

# Blur Hit-Miss Transform and its use in Document Image Pattern Detection

Dan S. Bloomberg and Luc Vincent

Xerox Palo Alto Research Center, Palo Alto, CA 94304, and  
Xerox Imaging Systems, Peabody, MA 01960

[*Proc. SPIE Vol. 2422, "Document Recognition II", pp. 278-292, San Jose CA, February 1995*]

## Abstract

The usefulness of the hit-miss transform (HMT) and related transforms for pattern matching in document image applications is examined. Although the HMT is sensitive to the types of noise found in scanned images, including both boundary and random noise, a simple extension, the Blur HMT, is relatively robust. The noise immunity of the Blur HMT derives from its ability to treat both types of noise together, and to remove them by appropriate dilations.

In analogy with the Hausdorff metric for the distance between two sets, metric generalizations for special cases of the Blur HMT are derived. Whereas Hausdorff uses both directions of the directed distances between two sets, a metric derived from a special case of the Blur HMT uses just one direction of the directed distances between foreground and background components of two sets. For both foreground and background, the template is always the first of the directed sets. A less restrictive metric generalization, where the disjoint foreground and background components of the template need not be set complements, is also derived. For images with a random component of noise, the Blur HMT is sensitive only to the size of the noise, whereas Hausdorff matching is sensitive to its location. It is also shown how these metric functions can be derived from the distance functions of the foreground and background of an image, using dilation by the appropriate templates.

The Blur HMT is implemented efficiently with boolean image operations. The FG and BG images are dilated with structuring elements that depend on image noise and pattern variability, and the results are then eroded with templates derived from patterns to be matched. Subsampling the patterns on a regular grid can improve speed and maintain match quality, and examples are given that indicate how to explore the parameter space. The Blur HMT can be used as a fast heuristic to avoid more expensive integer-based matching techniques. Truncated matches give the same result as full erosions and are much faster. Keywords: pattern matching, scanned image, hit-miss transform, Hausdorff distance, image morphology, OCR

## 1 Introduction

Pattern matching techniques are critical for all aspects of the analysis of document images. Documents are typically scanned into a binary image, and many of the operations subsequently performed, both for page segmentation and character identification, use the pattern matching techniques (e.g., *erosion* and its dual, *dilation*) of binary image morphology. There are several reasons: they are implemented by fast boolean operations; they can be used either for extracting or extending pixel aggregations, both for direct use in later image processing and for subsequent analysis; they are translationally invariant; they can be used to treat both foreground (FG) and background (BG) simultaneously; and they can be used without regard to connected component analysis. Further, there exist a variety of methods for controlling the noise immunity of these operations.

One of the most important uses of pattern matching is in the analysis of character shapes. For binary input, the result of image processing can be either binary or gray (integer value) images. Binary results are much faster to compute, but they contain less information. Even for binary output, the internal operations can be integer or boolean. For integer operations, such as convolution and thresholded convolution (rank order filters[8]), some level of noise immunity is achieved, but at the price of doing expensive arithmetic operations on each pixel.

We use the term *template* to refer to the pattern of FG and BG pixels that are to be matched in the image. The *hit-miss transform* (HMT) is a faster boolean operation that performs translationally-invariant matching between both the FG and BG of template and image sets. However, it is prone to error from noise because exact

matches are required between image and template in both the FG and BG. Because of the simplicity and power of the HMT, there have been many attempts to use it for pattern matching. The usual approach is to choose a subset of the template pixels, typically sparse. We cite a few examples.

Zhao and Daut[3] gained noise immunity, relative to a HMT, by using either boundary pixels of eroded FG and BG templates, or skeletons of these templates, as structuring elements for the HMT. Wilson[13] automated the design of the structuring elements through a training process that searched for the smallest subset of pixels that would attain the desired level of discrimination. Kraus and Dougherty[5] generated a sparse set of structuring elements by thresholding a single grayscale instance of each character. Appropriate choice of thresholds is the critical element: if chosen too conservatively, the subset is too sparse and lacks discriminatory ability; if chosen too tightly, instances with atypical variation are missed. Gillies[4] took a somewhat different approach, accumulating statistics from instances of each character, and thresholding the aggregates to generate non-sparse structuring elements. From these, multi-pixel features were extracted and used to train a classifier for character discrimination.

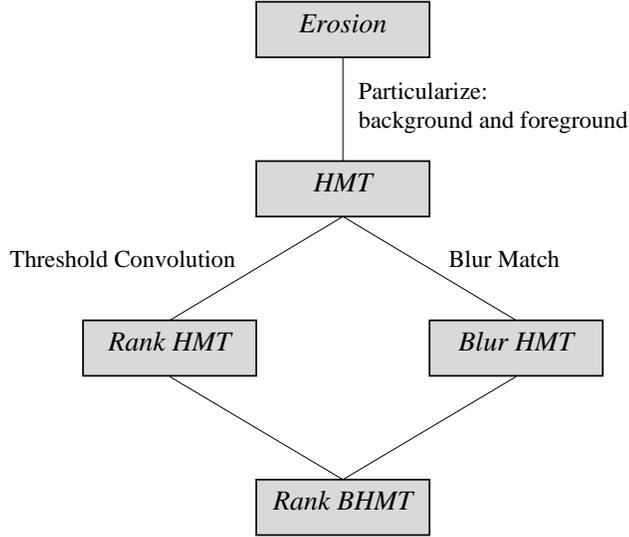
One characteristic that these methods have in common is an attempt to compensate for *image noise* by altering the *template*, and leaving the image alone. We argue here that although it is useful to choose a subset of template pixels, it is important to alter the image before performing the HMT. The *blur hit-miss transform* (BHMT) has been introduced to do precisely this[1]. Unlike the HMT, the BHMT performs the match between template and image, for both FG and BG, with a variable degree of tolerance to alignment of image and template pixels. The “blur” parameter specifies the maximum distance allowed between a template pixel and the nearest image pixel, in order to constitute a match for that template pixel. Stated this way, there is an interesting relation between the BHMT and the Hausdorff metric for the distance between two sets, but the differences are important for their uses in applications.

To understand the usefulness of the BHMT, it is necessary to consider the origin of noise in scanned document images. We postulate a simple model, where variability between instances in the image is caused by two different processes. One type is *boundary noise*, caused by the binarization process along the edge of an object. Depending on the sub-pixel alignment of scanned objects with scanner pixels, considerable edge variation occurs. This boundary noise is typically restricted to a width of two pixels, including both FG and BG boundary pixels. The second type is *random noise*, either generated in printing or due to scanner defects such as dirt on the platten. This is assumed to occur independently of the pixels in the scanned object, and is most often observed as isolated FG “pepper” pixels surrounded by BG. It should be noted that both types of noise occur in *boundary pixels*, defined to be pixels of either FG or BG that are adjacent to a pixel of the opposite type. Thus, operations that treat boundary pixels appropriately will influence both types of noise.

Even without random noise, boundary noise will defeat an HMT that uses boundary pixels in the template. Therefore, when computing matches, it is necessary to give little or no weight to the boundary pixels. On the other hand, the non-boundary pixels, because of their high correlation between template and image, are critical for matches. Differences occurring between non-boundary pixels in image and template, although relatively rare, will defeat both a matching technique like HMT, that requires an exact match of all pixels, and a metric such as Hausdorff, that is sensitive to such “outliers”.

The paper is organized as follows. In Sec. 2.1 the BHMT is defined, and the method in which it provides immunity to both types of noise is described qualitatively. In Sec. 2.2, the Hausdorff metric is introduced, and the connection between this measure on sets and operations using morphology is made. Also, an illustration is given to show why the Hausdorff metric is not appropriate for matching templates to noisy images. In Sec. 2.3, two metric functions are constructed, that are related to special cases of the BHMT. The same example is then used to show how the BHMT succeeds in matching templates to noisy images. Then in Sec. 2.4, the BHMT metric functions are again derived, this time from dilations by the template of the distance function for the image. In Sec. 2.5, several methods for implementing the BHMT are described, including subsampling the template. Then in Sec. 3, some results are given to illustrate the use of the BHMT in identifying characters. Finally, the major results are summarized.

We end this section with an illustration, in Fig. 1, of the family of rank and blur template matching operations. The HMT generalizes the erosion to operations that match in both FG and BG. The rank operations take thresholds on convolutions, whereas the blur operations remove boundary pixels appropriately before doing strict matching. The *rank* HMT, a relatively expensive integer operation, requires co-location of image and template pixels, but eases the constraint on the number of matches. The rank HMT and the BHMT can also be combined into the *rank* BHMT, in which a match is accepted if only a given number of template pixels are within a given



*Figure 1: Family of rank and blur template matching operations.*

distance of the nearest image pixel. In the sequel, we concentrate on the BHMT.

## 2 Blur HMT

### 2.1 Basic definitions

We are strictly interested in the discrete case of sets and functions defined on  $\mathbb{Z}^2$ , although extensions can be made to the continuous case or higher dimensions. The basic morphological operations are *erosion* and *dilation*. The erosion of a binary image  $X$  by a structuring element (SE)  $B$  is the set operation defined by

$$X \ominus B = \bigcap_{b \in B} X_{-b} = \{x \in \mathbb{Z}^2 \mid B_x \subseteq X\} \quad (1)$$

where  $X_{-b}$  is the translation of image  $X$  by  $-b$ . The second definition states that erosion generates a set with a non-empty result at every location where the translate of  $B$  fits entirely within  $X$ . The dilation of an image  $X$  is defined

$$X \oplus B = \bigcup_{b \in B} X_b = \bigcup_{x \in X} B_x = \{x + b \in \mathbb{Z}^2 \mid x \in X, b \in B\} \quad (2)$$

The first and second definitions state that dilation generates a set composed of the union of translations of  $X$  by elements in  $B$ , and *v.v.* Note that  $\ominus$  and  $\oplus$  are *not* the original definitions of Minkowski subtraction and addition, respectively, which require an inversion of the SE about its center[11].

The HMT is a morphological template matcher whose definition is based on the erosion operator[11]. The HMT of a binary image  $X$  by a disjoint pair  $(\tau_f, \tau_b)$  of SEs is defined as the set transformation

$$X \otimes (\tau_f, \tau_b) = (X \ominus \tau_f) \cap (X^c \ominus \tau_b) \quad (3)$$

where  $X^c$  is the set of BG pixels of  $X$ . The HMT generates a set with non-empty result at every location where both the FG SE  $\tau_f$  fits entirely within  $X$  and the BG SE  $\tau_b$  fits entirely within  $X^c$ , the complement of  $X$ . It is common to speak of the elements in  $\tau_f$  as *hits*, of elements in  $\tau_b$  as *misses*, and elements not in their union as *don't-cares*.

There are several methods for reducing the sensitivity to boundary noise. We can erode the template SEs by the blur SEs, or dilate the image by the blur SEs, all prior to the HMT. Eroding the template removes its boundary pixels from consideration, whereas dilating the image removes the image boundary pixels. The random

noise pixels are also affected: eroding the template opens up FG and BG holes, so they are not included in the match; dilating the image closes up holes (i.e., removes salt and pepper) in FG and BG. The results of these operations (followed by the HMT) differ, and the choice must be made based on the statistics of expected noise. For document images, the template is expected to be free of salt and pepper noise in situations where it can be generated by averaging a large set of instances. However, this is *not* true for the image. Salt and pepper noise in the image will prevent matches between FG and BG of the template, respectively. Generally, it is imperative to remove noise pixels from both the template and image before doing the HMT. For situations where random noise is more frequent in the image than in the template, we thus define the BHMT of a binary image  $X$  using the SE pair  $(\tau_f, \tau_b)$  for the template and the SE pair  $(\beta_f, \beta_b)$  for blur as[1]

$$X \otimes (\tau_f, \tau_b; \beta_f, \beta_b) = (X \oplus \beta_f) \ominus \tau_f \cap (X^c \oplus \beta_b) \ominus \tau_b \quad (4)$$

It can be noted that there is no requirement that the SEs used for blur be symmetric about their center. Translation of the center of a SE simply results in translation of the dilated image. The only requirement is for alignment between FG and BG: the FG and BG centers must be in the same (approximate) relation to the elements of their SEs. Because the BHMT allows different  $\beta_f$  and  $\beta_b$ , this may not be exactly true. For example, the center of a  $3 \times 3$  SE can be placed in the geometrical center of the elements, whereas this is not true of a  $2 \times 2$  SE. The difference can be kept to less than one pixel, which is insignificant.

## 2.2 Relation between Hausdorff metric and morphology

The Hausdorff metric is a distance between sets that allows one to define a topology on the set of all possible sets in the plane[9]. In image terminology, it is a distance between the FG of two images. The relation between the Hausdorff metric and a pair of blurred template matches has been noted previously[2], and we present the connection here.

Define the distance function[10] from a point  $p$  to the nearest point in a set  $X$  to be  $d(p, X)$ . If  $p \in X$ , then  $d(p, X) = 0$ . For two sets  $T$  and  $I$ , define the *directed* Hausdorff distance[6] from  $T \Rightarrow I$  as the maximum over the pixels in the set  $T$  of the distance from the pixel in  $T$  to its nearest pixel in  $I$ :

$$D(T, I) = \sup_{t \in T} d(t, I) \quad (5)$$

For applications to document images, consider  $T$  to be a FG template and consider  $I$  to be a *windowed* subset of the image  $X$  with support equal to that of the template. Then for each position in the plane represented by  $X$ , there exists a windowed subset  $I \subset X$  and a directed Hausdorff distance  $D(T, I)$  between  $T$  and the co-located  $I$ . Suppose this distance is  $\delta$ . If the set  $I$  is dilated by a disk of radius  $\delta$ , the distance between the dilated set and  $T$  will be zero. Consequently, an erosion of the dilated  $I$  by  $T$  will give a non-empty result.

The Hausdorff metric  $D_H$  is formed symmetrically between  $T$  and  $I$ , as the maximum of the two directed Hausdorff distances:

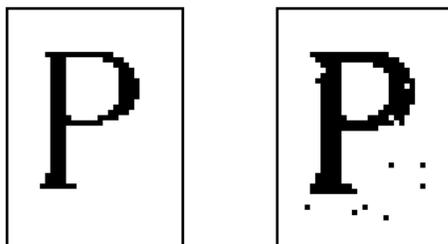
$$D_H(T, I) = \max\{D(T, I), D(I, T)\} \quad (6)$$

Its relation to the blurred match between template and windowed image subset is[2]

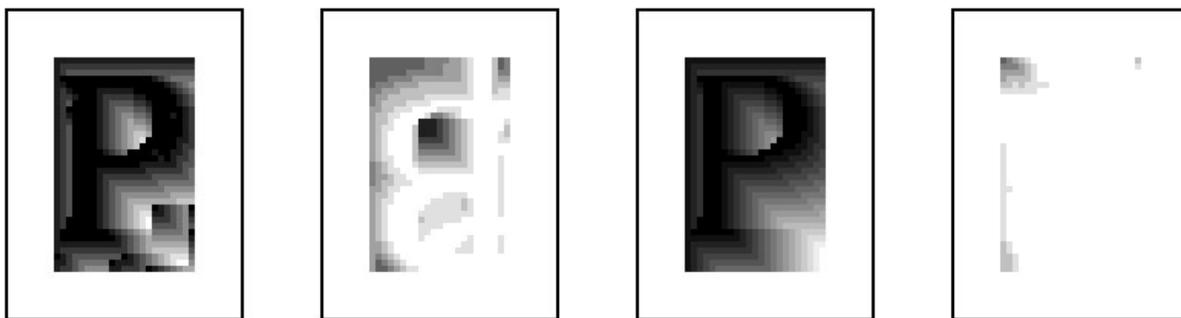
$$D_H(T, I) = \inf\{\rho \geq 0 \mid T \subseteq (I \oplus \rho B) \text{ and } I \subseteq (T \oplus \rho B)\} \quad (7)$$

where  $B$  is the unit disk SE. This is a symmetrical relation between FG sets. If  $I$  and  $T$  are very similar, small dilations act only to reduce the distance contributions from boundary pixels. Thus, a single non-boundary noise pixel in either  $I$  or  $T$  can render the Hausdorff distance quite large.

The effects of noise are illustrated by the two sets shown in Fig. 2. Call the sets on the left and right  $T$  and  $I$ , respectively. We have chosen the template to be less noisy and slightly eroded with respect to the image. The directed Hausdorff distances generated by these sets, with translates of the first set, are shown in Fig. 3. The first and third frames are distance functions that will be discussed later. The second frame is the directed distance  $D(T_{(x,y)}, I)$  from  $T \Rightarrow I$ , evaluated at each possible location of  $T$  with respect to  $I$ . Darker values represent shorter distances. The best match, with a distance of 0, is from the dark region near the center. However, because of the noise in  $I$ , the distance  $D(I_{(x,y)}, T)$  from  $I \Rightarrow T$ , shown in the fourth frame, has a very large value at that location. In fact, the smallest distances in  $D(I_{(x,y)}, T)$  are found near the boundaries, due to clipping.



*Figure 2: Two sets used to illustrate effect of noise in Hausdorff distance.*



*Figure 3: Hausdorff distances generated from sets in Fig. 2. The first and third frames are distance functions for each set. The second and fourth frames are the directed distances  $D(T_{(x,y)}, I)$  and  $D(I_{(x,y)}, T)$ , respectively.*

This clipping effect is another complication of using a windowed directed Hausdorff distance from large image to a small template. For this example, the Hausdorff distance  $D_H(T, I)$  is identical to the directed distance  $D(I_{(x,y)}, T)$  for every translate of  $I$ , because the match between the two sets is entirely obscured by the noise in  $I$ .

### 2.3 Blur HMT metric

The BHMT produces a binary image representing locations of a match with a given amount of FG and BG blurring. In this section, we construct two BHMT-related distance metrics, in analogy with the Hausdorff metric. These are a generalized distance between image and template that, when thresholded, produce a BHMT for the value of FG and BG blur equal to the threshold.

When the template SE  $T$  is located on some windowed subset  $I$  of  $X$ , its center falls on coordinates  $(x, y)$ . Label each subset  $I$  of  $X$  by this location  $(x, y)$ . Then form a FG/BG metric  $D_{FB}$ , in analogy to  $D_H$ , that measures the directed distance between the FG and BG parts of  $T$  and  $I$ . Choose the direction  $T \Rightarrow I$  because (1)  $T$  should have less non-boundary noise than  $I$  and (2)  $T$  is typically much smaller than  $X$ . Now dilate the full image  $X$  and use the  $(x, y)$  translate of  $T$ ,  $T_{(x,y)}$ , to compare  $T$  with each subset  $I$  of  $X$  in computing the metric:

$$D_{FB}(T_{(x,y)}, X) = \inf\{\rho \geq 0 \mid T_{(x,y)} \subseteq (X \oplus \rho B) \text{ and } T_{(x,y)}^c \subseteq (X^c \oplus \rho B)\} \quad (8)$$

$$= \max\{D(T_{(x,y)}, X), D(T_{(x,y)}^c, X^c)\} \quad (9)$$

To understand the relation between  $D_{FB}$  and the BHMT, consider the BHMT in its most simple form, with two disk SEs of equal radius for the blur and two SEs for the templates that are set complements. Setting  $\tau_f = T$  and  $\tau_b = T^c$ , the BHMT is found by thresholding  $D_{FB}$ :

$$X \otimes (\tau_f, \tau_f^c; rB, rB) = \{(x, y) \in \mathbb{Z}^2 \mid D_{FB}(T_{(x,y)}, X) \leq r\} \quad (10)$$

The restriction in  $D_{FB}$  that the two SEs for the templates are set complements can easily be relaxed to the disjoint constraint for SEs in the HMT, by requiring only that  $\tau_f \subseteq T$  and  $\tau_b \subseteq T^c$ . Then the metric  $D_{FB}$  is generalized to

$$D_{BHMT^*}(\tau_{f(x,y)}, \tau_{b(x,y)}, X) = \inf\{\rho \geq 0 \mid \tau_{f(x,y)} \subseteq (X \oplus \rho B) \text{ and } \tau_{b(x,y)} \subseteq (X^c \oplus \rho B)\} \quad (11)$$

$$= \max\{D(\tau_{f(x,y)}, X), D(\tau_{b(x,y)}, X^c)\} \quad (12)$$

with the thresholding relation

$$X \otimes (\tau_f, \tau_b; rB, rB) = \{(x, y) \in \mathbb{Z}^2 \mid D_{BHMT^*}(\tau_{f(x,y)}, \tau_{b(x,y)}, X) \leq r\} \quad (13)$$

We use the notation ‘‘BHMT\*’’ to indicate the special case where the same dilation operator is used for both FG and BG. For the general case there are two BHMT distance metrics, one for FG and one for BG, and the BHMT is derived from from them by thresholding each separately and AND-ing the results.

For reasons of both efficiency and effectiveness we are usually interested in BHMT where  $\tau_b \neq \tau_f^c$  and  $\beta_f \neq \beta_b$ . Examples will be given in Sec. 3. In Sec. 2.4, we arrive at a distance metric generalization for the BHMT by a different route.

Whereas  $D_H$  has bi-directional symmetry between two FG sets, and ignores the BG,  $D_{FB}$  has FG/BG symmetry but imposes a directionality on the relation between the two sets  $T$  and  $X$ . Unlike the Hausdorff metric,  $D_{FB}$  and  $D_{BHMT^*}$  are relatively immune to salt and pepper noise pixels in  $X$ . However, they are sensitive to noise in  $T$ . If there is salt and pepper noise in  $T$  as well as in  $X$ , then some dilation of the FG and BG of  $T$  must also be considered.

Referring to Fig. 4, illustrating the BHMT for the same example as previously shown for Hausdorff, the directed distance  $D(T, I)$  in the first frame is identical to the directed Hausdorff  $D(T, I)$  in Fig. 3. (Although we now omit the  $(x, y)$  label on  $T$ , remember that these functions are defined over the set  $(x, y) \in \mathbb{Z}^2$  of translates of  $T$ .) However, the second frame gives the directed distance for the BG,  $D(T^c, I^c)$ . The effect of noise on this

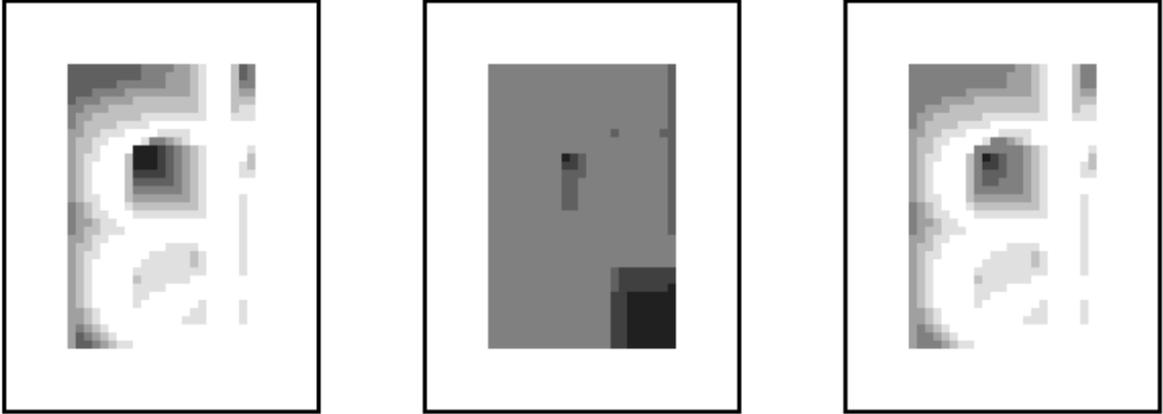


Figure 4: BHMT distances generated from sets in Fig. 2. The first and second frames are the directed distances  $D(T, I)$  and  $D(T^c, I^c)$ , respectively. The third frame is  $D_{FB}$ , the maximum of the two directed distances.

distance is small: it is determined by the *size* of the noise in  $I$ , rather than the *distance* from the noise to  $T$ . The contribution from the BG does affect the overall match to a small extent, as shown by  $D_{FB}$  in the third frame. This is the special case of the BHMT metric  $D_{BHMT^*}$  where the FG and BG templates are set complements. Because the BHMT distances are always directed  $T \Rightarrow X$  from the template to the large image, boundary effects occur only on the boundary of the image  $X$ .

Geometrically, the BHMT as we have defined it has the following interpretation. The blur dilations in the image  $X$  create an *umbra* of pixels near the boundaries, with polarity required to match the FG and BG templates in the HMT. Dilation of the FG of  $X$  removes salt noise and expands the FG to improve the FG template match near the boundaries. Likewise, dilation of the BG of  $X$  removes pepper noise and expands the BG, this time to improve match of the BG template. Both FG and BG matches must be perfect to contribute to the result set of the BHMT.

#### 2.4 Blur HMT metric derived from distance function

The directed Hausdorff distance  $D(T_{(x,y)}, X)$  for all translations  $(x, y)$  of the the template  $T$  can be found by first computing the distance function in  $X$  and then dilating by  $T$ . The distance function in  $X$  is 0 for points within the set, and otherwise is the smallest number of steps required to reach a point in  $X$ . We use an *asymmetric* 8-connected distance function that can be generated from successive dilations of the image by a  $2 \times 2$  brick SE. More efficiently, this distance function can be generated by a single raster scan using the maximum value of three neighbors to the North, West, and NorthWest to determine the value of each pixel in the distance function[12]. (The reason we do not use a symmetric distance function, which can be generated from successive *symmetric* dilations by a  $3 \times 3$  SE, is that the granularity to which distances between sets can be measured is twice as coarse with the symmetric distance.) The first and third frames of Fig. 3 show this distance function for the two sets in Fig. 2.

Dilation of this distance function by  $T$  gives the maximum value of this function for any pixels under  $T$ ; i.e., the maximum distance from the FG of  $X$  of any pixels that are within the translate of  $T$ . This distance, of course, is exactly the minimum dilation of  $X$  necessary to get a blur match from  $T$ . It is the distance metric for the FG part of the BHMT.

The distance  $D_{BHMT^*}$  requires taking the maximum of two such directed distances, one computed for the FG and one for the BG. Because the distance function is asymmetric, the *location* of the result is translated to the SouthEast, relative to  $X$ , by an amount equal to the distance function itself. Thresholding  $D_{BHMT^*}$  with some

value  $r$  generates the identical set as using blur dilation with an  $r + 1 \times r + 1$  SE on BG and FG before the HMT.

This set of relations is illustrated in Fig. 5, which shows the sequence of operations that generate BHMT distance metrics and BHMT images. We start with the image  $X$  and FG and BG templates. Grid spacings of 2 in  $x$  and  $y$  directions are used for generating the FG and BG templates. In the FG, the distance function is found for  $X$ , and dilated with the FG template, giving the FG directed distance metric. The dual process in the BG yields the BG directed distance metric. The maximum of these gives the BHMT\* metric, and for this example it is possible to find a threshold that yields a single match in the BHMT\* set. The same result can be derived using the threshold individually on the FG and BG metrics, and AND-ing the result. With the threshold chosen, it should be noted that the thresholded FG metric yields matches in two locations, one of which is between the template and the “g” in the image. This match was not seen in the BG, which removed it from the BHMT. Details of the BHMT\* metric are shown in Fig. 6.

In the general case, one would choose different threshold values (blur SEs) in FG and BG, in order to make the matching process more robust. This is the difference between the BHMT and the BHMT\*. When an asymmetric distance function is used, which is analogous to the use of different asymmetric SEs for FG and BG blur, the thresholded binary images, which have different translations of the match with respect to the image, must be re-aligned before being AND-ed.

## 2.5 Efficient implementations

Our interest is in finding relatively efficient implementations of the BHMT that are effective at locating matches without a large number of false positives. For character recognition, for example, the purpose is not to use the best possible pattern matchers, such as those used to estimate probabilities for templates in a maximum likelihood calculation[7]. Instead, we might want information that is good enough to be used as a heuristic for narrowing the search space for more computationally-intensive methods that do a better job of identifying characters.

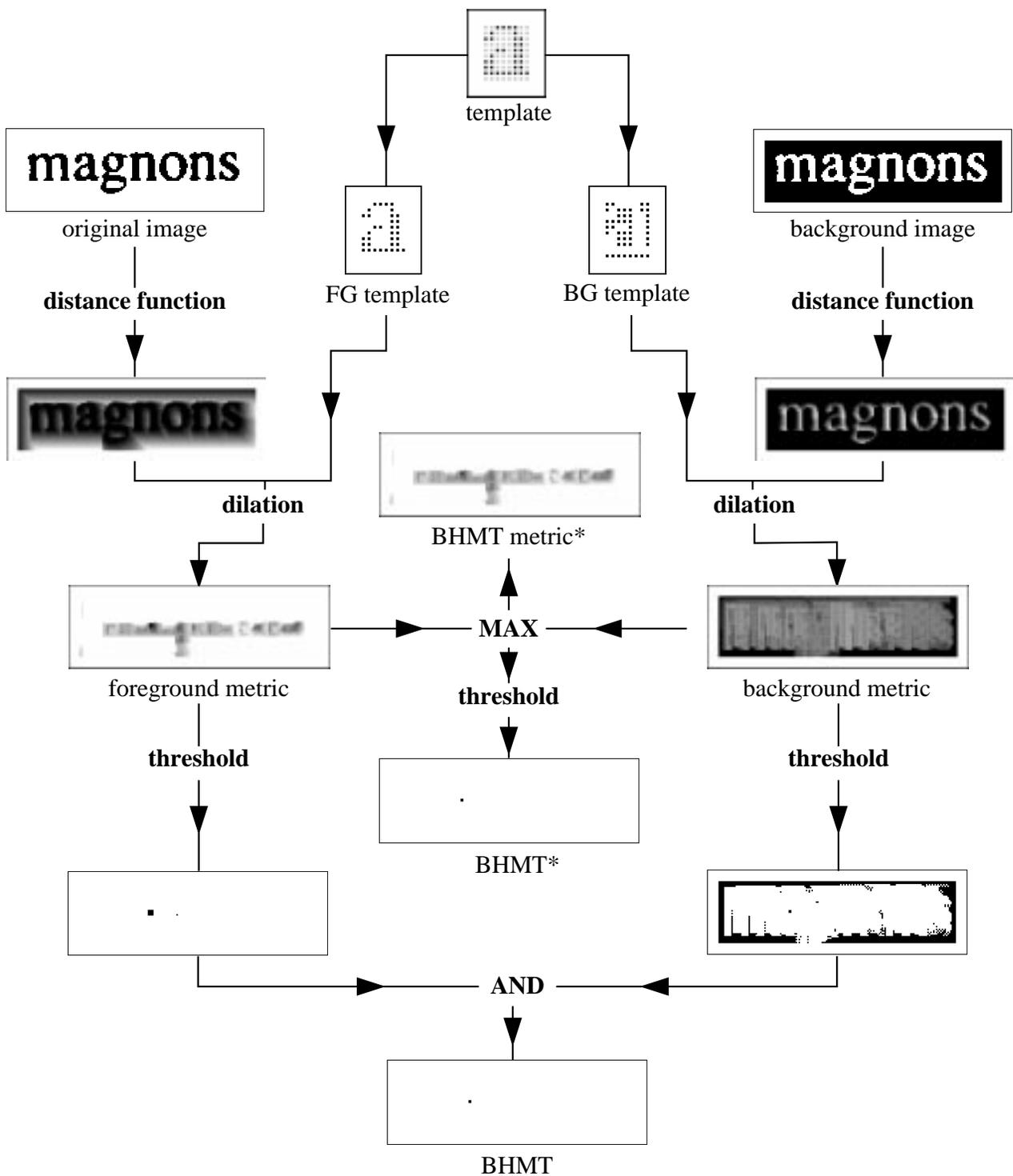
The rank operations, such as the rank BHMT, are less efficient than the BHMT because they require two (integer) convolutions by SEs, followed by thresholding. The BHMT uses only boolean operations. To improve the efficiency, we can also

- *Scale down both image and template.* Because we match all template pixels at each image position, the total number of pixels to be matched varies as the fourth power of the scaling parameter. The actual reduction in computation will be between the second power and the fourth power, depending on the implementation.
- *Subsample the template.* Suppose each template is subsampled by imposing a regular grid, with subsampling factors  $n_x$  and  $n_y$ . This has two effects. First, it decreases the computation required. The reduction varies from approximately  $n_y$  to the product  $n_x \times n_y$ , depending on the implementation. The second effect is that the subsampling tends to reduce the overall template dimensions, effectively augmenting the blur in the image.

The template was subsampled in two ways: choosing a random subset of template pixels and using a rectangular grid. For the same fraction of template pixels chosen, matches were much more accurate when a rectangular grid was chosen. Results using this method of subsampling are given in Sec. 3, where we typically subsampled the templates choosing  $n_x$  and  $n_y$  in the range between 1 and 5. Fig. 7 shows an example of subsampled FG templates, where  $n_x$  increases to the right and  $n_y$  increases downward. Another set is generated for the BG templates.

It is necessary to implement both the BHMT and the labelling process for the matches. These can be done sequentially or together. The first step is to perform the blur dilations on both the FG and BG of the image. This can then be used for a multiplicity of templates. The HMT can be implemented in several ways. For fully parallel methods, each erosion can be formed separately by the usual set of translations and ANDs, where the unit of operation can be anything from the pixel to the entire image. After taking the intersection of the FG and BG results, it is necessary to scan the image for connected components. For the sparse BHMT images, this is relatively fast compared to the BHMT itself.

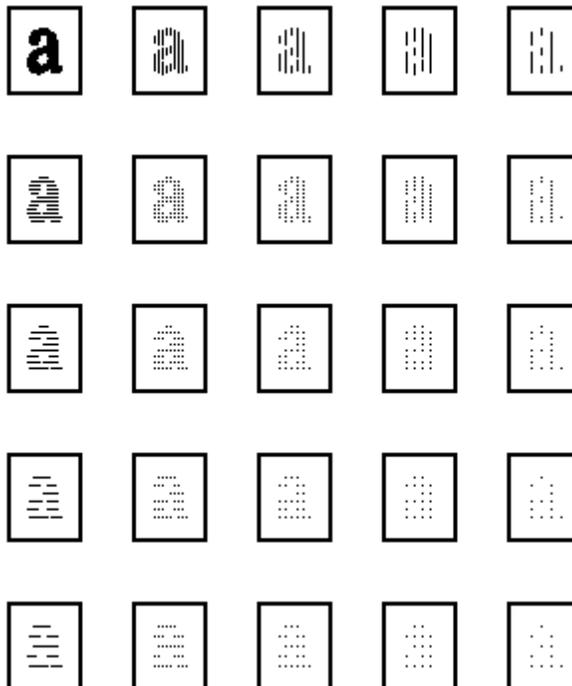
There is another implementation of the HMT that is faster, somewhat serialized, and largely circumvents the labelling process itself. The idea is to truncate the matching process in each location at the first instance of failure. Each template can be composed of an array of words, with each word representing the pixels in a template row. Suppose both FG and BG templates are to be tested at some location  $(x,y)$ . Choose the FG template and align its first row with the image to test for a match. (The test requires only three boolean operations: AND between template and image; XOR between this result and the template; test for 0.) If a line match is found,



*Figure 5: Sequence of operations that generate BHMT distance metrics and BHMT images.*



*Figure 6: Detail of BHMT\* metric for example in Fig. 5*



*Figure 7: Gridded FG templates, for  $(n_x, n_y)$  varying from  $(1,1)$  in the upper-left template to  $(5,5)$  in the lower-right template. Grid spacing  $n_x$  increases to right;  $n_y$  increases downward.*

proceed to the next template row. Whenever a line match is not found, quit the process at (x,y) and move to the next image location. If a full FG match is found, repeat with the BG template. If both matches succeed, record the location (the labelling process). Matches to a single image feature typically occur contiguously within some region that is smaller than the blur size. The serial aspect is required to avoid recording multiple positions for an image feature that has already been matched. When a match occurs, move several pixels away before looking for the next match. For the same reason, when scanning successive image lines, it is useful to avoid regions where a match was found proximally on lines above. Truncation of the matching process reduces the computation time by a factor proportional to the number of lines in the FG and BG templates that have ON pixels.

A modification of this approach also gives an efficient implementation of the rank BHMT. Rather than testing for zero, count the ON pixels in the template line that are OFF in the image. Accumulate this sum over successive template lines until either the rank threshold for that template is exceeded or the full template is matched. Do this for both FG and BG, which generally have different thresholds. As in the BHMT, avoid searching for matches in the vicinity of any location where a full match has already been found.

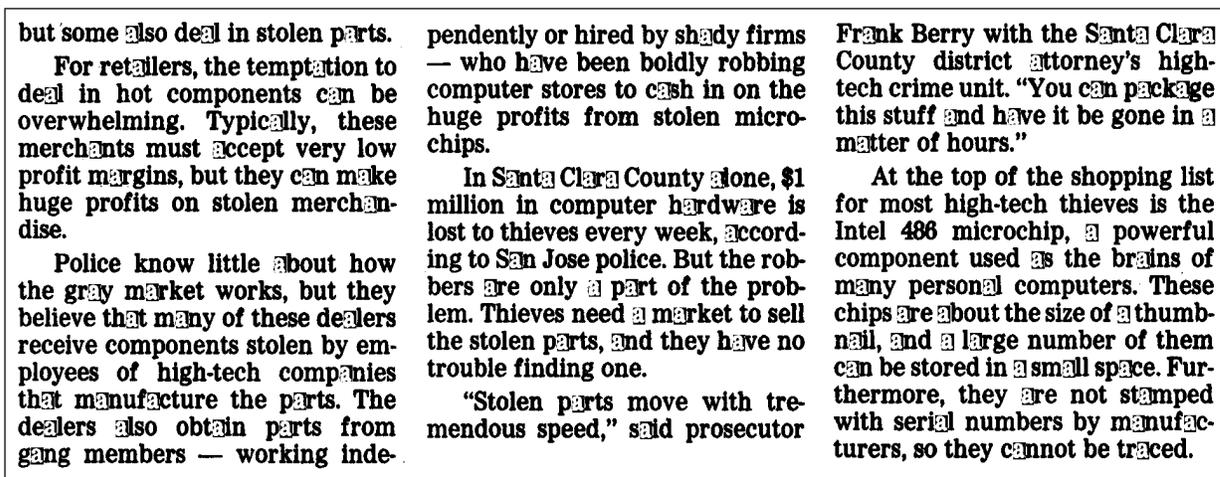
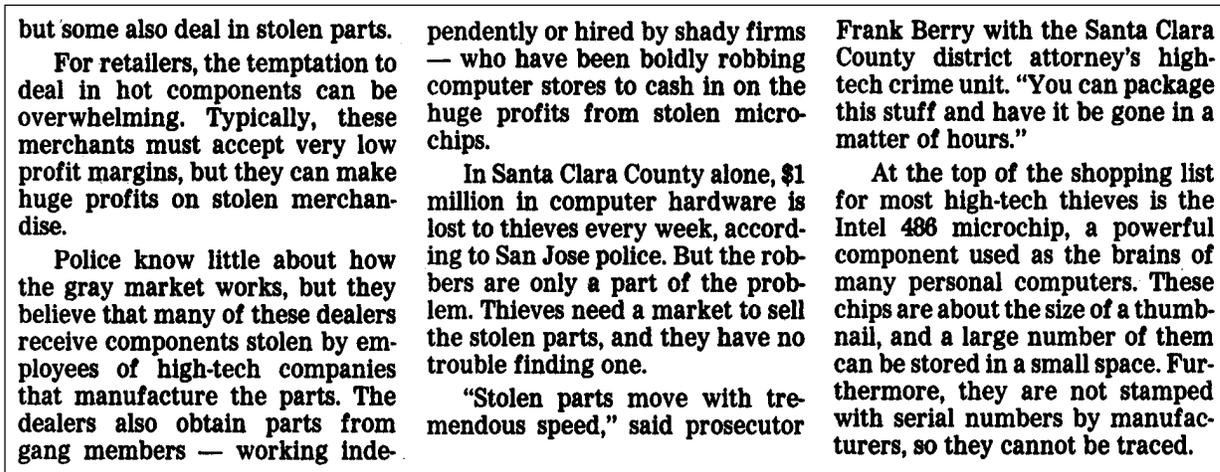


Figure 8: Top: Example image. Bottom: characters matched by the BHMT for the templates in Fig. 7 (and the BG templates as well) have been highlighted.

The efficiency of the truncated approach to the HMT can be estimated within a factor of two or so, depending on implementation details and the hardware. For most locations of the template(s), the match will fail on the

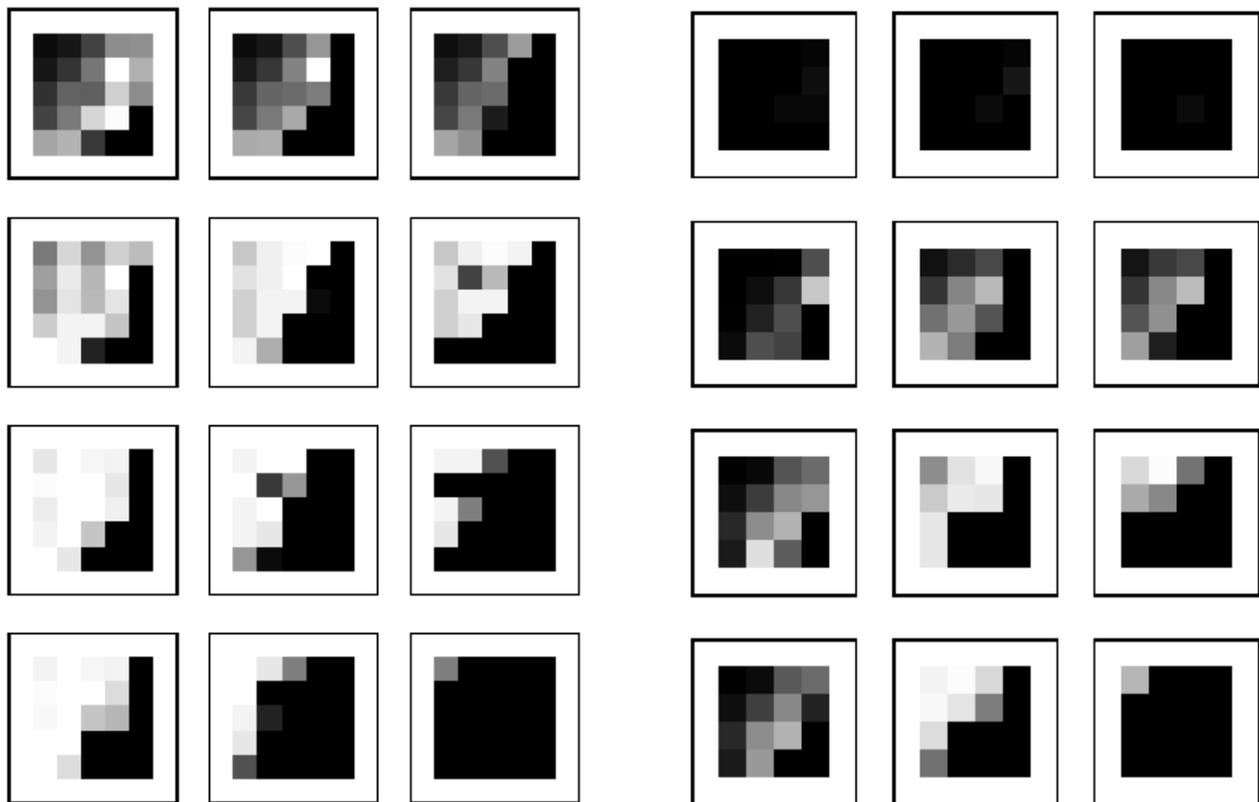
first line, requiring about 10 machine instructions (MIs). Suppose on average that 20 MIs are required for each location. A 40 MIPS machine can then match 2 million positions/second. For a document image where the vertical location of text baselines is known within  $\pm 2$  pixels, and where the textline width is 2000 pixels, matches are required at 10,000 positions for each textline. For a full page with 50 textlines, the matching time is about 250 ms.

### 3 Illustrative results for BHMT

In this section, the use of the BHMT for matching image characters is briefly explored. A number of parameters can be varied independently: the FG and BG blur of the image, the  $x$  and  $y$  grid subsampling of the template, and scale reduction of both image and template. We also need to define a criterion for scoring the matches.

As an example of the effect of blur and template gridding, consider the image at the top of Fig. 8 and the set of gridded templates in Fig. 7. There are 88 instances of the character “a” in the image, that are highlighted in the bottom image, having been identified by a BHMT using the blur parameters  $(\beta_f, \beta_b) = (3, 2)$  and grid spacings  $(n_x, n_y) = (4, 2)$ . With this combination, no false-positives were identified.

More generally, we define the *score* for a BHMT as the number of “a”s correctly found, decreased by the number of false-positives weighted by a penalty factor (we use 4 in Fig. 9). For visualization, we apply a floor of 0 to the scores, and scale the result from 0 to 255. Expressed as an image, white (255) is a perfect score, and black (0) is the lowest score recorded. For example, a BHMT that finds all 88 correct instances and has 22 or more false-positives receives a score of 0. The penalty chosen for false-positives should depend on the application.



*Figure 9: BHMT scores for 12 different combinations of FG and BG blur, at full resolution on left and reduced scale (0.8) on the right. At full scale on the left, each of the 12 frames shows scores from 25 different griddings of FG and BG templates. At reduced scale on the right, each frame shows 16 different griddings.*

The score results for this image are shown in Fig. 9, for different values of blur and template gridding. The left

half has BHMT scores for 12 different combinations of FG and BG blur, all found at full resolution; the right half has BHMT scores at scale reduced to 0.8. Each of the 12 frames (3 columns, 4 rows) on the left shows 25 different griddings of FG and BG templates; on the right, each frame shows 16 different griddings. At full resolution, FG blur increases from 2 to 4 to the right and BG blur increases from 2 to 5 downward. At 0.8 scale, smaller blur is used: FG blur increases from 1 to 3 to the right and BG blur increases from 1 to 4 downward. Within each frame at full resolution, the 25 squares correspond to the score from each of the gridded templates in Fig. 7, with grid spacings  $n_x$  increasing to the right and  $n_y$  increasing downward. At 0.8 scale, the full template is reduced to  $14 \times 20$  pixels, and smaller grid subsampling is used, with grid spacings  $n_x$  increasing from 1 to 4 to the right and  $n_y$  increasing from 1 to 4 downward. For both full scale and 0.8 scale, some blur is required. Cases with FG and BG blur less than 2 are not shown at full scale; at 0.8 scale, results are poor without some BG blur.

The BHMT is typically more efficient with coarser grid spacing, particularly with coarser y-grids; however, scores tend to get worse with coarse gridding, and cases with large blur generally do the worst with coarse gridding. Nevertheless, looking at any particular grid spacing, the score is typically maximized for an *intermediate* value of blur. The asymmetry between FG and BG blur is striking. It occurs in part because the image has no large solid black areas that could give false-positive matches to every FG template. When such regions exist, excessive BG dilation contributes significantly to these errors.

At full resolution, the most robust choice of grid and blur with BG blur not larger than 3 is around  $(n_x, n_y) = (3, 2)$  and  $(\beta_f, \beta_b) = (3, 3)$ . At 0.8 scale, the FG blur is for most griddings optimized at 2. The most robust choice of grid and blur is around  $(n_x, n_y) = (2, 2)$  and  $(\beta_f, \beta_b) = (2, 2)$ . When matching multiple characters in a font and size, an optimum pair of FG and BG blur SEs can be chosen, and the template grid spacings can be individually optimized for each character, including use of different grid spacings for FG and BG templates.

The immunity of the BHMT to random noise is striking. For example, when random noise at the high (about one percent) level shown in Fig. 2 is added to the image, the scores from the BHMT shown in Fig. 9 are not significantly changed.

## 4 Summary

Translationally invariant methods for pattern matching in scanned document images have no dependencies on pixel connectedness in either the image or template. We have focused on the most efficient techniques, that require only boolean operations. The basic operation, the HMT (which should be called the Hit-*And*-Miss transform!), is maximally sensitive to noise in *both* FG and BG.

Fortunately, there are ways to increase the noise immunity of the HMT. For extensions that use only boolean operations (as opposed to linear convolution and rank order filters), and considering the nature of binary pixel noise in both images and templates, we have argued that the BHMT is the best choice.

Considerable attention was devoted to distance metrics that can be derived from (special cases of) the BHMT. We began with the well-known relation between the Hausdorff metric and blurred template matches, and derived similar metrics for the BHMT. *This distance provides a measure of the goodness of fit of the template at every location in the image.* Comparing the Hausdorff and BHMT mechanisms of action on noisy document images, we showed why (1) the bi-directional symmetry of Hausdorff is so problematic and (2) the uni-directional but FG/BG-symmetric BHMT provides immunity to both boundary and random pixel noise. An intuitive presentation of these differences is a primary goal of this paper.

We also showed how the BHMT\* metric can be derived in the grayscale regime starting with distance functions for the FG and BG image. For most sensitivity, we choose an 8-connected asymmetric function that is generated in one raster scan and increments the distance to the South and East. These distance functions are then dilated with the FG and BG templates, and combined using the pixelwise Max operator.

The BHMT is useful for any document image pattern matching task. We have shown the results of a few experiments on pattern matching for characters, to illustrate the effects of FG and BG blur, and of regular subsamplings of the templates. Regular gridding of the templates gives much better results than using random subsets, for the same number of elements chosen. We also discussed methods for truncating the matches, that give much greater efficiency than using full erosions at every location. These truncation methods are also applicable to rank operations, particularly if only a small number of matching failures is permitted.

## 5 References

1. D. S. Bloomberg and P. Maragos, "Generalized hit-miss operations", *SPIE Conf. Image Algebra and Morphological Image Processing, Vol. 1350*, San Diego, CA, July 1990, pp. 116-128.
2. H. J. A. M. Heijmans, *Morphological Image Operators*, Acad. Press, 1994, see p. 296.
3. D. Zhao and D. G. Daut, "Shape recognition using morphological transformation", *J. Visual Comm. and Image Rep.* **2**, pp. 230-243, Sept 1991.
4. A. M. Gillies, "Automatic generation of morphological template features", *SPIE Conf. Image Algebra and Morphological Image Processing, Vol. 1350*, San Diego, CA, July 1990, pp. 252-261.
5. E. Kraus and E. Dougherty, "Segmentation free morphological character recognition", *SPIE Conf. Document Recognition, Vol. 2181*, San Jose, CA, Feb 1994, pp. 14-23.
6. D. P. Huttenlocher, G. A. Klanderman and W. J. Rucklidge, "Comparing images using the Hausdorff distance", *IEEE Trans. PAMI* **15**, pp. 850-863, Sept 1993.
7. G. Kopec and P. Chou, "Document image decoding using Markov source models", *IEEE Trans. PAMI* **16**, pp. 602-618, June 1994.
8. P. Maragos and R. W. Schafer, "Morphological filters - Part II: their relations to median, order-statistic, and stack filters," *IEEE Trans. Acoust. Speech Signal Process.*, ASSP-35, pp. 1170-1184, Aug. 1987.
9. G. Matheron, *Random Sets and Integral Geometry*, J. Wiley and Sons, NY, 1975.
10. A. Rosenfeld and J. L. Pfalz, "Distance functions on digital pictures", *Pattern Recognition* **1**, pp. 33-61, 1968.
11. J. Serra, *Image Analysis and Mathematical Morphology*, Acad. Press, 1982.
12. L. Vincent, "New trends in morphological algorithms", *SPIE Conf. Nonlinear Image Processing II, Vol. 1451*, San Jose, CA, Feb 1991, pp. 158-169.
13. S. S. Wilson, "Training structuring elements in morphological networks", *Mathematical Morphology in Image Processing*, ed. E. R. Dougherty, Marcel Dekker, NY 1992.