DISPARITY ESTIMATION BY PATCH MATCHING

Fatih M. Porikli and Yao Wang
Electrical Engineering Department
Polytechnic University
Brooklyn, NY, 11201

Abstract

In this paper, a stripe-mesh type image representation scheme and a disparity estimation algorithm based on the stripe-mesh structure are proposed. The mesh structure consists of stripes that are composed of triangular patches, and lay along the epipolar direction. The borders of the patches coincide with the depth discontinuities on the image plane. Thus, inside each patch the depth of the physical scene can be approximated by planar surfaces such that each surface is modeled by the disparity values of the patch corners. The stereo matching algorithm acquires the structure of the scene by estimating the disparities within the mesh patches, using a modified exhaustive search algorithm.

1 Introduction

Extracting the structure of a scene from stereo images has drawn significant attention in image processing and computer vision. Similar to the human visual perception, 3-D structure of a scene is derived by establishing correspondence between the points, representing the same 3-D point which is visible in the two views of the same scene. Disparity vectors, the relative imaging plane difference between the point correspondences, are estimated by image matching.

In terms of the density of the estimated disparity vectors for the image pair, majority of the stereo matching algorithms can roughly be classified as area-based and feature-based methods.

Area-based methods aim to assign a disparity value to every point in the image, and employ point characteristics such as intensity similarity or phase information as a matching exemplar [1]. Matching is evaluated in the neighborhood of each point as in regular block matching. Block matching algorithms approximate each local patch of disparity surface with a horizontal planar patch. This type of approximation, however, fails to consider the surface orientation of each local patch. In spite of its advantage of generating dense disparity map, area-based process is known to be inherently ambiguous and ill-posed. For instance, computation of disparity fails when matching is performed on edges that are parallel to the epipoles [3], [4]. As a result, several different smoothness constraints (either in the estimation stage or as a post-processing filter) have been appended to overcome ill-conditioning, but over-exposing smoothness constraints also flattens disparities at discontinuities [6]. Some robust statistical methods offer powerful alternative to smoothness and regularization to lessen the effects of modeling errors [9]. Most area-based methods are sensitive to both noise and photometric non-linearities due to the space-varying characteristics of illumination. Histogram equalization is used to enhance the similarities between the images; yet, it cannot handle local intensity divergences. Furthermore, area-based analysis fails in the regions where texture is absent or occlusion occurs. Area-based methods that use phase difference or phase correlation are claimed to robust to mild shading and stable upon the geometric deformations that occur with perspective projections of 3-D scenes [5].

Unlike area-based methods, feature based methods assign disparities only to the feature points such as edges, corner points or zero crossings of the image. Feature-based analysis provides more precise positioning for the feature in individual images, and it can attain relatively higher accuracy for its 3-D correspondences. As features are extracted and described based on information gathered in a rather large region, they are less ambiguous than the point exemplars of the area-based methods, hence make the stereo matching more reliable than the former. The problem, however, is that as features extracted within a larger region, their shapes might be deformed by the effect of the perspective projection. Since feature correspondences should be based on high level invariant that is derived from global-wise information, feature-based analysis is usually sensitive to

the performance of segmentation and suffer from occlusion. Such methods require interpolation as well as some techniques for modeling of occlusion. In addition, the feature-based process may be confused by large local change in disparity. It is also very difficult to incorporate the smoothness constraint into the matching strategy, because it is most likely to be violated at the edges [8]. They have trouble in treating scenes containing transparency; as inherently the opacity of the scene is assumed, appropriate features are hard to obtain by conventional feature detectors, such as edge and corner detectors [7].

In this paper, we describe a mesh-based disparity estimation algorithm that uses patch matching to acquire disparity vectors for all the points in the scene. A mesh structure that fits the depth discontinuities is developed [2]. This mesh structure takes the advantage of area matching by calculating matching error criterion on rather large regions, i.e., patches, while imposing smoothness constraints by keeping the general structure of the mesh undeformed. Such patches are chosen that their borders overlap with the edges where the probability of having depth discontinuity is quite higher than the other regions of the image. This way, mesh structure prevents excessive smoothing of disparity field at the disparity discontinuities. The proposed mesh structure is composed of stripes along the epipolar lines, and later each stripe is divided into triangular patches. One rationale of using such striped-mesh is to apply uniqueness and epipolar line constraints more efficiently. Since all possible matches for a point constitute a line segment parallel to the epipoles, the set of all possible matches for an arbitrary patch will build up a stripe. This stripe has the length of maximum possible disparity value, and the width equals to the patch height perpendicular to the epipolar direction. Thus, application of uniqueness for the neighboring patches intuitively suggests using stripes. To simplify matching scheme, it is also desirable to generate patches such that their mutual estimation range forms a simple geometric shape. The rectangular stripes permit to attain this simplicity. Furthermore, the edges parallel to the epipoles are forced to coincide with the up and down sides of the patches, so that ill-conditioning caused by matching such edges is minimized.

The proposed disparity estimation algorithm takes the advantages of the stripe mesh to find the structure of a scene from stereo image pairs. After a preprocessing step which provides initial disparity values, disparity range and visible image

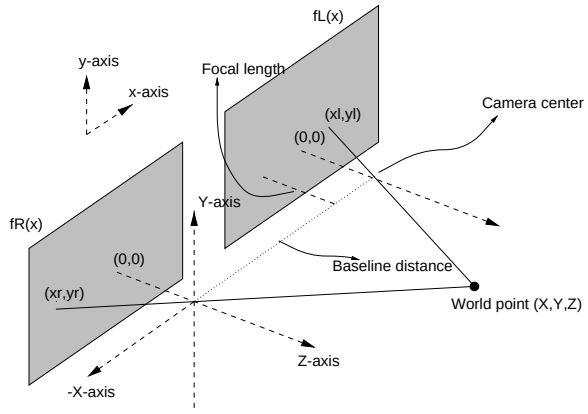

Figure 1: Stereo system

regions, the disparity values of the image points are estimated by matching the patches in the same stripe. For a patch corner, matching is done by minimizing squared-frame difference error in a window that consists of certain neighboring patches around the corner point. The minimum error that represents the best match is determined by using a modified exhaustive search algorithm. The disparity values of the subsequent stripes in the image are initialized by the disparity values obtained for the previous ones. The estimation is done in a hierarchical manner from coarse to finer resolution levels.

In the following section are provided the formulation of stereo imaging system and mesh structure. Section III describes patch matching algorithm. In Section IV, the results are discussed.

2 Formulation

2.1 Stereo System

The stereo-scopic camera system is centered in Cartesian coordinate system (X, Y, Z) and viewing direction is chosen as the Z-axis. Both of the camera image planes (x, y)'s are located normal to the Z-axis at a focal length, f, distance from the origin as shown in Fig.1. Let the baseline, B, of the stereo-scopic camera system be parallel to the X-axis, and image planes be coplanar, thus image coordinate plane's x and y axes are parallel to the X and Y axes respectively. Then, perspective projections of a 3-D point are

$$x_L = f\frac{X}{Z}, \qquad y_L = f\frac{Y}{Z}$$
$$x_R = f\frac{X+B}{Z}, \qquad y_R = f\frac{Y}{Z}.$$

From the above relations, we can derive the disparity, d, as

$$d = x_R - x_L = f\frac{B}{Z}$$

Let $f_R(\mathbf{x})$ and $f_L(\mathbf{x})$ be $M \times N$ left and right images. Then, f_R and f_L are related by

$$f_R(\mathbf{x}) = f_L(\mathbf{x} - d(\mathbf{x}))$$

where $d(\mathbf{x})$ is the disparity vector function. If we assume that the depth of the scene in each patch corresponds to a planar surface

$$Z = aX + bY + c$$

then, we obtain disparity for the same patch in terms of image plane coordinates (x_L, y_L) as

$$d = -\frac{B}{c}(ax_L + by_L - f)$$

which is a planar equation. In other words, if the depth of the physical scene corresponding to a mesh patch is a planar surface, the disparity values of the patch can be affine modeled. This enables us to estimate disparity by using simple affine functions while adequately modeling the 3-D structure on planar surfaces.

2.2 Mesh Generation

The stripe-mesh is constructed from the edge maps of the left image. The vertical and horizontal edge maps, $V(\mathbf{x})$ and $H(\mathbf{x})$, are obtained by using 3×3 sobel operators, and later, refined by a confident algorithm to enhance the line continuity. A sparse and global block matching is used to derive the maximum, minimum and average disparities within a predetermined range. For each row, an edge strength score, h_i, is calculated by adding $H(\mathbf{x})$ for $x = 1 \cdots M, y = i$. This score provides an insight on the total magnitudes of the edges that are oriented horizontally and intersect the corresponding row. The up and down borders of a stripe, s, are decided by choosing the biggest h_i's by simultaneously constraining them with a closeness and a minimum stripe width criteria. Within each stripe, the strength of each possible abutting line segments, l_j, is calculated similarly by projecting $H(\mathbf{x})$ and $V(\mathbf{x})$ on the possible line directions. The biggest l_j's, exceeding a strength threshold and representing the line segments that are sufficiently away from the previously selected ones, are chosen as quadrilateral patch borders. In the decision-making process, the selections causing irregular triangularization are also discarded.

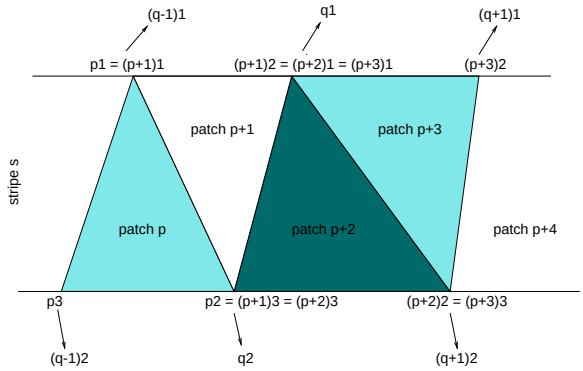

Figure 2: Patch indexing

To achieve vertical continuity of the line segments from stripe to stripe, l_j's are weighted by such a function that if there is a line segment chosen as patch border in the previous stripe which is close to the lines corresponding to l_j's in the present stripe, the l_j's are increased. This weighting forces patch borders in conjoint stripes to align as much as possible. Then each quadrilateral patch is divided into two triangle patches, p's. These patches form stripe-mesh $M(s, p_n)$ where s stands for stripe number, p stands for patch number, and n is one of the three corners, i.e. nodes. Note that either $M(s, p_1) = M(s, p + 1_1), M(s, p_2) = M(s, p + 1_3)$ or $M(s, p_2) = M(s, p+1_1), M(s, p_3) = M(s, p+1_3)$ holds for a regularly connected mesh structure as shown in Fig.2.

The nodes at which horizontal edges does not exist in a neighborhood are marked. Disparity field at such a node is supposed to be changing smoothly in vertical direction. Thus, when the estimation error is calculated, up and down neighboring patches will also be included in the matching window.

3 Matching Algorithm

After mesh is generated, the disparity values of the nodes are estimated one stripe at a time by minimizing frame difference errors. First, node disparity values are initialized. At each node, initial disparity value is obtained by block matching for 8×8 window size. This estimation supports disparity values only for node points without considering the disparity fluctuation inside the patches.

For horizontally continuous disparity estimation, mesh is transformed to a simpler, connected form M_c as $M(s, 2q - 1_1) \mapsto M_c(s, q_1)$ and $M(s, 2q - 1_3) \mapsto M_c(s, q_2)$ for $q = 1 \cdots \frac{p}{2}$. Then, q_1, q_2 are processed at each step. While anchoring the nodes $q - 1_1, q - 1_2, q + 1_1$, and $q + 1_2$ on the preceding and succeeding borders, matching er-

rors are calculated for all possible combination of disparity values for the current two nodes q_1, q_2. The disparity values of the points inside a patch are interpolated from the disparity values of the nodes. Matching error is computed over patches $p+1, p+2, p+3$ for node q_1, and over patches $p, p+1, p+2$ for node q_2. See Fig.3. The neighboring patches in the upper and lower stripes are also included if the current nodes are marked accordingly. Sum of squared frame difference is used as matching error measure,

$$E_{i=1,2} = \frac{1}{2} \sum_{p \in N_{q_i}} \sum_{x \in R_p} (f_L(x) - f_R(\mathbf{x} + g(\mathbf{x}, d(M(s, p_n)\mid_{n=1,2,3}))))^2$$

where N_q is the neighborhood of the corresponding q as given in Fig.3, R_p is the set which includes all the points inside the patch p. The function g is a linear interpolator, which derives disparity values of interior points from the disparity values of the nodes.

A penalty term is added in the matching error in order to prevent the estimated left and right borders from colliding or crossing. The disparity combination which gives the minimum error is chosen as the disparity estimates. While the following node pair is processed, previously estimated values are used instead of pre-initialized values. After minimization is done for the whole stripe, it is repeated to refine the results.

If continuity is to be dismissed, then matching is done for each triangular patch separately. The error terms for the neighboring patches are also included in the matching error by weighting these terms depending on the degree of the discontinuity.

The next stripe is initialized with the estimated values of the previous one. The upper nodes in the next stripe are initialized with the disparities at the image coordinates of these nodes, if the nodes are marked respectively.

The above algorithm is applied in a hierarchical manner from coarse to finer resolution levels. The disparities obtained in the previous resolution level are weighted and added to initial values obtained by block matching algorithm.

4 Experimental Results

The proposed algorithm has been tested with natural and artificial image pairs. Fig.4 shows a 320×240 left image with 32 pixels maximum disparity. As seen in Fig.5, the patch borders of the generated mesh match well with the edges of the original image. The signal-to-noise ratio's for

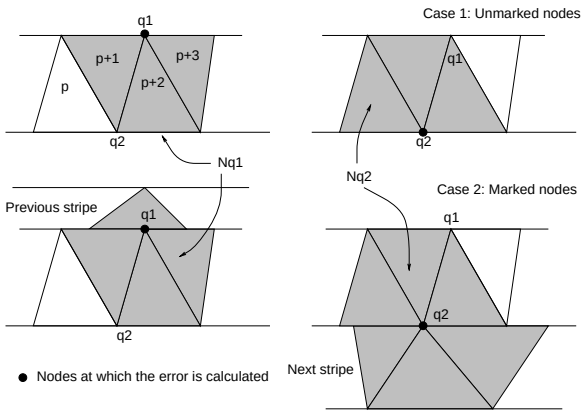

Figure 3: Matching regions for continuous disparity estimation

	SNR	PSNR
Originals	6.00	14.02
Input image size: 160×120		
8×8 Block matching	17.85	25.83
Patch matching (1^{st} ite.)	19.76	27.86
Patch matching (2^{nd} ite.)	20.31	28.42
Input image size: 320×240		
8×8 Block matching	19.63	27.65
Patch matching (1^{st} ite.)	19.07	27.00
Patch matching (2^{nd} ite.)	19.33	27.26

Table 1: Estimation errors

different resolution input images are presented in Table 1. The density of the patches and the total number of stripes can be increased by decreasing the strength thresholds and closeness distance. Although using dense meshes allows to approximate disparity more accurately in patches, it slows the calculation process. Besides, the probability of mismatch rises as the average patch region, hence spatial information entering the matching window decreases. The resultant depth map (Fig.7) is continuous both vertically and horizontally. In addition, it doesn't have major estimation errors as in block matching of texture-free image segments (Fig.6). The experiments show that the accuracy of resultant depth fields depends on the resolution of the stripe mesh (Compare Figs. 5-7 with Fig.8). Furthermore, the depth fields are more similar to the actual ones and less noisy than those obtained by the block matching.

Figure 4: Original 320×240 left image

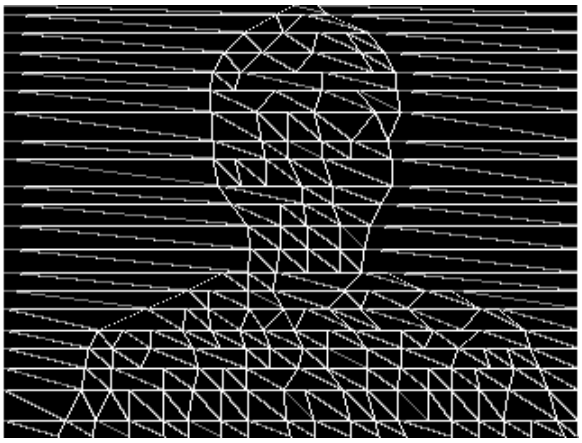

Figure 5: Generated stripe mesh

5 Discussion

In this paper, we present a disparity estimation algorithm that uses stripe mesh to improve the accuracy of matching process. Instead of using regular blocks, an image-dependent, edge-adaptive segmentation scheme is employed to obtain the matching windows. This way, estimation errors caused by matching blocks that contain discontinuous and non-constant disparity values as a result of occlusion and depth changes are minimized. A mesh is generated from the image pair such that borders (line segments connecting mesh nodes) overlap with the edges of the base image. Since disparity space is bounded by the epipolar line constraint, for the stereo system, the mesh nodes are organized to form stripes along the epipolar direction. Using stripes eases the task of checking uniqueness constraint that states a point can only have one correspondence in the other image, and eliminates interfering with the previously processed patches in the estimation algorithm. Furthermore, allocation of nodes as proposed achieves a good trade-off between the accuracy of the estimation and the number of mesh nodes or mesh patches.

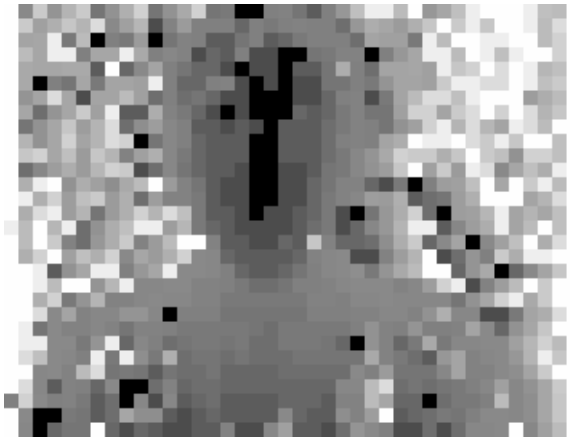

Figure 6: Estimated depth map by 8×8 block matching

Figure 7: Estimated depth map by patch matching

References

[1] S. D. Cochran and G. Medoni, "3-D surface description from binocular stereo," *IEEE Trans. Pattern Anal. Machine Intell.*, vol 14, no. 10, October 1992.

[2] F.M. Porikli, Y. Wang and C. Swain, "Adaptive stripe based patch matching for depth estimation," *ICASSP97*, Munich, April 1997.

[3] H. K. Nishihara, "PRISM: A practical real-time imaging stereo matcher," *Opt. Eng*, vol. 23, no. 5, 1984.

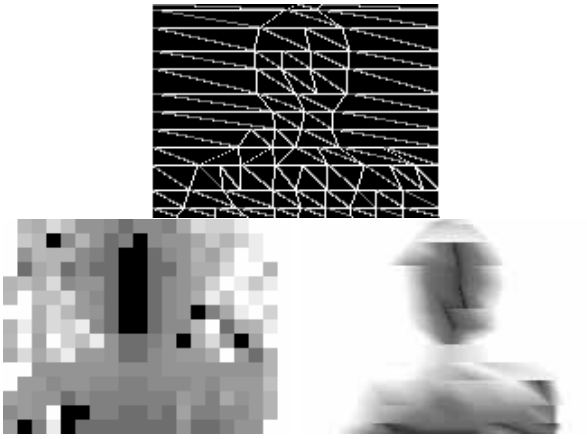

Figure 8: Generated mesh, estimated depth maps by block matching and by patch matching for 160× 120 input images

[4] E. Grosso, G. Sandini and M. Tistarelli, "3D object reconstruction using stereo and motion," *IEEE Trans. Systems, Man. and Cyber.*, vol. 19, no. 6, 1989.

[5] D. J. Fleet and A. D. Jepson, "Stability of phase information," *IEEE Trans. Anal. Mach. Intell.*, vol. 15, no. 12, 1991.

[6] D. V. Papadimitrio and T. J. Dennis, "Nonlinear smoothing of stereo disparity maps" *Electronic Letters,* vol. 30, no. 5, 1994.

[7] M. Shizawa, "Direct estimation of multiple disparities for transparent multiple surfaces in binocuolar stereo", *ICCV*, Berlin, 1993.

[8] X. Lin, Z. Zhu and W. Deng, "A stereo matching algorithm based on shape similarity for indoor environment model building", *ICRA*, April 1996.

[9] X. Zhuang and Y. Huang, "Robust 3D-3D pose estimation", *IEEE Trans. Anal. Mach. Intell.*, vol. 16, no. 8, 1994.