

主成分分析在影像處理上的應用

last modified May 9, 2006

數位影像是屬於大型的資料，一張 1024×1024 的灰階照片大小等於 1Mbytes。當影像更大、色彩更豐富、數量更多時，不管在儲存、傳輸或是特徵抽取 (feature extraction) 上都造成困擾。基於影像資料中，鄰近像素的相關性頗高 (如圖 1)，表示這類的資料本身即存在著一些「多餘 (redundant) 資料」，利用主成分分析的原理可以有效的去除這些多餘的資料，讓資料量變小，在影像處理上則稱為「影像壓縮」。經過壓縮處理過的影像適合保存、傳輸等用途，但被刪除的資料畢竟仍是影像的一部份，在影像復原時會有某種程度上的損失，不過相較於影像大小的縮減，在某些應用上仍是值得的。

本單元將以實際的資料展現主成分分析在資料壓縮與上的功能，雖然有時候效果不是很好，卻可以當作其他方法的前置作業處理 (preprocessing)。



圖 1: 影像資料中鄰近像素間的高相關性: 右圖為左圖Lena 左眼部位的放大圖，可以清楚的看到每個像素及其灰度。

1 背景介紹

1.1 主成分分析

主成分分析的原理係將原變數向量 $\mathbf{x}$ (含 p 個變數), 正交投射 (Orthogonal projection) 到維度較低 ($q < p$) 的子空間 (Subspace) V , 成為 $\mathbf{x}_q$, 如圖 2 所示, 其數學的表示法寫成

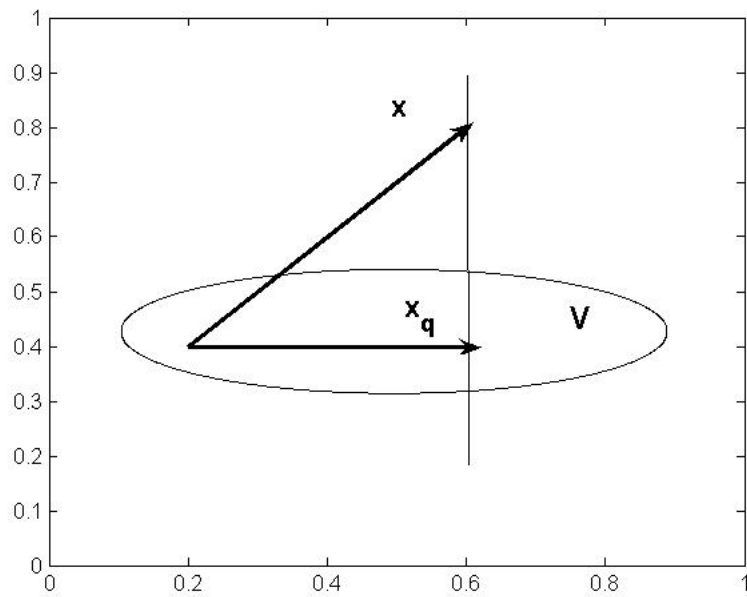


圖 2: 主成分分析的正交投射原理

$$\mathbf{x}_q = P\mathbf{x} \quad (1)$$

其中 P 即所謂的投射矩陣, 在此也稱為正交投射矩陣 (Orthogonal Projection Matrix)。主成分分析為滿足其「主成分」的目的, 定義了投射的目標區 (子空間 V) 及 P 的選擇。 V 的定義及 P 的選擇如下:

$$\begin{aligned} V &= \text{span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_q\} \\ P &= V_q V_q^T \end{aligned} \quad (2)$$

其中 $V_q = [\mathbf{v}_1 \ \mathbf{v}_2 \ \cdots \ \mathbf{v}_q]$, $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_q$ 為變數向量 $\mathbf{x}$ 的共變異矩陣 Σ_X 的前 q 個特徵向量 (經標準化後) 即

$$\Sigma_X \mathbf{v}_k = \lambda_k \mathbf{v}_k, \quad 1 \leq k \leq q, \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_q \quad (3)$$

其中 $p \times q$ 的矩陣 V_q 也稱為 Orthonormal matrix, 因其滿足 $V_q^T V_q = I_q$ 。 「主成分分析」將資料變數從 $\mathbf{x}$ 投射(轉置) 到 $\mathbf{x}_q$ 的主要目的有二: 其一, 從空間幾何的角度來看, 子空間的座標軸的選擇係依據資料成分 (能量或變異) 的分佈, 資料呈現在其間的分佈與原座標軸所呈現出的分佈不同, 在某些應用上, 如群組分析、變數的選擇, 可以得到好處。經轉換後的資料在子空間 V 的座標為

$$\mathbf{z}_q = V_q^T \mathbf{x} \quad (4)$$

$\mathbf{z}_q$ 為 $q \times 1$ 的向量, 當 $q = 1, 2$ 時, 常可從資料分佈圖中, 發現一些潛藏的資訊。

其二: 資料量變小。原變數 $\mathbf{x}$ 與新變數 $\mathbf{x}_q$ 雖同為 $p \times 1$ 的向量, 但 $\mathbf{x}_q$ 所在的空間較小, 在資料的儲存上通常以下列方式來表達

$$\mathbf{x}_q = \alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \cdots + \alpha_q \mathbf{v}_q = V_q \alpha \quad (5)$$

每個樣本都是基底向量的線性組合, 儲存上僅需為每一個樣本保留其組合係數, 及一組共用的基底。當樣本數愈大, 節省的空间也相對可觀。式 (5) 的 α 其實就是式 (4) 的 $\mathbf{z}_q$ 。

當然儲存空間變小並非全無代價, 根據主成分的原理, $\mathbf{x}_q$ 僅保留原變數一定比例的資訊 (能量), 其餘部分 (假設為誤差) 完全捨棄, 這項誤差表示為

$$\mathbf{e} = \mathbf{x} - \mathbf{x}_q = (I - P)\mathbf{x} \in V^\perp, \text{ where } V \oplus V^\perp = R^p$$

很明顯的, q 的選擇決定了誤差的大小, 而且並無一定的標準, 完全視應用的情況而定。譬如影像資料的壓縮若著眼於視覺感官的反應, 若干資料的損失往往是視覺上所能忍受的, 此時 q 的選擇常以目視來決定。

主成分分析在實際資料的計算上與著名的 SVD 矩陣分解有密切的關係，先就 SVD 說明如下，再來分析之間的關係。

1.2 SVD: Singular Value Decomposition

假設 A 為一個 $m \times n$ 的矩陣，其 SVD 表示法為

$$A = \sum_{k=1}^r \sigma_k \mathbf{u}_k \mathbf{v}_k^T \quad (6)$$

其中 $r = \text{Rank}(A) \leq \min(m, n)$, $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r$ 稱為 singular values, $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_r$ 為 $m \times 1$ 的 orthonormal vectors, 也稱為 left singular vectors, 而 right singular vectors 指的是 $n \times 1$ 的 orthonormal vectors $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_r$ 。這些 singular values 及 singular vectors 來自以下的 eigenvalue-eigenvector 分析:

$$\begin{aligned} AA^T \mathbf{u}_k &= \lambda_k \mathbf{u}_k = \sigma_k^2 \mathbf{u}_k \\ A^T A \mathbf{v}_k &= \sigma_k^2 \mathbf{v}_k \end{aligned} \quad (7)$$

式 (7) 說明 singular value σ_k^2 的平方是 AA^T 的特徵值，而 left singular vector $\mathbf{u}_k$ 為其相對的特徵向量。Right singular vectors $\mathbf{v}_k$ 為 $A^T A$ 的特徵向量，相對應的特徵值也是 σ_k^2 的平方。式 (6) 的 SVD 表示法可以矩陣的方式改寫為

$$A = U \Sigma V^T \quad (8)$$

其中

$$\begin{aligned}
U &= \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_m \end{bmatrix} \\
\Sigma &= \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_r & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{bmatrix} \\
V &= \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix}
\end{aligned}$$

U, V 爲 $m \times m$ 及 $n \times n$ 的 orthonormal matrix, 即 $U^T U = U U^T = I_m, V^T V = V V^T = I_n$ 。 Σ 爲一 $m \times n$ 的對角矩陣, 對角的位置自 $r + 1$ 之後皆爲 0。

1.3 主成分分析與 SVD

之前對於主成分分析的探討僅止於變數, 現考量其實際的樣本資料。假設原始變數表示爲 $p \times 1$ 的向量 $\mathbf{x}$, 現有 N 個樣本, 疊成一個 $N \times p$ 的資料矩陣 X , 即

$$X = \begin{bmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \vdots \\ \mathbf{x}_N^T \end{bmatrix} \quad (9)$$

對 $\mathbf{x}$ 的共變異矩陣 Σ_X 的估計, 寫成

$$C_x = \frac{1}{N-1} X^T X \quad (10)$$

假設資料矩陣 X 的 SVD 表示爲

$$X = U \Sigma V^T \quad (11)$$

根據式 (7), 上式 V 中的 right singular vectors v_k 即為式 (10) 中共變異矩陣的特徵向量, 當式 (11) 改寫為

$$XV = U\Sigma = Z = \begin{bmatrix} \mathbf{z}_1^T \\ \mathbf{z}_2^T \\ \vdots \\ \mathbf{z}_N^T \end{bmatrix} \quad (12)$$

上式的資料矩陣 Z 就是原始資料經過座標軸轉換過的新座標。當只取 $q (q < p)$ 個主成分時, 上式寫成

$$XV_q = U_q\Sigma_q = \begin{bmatrix} \sigma_1\mathbf{u}_1 & \sigma_2\mathbf{u}_2 & \cdots & \sigma_q\mathbf{u}_q \end{bmatrix} = Z_q \quad (13)$$

上式其實就是式 (4) 的樣本值表示法, 利用 SVD 的方式可以看得更清楚, 計算上也更方便。前面說過, 當 $q < p$ 時, 原始資料並無法完全復原, 只能得到近似值 X_q , 以資料矩陣 X 的 SVD 的表示法

$$X_q = U_q\Sigma_qV_q^T = \sum_{k=1}^q \sigma_k\mathbf{u}_k\mathbf{v}_k^T = Z_qV_q^T \quad (14)$$

稱為原始資料 X 的 Rank q approximation。或是針對每個樣本的近似值

$$\tilde{\mathbf{x}}_k = V_q\mathbf{z}_k = \mathbf{z}_k(1)\mathbf{v}_1 + \mathbf{z}_k(2)\mathbf{v}_2 + \cdots + \mathbf{z}_k(q)\mathbf{v}_q, \quad 1 \leq k \leq N \quad (15)$$

當 X 代表一張影像資料時, X_q 便是其壓縮後還原的影像, 其失真程度與 q 的選擇有關。以下的範例可以幫助理解。

2 練習

本單元著重在影像處理上的應用, 如圖3秀出50張手寫的「3」(資料來源[?], 檔案3.mat), 每張大小為 16×16 灰階影像, 總共有658張。應用主成分分析的技巧, 可以以較低的容量來存放這些圖片, 而且保留相當程度的外觀。本練習將實際計算出 q 的選擇與容量大小的縮減及最後外觀上的差別。

範例1: 每張圖都是一個樣本, 是 16×16 的矩陣, 那麼前面所說的 $p \times 1$ 的變數 $\mathbf{x}$ 中的 p 該是如何界定? 應用前述主成分分析的推論, 一般的作法是將影像的矩陣資料展開成向量, 即將 16×16 的矩陣拆成 256×1 或 1×256 的向量。以資料檔案3.mat 為例, 其中的每一列 (1×256) 代表一個樣本, 也就是一張 16×16 的圖。在本範例中, 試著先寫一支小程序秀出其中的一張圖, 熟悉在 MATLAB 中秀出一張圖的指令, 成功之後, 再將程式擴大為可以秀出多張樣本圖的「大圖」, 以便能多觀察樣本影像的異同。

要將 1×256 的向量顯示成圖形, 首先要先恢復成矩陣, 譬如使用 reshape 的指令, 重新調整資料的排列。在此仍建議 MATLAB 初學者, 在使用每一個指令之前, 最好都要詳讀該指令的使用說明, 如果可以找到幾個範例, 也一定要試著跑看看, 才能準確的掌握正確的使用方式, 省去一些嘗試錯誤的時間。萬一找不到範例, 也應該試著以一些簡單的數據去摸索。將影像重排成矩陣後 (假設為 $\mathbf{Y}$), 以下的指令可以秀出第 i 張「3」的影像圖:

```
load 3
i=1;
Y=reshape(X(i,:),16,16)'; %本範例的資料需經轉置後才是正的
imagesc(Y);
colormap(gray) %以灰階表現影像色彩
```

指令 imagesc 將影像資料 $\mathbf{Y}$ 自動調整為現行「調色盤 (colormap)」的色彩範圍。通常為配合圖形本身的色彩 (調), 在 imagesc 後面, 還會特別指定該影像適合的調色盤, 也就是指令 colormap 的功能。

範例2: 研究者在分析資料前, 總希望能對手上的資料有更清楚的瞭解, 就本練習的資料3.mat 來說, 共有 658 張圖, 最好都能一一看過, 知道資料的變異情況, 方有助於資料分析時所採取的策略。畫出如圖 3 的 50 個影像樣本, 排列成 5×10 的大張影像圖。



圖 3: 50個數字影像樣本

圖3展示的50張圖，每張都有數字3的樣子（至少目視可以辨認），當然也有其變異處。主成分分析的意涵有保留最大的共同處，去除其他變異較大的部分。為方便觀察資料的變異情況，最好能把圖並列顯示，在程式的作法上，其實是將矩陣擴大，將原來一張張影像的小矩陣疊成適當的大矩陣，當作一張比較大的圖來顯示，這就是圖3的樣子了。

將小矩陣疊成大矩陣的程式寫法有很多種，這個技巧非常普遍，經常會用到。以 MATLAB 而言，寫法更是五花八門，有的比較好寫（比較直覺），但往往速度不理想，講究速度快的，寫起來常要費些手脚。下面的程式提示可以很快發展成秀出圖3的程式，

```
範例一： 假設 Y1,Y2 為 p x p 的矩陣
Y=[Y1 Y2]           % 疊成 p x 2p 的矩陣
Y=horzcat(Y1,Y2)     % 同上的功能
Y=[Y1;Y2]           % 疊成 2p x p 的矩陣
Y=vertcat(Y1,Y2)     % 同上的功能
```


範例二

```
Y=[]; %先設定 Y 為空矩陣
for i=1:5
    Y=[Y rand(3,3)];
end
%最後 Y 疊成為 3x15的亂數矩陣
%練習：用這個方式將10個 3x3的矩陣疊成 6x15的大矩陣（即兩列五行）。
```

範例三：

```
Y=zeros(6,15); %先將 Y 最後的大小設定好, 其值先預設為0
for i=1:2
    for j=1:5
        Y((i-1)*3+1:3*i, (j-1)*3+1:3*j)=rand(3,3);
    end
end
end
```

以上三個範例是常用的矩陣堆疊的技巧。使用上視矩陣的大小及數量而定。圖 3 秀出 658張圖的其中 50張, 程式可以這樣寫

```
load 3
Y=[];YY=[];
r=10;c=5; N=r*c; %to create a big image with cxr images.
Y_pad=-1*ones(8,16); % padding spaces between image patches
for i=1:N %自上而下, 由左而右排列 N 張 16x16的影像
    Y=[Y;reshape(X(i,:),16,16)';Y_pad];
    if rem(i,c)==0, YY=[YY Y]; Y=[];end %判斷是否換行
end
imagesc(YY);
colormap(gray)
```

上面的程式值得的是第四行，定義了一個矩陣 Y_{pad} ，其值均為-1，目的在使每張影像堆疊時中間也多餘的空間，看起來比較不那麼擁擠。讀者當然可以試著將第六行改為

```
Y=[Y;reshape(X(i,:),16,16)'];
```

看看堆疊出來的大圖長相如何。第七行對程式設計不熟悉者而言也是有點困難，它是利用餘數為零來做換行的動作；這是當每行堆滿 c 張圖後，進行換行的動作。

範例3：圖形資料檔3.mat 以式 (9) 的方式儲存658張影像在變數 X ，共658列256行。以SVD的方式(式(14))取前 q 個主成分，再進行影像還原，合成如圖4展示的五張樣本影像，其中原影像在最左行，其次為合成 $q = 1, 2, \dots, 10$ 個主成分的還原影像。



圖 4: 取 q 個主成分後的影像還原 (壓縮) 效果: 第2 10 欄為 $q = 1 \sim 10$

這張圖將不同的 q 值所呈現的影像並列，方便研究者觀察「對於不同的影像，不同的 q 值在回復影像時的效果。」當加進來的成分愈多時（往右），影像愈清楚。不過仍可以看出來第一個主成分已經展現原影像的大致樣貌，後面的成分做了更多的修飾，讓影像更接近原圖。其實主成分分析將658張不同的「3」影像，分析出共同的 256個成分，每張影像是這 256個成分的不同的線性組合，圖4幫助我們瞭解不同成分扮演的角色，你可以看出從加進第幾個成分後，這五張圖才有明顯的不同呢？這或許可以幫助決定影像壓縮時 q 該取多少？

範例4: 式(15) 說明每個壓縮過的樣本是獨立向量（主成分） $v_1, v_2, \dots, v_q$ 的線性組合。這是否意味著，這些獨立成分也有「3」的味道？畫出來看看便知有沒有，如圖5。這樣的試驗，其實也是研究工作中經常會做的，有時只是證明自己的想法很天真，有時卻有令人意外的驚奇。不管結果如何，在觀念上都有深植的效果。

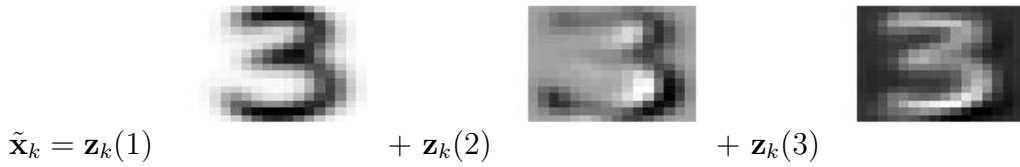


圖 5: 前3個主成分的「長相」

圖5秀出前三個主成分代表的圖形，確實看出「3」的模樣，雖然愈往後的向量其影像愈模糊，不過長相卻略有不同。完整的成分應該有 256個（可以試著秀出如圖6的 100個），也可以解釋為數字影像「3」的特徵，當然主成分分析的原理也說明，後面的特徵比較「不重要」。而組合的係數就是該樣本對於每個特徵的權重，藉以修飾初步同樣貌的影像。

範例5: 接續上面的練習，上圖中的組合係數 $\mathbf{z}_k$ 與特徵值的分佈有關（見 (12)），這些係數是否也呈現由大變小的分佈？是否呈現出特殊的分配模式？值得畫出來看看。如圖7列出其中的三組係數，左圖依序畫出其大小，右圖則是畫出相對的直方圖。這樣的分配比較稀疏 (sparse)，接近雙指數分配 (double exponential)。而大小的順序像 damped exponential function。

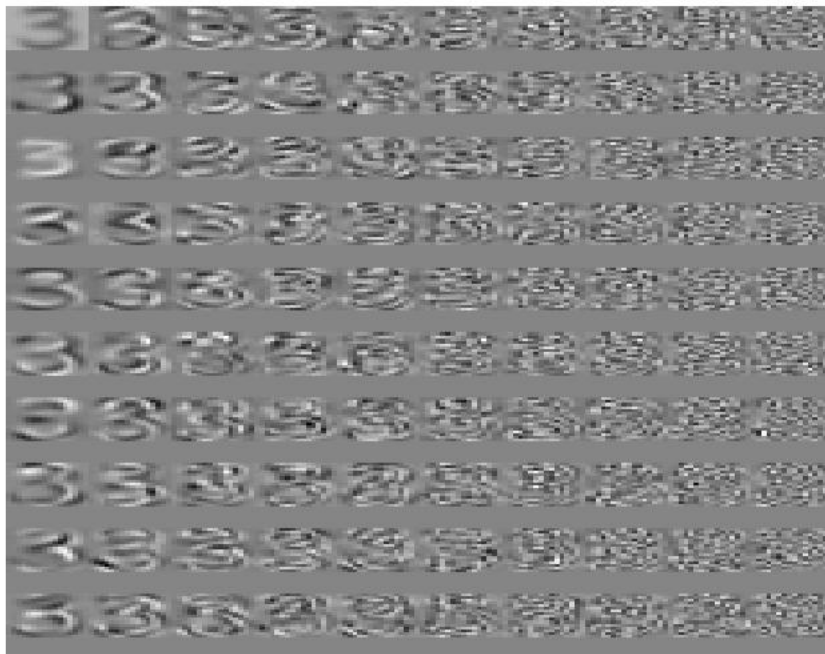


圖 6: 前100個主成分的「長相」。

範例6: 前面的練習中，每張影像代表一個樣本，主成分分析從總共658個樣本中分析出組合成一張 16x16 的影像「3」的成分（特徵）。將這個觀念更大膽的延伸到對任何一張 $m \times n$ 的影像抽取特徵，或許也可以作為單一——張影像的壓縮。如同 JPEG 或 TIFF 圖形檔一般，是從原圖壓縮來的。圖8。利用著名的 Lena 圖展示這個想法的結果。

當然主成分分析並非針對單一樣本，對於一張影像而言，可以將影像切出許多的 image patches，即小影像，將這些影像當作樣本，並且假設這些樣本有著共同的成分。將一張圖切成許多小圖，並假設這些圖有著相關性，這個想法或許理論根據稍薄弱，但實際上的效果如何，還是得做做看才知道。圖8為著名的 Lena 圖，大小 256×256 。剪成 16×16 的 patch image，共有256個樣本，經主成分分析後，取 q 值分別為16

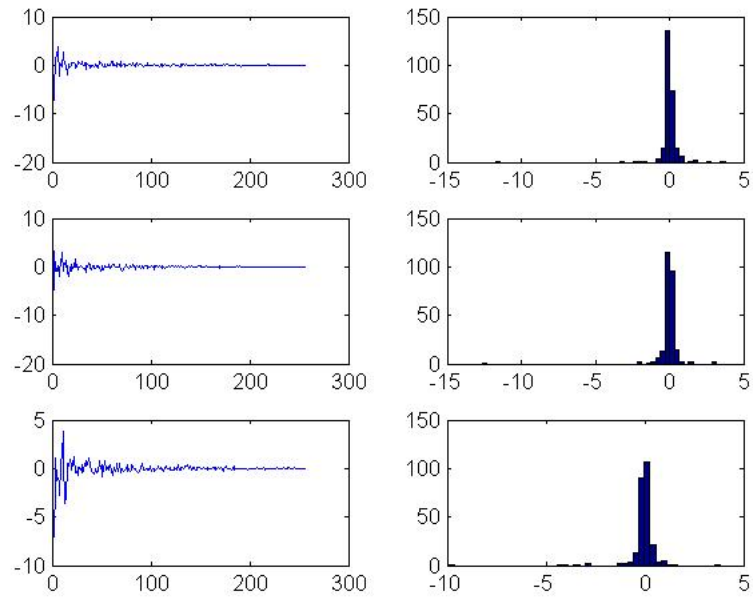


圖 7: 組合係數 $\mathbf{z}_k$ 的大小與分配

及 64 並回復影像，其結果如第二及第三張圖所示，資料壓縮的比例分別為 8 倍及 2 倍。



圖 8: 左邊為 256x256 的 Lena 原圖，中間為壓縮圖採 $q = 16$ (壓縮比 8 倍)，右邊採 $q = 64$ (壓縮比 2 倍)

將一張圖切割成連續的小圖 (patch)，當成樣本並且整理成像上面練習用的檔案 3.mat 的矩陣型態，需要寫一小段程式。譬如：

```

%以下變數的意義:p 為所選擇 patch image 的長,mxn 代表原影像的大小
X=[];
for j=1:m/p
    for i=1:n/p
        %由左而右, 由上而下擷取影像片段
        x=reshape(Y((j-1)*p+1:j*p,(i-1)*p+1:i*p)',1,p*p);
        X=[X;x]; %組合成資料矩陣
    end
end
end

```

Patch image 大小的選擇並非隨意，取不同大小會發現效果略有不同，一般而言取 $p = 12$ 是比較有根據的。本練習取 $p = 16$ 純粹是為配合原圖的尺寸，方便程式的撰寫，不必處理圖形長寬不能為 p 整除的麻煩。將最後組合成的矩陣送入之前練習的程式，便可以直接進行主成分分析。當然可以儲存起來，方便以後的程式直接使用，無須轉換。儲存的指令如下

```
save lena X
```

將矩陣 X 存放在檔案 lena.mat。將來止只需要 load lena 便可以直接取用 X 。

3 觀察

1. 式 (4) 同時代表座標軸的轉換與空間的縮減。如果不考慮空間的縮減，即子空間 V 等於原空間，或是表示為 $V = \text{span}\{\mathbf{v}_1 \mathbf{v}_2 \cdots \mathbf{v}_p\}$ ，則式(4) 寫成

$$\mathbf{z} = V^T \mathbf{x}$$

$$\mathbf{x} = V \mathbf{z} = [V_q \ V_{p-q}] \begin{bmatrix} \mathbf{z}_q \\ \mathbf{z}_{p-q} \end{bmatrix} = V_q \mathbf{z}_q + V_{p-q} \mathbf{z}_{p-q} = \mathbf{x}_q + \mathbf{e}$$

更明白的表示出 q 的選擇與誤差間的關係。

2. 從圖3觀察主成分代表的意義。五張「3」的影像除了看起來都像數字「3」的共同點外，各有其不同的特色，主成分分析中的每個成分是否都代表著組成「3」的

成分？當 $q = 1, 2, \dots, 10$ 將主成分一個個加進來時，回復出來的影像似乎也逐漸從相似 ($q = 1$: 第二行的每一張圖) 到突顯出特色 ($q = 10$: 最後一行的每一張圖)。

3. 式 (5) 或式 (15) 也提供一個觀點：每個樣本都來自同一組向量的線性組合，如範例4所示。如果我們任意的組合這些基底向量，是不是可以製造出新（人造）的樣本？換句話說，以本單元所使用的數字3的影像資料，從樣本分析出的基底向量，可否組成一個新的手寫「3」的人造影像？範例5中關於組合係數的大小與分佈，提供了些線索。以下這張圖是人工組合出來的，不與原樣本的任一張圖相同。



圖 9: 人造「3」

4 作業

1. 證明式 (1) 中的 P 是一個正交投射矩陣，其定義如下

定理 1. P is an orthogonal projection if and only if

$$P^2 = P \quad \text{and} \quad P = P^T$$

2. 根據式 (4)，證明變數 $\mathbf{z}_q(i)$ 與 $\mathbf{z}_q(j)$ 不相關，即 $E(\mathbf{z}_q(i)\mathbf{z}_q(j)) = 0$ 。
3. 假設樣本數為 N ，根據式 (4) 計算主成分分析在儲存空間上節省的比例。
4. 證明式 (4) 中的組合係數等於新的座標值，即 $\alpha = \mathbf{z}_q$ 。
5. 式 (7) 隱含一個事實：矩陣 AA^T 的特徵值為實數並且大於等於0，試證明之。
6. 假設 $\mathbf{x}$ 為一 $p \times 1$ 的向量，欲將 $\mathbf{x}$ 投射到另一個維度為 $q (q < p)$ 的子空間 V ，成為 $\mathbf{x}_q$ ，如圖2所示。假設 $w_1, w_2, \dots, w_q$ 為構成子空間 V 的 orthonormal

basis, 即 $V = \text{span}\{w_1, w_2, \dots, w_q\}$, 證明其正交投射矩陣為 WW^T , 即 $\mathbf{x}_q = WW^T \mathbf{x}$ 。其中 $W = [w_1 \ w_2 \ \dots \ w_q]$

7. 同範例3, 但多觀察幾個不同的 q 值, 直到壓縮的影像與原圖幾乎一樣。計算在不同的 q 值, 壓縮的影像所需的儲存空間, 及與原影像比較時, 壓縮的比例為何?
8. 同上, 試試不同的資料檔。
9. 如「觀察3」所言, 試著從式 (15) 模擬幾個人造的影像「3」出來。
10. 畫出 Lena 圖的前100個主成分影像圖。

參考文獻

- [1] J. Latin, D. Carroll, P. E. Green, "Analyzing Multivariate Data," 2003, Duxbury.
- [2] A. C. Rencher, "Multivariate Statistical Inference and Applications," 1998, John Wiley and Sons.
- [3] T. Hastie, R. Tibshirani, J. Friedman, "The Elements of Statistical Learning: Data Mining, Inference, and Prediction".