ACCELERATED TEMPLATE MATCHING USING LOCAL STATISTICS AND FOURIER TRANSFORMS

F. WEINHAUS[1]

Abstract – **This paper presents a method to accelerate correlation-based image template matching using local statistics that are computed by Fourier transform cross correlation. This approach is applicable to several different metrics. The concept is based upon equivalent spatial and frequency domain principles. Each metric is computed completely in the frequency domain using Discrete Fourier Transforms. Timing results are shown to be independent of the size of the smaller template image.**

1. INTRODUCTION

Image registration is an operation that aligns the pixels of one image to the corresponding pixels of another image. There are many goals that are typical of image registration. Some of these include: detecting changes between images (as in vegetation analysis in remote sensing and industrial parts quality control), aligning multiple images prior to creating a mosaic (in remote sensing) and looking for similar images (for content based image retrieval and fingerprint analysis). Numerous approaches have been proposed, which include: pixel-based template matching, feature matching,

[1] Sunnyvale, CA

area matching, shape matching, transform analysis matching and heuristics matching. Detailed descriptions can be found in numerous books and survey papers [1]-[7].

2. BACKGROUND

This paper focuses on pixel-based template matching via correlation metrics. This is an old and traditional method where a small image is moved one pixel at a time over a larger image. For each shift position, a metric is computed pixel by pixel between the small image and the correspondingly sized region of the larger image. The position where the metric value is largest or smallest, depending upon the metric, identifies the shift position for which the small image best matches with the large image.

One of the most common metrics is the normalized cross correlation (NCC), which can be expressed in the spatial domain as

$$NCC(h,k) = \frac{\sum_{i,j}\left(S(i,j) - M_S\right)\left((L(i+h,j+k) - M_L)\right)}{\left\{\sum_{i,j}\left(S(i,j) - M_S\right)^2 \sum_{i,j}\left(L(i+h,j+k) - M_L\right)^2\right\}^{0.5}}. \quad (1)$$

Here S(i,j) is the small image, L(i,j) is the large image, M_S is the mean of the small image, $M_L \equiv M_L(h,k)$ is the mean of the subsection of the large image at offset (h,k), N is the number of pixels in the small image and NCC(h,k) is the normalized cross correlation metric at offset (h,k). The numerator is essentially a simple cross correlation, but using a zero mean small image and zero mean subsections of the larger image. The mean subtraction mitigates

brightness differences between the two images. The denominator is included so that the resulting correlation metric ranges from -1 to 1. A perfect match has a value of 1.

Normalized cross correlation, as described by equation (1), is computationally intensive and slow. Part of the complexity has to do with evaluating the numerator correlation in the spatial domain when the template image is large. The other aspect that adds to the complexity is the computation of the mean and standard deviation of each subsection of the larger image.

A number of techniques have been used to speed up these computations. A simple approach uses a coarse to fine search strategy. The images are reduced in size and the correlation metric is evaluated and the best match found. Then the matching is repeated at full resolution, but only in the neighborhood of the coarse match location [8][9]. A variation on this theme involves pyramidal search techniques [10][11].

The Bounded Partial Correlation method uses a sufficient condition test at each shift position to rapidly skip most of the expensive calculations involved in the NCC scores at those points that cannot improve the best score found so far [12].

Another approach skips the normalization and computes the simple cross correlation, C(h,j), using forward and inverse Fourier transforms. A basic principle of Fourier transforms is that convolution in the spatial domain is equivalent to multiplication in the frequency domain. Likewise, correlation

in the spatial domain is equivalent to multiplication in the frequency domain using the complex conjugate of one of the transformed images.

For simple cross correlation, the Fourier transform procedure is as follows. First pad the smaller image with zeros at the bottom and right sides to fill it out to the size of the larger image. Next, apply the Fourier transform to the both the padded small image and the large image. Then, take the complex conjugate of one of them and multiply the two together. Finally, take the inverse Fourier transform. This process is much faster than doing the un-normalized correlation in the spatial domain. This spatial and frequency domain equivalents may be expressed as

$$C(h,j) = \sum_{i,j} S(i,j)(L(i+h,j+k) = F^{-1}\left\{F^{*}(S)F(L)\right\} \equiv S \otimes L, \qquad (2)$$

where F is the Fourier transform, F^{*} is the complex conjugate of the Fourier transform, F^{-1} is the inverse Fourier transform and S is padded with zeros to the same size as the large image. A⊗B, is defined as a shorthand notation for the forward and inverse Fourier transform cross correlation process between any two images A and B. [2] This nomenclature will be used extensively in the subsequent sections.

If the Fourier transforms of the two images are divided by their magnitudes as a form of normalization, then the inverse Fourier transform of the product is called phase correlation [13]. The downside here is that it bypasses the

[2] To avoid normalization corrections, it is best that the internal Fourier Transform normalization, (1/total pixels) is computed in the inverse Fourier Transform

proper normalization. Furthermore, it is based only on phase information and is insensitive to changes in the image's intensity.

Lewis [14][15] used a mixed spatial and Fourier transform approach to compute the NCC. He pointed out that (1) can be expressed as

$$NCC(h,k) = \frac{\sum_{i,j}\big(S(i,j) - M_S\big)\big(L(i+h,j+k)\big) - M_L\sum_{i,j}\big(S(i,j) - M_S\big)}{\left\{\sum_{i,j}\big(S(i,j) - M_S\big)^2\sum_{i,j}\big(L(i+h,j+k) - M_L\big)^2\right\}^{0.5}}. \qquad (3)$$

Furthermore, he noted that the second term in the numerator is zero, because $\Sigma S(i,j) = \Sigma M_S = NM_S$. This allowed him to compute the numerator with the Fourier transform cross correlation as in (2) after subtracting the mean from the small image. On the other hand, he computed the large image's denominator term in the spatial domain using summed area tables [16] to speed up that part of the computation. The summed area tables were used to evaluate the mean and mean squared of each subsection of the large image very quickly.

Lastly, others have used variations the Lewis technique using summed area tables to compute the denominator. But they have also computed the numerator using summed area tables [17] or with a weighted sum of basis functions[18].

The method described in the following sections computes the NCC and other metrics using Fourier transform correlations alone.

3. LOCAL STATISTICS

In (3), the two denominator terms can be separate and by definition are just the standard deviation of the small image, σ_S, and the standard deviation of the large image's subsections, $\sigma_L \equiv \sigma_L(h,k)$. Therefore, (3) may be expressed as

$$NCC(h,k) = \frac{\sum_{i,j} S'(i,j)\big(L(i+h,j+k)\big)}{N\sigma_S\sigma_L}, \tag{4}$$

where $S'(i,j) \equiv S(i,j) - M_S$.

In the spatial domain, the local sums of L at each offset position of the small image relative to the large image can be computed as a correlation of the large image with a rectangular kernel, U, the size of the small image having unit weights at each element. The mean values, M_L are then achieved by dividing the sums by N. The important factor here is that the local mean image is achieved from just a correlation with a uniform kernel of unit value weights. This can be expressed as

$$M_L(h,k) = \left(\frac{1}{N}\right)\sum_{i,j} U(i,j)L(i+h,j+k). \tag{5}$$

As mentioned earlier, correlation in the spatial domain is equivalent to multiplication in the frequency domain (with one Fourier transform term conjugated). The corresponding frequency domain image, U, is then just an

image the size of the small image with all pixel values equal to unity, but padded with zeroes at the bottom and right sides to the size of the large image. This is in analogy to the padding process used when cross correlating the small and large images using Fourier transforms as described earlier. Therefore, the Fourier transform analogy to (5) is just

$$M_L(h,k) = M_L = \left(\frac{1}{N}\right)(U \otimes L). \tag{6}$$

For the standard deviation, σ_L, one may recast it in variance form as

$$\sigma_L(h,k) = \sigma_L = \left[\left(\frac{1}{N}\sum_{i,j} L(i+h,j+k)^2\right) - \left(\frac{1}{N}\sum_{i,j} L(i+h,j+k)\right)^2\right]^{0.5}. \tag{7}$$

By definition, the standard deviation of x is just the square root of the variance of x, which is equal to the mean of the square of x minus the square of the mean of x. Therefore, the standard deviation of each subsection of L can be expressed, using the shorthand notation for the Fourier transform correlation process, as

$$\sigma_L(h,k) = \sigma_L = \left[\frac{(U \otimes L^2)}{N} - \frac{(U \otimes L)^2}{N^2}\right]^{0.5}. \tag{8}$$

4. CORRELATION METRICS

In this section, several different correlation metrics will be expressed using the Fourier transform correlation method.

4.1 Normalized Cross Correlation

Equations (6) and (8) may be substituted into equation (4) to give the final Fourier transform format of the Normalized Cross Correlation for a grayscale image.

$$NCC(h,k) = \frac{S' \otimes L}{\sigma_S \left\{ N(U \otimes L^2) - (U \otimes L)^2 \right\}^{0.5}}. \qquad (9)$$

Equation (9) shows that the normalized cross correlation can be evaluated using only 3 simple correlations via Fourier transforms. For color images, either the images are converted to grayscale first or the correlation is performed on each color channel and the results combined. This author first used this approach[3] in a template matching study of pairs of images where one image was a photograph and the other was a synthetic thermal image. Test condition variations included different thermal wavelengths, lighting conditions and noise levels [19].

Sun, et. al. [20] and later Papamakarios [21] used a similar local statistics method to perform normalized cross correlation solely using Fourier transforms. However, they reduced the computational complexity to what they called 2.5 FFTs. Their approach combined L and L^2 into one complex expression, $(L+iL^2)$, before applying a forward Fourier transform, $F(L+iL^2)$, and then recovered the separate correlations from
$F^{-1}(U*L) + iF^{-1}(U*L^2)$.

4.2 Root Mean Squared Error

[3] Technique only reported verbally at the conference presentation

For a grayscale image, the root mean squared error metric may be expressed as

$$RMSE = \left\{ \frac{1}{N} \sum_{i,j} \left(S(i,j) - L(i+h, j+k) \right)^2 \right\}^{0.5} \tag{10}$$

or with expansion as

$$RMSE = \left\{ \frac{1}{N} \sum_{i,j} \left(S^2(i,j) - 2S(i,j)L(i+h, j+k) + L^2(i+h, j+k) \right) \right\}^{0.5}. \tag{11}$$

Since the small image is independent of (h,k), its squared sum is a constant. So, one may fill out a new small sized image, T, with uniform values of this sum and then pad it with zeroes to the size of the larger image. Then, (11) may be converted to Fourier transform correlation form as

$$RMSE = \left\{ \frac{1}{N} \left(T - 2(S \otimes L) + (U \otimes L^2) \right) \right\}^{0.5}. \tag{12}$$

For a color image, the argument inside the radical would be evaluated for each channel, added together and divided by the number of channels. The RMSE metric is unbounded and a perfect match has a score of 0.

4.3 Dot Product Correlation

When the two images are dissimilar in sensors or lighting, such as the case in [19], it becomes advantageous to use the intensity values of edge-extracted images rather than the raw image values. Either of the above two

metrics may be used in this case. However, if edge directions are also extracted, then a dot product like metric may provide better results. Such a metric may be expressed as

$$DPC = \sum_{i,j} \frac{S_X(i,j)L_X(i+h,j+k) + S_Y(i,j)L_Y(i+h,j+k)}{NS_M(i,j)L_M(i+h,j+k)}, \tag{13}$$

where subscripts X,Y, M correspond to the X gradient direction image, the Y gradient direction image and the gradient magnitude image. The latter is simply the square root of the sum of squares of the two gradient direction images. Each gradient derivative component may be divided by its respective magnitude. This will be indicated below with an apostrophe. If the resulting small images are padded with zeros to the size of the large image, then (13) may be converted to Fourier transform correlation form as

$$DPC = \frac{1}{N}\left(S'_X \otimes L'_X + S'_Y \otimes L'_Y\right). \tag{14}$$

A variation of this dot product correlation metric was also used in [19]. The DPC metric has a range of values between -1 and 1 and a perfect match has a value of 1.

5. RESULTS

Equations (9), (12) and (14) were each implement as Unix bash shell scripts using the open source, cross-platform, image processing suite called Imagemagick [22]. It utilizes the open source FFTW [23] package to perform Fourier transforms. The Imagemagick suite includes brute-force

spatial domain normalized cross correlation and root mean squared error metrics for template matching. A 2010 vintage 2.66 GHz Intel Core 2 Duo Mac Mini was used for testing. Imagemagick, which can be configured for multi-threaded operation via OpenMP, was limited to one thread for most of these tests.

Figures 1 show the metric surfaces computed with NCC, RMSE and DPC correlation methods, respectively. The large image had dimensions of 256x256 and the small image was a 128x128 subsection located at coordinates (64,52). Each metric successfully found the correct match location. The NCC score was 1.00, the RMSE score was 0.00 and the DPC score was 1.00 to two decimal places. No tests were performed in this study for robustness against noise, image distortions or different images. For the RMSE surface, the image in Figure 1 has been inverted to show bright values for the best match. The Sobel edge detector was used to create the X and Y edge directional derivatives needed for the DPC. The DPC metric surface shows only a very small white dot at the correct locations, whereas the surfaces for the other methods have a wider peak.

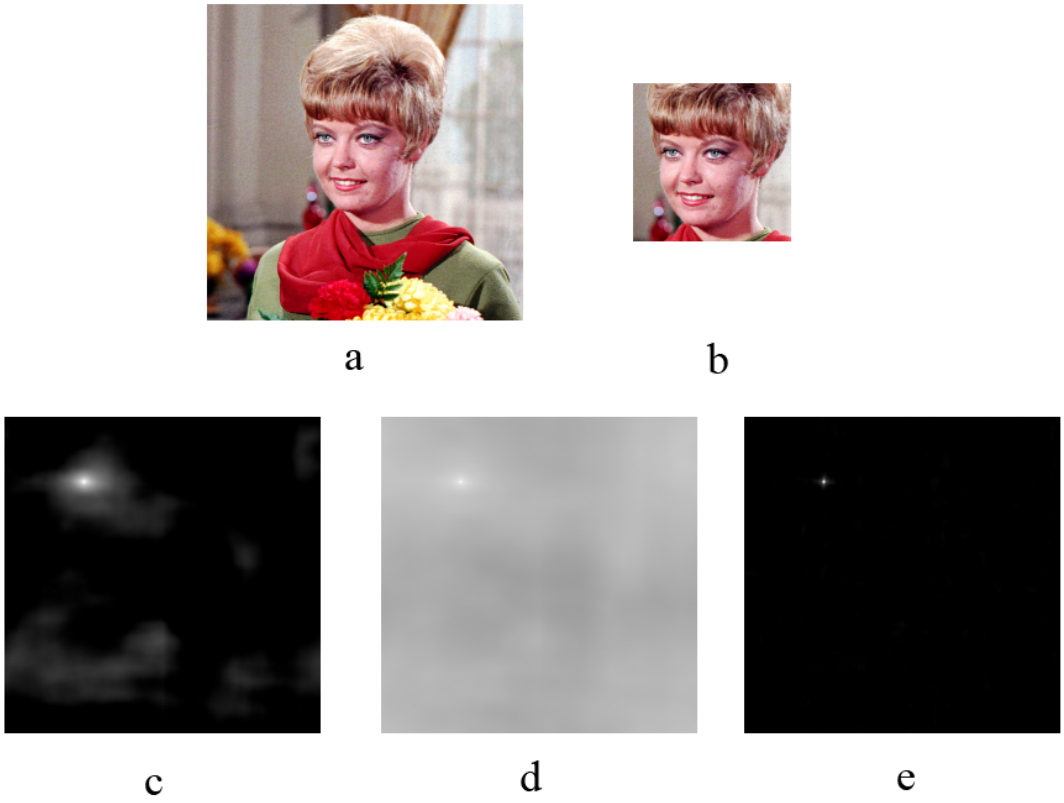

Figure 1 (a) 256x256 large image, (b) 128x128 small image, (c) NCC metric surface, (d) RMSE metric surface and (e) DPC metric image.

Tests, also, were performed to compare the run-times of both the brute force (1) and the Fourier transform (9) approaches for the NCC metric. Comparisons were made for a 500x500 color large image and various square sizes for the small image ranging from 10x10 to 450x450. The resulting CPU times are shown in Table 1 along with their ratios, which characterize the speed-up factor associated with the frequency domain method compared to the spatial domain method. The last columns shows the estimated ratio between the brute force method and Lewis's method, where the denominator is computed with summed area tables and the numerator is computed with the Fourier transform evaluation of the cross correlation. The estimates are based upon the computational burden analysis presented in Lewis's papers and summarized in Table 2. The estimates are based solely on counts of

additions, subtractions and multiplications, each weighted equally. The estimates do not take into account memory and other limiting factors. The computation of the Fourier transform cross correlation is also subject to type of radix approach used in actual implementation. Consequently, these estimated speed-up factors should be considered with liberal uncertainty. Nevertheless, Table 1 shows that substantial performance improvements can be achieved by Lewis's method. However, the method proposed in this study using Fourier transforms to evaluate both the numerator and denominator would appear to be even faster and independent of the size of the smaller image. Table 3 shows the same data, but for dual threading. Table 4 show the same kind of data, but RMSE matching.

Table 1. Single threaded comparison of run times between spatial and Fourier domain normalized cross correlation approaches.

Large Image Size	Small Image Size	NCC Spatial Domain Run Times seconds	NCC Frequency Domain Run Times seconds	Ratio Spatial vs Frequency Domains Timings	Estimated Ratio Spatial Domain vs Lewis Method
500x500	10x10	22.1	1.6	14	1
500x500	25x25	39.4	1.6	25	5
500x500	50x50	91.7	1.5	62	19
500x500	100x100	247.2	1.4	174	59
500x500	150x150	411.4	1.4	286	102
500x500	200x200	536.1	1.5	353	133
500x500	250x250	593.4	1.5	406	145
500x500	300x300	554.8	1.5	375	134
500x500	350x350	427.0	1.5	285	103
500x500	400x400	262.4	1.5	172	60
500x500	450x450	84.3	1.6	54	19

Table 2. Estimated operation counts between brute force spatial domain and Lewis' method. This is the basis of the numbers in the last column of Table 1.

D=large image square dimension and d=small image square dimension	NCC Estimated Operations Spatial Domain	NCC Estimated Operations Lewis Method
numerator	2*(d*d)*(D-d+1)*(D-d+1)	30*D*D*log2(D*D)
denominator	3*(d*d)*(D-d+1)*(D-d+1)	4*D*D + 8*(D-d+1)*(D-d+1)

Table 3. Double threaded comparison of run times between spatial and Fourier domain normalized cross correlation approaches. Only a slight gain in speed seems to be had between single and double threading.

Large Image Size	Small Image Size	NCC Spatial Domain Run Times seconds	NCC Frequency Domain Run Times seconds	Ratio Spatial vs Frequency Domains
500x500	10x10	15.2	1.3	11
500x500	25x25	23.4	1.3	18
500x500	50x50	51.4	1.3	39
500x500	100x100	130.3	1.3	98
500x500	150x150	215.4	1.3	161
500x500	200x200	289.8	1.4	215
500x500	250x250	321.7	1.4	235
500x500	300x300	289.2	1.4	210
500x500	350x350	289.2	1.4	207
500x500	400x400	150.3	1.6	93
500x500	450x450	51.7	1.5	36

Table 4. Single threaded comparison of run times between spatial and Fourier domain root mean squared error correlation approaches.

Large Image Size	Small Image Size	RMSE Spatial Domain Run Times seconds	RMSE Frequency Domain Run Times seconds	Ratio Spatial vs Frequency Domains
500x500	10x10	17.6	1.2	15
500x500	25x25	20.4	1.1	18
500x500	50x50	27.6	1.1	25
500x500	100x100	48.8	1.1	44
500x500	150x150	70.5	1.1	62
500x500	200x200	88.9	1.1	79
500x500	250x250	108.4	1.1	95
500x500	300x300	106.3	1.2	92
500x500	350x350	82.8	1.2	71
500x500	400x400	60.6	1.2	50
500x500	450x450	19.1	1.2	16

6. CONCLUSION

This paper has presented a method of performing several types of correlation-based template matching, where all major computations are done in the Fourier Domain. This approach has proved both efficient and flexible. Run times are one to two orders of magnitude faster than doing the same types of correlations in the spatial domain.

REFERENCES

[1] A. Ardeshir Goshtasby, 2-D and 3-D Image Registration: for Medical, Remote Sensing, and Industrial Applications, Wiley, 2005.

[2] Barbara Zitova , Jan Flusser, Image Registration Methods: A Survey, Image and Vision Computing 21 (2003) 977–1000.

[3] T. Mahalakshmi, R. Muthaiah and P. Swaminathan, Review Article: An Overview of Template Matching Technique in Image Processing, Research Journal of Applied Sciences, Engineering and Technology 4(24): 5469-5473, 2012.

[4] Richard Szeliski, Image Alignment and Stitching: A Tutorial, Foundations and Trends in Computer Graphics and Vision Vol. 2, No 1 (2006) 1–104, 2006.

[5] Medha V. Wyawahare, Dr. Pradeep M. Patil, and Hemant K. Abhyankar, Image Registration Techniques: An Overview, International Journal of Signal Processing, Image Processing and Pattern Recognition Vol. 2, No.3, September 2009.

[6] Lisa Gottesfeld Brown, A Survey of Image Registration Techniques, Department of Computer Science, Columbia University, New York, NY 1007, January 1992. (http://iu1.bmstu.ru/Public/Books-bkp/lizabrwn.pdf)

[7] Guido Bartoli, Image Registration Techniques: A Comprehensive Survey, Visual Information Processing and Protection Group, Universita degli Studi de Siena, June 2007. (http://clem.dii.unisi.it/~vipp/projects/firb/files/Registration.pdf)

[8] Rosenfeld, A. and G. J. Vanderbrug, Coarse-fine template matching, *IEEE Trans. Systems, Man, and Cybernetics*, 104–107 (1977).

[9] Goshtasby, A., S. H. Gage, and J. F. Bartholic, A two-stage cross correlation approach to template matching, *IEEE Trans. Pattern Analysis and Machine Intelligence*, **6**(3):374–378 (1984).

[10] S.L. Tanimoto, Template matching in pyramids, Computer Graphics and Image Processing, vol. 16(4), 1981, 356-369.

[10] W. James MacLean · John K. Tsotsos, Fast pattern recognition using normalized grey-scale correlation in a pyramid image representation, Machine Vision and Applications, Springer-Verlag 2007, DOI 10.1007/s00138-007-0089-8

[11] W. James MacLean and John K. Tsotsos. Fast Pattern Recognition Using Gradient-Descent Search in an Image Pyramid. *Proceedings of 15th Annual International Conference on Pattern Recognition*, volume 2, 877–881, Barcelona, Spain, September 2000.

[12] Luigi Di Stefano, Stefano Mattoccia, Fast template matching using bounded partial correlation, Maschine Vision and Applications (2003) 13: 213–221.

[13] Kuglin, C. D. and Hines, D. C., The Phase Correlation Image Alignment Method. *Proceeding of IEEE International Conference on Cybernetics and Society*, pp. 163-165, 1975, New York, NY, USA.

[14] J.P. Lewis, Fast Template Matching, Vision Interface 95, Canadian Image Processing and Pattern Recognition Society, Quebec City, Canada, May 15-19, 1995, 120-123.

[15] J. P. Lewis, "Fast Template Matching", *Vision Interface,* p. 120-123, 1995.

[16] F. Crow, "Summed-Area Tables for Texture Mapping", *Computer Graphics*, vol 18, No. 3, pp. 207- 212, 1984.

17] D. M. Tsai and C. T. Lin, "Fast normalized cross-correlation for defect detection," *Pattern Recognition Letters*, vol. 24, pp. 2625-2631, 2003

[18] K. Briechle and U.D. Hanebeck, Template matching using fast normalized cross correlation, Optical Pattern Recognition XII, vol. SPIE-4387, The International Society for Optical Engineering, Bellingham, WA, USA, 2001, pp. 95-102.

[19] F. Weinhaus and G. Latshaw, Edge Extraction Based Image Correlation, *Proceedings SPIE*, Vol. 205, 67-75, 1979.

[20] Xiaobai Sun, Nikos P. Pitsianis and Paolo Bientinesi, Proc. of SPIE Vol. 7074, 2008.

[21] Georgios Papamakarios, Georgios Rizos, Nikos P.Pitsianis and Xiaobai Sun, SPIE Vol. 7444, 2009.

[22] http://www.imagemagick.org/index.php

[23] http://www.fftw.org/