An Alternation-based Approach to Super-Resolution for Deformable Surfaces

Dawei Liu   Adrien Bartoli   Julien Peyras

LASMEA - CNRS and Université Blaise Pascal - Clermont-Ferrand

Abstract

Conventional Super-Resolution from image sequences fail when dealing with deformable surfaces. This is because only planar homography is used to register the low resolution sequence. This paper makes an effort to break this limitation. For this we use Free-Form Deformation for deformable surfaces registration. We also use an alternation-based approach to improve the registration and the Super-Resolution estimate.

Keywords

Super-Resolution, Deformable Surface, Free-Form Deformation.

1 Introduction

Due to the limitations of optical imaging systems we often obtain images with insufficient resolution. Typical examples include video surveillance images or cell phone camera images. Poor resolution degrades the aesthetic quality of images for media publication or hampers higher vision tasks such as object recognition or localization. Multi-frame Super-Resolution (SR) is a technique that takes advantage of the redundancy information among a sequence of Low Resolution (LR) images and tries to reconstruct a higher resolution image. Numerous algorithms have been proposed since [1] first addressed the problem. These algorithms generally contain two main steps:

1. **Registration**: Aligning the LR frames.

2. **Fusion**: Combining the information in all the LR frames to construct an SR image.

Most authors use planar homographies in the registration step to describe the inter-image warp. This prevents the algorithm from dealing with images containing deformable surfaces.

Contributions. The main contributions of this article are:

- The use of Free-Form Deformation (FFD) registration to replace the planar homography.

- An alternation-based approach to improve the registration and the recovered SR image.

Our method achieves Super-Resolution for images of deformable surfaces. Figure 1 shows a comparison of the different type of images handled by classical SR methods and our algorithm.
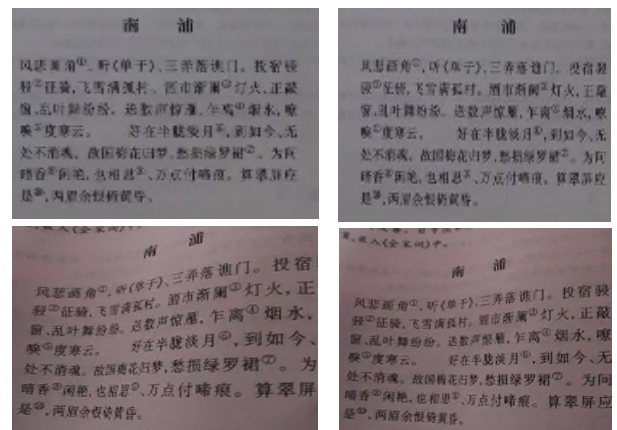

Figure 1 – *(Top) Classical SR methods use planar homography motion models; (Bottom) Our method handles deformable surfaces with local image motion.*

Plan. Section 2 introduces the state-of-the-art SR methods. Section 3 explains the basic methods of non-rigid registration with a focus on B-spline based FFD. Section 4 gives our algorithm and section 5 shows experimental results.

Notation. Our notations are given in Table 1.

2 The Super-Resolution Problem

In this section we describe the SR problem and introduce a popular matrix notation. We show that the limitations of most SR methods come from the motion model between the SR frame and the LR frames.

SR is the process of combining a sequence of LR noisy blurred images to produce an HR image. The idea of using multiple frames for SR was first proposed in [1]. They adopted a frequency domain approach which was later extended by others, such as [2]. Frequency domain methods are intuitive, simple and computationally cheap, but they are extremely sensitive to model errors, limiting their use. Also, by definition, only pure translational motion can be treated with such tools and even small devi-

Notations	
$\mathcal{R}$	Region of interest
$\mathbf{X}$	The unknown high resolution image (vectorized)
$\underline{\mathbf{X}}$	the estimated high resolution image (vectorized)
$\mathbf{Y}_k$	the kth low resolution image (vectorized)
$\mathbf{V}$	the additive Gaussian noise
$\mathcal{W}$	the Free-Form-Deformation warping function
H_{cam}	the camera blur operator
F	the geometric motion operator
D	the decimation operator
n	the number of frames in the low resolution sequence
$\mathbf{u}$	the Free-Form-Deformation warping parameter
γ	the downsampling scale

Table 1 – *Table of notations.*

ations from translational motion significantly degrade performance. Another popular approach solves the problem of resolution enhancement in the spatial domain, such as [3] and [4]. Many methods makes use of an optical flow estimated from the low resolution images (although this does not seems very successful so far) such as [5] and [6]. [7] and [8] use iterative computation of the optical flow and the super-resolution which resemble our approach. Recent methods [9] have been proposed to try to avoid estimating a warp or the optical flow by using Nonlocal-Means (NLM) algorithm.

The forward imaging model. Now we introduce the forward imaging model found in many of the SR literature. According to this model, a real scene with continuous intensity distribution $\underline{\mathbf{X}}(x, y)$ is warped and blurred by the camera lens. It is then discretized at the CCD resulting a digitized noisy LR frame. This process is illustrated in Figure 2.

The most common matrix notation found in the literature for this process is:

$$\mathbf{Y}_k = \mathsf{D}_k \mathsf{H}_k \mathsf{F}_k \mathbf{X} + \mathbf{V}_k, \quad k = 1, ..., n \qquad (1)$$

where $\mathbf{X}$ is the SR image and is a discretized representation of the real world scene; F_k is the geometric motion operator that warps pixels from $\mathbf{X}$ to the kth LR frame $\mathbf{Y}_k$. The images are rearranged in lexicographic order. H_k models the combined effect of the camera's Point-Spread Function (PSF) and atmospheric blur. In this article we use a shift-invariant Gaussian kernel. D_k is the decimation operator and $\mathbf{V}_k$ is the additive Gaussian noise. n is the total number of LR frames.

In most of the literature, the geometric motion operator F_k is modeled as a planar homography. Clearly this model cannot handle images of deformable surfaces in which the relation between the HR frame $\mathbf{X}$ and kth LR frame $\mathbf{Y}_k$ can not be described by a homography. In this paper we propose to use deformable registration to replace planar homographies.

Maximum Likelihood Estimate of the SR frame. After the non-rigid registration we proceed to make a Maximum

Likelihood Estimate (MLE) of the SR frame which in that case is the average of the registered images:

$$\underline{\mathbf{X}}(\mathbf{q}) = \frac{1}{n} \sum_{k=1}^{n} \mathbf{Y}_k(\mathcal{W}(\gamma \mathbf{q}; \mathbf{u}_k)) \qquad (2)$$

in which the $\mathbf{q} \in \mathbb{R}^2$ is a pixel and $\mathbf{Y}_k(\mathcal{W}(\gamma \mathbf{q}; \mathbf{u}_k)), k = 1, 2, ..., n$ are the registered frames. $\mathcal{W}$ is the warp that we will discuss in the next section.
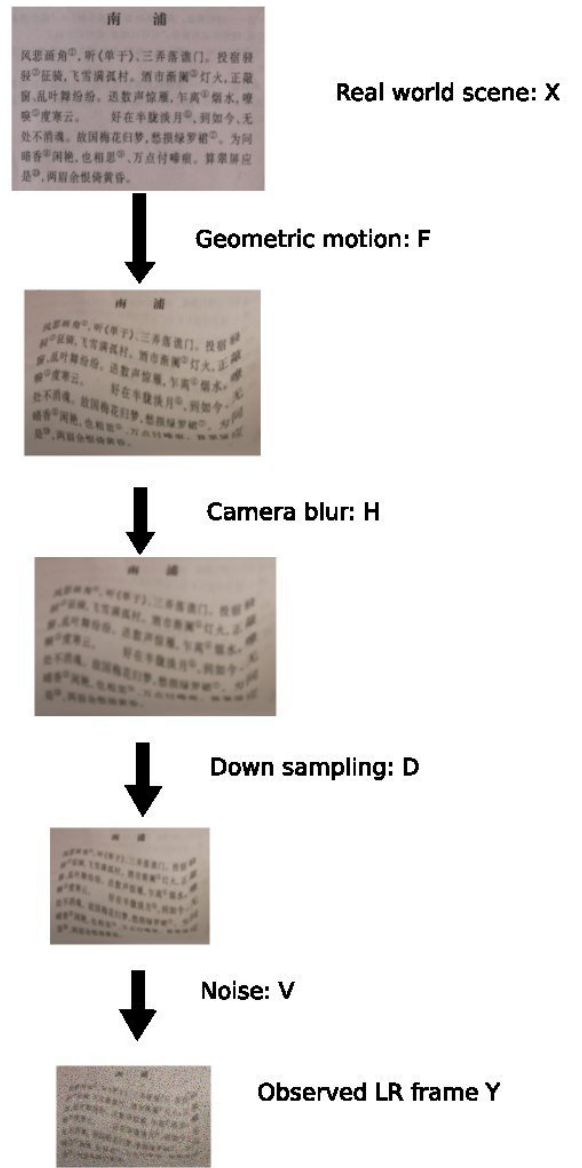

Figure 2 – *The forward imaging model.*

3 Non-rigid Registration Using Free-Form Deformation

3.1 Parametric Image Warps

Parametric image warps are often used for deformable image registration. A parametric image warp maps a point $\mathbf{q}$ from a source image to the corresponding point $\mathbf{q}'$ in the target image. It is written as a function $\mathcal{W} : \mathbb{R}^2 \times \mathbb{R}^p \to \mathbb{R}^2$ of the point coordinates $\mathbf{q} \in \mathbb{R}^2$ and a parameter vector $\mathbf{u} \in \mathbb{R}^p$ as follows:

$$\mathbf{q}' = \mathcal{W}(\mathbf{q}; \mathbf{u}) \tag{3}$$

The parameter vector $\mathbf{u}$ may encapsulate many different kinds of parameters, depending on the nature of the warp, which typically are image control points for 2D warps and 3D control points, surface and camera parameters for 3D warps.

Typical image warps include Thin-Plate Spline, Radial Basis Function warps and Free-Form Deformation (FFD). In this paper we choose FFD for its excellent ability to model local deformation. It is also computationally efficient compared to other types of warps.

3.2 Free-Form Deformations

We replace the above planar homography by an FFD. FFD warps are based on the tensor product between two $\mathbb{R} \mapsto \mathbb{R}$ splines. The *source control points* thus lie on a regular grid. It was proposed in [10] for computer graphics applications and has been extensively used since then. Early work with FFD warps for image registration is [11]. Most recently FFD warps have been used for shape registration in [12].

Consider two one-dimentional sets of evenly spaced knots with unity inter-knot distance. Assume one set spans the horizontal, x axis of the image, and the other one spans the vertical, y axis. These two sets of knots define a regular grid whose vertices are the source control points. A scalar target value $z_{u,v}$ is associated to each source control point $\begin{pmatrix} u & v \end{pmatrix}^\mathsf{T} \in \mathbb{N}^2$.

For a point $\mathbf{q} \in \mathbb{R}^2$ with $\mathbf{q}^\mathsf{T} = \begin{pmatrix} x & y \end{pmatrix}$, the tensor product writes:

$$\sum_{a=0}^{3} \sum_{b=0}^{3} B_a(x - \lfloor x \rfloor) B_b(y - \lfloor y \rfloor) z_{\lfloor x \rfloor + a, \lfloor y \rfloor + b} \tag{4}$$

The $\mathbb{R}^2 \mapsto \mathbb{R}^2$ warp is obtained by stacking two such tensor products sharing their source control points, or equivalently, by replacing the scalars $z_{u,v}$ by so-called *target control points* $\mathbf{c}_{u,v}$, giving the FFD warps as:

$$\mathcal{W}(\mathbf{q}) \stackrel{def}{=} \sum_{a=0}^{3} \sum_{b=0}^{3} B_a(x - \lfloor x \rfloor) B_b(y - \lfloor y \rfloor) \mathbf{c}_{\lfloor x \rfloor + a, \lfloor y \rfloor + b} \tag{5}$$

Figure 3 shows an example of using FFD to find the best warp bewteen two deformed images. In our algorithm, the first LR frame $\mathbf{Y}_1$ is chosen as the reference frame and warp the rest of LR frames to it. This process minimize the squared intensity difference between the warped image and the original image:

$$\mathbf{u}_k = \arg \min_{\mathbf{u}_k \in \mathbb{R}^p} \sum_{\mathbf{q} \in \mathcal{R}} \|\mathbf{Y}_1(\mathbf{q}) - \mathbf{Y}_k(\mathcal{W}(\mathbf{q}; \mathbf{u}_k))\|^2 \tag{6}$$

where $\mathbf{u}_k$ are the parameters of the warp between the first and the kth frame; and $\mathcal{R}$ is the region of interest.
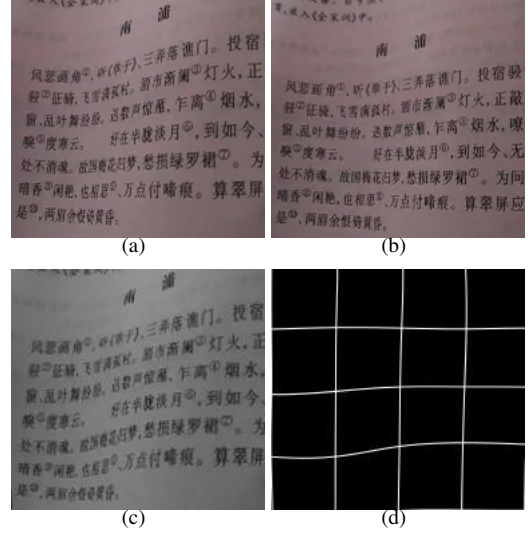

(a)　　　　(b)

(c)　　　　(d)

Figure 3 – *An example of image registration with FFD. A warp between input image (a) and (b) is calculated and then applied on input image (b) to obtain the transformed image (c). The grid (d) illustrates the result.*

4 An Alternation Approach

In the registration step, the first LR frame is chosen as the reference and the rest of the LR frames are warped to it. But since the first LR frame does not contain all the information to generate the other LR frames, this is only an approximation. The ideal choice would be the SR frame itself that we are looking for. Therefore we face a "chicken-and-egg" situation. We can solve this by an alternation-based approach:

1. **Initialize the warps** from the reference to the other images. This is done in a sequential manner: we set $\mathbf{u}_2 \leftarrow \mathbf{u}_1$ ($\mathbf{u}_1$ is the identity warp) and proceed to refine $\mathbf{u}_k$ by nonlinear minimization of the intensity discrepancy:

$$\mathbf{u}_2 \leftarrow \arg \min_{\mathbf{u}_2 \in \mathbb{R}^p} \sum_{\mathbf{q} \in \mathcal{R}} \|\mathbf{Y}_1(\mathbf{q}) - \mathbf{Y}_2(\mathcal{W}(\mathbf{q}; \mathbf{u}_2))\|^2 \tag{7}$$

This is done with Gauss-Newton iterations. We then proceed to estimate $\mathbf{u}_3$ by setting $\mathbf{u}_3 \leftarrow \mathbf{u}_2$ and minimizing the discrepancy to the reference image and so on and so forth.

2. **Compute the SR image** from the registered input images as in other classical SR algorithms.

$$\underline{\mathbf{X}}(\mathbf{q}) = \frac{1}{n} \sum_{k=1}^{n} \mathbf{Y}_k(\mathcal{W}(\gamma\mathbf{q}; \mathbf{u}_k)) \qquad (8)$$

3. **Refine the warps** from the SR to the other images (except the reference one, to which it is fixed). This is done very similarly as in the initialization step, except that one does not need to proceed sequentially since one has an initial guess. For each image $k = 2, ..., n$, solve:

$$\mathbf{u}_k \leftarrow \arg\min_{\mathbf{u}_k \in \mathbb{R}^p} \sum_{\mathbf{q} \in \mathcal{R}} \|\underline{\mathbf{X}}(\mathbf{q}) - \mathbf{Y}_k(\mathcal{W}(\gamma\mathbf{q}; \mathbf{u}_k))\|^2 \qquad (9)$$

5 Experiments

5.1 The Paper Sequence

The paper sequence contains 15 images of size 190×180. In the registration step we set the grid size to 5×5. It takes about 5 seconds to calculate the warp between each pair of images. Figure 4 shows the improved readibility of the text sequence.
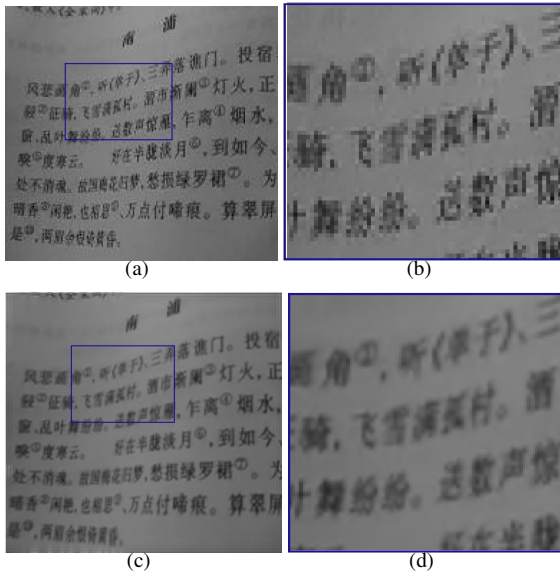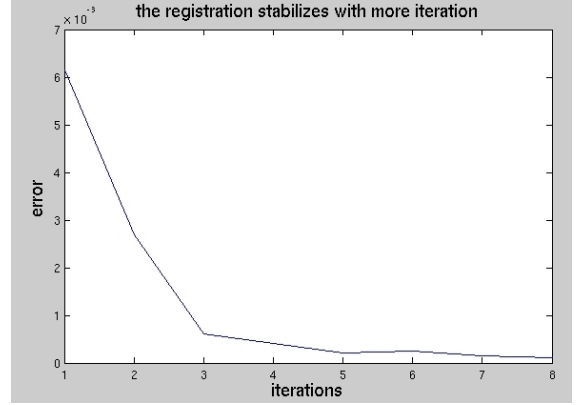
Figure 5 – *After 8 iterations, the intensity squared difference between each consecutive pair of SR estimations diminishes until below 1×10^{-3}.*

5.2 The T-Shirt Sequence

The t-shirt sequence contains 14 images of size 251×201. We use a slightly larger grid of size 6×6. The t-shirt sequence shows a deformation different than that of the paper sequence. It involves stretching and shearing. The registration for this sequence is more accurate since there is less illumination variation than for the paper sequence.
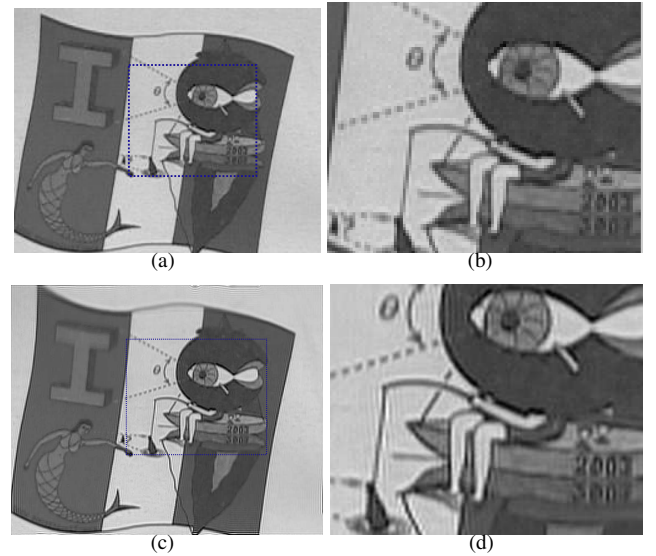
Figure 4 – *The paper sequence. (a-b):The original image (c-d): SR result with up-sampling factor 2.*

Figure 6 – *The t-shirt sequence. (a-b): The original image (c-d): SR result with up-sampling factor 3.*
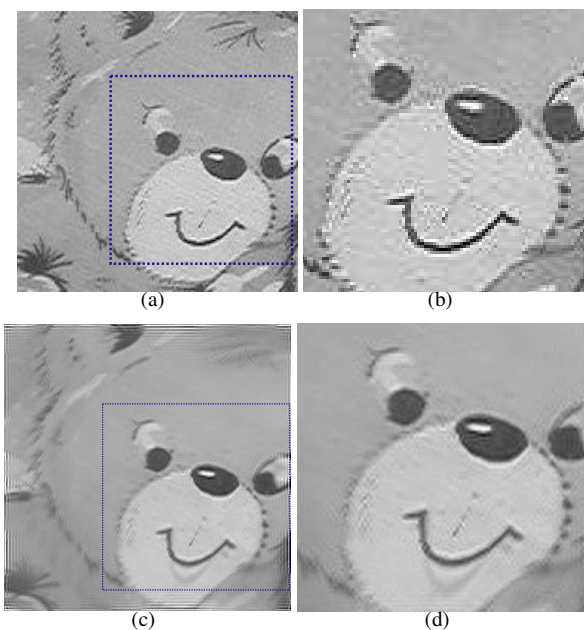
Figure 7 – *The cloth sequence. (a-b): The original image (c-d): SR result with up-sampling factor 2.*

5.3 The Cloth Sequence

The cloth sequence in Figure 7 contains 16 images of size 150×150. The visible registration error around the mouth of the bear is due to drastic movements between consecutive frames in the sequence. This can be improved by more iterations in the FFD registration step.

6 Conclusion

This paper extends the conventional SR algorithm by using FFD to register low resolution images with deformable surfaces and then adopt an alternation-based procedure to improve the registration. Experiments show that under small illumination variations the resulting SR estimation is quite encouraging.

The registration results can be further improved by considering the material properties of the deformable surfaces. Concerning the experiments, since conventional methods such as those that use globally affine registration do not produce sensible results, it is difficult to compare with competing methods. We envision a simutaneous approach to SR estimation in the future work.

References

[1] R. Tsai and T. Huang. Multiframe image restoration and registration. *Advances in Computer Vision and Image Processing*, 1:317–339, 1984.

[2] N. K. Bose, H. C. Kim, and H. M. Valenzuela. Recurcive implementation of total least squares algorithm for image reconstruction from noisy, undersampled multiframes. *Proc. IEEE Int. Conf. Acoustics,Speech, and Signal Processing*, 5:269–272, 1993.

[3] D. P. Capel. *Image Mosaicing and Super-Resolution*. Springer,New York, NY, USA,, 2004.

[4] L. C. Pickup, S. J. Roberts, and A. Zisserman. Optimizing and learning for super-resolution. *Proceedings of the 17th British Machine Vision Conference (BMVC '06)*, 2006.

[5] Brian C. Tom and Aggelos K. Katsaggelos. Resolution enhancement of monochrome and color video using motion compensation. *IEEE Trans. Image Processing*, 10:278–287, 2001.

[6] Sina Farsiu, Dirk Robinson, Michael Elad, and Peyman Milanfar. Advances and challenges in super-resolution. *International Journal of Imaging Systems and Technology*, 14:47–57, 2004.

[7] Zhang B. Huang H.F. Shen, L.P and P.X. Li. A map approach for joint motion estimation, segmentation, and super resolution. *IEEE Trans. On Image Process*, 16:479–490, 2007.

[8] K.J. Barnard R.C. Hardie and E.E. Armstrong. Joint map registration and high-resolution image estimation using asequence of undersampled images. *IEEE Trans. Image Processing*, 6:1621–1633, 1997.

[9] M. Takeda H. Milanfar P. Protter, M. Elad. Generalizing the nonlocal-means to super-resolution reconstruction. *Image Processing, IEEE Transactions on*, 2009.

[10] T. W. Sederberg and S. R. Parry. Free-form deformation of solid geometric models. *SIGGRAPH*, 1986.

[11] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, and D. J. Hawkes. Nonrigid registation using free-form deformations: Application to breast mr images. *IEEE Transactions on Medical Imaging*, 18:712–721, 1999.

[12] X. Huang, N. Paragios, and D. Metaxas. Shape registration in implicit spaces using information theory and free form deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28, 2006.