# Image Sequence Enhancement Using Sub-pixel Displacements

Danny Keren, Shmuel Peleg, Rafi Brada

Dept. of computer science
The Hebrew University of Jerusalem
91904 Jerusalem, Israel

## Abstract

Given a sequence of images taken from a moving camera, they are registered with sub-pixel accuracy in respect to translation and rotation. The sub-pixel registration enables image enhancement in respect to improved resolution and noise cleaning. Both the registration and the enhancement procedures are described. The methods are particulary useful for image sequences taken from an aircraft or satellite where images in a sequence differ mostly by translation and rotation. In these cases the process results in images that are stable, clean, and sharp.

## I. Introduction

This paper is an extension of [P.K.S], where it was shown that given a number of low-resolution images of a scene and their exact sub-pixel displacements an image of higher resolution can be obtained. The problem was treated elsewhere ([Hua1], [G]), mostly using the Fourier transform. We, however, do not use the Fourier transform at any stage as it proved to be sensitive to noise in our noisy enviroment.

This resolution improvement can not, of course, bring back frequencies that have been lost in the coarse sampling. But there are many spatial features like edges or lines that could be located more accurately. Figure 1 shows in one dimension how two coarse samplings with sub-pixel displacements could yield a better resolution estimation. This indicates that the process should be performed in the spatial domain rather then the frequency domain.

Section 2 describes the methods used for exact image registration. It should be noted that methods using the frequency domain ([Hua2], [Ca]) were found to be robust but not accurate enough for our sub-pixel purposes. Section 3 describes the methods used for improving the resolution and the reduction of noise.

Special mention is given here to the *interlace*, a feature in most video systems, where the odd lines are scanned first followed by the even scan lines. Due to this anti-flickering technique every digitized image is actually composed of two images: the odd-lines image and the even-lines image taken 1/50 of a second apart. With a moving camera this can represent substantial spatial distances, and therefore such a digital image cannot be treated as one image. Figure 2 shows the interlace problem in a picture taken from a moving camera. The simplest method to overcome this problem is to treat each frame as two separate images. These images will have in the x-direction double the sampling density than in the y-direction. This method is satisfactory when the camera has no - or little - acceleration during the frame. Figure 2 also shows the result of separating an image into two: the odd and even scan lines, followed by recombining these two images using the methods developed in the following sections.

**This work has been supported by a grant from AFIRST and the Israeli NCRD**

## II. Sub-pixel Registration

Image registration is an extensively treated subject ([Hua2], [Ca]). Most methods, however, are not accurate enough for our sub-pixel accuracy. The following image registration has been found to be the most accurate and robust for our purposes.

We look at the frames as two functions $f$ and $g$, where the following relation holds for the horizontal shift $a$, the vertical shift $b$ and the rotation angle around the origin $\theta$:

$$g(x,y) = f(x\cos(\theta) - y\sin(\theta) + a \ , \ y\cos(\theta) + x\sin(\theta) + b) \quad (1)$$

If we expand $\sin(\theta)$ and $\cos(\theta)$ to the first two terms in their taylor series we will get -

$$g(x,y) \approx f(x+a-y\theta-x\theta^2/2 \ , \ y+b+x\theta-y\theta^2/2) \quad (2)$$

Expanding $f$ to the first term of its own taylor series gives the following first order equation -

$$g(x,y) \approx f(x,y) + (a - y\theta - x\theta^2/2)\frac{\partial f}{\partial x} \quad (3)$$

$$+(b + x\theta - y\theta^2/2)\frac{\partial f}{\partial y}$$

The error function between $g$ and $f$ after rotation by $\theta$ and translation by $a$ and $b$ can then be approximated by -

$$E(a,b,\theta) = \sum[f(x,y) + (a - y\theta - x\theta^2/2)\frac{\partial f}{\partial x} + \quad (4)$$

$$(b + x\theta - y\theta^2/2)\frac{\partial f}{\partial y} - g(x,y)]^2$$

where the summation is over the overlapping part of $f$ and $g$ (actually we need a much smaller part).

If we look for the minimum of $E(a, b, \theta)$ by computing its derivatives by $a$, $b$ and $\theta$ and comparing them to zero, then after neglecting the non-linear terms and some small coefficients we get the following system of linear equations, where $R$ is an abbrevation for $x\frac{\partial f}{\partial y} - y\frac{\partial f}{\partial x}$ and the summation is over the overlapping area:

$$[\sum(\frac{\partial f}{\partial x})^2]a + [\sum\frac{\partial f}{\partial x}\frac{\partial f}{\partial y}]b + [\sum R\frac{\partial f}{\partial x}]\theta =$$

$$\sum\frac{\partial f}{\partial x}(f - g)$$

$$[\sum\frac{\partial f}{\partial x}\frac{\partial f}{\partial y}]a + [\sum(\frac{\partial f}{\partial y})^2]b + [\sum R\frac{\partial f}{\partial y}]\theta = \quad (5)$$

$$\sum\frac{\partial f}{\partial y}(f - g)$$

$$[\sum R\frac{\partial f}{\partial x}]a + [\sum R\frac{\partial f}{\partial y}]b + [\sum R^2]\theta =$$

$$\sum R(f - g)$$

Due to the approximations made when obtaining (3), the expression is correct only for small values of ($a$, $b$, $\theta$). Therefore we perform the following iterative process - solve the equations, "push" $g$ (using formula (1)) with the solutions obtained - to sub-pixel accuracy (i.e in this "pushing" process we do sub-pixel interpolation) and continue with the new $g$, either a fixed number of iterations or until the solutions are very small. In order to keep accuracy we always "push" the original $g$ by the accumulated values of $a$, $b$, and $\theta$.

It should be noted that we need to compute nine of the twelve equation parameters only once - we need to change only the scalar part where $g$ occurs, because $f$ is always the same. This saves a lot of time both in the iterations and further if we wish to compute the motion parameters of many different $g$'s relative to the same $f$.

In order to increase speed and robustness we use a coarse-to-fine structure of the image, called a gaussian pyramid ([R]). In this scheme, the original image of size $N \times N$ is filtered by a gaussian and sub-sampled to give an image of size $N/2 \times N/2$. This process is repeated until an image of one pixel is reached. We first compute the motion parameters for a small image (usually 64 x 64). Even big translations are small at this reduction level. We then interpolate the found parameters into the larger image, correct this guess by one or two iterations, and interpolate again to the next resolution. This process continues until the original images are reached. The complexity of the entire process is like computing two iterations on the original images.

It proved useful to run some kind of smoothing on $f$ and $g$ before applying the algorithm. This lowers the high derivatives and thus the functions are closer to the taylor expansions.

Running the process over various real and computer simulated motion frames showed good results. With the simulated motion, where we know the real parameters exactly, we got errors of about 0.03 pixels in the x and y shifts and 0.0005 radians ($0.03^0$) in the rotation angle. The method is also very immune to noise. However, it is good only for small rotations - usually not more than 0.3 radians ($6^0$). Since, however, we are discussing video frames taken 50 a second this is enough for many practical purposes. Feature-based matching methods ([L.D]) can be used to bring the images to better angular alignment for larger rotations.

Given the sequence of images with their exact registrations, we can proceed in the enhancement task which includes noise reduction and improving the resolution.

# III. Image Improvement

## A. Noise Reduction

Given k frames $\{P^k\}$ of size $N \times N$ each with sub-pixel registration, we create a combined image of size $\lambda N \times \lambda N$ , $\lambda \geq 1$. We will describe the process for $\lambda = 2$, but as the accuracy of our registration increases $\lambda$ can increase, and the results still be significant.

Let $Q$ of size $2N \times 2N$ be the combined image. Its computation can be visualized as follows: pile the images $\{P^k\}$ on top of each other, registered accurately using their sub-pixel displacement. Each pixel in these images will represent an area of $d \times d$. On top of these images put $Q$, whose pixels represent an area of $d/2 \times d/2$ each. Given a pixel $q \in Q$, let $p^k$ be the pixel in $P^k$ whose area includes the center of $q$. It can be viewed like pushing a needle down through the center of $q$, and $p^k$ is the pixel in $P^k$ which is pierced by the needle. The value of $q$ is then computed from the set of $p^k$'s.

Given the set of $p^k$'s, the set of pixels from the original images that correspond to $q$, the grey level of $q$ is computed in two stages. First, the $p^k$'s with extreme values are eliminated, and then the remaining pixels are averaged. Figure 3 shows the effect of such a process for combining 32 original images of size 200 × 50 to an image of size 400 × 200, where each point in $Q$ is the average of the 14 central values of the corresponding $p^k$'s after the elimination of the highest 9 and lowest 9 values. Elimination of extreme values treats well non-linear noise, while the averaging eliminates linear noise.

## B. Resolution Improvement

It is not hard to follow the imaging process and see that the images obtained by combining several low-resolution images after registration can be viewed as a gaussian blurred version of the original image. This calls for deblurring ([S.P], [Huml], [Hum2]), and by using the high-pass filtering suggested in [Hum2] we can get substantial improvement. The effect of this high-pass filtering is shown in Figure 4. on a sequence of very noisy images.

Deblurring the combined image is simple and effective, but does not use all the information available from the original images and their known displacements. A model of the imaging and restoration process is mentioned in [P.K.S], and will be described in section 3.3.

## C. Optimization Process

We try to find a high-resolution image P such that the set of low-resolution images obtained from P by a simulated imaging process is closest to the given set of images - that is, the set $\{P^k\}$ of section 3.1. We start with an initial guess for the high-resolution image P, usually the combined image of section 3.1, obtained by registrating and averaging the $P^k$'s. Let's denote this guess by $P_0(x, y)$. Now we simulate the imaging process to get a set of low-resolution images, $\{S_0^k\}$, and we would like them to be as close as possible to the $P^k$'s. The error of the image $P_0(x, y)$ can thus be defined as

$$E_0 = \sum_k \sum_{(x,y)} | S_0^k(x, y) - P^k(x, y) | \qquad (6)$$

The high-resolution image will be the one minimizing the error in (6). A primitive minimization can be to examine each pixel in the guess $P_0$. If its current grey level is $P_0(x, y) = l$, we consider the three possibilities $\{l-1, l, l+1\}$ for the value of $P_0(x, y)$. For each of these values we compute the change in the $S_0^k$ images and the change of the error in expression (6). $P_0(x, y)$ is then assigned that value which results in the minimal error. The process is continued iteratively until no further improvement can be obtained in the error function, or until the maximum of allowed iterations is reached. Figure 5. shows the application of this approach to increase the resolution of an image by three.

# IV. Concluding Remarks

We have presented a robust image registration scheme, that can be used to stabilize and enhance noisy images in a sequence taken with a moving camera. The method works best when the camera has little or no acceleration, and when the projection is almost parallel. An immediate extension can be to any projective projection of three dimensional objects. The methods tested successfully on real noisy images.

# References


[P.K.S] - Peleg S., Keren D., Schweitzer L. (1987), *Improving image resolution using sub-pixel motion*, Pattern recognition letters 5, 223-226.

[Hua1] - Huang T.S and Tsai R.Y. (1984). *Multi-frame image restoration and registration*, Advances in computer vision and image processing, edited by Thomas S. Huang, vol. 1, 317-339.

[Hua2] - Huang T.S and Tsai R.Y. (1981), *Image sequence analysis: motion estimation*, Image sequence analysis, edited by Thomas S. Huang, 2-6.

[Ca] - De Castro E. and Morandi C. (1987), *Registration of translated and rotated images using finite Fourier transforms*, IEEE Trans. on PAMI, September 1987, 700-703.

[Huml] - Hummel R. and Moniot R. (1987), *Solving ill-conditioned problems by minimizing equation error*, Proceedings of first international conference on computer vision, June 8-11 London, 527-533.

[Hum2] - Hummel R. and Kimia B. (1986). *Deblurring gaussian blurr*, CVGP 1986, in galleys.

[G] - Gross D. (1986), *Super resolution from subpixel shifted pictures*, M.Sc thesis, Tel-Aviv university October 1986.

[R] - Rosenfeld A. (1984), *Multiresolution image processing and analysis*, Edited by A. Rosenfeld, Springer-Verlag 1984.

[L.D] - Lalita L. and Dutta Majumder D. (1987), *Image registration using pseudo control points on linear features*, Proceedings of the 5th Scandinavian conference on image analysis, Stockholm, June 1987, 335-343.

[S.P] - Shvaytser H. and Peleg S. (1987), *Inversion of picture operators*, Pattern recognition letters, January 1987, 49-61.

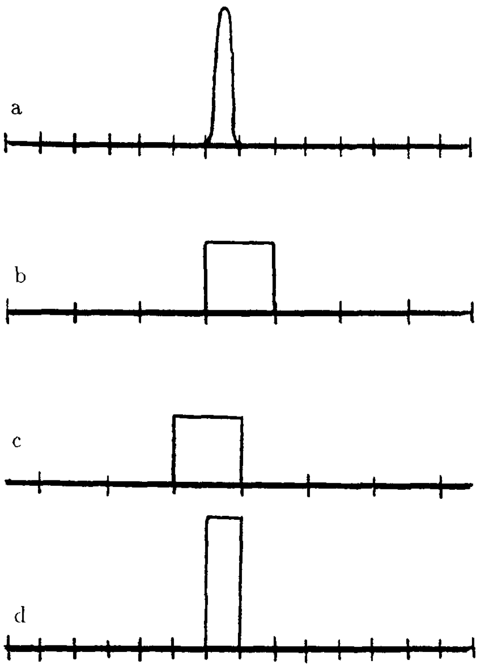



Figure 1: A simple example of resolution improvement.

(b) and (c) are two coarse samplings, where each sampling point is a simple average of a region of length $e$, and the points are distance $e$ apart. The sampling in (c) is shifted $e/2$ apart from the sampling in (b). (d) shows a higher resolution estimate to the original signal (a) that can be inferred from both (b) and (c), knowing their exact registration.

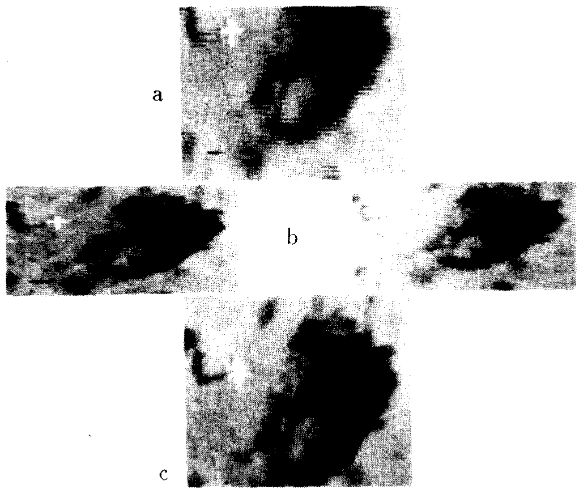



Figure 2: The interlace effect of images taken with a moving camera.

a) Original image.
b) Separating into odd and even line images.
c) Recombining the two images in (b) using the methods described in following sections.

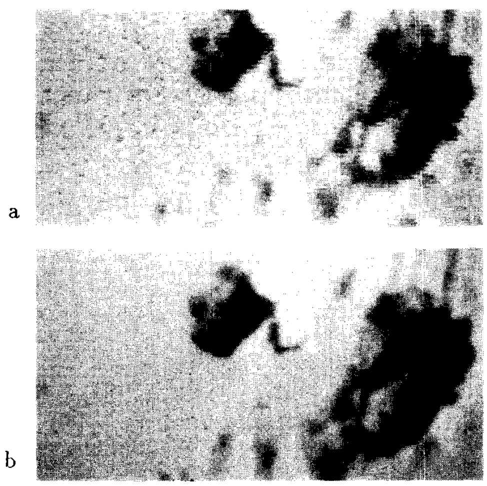



Figure 3: Combining a sequence after registration.

a) Sample image from sequence, size $200 \times 100$ (before interlace separation).
b) Combination of 32 images of size $200 \times 50$ (after interlace separation of 16 original images of size $200 \times 100$) giving an image of size $400 \times 200$. Each new pixel is computed from the average of 14 central values (see text).

a

b

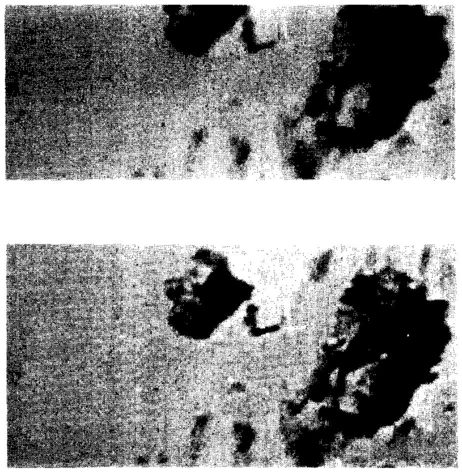
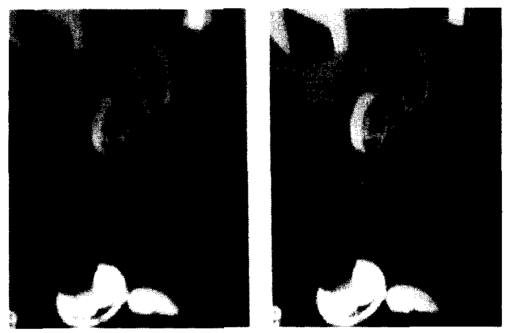


Figure 4: Sharpening using high-pass filtering.
a) Figure 3.b after high-pass filtering.
b) Preforming high-pass filtering before combining the images.

a



b c

Figure 5: Improved resolution by a factor of three using optimization.

a) One image from a sequence, size $70 \times 100$.
b) Combined image from a sequence of 15, size $210 \times 300$. Each pixel is the result of averaging the 11 central values.
c) Result after running several iterations of the optimization.

# SESSION 3.2 – MORPHOLOGY

**Chair**

**R. Haralick**
**University of Washington**