Journal of Engg. Research Vol. 7 No. (2) June 2019 pp. 178-195

Tangential and exponential kernel weighted regression model for multi-view face video super resolution

Amar B. Deshmukh* and Dr. N Usha Rani**

**Research Scholar, Vignan University, Guntur, Andhra Pradesh 522213, India*
***Professor & Head, Department of ECE, Vignan University, Guntur, Andhra Pradesh 522213, India*
**Corresponding Author: amarbdeshmukh@gmail.com.*

ABSTRACT

The necessity of recognizing a person from a low-resolution non-frontal image is a challenging problem in video surveillance. In order to alleviate the problem of recognition in the low-resolution image, the literature presents different techniques for face recognition after converting the low-resolution image to high resolution. Accordingly, this paper presents a technique for multi-view face video super resolution using the tangential and exponential kernel weighted regression model. In this paper, a new hybrid kernel is proposed to perform non-parametric kernel regression model for estimation of a neighbor pixel in the super resolution after the face detection is performed using Viola-Jones algorithm. The experimentation is performed with the UCSD face video databases, and the quantitative results are analyzed using the SDME with the existing techniques. From the resulting outcome, we prove that the maximum SDME of 77.3db is obtained for the proposed technique as compared to the existing techniques like nearest interpolation, bicubic interpolation, and bilinear interpolation.

Keywords: Super resolution, face video, face detection, kernel, Viola-Jones algorithm, a second-derivative-like measure of enhancement (SDME).

INTRODUCTION

In recent years, the intelligent surveillance system is widely applied in several fields, which include security and protection camera, where the resolution of the required face in the picture is very low and so it cannot provide the desired information. Moreover, there are some constraints in the imaging conditions in certain situations, and so acquiring face images with high resolution is not always possible. Consequently, the face images captured by the camera may miss many facial feature details the human needs to be identified. So, image resolution enhancement techniques, particularly, human face resolution called face super resolution, are getting more attention (Qu *et al.,* 2014). Face super resolution (SR), which is also called face hallucination, means hallucinating the high-resolution (HR) face image from its low-resolution (LR) image. Baker and Kanade proposed the term face hallucination (Baker & Kanade, 2000).

In general, face super resolution techniques (Li *et al.,* 2010; Kim & Kown, 2010) are mainly classified into two general categories, that is, learning-based techniques (Zhuang *et al.,* 2007; Chang et al., 2004) and reconstruction-based techniques (Park & Lee, 2008; Farsiu *et al.,* 2004). In learning-based techniques, the prior information about the deviation of the high-resolution images from the low-resolution images is used to construct the model, and the constructed model is used in performing face super resolution. Here, the main problem in learning based super resolution method is using precise prior information for the reconstructing high-resolution image. To overcome this issue, reconstruction-based techniques are used. There are two approaches in reconstruction based techniques, that is, global approach and local patch based approach. Here, the images are divided into overlapped patches and for each patch, the nearest neighbors are used as the prior to creating the required high-resolution patches. In global approach method, the whole face images are considered as a global model and in local approach method, the faces for arranged patches are considered as a local model. In simulation, these methods achieve good results when applied to very low-resolution faces (Lu *et al.,* 2013a).

The local approach has good subjective quality than the global approach due to the smooth face image (Tao *et al.,* 2012; Jiang *et al.,* 2014). As the local image patches are similar in local approach, one image patch can be represented using a few neighbor patches, which result in a local representation of image patches. Furthermore, there is no noise in locality based representation as the noisy image patch is replaced with a similar clear image patch without synthesizing the noisy image patch as in LSR and SR (Jiang *et al.,* 2014). Traditional parametric super resolution methods are based on a certain model of the signal of interest and try to calculate the model parameters in the presence of noise (Takeda *et al.,* 2007). Then, a generative model based on the estimated parameters is generated as the best estimate of the original image. On the contrary, nonparametric methods are based on the data itself to explain the structure of the model, and this model is referred to as a regression function (Wand & Jones, 1995). Due to the arrival of recent machine learning techniques, kernel methods are becoming more popular and most commonly used for pattern detection and discrimination problems (Yee & Haykin, 1993).

In this paper, we have presented a new hybrid kernel for non-parametric estimation of pixels using local patching process. The proposed technique performs multi-view face video super resolution using tangential and exponential weighted regression model. At first, the input face video is read out, and the frames are extracted from the face video. Once the frames are extracted, the face detection is performed using Viola-Jones algorithm, which contains the important process of rectangular feature extraction, training, and testing of AdaBoost classifier. Once the face region is detected, the low-resolution face part is given to the final step where the multi-kernel regression model is applied to obtain super resolution of the face part. The super resolution image is then analyzed with the evaluation called SDME.

The major contribution of this paper is given as follows:

- **A novel hybrid kernel**: The hybrid kernel matrix is the usage of both the exponential kernel and the tangential kernel, which is designed to perform the video super resolution using the non-parametric regression model that is based on the weights of the neighboring pixels. The tangential kernel preserves the singularity within the disc resolved in the shape, whereas the exponential kernel estimates the pixels.

The paper is organized as follows. Section 2 presents the review of the literature and Section 3 presents the motivation behind the approach. Section 4 discusses the multi-kernel based regression model face multi-view face super resolution. Section 5 presents the detailed experimentation and comparative results. Finally, the conclusion is given in section 6.

LITERATURE REVIEW

Table 1 reviews the recent literature related to face super resolution. Most of the techniques utilized the local neighbourhood-based estimation (Qu *et al.,* 2014; Jiang *et al.,* 2014; Hu *et al.,* 2010) to perform the super resolution. Some techniques (Wang *et al.,* 2014; Lu *et al.,* 2013b) utilize the nonnegative matrix factorization to estimate the pixel even though it is an overhead computation process. Also, some methods (Tao *et al.,* 2012) utilize the learning algorithm for estimation of super resolution image even though the image prior information is required. From the literature, we identify that the estimation technique without the need of image prior and the efficient computational algorithms can be a better choice to proceed further in face video super resolution.

Xiao Zeng and Hua Huang (2012) have proposed a regression based method for face super resolution. It has better subjective quality due to more smooth face image but it requires accurate prior for high-resolution image reconstruction. Lu Tao *et al.* (2012) have proposed shape clustering and subspace learning-based model for face super resolution. It is robust to noise but it maintains the shape in the reconstruction process, which is a problem worth to explore in the future. Shenming Qu *et al.* (2014) have proposed a method for face super resolution. The position-based patch neighborhood considers the spatial distance to obtain improved quality but finding the distinct patch requires more time. Xiaofeng (Wang *et al.,* 2014) have proposed a non-negative matrix factorization for face super resolution. This method maintains the relation between high-resolution residue and low-resolution residue to better preserve high frequency details but selecting NMF basis images corresponding to specific face parts is more challenging.

Xiang Ma *et al.* (2015) have developed a face super resolution by considering the redundant transformation with diagonal loading. It offers robustness when dealing with the inputs that have different expressions, head poses, and illuminations but this approach is computationally intensive and sensitive to training examples. Junjun Jiang *et al.* (2014) have proposed a locality-constrained representation for face super resolution, and it is very robust against noise in real surveillance scenarios but fails to discover the intrinsic geometrical structure of the data set. Tao Lu *et al.* (2013b) have proposed a nonnegative matrix factorization for face image super resolution. It has a better performance on local facial detailed features due to nonnegative part-based features but factorization requires much computational effort. Yu Hu *et al.* (2010) have proposed a local pixel structure-based method for face super resolution, and it exhibits an impressive ability to infer the fine facial details and to generate plausible HR facial images from very small LR input but it requires reference samples for HR estimation.

Table 1. Literature review.

Author	**Method**	**Advantages**	**Disadvantages**
Xiao Zeng and Hua Huang (2012)	Regression based method	Has better subjective quality due to the more smooth face image	Require accurate prior for high-resolution image reconstruction
Lu Tao *et al.* (2012)	Shape clustering and subspace learning	Robust to noise	Maintain the shape in the reconstruction process is a problem worth to explore in future.
Shenming Qu *et al.* (2014)	Position patch neighborhood preserving	Considering spatial distance helps to obtain improved quality	finding the distinct patch requires more time
Xiaofeng Wang *et al.* (2014)	Non-negative matrix factorization	The relation between high-resolution residue and low-resolution residue to better preserve high frequency details	Selecting NMF basis images corresponding to specific face parts is more challenging
Xiang Ma *et al.* (2015)	Redundant transformation with diagonal loading	Offers robustness when dealing with the inputs that have different expressions, head poses, and illuminations	This approach is computationally intensive and sensitive to training examples
Junjun Jiang *et al.* (2014)	Locality-constrained Representation	It is very robust against noise in real surveillance scenarios	Fails to discover the intrinsic geometrical structure of the data set
Tao Lu *et al.* (2013b)	Nonnegative matrix factorization	Better performance on local facial detailed features due to nonnegative part-based features	Factorization requires much computational effort
Yu Hu *et al.* (2010)	Local pixel structure-based method	Exhibits an impressive ability to infer the fine facial details and to generate plausible HR facial images from very small LR input	Require reference samples of HR estimation

Junjun Jiang *et al.* (2016)	SR-based face image super resolution approach	It exhibits better performance both quantitatively and qualitatively even in the presence of high level of noise in the input LR face	Takes huge time to represent and reconstruct the overlap patch and this method lacks parallel computation
Junjun Jiang *et al.* (2016)	Missing intensity interpolation method based on smooth regression with local structure prior (LSP), named SRLSP	Facilitates parallel computation that reconstructs the HR patch of the target independently.	Not applicable for real time face recognition and 3D face synthesis as they consume large amount of time.
Junjun Jiang *et al.* (2016)	A face SR method based on Tikhonov regularized neighbor representation (TRNR)	Enables the relevant selection of the patches even under the presence of the noise.	Did not concentrate on recovering the frontal HR face image and did not enable parallel computation.
Junjun Jiang *et al.* (2016)	Super resolution via Locally Regularized Anchored Neighborhood Regression and Nonlocal Means	Possess the capacity to generate the output with the sharp edges and rich textures. Highly reliable SR construction.	Requires the filter to improve the SR-output.
Zhiliang Zhu *et al.* (2014)	Super resolution via Self-Example Learning and Sparse Representation	Does not require HR training set and hence, it exploits image patches within a single image and sparse representation, using a single learned dictionary. This method is more practical.	The performance of this method is similar to the existing SR approaches in terms of the visual effect.

MOTIVATION BEHIND THE APPROACH

Problem definition

The main objective of this paper is to perform the super resolution on the face region, which is obtained from the face video. So, the input for the proposed system is the face video, which contains multiple of frames. Every frame is given to the face detection process, which detects the face region. Once the face region is extracted, the super resolution process is performed on the face region to increase the number of pixels based on the upscaling factor without compromising the visual quality. As such, the input video V_i is represented as N frames as

$$V = \left\{ V_i , 1 \le i \le N \right\}$$

From every frame of the video, the face region should be identified and extracted.

$$I = VI(V_i)$$

where $VI(\bullet)$ is the function to extract the face region. The extracted face regions can be represented as

$$I_{face} = \{I_{ij} ; 1 \le i \le ; 1 \le j \le n\}$$

where $m \times n$ is the size of the face region. The problem considered here is to perform the super resolution of the face region by increasing the size of the region ($m \times n$) to ($m * r \times n * r$), where r is the upscaling factor. The image obtained after performing super resolution is given as follows:

$$I_{super} = \left\{ I^s{}_{ij} ; 1 \le i \le m \times r ; 1 \le j \le n \times r \right\}$$

Challenges

Face super resolution is an active area of research due to the wide applicability of the method in video surveillance system, especially for security. When analyzing the recent works available in the literature for face super resolution (Hu *et al.,* 2010; Viola & Jones, 2004), the following challenges are identified.

Most of the super resolution frameworks require accurate image prior for high-resolution image reconstruction. This poses a challenge of identifying the accurate training face samples for performing effective learning task.

Utilizing the shape metrics for reconstruction can have good performance but the storing and maintaining of shape metrics are very difficult to handle.

Selecting the suitable size of neighbourhoods and shape of the neighbourhoods is a challenging problem in super resolution because it is directly related to the estimating behaviour of the pixels.

Even though most of the methods are a face image super resolution, the important challenge in the surveillance camera is constructing super resolved faces from the low-resolution videos because the face should be detected correctly from multiple views, and the estimated super resolved face should be useful for improving the performance of face recognition task.

PROPOSED METHODOLOGY: TANGENTIAL AND EXPONENTIAL KERNEL WEIGHTED REGRESSION MODEL FOR MULTI-VIEW FACE VIDEO SUPER RESOLUTION

This section presents the multi-view face video super resolution using tangential and exponential kernel weighted regression model. Here, local patch estimation and the multi-kernel weighted regression model are included for generating super resolved pixels. Overall, the proposed technique consists of three major steps. In the first step, the video is directly read out, and it is converted into a set of frames. In the second step, every frame is serially taken, and the face part is detected using Viola–Jones object detection (Viola & Jones, 2004), which is one of the popular methods applied for face detection. Once the face part is detected for the input frame, the resolution of the face detected part is improved using the proposed face super resolution method in the third step. The proposed super resolution method integrates the process of the multi-kernel weighted regression model and local patching process. Here, based on the scaling factor, the local patch size is decided, and the estimation of pixels is performed using the exponentially weighted regression model. This step is repeated for every frame, and the super resolved faces are separately stored for the analysis part. The block diagram of the proposed multi-view face video super resolution is given in Figure 1.

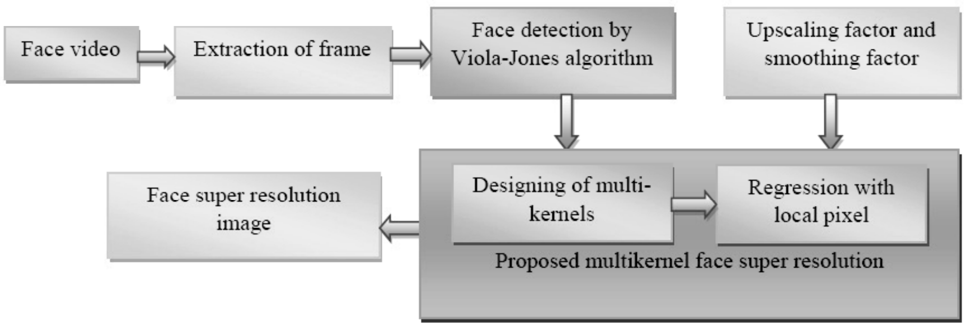


Fig. 1. Block diagram of the proposed multi-view face video super resolution.

Video reading and frame extraction

The input for the proposed multi-view face video super resolution is the video that is represented as the multiple frames. The video, V, is directly read out and every frame V_i is taken out to find the face region. The input video may be in any of the file formats like AVI, MPEG, 3GPP, and so on. The reading of video is performed by constructing

VideoReader object and extracting one frame at a time associated with. Here, i denotes the frame index and N represents the number of frames in the input video.

Face detection by Viola-Jones algorithm

The second step of the proposed multi-kernel weighted regression model is to detect the faces from every frame using Viola-Jones algorithm (Viola & Jones, 2004), which is the popular method for face detection. Even though many different methods are presented in the literature (Fleuret & Geman, 2001; Roth *et al.,* 2000), the Viola-Jones algorithm is taken here for the face detection because it can detect the multi-view faces effectively. This algorithm performs the face detection process using three important steps, that is, i) feature extraction, ii) training of AdaBoost classifier, and iii) detection of the face region.

i) Feature extraction: At first, the input frame $I(a,b)$ is given to the feature extraction step, which finds the rectangular features using an integral frame. The integral frame is computed by finding the summation of the pixels intensity from the above and left parts of the location of the pixels. The process of finding the integral frame is given as follows:

$$II(a,b)=\sum_{a^{'}\le a,b^{'}\le b} I\left(a^{'},b^{i}\right)$$

where $II(a,b)$ is the integral frame and $I(a,b)$ is the original frame. Using the following pair of recurrences,

$$s(a,b)=s(a,b-1)+I(a,b)$$

$$II(a,b)=II(a-1,b)+s(a,b)$$

where $s(a,b)$ is the cumulative row sum, $s(a,-1)=0$, and $II(-1,b)=0$), the integral frame can be computed in one pass over the original frame. This integral frame is computed for every rectangular part of the images, and it is stored as rectangular features. The dimension of the rectangular feature is usually high, so the feature selection is important to avoid the computational overhead. Here, AdaBoost classifier (Freund & Schapire, 1995) is utilized to select the important features and further to perform the classification task.

ii) Training of AdaBoost classifier: Once the features are extracted, the learning algorithm, called AdaBoost classifier (Freund & Schapire, 1995), is trained based on the positive and negative samples. The training of classifier is performed by updating the weights of every iteration towards reaching the minimum error value. The error values are computed from the original class information with the ground truth label. The step behind the training of classifier is as shown below:

a) Initialization: Initialize weights $w_{1,k}=\frac{1}{2p},\frac{1}{2q}$ for $z_i=0,1$, respectively, where p and q are the negative and positive numbers, respectively.

b) Normalization: Normalize the weights, $w_{t,k}\leftarrow\frac{w_{t,k}}{\sum_{l=1}^{m} w_{t,l}}$.

c) Selection of weights: Select the best weak classifier with respect to the weighted error

$$\in_t=\min_{g,r,\theta}\sum_k w_k\left|J(z_k,g,r,\theta)-z_k\right|.$$

where g_t, r_t and θ_t are the minimizers of $\in_t$.

d) Update the weights: $w_{t+1,k}=w_{t,k}\alpha_t^{1-f_k}$ where $f_k=0$ if example z_k is classified correctly, $f_k=1$ otherwise, and $\alpha_t=\frac{\in_t}{1-\in_t}$.

The above step is repeated for the required number of iterations, and the final weight is selected to perform the classification task. The classification of the input rectangular feature is performed using the following equation.

$$C(z)=\begin{cases}1 & \sum_{t=1}^{T}\beta_t J_t(z)\geq \frac{1}{2}\sum_{t=1}^{T}\beta_t \\ 0 & otherwise\end{cases}$$

where $\alpha_t = \log\frac{1}{\beta_t}$. After performing the above process, face regions are extracted from every frame. The detected face region can be represented as

$$I_{face}=\{I_{ij};1\leq i\leq;1\leq j\leq n\}$$

where $m\times n$ is the size of the face region.

Face super resolution by tangential and exponential kernel weighted regression model

This section represents the tangential and exponential kernel weighted regression model for face super resolution. The process of super resolution is first performed by deriving the multi-kernel matrix based on the weights to be assigned for the neighbouring pixel values. Then, the derived kernel matrix is given for performing the face super resolution. The reason behind the selection of kernel methods is that it is well known and regularly used for pattern detection and discrimination problems (Tao *et al.,* 2012). Even though the kernel regression methods are familiar in data mining works, image and video processing literature did not much utilize these methods. Also, kernel regression is the nonparametric estimation that allows for tailoring the estimation problem to the local characteristics of the data, whereas the standard parametric model is intended as a more global fit. Second, in the estimation of the local structure, higher weight is given to the nearby data as compared to samples that are farther away from the centre of the analysis window. Also, this approach does not specifically require the data to follow a regular or equally spaced sampling structure (Takeda *et al.,* 2007).

A precise kernel is more significant than a sophisticated prior for image super resolution. The one-pass convolution with small kernels is very efficient for reconstruction and restoration, and also it improves the resolution and fidelity [41]. In super resolution, the restoration and reconstruction can be decreased by means of constraining the spatial support of the filter to a small kernel. Besides, small kernel approach capitulates major quantitative and qualitative improvements [43]. The steering kernel regression preserves the restore details with minimal assumptions on noise models and local signal. Moreover, steering kernel weights effectively take local image structures and capture the local signal structures, which include both spatial and temporal edges [42]. The super resolution for image and video restoration in the non-local kernel regression process combines both the local structural regularity and non-local self-similarity. The non-local kernel regression framework is robust, and it is applied to several images and video restoration tasks. In multidimensional kernel regression, each pixel in the video sequence captures the essential local behavior of the spatiotemporal neighborhood. The kernels used in the super resolution context are an alternative to the bicubic kernel for interpolation, and they are used in many algorithms [44]. Moreover, kernels used in many super resolutions play an important role in the final super-resolved HR images. Even though various kernel functions are presented for super resolution, exponential and tangential kernel functions played a major role for face resolution. The exponential kernel is directly associated with the Gaussian kernel, with only the square of the norm being left out, and it is also known as a radial basis function kernel. Generally, the exponential kernel is acquired from the spectral representation with the spectral density. In kernel machine fields, the squared exponential is very smooth, and it is the most widely used kernel [45]. The tangential kernel preserves the singularity within the disc resolved in the shape. The tangential kernel estimates the first and second order derivative factors using the tangential acceleration functions [50], which can easily estimate the neighbour pixels.

a) Generation of multi kernel matrix

The first step in the regression process is to build up the kernel matrix, which does not rely on the sample's location and density. The shape of the regression kernel is square in size, and the size of the kernel fixed is based on the user specified parameters. Here, the size of the kernel is fixed based on the upscaling parameter given by the user. Then, the matrix F is constructed by filling up the distance integer from the centre pixel. Every element in this matrix consists of integers based on the distance from the centre pixel. The matrix F can be represented as follows:

$$F = \{f_{ij}; 0 \le i, j \le v\}$$

The distance-based integer matrix is then given for the kernel model exponential and tangent function to convert the distance values into a kernel space. The exponential kernel function is represented as follows:

$$K_1 = \exp\left(-\left(\frac{1}{2 \times h^2}\right) * F\right)$$

where h is a smoothing factor. The tangent kernel function is represented as follows:

$$K_2 = \tanh\left(-\left(\frac{1}{2 \times h^2}\right) \bullet F\right)$$

These two kernel matrices are effectively combined with the weighted formulae using the parameter called alpha. The hybridization of two kernels is the new contribution proposed in this paper for kernel design. The advantage of the hybridized form of kernel ensures the advantage of exponential and tangent function in estimating the local neighbour values. The hybridized form of the kernel function is given as follows:

$$K = \alpha K_1 + (1 - \alpha) K_2$$

where α is weighted constants

b) Generation of super resolution image

Once the multi-kernel matrix is designed, the super resolution is performed by doing the interpolation with the extracted face image. At first, every pixel belonging to the $I^s(i, j)$ is generated by finding the sub-image of $I^{ne}(a,b)$ corresponding to (i, j) the pixel. The size of this sub-image $I^{ne}(a,b)$ is dependent on the size of the upscaling factor, and the pixels without having the intensities are filled out with the neighbouring values. Once the sub-image is found, the kernel matrix is multiplied with this sub-image and the summation is taken as the representative pixel of the super resolution image, $I^s(i, j)$. The equation used to find the pixel of the super resolution image is given as follows:

$$I^s(i, j) = \frac{1}{k_1 * k_2} \sum_{a=1}^{k_1} \sum_{b=1}^{k_2} I^{ne}(a,b) * K(a,b)$$

where $I^{ne}(a,b) \in (i, j)$; $k(a,b) \in (i, j)$, k_1 and k_2 are the size of the rows and columns of the sub-image.

Algorithmic description:

i. Get the input video V and the input parameters, α, h.

ii. Extract the frames from the input video.

iii. Detect the face using Viola-Jones algorithm, I_{face} which is extracted face region.

iv. Design multi-kernel matrix K using exponential and tangent kernel, α, h.

v. Generate super resolution face image I^s.

vi. Repeat step 3 to 5 for every frame.

RESULTS AND DISCUSSION

This section presents the experimental results and the detailed analysis of the proposed multi-kernel-based regression model of face super resolution.

Experimental setup

The proposed multi-view face video super resolution is implemented using MATLAB, and the performance of the proposed technique and the existing technique is validated using SDME (Panetta *et al.,* 2011). Here, the performance comparison is done with the existing methods like nearest-neighbor interpolation, bicubic interpolation, and bilinear interpolation. For the nearest neighbor interpolation, the block uses the value of the nearby translated pixel values for the output pixel values in the super resolution image. For bilinear interpolation, the block uses the weighted average of two translated pixel values for each output pixel value. Bicubic interpolation is an extension of cubic interpolation for interpolating data points on a two-dimensional regular grid. The interpolated surface is smoother than the corresponding surfaces obtained by the bilinear interpolation or nearest-neighbor interpolation. Bicubic interpolation can be accomplished using Lagrange polynomials, cubic splines, or cubic convolution algorithm.

Dataset description: The experimentation is performed with the face video database available in (SD 2014), which is UCSD face video database. From the database, we have taken two different videos, which have the multi-view of a person.

Evaluation metrics: The definition of SDME (Panetta *et al.,* 2011) is given as follows:

$$SDME = \frac{-1}{b_1 * b_2} \sum_{i=1}^{b_1} \sum_{j=1}^{b_2} 20 \ln \left| \frac{I_s^{\max,i,j} - 2 * I_s^{centre,i,j} + I_s^{\min,i,j}}{I_s^{\max,i,j} + 2 * I_s^{centre,i,j} + I_s^{\min,i,j}} \right|$$

where the super resolution image is divided into b1 × b2 blocks, $I_s^{\max,i,j}$ and $I_s^{\min,i,j}$ are the maximum and minimum values of the pixels in each block separately, and $I_s^{centre,i,j}$ is the intensity of the centre pixel in each block. Thus, the size of the blocks should be composed of an odd number of pixels such as 5 × 5, 7 × 7 or 9 × 9. SDME is the enhanced parameter designed to overcome the disadvantages of the existing measures. The existing measures, such as EME (measure of enhancement), EMEE (measure of enhancement by entropy) (Agaian *et al.* 2001), AME (Michelson law measure of enhancement), and AMEE (Michelson law measure of enhancement by entropy) (Agaian *et al.* 2007), are very sensitive to the noise and steep edges in the images due to the measurement of the maximum and the minimum values of the blocks in the image. Thus, SDME serves as an advanced measure that integrates the idea of the second-derivative-like visibility operator (S. DelMarco and S. Agaian, 2009) with the strengths of the existing measures.

The definition of PSNR (peak signal to noise ratio (PSNR) is given as follows:

$$PSNR = 20 * \log_{10}(MAX_I) - 10 * \log_{10}(MSE)$$

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

where MAX_I is the maximum possible pixel value of the image I ; I and K are the input and output images.

The definition of SSIM (Structural similarity index) is calculated on various windows of an image. The measure between two windows x and y of common size N×N is given as follows:

$$SSIM(x,y) = \frac{(2\mu_x \mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu^2 x + \mu^2 y + c_1)(\sigma^2 x + \sigma^2 y + c_2)}$$

where x and y denote the measurement between the two windows, μ_x is the average of x; μ_y is the average of y; $\sigma^2 x$ is the variance of x ; $\sigma^2 y$ is the variance of y ; σ_{xy} is the covariance of x and y ; c_1 and c_1 are two variables to stabilize the division with weak denominator; L is the dynamic range of the pixel values.

Parameters to be fixed: The proposed face super resolution contains three important parameters, that is, smoothing factor *h*, weighted constant *α*, and upscaling factor *r*. These parameters are extensively analyzed in the following section to find the best parametric value.

Experimental results

This section represents the experimental results of the proposed face super resolution technique. Figure.2 shows the intermediate results of video 1 for the straight position. Here, figure 2.a is the frame extracted from the input video, and figure 2.b shows the face detected from the output of the Viola–Jones algorithm. Figure 2.c is the extracted face region from the frame, and figure 2.d is the outcome of the super resolution for the upscaling factor of three using the methods using the nearest method, bicubic, bilinear, and the proposed multi-kernel-based regression model. Similarly, the frame extracted from video 1 in the side view position is given in figure 3.a. Figure 3.b shows the face detected image through Viola–Jones algorithm, and figure 3.c represents the extracted face region from the frame. Finally, the output image of the face super resolution using the nearest method, bicubic, bilinear, and the proposed multi-kernel-based regression model, is given in figures 3.d, 3.e, 3.f, and 3.g.

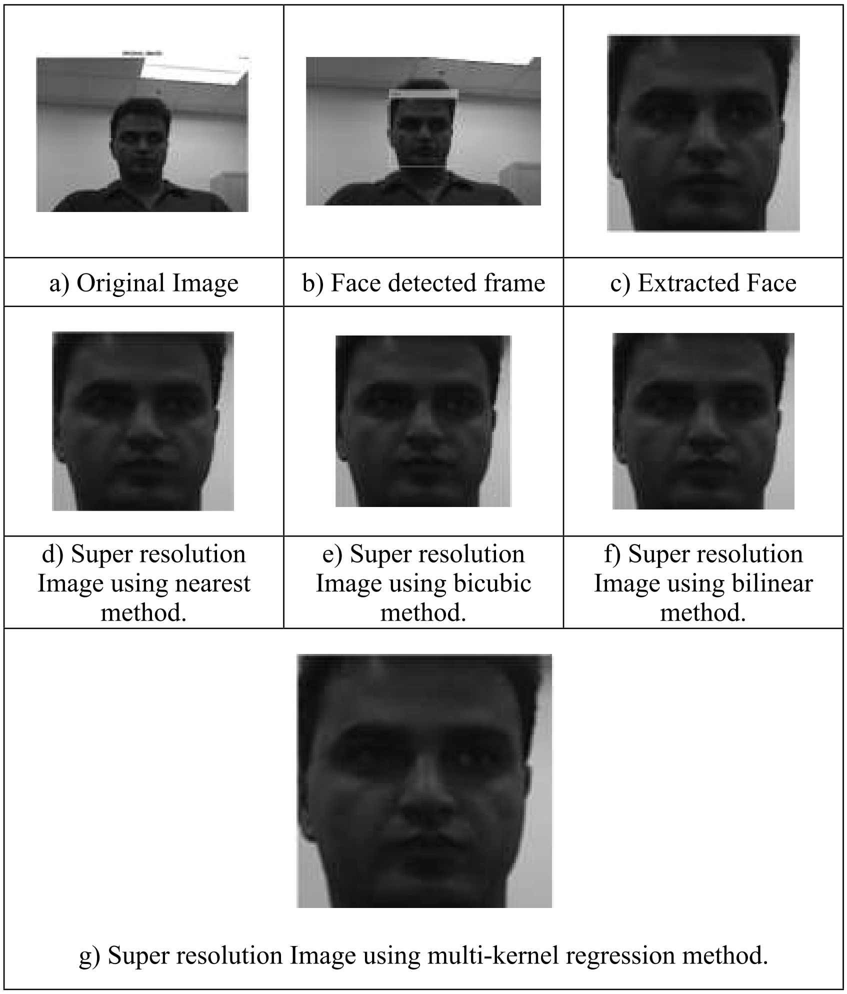


Fig. 2. Intermediate results of video 1 for straight position using the scaling factor, $\delta = 2$.

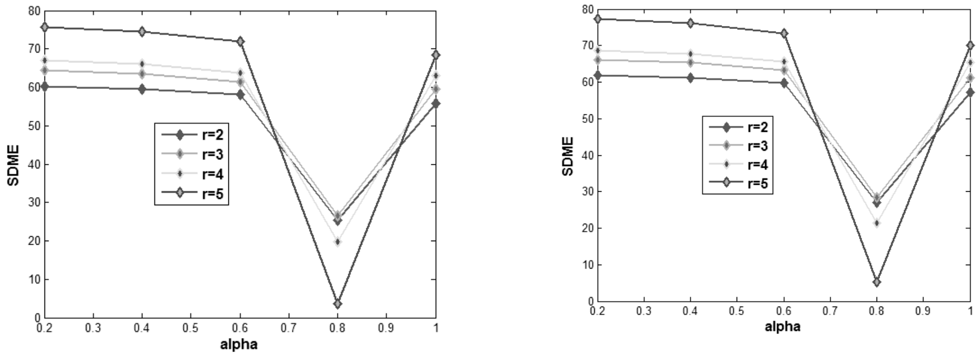
		
a) Original Image	b) Face detected frame	c) Extracted Face
		
d) Super resolution Image using nearest method.	e) Super resolution Image using bicubic method.	f) Super resolution Image using bilinear method.

g) Super resolution Image using multi-kernel regression method.

Fig. 3. Intermediate result of video 2 using the scaling factor, $\delta = 2$.

Performance evaluation

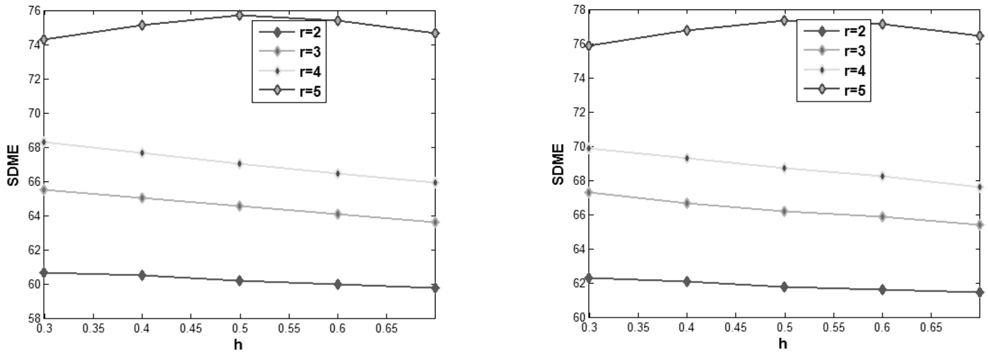
This section presents the performance evaluation of the proposed multi-kernel regression model. Here, two parameters related to the proposed technique are analyzed to identify the best value for comparative analysis. The first parameter, called alpha, is taken, and the values are changed from 0.2 to 1 to find the optimal value. Figure 4.a shows the performance analysis of the proposed face super resolution method based on alpha for video 1. Here, the upscaling factor is fixed from 2 to 6, and the results are evaluated. When analyzing Figure 4, we understand that the better performance is achieved when the upscaling factor is increased. For the upscaling factor of five, the proposed technique obtained the SDME value of 75.7db. Also, for the upscaling factor of two, the proposed technique obtained the SDME value of 60.18db. A similar type of behaviour can be seen in Figure 4.b. From Figure 4.b, the proposed technique obtained the maximum SDME when the upscaling factor is fixed to five. Until the alpha is equivalent to 0.8, the SDME value is decreased when the alpha is increased. After that, the performance is increased. For the alpha value of 1, the proposed technique obtains the SDME value of 70.01db for the upscaling factor of five.



(a) (b)

Fig. 4. Performance analysis based on alpha, a) video 1, b) video 2.

Figure 5.a shows the performance analysis of the proposed technique for the global smoothing factor of video 1. From the results, we understand that when the smoothing factor is increased, the performance is also increased until a specific value. After that, the performance is decreased even though the smoothing factor has increased. From the figure, we have seen that the performance in terms of SDME is increased until the factor is equal to five for the upscaling factor of five. After that, the performance is decreased slightly. So, the global smoothing factor can be fixed to five to get the maximum results when the upscaling is equal to five. Similarly, the performance in terms of SDME is plotted in figure 5.b. From the figure, we understand that the performance behaviour is almost similar in both videos. The maximum performance is achieved when the upscaling factor is equal to five and the smoothing factor is equal to 0.5.



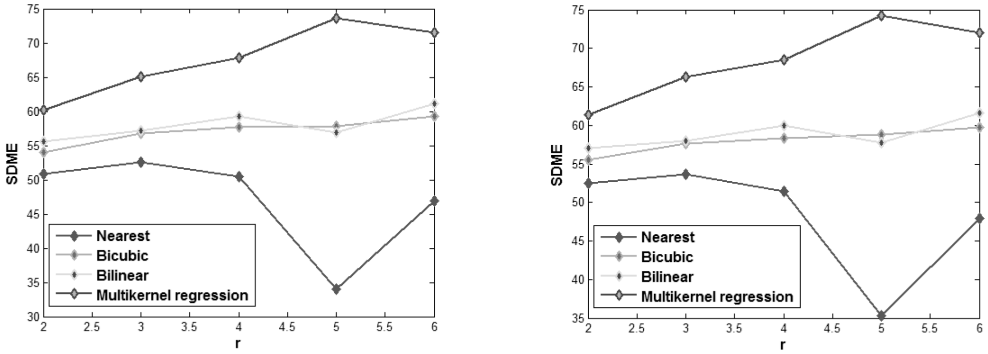
(a) (b)

Fig. 5. Performance analysis based on global smoothing factor, a) video 1, b) video 2.

Comparative analysis

This section discusses the comparative analysis of the proposed regression model for super resolution with the existing methods, that is, nearest interpolation, bicubic interpolation, and bilinear interpolation. Figure 6.a shows the performance of the proposed technique with the existing methods for various values of an upscaling factor. For all the different values of the upscaling factor, the proposed technique shows better performance as compared to the existing technique. When the upscaling factor is fixed to 2, the SDME of the nearest interpolation, bicubic interpolation, bilinear interpolation, and proposed technique, is 50.8 dB, 54 dB, 55.5 dB, and 60.19 dB. Similarly, the performance of SDME for those techniques for the upscaling factor of six is 46.96 dB, 59.3 dB, 61.1 dB, and 71.57 dB. The performance of SDME in video 2 is plotted in figure 6.b. The similar kind of behaviour can be found in video 2 also. The better performance of 74.2 dB is achieved for the proposed technique when the upscaling factor is fixed to five.

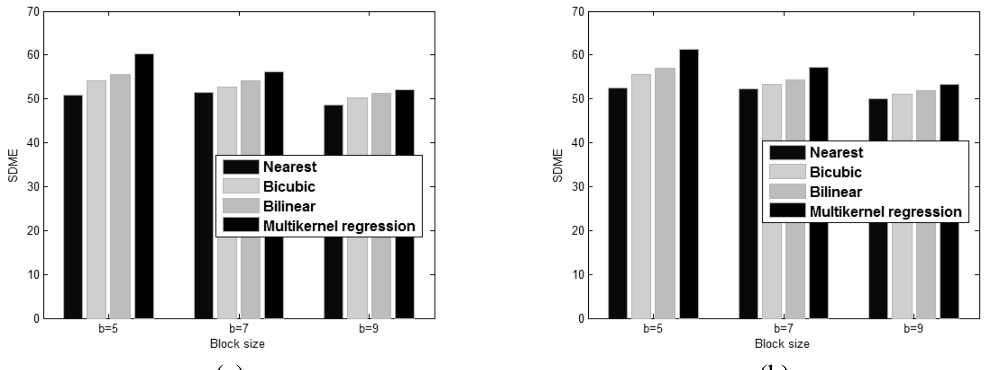
The figure shows that the performance of the techniques is increased when the upscaling factor is increased until it reaches a specific value. After that, the performance is decreased even though the upscaling factor is increased.



(a) (b)

Fig. 6. Comparative analysis based on upscaling factor, a) video 1, b) video 2.

Figure 7 shows the comparative analysis of four different techniques for the different size of blocks, which is the parameters in SDME. The block size is varied from five to nine, and the results are analyzed. For the block size of five, the SDME of the nearest interpolation, bicubic interpolation, bilinear interpolation, and proposed technique, is 50.8 dB, 54 db, 55.5 dB, and 60.19 dB. Similarly, the nearest interpolation, bicubic interpolation, bilinear interpolation, and proposed technique, obtained the SDME of 51.48 dB, 52.76 dB, 54.08 dB, and 56.2 dB when the block size is equal to seven. The better performance in block size of nine for the proposed technique is 52.17 dB, which is higher than all the existing techniques. Similarly, for video 2, the performance is analyzed for all the techniques using SDME in Figure 7.b. From the figure, the nearest interpolation, bicubic interpolation, bilinear interpolation, and proposed technique, obtained the SDME of 52.49 dB, 55.5 dB, 57.08 dB, and 61.39 dB when the block size is equal to five. Overall, the proposed technique outperformed the existing techniques for all different sizes of blocks.



(a) (b)

Fig. 7. Comparative analysis based on block size, a) video 1, b) video 2.

Comparative analysis with existing works

Table II illustrates the PSNR, SSIM, and SDME for different methods with the sampling factors 2, 3, 4, 5, and 6 of video 1. Here, the upsampling factor is fixed from 2 to 6, and the results are evaluated. This section analyzes the performance of the proposed method with the existing works, that is, J. Yang *et al.* (2008), A. Marquina and S. J. Osher (2008), Weisheng Dong *et al.* (2011), and Victor may *et al.* (2016). Here, video 1 and video 2 are taken from UCSD video database. For the upsampling factor of 6, the PSNR, SSIM, and SDME values for the proposed multi-kernel regression model are 42.5573 dB, 0.9612, and 73.71 dB. Similarly, PSNR, SSIM, and SDME values for J. Yang *et al.* at upsampling factor 6 is 32.12dB, 0.8221, and 65.8dB. From Table II, we clearly understand that our proposed method has greater PSNR, SSIM, and SDME values at all sampling factors.

Table III indicates the PSNR, SSIM, and SDME for video 2 of different methods with different upsampling factors 2, 3, 4, 5, and 6 of video 2. For the upsampling factor of 2, PSNR, SSIM, and SDME values for the proposed multi-kernel regression method are 47.12dB, 0.9819, and 75.32dB for upsampling factor of 2, whereas for the other existing methods, the values of PSNR, SSIM, and SDME are less. From table III, we clearly understand that our proposed method has greater PSNR, SSIM, and SDME values at all the sampling factors.

Table 2. Performance analysis based on upsampling factor of video 1.

	Upsampling factor	Multi-kernel regression	J. Yang, J. Wright *et al.* (2008)	A. Marquina and S. J. Osher (2008)	Weisheng Dong *et al.* (2011)	Victor may *et al.* (2016)	Ce Liu and Deqing Sun (2011)	Kappeler *et al.* (2016)
PSNR (dB)	2	47.3510	34.12	29.12	30.92	21.31	27.11	31.18
	3	41.2137	34.00	29.10	30.54	21.03	27.01	31.02
	4	42.0937	33.98	28.99	30.27	20.75	26.98	31.00
	5	41.8749	32.96	28.68	30.06	20.11	26.87	30.94
	6	42.5573	32.12	28.51	29.52	20.01	26.81	30.87
SSIM	2	0.9251	0.9122	0.9	0.8936	0.794	0.842	0.8951
	3	0.9264	0.900	0.87	0.8871	0.7521	0.835	0.8864
	4	0.9666	0.8997	0.82	0.8725	0.7495	0.825	0.8766
	5	0.9623	0.8494	0.84	0.8233	0.7255	0.812	0.8523
	6	0.9612	0.8221	0.82	0.8111	0.71	0.801	0.8112
SDME (dB)	2	74.56	68.9	69.1	71.2	74.56	74.35	73.86
	3	74.35	67.5	69.8	70.8	67.5	73.86	67.2
	4	74.01	66.5	68.2	70.7	66.2	73.86	67.86
	5	73.86	66.4	67.2	69.8	65.15	69.15	67.2
	6	73.71	65.8	67.15	68.5	64.2	68.2	66.2
Time(in sec)	-	50	121	131	100.5	59.5	57.8	61.1

Table 3. Performance analysis based on upsampling factor of video 2.

	Upsampling factor	Multi-kernel regression	J. Yang, J. Wright *et al.* (2008)	A. Marquina and S. J. Osher (2008)	Weisheng Dong *et al.* (2011)	Victor may *et al.* (2016)	Ce Liu and Deqing Sun (2011)	Kappeler *et al.* (2016)
PSNR (dB)	2	47.12	33.12	29.12	30.92	21.92	26.99	31.08
	3	46.1235	33	29.02	30.67	21.52	26.21	30.97
	4	45.8874	32.68	28.99	29.65	21.02	26.01	30.54
	5	43.8577	32.24	28.14	29.32	20.63	25.65	30.01
	6	42.1247	31.87	28.02	28.96	19.15	25.24	29.58
SSIM	2	0.9899	0.9623	0.9512	0.9101	0.8857	0.833	0.8858
	3	0.9455	0.9421	0.9412	0.9201	0.8824	0.821	0.8721
	4	0.8965	0.9320	0.9401	0.8935	0.7891	0.801	0.801
	5	0.8688	0.9122	0.9322	0.8962	0.7718	0.799	0.7891
	6	0.8654	0.899	0.9222	0.8721	0.7611	0.769	0.7711
SDME (dB)	2	75.32	68.19	59.1	61.27	74.56	64.312	63.86
	3	75.12	67.45	59.012	60.81	67.5	63.812	57.211
	4	74.32	66.55	58.21	60.97	66.2	63.834	57.823
	5	72.18	66.44	57.28	60.82	65.15	62.154	57.221
	6	71.11	65.238	57.157	58.51	64.2	61.56	56.223
Time(in sec)	-	53.1	113.1	122.3	105.5	69.5	77.9	85.5

Tables II and III present the computational time of all the SR methods discussed in the paper. The time complexity of the proposed method is less when compared with all other methods. For video 1, the computational time is 50 secs for the proposed multi-kernel regression model, whereas for the existing methods, the computational time is greater. Similarly, the computational time of the proposed method for video 2 is 53.1 secs, which is less than that of the existing methods.

CONCLUSION

This paper presented a tangential and exponential kernel weighted regression model for multi-view face video super resolution. The ultimate contribution we made in this work is to design a hybrid kernel for the designing of the regression model. The hybrid kernel utilizes the exponential and tangent functions with the weighted parameters. The designed hybrid kernel is then utilized to perform the super resolution of the face regions, which is extracted from the input video using Viola-Jones algorithm. The proposed face super resolution is experimented with UCSD face video dataset, and the performance is analyzed with the help of SDME. Also, the detailed parametric analysis is carried out to find the better value to perform a comparative analysis with the existing techniques, that is, nearest interpolation, bicubic interpolation, and bilinear interpolation. From the results, we proved that the proposed techniques obtained the maximum SDME of 77.3db as compared to the existing techniques. In future, the optimization algorithms can be included to steer the regression model for better performance. The proposed method is applicable to the video-based face super resolution but the future work concentrates on improving the SR performance using the spatial temporal information of consecutive video frames.

REFERENCES

Qu, S., Hu, R., Chen, S., Chen, L. & Zhang, M. 2014. Robust face super-resolution via position-patch neighborhood preserving, *in proceedings of IEEE International Conference on Multimedia and Expo Workshops (ICMEW),* 1–5.

Baker, S. & Kanade T. 2000. Hallucinating faces, In: *IEEE international conference on automatic face and gesture recognition,* 83–88.

Li, X., Hu, Y., Gao, X., Tao, D. & Ning, B. 2010. A multi-frame image superresolution method," Signal Process., **90**: 405–414.

Kim, K.I. & Kown, Y. 2010. Single-image super-resolution using sparse regression and natural image prior," IEEE Trans. Pattern Anal. Mach.Intell., **32**(6): 127–1133.

Zhuang, Y., Zhang, J. & Wu, F. 2007. Hallucinating faces: LPH super-resolution and neighbor reconstruction for residue compensation, *Pattern Recogn,* **40**: 3178–94.

Chang, H., Yeung, D. & Xiong, Y. 2004. Super-resolution through neighbor embedding, *In: IEEE computer society conference on computer vision and pattern recognition*, 275–282.

Park, J. & Lee, S. 2008. An example-based face hallucination method for single-frame, low-resolution facial images, *IEEE Transaction on Image Processing,* **17**: 1806–1816.

Farsiu, S., Robinson, M., Elad, M. & Milanfar, P. 2004. Fast and robust multiframe super-resolution, *IEEE Transaction on Image Processing,* **13**: 1327–1344.

Lu, T., Hu, R., Jiang, J. & Wei H.Y.Z. 2013a. Super-Resolution for Surveillance Facial Images via Shape Prior and Residue Compensation, *International Journal of Multimedia and Ubiquitous Engineering,* **8**(6): 47-58.

Tao, L., Ruimin, H., Zhen, H., Yang, X. & Shang, G. 2012. Surveillance Face Super-Resolution via Shape Clustering and Subspace Learning, *International Journal of Signal Processing, Image Processing and Pattern Recognition,* **5**(4).

Jiang, J., Hu, R., Wang, Z. & Han, Z. 2014. Noise Robust Face Hallucination via Locality-Constrained Representation, *IEEE transactions on multimedia,* **16**(5): 1268-1281.

Takeda, H., Farsiu, S. & Milanfar, P. 2007. Kernel Regression for Image Processing and Reconstruction, *IEEE transactions on image processing*, **16**(2): 349-366.

Wand, M.P. & Jones, M.C. 1995. Kernel Smoothing, ser. Monographs on Statistics and Applied Probability. *New York: Chapman & Hall.*

Yee, P. & Haykin, S. 1993. Pattern classification as an ill-posed, inverse problem: a regularization approach, *in Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing,* **1993**(1): 597–600.

Hu, Y., Lam, K.M., Qiu, G. & Shen, T. 2010. From Local Pixel Structure to Global Image Super-Resolution: A New Face Hallucination Framework, *IEEE Transactions on Image Processing,* **20**:(2): 433 – 445.

Wang, X., Lin, H. & Xu, X. 2014. Parts-based face super-resolution via non-negative matrix factorization, *Computers and Electrical Engineering,* **40**(8): 130–141.

Lu, T., Hu, R., Han, Z., Jiang, J. & Zhang, Y. 2013b. From Local Representation to Global Face Hallucination: A Novel Super-Resolution Method By Nonnegative Feature Transformation, *in proceedings of Visual Communications and Image Processing (VCIP),* 1 – 6.

Zeng, X. & Huang, H. 2012. Super-Resolution Method for Multiview Face Recognition from a Single Image Per Person Using Nonlinear Mappings on Coherent Features, *IEEE signal processing letters,* **19**(4): 195-198.

Ma, X., Song, H. & Qian, X. 2015. Robust Framework of Single-Frame Face Superresolution Across Head Pose, Facial Expression, and Illumination Variations, *IEEE transactions on human-machine systems,* **45**(2): 238-250.

viola, P. & jones, M.J. 2004. Robust Real-Time Face Detection, *International Journal of Computer Vision*, **57**(2): 137–154.

Fleuret, F. & Geman, D. 2001. Coarse-to-fine face detection, Int.*J. Computer Vision,* **41**: 85–107.

Osuna, E., Freund, R. & Girosi, F. 1997. Training support vector machines: An application to face detection, *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.*

Roth, D., Yang, M. & Ahuja, N. 2000. A snow based face detector, *In Neural Information Processing,* 12.

Freund, Y. & Schapire, R.E. 1995. A decision-theoretic generalization of on-line learning and an application to boosting, *In Computational Learning Theory: Eurocolt 95, Springer-Verlag,* 23–37.

Panetta, K., Zhou, Y., Agaian, S & Jia, H. 2011. Nonlinear Unsharp Masking for Mammogram Enhancement. *IEEE transactions on information technology in biomedicine,* **15**(6): 918-928.

SD,U.C. 2014. The UCSD face video Database from, "http://vision.ucsd.edu/datasets/leekc/disk2/VideoDatabase/testing/".

J. Jiang, X. Ma, C. Chen, T. Lu & R. Hu. 2016. Single Image Super-Resolution via Locally Regularized Anchored Neighborhood Regression and Nonlocal Means. *IEEE Transactions on Multimedia*.

J. Jiang, J. Ma, C. Chen, Z. Cai & R. Hu. 2016. Noise Robust Face Image Super-resolution through Smooth Sparse Representation. *IEEE Transactions on Cybernetics*.

Z. Zhu, F. Guo, H. Yu & C. Chen. 2014. Fast Single Image Super-Resolution via Self-Example Learning and Sparse Representation. *IEEE Transactions on Multimedia*, **16**(8):2178-2190.

J. Jiang, C. Chen, J. Fu & R. Hu. 2016. SRLSP: A Face Image Super-Resolution Algorithm using Smooth Regression with Local Structure Prior. *IEEE Transactions on Multimedia*.

J. Jiang, C. Chen, K. Huang, Z. Cai & R. Hu. 2016. Noise Robust Position-Patch based Face Super-Resolution via Tikhonov Regularized Neighbor Representation. *Information Sciences*. **367-368**: 354-372.

S. S. Agaian, K. Panetta & A. M. Grigoryan. 2001. Transform-based image enhancement algorithms with performance measure. *IEEE Transactions on Image Processing*. **10**(3):367–382.

S. S. Agaian, B. Silver & K. A. Panetta. 2007. Transform coefficient histogram-based image enhancement algorithms using contrast entropy. *IEEE Transactions on Image Processing*. **16**(3):741–758.

S. DelMarco &S. Agaian. 2009. The design of wavelets for image enhancement and target detection. *in proceedings of the Mobile Multimedia/Image Processing and Security applications*. pp. 735103-1–735103-12.

Armin Kappeler, Seunghwan Yoo, Qiqin Dai & Aggelos K. Katsaggelos. 2016. Video Super-Resolution with Convolutional Neural Networks. *IEEE Transactions on Computational Imaging.* **2**(2).

Ce Liu & Deqing Sun. 2011. A Bayesian approach to adaptive video super resolution. *in proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).*

J. Yang, J. Wright, Y. Ma & T. Huang. 2008. Image super-resolution as sparse representation of raw image patches, *IEEE Computer Vision and Pattern Recognition*.

A. Marquina & S. J. Osher. 2008. Image super-resolution by TV-regularization and Bregman iteration. *Journal of Science and Computing*. **37**:367-382.

W. Dong, L. Zhang, G. Shi & X. Wu. 2011. Image Deblurring and Super-Resolution by Adaptive Sparse Domain Selection and Adaptive Regularization. *IEEE Transactions on Image Processing*, **20**(7): 1838-1857.

V. May, Y. Keller, N. Sharon & Y. Shkolnisky. 2016. An Algorithm for Improving Non-Local Means Operators via Low-Rank Approximation. *IEEE Transactions on Image Processing*. **25**(3):1340-1353.

Submitted: 17/12/2016
Revised: 03/06/2017
Accepted: 17/07/2017

نموذج انحدار لنواة موزونة أسية وملموسة من خلال عرض فيديو فائق الدقة لوجوه بزوايا دوران متعددة

***عمار ب. دشمخ و** إن. أوشا راني**
*جامعة فيجنان، جونتور، أندرا براديش 522213.
** رئيس قسم ECE، جامعة فيجنان، جونتور، أندرا براديش 522213.

الخلاصة

يُعتبر التعرف على الشخص من صورة غير أمامية منخفضة الدقة مشكلة صعبة أثناء المراقبة عن طريق الفيديو. ولحل تلك المسألة، قدمت البحوث تقنيات مختلفة للتعرف على الوجوه بعد تحويل الصورة منخفضة الدقة إلى صورة عالية الدقة. وبناءً على ذلك، يقدم هذا البحث تقنية لفيديو عالي الدقة لعرض وجوه بزوايا دوران متعددة باستخدام نموذج انحدار النواة الموزونة الأسية والملموسة. تم اقتراح نواة هجينة جديدة لنموذج انحدار نواة لا معلميه لتقدير بكسل الجوار في الدقة الفائقة بعد تمييز الوجه باستخدام خوارزميات فيولا جونز. تم إجراء التجارب باستخدام قواعد بيانات فيديو الوجه UCSD وتم تحليل النتائج الكمية باستخدام SDME مع التقنيات الحالية. وقد أثبتت النتائج أن التقنية المقترحة حصلت على الحد الأقصى 77.3db بالمقارنة مع التقنيات الحالية مثل، الاستكمال الأقرب، الاستكمال ثنائي التكعيبية والاستكمال ثنائي الخطية.