DOMAIN INDEPENDENT OBJECT DESCRIPTION AND DECOMPOSITION

Tod S. Levitt

Advanced Information & Decision Systems
201 San Antonio Circle, Suite 286
Mountain View, CA 94040

ABSTRACT

A significant problem in image understanding (IU) is to represent objects as models stored in a machine environment for IU systems to use in model driven pattern matching for object recognition. This paper presents a technique for autonomous machine description of objects presented as spatial data, i.e., data presented as point sets in Euclidean n-space. This general definition of objects as spatial data encompasses the cases of explicit listings of points, lines or other spatial features, objects defined by light pen in a CAD system, generalized cone representations, polygonal boundary representations, quad-trees, etc. The description technique decomposes an object into component sub-parts which are meaningful to a human being. It is based upon a measure of symmetry of point sets. Most spatial data has no global symmetry. In order to arrive at a reasonable description of a point set, we attempt to decompose the data into the fewest subsets each of which is as symmetric as possible. The technique is based upon statistics which capture the opposing goals of fewest pieces and most symmetry. An algorithm is proposed which operates sequentially in polynomial time to reach an optimal (but not necessarily unique) decomposition. The semantic content of the descriptions which the technique produces agrees with results of experiments on qualitative human perception of spatial data. In particular, the technique provides a step toward a quantitative measure of the old perceptual Gestalt school of psychology's concept of "goodness of figure".

1. INTRODUCTION

A significant problem in image understanding (IU) is to represent objects as models stored in a machine environment for IU systems to use in model driven pattern matching for object recognition. This paper presents a technique for autonomous machine description of objects presented as spatial data, i.e., data presented as point sets in Euclidean n-space, E^n. This general definition of objects as spatial data encompasses the cases of explicit listings of points, lines or other spatial features, objects defined by light pen in a CAD system, generalized cone representations (Brooks, 1981), polygonal boundary representations, quad-trees (Samet and Webber, 1983), etc.

The description technique decomposes an object into component sub-parts which are meaningful to a human being. It is based upon a measure of symmetry of point sets in E^n. Symmetry is quantified as the reciprocal of the number of reflections and rotations under which the point set is invariant relative to some fixed point (the center of reflection or rotation). For instance, the regular k-sided polygon in E^2 remains invariant under k rotations and k reflections relative to its center for a total of 2k invariant transformations. Thus, the symmetry measure of the regular k polygon in E^2 is $\frac{1}{2k}$. Note that the identity transformation (the null rotation) leaves any point set fixed so that the symmetry value of any point set is always defined and bounded above by one. A circle, relative to its center, has infinitely many invariant transformations, resulting in a minimum possible symmetry measure of zero.

Most spatial data has no global symmetry. In order to arrive at a reasonable description of a point set, we attempt to decompose the data into the fewest subsets each of which is as symmetric as possible. The technique is based upon statistics which capture the opposing goals of fewest pieces and most symmetry. An algorithm is proposed which operates sequentially in polynomial time to reach an optimal (but not necessarily unique) decomposition.

Standard applications of the technique are in E^2 for 2D images and E^3 for 3D object modeling. The technique allows a machine to provide a more meaningful description of spatial data than a simple list of points. This description can be represented as a model and may subsequently be used to match the model to observed instances of the object in imagery. The technique could potentially relieve humans of the need to manually indicate sub-parts of objects which we want the machine to model. Its effectiveness depends upon the semantic reality of the decomposition, and the relationship of that decomposition to autonomous image segmentation techniques. That is, the machine should produce descriptions of objects which indicate much the same "parts" decomposition which a human would provide, and, if matching objects in imagery is desired, the spatial data output from segmentation must lead to a decomposition which is capable of being matched to the stored model.

The semantic content of the descriptions which the technique produces agrees with the results of experiments on qualitative human perception of spatial data. In particular, the technique provides a step toward a quantitative measure of the old perceptual Gestalt school of psychology's concept of "goodness of figure" (Allport, 1955). In a machine environment, we can combine this technique for quantifying the qualitative aspects of perception with the physiologically based structuralist approach to image segmentation by utilizing points of greatest change (i.e. most information) extracted during segmentation as the input to the description and matching process. This differs from the wholly structuralist approach to IU, as typified by Marr and his associates (Marr, 1979), which uses the principle of minimum energy to obtain object descriptions from high information point sets. The (mathematical) relationship between the symmetry-based and information-based approaches is an open question, although Hoffman's work (Hoffman, 1966), provides an excellent first step.

The second section of this paper presents the definition of quantization of symmetry of a single point set along with some mathematical preliminaries. In Section 3, the notion of symmetry is extended to multiple point sets, and we show conditions for spatial data imbedded in E^2 under which decompositions into regular-polygonal subsets are optimal. We use the optimality criterion of regular-polygonal decompositions to derive an algorithm in Section 4 for constructing decompositions in finite points sets in E^2.

2. SYMMETRY OF SINGLE POINT SETS

Many techniques exist to quantify various aspects of regularity in spatial data including texture measures (Wechsler, 1980), analysis in the frequency domain (Crowley, 1984) and information theoretic approaches (Green and Courtis, 1966). Symmetry is one of the most striking forms of regularity in human perception, yet it is not well quantified by any of the methods mentioned above. Most previous attempts (Birkoff, 1932, Eysenck, 1968), to quantify symmetry focused on brute-force delineation of point, side and angle relationships in polygons. Weyl (Weyl, 1952) provided a mathematical description of symmetry of a point set, but even this was not a general enough framework for machine description (nor was that the intention of the work), and his research did not attempt to define symmetry in multiple point sets.

Symmetry may be defined in terms of three types of linear transformations in E^n: reflections, rotations and translations. A point set in E^n is symmetric with respect to a linear transformation if it remains invariant under that transformation. That is, if T maps E^n to E^n, and S is a subset of E^n, $S \subseteq E^n$, then we say S is symmetric with respect to T if T(S) = S. For brevity we will use "wrt" to mean "with respect to".

Rotations leave only the point in E^n which is the center of rotation fixed, and reflections have only the line or (hyper) plane of reflection fixed. Note that the line or plane of reflection uniquely determines the corresponding transformation.

Non-null translations have no fixed points in E^n, and it can be shown that no finite point set in E^n is symmetric wrt a translation. (Some spaces other than E^n, such as a torus obtained by identifying opposite sides of a rectangle as in CRT screen wrap-around, admit finite symmetric subsets wrt translation, but we do not address those issues here.) Finally, by allowing centers of rotation to be translated about, we may also subsume any need for separate translation transformation in defining symmetry. For these reasons we do not explicitly consider translations further in the quantification of symmetry.

We denote the set of reflections and rotations of E^n relative to some fixed point, c, by O(n,c). When c is the origin this is more commonly denoted by O(n) for the orthogonal group of (rigid) linear transformations in E^n. Stretching and contraction are not permitted in this subset of the larger set of all linear transformations which map E^n onto E^n. Note that O(n,c) is just another copy of O(n) with the origin translated to c, so arguments regarding O(n) apply to O(n,c).

Possession of the properties of closure, identity, inverse, and associativity defines O(n) to be a group in the mathematical sense. The order of a group is the number of distinct elements it contains. If G is a group, then we write o(G) for the order of G. (See (Weyl, 1939), or other standard texts for detail on O(n) and groups.) The structure of the group O(n) and its subgroups is well-understood. If a connection between the description of spatial data and O(n) is made, then the structure of O(n) may provide additional insight into the description. This is precisely our program in the following.

Let S be a point set in E^n, and $c \in E^n$. Let $sym(S,c) = \{T \in O(n,c) | T(S) = S\}$, then sym(S,c) is the set of orthogonal transformations under which S is invariant relative to the point c. It is not hard to show that sym(S,c) is a subgroup of O(n,c). (See (Weyl, 1939, 1952).)

We define the symmetry measure of S wrt a point $c \in E^n$, m(S,c), to be the reciprocal of the order of the group of invariant orthogonal transformations of S. That is $m(S,c) = [o(sym(S,c))]^{-1}$. The purpose of the reciprocal is to obtain a bounded measure. Since we always have $I \in sym(S,c)$, we know $o(sym(S,c)) \geq 1$. If o(sym(S,c)) is not finite then we define m(S,c) = 0. It follows that for any S and c, $0 \leq m(S,c) \leq 1$.

To see the usefulness of quantifying symmetry within a well-known mathematical object, consider the case where S is a square in E^2 and c is its center. There are four reflections which leave S invariant, one each across the vertical and horizontal bisectors of the sides, and one across each diagonal. There are also four 90^o rotations (including the identity) which leave S invariant. Thus, o(sym(S,c)) = 8, and $m(S,c) = \frac{1}{8}$.

For the case of a square, sym(S) is a mathematical object well studied in the nineteenth century, called the dihedral group and denoted D(8). As in all groups, the subgroups of D(8) form a partially ordered hierarchy. For D(8) this hierarchy is pictured in Figure 2-1. The point is that subgroups correspond to different ways to decompose S into parts. The vertical and horizontal reflections together imply quartering the square, while each alone implies halving it into rectangles. The diagonal reflections similarly give rise to spatial partitions of the square into two or four triangles. Thus, we can map the structure of D(8) back to S to extract the structure inherent in the square.

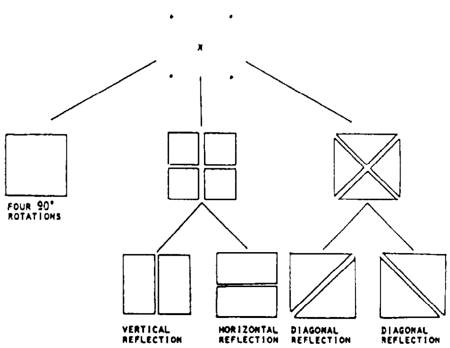

Figure 2-1: Hierarchical Decomposition Example

In fact, if we consider, rather than the whole square, just the corner points (which have high information content) and then include all midpoints between pairs of corner points in order to encode reflectively information, we obtain the dot diagram in Figure 2-1. The decompositions which humans tend to make of this diagram (Zusne, 1970) correspond nicely to those predicted by the subgroup structure of D(8). However, regardless of the nature of human perception, this technique provides an approach to more semantically based machine perception of objects.

3. DECOMPOSITION OF SPATIAL DATA

Most point sets in E^n have no global symmetry. The technique outlined in Section 2 is therefore not sufficient by itself to allow a machine to derive a meaningful description of complex spatial data. Our objective is to decompose the spatial data into subsets which are inherently more identifiable than the object represented by the total data. For instance, the side view of a car might be roughly described as a smaller rectangle, (the roof and windows), over a larger rectangle (the car body), over two circles (the wheels). The technique presented here allows a machine to generate this description for itself.

Following guidelines suggested by research in qualitative visual perception, we seek a decomposition of a set $S \subset E^n$ into as few subsets as possible, each of which is as symmetric as possible. (Other criterion such as relative size and clustering of subsets could also be considered, but are beyond the scope of this paper.)

This approach suggests that we search for a decomposition of S into k subsets each with associated point c_i: $\{(S_i, c_i) \mid i=1 \text{ to } k\}$ which minimizes the evaluation function:

$$E(\{(S_i,c_i) \mid i=1 \text{ to } k\}) = \frac{k}{\sum_{i=1}^{k} [m(S_i,c_i)]^{-1}} = \frac{k}{\sum_{i=1}^{k} o(sym(S_i,c_i))}$$

Note that since $[m(S_i, c_i)]^{-1} = o(sym(S_i, c_i)) \geq 1$,

we have $\sum_{i=1}^{k} [m(S_i,c_i)]^{-1} \geq k$ so that,

$$E(\{(S_i,c_i) \mid i=1 \text{ to } k\}) = \frac{k}{\sum_{i=1}^{k} m(S_i,c_i)^{-1}} \leq 1.$$

Since k and $m(S_i,c_i)$ are always positive, we have $0 \leq E(\{(S_i,c_i) \mid i=1 \text{ to } k\}) \leq 1$. Thus, E is a bounded evaluation function. E is not the only function we might choose. E is in fact a mean statistic satisfying the properties of means as defined, for instance, in (Mays, 1983). Any mean of the set $\{m(S_i,c_i) \mid i=1 \text{ to } k\}$, such as the standard arithmetic mean, would serve as well, although results will differ. The mean E, as defined above, however, is particularly tractable because of its linearity in the $o(sym(S_i,c_i))$. We take advantage of this in the following proposition. Recall that the regular n polygon in E^2 is the polygon which has n equal sides.

Proposition: Let $S \subset E^2$ be a finite set and let $D = \{(S_i,c_i) \mid i=1 \text{ to } k\}$ be a decomposition of S with none of the S_i being a regular polygon. If $P = \{(P_i,d_i) \mid i=1 \text{ to } m\}$ is another decomposition of S where all the P_i are regular polygons, then

$$E(P) < E(D)$$

Furthermore, if $P = \{(P_i,d_i) \mid i=1 \text{ to } m\}$ and $Q = \{(Q_i, e_i) \mid i=1 \text{ to } n\}$ are two different regular polygonal decompositions, then the one with fewer subsets will have the smaller evaluation function. The proof of this proposition is omitted due to lack of space.

This proposition shows that any regular polygonal decomposition of a set is better than any decomposition which has no regular polygons, and that the regular polygonal decomposition with fewest polygons is superior to other regular polygonal decompositions. Figure 3-1 shows the value of the symmetry measure, E, for several possible decompositions of the point set S pictured in Figure 3-1a. Notice that E is minimized by the regular polygonal decomposition into two squares.

However, the case of decompositions with some, but not all, regular polygons is subtler. The measure captures a trade-off between number of objects and relative symmetry of objects. It favors more objects with higher individual symmetry, if their symmetries are close to the same value, but favors few objects if there must be a great disparity in their relative symmetry.

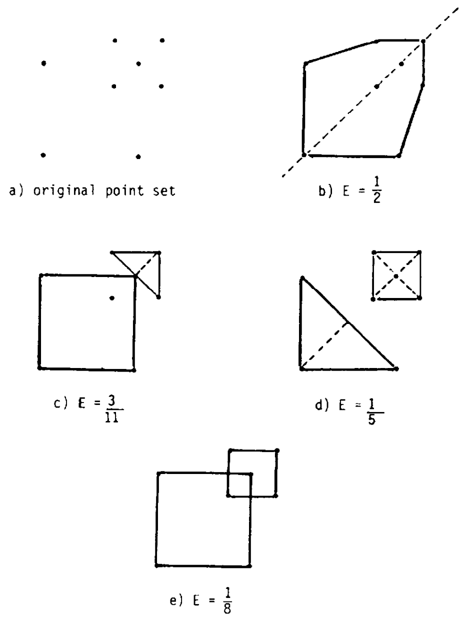

Figure 3-1: Evaluation of Point Set Decompositions

These statements can be explicitly quantized by observing the value of E on decompositions of a set S of N points. Suppose S has a regular polygonal decomposition into m polygons. Then this decomposition has E equal to $\frac{m}{2N}$. If we lump m-k of the polygons into a single set (which is not a regular polygon) then the new decomposition has E equal to $\frac{k}{2n+x}$ where n is the number of points left in regular polygons, and x is between 1 and 4. (This fact is a by-product of the omitted proofs.) We see that $\frac{k}{2n+x} < \frac{m}{2N}$ when $N - (\frac{m}{k})n < (\frac{m}{k})(\frac{x}{2})$.

Notice that $\frac{m}{k}$ measures how many more objects there are in the regular polygonal decomposition, while n (n<N) is the number of points left in regular polygons. So, for instance, if it is possible to lump together at least half the regular polygons using no more than half the total points, it always pays to do so, according to the measure E.

4. ALGORITHM FOR DECOMPOSITIONS OF 2D FINITE POINT SETS

In this section, the evaluation function, E, is as defined in Section 2. We also assume familiarity with the Hough transform method of line finding in spatial data (Duda and Hart, 1972).

The decomposition algorithm presented here depends upon the proposition in Section 3, and also on the following observation. If S is symmetric with respect to a reflection, then the midpoints of the pairs of points in S which are reflected onto each other will lie on the line of reflection. Furthermore, the line joining any two reflection-related points will be perpendicular to the line of reflection.

These observations motivate the following outline of the algorithm. We take the set of midpoints of all $\binom{N}{2}$ pairs of points in S, and associated with each midpoint the orientation of the line of reflection it would lie on if the pair of points associated to the midpoint were in a subset of decomposition induced by a reflection. Two coincidental midpoints, as illustrated in Figure 4-1, where points A and B have the same midpoint and orientation as points C and D, must also be distinguished by an appropriate data handling mechanism. To points already in S we associate all possible (quantized, therefore, finitely many) orientations. We then apply the Hough transform line finding algorithm to this set of $N + \binom{N}{2}$ points from the original N points and their midpoints with their associated (possibly multiple) orientations. Lines found by the Hough algorithm are candidates for axes of reflection. Error can be adjusted here in the quantization cells of the Hough transform to allow detection of reflectional symmetry to be as loose or tight as desired. Lines with high Hough transform values are good candidates for axis of reflectional symmetry because many midpoints lie on them.

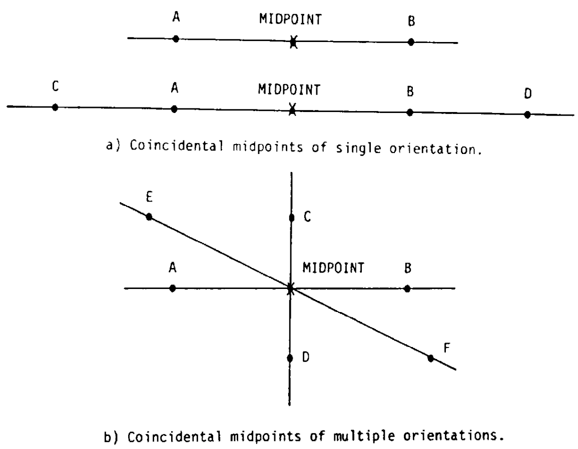

Figure 4-1: Coincidental Midpoints Must Be Distinguished

To each line found by the Hough technique, a subset of the points of S can be uniquely associated which have reflectional symmetry with respect to that line, namely the points which lie on the line or those pairs of points whose associated midpoint(s) lie on the axis of reflection. We now cluster lines of reflection by grouping lines associated to midpoints which appear with multiple orientations.

In each line group we compute the pairwise composition of the reflections across each pair of lines. Since the composition of two reflections is a rotation, this gives a set of angles of rotation associated to each line group. These angles are searched to determine sequences of multiples, i.e., sequences of angles $\{\theta, 2\theta, 3\theta, \ldots, k\theta = \pi\}$. The intersection of the subsets of the points associated to the lines associated to the angles, yield either regular k-polygons or lines of points in the original set S, (if k=1).

Sort the subsets of S by their symmetry measures. We now choose a decomposition of S by beginning with the subset of smallest measure, and adding them in order of increasing measure as long as at least one point not already included is added in.

When all points are in at least one chosen subset, we have, say, m subsets which together contain N points. We now sequentially remove the subsets of largest measure until the function $N - (\frac{m}{m-k})(n+1) > 0$ where k is the number of subsets removed and n is the number of points left. (This is an application of the "lumping together" criteria from section 3.) An optimal decomposition is given by the remaining regular polygons and lines, and the single subset obtained by lumping together all the removed subsets. The worst case step in this procedure is order $(\binom{N}{2})^4$, so the algorithm can be completed in polynomial $(o(N^8))$ time.

5. SUMMARY

We have developed a mathematical model of the concept of the symmetry of objects presented as spatial data. This model provides a domain independent approach to autonomous machine decomposition of objects into component parts. A polynomial time algorithm for performing such decompositions on finite point sets in E^2 was presented. Furthermore, there is reason to believe, based on simple examples and studies in qualitative perception by experimental psychologists, that these decompositions are similar to those which humans would choose.

REFERENCES

[1] Allport, F.H., Theories of Perception and the Concept of Structure, Wiley & Sons, Inc., New York, 1955.

[2] Birkhoff, G.D., Aesthetic Measure, Harvard University Press, Cambridge, MA, 1932.

[3] Brooks, R., "Symbolic Reasoning Among 3-Dimensional Models and 2-Dimensional Images," Artificial Intelligence, 1981.

[4] Crowley, J.L., "A Representation of Shape Based on Peaks and Ridges in the Difference of Low-Pass Transform," IEEE-PAMI, Vol. PAMI-6, No. 2, March 1984.

[5] Duda, R.O. and Hart, P.E., "Use of the Hough Transformation to Detect Lines and Curves in Pictures", Communications ACM, Vol. 15, No. 1, January, 1972.

[6] Eysenck, H.J., "An Experimental Study of Aesthetic Preference for Polygonal Figures," The Journal of General Psychology 79, 1968.

[7] Green, R.I. and Courtis, M.C., "Information Theory and Figure Perception: The Metaphor that Failed," Acta Psychologica 25, North-Holland Publishing Co., 1966.

[8] Hoffman, W.C., "The Lie Algebra of Visual Perception," Journal of Mathematical Psychology, Vol. 3, 1966.

[9] Marr, D., "Representing and Computing Visual Information," in Artificial Intelligence: An MIT Perspective, Vol. 12, P. Winston and R.H. Brown [Eds.], MIT Press, 1979.

[10] Mays, M.E., "Functions Which Parametrize Means," American Mathematical Monthly, Vol. 90, No. 10, December 1983.

[11] Samet, H. and Webber, R. E., "Using Quadtrees to Represent Polygonal Maps," Proceedings of Computer Vision and Pattern Recognition 83, Washington, D.C., June 1983.

[12] Wechsler, H., "Texture Analysis-A Survey," Signal Processing, Vol. 2, North-Holland Publishing Co., 1980.

[13] Weyl, H. The Classical Groups, Princeton University Press, 1939.

[14] Weyl, H. Symmetry, Princeton University Press, 1952.

[15] Zusne, L., Visual Perception of Form, Academic Press, NY, 1970.