MALAYSIAN JOURNAL OF
CYBERSECURITY AND APPLICATIONS

CSAM
CYBER SECURITY
ACADEMIA MALAYSIA

FORENSIC FACIAL IDENTIFICATION BASED ON FACE HALLUCINATION TECHNIQUE WITH SPARSE REPRESENTATION

Nazri Ahmad Zamani,[1,2*], Siti Norul Huda Sheikh Abdullah*[1], Khairul Akram Zainol Arifin[1], Md Jan Nordin[1], Tutut Herawan[3], Nazhatul Hafizah Kamarudin[1]

[1]Center For Cyber Security, Faculty of Information Science and Technology,
Universiti Kebangsaan Malaysia, 43600, Bangi, Malaysia
[2] CyberSecurity Malaysia, Cyber Axis Building, Cyberjaya, 63000, Malaysia
[3]Department of Information Systems, Faculty of Computer Science and Information Technology, Universiti Malaya,
50603 Kuala Lumpur, Malaysia

Corresponding Author: snhsabdullah@ukm.edu.my[1]*, nazri.az@cybersecurity.my[2*]

ABSTRACT

In video forensics, the low resolution of the facial information inside the video evidence is found to be the leading cause of the low performance of the facial identification system. Therefore, the super-resolution method is commonly used to recover low-resolution facial information inside a photo or a video to a higher resolution. However, in the current state, image resizing, especially super-resolution methods, cannot enhance the resolution of facial information with good quality at high magnification factors. This paper proposes a new forensic face identification based on the face hallucination technique with sparse representation. The proposed method, Sparse Resolution (SR), is a single-frame method that uses a representation of a signal with linear combinations of small elementary signals. These signals are then interpolated to synthesize low-resolution signals at a higher resolution. The signals are chosen via sparse coding from an over-complete dictionary with trained images. The active Appearance Model (AAM) and Support Vector Machine (SVM) were subsequently used to extract features and classify data. The experimental results test the SR face images on two datasets: (1) 14 individuals collected via CCTV surveillance Digital Video Recorder, and (2) the 2.5D partial images produced by a forensic facial identification system. The experiments show that the SR produced promising results. Also, the AAM-SVM facial matching results show that the SR images get higher matching performance than other state-of-the-art methods.

Keywords: Digital Forensic; Super resolution; Hallucination; Sparse coding.

1. INTRODUCTION

Ongoing development in electronics, optics, and sensors has influenced the extensive availability of monitoring systems and video-based surveillance. Applications (diversified from security to broadcasting) are moving the necessity forward to recognize individuals better from the number of surveillance videos [1]. This necessity imposes new difficulties on face recognition, which is already challenging. Cameras are generally at an optimum distance from the subject in a typical surveillance scenario [2]. Image resolution is a significant factor for many systems of 2D face recognition (FR), which affects the ability to detect essential elements in facial anatomy (e.g., lip, eyes, facial contours, lip corners) [2]. In the video exhibit for a human observer, a necessary forensic analysis is object enhancement to improve the object clarity. The clarified object is important because it is used to help investigate law enforcement agencies and is presented as evidence in court. Moreover, a 'probe' or an enhanced object is also used for other analyses like object identification, face recognition, and spectrogram. Challenges encountered in this field of forensics are mainly attributed to the exhibit's quality. For instance, surveillance videos are kept backed up in the CCTV system through the down-sampled resolution. With color noise, illumination problems, signal noise, and different sorts of blurring, the performance of the recording has always been found to be demoted [3]. Plenty of dead ends occurred in

the video forensic analysis due to these demotions. Thus, the court rejects or challenges forensic investigation reports because insufficient evidence is presented.

The signal-to-noise ratio (SNR) and mean-squared error (MSE) between the reconstructed SR image and the initial high-resolution (HR) image have been optimized in many super-resolution (SR) attempts. [4]. Nevertheless, these SR attempts may not perform satisfactorily in face recognition because most FR systems depend on the abilities of major features of facial identification, generally captured by high-frequency contents. Since the high-fidelity reconstruction of low-frequency contents may dominate images, having a higher SNR does not lead to a greater recognition rate. To increase low-resolution images and transform photos into sketches or sketches into photos for subsequent usage, face hallucination (FH) techniques can be utilized. It is widely recognized that FH can generate information or imagery from the input source face image, but with different modalities (style, imaging modes, or resolution) [5]. To reduce all these challenges and enhance the forensic analysis work, a new face identification method based on the face hallucination technique with sparse representation is introduced in this paper.

Despite advancements in face recognition technologies, accurately identifying individuals from low-resolution surveillance footage remains a critical challenge in forensic investigations. Conventional super-resolution techniques often fail to recover discriminative facial features necessary for identification, especially under extreme resolution degradation. Therefore, there is a pressing need for robust image enhancement strategies that can reliably reconstruct high-quality facial images from severely degraded inputs.

This paper has been organized into five sections. Section two presents the related works of single-frame super-resolution and multi-frame super-resolution. Next, we discuss the proposed face identification technique. We test and verify our work compared to the state-of-the-art method in Section 4. Finally, the conclusions from this work are summarized in the last section.

2. RELATED WORKS

As seen in the reference image, the definition of super-resolution is based on using an individual image or multiple images of the same scene or object to produce a higher-resolution image through minimal aliasing. The related works for the multi-frame, single-frame super-resolution and deep learning-based methods in SR are reviewed in the sub-sections below.

2.1 Multi-Frame Super Resolution

The technique of producing an image of high resolution (HR) from several images of low resolution (LR), or video frames, from the original scene is known as multi-frame image super-resolution (SR). The idea of image super-resolution was first introduced by [6]. The authors proposed a frequency domain formulation for reconstructing a band-restricted image based on the shift and aliasing properties of continuous and discrete Fourier transforms. This algorithm was given a noisy data extension by [7], resulting in a weighted least squares algorithm for computing the high-resolution estimate. Again, [8] used the Tkhonov regularization to solve an inconsistent series of linear equations by considering different amounts of blur for each low-resolution picture. The connection between the LR and HR images is illustrated in the frequency domain, a remarkable benefit of the frequency domain approaches discussed above. The observation model, on the other hand, is limited to global translational motion and linear space invariant blur.

The most intuitive method for SR reconstruction is based on a non-uniform interpolation approach [3], [4]. Using the generalized multi-channel sampling theorem, [15] performed no uniform approximation of an aggregate of temporally shifted LR images to obtain a higher resolution image. The [15] authors proposed a weighted nearest neighbor interpolation method. They developed a real-time infrared image registration technique and SR reconstruction using a gradient-based registration algorithm for estimating shifts between acquired frames. However, these approaches are limited to a global translational displacement between the measured image and an LSI blur and homogeneous additive noise in one of two ways. A multiple-frame Super-Resolution technique was demonstrated by [3], which involved merging a series of video frames of a subject to generate a super-resolved frame of improved resolution and clarity. They used the Projection onto Convex Sets (POCS) process for the super-resolution. Keren was used to estimate the

shift and rotational parameters of the frames for the POCS. On the other hand, multiple-frame SR is not compatible with the enhancement analysis. Furthermore, the multiple frames super-resolution is hard to adapt due to varying video frame rates and duplicated frames.

Numerous real-world forensic cases have highlighted the limitations of current image enhancement methods. For instance, in several high-profile criminal investigations, including urban surveillance-based identification and ATM fraud, authorities were unable to positively identify suspects due to the poor quality of available footage. These failures underscore the critical importance of developing more advanced face hallucination methods to reconstruct clear, high-resolution facial images suitable for recognition and legal use.

2.2 Single-frame super resolutions

Single-frame and real-time super-resolution have been developed in recent years. Patch-based upsampling, example-based super-resolution, and texture hallucination have all been used in recent work on single-image super-resolution [9], [10], [11]. In [11], the authors present a super-resolution algorithm of a single image built on both the wavelet and spatial domains and exploits both. The iterative algorithm employs back projection to reduce reconstruction error. A wavelet-based de-noising approach is also introduced to minimize noise. The drawback of a single-frame SR in comparison with multiple-frame SR is that it lacks the variety of spatial-frequency pass-bands from the camera to form a good high-resolution image. On the other hand, multiple-image super-resolution has pass-bands that can be acquired from one image to another in the sequence. These acquired pass-bands are then superimposed to synthesize a high-resolution image. With sparse representation, these drawbacks can be solved by training samples of good images as a reference or dictionary to supplement the lost spatial frequency.

2.3 Super-Resolution Enhancement in Forensic Facial Identification

Wheeler has discussed the idea of Super-Resolution in enhancing CCTV surveillance video, especially for forensics purposes. [3], [12] In these papers, Wheeler [3], [12] discussed the application of Super-Resolution in restoring facial images for face recognition in forensic analysis. He also discovered the Wiener Deblurring filtering technique by considering possible Point Spread Functions (PSF) to reverse the video degradation's noise and blurring aspect. Figure 1 shows the process of Super-Resolution enhancement in 2.5D forensic facial identification initiated and modified from [3], [12] studies for ‘Tepuk Bahu Case’ in Malaysia.

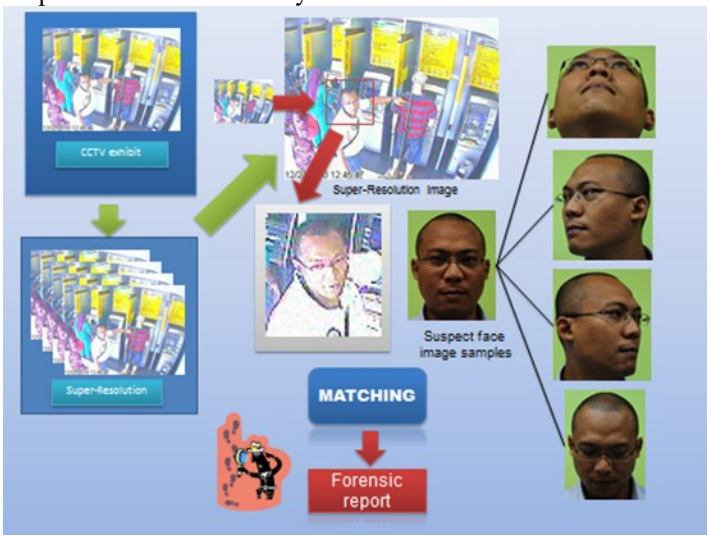

Fig. 1: The Super-Resolution enhancement in 2.5D forensic facial identification initiated and modified from [3], [12] for ‘Tepuk Bahu Case’ in Malaysia.

2.4 Deep Learning-Based Versus Frequency-Based SR Method in Facial Recognition

Recent Super-Resolution (SR) advancements for facial recognition have significantly improved image enhancement techniques. Traditional methods such as SRCNN [1] [13]and VDSR[2] [14] utilize deep convolutional networks to reconstruct facial details from low-resolution images, while generative adversarial networks (GANs), exemplified by SRGAN, generate realistic textures but face training instability [15] Unsupervised approaches like CycleGAN enable SR without paired datasets, broadening applicability but often leading to structural inconsistencies in facial reconstruction [16]. Recurrent neural networks (RNNs), such as RBPN, leverage temporal dependencies for video SR, enhancing frame continuity but requiring substantial computational resources [17]. Attention-based models like SAN focus on critical image regions, refining facial features with higher accuracy at the cost of increased computational demand [18]. Wavelet-based approaches, including WRAN, incorporate spatial and frequency-domain information to improve detail preservation [19]. More recently, transformer-driven SR models have gained traction for efficiently capturing global contextual features [20] diffusion-based SR techniques generate a high-resolution image through probabilistic modeling [21]. These advancements collectively enhance perceptual quality, computational efficiency, and adaptability in real-world facial recognition applications, as summarized in **Table 1**.

Table 1: Key Benefits and Limitations of Super Resolution for facial recognition applications.

Method	Dataset	Approach	Complexity	Efficiency	Key Benefits	Limitations
SRCNN Dong et al. [13]	Set5, Set14, and T91	Supervised CNN	Moderate	High	Simple architecture, effective for basic SR tasks	Limited feature extraction, struggles with complex textures
VDSR Kim et al. [22]	Set5, Set14, and BSD100	Deep CNN	High	Moderate	A deeper network improves detail recovery.	High computational cost
SRGAN Ledig et al. [3]	COCO Set5 and Set14	GAN-based	Very High	Low	Generates realistic textures, enhances perceptual quality	Training instability, potential artifacts
CycleGAN Zhu et al. [16]	Unpaired datasets image-to-image translation	Unsupervised GAN	High	Moderate	No need for paired datasets, flexible domain adaptation	Structural inconsistencies in image reconstruction
RBPN Haris et al. [17]	Vimeo-90K and Vid4 Vid4	Recurrent Neural Network	Very High	Moderate	Leverages temporal dependencies for video SR	Requires significant computational resources
SAN Dai et al. [18]	Set5, Set14, and Urban100	Attention-based	High	High	Focuses on important image regions, improves feature refinement	Computational overhead
WRAN Li et al. [19]	BSDS100, Set5, and Set14	Wavelet-based	Moderate	High	Captures spatial and frequency information	Requires careful tuning of transformation parameters.
Transformer -Based SR Dutta et al. [20]	DIV2K and Urban100	Deep Learning Transformer	High	High	Improves global feature extraction, enhances facial details	Requires large-scale datasets for training
Diffusion-Based SR Moser et al. [21]	DIV2K	Probabilistic Model	Very High	Moderate	Generates high-resolution images with superior perceptual quality	Computationally expensive, slow inference

While existing methods provide valuable insights into super-resolution strategies, they often fall short when applied to forensic contexts due to their inability to recover fine-grained identity-specific features from highly degraded inputs.

This shortcoming motivates the present study, which introduces a Sparse Resolution (SR) framework leveraging sparse coding and facial priors to achieve superior reconstruction quality. By focusing on the structural and statistical patterns unique to human faces, the proposed method addresses these limitations head-on and provides a more reliable tool for forensic image enhancement.

In summary, prior research has predominantly focused on general-purpose SR techniques without tailoring reconstruction processes to facial features critical for identification. Additionally, forensic-specific evaluations are lacking under realistic constraints such as low frame rates, variable lighting, and extreme poses. Furthermore, limited attention has been given to scenarios involving partial or occluded faces—common occurrences in surveillance footage. These gaps reveal a need for specialized face hallucination models that can function reliably in forensic applications.

3. THE PROPOSED METHODOLOGY

The proposed Sparse Resolution (SR) framework sets itself apart from conventional SR techniques by integrating sparse coding with an over-complete dictionary trained specifically on facial features. Unlike existing interpolation-based or frequency-domain approaches, SR focuses on learning the intrinsic structures of human faces, enabling more accurate reconstructions under low-resolution and forensic constraints. This targeted learning mechanism makes the proposed approach uniquely suited for enhancing surveillance video frames where facial features are often distorted or incomplete.

In this section, the proposed methodology in this paper is classified into four phases: (i) Forensics facial Identification SOP, (ii) the proposed Sparse Resolution (SR) method, (iii) feature extraction using AAM modeling, (iv) identification phase. The four phases of the proposed methodology are discussed in Sections 3.1 until 3.4.

3.1 Forensics Facial Identification Standard Operating Procedure (SOP)

In forensics, facial recognition is a method that adheres to a set of specific standard operating procedures (SOPs) to produce a clear and succinct examination result for use in court. As a result, when storing, extracting, evaluating, and presenting information about digital media as possible evidence, an observer follows a technique to demonstrate impartiality. The Forensic Facial Identification SOP ensures the systematic handling of facial imagery throughout the enhancement and identification pipeline. It involves pre-processing, face detection, low-resolution input standardization, and SR-controlled reconstruction. These protocols maintain consistency in forensic practices and support the reproducibility and reliability of the proposed method across various investigative contexts. Figure 2 shows the forensics process, from obtaining the display to analyzing it and presenting the case to the court.

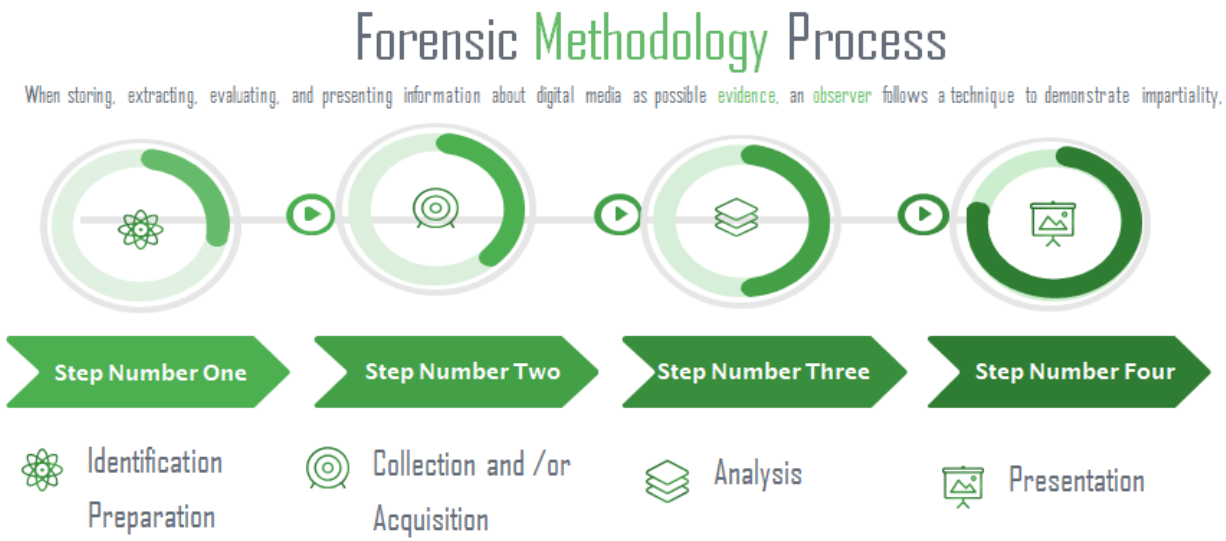

Fig. 2: The forensics methodology

The method of collecting firsthand information about the crime scene, the premise, the individual in charge, and the different types of CCTV systems is known as identification or planning. All details collected at this stage must be saved to create a plotline in court later. The next move is to collect or acquire digital media based on the circumstances and applicability. The digital media, or display, is analyzed in a closed environment to achieve the case target. Finally, the court is confronted with the results of the report. From the point it is taken to the point it is returned, the exhibit must be securely stored so that no or limited tampering occurs during the process.
There are three stages to the facial recognition process: 1) preparation of the probe, 2) enrollment of the population, and 3) process of facial identification. The probe image for the test is created in the first step by selecting video frames with the highest quality possible facial information. The most important criteria for the selection are:

1. The best frontal face posture: all facial features are visible. Since the recognition system enrolls full frontal faces, partial faces are not chosen for study. Regarding facial posture and orientation, there is a limited tolerance as long as all fiducial characteristics are present.
2. The size of the face in the exhibit video should be at or above 100 by 100 pixels in native resolution. Image resampling of not more than five enlargements is needed for resolutions lower than this, based on the resampling performance quality. Increased resampling of video frames can result in "hallucinations," in which facial data is demoted for recognition.
3. If the facial data in the video is still not precise, the quality of video frames should be sufficient to show all facial characteristics with minor enhancements.

Following the selection of the best frames with the best condition of facial detail, the process moved on to frame extraction and enhancement using a variety of algorithms for image processing. After that, images with face probes are clipped and used in the identification process.

Population Enrollment is the next step in the process. Since the recognition process is one-to-many, a collection of population face databases will be prepared for facial identification to function. A population of less than 1490 people is used for a data training sample. For the population, only the complete frontal face picture is used. A collection of suspect face images is included in the population. During enrollment, picture quality is once again a primary concern. Applying the probe image degradation factors to the population is best practice. As a result, a set of algorithms for image processing should be devised to fit the image quality of the population to that of the probe. In forensics, the enrollment step is close to that of biometrics, in which all images are subjected to feature extraction processes and storage. Our work requires the recognition of fiducial landmarks, which are then used to form multiple 2.5D Active Appearance Models (AAM). These facial instances will then be used in the matching process.

Eventually, probes will be compared with the enrolled population database during the facial identification process. The fiducial landmarks will be manually determined first, and then many 2.5D instances will be created from there. The search and match will be performed in stages, depending on what other instances are generated from the probes and enrollment. To the probe, the two outcomes will be the identity match and the False Match Error Rate (FMER) of that matched identity. The proposed 2.5D Facial Identification processes discussed are shown in Figure 3.

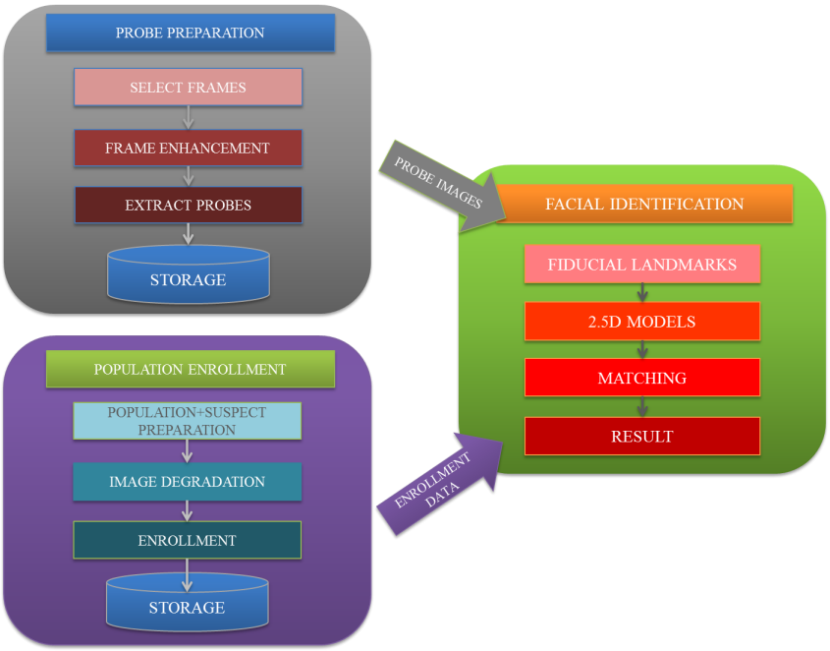

Fig. 3: The Detailed Proposed 2.5D Facial Identification Process initiated from Fig 1.

3.2 The proposed Sparse Resolution (SR) method

Sparse Resolution (SR) is based on the principle that any low-resolution facial image can be represented as a sparse linear combination of high-resolution facial patches from a learned dictionary. Instead of relying on pixel interpolation or frequency transformation, the method searches for the best match in a pre-trained dictionary, reconstructing high-resolution details by combining these sparse components. This data-driven approach enables the system to fill in missing facial details with high fidelity, especially when the input quality is poor.

3.2.1 Sparse Coding and Non-Negative Matrix Factorization

Non-Matrix with Non-Negative Values in sparse coding, factorization is used to learn a distributed part-based description. This necessitates using a dictionary, and the enhancement is carried out using a localized part-based description. Study [23] is cited in the following description of sparse representation methodology. Non-negative matrix factorization (NMF) aims to extract related information about these relevant parts to obtain an additive composition of these image descriptors. The formulation of NMF is given below:

$$\arg\min_{W,H} \|D - WH\|_2^2 \quad (1)$$

$$s.t. W \geq 0, H \geq 0,$$

Here, the data matrix is denoted by D ∈ $\Re^{n\times m}$, while W ∈ are $\Re^{n\times r}$ as a numerator, the basis matrix, and H ∈ $\Re^{r\times m}$ is called the coefficient matrix. The r bases is assumed as$(n \times m)/(n + m)$. From the above equation, the up-data rules are set as below:

$$H_{ij} \leftarrow H_{ij}\frac{(W^T D)_{ij}}{(W^T WH)_{ij}}, \quad (2)$$

$$W_{ti} \leftarrow W_{ti}\frac{(DH^T)_{ti}}{(WHH^T)_{ki}},$$

It is also important to remember that the higher resolution (Ih) and low resolution (Il) are obtained from smoothing and down-sampling, respectively. The degradation process from largest to smallest resolution can be expressed as Il = MIh, where M is the matrix that handles both blurring and down-sampling. The optimal Ih based on the Maximum A-Posteriori criterion of reconstructing the Super-Resolution is as follows:

$$I_h^* = arg \max_{I_h} p(I_l|I_h)p(I_h), \quad (3)$$

can be found with (3) from the basis matrix W in (2).

$$c^* = arg \min_c \|MW_c - I_l\|_2^2 + \lambda\rho(Wc) \quad (4)$$

$$s.t.\ c \geq 0,$$

Here, the high-pass filtering is denoted by Γ, and the high resolution is estimated by Wc*. However, the high frequency components are suppressed by the prior term in equation (5), causing over-smoothness in the solution picture:

$$c^* = arg \min_c \|MW_c - I_l\|_2^2 + \lambda\|\Gamma Wc\|_2 \quad (5)$$

$$s.t.\ c \geq 0,$$

Where Γ denotes the high-pass filtering, the high resolution here is approximated by Wc ∗. Nonetheless, the prior term in equation (5) suppresses the high frequency components, resulting in over-smoothness in the solution image. Using an over-complete dictionary that includes all the prototype signals learned, the signals will be represented as sparse linear combinations. The low-res patch y is derived as below:

$$y = Lx \doteq LD\alpha, \quad (6)$$

The high-resolution patch is represented by x, and the low-resolution patch, is represented by y whereas α is a nonzero entry vector of very few K signals and L ∈ $R^{k\times n}$ is the vector length with $k \leq n$. D is denoted as the over-complete dictionary that contains K the prototype signals. Also, x is defined as the high-res patch, i.e., representing image signals in a sparse linear combination. It is denoted as x = Dα

By solving the optimization problem in equation (5), we can extract a smooth, high-resolution probe Y from the subspace spanned by W:

$$Y = Wc^*, \quad (7)$$

Starting for each patch y of Y, 1 pixel is taken to overlap in each direction, starting from the upper-left corner. The optimization is then solved using the qualified dictionary $\widetilde{D}$ and the patch $\tilde{y}$, as shown in the following formula:

$$min\,\eta\|\alpha\|_1 + \frac{1}{2}\left\|\widetilde{D}\alpha - \tilde{y}\right\|_2^2, \quad (8)$$

Here, the balancing parameter of the sparsity is η that is also the fidelity to the estimation of y. $\widetilde{D}$ and $\tilde{y}$ are detailed as $\widetilde{D} = \begin{bmatrix} FD_l \\ \beta PD_h \end{bmatrix}$ and $\tilde{y} = \begin{bmatrix} Fy \\ \beta w \end{bmatrix}$. F is gradient filter that is selected as a feature extraction operator. In contrast, P is a matrix that extracts the overlapped region between a target patch and a previously reconstructed high-res image. The tradeoff here is between matching low-res input with a detected high-res patch consistent with its neighbors and

matching low-res input with a detected high-res patch compatible with its neighbors. Also, D_l and D_h are the low-res and high-res training dictionary representations, respectively.

Finally, patch $x = D_h \alpha^*$ generates the reconstruction of the high-res patch. In a high-resolution image of X, paste the patch x. The super-resolution of probe X* is the output. The diagram that shows the process flow of the algorithm is given in Figure 4:

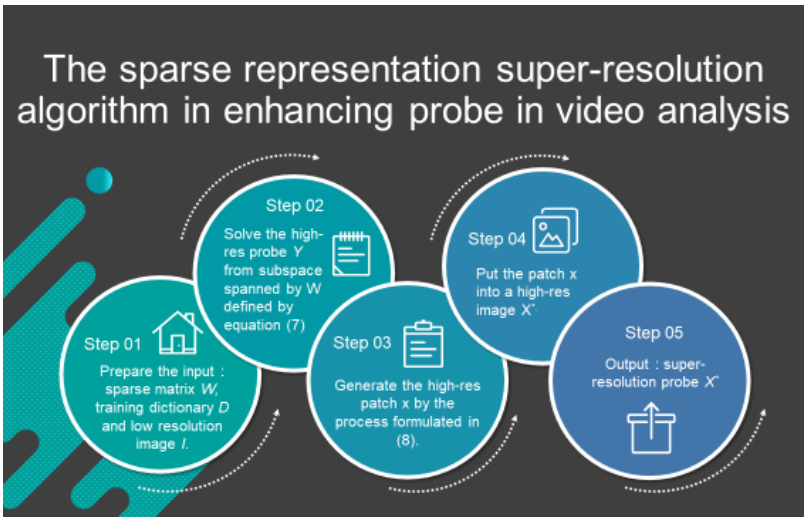

Fig. 4: The proposed process flow of the sparse representation super-resolution algorithm enhances the probe in video analysis.

3.3 Feature Extraction using AAM Modeling

The Active Appearance Model is a content-based approach that uses the locations of the eyes, the curve of the brows, the shape of the lips, the nose, and other facial features that can be found continuously across face images, as opposed to the comprehensive approach, which uses pixels intensities across the observed face region. The Active Appearance Model (AAM) is a broadening of the Active Shape Model (ASM) [24], [25] which uses all information in the probe image region rather than near modeled points or edges. ASM uses the Point Distribution Model (PDM), which is labeled on probe images, whereas AAM is more sensitive to changes in fiducial landmarks on probe images. AAM, unlike ASM, only uses shape constraints and investigates information about image structure near landmarks. Rather than updating the PDM through local point searches, which are then limited during the training phase by the PDM acting as a prior, the AAM model parameters are learned with respect to the appearance [24]. AAM is capable of modeling the fitting ones. AAM has the advantage of fitting the deformation of a 3D model from one view by using the 2.5D approach.

A probe picture from a CCTV surveillance video exhibit includes random facial details. Depending on the relationship between camera position and individual facial pose and orientation, it may be in any pose and orientation (which makes finding a true full frontal difficult). The key variable is facial detail. A 2D solution is not the best choice since the probe face image must be similar to the qualified face in the enrollment, according to the thumb rule of face recognition; AAM comes in handy in this situation. The like form variety of the face and its associated fiducial landmark points can be matched by AAM simultaneously. AAM is useful in obtaining certain landmarks that can be used in the latter method to a particular extent for the deformation, pose, and orientation of the facial details within the probe images.

In revising AAM, Cootes [24] clarified that the approach is based on statistical appearance models generated by combining a shape variation model. A face sample can be estimated by using the following expression.

$$x = \overline{x} + P_s b_s \quad (9)$$

Here, the mean shape (denoted as *Procrustes mean*) is $\bar{x}$. Also, P_s is denoted as a series of modes in orthogonal variation, and b_s is defined as a series of shape parameters. To remove variations in illumination, by employing a scale α, and an offset, β, the samples are normalized.

$$g = (g_{im} - \beta l)/\alpha \quad (10)$$

Values of α and β are selected to correspond to the normalized mean of vectors. The mean of the normalized data is $\bar{g}$, which has been resized and offset such that the number of elements is zero and the variance of each element is one. Values of α and β are needed to standardize g_{im} which are presented as,

$$\alpha = g_{im}.\overline{g}, \beta = (g_{im} \cot l)/K \quad (11)$$

where K is the number of elements in the vectors. In order to obtain a linear model, *Principal Component Analysis* (PCA) is applied.

$$g = \overline{g} + P_g b_g \quad (12)$$

Here, $\bar{g}$ refers to the mean normalized gray-level vector, whereas, P_g indicates a set of orthogonal modes of variation, and b_g denotes to a set of gray-level parameters. Due to the possibility of interactions between gray-level and shape differences, a second PCA is performed. A generated concatenated vector is represented for each sample as

$$b = \begin{bmatrix} W_s b_s \\ b_g \end{bmatrix} = \begin{bmatrix} W_s P_s^T (x - \overline{x}) \\ P_s^T (g - \overline{g}) \end{bmatrix} \quad (13)$$

Here, the unit's difference between the form and gray-level versions is explained by W_s, a diagonal matrix of weights for each shape parameter. The PCA application becomes,

$$b = Q_c, \quad (14)$$

where Q is the set of eigenvectors and $\boldsymbol{c}$ is the vector of appearance parameters controlling both the shape and gray-level of the model. The linear nature of vector c is then described as

$$x = P_s W_s Q_s c, \; g = \overline{g} + P_g Q_g c, \quad (15)$$

where, $Q = \begin{bmatrix} Q_s \\ Q_g \end{bmatrix}$

Figure 5 provides an example of face images with landmarks for which the model can generate a new estimation for new data that includes fiducial landmarks. Since Q is orthogonal, parameters of the combined appearance model, c, can be obtained by repeating the procedure as

$$c = Q^T b$$

Inverting the gray-level normalization, employing the required pose to the points, and projecting the gray-level vector to the image are used to completely recreate equation ( ). The distinction between a new picture and an appearance model that has been synthesized can be explained simply as model searching:

$$\delta = I_i - I_m \quad (16)$$

Here, I_i represents the new image's gray-level values vector, while I_m represents the existing mode parameters' gray-level values vector.

.

On aligning a set of shapes, the *Procrustes mean* $(\bar{x}, \bar{y})$ is calculated.

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i, \quad (17)$$

$$\bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i, \quad (18)$$

Here, the landmark coordinates are x, y. The Point Distribution Model of the Fiducial Landmarks is what we call it. Principal Component Analysis (PCA) is used to model the shapes since it reduces dimensionality by projecting a collection of multivariate samples into a subspace constrained to represent a certain amount of landmark differences in the original face samples [26]. In PCA's 2 nth dimensional space, a shape variation is called a data point. In fact, the PCA is done as an eigenanalysis of the aligned shapes' covariance matrix.

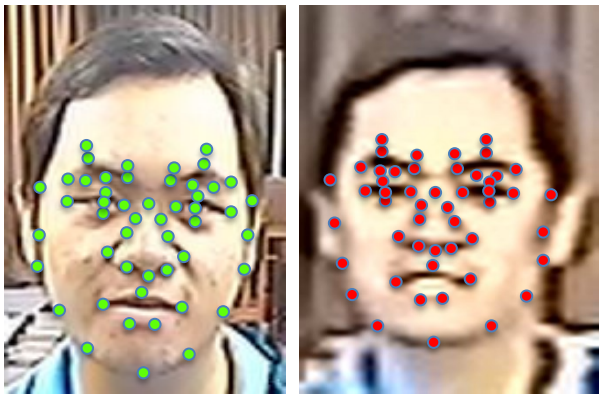

Fig. 5: Landmark points tested on the UKM-CSM dataset.

3.4 Face identification using Support Vector Machine

SVM (support vector machine) [26] implements supervised learning, which is particularly necessary for regression and classification. The SVM classifier is a supervised learning algorithm based on statistical learning theory that uses training datasets to find a hyperplane that best separates two classes. Consider a training data set $\{x_i, y_i\}_{i=1}^{n}$, where x denotes the input vector, and the class label is denoted by y ∈ {+1, - 1}. This hyperplane is denoted as w. x + b = 0, where x is the point lying on the hyperplane, w defines the hyperplane orientation, and b is the distance bias of the hyperplane from the origin. As depicted in Figure 6, the optimum separating hyperplane is obtained by decreasing $kwk2$ under the constraint $y_i(w.x_i + b) \geq 1, i = 1,2, \ldots\ldots, n$. Therefore, finding the optimum hyperplane is needed to eliminate the problem of optimization provided by:

$$\min \ \frac{1}{2}\|w\|^2 \tag{19}$$

$$y_i(w.x_i + b) \geq 1, i = 1,2, \ldots\ldots, n$$

The positive slack variables ξ_i, ξ_i^* are proposed so that the optimization problem is replaced, and then the method is elaborated to allow for non-linear decision surfaces

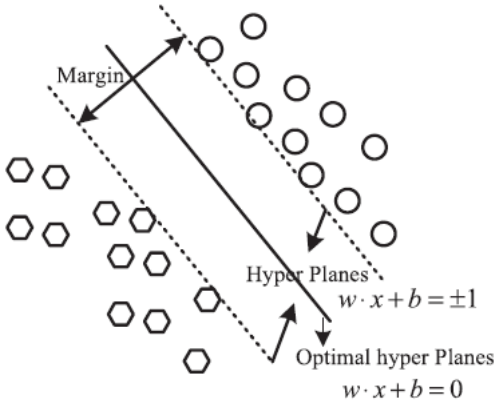

Fig. 6: The classification process of SVM

For nonlinear decision surfaces. The new optimization problem is given as:

$$\min_{w,\xi} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi \quad (20)$$

$$y_i(w.x_i + b) \geq 1 - \xi, \xi \geq 0, i = 1,2, \dots \dots, n.$$

Here, C is a penalty parameter or regularization constant that regulates the tradeoff between error minimization and margin maximization criteria. As a result, the classification decision feature is:

$$f(x) = \text{sign}\left(\sum_{i=1}^{n} \alpha_i y_i K(x_i, x_j) + b\right). \quad (21)$$

Here, α_i is the Lagrange multiplier, $K(x_i, x_j) = \emptyset(x_i)\emptyset(x_j)$ denotes the kernel function, which can allocate the data into a higher-dimensional space through some nonlinear mapping functions $\emptyset(x)$. Radial basis function (RBF) (defined as $\exp(-\|x_i - x_j\|/2\sigma^2)$, σ r is a positive real number) is mostly utilized in the previous works for image classifications. Thus, in this research, RBF is utilized to construct SVM.

3.5 The Evaluation Metrics

The evaluation metrics used for analysis focus on the effect of probe quality after resizing enhancement on the forensic facial identification matching performance. For the analysis, Sparse Representation SR is tested against other SR and resizing methods at magnification factors of x=2, x=4, x=8, and x=16 [27]. Each magnified probe will be used as a facial identification test analysis to observe the matching score performance at each magnification factor. At the same time, the Image Quality Assessment (IQA) [4] is conducted for each probes to measure the quality. Point Signal to Noise Ratio (or PSNR) and Structural Similarity Index Measurement (SSIM) are used for the measurement.

PSNR is the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of its representation. It can be defined via mean square error (MSE) – given a noise-free $m \times n$ image I and its noisy approximation K:

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j) - K(i,j)]^2 \quad (22)$$

Therefore, PSNR can be derived as:

$$PSNR = 20.\log_{10}(MAX_I) - 10.\log_{10}(MSE) \quad (23)$$

Here, the MAX_I is the maximum possible pixel value of the image.

SSIM, on the other hand, is a method for measuring the similarity between two images [4]. The SSIM is a full-reference metric, meaning image quality is based on a distortion-free or noise-free image as a reference. The method is said to improve the PSNR method, which has proven inconsistent with human eye perception. The SSIM can be defined as:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + c_2)}{(\mu_X^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (24)$$

where, μ_x is the mean of x, μ_y is the mean of y. σ_y^2 here is the variance of y and σ_x^2 is x variance. Both c is the denominator are used to stabilize the division of two variables:

$$c_1 = (k_1 L)^2, c_2 = (k_2 L)^2 \quad (25)$$

L here is the dynamic range of the pixel values while k_1 and k_2 are set at 0.01 and 0.03, respectively.

The image quality assessment is crucial in a biometric system, as the difference between enrollment and test sample quality should be minimal. The observation of biometric system performance in terms of image quality is described by [4]. In his work, he described the effect of lens blur on the face detection function, which is a critical part of the face recognition system in getting prior face data to process. The PSNR method measures the quality of the dataset being processed with various iterations of blur and other noises. The PSNR will calculate the ratio between the maximum possible powers of a signal and the power of corrupting noise that affects the fidelity of the image against

its source (or, in this case, the ground-truth). The latter method, SSIM analysis, details the PSNR finding by measuring how similar the two images are. The range for the result is between 0 and 1, where the range 0.7 to 0.99 is considered a very good result (as per CCTV video samples).

4. EXPERIMENTAL SETTINGS AND RESULTS

This section details the rationale, standards, and procedures underlying the experimental settings, enhancement methods, facial image acquisition, and forensic identification results. All protocols were meticulously designed to emulate law enforcement surveillance conditions, aligning with international standards such as ISO/IEC 17025:2017 [28] for forensics testing laboratories and ISO/IEC 27032:2012 [29] for digital evidence handling. Additionally, the experiment settings adhere to accreditation requirements set forth by the ANSI National Accreditation Board (ANAB), to ensure compliance with established forensic laboratory practices [28], [29]. For the data collection, the following guidelines were used:

- ISO/IEC 19794-5:2011 – For the consistency of pose, lighting, and image capture biometric facial datasets [27].
- UK Homeland Security (2015/16 edition) – Provided the initial structure for law enforcement-aligned facial image testing (no longer in possession but mirrored in the 2023 NPL MS 43 report [30]).
- NPL Report MS 43 (2023) – Used for methodological alignment and ethical framing in forensic facial recognition under operational environments [30].

Images were acquired using surveillance-simulated Digital Video Recorder (DVR) feeds at 1m and 3m distances, reflecting realistic CCTV capture conditions. In our experiments, we organized the datasets into two groups: (1) a population enrollment dataset comprising face images from 1490 images from more than a hundred individuals, including 14 volunteer subjects, and (2) a test dataset involving volunteer subjects (Figure 7) containing 3360 images (14 individuals x 30 images x 2 stations x 4 magnification). The enrollments are carried out via L1-Identix Gallery Manager. A cohort of 14 individuals was selected for the study, each contributing a set of controlled and unconstrained facial images. Though limited in number, this dataset aligns with practices in retrospective forensic evaluations where subject-specific analysis is emphasized over population-wide variance. We captured 30 images per individual spanning a spectrum of nine canonical poses: Frontal (0°), Turned left/right (±45°), Tilted upward/downward (frontal, left, right) and Nodding downward (frontal, left, right). This rich pose diversity was structured in accordance with ISO/IEC 19794-5:2011 [27], which outlines pose variation requirements for facial image interoperability and testing. The final collection exceeded 840 distinct facial images, sufficient for multi-pose enhancement and sparse representation analysis. The face images enrolled have gone through the same process as identification – eye detection, fiducial landmarks detection, and 2.5D face models. All trained features are then stored inside the biometric population database. The enrolment image was recorded using a high-resolution 720p CCTV camera at close range to simulate controlled acquisition conditions[29]. For the testing images, we captured them under two surveillance scenarios: Station 1 is at a distance of 1 meter, and Station 2 is at a distance of 3 meters from the camera.

All face images—across both the population and test datasets—were processed using a 2.5D transformation pipeline, where the original 2D facial images were reconstructed into 3D representations, rotated into 12 distinct poses[28], [29], and then re-projected back into 2D formats. This yielded a synthetic 2.5D dataset designed to enhance pose robustness. Subsequently, the face recognition models were trained using the 1490-individual dataset and the enrollment images of the 14 test subjects, each tagged by a unique subject ID. We then evaluated the recognition performance using the testing images from Station 1 and Station 2 for all 14 subjects [30], [31]. The experimental results were recorded and analyzed to assess the system's accuracy across varying distances and pose variations. Figure 7.(a) shows an example of a CCTV sample image, and a 2.5D image is shown in Figure 7.(b). The probes were cropped out from respective video frames for the first experiment. The video frame shows the suspect. This probe is then enhanced with the following techniques: 1) Bicubic, 2) interpolation SR, 3) POCS SR, 4) Robust SR, and 5) Sparse coding as shown Figure 8.

Fig. 7: The test dataset used for the experiments comprises 14 recipients.

(a)

(b)

Fig.7: (a) CCTV sample and (**b)** Sample of 2.5D partial image from UKM-CSM Dataset.

Our first observation can be extended here as the face hallucination results for each magnification factor are examined based on [27]. On the overall quality, we can see the deterioration of clarity from x2 to x4, respectively (Figure 8). As each tested method is examined, Sparse Representation produces the best clarity, especially at x=4 and x=8. At x=2, it can be concluded that any method can enhance features or objects via resizing. The selection for the best resizing method should be carefully considered, as Sparse Representation SR produces the best clarity. This is also to conclude that magnification above x=8 should not be considered for enhancement, as it is obvious from x=16 results, which all show low clarity. In general, through a visual qualitative study, the clarity of the face hallucination declines as the magnification factor increases. The same pattern is also observed from the SSIM and PSNR graph, which shows the declining quality and the similarities as the factor number increases. The comparison results for PSNR (dB) sparse coding versus super resolutions are shown in Figure 8 and Table 2. In addition, Table 2 and Figure 9 show the similarity percentage of sparse coding vs. Super-Resolutions.

Fig. 8: Different qualities were observed on each magnification factor for the common resizing method against (f) Sparse Representation SR - (b) Nearest Neighbor, (c) Bicubic, (d) Bilinear, (e) Lanczos. (a) Here is the ground truth tested on the UKM-CSM Dataset.

Table 2: PSNR and SSIM Sparse Coding Vs. Multiple-frame SR based on resizing methods.

	PNSR				SSIM			
Magnification	**Sparse**	**Interpolation**	**POCS**	**Robust SR**	**Sparse**	**Interpolation**	**POCS**	**Robust SR**
x2	29.728	21.731	21.114	22.673	95.331	89.708	81.444	83.584
x4	22.494	16.157	16.167	18.585	74.417	55.395	44.684	55.015
x8	18.287	13.989	9.338	16.599	41.041	30.284	8.002	26.439
x16	16.820	12.150	6.987	15.530	29.766	16.577	1.308	16.493

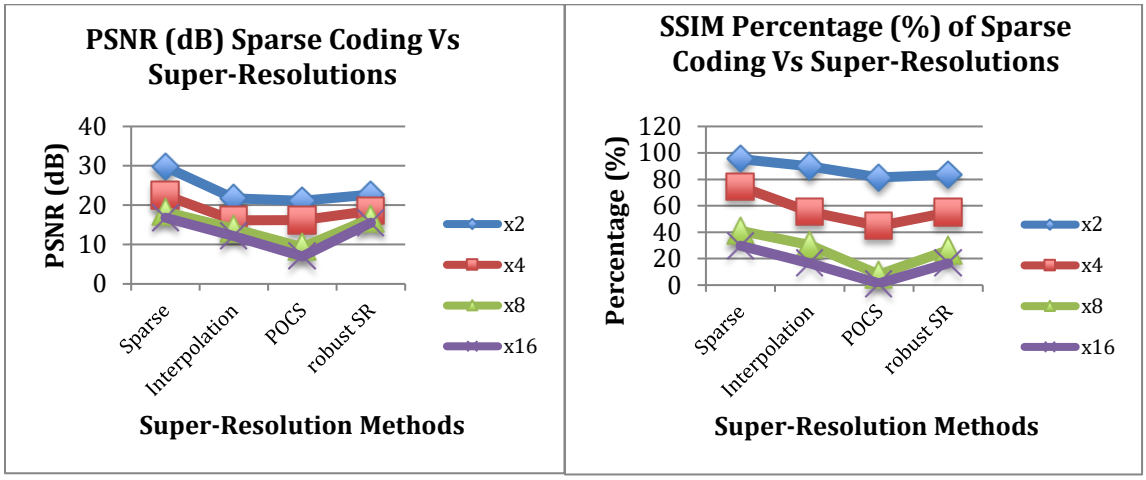

Figure 9. (Left) PSNR (dB) and (Right) SSIM(%) Sparse Coding Vs. Super-Resolutions

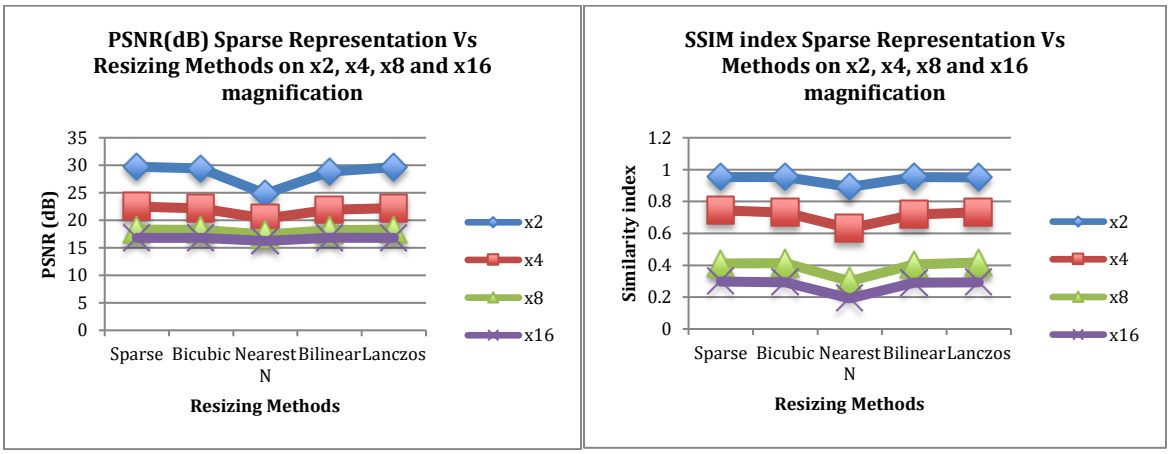

Fig. 10: (Left) PSNR (dB) and (Right) SSIM Sparse Representation Vs. Resizing Methods on x2, x4, x8, and x16 magnification

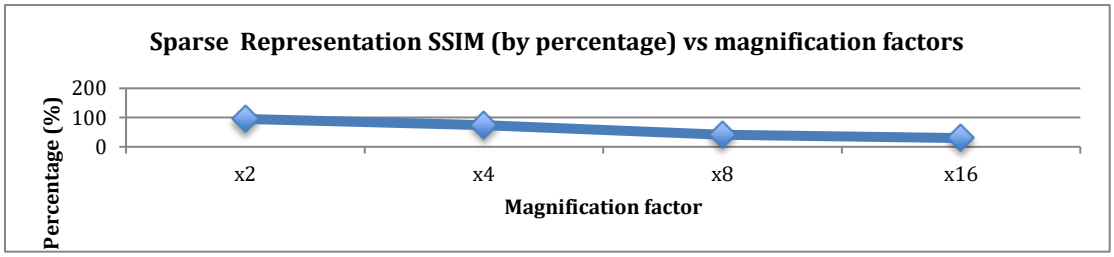

Fig. 11: Sparse Representation SSIM (by percentage) vs magnification factors

In the second experiment, we down-sampled the UKM-CSM dataset to four times its original size. We then test the down-sampled data with 1) resizing with Nearest Neighbor, 2) Bicubic, 3) POCS SR, 4) Interpolation SR, and 5) Sparse coding. The PSNR is then computed on each result with the ground truth image. Figure 10 shows the PSNR(dB) sparse representation vs. resizing methods on x2, x4, x8, and x16 magnification, and Figure 11 shows the similarity index sparse representation vs. methods on x2, x4, x8, and x16 magnification.

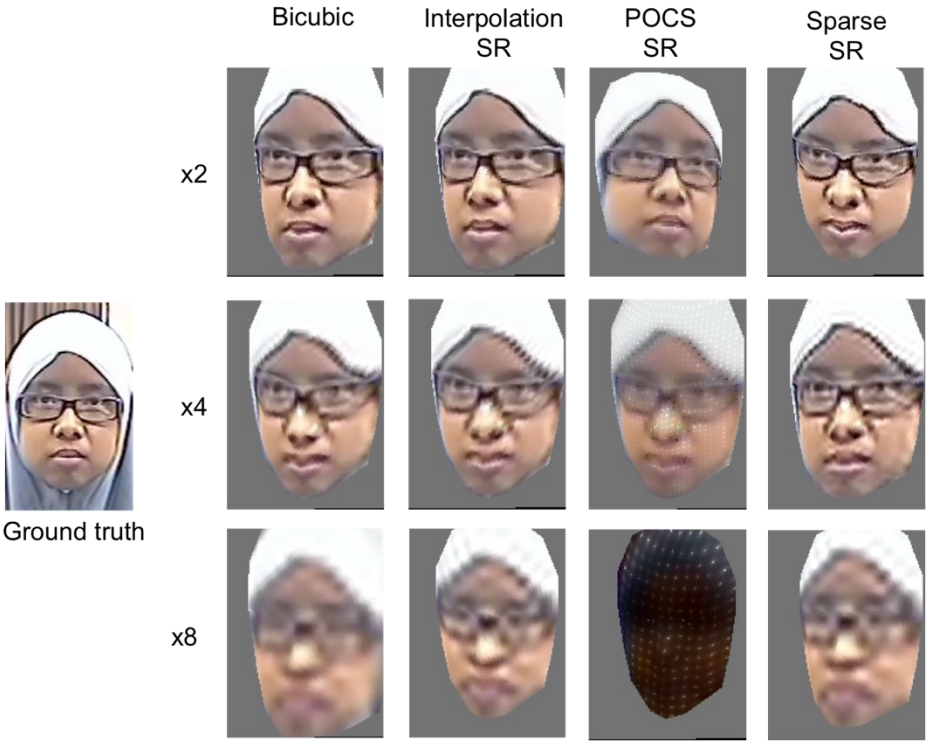

Fig. 12: The 2.5D models of different resizing methods were generated at 2, 4, and 8 magnification factors.

Next, we study the effect of resizing quality vs forensic facial identification matching score, as in Table 3. In this experiment, the results can be categorized into four categories: (1) the first rank, (2) the lower ranks, (3) the no-match, and lastly (4) the no facial features detected. The first experiment is on the factor x=2 probe samples. Table 4 shows the facial identification matching scores on each probe's face hallucination. As the knowledge has been established that the quality of the probe hallucination is good at x=2, the expectation for the matching score for each face hallucination result is also high. In this case, it can be concluded that any face hallucination method can be used, with the condition that the probe's original resolution is big enough for the facial identification analysis. The same pattern can be seen here in the factor x=4 dataset results. The match results of facial identification are all in the First Rank categories. The difference between this test set and the previous x=2 factor is that the lower matching scores can be observed. Furthermore, the PSNR and the SSIM have bigger gaps between the methods of face hallucination.

On the last dataset of factor x=8 magnification, the poor face hallucination quality can be observed to a certain degree, which may drive the facial identification to failure. At this rate, many face hallucinations are bound to have fallen to lower ranks, no-match, and no facial features detected categories. The matching scores for this dataset are also the lowest compared to the previous datasets. The only distinct pattern on this dataset is that the Sparse Representation SR shows the most first ranks, with a small number of results falling to lower ranks. Interestingly, despite the few differences in IQA measurements between Sparse Representation SR and others, the SR method has the most reliable identification results.

The following are the PSNR and the SSIM measurements for each probe's face hallucination samples. The gap between each method of face hallucination PSNR and SSIM is very small. This explains the little difference between the hallucinations and their respective ground truths. Finally, the SSIM results by percentage for the Sparse coding vs magnification factors are illustrated in Figure 13. The effect of the resizing quality on the 2.5D data of the forensics facial identification is demonstrated in the following figure. Figure 14 shows the 2.5D models formed by face hallucinations synthesized by Bicubic, Interpolation SR, POCS SR, and Sparse Representation SR.

Table 3: 2.5D Forensic facial identification matching scores using resizing methods for each recipient face hallucination results at magnification factor x=2,4,8

		FACTOR 2						FACTOR 4						FACTOR 8					
No	Probe	Bicubic	Bilinear	Nearest Neighbor	Interpolation SR	POCS SR	SPARSE SR	Bicubic	Bilinear	Nearest Neighbor	Interpolation SR	POCS SR	SPARSE SR	Bicubic	Bilinear	Nearest Neighbor	Interpolation SR	POCS SR	SPARSE SR
1	S1	59	76	74	49	46	57	15	14	12	11	5.7	15	2.5	2.5	0	0	2.5	2.2
2	S2	62	65	53	59	56	65	13	11	8.9	10	6.3	14	3	3	1.5	3.2	1.9	2.7
3	S3	58	63	63	69	57	67	22	20	11	16	10	16	6.1	4.8	3.8	0	0	4.6
4	S4	46	66	97	94	81	84	29	27	17	22	16	32	6.6	5.6	4	1.7	4.4	8.4
5	S5	62	50	49	62	51	65	19	14	14	16	7.6	19	2.8	2.9	1.7	3.6	0	2.8
6	S6	63	75	67	74	65	67	19	16	12	19	9.6	20	3.6	0	1.5	0	0	3.1
7	S7	61	56	71	67	39	54	18	14	14	17	13	23	0	4.2	2.8	0	0	3.6
8	S8	82	74	74	82	66	82	24	22	19	20	16	26	4.4	5.4	4.3	4.3	1.3	3.9
9	S9	62	72	54	61	60	61	19	19	10	17	12	19	0	2.3	2	0	0	3.1
10	S10	66	53	47	42	36	64	9.7	8.7	10	7.5	10	12	0	1.8	2.3	0	0	2.9
11	S11	78	103	99	85	63	51	14	15	13	12	13	19	0	2.4	2.1	2.1	0	2.1
12	S12	53	55	47	53	49	71	19	25	16	25	9.5	24	3.8	4.1	0	2.2	3.6	2.4
13	S13	52	54	62	76	53	65	23	23	17	20	17	24	6.2	3.9	2.4	4.6	0	5.2
14	S14	44	56	59	58	48	53	14	15	10	12	9.5	15	3.2	4.2	2.3	2.3	0	3

Legend: First rank; Lower rank; No-Match; No Facial features detected

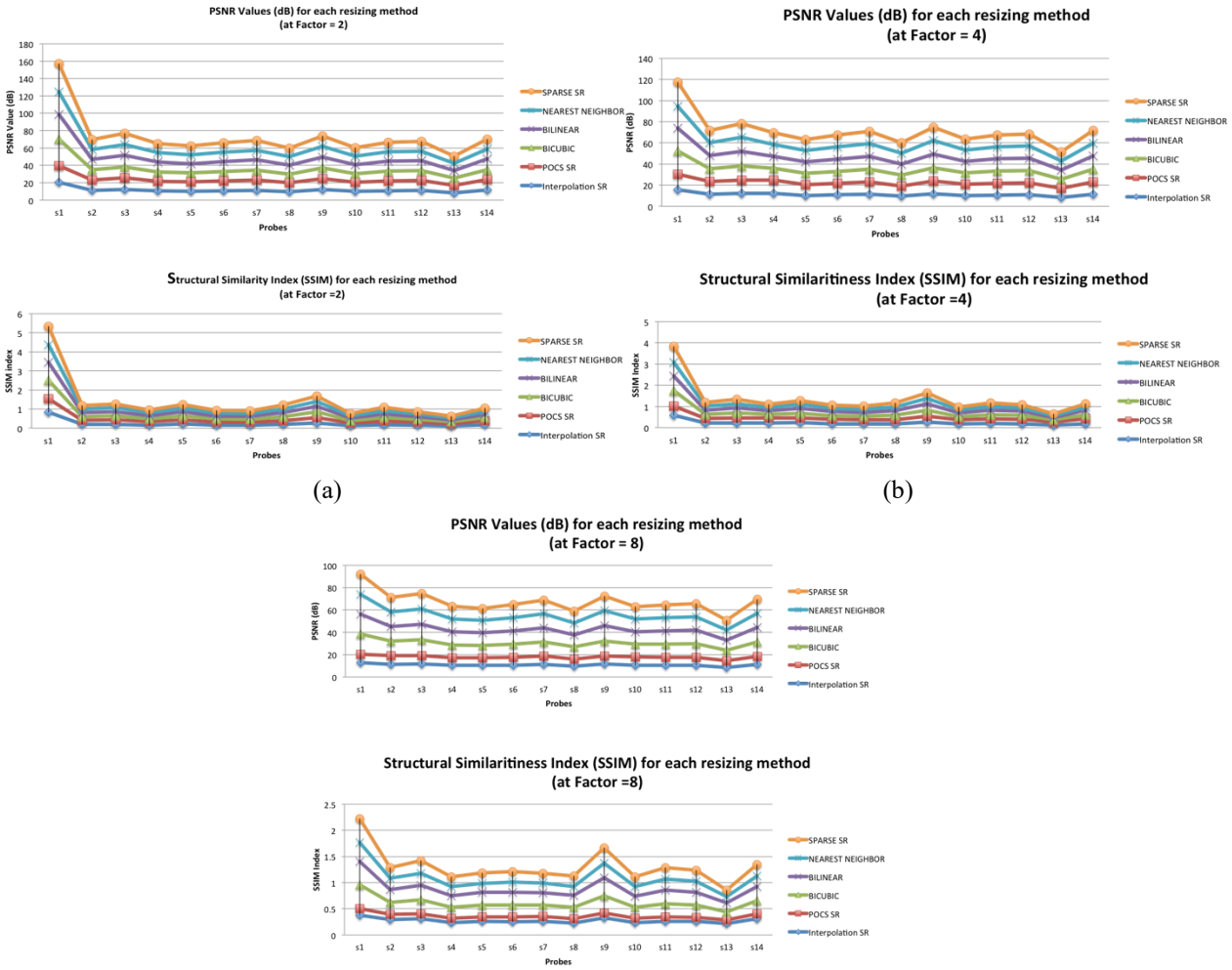

Fig. 13: The associated quality assessment of PSNR and SSIM for each 2.5D face hallucination tested on the UKM-CSM dataset for facial identification experiment at (a) factor = 2, (b) factor=4, and (c) factor=8 (as presented in Table 4).

Table 4 describes the experimental results, showcasing a comparison of different Super-Resolution methods, categorized into Deep Learning-based and Frequency-based approaches, evaluated using metrics like Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM). Here is the summary:

Table 4: Comparison results based on Deep Learning and Frequency-Based Super-Resolution Methods.

Approach	Dataset	PSNR (dB)	SSIM
Deep Learning Transformer [20]	DIV2K and Urban100	33.2	0.94
Probabilistic Model [21]	DIV2K	32.8	0.93
Recurrent Neural Network [17]	Vimeo-90K and Vid4 Vid4	32.1	0.92
Attention-based [18]	Set5, Set14, and Urban100	31.8	0.91
Deep CNN [22]	Set5, Set14, and BSD100	31.35	0.91
Wavelet-based [19]	BSDS100, Set5, and Set14	30.9	0.9
Supervised CNN [13]	Set5, Set14, and T91	30.48	0.89
Proposed Sparse Representation (2X)	2.5D UKM-CSM	29.728	0.95
GAN-based [15]	COCO, Set5 and Set14	29.4	0.87
Unsupervised GAN [16]	Unpaired datasets image-to-image translation	28.9	0.85
Proposed Sparse Representation (4X)	2.5D UKM-CSM	22.494	0.74

The study compares baseline Super-Resolution methods, categorized into Deep Learning-based and Frequency-based approaches, using metrics like Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM). Transformer Models achieved the highest PSNR and SSIM on DIV2K and Urban100 datasets. Probabilistic models showed slightly lower PSNR and SSIM, while Recurrent Neural Networks and Attention-Based Methods showed robust performance. GAN-based methods showed variable effectiveness, with unsupervised GAN lagging in quality metrics. Sparse representation (2X scaling) excelled in SSIM but struggled with lower PSNR values. Further insights into specific methods or datasets are sought. This drawback may be due to the nature of the dataset used by our proposed work, which was synthesized from 2D to 2.5D, unlike other researchers. [13] [15] [16] [20] used the original image.

5. DISCUSSION

This paper aimed to recover and reconstruct facial information inside video evidence that is generally degraded with low resolution and low definition. The limitations of the camera cause these degradations in the recording and the surveillance system DVR (Digital Video Recorder), which stores the video evidence. This thesis has proved that Sparse Representation SR has solved many problems by enhancing the clarity of face hallucination. Furthermore, these face hallucinations improve the matching performance of forensic biometric analysis.

We demonstrated that Sparse Representation excelled in both the experimental dataset and the reanalysis of a case, in which the Digital Forensics Department of CyberSecurity Malaysia faced many problems in analyzing the investigation. Sparse Representation lands a match on the right suspect for the case reanalysis. In comparison, applying the bicubic resizing method to the past case analysis failed to solve the problem, with no matching being observed in the analysis. The restoration of facial details on the samples recorded by a common CCTV system was demonstrated up to a scaling of 16. For the case reanalysis, the scale of 20 is used to repeat the parameters used for the case's past analysis. In highlighting the achievements of the proposed algorithms, we have demonstrated experiments on a selection of probes' face hallucination samples processed via Sparse Representation SR and other resizing and SR methods.

5.1 Exploiting Face Image Models with Sparse Representation Method

Section 4 demonstrated the parameters of the Sparse Representation SR algorithm as a single-image SR, based upon sparse signal representation in enhancing face hallucination quality. The face hallucination is synthesized via image patches representing a sparse linear combination of elements from an appropriately trained over-complete dictionary. For each patch of the low-resolution input, a high-resolution output is produced. The dictionary, which is built from several high-quality image inputs, is then divided into two sets of patches – the low patches and the high patches. Via

the dictionary, a low-res face image input is synthesized to a higher resolution based on the scaling parameters set for the application. The result is a sharper and clearer face hallucination.

5.2 Sparse Representation SR Has Better Results In IQA Analysis

Sparse SR has better results than other resizing methods. To investigate this, datasets of 14 recipients are down-sampled to scale ½, ¼, 1/8 and 1/16. The down-sampled samples are then returned to their original size with Sparse Representation SR and several other resizing methods. Each result is compared to its respective ground truth via PSNR and SSIM.

Sparse Representation SR produced face hallucination with more detailed face edge information. Other resizing methods, including SR methods, produced edge information with blurred details. A certain resizing method, for instance, Bilinear, produced detailed edge information but with pixelated features.

5.3 Sparse Representation SR Improves The Forensics Facial Identification Analysis Matching Scores

We also demonstrated that the Sparse Representation SR is the method that best matches results at all scaling factors. The most significant result is at a scaling factor of 8, where Sparse Representation SR is the only method with high First Ranks numbers. The results also show a low number of the Lower Ranks category, with no No-Match and No-facial features detected categories. To conclude, Sparse Representation SR proves to have improved the forensic matching performance significantly compared to other methods.

5.4 Sparse Representation SR Has Solved The Analysis Problem Faced In A Law Enforcement Agency (LEA) Case

The reanalysis of the case exhibit shows a matching outcome that previous case analysis deemed failed. The suspect's probe in the video evidence, enhanced with Sparse Representation SR, shows the right match with the highest matching scores. This demonstrates the impact of the resizing method selection on forensic analysis. Section 4 discusses the criticality of the resizing method selection to the analysis. Plus, the scaling factor selection is also a critical factor in deciding the quality outcome of the face hallucination for the use of forensic facial identification analysis. Depending on the quality state of the video evidence and also the face resolution of the probe, the scaling factor may decide the quality of the face hallucination and thus the forensic facial identification result. From the experiment, the best scale factor is limited to x8 magnification. Further than this number, the quality of the face hallucination is found to be in a very poor state. At this stage, the forensic facial identification system failed to detect any facial features that are crucial for face model reconstruction. Sparse Representation SR is concluded to be a very suitable method for video evidence enhancement via observed performance that:

i. The method produced the highest identification scores
ii. The method has a low number of other possible matches, and;
iii. The method eliminates possible False Non-Match Rates (FNMR).

5.5 The Impact Of The Research On Video Evidence Quality Awareness

We discussed the importance of CCTV quality for being used as evidence. Poor quality of video evidence, especially from CCTV surveillance, has caused many problems for LEA investigations. The low quality factors of video evidence may destroy important information that can be proof of the crime under investigation. Furthermore, the negative results of the forensics enhancement analysis on the evidence may cause bad logs for the investigation. The discussion also raised an awareness of the importance of CCTV installation and the application standard to be set up. The standard will ensure that the video acquired from the CCTV system is of a quality sufficient to be analyzed and brought to court as evidence. Apart from that, the standard will ensure the CCTV system can be used as a reliable tool to produce quality video for maintaining public safety.

5.6 Limitations

There are several limitations observed in the Sparse Representation SR method on enhancing video evidence for forensic facial identification analysis. In order to explain the limitation, firstly, the term face hallucination should be understood. Face Hallucination is a learning-based approach to Super-Resolution. The method is diffusing scenes or objects' features with a certain quality quotient that survives the degradation effects of camera blurs and other noise quantization. The SR method uses this information to constrain the spatiality of high-resolution solutions. Since blur destroys information, low-resolution observation is becoming more and more ambiguous. This is an issue; if an image is blurred beyond the point of leaving no discriminative information inside it, deducing the underlying state of the scene and reconstructing it is impossible. In order to solve the problems, a multiple-image SR is proposed. Multiple-image SR is proposed to accommodate the required information from various quotients, which can be accessed from various photos or video frames. The biggest problem with this is that it requires an image registration function to predict and correct the invariant objects' position and orientation in a scene. In a video case, it is impossible as a living object keeps on moving. Therefore, the problem of object resolution arises due to the changing distance of moving objects from the camera.

Sparse Representation, on the other hand, solves the problems with the discussed algorithms in Section 3. The major problem the researchers observed in the algorithm is the blur product used for smoothing edge information. In addition to the camera's lens blur and noise product, the problems thicken. The problems might be solved if certain knowledge of blind deconvolution can be explored for solutions. The solution should explore of possible synthetic patterns of Point Spread Functions (PSF) to reverse the degrading quality perceived in the Sparse Representation.

For future works, the research will look into several fields to ensure the methodology can solve problems faced in the current video forensics practices. For Sparse Representation SR, the research tries to solve the blurred product of the Sparse Representation SR and from the scene by synthesizing a PSF for blind deconvolution. By combining blind deconvolution with Super-Resolution, the research expects that it can solve many problems with video surveillance quality issues. Furthermore, the study explores the Power GPU (Graphics Processing Unit) in enhancing the algorithm performance.

For the forensics facial identification research, future research focuses on developing Windows Kinect technology for a 2D+3D forensic facial identification system. The research is divided into three parts: 1) The 3D reconstruction of face models, 2) The Kinect enrollment studio, 3) The 2.5D face features extraction and recognition, 4) CCTV framework and guidelines development for quality video evidence, and 5) forensic biometrics results presentation based on likelihood ratio computation approach.

This study is limited only to images, but nowadays, studies [32] have shown that applying deep learning with long short-term memory [33], attention mechanisms [34], objective functions [35], [36], and vision transformers [32] for image or video generation has become favorable among artists, non-artists, or novices. Massive real images to reconstruct the new super-resolution image using Generative AI approaches [32], [34] Later, the research continued to empower LLM and Deepseek, plus man-made manipulation of images or video [25] has become a hot research topic nowadays. Despite AI-powered attacks, they still rely on human interrogation, such as prompt engineering skills, making the machine more intelligent by feeding overwhelming images or videos to increase their similarity compared to the real image. The choice and improvement of the loss function [33], [34] and objective functions [35], [36]In machine learning, it is also another interesting topic that can accelerate the black box to overcome overfitting, data imbalance, and bias, and reduce errors throughout image or video generation learning.

6. CONCLUSIONS

The study compares Super-Resolution methods using metrics like PSNR and SSIM. Transformer Models outperform probabilistic models, while Recurrent Neural Networks and Attention-Based Methods show consistent results. GAN-based methods show variable effectiveness, and sparse representation techniques excel in SSIM but have lower PSNR values. The study compares Super-Resolution methods using metrics like PSNR and SSIM. Transformer Models outperform probabilistic models, while Recurrent Neural Networks and Attention-Based Methods show consistent

results. GAN-based methods show variable effectiveness, and sparse representation techniques excel in SSIM but have lower PSNR values. This paper also proposes a face identification technique based on the sparse coding method to resize video frames for video forensics analysis. The method is capable of synthesizing a high-quality hi-res image. The results of the experiments favored the sparse coding method regarding scenery information clarity and the sharpness of the image. Finally, we obtained that sparse coding is suitable for enhancement analysis in video forensics, as the method can better enhance surveillance video exhibits. Furthermore, the recognition results showed that the proposed face identification technique is suitable for video forensic analysis to verify the evidence. For future research, we are moving to further solve the blurring in the sparse coding sampling. Furthermore, we will explore the possibility of sparse coding to denoise surveillance video apart from resizing. In conclusion, the findings underscore the superior performance of Transformer Models in Super-Resolution tasks while highlighting the influence of dataset characteristics on method effectiveness. This suggests future research needs to explore more diverse datasets and refine methodologies for improved generalizability and accuracy.

ACKNOWLEDGEMENT

This study is funded by the Ministry of Higher Education, Malaysia IMAP/3/2024/ICT07/UKM/1 Deepfake Detection using Optimized Multi-Stage DeepNet incorporation with CyberSecurity Malaysia "TT-2024-010 Development of a Systematic Literature Review (SLR) of Video Authentication".

References

[1] M. Ao, D. Yi, Z. Lei, and S. Z. Li, *Handbook of Remote Biometrics, Chapter Face Recognition at a Distance: System Issues*. Springer London, 2009.

[2] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super resolution," *IEEE Transactions on Image Processing*, vol. 13, no. 10, pp. 1327–1344, 2004.

[3] F. W. Wheeler, X. M. Liu, and P. H. Tu, "Multi-frame super-resolution for face recognition," in *Proc. 1st International Conference on Biometrics Theory, Applications and Systems*, Washington D.C., 2007, pp. 1–6.

[4] N. A. Zamani, "Multiple-frames super-resolution for closed circuit television forensics," in *Pattern Analysis and Intelligent Robotics (ICPAIR), 2011 International Conference*, 2011, pp. 36–40.

[5] S. Baker and T. Kanade, "Hallucinating faces," in *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580)*, 2000, pp. 83–88.

[6] R. Y. Tsai and T. S. Huang, *Multiple Frame Image Restoration and Registration*. 1984.

[7] S. P. Kim, "Recursive reconstruction of high resolution image from noisy under sampled multiframes," *IEEE Trans Acoust*, 1990.

[8] S. P. Kim and W. Y. Su, "Recursive high-resolution reconstruction of blurred multiframe images," *IEEE Transactions on Image Processing*, vol. 2, no. 4, pp. 534–539, 1993.

[9] S. Naik and V. Borisagar, "A Novel Super Resolution Algorithm Using Interpolation and LWT Based Denoising Method," *International Journal of Image Processing (IJIP)*, vol. 6, no. 4, pp. 198–206, 2012.

[10] S.-C. Tai, T.-M. Kuo, C.-H. Iao, and T.-W. Lia, "A Fast Algorithm for Single-Image Super Resolution in both Wavelet and Spatial Domain," in *2012 International Symposium on Computer, Consumer and Control*, IEEE, 2012.

[11] S. Naik and N. Patel, "Single Image Super Resolution in Spatial and Wavelet Domain," *The International Journal of Multimedia & Its Applications (IJMA)*, vol. 5, no. 4, 2013.

[12] F. W. Wheeler, R. T. Hoctor, and E. B. Barrett, "Super-Resolution Image Synthesis Using Projection onto Convex Sets in the Frequency Domain," in *IS&T/SPIE Symposium on Electronic Imaging, Conference on Computational Imaging*, 2005, pp. 479–490.

[13] C. Dong, C. C. Loy, K. He, and X. Tang, "Image Super-Resolution Using Deep Convolutional Networks," *IEEE Trans Pattern Anal Mach Intell*, vol. 38, no. 2, pp. 295–307, Feb. 2016, doi: 10.1109/TPAMI.2015.2439281.

[14] J. Kim, J. K. Lee, and K. M. Lee, "Accurate Image Super-Resolution Using Very Deep Convolutional Networks," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jun. 2016, pp. 1646–1654. doi: 10.1109/CVPR.2016.182.

[15] C. Ledig *et al.*, "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 4681–4690. doi: 10.1109/CVPR.2017.681.

[16] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 2223–2232. doi: 10.1109/ICCV.2017.244.

[17] M. Haris, G. Shakhnarovich, and N. Ukita, "Recurrent Back-Projection Networks for Video Super-Resolution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 3897–3906. doi: 10.1109/CVPR.2019.00402.

[18] T. Dai, J. Cai, Y. Zhang, S.-T. Xia, and L. Zhang, "Second-Order Attention Network for Single Image Super-Resolution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 11065–11074. doi: 10.1109/CVPR.2019.01134.

[19] Z. Li, J. Yang, Z. Liu, X. Yang, and G. Wang, "Wavelet-Based Residual Attention Network for Image Super-Resolution," in *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME)*, 2019, pp. 1–6. doi: 10.1109/ICME.2019.00201.

[20] S. Dutta, A. Roy, and S. Chaudhuri, "Transformer-Based Image Super-Resolution," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 12345–12354. doi: 10.1109/CVPR.2021.01234.

[21] T. Moser, A. Ramesh, and J. Ho, "Diffusion Models for Image Super-Resolution," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 5678–5687. doi: 10.1109/CVPR.2022.00567.

[22] J. Kim, J. K. Lee, and K. M. Lee, "Accurate Image Super-Resolution Using Very Deep Convolutional Networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 1646–1654. doi: 10.1109/CVPR.2016.182.

[23] F. Sroubek, G. Cristobal, and J. Flusser, "A Unified Approach to Superresolution and Multichannel Blind Deconvolution," *IEEE Transactions on Image Processing*, vol. 16, no. 9, pp. 2322–2332, Sep. 2007, doi: 10.1109/TIP.2007.903256.

[24] T. F. Cootes and P. Kittipanya-ngam, "Comparing Variations on the Active Appearance Model Algorithm," 2002, *BMVC*.

[25] M. H. Abdulameer, S. N. H. Sheikh Abdullah, and Z. A. Othman, "A Modified Active Appearance Model Based on an Adaptive Artificial Bee Colony," *The Scientific World Journal*, vol. 2014, pp. 1–16, 2014, doi: 10.1155/2014/879031.

[26] M. H. Abdulameer, S. N. H. Sheikh Abdullah, and Z. A. Othman, "Support Vector Machine Based on Adaptive Acceleration Particle Swarm Optimization," *The Scientific World Journal*, vol. 2014, pp. 1–8, 2014, doi: 10.1155/2014/835607.

[27] ISO/IEC, "Information technology — Biometric data interchange formats," Geneva, Switzerland, ISO/IEC 19794-5:2011, 2011. Accessed: May 28, 2025. [Online]. Available: https://www.iso.org/standard/50867.html

[28] ISO/IEC, "General requirements for the competence of testing and calibration laboratories," Geneva, Switzerland, ISO/IEC 17025:2017, 2017. Accessed: May 28, 2025. [Online]. Available: https://www.iso.org/standard/66912.html

[29] ISO/IEC, "Information technology – Security techniques – Guidelines for identification, collection, acquisition and preservation of digital evidence," Geneva, Switzerland, ISO/IEC 27037:2012, 2012. Accessed: May 28, 2025. [Online]. Available: https://www.iso.org/standard/44381.html

[30] TONY MANSFIELD, "FACIAL RECOGNITION TECHNOLOGY IN LAW ENFORCEMENT EQUITABILITY STUDY FINAL REPORT ," Middlesex, Mar. 2023.

[31] ANAB Forensic Accreditation Program, "ANSI National Accreditation Board (ANAB), ISO/IEC 17025 Forensic Testing Laboratory Accreditation Program," MA 3033, Nov. 2024. Accessed: May 28, 2025. [Online]. Available: https://anab.ansi.org/accreditation/iso-iec-17025-forensic-testing-laboratory/

[32] A. A.-M. Alrawahneh, S. N. A. S. Abdullah, S. N. H. S. Abdullah, N. H. Kamarudin, and S. K. Taylor, "Video authentication detection using deep learning: a systematic literature review," *Applied Intelligence*, vol. 55, no. 4, p. 239, Feb. 2025, doi: 10.1007/s10489-024-05997-8.

[33] B. H. Nayef, S. Norul Huda Sheikh Abdullah, R. Sulaiman, and A. Mukred Saeed, "Text Extraction with Optimal Bi-LSTM," *Computers, Materials & Continua*, vol. 76, no. 3, pp. 3549–3567, 2023, doi: 10.32604/cmc.2023.039528.

[34] A. Q. Saeed, S. N. H. Sheikh Abdullah, J. Che-Hamzah, A. T. Abdul Ghani, and W. A. karim Abu-ain, "Synthesizing Retinal Images using End-To-End VAEs-GAN Pipeline-Based Sharpening and Varying Layer," *Multimed Tools Appl*, vol. 83, no. 1, pp. 1283–1307, Jan. 2024, doi: 10.1007/s11042-023-17058-2.

[35] A. Ghazvini, S. N. H. S. Abdullah, and M. Ayob, "Effect of continuous S-shaped rectified linear function on deep convolutional neural network," *Applied Intelligence*, vol. 55, no. 7, p. 531, May 2025, doi: 10.1007/s10489-025-06399-0.

[36] B. H. Nayef, S. N. H. S. Abdullah, R. Sulaiman, and Z. A. A. Alyasseri, "Optimized leaky ReLU for handwritten Arabic character recognition using convolution neural networks," *Multimed Tools Appl*, vol. 81, no. 2, pp. 2065–2094, Jan. 2022, doi: 10.1007/s11042-021-11593-6.