VIDEO SUPER-RESOLUTION RECONSTRUCTION USING A MOBILE SEARCH STRATEGY AND ADAPTIVE PATCH SIZE

Ming-Hui Cheng, Hsuan-Ying Chen and Jin-Jang Leou

Department of Computer Science and Information Engineering, National Chung Cheng University
Chiayi, Taiwan 621, Republic of China

Keywords: Super-resolution (SR) Reconstruction, Fuzzy Motion Estimation, Mobile Search Strategy, Adaptive Patch Size.

Abstract: In this study, a new video super-resolution (SR) reconstruction approach using a mobile search strategy and adaptive patch size is proposed. Based on the nonlocal-means (NLM) SR algorithm, the mobile strategy for search center and adaptive patch size are proposed to reduce the computational complexity and improve the visual quality, respectively. Based on the experimental results obtained in this study, the performance of the proposed approach is better than those of two comparison approaches.

1 INTRODUCTION

Obtain high-resolution (HR) video frames (images) from multiple observed low-resolution (LR) video frames (images) by using image/video processing techniques can be called video super-resolution (SR) reconstruction, which can be employed in many applications, such as video surveillance and medical imaging. Video SR reconstruction usually consists of three steps: registration (motion estimation), interpolation, and restoration. Based on the three steps implemented simultaneously or individually, there are two kinds of video SR reconstruction approaches, namely, simultaneous and asynchronous (Park, Park, and Kang, 2003). Simultaneous video SR reconstruction approaches include frequency-domain SR reconstruction, regularized SR reconstruction, …, etc. Frequency-domain SR reconstruction (Tsai and Huang, 1984), limited to global translation models, is not suitable for local motion models including spatial variations. Video SR reconstruction, an ill-posed problem, can be processed by regularization. Zibetti and Mayer (2007) proposed a video SR reconstruction approach exploiting the correlation among a video sequence to obtain HR video frames. Costa and Bermudez (2008) proposed a strategy to reduce the outlier effect on the reconstructed video sequence.

+ This work was supported in part by National Science Council, Taiwan, Republic of China under Grants NSC 96-2221-E-194-033-MY and NSC 98-2221-E-194-034-MY33.

Regularized SR reconstruction may obtain good SR reconstruction results, whereas it is computationally expensive.

The performance of asynchronous video SR reconstruction is comparable to that of simultaneous video SR reconstruction, whereas asynchronous video SR reconstruction is usually intuitive and simple. Narayanan, Hardie, Barner, and Shao (2007) proposed a computationally efficient SR reconstruction algorithm using partition-based weighted sum (PWS) filters. Protter, Elad, Takeda, and Milanfar (2009) used the nonlocal-means (NLM) algorithm to perform video SR reconstruction without explicit motion estimation. Protter and Elad (2009) presented a video SR reconstruction framework using probabilistic and crude motion estimation.

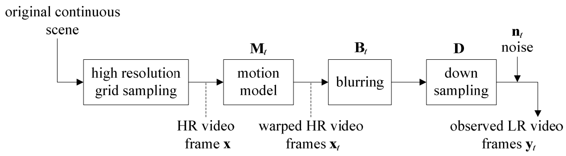

Figure 1: Observation model for video SR reconstruction (Park, et al., 2003).

In this study, an observation model describing the relationship between HR and LR video frames for video SR reconstruction is shown in Fig. 1 (Park et al., 2003). If $\mathbf{x}$ denotes an HR video frame that is sampled from the original continuous scene. Then,

Cheng M., Chen H. and Leou J. (2010).
VIDEO SUPER-RESOLUTION RECONSTRUCTION USING A MOBILE SEARCH STRATEGY AND ADAPTIVE PATCH SIZE.
In *Proceedings of the International Conference on Signal Processing and Multimedia Applications*, pages 108-111
DOI: 10.5220/0002945301080111

the t-th observed LR video frame $\mathbf{y}_t$, processed by the warping ($\mathbf{M}_t$), blurring ($\mathbf{B}_t$), downsampling ($\mathbf{D}$), and noise ($\mathbf{n}_t$) operators, can be obtained by

$$\mathbf{y}_t = \mathbf{D}\,\mathbf{B}_t\,\mathbf{M}_t\,\mathbf{x} + \mathbf{n}_t. \qquad (1)$$

2 PROPOSED APPROACH

In this study, based on the NLM SR algorithm (Protter et al., 2009), a "mobile" strategy for motion search center and adaptive patch size are proposed to reduce the computational complexity and to improve the visual quality, respectively.

2.1 NLM SR Algorithm

The T LR video frames $\mathbf{y}_t$, $t \in [0,...,T-1]$, in a video sequence are interpolated by Lanczos interpolation with a magnification factor (MF) into $\mathbf{Y}_t$. Then, the NLM SR algorithm will perform two iterations, in which each iteration consists of fuzzy motion estimation and deblurring.

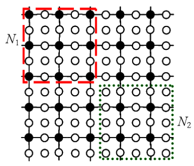

Figure 2: The HR video frame consists of known pixels (●) from a reference LR video frame and the pixels to be interpolated (○) with $MF = 2\times 2$ whereas N_1 and N_2 denote two 5×5 search neighbourhoods centered at ● and ○, respectively.

Fuzzy motion estimation finds several matching patches based on the spatial or temporal redundancy and fuses the matched patches using designed weights. As the illustrated example shown in Fig. 2, an HR video frame consists of known pixels (●) from a reference LR video frame and the pixels to be interpolated (○) with $MF = 2\times 2$, whereas N_1 and N_2 denote two 5×5 search neighbourhoods centered at ● and ○, respectively. $N_t(x,y)$ denotes the search neighbour centered at a pixel (●(x,y) or ○(x,y)) in the t-th HR video frame $\mathbf{Y}_t$, whereas ●$(i,j) \in N_t(x,y)$ means that a know pixel ●(i,j) within $N_t(x,y)$. Then, the reconstructed pixel $\mathbf{s}_{ref}(x,y)$ in the current HR video frame by fuzzy motion estimation, $ref \in [0,...,T-1]$, is given by (Protter, et al., 2009).

$$\mathbf{s}_{ref}(x,y) = \frac{\sum_{t\in[0,...,T-1]}\sum_{\bullet(i,j)\in N_t(x,y)} w_t(i,j)\mathbf{Y}_t(i,j)}{\sum_{t\in[0,...,T-1]}\sum_{\bullet(i,j)\in N_t(x,y)} w_t(i,j)}, \qquad (2)$$

where $w_t(i,j)$, the fusion weight for a given pixel ● (i,j) in $\mathbf{Y}_t$, is determined as

$$w_t(i,j) = \exp\left(\frac{-\| P_{ref}(x,y) - P_t(i,j)\|^2}{2\sigma_{\mathrm{NLM}}^{\;2}}\right), \qquad (3)$$

where $P_{ref}(x,y)$ and $P_t(i,j)$ denote two 13×13 patches centered at (x,y) and (i,j) extracted from $\mathbf{Y}_{ref}$ and $\mathbf{Y}_t$, respectively, and σ_{NLM} is a control parameter. Note that if the patch $P_t(i,j)$ in $\mathbf{Y}_t$ is similar to the reference patch $P_{ref}(x,y)$ in $\mathbf{Y}_{ref}$, a high fusion weight $w_t(i,j)$ will be obtained.

Additionally, the total-variation deblurring approach (Rudin, Osher, and Fatemi, 1992) is used for regularization in the NLM SR algorithm. Then, each processed HR video frame generated by the first iteration is treated as the input of the second iteration.

2.2 Mobile Search Strategy

The visual quality of video SR reconstruction by the NLM video SR algorithm is good for static or small-motion areas, but it is degraded for other types of areas, if a small size (such as 7×7) of search neighbourhoods $N_t(x,y)$ is employed. To improve its performance, a large size (such as 63×63) of search neighbourhoods $N_t(x,y)$ can be used, which will make the NLM video SR algorithm computationally expensive. As shown in Fig. 3, to overcome the weakness of the NLM video SR algorithm using T "static" search neighbourhoods in T video frames, the proposed approach uses T "mobile" search neighbourhoods in T video frames so that the performance of the proposed approach is good even a small size (here 7×7) of search neighbourhoods is used (with a low computational complexity).

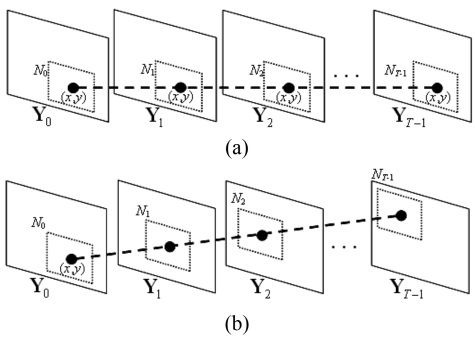

Figure 3: (a) The T "static" search neighbourhoods in the NLM SR algorithm; (b) the T "mobile" search neighbourhoods in the proposed approach.

In this study, the search neighbourhood $N_t(x,y)$ in the t-th HR video frame $\mathbf{Y}_t$ will move to $N_{t+1}(i,j)$ in the $(t+1)$-th video frame $\mathbf{Y}_{t+1}$ if $w_{t+1}(i,j)$ is the maximum value among all computed $w_{t+1}(i,j)$'s for $N_t(x,y)$. Similarly, the search neighbourhood $N_t(x,y)$ in the t-th HR video frame $\mathbf{Y}_t$ will move to $N_{t-1}(i,j)$ in the $(t-1)$-th HR video frame $\mathbf{Y}_{t-1}$ if $w_{t-1}(i,j)$ is the maximum value among all computed $w_{t-1}(i,j)$'s for $N_t(x,y)$.

2.3 Video SR Reconstruction using Adaptive Patch Size

In this study, a variation detection algorithm is proposed to extract non-translation motion areas. Then, the morphological dilation and erosion operators are employed to complete these non-translation motion areas. Finally, a video SR reconstruction approach using adaptive patch size is proposed to generate the final video SR reconstruction results.

If a patch $P_{ref}(x,y)$ in $\mathbf{Y}_{ref}$ is locates in a static or translation motion area, it may contain many similar patches in $\mathbf{Y}_t$, $t \in [0,\ldots,T-1]$. On the contrary, if the patch $P_{ref}(x,y)$ in $\mathbf{Y}_{ref}$ locates in a non-translation motion area, it may find only one or few sub-similar patches in $\mathbf{Y}_t$, $t \in [0,\ldots,T-1]$. These sub-similar patches have high fusion weights. In this study, the fusion weights can be utilized to detect non-translation motion areas. The binarized pixel $v_{ref}(x,y)$ in the variation video frame v_{ref} can be defined as

$$v_{ref}(x,y) = \begin{cases} 1 \text{ (white), if } \sigma_{ref}(x,y) > T_v \text{ and } (x,y) \neq (i,j), \\ 0 \text{ (black), otherwise,} \end{cases} \quad (4)$$

where

$$\sigma_{ref}(x,y) = \sqrt{\left(\sum_{l=0}^{n-1} (\hat{w}_l - \mu(x,y))^2\right)/n}, \quad (5)$$

and

$$\mu(x,y) = \left[\sum_{l=0}^{n-1} \hat{w}_l\right]/n, \quad (6)$$

T_v is a threshold, $\hat{w}$ is the normalized fusion weight over n fusion weights, and n is the total number of fusion weights in T video frames. Note that a patch $P_{ref}(x,y)$ having a high $\sigma_{ref}(x,y)$ may have one or few sub-similar patches, i.e., $P_{ref}(x,y)$ locates in a non-translation motion area in $\mathbf{Y}_{ref}$.

The variation video frame of the 10th video frame of the "Foreman" sequence is shown in Fig. 4(a). Because v_{ref} usually contain discontinuous parts, the morphological dilation and erosion operators are used to complete the detected (white) areas, as the illustrated example shown in Fig. 4(b)-(c).

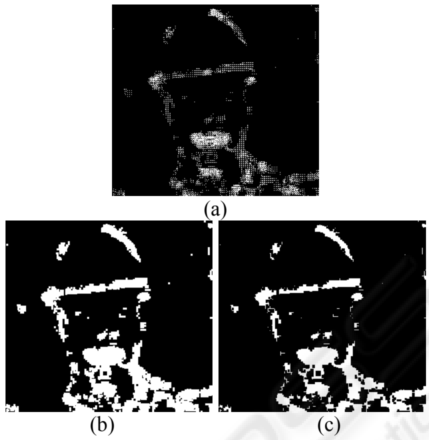

Figure 4: (a) The variation video frame v_{ref} of the 10th video frame of the "Foreman" sequence; (b)-(c) the processed variation video frames, d_{ref} and e_{ref}, after applying the morphological dilation and erosion operators, respectively.

For $P_{ref}(x,y)$ locating in a non-translation motion area in $\mathbf{Y}_{ref}$, using a small patch size usually has more similar patches. Thus, in this study, $P_{ref}(x,y)$ locating in a static or small-motion area will use the standard patch size, whereas $P_{ref}(x,y)$ locating in a non-translation motion area will use a small patch size. Here, the proposed video SR reconstruction approach using adaptive patch size is described as follows.

Step 1:	Compute $e_{ref}(x,y)$ from $\mathbf{Y}_{ref}$.
Step 2:	Reduce the patch size by p pixels and go to Step 1 if $e_{ref}(x,y)$=1 and the patch size is larger than 5×5.
Step 3:	Reconstruct each pixel $s_{ref}(x,y)$ (Eq. (2)) using the new patch size.

3 EXPERIMENTAL RESULTS

In this study, three video sequences, namely, "Foreman," "Miss America," and "Suzie," are used to evaluate the performance of the proposed approach. Two comparison approaches, namely, Lanczos interpolation and NLM SR algorithm (Protter et al., 2009), are implemented. The LR video frames in the observation model are obtained as follows: (1) each ground truth video frame is blurred using a 3×3 uniform mask, (2) decimated by a factor of 3×3, and (3) added by an additive

Gaussian noise with zero mean and variance $\sigma^2 = 4$. In this study, T=30, the initial (standard) patch size= 13×13, the size of search neighbourhoods= 7×7 , $\sigma_{\mathrm{NLM}} = 2.2$, T_v=0.018, and p=8.

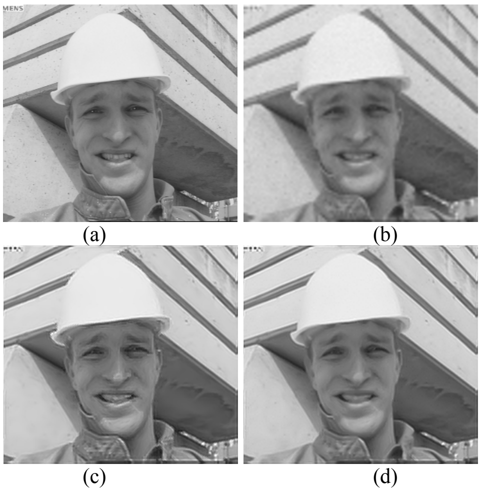

Figure 5: (a) The ground truth (the 10th video frame); (b)-(d) the processed video frames by Lanczos interpolation, NLM video SR algorithm, and the proposed approach, respectively, with $MF = 3\times3$.

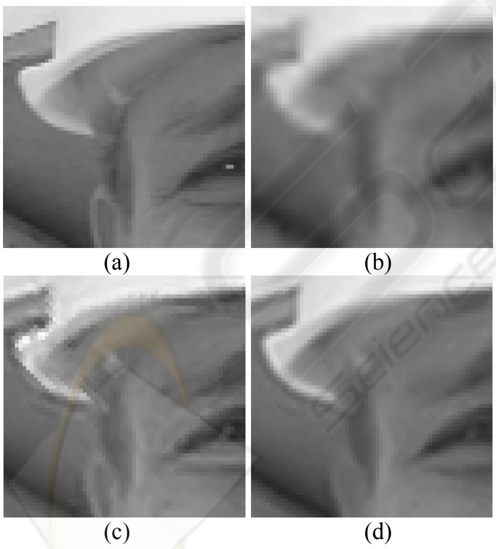

Figure 6: Details of (a) the ground truth; and (b)-(d) the processed video frames by Lanczos interpolation, NLM video SR algorithm, and the proposed approach, respectively, with $MF = 3\times3$.

The video SR reconstruction results of the 10th video frame of the "Foreman" sequence by the two comparison approaches and the proposed approach are shown in Fig. 5, whereas the details of the processing results shown in Fig. 5 are shown in Fig. 6. The visual quality of the processed results by the proposed approach is indeed better than those by the two comparison approaches. In terms of average PSNR (peak-signal-to-noise-ratio) in dB, the performance of the proposed approach for the three video sequences are better than those of Lanczos interpolation and the NLM video SR algorithm about 0.8 dB and 0.5 dB, respectively.

4 CONCLUDING REMARKS

In this study, a new video SR reconstruction approach using a mobile strategy and adaptive patch size is proposed. Based on the NLM SR algorithm, the mobile strategy for motion search center and adaptive patch size are used to reduce the computational complexity and improve the visual quality, respectively. Based on the experimental results obtained in this study, the performance of the proposed approach is better than those of two comparison approaches.

REFERENCES

Costa, G. H. and Bermudez, J. C. M., 2008. Informed choice of the LMS parameters in super-resolution video reconstruction applications. *IEEE Trans. on Signal Process.*, 56(2), 555-564.

Narayanan, B., Hardie, R. C., Barner, K. E., and Shao, M., 2007. A computationally efficient super-resolution algorithm for video processing using partition filters. *IEEE Trans. Circuits Syst. Video Technol.*, 17(5), 621–634.

Park, S., Park, M., and Kang, M. G., 2003. Super-resolution image reconstruction: a technical overview. *IEEE Signal Process. Mag.*, 20(5), 21–36.

Protter, M. and Elad, M., 2009. Super resolution with probabilistic motion estimation. *IEEE Trans. Image Process.*, 18(8), 1899–1904.

Protter, M., Elad, M., Takeda, H., and Milanfar, P., 2009. Generalizing the nonlocal-means to super-resolution reconstruction. *IEEE Trans. Image Process.*, 18(1), 36–51.

Rudin, L., Osher, S., and Fatemi, E., 1992. Nonlinear total variation based noise removal algorithms. *Phys. D*, 60, 259–268.

Tsai, R. Y. and Huang, T. S., 1984. Multiple frame image restoration and registration. in *Advances in Computer Vision and Image Process.*, 1, 317-339.

Zibetti, M. V. W. and Mayer, J., 2007. A robust and computationally efficient simultaneous super-resolution scheme for image sequences. *IEEE Trans. Circuits Syst. Video Technol.*, 17(10), 1288-1300.