

계층적 신경망을 이용한 객체 영상 분류

김종호*, 이재원*, 김상균**

*인제대학교 전산학과

**인제대학교 컴퓨터공학부

e-mail : lucky@cs.inje.ac.kr

An Object Image Classification Using Hierarchical Neural Network

Jong-Ho Kim*, Jae-Won Lee*, Sang-Kyoon Kim**

*Dept of Computer Science, Inje University

**Dept of Computer Engineering, Inje University

요 약

본 연구는 웨이블릿 변환을 통하여 객체 영상에서 질감 특징값을 추출하고, 신경망을 계층적으로 구성하여 분류하는 방법을 제안한다. 기존의 신경망을 이용한 영상의 분류는 단일 신경망을 이용하는 것이 대부분이었다. 하지만 단일 신경망은 분류하고자 하는 클래스의 수가 많거나 분류하고자 하는 대상이 유사한 입력패턴을 가질 경우 학습시간이 오래 걸리고, 인식률이 크게 떨어지는 문제를 가지고 있다. 그래서 본 연구에서는 효과적인 객체 영상 분류를 위해서 여러개의 단일 신경망을 계층적으로 결합하는 방법을 제안한다. 실험결과 분류 대상 클래스가 증가함에도 불구하고 단일 신경망에 비해 학습시간이 단축되고, 높은 인식률을 보여주었다.

1. 서론

신경망은 복잡하게 상호 연결된 뉴런의 망으로 구성된 생체 학습 시스템에서 영감을 받은 것으로, 이런 생물학적 처리과정을 빌어 매우 효율적인 기계 학습을 얻을 수 있다. 이러한 학습 특성 때문에 신경망은 여러 분야에서 사용되는데, 특히 학습을 통한 인식 및 분류 분야에서 탁월한 성능을 발휘한다.

정보 분류에 있어서 신경망을 이용한 기존의 연구방법들은 단일 신경망을 사용하거나, 신경망의 초기 가중치나 위상 구조를 다양하게 하거나, 각 신경망을 서로 다른 학습 데이터나 학습 알고리즘을 사용하고 각 신경망에서 나온 결과를 투표나 평균 방법을 이용하여 결합해서 결과를 도출하는 방법을 사용하고 있다 [1][2].

그러나 기존의 방법들은 분류하고자하는 대상 클래스의 수가 증가하면 뉴런의 연결 강도 수가 급격하게 증가해서 학습시간이 오래 걸리고 학습효율 또한 떨어진다[3]. 분류하고자 하는 클래스 중에 유사한 입력 패턴을 가진 클래스가 많을 경우에도 분류

효율이 떨어진다.

그래서 본 연구에서는 이러한 단점을 극복하기 위해서 기존의 단일 신경망을 계층적으로 결합한다. 개별 신경망의 분류 클래스 수를 줄여서 학습 시간을 단축하고, 유사한 입력 패턴을 가진 대상을 분리해서 학습시킴으로써 인식률을 높이는 계층적 신경망을 제안한다.

2. 영상 분류기의 구성

본 연구에서 제시하는 영상 분류기의 전체 구성 절차는 그림 1과 같다.

먼저 정규화 단계에서는 인터넷에서 획득한 JPEG 포맷의 영상 파일을 BMP 포맷으로 변화한다. 그리고 영상 정보를 신경망의 학습패턴으로 사용하기 위해 영상 크기를 128×128로 정규화 한다. 다음으로 영상에서 특징값 추출을 위한 전처리로써 1단계 웨이블릿 변환을 한다. 그리고 RGB 컬러모델에서 HSI 컬러모델로 변환한 뒤 명도(intensity)값을 사용한다. 다양한 질감 특징들 중 실험적으로 가장

좋은 인식률을 보여주는 대각 모멘트(diagonal moment) 특징값을 추출한다. 추출된 49개의 특징값들을 학습패턴으로 구성하고, 최종적으로 이 학습패턴과 신경망 분류기를 이용해서 학습 및 분류한다

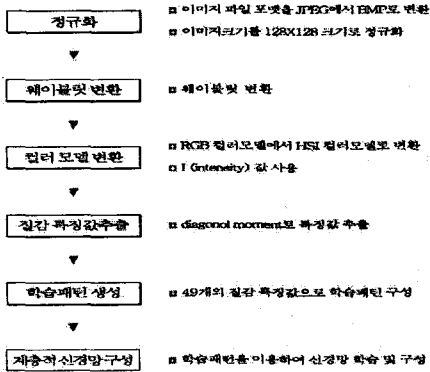


그림 1. 시스템 전체 구성도

2.1 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 정규화한다.

2.2 웨이블릿 변환

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

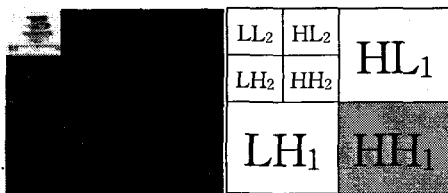


그림 2. 웨이블릿 분할영역

웨이블릿 변환을 이차원 영상에 2단계 변환하였을 때 그림 2와 같이 7 개의 성분으로 나누어진다. 이 7 개의 성분은 시각에 민감한 정보를 가지고 있

으며, 크게 저주파 성분과 고주파 성분으로 나누어지며, 고주파 성분에는 또다시 수직 성분, 수평 성분, 대각 성분으로 나누어진다. 본 연구에서는 가장 정보의 양이 많은 대각 성분(HH1)을 사용하였다.

2.3 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.

따라서 본 연구에서는 많은 영상 관련 어플리케이션에서 사용되고 있는 HSI 컬러 모델로 변환하여 그 중 명도(intensity) 값을 사용하였다.

2.4 질감 특징 추출

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[4][5].

본 연구에서는 이러한 질감 특징들 중 실험적으로 가장 높은 인식률을 보여준 대각 모멘트(diagonal moment)를 사용하였다.

$$Diagonal\ Moment = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \sqrt{\frac{|i-j|P(i,j)}{2}} \quad (1)$$

질감 특징값은 그림 3과 같이 추출된다. 웨이블릿 변환후 생성되는 64×64 크기의 대각 영역을 대상으로 특징값을 추출한다. 16×16 크기의 윈도우에서 하나의 대각 모멘트값을 획득한다. 그리고 이 윈도우를 좌에서 우로, 위에서 아래로 8픽셀씩 중첩

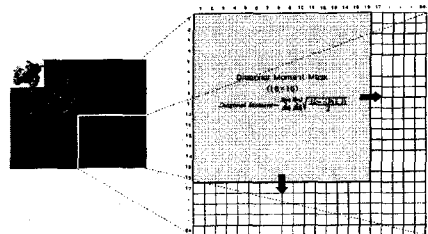


그림 3. 질감 특징값 추출 예

(overlap) 시키면서 이동하여 각각의 영역에서 특징값을 추출한다. 최종적으로 하나의 영상에서 총 49개의 질감 특징값을 얻을 수 있다.

2.5 분류기의 구성

본 연구에서는 신경망 학습 알고리즘들 중에서 여러개의 역전파(BP:Back Propagation) 알고리즘을 계층적으로 구성하여 사용하였다. 역전파 알고리즘은 순방향 다층 신경망의 학습에 효과적으로 사용할 수 있어서 다양한 분야에 가장 널리 활용되는 학습 알고리즘이다. 출력층의 오차 신호를 이용하여 은닉층과 출력층간의 연결 강도를 변경하고, 또한 출력층의 오차 신호를 은닉층에 역전파하여 입력층과 은닉층과의 연결강도를 변경하는 학습 방법이다. 계층적 신경망의 구성은 그림 4와 같다.

본 연구에서 그림 4와 같이 역전파 알고리즘을 계층적으로 사용한 이유는 여러개의 신경망을 동시에 학습시킴으로써 학습시간을 단축시킬수 있고 개별 신경망의 분류수를 줄일수 있어서 분류 효율성을 높일수 있다. 또한 새로운 클래스가 추가되었을 때 단일 신경망의 경우에는 전체를 다시 학습시켜야 되는데 반해, 계층적 신경망은 상위레벨과 하위레벨의 추가된 클래스에 대한 신경망만 학습하면 된다. 그리고 단일 신경망의 경우 그림 5와 같이 유사한 질감 특징값을 가지는 영상의 경우 분류가 힘들었으나 계층적 신경망에서는 유사한 질감 특징값을 가지는 영상을 분리하여 학습시킴으로써 분류 효율성을 높

일수 있다.

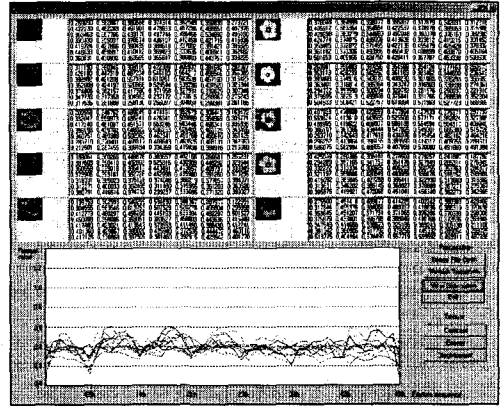


그림 5. 유사한 질감 특징값을 가지는 예

3. 실험 및 결과 분석

3.1 실험환경

이 실험은 단일 신경망을 사용했을 때와 본 연구에서 제안한 계층적 신경을 사용했을 때와의 인식률과 학습시간을 비교한다.

이 실험을 위하여 펜티엄 III 1GHz PC, 윈도우즈 환경에서 비주얼 C++을 사용하여 실험 및 분석을 위한 시스템을 구성하였다.

그림 6은 신경망 분류기의 학습 및 분류에 사용한 영상 종류들이다. 인터넷 상의 다양한 영상들 중에 객체와 배경의 구분이 있는 객체 영상들을 실험 데이터로 사용하였다. 그리고 각 종류의 질감 특징값을 분석해서 유사한 질감 특징값을 가지는 영상은 분리해서 학습하였다. 실험에 사용한 영상의 종류는 30종이다.

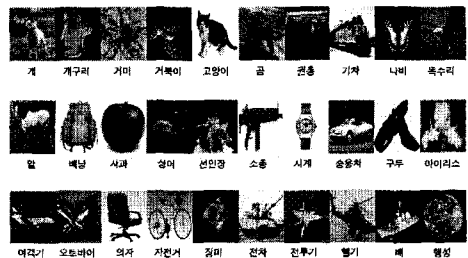


그림 6 실험에 사용한 영상의 종류

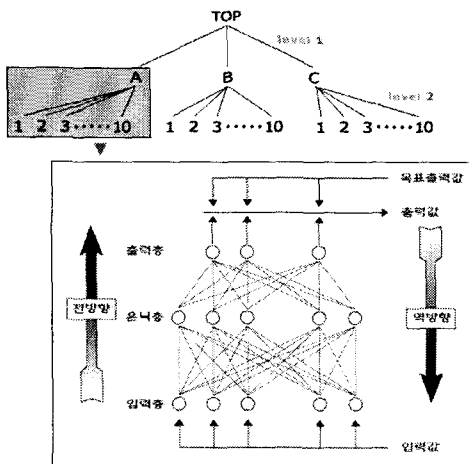


그림 4. 계층적 신경망

3.2 결과 분석

영상 종류당 10개씩 30종류, 총 300개의 영상을 사용하여 300개 전부를 하나의 신경망으로 학습시키

고 분류하는 단일 신경망과 첫 번째 레벨에서는 3종류, 그리고 두 번째 레벨에서는 10종류를 분류하는 3개의 신경망을 가지는 계층적 신경망으로 학습과 분류 실험을 하였다. 두 가지의 실험을 비교한 결과는 표 1과 같다.

거나 복잡한 경우 인식률이 상당히 저하되었다. 이 문제를 해결하기 위해 차후 영상에서 배경을 제거하여 객체의 인식률을 높이는 방법을 연구해보고자 한다. 그리고, 보다 많은 영상 종류를 계층적 신경망을 이용하여 효율성을 실험해 보고자 한다.

표 1. 단일 신경망과 계층적 신경망의 실험결과 비교

단일 신경망(에러율 0.0001)				
학습시간		학 습 데이터	테 스트 데이터	총 데이터
123h		217/300 (72.3%)	201/300 (67.0%)	418/600 (69.7%)
계층적 신경망(에러율 0.0001)				
학습시간		학 습 데이터	테 스트 데이터	총 데이터
1	0.5h	300/300 (100%)	289/300 (96.3%)	589/600 (98.1%)
2-1	21h	82/100 (82.0%)	76/100 (76.0%)	158/200 (79.0%)
2-2	25h	84/100 (84.0%)	79/100 (79.0%)	163/200 (81.5%)
2-3	23h	79/100 (79.0%)	76/100 (76.0%)	155/200 (77.5%)

참고문헌

- [1] Giorgio Giacinto, Fabio Roli, Luca Didaci "Fusion of multiple classifiers for intrusion detection in computer networks" Pattern Recognition Letter 2003
- [2] 안준현, 조성배 "적합도 공유에 의해 진화한 중분화 신경망" 한국정보과학회 2000
- [3] 오창석 "뉴로 컴퓨터 개론" 내하출판사, 2000
- [4] Ramesh Jain, Rangachar Kasturi, and Brain G. Schunck "Machine Vision" ISBN 0-07-032018 1995
- [5] 김희승 "영상인식-영상처리, 컴퓨터비전, 패턴 인식, 신경망" 생능출판사, 1994

실험 결과, 단일 신경망의 경우 학습시간이 123시간이 걸렸지만 계층적 신경망의 경우 산술적으로 계산하더라도 학습시간을 1/2정도로 줄일 수 있다. 그리고 인식률에서도 계층적 신경망이 효율적인 것을 확인할 수 있다. 그리고 본 연구에서 학습시간이 오래 걸리는 이유는 서론에서도 언급했듯이 인식률을 높이기 위해 입력층, 은닉층, 출력층 각 층마다 다수의 노드를 사용했기 때문이다.

4. 결론

본 연구에서는 웨이블릿 변환을 이용하여 영상내의 중요한 정보를 획득하고, 여러 가지 질감 특징들 중에서 실험적으로 가장 우수한 분류 결과를 나타내는 대각 모멘트(diagonal moment)를 추출하여 이를 학습패턴으로 하는 계층적 신경망을 통해 객체 영상을 분류하는 방법을 제안하였다. 실험 결과, 단일 신경망으로 실험하였을 때보다 학습시간이 단축되고, 높은 인식률을 보여주었다.

하지만 배경영역이 객체와 유사한 질감값을 가지