Chapter 2

Prior Arts of Image Registration and HDR Application

2.1 Introduction

Image registration has already been studied for a long time. It was an important technique for a great variety of applications such as aerial image analysis [10-11] or medical image processing. [12] These applications require high accurate fusion thus many unknowns are needed but the cost of computation is high. Above applications are not applicable for DSC because of the long process. Therefore, in order to be applicable in DSCs, a simple and fast registration algorithm including less multiplication, less memory buffer, and higher computable efficiency are needed. In section 2.2, we will briefly introduce the basic concept of the prior arts of registration algorithms such as aerial image analysis which requires four unknowns and three appropriate algorithms for DSCs. In section 2.3, the basic concept of fusion for HDR will be expressed. Section 2.4 will be the summary.

2.2 The Basic Concept of Image Registration Algorithm

2.2.1 Aerial Image Analysis

Aerially remote measurement was described by a computational approach to the estimation of 2-D translation, rotation, and scale from two partially overlapping images. When the camera moved along a specific vector with the effects of wind and shake of aircraft, the two

images not only shifted linearly by the translation parameters (ΔX, ΔY) but also rotated by an angle θ. Sometimes they may also be changed by the scaling factor (s) along the z direction. The geometry of two images, taken by an aircraft with a down-looking camera at time t1 and t2, is schematically illustrated in **Fig. 2-1**. The relationship between a point (X, Y) in one image and its corresponding point ($\hat{X}$,$\hat{Y}$) in the other image can be expressed by a 2-D transformation defined in equation 2-1.

$$\begin{pmatrix} \hat{X} \\ \hat{Y} \end{pmatrix} = s \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix} + \begin{pmatrix} \Delta X \\ \Delta Y \end{pmatrix} \qquad (2\text{-}1)$$

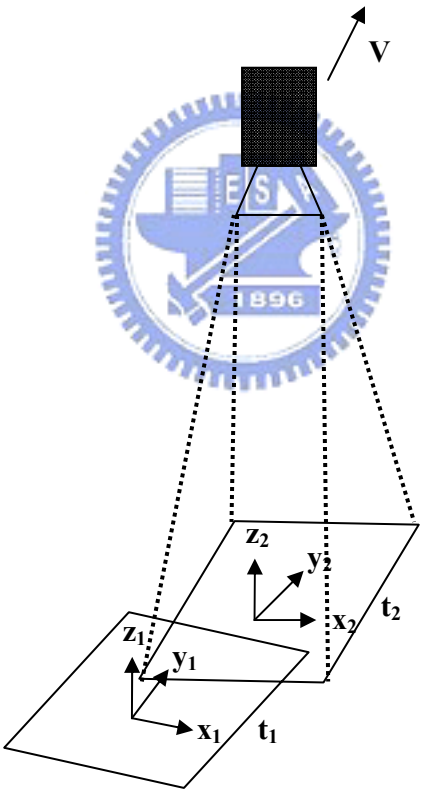


Fig. 2-1 Geometry of aerially remote measurement. Images are taken by a down-looking camera at time t_1 and t_2.

In practice, the variation of the two pictures continuously taken by the commercial DSC is

much less than by the camera of aerially remote measurement. Moreover, to estimate four unknowns by only using pixel values of two images will also be an issue.

In order to simplify the algorithm, both the rotation angle (θ) and the scaling factor (s) can be ignored because the misalignment of the images taken continuously was not serious. Such a tiny rotation angle which affects the blur edges after fusion is as less as viewer can distract, and the hand-shake always happens in the x-y plane thus the effect of scaling factor is very small. Only the translation parameters (ΔX, ΔY), global motion vector (GMV), are what we need. In the following, three different kinds of registration algorithms for estimating GMV will be introduced. The main registration procedure will be divided into two parts: The first is pattern-selection and the second is GMV-estimation. We will concisely introduce these two parts of each algorithm and discuss the issues related to the efficiency, complexity, and adaptation for different exposed images of each other.

2.2.2 Zhong Zhang's Algorithm

(1) Pattern-selection：

Zhang [13] provided a feature-based method detecting the stronger edges from the images and obtained the patterns. **Fig. 2-2 (a)** shows a mask contains 3x3 pixels and Z_i are the relative positions. The masks and their gradient of three well-known edge detection filters including Sobel, Prewitt, and Roberts are shown in **Figs. 2-2 (b), (c),** and **(d)** respectively. The pattern of edges is calculated by equation 2-2. P_{Gx} and P_{Gy} are the two patterns filtered by gradient, G_x and G_y. All patterns of one specific image filtered by above masks and the extra "Canny" edge detection are shown in **Fig. 2-3**. [15] Quantizing the intensity of image into three levels using hysteresis thresholds and smoothing by Gaussian mask can extract the feature-based pattern. **Fig. 2-4 (a)** is one of the test images and **Fig. 2-4 (b)** shows the pattern of edges filtered by "Canny" filter. **Fig. 2-4 (c)** is the final pattern of the stronger edges.

$$Pattern = \sqrt{P_{G_x}{}^2 + P_{G_y}{}^2} \qquad (2\text{-}2)$$

Z_1	Z_2	Z_3
Z_4	Z_5	Z_6
Z_7	Z_8	Z_9

(a)

-1	-2	-1
0	0	0
1	2	1

-1	0	1
-2	0	2
-1	0	1

$G_x = (Z_7 + 2Z_8 + Z_9) - (Z_1 + 2Z_2 + Z_3)$ $\quad$ $G_y = (Z_3 + 2Z_6 + Z_9) - (Z_1 + 2Z_4 + Z_7)$

(b)

-1	-1	-1
0	0	0
1	1	1

-1	0	1
-1	0	1
-1	0	1

$G_x = (Z_7 + Z_8 + Z_9) - (Z_1 + Z_2 + Z_3)$ $\quad$ $G_y = (Z_3 + Z_6 + Z_9) - (Z_1 + Z_4 + Z_7)$

(c)

-1	0
0	1

0	-1
1	0

$G_x = Z_9 - Z_5$ $\quad$ $G_y = Z_8 - Z_6$

(d)

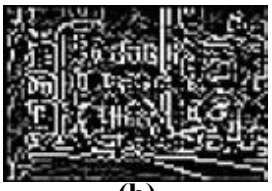
Fig. 2-2 (a) The mask and its relative position, Z_i. (b) Sobel, (c) Prewitt, and (d) Roberts are three different masks and their gradient.

(a)

(b) **(c)**

(d) **(e)**

Fig. 2-3 (a) The target image. The patterns of edges are filtered by (b)Sobel, (c) Prewitt, (d) Roberts, and (e) Canny masks.



(b)

(a) **(c)**

Fig. 2-4 (a) The target image. (b) The pattern of its edges. (c) The pattern of its stronger edges.

.

(2) GMV-estimation ：

To compute the gravity centre (C_x, C_y) of above pattern M(x, y) by equation (2-3) was simple.

$$c_x = \frac{\Sigma_{x,y} xM(x,y)}{\Sigma_{x,y} M(x,y)} \quad ; \quad c_y = \frac{\Sigma_{x,y} yM(x,y)}{\Sigma_{x,y} M(x,y)} \qquad (2\text{-}3)$$

By estimating the relative position, ΔX and ΔY, between two gravity centers (C_{tx}, C_{ty}) of the target image and (C_{rx}, C_{ry}) of the reference image, the GMV could be estimated.

(3) Discussions：

Detecting the stronger edges of two images needs a great deal of calculation especially multiplication. This complicated procedure results in lower efficiency and greater complexity. Moreover, due to the effect that the noise under low exposure is more than under the high exposure, two patterns with high and low exposures are quiet different. Therefore, the GMV-estimation is unreliable for DSC applications.

2.2.3 W.-C. Kao's Algorithm

(1) Pattern-selection：

W.-C. Kao proposed the global motion stabilization adapted to fuse different exposed images. Two patterns were captured directly by color filter array (CFA) which is made up of the Bayer pattern shown in **Fig. 2-5**. Then, low exposed one was multiplied by the exposure ratio between two images because the intensity was changed with different exposures.

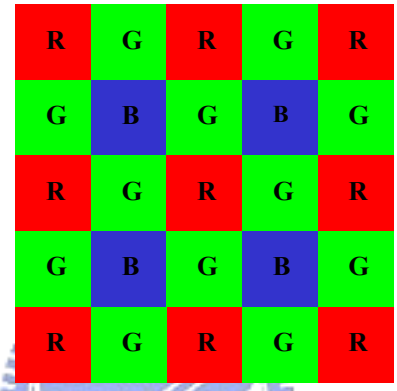


Fig. 2-5 The Bayer pattern.

(2) GMV-estimation：

The GMV-estimation was to cut each pattern into macro blocks. Thereafter, each block was shifted by a search window contained 16×16 pixels. By calculating the variations of sum of absolute difference (SAD) in each shift step, the minimum value of SAD was the best motion vector in this block. The SAD was derived by the equation 2-4 where $a_{p,q}(i, j)$ and $\hat{a}_{p,q}(i, j)$ can be further derived by equation 2-5 and equation 2-6.

$$S(i,j) = \frac{1}{(N+1)^2} \sum_{p=-N/2}^{N/2} \sum_{q=-N/2}^{N/2} \left| a_{p,q}(i,j) - \hat{a}(i,j) \right| \qquad (2\text{-}4)$$

$$a_{p,q}(i,j) = \sum_{u=16i}^{16i+15} \sum_{v=16i}^{16i+15} \left| \hat{H}(u,v) - D(u+p, v+q) \right| \qquad (2\text{-}5)$$

$$\hat{a}(i,j) = \sum_{p=-N/2}^{N/2} \sum_{q=-N/2}^{N/2} a_{p,q}(i,j)/(N+1)^2 \qquad (2\text{-}6)$$

H and D in equation 2-5 are the pixel values of two images and N is the size of search window. After all motion vectors of blocks are estimated, the GMV is achieved by the maximum count of them.

(3) Discussions：

SAD is inefficient because it was subtracted by pixel values of two images containing 8-bits, and the amount of all shifted steps is N×N×blocks. There are fewer multiplication in equations 2-4, 2-5, and 2-6 so that this algorithm can not be complicated. Although the initial multiplication of the low exposed pattern was to make it the same intensity as high exposed pattern, the intensity of two patterns was not totally equal because of the nonlinearity of cameras.

2.2.4 Ward's Algorithm

(1) Pattern-selection：

The pattern-selection method of Ward [14] is to obtain each medium threshold bitmap (MTB) to achieve similar patterns under different exposed images. **Figs. 2-6 (a)** and **(b)** show two images with high and low exposures separately. **Figs. 2-7 (a)** and **(b)** are the MTBs of the two images. The two patterns are similar because the medium value was changed through the different exposed images.

(a) **(b)**

Fig. 2-6 Two images with (a) high exposure and (b) low exposure.

(a) **(b)**

Fig. 2-7 (a) The MTB pattern of Fig. 2-6(a). (b) The MTB pattern of Fig. 2-6(b).

(2) GMV-estimation：

The alignment step was based on the multi-scale techniques which were well-known in the computer vision and image procession communities. **Fig. 2-8** is the pyramid decomposition of MTB with $\log_2$(max_offset) levels pasting the base resolution. Each level was operated by the shift block to obtain the GMV and doubled to the next level until the final GMV-estimation was completed.

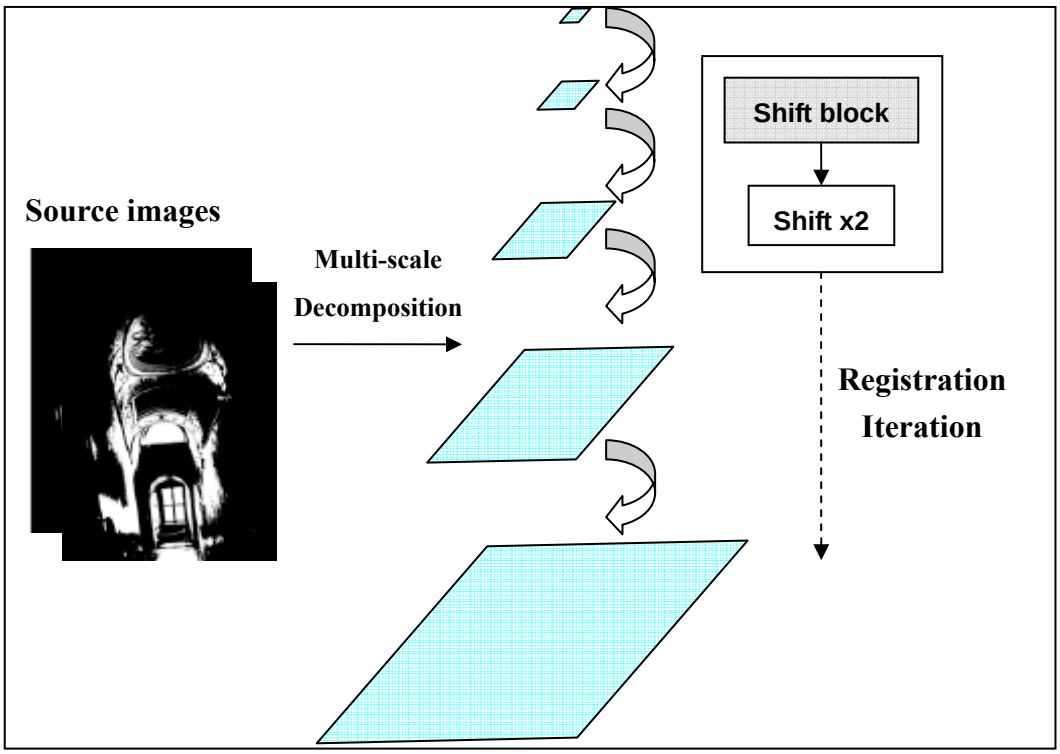


Fig. 2-8 The pyramid decomposition of estimating GMV.

(3) Discussions：

MTB was the binary pattern so that estimating the shifted value of each level was more efficient comparing with the algorithm of W.-C. Kao. Moreover, two patterns of MTBs with different exposures were similar, thus can be adapted for fusing high and low exposed images. However, the processing of pyramid decomposition was complicated because of the procedures including down-sampling and filtering.

2.2.5 Comparison

According to the above discussions of three algorithms of efficiency, complexity, and adaptation for different exposed images, the pros and cons of each algorithm are listed in **Table 2-1**. The additional discussion of memory buffer was needed since less memory requirement results in low cost. W.-C. Kao's algorithm which was processed by one block each time is better than others for memory buffer. Therefore, our improved algorithm should

be proposed by combining above merits such as less multiplication, less memory buffer, high efficiency and adaptation for the application.

Table 2-1 The comparison of Zhang's, W.-C. Kao's, and Ward's methods.

	Efficiency	Memory buffer	Complexity (multiplication)	For different exposures
Zhang	×	×(All size)	×	×
W.-C. Kao	Δ	O(16×16)	O	Δ
Ward	O	×(All size)	Δ	O

×：bad ； Δ：acceptable ； O：good

2.3 The Basic Concept of Fusion for HDR Image

2.3.1 The Definition of HDR

Scenes from the real world can exhibit a broad range of light variations especially containing both areas of low and high illumination. In human vision, people can feel the difference even with tiny change of light intensity. For example, as we take the image in dark room with a sunlit window, it is easy for us to observe the details of outside scenery and the inside objects. It is clear that the dynamic range, the term referring to the ratio of the highest and lowest recorded level of light, in such scenes can be very high. In the case of DSCs, the dynamic range is usually described in terms of the ratio between the maximum charges that the CCD/CMOS sensor can collect (full well capacity) and the minimum charges that the sensor can just overcome the noises (noise floor). There are two factors which affect the dynamic range of DSC. One is the pixel size of CCD/CMOS sensor which affects the capacity for charges collection. Moreover, the low noise floor is affected by the ability of noise suppression for the sensor. Second is the procedure of quantification. The light values

captured by the sensor are finally quantized into a range between 0~255 to produce an 8-bits per pixel storage, even if more bits were initially used by the sensor analog to digital converter (ADC) to cover the input signal. Therefore, information loss in highly and dimly illuminated areas is observed because all light variations are mapped to the saturated value and signals near noise floor are unclear respectively. **Fig. 2-9** shows the portrait that was photographed by the user inside the building of bright background. Obviously, the face is unclear and the sky is saturated without cloud.

Fig. 2-9 The portrait taken by the user inside the building of bright background.

Nowadays, most of the DSCs are incorporating with automatic gain control (AGC) functioning along with a metering technique to allow the users to take pictures without manual setting of the device. Choosing “exposure bracket” mode can easily take pictures continuously with various exposure values (EV) until obtaining a subjective satisfactory representation. However, once portions of the dynamic range have been definitively lost, there is impossible to restore them, even using some sort of post-processing techniques. The improved method is to combine multiple pictures of the same scene using different exposure settings, each of which will reveal different details. Pictures taken with low exposure provide highlight information and vice versa. Two images with low and high exposure settings are

shown in **Figs. 2-10 (a) and (b)**.

(a)

(b)

Fig. 2-10 Two images with different exposure settings: (a) low exposure and (b) high exposure.

The image fusion procedure is to swap the saturated area from low exposed image which contains more information, such as the cloud in the sky, as shown in **Fig. 2-10 (a).** Then to replace the dark zone by using high exposed image shown in **Fig. 2-10 (b)** obtains clear portrait. The fusion result for a HDR image will be discussed more detail in chapter 5.

2.3.2 Nonlinear Mapping in Camera

In the conventional camera, film is exposed to the light to form a latent image. The film is then developed to change this latent image into variations in transparency or density on the film. After developing, the film scanner which projects light through the film onto an electric light-sensitive array will convert the image (analog signal) to electrical voltage (digital signal). These digitized voltages manipulated before finally being written to the storage medium. The scanning process can also introduce nonlinear mappings. The image acquisition pipeline shown in **Fig. 2-11** illustrates that how scene radiance becomes pixel values for both film and digital camera.

In the first stage of the process, the film responses to variations in exposure X (which is $E\Delta t$, the product of the irradiance E the film receives and the exposed time Δt) is a nonlinear function, called the "characteristic curve" of the film. The development, scanning and digitization processes usually introduce their own nonlinearities, is composed of the nonlinear relationship between the image pixel exposures X and their values Z.

Comparing with film camera, the digital still camera (DSC) using charge coupled device (CCD) arrays to achieve the information for the radiance of the scene is prone to the same difficulties. Although the collected charges are proportional to its irradiance, most digital cameras apply a nonlinear mapping to the CCD outputs before writing to the storage medium. The nonlinear mapping is used in various ways to mimic the response characteristics of film anticipates nonlinear responses in the display device.

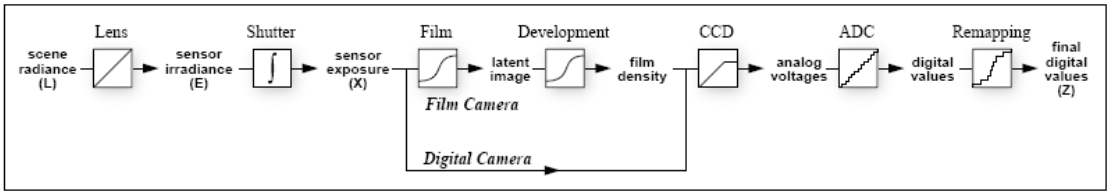


Fig. 2-11 Image acquisition pipeline

As a result, when we photograph a scene either with film or digital camera, the digitized 8-bit pixel values are rarely truly measured of the relative radiance in the scene. For example, if one pixel has twice of the value of another, it means that twice of the radiance will not be received.

2.3.3 HDR Image Recovering Using Exposure Times

By using differently exposed images, the measured brightness change with the exposure time while radiance remains constant, where the radiance (photo quantity) of the scene can be estimated. The fused concept [14] is shown in **Fig. 2-12**. Multiple images which are 8-bit pixel values with different exposures are taken. Transforming images by camera response curve can obtain each photo quantities. These photo quantities are combined by their relative weighting functions and then the radiance of this scene is achieved. From the estimated photo quantity, one or more output images may be generated by multiplying the desired synthetic exposed ratio and passing the result through the estimated camera response curve.

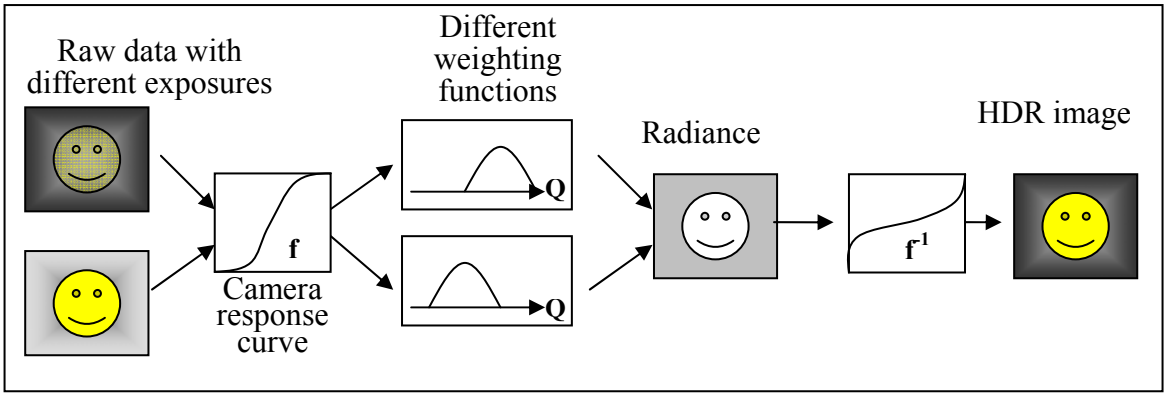


Fig. 2-12 The flow chart of fused concept.

Therefore, once the camera response curve *f* was known, it is possible to transfer the radiance to the 8-bit image for high dynamic range.

2.4 Summary

The prior arts of registration and HDR application were introduced. For image registration, three applicable algorithms were discussed and compared of efficiency, complexity, memory buffer and adaptation for HDR application. Moreover, the definition of high dynamic range and the briefly introduction of fusion by using two different exposed images were presented.