

신경망을 이용한 내용기반 영상 분류

이재원[†] · 김상균^{††}

요 약

본 논문에서는 내용기반 영상 분류를 위한 방법론으로써 신경망을 이용한 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 특징은 웨이블릿 변환 후 추출된 형태 특징과 질감 특징을 이용한다. 추출된 특징 값들을 이용하여 영상들에 대한 학습패턴을 생성하고 신경망 분류기를 구성한다. 신경망의 학습 알고리즘은 역전파 알고리즘을 사용한다. 가장 효과적인 질감특징을 선택하기 위한 실험에서는 대각 모멘트가 가장 높은 분류율을 보여주었다. 배경을 제거하고 대각 모멘트를 특징으로 사용하여 실험하였을 때, 30종류에서 각 10개씩 총 300개의 학습 데이터와 300개의 테스트 데이터에 대하여 각각 72.3%와 67%의 정분류율을 보였다.

A Content-Based Image Classification using Neural Network

Jae-Won Lee[†] and Sang-Kyoon Kim^{††}

ABSTRACT

In this paper, we propose a method of content-based image classification using neural network. The images for classification are object images that can be divided into foreground and background. To deal with the object images efficiently, object region is extracted with a region segmentation technique in the preprocessing step. Features for the classification are texture and shape features extracted from wavelet transformed image. The neural network classifier is constructed with the extracted features and the back-propagation learning algorithm. Among the various texture features, the diagonal moment was more effective. A test with 300 training data and 300 test data composed of 10 images from each of 30 classes shows correct classification rates of 72.3% and 67%, respectively.

Key words: classification, content-based, wavelet transform, texture feature, neural network classifier

1. 서 론

컴퓨터 기술이 급속도로 발전하고 인터넷이 보편화되면서 영상, 오디오, 비디오, 그래픽, 문자 등과 같은 멀티미디어 데이터의 활용이 증가하고 있다. 그 중에서도 영상 데이터를 처리하고 검색하는 방법론에 대한 연구가 활발히 진행되고 있다.

기존의 영상 검색은 각 영상에 대한 개별적인 주석을 이용한 문자기반 영상 검색(text-based image retrieval)에 의존해왔다. 이 방법은 영상 파일의 파일명, 표제, 키워드, 그리고 문서내의 내용을 이용하여 영상을 색인한다. 이는 바이트마다 숫자나 문자라는 의미가 부여되는 문자와는 달리, 단순한 비트열인 영상 데이터를 문자처럼 정보화하기 위해서, 사람이 직접 영상 데이터를 해석해서 표현하는 것과 유사한 방식이다.

그러나 문자기반 영상 검색 방법은 많은 양의 데이터에는 적합하지 않으며, 키워드의 자동적인 생성

접수일 : 2002년 4월 8일, 완료일 : 2002년 6월 27일
본 연구는 2000년도 인제대학교 학술연구조성비 단독보조로 수행되었음.

[†] (주)투원정보시스템

^{††} 인제대학교 정보컴퓨터공학부 조교수

이나 다양한 종류의 영상들을 구분하기 위한 특징 추출이 어렵다. 즉 사람에 의해서 키워드가 생성되므로 매우 주관적이고, 수작업에 따른 막대한 정보화 비용이 요구된다.

이러한 문제점들을 해결하기 위하여 1990년대 초에 내용기반 영상 검색(content-based image retrieval)이 제안되었다. 이 방법은 영상 데이터로부터 표현되는 색상, 모양, 질감, 스케치 등의 특징들을 자동으로 추출하여 이 정보를 분류 및 검색에 사용한다[1-4]. 이러한 검색 시스템들은 유클리디안 거리를 유사성 척도로 사용하여 영상 데이터베이스에서 질의 영상과 매칭되는 영상을 검색한다.

그러나 이러한 방법들은 각 영상들을 특정 카테고리에 따른 분류 없이 개별적으로 저장 및 관리한다. 이는 대용량 영상 데이터의 저장 및 관리, 특징 추출 및 검색에 있어 효율성이 떨어진다. 이러한 문제점을 해결하기 위하여 사전에 전체 영상을 그룹핑 및 분류하여 검색 시스템의 성능을 개선시키는 방법론에 대한 연구가 필요하다.

본 연구에서는 효율적인 영상의 분류를 위해서 신경망을 이용한 방법을 제안한다. 분류 대상 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. 그리고 웨이블릿 변환을 이용하여 영상에서 형태, 질감의 정보를 추출하며, 특징 추출시 객체의 각 지역영역에서 구조적 특징을 추출함으로써 보다 많은 내용 정보를 획득 및 반영한다. 객체 영상들을 분류하기 위한 분류기로서는 패턴 인식 분야에서 인식기, 분류기로서 안정적인 성능을 보이는 신경망을 사용하였다.

2. 영상 분류기의 구성

본 연구에서 제안하는 영상 분류기의 전반적인 구조는 그림 1과 같다. 본 영상 분류기는 크게 전처리 모듈, 특징 추출 모듈, 분류 모듈의 3가지로 구성된다.

전처리 모듈에서는 배경 제거와 정규화가 이루어진다. 먼저 지역 분할(region segmentation)을 이용하여 영상에서 배경을 제거한다. 다음으로 인터넷에서 획득한 JPEG 포맷의 영상 파일을 BMP 포맷으로

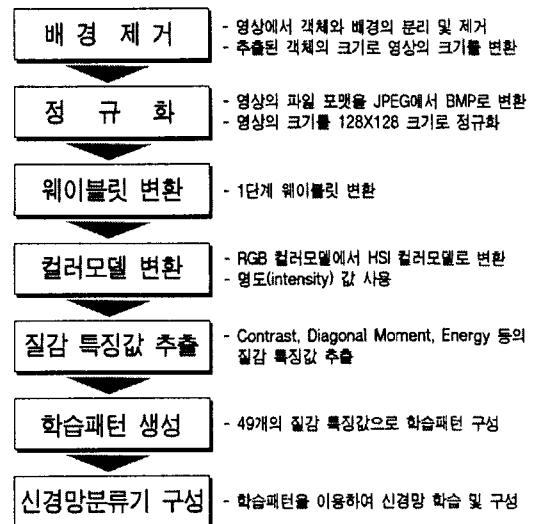


그림 1. 제안된 영상 분류기의 전체 구조

변환한다. 그리고 영상 정보를 신경망의 학습패턴으로 사용하기 위해 영상의 크기를 128×128로 정규화한다.

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

분류 모듈에서는 학습패턴을 생성하고, 신경망 분류기를 구성한다. 학습패턴 생성 과정에서는 구조적으로 추출된 49개의 특징값들을 이용하여 학습패턴을 구성하고, 최종적으로 이 학습패턴을 사용하여 신경망 분류기를 학습 및 구성한다.

2.1 배경 제거

인터넷에 존재하는 다양한 영상들 중 객체 영상들은 대부분 복잡한 배경과 함께 전체 영상을 구성하고 있다. 본 연구에서는 분류기의 인식률을 높이기 위하여 배경에 의한 잡음을 최소화하는 배경 제거를 전처리 단계로서 수행한다. 본 연구에서 제안하는 배경 제거의 단계는 그림 2와 같다.

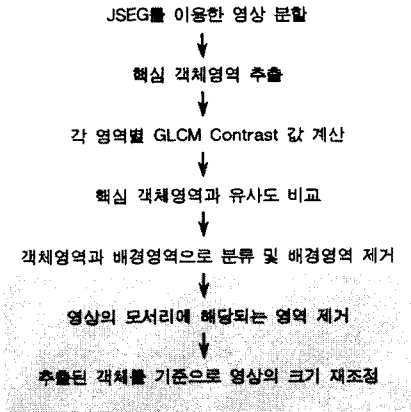


그림 2. 배경 제거의 단계별 흐름

2.1.1 JSEG를 이용한 영상 분할

배경을 제거하기 위해서는 우선 영상 분할(image segmentation) 과정이 필요하다. 일반적인 분할에서는 픽셀 단위의 정교한 영역 경계 추출이 요구되지만, 영상 검색 및 그룹핑을 위한 영상 분할에서는 영역 경계는 다소 부정확하더라도 특징적인 영역들을 고속으로 추출하는 것을 목적으로 한다. 이때, 질감에 의해 영역이 너무 세분화되는 것은 바람직하지 못하다.

질감에 무관하게 영역 분할을 하는 방법으로서 [5]에서 제안된 것을 이용할 수 있다. 내용기반 영상 검색 시스템인 넷트라(Netra)에서 사용되고 있는 방법인 JSEG[5]에서는 색상과 질감 분포 유사도를 반영하는 J 값을 정의하여 영역을 분할하는데 사용한다. 이 방법은 처리 속도는 다소 떨어지나 좋은 영역 분할 성능을 보여준다.

JSEG 영상 분할 방법은 먼저 입력 영상을 양자화(quantization)한다[6]. 이후 양자화된 컬러값들을 레이블링하여 class-map을 생성한다. 이후 영상내 컬러의 공간적 특징을 반영하는 J 값의 정의에 따라 J -image를 생성하고 J -image에 대해서 영역 확장(growing)과 병합(merge)을 거쳐 분할된 영상을 생성한다. 식 (1)은 J 값의 정의에 따른 수식이다. 여기서 J 값은 class-map들의 응집된 정도를 수치화한 값이며, S_W 는 class내의 응집도, S_T 는 class간의 분할된 정도, m 은 class-map내에서 모든 class들의 평균 위치를 의미하며, Z 는 class-map내에서 모든 N 개의 픽셀들의 집합을 의미한다[5].

$$\begin{aligned}
 m_i &= \frac{1}{N_i} \sum_{z \in Z_i} z, & m &= \frac{1}{N} \sum_{z \in Z} z \\
 S_W &= \sum_{i=1}^C S_i = \sum_{i=1}^C \sum_{z \in Z_i} \|z - m_i\|^2 \\
 S_T &= \sum_{z \in Z} \|z - m\|^2 \\
 J &= S_B / S_W = (S_T - S_W) / S_W
 \end{aligned}
 \tag{1}$$

그림 3은 자동차 영상을 JSEG를 이용해서 영역 분할을 한 결과이다

2.1.2 핵심 객체영역 추출

이 영역 분할 결과를 이용하여 객체 영역과 배경 영역을 분류한다. 먼저 핵심 객체영역을 추출한다. 본 연구에서 분류하고자 하는 대상 영상은 객체 영상이다. 이 객체 영상들은 객체가 영상의 중심에 위치하는 경우가 많다는 점에서 원본 영상의 중심에서 핵심 객체영역을 추출한다. 그림 4에서와 같이 원본 영상의 1/2 크기 윈도우를 관심 영역으로 설정한다. 그리고 이 윈도우 내에 포함되는 각 영역들의 면적을 구하여 이 중 가장 면적이 넓은 영역을 핵심 객체영역으로 선정한다. 그림 4에서 검은색으로 표시된 영역이 추출된 핵심 객체영역이다.

2.1.3 객체영역과 배경영역 분류

다음으로 각 영역의 질감 특징값을 추출한다. 본 논문에서는 다양한 질감 특징들 중에서 식 (2)와 같이 GLCM(Glay Level Co-occurrence Matrix) 대비(contrast)를 사용하였다. 식 (2)에서 $GR[i,j]$, $GG[i,j]$, $GB[i,j]$ 는 RGB 각각의 GLCM을 의미한다.

$$\begin{aligned}
 Contrast(Red) &= \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (i-j)^2 GR[i,j] \\
 Contrast(Green) &= \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (i-j)^2 GG[i,j] \\
 Contrast(Blue) &= \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (i-j)^2 GB[i,j]
 \end{aligned}
 \tag{2}$$

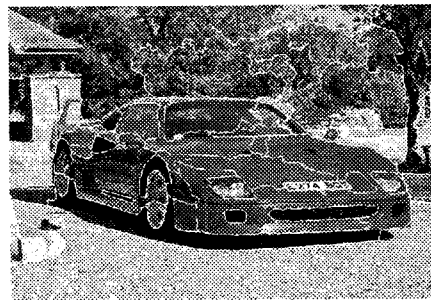


그림 3. JSEG 영역 분할의 예

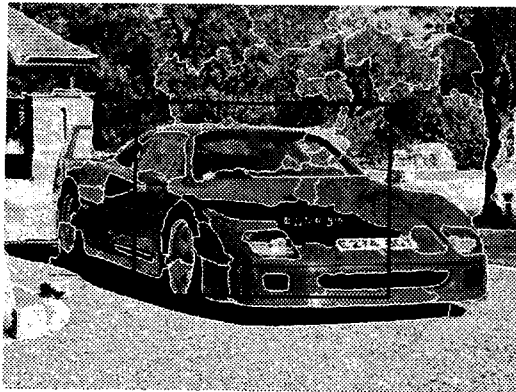


그림 4. 핵심 객체영역의 추출

통계적인 질감 분석 방법인 GLCM은 그레이 레벨 영상에서 정의한 변위 벡터의 거리와 방향이 일치하는 픽셀쌍의 빈도수를 표시하는 빈도수 매트릭스이다. GLCM은 거리가 1인 픽셀에서는 그림 5와 같이 8방향 인접(8-neighbor)으로 표현 될 수 있다.

그림 5에서 1과 5는 0도를, 2와 6은 45도를, 3과 7은 90도를, 그리고 4와 8은 135도를 각각 나타낸다. 네 방향에 따른 매트릭스는 4가지 종류가 구해지게 되고 이를 이용한 데이터를 생성하게 된다[7,8].

이렇게 추출된 각 영역의 질감 특징값들을 이전에 선정된 핵심 객체영역의 질감 특징값과 유사도를 비교하여 객체영역과 배경영역으로 분류한다. 유사성 척도로서는 식 (3)의 유클리디안 거리를 사용하였다. CR, CG, CB는 식 (2)에서 계산된 RGB 각각의 GLCM Contrast를 의미한다.

$$D = \sqrt{(CR - CR')^2 + (CG - CG')^2 + (CB - CB')^2} \quad (3)$$

그림 6은 핵심 객체 영역과의 유사도 비교 결과 유사도 문턱치(threshold value)가 1.948979 이상인 영역을 배경 영역으로 분류하고, 제거한 결과이다. 이 문턱치는 실험적 경험에 의해 유클리디안 거리값의 분포의 급변하는 구간에서 선택하였다.

2.1.4 모서리영역 제거

최종적으로 보다 정확한 배경의 제거를 위하여 각 모서리에 해당되는 영역을 배경영역으로 분류 및 제거한다. 그림 7은 그림 6에서 모서리 영역을 제거한 결과이다.

135도	90도	45도	
4	3	2	
5	1	1	0도
6	7	8	

그림 5. GLCM의 방향성

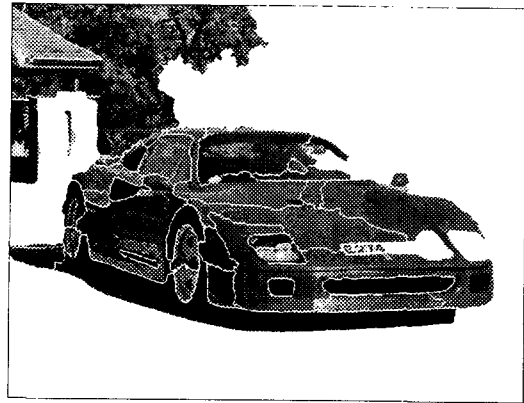


그림 6. 핵심 객체 영역과의 유사도 비교 결과

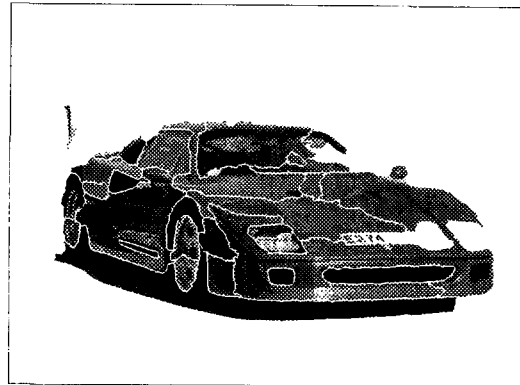


그림 7. 모서리 영역 제거 결과

2.1.5 영상의 크기 재조정

영상 내에서 객체의 크기와 위치는 다양하다. 이러한 다양한 크기와 위치에 무관한 특징값을 추출하기 위해선 GLCM을 사용하여 영상 전체에 대한 하나의 특징값을 추출하는 방법이 있다. 하지만 본 연구에서는 영상에서 보다 많은 정보를 추출하기 위하여 구조적 특징을 추출하였다.

하지만 구조적 특징값들을 추출할 경우 객체의 크기와 위치에 따라 서로 다른 특징값이 추출된다는

문제점이 있다. 이러한 문제점을 해결하기 위해서 본 연구에서는 위의 과정으로 추출된 객체를 기준으로 영상의 크기를 재조정(resizing) 한다. 그림 8은 그림 7의 크기를 객체를 기준으로 재조정된 결과이다.

2.2 정규화

인터넷에서 획득한 영상 파일들은 대부분 JPEG 포맷이다. 본 연구에서는 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. 이러한 획일적 정규화는 영상이 정방형이 아닐 경우, 특정 방향성 질감 정보의 손실을 초래하는 단점이 있다.

2.3 웨이블릿 변환

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

웨이블릿 변환 방법은 웨이블릿이라 일컫는 기저함수(basis function)를 이동(translation), 확장(dilating)함으로써 영상 데이터를 주파수 영역으로 변환한다[9]. 변환 후 생성되는 부밴드들을 분석함으로써 영상에 대한 정보를 추출할 수 있다.

웨이블릿 변환을 이차원 영상에 적용하였을 때 네 개의 성분으로 나누어진다. 이 네 개의 성분은 시각에 민감한 정보를 가지고 있으며, 크게 저주파 성분과 고주파 성분으로 나누어지며, 고주파 성분에는 또 다시 수직 성분, 수평 성분, 대각 성분으로 나누어진다. 그림 9는 2-채널 필터로 2단계 웨이블릿 변환한 후 생성되는 분할 영역이다.

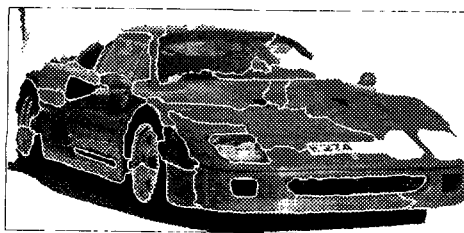


그림 8. 크기 재조정 결과

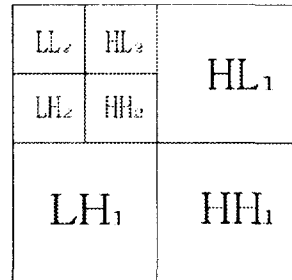


그림 9. 웨이블릿 분할 영역

대부분의 관련 연구에서는 2단계 웨이블릿 변환의 결과에서 HH2 영역만을 사용한다. 그리고, HL1, HH1, LH1, HL2, HH2, LH2, 6개의 모든 고주파 영역을 사용한 예도 있다. 하지만 본 연구에서는 배경 제거 단계에서 객체를 기준으로 크기를 재조정하고, 정규화 단계에서 영상의 크기를 128×128로 재조정함으로써, 이미 두 차례 영상 크기를 축소하였기 때문에 웨이블릿 변환된 결과에서는 가장 정보의 양이 많은 HH1 영역을 특징 추출 대상 영역으로 사용하였다.

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[10,11].

따라서 본 연구에서는 많은 영상 관련 어플리케이션에서 사용되고 있는 HSI 컬러 모델로 변환하여 그 중 명도(intensity) 값을 사용하였다. 식 (4)는 RGB 컬러 모델에서 HSI 컬러 모델로의 변환 식이다[10].

$$\begin{aligned}
 H &= \cos^{-1} \left(\frac{\frac{1}{2} [(R-G) + (R-B)]}{[(R-G)^2 + (R-B)(G-B)]^{1/2}} \right) \\
 S &= 1 - \frac{3}{(R+G+B)} [\min(R, G, B)] \\
 I &= \frac{(R+G+B)}{3}
 \end{aligned}
 \tag{4}$$

변환된 HSI 컬러 모델에서 명도를 사용함으로써 얻을 수 있는 유용성은 영상에서 색상 정보로부터 분리되어 있다는 점이다[7].

예를 들어 사과는 빨간색, 노란색, 초록색 등의 서로 다른 색상을 가지고 있다. 이런 색상이 다른 사과 영상들을 각각 세분화하여 분류한다면 인터넷 상의 수많은 영상들의 분류 종류는 엄청난 수가 될 것이

다. 하지만 영상의 의미 정보 중 색상 대신 명도를 사용함으로써 이 세 가지 색상의 사과를 “사과”라는 하나의 종류로 분류할 수 있다.

그림 10은 그러한 예를 보여주는 결과로서, 세 가지 색상의 사과들의 특징값 분포를 보여준다.

2.5 질감 특징 추출 및 학습패턴 생성

일반적으로 영역 서술의 중요한 방법 중 하나는 영역의 질감(texture)의 내용을 수량화하는 것이다 [7]. 질감은 통계적인 특징과 구조적인 특징, 스펙트럼 특징을 이용하여 표현된다. 통계적인 특징은 영상의 거칠고 부드러운 정도를 나타낸다. 구조적인 특징은 영상 내의 수평선의 반복과 같은 규칙적인 배열을 표현한다. 스펙트럼 특징은 높은 에너지 성분을 분석함으로써 영상 내의 전체적인 주기성을 알아내는데 사용된다[8,13].

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. 식 (5)는 본 연구에서 실험한 7가지, contrast, diagonal moment, energy 등의 질감 특징값들을 구하는 식이다.

$$\begin{aligned}
 Contrast &= \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (i-j)^2 P(i, j) \\
 Diagonal\ Moment &= \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \sqrt{\frac{|i-j| P(i, j)}{2}} \\
 Energy &= \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P(i, j)^2 \\
 Entropy &= - \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P(i, j) \log(P(i, j)) \\
 Homogeneity &= \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \frac{P(i, j)}{1 + (i-j)^2} \\
 Second\ Diagonal\ Moment &= \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \frac{|i-j| P(i, j)}{2} \\
 Uniformity &= \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \frac{P(i, j)}{1 + |i-j|} \quad (5)
 \end{aligned}$$

본 연구에서는 이러한 다양한 질감 특징들 중 일부를 사용하여 실험을 하였다. 실험 결과 대각 모멘트(diagonal moment)가 가장 높은 인식률을 보여주었으며, 따라서 본 연구에서는 질감 특징으로써 대각 질감 특징값들은 그림 11과 같이 추출된다. 웨이

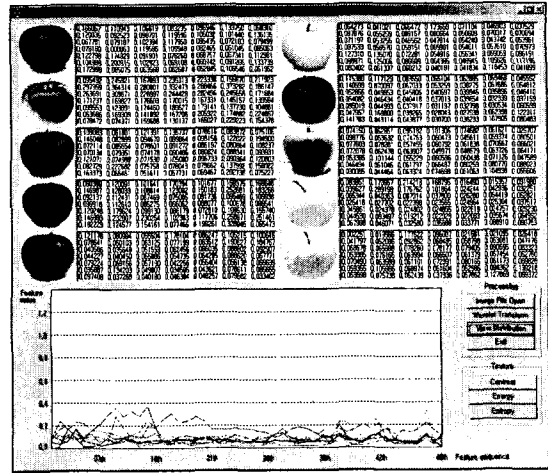


그림 10. 세 가지 색상의 사과의 질감 특징값 분포

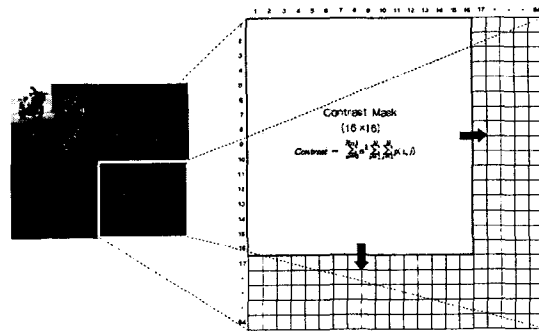


그림 11. 질감 특징값을 추출하는 예

모멘트를 사용하였다.

블러 변환후 생성되는 64×64 크기의 HH1 영역을 대상으로 특징값을 추출한다. 16×16 크기의 윈도우에서 하나의 질감 특징값을 획득한다. 그리고 이 윈도우를 좌에서 우로, 위에서 아래로 8 픽셀씩 중첩(overlap) 이동하며 각각의 영역에서 특징값을 추출한다. 최종적으로 하나의 영상에서 총 49개의 질감 특징값을 얻을 수 있다. 윈도우의 크기와 중첩 이동되는 픽셀수는 실험적 경험에 의해 가장 효율적인 수치를 적용하였다.

이와 같이 추출된 49개의 질감 특징값은 신경망을 학습시키기 위한 입력값이 되며, 30개의 목표값과 함께 학습패턴으로 구성된다.

2.6 신경망 분류기 구성

신경망은 인간의 두뇌를 수학적으로 모델링하여 만든 인공적인 두뇌를 공학이나 기타 여러 분야에

적용시킨 것이다[14]. 신경망 학습 알고리즘 중 역전파(back-propagation) 알고리즘은 LMS 법칙을 일반화시킨 것으로 신경망의 실제 출력값과 목표값과의 평균 차이를 최소화하는 경사 탐색기법을 사용한다. 신경망의 구성은 그림 12와 같다[15].

신경망 분류기는 49개의 입력노드를 가진 입력층(input layer)과 49개의 은닉노드로 구성된 은닉층(hidden layer), 그리고 분류할 영상 종류의 개체수인 30개의 출력노드로 구성된 출력층(output layer)으로 설계하였다. 신경망의 학습은 에러값(minimum error)이 0.1이 될 때까지 연결강도(weight)를 조정하여 학습하였다.

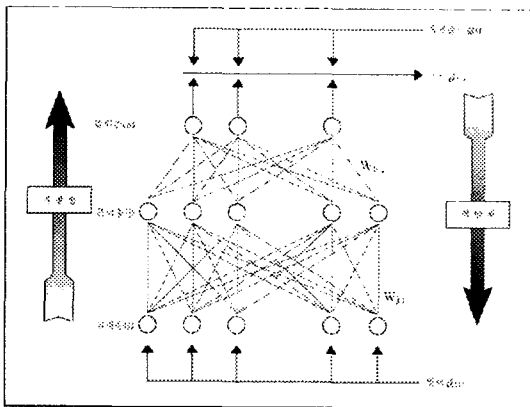


그림 12. 신경망의 구성도

3. 실험 및 결과 분석

3.1 실험 환경

이 실험은 다양한 질감 특징들을 사용하였을 때 각각의 인식률을 비교하고, 본 연구에서 제안한 배경을 제거했을 때와 배경을 제거하지 않았을 때의 인식률 차이를 비교한다. 그리고 각 영상 종류별 특성을 분석하여 인식 결과가 높은 영상 종류와 인식 결과가 낮은 영상 종류에 대하여 그 원인을 분석한다.

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

그림 13은 신경망 분류기의 학습 및 실험에 사용한 영상 종류들이다. 인터넷 상의 다양한 영상들 중에 객체와 배경의 구분이 있고, 객체가 영상의 중심에 위치한 객체 영상들을 실험 데이터로 사용하였다.

실험에 사용한 영상의 종류는 총 30종이다.

그림 14는 질감 특징값을 추출하기 위하여 구현한 프로그램의 예이다. 총 10개의 영상의 결과를 보여주며, 각 영상별 좌측에는 원본 영상의 웨이블릿 변환된 결과를 보여주며, 우측에는 추출된 질감 특징값들이 출력된다. 하단에 있는 그래프는 각 영상들의 특징값들의 분포를 보여준다.

그림 15는 위와 같이 추출된 특징값을 학습패턴으로 사용하여 신경망을 학습시키는 프로그램을 실행한 예이다.



그림 13. 학습 및 실험에 사용한 영상의 종류

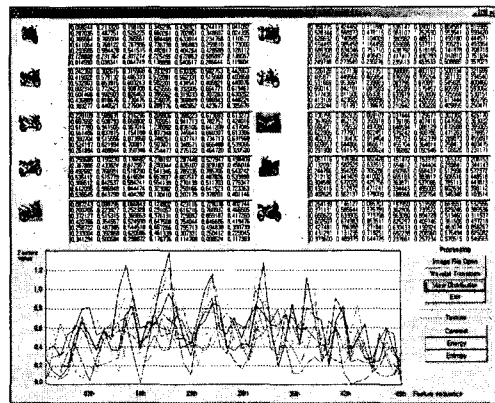


그림 14. 특징값을 추출하는 프로그램의 예

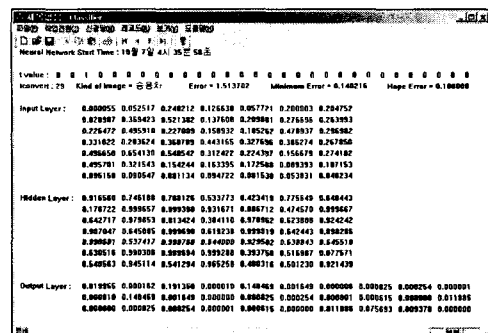


그림 15. 신경망 학습 프로그램의 예

그림 16은 최종적으로 학습을 마치고 구성된 신경망 분류기를 이용하여 오토바이 영상들을 테스트한 예이다. 우측에 있는 값들이 신경망 분류기를 통하여 출력되는 결과값이며 이 결과값을 이용하여 종류를 분류한다. 분류된 결과는 텍스트박스의 하단에 표시된다.

3.2 결과 분석

영상 종류당 10개, 30종류, 총 300개의 영상을 신경망에 학습시켜서 학습된 영상 300개와 테스트용으로 사용한 영상 300개, 총 600개의 영상을 실험 데이터로 사용하였다. 실험에 사용한 질감 특징들은 contrast, diagonal moment, energy, entropy, homogeneity, 2nd diagonal moment, uniformity, 7가지이다. 각 질감 특징을 사용하여 배경 제거를 하지 않았을 때의 결과는 표 1, 배경을 제거했을 때의 결과는 표 2와 같다.

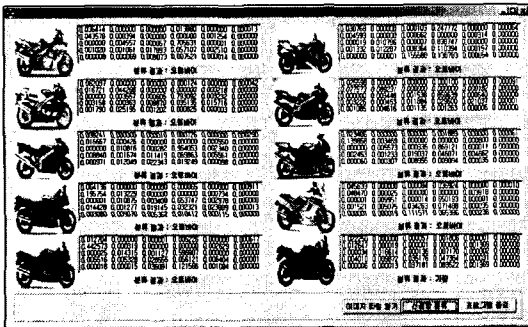


그림 16. 신경망 분류 실험 프로그램의 예

표 1. 배경 제거를 하지 않았을 때의 인식률 결과

질감 특징	학습데이터	테스트 데이터	총 데이터
Contrast	178/300 (59.3%)	159/300 (53.0%)	337/600 (56.2%)
Diagonal Moment	217/300 (72.3%)	201/300 (67.0%)	418/600 (69.7%)
Energy	163/300 (54.3%)	144/300 (48.0%)	307/600 (51.2%)
Entropy	153/300 (51.0%)	135/300 (45.0%)	288/600 (48.0%)
Homogeneity	126/300 (42.0%)	110/300 (36.7%)	236/600 (39.3%)
Second DM	182/300 (60.7%)	166/300 (55.3%)	348/600 (58.0%)
Uniformity	173/300 (57.7%)	157/300 (52.3%)	330/600 (55.0%)

표 2. 배경을 제거했을 때의 인식률 결과

질감 특징	학습데이터	테스트 데이터	총 데이터
Contrast	204/300 (68.0%)	188/300 (62.7%)	392/600 (65.3%)
Diagonal Moment	245/300 (81.7%)	230/300 (76.7%)	475/600 (79.2%)
Energy	194/300 (64.7%)	170/300 (56.7%)	364/600 (60.7%)
Entropy	185/300 (61.7%)	163/300 (54.3%)	348/600 (58.0%)
Homogeneity	158/300 (52.7%)	145/300 (48.3%)	303/600 (50.5%)
Second DM	210/300 (70.0%)	197/300 (65.7%)	407/600 (67.8%)
Uniformity	201/300 (67.0%)	188/300 (62.7%)	389/600 (64.8%)

실험을 한 결과, 배경 제거를 안했을 때보다는 했을 때가 인식률이 더 높았으며, 7가지 질감 특징들 중 대각 모멘트(diagonal moment)가 가장 높은 인식률을 보여주었다. 학습 데이터의 경우엔 72.3%, 테스트 데이터는 67.0%, 총 600개의 모든 데이터에 대하여 69.7%의 인식률을 보여주었다.

결과를 분석하여 보았을 때 인식 결과가 높은 영상의 종류는 손목시계, 의자, 케이크, 오토바이 등이었다. 그리고 인식 결과가 낮은 영상의 종류는 전차, 헬기, 조개, 말, 금강화 등이었다.

다양한 영상의 종류들을 분석한 결과, 인식률이 높은 종류는 그림 17과 같이 특징값이 규칙적으로 분포되어 있는 경우에 좋은 결과를 보여주었다.

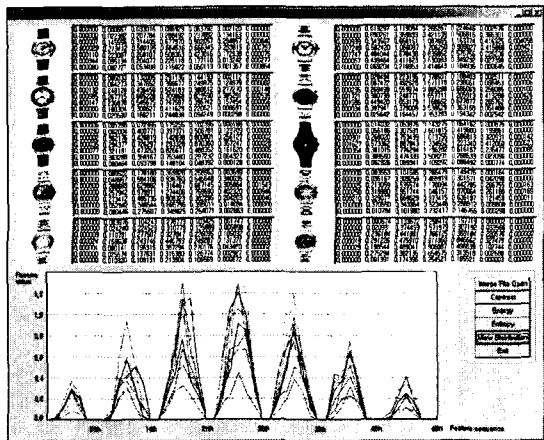


그림 17. 특징값 분포가 규칙적인 예

낮은 인식률을 보인 종류에 대해서는 다음과 같은 영상 데이터의 분석 및 이해에 따른 어려움이 원인이 되었다.

첫째, 다양한 영상의 종류들 중에선 서로 형태와 질감이 유사한 종류들이 다수 존재한다는 점이다. 그림 18은 그러한 예로써 사과와 토마토의 경우에 그 형태와 질감이 유사하고 추출된 특징값 또한 그 분포가 균일함을 볼 수 있다.

둘째, 영상내의 객체 자체가 형태와 질감이 복잡한 경우이다. 그림 19는 전차의 예로서 그 형태와 질감이 복잡하여 특징값 분포에서 특정 규칙을 찾기가 어려움을 알 수 있다.

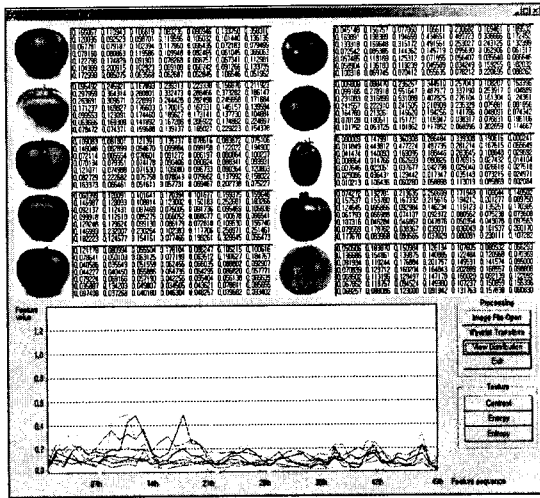


그림 18. 서로 형태와 질감이 유사한 예

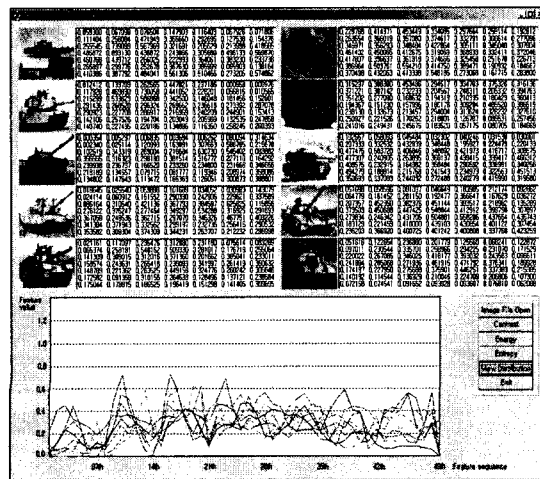


그림 19. 객체의 형태와 질감이 복잡한 예

4. 결 론

본 논문에서는 배경을 제거하여 오분류의 원인이 되는 잡음을 최소화하고, 웨이블릿 변환을 이용하여 영상의 형태 정보를 획득하며, 여러 가지 특징들 중 질감 특징을 구조적으로 추출하여 이를 학습패턴으로 하는 신경망을 통해 영상을 분류하는 방법을 제안하였다. 실험 결과, 배경 제거를 했을 때 보다 높은 인식률을 보여주었으며, 다양한 질감 특징들 중에서는 대각 모멘트(diagonal moment)가 가장 높은 인식률을 보여주었다.

하지만 위의 실험 결과 분석에서와 같이 영상 데이터의 분석 및 해석상의 어려움으로 인하여 인식률이 상당히 저하되었다. 이는 내용기반 영상 검색에서의 기본적인 난점이기도 하다. 하지만 인터넷상의 방대한 영상들을 자동 분류하여 보다 효율적으로 검색하고 관리할 수 있는 가능성을 제시하였다.

앞으로의 연구과제는 배경 제거에 사용한 영역 분할을 개선하여 보다 정확히 배경을 제거하여 분류기의 인식률을 높이고자 한다. 또한 특징 추출에서는 전체 영상에서 각각의 질감 특징값들을 하나씩 추출하고, 이 값들을 학습패턴으로 사용하는 방법에 대해 연구해보고자 한다. 그리고 인터넷상의 다양한 영상들을 분류해내기 위해서는 각 영상 종류들의 특징값 분포에 따른 트리구조의 계층도를 구성하고, 각각의 그룹들을 현재의 단일 신경망을 이용하여 분류하는 계층적 신경망(hierarchical neural network)을 구성하여 보다 많은 종류의 영상들을 대상으로 높은 검색 결과를 얻을 수 있도록 개선하고자 한다.

참 고 문 헌

- [1] J. P. Eakins and M. E. Graham, "Content-based Image Retrieval," *Technical Report, University of Northumbria at Newcastle, UK*, 1999.
- [2] A. Yoshitaka and T. Ichkawa, "A Survey on Content-based Retrieval for Multimedia Database," *IEEE Trans. on Knowledge and Data Engineering*, Vol. 11, No. 1, pp. 81-93, 1999.
- [3] Batber, W. Equitz, C. Faloutsos, "Query By

Content for Large On-Line Image Collection," *IEEE*, 1995.

[4] Ang, Y. H., Li, Z. and Ong, S. H., "Image Retrieval based on Multimensional Feature Properties," *Proceedings of SPIE Storage and Retrieval for Image and Video Database III*, No. 2420, pp. 47-57, 1995.

[5] Y. Deng, B. S. Manjunath and H. Shin, "Color Image Segmentation," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 446-451, 1999.

[6] Y. Deng, C. Kenney, M. S. Moore and B. S. Manjunath, "Peer Group Filtering and Perceptual Color Image Quantization," *IEEE International Symposium on Circuits and Systems*, pp. 21-24, 1999.

[7] Rafael C. Gonzalez, Richard E. Woods, *Digital Image Processing*, Addison-Wesley, ISBN 0-201-60078-1, pp. 512-523, 1998.

[8] Ramesh Jain, Rangachar Kasturi, and Brain G. Schunck, *Machine Vision*, ISBN 0-07-032018-7, pp.234-248, 1995

[9] 안영아, 장주만, 김지균, 김태윤, "2차원 웨이블릿을 이용한 이미지 검색 시스템 설계," 한국정보과학회 '99 춘계학술발표논문집, 제6권, 제1호, pp. 1390-1412. 1999.

[10] 김진숙, 김호성, "색과 공간정보를 이용한 영상 검색 기법," 한국정보과학회 '97 추계학술발표논문집, 제24권, 제2호, pp. 483-486, 1997.

[11] 강영미, "Wavelet과 신경망을 이용한 내용기반 얼굴 검색 시스템," 창원대학교 석사논문, 1998.

[12] Milan Sonka, Vaclav Hlavac, Roger Boyle,

Image Processing, Analysis, and Machine Vision 2/E, PWS Press, ISBN 0-534-95393-X, pp. 651-653, 1999.

[13] 김희승, 영상인식-영상처리, 컴퓨터비전, 패턴인식, 신경망, 생능출판사, 1994.

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

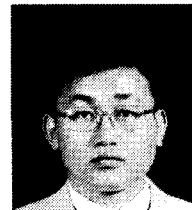
[15] L. Fausett, *Fundamentals of Neural Networks*, Prentice Hall, pp. 289-333, 1994.



이 재 원

2000년 인제대학교 전산학과 졸업(이학사)
 2002년 인제대학교 대학원 전산학과 졸업(이학석사)
 2002년~현재 (주)투원정보시스템 근무

관심분야 : 정보검색, 패턴인식, 영상처리
 E-mail : jwviolet@shinbiro.com



김 상 균

1991년 경북대학교 통계학과 졸업(이학사)
 1994년 경북대학교 대학원 컴퓨터공학과 졸업(공학석사)
 1996년 경북대학교 대학원 컴퓨터공학과 졸업(공학박사)
 1996년~현재 인제대학교 정보컴퓨터공학부 조교수

관심분야 : 정보검색, 정보보호, 패턴인식
 E-mail : skkim@cs.inje.ac.kr