

# ỨNG DỤNG MẠNG NEURON NHÂN TẠO NHẬN DẠNG ĐỐI TƯỢNG ẢNH CHUYỂN ĐỘNG

Hoàng Mạnh Hà, Vũ Đình Thành

Khoa Điện – Điện tử, Trường Đại Học Bách Khoa – Đại học Quốc gia Tp. HCM

(Bài nhận ngày 14 tháng 7 năm 2005, hoàn chỉnh sửa chữa ngày 20 tháng 9 năm 2005)

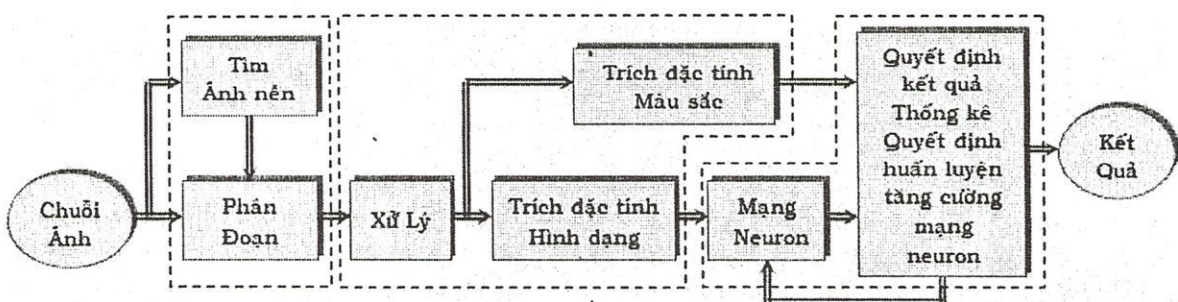
**TÓM TẮT:** Bài báo giới thiệu một giải thuật xử lý ảnh cho phép tách đối tượng ảnh chuyển động ra khỏi vùng nền cố định. Việc phân lớp đối tượng dựa trên đặc trưng về hình dạng và màu sắc của đối tượng. Mạng neuron được sử dụng cho việc phân lớp. Quá trình huấn luyện cho mạng neuron được thực hiện dựa trên các mẫu cho trước. Đồng thời, trong quá trình hoạt động, mạng neuron còn được tự động huấn luyện tăng cường, nhờ đó giải thuật nhận dạng có tính thích nghi với những thay đổi về hình dạng của đối tượng.

## 1. GIỚI THIỆU

Mục tiêu của quá trình tự động hoá sản xuất luôn luôn là sự hiệu quả và tối ưu về kinh tế, về năng suất và thời gian, đồng thời cũng yêu cầu hệ thống đạt được độ chính xác và tính khách quan trong quá trình xử lý. Trong số các bước xử lý căn bản này, việc tự động nhận dạng một vật thể chuyển động (một sản phẩm trên dây chuyền, một đối tượng đặc trưng cần được phân tích trong tập hợp nhiều đối tượng,...) là một trong những nội dung có ý nghĩa quyết định.

Việc nhận dạng đối tượng ảnh là một vấn đề khá phức tạp và không có lời giải tổng quát cho mọi trường hợp. Trong mỗi bài toán cụ thể, chúng ta chỉ có được giải thuật áp dụng cho một số vật thể và với các giới hạn của bài toán đó [1]. Bài viết giới thiệu một mô hình hệ thống ứng dụng mạng neuron nhân tạo trong việc nhận dạng đối tượng ảnh chuyển động. Phương pháp dùng neuron không phải là vấn đề hoàn toàn mới trong kỹ thuật, nhiều giải thuật xử lý và nhận dạng dùng neuron đã được tiến hành, nhưng các mô hình và thuật toán tái huấn luyện trong từng trường hợp cụ thể là khác biệt nhau [2]. Mục tiêu của các phép nhận dạng ở đây là sự thích nghi với quá trình thay đổi hình dạng của đối tượng (vị trí quan sát khác nhau, góc nghiêng khác nhau, độ chiếu sáng khác nhau,...).

Quy trình nhận dạng bao gồm: tách đối tượng chuyển động ra khỏi vùng nền cố định, xử lý về mặt hình học, chuẩn hoá, trích các đặc trưng về hình dạng, đặc trưng về màu sắc, huấn luyện mạng neuron từ những mẫu cho trước, và sau cùng là đưa ra một giải pháp để hệ thống nhận dạng có thể tự động huấn luyện tăng cường cho mạng neuron làm tăng tính thích nghi. Hình 1 mô tả sơ đồ khối của hệ thống này, gồm 3 phần chính: Tách đối tượng; Xử lý ảnh đối tượng và trích đặc tính; Mạng neuron, huấn luyện và huấn luyện tăng cường mạng neuron. Các kết quả ở ngõ ra được so sánh với chuỗi ảnh ngõ vào.



Hình 1

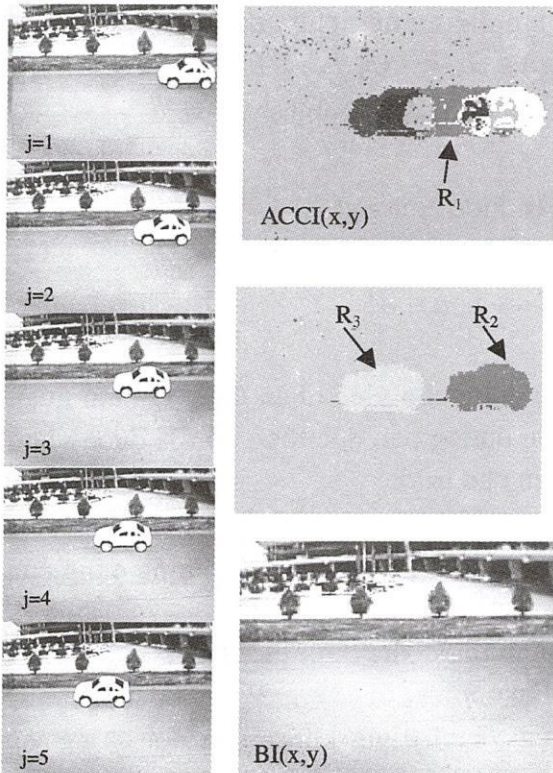
## 2. TÁCH ĐỐI TƯỢNG

Việc tách đối tượng ảnh chuyển động ra khỏi vùng nền cố định sẽ được thực hiện một cách dễ dàng nếu như ta có được ảnh nền chỉ chứa các đối tượng tĩnh. Việc tìm ảnh nền được thực hiện thông qua việc xác định *ma trận tích lũy sai biệt* như sau [3, 4]:

Giả sử chuỗi ảnh vào là  $I_1(x,y), I_2(x,y), \dots, I_n(x,y)$  với  $x = \overline{1, Xsize}; y = \overline{1, Ysize}$ .  $Xsize, Ysize$  là kích thước của mỗi ảnh. Ảnh đầu tiên  $I_1(x,y)$  được chọn làm ảnh tham chiếu. *Ma trận tích lũy sai biệt*  $ACCI(x,y)$  có kích thước bằng với kích thước ảnh. Ban đầu, mỗi phần tử của ma trận tích lũy sai biệt được gán trị zero. Mỗi ảnh tiếp sau đó được so sánh với ảnh tham chiếu:  $\Delta = |I_j(x,y) - I_1(x,y)|, j \geq 2$ .

Nếu  $\Delta \geq \theta_0$  thì phần tử tương ứng trong ma trận tích lũy được tăng 1 đơn vị:  $ACCI_j(x,y) = ACCI_{j-1}(x,y) + 1$ ,  $\theta_0$  là ngưỡng cho trước. Nếu  $\Delta < \theta_0$  và  $ACCI_{j-1}(x,y) > 0$  thì gán  $ACCI_j(x,y) = -1$ . Như vậy chúng ta thấy rằng các phần tử trong ma trận tích lũy sai biệt được gán trị -1 tương ứng với những pixel trước đó có sự sai biệt nhưng tại thời điểm hiện tại không còn sự sai biệt so với ảnh tham chiếu, chúng tương ứng với miền trống mà đối tượng để lại sau khi đối tượng di chuyển ra khỏi vị trí ban đầu của nó. Miền này được ký hiệu là R1. Khi miền R1 có diện tích lớn hơn ngưỡng diện tích  $S_{\theta}$  cho trước thì quá trình dừng lại.

Giả sử tại thời điểm k, ảnh  $I_k(x,y)$  đang được xem xét ( $k < n$ ).



Hình 2

Lúc này, ma trận tích lũy sai biệt  $ACCI(x,y)$  chứa các miền sau: miền R1; miền R0 tương ứng với các đối tượng tĩnh ( $ACCI(x,y)=0$ ); miền R2 tương ứng với miền đối tượng choán chỗ trong ảnh đầu tiên ( $ACCI(x,y)>0$ ); miền R3 tương ứng với miền đối tượng choán chỗ trong ảnh thứ k ( $ACCI(x,y)>0$ ).

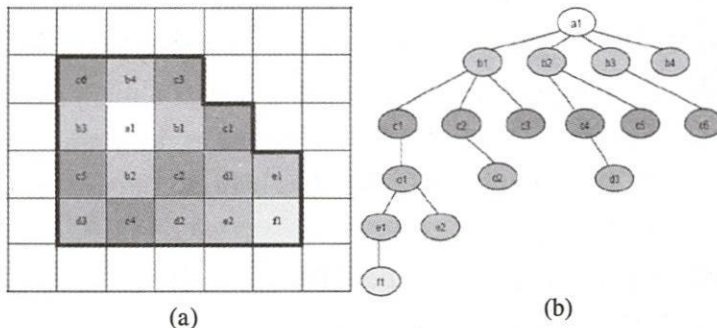
Miền R2 sẽ chứa giá trị lớn nhất của ma trận tích lũy sai biệt  $\max(ACCI(x,y))$ , đó chính là yếu tố để phân biệt miền R2 với miền R3. Việc sao chép các pixel tương ứng với miền R2 từ ảnh  $I_k(x,y)$  vào ảnh  $I_1(x,y)$  sẽ tạo ra được ảnh nền  $BI(x,y)$ . Hình 2 minh họa quá trình tìm ảnh nền từ chuỗi ảnh có chiếc ô tô di chuyển theo phương ngang.

Khâu phân đoạn trong hệ thống này thực hiện việc so sánh từng ảnh trong chuỗi ảnh với ảnh nền để có được ảnh sai biệt  $DI(x,y) = |I_j(x,y) - BI(x,y)|, j = \overline{1, n}$ . Ảnh của đối tượng  $OI(x_1, y_1)$  được tách ra từ ảnh  $I_j(x,y)$  dựa vào miền  $DI(x,y) > \theta_1$ , với  $\theta_1$

là ngưỡng cho trước. Ảnh của đối tượng có kích thước phụ thuộc vào kích thước của đối tượng trong ảnh  $I_j(x,y)$ , với  $x_1 = \overline{1, Xsize1}, y_1 = \overline{1, Ysize1}$  và  $Xsize1 < Xsize, Ysize1 < Ysize$ .

Những giải thích ở trên đã được đơn giản hoá chỉ cho ảnh đơn sắc. Trong thực tế, thuật toán có thể được dùng để xử lý trên chuỗi ảnh màu với 3 thành phần màu R,G,B.

Bên cạnh đó, thuật toán mở rộng miền [4] được sử dụng để xác định các miền liên kết. Thuật toán này được minh họa trong hình 3. Ban đầu một pixel thuộc miền liên kết được xác định (pixel  $a_1$ ) và được xem là pixel hạt nhân. Lần lượt, bốn pixel kế cận ( $b_1, b_2, b_3, b_4$ ) được xét có thuộc miền liên kết này hay không. Nếu nó thuộc vào miền liên kết này, nó lại được xem như pixel hạt nhân và cứ như vậy quá trình lặp lại, vùng xác định của miền được mở rộng. Hình 3 (b) minh họa cây với các nút là các pixel tương ứng trong hình 3 (a).



Hình 3

### 3. XỬ LÝ ẢNH ĐỐI TƯỢNG, TRÍCH ĐẶC TÍNH

#### 3.1. Xử lý ảnh đối tượng

Ảnh đối tượng sau khi được tách ra từ chuỗi ảnh có đường bao không trơn, chứa nhiều gai nhọn, rãnh nứt và các lỗ hổng. Khâu xử lý về mặt hình học loại bỏ các yếu tố không mong muốn kể trên thông qua hai thuật toán cơ sở : mở rộng đường bao (dilation) và co hẹp đường bao (erosion) [4, 5].

Xét hai tập điểm trong không gian  $Z^2$ : A là tập điểm miền chứa đối tượng, B là tập điểm tác động. Xét các điểm thành phần:  $a = (a_1, a_2) \in A$ ,  $b = (b_1, b_2) \in B$ .

Thuật toán mở rộng đường bao tập A bằng tập B được định nghĩa:

$$A \oplus B = \{x | (\hat{B})_x \cap A \neq \Phi\} \tag{1}$$

Thuật toán co hẹp đường bao tập A bằng tập B được định nghĩa:

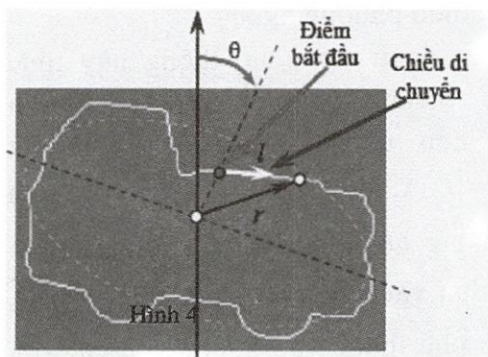
$$A \ominus B = \{x | (B)_x \subseteq A\} \tag{2}$$

Với :  $\hat{B} = \{x | x = -b, b \in B\}$  và:  $(A)_x = \{c | c = a + x, a \in A\}$

#### 3.2. Trích đặc tính

Đặc tính về hình dạng của đối tượng được sử dụng cho việc nhận dạng là hàm quan hệ giữa khoảng cách từ trọng tâm của đối tượng đến đường bao dọc theo độ dài đường bao tính từ một điểm bắt đầu cho trước và theo một chiều cho trước:

$$r = r(l) \tag{3}$$



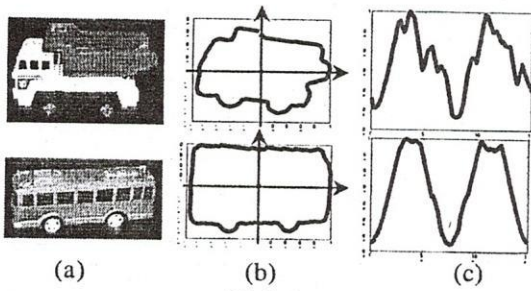
Hình 4

Hình 4 minh họa việc trích đặc tính hình dạng của đối tượng.

Điểm bắt đầu được xác định thông qua việc chuẩn hoá về chiều hướng đối tượng. Việc chuẩn hoá này dựa trên việc làm trùng khớp giữa đối tượng với hình elip có momen quán tính bằng với momen quán tính của đối tượng [6]. Từ đó, góc  $\theta$  tương ứng với điểm bắt đầu được suy ra.

Hình 5 minh họa 2 hàm quan hệ giữa  $r$  và  $l$  của 2 loại xe khác nhau. Hình 5(a) là ảnh đối tượng, hình 5(b) là đường bao của đối tượng đã

được chuẩn hoá, hình 5(c) là hàm  $r(l)$ .



Hình 5

Từ hàm  $r(l)$ , 150 điểm được trích ra hình thành một vector P. Nó chính là dữ liệu được đưa vào mạng neuron để nhận dạng. Nếu đối tượng có tính đối xứng thì việc đảo thứ tự các thành phần của vector P sẽ xác định được đặc tính hình dạng của đối tượng nhìn từ phía đối diện.

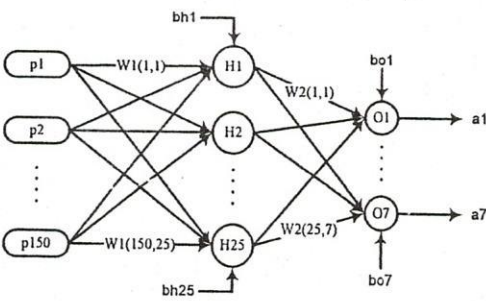
Ngoài đặc tính hình dạng, đặc tính màu sắc cũng được sử dụng trong việc phân biệt các đối tượng.

Thông qua việc chuyển đổi cơ sở không gian

màu từ RGB sang HSV (hue, saturation, value) [6, 8], màu sắc của đối tượng được định nghĩa bởi 9 sắc thái cơ bản: đen, xám, trắng, đỏ, vàng, xanh lá, xanh lơ, xanh dương, hồng. Thuật toán này có thể xác định được 3 mảng màu có diện tích lớn nhất trong ảnh của một đối tượng.

#### 4. MẠNG NEURON, HUẤN LUYỆN VÀ HUẤN LUYỆN TĂNG CƯỜNG MẠNG NEURON

Mô hình hệ thống ở hình 1 có tính tổng quát. Với mỗi ứng dụng nhận dạng cụ thể, chúng ta có thể xây dựng một mạng neuron để phân lớp đối tượng dựa trên đặc tính trích ra từ ảnh đối tượng. Trong bài báo này, một mạng neuron có cấu trúc như hình 6 được xây



Hình 6

dựng để phân lớp đối tượng dựa trên đặc tính về hình dạng của đối tượng. Mạng neuron này có 3 lớp: 1 lớp nhập tương ứng với 150 phần tử của vector P đã được nói ở trên, 1 lớp xuất bao gồm 7 neuron tương ứng với 7 lớp mẫu có thể phân biệt được, và 1 lớp ẩn bao gồm 25 neuron. Số neuron trong lớp ẩn được chọn lựa trong quá trình huấn luyện mạng neuron nhằm đảm bảo tối thiểu độ phức tạp trong cấu trúc mạng và sự hội tụ khi huấn

luyện mạng với tập mẫu cho trước [1], [5], [7]. Hàm truyền của các neuron lớp xuất và lớp ẩn là hàm sigmoid:  $f(x) = 1/(1 + e^{-x})$ . Với cấu trúc này, đáp ứng của các ngõ ra  $a_1 \div a_7$  đối với 150 phần tử ở ngõ vào  $p_1 \div p_{150}$  của vector P là:

$$a_i = f \left[ \sum_{j=1}^{25} W_2(j,i) \cdot f \left( \sum_{k=1}^{150} W_1(k,j) \cdot p_k + bh_j \right) + bo_i \right], \quad i = \overline{1,7} \quad (4)$$

Với  $W_1, bh$  là bộ trọng số của các neuron lớp ẩn,  $W_2, bo$  là bộ trọng số của các neuron lớp xuất.

Quá trình huấn luyện mạng neuron được thực hiện dựa trên 21 ảnh của 7 đối tượng mẫu. Mỗi đối tượng mẫu được chụp ở 3 góc độ khác nhau. Từ ảnh đối tượng mẫu, đặc tính hình dạng, cụ thể là vector P, được trích ra và đó chính là dữ liệu huấn luyện cho mạng neuron. Trong nghiên cứu này, 7 đối tượng mẫu là 7 loại xe khác nhau. Chúng có tính đối xứng do đó đặc tính hình dạng của chúng nhìn từ phía đối diện có thể được suy ra bằng cách đảo chiều các phần tử vector P. Ngoài các dữ liệu được trích trực tiếp từ ảnh đối tượng mẫu và bằng việc đảo chiều các phần tử vector P, tập dữ liệu huấn luyện còn chứa các thành phần có thêm các yếu tố nhiễu nhằm làm phong phú thêm tập dữ liệu huấn luyện và làm cho mạng neuron “quen” với những ảnh hưởng của nhiễu tác động lên đối tượng.

Việc huấn luyện mạng neuron là có giám sát. Tức là, đối với 1 vector P ở ngõ vào phải luôn kèm với 1 kết quả mong muốn ở ngõ ra – vector T. Việc huấn luyện xác lập bộ

trọng số của các neuron lớp xuất và lớp ẩn  $W_2, b_o, W_1, b_h$  làm cho mạng neuron có thể phân lớp được 7 lớp đối tượng tương ứng với 7 loại xe khác nhau.

Quá trình nhận dạng dùng mạng neuron có thể được huấn luyện tăng cường nếu hệ thống gặp phải những đối tượng có hình dạng thay đổi so với tập đối tượng mẫu đã được huấn luyện trước. Sự thay đổi đặc tính hình dạng của đối tượng dẫn đến có sai số nhất định tại ngõ ra mạng neuron. Nếu sai số này lọt trong một khoảng ngưỡng cho trước và đối tượng xuất hiện trong nhiều ảnh liên tiếp thì nó sẽ kích hoạt cơ chế tự động huấn luyện tăng cường cho mạng neuron theo đặc tính hình dạng của đối tượng đó. Điều này làm cho

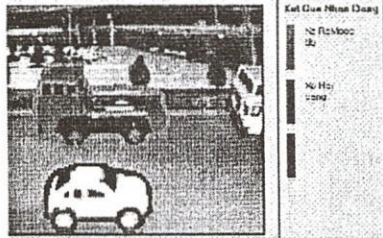


Hình 7

hệ thống nhận dạng có được tính thích nghi theo những thay đổi về hình dạng của đối tượng cũng như thích nghi được với ảnh hưởng của môi trường làm thay đổi hình dạng đối tượng. Ở hình 7 minh họa 1 trường hợp đối tượng có hình dạng thay đổi so với hình dạng của đối tượng mẫu ban đầu (xe ben với ben chưa hạ xuống).

## 5. KẾT QUẢ MÔ PHỎNG

Trong nghiên cứu này, hệ thống nhận dạng được xây dựng để nhận dạng các đối tượng chuyển động. Cụ thể, hệ thống có thể phân biệt 7 lớp đối tượng khác nhau. Hệ thống



(a)



(b)

Hình 8

có thể nhận dạng được nhiều đối tượng trong cùng một ảnh. Kết quả nhận dạng không bị ảnh hưởng bởi vị trí, kích thước và chiều hướng của đối tượng trong ảnh. Kết quả nhận dạng của hệ thống cũng ít bị ảnh hưởng bởi nhiễu. Hình 8 minh họa một số kết quả.

Tỉ lệ nhận dạng đúng dựa trên đặc tính hình dạng của hệ thống này là 98,4% (Hệ thống nhận dạng được kiểm tra với 250 ảnh của các đối tượng được trích ra từ chuỗi ảnh, số ảnh nhận dạng đúng là 246).

Với trường hợp đối tượng có hình dạng thay đổi làm kết quả nhận dạng có sai số nhất định, kết quả này sẽ kích hoạt cơ chế tự động học tăng cường của hệ thống. Sau khi được huấn luyện tăng cường, hệ thống hoàn toàn nhận dạng chính xác đối tượng đó nếu nó gặp đối tượng đó sau đó.

Tốc độ xử lý tách đối tượng và nhận dạng đạt trung bình 0.72giây/đối tượng. Tốc độ này được xét khi thử nghiệm giải thuật bằng phần mềm Matlab7.0 trên máy tính PentiumIII 866MHz tốc độ bus 133MHz và kích thước mỗi ảnh 120x160 pixel (RGB).

Mặc dù các nhận dạng dùng mạng neuron cho kết quả tương đối tốt như các mô phỏng ở trên, nhưng hệ thống còn một số giới hạn sau:

- Hệ thống chỉ làm việc chính xác khi cảnh nền cố định hoặc chỉ có sự thay đổi nhỏ.
- Việc nhận dạng chỉ dựa trên đặc tính trích từ hàm  $r(l)$  và đặc tính màu sắc do đó thiếu thông tin về cấu trúc của đối tượng dẫn đến việc số lớp có thể phân biệt được bị giới hạn.
- Hệ thống không tách được ảnh đối tượng khi chúng chồng lấp lên nhau hoặc chúng có sắc thái giống với sắc thái của phông nền.
- Đặc tính màu sắc còn phụ thuộc vào điều kiện chiếu sáng. Góc độ chiếu sáng cũng bị hạn chế để tránh việc gây ra bóng của đối tượng.

## 6. KẾT LUẬN

Giải thuật nhận dạng dùng mạng neuron được mô phỏng ở trong bài viết đã minh họa được khả năng thích nghi với những sự thay đổi về hình dạng của vật thể đối tượng, có khả năng tách vật thể chuyển động ra khỏi bức ảnh nền chứa các vật thể tĩnh. Quá trình huấn luyện mạng neuron là tự động, giúp cải tiến từng bước khả năng linh hoạt của quá trình điều khiển, phù hợp với các hệ thống tự động hoá. Giải thuật có thể được phát triển ứng dụng vào nhiều lĩnh vực thực tế như:

- Tự động nhận dạng loại xe tại trạm kiểm soát hoặc thu phí giao thông tự động.
- Tự động thống kê, phân loại trong ngành vi sinh học, huyết học.
- Tự động phân loại, đánh giá chất lượng sản phẩm trong dây chuyền sản xuất công nghiệp.
- Tự động phát hiện, nhận dạng mục tiêu trong quân sự ....

### APPLICATIONS OF NEURAL NETWORK FOR RECOGNISING MOVING OBJECTS FROM DYNAMIC IMAGES

Hoang Manh Ha, Vu Dinh Thanh

Department of Electrical & Electronics Engineering, University of Technology – VNU-HCM

**ABSTRACT:** *This paper presents an object recognition algorithm which extracts the moving objects out of its background region from a dynamic image. Shape feature and color feature are employed for objects characterising, and neuron network is used for their classifying. Initially, the neuron network is trained with a priori defined patterns for its recognition. Then, if the system finds any modification in object shape, the neural network will automatically learn these features. This will enable the system to adapt to modification of the object shapes while maintaining its recognising capability*

### TÀI LIỆU THAM KHẢO

- [1]. C. M. Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press, 1995.
- [2]. Carl G. Looney, *Pattern Recognition Using Neural Networks*, Oxford University Press, 1997.
- [3]. Eric D. Lester, *Feature Extraction, Image Segmentation, and Surface Fitting: The Development of a 3D Scene Reconstruction System*, The University of Tennessee, Knoxville, 1998.
- [4]. Rafael C. Gonzalez, Richard E. Woods, *Digital Image Processing*, Addison-Wesley, 1999.
- [5]. Sing-TZE Bow, *Pattern Recognition and Image Preprocessing*, Marcel Dekker, Inc, 2002.

- [6]. The Mathworks, *ImageProcessing Toolbox User's Guide*, The MathWorks Inc, 2003.
- [7]. The Mathworks, *Neural Networks Toolbox User's Guide*, The MathWorks Inc, 2003.
- [8]. Xiangjian HE, *2D\_Object Recognition with Spiral Architecture*, University of Technology, Sydney Australia, 1999.