

지리정보시스템에서 등고선 벡터화를 위한 최적 세선화 알고리즘에 대한 고찰

원남식¹ · 전일수¹ · 이두한² · 부기동¹

A Consideration of the Optimal Thinning Algorithm For Contour Line Vectorizing in the Geographic Information System

Nam-Sik Won¹ · Il-Soo Jeon¹ · Doo-Han Lee² · Ki-Dong Bu¹

요 약

지리정보시스템은 지리정보를 가장 효과적으로 저장하여 검색할 수 있고, 가공 처리할 수 있는 유용한 도구이다. 지리정보시스템을 구축하는데 있어서 가장 많은 시간과 비용이 소요되는 것은 입력 부분이기 때문에 벡터라이징 시스템의 개발은 매우 중요하다. 지리정보시스템에서 처리되는 모든 종류의 도면 중에서 특히 등고선도가 가장 많은 노력을 요한다. 본 논문에서는 지리정보시스템에서 등고선 벡터라이징을 위한 최적 세선화 알고리즘에 관하여 고찰하였고, 실험 결과로서 연결값을 이용한 세선화 알고리즘이 유사도가 가장 우수한 알고리즘임을 입증하였다.

주요어: 세선화 알고리즘, 등고선, 벡터라이징

ABSTRACT

Geographic Information System(GIS) which facilitates efficient storage and retrieval of geographic information is very useful tools . It is of extreme importance to develop automated vectorizing system as input method for GIS, because it takes a large amount of time and effort in constructing a GIS. In all kinds of map processed by GIS, contour line map specially takes a large amount of effort. In this paper we have considered an optimal thinning algorithm for the contour line vectorizing in the GIS. Based on the experimental results, it has been proved that thinning algorithm using the connection value is most excellent algorithm in the similarity and connectivity.

KEYWORDS: Thinning Algorithm, Contour Line, Vectorizing

1999년 3월 9일 접수 Received on March 9, 1999

¹ 경일대학교 컴퓨터공학과({nswon, isjeon, kdbu}@bear.kyungil.ac.kr)
Dept. of Computer Engineering, Kyungil University

² 경동전문대학 전산정보처리학과(dhlee@serve.kyungdong-c.ac.kr)

서 론

지리정보를 사용하는데 있어서 가장 경제적이고 효율적인 판단을 내릴 수 있도록 도와주는 의사결정 지원 도구로서 지리정보시스템(Geographic Information System : GIS)은 여러 가지 형태의 지리정보를 수집하여 입력, 저장시킨 후, 관리 및 분석하기 위해 사용되고 있다. GIS는 1960년대 미국, 캐나다 등에서 시도 후, 현재 세계 각국에서 여러 분야에 걸쳐서 다양한 목적으로 널리 이용되고 있으며 우리나라에서도 1980년대부터 지도 제작, 군사 목적, 도시계획, 지질조사, 환경 자원 분야 등에 이용되고 있으며, 최근에 사회 여러 분야에서 지리정보를 효과적으로 저장하여 검색할 수 있도록 하는 지리정보시스템에 대한 관심이 더욱 고조되고 있다.

일반적으로 GIS는 입력, 관리, 처리와 분석, 그리고 출력의 4가지 기본 요소로 구성되며, 이 중에서 자료의 입력은 실제계의 데이터를 지리정보시스템에서 사용되는 자료 형태로 변환하는 과정으로 가장 많은 시간과 비용이 소요되므로 효율적인 자료 입력 방법의 선택은 매우 중요하다. 벡터라이징 시스템의 입력 방법으로는 스캐너를 통해 얻은 래스터 영상에 대해 벡터라이징을 수행하는 것이다.

지리 도면은 종류와 사용 용도에 따라 그 특성이 다양하기 때문에 이를 잘 반영하는 벡터라이징 시스템을 설계하는 것이 중요하다. 기존에 발표된 대부분의 시스템은 모든 도면에 대하여 동일한 방법을 사용하면서 사용자가 도면에 따라 적절히 매개변수를 설정하여 벡터 데이터를 생성하고 있다. 이러한 벡터 데이터를 생성하기 위해서는 입력된 지리 도면의 영상을 한 화소 두께의 골격선을 형성하기 위한 세선화 알고리즘(Thinning Algorithm)(Arcelli 등, 1981; Lam 등, 1992; Kasturi 등, 1990; Pavlidis, 1986; Xu 등, 1987)이 필요하게 된다.

일반적으로 지리도면은 선분 요소와 문자, 숫자, 기호 등의 심벌로 이루어진다. 이중에 지리도면의 자동 벡터라이징에서 처리하는 부분은 선분요소이고, 선분 요소는 직선과 곡선으로 이루어지기 때문에 도면의 종류를 곡선 위주의 등고선 도와 다각형 위주의 지적도, 그리고 직선 위주의 상하수도로 구분하여 처리된다. 이와 같이 지리 도면은 그 종류에 따라 다양한 특성을 가지므로 도면의 종류에 따라 적합한 벡터 데이터를 생성할 수 있는 세선화 알고리즘(이경호 등, 1998; 이경호 등, 1996; Boatta 등, 1992; Suzuki, 1990)이 선택되어야 한다.

세선화 하는 목적은 도면 영상을 한 화소 두께의 골격선으로 나타내어 자료의 형태를 단순화하여 지도 영상의 이미지를 유지하는데 필요한 정보의 양을 줄이기 위함이다. 그러므로 각 도면에 가장 적합한 세선화 알고리즘을 적용하여 지도 영상을 벡터라이징하는 것은 GIS의 질을 결정하는 가장 중요한 부분이 된다.

도면을 세선화하여 추출된 골격선에는 패턴 주의의 먼지, 잡영 등에 의해 불필요한 잡영이 나타나기도 하고, 세선화 알고리즘의 특성에 의한 연결 선분의 끊어짐, 두 화소 두께의 골격선 형성, 일부 선분의 소멸현상, 불필요한 잡 가지의 발생 등 많은 문제들이 나타나므로, 이러한 문제 발생을 최소화하기 위한 세선화 알고리즘의 개발은 중요한 과제가 된다.

본 논문에서는 기존의 시스템에서 처리하는데 가장 많은 시간을 요하는 곡선 위주의 등고선도 처리에 적합한 알고리즘으로서 연결값을 이용한 세선화 알고리즘(원남식과 손운구, 1995)을 제안하고, 이를 구현하여 그 결과를 기존의 잘 알려진 세선화 알고리즘들(Mandalia 등, 1992; Lu, 등, 1985; Zhang 등, 1984; Wang 등, 1986)과 성능을 비교하였다.

연결값을 이용한 세선화 알고리즘

세선화 알고리즘은 영상 패턴의 특징을 파괴하지 않는 조건으로, 배경과 인접한 경계화소(contour pixel)를 제거하는 연산 과정이다. 즉, 인식 대상 영상을 양자화 하여 이를 구성하는 가장 기본 요소인 화소를 다루는 작업으로서 인식 대상 영상의 특징 추출에 무관한 화소를 제거하여 골격선에서 화소의 두께가 1인 골격을 형성하는 과정이다. 세선화 알고리즘이 만족해야 하는 일반적인 사항으로는 위상적인 특성, 기하학적인 특성, 등방성, 복원력, 빠른 처리 속도 등이 있다(Lam 등, 1992).

특히, 문자인식에서 세선화된 골격선의 기하학적인 특성 유지는 매우 중요한 문제로서 이는 문자의 인식률에 직접적인 영향을 준다. 그러나 대부분의 세선화 알고리즘에서는 세선화된 결과에서 골격의 한 부분이 사라지는 골격소멸현상이 발생하여 기하학적인 특성 유지가 어려워지게 되므로 세선화 알고리즘에서는 이러한 현상을 제거하기 위한 특별한 방법을 사용한다.

대부분의 세선화 알고리즘에서는 국부적인 이웃 화소만을 이용하여 처리하므로, 이러한 정보로는 잡영 가지와 끝점의 구별에 필요한 정보를 유지하지 못하기 때문에 이상적인 기하학적인 특성의 유지는 매우 어려운 문제가 된다. 그러므로 기하학적인 특성 유지를 위한 방법으로서 심한 침식과 잡영 가지의 생성을 동시에 방지하기 위해서 끝점 조건을 없애는 방법(Beun, 1973), 조건을 더 광범위하게 적용시키는 방법(Lu 등, 1985), 또는 세선화의 마지막 단계에서만 조건을 적용시키는 방법(Suzuki와 Abe, 1987) 등 다양한 방법이 시도되었다. 그러나, 이러한 형태의 모든 수정을 적용시켰을 때에는 상층 관계를 가지므로 이를 위한 부가적인 정보가 필요하게 된다. 이러한 이유 때문에 거리 변환을 사용하는 다양한 기준이 도입되었다(Rosenfeld 등, 1976; Arcelli

등, 1981; Chuei 등, 1986).

기존의 세선화 알고리즘들은 이미지 화소를 단순히 0과 1의 상태 값으로만 처리하고 화소 제거를 위한 지역연산 과정에서 3×3 윈도우를 사용하므로 제거 대상 화소에서 이상적인 골격 추출을 위한 주변 화소에 대한 정보 부족으로 인하여 세선화 작업에 많은 제약 요인이 되고 있다. 이러한 제약 요인을 해결하기 위한 방법으로 지역 연산을 위해 확장된 윈도우(3×4, 4×5, 5×5)를 사용하기도 한다. 그러나 이는 알고리즘의 복잡성과 효율을 저하시키는 주된 요인이 되고 있다. 이와 같은 문제를 해결하고, 품질이 우수한 골격선을 추출할 수 있는 4-이웃 연결값을 이용한 세선화 알고리즘(Lu와 Wang, 1985)이 제안된 바 있다.

1. 4-이웃 연결값

4-이웃 연결값이란 그림 1과 같이 한 화소(P)를 중심으로 수평, 수직의 4방향으로 연결된 화소(P_0, P_2, P_4, P_6)들의 합을 4-이웃 연결값(4-neighbors connection value: V_{c4})이라 정의한다. 그러므로 모든 화소가 갖는 연결값은 0에서 4까지의 값 중 하나가 된다. 그림 2에서 (a)의 데이터를 4-이웃 연결값으로 변환한 상태는 그림 2의 (b)와 같다.

$$V_{c4} = \sum_{i=0}^3 P_{2i} \cdot \cdot \cdot \cdot \cdot \cdot \cdot \quad (1)$$

P_7	P_0	P_1
P_6	P	P_2
P_5	P_4	P_3

FIGURE 1. The window for the proposed algorithm

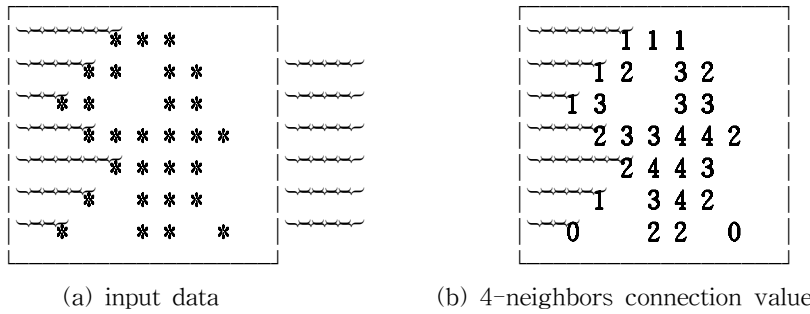


FIGURE 2. Conversion to the 4-neighbors connection value V_{c4}

2. 세선화 알고리즘 WPTA

기존의 세선화 알고리즘들에서 흔히 발생하는 문제점은 세선화된 결과가 두 화소 두께로 남는 것과 연결성이 끊어지는 것과 골격소멸현상 등이다. 이러한 현상은 문자인식 과정에서 여러 가지 나쁜 영향을 미치게 되므로 반드시 제거되어야 한다.

알고리즘 WPTA(Won's Parallel Thinning Algorithm)(원남식과 손윤구, 1995)는 세선화된 결과가 완전-8 연결 형태로 남으며, 연결성이 끊어지지 않는 세선화 알고리즘이다. WPTA 알고리즘의 화소 보존 조건은 골격의 연결성 유지를 위한 연결점(Connection point) 조건 세 가지와 끝점(End point) 유지를 위한 끝점 조건 하나를 갖는다. 그러므로 네 가지 조건 중 하나라도 만족되면 그 화소는 제거될 수 없다. WPTA에서 화소를 제거할 수 없는 네 가지 조건은 다음과 같다.

▶ WPTA 알고리즘에서 화소를 제거할 수 없는 네 조건

[조건 1] 연결점 조건 1

$$(P_0 \vee P_1 \vee P_7) \wedge (P_3 \vee P_4 \vee P_5) \wedge \overline{(P_2 \vee P_6)} = 1 \dots \dots (2)$$

[조건 2] 연결점 조건 2

$$(P_1 \vee P_2 \vee P_3) \wedge (P_5 \vee P_6 \vee P_7) \wedge \overline{(P_0 \vee P_4)} = 1 \dots \dots (3)$$

[조건 3] 연결점 조건 3

$$\sum_{i=0}^3 P_{2i} \cdot P_{(2i+2) \text{ MOD } 8} \cdot P_{(2i+5) \text{ MOD } 8} \geq 1 \dots \dots (4)$$

[조건 4] 끝점 조건

$$B(P) = \sum_{i=0}^7 P_i = 1 \cdot \dots \dots (5)$$

3. 4-이웃 연결값을 이용한 세선화 알고리즘(WPTA4)

연결값 V_{c4} 와 WPTA를 결합한 4-이웃 연결값을 이용한 세선화 알고리즘을 WPTA4로 표시한다. WPTA4의 알고리즘 수행 과정은 입력 이미지 데이터를 연결값으로 변환하고 연결값 0과 1을 갖는 화소를 먼저 제거한 다음, 이를 다시 연결값 V_{c4} 로 변환한다. 그리고 WPTA에 의해 연결값이 0에서 4까지의 화소를 연결성이 낮은값 부터 순서적으로 제거한 다음, 남은 화소 이미지 데이터를 다시 연결값으로 변환한다. 이 과정은 WPTA에 의해 제거되는 화소가 없을 때까지 반복한다.

▷ 알고리즘 WPTA4

{ 입력: 이미지 데이터
출력: 세선화된 골격

$$V_{c4} = \sum_{i=0}^3 P_{2i}$$

begin

for each pixel p do calculate V_{c4} ;

for each pixel p do

```

if  $V_{ca} = 0$  or  $V_{ca} = 1$  then delete  $p$ ;
for each pixel  $p$  do calculate  $V_{ca}$ ;
repeat
  for  $k = 0$  to 3 do
    for each pixel  $p$  do
      if ( $V_{ca} = k$ ) then WPTA;
    for each pixel  $p$  do calculate  $V_{ca}$ ;
  until ( no more pixels deleted );
end;

```

▷ WPTA

{입력 : 제거 대상 화소

출력 : 보존 화소}

begin

```

if ( 조건1 or 조건2 or 조건3 or 조건
4) = false then delete  $p$ ;

```

end;

알고리즘과 비교 분석

본 장에서는 WPTA4에서의 수행결과를 그림 4에 나타내었고, 잘 알려진 병렬세화 알고리즘으로서 Zhang 과 Suen(1984)이 제안한 알고리즘(ZS)과 ZS 알고리즘에서의 골격소멸 및 두 화소 두께로 남는 문제와 연결성을 유지하기 위해 ZS를 개선한 Lu와 Wang(1985)이 제안한 알고리즘(LW)과 Wang 등(1986)이 제안한 알고리즘(WHF)과 Mandalia 등(1992)이 제안한 알고리즘(MPS)의 수행 결과를 그림 5에서 그림 8에 나타내었다.

WPTA4 알고리즘의 수행 결과는 골격선의 연결성이 유지되고, 그 형태는 이상적인 완전-8 연결 형태로 남으며, 최종 골격선의 화소는 1에서 3까지의 수치 정보로 표현되므로 문자 인식 과정에서 다양한 형태의 정보로 이용될 수 있다. 특히, 곡선 획의 형태를 잘 나타내므로 등고선도 벡터라이징을 위한 세션화 알고리즘으로 잘 적용될 수 있을 것으로 생각된다. 그리고 각 알고리즘의 성능 분석은 표 1, 표 2에서 최종 골격선의 유사도와 연결성으로 서로 비교하였다.

1. WPTA4와 기존 알고리즘들의 세션화된 결과



FIGURE 3. Contour line input image

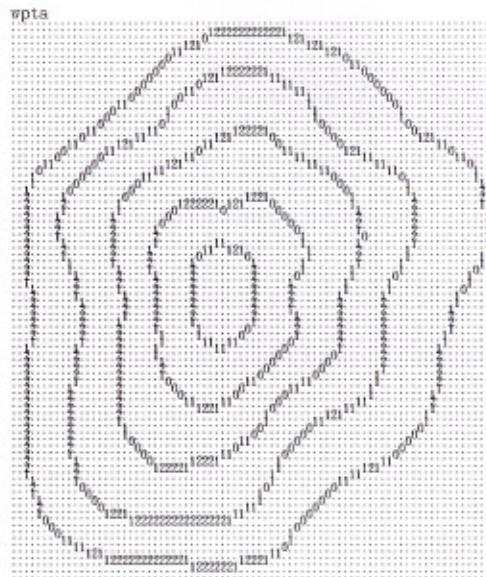


FIGURE 4. Result of the WPTA4 algorithm

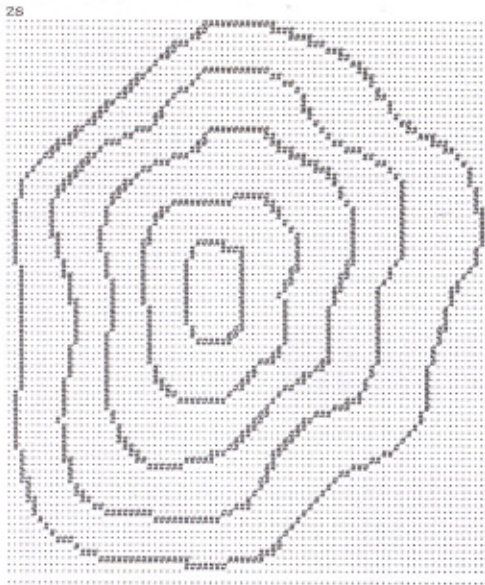


FIGURE 5. Result of the ZS algorithm

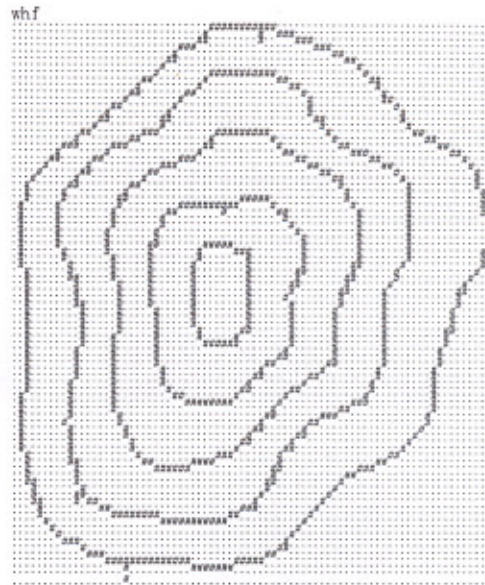


FIGURE 7. Result of the WHF algorithm

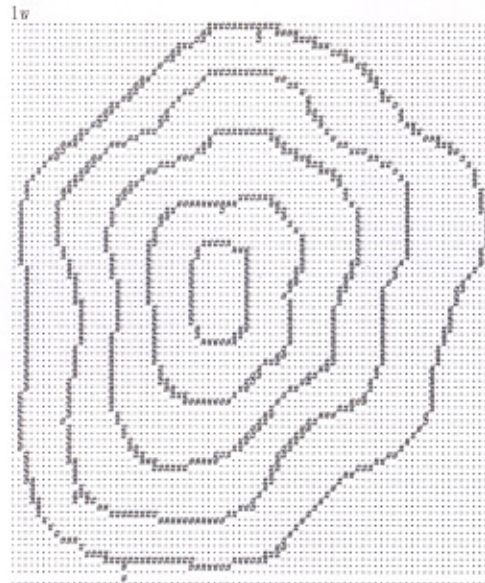


FIGURE 6. Result of the LW algorithm

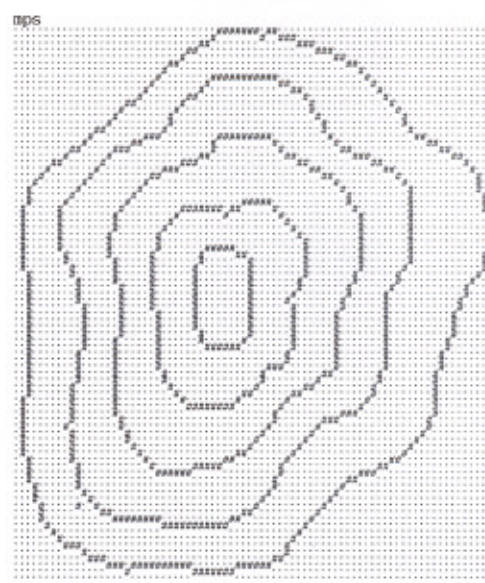


FIGURE 8. Result of the MPS algorithm

WPTA4와 기존 알고리즘과의 성능 분석

지리도면의 벡터라이징에서는 골격선이 패턴의 위상적 또는 기하학적인 정보를 유지하는데 불필요한 잡영에 둔감하여, 골격선으로부터 특징점을 쉽게 추출할 수 있어야 한다. 자동 벡터라이징 시스템을 개발하는데 있어서 세선화는 중요한 전처리 과정이다. 본 절에서는 지리 도면의 자동 입력을 위한 최적의 세선화 알고리즘을 찾기 위한 연구 중 각 알고리즘을 등고선도로 처리한 결과를 고찰하였다.

골격선의 질 면에서 등고선의 경우 도면에 실선과 점선이 결합되어 있어 이러한 도면을 스캐닝 하면 요철이 심한 영상을 얻게 된다. 이러한 점에서 볼 때 패턴 주위의 잡영에 둔감하면서 끝점의 침식 정도가 작고, 패턴 주위의 잡영에 둔감하고, 최종 세선화된 결과는 한 화소 두께의 완전 8연결 형태로 남아야 한다. GIS에 적용하기 위한 세선화 알고리즘의 가장 중요한 요인은 원래의 도면을 정확히 나타낼 수 있는 유사도(Lam 등, 1992)가 된다. 유사도는 인간 전문가에 의해

서 작성된 골격선과 비교하여 일치하는 정도를 백분율로 나타내는 것이다. 그러므로, 본 논문에서는 최종 골격선의 유사도 측정과 연결성에 의해서 각 알고리즘의 성능을 표 1, 표 2에 나타내었다.

결 론

일반적으로 지리도면은 선분 요소와 문자, 숫자, 기호 등의 심벌로 이루어진다. 이중에 지리도면의 자동 벡터라이징에서 처리하는 부분은 선분 요소이고, 선분 요소는 직선과 곡선으로 이루어지기 때문에 도면의 종류를 곡선 위주의 등고선도와 다각형 위주의 지적도, 그리고 직선 위주의 상하수도로 구분된다. 그리고 지도 영상을 구성하고 있는 패턴의 두께와 기울기, 그리고 잡영의 정도에 따라 도면 영상을 분류하고, 각 도면에 따라 벡터 데이터의 생성 기준을 다르게 적용하고 있다.

지도 영상으로부터 벡터 데이터를 생성하는 과정은 크게 나누어 네 단계로 이루어진다. 먼저 도면 영상을 세선화하여 골격선 영상을 얻는다. 두 번째로 골격선 영상을 그래프로 변환하고, 세 번째로 골격선의 위상을 유지하는데 불필요한 에지를 그래프로부터 제거한다. 마지막으로 그래프로부터 특징 점을 구하고 직선으로 연결한다. 이 과정 중에서 첫 번째 과정인 세선화 과정에 의해서 최종 데이터 질에 결정적인 영향을 미치게 된다.

도면 영상을 한 화소 두께의 골격선으로 세선화 하는 목적은 자료의 형태를 단순화하며, 지도 영상의 구조적 정보를 유지하는데 필요한 정보의 양을 줄이기 위함이다. 지

TABLE 1. Comparison of the similarity for the referenced skeleton

algorithm	similarity (%)
WTA4N	96.9
ZS	78.1
LW	78.9
WHF	83.3
MPS	88.9

TABLE 2. The connectivity of skeleton

algorithm	WTA	ZS	LW	WHF	MPS
connectivity	perfectly-8 connectivity	imperfectly-8 connectivity	imperfectly-8 connectivity	imperfectly-8 connectivity	perfectly-8 connectivity

리도면은 그 종류에 따라 다양한 특성을 가지므로 이에 적합한 세선화 알고리즘을 적용하여야 하고, 또한 벡터 데이터를 생성하는데 있어서 이러한 특성을 잘 반영하는 것이 중요하다.

본 논문에서는 연결값을 이용한 세선화 알고리즘은 골격선의 끊어짐이 없고, 잡영 발생이 적고, 세선화된 결과가 완전 8연결 형태로 남으며, 유사도가 뛰어나므로 등고선도를 벡터라이징하는데 매우 적합한 알고리즘임을 입증하였다.

KAGIS

참고문헌

- 원남식, 손윤구. 1995. 4-인접 연결값을 이용한 병렬 세선화알고리즘. 한국정보과학회논문집 22(7):1047-1056.
- 이경호, 김광수, 조성배, 최윤철. 1998. 지리정보시스템의 자동 벡터화를 위한 지식기반 시스템. 한국정보과학회논문집(B) 25(1):193-203.
- 이경호, 조성배, 최윤철. 1996. 지리정보 시스템을 위한 영상 세선화 알고리즘의 성능평가. 한국정보과학회 춘계학술논문 발표집, 23(1):3-6.
- Arcelli,C.and G.S.D.Baja. 1981. A Thinning algorithm based on Prominence detection. Pattern Recognition 13(13):225-235.
- Boatta,L., V.Consorti, M.del Buono, S.D.Zenzo, V.Eramo, A.Esposito, F.Melcarne, M.Meucci, A.Morelli, M.Mosciatti, S.Scarci and M.Mosciatti. 1992. An interpretation system for land register maps. IEEE Computer 25(7):25-33.
- Beun,M. 1973. A flexible method for automatic reading of handwritten numerals. Pilips Technical Review 33(5):89-101;130-137.
- Chuei,N., T.Y.Zhang and C.Y.Suen.1986. New algorithm for thinning binary image and Chinese character. Computer Processing of Chinese Oriental Languages 2(3):169-179.
- Kasturi,R., S. T. Bow, W.E.Masri, J. Shah, J.R.Gattiker and U.B.Mokte. 1990. A system for interpretation of line drawings. IEEE Trans. on Pattern Recognition and machine Intelligence 12(10):973-992.
- Lam,L.S.,W.Lee and C.Y.Suen. 1992. Thinning methodologies - a comprehensive survey. IEEE Trans. Pattern Analysis and Machine Intelligence 14(9):869-885.
- Lu,H.E. and P.S.P.Wang, 1985. An improved fast parallel thinning algorithm for digital patterns. In Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. pp.364-367.
- Mandalia,A.D., A.S.Pandya and R.Sudhaker. 1992. Modified fast parallel thinning algorithm for noisy handprinted characters. 92 Proceedings of the 2nd Singapore international conference on image processing pp.7-11.
- Pavlidis,T. 1986. A vectorizer and feature extractor for document recognition. Computer Vision Graphics and Image Processing 35:111-127.
- Rosenfeld,A. and L.S.Davis. 1976. A note on thinning. IEEE Trans. on Systems. Man and Cybernetics 25:226-228.
- Suzuki,S. 1990. MARIS: Map recognition input system. Pattern Recognition 23(8)919.
- Suzuki,S. and K.Abe. 1987. Binary picture thinning by an iterative parallel two-subcycle operation. Pattern Recognition 10(3):297-307.
- Wang, P.S.P., L.Hui,and T.Fleming Jr. 1986. Further improved fast parallel thinning algorithm for digital patterns. In Computer Vision, Image Processing and communications systems and application. pp. 37-40.

Xu, W., C. Wang. CGT. 1987. A fast thinning algorithm implemented on a sequential computer. IEEE Trans. on Systems, Man and Cybernetics 17(5):847-851.

Zhang, T.Y. and C.Y. Suen. 1984. A fast parallel algorithm for thinning digital patterns. Communications of the ACM, 27(6):236-239. [KAGIS](#)