3D RECONSTRUCTION BASED ON A HYBRID DISPARITY ESTIMATION ALGORITHM

Xiaodong Huang and Eric Dubois

School of Information Technology and Engineering (SITE)
University of Ottawa
Ottawa, ON, K1N 6N5 Canada
Email: {xhuang, edubois}@site.uottawa.ca

ABSTRACT
This paper presents a hybrid disparity estimation algorithm which combines the pixel-based and region-based approaches. In the pixel-based approach, we use the Gabor transform and variational refinement, and the resulting disparities are combined with region information obtained from image segmentation so that a region matching scheme can be further applied. It will be shown with 3D reconstruction that our hybrid algorithm can give disparities with very high quality, in which some standard problems in disparity estimation can be solved.

Index Terms—Transforms, stereo vision, modeling, rendering

1. INTRODUCTION
3D reconstruction from the depth information obtained by disparity estimation on stereo images is a long standing problem. Due to its ill-posed nature, it still remains as a challenging issue with some typical problems unsolved like the matching of untextured and slanted surfaces, objects with complex geometries like trees and shrubs with sky as background, etc.

Most existing disparity estimation algorithms are pixel-based, in which the disparities are estimated pixel-by-pixel. This is one common feature for various otherwise quite different disparity estimation algorithms. While this approach is convenient, it does not make use of the information associated with the shape of areas in the images. The main problem associated with pixel-based approaches is that they cannot effectively handle untextured surfaces. Usually the estimated disparities on such surfaces are noisy. Although regularization has been introduced to deal with this problem, e.g. [1], it still cannot solve the issue completely. On the other hand, using too much regularization will also bring distortions to object boundaries. Thus, regularization cannot be used as the only solution to increase the performance of disparity estimation. Compared to the large number of papers on pixel-based disparity estimation algorithms, there are only a few dealing with region-based disparity estimation [2, 3]. In [2], the mean shift segmentation algorithm developed in [4] was used to segment the images into different regions. However, in the next steps, oversegmentation was applied to each region rather than a direct region matching. Alternatively we can suppose that each region in one image of the stereo pair can be considered as an affine transform from the same region in another image, and the region-based disparity estimation is thus converted to the estimation of the affine parameters for each region.

In this paper, we present a hybrid disparity estimation algorithm which combines the pixel-based and region-based approaches. A pixel-based approach based on the Gabor transform and variational regularization is performed first, and then the region information from the segmentation by mean shift is combined with the pixel-based disparity results so that a region matching scheme using affine transform can be applied. A novel contribution is a method to analyze the changes of disparity histograms after region matching, so that the regions with occlusions can be detected and the true disparities for such regions can be determined. The high quality of our results will be shown by both disparity maps and 3D reconstructions.

2. SYSTEM ARCHITECTURE
As shown in Fig. 1, our system starts by filtering the left image \( I_l \) and the right image \( I_r \) of a pair of stereo images with a set of Gabor filters. Also \( I_l \) is put through a segmentation process using the mean shift algorithm in which each region is formed by grouping pixels with similar color values and is represented by one color value for this region. The filter outputs of \( I_l \) and \( I_r \) are compared and a coarse disparity map is estimated. Then a variational regularization using an edge-preserving functional is applied on this coarse disparity map as a refinement process. After variational refinement, the disparity values in each region of \( I_l \) (obtained from the segmentation) are used to estimate a set of affine transform parameters by least squares, so that the matching relation for the pixels in this region with their corresponding pixels in \( I_r \) can be represented by the resulting affine transform. The affine parameters for each region are further adjusted using a descent-based region matching technique, and these adjusted affine parameters can be used in turn to calculate a more refined disparity map for 3D reconstruction.
3. PIXEL-BASED DISPARITY ESTIMATION

In this section, we briefly introduce the pixel-based disparity estimation based on the Gabor transform, as well as the refinement result using variational regularization. The detailed information can be found in [5].

We used a set of quadrature-pair Gabor filters tuned to different directions with different frequencies to filter $I_l$ and $I_r$, and the filter coefficients for the two images are compared for a coarse disparity estimation. In our experiments, we used two standard stereo image sets – Tsukuba and Flower Garden. The original left images of Tsukuba and Flower Garden are shown in Fig. 2(a) and Fig. 3(a) respectively, while their coarse disparity maps $d(x, y)$ estimated by the method in [5] are shown in Fig. 2(b) and Fig. 3(b) respectively.

The refinement functional including regularization function, as well as the procedure to solve its associated Euler-Lagrange equation iteratively, can be found in [5]. The $d(x, y)$ for Tsukuba and Flower Garden after variational refinement are shown in Fig. 4(a) and (b) respectively.

4. REGION MATCHING

4.1. Segmentation of $I_l$

We applied the mean shift segmentation algorithm [4] to the image $I_l$, and the segmentation results for the above two images are shown in Fig. 5. Each region is indicated by one color value. Comparing Fig. 5(b) with its original in Fig. 3(a), we can find that the mean shift algorithm could not identify some tiny features, which are missing after segmentation (e.g., some twigs on the tree, and part of the shrubs). To alleviate such a problem, we have performed an edge-detection by Canny detector on $I_l$ and on Fig. 5(b), and then compare the detected edges between the two images to pick out the missing tiny contours. The new segmentation result for Flower Garden is shown in Fig. 6. Although we get most of the missing tiny contours back, this method also introduces some extra contours on some existing regions.

4.2. Representing Disparity by Affine Transform

We assume that the coordinates $(x, y)^T$ of each pixel in a region in $I_l$ is related to its corresponding pixel $(x_r, y_r)^T$ in $I_r$ by an affine transform. We found that most of the regions segmented by mean shift can be seen as planar patches, so the choice of affine transform is reasonable. In this case of parallel stereo without vertical displacement ($y = y_r$), we have:

$$x_r = a_{11}x + a_{12}y + a_{13}.$$  

(a)\(\hspace{8cm}\) (b)

Fig. 2. (a) Original left image of Tsukuba. (b) The coarse disparity estimated by Gabor transform.

(a)\(\hspace{8cm}\) (b)

Fig. 3. (a) Original left image of Flower Garden. (b) The coarse disparity estimated by Gabor transform.

(a)\(\hspace{8cm}\) (b)

Fig. 4. Disparities after variational refinement: (a) Tsukuba. (b) Flower Garden.

(a)\(\hspace{8cm}\) (b)

Fig. 5. Segmentation by mean shift: (a) Tsukuba. (b) Flower Garden.

(a)\(\hspace{8cm}\) (b)

Fig. 6. (a) New segmentation result for Flower Garden after edge-detection by Canny detector. (b) Original image for Flower Garden.

4.2. Representing Disparity by Affine Transform

We assume that the coordinates $(x, y)^T$ of each pixel in a region in $I_l$ is related to its corresponding pixel $(x_r, y_r)^T$ in $I_r$ by an affine transform. We found that most of the regions segmented by mean shift can be seen as planar patches, so the choice of affine transform is reasonable. In this case of parallel stereo without vertical displacement ($y = y_r$), we have:

$$x_r = a_{11}x + a_{12}y + a_{13}.$$  

1026
Therefore, the disparity \(d(x, y)\) is related to these affine parameters by
\[
d(x, y) = x - a_{11}x - a_{12}y - a_{13}.
\]
(2)
Thus, the estimated \(d(x, y)\) for each pixel in one region from the previous variational refinement can be grouped and used as known variables so that the affine parameters can be estimated from (2). Since each pixel in the region gives a set of equations as in (2), and for most of the cases, the number of pixels in a region is larger than the number of affine parameters (three for 1-D affine transform), the estimation of the three parameters \((a_{11} \sim a_{13})\) can be done by least squares, implemented using singular value decomposition (SVD). Then, once the affine parameters are estimated, a new disparity \(d(x, y)\) for each pixel in the region can be in turn calculated by (2).

The new results for Tsukuba and Flower Garden from the above procedure are shown in Fig. 7. We can see that this kind of parameterized estimation process can give more reasonable results in which the noise in each region is somewhat removed, but non-smoothness exists among some adjacent regions. To solve this problem, we use region matching to improve the affine parameters.

**4.3. Further Refinement by Region Matching**

\[
E(\Delta \hat{a}) = \sum_{(x,y) \in W_i} [\psi^T \Delta \hat{a} - D]^2
\]
(4)
where \(\psi = I_r(x, y)X\) and \(D = I_l(x, y) - I_r(\hat{a}^T X, y)\). The iterative solution of (4) by least squares is [6]:
\[
\Delta \hat{a} = [\sum_{(x,y) \in W_i} \psi \psi^T]^{-1} \sum_{(x,y) \in W_i} D \psi.
\]
(5)
The resulting disparities obtained from the new affine parameters updated using (5) are shown in Fig. 8. Compared with Fig. 7, the new disparities have improvements for most areas. However, some regions where there are occlusions give worse effects than the same areas in Fig. 7, e.g., the sky areas with twigs and shrubs as foreground objects. This is because the minimization of (4) through (5) is trying to minimize the squared intensity difference between all the pixels (including occluded pixels) in a region of \(I_l\) with their corresponding pixels in \(I_r\). Thus the region matching technique can make such regions shift along with their foreground objects. This effect can be detected by comparing patterns in the disparity histogram for a region before and after region matching, and the correct disparities can thus be determined from these patterns. For example, as shown in Fig. 9, the disparity histogram for a sky region (between a twig and the tree) has one peak near zero value before region matching, and after region matching there are two peaks. The second peak with a higher value comes out from those occluded pixels near foreground objects. Also, since most of the regions we are dealing with are smooth regions, there is high possibility that the regions

**Fig. 7.** New results by applying the affine parameters to the calculation of the disparities for each region: (a) Tsukuba; (b) Flower Garden.

The error function that we need to minimize for each region is:
\[
E = \sum_{(x,y) \in W_i} [I_r(a_{11}x + a_{12}y + a_{13}, y) - I_l(x, y)]^2
\]
(3)
where \(W_i\) represents a region. We need to minimize (3) by updating affine parameters \(a = [a_{11}, a_{12}, a_{13}]^T\) iteratively using least squares with Taylor expansion. Assume \(X = [x, y, 1]^T\). Let \(\hat{a}\) be the current estimate of affine parameters, and \(a = \hat{a} + \Delta \hat{a}\). Then expand \(I_r\) around the current estimate \(I_r(a^T X, y) \approx I_r(\hat{a}^T X, y) + \Delta \hat{a}^T X I_r(x, y)\), and this first order expansion is valid only when \(\hat{a}\) is close to \(a\). This is the reason that we start region matching with the result from pixel-based approach, rather than doing it from very beginning without pixel-based results. Substituting the above first order expansion into (3), the error function becomes:

**Fig. 8.** New disparities after region matching: (a) Tsukuba; (b) Flower Garden.
with two peaks in histogram have wrong disparity values. Once such pattern changes have been detected, the real disparity values for such regions can be determined by the lower peak in Fig. 9(a), and verified if the *mean absolute difference* between $I_l$ and $I_r$ for that region (excluding those occluded pixels which are detected from the technique in [5]) is less or equal to its value by using Fig. 9(b). Fig. 10 shows the final results after making use of such pattern detections, where we can see some regions containing occlusions, like the regions left of the lamp and of the neck-of-head in *Tsukuba*, and the sky regions in *Flower Garden*, have been identified and the correct disparities have been assigned.

![Histograms](image)

**Fig. 9.** Histogram change for a sky region of *Flower Garden*: (a) before region matching; (b) after region matching.

The 3D models based on disparities in Fig. 10 can be set up using OpenGL, and novel views can be rendered, as shown in Fig. 11 for *Flower Garden* which contains more complex geometric structures. Fig. 11(a) is the direct reconstruction for the original image from the model. Fig. 11(b) is the novel reconstruction by rotating $35^\circ$ from the original viewpoint, where we can see that the sky has much more shifting than the foreground scenes since it has the largest depth, and the occlusion from the foreground trees and shrubs on the sky can be clearly seen (black areas). Also, the occlusion from the tree (with supporting frames) right in front of the house can be identified from the splitting effect on the house in that area, which means even the tiny depth difference between that tree and the house is detected (otherwise that tree should “stick” to the house and rotate together). All these facts indicate that the complex geometric structures detected by our algorithm are largely correct.

![3D reconstruction](image)

**Fig. 11.** 3D reconstruction: (a) original; (b) rotation with $35^\circ$.

### 5. CONCLUSION

We developed a hybrid disparity estimation algorithm which combines pixel-based and region-based approaches. The novelty of our algorithm lies in the fact that it provides a robust method to solve some long standing issues in disparity estimation, like the smoothness of surfaces while keeping object boundaries sharp and clear, and the identification of occluding regions to recover their true disparities by analyzing the histograms from pixel-based and region-based approaches. These problems cannot be solved by either approach separately.

### 6. ACKNOWLEDGEMENTS

This work was supported by the Natural Sciences and Engineering Research Council of Canada under the research network *LORNET*.

### 7. REFERENCES


