



# 利用小波描述子解決三維物體姿態測定之有效方法

## Fast 3D Object Pose Determination Using Wavelet Descriptors

計畫編號：NSC90 – 2213 - E009 - 120

執行期限：90 年 8 月 1 日至 91 年 7 月 31 日

主持人：陳稔 國立交通大學資工系教授

E-mail : [zchen@csie.nctu.edu.tw](mailto:zchen@csie.nctu.edu.tw) Tel : 03-5731875

### 1. 中文摘要

本計畫主要之目的在利用 Gabor wavelet 多重尺度(multi-scale)及多重方位(multi-orientation)的特性來進行對於尺度與旋轉不變量之特徵點萃取與描述,並藉由特徵點群之 Delaunay 三角網格化對應來進行姿態測定之研究,並提出一有效之方法來解決物件被部分遮蔽之問題。

**關鍵詞**：姿態測定、特徵點、Delaunay triangulation

### Abstract

In this research we propose to use the Gabor wavelet to extract feature points in the scale-orientation space. The extracted feature points are robust to the variations in scaling, rotation, translation, and minor projection distortion. We construct the Delaunay triangle from the set of feature points to describe the geometrical structure of the object. The Delaunay triangle representation can be used to determine the pose of the object in the 3D space, in particular, under the circumstance of partial object occlusion.

**Keywords**: Pose determination, feature point, Delaunay triangulation

### 2. 緣由與目的

利用不同時間或不同攝影機拍攝所得的影像來求取一個 3D 物體的位置(position)與方位(orientation)的姿態測定(pose determination),在許多應用上扮演著相當關鍵性的角色,這已成為近幾年來相當熱門的研究主題。

姿態測定的應用,依其特徵來源之不同,大致上可分為三類:

- (1)3D to 3D 特徵對應[1-2]:利用立體視覺方法或其他深度獲得之設備,來估算物體之三度空間位置或動作。
- (2)2D to 3D 特徵對應[3-6]:利用已知位置之特徵 3D 資訊及其在影像平面之投影位置來進行相機校正或被動式飛彈導航。

(3)2D to 2D 特徵對應[7-12]:利用不同時間對同一物體拍攝所得之影像序列來決定物體之相對姿態或移動資訊。

姿態測定的做法主要分為二類: 其一是利用物體的輪廓線(silhouettes 或 contours)來進行,這即所謂 shape from silhouette(or contour)的技巧[9-12]用來描述物件的輪廓線包含 chain code、moment、spline、Fourier descriptor(FD)[11-14]等,其中又以 FD 最廣泛被近來的研究使用,其做法是將輪廓線先參數化成曲率(curvature)或空間座標(spatial coordinate)等特徵後再做 Fourier 轉換,以其 Fourier 係數來描述輪廓。由於影像中存在著雜訊及物體被遮蔽的可能,我們只能利用區域性的資料來進行比對,因此,FD 全域之特性已不再適用。雖然近年來已有相當多的研究利用 wavelet transform[15-17]多重解析度概念來描述輪廓線以提供區域性的比對資訊,但其對於起始點及遮蔽的問題仍然無法獲得有效的解決方案。

另一類是利用特徵來進行比對,用來比對之特徵可分為點、直線、曲線及邊角等幾種,利用此類特徵將面臨到在影像間存在不同尺度、旋轉方位之情況下的特徵萃取及對應問題。

本計畫主要目的在利用 Gabor wavelet 多重尺度(multi-scale)及多重方位(multi-orientation)的特性來進行對於尺度與旋轉不變量之特徵點萃取與描述,進一步藉由特徵點群之 Delaunay 三角網格化對應來進行姿態測定之研究,並提出一有效之方法來解決物件被部分遮蔽之問題。

### 3. 研究方法

本研究為 2D to 2D 特徵對應的姿態測定方法,做法係預先建立一參考影像庫(reference image library),其中,在影像庫內每一影像之姿態為已知,在獲得感測影像(sensory image)後,利用物件之特徵描述與預先處理之參考影像庫特徵進行比對,以相似度評估法則找出影像庫中最接近的參考影像,再對照出其姿態資訊。

由於感測影像與參考影像庫中之影像係利用攝影機對於不同姿態的物件拍攝而得，因此，我們可以有以下二點假設：

1. 在攝影機鏡心不動的情況下，不同距離或旋轉角度所拍攝的影像間存在著平面投影轉換矩陣  $M_{3 \times 3}$  關係，而當物體大小與攝影機到物體之距離比例很小的情況下，則通常可簡化為 affine 轉換之關係。
2. 相同姿態物件所拍攝影像，由於感測影像與對應之參考影像間幾乎只存在旋轉及均勻尺度 (uniform scaling) 的變化 (如果感測影像中的物體不在相機光軸上則會有些許差異)，如果我們可以獲得其間之轉換矩陣，則可以此轉換矩陣之數值來作為相似性評估之量度。

二張影像間之轉換矩陣可以由三個以上匹配之點對應來求取，一般而言，點對應包含初始比對與消除錯誤比對之程序 (relaxation, maximum clique 等)，此種全域搜尋的做法往往相當耗時；因此，在本計畫中，我們提出以三角形比對取代傳統點對應的方法，這種做法除了計算速度較快外，其最主要它可以解決物件被部分遮蔽的問題。另外，我們並考慮雜訊影響所造成之特徵點位置飄移現象，進一步藉由 ICMP (Iterative Closest Matching Pair) 方法以疊代方式逐步引入其他對應點來求得較精確之轉換矩陣。

### 3.1 參考影像庫之建立 (Training stage)

圖 1 顯示之參考資料庫，其建立方式係由攝影機光軸對著物件中心，在攝影機不動的情況下，對不同的物件姿態拍攝而得，其中垂直軸方向係每隔  $5^\circ$  變更 roll 的角度拍攝，水平軸方向則每隔  $5^\circ$  變更 pitch 的角度拍攝，對於參考影像庫中的每張影像，可以事先離線進行以下之特徵粹取與描述處理，以進行後續之比對作業。

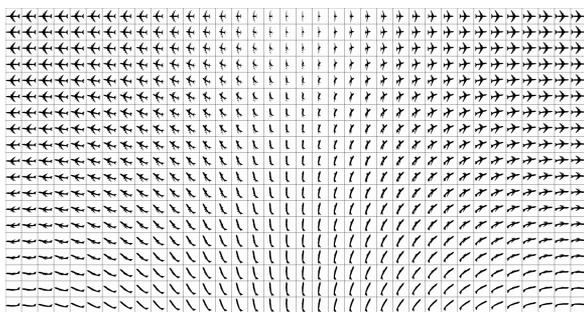


圖 1 參考影像庫

#### 3.1.a 特徵點粹取與描述：

利用一組多重尺度 (multi-scale) 及方向 (orientation) 之 Gabor wavelet (Eq.1) 對影像進行濾波

(filtering) 處理，從其所獲得的 Gabor-filtered image 的特性進行分析，我們發現當在適當的參數設計下，在物件的輪廓線上將會有 isolated dominant points 存在，而這些 dominant points 具有與平移，旋轉及尺度不變量之特性，可進行粹取為特徵點。

$$g^{s, \theta, l}(x, y) = \frac{1}{2\pi\alpha\sigma_s^2} \exp\left\{-\left[\frac{x'}{\sigma_s}\right]^2 + \left[\frac{y'}{\alpha\sigma_s}\right]^2\right\} \cos(jw_s x') \quad (1)$$

$$\text{where } \begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos\theta_l & \sin\theta_l \\ -\sin\theta_l & \cos\theta_l \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

每一個特徵點可以由一組特徵向量  $\vec{V}(x, y)$  (Eq.2) 來描述，其是由 filter response 所組成的尺度-方向空間 (scale-orientation space) 中，取其對應的 principal scale ( $s_d$ ) 且以 principal orientation ( $l_d$ ) 當起始點作 cyclically rotation 的 filter response  $R^{(s_d, l_d)}(x, y)$  所組成，這使得點對應的比對可以應付平移、旋轉與尺度變化的問題。

$$\vec{V}(x, y) = [R^{(s_d, l_d)}(x, y), R^{(s_d, l_d+1)}(x, y), \dots, R^{(s_d, l_d+L)}(x, y)] \quad (2)$$

#### 3.1.2 特徵點群之 Delaunay 三角網格化

離散的特徵點群可以進一步藉由 Delaunay triangulation 的方法 [18] 來建立三角網格，用以呈現出特徵點群間的 2D 空間結構關係 (也就是 object structure)，Delaunay triangulation 的特性是在相同的點集合下，具有建構結果的唯一性，而當點遺失時只會破壞原來與此點相連之三角網，其他三角網之結構仍會保存，將其運用在比對之用途上，功能為：

1. 提供 2D 的物件結構描述，較之 curvature 1D 的做法有較高的物件區辨能力。
2. 三角形的夾角在旋轉及尺度的變化下是不變的，可以解決感測影像與參考影像間存在著旋轉及尺度變化之比對問題。
3. 對於不同姿態的物件，由於其 Delaunay triangulation 的結果，三角形的外型會有相當程度的差異，因此，可以藉由較快速的三角形比對方法很快過濾掉。
4. 對於相似性的姿態，則只要找到其中一個對應的三角形，即可藉由此三角形的三個點座標來求取轉換矩陣。
5. 當點遺失時只會破壞原來與此點相連之三角網，其他三角網之結構仍會保存，這在解決物體被遮蔽 (特徵點遺失) 時的比對有相當好的適用性。

#### 3.2 姿態決定 (Pose determination) stage：

在獲得感測影像 (sensory image) 後，即進行特徵點之粹取及特徵點群之 Delaunay 三角網格化處

理，以據以與參考影像庫進行比對找出影像庫中最接近的參考影像，再對照出其姿態資訊，比對的做法依序為：

### 3.2.1 Delaunay 三角形比對

假設  $\Delta_i$  與  $\Delta_j$  分別為感測影像與參考影像上的一個三角形，其三角形三個頂點之特徵點編號分別為  $(p_{i1}, p_{i2}, p_{i3})$  ( $q_{j1}, q_{j2}, q_{j3}$ )，而三角形的三個夾角依大小排序分別為  $(\theta_{i1}, \theta_{i2}, \theta_{i3})$  ( $\phi_{j1}, \phi_{j2}, \phi_{j3}$ )，此三角形的夾角在旋轉與尺度變化的情況下為不變的，因此，可以作為三角形比對之依據，但缺點是對於具有尺度差異的類似三角形並無法區分，此時，可以再藉由特徵點所具有的 principal orientation 資訊再進一步區分。

令特徵點  $(p_{i1}, p_{i2}, p_{i3})$  ( $q_{j1}, q_{j2}, q_{j3}$ ) 之 principal orientation 分別為  $(l_{p_{i1}}, l_{p_{i2}}, l_{p_{i3}})$  與  $(l_{q_{j1}}, l_{q_{j2}}, l_{q_{j3}})$ ，則在任二個 principal orientation 不平行的情況下，此二組 principal orientation 可以分別夾出三角形  $\Delta_{ii}$  與  $\Delta_{ij}$ 。

則  $\Delta_i$  與  $\Delta_j$  被認為是匹配的準則為：

1.  $|\theta_{i1} - \phi_{j1}| + |\theta_{i2} - \phi_{j2}| < \text{Threshold}$
2. 如果  $(l_{p_{i1}}, l_{p_{i2}}, l_{p_{i3}})$  與  $(l_{q_{j1}}, l_{q_{j2}}, l_{q_{j3}})$  均可夾出三角形  $\Delta_{ii}$  與  $\Delta_{ij}$ ，則  $\frac{\text{area}(\Delta_{ii})}{\text{area}(\Delta_j)} \cong \frac{\text{area}(\Delta_{ij})}{\text{area}(\Delta_j)}$

Otherwise,  $|\theta_{ii} - \phi_{ij}| < \text{Threshold}$

其中  $\theta_{ii}, \phi_{ij}$  分別為平行之 principal orientation 與第三個 principal orientation 的夾角。

如果  $\Delta_i$  與  $\Delta_j$  為一對應之三角形對，則可將其頂點  $(p_{i1}, p_{i2}, p_{i3})$  ( $q_{j1}, q_{j2}, q_{j3}$ ) 依角度大小排序得到三組點對應求取其間之 Affine 轉換矩陣  $M$

$$\begin{bmatrix} x_{p_i} \\ y_{p_i} \\ 1 \end{bmatrix} = M_{3 \times 3} \begin{bmatrix} x_{q_j} \\ y_{q_j} \\ 1 \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_{q_j} \\ y_{q_j} \\ 1 \end{bmatrix}$$

### 3.2.2 轉換矩陣之再精進處理

進一步考慮 noise 的影響所造成特徵點位置飄移，使得求得之轉換矩陣不精確之可能性，可以把先前求得之 Affine 轉換矩陣  $M$  當作一初始解，再利用 ICMP 以疊代方式逐步引入其他對應點來求得較精確之轉換矩陣，演算法為：

#### Algorithm ICMP

Input:

1. Two point set  $Q_c = \{q_1, q_2, \dots, q_{N_q}\}$  and

$P_c = \{p_1, p_2, \dots, p_{N_p}\}$  with their feature vectors.

2. The initial estimate of the transformation  $M_0$ .

Output: The transformation  $M_f$

Initial:  $k=0$ ;

Begin

Repeat until  $CPS_k \cup CPS_{k-1} = CPS_{k-1}$

a. Build up the set of corresponding points  $CPS_k$

$$CPS_k = \bigcup_{i=1}^{N_q} (q_i, CC(T_{k-1}(q_i), P_c))$$

where  $CC(T_{k-1}(q_i), P_c) = p_j$

if

(a)  $\min_{p_j} |T_{k-1}(q_i) - p_j|$  with  $|T_{k-1}(q_i) - p_j| \leq d_k$

where  $d_k = d_0 / 2^k$

and  $S(q_i, p_j) > \text{Threshold}$

where  $S(\vec{v}_{q_i}, \vec{v}_{p_j}) = \vec{v}_{q_i} \cdot \vec{v}_{p_j} / \|\vec{v}_{q_i}\| \cdot \|\vec{v}_{p_j}\|$  with the vector factors  $\vec{v}_{q_i}, \vec{v}_{p_j}$  of  $q_i, p_j$

(b)  $(\vec{e}_{Q_k} \cdot \vec{e}_{Q_{k-1}})(\vec{e}_{P_j} \cdot \vec{e}_{P_n}) \geq 0$  (Convexity compatibility check)

b. Compute the new transformation  $M_k$  using all points in  $CPS_k$

c.  $k = k + 1$

end

$M_f = M_k$

End

### 3.2.3 相似度量測

感測影像  $I_s(x_p, y_p)$  與參考影像  $I_r(x_q, y_q)$  間存在著 Affine 轉換關係

$$\begin{bmatrix} x_{p_i} \\ y_{p_i} \end{bmatrix} = \begin{bmatrix} s_1 & 0 \\ 0 & s_2 \end{bmatrix} \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x_{q_j} \\ y_{q_j} \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix}$$

$$= \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} \begin{bmatrix} x_{q_j} \\ y_{q_j} \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix}$$

由於  $\begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} = \begin{bmatrix} s_1 \cos\theta & s_1 \sin\theta \\ -s_2 \sin\theta & s_2 \cos\theta \end{bmatrix}$  係描述其間

之旋轉及尺度的變化關係，而感測影像與正確對應之參考影像中亦幾乎只存在著旋轉與均勻之尺度變化 (uniform scaling)，因此，我們可以下式之類似度量測式中將旋轉因素消除，只保留尺度之變化，利用判斷均勻尺度變化之程度來作為相似性量度之依據。

$$S(I_s, I_r) = \min \left( \left| \frac{m_{11} \times m_{12}}{m_{22} \times m_{21}} \right|, \left( \frac{m_{11} \times m_{12}}{m_{22} \times m_{21}} \right)^{-1} \right)$$

where  $\frac{m_{11} \times m_{12}}{m_{22} \times m_{21}} = \frac{s_1 \cos\theta}{s_2 \cos\theta} \times \frac{s_1 \sin\theta}{-s_2 \sin\theta} = -\frac{s_1^2}{s_2^2}$

and  $0 \leq S(I_s, I_r) \leq 1$

當感測影像與參考影像完全吻合時 (uniform

scaling),  $S(I_s, I_r) = 1$ ; 雖然我們會面臨感測影像中的物體不在相機光軸上, 也就是成像不在影像中間區域之可能, 使得與理想情況有點差異, 但由於物件在視域(field of view)的範圍內位移量有限, 根據實驗驗證仍可以此相似性量度方法區辨出來。

#### 4. 實驗結果

我們針對二個不同姿態的感測影像進行實驗：  
實驗一：

感測影像(圖 2(a))與實際對應的參考影像間具有旋轉與均勻尺度變化的關係, 且物體不在相機光軸上, 圖 2(b)顯示所粹取之特徵點位置與其 principal scale 及 principal orientation, 圖 2(c)則為特徵點群之 Delaunay triangulation 結果

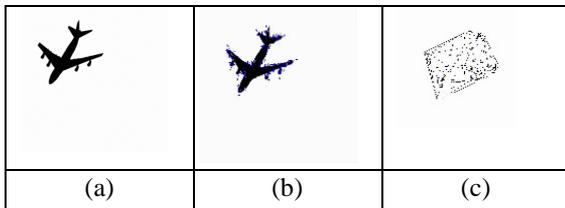


圖 2 (a) 感測影像 (b) 特徵點位置與其 principal scale 及 principal orientation (c) 特徵點群之 Delaunay triangulation 結果

參考影像庫搜尋的結果, 圖 3(e) 的物件獲得最高之比對係數, 我們可以得到感測影像之姿態為圖 3(e) 所對應的姿態參數, 另為表現相似圖形的區辨能力, 圖 3 中其他圖形(a-d, f-i) 則顯示相鄰姿態的比對係數。圖 4 則顯示其他姿態差異較大之參考影像比對結果。

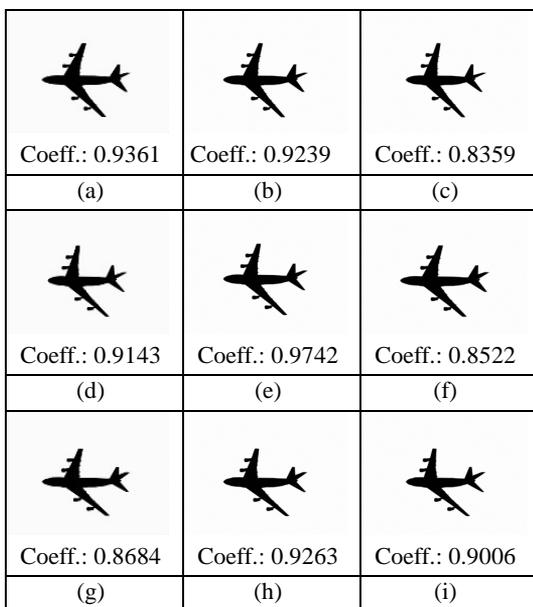


圖 3 (e) 與感測影像最近似的參考影像, (a)-(d), (f)-(h) 其他相鄰姿態的參考影像與比對係數

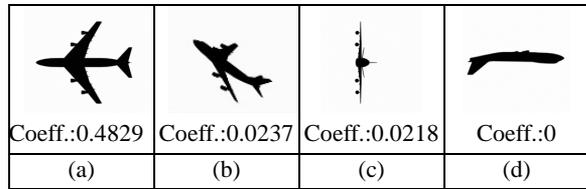


圖 4 (a)-(d) 其他姿態差異較大之參考影像比對結果

實驗二：

圖 5(a) 顯示一個被部分遮蔽的物件, 其特徵點群之 Delaunay triangulation 結果如圖 5(b) 所示, 與參考影像庫比對的結果, 顯示圖 5(c) 為比對係數最高的物件, 其比對係數為 0.9562, 由二者的 Delaunay triangulation 結果(圖 5(b), (d)), 我們可以觀察到除了在因遮蔽所遺失點附近之三角網被破壞外, 其於地方的三角網結構仍然被保存, 因此, 藉由未被破壞的三角網我們仍能進行比對工作。

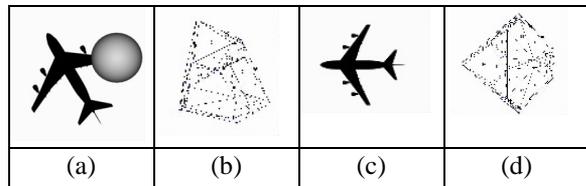


圖 5 (a) 被部分遮蔽的物件. (b) 圖(a)之 Delaunay triangulation 結果. (c) 比對到的參考影像 (d) 圖(c)之 Delaunay triangulation 結果.

#### 5. 結論

本計畫為 2D to 2D 特徵對應的姿態測定方法, 做法是預先建立一參考影像庫, 在獲得感測影像後, 即進行特徵點之粹取及特徵點群之 Delaunay 三角網格化, 據以與參考影像庫進行比對以找出影像庫中最接近的參考影像, 再對照出其姿態資訊。

我們提出一個以特徵點群所組成 Delaunay triangulation 的快速比對方法, 這種方法提供 2D 的物件結構描述, 較之 curvature 1D 的做法有較高的物件區分能力, 而且當點遺失時只會破壞原來與此點相連之三角網, 其他三角網之結構仍會保存, 實驗證明這對於解決物體被遮蔽(特徵點遺失)時的比對有相當好的成果。

#### 6. 計畫成果自評

研究成果與計畫所提之預期目標相符, 即將投稿至國際期刊發表。

#### 7. 參考文獻

[1] Stanilav and Franjo, "Planning sequences of views for 3-D object recognition and pose determination", *Pattern Recognition*, Vol.31, No.10, pp.1407-1417, 1998.  
[2] F. Quek, R. Jain et al. "An abstraction-based approach to 3-D pose determination from range

- images”, *IEEE Transactions on PAMI*, Vol.15, No.17, pp.722-736, 1993.
- [3] S. Linnainmaa, D. Harwood et al. “Pose determination of a three-dimensional object using triangle pairs”, *IEEE Transactions on PAMI*, Vol.10, No.5, pp.634-647, Sep 1988.
- [4] J. Batista, H. Araujo et al. “Iterative multistep explicit camera calibration”, *IEEE Transactions on Robotics and Automation*, Vol.15, No.5, pp.897-917, 1999.
- [5] S. Gold, A. Rangarajan et al. “New algorithms for 2D and 3D point matching - pose estimation and correspondence”, *Pattern Recognition*, Vol.31, No.8, pp.1019-1031, 1998.
- [6] S. H. Joseph, “Optimal Pose Estimation in Two and Three Dimensions”, *Computer Vision and Image Understanding*, Vol.73, No.2, pp.215-231, 1999.
- [7] S. Sullivan and J. Ponce, “Automatic model construction and pose estimation from photographs using triangular splines”, *IEEE Transactions on PAMI*, Vol.20, No.10, pp.1091-1096, Oct 1996.
- [8] Y. A. Cass, “Robust affine structure matching for 3D object recognition”, *IEEE Transactions on PAMI*, Vol.20, No.11, pp.1265-1274, Nov 1998.
- [9] K. Stark, S. Fuchs, “A method for tracking the pose of known 3-D objects based on an active contour model”, *IEEE Proceedings of ICPR*, pp.905-909.
- [10] M. Jones and J. P. Oakley, “Registration of image sets using silhouette consistency”, *IEE Proc.-Vis. Image signal process.*, Vol.147, No.1, February 2000.
- [11] K. Arbter et al., “Application of Affine-Invariant Fourier Descriptor to Recognition of 3-D Objects”, *IEEE Transactions on PAMI*, vol.12, No.7, pp. July 1990.
- [12] Z. Chen and S. Y. Ho, “Computer vision for robust aircraft recognition with fast library search”, *Pattern Recognition*, Vol.24, No.5, pp.375-390, 1991.
- [13] E. Bakry, Abo-Elsoud, Kamel, “Automatic face recognition system using neural networks”, *Circuits and Systems, 2000. Proceedings .ISCAS 2000 Geneva*, vol.3, 543–546, 2000.
- [14] V. Solachidis, N. Nikolaidis, I. Pitas, “Watermarking polygonal lines using Fourier descriptors”, *Acoustics, Speech, and Signal Processing, 2000. ICASSP '00.*, Volume: 4, pp. 1955–1958, 2000
- [15] G. C. H. Cheng et al., “Wavelet descriptor of planar curves : theory and applications”, *IEEE Transaction on Signal Processing*, Vol.5, No.1, pp.56-70, Jan 1996.
- [16] Q. M. Tieng and Wageeh, “Wavelet-based affine invariant representation: a tool for recognizing planar objects in 3D space”, *IEEE Transactions on PAMI*, Vol.19, No.8, pp.846-857, Aug 1997.
- [17] H. S. Yang and S. U. Lee, “Recognition of 2D Object Contours Using Starting-Point-Independent Wavelet Coefficient Matching”, *Journal of visual communication and image representation*, Vol.9, No.2 pp.171-181, June 1998
- [18] Y. Tao and W. I. Grosky, Image Indexing and Retrieval Using Object-based Point Feature maps, *Journal of Visual Languages and Computing*, vol.11(2000), pp.323-343.