

2010

Wavelet based nonlocal-means super-resolution for video sequences

H Zheng

University of Wollongong, hz384@uow.edu.au

A Bouzerdoun

University of Wollongong, bouzer@uow.edu.au

S L. Phung

University of Wollongong, phung@uow.edu.au

Publication Details

H. Zheng, A. Bouzerdoun & S. L. Phung, "Wavelet based nonlocal-means super-resolution for video sequences," in 2010 IEEE 17th International Conference on Image Processing (ICIP), 2010, pp. 2817-2820.

Wavelet based nonlocal-means super-resolution for video sequences

Abstract

Video sequence resolution enhancement became a popular research area during the last two decades. Although traditional super-resolution techniques have been successful in dealing with image sequences, many constraints such as global translation between frames, have to be imposed to obtain good performance. In this paper, we present a new wavelet-based nonlocal-means (WNLM) framework to bypass the motion estimation stage. It can handle complex motion changes between frames. Compared with the nonlocal-means (NLM) super-resolution framework, the proposed method provides better result in terms of PSNR and faster processing.

Keywords

sequences, wavelet, video, resolution, super, means, nonlocal

Disciplines

Physical Sciences and Mathematics

Publication Details

H. Zheng, A. Bouzerdoum & S. L. Phung, "Wavelet based nonlocal-means super-resolution for video sequences," in 2010 IEEE 17th International Conference on Image Processing (ICIP), 2010, pp. 2817-2820.

WAVELET BASED NONLOCAL-MEANS SUPER-RESOLUTION FOR VIDEO SEQUENCES

H. Zheng, A. Bouzerdoun, *IEEE Senior Member*, and S. L. Phung, *IEEE Member*

School of Electrical, Computer and Telecommunications Engineering
University of Wollongong
Wollongong, NSW 2522, Australia

ABSTRACT

Video sequence resolution enhancement became a popular research area during the last two decades. Although traditional super-resolution techniques have been successful in dealing with image sequences, many constraints such as global translation between frames, have to be imposed to obtain good performance. In this paper, we present a new wavelet-based nonlocal-means (WNLM) framework to bypass the motion estimation stage. It can handle complex motion changes between frames. Compared with the nonlocal-means (NLM) super-resolution framework, the proposed method provides better result in terms of PSNR and faster processing.

Index Terms— super-resolution, wavelet transform, nonlocal-means, probabilistic model, motion estimation

1. INTRODUCTION

The purpose of super-resolution (SR) image reconstruction is to combine a set of low resolution (LR) images (or frames) into a set of high resolution (HR) images. It has become a popular area in the last 20 years since Tsai and Huang first introduced SR concept into image processing [1]. Although, SR techniques vary from each other in details, most of them are still based on a conventional three-step procedure: image registration (motion estimation), interpolation, and noise/blur removal [2]. The first step, image registration, is the most important step as many advanced SR techniques, such as POCS (Projection Onto Convex Sets) [3] and MAP (Maximum A Posteriori) [4], require a precise set of motion parameters to obtain reconstructed image of reasonable quality. Therefore, how to estimate the motion change becomes crucial.

To overcome the registration error problem, some researchers change the traditional three-step SR technique into two steps, where the motion estimation and the interpolation are combined [5]-[7]. Hardie *et al.* [5] and Michael *et al.* [6] introduce iterative schemes to estimate HR images and motion parameters alternately to reduce the registration error and improve the quality of the HR image. Instead of estimating the HR image and motion parameters alternately, He *et al.* [7] estimate the HR image and motion parameters simultaneously. Their method provides a better PSNR

than the methods of [5] and [6]. To simplify motion estimation and improve the quality of the estimated HR image, researchers make assumptions about the motion changes, for example, only global translational motion exists in different LR frames. Many SR approaches [1]-[6] are based on this constrained motion model. Unfortunately, this assumption limits the application of the SR technique because the motion changes are complex in practice. Therefore, most current SR techniques cannot give a satisfactory performance when complex motion changes occur between frames, such as in a video sequence.

To conquer the above restriction and make SR more realistic, a completely new method has to be created. Inspired by the recent breakthrough in 3-D denoising, Protter *et al.* [8] proposed a new SR framework, which skips the motion estimation stage by bringing the concept of NLM into SR field. As NLM-SR technique has already shown great potential in dealing with video sequence resolution enhancement, here we propose a hybrid SR model combining wavelet and NLM together, and use NLM to estimate the detail coefficients contained in wavelet sub-bands. The proposed WNLM-SR approach gives better numerical and visual quality than NLM-SR in our experiments.

The outline of the paper is as follows. A short review of NLM and its application in SR field is given in Section 2. A novel NLM-SR model based on wavelets is presented in Section 3. The experimental results of WNLM-SR are compared with other resolution enhancement techniques in Section 4. Conclusion is provided in Section 5.

2. NONLOCAL-MEANS FILTER AND SUPER-RESOLUTION

Here, a short review of NLM filtering is given first; more details can be found in [9]. Then, an NLM based SR technique [8] will be briefly introduced.

2.1. Nonlocal-Means Filtering

This algorithm is based on the assumption that a small patch of pixels appears several times in one image and also in different frames. Figure 1 shows an example. The region delimited



Fig. 1. An example of NLM. From left to right are the first three frames captured from video sequence “Miss America”.

with a dotted square represents the search area for the similarity between the target patch and its neighboring patches. In this area, similar patches marked by rectangular boxes can be found in different positions due to information redundancy. This feature provides a way to estimate the pixel value from a noise contaminated image. In the 3-D NLM algorithm, the estimate of the pixel at position (i, j) is given by

$$x(i, j) = \frac{\sum_{t=1}^T \sum_{(k,l) \in N(i,j)} w_t(k, l) y_t(k, l)}{\sum_{t=1}^T \sum_{(k,l) \in N(i,j)} w_t(k, l)}, \quad (1)$$

where y is the original pixel contaminated with white Gaussian noise, t is the frame index, and the w 's are the filter weights. The weight w is given by [8]

$$w(k, l) = \exp \left\{ -\frac{\|p_{y(i,j)} - p_{y(k,l)}\|_2^2}{2\sigma^2} \right\} \cdot f(\sqrt{(i-k)^2 + (j-l)^2 + (t-1)^2}), \quad (2)$$

where the operator p extracts patches centered at $y(i, j)$ and $y(k, l)$, and $\|\cdot\|_2$ denotes the Euclidean distance between the two patches. The second term $f(\cdot)$ is a geometric distance function; the weight is inversely proportional to the distance. This denoising algorithm is more robust than other methods because it utilizes information from several frames.

2.2. Nonlocal-Means Super-Resolution

When NLM is applied to the SR problem, several changes have to be made to achieve an improved resolution. The SR problem can be written as

$$Y_t = D H F X + n, \quad (3)$$

where Y is the vectorized LR frame, t is the index of LR frame, D is the decimation operator, H is the blurring matrix, F is the warping matrix, X is the vectorized HR image and n denotes Gaussian white noise. The aim of SR is to restore the HR image X from a series of LR frames Y . A penalty function can be defined as [8]

$$\epsilon^2 = \frac{1}{2} \sum_{t=1}^T \|D H F x - y_t\|_2^2 + \lambda C(x), \quad (4)$$

where C is a regularization term. In [8], the total Variation (TV) kernel is chosen to replace C , acting as an image deblurring kernel. To simplify the algorithm, a separation of the problem in (4) is done by first minimizing

$$\epsilon_{fusion}^2(Z) = \frac{1}{2} \sum_{t=1}^T (D F Z - Y_t)^T W_t (D F Z - Y_t), \quad (5)$$

which is the fusion step, where W_t is the weight matrix, followed by minimizing a deblurring equation

$$\epsilon_{db}^2(X) = \|H X - Z\|_2^2 + \lambda C(X). \quad (6)$$

A pixel-wise solution of Eq. (5) can be obtained as

$$\hat{z} = \frac{\sum_{t=1}^T \sum_{(k,l) \in N(i,j)} w_t^h(k, l) y_t^h(k, l)}{\sum_{t=1}^T \sum_{(k,l) \in N(i,j)} w_t^h(k, l)}, \quad (7)$$

where the superscript h refers to the HR coordinate. Instead of estimating the target pixel position in nearby frames, this algorithm considers all possible positions where the pixel may appear. Therefore, explicit motion estimation can be avoided. In fact, the above process can be considered as a fuzzy motion estimation [8]. Although, Eq. (7) has a similar form to (1), there are several important differences. First, weight estimation in (2) should be modified because w 's corresponding matrix W has to be of the same size as the HR image. So a simple up-scaling process to patch $p_{y(\cdot)}$ is needed before computing w ; Second, the total number of pixels y in (7) should be equal to the number of weights w . Thus, a zero-padding interpolation is applied to Y before the fusion stage.

3. WAVELET BASED NONLOCAL-MEANS SUPER-RESOLUTION

We propose a wavelet based nonlocal-means super-resolution (WNLM-SR) algorithm to explore further NLM ability in solving SR problems.

Figure 2 shows the framework of WNLM-SR. The LR image is first decomposed by wavelet transform. Here, we choose CDF 9/7 filter combination as the basis wavelets. Second, the HR sub-bands HL' and LH' are estimated from HL and LH by NLM-SR, while the LL sub-band is directly obtained from the LR image. To speed up the computation, the sub-band HH' is directly filled out with zeros. Then the inverse wavelet transform is applied to get an HR image, which may be noisy and blurred. Finally, a denoising and an iterative deblurring process are implemented.

To adapt to the above wavelet framework, several changes have to be made. A major changes is the way of estimating the weights. Because the input LR images and the output HR image have different sizes, the weight matrix W should have

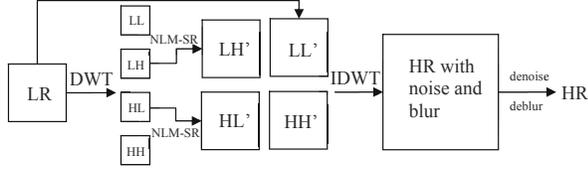


Fig. 2. Wavelet-based nonlocal-means super resolution

the same size as the HR image. Protter *et al.* [8] provide a simple way to solve this problem. The LR frames are first scaled up by Lanczos interpolation, then the weights are calculated from (2). In our case, two sets of weights should be calculated for sub-bands HL and LH (weights for HH can be skipped). The wavelet interpolation method is chosen from [10], as it offers a good sub-band approximation without increasing the computational time too much.

Another difference compared with NLM-SR is that the coefficients with large magnitudes are preserved, that is coefficients satisfying

$$|M_{coef}| > \alpha \times std(|M_{coef}|), \quad (8)$$

where α is a positive constant and std is for the standard deviation. As a result, not all the coefficients need to be re-estimated. Noted that, in our proposed approach, NLM only brings the information from nearby coefficients to provide a better estimation of the target coefficient. Therefore, there is no significant denoising ability compared with NLM-SR. A denoising filter needs to be added after taking the inverse wavelet transform. Here, we choose the 3D-NLM filter, though a 2D-NLM filter or other fast denoising methods can be adopted as well.

Instead of using the TV regularizer for deblurring [8], a Laplacian regularizer is applied in (6) for simplicity. The entire algorithm is summarized as follow:

Step 1: Wavelet decomposition

- Decompose LR frame using CDF 9/7 wavelet.

Step 2: Wavelet sub-band estimation

- Apply wavelet interpolation in the sub-bands HL and LH to obtain the up-scaled versions;
- Calculate the weight matrix W for every coefficient in the up-scaled sub-bands according to (2);
- Estimate the sub-band coefficients by weighted averaging the relevant coefficients with the weights computed in (7);
- Fill the sub-band HH' with zeros, and use the original LR frame for the sub-band LL';
- Apply the inverse wavelet transform to get a HR frame;
- Repeat the above steps for each frame in the sequence.

Step 3: Post-processing

- Use the 3D-NLM denoising filter to process the HR sequence generated in Step 2;
- Apply deblurring to each frame.



Fig. 3. Rows 1 to 4 are the results for 1st, 5th, 10th and 15th frames from “Susie” video sequence. Columns 1 to 4 are the HR images (ground truth), LR images, results of NLM-SR and results of proposed WNLM-SR.

4. EXPERIMENTAL RESULTS

The performance of WNLM-SR is compared to the performance of NLM-SR. To show the level of improvement of the proposed method, the single frame wavelet interpolation technique in [10] is also included in the comparison. It should be pointed out that the comparison in [8] is slightly unfair because NLM-SR has the ability to remove noise, while the other approaches do not use filtering to remove noise, a step that can easily be added. To be fairly, the 3-D NLM denoising filter is adopted after the single frame interpolation. The same regularization method, with a Laplacian regularizer, is implemented for all methods in the final deblurring stage. Other SR methods are not implemented in this comparison because those methods make some assumptions about motion, which lead to bad performance in practical video resolution enhancement [8].

The first 15 frames of Foreman, Susie and Miss America sequences are used in the experiment. However, to accelerate computation, only 5 frames are combined to estimate one frame; for example, 3rd, 4th, 5th, 6th, 7th frames participate in computing the 5th frame. To obtain the LR video sequences, the original HR frames are first filtered with a 3×3 low-pass filter, then decimated by 2. Additive Gaussian White Noise is added with a stand deviation $\sigma = 2$. The size of the similarity block for calculating the weights is 13×13 for all the sequences. The size of the search area is different for different sequences: 13×13 in “Miss America”, 21×21 in “Foreman” and 31×31 in “Susie”. To simplify the threshold selection, we preserve 6% largest coefficients in

Table 1. Mean PSNR results for the test sequences. The five methods are: Wavelet Interpolation (WI), deblurred Wavelet Interpolation (D-WI), denoised and deblurred Wavelet Interpolation (D-D-WI), NLM-SR and Proposed WNLM-SR.

Sequence	WI	D-WI	D-D-WI	NLM SR	WNLM SR
Miss-A	36.41	36.61	37.93	37.39	37.96
Foreman	31.63	32.02	32.89	32.87	33.02
Susie	33.16	33.36	34.43	34.13	34.47

each sub-band. The peak signal to noise ratio (PSNR) is used to evaluate the quality of the restored images.

Table 1 shows that the proposed WNLM-SR provides the highest PSNR. We should point out that the D-D-WI was not considered in [8]; however, it achieves better results than D-WI, which includes deblurring only after interpolation. If all the sub-band coefficients are preserved by setting $\alpha = 0$, WNLM-SR becomes D-D-WI. Figure 3 shows sample frames restored by both NLM-SR and WNLM-SR; the results of WNLM-SR appear to be sharper than the results of NLM-SR. Figure 4 illustrates the errors generated in the facial region of “Susie” image. Compared with NLM-SR, WNLM-SR generates fewer errors in the edge areas, especially within the areas highlighted with red rectangles.

In terms of computation speed, WNLM-SR is generally faster than NLM-SR. Only half of the image needs to be estimated, which is the detail coefficients in sub-bands HL and LH. The wavelet transform is very fast to compute. Although WNLM-SR includes 3D-NLM denoising, the flexibility of this hybrid framework allows implementation of fast denoising techniques, such as wavelet shrinkage.

5. CONCLUSIONS

This paper contributes a hybrid framework for video sequence super-resolution. NLM is used to estimate the wavelet sub-band coefficients. The proposed WNLM-SR outperforms the previous methods including NLM-SR, single frame wavelet interpolation. The computational time for WNLM-SR is also reduced. The performance of WNLM-SR can be further improved by considering HH sub-band and introducing adaptive threshold for each frame, which is left for future study.

Acknowledgments: This work is supported in part by a grant from the Australian Research Council (ARC).

6. REFERENCES

- [1] T. S. Huang and R. Y. Tsai, “Multi-frame image restoration and registration,” *Advances in Computer Vision and Image Processing*, vol. 1, pp. 317–339, 1984.
- [2] 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, 2003.
- [3] A. J. Patti, M. I. Sezan, and A. Murat Tekalp, “Super-resolution video reconstruction with arbitrary sampling lattices and nonzero aperture time,” *IEEE Transactions on Image Processing*, vol. 6, no. 8, pp. 1064–1076, 1997.
- [4] Y. He, K.-H. Yap, L. Chen, and L.-P. Chau, “Blind super-resolution image reconstruction using a maximum a posteriori estimation,” *IEEE International Conference on Image Processing*, pp. 1729–1732, 2006.
- [5] 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 Transactions on Image Processing*, vol. 6, no. 12, pp. 1621–1633, 1997.
- [6] K. N. Michael, K. Jaehoon, and N. K. Bose, “Constrained total least-squares computations for high-resolution image reconstruction with multisensors,” *International Journal of Imaging Systems and Technology*, vol. 12, no. 1, pp. 35–42, 2002.
- [7] Y. He, K.-H. Yap, L. Chen, and L.-P. Chau, “A nonlinear least square technique for simultaneous image registration and super-resolution,” *IEEE Transactions on Image Processing*, vol. 16, no. 11, pp. 2830–2841, 2007.
- [8] M. Protter, M. Elad, H. Takeda, and P. Milanfar, “Generalizing the nonlocal-means to super-resolution reconstruction,” *IEEE Transactions on Image Processing*, vol. 18, no. 1, pp. 36–51, 2009.
- [9] A. Buades, B. Coll, and J. M. Morel, “Denoising image sequences does not require motion estimation,” *IEEE Conference on Advanced Video and Signal Based Surveillance*, pp. 70–74, 2005.
- [10] A. Temizel and T. Vlachos, “Wavelet domain image resolution enhancement,” *IEE Proceedings on Vision, Image and Signal Processing*, vol. 153, pp. 25–30, 2006.

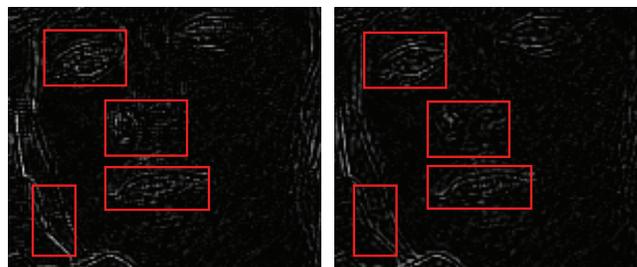


Fig. 4. Error comparison of NLM-SR (left) and WNLM-SR (right) on a facial area.