

Joint transform optical correlation applied to sub-pixel image registration

Thomas J Grycewicz^{*a}, Brian E Evans^{a,b}, Cheryl S Lau^{a,c}

^aThe Aerospace Corporation, 15049 Conference Center Drive, Chantilly, VA 20151-3824

^bPhysics Department, Virginia Tech, Blacksburg, VA 24061-0435

^cColumbia University Department of Computer Science, New York, NY 10027-7003

ABSTRACT

The binary joint transform correlator (BJTC) can provide sub-pixel correlation location accuracy for a pair of almost identical inputs, as is the case when computing the registration offset between two overlapping images from the same sensor. Applications include noise cancellation, motion compensation, super-resolution processing, and image splicing. We experimentally demonstrated sub-pixel registration and image co-addition. Our results show a resolution improves by a factor of almost two compared to normal integration. This paper details early results in an ongoing project.

Keywords: Joint Transform Correlator, image registration, co-addition, super-resolution processing

1. INTRODUCTION

Optical binary joint transform correlators (BJTCs)^{1,2} are ideal processors for computing the shift of an image or sub-image from one frame of data to the next.³ The shift amount can be measured to a small fraction of a pixel.⁴ This registration process can be used to align whole images for noise reduction and/or resolution enhancement, image splicing, or scene-based adaptive optics.⁵ Through the parallel nature of optical processing, computation is done very quickly and uses very little power. Computation time is determined by the integration time on the camera and can be a fraction of a millisecond. System response is dominated by camera and spatial light modulator (SLM) data transfer rates, which can be more than a thousand frames per second with current equipment. This makes an optical coprocessor an attractive option for high-speed image processing.

The primary motivation for a fast image registration system investigated in this work is camera motion compensation and resolution enhancement^{6,7} through co-adding frame data. The basic idea is to start with a series of images taken at a very high frame rate. The series of frames is then aligned and added together to form a single still image. As the input frame rate approaches or exceeds 100 frames per second, the processing required quickly exceeds the real-time capability of small digital processors. If the JTC is incorporated in a pipeline processor in a fast focal plane camera system, processing speed can be extended significantly. Co-addition of frames can be used to get the high signal-to-noise of a long exposure while eliminating motion effects. This enables new system performance tradeoffs. When registration is done at several widely spaced locations in the image, translation, rotation, and some simple distortions can be removed. Co-addition also relaxes requirements for holding the camera steady while capturing an image. Summing images digitally reduces the motivation for high resolution analog to digital conversion at the focal plane.

This paper reports our initial efforts to investigate the advantages of high-speed JTC-based image registration at The Aerospace Corporation. While we have been pursuing simulation studies for over a year, the Optical Correlation Laboratory at Aerospace's Chantilly location has only been open for a month. Initial registration experiments have verified resolution enhancement. Results closely follow predictions from simulation. Planned experiments will investigate input and Fourier plane pre-processing techniques, and applications in image splicing, scene-based adaptive optics, and moving target detection.

* thomas.j.grycewicz@aero.org; phone (703)324-8857

2. THEORY

A brief description of JTC operation follows. For an in-depth mathematical description, many references are available.⁸ A block diagram of a binary JTC is shown in Fig. 1. Joint transform correlation is a two-stage process. In both stages the image of a coherently illuminated input passes through a lens to take its Fourier transform. The input device is usually a SLM, and the output device is usually a CCD camera. The camera in the Fourier plane can be combined with the second stage input in an optically addressed SLM. In the first stage input the reference and target scene images are displayed side by side. Coherent light is used to illuminate the input, and a lens is used to produce the joint Fourier transform. The joint power spectrum is detected in the Fourier plane, and is the input to the second stage. The second Fourier transform produces the correlation output. The binary joint transform correlator (BJTC) adds a binarization operation between detection of the joint power spectrum in the first stage and the input of the second stage.

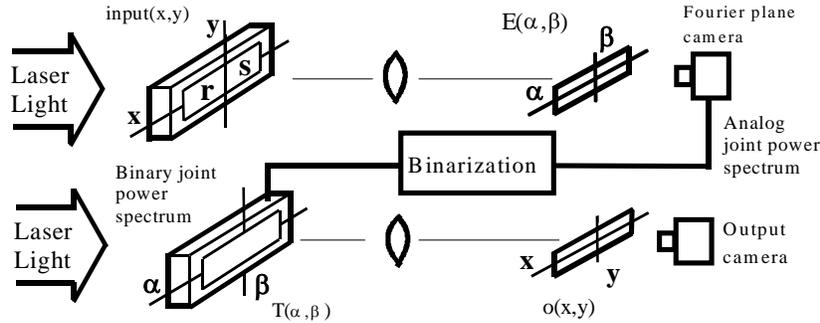


Figure 1. Binary Joint Transform Correlator. Illustration courtesy U.S. Air Force.

If the input to a JTC is made up of a reference and a scene image separated by a distance $2\Delta x$:

$$input(x, y) = r(x - \Delta x, y) + s(x + \Delta x, y). \quad (1)$$

Then the joint power spectrum captured by the camera in the Fourier plane is

$$E(\alpha, \beta) = |R(\alpha, \beta)|^2 + |S(\alpha, \beta)|^2 + R^*(\alpha, \beta)S(\alpha, \beta)e^{j4\pi\Delta x\alpha} + R(\alpha, \beta)S^*(\alpha, \beta)e^{-j4\pi\Delta x\alpha}. \quad (2)$$

Here α and β include dimensional scaling factors determined by the lens focal length, the wavelength of the light, and the pixel pitches on the camera focal plane and SLM. Binarization or other nonlinear processes can be applied prior to displaying the joint power spectrum on the second stage SLM. Since the first and second stages carry out the same process, it is common to use one SLM and camera to implement both stages. Once the joint power spectrum is captured, it is loaded on the SLM, replacing the inputs. The second stage output is the correlation plane:

$$\begin{aligned} o(x, y) = & r(x, y) \otimes r(x, y) + s(x, y) \otimes s(x, y) \\ & + r(x, y) \otimes s(x, y) \otimes \delta(x - 2\Delta x) + s(x, y) \otimes r(x, y) \otimes \delta(x + 2\Delta x). \end{aligned} \quad (3)$$

When the reference and scene contain the same information with a slight displacement, the displacement can be determined by finding the displacement of the cross-correlation peaks centered at $\pm 2\Delta x$. Depending on the preprocessing done these correlation peaks can be quite sharp. For the BJTC, the shape of these peaks approaches a delta function centered at the displacement distance. In a practical system, the location of this peak can easily be determined with sub-pixel accuracy.

3. SIMULATION

Image registration has been modeled through computer simulations written in the IDL programming language. Our initial simulations used images differing by translation, rotation, and addition of noise. Simulations dealing with translation are described below. Images we used to test resolution accuracy are shown in Figure 2. The first image is a subset of a New York City image supplied with IDL and the second an USGS image of the terminal area of Reagan National Airport. Higher fidelity simulations were later done using an image of an USAF 1951 test pattern as an input. These simulations directly model work which has been or will be done in the lab.

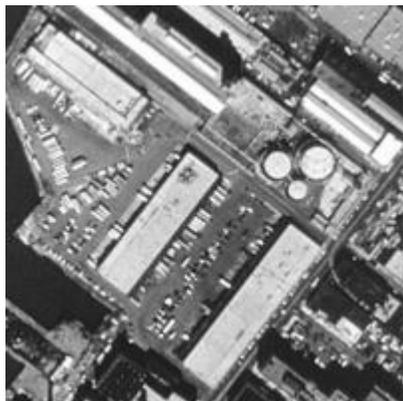


Figure 2.a. New York City.
Illustration courtesy NASA.



Figure 2.b. Reagan National Airport
Illustration courtesy USGS.

To model sub-pixel image shifts, the image was shifted by between zero and two pixels in the x and y directions, with new pixel values determined by linear interpolation. The pixels were then binned together in groups of four, reducing the image resolution by a factor of two in both directions. If we assume Nyquist sampling for the original image, this produces an accurate modeling of sub-pixel translation.

The goals of the simulation study was to test accuracy of sub-pixel translation measurement on a JTC processor. Our simulations modeled the basic JTC, and variants where simple filtering is applied at the input and Fourier planes. Only processes efficiently implemented digitally were considered. These included frame subtraction⁹ and convolution filtering.¹⁰ Convolution kernels used were 1x3 and 3x3 pixels in size. Fourier plane math operations were limited to shift (multiplication by two), add, subtract, and compare. These limitations are put in place so that one can easily envision this preprocessing done at video rate by a FPGA chip in-line with the camera. Computer frame grab cards with this capability are available commercially. Output processing uses peak detection followed by centroid location on a 5x5 pixel square.

Registration used correlation of small sub-images. The size of the sub-images registered ranged from 8x8 to 128x128. Even with sub-image sizes as small as 16x16 pixels, image registration with an RMS accuracy of 1/10 pixel was achieved. Correlation of small sub-images is important if the goal is to measure image rotation or distortion, and multiple correlation points are needed. Small shifted sub-images were then registered with the un-shifted image using a JTC model. It was found that measurement accuracy on the order of 1/8 pixel was easily obtained when the input sub-images were at least 16x16 in size for all cases where a preprocessing algorithm was applied to the input plane. Without preprocessing, the correlator generally failed. The results are tabulated in Table 1 below:

Root mean position error (pixels) vs. image size and processing for NYC Landsat data						
		Input image size				
Input Processing	Fourier plane	16x16	20x20	32x32	64x64	128x128
3x3 Laplacian, Binary	linear	0.1102	0.0999	0.1076	0.1005	0.1124
3x3 Laplacian	linear	0.0810	0.0854	0.1039	0.1114	0.1104
none	linear	fail	fail	fail	fail	fail
3x3 Laplacian	binary	0.1145	0.1190	0.1201	0.1244	0.1215
3x3 Laplacian, Binary	binary	0.1158	0.143	0.1357	0.125	0.1265
none	binary	0.3340	0.2662	0.1563	0.1311	0.1166

Localization based on centroid of 5x5 region in output plane
Study area is centered in downtown Manhattan

Table 1a. Registration accuracy for New York City data.

Root mean position error (pixels) vs. image size and processing for DCA USGS						
		Input image size				
Input Processing	Fourier plane	8x8	12x12	16x16	32x32	64x64
3x3 Laplacian, Binary	linear	0.3328	0.1659	0.1426	0.0577	0.0529
3x3 Laplacian	linear	0.1472	0.1319	0.1216	0.0644	0.0483
none	linear	fail	fail	fail	fail	fail
3x3 Laplacian	binary	0.1522	0.1510	0.1280	0.1272	0.1233
3x3 Laplacian, Binary	binary	fail	0.1414	0.0991	0.0901	0.0965
none	binary	fail	0.3441	0.2822	0.2726	0.1808

Localization based on centroid of 5x5 region in output plane < 0.125
Study area is a terminal at Reagan National Airport < 0.25
> 0.25

Table 1b. Registration accuracy for Reagan National Airport data.

These simulations were done using a number of approaches to input preprocessing, and compared linear to binary presentation in the Fourier plane (the JTC vs. the BJTC). The convolution kernel

$$L = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix} \quad (4)$$

was used for Laplacian edge enhancement in the input plane. For a binary input, a threshold of zero was used after convolution. Binarization in the Fourier plane used a frame subtraction algorithm⁹. The output peak location (registration) was found by calculating the centroid of a 5x5 region around the peak in the output plane.

The classical linear JTC performance poorly in all situations. It was impossible to reliably isolate the correlation location to within a pixel. In Table 1, a measurement series was labeled as a “fail” if any one correlation location was more than a pixel off. Edge enhancement of the input, binarization in the Fourier plane, or the implementation of both resulted in good sub-pixel tracking of registration.

Several of the cases above were repeated with Gaussian white noise added to the image data. Using a sub-image size of 32x32 at the input, tracking accuracy was found to be very robust in the presence of noise, with signal-to-noise ratios as high as 1:1 at the input pixels. The tracking accuracy degraded very little with this noise, but when the noise was large enough (around 1:1), tracking failed entirely.

A second set of simulation experiments used an USAF 1951 resolution pattern. The original was a 2312x2256 pixel JPEG image. This image was cropped to a 2300x2200 PICASSO input file with one byte per pixel gray scale resolution. The Parametric Image Chain Analysis and Simulation Software (PICASSO) program was developed by The Aerospace Corporation for simulation of satellite imaging systems, but can be applied to any digital camera. In this case the parameters have been set to model a Uniq UF-1000CL CCD camera with a 16 mm lens operated at a range of about a meter. A 15 pixel by 15 pixel region in the input image maps to a single camera pixel. Sub-pixel jitter is modeled by randomly translating the input image by zero to fifteen pixels prior to using PICASSO to model the camera’s optical transfer function (OTF). PICASSO then captures the image at the detector pixel pitch and simulates detector noise in the focal plane. A series of 32 images with random jitter were generated. These images were used as inputs for both simulated and experimental registration. A typical 153x146 pixel input image is shown in Figure 3.

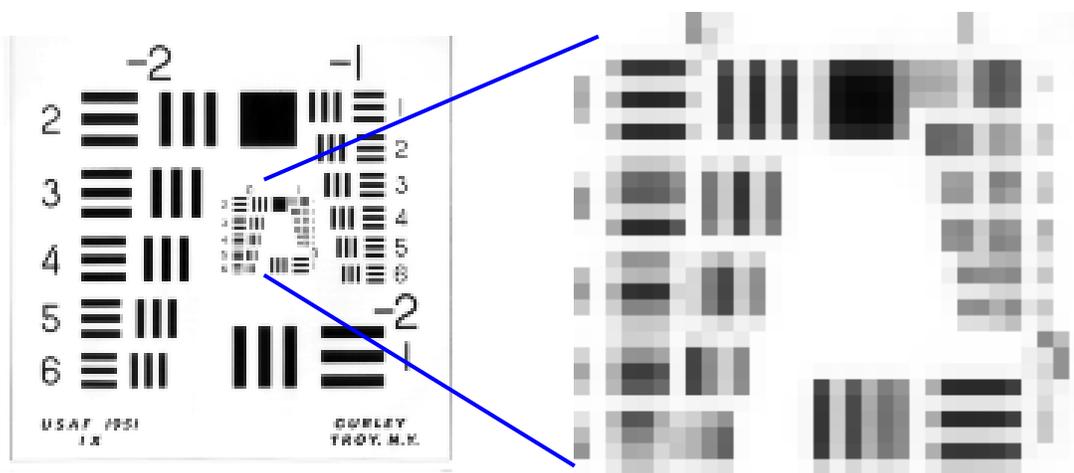


Figure 3. Typical USAF 1951 test pattern input used in both simulation and experiment. The input was created starting with a high-resolution image and degrading it to camera resolution using the PICASSO image chain analysis tool. The expanded area shows the central 0 and 1 order bar patterns, from which a resolution of 0,2 is easily read.

In a static image, lack of resolution can be improved somewhat by simply integrating the image, either through time exposure or through summing multiple frames. Figure 4 shows the result of summing the first five images and for summing the full sequence of 32 images. Since the sub-pixel jitter in these images is uniformly distributed across an area of one pixel, this result is obtained if registration is applied as a nearest neighbor match for images with large (multi-pixel) misregistration. The resolutions read from the USAF 1951 pattern are 0,1 and 0,3. Summing a short sequence of five jittered images results in an increase in the blur seen in the single image. As the number of frames is increased, averaging effects result in a resolution improvement.

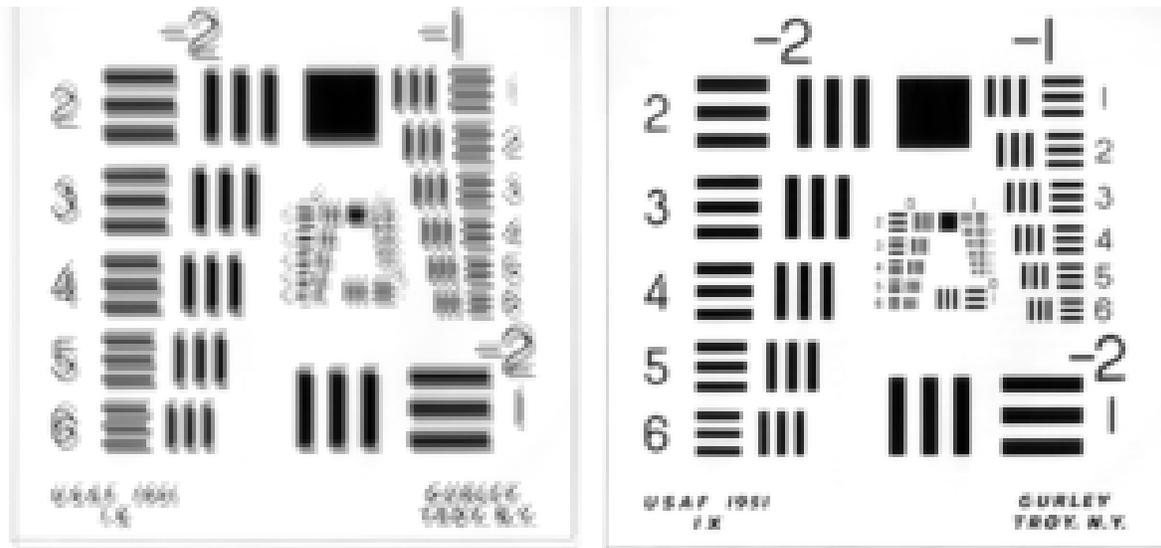


Figure 4. Results of co-addition without registration of the images. The left illustration is the sum of five images with resolution 0,1, while the right output is the sum of 32 images with resolution 0,3.

The effect of registering the image to the nearest half pixel is shown in Figure 5. To generate the image, the number of pixels in each dimension was doubled, resulting in four output pixels for each input pixel. This separated each pixel to four quadrants. Based on the jitter displacement, camera values were assigned to one of the four quadrants and averaged. It is easily seen that even with only a few images the output resolution is greatly improved. The resolution readings from the USAF 1951 are 0,6 and 1,2 in this case. For the 3- image composite, the resolution is almost double the resolution achieved without performing the sub-pixel registration.

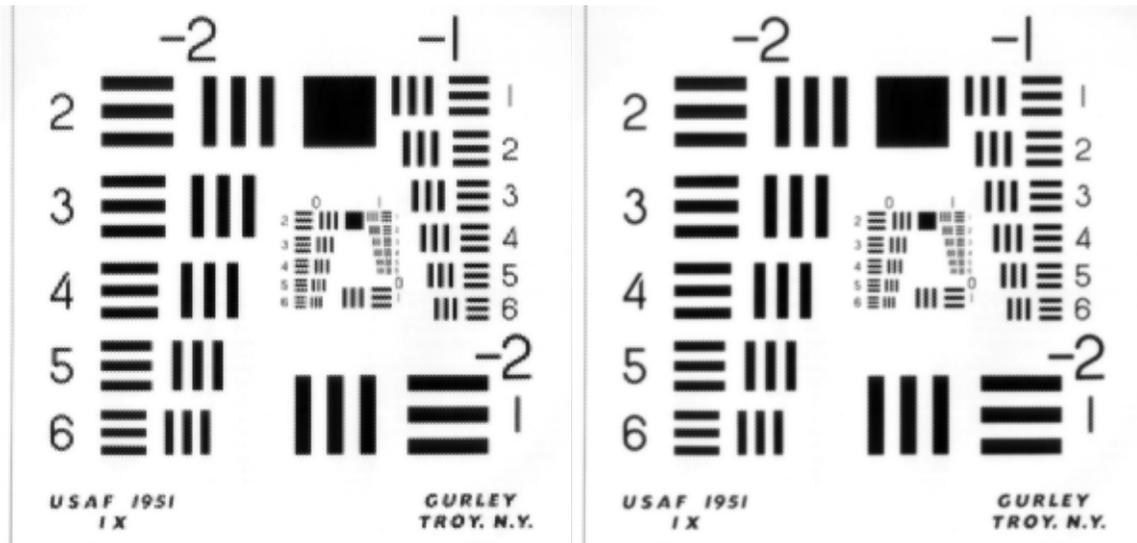


Figure 5. Results of co-addition with sub-pixel registration of the images. The output image has four pixels for every pixel in the input image. The left illustration is the sum of five images and has a resolution of 0,6. The right output is the sum of 32 images and has a resolution of 1,2—twice the resolution of the input image.

In order to perform this sub-pixel registration, it is necessary to accurately measure the registration. The registration was measured by comparing the first image in the series to all of the others using a joint transform correlator. The side edges of the pattern were cropped, and the resultant 128x146 pixel images were placed side-by-side in the 256x256 pixel input plane of a joint transform correlator. No input preprocessing was performed. The joint power spectrum was binarized using a frame subtraction algorithm. The results of this simulation are presented along with the experimental results in the next section. The RMS error between the actual and measured pixel location was 0.192 pixels.

4. EXPERIMENTAL RESULTS

An experimental setup was built based on a Boulder Nonlinear Systems 256x256 ferroelectric SLM and a Uniq model UF-1000CL progressive scan CCD camera in the layout shown in Figure 6. The laser used was a 632.8 nm HeNe, the focal length for the Fourier transform lens was 175 mm, and a polarizing beamsplitter cube was used. HWP is a half-wave plate used to rotate the input polarization to the SLM.

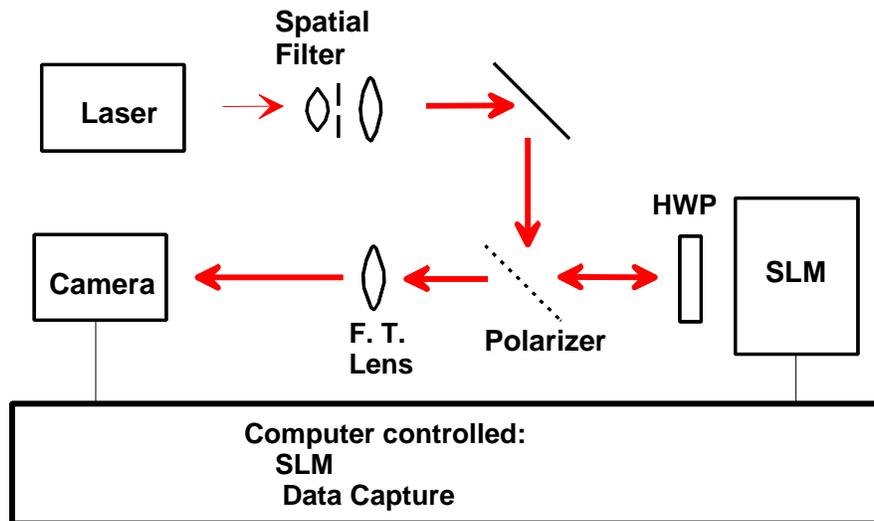


Figure 6. Experimental layout.

The input images developed in the simulation study were used as the JTC input stack. Input preprocessing was not applied. The joint power spectrum was captured on the camera and saved to file. Both convolution filtering and frame subtraction were used as Fourier plane processing methods. The results were almost identical. The simpler convolution filtering method was used for most of the study and for the results presented here. We used the three pixel kernel $[-1, 2, -1]$ for the convolution filter, and binarized the output with a threshold of zero. (This is done by comparing twice the value of the center pixel with the sum of the two nearest neighbors.) The result is displayed on the SLM and the output captured on the camera. The location of the output peaks is found by computing the centroid of a 5x5 pixel region around the brightest pixel in the output region. The experimental registration result is compared to the true location and the location determined through simulated JTC correlation in Figure 7. The RMS error of the registration match is 0.259 pixels. The results of using these registrations to co-add the input images are shown in Figure 8. The resolution read from the USAF 1951 pattern is 1,2. The resolution is improved by nearly a factor of two when compared to the resolution of a single image or when the frames were added without removing the sub-pixel jitter. The same resolution improvement is seen when actual rather than measured registration data was used to align the images.

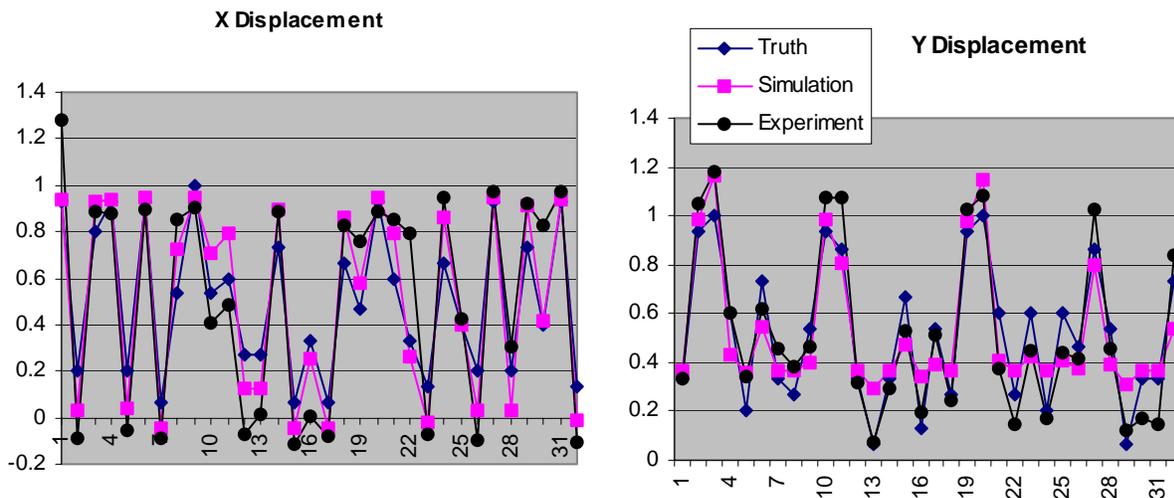


Figure 7. Registration displacement of input data compared to the registration measures for both the simulated and experimental JTC correlator.

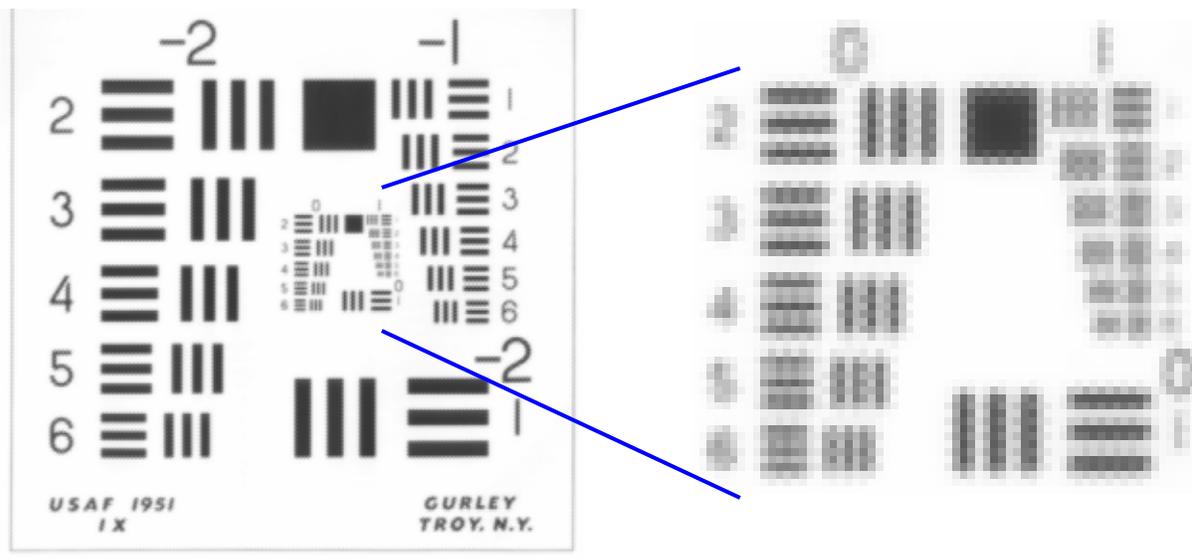


Figure 8. Results of co-adding frames using experimental registration measurements. The resolution achieved was 1,2.

5. CONCLUSIONS

The binary joint transform correlator can provide very accurate sub-pixel correlation locations when presented with a pair of almost identical inputs. This makes it ideal for calculating the registration of overlapping images captured with the same camera at nearly the same time. Applications of this are noise cancellation, motion compensation, super-resolution processing, and image splicing. This study focuses on super-resolution processing.

We have experimentally demonstrated sub-pixel registration using the binary joint transform correlator. We have shown application of sub-pixel registration to image co-addition, illustrating an improvement in the resolution of the image by a factor of almost two. We have shown that input pre-processing, with the purpose of enhancing edges, and Fourier plane binarization are critical to high-resolution correlation.

This paper details early results in an ongoing project. Our current experimental registration error is on the order of a quarter pixel. We hope to improve this accuracy through optimizing our experimental set-up. We currently capture each Fourier Plane and output image manually. However, our long-term goal is to demonstrate real-time video-rate registration with a live camera input. To that end, we will be automating the processes of capturing and correlating data. We also plan to address estimating image rotation, which requires measuring the translation at multiple points on each pair of overlapping images.

REFERENCES

- ¹ C.S. Weaver and J.W. Goodman, "A technique for optically convolving two functions," *Appl. Opt.* 5, 1248-1249 (1966).
- ² B. Javidi, C.J. Kuo, "Joint transform image correlation using a binary spatial light modulator in the image plane," *Applied Optics*, 27, 663-665, 1988.
- ³ K.L. Scherer, M.G. Roe, and R.A. Dobson, "Rapid tracking of a human retina using a nonlinear joint transform correlator," *Proc. SPIE 1959, Optical Pattern Recognition IV*, Orlando, FL, April 1993.
- ⁴ T.J. Grycewicz, "Sub-micron position resolution using the chirp-modulated single-lens joint transform correlator," *Optical pattern recognition VIII, Proc. SPIE 3073*, p. 416-420, Orlando, FL, April 1997.
- ⁵ L.A. Poyneer, "Scene-based Shack-Hartmann wave-front sensing: analysis and simulation," *App. Opt.* 42, 5807-5815 (2003).
- ⁶ I.E. Abdou, "Practical approach to the registration of multiple frames of video images," *IS&T/SPIE Conference on Visual Communications and Image Processing '99*, San Jose, CA, Jan 1999.
- ⁷ H. Foroosh, J.B. Zerubia, M. Berthod, "Extension of Phase Correlation to Subpixel Registration," *IEEE Transactions on Image Processing*, Vol. 11, No. 3, p. 188-200 (2002).
- ⁸ Selected papers on *Optical Pattern Recognition Using Joint Transform Correlation*, SPIE Milestone Series MS-157, M.S. Alam, ed., Bellingham, WA, 1999.
- ⁹ T.J. Grycewicz, "Applying time modulation to the joint transform correlator," *Opt. Eng.* 33, pp. 1813-1820 (1994).
- ¹⁰ T.J. Grycewicz, "Fourier plane windowing in the binary joint transform correlator for multiple target detection," *Appl. Opt.* 34, 3933-3941 (1995).