PARAMETRIC CORRESPONDENCE AND CHAMFER MATCHING:
TWO NEW TECHNIQUES FOR IMAGE MATCHING

H.G. Barrow J.M. Tenenbaum, R.C. Bolles, H.C. Wolf
SRI International, Menlo Park, California 94025

Abstract

Parametric correspondence is a technique for matching images to a three dimensional symbolic reference map. An analytic camera model is used to predict the location and appearance of landmarks in the image, generating a projection for an assumed viewpoint. Correspondence is achieved by adjusting the parameters of the camera model until the appearances of the landmarks optimally match a symbolic description extracted from the image

The matching of image and map features is performed rapidly by a new technique, called "chamfer matching", that compares the shapes of two collections of shape fragments, at a cost proportional to linear dimension, rather than area These two techniques permit the matching of spatially extensive features on the basis of shape, which reduces the risk of ambiguous matches and the dependence on viewing conditions inherent in conventional image-based correlation matching.

Introduction

Many tasks involving pictures require the ability to put a sensed image into correspondence with a reference image or map. Examples include vehicle guidance, photo interpretation (change detection and monitoring) and cartography (map updating). The conventional approach is to determine a large number of points of correspondence by correlating small patches of the reference image with the sensed image. A polynomial interpolation is then used to estimate correspondence for arbitrary intermediate points [Bernstein]. This approach is computationally expensive and limited to cases where the reference and sensed images were obtained under similar viewing conditions. In particular, it cannot match images obtained from radically different viewpoints, sensors, or seasonal or climatic conditions, and it cannot match images against symbolic maps.

Parametric correspondence matches images to a symbolic reference map, rather than a reference image. The map contains a compact three dimensional representation of the shape of major landmarks, such as coastlines, buildings and roads. An analytic camera model is used to predict the location and appearance of landmarks in the image, generating a projection for an assumed viewpoint. Correspondence is achieved by adjusting the parameters of the camera model (i.e. the assumed viewpoint) until the appearances of the landmarks optimally match a symbolic description extracted from the image

The success of this approach requires the ability to rapidly match predicted and sensed appearances after each projection. The matching of image and map features is performed by a new technique, called "chamfer matching", that compares the shapes of two collections of curve fragments at a cost proportional to linear dimension, rather than area.

In principle, this approach should be superior, since it exploits more knowledge of the invariant three dimensional structure of the world and of the imaging process. At a practical level, this permits matching of spatially extensive features on the basis of shape, which reduces the risk of ambiguous matches and dependence on viewing conditions.

Chamfer Matching

Point landmarks such as intersections or promontories, are represented in the map with their associated three dimensional world coordinates. Linear landmarks, such as roads or coastlines are represented as curve fragments with associated ordered lists of world coordinates. Volumetric structures, such as buildings or bridges, are represented as wire frame models.

From a knowledge of the expected viewpoint, a prediction of the image can be made by projecting world coordinates into corresponding image coordinates, suppressing hidden lines. The problem in matching is to determine how well the predicted features correspond with image features, such as edges and lines.

The first step is to extract image features by applying edge and line operators or tracing boundaries. Edge fragment linking [Nevatia, Perkins] or relaxation enhancement [Zucker, Barrow] is optional. The net result is a feature array each element of which records whether or not a line fragment passes through it. This process preserves shape information and discards greyscale information, which is less invariant.

To correlate the extracted feature array directly with the predicted feature array would encounter several problems: The correlation peak for two arrays depicting identical linear features is very sharp and therefore intolerant of slight misalignment or distortion (e.g., two lines, slightly rotated with respect to each other, can have at most one point of correspondence) [Andrus]; A sharply peaked correlation surface is an inappropriate optimization criterion because it provides little indication of closeness to the true match, nor of the proper direction in which to proceed; Computational cost is heavy with large feature arrays.

A more robust measure of similarity between the two sets of feature points is the sum of the distances between each predicted feature point and the nearest image point. This can be computed efficiently by transforming the image feature array into an array of numbers representing distance to the nearest image feature point. The similarity measure is then easily computed by stepping through the list of predicted features and simply summing the distance array values at the predicted locations.

The distance values can be determined in two passes through the image feature array by a process known as "chamfering" [Munson, Rosenfeld]. The feature array (F[i j]. i.j=1 N) is initially two-valued: 0 for feature points and infinity otherwise. The forward pass modifies the feature array as follows:

```
FOR i ← 2 STEP 1 UNTIL N DO
  FOR j ← 2 STEP 1 UNTIL N DO
    F[i.j] ← MINIMUM(F[i.j]. (F[i-1 j]+2).
              (F[i-1.j-1]+3). (F[i,j-1]+2),
              (F[i+1.j-1]+3));
```

Similarly, the backward pass operates as follows:

```
FOR i ← (N-1) STEP -1 UNTIL 1 DO
  FOR j ← (N-1) STEP -1 UNTIL 1 DO
    F[i,j] ← MINIMUM(F[i,j], (F[i+1,j]+2),
              (F[i+1,j+1]+3), (F[i,j+1]+2).
              (F[i-1.j+1]+3));
```

The incremental distance values of 2 and 3 provide relative distances that approximate the Euclidean distances 1 and the square root of 2.

Chamfer matching provides an efficient way of computing the integral distance (i.e. area) or integral squared distance. between two curve fragments, two commonly used measures of shape similarity. Note that the distance array is computed only once, after image feature extraction.

Parametric Correspondence

Parametric correspondence puts an image into correspondence with a three dimensional reference map by determining the parameters of an analytic camera model (3 position and 3 orientation parameters).

The traditional method of calibrating the camera model takes place in two stages: first, a number of known landmarks are independently located in the image, and second. the camera parameters are computed from the pairs of corresponding world and image locations, by solving an over-constrained set of equations [Sobel, Quam, Hannah].

The failings of the traditional method stem from the first stage. The landmarks are found individually, using only very local context (e.g. a small patch of surrounding image) and with no mutual constraints. Thus local false matches commonly occur. The restriction to small features is mandated by the high cost of area correlation, and by the fact that large image features correlate poorly over small changes in viewpoint.

Parametric correspondence overcomes these failings by integrating the landmark-matching and camera calibration stages. It operates by hill-climbing on the camera parameters. A transformation matrix is constructed for each set of parameters considered, and it is used to project landmark descriptions from the map onto the image at a particular translation, rotation, scale and perspective. A similarity score is computed with chamfer matching and used to update parameter values. Initial parameter values are estimated from navigational data.

Integrating the two stages allows the simultaneous matching of all landmarks in their correct spatial relationships. Viewpoint problems with extended features are avoided because features are precisely projected by the camera model prior to matching. Parametric correspondence has the same advantages as rubber-sheet template matching [Fischler, Widrow] in that it obtains the best embedding of a map in an image, but avoids the combinatorics of trying arbitrary distortions by only considering those corresponding to some possible viewpoint.

An Example

The following example illustrates the major concepts in chamfer matching and parametric correspondence. A sensed image (Figure 1) was input along with manually derived initial estimates of the camera parameters. A reference map of the coastline was obtained. using a digitizing tablet to encode coordinates of a set of 51 sample points on a USGS map. Elevations for the points were entered manually. Figure 2 is an orthographic projection of this three dimensional map.

A simple edge follower traced the high contrast boundary of the harbor, producing the edge picture shown in Figure 3. The chamfering algorithm was applied to this edge array to obtain a distance array. Figure 4 depicts this distance array; distance is encoded by brightness with maximum brightness corresponding to zero distance from an edge point.

Using the initial camera parameter estimates, the map was projected onto the sensed image (Figure 5). The average distance between projected points and the nearest edge point, as determined by chamfer matching, was 25.8 pixels.

A straightforward optimization algorithm adjusted the camera parameters, one at a time, to minimize the average distance. Figures 6 and 7 show an intermediate state and the final state, in which the average distance has been reduced to 0.8 pixels. This result, obtained with 51 sample points, compares favorably with a 1.1 pixel average distance for 19 sample points obtained using conventional image chip correlation followed by camera calibration. The curves in Figure 8 characterize the local behavior of this minimum, showing how average distance varies with variation of each parameter from its optimal value. Approximately 60 iterations (each involving a parameter adjustment and reprojection). were required for this example. The number of iterations could be reduced by using a better optimization algorithm for example, a gradient search.

Discussion

We have presented a scheme for establishing correspondence between an image and a reference map that integrates the processes of landmark matching and camera calibration. The potential advantages of this approach stem from 1) matching shape, rather than brightness, 2) matching spatially extensive features, rather than small patches of image, 3) matching simultaneously to all features, rather than searching the combinatorial space of alternative local matches, 4) using a compact three dimensional model, rather than many two dimensional templates.

Snape has proved to be much easier to model and predict than brightness. Shape is a relatively invariant geometric property whose appearance from arbitrary viewpoints can be precisely predicted by the camera model. This eliminates the need for multiple descriptions, corresponding to different viewing conditions, and overcomes difficulties of matching large features over small changes of viewpoint.

The ability to treat the entirety of the relevant portion of the reference map as a single extensive feature reduces significantly the risk of ambiguous matches, and avoids the combinatorial complexity of finding the optimal embedding of multiple local features.

A number of obstacles have been encountered in reducing the above ideas to practice The distance metric used in chamfer matching provides a smooth, monotonic measure near the correct correspondence, and nicely interpolates over gaps in curves. However, scores can be unreliable when image and reference are badly out of alignment. In particular, discrimination is poor in textured areas, aliasing can occur with parallel linear features, a single isolated image feature can support multiple reference features.

The main problem is that edge position is not a distinguishing feature and consequently many alternative matches receive equal weight. One way of overcoming this problem, therefore, is to use more descriptive features brightness discontinuities can be classified, for example, by orientation, by edge or line, and by local spatial context (texture versus isolated boundary). Each type of feature would be separately chamfered and map features would be matched in the appropriate array. Similarly, features at a much higher level could be used, such as promontory or bay, area features having particular internal textures or structures, and even specific landmarks, such as "the top of the Transamerica pyramid". Ideally, with a few highly differentiated features distributed widely over the image the parametric correspondence process would be able to home in directly on the solution regardless of initial conditions.

Another dimension for possible improvement is the chamfering process itself. Determining for each point of the array a weighted sum of distances to many features (e.g. a convolution with the feature array), instead of the distance to the nearest feature, would provide more immunity from isolated noise points. Alternatively, propagating the coordinates of the nearest point instead of merely the distance to it, it becomes possible to use characteristics of features, such as local slope or curvature, in evaluating the goodness of match. It also makes possible a more directed search, since corresponding pairs of points are now known, an improved set of parameter estimates can be analytically determined.

Chamfer matching and parametric correspondence are separable techniques. Conceptually, parametric correspondence can be performed by re projecting image chips and evaluating the match with correlation. However, the cost of projection and matching grows with the square of the template size: The cost for chamfer matching grows linearly with the number of feature points. Chamfer matching is an alternative to other shape matching techniques, such as chain-code correlation [Freeman], Fourier matching [Zahn], and graph matching [e.g. Davis]. Also, the smoothing obtained by transforming two edge arrays to distance arrays via chamfering can be used to improve the robustness of conventional area-based edge correlation.

Parametric correspondence, in its most general form, is a technique for matching two parametrically related representations of the same geometric structure. The representations can be two- or three-dimensional, iconic or symbolic; the parametric relation can be perspective projection, a simple similarity transformation, a polynomial warp, and so forth. This view is similar to rubber-sheet template matching as conceived by Fischler and Widrow [Fischler, Widrow]. The feasibility of the approach in any application, as Widrow points out, depends on efficient algorithms for "pattern stretching, hypothesis testing, and pattern memory", corresponding to our camera model, chamfer matching, and three dimensional map.

As an illustration of its versatility, the technique can be used with a known camera location to find a known object whose position and orientation are known only approximately. In this case, the object's position and orientation are the parameters; the object is translated and rotated until its projection best matches the image data. Such an application has a more iconic flavor, as advocated by Shepard [Shepard], and is more integrated than the traditional feature extraction and graph matching approach [Roberts, Falk and Grape].

As a final consideration, the approach is amenable to efficient hardware implementation. There already exists commercially available hardware for generating parametrically specified perspective views of wire frame models at video rates, complete with hidden line suppression. The chamfering process itself requires only two passes through an array by a local operator, and match scoring requires only summing table lookups in the resulting distance array.

Conclusion

Iconic matching techniques, such as correlation, are known for efficiency and precision obtained by exploiting all available pictorial information, especially geometry. However, they are overly sensitive to changes in viewing conditions and cannot make use of non-pictorial information. Symbolic matching techniques, on the other hand, are more robust because they rely on invariant abstractions, but are less precise and less efficient in handling geometrical relationships. Their applicability in real scenes is limited by the difficulty of reliably extracting the invariant description. The techniques we have put forward offer a way of combining the best features of iconic and symbolic approaches.

Acknowledgments

This work was supported by ARPA under contract DAAG29-76-C-0012. Additional support was provided by NASA under contract NASW-2865.

REFERENCES

Andrus J.F., Campbell, C.W. and Jayroe R.R., "A Digital Image Registration Method Using Boundary Maps", IEEE Trans. Comp., Sept. 1975, p. 935-939.

Barrow, H.G., "Interactive Aids for Cartography and Photo Interpretation", Interim Report of ARPA Project DAAG29-76-C-0057, Artificial Intelligence Center, Stanford Research Institute, Menlo Park, California. Dec. 1976.

Bernstein, R., "Digital Image Processing of Earth Observation Sensor Data", IBM Journal of Research and Development, Vol. 20, No. 1, 1976.

Davis, L., "Shape Matching Using Relaxation Techniques". TR-480. Computer Science Dept., University of Maryland, Sept. 1976.

Falk, G., "Interpretation of Imperfect Line Data as a Three Dimensional Scene," Artificial Intelligence, Vol. 3, 1972, p.101-144.

Feder, J. and Freeman, H., "Segment Fitting of Curves in Pattern Analysis Using Chain Correlation", AD619525, March 1965.

Fischler, M. and Elschlager, R., "The Representation and Matching of Pictorial Structures", IEEE Trans. Comp., no. 22, 1973, p.67-92.

Grape, G.R. "Model Based (Intermediate Level) Computer Vision", Stanford AI Memo AIM-201, May 1973, Ph.D. Thesis.

Hannah, M.J., "Computer Matching of Areas in Stereo Images", Stanford AI Memo AIM-239, July 1974, Ph.D. Thesis.

Munson, J., Internal memo, Artificial Intelligence Center, Stanford Research Institute, Menlo Park, California, Dec. 1973.

Perkins, W.P., "Multi-level Vision Recognition System", Proc. Tnird IJCPR Nov. 1976, p.739-744.

Quam, L.H., "Computer Comparison of Pictures", Stanford AI Project Memo AIM 144, May 1971, Ph.D. Thesis.

Roberts, L.G., "Machine Perception of Three-dimensional Objects", in: Tippet, J.T., et al. (Eds.), "Optical and Electro-Optical Information Processing", MIT Press, 1965, p.159-197.

Rosenfeld, A. and Pflatz, J.L., "Distance Functions on Digital Pictures", Pattern Recognition, Vol. 1 No. 1, July 1968, p.33 62.

Shepard, R.N. and Metzler, J., "Mental Rotation of Three-Dimensional Objects", Science, 1971, p.701-703.

Sobel, I., "On Calibrating Computer Controlled Cameras for Perceiving 3-D Scenes", Artificial Intelligence, Vol. 5, 1974, p.185-198.

Widrow, B., "The 'Rubber-Mask' Technique", Pattern Recognition, Vol. 5, 1973, p.175-211.

Zahn, C. and Roskies, R., "Fourier Descriptors for Plane Closed Curves", IEEE Trans. Comp., no. 21, 1972, p.269-281.

Zucker, S., Hummel, R. and Rosenfeld, A., "An Application of Relaxation Labeling to Curve and Line Enhancement", IEEE Trans. Comp., no. 25, 1976.

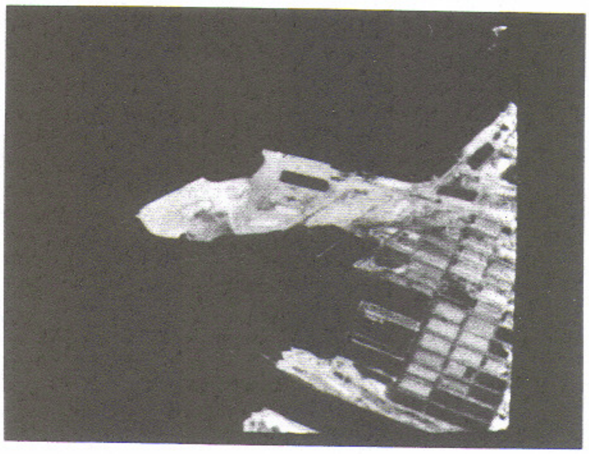

Figure 1. An aerial image of a section of coastline.

Figure 2. A set of sample points taken from a USGS map.

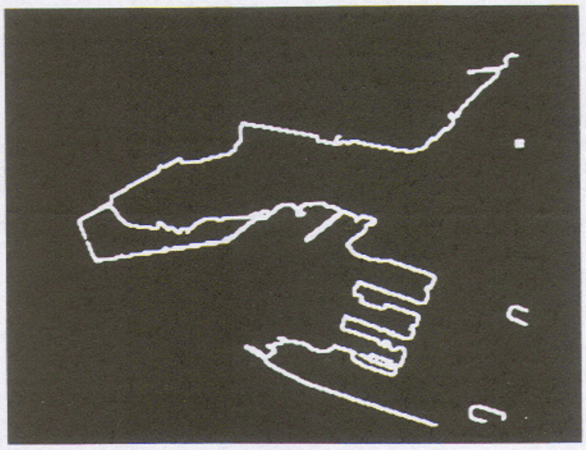

Figure 3. The traced boundary of the coastline.

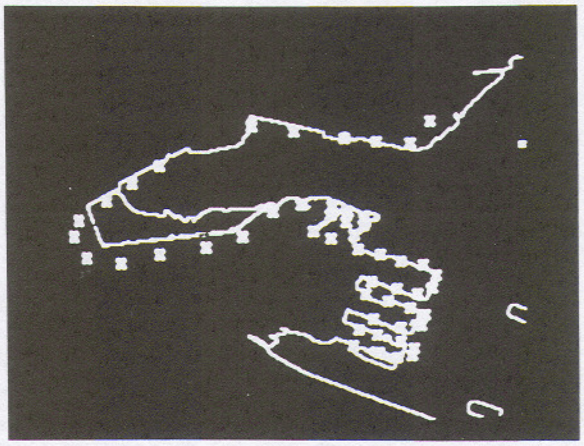

Figure 6. Projection of map points onto the image after some adjustment of camera parameters.

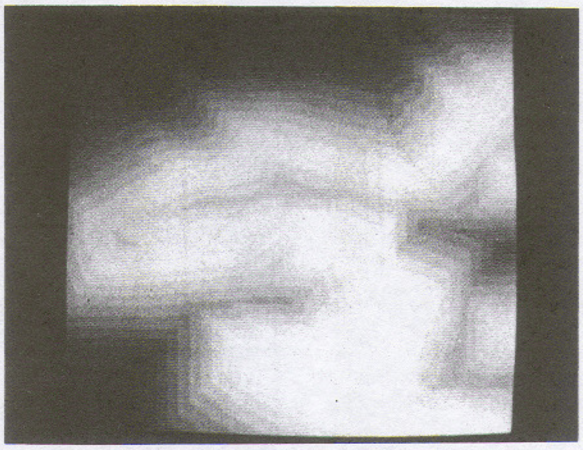

Figure 4. The distance array produced by chamfering the boundary.

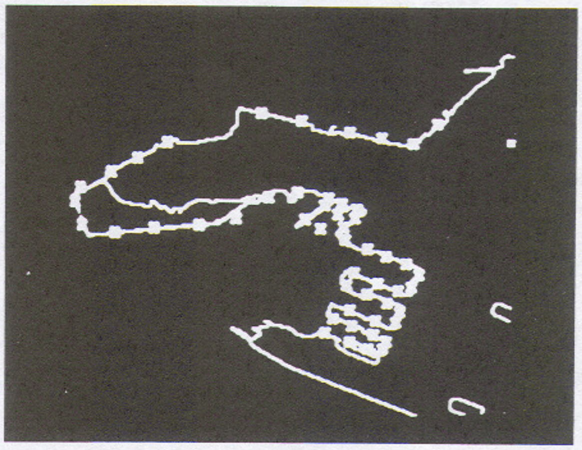

Figure 7. Projection of map points onto the image after optimization of camera parameters.

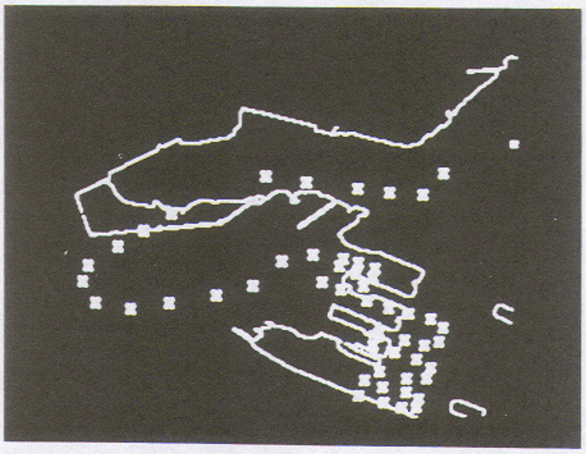

Figure 5. Initial projection of map points onto the image.

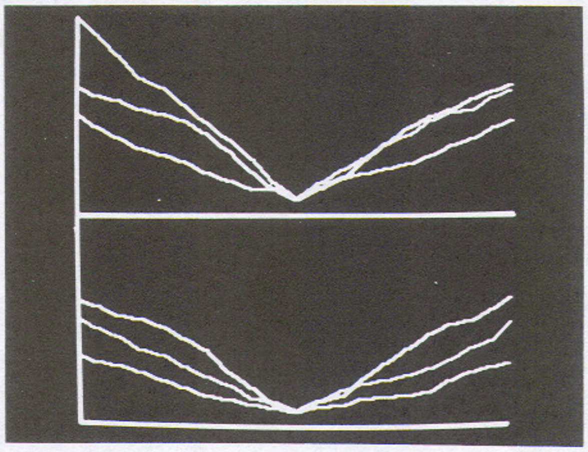

Figure 8. Behavior of average distance score with variation of the six camera parameters from their optimal values.