

확장된 Fuzzy Clustering 알고리즘을 이용한 자동 목표물 검출

(Automatic Target Detection Using the Extended Fuzzy Clustering)

金 秀 桓*, 姜 景 辰**, 李 太 遠**

(Su Hwan Kim, Kyung Jin Kang, and Tae Won Rhee)

要 約

본 논문에서는 특징 공간에서 데이터 분류에 널리 사용되고 있는 기존의 퍼지 clustering 알고리즘을 확장하여 분류하고자 하는데 데이터의 성질을 고려할 수 있는 새로운 알고리즘을 제안한다.

이 알고리즘은 데이터의 성질에 따라 알고리즘내에서 가중치를 주어 각 클러스터의 중심값이 가중치가 높은 위치로 이동하는 특징을 갖도록 하였다. 물론, 데이터의 성질이 모두 똑같은 경우 가중치가 단위 크기를 갖으며 이 경우는 기존의 퍼지 clustering 알고리즘과 같은 clustering 효과를 나타낸다. 또한 퍼지 clustering 알고리즘의 성능에 중요한 영향을 미치는 초기 중심값 설정을 위한 새로운 알고리즘을 포함시키므로써 알고리즘의 효율성을 높였을 뿐 아니라 기존의 퍼지 clustering 알고리즘의 확장에도 불구하고 소속 함수(membership function) 크기의 증가는 없으며 계산량의 증가는 무시할 정도이다. 한편, 입력 영상내 데이터들의 밝기 정도를 가중치로 사용하여 이 알고리즘을 자동 목표물 검출에 응용할 경우 목표물간 경계의 모호함과 목표물과 배경간 경계의 모호함을 해결할 수 있을 뿐 아니라 절대적 밝기가 아닌 주위 배경에 대한 상대적 밝기에 따라 결정되는 목표물은 적응적으로 탐색할 수 있음을 실험을 통해 입증한다.

Abstract

The automatic target detection which automatically identifies the location of the target with its input image is one of the significant subjects of image processing field.

Then, there are some problems that should be solved to detect the target automatically from the input image. First of all, the ambiguity of the boundary between targets or between a target and background should be solved and the target should be searched adaptively. In other words, the target should be identified by the relative brightness to the background, not by the absolute brightness. In this paper, to solve these problems, a new algorithm which can identify the target automatically is proposed. This algorithm uses the set of fuzzy for solving the ambiguity between the boundaries, and using the weight according to the brightness of data in the input image, the target is identified adaptively by the relative brightness to the background. Applying this algorithm to real images, it is experimentally proved that it is can be effectively applied to the automatic target detection.

**正會員, *準會員, 高麗大學校 電子工學科

(Dept. of Elec. Eng., Korea Univ.)

接受日字: 1991年 7月 15日

I. 서 론

인식과 분류의 과정이 인간 활동의 가장 기본적인 과정으로 인식 되어지고 있다. 특히 같은 특징을 갖

는 항목을 하나의 묶음으로 분류하는 과정은 인간을 포함한 모든 동물의 가장 기본적인 활동이다. 이러한 분류 과정을 수학적으로 구현하는 방법에는 여러가지가 있으며 이중 clustering 알고리즘이 일반적으로 사용된다.^{[1][8]}

Clustering 알고리즘은 주어진 항목들로부터 자연스러운 묶음을 발견하는 방법으로 특징 공간상에서 가까이 있는 데이터들을 같은 클러스터에 존재한다고 가정한다. 이러한 clustering 알고리즘은 영상 신호처리 분야에 적용되어 입력 영상에서 영역이나 물체를 분할하고 인식하며 라벨을 붙일 수 있는 시스템이 최근 몇년 사이에 상당한 관심을 불러 일으키고 있다.^{[3-4][8][9][11][15][17]}

특히 입력 영상내의 자동 목표물 검출은 상당한 진전을 보이고 있는 분야이다. 여기서 자동 목표물 검출이란 입력 영상에서 가능한 목표물의 위치를 자동으로 결정하는 것을 말한다. 실제로 입력 영상의 배경으로부터 가능한 목표물을 분류하는 많은 알고리즘들이 개발되어 적용되어지고 있다.^[13]

이러한 알고리즘을 다시 세부적으로 나누면

(a) 목표물과 배경의 스펙트럼적인 특징을 특징 공간상에서 clustering함으로써 입력 영상의 배경으로부터 목표물을 분류하는 방법

(b) 목표물과 배경의 공간적인 특징을 특징 공간상에서 clustering함으로써 입력 영상의 배경으로부터 목표물을 분류하는 방법

(c) 목표물과 배경의 스펙트럼적인 특징과 공간적인 특징을 동시에 이용하여 특징 공간상에서 clustering 함으로써 입력 영상의 배경으로부터 목표물을 분류하는 방법 등이 있다.^[6]

그러나 입력 영상에서 목표물을 효과적으로 자동 탐색하기 위해서는 몇가지 해결해야 할 점이 있다. 우선 목표물간 경계의 모호함과 목표물과 배경간 경계의 모호함을 해결해야 할 뿐만 아니라 목표물이 절대적 밝기가 아닌 주위 배경에 대한 상대적 밝기에 따라 결정되기 때문에 목표물은 적응적으로 탐색되어져야 한다.^{[6][10][13]}

따라서 본 논문에서는 이러한 문제점을 해결하기 위하여 입력 영상내의 데이터들의 밝기 정도에 따라 가중치로 주는 새로운 알고리즘을 이용하며 입력 영상의 배경으로부터 목표물을 효율적으로 분류하여 검출한다. 각 장에 대한 설명을 다음과 같다. 2장에서는 퍼지 집합의 특징을 설명하고 3장에서는 일반적인 퍼지 clustering 알고리즘에 대해서 설명하며 4장에서는 본 논문에서 제안하는 확장된 퍼지 clustering 알고리즘에 대해서 설명한다. 그리고 5장과 6장

에서는 이 알고리즘을 실험적으로 증명하고 결론을 기술한다.

II. 퍼지 집합의 특징

퍼지 집합이란 집합에 속하는 member를 확실히 구분할 수 없는 집합을 말한다. 예를 들면 “집합 A는 10에 근접한 실수”라는 퍼지 집합은 다음과 같은 순서쌍으로 표현된다.

$$A = \{ (x, u_A(x)) : x \in X \}$$

$$u_A(x) = \begin{cases} 0 & x < 10 \\ (1 + (x - 10)^2)^{-1} & x \geq 10 \end{cases} \quad (1)$$

여기서 $u_A(x)$ 는 소속함수(memberhip function)를 나타내고, 그림으로 표현하면 아래와 같다.

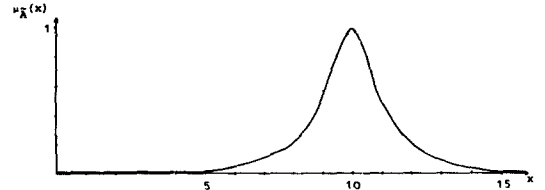


그림 1. “10에 근접한 실수”의 소속함수
Fig. 1. Real numbers close to 10.

기존의 집합과의 차이점은 집합을 규정하는 특성 함수(즉, membership function)이 0과 1사이에 고르게 분포되는 특징이 있으며 다음과 같은 수식으로 나타낼 수 있다.

$$u_A(x) : A \rightarrow \{0, 1\} \text{ 기존의 집합}$$

$$u_A'(x) : A \rightarrow [0, 1] \text{ 퍼지 집합} \quad (2)$$

퍼지 집합의 장점은 기존의 crisp 집합이 표현하기 어려운 애매한 문제를 쉽게 설명할 수 있다. 예를 들어, “오늘 날씨가 맑은가?”라는 문제에서 Crisp 집합의 경우는 하늘에 구름이 몇% 있을 때 맑다는 등을 상세히 정해 주어야 결론을 내릴 수 있다. 그러나 퍼지 집합은 하늘에 구름이 적으니까 맑다는 결론을 쉽게 내린다. 이러한 이유로 퍼지 집합은 시스템의 모델이 복잡한 경우에도 쉽게 제어할 수 있다.

현재 모든 시스템의 예측을 위해서는 주로 확률을 사용하여 결정을 내리고 있다. 그러나 통계 처리를 할 수 없거나, 할 수 있어도 정확한 확률을 위해서 샘플을 많이 취해야 하는 경우에는 아무래도 확률을 가지고는 부족하다는 감이 든다. 사실, 통계처리를 하는 것이 현재에는 가장 정확하다고 할 수 있지만,

전술한 문제를 위해서는 가능성이라는 개념을 도입하게 된다. 전문가의 충분한 경험을 바탕으로 예측한 가능성은 확률보다 정확하다고 할 수 있다. 쉽게 이야기해서, 결국 이러한 가능성은 퍼지 집합에서의 소속함수가 되기 때문에 퍼지 이론의 장점이 되는 것이다. 즉, 확률적 처리가 불가능한 시스템에서도 퍼지 이론을 이용하면 간단하게 처리된다.

III. 퍼지 Clustering 알고리즘

일반적인 clustering 알고리즘은 특정 공간내의 데이터들 사이의 유사도를 측정하기 위한 수학적 도구이다. 즉, 특정 공간내의 데이터들은 서로의 유사도를 계산하고 이를 근거로하여 몇개의 부분집합으로 분류한다. 예를 들어 주어진 문제에서 특징 추출이 되어 있다고 하자. 여기서 문제 해결을 위해서는 "P"로 표시되는 특성을 가지는 n개의 개체 $x \in X$ 를 클러스터라고 불리는 동질성을 가지는 $c(2 \leq c \leq n)$ 개의 부류로 나누는 것이다. 하나의 클러스터에 속하는 개체는 비슷한 성질을 가져야 하며, 다른 클러스터에 속하는 개체는 가능한 한 상이한 성질을 가져야 한다. 이를 위해서는 먼저 데이터의 유사도를 측정하기 위해서 어떤 수학적 성질은 사용할 것인가와 클러스터를 구분하기 위해 이러한 성질을 어떤 방법으로 사용할 것인가에 문제가 있다. 이러한 문제는 주어진 문제에 따라 달라질 수 있는데, 그 이유는 일반적으로 최적의 clustering방법에 관한 기준이 아직 없기 때문이다. 현재까지 연구되어지고 있는 clustering 방법은 계층적(hierachical), 그래프 이론적(graph theoretic) 또는 목적 함수적(objective functional) 방법으로 구분된다. 그 중 목적함수 방법은 대부분의 clustering 알고리즘의 기준 설정에 비교적 정확한 방법을 제공하는 것으로 알려져 있다.

1. 전통적인 clustering 알고리즘과 퍼지 clustering 알고리즘의 비교

전통적인 clustering 알고리즘의 각 개체는 오직 하나의 클러스터에 할당된다. 따라서 개체는 중간적인 클러스터의 특징을 가짐으로써 정확히 하나의 클러스터에 할당되기 어려운 경우가 있다. 이에 반해 최근 활발히 연구되어 지고 있는 퍼지 clustering 방법에서는 좀 더 적절한 clustering 방법을 제공함으로써 퍼지 clustering 알고리즘은 퍼지 집합 이론의 성공적인 응용 사례중 하나로 받아들여지고 있다. 전통적인 clustering 방법과 퍼지 clustering 방법의 차이점은 나비 문제로 불리는 문제점에서 찾아볼 수 있다.

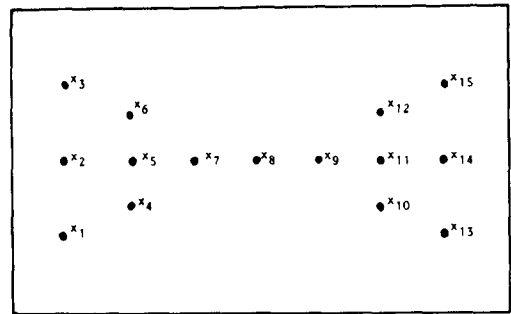


그림 2. 나비
Fig. 2. The butterfly.

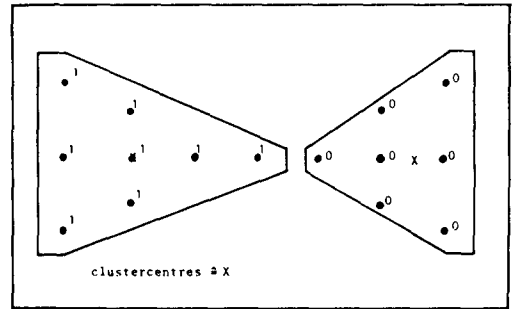


그림 3. 나비의 Crisp 클러스터들
Fig. 3. Crisp clusters of the butterfly.

그림2는 15개의 점으로 구성되는 데이터 집합X를 표현하는 평면 공간을 나타내고 있다. 전통적인 clustering 방법에 의해 이들 점들은 그림3에 나타낸 바와 같이 분류된다. 여기서 '1'은 왼쪽 클러스터에 속하는 것을 표시하며 '0'은 오른쪽 클러스터에 속하는 점을 표시한다.

반면 그림4와 그림5는 각각 퍼지 clustering 방법에 의해서 각 점들에 대해 왼쪽과 오른쪽 클러스터에 속하는 귀속 정도(degree of membership)를 나타낸다.

그림4와 그림5에서 알 수 있듯이 나비 모양의 점들은 대칭이지만 그림3에서는 비대칭이다. 즉, 정확히 가운데의 점이 반드시 왼쪽 또는 오른쪽으로 귀속되어야 하기 때문에 비대칭이다. 그러나 그림4와 그림5에서는 이 점의 귀속도가 0.5이므로 이러한 문제점을 극복할 수가 있다.

2. 퍼지 평균과 퍼지 공분산 최근들어 많은 퍼지 clustering 알고리즘이 데이터

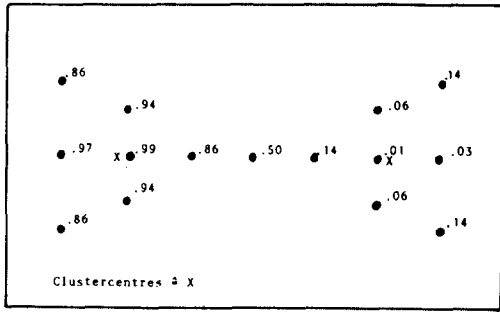


그림 4. 나비의 첫번째 클러스터
Fig. 4. Cluster 1 of the butterfly

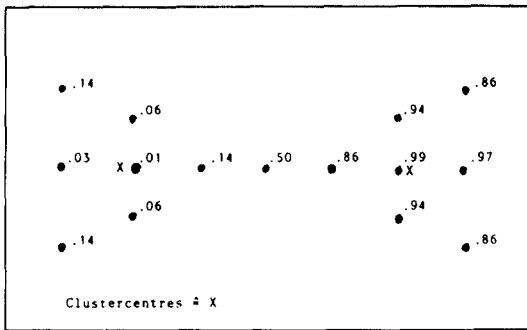


그림 5. 나비의 두번째 클러스터
Fig. 5. Cluster 2 of the butterfly.

집합의 애매한 분할 문제를 해결하기 위해 제안되어 지고 있다. 이들 알고리즘들은 전형적인 평균이나 공분산 대신에 퍼지 평균과 퍼지 공분산을 사용하며 소속 함수의 특징에 따라 특징지어 진다.

확률 이론에서 사건A가 전체 집합 R안에서 정확하게 정의된 점들의 집합이라 하고 $\int_R dp=1$ 이라고 할 때 사건A의 확률값P(A)는 적분형태로 표시하면 다음과 같다.

$$P(A) = \int_A dp \tag{3}$$

또는 다르게 나타내면

$$P(A) = \int_R x_A(x) dp \tag{4}$$

여기서 x는 전체 집합 R안의 한 점을 의미하며 x_A 는 사건A의 특성함수 ($x_A(x) = 0$ 또는 1)이다.

사건A가 애매한 사건일 때 즉, 전체 집합R안의 점들의 퍼지 집합일 때 사건A의 확률값은 다음과 같이 정의된다.

$$P(A) = \int_R f_A(x) dp \tag{5}$$

여기서 f_A 는 사건A의 소속 함수이며 $0 \leq f_A(x) \leq 1$ 의 범위에 존재한다. 이 정의는 식(4)의 일반화이다. 사건A안에서 한점x의 부분적 소속은 P(A)를 결정하는데 고려된다. 비슷하게 확률값 P(A)와 관련있는 퍼지 사건의 평균과 분산은 다음과 같이 표현될 수 있다.

$$V_A = \frac{1}{P(A)} \int_R x f_A(x) dp \tag{6}$$

그리고

$$\sigma_A^2 = \frac{1}{P(A)} \int_R (x - v_A)^2 f_A(x) dp \tag{7}$$

식(6)과 식(7)에 대한 기본적인 생각은 많은 데이터들이 하나의 사건에 속해 있거나 많은 점들이 그 사건의 평균과 분산을 계산하는데 영향을 미친다는 것이다.

이와 같은 방법으로 계산되어 지는 평균과 분산을 퍼지 분산이라고 부른다. 식(6)과 식(7)로부터 discrete 형태의 퍼지 평균과 퍼지 공분산 행렬이 얻어진다.

즉, 퍼지 평균은 식(8)과 같이 표현된다.

$$v_i = \frac{\sum_{k=1}^n f_i(x_k) x_k}{\sum_{k=1}^n f_i(x_k)} \tag{8}$$

여기서 n은 데이터 수, f_i 는 class i의 소속 함수이며 x_k 는 데이터를 나타내는 벡터이다.

퍼지 공분산 행렬은 식(9)와 같이 표현된다.

$$\Sigma_i = \frac{\sum_{k=1}^n f_i(x_k) (x_k - v_i) (x_k - v_i)^T}{\sum_{k=1}^n f_i(x_k)} \tag{9}$$

즉, 퍼지 평균과 퍼지 공분산 행렬은 전형적인 평균과 공분산 행렬의 확장으로 이해될 수 있다. $f_i(x)$ 가 0 또는 1의 값을 가질 때 식(8)과 식(9)는 전형적인 평균과 공분산 행렬의 정의이다.

3. 퍼지 c-means 알고리즘

퍼지 c-means 알고리즘은 식(10)과 같이 일반화된 그룹내의 여러 제품 합을 나타내는 목적함수를 최소화할 수 있도록 데이터 집합을 분할하는 알고리즘으로 Dunn에 의해서 소개되었으며 Bezdek에 의해 일반화된 뒤 순수 이론적 측면뿐 아니라 패턴 인식의 응용 분야에서 많이 이용되고 있다.

$$J_m(U, v) = \sum_{i=1}^c \sum_{k=1}^n (u_{ik})^m \|x_k - v_i\|^2, 1 \leq m < \infty \tag{10}$$

m이 1보다 큰 경우에 모든 i, k에 대해서 $x_k \neq v_i$ 를 만족한다고 가정하면 다음 조건식들을 만족할 때만 (U, v)가 J_m 의 최소화를 가능하게 한다.

모든 i, k에 대해서

$$u_{ik} = \left\{ \sum_{j=1}^c \left(\frac{\|x_k - v_j\|_A}{\|x_k - v_i\|_A} \right)^{2/(m-1)} \right\}^{-1} \tag{11}$$

모든 i에 대해서

$$v_i = \frac{\sum_{k=1}^n (u_{ik})^m x_k}{\sum_{k=1}^n (u_{ik})^m} \tag{12}$$

이 알고리즘은 조건을 나타내는 식(11)과 식(12)을 포함하는 과정을 반복함으로써 J_m 은 어떤 정해진 값에 수렴하게 된다. 즉, 퍼지 c-means 알고리즘은 아래의 1)부터 4)까지의 과정을 반복함으로써 수행된다.

1) 클러스터의 수 c를 $2 \leq c \leq n$ 의 범위 안에서 고정하고 $U^{(0)}$ 를 임의로 초기화 한다. 또한 p는 1로 초기화 한다.

2) $U^{(p-1)}$ 와 식(12)를 이용하여 클러스터의 중심 $\{v_i^{(p)}\}$ 를 계산한다.

3) $\{v_i^{(p)}\}$ 와 식(11)을 이용하여 $U^{(p)}$ 를 다시 개선한다.

4) 일반적인 norm을 이용하여 $U^{(p)}$ 와 $U^{(p-1)}$ 를 비교한다. 만약 $\|U^{(p)} - U^{(p-1)}\| \leq \epsilon$ 를 만족한다면 끝내고 그렇지 않다면 p를 1만큼 증가시키고 2)로 복귀한다.

4. 퍼지 MLE 알고리즘

퍼지 MLE 알고리즘의 소속 함수는 전형적인 평균이나 공분산 대신에 퍼지 평균 v과 퍼지 공분산 Σ 를 사용한 MLE 알고리즘을 이용하여 구한다. 식(13)은 퍼지 MLE 알고리즘의 소속함수를 나타낸다.

$$f_i(x) = \frac{P(x, v_i)}{\sum_{j=1}^c P(x, v_j)} \tag{13}$$

여기서

$$P(x, v_i) = \frac{1}{(2\pi)^{1/N} |\Sigma_i|^{1/N}} \cdot \exp \left[-\frac{1}{2} (x - v_i)^T \Sigma_i^{-1} (x - v_i) \right]$$

이고 N은 데이터를 나타내는 벡터의 차원이다. 또한 c는 미리 정해진 클러스터의 수이며 i의 범위는 $1 \leq i \leq c$ 이다.

그러나 일반적으로는 사용되는 퍼지 MLE 알고리즘은 소속 함수를 정할 때 주어진 데이터가 c개의 클러스터 중 어느 클러스터를 속할지를 미리 결정할 수 있는 이전 확률값 B_i 를 포함한다. 즉, 일반적으로 사용되는 퍼지 MLE 알고리즘은 아래의 1)부터 4)까지의 과정을 반복함으로써 수행된다.

1) 클러스터의 수 c를 $2 \leq c \leq n$ 의 범위 안에서 고정하고 $U^{(0)}$ 를 임의로 초기화한다. 또한 p는 1로 초기화 한다.

2) $U^{(p-1)}$ 를 이용하고 식(4), 식(15) 그리고 식(16)을 계산함으로써 $\{B_i\}, \{v_i\}$, 그리고 $\{\Sigma_i\}$ 를 구한다.

3) 위에서 구한 $\{B_i\}, \{v_i\}$ 그리고 $\{\Sigma_i\}$ 를 이용하여 식(17)에 따라 새로운 $U^{(p)}$ 를 구한다.

4) 일반적인 norm을 이용하여 $U^{(p-1)}$ 와 $U^{(p)}$ 를 비교한다. 만약 $\|U^{(p)} - U^{(p-1)}\| \leq \delta$ 를 만족한다면 끝내고 그렇지 않다면 p를 1만큼 증가시키고 2)로 복귀한다.

모든 i에 대해서

$$B_i = \sum_{k=1}^n u_{ik}(x) / n \tag{14}$$

$$v_i = \sum_{k=1}^n u_{ik}(x) x_k / \sum u_{ik}(x) \tag{15}$$

$$\Sigma_i = \sum_{k=1}^n u_{ik}(x) (x_k - v_i) (x_k - v_i)^T / \sum u_{ik}(x) \tag{16}$$

모든 i, k에 대해서

$$u_{ik}(x) = \frac{P(x_k, v_i)}{\sum_{j=1}^c P(x_k, v_j)} \tag{17}$$

여기서

$$P(x_k, v_i) = \frac{B_i}{|\Sigma_i|^{1/2}} \cdot \exp \left[-\frac{1}{2} (x_k - v_i)^T \Sigma_i^{-1} (x_k - v_i) \right]$$

이고 c는 미리 정해진 클러스터의 수이며 i의 범위는 $1 \leq i \leq c$ 이다.

IV. 확장된 퍼지 Clustering 알고리즘

Clustering 알고리즘은 주어진 데이터를 어느 정도의 동질성을 갖는 몇 개의 클러스터로 나눈다. 전통적인 clustering 알고리즘은 클러스터들간 경계가 명확하다는 가정하에서 각각의 데이터를 오직 하나의 클러스터에 할당한다. 이 경우 클러스터간 경계가 애매하고 하나의 데이터가 하나의 클러스터에 속한다고 명확하게 말할 수 없는 실제 데이터의 특징을 반영하지 못한다.

이러한 문제를 해결하기 위하여 기존의 clustering 알고리즘을 퍼지 집합론에 근거를 두고 확장한 퍼지 clustering 알고리즘이 제안되어 기존의 clustering 알고리즘에 비해 인간의 사고에 가깝게 주어진 데이터를 분류하고 있다. 또한 최근에는 최적 퍼지 clustering 알고리즘이 제안되는 등 퍼지 clustering 알고리즘의 성능을 향상시키기 위한 노력이 계속되어 지고 있다.

본 연구에서는 이에 대한 노력의 일환으로 주어진 데이터의 분류에 널리 사용되고 있는 기존의 퍼지 clustering 알고리즘을 확장하여 분류하고자 하는 데이터의 성질을 고려할 수 있는 새로운 알고리즘을 제안한다.

이 알고리즘은 데이터의 성질에 따라 알고리즘내에서 가중치를 주어 각 클러스터의 중심값이 가중치가 높은 위치로 이동하는 특징을 갖도록 하였으며 이때 중심값은 데이터의 성질을 고려한 각 클러스터의 대표값의 의미를 갖는다. 물론, 데이터의 성질이 모두 똑같은 경우 가중치가 단위 크기를 갖으며 이 경우는 기존의 퍼지 clustering 알고리즘과 같은 데이터 분할 효과를 나타낸다. 또한 이 알고리즘의 성능을 향상시키기 위해 퍼지 clustering 알고리즘의 효율성에 중요한 영향을 미치는 초기 중심값 설정을 위한 새로운 알고리즘을 포함시켰으며 최적의 클러스터의 수를 정하기 위해 기존의 clustering 알고리즘에서 사용하는 평가함수에 가중치를 적용하여 사용하였다. 물론 기존의 퍼지 clustering 알고리즘의 확장으로 인한 소속함수 크기의 증가는 없으며 계산량의 증가는 무시할 정도이다.

한편 확장된 퍼지 clustering 알고리즘의 각 단계는 다음과 같다.

- 1) 시작점 결정 알고리즘에 의해 각각 클러스터의 중심값을 가정한다.
- 2) 이 초기값을 확장된 퍼지 c-means 알고리즘에 이용함으로써 각 클러스터 중심을 정한다.
- 3) 이 때 구해진 소속 함수에 확장된 퍼지 MLE 알고리즘을 적용함으로써 각 클러스터를 최적화 한다.
- 4) 최적화된 정도에 대한 성능 평가를 한다.
- 5) 클러스터 수를 1만큼 증가시키고 주어진 데이터가 최적으로 분류될 때까지 위 과정을 반복한다.

1. 시작점 설정 알고리즘

퍼지 clustering 알고리즘과 마찬가지로 확장된 퍼지 clustering 알고리즘도 각 클러스터의 중심값을 어떻게 예측하는가에 따라 데이터는 다르게 분류된

다. 많은 실질적인 상황에서 각 클러스터의 중심값을 예측하기는 불가능하며 자동적으로 각 클러스터의 중심값을 예측하는 것이 필요하다.

본 논문에서 제안하고자 하는 시작점 결정 알고리즘의 기본적인 생각은 주어진 데이터가 k개로 분류되었을 때 각 클러스터가 multivariate 정규 분포를 갖는다는 가정하에 이 클러스터내에 포함된 데이터 수에 대한 일정 확률 밀도 경계에 포함된 데이터의 비율이 가장 낮은 클러스터내에 (k+1)번째 클러스터의 중심이 존재한다는 것이다. 따라서 이 클러스터를 두개로 분할하여 새로운 중심값으로 사용한다.

한편 일정 확률 밀도 경계는 주어진 데이터 X가 평균이 v고 분산이 Σ인 정규 분포를 따른다고 할 때

(a) $(x-v)^T \Sigma^{-1} (x-v)$ 는 자유도가 p인 chi-square 분포를 따르며

(b) 평균으로부터 $\{x: (x-v)^T \Sigma^{-1} (x-v) \leq \chi_p^2(\alpha)\}$ 의 거리안에 $1-\alpha$ 만큼의 데이터를 포함한다.

다음은 시작점 결정 알고리즘의 각 단계이다.

- 1) 주어진 데이터의 성질을 고려하여 전체 평균과 공분산을 구한다.
- 2) 각 클러스터 중 비율이 가장 낮은 클러스터를 선택한 후 이 클러스터를 두개로 분할하여 새로운 중심값으로 사용한다.
- 3) 확장된 퍼지 c-means 알고리즘과 확장된 퍼지 MLE 알고리즘을 차례로 적용하여 최적의 클러스터를 구한다.
- 4) 클러스터의 수가 주어진 최대 클러스터 수보다 작으면 단계2)로 돌아가고 그렇지 않으면 멈춘다.

2. 확장된 퍼지 c-means 알고리즘

확장된 퍼지 c-means 알고리즘은 기존의 퍼지 c-means 알고리즘이 사용하는 목적함수를 각 데이터의 성질을 고려할 수 있도록 확장시켰으며 이 때 사용하는 목적함수는 식(18)과 같다.

$$J_{wm}(U, v) = \sum_{i=1}^c \sum_{k=1}^n \text{가중치}(k) \times (u_{ik})^m \|x_k - v_i\|^2, \quad 1 \leq m < \infty \tag{18}$$

m이 1보다 큰 경우에 모든 i, k에 대해서 $x_k \neq v_i$ 을 만족한다고 가정하면 다음 조건식들을 만족할 때만 (U, v)가 J_{wm} 의 최소를 가능하게 한다.

모든 i, k에 대해서

$$u_{ik} = \left[\sum_{j=1}^c \left[\frac{\|x_k - v_j\|_A}{\|x_k - v_i\|_A} \right]^{2/m-1} \right]^{-1} \tag{19}$$

모든 i 에 대해서

$$v_i = \frac{\sum_{k=1}^n \text{가중치}(k) \times (u_{ik})^m x_k}{\sum_{k=1}^n \text{가중치}(k) \times (u_{ik})^m} \quad (20)$$

- 1) 모든 데이터의 성질에 따라 가중치를 준다.
- 2) 클러스터의 수 c 를 $2 \leq c \leq n$ 의 범위안에서 고정하고 $\{v_i\}$ 를 시작점 결정 알고리즘에 따라 초기화한다.
- 3) $\{v_i\}$ 를 이용하고 식(19)를 계산함으로써 $U^{(0)}$ 를 초기화 한다. 또한 p 는 1로 초기화 한다.
- 4) $U^{(p-1)}$ 를 이용하고 식(20)를 계산함으로써 $\{v_i\}$ 를 다시 개선한다.
- 5) $\{v_i\}$ 를 이용하고 식(19)를 계산함으로써 $U^{(p)}$ 를 다시 개선한다.
- 6) 일반적인 norm를 이용하여 $U^{(p)}$ 와 $U^{(p-1)}$ 를 비교한다. 만약 $\|U^{(p)} - U^{(p-1)}\| \leq \epsilon$ 를 만족한다면 끝내고 그렇지 않다면 p 를 1만큼 증가시키고 2)로 복귀한다.

3. 확장된 퍼지 MLE 알고리즘

확장된 퍼지 MLE 알고리즘은 기존의 퍼지 MLE 알고리즘에서 사용하는 퍼지 평균과 퍼지 공분산 대신 각 데이터의 성질을 고려한 퍼지 평균과 퍼지 공분산을 사용한 알고리즘이다.

- 1) 모든 데이터의 성질에 따라 가중치를 준다.
- 2) 클러스터의 수 c 를 $2 \leq c \leq n$ 의 범위안에서 고정하고 확장된 퍼지 c-means 알고리즘의 최종적인 소속 함수를 이용하여 $U^{(0)}$ 를 초기화 한다. 또한 p 는 1로 초기화 한다.
- 3) $U^{(p-1)}$ 를 이용하고 식(21), 식(22) 그리고 식(23)을 계산함으로써 $\{B_i\}$, $\{v_i\}$ 그리고 $\{\Sigma_i\}$ 를 구한다.

모든 i 에 대해서

$$B_i = \sum_{k=1}^n \text{가중치}(k) \times u_{ik}(x) \times x_k / \Sigma \quad (21)$$

$$v_i = \frac{\sum_{k=1}^n \text{가중치}(k) \times u_{ik}(x) \times x_k / \sum_{k=1}^n \text{가중치}(k) \times u_{ik}(x)^m}{\sum_{k=1}^n \text{가중치}(k) \times u_{ik}(x)^m} \quad (22)$$

$$\Sigma_i = \frac{\sum_{k=1}^n \text{가중치}(k) \times u_{ik}(x) \times (x_k - v_i)^T / \sum_{k=1}^n \text{가중치}(k) \times u_{ik}(x)}{\sum_{k=1}^n \text{가중치}(k) \times u_{ik}(x)} \quad (23)$$

- 4) 위에서 구한 $\{B_i\}$, $\{v_i\}$ 그리고 $\{\Sigma_i\}$ 를 이용하고 식(24)를 계산함으로써 새로운 $U^{(p)}$ 를 구한다.

모든 i, k 에 대해서

$$u_{ik}(x) = \frac{P(x_k, v_i)}{\sum_{j=1}^c P(x_k, v_j)} \quad (24)$$

여기서

$$P(x_k, v_i) = \frac{B_i}{|\Sigma_i|^{1/2}} \cdot \exp\left[-\frac{1}{2}(x_k - v_i)^T \Sigma_i^{-1}(x_k - v_i)\right]$$

이고 c 는 미리 정해진 클러스터의 수이며 i 의 범위는 $1 \leq i \leq c$ 이다.

- 5) 일반적인 norm을 이용하여 $U^{(p-1)}$ 와 $U^{(p)}$ 를 비교한다. 만약 $\|U^{(p)} - U^{(p-1)}\| \leq \delta$ 를 만족한다면 끝내고 그렇지 않다면 p 를 1만큼 증가시키고 3)으로 복귀한다.

4. 주어진 데이터의 클러스터 수 결정문제

일반적으로 clustering 알고리즘은 클러스터의 수를 알고 있다는 가정하에서 주어진 데이터를 분할한다. 그러나 실제에 있어서 본질적으로 알려지지 않은 데이터 집합을 분류하는데 있어서 이러한 가정은 매우 부자연스러운 것이다. 따라서 clustering 문제를 다룰 때 항상 발생하는 중요한 문제는 얼마나 많은 클러스터가 존재하는지 결정하는 일이며 unsupervised clustering 알고리즘에서 필수적이다.

이 문제를 해결하기 위한 많은 방법들이 제안되고 있으나 본 논문에서는 클러스터의 갯수를 점차 증가시키면서 클러스터 평가함수의 값의 변화를 조사하는 방법을 선택하였다.

이 방법은 우선 주어진 클러스터 갯수에 대하여 목적함수가 최소가 되도록 주어진 데이터를 분할하고 분할된 결과에 대해 평가 함수를 적용한다. 그리고 이 과정을 클러스터의 갯수를 증가시켜 가면서 평가함수 값의 변화를 조사한다.

본 논문에서는 사용한 에러 제곱합의 평가 함수 J 는 c 가 증가함에 따라 단조적으로 감소한다. 그러나 만약 n 개의 데이터가 실제로 c 개의 잘 분할된 클러스터로 구성되어 있다면 클러스터의 갯수가 c 개일 때까지 J 는 빠르게 감소하다가 클러스터의 갯수가 c 개 이후부터는 매우 천천히 감소한다. 마지막에는 클러스터의 갯수가 n 일때 J 값은 0이 된다.

식(25)는 기존의 퍼지 clustering 알고리즘을 위한 에러 제곱합의 평가함수이며 식(26)은 확장된 퍼지 clustering 알고리즘을 위한 평가 함수이다.

$$J_e = \sum_{i=1}^c \sum_{j=1}^n u_{ij} \times \|x - m_i\|^2 \quad (25)$$

$$J_{we} = \sum_{j=1}^c \sum_{i=1}^n \text{가중치}(j) \times u_{ij} \|x_j - m_i\|^2 \quad (26)$$

V. 확장된 퍼지 Clustering 알고리즘을 이용한 자동 목표물 탐색

자동 목표물 탐색에 입력 영상의 밝기 정도를 데이터의 성질로 이용한 확장된 clustering 알고리즘을 적용할 경우 이 알고리즘은 목표물의 절대적 밝기에 대해 상대적으로 밝은 부분을 목표물로 검출하며 목표물간 경계의 모호함과 목표물과 배경간 경계의 모호함을 해결한다.

그림6과 그림7은 확장된 퍼지 clustering 알고리즘을 이용한 자동 목표물 탐색의 효과를 입증하기 위해 사용한 입력 영상들로 0부터 255까지의 밝기 정도를 갖는다. 이들 입력 영상내 목표물을 모두 주변에 비해 밝다는 특성을 가지고 있으며 특히 그림7의 입력 영상내 목표물들의 주위 잡음과 밝기 정도는 비슷하지만 상대적으로 더 밝은 화소들이 뭉쳐 있기 때문에 목표물로서의 의미를 갖는다.

그림8과 그림9는 그림6과 그림7의 입력 영상에 대한 최적의 목표물 수 결정을 위한 평가함수 적용 결과이다. 이때 사용한 평가 함수는 식(26)의 J_{we} 이다.

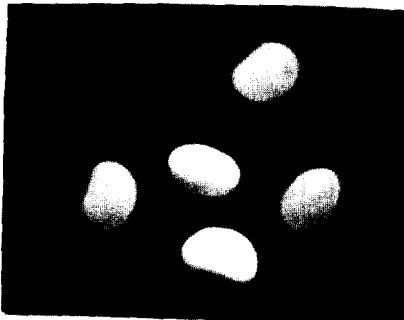


그림 6. 입력 영상(1)
Fig. 6. Input image(1).

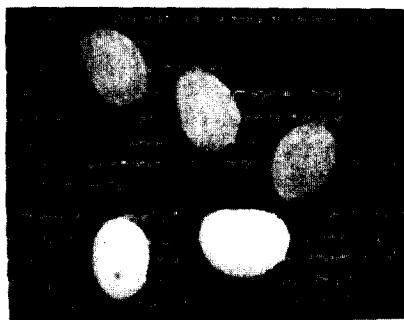


그림 7. 입력 영상(2)
Fig. 7. Input image(2).

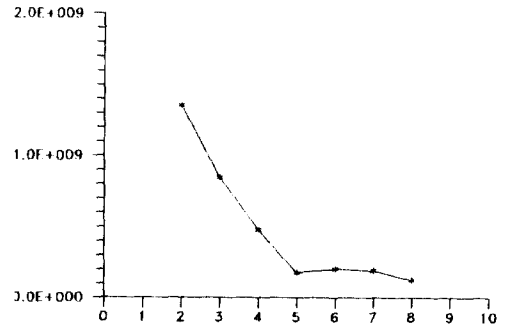


그림 8. 입력 영상 (1)에 대한 최적의 목표물 수
Fig. 8. Number of targets in input image(1).

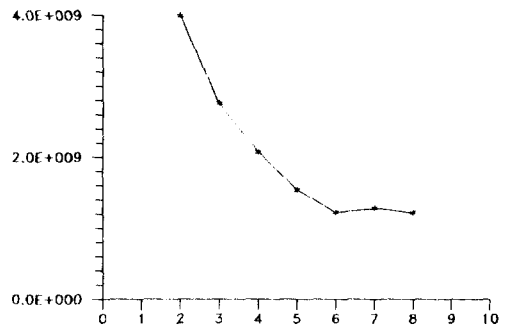


그림 9. 입력 영상 (2)에 대한 최적의 목표물 수
Fig. 9. Number of targets in input image (2).

평가 함수 결과 그림6과 그림7의 입력 영상내 최적의 목표물수는 모두 5개씩임을 나타내고 있다.

그림10과 그림11은 그림6과 그림7의 입력 영상에 대한 자동 목표물 탐색 결과이다. 이때 각 폐곡선은 목표물이 존재할 가능성이 있는 영역을 나타내고 있다.

VI. 결 론

본 논문에서는 특정 공간에서 데이터 분류에 널리 사용되고 있는 기존의 퍼지 clustering 알고리즘을 확장하여 분류하고자 하는 데이터의 성질을 고려할 수 있는 새로운 알고리즘을 제안했다.

또한 입력영상내 데이터들의 밝기 정도를 이 알고리즘내에서 가중치로 사용하여 자동 목표물 검출에 응용할 경우 입력 영상내의 목표물을 분할하는데 있어 절대적인 밝기에 영향받지 않고 배경에 비해 상

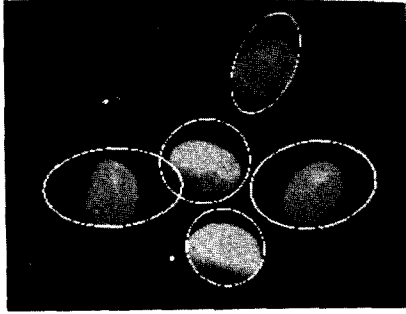


그림10. 입력 영상 (1)에 대한 목표물 탐색 결과
Fig. 10. Result of detection in input image (1).

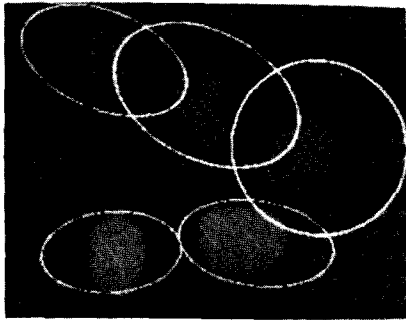


그림11. 입력 영상 (2)에 대한 목표물 탐색 결과
Fig. 11. Result of detection in input image (2).

대적으로 밝은 부분이 목표물이 있을 가능성이 높은 부분으로 인식했다. 즉, 목표물을 인식하는데 있어서 일반적으로 사용하는 잡음제거와 가능성 있는 목표물의 밝기의 개선과 같은 전처리 단계를 필요로 하지 않으면서 주위 배경의 가중치에 비해 목표물에 가중치를 더 줌으로써 목표물 검출에 있어서 주위 배경에 대한 상대적 밝기에 따라 목표물을 적응적으로 인식했다. 뿐만 아니라 퍼지 집합의 개념을 사용하므로써 목표물간 경계의 모호함과 목표물과 배경간의 경계의 모호함을 해결했다.

參 考 文 獻

- [1] R.V. Duda and P.E. Hart, *Pattern Classification and Scene Analysis*, Wiley-Interscience, 1973.
- [2] James C. Bezdek and Joseph C. Dunn, "Optimal fuzzy bartition: A heuristic for estimating the parameters in a mixture of normal distributions," *IEEE Trans. Computers*, August 1975.
- [3] Guy B. Coleman and Harry C. Andrews, "Image segmentation by clustering," *Proc. IEEE*, vol. 67, no. 5, May 1979.
- [4] Dana H. Ballard, Christopher M. Brown, *Computer Vision*, Prentice Hall 1982.
- [5] Michael P. Windhan, "Cluster validity for the fuzzy c-means clustering algorithm," *IEEE Trans. PAMI*, vol. 4, no. 4, July 1982.
- [6] L. Sevigny, M. Bohner, E. Ostevoid and S. Grinaker, J. Dehne, "Discrimination and classification of vehicles in natural scenes from thermal imagery," *CGIP* 24, 229-243, 1983.
- [7] Lawrence L'Gorman and Arthur C. Sander-son, "The converging squares algorithm : An efficient method for locating peaks in multidimensions," *IEEE Trans. PAMI*, vol. no. 3, May 1984.
- [8] T.L. Huntsberger, C.L. Jacobs and R.L. Cannon, "Iterative fuzzy image segmentation," *Pattern Recognition*, vol. 2, no. 2, pp. 131-138, 1985.
- [9] Terano, Asai, Sugeno 지음, 박민용, 최항식 역, pp. 139 - 158, 대영사, 1985
- [10] Bir Bhanu, "Automatic target, recorgnition: state of the art survey," *IEEE Trans. Aerospace and Electronics Systems*, vol. 22, no. 4, July 1986.
- [11] Mohan M. Trivedi and James C. Bezdek, "Low-level of aerial images with fuzzy clustering," *IEEE Trans. PAMI*, vl. 16, no. 4, July/August 1986.
- [12] O.A. Zunga and R.M. Haralick, "Integrated Directional derivative gradient operator," *IEEE Trans. SMC*, vol. 7, no. 3, pp. 508-517, May/June 1987.
- [13] James M. Keller, Gregory Hobson, John Wootton , Asghar Nafarieh, and Kent Luetkemeyer, "Fzzy confidence measures in middle vision," *IEEE Trans. SMC*, vol. 17, no. 4, July/August 1987.
- [14] Richard A. Johnson and Dean W. Wichern, *Applied Multivariate Statistical Analysis*, Prentice-Hall, 1988.
- [15] J.K. Kent and K.V. Mardia, "Spatail classification using fuzzy membership models," *IEEE Trans. PAMI*, vol. 10, no. 5, September 1988.
- [16] I. Gath and A.B. Geva, "Unsupervised optimal clustering," *IEEE Trans. PAMI*,

vol. 11, no. 7, July 1989.

[17] Fangju Wang, "Fuzzy supervised classification of Remote sensing images," *IEEE Trans. Geoscience and remote sensing*, vol.

28, no. 2, March 1990.

[18] Stephen P. Bonks, *Signal processing Image Processing and Pattern Recognition*, Prentice-Hall, 1990.

著 者 紹 介



姜 景 辰 (正會員)

1962年 10月 31日生. 1985年 2月 고려대학교 전자공학과 졸업. 1990年 8月 고려대학교 산업과학 대학원 전자통신공학과 석사 학위 취득. 1984年 12月 ~ 현재 금성사 가전 연구소 주임 연구원

1991年 3月 ~ 현재 고려대학교 대학원 전자공학과 박사과정 재학중. 주관심분야는 디지털 이미지 프로세싱, HDTV, 컴퓨터 아키텍처 등임.

金 秀 桓 (準會員) 第28卷 B編 第5號 參照

현재 고려대학교 대학원 전자공학과 석사과정 재학중.

李 太 遠 (正會員) 第25卷 第2號 參照

현재 고려대학교 전자공학과 교수