# Adaptive Bilateral Filtering for Super-Resolution Reconstruction of Video Sequences

Giaime Ginesu, Tiziana Dessì, Luigi Atzori, and Daniele D. Giusto

Department of Electronic Engineering,
University of Cagliari, Italy
{g.ginesu,tiziana.dessi,l.atzori}@diee.unica.it,
ddgiusto@unica.it

**Abstract.** The current multimedia consumer market is characterized by the advent of cheap but rather high-quality high definition displays, mostly for home theater applications. This trend is only partially supported by the deployment of high-resolution multimedia services, either over the Internet or through satellite channels. To address the resulting disparity between content and display formats, video super-resolution techniques represent a major solution. This subject is addressed in this paper, by exploiting the use of the bilateral filtering. This is a spatial filtering operator that relies on dynamically calculating a FIR kernel which has the major advantages of video content adaptability and edge preserving. Results are encouraging and suggest that the proposed method could be practically implemented.

**Keywords:** super resolution reconstruction, high-definition, image enhancement.

## 1 Introduction

During the last couple of years, the multimedia consumer market has been characterized by the advent of cheap but rather high-quality HD (High Definition) displays, mostly for home theater applications. This process is bound to continue at least in the near future, with the introduction of displays of even higher spatial resolution formats, such as DigitalCinema or UHD/UHDTV (Ultra High Definition). This phenomenon is only partially supported by the deployment of high-resolution (spatial and temporal) multimedia services, either over the Internet or through satellite channels. Indeed, the content generation and distribution sector seems not to be able to keep pace with the display technology, which is characterized by a significant decrease of the cost per pixel. Conversely, the cost of transmitting one bit of video information is not going to decrease, at least when sending it at the quality of service level required by the streaming applications. The advances in the video compression domain, which proceeds by roughly doubling the compression rate every 5 years, do not allow for decreasing such cost significantly. Moreover, older productions need to be either re-mastered or post-processed in order to be broadcasted for HD exploitation. The decoding of low-resolution multimedia content then thwarts the

J. Rodriguez, R. Tafazolli, C. Verikoukis (Eds.): MOBIMEDIA 2010, LNICST 77, pp. 39–47, 2012.

benefits of high-resolution displays and involves the use of appropriate signal processing procedures. Low resolution frames then need to be enlarged through super-resolution techniques, with zooming factors that may increase considerably during the next few years.

The present paper focuses on this problem by proposing a solution which resorts on the use of the bilateral filtering [16]. This is a spatial filtering operator that relies on dynamically calculating a FIR kernel. Edge preserving nature and adaptability are the main advantages of this kind of filter. Whereas it has already been adopted to address the super-resolution problem, its application has been mostly restricted to the case of still-pictures. Herein, we propose its use to tackle the video sequences super-resolution problem and, accordingly, we propose several changes in its use. The first change is related to its extension to the time domain through the use of a group of frames when estimating the super-resolution version of each frame. This operation goes in the direction of both strengthening the local visual information sketch and compensating (thus reducing) the local noise in the current frame with that of previous ones. It may be argued that using frames other than the one to be processed may introduce some distortions due to differences between adjacent frames. However, given an adequately small time window, these differences do not modify significantly the local visual structure, as shown by the high correlation between adjacent frames in Fig. 1. This graph plots the average interframe correlation of 10 CIF (Common Intermediate Format, 352×288pixels, 29fps) test sequences with no less than 300 frames, computed for a window of 31 frames. The correlation curve shows that on average 3 consecutive frames have a correlation higher than 0.9 and 5 consecutive frames are correlated as much as 0.85.

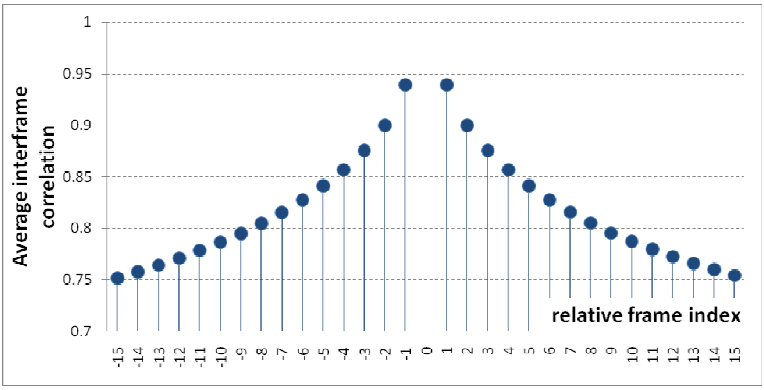


**Fig. 1.** Average interframe correlation

Instead of relying on a classical motion compensation algorithms, the proposed method implements a 3D sample estimation and filtering.

A second major change we propose is related to the preliminary estimation of the pixel that are added to increase the resolution. While the procedure itself is aimed at estimating these values, these are needed to bootstrap the bilateral interpolation when

computing the filter kernels. To address this problem we make use of a gradient based edge-preserving interpolation.

The paper is organized as follows. Section 2 gives an overview of the state of the art of super-resolution techniques applied to image sequences. Section 3 illustrates the proposed technique. Experimental results are discussed in Section 4. Conclusions are drawn in Section 5.

## 2 Past Work

This section gathers some of the most significant approaches addressing the issue of super-resolution. In the following, LR and HR refer to low resolution and high resolution frames respectively. The former represents the starting point of the signal processing procedure, whereas the latter corresponds to its output. It is assumed that the LR nature of input frames can derive from a low-resolution original source or be the result of sub-sampling the original frames to meet storage or transmission requirements.

Among single frame approaches, we focus on bilateral filtering techniques. A novel algorithm that integrates bilateral filtering and back-projection is presented in [1]. The former achieves edge-preserving image smoothing while the latter minimizes the reconstruction error with an edge-based iterative procedure. In [2], the authors find the connection between the soft edge smoothness and a soft cut metric through a generalization of the Geocuts method. This term is incorporated into an objective function to produce smooth soft edges and it is applied on alpha channel.

Among frequency-based multi-frame approaches, Tsai and Huang. [3] present an algorithm that improves the resolution of Landsat image data by modeling the observed images as under-sampled versions of an unchanging scene undergoing global translational motion. Several limitations of the such method are addressed by Tekalp, Ozkan and Sezan in [4]. Periodic sampling is still assumed and a translation-only motion model is used. Kim, Bose and Valenzuela [5] exploit the frequency domain theoretical framework and the global translation observation model proposed in [3].

Among spatial domain methods, Keren, Peleg and Brada [6] propose an approach to image registration based on a global translation and rotation model. Irani and Peleg [7, 8] extend the earlier work by improving the means of backprojecting the error between the simulated LR images and the observed data. A very general procedure for super-resolution reconstruction is proposed in [9], for scenes which contain arbitrary independent motion.

Among spatio-temporal approaches, [10] consists in an adaptive weighted bidirectional algorithm that uses multiple frames to enhance the accuracy of motion estimation. In [11] a directional fuzzy filter is applied to intra- and inter-frame pixels to reduce artifacts in compressed image sequences. In [12] the idea of super-resolution reconstruction from a set of globally translated images of an unchanging 2D scene is considered and compared to a global translation and rotation model used

in [6]. A dynamic super-resolution sequence reconstruction from a LR sequence containing sub-pixel shifts is presented in [13]. An iterated back-projection based (IBP) algorithm is presented in [14], based on the uncertainty degree metric, used in pixel reconstruction and error correction and improved by adaptive techniques.

Probabilistic methods are also considered. Since super-resolution is an ill-posed inverse problem, techniques which are capable of including a-priori constraints are well suited to this application. Schultz and Stevenson developed an estimator based on the maximum a posterior probability (MAP) with both the spatial and temporal information [15].

## 3 Proposed Approach

The proposed technique is aimed at reconstructing each HR frame from a limited number of frames extracted from a LR sequence, without any preliminary knowledge of the high-definition data. For any given frame, a sliding time-window determines the set of LR frames (from 2 to $N$) to be processed in order to produce the output stream. The window is shifted forward to produce successive HR frames of the output sequence, as shown in Fig. 2.

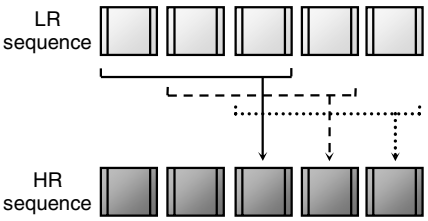


**Fig. 2.** Sliding time-window

Not to delay the display of the frames, each HR is generated by considering only previous frames. A space-time 3D filter is then applied to such partitioning of the original signal; the filter is developed from the bilateral filter solution with the introduction of sample estimation through local analysis, involving smooth and edge area classification and exploitation.

### 3.1 Background

The proposed interpolation is based on bilateral filtering [16], which relies on dynamically calculating a FIR kernel from known pixels through spatial distance ($W_S$) and amplitude distance ($W_R$) weighting contributions:

$$\begin{cases} W_S^{i,j}(h,k) = \exp\left(-\dfrac{d^2([i,j],[i-h,j-k])}{2\sigma_S^2}\right) \\ W_R^{i,j}(h,k) = \exp\left(-\dfrac{(I(i,j)-I(i-h,j-k))^2}{2\sigma_R^2}\right) \end{cases} \tag{1}$$

where: $(i,j)$ denotes the kernel center; $I(x,y)$ is the signal amplitude at coordinates $p_{x,y}$; $d(x,y)$ is the Euclidean distance function; $\sigma_S^2$ and $\sigma_R^2$ are the spatial and the amplitude variance, respectively. The kernel coefficients are then computed as follows:

$$W^{i,j}(h,k) = \frac{W_S^{i,j}(h,k)\cdot W_R^{i,j}(h,k)}{\sum_{x,y\in K} W_S^{i,j}(x,y)\cdot W_R^{i,j}(x,y)} \tag{2}$$

where $K$ represent the set of pixels belonging to the filtering kernel. Fig. 3 illustrates the two weighting contributions and the final kernel shape $W$. It can be seen that the $W_S$ contribution has a symmetric shape depending only on the distance from the kernel center, while the $W_R$ contribution is modeled by the amplitude distance from the central sample.

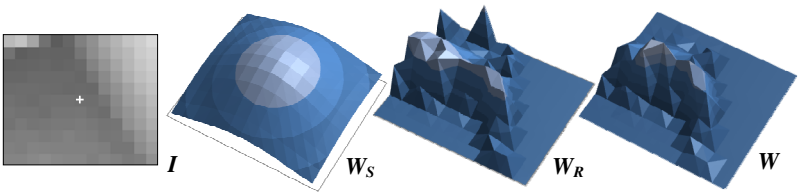


**Fig. 3.** Filter kernel shape related to an edge area

### 3.2 Super Resolution

In the proposed technique, we extend the bidimensional bilateral filter described in the previous section into a tridimensional filter adding the temporal axis. Additionally, we make use of kernel with three equal edges. Given the size of the sliding time-window, $N$, the linear size of the LR kernel, $s_W$, and the linear zoom factor, $zf$, the cubic filter kernel will entail a local lattice with size:

$$N \cdot s_W^2 \cdot zf^2 \tag{3}$$

It can be observed (Fig. 4) that only $N \cdot s_W^2$ samples are known from the original signal. The bilateral interpolation then consists in reconstructing the current (unknown) sample through the bilateral formulation.

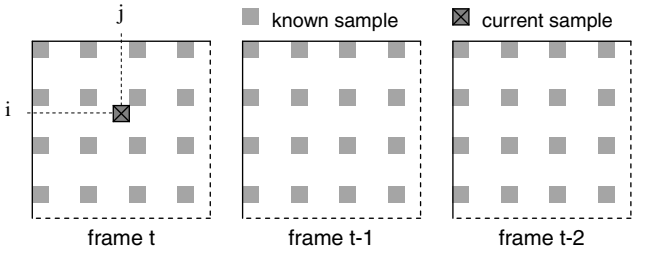


**Fig. 4.** HR image lattice for the kernel support

However, while the spatial term, $W_S$, can be easily computed by considering the spatial distances in the HR lattice, the amplitude term, $W_R$, lacks the definition of the sample value itself. In order to process the signal, such value must be estimated. Given $\hat{I}$, the amplitude estimate, $(i, j, t)$ spatial (intra-frame) and temporal (inter-frame) dimensions respectively, (1) becomes:

$$\begin{cases} W_S^{i,j,t}(h,k,l) = \exp\left(-\dfrac{d^2([i,j,t],[i-h,j-k,t-l])}{2\sigma_S^2}\right) \\ \hat{W}_R^{i,j,t}(h,k,l) = \exp\left(-\dfrac{\left(\hat{I}(i,j,t) - I(i-h,j-k,t-l)\right)^2}{2\sigma_R^2}\right) \end{cases} \tag{4}$$

In order to estimate the current sample value, a local analysis is performed, based on the LR edge map. The process is graphically described in Fig. 5. Both edge magnitude and orientation are firstly computed through a gradient operator. Only strong edges are considered by applying a threshold to the edge magnitude values. For each neighborhood, a linear edge model is derived through the computation of the local edge center of mass and the average edge normal angle:

$$\begin{cases} i_c, j_c = \dfrac{1}{N_{p_{h,k}}} \sum\sum_{p_{h,k} \in edge} h_{p_{h,k}}, \dfrac{1}{N_{p_{h,k}}} \sum\sum_{p_{h,k} \in edge} k_{p_{h,k}} \\ \theta_c = \dfrac{1}{N_{p_{h,k}}} \sum\sum_{p_{h,k} \in edge} \theta_{p_{h,k}} \end{cases} \tag{5}$$

with $i_c, j_c$ coordinates of the edge center of mass, $\theta_c$ average edge angle, $h_{p_{h,k}}, k_{p_{h,k}}$ coordinates of edge pixels, $\theta_{p_{h,k}}$ edge pixel angle and $N_{p_{h,k}}$ number of edge pixels.

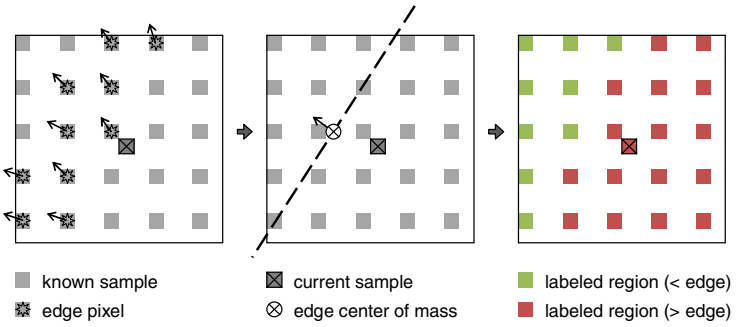


**Fig. 5.** Local neighborhood analysis

Known samples are then classified as belonging to either the same (SS) or the other side of the edge line (OS) in comparison with the current sample, according to the following rule:

$$\begin{aligned} &\text{if } (i-i_c) < (\arctan\theta_c + \pi/2)\cdot(j-j_c) \text{ then } i \in \text{SS}; \\ &\text{otherwise } i \in \text{OS} \end{aligned} \tag{6}$$

Once all known samples are classified, the current sample value is computed as the distance-weighted average among the samples from the same class. Notice that the complete process is applied to a time neighborhood of $N$ frames (Fig. 6).

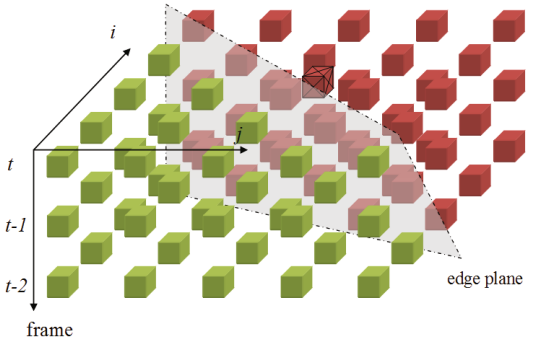


**Fig. 6.** Local neighborhood analysis; space-time structure

## 4 Results and Discussion

The proposed method has been evaluated on 12 1280×720 and 1920×1080 4:2:0 YUV video sequences, provided by [17]. The test sequences have been selected with the purpose of presenting a broad range of signal behaviors, in terms of different motion and scene complexity. A subsampled video sequence (Gaussian local filtering) is

preliminarily produced from the original video and is used as input sequence for the devised algorithm at any given zoom factor. Test parameters: $N = 3$, $\sigma_S = 10$ and $\sigma_R = 2$. A visual comparison between bicubic interpolation (left) and the proposed method (right) is provided in Fig. 7 for three different samples. The proposed algorithm shows a good behavior in both strong and weak edge regions, while highly textured areas are still challenging. Further developments are ongoing in order to deal with such problem through the exploitation of a more precise edge model.

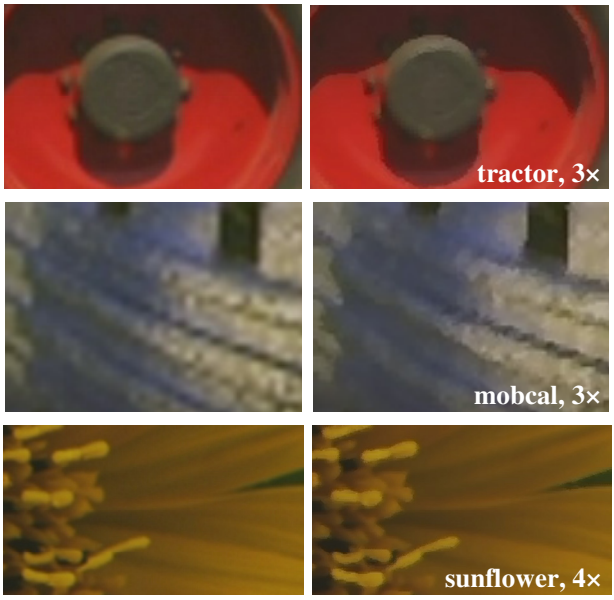


**Fig. 7.** Visual comparison

## 5 Conclusion

A technique for high-resolution reconstruction of low-resolution video sequences has been presented. The proposed algorithm extends the use of the bilateral filter through the exploitation of the space-time domain and the development of edge-based samples estimation, achieving promising results.

## References

1. Dai, S., Han, M., Wu, Y., Gong, Y.: Bilateral Back-Projection for Single Image Super Resolution. In: Proc. of IEEE Int. Conf. on Multimedia and Expo., Beijing, pp. 1039–1042 (July 2007)
2. Dai, S., Han, M., Wu, Y., Gong, Y.: Soft Edge Smoothness Prior for Alpha Channel Super Resolution. In: Proc. of IEEE Int. Conf. on Computer Vision and Pattern Recognition, Minneapolis, MN, pp. 1–8 (June 2007)

3. Tsai, R.Y., Huang, T.S.: Multiframe image restoration and registration. In: Tsai, R.Y., Huang, T.S. (eds.) Advances in Computer Vision and Image Processing, vol. 1, pp. 317–339. JAI Press Inc. (1984)
4. Tekalp, A.M., Ozkan, M.K., Sezan, M.I.: High-resolution image reconstruction from lower-resolution image sequences and space-varying image restoration. In: Proc. of the IEEE Int. Conf. on Acoustics, Speech and Signal Processing, San Francisco, CA, vol. 3, pp. 169–172 (1992)
5. Kim, S.P., Bose, N.K., Valenzuela, H.M.: Recursive reconstruction of high resolution image from noisy undersampled multiframes. IEEE Trans. on Acoustics, Speech and Signal Processing 38(6), 1013–1027 (1990)
6. Keren, D., Peleg, S., Brada, R.: Image sequence enhancement using subpixel displacements. In: Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, pp. 742–746 (June 1988)
7. Irani, M., Peleg, S.: Super Resolution From Image Sequences. In: Proc. of the 10th Int. Conf. on Pattern Recognition, Atlantic City, NJ, vol. 2, pp. 115–120 (June 1990)
8. Irani, M., Peleg, S.: Improving resolution by image registration. CVGIP: Graphical Models and Image Processing 53(3), 231–239 (1991)
9. Irani, M., Peleg, S.: Motion analysis for image enhancement: Resolution, occlusion and transparency. Journal of Visual Communications and Image Representation 4(4), 324–335 (1993)
10. Kang, S.-J., Yoo, D.-G., Lee, S.-K., Kim, Y.: Multiframe-based bilateral motion estimation with emphasis on stationary caption processing for frame rate up-conversion. IEEE Trans. on Consumer Electronics 54(4), 1830–1838 (2008)
11. Vo, D.T., Nguyen, T.Q., Yea, S., Vetro, A.: Adaptive Fuzzy Filtering for Artifact Reduction in Compressed Images and Videos. IEEE Trans. on Image Processing 18(6) (June 2009)
12. Peleg, S., Keren, D., Schweitzer, L.: Improving image resolution by using subpixel motion. Pattern Recognition Letters 5(3), 223–226 (1987)
13. Avrin, V., Dinstein, I.: Local Motion Estimation and Resolution Enhancement of Video Sequences. In: Proc. of the 14th IEEE Int. Conf. on Pattern Recognition, Washington, DC, USA, vol. 1, pp. 539–541 (August 1998)
14. Messina, G., Battiato, S., Mancuso, M., Buemi, A.: Improving Image Resolution by Adaptive Back-Projection Correction Techniques. Proc. of the IEEE Trans. on Consumer Electronics 48(3), 409–416 (2002)
15. Schultz, R.R., Stevenson, R.L.: Extraction of high-resolution frames from video sequences. IEEE Trans. on Image Processing 5(6), 996–1011 (1996)
16. Tomasi, C., Manduchi, R.: Bilateral Filtering for Gray and Color Images. In: Proc. IEEE Intl. Conf. on Computer Vision (ICCV), Bombay, India (January 1998)
17. Test video sequences, `http://media.xiph.org/video/derf/`