REVIEW

Template Matching and Measures of Match in Image Processing

G. S. Cox
Department of Electrical Engineering
University of Cape Town

July 12, 1995

Contents

Table of Contents **1**

1 Introduction **3**

1.1 Preliminary Information and Nomenclature 4

1.1.1 Measures of Match and Mismatch 4

1.1.2 Relative and Absolute Measures of Match 4

1.1.3 Additive Measures of Match 4

1.1.4 Distance Metrics 5

1.1.5 Template Matching 5

1.1.6 Matching Filters 5

1.2 Outline 6

2 Measures of Match **6**

2.1 Correlation Measures 6

2.2 Intensity Difference Measures 8

2.3 Sequential Similarity Detection Algorithms 9

2.4 Sign Change Criteria 10

3 Distortion Measures **10**

4 Matching Filters **12**

4.1 Optimal SNR Filters 12

4.1.1 The Matched Filter 13

4.1.2 The Least Mean Square Filter 13

4.1.3 Phase-Only Matched Filters 13

4.2 Nonorthogonal Image Expansion 13

4.2.1 Discriminative Signal to Noise Ratio 14

4.2.2 Expansion Matching 14
4.2.3 Multiple Template Expansion Matching 16

5 Efficient Template Matching Algorithms 17

5.1 Fast Fourier Transform Techniques 18
5.2 The Cumulative Mismatch Threshold 19
5.2.1 Ordered Application of the Template Pixels 20
5.2.2 The Adaptive Threshold 20
5.3 Two Stage Template Matching 20
5.4 Efficient Search Strategies 21

1 Introduction

In the majority of object recognition applications of computer vision, it is necessary to determine a measure of the similarity between different scenes, or parts of scenes, that have been presented to the computer vision system. In many implementations the scene information is represented by two-dimensional digitised images. This report discusses different quantitative measures of similarity, or measures of match, between similarly sized two-dimensional digital images. Much research in the last four decades [1, 2, 3, 4] has sought to register two-dimensional images by explicitly extracting features of objects in the images. Examples include features generated by segmentation and shape analysis, and texture analysis of local regions. In this report, approaches that involve any explicit feature extraction, or image understanding, are not considered since they require explicit knowledge of the content of the images. The only information used is the actual image data and statistical or mathematical models for the image and noise sources.

Consider the the general case of determining a measure of match between two digital images. Let $\tilde{f}(x, y)$ and $\tilde{g}(x, y)$ give the pixel intensity in the images as a function of position, with the origin centred in the images. Then let

$$f_{ij} = \tilde{f}(in, jn)$$

and

$$g_{ij} = \tilde{g}(in, jn)$$

where n is is the distance between points on the discretisation grid. Pixel intensities take on a value between 0 and I_m, where I_m is the maximum intensity. The intensity values are assumed to be continuous over this range.

For notational convenience we define the following operator:

$$\sum_A \equiv \sum_{i=-\frac{t}{2}}^{\frac{t}{2}} \sum_{j=-\frac{t}{2}}^{\frac{t}{2}}$$

for images of $t \times t$ pixels. $n = t \times t$ is the number of pixels in each image.

We also define the mean intensity values of f and g to be:

$$\begin{aligned} \bar{f} &= \frac{1}{n} \sum_A f_{ij} \\ \bar{g} &= \frac{1}{n} \sum_A g_{ij} \end{aligned}$$

In this document we will refer to measures of *match*, although some of the measures are actually measures of *mismatch* which are designed to increase with decreasing similarity between two images.

1.1 Preliminary Information and Nomenclature

Measures of match and template matching in two dimensional images are topics that have been investigated by researchers from many different fields; radar signal processing, medical imaging, remote sensing, surveying, astronomy to mention only a few. For this reason, similar research in this area has been presented using differing terminology and background. Differing backgrounds on the part of researchers has resulted in varied approaches to measuring similarity between digital images. This section attempts to provide a basis and nomenclature for the concepts that are presented later.

1.1.1 Measures of Match and Mismatch

A *measure of match* between two images is considered to be a metric that indicates the degree of similarity *or* dissimilarity between them. Unless it is specifically stated otherwise, this metric can be increasing or decreasing with the degree of similarity. Where the metric is specifically stated to be a *measure of mismatch*, it is a quantity that is increasing with the degree of *dis*similarity.

1.1.2 Relative and Absolute Measures of Match

The measure of match is used to classify the unknown samples, deciding whether they are a representation of the template or not. The actual classification procedure that uses measures of match to make a decision depends on the particular template matching problem. If it is known that there is exactly one instance of the template among the samples (for example, where a known object is to be located in a scene), the sample with the highest measure of match will be chosen in the absence of any other information. Similarly, if each sample is known to represent one of a fixed number of templates, the sample can be classified according to the template that yields the highest measure of match. In these situations, a *relative* measure of match is required.

A different situation arises if there are an unknown number of instances of a single template in the set of samples (and possibly none). It is necessary to consider the measure of match for each sample in isolation, and decide without reference to the measures of match of other samples whether it represents that template. For this more general problem, an *absolute* measure of match is required.

1.1.3 Additive Measures of Match

In the case where a measure of match between two images can be expressed as the sum of the result of an identical operation performed on each corresponding pair of pixels, we refer to it as an *additive* or *scalar* measure of match. As an example of an additive measure of match, consider the squared error

$$d(f,g) = \sum_{A} (f_{ij} - g_{ij})^2.$$

1.1.4 Distance Metrics

If we consider the space I consisting of equally sized digital images which we will encounter in a particular matching problem, the degree of similarity can be defined in terms of the distance between the the images in this space. The measure of match between digital images f and g, elements of I, can be constructed for a distance function $d(f, g)$ using

$$M_d = \frac{1}{1 + d(f, g)}.$$

For a perfect match $M = 1$ and for increasing mismatch $M \to 0$.

If the distance function satisfies the following properties

$$d(f, g) = 0 \Leftrightarrow f = g$$

$$d(f, g) = d(g, f)$$

$$d(f, g) \leq d(f, h) + d(h, g)$$

then $d(f, g)$ is a metric in the formal mathematical sense, and I is a metric space.

1.1.5 Template Matching

Template matching is the classification of unknown samples by comparing them to known prototypes, or templates. Here we consider the applications in image processing where the template and samples are digitised images. There are two common scenarios:

Search: A search for occurrences of a single template in an image.

Classification: Classification of a sample extracted from an image as one of a number of template prototypes.

In the first scenario the template or prototype remains fixed, while in the second the sample is fixed and there are a number of prototypes. For the purposes of this discussion we will consider the fixed object to be the template, whether it is a prototype or a sample.

Unless stated otherwise, the term *template matching* will refer to the detection of the $n \times n$ sub-image g within $N \times N$ search area s that best matches $n \times n$ template f.

1.1.6 Matching Filters

Exhaustive search template matching algorithms can be described in terms of a *matching filter*, where instances of a template $f(x, y)$ are to be detected in a search area $s(x, y)$. The

filter is designed to give maximal response where a region of the search area matches the template. The filter response, $z(x, y)$, is given by

$$z(x, y) = s(x, y) * h(x, y),$$

where $h(x, y)$ is the matching filter and $*$ denotes the convolution operation.

1.2 Outline

Measures of match for template matching are reviewed in section 2, distortion measures used for measuring image fidelity after image compression or reconstruction are discussed in section 3, and matching filters are reviewed in section 4. Methods for improving the computational efficiency of template matching algorithms are described in section 5.

2 Measures of Match

The measure of match, $M(f, g)$, is a metric that represents the degree of similarity between two digital images, f and g. The measures presented here have been divided into three categories: Correlation measures, inter-pixel distance measures, sequential similarity detection algorithms and sign change criteria. The term "measure of match" is usually used within the context of an image matching application, although some of the measures are used as distortion measures for evaluating lossy image compression and image reconstruction techniques.

2.1 Correlation Measures

The term *correlation measure* is often used in a broad sense to refer to measures of match in general. This stems from the fact that the first techniques considered for recognition in two-dimensional images were based on the matched filters used in one-dimensional signal analysis. The matched filter output at a particular position and cross-correlation of the filter kernel with the image are equivalent. The term *correlation* is used here in more narrow sense, referring to measures of match based on the cross-correlation of two images. Section 4.1 deals with matched filters in more detail.

Normalised cross-correlation: Consider continuous images $\tilde{f}(x, y)$ and $\tilde{g}(x, y)$. By the Cauchy-Schwarz inequality

$$\iint_A \tilde{f}(x, y)\tilde{g}(x, y)dxdy \leq \left[\iint_A \tilde{f}^2(x, y) \iint_A \tilde{g}^2(x, y)\right]^{\frac{1}{2}} \tag{1}$$

specifies an upper bound on $\iint fg$ which is attained when $f = Cg$, where C is a scalar constant.

Since f and g are zero outside A, the left hand side of equation 1 can be rewritten as

$$\iint_{-\infty}^{\infty} \tilde{f}(x,y)g(x,y)dxdy \tag{2}$$

which is the *cross-correlation* of f and g. This however, does not constitute a measure of match since the upper bound on the right hand side of equation 2 changes for different f and g. Therefore we use the *normalised* cross-correlation

$$\frac{\iint_{-\infty}^{\infty} \tilde{f}(x,y)g(x,y)dxdy}{\left[\iint_{-\infty}^{\infty} \tilde{f}^2(x,y) \iint_{-\infty}^{\infty} \tilde{g}^2(x,y)\right]^{\frac{1}{2}}} \tag{3}$$

For digitised images f and g the normalised cross-correlation measure of match is given by

$$M_c(f,g) = \frac{\sum_A f_{ij} \cdot g_{ij}}{\sqrt{\sum_A f_{ij}^2 \sum_A g_{ij}^2}}. \tag{4}$$

In the case of template matching where f is the template,

$$M_{ct}(f,g) = \frac{\sum_A f_{ij} \cdot g_{ij}}{\sqrt{\sum_A g_{ij}^2}}.$$

can be used as a measure of match since $\sum_A f_{ij}^2$ is constant.

Correlation coefficient: The normalised cross-covariance is the linear *correlation coefficient* of statistics, and is equivalent to the normalised cross-correlation of the zero mean equivalents of f and g. It is given by

$$C(f,g) = \frac{n \cdot \sum_A f_{ij} \cdot g_{ij} - \sum_A f_{ij} \sum g_{ij}}{n\sqrt{\left(\sum_A f^2 + \frac{n-2}{n^2}\left(\sum_A f\right)^2\right)\left(\sum_A g^2 + \frac{n-2}{n^2}\left(\sum_A g\right)^2\right)}}. \tag{5}$$

Statistical correlation measure: Pratt [5] proposed a "statistical correlation measure" that is based on the normalised cross-correlation. Referring to (4), the statistical correlation measure is defined as

$$M_{sc}(f,g) = M_c(f',g')$$

where

$$\begin{aligned} f'(i,j) &= f(i,j) * D_f(i,j) \\ g'(i,j) &= g(i,j) * D_g(i,j). \end{aligned}$$

$D_f(i,j)$ and $D_g(i,j)$ are "whitening" filters. An equivalent, but more efficient implementation is obtained by combining the effect of the two whitening filters into a single filter. That is

$$M_{sc}(f,g) = M_c(f'',g)$$

where

$$f''(i,j) = f(i,j) * D(i,j).$$

Pratt derived the convolution operator $D(i,j)$ under the assumptions that the rows and columns are separable, sensor noise is additive white noise, and the row and column elements of the images are samples of a Markov process

The effect of $D(i,j)$ depends on the degree of spatial correlation in the images f and g. If they are totally uncorrelated, the statistical correlation measure reduces to the normalised cross-correlation. At the other extreme, where f and g can be modelled by Markov processes with probabilities $\rho = 1$, $D(i,j)$ takes the form of a spatial differentiation operator.

2.2 Intensity Difference Measures

These measures of match are based on the pixel-by-pixel intensity differences between the two images f and g.

Root mean square distance (RMS): The RMS distance metric is a common measure of mismatch between two digital images. It is given by

$$d_{rms}(f,g) = \sqrt{\frac{1}{n}\sum_{A}(f_{ij} - g_{ij})^2} \tag{6}$$

Sum of squared differences or squared error (SSD): The SSD metric is equivalent to the RMS distance, but requires less computation.

$$d_{ssd}(f,g) = \sum_{A}(f_{ij} - g_{ij})^2 . \tag{7}$$

Sum of the absolute valued differences (SAVD):

$$d_{savd}(f,g) = \sum_{A}|f_{ij} - g_{ij}| \tag{8}$$

Variance of the differences (VOD): The VOD distance is the variance of the pixel by pixel differences between the template f and sub-image g. That is:

$$d_{vod}(f,g) = \frac{n \cdot \sum_A (f_{ij} - g_{ij})^2 - (\sum_A (f_{ij} - g_{ij}))^2}{n^2}. \tag{9}$$

Variance of the squared differences (VOSD): This is a *squared difference* version of the VOD distance.

$$d_{vosd}(f,g) = \frac{n \cdot \sum_A (f_{ij} - g_{ij})^4 - (\sum_A (f_{ij} - g_{ij})^2)^2}{n^2} \tag{10}$$

The VOSD distance is equivalent to an RMS distance measure using $\hat{f}$ and $\hat{g}$, where these are the zero mean equivalents of f and g respectively. That is,

$$\hat{f} = f - \mu_f$$

and

$$\hat{g} = g - \mu_g.$$

Variance of the absolute valued differences (VOAD): This is an *absolute-valued difference* version of the VOD distance. It is equivalent to the SAVD distance in the case where f and g have zero mean.

$$d_{voad}(f,g) = \frac{n \cdot \sum_A |f_{ij} - g_{ij}|^2 - (\sum_A |f_{ij} - g_{ij}|)^2}{n^2} \tag{11}$$

2.3 Sequential Similarity Detection Algorithms

Sequential similarity detection algorithms (SSDAs) were proposed by Barnea and Silverman [6] as a more efficient alternative to correlation measures (matched filters) for translational registration (template matching). The measure of match is based indirectly on an error measure for corresponding pixels in f and g, the images under comparison at any stage during the registration process. Barnea and Silverman use an error measure based on the L_1 norm, the SAVD (see section 2.2);

$$\epsilon_{ij} \equiv \left| (f_{ij} - \hat{f}) - (g_{ij} - \hat{g}) \right| \tag{12}$$

The error measure is calculated for pixel pairs, the coordinates (i,j) of which are selected in a random order. For each pair the error measure is calculated, and a cumulative error updated. When the cumulative error exceeds a predetermined threshold value T, the process stops and the measure of match is set to the number of pairs (i,j) that have been considered. That is,

$$M_{ssda} = \left\{ r | \min_{1 \le r \le n} \left\{ \sum_{n=1}^{r} \epsilon_{ij} \ge T \right\} \right\}.$$

A variation on this method uses a monotonic-increasing threshold sequence instead of a constant threshold. For the kth randomly selected corresponding pixel pair, the threshold value $T(k)$ is used. Barnea and Silverman describe a stochastic approach [6] to generating monotonically increasing threshold curves.

2.4 Sign Change Criteria

The sign change criteria were proposed by Venot *et al* [7] particularly for the registration of dissimilar images. Consider two images f and g which differ only by additive noise with zero mean and a symmetric density function. If we take the pixels in the difference image,

$$d_{ij} = f_{ij} - d_{ij},$$

in row-column order, there will be many sign changes between adjacent d_{ij}. Images which have differences significantly greater than the mean of the noise will not produce many sign changes between adjacent d_{ij}. This is the motivation behind the use of sign change criteria as a basis for measures of similarity between images.

We define the function

$$sgn(x) = \begin{cases} 0 & x \geq 0 \\ 1 & x < 0 \end{cases}$$

Stochastic sign change (SSC): A normalised SSC measure of similarity is given by

$$M_{ssc}(f,g) = \frac{\sum_i^t \sum_{j+1}^t sgn(d_{ij} \cdot d_{i,j-1})}{t(t-1)}$$

where $d_{ij} = f_{ij} - g_{ij}$. Recall that t is the number of pixels per row or column. $M_{ssc}(f,g)$ increases with the number of horizontal sign changes in the difference image, d_{ij}, and is therefore a measure of *match*.

Deterministic sign change (DSC): Without additive noise in the images, the SSC no longer functions as a measure of similarity. In this case the *deterministic* sign change can be used. A third image, $\acute{g}$ is generated as follows:

$$\acute{g}_{ij} = \begin{cases} g_{ij} + q & i+j \text{ even} \\ g_{ij} - q & i+j \text{ odd} \end{cases}$$

The SSC measure is then calculated for f and $\acute{g}$, that is

$$M_{dsc}(f,g) = M_{ssc}(f,\acute{g}).$$

3 Distortion Measures

Measures of match for image matching have a lot in common with distortion measures used for evaluating the performance of lossy image compression [8, 9] or image reconstruction techniques. The commonality lies even deeper for fractal image compression or vector quantisation (VQ), since image matching is used to specify the fractal mappings and to select

VQ codebook entries. Oakley and Cunningham [10] use a function space model to derive a reconstruction filter which is optimal in the sense that the original and reconstructed images are as similar as possible with respect to a particular fidelity measure, or measure of match.

We consider here measures that have been used to quantify the performance of lossy image compression techniques [8, p. 327]. The most common distortion measure is the *squared error* or squared Euclidean distance, given by

$$d(f,g) = \sum_A (f_{ij} - g_{ij})^2 \tag{13}$$

where f is the original image and g is the decoded image after lossy compression. The squared error is a member of the class of scalar distortion measures of the form

$$d_p(f,g) = \sum_A |f_{ij} - g_{ij}|^p$$

where $p = 2$. For $p = 1$, $d_p(f,g)$ is equivalent to the SAVD measure of match and is a distance based on the l_1 norm. The pth root of $d_p(f,g)$ is a distance based on the l_p norm. As $p \to \infty$, $d_p(f,g)$ approaches the *maximum* or l_∞ *norm* distortion measure, given by

$$d_{max}(f,g) = \max_i |f_i - g_i| .$$

If we represent the original and decoded images as vectors $\mathbf{f}$ and $\mathbf{g}$ respectively, then (13) can be written

$$\begin{aligned} d(\mathbf{f},\mathbf{g}) &= \|\mathbf{f} - \mathbf{g}\|^2 \\ &= (\mathbf{f} - \mathbf{g})^T (\mathbf{f} - \mathbf{g}) \end{aligned}$$

Another distortion measure of interest is the *weighted squared error* measure:

$$d(\mathbf{f},\mathbf{g}) = (\mathbf{f} - \mathbf{g})^T \mathbf{W} (\mathbf{f} - \mathbf{g}) \tag{14}$$

where $\mathbf{W}$ is a positive definite weighting matrix. The squared error distortion measure is a special case of (14) where $\mathbf{W} = \mathbf{I}$. $\mathbf{W}$ can be defined to simply weight the contributions of corresponding vector components if only the diagonal elements, w_{ii}, are non-zero, that is

$$d(\mathbf{f},\mathbf{g}) = \sum_i w_{ii} (f_i - g_i)^2. \tag{15}$$

The matrix $\mathbf{W}$ can also be specified according to the statistical properties of the input vector $\mathbf{f}$. For example, if $\mathbf{W}$ is the inverse of the covariance matrix of $\mathbf{f}$, $d(\mathbf{f},\mathbf{g})$ is the *Mahalanobis distortion measure*. If the components of $\mathbf{f}$ are uncorrelated, $\mathbf{W}$ reduces to a diagonal matrix with diagonal elements $w_{ii} = \sigma_i^{-2}$, where σ_i^{-2} is the variance of f_i. This is equivalent to the distortion measure in (15) where the contributions of corresponding vector components in $\mathbf{f}$ and $\mathbf{g}$ are weighted by the inverse of the variance for that component, that is

$$d(\mathbf{f},\mathbf{g}) = \sum_i \sigma_i^{-2} (f_i - g_i)^2.$$

An alternative to a weighting matrix that is based on the stochastic properties of $\mathbf{f}$, such as the Mahalanobis weighting matrix, is one that is based on the deterministic nature of the input vector $\mathbf{f}$. That is a weighting matrix $\mathbf{W}(\mathbf{f})$ that depends explicitly on the values of the input vector components. For example, consider the weighting matrix

$$\mathbf{W}(\mathbf{f}) = \|\mathbf{f}\|^{-2}\mathbf{I}.$$

This gives a distortion measure of

$$d(\mathbf{f}, \mathbf{g}) = \frac{\|\mathbf{f} - \mathbf{g}\|^{-2}}{\|\mathbf{f}\|^{-2}}$$

which weights the distortion according to the strength of the input vector.

4 Matching Filters

In sections 2 and 3 we considered measures of similarity between two images of the same size. Using one of these measures for exhaustive template matching, with template image $\phi(x, y)$, involves generating a match surface by calculating the measure of match for each position in the search area. The generation of the match surface, $m(x, y)$, could also be regarded as a filtering operation, given by

$$m(x, y) = s(x, y) * h(x, y),$$

where $s(x, y)$ is the search area and $h(x, y)$ represents the matching filter.

Matching filters are derived to be optimal in the sense of some match fidelity criterion. We consider two criteria here: Section 4.1 considers the *signal to noise ratio* (SNR), which is a ratio of the peak filter response to the filter response for additive noise. The SNR is used to derive the standard matched filter in the white noise case. Section 4.2 considers the *discriminative signal to noise ratio* (DSNR), proposed by Ben-Arie and Rao [11], which is a ratio of peak filter response to all off-peak filter response. The orthogonal image expansion filter is optimal with respect to the DSNR.

4.1 Optimal SNR Filters

Here we consider the search area to be the sum of the template image at (x_0, y_0) and additive noise, $\lambda(x, y)$, that is

$$s(x, y) = \phi(x - x_0, y - y_0) + \lambda(x, y),$$

then the filter response is given by

$$\begin{aligned} m(x, y) &= s(x, y) * h(x, y) \\ &= k\phi(x - x_0, y - y_0) * h(x, y) + \lambda(x, y) * h(x, y) \\ &= m_\phi(x, y) + m_\lambda(x, y). \end{aligned}$$

where k is a scaling factor.

The SNR is defined to be the ratio of the peak response at (x_0, y_0) to the variance of the noisy response [2, p. 59], that is,

$$SNR = 10 \log \frac{m(x_0, y_0)}{\sigma_\lambda^2}.$$

The filter that maximises the SNR is given by

$$H(u,v) = \frac{\overline{\Phi}(u,v) \exp[-j(ux_0 + vy_0)]}{S_\lambda(u,v)}, \tag{16}$$

where $\overline{\Phi}(u,v)$ is the complex conjugate of the Fourier transform of $\phi(x,y)$, and $S_\lambda(u,v)$ is the expected spectral density of the noise.

4.1.1 The Matched Filter

Matched filters, which were originally applied in one dimensional signal processing applications [12, p. 406], are optimal for detecting a signal in the presence of white noise. Equation (16) is an expression for the optimal matching filter in terms of the SNR. For the restricted case of wide-sense stationary white noise, it can be shown that the optimal SNR filter is the original template, rotated 180°, and translated to (x_0, y_0), that is,

$$h_{mf}(x,y) = f(y - y_0, x - x_0).$$

The matched filter response at (x', y') is equivalent to the cross-correlation (2.1) of $\phi(x,y)$ and the sub-image of $s(x,y)$ at (x', y').

4.1.2 The Least Mean Square Filter

The least mean square (LMS) filter [13] is a variation of the matched filter that is optimal in the presence of globally stationary noise with a $1/f^2$ power spectrum. It is basically a zero-mean matched filter, and is given by

$$h_{lms}(x,y) = h_{mf}(x,y) - \mu_h,$$

where is the mean of $h_{mf}(x,y)$. The LMS filter is equivalent to performing a zero-mean cross correlation of $\phi(x,y)$ and each position of $s(x,y)$.

4.1.3 Phase-Only Matched Filters

4.2 Nonorthogonal Image Expansion

Ben-Arie and Rao [11, 14] propose the use of non-orthogonal image expansion for template matching as an alternative to the matched filter. Expansion matching (EXM), as it is called, produces more localised peaks and reduces the number of spurious peaks in the match surface. This claim is supported by use of a new fidelity criterion: the discriminative signal to noise ratio (DSNR).

4.2.1 Discriminative Signal to Noise Ratio

Since the SNR does not take into account the filter response to signal other than at the position (x_0, y_0), optimal SNR filters produce broad peaks. Ben-Arie and Rao propose the DSNR, which considers as noise all off-peak filter response. Expansion matching is optimal with respect to the DSNR, which is given by

$$DSNR = 10 \log \frac{m(x_0, y_0)}{\sum_{i \neq x_0} \sum_{j \neq y_0} [m(i,j)^2]}.$$

4.2.2 Expansion Matching

The concept behind expansion matching with an $n \times n$ template in an $N \times N$ search area is to represent the search area by basis functions that are effectively the template translated to different positions. In this representation, a basis function with a high coefficient corresponds to a position of the template with a high degree of match. The procedure for deriving the EXM filter as follows:

Non-orthogonal template basis functions: Each of the N^2 basis function consists of an $N \times N$ image which is zero everywhere except for an instance of the $n \times n$ template at one of the N^2 positions. Where the template overlaps the boundary of the basis function (or basis image), it can either be wrapped around or truncated. Ben-Arie and Rao prove that circulant or truncated self-similar basis functions in a dense configuration, such as these, require only modest conditions to be a complete basis for discrete $L^2(\mathbb{R})$ signals such as the search area.

Basis function coefficients for the search area: We estimate the search area $s(x, y)$ as the linear combination of the basis functions $\phi_i(x, y)$, that is

$$\hat{s}(x, y) = \sum_{i=1}^{m} c_i \, \phi_i(x, y), \tag{17}$$

where $m = N^2$ and $\{c_i\}$ are the coefficients of the expansion. The $\{c_i\}$ are calculated to minimise the squared error between $s(x, y)$ and $\hat{s}(x, y)$. If we represent $s(x, y)$, $\hat{s}(x, y)$, and $\{\phi_i\}$ as the m-vectors $\mathbf{s}$, $\hat{\mathbf{s}}$ and $\{\Phi_i\}$ respectively, the squared error d is given by the inner product

$$\begin{aligned} d &= \langle (\mathbf{s} - \hat{\mathbf{s}}), (\mathbf{s} - \hat{\mathbf{s}}) \rangle \\ &= \left\langle (\mathbf{s} - \sum_{i=1}^{m} c_i \Phi_i), (\mathbf{s} - \sum_{i=1}^{m} c_i \Phi_i) \right\rangle. \end{aligned} \tag{18}$$

Minimising d in (18) amounts to solving a set linear equations, but a direct solution is computationally prohibitive, since it involves finding the inverse of an $N^2 \times N^2$ matrix. Fortunately, the image expansion in this case (that is, with template similar basis functions) is equivalent to "restoring" the search area where the blurring function is the template.

Wiener filter formulation for calculating basis coefficients: In order to simplify the notation, we will refer to one-dimensional signals here. Rewriting (17), we wish to determine the coefficients c_i in the following equation

$$s(x) = \sum_{i=-\infty}^{\infty} c_i \, \phi_i(x). \tag{19}$$

Since the basis functions $\phi_i(x)$ are translated versions of the template, (19) can be rewritten as

$$\begin{aligned} s(x) &= \sum_{i=-\infty}^{\infty} c_i \, \phi(x-i) \\ &= c(x) * \phi(x), \end{aligned} \tag{20}$$

where $*$ denotes a discrete convolution. In (20), $s(x)$ can be regarded as the signal $c(x)$, degraded by the blurring function $\phi(x)$. Determining the basis coefficients, is therefore equivalent to restoring $c(x)$. The Wiener filter [4, p. 354], $\theta(x)$, is the optimal filter, in a least squares sense, for restoring $c(x)$. It is assumed that the estimate,

$$\hat{c}(x) = \int \theta(x-t)s(t)dt$$

is a linear function of $s(x)$. The Wiener filter minimises the error

$$e^2 = E\left\{ \int [c(x) - \hat{c}(x)]^2 dx \right\}.$$

We are now in a position to determine the basis coefficient signal by filtering the search area with the optimal Wiener filter, which we shall refer to as the EXM filter.

The EXM filter: If we consider the search area estimate to be to composed of a linearly blurred signal with additive noise $\lambda(x)$, that is

$$s(x) = \int c(t)\phi(x-t)dt + \lambda(x)$$

and we assume that $\lambda(x)$ is zero-mean and uncorrelated with $c(x)$, it can be shown that the Wiener filter is given by

$$\Theta(w) = \frac{\overline{\Phi}(w) S_{cc}(w)}{S_c(w)|\Phi(w)|^2 + S_{\lambda\lambda}(w)}, \tag{21}$$

where $S_c(w)$ is the spectral density of basis coefficient signal, $c(x)$.

Optimality with respect to the DSNR: The EXM filter is made optimal with respect to the DSNR by specifying spectral density, $S_c(w)$, according to the ideal $c(x)$, that is

$$c(x) = \begin{cases} c_s & x = x_0 \\ 0 & \text{elsewhere.} \end{cases} \tag{22}$$

The Wiener filter is guaranteed to find the estimate, $\hat{c}(x)$, which is closest to $c(x)$ in the least squares sense, and by selecting $c(x)$ according to (22), the DSNR is maximised.

4.2.3 Multiple Template Expansion Matching

Rao and Ben-Arie also derive an expansion matching procedure for recognising samples as one of a predetermined set of template vectors, $\Psi_i; i = 1 \cdots N$ [14]. A filter, Θ, is designed to have an expected peak correlation,

$$u_i = E\{\Theta^T \mathbf{S}_i\}, \tag{23}$$

where $\mathbf{S}_i$ is a noisy sample vector given by

$$\mathbf{S}_i = \Psi_i + \Lambda; \quad i = 1 \cdots N,$$

and Λ is the noise vector. The basis coefficient vectors, $\mathbf{C}_i$ are calculated as the filter response using circulant correlation matrix $[\mathbf{S}_i]$, that is

$$\mathbf{C}_i = [\mathbf{S}_i]\Theta; \quad i = 1 \cdots N,$$

where

$$[\mathbf{S}_i] = [\Psi_i] + [\Lambda]$$

with circular correlation matrices $[\Psi_i]$ and $[\Lambda]$ for the template and noise respectively. The optimal filter, Θ, is derived by maximising the DSNR, that is, maximising the ratio of the power at the template centre to the total off-centre power. The DSNR for the i-th template in terms of expected values is given by

$$DSNR_i = 10 \ \log \frac{E\{(\Theta^T \mathbf{S}_i)^2\}}{E\{\mathbf{C}_i^T \mathbf{C}_i\} - E\{(\Theta^T \mathbf{S}_i)^2\}}.$$

This is equivalent to minimising the $E\{\mathbf{C}_i^T \mathbf{C}_i\}$ term in the denominator for each template, but since global minimisation of the DSNR for all templates is not possible, a weighted sum of the individual DSNRs,

$$J(\Theta) = E\left\{\sum_{i=1}^{N} \alpha_i \mathbf{C}_i^T \mathbf{C}_i\right\}; \quad \sum_{i=1}^{N} \alpha_i = 1,$$

is minimised. Typically $\alpha_i = 1/N$. The objective function $J(\Theta)$ is modified to include the constraints provided by equation (23)

$$J(\Theta) = E\left\{\sum_{i=1}^{N} \alpha_i \mathbf{C}_i^T \mathbf{C}_i\right\} + \sum_{i=1}^{N} \xi_i (E\{\Theta^T \mathbf{S}_i\} - u_i). \tag{24}$$

Rao and Ben-Arie minimise (24) and obtain an expression for Θ. The ξ_i are determined by enforcing the constraints in equation (23). A frequency domain formulation for this filter is given by

$$\Theta(w) = \frac{\sum_{j=1}^{N} \tilde{\xi}_j \mathbf{X}_j(w)}{\sum_{i=1}^{N} \alpha_i S_{\Psi\Psi i}(w) + S_{\lambda\lambda}(w)}$$

where $\mathbf{X}_i$ and and $S_{\Psi\Psi i}(w)$ are the DFT and spectral density respectively of the i-th template, and $\tilde{\xi}_i = -\xi/2$. The vector $\tilde{\Xi} = [\tilde{\xi}_1 \cdots \tilde{\xi}_N]^T$ is given by

$$\tilde{\Xi} = [\mathbf{A}]^{-1}\mathbf{U}$$

Table 1: Comparison of measure of match computational complexity

Measure of Match	Operations				
	Mult	Div	Add	Sub	Other
Normalised Cross-correlation	$3rc+1$	1	$3(rc-1)$	0	1 sqrt
Cross Covariance	$3rc+8$	1	$5rc-3$	1	1 sqrt
Root Mean Square	$rc+1$	0	$rc-1$	rc	1 sqrt
Summed Absolute Valued Diff's	0	0	$rc-1$	rc	rc abs
Variance of the Differences	$rc+2$	1	$2(rc-1)$	$rc+1$	0
Stochastic Sign Change	$r(c-1)$	1	$(r-1)(c-2)$	rc	$r(c-1)$ sgn
Deterministic Sign Change	$r(c-1)$	1	$rc+(r-1)(c-2)$	rc	$r(c-1)$ sgn

where $\mathbf{U} = [u_1 \cdots u_N]^T$ are the constraints and the real symmetric $N \times N$ matrix $[\mathbf{A}]$ is given by

$$[\mathbf{A}]_{ij} = \sum_w \left(\frac{\overline{\mathbf{X}}_i(w)\mathbf{X}_j(w)}{\sum_{k=1}^N \alpha_k S_{\Psi\Psi k}(w) + S_{\lambda\lambda}(w)} \right).$$

5 Efficient Template Matching Algorithms

Table 1 gives the computational complexity of the measures of match described in section 2. The sequential similarity detection algorithms discussed in section 2 are not represented in table 1, since techniques for increasing their efficiency are an integral part of the measure of match. The computational complexity is described in terms of the number of operations performed when calculating a measure of match between two images with r rows and c columns.

In this section we consider methods for improving the efficiency of template matching. For purposes of comparison, we assume the following scenario: Template f and sample g are digitised images with r rows and c columns. The sample images, g, are extracted from each position in a search area, s, which has R rows and C columns. Unless it is stated otherwise, we consider template matching using an exhaustive search, performed by translating the the template to each of the $(R-r+1)(C-c+1)$ positions in the image, and calculating a measure of match between the template and the sample at that point of overlap.

The operations required for template matching using the measures of match in table 1 can be calculated by simply multiplying the numbers of operations required by the size of the search area. However, it is immediately possible to increase efficiency by noting that, since the template does not change as it is translated, any term involving only the template (eg. the term $\sum_A f_{ij}^2$ in the cross-correlation) can be calculated beforehand. It is also possible to use redundancy in the calculation of terms for adjacent measures of match to improve efficiency. For example: Serial computation of the $\sum g_{ij}^2$ term for all positions in the image requires $(R-r+1)(C-c+1)rc$ multiplications and $(R-r+1)(C-c+1)(rc-1)$ additions. The computational requirement for $\sum_A g_{ij}^2$ can be reduced by making use of the fact that

when moving from g to g' one column to the right, only one column in g' holds information that wasn't in sample g. If we keep a separate partial sum, $S(j) = \sum g^2$, for each column j

$$S(j) = \sum_{i=1}^{r} g_{ij}^2$$

where

$$\sum_{i=1}^{r}\sum_{j=1}^{c} g_{ij}^2 = \sum_{j=1}^{c} S(j)$$

then $\sum g'^2$ for the next position (the sample g' one column to the right) can be calculated by subtracting the contribution of the first column in the previous position ($S(0)$) and adding the contribution of the last column ($S'(c)$) in the new position, to $\sum g^2$ (at the previous position).

The number of additions required for this algorithm [15] is

$$4RC - Rc - 3Cr + rc + 5C + 3R - 2c - 3r + 4.$$

Therefore, in order to calculate $\sum g^2$ for every position in a 512×512 image where the samples g are 32×32 (ie. $R = C = 512$ and $r = c = 32$), only 1 addition is required for every 240 additions required by the algorithm that does not eliminate the redundancy.

Further strategies for reducing the expected computational cost of various measures of match are described in the sections that follow.

5.1 Fast Fourier Transform Techniques

If template matching is to be performed in a search area, using cross-correlation and an exhaustive search (ie. translating the template over every position in the search area), Fast Fourier transform (FFT) operations can be used to calculate the cross-correlation surface [16]. Consider the cross-correlation Φ of one dimensional continuous signals f and s:

$$\Phi = \int f(t)s(t-k)dt.$$

By the convolution theorem it can be shown that

$$\Phi = FT^{-1}\left[F(f)S^*(f)\right]$$

where F and S are the Fourier transforms of f and s respectively.

The corresponding result for two dimensional digitised $n \times n$ template f (ie. $n = r = c$) and $N \times N$ search area s (ie. $N = R = C$) allows the cross-correlation surface to be calculated as

$$\Phi(k,l) = FT^{-1}\left[F(\underline{w})S^*(\underline{w})\right] \tag{25}$$

where

$$\begin{aligned} k &= 0, 1, \cdots, N-1 \\ l &= 0, 1, \cdots, N-1. \end{aligned}$$

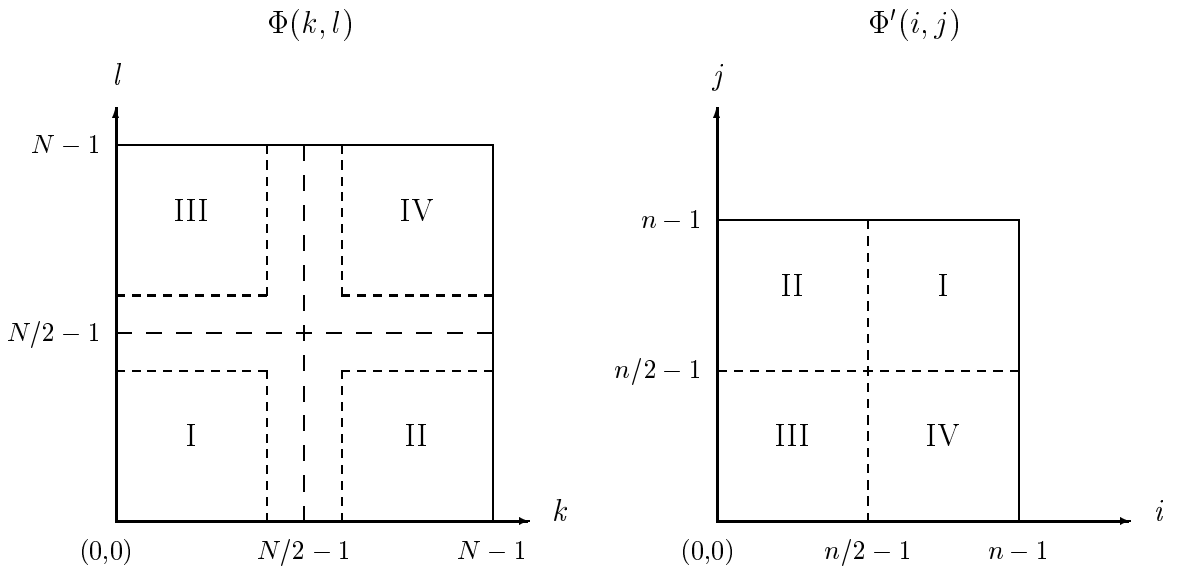

Figure 1: Extracting the cross-correlation surface from the Fourier Transform of the Image and padded template

In order to use an FFT, the image dimensions (N) must be powers of 2. Therefore it is necessary to pad the template with zeros in order to make it the same size as the image.

By using the Fourier transform, one is asserting that both the template and image are periodic, and this is represented in the result of (25). The non-cyclical portion of the cross-correlation surface $\Phi'(i,j)$, where

$$\begin{aligned} i &= 0, 1, \cdots, N-n \\ j &= 0, 1, \cdots, N-n, \end{aligned}$$

is extracted from $\Phi(k,l)$ as illustrated in figure 1.

5.2 The Cumulative Mismatch Threshold

Consider template matching with an additive measure of mismatch, where a threshold T can be specified such that values above T indicate a mismatch. As an example of an additive measure of mismatch, consider the SAVD:

$$d(f,g) = \sum_A |f_{ij} - g_{ij}|,$$

where f is the template and g is a template sized sub-image in the search area. If for this position of the template, a partial sum of the measures of mismatch for the individual pixels exceeds T, it is not necessary to continue, since it has already been determined that there

is a mismatch at this position. Therefore, the overall computation can be reduced if the computation saved by not applying the entire template at every position is more than the additional computation required to check whether the partial sum has exceeded the threshold after applying each pixel in the template.

5.2.1 Ordered Application of the Template Pixels

The computation could be reduced further if it were possible to process the template pixels in an order which results in the partial measure of mismatch exceeding the mismatch threshold after visiting as few pixels as possible. Nagel and Rosenfeld [17] use the expected difference as a function of pixel intensity for each template pixel for this purpose. The expected difference for pixel intensity i is estimated by

$$E[d(i)] = \sum_j p_i \cdot |i - j|$$

where p_i is the probability of encountering a pixel intensity of i in the image. Prior to the template matching process, $E[d(i)]$ is calculated for each template pixel, and during template matching the pixels in the template are visited in order of decreasing $E[d(i)]$.

5.2.2 The Adaptive Threshold

If is not possible to specify a threshold T, but a relative measure of mismatch is being used (ie. we are looking only for the best match in the search area), then an adaptive threshold can be used. This threshold is set to be the best measure of mismatch calculated in the search area to date.

5.3 Two Stage Template Matching

Vanderbrug and Rosenfeld [18] (using SAVD), and Goshtasby *et al* [19] (using cross-correlation) describe the following strategy for reducing the computation required of template matching: In each position, calculate a partial measure of mismatch for a subset of the template pixels. Only if the partial measure match exceeds a threshold T, is the measure of mismatch calculated for the entire template.

The computational cost of this procedure is given by

$$C = C_{subset} + p \cdot C_{remainder}$$

where C_{subset} is the computational cost of applying the subset, p is the probability of an above threshold match, and $C_{remainder}$ is the computational cost of applying the remainder of the template. Since in general C_{subset} increases, and $p \cdot C_{remainder}$ decreases for an increasing threshold, there should be an optimum value for T than minimises C. It must also be taken into account that the smaller the value of T, the higher the probability of overlooking a correct

match. The positions of the subset pixels can be chosen using a strategy similar to the one described in section 5.2.1, in order to minimise the expected computational cost.

Rosenfeld and Vanderbrug [20] describe a similar two stage approach, referred to as c*coarse-fine* template matching, where a low resolution template is applied in the first stage, followed by the full resolution template where the match threshold is exceeded.

5.4 Efficient Search Strategies

The three step search is a technique used in the coding of digital image sequences, which is useful where the search area is not much larger than the template itself, although the concept could be expanded for larger search areas. Instead of doing an exhaustive search, the position of best match is narrowed down in three steps. This process is illustrated in figure 2.

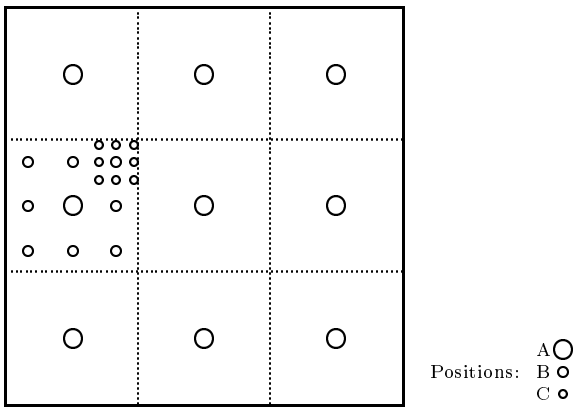

Figure 2: Order of operation in the three step search

Initially, template matching is done at the nine positions labelled as A. The second step compares the best match of the first step with template matching performed at the positions labelled as B. The best match is now narrowed down to a 3×3 region, and template matching at the C positions determines the best match.

A drawback of this search procedure is that it sometimes finds a local maximum on the measure of match surface.

References

[1] W. Niblack, *An Introduction to Digital Image Processing.* Prentice/Hall International, 1986.

[2] A. K. Jain, *Fundamentals of Digital Image Processing.* Prentice-Hall, 1989.

[3] R. C. Gonzales and R. E. Woods, *Digital Image Processing.* Addison-Wesley, 1992.

[4] W. K. Pratt, *Digital Image Processing.* John Wiley and Sons, 1991.

[5] W. K. Pratt, "Correlation techniques of image registration," *IEEE transactions on Aerospace and Electronic Systems*, vol. AES-10, pp. 353–8, May 1974.

[6] D. I. Barnea and H. F. Silverman, "A class of algorithms for fast digital image registration," *IEEE Transactions on Computers*, vol. C-21, pp. 179–86, Feb. 1972.

[7] A. Venot, J. F. Lebruchec, and J. C. Roucayrol, "A new class of similarity measures for robust image registration," *Computer Vision, Graphics, and Image Processing*, vol. 28, pp. 176–84, 1984.

[8] A. Gersho and R. M. Gray, *Vector Quantization And Signal Compression.* Communications and Information Theory, Norwell, Massachusetts: Kluwer Academic Publishers, 1992.

[9] Y. Fisher, E. Jacobs, and R. Boss, "Fractal image compression using iterated transforms," in *Image and Text Compression* (J. A. Storer, ed.), ch. 2, pp. 35–61, Norwell, Massachusetts: Kluwer Academic Publishers, 1992.

[10] J. P. Oakley and M. J. Cunningham, "A function space model for digital image sampling and its application in image reconstruction," *Computer Vision, Graphics, and Image Processing*, vol. 49, pp. 171–97, 1990.

[11] J. Ben-Arie and K. R. Rao, "A novel approach for template matching by nonorthogonal image expansion," *IEEE Transactions on circuits and Systems for Video Technology*, vol. 3, pp. 71–84, Feb. 1993.

[12] F. G. Stremler, *Introduction to Communications Systems.* Addison-Wesley series in electrical engineering, Addison-Wesley, 2 ed., 1982.

[13] M. S. Longmire, "Simulation of mid-infrared clutter rejection. 1. one-dimensional lms spatial filter and adaptive threshold algorithms," *Applied Optics*, vol. 21, p. 3819, Nov. 1982.

[14] K. R. Rao and J. Ben-Arie, "Nonorthogonal image expansion related to optimal template matching in complex images," *CVGIP: Graphical Models and Image Processing*, vol. 56, pp. 149–60, Mar. 1994.

[15] L. J. Siegel, H. J. Siegel, and A. E. Feather, "Parallel processing approaches to image correlation," *IEEE Transactions on Computers*, vol. C-31, pp. 208–18, Mar. 1982.

[16] P. E. Anuta, "Spatial registration of multispectral and multitemporal digital imagery using fast fourier transform techniques," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. GE-8, pp. 353–68, Oct. 1970.

[17] R. N. Nagel and A. Rosenfeld, "Ordered search techniques in template matching," *Proceedings of the IEEE*, pp. 242–4, Feb. 1972.

[18] G. J. Vanderbrug and A. Rosenfeld, "Two-stage template matching," *IEEE Transactions on Computers*, vol. C-26, pp. 384–93, Apr. 1977.

[19] A. Goshtasby, S. H. Gage, and J. F. Batholic, "A two-stage cross correlation approach to template matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-6, pp. 374–78, May 1984.

[20] A. Rosenfeld and G. J. Vanderbrug, "Coarse-fine template matching," *IEEE Transactions on Systems, Man, and Cybernetics*, pp. 104–7, Feb. 1977.