

A NEW AND SIMPLE SHAPE DESCRIPTOR BASED ON A NON-PARAMETRIC MULTISCALE MODEL

Max Mignotte

DIRO, Département d'Informatique et de Recherche Opérationnelle, CP 6128,
Succ. Centre-Ville, P.O. 6128, Montréal (Québec), H3C 3J7.

HTTP : WWW.IRO.UMONTREAL.CA/~MIGNOTTE/

E-MAIL : MIGNOTTE@IRO.UMONTREAL.CA

ABSTRACT

In this paper, we present a new and robust shape descriptor, which can be efficiently used to quickly prune a search for similar shapes in a large image database. The proposed shape descriptor is based on a multiscale representation of the discrete set of points, sampled from the internal and external contour points of the query and the candidate shapes. In this approach, dissimilarity between two shapes is defined as the reconstruction error, of the candidate shape, made by using multiscale elements of contours extracted from the query shape. This dissimilarity measure allows to quickly produce an accurate shortlist of candidate matches, ranked from the most similar to the least similar one, suitable for a more careful and more time consuming matching algorithm. Experiments on the Snodgrass & Vanderwart database allows to attest the discriminating power of this measure and its robustness to possible distortions, warping and occlusion artifacts.

1. INTRODUCTION

At a time of information overload and a potential explosion in digital content, the ability to compare and index digital images in an automated fashion is of growing importance in various application fields. This has great relevance, for example, to researchers, marketing departments, media and design companies, online auction companies, search engines and portals.

Along with color and texture, widely used in the content-based image retrieval community, shape information remains an important, although relatively under-exploited, cue for pattern recognition and retrieval purposes in large image database (see [1] for an extensive survey of shape matching in computer vision). Nevertheless, even when color and texture are absent, as in line drawing, the human visual system can recognize objects quite well and can also make efficiently coarse distinctions very quickly. This demonstrates the highly discriminative and descriptive power that is only conveyed by the information modeled by a single shape contour.

One shape-based approach, commonly used in an automatic image database retrieval system, is the deformable template-based framework [1][2][3]. This approach of shape matching through global deformations is now well known and gives quite good results. A drawback of this technique is its computational cost. Such an approach could not be viable on a database of several thousands objects. While a final stage of deformable matching may be necessary to make very fine distinctions, it would be interesting to do fast early pruning based on much cruder measures. This is the goal of this paper. The contribution of this paper is to propose a shape-

based similarity measure allowing to quickly produce an accurate shortlist of candidate matches, ranked from the most similar to the least similar one, suitable for a more careful (and also more time consuming) (possibly shape-based) matching engine.

In that prospect, Wei *et al.* have recently proposed an interesting multiscale model to synthesize a texture of arbitrary size from a given texture sample [4]. Contrary to classical parametric models, their approach, is simple, intuitive and works surprisingly well even for a wide variety of complex textures. This model allows to efficiently capture the local and global visual characteristics of a complex input texture and thus seems also well suited to compare some local geometric characteristics, at different scales, of two contour shapes.

This paper is organized as follows. In Section 2, we provide a brief overview of the multiscale model introduced by Wei *et al.* for the texture synthesis and we generalize their model, in Section 3, for defining an efficient shape similarity distance for the problem of rapid shape retrieval and shape indexing. We provide experimental results on the Snodgrass & Vanderwart database and we discuss the obtained results in Section 4. Finally, future work and conclusion are given in Section 5.

2. PREVIOUS WORK

Recently Wei *et al.* [4] have proposed an interesting non parametric multiscale model to synthesize an output image from a given input texture sample. Their approach works surprisingly well for a wide range of textures. Their algorithm involves a pyramidal structure of synthesized images and a multiresolution (Gaussian) pyramid built from the input textural sample. From coarse to fine, each level of the pyramid is synthesized as in Efros's algorithm [5];

- The output pyramidal structure is initialized with a random noise whose histogram is equalized with respect to the one of the Gaussian pyramid built from the input sample.

- For each pixel of each output level, in scan-line order, the already-synthesized values in neighborhood of current pixel z of a specific (fixed) size ($N(z)$) is considered and is compared with all possible neighborhoods $N(x)$ from the input sample (of the same pyramidal level). The value of the input pixel x with the most similar $N(x)$ is then assigned to pixel z .

The considered neighborhood uses spatial adjacent sites and sites belonging to the coarser scale. The similarity of the two considered neighborhoods is computed according to the L_2 norm. In order to obtain good synthesis results, the size of the neighborhood must be large enough to capture the scale of the texture elements of

the input sample. Although this searching process is deterministic, the random aspect of each synthesized texture is ensured by the initial random noise (first causing the boundary pixels to be assigned semi-stochastically and then using these first assignments as a seed). This method, combined with a tree-structured vector quantization, to accelerate the search for the nearest neighbor, provides surprisingly good results even on complex textures [4]. The synthesized textures are very similar to the original and seem to come from the same (underlying) generative process as the input textural samples. Wei *et al.*'s synthesis algorithm run also efficiently on input binary textural images (such as line drawings of a contour shape) (cf. Fig. 1).



FIG. 1 – Wei *et al.*'s texture synthesis algorithm result on a binary textural image. The resulting texture (right) is synthesized at twice the size of the original input sample (left) (see Ref. [4] for more examples on input colored textural samples).

Let $z = \{z_s, s \in S\}$ and $x = \{x_s, s \in S\}$ designate, at full resolution, the set of grey values associated to the N pixels of the input and output texture images respectively and located on a lattice S of N sites s . Given the initial textural sample z and the initial random noise of the output pyramidal, an output synthesized image x is a configuration for which the distance $\text{Dist}(x \| z)$, at full resolution, is low with,

$$\text{Dist}(x \| z) = \sum_{s \in S} \left(\min_{p \in S} D(N(x_p), N(z_s)) \right), \quad (1)$$

where $D(\cdot)$ is the L_2 distance and $N(\cdot)$ designates the spatial and scale causal neighborhood on the multiscale structure.

This multiscale model with this spatial and scale causal neighborhood offers the opportunity to capture very efficiently the larger scale characteristics of a given texture sample. This hierarchical model along with its neighborhood structure is also very close to the prior model used in Ref. [6], to capture the larger scale characteristics of sonar image contents, for the ultrasound image segmentation issue. Of course, a shape contour is fundamentally different of a textural sample. Shapes generally contain detailed and non-repeating variations. Nevertheless, the distance $\text{Dist}(x \| z)$ seems also well suited to capture some local geometric characteristics of a shape contour and could be used to infer and define an efficient shape descriptor for the shape indexing problem. We consider this in the next Section.

3. SHAPE SIMILARITY DISTANCE

The query shape describes only one of the possible instances of the considered shape class. In order to take into account the variability of each possible shapes related to the object class to be

detected in the database, we do not exploit directly the discrete set of points, sampled from the internal or external contours on the shape. Instead, we exploit an edge potential field. This field is determined by the positions of the edge points in the contour images. For a pixel (i, j) in the contour image I , we define its edge potential by,

$$\Phi_I(i, j) = \exp\left\{-\frac{1}{\rho} (\delta_i^2 + \delta_j^2)^{\frac{1}{2}}\right\},$$

where (δ_i, δ_j) is the displacement to the nearest edge point in the image and ρ is a smoothing factor which controls the degree of smoothness of the potential field and consequently the intrinsic variability of the object class to be detected¹. We compute this potential field both for the (internal and external) contours of the query shape and for the contours of the candidate shape. The query and candidate images are then decomposed into multiple resolutions by building two Gaussian pyramids in order to get the multiscale structure that will be used in our matching process. Example of this multiscale potential edge field is shown in Fig. 2.

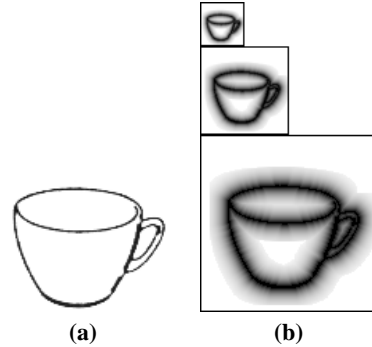


FIG. 2 – (a) Original line drawing stemming from the Snodgrass & Vanderwart database. (b) Multiscale edge potential map (with two levels of pyramid and $\rho = 4$).

Finally, in order to get a symmetric shape similarity measure, the distance $\text{Dist}(q, c)$ between a query (q) and a candidate (c) shape is defined as,

$$\text{Dist}(q, c) = \max\{D(q \| c), D(c \| q)\},$$

$$\text{with, } \text{Dist}(q \| c) = \sum_{s \in C_c} \min_{p \in C_q} D(N(\Phi_{q_p}), N(\Phi_{c_s})), \quad (2)$$

where the first summation is over the contour points $\{c_s\}$ of the candidate shape c . The 2nd summation is over the contour points $\{q_s\}$ of the query shape q which are not already selected. $N(\Phi_{q_p})$ and $N(\Phi_{c_s})$ designates respectively the (multiscale) set of edge potential values estimated at contour point p of the query image and potential values at contour point s of the candidate image. The shape similarity distance $\text{Dist}(q, c)$ can be also interpreted as a reconstruction error measure of the contour of the candidate shape

¹We can easily complete this edge potential field $\Phi_I(i, j)$ by adding to it a directional component in order to obtain a directional edge potential field,

$$\varphi_I(i, j) = \Phi_I(i, j) |\cos \beta(i, j)|,$$

where $\beta(i, j)$ is the angle between the tangent of the nearest edge and the tangent direction of the contour at (i, j) . This potential is similar to the one proposed in [2] for the globally deformable template-based matching approach.

using multiscale contour elements of the query shape (like a reconstruction given by a jigsaw made up of multiscale pieces) (see Fig. 3). In this way, we expect to obtain low distance values for similar shapes and high distance values for distinct ones.

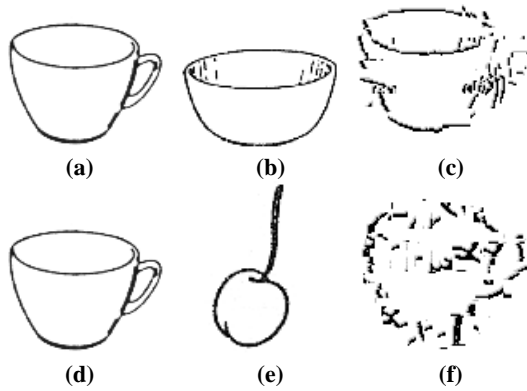


FIG. 3 – Examples of reconstructions between a given query shape and two candidate contour shapes ($\rho = 1$ in this example). (a)(d) Query shape q . (b)(e) Candidate shapes c . (c) Good reconstruction result and consequently low distance value ($\text{Dist}(c \parallel q) = 6.8 \times 10^5$) and good similarity between the two shapes (a) and (b). (f) Less good reconstruction result and consequently higher distance value ($\text{Dist}(c \parallel q) = 9.65 \times 10^5$) and low similarity between the two shapes (d) and (e).

4. EXPERIMENTAL RESULTS

We use one level of pyramid and a 17×17 spatial and symmetric neighborhood in the fine level with a symmetric 17×17 spatial neighborhood in the coarse level. We set $\rho = 4$. Finally, we use an approximate-nearest-neighbor search technique [7] to accelerate the search of the minimal of the L_2 norm $D(\cdot)$ in Eq. 2. In order to reduce the computational cost, the first and second summation of Eq. 2 is over a subset of (regularly spaced) contour points on the query and the candidate shapes. A similarity measure takes about 75ms to perform a single matching between two shapes on a PC workstation (1Ghz). Due to the local characteristic of this measure, an implementation on parallel machines remains possible and could greatly reduce the computational time.

Our shape similarity measure has been tested on the Snodgrass & Vanderwart line drawing database. This dataset, frequently used in the psychophysics community for tests with human subjects, contains line drawings of 260 commonly occurring and distinct objects (only one image per object) [8]. In this test, no pre-processing phase is required for normalizing the shape size or for extracting the edges since the images are only line drawings of fixed size. In the case of raw images, a canny edge detector [9] would allow to extract line features from the images and a pre-treatment would be required for normalizing the size of the shapes. In our application, we use only an Hotelling transform [10] in order to align the binary shapes with their principal axes. In order to also test the ability of our pruning method to deal with possible deformations and occlusions (i.e., contours partially detected) of the object class to be detected, we create, for each shape, a synthetic distorted and/or warped and/or occluded contour version which is added to the initial dataset.

Figs 4 and 5 show some query shapes and a shortlist of candidate matches (i.e., the four most similar shape ranked from 1 to 4)

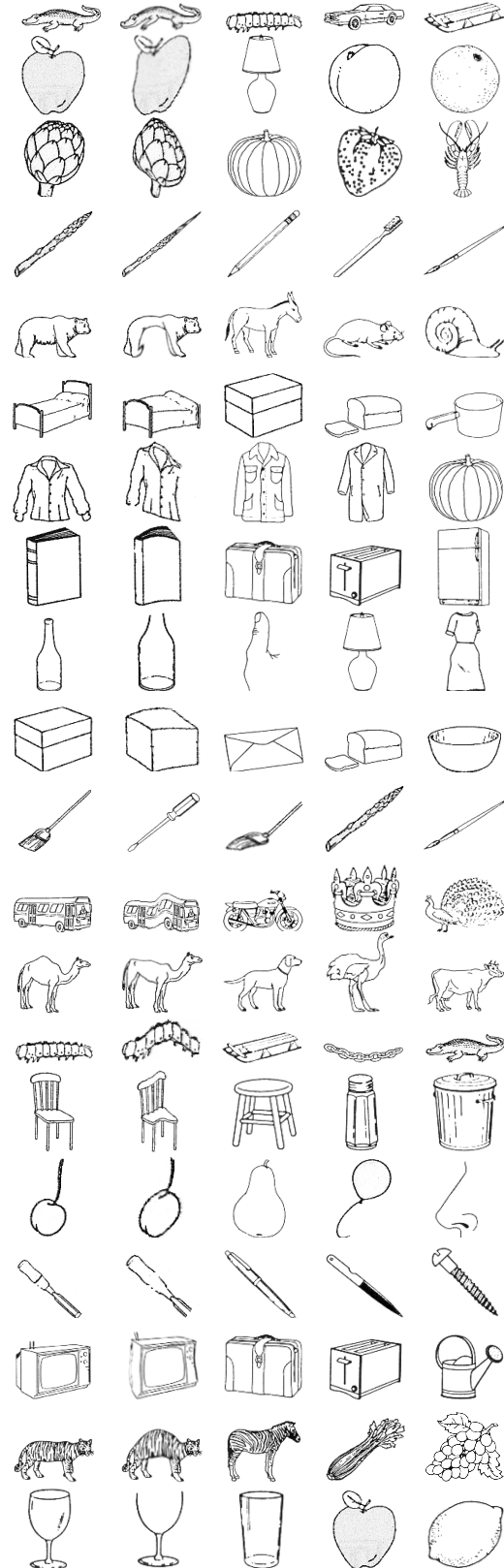


FIG. 4 – Retrieval results. The first column is the query shape. The remaining columns show the closest 4 matches to each query object.

proposed by our pruning method. Our shape similarity measure is quite insensitive to the deformations and the distorted and/or occluded version of each shape is correctly selected in the four closest matches. Moreover, our pruning method allows to find quite correctly visually similar shapes. Results of the ranking are quite consistent with the human visual system. Our method can be efficiently combined in a second stage with a more powerful and more time consuming comparison technique such as a deformable template-based matching algorithm, to only the shortlist, in order to refine the search process. Finally, let us recall that there is additional and interesting appearance information available that would not be captured by our retrieval system. e.g., textural and color information. A complete image database retrieval system should include these informations.

5. CONCLUSION

In this paper, we have presented a new and efficient shape descriptor for a quick and coarse shape matching. To this end, we have exploited the multiscale non parametric model introduced by Wei *et al.* for texture synthesis to define an intuitive similarity measure between two contour shapes. This technique has shown itself to be well suited to efficiently compare *some* local geometric characteristics at different scale of two contour shapes. This similarity measure allows to quickly produce an accurate shortlist of candidate matches, ranked from the most similar to the least similar one, suitable for a more powerful and more time consuming indexing tool such as a deformable template-based matching search engine. Experiments on contour shape database allows to attest the discriminating power of this shape descriptor and its robustness to distortions and occlusions artifacts.

6. REFERENCES

[1] R.C. Veltkamp and M. Hagedoorn. State of the art in shape matching. Technical Report UU-CS-1999-27, Utrecht, 1999.

[2] A.K. Jain, Y. Zhong, and S. Lakshmanan. Object matching using deformable templates. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 18(3) :267–278, 1996.

[3] M. Mignotte, C. Collet, P. Pérez, and P. Boutheymy. Hybrid genetic optimization and statistical model-based approach for the classification of shadow shapes in sonar imagery. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22(2) :129–141, February 2000.

[4] L.-Y. Wei and M. Levoy. Fast texture synthesis using tree-structured vector quantization. In *Proceedings of SIGGRAPH'2000*, pages 479–488, July 2000.

[5] A. A. Efros and T. K. Leung. Texture synthesis by non-parametric sampling. In *7th International Conference on Computer Vision, (ICCV'99)*, pages 1033–1038, 1999.

[6] M. Mignotte, C. Collet, P. Pérez, and P. Boutheymy. Sonar image segmentation using a hierarchical MRF model. *IEEE Trans. on Image Processing*, 9(7) :1216–1231, 2000.

[7] S. Arya, D. M. Mount, N.S. Netanyahu, R. Silverman, and A.Y. Wu. An optimal algorithm for approximate nearest neighbor searching in fixed dimensions. *Journal of the ACM*, 45(6) :891–923, 1998.

[8] J. G. Snodgrass and M. Verderwart. A standardized set of 260 pictures : Norms for name agreement, familiarity and visual complexity. *Journal of Experimental Psychology : Human Learning and Memory*, 6 :174–215, 1980. <http://www.cog.brown.edu/tarr/stimuli.html>.

[9] J. Canny. A computational approach edge detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 8(6), 1986.

[10] R.C. Gonzales and R.W. Woods. *Digital Image Processing*. Addison-Wesley Publishing Company, 1992.

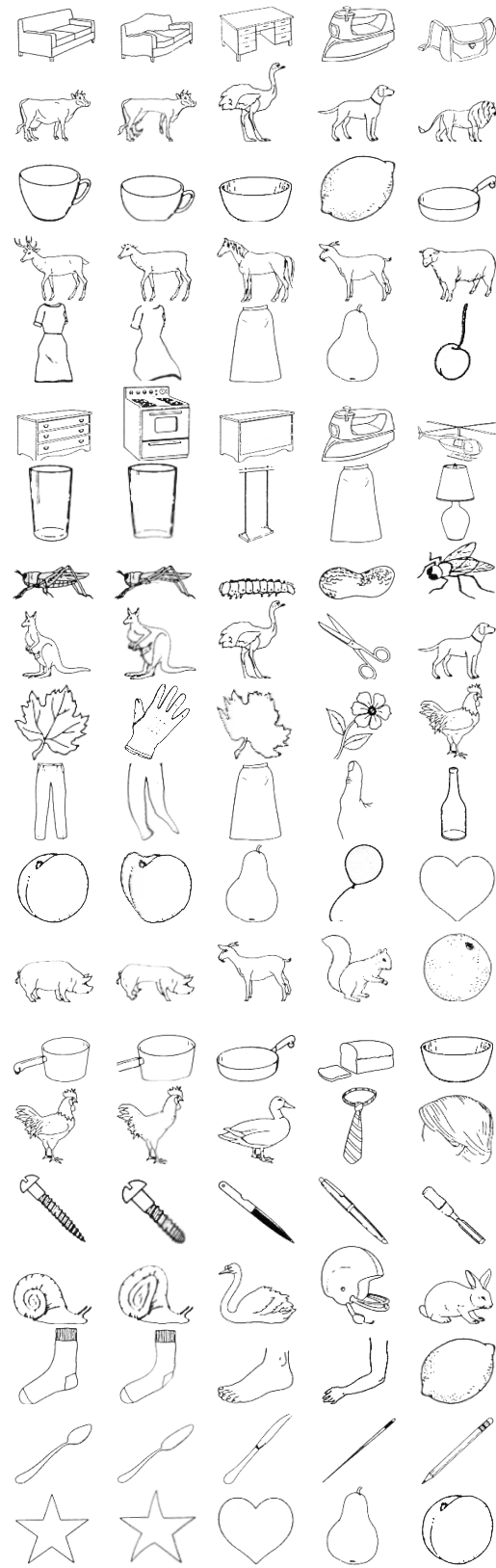


FIG. 5 – Retrieval results. The first column is the query shape. The remaining columns show the closest 4 matches to each query object.