

多焦距影像融合

曹博凱
朝陽科技大學
bokaicao@gmail.com

戴紹國
朝陽科技大學
sgdai@cyut.edu.tw

摘要

影像融合技術能夠將多張相同視景不同景深的影像融合成一張影像，而這張影像的每個地方都是由原來那些影像聚焦最清楚的部位所組成，所以產生的這張影像具有「到處都清楚」的特性。在這篇論文中，我們提出來一個新的影像融合的方法，結合了頻率域和空間域方法的優點，產生更精準的融合結果。其中頻率域的方法我們採用小波分解融合，而空間域的方法則加入了多階段的合併策略。實驗的部分，本實驗所提出的方法將與8種現存的方法進行比較，並且使用三種不同的評估方式來進行效率分析，結果證明我們的方法確實產生了較其他方法更佳的合併結果。

關鍵詞：小波轉換、空間域、頻率域、多階段合併。

1. 前言

現今的數位相機，特別是那些使用長焦距的數位相機會遭受到景深的問題，往往難以讓影像中的所有對象都獲得良好的聚焦。這些數位相機在拍攝影像時聚焦於某一事物上時，其他事物就會變成模糊。影像融合技術能夠將多張相同視景不同景深的影像融合成一張影像，而這張影像的每個物體都已個別取出最清楚的聚焦組成，所以當我們希望獲得影像能夠全部都聚焦的時候，這會是很有用的技術。尤其現今智慧型手機上的照相功已經漸漸的取代傳統數位相機。使用者可以很方便的利用手機上的App，在拍攝的同時直接進行影像的修改和美化，也因此可以在使用者未察覺的情形下手機直接拍攝多張不同景深的影像直接融合輸出。

目前影像融合方法分為兩個領域：空間域(Spatial domain)與頻率域(Frequency domain)。空間域算是影像融合方法中最簡單的領域，只需要計算原影像區塊的像素統計值來決定要選擇那張原影像，例如：Variance、Energy of image gradient (EOG)、Tenenbaum's algorithm (Tenengrad)、Energy of Laplacian of the image(EOL)、Sum-modified-Laplacian (SML)、Spatial frequency (SF)[14][10]，這些方法花費時間少且容易計算，然而這樣往往會造成不希望的副作用，例如由於他是區塊來選擇所以有時候選擇的區塊會包含失焦的地方。

所以近年來頻率域方面提出了許多新的方法，例如多分辨率影像融合[1][5][12][17]。然而有許多分辨率的方法，像是離散小波轉換

(DWT)[6][11][13][20][21]具有移變的特性(shift-variant)，例如一張原影像的物體對象有移動或是對焦失準的話這些算法的性能會變差。為了減緩這個問題的方法是使用離散小波框架轉換(DWFT)[18]或是雙樹複小波轉換(DT-CWT)[9]，DWFT是利用不同於DWT的方法，原始的DWT在經過轉換後，每個子帶(sub-band)的影像大小都只有原始影像的1/4，而DWFT在經過轉換後得到的子帶大小與原始影像大小相同。DT-CWT是將影像做轉換，然後以區塊計算的方式先建立一個區塊地圖(region map)，進行區域對區域的比較，在比較的結果進行影像融合。但是使用DWFT或DT-CWT很複雜且較費時。

所以近年來開始有學者結合頻率域與空間域的方法進行影像融合，在文獻中[3]是將DWT與DCT結合STD的方法進行融合。在本文中我們將使用基於文獻中[2]的方法結合空間域進行融合，本文的最終目標是希望結合空間域與頻率域的優點，產生更好的融合效果。

本文架構如下，在第二章節介紹空間域結合頻率域的多焦距影像融合方法，第三章實驗結果，最後是結論。

2. 多焦距影像融合方法

我們首先介紹頻率域融合 DWTFM 方法的原理與流程，接著是空間域多階段融合以及 DWTFM 與空間域結合的作法。

2.1 頻率域小波係數融合

離散小波轉換聚焦測量(Discrete Wavelet Transform Focus Measure; DWTFM)[2]主要是利用DWT轉換後的係數，進行聚焦測量，取得最佳聚焦點影像之後，再合成一張清晰的影像。令 $A = \{A_k, \text{for } k=1 \text{ to } N\}$ 是N張相同影像但是不同焦距的影像，將 A_k 進行DWT轉換後得到 B_{kvLL} 、 B_{kvLH} 、 B_{kvHL} 、 B_{kvHH} 等四個頻帶，其中V代表DWT的第i層分解，而 B_{kvLH} 及 B_{kvHL} 分別為水平與垂直方向的次高頻； B_{kvHH} 為高頻；而 B_{kvLL} 為低頻部分。因為清晰影像的部分，在高頻及次高頻的值會較高，低頻的值會相對較低；而模糊影像的部分，在高頻及次高頻的值會較低，低頻的值會相對較高。DWTFM是利用這種特性來計算聚焦點。

DWFT在聚焦測量部分，也是將這四個頻帶分別做計算及比較，但是DWFT是將各個頻帶各自取

最大值進行組合，雖然這樣可以得到最清晰的影像，但是會有一些影像的低階特徵受到影響，對於自動化辨識是不利的，所以DWTFM提出一個較不會影響自動化辨識所需特徵的聚焦測量方法來進行影像融合。下面公式是計算第k張影像的DWT第v層的座標(i,j)係數的聚焦點測量。

$$FM_{kvH}(i, j) = \frac{FM_{kvH}(i, j)}{FM_{kvL}(i, j)} \quad (1)$$

其中 FM_H 代表高頻部分的焦距度量

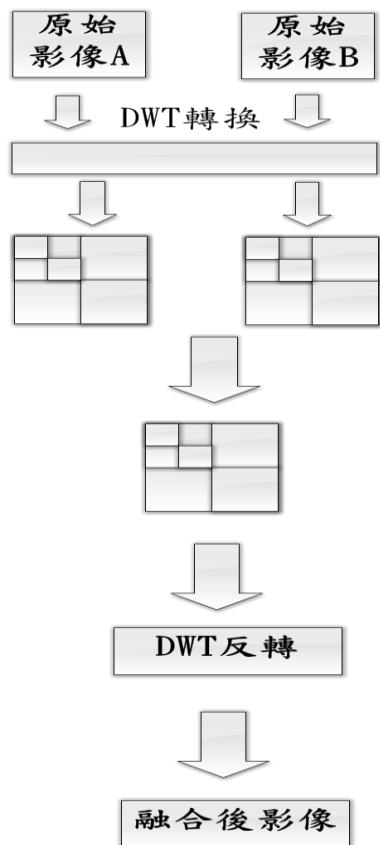
$$FM_{kvH}(i, j) = \frac{FM_{kvH}(i, j)}{\sqrt{B_{kvLH}(i, j)^2 + B_{kvHL}(i, j)^2 + B_{kvHH}(i, j)^2}} \quad (2)$$

而 FM_L 代表低頻部分的焦距度量

$$FM_{kvL}(i, j) = B_{kvLL}(i, j)^2 \quad (3)$$

在計算出所有影像第 v 層係數的 FM 值之後，對於每一個位置(i,j)選出一組最高的值，作為反向 DWT 之用，組合出第 v-1 層的係數，然後再重複剛才的步驟完成第 v-1 層的組合，直到反轉 DWT 轉換全部完成。

我們以圖 1 為例，首先將兩張原始影像個別進行 DWT 轉換再計算 DWTFM 找出聚焦最高的係數，將影像進行融合，反轉小波轉換 DWT，最後得到一張新的融合影像，我們稱為 D 圖。



圖一 DWTFM 影像融合流程圖

2.2 多階段空間域影像融合

雖然DWTFM解決了DWT與DWFT的問題，但是使用DWTFM融合後的新影像，雖然肉眼不易看出來差異之處，但是影像像素值幾乎都與原始影像的像素值有不小的差異，所以我們希望加入空間域的優點，讓融合後的像素值更接近原圖聚焦正確的部分，提高融合的品質。

我們進行的步驟分成多個階段，第一階段的目的在比對原圖和頻率域融合後的D圖，找出已經具有正確的融合結果的像素。為此，我們做了一個假設，如果D圖的像素和任一原圖非常接近，那我們就認為他是一個正確的融合結果。所以影像D在融合後對原圖A和B相減後取絕對值，若小於k值，那麼頻率域融合的像素會很接近我們想要的結果，我們會直接在D圖增減差值，使它和原圖像素一致，並且使用標記圖註明A或B來表示該像素是來自那一個原圖。此外若與每個原圖的差值都小於k，則表示每張原圖該部分都相差不遠，所以預測成哪一張原圖則不是那麼重要了。但是為了之後的運算我們另外使用一個特殊標記C來註明。因此這個階段除了修正D圖之外，並且會產生一個標記圖F，上面除了前述的A、B和C標記之外，另外以U來表示尚未決定的像素，如下公式。

$$F(i, j) \begin{cases} A & |D(i, j) - A(i, j)| \leq k \\ B & |D(i, j) - B(i, j)| \leq k \\ C & |D(i, j) - A(i, j)| \leq k \\ & \text{and } |D(i, j) - B(i, j)| \leq k \\ U & |D(i, j) - A(i, j)| > k \\ & \text{and } |D(i, j) - B(i, j)| \leq k \end{cases} \quad (4)$$

接下來之後的每一個階段都重複相同的程序，目的在針對第一階段未被決定的像素，進行判斷找出取自那一張原圖可以產生更正確的融合結果。這些像素值和每張原圖都有一個不小的差距，因此無法直接以像素值判斷。然而我們認為聚焦應該具有區塊性，也就是說不會有單一點聚焦不良，而其鄰居卻有良好聚焦的情形出現。也就是說，其最佳聚焦和鄰居有很大的關係。所以我們以未知點為中心，參考它的鄰居來確定其聚焦。以圖二為例，中心點是未決定的像素，利用8個鄰居點來判斷要選原圖A或是B。若未知點選擇B，而四周大多都是A，則這種聚焦情形機率很低；然而，若選擇A，則很可能是上下兩個A區塊邊緣接合在一起，這種聚焦情形的可能性就大很多了。所以在這種情形下，選擇原圖A會比較適合的。



圖二 3×3 鄰居示意圖

接下來描述詳細的做法：第一階段所產生的標示圖

F, 上面的可能標記包括 A、B、C 和 U。針對標示圖每一個 U 點, 檢視其鄰居的區塊, 來決定 U 應該更改為 A 或 B, 同時擷取其相對應原圖的值給 D 圖。令 N 為區塊中鄰居的數目, 以 3×3 為例, N 就等於 8。當區塊內鄰居標示 U 的數目高於 N/2, 則不做處理, 留待下一階段; 否則以公式 5 和 6 進行 F 圖的標示以及 D 圖的修改。其中公式 5 主要在找絕對多數的標示, n_A 、 n_B 和 n_C 分別是區塊內 A、B 和 C 標記的數目, 而 $n_{AC}=n_A+n_C$; $n_{BC}=n_B+n_C$ 。若 n_A 占絕對多數, 則該像素 A 圖的焦距會比較好的機率較高; 反之, 則是 B 圖的機率較高。因為標示 C 表示 A 或 B 都可能良好聚焦, 所以在統計絕對多數時, 也將他加入。若非絕對多數, 則使用公式 6 來計算 D 圖與 A 和 B 的相對差距, 如果和 A 的距離遠比和 B 的距離近, 則 A 有良好聚焦的機率就比較高; 反之, 則是 B 有良好聚焦的機率比較高。T 是一個小於 1 的臨界值, T 值設定越小, 表示要求與 A 和 B 的相對差距要越大。本研究對 T 的門檻值設 0.5, 如下所示。

$$F(i, j) = \begin{cases} A & (n_{AC} > n_{BC}) \text{ AND } (n_{AC} > \frac{N}{2}) \\ B & (n_{BC} > n_{AC}) \text{ AND } (n_{BC} > \frac{N}{2}) \\ X & \text{Otherwise} \end{cases} \quad (5)$$

$$X = \begin{cases} A & \begin{cases} |A(i, j) - D(i, j)| \\ |B(i, j) - D(i, j)| \end{cases} < T \\ B & \begin{cases} |B(i, j) - D(i, j)| \\ |A(i, j) - D(i, j)| \end{cases} < T \\ U & \text{otherwise} \end{cases} \quad (6)$$

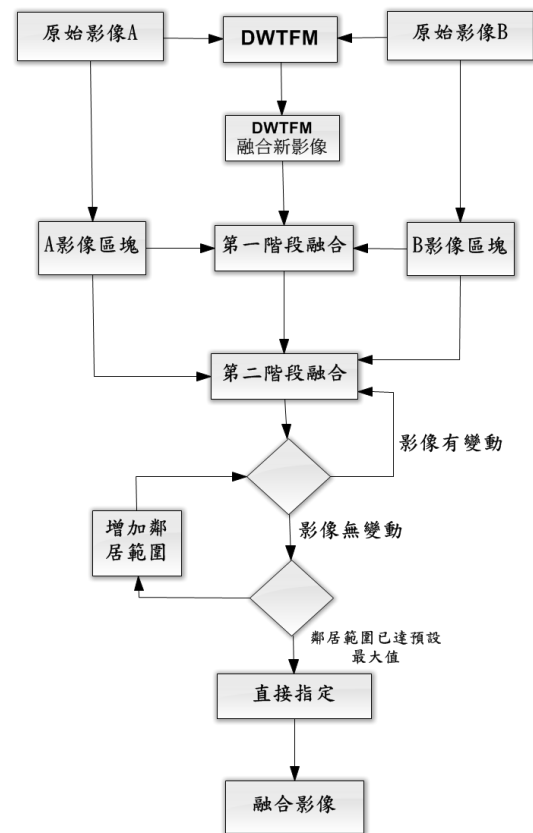
然而有可能 3×3 的區塊無法完成全部的融合, 例如周圍未知區塊過多或是 X 值大於 0.5 所造成無法決定的狀況, 所以可能到達某一階段之後, D 圖不會再變動。這個可能是我們設定鄰居區塊過小, 所以在融合未完成但 D 圖已不再變動時, 我們將範圍擴大, 一次考慮更多的鄰居。實驗時的鄰居區塊的設定接著包括 5×5、7×7、9×9 和 11×11 大小的區塊。但是我們也必須考慮到如果在 11×11 的鄰居區塊仍無法完成融合, 我們就直接進行指定, 公式如下。

$$F(i, j) = \begin{cases} A & N_{AC} \geq N_{BC} \\ B & N_{BC} > N_{AC} \end{cases} \quad (7)$$

2.3 整體方法架構

本研究架構如圖三所示。首先使用 DWTFM 融合方法取得頻率域融合影像, 這個影像已經有了可接受的融合結果, 但是距離實際完美融合仍有不小差距。我們希望融合完成後的影像的像素值是取自於 A 或 B 影像中聚焦清楚的部分。所以接下來便基於這個影像參考與 A 和 B 圖的距離以及區域聚焦性進行本研究的空間域運算, 整個過程分為兩個階段來進行處理, 第一階段和原圖十分接近的像素直接進行調整與原圖一致。第二階段, 參考鄰居區塊找出最可能的聚焦原圖進行修改, 若無任何修改則擴

大鄰居區塊, 再進行相同運算直到 11×11, 若 11×11 鄰居區塊仍無法完成融合的話, 就直接根據與 A 和 B 原圖的距離直接指定融合對象。



圖三 系統架構

3. 實驗與討論

本研究將使用個人電腦 (Intel(R) Core™ i7-2600 CPU, 4GB RAM) 進行操作, 並利用程式工具 Matlab 2011a 中所提供的影像處理工具為基礎, 撰寫影像融合之程式進行本文所提出的實驗。

本研究使用 3 張常使用在文獻中的影像來進行實驗, 分別為花朵、研究室、瓶子 (大小分別為 512×512、640×640、512×512)、與 8 組本研究自行拍攝影像 (大小 1520×1520)。

本研究將使用 8 種方法與本研究使用的方法進行比較。

3.1 客觀的評比方法

本文使用一種評比方法, 因為依照人的肉眼來觀察無法以客觀的角度來分析, 所以本文將使用 PSNR 來進行客觀的評比。

PSNR 中文名稱是峰值訊噪比, 經常被用來作為影像壓縮領域中信號重建質量方法 [7][8][19], 它主要是利用影像信號的最大值與影像中雜訊的筆直做為評估標準, 公式如下:

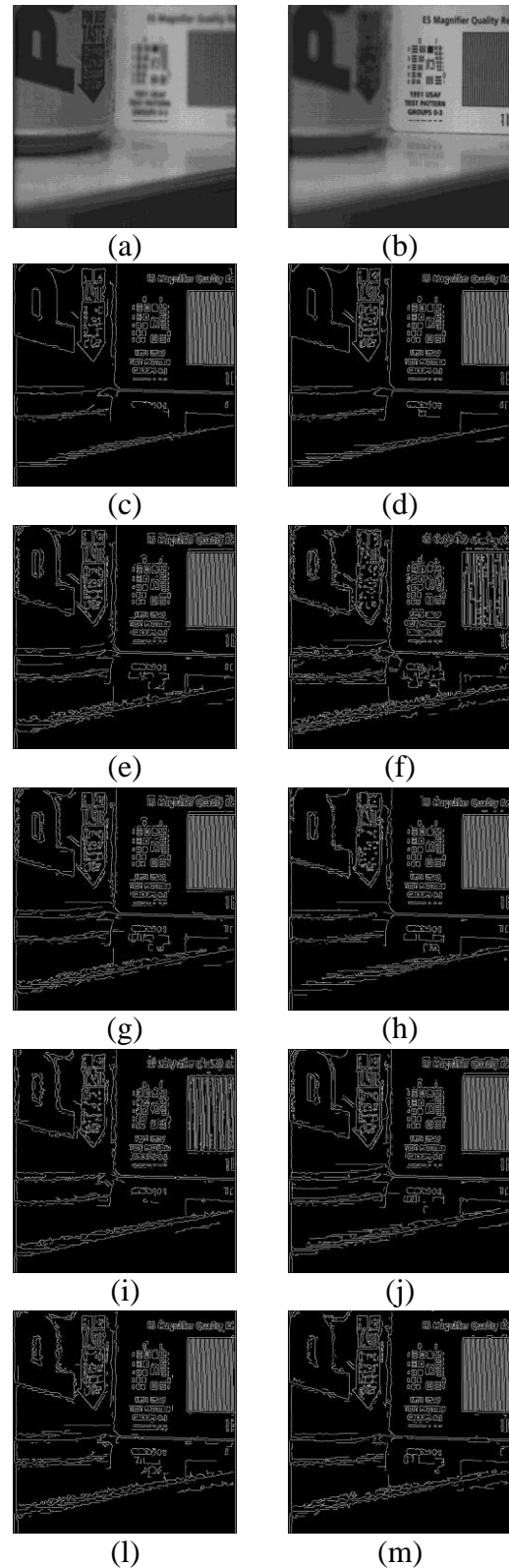
$$PSNR = 10 \log \frac{(2^n - 1)^2}{MSE} \quad (8)$$

這裡的 n 代表是影像點顏色的最大值，例如影像用 8 位元型態來表示，那麼 $2^n - 1$ 就是 255。

這三種方法皆是值越大融合效果越佳。

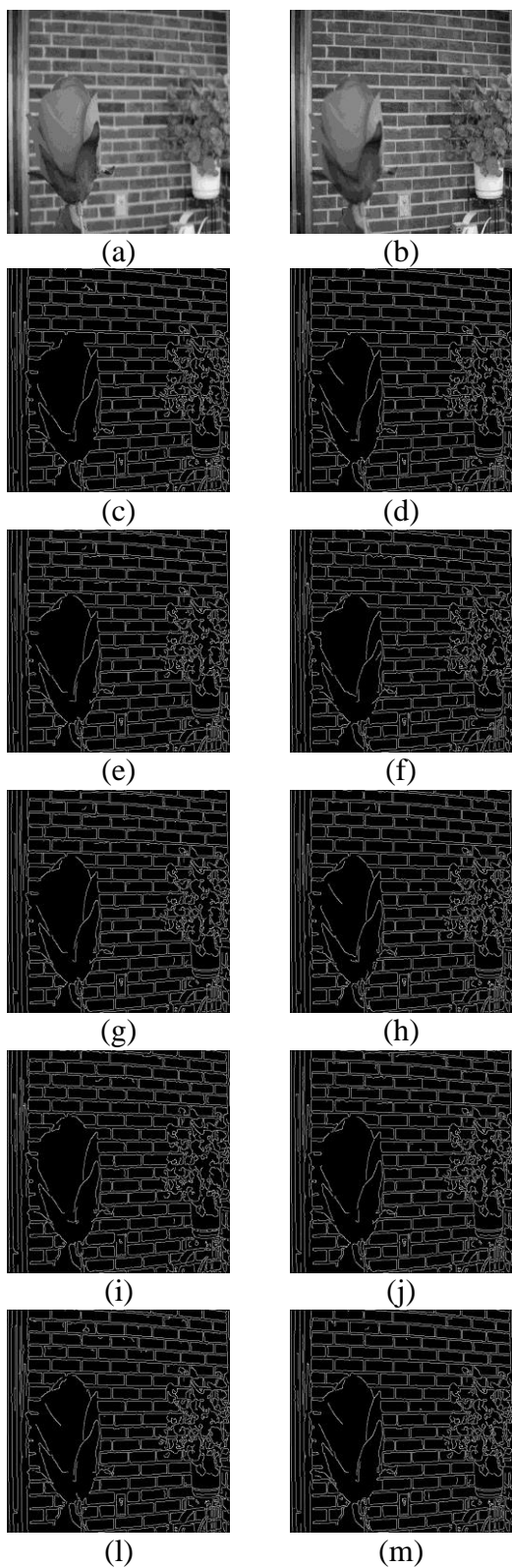
3.2 實驗結果

本研究將使用 6 種空間域方法、1 種頻率域、2 種頻率結合空間域方法進行比較，空間域的區塊大小本研究使用 4×4 的區塊大小進行實驗，PSNR 所使用的標準圖是使用人工的方式進行切割出來的，實驗結果如下：



圖四 使用二值化比較瓶子融合圖

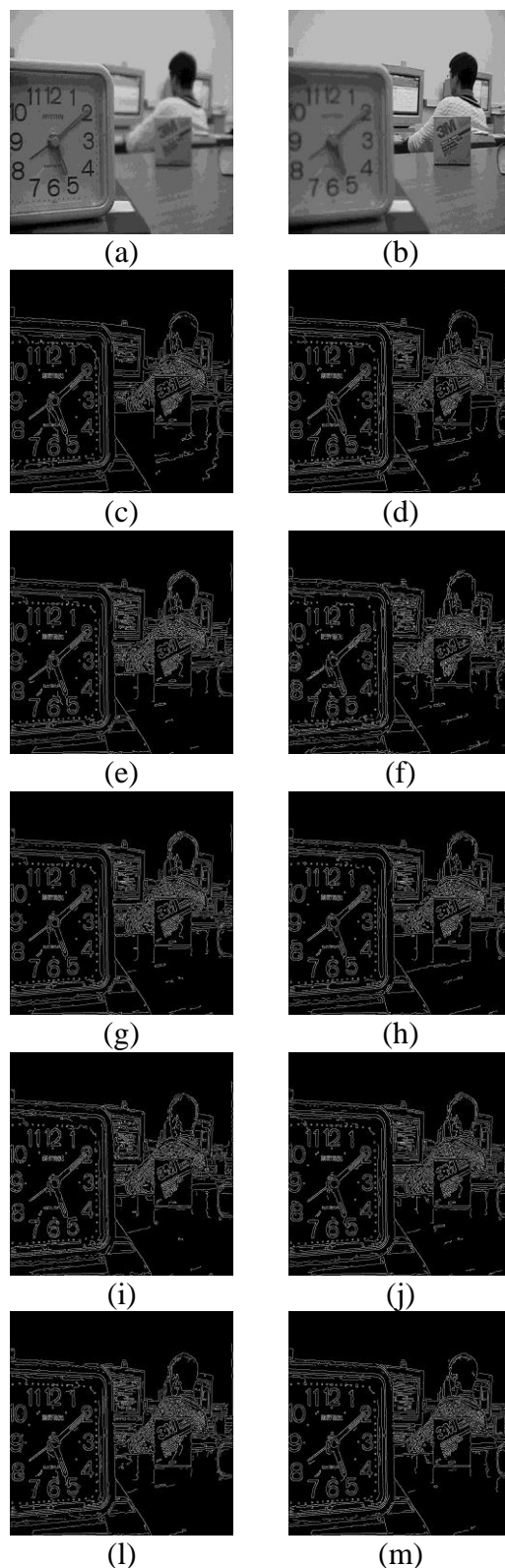
(a) 為原圖前景聚焦、(b) 為原圖後景聚焦、(c) 為 DCT+STD 方法、(d) 為 DWT+STD 方法、(e) 為 EOG 方法、(f) 為 EOL 方法、(g) 為 SF 方法、(h) 為 SML 方法、(i) 為 Tenengrad 方法、(j) 為 Variance 方法、(k) 為 DWTFM 方法，(m) 為本研究方法。我們發現 DCT 與 DWT 方法中的 "P" 字相當明顯，其餘的方法都有嚴重的鋸齒狀，但是以 (i) 來看原本鋸齒狀相當嚴重但是本研究將 DWTFM 方法進行改善後鋸齒狀有明顯的減少。



圖五 使用二值化比較花朵融合圖

(a)為原圖前景聚焦、(b)為原圖後景聚焦、(c)為DCT+STD方法、(d)為DWT+STD方法、(e)為EOG方法、(f)為EOL方法、(g)為SF方法、(h)為SML方法、(i)為Tenengrad方法、(j)為Variance方法、(k)為DWTFM方法、(m)為本研究方法。在花朵這組影像來看我們發現差異度不大，但是我們發現圖(l)上的磚

塊上有許多的雜質，在圖(m)上雜質減少許多。



圖六 使用二值化比較實驗室融合圖

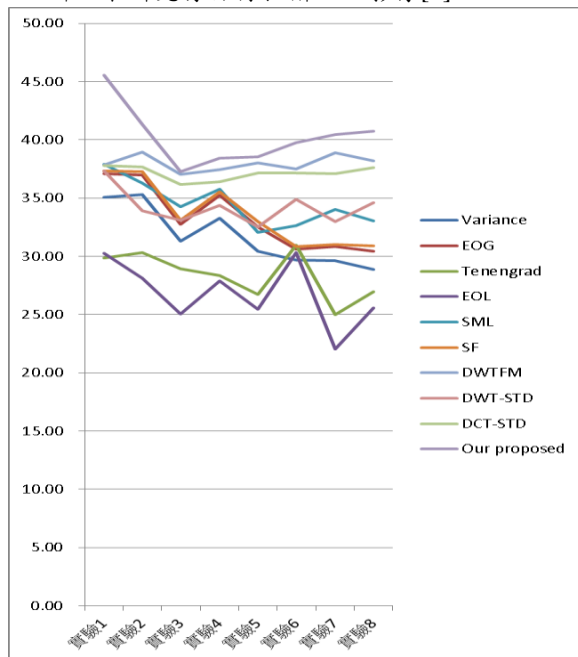
(a)為原圖前景聚焦、(b)為原圖後景聚焦、(c)為DCT+STD方法、(d)為DWT+STD方法、(e)為EOG方法、(f)為EOL方法、(g)為SF方法、(h)為SML方法、(i)為Tenengrad方法、(j)為Variance方法、(k)為DWTFM方法、(m)為本研究方法。我們發現在此組

有一個共通問題，在人的頭髮上有嚴重的鋸齒狀，但是圖(c)圖、(d)圖、(l)與圖(m)並沒有嚴重的鋸齒狀，我們在往開鐘查看發現圖(c)圖與(d)圖在六的上方有不明的圓圈但是本研究的影像沒有出現。

表一 三組影像數據圖

方法	花朵	研究室	瓶子
Variance	30.3994	32.8263	35.2427
EOG	32.6445	34.5926	35.8353
Tenengrad	29.3092	29.7304	30.2021
EOL	27.9968	28.1862	29.2686
SML	37.4488	35.9875	36.5549
SF	32.7812	35.0717	35.9046
DWTFM	34.71172	35.66214	37.06036
DWT-STD	37.2233	36.6976	36.6732
DCT-STD	36.2601	34.6764	37.1228
本研究方法	39.26324	38.22234	37.26775

由表一顯示，本研究的方法優於上述各種方法，但是由於影像過少不足以代表本研究的方法優於其他方法，所以本研究有自行拍攝 8 組影像[1]



圖七 研究實驗圖表

從這八組實驗中我們發現本研究優於DWT-STD與DCT-STD並且改善了DWTFM，但是如果DWTFM一開始融合完成後有大部分模糊區塊的話，本研究方法有可能會判斷成模糊區塊，所以未來需要再增加判斷機制。

4. 結論

本研究提一個方法，主要針對相同場景但不同聚焦的影像進行多焦距影像融合，我們提出來一個新的影像融合的方法，結合了頻率域和空間域方法的優點，產生更精準的融合結果。

在實驗結果中我們的方法確實優於 DWT-STD 及 DCT-STD 並且解決了 DWTFM 方法上產生像素範圍的變動，也優於其他的方法，本研究方法避免空間域的範圍性選取功能採用單點選取功能，這樣可以避免將空間域選擇到模糊地方的問題，但是正如剛剛所提到的在 DWTFM 上如果融合後就有問題的話，在接下來的判斷就都錯誤，所以在未來本研究要再加入各種措施，讓他在選擇區域點的時候有更完善的機制。

參考文獻

- [1] 本研究研究圖組 (https://drive.google.com/folderview?id=0B_NJ_Irg33S1sUVRqNmo5WDVVZwC&usp=drive_web)。
- [2] 吳彥緻，戴紹國，影像融合運用於肝癌病理切片自動診斷之研究，朝陽科技大學資訊管理系碩士論文，2010，台中。
- [3] 周政璋，李朱慧，植基於頻率域分析影像融合之研究，朝陽科技大學資訊管理系碩士論文，2010，台中。
- [4] Burt, Peter J., and Raymond J. Kolczynski. "Enhanced image capture through fusion," Computer Vision, 1993. Proceedings, Fourth International Conference on, IEEE, 1993. pp.173-182.
- [5] Burt, Peter, and Edward Adelson. "The Laplacian pyramid as a compact image code," Communications, IEEE Transactions on, Vol.31, No.4, 1983, pp.532-540.
- [6] Chipman, Laure J., Timothy M. Orr, and Lewis N. Graham. "Wavelets and image fusion," SPIE's 1995 International Symposium on Optical Science, Engineering, and Instrumentation. International Society for Optics and Photonics, 1995, pp.208-219.
- [7] H.R. Sheikh and A.C. Bovik, "Image Information and Visual Quality," IEEE Trans. Image Processing, Vol.15, No.2, 2006, pp.430-444.
- [8] H.R. Sheikh, M.F. Sabir and A.C. Bovik, "A Statistical Evaluation of Recent Full Reference Image Quality Assessment Algorithms," IEEE Trans. Image Processing, Vol.15, No.11, 2006, pp.3440-3451.
- [9] Hill, Paul R., Cedric Nishan Canagarajah, and David R. Bull. "Image fusion using complex wavelets," *BMVC*. 2002.pp.1-10.
- [10] Huang, Wei, and Zhongliang Jing. "Evaluation of focus measures in multi-focus image fusion,"

- Pattern Recognition Letters, Vol.28, No.4, 2007, pp.493-500.
- [11] Koren, Iztok, Andrew Laine, and Fred Taylor. "Image fusion using steerable dyadic wavelet transform," Proceedings, International Conference on, Vol.3, IEEE, 1995. pp.232-235.
- [12] Li, Hui, B. S. Manjunath, and Sanjit K. Mitra. "Multisensor image fusion using the wavelet transform," Graphical models and image processing, Vol.57, No.3, 1995, pp.235-245.
- [13] Li, Hui, B. S. Manjunath, and Sanjit K. Mitra. "Multisensor image fusion using the wavelet transform," Graphical models and image processing, Vol.57, No.3, 1995, pp.235-245.
- [14] Nayar, Shree K., and Yasuo Nakagawa. "Shape from focus," *Pattern analysis and machine intelligence, IEEE Transactions on*, Vol.16, No.8, 1994, pp.824-831.
- [15] Petrovi, V., and C. Xydeas. "On the effects of sensor noise in pixel-level image fusion performance." *Information Fusion*, 2000. FUSION 2000. Proceedings of the Third International Conference on, Vol.2, 2000, pp.WEC3/14 - WEC3/19
- [16] Qu, Guihong, Dali Zhang, and Pingfan Yan. "Information measure for performance of image fusion," *Electronics letters* Vol.38, No.7, 2002, pp. 313-315.
- [17] Toet, Alexander, Lodewik J. Van Ruyven, and J. Mathee Valetton. "Merging thermal and visual images by a contrast pyramid," *Optical Engineering*, Vol.28, No.7, 1989, pp.287789-287789.
- [18] Unser, Michael. "Texture classification and segmentation using wavelet frames," *Image Processing, IEEE Transactions on* Vol.4, No.11, 1995, pp.1549-1560.
- [19] Y. Chai, H. Li and X. Zhang, "Multifocus image fusion based on features contrast of multiscale products in nonsubsampling contourlet transform domain," *Journal of Optics*, Vol. 123, No. 7, 2012, pp. 569-581.
- [20] Yocky, David A. "Image merging and data fusion by means of the discrete two-dimensional wavelet transform, " *JOSA A*, Vol.12, No.9, 1995 pp.1834-1841.
- [21] Zhang, Zhong, and Rick S. Blum. "A categorization of multiscale-decomposition-based image fusion schemes with a performance study for a digital camera application," *Proceedings of the IEEE*, Vol.87, No.8, 1999, pp.1315-1326.