Multiple reference frames registration for super resolution

Kuybeda Oleg
Supervisors:
Dr. Meir Bar-Zohar, RAFAEL
Meir Shaked, TECHNION

Sponsored by RAFAEL,
Image processing department

Abstract

A registration of images deals with finding geometrical warp parameters between different images that contain some common intrinsic details. The registration plays one of the most important roles in the super-resolution reconstruction since the latter deals with recovery of one high-resolution image from warped, blurred and decimated replicas of it. Several algorithms were already proposed for the general solution of the image registration. In this paper we concentrate on a special case, where warps are pure translations. We use previous basic ideas and results to develop a new registration technique, which uses multiple reference frames, works with hardly aliased images and brings high subpixel precisions. This technique is intended to increase significantly the resolution-refining factor[*] in super-resolution reconstruction problems.

Introduction

As a consequence to contemporary growing demand to higher precision of image interpolation techniques, super-resolution reconstruction becomes apparently better investigated. A common sense of the interpolation means compensating missing data between fixed lattice nodes i.e. between the pixels, based on the existing image pixel values. Contrarily, super-resolution reconstruction means compensating this data by using additional source – such as another differently captured image or a set of *n* images. Only aliased images could be used for super resolution, because images, sampled according to Nyquist criteria, are completely self-consistent. Thus, for example, having a video sequence of 15 frames with a given resolution, one can construct a single frame with better resolution say as twice.

A typical super resolution reconstruction algorithm consists of two topics:

1) *Registration between given images*
2) *Super-resolution restoration*

Many approaches are developed regarding the second topic. ML, MAP & POCS based de-blurring techniques are used in the recent works [3,4,5,6]. LMS algorithm using frequency plane representation is described in [1]. Back projection algorithm is in [2]. However there are few advances regarding the registration of aliased low-resolution images. On the other hand, super-resolution registration precision is critical, because the estimation of the warp parameters is performed over low-resolution grid, and each registration error is magnified by resolution refining factor over the high-resolution grid. The situation is worsened further by strongly aliased nature of the measured data. Some of works [2,4,5] apply regular registration techniques to the super-resolution problem, which uses single reference low-resolution frame. This approach is taken as a generic registration solution and doesn't exploit additional super-resolution related features. Many other works assume apriory knowledge of the warp parameters without estimating them. As a result, the best practical resolution improvement is 2-3 times in each axis.

The new registration technique, presented in this work, enables to obtain resolution improvement by 5 –10 times in each axis. This technique is based on multiple reference frames, high-resolution autocorrelation function estimation and takes into consideration also the strong aliasing nature of the measured images. The primary accent was applied to the de-aliasing process, while other restoration aspects are left to the future expansion. During current research, we used synthetically generated data by random continuous displacements, with a weak additive gaussian white noise & without blur.

The organization of the rest paper as follows:
Section 1 is a description of the general model of the image generation process. Section 2 deals with special case reductions of the general model, used for the algorithm development. Section 3 presents basic problems of the single reference frame registration. Section 4 proposes the new registration algorithm. Section 5 presents simulative results.

[*] Feasibly recoverable decimation of the images

1 Imaging process – general description

This section presents the model of digital image acquisition. Let's denote $I_1..I_n$ to be the measured images ordered as column vectors. Then the imaging process for measuring an image I_k can be represented as:

$$I_k = D_k \cdot H_k \cdot G_k \cdot I^s + N_k \,,\; k = 1..n \quad (1)$$

The vector I^s is unknown high resolution image. Matrix G_k is spatial warp operation on I^s during the process of obtaining the image I_k. The matrix H_k represents blurring operation, originated from integration on finite size detectors. The matrix D_k , which is not square, stands for decimation operation to obtain I_k, which occurs due to a finite distance between detectors. For dealing with super resolution reconstruction, these distances have to be larger, than the desirable high-resolution pixel size. The vector N_k represents the gaussian uncorrelated noise, added to the low-resolution image during sampling process.

2 Special case simplifications

In this section we are going to impose some modifications to the model of the imaging process. For most cases, H_k and D_k are equal for all k = 1..n. We can treat H_k as LSI operator, due to uniformly symmetrical structure of the majority of CCD's. As long as in our case, G_k represents only translations, it is a linear space invariant operator (LSI), which commutes* with other LSI operator such as H_k. So we can rephrase the expression (1) as:

$$I_k = D_k \cdot G_k \cdot H_k \cdot I^s + N_k \quad (2)$$

Since the blurring and decimation are invariant between images, we can denote:

$$H = H_k \,;\; D = D_k \quad (3)$$

The noise isn't correlated neither spatially, nor between measured images $I_1..I_n$, so the model can be written as:

$$I_k = D \cdot G_k \cdot (H \cdot I^s + N^h) \quad (4)$$

Where N^h is some noise image, added to original high-resolution source and obeying conditions given for N_k. Let's denote $I^h = H \cdot I^s + N^h$, then:

$$I_k = D \cdot G_k \cdot I^h \quad (5)$$

Our goal now is to restore I_h – the blurred and noised version of the high-resolution image. Then the whole problem reduces to dealiasing process, leaving the rest restoration of I^s from I^h as a general image restoration topic, which can be treated by other classical methods, such as ML, MAP, LMS, POC used in the works referenced above.

The restoration of I^h , given the set of $I_1..I_n$ images and $G_1..G_n$, as known data, is straightforward. The simplest heuristic way is to place the images $I_1..I_n$, inversely, on high-resolution grid.

The resulting high-resolution image Ih comes out to be over non-uniform grid, because of irregularity of the spatial transformations G1..Gn. Therefore, uniform interpolation over non-uniform grid could be used, for uniform representation of the result.

3 Super-resolution registration problems

However, $G_1..G_n$ are unknown. The goal of registration is to estimate the spatial transformations from the given set of measured low-resolution images $I_1 .. I_n$.

3.1 Conventional method

The most common way to estimate geometrical warp is defined as following:

1. One image from the set $I_1..I_n$ is chosen to be the reference frame. Let's denote it I_r.
2. For each of the remained n-1 images, lets say I_k, we find a spatial warp G_k', the inverse of which, produces the best match between the image I_k and the reference frame I_r .
3. We convert the low-resolution grid related $G_2'..G_n'$ warps, to the corresponding $G_2..G_n$, defined over high-resolution grid (in case of displacements only, the converting is simply multiplication of each axis displacement parameter, by a corresponding decimation factor).

* The case of rotations could also be treated, but then G_k is not LSI. Therefore, working with cyclically symmetrical kernel of blurring in H_k, enables us to interchange H_k and G_k in this case too.

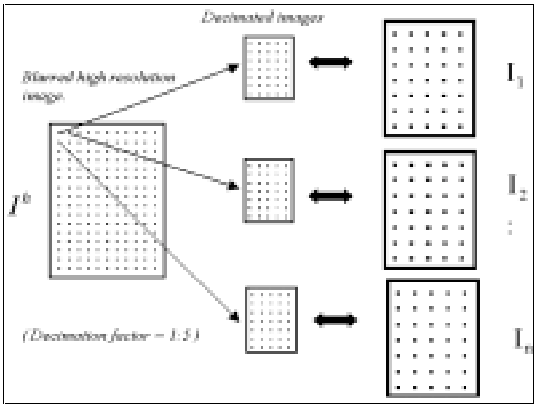

Figure 1 – small translations and decimation

Let's call the method of single reference frame described above, as conventional method. Conventional method brings relatively satisfying results, when working with low decimations - of order smaller, than 3. That means, that the best resolution improvement in ideal conditions (no blur, no noise) could be 3 times in each axis.

3.2 Insight to Conventional Method (CM) limitations

- *Why cannot we go further in decimation?*
- *Why cannot we improve the resolution 10 times in each axis, having big enough database of measured images?*

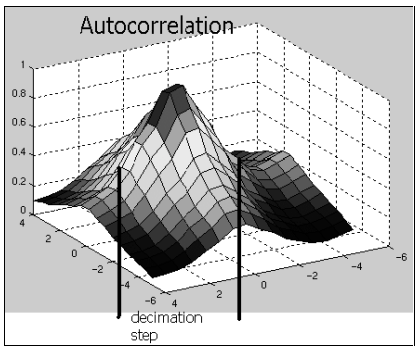

Figure 2 – an autocorrelation function of a typical high-resolution image

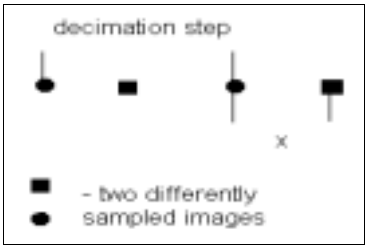

Figure 3 – heavy decimation process

Working with high resolution grid demands very high precision of registration between low resolution measured images I_1 .. I_k. So the super-resolution demands the registration error in low-resolution scale to be of order of $\frac{1}{d}$, where d is the decimation factor.

Working with heavily aliased images, introduces additional difficulties to the registration problem, as follows.

General images have strong correlation only in a small neighborhood of points. Let's consider an autocorrelation function of the high-resolution image. For most images, this function has a shape close to a gaussian, with $\sigma \approx 3..5$ pixels (see Figure 2). If we decimate such image with step $d = 2 \cdot \sigma$ for example, the autocorrelation function of the resulting images will be much narrower ($\sigma < 1$ pixel). That means there is almost no correlation between two neighboring points in that image.

Moreover, consider two images that were sampled as sown in Figure 3. The first image was sampled with decimated factor d. The second was sampled with the same decimation factor, but at a slightly different position. Let's denote the displacement value by x. So for images, where $x \approx \sigma < d$ & $d - x > \sigma$, it easily seen, that the maximum cross-correlation value that can be attained in the registration process is 0.5! In another words, there is no good optimal matching point between two heavily decimated images. And the conventional method has very low confidence value* in its results, which makes it unstable.

The visual sense of the described effect is shown in Figure 4. Six images were sampled at different positions with x as described above. You can see that they are locally different. And it is very difficult (if possible at all) to talk about subpixel registration precisions.

* Confidence value refers to the correlation value at the best matching position

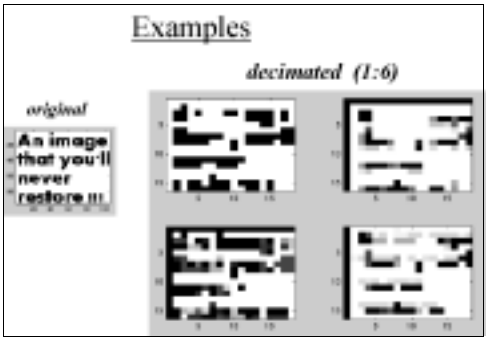

Figure 4 – small translations & decimation

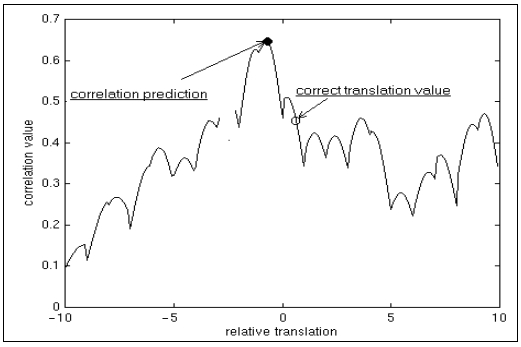

Figure 5 – deceptive behavior of cross-correlation between two heavily decimated images

Figure 5 shows deceptive behavior of cross-correlation of the first two frames from Figure 4. The maximum of the cross-correlation (matching) occurs at the wrong position. The graph is shown in high-resolution scale. MSE related algorithms behave the same way.

3.3 Conclusions and proposals

According to said above, we can roughly define decimation factor limit per each high-resolution image above which, the conventional registration method fails. Roughly it is about $2\cdot\sigma$. This makes all neighbor points of the resulting image uncorrelated. Actually, due to other image distortions like blur and noise, this factor is even smaller.

However, the proposed algorithm shows, that there is still possibility to register such heavily aliased images. The main idea is instead of using single reference low-resolution frame, to use the whole set of measured images, combined in one, iteratively pre-estimated high-resolution reference frame.

4 Registration algorithm

4.1 Assumptions

The proposed algorithm, given below, assumes some prior knowledge about measured images relationship, as following:

1. We know the autocorrelation function of the I^h (high-resolution blurred image). From the first sight this assumption seems to make no sense, since I^h is unknown. However, in **appendix A**, with aid of assumption 4, it is shown how we can estimate the mentioned autocorrelation function, using the given $I_1..I_n$.
2. The autocorrelation function of I_h is cyclically symmetrical.
3. The autocorrelation function decreases monotonically in the region, close to origin.
4. The cross-correlation between two high-resolution images approximately equals to cross correlation between their corresponding decimated replicas.
   So if

   $I_1=G_1\cdot I^h$; $I_1' = D\cdot G_1\cdot I^h$
   $I_2=G_2\cdot I^h$; $I_2' = D\cdot G_2\cdot I^h$, *then*

   $corr(I_1,I_2) \approx corr(I_1',I_2')$

   The validity of this assumption is shown in **appendix B**. The meaning of such statement is that decimation doesn't produce primary impact on the correlation value, whereas the correlation is strongly influenced by the displacement value of the original high-resolution image.
5. Let's denote *dx* and *dy* to be the displacement value in x and y directions respectively. The distribution of the displacements is expected to be as following:

   $dx \sim U[-d_x;d_x]$; $dy \sim U[-d_y;d_y]$,
   where d_x *and* d_y *are decimation factors in each axis*

   Actually, the displacements could be distributed over larger limits, but then, the conventional method is capable to compensate them by bringing all the images to the common frame with the it's precision of about $\pm d$.

As a result of these assumptions, we can significantly reduce the region of registration

search and avoid many deceptive extreme points. Thus, given two measured images – say I_1 and I_2, we can approximately determine the distance between them in the high-resolution scale. For example, if the autocorrelation function of high-resolution image I^h is as shown in Figure 6 and $corr(I_1,I_2) = 0.5$, then the distance between I_1 and I_2 is *2* pixels in some direction, i.e. : $\sqrt{dx^2 + dy^2} = 2$. So the region of possible translations between these two images is reduced to a circle with radius of *2*. It is advisable to add some safety margins to the circle and to work with ring of finite width, but to be careful - enlarging the region of search raises chances of undesirable artifacts, as shown in Figure 5.

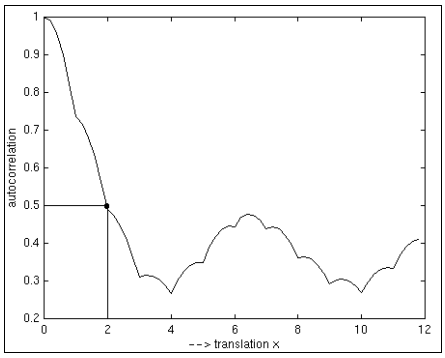

Figure 6 – an example of autocorrelation function

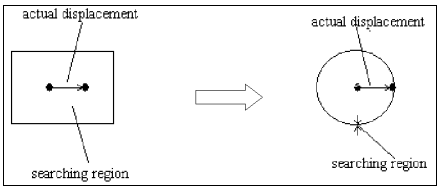

Figure 7 – reducing region of search

4.2 Registration algorithm – tools definition

In this section we show the proposed algorithm for registration process, which uses single, iteratively estimated high-resolution reference frame.

Let's start from a simulative experiment. Suppose we have some low-resolution image I_1. And let us produce four additional images by translating I_1 one pixel position in each of four directions respectively. Let's call them $I_{1.1}..I_{1.4}$. Each such pixel position corresponds to a distance of d in the high resolution. If we pick up another frame say I_2, so each of these four images gives us a separate circle. The radius of such circle is defined by a cross-correlation between corresponding $I_{1.i}$, $i=1..4$ and I_2, whereas the center - by the corresponding translation vector. The intersection point of all of these circles is the desired registration parameter between the I_1 and I_2. Figure 8 shows these circles, when $I_2 = I_1$. As expected, we get the intersection at the origin, since there is no movement between the I_1 and itself. Figure 9 shows the situation, when $I_1 \neq I_2$.

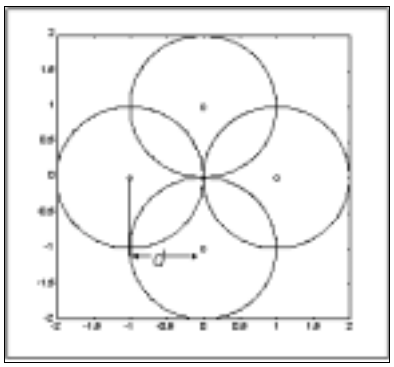

Figure 8 – I1 = I2

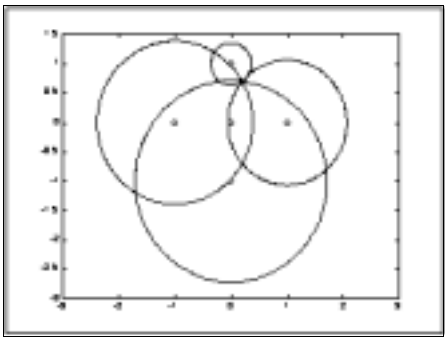

Figure 9 - I1 ≠ I2

4.3 Algorithm steps

1. Given a set of measured low-resolution images $I_1..I_n$, we choose one, say I_1 to be the reference frame for the first step.

- For each of the rest n-1 images, say I_k, we find intersection points of 5 circles - 4 as described in the experiment above and 1 with center at the origin. Each intersection point is a candidate to be the displacement G_k of I_k respectively to I_1. The coordinates are converted to the high-resolution scale.
- We choose one point with best correlation*, say G_k, which corresponds to I_k.
- Among all such n-1 results (as the number of remained frames), we choose one – say I_2

* Since the coordinates of examined intersection points are fractional, we use interpolation of I_1 for getting the correlations at those points.

with G_2, which is the best according to its correlation at that position.

2. Now we have two reference images – I_1 and I_2.
   - For each of the rest n-2 images, we have now 10 circles, produced by the both reference frames. We find all intersection points of these circles.
   - We choose one point with best correlation parameter. Although we have two reference frames, for getting correlation value, we use only one of them alone, chosen randomly, in order to avoid divergence of the algorithm due to consistent erroneous feedback.*
   - Among n-2 results we choose the best one, as in previous step.

3. Repeat step 2 with correct number of reference frames, until all displacements for all images are found.

As more circles created in the consequent steps, many more points we have to examine. Therefore, clusterizing is used for reducing the number of candidate points. Only the centers of clusters are examined. Small clusters are ignored. Figure 10 shows typical clusterizing results. This figure shows also the displacement found versus the originally generated displacement for a typical frame. It is possible to choose more than one frame with its displacement per step if their matching quality is the same or close enough.

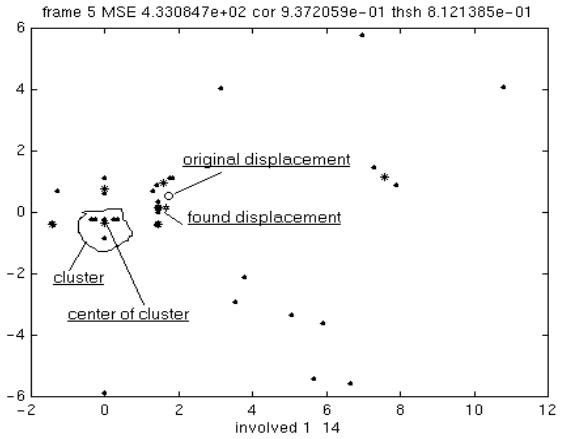

Figure 10 - clusterization of the intersection points

* Using all reference frames for calculating correlation parameters is problematic, because the errors would affect the results consistently from frame to frame, which produces the parasitic feedback.

4.4 The final stage of algorithm

In the previous steps, the resulting displacement vectors $G_1..G_2$ for each of measured images $I_1..I_2$, were determined pointwise. At this stage, we allow them to move in some rectangular environment, but over the whole combined pre-estimated high-resolution image I^h. Indeed, once we have estimated all displacements for $I_1..I_n$, we can build I^h over non-uniform grid. Now, for each of the $I_1..I_n$, we perform next steps:

- Let's say the mentioned image is I_k. We construct I^h using contribution of $I_1..I_{k-1}$ and $I_{k+1}..I_n$.
- For I_k, we refine its displacement vector G_k within a small rectangle with side length of $\frac{d}{2}$ around its previously estimated version. The match criterion is correlation of I_k at position G_k^{-1} over high-resolution image I^h. For each pixel in I_k, we find corresponding pixel in I^h by interpolation over its non-uniform grid. Note: I_k is low-resolution image, and the correlation is performed in low resolution.
- This new G_k will be counted in constructing I^h for refining the next G_{k+1}.
- The order of choosing G_k is according to matching criteria of that frame in the high-resolution image. Bad matching frames are refined first.

This process will run until low rate of refining corrections is reached. As was experimentally observed, steps 1-3 do the most of the work, whereas the final stage improves only extremely erroneous results for the $G_1..G_n$.

5 Results

Figures 11, 12 show the results of text and Lena images respectively. The text image source (a) was decimated 1:6. And 36 uniformly displaced images, some of them are shown in (b), were used to restore the original image. You can see the ideal restoration (d), which was obtained with the exact displacements. It is the best theoretically obtainable result. Due to random displacements and interpolations, this result differs from the source image. In (e) and (f) you can see the conventional method and the proposed algorithm results respectively. It should be noticed, that the proposed algorithm

brings significantly improved results, than conventional method, only in case of heavy aliasing. The decimation factor for such aliasing is defined by the image nature. Thus, in the Text image, more improvement achieved, than in Lena with the same decimation factor.

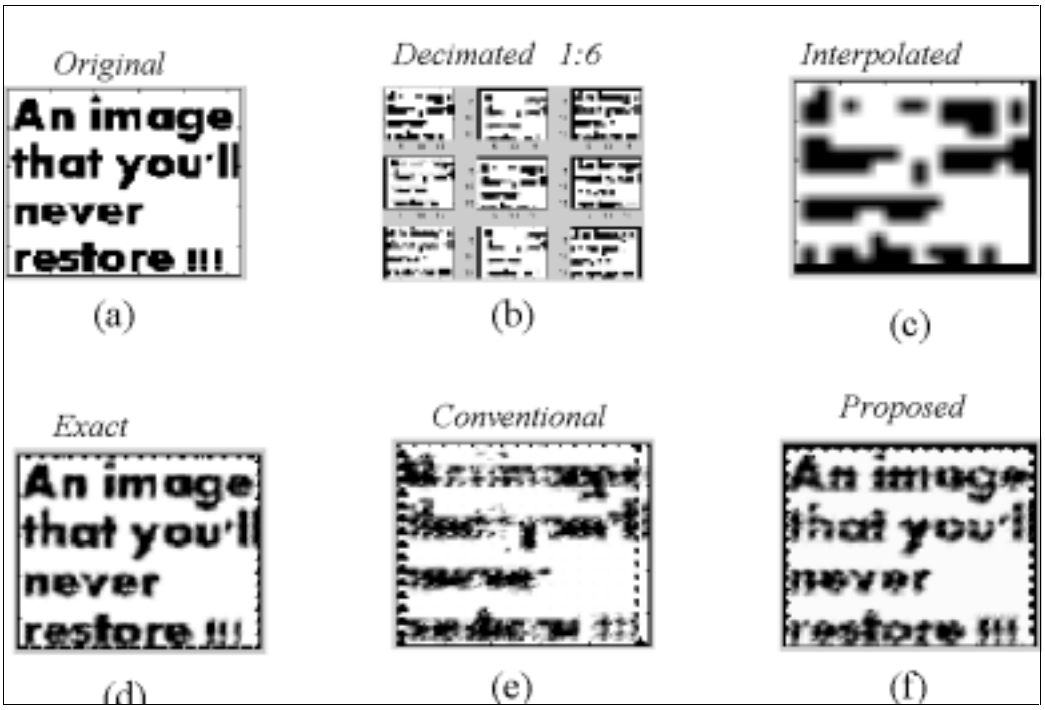

Figure 11

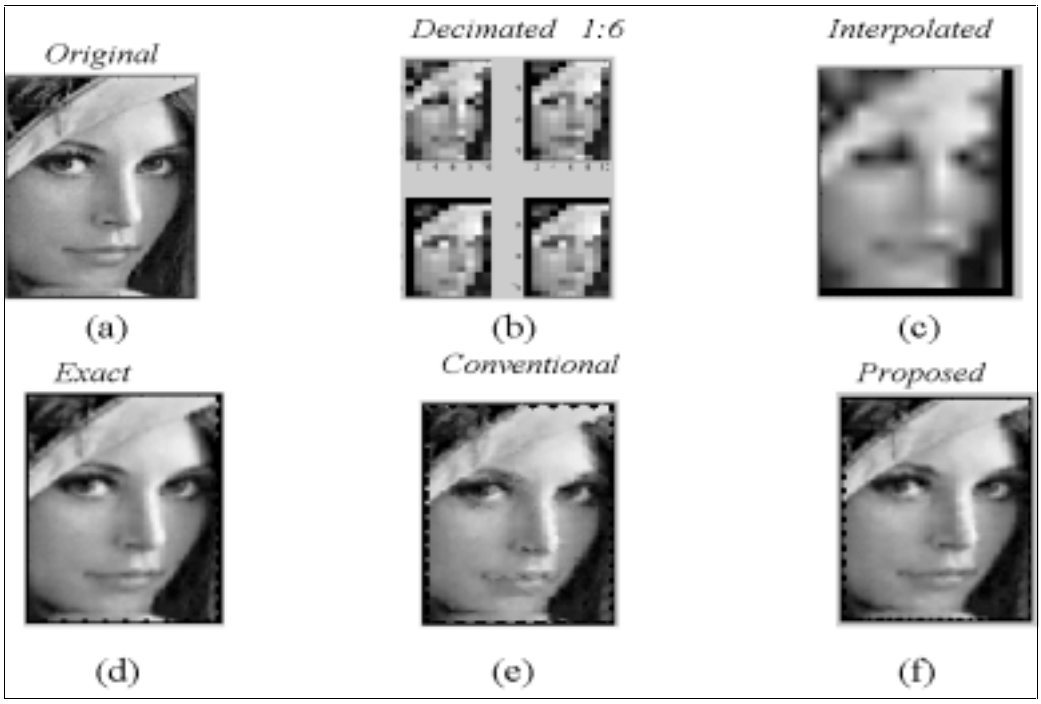

Figure 12

Appendix A

Estimation of high-resolution autocorrelation function, using measured images

Assumptions:

- All displacements $G_1..G_k$ are distributed uniformly on the decimation rectangle.
  Let G_k to be described by (dx_k, dy_k), and

  $dx_k \sim U[-d;d]$ $dy_k \sim U[-d;d]$

- The desired autocorrelation function is monotonic close enough to the origin - see Figure 13. Let's call it "reliable region".

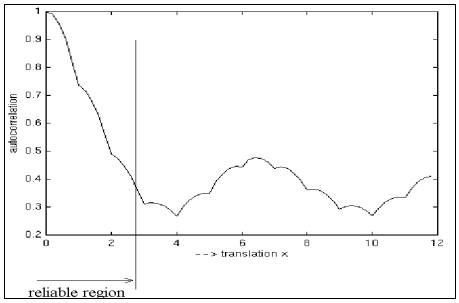

Figure 13 – the unknown autocorrelation function

Estimation steps:

Let's denote $r = \sqrt{dx_k^2 + dy_k^2}$

It's easy to show, that

$$f(r) = \begin{cases} \dfrac{\pi \cdot r}{2 \cdot d^2} & r < d \\ \dfrac{r}{4 \cdot d^2} \cdot \left(2 \cdot \pi - 8 \cdot \arccos\left(\dfrac{d}{r} \right) \right) & d < r < d \cdot \sqrt{2} \end{cases}$$

f(r) – density function of random variable *r*.

To estimate autocorrelation function we need to:

- choose one low resolution image – say I_1 to be the reference
- find correlation values between I_1 and all other images
- to distribute results according to *f(r)*, given above
- to interpolate the autocorrelation by a polynomial or any other smooth function

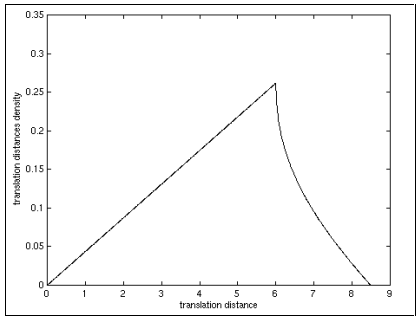

Figure 14 – density f(r)

The results of this process are shown in Figures 15 and 16. Only the region, which is close to the origin, is important. It is reliable region, were the correlation values are high enough to work with them and also the function behaves monotonically.

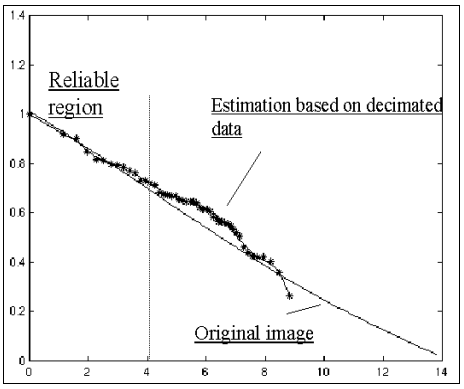

Figure 15 – autocorrelation estimation of Lena

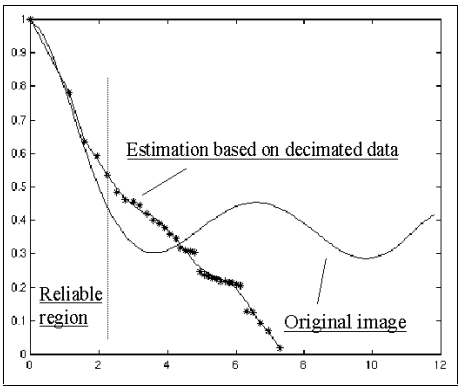

Figure 16 – autocorrelation estimation of Text image

Appendix B

Approximation of cross correlation between two high-resolution images by their low-resolution decimated replicas.

In this appendix we justify the following statement:

Let I_{h1}, I_{h2} be two images of the same size and I_{l1}, I_{l2} are their corresponding decimated replicas so that:

$I_{l1}=D\cdot I_{h1}$; $I_{l2}=D\cdot I_{h2}$ **, then**

$corr(I_{h1},I_{h2}) \approx corr(I_{l1},I_{l2})$

where *D* is a decimation operator, applied on the original images and "*corr(x,y)*" is cross-correlation between two corresponding images *x* & *y, defined as:*

$$corr(x,y)=\sqrt{\frac{\sum_i \left(x(i)-\bar{x}\right)\cdot\left(y(i)-\bar{y}\right)}{N\cdot\sqrt{\mathrm{var}(x)\cdot\mathrm{var}(y)}}} \quad \textbf{(1)}$$

$\bar{x}, \bar{y}$ - is the mean of x and y pixel values respectively.

Figure 1 spatially show the decimation process of the I_{h1} and I_{h2}. If not many small intrinsic image details are destroyed by the decimation, it is fair to assume, that:

$\overline{I_{l1}} \approx \overline{I_{h1}}$; $\overline{I_{l2}} \approx \overline{I_{h2}}$ and

$\mathrm{var}(I_{l1}) \approx \mathrm{var}(I_{h1})$; $\mathrm{var}(I_{l2}) \approx \mathrm{var}(I_{h2})$

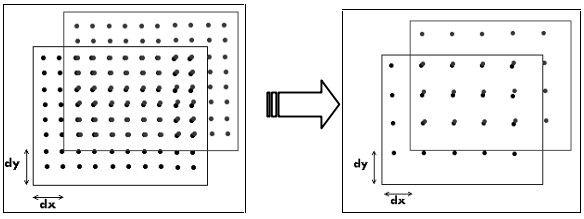

Figure 1

So, that:

$$corr(I_{l1},I_{l2}) \approx \sqrt{\frac{\sum_i \left(I_{h1}(d\cdot i)-\overline{I_{h1}}\right)\cdot\left(I_{h2}(d\cdot i)-\overline{I_{h2}}\right)}{\frac{N}{d}\cdot\sqrt{\mathrm{var}(I_{h1})\cdot\mathrm{var}(I_{h2})}}}$$

(2)

, where *d* is some decimation factor.

The difference between $corr(I_{l1},I_{l2})$ and $corr(I_{h1},I_{h2})$ is that the former includes a sum of less items, chosen with period of *d*, whereas the latter corresponds to the sum of all the items. Because of independence of the spatial choice of the items on their values, it is fairly could be assumed, that the result of the cross-correlation value isn't changed by reducing the number of samples.

The considerations, shown above do not represent a proof for the statement, but only speculations that could justify the approximation. This approximation was found valid in various simulations conducted on various images and decimations of $d = 5..10$

REFERENCES

1. *S. Kim, N. Bose and H. Valnzuella, "Recursive reconstruction of high resolution image from noisy undersampled multiframes" IEEE Trans. Acouss., Speech, Signal Processing, Vol 38, pp 1013 – 1027, June 1990*
2. *M. Irani and S. Peleg, "Improving Resolution by Image Registration" CVGIP Graphical Models AND Image Processing, Vol 53, NO. 3, pp 231 – 239, May 1991*
3. *Andre J. Patti, M. Ibrahim Sezan and A. Murat Tekalp "Superresolution Video Reconstruction with Arbutrary Sampling Lattices and NonzeroAperture Time" IEEE Trans., Image Processing, Vol. 6, NO 8, pp 1064 – 1076, August 1997*
4. *Russel C. Hardie, Kenneth J. Barnard, John G. Bognar, Edward A. Watson "High-resolution image reconstruction from a sequence of rotated and translated frames and its application to an infrared imaging system", Opt. Eng. 37(1), pp 247-260, January 1998*
5. *Russel C. Hardie, Kenneth J. Barnard and Ernest E. Armtromg, "Joint MAP Registrstion and High-Resolution Image Estimation Using a Sequence of Undersampled Images" IEEE Trans. Image proc. , Vol 6, NO 12, pp 1621 – 1633, December 1997*
6. *M. Elad and Y. Hel-Or "A Fast Super Resolution Reconstruction Algorithm For Pure Translational Motion And Common Space-Invariant Blur" IEEE Assembly, pp 402 – 405, 2000*