GENERALIZED DISTANCE BASED MATCHING OF NONBINARY IMAGES

Valery V. Starovoitov

Institute of Engineering Cybernetics
Surganova Str. 6, 220012 Minsk, Belarus
valerys@newman.basnet.minsk.by

Cenk Köse and Bülent Sankur

Boğaziçi University, EEE Department
80815 Bebek, Istanbul, Turkey
sankur@boun.edu.tr

ABSTRACT

We develope a technique to perform efficient and accurate matching to detect the occurence(s) of a template in a scene. The template may describe an object or a textured surface. Representative binary features such as edge points for objects and a set of intensity extrema for textured surfaces are extracted from both test and template images. The search is based on a three-stage "coarse-fine-final" matching. First, we match with the binarized template skipping positions in horizontal and vertical directions and find a number of highest ranking matching positions. Second, we search in neighborhoods centered at the stated positions and find improved matching positions. Finally, some error measure between the original, i.e. gray level or color, test and template images is computed at the fine positions as the final matching. It is demonstrated that the proposed scheme achieves significant reduction in computational complexity with virtually no loss of detection performance.

1. INTRODUCTION

Template matching is one of the basic operations in image processing and computer vision. Suppose we are given a test image **A** of size $N_1 \times N_1$ and a template image **B** of size $N_2 \times N_2$, where $N_1 \gg N_2$. For simplicity of notation we assume that images are square. The images may be in gray level or in color. In the classical approach the template is translated over the test image and for each of $(N_1 - N_2)^2$ positions the degree of matching is determined by some image distance measure, such as cross-correlation or the mean square difference [12].

Such exhaustive searches for the matching positions of a template are computationally very expensive. Many researchers have studied various ways to decrease this complexity. To this effect, parallel and quadtree structures, chamfer matching [7], edge based matching [2] were suggested.

Two-stage coarse-to-fine template matching scheme based on a reduced-resolution template was introduced in [10]. Other such approaches extract binary feature maps representative of template and test images and then match these, so-called binary feature fields. The binary features may be (a) edge points, [7, 2], (b) corners, (c) line or curve primitives [8]. Among these, edge points, have proven to be the most useful set of features for general template matching [2]. However, direct matching of edge maps by calculating pairwise point distances is still a very time consuming procedure. This brute force approach may take over 4 hours on Silicon Graphics R4000 [6].

To speed up these calculations the Distance Transform (DT) was successfully used [7, 2]. The DT is applied to the binarized template image, then the resulting distance map is superimposed with the binarized test image at all possible locations. Distances from the nearest edge points of the test image are easily computed and combined to give various indices of mismatch.

More robust and more accurate distance measures combined with efficient search methods have reduced the complexity of the matching task: D. Huttenlocher with co-workers developed ranked Hausdorff distances (Eq. 5) based on order statistics [2]. Twenty-four such functions combining the local distances in different ways were studied by M.-P. Dubuisson and A. Jain [3]. Further modifications of these ideas based on robust statistics were presented in [5].

The goal of our study is two-fold: First we try to develop a new reduced-cost coarse-to-fine matching scheme which is applicable to a broad range of images In this scheme the coarse stage runs on the binarized feature maps of both the template and the test images, while the fine stage validates the products of the coarse stage by recoursing to the original images. Secondly, we assess quantitatively the effectiveness of various distance measures (Eqs. 4 to 10).

The study of the first author was supported by TÜBİTAK-BAYG during his stay at Boğaziçi University.

The paper is organized as follows: Section 2 presents various matching measures. Section 3 introduces our matching scheme. In section 4 experimental results of two test cases along with a brief discussion are provided.

2. MEASURES OF MATCHING

2.1. Intensity based measures

A digital image **A** is a discrete function defined in a lattice domain of size $N \times N$ and taking values in the set of gray levels $\{0, 1, ..., G-1\}$. In this section, without loss of generality, we consider a gray scale image **A** as a set of pixels $\mathbf{A} = \{A_{ij}\}$, where every pixel is defined by its spatial coordinates (i, j) and gray value a_{ij}, i.e. a pixel is a point $A_{ij} = (i, j, a_{ij})$ in the *3-d* space.

A general criterion to deduce how much the two data sets look alike is the normalized correlation, formulated as:

$$CO(\mathbf{A}, \mathbf{B}) = \frac{\sum_{ij=1}^{N} a_{ij} b_{ij}}{\sqrt{\sum_{ij=1}^{N} a_{ij}^2 \sum_{ij=1}^{N} b_{ij}^2}} \qquad (1)$$

A second common error measure is the mean-square difference [12]:

$$MSD(\mathbf{A}, \mathbf{B}) = \frac{1}{N^2} \sum_{i,j=1}^{N} (a_{ij} - b_{ij})^2 \qquad (2)$$

which is a metric.

It is often desirable to have similarity measures that possess the metric property because it provides consistency of multiple measurements with qualitative human observations. However, square errors are very sensitive to phase (translation) therefore exhibit frequent oscillations over a small range of translations and this behavior hinders accurate localization of the global minimum.

2.2. Distance based measures

Let us consider two point sets $A = \{a_1, ..., a_{N_A}\}$ and $B = \{b_1, ..., b_{N_B}\}$. They may be the sets of black pixels from the binarized images A_b and B_b. The standard bidirectional Hausdorff distance between A and B is defined as

$$H(A, B) = max\{h(A \to B), h(B \to A)\}. \qquad (3)$$

The function h is called *the directed Hausdorff distance* from one set to another. In the case of binary images we calculate distances from the black pixels a to the set of the black pixels B. These distances may be calculated in different manners, for example,

$$h_1(A \to B) = max_{a \in A}\{d(a, B)\} = d_{N_A}^{sort}(a, B), \qquad (4)$$

$$h_2(A \to B) = K_{min}^{th}\{d_i^{sort}(a, B)\}, \qquad (5)$$

$$h_3(A \to B) = \frac{1}{N_A} \sum_{a \in A} d(a, B), \qquad (6)$$

$$h_4(A \to B) = \sqrt{\frac{1}{N_A} \sum_{a \in A} d^2(a, B)}, \qquad (7)$$

$$h_5(A \to B) = \frac{1}{N_A} \sum_{a \in A} \rho(d(a, B)), \qquad (8)$$

$$h_6(A \to B) = \frac{1}{K} \sum_{i=1}^{K} d_i^{sort}(a, B), K \leq N_A \qquad (9)$$

$$h_7(A \to B) = \frac{1}{(1-2\alpha)N_A} \sum_{i=\alpha N_A+1}^{N_A - \alpha N_A} d_i^{sort}(a, B), \qquad (10)$$

where $d_i^{sort}(a, B)$ means distances from points a to image B sorted in the increasing order; K_{min}^{th} denotes the K-th ranked distance value after sorting all distances $d(a, B)$; $\rho(d)$ is a simple limiting function, taken to be $\rho(d) = d$ for $d < \tau$ and $\rho(d) = \tau$ for $d \geq \tau$ (we used $\tau = 2.5$); α is the trimming parameter (we used 0.2). Similarly we define functions $h_l(B \to A)$, $l = 1, 2, .., 7$. All the distances from a point to a set use some basic norm, such as the city-block, Euclidean, chessboard distances on the $2D$ image space.

In the following we will denote bidirectional distances based on $h_l(.,.)$ as $H_l(.,.)$ for $l = 1, 2, .., 7$, consistent with Eqs. 4 to 10. Properties of these measures can be found in [2, 4, 3].

One advantage of using Hausdorff distance based dissimilarity measures is the ease of computation: Hausdorff distances express local distance transform (DT) information and approximate DTs are obtained by simple mask convolutions [7]. Furthermore distance maps are inherently not as sensitive to translations as the square-error based measures. Consequently error surfaces are smoother in contrast to error surfaces resulting from correlation criteria which tend to have spurious minima. This property of the error surfaces resulting from the Hausdorff measure enables one to carry out the search on a sparser set of points.

2.3. A combined distance-intensity based measure

The new dissimilarity measure proposed, $D(.,.)$, penalizes the discrepancies both between spatial coordinates and between gray values[11].

$$D(\mathbf{A},\mathbf{B}) = \sqrt{k\sum_{i,j=w}^{N_2-w}[d(a_{ij},\mathbf{B}_{W_{ij}}) + d(b_{ij},\mathbf{A}_{W_{ij}})]^2} \quad (11)$$

where $k = 1/2(N_2 - 2w)$, $\mathbf{A}_{W_{ij}}$ and $\mathbf{B}_{W_{ij}}$ are subimages of the images $\mathbf{A}$ and $\mathbf{B}$ limited in the spatial domain by the square window W of size $(2w+1)\times(2w+1)$ centered at the point (i,j). To avoid edge effects we let window centers roam only within $(N_2 - 2w)\times(N_2 - 2w)$. The distance function d has been computed as in Eqs. 12-15 .

$$d(a_{ij},\mathbf{B}) = min_{b_{lm}\in\mathbf{B}}\{d(a_{ij},b_{lm})\}. \quad (12)$$

$$d(b_{ij},\mathbf{A}) = min_{a_{lm}\in\mathbf{A}}\{d(b_{ij},a_{lm})\}. \quad (13)$$

Throughout experiments we used *the chess* and *the city-block metrics* as the distance function d between two pixels

$$d^{city}(a_{ij},b_{lm}) = (|i-l|+|j-m|)/N_1 + |a_{ij}-b_{lm}|/G, \quad (14)$$

$$d^{ch}(a_{ij},b_{lm}) = max\{|i-l|/N,|j-m|/N_1,|a_{ij}-b_{lm}|/G\} \quad (15)$$

as they are faster to compute. In Eqs. 15 and 16, N_1 is the size of the test image and G is the range of gray values. Note that the distance measure (Eq. 11) is more flexible in that it is tolerant of both shape deformations and gray level differences.

3. THE PROPOSED MATCHING SCHEME

3.1. Template analysis and image binarization

Given a template we may want to *determine* its nature as to whether it contains any objects with prominent edges surrounding homogeneous regions, whether it is textured or finally whether it is fine enough to be noise. A decision based on a simple density analysis of the image intensity peaks usually suffices [1]: We compute a peak map $Z = \{z_{ij}\}$ of our template image such that z_{ij} is the number of intensity extrema in a window of size $v \times v$ centered at the point (i,j). The noise/textured/nontextured decision can be made by applying two thresholds to the mean, $\overline{z}$, (over all windows) of the extrema density. No further processing is made if the template describes noise.

If the template represents one or more objects, the edges may be extracted and used as the binary features. For this purpose it was sufficient to use the output of the Canny edge detector (with $1.0 \leq \sigma \leq 2.0$.)

On the other hand for templates that are textured, it is difficult to come up with universal binary features for matching. In our scheme we simply extract intensity extrema: Such a binary map facilitates at least the coarse step of matching, and in our experiments it proved to be a robust indicator.

The binarization of the test image is carried out using the same parameters as for the template image. The output of this stage is two binary models A_b and B_b of the initial nonbinary images.

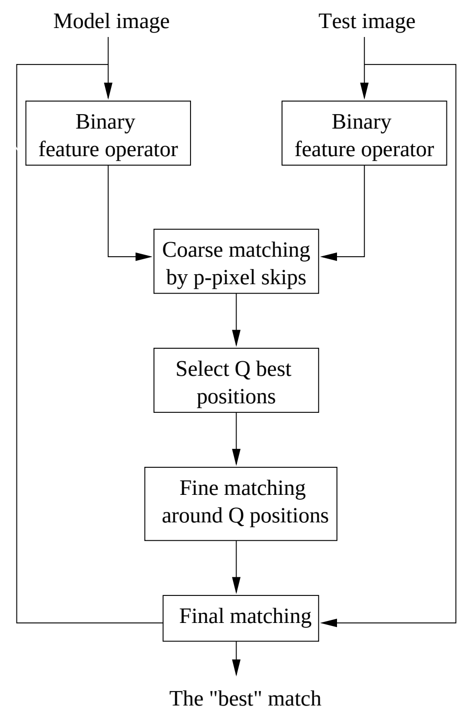

Figure 1: Three stage coarse-to-fine-final matching with recourse to original image pair.

3.2. Coarse matching

Coarse matching is performed between binary images A_b and B_b. First, we carry out the Distance Transform of the binarized template B_b to obtain the distance map V_{B_b}, using an integer valued metric such as one of the the city-block, chessboard or pseudo-Euclidean metrics.

Then we superimpose the map V_{B_b} on A_b at positions $(x,y) = (mp+(p-1)/2, np+(p-1)/2)$, for all possible integers m and n such that the translated binary map B_b stays within the borders of the test image. That is, we do not visit all possible locations one by one, but a regularly sampled subset of it, by skipping p units (pixels) in horizontal and vertical directions. The skipping parameter p is an odd integer greater than 1.

For each of the positions (x, y) indicated above, we calculate $H_l(A_b, B_b \oplus (x, y))$, $l = 1, 2..7$, and select Q positions yielding smallest values of H_l. Q denotes the ceiling on the number of match positions we want to investigate and we call these Q positions *coarse matches.*

3.3. Fine and final matchings and choice of parameters

At the fine matching stage we analyse the neighbourhoods of size $p \times p$ centered at the coarse positions and find the best matching position in every neighbourhood. We call these updated Q positions the *fine matches*. Straightforward template matching needs testing of $(N_1 - N_2)^2$ variants. Using skipping approach, we test just $(N_1 - N_2)^2 p^{-2} + Qp^2$ variants. For a given number, Q, of selected coarse positions the optimum skip parameter that minimizes the number of distance calculations is then given by

$$p_{opt} = \frac{(N_1 - N_2)^{1/2}}{Q^{1/4}} \qquad (16)$$

Skipping by this optimum amount, the total number of distance calculations becomes

$$C_{min} = 2Q^{1/2}(N_1 - N_2) \qquad (17)$$

which reflects an order of magnitude reduction in complexity, as opposed to exhaustive template matching with $C = (N_1 - N_2)^2$. In practice we use the closest odd integer to p_{opt} as the skipping parameter p. To give an example, for $N_1 = 256$, $N_2 = 32$, if $Q = 16$ is selected then $p_{opt} = 7$ and the number of distance calculations is $1,792$ whereas $50,176$ calculations are needed in straightforward template matching.

The skip parameter that is optimum for computational complexity is not necessarily optimum for matching, since it depends on the number Q of coarse matches hypothesized. With high p values, the sampling of the error surface is sparse and one needs a high Q value which is also limited by $((N_1 - N_2)/p)^2$. With low values of p, an occurence of the scene may cause spurious weaker matches in its surrounding, which may mask other true occurences, unless special care is taken in the list of coarse matches. In practice, a reasonable choice of p is

$$p = min\{p_{opt}, \frac{N_2}{2}\} \qquad (18)$$

where N_2 is the size of the template.

Finally we apply, to the original image pair $(\mathbf{A}, \mathbf{B})$, some measures of matching (section 2) for the selected positions output by the fine matching stage.

4. EXPERIMENTAL STUDY

Experiment 1: Matching an object. The 32×32 template "eight" is extracted from the first frame of the 240×240 "football" sequence and compared to the fifth frame of the same sequence using $H_l(.,.)$, $l = 1, 2, .., 7$, and $3 \leq p \leq 2p_{opt}$. The number of *coarse* matches was taken to be $Q = 10$. The optimum skip parameter is $p_{opt} = 7$.

Table 1 shows the results obtained with the Modified Hausdorff distance (MHD, $H_3(.,.)$) for which seven out of ten *fine* matches are found to be below a threshold of 1.90. The first column of the table gives the relative positions of the template with respect to the test image (fine matches), sorted in ascending order of $H_3(.,.)$ values which are in the second column. The third and fourth columns give the figures of the normalized correlation and the proposed distance measure (subsection 2.3) respectively. Figure 2 is the template overlaid with the test image at the best match (116,1).

(x,y)	$H_3(A_b, B_b^t)$	$CO(\mathbf{A}, \mathbf{B^t})$	$D(\mathbf{A}, \mathbf{B^t})$
(116,1)	1.570	0.944	0.051
(123,15)	1.622	0.878	0.104
(122,5)	1.689	0.885	0.102
(118,13)	1.700	0.914	0.108
(123,14)	1.705	0.885	0.097
(127,4)	1.725	0.898	0.099
(127,10)	1.857	0.898	0.091

Table 1: Results of experiment 1. Superscript ^{t} denotes translation.

Experiment 2: Matching pieces of texture. The 48×48 template is a high-pass filtered version of the textured surface taken from the Brodatz album (D79). The test image is an original 256×256 piece of the same texture. The number of *coarse* matches was taken to be $Q = 8$. The optimum skip parameter is $p_{opt} = 7$, however $p = 5$ was used in the experiment. Table 2 shows the results obtained with the MHD. Figure 3 is the template overlaid with the test image at two best positions, (74,63) and (123,52).

(x,y)	$H_3(A_b, B_b^t)$	$CO(\mathbf{A}, \mathbf{B^t})$	$D(\mathbf{A}, \mathbf{B^t})$
(74,63)	1.053	0.924	0.039
(76,62)	1.057	0.937	0.038
(123,52)	1.080	0.924	0.041
(108,64)	1.124	0.927	0.047
(126,52)	1.127	0.926	0.043
(137,52)	1.137	0.916	0.043
(54,49)	1.154	0.929	0.044
(48,188)	1.158	0.915	0.088

Table 2: Results of experiment 2.

Figure 2: Best match (116,1) of experiment 1.

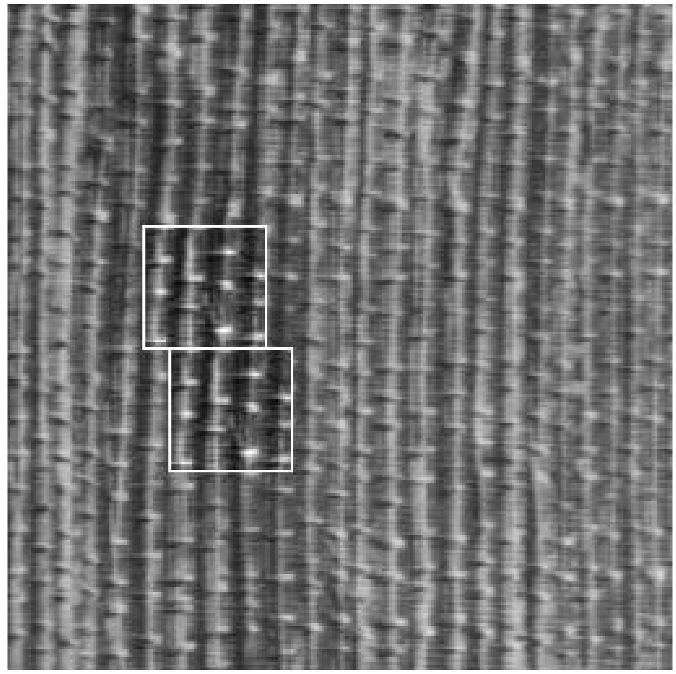

Figure 3: Best matches ((74,63) and (123,52)) of experiment 2.

The first experiment demonstrates the effectiveness of the coarse-to-fine matching approach. In our experiments, the true matching position is consistent for each of the skipping values in the range $(3, N_2/2)$. For $p = p_{opt}$, extremely fast operation was achieved.

The second experiment demonstrates a successful generalization of the approach to texture images: We achieved accurate detection of many occurences of the sample texture pattern in the original textured surface. Furthermore, the set of intensity extrema has proven to be a reasonable set of binary texture features for matching purposes.

5. CONCLUSIONS

In this study, we have proposed an efficient three-stage template matching scheme which is applicable to a wide range of images. In general, most accurate results are obtained using the pseudo-Euclidean metric for the DT, MHD as the geometrical mismatch measure and the city-block metric for the combined diversity measure. It was observed that the normalized correlation gave spurious responses to smooth regions in the test image whereas the combined diversity measure sharply distinguish the best match in almost all cases.

6. REFERENCES

[1] Karu K., Jain A.K. and Bolle R.M., Is there any texture in the image?, Pattern Recognition, Vol.29, 1437-1446, 1996.

[2] Huttenlocher D.P., et al., Comparing images using the Hausdorff Distance, IEEE Trans. on Pattern Anal. and Machine Intell., Vol. PAMI-15, No.9, 850-863, 1993.

[3] Dubuisson M.-P. and Jain A.K., A modified Hausdorff Distance for object matching, Proc. 12th Int. Conf. Pattern Recognition, 566-568, Israel, Oct.1994.

[4] You J. et al., A guided image matching approach using Hausdorff distance with interesting points detection, Proc. ICIP'94, Vol.I, 968-972, 1994.

[5] Kwon O.-K. et al., New Hausdorff Distances based on robust statistics for comparing images, Proc. ICIP'96, Vol.III, 21-24, 1996.

[6] Candocia F. and Adjouadi M., A similarity measure for feature matching, IEEE Trans. Image Proc., Vol.6, No.10, 1460-1464, 1997.

[7] Borgefors G., Hierarchical chamfer matching: A parametric edge matching algorithm, IEEE Trans. on Pattern Anal. and Machine Intell., Vol. PAMI-10, No.6, 849-865, 1988.

[8] Haralick R.N., Shapiro L.G., Computer and Robot Vision, Part II, Addison-Wesley, 1992.

[9] Goshtasby A., Gage S.H., and Bartholic J.F., A two-stage cross correlation approach to template matching, IEEE Trans. on Pattern Anal. and Machine Intell., Vol. PAMI-6, No.3, 374-378, May 1984.

[10] Rosenfeld A. and Vanderburg G.J., Coarse-fine template matching, IEEE Trans. on SMC, 104-107, Feb.1977.

[11] Starovoitov, V. (1995). Towards a measure of diversity between gray-scale images, Lecture Notes in Computer Science,970, 214-221.

[12] Pratt W.K., Digital Image Processing, Willey:N.Y.,1991.

[13] Gonzalez R.C. and Woods R.E., Digital Image Processing, Addison-Wesley, 1993.