Constraint-Based Stereo Matching

D.T. Kuan

FMC Corporation

Santa Clara, CA 95052

ABSTRACT

The major difficulty in stereo vision is the correspondence problem that requires matching features in two stereo images. In this paper, we describe a constraint-based stereo matching technique using local geometric constraints among edge segments to limit the search space and to resolve matching ambiguity. Edge segments are used as image features for stereo matching. Epipolar constraint and individual edge properties are used to determine possible initial matches between edge segments in a stereo image pair. Local edge geometric attributes such as continuity, junction structure, and edge neighborhood relations are used as constraints to guide the stereo matching process. The result is a locally consistent set of edge segment correspondences between stereo images. These locally consistent matches are used to generate higher-level hypotheses on extended edge segments and junctions to form more global contexts to achieve global consistency.

INTRODUCTION

Stereo vision is the process of reconstructing 3-D depth information from 2-D images. Depth information is crucial in passive navigation and scene interpretation applications. The key problem in stereo vision is to establish a correspondence between features in two images in order to calculate position in 3-D space according to stereo imaging geometry. A scene point will project to two image planes through cameras. The epipolar lines are intersections of the two image planes with an epipolar plane defined by the scene point and the two camera foci. Based on this relationship, if we locate one feature point in one image, then the corresponding feature in the other image must lie on the corresponding epipolar line. The displacement between two corresponding feature points is termed disparity.

There are basically two types of automated stereo matching techniques - area-based and feature-based. Early work on stereo vision used area-based cross-correlation techniques for image correspondence. For example, Moravec [5] used "interest points" and image intensity cross-correlation measure for stereo matching. An interest point is an image feature with significant intensity variation around it and is usually a corner point of an object. A set of interest points are extracted from one image and the corresponding features in the other image are searched using a hierarchical correlation technique. No global consistency is checked in the matching process. For images with similar intensity variation and no significant occlusion effect, the area-based stereo analysis technique works well. However, the technique may fail in the presence of repetitive features, surface discontinuity, and intensity variation.

The edge-based stereo technique first extracts edge features from both stereo images and uses various constraints to resolve the correspondence problem. Grimson [3] implemented an edge-based stereo algorithm with a coarse-to-fine strategy and the uniqueness and continuity constraints originally proposed by Marr and Poggio [4]. The zero-crossing edge pixels are used as features, and disparity similarity is used to determine the final matches. Recent work by Baker and Arnold [1,2] incorporates geometric constraints into the dynamic programming algorithm to match edge pixels in a single scanline and uses an edge connectivity constraint to guide the inter-scanline matching. Due to the limitation of the dynamic programming algorithm, edge pixel correspondence between the left and right images has a strict order sequence and edge reversal is not allowed. The advantages of edge-based stereo include faster processing speed (because it requires fewer features to match) more accurate results (because edges may be located with sub-pixel precision), and less sensitivity to intensity variation (because edges represent geometric features).

In this paper, we use the edge segment instead of the edge pixel as the primitive image feature. This choice has several advantages. The edge segment is a group of consistent edge pixels and is a more stable and robust feature primitive. The inter-scanline continuity constraint used in [2] is implicitly imbedded in the edge segment primitive. Vertical disparity problem that exists in edge-pixel based stereo algorithms does not happen here. In addition, the number of edge segments is much less than the number of edge pixels. This significantly reduces the computation time for stereo matching. Finally, geometric relations among edge segments can be used as constraint in stereo matching.

In our approach, edge pixels are first detected and linked into edge segments. Edge segment orientation and intensity profile are used as matching properties. Junctions are detected and classified according to their types. A junction is a place to propagate geometric constraints from one edge to all the connecting edges. The edge neighborhood relation is also used as a local geometric feature to propagate constraints among edges where a junction feature does not exist.

In the initial matching stage, standard epipolar constraint and individual edge properties are used to determine possible initial matches between edge segments in the left and right images. For each initial hypothesis, we then apply a set of geometric constraints to reduce matching ambiguity. The results of constraint checking are recorded and a maximum likelihood scheme is used to select the most likely match for each edge segment in stereo images. These

locally consistent matches then generate higher-level hypotheses on extended edges and junctions to form more global contexts. The number of higher-level hypothesis is very small compared to the number of initial hypothesis because local matches have constrained the possible global matches. If a higher-level hypothesis has enough supports from local matches, it becomes a global match and can enforce global consistency to correct inconsistent local matches in the same context.

IMAGE FEATURE EXTRACTION

The image feature extraction process proceeds concurrently on two stereo images. Edge segments are first extracted, and then junctions and edge neighborhood relations are extracted based on edge segment relations.

The Edge Feature

The two stereo images are first smoothed by using a Gaussian convolution mask to reduce image noise. Edges are detected by using an eight-directional fast compass edge detector. Edge direction and magnitude information are used to thin and link edge pixels into edge chains. Then a recursive line-fitting algorithm is used to represent each edge chain as a sequence of line segments. Each sequence of line segments is stored in an extended edge structure and each line segment is represented in an edge segment structure. For each edge segment, its starting and ending points, midpoint, orientation, length, and average intensities on two sides of the edge are stored in the edge segment structure. Edge segments are indexed by dividing an image into a set of square windows to provide fast access to neighboring edges. This spatial indexing scheme speeds up local geometric feature calculation significantly. Figure 1 shows a pair of stereo images. Figure 2 shows all edge segments detected in Figure 1.

The Junction Feature

Junctions are detected by searching for all edge segments intersecting a small window attached at the ends of an edge segment. Junction type, orientation, location, associated edges, and relative edge angles are stored in the junction structure. Junction type includes L, arrow, fork, T, and a complex junction containing more than three edges. Edges associated with an junction are ordered according their orientations. A junction is a place to propagate constraints from one edge to all the connecting edges.

The Edge Neighborhood Feature

Junctions are useful features to propagate constraints between edges. However, most images contain few junctions. We use the neighborhood relation among edges as another class of local features to propagate constraints among edges. Each edge has a set of left-neighboring edges and a set of right-neighboring edges. A neighboring edge is defined as an edge that has significant vertical overlap with and is adjacent to a specified edge. The relative orientation, interval between an edge and its neighboring edges, and their vertical overlap interval, are stored in the edge neighborhood structure as a local feature for stereo matching.

CONSTRAINTS FOR STEREO MATCHING

In principle, each edge segment in the left image can match any edge segment in the right image. In practice, edge segment properties, stereo imaging geometry, edge continuity, and local geometric relations greatly constrain the possible matches for each edge segment. We describe in this section the constraints used in our stereo matching process.

The Epipolar Constraint

The standard epipolar constraint specifies that if an image feature exists in the left image, then the corresponding feature in the right image must lie on the epipolar line of the right image. Without loss of generality, we assume the two stereo images are properly aligned such that the epipolar lines are the scan lines of the image. In this case, for each edge in the left image, we only need to consider those right image edges having vertical overlap with the left image edge. The epipolar constraint is a strong geometric constraint based on stereo imaging geometry and has to be satisfied by all matches.

The Disparity Range Constraint

For each edge segment in the left image, we only search candidate edges within the window defined by the maximum allowed disparity interval in the right image. This constraint is applied in the initial matching stage.

The Edge Orientation Constraint

In general, the corresponding edge segments in stereo images should have similar orientations. The edge orientation constraint restricts each edge segment in the left image to match only those edges in the right image within an orientation threshold.

The Continuity Constraint

In the line-fitting procedure, it is possible that the corresponding edge of an edge segment is segmented into several pieces. To deal with this situation, partial edge segment matches must be considered. One edge segment may match more than one edge as long as these matches do not overlap in the vertical direction. In addition, these matches should have similar disparity to guarantee the corresponding edges are colinear. In the continuity constraint, if an initial match between two edge segments in the stereo images exists, and the two edges only partially overlap in the vertical direction, then we must find continuity evidence to support the partial match. We currently implement this by finding an edge connected to and colinear with the partially overlapped edge in the direction of the required extension. This constraint also applies to the complete edge segment match case where two edge segments have significant vertical overlap. In this case, only continuity in the correct direction is checked.

The Disparity Compatibility Constraint

Edges that are connected or close to each other in the image usually have similar disparities. This is based on the smoothness constraint of physical objects. In this constraint, we first identify all the edges that are connected or close to a given edge in one image by using junction structure and edge neighborhood relations. We then look for supporting matches from this edge group with similar disparity to one of the matches of the given edge, and record this information for global consistency checking.

The Junction Constraint

In the junction constraint, if two edges in the left image are in the same junction, we then check the corresponding matches in the right image to see if they are in the same junction. The relative edge segment ordering and angles in the junction are also checked to further reduce ambiguity.

The Neighborhood Relation Constraint

Each edge segment has its left and right neighbors. Unless there is occlusion or edge reversal effects, this edge neighborhood relation will be preserved in both stereo images. In the neighborhood relation constraint, if two edges are neighboring edges in the left image, then the corresponding matches in the right image should also be neighboring edges. The local features in the edge neighborhood structure are used for matching.

STEREO MATCHING

We separate the stereo matching process into three stages. In the initial matching stage, for each edge segment in the left image, we apply the epipolar constraint, disparity range constraint, and edge orientation constraint to obtain a set of possible matches. In Figure 2, 21 edge segments have no match in the right image, 38 edges have a unique match, 63 edges have two matches, 55 edges have three matches, and 153 edges have more than three matches. Most no-match edges are small edge segments or horizontal edge segments that are not very useful for stereo depth reconstruction. The constraints applied in this stage only use single edge segment properties to limit the search space.

In the second matching stage, we propagate constraints through junctions and neighboring edges to resolve multiple matching ambiguity of each edge segment. For each edge segment, a weighted average of the results after applying constraints and the edge segment similarity measure is used to calculate the likelihood of a hypothesized match. The most likely match for each edge segment in the image is selected. This process is performed for each edge segment in the left and right images. Figure 3 shows all the edge segments that have consistent left-to-right and right-to-left matches. Most matches at this stage are correct matches.

In the final matching stage, we use consistent local matches to form higher-level hypotheses on extended edges and junctions that have more global contexts. Within each context, maximum global consistency is used as a criterion to correct incorrect local matches. This is currently implemented by summing up the strength of each supporting local matches. The most likely match is then selected and enforces global consistency to correct inconsistent local matches in the same context. This process does improve the results of the second matching stage. We are currently developing more sophisticated global consistency matching technique that examines the justifications of local matches.

CONCLUSIONS

In this paper, we presented a new stereo matching technique based on geometric constraints. Edge segment is used because it has more features to compare and is more stable compared to individual edge pixels. In addition, combinatorial search has been significantly reduced in the matching process. Constraints are applied locally and the most likely match is selected based on how well the constraints fit the data. Because we use no assumption about the order of edge segment correspondence, this technique can potentially deal with more difficult stereo matching cases. We are currently working on more sophisticated global consistency matching techniques for better stereo matching.

REFERENCES

[1] Arnold, David R., "Automated Stereo Perception," Stanford Artificial Intelligence Laboratory, AIM-351, March 1983.

[2] Baker, Henry Harlyn, "Depth from Edge and Intensity Based Stereo," Stanford Artificial Intelligence Laboratory, AIM-347, September 1982.

[3] Grimson, W. E. W., "From Images to Surfaces: A Computational Study of the Human Early Visual System," MIT Press, Cambridge, Massachusetts, 1981.

[4] Marr, D, and Poggio, T., "A Computational Theory of Human Stereo Vision," Proc. R. Soc. London, pp. 301-328, 1979.

[5] Moravec, Hans P., "Obstacle Avoidance and Navigation in the Real World by a Seeing Robot Rover," Stanford Artificial Intelligence Laboratory, AIM-340, Sept. 1980.

[6] Ohta, Y., and Kanade, T., "Stereo by Intra- and Inter-Scanline Search Using Dynamic Programming," IEEE Trans. on Pattern Analysis and Machine Intelligence, March 1985, pp. 139-154.
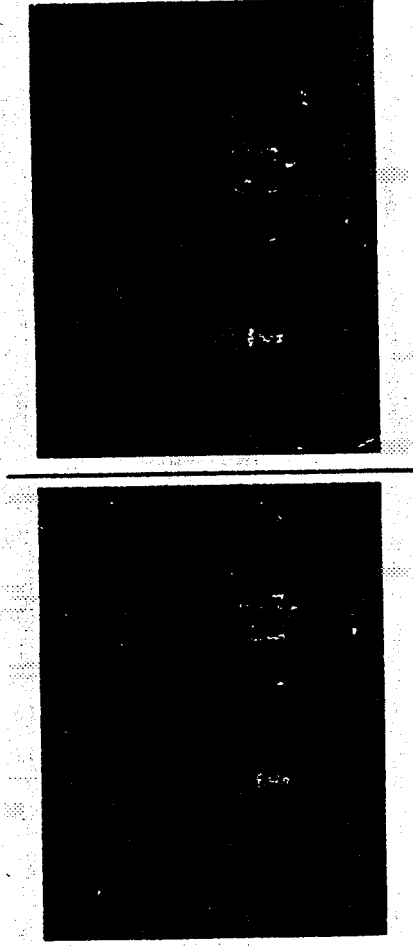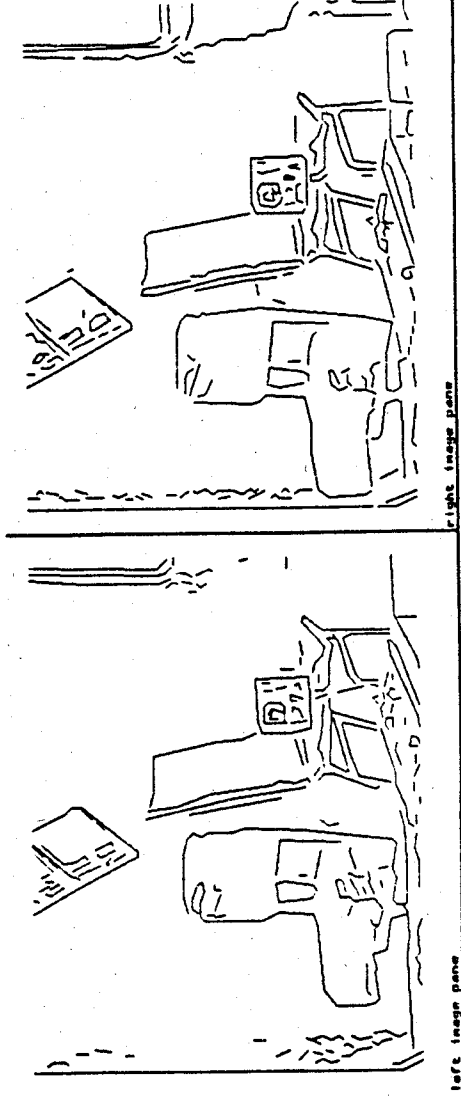
Figure 1: Stereo images.

right image pane

left image pane

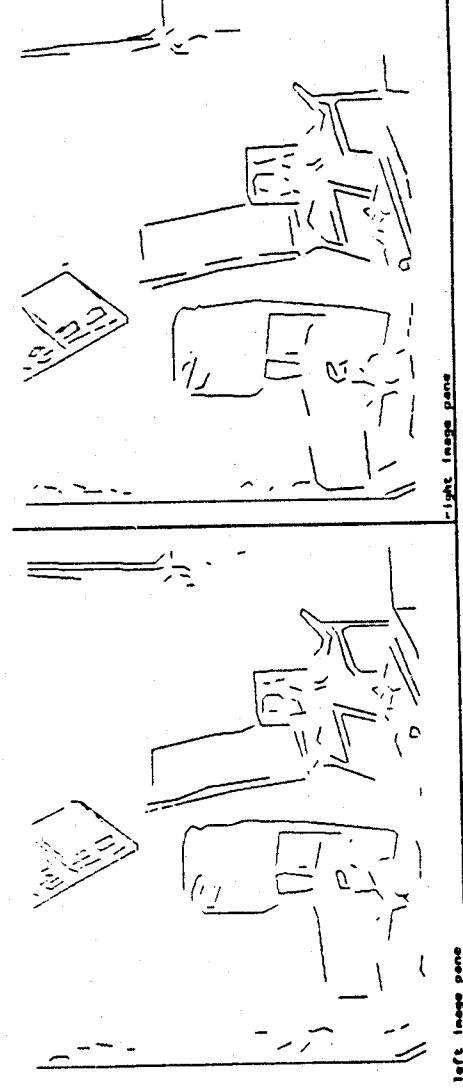Figure 2: Edge segments in Figure 1.

right image pane

left image pane

Figure 3: The final consistently matched edges in both stereo images.

134