

# Digital Visual Effects hw2 Report

R06944059 林昀宣 R06922033 陳映紅



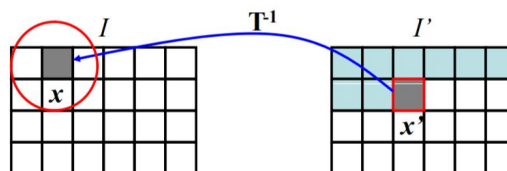
## 1. Taking Photo

我們使用到單眼相機，並且向助教借了腳架來使用，拍攝地點在台大醉月湖畔、天主教聖家堂以及台大社科院五樓。

## 2. Inverse Warping

將拍攝到的照片投影到 Cylindrical 空間上，使得照片之間能有適當的對齊。我們先將照片輸入 AutoStitch 這套軟體中，求得了每張照片對應到的 focal length，根據求得的 focal length 對每張照片分別做 inverse warping，使用上課提到的公式，從新的座標點推得其對應到原本照片的座標點，由於推得的結果不一定為整數，因此我們將求得的座標點去四捨五入找到最近的整數座標。

$$x' = s\theta = s \tan^{-1} \frac{x}{f}$$
$$y' = sh = s \frac{y}{\sqrt{x^2 + f^2}}$$



### 3. Feature Detection



我們依照上課提過的MSOP論文[1]實作feature detection和feature description。Feature Detection的實作流程如上圖，可分為以下四個部分：

- 建立3層的image pyramid，在每一層做適當的高斯模糊後取harmonic mean
- 在每一層對其harmonic mean取local maximum，並留下數值大於等於10的local maximum
- 對剩下的點做一次sub-pixel refinement，找更精確的位址
- 找到三層金字塔的所有特徵點後，做non-maximal suppression，留下數值夠大又夠分散的500個特徵點，作為這張圖最後輸出的特徵點。下圖左為沒有做non-maximal suppression的所有特徵點，下圖右為做完non-maximal suppression的500個特徵點



### 4. Feature Description



Feature Description分為以下四部分：

- 在每一層做適當的高斯模糊後，即可求出當層之特徵點的角度
- 以特徵點為中心張開一個40\*40的patch，並以方才求出的角度做適當旋轉
- 得到patch後，將整個patch分為64個5\*5的小方格，求出每個小方格之灰階值的mean，存成一個64維的向量
- 將此向量做標準化，使其平均值為0，標準差為1



原圖中的特徵點



取40\*40的patch



旋轉後的patch

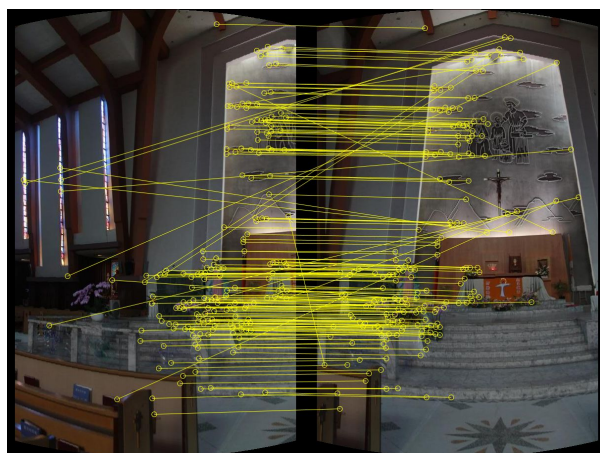


標準化後的64維向量



## 5. Feature Matching

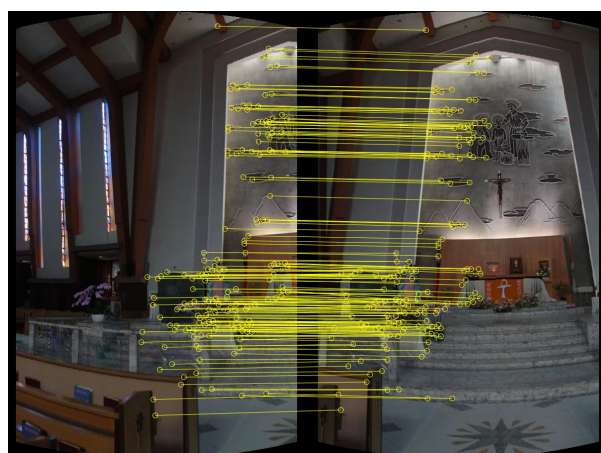
將每張圖取出的 500 個 feature point，根據他們的 64 維 descriptor 向量計算兩張圖之間每個點的 L2 distance 並取 distance 最小的當作一個 feature match。實際做法為了避免第  $i$  張圖的不只一個 descriptor 在第  $i + 1$  張圖中對應到同一個點，希望兩張圖之間的 descriptor 能得到 one-to-one 的對應，因此將兩張圖的所有 descriptor 彼此計算後，算出  $500 * 500$  個距離，再將其排序，從距離最小的開始當作 match，並且會將  $D(d_i, d_{j'}) / D(d_i, d_{k'}) > 0.8$  的 pair  $(d_i, d_{j'})$  去掉， $d_i$  為第一張照片的 descriptor， $d_{j'}$  和  $d_{k'}$  為第二張照片的 descriptor，且  $D(d_i, d_{j'})$ 、 $D(d_i, d_{k'})$  為所有第二張照片 descriptor 中與  $d_i$  距離最小與第二小者，因為這樣代表 descriptor 與圖二中不只一個 descriptor 都很接近，得到初步的 feature matching 結果如右圖。從右圖中很明顯有些配對是 outlier，於是接著使用 RANSAC 的方法將這些 outlier 去除，參數設定為  $(n, P, p, k) = (3, 0.5, 0.9999, 69)$ ，並隨機 sample 出的三個 match points，從這三個點取平均水平位移量  $dx$  和垂直位移量  $dy$ ， $dx$  和  $dy$  為第二張圖的 descriptor 要對應到第一張圖的 descriptor 平均該移動的量，下圖左為這 69 個 iteration 中 sample 到最好的三個點使得最多的 descriptor 是符合的，如下圖右，由圖中也可看出 RANSAC 有效的去除了大部分的 outlier。



初步 feature matching



sample 的 descriptor



在此 sample 下符合的 descriptor

## 6. Image Alignment

從上述 RANSAC 中求得的 feature matching 結果，這些 descriptor 座標依照以下矩陣方式填入後，第  $i + 1$  照片中的 descriptor 座標會填入 A 矩陣中，第  $i$  張照片的 descriptor 會填入 b 矩陣中，依照這個去解 overdetermined system 的 normal equation，求得最終的  $dx$  和  $dy$ 。

$$Ax = b$$

$$\begin{bmatrix} 1 & 0 & x_1 \\ 1 & 0 & x_2 \\ \dots & & \\ 1 & 0 & x_n \\ 0 & 1 & y_1 \\ 0 & 1 & y_2 \\ \dots & & \\ 0 & 1 & y_n \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} dx \\ dy \\ 1 \end{bmatrix} = \begin{bmatrix} x_1' \\ x_2' \\ \dots \\ x_n' \\ y_1' \\ y_2' \\ \dots \\ y_n' \\ 1 \end{bmatrix}$$

## 7. Image Blending

將第  $i$  張照片和第  $i + 1$  張照片接合時，重疊的部分挑選其中一部份做兩張照片的  $\alpha * \text{img}_i + (1 - \alpha) * \text{img}_{i+1}$  內插，挑選的範圍經實驗結果，我們大概使用重疊的部分最中間寬為 30 個 pixel 做interpolation，如果選擇重疊的部分全部去做的話，會有照片因為每個位置的角度不一樣而有明顯的鬼影，但當 interpolation 的 pixel 數量太少的話，就會有很明顯的斷點以及顏色明亮度上的落差。





## 8. Refinement

單純從 image alignment 和 blending 的結果，得到的全景圖會有以下兩種缺點（如下圖），由於這兩個缺點，使得得到的照片並不是一般大眾可以接受的好的結果。

1. 照片在接合過程中會因為慢慢累積的 dy 誤差而有逐漸下移（或上移）的結果。
2. 由於我們是將每張照片做 cylindrical projection，因此每張都呈現上下凸出的取線，因此在縫隙間會有許多黑色區塊。



因此我們先將整張 panorama 的結果先將左右照片校正回同個水平後，再實作了 Rectangling Panoramic Images via Seam carving 的方法，得到的結果明顯是彌補了以上兩個缺點。



## 9. Result

我們測試了許多組照片，分別是助教提供的兩組測試資料和自己拍的照片，Artifacts 是我們的結果，AutoStitch 的結果為對照。我們的程式在兩組測資上表現都非常好，而在自己拍的照片中，表現最好的為醉月湖那組圖。

Test data : Parrington

Artifacts



AutoStitch



Test data : Grail

Artifacts





AutoStitch



Our Photo : 台大醉月湖畔

Artifacts

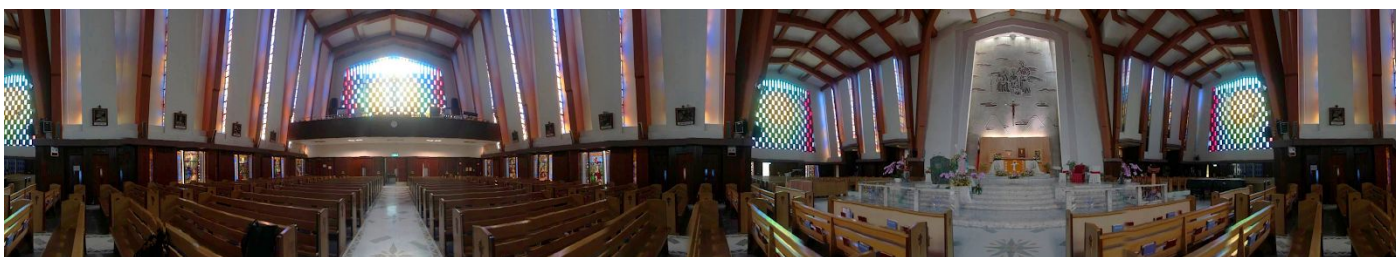


AutoStitch

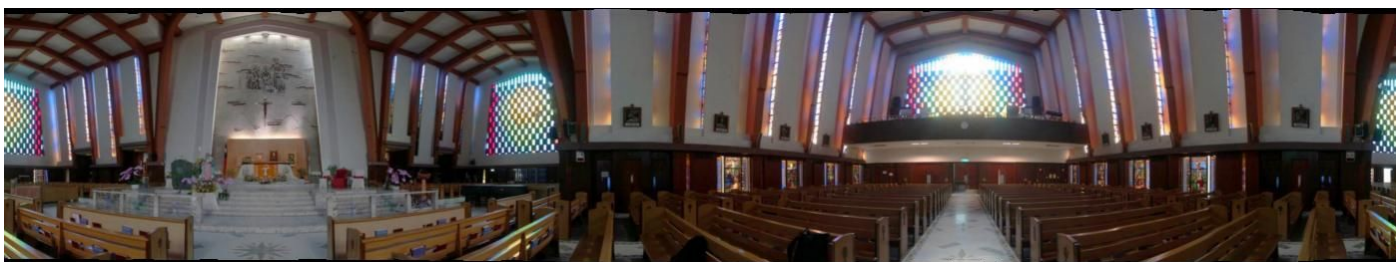


Our Photo : 天主教聖家堂

Artifacts



AutoStitch



Our Photo : 社科院五樓  
Artifacts



AutoStitch



## 10. Reference

1. Matthew Brown, Richard Szeliski and Simon Winder, "Multi-Image Matching using Multi-Scale Oriented Patches", CVPR 2005