

# 대화형 진화연산을 이용한 아바타 생성

이자용, 오재홍, 고희승, 강훈  
중앙대학교 전자전기공학부

## An Interactive Approach based on Genetic Algorithm Using Hidden Population and Simplified Genotype for Avatar Synthesis

Jayong Lee, Jaehong Oh, Hyeungseong Ko, Hoon Kang  
School of Electrical and Electronic Engineering, Chung-Ang University  
E-mail : jalnans@sirius.cie.cau.ac.kr

### Abstract

본 논문에서는 사용자 개개인에 최적화된 아바타를 생성하기 위해 대화형 진화 연산(Interactive Genetic Algorithm, IGA)을 적용하는 방법을 제안하고 있다. 대화형 진화연산은 사용자의 선택을 적합도 평가에 사용하는 방법이기 때문에, 사용자의 개인적인 취향을 아바타 생성 과정에 반영할 수 있다. 본 연구에서는 기존의 대화형 진화연산이 가지고 있는 단점을 극복하기 위해 hidden population 기법과, simplified genotype 기법을 제안한다. 이러한 방법들은 단시간 내에 최적화된 결과물을 생성하도록 유도함으로써 IGA 시스템의 최대 문제점인 사용자의 피로도를 최소화한다. 마지막으로, 제안하고 있는 알고리즘의 우수성을 증명하기 위해 사용자의 만족도나 신뢰도를 측정할 수 있는 독자적인 평가 방법을 소개하고 있다.

### I. 서론

대화형 진화 연산(Interactive Genetic Algorithm, IGA)은 주관적인 사용자의 평가를 GA 에 적용하여 최적화된 값을 찾아내는 방법이다. 간단히 말해서, IGA 는 적합도 함수 대신 사용자가 평가하는 GA 라고 할 수 있다. 일반적으로 IGA 가 적용된 시스템에서는 인간의 주관적인 평가를 적합도 평가에 사용하기 때문에 각 개인의 취향이 반영된 결과물을 얻을 수 있다. 이러한 관점에서 IGA 기법은 인간의 취향이나 감정, 심리적인 성향

등을 시스템에 반영할 수 있는 기법이라고 할 수 있겠다[3].

IGA 기법의 가장 큰 문제점은 사용자가 되풀이되는 평가 과정에서 쉽게 피로해질 수 있다는 점이다. 사용자의 피로를 고려해야 하기 때문에, 대부분의 IGA 기법에서는 다른 GA 기법과는 달리 많은 수의 개체를 사용하거나 많은 세대를 거치지 못하는 단점이 존재한다.

본 논문에서는 기존의 IGA 기법의 단점을 극복하기 위한 2 가지 해결책을 제안하고 있다. 개체군의 제한을 극복하기 위해, 사용자에게 직접 평가를 받지 않는 'hidden population'을 사용하였다. 또한, 결과물을 얻기까지의 세대 수를 줄이기 위한 방법으로 'simplified genotype'을 사용하였다. 전체 생성 과정의 복잡도를 줄이기 위해 기존의 유전형을 단순화한 simplified genotype 을 진화에 사용하여 보다 빠른 수렴을 유도하고 있다.

### II. IGA 에 기초한 아바타 생성

IGA 는 사용자의 입력을 적합도 평가에 사용한다. 따라서, 사용자가 평가해야 하는 개체가 많거나 최종 결과물을 얻기까지 많은 세대를 거쳐야 한다면 사용자는 쉽게 피로를 느끼게 될 것이다. IGA 를 적용한 시스템의 경우, 사용자의 피로도를 줄이기 위한 다양한 연구가 시도되고 있지만, 대부분 사용자 인터페이스를 개선하여 피로도를 줄이는 방법에 그치는 실정이다[4]. 인터페이스를 개선하는 경우 사용자의 피로도를 일부 줄여줄 수는 있지만, 근본적인 해결책이라고 볼 수는 없다.

본 논문에서는 사용자의 피로도를 줄이면서 효과적으로 최적화된 결과를 얻기 위해서 2 가지 방법을 제안한다.

- Hidden Population
- Simplified Genotype

### 2.1 Hidden Population

일반적으로 개체군의 크기가 큰 경우에는 작은 경우에 비해 최적화된 결과물을 얻는데 많은 시간이 소요되지만 보다 더 좋은 결과를 얻을 가능성이 크다. 반면, 개체군의 크기가 작은 경우에는 해공간에 대한 충분한 탐색이 이루어지지 않아 국소해에 빠질 가능성이 높다.

일반적으로 IGA 가 적용된 시스템은 모든 개체에 대해서 사용자의 평가를 받는 방법을 사용하고 있다. 이러한 방법의 경우, 사용자의 피로도를 고려해야 하기 때문에 전체 개체군의 크기를 크게 할 수 없다는 제한이 존재한다. 이러한 개체군의 크기 제한에 대한 문제들은 hidden population 을 이용하여 해결 할 수 있다. 전체 개체군은 hidden population 과 displayed population 으로 나뉜다. Displayed population 은 적합도 평가를 받기 위해 사용자에게 보여지는 개체를 말한다. 반면, hidden population 은 사용자에게 직접적인 적합도 평가를 받지

않는 개체를 뜻한다. 전체 개체군은 각각 미리 정의된 몇 개의 클러스터로 나뉜다.

사용자가 낮은 평가를 내린 개체는 전체 개체군에서 제거된다. 그대신 제거된 숫자만큼 높은 평가를 받은 개체가 그 자리에 복사된다. 따라서, 선택 받은 개체가 포함된 클러스터의 개체 수는 증가하고, 반대로 선택 받지 못한 개체가 포함된 클러스터의 개체 수는 감소한다. 룰렛 선택(roulette selection)에 의거하여, 많은 개체를 소유하고 있는 클러스터의 개체가 다음 세대에 유전자를 남길 확률이 높아진다.

그림 1 은 사용자의 평가가 실제로 클러스터의 적합도에 반영되는 과정을 나타내었다. 그림에서 전체 개체군의 크기는 15, displayed population 의 크기는 5, hidden population 의 크기는 10 이며, 클러스터는 4 개이고 초기값은 그림 1-(a)에 나타낸 것과 같다. 사용자가 5 개의 displayed population 중 클러스터 A 에 포함된 개체를 선택한 경우, 이것은 사용자가 클러스터 A 에 포함된 개체를 다른 클러스터에 포함된 개체보다 적합도가 높게 평가하고 있다고 판단할 수 있다. 따라서 이러한 경우 클러스터 B 와 C 에 포함된 displayed population 은 제거되고 그 자리에 사용자가 선택한 개체와 동일한 유전자를 가지고 있는 개체가 생성된다. 결과적으로 그림 1-(b)에 나타난 것처럼 클러스터 A 의 개체수가 증가하므로 다음 세대에 자신의 유전자를 남길 확률이 커진다. 그리고 사용자에게 선택 받은 개체는 엘리트(elite)로 간주되어 교차나 변이 연산을 거치지 않고 그대로 다음 세대에 전달된다. 클러스터 D 의 경우에는 displayed population 에 속하는 개체가 없기 때문에 클러스터의 크기에 변화가 없고, 따라서 확률적으로 다음 세대에서 클러스터의 크기를 유지할 가능성이 높다.

Hidden population 을 사용하는 경우, 각 개체의 적합도를 일일이 평가하는 대신에 사용자의 평가를 개체가 포함된 클러스터의 적합도 평가에 사용하기 때문에 전체 개체군의 크기에 제한이 없다. 다시 말해서, hidden population 을 사용하면 IEC 가 적용된 시스템 특유의 개체군 크기 제한을 극복할 수 있다.

### 2.2 Simplified Genotype

유전자의 길이는 진화 연산을 사용하는 시스템에서 수렴에 걸리는 시간에 영향을 준다[2]. 유전자형의 길이가 길다면 그만큼 탐색 영역이 넓어진다는 것을 의미하

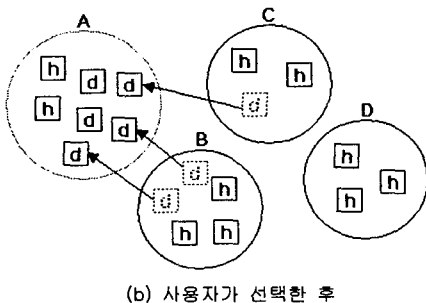
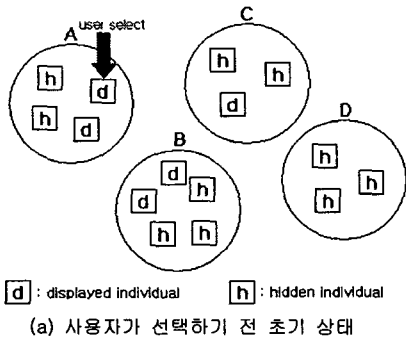


그림 1 사용자의 선택 의한 클러스터의 변화

Face Texture 2bits	Ear Texture 2bits	Eyebrow Texture 2bits	Nose Texture 2bits	Eye Texture 4bits	Lip Texture 4bits	Cluster (Pattern) 4bits	Morphing Rate 4bits
-----------------------	----------------------	--------------------------	-----------------------	----------------------	----------------------	----------------------------	------------------------

그림 2 Gene

고, 이것은 높은 복잡도로 인해 시스템이 결과를 산출하는데 오랜 시간을 필요로 한다고 할 수 있다.

그림 2 는 진화 연산에 사용되는 염색체의 유전자형(genotype)을 보여준다. 유전자형은 총 24 비트 길이이다. 상위 16 비트는 6 개의 부분에 사용될 텍스처의 종류를 표현한다. 다음의 4 비트는 소속된 클러스터를 나타내며, 나머지 4 비트는 클러스터의 기본 패턴과의 모핑 비율(morphing rate)를 나타낸다.

전문가에 의해서 미리 디자인된 다양한 패턴은 인상이나 화풍에 따라 몇 개의 클러스터로 분류된다. 이러한 패턴들은 비전문가인 사용자들의 선택을 돕는 역할을 하면서 동시에 화풍을 표현하는 역할을 한다. 또한, 불필요한 탐색 영역을 미리 제거함으로써 결과물을 얻기까지 걸리는 세대 수를 줄일 수 있다. Simplified genotype 의 사용은 결과적으로 사용자의 노력을 최소화 하면서 동시에 최적화된 결과물을 빠른 시간 안에 얻을 수 있게 해준다.

### 2.3 아바타 생성 시스템

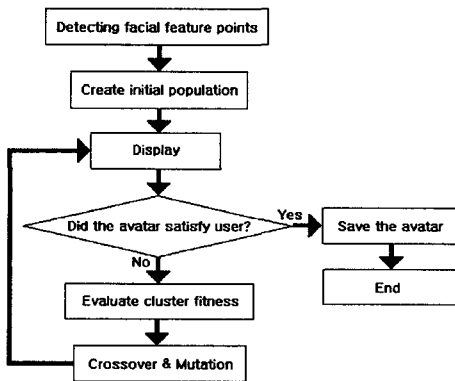


그림 3 System Block Diagram

그림 3 에 전체 아바타 생성 시스템의 block diagram 을 보였다. 우선, 입력된 사진으로부터 얼굴 각 부위의 특징점을 추출한다. 추출된 특징점들을 바탕으로 초기 개체군이 생성된다. 생성된 개체군 중에서 displayed

population 에 해당하는 개체는 사용자의 평가를 받기 위해 화면에 표시된다. 만약 사용자를 만족시키는 아바타가 화면에 존재한다면, 결과물을 저장하고 전체 프로세스가 종료된다. 만족할만한 아바타가 화면에 없는 경우에, 사용자는 화면에 표시된 displayed population 에 속하는 개체들을 평가하게 되고, 사용자의 평가는 각 클러스터를 평가하는데 사용된다. 마지막으로, 다음 세대를 구성할 새로운 개체군이 교차(crossover)와 변이(mutation) 과정을 거쳐서 생성된다. 이러한 과정을 사용자를 만족시키는 아바타가 나타날 때까지 반복한다.

## III. 실험 결과 및 분석

### 3.1 실험 환경

개체군은 100 개의 개체로 이루어진다. 100 개의 개체 중 12 개는 displayed population 에 속하고, 나머지 88 개가 hidden population 으로 사용된다. 일점 교차(one-point crossover)를 사용하였으며, 교차율(crossover rate)은 50%, 변이율(mutation rate)은 3%이다. 엘리트 보존 선택(elitism)에 의해, 사용자의 선택을 받은 개체는 다음 세대에서도 그대로 보존된다.

정확한 데이터 획득을 위해 각 데이터에 대한 신뢰도 테스트가 병행된다. 신뢰도는 현재 세대에서의 사용자의 선택을 최근 10 세대에서의 선택과 비교하여 사용자의 선택이 동일한 클러스터에 포함되는 비율로 계산된다. 신뢰도가 너무 낮은 경우에는 사용자가 일관된 선택을 하지 않는다고 간주 할 수 있다. 실험적인 경험에 의해, 신뢰도가 50% 이상인 결과에 대해서만을 실험 데이터로 인정한다.

### 3.2 적합도 평가

IGA 를 적용한 시스템의 경우 적합도 평가가 사용자의 선택에 의해 이루어지기 때문에, 적합도를 산술적으로 계산하기가 쉽지 않다. 따라서, 최종 결과물의 적

합도를 평가하기 위한 평가 기준이 필요하다. 적합도 평가를 위해 3 가지 기준을 사용한다.

첫번째 방법으로, 전체 탐색 영역에서 동일한 간격으로 샘플 아바타를 추출하여 피험자에게 무작위 순서로 보여주어 평가를 하게 함으로써 사용자의 취향을 대략적으로 선호도 데이터를 얻을 수 있다.

두번째 평가기준은 클러스터의 적합도이다. 사용자의 입력은 클러스터의 적합도를 평가하는데 사용되기 때문에, 개체가 소속된 클러스터의 적합도를 각 개체의 적합도 평가에 반영하는 것이 타당하다.

마지막으로 텍스처에 대한 선호도 평가를 평가한다. 사용자에게 아바타 생성과정에 사용된 텍스처들을 보여주고 이를 평가하게 한다.

각 개체의 적합도는 3 개의 평가 기준, 즉 샘플 아바타를 기준으로 얻어진 값과 사용된 텍스처에 대한 선호도 그리고 소속된 클러스터의 적합도를 합쳐서 평가된다. 각 개체의 적합도를 구하는 식은 다음과 같다.

$$f_i = \text{avr}(p_i, t_i, c_i)$$

$f_i$  : fitness function

$p_i$  : pattern fitness value

$t_i$  : texture fitness value

$c_i$  : cluster size value

### 3.3 실험 결과

표 1 에 실험 결과를 나타내었다. 표에서 보이는 것과 같이, 최적해를 얻기까지 걸리는데 평균적으로 13 세대가 걸리는 것을 알 수 있다. 즉, 본 시스템에서 사용자는 단지 13 번의 선택만으로 최적화된 결과를 얻을 수 있다. 동시에 가장 큰 클러스터가 50% 이하인 것으로 미루어 개체의 다양성을 유지한다는 사실을 알 수 있을 것이다. 최종적으로 만들어진 결과물의 일부를 그림 4 에 보였다.

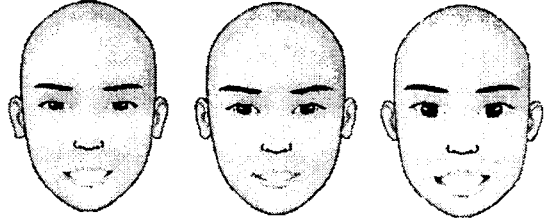


그림 4 Avatars

## IV. 결론

IGA 를 적용한 시스템은 사용자가 각 개체의 적합도를 직접 평가한다. 따라서, 전체 개체의 수가 많거나 최적화된 결과를 얻기까지 많은 세대를 거쳐야 한다면 사용자가 쉽게 피로해져서 원하는 결과를 얻기가 어려워 질 것이다. 본 논문에서는 기존의 IGA 의 단점을 극복할 수 있는 2 가지 방법을 제안하고 있다.

- Hidden Population
- Simplified Genotype.

IGA 를 적용한 시스템에 본 논문에서 제안하고 있는 방법들을 사용한다면 사용자의 피로도를 줄여줄 뿐만 아니라 더욱 효율적으로 최적화된 결과물을 생성해 줄 것이다. 제안된 방법들의 유용성은 확인하기위해 다양한 평가 기준과 방법을 마련하여 실험하였다. 여러 실험적인 결과를 보면, 본 논문에서 제안하고 있는 방법을 사용한 IGA 기법이 기존의 단점을 극복했을 뿐만 아니라 빠르게 만족도 높은 결과물을 생성해 낸다는 사실을 확인 할 수 있다.

## 참고문헌

[1] D.E.Golbber, *Genetic Algorithms in Search, Optimization, and Machine Learning*, Massachusetts, Addison-Wesley, 1989.  
 [2] P.Bentley, *Evolutionary Design by Computers*, California, Morgan Kaufmann, 1999.  
 [3] H.Takagi, "Interactive Evolutionary Computation: Fusion of the Capabilities of EC Optimization and Human Evaluation", *Proc. of the IEEE*, vol: 89 Issue: 9, pp. 1275-1296, 2001.  
 [4] H.Takagi, K.Ohya, "Discrete fitness values for improving the human interface in an interactive GA", *IEEE 3<sup>rd</sup> Int'l Conf. on Evolutionary Computation*, pp.109-112, 1996.

표 1 Experimental results.

	Max.	Min.	Average
Reliability	100	65	89
Best fitness	90	68	81
Average fitness	65	48	58
Largest cluster size	62	41	49
Generation	22	8	13