Invariance in Template Matching

G. S. Cox and G. de Jager (Member IEEE)
Department of Electrical Engineering
University of Cape Town
South Africa

Abstract

This paper introduces variations on the template matching theme that extend its usefulness by providing invariance to mean intensity-level variations, certain geometric transformations and partial obscurations of the target object in the image. First order statistics of the pixel by pixel differences between the template and the image are used as a match measure in order to provide invariance to mean intensity-level differences. It is shown how scale and translation invariance can be achieved by using the Fourier-Mellin transform. Rotation invariance is achieved by pre-processing with a polar coordinate transform before using a modified Fourier-Mellin transform. Tesselated sub-templates are used to ignore the influence of obscurations on an otherwise matching object when calculating the match measure.

1 Introduction

Template matching is an essential pattern recognition tool. The template is a pattern representing an object which is to be detected. In image processing, the two-dimensional template is applied to an image by determining a measure of match between the template and template-sized subimages centred on every coordinate position in the image. An intermediate image of the match measure for each position is created. Peaks in this intermediate image should represent positions where there is a high degree of match between the template and the image.

The traditional methods for determining the measure of match limit the application of template matching to problems where the scale and orientation of the target pattern is fixed, and where the mean intensity level of the template and subimage are similar. Applying multiple templates with patterns in various orientations and different scales, and with a range of offset intensity levels can overcome this problem to some extent.

We present a match measure, pre-processing transforms and a new way of applying the template that is invariant to different properties of the template and subimage. We propose a measure of match that is invariant to mean intensity-level differences using the *variance of the differences* between the template and the image. Scale and translation invariance is achieved by using the Fourier-Mellin transform. Invariance to rotation of the target pattern is achieved by using a polar coordinate transform, followed by a modified Fourier-Mellin transform. Additionally, a procedure for determining a match measure using multiple subtemplates that ignores a certain amount of obscuration of the target pattern is introduced.

In this paper *template matching* refers only to techniques that do not require the extraction of shape information from the template and the image. Therefore, we do not consider techniques such as Fourier descriptors which identify and use properties of the shape of objects in the template. We limit the discussion to two dimensional, single band images.

2 Traditional measures of match

Consider the application of the template to a single position, or template-sized subimage, in an image. Let $\tilde{f}(x,y)$ and $\tilde{g}(x,y)$ give the pixel intensity as a function of position with the origin the centred in the template and subimage respectively. 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.

Notationally we define

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

where the template and subimage are $N \times N$ pixels.

Common measures of mismatch are the *sum of the absolute differences* (SAVD)

$$\sum_A |f_{ij} - g_{ij}| \qquad (1)$$

and the *sum of squared differences* (SSD)

$$\sum_A (f_{ij} - g_{ij})^2 . \tag{2}$$

Since expression 2 can be written

$$\sum_A f_{ij}^2 + \sum_A g_{ij}^2 - \sum_A f_{ij} g_{ij}$$

and $\sum_A f_{ij}^2$ is constant,

$$\sum_A f_{ij} g_{ij}$$

is a measure of match as long as the average energy

$$\frac{\sum_A g_{ij}^2}{N^2}$$

in the image is constant.

In general, $\sum_A g_{ij}^2$ is not constant in the image and the *normalised cross correlation*

$$C = \frac{\sum_A f_{ij} g_{ij}}{\sqrt{\sum_A g_{ij}^2}} \tag{3}$$

is used.

3 Mean intensity level invariance

For mean intensity level invariant template matching we propose a measure of match based on the variance of the pixel by pixel differences between the template and the subimage it is applied to. The match measure is given by

$$M_{vod} = \frac{1}{1 + \sigma^2} \tag{4}$$

where σ^2, the variance of the differences (VOD), is given by

$$\sigma^2 = \frac{\sum_A \left[(f_{ij} - g_{ij}) - \mu_d\right]^2}{n - 1} \tag{5}$$

and μ_d is the mean of the differences, given by

$$\mu_d = \frac{\sum_A (f_{ij} - g_{ij})}{n}$$

where $n = N^2$ is the number of pixels in the $N \times N$ template. For a perfect match $M_{vod} = 1$ and for a mismatch $M_{vod} \to 0$.

Figure 1 shows two one-dimensional signals, f and g, that differ only by a constant offset. SAVD and SSD would give mismatches between f and g of $N(a-b)$ and $N(a-b)^2$ respectively, whereas VOD would indicate a perfect match.

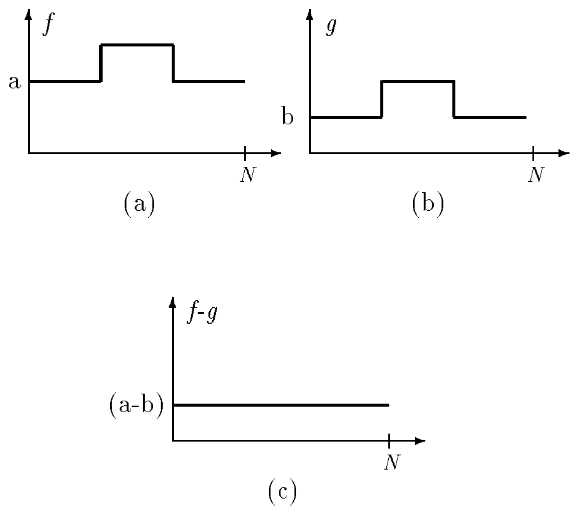


Figure 1: Constant-offset invariance for one dimensional signals a) Template f. b) Subimage g. c) $f-g$. $M_{vod} = 1$.

The more efficient algebraic equivalent to equation 5,

$$\sigma^2 = \frac{n \cdot \sum_A (f_{ij} - g_{ij})^2 - \left(\sum_A (f_{ij} - g_{ij})\right)^2}{n(n-1)},$$

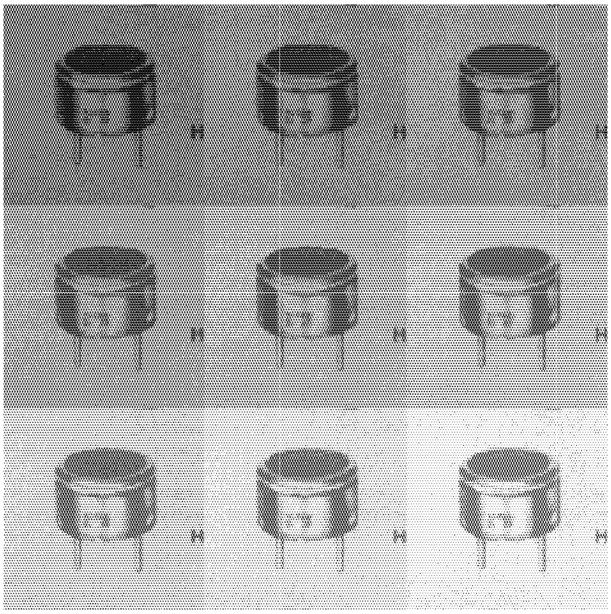
is used for computing the variance.

Figure 2 demonstrates the strength of the VOD match measure. An image of an electronic component was given nine different offsets and tessellated onto a larger image. This image was then degraded with Gaussian noise. The original undegraded template was applied to the image. Figures 2b and 2c show the match surfaces for the SAVD and VOD match measures respectively. The SAVD surface has only one peak where the mean intensity level of the template and image were similar, whereas the VOD surface has a peak for each object.

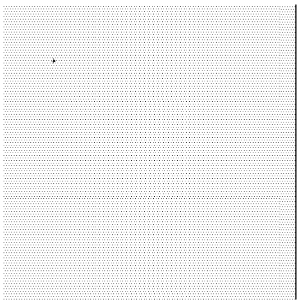
4 Scale invariance

In cases where the target objects in the image are sufficiently far apart and the background is uniform the Fourier-Mellin (FM) transform [1–3] can provide a scale invariant representation of the template and the subimages (at each position in the image) before a measure of match is calculated. The FM transform of an image consists of four stages:

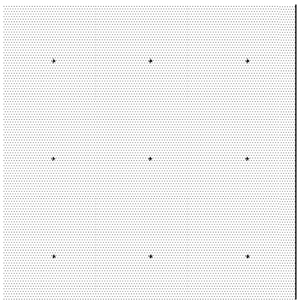
1. The Discrete Fourier Transform (DFT) magnitude of the image is calculated. By the circular shift property of the DFT, this is a centred, translation invariant representation of the original image. This implies that the translation limits are the borders of the image.



(a) Object images with nine different mean intensity levels



(b) SAVD match surface



(c) VOD match surface

Figure 2: Demonstration of VOD template matching.

2. The DFT magnitude image is normalised by its value at the origin in order to cancel the multiplicative effect of scale changes in the original image. Consider the continuous two dimensional signal $f(x,y)$ and its Fourier transform $F(u,v)$. If the scale factors a and b are applied to the x and y dimensions respectively, then by the scaling property

$$f(ax,by) \Longleftrightarrow \frac{1}{|ab|} F\left(\frac{u}{a}, \frac{v}{b}\right). \qquad (6)$$

Normalisation compensates for the effect of the factor $\frac{1}{|ab|}$.

3. The axes of the positive quadrant of the normalised image are warped onto a logarithmic scale. This converts scaling in the original image to translation. We consider again the continuous two dimensional signal $f(x,y)$. Let $\tilde{F}(u,v)$ be $F(u,v)$ with logarithmically scaled axes. That is,

$$\tilde{F}(u,v) = F(\ln u, \ln v) \qquad (7)$$

Using equations 6 and 7, it can be shown that the normalised Fourier transform of $f(ax,by)$ with logarithmically scaled axes is related to the normalised Fourier transform of $f(x,y)$ by

$$\tilde{F}_{norm}(u,v) = F_{norm}(\ln \frac{u}{a}, \ln \frac{v}{b}) \qquad (8)$$

The right hand side of equation 8 can be rewritten as

$$F_{norm}(\ln u - \ln a, \ln v - \ln b)$$

which is a translation of

$$F_{norm}(\ln u, \ln v).$$

Figure 3 shows an example of output for this stage of the FM transform.

4. The DFT magnitude of the logarithmically scaled image is calculated, providing a scale invariant representation of the original image.

The scale variations tolerated by the FM transform are limited by the resolution of the image and the size of the image window. Altmann and Rietbock [3] proposed the following criteria for no significant loss of information in the DFT of stage one:

$$n_t \leq N/4$$

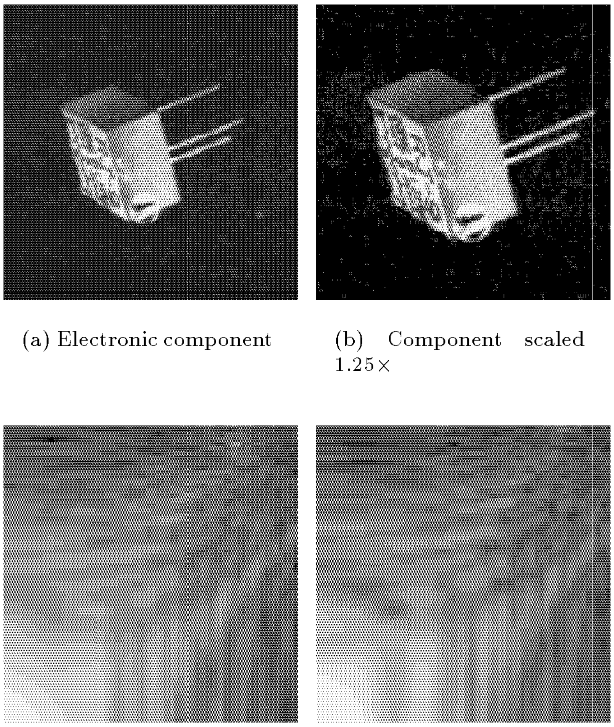
$$n_d \geq 4$$

for a $N \times N$ image , where n_t is the total size of the object and n_d is the size of its smallest detail.

Since the template and each subimage can be regarded as centred, translation invariant images, the DFT in stage one is not required for normal template matching. The template and subimage axes are warped onto a logarithmic scale, and Fourier transformed to obtain a scale invariant representation. The translation invariance property of the full FM transform can be utilised to reduce the computation that is required of normal template matching. Instead of applying the template exhaustively through the image, the $N \times N$ template can be moved at $N/2$ increments to narrow down the search space.

5 Rotation invariance

A rotation and scale invariant representation of the template and the subimage can be generated by simply preceding the procedure described in the previous



(a) Electronic component (b) Component scaled 1.25×

(c) (d)

Figure 3: Results of stage three of the Fourier-Mellin transform. The DFT followed by logarithmic scaling of the positive quadrant axes for an electronic component at two scales. Scale differences between (a) and (b) become translation in (c) and (d).

section by a polar coordinate transform, with the origin at the centre of the template or subimage. This converts rotation to a circular shift. The r axis in the polar coordinate system is then warped onto a logarithmic scale and scaling becomes translation. The DFT magnitude is then calculated for a rotation and scale invariant representation of the image.

Wechsler and Zimmerman [4] describe a similar approach in terms of a conformal complex-log mapping. The image is mapped onto a complex plane where each pixel (x, y) is represented mathematically by

$$z = x + jy.$$

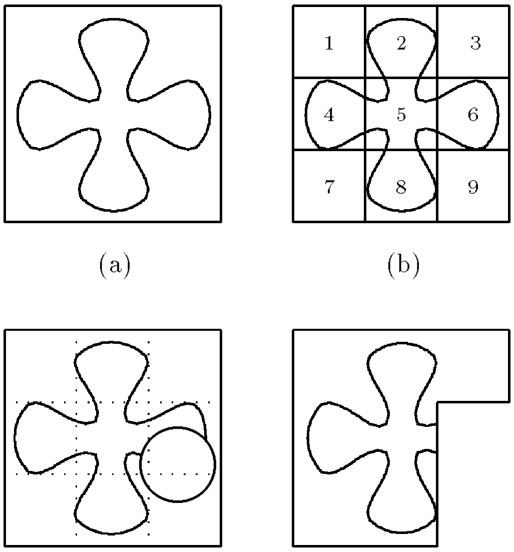
The complex-log mapping, which transforms an image from rectangular to polar exponential coordinates, is given by

$$w = \ln(z) = ln(|z|) + j\theta_z.$$

Rotation and scaling become translation in the transform domain.

6 Obscuration invariance

Problems arise in template matching when the target object is obscured by another object in the image. If the object is only partially obscured, it is possible to ignore the obscuration by dividing the template into n smaller subtemplates, applying them separately, and using only the $k < n$ best matching subtemplates to calculate the final measure of match. Ideally, the $n-k$ subtemplates that are discarded will contain the obscuration.



(a) (b)

(c) (d)

Figure 4: Obscuration invariance. a) Template of a flower. b) Tessellated subtemplates ($n = 9$). c) Image instance of partially obscured flower. d) Template after removing mismatching subtemplates ($k = 7$).

As an example consider figure 4. Figure 4a shows a template containing a flower. Figure 4b shows the template tessellated into nine subtemplates. Figure 4c shows a subimage containing an instance of the flower with one of the 'petals' partially obscured by a disc. The template matching procedure applies each of the nine templates. Assuming $k = 7$, the two subtemplates with the lowest measure of match would be discarded. These will be subtemplates 6 and 9, which contain the disc. The remaining seven subtemplates are combined (figure 4d) to recalculate an overall measure of match.

7 Conclusions

We have demonstrated a selection of techniques that provide invariance to certain properties of the template and the image. In each case, this invariance enables template matching to be applied under more general circumstances. We have proposed the *variance of differences* measure of match in order to provide mean intensity level invariance, and have shown how the Fourier-Mellin transform can be used in the context of template matching to provide scale and rotation invariance. Multiple subtemplate matching has been introduced as a strategy that provides a certain amount of obscuration invariance.

References

[1] D. Casasent and D. Psaltis, "Scale invariant optical correlation usin Mellin transforms," *Optical Communications*, vol. 17, pp. 59–63, Apr. 1976.

[2] D. Casasent and D. Psaltis, "New optical transforms for pattern recognition," *IEEE Proceedings*, vol. 65, pp. 77–84, Jan. 1977.

[3] J. Altmann and H. J. P. Rietbock, "A fast correlation method for scale- and translation-invariant pattern recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-6, pp. 46–57, Jan. 1984.

[4] H. Wechsler and G. L. Zimmerman, "2-D invariant object recognition using distributed associative memory," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 10, pp. 811–821, Nov. 1988.