

多媒體雕塑工作室：智慧型音視訊編修處理系統

視訊滌洗、偏好色調調整及場景內容分離萃取

計畫編號：98-2221-E-009-110-MY3

執行期限：2009.08.01 至 2012.07.31

主持人：王聖智 (交通大學電子工程系教授)

計畫參與人員：曾禎宇、許家豪、鄭心憫、陳柏翔、宋秉修、劉彥廷 (交通大學電子所研究生)

中文摘要

在本計畫中，我們提出了一套影像色調自動重建演算法，針對拍攝影像時過暗或是過亮的影像，採用局部性 gamma 調整，能夠依據影像本身亮度，自動調整至偏好的色調，同時增強影像對比。由於調整亮度之 gamma 圖必須與影像本身亮暗配合，否則容易出現調整的瑕疵現象，我們採用 matting Laplacian 作為估算的基礎，透過最佳化方法計算最佳之 gamma 值。為了降低運算複雜度，我們發展了一套快速 cell-based matting Laplacian 運算將相似像素合併進行 gamma 估算，最後再取得對應的 gamma 影像進行色調調整。

除此之外，本計畫提出了兩套影像深度估算演算法，目的是希望透過影像深度讓未來有更便利的影像編輯方式。為了估測影像深度，我們發展兩種不同的運算類型，第一套方法是用於視訊類型影片，利用時間累計影像資訊，透過物體間的遮蔽關係估算相對深度。另一套方法是利用拍攝多重對焦，透過對焦與深度關聯性估測影像深度。未來便可透過深度資訊進行影像編修。

關鍵詞：影像增強、去背拉普拉斯矩陣(matting Laplacian)、色調重建、深度重建、多重對焦。

Abstract

In this project, in order to improve the visual quality of over-exposed or under-exposed

images we present an automatic tone mapping algorithm based on a spatial-varying gamma adjustment. The proposed method aims to map the tone values of the input image into a preferred mud-tone value through a gamma map for enhancing the detailed contrast for those overly bright /dark image contents. The corresponding gamma map shall be aligned with the image content. Otherwise a certain adjusting artifacts will be introduced into the output image. To achieve that, we present a global optimization framework with the matting Laplacian for the estimation of the gamma map. In addition, in order to reduce the required computations we present a cell-based framework which groups pixels with similar colors. After the grouping, the gamma map can be computed in a more efficient way.

Besides, this project presents two schemes of depth reconstruction based on different input information. The first scheme is based on occluding cue extracted from a video acquired with a still camera. Temporally, moving objects provide occluding information about the scene, thus we can reconstruct the relative depth based on the occluding cue between objects. The second scheme is developed with a multi-focused image sequence where a number of images are acquired with varying focus

settings. Based on the focus measurement, we can reconstruct the corresponding depth image. We expect that the depth images will be useful for advanced image editing.

Keywords: contrast enhancement, matting Laplacian, tone reproduction, depth reconstruction, multi-focus.

1. 簡介

1.1 自動色調重建

在一般攝影過程，曝光亮度是影響所得影像之品質關鍵因素，過暗或是過亮的影像都會造成影像對比及細節之損失，本研究提出之演算法透過區域性之 γ 調整，改善影像亮度偏差問題，增強色彩飽和度，同時提昇影像局部對比及細節。

本篇論文所提出的 γ 調整技術主要會根據影像本身亮度自動進行調整，其核心概念在於將影像亮度值經過 γ 重新調配至適當之亮度。在[1]中，其作者提出區域性 γ 調整增強影像對比之技術，本研究主要以其概念為基礎發展。而當中調整之 γ 必須符合影像內容進行一致性的調整，否則容易出現調整後的瑕疵問題，如對比增強常見的光暈問題，或是高對比衰減問題。為了讓估算之 γ 配合影像內容，我們採用 Matting Laplacian (ML) 作為估算 γ 的核心基礎。ML 是由 Levin 等作者於[2]所提出解決自然影像去背問題，當中他們建立影像像素點間的關聯性作為前景背景資訊分離的依據，而像素的關聯性會建立於 ML 中。

在本研究中，由於亮度調整必須考量影像像素之間的關聯性進行一致性調整，以避免產生調整瑕疵問題，在此我們便以 ML 作為基礎進行整張影像之 γ 估算，以此所估算之

γ 能夠密切貼合影像內容，可有效改善亮度問題並增強影像對比，同時避免瑕疵。

另一方面，為了提升處理效能，我們利用影像像素點相似特質發展一套 cell-based matting Laplacian，事先將影像像素進行群聚分類，將多像素點整合為 cell 之後再針對 cell 進行估算，藉此省略大量的運算，最後只需再將像素所對應的 γ 值透過 cell 進行還原估算即可獲得高解析度的調整影像。

1.2 影像深度估算

針對深度估算我們提出兩套系統架構，第一種系統架構主要針對視訊影片估測場景深度，固定攝影機拍攝一段影片之後利用影片中物體移動後的遮蔽關係判斷影片中物體的相對深度。此外，我們會採用 Hoiem 在[6]所提出的單張影像深度估計演算法，與我們透過時間軸分析遮蔽性的方法結合，提供更準確的深度估測結果。

另一方面，我們針對多重對焦影像發展一套深度估測演算法，此方法利用單台相機調整不同對焦深度，拍攝多張不同對焦影像。再透過對焦程度分析影像中各物件聚焦對應深度，然而由於影像內容常出現不易對焦物件表面而造成估測上的誤判。為了克服這樣的問題，我們提出一套全域最佳化演算法，此最佳化主要是以事後機率最大化模型為基礎。在此機率模型中，由於影像深度有著空間上連續的特質，因此我們採用 ML 建立後述機率中的事前機率模型。為了降低運算複雜度，我們也採用了 cell-based ML 估算全域最佳解，再透過內插計算取得高解析度深度影像。

2. 自動化色調調整

2.1 區域性 gamma 調整

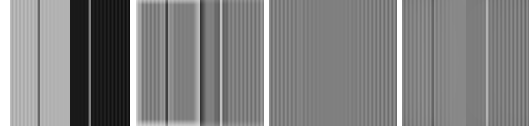
為了改善拍攝影像光線不適當所造成的影像品質衰減問題，我們提出採用區域性自動化 gamma 調整技術，將整張影像 $I(x,y)$ 各像素經由個別之 gamma map $\gamma(x,y)$ 調整為輸出影像 $O(x,y)$ ，表示如下

$$O(x, y) = I(x, y)^{\gamma(x,y)} \quad (1)$$

當中我們希望透過 gamma 調整能將影像亮度調整至適當亮度值，因此我們將影像色彩空間轉至 HSV 色彩空間，單純對於影像亮度 $Y(x,y)$ 進行 gamma 調整，此調整核心觀念為透過 gamma 將亮度調整至適當之中間亮度值 Y_0 ，如下所示

$$Y(x, y)^{\gamma_b(x,y)} = Y_0. \quad (2)$$

其中 γ_b 為將影像調整至中間亮度之 base gamma map。單純將影像亮度調整至中間亮度將會損失影像細節對比，因此在[1]中，作者發現損失之細節對比可透過移除 gamma map 中的對應細節而得到還原。為了移除 gamma map 中的細節，最簡單的方式是採用 Gaussian 濾波，但是 Gaussian 濾波會造成 gamma map 在影像亮暗劇烈變化邊界產生模糊，造成調整後的影像在亮暗交界處發生光暈瑕疵。因此在[1]中，作者採用 Bilateral 濾波[3][4]方式避免此問題。Bilateral 主要觀念在於模糊計算過程考量影像本身亮度，避開模糊劇烈變化之亮度，藉此排除光暈問題。如圖一所示，原始影像(a)左半部過亮，右半部過暗。透過 gamma 調整可使亮度調整至適當亮度同時增強細節。Bilateral 雖然可以避開亮度劇烈變化邊界產生光暈瑕疵，但卻會帶來另外的副作用，抑制高對比細節，而造成某些高對比細節喪失問題。因此在本研究中，我們採用 matting Laplacian 運算藉此調整影像亮暗並提升細節對比。



(a) Input (b) Gaussian (c) Bilateral (d) Ours

圖一、採用不同 gamma 調整之結果

2.2 最佳化 gamma 估算

在本研究中，與[1]不同我們並不直接對 gamma map 進行平滑濾波，主要原因在於 gamma 並非線性，受到影像亮暗影響變化甚鉅。因此本研究我們將對影像亮度進行平滑估算，取得一平緩之亮度影像 Y_s ，將影像中的微弱細節移除，再以此亮度影像估測對應之 smoothed gamma map，當中的影像細節便可被分離，避免調整後之細節衰減。傳統採用區域濾波方式如 bilateral 濾波並無法根據影像本質進行適應性的調整，因此可能造成前面所描述的高對比細節喪失問題。因此本研究採用全域性最佳化之方式，將估測平緩影像 Y_s 問題轉為一全域最佳化估測問題，如下所示

$$E(Y_s) = \lambda_\gamma (\mathbf{Y}_s - \tilde{\mathbf{Y}}_s)^T (\mathbf{Y}_s - \tilde{\mathbf{Y}}_s) + \mathbf{Y}_s^T \mathbf{L} \mathbf{Y}_s. \quad (3)$$

在式(3)中，前半段為資料回歸項，後半為平滑項，在資料回歸項中，我們首先利用 Gaussian 產生一模糊影像 $\tilde{\mathbf{Y}}_s$ 做為參考目標，我們希望最後所找到的 Y_s 能與之相近。然而為了避開強烈變化邊界產生模糊，在式(3)後半段我們採用了 matting Laplacian 限制平滑效果能貼近影像本身內容，matting Laplacian 矩陣 L 之定義為 $L=D-A$ ，其中 D 為 degree 矩陣而 A 為 affinity 矩陣，affinity 矩陣描述影像像素點間之連續性，如像素點 i 與 j 之連續性為 $A(i,j)$ ，其定義如下

$$\mathbf{A}(i,j) = \sum_{k|(i,j) \in w_k} \left(\frac{1}{|w_k|} \left(1 + (\mathbf{I}_i - \mu_k)^T \left(\Sigma_k + \frac{\varepsilon}{|w_k|} \mathbf{U}_3 \right)^{-1} (\mathbf{I}_j - \mu_k) \right) \right) \quad (4)$$

而 degree 矩陣 \mathbf{D} 為一對角矩陣，定義如下

$$\mathbf{D}(i,i) = \sum_{j=1}^N \mathbf{A}(i,j) \quad (5)$$

而式(3)後半之平滑項可進一步拆解為

$$\mathbf{Y}_s^T \mathbf{L} \mathbf{Y}_s = \sum_{i,j} \mathbf{A}(i,j) \|\mathbf{Y}_s(i) - \mathbf{Y}_s(j)\|^2 \quad (6)$$

當影像像素相近 $\mathbf{A}(i,j)$ 值較大，則會抑制 \mathbf{Y}_s 之變動量，透過最小化(3)可以找到較佳之平滑影像，其解如下所示。

$$(\mathbf{L} + \lambda_\gamma \mathbf{U}) \mathbf{Y}_s = \lambda \tilde{\mathbf{Y}}_s \quad (7)$$

取得 \mathbf{Y}_s 之後，我們根據此平滑亮度定義其對應之平滑 gamma map 如下

$$\gamma_s(x,y) = \frac{\log(Y_0)}{\log(Y_s(x,y))} \quad (8)$$

接著我們定義用來重塑亮度影像的 gamma map γ_{shaped} ，此 γ_{shaped} 之定義主要是將 γ_b 中關於影像細節的部分扣除，根據[2]所提出的 gamma map 處理發現，當扣除 gamma 中的細節會使所轉換出的影像細節對比增強。

$$\gamma_{shaped}(x,y) = \max(0, (\gamma_b(x,y) - c_1 \gamma_{detailed}(x,y))) \quad (9)$$

$$\text{where } \gamma_{detailed}(x,y) \equiv \gamma_b(x,y) - \gamma_s(x,y).$$

而最後調整所用之 gamma 定義如下

$$\gamma(x,y) = (1-w)\gamma_0 + w\gamma_{shaped}(x,y) \quad (10)$$

其中 w 為控制調整程度之參數，當 $w=1$ 則影像調整將會採用 γ_{shaped} 重塑影像亮度，而當 $w=0$ 則不會進行調整，其中 $\gamma_0=1$ 。最後我們將輸入影像轉至 HSV 色彩空間後，將亮度值透過 gamma 調整取得輸出影像之亮度值 Y_{output}

$$Y_{output}(x,y) = Y(x,y)^{\gamma(x,y)} \quad (11)$$

再轉回 RGB 色彩空間即可獲得輸出結果。

2.3 高效能運算架構

在前一小節中，我們為了估算平滑亮度影像採用了採用全域最佳化的方式，然而此方法是將整張影像像素點同時考慮，運算量過高，不利於一般應用。為了降低運算複雜度，我們提出一套高效能運算方法，此方法主要利用影像相鄰像素之相似性，將相似之像素點群聚成為 cell，而整個最佳化運算將從原始的像素為基礎的計算轉化為以 cell 為基礎的計算方式。由於 cell 的數目通常遠小於像素數目，運算量可大幅降低。

為了降低運算量並維持影像細節表現，我們將影像轉至一高維度空間，此空間由影像空間域與色彩域所構成。在此空間中，我們採用網格進行降取樣，在同一個網格點中的像素將會共同以 cell 表示。建立了像素轉換至 cell 之對應關係之後，我們便可針對 cell 進行最佳化亮度估算，如下式所示

$$(\mathbf{L}_c + \lambda_\gamma \mathbf{A}_c) \mathbf{Y}_c = \lambda \bar{\mathbf{Y}}_c \quad (12)$$

其中 \mathbf{L}_c 為 cell 所對應的 matting Laplacian，它是由原始像素之 \mathbf{ML} 矩陣經過 pixel-to-cell 轉換矩陣 \mathbf{m} 之壓縮所得，如下所示

$$\mathbf{L}_c = \mathbf{m}^T \mathbf{L} \mathbf{m} \quad (13)$$

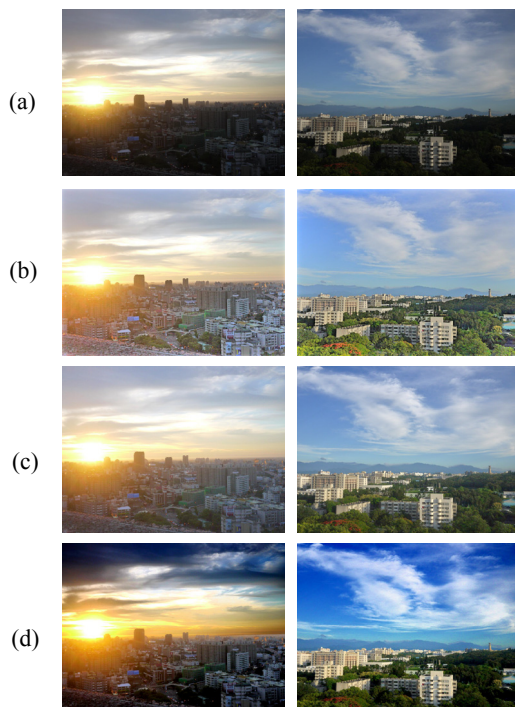
透過式(12)所估測之 \mathbf{Y}_c 為 cell 所對應的亮度值，我們進一步採用內插方法便可還原出高解析度影像之平滑亮度影像。這樣的處理架構可有效將原本最為複雜的最佳化運算以較低運算的方式達成。

2.4 實驗結果

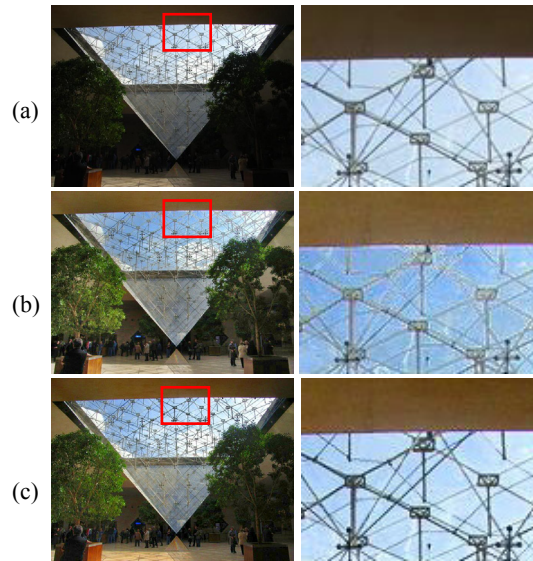
我們將所提出之演算法套用至自然影像進行調整，我們針對拍照環境亮度動態對比過高之情況，如圖二所示，畫面同時出現過亮及過暗之內容。我們比較現存相似演算法與我們所提出的方法結果比較。對於過亮的部分，傳統方

法並無法有效改善，而我們的處理可以有效提升對比與色彩鮮豔度。而過暗部分，我們也能有效提升其細節強度，同時避免過度調整，而造成影響影像內容之自然性。

針對高對比影像細節調整，傳統方法容易為了提升過暗區域的對比而犧牲高對比部分，造成高對比細節的喪失，如圖三(b)所示。而經過我們的 matting Laplacian 結構，可有效保留高對比細節如圖三(c)所示。



圖二、採用不同演算法調整之結果，(a)原圖、(b)使用[4]的結果、(c)使用[5]的結果、(d)我們方法的結果



圖三、採用不同演算法調整之結果，(a)原圖、(b)使用[1]的結果、(c)我們方法的結果

3. 深度影像估計

影像分層往往是進行影像編修的一項重要技術，為了讓影像能夠依據場景分層，深度資訊通常是一項重要資訊，若能取得影像深度，在編修時便可方便分離或是調整物件。為了估測影像深度，我們提出兩種不同架構針對不同影像內容進行處理。第一套架構主要針對固定場景之影片，利用影片中物件的移動過程，物件互相遮蔽的關係分辨相對深度。第二套架構利用多重對焦的影像序列判斷對焦與深度的關係，建立深度影像。下面我們將針對這兩套架構分別介紹。

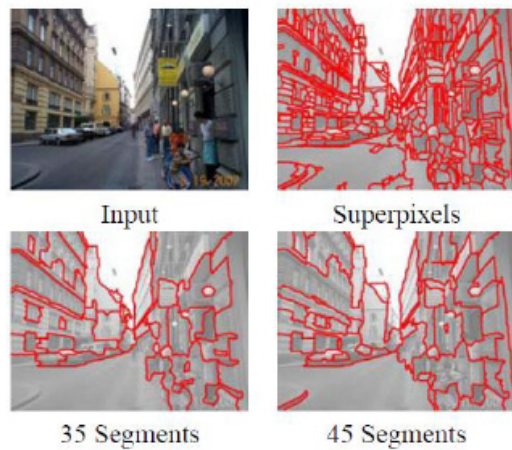
3.1 靜態攝影機拍攝影片之相對深度重建

在過去靜態影像之深度估測 Hoiem 在[6]中提出以影像分割邊界遮蔽性為基礎的相對前後深度分析方式，首先他們會對輸入影像產生各種不同數量的分割，尋找可能的影像區塊邊界，利用機器學習事先學習區塊互相的關聯性，將一些屬於同一物件的邊界去除後他們可以取得較符合物體前後關聯性的邊界，如圖五

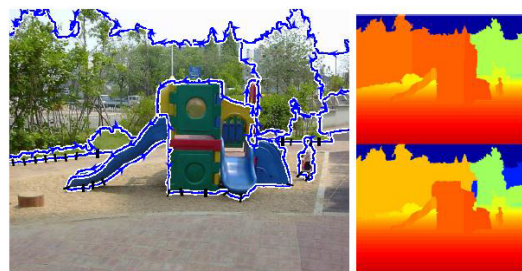
所示，他們利用影像分割的邊界，再透過機器學習判斷區塊間的前後關係，藉此建立相對深度圖。然而單張影像估測深度的準確性與穩定度仍然是相當困難，常會因為影像內容不同而有可能發生誤判，

因此在本計畫中，我們提出一套針對靜態攝影機，拍攝一段累積時間的影片，透過畫面中移動物體與場景物體互相遮蔽的現象，建立出分割遮蔽

物件的邊界，再與[6]的單張影像深度估測演算法結合，提供較穩定的效果。



圖四、Hoiem在[6]中所採用的多重分割。



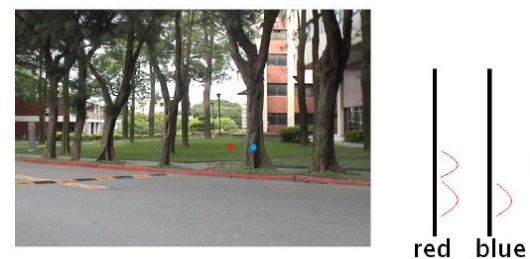
圖五、Hoiem在[6]中所得到的物件分割與深度估測結果。

本演算法採用固定攝影機拍攝一段時間，建立場景的深度影像，我們以圖六進行說明本系統的主要概念，圖六上列影像，中間影像出現一車被中間的樹幹所遮蔽，而車子遮蔽後方的建築與馬路，因此提供了樹木與後面背景的前後資訊。圖六下列影像中，中間影像出現了一個

行人遮蔽了中間的樹幹，因此可以判斷樹幹與行人走過路線的前後關係。我們便是利用這種隨著物件移動產生的遮蔽特質判斷遮蔽邊界，提供相對深度分析時所需要的資料。



圖六、場景移動物件遮蔽特性說明。



圖七、遮蔽物分析。

針對同一場景拍攝一段時間後，我們分析隨著時間經過的物件接地點位置統計值，當畫面中出現遮蔽物件，如圖七中的樹幹。統計行人走過的接地點分佈，圖七中紅點因為沒有遮蔽物出現，因此行人的接地點位置分佈可能會出現在馬路或是人行道上，呈現兩個可能的分佈統計區間。相對的，圖七中的藍點落在樹幹上，因此統計接地點只可能發生在樹幹前的行人，而不會看到樹幹後的行人，因此分佈圖上只有單一個分佈區間。我們藉由這樣的分佈關係判斷出影像的遮蔽物。

透過這樣的方式，我們可以偵測出遮蔽物件的分佈狀況，如圖八所示，依此判斷遮蔽物件的邊界位置，我們將此遮蔽物件的邊界與原始影像採用硬分割的分割邊界整合，可以取得較穩定的分割區塊。接著我們再針對這樣的分割區塊進行深度估算。

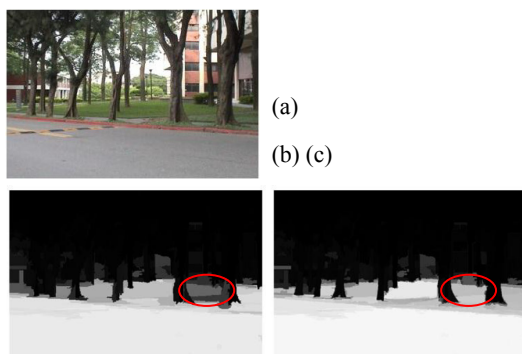


圖八、遮蔽物邊界分析。

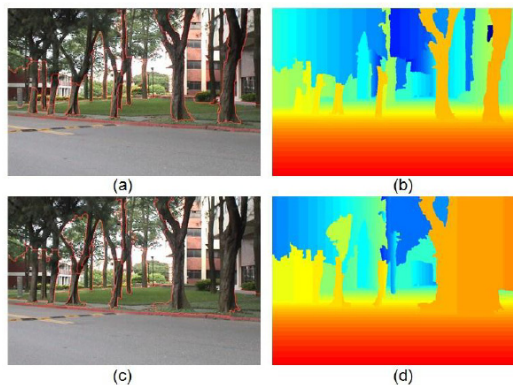
有了比較準確的分割結果，我們再採用 Hoiem 在[6]中所提出的單張影像深度估測演算法，將影像中的地面，天空與其他物件分離，再分別對各種類別進行深度的估算。此估算方式是利用機器學習，學習各種區塊對於各種類別可能的特徵值關聯性，判斷出各種類別之後再判斷區塊間的相對關係，進而判斷深度。

在圖九中我們比較使用我們方法對於深度估算的改善。圖中紅框標示地面偵測時，在圖九(b)使用[6]的方法會因為影像分割判斷誤判而導致錯誤估測。圖九(c)表示使用我們方法後可以排除原本的誤判。

最後的深度估算結果如圖十所示，途中所示透過我們所提出的方法，可以有效排除原本因為影像分割分析不正確所造成的誤判，提升深度估測的準確度。



圖九、地面影像偵測。(a)原圖，(b)使用[6]之估算結果，(c)使用我們方法的結果。



圖十、遮蔽物邊界偵測與深度估算結果。(a)所提出方法的遮蔽物邊界偵測，(b)所提出方法的深度估算結果，(c) [6]的遮蔽物邊界偵測，(d) [6]的深度估算結果。

3.2 多重對焦影像深度重建

在本系統中，我們利用改變攝影機的對焦距離，拍攝出數張不同對焦的影像，接著在影像中各個位置判斷其所屬的正確對焦範圍，透過正確對焦的位置重建影像深度資訊。在這樣的過程中，有個最困難的挑戰在於影像中有許多位置是屬於難以量測對焦值的區域，例如沒有紋理的表面，在處理這些表面往往會由於對焦量測誤判而導致深度估測的誤判。傳統區域性的方法[7][8]很容易受到這樣的問題而影響重建深度的準確度。為了克服這樣問題，我們提出一套全域性最佳化方法，我們將深度重建問題以一事後機率模型表示，當中我們會建立深度影像的事前機率模型，結合觀測的多重對焦影像，透過最大化事後機率得到全域最佳解。這樣的方法可以有效改善由於對焦不易判別而造成的錯誤問題。而為了建立事前機率模型，我們採用 matting Laplacian 運算作為背景，而為了提升運算效能，我們提出了一套有效的 cell-based 架構，將影像像素有效率的整合，大幅度降低運算量再還原回高解析度深度影像。

3.2.1 事後機率最大化之深度重建

在我們所提出的系統中，我們將深度重建以一最大事後機率模型表示如下

$$d^* = \arg \max_d \{p(d|I^{\text{set}})\} \quad (14)$$

上式中 d^* 為我們所希望取得的深度影像，它是透過最大化事後機率 $p(d|I^{\text{set}})$ 所取得，其中 I^{set} 為所拍攝的多重影像序列， d 為深度影像值。我們會進一步將事後機率模型以貝氏定理拆解如下

$$p(d|I^{\text{set}}) \propto p(I^{\text{set}}|d)p(d) \quad (15)$$

其中 $p(d|I^{\text{set}})$ 為可能性機率模型， $p(d)$ 為事前機率模型，我們透過可能性機率模型建立所拍攝的多重對焦影像與深度影像之關聯性，而深度影像本身所應有的空間連續性則會被我們建立於 $p(d)$ 。當處理畫面中無紋理的區域時，由於對焦難以辨別，因此系統會自動地以 $p(d)$ 作為深度估算的考量，利用 $p(d)$ 限定深度影像的空間連續性，有效排除誤判情況。接下來，我們將分別介紹(15)中所建立的可能性機率模型 $p(d|I^{\text{set}})$ 與為事前機率模型 $p(d)$ 。

建構可能性機率模型中，我們主要目的是建立觀測影像與實際深度影像之間的相聯性，我們首先會將此關聯性定義為觀測深度與實際深度的關聯性，如下所示

$$p(I_q^{\text{set}}|d_q) \equiv p(\tilde{\mathbf{d}}_q|d_q) \quad (16)$$

當中我們介紹了 $\tilde{\mathbf{d}}_q$ 表示從多重對焦影像序列所觀測的深度向量，這個深度會與真正理想的深度有個偏差，我們將此偏差以 Gaussian 模型建立如下

$$p(\tilde{\mathbf{d}}_q|d_q) \equiv N(\tilde{\mathbf{d}}_q|d_q, \Lambda_q^{-1}), \quad (16)$$

$$-\log p(\tilde{\mathbf{d}}_q|d_q) \equiv (\tilde{\mathbf{d}}_q - d_q)^T \Lambda_q (\tilde{\mathbf{d}}_q - d_q) \quad (17)$$

$$= \sum_{i \in \Omega_q} \lambda_i \|\tilde{d}_i - d_i\|^2. \quad (18)$$

其中每個像素所對應的深度值 d_i 都有一個自己所對應的精準度，這樣的設計在於模擬有些觀測深度可能與實際深度相當貼近，但有些觀測深度可能由於缺乏表面紋理，很容易受到雜訊干擾而偏離實際深度值。我們利用這樣的不確定性來建立觀察值與實際深度的相似可能性，若影像中出現難以對焦的狀況，此時實際深度就不會被鎖定在觀察值附近，之後便可透過實際深度的事前機率模型來估計。

為了建立事前機率模型，我們提出一套學習機制，透過局部的多重影像區塊學習深度影像像素間的關聯性。這個學習主要來自於線性估測，我們假設影像的深度可以由一些影像的特徵向量經過線性組合而得，如下所示

$$d_i^k = [\mathbf{v}_i^k]^T \boldsymbol{\beta} + \beta_0. \quad (19)$$

其中 d_i^k 表示在第 k 張對焦影像所推測在第 i 個像素位置的深度值，這個深度值是利用像素本身的特徵向量 \mathbf{v}_i^k 進行線性組合而得，而 $\boldsymbol{\beta}$ 與 β_0 分別為線性組合的參數。接著我們再將多重對焦影像合併出最後的深度值

$$d_i = \mathbf{p}_i [\mathbf{V}_i^T \quad 1] \begin{bmatrix} \boldsymbol{\beta} \\ \beta_0 \end{bmatrix} \quad (20)$$

$$= [\mathbf{x}_i \quad 1] \begin{bmatrix} \boldsymbol{\beta} \\ \beta_0 \end{bmatrix}, \quad (21)$$

$$\text{where } \mathbf{x}_i = \mathbf{p}_i \mathbf{V}_i^T. \quad (22)$$

其中 $\mathbf{p}_i = [p_i^1, \dots, p_i^K]$ 是合併多重影像的對應

機率向量， $\mathbf{V}_i = [\mathbf{v}_i^1, \dots, \mathbf{v}_i^K]$ 是多重影像的特

徵向量矩陣，透過(21)我們建立了從多重對焦影像預測深度的關係式。接著我們將(20)拓展至一個局部區塊，假設區塊中的像素共用相同

線性組合參數 β 與 β_0 ，我們可以將區塊深度預測表示如下

$$\mathbf{d}_q = \mathbf{X}_q \begin{bmatrix} \beta \\ \beta_0 \end{bmatrix} \quad (23)$$

其中 $\mathbf{d}_q = [dq_1, \dots, dq_j, \dots, dq_w]^T$ 表示一個區塊 q 中各像素深度值組合成的一深度向量，而 $\mathbf{X}_q = [\mathbf{x}'_{q1}, \dots, \mathbf{x}'_{qj}, \dots, \mathbf{x}'_{qw}]^T$ 表示區塊的特徵矩陣。

接著組合參數 β 與 β_0 可以透過最小化誤差方程式

$$E(\beta, \beta_0) = \left\| \mathbf{d}_q - \mathbf{X}_q \begin{bmatrix} \beta \\ \beta_0 \end{bmatrix} \right\|^2 + \lambda_\beta \beta^T \beta \quad (24)$$

經由微分求極值，式(24)的最佳解為

$$\begin{bmatrix} \beta \\ \beta_0 \end{bmatrix} = (\mathbf{X}_q^T \mathbf{X}_q + \lambda_\beta \mathbf{D}_\beta)^{-1} \mathbf{X}_q^T \mathbf{d}_q \quad (25)$$

最後將此組合參數代入(23)，我們可以得到深度估測方程式如下

$$\mathbf{d}_q = \mathbf{W}_q^T \mathbf{d}_q \quad (26)$$

$$\text{where } \mathbf{W}_q = \mathbf{X}_q (\mathbf{X}_q^T \mathbf{X}_q + \lambda_\beta \mathbf{I}_\beta)^{-1} \mathbf{X}_q^T \quad (27)$$

在(27)中 \mathbf{W}_q 為像素間互相預測的組合矩陣，我們可以想像在區塊 q 中的某一深度值可以由區塊內的其他深度值透過 \mathbf{W}_q 合併估算。換而言之，式(26)主要描述同區塊內的像素點彼此深度值的關聯性。我們便進一步將此關聯性建成區塊的深度事前機率模型

$$\begin{aligned} -\log(p(\mathbf{d}_q)) &= \left\| \mathbf{d}_q - \mathbf{W}_q^T \mathbf{d}_q \right\|^2 \\ &= \mathbf{d}_q^T (\mathbf{I}_q - \mathbf{W}_q)^T (\mathbf{I}_q - \mathbf{W}_q) \mathbf{d}_q \\ &= \mathbf{d}_q^T \mathbf{L}_q \mathbf{d}_q. \end{aligned} \quad (28)$$

在(28)中 \mathbf{L}_q 表示區塊的 Laplacian 矩陣。

當我們建立完區塊的可能性機率模型以及事前機率模型，我們便可以建立區塊的事後機率模型，如下所示

$$\begin{aligned} &-\log(p(\mathbf{I}_q^{\text{set}} | \mathbf{d}_q) p(\mathbf{d}_q)) \\ &= (\tilde{\mathbf{d}}_q - \mathbf{d}_q)^T \Lambda_q (\tilde{\mathbf{d}}_q - \mathbf{d}_q) + \mathbf{d}_q^T \mathbf{L}_q \mathbf{d}_q. \end{aligned} \quad (29)$$

最後整張影像的深度事後機率可以假設區塊之間彼此無關，因此全域機率可由統計各區塊而得，如下所示

$$\begin{aligned} &-\log(p(I^{\text{set}} | d) p(d)) \\ &= \sum_{q \in \Omega} -\log(p(\mathbf{I}_q^{\text{set}} | \mathbf{d}_q) p(\mathbf{d}_q)) \\ &= \sum_{q \in \Omega} (\tilde{\mathbf{d}}_q - \mathbf{d}_q)^T \Lambda_q (\tilde{\mathbf{d}}_q - \mathbf{d}_q) + \mathbf{d}_q^T \mathbf{L}_q \mathbf{d}_q \end{aligned} \quad (30)$$

式(30)可以再經過矩陣的整理而得到下式

$$\begin{aligned} &-\log(p(I^{\text{set}} | d) p(d)) \\ &= (\tilde{\mathbf{d}} - \mathbf{d})^T \Lambda (\tilde{\mathbf{d}} - \mathbf{d}) + \mathbf{d}^T \mathbf{L} \mathbf{d}. \end{aligned} \quad (31)$$

此事後機率最佳解可由微分求極值而得，如下

$$(\mathbf{L} + \Lambda) \mathbf{d} = \Lambda \tilde{\mathbf{d}}. \quad (32)$$

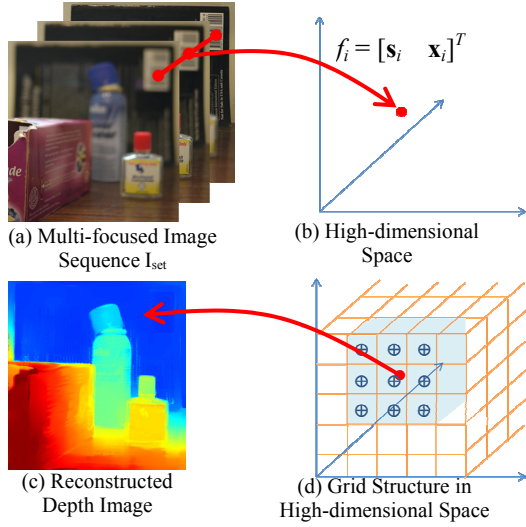
透過這樣的方式，我們可以解(32)而求得最佳的深度估測結果 \mathbf{d} 。

3.2.2 基於網格之快速演算法架構

在前一小節我們介紹了透過事後機率最大化的深度影像重建，然而此架構有一最大問題在於運算複雜度過於龐大，要求得整張影像的最佳解非常困難，為了克服這樣的問題，我們希望發展更有效率的演算法，因而提出了以基於網格的快速演算法架構。

我們所提出的網格演算法主要是根據影像本身有相當高的像素相似性特質，因為相似的像素若能合併計算，預期可以省去相當多的運算量而不影響效果。基於這樣的概念，如圖十一所示，我們所提出的演算法架構是將多重對焦影像轉至一高維度空間中，此高維度空間是根據影像的特徵向量與空間向量所建立，在此高

維度空間我們建立網格將相似的像素點合併，可避免單純使用降取樣可能造成影像細節損失的問題。將影像轉為網格點之後，我們便將問題轉化為估測網格點的深度值，而最後的輸出深度影像，則是由像素點與網格點間的內插運算求得，如此一來，原本最複雜的全域最佳化深度重建會以網格點為基礎，其運算量遠小於以像素為基礎的運算。



圖十一、基於網格之快速深度估算架構。

為了針對網格點進行深度重建，我們將事後機率模型改為針對網格點建立如下

$$p(g|I^{\text{set}}) \propto p(I^{\text{set}}|g)p(g) \quad (33)$$

其中 g 為網格點的深度值，同樣地我們分別對於網格點的可能性機率 $p(I^{\text{set}}|g)$ 與事前機率 $p(g)$ 分別建立模型。在建立可能性機率，我們假設網格點的真实深度與觀測深度之間的關係以 Gaussian 模型表示

$$-\log(p(I^{\text{set}}|g)) = (\tilde{\mathbf{g}} - \mathbf{g})^T \Lambda_g (\tilde{\mathbf{g}} - \mathbf{g}) \quad (34)$$

其中 $\tilde{\mathbf{g}}$ 表示由觀測多重影像所得的網格深度估計值，此觀測值是將原本從影像的觀測值透過像素至網格的轉換矩陣 m 所得，如下所示

$$\tilde{\mathbf{g}}_j = \frac{1}{w_j} \sum_{i=1}^N m(i, j) \tilde{\mathbf{d}}_i, \quad (35)$$

$$\text{where } w_j = \sum_{i=1}^N m(i, j).$$

其中像素 i 若會被轉換至網格點 j ，則 $m(i, j)=1$

反之 $m(i, j)=0$ 。式(34)中的 Λ_g 為一對角矩陣，

其中對角元素 $\Lambda_g(j, j)$ 表示網格點 j 觀察深度與真實深度的準確度，同樣此網格精準度是由像素點的精確度經過網格轉換而得。

$$\Lambda_g(j, j) = \frac{1}{w_j} \sum_{i=1}^N m(i, j) \Lambda(i, i). \quad (36)$$

經由式(34)我們可以建立網格點對於觀察深度值的可能性分佈情況。接著我們再建立網格點互相間的事前機率模型。這個建立過程與像素點的學習模型相同，是透過特徵向量的線性組合預測深度值。首先對於各網格點，我們建立它所對應的特徵向量，此特徵向量是由網格點所對應的所有像素特徵向量整合而得

$$\boldsymbol{\varphi}_j = \frac{1}{w_j} \sum_{i=1}^N m(i, j) \mathbf{x}_i. \quad (37)$$

取得了此特徵向量 $\boldsymbol{\varphi}_j$ 之後，我們進一步透過線性組合預測對應深度

$$\mathbf{g}_j = [\boldsymbol{\varphi}_j \quad 1] \begin{bmatrix} \boldsymbol{\beta} \\ \beta_0 \end{bmatrix}. \quad (38)$$

接著為了建立區塊性的深度預測，我們同樣取畫面中的區塊進行深度估測如下所示

$$\mathbf{g}_q = \boldsymbol{\Phi}_q \begin{bmatrix} \boldsymbol{\beta} \\ \beta_0 \end{bmatrix} \quad (38)$$

當中向量 $\mathbf{g}_q = [g_{q1}, \dots, g_{qw}]^T$ 表示在影像區塊 q 中的深度值組合向量，而這些區塊中的像素共用同樣組合參數。 $\boldsymbol{\Phi}_q = [\boldsymbol{\varphi}'_{q1}, \dots, \boldsymbol{\varphi}'_{qw}]^T$ 表示特徵向量組合的特徵矩陣。而組合參數可透過最小化估測誤差所得。最後我們可以得到區塊的

預估方程式為

$$\mathbf{g}_q = \mathbf{H}_q^T \mathbf{g}_q \quad (39)$$

$$\text{where } \mathbf{H}_q = \mathbf{\Phi}_q \left(\mathbf{\Phi}_q^T \mathbf{\Phi}_q + \lambda_\beta \mathbf{I}_\beta \right)^{-1} \mathbf{\Phi}_q^T. \quad (40)$$

其中 \mathbf{H}_q 為區塊估測組合矩陣。進一步我們依據(39)可以取得區塊的事前機率模型

$$\begin{aligned} -\log(p(\mathbf{g}_q)) &= \left\| \mathbf{g}_q - \mathbf{H}_q^T \mathbf{g}_q \right\|^2 \\ &= \mathbf{g}_q^T \mathbf{Q}_q \mathbf{g}_q. \end{aligned} \quad (41)$$

而全域事前機率可由各區塊合併而得

$$-\log(p(g)) = \mathbf{g}^T \mathbf{Q} \mathbf{g}. \quad (42)$$

在(42)中 \mathbf{Q} 為網格點所對應的 Laplacian 矩陣。網格點所對應的事後機率模型如下所示:

$$\begin{aligned} -\log(p(I^{set}|g)p(g)) \\ = (\tilde{\mathbf{g}} - \mathbf{g})^T \mathbf{\Lambda}_g (\tilde{\mathbf{g}} - \mathbf{g}) + \mathbf{g}^T \mathbf{Q} \mathbf{g}. \end{aligned} \quad (43)$$

此網格事後機率的極大值可由微分求得，其最佳解 \mathbf{g} 可由下式所得

$$(\mathbf{Q} + \mathbf{\Lambda}_g) \mathbf{g} = \mathbf{\Lambda}_g \tilde{\mathbf{g}}. \quad (44)$$

當我們估測出網格點所對影的深度值，我們便可以接著重建深度影像各像素點的深度值。其估測方式如下所示，我們建立網格點 j 與向量點 i 的條件機率模型，再利用此機率預測像素點的深度值，如下所示

$$d_i^* = \sum_{j \in N(i)} g_j^* \cdot p(f_j | f_i) \quad (45)$$

當中網格點 j 與向量點 i 的關係式如下

$$\begin{aligned} p(f_j | f_i) &= \frac{1}{F_i} \exp \left(-\frac{\|f_i - f_j\|^2}{\sigma_f} \right), \\ \text{where } F_i &= \sum_{j \in N(i)} \exp \left(-\frac{\|f_i - f_j\|^2}{\sigma_f} \right). \end{aligned} \quad (45)$$

藉由式(45)我們便可將所有影像像素點的深度值重建，而這樣的方式會遠比直接估計所有像

素點深度快速許多。

利用多重對焦所估測的深度影像結果如圖十二所示，我們比較我們的全域性處理效果與區域性深度重建演算法[9]所處理的結果，藉由我們的方法可以有效排除平坦區域可能面臨的深度誤判，同時在深度影像中有較佳之邊界銳利度表現。

圖十三是透過多重對焦重建的深度資訊，當中我們只利用三種不同對焦，藉由這樣的方式我們可以將前景物與背景物進行分離動作。未來影像編修便可以依據此深度資訊進行影像拆解與調整編修。

4. 結論

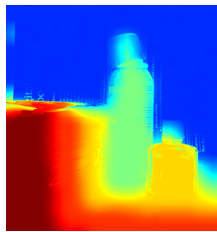
在本計畫中，我們提出了一套自動色調調整系統，能夠將影像拍攝時受到環境光線限制所導致的過亮或過暗部分，利用區域性的 gamma 調整至偏好色調，同時增強影像細節對比與色彩飽和度。為了自動估測對應於影像的適當 gamma，我們採用 ML 基礎的最佳化演算法，考量像素間的相似性估測較佳之 gamma，以避免影像產生調整上的瑕疵問題，獲得較為自然的調整結果。為了提升效能，我們也提出了一套高效能處理方式，透過像素點的合併，能夠大幅度降低最佳化運算所需之運算量。達到快速提升影像品質的效果。

另一方面，我們提出了兩套不同的深度影像重建系統，第一套系統針對固定攝影機拍攝一段時間的影片，利用畫面中移動物體與場景物件互相遮蔽關係分割前後景物邊界，建立相對深度關係圖。另一套系統是利用多重對焦影像，透過正確對焦位置與深度的關聯性重建深度影像，為了克服畫面中常見的平坦難以量測對焦的表面，我們提出一套事後機率最大化的最佳化深度重建演算法，能夠有效排除平坦表面的誤判情況。為了降低最佳化的運算複雜度，我們提出了基於網格的演算架構，將相似特徵值的影像像素點合併，大幅度提升運算效

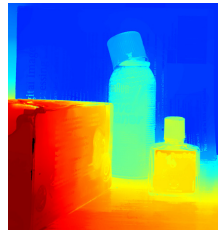
能。這些深度重建方法有助於未來進行影像編修時的物件分離與調整，影像編修可以透過影像深度有更多元的處理發展可能性。



(a) Image 1



(b) Result by [9]

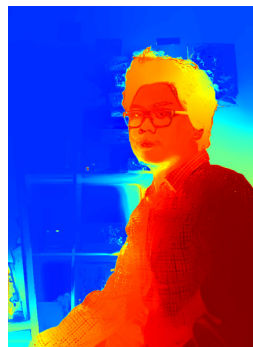


(c) Our result

圖十二、影像深度重建結果。



(a) Image



(b) Our result

圖十三、影像深度重建結果。

5. References

- [1] C. Tseng, S. Wang, and Y. Chen, "Image enhancement based on gamma map processing," *Proceedings of the SPIE*, vol. 7723, pp. 77230G-77230G-10, 2010.
- [2] A. Levin, D. Lischinski, Y. Weiss, "A Closed Form Solution to Natural Image Matting," *Proc. ICCV*, pp.61-68, 2006.
- [3] J. Tumblin and G. Turk, "LCIS: a boundary hierarchy for detail-preserving contrast reduction," *In Proc. ACM SIGGRAPH*, p.83-90, July 1999.
- [4] F. Durand and J. Dorsey, "Fast bilateral filtering for the display of high dynamic- range images," *In ACM Transactions on Graphics*, vol. 21, no. 3, pp. 257-266, 2002.
- [5] L. Tao and V. Asari, "Adaptive and integrated neighborhood-dependent approach for nonlinear enhancement of color images," *J. Electron. Imaging* 14, 043006, 2005.
- [6] D. Hoiem, A. A. Efros, and M. Hebert. "Recovering surface layout from an image," *IJCV*, vol. 75, no. 1, pp. 151-172, 2007.
- [7] S. K. Nayar and Y. Nakagawa, "Shape from focus," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 8, pp. 824-830, Aug. 1994.
- [8] N. Yokoya, T. Shakunaga, and M. Kanbara, "Passive range sensing techniques: Depth from images," *IEICE Trans. Syst. Inf. E82D*, vol. 3, pp. 523-533, 1999.
- [9] T. Aydin and Y. Akgul, "A new adaptive focus measure for shape from focus," in *BMVC*, 2008.
- [10] C. Tseng, and S. Wang, "Maximum-a-posteriori estimation for global spatial coherence recovery based on matting Laplacian," in *IEEE International Conference on Image Processing*, 2012.