IMAGE AND VIDEO SUPER-RESOLUTION VIA SPATIALLY ADAPTIVE BLOCK-MATCHING FILTERING

Aram Danielyan, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian

Department of Signal Processing, Tampere University of Technology
P.O. Box 553, 33101, Tampere, Finland
web: www.cs.tut.fi/~comsens email: firstname.lastname@tut.fi

ABSTRACT

In our recent work [6], we proposed an algorithm for image upsampling based on alternation of two procedures: spatially adaptive filtering in image domain and projection on the observation-constrained subspace in a wavelet domain. The nonlocal Block-Matching 3-D (BM3D) filter was used to suppress ringing and reconstruct missing detail coefficients. Here we generalize this method in two aspects. First, we reformulate observation model and reconstruction algorithm from wavelets to a general class of scaling transforms, and second, we extend applicability from single image upsampling to image and video super-resolution.

Experimental results demonstrate significant improvement over the upsampling method [6] and an overall performance on the level of the best existing super-resolution methods.

1. INTRODUCTION

Image *upsampling* or *zooming*, can be defined as the process of resampling a *single* low-resolution (LR) image on a high-resolution grid. Different resampling methods can be used to obtain zoomed images with specific desired properties, such as edge preservation, degree of smoothness, etc. In [6], we proposed a wavelet-domain image *upsampling* algorithm based on the iterative spatially adaptive Block-Matching 3D filtering (BM3D) [3], assuming that the given low-resolution image is obtained from the approximation subband of an orthonormal wavelet decomposition of a high-resolution (HR) original image. Its effectiveness was proven in a number of experiments, showing a significant improvement over some of the best methods known in the field [12], [11].

However, fine details missing or distorted in the low-resolution image cannot be reconstructed in the upsampled one. Roughly speaking, there is no sufficient information in the low-resolution image to do this. The situation changes when a number of LR images portraying slightly different views of the same scene are available. The reconstruction algorithm now can try to improve the spatial resolution by incorporating into the final HR result the additional new details revealed in each LR image. The process of combining a sequence of undersampled and degraded low-resolution images in order to produce a single high-resolution image is commonly referred to as a *super-resolution (SR) reconstruction*.

The classical SR approach is loosely based on the following three steps: 1) registration of the LR images to a HR coordinate grid, 2) warping of the LR images onto that grid by interpolation, and 3) fusion of the warped images into the final HR image. An additional deblurring step is sometimes considered to compensate the blur existing in the LR frames. Several algorithms based on such classical approach exist and detailed reviews can be found, e.g., in [9] and [14]. For successful reconstruction it is crucial to perform accurate registration. Most of the existing SR methods rely either on a parametric global motion estimation, or on a computationally intensive optical flow calculation**.** However, an explicit registration of the LR frames is often not feasible: on the one hand, if the registration map has few degrees of freedom, it is too rigid to model the geometrical distortions caused by the lens system; on the other hand, when many degrees of freedoms are available (e.g., a dense optical flow), reliable estimates of the registration parameters cannot be obtained. In either case, registration artifacts are likely to appear in the fusion, requiring heavy regularization (smoothing) for their concealment [9]. The situation becomes even more difficult when non-global motion is present in images. Modern SR methods depart from the classical approach and we specially mention two recent SR reconstruction algorithms [13], [7] based on the nonlocal means (NLM) filtering paradigm [1]. In these algorithms, instead of trying to obtain an explicit registration as a one-to-one pixel mapping between frames, a one-to-many mapping is utilized, where multiple pixels can be assigned to a given one, with weights typically defined by the similarity of the patches/blocks surrounding the pixels. The HR image is estimated through a weighted average of these multiple pixels (or of their surrounding patches) with their corresponding weights.

The increased redundancy of the NLM, which can exploit also multiple patches from a same frame, contributes significantly to the overall good performance of the methods [13].

The BM3D algorithm, used as iterative filter in [6], shares with the NLM the idea of exploiting nonlocal similarity between blocks. However, in BM3D a more powerful transform-domain modeling is used and, as shown in [10], the BM3D turns out to be a much more effective filter than the NLM. Thus, our image upsampling approach [6] is naturally suited to be generalized for successful application to the SR reconstruction problem. This is essentially achieved by replacing the BM3D *image* filter with the V-BM3D *video* filter [2]. As opposed to the BM3D, which searches and combines blocks only within a single image, the V-BM3D operates across multiple frames, exploiting similar blocks also from different frames.

Further, in order to accurately deal with a wider class of LR downsampling schemes, we reformulate our observation model and recursive reconstruction algorithm from the wavelet domain to a general class of scaling transforms (which include wavelets, DCT, DFT, as well as their corresponding blockwise counterparts).

The paper is organized as follows. In the next section, we introduce our observation modeling for the low-resolution images and provide a detailed description of the both *upsampling* and *super-resolution* algorithms in a unified form. Then, in Section 3, we present experimental results, demonstrating the feasibility of the approach, and provide a comparison with the state-of-the-art. Discussions and conclusion are given in the last section.

2. THE ALGORITHM

2.1 Preliminaries: scaling family of transforms

Let $\{\mathcal{T}_m\}_{m=0}^{M}$ be a family of orthonormal transforms of increasing sizes $x_m^{\mathrm{h}} \times x_m^{\mathrm{v}}$, $x_m^{\mathrm{h}} < x_{m+1}^{\mathrm{h}}$, $x_m^{\mathrm{v}} < x_{m+1}^{\mathrm{v}}$, such that for any pair m,m' with $m < m'$, up to a scaling factor $\beta_{m,m'} = \sqrt{\frac{x_{m'}^{\mathrm{h}} x_{m'}^{\mathrm{v}}}{x_m^{\mathrm{h}} x_m^{\mathrm{v}}}}$, the whole $\mathcal{T}_m$-spectrum can be considered as a smaller portion of the $\mathcal{T}_{m'}$-spectrum. In particular, this means that the supports Ω_m of

This work was supported by the Academy of Finland (application no. 213462, Finnish Programme for Centres of Excellence in Research 2006-2011, and application no. 118312, Finland Distinguished Professor Programme 2007-2010).

$$\begin{cases} \hat{y}_{r,0} = y_{\text{low}\,r} \\ \hat{y}_{r,m} = \hat{y}_{r,m}^{(k_{\text{final}\,m})} \\ \hat{y}_{r,m}^{(0)} = \mathcal{T}_m^{-1}\left(\mathcal{U}_{m-1,m}\left(\beta_{m-1,m}\mathcal{T}_{m-1}\left(\hat{y}_{r,m-1}\right)\right)\right) \\ \hat{y}_{r,m}^{(k)} = \mathcal{T}_m^{-1}\left(\mathcal{U}_{0,m}\left(\beta_{0,m}\mathcal{T}_0\left(y_{\text{low}\,r}\right)\right) + \mathcal{P}_{0,m}^{\perp}\left(\mathcal{T}_m\left(\Phi\left(r, \left\{\hat{y}_{r,m}^{(k-1)}\right\}_{r=1}^{R}, \sigma_{k,m}\right)\right)\right)\right) \end{cases} \tag{1}$$

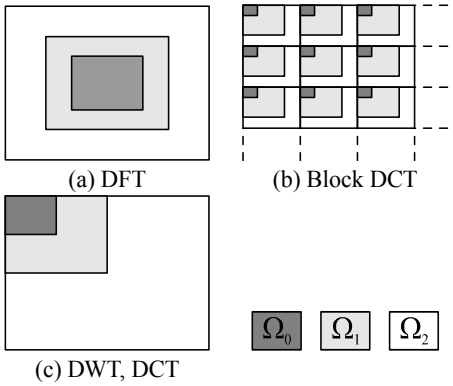


Figure 1: Nested support subsets. Ω_0 and Ω_1 are shown as a submatrices of matrix $\Omega_M = \Omega_2$ of coefficients of the transform $\mathcal{T}_M$.

the $\mathcal{T}_m$-transform coefficients form a *nested sequence* of subsets of Ω_M, i.e. $\Omega_0 \subset \cdots \subset \Omega_M$, where Ω_M is a set of the coefficients. The most notable examples of such $\{\mathcal{T}_m\}_{m=0}^{M}$ families are discrete cosine (DCT) or Fourier (DFT) transforms of different sizes, discrete wavelet transforms (DWT) associated to one common scaling function, as well as block DCT, DFT and DWT transforms. Figure 1 illustrates the nested sequence of supports for these families.

Depending on the specific transform family of choice, the sets Ω_m are commonly referred to as *lower-resolution, low-frequency,* or *coarser-scale subbands* of the $\mathcal{T}_M$-spectrum.
For $m < m'$ we define three operators:

- the *restriction operator* $|_{\Omega_{m,m'}}$ that, from a given $\mathcal{T}_{m'}$-spectrum, extracts its smaller portion defined on Ω_m, which can be thus considered as the $\mathcal{T}_m$-spectrum of a smaller image;
- the *zero-padding operator* $\mathcal{U}_{m,m'}$ that expands a $\mathcal{T}_m$-spectrum defined on Ω_m to the $\mathcal{T}_{m'}$-spectrum defined on the superset $\Omega_{m'} \supset \Omega_m$ by introducing zeros in the complementary $\Omega_{m'} \setminus \Omega_m$;
- the *projection operator* $\mathcal{P}_{m,m'}^{\perp}$ that zeroes out all coefficients of $\mathcal{T}_{m'}$-spectrum defined on Ω_m.

Note that $\mathcal{U}_{m,m'}(A)|_{\Omega_m} = A$ for any $\mathcal{T}_m$-spectrum A, and $B = \mathcal{P}_{m,m'}^{\perp}(B) + \mathcal{U}_{m,m'}\left(B|_{\Omega_m}\right)$ for any $\mathcal{T}_{m'}$-spectrum B. Thus, $\mathcal{U}_{m,m'}$ can be regarded as "dual" operator of $|_{\Omega_{m,m'}}$.

2.2 Observation model

Let us be given a sequence of $R \geq 1$ *low-resolution images* $\{y_{\text{low}\,r}\}_{r=1}^{R}$ of size $x_0^{\text{h}} \times x_0^{\text{v}}$, and assume each $y_{\text{low}\,r}$ being reconstructed from the subband of the corresponding $\mathcal{T}_M$ spectra of original *higher-resolution* images $\{y_{\text{hi}\,r}\}_{r=1}^{R}$ of size $x_M^{\text{h}} \times x_M^{\text{v}}$ in the following way:

$$y_{\text{low}\,r} = \mathcal{T}_0^{-1}\left(\beta_{0,M}^{-1}\,\mathcal{T}_M\left(y_{\text{hi}\,r}\right)|_{\Omega_{0,M}}\right), \tag{2}$$

where the scaling factor $\beta_{0,M}$ ensures that the means of $y_{\text{hi}\,r}$ and $y_{\text{low}\,r}$ are the same.

The problem is to reconstruct $\{y_{\text{hi}\,r}\}_{r=1}^{R}$ from $\{y_{\text{low}\,r}\}_{r=1}^{R}$. Clearly, for a fixed r, any good estimate $\hat{y}_r$ of $y_{\text{hi}\,r}$ must have its Ω_0 subband equal to $\beta_{0,M}\mathcal{T}_0\left(y_{\text{low}\,r}\right) = \mathcal{T}_M\left(y_{\text{hi}\,r}\right)|_{\Omega_{0,M}}$. Under this restriction, the estimates constitute an affine subspace $T_{y_{\text{low}\,r}}$ of codimension $x_0^{\text{h}}x_0^{\text{v}}$ in a $x_M^{\text{h}}x_M^{\text{v}}$-dimensional linear space T: $T_{y_{\text{low}\,r}} = \left\{\hat{y}_r \in T : T_M\left(\hat{y}_r\right)|_{\Omega_{0,M}} = \beta_{0,M}\mathcal{T}_0\left(y_{\text{low}\,r}\right)\right\}$.

For $R = 1$, the observation model (2) corresponds to the image upsampling problem (as presented in [6] for the specific case of a wavelet transform family). Whenever $R > 1$, we are instead in the image or video super-resolution setting.

2.3 Multistage iterative reconstruction

Scaling from $x_0^{\text{h}} \times x_0^{\text{v}}$ size to $x_M^{\text{h}} \times x_M^{\text{v}}$ is performed using the transform family $\{\mathcal{T}_m\}_{m=0}^{M}$, progressively across M *stages*, which are indicated using the subscript $m = 1, \ldots, M$. The complete algorithm is defined by the recursive system given in (1). At each stage, the images are being super-resolved from size $x_{m-1}^{\text{h}} \times x_{m-1}^{\text{v}}$ to $x_m^{\text{h}} \times x_m^{\text{v}}$. $\{y_{\text{low}\,r}\}_{r=1}^{R}$ serves as input for the first stage, and the output of the current stage $\left\{\hat{y}_{r,m}\right\}_{r=1}^{R}$ becomes an input for the next one. At each stage, the initial estimate $\hat{y}_{r,m}^{(0)}$ is obtained from $\hat{y}_{r,m-1}$ by zero-padding its spectra following the third equation in (1). During the subsequent iterations, the estimates are obtained according to the last equation in (1), where the superscript $k = 0, 1, 2, \ldots$ corresponds to the iteration count inside each stage, $\hat{y}_{r,m}^{(k)}$ is a sequence of estimates for $\hat{y}_{r,m}$, Φ is a spatially adaptive filter and $\sigma_{k,m}$ is a parameter controlling the strength of this filter. In other words, at each iteration we jointly filter the images $\left\{\hat{y}_{r,m}^{(k-1)}\right\}_{r=1}^{R}$ obtained from the previous iteration, perform a transform $\mathcal{T}_m$ for each r, substitute the $x_0^{\text{h}} \times x_0^{\text{v}}$ coefficients defined on Ω_0 with $\beta_{0,M}\mathcal{T}_0\left(y_{\text{low},r}\right)$, and take an inverse transform $\mathcal{T}_m^{-1}$ to obtain $\hat{y}_{r,m}^{(k)}$. The flowchart of the system (1) is presented in Figure 2. The iteration process stops at iteration $k_{\text{final}\,m}$ when the distance between $\left\{\hat{y}_{r,m}^{(k)}\right\}$ and $\left\{\hat{y}_{r,m}^{(k-1)}\right\}$ in some metric becomes less than a certain threshold δ_0, or if the maximum number of iterations $k_{\max\,m}$ is reached.

2.4 The filter

Detailed description of the V-BM3D denoising filter, used in the algorithm as the spatially adaptive filter Φ, can be found in [2]. In brief, the filter works as follows.

1. *Block-wise estimates*. Each image in a sequence is processed in sliding-window manner. For each block the filter performs:
   (a) *Grouping*. Searching within all images in the sequence, find blocks that are similar to the currently processed one, and then stack them together in a 3-D array (group).
   (b) *Collaborative hard-thresholding*. Apply a 3-D transform to the formed group, attenuate the noise by hard-thresholding of the transform coefficients, invert the 3-D transform to produce estimates of all grouped blocks, and return the estimates of the blocks to their original place.
2. *Aggregation*. Compute the estimates of the output images by weighted averaging all of the obtained block-wise estimates that are overlapping.

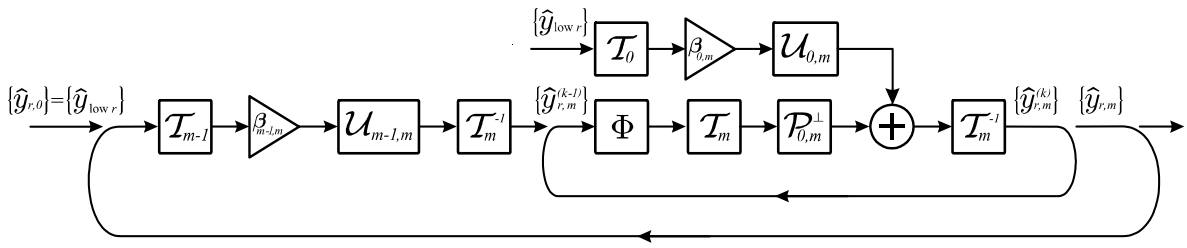


Figure 2: Multistage iterative reconstruction.The inner loop corresponds to an iteration process inside a stage; the outer loop corresponds to a transition to the next stage.

Due to the similarity between the grouped blocks, the transform can achieve a highly sparse representation of the true signal so that the noise or small distortions can be well separated by shrinkage. In this way, the collaborative filtering reveals even the finest details shared by grouped fragments and at the same time it preserves the essential unique features of each individual fragment.

For the purposes of this work, we do not perform the final collaborative Wiener filtering stage of the original V-BM3D denoising algorithm.

Here, the parameter $\sigma_{k,m}$ is used in place of the standard-deviation of the noise. This parameter controls the strength of the filter. In order to prevent smearing of the small details, the sequence $\{\sigma_{k,m}\}_{k=0,1,\ldots}$ should be decreasing with the progress of the iterations.

Parameters	stage	
	1	2
$k_{\text{final } m}$	20	20
$\sigma_{0,m}$	60	35
$\Delta_{\sigma\ m}$	2.5	1.5
block size	12	8
spatial search window	15	25
temporal search window (SR/upsampling)	9/1	9/1
number of blocks in a group (SR/upsampling)	128/32	128/32

Table 1: V-BM3D filter parameters used in the SR and upsampling algorithm.

3. EXPERIMENTAL RESULTS

3.1 General remarks

We consider two sets of experiments. First, in Section 3.3, we assess how well our SR algorithm can reconstruct a single HR image provided that the set of LR images covers the whole HR sampling grid. Second, in Section 3.4, we compare our algorithm for SR of video sequences, comparing against the state-of-the-art method [13].

As seen from the previous section, the algorithm formulations for upsampling and super-resolution coincide. In both cases, the algorithm performs reconstruction of each image of the input sequence, and the output sequence always contains the same number of frames as input. Whether the algorithm performs upsampling or SR reconstruction, depends on the parameter controlling the number of frames where V-BM3D searches for the similar blocks (so-called "temporal search window"), which is typically limited in order to contain complexity. When this parameter is set to one, the search is restricted to the current frame only and the algorithm independently *upsamples* each frame of the input sequence.

In all these experiments, we consider the same downsampling model, where the LR image is obtained from the HR one by first blurring using a 3×3 uniform kernel and then decimating by factor 3. A scaling family of transforms can be easily associated to this particular downsampling model, noticing that the LR images can be treated (up to a scaling factor $\beta_{0,M}=3$) to be composed of DC coefficients of some orthogonal 3×3 block transform. As we observed in [6], better results can be achieved if the image enlargement is performed gradually across multiple stages. Thus, the transform family $\{\mathcal{T}_m\}_{m=0}^{M}$ should consist of three 2-D block transforms with 1×1, 2×2, and 3×3 block sizes, which results in a progressive enlargement of 2 and 1.5 times, providing an overall enlargement of 3 times. As a particular family of transforms satisfying the above conditions, we choose the block DCT transforms.

3.2 Implementation details

For the presented experiments, the V-BM3D filter [2] inside the reconstruction algorithm is used without the collaborative Wiener filtering step. The filter's internal 3-D transform is a composition of a 2-D DCT transform applied to each block and of a 1-D Haar wavelet transform applied along the third dimension of the group. Other filter parameters are given in Table 1. Similar to [6], the block size is decreasing with the stages, within each stage, σ is decreased by $\Delta_{\sigma\ m}$ at every iteration. In terms both of smoothing and scale, this consistent with the "coarse-to-fine" approach analyzed in [5].

In order to avoid the influence of border distortions and provide more fair comparisons, in all experiments the PSNR and SSIM [15] values are calculated over the central part of the images, omitting a border of 15-pixel width.

3.3 Performance assessment and comparison with upsampling algorithm

To assess the potential of the algorithm, we performed an ideal experiment. The original ground-truth image is shifted horizontally by d^h pixels and vertically by d^v pixels, with $d^h, d^v = 0, 1, 2$, thus producing nine HR images. In particular, for these experiments, the first frame of the HR sequences from [13] is used as the ground-truth image. The nine HR images are then blurred and decimated as described above. In this way, the obtained LR sequence includes all pixels of a blurred HR image. Provided an ideal perfect registration, we can expect each reconstructed image to be at least not worse than the blurred HR one. During the SR reconstruction, block-matching search is performed in all 9 images.

As a comparison, we also performed upsampling of the same sequence, by restricting the algorithm to search only within the current frame. The reconstruction results are considered for the first frame only ($d^h, d^v = 0$) and are shown in Figures 3 and 4. The corresponding PSNR and SSIM values are given in Table 2. Besides the increased PSNR and SSIM values, one can see that the visual quality of the SR images is clearly superior than that of the upsampled ones: SR images are sharper, portray more fine details and do not exhibit ringing artefacts. Comparing to the blurred HR images, we can say that SR images are slightly sharper, but some minor details present in the blurred image have not been reconstructed properly in the SR images (e.g., the eyes and surrounding area, as shown in Figure 4). Overall, SR and blurred HR image qualities are close both visually and numerically.

	Nearest neighbor		HR blurred		Upsampled		Super-resolved	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Foreman	29.0	0.843	33.8	0.930	33.5	0.913	34.9	0.928
Suzie	30.3	0.812	33.7	0.891	32.8	0.863	33.6	0.883
Miss America	32.0	0.867	37.2	0.940	36.1	0.928	37.5	0.940

Table 2: PSNR (dB) and SSIM values of the super-resolved and upsampled images (see Section 3.3).

3.4 SR comparison

The method [13] provides a suitable ground for comparison for a number of reasons: firstly, it is based on the nonlocal means paradigm, thus exploiting the same kind of spatiotemporal redundancy as our algorithm; then, it reports high quality results; and finally, the HR, LR and SR reconstructed sequences are available online, making the experiments reproducible.

The observation model in [13] includes also presence of the additive noise. For this reason the LR images were contaminated with additive Gaussian noise with standard-deviation equal to 2. Since our observation model does not assume presence of the noise, we prefilter the noisy LR input sequence with the standard V-BM3D filter using default parameters [2]. The denoised LR sequences were then super-resolved by the algorithm described in this paper. The mean (over all 30 frames) PSNR and SSIM values for the reconstructed sequences are summarized in Table 3. The numerical results obtained by our algorithm are either comparable or superior than those of [13]. A visual comparison is provided in Figures 5 - 7. Although our SR images are not as sharp as those by the method [13], we can observe that in terms of artifacts, our estimates are indeed much cleaner, yet providing the same amount of reconstructed image details. Let us also note that while our implementation we use a temporal search window of 9 frames, in [13] all frames are involved in the reconstruction of each frame.

We remark that a direct comparison, between the numerical results of the experiments from this section and those from Section 3.3, is not possible because the input LR sequences for the two experiments are actually different.

4. DISCUSSION AND CONCLUSIONS

This papers extends results of our earlier paper [6] and continues the series of our works on iterative image reconstruction, started from [8], where a first attempt was made to apply iterative spatially adaptive filtering to the compressed sensing task. In fact, for $R = 1$, $M = 2$, the system (1) becomes equivalent to the recursive system used in [8], with the exception of a missing excitation-noise term. The presented experimental results show significant improvement of the super-resolved images over the upsampled ones. This means that filter is able to properly follow the local motion, and register and fuse subpixel-shifted image patches.

The results are promising and much work remains to be done. First of all, one would like to achieve better deblurring. Despite the algorithm implicitly relies on the knowledge of the blurring model (indeed, the blurring kernel is embedded as the basis function used for defining the coarse approximation coefficients), better visual quality can possibly be achieved, provided some additional deblurring as in [13]. The use of sharpening by collaborative alpha-rooting [4] within V-BM3D may also be considered. Future work shall also extend the observation model and the reconstruction algorithm to account for the presence of the noise in the LR observations.

REFERENCES

[1] Buades, A., B. Coll, and J. M. Morel, “A review of image denoising algorithms, with a new one”, *Multisc. Model. Simulat.*, vol. 4, no. 2, pp. 490-530, 2005.

[2] K. Dabov, A. Foi, and K. Egiazarian, “Video denoising by sparse 3D transform-domain collaborative filtering,” Proc. 15th European Signal Processing Conference, EUSIPCO 2007, Poznan, Poland, September 2007. `http://www.cs.tut.fi/~foi/GCF-BM3D`

[3] Dabov, K., A. Foi, V. Katkovnik, and K. Egiazarian, “Image denoising by sparse 3D transform-domain collaborative filtering”, *IEEE Trans. Image Process.*, vol. 16, no. 8, Aug. 2007.

[4] Dabov, K., A. Foi, V. Katkovnik, and K. Egiazarian, “Joint image sharpening and denoising by 3D transform-domain collaborative filtering,” *Proc. 2007 Int. TICSP Workshop Spectral Meth. Multirate Signal Process., SMMSP 2007*, Moscow, Russia, September 2007.

[5] Buades, A., B. Coll, J.M. Morel, C. Sbert, “Non local demosaicing”, Preprint CMLA Cachan, no. 2007-15 2007.

[6] Danielyan, A., A. Foi, V. Katkovnik, and K. Egiazarian, “Image upsampling via spatially adaptive block-matching filtering”, *Proc. of 16th European Signal Processing Conference, EUSIPCO2008,* Lausanne, Switzerland, Aug. 2008.

[7] Ebrahimi, M., E. R. Vrscay, “Multi-frame super-resolution with no explicit motion estimation”, *Proc. Int. Conf. on Image Process., Computer Vision, and Pattern Recognition, IPCV 2008*, Las Vegas, Nevada, USA, July. 2008.

[8] Egiazarian, K., A. Foi, and V. Katkovnik, “Compressed Sensing Image Reconstruction via Recursive Spatially Adaptive Filtering”, *Proc. IEEE Int. Conf. Image Process.*, *ICIP 2007*, San Antonio (TX), USA, pp. 549-552, Sep. 2007.

[9] Farsiu, S., D. Robinson, M. Elad, P. Milanfar, "Advances and challenges in super-resolution", *Int. J. of Imaging Systems and Technology*, vol. 14, no. 2, pp. 47-57, 2004.

[10] Lansel, S., D. Donoho, and T. Weissman, “DenoiseLab: a standard test set and evaluation method to compare denoising algorithms”, `http://www.stanford.edu/~slansel/DenoiseLab/`.

[11] Luong, H.Q., A. Ledda and W. Philips, “An Image Interpolation Scheme For Repetitive Structures”, *Int. Conf Image Analysis and Recognition, ICIAR 2006, Lecture Notes in Computer Science*, vol. 4141, pp. 104-115, Povoa De Varzim, Portugal, Sep. 2006.

[12] Mueller, N., Y. Lu, and M. N. Do, “Image interpolation using multiscale geometric representations”, *Proc. SPIE Electronic Imaging*, San Jose (CA), USA, 2007.

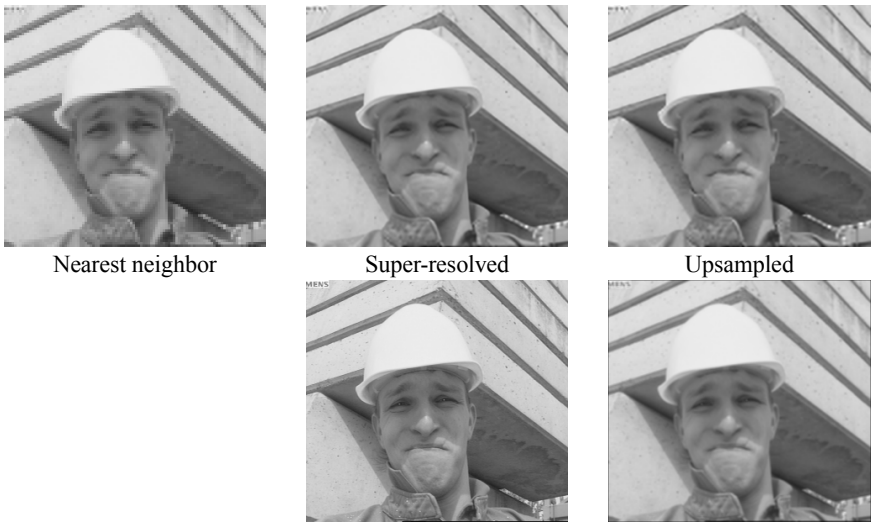
[13] Protter, M., M. Elad, H. Takeda, and P. Milanfar, “Generalizing the Non-Local-Means to Super-Resolution Reconstruction”, to appear in *IEEE Trans. Image Process.*, 2008. Test sequences available online at `http://www.cs.technion.ac.il/~matanpr/NLM-SR`

[14] Trimeche, M. , "Super-Resolution Image Reconstruction Using Non-Linear Filtering Techniques", *Ph.D. Thesis*, Tampere University of Technology, 2006.

[15] Wang, Z., A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, “Image quality assessment: From error visibility to structural similarity”, *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600-612, Apr. 2004.

	Nearest neighbor		Protter et al. [13]		Proposed	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Foreman	29.0	0.835	32.9	0.903	33.5	0.910
Suzie	30.3	0.800	33.0	0.880	33.0	0.862
Miss America	32.0	0.862	34.74	0.913	36.3	0.924

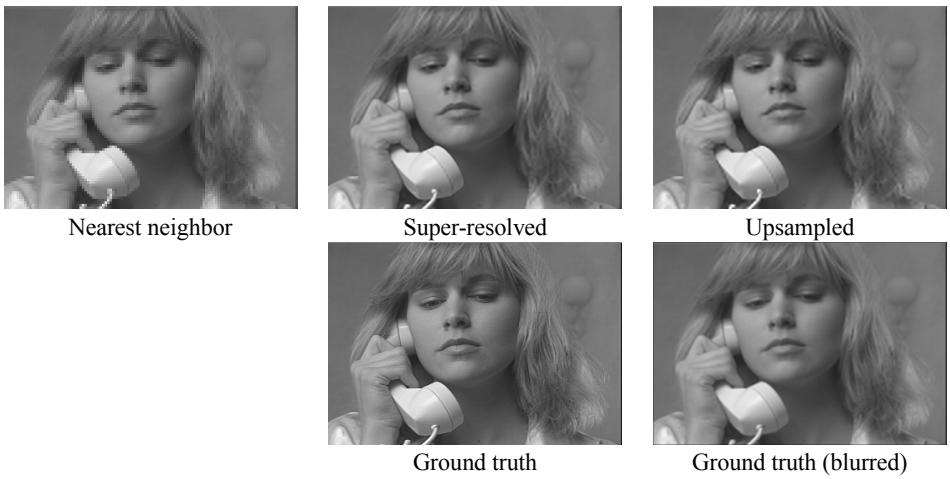
Table 3: Mean (over all frames) PSNR (dB) and SSIM values of the super-resolved video sequences (see Sec. 3.4).



Nearest neighbor | Super-resolved | Upsampled

Ground truth | Ground truth (blurred)

Figure 3: Comparison of the super-resolution reconstruction and upsampling results of the first frame of *Foreman* sequence (see Sec. 3.3).



Nearest neighbor | Super-resolved | Upsampled

Ground truth | Ground truth (blurred)

Figure 4: Comparison of the super-resolution reconstruction and upsampling results of first frame of *Suzie* sequence (see Sec. 3.3).

Figure 5: Results for the 1st, 8th, 13th, 18th, 23th, and 28th frame from the "Foreman" sequence. From left to right: low resolution image; original image (ground truth); Protter et al. [13] algorithm; result of the proposed algorithm.

Figure 6: Results for the 1st, 8th, 13th, 18th, 23th, and 28th frame from the "Suzie" sequence. From left to right: low resolution image; original image (ground truth); Protter et al. [13] algorithm; result of the proposed algorithm.

Figure 7: Results for the 1st, 8th, 13th, 18th, 23th, and 28th frame from the "Miss America" sequence. From left to right: low resolution image; original image (ground truth); Protter et al. [13] algorithm; result of the proposed algorithm.