

Motion Compensated Video Super Resolution

Sune Høgild Keller¹, François Lauze² and Mads Nielsen^{1,2}

¹ The Image Group, Department of Computer Science, University of Copenhagen,
Universitetsparken 1, DK-2100 Copenhagen, Denmark,
sunebio@diku.dk,

WWW home page: <http://www.image.diku.dk/sunebio>

² Nordic Bioscience A/S, Herlev Hovedgade 207, 2730 Herlev, Denmark

Abstract. In this paper we present a variational, spatiotemporal video super resolution scheme that produces not just one but n high resolution video frames from an n frame low resolution video sequence. We use a generic prior and the output is artifact-free, sharp and superior in quality to state of the art home cinema video processors. Unlike many other super resolution schemes, ours does not limit itself to just translational or affine motion, or to certain subclasses of image content to optimize the output quality. We present a link between image reconstruction and super resolution and formulate our super resolution constraint with arbitrary up-scaling factors in space from that.

1 Introduction

In this paper we introduce a novel approach to motion compensated video super resolution. In Fig. 1(e) it is shown how our method is able to recreate the ground truth (Fig. 1(a)) of a skewing bar sequence while spatial methods (Fig. 1(c) and Fig. 1(d)) currently used in consumer electronics produce blurred outputs.

Super resolution (SR) is an ill-posed problem: one wants to reconstruct a hypothetical higher resolution image (sequence) from an image sub-sampling at a lower resolution. The inevitable loss of high frequency information follows immediately from the Nyquist-Shannon sampling theorem and SR algorithms hence normally use multiple low resolution views, perform registration on them and then extract a high resolution image. Super resolution is a thoroughly investigated subject in image processing (see for instance [6]) whereas video super

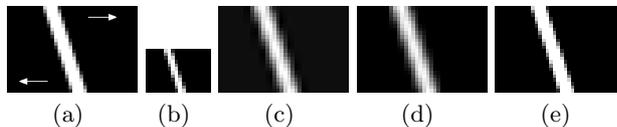


Fig. 1. Frame no. 3 of the five frame skew sequence: (a) ground truth (arrows show skew motion), (b) down sampled by 0.5x0.5. 2x2 SR on (b): (c) bicubic, (d) bilinear and (e) our algorithm.

resolution (VSR also known as multiframe super resolution) which by definition is the creation of n high resolution (HR) video frames from n low resolution (LR) video frames, is a lot less researched (see for instance [8]).

This paper is organized as follows. In Sec. 2 we present the low resolution image formation equation and the super resolution problem, and we discuss other work on super resolution and its relation to our work. In Sec. 3 we present our own motion compensated video super resolution scheme followed by a presentation of our experimental work in Sec. 4. Finally we conclude and look ahead in Sec. 5.

2 Overview of the Super Resolution Problem and Related Work

Following [12], a usual starting point for analysis of the SR problem is the continuous image formation equation that model the projection \mathcal{R} of an HR image into an LR one:

$$u_0(y) = \mathcal{R}(u)(y) + e(y) := \int B(x, y)u(x)dx + e(y) \quad (1)$$

where $u_0(y)$ is the low resolution irradiance field, $u(x)$ the high resolution one, $e(y)$ some noise and $B(x, y)$ is a blurring kernel. In general this kernel is assumed to be shift invariant and takes the form of a point spread function (PSF), $u_0 = B * u$. In the case of image sequences, x can be replaced by (x, t) where t is the time variable.

The general numerical formulation is $L = RH + E$ where $H \in \mathbb{R}^N$, $L \in \mathbb{R}^n$ with $n < N$, R a linear map $\mathbb{R}^N \rightarrow \mathbb{R}^n$ and $E \in \mathbb{R}^n$ a random vector. Inverting this equation is severely ill-posed and more information is needed to get a stable solution. First, more than one low resolution sample is used, and some prior on the distribution of HR images is introduced. A common type of prior imposes spatial or spatiotemporal regularity. Another one (that can be used simultaneously with the first one) relies on expected content of images. The general process can be described as follows: different LR images L_1, \dots, L_n are registered toward a common one, say L_1 , via sub-pixel displacement fields w_2, \dots, w_n such that $L_1(y) = L_i(w_i(y))$, hoping to get as close as possible to the ground truth H . The earliest approach, modelled in the frequency domain, was proposed by Tsai and Huang in [16]. A different family of methods also emerged directly in the spatial domain where a maximum likelihood estimator (ML) will produce HR output \hat{H} which minimizes the projection distances to all the sample LR images L_i . Adding priors will replace the ML estimator by a regularized ML or a maximum a posteriori (MAP) like estimator.

Borman and Stevenson have presented an extensive review of the different approaches in [2] and good references to recent work as well as another extensive bibliography can be found in [6]. In the work of Irani and Peleg [9] motion compensation (MC) is used to extract and register several frames from a given sequence and then create an HR image from them. Schultz and Stevenson [14]

also integrate motion compensation as well as prior smoothness constraints more permissive for edges than the Gaussian one. Here too, authors aim at reconstructing one HR frame from a video sequence input. Generalizations of these methods have been used for spatiotemporal super resolution of video sequences by Shechtman *et al.* in [15], from a set of low resolution sequences (multi-camera approach). The recent video super resolution algorithm by Farsiu *et al.* [8] is a typical example of simplified modelling of the optical flow that prevents it from producing high quality results like the one given in Fig. 1(e).

Super resolution has its limits as shown by Baker and Kanade in [1] and recently by Lin and Shum in [12]. In order to overcome these limitations, Baker and Kanade have proposed to add a generative, trained model part to the reconstruction, a choice not possible for us, as we want a generic VSR algorithm applicable to any kind of video content.

The starting point of our work is Bayesian inference. Using (MAP) we derive a variational formulation via an energy minimization whose optimization takes the form of a non linear reaction-diffusion both spatially, and temporally along the motion trajectories, which are themselves computed using variational methods in an integrated framework. This is a technique known from image sequence inpainting [7] and [10]. The use of a highly precise variational optical flow algorithms enables us to capture flows even more complex and detailed than the one in the skew sequence example in Fig. 1. Even the skew motion would be a source of problems to the widely used block matching motion estimators.

Lillholm *et al.* [11] have presented a method for the reconstruction of images from a finite set of features measured on the image (linear filter responses) under smoothness constraints in a variational framework, a problem very close to ours. We follow their approach to find the correct energy minimization by orthogonally projecting the suggested energy back onto the solution hyperplane as dictated by the super resolution constraint that follows from the modelling of the image formation process and the definition of R given above.

Our algorithm allows for arbitrary magnification factor but we will mainly look at two cases: a) The often presented 2x2 (magnification factor 2 in height and width) super resolution, and b) the upscaling from SD digital PAL video (576×720) to 720p HD video (720×1280). We can easily adapt our algorithm to do VSR to any other HD format (at least up to 1080×1920). In the two settings in focus we stay within the limits of SR and avoid the failure of to large magnification factors as shown in [1] and [12].

Our choice of focusing on video super resolution is motivated by real world needs. Today many high definition (HD) television formats exist along with the older standard definition (SD) formats like PAL, NTSC and SECAM. Screens for both broadcast and video viewing as well as computer screens have many different resolutions. People have large collections of SD DVDs and broadcasters have huge SD archives. The need for super resolution is there and as the display devices grow in size and viewing angle, the demand for high performance super resolution algorithms also grows. This also means that our main concern here is the output quality as judged by the human visual system (HVS). We focus on

producing outputs sharper than the input while possibly removing errors (e.g. flickering) present in the input, and with no new artifacts (the typical ringing and other, see e.g. results given in [1] and [8]).

So far we have discussed scientific work on super resolution focussing on getting as close a possible to the ground truth. In actual products for video processing like the Faroudja and DVDO video processors, and in build-in processors in high-end DVD-players and displays (plasmas, projectors etc.) focus is on visual quality, but the majority of the resources are typically spent on deinterlacing, noise filtering, correction of MPEG-2 errors and color corrections. Unfortunately bilinear interpolation as in Fig. 1(d) is the standard method used for video super resolution as it is cheap, easy to implement and does not create artifacts. The smoothing it produces is not severely unpleasing to the HVS, but given the trend of larger and better displays it will not suffice over time.

3 Motion Compensated Video Super Resolution

We wish to model the image sequence content and optical flow using probability distributions and thus start formulating our problem using Bayesian inference. This has been done by Lauze and Nielsen in [10] for simultaneous image sequence inpainting and motion recovery, leaving out the locus of missing data, which we do not need, we are left with

$$p(u, \mathbf{v}|u_0) \propto \underbrace{p(u_0|u)}_{P_0} \underbrace{p(u_s)}_{P_1} \underbrace{p(u_t|u_s, \mathbf{v})}_{P_2} \underbrace{p(\mathbf{v})}_{P_3}. \quad (2)$$

where we \mathbf{v} is the optical flow of the HR output sequence u , and u_0 is the input low resolution sequence. u_s and u_t are the spatial and temporal distribution of intensities respectively. On the left hand side we have the posterior distribution from which we wish to extract a MAP estimate. The right side terms are: P_0 , the image sequence likelihood, P_1 the spatial prior on image sequences, P_3 the prior on motion fields and P_2 a term that acts both as likelihood term for the motion field and as spatiotemporal prior on the image sequence. In the sequel we assume that P_0 comes from a noiseless image formation equation (1) (i.e. $e(y) = 0$). This means that the optimal pair (u, \mathbf{v}) minimizes the constrained problem

$$\begin{cases} E(u, \mathbf{v}) = E_1(u_s) + E_2(u_s, u_t, \mathbf{v}) + E_3(\mathbf{v}) \\ Ru = u_0 \end{cases} \quad (3)$$

where $E_i = -\log P_i$ and R the projection described in Sec. 2.

Applying Calculus of Variations, a minimizer for the above energy is characterized by the coupled system of equations, $\partial E(u, \mathbf{v})/\partial u = 0$ subject to $Ru = u_0$ and $\partial E(u, \mathbf{v})/\partial \mathbf{v} = 0$, in our case solved first for the flow and secondly for the intensities. To get a reasonable tradeoff between tractability and modelling accuracy we generally chose to use total variation term for E_i and for the optical flow related terms E_2 and E_3 we use the formulation of [3] as it is

also done in [10] for motion compensated inpainting. We use [3] because it has been shown to yield some of the most precise flows (see [4]).

For the actual video super resolution part of our problem, $\partial E(u, v)/\partial u = 0$, we neglect the gradient matching term of [3], as it introduces a 4th order term in the Euler-Lagrange equation and we do not believe it will improve our results in the current settings. We are left with minimizing

$$E(u) = \lambda_s \int_{\Omega} \psi(|\nabla u|^2) dx + \lambda_t \int_{\Omega} \psi(|u(x+v, t+1) - u(x, t)|^2) dx \quad (4)$$

using both backwards and forwards flows (warps) in the second term as in [10]. Ω is the spatiotemporal domain of u , u_t is the local temporal derivative of u and ∇ is the spatial gradient operator. Instead of using the $|\cdot|$ function which is non differentiable at the origin, we replace it by the approximation $\psi(s^2) = \sqrt{s^2 + \varepsilon^2}$, where ε is a small positive constant. the λ 's are positive constants weighing the two terms with respect to each other.

3.1 The Super Resolution Constraint

Many things can be said about modelling image acquisition, but one thing is certain: it is a process with many variables and it changes with cameras and telecines (film scanners). As we wish to operate on all types of video, we can only do very little in this modelling step.

The point spread function (PSF) describes how the incoming light is sampled over the pixel area. Gaussian distributions are – as always – a popular choice along with the model we have chosen to use, the uniform distribution. In 2x2 SR the two models yield identical results, but in cases like SD to 720p HD their outcomes will be different, although the regularization part of our SR algorithm might minimize the difference. Replacing the uniform distribution with a gaussian would complicate modelling and whether it would improve results is an open question.

A second key element in super resolution is the magnification factor and our model allows for arbitrary, independent magnification factors in both frame height and width (integer to integer of course). Thus we do not require any preservation of aspect ratio in our super resolution constraint, which is fully on purpose as it enables us to change between different pixel aspect ratios in video, e.g. from PAL widescreen 1:1.422 pixels to square 1:1 HD pixels.

The super resolution constraint simply is: *The average pixel intensity over an area of interest is kept constant.* This essentially means that the projection operator R is defined via an idealized point spread function B , whose support is an LR voxel (pixel + time spread), i.e. a moving average filter. This type of PSF is routinely used, often implicitly. We now describe the filters based on the super resolution constraint for both 1D and 2D signals.

1D Super Resolution Filter. The mapping $R_n^N : \mathbb{R}^n \rightarrow \mathbb{R}^N$ can be decomposed as a replication step $S : \mathbb{R}^n \rightarrow \mathbb{R}^{nN}$, where each source component is

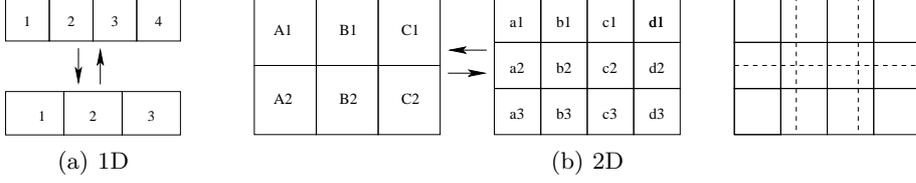


Fig. 2. Super resolution.

replicated N times, followed by an average step $T : \mathbb{R}^{nN} \rightarrow \mathbb{R}^N$, where consecutive blocks of n entries are replaced by their average. As an example shown in Fig. 2(a), we present the R_3^4 mapping and its action on a vector $u \in \mathbb{R}^3$:

$$R_3^4 = \frac{1}{3} \begin{pmatrix} 3 & 1 & 0 & 0 \\ 0 & 2 & 2 & 0 \\ 0 & 0 & 1 & 3 \end{pmatrix}, \quad R_3^4(u_1, u_2, u_3) = \frac{1}{3} (3u_1, u_1 + 2u_2, 2u_2 + u_3, 3u_3).$$

2D Super Resolution Filter. The 2D and higher dimensional filters of this form are separable, i.e. they can be obtained by cascading 1D filters along the appropriate dimensions. A 2D filter is simply obtained by defining the needed vertical and horizontal 1D filters, R_v and R_h , and taking the Kronecker product of them (vectorizing the image matrix lexicographically column by column) as given here for the $m \times n = 2 \times 3$ to $M \times N = 3 \times 4$ example in Fig. 2(b)

$$R = \frac{mn}{MN} R_h \otimes R_v. \quad (5)$$

For larger image sizes, e.g. going from 576×720 SD to a 720×1280 HD, our projection will map pixel blocks of size 4×9 to blocks of size 5×16 as given by the greatest common divisors in height and width.

3.2 Minimization Algorithm

The full algorithm we use for video super resolution is described here. First we calculate the forward and backward flow on the low resolution sequence using the method from [3], that is finding the optimal solution of $\partial E(u, \mathbf{v}) / \partial \mathbf{v} = 0$ by minimizing E_2 and E_3 in (3) using multiresolution. We then create the high resolution flow by applying R (from (5)) on the LR flow, and in the same way we initialize the HR image sequence, from which we then compute the final HR sequence. We need to fix some notations first: $\nabla_3 u$ will denote the spatiotemporal gradient of u , $V = (\mathbf{v}^t, 1)^T$ the spatiotemporal velocity field so that $u(x+v, t+1) - u(x, t) \approx \nabla u \cdot \mathbf{v} + u_t = \mathbf{V}^T \nabla_3 u$, with discretization suggested by the former expression (see [10] for details). We define $A(u) = 2\psi'(|\nabla u|^2)$ and $B(u) = 2\psi'(|V^T \nabla_3 u|^2)$. The gradient of the energy in (4) is

$$G(A(u), B(u)) := \frac{\partial E}{\partial u} = -\lambda_s \text{div}_2(A(u) \nabla u) - \lambda_t \text{div}_3(B(u)(V^T \nabla_3 u)V) \quad (6)$$

where div_2 and div_3 are the 2D and 3D divergence operators respectively.

Discretization is performed using schemes from [3] and [10]. In order to cope with nonlinearity in the gradient we use a fixed point approach. We only update the values $A(u)$ and $B(u)$ in each of a number of outer fixed point iterations in order to have a linear approximation of the discretized gradient coming from (6). For each outer iteration, we run a number of inner iterations on the now linearized system using a Gauss-Seidel relaxation scheme with the super resolution constraint incorporated by projection, a projection modulated by a positive weight, α , in order to respect intensity bounds. Denoting the null space of R by T and by P_T the orthogonal projection onto T , we sketch the algorithm here:

- Let u^0 be the initial guess for the HR sequence
- for $j = 0$ until $J - 1$
 1. Let $A^j := A(u^j)$, $B^j := B(u^j)$, $u^{j,0} := u^j$
 2. for $k = 0$ until $K - 1$
 - (a) from $u^{j,k}$, compute $\bar{u}^{j,k+1}$ by one Gauss Seidel sweep on $G(A^j, B^j)$
 - (b) form the update vector $\bar{\delta}^{j,k} := \bar{u}^{j,k+1} - u^{j,k}$
 - (c) project it on T , $\delta^{j,k} := P_T(\bar{\delta}^{j,k})$
 - (d) update current estimate $u^{j,k+1} := u^{j,k} + \alpha\delta^{j,k}$
 3. $u^{j+1} := u^{j,K}$
- output the HR sequence u^J .

Form classical linear algebra, the orthogonal projection is

$$P_T \equiv Id - R^T(RR^T)^{-1}R. \quad (7)$$

As mentioned in Sec. 3.1 the projection allows processing in blocks, so for any $M \times N$ block step 2(c-d) above is actually

$$u_l^{j,k+1} = u_l^{j,k} - \alpha \left(\bar{\delta}_l^{j,k} + \sum_{i=1}^{MN} P^T(l, i) \bar{\delta}_i^{j,k} \right), \quad l = 1, \dots, MN \quad (8)$$

The diffusion part of our algorithm follow the minimum-maximum principle ([13]) and keeps intensity values within bounds, but we need the weight α in (8) to ensure that the projection does not create values out of bounds. If a value in a block is detected out of bounds, α is recursively lowered from 1 to 0 in steps of 0.1, and all values in the block recalculated with the new α , stopping when all values in the block are within bounds. This can potentially stop the evolution of the sequence, but it was found not to do so in testing. The problem was limited to very few regions around extremely high contrast edges and thus the computational overhead was also neglectable.

4 Experiments

The goal of our research is to improve the perceived quality by human viewers of super resolved video and we will therefore evaluate the visual quality of the

output sequences. As mentioned in Sec. 2 we evaluate results at two different magnifications: The industry oriented 576p SD (576×720) to 720p HD (720×1280) and the classic 2x2 VSR. We work on the luminance channel (8 bit, [0-255]) of 5-20 frame sequences of standard DVD Video material telecined from 35mm film. First though, we wish to emphasize the importance of the super resolution constraint by a small experiment.

4.1 The Importance of the Super Resolution Constraint

There is no doubt that the super resolution constraint adds complexity to an already complex regularization scheme, so why not leave it out? An example of its importance as data attachment term is given here. In Fig. 3(a) we have one of four identical LR frames, a 4×9 matrix filled line by line with the values 1 to 36, and in Fig. 3(b) the corresponding HR frame, a 6×12 matrix initialized using the example matrix in (5).³ In Fig. 3(c) we see how pure regularization without the SR constraint destroys the image content, whereas it is preserved in Fig. 3(d) with the SR constraint on. The destruction process is significantly slowed down if the temporal regularization is given high weight and the flow is accurate.

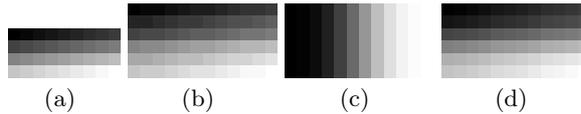


Fig. 3. The importance of the SR constraint.

4.2 Parameter Settings

We use the algorithm as given in Sec. 3.2 and give the used settings here.

After extensive testing we have found what we believe to be the best parameter setting of the optical flow calculations when used on real world SD resolution video data. We use 5 outer and 20 inner loops on a 100 level pyramid with a scale factor of 1.04 between the levels. The weight of the gradient constancy assumption is set to 200 and the weight on the smoothness term is 70 (see [3]). As there is no ground truth flow to compare to, we have no way of guaranteeing that this setting is optimal, but we believe it to be from extensive testing.

For the actual super resolution part, finding the optimal output sequence, we tested a wide range of settings for its four free parameters. First the number of inner and outer iterations as described in Sec. 3.2. We found that 5×10 (5 outer, 10 inner) iterations gives optimal results, but that even as little as 5×5 and 3×5 iterations yield practically the same quality. Even with the results

³ The HR initialization looks so fine because there is no sharp edges in this example.

given here using 5×10 iterations, this part is still much faster than the flow calculation part of our algorithm. For the spatial and temporal regularization weights λ_s and λ_t in (4) and (6) we have chosen four settings that we generally used in our experiments. They are: Temporal favoring S_t ($\lambda_s:\lambda_t = 1:5$), S_{st} with equal weight (1:1), spatial favoring S_s (5:1) and pure spatial S_{ss} (1:0).

4.3 Results

The output quality for our two test cases, 2x2 video super resolution and SD to 720p HD VSR, are very similar on all tested sequences. Of the four test settings, spatial favoring S_s gave the best results. First off one would think the S_t setting favoring temporal regularization would be best, but the simple flow initialization we use causes some blockiness in the initialization to be preserved as (false) sharp details. The equal weighing S_{st} does not have this problem and its output is often indistinguishable from the output of S_s . Switching off the temporal regularization completely in S_{ss} gives a small loss of overall sharpness in all tests and the outputs are far sharper than the bilinear outputs.

Besides the problems with S_t , which will be discussed later in this section, we get sharp outputs for all 20 sequences in test. The outputs have no artifacts from the upscaling process and subsampling artifacts present in the LR inputs are reduced to some degree. All the S_s (and most S_{st}) results are perceived quite a lot sharper than their LR inputs. Our evaluations was of course done on video but to give an impression of the quality obtained, Fig. 4 and Fig. 5 show some inputs and results produced with the S_s setting. Due limited space S_{st} and S_{ss} will not be show – they are practically the same as S_s – and S_{st} neither – they are not unsharp but tend towards the LR inputs in blockiness. We show the HR outputs and the LR input at the (relative) same size to clearly show the enhancement resulting from the upscaling process. This also illustrate why high resolution is essential to large displays used with large viewing angles.

In 4(a) we see part of one frame of the LR input test sequence **Front Grill**. The initialized 2x2 HR input to our algorithm looks just like the LR. The result of bilinear 2x2 SR shown in 4(b) is clearly not as sharp as the 2x2 VSR output of our algorithm in 4(c). The bicubic 2x2 SR shown in 4(d) is not as sharp as our output and white details seem greyish, but it is a bit better than bilinear. The example given here is the best performance observed from bicubic, that generally lies closer to bilinear than our method in quality. Figure 4(f) is the 720p HD VSR output with correct aspect ratio as created by our algorithm from the anamorph LR input in 4(e). A lot of flickering is removed in the far away part of the wire fence (to see this video can be found at the authors home page). Note the good continuation of the wires in the lower left corner behind the policeman, and the contour of, and text on, the helmet. The cracks between the boards in the boardwalk in the 720p output in 5(b) look natural compared to how they look in the LR input in 5(a). In 5(d) and 5(e) we see how the bullets are depicted much better than in the input 5(c), and in 5(g) it is clearly seen how the straw hat suddenly get it 'true' structure represented when compared to the LR input in 5(f). In the **straw hat** and **wire fence** sequences, and seen very

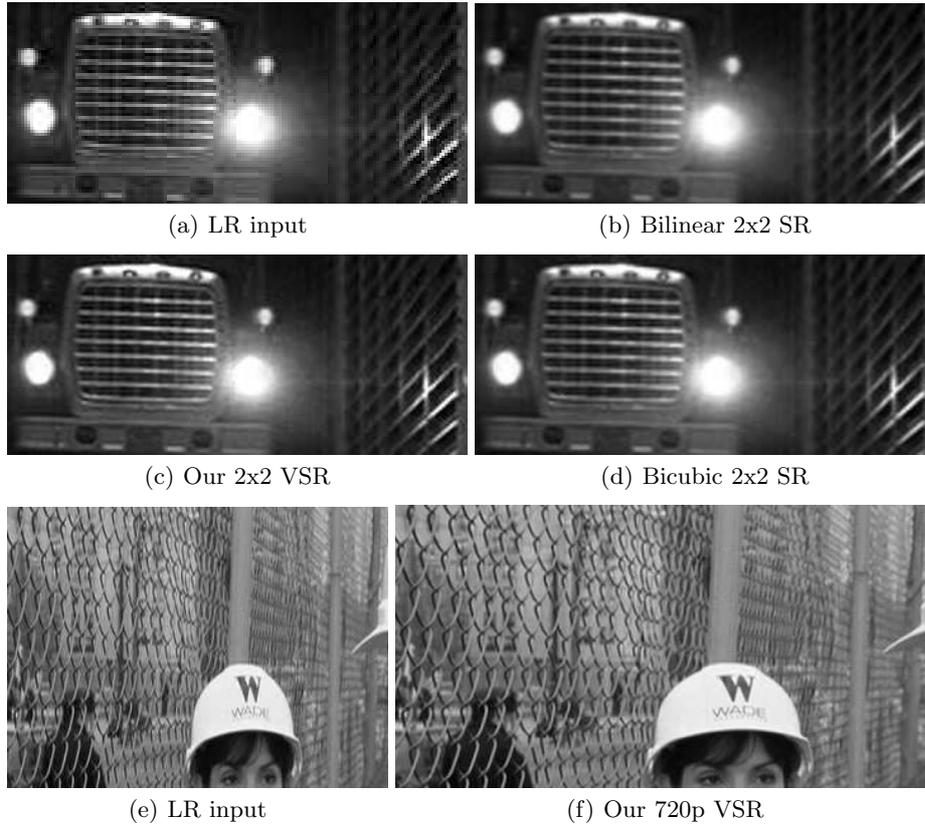


Fig. 4. Inputs and results I. Best viewed electronically (zoomed to 125% or more).

clearly in another test sequence (a helicopter flyby shot of Manhattan found as video on the authors home page), a major enhancement from the LR inputs to our HR outputs is flicker reduction, which cannot be seen in the figures.

We know the S_t setting favoring temporal input has the potential to improve on the current results, especially it can increase the sharpness. This can be obtained by replacing the simple HR flow initialization we use now with a simultaneous upscaling of the flow and intensities. The S_t setting already shows a tendency to push sharp details, although with the current simple flow it is easily overdone resulting in blockiness. The rather sharp S_s and S_{st} outputs show that we already gain from the temporal regularization as even our current flow will convey information of subsampled details when they are part of uniformly moving regions. The still relative sharp S_{ss} outputs show that edges at motion boundaries – where flow accuracy suffers the most from the simple upscaling – are handled reasonably well by the spatial regularization.

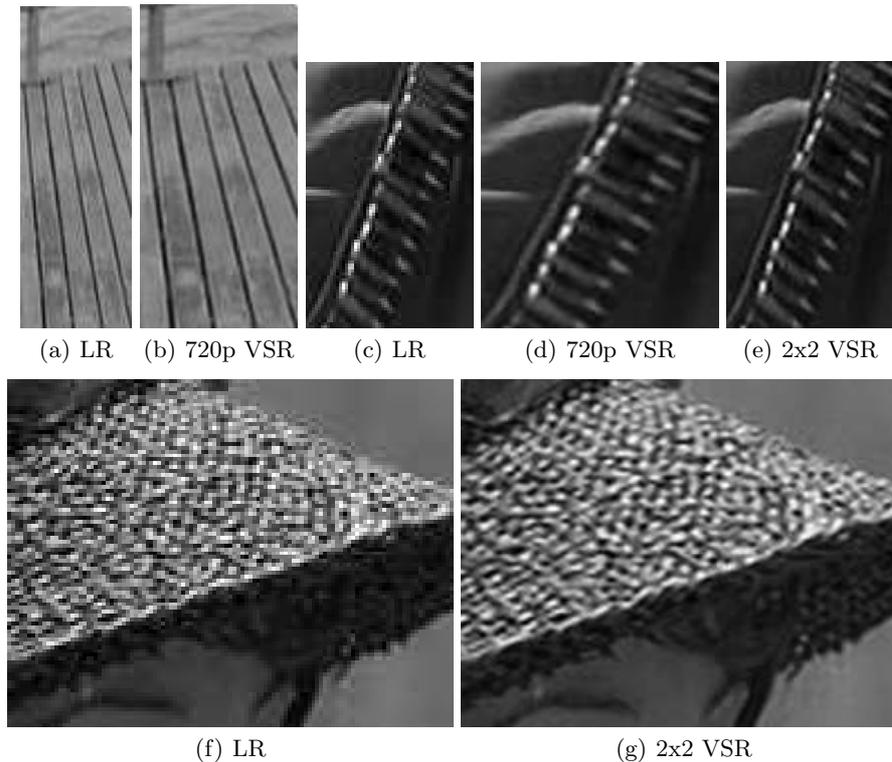


Fig. 5. Inputs and results II. Best viewed electronically (zoomed to 125% or more).

5 Conclusion and Future Work

We have developed and tested a novel motion compensated video super resolution scheme with arbitrary upscale factors in space, and a super resolution constraint based on image reconstruction. In comparison to many other video super resolution algorithms we allow for any kind of flow, and our modelling is based on a generic prior as we aim at general applicability in video processing pipelines. From n LR frames we produce n sharp and artifact-free HR frames with reduction of flicker, clearly outperforming the (industrial) standard bilinear interpolation.

The major drawback of our algorithm in its current form is that it does not employ the full potential of our framework. We are currently working on an algorithm simultaneously updating the flow and intensities and expect to give even sharper outputs, possibly with an additional reduction of flicker. We also intend to reformulate our framework to enable temporal super resolution as well as adapting and applying it to 3D and 4D medical data. In terms of running time we are a factor of 500-1000 from realtime in the intensity part of our algorithm using unoptimized code. The flow part of the algorithm is much slower, but real

time variational optical flow using a standard PC has been presented in [5]. Using modern, relatively cheap, HD capable field programmable gate arrays (FPGA) a real time implementation seems close at hand.

References

1. S. Baker and T. Kanade. Limits on Super-Resolution and How to Break Them. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 24(9):1167–1183, 2002.
2. S. Borman and R. Stevenson. Spatial Resolution Enhancement of Low-Resolution Image Sequences: A Comprehensive Review with Directions for Future Research. Technical report, Laboratory for Image and Sequence Analysis (LISA), University of Notre Dame, July 1998.
3. T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High Accuracy Optical Flow Estimation Based on a Theory for Warping. In T. Pajdla and J. Matas, editors, *Proceedings of the 8th European Conference on Computer Vision*, volume 4, pages 25–36, Prague, Czech Republic, 2004. Springer-Verlag.
4. A. Bruhn, J. Weickert, and C. Schnörr. Lucas/Kanade Meets Horn/Schunck: Combining Local and Global Optic Flow Methods. *International Journal of Computer Vision*, 61(3):211–231, 2005.
5. A. Bruhn, J. Weickert, C. Feddern, T. Kohlberger, and C. Schnörr. Variational Optic Flow Computation in Real-Time. *IEEE Trans. on Image Processing*, 14(5):608–615, 2005.
6. S. Chaudhuri, editor. *Super-Resolution Imaging*. The International Series in Engineering and Computer Science. Springer, 2001.
7. J.P. Cocquerez, L. Chanas, and J. Blanc-Talon. Simultaneous Inpainting and Motion Estimation of Highly Degraded Video-Sequences. In *Scandinavian Conference on Image Analysis*, pages 523–530. Springer-Verlag, 2003. LNCS, 2749.
8. S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar. Fast and Robust Multiframe Super Resolution. *IEEE Trans. on Image Processing*, 13(10):1327–1344, 2004.
9. M. Irani and S. Peleg. Motion Analysis for Image Enhancement: Resolution, Occlusion, and Transparency. *Journal on Visual Communications and Image Representation*, 4(4):324–335, 1993.
10. F. Lauze and M. Nielsen. A Variational Algorithm for Motion Compensated Inpainting. In S. Barman A. Hoppe and T. Ellis, editors, *British Machine Vision Conference*, volume 2, pages 777–787. BMVA, 2004.
11. M. Lillholm, M. Nielsen, and L. Griffin. Feature-Based Image Analysis. *IJCV*, 52(2/3):73–95, 2003.
12. Z. Lin and H-Y Shum. Fundamental Limits of Reconstruction-Based Superresolution Algorithms under Local Translation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 26(1):83–97, 2004.
13. P. Perona and J. Malik. Scale-Space and Edge Detection Using Anisotropic Diffusion. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 12(7):629–639, July 1990.
14. R.R. Schultz and R.L. Stevenson. Extraction of High Resolution Frames from Video Sequences. *IEEE Trans. on Image Processing*, 5(6):996–1011, 1996.
15. E. Shechtman, Y. Caspi, and M. Irani. Space-Time Super-Resolution. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 27(4):531–545, 2005.
16. R. Tsai and T. Huang. Multiframe Image Restoration and Registration. In *Advances in Computer Vision and Image Processing*, volume 1, 1984.