A DECISION STEP FOR SHAPE CONTEXT MATCHING

Mariano Tepper, Daniel Acevedo, Norberto Goussies, Julio Jacobo, Marta Mejail

Departamento de Computación,
Facultad de Ciencias Exactas y Naturales,
Universidad de Buenos Aires

ABSTRACT

This work presents a novel contribution in the field of shape recognition, in general, and in the Shape Context technique, in particular. We propose to address the problem of deciding if two shape context descriptors match or not using an *a contrario* approach. Its key advantage is to provide a measure of the quality of each match, which is a powerful tool for later recognition processes. We tested the proposed combination of Shape Context and the *a contrario* framework in character recognition from license plate images.

Index Terms— shape recognition, shape context, *a contrario* matching.

1. INTRODUCTION

One of the central problems in any automated recognition system is to have a tool to make a sound judgement about the accuracy of its output. However, the vast majority of related works concentrate on other parts of the recognition process [1]. The goal of this work is to treat this problem and we will focus on shape recognition.

Shape Context [2] is a successful method for shape recognition. Many applications proved its utility: in hand-drawn character recognition and trademark retrieval [2], in breaking CAPTCHA's [3] and in object recognition [4]. In all of them, the decision step is approached in an heuristic manner and thus experimentally fixing arbitrary thresholds.

The *a contrario* framework [5] was born to provide an intuitive and general threshold in recognition tasks as a part of the Computational Gestalt program. It has been used for curve recognition by Musé *et al.* (see [6] for a comprehensive account). This approach has the advantage of not using a priori information about the curves (shapes).

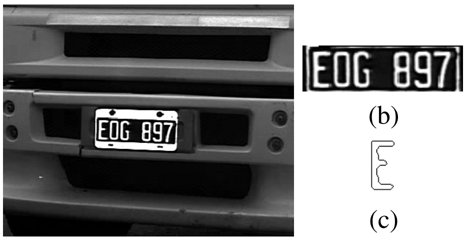
We propose to combine these two methods to decide whether two shape contexts match or not. We claim that this combination provides crucial information for later recognition refinement processes.

We test this method on a database of truck license plates, some of them being seriously deteriorated. The classical licence plate recognition approaches in [1] are unsuccessful in this case.

The organisation of this work is as follows. In Section 2 the shape context descriptor is introduced. Section 3 describes the *a contrario* framework and its application to shape context matching. Finally, an application for character recognition from license plates is provided and analyzed in Section 4, along with concluding remarks.

2. SHAPE CONTEXT

Let $I : D \to V$ be a discrete image, where $D \subseteq \mathbb{Z} \times \mathbb{Z}$ and $V \subseteq \mathbb{C}^n$. We are concerned with images that have a foreground object and background that might have been previously extracted with some detection algorithm. The contour of this object is extracted and we use it for shape recognition. Fig. 1 depicts an example of this process: the license plate identification process begins with a license plate detection stage followed by a contour extraction for the individual characters.



(b)

(c)

(a)

Fig. 1: Shape detection process. (a) Original grayscale image; (b) license plate detected; (c) first digit extracted from (b).

A novel way of describing shapes was introduced by Belongie *et al.* [2]. Let $\mathcal{T} = \{t_1, \ldots, t_n\}$ be the set of points of the contour (Fig. 2(a)). For each $t_i \in \mathcal{T}$, $1 \leq i \leq n$, we model the distribution of the positions of $n-1$ remaining points in $\mathcal{T}$ relative to t_i. We call this distribution, i.e. a 2D log-polar histogram, the Shape Context of t_i (SC_{t_i}).

Formally, be a polar space $[0, 2\pi] \times \mathbb{R}$ and let $\Theta = [0 = \alpha_0, \alpha_1, \ldots, \alpha_A = 2\pi]$ be a partition of $[0, 2\pi]$ and $\Delta = [0 = d_0, d_1, \ldots, d_D]$ be a partition of the range $[0, d_D]$ where d_D

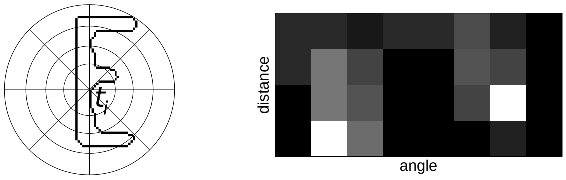
is the distance to the farthest point in $\mathcal{T}$ to t_i. $\Theta \times \Delta$ forms a partition of $[0, 2\pi] \times \mathbb{R}$.

Let us denote by $SC_{t_i}(\alpha_k, d_m)$ the number of points in the bin $[\alpha_{k-1}, \alpha_k) \times [d_{m-1}, d_m) \in \Theta \times \Delta$, $0 < k \leq A, 0 < m \leq D$. This quantity can be defined by

$$\begin{aligned} SC_{t_i}(\alpha_k, d_m) = \#\{t_j \in \mathcal{T}, j \neq i \ : \\ \alpha_{k-1} \leq \arg(t_j - t_i) < \alpha_k, d_{m-1} \leq ||t_j - t_i||_2 < d_m\} \end{aligned}$$

where $\arg(v)$ is the angle of the vector v.

Fig. 2 depicts both spatial and matrix representations of a shape context.



(a) (b)

Fig. 2: Shape context. (a) Partition into bins around the central point t_i; (b) matrix representation of SC_{t_i}.

The collection of the shape context for every point in the shape is a redundant and powerful descriptor for that shape. In order to achieve shape recognition with Shape Contexts it is mandatory to perform a matching and to have a measure of its accuracy. This measure involves the computation of thresholds, a hard task that will be addressed in the next section.

3. THE DECISION STEP

The decision step is the least studied of all the processes involved in visual recognition. Most methods use a nearest neighbor approach to match two sets of descriptors. For example, the method proposed by Lowe [7] to match SIFT descriptors performs an additional heuristic check between the first and the second nearest neighbor, that makes the match more robust .

In [2] a global dissimilarity minimization, via bipartite graph matching, is described. A faster method is depicted in [4], based on a weighted sum of costs of a generalization of the shape context descriptor described in the previous section.

All these efforts aim at reducing the number of false correspondences but are not truly successful: none of the above methods gives a clear-cut answer to the problem of deciding if two descriptors are similar. In our particular case, we need to find out whether two shapes look alike or not.

3.1. FORMAL DEFINITIONS

The *a contrario* framework gives a natural environment to address this problem. It was developed as a part of the Computational Gestalt project to detect events against a random situation. It is used in a wide range of applications, see [5] for a complete description. In [8] a decision method for shape recognition is proposed. The *a contrario* detection framework is based on the Helmholtz Principle that, for our application, states that a match is meaningful when it is not likely to occur in a context where noise overwhelms the information.

The aforementioned framework is specially suited for shape matching. Let $\{SC_i | 1 \leq i \leq n\}$ and $\{SC'_j | 1 \leq j \leq m\}$ be two sets of shape contexts from two different shapes. We want to see if both shapes look alike. The distances between SC_i and SC'_j can be seen as observations of a random variable D that follows some unknown random process.

What we would really want to do is to perform an hypothesis test, for each pair (SC_i, SC'_j), where
$\mathcal{H}_1$*: SC_i and SC'_j is observed because of some causality, i.e. because the shapes look alike.*
$\mathcal{H}_0$*: SC_i and SC'_j is observed only by chance, i.e. because the database is large.*

On one hand, $P(D|\mathcal{H}_0)$ can be modeled with relative ease, even if the model is not perfectly realistic. On the other, it is not possible to model $P(D|\mathcal{H}_1)$ because we assume no other information but the observed set of features. Hence, the full hypothesis test can not be done: we can not control type II errors.

However controlling type I errors, i.e. the number of false correspondences under $\mathcal{H}_0$, is enough to make a sound answer to our decision problem. In other words, low probabilities under $\mathcal{H}_0$ are not likely to happen by chance and are, on the contrary, causal.

The novel contribution in this work is to apply the above detection framework for shape context matching.

We define the distance between two shape contexts and estimate the probability of occurrence of a given match under $\mathcal{H}_0$. Following [8], it is essential to split the shape context into independent features (its importance will be clarified in Section 3.2).

Formally, let $\mathcal{F} = \{F^k | 1 \leq k \leq M\}$ be a database of M shapes. For each shape $F^k \in \mathcal{F}$ we have a set $\mathcal{T}^k = \{t_j^k | 1 \leq j \leq n_k\}$ where n_k is the number of points in the shape. Let $SC_{t_j^k}$ be the shape context of t_j^k, $1 \leq j \leq n_k, 1 \leq k \leq M$. We assume that each shape context is split in C independent features that we denote $SC_{t_j^k}^{(i)}$ with $1 \leq i \leq C$.

Let Q be a query shape and q a point of Q. We define

$$d_j^k = \max_{1 \leq i \leq C} d_j^{k(i)} \quad (1)$$

$$d_j^{k(i)} = d(SC_q^{(i)}, SC_{t_j^k}^{(i)}) \quad (2)$$

where $d(\cdot, \cdot)$ is some appropriately chosen distance.

We can now formally state the *a contrario* hypothesis
$\mathcal{H}_0$*: the distances d_j^k (resp. $d_j^{k(i)}$) are observations of identically distributed independent random variables D (resp. $D^{(i)}$) that follows some stochastic process.*

Then the probability of false alarms is defined as

$$P(D \leq \delta|\mathcal{H}_0) = P(\max_{1\leq i\leq C} D^{(i)} \leq \delta|\mathcal{H}_0) \qquad (3)$$

$$= \prod_{i=1}^{C} P(D^{(i)} \leq \delta|\mathcal{H}_0) \qquad (4)$$

The probabilities $P(D^{(i)} \leq \delta|\mathcal{H}_0)$ can be estimated empirically as the cumulated histogram of the distances $d_j^{k(i)}$, $1 \leq i \leq C$, $1 \leq k \leq M$ and $1 \leq j \leq n_k$.

Definition 1. *The number of false alarms of the pair* (q, t_j^k) *is*

$$\mathrm{NFA}(q, t_j^k) = N \cdot \prod_{i=1}^{C} P(D \leq d_j^k|\mathcal{H}_0) \qquad (5)$$

where $N = \sum_{k=1}^{M} n_k$.

If $\mathrm{NFA}(q, t_j^k) \leq \varepsilon$ then the pair (q, t_j^k) is called ε-meaningful match. This provides a simple rule to decide whether a single pair (q, t_j^k) does match or not.

From one side, this is a clear advantage over other matching methods since we have an individualized assessment for the quality of each possible match.

From the other, the threshold is taken on the probability instead of directly on the distances. Setting a threshold directly on the distances d_j^k is hard, since distances do not have an absolute meaning. If all the shapes in the database look alike, the threshold should be very restrictive. If they differ significantly from each other, a relaxed threshold would suffice.

Thresholding the probabilities $P(D \leq \delta|\mathcal{H}_0)$ is more robust and stable. More stable, since the same threshold is suitable for different database configurations. More robust, since we explicitly control type II errors.

Moreover, the NFA is a bound of the expected number of false alarms:

Proposition 1. *The expected number of* ε*-meaningful matches in a random set E of random matches is smaller than* ε.

Proof. A complete proof can be found in [8]. □

3.2. PARTITIONING THE SHAPE CONTEXT

As stated above, the features in which the shape context is splitted must be independent to go from Eq. 3 to Eq. 4.

A shape is represented by a set of sample points drawn from the contours of an object. In [2] the sampling is made somewhat uniformly along the contour. This responds to the assumption that the points follow a Poisson process [9]. This is a fundamental property to take advantage of when splitting the shape context.

The shape context can be directly splitted by grouping its bins. Since the points are assumed to be uniformly distributed, any way to group the bins (without overlapping), produce a set of independent features. Fig. 3 shows different ways to split the shape context (the 2D polar histogram boundaries are indicated by thick lines). We can see each group indicated with roman numerals: for $C = 4$, in Fig. 3 (a) distances and angles are halved, and in Fig. 3 (b) only angles are divided; in Fig. 3 (c), two ranges of distances and four quadrants amount to a total of $C = 8$ features.

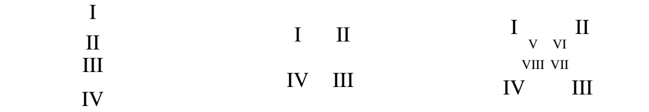
Once we know that the shape context can be splitted, the question is: why is it necessary to split it? The probabilities $P(D^{(i)} \leq \delta|\mathcal{H}_0)$ are estimated in practice using the cumulated histograms of the distances $d_j^{k(i)}$. Each bin can be at least $1/N$. If we take the number of features $C = 1$, from Def. 1 the NFA of any pair of features is greater than $N \cdot 1/N = 1$. This means that on the average we would have 1 false alarm per query, which is not by any means an acceptable bound.

It is therefore important to choose $C > 1$, so that the NFA of any pair of features is greater than $N \cdot 1/N^C = 1/N^{C-1}$. This means that we can reach lower values for the NFA by splitting the shape context into independent features.

4. RESULTS AND CONCLUSIONS

In order to test the machinery presented in Sec. 3, we work over a set of grayscale images of truck license plates captured with an infrared camera, some of them being seriously deteriorated. A license plate detection and character segmentation algorithm is applied to these images (such algorithm is developed to deal with this kind of images but it is out of the scope of this work). The binary images resulting from this process are used as input to our proposed method (see Fig. 4).

We develop an application that decides which character from the database corresponds to the query. For that, we count the number of ε-meaningful matches between the shape query and each shape from the database. The database shape that produces the biggest number of matches is selected. In case that there are no ε-meaningful matches for any database shape, a no-match decision is returned.



(a) $P_2^2 \Rightarrow C = 4$ (b) $P_1^4 \Rightarrow C = 4$ (c) $P_2^4 \Rightarrow C = 8$

Fig. 3: Different ways to split a shape context. Doted lines separate the bins and thick lines separate each bin grouping. The notation P_b^a means that we split uniformly the angles in a groups and the distances in b groups.

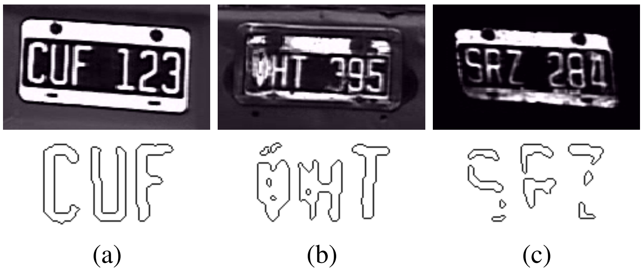


Fig. 4: (a) An example of a non-deteriorated license plate; (b) - (c) two examples of deteriorated license plates. On the second row, their corresponding segmentations.

Following the notation in Sec. 3, our database $\mathcal{F}$ is built of $M = 27$ characters from A to Z. The number of points of each $\mathcal{T}^k$ is approximately $n_k = 220$, for $1 \leq k \leq 27$. In our implementation, for $\Theta = [0 = \alpha_0, \alpha_1, \ldots, \alpha_A = 2\pi]$ we take $A = 18$, and for $\Delta = [0 = d_0, d_1, \ldots, d_D]$ we take $D = 10$. We split the shape contexts into $C = 8$ features, using the P_2^4 grouping because it showed the best behaviour among the three possibilities depicted in Fig. 3.

We emphasize the fact that very deteriorated license plates exist in our test set. In Fig. 4 (a) we show an example of a non-deteriorated license plate and the result of its segmentation; in Figs. 4 (b) and (c) two examples of deteriorated license plates and the result of their segmentation are shown. Due to the different nature of these two classes of images (deteriorated and non-deteriorated), the results are analyzed separately.

We tested our method with 352 character queries: 224 non-deteriorated and 128 deteriorated. We used 3 different values for ε: $1/N, 1/N^2$ and $1/N^3$. Comparative results are summarized in Table 1. Both for deteriorated and non-deteriorated images, the choice $\varepsilon = 1/N^2$ gives the best results, 67.19% and 88.39% respectively.

Without the proposed decision step, absolute thresholds should be fixed which is a painful task. Even if the thresholds could be tuned for our test set, we cannot assure that this approach would work in general.

With this simple application we show that the combination of the shape context method and the *a contrario* framework is a powerful tool. The simplicity of the application allows to evaluate solely and effectively the decision step. The recognition of 9 out of 10 characters is already a very good result as the first process in the recognition chain. Moreover, for severely deteriorated images, the proposed method also behaves relatively well. Indeed, the *a contrario* framework allows to quantitatively assess the quality of the matches between shape contexts.

The discrimination of good-quality and bad-quality matches is important as it allows to increase the robustness of the posterior steps that a complete shape recognition algorithm should incorporate. As an example, thin plate splines are used in [2] reaching high recognition rates. In future work we will add them to our algorithm to test the overall performance of a complete character recognition algorithm.

Classes of images	ε		
	$1/N$	$1/N^2$	$1/N^3$
non-deteriorated (%)	83.93	88.39	72.63
deteriorated (%)	66.41	67.19	39.84

Table 1: Character recognition percentage for each class using different values for ε. We tested 352 character queries: 224 non-deteriorated and 128 deteriorated.

5. REFERENCES

[1] C.-N.E. Anagnostopoulos, I.E. Anagnostopoulos, I.D. Psoroulas, V. Loumos, and E. Kayafas, "License plate recognition from still images and video sequences: A survey," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 9, no. 3, pp. 377–391, Sept. 2008.

[2] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, no. 4, pp. 509–522, April 2002.

[3] G. Mori and J. Malik, "Recognizing objects in adversarial clutter: breaking a visual captcha," *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, vol. 1, pp. I–134–I–141 vol.1, June 2003.

[4] G. Mori, S. Belongie, and J. Malik, "Efficient shape matching using shape contexts," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 27, no. 11, pp. 1832–1837, November 2005.

[5] A. Desolneux, L. Moisan, and J. M. Morel, *From Gestalt Theory to Image Analysis*, vol. 34, Springer, 2008.

[6] F. Cao, J. L. Lisani, J. M. Morel, P. Musé, and F. Sur, *A Theory of Shape Identification*, vol. 1948 of *Lecture Notes in Mathematics*, Springer, 2008.

[7] D. Lowe, "Distinctive image features from scale-invariant keypoints," in *International Journal of Computer Vision*, vol. 60, pp. 91–110. Springer, Nov. 2004.

[8] P. Musé, F. Sur, F. Cao, Y. Gousseau, and J. M. Morel, "An a contrario decision method for shape element recognition," *International Journal of Computer Vision*, vol. 69, no. 3, pp. 295–315, September 2006.

[9] B. D. Ripley, "Modelling spatial patterns," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39, no. 2, pp. 172–212, 1977.