

계층적 신경망을 이용한 객체 영상 분류

(Object Image Classification Using Hierarchical Neural Network)

김종호*, 김상균**, 신범주***

(Jong-Ho Kim, Sang-Kyoon Kim, Bum-Joo Shin)

요약 본 논문에서는 내용기반 영상 분류를 위한 방법론으로써 신경망을 이용한 계층적 분류 방법을 제안한다. 분류 대상 영상은 인터넷상의 다양한 영상들 중에서 전경과 배경의 구분이 있는 객체 영상이다. 전처리 과정에서 영역 분할을 이용하여 영상 내에서 배경을 제거하고 객체 영역을 추출한다. 분류를 위한 특징으로는 웨이블릿 변환 후 추출된 형태 특징과 질감 특징을 이용한다. 추출된 특징 값들을 Principal Component Analysis(PCA)와 K-means를 이용해서 군집화 시키고 유사한 군집들을 묶으면서, 5단계의 계층적 분류기를 구성한다. 계층적 분류기는 BP를 학습 알고리즘으로 사용하는 59개의 신경망분류기로 구성된다. 배경을 제거하고 질감특징 중 가장 높은 분류율을 보이는 대각 모멘트를 사용하여 실험하였을 때, 100종류에서 각 10개씩, 총 1000개의 학습 데이터와 1000개의 테스트 데이터에 대하여 각각 81.5%와 75.1%의 정분류율을 보였다.

핵심주제어 : 계층적 신경망, 내용기반 영상 분류, PCA, K-means

Abstract In this paper, we propose a hierarchical classifier of object images using neural networks for content-based image classification. The images for classification are object images that can be divided into foreground and background. In the preprocessing step, we extract the object region and shape-based texture features extracted from wavelet transformed images. We group the image classes into clusters which have similar texture features using Principal Component Analysis(PCA) and K-means. The hierarchical classifier has five layers which combine the clusters. The hierarchical classifier consists of 59 neural network classifiers learned with the back propagation algorithm. Among the various texture features, the diagonal moment was the most effective. A test with 1000 training data and 1000 test data composed of 10 images from each of 100 classes shows classification rates of 81.5% and 75.1% correct, respectively.

Key Words : Hierarchical Neural Network, Content Based Image Classification, PCA, K-means

1. 서론

신경망은 복잡하게 상호 연결된 뉴런의 망으로 구성된 생체 학습 시스템에서 영감을 받은 것으로, 이런 생물학적 처리 과정을 빌어 효율적인 기계학습을 제공한다. 이러한 학습 특성 때문에 신경망은 여러 분야에서 사용되는데, 특

히 학습을 통한 인식 및 분류 분야에서 탁월한 성능을 발휘한다.

정보 분류에 있어서 신경망을 이용한 기존의 연구 방법들은 주로 단일 신경망을 사용한다. 그리고 성능을 개선하기 위해 신경망의 초기 가중치나 위상 구조를 다양하게 하거나, 각 신경망을 서로 다른 학습 데이터나 학습 알고리즘을 사용하여 학습한다. 각 신경망에서 나온 결과를 투표나 평균 방법을 이용하여 결합해서 결과를 도출하는 방법을 사용하고 있다[1][2].

* 인제대학교 전산학과 대학원

** 인제대학교 컴퓨터공학부

*** 부산대학교 바이오시스템학부

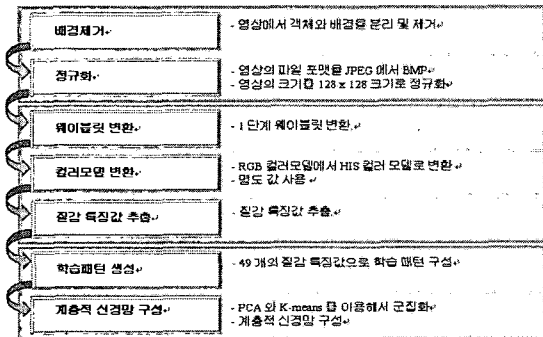
그러나 기존의 방법들은 분류하고자하는 대상 클래스의 수가 증가하면 뉴런의 연결 강도가 급격하게 증가해서 학습 시간이 오래 걸리고 학습 효율 또한 떨어진다. 따라서 다양한 종류의 영상을 분류하기 위한 방법론으로 기존의 단일 신경망을 직접적으로 적용하기 어렵다.

본 연구에서는 이러한 단점을 극복하기 위해 기존에 단일 신경망을 계층적으로 결합한 영상 분류기를 제안한다. 분류대상 영상은 인터넷상의 다양한 영상들 중에서 전경과 배경의 구분이 있는 객체 영상이다. 이러한 객체 영상에서 영역 분할을 이용하여 배경을 제거한다. 배경을 제거함으로써 객체 자체만의 정보를 추출하여 보다 정확한 분류를 한다. 웨이블릿 변환을 이용하여 영상에서 형태, 질감의 정보를 추출하며, 특징 추출 시 객체의 부분 부분에서 구조적 특징을 추출함으로써 보다 많은 내용정보를 획득한다. 추출된 특징을 PCA와 K-means를 이용해서 유사한 특징을 가지는 영상들로 군집화 시키고 계층적 분류기를 구성한다.

계층적 분류기는 5계층 59개의 신경망 분류기로 구성된다. 이렇게 함으로써 개별 신경망의 분류 클래스 수를 줄여서 학습시간이 단축되고 분류효율이 높아진다.

2. 객체 영상의 분류

본 연구에서 제시하는 영상 분류기의 전체 구성절차는 [그림 1]과 같다. 본 영상 분류기는 크게 전처리 모듈, 특징 추출 모듈, 분류 모듈의 3가지로 구성된다.



[그림 1] 제안된 분류기의 전체 구성

전처리 모듈에서는 배경 제거와 정규화가 이루어진다. 먼저 지역 분할(region segmentation)을 이용하여 영상에서 배경을 제거한다. 다음으로 JPEG 포맷의 영상 파일을 BMP 포맷으로 변환한다. 그리고 영상 정보를 신경망의 학습패턴으로 사용하기 위해 영상의 크기를 128x128로 정규화 한다.

특징 추출 모듈에서는 웨이블릿 변환과 컬러 모델 변환, 그리고 질감 특징 값을 추출한다. 정규화된 영상에서 형태, 질감 정보를 획득하기 위하여 1단계 웨이블릿 변환(1-level wavelet transform)을 한다. 그리고 RGB 컬러 모델에서 HSI 컬러 모델로 변환한 뒤 명도(intensity) 값을 사용한다. 다양한 질감 특징들 중 가장 높은 분류율을 보여주는 대각 모멘트(diagonal moment)를 사용하여 구조적인 특징 값을 추출한다.

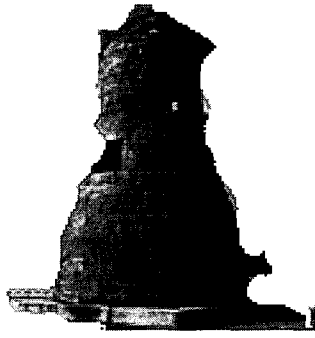
분류 모듈에서는 PCA와 K-means를 이용해서 영상들을 군집화 시키고 5계층 분류기를 구성한다. 최종적으로 이전단계에서 추출된 특징 값을 사용하여 5계층 신경망 분류기를 학습 및 테스트한다.

2.1 배경 제거

인터넷에서 획득한 영상 파일들은 대부분 복잡한 배경을 가지고 있다. 본 논문에서는 전처리 과정으로 영상에서 배경을 제거한다. 배경을 제거하기 위해 JSEG 영상 분할 알고리즘[3]을 이용해서 영상을 분할하고 객체 영상에서 객체는 영상의 가운데에 있다는 점을 이용해서 배경을 제거한다[4]. [그림 2]는 복잡한 배경을 가진 영상이고 [그림 3]은 배경영역을 제거한 영상이다.



[그림 2] 복잡한 배경을 가진 영상



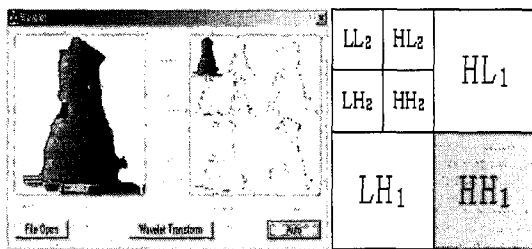
[그림 3] 배경이 제거된 영상

2.2 정규화

인터넷에서 획득한 영상 파일들은 대부분 JPEG 포맷이다. 본 연구에는 JPEG 포맷의 영상 파일을 BMP 포맷으로 변환한다. 변환된 영상 파일들은 다양한 크기를 가지고 있다. 그러나 영상 데이터를 신경망 분류기의 학습패턴으로 사용하기 위해서는 데이터가 동일한 크기여야 한다. 그래서 본 연구에서는 영상 파일의 크기를 모두 128×128로 정규화 한다.

2.3 웨이블릿 변환

본 연구에서는 영상의 질감 특징을 추출하기 위하여 먼저 웨이블릿 변환을 수행하여 저주파 성분과 고주파 성분의 부 밴드(subband)들로 분할한다.



[그림 4] 웨이블릿 분할 영역

웨이블릿 변환을 이차원 영상에 2단계 변환하였을 때 [그림 4]와 같이 7 개의 성분으로 나누어진다. 이 7 개의 성분은 시각에 민감한 정보를 가지고 있으며, 크게 저주파 성분과 고주파 성분으로 나누어지며, 고주파 성분은 또다시 수

직 성분, 수평 성분, 대각 성분으로 나누어진다. 본 연구에서는 가장 정보의 양이 많은 대각 성분(HH₁)을 사용한다.

2.4 컬러 모델 변환

컴퓨터 시스템에서는 RGB 컬러들이 주기적으로 복합되어 결과적인 컬러를 만들어 낸다. 이러한 RGB 요소들은 상호 관계가 너무 크기 때문에 영상 처리 알고리즘을 수행하기 어렵다 [5][6].

따라서 본 연구에서는 많은 영상 관련 어플리케이션에서 사용되고 있는 HSI 컬러 모델로 변환하여 그 중 명도(intensity) 값을 사용한다. 변환된 HSI 컬러 모델에서 명도 값을 사용함으로써 얻을 수 있는 장점은 영상 정보 중 하나인 색상(color)에 대한 간섭을 어느 정도 배제할 수 있다는 점이다. 식(1)은 RGB에서 HSI 컬러 모델로 변환하는 식이다.

$$H = \cos^{-1} \left(\frac{\frac{1}{2} [(R-G) + (R-B)]}{[(R-G)^2 + (R-B)(G-B)]^{1/2}} \right)$$

$$S = 1 - \frac{3}{(R+G+B)} [\min(R, G, B)]$$

$$I = \frac{(R+G+B)}{3} \quad (1)$$

2.5 질감 특징 값 추출 및 학습패턴 생성

일반적으로 영상의 내용 정보로 사용되는 질감 특징은 통계적인 특징, 구조적인 특징과 스펙트럼 특징을 이용하여 표현된다. 통계적인 특징은 영상의 거칠고 부드러운 정도를 나타낸다. 구조적인 특징은 영상내의 수평선의 반복과 같은 규칙적인 배열을 표현한다. 스펙트럼 특징은 높은 에너지 성분을 분석함으로써 영상 내의 전체적인 주기성을 알아내는데 사용된다[7].

질감의 특성을 표현하는 요소에는 질감의 미세 정도를 나타내는 거침(coarseness), 명도의 높낮이 분포를 나타내는 대비(contrast), 질감이 가지고 있는 특정한 방향을 나타내는 방향성(directionality) 등이 있다.

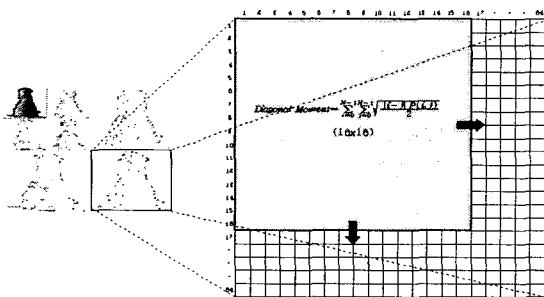
본 연구에서는 이러한 질감 특징들 중, contrast와 diagonal moment, energy, entropy, homogeneity,

second diagonal moment, uniformity를 각각 사용하여 분류기의 분류율을 실험해 보았으며, 실험 결과 그중에서도 diagonal moment가 가장 높은 분류율을 보여 주었다.

각 질감 특징들을 구하는 식은 수식 (2)와 같다.

$$\begin{aligned}
 \text{Contrast} &= \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (i-j)^2 P(i,j) \\
 \text{Diagonal Moment} &= \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \sqrt{\frac{|i-j| P(i,j)}{2}} \\
 \text{Energy} &= \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P(i,j)^2 \\
 \text{Entropy} &= - \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P(i,j) \log(P(i,j)) \\
 \text{Homogeneity} &= \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \frac{P(i,j)}{1+(i-j)^2} \\
 \text{Second Diagonal Moment} &= \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \frac{|i-j| P(i,j)}{2} \\
 \text{Uniformity} &= \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \frac{P(i,j)}{1+|i-j|} \quad (2)
 \end{aligned}$$

질감 특징 값은 [그림 5]와 같이 추출된다. 웨이블릿 변환 후 생성되는 64×64 크기의 대각 영역을 대상으로 특징 값을 추출한다. 16×16 크기의 윈도우에서 하나의 대각 모멘트 값을 획득한다. 그리고 이 윈도우를 좌에서 우로, 위에서 아래로 8픽셀씩 중첩(overlap) 시키면서 이동하여 각각의 영역에서 특징 값을 추출한다. 최종적으로 하나의 영상에서 총 49개의 질감 특징 값을 얻을 수 있다.



[그림 5] 질감 특징 값을 추출하는 예

2.6 계층적 분류기

PCA와 K-means를 이용하여 유사한 특징을

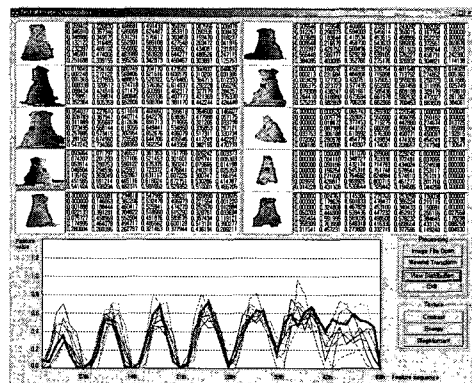
가지는 영상들끼리 군집화 시키고 각각의 군집들을 병렬적으로 학습 및 테스트하는 계층적 분류기를 구성한다.

PCA 변환은 차원 입력 벡터를 저 차원의 형태로 표현하는 방법인데, 이는 전체 영상의 데이터를 가지고 그것들의 분산이 큰 몇 개의 고유 방향에 대한 축으로 선형 투사시켜 차원을 줄이는 방법을 말한다. 이를 통해 데이터 분포에 대한 정보는 그대로 유지하지만 계산상의 부하를 감소시키고 노이즈를 제거하며, 데이터 압축과 같은 효과를 낼 수 있다[8].

K-means는 사전에 정해진 어떤 수의 클러스터를 통해서 주어진 데이터 집합을 분류하는 간단하고 쉬운 방법이다. 초기 k개의 seed point를 정하고 값이 들어올 때 마다 가장 가까운 seed point로 군집화 시키고 군집화 된 값들의 중심 값을 군집의 새로운 중심 값으로 한다. 이렇게 k개의 중심 값이 더 이상 변화가 없을 때 까지 연산을 계속한다[9].

























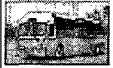



















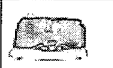



















































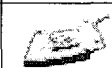



이전 단계에서 추출된 diagonal moment값을 PCA를 이용해서 다른 특징공간으로 변환 시켜서 principal components라는 서로 독립적인 새로운 값을 추출한다. 그중에서 설명 비율이 90% 이상되는 상위 4개의 component를 선택한다.

다음으로 각 영상종류를 대표할 수 있는 영상을 [그림 6]과 같은 특징 값 분포 곡선 중에서 평균인 분포를 나타내는 영상을 선택한다. 이전 단계에서 추출된 4개의 component를 K-means의 특징 값으로 사용해서 영상들을 군집화 시켜서 [그림 7]과 같은 계층도를 구성한다. [그림 8]은 인간 시각에 의한 계층도 구성 그림이다.



[그림 6] 특징 값 분포 예

[표 1] 실험에 사용된 객체영상의 예

									
11	12	13	14	15	16	17	18	19	20
									
21	22	23	24	25	26	27	28	29	30
									
31	32	33	34	35	36	37	38	39	40
									
41	42	43	44	45	46	47	48	49	50
									
51	52	53	54	55	56	57	58	59	60
									
61	62	63	64	65	66	67	68	69	70
									
71	72	73	74	75	76	77	78	79	80
									
81	82	83	84	85	86	87	88	89	90
									
91	92	93	94	95	96	97	98	99	100
									

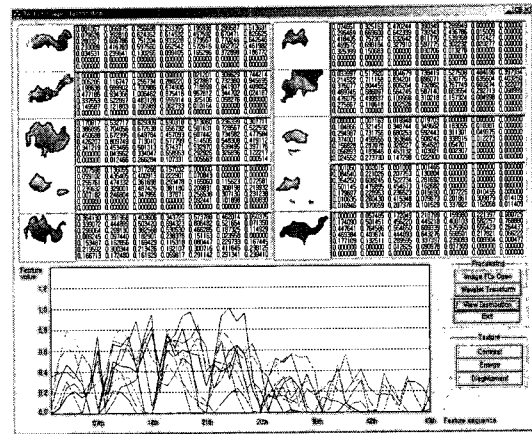
[표2] 30개 영상에 대한 단일분류기와 계층적 분류기의 대한 결과

단일 분류기			
학습시간	학습 데이터	테스트데이터	총 데이터
123h	217/300 (72.3%)	201/300 (67.0%)	418/600 (69.7%)
계층적 분류기			
학습시간	학습 데이터	테스트데이터	총 데이터
23h	255/300 (85.0%)	228/300 (76.0%)	483/600 (80.5%)

인간 시각의 의한 계층적 분류기의 경우 1000개의 학습데이터와 1000개의 테스트 데이터에서 총 53.8%의 분류율을 보여주었다. PCA와 K-means를 이용해서 계층도를 구성했을 때는 학습 데이터의 경우는 81.5%, 테스트 데이터의 경우는 75.1%의 정분류율을 보여주었다.

본 논문에서 제안하는 방법으로 계층도를 구성하였을때 더 좋은 결과를 보여주는 이유는 인간 시각에 의한 방법은 기능, 쓰임, 종류 등이 유사한 영상들을 군집화 시킨다. 그러나 실제 영상 처리에 사용되는 feature들은 기능적인 측면이 아니라 영상자체의 내부정보를 사용한다. 그래서 컴퓨터 시각에 근거하여 영상의 내용정보를 가지고 군집화 시키는 PCA와 K-means를 이용한 방법이 좀 더 높은 분류율을 나타내었다. [표 3]과 [표 4]는 인간 시각에 의한 분류와 PCA와 K-means를 이용한 분류 결과이다.

결과를 분석하여 보았을 때 배경과 객체가 뚜렷이 구분되고 객체내부가 복잡하지 않았을 때 객체영역이 잘 추출되고 배경의 영향을 적게 받아서 좋은 분류결과를 보였다. 그러나 [그림 9]와 같이 객체부분과 배경부분의 유사성과 객체의 복잡성으로 인해 배경을 충분히 제거하지 못했을 때 배경영역의 잡음에 의한 분류율 저하를 나타내었다.



[그림 9] 배경의 복잡성으로 인한 분류율 저하영상의 예

[표 3] 인간 시각에 의한 계층적 분류 결과

질감 특징	학습 데이터	테스트 데이터	총 데이터
Contrast	593/1000 (59.3%)	430/1000 (43.0%)	1023/2000 (51.2%)
Diagonal Moment	623/1000 (62.3%)	453/1000 (45.3%)	1076/2000 (53.8%)
Energy	543/1000 (54.3%)	444/1000 (44.4%)	987/2000 (49.4%)
Entropy	510/1000 (51.0%)	435/1000 (43.5%)	945/2000 (47.3%)
Homogeneity	420/1000 (42.0%)	410/1000 (41.0%)	830/2000 (41.5%)
Second DM	572/1000 (57.2%)	426/1000 (42.6%)	998/2000 (49.9%)
Uniformity	544/1000 (54.4%)	457/1000 (45.7%)	1001/2000 (50.1%)

[표 4] PCA와 K-means를 이용한 계층적 분류 결과

질감 특징	학습 데이터	테스트 데이터	총 데이터
Contrast	680/1000 (68.0%)	627/1000 (62.7%)	1307/2000 (65.4%)
Diagonal Moment	815/1000 (81.5%)	751/1000 (75.1%)	1566/2000 (78.3%)
Energy	644/1000 (64.4%)	567/1000 (56.7%)	1211/2000 (60.6%)
Entropy	617/1000 (61.7%)	543/1000 (54.3%)	1160/2000 (58.0%)
Homogeneity	536/1000 (53.6%)	483/1000 (48.3%)	1019/2000 (51.0%)
Second DM	712/1000 (71.2%)	627/1000 (62.7%)	1339/2000 (67.0%)
Uniformity	701/1000 (70.1%)	621/1000 (62.1%)	1322/2000 (66.1%)

4. 결 론

본 논문에서는 배경을 제거하여 오분류의 원인이 되는 잡음을 최소화하고, 웨이블릿 변환을 이용하여 영상의 형태 정보를 획득하며, 여러 가지 질감 특징을 구조적으로 추출한다. 추출된 질감 특징 값들을 PCA 변환 시킨 다음 K-means로 군집화 시키고, 계층도를 구성한 다음 신경망을 통해 영상을 분류하는 방법을 제안하였다.

계층도를 구성하고 각 계층에서 여러 개의 신경망 분류기가 병렬적으로 학습 및 테스트함으로써 학습 시간을 크게 단축할 수 있었다. 그리고 계층도를 구성함에 있어서 인간 시각에 의하지 않고 컴퓨터 시각으로 구성함으로써 분류 효율을 높일 수 있었다.

앞으로의 과제는 분류율을 높이기 위해 영상의 특징을 좀 더 잘 나타낼 수 있는 특징 값 추출과 복잡한 배경에서 객체를 추출하는 연구를 계속하고자 한다.

참 고 문 헌

[1] Giacinto, G., Roli F. and Didaci, L "Fusion of Multiple Classifiers for Intrusion Detection in Computer Networks." Pattern Recognition Letters, Vol.24, pp.1795-1803, 2003.

[2] 안준현, 조성배 "적합도 공유에 의해 진화한 중분화 신경망" 정보과학회논문지, 제29권, 제1호, pp.0001-0009, 2002.

[3] Deng, Y. and Manjunath, B.S. "Unsupervised Segmentation of Color-Texture Regions in Images and Video." IEEE Transaction on pattern analysis and machine intelligence, Vol.23, pp.800-810, 2001.

[4] Park, S.B., Lee, J.W. and Kim, S.K. "Content-based Image Classification using Neural Network." Pattern Recognition Letters, Vol.25. pp.287-300, 2004.

[5] 김진숙, 김호성, "색과 공간정보를 이용한 영상 검색 기법," 한국정보과학회 '97 추계

학술발표논문집, 제24권, 제2호, pp.483-486, 1997.

[6] 강영미, "Wavelet과 신경망을 이용한 내용 기반 얼굴 검색 시스템," 창원대학교 석사논문, 1998.

[7] Ramesh Jain, Rangachar Kasturi, and Brain G. Schunck "Machine Vision" ISBN 0-07-032018, 1995.

[8] Turk, M., and Pentland, A "Eigenfaces for Recognition" Journal of Cognitive Neuroscience, Vol.3. pp.71-86, 1991.

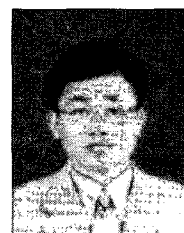
[9] Eari Gose, Richard Johnsonbaugh, and Steve Jost, "Pattern Recognition and Image Analysis" ISBN 0-13-236415-8, 1996.



김 종 호 (Jong-Ho Kim)

정회원

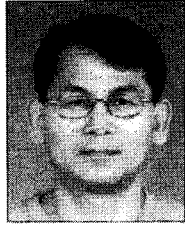
- 2002년 인제대학교 정보컴퓨터학부 졸업(정보컴퓨터학사)
- 2004년 인제대학교 대학원 전산학과 졸업(전산학석사)
- 2004년~현재 인제대학교 대학원 전산학과 박사과정
- 관심분야 : 정보검색, 패턴인식, 컴퓨터비전



김 상 균 (Sang-Kyoon Kim)

정회원

- 1991년 경북대학교 통계학과 졸업(이학사)
- 1994년 경북대학교 대학원 컴퓨터공학과 졸업(공학석사)
- 1996년 경북대학교 대학원 컴퓨터공학과 졸업(공학박사)
- 1996년~현재 인제대학교 컴퓨터공학부 부교수
- 관심분야 : 패턴인식, 정보검색, 정보보호, 컴퓨터비전



신 범 주 (Bum-Joo Shin)

정회원

- 1983년 2월 : 경북대학교 전자공학과 (공학사)
- 1991년 2월 : 경북대학교 컴퓨터공학과 (공학석사)
- 1998년 8월 : 경북대학교 컴퓨터학과 (공학박사)
- 2006년 3월~현재 : 부산대학교 바이오시스템학부 교수
- 관심분야 : 시스템소프트웨어