Chapter 3
Image Registration

3.1 Introduction

Image registration is a commonly used technique on two or more images which have overlapping regions with others. Because such images usually are taken from different view points and at different times, it is important to map these images to their correct positions to avoid unnecessary image artifacts before stitching them.

In general, the image registration can be separated into two steps, feature points extraction and feature matching. In this chapter, several procedures are proposed for the source image $\boldsymbol{I}_1$ and the target image $\boldsymbol{I}_2$ to register each feature point as accurately as possible by using the four-direction edge detection method [5] and then translate them referring to the matching pairs obtained by normalized cross-correlation method [1] [10].

3.2 Feature Points Extraction

Feature points extraction is a very important preprocessing step in many image recognition techniques, such as license plate registration [8][17] and face recognition [11]. It marks and locates the objects that with salient and distinctive features in images. For example, edges, corners, critical points or the eyes and nose in human faces which have some detectable special characteristics.

Since these features generally have abrupt changes in gray-level associating with high frequency components, the high-pass filtering technique is needed to attenuate low-frequency components, not to disturb high-frequency information. The gradient domain feature detectors are adopted to select feature points, which are invariant to the change of light conditions.

In this section, several edge operators will be used to extract a set of useful feature points for desired registration parameters. First of all, 2D spatial averaging filter masks such as the Gaussian smoothing masks [5] are commonly used for data smoothing to reduce the noise. The smoothed image $\boldsymbol{I}'(x, y)$ is produced by evaluating the convolution of the Gaussian smoothing masks $\boldsymbol{H}$ with input image $\boldsymbol{I}(x, y)$ which can be expressed as

$$\boldsymbol{I}'(x,y) = \sum_{i=-1}^{1}\sum_{j=-1}^{1}\boldsymbol{I}(x+i, y+j)\cdot\boldsymbol{H}(i,j) \tag{3.2-1}$$

where

$$\boldsymbol{H} = \frac{1}{16}\begin{bmatrix}1 & 2 & 1\\ 2 & 4 & 2\\ 1 & 2 & 1\end{bmatrix} \text{ or } \frac{1}{25}\begin{bmatrix}1 & 3 & 1\\ 3 & 9 & 3\\ 1 & 3 & 1\end{bmatrix}$$

After image smoothing, the first partial derivative, gradient, is used to extract the abrupt change of regions, the edges, in the image. The Sobel edge detection is the most popular method to find the sensible and reliable locations of edges by using the masks

$$\boldsymbol{S}_x = \begin{bmatrix}-1 & -2 & -1\\ 0 & 0 & 0\\ 1 & 2 & 1\end{bmatrix}, \boldsymbol{S}_y = \begin{bmatrix}-1 & 0 & 1\\ -2 & 0 & 2\\ -1 & 0 & 1\end{bmatrix}, \boldsymbol{S}_{d1} = \begin{bmatrix}-2 & -1 & 0\\ -1 & 0 & 1\\ 0 & 1 & 2\end{bmatrix} \text{ and } \boldsymbol{S}_{d2} = \begin{bmatrix}0 & -1 & -2\\ 1 & 0 & -1\\ 2 & 1 & 0\end{bmatrix}$$

to get the gradient images as

$$\begin{aligned}
\nabla_x I'(x,y) &= \sum_{i=-1}^{1}\sum_{j=-1}^{1}\boldsymbol{I}'(x+i, y+j)\cdot\boldsymbol{S}_x(i,j)\\
\nabla_y I'(x,y) &= \sum_{i=-1}^{1}\sum_{j=-1}^{1}\boldsymbol{I}'(x+i, y+j)\cdot\boldsymbol{S}_y(i,j)\\
\nabla_{d1} I'(x,y) &= \sum_{i=-1}^{1}\sum_{j=-1}^{1}\boldsymbol{I}'(x+i, y+j)\cdot\boldsymbol{S}_{d1}(i,j)\\
\nabla_{d2} I'(x,y) &= \sum_{i=-1}^{1}\sum_{j=-1}^{1}\boldsymbol{I}'(x+i, y+j)\cdot\boldsymbol{S}_{d1}(i,j)
\end{aligned} \tag{3.2-2}$$

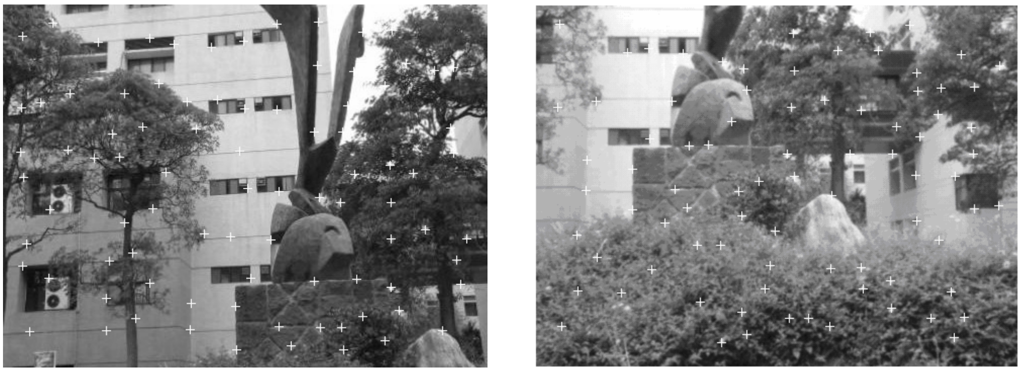
in the x direction, y direction and two diagonal directions, respectively. Because the pixels in the gradient images $\nabla_x\boldsymbol{I}'$, $\nabla_y\boldsymbol{I}'$, $\nabla_{d1}\boldsymbol{I}'$ and $\nabla_{d2}\boldsymbol{I}'$ represent the degree of changes in different direction, the modulus of all the gradient images can be obtained as

$$|\nabla I'(x,y)| = \sqrt{|\nabla_x I'(x,y)|^2 + |\nabla_y I'(x,y)|^2 + |\nabla_{d1} I'(x,y)|^2 + |\nabla_{d2} I'(x,y)|^2} \quad (3.2\text{-}3)$$

If all the local maxima of $|\nabla I'(x,y)|$ in the region with size 35×35 are located and greater than the threshold value T_1, they will be treated as the edge points and transformed by the binarilization to a binary image $\boldsymbol{I}_b(x,y)$, which has significant features and is expressed as

$$\boldsymbol{I}_b(x,y) = \begin{cases} 255 & |\nabla I'(x,y)| \geq T_1 \\ 0 & \text{otherwise} \end{cases} \quad (3.2\text{-}4)$$

The threshold T_1 is commonly selected according to experiment results. The points of $\boldsymbol{I}_b(x,y) = 255$, i.e., $|\nabla I'(x,y)| \geq T_1$, are called the feature points which can be used in feature matching method to find the matching pairs. In Figure 3.1, the center of the white crosses in the images represent the location of feature points extracted after the method proposed above using the gray-level images of the source image $\boldsymbol{I}_1$ and the target image $\boldsymbol{I}_2$ as input images.



(a)

(b)

Fig 3.1 The results of feature extraction process for the source images (left images) and the target images (right images)

In this section, the locations of the points with strong high-frequency information and local maxima values in gradient domain can be detected by the above method. After the feature points of the source image $\boldsymbol{I}_1$ and the target image $\boldsymbol{I}_2$ are extracted, next section will introduce the so-called feature matching method to get the matching pairs between these two input images.

3.3 Feature Matching

After the significant feature points with local maxima values in gradient domain are obtained, they could be also extracted in other images taking for the same scenes because the location represented by these feature points has the salient and distinctive information. In other words, if the sight of image changes due to the motion of camera, some scenes in the source image will no longer exist in the target image such that the related feature points become useless. Therefore, feature matching method is important to determine whether the feature points extracted in one image have the correspondents in the other image to form matching pairs and to eliminate the feature points which fail in matching.

Let $\boldsymbol{FP}_{\boldsymbol{I}_1}$ and $\boldsymbol{FP}_{\boldsymbol{I}_2}$ be denoted as two sets of coordinates of the feature points in the source image $\boldsymbol{I}_1$ and the target image $\boldsymbol{I}_2$ which can be expressed respectively as

$$\begin{aligned} \boldsymbol{FP}_{\boldsymbol{I}_1} &= \{(x,y) \mid \boldsymbol{I}_{b_1}(x,y) = 255\} \\ \boldsymbol{FP}_{\boldsymbol{I}_2} &= \{(x,y) \mid \boldsymbol{I}_{b_2}(x,y) = 255\} \end{aligned} \qquad (3.3\text{-}1)$$

where $\boldsymbol{I}_{b_1}$ and $\boldsymbol{I}_{b_2}$ are the binary images given in (3.2-4). Furthermore, if both the source image $\boldsymbol{I}_1$ and the target image $\boldsymbol{I}_2$ are possessed of the same sights, their feature points in the same relative positions will be extracted as matching pairs that with cross-correlation higher than other possible matches.

In order to find the matching pairs of all the feature points in $\boldsymbol{I}_1$ and $\boldsymbol{I}_2$, the similarity of each pair is measured by normalized cross-correlation algorithm [1] [10]. Let $p_i \in \boldsymbol{FP}_{\boldsymbol{I}_1}$ be the ith feature point located at $\left(p_x^i, p_y^i\right)$ in $\boldsymbol{I}_1$ and $q_k \in \boldsymbol{FP}_{\boldsymbol{I}_2}$ be the kth feature point located at $\left(q_x^k, q_y^k\right)$ in $\boldsymbol{I}_2$. Calculate the mean values of p_i and q_k as

$$
\begin{aligned}
E_{p_i} &= \sum_{m=-17}^{17} \sum_{n=-17}^{17} I_1\left(p_x^i - m, p_y^i - n\right) \\
E_{q_k} &= \sum_{m=-17}^{17} \sum_{n=-17}^{17} I_2\left(q_x^k - m, q_y^k - n\right)
\end{aligned}
\tag{3.3-2}
$$

each based on a 35×35 window around the feature point. Then, the auto-correlations of p_i and q_k are determined as

$$
\begin{aligned}
R_{p_i} &= \sum_{m=17}^{17} \sum_{n=-17}^{17} \left[I_1\left(p_x^i - m, p_y^i - n\right) - E_{p_i}\right]^2 \\
R_{q_k} &= \sum_{m=17}^{17} \sum_{n=-17}^{17} \left[I_2\left(q_x^k - m, q_y^k - n\right) - E_{q_k}\right]^2
\end{aligned}
\tag{3.3-3}
$$

and the cross-correlation between p_i and q_k is obtained as

$$
R_{p_i, q_k} = \sum_{m=-17}^{17} \sum_{n=-17}^{17} \left[I_1\left(p_x^i - m, p_y^i - n\right) - E_{p_i}\right] \cdot \left[I_2\left(q_x^k - m, q_y^k - n\right) - E_{q_k}\right] \tag{3.3-4}
$$

Consequently, the normalized cross-correlation can be expressed as

$$
\boldsymbol{U}(i,k) = \frac{R_{p_i, q_k}}{\sqrt{R_{p_i}} \cdot \sqrt{R_{q_k}}} \tag{3.3-5}
$$

with $\boldsymbol{U}(i,k) \in [0,\ 1]$, which is generally used for two images of gray-level. In fact, it is easy to extend (3.3-5) to two images of RGB-level as below

$$
\boldsymbol{U}(i,k) = \frac{1}{3} \sum_{c=R,G,B} \frac{R_{p_i, q_k}^c}{\sqrt{R_{p_i}^c} \cdot \sqrt{R_{q_k}^c}} \tag{3.3-5}
$$

where the superscript c is equal to R, G and B to represent the auto-correlation and the cross-correlation corresponding to the colors of red, green and blue, respectively. The scale $\frac{1}{3}$ is added to satisfy the condition $\boldsymbol{U}(i,k) \in [0,\ 1]$. By fixing sth feature point $p_s \in \boldsymbol{FP}_{\boldsymbol{I}_1}$, the maximum value among $\boldsymbol{U}(s,k)$ of all the feature points $q_k \in \boldsymbol{FP}_{\boldsymbol{I}_2}$ can be obtained. For example, if $U(s,t)$ is the one we have, i.e.,

$$U(s,t) = \max_{q_k \in \boldsymbol{FP}_{\boldsymbol{I}_2}} \{\boldsymbol{U}(s,k)\} \tag{3.3-6}$$

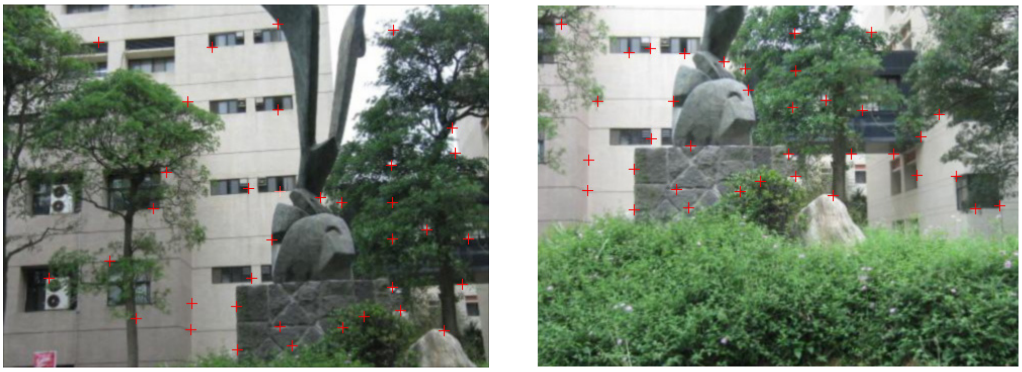
which implies $q_t \in \boldsymbol{FP}_{\boldsymbol{I}_2}$ is the best matching point to $p_s \in \boldsymbol{FP}_{\boldsymbol{I}_1}$. In other words, p_s and q_t are the best correlation pair.

After all the correlation pairs are found, choose the pairs satisfying $U(s,t) > T_c$, where T_c is the threshold value, as the matching pairs. Then, store the coordinates of all the matching pairs in order as $\boldsymbol{MP}_{\boldsymbol{I}_1,\boldsymbol{I}_2}$, described as

$$\boldsymbol{MP}_{\boldsymbol{I}_1,\boldsymbol{I}_2}(j) = \{(x_1^j, y_1^j, x_2^j, y_2^j) \mid (x_1^j, y_1^j) \in p_s, (x_2^j, y_2^j) \in p_t, U(s,t) > T_c\} \tag{3.3-7}$$

where j represents the jth matching pair of $\boldsymbol{MP}_{\boldsymbol{I}_1,\boldsymbol{I}_2}$. In Figure 3.2, the red crosses represent the matching pairs after applying the feature matching method. It is obvious that there are many feature points in the same relative positions of the source and target images are extracted as the matching pairs.

After applying the normalized cross-correlation method described above, the set storing the coordinates of matching pairs $\boldsymbol{MP}_{\boldsymbol{I}_1,\boldsymbol{I}_2}$ are obtained. However, it is inevitable that there still exist some wrong matching pairs which might influence the success of mapping. Therefore, an appropriate image mapping technique is introduced in the next chapter to find the optimal mapping from one image to the other and the false matching pairs can be found and eliminated automatically.



(a)

(b)

Fig 3.2 The results of feature matching process for the source images (left images) and the target images (right images)