

논문-00-5-1-02

MPEG-7 질감 기술자

강 호 경*, 정 용 주*, 유 기 원*, 노 용 만*, 김 문 철**, 김 진 응**

MPEG-7 Texture Descriptor

Ho-Kyoung Kang*, Yong Ju Jung*, Ki Won Yoo*, Yong Man Ro*, Munchul Kim**, and Jinwoong Kim**

요 약

본 논문에서는 MPEG-7 표준 비주얼 분야 기술자에서 영상의 질감 정보의 표현 방법에 대해 다룬다. 질감 기술자는 멀티미디어 콘텐츠 표현에 대한 국제 표준에서 칼라, 윤곽, 객체나 카메라 움직임 등과 함께 중요한 비주얼 특징으로 다루어지고 있으며 표준화의 대상이 되고 있다. 본 논문에서는 현재 MPEG-7 표준안으로 잠정 채택된 휴먼 비주얼 시스템(Human Visual System, HVS)에 기반한 질감 기술자에 대하여 논한다. 이 방법은 라돈(Radon) 변환을 통하여 영상을 시각인지 특성에 맞게 변환하고, 뇌의 비주얼 피질(Visual Cortex)의 반응과 일치하게 다수개의 주파수 채널들로 질감을 분해한다. 그리고 각 채널들에서의 평균 에너지와 에너지의 편차가 영상의 질감 기술자로 추출한다. 본 논문에서 질감 영상 검색에 있어서 현재 MPEG-7 실험 데이터베이스인 T1~T7에서 검색 유효를 측정하였다. 실험결과 본 영상 질감 기술자 표현 방법은 모든 실험 데이터베이스에서 우수한 검색 성능과 빠른 특징 추출 능력을 보였다.

Abstract

In this paper, we present a texture description method as a standardization of multimedia contents description. Like color, shape, object and camera motion information, texture is one of very important information in the visual part of international standard (MPEG-7) in multimedia contents description. Current MPEG-7 texture descriptor has been designed to fit human visual system. Many psychophysical experiments give evidence that the brain decomposes the spectra into perceptual channels that are bands in spatial frequency. The MPEG-7 texture description method has employed Radon transform that fits with HVS behavior. By taking average energy and energy deviation of HVS channels, the texture descriptor is generated. To test the performance of current texture descriptor, experiments with MPEG-7 Texture data sets of T1 to T7 are performed. Results show that the current MPEG-7 texture descriptor gives better retrieval rate and fast extraction time for texture feature.

I. 서 론

현재 컴퓨터와 인터넷의 발달과 더불어 멀티미디어 데이터양은 기하급수적으로 증가하고 있다. 이러한 방대한 양의

멀티미디어 콘텐츠에 대한 효율적인 검색, 저장, 관리, 전송 등을 위해 멀티미디어 콘텐츠를 효과적으로 표현 및 검색하기 위한 MPEG-7 표준화 작업이 진행되고 있고, MPEG-7 비주얼 분야는 주로 영상의 비주얼 특징(칼라, 윤곽, 질감, 움직임 등)에 대한 기술자(descriptor)를 표준화 대상으로 하고 있다^{[1][2]}. 여기서 기술자는 영상과 비디오의 내용 기반 특징을 표현하는 것으로, 특징 표현의 배열과 의미의 체계로 정의된다.

영상의 질감 정보는 색, 윤곽, 움직임 등과 함께 영상의

* 한국정보통신 대학원 대학교, 영상/비디오 연구실
Information and Communications University (ICU), IVY lab

** 한국전자통신연구원 무선방송기술연구소 방송미디어 연구소 방송미디어연구부
Broadcasting Media Technology Department, ETRI - Radio & Broadcast Technology Lab.

내용 정보를 기술하는 중요한 시각적인 특징의 하나로, 의료, 우주 항공, 영상 편집 및 대기 영상 등의 영상 분석, 객체 인식, 표면 분석 등과 관련하여 최근 30여 년 간 폭 넓게 연구되어오고 있다^{[3][4][5]}. 일상에서 자주 접하는 질감 정보로는 옷감의 무늬, 벽, 타일 등의 균질 또는 비 균질의 패턴을 예로 들 수 있으며, 이러한 질감은 영상 화소간의 주기성 또는 반주기성으로 설명된다. 이 질감정보는 영상의 구조나, 방향성, 거친 정도 등의 특성을 나타내고, 영상 또는 비디오 데이터를 내용 기반 색인(indexing)하거나 요약하는데 있어서 중요한 특징으로 사용된다.

이러한 질감 정보를 분석하려는 많은 시도들이 있어 왔는데 과거에는 화소간의 확률 분포^[3], 방향성 필터^[4], 마르코프 랜덤 필드(Markov random field)^[6]를 이용한 방법이 연구되었으며, 최근에는 사람의 시각인지 특성 즉 저주파 성분의 변화에 민감하고 고주파 성분의 변화에 둔감한 HVS 특성에 기반한 가보(Gabor) 필터^{[7][8]}와 웨이블릿 변환^[9]을 이용한 방법들이 활발히 연구되고 있다.

본 논문에서는 MPEG-7 비주얼 부분의 위원회 작업 초안(Working Draft)에 채택된 HVS 특성^{[10][11][12][13]}에 기반하여 영상의 질감 특징을 추출하는 방법과 함께 유사도 측정 과정에서 회전 불변, 배율 불변 방법으로 질감 영상을 검색할 수 있는 유사도 계산방법에 대한 소개를 한다^{[14][15][16][17]}.

본 논문의 구성은 다음과 같다. 2장에서는 질감 특징 추출 알고리즘에 대해 설명하고, 3장에서는 질감 검색 알고리즘에 대하여 설명하고 4장에서 실험 과정 및 결과를 설명한다. 그리고 마지막으로 5장에서 결론을 내린다.

II. 질감 특징 추출 알고리즘

1. 질감 특징을 위한 HVS 구조

시각의 인지에 대한 뇌의 기능을 규명하는 정신물리(Psychophysical) 실험에 따르면 비주얼 피질(Visual Cortex)의 외부 자극에 대한 반응은 대역폭이 제한된 공간 주파수의 몇 개의 밴드로 분할되는 것으로 알려지고 있다^[18]. 즉, 외부 물체는 몇 개의 공간 주파수 영역의 채널로 분해(decomposition)되어 뇌에서 인식된다. 여러 실험결과를 통해서 이제까지 알려진 가장 적절한 공간 주파수 분할은 각도 방향과 방사 방향으로 기본 축들을 구성하는 공간 주파수 영역에서 각도방향으로는 동일 각도 폭으로 180도의 전 각도영역을 5에서 6개로 분할, 방사방향에서는 전체 주

파수영역을 4에서 5개의 옥타브 밴드로 나누는 방법이다^[12]. 위의 HVS 특성을 따라서 본 논문에서는 질감 특징 추출에 사용되는 채널을 각도 방향과 방사 방향으로 다음과 같이 분할함으로써 얻었다.

그림 1은 방사방향에서 영역을 5개로 분할한 것을 그린 것으로 ω_n 은 n 번째 분할영역의 중심 주파수를 나타낸 것이다. 각 영역의 분할은 전체 주파수 범위를 N으로 하였을 때 전체 영역을 연속적인 옥타브 밴드로 나눔으로써 이루어진다.

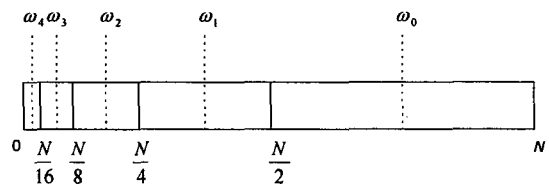


그림 1. 방사 방향으로 채널 분할.
Fig. 1. Channel division in radial direction

편 HVS 필터는 주파수 영역에서 각도 방향으로 180/P의 균등 분할로 나눈다. 여기서 P는 각 방향의 분할 수이며 본 논문에서는 기본 질감 패턴 실험을 통하여 6개로 각도 방향을 분할하였다.

이렇게 하여 분할된 공간주파수 영역은 그림 2에 나타내었다. 그림 2에서 보는 것처럼 각 분할 영역의 하나는 HVS 채널로 외부 물체를 뇌에서 분해 해석하는 기본 주파수 영역이 된다. 그림에서 보는 것과 같이 HVS 채널은 저주파 영역에서는 조밀한 채널들로 구성되고 고주파 성분에서는 상대적으로 넓은 영역의 채널들로 구성됨을 알 수 있다. 즉 시각 특성은 저주파 성분의 변화에 민감하고 고주파 성분의 변화에 둔감한 특성을 보인다. w_x 축이나 w_y 축은 주파수 공간에서의 좌표축을 의미하는데 실험적으로 영상

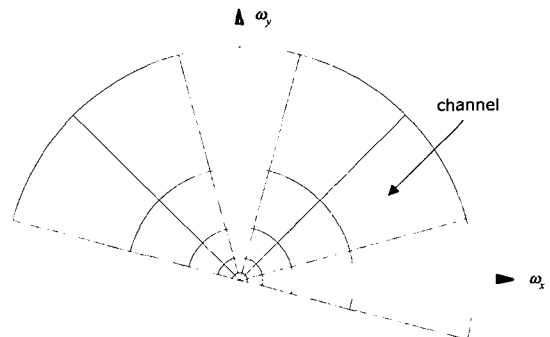


그림 2. HVS 채널에 의한 주파수 공간의 분해
Fig. 2. Frequency region division with HVS filter.

의 중요한 특징이 이 부분에 존재함으로 HVS 채널에서 이 영역을 포함하기 위하여 15° 정도 기울어지게 된다. 이런 채널 특성에 기반 하여 본 논문에서는 질감 특징을 효과적으로 추출한다.

2. HVS에서 데이터 샘플링

이 절에서는 위에서 설명한 HVS 특성에 맞게 영상을 변환하고 특징 기술자를 얻기 위한 방법을 설명한다. 그림 2에서 보듯이 각 채널의 모양은 부채꼴 모양이다. 주어진 영상 혹은 비디오 키 프레임에서 이런 모양의 채널에 알맞은 데이터를 얻기 위해 다음과 같이 라돈(Radon) 변환을 수행한다. 라돈 변환은 직교 좌표 계의 (x, y) 좌표를 (R, θ)로 표현되는 라돈 영역으로 변환하는 것으로 라돈 변환에 의해 2차원 입력 영상은 1차원의 시노그램(Sinogram)으로 변환된다. 라돈 변환의 수학적 의미는 어떤 각도 θ에서 선, L(R, θ)를 따르는 선 적분이며, 이것은 식 (1)로 표현된다.

$$p_{\theta}(R) = \int_{L(R, \theta)} f(x, y) dl \tag{1}$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \delta(x \cos \theta + y \sin \theta - R) dx dy$$

여기서 f(x, y)는 2차원 영상을 의미하며, pθ(R)은 각 θ에서 2차원 영상을 선 적분에 의하여 1차원 신호로 변환한 것으로 투영(Projection)데이터 또는 시노그램이라 부른다. 이러한 영상의 각 θ에서의 선 적분 과정을 라돈 변환이라고 한다.

이 시노그램을 1차원 푸리에 변환하면 원 영상의 2차원 푸리에 변환과 동일한 효과를 얻을 수 있다(Central slice theorem). 즉 그림 3과 같이 2차원 푸리에 영역에서 각 중심을 지나는 선은 그 각도에 대응하는 시노그램을 1차원 푸리에 변환 한 것과 같다.

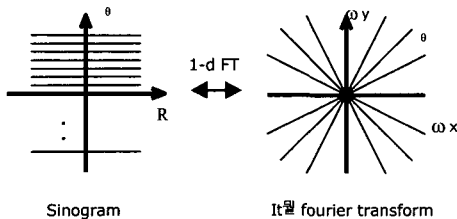


그림 3. 시노그램과 2차원 푸리에 영역과의 관계.
Fig. 3. Relationship between Sinogram and 2 dimensional Fourier domain.

그림 3의 시노그램에서 R은 원점에서의 거리, θ는 각도를 의미한다.

그림 4는 1차원 푸리에 변환을 수행한 시노그램을 직교 좌표계로 배치한 것이다. 직교 좌표 계에서 라돈 변환된 데이터는 저주파와 중간 주파수 영역에서 보다 조밀한 데이터 샘플링 구조를 가지게 되며, 고주파 영역에서는 상대적으로 밀도가 적은 샘플링 구조를 갖는다. 이것은 1절에서 논의한 질감에 대한 HVS 특성과 관련하여 분해된 각 채널들의 특성, 즉 저주파 정보변화에 대한 강조 및 고주파 정보 변화에 대한 상대적 중요도 감소와 일치한다. 따라서 라돈 변환이 수행된 영상 정보는 HVS 특성을 가지게 된다.

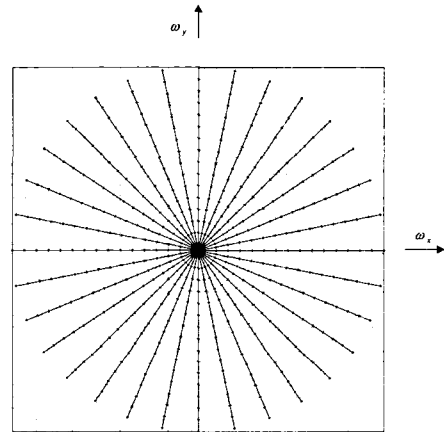


그림 4. Radon 변환을 통한 데이터 샘플링 구조.
Fig. 4. Data sampling structure using Radon transform.

결론적으로 HVS 채널에서의 질감을 기술하고자 하는 방법으로 다음 두 가지로 요약 할 수 있다. 첫째 HVS 특성을 갖는 채널들로 공간 주파수를 분해한다. 둘째, 라돈 변환된 영상 정보로부터 분해된 각 채널 내의 데이터를 구성한다.

그림 2와 같이 나누어진 HVS 채널들로 질감 특징을 표현하는데 각 채널사이에 완화효과(relaxation)를 주기 위하여 가보(Gabor) 필터를 사용하였다. 이 가보 필터는 주파수 영역을 분할함에 있어서 주위의 채널과 중복되게 된다. 즉 한 채널의 가보 필터는 주위의 채널에도 값을 가지게 되어서 채널의 중복이 일어난다. 이러한 중복에 의한 완화효과로 인하여 뒤에 설명할 회전 불변 알고리즘에서 특히 좋은 결과를 가져온다.

식 (2)는 극 좌표계에서 사용된 가보 필터를 표현한 것이다.

$$G_{mn}(\omega_r, \theta) = \exp\left\{-\frac{1}{2}\left[\frac{(\omega_r - \omega_m)^2}{\sigma_\omega^2} + \frac{(\theta - \theta_n)^2}{\sigma_\theta^2}\right]\right\} \quad (2)$$

여기서 $G_{mn}(\omega_r, \theta)$ 는 m번째 방사 방향과 n번째 각 방향에서의 가보 필터를 의미하고 σ_ω 와 σ_θ 는 각각 방사방향과 각도 방향에서의 분산으로 필터의 폭을 나타낸다. 필터 폭은 인접하는 다른 필터들과 최대 필터 높이의 1/2이 되는 위치에서 필터들의 폭이 인접하도록 설계한다. 그림 5는 각도방향으로 6분할, 방사방향으로 5분할로 구성된 가보 필터들을 그린 것이다. 그림 5는 필터의 높이가 최대높이의 1/2 이상 되는 부분만 그린 것이다.

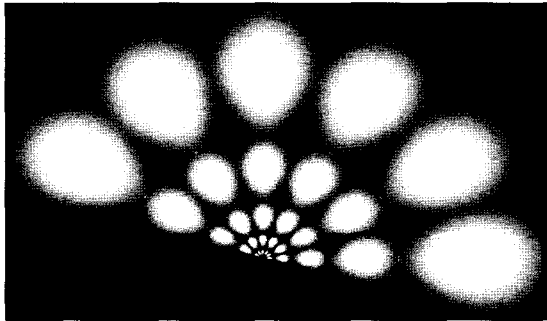


그림 5. 극 좌표계에서 표현된 5x6 가보 필터들
Fig. 5. 5x6 Gabor filters in polar coordinate system

3. 질감 특징 벡터의 추출

앞 절에서 언급한 HVS 구조하에 질감 특징 벡터의 추출은 다음과 같은 크게 둘로 나눌 수 있다.

첫 번째는 영상을 HVS 특성의 채널로 분해하기 전의 질감 특성을 나타내는 것이고, 두 번째는 HVS 특성의 채널로 분해 후 질감특성을 나타내는 것이다.

첫째, 입력 영상으로부터 전체 영상의 평균 에너지(dc)와 표준 편차(std)를 계산한다.

전체 영상의 평균 에너지 값은 영상의 밝기에 따라 달라지나 질감의 밝기 정보를 필요로 하는 검색에서는 중요한 요소가 된다. 그리고 영상의 전체 화소에 대한 표준 편차는 영상의 포괄적인 특성을 파악하는데 유용하다^[7].

둘째, 질감을 위한 HVS 구조 하에서 특징 벡터를 추출한다.

입력 영상 혹은 비디오의 프레임으로 식 (1)에서 설명한 라돈 변환을 수행하여 시노그램을 얻는다. 이후 시노그램은 그림 3에서 보이듯이 푸리에 변환하여 입력 영상의 2차원

푸리에 변환을 그림4의 데이터 샘플링 구조를 갖도록 한다.

다음으로 주파수 영역을 시각 특성에 맞게 그림2에서와 같은 채널들로 분해한다.

그림 2와 같이 분해된 각 채널 위에 그림 5의 가보 필터가 씌워진다. 가보 필터링 후의 특정 채널내의 신호를 P_θ (ω) 라고 한다면 질감 특징은 각 채널내의 이 신호들의 에너지와 에너지의 편차로 정의된다. 다음 식은 에너지와 에너지 편차를 나타낸 것으로 최종 특징 값은 로그 스케일링한 값으로 나타내었다.

$$p_{mn} = C_{mn} \sum_{\omega} \sum_{\theta} [P_\theta(\omega)]^2 \quad (3)$$

$$e_{mn} = \log(1 + p_{mn}) \quad (4)$$

$$q_{mn} = \sqrt{D_{mn} \sum_{\omega} \sum_{\theta} [(P_\theta(\omega))^2 - p_{mn}]^2} \quad (5)$$

$$d_{mn} = \log(1 + q_{mn}) \quad (6)$$

여기서, m과 n은 각각 방사 방향과 각도 방향에서 채널을 나타내는 지수로, 방사방향으로 저주파 영역에서 고주파 영역으로 0에서 4까지를, 0° 에서 180° 까지를 0 에서 5로 순서대로 구분하는 정수이다. C_{mn} 과 D_{mn} 는 각 채널에서 각각 정규화된 에너지 평균 및 편차를 구하기 위한 상수이다.

p는 하나의 채널에 포함된 신호의 평균 에너지를 나타내고, q는 에너지의 편차를 나타낸다. 이 값들은 로그 스케일링에서 e와 d를 얻는다.

따라서 질감특징 벡터의 구성은 5x6 HVS 채널에서 에너지와 에너지의 편차에 의하여 60개의 특징이 추출되고 전체 영상의 평균 에너지 값과 편차로 이루어진다. 결론적으로 식 (7)과 같이 총 62개의 질감 특징 값으로 이루어진다.

$$Descriptor_{texture} = [dc \quad std \quad e_{00} \quad e_{01} \dots e_{45} \quad d_{00} \quad d_{01} \dots d_{45}] \quad (7)$$

4. 질감 기술자의 양자화

식 (7)에서 정의되는 각 특징 벡터들은 실수형으로, 이러한 값들의 저장 및 전송을 위해서는 특징 벡터 당 4byte를 필요로 한다. 이때, 전체 질감 특징 벡터들의 크기는 248byte가 된다. 248byte의 저장 공간은 10만장의 영상으로 구성된 데이터베이스를 가정하였을 때, 질감 정보만으로 25Mbyte의 공간을 필요로 하게 된다. 이러한 큰 저장 공간

은 질감 정보의 저장 및 전송과 관련하여 큰 제약으로 작용한다. 이러한 큰 저장 공간은 특징 벡터를 양자화하는 과정에서 감소할 수 있다.

본 알고리즘에서는 특징 벡터들의 양자화 과정에서 각 특징 벡터들의 최대값과 최소값에 의한 256단계의 선형 양자화를 이용한다. 8비트로 질감 특징을 표현하면 양자화를 수행하기 전과 검색을에서 크게 변화가 없고 데이터베이스에 영향을 받지 않기 위해 선형 양자화를 수행하였다. 이때 각 특징 벡터들의 최대값과 최소값은 실험 데이터에서 가장 큰 값과 작은 값을 구하여 사용한다. 특징 벡터의 양자화 과정을 식 (8)에 나타내었다.

$$D_{quant} = \left[\frac{D_{nonquant} - \beta_{min}}{\beta_{max} - \beta_{min}} \times q_level \right] \times \frac{\beta_{max} - \beta_{min}}{q_level} + \beta_{min} \quad (8)$$

여기서 D_{quant} 는 양자화된 특징 벡터를 의미하고 $D_{nonquant}$ 는 실수 형식의 실제 특징 벡터를 의미한다. β_{max} 와 β_{min} 는 실험데이터에서 구한 각 특징 벡터들의 최대값과 최소값이며, q_level 은 양자화 단계의 수로 255로 정의된다.

식 (8)의 과정에 의해 질감 특징 벡터는 벡터 당 1byte의 크기만을 가지게 되며, 따라서 전체 질감 특징 벡터들의 크기는 62byte로 줄어들게 된다.

III. 질감 특징 검색 알고리즘

현재 MPEG-7 표준안으로 채택된 질감 기술자는 영상 및 비디오의 검색함에 있어서 자유도가 있다. 이번 장에서는 질감 기술자에 기초한 검색 알고리즘을 설명한다. 참고로 검색 알고리즘은 MPEG-7 표준화 분야는 아니다.

1. 유사도 측정

검색을 위하여 유사도 측정으로 i 와 j 의 두 영상간의 거리 함수 $d(i, j)$ 를 구한다. 거리함수는 식 (9)과 식 (10)에서 정의된다.

$$d(i, j) = \text{distance}(\text{Descriptor}_{texture}^i, \text{Descriptor}_{texture}^j) \quad (9)$$

$$= Wb + \sum_{m=0, n=0}^{S-1, K-1} \left[W_{e, mn} \left| \frac{e_{mn}^{(i)} - e_{mn}^{(j)}}{\alpha(e_{mn})} \right| \right] + \sum_{m=0, n=0}^{S-1, K-1} \left[W_{d, mn} \left| \frac{d_{mn}^{(i)} - d_{mn}^{(j)}}{\alpha(d_{mn})} \right| \right]$$

여기서

$$b = \left| \frac{dc^{(i)} - dc^{(j)}}{\alpha(dc)} \right| + \left| \frac{std^{(i)} - std^{(j)}}{\alpha(std)} \right| \quad (10)$$

식 (9)와 (10)에서 dc , std , emn , d_{mn} 는 식 (7)에서 보듯이 질감 특징 값을 나타내고 S 와 K 는 각각 5, 6로 정의되며 W 와 $W_{e, mn}$, $W_{d, mn}$ 는 특징 값간의 가중치를 부여하는 값이며, $\alpha(\cdot)$ 는 전체 데이터베이스에 대한 각 특징 값들의 평균으로 특징 값들의 정규화 과정에서 사용된다.

2. 밝기 불변의 유사도 측정

밝기 특징은 유사 영상의 평균 에너지 값과 관련되는 것으로, 밝기 불변에서의 유사도 측정은 식 (10)에서 dc 특징 값을 제외한 후 수행한다. 이것을 식 (11)에서 나타내었다.

$$b = \left| \frac{std^{(i)} - std^{(j)}}{\alpha(std)} \right| \quad (11)$$

3. 회전 불변 유사도 측정

MPEG-7 질감 특징 기술자에 영상의 회전에 상관없이 검색기능을 부여하기 위해 이 절에서는 회전 불변 검색 알고리즘에 제시한다. 만일 각도의 회전이 30° 내에서 이루어지면 1절에서 측정한 유사도 검사는 회전 불변을 요구할 때도 유효하다. 이는 각도방향으로 전체 180° 를 6개로 분할했기 때문이다. 30° 보다 큰 회전은 다음과 같은 방법으로 거리를 측정할 수 있다.

질감 특징 기술자는 그림 2에서 보듯이 2차원 푸리에 영역이므로 입력 영상의 회전에 대해서 채널들이 각도 방향으로 회전한 만큼의 각도로 이동하는 특성이 있다. 따라서 회전불변 검색을 위한 거리 측정에서 회전 각도를 고려하면 회전 불변의 검색을 수행할 수 있다.

거리 함수를 θ 에 의한 함수로 놓으면 식 (12)와 같이 된다.

$$d(i, j, m\theta) = \text{distance}(\text{Descriptor}_{texture}^i |_{m\theta}, \text{Descriptor}_{texture}^j) \quad (12)$$

여기서 $\text{Descriptor}_{texture}^i |_{m\theta}$ 는 i 영상을 $m\theta$ 각만큼 이동

시켜 얻은 질감기술자를 나타낸다. 각도 θ 는 $180/L$ 값으로 단위 각도가 된다. m 은 1에서 L 까지 정수이다. 회전 불변 질감 특징 유사도 검색은 식 (13)에 의한 거리 함수에 의하여 측정된다.

$$d(i, j) = \min\{d(i, j, m\phi) | m = 1 \text{ to } L\} \quad (13)$$

여기서 “min { }”은 0도에서부터 180도까지 단위각도로 증가하면서 회전한 영상과의 거리 중 최소값을 의미한다.

3.1. 계층적인 회전 불변 유사도 측정

식 (13)에 따르면 회전 불변 유사도를 측정을 위하여 원하는 회전 각도에 따라서 모두 거리 측정을 수행하여야 한다. 이러한 방법은 만약 5° 단위로 유사도 측정을 수행하는 경우 한 영상에 대하여 36번의 거리 측정을 수행해야 하는 단점이 존재한다.

따라서 회전 불변 유사도 측정시 측정 횟수를 단축하기 위하여 다음과 같은 계층적인 유사도 검색을 수행할 수 있다.

예로 30° 단위로 6번의 검색을 수행하여 가장 가까운 부분을 선택하고 선택된 부분으로부터 10° 단위로 4번의 검색을 수행한 뒤 5° 단위로 2번의 유사도 측정을 수행하여 3단계로 회전 불변 유사도 측정을 수행한다.

단계 1 : 다음 식을 만족하는 $\theta_{ref}^1 = m\phi$ 을 구한다.

$$\min\{d(i, j, m\phi) | m = 1 \text{ to } 6, \phi = 30^\circ\} \quad (14)$$

단계 2 : 다음 식을 만족하는 $\theta_{ref}^2 = \theta_{ref}^1 + \Delta\theta$ 를 구한다

$$\min\{d(i, j, \theta_{ref}^1 + \Delta\theta) | \Delta\theta = \pm 10^\circ, 0^\circ, \pm 20^\circ\} \quad (15)$$

단계 3 : 최종적인 거리를 측정한다.

$$d(i, j) = \min\{d(i, j, \theta_{ref}^2 + \Delta\theta) | \Delta\theta = 0^\circ, \pm 5^\circ\} \quad (16)$$

여기서 θ_{ref} 는 각 레벨에서 최소거리를 구할 때 기준이 되는 각도를 의미하고 $\Delta\theta$ 는 식 (15), (16)에서 보이듯이 2단계에서는 10° 단위로 3 단계에서는 5° 단위로 변화하면서 거리를 측정하게 된다.

그림 6은 3 단계의 계층적인 회전 불변 유사도 측정의

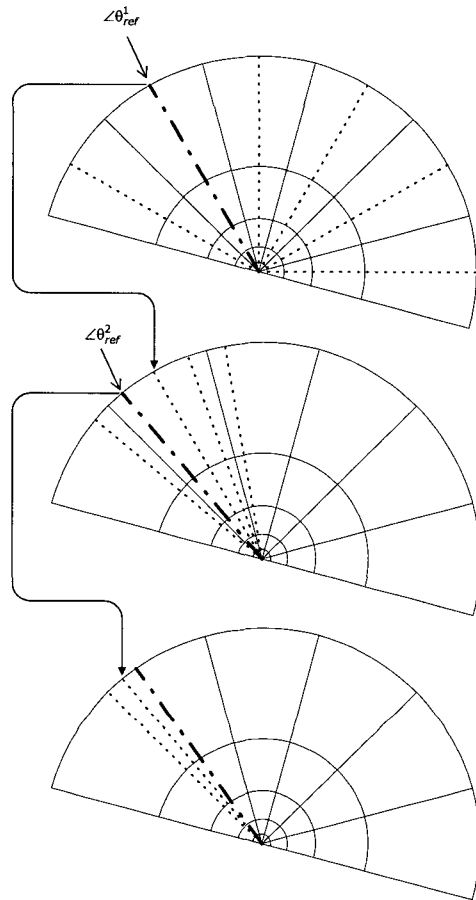


그림 6. 3 레벨 계층적인 유사도 측정

Fig. 6. 3-Level hierarchical minimum distance searching.

예를 보인다. 이 방법에 의하여 유사도 측정을 12번으로 줄일 수 있다.

4. 크기 불변 유사도 측정

크기 불변의 경우도 방사 방향으로 채널의 폭 넓이가 존재하므로 어느 정도 1절에 제시된 유사도를 이용하여 측정할 수 있다. 그러나 크기가 많이 변화한 경우 고주파부분 채널의 불일치가 발생되지 않도록 고주파 영역의 채널을 유사도 측정과정에서 배제함으로써 크기 불변의 질감 영상 검색을 수행할 수 있다. 크기 불변에서의 유사도 측정을 식 (17)에서 나타내었다.

$$d(i, j) = Wb + \sum_{m=1, n=1}^{S-1, K-1} \left[W_{e, mn} \left| \frac{e_{mn}^{(i)} - e_{mn}^{(j)}}{\alpha(e_{mn})} \right| \right] + \sum_{m=1, n=1}^{S-1, K-1} \left[W_{d, mn} \left| \frac{d_{mn}^{(i)} - d_{mn}^{(j)}}{\alpha(d_{mn})} \right| \right] \quad (17)$$

5. 회전 / 크기 불변 유사도 측정

회전 불변과 크기 불변의 유사도 측정을 동시에 수행할 때는 3절과 4절의 알고리즘을 동시에 수행한다. 이러한 과정을 식 (18), (19)에서 보인다.

$$d(i, j, k\theta) = Wb + \sum_{m=1, n=1}^{S-1, K-1} \left[W_{z, mn} \left| \frac{e_{mn, k\theta}^{(i)} - e_{mn}^{(j)}}{\alpha(e_{mn})} \right| \right] + \sum_{m=1, n=1}^{S-1, K-1} \left[W_{d, mn} \left| \frac{d_{mn, k\theta}^{(i)} - d_{mn}^{(j)}}{\alpha(d_{mn})} \right| \right] \quad (18)$$

$$d(i, j) = \min \{ d(i, i, k\theta) \} \quad (19)$$

6. 질감 특징 기술자의 계층화

질감 특징의 계층화는 전체 특징 벡터에서 높은 성능을 갖는 특징 벡터들에 대해 저장, 검색 및 전송 과정에서 우선 순위를 부여하는 것으로 전체 특징 벡터를 모두 사용하지 않고서도 일정 이상의 성능을 보이도록 한다.

계층화의 유용한 사용은 무선 인터넷과 같은 열악한 네트워크 환경에서의 사용자가 데이터베이스에 접근하여 검색을 수행하는 경우를 생각할 수 있다. 이러한 환경에서 전체 특징 벡터들의 일부만이 전송되는 경우가 종종 발생하게 되며, 이때 계층화는 사용자가 이미 전송된 일부 특징 벡터들만으로도 만족할 만한 검색이 가능하도록 한다. 실험에서 결정된 계층화는 질감 특징 벡터에 대해 다음의 식 (20)과 같이 정의된다.

$$Descriptor_{texture} = Descriptor_{texture}^{basic\ scale} + Descriptor_{texture}^{enhanced\ scale} \quad (20)$$

여기서 $Descriptor_{texture}^{basic\ scale}$ 는 질감 특징을 나타내는 기본 기술자이고 다음 식과 같이 HVS 채널의 평균 에너지 값으로 구성된다

$$Descriptor_{texture}^{basic\ scale} = [std\ dc\ e_{00}\ e_{01} \dots e_{45} \dots e_{45}] \quad (21)$$

더욱 검색율을 높이기 위해 개선 기술자인 $Descriptor_{texture}^{enhanced\ scale}$ 가 기본 기술자에 더하여져 사용하며 개선 기술자는 HVS 채널의 에너지 편차값으로 구성된다. 개선 기술자는 다음 식 (22)와 같이 나타낼 수 있다.

$$Descriptor_{texture}^{enhanced\ scale} = [d_{00}\ d_{01} \dots d_{45} \dots d_{45}] \quad (22)$$

IV. 실험 및 결과

앞장에서 언급된 질감 기술자의 성능을 측정하는 실험을 위하여 위해 본 장에서는 MPEG-7 질감 실험용 데이터베이스를 이용하여 검색율을 측정된 결과를 보인다.

1. MPEG-7 질감기술자 성능 실험

MPEG-7에서 질감과 관련한 실험데이터베이스는 모두 7개가 있으며, 각각 T1, T2, T3, T4, T5, T6, 그리고 T7이라 부른다.

1) T1 데이터베이스

T1 데이터베이스는 질감 패턴이 기록되어 있는 Brodatz 앨범의 영상들을 스캔한 것으로, 실험을 위해서는 스캔된 116개의 512x512 크기의 영상을 중복되지 않게 1856개 (=116x16)의 128x128 크기의 영상으로 나누어 구성된다.

이러한 데이터베이스를 가진 실험에서 검색율(RR: Retrieval Rate)은 데이터베이스의 각 영상에서 질의하였을 때 동일한 원 영상에 속한 나머지 15개의 영상을 몇 개나 찾는가에 의해 결정된다.

$$RR = \frac{\# \text{ of relevant retrieved images in the 15 matches}}{15} \quad (23)$$

식 (23)에서 분모의 수는 같은 질감으로 분류된 영상들 거리 함수에 의하여 15개중에 몇 개를 찾았는지를 나타내는 값이다. T1 데이터베이스에서의 검색율(AvRR: Average Retrieval Rate)은 각 영상에 대해 검색율을 체크한 후 이것을 평균한 값이 된다.

$$AvRR = \frac{\sum_{i=1}^{db_num} RR_i}{db_num} \quad (24)$$

식 (24)에서 RR_i은 한 영상에 대한 검색율을 뜻하고 db_num은 전체 질의된 영상의 수로 1856개이다. 그림 7은 T1 데이터베이스에 있는 특정 영상에의 검색 예를 보인다. 좌측 위에 있는 영상이 질의영상이고 나머지가 검색된 영상들이다.

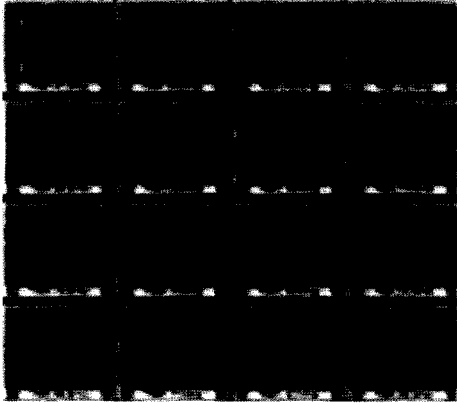


그림 7. T1 데이터베이스 검색.
Fig. 7. Retrieval image in T1 Data base.

2) T2 데이터베이스

T2 데이터베이스는 실험을 위해 한국정보통신대학원에서 자체 제작한 데이터베이스로 1999년 3월 MPEG Seoul meeting에서 정식 자료로 등록되었다. T2 데이터베이스는 자연 영상 및 인공물에서 질감부분을 디지털 카메라를 이용하여 획득한 52개의 512x512 크기의 영상들로 구성되어 있으며, T2 데이터베이스는 T1 데이터베이스에 있는 영상보다 전체적으로 균일하다. 이에 따라서 보다 좋은 검색을 가진다. 질감으로 실험을 위해 T1 데이터베이스와 동일한 과정에서 832개의 128x128 크기의 영상들로 나누어 사용한다. T2 데이터베이스에서의 성능 평가 또한 T1과 동일한 과정에서 수행된다.

그림 8는 T2 데이터베이스에서의 검색 예를 보인다. 마찬가지로 좌측 위에 있는 영상이 질의영상이고 나머지가 검색된 영상들이다.

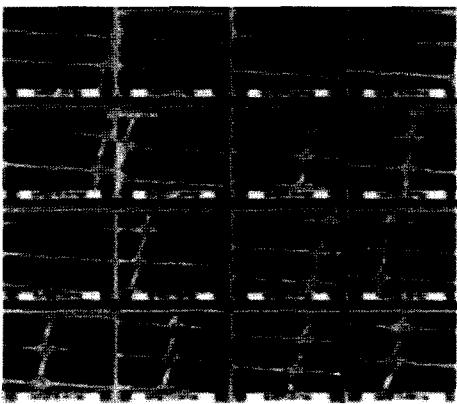


그림 8. T2 데이터베이스 검색.
Fig. 8. Retrieval image in T2 Data base.

3) T3 데이터베이스

T3 데이터베이스는 512x512 크기의 원 영상을 중o심에서 10, 15, 20, 30, 40, 50, 70, 75, 80, 100, 110, 130, 135, 140, 160, 170에 대하여 회전한 후 다른 위치에서 128x128 크기의 영상을 무작위로 잘라낸 것으로, T3 데이터베이스의 발생을 위해 30개의 T1 데이터베이스와 25개의 T2 데이터베이스를 이용한다. 이러한 과정에서 정의되는 전체 데이터베이스의 수는 880개 가 된다

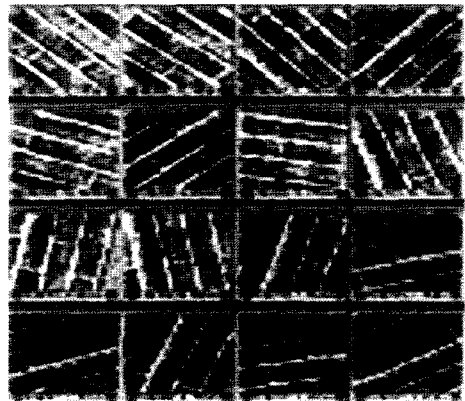


그림 9. T3 데이터베이스 검색.
Fig. 9. Retrieval image in T3 Data base.

T3 데이터베이스에서의 성능 평가는 55개의 30° 회전된 영상을 질의하였을 때, 원 영상에서 회전된 나머지 15개의 영상을 몇 개나 찾는가에 의해 결정된다. 그림 9는 T3 데이터베이스의 검색 예를 보인다.

4) T4 데이터베이스

T4 데이터베이스는 512x512 크기의 원 영상을 50에서

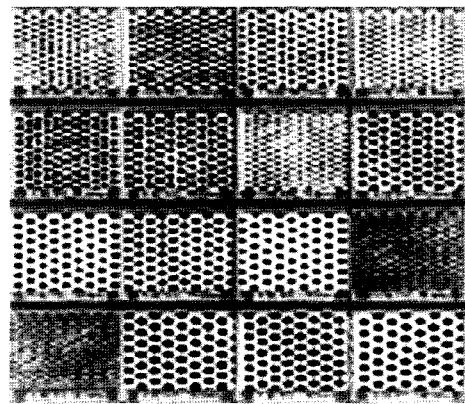


그림 10. T4 데이터베이스 검색.
Fig. 10. Retrieval image in T4 Data base.

150%까지 5%간격으로 크기를 변화한 후 다른 위치에서 128x128크기의 영상을 무작위로 잘라낸 것으로, T4 데이터베이스의 발생을 위해 116개의 T1 데이터베이스를 이용한다. 이러한 과정에서 정의되는 전체 T4 데이터베이스의 수는 2436개가 된다.

이렇게 발생된 T4 데이터베이스는 다시 크기의 변화 정도에 따라 T4a)에서 T4j)까지 10개의 데이터베이스로 재분류된다. T4a)는 100%의 원 영상에 5%의 크기를 변화한 95%와 105%의 영상들로 구성되며, T4b)는 T4a)에 10%의 크기를 변화한 90%와 110%의 영상을 추가하여 구성된다. 이러한 과정을 반복하여 T4i)는 T4h)에 45%의 크기를 변화한 영상으로 구성되며, 마지막으로 T4j)는 T4i)에 50%의 크기를 변화한 영상을 포함한 전체 T4 데이터베이스로 정의된다.

T4 데이터베이스에서의 성능 평가는 재분류된 10개의 T4 데이터베이스에서 116개의 128x128크기의 원 영상을 질의하였을 때, 크기가 변화된 나머지 영상을 몇 개나 찾는데 의해 결정된다. 그림 10은 T4 데이터베이스의 검색 예를 보인다.

5) T5 데이터베이스

T5 데이터베이스는 Corel 앨범의 영상에서 질감을 잘 표현하고 있는 영상을 고른 것으로 다른 데이터베이스와는 달리 한 영상에 대하여 15개의 비슷한 질감 특징이 존재하는 것이 아니라 각각의 검색하고자 하는 영상에 대하여 같은 질감 분류된 영상의 수가 일정하지 않다. 같은 질감 형태가 40개가 같은 것에서부터 4개만 같은 것까지 있다.

총 검색하고자 하는 수(Query)는 16개를 가지고 총 영상의 수는 2400개의 중에서 찾아야 되는 영상(Ground Truth)은 216개이다. 식 (25)은 현재 T5에 대한 검색을 측정 예를 보인다.

$$RR = \frac{\# \text{ of relevant retrieved images in the Ground Truth}}{\# \text{ of Ground Truth}} \quad (25)$$

그림 11는 T5 데이터베이스의 검색 예를 보인다.

위 검색 결과와 같이 경우 질의에 대하여 같은 질감 특징을 가지는 영상이 3개만 존재하는 경우이다. 따라서 위의 4개만 검색된 결과로 보면 된다.

6) T6 데이터베이스

T6 데이터베이스는 20개의 항공 지도 영상에서 128x128로 잘라낸 34,000개의 영상 중에서 T5와 같은 방식으로 질의 영상과 Ground Truth를 선택하여 검색을 수행한다.

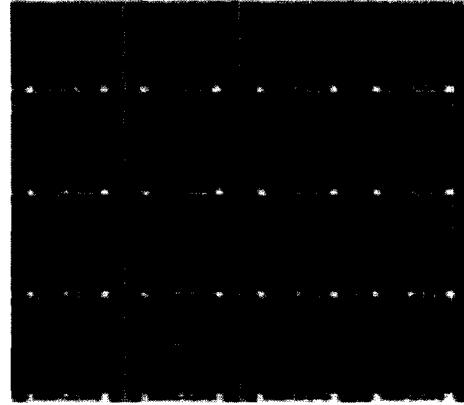


그림 11. T5 데이터베이스 검색.
Fig. 11. Retrieval image in T5 Data base.

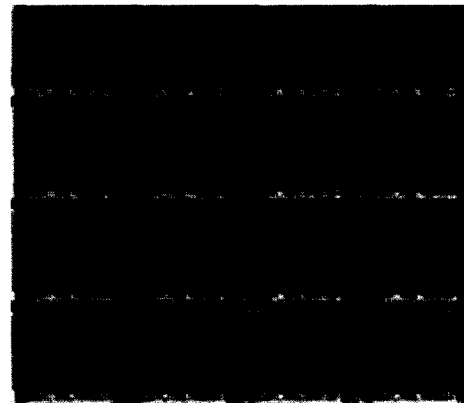


그림 12. T6 데이터베이스 검색.
Fig. 12. Retrieval image in T6 Data base.

그림 12는 T6 데이터베이스의 검색 예를 보인다.

7) T7 데이터베이스

T7 데이터베이스는 512x512 크기의 원 영상을 0, 8, 25, 55, 107, 131, 174 의 6개 각도로 회전한 후, 다시 90%, 80%, 70%, 60% 50%로 크기를 변화한 후, 다른 위치에서 128x128 크기의 영상을 무작위로 잘라낸 것으로, T7 데이터베이스의 발생을 위해 30개의 T1 데이터베이스와 25개의 T2 데이터베이스, 15개의 T5 데이터베이스를 사용하였다. 이러한 과정에서 발생하는 전체 T7 데이터베이스의 수는 2450개가 된다.

T7 데이터베이스에서의 성능 평가는 0° 회전의 70%로 크기를 변화한 영상들을 질의 영상으로 하여, 각 질의에서 나머지 34개의 영상을 몇 개나 찾는가에 의해 결정된다. 따라서, T7 데이터베이스에서의 실험 결과는 0° 회전에서

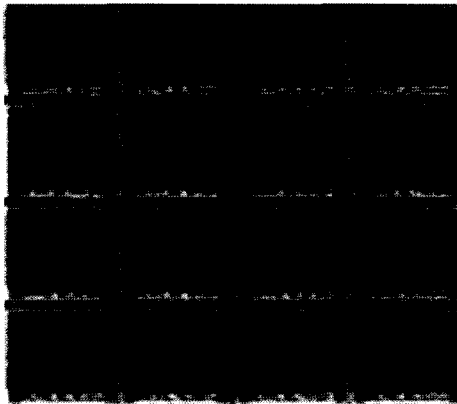


그림 13. T7 데이터베이스 검색.
Fig. 13. Retrieval image in T7 Data base.

5개의 크기 변화에서 각각 구해지는 AvRR로 주어진다. 그림 13은 T7 데이터베이스의 검색 예를 보인다.

2. 실험 결과

1) 검색율

표 1은 MPEG-7 실험 데이터베이스에서의 검색율을 나타낸다.

표 1. MPEG-7 실험 데이터베이스에서 검색율 측정.
Table 1. Retrieval Rate in MPEG-7 Test Database.

Data set	AVRR(검색율)	
T1	77.32	
T2	90.67	
T3	93.45	
T4	T4a	86.21
	T4b	87.50
	T4c	88.94
	T4d	87.07
	T4e	84.91
	T4f	84.55
	T4g	83.87
	T4h	82.60
	T4i	79.60
	T4j	76.72
T5	64.60	
T6	75.75	
T7	85.76	

표 2. T1(Brodarz DB)에서 검색율 비교.
Table 2. Comparison of Retrieval Rate in T1(Brodarz DB).

	본 논문	Gabor ^[5]	PWT ^[19]	TWT ^[19]	MRSAR ^[5]
AVRR	77.32	74.37	68.70	69.41	73.18

표 2는 T1 데이터베이스에 대하여 다른 논문의 결과와의 비교를 수행하였다. 질감 검색에서 그동안 우수하다 알려진 방법인 직교 좌표계에서의 가보 필터에 의한 방법은 74.37%의 검색율이 보고되고^[5], 또 다른 방법인 wavelet을 이용한 방법인 PWT(Pyramid-structured wavelet transform)와 TWT(Tree-structured wavelet transform)의 경우도 70% 이하의 검색율을 가지고 있다^[19]. 본 논문의 경우 같은 데이터베이스에서 77.32%의 검색율을 나타내는 것으로 보다 우수한 알고리즘임을 알 수 있다.

표 1에서 T2는 T1보다 규칙적인 질감이 더 많은 관계로 90.67%의 검색율을 가진다.

T3은 회전 불변 알고리즘으로 검색을 수행하여 92%의 검색율을 가진다. 이때 선택한 질감 영상은 매우 규칙적인 영상을 T1과 T2에서 선택을 하고 영상에 회전을 하여 구성되어 있으므로 회전 불변 알고리즘을 테스트하는 데이터베이스이다.

T4의 경우 T4a에서 T4j까지 크기 변화가 더 심해지는 데이터베이스임으로 크기의 변화가 작은 데이터베이스에서 상대적으로 높은 검색율을 가지고 T4j로 가면서 검색율이 점차 하락함을 볼 수 있다.

T5, T6은 기본적으로 검색하는 부분에서 차이가 있는데 다른 데이터베이스에서는 포함하고 있는 모든 영상에 대하여 거리 측정을 통하여 가장 가까운 영상 15개를 선택해야 하는데 T5, T6은 질의 영상이라는 거리 측정을 통하여 검색을 수행하는 영상이 고정적으로 정해져 있다. 일반적인 검색을 수행하려면 T5, T6의 데이터베이스에서 검색하는 것과 동일한 효과를 지닌다.

T7 데이터베이스는 회전과 크기를 변화시킨 데이터베이스임으로 회전 불변 알고리즘과 크기 불변 알고리즘에 대한 테스트 알고리즘을 위한 데이터베이스이다.

2) 특징 추출 시간

질감 특징 추출에 대하여 시간 또한 매우 중요한 요소이다. 본 알고리즘은 주파수 공간에서 특징을 추출하므로 Radon 변환에 걸리는 시간이 질감 특징 추출에 걸리는 시간의 대부분을 차지한다.

펜티엄 400MHz, 윈도우 NT환경에서 128x128 영상에 0.14초 정도의 빠른 특징 추출 시간을 가진다.

V. 결 론

본 논문에서는 질감 정보를 효과적으로 분석하기 위한 방법으로 HVS 특성에 기반 하여 질감 특징 추출 알고리즘과 유사도 측정 방법들을 설명하였다.

본 논문에서는 질감 특징을 추출하기 위한 주파수 분할에서 저 주파수에 민감하고 고 주파수에서 둔감한 인간의 시각인지 특성을 고려하여 저주파 대역에서는 세밀하게, 주파수에서 듬성하게 분할을 수행한다. 또한 라돈 변환은 저 주파수에서 많은 샘플링을 수행함 HVS 특성을 반영하고 있으며 극 좌표계 임으로 5x6의 주파수 분할을 하기 위하여 필요한 각도 즉 θ 방향으로의 분할도 쉽게 수행이 가능하여 질감 특징을 추출하는데 매우 효과적이다.

본 알고리즘은 MPEG-7의 실험 데이터베이스를 이용한 객관적인 성능 비교에서 우수한 성능과 빠른 특징 추출 시간을 가졌다. 또한 제안된 회전 및 배율 불변 알고리즘 역시 관련된 실험에서 높은 성능을 보였다.

향후 본 논문의 알고리즘은 시각적 인지도의 향상 및 알고리즘의 보안을 통한 객관적 성능의 향상, 구체적 응용 방안의 연구 개발을 통해 국제 표준 기술로의 채택을 위해 지속적인 연구가 수행될 예정이다. 이러한 과정에서 구체화된 기술은 실생활에서의 내용 기반 멀티미디어 데이터베이스의 검색뿐만 아니라 직물 영상의 데이터베이스 검색, 의료 영상, 항공 사진 검색, 대기 관측, 지질 탐사 등의 많은 영상에 대하여 질감 정보를 가지고 검색이 필요한 경우 매우 중요한 요소기술로 사용이 기대된다.

참 고 문 헌

- [1] "MPEG-7 Visual Part of Experimentation Model Version 4.0," *ISO/MPEG, m3068, Maui*, Dec. 1999.
- [2] AHG on Editing The Document of The MPEG-7 Visual Part of XM and WD, *ISO/MPEG, w3335, Noordwijkerhout*, May, 2000
- [3] O. D. Faugeras and W. K. Pratt, "Decorrelation Methods of Texture Feature Extraction," *IEEE Trans. Pattern Anal. and Machine Intell.*, vol. 2, pp. 323-332, Jul. 1980
- [4] I. Fogel and D. Sagi, "Gabor Filters as Texture Discriminator," *Biological Cybernetics*, vol. 61, pp. 103-113, 1989.
- [5] B. S. Manjunath and W. Y. Ma, "Texture Features for Browsing and Retrieval of Image Data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 8, Aug. 1996.
- [6] R. Chellappa, "Two-dimensional Discrete Gaussian Markov Random Field Models for Image Processing," *Pattern Recognition*, vol. 2, pp. 79-112, 1985.
- [7] P. Wu, W. Y. Ma, B. S. Manjunath, H. D. Shin, and Y. L. Choi, "A Texture Descriptor for MPEG-7," *ISO/MPEG*, pp 77, Lancaster, 1999.
- [8] Y. M. Ro, Matching Pursuit : Contents Featuring for Image Indexing, *Proceedings of SPIE*, vol. 3527 pp. 89-100, 1998.
- [9] J. R. Ohm and F. Bunjamin, "Descriptor for Texture in Wavelet Domain," *ISO/MPEG*, P566, Lancaster, 1999.
- [10] A. Saadane, H. Senane, D. Barba, "On the Design of Psychovisual Quantizers for a Visual Subband Image Coding," *SPIE*, vol. 2308, pp1446, 1994
- [11] A. Saadane, H. Senane, D. and Barba, "An Entirely Psychovisual Based Subband Image Coding Scheme," *SPIE*, vol. 2501, pp. 1702, 1995.
- [12] J. G. Daugman, "High Confidence Visual Recognition of Persons by a Test of Statistical Independence," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 15, no. 11, pp. 1148-1161, Nov. 1993.
- [13] C. J. Lambrecht, "A Working Spatio-Temporal Model of Human Vision System for Image Restoration and Quality Assessment Applications," *IEEE International Conference on ASSP, New York, NY, USA*, P. 2291-4 vol. 4., 1996
- [14] Y. M. Ro, S. Y. Kim, K. W. Yoo, M. C. Kim, and J. W. Kim, "Texture Descriptor Using Atoms for Matching Pursuit," *ISO/MPEG, P612, Lancaster 1999*.
- [15] Y. M. Ro, K. W. Yoo, M. C. Kim, and J. W. Kim, "Texture Description Using Radon Transform," *ISO/MPEG, m4703, Vancouver*, 1999.
- [16] Y. M. Ro, K. W. Yoo, M. C. Kim, and J. W. Kim, "Texture Description Using Radon Transform and Experimental Results on CT-5 Core Experiment Using Atoms for Matching Pursuit," *ISO/MPEG, m5152, Melbourne* 1999.

[17] Y. M. Ro, K. W. Yoo, M. C. Kim, J. W. Kim, B. S. Manjunath, D. G. Sim, H. K. Kim, and J. R. Ohm, "An Unified Texture Descriptor ISO/MPEG," *m5490*, Maui 1999.

[18] J. G. Daugman, "Complete Discrete 2-D Gabor Transforms by Neural Networks for Image Analysis and Compression," *IEEE Trans. ASSP*, vol. 36, pp. 1160-1179, Jul. 1988

[19] T. Chang, C. C. J. Kuo, "Texture Analysis and Classification with Tree-Structured Wavelet Transforms," *IEEE Trans. Image Processing*, vol. 2, no. 4 pp. 429-441, Oct. 1993

저 자 소 개



강 호 경

1998년 : 고려대학교 전자공학과, 학사
 1998년 9월 ~ 2000년 2월 : 한국전자통신연구원 위촉연구원
 1998년 9월 ~ 현재 : 한국정보통신대학원 석사과정
 주관심분야 : 영상처리, MPEG-4, MPEG-7, Data Hiding



정 용 주

1999년 : 홍익대학교 컴퓨터공학과, 공학사
 1999년 : (주)데이콤 위촉연구원
 1999년 3월 ~ 현재 : 한국정보통신대학원 석사과정
 주관심분야 : 영상처리, MPEG-7, Watermarking



유 기 원

1998년 : 고려대학교 전자공학과, 학사
 2000년 : 한국정보통신대학원, 석사
 2000년 2월 ~ 현재 : 삼성전자 중앙연구소(R&D coporate center) DMS lab,에서 MPEG-7관련 업무
 졸업논문 : Texture analysis and classification based on the Human Visual System



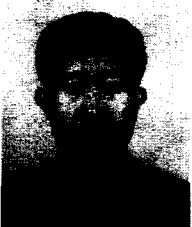
노 용 만

1985년 : 연세대학교 전자공학, 학사,
 1987년 : KAIST 전기 및 전자공학, 석사
 1992년 : KAIST 전기 및 전자공학, 박사
 1996년 ~ 1997년 : University of California, Berkeley, 연구원
 1988년 ~ 현재 : IEEE Senior member, SPIE, ISMRM member
 1997년 ~ 현재 : 한국정보통신대학원대학교, 교수(부교수)



김 문 철

1989년 : 경북대학교 전자공학과, 학사,
 1992년 : University of Florida, Electrical and Computer Engineering, 석사,
 1996년 : University of Florida, Electrical and Computer Engineering, 박사,
 1997년 1월 ~ 현재 : 한국전자통신연구원, 선임연구원
 2000년 3월 ~ 현재 : 한국전자통신연구원 방송미디어연구부 실감영상연구팀장
 주관심분야 : 멀티미디어 정보처리 및 검색, 멀티미디어 통신, 디지털 신호/영상처리



김 진 응

1981년 : 서울대학교 공과대학 전자공학과 학사,
 1983년 : 서울대학교 대학원 전자공학, 석사,
 1993년 : Texas A&M Univ. Electrical Engineering, 박사,
 1983년 3월 ~ 현재 : 한국전자통신연구원, 책임연구원.
 주관심분야 : 디지털 신호 처리, 영상 통신, 디지털 라이브러리, VLSI 신호처리