Video Noise Removal Method using Spatial Filtering Approach and Super Resolution Algorithm

Aarthi Rajagopalan[1], Aravinda R[2], Jayashree B[3], Mr.S.Balasundaram[4]

[1,2,3]Students, ECE Department, Meenakshi Sundararajan Engineering College, Tamilnadu, India
[4]Assistant Professor, ECE Department, Meenakshi Sundararajan Engineering College, Tamilnadu, India

---***---

Abstract - *Video denoising is an important field in digital image/video processing. Our work proposes a system to denoise the frames of a video and improve its resolution using MATLAB 2017 software. Video denoising is performed using spatial filtering and the resolution is enhanced by applying a super-resolution framework to single images. The obtained high resolution frames are then combined to form the output video. First, the input video is divided into frames. Salt and pepper noise is added to each of the frames. Then two filters namely, median filter and weighted average filter are applied to the frames to remove the added noise. The denoised frames are then given as input to the super resolution framework, entering which they are interpolated using bicubic interpolation. The proposed super resolution algorithm uses Single Image Super Resolution (SISR) technique, which recovers one high resolution image from a single low resolution image. The Super Resolution framework comprises of an Artifacts Convolutional Neural Network (ARCNN), to which the bicubic interpolated frames are provided as input, one at a time. ARCNN is used to reduce the blur artifact. Discrete Wavelet Transform (DWT) is applied to the extracted feature maps to obtain the approximation and detail coefficients matrices. Finally, Inverse Discrete Wavelet Transform (IDWT) is applied to obtain high resolution frames. Peak Signal to Noise Ratio (PSNR) and Root Mean Square Error (RMSE) are the parameters used to estimate the quality of output video.*

Key Words: Video denoising, spatial filtering, super-resolution framework, SISR, ARCNN, DWT, IDWT, PSNR, RMSE

1. INTRODUCTION

Video denoising basically deals with noise removal in a video. Noise is always present in digital images during image acquisition, coding, transmission, and processing steps. In the presence of noise, video processing, image analysis, and tracking, are adversely affected due to distortion and loss of image information. Hence video denoising plays an important role in the field of digital image/video processing. Video denoising methods are designed and tuned for specific types of noise like analog noise, digital noise, and film artifacts. The main purpose of the noise reduction algorithms is to restore the image from the degraded version of the original image. The most efficient algorithm is that which has the ability to yield image as close as possible to the original image. In short, meaningful information is recovered from noisy images in the process of noise removal in order to obtain high quality images. These images are then combined to reconstruct the original video. The noise reduction algorithms depend on the noise models. Noise models are classified into two types: additive and multiplicative noise model. Most of the natural video frames are assumed to have additive random noise which is modeled as a Gaussian. In addition to this, there are other noises (that are also modeled) which greatly degrade the video frames like Salt and Pepper noise, Poisson noise and Speckle noise [1]. This paper proposes a video noise removal method using spatial filtering approach and super resolution algorithm.

2. PROPOSED SYSTEM

The proposed system uses spatial filtering to minimize the effect of noise in a video. The super resolution algorithm is applied in order to enhance the resolution of the resultant noise-free video frames.

The proposed spatial filtering approach employs median filter and weighted average filter. The super resolution algorithm revolves around single image super resolution using Artifacts Reduction Convolutional Neural Network (ARCNN). ARCNN is an improved super resolution model as it reduces the undesirable noisy patterns in reconstruction.

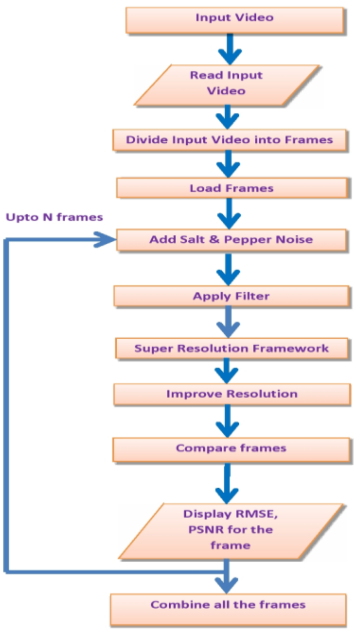

Fig-1: Flow chart representing the proposed system

2.1 SPATIAL FILTERING

Filtering is a technique for enhancing an image. For example, one can filter an image to highlight certain features or remove other features. Smoothing, sharpening and edge enhancement are some of the image processing operations that can be implemented with filtering. Spatial filtering is an image processing technique for changing the intensities of a pixel according to the intensities of the neighbouring pixels. The corresponding process is to convolve the input image I(i,j) with the filter function H(x,y) to produce new filtered image. To perform spatial filtering, we make use of two filters, one is the median filter and the other one is the weighted average filter. Median filter is a standard and a default filter that is used in most of the filtering techniques. This filter is mainly used to remove the noise caused due to sharp and sudden disturbances present in an image. This type of noise is known as Salt and pepper noise. The noise that we add to the frames in order to assess the performance of our proposed system is salt and pepper noise. Weighted average filter is used to control the blurring of an image. Spatial filters are more versatile as they are used in linear as well as non-linear filtering. Linear filtering is a technique in which the output pixel is a linear combination of the pixel values in the input pixel's neighbourhood. On the other hand, if the operation performed on the image pixels is non-linear, it is called as non-linear filtering.

2.1.1 Types

• Point to point (pixel to pixel) operations

• Mask based (Neighborhood) operations

1. Operation with 3*3 filter (E.g. Mean, max, min, etc.)

2. Correlation or Convolution

Point to point operations

Such operations are known as gray level transformations. All image processing techniques focus on gray level transformation as it operates directly on pixels. The gray level image involves 256 levels of gray and in a histogram, horizontal axis spans from 0 to 255, and the vertical axis depends on the number of pixels in the image. Some examples of point to point operations are: Gamma correction, Window-center correction, Histogram equalization.

Mask based (Neighbourhood) operations

The neighbourhood operations work with the values of the image pixels in the neighbourhood and the corresponding values of a sub-image that has the same dimension as neighbourhood. The sub-image is called as filter, mask, template, kernel or window. The values in a filter sub-image are referred to as coefficients, rather than pixels. The process consists simply of moving the filter mask from point to point in an image. At each point (x, y), the response of the filter at that point is calculated using a predefined relationship. For linear spatial filtering the response is given by a sum of products of the filter coefficients and the corresponding image pixels in the area spanned by the filter mask.

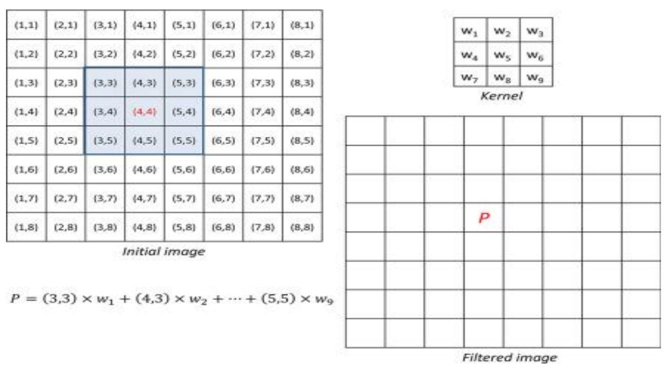

Fig-2: Illustration of Mask-based (Neighbourhood) operations on pixels

Convolution

Linear filtering of an image is achieved through an operation called convolution. It is a neighbourhood operation in which each output pixel is the weighted sum of the values of neighbouring input pixels. The matrix of weights is called a convolution kernel, which is otherwise known as filter. A convolution kernel is nothing but a correlation kernel rotated through 180 degrees.

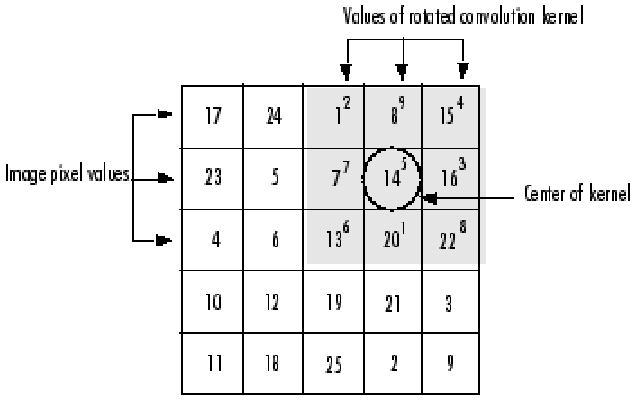

Fig-3: Calculating the (2,4) output of convolution

Correlation

Correlation is similar to convolution. In correlation, the value of output pixel is also calculated as a weighted sum of the neighbouring pixels. The only difference is that the correlation kernel is not rotated during the computation.

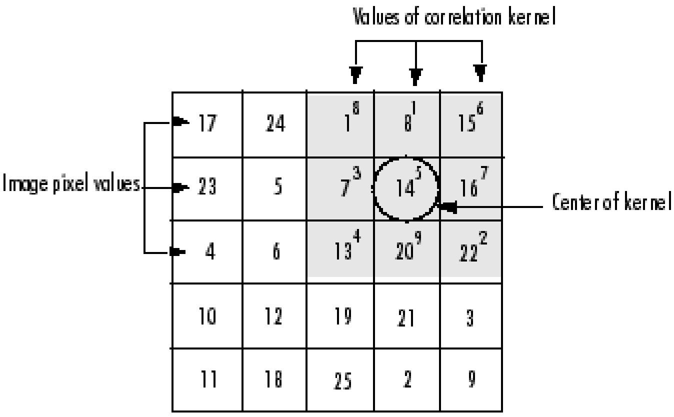

Fig-4: Calculating the (2,4) output of correlation

2.1.2 SMOOTHING SPATIAL FILTERS

Smoothing filters are used for blurring and for noise reduction. Blurring is used in preprocessing steps, such as removal of small details from an image prior to (large) object extraction, and bridging of small gaps in lines or curves. Noise reduction can be accomplished by blurring with a linear filter and also by nonlinear filtering. The main idea is to replace the value of every pixel in an image by the average of the gray levels in the neighborhood defined by the filter mask. This process results in an image with reduced "sharp" transitions in gray levels. Such filters reduce noise from the image but blur its edges.

Median filter

Median filtering is a nonlinear method used to remove noise from images. It is widely used as it is very effective at removing noise while preserving edges. It is particularly effective at removing 'salt and pepper' type noise. The median filter works by moving through the image pixel by pixel, replacing each value with the median value of neighbouring pixels. The pattern of neighbours is called the "window", which slides, pixel by pixel over the entire image. The median is calculated by first sorting all the pixel values from the window in numerical order, and then replacing the pixel being considered with the middle (median) pixel value.

A single very unrepresentative pixel in a neighbourhood will not affect the median value significantly. The median value must actually be the value of one of the pixels in the neighbourhood, so the median filter does not create new unrealistic pixel values when the filter straddles an edge. Hence the median filter is good at preserving sharp edges.

Averaging filter

Average (or mean) filtering is a method of 'smoothing' images by reducing the amount of intensity variation between neighbouring pixels. The average filter works by moving through the image pixel by pixel, replacing each value with the average value of neighbouring pixels, including itself. There are two types, which are as follows:

- Box filter
- Weighted average filter

Weighted Average filter

In weighted average filter, more weight is given to the center value, due to which the contribution of center becomes more than the rest of the values. Due to weighted average filtering, the blurring of image can be controlled.

$$\frac{1}{16} \times \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$$

Fig-5: An example of Weighted Average filter

2.2 SUPER RESOLUTION ALGORITHM

Super resolution is the process of creating high-resolution images from low-resolution images. In our proposed system, single image super resolution (SISR) is used, wherein one high-resolution image is recovered from one low-resolution image. Super resolution is used to increase the high frequency components and removing the degradations caused by the imaging process of the low resolution camera.

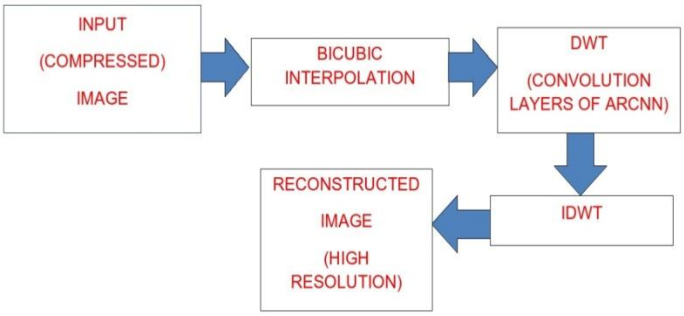

Fig-6: Block diagram representing the proposed super resolution algorithm

2.2.1 BICUBIC INTERPOLATION

Our proposed system uses bicubic interpolation to implement the single image super resolution algorithm. Bicubic interpolation goes one step beyond bilinear interpolation by considering the closest 4x4 neighborhood of known pixels for a total of 16 pixels. Since these are at various distances from the unknown pixel, closer pixels are given a higher weighting in the calculation [2]. In image processing, bicubic interpolation is often chosen over bilinear or nearest-neighbour interpolation for image resampling, when speed is not an issue. Bicubic interpolation produces noticeably sharper images than the nearest neighbour interpolation and bilinear interpolation methods.

2.2.2 ARTIFACTS REDUCTION CONVOLUTIONAL NEURAL NETWORKS (ARCNN):

Deep learning or convolutional neural network (CNN) is usually used for image classification. There are convolutional neural networks which perform specific tasks like super resolution. One such network is ARCNN which not only enhances the resolution but also reduces undesired image artifacts as follows:

- Blocking Artifacts
- Ringing Artifacts along the sharp edges
- Blurring

In our proposed system, we use ARCNN to reduce blurring, which is caused due to the attenuation of the high spatial frequencies during spatial filtering.

Framework of ARCNN

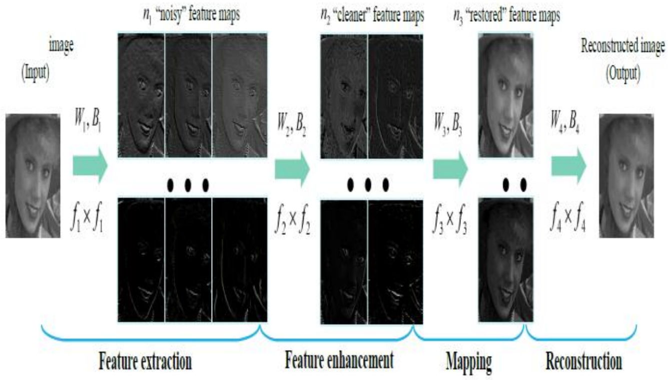

Fig-7: The ARCNN framework

The network comprises four convolutional layers, each of which is used to perform a specific operation. The four operations are:

1. Feature extraction
2. Feature enhancement
3. Mapping
4. Reconstruction

These four operations are optimized jointly in an end-to-end framework. The convolved features from each and every layer are called as feature maps. The convolutional layers play a vital role in detecting the various patterns in the image.

Feature extraction

This operation is similar to the patch extraction and representation operation in Super Resolution Convolutional Neural Network (SRCNN), in which the (overlapping) patches are extracted from a low-resolution image Y and each of them is represented as a high-dimensional vector. These vectors constitute a set of feature maps, of which the number equals to the dimensionality of the vectors.

First the interpolated image is provided as input to the ARCNN. Filters of size $f_1 \times f_1$ are applied to the input image for convolution. Features (such as edges, lines, corners, objects) of the input image are extracted in the form of feature maps (convolved features). The number of feature maps obtained is equivalent to the number of filters used.

Feature enhancement

To conduct feature enhancement, new features are extracted from the feature maps of the first layer, and then combined to form another set of feature maps. This operation can also be formulated as a convolutional layer.

This layer maps the “noisy” features to a relatively “cleaner” feature space, which is equivalent to denoising the feature maps. Due to this additional layer, ARCNN is a new and deeper network in comparison with SRCNN.

Mapping

In this operation, each of the vectors (feature maps) of the second layer is mapped non-linearly into another one. Each mapped vector is conceptually the representation of a high-resolution patch. These vectors constitute another set of feature maps.

This is equivalent to applying filters having a trivial spatial support 1×1. The 1×1 convolution introduces non-linearity to improve the accuracy. Using 1×1 ($f_3 \times f_3$) filters enable dimensionality reduction, decreasing the number of feature maps whilst retaining their salient features.

Reconstruction

This operation aggregates the above high-resolution patch-wise representations to generate the final high-resolution image. Usually, a convolutional layer is defined to produce the final high-resolution image similar to SRCNN.

However, in our proposed system, this reconstruction operation is not defined by a convolutional layer. Instead, Inverse Discrete Wavelet Transform is used to obtain the high-resolution output. In order to implement this, Discrete Wavelet Transform should first be applied to the feature maps obtained from the first three convolutional layers of ARCNN.

2.2.3 DISCRETE WAVELET TRANSFORM

Wavelets provide frequency information, space localization as well as high frequency details in an image viz. the horizontal, vertical and diagonal detail [3]. Wavelets are mathematical functions that split the data into different frequency components, after which each component is studied with a resolution matched to its scale.

The transform of a signal is just another form of representing the signal. It does not change the information content present in the signal. The Wavelet Transform provides a time-frequency representation of the signal. In numerical analysis and functional analysis, a discrete wavelet transform (DWT) is any wavelet transform for which the wavelets are discretely sampled. A wavelet, in the sense of the Discrete Wavelet Transform (or DWT), is an orthogonal function which can be applied to a finite group of data.

In DWT, the most prominent information in the signal appears in high amplitudes and the less prominent information appears in very low amplitudes. Data compression can be achieved by eliminating these low amplitudes. High compression ratios with good quality of reconstruction can be obtained by the application of wavelet transforms. The discrete wavelet transform employs low-pass and high-pass filters, h(n) and g(n) respectively, in order to expand a digital signal. These are known as analysis filters.

The coefficients c_k and d_k are obtained by convolving the digital signal with each filter. The output is then decimated. The c_k coefficients are obtained by the application of the low-pass filter, h(n), and are termed as coarse coefficients. The d_k coefficients are obtained by the application of the high-pass filter, g(n), and are termed as detail coefficients. Low frequency information is provided with the help of coarse coefficients, while detail coefficients provide high-frequency information. A tree-structured filter bank, known as analysis filter bank, can be used to describe the entire process. This can be illustrated with the help of the following diagram. The high and low pass filters divide the signal into a series of coarse and detail coefficients.

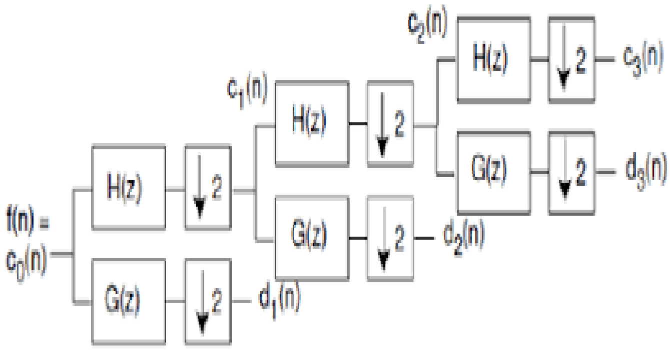

Fig-8: Analysis filter bank

After analyzing, or processing the signal in the wavelet domain, it is often necessary to return the signal back to its original domain. This is accomplished using synthesis filters and expanders. The wavelet coefficients are applied to a synthesis filter bank in order to retrieve the original signal.

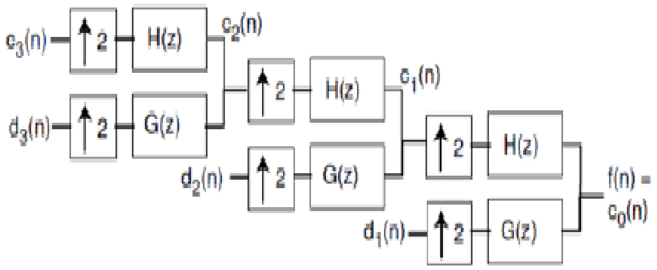

Fig-9: Synthesis filter bank

Two dimensional Discrete Wavelet Transform

The two dimensional discrete wavelet transform is essentially a one dimensional analysis of a two dimensional signal. It operates only on one dimension, at a time, by considering the rows and columns of an image separately. First, the analysis filters are applied to the rows of an image. This results in the formation of two new images, where one image corresponds to a set of coarse row coefficients, while the other corresponds to a set of detail row coefficients.

Next analysis filters are applied to the columns of each new image, to produce four different images which are referred to as subbands. When a high pass filter is applied to the rows and columns of an image, they are denoted by H. Similarly, when a low pass filter is applied to the rows and columns, they are denoted by L. For example, if a subband image is produced using a high pass filter on the rows and a low pass filter on the columns, it is called the HL subband.

Each sub band provides different information about the image. The LL subband is a coarse approximation of the image and removes all high frequency information. The LH subband removes high frequency information along the rows and emphasizes high frequency information along the columns. The result is an image in which horizontal edges are emphasized. The HL subband emphasizes vertical edges, and the HH subband emphasizes diagonal edges.

The four images produced from each level of wavelet decomposition are LL, LH, HL, and HH. The LL image is considered a reduced version of the original as it retains most details. In wavelet decomposition, only the LL image is used to produce the next level of decomposition.

When a high pass filter is used on an image, there are high variations in the gray level between the two adjacent pixels. So edges occur in image. Application of a low pass filter on an image, results in smooth variations between the adjacent pixels, due to which very few edges are generated.

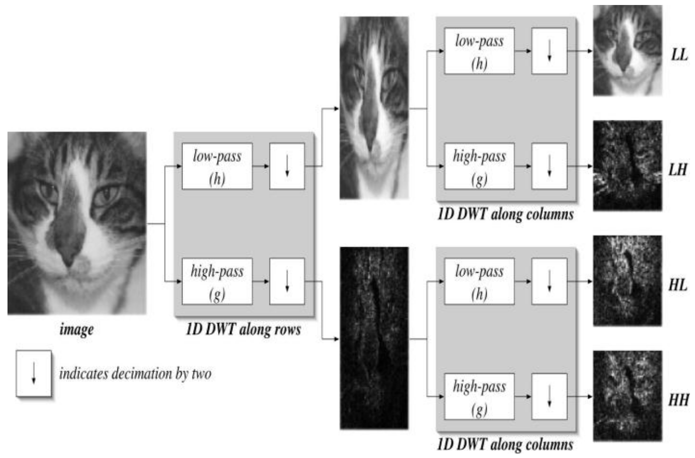

Fig-10: An example of Single Level Decomposition

In our proposed system, single level 2-D discrete wavelet transform is applied to the feature maps from the first three convolutional layers, which returns approximation coefficients matrix (cA) and detail coefficients matrices, cH, cV, and cD (horizontal, vertical, and diagonal, respectively). The approximation coefficients represent the low frequency information in an image, thus being equivalent to the LL subband. The detail coefficients represent the high frequency information along the horizontal, vertical and diagonal directions.

2.2.4 INVERSE DISCRETE WAVELET TRANSFORM

Inverse Discrete Wavelet Transform is a method used to reconstruct the obtained discrete wavelet coefficients back into the original signal or image. The idwt2 command in MATLAB performs a single level two-dimensional wavelet reconstruction.

Single level reconstructed approximation coefficients matrix (say X), are obtained on the basis of approximation matrix cA and details matrices cH, cV, and cD (horizontal, vertical, and diagonal, respectively). It is simply a reverse of DWT. First, the rows of each of the coefficient matrices are upsampled and the columns are convolved with reconstruction low pass and high pass filters. Next, the columns of the convolved output are upsampled, and the rows are convolved with reconstruction lowpass and highpass filters to finally arrive at the original image with enhanced resolution.

3. CONCLUSION

The proposed video denoising system was successfully implemented in MATLAB 2017 software using image processing toolbox. Though the existing noise reduction algorithms are efficient and robust in removing various types of noise, the application of filters can result in information loss. This drawback can be overcome in our proposed super resolution algorithm comprising of an Artifacts Reduction Convolutional Neural Network (ARCNN), which not only retains the important details but also reduces the undesired artifacts in the reconstructed frames. The use of ARCNN in the implementation of SISR helped in reducing blur artifact in the frames. The Peak Signal-to-Noise Ratio (PSNR) and Root Mean Square Error (RMSE) were estimated as a measure of the quality of the obtained output frames. In future, a deeper Convolutional Neural Network (CNN) can be designed to perform tasks like sobel edge detection and intensity level slicing on the video frames.

REFERENCES

[1] Aparna Tripathi, Om Prakash &Yadav, (2015) "Build an Effective Video Noise Removal Method using Combination of Intraframe Filters", International Research Journal of Engineering and Technology (IRJET), Volume: 02, Issue: 2.

[2] Mrs. Amruta Savgave, (2014) "Study of Image Interpolation", International Journal of Innovative Science, Engineering & Technology, Volume 1, Issue 10.

[3]Kulbir Singh and Gaurav Kumar,(2013) "Image Super Resolution on the basis of DWT and Bicubic Interpolation", International Journal of Computer Applications, Volume 65-No.15.

[4] Josep Arnal and Luis Súcar, (2019) "Hybrid Filter Based on Fuzzy Techniques for Mixed Noise Reduction in Color Images", Applied Sciences, MDPI.

[5] Lukasz Malinski and Bogdan Smolka (2019) "Fast adaptive switching technique of impulsive noise removal in color images", Journal of real time image processing', Volume 16, pp.1077-1098.

[6] Nalin Kumar and M. Nachamai, (2017) "Noise Removal and Filtering Techniques Used in Medical Images", Oriental Journal of Computer Science & Technology, Volume 10,No.(1): pp. 103-113.

[7]Ali Abdullah Yahya, Jieqing Tan, and Lian L. (2015) "Video Noise Removal Method Using Adaptive Spatial-Temporal Filtering", Hindawi Publishing Corporation, Discrete Dynamics in Nature and Society, Volume 2015, Article ID 351763.

[8]Sweta Patel, Karshan Kandoriya, (2016) ''Analysis of Various Single Frame Super Resolution Techniques for better PSNR", International Research Journal of Engineering and Technology (IRJET), Volume 03, Issue: 2.