Boundary-based Correspondence Computation Using the Topology Constraint

T. Rachidi* and L. Spacek
Department of Computer Science,
University of Essex,
Colchester C04 3SQ
U.K.

Abstract

This paper presents a novel approach to matching boundaries in images. Carefully chosen attributes of boundaries are used to build a *parameter space*. Potential matches are searched for in the parameter space, rather than in the topological space. The 'goodness' of each potential match is measured by means of an affinity function. Final matches are obtained using Simulated Annealing, in which the topology constraint is integrated to drive a global optimisation.
Our method delivers pairings which are topologically correct and does not require any of the commonly used constraints, such as the *maximum velocity, constant disparity gradient* or the *epipolarity constraint.* This characteristic makes the method applicable for motion and tracking, as well as for stereopsis. The current implementation is fully presented together with the results obtained, which are most satisfactory.

Keywords: *Boundaries, Parameter space, Correspondence computation, Topology constraint, Simulated annealing.*

1 Introduction

The choice of appropriate features is crucial to motion perception in general, and to the correspondence computation in particular. In order to solve the correspondence problem, one must first decide on the elements to be matched in the scene. There is a trade-off between the complexity of the monocular analysis, used to extract the elements to be matched, and the complexity of the matching process.

Low level features, like edgels, convey very little information, and often yield ambiguous output. Therefore, these features are difficult to match correctly between successive frames. Very high level features, such as surface patches or objects, are easy to match, but very difficult to detect.

Boundaries, however, are more structurally evolved features than edgels. They are geometrically unambiguous, involve implicit grouping and continuity information, and are relatively easy to extract from images. Matching boundaries between successive frames can yield quality information about the scene being viewed.

*This research was supported in part by an ORS award.

BMVC 1994 doi:10.5244/C.8.5

In this work, we present a computational approach to the correspondence problem which uses connected boundaries. We assume that boundaries do not undergo drastic variations (in shape, contrast, and length), and have very similar perspective projections in successive image frames. This does not imply a serious restriction on the domain of applicability of the approach. The similar shape of image projections of the same solid object curve is quite plausible, given an appropriate sampling rate.

Additionally, we assume that topology is conserved *almost everywhere* between two consecutive frames. Onset and end of occlusion between two consecutive frames are the cases where the topology assumption is violated. These events are rare, and can be turned into an advantage: locations where topology is not conserved between frames can be analysed to detect beginning and end of occlusion.

Attributes of boundaries are used to build a *parameter space*. Potential matches are searched for in the parameter space, rather than the topological space. The goodness of each potential match is measured by means of an affinity function. Final matches are obtained using Simulated Annealing, in which the topology constraint is integrated to drive a global optimisation. A fundamental difference between our approach and existing approaches is the use of the *topology constraint* to drive global matches. Final matches are expected to be topologically correct and yield sub-optimalities free correspondence.

2 Related work

Many researchers attempted the matching of line segments [4], or piecewise linear representations of boundaries [14, 13, 7, 17]. To limit the combinatorial search for pairings, additional constraints were needed, some of which restrict the domain of applicability to stereo correspondence. *Maximum velocity, constant disparity gradient* and *epipolarity* are among these constraints.

Deriche and Faugeras [3] used points of high curvature to guide the matching of curves. The epipolar constraint was used to limit the search area for potential matches of high curvature points. Gradient direction and curvature were used to select plausible matches. The corresponding curves of each element of the obtained pairs of high curvature points were associated. Ambiguities were resolved by taking as final matches the pair of curves with maximum matched high curvature points. The epipolar constraint limits the applicability of this technique to stereo matching. Moreover, this method relied on the existence of high-curvature points in the images. Hence, smooth boundaries were not used.

In their approach, Brint and Brady [1] used *elastic strings/snakes* to represent curves, and measured the amount of deformation the strings have undergone between corresponding curves in images. The amount of deformation was measured using a cross energy term composed of two parts: a disparity term, and a similarity term. The similarity term encourages corners to match against corners. The final stage consisted of a search to maximise the sum of the scores from all possible pairings. The combinatorial nature of the approach (trying all possible pairings), and the two optimisations needed for snake representation, made it a very expensive solution.

Wolfson [21] used *characteristic strings* (a string of real numbers representing pointwise curvatures) to match two given curves. Elements of the characteristic strings of each curve were used to increment a shift accumulator. Shifts in a band around local maxima were used to find the longest matching subcurve. As admitted by the author, this was not a robust step. Beside the combinatorial nature of the overall process, topology and contrast were not used.

Wilson and Hancock's junction matching method used topological information [20]. Although there is an element of duality between their approach and ours, in their approach, only boundaries terminating at junctions were used. Additionally, the topology constraint used does not account for imperfections at junctions.

In general, existing algorithms (with the exception of [20]) rely on sliding curves along each other to find common portions and/or positional constraints, such as the epipolar constraint or the maximum velocity constraint. This feature makes them unsuitable for motion analysis. The spatial connectivity between curves/segments, and topology, are not used.

3 Our approach

Boundary fragments are first extracted from Spacek's *first difference magnitude surface* [18], which combines both edge detection and boundary thinning.

The second stage consists of bridging gaps in boundaries, repairing boundary junctions, and constructing the *boundary adjacency graph.* Evidence from existing fragments and perceptual groupings are used by a local non-iterative algorithm to select the best joins and/or junctions that satisfy specific structural conditions. This algorithm is described in detail in [16]. The adjacency graph is updated accordingly to reflect the topological structure of the resulting boundaries. Loose connections at junctions, as in **Figure** 1, are also included in the graph.

The third stage, which is the main subject matter of this paper, consists of the correspondence matching.

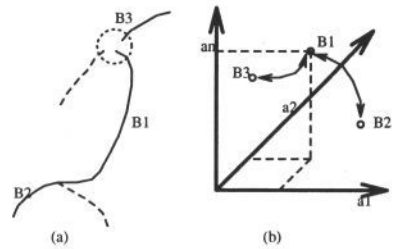

Figure 1. *(a) a set of boundaries* B_1, B_2, B_3. *(b) points* B_1, B_2, B_3 *respectively denote boundaries* B_1, B_2, B_3 *in the* n-D *space* $(a_1, a_2, \ldots, a_n)$. *Arcs represent the topology constraint.*

3.1 Parameter space

The *parameter space* was introduced by McCafferty to extend the iconic representation of visual data from $(x, y, grey)$ to $(x, y, grey, a_1, a_2, a_3, \ldots, a_n)$ [12]. In

contrast to McCafferty's representation, where the tokens are accessible only by their coordinates (x, y), we use a representation which allows access to boundaries either by their attributes or by their neighbourhood. **Figure** 1 shows the combined structure for representing neighbouring boundaries $\{B_1, B_2, B_3\}$.

The boundaries' attributes $(a_1, a_2, a_3, \ldots, a_n)$ are used to construct an n-dimensional *(n-D)* space. Boundaries with the same set of attributes are linked in a list anchored in the same location in the *n-D* space. To capture topological properties, the 'connected-to' relationships are included as pointers to the structures within *n-D* space. As a consequence of this representation, two neighbourhoods can be defined for each token:

- *Topological Neighbourhood:* is the set of first neighbours in the boundary fragments' adjacency graph.
- *Attributive Neighbourhood:* is the hypervolume defined by the elementary displacements $(da_1, da_2, \ldots, da_n)$ around a given token $(a_1, a_2, \ldots, a_n)$.

We use length and contrast to build a *2-D* space PS_p for every image p. Other attributes, such as moments [11] and energy, can also be used.

3.2 Potential pairings

For a given token $t \in p$, the search space $PM(t)$ for candidate matches is given by the area $[L(t) - \alpha L(t), L(t) + \alpha L(t)] \times [C(t) - \beta C(t), C(t) + \beta C(t)] \subset PS_{Other(p)}$, where α, β are the ratios of changes accepted for the length $L(t)$, and the contrast $C(t)$, for the token t.

3.3 Affinity function

The affinity function is a means to evaluate the *goodness* of a match. Matches with a low *goodness* value are most likely to be discarded. The affinity function $Aff(t, t')$ compares the characteristics of two tokens, one from each image, and returns their degree of similarity. Let t be a boundary in image p, and t' a potential match in q, $t' \in PM(t)$:

$$Aff(t,t') = \left(\frac{|L(t) - L(t')|}{MmL}\right)^{w_1} \left(\frac{|C(t) - C(t')|}{MmC}\right)^{w_2} S(t,t')^{w_3}\, T(t,t')^{w_4}$$

where $L(t)$ is the length of the boundary t, and $C(t)$ its contrast, $MmL = |max_e(L(e)) - min_e(L(e))|$ is a normalising factor for the length attribute, $MmC = |max_e(C(e)) - min_e(C(e))|$ is a normalising factor for the contrast attribute, $T(t,t')$ is the topology factor (see **Section** 3.3.2) and $S(t,t')$ is the dissimilarity in shape between t and t' (see **Section** 3.3.1). Weights (w_1, w_2, w_3, w_4) in the form of exponents are used to adjust the relative strengths with which each attribute affects the final result. Note that $Aff(t,t') \in [0..1]$, and $\sum_i w_i = 1$. $w_i < w_j$ implies that more weight is given to the attribute corresponding to w_i.

3.3.1 Shape

The human visual system is excellent at comparing shapes in general, and curves in particular, independently of their size, contrast and length. It also seems that the more regular the curves are, the quicker is the comparison.

Many techniques for shape comparison exist in the literature. Their performances depend on the representation adopted for boundaries. Moment based techniques [19] work in real time, but are very sensitive to noise. Polynomial representations are not translation and rotation invariant. The naive comparison $(\sum_{i=0}^{n} |a_i - b_i|^2)^{\frac{1}{2}}$, where $(a)_i$ and $(b)_i$ are the polynomial coefficients of the two boundaries, does not work. Fourier descriptors based similarity measures [15] give good results, but are expensive to compute. Generally, one has to translate, rotate and expand one of the curves until they are superimposed, and then compute the area between them. The smaller the area is, the more similar are the curves. However, it is not possible to achieve the superimposition at this stage, given that expansion, rotation and translation are unknown yet.

We have chosen the chain codes [6] for representing boundaries in order to obtain an approximate invariance under rotation, translation, and expansion. The comparison of the boundary shapes is based on the Mahalanobis distance between the frequencies of various types of boundary corners, represented by first differences of adjacent chain code digits. Barring quantisation problems, this is fairly stable.

The dissimilarity measure $S(B, B')$ between curves B and B' is defined to be:

$$S(B, B') = \frac{1}{N} \sum_{i=0}^{N-1} \left(\frac{c_i(B)}{L(B)} - \frac{c_i(B')}{L(B')} \right)^2$$

where $L(B)$ is the length of curve B, $c_i(B)$ is the number of corners of type i on B, and N is the number of distinct corner types (N depends on grid connectivity).

If $S(B, B')$ is below a certain threshold Ψ, the two curves are judged similar, and $S(B, B')$ is used in the computation of the affinity. However, $S(B, B')$ is ignored for pairs for which $S(B, B') > \Psi$. This shape dissimilarity measure is $O(n)$, where n is the number of points of the two boundaries.

3.3.2 Topology

The use of topology consists of taking into account the spatial interconnections between curves in the affinity function. A potential match (t, t') is *n-consistent*, if there are at most n t's spatial neighbours that have as potential matches neighbours of t'. Let (t, t') be an *n-consistent* potential match, we define the topology factor $T(t, t')$ in the following manner:

$$T(t, t') = (n + 1)/(g + 1)$$

where g is the number of t's neighbours. Thus, if t has n neighbours, the contribution of the topology factor to the match (t, t') is high if the match is *n-consistent* (*i.e.*, all neighbours of t have neighbours of t' as potential matches), and low if the match is *0-consistent*.

3.4 Global considerations

At this stage, the problem becomes as follows: for every boundary fragment in one image, select its correct match from the set of its potential matches.

One way of using the affinities would be to select those matches with the highest affinity values as the correct ones. Selecting matches in this way amounts to only

looking at local information (*i.e.*, local affinity information), since the information provided by the comparison of the attributes comes only from two tokens. The selection of a single wrong match makes two tokens unavailable for the matches they really belong to. A wrong match may cause further incorrect matches in a knock-on effect. Thus, a global approach is necessary to avoid such problems.

Simulated Annealing [10] was retained as a plausible strategy to our approach. Unlike the genetic approach [8], it is not necessary to work with more than one state at a time. Additionally, finding a suitable crossover operator for combining aspects of each state is problematic. On the other hand, simulated annealing was also preferred to a stochastic relaxation strategy [5] on the basis of the results reported by Jones [9]. Moreover, simulated annealing has been successfully used in various machine vision problems such as: perceptual grouping [12], image processing [2] and region-based correspondence computation [9].

Two entities need to be defined when applying simulated annealing to a problem:

- *The search space:* It is the set of all possible one-to-one pairings of boundaries (see **Figure** 2). $(t_1, t_2) \in f_1 \times f_2$ is chosen as a potential pair only if $t_1 \in PM(t_2)$ and $t_2 \in PM(t_1)$, where $PM(b)$ is the set of the possible matches of b.
- *The energy function:* The energy function is simply the sum of the energies of all current pairings. The energy of a pair (t_1, t_2) is defined as $Aff(t_1, t_2) * Aff(t_2, t_1)$. The energy function is to be maximised rather than minimised (assuming an affinity of 1 represents a perfect match, and 0 represents no match).

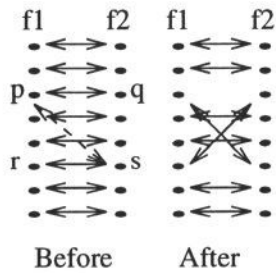

Figure 2. *Ensuring a 1:1 match.*

Ensuring a 1:1 mapping of boundaries requires the following scheme when creating new situations: Let (p, s) be the desired new pair (see **Figure** 2). If $s = q$, no change is done. If, however, $q \neq s$, subtract the energy due to the pair (p, q) from the energy of the current situation, and undo the pair (p, q). Let, if it exists, r be the node in frame f_1 that has s as match in the current situation. Undo the pair (r, s), and subtract its energy from the energy of the current situation. Create the pair (p, s), and add its energy to the energy of the current situation. If possible (*i.e.*, $r \in PM(q)$ and $q \in PM(r)$), pair r and q, and add the energy of the pair to the energy of the current situation.

3.5 Constrained simulated annealing

One way of implementing the topology constraint in simulated annealing is by selecting situations from the pool of consistent pairings. The stochasticity of such

process is questionable. Moreover, computationally speaking, generating the pool of consistent matches is as expensive as graph matching techniques.

Instead, we use strong matches to guide the matching of their neighbouring boundaries at every iteration. A match (p, q) is said to be *strong* iff $PM(p) = \{q\}$ and $PM(q) = \{p\}$. When selecting a new match for a node p from the set of its potential matches $PM(p)$, we select the one with the highest number of strong neighbours, even if this selection may result in lowering the energy. Such a match is marked *strong*, and may guide the matching of other nodes throughout the iterations. If there are no strong neighbours, the selection is done at random, as in classical simulated annealing.

4 The results

The parameters that govern the matching process are: min_l and min_c for the minimum length and contrast of boundaries used; α and β for the ratios of changes allowed on length and contrast; w_1, w_2, w_3, w_4 the respective weights associated with length, contrast, similarity, and local topology attributes in computing affinities; Ψ, the threshold used to declare curves similar. The values for these parameters were obtained empirically throughout a series of runs over diverse images. Typically, $min_l = 10$, $min_c = 5$, $\Psi = 0.007$, $\alpha = 0.15$, $\beta = 0.15$, and $(w_1, w_2, w_3, w_4) = (0.3, 0.4, 0.2, 0.1)$. However, these values can be easily changed to adapt the system to specific situations.

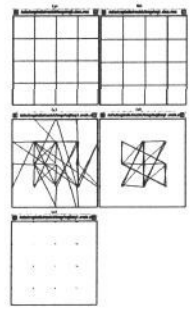

Figure 3. *(a) & (b) Expanding chessboard 64x64 to 68x68. (c) Pairings obtained with $w_4 = 0$, and with non-constrained simulated annealing. (d) Pairings obtained with $w_4 = 0.09$, and with non-constrained simulated annealing. (e) Pairings obtained with constrained simulated annealing.*

The above synthetic sequence was generated to test the topology constraint. The average number of potential matches per boundary in this experiment is eight. This is due to the fact that all boundaries found in the *chessboard* are very similar in shape, length and contrast. Pairings of **Figure** 3(c) were obtained without the topology component when computing local affinities. Only the four boundaries in the corners were paired correctly, since each of them has only one potential match which is the true match. When using the topology factor in computing local

affinities, a greater number of correct pairings were obtained (see **Figure** 3(d)). This was expected because the topology factor introduced in the computation of affinities is of a local nature. **Figure** 3(e) shows the matches obtained when applying constrained simulated annealing. Only four iterations were necessary.

Figures 4–6 and **Figures** 7–9 show experiments with real pairs of images. The maximum number of potential matches per boundary is two (in each case). The final (correct) pairings were obtained respectively in four and three iterations. The process is conservative, delivering correct matches for the paired boundaries. The degree of its success on a particular image is thus best critically judged by the number and nature of the remaining unmatched boundaries.

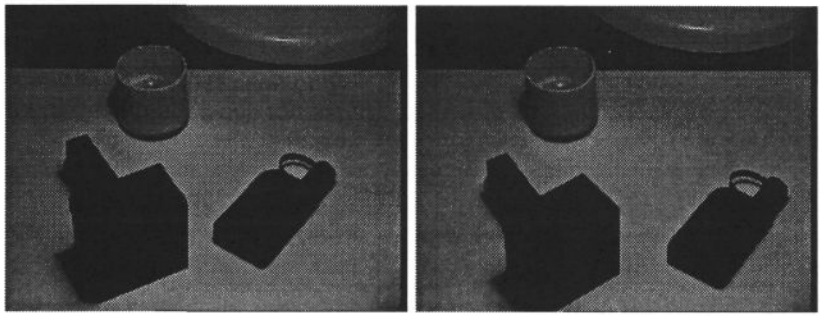

Figure 4. *Image of several objects.*

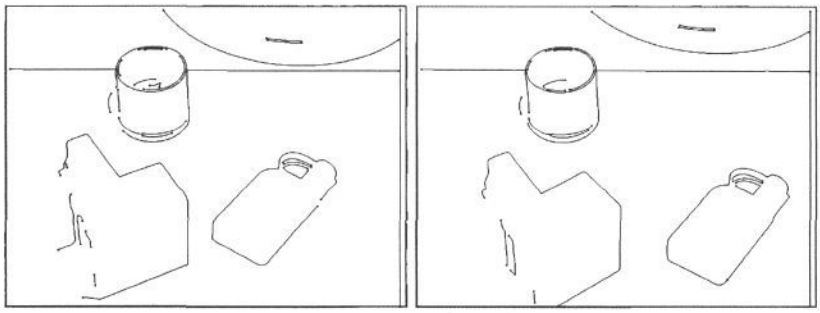

Figure 5. *Input boundaries.*

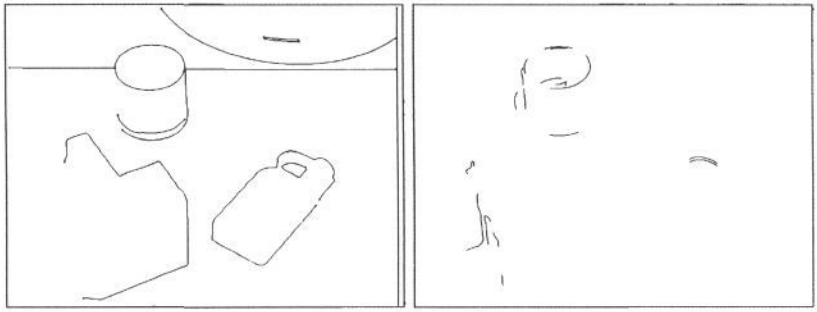

Figure 6. *Matched and unmatched boundaries.*

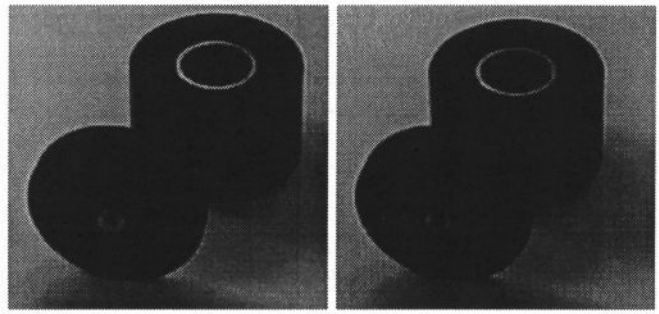

Figure 7. *Image of two objects. The image on the right was obtained after moving the object behind.*

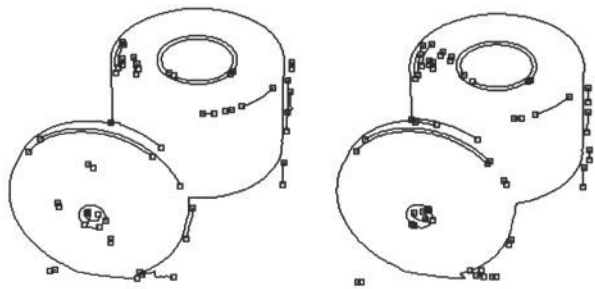

Figure 8. *Squares indicate end points and junctions.*

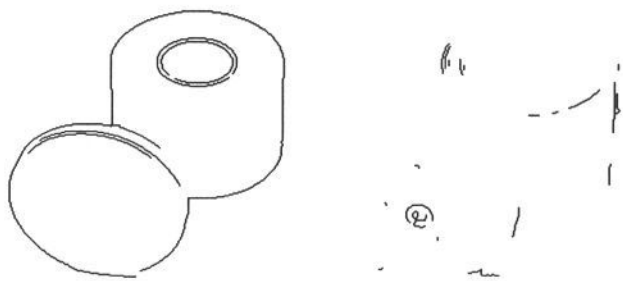

Figure 9. *Matched and unmatched boundaries.*

5 Conclusion

We presented an approach to boundary-based correspondence computation which uses attributes of boundaries and integrates the topological structure of boundaries into the computation.

Many further useful results can be derived from the obtained pairings. Unmatched boundaries together with points where the topology assumption is broken have been identified and can be used for *ghost* boundaries detection and for

occlusion analysis. This leads, in the first instance, to much cleaner boundary representation of the image. Analysing junctions, their associated boundaries, and their matches, leads to the separation of objects in cluttered scenes (the solution to the bin-picking problem). Boundary pairings can also be used to recover reliably the parameters of motion.

References

[1] Andrew T. Brint and Michel Brady. Stereo matching of curves. In *5th Alvey Vision Conference*, pages 187–192, September 1989.

[2] P. Carnevali, L. Coletti, and S. Patarnello. Image processing by simulated annealing. In *Readings in Computer Vision*, pages 551–561. Morgan Kaufmann Publishers Inc., 1987.

[3] R. Deriche and O. Faugeras. 2-d matching using high curvature points: Application to stereo vision. In *INRIA*, May 1988.

[4] R. Deriche and O. Faugeras. Tracking line segments. *INRIA*, 8(4):261–270, November 1990.

[5] G. Fekete, O. Eklundh, and A. Rosenfeld. Relaxation: Evaluation and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 3(4):549–469, 1981.

[6] H. Freeman. Boundary encoding and processing. In B. S. Lipjin and A. Rosenfield, editors, *Picture Processing and Psychopictorics*, pages 241–266, New York, 1970. Academic Press.

[7] S. L. Gazit and G. Medioni. Contour correspondences in dynamic imagery. In *Proceeding of the Image Understanding Workshop*, volume 1, pages 423–432, 1988.

[8] J. H. Holland. *Adaptation of Natural and Artificial Systems.* University of Michigan Press, Ann Arbor, 1975.

[9] A. G. Jones. Region-based correspondence computation. Master's thesis, University of Essex, UK, 1990.

[10] S. Kirkpatrick, C. D. Gelatt-Jr, and M. P. Vecchi. Optimisation by simulated annealing. *Science*, 220(4598):671–680, 1983.

[11] S. Maitra. Moment invariants. In *Proceedings of the IEEE*, volume 67, pages 697–699, 1979.

[12] J. D. McCafferty. *Human and Machine Vision: Computing Perceptual Organisation.* Ellis Horwood Ltd, 1990.

[13] G. Medioni and R. Nevatia. Segment based stereo matching. *Computer Vision, Graphics, and Image Processing*, 31:2–18, 1985.

[14] Gerard Medioni. Maching images using linear features. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(6):675–685, November 1984.

[15] E. Persoon and K. S. Fu. Shape discrimination using fourier descriptors. In *IEEE Transaction on Systems, Man and, Cybernetics*, volume SMC-7, pages 170–179, 1977.

[16] T. Rachidi and L. Spacek. Constructing coherent boundaries. In *Proceedings of the 5th British Machine Vision Conference*, York, September 1994.

[17] Gerard Medioni Salit Levy Gazit. Multi-scale countour matching in a motion sequence. In *The DARPA Image Understanding Workshop*, pages 934–943, 1989.

[18] L. Spacek. Thinning image boundaries. CSM 187, Dept of Computer Science, University of Essex, Colchester C04 3SQ UK., 1993.

[19] C. H. Teh and R. T. Chin. On digital approximation of moment invariants. *Computer Vision, Graphics, and Image Processing*, 33:318–326, 1986.

[20] R. Wilson and E. R. Hancock. A topological constraint corruption process for hierarchical graph-matching. *Czech Pattern Recognition Workshop*, pages 18–25, November 1993.

[21] Haim Wolfson. On curve matching. In *IEEE Computer Society Workshop On Computer Vision*, pages 307–310, 1987.