CHAPTER 3

DISPARITY AND DEPTH MAP COMPUTATION

In this chapter we will discuss the process of disparity computation. It plays an important role in our caricature system because all 3D coordinates of nodes on facial model will be recovered at this stage.

3.1 Depth map recovery by stereo vision

Stereo algorithms are usually divided into two categories [15]:

- Feature matching

  This kind of algorithm is implemented by extracting some features, such as segment of edge or contours, from each input image and then comparing the corresponding part in these images. It is quite fast since a small portion is used for matching, but the disparity map is often sparse.

- Correlation

  Choose a specific block in a certain image, and define a block with the size called "window" in another image. The process slides the window in corresponding image to find the block which holds a minimum difference of intensity with the specific one. The disadvantages are slow and a mismatch may occur in the area lacking salient features.

A stereo pair is two images with disparity. It simulates human's vision system to get depth information. The technology to reconstruct stereo scene from stereo pair is known as "binocular fusion"。

There are three stages involved in binocular fusion:

- Preprocessing

  In camera geometry, the specific configuration of each camera is described by its parameters. Intrinsic camera parameters represent the unique properties of each camera, such as focal length, central coordinate, valid pixels, and distortion coefficients, and extrinsic camera parameters show the relations between real world and camera, including rotation matrix and translation vectors. Our algorithm requires the images to be rectified before correlation because we assume the epipolar lines are horizontal. Therefore, we will calibrate cameras to estimate these parameters, and use the information to rectify the image pairs for future computation,

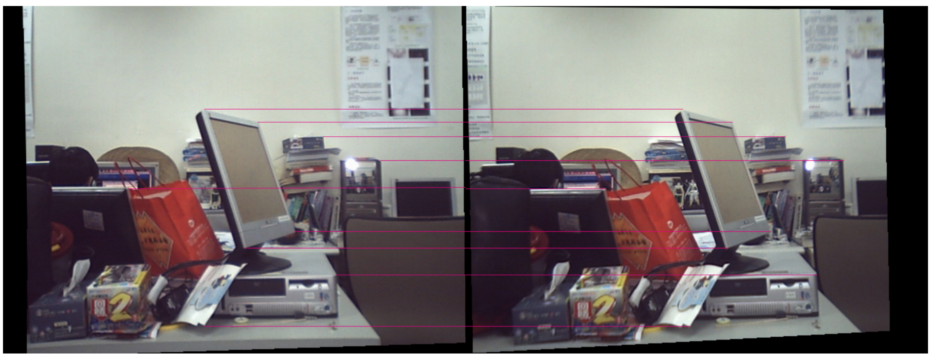


Fig. 3-1: A parallel image pair after undistortion and rectification

- Correspondence problem

  After preprocessing is complete, what we have to do is to match corresponding points between two images. In this thesis, we adopt a modified correlation-based method to find the correspondence.

- Depth map recovery

    If we find corresponding points between image pair, we can recover the depth information from their disparity. By observing human's vision system, images taken from left eye and right eye are not the same, and the position of the corresponding points in each image is a little different. The difference is called "disparity".

    As shown in Fig. 3-2, the farther the object is, the smaller the disparity will be. On the contrary, if the object is closer to the observer, the disparity will be larger.

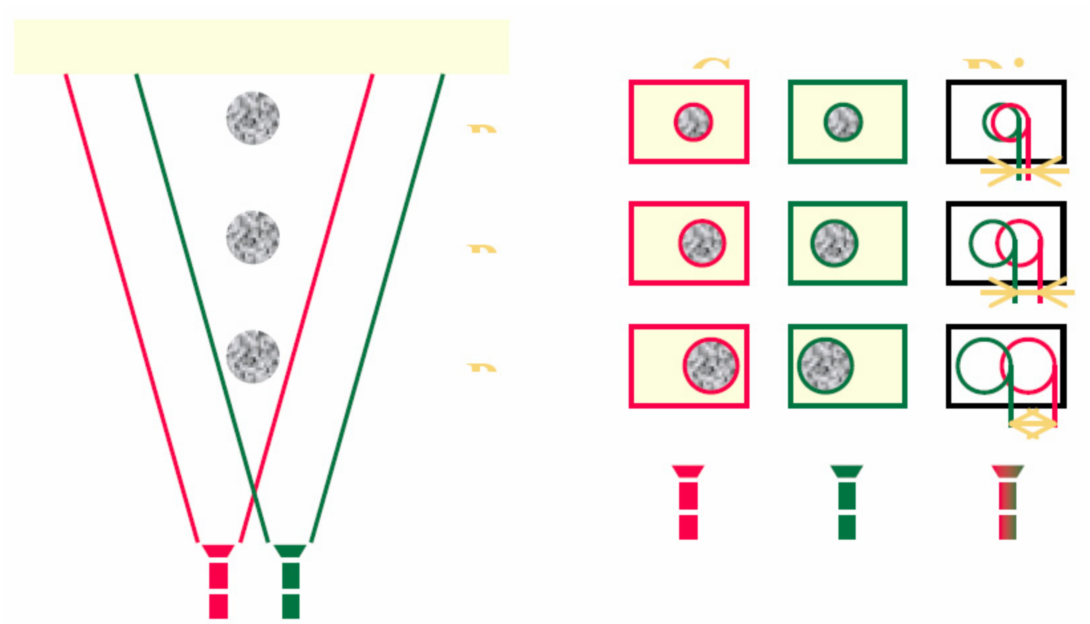


Fig. 3-2: Relation between depth and disparity

3.2 Disparity computation

In this thesis, we use intensity information for matching. Despite Muhlmann *et al.* [16] argued that incorporating color for matching will yield more precise result, it will not have a significant impact in our research as we are dealing with human faces, which usually have uniform colors. The method we adopt is introduced by Fua [17]. We define a score s to evaluate the similarities between two images.

$$s = \max(0, 1 - c) \tag{7}$$

$$c = \frac{\sum_{i,j}((I_1(x+i, y+j) - \overline{I_1}) - (I_2(x+dx+i, y+dy+j) - \overline{I_2}))^2}{\sqrt{(\sum_{i,j}(I_1(x+i, y+j) - \overline{I_2})^2)\sum_{i,j}(I_2(x+dx+i, y+dy+j) - \overline{I_2})^2)}} \tag{8}$$

dx, dy: displacement of x and y $\bar{I}$: mean of I

I_1 and I_2 represent the intensity of left image and right image. Since images are rectified, so points in the left image will be found on the same epipolar line in the right image. Therefore, we just have to slide the window alone the same height, and find the point with the maximum s.

In order to avoid mismatch, after left image is computed, we will take right one as the host image and repeat the matching process. We then preserve the consistent points in the two images by double checking. As a result, not every pixel has a corresponding point and the computed disparity map will be sparse. To solve this problem, we enlarge the sliding window in different size and repeat the search process. In general, a larger window size will result in a more dense disparity map, but the accuracy is not as good [18]. A multi-layer algorithm that is designed to overcome this problem will be discussed subsequently.

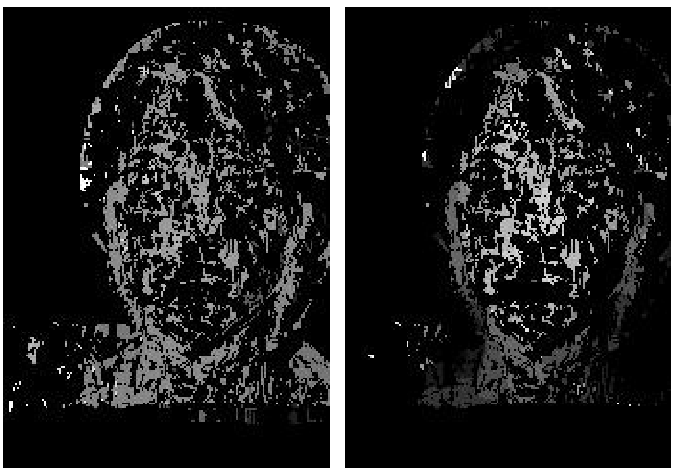


Fig. 3-3: (left) A disparity map computed by 7x7 window

(right) Result after correction

By observing the anatomy of human faces, point with maximum disparity value should be found near the nose tip, where the minimum should be close to the ears. The disparity range thus identified will serve as the reference for correcting the previously computed disparity map, as depicted in Fig.3-3.

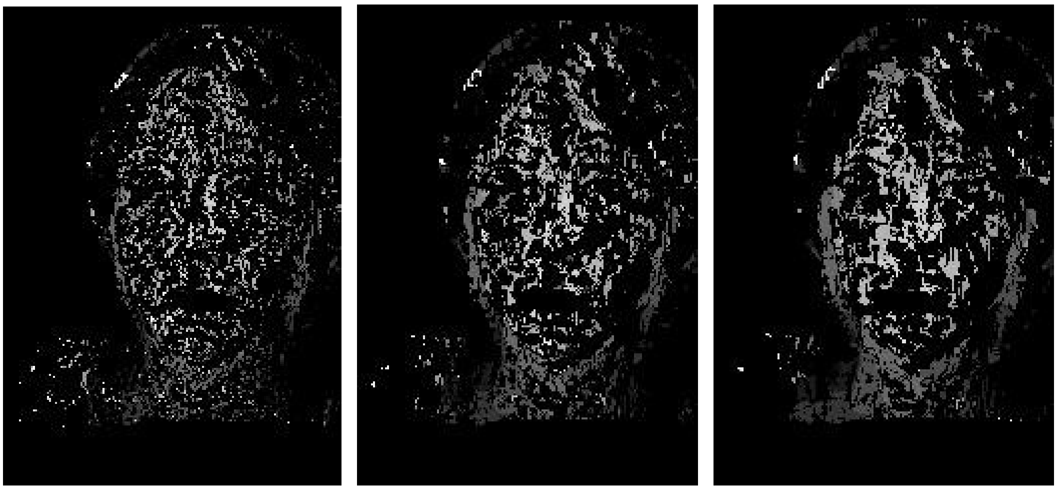


Fig. 3-4: Depth map obtained by different window size. From left to right - 3x3, 5x5, and 7x7.

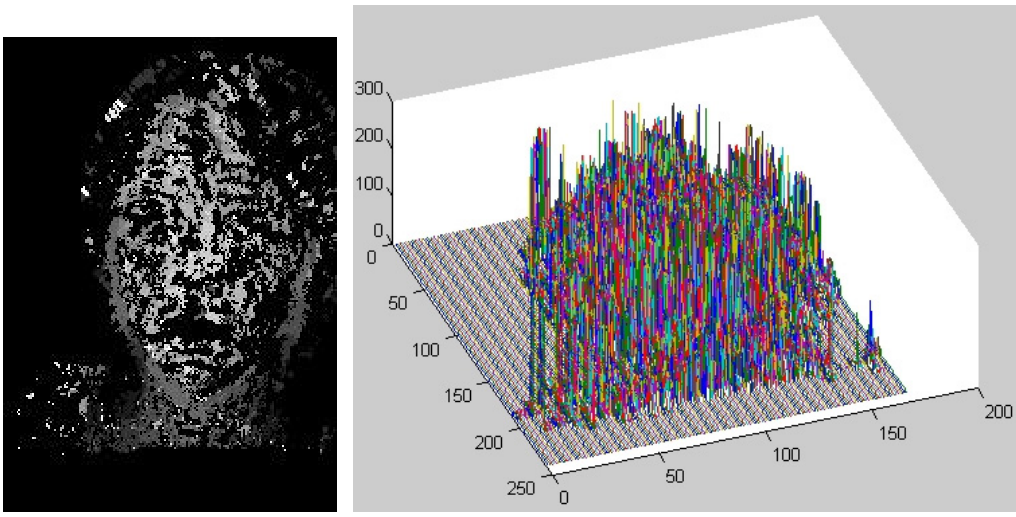
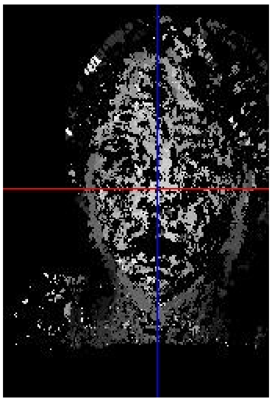
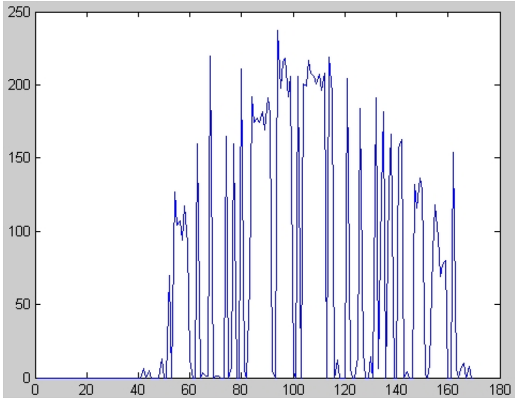


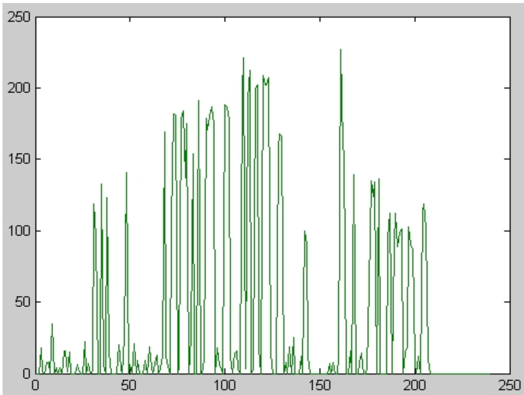
Fig 3-5: Result of multi-layer algorithm (left) and its histogram (right)



(a)



(b)



(c)

Fig 3-6: (b) Profile of the red line in (a)

(c) Profile of the blue line in (a)

At this stage, we can obtain a robust disparity map. However, a sparse depth map is not sufficient for modeling. Moreover, there is still some incorrect depth information. A strategy to compute a dense and accurate disparity map based on the feature nodes will be presented to address this issue.

As discussed previously, face components are automatically extracted by the AAM we built. Thus, we can choose a set of control points to separate the whole face into some small regions, as shown in Fig 3-7. In so doing, the corresponding points are restrained in the correspondent areas, and the range to be searched is greatly reduced. We suggest computing a linear mapping of each point to find its matching pair in another image, and searching a 5x5 square by taking the matching point as the center. The result is shown in Fig. 3-8.

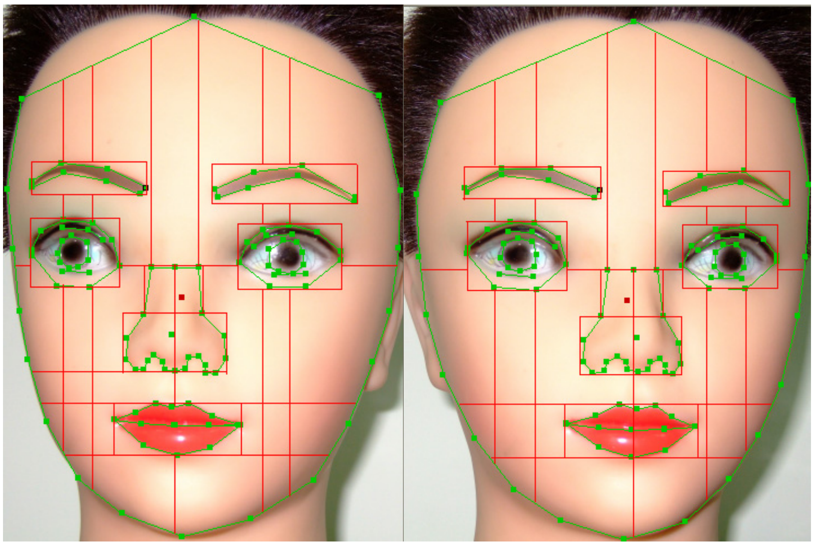
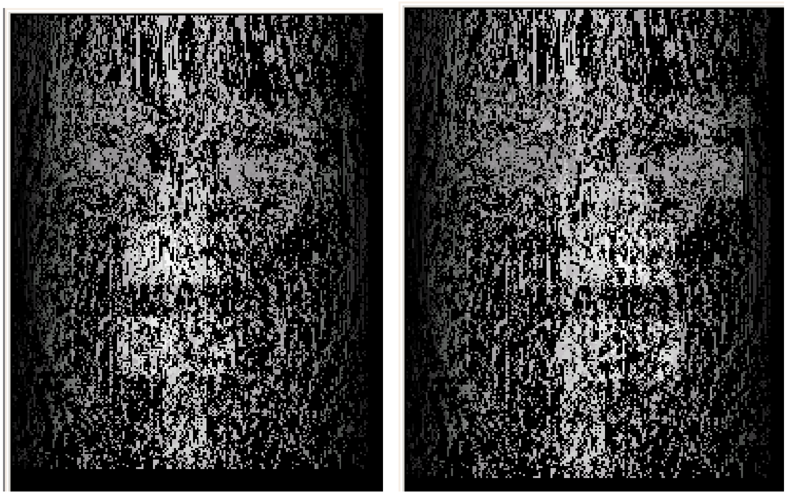


Fig. 3-7 Faces in an image pair are separated into small regions.



(a) (b)

Fig. 3.8: Disparity map obtained by the modified correlation algorithm.

(a) left image (b) right image

3.3 Interpolation and smoothing

When a dense disparity map is available, the next step is to eradicate the invalid points according to the disparity value of its neighborhood by interpolation. The interpolation is done using a weighted mask given by (9) where the pixels near the center are :assigned larger weights.

$$\begin{bmatrix} 1 & 2 & 3 & 2 & 1 \\ 2 & 4 & 6 & 4 & 2 \\ 3 & 6 & 9 & 6 & 3 \\ 2 & 4 & 6 & 4 & 2 \\ 1 & 2 & 3 & 2 & 1 \end{bmatrix} \tag{9}$$

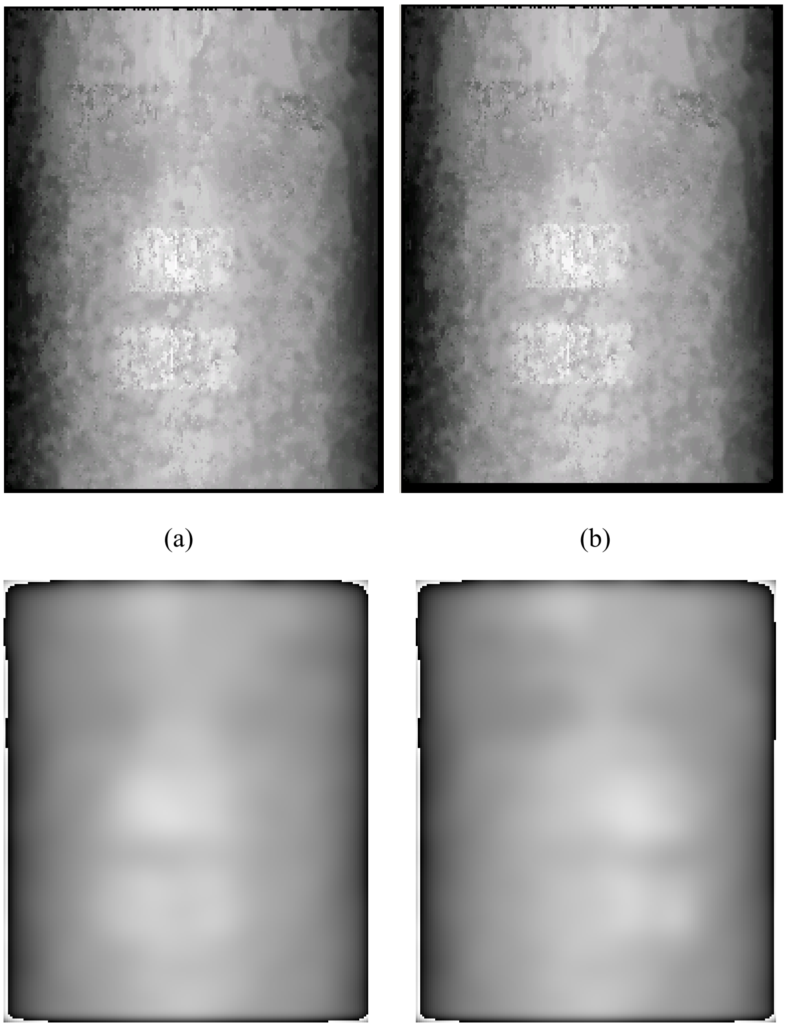
The rule for filtering unreliable nodes is summarized as follows:

```
do{
   for ( p(i,j) is invavlid )
     if ( number of valid points in kernel > 5 && weighted sum > 30 )
           fill p(i,j)
     else
           discard the result of interpolation at p(i,j)
}while( all points are valid )
```

When all initially invalid points are filled with correct disparity values, we adopt a 5x5 mean kernel, which is defined as:

$$\frac{1}{25}\left[1_{5\times 5}\right] \tag{10}$$

to smooth the disparity map.



(a) (b)

(b) (d)

Fig. 3-9: (a)(b) depth maps after interpolation
(c)(d) depth maps after smoothing.

After left and right image with a smoothing disparity map are generated, we proceed to obtain a frontal view image for caricature generation using view morphing techniques [6].

In a general view morphing process, 3 steps are required, namely, pre-warp, interpolate, and post-warp. But in our case, the image pair is already rectified, so an intermediate view can be interpolated directly according to:

$$p_{(1-s)} = p_L + sp_R \tag{11}$$

where s, p_L,p_R denotes morphing ratio, point in left and right images, respectively.

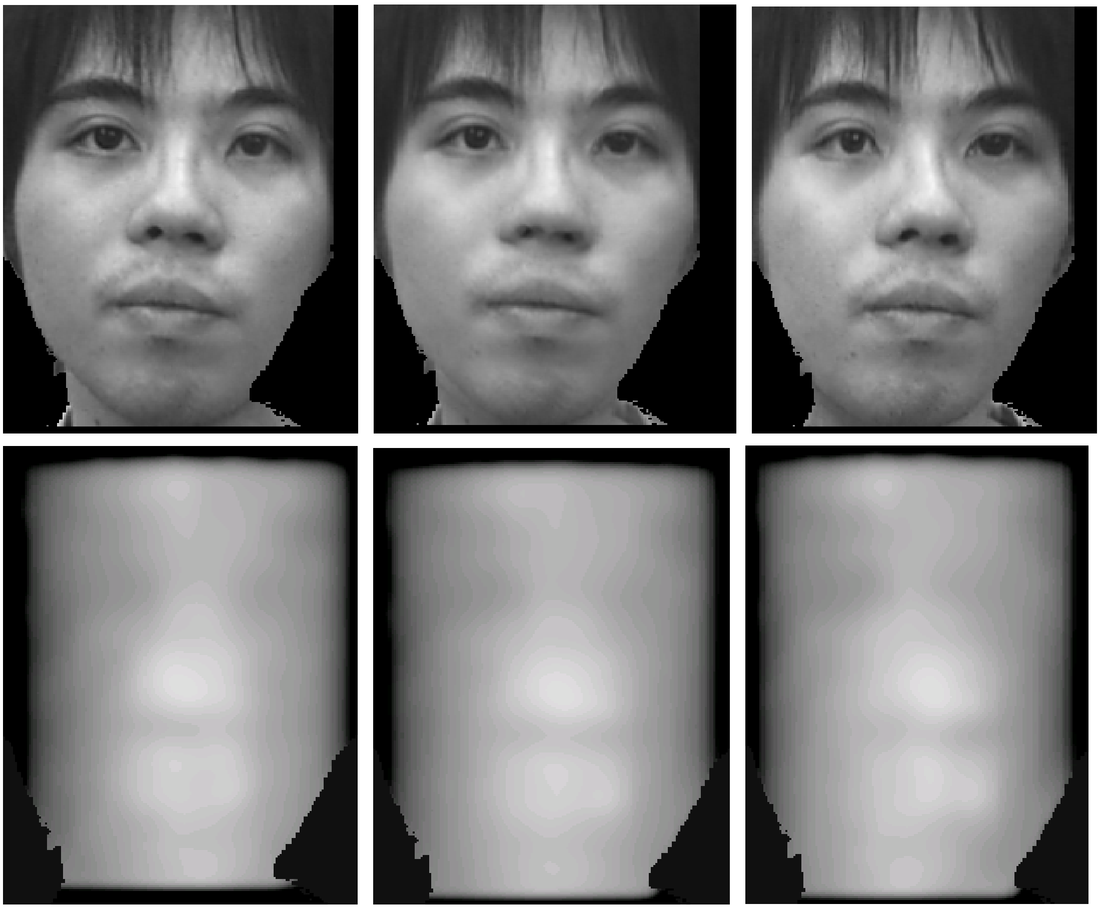


Fig. 3-10: Images with different view and their corresponding disparity map.