SMART SENSOR FOR TERMINING HOMING

Robert Hummel, Durga Panda, and Raj Aggarwal

Honeywell Systems & Research Center
2600 Ridgeway Parkway
Minneapolis, Minnesota 55413

Abstract

The practical scene matching problem presents certain complications which must extend classical image processing capabilities. In this paper, we consider certain aspects of the scene matching problem which must be addressed by a smart sensor for terminal homing. In the first section, we outline a philosophy for treating the matching problem for the terminal homing scenario. Later, we consider certain aspects of the feature extraction process and symbolic pattern matching.

Fundamentals

We begin with a reference image which contains as much detail as possible about a region of space near the target ground space. The form of the representation is a central question, but the ultimate effect is that a representative description of the expected scene can be constructed from a range of viewing positions and orientations.

From a specified sensor position and orientation, a model of the expected sensor output can be built. Since illumination models present additional complications, pixel intensity information is less important than the higher order constructs, such as edges, corners, and regions, represented in a predicted reconstruction. Thus the stored information that the edge of a building connects locations (x_1,y_1,z_1) with (x_2,y_2,z_2), together with sensor position parameters, yields a prediction of an edge on the sensor image plane connecting points (X_1,Y_1) with (X_2,Y_2). The three dimensional structure of the reference is important when the sensor is close to scenes containing tall objects, such as buildings, and in that case the reconstruction process must include hidden surface suppression. For relatively far range scenes, or relatively flat target areas, it generally suffices to represent the reference data in terms of its two dimensional domain coordinatized by a ground based coordinate system.

The stored data might include an approximated intensity image, perhaps a registered array of expected edge responses, a symbolic description of the probable objects and their features, and relational descriptors modifying the objects. Relational descriptors about objects should be generated in the reconstruction, although lists might be maintained pointing to pairs of objects about which a relation description should be generated.

The perspective transform which converts the reference data into a predicted scene depends on six parameters. [1] In practice, searching the (six dimensional) parameter space for the best match with the sensor image is unrealistic. Instead, symbolic techniques should be used to identify match points and match lines, which provide relations which constrain the parameter values.[2] In some implementations, investigators have used affine linear transformations or polynomial coordinate transforms to approximate the six parameter perspective transformation, as determined by match points and match lines.[3] Although this is apparently an unnecessary complication, in practice the qualitative behavior of the transformation varies smoothly with changes in the match point information, whereas the nonlinear solution equations to obtain the sensor viewing position parameters using perspective transforms can be unstable. When one uses a smooth approximating transformation to model the perspective transformation, the sensor parameters are of secondary importance. Instead, the approximating transformation can be used to locate the target position in the sensed image, and provide incremental update guidance to the missile's navigation system.

Because of the complexities of the matching problem, it is probably wiser to attempt to solve for an approximating transformation, rather than backsolving for the six sensor parameters.

Perhaps as matching techniques become more sophisticated, and sensors become smarter, the terminal homing device will be able to provide the navigation system with precise location information.

Extracting Symbolic Descriptions

The symbolic description of the sensed scene is formed from a small set of primitive scene constructs. The minimum set of constructs needed for symbolic descriptions consists of straight lines, representing edges and lines observed in the sensed images, and corner descriptors, which express relational features between the lines.

Curved lines, edges, regions and spots could also be included as primitives. Typically, however, straight line and corner primitives are used to build symbolic descriptions, which can include configurations and regions formed from the primitives.

There are many edge and line detectors available which can be used by a long line extractor to locate long lines and edges. For example, one can use a Sobel edge detector, followed by non-maxima suppression and thresholding, as input to a merger which locates initial and terminal points of aligned responses, optionally bridging gaps.[2]

These algorithms are essentially non-adaptive, unintelligent functional operators which operate on pixel intensities. An intelligent sensor should be able to "look more closely" for edges in certain regions, or certain orientations. It is frequently more important to extract weak edges, if they are long, while rejecting edge segments which might be strong but not sufficiently long or prominent in the scene. As conditions change, the edge extraction process must be easily adjusted automatically.

One way to perform local edge and line extraction, making use of a knowledge base which can adjust for conditions or cueing, is to use relaxation labeling.[4] Earlier studies have shown that relaxation labeling can be used for line and curve extraction.[5] The relaxation labeling process performs a general labeling function, by assigning labels from a set of possible labels to each of many objects. In line and curve extraction, each pixel is an object and the labels consist of line descriptors specifying one of quantized set of orientations, or a no-line label.

The first step of a relaxation labeling process to extract lines and edges is to obtain initial estimates for label assignment values. This can be done by applying local operators at every point in the image to obtain an edge or line response for each orientation at every point. The initial estimate of the no-line label should be inversely related to the strength of the edge and line responses at the point. Thus at iteration zero, at every pixel (say, pixel number i), we have m distinct line orientation responses, $P_i(1)$, $P_i(2)$, ..., $P_i(m)$, corresponding to the m quantized orientations, and the no-line value $P_i(m+1)$. These values are normalized so that at every pixel, the sum of the m+1 values is 1.

A relaxation labeling algorithm is then used to update these values. The process is conceptually parallel and iterative, with each pixel updating its assignment values for each label independent of the processing at other pixels. At the heart of relaxation labeling is a set of compatibilities matrices.

Suppose that i and j are distinct pixels, and ℓ and ℓ' are labels. The coefficient $r_{ij}(\ell,\ell')$ is used to denote the compatibility of pixel i having label ℓ with pixel j having label ℓ'. For example, if pixels i and j are horizontally adjacent, and ℓ and ℓ' are both horizontal edge labels, then $r_{ij}(\ell,\ell')$ should be large and positive. On the other hand, if i and j are the same two pixels, ℓ is a horizontal edge label, and ℓ' is a vertical edge label, then $r_{ij}(\ell,\ell')$ should be negative, representing the fact that straight lines do not take sudden right angle turns. When there is no influence of the object label pair (j, ℓ') on the object label (i, ℓ), then $r_{ij}(\ell, \ell')$ should be zero. For example, when pixels i and j are far apart, the entire matrix of values should be zero.

Using these matrices, an updating evidence value is obtained for each object-label at every iteration according to the formula $q_i(\ell) = \sum_j \sum_{\ell'} r_{ij} (\ell, \ell') P_j (\ell')$.

The $P_j (\ell')$ values represent the current assignment values. Generally, a positive $q_i(\ell)$ means that $P_i(\ell)$ should be increased, while a negative $q_i(\ell)$ means that $P_i(\ell)$ should be decreased for the next iteration. This updating must be done subject to the normalization constraint $\sum P_i (\ell) = 1$, $P_i (\ell) \geq 0$, at every iteration. A heuristic updating formula was introduced in (4), but more recent studies have led to a better updating method (6).

Essentially, the relaxation process applies local constraints, codified by a matrix of compatibilities, to iteratively update initial labeling assignment values. As information in the scene propogates from one portion of the image to another by repeated application of local constraints, eventually the assignments relax to a labeling which is everywhere locally consistent (see [6] for recent theoretical results).

Further processing is still required to analyze global structures, extracting symbolic representations, and to perform matching. However, relaxation labeling permits one to include local processing as part of the sensor, or immediate post-sensor function. By varying the assignment compatibility parameters of the relaxation process, the extraction of lines and edges can be adjusted dynamically for varying conditions. With modern VHSIC designs real time implementation of relaxation labeling have become a real possibility[7].

Symbolic pattern matching

Using a symbolic representation of the sensed scene, a comparison can be made against the reference image in order to locate a match and establish a coordinate transformation. Symbolic techniques are crucial to the terminal homing scenario because of the unreliability of pixel intensities, and because of the high computational overhead required for correlation techniques.

Unfortunately, the matching problem in the presence of noise, ambiguity, incorrect descriptions and missed constructs is far from trivial. Standard techniques try to measure a mismatch in a proposed correspondence by computing an average value of the difference between certain feature values of lines, regions, and edges of the sensed and reference scene.[8] Search techniques can then be used to find the correspondence with the least mismatch.

Using artificial intelligence theories, the possible alternate approaches to the matching problem can become extremely sophisticated. Recent studies of the related correspondence problem in motion and steriopsis studies suggest that matching can be treated as a labeling problem.[9] In the following, we show how the matching problem can be formulated in the general framework of relaxation labeling.

Suppose that $a_1, a_2, \ldots, a_n$ are object primitives extracted from the sensed scene. There are certain objects in the reference scene, $\ell_1, \ldots\ {}_{m-1}$, and some of the sensed objects a_i may be spurious, or outside of the reference area, and thus should be labeled with a no-match label, ℓ_m. The goal of the pattern matching process is to label each object a_i with a label $\ell_{k(i)}$, where $1 \leq k(i) \leq m$ for each i.

Certain constraints can be imposed. For instance, if the objects in the reference are distinct, then two different objects a_i and a_j should not be labeled by the same ℓ_k, with $1 \leq k \leq m-1$. That is, if it is known with a large degree of certainty that object a_i is label ℓ_k, with $k \leq m-1$, then no other object a_j should have label ℓ_k.

There are also positive constraints. A match between object a_i and label ℓ_j should enhance the labeling of nearby objects with labels with are close to label ℓ_j in the reference scene. Relational structures should be taken into account when assessing the degree of enhancement, and current estimates of the coordinate transformation are also helpful in assessing the compatibility of two neighboring matches.

In terms of the matrix of compatibilities, we can translate these constraints into numerical values. For example, $r_{ij}(\ell,\ell')$ should always be negative, reflecting the bias toward label uniqueness. On the other hand, if objects a_i and a_j are neighbors, and the reference objects ℓ and ℓ' are similarly close, then $r_{ij}(\ell,\ell')$ should be positive, especially if the object relational descriptors between a_i and a_j and between ℓ and ℓ' match. For example, if the objects are straight lines, the relational descriptor might be the angle at which the lines meet.

Initial matching probabilities can be assigned by labeling each object a_i with a set of labels ℓ_j, where every reasonable match is given a nonzero possibility. Updates are made iteratively, using local constraints, until objects are matched, or everything is labeled by ℓ_m, "no match".

Implementations of this sort have been proven feasible in the image plane for disparity analysis[10], and symbolic pattern matching using relaxation labeling is currently under further investigation at Honeywell.

Conclusions

As increased processing capabilities become available for sensor and image processing systems, more sophisticated techniques for object extraction and pattern matching become appropriate. In the future, general ideas from artificial intelligence, will be more and useful for terminal homing requirements of fast scene recognition and pattern matching. Honeywell has been investigating relaxation labeling as an application of artificial intelligence to the symbolic pattern matching problem for terminal homing.

Acknowledgment

The authors wish to thank Dr. M. Geokezas, the Manager of Information Science at Honeywell for helpful discussion and encouragement, and Deborah Johnson for skillful typing in preparing this manuscript.

References

1. Rogers, D.F. and Adams, J.A., Mathematical Elements for Computer Graphics; McGraw-Hill, 1975.

2. Narendra, P.M., "Advanced Pattern Matching Techniques for Autonomous Acquisition," First Quarter Progress Report., Honeywell, Inc., Contract No. DAAK70-79-C-0114.

3. Clark, C.S., et. al. "High Accuracy Model Matching for Scenes Containing Man-Made Structures," Proceedings of the SPIE, Vol. 186, May 22-24, 1979, p. 54.

4. Rosenfeld, A.; Hummel, R.A.; and Zucker, S.W. "Scene Labeling by Relaxation Operations," IEEE - Trans. SMC, SMC-6, 1976, p. 420.

5. Zucker, S.W.; Hummel, R.A.; and Rosenfeld, A. "An Application of Relaxation Labeling to Line and Curve Enhancement," IEEE - Trans. on Computers, C-26, 1977, p. 394.

6. Hummel, R.A. and Zucker, S.W., "On the Foundations of Relaxation Labeling Processes," Computer Vision Lab, Dept. of Elec. Eng., McGill University, Tech. Rep. TR-80-7, Montreal, Quebec.

7. Honeywell, Inc., "Signal and Image Processing, "VHSIC, Electro Optical System," May, 1980, 2600 Ridgeway Parkway, Minneapolis, MN 55413.

8. Belsher, J.F.; Williams, H.F.; and Kin, R.H., "Scene Matching with Feature Detection," in Proceedings of the SPIE, Vol. 189, May 22-24, 1979, p; 12.

9. Batali, J. and Ullman, A., "Motion Detection and Analysis," in Proceedings of the Image Understanding Workshop, Nov., 1979, p. 69.

10. Barnard, S.T. and Thompson, W.B., "Disparity Analysis of Images," IEEE Trans. on Pattern Analysis and Machine Intelligence, PAMI - 2, 1980, p. 333.