AN ADAPTIVE L_1-L_2 HYBRID ERROR MODEL TO SUPER-RESOLUTION

Huihui Song [a], Lei Zhang [b], Peikang Wang [a], Kaihua Zhang [b] and Xin Li [c]

[a] Dept. of EEIS, University of Science and Technology of China, Hefei 230027, P. R. China
[b] Dept. of Computing, The Hong Kong Polytechnic University, Hong Kong
[c] Lane Dept of CSEE, West Virginia University

ABSTRACT

A hybrid error model with L_1 and L_2 norm minimization criteria is proposed in this paper for image/video super-resolution. A membership function is defined to adaptively control the tradeoff between the L_1 and L_2 norm terms. Therefore, the proposed hybrid model can have the advantages of both L_1 norm minimization (i.e. edge preservation) and L_2 norm minimization (i.e. smoothing noise). In addition, an effective convergence criterion is proposed, which is able to terminate the iterative L_1 and L_2 norm minimization process efficiently. Experimental results on images corrupted with various types of noises demonstrate the robustness of the proposed algorithm and its superiority to representative algorithms.

***Index Terms*—**Super-resolution, L_1 norm, L_2 norm, convergence criterion.

1. INTRODUCTION

Image interpolation and superresolution (SR) aims to reconstruct high-resolution (HR) images from their low-resolution (LR) counterparts [1-3, 9-10]. In particular multi-frame video SR could reproduce HR frames from a sequence of LR frames that have sub-pixel shifts [3]. In the past decades, SR reconstruction has been widely studied and various algorithms have been developed in applications such as image/video resolution enhancement, medical imaging, remote sensing and video surveillance, etc., [1]. Usually, the LR frames need to be previously registered via motion estimation, and the point spread function (PSF) estimation and photometric correction, etc., need to be performed in advance [2][3][5]. A mathematical model is then used to bridge the LR observations with the unknown HR scene. Finally, Maximum Likelihood (ML) estimation techniques are often used to solve the SR inverse problem to reconstruct the HR scene [2][5][6][7].

Most of the existing ML estimators assume that the noise in the LR observations is Gaussian distributed [6][7]. However, for many real world image sequences, the Laplacian distribution is more accurate to model the impulsive noise (such as the salt and pepper noise) inside them [2][5]. Generally speaking, the ML estimators for signals with Gaussian distributed noise can be seen as mean filters, which have good performance at smoothing noise (but may smooth edges as well), while the ML estimators for signals with Laplacian distributed impulsive noise corresponds to the median filters, which has good performance at preserving the image edges (but may not be able to smooth out Gaussian additive noise) [2][4].

Often the noise corrupted in real image sequence is a combination of additive Gaussian noise and impulsive Laplacian noise. In order to better suppress them in the SR image reconstruction process, in this paper we propose a hybrid error model which incorporates the advantages of the ML estimators for signals with Gaussian additive and Laplacian impulsive noise. The estimators are adaptively tuned according to the noise distribution. Note that in the iterative SR reconstruction process, the noise and its distribution will also change. Moreover, we propose an effective adaptive convergence criterion (ACC), which is able to terminate the iteration efficiently. The experiments demonstrate that the proposed method can effectively reconstruct the HR images from the LR frames, preserving well the image details while suppressing the noise.

Section 2 introduces the background of the algorithm. Section 3 describes our method in detail. Section 4 presents experimental results and Section 5 concludes the paper.

2. THE IMAGING MODEL

In this paper, we adopt the following linear observation model for LR image formation [2][3]

$$y^{(k)} = D_k H_k F_k x + N_k,\ k = 1,...,K. \quad (1)$$

where K is the number of the LR frames, x is the HR frame, and $y^{(k)}$ is the k^{th} LR frame. The matrices D_k and H_k represent the sensor spatial sampling and the system PSF, respectively. F_k is the geometric motion operator between $y^{(k)}$ and x, and N_k is the system noise.

The motion model and the PSF need to be well estimated to register the LR images. The planar projective (8DoF) motion model [3] is very suitable to model planar objects captured from a variety of angles, even for the case that the camera centre rotates about its optical centre. The motions in small regions and short time sequences can also be adequately approximated by the 8DoF transform even when the true underlying motion of the overall scene is not completely described. The PSF can be decomposed into factors representing the blurring caused by camera optics and the spatial integration by CCD sensor. As in many literatures [3], here we set the PSF as an isotropic 2D Gaussian kernel with variance σ_{PSF}^2 .

3. METHODOLOGY

3.1. Problem Formulation

The noise N_k in Eq. (1) is often assumed to be Gaussian or Laplacian distributed, and the corresponding HR ML estimate can be calculated by the following L_p norm minimization criterion [2]

$$\hat{x} = \arg\min_{x} \left(\sum_{k=1}^{N} \| y^{(k)} - D_k H_k F_k x \|_p^p \right) \quad , p = 1, 2. \tag{2}$$

where p = 1 corresponds to Laplacian distribution, and p = 2 corresponds to Gaussian distribution.

Since super-resolution is an ill-posed problem, the solution to Eq. (2) is not unique. Some regularization methods are often used to yield a stable solution. In [2], Farsiu *et al.* proposed a so-called Bilinear Total Variation (BTV) regularization method, which is not only computationally efficient, but also good at edge preservation. This regularization method is employed in the proposed hybrid error model.

3.2. The Hybrid Error Model (HEM)

The probability density functions (PDF) of Gaussian and Laplacian noise are respectively as follows

$$\begin{cases} P_G(r) = \dfrac{1}{\left(2\pi\sigma_G^2\right)^{N/2}} \exp\left(-\sum_{i=1}^{N} \left(r(i) - m_G\right)^2 \Big/ 2\sigma_G^2 \right), \\ P_L(r) = \dfrac{1}{\left(2\sigma_L\right)^N} \exp\left(-\sum_{i=1}^{N} \left| r(i) - m_L \right| \Big/ \sigma_L \right), \end{cases} \tag{3}$$

where r denotes an N-dimensional noise vector. The ML estimates of σ_G, m_G, σ_L, and m_L, denoted by $\hat{\sigma}_G, \hat{m}_G, \hat{\sigma}_L, \hat{m}_L$, can be computed as follows

$$\begin{cases} \hat{\sigma}_G = \sqrt{\dfrac{\sum_{i=1}^{N} \left(r(i) - \hat{m}_G\right)^2}{N}}, \hat{m}_G = \text{mean}(r), \\ \hat{\sigma}_L = \dfrac{\sum_{i=1}^{N} \left| r(i) - \hat{m}_L \right|}{N}, \ \hat{m}_L = \text{median}(r). \end{cases} \tag{4}$$

We use the following *generalized likelihood ratio test* (*GLRT*) [5] to decide which noise model is preferable:

$$\frac{P_G(r; \hat{\sigma}_G, \hat{m}_G)}{P_L(r; \hat{\sigma}_L, \hat{m}_L)} > 1. \tag{5}$$

Combining Eqs.(3), (4), (5), and defining $\gamma \triangleq \hat{\sigma}_L / \hat{\sigma}_G$, we have

$$\gamma > \sqrt{\pi / 2e} \approx 0.7602. \tag{6}$$

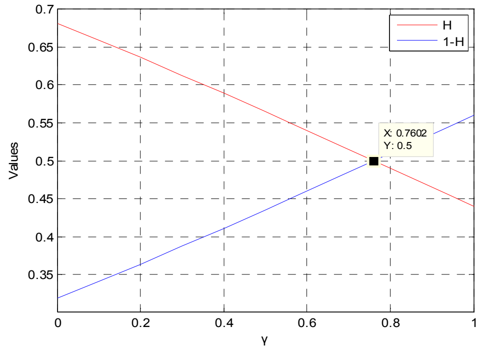


Fig.1. Membership functions of Gaussian model (blue solid curve) and Laplacian model (red solid curve).

Therefore, if the ratio γ > 0.7602, the Gaussian model is preferable to the Laplacian one, and vice versa.

If the Gaussian model is selected, then in Eq. (2) the L_2-norm is used, while if the Laplacian model is selected, then in Eq. (2) the L_1-norm is used. Both the L_2-norm and L_1-norm minimization have their own advantages. The L_2-norm for Gaussian noise model have a good performance of smoothing image, while the L_1-norm for Laplacian noise model can well preserve the details of image, such as edges. In order to adaptively balance between the image smoothing and edge preserving, we define the following membership function for the Laplacian distribution, which adaptively changes according to the ratio γ (see the red curve in Fig.1).

$$H(\gamma) \triangleq \frac{1}{1 + \exp(\gamma - 0.7602)}. \tag{7}$$

Obviously, the membership function corresponding to the Gaussian model is $1-H(\gamma)$ (see the blue curve in Fig.1).

The objective function of the data term of our model is as follows:

$$L_e = \left(1 - H(\gamma)\right) \sum_{k=1}^{N} \tfrac{1}{2\hat{\sigma}_G^2} \| r^{(k)} \|_2^2 + H(\gamma) \sum_{k=1}^{N} \tfrac{1}{\hat{\sigma}_L} \| r^{(k)} \|_1^1, \tag{8}$$

where $r^{(k)} = y^{(k)} - D_k H_k F_k x$.

In order for a stable solution and for edge preservation, the Bilinear Total Variation (BTV) [2] is used here as the regularization term, whose objective function is as follows

$$L_r = \sum_{l,m=-P}^{P} \alpha^{(|l|+|m|)} \| x - S_x^l S_y^m x \| \tag{9}$$

where $0 < \alpha < 1$, P is a fixed parameter, S_x^l and S_y^m are matrices accounting for the shift operations in the horizontal and vertical directions by l and m pixels, respectively. The final objective function is as follows

$$L = L_e + \nu L_r, \tag{10}$$

where ν is a regularization parameter that weights the first term against the second term.

The objective function in Eq. (10) can be solved by an iterative procedure. The reconstructed SR image in previous iteration is an initialization for the current iteration, and the reconstructed SR image will gradually become better. So the intensity of the noise in the reconstructed image will change during the iteration procedure. The ratio γ in Eq. (6) can be used to indicate this variance. The membership functions w.r.t the ratio γ, which correspond to the Gaussian and Laplacian models, respectively, can adaptively select which model is more suitable during the iteration procedure.

3.3. Adaptive Convergence Criterion (ACC)

We use the scaled conjugate gradient (SCG) algorithm [8] to optimize the objective function in Eq. (10), which is computationally efficient. Usually, the convergence is assumed to be reached by setting a maximum iteration number N_{max} and comparing the values of the objective function between two successive iterations. However, this method is not general and sometimes difficult to choose proper threshold, which limits its application. Moreover, from experiments, we found it is unnecessary to stop until N_{max} is reached, since the reconstructed SR image can be good enough before N_{max} is reached.

We propose an adaptive convergence criterion (ACC) by using the change of ratio γ in Eq. (6) as an indicator, which is robust and very efficient. If the absolute value variance of the ratio γ is successively less than a fixed threshold ε more than four times, the convergence is assumed to be reached.

The pseudocode of the convergence algorithm is as follows.

Algorithm of ACC

While iteration times < N_{max}, do
  Compute the absolute value variance of γ during two successive iterations, i.e., $abf\gamma = |\gamma_{back} - \gamma_{forword}|$.
  If *abfy* is successively less than ε more than four times
    End iteration.
  Else, continue until the iterations reach N_{max}.

4. EXPERIMENTAL RESULTS

In the experiments we consider three common types of noise in SR: Gaussian noise, impulse noise (e.g. salt and pepper noise), and the mixed noise of them including estimation error. We name the L_1-norm with BTV regularized method L1BTV [2], while the L_2-norm with BTV regularized method L2BTV [3]. The code of the proposed method can be downloaded at http://www4.comp.polyu.edu.hk/~cslzhang/code.htm.

In the first experiment, we validate the ACC (adaptive convergence criterion) algorithm by comparing it with the commonly used method in [3]. We added the salt and pepper noise to the Lena image with intensity of 0.08 and recover it using the L1BTV method. Fig.2 (a) shows the difference of the objective function (in Eq.(10)) between two successive iterations. When the maximum iterations N_{max} is reached, the difference between the last two successive iterations is still greater than 10^{-2}. In other words, the convergence cannot be reached if we set the threshold $\varepsilon < 10^{-2}$. As Fig.2(c) shows, the difference of ratio γ can be very small before N_{max} is reached. In this experiment, we set $\varepsilon = 10^{-5}$, and according to the algorithm of ACC, iteration ends when the iteration number is 14. The reconstructed SR image by ACC is shown in Fig.2 (d), while the result by L1BTV is shown in Fig.2 (b) when N_{max} is reached. The RMSE (root mean square error) of the two results are the same (7.05) but the proposed ACC method can save much computational cost.

In the second experiment, we evaluated the performance of our method based on the L_1-L_2 hybrid norm by removing the Gaussian and the salt and pepper noises. A sequence of five LR frames is generated from the HR Lena image as follows. We applied random affine transformation (8DoF) and Gaussian blurring to the HR Lena image, and then down-sampled it with a factor of two. Finally, we added Gaussian noise and salt and pepper noise to the down-sampled images respectively. By adding different Gaussian noise, the signal-to-noise ratios (SNR) are 14dB, 10dB and 7.6dB, respectively, while the salt and pepper noise have different densities of 0.06, 0.08 and 0.1. Table 1 lists the RMSE results by the proposed hybrid error model (HEM) and the L1BTV method. It can be seen that the proposed method outperforms the L_1-norm method in reducing noise. Fig. 3 shows the results by the two methods when the salt and pepper noise is added (density is 0.08).

Table 1.The RMSE results

Noise	Gaussian			Salt&pepper		
	14(dB)	10(dB)	7.6(dB)	0.06	0.08	0.1
L1BTV	6.55	7.64	8.69	8.69	7.05	9.74
HEM	**6.24**	**7.31**	**8.46**	**8.09**	**5.59**	**7.72**

In the last experiment, a real benchmark video sequence (the Foreman sequence) with unknown camera PSF and transformation information is used to demonstrate the superiority of our method to the L2BTV and L1BTV methods. We extracted 5 frames (frames from 52 to 56) with size 144×176 from the video to reconstruct the middle frame (frame 54) with resolution enhancement factor of two. These 5 frames approximately follow the global translational and rotation motion model, both in the man and the background. For simplicity, we just estimated the three parameters of translation and rotation in the 8DoF matrix, i.e., the horizontal and vertical translations, and the angle of the rotation. We applied transformation with various scale parameters until the correlation between the middle and the aligned image is acceptable to compute the assumed scale parameters. The unknown camera PSF is assumed to be a normalized 5×5 Gaussian kernel with zero mean and variance 0.5, which can yield a good result in our experiment. The original middle frame is shown in Fig.4 (a). The results of the super-resolution reconstruction using the three methods are depicted from Fig.4 (b) to (d). In the three methods, we tuned parameters to get the best visual results. From Fig.4 (b), we can observe that the L2BTV method loses many subtle details (e.g. the eye part), and the L1BTV method yields some chessboard effects, while the HEM method can yield a smoother result without losing distinct details. This is the contribution of the adaptive adjustment of the L_2-norm and the L_1-norm.

5. CONCLUSION

This paper presented a novel hybrid error model for SR reconstruction, which combines the advantages of Gaussian model and Laplacian model. The hybrid error model integrates the Gaussian and Laplacian models by their corresponding membership functions, which are varying according to the noise intensity distribution during the iteration procedure. Moreover, an adaptive convergence criterion (ACC) was proposed, which can effectively and efficiently end the iteration. Comparisons with the L1BTV and L2BTV on images with different noises demonstrated the superiority of the proposed algorithm.

ACKNOWLEDGEMENT

This research is supported by the Hong Kong RGC General Research Fund (PolyU 5375/09E).

6. REFERENCES

[1] S.Park, M.Park, and M.Kang, “Supper-Resolution Image Reconstruction: A Technical Overview”, *IEEE Signal Processing Magazine*, vol.20, pp.21–36, 2003.

[2] S. Farsiu, M. Robinson, M. Elad, and P. Milanfar, “Fast and robust multiframe super resolution”, *IEEE T-IP*, vol.13, pp.1327–1344, Oct. 2004.

[3] L.Pickup, “Machine Learning in Multi-frame Image Super-resolution”, *Ph.D. Thesis,* University of Oxford, England, 2007.

[4] W.Burger and M.Burge, “Digital Image Processing”, Springer London, 2008.

[5] S.Farsiu, D.Robinson, M. Elad, and P. Milanfar, “Robust shift and add approach to super-resolution”, *Proc. SPIE,* vol. 5203, 121, 2003.

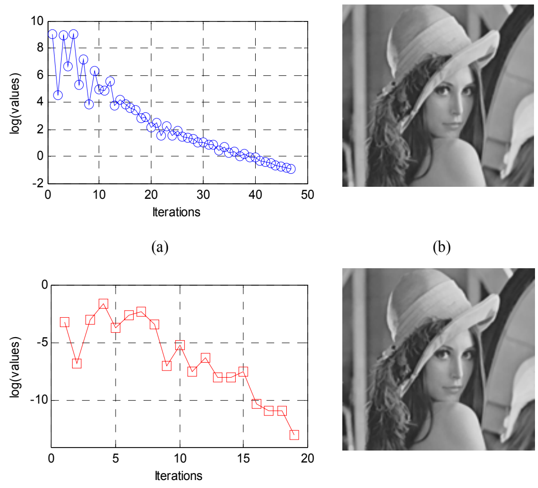
[6] M.Elad and Y.Hel-Or, “A fast super-resolution reconstruction algorithm for pure translational motion and common space invariant blur”, *IEEE T-IP*, vol.10, pp.1187-1193, 2001.

[7] S.Babacan, R.Molina, A.Katsaggelos, “Total Variation Super Resolution Using A Variational Approach”, *ICIP*, pp.641-644, 2008.

[8] I.Nabney, “Netlab algorithms for pattern recognition”, Springer, 2002.

[9] X. Li and M. Orchard, “New Edge-Directed Interpolation ,” *IEEE Trans. on Image Processing*, vol. 10, no. 10, pp. 1521-1527, Oct. 2001.

[10] L. Zhang and X. Wu, “An edge-guided image interpolation algorithm via directional filtering and data fusion,” *IEEE Trans. on Image Processing,* vol. 15, pp. 2226-2238, Aug. 2006.



(c) (d)

Fig.2. Demonstration of the effectiveness of ACC. (a): The difference of objective function between two successive iterations. (c): The difference of ratio γ between two successive iterations. (b): result when N_{max} is reached (RMSE = 7.05). (d): result of ACC (RMSE = 7.05).

Fig. 3. Comparisons between L1BTV and our method on images with salt and pepper noise. Left column: one of the LR images. Middle column: result of L1BTV (RMSE = 7.05). Right column: result of our method (RMSE = 5.59). Noise intensity is 0.08.

(a) The 54^{th} frame of the foreman sequence (b) result of L2BTV (c) result of L1BTV (d) result of HEM

Fig. 4. Comparisons for the methods of L2BTV, L1BTV, and HEM.