

Using Structural Similarity Error Measurement In Multi-frame Image Super-Resolution

M. Amintoosi, M. Fathy and N. Mozayani
Computer Engineering Department,
Iran University of Science and Technology
Narmak, Tehran, Iran
{mAmintoosi,mahFathy,Mozayani}@iust.ac.ir

Abstract

The Iterated Back Projection (IBP) is a famous reconstruction method in Super-Resolution context, which is based on the sum of squared differences of two images. It is commonly known that the mean square error does not accurately reflect the subjective image quality for most image and video enhancement tasks. Among the various image quality metrics, Structural Similarity provides remarkably good prediction for subjective scores. In this paper a new version of IBP method based on contribution of this measurement to IBP formulation is proposed. The proposed approach has been tested over the classical IBP approach and the robust super resolution method, which is a version of IBP method. The new methods applied to a video super resolution problem, successfully. Various objective and subjective comparisons show the superior performance of the proposed approach.

Keywords: Super-Resolution, Iterated Back Projection, Structural Similarity.

1 Introduction

The Super-Resolution (SR) techniques fuse a sequence of low-resolution (LR) images to produce a higher resolution (HR) image. The low resolution images may be noisy, blurred and have some displacement with each other. A common matrix notation which is used to formulate the super-resolution problem [6, 7] is as follows:

$$\underline{Y}_k = DHF_k \underline{X} + \underline{V}_k, \quad k = 1, \dots, N \quad (1)$$

where:

- \underline{X} is the high-resolution frame and
- \underline{Y}_k is the k^{th} low-resolution frame which are rearranged in lexicographic order,

- F_k is the geometric motion operator between the high-resolution frame \underline{X} and the k^{th} low-resolution frame \underline{Y}_k ,
- r is the resolution enhancement factor,
- H is the camera's point spread function (PSF),
- D represents the decimation operator,
- Vector V is the system noise and
- N is the number of available low-resolution frames.

We assumed that the decimation operator D and blur matrix H is same for all images.

The vast majority of the super-resolution restoration algorithms – named as reconstruction methods – use a short sequence of low-resolution input frames to produce a single super-resolved high-resolution output frame. Perhaps the mean square error is the most common objective criterion for measuring the differences in the image and video domains for several years. For this reason many super-resolution methods try to minimize an error function, reflecting the difference between down-sampled of the estimated high resolution image and low resolution frames.

The Iterated Back Projection technique, initially introduced by Michal Irani and Schmuël Peleg [12] is one of the pioneering works in the field of super-resolution. This method uses averaged projections in HR grid to iteratively solve the HR image X , by minimizing an error function. The method of Zomet et al. [20] -also known as the Robust Super-resolution (SR)- is a version of IBP in which instead of averaging, uses median operator. These methods are the basis or a part of some later methods [11, 8, 7, 9, 16, 17].

Minimization of the mentioned error function, alone can lead to excessive noise magnification in some applications due to the ill-posed nature of the SR problem[11]. In some super-resolution approaches, such as Shift & Add approach of Farsiu et al. [7], Bayesian approach of Cheeseman et al. [5] and example based approach of Freeman et al. [10], a kind of pixel neighbour correlation is included in the model as a way to smooth out the noise without bullring the real features. Hardie et al. [11] and Nguyen et al. [15] added a regularization term to the objective function, in which makes the resulting HR image smoother.

In this paper a new approach for including the neighbor correlation into the IBP model via Structural SIMilarity (SSIM) error measurement [19] is proposed. Recently by incorporating the *SSIM* into the objective function of Lucas-Kanade approach [14], this idea has been applied to image registration by us, successfully [1], .

The reminder of this paper is organized as follows: in section 2 we will have a brief look to IBP approach and SSIM error measurement criterion. In section 3 the proposed method and in section 4 experimental results are provided. The last section describes concluding remarks.

2 Previous Art

Since the proposed method is based on IBP approach and the SSIM criterion, at first we will have a quick review to these concepts.

2.1 Iterated Back Projection

The Iterated Back Projection approach uses averaged projections in HR grid to iteratively solve the HR image X , by minimizing the following error function [20]:

$$L(\underline{X}) = \frac{1}{2} \sum_{k=1}^N \|\underline{Y}_k - DHF_k \underline{X}\|_2^2 \quad (2)$$

Minimizing the error function has been done with taking the derivative of $L(\cdot)$, with respect to X ; the gradient of $L(\cdot)$ is the sum of gradients computed over the input images:

$$\underline{B}_k = F_k^T H^T D^T (DHF_k \underline{X} - \underline{Y}_k) \quad (3)$$

$$\nabla L(\underline{X}) = \sum_{k=1}^N \underline{B}_k \quad (4)$$

The estimated HR is updated in each iteration using gradient based minimization method:

$$\underline{X}^{n+1} = \underline{X}^n - \lambda \nabla L(\underline{X}) \quad (5)$$

where λ is a scale factor defining the step size. In each iteration, the HR estimate is re-sampled in the lattices of the input images. The difference between this re-sampled image and the input image is projected back to the HR lattice.

Zomet *et. al.* [20] replaced the sum of images in Eqn. (3) with a scaled pixel-wise median in their Robust Super-resolution (RS) approach:

$$\nabla L(\underline{X})(x, y) \approx n \cdot \text{median} \underline{B}_k(x, y)_{k=1}^n \quad (6)$$

2.2 Structural Similarity Error Measurement

Wang et al. [19] introduced a method for measuring the structural similarity of two images. The system separates the task of similarity measurement into three comparisons: luminance, contrast and structure. The luminance and contrast are estimated based on the mean and standard deviation of pixel intensities. The similarity is estimated based on the correlation between two image patches. Suppose x, y are two non-negative image signals, based on these three factors, Structural SIMilarity (*SSIM*) measurement of two image patches is defined as follows:

$$SSIM(x, y) = \frac{(2\mu_x \mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (7)$$

where C_1 and C_2 are some constants for avoiding instability, μ_x, σ_x and σ_{xy} are estimates of local statistics. SSIM closes to 1 when x, y became more similar.

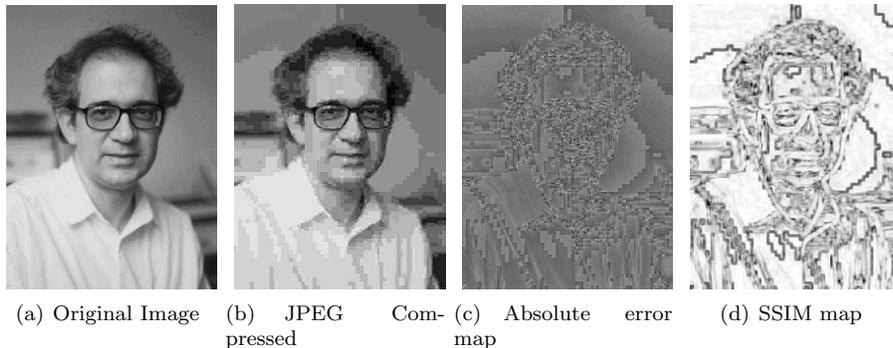


Figure 1: Comparing the error map of two images based on MSE and SSIM. The images are taken from [4].

The Mean Structural SIMilarity ($MSSIM$) is defined for structural error measurement of two images as follows:

$$MSSIM(X, Y) = \frac{1}{M} \sum_{j=1}^M SSIM(x_j, y_j) \quad (8)$$

Where X and Y are the reference and the distorted images, respectively; x_j and y_j are the image contents at the j^{th} local window; M is the number of local windows of the image. The higher values of $MSSIM$ mean more structural similarity of X and Y .

3 The Proposed Method

The contrast inverted form of SSIM highlights the structural differences of two images, much better than the absolute error map, in particular when one image is distorted. Figure 1 shows a reference image, its JPEG compressed version; the absolute error map and the SSIM map between the original and its distorted version. As can be seen the structural differences are more clear in the SSIM image map.

The chief idea of the proposed method is incorporating the $SSIM$ as a weighting term to the mentioned IBP formulation. Hence in addition to absolute error map of two images in eq. (2), the structural difference of two images is considered. Since the definition of SSIM is based on a block around of each pixel, the neighbor information of each pixel is included in the formulation, implicitly.

But the SSIM is for measuring similarity, while we look for a criterion for difference or dissimilarity of images. Hence here $SDIS$ as Structural DISsimilarity is defined as follows, which represents the difference of two images:

$$SDIS(x, y) = 1 - SSIM(x, y) \quad (9)$$

More structural difference leads to a higher value of $SDIS$. The error map of two images X and Y based on $SDIS$ is called $SDIS$ map image here and denoted by $E_{SDIS}(X, Y)$ or simply E_{SDIS} .

In eq. (5), the step size factor λ is equal for all of the image pixels; here we multiply it by E_{SDIS} , so the step size of gradient method will be variable, with respect to the structural differences of corresponding pixels. The max value of $SSIM$, when two images are identical, is equal 1; so the min value of $SDIS$ is zero. More structural differences of two images, lead to a higher value of $SDIS$; and thus have more contribution to $\Delta \underline{X}$.

We rewrite the eq. (5) as follows:

$$\underline{X}^{n+1} = \underline{X}^n - \lambda \sum_{k=1}^N [F_k^T H^T D^T E_{SDIS}(DHF_k X, Y_k)] \nabla L(\underline{X}) \quad (10)$$

In the above equation $E_{SDIS}(DHF_k X, Y_k)$, denotes the structural dis-similarity of LR frame Y_k , and the current estimation of HR image X , affected by the imaging model. The reverse of the imaging model should be run on E_{SDIS} so that the resulting matrix would be summable with \underline{X}^n .

For speed-up the process and having an equal formulation with the original IBP algorithm, we move the structural difference term from eq. (10) to the eq. (3)¹. Hence eq. (3) changes to:

$$\underline{B}_k = F_k^T H^T D^T E_{SDIS}(DHF_k X, Y_k)(DHF_k \underline{X} - \underline{Y}_k) \quad (11)$$

and the estimated HR is updated in each iteration as before:

$$\underline{X}^{n+1} = \underline{X}^n - \lambda \nabla L(\underline{X}) \quad (12)$$

where :

$$\nabla L(\underline{X}) = \sum_{k=1}^N \underline{B}_k \quad (13)$$

We name this approach as 'IBP-SSIM'².

The mentioned method can be combined with the robust super resolution, where in eq. (6), B_k , is computed based on the proposed method in eq. (11). We name this modified robust SR as 'RS-SSIM'.

¹The reason for speed-up gain of moving the structural difference term from eq. (10) to the eq. (3) is because that in the case of eq. (10), the reverse of the imaging model ($F_k^T H^T D^T$) should be run two times (in equations (10) and (3)); but with the mentioned modification, the reversing process appears only in eq. (3).

² E_{SDIS} does not included in the error function (2); because, in this case -as we will see in Appendix A- the term E_{SDIS} vanishes in the update stage. Hence this term is multiplied directly in the gradient term in eq. (10).

Table 1: Description of test sequences. The differences between HR and LR frames are clear with zooming on the electronic version of the paper.

Sequence Name:	Tehran Park	Tokyo	LSMS Closing
First HR Frame			
Frames	69	60	4
Resolution	360×288	320×240	512×384
Device:	Panasonic NV-GS75	Sony HDR-SR12E	Sony DSC-W30
First LR Frame			
Resolution	180×144	160×120	256×192
Blurred	Yes	Yes	Yes
Down-sampled	Yes	Yes	Yes
Noisy	Yes	Yes	No

4 Experimental Results

We categorized the experiments into two cases. In the first case we use some real video sequences, which are corrupted with noise for quantitative comparisons. In the second part we will use 4 synthesized LR image from one HR image for visual comparison of the proposed method against some others. The test sequences are illustrated in table 1.

4.1 Quantitative Comparison

We applied our modified IBP algorithms on ‘Tehran Park’ and ‘Tokyo’ sequences and compared its performance with the original algorithms. Each sequence is blurred, down-sampled and corrupted with noise for producing low-resolution video frames with SNR value equal 40db.

For video enhancement by SR algorithms, we used the ‘sliding-window’ technique [3, 2] with the Iterated Back-Projection (IBP) [12] and Robust Super-resolution (RS)[20] and their modifications (IBP-SSIM and RS-SSIM) as reconstruction methods. Computing the motion parameters between frames has been done using the registration method of Keren *et. al.*[13] and the initial conditions are the same for all of the methods. The implementations has been done in MATLAB. The magnification factor r and the window size were set to 2 and 4 respectively.

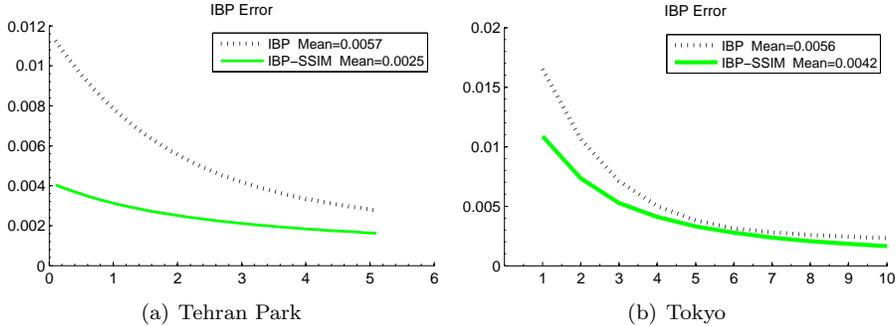


Figure 2: The average error over image sequences in each iteration.

4.1.1 Convergence Comparison

The back projection algorithm cycles until reaches to a predefined maximum iterations allowed or $\|e\| > \epsilon$, where:

$$e = \frac{\|\underline{X}^i - \underline{X}^{i-1}\|}{\|\underline{X}^i\|} \quad (14)$$

Figure 2 shows the average error over video frames in each iteration for two mentioned sequences. For each video, the average of $\|e\|$ over all of its frames for each iteration is computed and illustrated. To avoid the results being biased by cases when one method converges sooner, we included in the averaging only those iterations, in which both of the two algorithms are not terminated. Better convergence of the proposed approach for both sequences, is obvious.

The convergence plots of RS and RS-SSIM – not shown here - was similar to the mentioned results.

4.1.2 Running Time Comparison

Since the computing of *SSIM* is not complicated, it does not increase the overall time significantly. For instance in average each cycle of the IBP and IBP-SSIM algorithms for ‘Tehran Park’ sequence, took 1.055 and 1.085 seconds, respectively; which means about 2.9% increase of running time. In other words, the time taken by IBP algorithm for 15 cycles is equal to required time of IBP-SSIM approach for 14.58 cycles. Hence the minor overhead time of computing *SSIM* is negligible.

4.1.3 Comparing with different supper-resolution methods

Table 2 shows quantitative comparisons of the mentioned methods based on Mean Absolute Error (MAE), Power Signal to Noise Ratio (PSNR) and SSIM

Table 2: MAE, PSNR and SSIM comparisons of the original IBP and RS methods and their modifications based on SSIM over different sequences. The best score is highlighted with **Bold** letters in each row and for each category.

	Iterated Back Projection		Robust Super Resolution	
Video	IBP	IBP-SSIM	RS	RS-SSIM
	MAE($\times 10^{-3}$)			
Tehran Park	45.91	37.54	46.72	39.47
Tokyo	44.83	38.65	42.75	35.80
	PSNR			
Tehran Park	23.05	24.84	22.76	24.21
Tokyo	22.74	24.09	23.09	24.69
	SSIM			
Tehran Park	0.67	0.74	0.65	0.70
Tokyo	0.62	0.66	0.63	0.69

for the test sequences, in which:

$$MAE = \frac{\sum_{i=1}^N \sum_{j=1}^M \sum_{q=1}^Q |F^q(i, j) - \hat{F}^q(i, j)|}{N.M.Q} \quad (15)$$

and

$$PSNR = 10 \times \log \left(\frac{255^2}{\frac{1}{N.M.Q} \sum_{i=1}^N \sum_{j=1}^M \sum_{q=1}^Q (F^q(i, j) - \hat{F}^q(i, j))^2} \right) \quad (16)$$

where M , N are the image dimensions, Q is the number of channels of the image ($Q = 3$ for color image), and $F^q(i, j)$ and $\hat{F}^q(i, j)$ denote the q th component of the original image vector and the distorted image, at pixel position (i, j) , respectively. In these experiments, the mentioned criteria has been computed over gray scale version of images ($Q=1$).

The best score is highlighted with **Bold** letters for each sequence and for each category (Iterated Back Projection and Robust Super Resolution), in table 2. As shown in table 2 the proposed methods (IBP-SSIM and RS-SSIM) have the highest performance for all MAE, PSNR and SSIM criteria.

The ‘Tehran Park’ sequence has been used here for further comparisons. Figure 3 shows MAE, PSNR and SSIM of the original IBP approach and the proposed method over 69 frames of ‘Tehran Park’ sequence. The superior performance of the proposed modification of original IBP is obvious. The experimental results for ‘Tokyo’ sequence was alike to ‘Tehran Park’ sequence. Similar results have been achieved for robust super resolution and its modification (RS-SSIM).

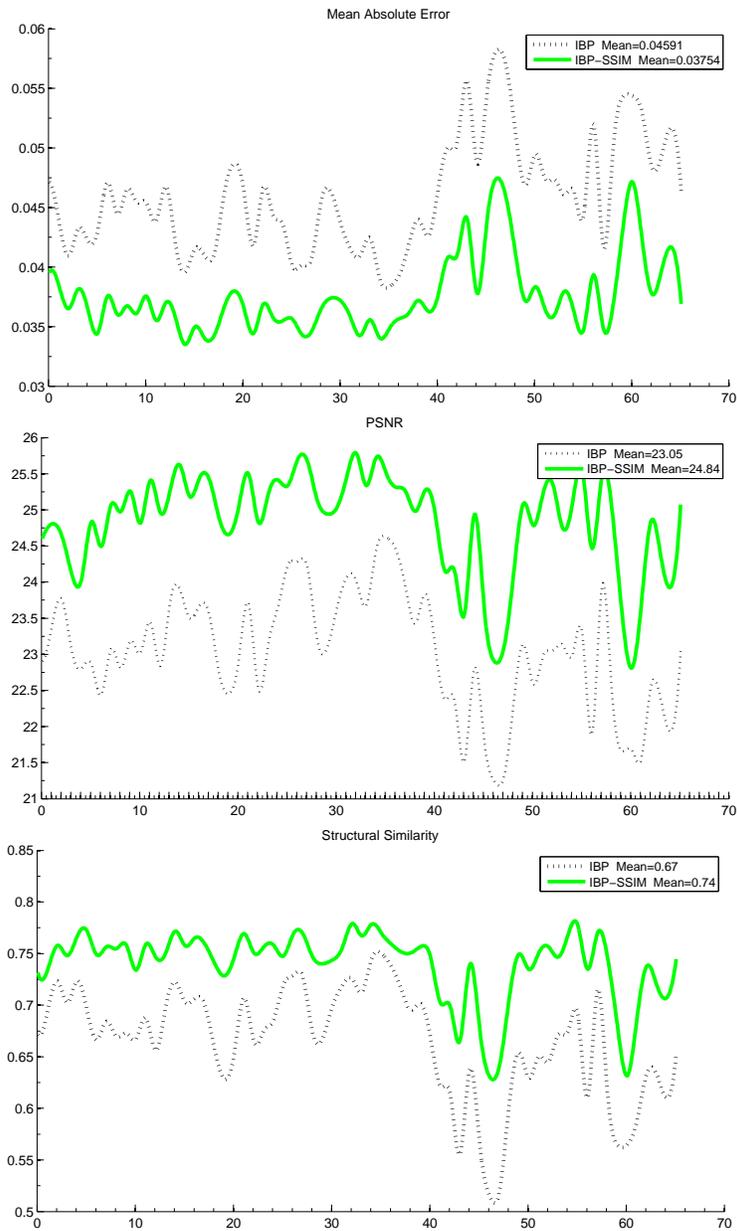


Figure 3: MAE, PSNR and SSIM comparison of the original IBP method ('IBP') and its modification based on SSIM ('IBP-SSIM') for 'Tehran Park' sequence.

4.2 Visual Comparison

The previous two test sequences do not have good details for visual comparison purposes. Hence we synthesized 4 LR images from a HR image shown in table 1 as ‘LSMS Closing’ sequence . Each frame in addition to blurring and down-sampling has some differences about horizontal and vertical shifts and rotation angles relating to the HR image .

Figure 4 shows close-up demonstrations of an instance frame, produced by some different methods. The original HR frame, the nearest and the bicubic resized versions of that frame has been shown for comparison purposes. Note the better quality of the proposed method (4(e)) with respect to 4(d) due to incorporating of the neighbor pixels via SSIM criterion. The *SDIS error map image* between two results shown in figures 4(d) and 4(e) and the HR image 4(a) is shown in figures 4(f) and 4(g). Brighter pixels indicate more structural differences. As can be seen the proposed method has less differences with the original HR image.

5 Concluding Remarks

Iterated Back Projection (IBP) and its successor, Robust Super resolution (RS) are two famous multi-frame super resolution reconstruction methods which use the Sum of Squared Differences (SSD) of two images as error function. The SSD between the low resolution frames and down-sampled of the estimated high resolution image does not reflect the pixels’ neighboring correlation.

In this paper the Structural Similarity of two images, which is based on the local statistical information of each pixel, is contributed to the formulation of IBP and RS methods. In contrast to the original approaches, in which the step size of the gradient based minimization method is fixed and equal for all of the pixels, here it is variable. The step size of each pixel of estimated high resolution image is affected by the structural differences of two corresponding patches. Hence more structural differences leads to a larger step size in updating the estimated high resolution image and consequently faster convergence in the proposed approach. Various objective and subjective comparisons showed the superior performance of the proposed modifications over original IBP and RS approaches.

Acknowledgment

The authors are indebted to Dr. Vandewalle [18] for his Super-Resolution package, Dr. Wang [19] for his *SSIM* function and Dr. Gh. Minaei for providing ‘Tokyo’ sequence.



(a) Original HR frame



(b) LR frame (Nearest)



(c) LR frame (Bicubic)



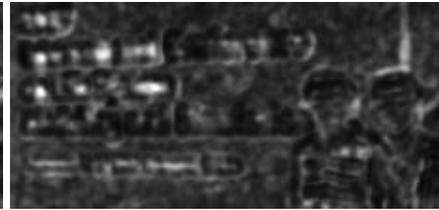
(d) Iterated Back-projection



(e) IBP-SSIM



(f) IBP's SDIS error map



(g) IBP-SSIM's SDIS error map

Figure 4: Close-up views of the original HR image, replication (nearest) and bicubic resized versions of the first LR image, super-resolution reconstruction methods: Iterated Back-projection[12] and the proposed method (IBP-SSIM) on 'LSMS Closing' sequence. The last row illustrates *SDIS error map image* between two results shown in the third row and the HR image. Brighter pixels indicate more structural differences. The lower error of the proposed approach is obvious, specially near the edges.

References

- [1] M. Amintoosi, M. Fathy, and N. Mozayani. Precise image registration with structural similarity error measurement applied to super-resolution. *EURASIP Journal on Applied Signal Processing*, 2009:7 pages, 2009. Article ID 305479. [2](#)
- [2] M. Amintoosi, M. Fathy, and N. Mozayani. Video resolution enhancement in the presence of moving objects. In *International Conference on Image Processing, Computer Vision, and Pattern Recognition*, pages 456–460, Las Vegas, USA, July 2009. [6](#)
- [3] Sean Borman. *Topics in Multiframe Superresolution Restoration*. PhD thesis, University of Notre Dame, Notre Dame, IN, May 2004. [6](#)
- [4] Alan Brooks. What makes an image look good?, March 17 2005. URL http://dailyburrito.com/projects/ImageQuality_Brooks2005.pdf. Presentation for the Image & Video Processing Lab (IVPL) at Northwestern University for ECE 510 Video Processing. [4](#)
- [5] Peter Cheeseman, Bob Kanefsky, Richard Kraft, John Stutz, and Robin Hanson. Super-resolved surface reconstruction from multiple images. In Glenn R. Heidbreder, editor, *Maximum Entropy and Bayesian Methods*, pages 293–308. Kluwer Academic Publishers, Dordrecht, the Netherlands, 1996. [2](#)
- [6] M. Elad and A. Feuer. Restoration of single super-resolution image from several blurred, noisy and downsampled measured images. *IEEE Trans. Image Processing*, 6:1646–1658, Dec 1997. [1](#)
- [7] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar. Robust shift and add approach to super-resolution. In *Proc. of the 2003 SPIE Conf. on Applications of Digital Signal and Image Processing*, pages 121–130, Aug 2003. [1](#), [2](#)
- [8] Sina Farsiu. *A Fast and Robust Framework for Image Fusion and Enhancement*. PhD thesis, Electrical Engineering, UC Santa Cruz, December 2005. [2](#)
- [9] Sina Farsiu, M. Dirk Robinson, Michael Elad, and Peyman Milanfar. Fast and robust multiframe super resolution. *IEEE Transactions on Image Processing*, 13(10):1327–1344, 2004. [2](#)
- [10] William T. Freeman, Thouis R. Jones, and Egon C Pasztor. Example-based super-resolution. *IEEE Comput. Graph. Appl.*, 22(2):56–65, 2002. [2](#)
- [11] R.C. Hardie, K.J. Barnard, and E.E. Armstrong. Joint map registration and high resolution image estimation using a sequence of undersampled images. *IEEE Trans. on Image Processing*, 6(12):1621–1633, December 1997. [2](#)
- [12] Michal Irani and Shmuel Peleg. Improving resolution by image registration. *CVGIP: Graph. Models Image Process.*, 53(3):231–239, 1991. [2](#), [6](#), [11](#)
- [13] D. Keren, S. Peleg, and R. Brada. Image sequence enhancement using sub-pixel displacement. In *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 742–746, 1988. [6](#)

- [14] B.D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *IJCAI81*, pages 674–679, 1981. [2](#)
- [15] N. Nguyen, P. Milanfar, and G.H. Golub. A computationally efficient superresolution image reconstruction algorithm. *IEEE Trans. on Image Processing*, 10(4): 573–583, April 2001. [2](#)
- [16] Tuan Q. Pham. *Spatiotonal Adaptivity in Super-Resolution of Under-sampled Image Sequences*. PhD thesis, aan de Technische Universiteit Delft, 2006. [2](#)
- [17] Tuan Q. Pham, Lucas J. van Vliet, and Klamer Schutte. Robust fusion of irregularly sampled data using adaptive normalized convolution. *EURASIP Journal on Applied Signal Processing*, pages ID 83268, 12 pages, 2006. [2](#)
- [18] Patrick Vandewalle, Sabine Süsstrunk, and Martin Vetterli. A Frequency Domain Approach to Registration of Aliased Images with Application to Super-Resolution. *EURASIP Journal on Applied Signal Processing (special issue on Super-resolution)*, 2006:Article ID 71459, 14 pages, 2006. [10](#)
- [19] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Processing*, 13(4):600–612, April 2004. [2](#), [3](#), [10](#)
- [20] A. Zomet, A. Rav-Acha, and S. Peleg. Robust super resolution. In *Proceedings of the Int. Conf. on Computer Vision and Patern Recognition (CVPR)*, pages 645–650, Dec 2001. [2](#), [3](#), [6](#)

A Why *SSIM* does not contribute into error function, $L(\underline{X})$

In eq. (10), *SSIM* was multiplied by the gradient term, here we discuss about why *SSIM* does not contribute into the objective function of IBP approach.

Let we multiply *SSIM* into the error function, hence eq. (2) is rewritten as follows:

$$L(\underline{X}) = \frac{1}{2} \sum_{k=1}^N \underline{E}_{SDIS} \cdot \|\underline{Y}_k - DHF_k \underline{X}\|_2^2 \quad (17)$$

where dot denotes the element by element multiplication as ‘.*’ operator in MATLAB. For minimizing (17) in an iterative manner, we have to minimize the following function:

$$L(\underline{X} + \Delta \underline{X}) = \frac{1}{2} \sum_{k=1}^N \underline{E}_{SDIS} \cdot [\underline{Y}_k - DHF_k(\underline{X} + \Delta \underline{X})]^2 \quad (18)$$

where \underline{E}_{SDIS} evaluates the structural dissimilarity of \underline{Y}_k and $DHF_k \underline{X}$. Finding the optimum value of $\Delta \underline{X}$ can be done by differentiating (18) with respect to $\Delta \underline{X}$, setting the result to equal zero and solving it:

Derivation of this new error function with respect to $\Delta \underline{X}$, yields:

$$\begin{aligned} \frac{\partial L}{\partial \Delta \underline{X}} &= \sum_{k=1}^N F_k^T H^T D^T (E_{SDIS} \cdot [DHF_k(\underline{X} + \Delta \underline{X}) - \underline{Y}_k]) \\ &= \sum_{k=1}^N F_k^T H^T D^T (E_{SDIS} \cdot [DHF_k \Delta \underline{X} + (DHF_k \underline{X} - \underline{Y}_k)]) \end{aligned} \quad (19)$$

$$\frac{\partial SSD}{\partial \Delta \underline{X}} = 0 \Rightarrow$$

$$\sum_{k=1}^N F_k^T H^T D^T E_{SDIS} \cdot DHF_k \Delta \underline{X} + \sum_{k=1}^N F_k^T H^T D^T E_{SDIS} \cdot (DHF_k \underline{X} - \underline{Y}_k) = 0 \quad (20)$$

Hence we have:

$$\sum_{k=1}^N F_k^T H^T D^T E_{SDIS} \cdot DHF_k \Delta \underline{X} = - \sum_{k=1}^N F_k^T H^T D^T E_{SDIS} \cdot (DHF_k \underline{X} - \underline{Y}_k) \quad (21)$$

dropping of $F_k^T H^T D^T E_{SDIS}$ from two side of eq. (21) and multiplying each side of the result by $F_k^T H^T D^T$ leads to:

$$\Delta \underline{X} = - \sum_{k=1}^N F_k^T H^T D^T (DHF_k \underline{X} - \underline{Y}_k) \quad (22)$$

which is as same as eq. (3). Note that (i) the dot operator after E_{SDIS} indicates element by element multiplication and (ii) the negative sign is included in eq. (5).

Thus the term E_{SDIS} is vanished in the update stage. Hence instead of contributing the $SSIM$ into the error function, it is multiplied directly in the gradient term in eq. (10).