신경망을 이용한 영상의 웨이블렛 특징공간과 감성공간의 매핑

정윤경, 조성배
연세대학교 컴퓨터과학과

Mapping Wavelet Feature Space to KANSEI Space in Image Using Neural Networks

Yun-Gyoung Chong and Sung-Bae Cho
Computer Science Department, Yonsei University

요 약

복합적인 감성기반 영상 검색 시스템을 구축하기 위해서는 감성속성으로 영상을 찾는 검색은 물론이고, 주어진 영상의 감성특성을 알아내는 과정이 필요하다. 본 논문에서는 영상의 특성으로부터 감성을 매핑하는 신경망을 구축하고 다양한 실험으로 그 가능성을 보인다. 여기에서 영상특징으로 웨이블렛계수와 위치정보를 사용했고, 감성공간으로는 SD법으로부터 14개의 형용사쌍을 추출했다. 이 두 공간의 매핑에 사용된 신경망의 입력으로 영상에서 얻은 RGB색상당 36개의 총 108개의 웨이블렛 개수를 사용했고, 출력은 14개의 감성속성당 7등급으로 총 98개로 구성했다. 총 6명이 영상을 보고 평가한 감성평가데이터중에서 2명이 각각 평가한 500개씩의 데이터로 신경망을 학습시키고 나머지 100개로 테스트한 결과중에서 96%의 인식률을 보였다. 4명이 각각 90개씩 평가한 데이터로 신경망을 학습시키고 나머지 10개로 테스트한 경우는 90%의 인식률을 보였다. 또한 공통된 감성을 신경망을 통해 인식할 수 있는지 판단하기 위해 600개씩 2명으로부터 얻은 1200개의 데이터에 대해서 1000개를 학습시키고 200개를 테스트하고, 100개씩 4명으로부터 400개의 데이터에 대해서 360개를 학습시키고 40개를 테스트해 본 결과, 전자의 경우 오류율 8, 후자의 경우 0.7~0.8범위였다.

1. 서론

멀티미디어 기술이 발달하고 하드웨어적인 요소가 발전함에 따라 대용량의 영상 데이터가 생성되고 있으며 이를 보다 효율적으로 유지하고 관리하는 기술이 필요하게 되었다. 영상검색 방법은 색인어 방법, 영상의 질감, 색상, 형태 등의 특징을 이용하는 방법 등을 들 수 있다[1]. 위 방법들은 구체적인 대상에 대해서 검색이 가능하지만 추상적이거나 사람의 주관이 들어 있는 대상의 경우 검색이 불가능하다.

이 문제에 대해 Tagaki는 감성을 이용한 영상검색 방법을 제시했다[2]. 그림1은 그가 제시한 신경망과 유전자 알고리즘을 이용하여 영상공간과 감성공간의 매핑구조를 보여준다. 본 논문은 신경망을 이용하여 감성 공간과 영상을 매핑하여 감성과 영상의 연관성을 찾고 감성을 통해 영상을 검색하는 방법을 제시하고, 실험을 통해 그 가능성을 보인다.

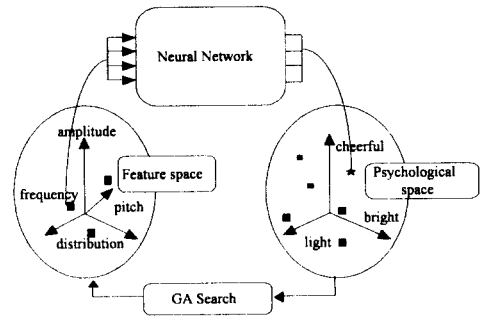

그림 1. 영상공간과 감성공간 매핑

2. 관련연구

2.1 감성

이제까지 감성은 단지 이성을 보조하기 위한 수단으로 여겼다. 현재 감성에 대한 연구는 생활을 다양하게 변화시킨다는 점에서 그 의미가 커지고 있다[3]. 감성을 단지 정보 획득의 수단으로 여기던 과거와 달리 다양한 삶의 욕구를 충족시키기 위한 감각, 감성, 정서, 욕구 등의 종합적인 개념으로 보게 되었으며 감성을 색채 패턴 분류, 의류 제조 등과 같은 공학적인 문제에 이용하게 되었다[3,4].

2.2 웨이블렛 변환

웨이블렛 변환은 그 진폭이 양방향으로 급격하게 0으로 줄어드는 특성을 갖는 웨이블렛을 기저함수로 이용하여 데이터를 변환시키는

것을 말한다[5]. 이것은 영상을 적은 수의 매개변수로 변환시켜도 전체 영상을 포괄할 수 있다. 또한 속도가 빠르고, 압축율이 높기때문에 음성인식, 정영상 및 동영상 압축, 오디오 압축, 음악 해석 등에 이용할 수 있다[6].

3. 특징공간과 감성공간의 매핑

시스템의 입출력공간으로 영상과 감성의 공간을 구성했으며, 신경망을 학습시키고 테스트하기 위해 사용자가 영상에 대해 직접 입력한 평가데이터를 이용했다. 평가데이터에 사용된 영상은 2000개중에서 무작위로 선택했다. 사용자는 선택된 영상에 대해 14쌍의 감성속성을 평가했다. 1쌍의 감성속성에 대한 평가단계로 7등급을 두었다. 시스템은 사용자가 직접 평가한 감성데이터와 영상 특성데이터를 이용하여 학습시켰다. 학습이 끝난후 테스트데이터에 대해 시스템이 평가한 감성데이터를 얻었다. 시스템에 의해 생성된 감성데이터는 데이터베이스에 저장된다. 전체적인 시스템 구성은 그림2와 같다.

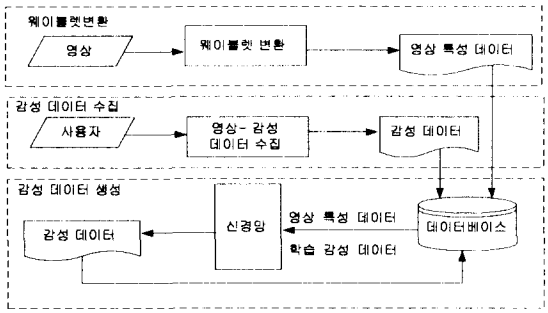

그림 2. 시스템구조

3.1 영상특징 추출

영상특징으로는 원영상으로부터 Haar 웨이블렛 변환을 이용하여 생성된 웨이블렛계수를이용했다. 이때 웨이블렛계수 일부로 전체영상을 표현하는 성질을 이용하기 위해 웨이블렛 변환시킨 $r \times r$ 영상행렬에 대해 r을 변화시켜가며 원영상의 에너지가 높은 계수들의 분포를 살펴보았다. 그림3에서 보는 것과 같이 6×6과 7×7행렬일 때 원영상의 에너지를 표현함을 볼 수 있었다. 따라서 본 논문은 웨이블렛 변환된 6×6영상을 특징으로 사용하였다.

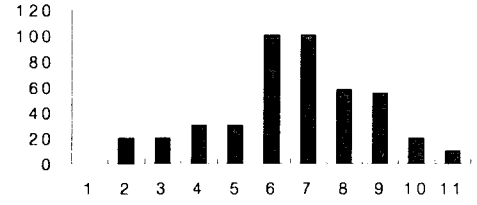

그림 3. 웨이블렛계수 에너지 분포

3.2 감성공간 선택

감성공간 구성의 경우 Tagagi의 연구에서처럼 SD법을 사용하여 얻은 14개의 형용사 쌍을 이용했다[2]. 형용사쌍은 그림4와 같다. 형용사쌍의 등급은 7개인데, 예를 들어 '밝다,' '어둡다'의 형용사에 대해서 7개의 등급은 '매우 밝다,' '밝다,' '약간 밝다,' '보통,' '약간 어둡다,' '어둡다,' '매우 어둡다'의 등급이 된다.

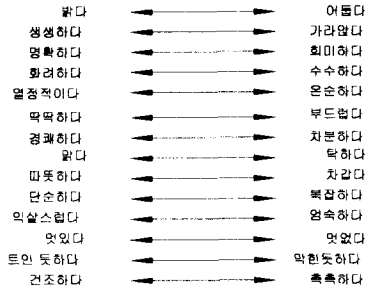

그림 4. 14개의 감성속성

3.3 감성평가 데이터 수집

감성평가 데이터는 6명의 사용자로부터 수집했다. 감성데이터집합이 많은 부류와 작은 부류로 나누어 6명의 사용자중 2명에 대해서는 600개의 영상 데이터를, 4명에 대해서는 100개의 영상을 평가하도록 하였다. 6명의 사용자에 대해서 이미지 파일은 데이터베이스의 2000개의 이미지 파일목록에서 무작위로 각각 다른 영상을 선택했다. 평가레벨과 감성평가 공간은 시스템에서 사용하는 감성공간을 이용했다.

3.4 신경망 매핑

신경망은 인간의 두뇌를 수학적으로 모델링하여 만든 인공적인 두뇌를 공학이나 기타 여러분야에 적용시킨 것이다[7]. 신경망 중 BP(Back Propagation)은 LMS법칙을 일반화시킨 것으로 신경망의 실제 출력과 이상적인 출력사이의 평균차이를 최소화하는 경사 탐색기법을 사용한다. 알고리즘은 다음과 같다[8].

① 입력층 신경 세포에 입력 패턴을 제시

② 신경망을 동작(바이어스를 사용할 경우 다음을 포함시킴)

$$a_j = \frac{1}{1+\exp(x_j + bias_j)}$$

③ 출력층 신경 세포들의 오류 e_j와 델타 δ_j를 구해 이를 은닉층으로 역전파시킴: $e_j = t_j - a_j$, $\delta_j = a_j(1-a_j)e_j$

④ 역전파된 델타로부터 은닉층 신경 세포들의 에러와 델타를 구해 이를 다시 역전파시킴: $e_j = \sum_k W_{jk}\delta_k$, $\delta_j = a_j(1-a_j)e_j$

⑤ ④과정을 입력층 바로 위층까지 반복

⑥ 출력층과 모든 은닉층의 오류와 델타를 구함

위 과정을 신경망이 완전히 학습될 때까지 반복한다.(δ_j: 신경세포 j의 델타값, t_j: 출력층의 신경세포 j의 목표값, a_j는 출력층 신경세포의 활성값, e_j: 신경 세포 j의 에러값)

본 논문에서는 입력층으로 RGB에 대한 영상 특성 36개를 모아 총 108개 노드를 구성했으며 출력층으로 98개의 감성 특성노드를 사용했다.

4. 실험

4.1 실험환경

이 실험은 신경망을 통해 개인의 영상과 감성의 연관성을 찾고, 여러 사용자에 대한 일관된 감성을 찾는 실험이다. 시스템은 그림5부터 그림8까지 4가지 상황에 대해 학습을 시키고성능을 분석한다.

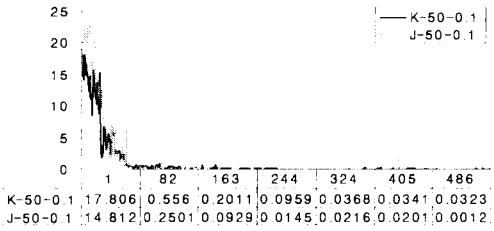

그림 5. 2명이 평가한 감성데이터중 500개의 학습데이터, 100개의 테스트데이터로 사용시 오류율

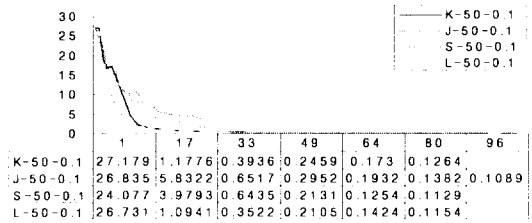

그림 6. 4명이 평가한 감성데이터중 90개의 학습데이터, 10개의 테스트데이터로 사용시 오류율

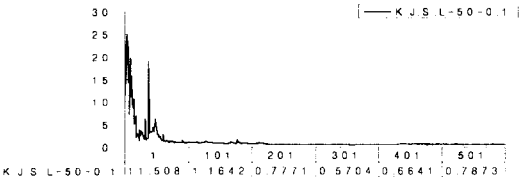

그림 7. 2명의 감성데이터 400개중 360개를 학습데이터, 40개를 테스트데이터로 사용시 오류율

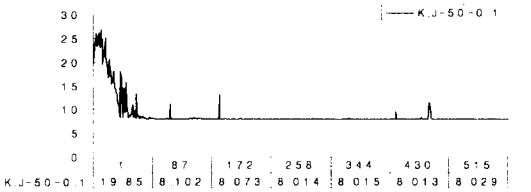

그림 8. 2명의 감성데이터 1200개중 1000개를 학습데이터, 200개의 테스트데이터로 사용시 오류율

4.2 결과분석

각각 2명에 대해 500개의 데이터를 학습시키고 100개의 데이터를 테스트 했을 경우 96% 이상의 인식율을 보였다. 4명에 대해 90개의 데이터를 학습시키고 10개를 테스트했을 때 3명에 대해서는 90%이상의 인식율을 보였으며, 1명에 대해서는 80%의 인식율을 보였다. 이것은 학습시킨 영상데이터 중 테스트영상과 유사한 영상이 없었기 때문이라 추정된다. 다른 사람과 감성데이터를 섞었을 경우 2명에 대해서 8.0, 4명에 대해서 0.7~0.9로 오류율로 여러 사용자에 대해서 공통된 영상에 대한 감성평가를 얻을 수 없었다.

그림9는 그림5의 경우에 대해 시스템을 학습시키고 난 뒤 테스트 데이터에 대해 평가한 감성데이터로 찾은 '따뜻한'영상이다.

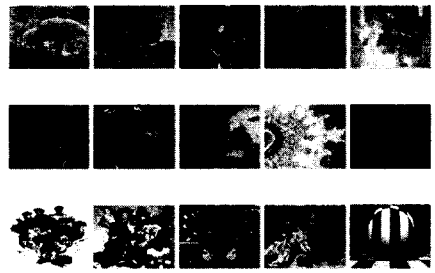

그림 9. 신경망이 평가한 '따뜻한' 영상의 예

5. 결론

실험결과 우리는 개개인에 대한 학습으로 영상을 감성추출하고, 신경망을 통해 평가한 감성으로 영상을 검색할 수 있음을 알았다. 개개인의 경우 90%이상의 인식율을 보여 개인의 감성을 추출할 수 있었으나 여러 사람의 공통된 감성을 추출하지 못했다. 80%의 감성축출율을 보인 사용자는 테스트에 사용된 영상특성이 학습에 사용된 영상의 특성중에 포함되지 않을 가능성이 있다. 이것은 사용자의 평가데이타를 수집할 때 영상을 무작위로 선정하여 영상특징이 한쪽으로 치우쳤기때문이라 사료된다. 향후에는 본 논문의 결과로 얻어진 영상-감성매핑 모듈과 대화형 유전자 알고리즘으로 얻어진 감성-영상 매핑 모듈을 결합하여보다 완벽한 감성기반 영상 검색 시스템을 구축하여야 할 것이다.

참고 문헌

[1] V. Gudivad and V. V. Raghavan, "Content-based Image Retrival Systems," *IEEE Computer*, pp. 18~32, 1995.

[2] H. Takagi, T. Noda and S.-B. Cho, "The Psychological Space of Common Media Impressions Held in a Media Database Retrieval System," *SMC*, pp. 263~268, 1999.

[3] 이순요, 양선모 공저, 감성공학, 청문각, pp. 5~11, 1996.

[4] 두성한, 엄경배, 이준환, "적응 퍼지 시스템을 이용한 칼라 패턴 감성 평가 모델에 관한 연구," 한국 퍼지 및 지능시스템학회 논문지, Vol. 9, No. 5, pp. 526~537, 1999.

[5] T. Edwards, *Discrete Wavelet Transforms: Theory and Implementation*, Stanford University, 1991.

[6] 이문호, 염재훈 공저, 영상신호처리, 대영사, 1998.

[7] R. P. Lippmann, "An Introduction to Computing with Neural Nets," *IEEE ASSP Magazine*, pp. 4~22, 1987.

[8] L. Fausett, *Fundamentals of Neural Networks*, Prentice Hall, pp. 289~333, 1994.