progress in 3D face recognition has enabled the capture and analysis of facial features in three dimensions, presenting significant potential for security and authentication systems.

Researchers have been diligently addressing the challenges posed by occlusion and pose variations in face detection algorithms, striving to bolster robustness across various real-world scenarios. Concurrently, within the domain of edge computing and deployment, a discernible trend has emerged: the development of lightweight models tailored for deployment on edge devices. This advancement facilitates the implementation of face detection and tracking capabilities even in resource-constrained environments. Additionally, there is a heightened focus on the ethical implications and privacy considerations associated with facial recognition technology. Researchers are

*Corresponding author:
Alex Kok Bin See <alex_see@np.edu.sg>
https://doi.org/10.56532/mjsat.v4i3.337

actively exploring strategies to alleviate biases and ensure the protection of individuals' privacy, reflecting a commitment to advancing technology responsibly.

Effective face localization and tracking are crucial tasks in law enforcement and camera surveillance systems, where the CAMSHIFT (Continuous Adaptive MeanShift) model has proven successful. Recent research, such as that discussed in [1] by a Chinese researcher, he exploited and improved the CAMSHIFT algorithm with a Kalman Filter-based tracking system. This hybrid approach integrates Kalman Filter predictions to refine the tracking window for subsequent frames. Evaluation using the Large-Scale Single Object Tracking benchmark indicates that researcher's hybrid system achieves some improved success rates in specific object categories and benchmark attributes compared to the original CAMSHIFT method. However, the readers are to take note that his experimental findings were focused on certain categories where the hybrid tracker outperforms the original CAMSHIFT tracker. These categories included mainly mobile objects such as flying drones, leopards, electric-fan, airplane, car, kite, racing cars, hippopotamus, and swing, as reported in that literature. Human subject and faces were not studied and reported in that research paper. As a result of that, no direct comparison of our findings could be made with that mentioned research paper.

In reference [2], the authors detailed a remote sensing vision target recognition system tailored for racket ball sports at People Republic of China (P.R.C). Central to their approach is the utilisation of CAMSHIFT algorithm, employed for detecting and tracking moving targets within racket ball games-related contexts. Their study emphasizes image processing and recognition, with Kalman filtering utilized to predict the approximate trajectory of targets across successive frames.

Separately, in [3], the researchers presented research on motion detection and tracking within vehicular flow. Their work introduces a vehicular flow statistics system for vision surveillance, integrating CAMSHIFT and Kalman Filter technique. Utilizing traffic walkway cameras, the system monitors and analyses traffic patterns over defined periods, offering real-time insights to users. The study also explores the synergies between CAMSHIFT and Kalman filter algorithms for multi-target tracking, proposing enhancements specifically for vehicular tracking. The findings underscore the system's efficacy in surveillance design, addressing challenges such as moving object detection, tracking accuracy, and occlusion management.

As in [4], Li et al reported their research paper proposing a method to extract fish trajectories from underwater videos by leveraging object detection models like Faster Regions with Convolutional Neural Networks (RCNN), which is reported elsewhere. It first predicts fish bounding boxes across all video frames using Faster RCNN, then correlates these predictions across consecutive frames based on criteria like Intersection over Union (IoU), center distance, and probability similarity. This correlation is done using either a greedy algorithm or the optimal Hungarian algorithm to form trajectory segments. To improve performance, the method links separated trajectory segments caused by missed detections using interpolation, and removes false detections based on a probability threshold. Experiments on a fish dataset show this object detection-based approach can effectively extract accurate fish trajectories, increasing the average precision from 74.75% to 80.94% after applying the proposed techniques.

In [5], researchers from Thailand, Jaichuen et al presented a paper real-time target human face detection and tracking reported elsewhere. The research conference paper published in 2023, proposes the BLUR & TRACK system that anonymizes detected faces in videos through pixel blurring while enabling efficient retrieval of specified faces. The system performs real-time face detection on video frames using models like RetinaFace and YOLOv5Face, extracts the face regions (ROIs), and applies a pixel blurring technique where the ROI is resized to 5x5 pixels using bilinear interpolation, then enlarged back to original size using nearest neighbour interpolation, resulting in a pixelated blurred effect on the face. A graph database stores user credentials, video metadata, and frame URIs, linking users to permitted videos and frames for quick retrieval based on date, time, or camera properties while enforcing access control. Evaluation on the NTU CCTV-Fights dataset showed RetinaFace achieved higher accuracy (72.5%) and recall (71.2%) compared to YOLOv5Face, but was 9 times slower in processing time. The key conclusion is that BLUR & TRACK enables real-time face blurring while allowing efficient retrieval of specific faces by authorized users, addressing privacy concerns raised by Thailand’s regulations like GDPR and PDPA.

As in [6], the researchers reported on a target tracking system for a novel target tracking system for an amphibious robot based on an improved CAMSHIFT algorithm combined with a Kalman filter. The amphibious robot platform is designed with a spherical structure capable of moving on land and water, equipped with an RGB-D camera and edge computing capabilities. The CAMSHIFT algorithm is used to track targets by continuously adjusting the search window size based on the target's scale changes, while the Kalman filter predicts the target's position in the next frame to improve tracking accuracy and robustness. Experiments demonstrate that the proposed system can accurately identify and track targets, with the tracking delay controlled within one second. The improved algorithm outperforms the traditional CAMSHIFT algorithm, as it is less affected by colour interference and occlusion, maintaining robust target tracking even when the target's color is similar to the surrounding environment. The system's performance is evaluated through pixel-level accuracy analysis, showing its effectiveness for applications such as marine rescue, marine debris search, and aquaculture surface target tracking.

In another research paper [7], that research team introduced a face detection system that integrates the CAMSHIFT with Single Shot MultiBox Detector (SSD) algorithm. The SSD algorithm, well known for its speed and accuracy in object detection, predicts objects and their bounding boxes within a single framework. It was employed to detect faces in a masked-face dataset, making it particularly suitable for applications such as fatigue driving detection. This system necessitates prior training of the SSD model, using collected driver face images to build the model. Their research adopted a two-step approach: initially, the enhanced SSD algorithm detects the face area, then this information is passed to CAMSHIFT, complemented by Kalman filtering for improved tracking.

In [8], the paper entitled “*Detection and Tracking of Human Motion* Targets *in Video Images Based on CAMSHIFT Algorithms*" explores enhancing human motion detection and tracking using the CAMSHIFT algorithm, combined with background and frame subtraction techniques. The study utilizes datasets like KTH Human Video and CAVIAR for testing and implements the algorithm with Visual C++ and OpenCV, alongside MATLAB for performance evaluation. The detail research work is reported elsewhere. The simulation and experiment in that work are carried out in three aspects: (1)

verifying the reliability of the algorithm in KTH human video dataset; (2) verifying the effect of human detection and tracking based on Context Aware Vision using Image-based Active Recognition (CAVIAR) dataset. (3) the superiority of their research method is verified by the performance of the algorithm compared with CAMSHIFT algorithm. The proposed method and experimental results demonstrated improved accuracy, robustness, and computational efficiency compared to traditional MeanShift and CAMSHIFT algorithms. Key findings highlight its adaptability to changes in scale, illumination, and occlusion, making it suitable for real-time applications requiring high precision in tracking human motion targets

Salankar and Bankar [9] reported in their paper entitled " *"A Vision Based* Face *Tracking using CAMSHIFT with BLBP Algorithm in Head Gesture Recognition System*". The authors proposed a new technique of integrating an Adaboost algorithm and CAMSHIFT with Block Local Binary Pattern (BLBP). They compared their proposed method experimentally with several other techniques as follows:

- Basic CAMSHIFT algorithm
- CAMSHIFT with SIFT algorithm
- CAMSHIFT with KLT algorithm
- CAMSHIFT with UKF algorithm
- Adaboost and CAMSHFT with BLBP algorithm

From their extensive experimentation and study, Salankar and Bankar concluded that their Adaboost algorithm and CAMSHIFT with BLBP algorithm's performance accuracy of face tracking and fast processing time is far more superior than the other techniques mentioned earlier. They reported that their proposed method is found to be 87% in performance accuracy for face tracking.

In [10], researchers reported in their paper "*Optimization of Face Tracking Based on KCF and CAMSHIFT*". They reported on optimizing face tracking by synergistically combining the Kernel Correlation Filter (KCF) and CAMSHIFT algorithms. The KCF algorithm is a visual tracking method that employs a linear kernel to learn a correlation filter for locating the target in subsequent video frames. It is renowned for its high-speed performance and accuracy in tracking various objects, including faces. The algorithm learns the correlation filter from training samples of the target object and then utilizes it to detect the object in ensuing video frames. The KCF algorithm has garnered widespread adoption in visual tracking applications due to its efficiency and effectiveness. The paper presents a method to optimize face tracking through the amalgamation of the KCF and CAMSHIFT algorithms, resulting in improved tracking accuracy and reduced failure rates compared to the KCF algorithm alone. The mathematical formulations of the KCF algorithm and CAMSHIFT method are comprehensively reported in the paper, precluding the need for further elucidation in this work.

Skin of human being has emerged as a potent feature in diverse applications, ranging from face detection to hand tracking, owing to its proven efficacy. Despite the variations in skin tones across different ethnic groups, such as Asians, Africans, and Caucasians, several studies have demonstrated that the primary distinction lies predominantly in intensity rather than chrominance. Numerous colour spaces, including RGB, normalized RGB, HSV, and YCrCb, have been employed to classify pixels as skin, underscoring the versatility and robustness of this approach in various domains.

In the original work originally reported by Bradski [11] on CAMSHIFT, human faces are tracked by projecting the face color distribution model onto the colour frame and moving the search window to the mode (peak) of the probability distributions by climbing density gradients. Bradski developed his novel algorithm based on a robust non-parametric technique for climbing density gradients to find the mode (peak) of probability distributions called mean shift algorithm. Bradski modified the mean shift algorithm to deal with dynamically changing colour probability distributions derived from video frame sequences. This modified algorithm is called Continuously Adaptive Mean Shift (CAMSHIFT) algorithm. Hue Saturation Value (HSV) colour system is used to correspond to projecting standard Red, Green, Blue (RGB) colour space along its principal diagonal from white to black Tracking of non-rigid objects is done through finding the most probable target position by minimizing the metric based on Bhattacharyya coefficient between the target model and the target candidates as in [12]. Bhattacharyya coefficient is a popular method that colour histogram to correlate images.

The field of computer vision dealing with face detection, tracking algorithm development and analysis presents researchers with intricate challenges arising from the complex nature of these diverse environmental conditions. Most often occurring conditions are listed as in the following:

- Occlusions from hands
- Tracking under varying illumination environment
- Disturbance from multiple faces

This paper presents the discussion on face detection, tracking and localization in video images using a proposed robust & resilient enhanced CAMSHIFT model that we have developed. The improvised model must be able to track face efficiently with very little error such as occlusion. Since skin colour model is being utilized in implementing this algorithm, hand occlusion (i.e. hand covering the whole face) will prove to be a challenge. The algorithm will also further enhance the technique to obtain a better performance and stability in face localization and tracking system. Several experimental tests will be conducted to fairly compare its robustness and resilient with Bradski's original CAMSHIFT algorithm.

In this paper, the key motivation is to ensure that the proposed enhanced CAMSHIFT algorithm with perceptual grouping, weighted histogram with selective adaptation and Kalman Filtering is presented to the reader with good clarity. Mathematical treatment of the theory of the algorithms are presented. Software flowcharts are presented and reported in this paper as well. Experimental data and analysis are presented and discussed in this paper.

After preliminary literature search, the authors postulated that an extensively developed enhanced CAMSHIFT model's capability with experimental results has not been reported in the literature to date, and at the time of this writing. It is believed that this proposed technique and implementation through the combination of these Perceptual Grouping method, weighted histogram with Selective Adaptation technique and Kalman Filtering have not been reported anywhere yet. Very little is known about the specific effects combining these methodologies prompting the need for further investigation. The research setup and experimental results presented in this paper significantly contribute to the existing body of knowledge in face detection and tracking of human subjects. The enhanced CAMSHIFT algorithm, combined with Perceptual Grouping, weighted histogram with Selective Adaptation techniques and Kalman Filtering has demonstrated improved accuracy,

robustness, and effective face tracking. The authors of this paper postulated that this research outcome will likely contribute to advancement, additionally it has substantial potential applications in various suggested application fields:

- **Healthcare**: Monitoring patient activities, detecting facial emotions or stress levels, ensuring patient safety in hospital premises
- **Marketing and Retails**: Analysing customer behaviour, optimising store layouts, and personalising shopping experiences based on demographic data
- **Automotive industry**: Enhancing driver assistance systems by monitoring driver attention and fatigue level.
- **Law Enforcement:** Assisting in identifying individuals in public spaces, aiding crime scene investigations if occurred and ensuring general surveillances.

In this paper, these contributions underscore the importance and versatility of face detection and tracking technologies in numerous practical applications previously mentioned. The remaining paper is structured as follows: with section 2 on literature review, section 3 describes the methodology used, section 4 focuses on theoretical analysis and implementation. Section 5 presents the results and discussion, some limitation of this study, Section 6 concludes the paper.

2. LITERATURE REVIEW

Numerous face detection strategies, localization, and tracking methodologies are extensively utilized by researchers globally. This section aims to review a selection of relevant journals and conference papers focusing primarily on these techniques. It should be noted that this review is not intended to be exhaustive. As reported by Soni and Waoo [13], a review paper published in the International Journal of Current Engineering and Technology in 2023, entitled "*A review of recent advances methodologies for face Detection*" provides a comprehensive review discussion about recent advances in face detection methodologies. That paper examines various face detection algorithms, encompassing statistical and neural network approaches. It discusses their advantages, limitations, and applications in fields beyond face detection. The authors conclude that no single algorithm is perfect for face detection; however, the comparative analysis provided in the paper aids in selecting an appropriate algorithm for specific use cases. A general classification of face detection techniques is depicted in Figure 1.

Further scanning of the available literature in [14] as reported by authors Himeur et al, another review paper "*Face Mask Detection in Smart Cities Using Deep and Transfer Learning: Lessons Learned from the COVID-19 Pandemic*" published in Systems in 2023. The research paper provides a comprehensive review of the advancements in face mask detection technologies, particularly leveraging deep learning (DL) and deep transfer learning (DTL) techniques. It emphasizes the importance of such technologies in ensuring public safety during the COVID-19 pandemic by automatically detecting face masks in smart city environments. The key objectives and relevance focuses on the critical role of face mask detection in mitigating the spread of COVID-19. It underscores the necessity of integrating modern technologies, such as DL and DTL, in smart cities to monitor compliance with mask-wearing regulations effectively.

For **Datasets and Methodologies**: this paper reviews various datasets used for face mask detection and highlights their characteristics and limitations.

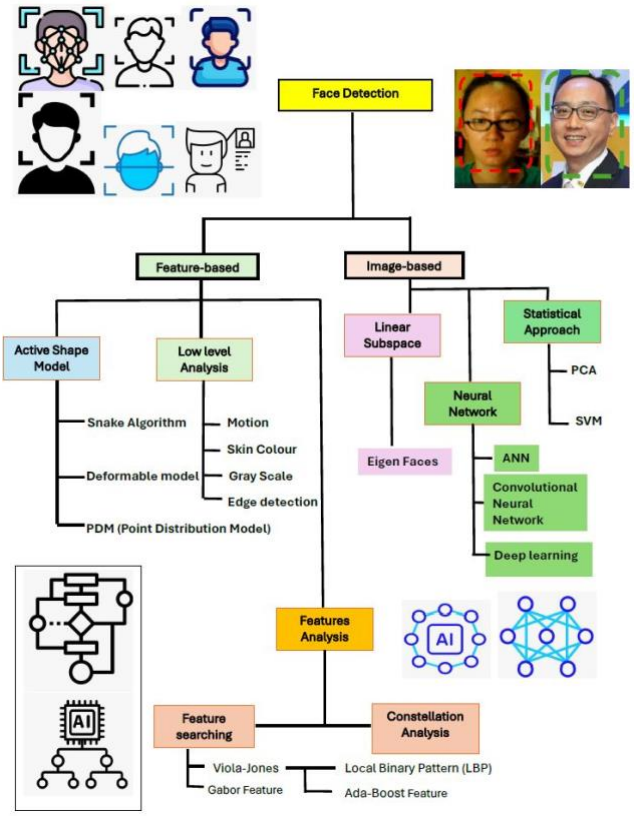

Fig. 1. Face detection techniques classification

For **Techniques and Algorithms**: The study discusses the application of various DL techniques, including YOLO and MobileNet, for face mask detection. It explores the use of deep domain adaptation and other advanced methods to enhance detection accuracy.

For **Performance Analysis**: The study of Benchmarking results for different face mask detection models are summarized. The authors provide a critical analysis of the performance, advantages, and limitations of the reviewed models.

For **Challenges and Future Direction**: The paper identifies key challenges, such as the need for lightweight detection methods that respect citizens' privacy while processing video streams in real-time. It suggests potential research directions and applications that could benefit from further investigation and development.

The authors concluded that while significant progress has been made in face mask detection using DL and DTL, the study concludes that there is no one-size-fits-all solution. The comparative analysis provided aims to guide researchers and practitioners in selecting the most suitable algorithms for specific applications.

As in [15], the authors presented a review paper entitled "*Face Detection Techniques: A Review*" This comprehensive study covers a wide range of face detection techniques, detailing their underlying principles. It discusses the application of mathematical algorithms, such as the Viola-Jones algorithm, and deep learning-based methods for real-time face detection. The paper also addresses the challenges and limitations inherent

in face detection algorithms, including issues related to image quality, face variability, and different face angles. Furthermore, this paper compares feature-based and image-based approaches for detecting faces in digital images, highlighting the strengths and weaknesses of each method. It examines the use of color models and statistical shape models for face detection as well. Overall, the review paper provides a detailed overview of advancements, challenges, and future research directions in the field of face detection.

The authors Li et al [16] reported a review paper entitled "*Face detection and tracking using Neural Network*". Face detection and tracking technology is used in transportation, security, military fields. In view of the traditional face detection and tracking technology is easy to be affected by light, which leads to low detection accuracy, in that paper, they proposed and used Retinaface and CAMSHIFT algorithm to face detection, and realizes real time face detection and tracking by P control steering gear in PID control. The detail of this paper is reported elsewhere. Through tests in different environments, the detection accuracy of the Retinaface algorithm and the CAMSHIFT algorithm is claimed to be above 99%. The camera is rotated through P to ensure that the face can be captured by the camera, and the camera response time can reach 0.1s.

The MeanShift algorithm remains crucial in image processing and cluster analysis within computer vision, focusing on identifying peaks of a designated density function. Initially conceptualized by Fukunaga and Hostetler [17], this method has established itself as a significant approach in computer vision, leading to numerous developments and adaptations. In clustering applications, the MeanShift method operates on the premise that each point represents samples associated with a probability density function. Regions with higher density are interpreted as local maxima of the distribution. The algorithm enables points to attract each other, akin to a minimal gravitational force, causing convergence toward higher density regions. This results in the amalgamation of points at various locations, revealing the local maxima in these convergent zones. The critical aspect of identifying local maxima is achieving optimal resolution, especially in 3D face tracking, where it facilitates maximum pixel dispersion. The MeanShift algorithm effectively pinpoints local maxima by shifting a window toward areas of highest density. As a gradient ascending algorithm, it involves iteratively moving kernels toward regions of increased density until convergence is achieved, thereby underscoring its role in clustering complexities.

3. METHODOLOGY

In this section, the authors will report an overview with an emphasis on the proposed and developed robust and resilient CAMSHIFT algorithm. The overall of this proposed CAMSHIFT face detection and tracking algorithm is divided into six (6) stages/processes, which will be described and discussed here. The hardware and computer in used for this work are described in subsection 3.1.

3.1 Computer Specification and hardware usage

For this research project, a webcam and computer were conveniently used primarily to develop software algorithms, collect data, and perform analysis.

- Computer equipped with Intel Pentium M processor 1.73GHz and 1GB DDR2 RAM
- WebCam, Creative Live! 640x480 VGA CCD sensors
- 4X digital Zoom
- 640x480 resolution video up to 30 FPS(frames/sec)
- USB 2.0 Hi-Speed connectivity

With a WebCam frame rate at up to 30 frames per second, the sampling time is at 33 ms throughout all the experimental work reported in this paper.

3.2 Software

The software development and coding were developed using graphical system design LabVIEW software and National Instruments NI-IMAQ. The developed software codes are upward compatible with the latest version of the LabVIEW and NI-IMAQ toolkit. Additionally, MATLAB software was utilised for result analysis.

One of the LabVIEW front panels is depicted in Figure 2. This front panel is shown in partial because there are several testing and data analysis programs which are known as LabVIEW subVIs or in layman's terms known as subroutines. The front panel has programming features where the user can interact with the software program through other dialogue box. The CAMSHIFT window, bounding rectangles can be seen overlaid onto the video frame. The probability distribution image is shown in Figure 3 as dark grey scale colour picture of the human face

LabVIEW formulae node was invoked and used. The C codes were embedded into the LabVIEW formulae node to ensure that the developed software program was executing efficiently.

3.3 Overview of Six (6) Stages of the Enhanced CAMSHIFT with Perceptual Grouping, weighted histogram with Selective Adaptation and Kalman Filtering Algorithm

The entire process is divided into six main stages/process. With reference to the version of the flow chart of this detection and tracking algorithm as depicted in figure 11. The first stage requires acquisition of images with conversion of its colour space. The next stage ensures initializing of search window and obtaining the colour histogram. The third stage performs segmentation of skin and usual of perceptual grouping technique. This follows by application of connected component operators. Fifth stage focuses on the original CAMSHIFT algorithm. Finally, Weighted Adaptive Colour Histogram with Selective Adaptation is utilized. Each stage will be described in detail over the next few sections found in this paper.

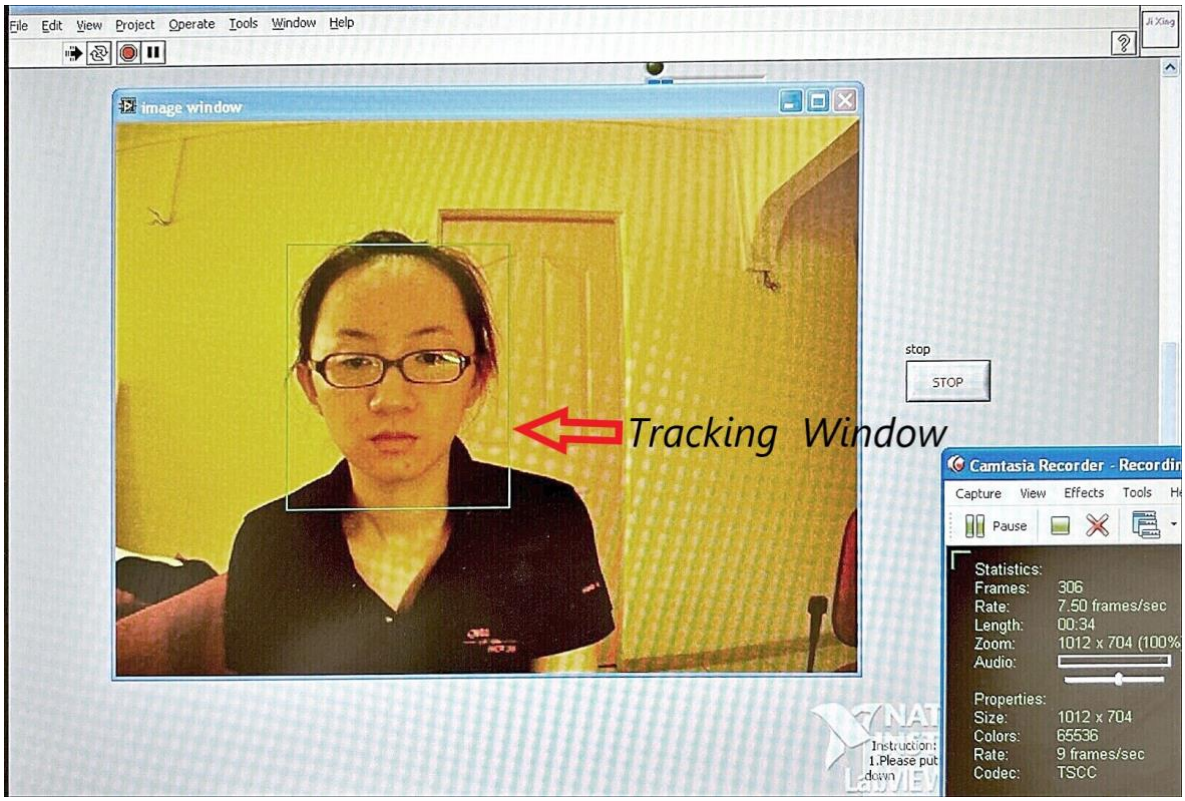

Fig. 2. LabVIEW software front panel with a human subject with a tracking window

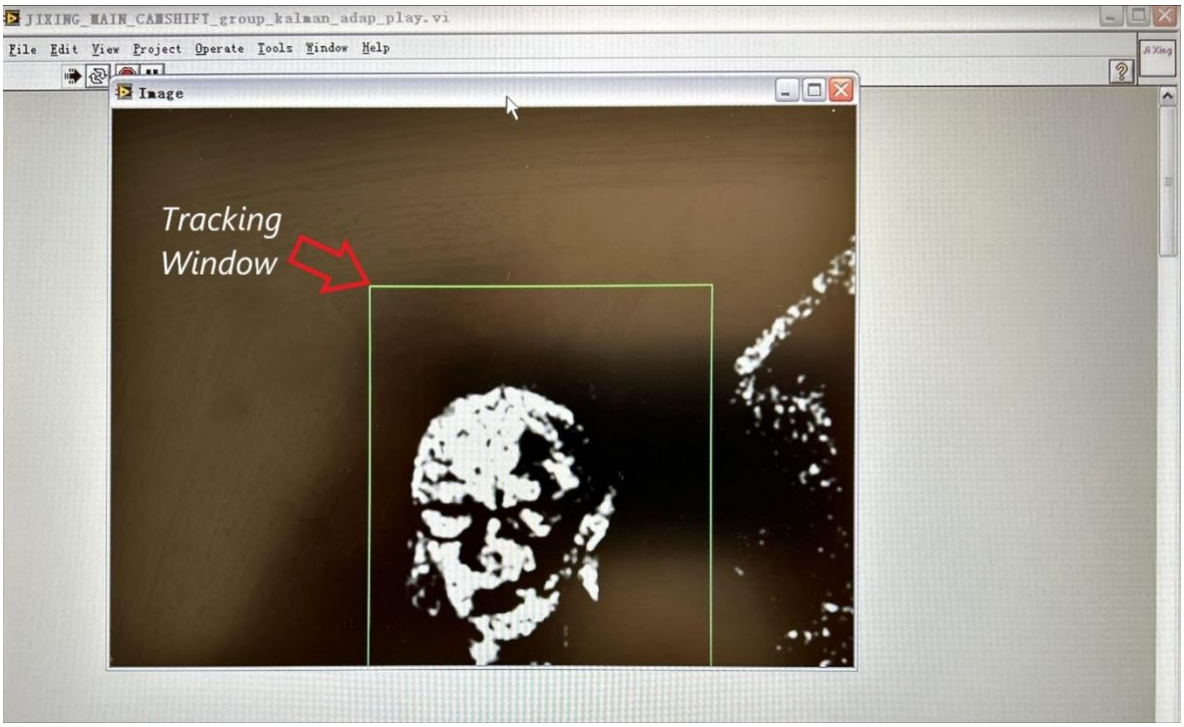

Fig. 3. A sample of the probability distribution image of the detected face inside a tracking window

3.4 Image Acquisition Stage

Most standard conventional color cameras have rather standard output with RGB (Red, Green, Blue) signal as reported [18]. However, numerous researchers have already reported that the RGB color space proves unreliable under varying illumination conditions, as the intensity distribution across RGB values causes the face's color distribution to change with scene brightness [19]. The readers are encouraged to read other literature works not further reported here. To achieve

illumination invariance, the RGB signal must be converted to the HSL color space, which comprises hue, saturation, and luminance. Given the highly curved surface of a human face, the observed intensity varies significantly [20]. Consequently, many face-tracking applications exclude, minimise or eliminate the luminance component of the color space, relying instead on the chrominance component for robust skin color region segmentation. Chrominance, represented by the hue and saturation in the HSL model, enhances tracking efficiency. Thus, a 2-D HS space distribution provides a colour model invariant to illumination changes.

3.5 Initialization Stage

The face detection algorithm introduced in this project is mainly based on skin color. There is a need to obtain the skin sample of the user to be tracked. Therefore, at the beginning of the algorithm, a 30 x 30 search window is drawn in the centre of the image plane. At this time, the target human subject must prompt his/her face within the search window. The algorithm will take a sample of the skin automatically as depicted in Figure 4. If the human subject failed to do so, the program will not have the sample skin for histogram binning and formation as a lookup table for subsequent processing.

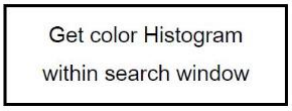

Fig. 4. Skin sample for the formation of color histogram in a lookup table form

At this time, the program will be terminated, and the user must reinitiate this procedure. Two small windows, overlaying on the screen, are created for this purpose. One of them is known as the search window, which indicates the region of the tracked face. The tracking window is shown in green colour. The other one, which is slightly larger in size than the search window is solely used for computation or calculation purposes only. The larger computation window may be turned on/off programmatically. The reason is to provide an accurate calculation of the color histogram. This colour histogram is vital for the formation of a lookup table for subsequent usage in the algorithm. The search window will contain the detected face and track it accordingly. Technique for auto detection of user face may be considered as a future enhancement. The limitation of auto detection at the start will need to be addressed for future work.

3.6 Skin Segmentation Stage

One of the colour spaces, namely the Hue values are sampled and binned into a 1-D histogram for array processing in the software algorithm. This skin color histogram forms a lookup table mainly for distinguishing skin color pixel values from non-skin color pixel values as depicted in Figure 5.0. The lookup table comprises 256 elements, each corresponding to a pixel value used for segmentation. Pixels that do not resemble skin are assigned a value of 0 based on this lookup table. The result of this processing generates a probability distribution image, resulting in a binary mask that highlights the skin color areas in the segmented image [21]. Further, a probability distribution image indicates the Hue image into a form of gray-scale image as depicted in Figure 3.0. Therefore, only skin region is probabilistically shown in the image plane. It is important to note that the image is not perfect, as it consists of not only the desired skin region but also the unwanted noise, such as objects with the background environment.

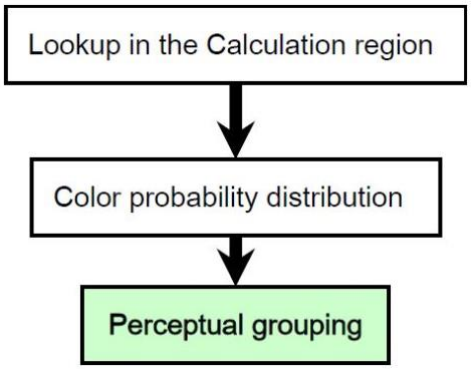

Fig. 5. Utilization of the lookup information to compute the color probability distribution and perceptual grouping

Unfortunately, the probability distribution image generated may consist of background noise such as pixel form by skin-like objects. Furthermore, the CAMSHIFT algorithm proposed by Bradski relies on colour probability distribution image alone. There will be error in the tracking whereby illumination condition is too bright or dim. As a result, the authors of this paper have proposed and implemented the additional filtering technique known as Perceptual Grouping as depicted in Figure 5. Perceptual grouping is introduced to filter the image. It is a process whereby a vision system organizes image regions into emergent boundary structures that aim to separate objects and scene background. Perceptual grouping uses morphological operator or erosion technique which is applied to the original probability distribution image. Furthermore, the image is subsequently convoluted by using a low-pass filter.

3.7 MeanShift Algorithm Overview

Tracking a face based solely on skin color is insufficient, as demonstrated by researchers [22]. This inadequacy arises from potential interference by skin-like pixels, such as those from the user's hand, which can cause occlusion during tracking. If these errors are not identified, the color model may adapt to image regions unrelated to the actual target, increasing the risk of the tracker losing the intended object. An iterative process is required to determine if convergence has occurred or not, as depicted in Figure 6.

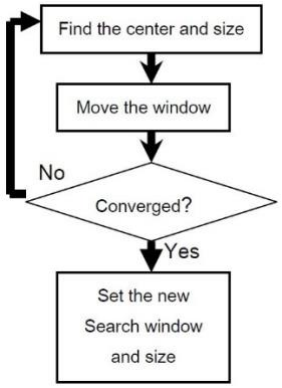

Fig. 6. The MeanShift Algorithm

MeanShift algorithm is the most vital concept in CAMSHIFT model. It is a proven robust, non-parametric technique for gradient climbing of a probability distribution image to find the mode (peak) of the target distribution. A detailed mathematical treatment is given in section 4 in this

paper. In later section, there is detail explanation on how to find the Zeroth moment, first moment and the size of the tracking window.

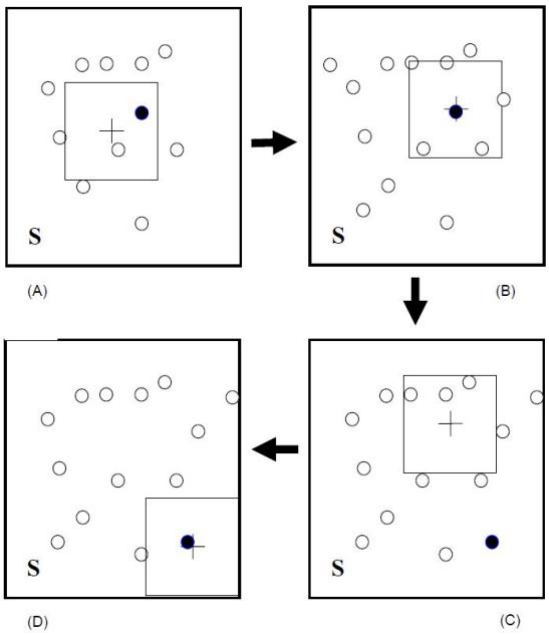

Fig. 7. Static sequences of MeanShift steps: Sequence (A) to (D), moving clockwise. Smaller inner box is the kernel, the bigger box depicts the Euclidean space, X, the empty dots are the vectors.

In Figure 7, the shown smaller inner box is the kernel, the larger box depicts the Euclidean space, X, the empty dots are the vectors. The dots within the kernel are the data points. The black dot refers to the centroid position computed based on the data points within the search kernel. Sequence (A) of Figure 7, assume at the initial state, the black dot is inside the kernel, let the centre of the kernel be $\boldsymbol{x}$, and the black dot be $\boldsymbol{m}(\boldsymbol{x})$. MeanShift algorithm shifts $\boldsymbol{m}(\boldsymbol{x})$ to $\boldsymbol{x}$ and becomes Sequence (B) of same figure. After MeanShift, the black dot $\boldsymbol{m}(\boldsymbol{x})$ is the centre of the kernel, $\boldsymbol{x}$. In Sequence (C), as the data points which correlates to the detected face already converted into probability distribution image are not static in nature, movement occurs causing displacement outside of the kernel. Repeatedly, the iterative process continues the kernel will be moved to black dot $\boldsymbol{m}(\boldsymbol{x})$ is the centre of the kernel. The full MeanShift procedure iterates until it finds a fixed point T=M(T). The difference $\boldsymbol{m}(\boldsymbol{x}) - \boldsymbol{x}$ is called mean shift. The repeated movement of data points to the sample means is called MeanShift algorithm. The steps will be continuously repeated until a convergence occurs.

3.8 CAMSHIFT Algorithm Stage

MeanShift algorithm was initially applied to mode seeking by Cheng [23]. Unlike MeanShift, which addresses static distributions, CAMSHIFT (Continuously Adaptive Mean Shift) is designed for dynamically changing distributions. The mean location, or Centroid, is identified within the search window of the discrete probability image calculated earlier using moments. The Zeroth, First, and Second moments, all determined within the Min-Max box, are used to compute the Centroid of the search window. The MeanShift process iteratively recalculates new window position values derived from the previous frame until there is no significant shift in position. The MeanShift algorithm is designed for static distribution and the new method proposed by Bradski is integrated with the MeanShift algorithm, the whole process of adaptively changing the position and size of the tracking window and named CAMSHIFT. The key contrast between the former and the latter is that CAMSHIFT is used to deal with dynamically changing colour probability distributions derived from video frame sequences. A figurative illustration is depicted in Figure 8.

In this work, authors of this manuscript are acutely aware of the limitation of CAMSHIFT and since histogram is fixed and based on the sample data that was collected at the initialization step. The fixed histogram is not ideal to handle illuminating variations and the skin likelihood background. Hence, predicting the centre position and selective adaptation steps are proposed and implemented in this research paper.

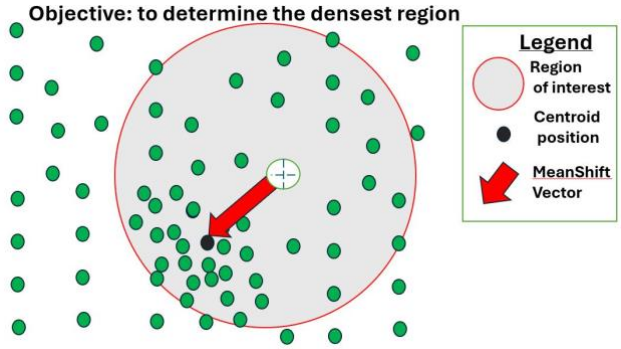

Fig. 8. An illustration on the adaptation of CAMSHIFT with dynamically changing Centroid position.

3.9 Selective Adaptation with Weighted Adaptive Colour Histogram

The MeanShift algorithm alone is not effective as a tracker. If the initially chosen region includes pixels outside the facial area, such as background pixels, the 2-D probability distribution image can be skewed by their prevalence in the histogram back-projection. Pixels near the boundary of the search window are generally less reliable and should be excluded. Thus, lower weights can be assigned to pixels farther from the center of the search window, with the highest weights given to pixels at the center, as described by See and Liaw [24]. In earlier research by one of the current authors, a weighted distribution mechanism was proposed and developed, assigning higher weights to pixels closer to the region center. A weighted histogram may be used to calculate the target histogram, as postulated by Comaniciu et al [25]. Figure 9 illustrates the weight distribution within a search/tracking window of a captured image.

Fig. 9. The weight distribution of the human face within the search/tracking window.

A fixed/static skin color model is highly susceptible to changes in illumination, often resulting in a loss of tracking. Adapting the skin color model is essential for managing varying lighting conditions. However, adjusting a color model during

tracking presents a significant challenge due to the lack of reliable ground-truth. Any color-based tracker faces the risk of losing the object, especially when occluded by other elements. Without effective error detection, the system may incorrectly adapt to image regions that do not correspond to the intended target. To mitigate this challenging issue, observed log-likelihood measurements can be used to identify video frames with errors. Colour data from these frames is excluded from adapting the object's colour model. When the face tracker loses the object, there is often a sudden, substantial decrease in log-likelihood value. Adaptation is suspended until the object is successfully tracked with a sufficiently high likelihood. Selective adaptation is vital, and it serves a decision-maker determining whether new skin model is allowed to be adapted or not.

3.10 Prediction of Centre Position using Kalman Filtering

The main concept in this step is the Kalman Filtering technique, which is a renowned method with numerous applications. It utilizes a set of mathematical equations that provides an efficient computational (recursive) means to estimate the state of a process, in an approach that minimizes the mean of the squared error. As part of this integration of this Kalman Filtering technique into this combined CAMSHIFT face tracking work, a flow diagram is depicted in Figure 10.

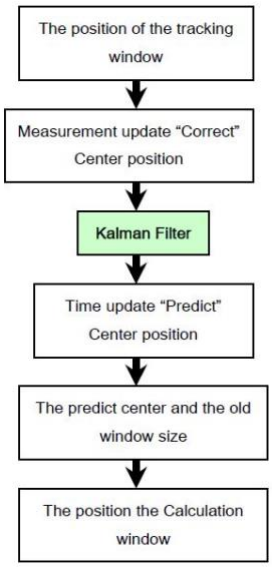

Fig. 10. The flow chart of the predict centre/centroid position of the face detection

The Kalman Filter estimates a process by using a form of feedback control. This filter estimates the process state at some time and further obtains feedback in the measurements. As such, the equations for Kalman Filter can be categorized into two distinct groups namely: time update and measurement update equations. The time update equations can be considered as predictor equations, while measurement update equations can be thought of as corrector equations. Indeed, the final estimation algorithm resembles that of a predictor-correct algorithm for solving several real-world applications. Detailed mathematical treatment Kalman Filtering applied in this research work on face detection will be elaborated and presented in Section 4.0. The final overview of the combined algorithm

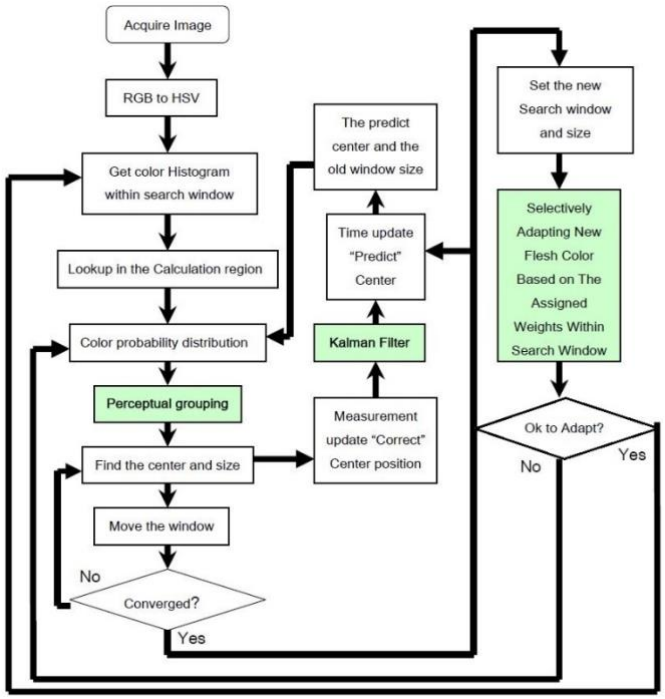

Fig. 11. The detailed flowcharts with the combination of all the six (6) stages of this proposed, developed and implemented enhanced CAMSHIFT model with Perceptual Grouping, Selective Adaptation and with Kalman Filtering.

4. THEORETICAL ANALYSIS AND IMPLEMENTATION

The mathematical formulation and theoretical analysis of each individual method used in each stage are presented in this section. Firstly, the authors will discuss on how the colour Probability Distribution is formed. This is followed by the introduction of Perceptual Grouping, which is effective in skin segmentation. The basic MeanShift algorithm will also be presented detailing the mathematical manipulation. With the basic algorithm explained, the steps to apply Continuously Adaptive Mean Shift Algorithm (CAMSHIFT) operation are derived. The application of weighted adaptive colour histogram and selective adaptation are explained and demonstrated clearly in the following. Finally, the Kalman Filtering theoretical and implementation concepts are reported.

When the program is executed, the user will see a countdown timer of 5 seconds before actual tracking begins. The LabVIEW application software program user will be prompted for skin sample acquisition. The human subject will be required to follow the instruction to place his/her face in the field of view (FOV) of the acquisition camera. The skin sample will be obtained in the initial window. The search window is set at 30 x 30 pixels (i.e. a small bounding box for skin sampling) as depicted in Figure 12.

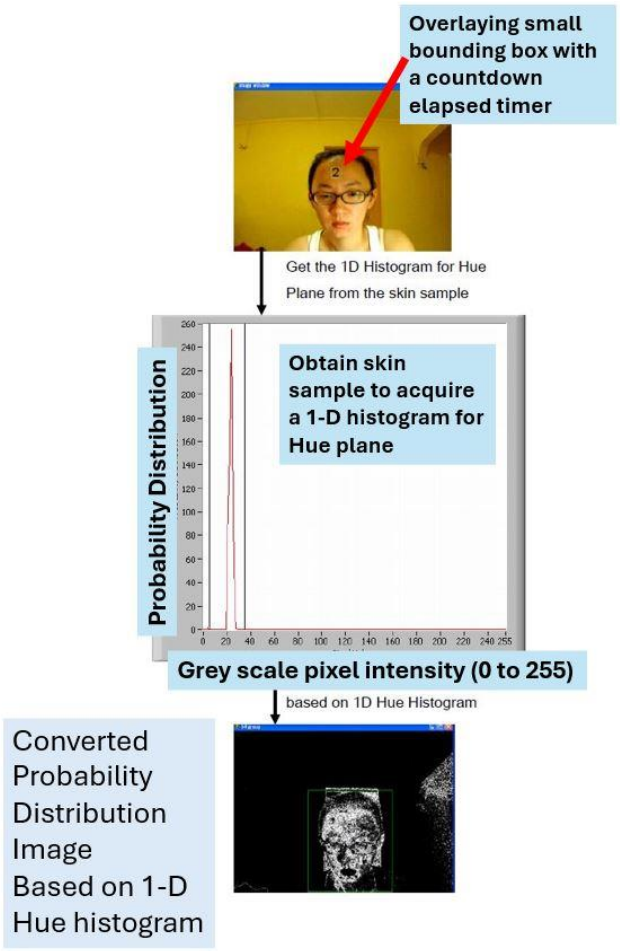

Fig. 12. The process of acquiring sample skin and binning into a 1-D histogram/array. This figure illustrates the noisy probability distribution grey scale image after conversion.

4.1 Colour Probability Distribution

Probability of an object of interest is computed based on Baye's rule [26]. Considering each of the pixel and a single measurement vector m_k:

$$p(o_n|m_k) = \frac{p(m_k|o_n)p(o_n)}{\sum_i p(m_k|o_i)p(o_i)} \quad (1)$$

Equation 1 converts the frame of each incoming image into probability distribution image. The measurement vector m_k is considered as Hue vector of the image plane in this case. Conditional probability of the desired object occurs when the Hue translation has already occurred. In other words, colour plane of an image has already been converted into probability of Hue color plane, provided the Hue of that pixel has been detected first. Furthermore, equation 1 is used to determine the probability of each pixel that considered as the desired object. In similar context, this equation can be analogized by considering generated histogram of skin and the entire image.

The probability of a colour vector (Hue, Saturation), or simply (h, s), for a given skin is approximated by:

$$p(h,s|skin) \approx \frac{h_{skin}(h,s)}{N_{skin}} \quad (2)$$

where $h_{skin}(h,s)$ is the histogram of skin of Hue and Saturation channel, and N_{skin} is the total number of skin pixel.

The probability of a skin pixel in an image can be approximated by a fraction of observed pixels known to be skin, as shown below:

$$p(skin) \approx \frac{N_{skin}}{N_{Total}} \quad (3)$$

where N_{total} is the total number of pixel of the image

The probability of a colour vector is approximated by:

$$p(h,s) \approx \frac{h_{total}(h,s)}{N_{total}} \quad (4)$$

Based on Baye's rule, the probability of skin given a colour vector can be found by:

$$p(skin|h,s) = \frac{p(h,s|skin)\cdot p(skin)}{p(h,s)} \quad (5)$$

Equation 5 is similar to 1, can be further simplified to the ratio of the two histograms:

$$p(skin|h,s) \approx \frac{h_{skin}(h,s)}{h_{total}(h,s)} \quad (6)$$

The result computed above forms equation 6, which is used as the lookup table to transform each pixel of every image frame into probability distribution image. It is suggested that more complicated measure sets for more object discrimination can be considered elsewhere as suggested by earlier researchers, Fukunaga and Hostetler, paper cited in the reference section.

The probability of many local measurement vectors over a region of the image can be determined by:

$$p(o_n|\Lambda_k m_k) = \frac{\prod_k p(m_k|o_n)p(o_n)}{\sum_i \prod_k p(m_k|o_i)p(o_i)} \quad (7)$$

Note that equation 7 is not applied in the proposed algorithm. The reason is because equation 1 or equation 5 is sufficient in producing the probability distribution image.

In enabling face tracking using a flesh color model, the system employs a user-guided approach. Users are instructed to center their face within an on-screen box, allowing the system to sample flesh areas effectively. A search window of 30 x 30 pixels is utilized, as illustrated in the previously reported Figure 12, which depicts both the RGB color plane and the transformed hue image plane. The hue values corresponding to the flesh pixels within this search window are then sampled from the hue (H) channel of the image. These sampled hue values are organized into a one-dimensional (1D) array histogram, simplifying the computational and space complexities of the model. This histogram-based representation facilitates the clustering of similar color values, as postulated by researchers [27], enabling an efficient modeling of the flesh color distribution for face tracking purposes. Histogram back-projection is a primitive operation that associates the pixel values in the image with the value of the corresponding

histogram bin. The back-projection of the target histogram with any consecutive frame generates a probability distribution image where the value of each pixel characterizes probability that the input pixel belongs to the histogram that was used.

Given that *m*-bin histograms are used, we define the *n* image pixel locations $\{x_i\}_{i=1\ldots n}$ and the histogram $\{\hat{q}_u\}_{u=1\ldots m}$. The author also defines a function $c: \Re^2 \rightarrow \{1\ldots m\}$ that associates to the pixel at location x_i^* the histogram bin index $c(x_i^*)$. The unweighted histogram is computed based on Allen et al as:

$$\hat{q}_u = \sum_{i=1}^{n} \delta[c(x_i^*) - u] \tag{8}$$

In all cases the histogram bin values are scaled to be within the discrete pixel range of the two-dimensional array (2D) probability distribution image as follows:

$$\left\{\hat{p}_u = min\left(\frac{255}{max(\hat{q})}\hat{q}_u, 255\right)\right\}_{u=1\ldots m} \tag{9}$$

In other words, the histogram bin values are rescaled from $[0, max(\,q)]$ to the new range [0, 255], where pixels with the highest probability of being in the sample histogram will map as visible intensities in the 2D histogram back-projection image.

4.2 Perceptual Grouping

In this study, the face is regarded as a point of selective attention on an image screen, as it is chosen for tracking. To fully rely and depend solely on a colour probability distribution image proves inadequate for face tracking. The probability map generated by a skin colour model may contain background noises, such as skin-like background pixels. Hence, perceptual grouping is introduced as a method to filter the image. Perceptual grouping involves the organization of image regions by a vision system into emergent boundary structures, with the goal of distinguishing objects from the scene background. The procedure is shown below:

1) *Compute log probabilities of the foreground in the image. This results in a probability distribution image,* I^o.
2) *Apply morphological erosion to* I^o. *This reduces noise and erroneous foreground and yields image* I^{er}.
3) *Let* $I^* = I^{er}$ *, then iterate the following operation to a desire number of times:*

$$I^* = \frac{1}{2}(I^* \otimes \text{low-pass filter} + I^o) \tag{10}$$

where $\otimes$ *denotes convolution*

As shown in Figure 13, such an operation effectively performs perceptual grouping in the resulting image I^* . The number of iterations was set at 20 with much lesser noise as observed in the probability distribution image. Previously, one of the authors of this paper has reported in some details about similar technique of perceptual grouping. As postulated by See and Goh [28], perceptual grouping technique is a powerful approach for eroding noises within the probability distribution images. The paper was published in this journal in January 2024.

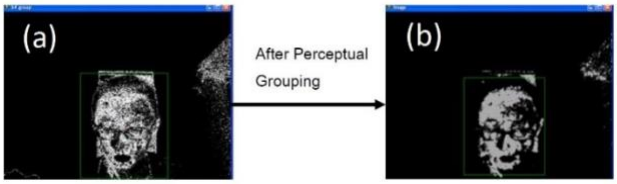

Fig. 13. (a) the original probability distribution image and (b) shows a processed image after perceptual grouping which eliminated much noise.

4.3 MeanShift Algorithm

Mean shift is a nonparametric, iterative mode-seeking algorithm introduced by Fukunaga and Hostetler in 1975 for locating the maxima (modes) of a density function represented by a set of samples S. The key idea is that the value of a density function at a point can be estimated using the samples within a small neighbourhood around that point. In 1995, Cheng revisited and generalized the mean shift procedure, providing a more comprehensive formulation and demonstrating its applications in clustering and global optimization. The generalized mean shift procedure iteratively computes the weighted mean shift vector m(x) at the current point x using a kernel function K(x) that assigns weights to the data points based on their distance from x, and then translates x to the new location m(x). This process is repeated until convergence, i.e., m(x) = x, at which point x is considered a mode of the underlying density function. Mathematically, the mean shift algorithm initializes x (e.g., with a data point from S), computes the mean shift vector m(x) using a weighted average of the data points in the neighbourhood of x, translates x to the new location m(x), and repeats these steps until convergence. The convergence points of the mean shift procedure correspond to the modes (local maxima) of the underlying density function represented by the data set S, making mean shift useful for various applications, including clustering, segmentation, and tracking.

Let X be an n-dimensional real Euclidean space and S a set of sample vectors in X. Let w be a weight function from a vector in X to a nonnegative real. Let the sample mean m with kernel K at $x \in X$ *be defined such as the following:*

$$m(x) = \frac{\sum_{s\in S} K(\|s-x\|^2)w(s)s}{\sum_{s\in S} K(\|s-x\|^2)w(s)} \tag{11}$$

Let $M(T) = \{m(t): t \in T\}$, Let T $\subset$ X be a finite set (the "cluster centers"). One iteration of MeanShift is given by $T \leftarrow M(T)$. The full MeanShift procedure iterates until it finds a fixed point $T = M(T)$. The difference $m(x) - x$ is called mean shift. The repeated movement of data points to the sample means is called MeanShift algorithm. In each iteration of the algorithm, $s \leftarrow m(s)$ is performed for all $s \in S$ simultaneously. Further detailed discussion of the procedure, its definitions, and constraints, reader is invited review articles cited earlier from Fukunaga and Hostetler or Cheng. The concept of a kernel is fundamental to the Mean-Shift procedure and, indeed, Mean Shift is conventionally defined in terms of a kernel, postulated by authors reported in [29].

The concept of a kernel is fundamental to the Mean Shift procedure, as it helps provide a density estimate of the data points, and different kernels can lead to different results. If a flat kernel is used as the search window, all data points within the kernel share the same weight, while data points outside are rejected. With a Gaussian kernel, all data points within the

kernel are considered, but each point has its own density distribution, with points farther from the center contributing less to the computation of the sample mean due to lower weights assigned by the Gaussian kernel. Cheng postulated that the Mean Shift algorithm using a special type of kernel, known as the "shadow" kernel, will be in the gradient direction of the density estimate, satisfying the condition h'(r) = ck(r), where h(r) is the shadow kernel, k(r) is the original kernel, r is the distance from the center, and c is a positive constant. When the Mean Shift algorithm is performed with a kernel K that has a corresponding shadow kernel H satisfying this condition, the mean shift vector at a point x is equivalent to the gradient of the density estimate obtained by convolving the data points with the shadow kernel H, allowing the Mean Shift procedure to follow the gradient direction of the underlying density estimate and providing a principled approach to mode-seeking and clustering.

Shifting of sample mean in the gradient direction means that the sample mean is calculated in each iteration will climb the slope of gradient along the distribution until it reaches the mode. Epanechnikov kernel, as described by Fashing and Tomasi is selected in the tracking algorithm. Since it is a type of "shadow" kernel, the search window will be shifted along the gradient direction until it reaches the mode of the distribution. When mean shift reaches the mode (or peak), the algorithm converges. The proven mathematical procedure can be found in research paper by Fashing and Tomasi as already cited in earlier reference.

MeanShift algorithm is designed for static distributions. It cannot be used to handle face tracking issue/problem which involve dynamically changing distribution, such as objects in real time video sequences. Furthermore, object that moves so that the size and location of the distribution changes in time cannot be supported by MeanShift. In the case of dynamic distribution, the mode of the distribution keeps on changing the location and causes the kernel travels aimlessly without reaching the final goal, leading to poor localization, as postulated by Collins [30]. It also discussed that a kernel that has too big in size will include background clutter as well as the foreground object pixels. Moreover, large, big kernel can also fail by encompassing multiple modes. To compensate the limitation of mean shift algorithm, it has been modified to become CAMSHIFT (Continuously Adaptive Mean Shift) in which the size of the kernel is changing adaptively to deal with both static and dynamically distribution as postulated by Bradski, cited in earlier reference. The details on CAMSHIFT are explained in the next section.

4.4 CAMSHIFT Algorithm

The Continuously Adaptive MeanShift algorithm (CAMSHIFT) is an adaptation of the MeanShift algorithm for object tracking that is intended as a step towards head and face tracking for a perceptual user interface reported by Allen et al earlier. The CAMSHIFT algorithm proposed by Bradski can be summarized in the following steps:

Step (1) - Set the region of interest (ROI) of the probability distribution image to the entire image.

Step (2) - Select an initial location of the MeanShift search window. The selected location is the target distribution to be tracked.

Step (3)- Calculate a colour probability distribution of the region centered at the MeanShift search window.

Step (4) - Iterate MeanShift algorithm to find the centroid of the probability image. Store the Zeroth moment (distribution area) and centroid location.

Step (5) -For the following frame, center the search window at the mean location found in Step 4 and set the window size to a function of the Zeroth moment. Go to Step 3.

As discussed earlier in section 3.7, the MeanShift (centroid) within the search window of the discrete probability image computed in Step 3 can be calculated using the Zeroth, First and Second moments. Given that is the intensity of the discrete probability image at within the search window as follows:

Zeroth moment,

$$M_{00} = \sum_x \sum_y I(x, y) \tag{12}$$

First moment,

$$M_{10} = \sum_x \sum_y xI(x, y) \tag{13}$$

Second moment,

$$M_{01} = \sum_x \sum_y yI(x, y) \tag{14}$$

Further, the mean search window location (Centroid) is:

$$x_c = \frac{M_{10}}{M_{00}}; y_c = \frac{M_{01}}{M_{00}} \tag{15}$$

The MeanShift component of the algorithm is implemented by continually recomputing new values of (x_c, y_c) for the window position computed in the previous frame until there is no significant shift in position reported by Allen et al, cited in the reference. The algorithm must terminate in the case where M_{00} is zero, which corresponds to a window consisting entirely of zero intensity. For 2D colour probability distributions where the maximum pixel value is 255, the window size s should be set to the following:

$$s = 2 \times \sqrt{\frac{M_{00}}{256}} \tag{16}$$

A sample example illustration of the CAMSHIFT in operation can be seen in the next two figures 14-15, one of them in RGB images while the other is in probability distribution images in a video sequence respectively.

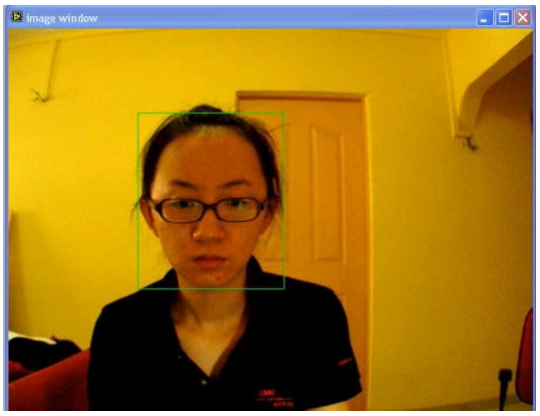

Fig. 14. An illustration of the standard CAMSHIFT tracking algorithm in RGB images

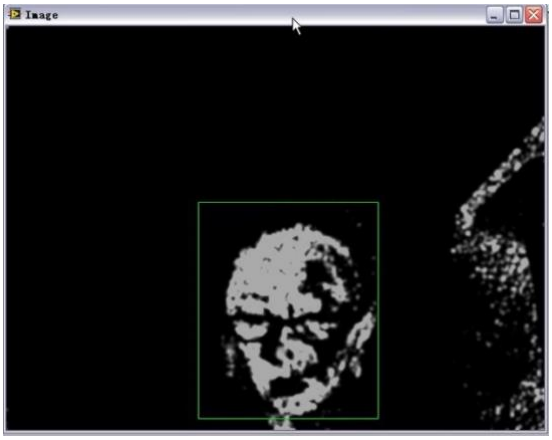

Fig. 15. An illustration of the standard CAMSHIFT tracking algorithm in probability distribution images.

4.5 Weighted Histogram

As mentioned earlier in this paper, the tracker would fail if MeanShift algorithm was implemented alone. It may consist of useless information such as background pixels within the initial search window or during the tracking time which was presented by the current authors, Alex See and Liaw in another earlier cited work in our reference. The generated 2D probability distribution image will be influenced by the untrusted lookup information. Hence, an isotropic kernel, with a convex and monotonic decreasing kernel profile $k(x)$, is chosen to assign smaller weights to pixels farther from the center, as postulated by Comaniciu et al's work cited earlier in the reference. The profile of a kernel K is defined as a function $k: [0, \infty] \rightarrow \Re$ such that $K(x) = k(\|x\|^2)$. The following equation indicates one of such kernel, called Epanechnikov Kernel, which was mentioned earlier in section 4.4, from author Y. Cheng's work.

$$K(x)\begin{cases}(1-\|x\|^2) & if\|x\| \leq 1\\ 0 & if\|x\| > 1\end{cases} \quad (17)$$

The weight distribution within a search window of a captured image was illustrated clearly in figure 7 earlier. Using these weights increases the robustness of the density estimation since the peripheral pixels are the least reliable, being often affected by occlusions (clutter) or interference from the background in consistent with other workers, Comaniciu et al cited earlier. The function $b: \Re^2 \rightarrow \{1 \ldots m\}$ associates to the pixel at location x_i^* the index $b(x_i^*)$ of its bin in the quantized feature space. The probability if the feature $u = 1 \ldots m$ in the target model is further computed as follows:

$$\hat{q}_u = C \sum_{i=1}^{n} k(\|x_i^*\|^2)\, \delta[b(x_i^*) - u] \quad (18)$$

where δ is the Kronecker delta function and C is the normalization constant derived by imposing the condition

$$\sum_{u=1}^{m} \hat{q}_u = 1 \,; C = \frac{1}{\sum_{i=1}^{n} k\left(\|x_i^*\|^2\right)} \quad (19)$$

since the summation of delta functions for u = 1…m is equal to one.

4.6 Selective Adaptation

As discussed in section 3.9, in practical face tracking video sequence, where unfortunately, probability distribution images are signficantly affected by background noise such as pixel form by skin-like objects, such as hand occlusion which cause erroneous frames and tracking loss. In this work, the authors proposed the selective adaptation algorithm imbued into the existing CAMSHIFT algorithm. There is a need to decide whether the new skin model is allowed to adapt or not. The adaptive mixture model seeks to maximize the log-likelihood of the color data over time. The normalized log-likelihood, $L^{(t)}$, of the data, $X^{(t)}$, observed from the object O at time t is given by McKenna et al cited earlier as in the reference.

$$L^{(t)} = \frac{1}{N^{(t)}} \sum_{x \in X^{(t)}} log\, p(x \mid 0) \quad (20)$$

At each time frame, $L^{(t)}$, is evaluated. A sudden large drop in its value will be observed if the tracker loses tracking the object. Adaptation will be suspended until the object is again tracked with sufficiently high likelihood in consistent with McKenna et al's work.

A temporal filter was used to compute a threshold, T_t. Adaptation was only performed when $L^{(t)} > T_t$. The median, v, and standard deviation, σ , of L were computed for the $n = 2f$ most recent above-threshold frames, where $n \leq L$. The threshold was set to the following:

$$T = v - k\sigma \quad (21)$$

where k was a constant found to be 1.2 and denotes the frame rate in Hz. Further experimental results and analysis will be described in section 5.5 of this paper.

4.7 Kalman Filtering

Kalman filtering, is also known as linear quadratic estimation (LQE), is an algorithm used in statistics and control theory. It is designed to estimate unknown variables based on a series of measurements observed over time, taking into account statistical noise and other inaccuracies. Kalman Filtering is a widely used estimation algorithm/technique. In face detecton and tracking, the region with the largest number of pixels both quantify and qualifies to be considered dominant face presence. As an illustration depicted in Figure 16, showing projection of histogram over vertical and horizontal planes/axes. The non-dominant faces are masked out and two one dimensional histograms are obtained by projecting the gray scale map along the x and y directions/axes.

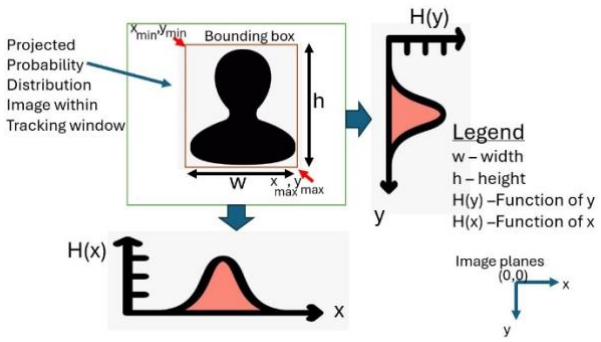

Fig. 16. An illustration of projection of histogram over two vertical and horizontal planes/axes

The Kalman Filter is integrated with CAMSHIFT to predict the centroid of the face. With reference to Figure 10, where the position of the tracking window and measurement update centroid position is fed into the Kalman Filter as inputs. The instantaneous face centre position (x_c, y_c and size (w, h) are estimated using the following equation [31]:

$$x_c = \frac{\sum_i x_i h_x(x_i)}{\sum_i h_x(x_i)}, y_c = \frac{\sum_j y_j h_y(y_j)}{\sum_j h_y(y_j)} \quad (22)$$

$$w = \alpha\sqrt{\frac{\sum_i (x_i - x_c)^2 h_x(x_i)}{\sum_i h_x(x_i)}}, \; h = \beta\sqrt{\frac{\sum_j (y_j - y_c)^2 h_y(y_j)}{\sum_j h_y(x_j)}} \quad (23)$$

where α and β are scaling factors.

The algorithm creates a bounding box around the dominant target tracked face. With reference to Figure 16, the bounding box is completely defined by two main points, which are namely P1 =(x_{max}, y_{max}) and P2 =(x_{min}, y_{min}). These two image coordinate points resides on the bounding box and they are diagonally opposite points of the box, namely, upper top left and the bottom right corners of the bounding box. These two points, P1 and P2 are derived from the expression below:

$$P1(x_{max}, y_{max}) = (x_c + {}^{w}/_{2}, y_c + {}^{h}/_{2}) \quad (24)$$

$$P2(x_{min}, y_{min}) = (x_c - {}^{w}/_{2}, y_c - {}^{h}/_{2}) \quad (25)$$

The system process model is given by the governing equations as follows:

$$X_{k+1} = AX_k + W_k \quad (26)$$

$$Z_k = HX_k + V_k \quad (27)$$

Equation (26) refers to the time update and equation (27) is the measurement update. W_k *and* V_k represent the process noise and measurement noise respectively. The White Gaussian noise is used for process noise and the measurement noise is negligible and ignore. *X* and *Z* are the state and measurement vectors respectively. X_{k+1} is the state vector at time step k+1.

During the face target tracking, the moving of the face object can be assumed as a constant velocity due to the fact that the time differences between two consecutive video frames is very small and insignificant. The following are the filter parameters:

$$X_k = x_k, y_k, v_{xk}, v_{yk} \quad (28)$$

$$Z_k = (x_k, y_k)^T \quad (29)$$

$$\hat{X}_k = \hat{x}_k, \hat{y}_k, \hat{v}_{xk}, \hat{v}_{yk} \quad (30)$$

Where $X_k = x_k, y_k, v_{xk}, v_{yk}$ represents the centroid position and the velocity of face movement in X and Y axes/planes.

where $\hat{X}_k = \hat{x}_k, \hat{y}_k, \hat{v}_{xk}, \hat{v}_{yk}$ represents the predicted centroid position and the velocity of face movement in X and Y axes/planes. The A and H are state transition matrix and measurement matrix respectively. They are further represented below:

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \quad (31)$$

$$A = \begin{bmatrix} 1 & 0 & T & 0 \\ 0 & 1 & 0 & T \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (32)$$

5. RESULTS AND DISCUSSIONS

In this section, the experimental results and discussions will focus on addressing these key challenges. They are as follows:

- Tracking under random movement
- Tracking with hand occlusion
- Tracking with multiple faces with occlusion
- Tracking under varying illumination
- Conventional CAMSHIFT model
- Analysis and comparing predicted versus actual positions

5.1 Tracking under random movement

After system initialisation, the skin is sampled/taken for processing and the system starts to track. As depicted in Figure 17, the green window is the tracking window. Two overlaying tracer dots are created within the RGB imaging. These two tracer dots were created by the developer for the purpose of performance measurement and comparison for the Kalman Filtering which will be described. The Blue tracer dot is the measured centroid position, and the red tracer dot is the centroid position predicted by Kalman filter. The red dot is also the center of the calculation window, and the size of the calculation window is the same as the tracking window at previous frame. The calculation is not shown in the image. Due to the nature of the smaller number of pixels for each tracer dots, the top image of Figure 17 depicted a magnified view/version with caption to explain it clearly to the reader.

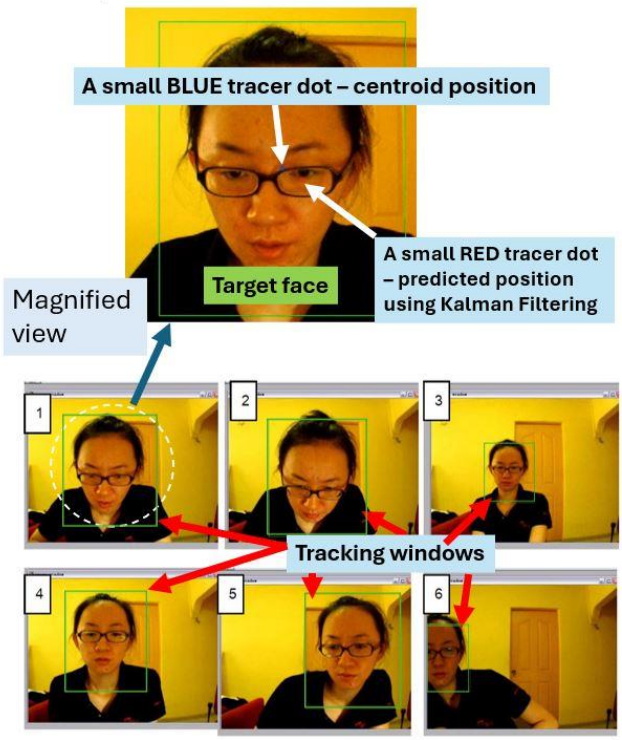

Fig. 17. Running sequence of RGB images of implemented enhanced CAMSHIFT algorithm with Kalman Filtering Tracker with tracking window and two tracer dots in the image.

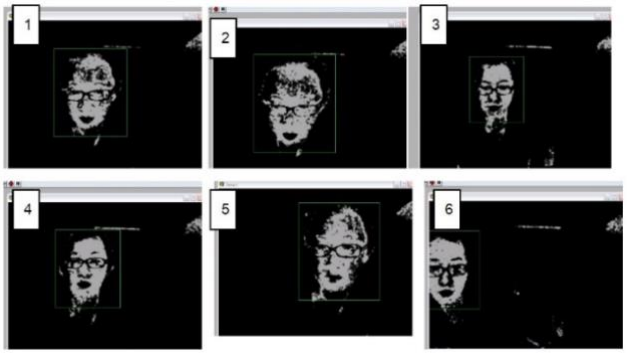

Fig. 18. Probability distribution images of implemented enhanced CAMSHIFT algorithm with Kalman Filtering tracker with tracking windows shown in green colour.

As the interval time between the consecutive image frames is very small, the moving of the tracked face target can be assumed to be of constant speed. The Kalman Filter uses this assumption to predict the object position for the next frame. For the CAMSHIFT, the tracking window and the calculation window are the same and called search window. If conventional without any prediction algorithm put in place, the search window is always behind the face moving. In the current proposed algorithm implementing Kalman Filtering with position prediction, the calculation window's moving is almost the same as the face, which make the calculation for the zeroth movement significant and highly accurate. Our similar findings and in agreement with reference work conducted by Chen et al [32], where the researchers reported using Kalman Filtering with CAMSHIFT and Adaboost algorithm. Although the traditional CAMSHIFT algorithm can track the moving object quite well, however it fails to track the object easily while the object is occluded and interfered by the same colour obstructions. Further, our observed experimental results has been shown that an accelerating object can be tracked by adding a Kalman estimator of target position and velocity coinciding with researchers work as reported in [33-35].

5.2 Tracking in the presence of hand occlusion

The CAMSHIFT algorithm fails to track face easily while it is occluded has been reported in many literatures, which is reported elsewhere. The CAMSHIFT algorithm, while effective in many object tracking scenarios, encounters significant limitations when dealing with slow-moving occlusions. Specifically, if a human subject's hand moves slowly across their face, CAMSHIFT tends to shift its focus from the face to the hand. This occurs because the hand's color distribution becomes dominant within the search window, leading the tracker to incorrectly follow the hand instead of the face.

In this paper, the authors hope to address this limitation with an improved algorithm integrating a Kalman filter has been proposed in this work. The Kalman filter enhances the system's ability to predict the movement of the tracked object, allowing it to maintain focus on the face even when it is occluded by slowly moving hands. The predictive capability of the Kalman filter ensures that the tracking window remains correctly centered on the face, regardless of the hand's movement speed. Experimental results shown in Figure 19 and 20 demonstrate the effectiveness of the proposed algorithm. In tests where the subject's hand moves slowly across the face, the Kalman filter successfully predicts the face's position, preventing the tracker from shifting to the hand. This results in the tracking window adjusting appropriately to accommodate the occlusion, maintaining accurate face tracking. Overall, the integration of the Kalman filter overcomes the CAMSHIFT algorithm's limitation in handling slow-moving occlusions. The enhanced system reliably tracks the face, ensuring robust performance even in complex visual environments where slow occlusions occur. This improvement significantly enhances the reliability and accuracy of object tracking systems, which is also consistent with other researchers report as in [36].

With reference to Figure 19, labelled image 4, the tracking window covers the face and hand. As the face does not move and the speed of the face is zero, the predict center position by Kalman filter is still on the face when the hand moves away from the face in a very slow motion. The calculation window is still on the face, and the centroid position calculated by CAMSHIFT will feedback to Kalman filter and prove the prediction of Kalman filter is right. So in the labelled images 4 & 5 of Figure 20, the tracker still effectively and efficiently follows the face when hand moves away from the tracking window.

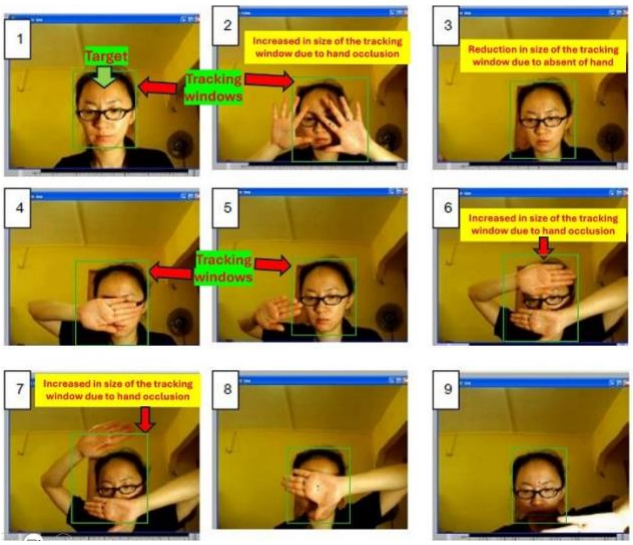

Fig. 19. RGB images for face tracking with hand occlusions with the enhanced CAMSHIFT algorithm.

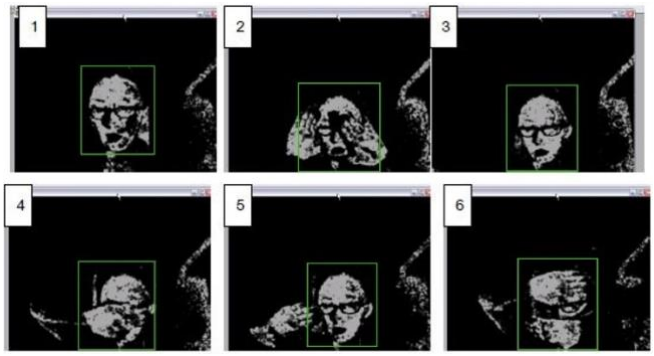

Fig. 20. Observed probability distribution images for tracking with hand occlusions with the enhanced CAMSHIFT algorithm.

Explanation of what happen during hand occlusion and the observations are described in the following. Figure 21 depicts a detailed graphical and text explanation on hand over face occlusion situation. In Figure 21, the initial analysis shows that a hand obscuring the face, the elliptic shape window symbolises the target face, the centroid position, hand symbol and the tracking window are clearly visible in the legend of this picture. The interpretation of the consecutive sequences of illustrated video frames are from starts from frame number 1 and ends at frame number 4. The employment of both CAMSHIFT and Kalman Filtering are effective and from the observation of the analysis, it becomes very clear that the CAMSHIFT feedback of the centroid position to the Kalman Filter, which facilitated computation and thus produced the predicted centroid position of the target face that is consistent and remain unchanged. This allows the consistency of providing the bounding tracking window to remain within the desired target face tracking perimeter. This becomes evident when the hand exits the camera field of view and leaves the tracking window as depicted in Figure 21. The centroid position is computed as the centre of the target face, and this is feedback by CAMSHIFT to the Kalman Filter as described earlier.

In the surveyed two studies [37-38], research papers on real-time hand detection during hand-over-face occlusion and face tracking with occlusion present significant advancements in the field of visual tracking algorithms. The first study proposes a robust hand detection system that addresses occlusion issues by employing a combination of skin colour detection, ellipse template matching, CAMSHIFT tracking algorithm, force field method, and Sobel edge extraction. This integrated approach enables accurate segmentation of hands even when they partially cover the face, facilitating effective cursor control on computers.

The second study enhances face tracking capabilities by integrating the CAMSHIFT algorithm with the GM(1,1) model, leveraging motion vector information to improve occlusion handling. This hybrid method predicts face positions using historical motion data, thus maintaining tracking accuracy despite occlusions. The combination of these techniques significantly reduces iteration times and enhances real-time performance, ensuring robust face tracking even under severe occlusion conditions.

Together, these papers highlight the critical importance of combining multiple algorithms and prediction models to tackle the challenges of occlusion in real-time hand and face detection systems. Their proposed methods show promising results in maintaining tracking accuracy and system robustness, which are essential for applications in human-computer interaction, video conferencing, and security monitoring. In retrospection similar to the earlier researchers, the current authors' implemented enhanced CAMSHIFT with Perceptual Grouping, weighted histogram and Selective Adaptation and with Kalman Filtering have collective effort to further enhance face tracking stability and robustness in handling occlusions to the target tracking face.

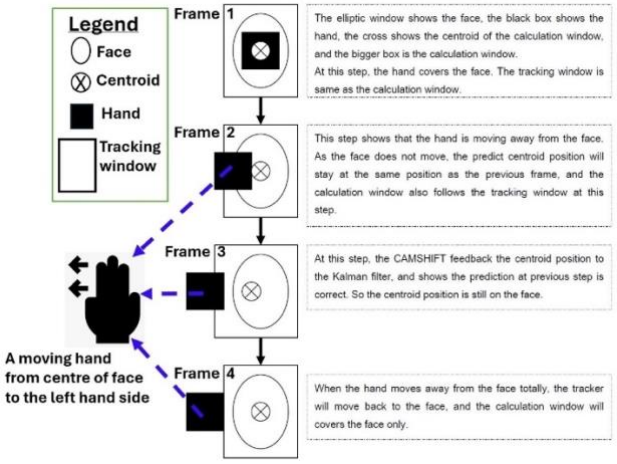

Fig. 21. A detailed graphical and text explanation on hand occlusion.

5.3 Tracking under multiple faces with occlusion

In the next experimental test on the robustness of the enhanced CAMSHIFT algorithm under external face disturbance due to the presence of multiple faces in the foreground. Two other human faces were introduced and deliberately seen moving randomly in front of the targeted face. This can be challenging since most human faces have the same hue value and the tracker rely heavily on this to track.

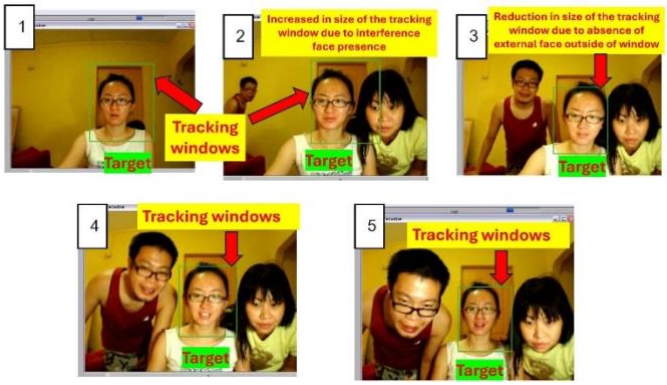

Fig. 22. Running sequence of RGB images in a video running sequence using enhanced CAMSHIFT model subjected to disturbance from multiple faces

Fig. 23. Probability distribution images in a video sequence for disturbance from multiple faces experimentation using Original CAMSHIFT

As shown in the RGB video sequence in Figure 22, labelled image 1 depicts a single target face who is the subject on interest. In labelled image 2, it is evident that an interference female subject has partially entered into the tracking window, which resulted in a slight increase in the size of the tracking window. The probability distribution images are depicted in Figure 23. The tracking managed to stay focus on its target. Since this enhanced CAMSHIFT algorithm model with add-on features with key Kalman Filtering technique can effectively cope and ignore outliers in the field of view of the camera. The sequence of images as shown in the probability distribution profiles suggest that the other two interference faces are present with increase presence of the grey scale patches of face distributions. The authors of this paper postulate that the built-in weighted histogram and weight assignment in this algorithm has played a significant role in battling multiple faces occlusion/interference. The pixels that are further away from the center of tracking window are assigned lower weights. The algorithm will disregard those pixels that lie below an acceptable threshold weight value and deemed them as non-trustable and unreliable. As depicted in Figure 22, labelled image 4 to image 5, when the dominant target face moves away from the camera, the target face becomes smaller, resulting in reduce size of the tracking window. In addition, the other interference faces were not seen to be moving and only that dominate face has moved backward. The Kalman Filter predicts the centroid position at the next frame.

The authors experimented with the new robust & resilient enhanced CAMSHIFT model with Kalman Filtering, face tracking has been reported to be very successful even with disturbance and interferences from multiple faces.

In the following, the authors will present the sub-optimal performance of conventional/original CAMSHIFT which has failed to track the target face. The video sequence is running images in sequential order from 1 to 6 as depicted in Figure 24. The converted probability distribution images of the same experiment is depicted in Figure 25. It is evident and clear that traditional CAMSHIFT algorithm can track individual face target. As soon as when the human face starts moving away from the camera and into a nearby region of another person with skin-alike histogram colour space, the search window has two modes of distribution to climb if there is no weight differences between the pixels of concern. One such possibility and probability is that the tracker will follow and climb the wrong gradient/hill of the distribution. As a result, the wrong face target is tracked that leads to tracking failure as depicted in both Figures 24 and 25.

Fig. 24. RGB images in a video sequence using conventional/original CAMSHIFT subjected to a disturbance from an interference face. Tracking was lost on labelled image 6.

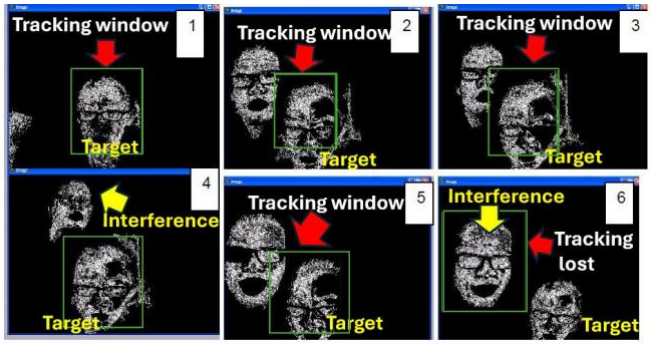

Fig. 25. Probability distribution images in a video sequence using conventional/original CAMSHIFT subjected to a disturbance from an interference face. Tracking was lost on labelled image 6.

5.4 Tracking under lighting illumination variation

As shown in the RGB video sequence in Figure 26, labelled image 1 depicts a single target face who is the subject on interest. The sources of strong lighting illumination in the environment and its background are clear as depicted in labelled images 1 to 6. The enhanced CAMSHIFT model was tested where the human subject was placed under an environmental background with disturbances from strong interferences of light sources.

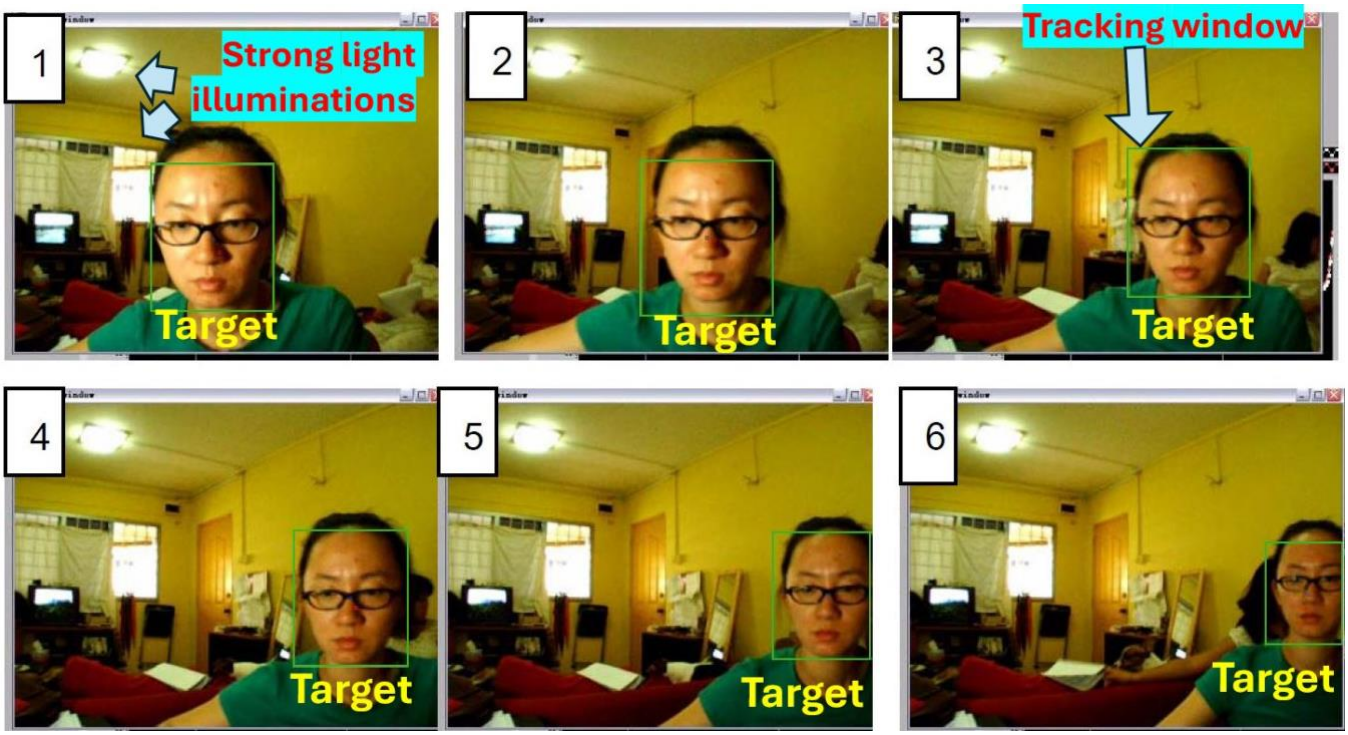

Fig. 26. RGB colour images in a video sequence using enhanced CAMSHIFT with perceptual group, weighted histogram, selective adaptation and with Kalman Filtering. Labelled image (1) corresponded to Frame no. 20, image (2) to Frame no. 56, image (3) to Frame no. 107, image (4) to Frame no. 171, image (5) to Frame no. 192 and image (6) to Frame no. 214.

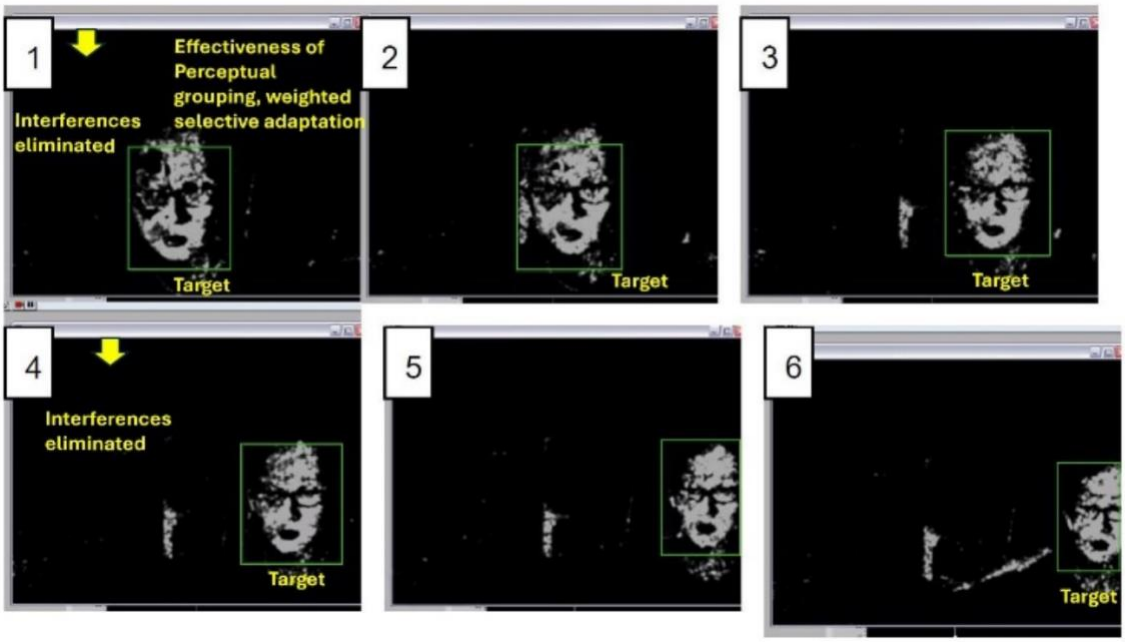

Fig. 27. Probability distribution images in a video sequence using enhanced CAMSHIFT with perceptual group, weighted histogram, selective adaptation and with Kalman Filtering. Labelled image (1) corresponded to Frame no. 20, image (2) to Frame no. 56, image (3) to Frame no. 107, image (4) to Frame no. 171, image (5) to Frame no. 192 and image (6) to Frame no. 214.

Further observations from Figure 27 revealing the probability distribution images can be seen that perceptual grouping technique has successfully eliminated the interferences of bright light sources on the left-hand side of the images. It is timely to perform an analysis of the dynamic Selective Adaptation technique presented in this work.

In conventional CAMSHIFT developed by original researcher, Bradski, his CAMSHIFT model does not have any form of adaption, as all lookup processes are based only an initial fixed histogram which is only binned once at the beginning of initialisation part of the program. In real-world application where the lighting illuminations maybe dynamically changing all the time.

In the development and implementation by the current authors of this paper, the enhanced CAMSHIFT model is boasted of its robustness and resilience to environmental light illuminations and occlusions. The dynamic adaptation receives new information about the Hue value over the whole range of brightness intensity.

The Hue in the brightest and the lowest lighting intensity condition are sampled in real time to form the lookup table. If no adaptation of colour skin is implemented, the face tracker will be losing its tracking capability when subjected to challenging illumination conditions. Furthermore, adaptation of new pixels may include pixel such as unrelated background pixels or other object pixels within the tracking window during the tracking process. If the error is not detectable, the lookup will result in the non-skin segmentation and eventually lost the tracking. However, if error is detected, adaptation should be suspended and the lookup should resume back to the previous accepted skin model to form the probability distribution image of the next video frame, and this is generally the genuine selective adaptation algorithm.

The entire Selective Adaptation algorithm/process was applied accordingly to track a human subject's face, depicting a video sequence containing probability distribution images as shown in figure 27. In figure 26, lost of tracking was analysed by examining the dynamic selective adaption plot versus image frame number, where the the normalized log-likelihood was lesser than the threshold value, $L^{(t)} < T_t$ occurring three times, at marked labelled position in Figure 28. The dynamic selective adaptation occurred at first at frame 20 (**Point A**), second time occurred between frame no. 170 (**Point B**) and frame no. 175 (**Point C**) and final time at frame no. 219 (**Point D**). At these few occurrences, it was examined that there are few changes in the normalised log-likelihood values. Since adaptation is performed only when $L^{(t)} > T_t$, therefore adaptation is suspended until the target is tracked again with sufficiently high likelihood, which in these cases occurred at Point A, between B and C as well as Point D. In order for the resumption of adaptation, the Hue value will be increased for effective tracking

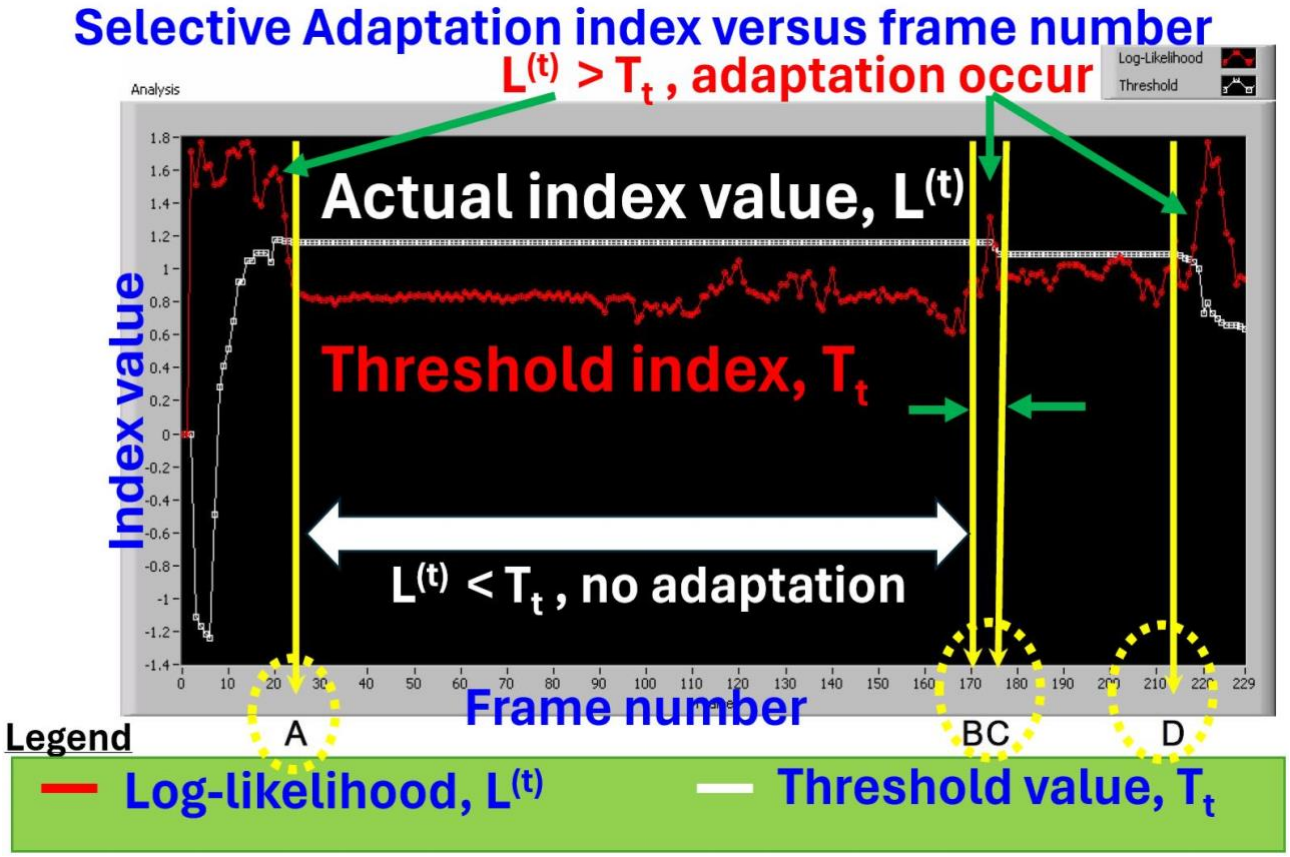

Fig. 28. Analysis of the Selective Adaptation examining in detail frame by frame of the images and the computed log likelihood index values and threshold value

5.5 Performance of Original CAMSHIFT model

For a comparative experimental investigation, the conventional original CAMSHIFT developed by Bradski, was put to test. The conventional CAMSHIFT model is a simple, yet computationally efficient face and colour object tracker. After numerous tests, there are observable performance limitations encountered by the original CAMSHIFT model. Two comparative experimental tests were performed using purely the original CAMSHIFT model. In Figure 29 and 30, they depicted the original CAMSHIFT algorithm which is subjected to hand occlusion. If a user places her hand, the covers her face in an extremely slow motion, it can be observed in Figures below that the CAMSHIFT tracker will be misled, climb onto the wrong mode (peak) of the probability distribution. Therefore, it will be misguided into tracking the hand as opposed to correctly tracked the face.

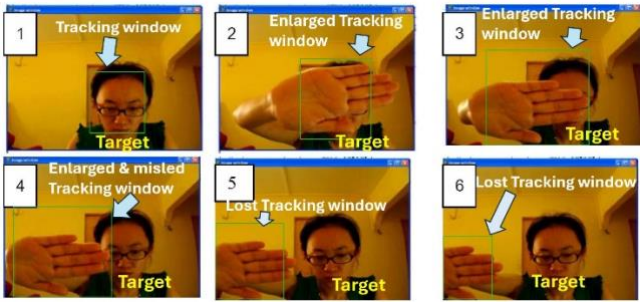

Fig 29 depicts RGB colour images in a video sequence using original CAMSHIFT

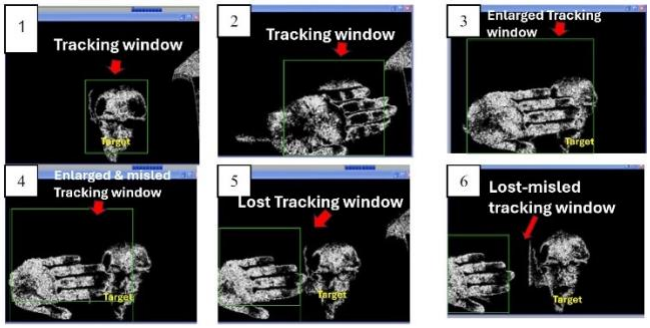

Fig. 30. Probability distribution images in a video sequence using original CAMSHIFT

In a separate experimental test, this original CAMSHIFT model developed by Bradski was subjected to environmental condition which has varied lighting illuminations. Figures 31 and 32 depict the RGB colour images and probability distributions images when this original CAMSHIFT was subjected to varying lighting illumination. Additionally, the background exists a door with skin alike colour, which has similar Hue value when examined in terms of histogram distribution. As it may be seen from the Figures 31 & 32, if there are many objects with skin alike colour tone, the conventional CAMSHIFT may consider that the search window has two or more modes of distribution to gradient climb if there are no weights assignment differences between the pixels. This original CAMSHIFT unlike the enhanced version, it does not have Perceptual Grouping filtering technique. As a result in Figure 32, the intensity of the grey scale distribution of the door, which has skin alike colour tone seems very dominant. This causes severe CAMSHIFT performance degradation. Our findings in this paper have good consistency with researchers in [39], who applied other pre-processing techniques, such as Gauss foreground detection in conjunction with CAMSHIFT and Kalman filter to ensure tracking stability and performance enhancement in moving object detection and tracking. Further in research papers [40-41] further resonated the importance of face detection and tracking, as much as the current authors' contributions in this field of study. The two papers combined provide insights into the application of the CAMSHIFT algorithm in different contexts. The first paper, "*Fuzzy System Based Face Tracking for Head Movement Control in Progressive Health Care*," highlights the use of CAMSHIFT for face tracking in a healthcare setting. This system integrates CAMSHIFT with fuzzy logic control to enhance the accuracy and efficiency of face tracking for head movement control, demonstrating improvements in user interface and motor control for real-time applications. The second paper, "*High Precision Indoor Positioning Method Based on Visible Light*" although primarily focused on indoor positioning, supports the integration of CAMSHIFT by providing robust object tracking mechanisms under varying conditions. The combined insights underscore the versatility of CAMSHIFT in handling dynamic tracking scenarios, particularly when enhanced with additional computational techniques such as fuzzy logic for precise and reliable performance in both healthcare and positioning systems. However, as postulated by the current authors and other researchers, the two performance limitations of original CAMSHIFT model are namely: changes in lighting illuminations and occlusions, which are major barriers to adoption if used on its own.

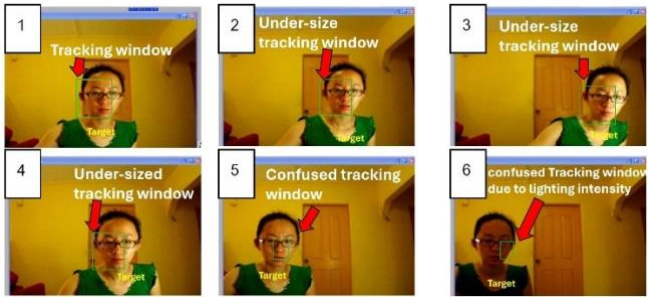

Fig. 31. RGB colour images in a video sequence using original CAMSHIFT subjected to varying lighting illumination.

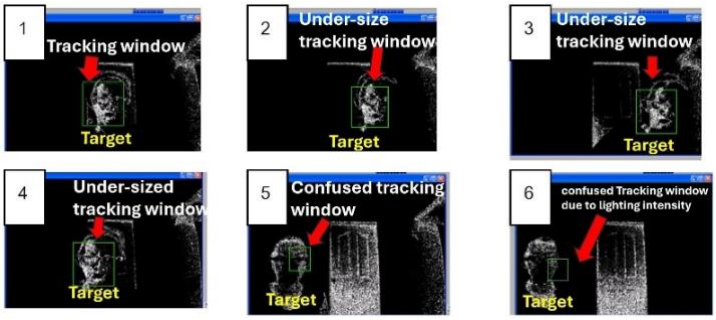

Fig. 32. Probability distribution images in a video sequence using original CAMSHIFT subjected to varying lighting illumination.

5.6 Analysis and comparing predicted versus actual x, y coordinate positions with Kalman Filtering

The Figure 33 shows the actual centroid position versus predicted centroid position. The red line shows the predicted data path, and the black line shows the measurement data path (also known as actual data path). To make the figure readable, only 11 frames data are used to plot the figure. The data of actual position is collected from frame 170 to frame 181, the data of predict position is collected from frame 169 to frame 180, and the data analysis is also based on those 11 pairs of data and shows in table 1 and table 2. The frame 170 was not included because from our analysis, the image was in sudden transient movement which led to a huge spike in Y coordinate value. It is considered as an outlier data point which is ignored in this following calculation.

Mean Absolute Percentage Error (MAPE) is one of the most popular measurement metrics and it is used in the measurement of the forecast accuracy as reported in International Journal of Forecasting cited in [42]. This metric is used and applied in this study. From the earlier mentioned literature, MAPE formula is adapted and reported as follows:

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{Actual_{Measured} - Predicted_{value}}{Actual_{Measured}}\right| X\ 100\ \% \quad (33)$$

Table 1. Mean Absolute Percentage Error (MAPE) calculation for x coordinates

Image Frame No.	Actual Measured (x) versus Predicted ($\hat{X}$)			
	Measured (x)	*Predicted($\hat{X}$)*	*Prediction Error (x- $\hat{X}$)*	*MAPE (%)*
170	276.3	274.8	1.6	
171	251.4	260.8	-9.5	
172	229.5	226.9	2.6	
173	186.3	186.8	-0.5	
174	165.8	167.5	-1.7	
175	118.8	130.9	-12.1	9.32%
176	80.6	84.7	-4.1	
177	56.9	63.2	-6.3	
178	47.6	45.1	2.4	
179	50.4	68.5	-18.1	
180	45.8	49.6	-3.8	
181	49.9	60.3	-10.4	

*Note: N =11 sample points, where image frame no. 170 is considered as an outlier, data point is ignored.

Table 2. MAPE calculation for Y coordinates

Image Frame No.	Actual Measured (x) versus Predicted ($\hat{Y}$)			
	Measured (y)	*Predicted($\hat{Y}$)*	*Prediction Error (y-$\hat{Y}$)*	*MAPE (%)*
170	274.8	175.2	99.5*	
171	260.8	258.7	2.1	
172	226.9	220.9	6.0	
173	186.8	199.0	-12.2	
174	167.5	176.5	-9.0	
175	130.9	120.9	10.0	9.70%
176	84.7	80.6	4.1	
177	63.2	57.8	5.4	
178	45.1	49.4	-4.2	
179	68.5	80.0	-11.5	
180	49.6	64.2	-14.7	
181	60.3	51.4	8.9	

*Note: N =11 sample points, where image frame no. 170 is considered as an outlier, data point is ignored.

It may be seen that the calculated MAPE values for x and y coordinates are 9.32% and 9.07% respectively. Thus, the MAPE for the x-coordinate is 9.32%, indicating that on average, the forecast/predicted by Kalman Filtering algorithm deviates from the actual values by 9.32%. Similarly, the forecast/predicted for the y-coordinate is 9.07% is an average deviation between predicted from actual measured values of the centroid position of the tracked face target. These low percentages are considered reasonable and good performance when Kalman Filtering is integrated with CAMSHIFT. The overall prediction deviation from actual measurement is less than 9.5%. So far to date, to the best of authors' knowledge, such MAPE computation or values have not been reported elsewhere yet.

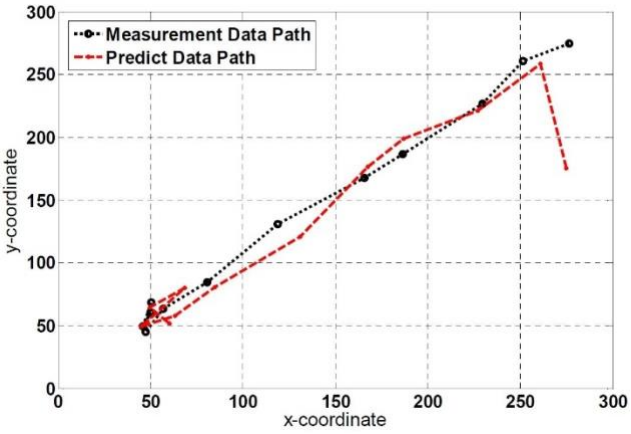

Fig. 33. The experimental measurements of acquired actual centroid position versus predicted centroid position. The red line shows the predicted data path, and the black line shows the measurement data path (also known as actual data path)

A secondary analysis was performed and the use of Root Mean Squared Error (RMSE) computation.

The formula for the RMSE is given as follows:

$$RMSE = \sqrt{\frac{\sum_{i}^{n}[(Measured(x_i) - Predicted(\hat{X}_i)]^2}{n}} \quad (34)$$

Where n is the number of data points, Measured(x_i) and Predicted($\hat{X}_i$) are the individual sample points indexed by i. x and y are X and Y pixel coordinates datasets respectively.

For this analysis, we have separately offline computed the RMSE for the X coordinates to be equal to 8.3 pixel from the observed range of predicted ($\hat{X}$) values 45.1 to 260.8. Similarly, from a Microsoft Excel spreadsheet offline computation of the RMSE for the Y coordinates, it was calculated to be 8.8 pixel from an observed range of predicted ($\hat{Y}$) values 49.4 to 258.7.

So RMSE essentially calculates the root of the average squared error between the predicted and measured values. It provides a measure of the typical magnitude of the prediction errors in the same units as the original values. The RMSE is a widely used metric to evaluate the performance of regression and forecasting models. A lower RMSE indicates better model performance and more accurate predictions on average. RMSE is always non-negative, with a value of 0 representing a perfect fit between predicted and measured. In this work, the low RMSE values of X and Y coordinates reported as 8.3 pixel and 8.8 pixel suggest that the Kalman Filtering predicted values are reliable

and this Kalman Filtering prediction model is dynamic with good accuracy when it is integrated with CAMSHIFT algorithm.

It should be further observed from both Table 1 and 2, the distribution of errors: The x-errors and y-errors have both positive and negative values, suggesting no systematic bias in over- or under-prediction in this developed Kalman Filtering model.

5.7 Limitation of this study

The limitation of this study includes the fact that there is no database of the skin sample, a human subject must be present to be prompted to sample the face skin information before this algorithm can even work properly. Besides, the image acquisition must be taken at the frontal image with the face, two eyes directly facing the web-camera. If only a partial face or even with the back of the head facing the camera, then skin sampling will not be achieved successfully. These are current challenges/barriers which can be further addressed in future work/research.

Separately, one other limitation in this study is that throughout the experimental work presented, the occlusions are mainly mobile hand or multiple interference faces moving around in the presence of the target face, which is not really in fast motion. We have explored a scenario where initially the moving hand occlusion is deliberately obscuring the target face. Once the hand occlusion has occurred such as the target face, which deliberately moves away from the tracking window. This confuses and led to a failed tracking scenario. In order to migitate and track the face after it has been blocked/obscured, suggest that an alternative algorithm should be able to detect the hand object, which has identical skin tone colour as the originally sampled skin. The shape of the initial face should be detected and recorded with more intelligent techniques, perhaps using deep learning, machine learning or AI approaches. This should provide computer vision researchers who may be keen to explore as another research project.

6. CONCLUSION AND FUTURE WORKS

In this paper, the authors have successfully presented a working experimental enhanced CAMSHIFT model with Perceptual Grouping, weighted histogram assignment with Selective Adaptation and Kalman Filtering technique for face detection and tracking. Besides, the authors of this paper have presented a detailed and comprehensive discussion on both the theoretical and experimental aspect of face detection and tracking in this implementation. It is evident that in this research contribution work, the combined use of Perceptual Grouping, weighted histogram with selective adaptation is very effective in eliminating background noise interference especially from lighting illumination or objects with similar skin alike tone colour.

In experimental analysis, it is evident that this proposed system can successfully outperform the original CAMSHIFT model in the key five different scenarios as follows:

- Tracking under random movement/motion
- Tracking in the presence of hand occlusion
- Tracking under multiple faces with occlusion
- Tracking under lighting illumination variation
- Performance of original CAMSHIFT model
- Analysis and comparing predicted versus actual x, y coordinate positions with Kalman Filtering

Further, from the experimental data analysis for selective adaptation, it can be said that the normalised log-likelihood index, $L^{(t)}$ is a powerful, authentic marker/indicator to analyse as a sudden decrease in this value falling below the threshold target will imply that the face tracking has been suspended or discontinued.

For occlusions, this enhanced CAMSHIFT algorithm has proven its robustness defined as its ability to maintain stability and functionality in its continuous tracking through the use of Kalman Filtering predicted values, adaptation of the target face, despite external interferences such as hand and face which cause occlusion. This model has also proven its merit to be robust, which is defined as system's ability to continue its performance despite external interferences and outperform the traditional CAMSHIFT algorithm developed by Bradski. From the experimental data and analysis of the tracking with prediction, the enhanced CAMSHIFT model have achieved low Mean Absolute Percentage Error (MAPE) both predicted ($\hat{X}$) and Predicted ($\hat{Y}$) at 9.32 % and 9.70 % respectively. This strongly suggest and indicate that on average, the forecast/predicted by Kalman Filtering algorithm deviates from the actual values by low margin and this enhanced CAMSHIFT model is effective and reliable in predicting and tracking the face target.

Further, the RMSE for the calculated X coordinate is equal to 8.3 pixel from the observed range of predicted ($\hat{X}$) values 45.1 to 260.8. For the Y coordinates, it was calculated to be 8.8 pixel from an observed range of predicted ($\hat{Y}$) values 49.4 to 258.7. This strongly suggests low values of the RMSE provided a measure of the low magnitude of the prediction errors in the same units as the original values, therefore accuracy is ascertained.

In conclusion, this experimental research significantly advances the body of literature on face detection and tracking, specifically through the application of the CAMSHIFT algorithm. The use of a webcam in this study offered practical advantages for software development. However, future research would benefit from the incorporation of high-performance camera systems, which are expected to enhance the algorithm's efficacy and overall performance.

ACKNOWLEDGEMENT

The authors would like to extend their sincere gratitude to Ngee Ann Polytechnic, School of Engineering, and the University of Newcastle, School of Electrical Engineering and Computer Science, Australia, for their support in this work.

REFERENCES

[1] S. Du. CAMShift-Based Moving Object Tracking System. IEEE 3rd International Conference on Electronic Technology, Communication and Information (ICETCI) 2023.

[2] Z. Jiang, R. Li and C.Z. Zhu. Remote Sensing Image Target Recognition System of Tennis Sports based on CAMSHIFT Algorithm. International Conference on Information System, Computing and Educational Technology (ICISCET) 2022.

[3] H. Sun, H.H. Chen, X. Cui and J. X. Wang. Vehicle Flow Statistics System in Video Surveillance based on CAMSHIFT and Kalman Filter. International Conference on the Software Process 2021;362–6.

[4] X. Li, M. Liu, S. Zhang and R. Zheng, "Fish Trajectory Extraction Based on Object Detection," 2020 39th Chinese Control Conference (CCC), Shenyang, China, (2020), pp. 6584-6588, doi: 10.23919/CCC50068.2020.9188642.

[5] T. Jaichuen, N. Ren, P. Wongapinya and S. Fugkeaw, "BLUR & TRACK: Real-time Face Detection with Immediate Blurring and Efficient Tracking," 2023 20th International Joint Conference on Computer Science and Software Engineering (JCSSE), Phitsanulok, Thailand, (2023), pp. 167-172, doi: 10.1109/JCSSE58229.2023.10202064.

[6] S. Guo, C. Handong, J. Guo and J. Xu J. A Novel Target Tracking System for the Amphibious Robot based on Improved Camshift Algorithm. IEEE International Conference on Mechatronics and Automation (ICMA) (2021); pg 1419-1424.

[7] X. Hu and B. Huang. Face Detection based on SSD and CamShift. IEEE Joint International Information Technology and Artificial Intelligence Conference 2020, pg 2324-2328.

[8] Y. Zhang. Detection and Tracking of Human Motion Targets in Video Images Based on Camshift Algorithms. IEEE Sensors Journal 2020;20:11887–93.

[9] S. Salankar and R. Bankar, "A Vision Based Face Tracking using Camshift with BLBP Algorithm in Head Gesture Recognition System," 2021 6th International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India, (2021), pp. 861-865, doi: 10.1109/ICICT50816.2021.9358481.

[10] N. Zhang, J. Zhang, Optimization of Face Tracking Based on KCF and Camshift, Procedia Computer Science,Vol. 131, 2018, pg 158-166, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2018.04.199.

[11] G.R. Bradski. Real time face and object tracking as a component of a perceptual user interface. In: Applications of Computer Vision. 1998. p. 214–219.

[12] M.S. Khalid, M.U. Ilyas, M.S. Sarfaraz and M. A. Ajaz. Bhattacharyya Coefficient in Correlation of Gray-Scale Objects. Journal of Multimedia 1(1); (2006): pg57-61.

[13] L. Soni and A. Waoo. A Review of Recent Advances Methodologies for Face Detection, International Journal of Current Engineering and Technology (2023): 13. 86-92. 10.14741/ijcet/v.13.2.6.

[14] Y. Himeur, S. Al-Maadeed, I. Varlamis, N. Al-Maadeed, K. Abualsaud, A. Mohamed. Face Mask Detection in Smart Cities Using Deep and Transfer Learning: Lessons Learned from the COVID-19 Pandemic. Systems. 2023; 11(2):107. https://doi.org/10.3390/systems11020107

[15] A. Kumar, A. Kaur and M. Kumar. Face detection techniques: a review, Artificial Intelligence Review (2019): 52, 927-948. https://doi.org/10.1007/s10462-018-9650-2

[16] J. Li, C. Yang, F. Yang, J. Huang, W. Wei, S. Zhang, X. Zuo and S. Zhang. "Face Detection and Tracking Based on Neural Network," 3rd International Conference on Information Science, Parallel and Distributed Systems (ISPDS), Guangzhou, China, (2022), pp. 257-260, doi: 10.1109/ISPDS56360.2022.9874114.

[17] K. Fukunaga, L.D. Hostetler, "The Estimation of the Gradient of a Density Function, with Applications in Pattern Recognition", in IEEE Transactions on Information Theory (1975), Vol. 21, Issue 1, January 1975, 32-40.

[18] S.J. McKenna, Y. Raja, S. Gong, "Tracking colour objects using adaptive mixture models", in Image and Vision Computing (1999), vol. 17, pp. 225-231.

[19] S. Gong, S.J. McKenna, A. Psarrou, "Review of Dynamic Vision: From Images to Face Recognition", in Imperial College Press. 2000.

[20] K. Schwerdt, J.L. Crowley, "Robust Face Tracking using Color", in Automatic Face and Gesture Recognition (2000), Fourth IEEE International Conference on 2000, page(s): 90-95, ISBN: 0-7695-0580-5

[21] S. Spors, R. Rabenstein, "A Real-Time Face Tracker For Color Video", in Acoustics, Speech, and Signal Processing (2001), Proceedings of IEEE International Conference (ICASSP '01), Volume 3, 7-11 May 2001, Page(s): 1493-1496, ISBN: 0-7803-7041-4

[22] T. Wang, Q. Diao, Y. Zhang, G. Song, C. Lai, G. Bradski, "A Dynamic Bayesian Network Approach to Multi-cue based Visual Tracking", in Pattern Recognition (2004), ICPR 2004, Proceedings of the 17th International Conference, Vol. 2, 23-26 Aug 2004, Page(s)167-170, ISSN: 1051-4651, ISBN: 0-7695-2128-2.

[23] Y. Cheng, "Mean Shift, Mode Seeking, and Clustering", in IEEE Transactions on Pattern Analysis and Machine Intelligence (1995), Vol. 17, Issue 8, August 1995, Page(s) 790-799

[24] Alex See Kok Bin and Yee Kang Liaw, Face Detection and Tracking Utilizing Enhanced CAMSHIFT Model, International Journal of Innovative Computing(IJICIC), Vol. 3, Issue 3, pp 597-608, ISSN 1349-4198.

[25] D. Comaniciu, V. Ramesh, P. Meer, "Kernel-Based Object Tracking", in IEEE Transactions on Pattern Analysis and Machine Intelligence (2003), Volume 25 Issue 5, May 2003, Page(s): 564-577, ISSN: 0162-8828.

[26] B.W. Silverman, "Density Estimation for Statistics and Data Analysis", in Monographs on Statistics and Applied Probability, London: Chapman & Hall (1986)

[27] J.G. Allen, R.Y.D. Xu, J.S. Jin, "Object Tracking Using CamShift Algorithm and Multiple Quantized Feature Spaces", in Proceedings of the Pan-Sydney area workshop on Visual information processing (2004), Pages: 3 – 7, ISBN ~ ISSN:1445-1336, 1-920682-18-X

[28] A.K.B. See and H.W. Goh, "Robust, Resilient Enhanced CAMSHIFT Model: Advancing Face Detection and Tracking Stability in Challenging Environments", Malaysian J. Sci. Adv. Tech., vol. 4, no. 1, pp. 44–67, Jan. 2024. https://doi.org/10.56532/mjsat.v4i1.238

[29] M. Fashing, C. Tomasi, "Mean Shift is a Bound Optimization", in IEEE Transactions on Pattern Analysis and Machine Intelligence (2005), Volume 27, Issue 3, March 2005, Page(s) 417-474

[30] R.T. Collins, "Mean-shift Blob Tracking Through Scale Space", in Computer Vision and Pattern Recognition (2003), IEEE Computer Society Conference, Volume 2, 18-20 June 2003 Page(s): II - 234-240

[31] P.K. Turaga, G. Singh and P.K. Bora, "Face Tracking using Kalman Filter with Dynamic Noise Statics", IEEE TENCON (2004). IEEE Region 10 Conference, Volume A, 21-24.

[32] K. Chen, C. Liu and Y. Xu. Face Detection and Tracking Based on Adaboost CamShift and Kalman Filter Algorithm. In: Fei, M., Peng, C., Su, Z., Song, Y., Han, Q. (eds) Computational Intelligence, Networked Systems and Their Applications. ICSEE LSMS (2014) Communications in Computer and Information Science, vol 462. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-45261-5_16

[33] A. Salhi, Y. Moresly, F. Ghozzi and A. Fakhfakh, "Face detection and tracking system with block-matching, meanshift and camshift algorithms and Kalman filter," (2017) 18th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering (STA), Monastir, Tunisia, 2017, pp. 139-145, doi: 10.1109/STA.2017.8314915.

[34] C.J. Hsieh and K.Y. Lum, "Automated image tracking based on the CAMSHIFT algorithm with adaboost and target trajectory and size estimation," 11th IEEE International Conference on Control & Automation (ICCA), Taichung, Taiwan, (2014), pp. 918-923, doi: 10.1109/ICCA.2014.6871044.

[35] S. Huang and J. Hong, "Moving Object Tracking System Based On Camshift And Kalman Filter", International Conference on Consumer Electronics, Communications and Networks (CECNet), 16–18 April 2011, pp. 1423–1426.

[36] B. Yang, H. Zhou, and X. Wang, "Target Tracking using Predicted CamShift", 7th World Congress on Intelligent Control and Automation, WCICA 2008. 7th World Congress on 25–27 June 2008, pp. 8501– 8505.

[37] J. Xu and X. Zhang, "A Real Time Hand Detection System during Hand over Face Occlusion", International Journal of Multimedia and Ubiquitous Engineering Vol.10, No.8 (2015), pp.287-302 http://dx.doi.org/10.14257/ijmue.2015.10.8.29

[38] J. Tang and J. Zhang, "Face Tracking with Occlusion," 2009 International Conference on Measuring Technology and Mechatronics Automation, Zhangjiajie, China, 2009, pp. 465-468, doi: 10.1109/ICMTMA.2009.57.

[39] M. Liu, H. Chen, Z. Qiu and X. Ren, "Moving Target Location Method Based on Euclidean Distance and Camshift Algorithm," 2018 Eighth International Conference on Instrumentation & Measurement, Computer, Communication and Control (IMCCC), Harbin, China, 2018, pp. 558-563, doi: 10.1109/IMCCC.2018.00123

[40] C. Pahl, T. Y. Oon and E. Supriyanto, "Fuzzy system based face tracking for head movement control in progressive health care," 2015 20th International Conference on Methods and Models in Automation and Robotics (MMAR), Miedzyzdroje, Poland, 2015, pp. 880-885, doi: 10.1109/MMAR.2015.7283993.

[41] W. Mao, H.Y Xie, Z.Q Tan, Z.P Liu and M.X Liu, “High precision indoor positioning method based on visible light communication using improved Camshift tracking algorithm, Optics Communications,Volume 468,2020,125599,ISSN 0030-4018, https://doi.org/10.1016/j.optcom.2020.125599.

[42] S. Kim, H. Kim,"A new metric of absolute percentage error for intermittent demand forecasts", International Journal of Forecasting, Vol. 32, Issue 3 (2016) ,pp 669-679, ISSN 0169-2070, https://doi.org /10.1016/j.ijforecast.2015.12.003.