A Novel Approach for Super Resolution of Video Frame using Spatially Adaptive Total Variation

Vinod Kumar Banse
M.Tech Scholar, Department of CSE
Technocrats Institute of Technology, Bhopal
RGPV, Bhopal

Kamlesh Chandravanshi
Associate Professor, Department of CSE
Technocrats Institute of Technology, Bhopal
RGPV, Bhopal

ABSTRACT

Super resolution (SR) for real-life video sequences is a challenging problem due to complex nature of the motion fields. In this paper, a novel blind SR method is proposed to improve the spatial resolution of video sequences, while the overall point spread function of the imaging system, motion fields, and noise statistics are unknown. The high-resolution frames are estimated using a cost function that has the fidelity and regularization terms of type Huber–Markov random field to preserve edges and fine details. The fidelity term is adaptively weighted at each iteration using a masking operation to suppress artifacts due to inaccurate motions. Very promising results are obtained for real-life videos containing detailed structures, complex motions, fast-moving objects, deformable regions, or severe brightness changes. The proposed method outperforms the state of the art in all performed experiments through both subjective and objective evaluations.

Keywords

Video super resolution, blur de-convolution, blind estimation, Huber Markov random field (HMRF).

1. INTRODUCTION

These Signal processing techniques to augment the visual quality of images and videos are nowadays particularly appealing. A first reason for this assertion is due to the technological progress that has raised the standards and the user expectations when enjoying multimedia contents. The past decade, in fact, has witnessed a revolution in large-size user-end display technology: consumer markets are currently flooded with television and other display systems - liquid crystal displays (LCDs), plasma display panels (PDPs), and many more -, which present very high-quality pictures with crystal-clear detail at high spatial and temporal resolutions. And the trend does not appear to be declining: only a few months ago, on the occasion of the annual Consumer Electronics Show (CES) held in January 2014, the company Seiki announced new 65-inch 4K TV models, which will be possible to buy at affordable prices. The recently adopted new compression standard HEVC, too, considered 4K videos for the tests, even if only on cropped areas.

However, despite the increasing interest towards them, high-quality contents are not always available to be displayed. Video source data are unfortunately often at a lower quality than the desired one, because of several possible causes spatial and temporal down-sampling that may be necessary, noise degradation, high compression, etc. Moreover, the new sources of video contents, like the

Internet or mobile devices have generally a lower picture quality than conventional TV broadcasting. Lower-quality video sources pose a formidable obstacle to high-definition video display, and unfulfilled results are paradoxically even

more accentuated by the new high-technology display systems. When we consider still images, things seem to be better. Modern cameras, even the handy and cheap ones, allow any user to easily produce breathtaking high-resolution photos. However, if we consider the past production, there is a tremendous amount of user-produced images collected in the years that are valuable but may be affected by a poor quality. A need to improve the image quality can then be remarked also in this case.

Apart from the user experience, reasons for the need of augmenting the resolution of a video or an image can also be imposed by the particular application context considered. Many applications, e.g. video surveillance and remote sensing, in fact, require the display of images at a considerable resolution, possibly for specific tasks like object recognition or zoom-in operations. Furthermore, compatibility requirements between the different video standards can sometimes lead to the necessity of implementing techniques of resolution up-conversion and de-interlacing.

Image resolution is determined by two main factors. Blurring, due to optical limits and various other processes (like the effect of the atmosphere and motion blur, for example), results in soft images, while low-sensor density of the imaging device causes aliasing. Signal processing based super-resolution (SR) methods are typically concerned with overcoming the resolution limitation resulting in aliasing (although such techniques do take blur into consideration). In this context, 'resolution' refers to the sampling interval, or pixel size. Coarse sampling (pixels of relatively large size) results in 'low resolution' images, while 'high resolution' images correspond to fine sampling (pixels of relatively small size) 1. This is in contrast to optical super-resolution where the aim is to beat the diffraction limit2. Optical SR methods are expensive and are usually developed to enhance the resolution of an already expensive imaging system that is capable of producing very high resolution images (up to the diffraction limit). Henceforth, the term 'super-resolution' shall be used exclusively to refer to the process of overcoming the sensor density limitation using signal processing methods3.

2. LITERATURE SURVEY

Esmaeil Faramarzi et. al. [1] "Blind Super Resolution of Real-Life Video Sequences" To estimate the blur(s), first, a non-uniform interpolation SR method is utilized to up-sample the frames, and then, the blur(s) is (are) estimated through a multiscale process. The proposed method outperforms the state of the art in all performed experiments through both subjective and objective evaluations. A method for blind deconvolution and super resolution from one low-resolution video is introduced in this paper. A masking operation is applied during each iteration of the final frame reconstruction to successively suppress artifacts resulted by inaccurate motion estimation. The complicated nature of motion fields in

real-life videos makes the frame and blur estimations a challenging problem.

Sunghyun Cho et. al. [2] "Video Deblurring for Hand-held Cameras Using Patch-based Synthesis" This paper presents a video deblurring method that can effectively restore sharp frames from blurry ones caused by camera shake. Videos captured by hand-held cameras often contain significant camera shake, causing many frames to be blurry. In proposed method author detects sharp regions in the video, and uses them to restore blurry regions of the same content in nearby frames. The camera motion in a hand-held sequence often causes some portion of video frames to be more blurry than others. Since author gives solution only involves forward convolution and patch-based image synthesis, it is robust enough to handle a wide range of real world videos.

Ronald Fevig et. al. [3] "Super-Resolution Mosaicking of UAV Surveillance Video" This paper explains and implements our efficient multi-frame super-resolution mosaicking algorithm. A generative model is then adopted and combined with Maximum Posteriori estimation to construct the underdetermined sparse linear system. Finally, multi-band blending is used to stitch these resolution-enhanced frames to form the final image. In this paper, we present a robust strategy for producing a resolution-enhanced panorama from a UAV video clip. It is difficult to directly compare the run-time for our technique relative to other algorithms since our method is so unique, dove-tailing both super-resolution and mosaicking into a single construct.

Tomonari Yoshida et. al. [4] "Robust Face Super-Resolution Using Free-Form Deformations For Low-Quality Surveillance Video" The proposed method can generate a high-resolution face image from low-resolution video frames including non-rigid deformations caused by changes of face poses and expressions without using any positional information of facial feature points. This enables super-resolution of face images from low-resolution videos. Author proposed a video-based super-resolution method for low-quality face images that can deal with non-rigid deformations caused by changes of face poses and expressions without any positional information of facial feature points. The results of image quality evaluation demonstrated that the proposed method could provide high-quality super-resolution images. Authors confirmed the effectiveness of the proposed method for face recognition from low-resolution videos.

Richard R. Schultz et. al. [5] "Super-Resolution Mosaicking of Unmanned Aircraft System (UAS) Surveillance Video using Levenberg Marquardt (LM) Algorithm" This paper presents a novel algorithm for the construction of super-resolution mosaicking. The algorithm is based on the Levenberg Marquardt (LM) method. Authors present the results with synthetic and real UAS surveillance data, resulting in a great improvement of the visual resolution. For that, author first compute the LR mosaic which becomes the input to the SR LM algorithm the results for synthetic and real frames from UAS show a great improvement in the resolution.

Ashish Gehani et. al. [6] "Super-Resolution Video Analysis for Forensic Investigations" This paper presents a super-resolution algorithm that differs from its counterparts in two important respects. Super-resolution algorithms typically improve the resolution of a video frame by mapping and performing signal processing operations on data from frames immediately preceding and immediately following the frame of interest. This paper describes how temporally proximal frames of a video sequence can be utilized to aid forensic video analysts by enhancing the resolution of individual frames.

Erman Engin et. al. [7] "Moving Target Detection Using Super-Resolution Algorithms with an Ultra Wideband Radar" present subspace based detection algorithms for detecting moving targets applied to the data collected by an ultra wideband radar system. Authors investigated methods to resolve the ambiguities in target detection and we performed experiments using two targets to test the effectiveness of the algorithms. The performance of the signal processing algorithms and our processing scheme is tested with the data acquired by a UWB radar system in a free space environment with two moving targets. Authors explained three methods to eliminate ambiguous cases in general. 1st method is the combining the processed results at multiple frequencies, 2nd method is the processing the echoes that are distant enough from each other in time separately and the third method is using the proper antenna placement in the array.

Jorge Núñez et. al. [8] "Super-Resolution of Remotely Sensed Images with Variable-Pixel Linear Reconstruction" In this paper, we describe its development for remote sensing purposes, show the usefulness of the algorithm working with images as different to the astronomical images as the remote sensing ones, and show applications to set of simulated set of multispectral real. In this paper we have presented the super-resolution algorithm SRVPLR and two examples of applications aimed at recognition of objects with sizes approaching the limiting spatial resolution scale. SRVPLR uses a nonuniform interpolation algorithm with low computational load, thus enabling real-time applications.

Y i Wang et. al. [9] "Panorama Recovery from Noisy UAV Surveillance Video" In this algorithm, the Eigen-space based neighborhood region will be introduced with our novel feature-based random M least squares (RMLS) registration technique. Finally, the sub-region in each frame which is applicable to the multi-frame sampling will be stitched utilizing multi-resolution blending. In this paper, we present a robust strategy for producing a noise-eliminated panorama from a noisy UAV video clip. And this algorithm is especially useful for the refining the data from a large set of poor observations.

3. RELATED WORK

A non-uniform interpolation SR method is utilized to up-sample the frames, and then, the blur(s) is (are) estimated through a multiscale process. The proposed method outperforms the state of the art in all performed experiments through both subjective and objective evaluations [1]. In proposed method author detects sharp regions in the video, and uses them to restore blurry regions of the same content in nearby frames. The camera motion in a hand-held sequence often causes some portion of video frames to be more blurry than others. Since author gives solution only involves forward convolution and patch-based image synthesis, it is robust enough to handle a wide range of real world videos [2].

4. THEORY

The several super-resolution (SR) approaches that are used to fuse these LR images together. We categorize these approaches into frequency domain and spatial domain approaches and discuss these in some detail. Closest contemporary work dealing with spatial-temporal SR and the need for accurate alignment and registration between the video sequences.

4.1. Imaging Model and Low Resolution Image Generation

Super-resolution (SR) is the process of combining multiple low resolution (LR) images or videos to form a higher resolution (HR) image. Nearly all SR algorithms are based on the premise that the LR images are blurred and down-sampled versions of the SR images.
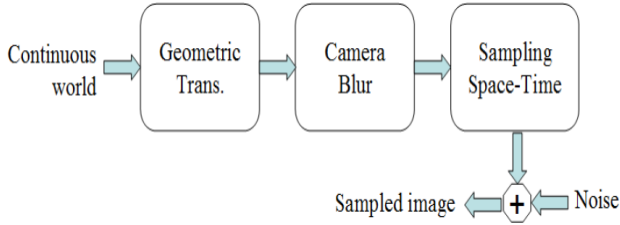

Fig.1 General imaging model of the generation of low resolution images

A general model of the generation of LR images is shown in Fig.1, where a continuous HR scene undergoes a geometric transformation from the world coordinates to the camera coordinates and looses spatial resolution as the camera's point spread function leads to some blurring. The scene is also sampled in both space and time and we can assume that through this entire process of acquisition some noise gets added to the image, leading to a LR representation of the real scene.

4.2. Categorization of Super Resolution Imaging Techniques

Past research in SR imaging can be categorized in two ways. The first categorization looks at the problem domain that is addressed – namely, image super-resolution or video super-resolution. We call this categorization 'image versus video SR'. The second categorization looks at the techniques or methods for SR imaging – namely, spatial versus frequency domain SR techniques. These two categorizations are discussed in further detail in the following sections.
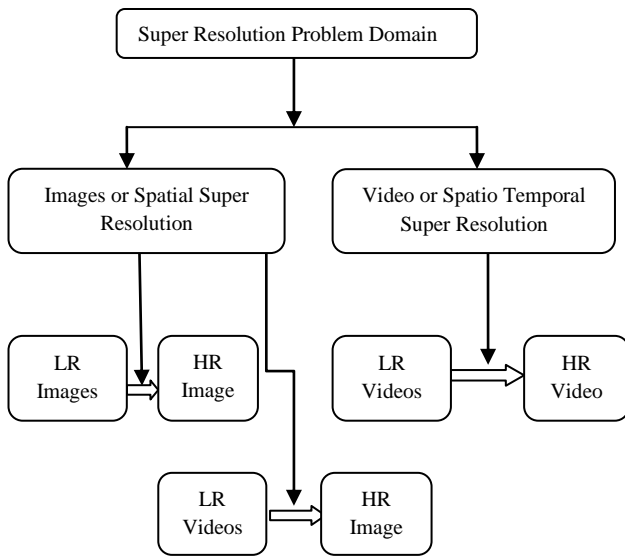
Fig.2 Super-resolution categories based on problem domain - Image SR versus Video SR

5. PROPOSED METHODOLOGY

In the given figure3, firstly we took an original image and then apply down sampling algorithm to design and then we extract spatial information of the sampled data and then we add noise to the sampled data like Gaussian noise and then filtering the data with the help of clustering algorithm.

5.1. RSATV Algorithm:

(1)   Initialization:
$$u^o = g, b^o = 0, d^o = 0, \sigma^o, \sigma_{prev} >> 1$$

(2)   While $(|\sigma_{prev} - \sigma| > \varepsilon_1)$ repeat

$$u^{k+1} = \arg\min_u \frac{1}{2} \| H_a u - g \|_2^2 + \frac{\beta}{2} \| d^k - \nabla u - b^k \|_2^2 + \frac{\mu}{2} [u]_2^2$$

Where u represents the number of LR images, $\| d^k - \nabla u - b^k \|_2^2$ is the data fidelity item,

$$d^{k+1} = \arg\min_d \lambda \sum_{if \in 0} p_{ij} \sqrt{(d_{ij}^k)^2 + (d_{ij}^v)^2} + \frac{\beta}{2} \| d - \nabla u^{k+1} - b^k \|_2^2$$

Where d_{ij}^k And d_{ij}^v are linear operators corresponding to the horizontal and vertical first-order differences. P represents the number of LR images. β is a constant factor. The direction of the gradient ∇u
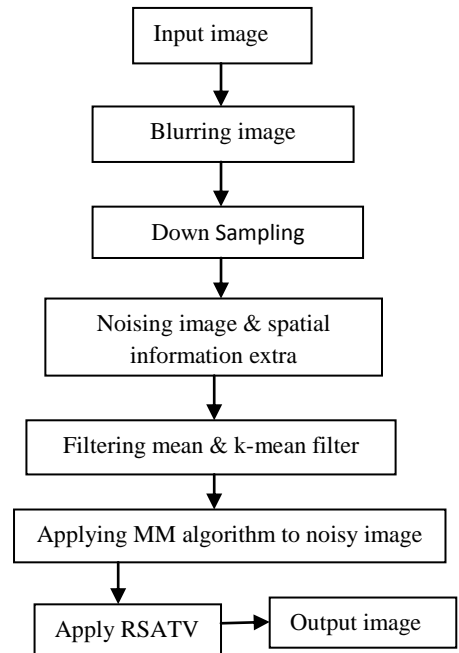
$$b^{k+1} = b^k + (\nabla u^{k+1} - d^{k+1})$$

Set $\sigma_{prev} = \sigma$, and solve

$$\sigma = \arg\min_\sigma \frac{1}{2} \| H_\sigma u^{k+1} - g \|_2^2 + r \| \nabla h_\sigma \|_2^2$$

End

Output u^{k+1}

Fig.3 Output Procedure Block Diagram

Calculate K means-clustering of an image sample. After this apply MM algorithm for the noisy image to de-noise of data sample and then apply RSATV algorithm get high resolution image output as show on figure3.
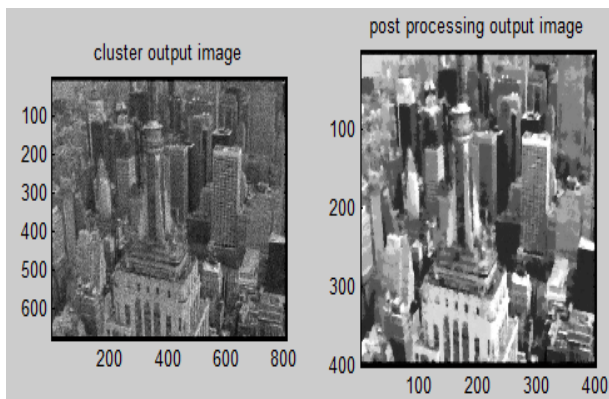
6. SIMULATION RESULTS

In this research work proposed algorithm for blind super-resolution. This proposed algorithm results are shown below:
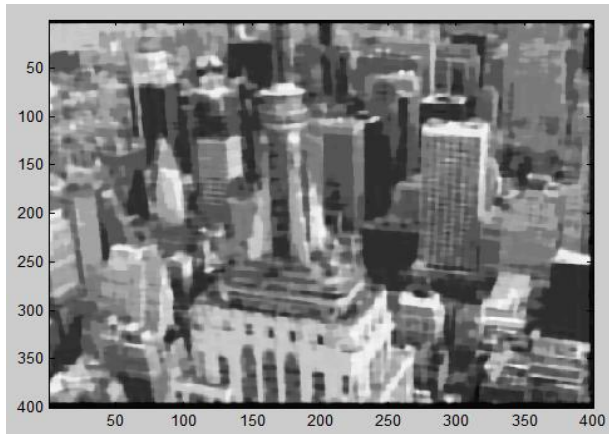
Fig.4 Majorization and Minimization

This fig.4 represents the Majorization and Minimization image. In this Applying MM Algorithm to Noisy Image then we get that Image.

Fig.5 Cluster and Post Processing Output Image

This fig.5 shows the output image. In the output there are two output images that is Cluster Output Image and Post Processing Output Image.

Table 1 Comparison Table of City Image

Parameter	Base paper	Proposed
PSNR	32.4	39.89
SSIM	0.8672	0.9899
Simulation Time	15.45	11.69 sec
NMSE	0.08	0.071

In the above table represents comparison of previous method PSNR value and proposed method PSNR value. In this table also SSIM and simulation time are calculated.
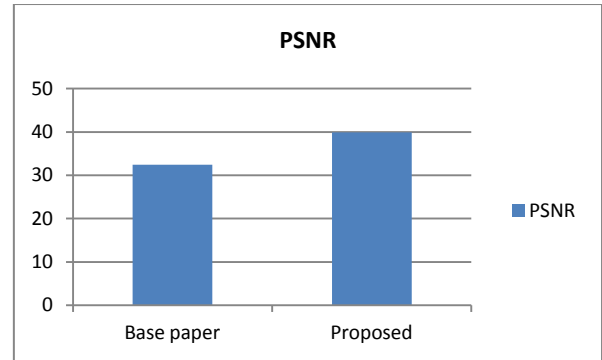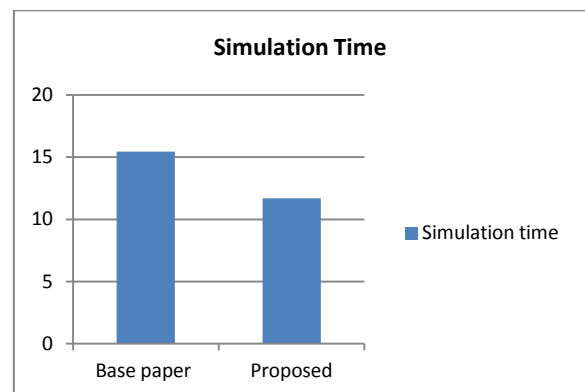

Fig.6 Graphical Representation of PSNR

Fig.6 Demonstrates the Graphical Representation of PSNR values. In previous method PSNR value is lower than the proposed method PSNR value. In the base paper quality metric is low that is 32.4 quality is poor than the proposed method result. The proposed method PSNR is 39.89 high than base paper PSNR, so the result image quality is better. PSNR is the ratio of maximum squared intensity of the signal to the mean square error. The term peak signal-to- noise ratio (PSNR) is an expression for the ratio between the maximum possible value (power) of a signal and the power of distorting noise that affects the quality of its representation.
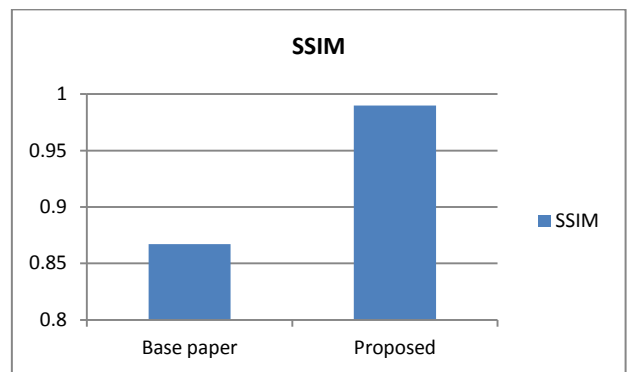
Fig.7 Graphical Representation of SSIM

Fig.7 shows graphical representation of SSIM. The base paper SSIM value 0.9899 is high and proposed method SSIM value 0.8672 is low. So the proposed method result is better than the previous method result. The structural similarity (SSIM) index measurement is a method for predicting the perceived quality of digital television and cinematic pictures, as well as other kinds of digital images and videos.

Fig.8 Graphical Representation of Simulation Time

Fig.8 demonstrates the graphical representation of simulation time. In previous method simulation time that is project running time is 15.45 high and proposed method simulation time that is project running time is 11.69 low. So this result is better than previous result. The time to complete each iteration of the process (each case) and to complete each task in each iteration is gathered as data is known as simulation time.
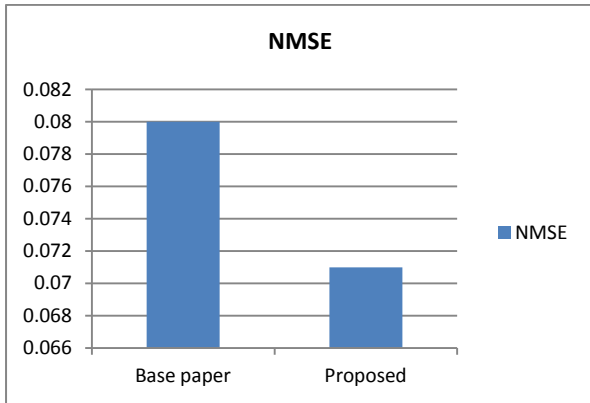

Fig.9 Graphical Representation of NMSE

Fig.9 shows the graphical representation of NMSE. Normalized Mean Square Error (NMSE) is an estimator to the whole deviations between predicted and calculated values. The base paper NMSE value is 0.08 and proposed method NMSE value is 110.17.
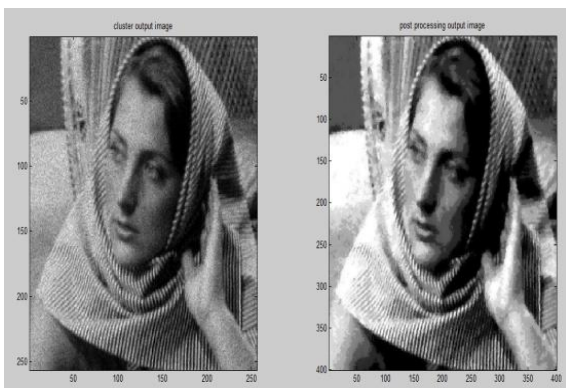

Fig.10 Majorization and Minimization

Fig.11 Cluster and Post Processing Output Image

Table 2 Comparison Table of Barbara Image

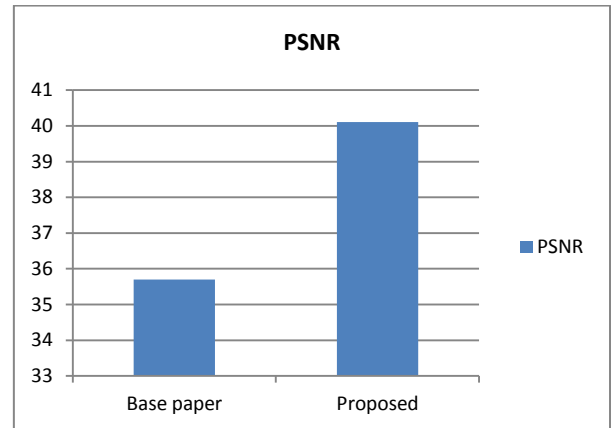| Parameter	Base Paper	Proposed
PSNR	35.7	40.11
SSIM	-	0.9069
Simulation Time	-	51.512 secs

Fig.12 Graphical Representation of PSNR

7. CONCLUSION

The conventional spatially adaptive total variation model has the inadequacy of being weak to noise, and it performs inadequately in high noise intensity conditions. To beat this, in this paper, we propose a Regional spatially adaptive (RSATV) super-resolution calculation with spatial data filtering and clustering. The spatial data is initially extricated for every pixel, and after that the spatial data filtering procedure and spatial weight clustering methodology are included. With these two courses of action, the regularization quality of the total variation model is balanced for every area with distinctive spatial information, instead of for every pixel, as in the conventional spatially adaptive TV model. The real data information examinations introduced in above Section demonstrate that the proposed RSATV model can better suppress the noise, without losing the edge and flat region data.

8. FUTURE WORK

In this paper future exploration, we will concentrate on adaptive parameter selection for the strategy and will also investigate the utilization of more effective optimization algorithm to quicken the solution speed of the RSATV model, for example, the FISTA and MFISTA algorithms. Furthermore, some noise robust spatial feature indicator, for example, steering weights, will likewise be considered, to further enhance the spatial weight construction procedure of the proposed algorithm.

9. REFERENCES

[1] Faramarzi, E., Rajan, D., Fernandes, F. C., & Christensen, "Blind Super Resolution of Real-Life Video Sequences". IEEE Transactions on Image Processing, 25(4), 1544-1555. (2016)

[2] Cho, Sunghyun, Jue Wang, and Seungyong Lee. "Video deblurring for hand-held cameras using patch-based synthesis" ACM Transactions on Graphics (TOG) 31, no. 4 (2012): 64.

[3] Wang, Yi, Ronald Fevig, and Richard R. Schultz. "Super-resolution mosaicking of UAV surveillance video" In Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on, pp. 345-348. IEEE, 2008.

[4] Yoshida, Tomonari, Tomokazu Takahashi, Daisuke Deguchi, Ichiro Ide, and Hiroshi Murase. "Robust face super-resolution using free-form deformations for low-quality surveillance video" In Multimedia and Expo (ICME), 2012 IEEE International Conference on, pp. 368-373. IEEE, 2012.

[5] Camargo, Aldo, Richard Schultz, and Qiang He. "Super-resolution mosaicking of unmanned aircraft system (UAS) surveillance video using Levenberg Marquardt (LM) algorithm" Advances in Visual Computing (2010): 698-706.

[6] S. Ebihara, M. Sato, and H. Niitsuma, "Super-Resolution of Coherent Targets by Directional Borehole Radar" IEEE Transactions on Geoscience and Remote Sensing, vol. 38, no. 4, pp. 1725_1732, Jul. 2000.

[7] A. Gehani and J. H. Reif, "Super-Resolution Video Analysis for Forensic Investigations" in IFIP WG 11.9 International Conference on Digital Forensics, ser. IFIP, vol. 242. Springer, 2007, pp. 281_299.

[8] Y. Wang, R. Fevig, and R. R. Schultz, "Super-resolution mosaicking of UAV surveillance video" in IEEE International Conference on Image Processing (ICIP), 2008, pp. 345_348.

[9] T. S. Huang and R. Y. Tsai, _Multiframe image restoration and registration,_ Advances in Computer Vision and Image Processing, vol. 1, no. 7, pp. 317_339, 1984

[10] M. Irani and S. Peleg, "Super resolution from image sequences" in 10th Intern. Conf. on Pattern Reco. vol. 2, Jun. 1990, pp. 115_120.

[11] S. Borman and R. Stevenson, "Super-Resolution from Image Sequences-A Review" in Midwest Symposium on Circuits and Systems, Notre Dame, IN, USA, 8 1998, pp. 374_378.

[12] A. Papoulis, "Generalized Sampling Expansion" IEEE Trans. on Circuits and Sys-tems, vol. 24, no. 11, pp. 652_654, Nov. 1977.

[13] N. K. Bose and N. A. Ahuja, "Super resolution and Noise Filtering Using Moving Least Squares" IEEE Trans. on Image Processing, vol. 15, no. 8, pp. 2239_2248, Aug. 2006.

[14] A. J. Patti, M. I. Sezan, and A. M. Tekalp, "Superresolution Video Reconstruction with Arbitrary Sampling Lattices and Nonzero Aperture Time" IEEE Trans. on Image Processing, vol. 6, no. 8, pp. 1064_1076, Aug. 1997.

[15] S. Rhee and M. G. Kang, "DCT-Based Regularized Algorithm for High-Resolution Image Reconstruction" in IEEE International Conference on Image Processing (ICIP). Los Alamitos, CA: IEEE, Oct. 1999, pp. 184_187.