International Journal of Safety and Security Engineering
Vol. 13, No. 4, August, 2023, pp. 693-700
Journal homepage: http://iieta.org/journals/ijsse

Enhancing Safety and Security: Face Tracking and Detection in Dehazed Video Frames Using KLT and Viola-Jones Algorithms

Vijaya Kumar Gurrala[1*], Srinivas Talasila[1], Pulagari Madhuri[1], Sandela Nithish Varma[1], Lekkala Puneeth[1], Padmaraju Koppireddi[2]

[1] Department of Electronics and Communication Engineering, VNR Vignanajyothi Institute of Engineering and Technology, Hyderabad 500090, India
[2] Department of Electronics and Communication Engineering, JNTU Kakinada, Kakinada 533003, India

Corresponding Author Email: vijayakumar_g@vnrvjiet.in

https://doi.org/10.18280/ijsse.130411

Received: 4 April 2023
Revised: 13 July 2023
Accepted: 25 July 2023
Available online: 28 September 2023

Keywords:
video enhancement, tracking, dark channel prior (DCP), dehaze, face tracking, Kanade Lucas Tomasi (KLT), Viola-Jones, haze removal, dehazing techniques

ABSTRACT

In the context of safety and security, the ability to track and identify faces in hazy conditions presents a significant challenge. The deleterious effects of haze on video quality, such as the diminution of detail, reduction in contrast, distortion of color, and complications in depth estimation, impede effective facial recognition. Additionally, the complexity of live video tracking is exacerbated by factors such as occlusion, positional variations, and lighting changes. Despite these challenges, video sequences offer an abundance of information, surpassing static images in terms of potential data extraction. In this study, a dual approach strategy is employed to detect and track faces in hazy conditions. The Kanade-Lucas-Tomasi (KLT) algorithm, celebrated for its adept feature tracking capabilities, is deployed to execute face tracking. The effectiveness of this algorithm lies in its ability to accurately trace points across successive image frames, a crucial aspect of reliable face tracking. Concurrently, the Viola-Jones algorithm is utilized for face detection. The algorithm harnesses Haar-like features to efficiently discern faces in real-time, effectively overcoming the challenge of identifying faces within video frames. To further enhance the quality of the video, the dark channel prior (DCP) image dehazing technique is employed. This technique improves visibility by increasing contrast and color saturation, whilst concurrently identifying and eliminating air haze from the video frames.

1. INTRODUCTION

Hazy conditions, dependent on their intensity and duration, can significantly impede safety and security measures, particularly in areas where surveillance is paramount. Under such circumstances, the quality of images captured by security cameras can be severely compromised, complicating the identification of individuals or objects. This in turn presents a challenge to the automated video processing systems, such as those used in traffic surveillance, criminal justice, and other applications requiring detection, analysis, identification, and recognition. Video enhancement, a process aimed at elevating the visual quality of a video or providing a superior transform representation for future automated video processing, proves indispensable in this context. The application of image and video enhancement techniques can elevate the quality of images and videos across varied fields, including medical imaging, satellite and aerial photography, and real-world photography, particularly those afflicted by low contrast and noise. By boosting contrast and eliminating noise, the overall video quality can be significantly improved. The initial steps in tracking and recognition processes involve the detection and localization of human faces in videos, a task that is often impeded by various factors such as poor lighting, positional fluctuation, occlusions, low resolution images, and unfavorable atmospheric conditions such as fog. Despite these challenges, video enhancement continues to be a critical procedure, not only elevating the visual quality of a video to increase viewer engagement, but also serving as a pivotal tool in safety and security measures. The field of video enhancement has witnessed significant advancements in recent years with the development of various methods and algorithms. By augmenting the clarity of video footage, it offers improved surveillance of high-risk areas, aiding in potential security concerns identification. Furthermore, investigators are enabled to interpret incidents more accurately through the review of enhanced security camera footage, a tool that could assist in crime resolution and future prevention. In the presented study, the dark channel prior (DCP) image dehazing technique is employed to remove atmospheric haze from the video frames, thereby improving image clarity. Figure 1 presents a comparison of a hazy and a dehazed image, illustrating the effectiveness of the DCP technique in enhancing video quality.

The dark channel prior (DCP), a dehazing technique, is employed as a pivotal tool in this work to enhance the video's clarity, color saturation, and contrast. An integral component of this video enhancement process is facing detection, a task that necessitates the identification and tracking of individuals' faces within the video. The Kanade-Lucas-Tomasi (KLT) algorithm, a feature-based tracking method, is utilized to monitor the evolution of specific features across consecutive video frames. Concurrently, the Viola-Jones algorithm, a cascaded classifier, is deployed for efficient and instantaneous face recognition. Through the utilization of these facial detection technologies, security cameras are equipped to identify suspects either in real-time or retrospectively.

This enhancement can facilitate immediate detection and deterrence of unauthorized individuals in critical locations, assisting law enforcement in promptly identifying and apprehending criminals, and thus mitigating potential harm or loss. Additionally, the technology enables crowd monitoring, allowing for the identification of potential disruptors, a feature particularly beneficial in densely populated settings with heightened security concerns.

The primary objectives of this work are as follows:

- Employ the DCP dehazing technique to improve contrast visibility.
- Detect (using the Viola-Jones algorithm) and track (using the KLT algorithm) faces in video frames.
- Evaluate the performance of face detection and tracking both with and without the application of video enhancement.

In this research, a comprehensive analysis of video enhancement techniques utilizing the DCP, in conjunction with face detection via the KLT and Viola-Jones algorithms, is presented. The various steps involved in each of these techniques, and their collective impact on the visual quality of the video, are discussed. Experimental results are provided to demonstrate the efficacy of these techniques in enhancing video quality.

The remainder of the paper is structured as follows: Section 2 offers a review of related work in the fields of video enhancement, face detection, and tracking. Section 3 outlines the methodology adopted for video enhancement using the DCP, face detection through the Viola-Jones algorithm, and face tracking via the KLT algorithm. Section 4 presents the experimental results and analysis. Finally, Section 5 delineates potential areas for future work.

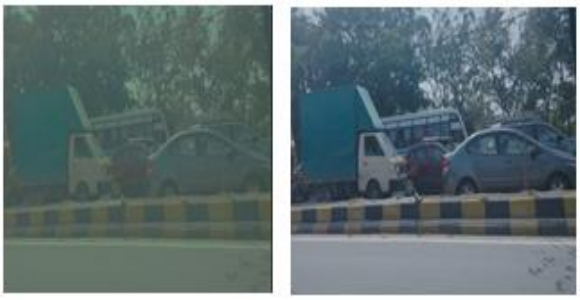


Figure 1. Hazy and dehazed image

2. LITERATURE REVIEW

The feature-based face detection approach, as discussed by Lee et al. [1], entails analyzing four fundamental facial features: the face as a whole, the nose, the lips, and the eyes. According to this approach, an image cannot be classified as a face unless all of these features are identified. However, a significant limitation associated with this method is its inability to account for a multitude of facial characteristics and its inefficacy in detecting faces in motion. On the other hand, the work of Narasimha and Batur [2] introduced a real-time High Dynamic Range (HDR) video camera, which, while beneficial, comes with its own set of challenges. These HDR video cameras tend to be more expensive than conventional cameras or the equipment required for the implementation of the dark channel prior, primarily due to the additional processing power and components necessary for HDR video capture. Moreover, the need for rapid processing and capture of video streams often results in these cameras offering lower resolutions than their traditional HDR counterparts. This discrepancy can lead to an inferior image quality when compared to standard cameras capable of recording video at higher resolutions. Furthermore, the dynamic range provided by real-time HDR video cameras may be less extensive than that offered by the dark channel prior. The inability of the camera's technology to handle scenes with high luminosity or darkness can result in overexposed or underexposed areas within the video. Additionally, owing to the intensive processing requirements, real-time HDR video cameras might not be compatible with all hardware or operating systems, which could limit the camera's adaptability and utility in certain scenarios. Conversely, the dark channel prior emerges as a simple yet potent HDR video processing technique that addresses these limitations. For instance, it can manage a wider dynamic range and generate high-quality video without the need for specialized hardware.

"A convolutional cascade neural network for face detection" by Li et al. [3], proposes a real-time face detection method that uses a convolutional neural network to detect faces in images and videos. Compared to the Viola-Jones algorithm with KLT tracker, the convolutional cascade neural network presented by Li et al. [3] is more complicated. Several layers of convolution and pooling are used in the network, which can be computationally expensive and demand additional processing power. For the convolutional cascade neural network to recognize faces with high accuracy, a lot of training data must be collected. Obtaining this can be difficult, and labelling training images may take a lot of human labour. While the Viola-Jones algorithm with KLT tracker has shown to be more resilient in handling occlusion and pose fluctuation, the convolutional cascade neural network has demonstrated great accuracy in recognizing faces in pictures and videos, faces that are partially obscured or at sharp angles may be difficult for the network to recognize. However, compared to the convolutional cascade neural network, the Viola-Jones algorithm with KLT tracker is a more straightforward method that is better able to manage occlusion and posture variation.

The study by Li et al. [3] introduces a real-time face detection approach that employs a convolutional cascade neural network to identify faces in images and videos. This approach is significantly more complex compared to the Viola-Jones algorithm in conjunction with the KLT tracker. The convolutional cascade neural network incorporates several layers of convolution and pooling, which can be computationally expensive and require substantial processing power. Furthermore, to achieve high accuracy in face recognition, an extensive collection of training data is necessitated. The collection and labeling of such data could entail considerable human effort. Although the network has demonstrated impressive accuracy in recognizing faces in images and videos, it may struggle to identify faces that are partially obscured or presented at sharp angles. In contrast, the Viola-Jones algorithm with the KLT tracker, though simpler, exhibits greater resilience in handling occlusions and pose variations. In a different vein, the method titled "Real-Time Temporally Coherent Local HDR Tone Mapping" by Croci et al. [4] necessitates more intricate processing requirements compared to the dark channel prior. This process includes steps such as tone mapping, contrast enhancement, and local image statistics estimation, which collectively contribute to increased computational demands for real-time implementation. While this approach provides benefits such as color enhancement, it may lead to inaccurate color reproduction and artifacts like color shifts, potentially compromising the overall image quality. Moreover, its capability to handle scenes with extreme brightness or darkness is limited due to its focus on local tone mapping. Conversely, the dark channel prior, although it might not excel in preserving details within bright areas of the image, provides a simpler HDR tone mapping technique that can manage a broader dynamic range and achieve more accurate color reproduction.

A subsequent study by Zhang et al. [5] proposes a real-time face detection method that uses a cascade of deep convolutional neural networks to detect and align faces in unconstrained environments. This approach is more sophisticated compared to the Viola-Jones algorithm with the KLT tracker, involving multiple layers of convolution and pooling which can be computationally expensive and demand considerable processing power. To detect and align faces with high accuracy, this method necessitates a large volume of training data. The collection and labeling of such data can be challenging and labor-intensive. Despite exhibiting impressive accuracy in recognizing and aligning faces in images and videos, the multitask cascaded convolutional networks might not be as resilient to occlusion and pose changes as the Viola-Jones approach combined with the KLT tracker.

"Real-time Video Super-Resolution with Spatio-temporal Networks and Motion Compensation" by Caballero et al. [6], which proposes a real-time video super-resolution method that uses spatio-temporal networks and motion compensation to enhance the resolution of low-resolution videos. The multi-step process known as "Real-time Video Super-Resolution with Spatio-temporal Networks and Motion Compensation" includes feature extraction, motion estimation, and spatio-temporal filtering. The dark channel prior, on the other hand, is a more straightforward technique that entails computing the dark channel prior and using a straightforward atmospheric scattering model. Under difficult lighting settings or for videos with complicated motion, it may not always perform well and it is expensive. In the realm of video super-resolution, Caballero et al. [6] proposed a method that utilizes spatio-temporal networks and motion compensation to enhance the resolution of low-resolution videos in real-time. This multi-step procedure, termed "Real-time Video Super-Resolution with Spatio-temporal Networks and Motion Compensation," encompasses feature extraction, motion estimation, and spatio-temporal filtering. In contrast, the dark channel prior is a simpler technique which involves the computation of the dark channel prior and the employment of a straightforward atmospheric scattering model. Despite its simplicity, it may not always deliver optimal performance under challenging lighting conditions or during the processing of videos with complex motion. Furthermore, it can be computationally expensive. In a similar vein, Ren et al. [7] introduced "Faster R-CNN," a more complex approach towards object detection, including face detection. This method involves constructing a deep convolutional neural network (CNN) for object detection and region suggestion, necessitating substantial processing resources for both training and inference. On the other hand, the Viola-Jones algorithm coupled with the KLT tracker presents a less complicated methodology, utilizing a set of predefined Haar-like features and a straightforward tracking procedure. Training a CNN for object detection and region proposal requires a large and diverse collection of images with annotated objects, which can be both time-consuming and costly to gather and label. Conversely, the Viola-Jones algorithm employs a smaller dataset of positive and negative samples for training. While "Faster R-CNN" can detect objects with high precision, it may not always be as swift or as computationally efficient as the Viola-Jones algorithm with the KLT tracker. The Viola-Jones approach with the KLT tracker, being faster and less computationally demanding, can be employed for a variety of object detection tasks, including face detection. Lastly, Zarkasi et al. [8] proposed a well-known technique for finding a matching region in an image or video frame for face detection, known as template matching. This method compares a template image of a recognized face with the target image or video frame. Despite its potential utility in certain situations, template matching presents a series of limitations. These include sensitivity to changes in lighting, pose, and scale, difficulties with occlusions, and the challenge of handling partial views of faces. Moreover, template matching can be computationally expensive when processing large datasets or real-time video.

The emergence and subsequent evolution of neural network-based methods for face detection have demonstrated promising outcomes. Bhandiwad et al. [9] constitute a prime example of such advanced methodologies. Despite the impressive results, these techniques do present notable disadvantages when juxtaposed with traditional methods like the Viola-Jones algorithm and the KLT tracker. Firstly, the complexity of neural network-based methods surpasses that of traditional methods. They necessitate substantial computational resources for both training and inference. Secondly, these neural networks demand extensive training data to effectively learn face detection, a requirement that can pose challenges in certain scenarios. Lastly, the interpretability of neural networks is often obfuscated, rendering it difficult to deduce how these 'black boxes' arrive at their decisions. Traditional methods like Viola-Jones and KLT, conversely, are founded on more interpretable mathematical models.

In the realm of face recognition algorithms, Zhao et al. [10] propose a method that hinges on optimal feature selection. This approach, however, requires a copious amount of training data to function optimally. The collection and labelling of such data can be both time-consuming and costly, representing a potential drawback when compared to the Viola-Jones approach with the KLT tracker. This traditional technique is simpler and more widespread, requiring less training data. Choudhary et al. [11] employ various techniques, such as deep learning, to eliminate haze from images. Despite its effectiveness, this method invites certain limitations. These include the necessity for a noise-free training dataset, the need for frequent alterations to the training dataset which slows down processing, and a dependency on several other factors such as the loss function's reliance on optimization. The dark channel prior (DCP), on the other hand, was expressly designed to enhance images and videos affected by atmospheric haze or fog. It operates by estimating the haze content in the image using the dark channel before eradicating it to improve visibility and contrast. As a computationally efficient method, DCP can be employed for real-time video processing and has been proven to deliver superior results.

Recent developments in video enhancement technology have heralded the introduction of polarization-based methods, which utilize polarized light to capture and process images or videos. An early application by Hu et al. [12] demonstrated its efficacy in enhancing underwater videos by reducing backscatter and augmenting image contrast. Despite its advantages, polarization-based enhancement introduces notable limitations. These include the requirement for specialized equipment like polarizing filters and cameras capable of recording polarized light. This necessitates implementation challenges and costs, particularly in practical scenarios. Furthermore, this method may not be universally applicable, such as in low-light conditions or when the subject lacks polarization.

3. METHODOLOGY

The proposed methodology can be implemented in five steps as shown in Figure 2. The first step is to take the input video that needs to be processed the second step is to apply the dark channel prior-based video enhancement algorithm to the input video. This involves calculating the dark channel prior for each frame of the video, estimating the transmission map and scene radiance, and applying temporal filtering to the estimated radiance. The output of this step is an enhanced video with improved brightness, contrast, and color fidelity. The third step is to perform face detection on the enhanced video using the Viola-Jones algorithm. This involves applying a set of pre-defined Haar-like features to detect faces in the video frames. The output of this step is a set of bounding boxes that indicate the location of each detected face. The fourth step is to track the detected faces using the KLT algorithm. This involves selecting a set of feature points within the bounding box of each detected face, and tracking those points across subsequent frames of the video using the KLT algorithm. The output of this step is a set of tracked feature points that represent the motion of each detected face across the video frame. The final step is to generate an output video that shows the original input video with the detected faces highlighted and tracked. This can be done by drawing bounding boxes around the tracked feature points for each detected face in each frame of the video.

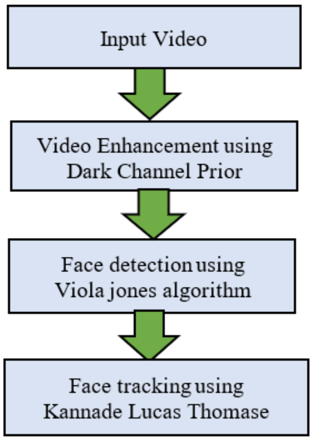

Input Video
Video Enhancement using Dark Channel Prior
Face detection using Viola jones algorithm
Face tracking using Kannade Lucas Thomase

Figure 2. Flow chart of the methodology

3.1 Dark channel prior

The proposed methodology for video enhancement using dark channel prior is an effective method for improving photos since it lessens the effects of haze and boosts contrast. It is a powerful algorithm that can handle a variety of scenarios with dynamic lighting and moving objects such as lighting setups, and camera settings is the dark channel prior. When used with video, it can offer a number of benefits including adaptability, robustness, consistency, speed, and increased visual quality. As a video is essentially a series of images, adding the dark channel before each frame guarantees that the video has a unified appearance and feel.

The basic premise is that there are some regions in a hazy or foggy image that have extremely low intensity values in comparison to the rest of the image. These regions are typically in the shadows. The "dark channel" of the image refers to these regions. Using the dark channel, the dark channel prior algorithm determines the amount of haze present in an image. It is predicated that the minimum intensity value in the dark channel is exactly related to the degree of haze in an image. The programme can produce a crisper image by evaluating how much haze is present in the image and then removing it.

The dark channel of the image is initially calculated in order to operate the dark channel prior method. Finally, using the image's brightest pixels, it calculates the amount of atmospheric light present in the image. It creates a transmission map that depicts the level of haze in the image using these variables. The image is then made clearer by using the transmission map to eliminate haze from it [13, 14]. Image dehazing, underwater photography, and remote sensing are just a few of the areas where dark channel prior has been proven to be successful. It is a well-liked algorithm for real-time applications since it is comparatively easy to use and computationally effective.

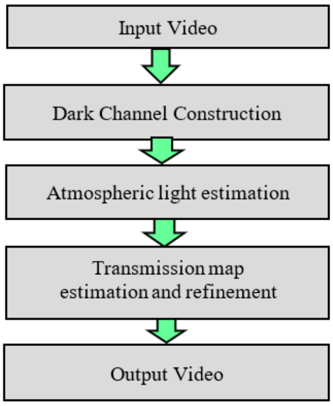

Input Video
Dark Channel Construction
Atmospheric light estimation
Transmission map estimation and refinement
Output Video

Figure 3. Block diagram of the dark channel prior

A step by step flow of video dehazing using dark channel prior is depicted in Figure 3. The video that needs to be dehazed must first be entered. Any format that the algorithm can read can be used for this video. The dark channel should be calculated for each video frame. Hence, for each frame of the video, a dark channel image will be an outcome. Calculate the atmospheric light for each frame after receiving the dark channel photos. The pixel with the greatest intensity value among the top 0.1% of pixels in the dark channel is used to calculate atmospheric light. The brightest and least haze affected pixel is presumed to be this one. Calculate the transmission map for each frame using the expected atmospheric light and dark channel. The equation 1 is used to compute the transmission map:

$$t(x) = 1 - w * min\_c\{I(x)/A\_c\} \tag{1}$$

where, $I(x)$ is the original image, $t(x)$ is the transmission map, w is a weighting factor (usually set to 0.95), and A_c is the atmospheric light in channel c.

The transmission map calculated in the earlier stage could have omissions and mistakes. Apply a guided filter on the transmission map, using the original image as the guidance image, to improve it. Lastly, to clear the haze from the original image, by utilising the transmission map and atmospheric light using Eq. (2):

$$J(x) = \frac{I(x) - A}{max\{t(x), t_{min}\}} + A \tag{2}$$

where, A is the ambient light, $J(x)$ is the dehazed image, and t_{min} is a tiny constant that is commonly set to 0.1 to prevent division by zero. The output of the dehazed video is the last stage.

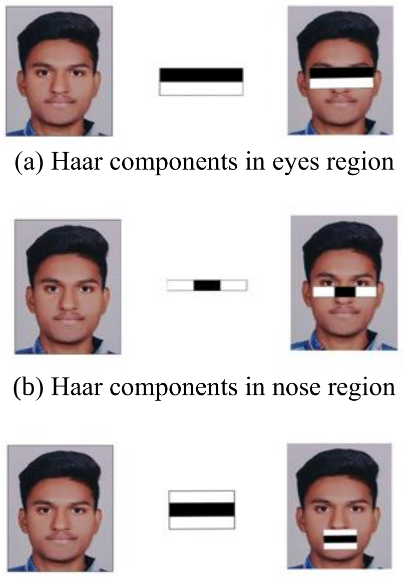
3.2 Viola jones-face detection

A quick and easy algorithm that can instantly identify faces is the Viola-Jones algorithm. To swiftly recognise the presence of a face in an image, it combines Haar-like features and an Adaboost classifier. Even in the presence of occlusion, partial obstructions, and variations in facial position, the Viola-Jones algorithm can identify faces. It has been demonstrated that the Viola-Jones algorithm detects faces [15] with a high degree of accuracy. It can accurately and reliably recognise faces, which is crucial for many applications. The Viola-Jones technique is appropriate for real-time applications on hardware with constrained processing capacity due to its comparatively low computing requirements. A human can easily recognise any face in a photograph or image, but a computer or robot will always require input and pressure. Faces cannot move sideways, thus for this purpose, Viola-Jones requires a strong front view against the camera [16].

Haar features are straightforward rectangular features that are the sum of pixels from different places inside the rectangle. This rectangle has the ability to scale the image and may be placed anywhere in the frame. 2-rectangle feature is the name of this modified feature set. Each feature type can reveal the presence or absence of specific details in the frame, like edges or texture changes.

To determine the face features, these haar features are used. For instance, in Figure 4, the black coloured portion indicates the presence of a nose, which is situated in the middle of the face. This part is utilised to detect this characteristic. When the white part is designated as -1 and the black part as +1. By deducting the total of pixels under the white rectangle from the total of pixels under the black rectangle, the outcome is determined. For specific features, a threshold is initially set. Calculate the average total of each black and white. Next, a threshold check is done on the difference. The value is detected as a relevant feature if it exceeds or equals the threshold.

To add all the pixels in a specific box to its left and upper ones, utilise the integral image component. It is necessary to determine the area's four corner values. This prevents the region's pixels from being added together. The sole purpose of this integral image conversion procedure is to accelerate the calculation of pixels. The formula for calculating the total number of pixels in component D of the above Figure 4 is (1+4)-(2+3), or [A+(A+B+C+D)] - [(A+B+A+C)], which results in D is depicted in Figure 5.



(a) Haar components in eyes region

(b) Haar components in nose region

(c) Haar components in mouth region

Figure 4. Haar feature components

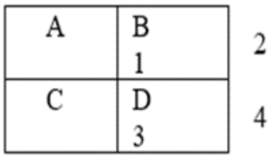


Figure 5. Integral image calculation

Following with the face features determination, the process used for identifying important and unimportant traits is Adaboost. The main principle behind AdaBoost for face identification is that it iteratively concentrates on difficult-to-classify samples and accentuates them during the training process. The method improves its accuracy over iterations by integrating numerous weak classifiers into a strong classifier. Each facial image is provided to the system with the same start weights as the others, with yn=1 for facial and yn=0 for non-facial photos. The classifier's weights in the images have now been adjusted. All of the images are used to train a classifier using a single feature, and the error is calculated. If a face feature is found, the mistake is 0, otherwise it is 1. The weights are updated and the lesser mistakes are picked. In the strong feature, a feature with little inaccuracy is therefore given more weight. Last but not least, a strong classifier exists when some of the weighted features are greater than 50% of the total weights. Non-faces are eliminated in order to shorten the computation time.

Cascading phase is added to the procedure shown in Figure 6 is to expedite it and produce accurate results. This process is broken down into multiple steps, each of which contains a powerful classifier. Each feature is divided into several levels. By moving a window over a frame, it can identify faces in the picture. After performing the Adaboost training, the fast way to check if the window contains a facial feature is to cascade the classifiers. The first classifier is the highest weight found than compared to the earlier ones. If the first feature is approved then it moves on for the second classifier until all of the features are approved, then a face is detected.

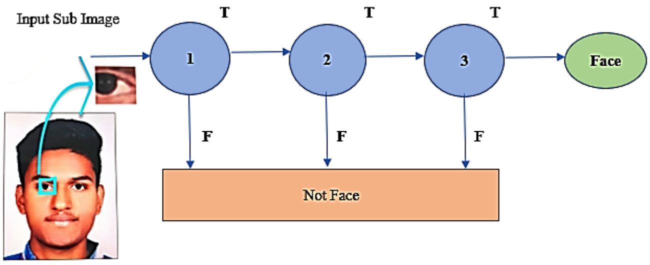


Figure 6. Cascading process

3.3 Kannade Lucas Thomase-face tracking

For face tracking and alignment, the Kanade-Lucas-Tomasi (KLT) tracker is frequently used in conjunction with the Viola-Jones algorithm. Face tracking across frames is made possible by this combination of KLT and Viola-Jones, finds sophisticated applications like facial tracking and recognition. KLT strategy depends on the optical movement of points from frame to frame (optical flow) to enable access to component abstraction since KLT is so much faster than other approaches and old ways are so much more expensive. By calculating the optical flow of a specified area over time, this technique is mostly used to continually monitor faces in videos that are being streamed live or that have already been captured. It is possible to follow a variety of diverse locations across time using this technique. By employing the KLT, the faces are tracked in a straight line from frame to frame.

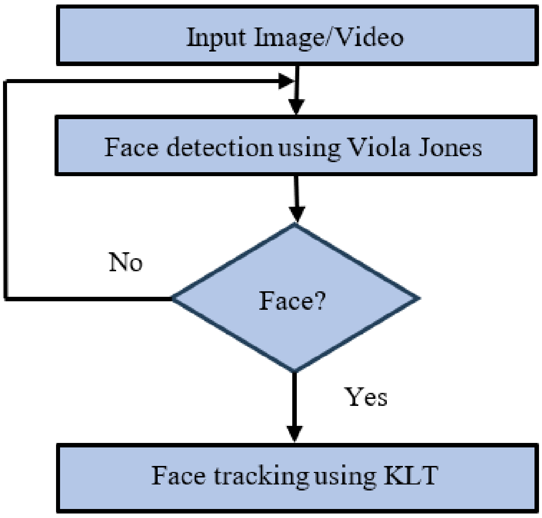


Figure 7. Flow chart for face detection

The KLT method is presented in Figure 7. It counts the constituents, or points, in one frame, locates those identical points in another frame, and then calculates the distance between the points in the two frames. Throughout this ongoing process, the KLT algorithm keeps track of the points until the execution is finished. This process is relatively straightforward in comparison to the prior techniques. Thus, locating the targeted areas and evaluating movement constitute the essential duty. This method is first used to identify Harris corners in the first frame. The KLT approach then employs a tracker to observe the light flow around any corners or points. The mobility of the pixels is assessed throughout this process. The tracker can instantly identify the unique points or pixels that each moving object in a movie comprises. The KLT tracker, which can also measure the motion of each pixel, can track each Harries point. The Harries corners detection tracker is used to properly evaluate the corners of each frame and assist in precisely detecting the points every 10 to 15 frames is depicted in Figure 8.

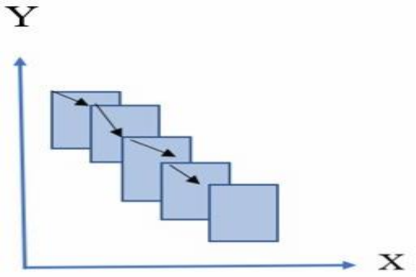


Figure 8. Movement of frames

4. RESULTS ANALYSIS

The use of the Dark channel before utilising the Viola Jones and Kannade Lucas Thomase algorithms helps in improving the quality of the video. The Figure 9 displays a face that was discovered using the Viola Jones method, and further features tracking was carried out using the KLT algorithm.

The feature points are located based on the kannade lucas thomase algorithm, and the below Figure 10 demonstrates multiple face detection using the viola jones algorithm as well as feature point detection. These feature points assist in tracking the face in the live video when the faces tilt or when faces enter and leave the video. Figure 11 helps in understanding that the video is first enhanced using dark channel prior (DCP), and then face detection and feature point extraction are carried out using both the Viola Jones and the KLT algorithms.

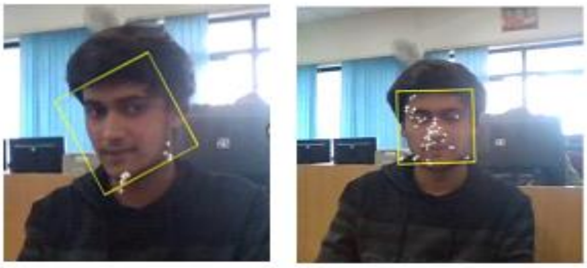


Figure 9. Frame considered from a video showing single face detection

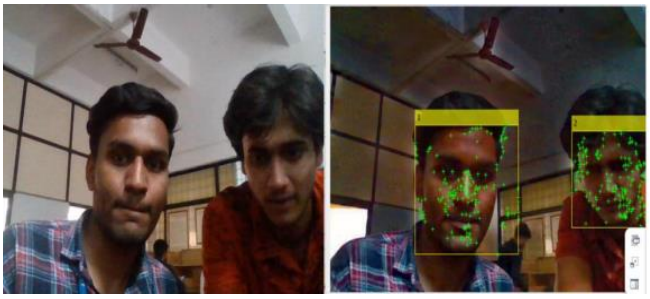


Figure 10. Frame considered from a video showing multiple face detection

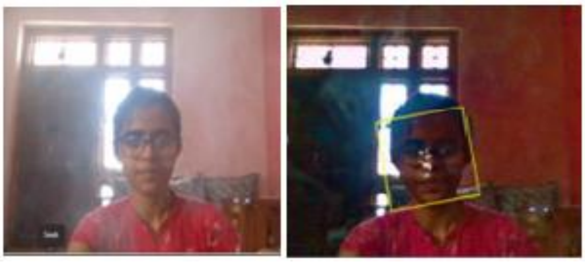


Figure 11. Frame considered from a video showing face detection in an unfavourable weather condition

Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) are two objective measures that can be used to assess the effeciency of the dark channel prior (DCP) algorithm. PSNR and SSIM are metrics used to quantitatively assess the quality of images, but they focus on different aspects. PSNR primarily measures the pixel-wise difference between the original and distorted images, while SSIM takes into account structural similarity and human perception. Higher PSNR numbers denote greater image quality. A value of 1 denotes perfect structural similarity between the original and denoised images. SSIM measures this structural similarity and SSIM lies between 0 and 1. A successful denoising method typically yields high PSNR and SSIM values. However, depending on the image content and the unique application needs, the precise values of these indicators that signify optimal performance can change.

In general, good performance for image denoising methods is regarded as an SSIM value closer to 1, and higher values denote greater image quality. The type and amount of noise present in the image, as well as the methods employed for image compression and processing, can all affect the actual range of PSNR values. It is determined using the *psnr* function as folllows:

$$psnr(noisyimage, denoisedimage, MAX\_I) \qquad (3)$$

where, *MAX_I* is the image's highest possible pixel value, noisy image is the image captured from the noisy video, and denoised image is the enhanced image that is captured from the enhanced video. The *PSNR* value is returned by the *psnr* function in decibels (dB).

The *ssim* function calculates the structural similarity index between two pictures. The *ssim* function syntax is as follows:

$$ssim(noisyimage, denoisedimage) \qquad (4)$$

The SSIM value, which ranges from 0 to 1, is what the ssim function returns when comparing two photos. Table 1 shows the PSNR and SSIM values of different set of input noisy and denoised images that are captured from a video.

The Table 2 represents the information about the number of faces detected in an image shown in Figure 12. It is enhanced using the DCP and further the face detection is done using the viola jones combined with the KLT and it shows that the faces detected are more when we combine the viola jones algorithm with the KLT rather using the viola jones algorithm alone. It also helps with security as it will be simpler to spot criminals in a clear video than compare to the video in unfavourable weather conditions.

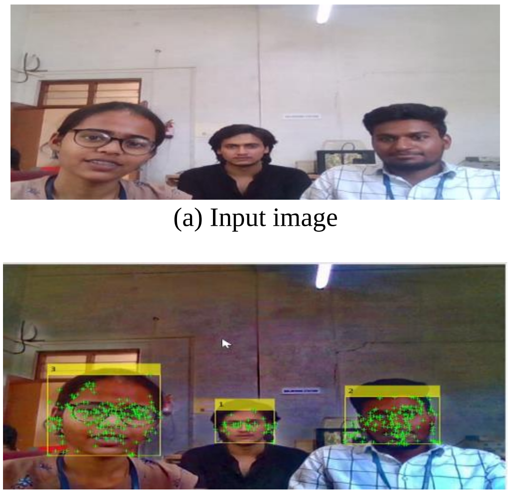
Overall, the Viola-Jones algorithm with KLT tracker is a straightforward, reliable, and effective technique for tracking faces that has been widely adopted in numerous applications. The approach given here for face detection and tracking minimises the calculation time producing results with great accuracy. KLT algorithm is utilised to track faces in video sequences, whereas Viola Jones is employed to identify facial features. It has been tested utilising a webcam for live video as well as video sequences.

Table 1. PSNR and SSIM values of set of noisy and denoised images

Sample Images	PSNR (in dB)	SSIM (in Decimal)
A frame consists single face	14.7350	0.8085
A frame consists multiple faces	14.7735	0.8812
A frame consists unfavorable weather condition	10.5936	0.6460

Table 2. Number of faces detected in an image

Detection View	Viola Jones	Viola with KLT
faces detected	2	3
faces not detected	1	0
false detection	1	0



(a) Input image

(b) Output image with detected faces

Figure 12. Faces detection from video frame

5. CONCLUSIONS

The problem of improving the visibility and detecting the faces is limited to images only but, in this work the detection and tracking is applied for the live stream video. The method for face detection and tracking described here reduces computing time, while generating highly accurate results. KLT algorithm is utilised to track faces in video sequences, whereas Viola Jones is employed to identify facial features. It has been tested utilising a webcam for live video as well as video sequences. In the near future, an item other than faces can be detected using these techniques such as leaf diseace detection [17] and biological deseace detections [18]. Future study will focus on the same area but track a specific face in a video clip. That is equivalent to ignoring all faces other than the one that is necessary.

REFERENCES

[1] Lee, T., Park, S.K., Park, M. (2005). A new facial

features and face detection method for human-robot interaction. Proceedings of the 2005 IEEE International Conference on Robotics and Automation, Barcelona, Spain, pp. 2063-2068. https://doi.org/10.1109/ROBOT.2005.1570417

[2] Narasimha, R., Batur, U. (2015). A real-time high dynamic range HD video camera. 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Boston, MA, USA, pp. 35-41. https://doi.org/10.1109/CVPRW.2015.7301364

[3] Li, H., Lin, Z., Shen, X., Brandt, J., Hua, G. (2015). A convolutional neural network cascade for face detection. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, pp. 5325-5334. https://doi.org/10.1109/CVPR.2015.7299170

[4] Croci, S., Aydın, T.O., Stefanoski, N., Gross, M., Smolic, A. (2016). Real-time temporally coherent local HDR tone mapping. 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, pp. 889-893. https://doi.org/10.1109/ICIP.2016.7532485

[5] Zhang, K., Zhang, Z., Li, Z., Qiao, Y. (2016). Joint face detection and alignment using multitask cascaded convolutional networks. IEEE Signal Processing Letters, 23(10): 1499-1503. https://doi.org/10.1109/LSP.2016.2603342

[6] Caballero, J., Ledig, C., Aitken, A., Acosta, A., Totz, J., Wang, Z., Shi, W. (2017). Real-time video super-resolution with spatio-temporal networks and motion compensation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4778-4787. https://doi.org/10.1109/CVPR.2017.304

[7] Ren, S., He, K., Girshick, R., Sun, J. (2017). Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(6): 1137-1149. https://doi.org/10.1109/tpami.2016.2577031

[8] Zarkasi, A., Nurmaini, S., Stiawan, D., Firdaus, Ubaya, H., Sanjaya, Y., Kunang, Y.N. (2018). Face movement detection using template matching. 2018 International Conference on Electrical Engineering and Computer Science (ICECOS), Pangkal, Indonesia, pp. 333-338. https://doi.org/10.1109/ICECOS.2018.8605215

[9] Bhandiwad, V., Tekwani, B. (2017). Face recognition and detection using neural networks. 2017 International Conference on Trends in Electronics and Informatics (ICEI), Tirunelveli, India, pp. 879-882. https://doi.org/10.1109/ICOEI.2017.8300832

[10] Zhao, K., Wang, D., Wang, Y. (2019). A face recognition algorithm based on optimal feature selection. Revue D'intelligence Artificielle, 33(2): 105-109. https://doi.org/10.18280/ria.330204

[11] Choudhary, R.R., Jisnu, K.K., Meena, G. (2020). Image DeHazing using deep learning techniques. Procedia Computer Science, 167: 1110-1119. https://doi.org/10.1016/j.procs.2020.03.413

[12] Hu, K., Weng, C., Zhang, Y., Jin, J., Xia, Q. (2022). An overview of underwater vision enhancement: from traditional methods to recent deep learning. Journal of Marine Science and Engineering, 10(2): 241. https://doi.org/10.3390/jmse10020241

[13] Naveena, M., HemanthaKumar, G., Prakasha, M., Nagabhushan, P. (2015). Partial face recognition by template matching. 2015 International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT), Mandya, India, pp. 319-323. https://doi.org/10.1109/ERECT.2015.7499034

[14] Berbar, M.A., Kelash, H.M., Kandeel, A.A. (2006). Faces and facial features detection in color images. Geometric Modeling and Imaging--New Trends (GMAI'06), London, UK, pp. 209-214. https://doi.org/10.1109/GMAI.2006.18

[15] Lang, L., Gu, W. (2009). Study of face detection algorithm for real-time face detection system. 2009 Second International Symposium on Electronic Commerce and Security, Nanchang, China, pp. 129-132. https://doi.org/10.1109/ISECS.2009.237

[16] Tarel, J.P., Hautière, N. (2009). Fast visibility restoration from a single color or gray level image. 2009 IEEE 12th International Conference on Computer Vision, Kyoto, Japan, pp. 2201-2208. https://doi.org/10.1109/ICCV.2009.5459251

[17] Talasila, S., Rawal, K., Sethi, G. (2023). Deep learning-based leaf region segmentation using high-resolution super HAD CCD and ISOCELL GW1 sensors. Journal of Sensors, 2023: 1085735. https://doi.org/10.1155/2023/1085735

[18] Gurrala, V., Yarlagadda, P., Koppireddi, P. (2021). Detection of sleep apnea based on the analysis of sleep stages data using single channel EEG. Traitement du Signal, 38(2): 431-436. https://doi.org/10.18280/ts.380221