OPTIMAL MORPHOLOGICAL APPROACHES TO IMAGE MATCHING AND OBJECT DETECTION

Petros Maragos

Division of Applied Sciences, Harvard University, Cambridge, MA 02138

ABSTRACT: It is shown that minimizing the ℓ_1 matching error is equivalent to maximizing a nonlinear signal correlation (a sum of minima), which is related to morphological filtering. This approach is optimum in many formulations of the image matching and object detection problem. Further, a closely related approach is outlined for object detection using rank order filters.

1 Introduction

Image matching is a fundamental issue that arises in many problems of computer vision including stereomapping, motion detection, template matching, and object detection. A coarse classification of image matching problems involves two cases: 1) The two images (or image parts) to be matched have similar sizes; a characteristic example is the correspondence problem in binocular stereopsis, photogrammetry, and motion detection. 2) The two images have different sizes, where the small image is a pattern (feature, object) to be matched against or detected in the larger image; examples include template matching and object detection in the presence of noise. We refer to the first class of problems as *image matching* whereas we view the second class as *object detection.* So far the vast majority of efforts to solve these problems has been based on minimizing the mean squared error, which leads to maximizing an image correlation [1,2]. The popularity of this approach is due to the mathematical tractability of the squared error metric and the easy implementations of the correlators. It was motivated by the success of the equivalent theory of matched filtering in communications. Some researchers, e.g., [3], have used the ℓ_1 matching metric. Despite the attractive mathematical tractability of the ℓ_2 norm, the ℓ_1 norm has some important practical advantages.

In this paper we link the ℓ_1 matching metric with a nonlinear image correlation, which consists of a sum of minima. We formally define this min-correlation and relate it to morphological filtering [4]-[7] as well as to linear correlation and Fourier analysis. Then we develop some approaches for image matching and object detection using maximization of this min-correlation, morphological and rank order operations.

This research was supported by the National Science Foundation under Grant MIPS-86-58150 with matching funds from Bellcore, Xerox, and an IBM Departmental Grant, and in part by ARO under Grant DAALO3-86-K-0171.

2 Image Matching via the ℓ_1 Norm

Let $f(\vec{n})$ and $g(\vec{n})$ be the two real-valued image signals to be matched. Both f and g are defined on the discrete Cartesian grid of integer coordinate pairs $\vec{n} = (n_1, n_2) \in \mathbf{Z}^2$. The general restrictions on f and g are

A1. *Summability:* $||f||_1, ||g||_1 < \infty$, where $||h||_1 = \sum_{\vec{n}\in \mathbf{Z}^2} |h(\vec{n})|$ is the ℓ_1 norm of the 2-D sequence h.

A2. *Nonnegativity:* $f(\vec{n}), g(\vec{n}) \geq 0 \ \ \forall \vec{n} \in \mathbf{Z}^2$.

The signals f and g represent either two images of similar size to be matched or an image f and a smaller template/object g whose presence in f is to be detected. In general, we expect that $f(\vec{n})$ must be shifted by a pixel vector $\vec{k} = (k_1, k_2)$ to bring it to registration with g or to detect the presence of g in f at a certain image location. (We also assume that rotational or scaling distortions of f or g either do not occur or have already been undone.) As a matching measure then between g and the shifted f we consider the sum of absolute differences

$$E(\vec{k}) = \sum_{\vec{n}\in W} |f(\vec{n}+\vec{k}) - g(\vec{n})| \tag{1}$$

where W is a subset of $\mathbf{Z}^2$ related to the domains of f and g. An obvious choice for W is to be equal to the whole plane $\mathbf{Z}^2$ in which case E becomes the ℓ_1 norm of the difference signal between g and f (shifted). To minimize the ℓ_1 error norm $||f-g||_1$ rather than the popular ℓ_2 error norm $||f-g||_2$, where $||f||_2 = \sqrt{\sum_{\vec{n}\in \mathbf{Z}^2} f^2(\vec{n})}$, has the following three advantages: First, the ℓ_1 norm is easier to compute than the ℓ_2, because the former involves only additions and sign changes whereas the latter involves additions and multiplications. Second, data approximation based on the ℓ_1 norm is more ***robust*** in the presence of outliers or non-Gaussian noise distributions. Third, since the ℓ_1 is a larger norm than ℓ_2, minimizing the ℓ_1-based error $||f-g||_1 > ||f-g||_2$ may result in a better match between f and g than minimizing $||f-g||_2$.

Since $|a-b| = a + b - 2\min(a,b)$ for any reals a, b,

$$E(\vec{k}) = \sum_{\vec{n}\in W} f(\vec{n}+\vec{k}) + \sum_{\vec{n}\in W} g(\vec{n}) - 2\sum_{\vec{n}\in W} \min[f(\vec{n}+\vec{k}), g(\vec{n})] \tag{2}$$

Obviously, a "good" matching between f and g will manifest itself via a minimum matching error $E(\vec{k})$. There are

CH2664-1/88/0000/0695$01.00

many ways to make the sum $\sum_{\vec{n}} f(\vec{n}+\vec{k}) + g(\vec{n})$ not affect the minimization of $E(\vec{k})$. In such cases minimizing $E(\vec{k})$ is equivalent to maximizing the nonlinear correlation $\sum_{\vec{n}} \min[f(\vec{n}+\vec{k}), g(\vec{n})]$, which we call *morphological correlation* for reasons explained in Section 3.

Case 1: Let f and g be images of similar size which is much larger than the possible shifts $\vec{k}$. Since f and g are zero outside their domains, the matching window W can be set equal to $\mathbf{Z}^2$. In this case the factor $\sum_{\vec{n}\in W} f(\vec{n}+\vec{k})+g(\vec{n})$ equals the sum of the areas under f and g; hence it is constant and does not affect the minimization of $E(\vec{k})$ with respect to $\vec{k}$. Then

$$\min_{\vec{k}} \; E(\vec{k}) \Longleftrightarrow \max_{\vec{k}} \; \sum_{\vec{n}\in W} \min[f(\vec{n}+\vec{k}), g(\vec{n})] \, . \tag{3}$$

Thus the shift of f with respect to g that minimizes their ℓ_1 matching error norm is the shift that maximizes their *global* morphological correlation.

Case 2: Assume (e.g., as in solving correspondence problems in stereo or motion) that f and g are two small image segments whose size is similar but comparable to the shifts $\vec{k}$. Now we set W equal to the finite domain of $g(\vec{n})$; hence the factor $\sum_{\vec{n}} g(\vec{n})$ remains constant. *(a)* If the factor $\sum_{\vec{n}} f(\vec{n}+\vec{k})$ does not vary much, we can consider the minimization of $E(\vec{k})$ equivalent to maximizing the *local* (i.e., over W) morphological correlation of the segments $g(\vec{n})$ and $f(\vec{n}+\vec{k})$. *(b)* A more robust approach is to minimize a normalized error $E_{norm}(\vec{k}) = E(\vec{k})/[\sum_{\vec{n}\in W} f(\vec{n}+\vec{k}) + g(\vec{n})]$. Then from (2)

$$\min_{\vec{k}} E_{norm}(\vec{k}) \Longleftrightarrow \max_{\vec{k}} \frac{\sum_{\vec{n}\in W} \min[f(\vec{n}+\vec{k}), g(\vec{n})]}{\sum_{\vec{n}\in W} f(\vec{n}+\vec{k}) + g(\vec{n})} \tag{4}$$

Thus to minimize the normalized mean absolute error (locally over W) is equivalent to maximize the morphological correlation between f and g *normalized* by the sum of the areas under the two image segments over the moving matching window W. Note that this normalized morphological correlation has always values in the interval $[0, 0.5]$. *(c)* Alternatively, by subtracting from f and g their local averages over W and thus creating new image segments f' and g', the factor $\sum_{\vec{n}\in W} f'(\vec{n}+\vec{k}) + g'(\vec{n})$ becomes zero. Then (3) applies again if we replace f, g, E with $f', g', E' = ||f'-g'||_1$.

We conducted some preliminary experiments to compute disparities from stereoscopic image pairs by using the approaches (a),(b),(c) and to compare the latter with the corresponding approaches that use linear correlation. These experiments indicate that in approach (a) morphological correlation usually gives *sharper match peaks* than the linear. In (b),(c) the two correlations perform similarly, but the morphological is always computationally *faster* because it does not use multiplications.

3 Morphological Correlation

We define the *morphological cross-correlation* of $f(\vec{n})$ and $g(\vec{n})$ as the 2-D sequence

$$M_{fg}(\vec{k}) = \sum_{n\in\mathbf{Z}^2} \min[f(\vec{n}+\vec{k}), g(\vec{n})] \tag{5}$$

Replacing g with f in (5) gives us the *morphological autocorrelation*, $M_{ff}(\vec{k})$, of f. Note that $M_{ff}(\vec{k}) \leq M_{ff}(\vec{0}) = \sum_{\vec{n}} f(\vec{n})$. We use the term "morphological correlation", because $M_{ff}(\vec{k})$ is equal to the area under the morphological erosion [4] of f by the 2-point structuring element $\{\vec{0}, \vec{k}\}$. Recently, areas of erosions have been used in texture description [8].

The classical linear cross-correlation of f and g is $L_{fg}(\vec{k}) = \sum_{\vec{n}\in\mathbf{Z}^2} f(\vec{n}+\vec{k})g(\vec{n})$. To relate the morphological with the linear correlation and Fourier analysis we decompose the images f and g into their *threshold binary images*

$$f_a(\vec{n}) = \begin{cases} 1, & f(\vec{n}) \geq a \\ 0, & f(\vec{n}) < a \end{cases} \tag{6}$$

where the amplitude gray level $a = 0, 1, 2, \ldots$ spans *all* the bounded range of f. Then from *all* its threshold binary images f_a, $a \geq 1$, we can exactly reconstruct f since

$$f(\vec{n}) = \sum_{a\geq 1} f_a(\vec{n}) \, , \;\; \forall \vec{n} \, . \tag{7}$$

Now for any $\vec{m}, \vec{n}$

$$\min[f(\vec{m}), g(\vec{n})] = \sum_{a\geq 1} \min[f_a(\vec{m}), g_a(\vec{n})] = \sum_{a\geq 1} f_a(\vec{m}) g_a(\vec{n}) \tag{8}$$

The second equality results because the minimum of two binary numbers is equal to their product; hence, *the morphological and linear correlation of any two binary signals concide.* (The relation between linear autocorrelation of binary images and erosion was used in [4] for structural image analysis via geometric probabilities.) Thus from (5) and (8) we obtain

$$M_{fg}(\vec{k}) = \sum_{a\geq 1} L_{f_a g_a}(\vec{k}) = \sum_{a\geq 1} \sum_{\vec{n}\in\mathbf{Z}^2} f_a(\vec{n}+\vec{k}) g_a(\vec{n}) \, . \tag{9}$$

Let now $\vec{\omega} = (\omega_1, \omega_2)$ be a frequency vector and $\stackrel{\mathcal{F}}{\longleftrightarrow}$ denote a Fourier transform pair. Then, given the pairs $f(\vec{n}) \stackrel{\mathcal{F}}{\longleftrightarrow} F(\vec{\omega})$, $g(\vec{n}) \stackrel{\mathcal{F}}{\longleftrightarrow} G(\vec{\omega})$, $f_a(\vec{n}) \stackrel{\mathcal{F}}{\longleftrightarrow} F_a(\vec{\omega})$, $g_a(\vec{n}) \stackrel{\mathcal{F}}{\longleftrightarrow} G_a(\vec{\omega})$, and $L_{fg}(\vec{k}) \stackrel{\mathcal{F}}{\longleftrightarrow} F(\vec{\omega})G(-\vec{\omega})$, the result (9) leads to

$$M_{fg}(\vec{k}) \stackrel{\mathcal{F}}{\longleftrightarrow} \sum_{a\geq 1} F_a(\vec{\omega}) G_a(-\vec{\omega}) \, . \tag{10}$$

Due to the Hermitian symmetry of $F_a(\vec{\omega})$, we get

$$M_{ff}(\vec{k}) \stackrel{\mathcal{F}}{\longleftrightarrow} \sum_{\vec{k}\in\mathbf{Z}^2} M_{ff}(\vec{k}) e^{-j\vec{\omega}\cdot\vec{k}} = \sum_{a\geq 1} |F_a(\vec{\omega})|^2 \, . \tag{11}$$

Hence, the morphological correlation between f and g is equal to the sum (over all gray levels a) of the linear (or morphological) correlations between the threshold binary images f_a and g_a. The sum of products of the Fourier transforms of f_a and g_a is the Fourier transform of the morphological correlation between f and g. Thus, the Fourier transform of the morphological autocorrelation of f, is equal to the sum of the power spectra of the thresholded versions of f.

Similarly, the mean absolute difference between f and g equals the sum (over all gray levels a) of the mean squared (or absolute) differences between the threshold binary images f_a and g_a:

$$||f-g||_1 = \sum_{a\geq 1} ||f_a - g_a||_1 = \sum_{a\geq 1} (||f_a - g_a||_2)^2 . \quad (12)$$

The proof of (12) proceeds as follows:

$$\begin{aligned} E(\vec{k}) &= \sum_{\vec{n}\in W} \Big|\sum_{a\geq 1} f_a(\vec{n}+\vec{k}) - \sum_{a\geq 1} g_a(\vec{n})\Big| \quad (13)\\ &= \sum_{\vec{n}\in W}\sum_{a\geq 1} |f_a(\vec{n}+\vec{k}) - g_a(\vec{n})| \quad (14)\\ &= \sum_{a\geq 1}\sum_{\vec{n}\in W} |f_a(\vec{n}+\vec{k}) - g_a(\vec{n})| . \quad (15)\end{aligned}$$

Summation and absolute value can be interchanged in (13), (14) because all the numbers $f_a - g_a$ are of the same sign [9]. Further, since $x = |f_a(\vec{n}+\vec{k}) - g_a(\vec{k})| \in \{0,1\}$, $x = x^2$ and hence the proof of (12) is complete.

4 Template Matching

Here g has a domain G much smaller than f and is viewed as a template (e.g., an image feature) whose presence is to be detected at various locations $\vec{k}$ in the larger image f. A flexible and robust approach is to set the matching window W equal to a finite set that contains G (see Fig. 1 for an example). This corresponds to requiring not only that g matches f over G but also that the surroundings of g match the surroundings of f over the window $W \setminus G = \{\vec{n} \in W : \vec{n} \notin G\}$. This causes no loss of generality since we can always set $W = G$. Next we present various approaches to solve this template matching problem. Some of the ideas discussed in the section on image matching apply here too.

Dual Template Approach

Since for a fixed template g $\sum_{\vec{n}\in W} g(\vec{n})$ is a constant factor equal to the area under g, to minimize $E(\vec{k})$ of (2) is equivalent to minimize

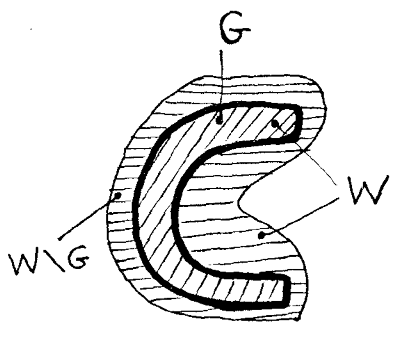

Figure 1. *G is the domain of a template g. W is a window set containing G. $W \setminus G$ is the local background of G.*

$$E_g(\vec{k}) = \sum_{\vec{n}\in W} f(\vec{n}+\vec{k}) - 2\sum_{\vec{n}\in W} \min[f(\vec{n}+\vec{k}), g(\vec{n})] . \quad (16)$$

Now threshold f and g, consider their threshold binary signals, and define a *dual template* $g^\star$ implicitly via its threshold binary versions

$$g_a^\star(\vec{n}) = 1 - g_a(\vec{n}),\ \vec{n}\in W;\quad g_a^\star(\vec{n}) = 0,\ \vec{n}\notin W, \quad (17)$$

for all $a \geq 1$; see Figs. 2a,b. Then from (8) we get

$$\begin{aligned} E_g(\vec{k}) &= \textstyle\sum_{\vec{n}}\sum_a [f_a(\vec{n}+\vec{k}) - 2 f_a(\vec{n}+\vec{k}) g_a(\vec{n})]\\ &= \textstyle\sum_{\vec{n}}\sum_a [f_a(\vec{n}+\vec{k})[g_a(\vec{n}) + g_a^\star(\vec{n})] - 2 f_a(\vec{n}+\vec{k}) g_a(\vec{n})]\\ &= \textstyle\sum_{\vec{n}}\sum_a [f_a(\vec{n}+\vec{k}) g_a^\star(\vec{n}) - f_a(\vec{n}+\vec{k}) g_a(\vec{n})]\\ &= -M_{fg}(\vec{k}) + \textstyle\sum_{\vec{n}\in W} h(\vec{n})\end{aligned}$$

where $h(\vec{n}) = \sum_{a\geq 1} f_a(\vec{n}+\vec{k}) g_a^\star(\vec{n})$. Again minimizing $E_g(\vec{k})$ is equivalent to maximizing the difference $M_{fg}(\vec{k}) - \sum_{\vec{n}} h(\vec{n})$, but the signal $h(\vec{n})$ cannot in general be expressed in terms of f and g except for the special case below:

Two-level template: Let $g(\vec{n}) = T \geq 1$ for $\vec{n} \in G$ and 0 elsewhere. Then all the threshold binary versions of g become identical, and likewise for its dual template; see Figs. 2c,d. This leads to the dual template having a simple expression in terms of g, i.e., $g^\star(\vec{n}) = T - g(\vec{n})$ for $\vec{n} \in W \setminus G$ and 0 elsewhere. Then the signal h takes the simple form $h(\vec{n}) = \min[f(\vec{n}+\vec{k}), g^\star(\vec{n})]$. From all the above we get

$$-E_g(\vec{k}) = M_{fg}(\vec{k}) - M_{fg^\star}(\vec{k}) . \quad (18)$$

Hence, minimizing $E_g(\vec{k})$ is equivalent to maximizing the difference between the morphological correlations of f with the original template from that with the dual template $g^\star$.

Binary Template/Binary Image: Assume now that g is binary, i.e., two-level as in (18) but with $T = 1$; further f

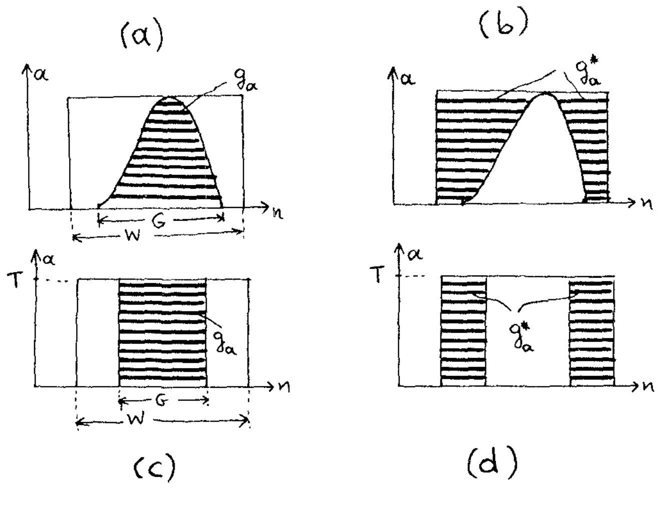

Figure 2. (a) A 1-D multilevel template g and its binary threshold versions g_a. (b) The dual binary threshold versions $g_a^\star$. (c)-(d) Same as in (a)-(b) but for a two-level template g.

is binary. Then, (18) becomes

$$-E_g(\vec{k}) = \sum_{\vec{n}\in W} f(\vec{n}+\vec{k})[g(\vec{n}) - g^\star(n)]\,. \qquad (19)$$

As is well-known [2], the binary $g^\star$ can be viewed as the "negative" template of g, and the minimization of $E_g(\vec{k})$ is equivalent to maximizing the linear correlation between f and the template $g-g^\star$; the latter is equal to 1 over G, to -1 over $W \setminus G$, and to 0 elsewhere. Thus for binary templates to be matched against binary images minimizing an ℓ_1 error (and hence maximizing a morphological correlation) is identical to minimizing an ℓ_2 error (and hence maximizing a linear correlation).

Hit-or-Miss Transforms

Assume a binary image f and a binary template g. Let W be a finite window set containing the domain G of g and let $g^\star(\vec{n}) = 1 - g(\vec{n})$ be the dual template of g with domain $G^\star = W \setminus G$. The locations of exact occurrences of g inside f (i.e., pixels $\vec{n}$ at which f matches exactly g and $g^\star$ over G and $G^\star$, respectively, shifted at $\vec{n}$) are the 1-valued pixels of the binary image

$$[f \circledast (g,g^\star)](\vec{n}) = MIN[\min_{\vec{m}\in G} f(\vec{n}+\vec{m})\ ,\ \min_{\vec{m}\in G^\star}(1-f(\vec{n}+\vec{m}))] \qquad (20)$$

This is called a *hit-or-miss transform* and was introduced in [4] for general feature detection. It is directly related to the linear correlation in (19) if the latter is followed by some thresholding. The hit-or-miss transform is a very general Boolean matched filter. Some special cases were used in [5] for object detection.

Noise is unavoidable in images. An exact Boolean matched filter as in (20) has a poor performance in the presence of noise, or occlusion, or inexact knowledge of the features to detect. To compensate for uncertainties, it was suggested in [5] to use parallel combination of matched filters (20) each by a slightly different distorted version of the correct template. However, such an approach would not be able to accomodate all possible distortions and still maintain a reasonably small number of matching templates. Hence we propose here an alternative approach which allows both for *partial* and exact occurrence of the templates $g, g^\star$. Thus, we replace the binary value $[\min_{\vec{m}\in G} f(\vec{n}+\vec{m})] \in \{0,1\}$ with the real value $\alpha(\vec{n}) = [\sum_{\vec{m}\in G} f(\vec{n}+\vec{m})]/A(g)$, where $A(g)$ is the area (number of pixels) of the template g. Hence, α is the area portion of f (over G shifted at $\vec{n}$) that matches g. Likewise, we replace the second local min in (20) with $\beta(\vec{n}) = 1 - [\sum_{\vec{m}\in G^\star} f(\vec{n}+\vec{m})/A(g^\star)]$. This leads to the *modified hit-or-miss transform*

$$[f \circledast_p (g,g^\star)](\vec{n}) = \psi[\alpha(\vec{n}),\beta(\vec{n})]\,, \qquad (21)$$

where ψ is a real-valued function such that $\psi(\alpha,\beta) \in [0,1]$. A final decision would require comparing ψ with some threshold. If $\psi(,) = \min(,)$, (20) becomes a special case of (21); another choice for ψ could be a linear average. Both ratios α and β take values in the interval $[0,1]$ and represent a "confidence score" (e.g., a probability) that the specific template-object occurs in f at various locations. Hence (21) can handle noise and other uncertainties.

5 Object Detection

Next we approach the problem of image object detection in the presence of noise from the viewpoint of statistical hypothesis testing and rank order filtering. Some of the previous ideas in template matching apply here too.

Statistical Approach

Assume that the observed binary image $f(\vec{n})$ has been generated under one of the following two probabilistic hypotheses:

H_0: $\quad f(\vec{n}) = e(\vec{n}),\ \ \vec{n} \in W.$
H_1: $\quad f(\vec{n}) = |g(\vec{n}-\vec{k}) - e(\vec{n})|,\ \ \vec{n} \in W.$

Hypothesis H_1 (H_0) stands for "object present" ("object not present") at pixel location $\vec{k}$. The "object" $g(\vec{n})$ is a deterministic binary template. $e(\vec{n})$ is a stationary binary noise random field which is a 2-D sequence of independent identically distributed (i.i.d.) random variables taking value 1 with probability $p < 0.5$ and 0 with probability $1-p$. W is a finite set of pixels equal to the domain of g shifted to location $\vec{k}$ at which the decision is taken. The absolute-difference superposition between g and e under H_1 forces f to always have values 0,1. Intuitively, such a signal/noise superposition means that the noise e *toggles* the value of g from 1 to 0 and from 0 to 1 with probability p at each pixel. This noise model can be viewed either as the common binary symmetric channel noise in signal transmission or as a binary version of the salt-and-pepper noise. To decide whether the object g occurs at $\vec{k}$ we use a Bayes decision rule that minimizes the total probability of error and hence leads to the *likelihood ratio test*

$$\frac{Pr(f/H_1)}{Pr(f/H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{Pr(H_0)}{Pr(H_1)} = \eta \qquad (22)$$

where $Pr(f/H_i)$ are the likelihoods of H_i with respect to the observed image f, and $Pr(H_i)$ are the a priori probabilities. Due to the i.i.d. assumption about e,
$Pr(f/H_0) = \prod_{\vec{n}\in W} p^{f(\vec{n})}(1-p)^{1-f(\vec{n})}$.
Under H_1 and since f, g, e are binary, $e(\vec{n}) = |f(\vec{n}) - g(\vec{n}-\vec{k})|$ for $\vec{n} \in W$ and hence
$Pr(f/H_1) = \prod_{\vec{n}\in W} p^{|f(\vec{n})-g(\vec{n}-\vec{k})|}(1-p)^{1-|f(\vec{n})-g(\vec{n}-\vec{k})|}$.
Substituting these into (22) and taking logarithms of both sides yields

$$\sum_{\vec{n}\in W} f(\vec{n}) - \sum_{\vec{n}\in W} |f(\vec{n}) - g(\vec{n}-\vec{k})| \underset{H_0}{\overset{H_1}{\gtrless}} \eta_0$$

where $\eta_0 = \log\eta/\log(1-p/p)$. Expanding the absolute difference and cancelling common terms gives

$$M_{fg}(\vec{k}) = L_{fg}(\vec{k}) = \sum_{\vec{n}\in \mathbf{Z}^2} \min[f(\vec{n}), g(\vec{n}-\vec{k})] \underset{H_0}{\overset{H_1}{\gtrless}} \theta \qquad (23)$$

where $\theta = [\eta_0 + \sum_{\vec{n}} g(\vec{n})]/2$. Thus, the selected statistical criterion and noise model lead to compute the morphological (or equivalently linear) binary correlation between a noisy image and a known image object and compare it to a threshold for deciding whether the object is present.

Rank Order Filtering

From the previous analysis, to detect in a binary image f the presence of a binary object g at $\vec{k}$ we compare the binary correlation between f and g to a threshold θ. Theoretically this is equivalent to performing a r^{th} *rank order* filter on f defined by

$$[\text{RO}_r(f;G)](\vec{k}) = r^{\text{th}} \text{ largest of } f(\vec{n}),\ (\vec{n}-\vec{k}) \in G, \quad (24)$$

where G is the domain of g containing $|G|$ pixels and $1 \leq r \leq |G|$. The filter (24) can be applied to graytone images too, where for rank $r = 1$ ($r = |G|$) we obtain the local max (min) filter. For (23) and (24) to be equivalent we must select r proportional to θ. Thus the rank r reflects the area portion of (or a probabilistic confidence score for) G existing at pixel $\vec{k}$.

This link between binary correlation followed by thresholding and binary rank order operations aids us in approaching the following problem: Detect in a graytone image f corrupted by noise the presence of a *graytone image* object g of known domain G and unknown but fairly uniform gray level. Assuming impulse noise models and statistical criteria as before leads to comparing the morphological correlation of f and g to some threshold. Since, however, the exact amplitude of g is not accurately known, the previous approach is not practical to apply. Instead we apply the rank order operations. Namely, (12) implies that $||f-g||_1$ is the sum of $||f_a - g_a||_1$ for all gray levels a. Mininizing $||f_a - g_a||_1$ at all a leads to comparing the correlation of f_a and g_a to θ, which can be approached through the binary rank order filter $\text{RO}_r(f_a; G)$. Further,

$$[\text{RO}_r(f;G)](\vec{k}) = \sum_{a \geq 1} [\text{RO}_r(f_a;G)](\vec{k}). \quad (25)$$

Thus noise affects all levels f_a, g_a. At each level we apply statistical decision (23). A decision for 'object present' is equivalent to the binary image $\text{RO}_r(f_a;G)$ having value one at pixel $\vec{k}$. Applying the decision on all levels, due to (12) and (9), is equivalent to summing the outputs of the binary rank order filters from all levels; this, in turn, is equivalent to performing a gray rank order filter on f because of (25). The larger the value of $\text{RO}_r(f;G)$ at $\vec{k}$, relative to the levels of the other neighbor pixels in G, the greater the probability of the object's presence there.

Further, with rank order filters not only can we detect the location of such a g, but we can also 'redraw' (estimate) g. Specifically, the morphological opening [4,6,7] $f \circ G = (f \ominus G) \oplus G$ eliminates from f all parts inside which no shifted (in argument and/or in level) version of G can fit. It consists of an erosion $\ominus$ (detects the locations of G) and a dilation $\oplus$ (redraws G). However, the erosion is a local min, i.e., a rank order filter with $r = |G|$. This imposes a very strict requirement on the detection algorithm since it searches for exact shapes like G. Instead we propose a *rank opening* consisting of a r^{th} rank order filter followed by a dilation (a $r = 1$ rank order filter):

$$[\text{RO}_r(f;G)] \oplus G\ ,$$

where by varying r (controlled by θ in (23)) we can accomodate uncertainties, noise, or occlusion.

References

[1] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*, NY: Wiley, 1973.

[2] A. Rosenfeld and A. C. Kak, *Digital Picture Processing*, Acad. Press, 1982.

[3] G. J. Vanderbrug and A. Rosenfeld, "Two-Stage Template Matching," *IEEE Trans. Comp.*, 26, Apr. 1977.

[4] J. Serra, *Image Analysis and Mathematical Morphology*, Acad. Press, 1982.

[5] T. R. Crimmins and W. M. Brown, "Image Algebra and Automatic Shape Recognition," *IEEE Trans. AES-21*, Jan. 1985.

[6] P. Maragos and R. W. Schafer, "Morphological Filters - Part I: Their Set-Theoretic Analysis and Relations to Linear Shift-Invariant Filters. Part II: Their Relations to Median, Order-Statistic, and Stack Filters," *IEEE Trans. ASSP-35*, Aug. 1987.

[7] R.M. Haralick, S.R. Sternberg, and X. Zhuang, "Image Analysis Using Mathematical Morphology," *IEEE Trans. PAMI-9*, July 1987.

[8] M. Werman and S. Peleg, "Min-Max Operators in Texture Analysis," *IEEE Trans. PAMI-7*, Nov. 1985.

[9] E. J. Coyle, "Rank Order Operators and the Mean Absolute Error Criterion," *IEEE Trans. ASSP-36*, Jan. 1988.