Chapter 4

Stereo Correspondence of Face Images Using Adaptive Support Weights
Chapter 4

Stereo Correspondence of Face Images Using Adaptive Support Weights

The stereo correspondences of human faces are often very difficult to achieve because of uniform texture, slow changes in depth and occlusion. In this chapter, we introduce an adaptive weight based stereo correspondence method for face images. We estimate the support weights of the pixels in a given support window based on a color similarity and proximity to reduce the fattening effect. The experiments are carried out on stereo images of face database and Middlebury database. The experimental results show that the proposed algorithm produces a smooth disparity map while preserving sharp depth discontinuities accurately.

4.1 Introduction

The color segment based stereo correspondence for face images using self adaptive dissimilarities in frequency domain is introduced in chapter 3. This method solves the problem of obtaining accurate results at depth discontinuities at boundaries and in textureless region. This method also
resolves the problem in finding the optimal support window with an arbitrary shapes and sizes in low textured images. But it is very difficult to dealing with highly textured images. The adaptive weights for stereo correspondence method tries to overcome the drawbacks of traditional window based method. The use of adaptive weight in segmentation based stereo correspondence enforce smoothness over textured planes as well as handles the depth discontinuities very efficiently when compared to method discussed in chapter 3. This method decreases the number of artifacts along textured region which lie at same depth.

Repetitive textured regions and regions with depth discontinuity fail to match correctly and this result in the "foreground-fattening" phenomenon. To obtain accurate results at depth discontinuities and also in homogeneous regions, an appropriate support weight should be assigned for each pixel adaptively. Some of the methods (Yoon et al., 2006; Hosni et al., 2009; Gu et al., 2008; Yoon et al., 2005) will try to assign appropriate support weights to the pixels in a support window while fixing the shape and size of a local support window. Yoon and Kweon (2006) finds the adaptive support weights to the pixels in a support window whose size and shape are usually fixed. These are assigned to pixels in the window, based on both color proximity and geometric distance to the center pixel of interest. This method gives good matching results at discontinuities and insensitive to brightness variation. Gu et al.(2008) introduced an algorithm which adopted both an adaptive support weight and a rank transform method to acquire an initial disparity map. The adaptive weights are calculated for each pixel using color similarity and geometric proximity.

Li et al. (2011) proposed an adaptive cost aggregation strategy based on a generalized bilateral filter model. The modified truncated L1 norm is used as initial cost function to improve robustness to noise and outliers. The final disparity map is obtained by a winner take all strategy. Psota et al. introduced stereo matching algorithm that performs iterative refinement on the results of adaptive support weight stereo matching. which uses
the set of adaptive support weights for both cost computation and iterative refinement of the disparity map. The iterative refinement method assigns a cost penalty whenever a pixel's disparity differs from the disparity of its neighboring pixels, where the level of similarity is measured using adaptive support weights. Thus, the adaptive support weights computed for matching costs are reused to determine the magnitude of the iterative refinement cost. The disparity penalty cost is added to the original matching costs at each iteration, and a combination of both dynamic programming and winner-take-all strategies is used to determine reliable matches.

In this chapter, we introduce an adaptive support aggregation strategy which deploys segmentation information in order to increase the reliability of stereo correspondence method for face images to get accurate results at depth discontinuities as well as in homogeneous regions. The flow diagram of proposed method is shown in Figure 4.2. The images are first rectified using uncalibrated rectification method (Fusiello et al., 2008) to suppress the vertical displacement. The support weights of the pixels are estimated in a given support window based on a color similarity and proximity to reduce the fattening effect. The reference image is segmented using mean shift segmentation method (Comaniciu et al., 2002). The self-adapting dissimilarity measure (Klaus et al., 2006) is used to estimate the initial disparity and then dissimilarity is computed using the raw matching costs and support weights. The proposed method is composed of three parts: adaptive support weight computation, dissimilarity computation based on the support weights and finally, disparity selection using winner take all algorithm.

The organization of the chapter is as follows: The stereo correspondence method is presented in section 4.2. The experimental results are illustrated in section 4.3. Finally, section 4.4 draws a chapter summary.
4.2 Stereo correspondence

4.2.1 Adaptive Support Weights Estimation

The difference between pixel colors is measured in the CIELab color space, mainly because it strongly correlates with human color discrimination and has perceptually meaningful measure of color similarity (Yoon et al., 2005). As the distance in the CIELab color space between two points increases, it is reasonable to assume that the perceived color difference between the stimuli that the two points represent increases accordingly. Especially, short
Euclidean distances correlate strongly with the human color discrimination performance. The proximity weights are assigned to the pixels based on their color difference and proximity. When $C_{pq}$ represents the euclidean distance between two colors, $c_p = [L_p, a_p, b_p]$ and $c_q = [L_q, a_q, b_q]$ in the CIELab color space. The color similarity of $p$ and $q$ pixel is defined as

$$\Delta C_{pq} = \sqrt{(L_p - L_q)^2 + (a_p - a_q)^2 + (b_p - b_q)^2},$$  \hspace{1cm} (4.1)$$

$$f_s(\Delta C_{pq}) = \exp\left(-\frac{\Delta C_{pq}}{\gamma_c}\right).$$  \hspace{1cm} (4.2)$$

The proximity between $p(x_p, y_p)$ and $q(x_q, y_q)$ is defined as

$$\Delta g_{pq} = \sqrt{(x_p - x_q)^2 + (y_p - y_q)^2},$$  \hspace{1cm} (4.3)$$

According to the gestalt principle of proximity, the support weight of a pixel decreases as the spatial distance to the pixel under consideration increases. The small spatial distances strongly correlate with the human discrimination performance. The strength of grouping by proximity is defined using the Laplacian kernel as follows

$$f_s(\Delta g_{pq}) = \exp\left(-\frac{\Delta g_{pq}}{\gamma_p}\right).$$  \hspace{1cm} (4.4)$$

The support weight of a pixel expressed as

$$w(p, q) = \exp\left(-\left(\frac{\Delta C_{pq}}{\gamma_c} + \frac{\Delta g_{pq}}{\gamma_p}\right)\right),$$  \hspace{1cm} (4.5)$$

where $\gamma_c$ and $\gamma_p$ are user defined parameter.

### 4.2.2 Estimation of Initial Disparity

To estimate the initial disparity of reference image, the rectified reference image is first divided into regions of homogeneous color using mean shift segmentation. The color image has three channels: red, green and blue.
The blue channel of an image is used to represent intensities between 0 and 255. Then disparity map is estimated by considering sum of squared difference of intensity and gradient of each pixel of stereo pairs. The dissimilarity measure based on the gradient fields and intensities of the images yields good accuracy in the resulting disparity map when compared to classical intensity based method. The main reason for this accuracy enhancement is the better robustness of the gradient against differences in sampling and local brightness changes between cameras. That is defined as

$$SSD(p, q) = \sum_{p, q \in W(x, y)} (b_1(p) - b_2(q))^2,$$

(4.6)

$$Grad(p, q) = \sum_{p, q \in W_x(x, y)} |\nabla_x b_1(p) - \nabla_x b_2(q)|^2 + \sum_{p, q \in W_y(x, y)} |\nabla_y b_1(p) - \nabla_y b_2(q)|^2,$$

(4.7)

where $W$ is window size. The color weighting makes the match scores less sensitive to occlusion boundaries by using the fact that occlusion boundaries most often cause color discontinuities as well. The left to right cross checking is used for occlusion handling and to find reliable correspondences. The resulting dissimilarity measure is given by

$$C(p, q) = SSD(p, q) + Grad(p, q).$$

(4.8)

### 4.2.3 Adaptive Support-Weight Stereo Matching

The dissimilarity between pixels is then measured by aggregating the raw matching costs with the support-weights in the support windows. We considered the support weights of both the reference and target support windows. When only considering the weights of the reference support window, the dissimilarity measure may be erroneous because the target support window may have pixels from different depth. In order to minimize the effect of such pixels, we compute the dissimilarity between pixels by combining the
support weights of both support windows. The combined support weights encourage the points that are likely to have similar disparities with the centered pixels in both images. The disparity between pixel $q$ and $q_d$ can be expressed as

$$D(q, q_d) = \sum w(p, q)w(p', q_d) C(q, q_d).$$  (4.9)

After the dissimilarity computation, winner takes all optimization is used to choose the disparity of a pixel with lowest matching cost.

$$D_p = \min_{d \in S_d} D(p, q),$$  (4.10)

where $S_d = \{d_{\text{min}} \cdots d_{\text{max}}\}$ is a set of all possible disparity values. In our method, instead of assigning disparity value to each pixel, the disparity plane is estimated by using disparities of each segment. The estimated disparity plane is assigned to each segment using least square plane fitting method (Chapter 3).

4.3 Experimental Results

We have conducted experiments to evaluate the performance of the proposed stereo correspondence algorithm on the test images of stereo face database and middlebury stereo database. The ground truth disparity map of faces of stereo face database is not available, so that we applied our proposed method to some of the images of Middlebury stereo database to estimate Root Mean Square Error (RMSE) in order to evaluate the performance.

4.3.1 Experiments on Stereo Face Images

To evaluate the performance of the proposed stereo correspondence algorithm, we have carried out experiments on test images of stereo face
database. The stereo images are first rectified using uncalibrated rectification to suppress the vertical displacement. The rectified images are cropped to extract face region by detecting nose tip. The cropped face region of reference image (right image) is segmented into homogeneous region using mean shift segmentation method. The results of adaptive weights based stereo correspondence technique is compared with segment based approach (Chapter 3). The disparity map obtained from adaptive weights based method is shown in second column of Figure 4.2 and the disparity map produced by proposed method in Chapter 3 is shown in third column of Figure 4.2. The results show that the use of adaptive weights in segmentation based stereo correspondence enforce smoothness over textured planes as well as handles the depth discontinuities very efficiently and decreases the number of artifacts along textured region which lie at same depth when compared to other stereo correspondence method discussed in chapter 3.

We have conducted experiments by varying the window size $3 \times 3$, $5 \times 5$, $7 \times 7$ and $9 \times 9$ to show the efficiency of our approach. From Figure 4.3, it is observed that even after varying the window size the small variation in the disparity map.

**Pose Variation**

The face images suffer from pose variation due to rotation of head. The images taken at two different viewpoints of the same subject may appear more different than two images taken from the same view point for two different subjects. We have carried out experiment on face images of each subject with eight different poses of stereo database to evaluate effects of our approach. There is a pose variation in left as well as right images. Figure 4.4 shows the stereo face images with pose variation and the corresponding depth map computed using our approach.

**Illumination Variation**

In order to capture the faces in controlled environment is very difficult. The illumination problem arises, when the same face appears differently due to the change in lighting variation. We have conducted experiments on stereo
face database with varying illumination. To show the effectiveness of our approach under illumination variation, the illumination of left image is varied using Adobe Photoshop software and right image is kept unaltered. The Figure 4.5 shows the pair of face images of stereo database used in the experiments with illumination variation and the corresponding estimated disparity map using our approach.
Figure 4.3: Disparity Map computed using our approach with different window size: First column - $3 \times 3$; second column - $5 \times 5$; third column - $7 \times 7$; fourth column - $9 \times 9$.

4.4 Experiments on Middlebury database

In order to evaluate the result of our algorithm with ground truth, the experiments are performed on some of the stereo test images from the Middlebury stereo database. We have conducted experiments on some of the stereo images of middlebury database such as Teddy and Cones (First Column of Figure 4.6). The second and third column of Figure 4.6 shows the corresponding image ground truth and result of proposed method. Root
mean square (RMS) method is a quantitative way to estimate the quality of the computed correspondences. The Table 4.1 shows the root mean square error estimated for results obtained by proposed method and SAD, SSD and NCC algorithms. The root mean square error of SAD, SSD and NCC algorithms are more than root mean square error of proposed method.

### 4.5 Observations

From all the experimental results, the following observations were made:

- The proposed stereo correspondence approach effectively produces smooth disparity maps while preserving depth discontinuities well compared to conventional correlation measures for stereo face images.
- The color-weighted based stereo correspondence approach yields improved treatment of radiometric distorted underwater images compared to directly using radiometric insensitive match measures on gray-scale images.
- The proposed method works well for images with little image noise, it may produce an erroneous result when there is severe image noise because the color difference used for the support-weight computation is measured by using an individual pixel color.
- Our approach does not depend on the initial disparity estimation, because the adaptive support-weight is computed non-iteratively based on the contextual information within a given support window.
• The proposed method, however, is computationally a little more expensive than other correlation methods for the pixel-wise adaptive support-weight computation step.

4.6 Chapter Summary

In this chapter, we introduce an adaptive weight-based cross-correlation stereo correspondence approach to estimate disparity map for face color stereo images. The experiments are conducted on stereo face images, which are captured in different directions. The rectified color stereo face images are cropped to have only a face region and then rectified. The cropped image pairs are used to estimate the disparity map using adaptive weight-based stereo correspondence method. The adaptive support weight of the pixel in a support window is computed by measuring the strength of grouping by geometric proximity and color similarity. The experiments are conducted to show the effectiveness of our approach for variation in pose and illumination. The experimental result shows that our approach yields a smooth disparity map which is insensitive to illumination variation and reduces the depth discontinuities at borders for images. The visual comparison shows that adaptive weight based approach outperforms the other two proposed methods.
Figure 4.4: Disparity Map computed using our approach for images with pose variation: first row of (a), (b), and (c) - left images with pose variation, second row of (a), (b) and (c) - right images with pose variation, third row of (a), (b) and (c) - obtained disparity map.
Figure 4.5: Disparity Map computed using our approach for images with illumination variation; first column - left image with illumination variation, second column - right image, third column - obtained disparity map.
Figure 4.6: Results on Middlebury datasets. From top to bottom: Teddy, Cones. From left to right: reference images, ground truth disparities, the results of the proposed algorithm.