International Conference on Pattern Recognition 2002, Québec, August 11-15, pp. 689-692 (2002)

Variable Neighborhood Search for Geometrically Deformable Templates

Marc Lalonde Langis Gagnon

Vision and Imaging Team, R&D Dept., CRIM
550 Sherbrooke St. West, #100
Montreal, Qc, Canada, H3A 1B9
{mlalonde,lgagnon}@crim.ca

Abstract

This paper proposes two modifications to the geometrically deformable template model. First, the optimization stage originally based on simulated annealing is replaced with a meta-heuristic called Variable Neighborhood Search that treats simulated annealing as a local search tool. Second, an affine deformation energy is introduced to improve the quality of the search. An example of optic disc segmentation in an ophthalmic image is given.

1. Introduction

Deformable templates, or parametric deformable models [5], have been a subject of intensive research over the past few years. Unlike their free-form counterparts (e.g. snakes), they usually rely on probabilistic shape models that 'embody' the knowledge about the typical shape of objects to be located in the image. They are especially useful in the field of medical image analysis where objects are e.g. anatomical structures and knowledge about their shape is readily available. A significant difficulty with most models is the high complexity of the search procedure, which is usually based either on local optimization, with the risk of getting trapped in a local extremum, or on global optimization with the well-known issue of computational complexity.

An interesting template model recently proposed by Rueckert [7] is called a "geometrically deformable template" and it aims to refine a coarse match obtained from multiresolution analysis. Its deformation model is based on the use of thin-plate splines [1] and has the appealing property of being invariant under affine transformations. However, on one hand the proposed search method is computationally costly because it relies on simulated annealing (SA), and on the other hand such invariance to affine transformations may cause the search algorithm to retain invalid solutions (e.g. ellipses when searching for circles). This paper addresses these concerns by exploring the use of a meta-heuristic called Variable Neighborhood Search [4] and also by redefining the shape energy so that affine transformations are taken into account. The paper is organized as follows. Section 2 introduces the basic template model. In Section 3, suggested modifications are presented. Section 4 reports on some results obtained with an ophthalmic image and finally, Section 5 discusses some advantages, limitations and potential enhancements of the new method.

2. Basic Model

Within the basic framework of geometrically deformable templates, and using the notation of [1], the contour of an object M is modeled as a set of n vertices $P_i = (x_i, y_i)$, $i = 1..n$, which may correspond to particular landmarks of the object. These vertices form the equilibrium shape (or undeformed prototype shape) of the model when no external forces are applied to it. External forces caused by the presence of edges in the image induce a deformation of the shape, with vertices being moved to new locations (x'_i, y'_i) collected in the matrix $\mathcal{V}$:

$$\mathcal{V} = (v_1, \ldots, v_n) = \begin{bmatrix} x'_1 & x'_2 & \ldots & x'_n \\ y'_1 & y'_2 & \ldots & y'_n \end{bmatrix}$$

Let us define r_{ij} as the distance between two vertices, i.e. $r_{ij} = |P_i - P_j|$, and the following matrices:

$$K = \begin{bmatrix} 0 & U(r_{12}) & \ldots & U(r_{1n}) \\ U(r_{21}) & 0 & \ldots & U(r_{2n}) \\ \ldots & \ldots & \ldots & \ldots \\ U(r_{n1}) & U(r_{n2}) & \ldots & 0 \end{bmatrix}$$

$$P = \begin{bmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ \ldots & \ldots & \ldots \\ 1 & x_n & y_n \end{bmatrix}, \quad L = \begin{bmatrix} K & P \\ P^T & O \end{bmatrix}$$

where O is a 3x3 matrix of zeros and the function $U(r) = -r^2 \log r^2$ is the fundamental basis function of the thin-plate spline model (see [1] for a discussion on this function).

It is then possible to define a pair of mapping functions $f_x(x, y)$ and $f_y(x, y)$, also called thin-plate spline mapping functions, having the form:

$$f(x, y) = a_1 + a_x x + a_y y + \sum_{i=1}^{n} w_i U(|P_i - (x, y)|)$$

The affine coefficients a_1, a_x and a_y as well as the weights w_i of the non-affine part of the mappings are obtained by the relation $L^{-1}Y = (W \mid a_1\, a_x\, a_y)^T$ where $W = (w_1, ..., w_n)$ and $Y = (\mathcal{V} \mid 0\,0\,0)^T$. Y is the $(2 \times (n+3))$ augmented vector of the new positions of the vertices. Since the bending energy of the thin plate at (x, y) is proportional to $\int\int_R \left[(\frac{\partial^2 f}{dx^2})^2 + 2(\frac{\partial^2 f}{dxdy})^2 + (\frac{\partial^2 f}{dy^2})^2\right]$, it has the property of being invariant to affine transformations. For example, an equilibrium shape bent toward either a circle or an ellipse will have the same bending energy. Searching for an instance of a given shape model in an image thus implies moving the landmarks to locations that globally minimize this deformation energy as well as some external energy that is function of the image gradient.

3. Proposed modifications

Two aspects of the basic model of [7] are revisited in this paper. First, since the affine invariance of the internal deformation energy affects the quality of the search in the sense that visually poor solutions may be achieved despite their low bending energy, we propose to add another energy term, as described in Section 3.1. Second, due to the huge size of the search space, Rueckert proposes to constrain the search for new landmark positions along the normal at each location (x_i, y_i) and resorts to stochastic relaxation (SA) to find the optimal configuration of points. Here, we propose a more elaborate search procedure, guided by a meta-heuristic called Variable Neighborhood Search (VNS), in which SA-based search is viewed as a local search in a restricted, yet variable, neighborhood centered on a given configuration of points.

3.1. Energy function

The energy of a configuration of points is a weighted contribution of its intrinsic (shape) energy as well as the gradient energy in the image: $E_T = \alpha_1 E_{shape} + \alpha_2 E_{image}$. For the basic model, the term E_{shape} captures the amount of bending energy and is proportional to the trace of the product WKW^T [7], while E_{image} is simply made proportional to the negative of the sum of the gradient magnitude at each location v_i of the image $\mathcal{I}$. As mentioned above, an affine deformation energy term has been added to penalize configurations that would be unacceptably far from the shape model despite a low non-affine deformation energy. The affine parameters linking the landmarks P_i of the reference object model M to a configuration $\mathcal{V}$ are given by the singular value decomposition of the matrix containing the affine terms a_i of the thin-plate spline mapping functions f_x and f_y (translations $a_1^{(x)}$, $a_1^{(y)}$ are ignored; the subscripts (x) and (y) refer to the functions f_x and f_y):

$$[U, S, V] = \text{svd}\left(\begin{bmatrix} a_x^{(x)} & a_y^{(x)} \\ a_x^{(y)} & a_y^{(y)} \end{bmatrix}\right)$$

As pointed out by Bookstein [1], the singular values of the SVD represent scale factors between S and $\mathcal{V}$ in the x and y directions, and rotation matrices U and V describe the relationship between the two configurations in terms of rotations (reflections are ignored). If one expects the configuration $\mathcal{V}$ to be quite similar to M, then a simple penalty term such as $E_{affine} = -[g(S_x)g(S_y) + \Theta(\theta_U, \theta_V)]$ can be constructed, where $g(\cdot)$ and $\Theta(\cdot)$ are some weighting (e.g. Gaussian-shaped) functions while S_x, S_y are the scale factors and θ_U, θ_V are the angles associated to the rotation matrices U and V. The term E_{affine} should be minimal for $S_x = S_y = 1$ and $\theta_U = \theta_V$. This affine term represents some *a priori* information about the shape model and to that regard, the parameters of the weighting functions might be learned from examples, e.g. as in [8],[2]. The total energy function now becomes:

$$E_T = \underbrace{\alpha_1 \text{tr}(WKW^T)}_{E_{shape}} - \underbrace{\alpha_2 \sum_i |\nabla \mathcal{I}(v_i)|}_{E_{image}} - \underbrace{\alpha_3 \left[g(S_x)g(S_y) + \Theta(\theta_U, \theta_V)\right]}_{E_{affine}}$$

3.2. Search

VNS is one of the few meta-heuristics that have been proposed as a means to combine efficient local optimization procedures (e.g. gradient descent) with heuristics having the ability to cope with local optima. VNS works by exploring increasingly distant neighborhoods around a starting point x, possibly restarting the search in case a locally optimal solution happens to improve the best solution found so far. The algorithm drawn from [4] is given below.

In the current context, s is a coarse solution obtained for example from multiresolution analysis (see [5], [7] for instance), s' is some configuration of points (x', y') randomly picked (see below for more detail) and s'' is the best solution found in the image by the local search constrained within a given neighborhood. In case local search fails to improve the solution, the neighborhood is enlarged. One simple definition may be purely spatial, i.e. $\mathcal{N}_k$ would include candidate points in the image that are located no farther than

Initialization. Select the set of neighborhood structures $\mathcal{N}_k$, $k = 1, ..., k_{max}$, that will be used in the search; find an initial solution s; choose a stopping condition;
Repeat the following until the stopping condition is met:
(1) Set $k \leftarrow 1$; (2) Until $k = k_{max}$, repeat the following steps:
(a) *Shaking*. Generate a configuration s' at random from the k^{th} neighborhood of s ($s' \in \mathcal{N}_k(s)$);
(b) *Local search*. Apply some local search method with s' as initial solution; denote with s'' the so obtained local optimum;
(c) *Move or not*. If this local optimum is better than the incumbent, move there ($s \leftarrow s''$), and continue the search with $\mathcal{N}_1(k \leftarrow 1)$; otherwise, set $k \leftarrow k+1$;

$k \cdot d$ pixels from any point (x, y), where k is the neighborhood number and d is some constant setting the width of the neighborhood (a few pixels typically).

- *Shaking : tacking points*. The shaking step of the VNS algorithm consists in choosing a random configuration of points s' within the $\mathcal{N}_k$ neighborhood. In short, the suggested strategy is to pick three points from s and tack them to some edges that intersect their normal (if more than one intersection is found, a random selection is made). Moving these points to new locations induces a deformation of the shape model s; the new equilibrium shape s' can be computed using the parameters of the affine transformation that resulted from the movement of the three tacked points. The model s' is the starting point for the local search.
- *Local search : simulated annealing*. The local search tries to move the free points (i.e. those that are not tacked) on neighboring edges so as to minimize the total energy E_T. The free points are moved along their normal. A discrete implementation of simulated annealing is used as the optimization method, but the search space is kept small (a few pixels) in order to reduce the computational burden. The matrix P used in the computation of E_{shape} and E_{affine} is constant and contains the landmarks of the reference model M.
- *Move or not*. If the local search finds a configuration of points with lower energy than the current best solution, the new configuration is the starting point for another pass of the VNS algorithm ($k = 1, ..., k_{max}$), otherwise, new shaking takes place within an increased neighborhood (i.e. some points of the current solution are tacked farther as k increases).

During shaking, the tacking step must obey two simple rules, otherwise a neighborhood change is triggered. First, since an affine transformation is applied to the free points after tacking, the selection of the triplet of points to be tacked must be done carefully. The stability of various combinations is assessed, the most stable triplets are retained and a random selection is made. Stability is defined as the ratio between the area of the triangle defined by the triplet with the area of the convex hull of the configuration of points [3]. Second, the new equilibrium shape s' should be relatively similar to the shape being sought, in order to avoid an irrecoverable drift of the search algorithm. A loose bound on the affine deformation energy of the equilibrium shape is used to constrain the search.

4. Experiments

In order to test the search procedure, two experiments have been carried out in the context of optic disc segmentation in ophthalmic images. Deformable templates are well adapted to this problem since a priori knowledge about the shape of the optic disc is available. (The current work is in fact a logical follow-up to previous work on rigid template matching in the same application domain [6]). In the first experiment, an initial solution for the optic disc location is shown in Fig. 1, left (total energy=17.63, essentially due to the E_{image} term). This initial solution was given manually for testing purposes but may be provided by automatic means (e.g. [6]) in a realistic application. During the search, seven solutions have been found in various neighborhoods ($k = 1, ..., 5$; $d = 6$ pixels for a 250x200 image). The best result is shown in Fig. 1, middle (total energy=12.27). Note that since the reference model is a circle, E_{affine} is simply a function of the scale factors S_x and S_y with $g(\cdot)$ defined as a 'tent' (triangle) function centered on 1.

The second experiment is designed to show how the search procedure can recover from being trapped in a local minimum and still find the global solution. An image has been altered manually by inserting an opaque circle inside the optic disc (Fig. 1, right). The deformable template initially captures most of the contour of the opaque circle (total energy > 20) but keeps looking for a better solution by enlarging the search region (i.e. increasing k). It finally captures the true contour of the optic disc (total energy = 12.44). The correct solution is found despite the false target's high similarity with the reference model and its strong edges in the edge map. This is a case where the contribution of the affine deformation energy makes a difference between a good and a bad solution.

As far as computation speed is concerned, the algorithm was implemented in Matlab, and the execution of the compiled version on a dual PIII 930 MHz running Linux yields processing times around 3-4 minutes. It is believed that a careful implementation in C along with the optimizations suggested in [7] will lower the processing time substantially.

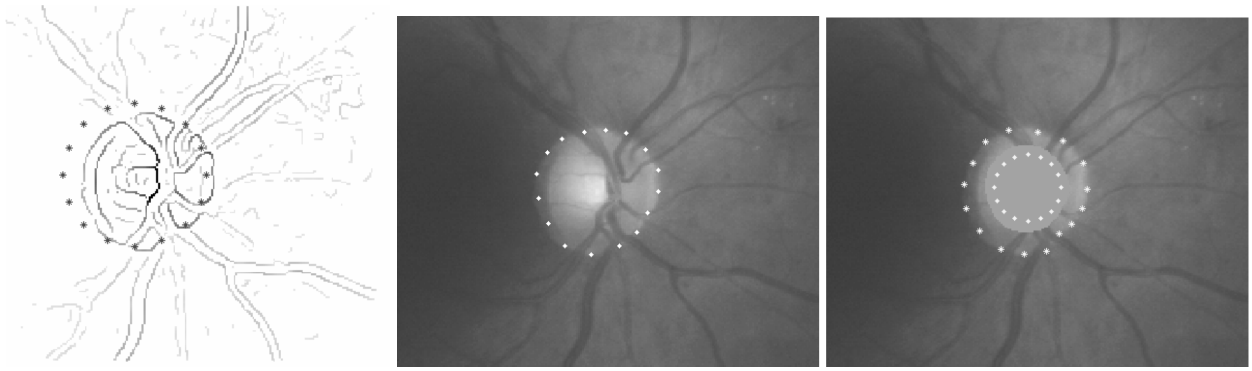

Figure 1. Left: initial configuration overlaid on the edge map (obtained using the Canny operator). Middle: final solution. Right: Segmented optic disc despite a false target inside the disc (the initial position of the deformable template is inside the false target).

5. Discussion and Conclusion

The experiments we have performed up to now on synthetic as well as real ophthalmic images indicate that the proposed method is capable of recovering the correct solution despite an unfavorable initialization (due to inappropriate translations and/or scale factors) and a complex edge map (broken edges due to vessels branching out of the optic disc), provided that the neighborhoods are large enough. As for robustness, one area of concern might be the discrete nature of the movements of landmarks. (Landmarks are moved to locations where their normals intersect edges). What happens when the object contours are broken/missing? The reason behind the choice of moving landmarks onto edges is simply to minimize the computational cost of the local search (simulated annealing). Indeed, some of these points may not move during the local search but the process of tacking points (selected randomly) and of moving the others with an affine transformation that restores the equilibrium shape insures that the set of points move close to a potential solution. This process of 'shaking by tacking' can be made more sophisticated. For instance, instead of tacking points onto edges, one may imagine tacking salient higher-level features (e.g. shape corners) onto similar edge groupings in the image. A post-processing step might also be added: points of the best solution that do not lie on edges would be revisited and possibly moved in a way that minimizes the deformation energy (affine and non-affine).

Finally, one might argue that such a landmark-based approach that reduces a complex object to a set of points coarsely describing its contour has intrinsic limitations in terms of reliability. For many applications an interesting direction to pursue might then be a migration toward an appearance-based model that would include energy terms based on texture or color information in addition to shape.

Acknowledgements

This work was partially supported by the Canadian Foundation for Innovation and by the National Sciences and Engineering Research Council of Canada (NSERC).

References

[1] F. Bookstein. Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Trans. Patt. Anal. Mach. Intell.*, 11(6):567–585, 1989.

[2] D. Cooper, T. Cootes, C. Taylor, and J. Graham. Active shape models - their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, 1995.

[3] M. Costa, R. Haralick, and L. Shapiro. Optimal affine-invariant point matching. In *Proc. 10th Int. Conf. on Pattern Recognition*, pages 233–236, 1990. Atlantic City.

[4] P. Hansen and N. Mladenovic. An introduction to variable neighborhood search. In S. Voss, S. Martello, I. Osman, and C. Roucairol, editors, *Meta-heuristics: advances and trends in local search paradigms for optimization*, pages 433–458. Kluwer Academic Publishers, 1999.

[5] A. Jain, Y. Zhong, and S. Lakshmanan. Object matching using deformable templates. *IEEE Trans. Patt. Anal. Mach. Intell.*, 18(3):267–278, 1996.

[6] M. Lalonde, M. Beaulieu, and L. Gagnon. Fast and robust optic disk detection using pyramidal decomposition and Hausdorff-based template matching. *IEEE Trans. Med. Imaging*, 20(11):1193–1200, Nov. 2001.

[7] D. Rueckert. *Segmentation and Tracking in cardiovascular MR Images using Geometrically Deformable Models and Templates*. PhD thesis, Imperial College of Science, Technology and Medicine, University of London, 1997.

[8] Y. Wang and L. Staib. Boundary finding with correspondence using statistical shape models. In *Proc. Conf. Computer Vision and Pattern Recognition*, pages 338–345, 1998. Santa Barbara, California.