Variational Reconstruction and Restoration for Video Super-Resolution

Jordi Salvador[†] Daniel Rivero[†‡] Axel Kochale[†] Javier Ruiz-Hidalgo[‡]

[†]*Technicolor R&I Hannover* [‡]*Universitat Politècnica de Catalunya*

Abstract

This paper presents a variational framework for obtaining super-resolved video-sequences, based on the observation that reconstruction-based Super-Resolution (SR) algorithms are limited by two factors: registration exactitude and Point Spread Function (PSF) estimation accuracy. To minimize the impact of the first limiting factor, a small-scale linear inpainting algorithm is proposed to provide smooth SR video frames. To improve the second limiting factor, a fast PSF local estimation and total variation-based denoising is proposed. Experimental results reflect the improvements provided by the proposed method when compared to classic SR approaches.

1 Introduction

Image and video Super-Resolution (SR) has been a very active research topic during the last three decades [12, 5, 1]. The purpose of the vast amount of techniques contributing to this field is indeed very attractive: a high-quality capture of a scene can be *computed* by applying signal processing algorithms on a number of severly corrupted captures (*e.g.* with low-resolution, aliased, noisy, defocused), based on exploiting accurate image formation models.

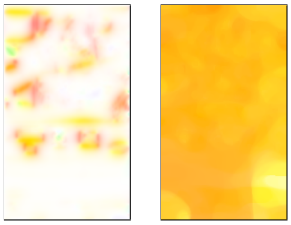
The above technqiues are more precisely classified as reconstruction-based SR, as opposed to example-based SR [2, 9], where the image (or video) enhancement is obtained by exploiting prior knowledge on the capture process or the scene structure. The main limiting factors of reconstruction-based SR are errors due to inaccurate image registration and errors in the estimation of the blur kernel or *Point Spread Function* (PSF). [6] shows how the problem of blind deconvolution (deblurring with unknown kernel) can benefit from exploiting the asimetry (between the number of unknowns and a reasonable number of available pixels) when attempting to estimate the blur kernel (PSF) separately from deblurring the image.

Interest in reconstruction-based SR is still high due to the fact that, in presence of notable aliasing and noise, robust reconstruction-based techniques [1] can be better suited. Example-based SR does not, in general, reconstruct actual fine detail but rather synthesizes plausible one. Furthermore, new achievements can be integrated in hybrid approaches, extending *e.g.* [3].

The reconstruction-based SR technique proposed in this paper is based on an image formation model given by the well-known expression $\mathbf{y}_i = \mathbf{D}_i\mathbf{B}_i\mathbf{F}_i\mathbf{x} + \mathbf{n}_i$, where $\mathbf{y}_i$ are low-resolution captures, $\mathbf{D}_i$ is a decimation operator, $\mathbf{B}_i$ models the image blur, $\mathbf{F}_i$ is the sub-pixel displacement, $\mathbf{x}$ is the desired high-resolution image and $\mathbf{n}_i$ is additive noise. Then, we subdivide the video SR problem in (1) estimating the sub-pixel shifts between video frames (registration); (2) warping neighbor frames following the registration results and inpainting empty gaps in the resulting image (reconstruction); (3) estimating the PSF; (4) applying the previous PSF for deblurring; and (5) denoising based on total variation (TV) [8]. This is done in order to apply a suitable conditioning to each stage, rather than assuming a single arbitrary regularizer suits the conditions of all subproblems. This paper is organized as follows; Section 2 explains steps (1) registration and (2) reconstruction of the proposed method. Section 3 focuses on steps (3) PSF estimation, (4) deblurring and (5) denoising. Section 4 shows some experimental results. Finally, conclusions are drawn in Section 5.

Table 1. Registration error (average endpoint, in pixels) of different optical flow approaches with global and local motion

Sequence	Type	PLK	HS	C+NL
camman	rigid	0.0032	0.4658	0.0018
teddy	non-rigid	3.80	1.51	0.49



HS C+NL

Figure 1. Detail of the color-encoded estimated optical flow in a scene with complex motion. C+NL contains a clearly reduced amount of outliers

2 Registration and Reconstruction

The goal of these two stages is to combine the available information in a set of N temporally close video frames in order to generate the super-resolved version of each frame in the video sequence. We will first obtain the sub-pixel misalignment of each pixel in each of the involved frames (by using a contemporary optical-flow estimation scheme) and then warp each frame using this registration in order to fill-in a high-resolution (HR) grid. Finally, missing data in the HR grid is obtained by means of a small-scale inpainting approach.

2.1 Registration

In our tests, we have compared the performance of *Classic Non-Local Optical Flow* (C+NL) [10] to that of Horn-Schunck (HS) [4] (dense motion) and also to Pyramidal Lucas-Kanade (PLK) [7].

In Table 1, we can observe how, even for rigid camera motion (*camman*), C+NL outperforms the classical approaches. When motion becomes more complex, C+NL also clearly outperforms HS with a greatly reduced amount of outliers, as shown in Fig. 1, while PLK just provides an estimate of the average scene motion over the whole image. Thus, our choice for providing the registration in video sequences with complex motion is C+NL, which basically consists in a multi-scale variation of the original Horn-Schunck method with a novel non-linear regularizer (extending the functionality of a median filter) instead of the original smooth one.

2.2 Reconstruction

Our proposed reconstruction scheme is based on, first, warping the neighbor frames with the registration results and, then, inpainting empty pixels in the resulting high-resolution image.

Warping. Neighbor frames are forward-warped into an initially empty HR grid with a number of pixels equal to the desired size of the SR image. The contribution to each of the four closest pixels is modulated by bilinear weights regarding their proximity to the warped and scaled position (*e.g.* $(x - \lfloor x \rfloor)(y - \lfloor y \rfloor)$ for the bottom-right pixel). After accumulating the contributions from all frames, the resulting HR grid is normalized by the accumulated weights and a mask M is obtained, indicating where a pixel has not received any contribution.

Small scale inpainting. Our small-scale inpainter is simply formulated as $\hat{\mathbf{x}}_r := \min_{\mathbf{x}_r} R(\mathbf{x}_r)$, with the only varying elements of $\mathbf{x}_r$ (the vectorized reconstructed image) being those where the mask M obtained in the previous stage

equals one. $R()$ is a regularizer regarding a-priori assumptions about the image structure. After experimenting with TV and Tikhonov regularizers, we assessed the latter was better conditioned for this task. The method employed to obtain the result is gradient descent $\mathbf{x}_r^{t+1} := \mathbf{x}_r^t - \mu M(\Delta \mathbf{x}_r^t)$ with constant update step μ. Δ is the Laplacian operator. With this approach for small-scale inpanting, we implicitly assume noiseless warped pixels (noise is treated at a later stage with a dedicated suitable regularizer) and we simplify the variational formulation by removing the data term.

3 Restoration

The outcome of the method described in the previous section is a smooth image containing a richer amount of detail than any of the single frames employed for obtaining it, but with the problem of potentially containing a substantial amount of blur and additive noise. Therefore, with the techniques presented in this section we aim at restoring a high-quality SR image by deblurring and denoising.

3.1 PSF estimation

Prior to attempting to deblur the image, we need a suitable estimate of the PSF (or blur kernel). Please note that the estimation also benefits from the choice of a linear regularizer in the previous stage, for a linear process can be properly described by means of a convolution kernel (this would not hold with *e.g.* TV regularization, which would require more complex deblurring techniques [11]).

In order to estimate the PSF, we follow an approach similar to that described in [6], but with practical modifications in two important parts. First, an estimate of the sharp image is obtained with a simpler strategy (which in our experiments appears to be robuster when processing images with very low resolution), consisting in: (1) *Canny* edge detection; (2) edge-transversal local binarization (to the maximum and minimum pixel values at each side of the edge) and (3) elimination of the linear transition between the extrema. This responds to the fact that, in heavily downsampled images, a single pixel already covers a large area around a contour point.

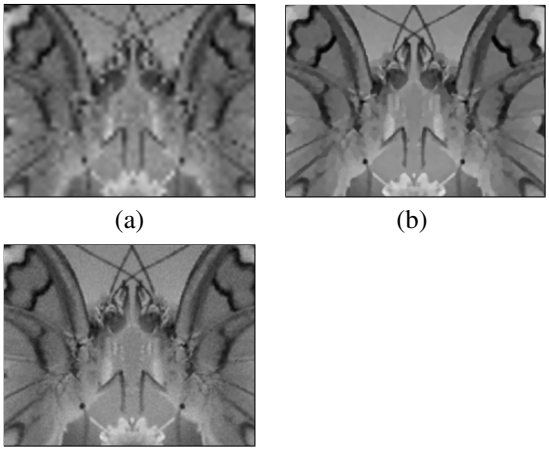
Once the sharp image x_d has been estimated (around edges) by this method, it is used for estimating the PSF in a direct formulation (possible by resorting to a Tikhonov –linear– regularizer). This contrasts with the gradient descent approach chosen in [6] and provides a closed-form solution. For the sake of readability, in the following formulas we avoid referring to indicator functions (masks) for estimated sharp pixels in the neighborhood of contour points.

Let h be the PSF to be estimated, x_r the reconstructed SR image and $R(h)$ a Tikhonov regularizer. We look for $\hat{h} := \min_h \|x_r - x_d * h\|_2^2 + \lambda_h R(h)$. We note that the convolution $x_d * h$ is identical to $h * x_d$, which leads to a closed-form solution when we formulate it as a matrix-vector product. It can be shown that $\hat{\mathbf{h}} = (\mathbf{X}_d^T \mathbf{X}_d + \lambda_h \Delta)^{-1} \mathbf{X}_d^T \mathbf{x}_r$, where $\hat{\mathbf{h}}$ is the vectorized estimated PSF, $\mathbf{X}_d$ is the convolution matrix form of the estimated sharp image (containing as many warps of the image –columns– as elements has h) and Δ is a linear operator obtained from the matrix form of the forward-difference derivative operators ($\Delta := D_u^T D_u + D_v^T D_v$, with $D_u \mathbf{x} = \mathrm{vec}(\nabla_u x)$ being the vectorized horizontal gradient and $D_v \mathbf{x}$ the corresponding vectorized vertical gradient).

3.2 Deblurring

Once the shape of the PSF is known, we marginalize on the other unknown of the blur model $\hat{x}_d := \min_{x_d} \|x_r - x_d * h$
with λ_d a small regularization factor included for numerical stability and $R(x_d)$ a Tikhonov regularizer. By using this linear regularizer (which has minimal influence on the result) we can once again obtain a closed-form solution, with better accuracy than gradient methods. Leaving apart considerations about the contour conditions, the solution to this problem is now $\hat{\mathbf{x}}_d = (H^T H + \lambda_d \Delta)^{-1} H^T \mathbf{x}_r$, where $\hat{\mathbf{x}}_d$ is the vectorized deblurred image and H is the convolution matrix form of the PSF h estimated in the previous stage.

The whole process for PSF estimation and deblurring benefits from being applied to small windows within the whole image. Indeed, the effect of the camera lens is not uniform across the image, and different types of motion blur can be present in the image. The practical limitation in this case is the sparser availability of contour data for correctly estimating the PSF.



(a) (b)

(c)

Figure 2. SR with rigid motion. SR image using (a) bicubic interpolation; (b) a robust SR method (Farsiu) [1]; and (c) our approach (better viewed when zoomed in)

3.3 Denoising

In contrast to deblurring, the effect of the regularizer has a clear impact on the obtained results when denoising. TV is known to be a powerful regularizer for this task. The resulting L_2-TV formulation is $\hat{x}_f := \min_{x_f} \|x_d - x_f\|_2^2 + \lambda_f R($ with $R(x_f)$ the anisotropic TV regularizer.

The practical consequence of using this type of regularizer is the impossibility of obtaining a closed-form solution. Here we propose using a gradient-descent approach like the one used for small-scale inpainting, with the differences of the shape of the regularizer and the inclusion of a data similarity term.

The iteration, in vectorial formulation, is defined as $\mathbf{x}_f^{t+1} := \mathbf{x}_f^t - \mu^t \left(2(\mathbf{x}_d - \mathbf{x}_f^t) + \lambda_f R'(\mathbf{x}_f^t)\right)$, with $R'(\mathbf{x}^t) := D_u^T$ and D_u and D_v defined as in PSF estimation. Please note that, in this case, we consider an exponentially decaying step μ^t. This is done to obtain faster, yet inaccurate initial variations with a large step when we are far from the optimal solution and slower but accurate final variations, when we aim at stabilizing around the optimal point.

4 Experimental Results

The presented algorithms have been implemented in MATLAB. Even though any real magnification factor can be applied with our method, we use $\times 2$ in our examples, in order to keep it comparable with other approaches. Concretely, we compare our method to bicubic interpolation of each single frame (*bicubic*) and the reference *Fast and Robust* SR method from [1] (*Farsiu*).

Using ground-truth data (synthetic sub-pixel-shifted image sequences), as in Fig. 2, we subjectively and quantitatively assess that our approach provides better accuracy than both *bicubic* and *Farsiu*. Indeed, averaging the results in 6 examples including the cited one, the SSIM index and Y-PSNR are: 0.45 and 18.09 dB for *bicubic*, 0.56 and 18.41 dB for *Farsiu* and 0.74 and 22.47 dB for our approach. This experiment, with fairly simple registration, allows us to assess the quality of the reconstruction and restoration algorithms.

Using real-data (low-quality video), we can also see how our method is capable of retrieving better results than the other references. This is due to the superior performance of the registration stage and also to the improvements due to our proposed reconstruction, denoising and deblurring algorithms. An example using a real video sequence with complex motion is shown in Fig. 3. The chosen registration technique allows to correctly measure the subpixel misalignment even when some objects move freely in the video sequence.

(a) (b) (c)

Figure 3. Detail of SR results with (a) bicubic interpolation, (b) Farsiu and (c) our proposed method for a real video-sequence (better viewed when zoomed in)

5 Conclusions

We have presented a modular variational approach for solving the subtasks of reconstruction and restoration in reconstruction-based SR. This, combined with contemporary state-of-the-art optical-flow estimation, allows us to obtain deblurred and denoised SR frames in video sequences showing spatial aliasing, even in presence of complex scene motion.

The main advantage of proceeding in a modular manner is that we are able to introduce a suitable regularizer for each substage. This allows us to obtain closed-form solutions to some of the tasks when employing the right formulation (*e.g.* L_2 data fidelity term with Tikhonov regularization) and consider costlier non-linear processes when required (*e.g.* in denoising).

For our future work, we plan to extend the effective magnification by using our results as input in single-image SR, with the latter benefiting from reduced aliasing and noise and the recovered high-frequency detail.

References

[1] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar. Fast and robust super-resolution. In *Proc. IEEE Int. Conf. on Image Processing*, volume 3, pages 291–294, 2003.

[2] W. Freeman, T. Jones, and E. Pasztor. Example-based super-resolution. *IEEE Computer Graphics and Applications*, 22(2):56–65, March/April 2002.

[3] D. Glasner, S. Bagon, and M. Irani. Super-resolution from a single image. In *Proc. IEEE Int. Conf. on Computer Vision*, pages 349–356, 2009.

[4] B. K. P. Horn and B. G. Schunck. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.

[5] M. Irani and S. Peleg. Super resolution from image sequences. In *Proc. Int. Conf. on Pattern Recognition*, volume 2, pages 115–120, 1990.

[6] N. Joshi, R. Szeliski, and D. Kriegman. PSF estimation using sharp edge prediction. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pages 1–8, 2008.

[7] B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proc. Int. Joint Conf. on Artificial intelligence*, pages 674–679, 1981.

[8] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Phys. D*, 60(1-4):259–268, Nov. 1992.

[9] O. Shahar, A. Faktor, and M. Irani. Space-time super-resolution from a single video. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pages 3353–3360, 2011.

[10] D. Sun, S. Roth, and M. Black. Secrets of optical flow estimation and their principles. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pages 2432–2439, 2010.

[11] H. Takeda, S. Farsiu, and P. Milanfar. Deblurring using regularized locally adaptive kernel regression. *IEEE Trans. on Image Processing*, 17(4):550–563, Apr. 2008.

[12] R. Y. Tsai and T. S. Huang. Multiframe image restoration and registration. *Advances in Computer Vision and Image Processing*, 1984.