International Journal of Innovative Computing, Information and Control
Volume **14**, Number **5**, October **2018**

ICIC International ©2018 ISSN 1349-4198
pp. 1969–1976

VIDEO RESOLUTION ENHANCEMENT USING DEEP NEURAL NETWORKS AND INTENSITY BASED REGISTRATIONS

Gholamreza Anbarjafari[1,2]

[1]iCV Research Group
Institute of Technology
University of Tartu
Nooruse Street, Tartu 50411, Estonia
shb@icv.tuit.ut.ee

[2]Faculty of Engineering
Hasan Kalyoncu University
Gaziantep, Turkey

Received January 2018; revised May 2018

Abstract. *Thanks to the recent rapid improvements made to the maximum possible resolution of display devices, higher qualities of experience have been made possible, which necessitates either producing and transmitting considerably higher volumes of data or super-resolving lower-resolution contents at the display side, where the former might not be practically feasible. Therefore, aiming at the latter, this paper proposes a novel super-resolution technique, which takes advantage of convolutional neural networks. Each image is registered into a window consisting of two frames, the second one standing for the reference image, using various intensity-based techniques, which have been tested and compared throughout the paper. According to the experimental results, the proposed method leads to substantial enhancements in the quality of the super-resolved images, compared with the state-of-the-art techniques introduced within the existing literature. On the Akiyo video sequence, on average, the result possesses 5.38dB higher PSNR values than those of the Vandewalle registration technique, with structure adaptive normalised convolution being utilized as the reconstruction approach.*
Keywords: Super resolution, Deep learning, Convolutional neural network, Intensity based registration, Video resolution enhancement

1. **Introduction.** The issue of high quality image reconstruction from the low quality version of the same image or the sequence of similar video frames has been extensively studied for several decades, and it is still an ongoing research [1, 2, 3, 4, 5, 6]. High demand for the images of high resolution and good visual quality have inspired the proposal of many new approaches, both in the fields of image processing and in camera hardware technologies [7, 8, 9, 10, 11, 12, 13].

Some research work has been conducted in order to study the effect on motion blurring on the quality of super resolution (SR) algorithm performance [13, 14, 15, 16, 17, 18, 19, 20, 21]. In [22], it has been shown that the elimination of the motion blur effect from the initial low resolution video frames has a great impact on the SR algorithm performance. This research inspired authors to propose a novel jitter camera, which allows for the motion blur effect to be minimised in the low resolution video frames for stationary or slow moving objects. As a result, it has been shown that the new jitter camera technology allows for an improvement in the performance of the SR algorithms applied to the low resolution camera frames obtained using the jitter camera.

DOI: 10.24507/ijicic.14.05.1969

Although many novel enhanced technologies in the camera hardware field are proposed every year, there are still limitations to the quality of the images which cameras can capture. Due to those limitations, many SR algorithms are being proposed in the area of image processing. A technical overview of the SR area of image processing has been presented in [23]. In this work various applications of the SR algorithm are outlined with the intent to explain the high demand for those algorithms. This work also provides an overview of the main approaches in SR algorithm design and presents the general SR algorithm scheme, which consists of the following steps: registration, interpolation, restoration.

In the general scheme of the super resolution (SR) algorithm [23] the first step is the registration of low resolution frames of some sequence. Many different approaches can be applied in order to perform registration. For example, the scale-space model for image registration has been proposed in [24]. The scale-space integrated Lucas-Kanade registration is an area-based registration method which, unlike other conventional registration methods, can simultaneously estimate translation, rotation, and scale parameters of the images that have different resolutions.

Another work that outlines the importance of a reliable registration technique is presented in [25]. According to this method, in order to overcome the problem of poorly registered pixels, the registration residuals are used to assign the weights to the registered pixels. The weight value is used to evaluate how reliable is the output of the registration technique for each particular pixel. Pixels with smaller weight values are considered to be poorly-registered and will have smaller influence in the further SR algorithm steps.

In [26] a Bayesian approach to adaptive video SR was proposed. According to the proposed Bayesian approach, the simultaneous estimation of underlying motion, blur kernel and noise levels is performed on the sequence of low resolution frames in order to reconstruct the high resolution frame. A different approach has been proposed in [27]. In this work, a generalised Gaussian Markov random field was proposed to be used for the SR purpose.

The important aspect, which needs to be taken into consideration when designing the SR algorithm, is the fact that the visual quality of the super resolved video frames often suffers from the significant loss of high frequency components. In this paper, a video super resolution technique is proposed which is addressing the aforementioned issue by adopting convolutional neural network SR followed by intensity-based registration on video streams.

The underlying notion of SR algorithms, and the most challenging issue when devising them is to recover the most detailed representation of the previously compressed content, so that the visual quality would be convincing to the eyes of the viewer. In other words, since most of the existing image compression techniques are lossy, achieving the exact amount of the details that have been available in the original image or video sequence may be impossible, which means that the mission of the algorithm, as well as the criterion for evaluating its performance, lies in improving the strength of the whole pipeline in extracting the foregoing information in the most authentic and reliable manner.

The video super-resolution technique proposed in this paper is superior to its conventional and state-of-the-art counterparts in that it can derive the high-frequency data required for the above purpose more efficiently. The latter improvement originates from two fundamental ideas, namely, making use of a deep-learning convolutional neural network and applying intensity-based registration. More clearly, in order to enable the algorithm to extract the high-frequency information most effectively, after super-resolving frames, every second frame is registered onto the preceding frame while the former frame is the reference.

The structure of this paper is organised as follows. The proposed method is described in detail in Section 2, the experimental results and discussion are presented in Section 3, and finally, the conclusion of the proposed work is presented in Section 4.

2. **The Proposed Method.** In this work firstly each frame is super resolved and the upsampled frames are registered in order to enhance the visual quality of the final super resolved streams. In order to upsample each frame, the frame is divided into the patches and for each patch the high resolution patch is obtained by using deep learning based patch mapping [28]. In order to enhance the sharpness of the upsampled frames and to find a high resolution video sequence intensity based image registration is employed. Each of these steps is explained in detail within this section. The following is a description of our method in two parts of SR and registration. A flowchart of the proposed method is illustrated in Figure 1.

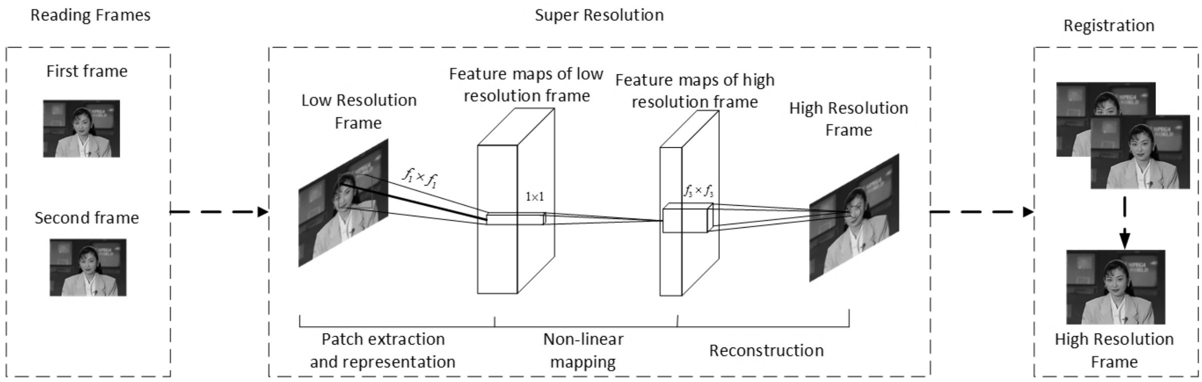

FIGURE 1. The flowchart of the proposed system

Consider a frame of a low resolution video stream, denoted by f_L. First, a bicubic interpolation technique is used for upscaling the frame to the desired size. Denote the interpolated frame as $\tilde{f}_H$, so one can write:

$$\tilde{f}_H = \Gamma^B_\alpha(f_L) \tag{1}$$

where Γ^B is the interpolation function with bicubic kernel and α is the enlargement factor. The goal is to recover from $\tilde{f}_H$ a frame $\boldsymbol{F}\left(\tilde{f}_H\right)$ which is as similar as possible to the ground truth high resolution frame, f_H. In [28], $\tilde{f}_H$ is still called a "low resolution"frame for the ease of presentation, although it has the same size as f_H. The aim is to learn a mapping function, $\boldsymbol{F}$, which conceptually consists of three operations:

- **Patch extraction and representation:** this operation extracts (overlapping) patches from the low resolution frame $\tilde{f}_H$ and represents each patch as a high-dimensional vector. These vectors comprise a set of feature maps, the number of which equals the dimensionality of the vectors.
- **Non-linear mapping:** this operation nonlinearly maps each high-dimensional vector onto another high-dimensional vector. Each mapped vector is conceptually the representation of a high resolution patch. These vectors comprise another set of feature maps.
- **Reconstruction:** this operation aggregates the above high resolution patchwise representations to generate the final high resolution frame, $\hat{f}_H$. This image is expected to be similar to the ground truth f_H.

The learning proposed in [28] is using convolutional neural network (CNN), which is also adopted in this paper. The middle box in Figure 1 shows an overview of the CNN SR technique used in the proposed method.

After obtaining the upsampled high resolution frame, $\hat{f}_H$, every two consecutive frames are being registered in order to enhance the resolution of the frames more by debluring effects in the frames. For this purpose, in the proposed algorithm, the consistent intensity (CI) based registration algorithm [29, 30] using thin plate splin (TPS) regularisation [31] is employed. CI based registration algorithm compares intensity patterns in images via similarity measurements. For the two consecutive super resolved frames, $\hat{f}_{H_i}$ and $\hat{f}_{H_{i+1}}$, and the transformation T, the similarity can be defined by using various means such as second norm or correlation coefficient. If the similarity measure is denoted by $sim\left(\hat{f}_{H_i}, T\left(\hat{f}_{H_{i+1}}\right)\right)$, then the formula turns into an optimisation problem which tries to maximise the similarity in order to obtain the best transformation, $\hat{T}$, as shown below:

$$\hat{T} = \arg\max_{T} sim\left(\hat{f}_{H_i}, T\left(\hat{f}_{H_{i+1}}\right)\right) \tag{2}$$

CI-TPS can be applied to the entire frame or sub-frame. In this work the registration is adopted to the entire frame. The conducted experimental results, which are reported in the next section, show the significant qualitative and quantitative improvement on enhancing the resolution of the videos.

3. **Experimental Results and Discussions.** For the sake of evaluating the efficiency of the proposed resolution enhancement method, it has been compared with a selection of following conventional and state-of-the-art SR techniques suggested:

- Keren et al. [32]: they described a spatial domain procedure to perform image registration using a global translation and rotation model.
- Lucchese and Cortelazzo [33]: they presented a method for estimating planar roto-translations that operates in the frequency domain and hence is not based on features.
- Marcel et al. [34]: they computed the rotation parameters from the motion in the picture. The properties of translation and rotation in the frequency domain of the Fourier transform were used.
- Vandewelle et al. [35]: they proposed a frequency domain technique to precisely register a set of aliased images, based on their low-frequency, aliasing-free part. The high-resolution image was reconstructed using cubic interpolation.

by Keren et al. [32], Lucchese and Cortelazzo [33], Marcel et al. [34] and Vandewelle et al. [35], which make use of registration, followed by interpolation, iterated back projection and structure adaptive normalised convolution techniques for reconstruction. In order to ensure the reliability of the proposed method under wide and various ranges of scenarios, different well-known databases of video sequences from [36] have been taken into account. Namely, they are Akiyo, container and mother & daughter. Each video sequence consists of 300 frames.

For examining the performance of the proposed method, the PSNR is used. In order to do so, first the original high resolution video frames have to be converted to low resolution ones which in our experiments, have been performed by resorting to downsampling by means of the nearest neighbour sampling technique. More clearly, the original high resolution sequence is deemed the ground truth for calculating the PSNR achieved by using each of the foregoing approaches, where the downsampled sequence, after undergoing SR, is compared against the original high resolution sequence. This comparison is undertaken in order to measure the extent of the authenticity of the information contained in the super resolved video sequence, standing for the effectiveness of the SR method. In the context of the experiments reported in this paper, the low resolution video sequences have the size of 128×128, and the super resolved sequences are of the size 256×256.

The values reported in Table 1 represent the average PSNR over the 300 frames of each of the above databases of video sequences, using the proposed method, along with the aforementioned techniques utilised for evaluation and comparison purposes. Figure 2 shows the changes of the PSNR over the 300 frames of all the video sequences for the proposed video SR technique. The average PSNR values associated with the proposed method, which are shown in the aforementioned table, clearly demonstrate the fact that it outperforms the other techniques, including both classical and the most recent ones, which have been used for evaluating this method's efficiency. For better demonstration of the performance of the proposed method over the other method Figure 3 has been prepared. Efficiency has been the aim from the outset, and is based on the modifications and enhancements made to the structure of the underlying functionality, as discussed thoroughly in the preceding sections.

Table 1. The average peak signal to noise ratio (PSNR) in dB, structural similarity index (SSIM), visual information fidelity in pixel domain (VIFP), universal quality index (UQI), noise quality meature (NQM) and weighted signal to noise ratio (WSNR) values of different resolution enhancement techniques on the test video sequences

		PSNR	SSIM	VIFP	UQI	NQM	WSNR
Vandewalle	container	24.48	0.8339	0.3081	0.6351	19.11	29.12
	Akiyo	30.70	0.9435	0.5596	0.8205	22.72	30.28
	mother & daughter	28.98	0.9097	0.4681	0.7361	19.97	29.73
Marcel	container	24.37	0.8355	0.3113	0.6392	19.33	29.33
	Akiyo	30.55	0.9437	0.5608	0.8225	22.49	29.95
	mother & daughter	28.97	0.9065	0.4686	0.7224	18.78	28.72
Lucchese	container	24.18	0.8299	0.3101	0.628	19.56	29.38
	Akiyo	30.23	0.9417	0.5582	0.8164	22.29	29.59
	mother & daughter	28.95	0.9016	0.4276	0.7122	18.67	28.71
Keren	container	24.44	0.8372	0.3149	0.6431	19.39	29.45
	Akiyo	30.47	0.9439	0.5619	0.8229	22.52	29.94
	mother & daughter	28.98	0.9096	0.4731	0.7364	18.78	28.73
Proposed Method	container	**28.71**	**0.9134**	**0.4736**	**0.7848**	**29.52**	**39.62**
	Akiyo	**36.08**	**0.9748**	**0.7366**	**0.9024**	**30.25**	**38.42**
	mother & daughter	**34.30**	**0.945**	**0.611**	**0.8568**	**25.16**	**36.02**

4. **Conclusion.** This paper proposed a new video SR technique by applying CNN based SR, followed by intensity based image registration. The proposed method preserved higher details after super resolution, thereby resulting in sharper super resolved video sequence. The experimental results conducted on several well-known video sequences showed the superiority of the proposed method over the conventional and state-of-the-art techniques.

As future work, one can study the effect of the proposed video resolution enhancement technique on some other computer vision applications such as action recognition using surveillance videos.

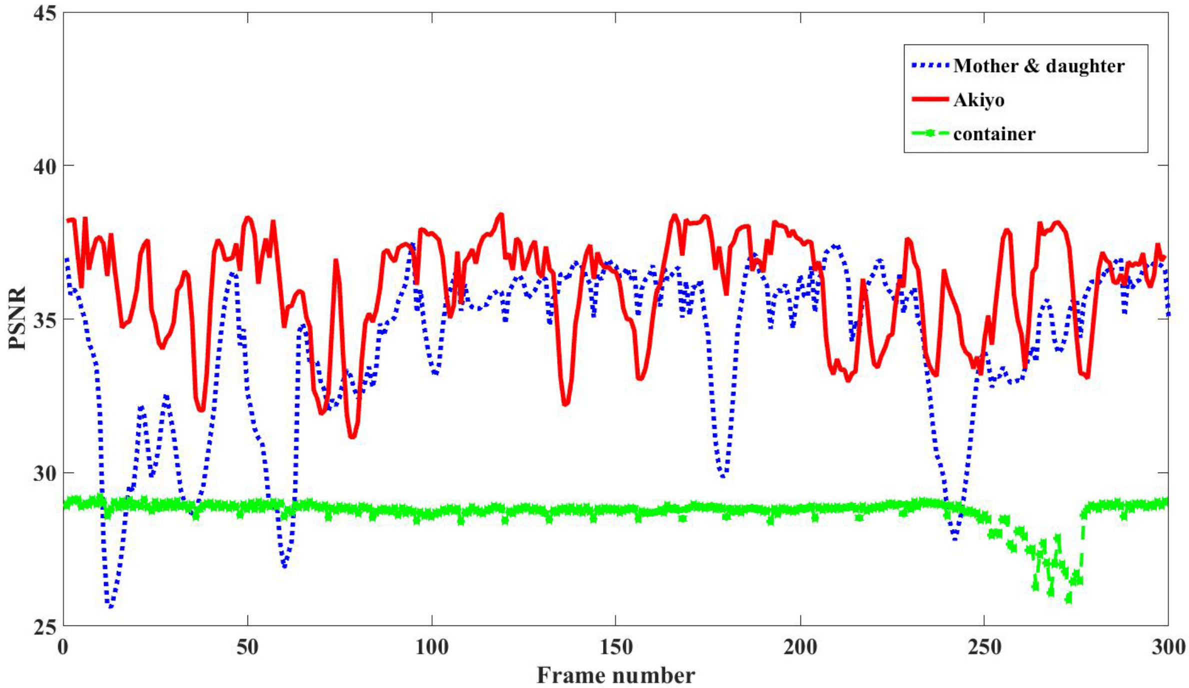

FIGURE 2. The change of PSNR per frame for mother & daughter, Akiyo and container video sequences

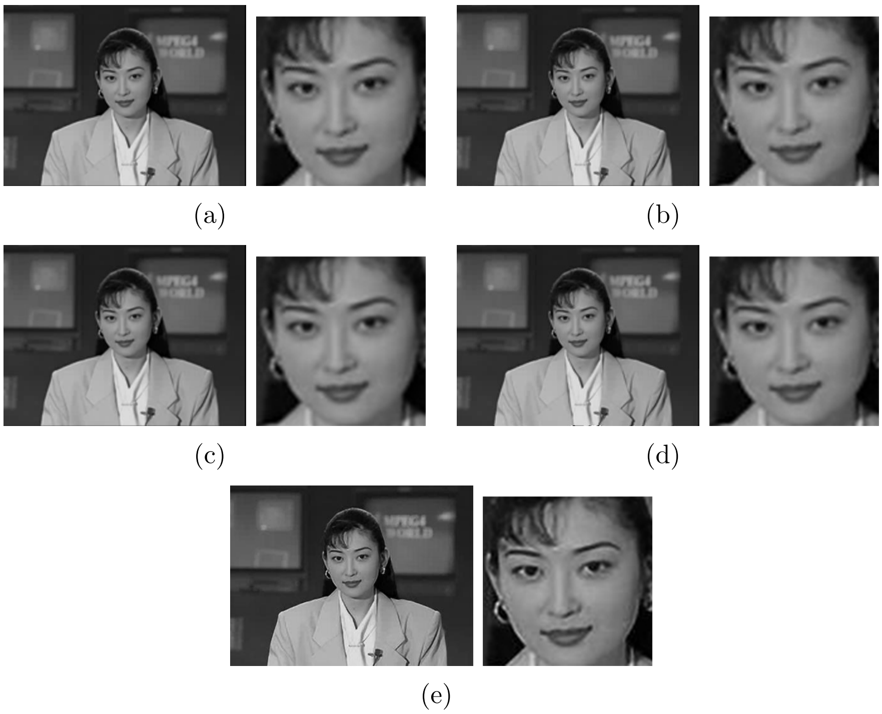

FIGURE 3. Visual comparison of the proposed method with other algorithms, where (a), (b), (c), (d) and (e) are the fifth frame of Akiyo video for Keren, Lucchese, Marcel, Vandewalle and proposed method respectively

Acknowledgement. This work has been partially supported by the Estonian Research Council Grant (PUT638) and the Scientific and Technological Research Council of Turkey (TÜBTAK) (Project 1001-116E097).

REFERENCES

[1] J. Van Ouwerkerk, Image super-resolution survey, *Image and Vision Computing*, vol.24, no.10, pp.1039-1052, 2006.

[2] F. Qin, X. He, W. Chen, X. Yang and W. Wu, Video superresolution reconstruction based on subpixel registration and iterative back projection, *Journal of Electronic Imaging*, vol.18, no.1, 2009.

[3] H.-K. Chen, C.-H. Chen and S.-J. Wang, A fast and robust map-based superresolution reconstruction method using CDHS and block-wise motion vector selection, *International Journal of Innovative Computing, Information and Control*, vol.6, no.7, pp.3105-3120, 2010.

[4] G. Cheng, J. Huang, Z. Liu and L. Cheng, Phase regularisation for single image super-resolution reconstruction, *Electronics Letters*, vol.46, no.14, 2010.

[5] S. E. El-Khamy, M. M. Hadhoud, M. I. Dessouky, B. M. Salam and F. A. El-Samie, Wavelet fusion: A tool to break the limits on LMMSE image super-resolution, *International Journal of Wavelets, Multiresolution and Information Processing*, vol.4, no.1, pp.105-118, 2006.

[6] E. D. Lee, G. Lee, H. Lee, Y. S. Kim and N. Hur, Upscaled sub-pixel multiplexing of multi-view images based on weighted view overlay, *Electronics Letters*, vol.51, no.11, pp.828-830, 2015.

[7] J. Tian and K.-K. Ma, A survey on super-resolution imaging, *Signal, Image and Video Processing*, vol.5, no.3, pp.329-342, 2011.

[8] H. Demirel, G. Anbarjafari, C. Ozcinar and S. Izadpanahi, Video resolution enhancement by using complex wavelet transform, *The 18th IEEE International Conference on Image Processing (ICIP)*, pp.2093-2096, 2011.

[9] G. Chen, S.-E. Qian, J.-P. Ardouin and W. Xie, Super-resolution of hyperspectral imagery using complex ridgelet transform, *International Journal of Wavelets, Multiresolution and Information Processing*, vol.10, no.3, 2012.

[10] G. Anbarjafari, S. Izadpanahi and H. Demirel, Video resolution enhancement by using discrete and stationary wavelet transforms with illumination compensation, *Signal, Image and Video Processing*, vol.9, no.1, pp.87-92, 2015.

[11] P. Rasti, T. Uiboupin, S. Escalera and G. Anbarjafari, Convolutional neural network super resolution for face recognition in surveillance monitoring, *International Conference on Articulated Motion and Deformable Objects*, pp.175-184, 2016.

[12] L. Jovanov, H. Luong and W. Philips, Multiview video quality enhancement, *Journal of Electronic Imaging*, vol.25, no.1, 2016.

[13] P. Rasti, K. Nasrollahi, O. Orlova, G. Tamberg, T. B. Moeslund and G. Anbarjafari, Reducible dictionaries for single image super-resolution based on patch matching and mean shifting, *Journal of Electronic Imaging*, vol.26, no.2, 2017.

[14] B. Bascle, A. Blake and A. Zisserman, Motion deblurring and super-resolution from an image sequence, *Proc. of the 4th European Conference on Computer Vision*, pp.573-582, 1996.

[15] S. Izadpanahi, Ç. Özçinar, G. Anbarjafari and H. Demirel, Resolution enhancement of video sequences by using discrete wavelet transform and illumination compensation, *Turkish Journal of Electrical Engineering & Computer Sciences*, vol.20, no.Sup.2, pp.1268-1276, 2012.

[16] K. Nasrollahi and T. B. Moeslund, Super-resolution: A comprehensive survey, *Machine Vision and Applications*, vol.25, no.6, pp.1423-1468, 2014.

[17] P. Rasti, K. Nasrollahi, O. Orlova, G. Tamberg, C. Ozcinar, T. B. Moeslund and G. Anbarjafari, A new low-complexity patch-based image super-resolution, *IET Computer Vision*, vol.11, no.7, pp.567-576, 2017.

[18] C. Liu and D. Sun, On bayesian adaptive video super resolution, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.36, no.2, pp.346-360, 2014.

[19] G. Anbarjafari, An objective no-reference measure of illumination assessment, *Measurement Science Review*, vol.15, no.6, pp.319-322, 2015.

[20] Z. Ma, R. Liao, X. Tao, L. Xu, J. Jia and E. Wu, Handling motion blur in multi-frame super-resolution, *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.5224-5232, 2015.

[21] K. Nasrollahi, S. Escalera, P. Rasti, G. Anbarjafari, X. Baro, H. J. Escalante and T. B. Moeslund, Deep learning based super-resolution for improved action recognition, *International Conference on Image Processing Theory, Tools and Applications (IPTA)*, pp.67-72, 2015.

[22] M. Ben-Ezra, A. Zomet and S. K. Nayar, Video super-resolution using controlled subpixel detector shifts, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.27, no.6, pp.977-987, 2005.

[23] S. C. Park, M. K. Park and M. G. Kang, Super-resolution image reconstruction: A technical overview, *Signal Processing Magazine*, vol.20, no.3, pp.21-36, 2003.

[24] J. Lee, R. Gutierrez-Osuna and S. S. Young, Silk: Scale-space integrated lucas-kanade image registration for super-resolution from video, *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp.2282-2286, 2013.

[25] X. Zhang, R. Xiong, S. Ma, G. Li and W. Gao, Video super-resolution with registration-reliability regulation and adaptive total variation, *Journal of Visual Communication and Image Representation*, vol.30, pp.181-190, 2015.

[26] C. Liu and D. Sun, A bayesian approach to adaptive video super resolution, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.209-216, 2011.

[27] J. Chen, J. Nunez-Yanez and A. Achim, Video super-resolution using generalized Gaussian Markov random fields, *Signal Processing Letters*, vol.19, no.2, pp.63-66, 2012.

[28] C. Dong, C. C. Loy, K. He and X. Tang, Image super-resolution using deep convolutional networks, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.38, no.2, pp.295-307, 2016.

[29] G. E. Christensen, Consistent linear-elastic transformations for image matching, *Information Processing in Medical Imaging*, pp.224-237, 1999.

[30] G. E. Christensen and H. J. Johnson, Consistent image registration, *IEEE Trans. Medical Imaging*, vol.20, no.7, pp.568-582, 2001.

[31] G. Donato and S. Belongie, Approximate thin plate spline mappings, *European Conference on Computer Vision*, pp.21-31, 2002.

[32] D. Keren, S. Peleg and R. Brada, Image sequence enhancement using sub-pixel displacements, *Proc. of the Computer Society Conference on Computer Vision and Pattern Recognition*, pp.742-746, 1988.

[33] L. Lucchese and G. M. Cortelazzo, A noise-robust frequency domain technique for estimating planar roto-translations, *IEEE Trans. Signal Processing*, vol.48, no.6, pp.1769-1786, 2000.

[34] B. Marcel, M. Briot and R. Murrieta, Calcul de translation et rotation par la transformation de fourier, *Traitement du Signal*, vol.14, no.2, pp.135-149, 1997.

[35] P. Vandewalle, S. Süsstrunk and M. Vetterli, A frequency domain approach to registration of aliased images with application to super-resolution, *EURASIP Journal on Applied Signal Processing*, vol.2006, p.233, 2006.

[36] *Xiph.org Test Media*, http://media.xiph.org/video/derf/, 2016.