



Bài 4

THU THẬP DỮ LIỆU GIS

Ths Trần Mạnh Trường
truongtm@gmail.com

Hà Nội – 2020





Các chủ đề nghiên cứu

1

Giới thiệu về GIS

2

Biểu diễn dữ liệu địa lý

3

Định vị và tham chiếu không gian

4

Thu thập dữ liệu không gian

5

Phân tích dữ liệu không gian

6

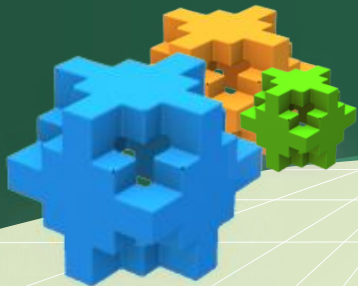
Hiển thị dữ liệu địa lý

7

Các chủ đề GIS nâng cao

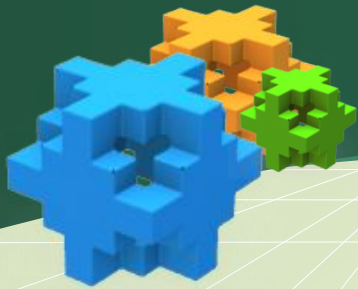
8

9



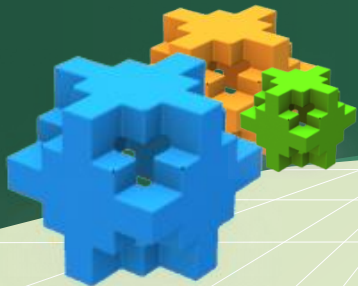
Nội dung

- ❖ Giới thiệu
- ❖ Thu thập dữ liệu địa lý trực tiếp
- ❖ Thu thập dữ liệu địa lý gián tiếp
- ❖ Chuyển đổi dữ liệu địa lý
- ❖ Thu thập thành phần dữ liệu thuộc tính
- ❖ Tổng kết bài



1. Giới thiệu

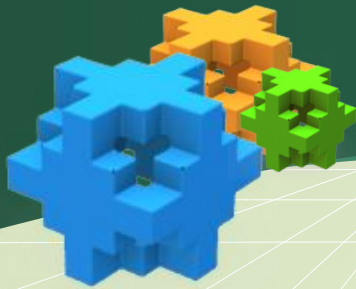
- ❖ Thu thập dữ liệu là nhiệm vụ tốn kém thời gian nhất, tốn kém kinh phí nhất, nhưng là công việc quan trọng nhất trong quá trình phát triển hệ thống GIS.
- ❖ Có nhiều nguồn dữ liệu khác nhau và nhiều phương pháp thu thập dữ liệu khác nhau.
- ❖ Hai phương pháp chính để thu thập dữ liệu raster và véctor, bao gồm:
 - Thu thập (đo trực tiếp – *primary*): Dữ liệu thu thập trực tiếp từ môi trường gọi là dữ liệu sơ cấp (primary data)
 - Chuyển đổi dữ liệu (thu thập gián tiếp từ các nguồn dữ liệu khác nhau – *secondary*): Dữ liệu thứ cấp (secondary data) là dữ liệu thu thập không trực tiếp từ môi trường.



Giới thiệu

❖ Phân loại thu thập dữ liệu địa lý

	Raster	Vector
Primary	<ul style="list-style-type: none">• Digital satellite remote-sensing images• Digital aerial photographs	<ul style="list-style-type: none">• GPS measurements• Survey measurements
Secondary	<ul style="list-style-type: none">• Scanned maps or photographs• Digital elevation models from topographic map contours	<ul style="list-style-type: none">• Topographic maps• Toponymy (placename) databases

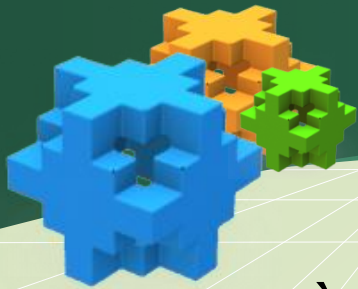


Giới thiệu

- ❖ Ví dụ về phân tích giá đầu tư cho hai hệ thống GIS theo mô hình client-server (đơn vị tính \$1000)

	10 Seats		100 Seats	
	\$	%	\$	%
Hardware	30	3,4	250	8,6
Software	25	2,8	200	6,9
Data	400	44,7	450	15,5
Staff	440	49,1	2000	69,0
Total	895	100,0	2900	100,0

- Ghi chú: *Hardware bao gồm máy khách và máy chủ (chưa tính đến mạng), Staff làm việc 100% thời gian*
- ❖ Kinh phí dành cho thu thập dữ liệu chiếm đến 85% tổng kinh phí của dự án GIS (Hardware+ Software+Staff).



Giới thiệu

❖ Luồng công việc thu thập dữ liệu địa lý: 6 công đoạn

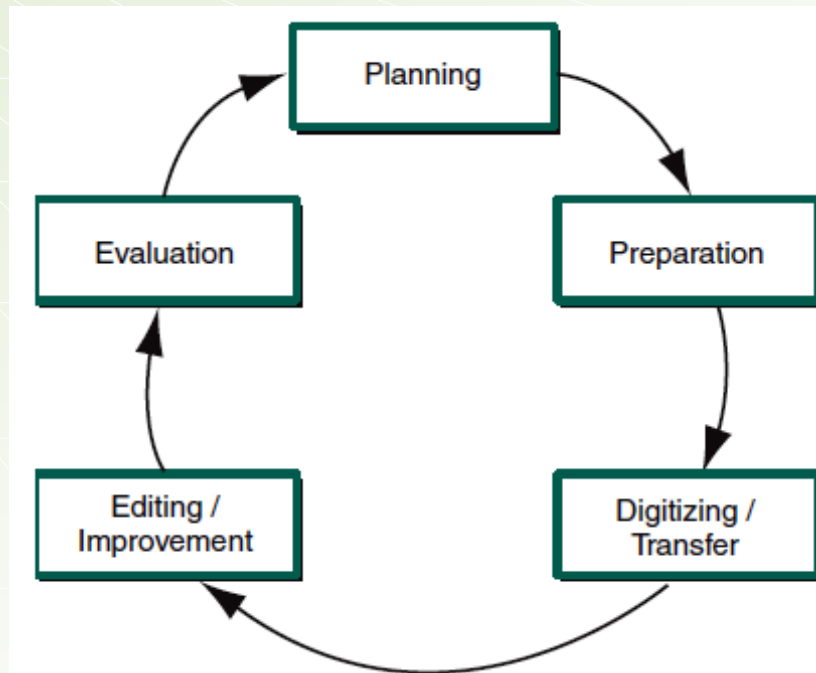
■ Planning:

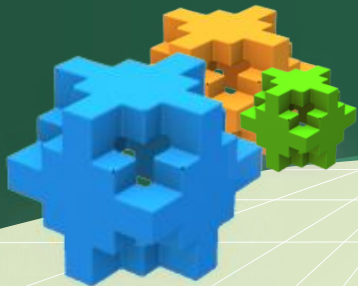
- Thu thập yêu cầu người sử dụng,
- Thu thập tài nguyên (*staff, hardware, và software*)
- Xây dựng kế hoạch dự án

■ Preparation:

- Thu thập dữ liệu
- Phác họa lại các bản đồ nguồn kém chất lượng
- Chỉnh sửa các ảnh quét bản đồ, khử nhiễu
- Lắp đặt hệ thống phần cứng/phần mềm để chuẩn bị nhập liệu.

■ ...





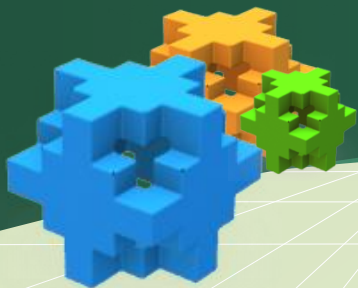
Giới thiệu

- ❖ Luồng công việc thu thập dữ liệu địa lý: 6 công đoạn (tt)
 - Digitizing/transfer:
 - Công đoạn này mất nhiều công sức nhất
 - Không chỉ có số hóa mà còn có nhiều nhiệm vụ khác (sẽ đề cập sau)
 - Editing/improvement:
 - Phê chuẩn (validate) dữ liệu
 - Sửa lỗi nếu có
 - Nâng cao chất lượng dữ liệu.
 - Evaluation:
 - Tiến trình nhận biết dự án thành công hay thất bại.
- ❖ Đối với các dự án lớn, luồng công việc được thực hiện lặp để có được kết quả vòng lặp sau tốt hơn vòng lặp trước cho đến khi người dung thỏa mãn.



2. Thu thập dữ liệu primary

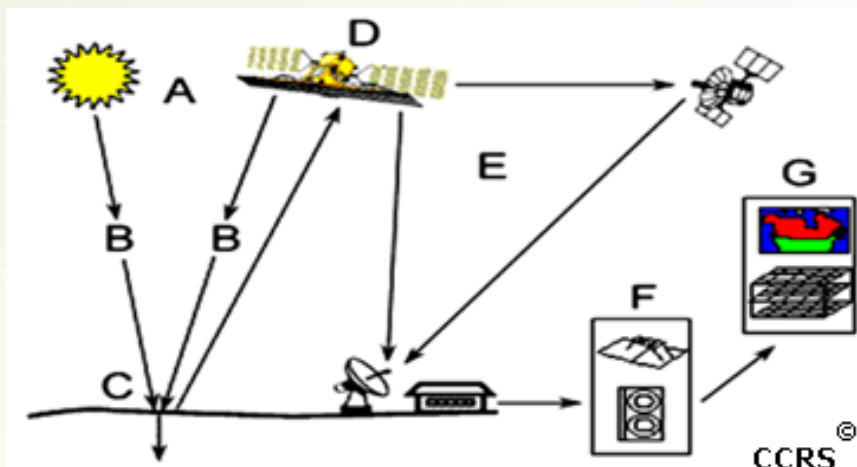
- ❖ Thực hiện thu thập trực tiếp các đối tượng
- ❖ Dữ liệu nhập trực tiếp vào CSDL hay vào tệp tạm thời
- ❖ Các giải pháp nghiên cứu đề xuất nhằm giảm thời gian nhập liệu và giảm tỷ lệ lỗi xảy ra.
- ❖ Hai loại dữ liệu thu thập trực tiếp được nghiên cứu ở đây:
 - Dữ liệu raster
 - Dữ liệu véctor

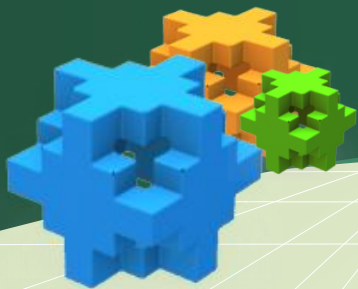


Thu thập dữ liệu raster

- ❖ Nguồn dữ liệu chính là ảnh vệ tinh, ảnh hàng không
- ❖ Viễn thám (*Remote Sensing*) được xem như một khoa học thu nhận thông tin về đối tượng, khu vực hay hiện tượng trên bề mặt Trái đất mà không tiếp xúc trực tiếp với chúng.
- ❖ Quy trình viễn thám bao gồm:

A - Energy Source or Illumination
B - Radiation and the Atmosphere
C - Interaction with the Target
D - Recording of Energy by the Sensor
E - Transmission, Reception, and Processing
F - Interpretation and Analysis
G - Application





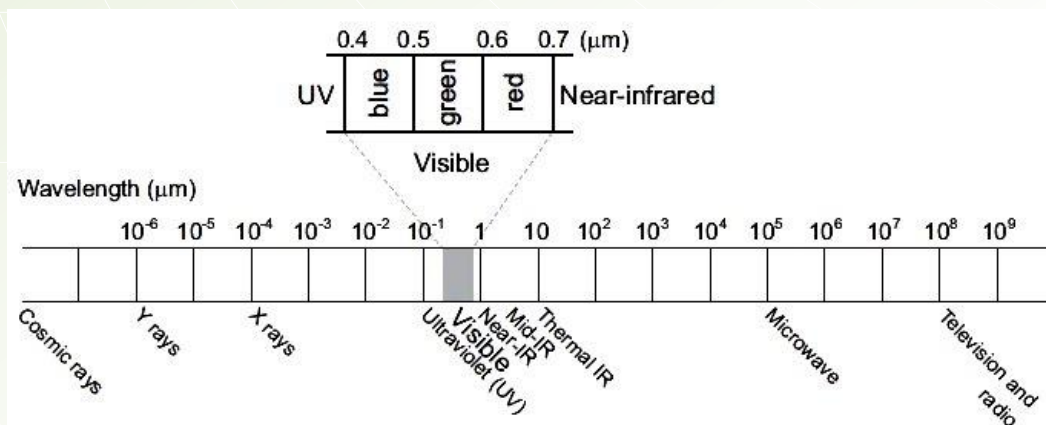
Qui trình viễn thám

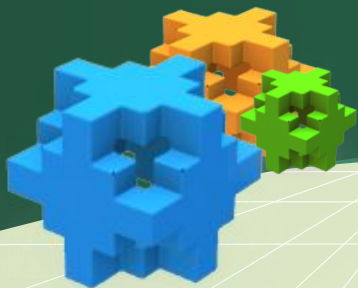
❖ Bức xạ điện từ

- Nguồn năng lượng trong viễn thám dưới dạng bức xạ điện từ (*electromagnetic radiation*)
 - Nó bao gồm trường điện và trường từ
 - Cả hai lan truyền trong không gian với tốc độ ánh sáng (C) xấp xỉ 299.799.000 m/s
- Vệ tinh có sensor đo bức xạ ánh sáng mặt trời, Trái đất và chính nguồn sáng của nó.

❖ Phổ điện từ

- Nhiều vùng được sử dụng trong viễn thám
 - Đá, khoáng trên mặt đất phát sáng khi chiếu tia cực tím.

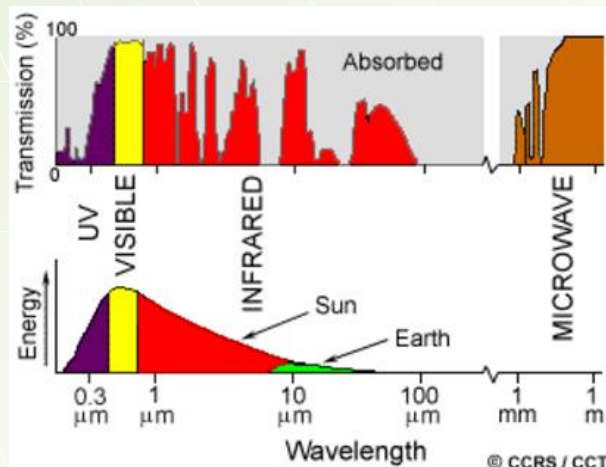




Qui trình viễn thám

❖ Tương tác giữa bức xạ và khí quyển

- Các hạt và khí (gas) trong khí quyển ảnh hưởng đến ánh sáng và bức xạ đi qua nó
- Sự ảnh hưởng của khí quyển trên bức xạ theo cơ chế tán xạ (*scattering*) hay hấp thụ (*absorption*)
- Những vùng phổ điện từ có bước sóng ngoài băng tần hấp thụ của các thành phần khí quyển được sử dụng trong viễn thám.
- Cửa sổ khí quyển:

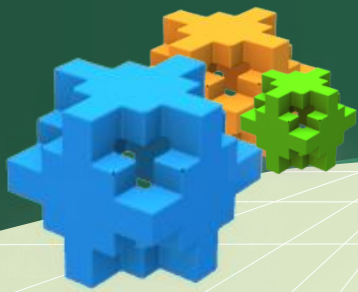




Quy trình viễn thám

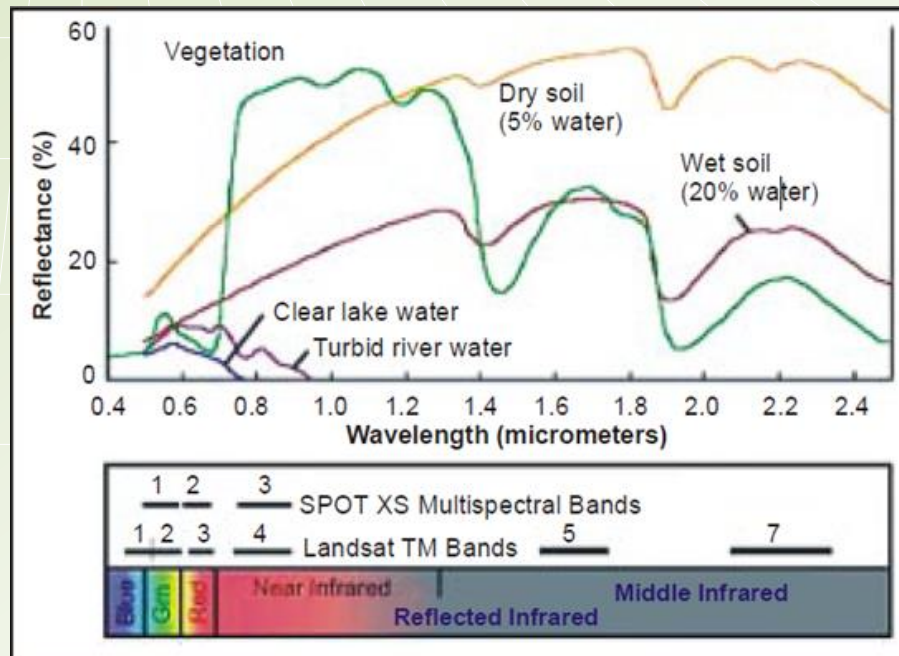
❖ Tương tác giữa bức xạ và đối tượng đích

- Bức xạ không bị hấp thụ hay tán xạ trong không khí sẽ tương tác với bề mặt Trái đất.
- Toàn bộ năng lượng sẽ tương tác với bề mặt theo một trong ba cách thức như sau:
 - Hấp thụ (*A-Absorption*),
 - Truyền qua (*T-Transmission*)
 - Phản xạ (*R-Reflection*)
- Có thể phân tích, nhận dạng đối tượng trên bề mặt Trái đất thông qua đáp ứng phổ của chúng



Đáp ứng phổ

- ❖ Ví dụ đồ thị đường cong phổ của thực vật, đất và nước.
 - Hình dạng đường cong phổ phụ thuộc vào:
 - Dải sóng năng lượng được thiết bị viễn thám ghi lại
 - Tính chất của đối tượng đích trên mặt Trái đất.

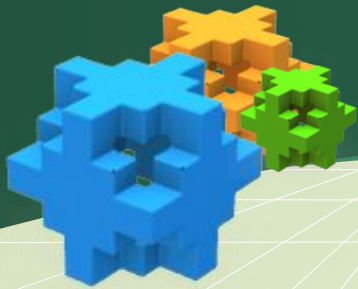




Ảnh viễn thám

- ❖ Độ phân giải là đặc trưng chính của ảnh viễn thám:
 - Phân giải không gian
 - Phân giải phổ
 - Phân giải thời gian.
- ❖ Một đặc trưng quan trọng khác là khả năng thu thập ảnh stereo để xây dựng mô hình số 3D.





Đặc trưng của ảnh viễn thám

❖ Phân giải không gian:

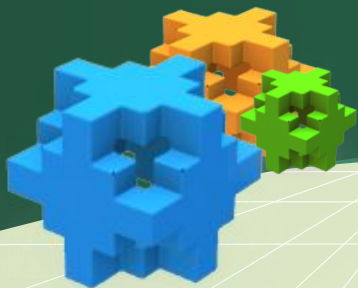
- Phân giải không gian là đề cập đến vùng được đo, kích thước đặc trưng nhỏ nhất trên mặt đất mà sensor có thể nhận biết.
- Ảnh vệ tinh: 0,5m – 1km,
- Không ảnh: 0,1m-5m;
- Kích thước ảnh từ 900x900 đến 3000x3000 pixels,
- Diện tích chụp: 9x9 đến 200x200Km

❖ Phân giải phổ:

- Là đề cập đến phổ sóng điện từ mà *sensor* có thể đo.
- Các loại: Ảnh đơn phổ; ảnh đa phổ;

❖ Phân giải thời gian:

- Mô tả tần số mà ảnh được thu thập cho cùng một vùng.



Ví dụ ảnh vệ tinh SPOT

❖ Một số tính chất của ảnh thu từ vệ tinh SPOT

Vệ tinh	Kênh	Giải phổ	Độ phủ	Phân giải
Spot 2	P: Panchromatic	0.50 - 0.73 μm	60 × 60 km	10 meter
	B1: Green	0.50 - 0.59 μm		20 meter
	B2: Red	0.61 - 0.68 μm		20 meter
	B3: Near Infrared	0.78 - 0.89 μm		20 meter
Spot 4	M: monospectral	0.61 - 0.68 μm		10 meter
	B1: Green	0.50 - 0.59 μm		20 meter
	B2: Red	0.61 - 0.68 μm		20 meter
	B3: Near Infrared	0.78 - 0.89 μm		20 meter
	B4: Short-wave Infrared	1.58 - 1.75 μm		20 meter
Spot 5	P: Panchromatic	0.48 - 0.71 μm	60 × 60 km	2.5 or 5 meter
	B1: Green	0.50 - 0.59 μm		10 meter
	B2: Red	0.61 - 0.68 μm		10 meter
	B3: Near Infrared	0.78 - 0.89 μm		10 meter
	B4: Short-wave Infrared	1.58 - 1.75 μm		20 meter



Thu thập dữ liệu vector

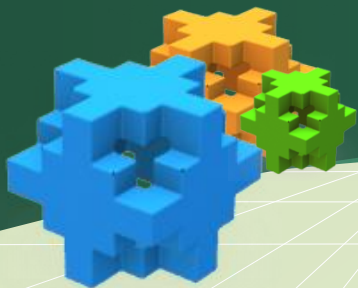
- ❖ Hai phương pháp được sử dụng: đo đạc mặt đất và GPS
- ❖ Đo đạc mặt đất
 - Nguyên lý: Vị trí 3-D của bất kỳ điểm nào cũng có thể được xác định bởi đo góc và khoảng cách đến các điểm biết trước.
 - Đo đạc bắt đầu từ điểm chuẩn (*benchmark*)
- ❖ Sử dụng máy toàn đạc (*Total stations*) để đo góc và đo khoảng cách với độ chính xác rất cao.
 - Có khả năng tự động hình thành các đối tượng điểm, đường và vùng vector.





Thu thập dữ liệu vector

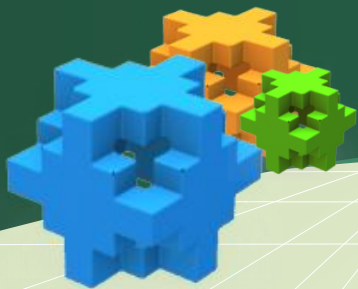
- ❖ Nguyên lý đo đạc mặt đất rất ít thay đổi trong vòng 100 năm nay, nhưng công nghệ đã thay đổi nhanh, làm gia tăng độ chính xác và giảm thời gian đo.
- ❖ Nhận xét:
 - Đo đạc mặt đất là công việc rất tốn kém thời gian, và giá thành cao
 - Tuy nhiên, đây là giải pháp tốt nhất để có giá trị đo chính xác
- ❖ Ví dụ ứng dụng:
 - Đo đạc mặt đất được sử dụng thu thập số liệu đo chính xác các ranh giới của ngôi nhà, mảnh đất...
 - Giá trị đo đạc mặt đất còn được sử dụng làm các điểm tham chiếu địa lý của ảnh viễn thám hay ảnh hàng không.



3. Thu thập dữ liệu thứ cấp

- ❖ Thu thập gián tiếp dữ liệu địa lý là tiến trình tạo ra các tệp, CSDL raster, vector từ bản đồ, ảnh chụp hay các tài liệu in khác.
- ❖ Thu thập dữ liệu raster từ máy quét
 - Sinh ra các pixel đến 3 bytes màu R,G,B
 - Độ phân giải: 200dpi (8 dot/mm) đến 2.400dpi (96 dot/mm)
 - Thông thường GIS sử dụng: 400-900dpi
 - Thời gian quét từ: 30s đến 30 min/map
 - Bản đồ và các tài liệu quét được sử dụng nhiều trong GIS để làm bản đồ nền và kho dữ liệu.





Bản số hóa

❖ Thu thập dữ liệu vector (bằng máy Digitizer)

- Là tiến trình số hóa (digitizing) các đối tượng vector từ bản đồ hay các nguồn dữ liệu địa lý khác
- Các phương pháp hay sử dụng
 - Số hóa thủ công
 - Số hóa heads-up
 - Vector hóa...
- Số hóa thủ công
 - Sử dụng Digitizer, độ chính xác từ 0,01mm đến 0,25mm
 - Kích thước bàn số hóa từ 30x60cm đến 112x152cm
 - Đây là giải pháp đơn giản nhất, rẻ nhất, dễ nhất để thu thập dữ liệu vector từ bản đồ
 - Hệ tọa độ sử dụng: Hệ tọa độ tương đối (thiết bị)





Thu thập dữ liệu secondary

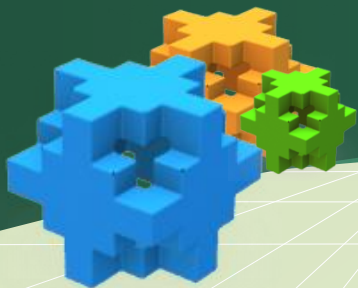
❖ Thu thập dữ liệu vector (tt)

■ Heads-up digitizing và vector hóa



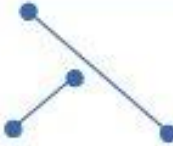



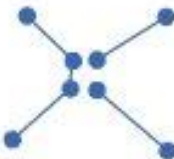





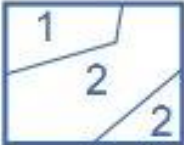
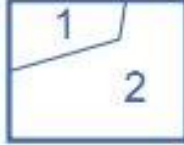

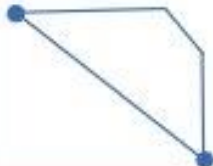
- Nhiệm vụ là chuyển đổi dữ liệu bản đồ raster sang dữ liệu vector. Hay nói cách khác là thực hiện vector hóa.
- Hai giải pháp
 - Heads-up digitizing: Sử dụng chuột để thu thập các điểm bản đồ trên màn hình
 - Vector hóa tự động
- Ví dụ ứng dụng: thu thập dữ liệu về mảnh đất, ngôi nhà, tiện ích...

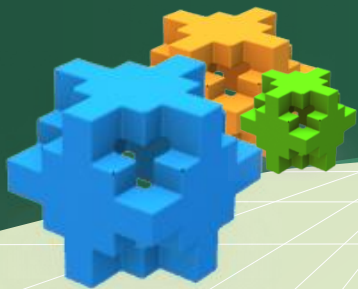
■ Vector hóa tự động

- Sử dụng phần mềm để sinh các đối tượng vector từ các giá trị pixel; sau đó tạo lập topo.
- Nhiều lỗi có thể xảy ra, dẫn tới phải chỉnh sửa sau vector hóa



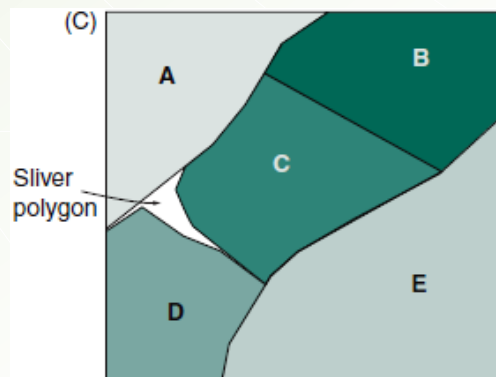
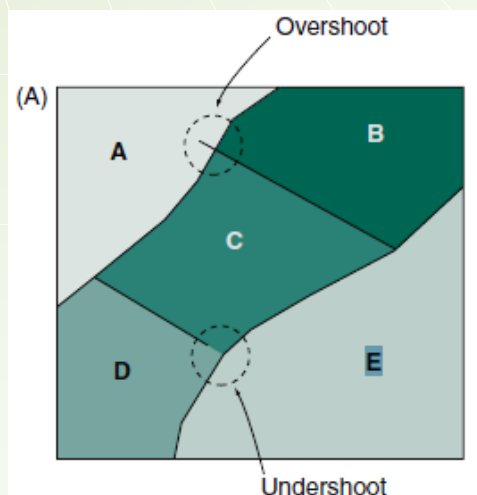
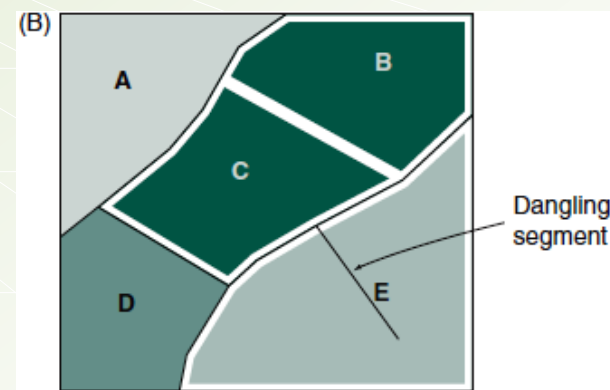
Lỗi thu thập dữ liệu gián tiếp

Before cleanup	After cleanup	Description	Before cleanup	After cleanup	Description
		Erase duplicates or sliver lines			Extend undershoots
		Erase short objects			Snap clustered nodes
		Break crossing objects			Erase dangling objects or overshoots
		Dissolve polygons			Dissolve nodes into vertices



Lỗi thu thập dữ liệu gián tiếp

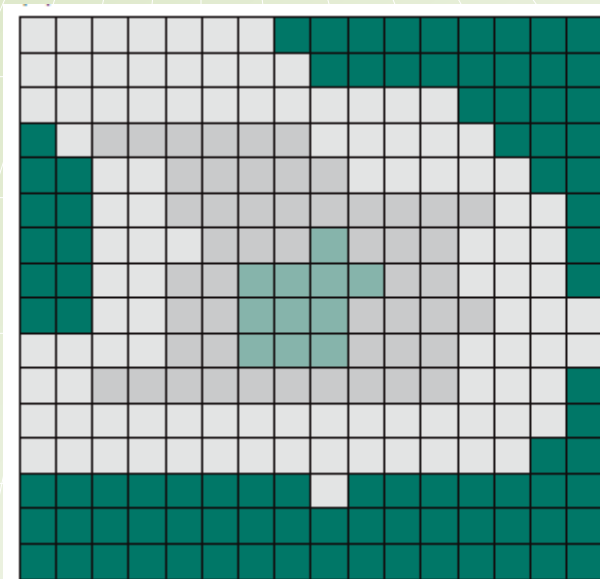
- ❖ Công việc số hóa là rất nặng nhọc và dễ gây ra lỗi
- ❖ Phần mềm GIS cần có chức năng hỗ trợ
 - Tìm ra nơi có lỗi
 - Tự động chỉnh sửa lỗi
 - Chỉnh sửa lỗi thủ công



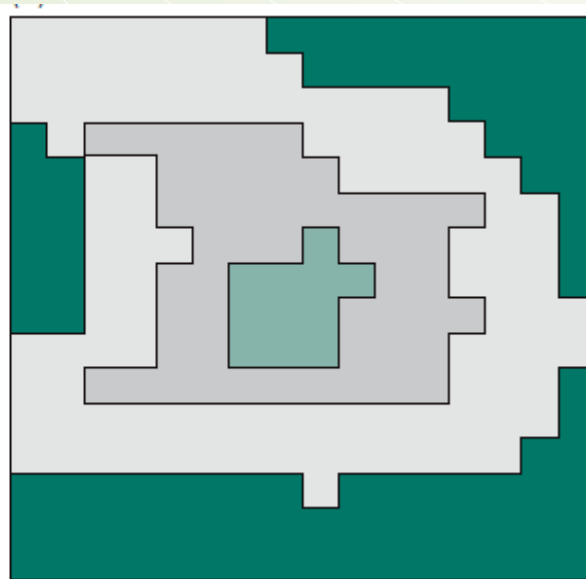


Véc tơ hóa tự động

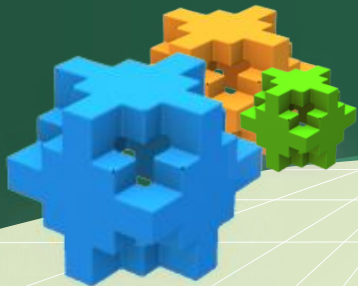
- ❖ Tụ hợp các tế bào raster có cùng thuộc tính
- ❖ Dò biên (tìm đường bao) các đa giác để tạo lập các đường gấp khúc vector.



Bản đồ gốc

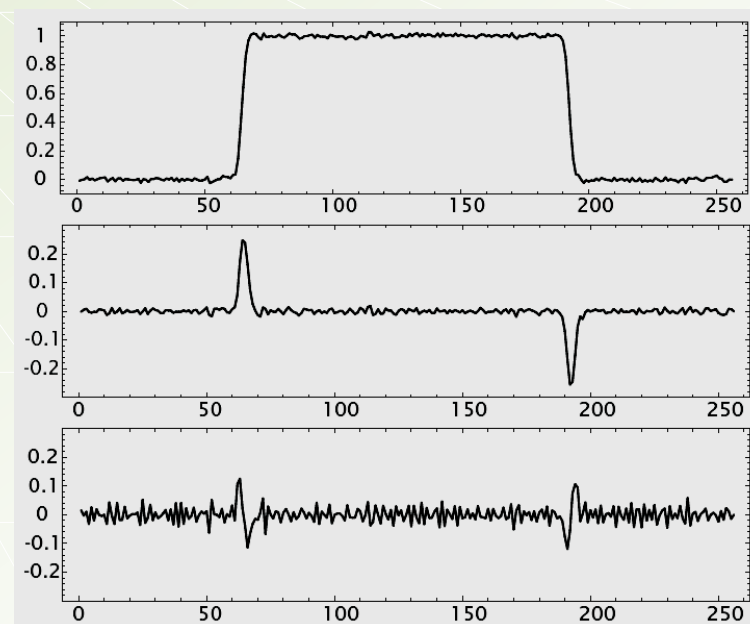
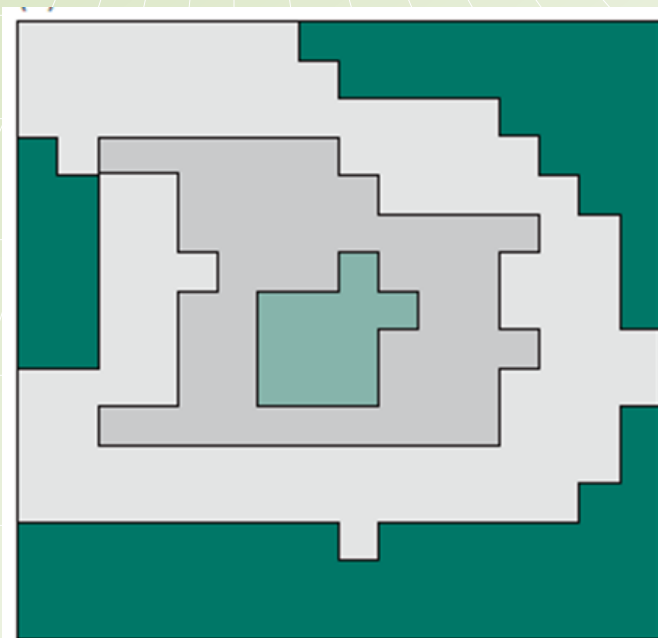


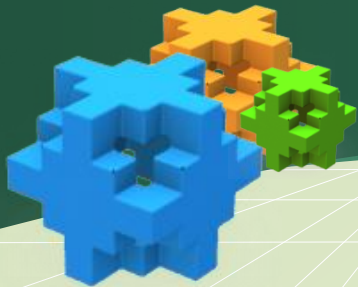
Đa giác sau vector hóa



Véc tơ hóa tự động

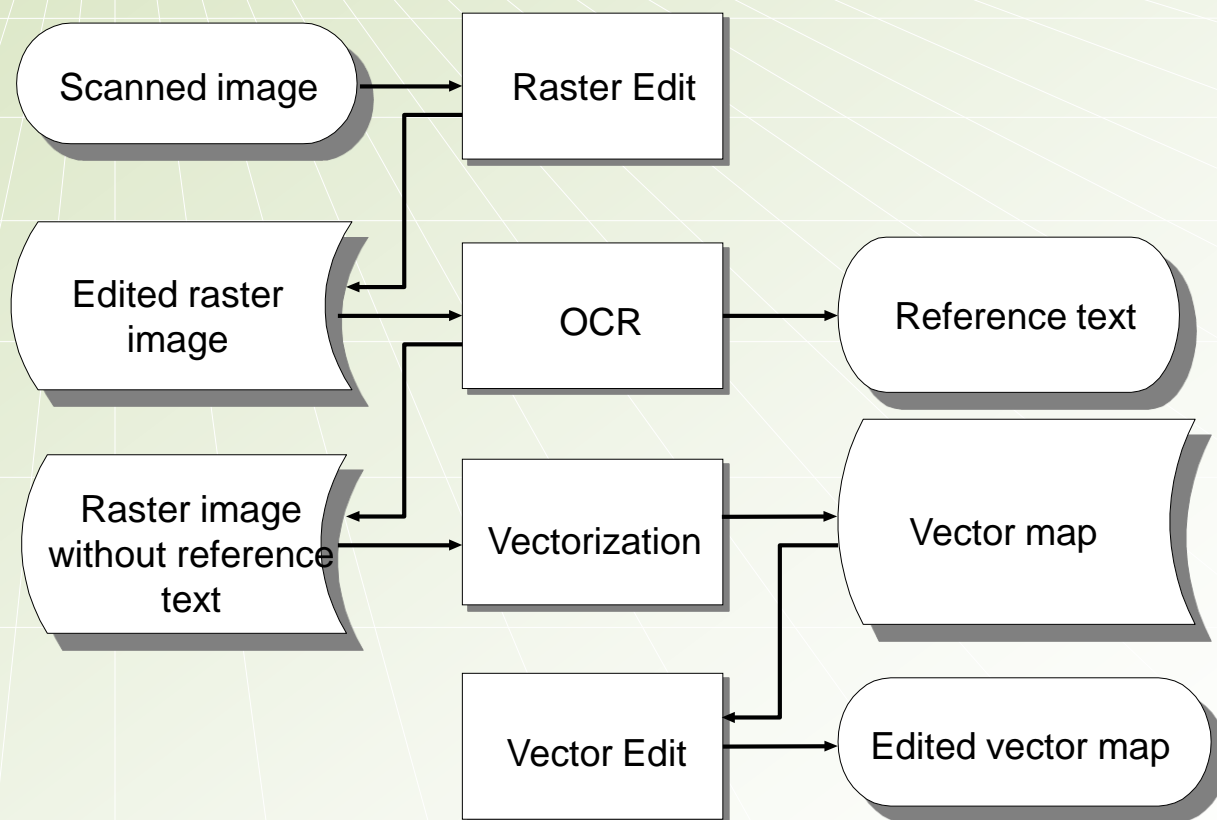
- ❖ Dò biên vùng bằng đạo hàm bậc nhất hoặc đạo hàm bậc hai





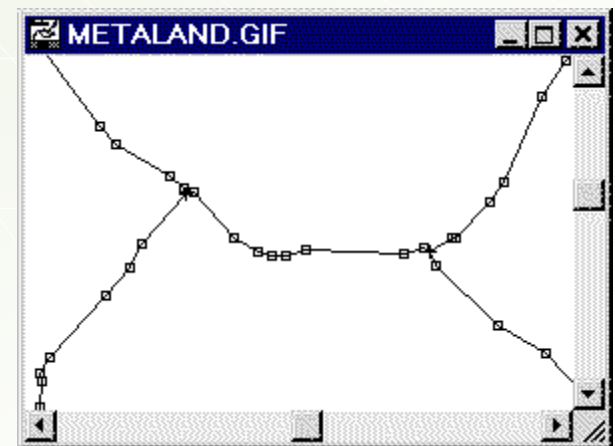
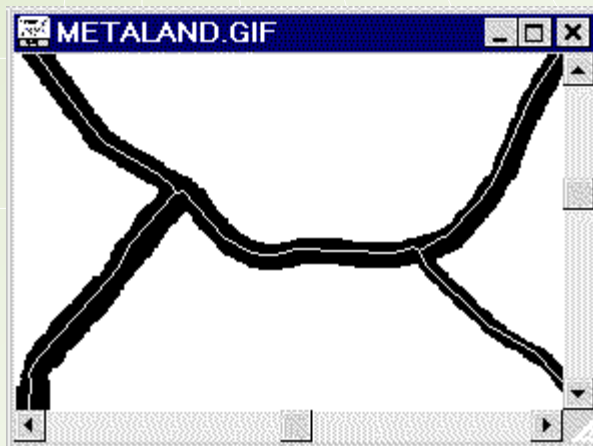
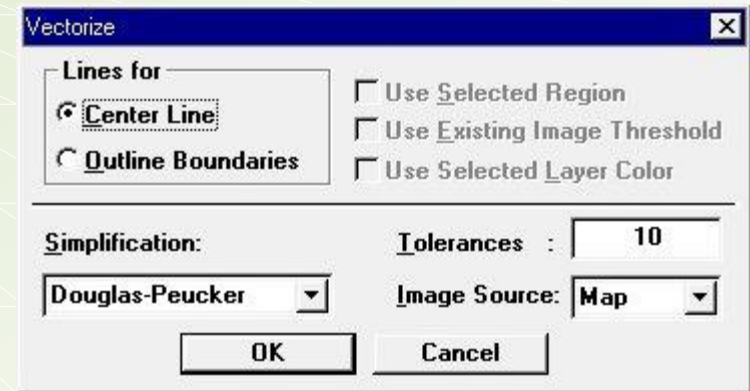
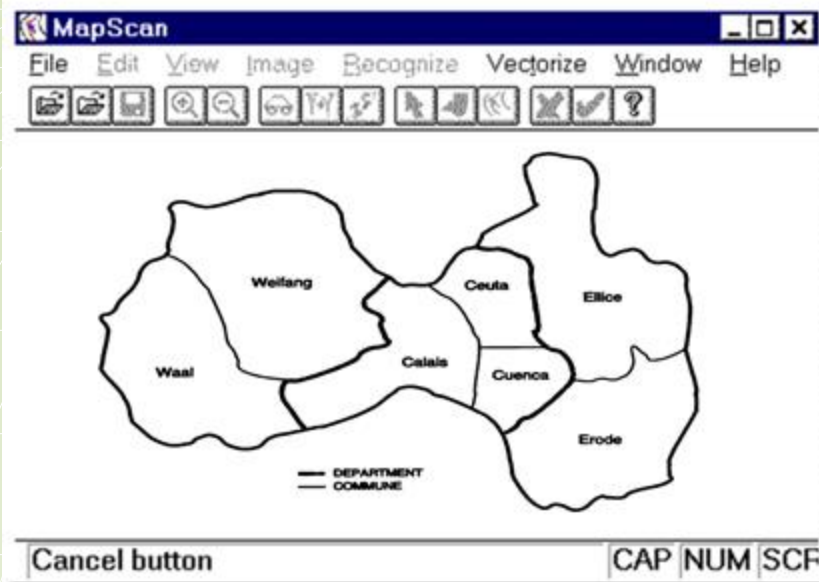
Ví dụ véc tơ hóa tự động

❖ Phần mềm MapScan





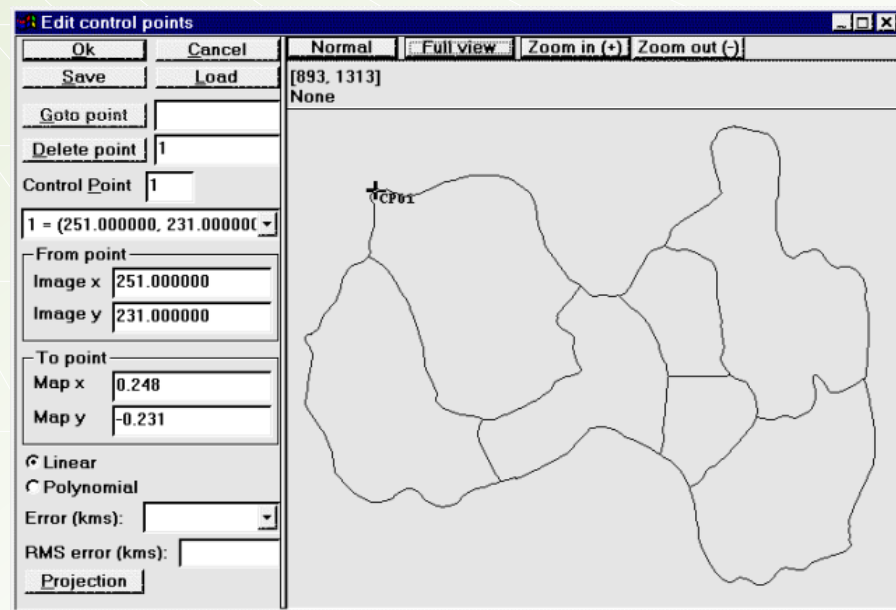
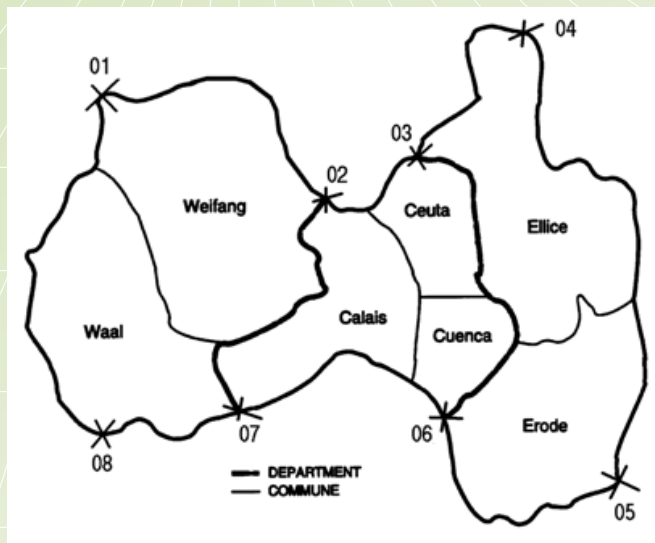
Ví dụ phần mềm MapScan

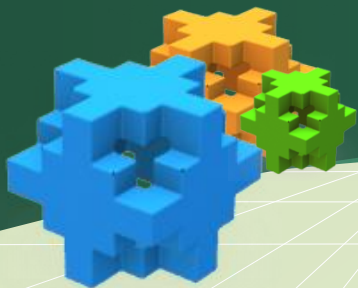




Ví dụ phần mềm MapScan

❖ Tham chiếu địa lý

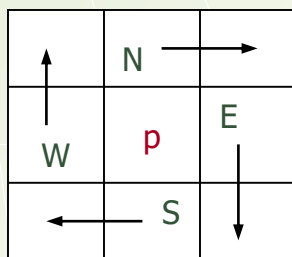




Thuật toán làm mảnh

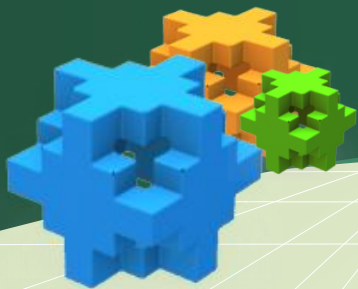
❖ Thuật toán Zhang-Suen

- Giả sử có ảnh nhị phân kích thước $m \times n$: 0 – trắng, 1-đen và nhiệm vụ là làm mảnh màu đen, p – điểm trên ảnh
- Gọi số nguyên $N(p)$ là tổng giá trị của 8 pixels láng giềng p
- p_N, p_S, p_E, p_W là các giá trị pixel trên, dưới, phải, trái của p
- $T(p)$ là số đếm lần biến đổi từ 0 sang 1 của các pixel láng giềng theo chiều kim đồng hồ
 - Ví dụ: $N(p) = 5, p_N = p_W = 1, p_S = p_E = 0$ và $T(p) = 2$.



1	1	1
1	p	0
0	0	1

(Worboys M.F., *GIS: A Computing Perspective*, Taylor&Francis, 1997)



Thuật toán làm mảnh

❖ Thuật toán Zhang-Suen (*tt*)

- Input: Ảnh raster nhị phân kích thước $m \times n$ (0- trắng, 1-đen)
- Output: Ảnh raster nhị phân đã làm mảnh kích thước $m \times n$
- Procedure:

Bước 1. Với mỗi điểm ảnh p hãy thực hiện:

Nếu $2 \leq N(p) \leq 6$ và $T(p)=1$ và $p_N.p_S.p_E = 0$ và $p_W.p_E.p_S = 0$ thì đánh dấu p

Bước 2. Với mỗi điểm p của ảnh hãy thực hiện:

Nếu có điểm nào bị đánh dấu thì gán giá trị 0 cho chúng, nếu không thì dừng.

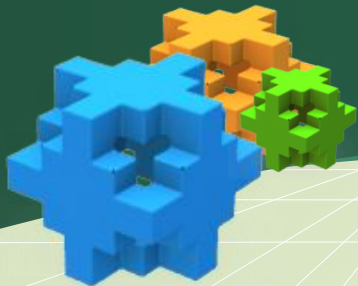
Bước 3. Với mỗi điểm ảnh p hãy thực hiện:

Nếu $2 \leq N(p) \leq 6$ và $T(p)=1$ và $p_N.p_S.p_W = 0$ và $p_W.p_E.p_N = 0$ thì đánh dấu p

Bước 4. Với mỗi điểm p của ảnh hãy thực hiện:

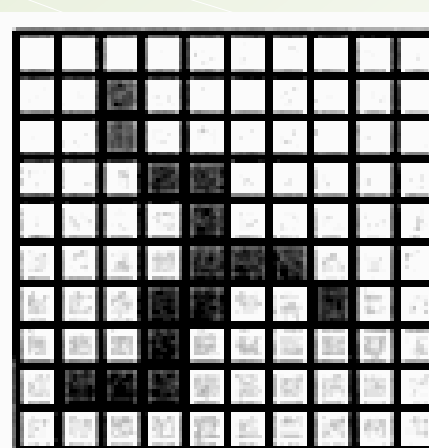
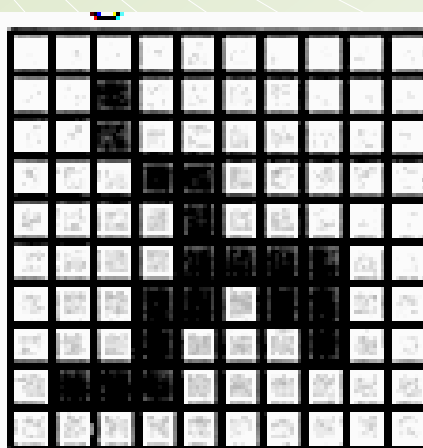
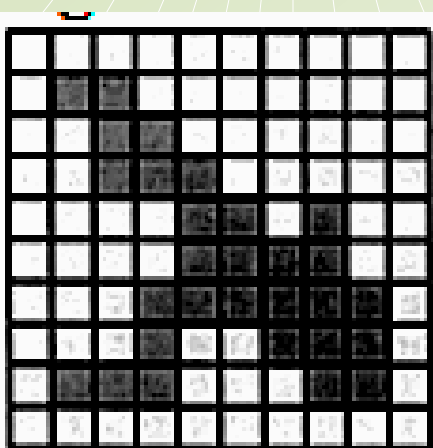
Nếu có điểm nào bị đánh dấu thì gán giá trị 0 cho chúng, nếu không thì dừng.

Bước 5. Thực hiện bước 1.



Thuật toán làm mảnh

❖ Ví dụ thuật toán Zhang-Suen





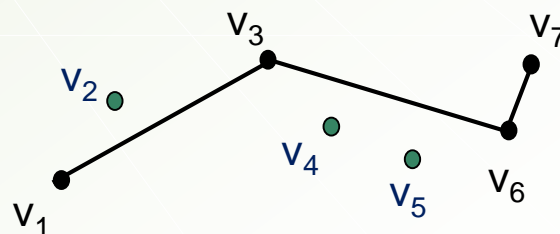
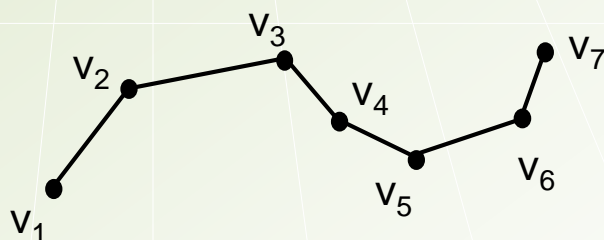
Khái quát hóa bản đồ

- ❖ **Khái quát hóa bản đồ:** tiến trình làm giảm mức độ chi tiết của bản đồ nhưng vẫn giữ được lượng thông tin.
- ❖ **Khái quát hóa đường cong:** là bộ phận của khái quát hóa bản đồ.
 - *Đơn giản hóa:* loại bỏ các cặp tọa độ không cần thiết theo tiêu chí hình học nào đó như khoảng cách giữa hai điểm, độ lệch khỏi đường trực
 - *Làm trơn:* qui trình xếp đặt lại hay dịch chuyển tọa độ để làm trơn tru các đoạn gồ ghề để hiển thị (làm tăng các cặp tọa độ)
 - *Dịch chuyển đặc trưng:* dịch chuyển các đặc trưng ở bản đồ tỷ lệ nhỏ để tránh chập dính hay phủ nhau
 - *Tăng cường/kết cấu bề mặt:* cho khả năng phát sinh chi tiết từ tập dữ liệu đã đơn giản hóa, phát sinh điểm răng cưa cho đường bờ biển
 - *Trộn:* trộn hai đặc trưng song song trong bản đồ tỷ lệ nhỏ, ví dụ hai cạnh đường quốc lộ, hai bờ sông sẽ được trộn.



Đơn giản hóa đường cong

- ❖ Đặc trưng sông ngòi, đường quốc lộ, các đường biên hành chính được hình thành từ các khâu đoạn thẳng.
- ❖ Dữ liệu không gian được nhập vào bằng bàn số hóa hay máy quét, chúng phát sinh vô số dữ liệu dư thừa không cần thiết
- ❖ Có nhiều thuật toán hỗ trợ đơn giản hóa đường cong. Yêu cầu chung đòi hỏi thuật toán là đường cong phải ít biến dạng.
- ❖ Thí dụ đơn giản hóa đường cong:
 - Giả sử ta có đường gấp khúc C với tập đỉnh V tùy ý
 - Đơn giản hóa C thành C': làm giảm số đỉnh V thành V' và vẫn giữ nguyên hai đầu mút. Như vậy V' là tập con của V. Không tạo ra đỉnh mới.

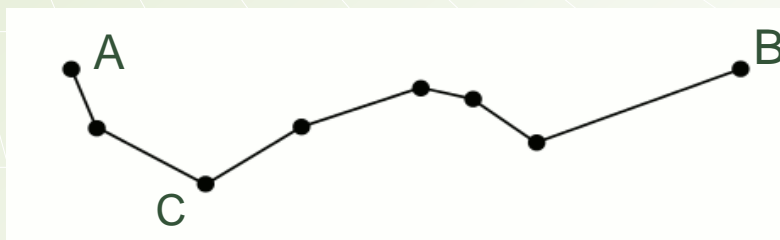




Đơn giản hóa đường cong

❖ Thuật toán *Douglas-Peucker* (1973)

- Ta xét polyline với đầu mút A, B và một sai số xấp xỉ $\varepsilon > 0$
- Dựng đường thẳng nối hai đầu mút A-B
- Tìm đỉnh C trên polyline có khoảng cách h lớn nhất đến A-B
- Nếu khoảng cách $h < \varepsilon$ thì xấp xỉ polyline bởi đoạn thẳng A-B
- Nếu không, lặp lại quá trình trên với đoạn thẳng A-C và C-B cho đến khi thu được đường xấp xỉ cuối cùng



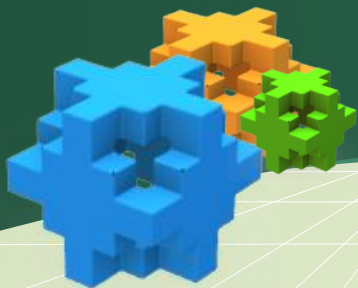
❖ Khoảng cách từ một điểm (x_0, y_0) đến đường thẳng xác định bởi $P_1=(x_1, y_1)$ và $P_2=(x_2, y_2)$

$$\text{distance}(P_1, P_2, (x_0, y_0)) = \frac{|(y_2 - y_1)x_0 - (x_2 - x_1)y_0 + x_2y_1 - y_2x_1|}{\sqrt{(y_2 - y_1)^2 + (x_2 - x_1)^2}}.$$



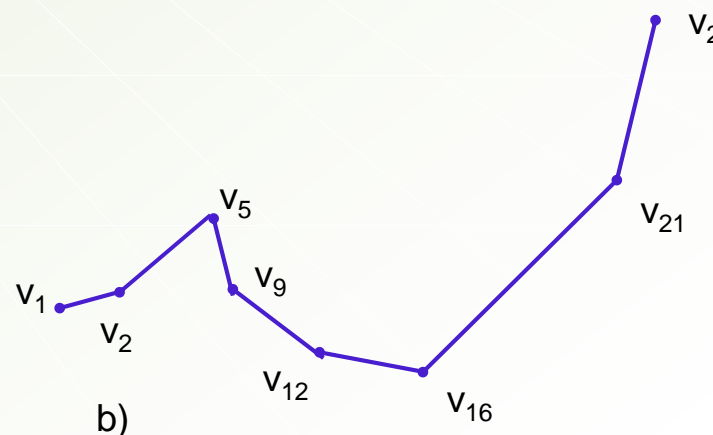
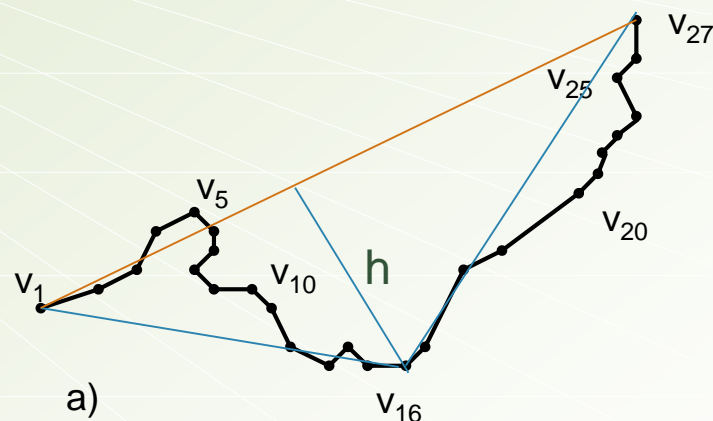
Thuật toán *Douglas-Peucker*

```
function DouglasPeucker(PointList[], epsilon)
    // Find the point with the maximum distance
    dmax = 0; index = 0;
    for i = 2 to (length(PointList) - 1) {
        d = shortestDistanceToSegment(PointList[i], Line(PointList[1], PointList[end]))
        if ( d > dmax ) then
            index = i; dmax = d ;
        endif;
    endfor;
    // If max distance is greater than epsilon, recursively simplify
    if dmax > epsilon then // Recursive call
        recResults1[] = DouglasPeucker(PointList[1...index], epsilon)
        recResults2[] = DouglasPeucker(PointList[index...end], epsilon)
        // Build the result list
        ResultList[] = {recResults1[1...length(recResults1)-1]
                        recResults2[1...length(recResults2)]}
    else
        ResultList[] = {PointList[1], PointList[end]}
    endif;
    // Return the result
    return ResultList[]
end. /*Function*/
```

Đơn giản hóa đường cong

- ❖ Thuật toán đơn giản hóa đường cong phân cấp (Cromley, 1991)
 - Ví dụ có đường cong với 27 điểm
 - Tính toán tương tự như thuật toán Douglas-Peucker: nối v_1 với v_{27} . Vì v_{16} ở xa đường cơ sở nhất, hãy chia đôi đường cong thành v_1, v_{16} và v_{16}, v_{27}
 - Gán điểm v_{16} và giá trị khoảng cách h (ví dụ 180,27) vào đỉnh của cây
 - Thực hiện lặp tính toán khoảng cách các điểm xa đường cơ sở nhất cho cả hai nhánh v_1-v_{16} và $v_{16}-v_{27}$, đồng thời điền vào cây:
 - Ví dụ, chọn v_5 và v_{21} cho mức 2
 - chọn v_2, v_{12}, v_{17} và v_{25} cho mức 3...



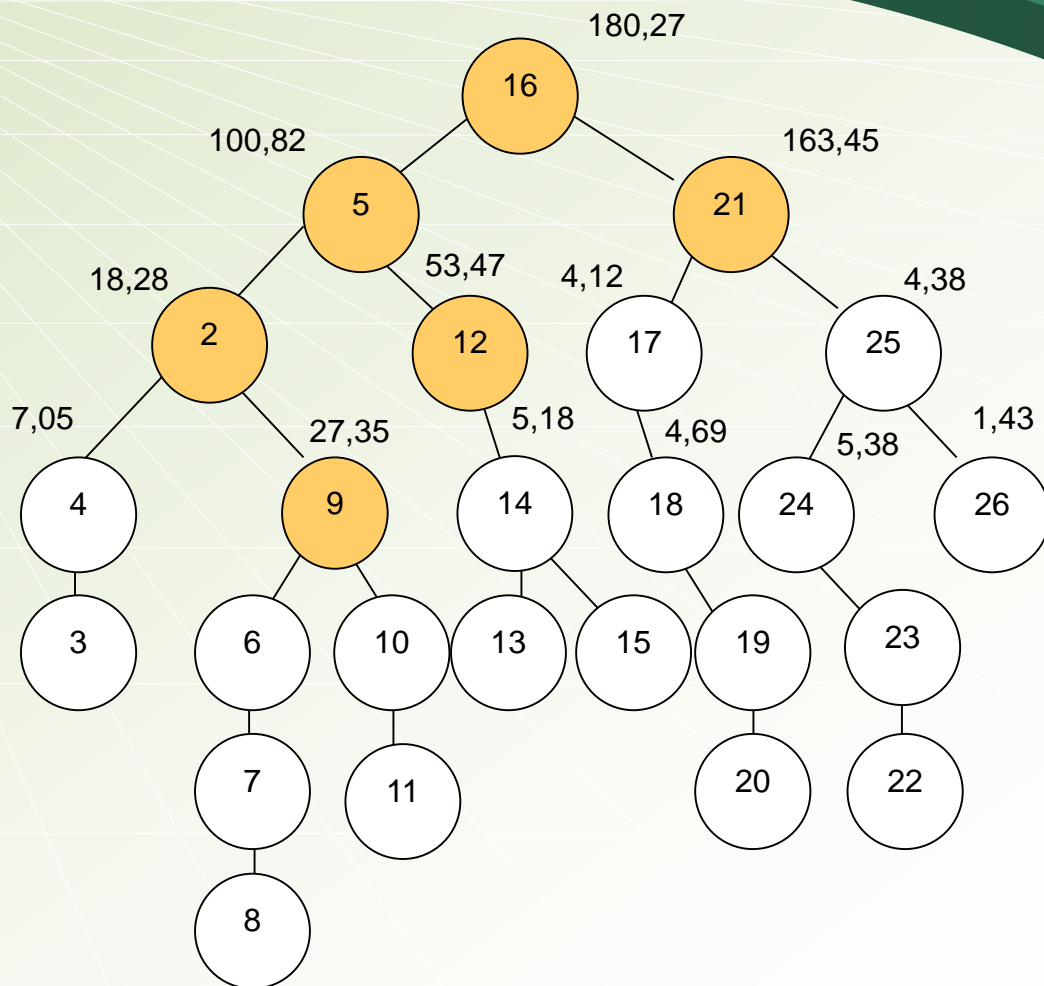


Đơn giản hóa đường cong

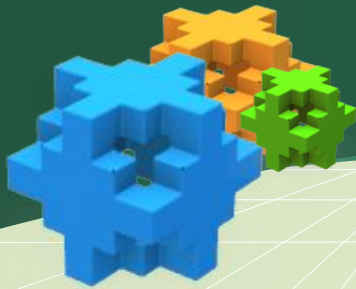
❖ Thuật toán đơn giản hóa đường cong phân cấp (tt)

■ Chọn mức độ khái quát hóa sau khi dựng xong cây phân cấp

- chỉ những cặp tọa độ có giá trị lớn hơn ε cho trước thì được giữ lại
- Ví dụ nếu sai phân là 17 thì các đỉnh $v_{16}, v_5, v_{21}, v_2, v_{12}, v_9$ của đường cong được giữ lại để hình thành đường cong thay thế

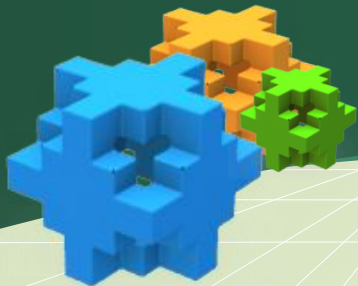


(Robert B. McMaster, K.Stuart Shea, *Generalization in Digital Cartography*, A.A.G. 1992)



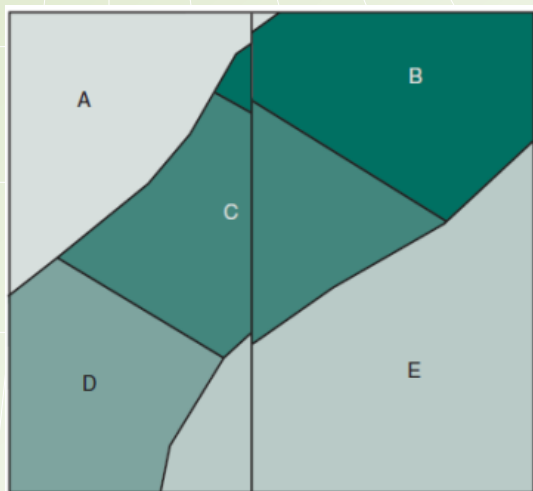
Đánh giá đơn giản hóa đường cong

- ❖ Có nhiều độ đo khác nhau được sử dụng để **đánh giá** tiến trình đơn giản hóa
 - đo thuộc tính đường đơn giản hóa
 - đo độ dịch chuyển (thay đổi) diện tích
- ❖ *Đo thuộc tính*: áp dụng cho đường đơn giản hóa như đo độ dài của đường, góc và độ cong
 - áp dụng cho các đường cơ sở hay toàn bộ đường
- ❖ *Đo dịch chuyển diện tích (so sánh)*: đánh giá sự khác nhau giữa đường cơ sở và đường đơn giản hóa
- ❖ Kết quả các phép đo trên cho thấy sự khác nhau giữa các thuật toán trong việc bảo toàn các đặc tính hình học quan trọng của dữ liệu.



Lỗi thu thập dữ liệu gián tiếp

- ❖ Lỗi khi ghép các mảnh bản đồ với nhau và lỗi bản đồ bị méo khi chụp, scan hay số hóa.
- ❖ Kỹ thuật tấm cao su có thể loại bỏ các lỗi này
 - Kỹ thuật này còn sử dụng nắn chỉnh các bản đồ chụp, bản đồ số hóa từ bản số hóa





Nắn chỉnh hình học

❖ Hiệu chỉnh hình học bằng kỹ thuật "*Rubber-Sheeting*"

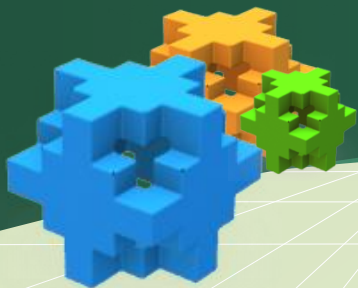
- Thực hiện phép biến đổi $T(c,r) \rightarrow (x,y)$:

$$x = \sum_{i=0}^N \sum_{j=0}^N a_{ij} P_i(c) P_j(r)$$

$$y = \sum_{i=0}^N \sum_{j=0}^N b_{ij} P_i(c) P_j(r)$$

- P - đa thức *Chebyshev*,
- a_{ij} và b_{ij} - ước lượng bằng phương pháp bình phương cực tiểu hoặc giá trị trung bình nhỏ nhất trên tập dữ liệu thu được từ các điểm điều khiển.





Nắn chỉnh hình học

- ❖ Hiệu chỉnh hình học bằng kỹ thuật "*Rubber-Sheeting*"
 - Cần $k=(N+1)(N+2)/2$ điểm điều khiển để xác định đa thức bậc N
 - Với biến đổi *Affine* bậc nhất sáu tham số (*Sprinsky*)

$$u = \beta_0 x + \beta_1 y + \beta_2$$

$$v = \beta_3 x + \beta_4 y + \beta_5$$

$\beta_0 \dots \beta_5$ được ước lượng bằng phương pháp bình phương tối thiểu

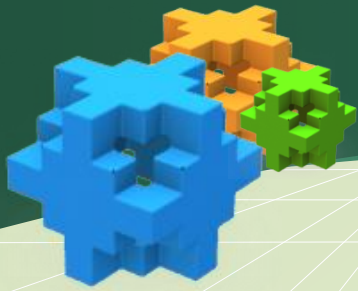
- Ước lượng β_i với n điểm điều khiển, như sau:

$$\bar{x} = \frac{\sum_{i=1}^{i=n} x_i}{n}$$

$$\bar{y} = \frac{\sum_{i=1}^{i=n} y_i}{n}$$

$$p = \frac{\sum_{i=1}^{i=n} [(x_i - \bar{x})(y_i - \bar{y})]}{\sum_{i=1}^{i=n} (y_i - \bar{y})^2}$$

$$q = \frac{\sum_{i=1}^{i=n} [(x_i - \bar{x})(y_i - \bar{y})]}{\sum_{i=1}^{i=n} (x_i - \bar{x})^2}$$



Nắn chỉnh hình học

❖ Hiệu chỉnh hình học bằng kỹ thuật "*Rubber-Sheeting*"

- Tính các tham số β_i

$$\beta_0 = 1 - q\beta_1 \frac{\sum_{i=1}^{i=n} [(x_i - \bar{x})(u_i - x_i)]}{\sum_{i=1}^{i=n} (x_i - \bar{x})^2}$$

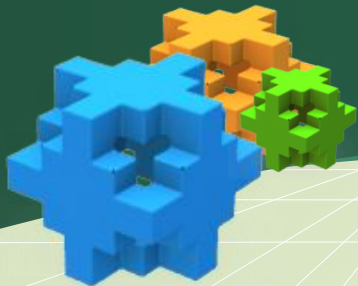
$$\beta_2 = \frac{1}{n} \sum_{i=1}^{i=n} (u_i - x_i) + \bar{x} - \beta_0 \bar{x} - \beta_1 \bar{y}$$

$$\beta_4 = 1 - p\beta_3 + \frac{\sum_{i=1}^{i=n} [(y_i - \bar{y})(v_i - y_i)]}{\sum_{i=1}^{i=n} (y_i - \bar{y})^2}$$

$$\beta_1 = \frac{\sum_{i=1}^{i=n} [(y_i - \bar{y})(u_i - x_i)] - q \sum_{i=1}^{i=n} [(x_i - \bar{x})(y_i - \bar{y})]}{\sum_{i=1}^{i=n} (y_i - \bar{y})^2 - q \sum_{i=1}^{i=n} [(x_i - \bar{x})(y_i - \bar{y})]}$$

$$\beta_3 = \frac{\sum_{i=1}^{i=n} [(x_i - \bar{x})(v_i - y_i)] - p \sum_{i=1}^{i=n} [(y_i - \bar{y})(v_i - y_i)]}{\sum_{i=1}^{i=n} (x_i - \bar{x})^2 - p \sum_{i=1}^{i=n} [(y_i - \bar{y})(x_i - \bar{x})]}$$

$$\beta_5 = \frac{1}{n} \sum_{i=1}^{i=n} (v_i - y_i) + \bar{y} - \beta_4 \bar{y} - \beta_2 \bar{y}$$



Nắn chỉnh hình học

❖ Hiệu chỉnh hình học bằng kỹ thuật "*Rubber-Sheeting*"

- Tính các tham số α_i

$$\alpha_0 = \frac{\beta_5}{\beta_4\beta_0 - \beta_3\beta_1}$$

$$\alpha_1 = \frac{-\beta_1}{\beta_4\beta_0 - \beta_3\beta_1}$$

$$\alpha_2 = \frac{\beta_1\beta_5 - \beta_2\beta_4}{\beta_4\beta_0 - \beta_3\beta_1}$$

$$\alpha_3 = \frac{-\beta_3}{\beta_4\beta_0 - \beta_3\beta_1}$$

$$\alpha_4 = \frac{\beta_0}{\beta_4\beta_0 - \beta_3\beta_1}$$

$$\alpha_5 = \frac{\beta_2\beta_3 - \beta_5\beta_0}{\beta_4\beta_0 - \beta_3\beta_1}$$

- Tính các giá trị hệ tọa độ (x, y) - sau khi nắn chỉnh:

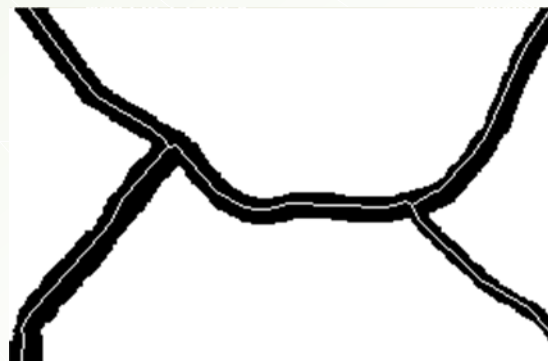
$$x = \alpha_0 u + \alpha_1 v + \alpha_2$$

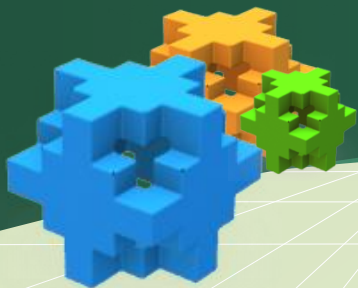
$$y = \alpha_3 u + \alpha_4 v + \alpha_5$$



Quy trình tự động thu thập

- ❖ Các bước chính chuyển đổi ảnh raster sang vectơ
 - Threshold: Chuyển ảnh đa mức xám về ảnh 2 màu
 - Smooth: Loại bỏ nhiễu ngẫu nhiên
 - Thin: Làm mảnh đến 1 pixel
 - Chain code: Chuyển đổi ảnh raster mảnh thành cung (xâu pixel)
 - Vector-reduction: Mỗi xâu pixel được chuyển đổi thành trình tự các vectơ.
 - Simplification: Đơn giản hóa đường cong
 - Rectification: Nắn chỉnh,
 - Georeferencing: Tham chiếu địa lý
 - Tạo lập topo

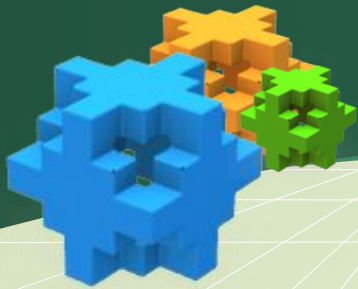




4. Chuyển đổi dữ liệu địa lý

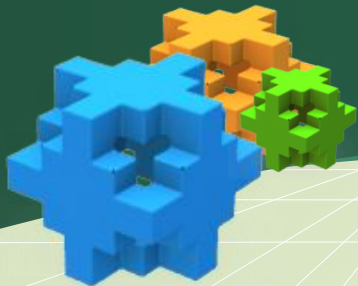
- ❖ Cân nhắc nguồn dữ liệu địa lý khi phát triển dự án
- ❖ Cách tốt nhất để tìm kiếm nguồn dữ liệu địa lý là tìm trên Internet trong *Geolibraries* hoặc *Geoportals*.

Source	URL	Description
AGI GIS Resource List	www.geo.ed.ac.uk/home/giswww.html	Indexed list of several hundred sites
The Data Store	www.data-store.co.uk/	UK, European, and worldwide data catalog
Geospatial One-Stop	www.geodata.gov	Geoportal providing metadata and direct access to over 50 000 datasets
MapMart	www.mapmart.com/	Extensive data and imagery provider
EROS Data Center	edc.usgs.gov/	US government data archive
Terraserver	www.terraserver-usa.com/	High-resolution aerial imagery and topo maps
Geography Network	www.GeographyNetwork.com	Global online data and map services
National Geographic Society	www.nationalgeographic.com	Worldwide maps
GeoConnections	www.connect.gc.ca/en/692-e.asp	Canadian government's geographic data over the Web
EuroGeographics	www.eurogeographics.org/eng/01_about.asp	Coalition of European NMOs offering topographic map data
GEOWorld Data Directory	www.geoplace.com	List of GIS data companies
The Data Depot	www.gisdatadepot.com	Extensive collection of mainly free geographic data depot



Chuyển đổi dữ liệu địa lý

- ❖ Sử dụng phần mềm để chuyển đổi (nhập) dữ liệu từ các nguồn khác
 - phần mềm chuyển đổi độc lập
 - chức năng chuyển đổi tích hợp trong các GIS
- ❖ Ngày càng có sẵn nhiều dữ liệu lưu trữ trên các phương tiện từ
 - dữ liệu bản đồ số của USGS (DLG – Digital Line Graph)
 - mô hình số độ cao (DEM – Digital Elevation Model)
 - dữ liệu liên quan đến tổng điều tra dân số (TIGER)
 - dữ liệu từ các hệ thống CAD/CAM (AutoCAD, DXF)
 - dữ liệu từ các GIS khác (Shp, MIF, DGN)



Thu thập dữ liệu thuộc tính

- ❖ Có thể thu thập đồng thời dữ liệu hình học và dữ liệu thuộc tính của đối tượng địa lý
- ❖ Thu thập tách biệt có thể hiệu quả hơn
 - Thu thập dữ liệu thuộc tính là công việc đơn giản hơn
 - Cần quan tâm đến Id để liên kết dữ liệu thuộc tính với dữ liệu hình học
- ❖ Thu thập meta-data (dữ liệu của dữ liệu)
 - Identification information: Nguồn dữ liệu, thời gian thu thập...
 - Data quality information: Độ chính xác về vị trí, thời gian và thuộc tính
 - Entity and attribute information: Các thuộc tính liên quan, độ đo



5. Tổng kết bài

- ❖ Mô tả luồng công việc thu thập dữ liệu
- ❖ Các kỹ thuật sử dụng để thu thập dữ liệu trực tiếp trong viễn thám và trắc địa
- ❖ Các kỹ thuật cơ bản trong thu thập dữ liệu gián tiếp như scanning, digitizing và vectorization
- ❖ Nguyên tắc chuyển đổi dữ liệu, nguồn dữ liệu địa lý, khuôn mẫu dữ liệu địa lý.

Câu hỏi?