

High-Quality Still Images from Video Frame Sequences

T. Daubos^a and F. Murtagh^a

^a School of Computer Science, Queen's University Belfast, Belfast BT7 1NN, Northern Ireland

ABSTRACT

This article tackles the classic super-resolution (SR) problem¹ of obtaining a high-resolution (HR) still image from a sequence of low-resolution (LR) images that have been warped and sub-sampled. The goal here is to recover frequencies higher than the Nyquist frequency by merging the LR information. We focus on the critical step of the SR process before any fusion technique can be applied which consists of the registration of LR images with an arbitrary reference image at sub-pixel accuracy. We propose a registration algorithm for color images, derived from the one described by Djamdji and Bijaoui in Ref. 2. This algorithm achieves automatic feature-based registration at sub-pixel accuracy. It seeks to take advantage of the multi-band (RGB) information in a color image to improve the robustness and accuracy compared to more usual greyscale registration. The fusion of the data from LR images into a higher resolution image is then carried out through thin plate spline interpolation. The results show the algorithm's performance for simulated image sets. The influence of several parameters on the registration algorithm is described.

Keywords: super-resolution, image registration, wavelet transform, thin plate spline

1. INTRODUCTION

Usually SR reconstruction methods involve three steps: (i) registration of a set of LR images, (ii) interpolation of the LR data and (iii) restoration by removal of noise as well as sensor and optical blur. These three steps can be considered separately, as in Ur and Gross,³ when the blur can be assumed to be the same for all LR frames. However, this approach is considered suboptimal since the interpolation and restoration stages have mutual influence on each other. Therefore many more recent algorithms aim to solve the interpolation and restoration problems simultaneously. In the class of image domain algorithms, we find methods based on projections onto convex sets (POCS).^{4,5} Irani and Peleg propose a similar method of averaging projections to iteratively solve for the SR image.^{6,7} There also exist Bayesian methods for SR image reconstruction that use statistical *a priori* models for the SR image. Here the restoration and interpolation problems are solved simultaneously by using a maximum *a posteriori* (MAP) or maximum likelihood (ML) formulation.^{8,9} While there are several approaches for solving stages (ii) and (iii) of the SR problem, one still needs to know with the best possible accuracy what is the motion of the imaging system between the LR frames.

In the image registration process, we assume that two or more observed sampled images represent the same scene. The registration task consists of finding the best spatial fit that matches one or more images with a reference image. The distortions we are interested in are dynamic (i.e., distortions occur in a non-predictable way for each image, unlike systematic distortions that could be corrected through a calibration process). They are also assumed to be external spatial distortions, due, usually, to the movement of the platform of the imaging system. Registration then becomes a matter of estimating the parameters of a mathematical model which restricts the type of deformations to be corrected. The choice of the mathematical model is closely related to the assumptions made about the set of images to be registered. Here it is assumed that the images were taken by a single imaging system, with the same imaging direction and over a short period of time (therefore no significant change in the lighting conditions is expected). This algorithm is feature-based. The features detected are significant structures of high intensity in the three Red-Green-Blue bands of the color images.

Further author information: (Send correspondence to T. Daubos)

T Daubos: E-mail: t.daubos@qub.ac.uk, Telephone: +44 28 90274726

F Murtagh: E-mail: F.Murtagh@qub.ac.uk, Telephone: +44 28 90274620

2. GEOMETRICAL REGISTRATION

Let us call “reference image” the image with regard to which all other images are to be registered, and “input image” the current image to be registered. The registration process is usually performed in two steps:

- Measurement of a set of strong features present simultaneously in both the reference and input images. We will call these features ground control points (GCPs).
- Determination of the parameters of a mathematical deformation model which determines the warping between the reference and input image.

Optionally a third step can be considered which consists of creating a corrected image from the input image once the deformation model is known. This can be done by output-to-input mapping and involves interpolation techniques. This step is usually implemented to allow direct comparison of intensities between the reference image and an interpolated registered image. However, in this article, the performance of the algorithm is evaluated using the errors between the expected positions of the pixels after correction of the input image and the positions actually found by our registration algorithm.

2.1. Choice of the deformation model

From the assumptions made that the images of the video sequence have been taken by a single imaging system, with the same imaging direction and over a short period of time, we chose to use as a deformation model a pair of first order, bivariate polynomials of type II as follows:

$$\begin{aligned} x' &= P(X, Y) = a_x X + b_x Y + c_x \\ y' &= Q(X, Y) = a_y X + b_y Y + c_y \end{aligned} \quad (1)$$

where (X, Y) are the coordinates of the GCPs in the reference image and (x', y') the transformed coordinates of the GCPs in the input image. $\{a_x, b_x, c_x\}$ and $\{a_y, b_y, c_y\}$ are the parameters of the two deformation polynomials P and Q. This deformation model corresponds to a general 2D affine transformation that allows us to correct distortions like translations, rotations, skew and aspect ratio.¹⁰

2.2. Wavelet transform

The wavelet transform used is the discrete, redundant (i.e., no decimation is carried out), “à trous” algorithm.¹¹ We briefly describe the “à trous” wavelet transform in 1D. One assumes that the sampled data $\{c_0(k)\}$ are the scalar products, at pixel k of the function $f(x)$, with the scaling function $\varphi(x)$. The scaling function, φ , is defined by:

$$\varphi_{j,k}(x) = \varphi\left(\frac{x-k}{2^j}\right) \quad (2)$$

where 2^{-j} is the dyadic scale parameter at scale j and k is the position parameter. The scaling function φ must also verify the dilation equation:

$$\psi(x) = 2 \sum_k h(k) \varphi(2x - k) \quad (3)$$

where $h(k)$ is a low-pass filter.

The first filtering is performed with a twice-magnified scale and leads to the set of coefficients $\{c_1(k)\}$ for the first scale of resolution. At each scale the signal is smoothed and information is lost. The information lost between these two scales is retained in the signal difference $\{c_0(k)\} - \{c_1(k)\}$ which is the set of coefficients corresponding to the wavelet transform $\varphi(x)$. The associated wavelet function $\psi(x)$ which generates the “details” or wavelet space can therefore be defined by:

$$\frac{1}{2} \psi\left(\frac{x}{2}\right) = \varphi(x) - \frac{1}{2} \varphi\left(\frac{x}{2}\right) \quad (4)$$

The distance between two samples increasing by a factor 2 between scale j and the next one, $c_{j+1}(k)$ is calculated by:

$$c_{j+1}(k) = \sum_l h(l)c_j(k + 2^j l) \quad (5)$$

and the discrete wavelet coefficients by:

$$w(j + 1, k) = c_j(k) - c_{j+1}(k) \quad (6)$$

The “à trous” algorithm is easily extended to 2D space. The scaling function used here is the piecewise polynomial B_3 -spline function. In order to speed up the computation, we assume separability in the 2D case. This leads to a row-by-row convolution by the mask of coefficients $(\frac{1}{16}, \frac{1}{4}, \frac{3}{8}, \frac{1}{4}, \frac{1}{16})$, followed by a column-by-column convolution. A mirror boundary condition is applied for the $c_j(k)$ (i.e., $c_j(k + N) = c_j(N - k)$).

2.3. Choice of the color system: RGB vs HLS

Both color systems were tested with the registration algorithm. While the intensity-hue-saturation (HLS) system is widely used in the field of data fusion,¹² the red-green-blue (RGB) system showed better performance for feature based registration. This arises from the fact that the hue band, which refers to the dominant or average wavelength that contributes to a color, is arbitrarily distributed between 0 and 360 degrees. (Both 0° and 360° correspond to red, 120° to green and 240° to blue). Therefore, in this band, strong features will be pixels with color somewhere between magenta and red, which is of no particular relevance for defining reliable ground control points.

2.4. Registration algorithm

The basic idea behind wavelet-based registration is that matching between the two sets of GCPs can be done unambiguously and automatically using the property that the sampling step is proportional to the scale.¹³ Usually, when one tries to match two sets of points, there are always ambiguities since a given point can only be associated with a small region, R , around this point considering the prior geometric uncertainty. Nevertheless, using the wavelet transform, one can ensure that, provided a sufficient number of scales is used, only one GCP will be detected in this region R .

- For the three Red-Green-Blue bands of both reference and input images, we compute the wavelet transform up to scale N .
- The noise behavior is modeled as a stationary Gaussian noise with standard deviation σ . The standard deviations of the noise in the wavelet transforms are estimated, separately for the three RGB bands, from the first scale by 3-sigma clipping.
- For each scale from 1 to N :
 - The wavelet coefficients are thresholded at 3σ (taking in account the noise behavior in the multiresolution space).
 - The structures (areas at least 3×3 pixels wide with local maximum at the center) are detected. The local maxima of the structures play the role of GCPs.
- Starting with the lowest resolution scale N , pairs of GCPs are created between the reference and working image, the radius of the matching circle being half the size of the support of the wavelet at the scale considered.
- Pairs with a GCP falling on the border of the wavelet plane are discarded.
- Using all the pairs of GCPs from the three bands, a first estimate at scale N of the deformation polynomial’s parameters is computed by least square fit.

- From scale $N-1$ to scale 2:
 - GCPs of the working image are transformed using the first approximation of the deformation polynomials.
 - Pairs for which the residual distance between GCP in the reference scale and transformed GCP in the working scale is greater than the residual threshold (half the size of the wavelet with a lower bound at 1.0) are discarded.
 - Finally, the redundant pairs (different pairs making use of the same GCP) are also discarded, the pair with the lowest residual being kept.
 - All pairs from the three RGB bands having been checked, they are agglomerated together and the final estimation of the deformation polynomial’s parameters for this scale is carried out.
- Once the process has been iterated through every wavelet plane but the last one, the final deformation polynomials are output.

3. NUMERICAL EXPERIMENTS ON REGISTRATION

3.1. Simulation of test sets of images

In order to assess the accuracy of registration, test sets of images were created, simulating the successive frames of a short video sequence. We used Lena color images of various sizes as reference images. Sets of images with specific random deformations (shifts, rotations and scale variations) were created for which we know the true deformations applied with regard to the reference image. Test sets of greyscale images with identical deformations were also created to compare the performances of greyscale and color registration.

3.2. Measurement of the accuracy of registration

The measurement of registration accuracy over a whole test set was computed as follows:

- For every registered image in the test set:
 - For every pixel in the considered image:
 - * Compute the new coordinates of the pixel according to the deformation polynomials obtained from the registration process.
 - * Compute the “true” coordinates using the reverse transform of the known deformation polynomials used to create the test image.
 - * Registration error for this pixel corresponds to distance between these two positions.
 - Registration accuracy for the considered image is the mean error of all pixels.
- Registration accuracy for the test set is the mean error of all images in the set.

$$\text{Registration}_{error} = \frac{1}{N} \sum_{i=0}^N \frac{\sum_{j=1}^{X_{dim}} \sum_{k=1}^{Y_{dim}} Err_{j,k}^i}{X_{dim} Y_{dim}} \quad (7)$$

N is the number of images in the test set

X_{dim} and Y_{dim} are the dimension of the images.

Err^i is defined by:

$$Err_{j,k}^i = \sqrt{(P[i]_{reg}(j, k) - P[i]_{true}(j, k))^2 + (Q[i]_{reg}(j, k) - Q[i]_{true}(j, k))^2} \quad (8)$$

$P[i]_{true}, Q[i]_{true}$ is the reverse of the known deformation polynomials used to produce the i^{th} image.

$P[i]_{reg}, Q[i]_{reg}$ are the i^{th} deformation polynomials as found by the registration algorithm.

Table 1: Behavior of the number of GCPs across scales

Scales	Number of GCPs	Accuracy of registration	RMSE
5	9.0	1.297	0.926
4	68.6	0.521	1.456
3	218.2	0.179	1.036
2	579.8	0.057	0.561
1	461.4	0.147	0.280

Table 2: Influence of the dimensions of the images on accuracy for shifts

Nbr of scales	Dimension of images	Accuracy of registration	
		Color images	Greyscale images
4	64 x 64	0.137	0.170
5	128 x 128	0.064	0.093
5	256 x 256	0.051	0.062

3.3. About the use of the higher-resolution scale

One notices that no use is made of the last wavelet scale of highest resolution (scale 1). This arises from the fact that this scale also holds most of the noise. Therefore the threshold used for the wavelet coefficients must be higher than those used for the other scales. As a consequence, the number of detected GCPs in this last scale drops significantly. When computing the deformation polynomials' parameters, we are dealing only with the integer coordinates (in pixel unit) of the GCPs. The sub-pixel accuracy of the parameter values comes only, provided the pairs have been properly matched, from the number of pairs used for the least square fit. For this reason, the best accuracy is actually achieved at scale 2 instead of scale 1 since the number of GCPs is lower for the latter.

Table 1 gives an example for a test set of 10 (128×128 pixels) images with random shifts in a range within ± 1.0 pixel. A five scale wavelet transform is used. The number of GCPs detected in each scale is shown along with the registration accuracy with regard to the true deformations and the root mean squared error (RMSE) of the least square polynomial fit. The RMSE is defined as:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=0}^N x_{residual}(i)^2 + y_{residual}(i)^2} \quad (9)$$

N being the number of detected GCPs.

3.4. Influence of the dimension of the images

In order to measure the influence of the dimension of the images, we used several test sets showing the same deformations applied to images of increasing size. In Table 2, the deformations were random shifts in a range within ± 1.0 pixel along both axes. Table 3 shows results for random rotations, around the center of the image, in the range $\pm 5^\circ$. Table 4 shows results for random scale variations in the range $\pm 5\%$ of the original image size. In every case, the number of scales used was 4 for the 64×64 pixel images, and 5 for the other dimensions.

The dimensional factor has different influences depending on the type of deformation considered. For shifts, the larger the image the better the accuracy. This is due to the increase in the number of GCPs involved in the least square fit that provides more information and therefore better accuracy. For rotations, on the other hand, the accuracy drops when the image dimension increases. In this case, a greater dimension induces larger displacements of the GCPs. As far as accuracy is concerned, the benefit of a greater number of GCPs does not compensate for the increase in GCP displacements.

Table 3: Influence of the dimensions of the images on accuracy for rotations

Nbr of scales	Dimension of images	Accuracy of registration	
		Color images	Grey-scale images
4	64 x 64	0.240	0.249
5	128 x 128	0.304	0.316
5	256 x 256	0.563	0.565

Table 4: Influence of the dimensions of the images on accuracy for scale

Nbr of scales	Dimension of images	Accuracy of registration	
		Color images	Grey-scale images
4	64 x 64	0.129	0.228
5	128 x 128	0.053	0.088
5	256 x 256	0.041	0.054

3.5. Simulations for shifts

This section aims to evaluate the registration accuracy with regard to the amplitude of the shifts applied on the images on both axes. The test sets used were 10, 128×128 pixel, images and registration was carried out using a 5 scale wavelet transform. Results are shown in Table 5.

There is no significant change in registration accuracy when the amplitude of the shifts increases. The results for color registration are always better than for greyscale registration. The main difference comes from the number of misregistrations which increases in the case of greyscale registration, while no failure is detected with color registration.

3.6. Simulations for rotation

Here we investigate the sensitivity of the registration algorithm to the amplitude of rotations. We used three test sets each containing 10, 128×128 pixel, images. The deformations were rotations of random amplitudes ranging between $\pm 5^\circ$, $\pm 10^\circ$ and $\pm 15^\circ$ respectively for the three sets. The rotation axis is located at the center of the image. Again, registration was carried out using a 5 scale wavelet transform. Results are shown in Table 6.

In this case, the registration accuracy is significantly lower than in the case of shifts, with no great difference between color and greyscale registration. However, with regard to the number of misregistrations, color registration proves more robust than greyscale registration.

3.7. Simulations for scale deformations

In this section we deal with the accuracy with regard to scale deformations. The deformations were independent scaling along x and y axes. For the three test sets used, amplitudes ranged between $\pm 5\%$, $\pm 10\%$ and $\pm 15\%$ of the original image dimensions (128×128 pixels). Registration was carried out using a 5 scale wavelet transform. Results are shown in Table 7.

Table 5: Influence of the magnitude of the shifts on accuracy

Shift magnitude	Color images		Grey-scale images	
	Accuracy	Misregistration	Accuracy	Misregistration
± 1.0 pixels	0.070	0/10	0.093	0/10
± 2.0 pixels	0.062	0/10	0.085	2/10
± 3.0 pixels	0.070	0/10	0.084	3/10

Table 6: Influence of the magnitude of the rotations on accuracy

Rot magnitude	Color images		Grey-scale images	
	Accuracy	Misregistration	Accuracy	Misregistration
$\pm 5.0^\circ$	0.204	0/10	0.231	1/10
$\pm 10.0^\circ$	0.459	2/10	0.517	3/10
$\pm 15.0^\circ$	0.594	3/10	0.692	4/10

Table 7: Influence of the magnitude of scaling on accuracy

Scale magnitude	Color images		Grey-scale images	
	Accuracy	Misregistration	Accuracy	Misregistration
$\pm 5\%$	0.051	0/10	0.094	5/10
$\pm 10\%$	0.058	0/10	0.060	9/10
$\pm 15\%$	0.072	0/10	0.085	7/10

Results show the same degree of accuracy as for shift deformations for color registration. Again with regard to the number of misregistered images, color registration is more robust than greyscale registration.

3.8. Influence of the noise

Finally we tested the algorithm in the presence of Gaussian noise added separately in the three bands of the color images. The noise-free test set was one of 10 images with random shifts in the range ± 2.0 pixels. From this set, we created four others with different signal to noise ratio (SNR). Noisy reference images with corresponding amounts of noise were also created. Table 8 shows the results for color and greyscale registration.

4. FUSION OF THE DATA FROM LOW-RESOLUTION FRAMES

We are given N low-resolution frames, of dimension $M \times M$, from the video sequence. From the registration step, we get N pairs of deformation polynomials P_i and Q_i which define the geometrical warp between the i^{th} low-resolution image and the reference image. We want to compute for each pixel of the low-resolution image the intensity surface on a finer grid using the registered positions of the $N \times M^2$ low-resolution pixels.

We consider the pixel (k, l) , $1 \leq k \leq M, 1 \leq l \leq M$ of the low-resolution grid. Let n be the number of the $N \times M^2$ low-resolution pixels which registered positions verifying:

$$\begin{aligned} k < x \leq k + 1 \\ l < y \leq l + 1 \end{aligned} \tag{10}$$

Provided there are at least 7 pixels (with non-colinear positions) in this area, we can compute the parameters of a thin plate spline (TPS) that interpolates the intensity surface of the pixel between these n irregularly sampled

Table 8: Influence of Gaussian noise on accuracy of registration

SNR	Color images		Grey-scale images	
	Accuracy	Misregistration	Accuracy	Misregistration
<i>no noise</i>	0.062	0/10	0.085	2/10
5	0.134	0/10	0.198	3/10
4	0.153	0/10	0.220	3/10
3	0.202	0/10	0.350	2/10
2	0.423	1/10	0.384	3/10

Figure 1: LR (128×128) Lena image used as reference image for registration.



data values. A TPS is a smooth function (i.e., with continuous first partial derivatives) which can be defined, given the n points (x_i, y_i) , $1 \leq i \leq n$ in the plane, as:

$$f(x, y) = a_0 + a_1x + a_2y + \frac{1}{N} \sum_{i=0}^{n-1} b_i r_i^2 \log r_i^2 \quad (11)$$

where $r_i^2 = (x - x_i)^2 + (y - y_i)^2$ and with the constraints:

$$\sum_{i=0}^{n-1} b_i = \sum_{i=0}^{n-1} b_i x_i = \sum_{i=0}^{n-1} b_i y_i = 0 \quad (12)$$

Using the TPS interpolation, we compute for this pixel the intensity surface on a finer grid. This process is carried out for every pixel of the low-resolution image leading to an over-sampled image of dimension $L \times L$ where $L > M$ (e.g., $L = 10 M$). Then assuming a Rect PSF (rectangular point spread function) of the size of a low-resolution pixel for the imaging system, we deconvolve this over-sampled image by the PSF using the Maximum Likelihood method. This step aims to correct for the finite detector size which is the primary contributor to the imaging system PSF. The resulting deconvolved image is then down-sampled to the desired super-resolution size (typically of dimension $2M \times 2M$).

For the SR experiments, we used a test set of 25, 128×128 Lena images with random shifts along x and y axis in the range ± 2 pixels. Figure 1 shows the reference LR image, while Figure 2 shows the final SR image and, for visual comparison, Figure 3 represents the original 256×256 Lena image.

5. CONCLUSION

In this article, the super-resolution problem for color video frame sequences has been addressed using an extension to color images of the multi-resolution registration algorithm described by Djamdji and Bijaoui². The algorithm makes use of the information available in the three bands to improve both robustness and accuracy compared to the greyscale registration algorithm. The results of tests on the performance of the registration algorithm were presented. They demonstrate the sensitivity of the algorithm to the relevant parameters of the method as well as to the general conditions of use (dimension of the images, level of noise). The accuracy measured at the registration stage provides reliable data for deciding to proceed further with the fusion of the data in the super-resolution process.

Figure 2: Super-resolved (256×256) Lena image obtained by the algorithm.



Figure 3: Original (256×256) Lena image for comparison.



REFERENCES

1. M. Elad and A. Feuer, "Superresolution restoration of an image sequence: Adaptive filtering approach," *IEEE Trans on Image Processing* **8**, pp. 412–422, March 1999.
2. J. Djamdji, A. Bijaoui, and R. Maniere, "Geometrical registration of images: the multiresolution approach," *Photogrammetric Engineering and Remote Sensing* **a**(no. 59), pp. 645–653, 1993.
3. H. Ur and D. Gross, "Improved resolution from subpixel shifted pictures," *CVGIP: Graphical Models and Image Processing* **54**, pp. 181–186, March 1992.
4. H. Stark and P. Oskoui, "High resolution image recovery from image-plane arrays, using convex projections," *J. of Opt. Soc. Am.* **6**(11), pp. 1715–1726, 1989.
5. A. Tekalp, M. Ozkan, and M. Sezan, "High-resolution image reconstruction from lower-resolution images sequences and space-varying image restoration," *IEEE Int. Conf Acoust, Speech and Signal Proc.* **III**, pp. 169–172, March 1992.
6. M. Irani and S. Peleg, "Improving resolution by image registration," *Computer Vision, Graphics, and Image Processing* **53**(No. 3), pp. 231–239, 1991.
7. M. Irani and S. Peleg, "Motion analysis for image enhancement: Resolution, occlusion and transparency," *J. of Visual Comm. and Image Representation* **4**, pp. 324–335, December 1993.
8. P. Cheeseman, B. Kanefsky, and R. Hanson, "Super resolved surface reconstruction from multiple images," tech. rep., NASA Technical Report, January 1993.
9. R. Schultz and R. Stevenson, "Improved definition video frame enhancement," *IEEE Int. Conf. Acoust. Speech and Signal Proc.* , pp. 2169–2172, May 1995.
10. B. L., "A survey of image registrations techniques," *ACM Computing Surveys* **24**(4), pp. 315–376, 1992.
11. M. Holdschneider, R. Konland-Martinet, J. Morlet, and P. Tchamitchan, *Wavelets, Time-Frequency Methods and Phase Space*, Springer-Verlag, 1989.
12. J. Nunez, X. Otazu, O. Fors, and A. Prades, "Image fusion with additive multiresolution wavelet decomposition. Application to SPOT and Landsat images," *Journal of the Optical Society of America* **16**, pp. 467–474, March 1999.
13. J. Starck, F. Murtagh, and A. Bijaoui, *Image Processing And Data Analysis: The Multiscale Approach*, Cambridge University Press, 1998.

APPENDIX A. FLOW-CHART OF COLOR REGISTRATION ALGORITHM

