Machine Vision and Applications (1995) 8:173–178

Machine Vision and Applications

An entropy based method for extracting robust binary templates

Claudio De Stefano, Francesco Tortorella, Mario Vento*

Department of Computer and Systems Engineering, University of Naples "Federico II", Via Claudio 21, I-80125, Naples, Italy*

Abstract. In applications based on template matching, the design of the template is a critical point and has a considerable effect on the overall performance that can be obtained. In this paper we present a method, applicable to binary images, for extracting robust templates. Starting from a provisional prototype (defined as the image that minimizes its overall distance from the samples of a training set), the template is obtained by eliminating unreliable pixels determined by means of an entropy-based criterion. The method is compared experimentally with the matched filtering technique in the recognition of symbols on topographic maps and shows promising results regarding the recognition rate and the computational cost of the matching process.

Key words: Template matching – Symbol recognition – Entropy-based methods – Map processing

1 Introduction

Template matching is a valuable tool in image analysis and is widely used in fields (e.g. recognition of symbols in documents and on maps, image registration, and industrial inspection) in which more sophisticated approaches, such as those based on structural descriptions, are not particularly helpful. Several methods have been proposed to reduce the computational cost of this technique through new algorithms (Davies 1988; Margalit and Rosenfeld 1990; Vanderbrug and Rosenfeld 1977) or specialized hardware (Prasanna Kumar and Krishman 1989; Zapata and Benavides 1990) and to overcome the problems induced by rotated images (Goshtasby 1985; Ryall and Sandor 1989).

However, extracting the template so as to ensure good recognition performance is still an on-going problem, even in the case of binary images. In real applications, the various occurrences of the pattern to be recognized generally exhibit shape variations and are affected by the noise introduced in the acquisition phase. Consequently in the determination of the template only the information needed to recognize reliably the real occurrences of the objects should be preserved, while unreliable information related to nonmeaningful shape variations should be discarded. At present, the most commonly used approaches are those based on matched filtering (Davies 1992) or manual design. In the first case, however, the extracted template is composed of real numbers, and expensive floating-point operations are needed in the matching stage, while in the second case the template is frequently unreliable because of the subjective criteria applied.

In the proposed approach, we start from a training set containing various occurrences of a given symbol to evaluate a provisional prototype P, defined as the pattern that minimizes its overall distance from the samples of the training set considered. Successively, unreliable pixels, characterized by an entropy greater than a fixed threshold H_0^*, are discarded. The threshold H_0^* is determined so as to maximize the discriminating power between real occurrences of the symbol and other incorrect pixel configurations. The final template T, using only the most reliable pixels of P, ensures good recognition reliability and does not require a computationally expensive matching phase, since it contains zeros and ones.

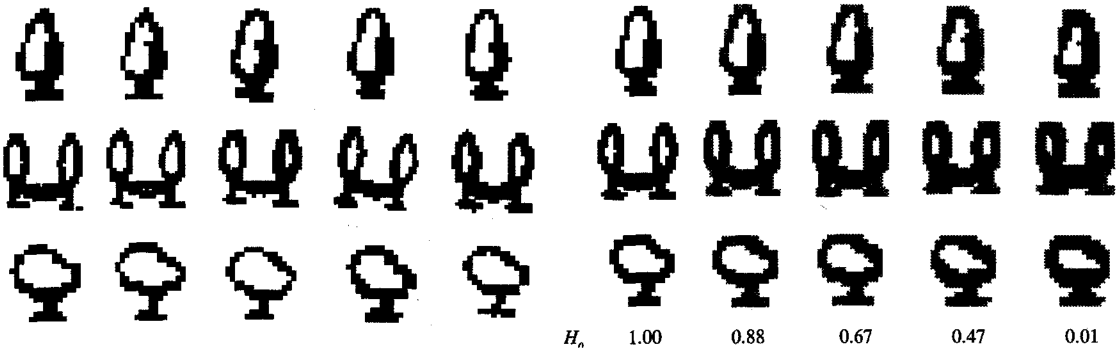
2 The proposed method

For a given symbol W, the provisional prototype P is determined starting from a training set of real samples adequately representing the real occurrences of the symbol. Denoting the training set of W with $\{S_k\}$, $k = 1..N$, the provisional prototype P is defined as the pattern that minimizes the overall distance between itself and the samples S_k. P can be formally defined as the pattern that minimizes the quantity $\sum_k d(P, S_k)$ with $k = 1 \dots N$, where the distance d is the sum of absolute pixel differences normalized with respect to the size:

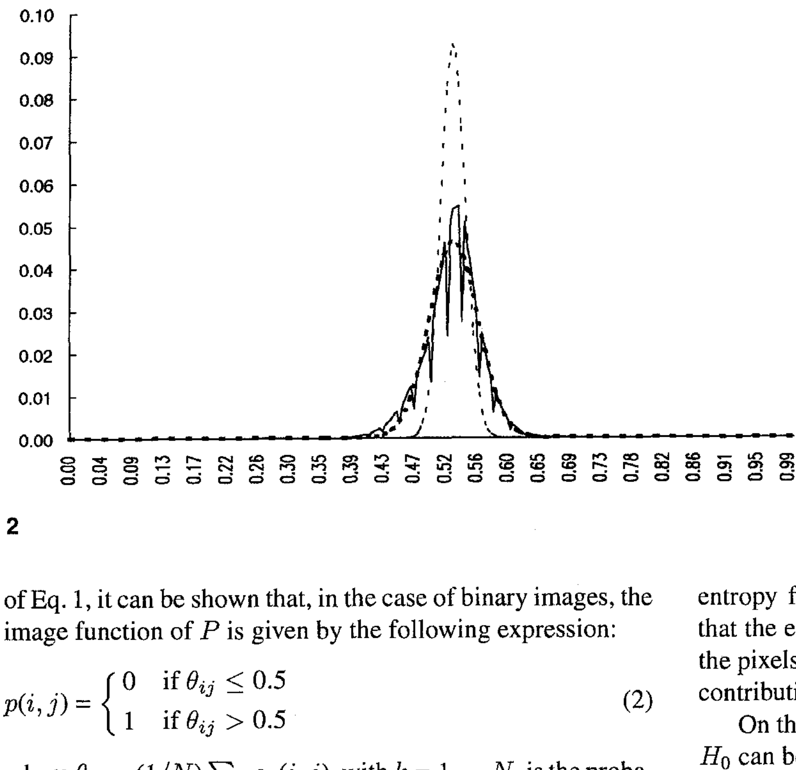
$$d(A, B) = (1/Z) \sum_{ij} |a(i, j) - b(i, j)| \tag{1}$$

where $a(i, j)$ and $b(i, j)$ are the image functions of two generic patterns A and B, respectively, and Z is their size. This distance expression has been chosen because of its peculiar features of robustness towards noise, as can be seen from a comparison pre sented in Secilla et al. (1988). By using the distance

* e-mail: vento@nadis.dis.unina.it
Correspondence to: M. Vento



1



2

Fig. 1. Some samples of the three symbols A, B, and C (*left*) and some templates of the same symbols (*right*), obtained by using different values for the entropy threshold H_0 (*dotted areas* indicate *don't care* points). Note that the symbols are not represented with the same scale; refer to Sect. 3 for details regarding their real dimensions

Fig. 2. The actual distribution of the distance between a template and an image containing only background, compared with the approximation obtained using Vanderbrug and Rosenfeld's model (V&R model) and the modified V&R model. — Actual distribution; - - - V&R Model; - - - Modified V&R model

of Eq. 1, it can be shown that, in the case of binary images, the image function of P is given by the following expression:

$$p(i,j) = \begin{cases} 0 & \text{if } \theta_{ij} \leq 0.5 \\ 1 & \text{if } \theta_{ij} > 0.5 \end{cases} \quad (2)$$

where $\theta_{ij} = (1/N)\sum_k s_k(i,j)$, with $k = 1 \ldots N$, is the probability that the pixel (i,j) of P assumes the value 1 and $s_k(i,j)$ is the image function of S_k.

Once P has been determined, the reliability of its single pixels is evaluated to obtain the final template T. To this end, consider that, if the relative frequency of a pixel is close to 0.5, any value it assumes is characterized by a considerable uncertainty which increases the probability of supplying a wrong contribution in the matching stage. However, the values of pixels denoted by relative frequencies close to 0 or 1 are much more invariant in the training set and consequently relatively certain.

To obtain a quantitative and concise evaluation of the reliability of the pixels of P, the entropy related to the single pixel is considered:

$$H_{ij} = -\theta_{ij}\log_2\theta_{ij} - (1-\theta_{ij})\log_2(1-\theta_{ij}) \,. \quad (3)$$

It follows from the definition that the most reliable pixels, corresponding to $\theta_{ij} \to 0$ or $\theta_{ij} \to 1$, have values of H_{ij} close to 0, while for totally uncertain pixels, $\theta_{ij} \to 0.5$, the entropy function reaches its maximum. It is therefore clear that the entropy function can be used successfully to classify the pixels of P as reliable or unreliable, thus making a useful contribution to the evaluation of the final template T.

On the basis of these considerations, an entropy threshold H_0 can be fixed, below which the corresponding reliability is acceptable. In this way, only the pixels of P with an entropy value lower than H_0 are used to build the template T, while the remaining pixels are considered unreliable and are marked as *don't care* in T. It is obvious that the bit map of the obtained template T is consequently dependent on the value of H_0. The pixels of $T = T(H_0)$ are found according to the following rule:

$$t(i,j) = \begin{cases} 0 & \text{if } H_{ij} < H_0 \text{ and } \theta_{ij} \leq 0.5 \\ 1 & \text{if } H_{ij} < H_0 \text{ and } \theta_{ij} > 0.5 \\ \text{don't care} & \text{if } H_{ij} \geq H_0 \end{cases} \,. \quad (4)$$

It is worthwhile pointing out that for $H_0 = 1$ all the pixels in the provisional prototype P will be considered in the template T (i.e., no pixel of P is marked as *don't care*), thus giving $T = P$. Conversely, if $H_0 = 0$, the template T contains only pixels of P corresponding to values of $\theta_{ij} = 1$ or $\theta_{ij} = 0$ (i.e. having the same value in all the samples of the training set), while all the remaining pixels of T are marked as *don't care*. In Fig. 1 some samples belonging to the training sets of the experiments are shown together with the templates obtained

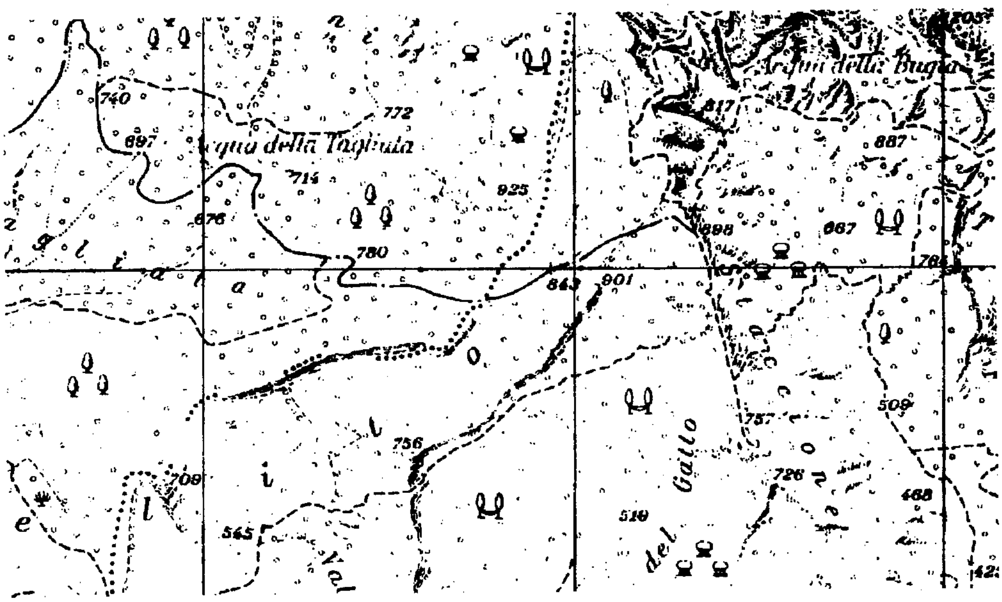


Fig. 3. A portion of a map used

for some values of H_0. Of course, the value of the threshold H_0 must be determined with care. If H_0 is too high, the template might contain too many unreliable pixels; if H_0 is too low, the template will be composed of pixels that are more reliable, but too few in number. Clearly, in both cases the matching process could be seriously affected by erroneous recognitions. The optimal value of H_0 is defined as that threshold value H_0^* corresponding to the best distinction between real occurrences of W and the background.

To this end, it is necessary to analyze two kinds of distribution functions jointly: the distribution of the distance between T and the samples of W, and the distribution of the distance between T and the background. With the introduction of *don't care* pixels, Eq. 1 is no longer a distance because the triangular inequality is not verified. However, for the sake of simplicity, the term "distance" is still used in the following, since the lack of this property does not affect our analysis.

A good estimate of the actual separation between the distributions is given by the separability index ρ (Venkatesh and Psaltis 1989) defined by the following expression:

$$\rho = (D_S - D_B)^2/(\sigma_S^2 + \sigma_B^2)\,. \tag{5}$$

In Eq. 5, $D_S(D_B)$ is the mean distance between T and the samples of the training set (between T and the background), while $\sigma_S(\sigma_B)$ is the relative standard deviation. It follows from the definition that the higher ρ is, the easier it is to distinguish the real occurrences of W from the background, thus giving a more robust match.

The parameters D_S and σ_S can be evaluated by matching T with the samples of W. In a similar way D_B and σ_B could be found by matching T with portions of the background of the images on which the occurrences are searched for, but in this case the process would be computationally very expensive. Therefore, an alternative approach has been used: a statistical model of the distribution of the distance between T and the background on the basis of some known features the images has been considered. The adopted model was introduced in Vanderbrug and Rosenfeld (1977) and assumes that the distance distribution can be approximated by a gaussian one with mean and standard deviation:

$$D_B = \gamma - 2\gamma\delta + \delta \tag{6a}$$

$$\sigma_B = \sqrt{\gamma(1-\gamma)/n} \tag{6b}$$

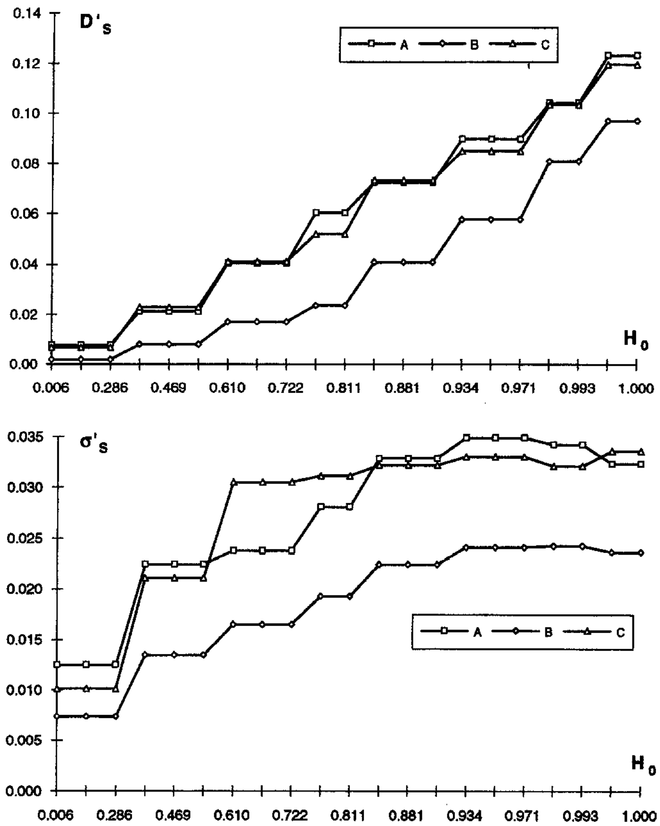
where γ and δ are the probabilities that the value 1 will occur in the background and in the template, and n is the number of points in the template. As reported by Vanderbrug and Rosenfeld (1977), the larger n is, the more accurate the approximation. Since, in our case, the template sizes are quite small, a correcting factor should be introduced into the model to obtain an approximation that is more accurate. The experiments have pointed out that this is ensured if the following expression is used for σ_B:

$$\sigma_B = \alpha\sqrt{\gamma(1-\gamma)/n} \tag{6b$'$}$$

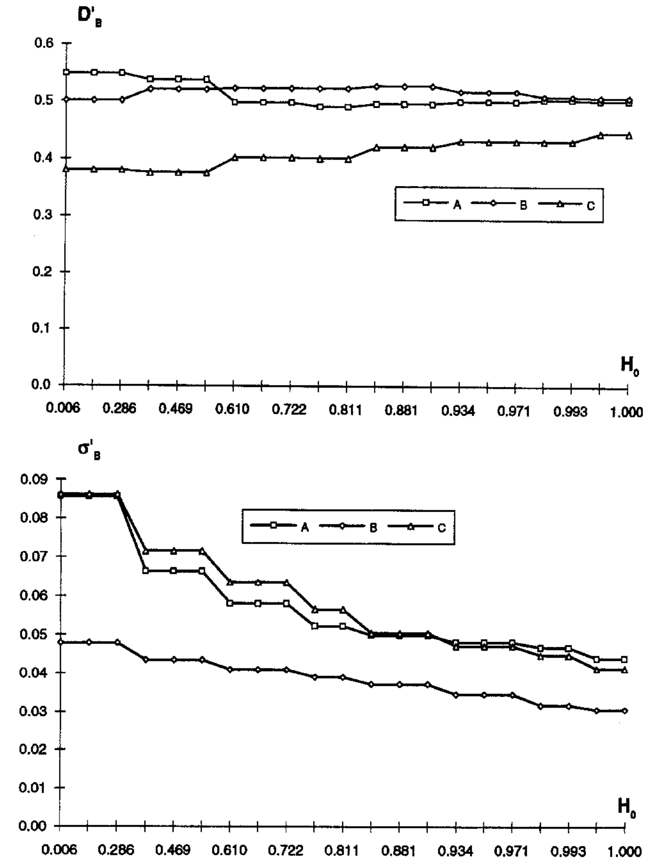
where α is a constant with a value close to 2. In Fig. 2 the actual distribution of the distance between a given template and the background of an image is shown, together with two approximating gaussian distributions, for both of which Eq. 6a has been used to estimate the mean, while Eq. 6b and 6b′ have been employed for the standard deviations. It can be clearly seen that the second expression for σ_B gives much better results than the first one.

From the previous discussion, it follows that the value of H_0 maximizing ρ should be chosen for H_0^*. Since the trend of ρ as a function of H_0 is not known analytically, H_0^* has been determined by means of a numerical procedure consisting of the following steps:

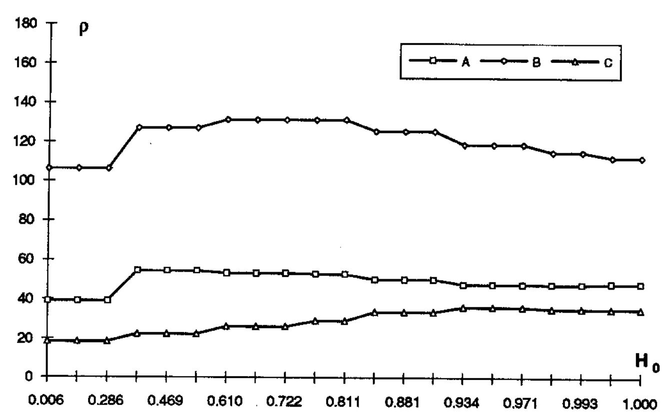
1. For each value of H_0 ranging from 0 to 1 with a fixed step, the corresponding template $T(H_0)$ is evaluated.
2. For each H_0 considered in step 1, the separability index is evaluated on the basis of the parameters D_B, σ_B, D_S and σ_S.



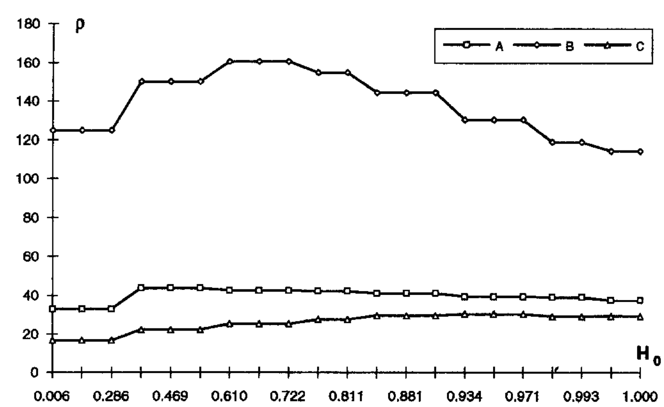
4



5



6



7

Fig. 4. Trends of the mean D'_B and of the standard deviation σ'_B of the distance between the template and the test set samples

Fig. 6. Trends of the separability index ρ vs. the threshold H_0, calculated for each of the symbols, starting from the curves in Figs. 4 and 5

Fig. 5. Trends of the mean D'_B and of the standard deviation σ'_B of the distance from the template to some parts of the map without the given symbols

Fig. 7. Trends of the separability index ρ vs. the threshold H_0, estimated for each symbol by means of the model presented in Sect. 2. Note that the maxima are reached at the same points as in Fig. 6

3. From the values of ρ calculated in step 2, the maximum is picked out, and the relative H_0 is chosen as H_0^*.

Finally, the template corresponding to the obtained H_0^* is assumed as the template for the symbol W.

3 Experimental results

The method has been tested on the recognition of symbols on topographic maps (scale 1:25 000) of the Italian Army Geographical Institute (IGM), scanned with a resolution of 200 dpi. This kind of maps (whose size is 40 cm × 40 cm) generally contains about 20 different symbols of fixed size (typically 2–4 mm). Because of the distorsions introduced by the printing process and the high density of cartographic entities (lines representing roads and rivers, level curves, letters, etc.), the symbols are often superimposed and/or similar to particular background configurations (Fig. 3); consequently, the application of template matching techniques becomes quite problematic. For our experiments, we have chosen three symbols (say A, B, and C) whose recognition was particularly critical. For each symbol, a training set (containing 30 samples) and a test set (containing 150 samples) were extracted from the maps. The dimensions in pixels of the three symbols are approximately 10×19 for the first symbol, 25×24 for the second one and 14×15 for the third one. Some occurrences of the symbols have already been presented in Fig. 1, and an example of the used maps is shown in Fig. 3.

The first issue examined in the experiments was the correlation between H_0 and the performance of the template in the matching phase. In other words, the template's ability to recognize real occurrences and reject the false ones was evaluated by varying the threshold H_0 to verify the agreement with the theoretical model. For this purpose, the mean D'_S and the standard deviation σ'_S of the distance between the template extracted with a given H_0 and the samples of the test set were evaluated for each symbol with H_0 ranging from 0 to 1. The curves obtained (Fig. 4) help in estimating the robustness of the template with respect to the variations affecting the real occurrences, which make their bit maps quite different from those of the samples belonging to the training set. Low values for both parameters denote good recognition performance. In fact, in this case the distance distribution is characterized by low values concentrated near the mean. The diagrams also point out that D'_S decreases, while σ'_S remains approximately constant when H_0 changes from 1 to lower values. This is due to the removal of the least reliable pixels from the template. If H_0 is further decremented, D'_S holds its trend, while σ'_S begins to decrease. This is easily explained by the increase of *don't care* points, so the samples are matched on the basis of ever more reliable pixels.

To assess the reliability of the template in discriminating the real occurrences from the background as H_0 varies, the mean D'_B and the standard deviation σ'_B of the distance between the template and some portions of maps without instances of the given symbols were evaluated versus H_0. Obviously, the higher the mean and the lower the standard deviation, the more distinguishable the background is from the real occurrences.

The diagrams presented in Fig. 5 show that, when H_0 decreases, D'_B is fairly constant, while σ'_B increases considerably. This means that when H_0 (and consequently the number of points in the template) is decremented, the distance between the template and the background is characterized by lower values that are also more scattered with respect to the mean, thus weakening the discriminating power of the template. Since the performance, in terms of robustness and reliability, has opposite trends with respect to H_0 (as theoretically stated in the previous section), ρ must be analyzed as a function of H_0

Table 1. The numbers of points in the templates vs. H_0

H_0	A	B	C
0.006	43	235	15
0.169	65	307	23
0.469	94	377	79
0.544	102	402	114
0.669	137	416	127
0.811	151	465	159
0.954	178	528	183
1.000	187	588	204

H_0, entropy threshold; A, B, and C, recognition test symbols

to determine the value of the entropy threshold that ensures the best overall performance. Figure 6 shows the trends of ρ calculated for each symbol starting from the previous curves. It can be seen that the template drawn by means of the proposed method gives a performance that is always better than the prototype P. In fact, the value of ρ associated to the latter corresponds to $H_0 = 1$ and is always lower than the value related to T.

The effectiveness of the method can be roughly, but meaningfully evaluated by comparing the analyzed curves with the trends of ρ obtained by using the estimates described in the previous section, which are shown in Fig. 7.

Note that, even though the scales are different, the trends are practically the same, and the maxima are reached at the same values of H_0. This means that the method is extremely reliable in estimating the value of H_0 corresponding to the highest separability that can be achieved for real images.

According to the criterion illustrated in the previous section, the templates selected are those corresponding to the values of H_0 maximizing ρ, i.e. 0.469, 0.669, and 0.954 for A, B, and C respectively. Table 1 illustrates how the size of the templates change when H_0 varies.

Figure 8 shows the percentages of correct recognitions achieved by using the templates drawn for the values of H_0 on 25 test images of size 600×600 pixels. There are 650 occurrences on the images for symbol A, 527 for symbol B, and 688 for symbol C. As can be clearly seen from the diagrams shown, the best performance for each symbol is achieved with the templates associated with the values of H_0 maximizing ρ.

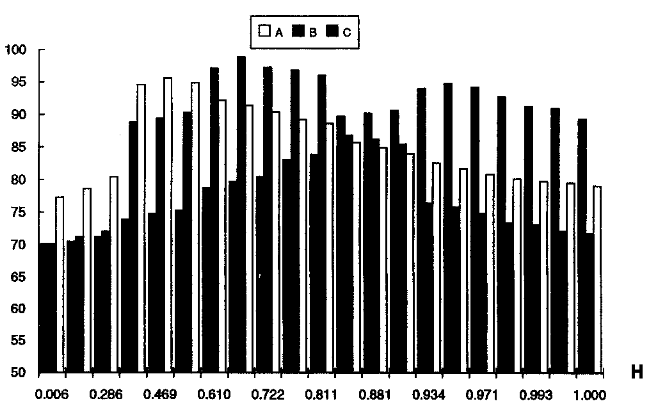


Fig. 8. The recognition rates vs. the threshold H_0 for each of the symbols

Table 2. Performance of the proposed method compared to the matched filtering technique

Symbol	Proposed method				Matched filtering			
	Percentage recognized	T_{sun}	T_{pc}	Size	Percentage recognized	T_{sun}	T_{pc}	Size
A	95.54	10.54	30.94	94	95.10	76.53	332.3	120
B	98.82	13.26	41.23	416	96.63	96.12	467.7	489
C	94.73	11.50	33.16	183	94.12	80.25	361.2	197

T_{sun} and T_{pc} represent the average times (in seconds) necessary to process a 600×600 pixels image on the SUN and the PC 486, respectively

To evaluate the performance of the method with respect to other techniques, the same test was carried out with templates drawn according to the matched filtering approach. The recognition rates obtained and the average times employed in the matching phase by the two methods on the test images are presented in Table 2, together with the sizes of the templates.

Examination of Table 2 shows that the proposed method seems to be practically equivalent to the matched filtering technique regarding the recognition rate, while the processing time is significantly reduced. This speed-up depends on two factors. First, the reduction of the number of pixels in the template T reduces the operations needed and second, the use of only zeros and ones in the template instead of reals greatly reduces the computational load.

This improvement is machine dependent since it is related to the ratio of the time spent on a floating-point operation to the time needed for an integer arithmetic operation, which is clearly dependent on the architecture employed. This aspect is present in the results shown in Table 2, where the processing times corresponding to a SUN SPARCstation *IPX* and a PC 486 are reported. It is worth noting that, for the SUN system, equipped with a powerful floating-point unit, the improvement is somewhat lower than for the PC.

4 Conclusions

In this paper, a new method is proposed for the extraction of robust templates from binary images. Starting from a provisional prototype P, which can be evaluated straightforwardly from a training set of samples, unreliable pixels characterized by an entropy greater than a fixed threshold H_0^* are successively discarded. The threshold H_0^* is determined so as to maximize the discrimination power between real occurrences of the symbol and other incorrect pixel configurations.

The proposed method has been tested for the recognition of symbols on topographic maps, and the results have been compared with those obtained by another well known technique (matched filtering). The experimental results proved interesting: even though the recognition rate is comparable to that achieved for the matched filtering technique, the proposed approach is more convenient in terms of the computational load of the matching process.

Acknowledgements. The helpful comments of the referees in improving the earlier draft of this paper are gratefully acknowledged.

References

Davies ER (1988) Tradeoffs between speed and accuracy in two-stage template matching. Signal Processing 15:351–363

Davies ER (1992) Procedure for generating template masks for detecting variable signals. Image Vis Comput 10:241–249

Goshtasby A (1985) Template matching in rotated images. IEEE Trans Patt Anal Machine Intell 7:338–344

Margalit A, Rosenfeld A (1990) Using probabilistic domain knowledge to reduce the expected computational cost of template matching. Comput Vis Graph Image Processing 51:219–234

Prasanna Kumar VK, Krishman V (1989) Efficient parallel algorithms for image template matching on hypercube SIMD machines. IEEE Trans Patt Anal Machine Intell 11:665–669

Ryall TG, Sandor J (1989) Statistical pattern matching. Patt Recogn Lett 9:163–168

Secilla JP, Garcia N, Carrascosa JL (1988) Template location in noisy pictures. Signal Processing 14:347–361

Vanderbrug GJ, Rosenfeld A (1977) Two-stage template matching. IEEE Trans Comput 26:384–393

Venkatesh S, Psaltis D (1989) Binary filters for pattern classification. IEEE Trans Acoustics Speech Signal Processing 37:604–611

Zapata EL, Benavides JI (1990) Image template matching on hypercube SIMD computers, Signal Processing 21:49–60

Claudio De Stefano was born in Naples (Italy) in 1961. He received the laurea degree in Electronic Engineering in 1990 and the Ph.D. degree in Electronics and Computer Engineering in 1994 both from the University of Naples "Federico II". His interests are in the fields of Optical Character Recognition, Neural Networks and Document Processing.

Francesco Tortorella was born in Salerno (Italy) in 1963. He received the laurea degree in Electronic Engineering in 1991 and he is currently a Ph.D. student at the Dipartimento di Informatica e Sistemistica of the University of Naples "Federico II". His interests are in the field of Optical Character Recognition, Map and Document Processing.

Mario Vento was born in Naples (Italy) in 1960. He received the laurea degree in Electronic Engineering in 1984 and the Ph.D. degree in Electronics and Computer Engineering in 1989 both from the University of Naples "Federico II". Currently he is an Assistant Professor at the Dipartimento di Informatica e Sistemistica of the same University. His interests are in the fields of Optical Character Recognition, Neural Networks and Document Processing.