SUPER-RESOLUTION RECONSTRUCTION OF VIDEO SEQUENCES BASED ON BACK-PROJECTION AND MOTION ESTIMATION

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

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

ABSTRACT

Bandwidth limitation and service costs are important factors when dealing with mobile multimedia contents' fruition. Super-resolution reconstruction might be a relevant solution, since it allows for restoring the original spatial resolution from low-resolution compressed data. In this way, both content and service providers, not to tell the final users, are relieved from the burden of providing and supporting large multimedia data transfer. In the proposed work, the high-resolution video sequence is reconstructed through iterative processing of inter-frame information and interpolative techniques. Resolution enhancement is based on iterative update of the high-resolution image estimate through motion and scene change detection. The devised technique is derived from the work of Irani and Peleg [8]. The main contribution of the proposed approach consists in the generalization of the transformation model through the exploitation of local change. Results are encouraging and prove that the devised scheme outperforms alternative techniques.

1. INTRODUCTION

Nowadays, the incredible production of user-generated multimedia contents is leading to serious issues related to their management and maintenance. The combination of increasing bandwidth availability and the development of software technologies allowing for the distributed and collaborative creation of multimedia objects has led to the proliferation of user-generated video communities and, more generally, multimedia information sharing. Although it is generally believed that we are only at the beginning of this era and we are probably going to experience the exponential growth of multimedia content repositories and traffic for the next decade, the massive production and distribution of multimedia data is already a relevant issue for service providers, device designers and software developers. The former are requested to satisfy the ever-growing bandwidth demand; device designers must face the challenge of developing more powerful and compact devices. The latter have to provide better applications and programming frameworks. A fourth category comprises image processing and compression standards professionals, who try to develop better algorithms for coding and reconstructing/recovering the multimedia signals.

Given such scenario, the archetypal use case is that of a user browsing through a huge video database. In order to minimize the bandwidth requirement and the latency, the video streams should be efficiently coded and transmitted at low spatial resolution. Then, the idea is to increase the video stream resolution through super-resolution reconstruction in order to provide the user with additional details that enhance the quality of multimedia browsing, preventing the transmission of additional overhead.

Within the devised context, this paper addresses the problem of super-resolution restoration of video sequences by proposing an approach based on back projection and motion estimation. The resolution enhancement is performed from multiple under-sampled and degraded frames by taking advantage of the additional spatio-temporal data available in the image sequence. In particular, the motion of both scene and camera is the cause for contiguous frames containing similar, but not identical information. The reconstruction of visually superior frames at higher resolution is then based on the exploitation of such inter-frame information. A particular example is provided in Fig. 1.

In this case, the resolution of the left frame can be improved by exploiting the details provided by a subsequent frame (right frame), as highlighted by the rectangle with broken line.

Given the observer's motion, each frame shows further details if compared to adjacent frames. Then, resolution enhancement can be achieved by identifying the corresponding image portions through motion estimation and combining the information from a limited number of frames. Although the provided example constitutes an ideal case since the observer's motion results in the natural zooming of the scene, similar considerations are also

Digital Object Identifier: 10.4108/ICST.MOBIMEDIA2009.7469
http://dx.doi.org/10.4108/ICST.MOBIMEDIA2009.7469

possible when dealing with different motion models and sub-pixel reconstruction.

The organization of the paper is the following. Section 2 gives an overview of the state of the art of super-resolution techniques applied to image sequences and describes the most important past works. Section 3 illustrates the proposed techniques, with the procedure for block-based motion estimation and super-resolution frame reconstruction. Experimental results are discussed in Section 4. Conclusions are finally drawn in Section 5.

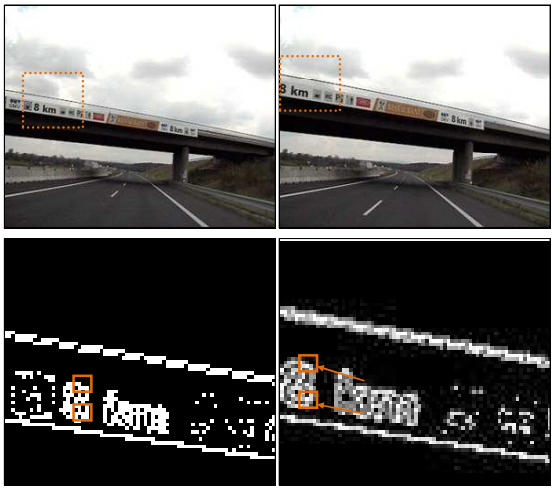

Fig. 1. Example of inter-frame change analysis.

2. PAST WORKS

This section presents a selection of the most significant approaches that address the issue of video 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.

A classification of super-resolution approaches is based on the number of input images, resulting in Multi-Frame and Single-Frame techniques (Fig. 2). Multi-frame techniques differ on the basis of whether the input data are static images (MISO - Multiple Input Single Output) or motion pictures (MIMO - Multiple Input Multiple Output). In both cases, the resolution improvement is obtained through the fusion of information coming from different frames at low resolution and generally implies inter-frame de-correlation through differencing or motion estimation. Although the focus of this work will be on video MIMO techniques, MISO and MIMO methods often blend into similar concepts.

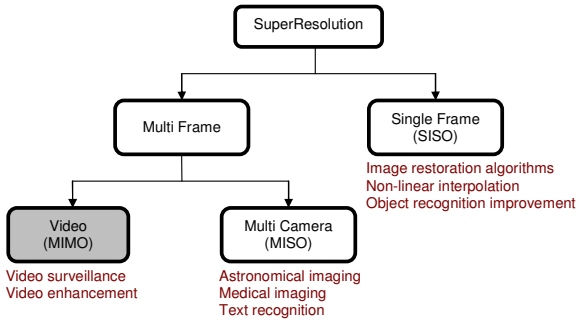

Fig. 2. Classification of Super-Resolution methods.

Multi-Frame approaches are generally divided into frequency and spatial domain techniques. Among the former, Tsai and Huang. [1] present an algorithm that improves the resolution of Landsat image data. Landsat acquires several images of partially overlapping areas of the earth in the course of its orbits, thus producing a sequence of similar, but not identical images. Observed images are modeled as under-sampled versions of an unchanging scene undergoing global translational motion. In [2] a frequency domain formulation is proposed, based on the shift and aliasing properties of the continuous and discrete Fourier transforms for the reconstruction of a band-limited image from a set under-sampled, and therefore aliased, observations. Several limitations of the Tsai-Huang method are addressed by Tekalp, Ozkan and Sezan in [3]. The authors propose a frequency domain approach which extends [1] by including the effects of a LSI PSF as well as observation noise. Periodic sampling is still assumed and a translation-only motion model is used. Kim, Bose and Valenzuela [4] exploit the frequency domain theoretical framework and the global translation observation model proposed in [1] and consider observation noise as well as the effects of spatial blurring.

Among spatial domain methods, Keren, Peleg and Brada [5] propose an approach to image registration based on a global translation and rotation model, as well as a two stage approach to super-resolution reconstruction. The first stage is a simple interpolation technique and the second consists in a motion estimation algorithm. Irani and Peleg [6, 7] extend the earlier work by improving the means of backprojecting the error between the simulated low-resolution images and the observed data. Then, a very general procedure for super-resolution reconstruction is proposed in [8], for scenes which contain arbitrary independent motion, by the incorporation of a multiple motions tracking algorithm which allows super-resolution reconstruction for partially

Digital Object Identifier: 10.4108/ICST.MOBIMEDIA2009.7469
http://dx.doi.org/10.4108/ICST.MOBIMEDIA2009.7469

occluded objects, transparent objects or some object of interest. An interpolation based technique is proposed by Aizawa, Komatsu and Saito [9]. They examine the problem of acquiring high-resolution imagery from stereo cameras. By considering the possibility of sampling at spatial positions between the array pixels, it is demonstrated that the effective frequency response of the combined (double image) system is increased.

An early algebraic tomographic filtered backprojection approach to super-resolution reconstruction is that of Frieden and Aumann, [10]. The authors consider the problem of super-resolution image reconstruction from multiple 1D scans of a stationary scene by a linear imaging array. The imaging geometry consists in overlapping scans of a given scene area, enabling reconstruction at a resolution higher than the limiting spatial sampling rate of the sensor array. In [11] 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 [5]. A dynamic super-resolution sequence reconstruction from a lower resolution sequence containing sub-pixel shifts is presented in [12]. The main features of this work are related to: the local motion estimation performed using the group delays of local adaptive linear prediction filters, in order to obtain a motion vector for each pixel in the image; and the application of the super-resolution improvement to the sequence images rather than to a prototype image. An iterated back-projection based (IBP) algorithm, improved by adaptive techniques, is presented in [13]. The proposed approach is based on the uncertainty degree metric, used in pixel reconstruction and error correction. A global motion estimation algorithm based on extracting 1-dimensional characteristic curves from subsequent frames with sub-pixel displacement values is considered.

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) [14]. Both the spatial and temporal information present in a short image sequence is used to create a single high-resolution video frame.

Another approach to super-resolution reconstruction is based on projection onto convex sets. In [15] the scanning linear array problem originally discussed by [10] is addressed, as well as the problem of restoring a super-resolution image from multiple plane array images. A variant of the POCS based formulation using an ellipsoid to bound the constraint sets has been investigated by Tom and Katsaggelos [16-18].

3. PROPOSED APPROACH

The proposed technique is aimed at reconstructing a high-definition video from a limited number of frames extracted from a low-resolution sequence, without any preliminary knowledge of the high-definition data. The process is based on backprojection and motion estimation. For any given frame, a sliding window determines the set of low resolution frames to be processed in order to produce the output stream. The window is shifted forward to produce successive super-resolution frames of the output sequence, as shown in Fig. 3.

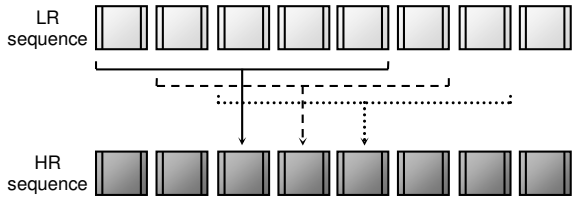

Fig. 3. Super-resolution video enhancement from a LR image sequence.

The main idea is that each pixel in a LR frame is a "projection" of a region in the scene. The HR image is constructed using an approach similar to the back projection method used in CAT (Computed Aided Tomography). Accurate knowledge of the relative scene locations sensed by each pixel in the observed images (LR) is necessary for super-resolution. This information is available in image regions where local deformation can be described by some parametric functions.

In the following paragraphs, the proposed algorithm is described is terms of imaging process, super-resolution and motion estimation in order to achieve the reconstruction of higher resolution frame which approximates the original one as accurately as possible.

3.1. Background

The first approach to super-resolution based on the iterated process of backprojecting the error between the estimated LR images and the observed data was proposed in [5] and further extended in [6-8]. The algorithm performs an initial estimate of the high resolution image; then, the subsampling/degradation process is simulated in order to deduce the set of LR frames which correspond to the observed input images. The difference (error) between the simulated and the observed frames is computed in order to update the initial HR frame estimate through the error backprojection. The process is iterated in accordance to an error minimization criterion. Only translation and rotation

Digital Object Identifier: 10.4108/ICST.MOBIMEDIA2009.7469
http://dx.doi.org/10.4108/ICST.MOBIMEDIA2009.7469

were considered for modeling the HR estimate and LR subsampling.

The relative displacements of the input images at subpixel accuracy are computed and an iterative refinement is adopted to improve accuracy. It is assumed that the imaging process for the observed image sequence (LR) is modeled by:

$$g^{(n)}(\vec{y}) = \sum_{\vec{x}} f^{(n)}(\vec{x}) \cdot h^{PSF}(\vec{x} - \vec{z}_{\vec{y}}) \quad (1)$$

where $g^{(n)}$ is the LR image obtained by applying the simulated imaging process to $f^{(n)}$; $f^{(n)}$ is the approximation of f obtained after n iterations; $\vec{x}$ and $\vec{y}$ denote HR and LR pixels respectively, the latter influenced by $\vec{x}$; $\vec{z}_{\vec{y}}$ is the center of the receptive field of $\vec{y}$ in $f^{(n)}$; h^{PSF} is the point spread function of the imaging blur; f is the target HR image to be constructed (unknown).

The iterative update scheme to estimate the HR image f is then:

$$f^{(n+1)}(\vec{x}) = f^{(n)}(\vec{x}) + \sum_{\vec{y} \in U_k Y_{k,\vec{x}}} (g_k(\vec{y}) - g_k^{(n)}(\vec{y})) \cdot \frac{(h_{\vec{x}\vec{y}}^{BP})^2}{c \sum_{\vec{y}' \in U_k Y_{k,\vec{x}}} (h_{\vec{x}\vec{y}'}^{BP})} \quad (2)$$

where $Y_{k,\vec{x}}$ is the set $\{ \vec{y} \in g_k \mid \vec{y}$ influenced by $\vec{x} \}$; c is a constant normalizing factor and $h_{\vec{x}\vec{y}}^{BP} = h^{BP}(\vec{x} - \vec{z}_{\vec{y}})$.

The error function to be minimized is:

$$\varepsilon^{(n)} = \sqrt{\sum_{k} \sum_{m_1, m_2} \left(g_k(y_1, y_2) - g_k^{(n)}(y_1, y_2) \right)^2} \quad (3)$$

Since the choice of the initial estimate does not influence the performance of the algorithm, the average of the LR frames is used as $f_i^{(0)}$; then, it is assumed that $h^{BP} = h^{PSF}$.

3.2. Super-Resolution

Starting from the devised scheme, the proposed work introduces several changes in order to outperform alternative techniques.

Let f_i denote the target frame to be reconstructed through the super-resolution method; we then extract k frames from the original LR video sequence: $(k-1)/2$ past and $(k-1)/2$ future frames. Differently from Peleg and Irani's method, the initial estimate for the high-resolution frame ($n = 0$) is a linear interpolation of the low-resolution one, g_i. The blocks of the neighboring LR frames which are found to be significantly similar to those of the reference frame are merged into the HR approximation. Such process resembles a projection, and is done according to the zoom factor and the estimated motion, in order to reconstruct the high-definition data. The block-based motion estimation is described in detail in Section 3.3. Residual information is restored through linear interpolation.

Such process is repeated for each of the k LR frames in order to obtain an approximation of k HR frames. They are then subsampled with the h^{PSF} filter to obtain the simulated LR frame sequence $g_{i-((k-1)/2)}^{(n)}, \ldots, g_i^{(n)}, \ldots, g_{i+((k-1)/2)}^{(n)}$. The difference between simulated and reconstructed LR frames are computed and the error is backprojected into the HR estimate in order to refine the restoration process.

The procedure is iterated until the error becomes appreciably small or a maximum number of iterations, n, is reached. Finally the reconstructed frame, $f_i^{(n)}$, is assumed as high-resolution approximation of g_i, $f_i^{(n)} \cong f_i$.

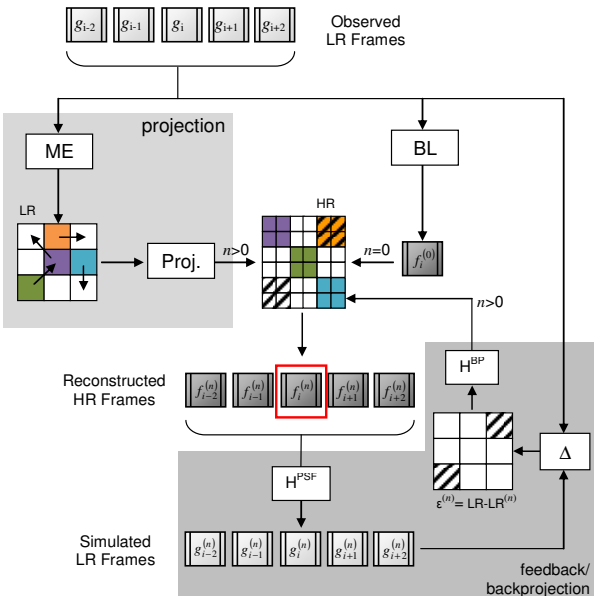

Fig. 4. Block scheme of the proposed method.

3.3. Motion Estimation

Motion estimation plays an important role in the video super-resolution reconstruction. Among motion estimation techniques, [19] and [20] have been considered, in which a

Digital Object Identifier: 10.4108/ICST.MOBIMEDIA2009.7469
http://dx.doi.org/10.4108/ICST.MOBIMEDIA2009.7469

three-step search algorithm is proposed. It employs a center-biased checking point pattern in the first step, which is derived by making the search adaptive to the motion vector distribution, and a halfway-stop technique to reduce the computational cost. In details, by considering a block of size $N \times N$, the block motion estimation searches for a motion vector in a previous frame that yields the minimum block distorsion measurement (BDM) in the scanning area. To do this, a multiple stage search is implemented:

1. the central point, the 8 points at $p/2$ in the scanning area ($p \times p$) and 8 extra neighbours points at 3-pixel distance are checked;
2. a halfway stop technique is used to estimate the stationary or quasistationary block's motion:
   a. if the minumum BDM in step 1 occurs at the search windows center, stop the search (first step stop);
   b. if the minumum BDM point in step 1 is one of the 8 neighbours of the window center, the search is performed for the 8 neighbouring points of the minimum only (second step stop).

A complete three step search is only performed when the minimum BDM point at the first step is not the window center, nor any of its 8 neighbours.

3.4. Method's parameters

In this section an overview of the method's main parameters is given. In particular:

- $f_i^{(0)}$ is the initial HR frame estimate. It is computed as linear interpolation of the corresponding LR frame only, g_i.
- h^{PSF} is the point spread function of the imaging system. In this implementation it represents a Gaussian filter.
- For each HR frame, $k = 5$ LR frames are considered for processing.
- A number of $n = 5$ maximum iterations is imposed.
- The zoom factor tested are $4\times$ and $8\times$. The blocks are 4×4 and 8×8 pixel wide respectively.
- The scanning area is 16×16 pixel.
- The mean square error is considered for BDM.

4. EXPERIMENTAL RESULTS

The proposed method has been evaluated with a test set of 7 video sequences in the 4:2:0 YUV format, chosen among classical video processing test sets [21]. The test video sequences have been selected with the purpose of presenting a broad range of signal behaviors, in terms of different motion.

To provide objective results, a subsampled video sequence is preliminarily produced (LR) from the original video (OR) and used as input sequence for the devised algorithm at any given zoom factor. Then, PSNR is computed between the original and the reconstructed (HR) signal (Fig. 5).

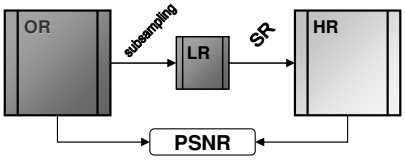

Fig. 5. Evaluation of the objective results.

The proposed method (SR) is computed for two different parameters sets (SR_A: $\sigma = 10; c = 0.1$, SR_B: $\sigma = 2; c = 0.5$) and is compared with the nearest neighbor (NN) and bilinear (BL) interpolation. Results are expressed in terms of overall average PSNR among the complete test set, at 4 and 8× zoom factor (Fig. 6 and 7 respectively).

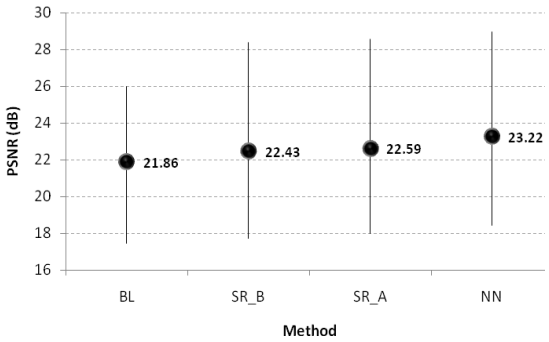

Fig. 6. Average PSNR results, 4× zoom factor.

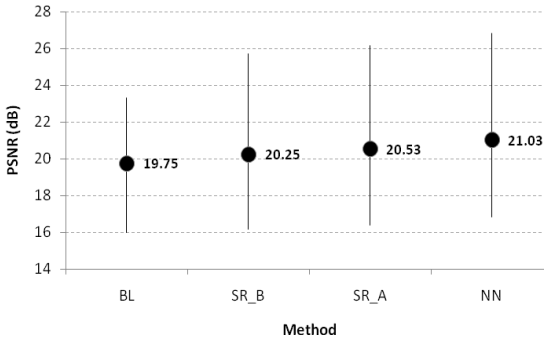

Fig. 7. Average PSNR results, 8× zoom factor.

Digital Object Identifier: 10.4108/ICST.MOBIMEDIA2009.7469
http://dx.doi.org/10.4108/ICST.MOBIMEDIA2009.7469

As expected, reconstruction quality decreases as the zoom factor increases. Such behaviour characterizes all methods and derives from the increasing lack of information. In fact, the subsampled frames constitute the only piece of known information for all reconstruction methods.

Objective results show that the proposed method lies between the nearest neighbour and the bilinear interpolation. The superior performance of the nearest neighbour interpolation is easily explained. Since subsampling is carried out through block averaging, nearest neighbour substitution simply assigns the local average to unknown pixels, thus approximating their value with the best estimation in terms of mean square error, thus PSNR.

Then, the proposed method outperforms bilinear interpolation by 0.65dB at 4× zoom factor and 0.89dB at 8× zoom factor on average. It must be observed that the proposed method's performance increases at higher zoom factors if compared to bilinear interpolation.

The results for two sequences are reported in Fig. 8 and 9.

As previously noticed, nearest neighbor generally outperforms any competing method. Then, PSNR is not fully adequate in providing a reliable quality index. It is a measure that provides an approximate performance indication and cannot be considered as an accurate indication of reconstruction quality. In fact, it does not take into account the issues related to the human visual system and subjective scene interpretation.

Visual results are provided in Figs. 10 and 11 in order to subjectively evaluate the proposed method. The proposed method results appear visually more pleasant than the competitors. In particular, a detail of the Y component of the "Bus" sequence is reconstructed through NN, SR and BL and is shown in Fig. 10 for comparison. The SR reconstruction appears sharper and more defined than the competing techniques.

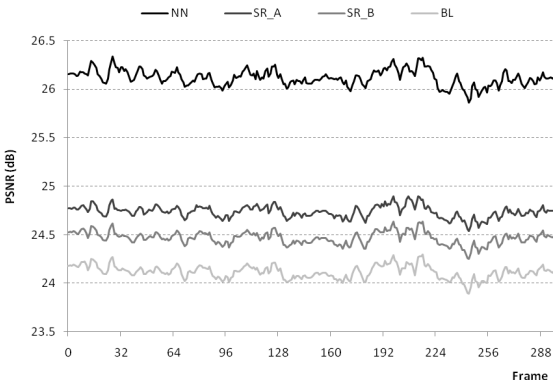

Fig. 8. PSNR results for the sequence "Silent" at 4× zoom factor.

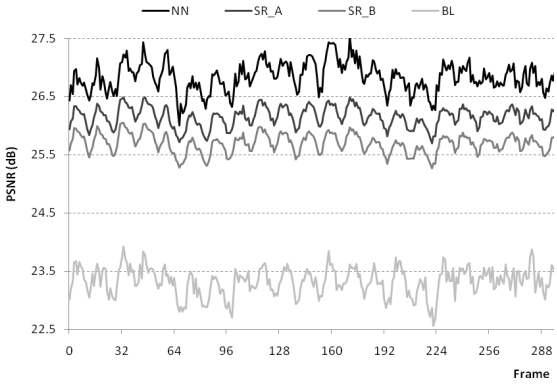

Fig. 9. PSNR results for the sequence "Highway" at 8× zoom factor.

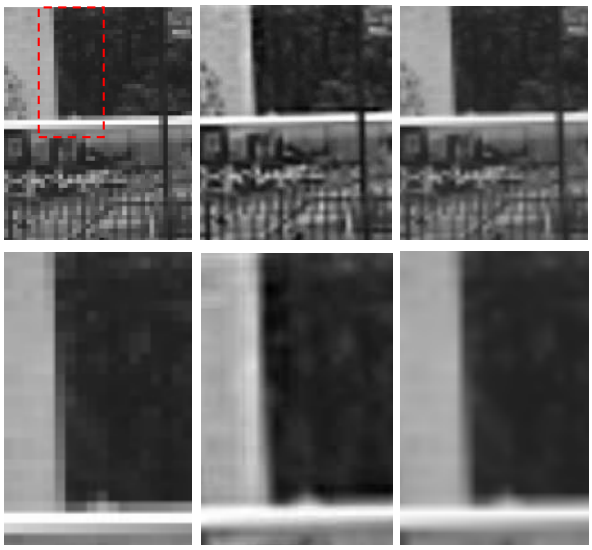

Fig. 10. A detail from the sequence "Bus" at 4× zoom factor; from left: NN, SR, BL.

Digital Object Identifier: 10.4108/ICST.MOBIMEDIA2009.7469
http://dx.doi.org/10.4108/ICST.MOBIMEDIA2009.7469

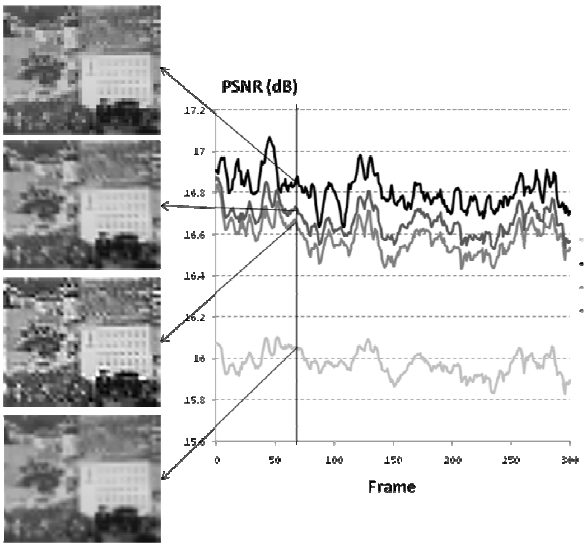

Fig. 11. Visual results for the sequence "mobile" at 8× zoom factor.

5. CONCLUSIONS

In this work, an iterative technique for high-resolution reconstruction of low-resolution video sequences has been presented. The proposed algorithm is developed from Peleg and Irani's works with the generalization of the motion estimation model and modifications to the system architecture. The proposed reconstruction method uses block-based motion estimation and backprojection of the error between the restored frame and the simulated one. Results are promising, especially when considering high zoom factors. Future developments may exploit different motion models and investigate the integration of novel human visual system-based techniques.

REFERENCES

[1] R.Y. Tsai and T.S. Huang, "Multiframe image restoration and registration," in Advances in Computer Vision and Image Processing, R.Y. Tsai and T.S. Huang, Eds., vol. 1, pp. 317–339. JAI Press Inc., 1984.

[2] R.A. Roberts and C.T. Mullis, Digital Signal Processing, Addison-Wesley, 1987.

[3] A. M. Tekalp, M. K. Ozkan, and M. I. Sezan, "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.

[4] S.P. Kim, N.K. Bose, and H.M. Valenzuela, "Recursive reconstruction of high resolution image from noisy undersampled multiframes," IEEE Trans. on Acoustics, Speech and Signal Processing, vol. 38, no. 6, pp. 1013–1027, 1990.

[5] D. Keren, S. Peleg, and R. Brada, "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.

[6] M. Irani and S. Peleg, "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.

[7] M. Irani and S. Peleg, "Improving resolution by image registration," CVGIP: Graphical Models and Image Processing, vol. 53, no. 3, pp. 231–239, May 1991.

[8] M. Irani and S. Peleg, "Motion analysis for image enhancement: Resolution, occlusion and transparency," Journal of Visual Communications and Image Representation, vol. 4, no. 4, pp. 324–335, Dec. 1993.

[9] K. Aizawa, T. Komatsu, and T. Saito, "Acquisition of very high resolution images using stereo cameras," Visual Communications and Image Processing Conf., 1991, Proc. SPIE, pp. 318–328, 1991.

[10] B.R. Frieden and H.G. Aumann, "Image reconstruction from multiple 1-D scans using filtered localized projection," Applied Optics, vol. 26, no. 17, pp. 3615–3621, Sept. 1987.

[11] S. Peleg, D. Keren, and L. Schweitzer, "Improving image resolution by using subpixel motion," Pattern Recognition Letters, vol. 5, no. 3, pp. 223–226, Mar 1987.

[12] V. Avrin and I. Dinstein, "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, Aug. 1998.

[13] G. Messina, S. Battiato, M. Mancuso and A. Buemi, "Improving Image Resolution by Adaptive Back-Projection Correction Techniques," in Proc. of the IEEE Trans. on Consumer Electronics, vol. 48, no. 3, pp 409–416, Aug. 2002.

[14] R.R. Schultz and R.L. Stevenson, "Extraction of high-resolution frames from video sequences," IEEE Trans. on Image Processing, vol. 5, no. 6, pp. 996–1011, June 1996.

[15] H. Stark and P. Oskoui, "High-resolution image recovery from image-plane arrays, using convex projections," Journal of the Optical Society of America, Journal, A: Optics and Image Science, vol. 6, no. 11, pp. 1715–1726, 1989.

[16] B.C. Tom and A.K. Katsaggelos, "An Iterative Algorithm for Improving the Resolution of Video Sequences," in Visual Communications and Image Processing, Orlando, FL, Proc. SPIE, vol. 2727, pp. 1430–1438, Mar. 1996.

[17] B.C. Tom and A.K. Katsaggelos, "Resolution enhancement of video sequences using motion compensation," in Proc. of the IEEE Int. Conf. on Image Processing, Lausanne, Switzerland, vol. 1, pp. 713–716, Sept. 1996.

[18] A.K. Katsaggelos, "A Multiple Input Image Restoration Approach," Journal of Visual Communication and Image Representation, vol. 1, no. 1, pp. 93–103, Sept. 1990.

[19] R. Li, B. Zeng and M.L. Liou, "A new three-step search algorithm for block motion estimation," IEEE Trans. on Circuits and Systems for Video Technology, vol. 4, no. 4, pp 438–442, Aug. 1994.

[20] X. Jing and L.P. Chau, "An efficient three-step search algorithm for block motion estimation," IEEE Trans. on Multimedia, vol. 6, no. 3, pp 435–438, June 2004.

[21] Test video sequences: http://trace.eas.asu.edu/yuv/index.html.

Digital Object Identifier: 10.4108/ICST.MOBIMEDIA2009.7469
http://dx.doi.org/10.4108/ICST.MOBIMEDIA2009.7469