# COARSE-TO-FINE CONTROL STRATEGY FOR MATCHING MOTION STEREO PAIRS

Gang Xu, Saburo Tsuji and Minoru Asada

Department of Control Engineering
Osaka University
Toyonaka, Osaka 560, Japan

## ABSTRACT

This article proposes a slider stereo matching method, which employs a coarse-to-fine control strategy to overcome the false targets problem. At first, a stereo pair is taken at a short baseline, and a match is assigned to it. The resulting disparity map is then used to restrict search range and to predict occlusion for efficiently and reliably matching the succeeding stereo pair taken at a longer baseline. The system iterates the sliding and matching to obtain an enough disparity range.

## 1. INTRODUCTION

It has been pointed out by the early researches[1,2] on stereopsis that at the heart of the matching problem lies the problem of false targets, whose severity is enhanced by both a wide disparity range and a fine resolution. To obtain both, which are necessary for accuracy, Marr and Poggio [2] proposed a zero-crossing algorithm (implemented by Crimson [3]) based on the coarse-to-fine strategy. The stereo images were filtered at different scales, yielding a set of primitive image pairs of different resolutions. A match of a wide disparity range but of a coarse resolution was assigned first, and this result was then used to reduce search space, for matching the finer detailed primitive pairs. While the algorithm was successfully applied to a variety of images, it resulted in incorrect matches along occluding zero-crossings, because the continuity constraint was used along those depth-discontinuous contours.

Tn this paper, we present a new stereo matching method, which is based upon a different coarse-to-fine strategy. A camera is slid along a straight line to take stereo pairs at different baselines, resulting in a set of stereo pairs of different disparity ranges, images are convolved with a small-sized Laplacian-Caussian operator, which assures us of a fine resolution, and the zero-crossings in the filtered images are extracted along horizontal scan lines. Then the matching process proceeds in a coarse-to-fine iterative manner. The pair taken at the shortest baseline is matched first; the matching is easy, at the expense of reduced disparity range. This result, a narrow range disparity map, can then be used to reduce search space for matching the longer baseline pair, again making the process easier. We iterate the sliding, imaging and matching several times to obtain a wide disparity range. One important advantage of our method over the conventional ones is that we can predict occlusion, helping avoid mismatches.

Motion stereo methods have been studied for a long time by Moravec and Nevatia. The idea, however, was used to make reliable depth measurement[4 ] and to save computing time[5], while the disparity information available from the already matched pairs was not utilized in matching the other pairs.

## 2. COARSE-TO-FINE RULE

Both our algorithm and Marr and Poggio's are based on the coarse-to-fine control strategy, aiming at obtaining both a wide disparity range and a fine resolution at last. The two algorithms, however, take different routes to reach the goal. Fig.l helps to understand the difference. While the depth accuracy in Marr and Poggic/s algorithm is improved by making the resolution finer over a wide disparity range, that in ours is improved by widening the disparity range in a fine resolution.

It is important to note that, the only physical constraint we use here is that features should not disappear, nor features newly appear, through the sliding, because the scene and the illumination condition are supposed not to change during this period. In general, it is satisfied except the cases where some feature points are occluded, as

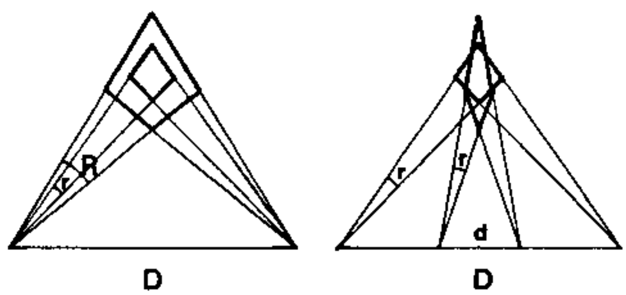


Fig.l There are two routes leading to the final goal, the accurate depth measurement. The diamonds indicate the possible errors in estimates of location of a point in scene, (a) The vertical diagonal of the diamond becomes shorter with the finer resolutions (R->r), while a wide disparity range(D) remaining unchanged, (b) The vertical diagonal becomes shorter with the disparity range getting wider(d->D)), while a fine resolution(r) remaining unchanged.

the camera is slid. These feature points, however, are assigned with disparities before occluded and the probability of occlusion is predicted from the disparity information obtained already. As a consequence, not only can we avoid the incorrect-matches due to occlusion, but it is also possible to obtain more disparity information about the feature points occluded finally than the other stereo matchers.

## 3. SLIDING AND MATCHING

A camera is slid from left to right, along a straight line, taking a sequence of images at predetermined intervals. The sliding is kept parallel to the horizontal axis *of* the camera coordinate system, so that the epipolar lines, on which correspondences can be found, are completely horizontal. Let the first image be the only left image and the remainders right ones. Note that *such* an arrangement is purely for simplifying our description, and the following matching algorithm can be easily modified for other arrangements.

Images are convolved with the Laplacian-Gaussian filter, and zero-crossings in the filtered images are found. For a fine resolution, we have chosen 4 as the filter's central region width W. Besides the location, an attribute of the zero-crossings, the contrast sign, is also recorded, and used as a criterion for matching.

The first stereo pair, due to its reduced disparity range, can be easily matched by simply searching a small region without encountering false targets. Below, we discuss how to use the obtained disparity information for matching the second stereo pair. Suppose that the baseline of the second stereo pair is k times that of the first one. Then for a zero-crossing in the left image at position (x,y), whose disparity obtained by matcing the first stereo pair is D(x,y), its correspondence's position in the right image is estimated by

(x-kD(x,y),y).

Taking into account the error *Ad* in the disparity obtained already, the position is revised as

(x-kD(x,y)+kAD,y).

Therefore, we can constrain the search for the corresponding zero-crossing in the right image of the second stereo pair to the region

{(x',y)| x-kD(x,y)-k^$D_{max}$<_x»^x-kD(x,y)+k$AD_{max}$,
D ^$^{D}$maxl.

The size of search region should be selected so that the probability for double targets in one search region is not high. A conservative value, 7 pixels, is selected for the width of the search region. Then the worst case probability of double targets is 0.5 (estimated from the result by Crimson [3]), and true matches can found by utilizing, "oulling effect" [2].

Now let us consider how long we can slide the camera in each step of the successive imaging. Given $D_{max}$, We can determine k, the ratio of the baselines, as 3/$AD_{max}$. AD results from many factors, such as errors in locating zero- crossings in quantized images and the physical inaccuracy in sliding the camera. We have not estimated all these factors yet. The experimental results, however, indicate that AD $_x$ does not exceed 1.5 pixels in most cases; 2 is currently selected for k. Clearly, we need less sliding times for obtaining an enough disparity range than conventional motion stereo methods [4,5], which take pictures at equal intervals.

## 4. OCCLUSION PREDICTION

In this section, we consider how the occlusions in the successive stereo pair is predicted from the disparity map obtained already. Suppose that there are two zero-crossings $P_1$ and $P_2$ at $(x_1,y)$ and $(x_2,y)$ on a scan line in the left image. If they are predicted as changing to the opposite order in the right image, i.e.,

$^{x}$l$^{<x}$2»

x,-kD(xj,y)>$x_2$-kD($x_2$,y),

then the probability for $P_1$ being occluded is high.

The prediction is not always true, because of the error in estimating the zero-crossing positions. Therefore, the following method is used to avoid mismatches. If the estimated distance between the occluding and occluded zero-crossings is above a threshold, then the occlusion is certain and we search for only $P_1$. If below, we search for both. In the latter case, if there is only one matchable zero-crossing, then it is considered as corresponding to $P_2$, the one closer to the camera.

## 5. EXPERIMENTS AND DISCUSSION

We have tested the described method for a blocks-world and an indoor scene. Fig.2 (a), (b) and (c) illustrate the first, second and fifth images of an image sequence of the blocksworld, respectively. Fig.2 (d) shows the zero-crossing contours of Fig.2 (c). The contours are displayed as bright if the contrast sign is positive, and as dark if negative. Fig.2 (e) shows the differences between the estimated zero-crossing positions and the real ones. It can be seen that they agree with each other quite well. The dark points indicate the estimated locations of zero-crossings and those predicted as occluded are displayed with larger spots. The bright points are the actual positions in the image. Fig.2 (f) shows the top view obtained by matching Fig.2 (a) to Fig.2(e). Fig.3(a) shows an input image of the indoor scene, and Fig.3(b) shows the disparity map, in which the disparity is displayed as proportional to the brightness.

The experiments indicate that most zero-crossings appear in the search pools, provided that the camera sliding is accurate enough. The disparities obtained along occluding zero-crossing contours also demonstrate that there are few mismatches resulting from occlusion.

One difficulty in matching the stereo pair is that the contrast sign along occluding zero-crossing contours sometimes changes as the camera is slid. Therefore, the contrast sign, which played a significant role in the implementation of Marr and Poggio's algorithm by Crimson, results in mismatches in this method. We consider that our method can be improved to accomodate such cases by predicting the changes in the background's intensity.

At present, the camera motion is along a straight line. We can, however, modify the method so as to move the camera more freely in space, and give the system intelligence to plan motion by observing the coarse disparity map obtained already.

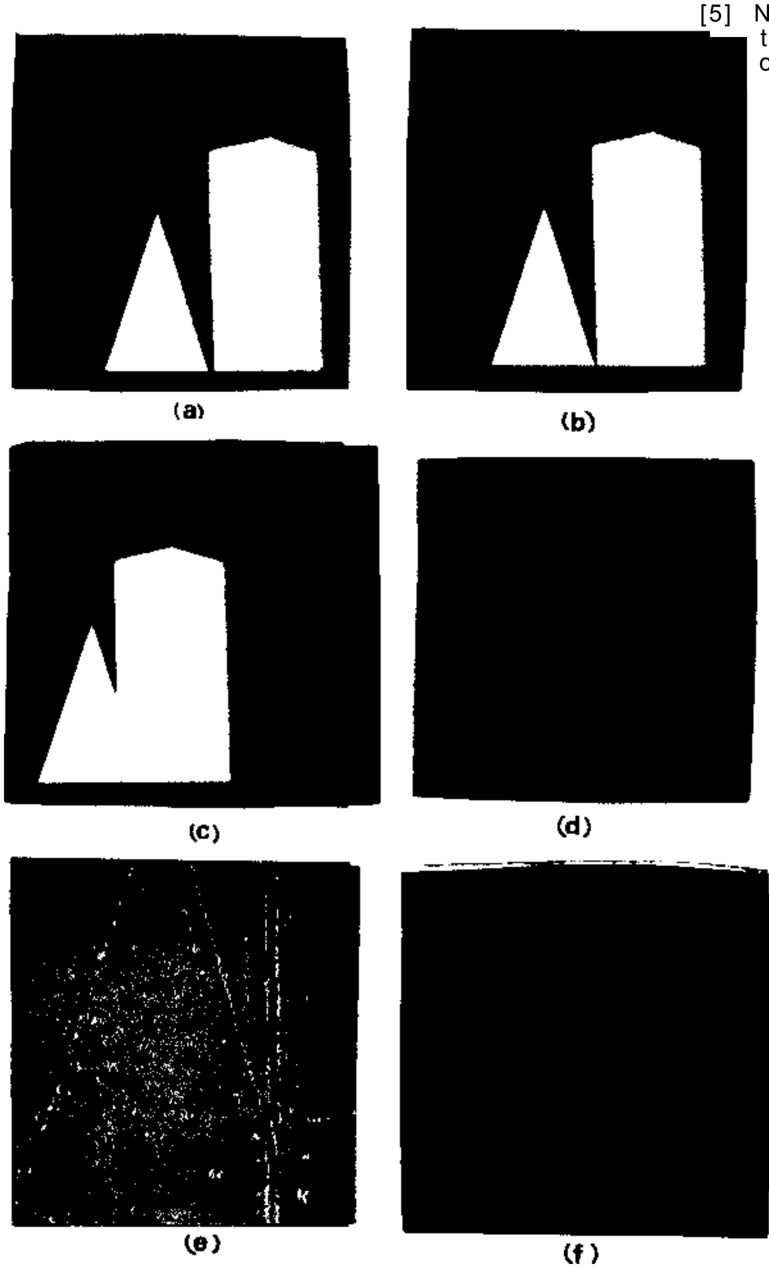

Figure 2

REFERENCES

[1] Marr, D. and Poggio, T. "Cooperative Computation of Stereo Disparity" Science, N.Y.194, 1976, pp.283-287

[2] Marr, D. and Poggio, T. "A Computational Theory of Human Stereo Vision" Proc. R. Soc. Lond. B204, 1979, pp.301-328.

[3] Crimson, W. E. L. "From Images To Surfaces: A Computational Study of the Human Early Vision System", Cambridge, MA., MIT Press, 1981

[4] Moravec, H. "The Stanford Cart and the CMU Rover" Proc. of IEEE, Vol .71,No.7,1983 pp.872-884

[5] Nevatia, R. "Depth Measurement by Motion tereo" Comput. Craphics & Image Proc, ol.G, 1976.

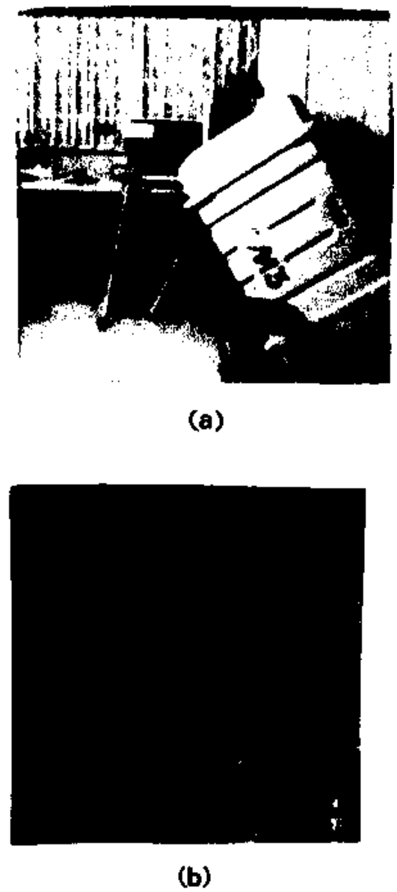

Figure 3