A Survey on Blind Super Resolution of Real-Life Video Sequences

A. Susmitha[1*], V. Kiran Kumar[2], P. Anitha[3]
[1]P. G. Student, Malla Reddy College of Engineering & Technology, Secunderabad, Telangana, India.
Email: susmitha221@gmail.com
[2]Associate Professor, ECE Department, Malla Reddy College of Engineering & Technology, Hyderabad, Telangana, India. Email: kiran92929@gmail.com
[3]Associate Professor, ECE Department, Malla Reddy College of Engineering & Technology, Hyderabad, Telangana, India. Email: anithakamal.ghanta@gmail.com
*Corresponding Author

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 pointspread function of the imaging system, motion fields, and noise statistics are unknown. To estimate the blur(s), first, a nonuniform interpolation SR method is utilized to upsample the frames, and then, the blur(s) is(are) estimated through a multiscale process. The blur estimation process is initially performed on a few emphasized edges and gradually on more edges as the iterations continue. Also for faster convergence, the blur is estimated in the filter domain rather than the pixel domain. 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: Blind estimation, Blur deconvolution, Huber Markov Random Field (HMRF), Video super resolution.

I. Introduction

MULTI-IMAGE Super Resolution (SR) is the process of estimating a High Resolution (HR) image by fusing a series of Low-Resolution (LR) images degraded by various artifacts such as aliasing, blurring, and noise. Video super resolution, by contrast, is the process of estimating a HR video from one or multiple LR videos in order to increase the spatial and / or temporal resolution(s). The spatial resolution of an imaging system depends on the spatial density of the detector (sensor) array and the Point Spread Function (PSF) of the induced detector blur. The temporal resolution on the other hand, is influenced by the frame rate and exposure time of the camera. Spatial aliasing appears in images or video frames when the cut-off frequency of the detector is lower than that of the lens. Temporal aliasing arises in video sequences when the frame rate of the camera is not high enough to capture high frequencies caused by fast moving objects. The blur in the captured images and videos is the overall effect of different factors such as defocus, motion blur, optical blur, and detector's blur resulting from light integration within the active area of each detector in the array. The references provide overviews of different SR approaches.

One way to increase the resolution of a video is by overlaying a sliding window upon each frame and combining all frames falling inside the window to build the corresponding HR frame. Then the window slides to the location of the other frames and the process repeats. For this system to work, usually a local registration method (such as optical flow, block-based, pel-recursive, or Bayesian) is required to accurately estimate the displacement vector of each pixel or block within the frames. However, local registration may not be reliable in some cases, especially when there are complex dynamic changes (e.g. complex 3D motions), nonrigid deformations (e.g. flowing water, flickering fire), or changes in illumination.

Another class of single-video SR techniques is the one known as learning-based, patch-based or example-based video SR. The basic idea is that small space-time patches within a video are repeated many times inside the same video or other videos, at multiple spatio-temporal scales. Therefore, by replacing LR patches in the input video with equivalent HR patches from internal / external sources, the resolution can be improved. The major advantage of patch-based image / video

SR methods is that motion estimation and object segmentation are not required. However, techniques of this group often have high computational complexity and most of them need offline database training. Furthermore, it is necessary that LR patches are generated from HR patches by a known PSF.

II. Existing Method

We present experimental results of Digital Super Resolution (DSR) techniques on low resolution data collected using PANOPTES, a multi-aperture miniature folded imaging architecture. The flat form factor of PANOPTES architecture results in an optical system that is heavily blurred with space variant PSF which makes super resolution challenging. We also introduce a new DSR method called SRUM (Super-Resolution with Unsharpenning Mask) which can efficiently highlight edges by embedding an unsharpenning mask to the cost function. This has much better effect than just applying the mask after all iterations as a post-processing step.

Disadvantage

1. Not Applicable on motion blur videos.
2. Less performance on high quality videos.

III. Proposed Method

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. To estimate the blur(s), first, a non-uniform interpolation SR method is utilized to upsample the frames, and then, the blur(s) is(are) estimated through a multiscale process. The blur estimation process is initially performed on a few emphasized edges and gradually on more edges as the iterations continue. Also for faster convergence, the blur is estimated in the filter domain rather than the pixel domain. 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.

Advantages

1. Applicable on motion blur videos.
2. Better performance on high quality videos.

Disadvantage

1. Robustness in sharpness is not much better.

In This Paper the Blind Super Resolution Contains Two Parts:

1. Blur Estimation.
2. Final Frame Estimation.

IV. Blur Estimation

In a multi-channel BD problem, the blurs could be estimated accurately along with the HR images. However, in a blind SR problem with a possibly different blur for each frame, some ambiguity in the blur estimation is inevitable due to the downsampling operation. By contrast, in a blind SR problem in which all blurs are supposed to be identical or have gradual changes over time, such an ambiguity can be avoided. Moreover, the assumption of identical (or gradually changing) blurs makes it possible to separate the registration and upsampling procedures from the deblurring process which significantly decreases the blur estimation complexity. In the NUI method to reconstruct the upsampled frame is explained. This upsampled yet-blurry frame is used to estimate the PSF(s) and the deblurred frames through an iterative Alternative Minimization (AM) process. The blur and frame estimation procedures are discussed. The estimated frames are used only for the deblurring process and so omitted thereafter. Finally, the overall AM optimization process is described.

A. Frame Upsampling

We discuss the situations in which the warping and blurring operation are commutable. Although for videos with arbitrary local motions this commutability does not hold exactly for all pixels, however we assume here that this is approximately satisfied. The ultimate appropriateness of the approximation is validated by the eventual performance of the algorithm that is derived based on this model. With this assumption, (2) can be rewritten as:

gi =DMk,iHfk +ni =DMk,izk +ni (1)

where $z_k = Hf_k$ is the upsampled but still blurry frame. Equation (4) suggests that we can first construct the upsampled frames z_k using an appropriate fusion method and then apply a deblurring method to z_k to estimate f_k and h. If noise characteristics are also the same for all frames, an appropriate way to estimate z_k is using the NUI method. In NUI, the pixels of all LR frames are projected on to the HR image grid according to their motion fields, and then the intensities of the true locations on the grid are computed via interpolation. Our experiments show that using NUI for upsampling the frames leads to better estimates of f and h compared to when z_k is estimated iteratively from the LR frames g using a MAP (Maximum A Posteriori) or ML (Maximum Likelihood) method.

B. Frame Deblurring

After upsampling the frames, we use the following cost function, J, to estimate the HR frames f_k having an estimate of the blur h (or H):

$$J(f_{k)=||} \rho\ (Hf_k\text{-}z_k)||+\lambda^n\Sigma||\rho(\Delta_j f_k)|| \tag{2}$$

where·1 denotes the l1 norm (defined for a sample vector x with elements xi as x= i |xi|), λ^n is the regularization coefficient, ρ(·) is the vector Huber function, ρ(·)1 is called the Huber norm, and Δ_j (i = 1,...,4) are the gradient operators in 0°, 45°, 90° and 135° spatial directions. The first term in is called the fidelity term which is the Hubernorm of error between the observed and simulated LR frames. While in most works the l2-norm is used for the fidelity term, we use the robust Huber norm to better suppress the outliers resulting from inaccurate registration. The next two terms are the regularization terms which apply spatiotemporal smoothness to the HR video frames while preserving the edges.

Each element of the vector function ρ(·) is the Huber function defined as:

$$p(x) = x^2 \quad \text{if } |x|\leq T$$

$$2T|x|-T^2 \quad \text{if } |x|> T, \tag{3}$$

The Huber function ρ(x) is a convex function that has a quadratic form for values less than or equal to a threshold T and a linear growth for values greater than T. The Gibbs PDF of the Huber function is heavier in the tails than a Gaussian. Consequently, edges in the frames are less penalized with this prior than with a Gaussian (quadratic) prior.

To minimize the cost function in, we use the Conjugate Gradient (CG) iterative method because of its simplicity and efficiency. Compared to some other iterative methods such as Gauss-Seidel (GS) or SOR that need explicit derivation of matrix A when solving a linear equation Ax = b, CG can decompose the matrix A to concatenation of filtering and weighting operations. However, CG can only be used with linear equation sets, whereas the cost function is nonquadratic and so its derivative is nonlinear. To overcome this limitation, we use lagged diffusivity Fixed-Point (FP) iterative method to lag the diffusive term by one iteration. Using this method for a sample vector x, at tenth iteration the non-quadratic Huber-norm $\rho(x^n)1$ is replaced by the following quadratic form:

$$||\ \rho(x^n)||=(x^n)^T\ V^n(x^n)=||x^n||^2_{Vn} \tag{4}$$

where V^n is the following diagonal matrix.

C. Blur Estimation

Within an image or video frame, non-edge regions and weak structures are not appropriate for blur estimation. Hence, more accurate results would be obtained if the estimation is not performed in such regions. For this reason the user should first manually select a region with rich edge structure, whereas the most salient edges are automatically chosen. Moreover, sharpening salient edges would also improve the accuracy of blur estimation. The authors of leveraged these two strategies by preprocessing blurred images with the shock filtering method proposed in. Shock filtering is an edge preserving smoothing operation by which soft edges gradually approach step edges within a few iterations while non-edge regions are smoothed. Since shock filtering is sensitive to noise, sometimes a pre-filtering operation is applied to first suppress noise. For example, in bilateral filtering (proposed by) is used and a lowpass Gaussian filtering is utilized before shock filtering. A similar concept for the blur estimation is exploited in which the image is first sharpened by redistributing the pixels along the edge profiles in such a way that antialiased step edges are produced. Having the sharpened image and the blurry input image, the blur is then estimated using a maximum a posteriori (MAP) framework.

D. Final HR Frame Estimation

After the PSF estimation is completed, the final HR frames are reconstructed through minimizing the cost function.

IV. Conclusion

A method for blind deconvolution and super resolution from one low-resolution video is introduced in this paper. The complicated nature of motion fields in real-life videos make the frame and blur estimations a challenging problem. To estimate the blur(s), the input frames are first upsampled using Non-Uniform Interpolation (NUI) SR method assuming that the blurs are either identical or have slow variations over time. Then the blurs are determined iteratively from some enhanced edges in the upsampled frames. After completion of blur estimation, the reconstructed frames are discarded and a non-blind iterative SR process is performed to obtain the final reconstructed frames using the estimated blur(s). A masking operation is applied during each iteration of the final frame reconstruction to successively suppress artifacts resulted by inaccurate motion estimation. Comparison is made with the state of the art and the superior performance of our proposed method is confirmed through different experiments**.**

References

[1] E. Faramarzi, V. R. Bhakta, D. Rajan, and M. P. Christensen, "Super resolution results in PANOPTES, an adaptive multi-aperture folded architecture," In *Proc. 17th IEEE Int. Conf. Image Process. (ICIP),* pp. 2833-2836, Sep. 2010.

[2] E. Faramarzi, D. Rajan, and M. P. Christensen, "Unified blind method for multi-image super-resolution and single/ multi-image blur deconvolution," *IEEE Trans. Image Process.,* vol. 22, no. 6, pp. 2101-2114, Jun. 2013.

[3] S. Borman, and R. L. Stevenson, "Spatial resolution enhancement of lowresolution image sequences: A comprehensive review with directions for future research," Dept. Elect. Eng., Univ. Notre Dame, Notre Dame, IN, USA, Tech. Rep., Jul. 1998.

[4] S. Borman, and R. L. Stevenson, "Super-resolution from image sequences: A review," In *Proc. Midwest Symp. Circuits Syst.,* Notre Dame, IN, USA, pp. 374-378, Aug. 1998.

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

[6] R. R. Schultz, L. Meng, and R. L. Stevenson, "Subpixel motion estimation for super-resolution image sequence enhancement," *Journal of Visual Communication and Image Representation,* vol. 9, no. 1, pp. 38-50, Mar. 1998.

[7] A. M. Tekalp, "Digital video processing," Englewood Cliffs, NJ, USA: Prentice-Hall: Signal Processing Series 1995.

[8] Y. Caspi, and M. Irani, "Spatio-temporal alignment of sequences," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 24, no. 11, pp. 1409-1424, Nov. 2002.

[9] O. Shahar, A. Faktor, and M. Irani, "Space-time super-resolution from a single video," In *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR),* pp. 3353-3360, Jun. 2011.

[10] V. Cheung, B. J. Frey, and N. Jojic, "Video epitomes," *International Journal of Computer Vision,* vol. 76, no. 2, pp. 141-152, 2008.

[11] R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. T. Freeman, "Removing camera shake from a single photograph," *ACM Trans. Graph.,* vol. 25, no. 3, pp. 787-794, 2006.

[12] Q. Shan, J. Jia, and A. Agarwala, "High-quality motion deblurring from a single image," *ACM Trans. Graph.,* vol. 27, no. 3, p. 73, 2008.

[13] S. Cho, and S. Lee, "Fast motion deblurring," *ACM Transactions on Graphics,* Art. ID 145, vol. 28, no. 5, 2009.

[14] L. Xu, and J. Jia, "Two-phase kernel estimation for robust motion deblurring," In *Proc. 11th Eur. Conf. Comput. Vis.,* pp. 157-170, 2010.

[15] T. F. Chan, and C.-K. Wong, "Total variation blind deconvolution," *IEEE Transactions on Image Processing,* vol. 7, no. 3, pp. 370-375, Mar. 1998.