

第四章 知覺分析器 (perceptual analyzer)

由感覺分析器所得到的資訊，將會傳送至知覺分析器。而知覺分析器主要有兩個模組，一為 STA (spatial-temporal attention) 類神經網路模組，另一個為類別特徵擷取模組。以下先介紹如何利用 STA 類神經網路來模擬注意焦點集中的過程，並產生注意力圖像。之後再討論如何從注意力圖像中擷取出類別特徵的方式。

4.1 STA (spatial-temporal attention)類神經網路模組

STA 類神經網路分為兩個階層 (layers)，一為輸入層 (input layer)，另一為輸出層 (output layer)，亦稱為 attention layer。圖 4-1 為 STA 類神經架構圖。輸出層中的每個神經元都與其鄰近的神經元有連接，此種連結稱之為階層內部連結 (within-layer connections)，主要用來傳遞抑制訊息。輸入層內的神經元彼此之間並無連結存在，然而輸入層的每個神經元與輸出層的每個神經元皆有連結存在，且為完全連結 (full-connected)，此種連結為階層間連結 (between-layer connections)，主要用來傳遞激發訊息。

輸入層與輸出層的神經元皆是以 2D 陣列形式組織在一起。而陣列的大小即為輸入影像的大小。令 w_{ij} 代表輸出層神經元 n_i 與輸入層神經元 n_j 之間連結權重 (linking weight)。對輸出層神經元 n_i 而言，其權重向量為 $\mathbf{w}_i = (w_{i1}, w_{i2}, \dots, w_{im})$ ，其中 m 為輸入層的神經元個數。當輸入一刺激源 \mathbf{x} ，對輸出層神經元 n_i 而言，其所收到來自輸入層神經元之訊息可寫成

$$I_i^v = \mathbf{w}_i \cdot \mathbf{x} = \sum_{j=1}^m w_{ij} x_j \quad (4-1)$$

輸入層與輸出層間的連結權重之給定方式，如圖 4-2 所示。 n_j 為輸入層中之神經元，而 n_i 為其所對應之輸出層神經元。假設 2D Gaussian 函數 G 的中心落在輸出層 n_i 上，則輸入層 n_j 與輸出層 n_k 之間的連結權重 w_{kj} 為 $w_{kj} = G(\mathbf{r}_{ki})$ ，其中 \mathbf{r}_{ki} 為神經元 n_k 相對於神經元 n_i 的位置向量 (position vector)。

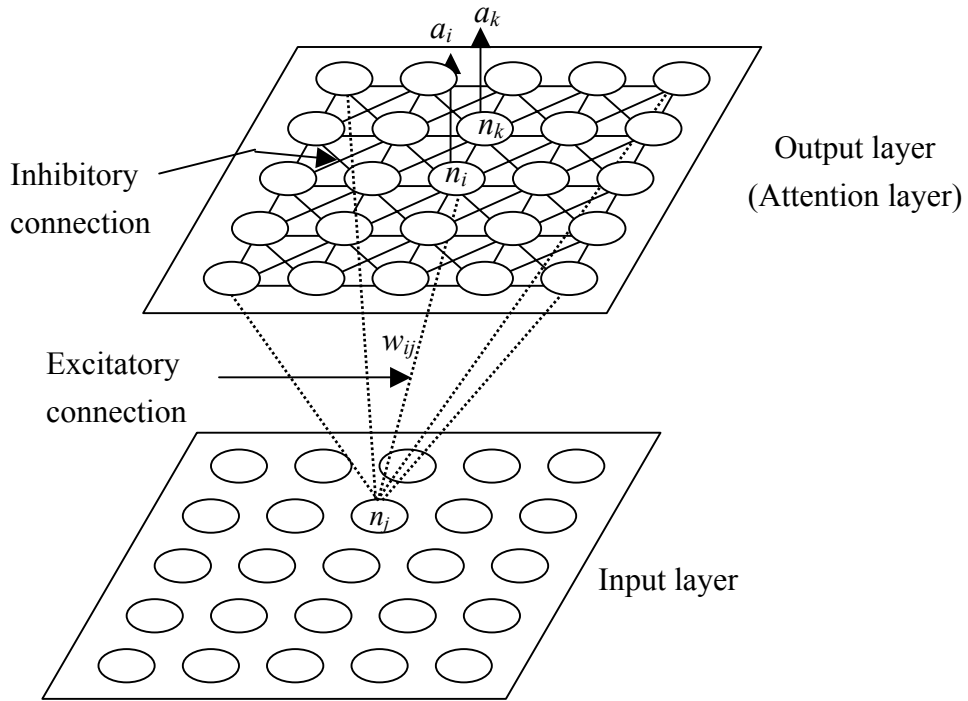


圖 4-1、STA 類神經網路架構圖。

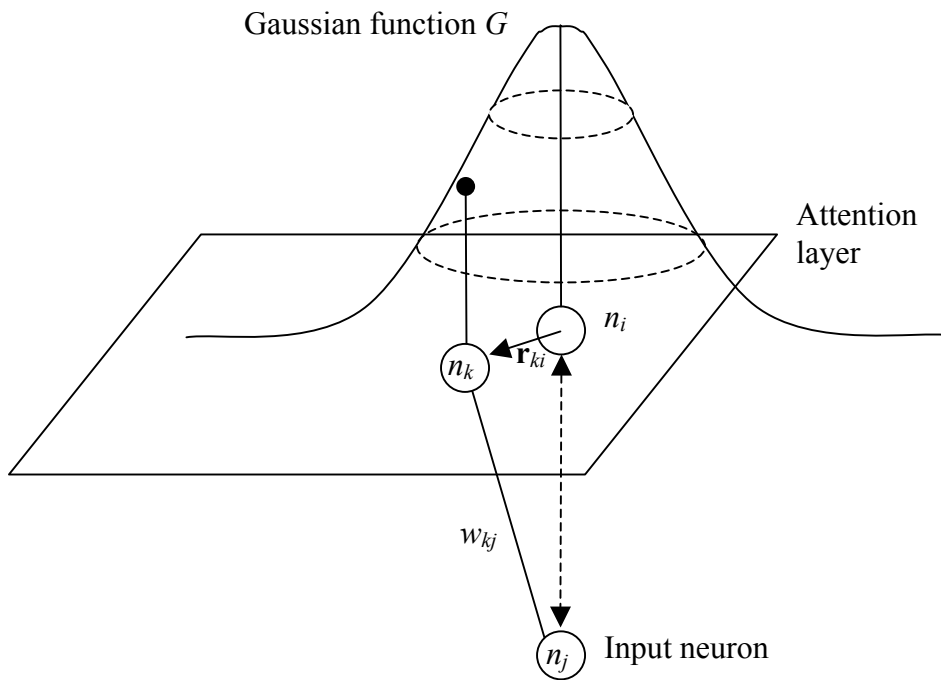


圖 4-2、輸入層與輸出層的連結權重。

在競爭學習時，輸出層的神經元會互相抑制，而這種抑制的機制可以用“墨西哥帽 (Mexican-hat)”函數來模擬，以 $M(\mathbf{r})$ 表示，其中 \mathbf{r} 為距離函數中心的位置。

置向量，如圖 4-3 所示。也就是說，當一輸出層神經元被激發時，在此神經元鄰近小範圍的神經元也都會被激發，但是超過某一段距離的神經元則會受到抑制，之後再隔一段距離的神經元又會受到激發的影響。通常“墨西哥帽(Mexican-hat)”函數會利用 Laplacian of Gaussian (LOG) $\nabla^2 G(\mathbf{r})$ 或 difference of Gaussian (DOG) $G_1(\mathbf{r}) - G_2(\mathbf{r})$ 等函數來近似。則輸出層神經元 n_i 所收到來自同層神經元之訊息為

$$I_i^l = \sum_{k \in N_i, k \neq i} [u_{ik} M(\mathbf{r}_k - \mathbf{r}_i) a_k] \quad (4-2)$$

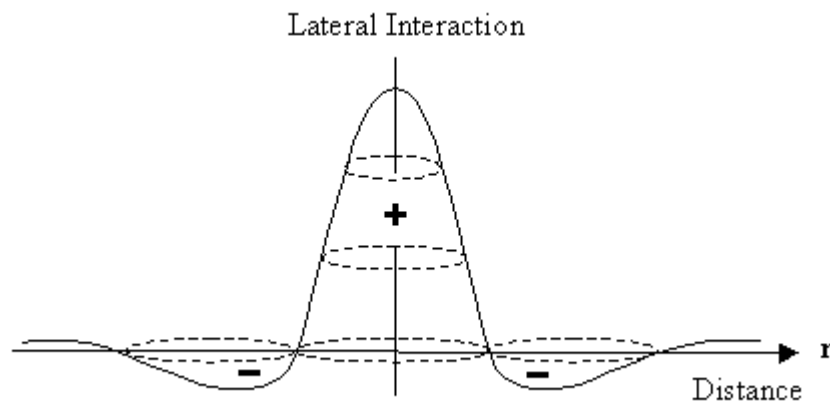


圖 4-3、墨西哥帽 (Mexican-hat) 函數。

其中 N_i 代表輸出層神經元 n_i 之鄰近神經元集合； u_{ik} 為神經元 n_i 與神經元 n_k 之間的連結強度，其位置向量分別為 \mathbf{r}_i 與 \mathbf{r}_k ； $a_k = \psi(\text{net}_k)$ 代表輸出層神經元 n_k 的激發值 (activation)，其中 $\psi(\cdot)$ 為轉換函數 (transfer function)，通常用 sigmoid function 來模擬，以下兩種方程式均可使用。

Fermi function:
$$\psi(x) = \frac{1}{1 + \exp(-\lambda x)} \text{ , 以及}$$

Glauber function:
$$\psi(x) = \frac{1}{2} [1 + \tanh(-\lambda x)] \text{ ,}$$

其中參數 λ 控制函數 $\psi(x)$ 的陡峭程度。

輸出層神經元 n_i 的淨輸入值之定義如下：

$$\text{net}_i = a_i + A(-ba_i + cB(I_i^v + I_i^l - \Gamma)) \text{ ,} \quad (4-3)$$

其中 b 與 c 為正值的常數； b 代表激發值 a_i 的衰減速率，而 c 代表從輸入層神經

元及神經元 n_i 的鄰近神經元所得到的總輸入之權重。門檻值 (threshold) Γ 可以過濾掉雜訊的干擾。函數 $B(x)$ 與 $A(x)$ 定義如下[Mar91]：

$$B(x) = \begin{cases} x & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases}, \quad (4-4)$$

$$A(x) = \begin{cases} x & \text{if } x > 0 \\ dx & \text{if } x \leq 0 \end{cases}, \quad (4-5)$$

其中 $1 > d > 0$ 。圖 4-4 說明輸入刺激源後，輸出層神經元激發值的活動情形。如果淨輸入值在 Δt 時間內超過門檻值 Γ ，則需要花 $1/b$ 時間來達到最高激發值，且在沒有其他刺激輸入的情況下衰減約需 $1/bd$ 時間。因為 $d < 1$ ，所以衰減的時間會比上升的時間要來得久。

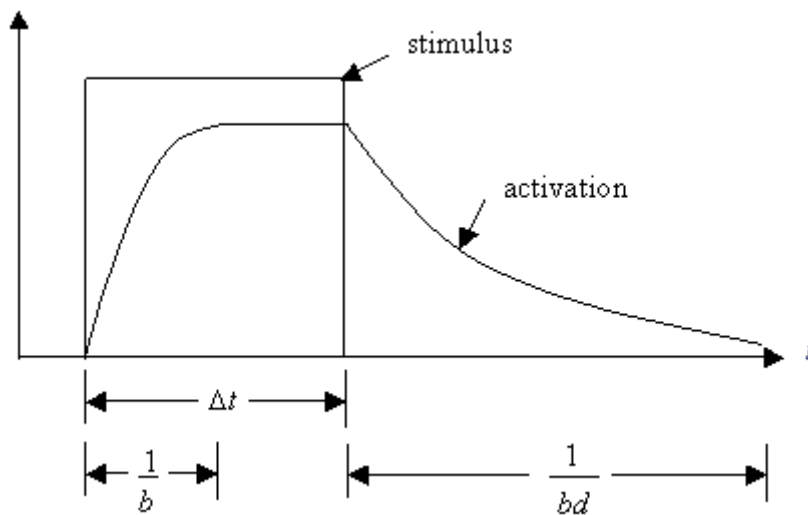


圖 4-4、輸入刺激源後，輸出層神經元激發值的活動情形。

假設某一點刺激了輸入層的神經元，接下來它便會影響到輸出層的神經元。每個輸出層的神經元受影響的程度並不相同，是由 Gaussian 函數來控制 (如圖 4-2)。Gaussian 函數可以使得某個輸入神經元的刺激在輸入後會刺激到輸出層對應神經元附近的神經元，產生刺激分散 (spreading) 的現象。因為輸出層神經元之間的相互抑制情形是由“墨西哥帽 (Mexican-hat)”函數所控制，所以輸出層的神經元會產生激發狀態聚集的現象。在 STA 類神經網路中的分散 (spreading) 與聚集 (clustering) 機制，分別如型態學 (morphological) 中的擴張 (dilation)

與侵蝕 (erosion) 運算。這兩種機制的整合，增加了系統容忍輸入資料的不完整之能力。

參考圖 4-1，假定一個移動點的在時間 t 時刺激輸出層神經元 n_i ，在時間 $t+1$ 時刺激它鄰近的神經元 n_k 。假設輸出層神經元 n_i 在時間 t 時的激發值為 $a_i(t) = 0$ 。從式(4-1)至式(4-3)，神經元的輸入值中，來自於點刺激與同一層間神經元的影響分別為 $I_i^v(t) (> \Gamma)$ 與 $I_i^l(t) (= 0)$ ，輸入層神經元 n_i 的淨輸入值為 $net_i(t) = c(I_i^v(t) - \Gamma)$ ，激發值則為 $a_i(t) = \psi(net_i(t))$ 。此時，輸出層神經元 n_k 的激發值 $a_k(t)$ 仍然為零。若在時間 $t+1$ 時刺激點移動到輸出層神經元 n_k ，這時輸出層神經元 n_k 開始上升，而輸出層神經元 n_i 的激發值則開始衰減。因為神經元的衰減速率比上升速率來得小，所以一旦刺激持續移動，便會在 attention layer 中留下移動的軌跡。這軌跡保留了刺激點在時間和空間方面的訊息。

總而言之，如果環境中無任何改變的話，衍生影像中只包含一些小的 entries，這並不會激發在 STA 類神經網路中的輸出層神經元 (參考式(4-3)中的門檻值 Γ)。但是，如果在衍生影像中包含許多的 entries，則會在輸出層神經元上產生的激發樣式。激發樣式會與環境中的改變同步發展，然後達到最大值，在類神經網路的輸出層上形成注意焦點。

圖 4-5 為前方車輛欲變換車道至右方車道之部分路況 (圖 4-5 之第一行)，及其所對應之衍生影像 (圖 4-5 之第二行) 與注意力圖像 (圖 4-5 之第三行)。衍生影像為 STA 類神經網路中 input layer 之輸入；而注意力圖像為 STA 類神經網路中 attention layer 之輸出。在注意力圖像中，可以發現當前方車輛在變換車道時，漸漸形成某種樣式 (pattern)，當它足夠明顯時，就形成注意焦點。注意力圖像也提供了鄰近車輛的位置與大小的訊息。



圖 4-5、知覺分析器中 STA 類神經網路模組之輸出範例。

4.2 類別特徵擷取模組

針對具有危險動作的鄰近車輛或本車輛所可能產生的注意力圖像，本研究歸類為十三種類別，分別對應十三種不同的危險動向，如圖 4-6 所示。其中圖 4-6 (a) 至圖 4-6 (i) 等九種注意力圖像為鄰近車輛所造成的危險動向；圖 4-6 (j) 至圖 4-6 (m) 等四種注意力圖像則為本車輛變換車道所造成的危險動向。並非所有的鄰近車輛都具有威脅性，通常那些與我車距離越來越近者，所具的威脅性最高。而自己所駕駛之車輛可能因為駕駛者未注意不在視野內之車輛而任意變換車道，也是有可能造成他人行車安全的威脅。

由於鄰近車輛在運動時，不一定是等速或等加速，因此在類別特徵的擷取時最好能夠擷取出這些注意力圖像定性的特徵，作為概念分析時分類的參考。從圖 4-6 中可以發現亮點漸層的分佈是偵測鄰近車輛動向的主要特徵。亮點分布的位置代表鄰近車輛的位置；亮點分布的漸層方向代表鄰近車輛移動的方向。



(a) 前車減速接近本車

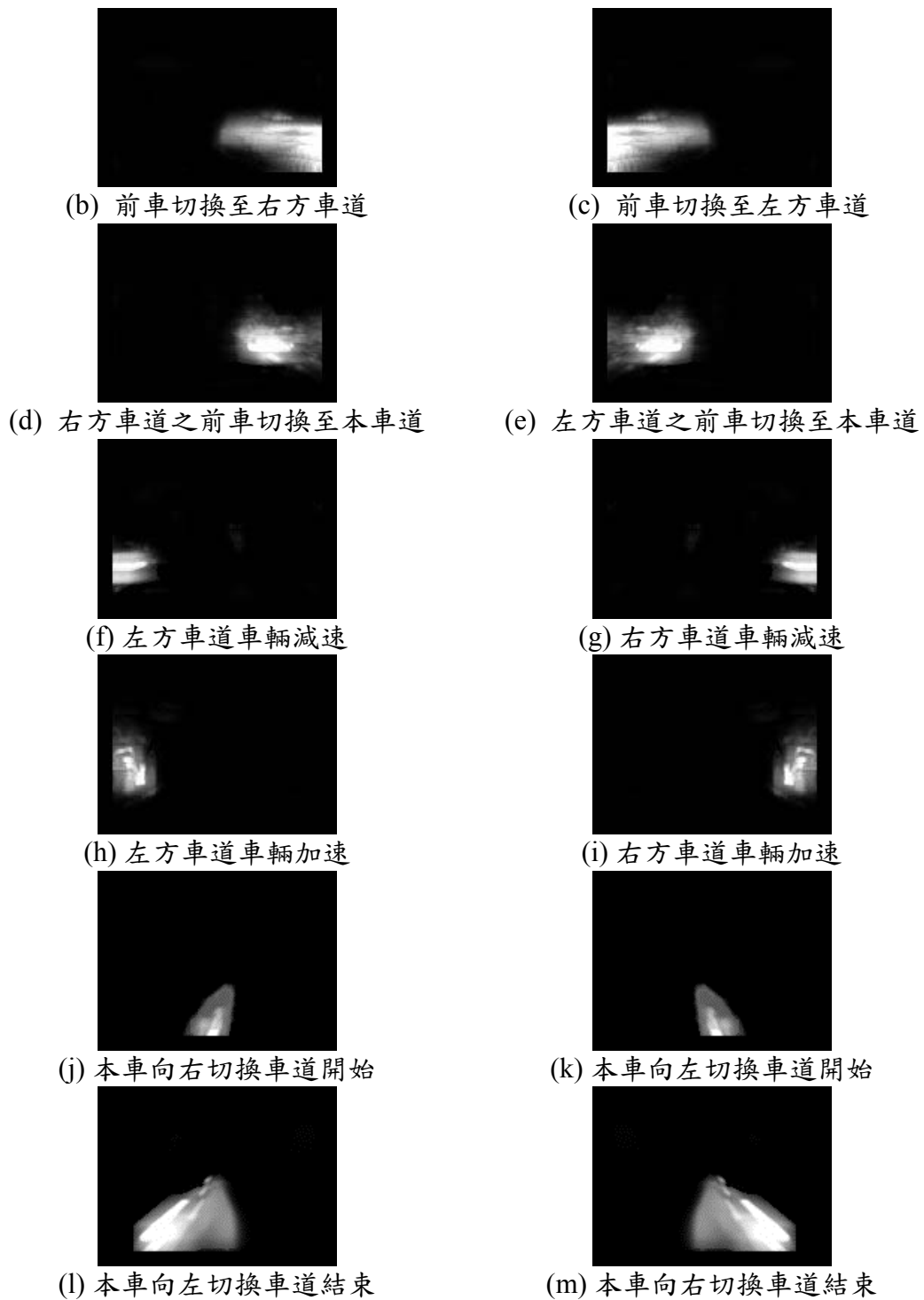


圖 4-6、單一鄰近車輛與本身車輛動向之注意力圖像。

在實際的狀況下可能同時有多台鄰近車輛在我車附近移動，因此容易導致多種危險動向同時發生。例如當前車切換至左方車道，並接近本車的同時，右方車道又跨越車道至我車道前方且接近本車。因此在做定性特徵之擷取時需考慮這些可能的情況。

4.2.1 分割視窗

有些危險動向只會發生在某些位置，因此我們希望利用不同位置的視窗來辨識不同的危險動向。如此一來，當有多種危險動向同時發生時，便可以利用不同視窗來辨識多個鄰近車輛動向類別。首先將所得到的注意力圖像分割成五個視窗，分別以 w_1 、 w_2 、 w_3 、 w_4 及 w_5 代表第一個、第二個、第三個、第四個及第五個視窗所切割出的影像。之所以切割成五個視窗，是根據本系統應用的需求。視窗一與視窗三涵蓋的範圍主要是左右車道的盡頭，與駕駛者的距離較遠，所造成的威脅也就比較少。視窗二與視窗四分別偵測左方車道與右方車道的車輛。這兩個車道的車輛可能忽然加速超越我車，或者即將切入我方車道前方，這些動作都可能對駕駛者造成威脅。視窗五主要在偵測車道前方的車輛。如果車道前方的車輛突然減速而造成與駕駛者的距離拉近，則其所造成的威脅最大，駕駛者須特別注意發生在此視窗中的危險動向。視窗的大小也是我們必須要考慮的條件之一，如果視窗面積越大，則越有可能涵蓋住整個注意焦點。其中 w_1 、 w_2 、 w_3 及 w_4 的大小為 100×80 像素 (pixel)，而 w_5 的大小為 80×80 像素 (pixel)。另外，之所以選擇重疊的視窗是因為希望不只一個視窗能夠涵蓋到注意焦點。這樣的話，我們便能夠整合各個視窗的分類結果以達到更好的辨識率。 w_1 、 w_2 、 w_3 、 w_4 及 w_5 之位置分佈圖，如圖 4-7。圖 4-8 即為「本車向右切換車道結束」注意力圖像之視窗分割情形。

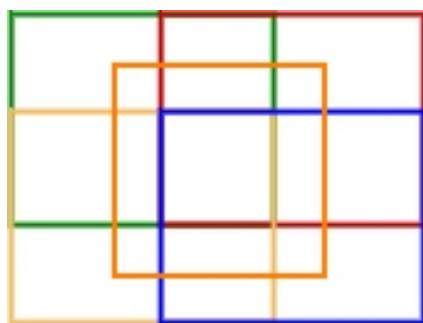


圖 4-7、所切割之五個視窗位置分佈圖：以綠色外框代表 w_1 ，以黃色外框代表 w_2 ，以紅色外框代表 w_3 ，以藍色外框代表 w_4 ，以橘色外框代表 w_5 。

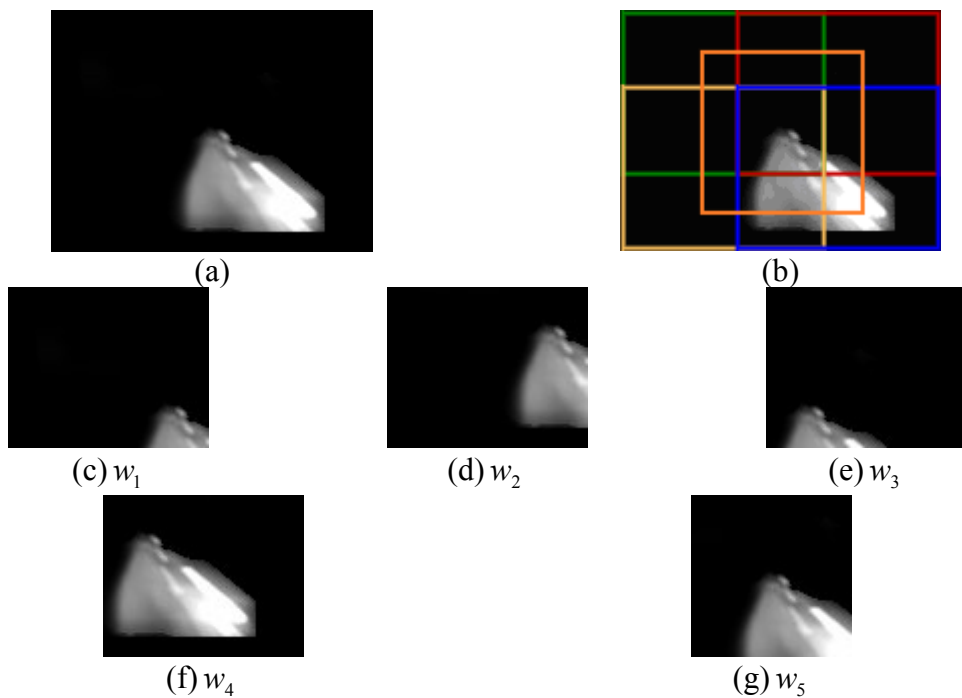


圖 4-8、(a)「本車向右切換車道結束」之注意力圖像，(b) (a)之視窗分割情形，(c)、(d)、(e)、(f)、(g)分別代表第一個至第五個切割的視窗。

4.2.2 特徵向量擷取

由於亮點分布的位置代表鄰近車輛的位置；亮點分布的漸層方向代表鄰近車輛移動的方向。因此我們認為利用偏態 (skewness) 的觀念可以擷取出符合上述條件要求的定性特徵。首先針對這五個視窗所分割出的部分注意力圖像 ($M \times N$) 做橫向切割成十個相等的區塊 ($M \times (N/10)$)，分割的方式如圖 4-9(b)。再將這十等分的區塊分別做縱向灰階平均，也就是每一行之灰階 (intensity) 平均值。經過這個步驟，任一區塊 i 皆可以由左至右求出 M 個灰階平均值，這些灰階平均值可以用向量來表示， $\mathbf{x}_i = \{x_{i1}, x_{i2}, \dots, x_{iM}\}$ ， $i = 1, 2, \dots, 10$ 。利用偏態的公式，可以求出單一向量本身的偏態程度。另 \bar{x}_i 表示第 i 個區塊之 M 個灰階平均值之平均所在位置， m_{i2} 與 m_{i3} 分別表示第 i 個區塊的二級動差 (second moment) 與三級動差 (third moment)，則經由下列公式可以算出偏態值 g_{i1} ：

$$g_{i1} = \frac{m_{i3}}{m_{i2} \sqrt{m_{i2}}},$$

$$\text{其中 } \bar{x}_i = \frac{\sum_{j=1}^M j \times x_{ij}}{\sum_{j=1}^M x_{ij}}, \quad m_{i2} = \frac{\sum_{j=1}^M (j \times x_{ij} - \bar{x}_i)^2}{\sum_{j=1}^M x_{ij}}, \quad \text{且 } m_{i3} = \frac{\sum_{j=1}^M (j \times x_{ij} - \bar{x}_i)^3}{\sum_{j=1}^M x_{ij}}。$$

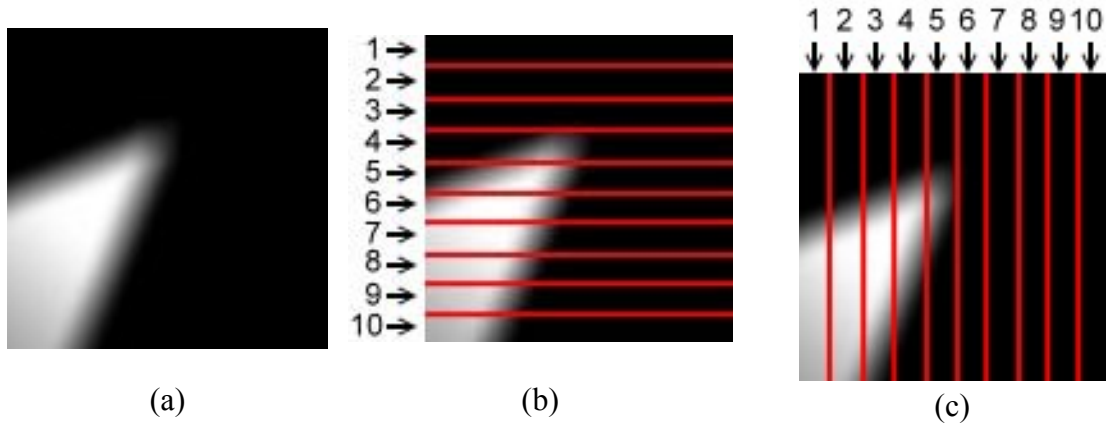


圖 4-9、(a)注意力圖像，(b) (a)之橫向切割圖，(c) (a)之縱向切割圖。

由上述公式觀察可知，當 $g_{il} > 0$ ，表示區塊中的亮點集中在左半部，即灰階值平均值之平均所在位置落在灰階平均值中位數所在位置的右邊，其所代表的是影像中的漸層分佈屬於正偏態。反之，若 $g_{il} < 0$ ，表示區塊中的亮點集中在右半部，即灰階值平均值之平均所在位置落在灰階平均值中位數所在位置的左邊，其所代表的是影像中的漸層分佈屬於負偏態。而當 $g_{il} = 0$ ，表示區塊中的亮點分佈為左右對稱或者無亮點分佈。圖 4-10 即顯示圖 4-9(b)之注意力圖像橫向切割所計算出的 g_{il} 值，其中第三至第十區塊之漸層分佈皆為正偏態，即亮點都集中在左半部。第一及第二區塊的偏態值皆為零，那是因為這兩個區塊皆無亮點存在。

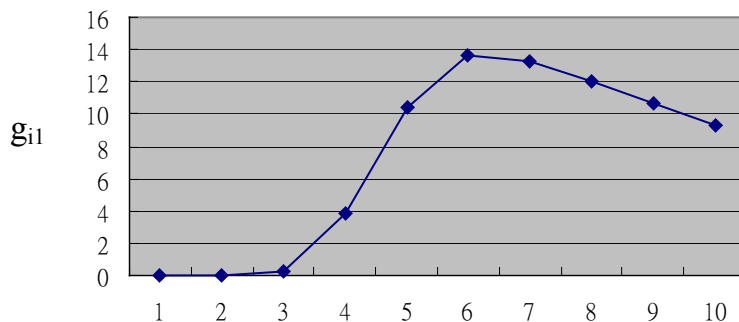


圖 4-10、圖 4-9(b)各區塊之偏態值。

同理將圖 4-9(a)做縱向切割均分為十等分的區塊，如圖 4-9(c)，再分別做橫向灰階平均，照上述的步驟亦可以求得十項偏態值，如圖 4-11 所示。其中第一至第四區塊之漸層分佈皆為負偏態，即亮點都集中在右半部。將縱向灰階平均與橫向灰階平均分別求出之十組偏態值以向量表示，就成為本論文所謂定性特徵的特徵向量，以 $(f_{w_1}^1, f_{w_1}^2, \dots, f_{w_1}^m)$ 、 $(f_{w_2}^1, f_{w_2}^2, \dots, f_{w_2}^m)$ 、 $(f_{w_3}^1, f_{w_3}^2, \dots, f_{w_3}^m)$ 、 $(f_{w_4}^1, f_{w_4}^2, \dots, f_{w_4}^m)$ 與 $(f_{w_5}^1, f_{w_5}^2, \dots, f_{w_5}^m)$ 分別代表第一至第五個視窗分割後所得的特徵向量。在本論文中 $m = 20$ ，前十項為橫向切割所得到之偏態值，後十項為縱向切割所得到之偏態值。

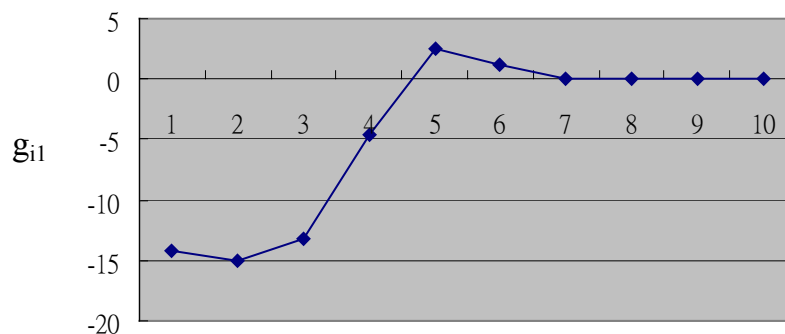
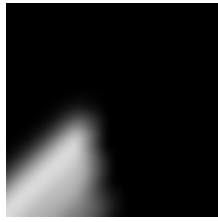
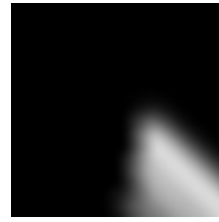


圖 4-11、圖 4-9(c)各區塊之偏態值。

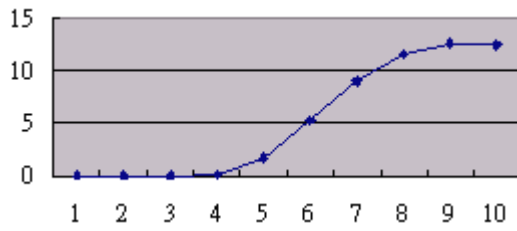
以圖 4-12 為例，可以發現圖 4-12 (a)與圖 4-12 (b)為左右對稱的注意力圖像。它們的橫向切割所得到之偏態值（圖 4-12 (c)與圖 4-12 (d)）是上下對稱的，至於它們的縱向切割所得到之偏態值（圖 4-12 (e)與圖 4-12 (f)）為左右對稱的。因此如果無法利用橫向切割所得到之偏態值來分辨它們的特徵，還可以利用縱向切割所得到之偏態值來分辨它們的特徵。如果這兩個方向還是無法辨別它們，亦可以考慮右斜或左斜 45 度角的方向切割，甚至其他方向角度的切割。但是在本研究中，橫向切割與縱向切割所提供的資訊已足夠，因此不考慮其他方向的切割。



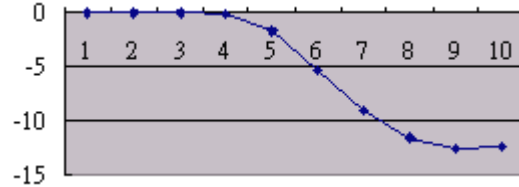
(a)



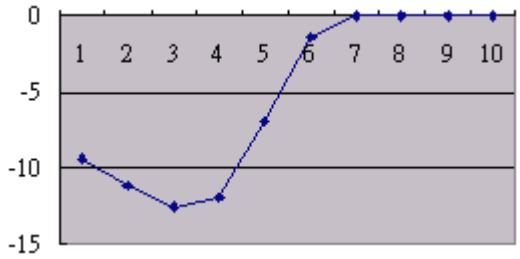
(b)



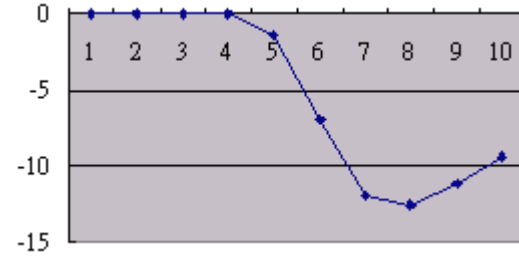
(c)



(d)



(e)



(f)

圖 4-12、(a)與(b)為對稱的注意力圖像，(c)與(d)各為(a)與(b)之橫向切割之偏態值，(e)與(f)各為(a)與(b)之縱向切割之偏態值。