一個受規範並以數學型態學為基礎之物件分割系統

計畫編號：NSC90-2212-E-002-109
執行期限：九十年八月一日至九十一年七月卅一日
主持人：吳家麟教授 國立臺灣大學資訊工程學系

一、 中、英文摘要

數學型態學對以幾何特性來分析信號及數位影像提供了完整而有系統的方法。因而，曾被廣泛的運用到各種信號與影像處理的應用中，例如：邊緣偵測，物件分割，雜訊壓制等等。

本計畫之主旨除了對數學型態學的各種運算子進行系統性之研討外，並推演出一些可運用於各種影像處理應用中的型態學濾波器。此外，根據所推演之型態學濾波器，本計畫提出一個受規範且以數學型態學為基礎的物件分割系統。

我們所提出的物件分割系統係由下列三個模組組成：

（1） 多尺度斜率計算模組

此模組依據較穩定之多尺度影像斜率計算方法，計算出斜率影像。

（2） 分水嶺轉換模組

根據（1）所得之斜率影像，本模組結合分水嶺演算法及極小值消去法提供較清晰的物件分割基礎單元。

（3） 適應性數學型態學運算模組

此模組提供具適應性能力之數學型態運算子，可精確的找到物件邊緣所在的範圍。

最後，影像位元及影像區域為基礎的分類法，將被用來區分物件邊界所在，以達成物件分割的目的。此外，該系統亦可結合物件追蹤法，將其運用範圍擴大到視訊物件分割的領域中。

關鍵詞：數學型態學，物件分割，多尺度斜率演算法，分水嶺轉換，適應性數學型態運算子。

Abstract

Mathematical morphology provides a systematic approach to analyze the geometric characteristics of signals or images, and has been applied widely to many applications, such as edge detection, object segmentation, noise suppression, and so on. The main purpose of this project is to provide an overview of mathematical morphology operators and develop some morphological filters which can widely be used in image processing. Furthermore, a morphology-based supervised object segmentation system has been proposed and realized.

The proposed system consists of three morphology-based modules: multiscale gradient algorithms, watershed transformation, and adaptive morphological operators. Multiscale gradient filters, on the basis of mathematical morphology, are robust to noise comparing to the traditional first-order differential or Laplacian operators. Two novel morphological gradient operators will be introduced and each has its pros and cons. The watershed transformation is a useful tool for initial segmentation but usually yields over-segmentation because of existing too many local minima. To avoid this problem, local minima elimination is involved and dominates

the result of segmentation. Some adaptive morphological operators have been derived to precisely find out the searching area. In this project the “maximizing edge strength” criterion will be proposed and examined to deal with the object segmentation problem. In the stage of “classification”, two kinds of approaches using hierarchical queue: pixel-based classification and region-based classification are studied in detail . Once extending to video object segmentation, it is our conjecture that the object with little motion can be tracked well due to the capability of adaptive morphological operators.

Keywords：Mathematical Morphology, Object Segmentation, Multiscale Gradient Algorithm, Watershed Transformation, Adaptive Morphological Operators.

二、 計畫之背景及目的：

From the perspective of multimedia data representation, object-based video representation provides a natural framework. Significant efforts have been devoted to object-oriented functionalities, including content-based indexing, content-based retrieval and content browsing systems. For instance, content-based indexing and retrieval involve a structuring of data in terms of low-level regions, semantic objects and so on. These features are addressed in the ongoing standard MPEG-7 and MPEG-21. In summary, video object segmentation has become a must-do preprocessing in view of multimedia data compression and representation.

There exist many algorithms for automatic image and video segmentation. Robert M. Haralick and Linda G. Shapiro [2] classified the image segmentation techniques as: measurement space guided spatial clustering, single linkage region growing schemes, hybrid linkage region growing schemes, centroid linkage region growing schemes, spatial clustering schemes, and split and merge schemes. Nikhil R. Pal and Sankar K. Pal [3] reviewed some traditional image segmentation techniques and included fuzzy set theoretic segmentation and neural network based techniques. Furthermore, most algorithms put emphasize on grayscale image, and some can extend to color image. In this section, we will focus on the algorithms for color image segmentation since it is the majority of image sequences.

Clustering-based algorithms have shown their capabilities on color image segmentation [4][5][6]. In [4], the k-means algorithm, a classical theme of supervised cluster analysis in statistical pattern recognition, is modified for unsupervised segmentation of color image in the chromaticity plane. It first makes use of only the chromatic information and performs 2D k-means algorithm in the u’v’ plane. Then the chromaticity clusters are associated with appropriate intensity value through the 1D version of the k-means algorithm. In [5], the proposed method detects image cluster in some linear decision volumes of the L*, a*, b* uniform color space. The detected clusters are projected onto the line of the Fischer discriminant for 1D thresholding. This algorithm is efficient because clusters are specified in linear decision volumes using only 1D histograms. A procedure similar to vector quantization is borrowed for cluster detection and a local feature analysis technique is also proposed in [6]. Some techniques based on mean-shift algorithm, which is a simple nonparametric procedure for estimating density gradients, have also been proposed [7][8].

Another trend of image segmentation is the fusion of different features to get a more reliable result [9][10]. In [9], the adaptive bayesian color segmentation based on Gibbs random field is performed first to form a segmentation map. At the same time, the edge map is built up by the color edge detector. Then regions in the segmentation map are split and merged by a region-labeling procedure to enforce their consistency with the edge map. Beside color information, motion is another useful feature for an image sequence. In [10], motion segmentation and color segmentation are integrated to get a promising result for extracting the moving vehicles on the road.

In the domain of supervised segmentation, snakes [11] had attracted lots of discussion. A snake is an energy-minimizing spline guided by external constraint forces and influenced by image forces that pull it toward features such as lines and edges. In [12], it extracts regions based on the multidimesional analysis of several image features by a spatially constrained fuzzy C-means algorithm. In the semantic step, the user add the points on the top of the semantic object. Many morphological approaches for image segmentation have also been proposed. One of powerful morphological tools for image segmentation is the watershed transformation [13][14]. Applying watershed transformation on gradient image could obtain an initial segmentation, but usually yields an over-segmented result. Thus, region merging can be performed as the post-processing, and efficient algorithms have been proposed [15]. A morphological clustering method has been illustrated in [16]. The histogram of RGB information in the color image is built up and then using the geometric property of closing to find out the dominant clusters. A morphology-based supervised segmentation is introduced in [17][18]. Its main concept is to create searching area where the real object resides by applying the morphological operators to the initial boundary given by the user. This project is based on this idea and includes adaptive morphological operators to improve the overall segmentation performance.

Moreover, mathematical morphology has also shown its power in video segmentation [19][20]. Actually, there already exist some systems combining the video object segmentation and the task of content-based retrieval [21][22][23]. Their main concept is to segment the first frame into homogeneous regions and form the semantic object defined by the user. In following frames, the semantic object is tracked by motion estimation and region aggregation.

The main purpose of this project is to provide an overview of mathematical morphology operators and derive some morphological filters which can widely be applied to various kinds of in image processing tasks. Furthermore, based on the derived morphological filters, a morphology based supervised object segmentation system has been proposed and realized.

三、 研究方法及進行步驟

A Supervised Object Segmentation System Based on Mathematical Morphology

1. The Proposed System

Figure 1 shows the flowchart of the proposed system. The aim of the "object definition" stage is to obtain the rough definition of the semantic object through user interaction. At the "region analysis" stage, the information of the features like edge strength in the image is extracted. Based on the information from "object definition" and "region analysis" modules, the "adaptive morphological operators" stage is devoted to properly create the band where the actual object boundary resides. Finally, the object boundary will be located at the "classification" stage.

In following sections, the functionalities of each stage are explained respectively.

2. Object Definition

During the past decades, many research works of image or video segmentation have concentrated themselves on unsupervised segmentation. However, the definition of "semantic" is ambiguous and only a human knows the real meaning of "semantic". That is why it is very difficult for the computer to accomplish automatic segmentation of semantic meaningful objects if lacking high level image understanding. On the other hand, a computer can perform a low level segmentation of an image into homogeneous regions based on some image features. Hence, the adoption of some form of user interaction in the segmentation process becomes commonly accepted and some research works of semi-automatic object segmentation have been proposed [11] [12] [17].

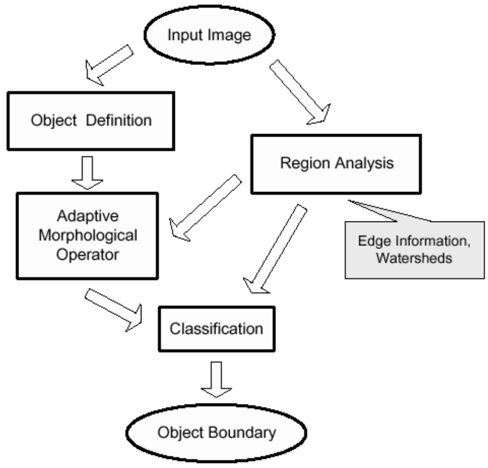

Figure 1: Flowchart of the proposed system

In this "object definition" stage, user assistance is involved to provide an approximation of the object boundary as the definition of the semantic object. In view of how a user interacts with the computer, two types of methods, pixel-based and contour-based approaches, are commonly used. In a pixel-based method, a user needs to input the positions of the boundary. On the other hand, only the key nodes to draw the outline of the boundary are required for contours-based method. Generally speaking, the contour-based approach is more intuitive and requires less human effort than the pixel-based one.

In this project, a contour-based polygonal method has been designed. A user clicks the mouse to provide the key nodes along the object. Then the key nodes are connected to form the polygonal approximation of the object boundary. Although the polygonal approach cannot approximate the complicated boundary well, it provides some information to define the semantic object.

3. Region Analysis

In the proposed system, the well-known watershed transformation is used to perform the region analysis. The reasons for choosing the watershed transformation include:

a. The watershed transformation can run in linear time using the algorithm proposed in [13].

b. The edge strength image is designed as an parameter of the adaptive morphological operators and it is also the only information required for the watershed transformation. Hence, no more calculation is needed for preparing the watershed transformation.

c. It is known that the watershed transformation usually has the problem of over-segmentation. From another point of view, the spot-like region can be used as the unit for the user to refine the segmentation result.

Figure 2: Flow diagram of typical watershed-based image segmentation

Figure 2 shows the procedures for applying watersheds to region analysis.

4. Adaptive Morphological Operators

Once obtaining the object definition by user assistance, the approximation of the real object boundary is not too far away. Hence, the task turns to creating the band of interest where the real object boundary resides. It is exactly the goal of "adaptive morphological operators". The concept is shown in Fig. 3.

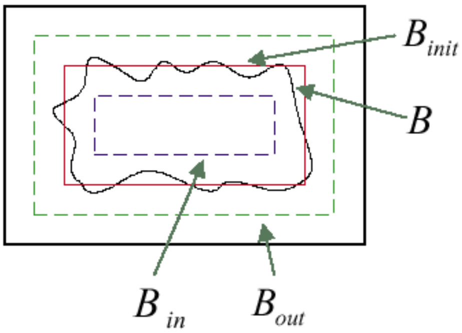

Figure 3 : Interior and Exterior Boundary

In Fig.3, B indicates the real object boundary, B_{init} is the initial boundary from the "object definition" stage. If we could find out the interior boundary B_{in} lying inside the object and the exterior boundary B_{out} lying outside the object, then we can limit the searching area where the real object boundary lies. In [17], the binary morphological operators have been applied to obtain B_{in} and B_{out}. While thinking each pixel inside the B_{init} as foreground and others as background, a binary image $I_{B_{init}}$ is created. Then binary dilation and binary erosion are performed to obtained two images $I_{B_{in}}$ and $I_{B_{out}}$:

$$I_{B_{in}} = (I_{B_{init}} \ominus SE)$$
$$I_{B_{out}} = (I_{B_{init}} \oplus SE)$$

where SE is the structure element.

By properly choosing the structure element SE, we can guarantee that

$$I_{B_{in}} \subseteq I_{B_{init}} \subseteq I_{B_{out}}$$

and, furthermore,

$$B_{in} \subseteq B \subseteq B_{out} \qquad (1)$$

The searching area then equals to $I_{B_{out}} - I_{B_{in}}$.

Note that the criterion in Eqn. (1) is quite strict and almost decides the accuracy of the segmentation result since wrong searching area will not contain right boundary. Moreover, the interior and exterior boundaries are obtained from the binary morphological operators; hence, the issue turns to how to choose the structure element properly. In this methodology, the size and shape of the structure element controls the tolerance of the displacement between the initial boundary and real object boundary. In [17], the structure element is chosen as a square of size 2. The ability of fault tolerance becomes relatively weak and careful user assistance is necessary. To tolerate the error between the initial boundary and real object boundary, an adaptive morphological operator is proposed to accurately create the band where the object boundary resides.

Most morphological operators for image or signal processing in the past belong to shift-invariant operators. In [21], V-operators have been investigated by us because the geometric interpretations of V-operators can be viewed as time-varying or space-varying versions generalized from those of shift-invariant ones. In fact, if an effective structure element assignment procedure is used, then a V-operator can adaptively change its processing scale according to the local characteristics of a signal [21].

The V-operator using progressive umbra-filling (PDF) procedure for adaptive assignment of structure element has been proposed in [21] for signal filtering. The reason why the adaptive morphological operators are employed in this project is to accomplish the aim stated in Eqn. (1)

Therefore, the procedure of adaptive assignment of structure element should be designed such that the domain of structure element whose center coincides with each pixel belonging to the initial boundary includes the real object boundary.

That is,

$$\forall(x,y)\in B_{init} \quad SE_{(x,y)} \cap B \neq 0 \qquad (2)$$

where SE(x,y) denotes to translate the structure element SE by a displacement vector (x,y).

5. Classification

After determining the searching area in which the real object boundary resides, we need to classify each pixel in the area as "object"(in) and "non-object"(out). Typically, the cluster centers need to be decided first, and then assign each undetermined unit to its cluster center that maximizes the similarity measurement.

In this stage, supposing that the cluster centers of "in"(object) and "out"(non-object) have already been located, two classification methods to group the unclassified units are illustrated in Fig. 4, pixel-wise classification and morphological watershed.

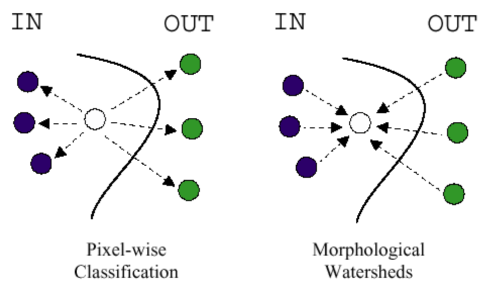

Figure 4 : Two types of classification

In pixel-wise classification, each pixel is considered independently so that the spatial relationship among neighboring pixels will not be take into account in the classification process. Hence, the result could be sensitive to noise, and the pixel geometrical relationship could be destroyed.

In [17], a multi-valued watershed algorithm is proposed. Its main concept is that, if a point is chosen for classification, it is the neighborhood of a marker, and the similarity between them is the highest at that time than any other pair of points and neighborhood markers. In addition, a hierarchical queue data structure is used to accomplish this task.

A hierarchical queue is a set of queues with different priorities and each queue is a first-in-first-out data structure. At any time, the element popped from the hierarchical queue is the one from the queue with the highest priority, and, if the queue with higher priority is empty, the element in the first non-empty of lower priority is pulled out.

While using the hierarchical queue for multivalued watershed, two steps are involved: initialization and flooding. In initialization, the neighborhood units of all "in" and "out" markers are pushed into the hierarchical queue according their similarity with the corresponding markers. In the flooding procedure, while the hierarchical queue is not empty, extract one unit and check if the unit has been classified to any marker. If not yet, the similarity between this unit and its neighboring markers is calculated and this unit is classified to the most similar marker. Then, all of the neighborhood of this unit are pushed into the hierarchical queue based on their similarity to the marker. The flooding process continues until the hierarchical queue is empty; that is, all the units in the searching area are already classified.

In the "classification" stage of the proposed system, two kinds of classification methods based on hierarchical queue data structure are introduced. Pixel-based classification mainly comes from the morphological watershed algorithm proposed in [17], except the modification of similarity measurement function and the flooding process. In region-based classification, the basic unit is no longer the image pixel but the regions created from the watershed transformation.

6. Tracking

There are several approaches to extend image segmentation to video segmentation. One of the most popular methodologies is the combination of spatial segmentation and temporal tracking [17][18][19][40]. Given an image sequence, the I-frame(Intra-frame) is segmented to obtain the object of interest, and the technique of I-frame segmentation can be either supervised or automatic, either spatial information based or temporal information based and, of course, spatio-temporal information based. In the following P-frames(Inter-frames), instead of re-segmenting the image, "tracking" the object based on the temporal information is introduced. Once if the temporal information needs to be compute, the motion information must be extracted by a motion estimator. Typically, a video object segmentation system consists of the stages, as shown in Fig. 5.

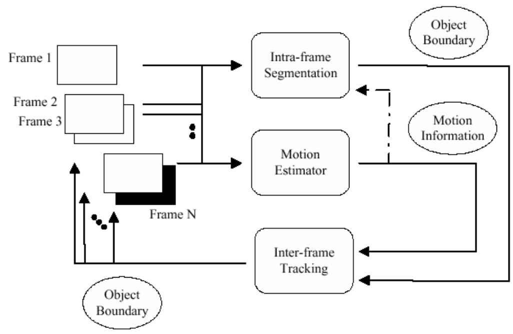

Figure 5: Flow diagram of a typical video object segmentation system

In [17], the P-frame tracking consists of four steps: motion prediction, motion estimation, boundary warping, and boundary adjustment. In "Boundary Adjustment", its main concept is that, given an approximation of real object boundary from "Boundary Warping", the method of I-frame segmentation can be employed to accomplish the boundary adjustment for non-rigid body boundary refinement.

A simplified tracking scheme (Fig. 6) is proposed based on the idea of boundary adjustment and the ability of adaptive morphological operators.

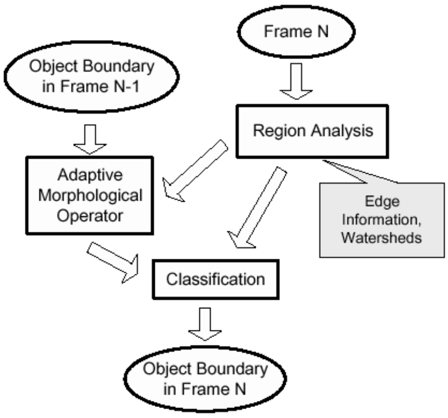

Figure 6: Flow diagram of proposed object tracking

The proposed algorithm is almost the same as the intra-frame segmentation, except the replacement of user interaction with previous segmentation result. Furthermore, the motion estimator is absent in this tracking scheme. The main idea is that, as long as the initial boundary is not too far away from the real object boundary, the adaptive morphological operators are able to properly find out the searching area where the object boundary resides. Hence, under the constraint that "the motion of the object is not too large for adaptive morphological operators to provide the correct searching area", the object boundary will eventually be found in the "classification" stage.

四、 完成之工作項目及具體成果：

- 完成工作項目：

（1） 完成一以數學型態學為基礎之物件分割系統，其中包括多尺度斜率計算模組，分水嶺轉換模組及適應性型態學運算子模組等三項新開發之單元模組。

（2） 上述分割系統可將數位影像及 中低移動速度’之視訊信號中經使用者協助所選定之物件分割完成。

（3） 在 P-Ⅲ750MHz CPU 之個人電腦上，每張彩色 256× 256size 之 Image 可在 0.25 秒內完成分割而相同 size 之數位視訊則可達到 2.5 frames/sec 速率。

- 對學術研究及其他應用之貢獻：

物件是 MPEG-4 標準中之“處理單元”，然而標準中對物件分割部分並未詳加界定（MPEG 只對解碼端加以標準化），本計畫之執行，對 MPEG-4 標準之編碼端相關技術進行研討。對國內多媒體及內容產業將有直接的貢獻。

- 發表論文

Wayne Lin, Chi-Hui Huang and Ja-Ling Wu, "A Novel Supervised Image/Video Segmentation System Based on Mathematical Morphology", Proceedings of 2001 IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing , Baltimore, USA.

五、 參考文獻

[1] ISO/IEC JTCI/SC29/WG11 N3056. MPEG-4 FDAM1, December 1999.

[2] Robert M. Haralick and Linda G. Shapiro. "Survey: Image Segmentation Techniques". Computer Vision, Graphics and Image Processing, 29:100{132, 1985.

[3] Nikhil R. Pal and Sankar K. Pal. "A Review on Image Segmentation Techniques". Pattern Recognition, 26(9):1277{1294, 1993.

[4] L. Lucchese and S. K. Mitra. "Unsupervised Sgementation of Color Image Based on k-means Clustering in Chromaticity Plane". IEEE. Proceedings, pages 74{78, 1999.

[5] M. Celenk. "Colour Image Segmentation by Clustering". IEE. Proceedings-E, 138(5):368{376, September 1991.

[6] Jens-Rainer Ohm and Phuong Ma. "Feature-Based Cluster Segmentation of Image Sequences". IEEE. Proceedings, pages 178{181, 1997.

[7] Dorin Comaniciu and Peter Meer. "Robust Analysis of Feature Spaces: Color image Segmentation". IEEE, pages 1{2, 1997.

[8] Jongwon Kim Ju Guo and C.-C. Jay Kuo. "Fast Video Object Segmentation Using A_ne Motion and Gradient-Based Color Clustering". IEEE, pages 750{755, 1999.

[9] A. Murat Tekalp Eli Saber and Gozde Bozdagi. "Fusion of Color and Edge Information for Improved Segmentation and Edge Linking". IEEE.. Proceedings, pages 2176{2179, 1996.

[10] Marie-Pierre Dubuisson and Anil K. Jain. "Object Contour Extraction Using Color and Motion". IEEE. Proceedings, pages 471{476, 1993.

[11] Andrew Witkin Michael Kass and Demetri Terzopoilos. "Snakes: Active Contour". International Journal of Computer Vision, pages 321{331, 1988.

[12] Touradj Ebrahimi Roberto Castagno and Murat Kunt. "Video Segmentation Based on Multiple Feature for Interactive Multimedia Applications". IEEE. Trans. on Circuits and System for Video Technology, 8(5):562{571, September 1998.

[13] Luc Vincent and Pierre Soille. "Watersheds in Digital Spaces: An Efficient Algorithm Based on Immersion Simulations". IEEE. Trans. on Pattern Analysis and Machine Intelligence, 13(6):583{598, June 1991.

[14] John M. Gauch. "Image Segmentation and Analysis via Multiscale Gradient Watershed Hierarchies". IEEE. Trans. on Image Processing, 8(1):69{79, January 1999.

[15] Serafim N. Nicos Maglaveras Kostas Haris and Aggelos K. Katsaggelos. "Hybrid Image Segmentation Using Watersheds and Fast Region Merging". IEEE. Trans. on Image Processing, 7(12):1684{1699, December 1998.

[16] Il Dong Yun Sang Ho Park and Sang Uk Lee. "Color Image Segmentation Based on 3D Clustering: Morphological Approach". Pattern Recognition, 31(8):1061{1076, 1998.

[17] Chuang Gu and Ming-Chieh Lee. "Semantic Video Object Segmentation and Tracking Using Mathematical Morphology and Perspective Motion Model". IEEE. Proceeding, pages 514{517, 1997.

[18] Chuang Gu and Ming-Chieh Lee. "Semiautomatic Segmentation and Tracking of Semantic Video Objects". IEEE. Trans. on Image Processing, 8(5):572{584, September 1998.

[19] Demin Wang. "Unsupervised Video Segmentation Based on Watersheds and Temporal Tracking". IEEE. Trans. on Circuits and System for Video Technology, 8(5):539{546, September 1998.

[20] T. Meier and K. N. Ngan. "Segmentation and Tracking of Moving Objects for Content-Based Video Coding". IEE. Proc.-Vis. Image Signal Processing, 146(8):144{150, June 1999.

[21] C. S. Chen. "Range Image Processing Using Mathematical Morphology and Its Application on 3D Object Recognition".

PhD thesis, National Taiwan University, 1996.

[22] Daehee Kim Munchurl Kim, Jae Gark Choi, Hyung Lee, Myoung Ho Lee, Chieteuk Ahn, and Yo-Sung Ho. "A VOP Generation Tool: Automatic Segmentation of Moving Object in Image Sequence Based on Spatial-Temporal Information". IEEE. Trans. on Circuits and System for Video Technology, 9(8):1216{1226, December 1999.