

# 신경망을 이용한 텍스타일 영상에서의 감성인식 시스템

## (Emotion Recognition System Using Neural Networks in Textile Images)

김나연<sup>†</sup>    신윤희<sup>†</sup>    김수정<sup>\*\*</sup>    김지인<sup>\*\*\*</sup>  
 (Nayeon Kim)    (Yunhee Shin)    (Soojeong Kim)    (Jeein Kim)

정갑주<sup>\*\*\*</sup>    구현진<sup>\*\*\*\*</sup>    김은이<sup>\*\*\*</sup>  
 (Karpjoo Jeong)    (Hyunjin Koo)    (Eunyi Kim)

**요약** 본 논문에서는 신경망을 이용하여 텍스타일 영상으로부터 인간의 감성을 인식할 수 있는 시스템을 제안한다. 자동감성인식 시스템의 구현을 위해 220장의 텍스타일 영상을 수집한 후, 일반인 20명을 대상으로 설문조사를 실시하였다. 이를 통해 텍스타일 영상에서의 패턴과 감성간의 상관관계, 즉 특정 패턴이 사람의 감성에 영향을 준다는 것을 발견하였다. 따라서 본 연구에서는 텍스타일 영상에 포함된 패턴의 인식을 위해 신경망을 이용하였으며, 이때 패턴 정보의 추출을 위해 두 가지 특징 추출 방법을 사용한다. 첫 번째는 auto-regressive method를 이용한 raw-pixel data extraction scheme(RDES)을 사용하는 것이고, 두 번째는 wavelet transformed data extraction scheme(WTDES)을 사용하는 것이다.

제안된 시스템의 유용성을 증명하기 위해서 실제 100장의 텍스타일 영상을 감성을 인식하는데 사용했다. 그 결과 RDES와 WTDES를 사용했을 때 각각 71%와 90%의 인식률로, WTDES를 사용했을 때가 RDES를 사용했을 때보다 더 좋은 성능을 보였다. 데이터 추출방법에 따라 다소 차이가 있었지만 전체적으로 약 81%의 정확도를 보였다. 이러한 실험 결과는 제안된 방법이 감성인식 기반으로 텍스타일 데이터를 검색 할 수 있는 시스템 및 다양한 산업 분야에 응용 가능성을 보여주었다.

**키워드** : 감성인식, 패턴인식, 신경망, 특징추출, 웨이블릿 변환

**Abstract** This paper proposes a neural network based approach for automatic human emotion recognition in textile images. To investigate the correlation between the emotion and the pattern, the survey is conducted on 20 peoples, which shows that a emotion is deeply affected by a pattern. Accordingly, a neural network based classifier is used for recognizing the pattern included in textiles. In our system, two schemes are used for describing the pattern: raw-pixel data extraction scheme using auto-regressive method (RDES) and wavelet transformed data extraction scheme (WTDES).

To assess the validity of the proposed method, it was applied to recognize the human emotions in 100 textiles, and the results shows that using WTDES guarantees better performance than using RDES. The former produced the accuracy of 71%, while the latter produced the accuracy of 90%. Although there are some differences according to the data extraction scheme, the proposed method shows the accuracy of 80% on average. This result confirmed that our system has the potential to be applied for various application such as textile industry and e-business.

**Key words** : Emotion recognition, neural networks, pattern recognition, feature extraction, wavelet transform

· 본 연구는 서울시 산학연 협력사업의 지원에 의해 수행되었음

<sup>†</sup> 학생회원 : 건국대학교 신기술융합학과  
 yeon0830@konkuk.ac.kr  
 ninharsa@konkuk.ac.kr

<sup>\*\*</sup> 정회원 : 건국대학교 인터넷 미디어공학부  
 cryolite@konkuk.ac.kr

<sup>\*\*\*</sup> 종신회원 : 건국대학교 신기술융합학과 교수

jnk@konkuk.ac.kr  
 jeongk@konkuk.ac.kr  
 eykim@konkuk.ac.kr

<sup>\*\*\*\*</sup> 정회원 : FITI 신뢰성평가팀 책임연구원  
 koohh@fiti.re.kr

논문접수 : 2006년 9월 13일  
 심사완료 : 2007년 6월 12일

## 1. 서 론

감성인식은 기계로 하여금 인간이 느끼는 감성을 자동으로 식별하게 하는 것이다. 이러한 감성인식은 최근 다양한 시스템에 결합되어 사용되고 있다[1,2]. 먼저 사용자의 얼굴표정과 제스처 및 음성으로부터 감성을 인식함을 통해 사용자에게 보다 편리하고 직관적인 인터페이스를 제공하기 위한 목적으로 사용되고 있다[3-7]. 감성인식을 기반으로 한 인터페이스는 일반 사용자뿐만 아니라 신체적인 불편함을 가진 사람 및 노인을 위한 인터페이스로도 활용될 수 있기 때문에 최근 활발한 연구가 이루어지고 있는 연구 분야이다.

또한 감성인식은 멀티미디어 검색 시스템과 결합하여 사용자에게 내용기반 검색 및 조작을 제공할 수 있다[8]. 즉 임의의 영상 데이터에 포함된 감성을 자동으로 인식하여 감성을 기준으로 영상을 색인하고, 사용자의 질의에 대해 데이터를 검색하여 서비스 할 수 있게 한다. 특히 텍스타일 영상에 대한 감성인식기반의 검색 시스템은 사용자가 바라는 감성의 천을 자동으로 선택할 수 있게 하기 때문에 사용자의 요구대로 직물의 기획과 생산을 가능하게 한다. 따라서 인터넷 쇼핑물의 패션업체나 섬유산업, 그리고 디자인 관련 분야 등의 많은 응용분야에 적용될 수 있다. 이러한 중요성과 다양한 응용 분야 때문에 최근 많은 연구자들로부터 주목을 받고 있다.

자동화된 감성기반의 텍스타일 인덱싱을 위해서는 주관적인 감성 표현은 객관적이고 일반적인 감성 정보로 변환되어야 한다. 또한, 대중적인 취향, 유행과 같은 일시적인 조건에 따라 변화되는 감성을 즉각적으로 처리하기 위한 능력이 요구되어 지는데, 이러한 요구를 해결하기 위해서는 효과적인 자동 시스템을 필요로 한다.

이와 관련한 연구로, Kobayashi는 색 등의 다양한 시각적 요소들로부터 받는 심리적인 영향에 대해 연구해 왔으며, 이러한 요소들 간의 상호관계를 조사하긴 했지만, 영상으로부터 이러한 요소들을 자동으로 추출하고, 추출된 정보를 자동으로 분류하는 것에 대한 연구는 이루어지지 않았다. Soen은 주파수성분과 밝기 및 색조 정보를 이용하여 13가지 감성을 인식하는 자동 시스템을 개발하였다[10]. 시스템 평가를 위해 10명에 대해 임의의 텍스타일에서의 감성을 조사하였고, 이를 이용하여 정확도를 조사하였다. 실험결과는 주파수 정보 및 컬러 정보가 감성을 인식하기 위한 유용한 정보이긴 하지만, 낮은 정확률 때문에 실제 문제에 적용하기에는 역시 한계를 가지고 있다.

이 외에, 자동으로 추출하는 연구가 진행되어 왔지만, 대부분의 연구에서는 단일 색상 및 패턴을 가지는 인위적인 영상을 사용하기 때문에, 실제 섬유산업 관련 영역에 적용하는데 문제가 있었다.

지난 연구에서 본 연구자는 색상 정보와 감성과의 상관관계를 조사하여, 색상 정보를 이용한 퍼지 기반의 자동감성인식 시스템을 개발하였다[2,9]. 그러나 이 연구에서는, 특정 감성만이 색상 정보에 의해서 영향을 받기 때문에 전체 텍스타일 영상을 감성에 의해 색인 하는 데는 한계가 있다는 것을 보여주었다[10].

본 논문에서는 임의의 입력된 텍스타일 영상에 포함된 감성을 자동으로 인식하기 위한 신경망 기반의 새로운 감성인식 시스템을 제안한다. 제안된 방법의 기본 가정은 영상에 나타나는 특정 패턴이 사람의 특정 감성에 많은 영향을 줄 수 있다는 것이다. 이러한 가정의 정당성을 증명하기 위해서 먼저 패턴과 감성과의 상관관계를 조사하였다. 패턴과 감성과의 관계 조사를 위해 먼저 패턴북<sup>1)</sup>으로부터 총 220개의 텍스타일 데이터를 수집한 후, 이 데이터를 패턴에 따라 분류하였다. 분류된 각각의 데이터에 대해 일반인 20명을 대상으로 설문조사를 실시하였으며, 그 결과의 분석을 통해서 특정 패턴이 몇 개의 감성과 높은 상관관계를 가지고 있다는 것을 증명하였다. 이러한 사실을 바탕으로 본 논문에서는 자동 감성인식 시스템을 개발한다.

본 논문에서는 영상에 포함된 패턴의 인식을 위해 신경망을 이용한다. 신경망의 입력을 위해 두 가지 추출 방법을 사용하여 텍스타일 영상으로부터 특징을 추출한다. 하나는 auto-regressive method를 이용한 raw-pixel data extraction scheme (RDES)을 사용하는 것이고, 다른 하나는 wavelet transformed data extraction scheme (WTDES)을 이용하여 생성된 sub-block들의 통계적인 특징을 입력으로 사용하는 것이다. 전자는 입력 영상에서 63×63 윈도우 내에 있는 raw-pixel data 중 중요한 화소만을 신경망의 입력으로 사용하는 것이고, 후자는 64×64의 입력 영상에 6차 웨이블릿 변환을 적용하여 각 sub-block들로부터 3가지의 대푯값을 추출하여 사용하는 것이다. 각각의 특징 추출 방법을 통해 역전파 알고리즘(back-propagation algorithm: BP)을 이용하여 신경망을 학습한다. 신경망의 학습을 위해 설문조사를 통해 수집된 220개의 데이터 중 120개를 사용한다.

제안된 방법의 공헌도는 다음의 두 가지이다. 하나는 본 연구는 인위적인 데이터가 아니라, 실제 의류영상을 대상으로 한 연구로 실제 의류영상에서의 패턴과 감성 간의 상관관계를 조사하여 관계 그래프를 생성한 것이다. 이는 실제 의류 영상에 적용을 용이하게 한다는 장점이 있다. 다른 하나는 도출된 패턴과 감성간의 상관관

1) Meller, Susan, "Textile designs : 200 years of European and American patterns for printed fabrics organized by motif, style, color, layout." Harry N. Abrams, 1991.

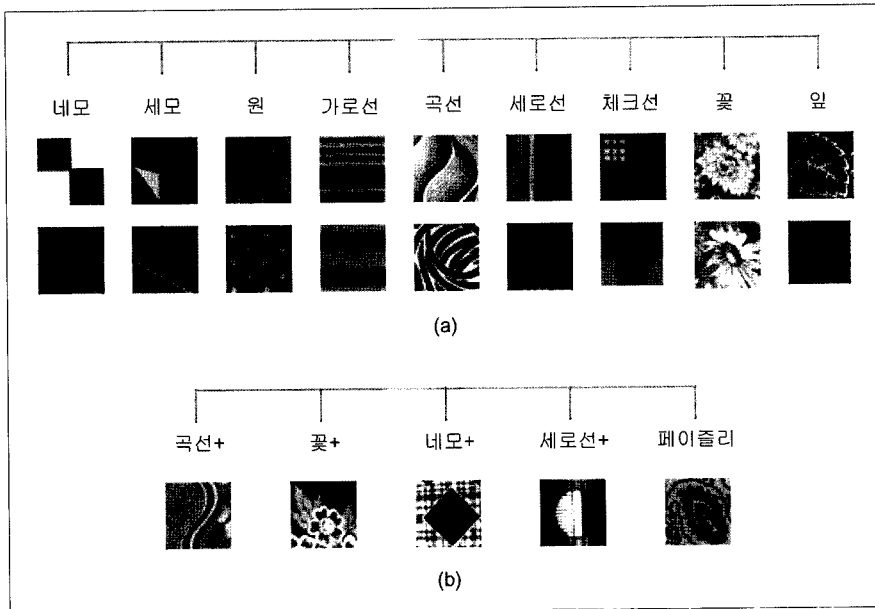


그림 1 패턴에 따른 텍스타일 분류 기준: (a) 단일 패턴 (b) 복합패턴

계를 이용하여 fully-automated system을 개발한 것이다. 실제 의류영상에서의 감성을 인식하기 위한 특징 추출을 위해 본 연구자의 기존 연구인 텍스트 추출 및 얼굴 특징 인식에서 사용한 RDES와 WTDES를 사용하였고, 추출된 특징들을 학습하고 인식하기 위해 대표적인 기계학습방법인 NN을 사용하였다[19]. 결과적으로 본 연구는 실제 의류영상 산업에 적용할 수 있는 자동화된 감성인식 시스템을 개발하였다는 독창성을 갖는다.

이러한 장점을 증명하기 위해서 100개의 실험 영상에 대해 테스트 하였다. 실험 결과는 각각 71%와 90%의 정확도를 보여주었다. 즉, 제안된 시스템은 웨이블릿 변환을 통해 추출된 특징을 사용했을 때 더 높은 성능을 보여주었다. 이러한 성능 차이는 웨이블릿 변환 시 각 화소의 색상 정보뿐만 아니라 수직, 수평, 대각선의 변환 성분 등을 사용하기 때문이다. 더욱이 정확도뿐만 아니라 입력 노드의 개수를 줄임으로써 계산량 역시 감소할 수 있음을 보여주었다.

특징 추출 방법에 따라서 성능의 차이가 있기는 하지만 평균적으로 81%의 인식률을 보여주었으며, 이는 제안된 인식 시스템이 산업 분야 및 감성 기반의 검색 서비스 시스템에 응용 가능성을 보여주었다.

본 논문의 구성은 다음과 같다. 먼저, 감성과 패턴과의 상관관계 조사를 위해서 실시한 데이터의 수집과정과 분석내용이 2장에서 소개되고, 3장에서는 이를 바탕으로 제안된 인식 시스템에 대한 설명이 자세히 나타난다.

다. 그리고 제안된 시스템을 이용한 실험 결과가 4장에서 설명되고, 마지막으로 5장에서는 결론을 말한다.

## 2. 데이터 수집 및 분석

제안된 감성인식 시스템을 구현하기 위해서 먼저 텍스타일 영상에 포함된 패턴과 감성과의 관계에 대한 설문 조사를 실시하였다. 이를 위해 먼저 텍스타일 영상 데이터를 수집하였다.

수집된 텍스타일 영상들은 영상에 포함된 패턴의 복잡도에 따라 크게 단일패턴과 복합패턴으로 분류된다. 단일패턴은 다시 형태에 따라서 '도형', '선', 그리고 '식물'로 분류되고, 각 단일 패턴은 다시 세부 항목으로 분류된다. 그 결과 총 9개의 그룹 ({네모, 세모, 원, 가로선, 세로선, 체크선, 곡선, 꽃, 잎})으로 텍스타일 영상은 분류된다. 복합패턴은 이러한 기본패턴 중 두개 이상의 결합으로 구성된다. 이러한 분류 기준과 각 대표 영상이 그림 1에 나타난다. 이렇게 분류된 데이터들에 대해서 표 1에 나열된 10가지 감성 기준<sup>2)</sup>에 따라 설문조사를 실시하였다. 설문조사는 패턴에 따라 분류된 총 220장의 영상에 대해 일반인 20명을 대상으로 실시되었다. 이 중 120장은 신경망의 학습을 위해, 나머지 100장은 시스템의 성능을 실험하기 위한 테스트 데이터로 사용되었다.

2) Kobayasi의 10가지 감성.

Shigenobu Kobayashi, "COLOR IMAGE SCALE," Kodansha, 1991.

표 1 설문 조사에 사용된 10가지 감성들

ROMANTIC	부드러운, 꿈을 꾸는 듯한
CLEAR	차갑고 맑은
NATURAL	자연스러운, 간소한
CASUAL	자유롭고 편한, 원색적인
ELEGANT	차분하고 우아한
CHIC	지적이고 세련된
DYNAMIC	생동감 있는
CLASSIC	중후하고 전통적인, 명품의
DANDY	튼튼하고, 남성적인
MODERN	쿨하고, 샤프한

설문 조사 시 각 조사 대상자는 각 텍스타일 영상에 대해 10가지 감성을 기준으로 -1, 0, 1의 점수를 매긴다. -1은 주어진 감성과 반대의 감성을 가지고 있다는 것을 나타내고, 0은 주어진 감성의 느낌이 아예 없다는 것을 나타낸다. 그리고 1은 주어진 감성의 느낌을 가지고 있다는 것을 나타낸다. 이렇게 각 영상에 대해서 얻어진 값들은 해당 감성에 따른 설문 조사 대상자들의 결과를 토대로 계속적으로 누적된다. 따라서 각 영상은 각 감성에 대해 최소 -20에서 최대 20까지의 정수 값을 가지게 된다.

이러한 설문 조사 결과는 9개의 단일패턴과 10가지 감성과의 상관관계를 조사하기 위해 그림 2와 같이 히스토그램으로 표현된다. 각 히스토그램은 임의의 감성과 패턴과의 관계를 나타내고, 가로축은 감성 값을, 세로축은 각 감성 값에서의 설문 조사자들의 응답 빈도수를 나타낸다. 히스토그램은 크게 3가지 형태를 갖는데, 그림 2(a)~(f)와 같이, 설문 조사자들의 응답 빈도수가 (-)축에 분포된 그래프와, (+)축에 분포된 그래프, 그리고 0을 중심으로 분포된 그래프로 나뉜다.

감성이란 개인의 환경이나 성향에 따라 빈번하게 변화하는 매우 주관적인 요소이므로 특정한 값을 갖기란 매우 어렵다. 설문조사에서의 분석 과정은 설문 조사자들의 주관적인 응답을 최대한 객관화 시켜야 하는 부분이기 때문에 많은 조사자들의 응답이 요구된다. 앞에서 언급된 'Soen'이나 'Kobayasi'와 같은 학자들의 연구에서는 10명 이하의 조사자들의 응답을 사용한 반면, 본 연구에서는 20명을 대상으로 설문조사를 수행하였다. 그렇기 때문에 시스템 개발을 위해 사용된 정보는 충분히 객관적이다. 또한 본 논문은 이런 애매한 기준을 명확하게 하기 위해 히스토그램 기반에 퍼지 이론을 적용시켰다.

퍼지는 특정 패턴에 따른 히스토그램의 분포를 입력값으로 하고, 소속 함수 그래프에서 입력값에 해당되는 변수가 속한 영역을 입력값의 소속함수(x)로 정한다. 이렇게 정해진 소속함수는 출력 함수 그래프에 매핑시킨 후 해당 감성에 대해 Non-Emotion, No correlation,

Emotion 중 매치되는 영역을 출력함수(y)로 정한다. 소속함수에서 출력함수로의 대응 관계는  $f: X \rightarrow Y$ 로 나타내고, 정해진 출력함수는 다음의 퍼지 룰을 이용해 결과를 낸다.

Rule 1. if  $-20 \leq \text{감성값} < -2$  then 해당 감성이 아니다.

Rule 2. if  $2 < \text{감성값} \leq 20$  then 해당 감성이 있다.

Rule 3. if  $-9 \leq \text{감성값} \leq 9$  then 해당 감성이 없다.

이러한 히스토그램 기반의 퍼지이론 분석을 통해 각각의 텍스타일을 감성 값에 따라 분류하였으며, 이로부터 특정 패턴과 몇 가지 감성 그룹간의 상관관계의 유무를 조사하였다. 그림 3은 패턴과 감성간의 상관관계를 트리구조로 도식화 한 것이다.

그림 3에서 보이듯이 각 도형마다 특정 감성이 매핑되어 있다. 예를 들어, '네모'와 '가로선' 패턴은 'dandy'를 주 감성으로 갖고 있고, '세모'와 '곡선' 그리고 '잎'의 패턴은 'dynamic'을 주 감성으로 갖고 있다. 또 '원' 패턴은 'casual'을 주 감성으로 갖고, '세로선'과 '체크' 패턴은 'chic'을 주 감성으로 갖는다. 그리고 '꽃' 패턴은 'romantic'을 주 감성으로 갖는다.

따라서 이러한 결과로부터 각 영상에 포함된 감성을 인식하기 위해서는 영상에 포함된 패턴의 인식이 필요하다. 본 논문에서는 이를 위하여 신경망을 이용한다.

### 3. 제안된 방법

설문 조사 분석을 통해 텍스타일 영상의 포함된 감성은 그 영상을 대표하는 패턴 정보에 많이 의존한다는 사실을 알 수 있었다. 따라서 패턴 정보를 인식하기 위해서 본 시스템에서는 신경망을 이용한다. 신경망의 입력으로는 크게 두 가지 방법을 사용하는데 하나는 RDES를 사용하는 것이고, 다른 하나는 WTDES를 사용하는 것이다.

#### 3.1 특징 추출

##### 3.1.1 Raw-pixel data extraction scheme (RDES)

Raw data는 말 그대로 가공되지 않은 데이터로써 원 영상 그대로를 사용함을 뜻하며 raw-pixel data는 텍스처를 표현하는 가장 쉽고 보편적인 방법으로 많이 사용되고 있다[11].

하지만 주어진 입력 윈도우  $M \times M$  내에 모든 화소 값을 신경망의 입력으로 사용하는 것은 입력 노드의 개수를 지나치게 크게 만들기 때문에 신경망 학습 및 테스트의 속도를 느리게 만드는 주된 요인이 된다. 따라서 입력 노드 개수를 줄이기 위해 본 논문에서는 autoregressive method를 사용하여  $M^2$ 의 노드 개수를  $4M-3$

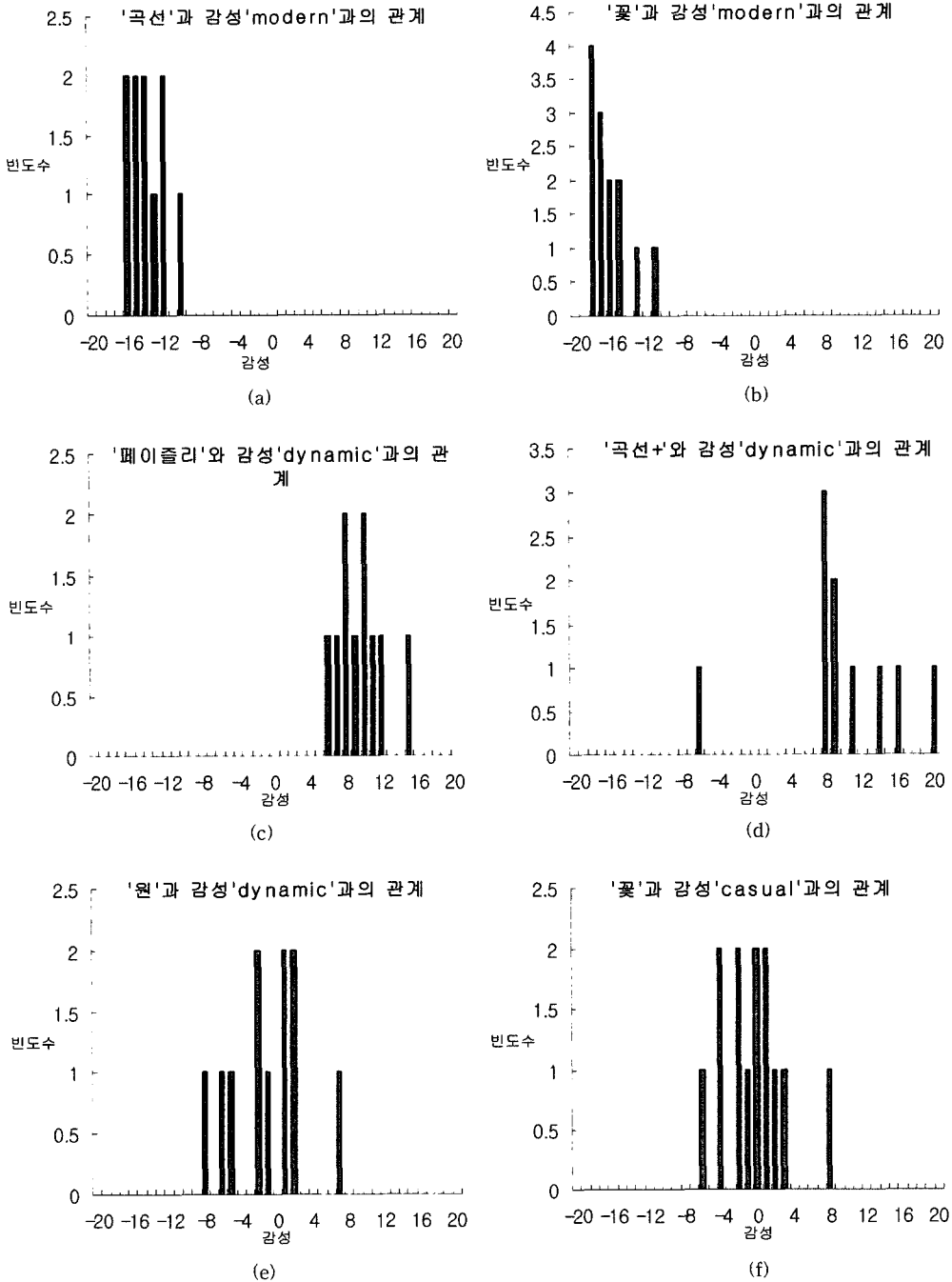


그림 2 패턴과 감성간의 상관관계 히스토그램의 예: (a) 곡선 vs. modern (b) 꽃 vs. modern (c) 페이즐리 vs. dynamic (d) 곡선+ vs. dynamic (e) 원 vs. dynamic (f) 꽃 vs. casual

개로 줄여 사용하기로 한다[12].

그림 4는 autoregressive method를 사용하였을 때, 63×63 윈도우 상에서의 신경망 입력 값을 나타낸다. 그림 4에서 회색으로 채워진 화소만이 신경망의 입력 값

으로 사용된다.

### 3.1.2 Wavelet transformed data extraction scheme (WTDES)

웨이블릿 변환은 입력영상을 각각 수직, 수평적 특징

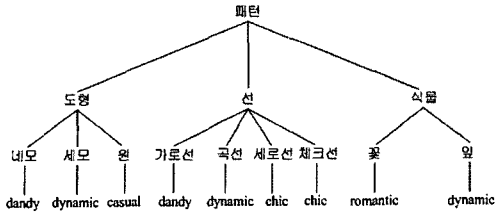


그림 3 패턴과 감성간의 상관관계 그래프

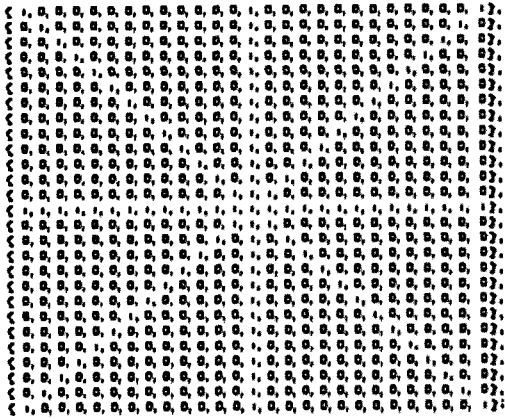


그림 4 Raw-pixel extraction scheme을 이용하였을 때의 신경망의 입력 값(회색에 대응하는 화소만이 입력 값으로 사용)

들을 나타내는 계수들로 구성된 4개의 sub-block LL, LH, HL, HH로 나눈다[13]. LL은 원 영상을 변환 레벨 만큼 압축한 것으로 모든 웨이블릿의 sub-block 중에서 가장 중요한 정보를 포함하고 있으며 나머지 sub-block은 수직, 수평, 대각선 방향에 대한 에지 정보를 나타내고 있다.

그림 5는 3차 웨이블릿 변환 과정을 적용하였을 때 웨이블릿으로 대역 분할된 영상을 나타낸다. 그림에서 보이듯이 차수가 증가함에 따라서 LL에 대해 계속해서 웨이블릿 변환이 수행된다.

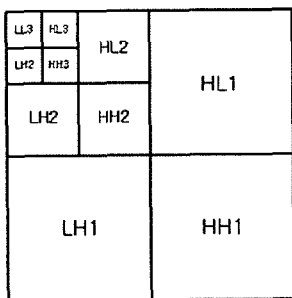


그림 5 2차원 영상의 웨이블릿 분해 결과

본 논문에서는 64×64의 입력 영상에 대해서 6차 웨이블릿 변환을 적용한다. 그 결과 생성된 총 24개의 sub-block에 대해서 각각 다음의 식을 이용하여 3가지 sub-block의 대표 값을 추출한다[14].

$$M(I) = \frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} I(i, j) \tag{1}$$

$$\mu_2(I) = \frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (I(i, j) - M(I))^2 \tag{2}$$

$$\mu_3(I) = \frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (I(i, j) - M(I))^3 \tag{3}$$

영상이 N×N이 주어졌을 때, 식 (1)은 평균값을 나타내고, 식 (2)와 (3)은 각각 2차와 3차 momentum을 나타낸다. 6차 웨이블릿 변환에서는 총 24개의 sub-block이 생성되기 때문에 64×64영상에 대해 72개의 특징 값이 나오며 이들은 신경망의 입력으로 사용된다.

이 특징 추출 방법은 RDES와 비교했을 때, 보다 적은 입력 노드의 수를 사용하기 때문에 신경망 학습 및 테스트에 필요한 시간을 감소할 수 있다.

### 3.2 신경망 기반의 인식

#### 3.2.1 신경망의 구조

본 논문에서는 다층 퍼셉트론(multilayer perceptron, MLP)을 이용하여 감성인식 시스템을 구현한다[15-17]. 제안된 신경망은 입력층, 은닉층, 출력층으로 구성되어 있으며 각 계층에서의 노드들은 완전 연결되어 있다.

본 논문에서는 신경망의 가중치의 학습을 위해 텍스타일 영상, I와 설문조사를 통해 그 영상에 색인된 감성 값, d의 쌍으로 구성된 (I, d)를 패턴으로 사용한다. 학습 알고리즘으로는 BP가 사용된다.

신경망의 인접한 레이어의 노드들을 연결하는 선 각각에는 가중치가 할당되어 있다. 이 가중치와 입력 노드의 값을 이용하여 은닉노드의 값을 계산하고 다시 은닉노드의 값과 가중치를 이용하여 출력노드의 값을 계산한다. 이러한 출력 값과 입력 패턴에 색인된 목적 값과의 차이를 보고 다시 가중치 값을 수정한다. 이러한 과정은 학습에 사용되는 전체 패턴의 오차 값이 최소 오차보다 작아질 때까지 또는 반복횟수가 미리 정해진 횟수에 도달할 때까지 계속된다.

#### 3.2.2 패턴 분류

본 논문에서는 3.2.1의 신경망을 이용하여 특정 감성을 인식할 수 있는 개별 감성 인식 시스템을 개발한다. 그림 6은 10개의 감성에 대응된 10개의 개별 감성 인식 시스템을 보여준다. 각 감성 인식기는 입력된 영상에 해당 감성의 존재 여부에 따라 텍스타일 영상을 분류한다. 각 감성 인식기는 사용된 특징 추출 방법에 따라서 두 가지로 나누어지며 이들은 특징 추출 방법의 우수성을

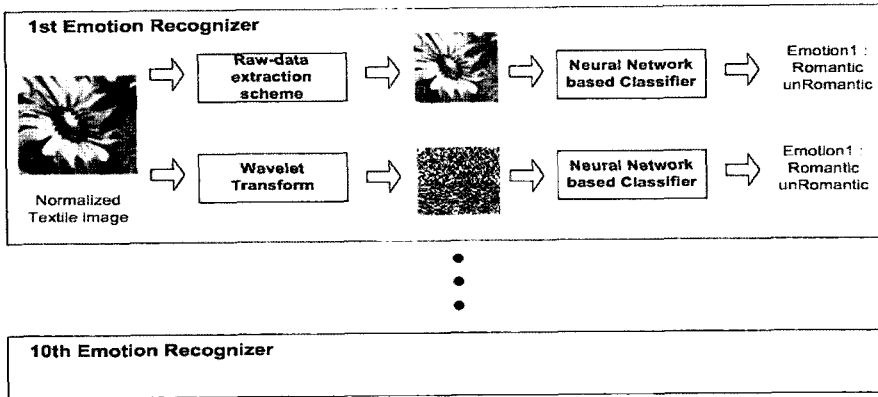


그림 6 신경망 기반의 개별 감성 인식 시스템

비교하기 위해서 사용된다.

주어진 시스템은 먼저 입력된 영상을 64×64의 크기로 정규화 한다. 그 후 두 가지 특징 추출 방법을 사용하여 신경망의 입력될 특징 값을 추출한다.

첫 번째 RDES를 이용하여 신경망의 입력 값을 추출하는 경우에는 63×63의 윈도우 크기가 주어졌을 때, 총 249개의 입력노드와 1개의 출력노드로 구성된 신경망을 이용한다. 이 때, 은닉노드의 수는 실험을 통해 정해진다.

두 번째 WTDES일 경우에는 6차 웨이블릿 변환 시 생성된 24개의 sub-block들로부터 총 72개의 특징 값이 나오므로 72개의 입력노드와 1개의 출력노드로 구성된 신경망을 인식기로 이용하고, 이 경우에도 은닉노드의 수는 실험을 통해 정해진다.

신경망의 출력 값은 0에서 1 사이의 실수로 정규화되며, 0.5보다 크면 해당 감성이 있다고 판단하고 그렇지 않은 경우에는 해당 감성이 없다고 본다.

#### 4. 실험 및 결과

제안된 시스템은 Window XP 환경에서 Visual C++를 사용하여 구현하였다. 구현된 시스템을 평가하기 위해서 설문조사로부터 얻어진 141개의 데이터를 사용하여 시스템의 성능을 평가하였다. 테스트에 사용된 데이터는 크게 두 가지로 분류된다. 하나는 학습에서 사용된 것과 같은 단일패턴을 가진 영상이고, 다른 하나는 학습에서 사용되지 않은 복합패턴을 가진 영상이다. 후자의

경우는 제안된 방법이 얼마나 일반화된 방법인가를 증명하기 위해서 사용되었다. 복합패턴은 단일패턴과는 다르게, 동대문 광장시장에서 실제 의류제작에 쓰이는 천들을 디지털 카메라로 촬영하였고, 총 41개의 텍스타일 데이터를 수집했다. 그림 7은 복합패턴을 갖는 텍스타일 영상의 샘플 데이터를 보여준다. 이렇게 얻은 41개의 복합패턴 데이터에 대해서 전문 디자이너 22명을 대상으로 설문조사를 실시하였고, 단일패턴 때와 동일한 방법으로 설문조사를 분석했다.

신경망의 학습시 중요한 파라미터인 오차율, momentum, 그리고 iteration은 각각 0.001, 0.5, 5000으로 설정된다. 그림 8(a)는 제안된 시스템을 통해서 감성 'chic'과 'unchic'으로 분류된 영상의 예를 보여주고, 그림 8(b)는 감성 'dandy'와 'non-dandy'로 분류된 영상의 예를 보여준다. 그림 8(a)에서 알 수 있듯이 감성 'chic'으로 분류된 영상의 대부분에는 '세로선'과 '체크선'이 포함되어 있으며 감성 'unchic'으로 분류된 영상에는 '직선'이 아닌 형태의 도형이 포함된다. 또한 감성 'dandy'로 분류된 영상 대부분에는 '가로선'과 '네모'가 들어가 있는 반면 감성 'non-dandy'로 분류된 영상에는 대부분 '직선'이 아닌 '곡선' 형태의 도형이 들어가 있다. 즉, 이러한 결과는 설문조사를 통해 얻어진 그림 3의 트리에서 설명된 특정 패턴에 대해 특정 감성이 매핑 된다는 사실을 증명하고 있으며, 제안된 신경망 기반의 인식이 영상에 포함된 패턴을 정확히 인식하고 있음을 보여준다.

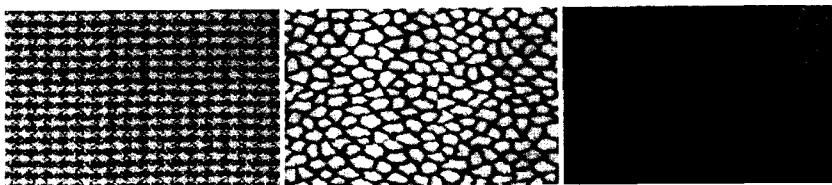


그림 7 복합패턴 영상의 예

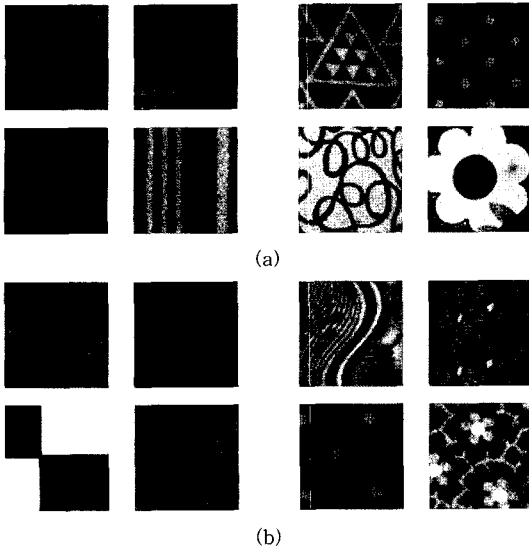


그림 8 감성 분류 별 인식 결과: (a)감성 'chic' vs. 감성 'unchic', (b)감성 'dandy' vs. 감성 'non-dandy'

표 2는 제안된 감성 인식 시스템의 전체적인 정확도를 보여준다. 성능 분석을 위해 precision과 recall, 이 두 가지 평가지표를 사용하는데, 각각은 다음과 같이 정의된다.

$$precision(\%) = \frac{\# \text{ of correctly detected textile image}}{\# \text{ of detected textile image}} \times 100 \quad (4)$$

$$recall(\%) = \frac{\# \text{ of correctly detected textile image}}{\# \text{ of textile image}} \times 100 \quad (5)$$

표 2에서 감성인식 시스템은 크게 두 가지 데이터 추출 방법에 의해서 구분이 되는데 WTDES를 사용한 결과가 RDES를 사용한 결과보다 더 좋은 결과를 보여주는 것을 알 수 있었다. 그리고 RDES를 사용했을 경우에는 윈도우 크기가 커질수록 조금 더 좋은 결과를 보여주었지만 그 차이가 그렇게 크지 않았다.

이러한 실험결과로부터 우리가 알 수 있는 것은 RDES를 사용한 경우 윈도우 크기가 커짐에 따라 정확도의 개선은 보이지만 그에 따르는 계산시간이 늘어나기 때문에 정확도와 속도 사이의 trade-off가 발생한다. 반면, WTDES를 사용하였을 경우에는 더 높은 인식률을 보일 뿐만 아니라 더 적은 입력 노드를 사용함으로써 계산 시간을 줄일 수 있어 속도 및 정확도면에서 더 우수한 성능을 보여준다.

이러한 성능 차이는 WTDES에서 각 sub-block내 입력 화소의 색상 정보만 이용한 것이 아니라 수직, 수평 방향의 에지 성분들을 입력으로 사용하고, 또한 6차 웨이블릿 변환을 사용함으로써 압축된 영상의 전체적인 특징을 사용하였기 때문이다.

WTDES를 사용하여 추출된 데이터를 사용하는 경우 전체적으로 높은 성능을 보이지만 특정 감성에 대해 79%의 낮은 인식률을 보인다. 그림 8은 이러한 오인식(false alarm)의 예를 보여준다. 그림 9(a)는 'romantic' 감성이지만 'unromantic'으로 분류된 예이며 그림 9(b)는 'casual' 감성이지만 'uncasual'로 분류된 예이다. 이러한 감성에 대해서 낮은 인식률을 보이는 이유는 감성 'romantic'과 'casual'의 경우 패턴의 형태뿐만 아니라 크기 및 조밀도에도 영향을 받기 때문이다. 따라서 이러한 오인식은 주파수 성분을 고려한다면 쉽게 해결될 것으로 추측된다.



그림 9 감성 인식기의 오인식의 예: (a) 감성 'romantic'의 영상을 감성 'unromantic'으로 오인식한 경우, (b) 감성 'casual'을 감성 'uncasual'로 오인식한 경우

표 2 단일패턴에서의 감성 인식 시스템의 성능 분석 (%)

	raw-pixel data(27×27)		raw-pixel data(63×63)		wavelet transform	
	recall	precision	recall	precision	recall	precision
ROMANTIC	57	56	79	100	79	100
CLEAR	79	100	79	100	86	100
NATURAL	64	100	71	80	86	100
CASUAL	57	100	64	63	79	100
ELEGANT	50	0	64	50	100	100
CHIC	71	80	64	67	100	100
DYNAMIC	79	100	79	75	86	100
CLASSIC	71	100	71	71	100	100
DANDY	64	100	71	71	100	100
MODERN	79	83	64	63	86	78
AVERAGE	67	82	71	74	90	98



표 3 복합패턴에서의 감성 인식 시스템의 성능 분석(%)

WTDES			
ROMANTIC	38	CHIC	3
CLEAR	47	DYNAMIC	86
NATURAL	50	CLASSIC	100
CASUAL	100	DANDY	97
ELEGANT	4	MODERN	100
AVERAGE	62		

표 3은 복합패턴에서의 감성인식 시스템의 성능을 보여준다. Raw 데이터는 가장 일반적인 방법으로, 웨이블릿과의 성능 비교를 위한 것이기 때문에, 복합패턴에 대해서는 WTDES를 사용했을 때만 테스트 했다. 복합패턴에 대한 실험 결과는 학습되어 있는 단일패턴 영상이 너무 단순하다는 점과, 새롭게 얻은 텍스타일 영상에 포함된 학습되지 않은 다양하고 복잡한 패턴들 때문에 단일패턴에 비해 62%라는 낮은 성능을 보인다.

감성 'romantic', 'clear', 'natural', 'casual'은 단일패턴의 실험에서도 다른 감성들에 비해 높지 않은 성능을 보여줬었고 또, 더 복잡한 패턴을 인식해야 하기 때문에, 단일패턴의 결과와 비교했을 때 위의 4가지 감성에 대한 2차 실험의 낮은 성능은 당연한 결과라고 볼 수 있다. 감성 'elegant'와 'chic'은 실험에 사용된 10가지 감성 중 가장 낮은 성능을 보이는데, 감성 'chic'의 경우 그림 3의 트리 그래프에서 알 수 있듯이, 패턴 '세로선'과 '체크선'을 포함하지만, 복합패턴의 실험을 위해 새롭게 추가된 데이터에는 위의 두 패턴을 포함하는 텍스타일 데이터가 존재하지 않았기 때문에 낮은 성능을 보일 수밖에 없었다. 또, 감성 'elegant'는 패턴과 감성간의 상관관계를 나타내는 트리 그래프에서 본 감성과 관련된 특정 패턴은 존재하지 않기 때문에, 감성 'elegant'를 인식할만한 패턴이 없었으므로 낮은 인식률을 보일 수밖에 없었다. 하지만 나머지 감성들에 대해서는 단일패턴의 실험 때와 같이 완벽에 가까운 성능을 보임으로써 인식이 제대로 동작한다는 것을 알 수 있다.

낮은 성능을 보이는 감성에 대해서는 상관관계가 있는 패턴을 조사하여 추가 학습을 시킨다면 지금보다 훨씬 높은 인식률을 기대할 수 있을 거라 예상하고, 좀 더 다양하고 복잡한 패턴들을 포함하는 텍스타일 영상들에 대한 추가 학습을 통해서 전체적으로 더 높은 정확도를 얻을 수 있을 것으로 보인다.

단일 패턴에 대한 실험 결과에서 보듯이, 특징 추출 방법에 따라 다소 성능의 차이가 있지만 제안된 시스템은 평균적으로 81%의 정확도를 가진다.

제안된 시스템의 성능을 객관적으로 평가하기 위해서는 다른 시스템과 성능 비교가 이루어져야한다. 기존의 색상 정보를 이용하였을 경우 특정 6가지 감성에 대해

서 72%의 성능을 보였다[8]. 반면 단일패턴정보를 이용한 제안된 시스템은 전체 10가지 감성에 대해서 80%이상의 정확도를 보였으며, 특히 WTDES의 특징 추출 방법을 사용하였을 경우에는 90%의 인식률을 보였다. 이는 패턴 정보를 이용하는 것이 감성 인식에 도움을 준다는 것을 보여준다.

다양한 영상에 대한 실험 결과는 제안된 시스템이 인터넷 쇼핑몰의 패션업체나 섬유산업, 그리고 디자인 관련 분야나 시각, 청각 및 지체 장애자들에게 원활한 인터페이스에 대한 어플리케이션에 적용할 수 있는 잠재력을 갖고 있는 것을 보여준다.

### 5. 결론

본 논문에서는 텍스타일 영상에 포함된 패턴과 감성과의 상관관계를 조사하고, 이를 바탕으로 신경망 기반의 감성 인식 시스템을 개발하였다. 상관관계 조사를 위해서 총 220장의 영상에 대해 일반인 20명을 대상으로 설문조사가 수행 되었으며, 이로부터 감성과 패턴간의 히스토그램을 도출해 냈다. 이러한 패턴과 감성의 상관관계로부터 감성의 인식은 패턴 인식으로 인해 수행될 수 있음을 밝혔고, 영상에 포함된 패턴을 인식하기 위해서 신경망 기반의 분류기를 개발하였다. 신경망의 입력으로는 raw-pixel data 또는 웨이블릿 변환된 데이터가 사용되었으며, 데이터 사용 방법에 따라 각각 71%와 90%의 성능을 보였다. 실험을 통해 raw-pixel data를 사용한 경우보다 웨이블릿 변환된 데이터를 사용한 경우가 속도와 인식률에서 더 우수함을 보였는데, 후자가 더 좋은 성능을 보인 것은 입력 화소의 색상 정보뿐만 아니라 수직, 수평 방향의 에지 성분들과 전체적인 특징들이 입력으로 사용되었기 때문으로 추측된다.

개발된 시스템은 전체적으로 좋은 성능을 보였지만, 몇 가지 특정 감성을 인식하는데 어려움을 보여주었다. 이러한 문제점은 이산 코사인 변환(Discrete Cosine Transform: DCT)을 이용하여 주파수 성분을 패턴 정보로 표현하는데 사용한다면, 쉽게 해결 될 것으로 보인다[18]. 따라서 현재 새로운 특징을 검출해내는 방법을 개발 중이며, 또한 단일 특징 추출법만이 아니라 여러 가지 추출 방법을 통합하여 사용하도록 시스템을 확장하고 있다.

### 참고 문헌

[1] Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., and Taylor, J. G., "Emotion recognition in human-computer interaction," IEEE Signal Processing magazine, Vol. 18, No. 1, pp. 32-80, Jan. 2001.

- [2] Bjorn Schuller, Gerhard Rigoll, and Manfred Lang, "Emotion recognition in the manual interaction with graphical user interfaces," IEEE International Conference on Vol. 2, pp. 1215-1218, June 2004.
- [3] G. Donato, M. S. Bartlett, J. C. Hager, P. Ekman, and T. J. Sejnowski, "Classifying facial actions," IEEE Trans. Pattern Anal. Machine Intell., Vol. 21, pp. 974-989, Oct. 1999.
- [4] L. Ma and K. Khorasani, "Facial Expression Recognition Using Constructive Feedforward Neural Networks," IEEE Transactions on System, Man, and Cybernetics - part B: Cybernetics, Vol. 34, No. 3, June 2004.
- [5] Hatice Gunes and Massimo Piccardi, "Fusing face and body gesture for machine recognition of emotions," IEEE International Workshop on Robots and Human Interactive Communication, 2005.
- [6] Aishah Abd Razak, Mohd Hafizuddin Mohd Yusof, Ryoichi Komiya, "Towards automatic recognition of emotion in speech," IEEE International Symposium on Signal Processing and Information Technology, 2003.
- [7] D. Ververidis, and C. Kotropoulos, "Automatic speech classification to five emotional states based on gender information," Proceedings of the EUSIP-CO2004 Conference, Austria, pp. 341-344, Sept. 2004.
- [8] Eun Yi Kim, Soo-jeong Kim, Hyun-jin Koo, Karpjoo Jeong, Jee-in Kim, "Emotion-based Textile Indexing using Colors and Texture," Lecture Notes in Computer Sciences, Vol. 3613, pp. 1077-1080, 2005.
- [9] 류형주, 채송아, 김수정, 김은이, 김지인, 정갑주, 구현진, "파지시스템을 이용한 텍스타일 인텍싱", 한국정보과학회 가을 학술발표논문집, Vol. 31, No. 2, 2004.
- [10] T. Soen, T. Shimada, and M. Akita, "Objective Evaluation of Color Design II", Color Res. Appl., Vol. 12, pp. 187-194, 1987.
- [11] Fu-Chiang Tsui, Mingui Sun, Ching-Chung Li, Robert J. Scabassi, "RECURRENT NEURAL NETWORKS AND DISCRETE WAVELET TRANSFORM FOR TIME SERIES MODELING AND PREDICTION", ICASSP International Conference on Vol. 5, pp. 3359-3362, May 1995.
- [12] Em Yi Kim, Sin Kuk Kang, Keechul Jung and Hang Joon Kim, "Eye Mouse- Mouse Implementation using Eye Tracking," IEEE International Conference on Consumer Electronics, pp. 207-208, 2005.
- [13] 종현, 박순영, 조완현, 오일석, "색상특징과 웨이블릿 기반의 질감특징을 이용한 영상검색", 정보과학회 논문지: 데이터베이스 제 30권 제 2호(2003.4)
- [14] Huiping Li, David Doermann, and Omid Kia, "Automatic Text Detection and Tracking in Digital Video", IEEE Transactions on Image Processing, Vol. 9, No. 1, January 2000.
- [15] Simon Haykin, Neural Networks a comprehensive foundation 2nd edition, Prentice Hall, pp. 10-13,

pp. 156-173, 1999.

- [16] H.-U. Bauer and T. Geisel, "Dynamics of signal processing in feedback multilayer perceptrons," in Proc. Int. Joint Conf. Neural Networks, pp. 131-136, 1990.
- [17] M. Brown, P. C. An, C. J. Harris, and H. Wang, "How biased is your multi-layer perceptron?," in World Congr. Neural Networks, pp. 507-511, 1993.
- [18] K.U. Barthel, J.Schiitemeyer, T.Voye, and P.Noll, "A new image coding technique unifying fractal and transform coding," IEEE ICIP Texas, pp. 112-116, 1994.
- [19] Eun Yi Kim; Sin Kuk Kang; Keechul Jung; Hang Joon Kim; Consumer Electronics, 2005. ICCE. 2005 Digest of Technical Papers. International Conference on 8-12 Jan. 2005 Page(s):207-208



김 나 연

2006년 건국대학교 인터넷미디어공학과 학사. 2006년~현재 건국대학교 신기술융합과 석사 재학중. 관심분야는 영상처리, 패턴인식, 컴퓨터비전



신 윤 회

2006년 건국대학교 인터넷미디어공학과 학사. 2006년~현재 건국대학교 신기술융합과 석사 재학중. 관심분야는 영상처리, 패턴인식, 컴퓨터비전



김 수 정

1998년 건국대학교 전자계산학과 학사  
2000년 건국대학교 컴퓨터정보통신공학과 석사. 2007년 건국대학교 컴퓨터정보통신공학 박사. 2007년~현재 건국대학교 서울시 산하 디지털 콘텐츠 연구센터 연구원. 관심분야는 인간과 컴퓨터 상호작용, 감성공학, 영상처리, 패턴인식



김 지 인

1982년 서울대학교 컴퓨터공학과 학사  
1984년 KAIST 전산학과 석사. 1993년 미국 University of Pennsylvania 전산학 박사. 1995년 건국대학교 컴퓨터공학과 교수. 2001년 건국대학교 인터넷미디어공학부 교수. 2006년~현재 건국대학교 신기술융합학과 교수. 관심분야는 Human Computer Interaction, Virtual Reality, Digital Contents 등



정 갑 주

1984년 서울대학교 컴퓨터공학 학사  
1986년 서울대학교 인공지능공학 석사  
1996년 New York University 분산컴퓨팅공학 박사. 1997년 건국대학교 컴퓨터공학과 조교수. 2001년 건국대학교 인터넷/미디어공학부 부교수. 2006년~현재

건국대학교 신기술융합학과 교수. 관심분야는 그리드컴퓨팅, eScience, eEngineering, distributed computing, fault tolerance 등



구 현 진

1989년 인하대학교 섬유공학 학사. 1993년 노스캐롤라이나 주립대 Textile Management&Technology 석사. 1996년 노스캐롤라이나 주립대 Fiber&Polymer Science 박사. 2002년~현재 FITI시험연구원 신뢰성평가그룹 책임연구원. 관심분야

는 패턴인식, 영상처리, 섬유제품 관능평가의 객관화



김 은 이

1997년 경북대학교 컴퓨터공학과 학사  
1999년 경북대학교 컴퓨터공학과 석사  
2001년 경북대학교 컴퓨터공학과 박사  
2002년 건국대학교 인터넷미디어공학부 교수. 2006년~현재 건국대학교 신기술융합학과 교수. 관심분야는 컴퓨터비전, 패

턴인식, 사용자인터페이스, 디지털콘텐츠, 생체인증시스템