

A Gradient DNA Coding for Searching Stable Solution

廉鶴松* · 金仁澤**
(HeSong Lian · Intaek Kim)

Abstract - This paper presents a novel method for searching stable solution using DNA coding scheme. Often there are more than one solutions that satisfy the system requirements. These solutions can be viewed as extremes in multimodal function. All extremes are not the same in that some of them are using sensitive to noise or perturbation. This paper address the method that selects a solution that meets the system requirements in terms of output performance and is tolerant to the perspective noise or perturbation. A new method, called a Gradient DNA coding is proposed to achieve such objectives several numerical examples and presented and comparing DNA coding with genetic algorithm is also given.

Key Words : DNA 코딩, 안정해 탐색, 유전자 알고리즘, 다봉함수

1. 서 론

최적화 도구인 유전자 알고리즘은 어떤 다봉함수(multimodal function)에서 전역해(global solution)를 찾는 문제에 많이 이용되었다[10],[11]. 다봉함수 문제는 주어진 성능을 만족하는 해가 여러 개 존재할 가능성이 있다. 따라서 이러한 문제는 굳이 어렵게 전역해를 구할 필요 없이 주어진 성능만을 만족하는 지역해(local solution)를 구해도 된다. 문제의 성능을 만족하는 해를 선택할 때 몇 가지 조건을 고려하게 된다. 예를 들면, 어떤 시스템을 설계할 때, 주어진 사양에 알맞은 해가 안정적인 영역에서 동작하는 것이 매우 중요하다. 안정해(stable solution)란 어떤 최적화문제에서 해를 선택할 때 어떤 파라미터 값이 작은 변화를 가져오더라도, 즉 해가 변하더라도, 성능의 변화가 최소화되는 것이 바람직하다. 예를 들어 어떤 성능을 출력하는 제어기가 있다고 가정하자. 이 출력을 발생할 수 있는 여러 개의 해 집합 중, 해의 변화에 대하여 출력의 변화가 최소화되는 해를 선택하여야 한다. 이와 같은 이유는 이 제어기를 사용할 때 선택되었던 해가 실제로 구현되고 나면 어느 정도의 범위 안에서 변화가 생길 수 있기 때문에 이 변화에 대한 출력의 변화는 가급적 둔감한 것이 요구된다.

그림 1은 전술한 상황을 간단하게 설명하여 준다. 여기서 x 는 함수의 영역을 나타내며, y 는 목적 함수이다. 해결하고자 하는 문제는 특정한 값(0.6) 이상의 출력을 발생시키는 x 를 찾는 것이다. 그림 1에서도 명확하듯이 point 1은 전역해이나,

x 의 미세한 변화에 대하여 시스템의 성능(y)이 상대적으로 급격한 변화를 보여준다. 따라서 point 1 또는 point 2는 최적의 성능을 발생하는 해이나, 그 해의 변화에 대하여 성능의 변화가 크기 때문에 바람직한 해가 아니다. 한편 point 3은 x 가 주어진 범위 내(Δx)에서 약간 변하더라도 Δy 가 허용범위에 들어갈 가능성이 많음을 쉽게 알 수 있다. 더욱이 point 3에서는 Δx 가 0에 수렴할수록 $\Delta y / \Delta x \approx 0$ 이 됨을 알 수 있다. point 4와 point 5는 해당 시스템의 성능(y)이 너무 적어 원하는 해가 아니다. 또한 $\Delta y / \Delta x$ 값도 크므로 제외한다. 즉 그림 1에서 주위 환경에 민감하지 않는 안정해 point 3이 가장 적당하다고 볼 수 있다.

자연계 생물들의 유전원리를 적용한 유전자 알고리즘을 이용하여 많은 수의 개체를 발생시켜 선택, 교배, 돌연변이, 재생성을 진행하여 조건에 적합한 해를 찾는다. 그림 1과 같은 문제에서 유전자알고리즘을 적용하면 일반적으로 전역해에 해당하는 point 1로 해가 수렴되고 집중한다. 그러나, point 1은 주위환경(x)에 너무 민감하므로 상대적으로 안정해를 가지는 point 3이 주어진 문제의 해가 되도록 알고리즘을 구현하는데 있어 제약 조건을 사용해야 할 필요성이 대두된다.

Niching 알고리즘은 유전자 알고리즘의 기본적인 틀 안에서 해 집단을 여러 개의 최적점들(optima)에 분산시키고자 노력한다[3],[8]. 이러한 방법에는 crowding 또는 preselection 과 같은 방법이 사용되기도 하였으나 최근에는 sharing으로 알려진 방법을 주로 사용하게 된다. 이 sharing을 사용하면 그림 1의 point 1에만 해가 몰려있는 것을 방지할 수 있는데, 이는 적합도를 계산할 때 sharing factor를 고려하여 적합도를 재조정하기 때문이다[5]. 그러나 sharing에 의한 효과는 해가 point 1부터 point 5까지에 모든 극대점에 분포하기 때문에 문제에서 요구하는 point 3에만 국한시키는데는 좋은 효과를 얻지 못한다. 따라서 본 연구의 궁극적인 목적은 해 값의 미세한 변화에도 안정적인 성능을 가진 point 3과 같은 점

* 廉鶴松 : 明知大學校 情報工學科 碩士

** 金仁澤 : 明知大學校 情報工學科 副教授 · 工博
接受日字 : 2002年 11月 18日
最終完了 : 2003年 1月 3日

을 찾는 것이다.

본 논문에서는 생물학적인 DNA와 인위적인 DNA의 개발 메커니즘에 근거를 둔 DNA 코딩 방법[1],[2]을 이용하여 원하는 안정해를 찾는 방법을 제안하고자 한다. 이 방법은 인위적인 DNA 코딩에 기술기 정보를 이용하기 때문에 기술기를 적용한 DNA 코딩(Gradient DNA coding)이라고 정의하자. 이 방법의 효율성을 증명하기 위해 같은 조건 하에서, 간단한 유전자 알고리즘과 제안된 알고리즘을 다봉함수에 적용하여 안정해를 찾고 그 결과를 비교 분석하고자 한다. 본 논문의 구성은 다음과 같다. 다음 장에서는 인위적인 DNA 코딩 방법을 설명한다. 제 3장에서는 본 논문에서 제안된 알고리즘을 언급하고, 제 4장에서는 시뮬레이션 결과에 대하여 설명하고 분석을 한다. 마지막 장은 본 연구의 결론과 앞으로의 연구 과제를 설명한다.

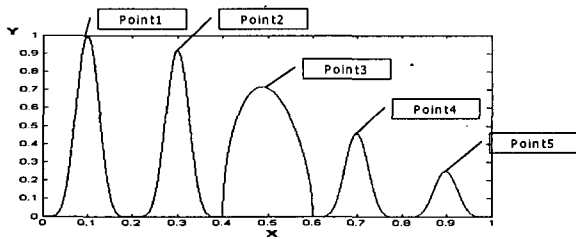


그림 1 다봉함수
Fig. 1 multimodal function

2. 인위적인 DNA 코딩 방법

본 논문에서는 해의 길이가 가변적이고 한 개 염색체 안에 여러 개 해가 들어있는 생물학적인 DNA와 유전자 알고리즘의 선택, 교배, 돌연변이 등 연산을 결합한 DNA 코딩방법을 사용한다. 이 방법은 염기들의 중복성과 그림 2에서 볼 수 있는 해들의 병렬성을 이용하여 주어진 공간을 더욱 효율적으로 탐색할 수 있다.

우리는 염기, 당 및 인산이 결합된 것을 뉴클레오타이드(nucleotides)라 하고 다수의 뉴클레오타이드가 결합되어 생성된 고분자 화합물을 핵산(nucleic acid)이라 한다. 핵산은 라이보 핵산(RNA)과 디옥시라이보 핵산(DNA) 두 가지가 있으며 이 중 DNA는 염색체에만 존재한다[1],[2],[13]. 모든 생물체는 각자 고유의 DNA를 가지고 있다. DNA는 개체의 특성을 발현하는 유전코드이다. DNA는 2중 나선의 구조를 가지고 있으며 A(Adenine), T(Thymine), G(Guanine), C(Cytosine)의 염기간의 수소 결합으로 연결되어 있다. 이들 중 세 염기의 연속된 배열을 한 의미 단위로 해석하는데, 생물학적 용어로 코돈(codon, 유전 정보의 최소단위)이라 부른다. DNA의 차이는 바로 A, T, G, C의 4종류의 염기배열 차이이며 총 4³=64 종류의 코돈은 20종류의 아미노산이 된다(표 1 참조)[6]. 매개의 아미노산은 각자의 의미를 가지며 유전자의 의미는 바로 이 아미노산의 결합에 의하여 결정된다. 모든 생물의 유전자가 DNA라는 점에서 유전자의 특징은 4종의 염기 배열의 방법이라고 할 수 있다. 따라서 DNA가 가진 유전정보(유전암호)란 DNA가 지닌 염기배열을 의미한다.

생물학에서 DNA는 RNA로 전사(transcription)되어 리보

솜(ribosome)에서 단백질로 번역(translation)된다. 즉 DNA의 배열에 따라 아미노산의 합성순서를 결정하여 여러 종류의 단백질을 만들어낸다. RNA가 단백질로 번역될 때 코돈 ATG에서 시작하여 코돈 TAG, TAA혹은 TGA에서 끝난다[13]. 시작 코돈에서 끝 코돈 사이 부분을 하나의 GENE라고 한다. 이는 코돈 ATG에서 시작하여 정해진 7개의 코돈을 이용한 알고리즘[6]과는 달리 끝 코돈도 찾아주어 GENE의 길이의 가변성을 충분히 이용하였다. 따라서 GENE은 몇 개의 코돈으로 구성되며 각 코돈은 하나의 아미노산에 대응된다(표 1 참조).

표 1 RNA(DNA) 코돈과 생성되는 아미노산[12]
Table 1 RNA(DNA) Codon and Amino Acid

코돈	아미노산	코돈	아미노산
UUU	Phe	UAU	Tyr
UUC		UAC	정지
UUA		UAA	
UUG	UAG		
CUU	Leu	CAU	His
CUC		CAC	Gln
CUA		CAA	
CUG		CAG	
AUU	Ile	AAU	Asn
AUC		AAC	Lys
AUA		AAA	
AUG	Met	AAG	
GUU	Val	GAU	
GUC		GAC	
GUA		GAA	Glu
GUG		GAG	
UCU	Ser	UGU	Cys
UCC		UGC	정지
UCA		UGA	
UCG		UGG	
CCU	Pro	CGU	Arg
CCC		CGC	
CCA		CGA	
CCG		CGG	
ACU	Thr	AGU	Ser
ACC		AGC	
ACA		AGA	Arg
ACG		AGG	
GCU	Ala	GGU	Gly
GCC		GGC	
GCA		GGA	
GCG		GGG	

표 2 아미노산 값
Table 2 Value assigned to each amino acid

Phe	0.25	Pro	0.50	His	0.75	Glu	1.00
Leu	1.25	Thr	1.50	Gln	1.75	Cys	2.00
Ile	2.25	Ala	2.50	Ans	2.75	Trp	3.00
Met	3.25	Tyr	3.50	Lys	3.75	Arg	4.00
Ser	4.25	Val	4.50	Asp	4.75	Gly	5.00

그림 2는 염색체의 변환과 매커니즘을 간단하게 보여준다. 그림에서 GENE1은 CGG, CAT, GAA 등 8개의 코돈으로 구성되어 이들은 각각 Arg, His, Glu 등 8개의 아미노산에 대응된다. GENE2도 마찬가지로 몇 개의 코돈으로 구성되며, 이 코돈 또한 몇 가지 종류의 아미노산에 대응된다. 그림 2에서 GENE1과 GENE2가 중복되어 나타나있음을 보여준다. 이러한 GENE의 중복 표현이 인위적인 DNA 코딩에서 가장 큰 장점중의 하나이며 유전자 알고리즘에 비해 다양한 표현이 가능함을 보여준다. 본 논문에서는 각각의 코돈에 대응하는 아미노산들이 문제를 해결하는데 맡은 역할이 상이하므로 경험적인 방법을 이용하여 아미노산에 표 2와 같이 값을 부여하였다. 또한 각 GENE의 값을 대응되는 각 아미노산에 부여한 값들의 합으로 정한 후, 이를 함수의 정의역에 mapping을 하여 시뮬레이션 하였다.

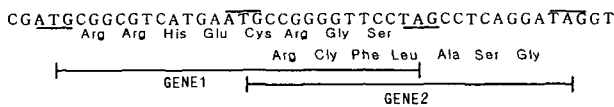


그림 2 GENE의 중복과 염색체의 변환
Fig. 2 Example of Genes overlapping and Chromosome Translation Mechanism

그림 3(a)은 교배의 예를 보여준다. 교배는 주어진 교배 확률에 의해 발생한다. 염색체의 길이를 600으로 하였을 때 각 염색체에 평균 10개 정도의 해를 가지고 있는데 본 논문에서는 좋은 해를 잃어버리는 단점을 보완하기 위해 각 염색체에서 적합도가 제일 큰 해를 살리는 엘리트 전략(elitist strategy)을 적용하였다. 그림 3(a)의 교배과정은 설명하기 위해, 첫 번째의 염색체를 A|B|C라고 하고 두 번째의 것을 abc라고 하자. 여기서 “|”는 교배점(crossover point)이다. 첫 번째 염색체에서 제일 좋은 해 Best GENE1(B)과 두 번째 염색체에서 제일 좋은 해 Best GENE5(b)를 모두 살려 교배했다. 교배의 결과는 그림 3(a)의 아래 그림과 같이 bAC와 Bac의 값을 가진 염색체를 발생시킨다. 한 세대에 약 600여 개 해가(집단크기 60) 존재하는데 교배확률에 따라 교배를 진행할 때 각 염색체에서 적합도가 제일 큰 해(GENE)만 그대로 남고 나머지 해들은 모두 변화를 가져올 수 있다. 변화된 해들 중에서 적합도 값이 큰 해(GENE)는 다음 세대에 살아남게 된다. 그림 3(b)은 유전자 알고리즘에서의 돌연변이의 결과를 예로 보여준다. 원래의 염색체에서 C가 G로 바뀌에 따라 많은 GENE의 개수와 그 값들이 변화한다. 본 논문에서는 염색체 내에 있는 A, T, G, C를 돌연변이의 확률에 따라 바뀌도록 하여 돌연변이를 진행하였다[6],[9].

따라서 인위적인 DNA 코딩 방법은 다음의 특징을 가지고 있다.

- (a) 한 염색체내에 다수의 해(GENE)가 존재하며 주어진 영역에 가능한 해를 충분히 잘 표현할 수 있다.
- (b) 코딩에 중복이 있다. 하나의 아미노산에 여러 종류의 코돈이 있다.
- (c) 해(GENE)의 길이가 가변적이다. 시작과 끝의 코돈이 정해지지 않았기에 해(GENE)의 길이는 항상 가변적이다.

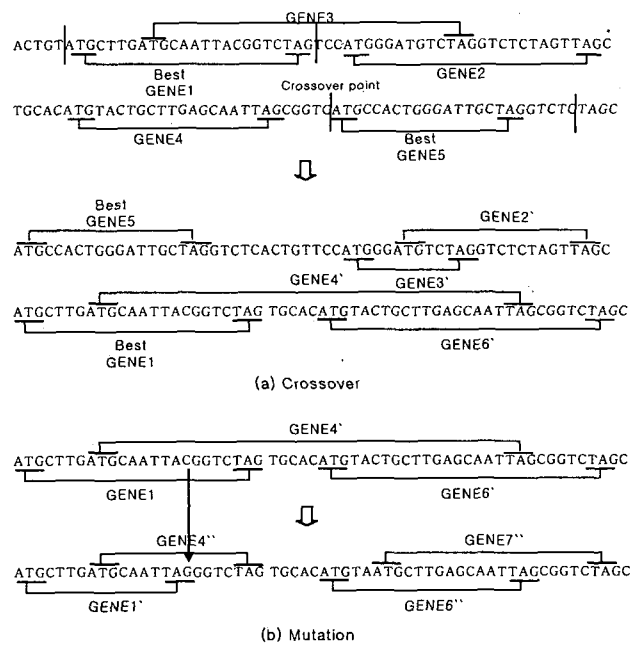


그림 3 교배와 돌연변이의 예
Fig. 3 Example of Crossover and Mutation

3. 알고리즘

주어진 문제를 풀기 위해 본 논문에서 제안하는 알고리즘은 아래와 같다.

단계1: 초기화

개체집단의 크기, 세대수, 염색체의 길이, 교배확률, 돌연변이 확률 등을 결정한다. 초기화 단계는 유전자의 일반적인 0과 1을 사용하는 2진수 초기화와는 달리 4가지 염기인 A(Adenine), T(Thymine), G(Guanine), C(Cytosine)로 초기화한다. 즉 4진수 초기화를 한다.

단계2: 잡음(noise)생성

원래 함수내의 파라미터에 추가 할 잡음을 random하게 생성을 하였다. 집단의 크기를 n 으로 초기화했다면 매 세대의 파라미터는 $X=(x_1, x_2, x_3, \dots, x_n)$ 로, 생성한 잡음을 $\Delta=(\delta_1, \delta_2, \delta_3, \dots, \delta_n)$ 으로 표시할 수 있다. 본 논문에서는 $X+\Delta$ 혹은 $X-\Delta$ 를 원 파라미터에 잡음을 추가하였다고 한다. 위에서 생성한 잡음 Δ 는 다음 단계에서 사용된다. 실제 문제에 적용할 때는 적용되는 문제에 따라 강도 변수를 사용하여 잡음의 크기를 조정하여야 한다. 본 논문에서는 강도 변수를 0.02로 사용하였다.

단계3: 기울기 평균 계산

각 염색체내의 매 GENE에 대응되는 매 점과 잡음 추가 후의 대응되는 점의 기울기를 계산한다. 같은 방법으로 잡음을 감한 후의 기울기를 계산하여 두 기울기 절대값의 평균을 구한다. 즉 함수에서 x_i 에 대응되는 한 점과 $x_i + \delta_i$ 에 대응되는 한 점을 이용하여 기울기 $S(x_i, x_i + \delta_i)$ 를 구하고 같은 방법으로 x_i 에 대응되는 점과 $x_i - \delta_i$ 에 대응되는 점을 이용하여 기울기 $S(x_i, x_i - \delta_i)$ 를 구하고 이 두 기울기 절대값의 평균 $(|S(x_i, x_i - \delta_i)| + |S(x_i, x_i + \delta_i)|)/2$ 을 구한다. 안정

해로 선정되려면 이 기울기 절대값 평균이 0에 가까워야 한다. 즉 안정해란 임의의 잡음 δ_i 를 추가하였을 때 그 기울기 변화가 작아야 한다. 따라서 우리는 이 기울기 절대값 평균은 다음 단계의 적합도 함수에 적용을 하였다.

단계4: 적합도 계산

각각의 염색체에서 얻은 기울기 평균값을 적합도 함수에 적용하여 기울기가 큰 값을 대응되는 적합도를 저하시켜 퇴화시키도록 하였다. 적합도 함수에는 또한 함수의 값도 같이 적용하여 가능한 기울기가 작고 함수 값이 큰 값을 찾고자 노력하였다.

단계5: 룰렛 휠 선택방법(roulette wheel selection)으로 최종 해가 될 가능성이 없는 해들을 삭제하고 가능성이 높은 해들만 보존하여 해를 계속 하여 진화시킨다. 또한 토너먼트 선택방법에 비해 전해영역에서 해를 선택할 수 있어 좋은 해를 놓칠 가능성이 적다.

단계6: 교배와 돌연변이

교배는 2점 교배를 하고 한 개 염색체 내에서 적합도 가장 큰 개체가 다음 세대로 변경되지 않고 전달되는 것을 보장하는 엘리트 전략(elitist strategy)을 적용하여 진행한다. 돌연변이는 4가지 염기를 돌연변이 확률에 따라 변이 시키는 일반돌연변이를 진행하였다.

단계7: 선택, 교배, 돌연변이 후 다시 단계 2로 넘어가 주어진 조건이 만족할 때까지(초기화한 세대수만큼) 반복한다.

앞에서 설명한 방법은 x_i 를 중심으로 기울기 $S(x_i, x_i - \delta_i)$ 와 기울기 $S(x_i, x_i + \delta_i)$ 를 고려하여 두 기울기 절대값의 평균인 $(|S(x_i, x_i - \delta_i)| + |S(x_i, x_i + \delta_i)|) / 2$ 가 큰 값은 적합도를 저하시켜 퇴화시키고 작은 값은 살려주는 알고리즘이다.

4. 시뮬레이션

앞 절에서 제안한 방법이 입증됨을 보여주기 위해 아래와 같은 세 개의 함수를 사용하여 안정해를 찾아보았다.

4.1 함수

$$Y_1(x) = \begin{cases} e^{-\frac{2 \ln 2}{0.4} \frac{x-0.1}{0.4}} |\sin(5\pi x)|^{0.5} & 0.4 < x \leq 0.6 \\ e^{-\frac{2 \ln 2}{0.8} \frac{x-0.1}{0.8}} \sin^6(5\pi x) & \text{otherwise.} \end{cases} \quad (1)$$

$$Y_2(x) = \begin{cases} 1.5 & 0.2 < x \leq 0.25 \\ 0.98 & 0.6 < x \leq 0.75 \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

$$Y_3(x) = \begin{cases} e^{-\frac{2 \ln 2}{1.6} \frac{(x-0.1)(x-0.1)}{(x-0.1)^2}} |\sin(5^2 x)|^{0.5} |\sin(5^2 x^2)|^{0.5} & (0.4 < x \leq 0.6) \\ & \& (0.4 < x^2 \leq 0.6) \\ e^{-\frac{2 \ln 2}{1.6} \frac{(x-0.1)(x-0.1)}{(x-0.1)^2}} |\sin(5^2 x)|^{0.5} \sin^6(5^2 x^2) & (0.4 < x \leq 0.6) \\ & \& (x^2 \leq 0.4 \text{ or } 0.6 > x^2) \\ e^{-\frac{2 \ln 2}{1.6} \frac{(x-0.1)(x-0.1)}{(x-0.1)^2}} \sin^6(5^2 x) |\sin(5^2 x^2)|^{0.5} & (x \leq 0.4 \text{ or } 0.6 > x) \\ e^{-\frac{2 \ln 2}{1.6} \frac{(x-0.1)(x-0.1)}{(x-0.1)^2}} \sin^6(5^2 x) \sin^6(5^2 x^2) & \& (0.4 < x^2 \leq 0.6) \\ e^{-\frac{2 \ln 2}{1.6} \frac{(x-0.1)(x-0.1)}{(x-0.1)^2}} \sin^6(5^2 x) \sin^6(5^2 x^2) & \text{otherwise} \end{cases} \quad (3)$$

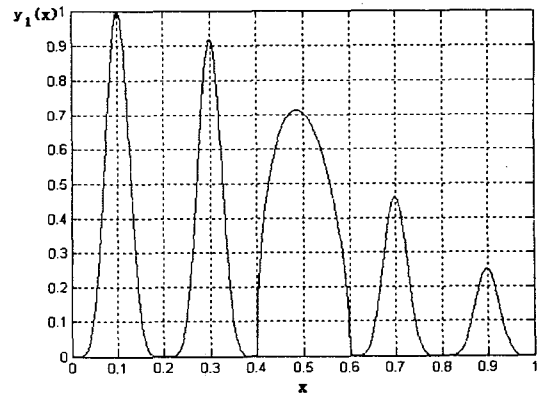
식 (1)의 함수는 일차원 다봉함수, 식 (2)는 간단한 계단 함수이며, 식 (3)은 이차원 다봉함수이다. 이들 함수는 각각 그림 4(a), 5(a), 6(a)에 보여준다. 식 (1)에서 전역해 $x=0.1$ 일 때 함수값 1.0을 가지고 안정해는 $x=0.486$ 일 때 0.715의 함

수값을 가진다[4]. 식 (2)는 x 가 0.2~0.25사이에서는 함수값 1.5를, 0.6~0.75사이에서는 함수값 0.98을 가진다. 식 (3)은 2차원 다봉함수 식 (1)을 2차원으로 확장한 다봉함수이다.

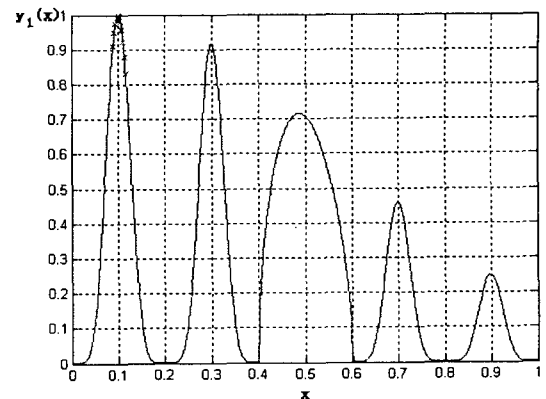
4.2 시뮬레이션 결과 및 분석

앞 절의 세 개의 함수들에 대하여 전역해와 안정해를 각각 구하는 시뮬레이션을 시행하였다. 시뮬레이션을 위해 사용된 변수는 세대수 50, 집단 크기 60, 교배확률은 80%, 돌연변이 확률 2%이다. 여기서 돌연변이확률 2%는 매우 큰 변이효과를 일으키는데 이는 염색체 길이가 유전자 알고리즘과 비교하여 상대적으로 길이가 길기 때문이다. 한편, DNA 염색체의 길이 600, 잡음의 강도는 0.05로 설정하여 위 함수들에 각각 30번씩 적용하여 시뮬레이션을 시행하였다.

그림 4와 그림 5는 본 논문에서 제안한 알고리즘을 식 (1)과 식 (2)에 각각 적용한 결과를 보여준다. 이 함수들에 DNA 코딩을 적용하기 위해서 한 개의 염색체에서 각 아미노산에 부여한 값들의 합을 그 염색체의 값으로 간주한 후, 그 염색체의 값을 두 함수의 정의역인 [0, 1]으로 매핑하여 시뮬레이션을 수행한다. 그림 4 (a)와 그림 5(a)는 일차원 함수의 예로 각 그림의 (b)는 단순 DNA 코딩 방법을 적용하였을 때 얻은 해의 분포를 보여준다. 여기서 단순 DNA 코딩이란 앞의 알고리즘에서 단계 2와 단계 3을 제외하고 수행한 방법을 의미하는 것으로 일반적인 DNA 코딩이다. 한편 각 그림의 (c)는 본 논문에서 제시한 기울기를 이용한 방법으로 50 세대 후의 해의 분포를 보여준다. 앞에서 언급한 바와 같



(a)



(b)

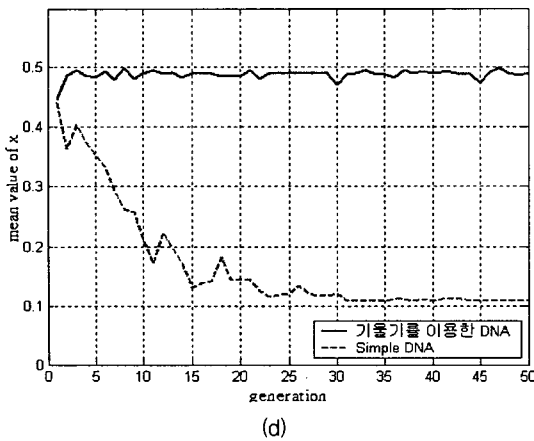
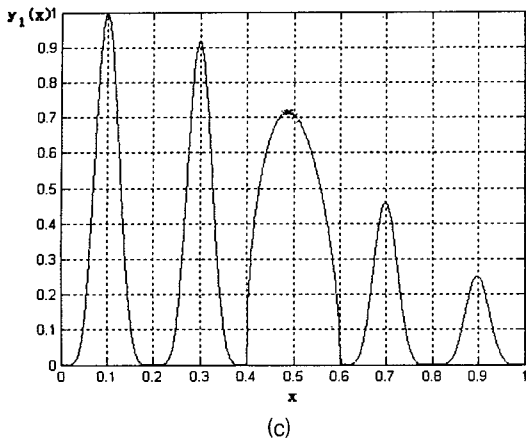


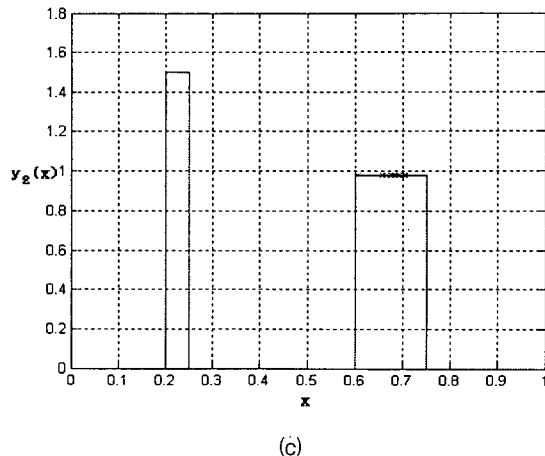
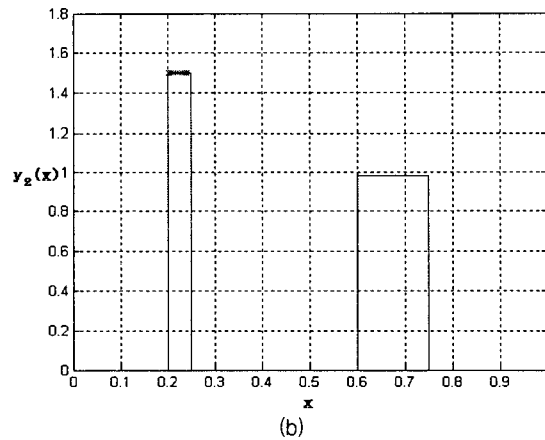
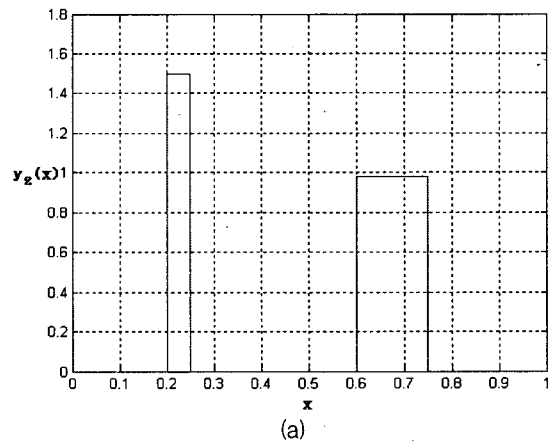
그림 4 (a) 원 함수 y_1 , (b) 단순 DNA 코딩에 의한 50세대 후의 해의 분포, (c) 기울기를 적용한 DNA 코딩 (Gradient DNA coding)에 의한 50세대 후의 해의 분포, (d) x (전체 집단) 평균의 변화.

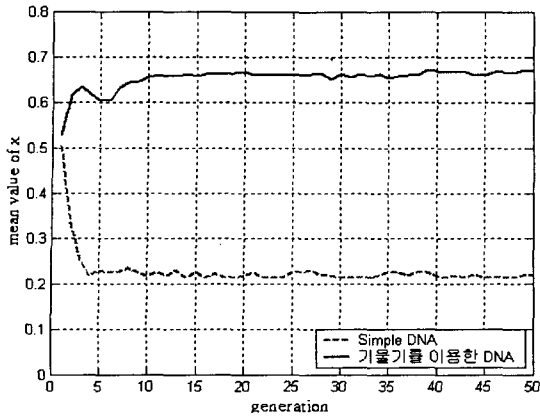
Fig. 4 (a) The original function y_1 , (b) A typical distribution of the individuals in function y_1 after 50 generations for the Simple DNA coding, (c) A typical distribution of the individuals in function y_1 after 50 generations for the Gradient DNA coding, (d) The variation of mean (over the population) value of x with function evaluations.

이 본 논문에서 찾고자 하는 해는 여러 가지 이유로 발생할 수 있는 해의 조그마한 변화(Δx)에 대하여 그 출력 성능이 비교적 안정된 값을 가져야한다. 그림 4와 그림 5의 각 (b)를 살펴보면 그 해의 위치가 일반적으로 주어진 문제의 최적해임을 알 수 있다. 그런데 이 해들을 살펴보면, 미세한 변화(Δx)에 대해 발생하는 출력 변화(Δy)가 크기 때문에 안정된 출력을 요구하는 문제의 해로 사용되기 곤란하다. 이에 반해 그림 4와 그림 5의 각 (c)를 보면 그 해가 최적해는 아니지만 $\Delta y/\Delta x$ 의 값이 비교적 적기 때문에 해의 미세한 변화에 대하여 발생하는 출력은 상당히 안정적이라 볼 수 있다.

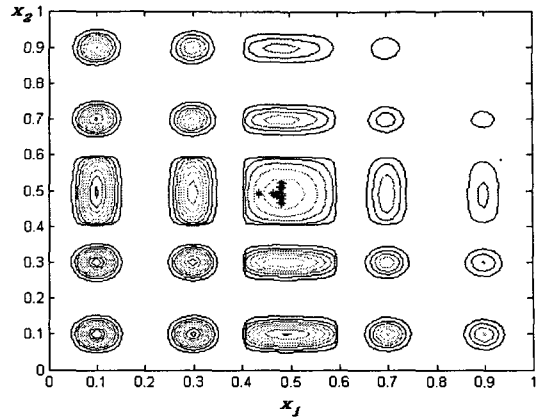
그림 4와 그림 5의 (d)는 세대가 지남에 따라 발생하는 해들의 평균값을 보여준다. 단순 DNA에 의한 결과는 점선으로 표현되어 있고, 기울기를 적용한 방법의 결과는 실선으로 나타나 있다. 이 두 경우 모두 해들이 비교적 특정한 값을 향하여 수렴해 가는 현상이 발견된다. 특별히 살펴볼 것은 기울기를 적용한 DNA 코딩은 몇 세대만에 해가 안정해로 수렴해 가는 것을 볼 수 있다.

그림 6은 식 (3)의 2차원 함수의 예를 보여준다. DNA 코딩을 적용하기 위해 한 개의 염색체를 두 부분으로 나누었다. 염색체의 전반부 아미노산값의 합 X1, 후반부 아미노산값의 합 X2를 각각 함수의 정의역인 [0, 1]에 매핑하였다. 그림





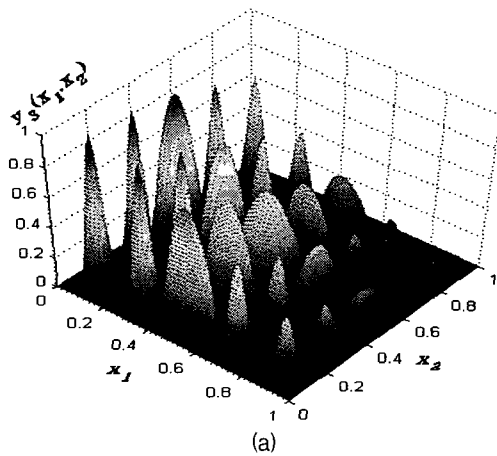
(d)



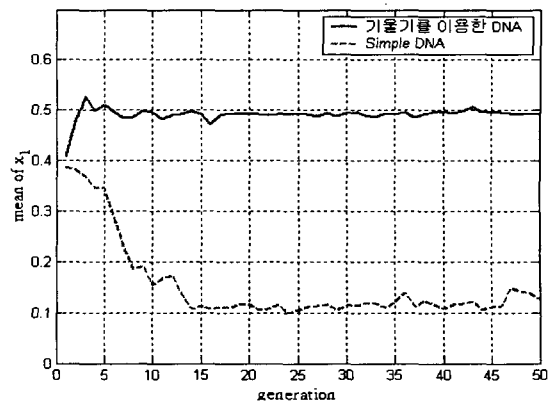
(c)

그림 5 (a) 원 함수 y_2 , (b) 단순 DNA 코딩에 의한 50세대 후의 해의 분포, (c) 기울기를 적용한 DNA 코딩 (Gradient DNA coding)에 의한 50세대 후의 해의 분포, (d) x (전체 집단) 평균의 변화.

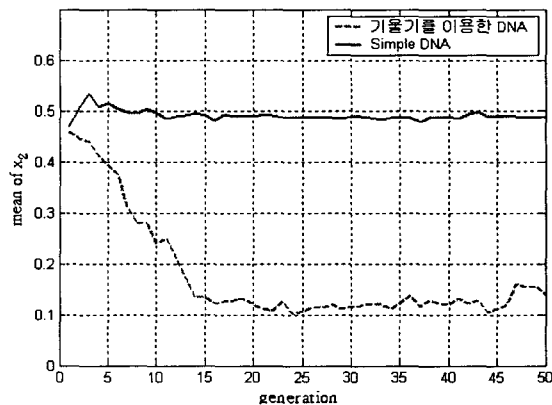
Fig. 5 (a) The original function y_2 , (b) A typical distribution of the individuals in function y_2 after 50 generations for the Simple DNA coding, (c) A typical distribution of the individuals in function y_2 after 50 generations for the Gradient DNA coding, (d) The variation of mean (over the population) value of x with function evaluations.



(a)



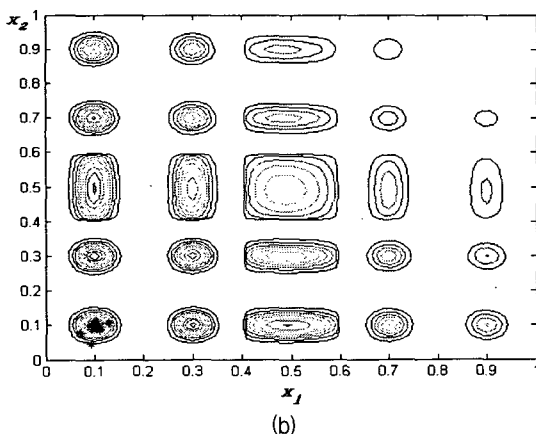
(d)



(e)

그림 6 (a) 원 함수 y_3 , (b) 단순 DNA 코딩에 의한 50세대 후의 해의 분포, (c) 기울기를 적용한 DNA 코딩 (Gradient DNA coding)에 의한 50세대 후의 해의 분포, (d) x_1 (전체 집단) 평균의 변화, (e) x_2 (전체 집단)평균의 변화

Fig. 6 (a) The original function y_3 , (b) A typical distribution of the individuals in function y_3 after 50 generations for the Simple DNA coding, (c) A typical distribution of the individuals in function y_3 after 50 generations for the Gradient DNA coding, (d) The variation of mean (over the population) value of x_1 with function evaluations, (e) The variation of mean (over the population) value of x_2 with function evaluations.

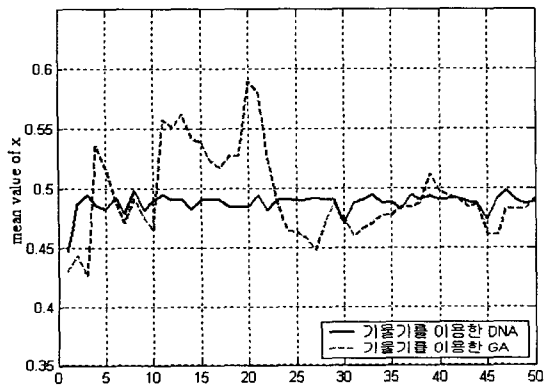


(b)

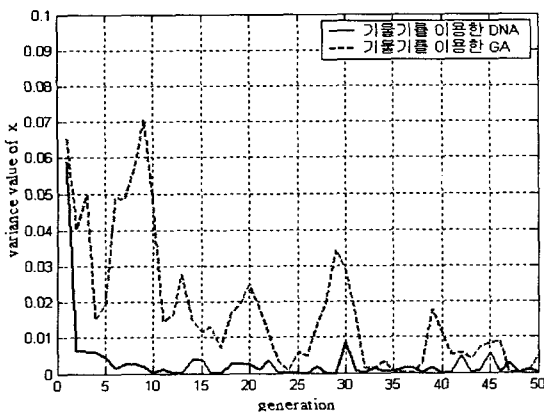
6(a)는 식 (3)에서 보여준 함수이며 이에 대한 등고선 그림이 그림 6의 (b) 또는 (c)에 나타나 있다. 그림 6의 (b)는 단순한 DNA 코딩을, (c)는 기울기를 이용한 DNA 코딩의 적용을 보여준다. 그림 4와 그림 5의 1차원 함수의 경우와 마찬가지로 그림 6(d)와 (e)는 각 세대마다 관찰된 해 집단 (x_1, x_2)의 평균이다. 2차원 함수의 경우도 기울기를 이용한 DNA 코딩이 매우 빠른 속도로 안정해에 접근하고 있다. 단순 DNA 코딩은 식 (3)이 가지고 있는 최적해로 수렴해 가고 있다.

이상에서 살펴본 결과에 의하면, 본 논문에서 제안한 기울기를 이용한 DNA 코딩이 주어진 문제의 안정해를 찾는 데 성공함을 알 수 있다. 한편 단순 DNA 코딩을 사용하면 안정해가 아닌 최적해로 모두 수렴하게 됨을 확인할 수 있었다.

그림 7 (a)와 (b)는 본 논문에서 적용한 DNA 코딩과 같은 방법을 이용하여 구현한 유전자 알고리즘에 적용하였을 때 발생하는 해 집단의 특성(평균과 분산)을 보여주고 있다. 동일한 세대수, 개체수, 교배확률, 그리고 돌연변이확률을 사용하여 시뮬레이션을 수행하였다. 그림 7은 DNA 코딩이 유전자 알고리즘에 비하여 탐색된 해들의 값이 안정되어 있음을 설명하고 있다. 그러나 DNA 코딩의 특성상 한 염색체 내에



(a)



(b)

그림 7 기울기를 이용한 DNA 코딩(Gradient DNA coding)과 같은 방법을 이용한 유전자 알고리즘의 비교 (a) 평균 (b) 분산

Fig. 7 Comparison of gradient DNA coding and gradient GA(genetic algorithm), (a) variation of mean (b) variation of variance.

존재하는 해의 수가 가변적이기 때문에 공정한 비교가 쉽지 않다. 다만 각 세대에서 얻은 해의 평균과 분산을 통해 그 경향을 짐작할 수 있을 정도이다.

5. 결 론

본 논문에서는 여러개의 해가 존재하는 다봉함수에서 안정해를 찾는 기울기를 적용한 DNA 코딩(Gradient DNA coding)을 제안하였다. 제안된 알고리즘은 해의 길이가 가변적이고 하나의 염색체에 다수의 해(GENE)가 들어있는 생물학적인 DNA와 유전자 알고리즘의 선택, 교배, 돌연변이 등 연산을 결합한 DNA 코딩방법이다.

본 알고리즘에서는 각 아미노산에 코드를 부여하는 대신 알고리즘과는 특정한 값을 부여하여 각 GENE에 속하는 아미노산 값들의 합을 정의역에 mapping하였다. 또한 잡음 $\Delta = (\delta_1, \delta_2, \dots, \delta_n)$ 을 생성하여 각 개체 x_i 를 중심으로 기울기 $S(x_i, x_i - \delta_i)$ 와 기울기 $S(x_i, x_i + \delta_i)$ 를 고려하여 두 기울기 절대값의 평균값 $(|S(x_i, x_i - \delta_i)| + |S(x_i, x_i + \delta_i)|) / 2$ 를 적합도 함수로 사용하였다. 즉 기울기의 절대값 평균이 큰 값을 퇴화시키고 적은 값을 선택하는 방향으로 진화시켰다. 또한 엘리트 전략(elitist strategy)을 적용하여 매 염색체의 가장 좋은 해들을 살리는 교배를 진행하였다.

제안한 알고리즘의 성능을 알아보기 위하여 세 종류의 함수에 적용하여 안정해를 찾는 시뮬레이션을 수행하였다. 그 결과 모든 함수에서 찾고자 하는 안정해가 비교적 빠른 세대 안에 발견됨을 확인할 수 있었다. 특히 기울기 정보를 사용하지 않은 경우는 안정해 대신 전역해로 수렴함을 알 수 있었다.

본 논문에서는 제안한 DNA 코딩방법을 같은 방법을 적용하여 구현한 유전자 알고리즘과 비교하였다. 해 집단의 특성을 살펴본 결과, DNA 코딩방법을 이용한 알고리즘이 일반 유전자 알고리즘보다 훨씬 더 좋은 방법임을 확인하였다. 하지만 시뮬레이션 수행속도에 있어서는 유전자 알고리즘이 DNA코딩방법에 비해 더 빠르다.

앞으로의 연구 과제는 본 연구에서 제안한 방법을 더욱 복잡한 실제문제에서 적용하여 주어진 조건의 성능을 만족하는 안정해를 구하는 것이다. 이를 위하여 현 알고리즘의 속도를 더 빠르게 하고 적당한 강도 변수와 정의역에 mapping하는 방법을 더 깊게 연구를 하는 것이다.

참 고 문 헌

- [1] Piotr Wasiewicz, Tomasz Janczak, J. Mulaka, "the Inference via DNA Computing," Proc. IEEE Int. Conf. on Evolution computation, pp. 988-993, 1999.
- [2] R. Deaton et. al, "A DNA Based Implementation of an Evolutionary Search for Good Encodings for DNA Computation," Proc. IEEE Int. Conf. on Evolution computation, Indianapolis, IN, USA, pp. 267-271, April, 1997.
- [3] B. Sareni, L. Krahenbuhl, A. Nicolas, "Niching Genetic Algorithms for Optimization in Electromagnetics I.

- Fundamentals," IEEE Transactions on Magnetics. Vol. 34, No. 5, pp. 2984-2987, 1998.
- [4] S. Tsutsui, A. Ghosh, "Genetic Algorithms with a Robust Solution Searching Scheme," IEEE Trans. on Evolutionary Computation, vol. 1, no. 3, pp. 201-208, 1997
- [5] Brad L. Miller and Michael J. Shaw, "Genetic Algorithms with Dynamic Niche Sharing for Multimodal Function Optimization," Proc. IEEE Int. Conf. on Evolutionary Computation, pp. 786-791, 1996.
- [6] T. Yoshikawa, T. Furuhashi, Y. Uchikawa, "Emergence of effective fuzzy rules for controlling mobile robots using DNA coding method," Proc. IEEE Int. Conf. on Evolutionary Computation, pp. 581-586, 1996.
- [7] M.R.M. Rizk, I.F. El-Arabawy, H.S. Khaddam, "An algorithm for optimum stability region of fuzzy control systems using genetic algorithms," Proc. American Control Conf., vol. 1, pp. 192-197, 2001.
- [8] O. Nasraoui, R. Krishnapuram, "A novel approach to unsupervised robust clustering using genetic niching," Proc. IEEE Ninth Int. Conf. on Fuzzy Systems, (San Antonio, TX) Vol. 1, pp. 170-175, May 2000.
- [9] H. Kawanaka, T. Yoshikawa, S. Tsuruoka, "Acquisition of fuzzy control rules for a mobile robot using genetic algorithm," Proc. 6th Int. Workshop on Advanced Motion Control, pp. 507-512, 2000.
- [10] Zhijiang Guo, Hongtao Zheng, Jinping Jiang, "A powerful modified genetic algorithm for multimodal function optimization," Proc. American Control Conf., vol. 4, pp. 3168-3173, 2002.
- [11] Chang-Su Park, Hungu Lee, Hyo-Choong Bang, Min-Jea Tahk, "Modified Mendel operation for multimodal function optimization," Proc. Conf. on Evolutionary Computation (CES 2001), vol. 2, pp. 1388-1392, 2001.
- [12] Ki-Youl Lee, Dong-Wook Lee, Kwee-Bo Sim, "Evolutionary neural networks for time series prediction based on L-system and DNA coding method," Proc. Conf. on Evolutionary Computation, vol. 2, pp. 1467-1474, 2000.
- [13] Tomohiro Yoshikawa, Takeshi Furuhashi, Yoshiki Uchikawa, "DNA Coding Method and a Mechanism of Development for Acquisition of Fuzzy Control Rules," Proc. IEEE Int. Conf. on Fuzzy Systems, vol. 3, pp. 2194-2200, 1996.

저 자 소 개



염 학 송(廉鶴松)

1971년 1월 7일생. 1994년 중국길림대학교 환경과학과 졸업. 2003년 명지대학교 정보제어공학과 졸업(석사). 관심분야 : 유전자 알고리즘, 워터마크.

Tel: 86-433-282-8690

E-mail: hslian@hanmail.net



김 인 택(金仁澤)

1960년 12월 25일생. 1984년 서울대학교 공과대학 전자공학과 졸업. 1986년 서울대학교 대학원 전자공학과 졸업(석사). 1992년 Georgia Institute of Technology, 전기공학과 졸업(공학박). 1993년-1995년 LG전자기술원 선임연구원. 1995-현재 명

지대학교 정보공학과 부교수. 2001-2002 미국농림부(USDA) Visiting Scientist. 연구분야: 센서융합, 정보보호, 영상분석
Tel: (031) 330-6475

E-mail: kit@mju.ac.kr