Hand-drawn symbol recognition in graphic documents using deformable template matching and a bayesian framework

Ernest Valveny, Enric Martí
Computer Vision Center - Computer Science Department
Universitat Autònoma de Barcelona
Edifici O - Campus UAB. 08193 Bellaterra. Spain
{ernest|enric}@cvc.uab.es

Abstract

Hand-drawn symbols can take many different and distorted shapes from their ideal representation. Then, very flexible methods are needed to be able to handle unconstrained drawings. We propose here to extend our previous work in hand-drawn symbol recognition based on a bayesian framework and deformable template matching. This approach gets flexibility enough to fit distorted shapes in the drawing while keeping fidelity to the ideal shape of the symbol. In this work we define the similarity measure between an image and a symbol based on the distance from every pixel in the image to the lines in the symbol. Matching is carried out using an implementation of the EM algorithm. Thus, we can improve recognition rates and computation time with respect to our previous formulation based on a simulated annealing algorithm.

1 Introduction

Deformable template matching and bayesian inference are very well-known techniques which have been applied to a wide number of applications in computer vision where noise, shape distortion or uncertainty make it difficult to identify the objects in an image [2].

In document analysis, and up to now, their application has been restricted to hand-written character recognition [3, 5]. In a previous work [7], we have proposed the application of deformable template matching and bayesian inference to the recognition of hand-drawn graphic symbols, such as those found in many kinds of graphic documents. We argue that these techniques are well-suited methods to handle noise, imprecision and uncertainty inherent to hand-drawing. They can help to overcome some of the drawbacks of previous methods for symbol recognition, usually based on vectorization, feature extraction and structural matching [1]. These methods decrease their efficiency and robustness as long as noise and distortion of hand-drawn symbols increase [6]. In deformable template matching, we work over the binary image. Therefore, we avoid errors and noise produced by vectorization and we can have more complete information. Matching is carried out by deforming the ideal shape of the symbol to get the best fit to the image while keeping some degree of fidelity to that ideal shape. Bayesian inference is used to model uncertainty in the shape of the symbols.

In this work, we modify our previous approach introducing a new similarity function between the symbol and the image based on the distance from every point in the image to the lines in the symbol. With this new measure we can formulate matching in a probabilistic way and we can solve it using an implementation of the EM algorithm. Then, we can achieve an stable solution in a lower time and we can improve recognition rates.

2 Bayesian formulation of symbol recognition

Given an input image I and a set of symbols $\{S_1, \ldots, S_n\}$, the general problem of symbol recognition is stated as the problem of finding the symbol S_i which maximises the probability $P(S_i|I)$, i.e., the probability that given image I we can identify symbol S_i in it. This probability can be related, using the development explained in [8], to the maximum of this expression:

$$P(I|D_i, S_i)P(D_i|S_i) \quad (1)$$

where D_i is any possible deformation of the ideal shape of symbol S_i; $P(D_i|S_i)$ is the prior information about the symbol and it expresses the probability that deformation D_i is a valid representation of it penalizing excessive distortions; $P(I|D_i, S_i)$ is the likelihood that image I corre-

0-7695-0750-6/00 $10.00

sponds to D_i and it measures the distance between the image and each of the deformations of the symbol

$\hat{D}_i$, the maximum of (1), is usually found searching for the minimum of the negative log of it:

$$\begin{aligned} \hat{D}_i &= \arg\min_{D_i} \left(-\log P(D_i|S_i) - \log P(I|D_i, S_i)\right) \\ &= \arg\min_{D_i} \left(E_{int} + E_{ext}\right) \end{aligned} \quad (2)$$

The problem of symbol recognition is reduced to the problem of minimising an energy function, E, composed of two terms: external energy, E_{ext}, which is related to likelihood and it represents a force which tries to deform the ideal symbol as much as possible to get the best match to the input image, and internal energy, E_{int}, which is related to prior probability and it represents a force which tries to keep the deformed symbol as close as possible to the ideal shape. The minimum of E is the equilibrium point between these two opposite forces and it corresponds to the shape of the symbol that best fits the image with the minimum amount of deformation. As the final value of E at this point is related to $P(S_i|I)$, it can be seen as a measure of the degree of correspondence between the input image and the symbol.

3 Definition of energy

3.1 Prior information

As we are concerned with the recognition of lineal symbols we represent the ideal shape of a symbol as a set of straight lines, not necessarily connected. We deform the shape of the symbol applying two kinds of transformations to these lines: global and local deformations. Global deformations translate, rotate and scale all the lines in the same way. They do not change the global shape of the symbol. Local deformations are generated by translating, rotating and scaling each of the lines of the symbol separately. They change the global shape of the symbol, adjusting it in a natural and intuitive way to the shapes produced by hand-drawing.

Prior probability and internal energy are defined assuming that each local deformation of every line follows a gaussian distribution of zero mean over the amount of deformation applied to the line, and that all local deformations of each line are independent. Then, we get the following expression for internal energy:

$$E_{int} = \sum_{i=1}^{n} \left(\frac{t_{x_i}^2}{2\sigma_{t_{x_i}}^2} + \frac{t_{y_i}^2}{2\sigma_{t_{y_i}}^2} + \frac{\sin^2\theta_i}{2\sigma_{\theta_i}^2} + \frac{s_i^2}{2\sigma_{s_i}^2} \right) + K \quad (3)$$

where n is the number of lines in the symbol; $t_{x_i}, t_{y_i}, \theta_i$ and s_i are the amount of translation, rotation and scaling applied to line i; $\sigma_{t_{x_i}}, \sigma_{t_{y_i}}, \sigma_{\theta_i}$ and σ_{s_i} are the standard deviations for translation, rotation and scaling for line i; and K is a constant.

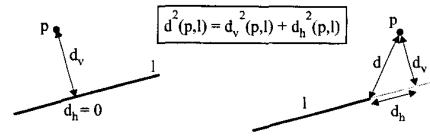

Figure 1. Compoments of distance function.

3.2 Likelihood

We define the similarity measure between an image and a symbol as an explanation measure, i.e, we try to explain all the pixels in the image with some nearby line of the symbol. In this way, we get the maximum exploration ability as we try to deform the symbol to get the best possible fit to all the pixels in the image. Because of this property we must segment the image in advance to avoid noisy pixels introduced by other elements or symbols in the drawing.

Then, likelihood is defined based on the distance between a point and a line. This distance can be expressed, as it is shown in figure 1, as the combination of two terms, d_v and d_h, which can be easily derived from the coordinates of the point p and the parameters defining the line l. We add to this definition a factor taking into account the similarity in orientation between the symbol line and the line in which the image point is located:

$$d'(p,l) = (1 + k_\alpha \cdot \sin^2(\alpha - \beta)) \cdot d(p,l) \quad (4)$$

where α is the orientation of line l and β is the orientation of the image at point p, measured from the analysis of the neighborhood of p; k_α is a factor which controls the influence of the orientation difference in the distance function.

To define likelihood we consider each of the symbol lines (l) as a generator of pixels in the image (p) with a gaussian distribution based on $d'(p,l)$. Then, each pixel has a global probability of being generated equal to the sum of generation probabilities for each of the lines. Assuming independence in the probability of every pixel, we get the following expression for likelihood:

$$P(I|S') = \prod_{i=1}^{n} P(p_i|S') = \prod_{i=1}^{n} \sum_{j=1}^{m} P(p_i|l_j) \quad (5)$$

where I stands for the image, S' for any possible deformation of symbol S, p_i for every pixel in the image, l_j for every line in S' and $P(p_i|l_j)$ is a normal distribution based on $d'(p_i, l_j)$.

4 Energy minimisation with the EM algorithm

Combining prior information (3) and likelihood (5) and following the formulation described in section 2 we get this

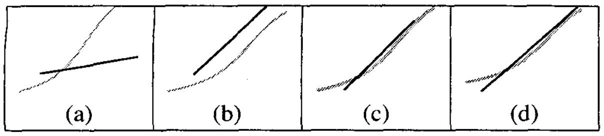

Figure 2. Aproximation of a line to a set of points: (a) Initial configuration. (b) Orientation. (c) Position. (d) Length

energy function to be minimised:

$$E = E_{int} + \gamma \cdot \sum_{i=1}^{n} - \log \sum_{j=1}^{m} P(p_i | l_j) \tag{6}$$

where γ is a factor which measures the relative weighting between internal and external energy. Higher influence of internal energy will derive in more rigid models.

Minimisation of this function is not a straightforward problem, although it can be seen as a problem with incomplete or missing information and then, an implementation of the EM algorithm [4] can be used to solve it, reformulating it as a problem with complete information. Here, the missing information is the correct association of each point in the image with its generating line. Knowing this information we could reformulate likelihood as the distance of each point to its generating line.

Then, in the expectation step of the algorithm, we can estimate association between points and lines from the generating probability of each pixel by every line, normalized to sum up to 1 for each pixel:

$$p_{ij} = \frac{P(p_i | l_j)}{\sum_{k=1}^{m} P(p_i | l_k)} \tag{7}$$

p_{ij} is the probability of correspondence between point p_i and line l_j.

Now, external energy can be defined as the sum of distances between the points in the image and the lines in the symbol weighted by the probability of correspondence p_{ij}:

$$E_{ext} = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m} p_{ij} \cdot d'^2(p_i, l_j) \tag{8}$$

In the maximisation step we must minimise the function resulting from internal and external energy defined in expressions 3 and 8. This function can be splitted in independent terms for each of the lines. Then, minimisation is done separately for each of them by successively finding the orientation, position and length of the line which minimise the energy funnction, as it is shown in figure 2. Each of these steps can be easily solved by deriving analiticallly the contribution of the line to the energy function.

These two steps are applied iteratively until convergence is reached, i.e, when the difference in the parameters of lines in two successive steps are small enough. At each step, the variance used in calculating p_{ij} is decreased based on the mean distance between the pixels and the symbol lines.

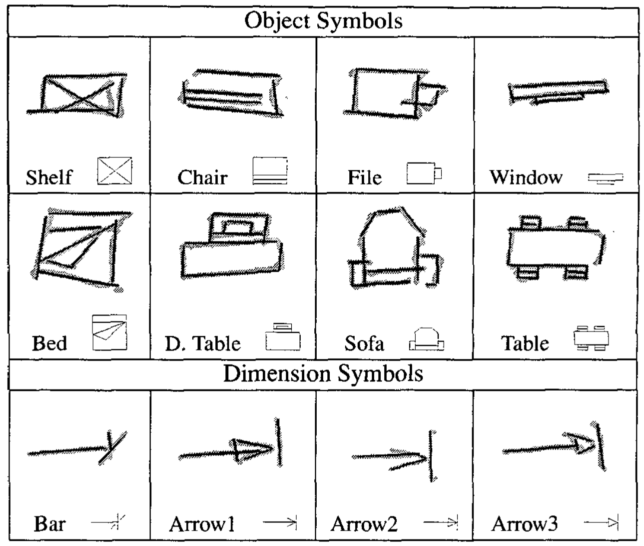

Figure 3. Matching of images (in gray) with their corresponding symbol (in black).

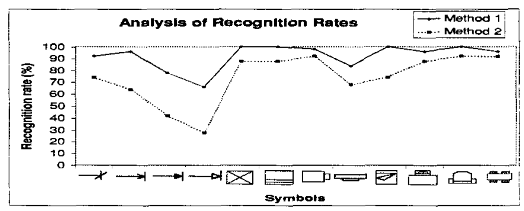

Figure 4. Recognition rate for each symbol.

5 Results and discussion

We have applied this method to the recognition of object and dimension symbols that can be found in architectural hand-drawn drawings. These symbols must be accurately identified to be able to get a complete semantic description of the drawing. We have worked with a sample set consisting of fifty images of each symbol drawn by ten different people without any drawing constraints in order to represent the greatest number of symbol distortions.

Figure 3 shows some representative examples of the matching for images of all the symbols. In it, we can see the different kinds of distortions yielded by hand-drawing. They include disconnected lines, non-straight lines, changes in orientation of lines, distortions at crossing points or corners, etc. The figure illustrates how the ideal model of the symbol (in black) is deformed in order to fit the pixels in the image (in gray).

Identification of an image with a symbol is done, as explained in section 2 classifying each image with the symbol with lowest energy value after matching it with all the sym-

	Method 1	Method 2
Object Symbols	96.75%	85.25%
Dimension Symbols	83%	52%
Dimension Symbols (*)	96%	88 %

Table 1. Recognition rates for object and dimension symbols

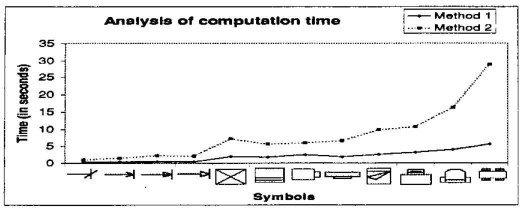

Figure 5. Recognition time for each symbol.

bols. Figure 4 shows the recognition rates for each symbol obtained after applying this criterion to all sample images. Table 1 shows the global recognition rates for all the symbols. It compares recognition rates for the method described on this paper (method 1) with results from our pervious approach (method 2) [7]. We have signifficantly improved them getting a 96.75% of global accuracy for object symbols. For dimension symbols, accuracy is lower due to the similarity among the three arrow symbols as it is shown in figure 3. To overcome this problem, we have done an experiment consisting of taking the three arrow symbols as the same symbol and classifying each image only as a bar or as an arrow. Results are showed in the last line. Accuacy increases to 96%, showing that although the method is confused with very similar symbols, it could be used to discriminate among different kinds of similar symbols.

Finally, in figure 5 we analyze the computation time of the algorithm. It shows the average time of matching all sample images with each symbol. Computational complexity is approximately linear in the number of lines of the symbol and computation time is lower than computation time with our previous approach. This is due to the fact that energy minimisation is easier in this new approach.

6 Conclusions and open issues

We have developed a new approach to deformable template matching for symbol recognition which improves our previous results in terms of recognition rates, computation time and stability of final solution. We have got an 96.75% of overall recognition accuracy over a set of 400 unconstrained images of eight symbols with a linear computational complexity in the number of lines of the symbol. This rate shows the feasibility of deformable template matching and bayesian inference to handle distortion produced in hand-drawn symbols.

Difference with our previous approach comes from the definition of a new similarity measure between an image and a symbol. This new measure allows to use the EM algorithm to minimize the energy function and therefore, stability of final solution is assured and computation time decreases. Working with all pixels in the skeleton of the image we can avoid a previous vectorization step which can introduce noise and errors in the representation of the image.

Open issues to be developed include the extension of this method to other kinds of primitives apart of straight lines, segmentation of the symbols in a whole drawing and validation of scalability of the method with a greater number of symbols and sample images and with more similar symbols, as it is the case of arrow symbols.

References

[1] A.K. Chhabra. Graphic symbol recognition: An overview. In K. Tombre and A.K. Chhabra, editors, *Graphics Recognition: Algorithms and Systems*, pages 68–79. Springer Verlag, Berlin, 1998.

[2] A.K. Jain, Y. Zhong, and M.P. Dubuisson-Jolly. Deformable template models: A review. *Signal Processing*, 71:109–129, 1998.

[3] A.K. Jain and D. Zongker. Representation and recognition of handwritten digits using deformable templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(12):1386–1391, Desember 1997.

[4] G. J. MacLachlan and T. Krishnan. *The EM algorithm and extensions*. John Wiley and Sons, Inc., 1997.

[5] M. Revow, C.K.I. Williams, and G. Hinton. Using generative models for handwritten digit recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(6):592–606, June 1996.

[6] K. Tombre. Graphics recognition - general context and challenges. *Pattern Recognition Letters*, 16(9):883–891, 1995.

[7] E. Valveny and E. Martí. Application of deformable template matching to symbol recognition in handwritten architectural drawings. In *Fifth IAPR International Conference on Document Analysis and Recognition ICDAR'99*, pages 483–486, 1999.

[8] E. Valveny and E. Martí. Deformable template matching within a bayesian framework for hand-written graphic symbol recognition. In *Proceedings of the Third International Workshop on Graphics Recognition*, 1999.