

Efficient Super-Resolution and Applications to Mosaics

Assaf Zomet Shmuel Peleg
School of Computer Science and Engineering
The Hebrew University of Jerusalem
91904 Jerusalem, Israel
{zomet, peleg}@cs.huji.ac.il

Abstract

Mosaicing and super resolution are two ways to combine information from multiple frames in video sequences. Mosaicing displays the information of multiple frames in a single panoramic image. Super-resolution uses regions which appear in multiple frames to improve resolution and reduce noise. The aim of this work is constructing a high resolution mosaic from a video sequence in an efficient way.

Simple combination of the two methods is problematic, since the alignment used in mosaicing may not be accurate enough for super resolution. Another issue is the efficiency of the super resolution algorithm, which requires heavy computations, especially when applied to large images such as panoramic mosaics.

This paper introduces two novelties. First, a framework for super resolution algorithms is presented, which enables the development of very efficient algorithms. Second, a method for applying super resolution to panoramic mosaics is presented. This method preserves the geometry of the original mosaic image, while improving its resolution.

1. Introduction

Video sequences of a scene can be compactly represented in a single image using panoramic mosaicing, by projecting the images to a common manifold [9, 8, 7, 12]. The overlap between the input images can be used to increase the resolution of the mosaic and reduce noise, by applying super resolution [13]. However, existing super resolution algorithms solve a very large optimization problem, and thus are computationally costly. They also require very accurate alignment over the entire image. This work handles these aspects of mosaicing and super resolution.

When constructing the mosaic, a region in the mosaic corresponds to regions in several images. There are two possible ways to determine the color details of this region: (i) Combining the aligned images by a median, average etc. (ii) Selecting a region from only one of the images.

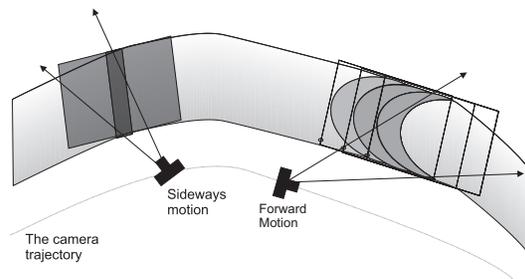


Figure 1. Manifold mosaicing is the process of aligning strips along the seams so the optical flow between each two strips is parallel. The result is general non-rectangular strips.

Method (i) requires accurate alignment over the entire image area, otherwise the resulting will be blurred [2]. Method (ii) requires alignment only along the seams. Thus it is more useful in cases where the camera motion, the scene geometry and the imaging conditions are challenging [9, 8, 7], as illustrated in Fig. 1.

The algorithm presented in Section 3 improves the resolution of the mosaic, even in cases where the alignment used for mosaicing is not accurate enough over the entire region. A preliminary presentation of this algorithm was made in [13].

An important part of this work is the development of efficient super resolution algorithms. The most common model for super resolution [4, 2, 11, 3, 6] presents the problem in the following way: The low resolution input images are the result of different imaging processes on a high resolution image. The imaging processes are usually modeled by a geometric transformation, blur, subsample, and additive noise. As a first stage, the imaging processes are modeled and estimated, based on motion computation between the input images, approximation of the blur caused by the camera and the motion, etc. Then an optimization function expressing the imaging process is defined. The large number of variables in this problem usually requires an iterative solution.

Several ways to solve to this optimization problem were proposed. Irani & Peleg [4] used Iterative Back Projection, adopted from computer-aided Tomography. Patti et. al. [6] used the non-linear Projection-Onto-Convex-Hull, with implementation similar to [4]. Some authors [1, 10] formulated the super resolution problem in probabilistic bayesian framework, and used MRF for modeling the prior, and finding the solution. A unifying framework for super resolution as a generalization of image restoration, was presented by Elad & Feuer [3]. They formulated the super resolution problem using matrix-vector notations, and showed that existing super resolution techniques are standard methods for solving linear equations. Based on this analysis, they proposed other sparse matrix optimization methods for the problem.

Recently, Shekarfroush & Chellapa proposed an algorithm for reconstructing a super resolved image analytically from the input images. They also described a way to estimate the camera blur function from the input images.

2. Super Resolution

Super resolution is the reconstruction of a high resolution image from several overlapping noisy low-resolution images. It can be presented as a large sparse linear optimization problem, and solved using explicit iterative methods [3, 2]. In the presented framework we use a matrix-vector formulation [3] in the analysis, but the implementation is by standard operations on images such as convolution, warping, sampling, etc. By alternating between the two formulations a considerable speedup in the super resolution computation is gained: This takes advantage of the two worlds: gradient-based optimization techniques are used, while the gradient is computed efficiently in the image domain instead of multiplying large sparse matrices.

In the analysis part we represent images as column vectors. The basic operations in the image formation model, such as convolution, sampling and warping are linear, and thus can be represented as matrices operating on these vector images. Given n input images $g_1..g_n$, the image formation process of g_k from the super resolved image f can be formulated in the following way [3]:

$$\vec{Y}_k = D_k C_k F_k \vec{X} + \vec{E}_k$$

where:

- \vec{X} is the high resolution image f of size $[L \times L]$, reordered in a vector of size $[L^2]$.
- \vec{Y}_k is the k-th input image g_k of size $[M_k \times M_k]$, reordered in a vector of size $[M_k^2]$.
- \vec{E}_k is the normally distributed additive noise of size $[M_k \times M_k]$, reordered in a vector of size $[M_k^2]$.

- F_k is the geometric warp matrix, of size $[L^2 \times L^2]$.
- C_k is the blurring matrix, of size $[L^2 \times L^2]$.
- D_k is the decimation matrix, of size $[M_k^2 \times L^2]$.

Stacking the vector equations from the different images into a single matrix-vector:

$$\begin{bmatrix} \vec{Y}_1 \\ \vdots \\ \vec{Y}_k \end{bmatrix} = \begin{bmatrix} D_1 F_1 C_1 \\ \vdots \\ D_k C_k F_k \end{bmatrix} \vec{X} + \begin{bmatrix} E_1 \\ \vdots \\ E_k \end{bmatrix} \iff \vec{Y} = A \vec{X} + E$$

We assume the noise is uncorrelated and has uniform variance. In this case, the maximum likelihood solution is found by minimizing the functional:

$$L(\vec{X}) = \frac{1}{2} \|\vec{Y} - A \vec{X}\|_2^2$$

deriving L with respect to \vec{X} :

$$\nabla L = 0 \implies A^T (A \vec{X} - \vec{Y}) = 0 \iff$$

$$\iff \sum_{k=1}^N F_k^T C_k^T D_k^T (D_k C_k F_k \vec{X} - Y_k) = 0$$

The matrices of this linear system are very large, and iterative methods are required for its solution. Gradient-based iterative methods can be used without explicit construction of these large matrices. Instead, the multiplication with A and $A^T A$ is implemented using only image operations such as warp, blur and sampling.

The matrix $A^T A$ operates on vectors \vec{X} , corresponding to an image of the size of the super resolution solution f ,

$$A^T A \vec{X} = \sum_{k=1}^N F_k^T C_k^T D_k^T D_k F_k C_k \vec{X}$$

The matrix A^T operates on vectors \vec{Y} , stacking of the input images $g_1 \dots g_N$ reordered in column vectors $\vec{Y}_1 \dots \vec{Y}_k$

$$A^T \vec{Y} = \sum_{k=1}^N F_k^T C_k^T D_k^T \vec{Y}_k$$

The matrices F_k, C_k, D_k model the image formation process, and their implementation is simply the image warping, blurring and subsampling respectively. The implementation of the transpose matrices is also very simple:

- D_k^T is implemented by upsampling the image without interpolation, i.e. zero padding.

- C_k^T - For a convolution blur, this operation is implemented by convolution with the flipped kernel, i.e. if H is the imaging blur kernel, then the flipped kernel is the matrix K such that $\forall i, j, K(i, j) = H(-i, -j)$. Otherwise, C_k^T is implemented by forward projection of the intensity values, using the weights of the original blur filter.
- F_k^T - If F_k is implemented by backward warping, then F_k^T should be the forward warping of the inverse motion.

The simplest implementation of this framework is using Richardson iterations [5], with iteration step:

$$X^{(n+1)} = X^n + \sum_{k=1}^N F_k^T C_k^T D_k^T (Y_k - D_k C_k F_k X^n)$$

This is a version of the Iterated Back Projection [4], using a specific blur kernel and forward warping in the back projection stage.

We have implemented the conjugate gradient in this framework, and achieved a considerable speedup in computation. The results are presented in Section 2.2.

2.1 Constrained Minimization

In many cases super resolution does not have a unique solution, and the matrix $A^T A$ is not invertible. This can be solved by introducing constraints on the solution, e.g. smoothness. If the constraints are linear, and can be easily computed from the image, (e.g. convolution), then they can be added to our framework, by minimizing:

$$L(\vec{X}) = \frac{1}{2} (\|\vec{Y} - A\vec{X}\|_2^2 + \lambda \|M\vec{X}\|_2^2)$$

Where M is the constraining operator, and λ is the regularization coefficient. Using the derivative of L with respect to \vec{X} :

$$A^T A \vec{X} + \lambda M^T M \vec{X} - A^T \vec{Y} = 0$$

The image corresponding to the added term $\lambda M^T M \vec{X}$ can be computed using image operations, by applying the constraining operator and its transpose, similarly to the blur operators mentioned in the previous section.

Non-linear constraints can also be used in this framework. an example is the POCS implementation in [6].

2.2 Results

To demonstrate the computational benefit of our framework, we implemented the conjugate gradient method using image operations. We compared its running time to another image-based non-constrained algorithm, the IBP of Irani & Peleg [4]. We captured images of a plane by a hand

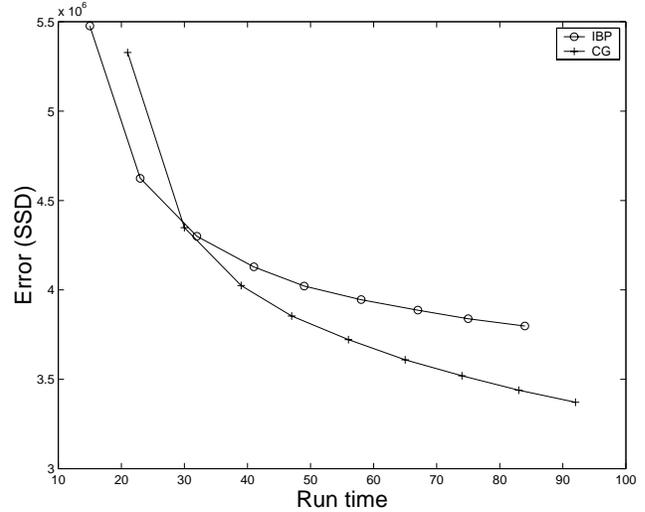


Figure 3. The minimization Error as a function of running time. The CG iteration times are marked by crosses, and the IBP ones are marked by circles.

held camera, and computed the projective-planar motion between the images. We then applied both methods of super resolution, and compared the computation time and the results.

The results of both of the algorithms after several iterations are presented in Fig. 3. Note the differences in the separation between the fingers of the player. The CG result at the 3rd iteration is similar to the IBP result at the 5th iteration, and after another iteration, the CG result is similar to the IBP result at the 9th iteration.

The graph in Fig. 2 shows these results quantitatively, presenting the minimization Error as a function of the program running time. The first iteration in the CG method is slower, since it requires additional multiplication with the matrix $A^T A$. The next iterations of both of the methods require a single multiplication with A and A^T , so the running time of each iteration is similar (with minor advantage to the CG method). This means that the comparison of the running time of these algorithm depends mainly on the convergence rate. One can see in the graph that the convergence of the CG method is much faster, yielding better results in very short time. This can be further accelerated by using efficient image operations in the computation of the gradient.

3. Combining Super Resolution and Mosaicing

Mosaicing algorithms such as manifold mosaicing emphasize image alignment in the seams between strips, whereas for super resolution algorithms alignment accuracy over the entire overlapping region is very important: image



Figure 2. A Panoramic mosaic with a detailed window before super resolution (a, b) and after super resolution (c, d). The camera motion included rotation and translation.

resolution can be improved only when an accurate alignment can be computed [2]. Thus the image alignment for mosaicing may not be accurate enough for super resolution. We present a method for combining super resolution into manifold mosaicing even in cases where the alignment used for mosaicing may not be accurate over the overlapping regions.

The proposed algorithm starts with creating the mosaic using generalized strips [8, 7]. As the first step in the super resolution algorithm, the mosaic is expanded to the desired resolution using some image interpolation method (for example a bilinear interpolation). The resolution is enhanced using the following steps for each strip in the panorama:

1. All images containing this strip are collected, and the location of this strip in these images is computed using the parameters found at the initial image alignment step.
2. All overlapping frames are aligned to the strip in the mosaic. This alignment is computed only for the strip and a small neighborhood, and may use a different alignment model than used for mosaicing. For example, even when the mosaic algorithm uses a rigid alignment with only image translation and image rotation (3 parameters), the alignment in this step may use an affine model (6 parameters) or a homography

(8 parameters). Since only a strip is aligned, and not the entire image, and since a more accurate model may be used, this alignment will be more accurate than the alignment used for mosaicing.

3. From all images aligned to the strip, images with poor quality or poor alignment are removed. Image quality may include sharpness, noise, dynamic range, and other quality criteria.
4. A super resolution algorithm is applied to the strip using the selected images.

The strip in the mosaic is enhanced without changing the geometry of the mosaic. The seams in the original panorama, the result of sampling artifacts, noise and gain differences, also disappear. An example is shown in Fig. 4.

4 Summary

For the construction of high resolution mosaic images, two novelties were presented. First, an algorithm for applying super resolution to mosaic images, using the enlarged panoramic image as the reference for registration. Second, a framework for fast implementation for super resolution was introduced. Our framework factors the super resolution optimization to two parts: the optimization technique, and the gradient computation by image operations. An efficient

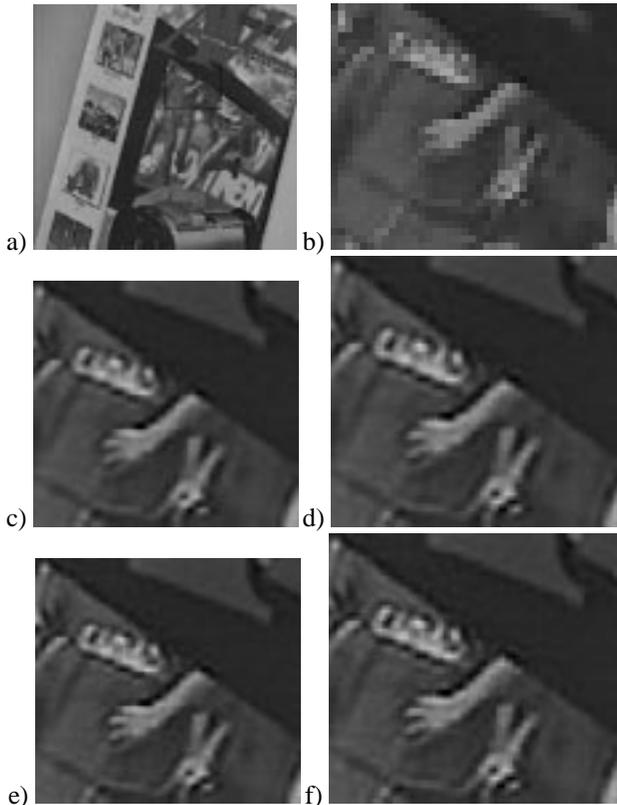


Figure 4. Super resolution results.

a) An original frame. b) A small region from the original frame, enlarged by NN interpolation. The region is marked in figure (a).

c) IBP method - 5th iteration. d) IBP method - 9th iteration.

e) CG method - 3rd iteration. f) CG method - 4th iteration.

algorithm can be implemented by choosing the appropriate gradient-based technique, and by optimizing the image operations.

By combining the two results, panoramic images with enhanced resolution can be constructed in an efficient way, without distorting its geometric structure. This process also smoothes the seams without degradation in the image quality.

References

- [1] M. Berthod, H. Shekarforoush, M. Werman, and J. Zerubia. Reconstruction of high resolution 3d visual information using sub-pixel camera displacements. In *CVPR94*, pages 654–657, 1994.
- [2] D. Capel and A. Zisserman. Automated mosaicing with super-resolution zoom. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Santa Barbara*, pages 885–891, June 1998.
- [3] M. Elad and A. Feuer. Restoration of a single super-resolution image from several blurred, noisy, and undersampled measured images. *IP*, 6(12):1646–1658, December 1997.
- [4] M. Irani and S. Peleg. Improving resolution by image registration. *GMIP*, 53:231–239, 1991.
- [5] C. Kelley. Iterative methods for linear and nonlinear equations. pages 11–32, 1995.
- [6] A. Patti, M. Sezan, and A. Tekalp. Superresolution video reconstruction with arbitrary sampling lattices and nonzero aperture time. *IP*, 6(8):1064–1076, August 1997.
- [7] S. Peleg and J. Herman. Panoramic mosaics by manifold projection. In *IEEE Conf. on Computer Vision and Pattern Recognition*, pages 338–343, June 1997.
- [8] B. Rousso, S. Peleg, and I. Finci. Mosaicing with generalized strips. In *DARPA Image Understanding Workshop*, pages 255–260, May 1997.
- [9] B. Rousso, S. Peleg, I. Finci, and A. Rav-Acha. Universal mosaicing using pipe projection. In *Int. Conf. on Computer Vision*, pages 945–952, January 1998.
- [10] R. Schultz and R. Stevenson. Extraction of high-resolution frames from video sequences. *IP*, 5(6):996–1011, June 1996.
- [11] H. Shekarforoush and R. Chellappa. Data-driven multi-channel super-resolution with application to video sequences. In *UMD*, 1998.
- [12] R. Szeliski. Video mosaics for virtual environments. *IEEE CGA*, 16(2):22–30, March 1996.
- [13] A. Zomet and S. Peleg. Applying super-resolution to panoramic mosaics. In *WACV98*, page Demo IV, 1998.