

# 공간 히스토그램과 웨이브릿 모멘트의 융합에 의한 영상검색

서 상 용<sup>o</sup>, 손 재 곤, 김 남 철  
경북대학교 전자.전기학부  
전화 : (053) 950-5530 / 팩스 : (053) 950-5505

## Image Retrieval Using the Fusion of Spatial Histogram and Wavelet Moments

Sang Yong Seo, Jae Gon Son, Nam Chul Kim  
School of Electronics and Electrical Engineering Kyungpook National University  
E-mail : syseo@palgong.knu.ac.kr

### Abstract

We present an image retrieval method that improves retrieval rate by using the fusion of histogram and wavelet moment features. The key idea is that images similar to a query image are selected in DB by using the wavelet moment features. Then the result images are retrieved from the selected images by using histogram method. In order to evaluate the performance of the proposed method, we use Brodatz texture database, MPEG-7 T1 database and Corel Draw photo. Experimental result shows that the proposed method is better than each of histogram method and wavelet moment method.

### I. 서론

최근 컴퓨터 및 네트워크 통신의 발달로 많은 양의 멀티미디어 정보를 다양한 형태로 접하고 있으며 이러한 방대한 양의 데이터를 관리하고 효율적으로 검색할 필요성을 느끼게 된다. 초기의 검색 시스템에서 많이 사용되는 텍스트 기반 검색(text-based retrieval)방법은 색인 구축이 수동으로 수행되는 불편이 있으며 데이터가 가지는 속성을 정확하게 표현하기 힘들다. 반면에 내용 기반 검색(content-based retrieval)은 색상, 질감, 모양과 같은 시각적 특징을 자동으로 추출하여 데이터베이스를 구축하므로 텍스트 기반보다 효율적으로 정보를 저장 및 검색할 수 있다.

내용기반 검색에 사용되는 특징중에 질감(texture) 특징은 객체의 표면이 가지는 특성을 표현하고 있기 때문에 영상의 내용 정보를 기술하는 가장 기본이 되는 특징이다. 이러한 질감 특징은 화소의 확률분포, 마코프 랜덤 필드(Markov random field)[1]를 이용하여 추출하기도 하고 최근에는 인간의 휴먼 비주얼 특성에 기반한 가보(Garbor)필터[2]와 웨이브릿 변환[3]을 통해서 질감 정보를 주파수 대역별로 표현한 정보로 변환하여 추출하기도 한다.

내용기반 검색에 사용되는 특징중에 화소의 확률분포를 이용한 히스토그램 기법은 가장 널리 이용되는 기법중 하나이며 영상의 회전(rotation)이나 이동(trans-

lation)에도 상당히 안정된 성능을 나타내고, 계산상 복잡도가 적어 검색속도가 빠른 장점도 있지만, 밝기와 같은 변화에 상당히 민감하며 공간 정보를 가지지 못하므로 유사한 히스토그램을 가지는 서로 다른 종류의 영상을 구분하지 못하는 단점이 있다. 다음은, 웨이브릿 변환을 이용한 내용기반 영상 검색은 인간의 시각 특성에 적합하게 영상을 다해상도(multi-resolution)로 분해하여 웨이브릿 계수들의 통계적인 특성을 이용한 방법이다[4]. 인간 시각은 세밀한 부분보다는 불연속성이나 갑작스런 밝기의 변화 등과 같은 해상도가 낮은 영상에 대해 더 민감한 반응을 보이므로, 이러한 특성을 이용하면 효과적인 검색결과를 얻을 수 있다. 그러나 웨이브릿 모멘트(moment)는 영상 내에서 질감을 이루는 기본 단위인 텍셀(texel)이나 물체(object)가 확대, 축소되거나 혹은 이동 및 회전을 하면 검색율이 떨어지는 단점을 가지고 있다.

본 논문에서는 공간 히스토그램과 웨이브릿 모멘트를 융합하여 질감 영상의 검색 효율을 향상시키는 방법을 제안한다. 이 방법은 웨이브릿 모멘트를 사용하여 질의 영상과 유사한 특징(feature)을 가진 영상들을 DB에서 가려내고 나서, 그 선택된 영상들을 다시 공간 히스토그램 기반법으로 검색하는 방식이다. 이렇게 함으로 유사한 히스토그램을 갖는 서로 다른 종류의 영상을 구분하지 못하는 히스토그램 기법의 단점을 공간 정보를 가진 웨이브릿 모멘트가 보완하여 줄 수 있다. 또한 웨이브릿 모멘트의 단점을 공간 히스토그램의 장점으로 보완하여 줄 수도 있다.

### II. 내용기반 영상 검색 기법

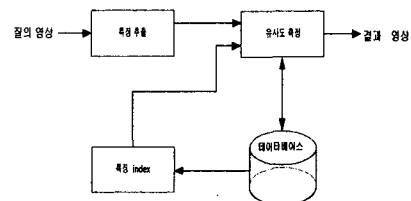


그림 1. 내용 기반 영상 검색 블록도

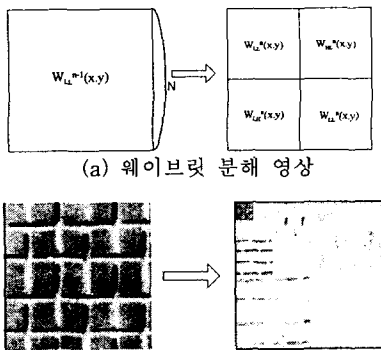
영상의 특징을 자동으로 추출하여 이를 기반으로 영상 데이터를 검색하는 내용 기반 영상검색은 그림 1과 같은 과정으로 수행된다. 질의 영상으로부터 특징을 추출하고 DB에 있는 영상들의 특징과 유사도를 비교하여 가장 유사한 영상 순서대로 배열하여 원하는 영상수만큼 출력한다. 각 블록별 기능을 살펴보면 다음과 같다.

1. 특징 추출

히스토그램기법의 특징은 영상의 그레이(gray) 값의 빈도수인 히스토그램을 이용한다. 그레이 레벨이  $n$ 인 영상의 히스토그램은 다음과 같이  $n$ 차원의 특징 벡터로 볼 수 있다.

$$f^h = [h(0), h(1), h(3) \cdots h(n)] \quad (1)$$

여기서,  $h(n)$ 은 입력영상  $I(x, y)$ 의 히스토그램이다. 웨이블릿 모멘트 특징 추출을 위한 웨이블릿 변환 [5]에 대하여 살펴보면, 그림 2 (a)에 입력 및 웨이블릿 분해된 영상을 보인다.  $W_{LL}^0(x, y)$ 를 입력영상이라 하면 이를 반복적으로 2차원 웨이블릿 변환하여 다해상도 피라미드를 구성한다. 각  $n$ 번째 웨이블릿 변환에서는 입력 영상  $W_{LL}^{n-1}(x, y)$ 에 수평 방향으로 저역 통과 필터 및 고역 통과 필터를 통과시키고, 변환된 계수를 간축(decimation)한다. 다음으로 수평방향으로 처리된 계수에 수직방향으로 저역 및 고역 통과 필터를 통과시키고, 변환된 계수를 간축하면, 한 단계 더 근사화 된 영상  $W_{LL}^n(x, y)$ 와 수직, 수평 및 대각선 방향의 영상 ( $W_{HL}^n(x, y)$ ,  $W_{LH}^n(x, y)$ ,  $W_{HH}^n(x, y)$ )를 얻게 된다.



(b) D1영상과 웨이블릿 분해 영상(3 level)

그림 2. 웨이블릿 분해과정 및 분해 영상

그림 2의 (b)는 Brodatz 질감 영상인 D1 영상을 3단계 웨이블릿 분해한 영상이다. 웨이블릿 변환 영역에서 웨이블릿 계수를  $W_m^n(x, y)$ 라 하자. 여기서  $n$ 은 차수를,  $m$ 은 LL, HL, LH, HH 대역을 나타낸다. 웨이블릿 계수의 절대치에 대한  $k$ 차 정규(regular) 모멘트는 다음과 같이 주어진다[4].

$$f_m^n(k) = \frac{1}{N} \sum_m \sum_n |W_m^n(x, y)|^k, \quad k=1, 2 \cdots \quad (2)$$

$k$ 가 1이면 변환계수의 절대치에 대한 평균이며 2차는 에너지를 나타낸다. 이 모멘트는 차수가 증가할수록 모멘트값이 기하급수적으로 증가하므로  $k$ 차 모멘트는

$k$ 제곱근하여 값을 줄여 준다. 정규화된 중앙(central) 모멘트는 다음과 같이 표현된다.

$$\mu_m^n = \frac{1}{N} \sum_m \sum_n \left| \frac{W_m^n(x, y)}{\sigma_{mu}^n} \right|, \sigma_m^n = \sqrt{\frac{1}{N} \sum_m \sum_n \left( \left| \frac{W_m^n(x, y)}{\sigma_{mu}^n} \right| - \mu_m^n \right)^2} \quad (3)$$

여기서,  $\sigma_{mu}^n$ 와  $\sigma_m^n$ 는 DB내의 모든 영상의  $mn$ 부대역이 가지는 1, 2차 모멘트의 표준편차이며 이것으로 각 부대역의 모멘트를 정규화함으로 각 부대역의 가중치를 동일하게 만든다. 이렇게 하므로 안정된 결과를 얻게 된다. 3단계 웨이블릿 분해 영상에서  $k$ 차 모멘트를 가지면 특징벡터의 차원은  $k \times 12$  (3 단계  $\times$  4 부밴드)가 된다.

2. 유사도 측정

대용량의 영상 DB를 검색하기 위해서는 각 특징 벡터에 알맞은 유사성 척도(distance measure)가 정의되어야 한다.  $k$  차원을 가진 입력영상의 특징 벡터와 DB영상의 특징 벡터사이의 유사도 척도는 다음과 같이 주어진다.

$$D_L(i, q) = ||f_i - f_q||^p \quad (4)$$

$p$ 는 메트릭(metric)의 차수를 나타내며  $L_1, L_2$  메트릭이 히스토그램기법 및 웨이블릿 모멘트기법에서 많이 사용된다.

3. 성능 평가

검색율(Retrieval Rate)은 검색된 영상들내에서 질의 영상과 동일한 부류에 속한 영상이 몇 개나 포함되어 있는가를 나타낸다. 이를 수식으로 나타내면 다음과 같다.

$$RR = \frac{R_r}{T_r} \quad (5)$$

여기서  $T_r$ 은 검색된 영상의 수이며  $R_r$ 은 질의 영상과 동일한 부류에 속한 영상 수이다. 따라서, 전체 DB에서의 평균 검색율(Averaged Retrieval Rate)은 각 질의 영상에 대한 검색율의 합을 전체 영상 수로 나눈 값으로 다음식과 같다.

$$AVRR = \frac{\sum_{i=1}^C RR_i}{C} \quad (6)$$

여기서  $C$ 는 DB내의 영상수를 나타낸다.

III. 공간 히스토그램과 웨이블릿 모멘트의 융합에 의한 영상검색

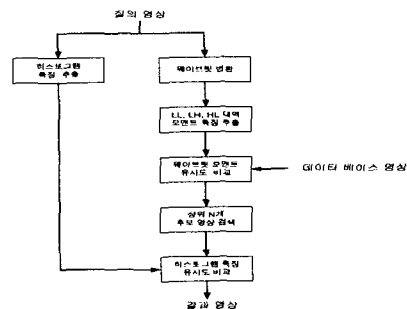


그림 3. 제안한 방법의 흐름도

그림 3은 제안한 방법의 블록도이다. 이 블록도의 흐름은 다음과 같다. 먼저, 질의 영상  $I(x,y)$ 가 들어오면 아래식과 같이  $n$  단계 웨이브릿 변환을 하여 그림 2와 같은 웨이브릿 분해 영상을 얻는다.

$$W_n^m(x,y) = WT[I(x,y)] \quad (7)$$

여기서  $WT$ 는 웨이브릿 변환을 나타내며, 질의 영상의 특징벡터를 웨이브릿 분해 영상에서 구한다. 각 부대역(HL, LH, HH)마다  $k$ 개의 모멘트를 가지므로  $n \times 3(HL, LH, HH) \times k$  차원의 특징벡터는 다음과 같다.

$$f^m = [\mu_{HL}^1 \sigma_{HL}^1 \mu_{LH}^1 \sigma_{LH}^1 \mu_{HH}^1 \sigma_{HH}^1 \dots \mu_{HH}^3 \sigma_{HH}^3] \quad (8)$$

그리고 DB에 있는 모든 영상 데이터의 웨이브릿 특징벡터  $f_q^m$ 와 유사도를 비교하여 유사도가 큰 상위  $N$ 개의 후보 영상을 가려낸다.

$$D_L(i,q) = |f_i^m - f_q^m| \quad (9)$$

그리고 선택된 후보 영상의 히스토그램 특징벡터와 질의 영상 히스토그램 특징벡터 사이의 유사도를 다시 구하여 유사도가 큰 순서대로 결과 영상을 얻는다.

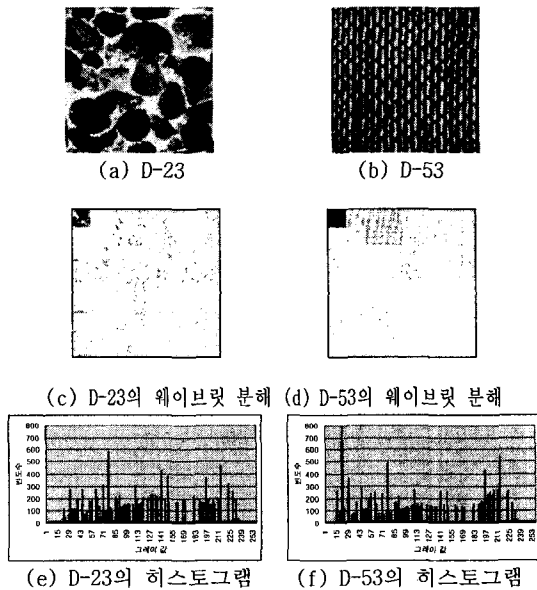


그림 4. Brodatz D-23과 D-53 영상의 웨이브릿 분해 영역 및 히스토그램

다음은 제안한 방법인 공간 히스토그램 특징과 웨이브릿 모멘트 특징을 융합한 방법을 분석한 내용이다. 그림 4(a)와 (b)는 Brodatz 질감 영상 중 23번째와 53번째의 데이터 영상이며, 이 두 영상의 질감은 서로 다르지만 히스토그램의 분포가 유사한 영상이다. 그림 4(a)영상을 질의 영상으로 입력하였을 경우, (a)의 히스토그램 분포와 유사한 그림 1(b)를 결과 영상으로

검색할 수가 있으므로 검색율이 떨어지게 된다. 그러므로 이러한 문제점은 히스토그램의 분포는 유사하지만 그림 4(c)와 (d)와 같이 웨이브릿 모멘트의 특징이 다르면 웨이브릿 모멘트 기법에 의해 사전에 걸러줄 수 있으므로 검색율을 증가시킬 수 있다. 여기서, 웨이브릿 특징벡터를 구할 때 LH, HL, HH대역의 모멘트만 사용하고 LL밴드의 모멘트는 사용하지 않는다. 이는 LL밴드는 공간 히스토그램의 특성과 비슷하여 히스토그램의 분포가 유사한 다른 부류의 영상을 배제시키지 못하기 때문이다.

#### IV. 실험결과 및 검토

실험에 사용한 DB영상은 Brodatz 영상 데이터 [6](111종:1776)와 MPEG-7의 실험 데이터로 제안되어 있는 Brodatz T1 영상 데이터[7](116종:1856)를 사용하였으며, 이들 영상 데이터는 모두 원영상(512 × 512)을 128x128크기의 영상 16개로 나누어서 동질한(homogeneous)질감을 갖는 영상들로 이루어져 있다. 그리고 Corel Draw Photo Album에서 분류되어 있는 자동차, 꽃, 경치, 마블로 구성된 128 × 128 크기의 영상 160장도 실험영상으로 사용하였다.

표 1은 웨이브릿 필터에 따른 검색율을 알아보기 위하여 Daubechies 16 탭 직교 웨이브릿 필터와 제안한 방법의 웨이브릿 필터로 사용된 Daubechies 9-7 탭 쌍직교(bi-orthogonal) 웨이브릿 필터 그리고 QMF 필터인 Irie 필터를 사용하여 3단계 웨이브릿 분해와 2차 모멘트를 이용하여 얻은 검색 결과이다.

표 1. 웨이브릿 필터에 따른 검색율 비교(Brodatz)

	Daubechies orthogonal 16tap	Daubechies Bi-orthogonal 9-7 tap	Irie 16 tap
AVRR	71.8%	72.2%	73.0%

여기서 보면 Daubechies 쌍직교 필터가 Daubechies 직교 필터보다 성능이 약간 우수하고, 그리고 Irie 필터의 성능이 제일 좋은 것으로 나타난다. 그러나, 복잡도도 함께 고려하여 이후의 실험에서는 Daubechies 9-7 탭 웨이브릿 필터를 사용하기로 한다.

표 2는 공간 히스토그램과 (1), (2)식의 웨이브릿 모멘트 그리고 제안한 방법으로 구한 결과를 보인 것이다. 공간 히스토그램 기법에서는 차원이 256이며  $L_{0.5}$  매트릭으로 구한 결과가  $L_1$  매트릭으로 구한 결과보다 6.5%정도 높음을 알 수 있으며, 웨이브릿 모멘트의 경우 3단계까지 웨이브릿 분해를 하였으며,  $L_{0.5}$ 과  $L_1$  매트릭에 의한 검색성능 차는 거의 없었다. (2)식의 중앙 모멘트를 특징벡터로 사용하면 (1)식의 정규 모멘트보다 검색율이 증가함을 알 수 있다. 그리고 제안한 방법에서 공간 히스토그램의 유사도를  $L_1$  매트릭으로 구한 경우, 공간 히스토그램 기법보다는 검색율이 7.3% 웨이브릿 모멘트를 이용한 기법보다 8.7% 향상되었다. 그리고  $L_{0.5}$  매트릭을 사용한 경우 5.3%와 3.0%가 각각 증가하였다. Corel Draw Photo Album의 자동차, 꽃, 경치, 마블 4종류의 영상에 대해서도 제안한 방법에 의한 검색율( $T_r=5$ )이 공간 히스토그램과 웨이브릿 모멘트만으로 검색한 결과보다 각각 21.1%와 2.3% 향상됨을 확인할 수 있었다. T1 데이터 경우 제안한 방법으로 구한 검색율이 오히려 공간 히스토그램

만으로 구한 검색율보다 떨어짐을 볼 수 있었다.

후보 영상의 개수는 공간 히스토그램 기법의 검색율과 관계가 있는데 공간 히스토그램의 검색율이 높으면 전체영상에 대한 후보영상의 비율을 크게하고 반대로 작으면 비율을 적게할수록 융합법에 의한 검색율이 증가함을 확인하였다.

표 2. 검색방법에 따른 검색율 비교

	DB	공간 히스토그램	웨이브릿 모멘트 (1)	웨이브릿 모멘트 (2)	제한한 방법	후보 영상 갯수
			(L <sub>1</sub> )	(L <sub>1</sub> )		
A	Brodatz	73.6% (L <sub>1</sub> )	65.0% (L <sub>1</sub> )	72.2% (L <sub>1</sub> )	80.9%	200
		79.9% (L <sub>05</sub> )			85.2%	200
V	MPEG-7 T1	97.7% (L <sub>1</sub> )	67.7% (L <sub>1</sub> )	74.8% (L <sub>1</sub> )	92.5%	200
		98.6% (L <sub>05</sub> )			92.7%	200
R	Corel Draw	56.1% (L <sub>05</sub> )	73.3% (L <sub>1</sub> )	74.9% (L <sub>1</sub> )	77.2%	9

공간 히스토그램 기법으로 Brodatz 데이터에 적용한 검색 결과는 73.6%인데 반해, MPEG-7의 T1데이터에 적용한 검색 결과는 97.7%로 상당히 높게 나왔다. 이는 그림 4의 히스토그램을 보면 쉽게 설명될 수 있다. 그림 4 (a)T001-1, (b)T001-2, (c)T001-6는 같은 원영상(512 × 512)으로부터 나온 16개의 동질한 영상 데이터(128 × 128)중 1, 2, 6번째 데이터의 히스토그램이다. 이들 히스토그램의 분포에서는 그레이 값이 거의 같은 위치에 있지만 그림 4의 (d)T001-1, (e)T003-4, (f)T007-11는 질감의 형태가 다른 부류에 속한 영상 데이터로 영상을 이루는 그레이 값이 영상마다 다르다는 것을 알 수 있다. 그러므로 이종간의 유사도는 더욱 크게 벌어진다. 이와 같은 원인에 의해 공간 히스토그램에 의한 검색 효율이 다른 방법에 비해 탁월하게 나왔다고 판단된다.

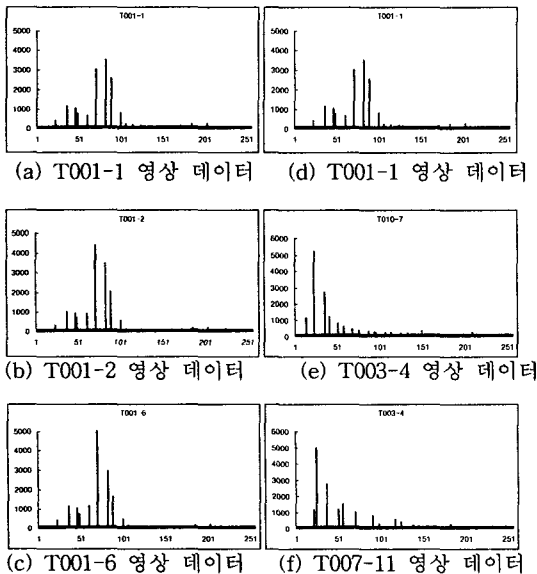


그림 5. T1영상 데이터의 히스토그램 분포 특성

그림 5는 (d)식의 T<sub>r</sub> 값을 증가시키에 따른 검색율

의 추이를 나타낸 그림이다. 웨이브릿 정규 모멘트를 이용한 방법이 가장 낮으며 그 다음은 공간영역의 히스토그램을 이용한 방법이고 그리고 제안한 방법의 검색율이 가장 높음을 알 수 있다. T<sub>r</sub>값이 60이상이면 검색율은 대체로 증가하지 않음을 볼 수 있다.

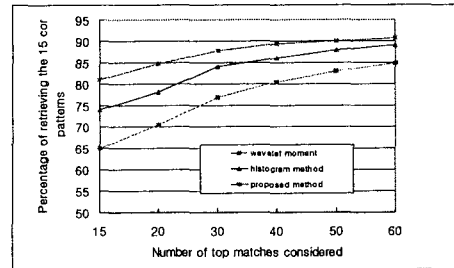


그림 6. 상위 매칭에 따른 검색 성능

### V. 결론

본 논문에서는 웨이브릿 모멘트 특징과 히스토그램 특징의 융합에 의하여 검색율을 향상시키는 방법을 제안하였다. 히스토그램 특징의 단점을 웨이브릿 모멘트 특징을 사용하여 보완하므로 성능 향상을 가져올 수 있었다. 실험은 Brodatz, MPEG-7 T1 질감 영상과 Corel Draw photo영상에서 수행하였으며, 웨이브릿 모멘트나 히스토그램만으로 검색한 결과보다 검색율이 향상됨을 확인하였다. 그러나 공간영역 히스토그램에 의한 검색 방법은 처리속도면에서는 빠르지만 영상의 특징 벡터의 길이가 다른 특징들에 비해 상당히 큰 편이므로 대량의 DB 구축시 많은 저장공간을 차지하게 된다. 그러므로 히스토그램의 차원을 줄여도 검색 효율이 저하되지 않는 연구가 향후 이루어져야 할 것으로 생각된다.

### 참고문헌

- [1] J. R. Smith and S.-F. Chang, Transform Features for texture classification and discrimination in large image databases, in *Proc. IEEE Int. Conf. on Image Proc.*, 1994.
- [2] B. s. Manjunath and W. Y. Ma, "Texture Features for Browsing and Retrieval of Image Data", *IEEE Trans on PAMI*, Vol 18, No. 8, pp. 837-841, Aug 1996
- [3] W. Y. Ma and B. S. Manjunath, "A comparison of wavelet transform feature for texture image annotation," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 1997.
- [4] M. K. Mandal, T. Aboulnasr, "Image Indexing Using Moments and Wavelets," *IEEE Trans on Consumer Electronics*, vol. 42, no. 3, pp. 557-564, Aug 1996.
- [5] I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis," *IEEE Trans. Information Theory*, vol. 36, pp. 961-1005, Sept. 1990.
- [6] P. Brodatz, Textures: a Photographic Album for Artists and Designers, Dover, New York, 1965. <http://www.ux.his.no/~tranden/brodatz.html>
- [7] "Core Experiments on MPEG-7 Color and Texture Descriptor," *ISO/IEC JTC1/SC29/WG11 N2691* 1999.