

다중 특징값의 조합을 이용한 자동적 이미지 카테고리화 방법

양승지, 윤정현, 노용만
한국정보통신대학교, 통신공학부, 멀티미디어그룹
e-mail : {yangzeno, sine, yro}@icu.ac.kr

Automatic Image Categorization using Combination of Multiple Features

Seungji Yang, Jeonghyun Yoon, Yong Man Ro
Multimedia Group, Information and Communications University

요 약

본 논문에서는 내용 기반 이미지 검색 및 필터링 시스템을 위한 카테고리 식별 방법을 제안한다. 제안된 방법에서는 식별 가능한 카테고리를 사전에 정의하고, 정의된 카테고리를 대표할 수 있는 이미지들을 수집한다. 다음으로, 이들로부터 다중의 내용 기반 특징값을 추출하고, 추출된 특징값들로 카테고리 데이터베이스를 구성한다. 카테고리를 식별할 질의 이미지가 입력으로 들어오면, 질의 이미지로부터 추출된 다중 특징값들을 각 카테고리의 단일 특징값과 각각 비교함으로써, 카테고리를 대표하는 다중의 유사도 거리값을 측정한다. 각 카테고리를 대표하는 다중의 유사도 거리값들은 두 가지 연산 방법에 의해 조합되는데, 조합 방법은 각각의 단일 특징값이 각 카테고리 식별에 미치는 영향을 고려하여 정의된다. 최종적으로, 각 카테고리의 조합된 유사도 거리값을 비교한 다음, 가장 유사도가 큰 카테고리를 해당 질의 이미지의 카테고리로 식별한다.

1. 서론

최근 인터넷과 개인용 통신 수단의 발전 및 보급으로 디지털 멀티미디어 콘텐츠의 제작 및 이용이 급속하게 증가하고 있다. 특히 디지털 지상파 및 위성 방송, 인터넷 방송, 디지털 도서관, 그리고 디지털 출판물에 이르기까지 디지털 콘텐츠의 이용 범위가 점점 더 다양해져서 사용자들은 하루에도 수 많은 디지털 멀티미디어 콘텐츠를 접하면서 생활하고 있다. 이제 콘텐츠의 디지털화 차원을 넘어서서, 무수한 디지털 콘텐츠들 가운데 사용자가 원하는 콘텐츠가 어디에 있는지, 그리고 그것을 얼마나 빠르고 효율적으로 찾을 수 있는지의 여부가 핵심 사항이 되고 있다. 이러한 요구 사항들은 콘텐츠의 적절한 카테고리화를 통하여 보다 효율적으로 수행될 수 있다.

콘텐츠의 카테고리화는 사용자가 원하는 콘텐츠를 검색하기 위해 검색 초기 과정에서 검색의 범위를 줄여줌으로써, 검색 수행 속도 뿐만 아니라 검색의 정확도를 향상시킬 수 있도록 한다. 뿐만 아니라, 사용자

가 원하지 않는 콘텐츠를 차단하는 콘텐츠 필터링 시스템을 위해서도 매우 유용한 방법이다.

기존의 카테고리화 방법은 대부분 텍스트에 기반한 방법으로서, 사람이 일일이 텍스트로 기술한 메타 데이터를 이용하여 콘텐츠를 카테고리화 하였다. 그러나, 이 방법은 콘텐츠의 양이 급속하게 증가하고 있기 때문에 존재하는 모든 콘텐츠의 카테고리 정보를 사람이 일일이 기술한다는 것이 거의 불가능하다는 점과, 텍스트 정보가 멀티미디어 콘텐츠를 기술하는데 효과적이지 못하다는 점에서 유용하지 못하다.

그러므로, 콘텐츠의 내용에 기반하여 추출된 특징값을 이용하여 멀티미디어 콘텐츠를 카테고리화 하는 방법이 제안될 수 있다 [1][2]. 본 논문에서는, 이미지 콘텐츠를 카테고리화 하기 위한 방법으로 MPEG-7 기술자를 이용하여 다중의 내용 기반 특징값을 추출하고, 이들을 이용하여 카테고리 대표 유사도 거리값을 얻는다. 얻어진 카테고리 대표 유사도 거리값들은 제안된 두 가지 연산 방법에 의해 조합되는데, 이들의 조합 방법은 각각의 단일 특징값이 각 카테고리 식별

에 미치는 영향에 따라 두 가지로 분류된다.

2. 다중 특징값 추출 및 유사도값 측정

단일한 시각적 특징 정보는 이미지를 카테고리화 하기 위한 충분한 구별 능력을 제공하기 어렵다. 따라서, 다중의 특징 정보를 카테고리를 식별 기준으로 사용하는 것이 보다 효과적이다. 본 논문에서는 이미지로부터 얻을 수 있는 대표적인 특징 정보인 질감, 색, 그리고 모양 특징 정보를 세 가지 MPEG-7 기술자를 이용하여 추출하였다 [3][4][5].

2.1 Homogeneous Texture Descriptor

이미지의 질감 정보는 이미지의 균질 또는 비균질의 패턴을 말하는 것이고, 인간의 시각적 인지의 기본적으로 중요한 특징이다. 이미지의 질감 특징 정보 표현을 위한 본 기술자는 주파수 영역을 6개의 방향 성분과 5개의 크기 성분으로 나누어, 모두 30개의 채널을 구성하고 각 채널에 대한 질감 특징을 추출한다. 마지막으로, 각 채널 에너지의 평균값과 표준편차값을 추출하여 62개의 특징 벡터 값을 추출한다. 본 기술자의 특징 벡터 \overline{TD} 의 식은 아래와 같다.

$$\overline{TD} = \{f_{DC}, f_{SD}, e_1, e_2, \dots, e_{30}, d_1, d_2, \dots, d_{30}\}, \quad (1)$$

여기서, f_{DC} 와 f_{SD} 는 전체 이미지의 평균값과 표준편차값을 나타내며, e 와 d 는 각 채널 에너지의 평균값과 표준편차값을 나타낸다. 카테고리 식별을 위하여, 질의 이미지의 특징 벡터 \overline{TD}_{query} 와 카테고리 데이터베이스 내의 이미지들로부터 특징 벡터 $\overline{TD}_{category}$ 를 추출한다. 추출된 두개의 특징 벡터들 간의 유사도 거리 측정은 다음과 같은 방법에 의해 수행된다.

$$d_{TD} = distance(\overline{TD}_{query}, \overline{TD}_{category}) = \sum_k \left| \frac{\overline{TD}_{query}(k) - \overline{TD}_{category}(k)}{a(k)} \right|, \quad (2)$$

2.2 Scalable Color Descriptor

이미지의 칼라 정보 또한 이미지 카테고리 식별을 위한 중요한 특징 값으로 사용된다. 이것은 이미지 내의 칼라의 분포를 나타내는 히스토그램으로 표현된다. 여기서, 이미지의 RGB 값들은 HSV 값으로 비선형 변환된다. 그리고, HVS 칼라 공간을 모두 256개의 빈(bin)으로 나누고, 각 빈에 속하는 픽셀의 수를 측정함으로써, 특징 벡터를 구성한다. 카테고리 식별을 위하여, 질의 이미지의 특징 벡터 \overline{CD}_{query} 와 카테고리 데이터베이스 내의 이미지들로부터 특징 벡터 $\overline{CD}_{category}$ 를 추출한다. 추출된 두개의 특징 벡터들 간의 유사도 거리 측정은 다음과 같은 방법에 의해 수행된다.

$$d_{CD} = distance(\overline{CD}_{query}, \overline{CD}_{category}) = \sum_k \left| \overline{CD}_{query}(k) - \overline{CD}_{category}(k) \right|, \quad (3)$$

2.3 Edge Histogram Descriptor

마지막으로, 이미지로부터 에지 정보를 추출하기 위해 본 기술자를 사용한다. 먼저 이미지를 16개의 하위 블록으로 나누고, 각 블록에 대해서 모두 5개의 에지 성분, 수직(vertical), 수평(horizontal), 45°, 135°, 그리고 무방향성(non-directional)의 에지 성분을 구한다. 결과적으로, 각 블록의 에지 성분들을 조합하여 한 이미지로부터 모두 80개의 local edge histogram을 얻을 수 있으며, 이미지의 16개 하위 블록의 local edge histogram을 조합하여, 각각 5개의 global edge histogram과 40개의 semi-global edge histogram을 구성한다. 그러므로, 각 에지 성분으로부터 모두 150개의 에지 히스토그램 빈을 구성하고 이를 특징값으로 정의한다. 에지 기술자의 특징 벡터 \overline{ED} 의 식은 아래와 같다.

$$\overline{ED} = \left\{ \begin{array}{l} f_{local_1}, \dots, f_{local_80}, \\ f_{semi-global_1}, \dots, f_{semi-global_40}, \\ f_{global_1}, \dots, f_{global_5} \end{array} \right\}, \quad (4)$$

여기서, f_{local_i} 는 i 번째 local edge histogram 빈을 나타내고, $f_{semi-global_j}$ 는 j 번째 semi-global edge histogram 빈을, f_{global_l} 은 l 번째 global Edge Histogram 빈을 각각 나타낸다.

카테고리 식별을 위하여, 질의 이미지의 특징 벡터 \overline{ED}_{query} 와 카테고리 데이터베이스 내의 이미지들로부터 특징 벡터 $\overline{ED}_{category}$ 를 추출한다. 추출된 두개의 특징 벡터들 간의 유사도 거리 측정은 다음과 같은 방법에 의해 수행된다.

$$d_{ED} = distance(\overline{ED}_{query}, \overline{ED}_{category}) = \sum_k \left| \overline{ED}_{query}(k) - \overline{ED}_{category}(k) \right|, \quad (5)$$

3. 다중 특징값의 조합을 통한 카테고리 식별

3.1 카테고리 데이터베이스 정의

먼저, 식별할 카테고리를 사전에 정의하고, 각 카테고리에 속하는 이미지들을 선별한다. 이때, 이미지들은 카테고리가 가질 수 있는 대표 특성을 가능한 많이 포함할 수 있도록 선택되어야 한다. 다중 특징값을 추출하고 데이터베이스에 저장한다.

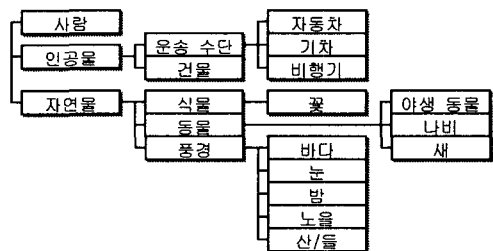


그림 1. 카테고리 데이터베이스의 구조

본 논문에서 카테고리 데이터베이스는 모두 14개의 카테고리로 구성된다. 그림 1은 데이터베이스에 정의한 카테고리의 종류를 나타낸다. 입력된 질의 이미지는 14개의 카테고리 가운데 하나의 카테고리로 식별된다. 그림 1에서 짙은 영역의 카테고리가 최종 식별되는 카테고리를 표현한다.

3.2 카테고리 유사도 측정

표 1은 다중 특징값을 이용하여 추출된 유사도 거리값을 각 카테고리별로 정리한 것이다. 각 카테고리마다 M 개의 이미지가 존재하며, 각 이미지마다 세 가지 기술자에 대한 유사도 거리값이 측정된다.

표 1. 각 카테고리 이미지들의 유사도

이미지	Homogeneous Texture	Edge Histogram	Scalable Color
C_1	$CI(1, 1)$	$d_{TP}(1, 1)$	$d_{EP}(1, 1)$
	$CI(1, M)$	$d_{TP}(1, M)$	$d_{EP}(1, M)$
C_2	$CI(2, 1)$	$d_{TP}(2, 1)$	$d_{EP}(2, 1)$
	$CI(2, M)$	$d_{TP}(2, M)$	$d_{EP}(2, M)$
C_n	$CI(N, 1)$	$d_{TP}(N, 1)$	$d_{EP}(N, 1)$
	$CI(N, M)$	$d_{TP}(N, M)$	$d_{EP}(N, M)$

C_n : 카테고리, $1 \leq n \leq N$
 $CI(n, m)$: n 번째 카테고리 내의 m 번째 이미지, $1 \leq m \leq M$
 $d_{TP}(n, m), d_{EP}(n, m), d_{CP}(n, m)$: $CI(n, m)$ 이미지의 유사도 거리값

이렇게 측정된 각각의 유사도 거리값은 질의 이미지와 각각의 이미지와의 유사도를 의미한다. 그러므로, 측정된 유사도 거리값이 해당 카테고리 특성을 대표한다고 말하기 어렵다. 본 논문에서는 주어진 유사도 거리값을 이용하여 카테고리 특성을 보다 잘 표현하기 위해서 이들을 조합한다.

본 논문에서는 일반적으로 내용 기반 검색에서 사용하는 다중 특징값 조합 방법과는 다르게, 다중 유사도 거리값의 조합을 단일 이미지에서 수행하지 않고, 각각의 단일 특징값을 기반으로 카테고리를 대표하는 이미지들을 찾고, 이들의 유사도 거리값을 조합하는 방법을 사용한다.

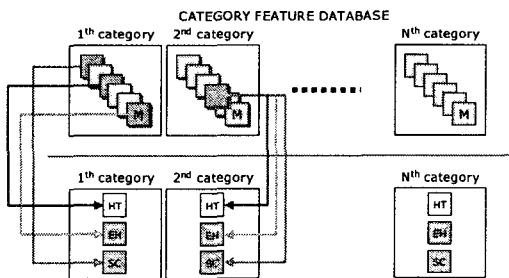


그림 2. 카테고리 대표 유사도 거리값 선택

그림 2에서와 같이, 해당 카테고리를 대표하는 서로 다른 이미지가 선택될 수 있다. 선택된 이미지들은

모두 같은 이미지일 수도 있고, 사용한 기술자의 종류에 따라 서로 다른 이미지일 수도 있다. 이 방법은 카테고리의 식별 능력이 떨어지는 경우, 해당 카테고리 데이터베이스 내의 이미지를 보다 많이 할당하여 식별 능력을 향상시키더라도, 다른 카테고리의 식별 수행 시간에 영향을 미치지 않기 때문에, 시스템의 수행 속도에 크게 영향을 주지 않는 장점이 있다. 최종적으로, 얻어진 카테고리 대표 유사도 거리값은 각 기술자마다 식 (6)과 같이 표현된다.

$$D(n) = \min\{d(n, m), 1 \leq m \leq M\}, \quad (6)$$

만일, 각각의 특징값이 서로 다른 카테고리에서 상대적으로 우수한 식별 능력을 갖는다면, 다중 특징값을 조합하는 방법은 매우 중요하다. 예를 들어, 정의된 카테고리 가운데, 색 특징 정보에 의해 식별이 잘 되는 카테고리도 있고, 질감이나 에지 특징 정보에 의해 식별이 잘 되는 카테고리도 존재할 수 있다. 다중 특징값 추출 방법이 이미 정해져 있는 상황에서는 이들을 얼마나 효율적으로 조합하느냐에 따라 전체적인 카테고리 식별 능력에 영향을 미친다. 그러므로, 가능한 한 단일 특징값의 카테고리 식별 능력을 최대화하는 동시에, 다중 특징값 조합으로 발생하는 공동상승 효과(synergy effect)를 얻을 수 있어야 한다.

본 논문에서는 단일 특징값이 가지는 카테고리 식별 능력을 고려한 두 가지 연산, OP_{AND} 와 OP_{OR} 를 식 (7)에서와 같이 정의한다.

$$\begin{aligned} OP_{AND}\{D_1(n), D_2(n)\} &= D_1(n) \times D_2(n), \\ OP_{OR}\{D_1(n), D_2(n)\} &= D_1(n) + D_2(n) \end{aligned} \quad (7)$$

OP_{AND} 는 임의의 두 특징값이 서로 다른 카테고리에서 우수한 식별 능력을 보일 때, 각각의 단일 특징값이 가지는 카테고리 식별 능력을 보존하기 위해 두 특징값을 곱한다. 따라서, OP_{AND} 의 결과는 두 특징값을 가운데, 작은 유사도 거리값에 영향을 더 많이 받는다. 이에 반하여, 임의의 두 특징값의 식별 능력이 전체 카테고리에 대하여 불균직한 성질을 보인다면, OP_{OR} 를 사용하여 두 특징값의 더한다.

본 논문에서 사용한 색, 질감, 모양 특징 정보를 가운데, 색과 모양 특징 정보는 서로 다른 카테고리에서 우수한 식별 능력을 나타내는데 반하여, 질감 특징 정보는 색이나 모양 특징 정보 가운데 특별한 관련을 찾을 수 없었다. 따라서, 본 논문에서는 식 (8)에서와 같이 세 가지 기술자의 유사도 거리값을 조합하였다.

$$D_{combined}(n) = OP_{AND}\{OP_{OR}\{D_{CD}(n), D_{ED}(n)\}, D_{TD}(n)\}. \quad (10)$$

그러므로, 질의 이미지의 최종 카테고리는 각 카테고리에서 얻어진 대표 유사도 거리값들 가운데 최소값을 가지는 카테고리로 결정된다.

4. 실험 결과

제안된 방법을 증명하기 위해, 각 카테고리마다 100

개의 이미지의 특징값을 추출하여 카테고리 데이터베이스를 구성하였다 [6]. 질의 이미지는 인터넷으로부터 임의로 수집한 이미지를, 각 카테고리마다 100개의 이미지로 분류하여 실험하였다. 본 실험에서 질의 이미지들은 모두 하나의 실제 카테고리를 갖는다고 가정한다. 표 2와 표 3은 각 카테고리마다 100개의 질의 이미지들에 대한 카테고리 식별 결과를 보여준다. 표 2는 단일 특징값만 사용한 경우의 카테고리 식별 결과이다. 카테고리 식별 비율의 평균값이 70% 정도이다. 단일 특징값 사용만으로는 만족할만한 카테고리 식별 결과를 얻을 수 없음을 알 수 있다.

표 2. 단일 특징값을 사용한 카테고리 식별 결과

질의 이미지의 실제 카테고리	Homogeneous Texture	Scalable Color	Edge Histogram
(1) 사람	45 / 100	75 / 100	60 / 100
(2) 자동차	82 / 100	83 / 100	89 / 100
(3) 기차	79 / 100	81 / 100	88 / 100
(4) 비행기	69 / 100	60 / 100	80 / 100
(5) 건물	70 / 100	68 / 100	62 / 100
(6) 꽃	61 / 100	81 / 100	81 / 100
(7) 야생 동물	45 / 100	80 / 100	69 / 100
(8) 나비	60 / 100	81 / 100	63 / 100
(9) 새	49 / 100	52 / 100	53 / 100
(10) 바다	62 / 100	88 / 100	66 / 100
(11) 눈 (snow)	64 / 100	81 / 100	51 / 100
(12) 노을	77 / 100	90 / 100	82 / 100
(13) 밤 (night)	70 / 100	89 / 100	77 / 100
(14) 산/들	56 / 100	81 / 100	68 / 100
카테고리 식별 비율 평균	889 / 1400 = 63.50%	1090 / 1400 = 77.86%	989 / 1400 = 70.64%

*올바르게 카테고리를 식별한 이미지 수 / 전체 질의 이미지 수

이에 반하여 표 3은 제안된 방법의 카테고리 식별 결과를 보여준다. 뿐만 아니라, 각 특징 정보의 카테고리 식별 능력을 고려하지 않고, OP_{AND} 와 OP_{OR} 연산만을 수행하여 유사도 거리값을 조합한 결과와 비교하였다. 제안된 방법이 카테고리 식별 능력을 크게 향상시켰음을 결과에서 확인할 수 있다.

표 3. 다중 특징값을 사용한 카테고리 식별 결과

질의 이미지의 실제 카테고리	OP_{OR} 만 사용한 경우	OP_{AND} 만 사용한 경우	제안된 방법
(1) 사람	86 / 100	87 / 100	90 / 100
(2) 자동차	89 / 100	91 / 100	98 / 100
(3) 기차	91 / 100	95 / 100	95 / 100
(4) 비행기	77 / 100	72 / 100	88 / 100
(5) 건물	71 / 100	70 / 100	93 / 100
(6) 꽃	89 / 100	90 / 100	91 / 100
(7) 야생 동물	68 / 100	74 / 100	94 / 100
(8) 나비	88 / 100	88 / 100	89 / 100
(9) 새	59 / 100	55 / 100	81 / 100
(10) 바다	85 / 100	93 / 100	91 / 100
(11) 눈 (snow)	79 / 100	78 / 100	92 / 100
(12) 노을	95 / 100	98 / 100	98 / 100
(13) 밤 (night)	95 / 100	95 / 100	96 / 100
(14) 산/들	79 / 100	86 / 100	90 / 100
카테고리 식별 비율 평균	1151 / 1400 = 82.21%	1172 / 1400 = 83.71%	1286 / 1400 = 91.86%

그림 3은 위의 카테고리 식별 결과를 도식적으로 비교한 것이다. 제안된 방법이 각 단일 특징값의 낮은 카테고리 식별 능력을 보완하고 높은 카테고리 식별 능력을 유지하는 결과를 보여준다.

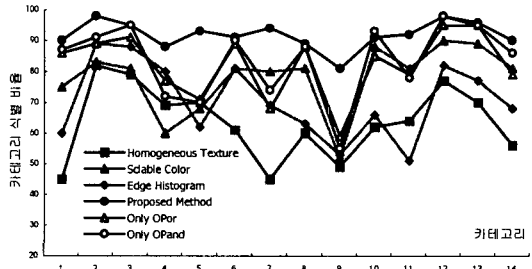


그림 3. 각 방법의 카테고리 식별 결과 비교

5. 결론

본 논문에서는 내용 기반 이미지 검색 또는 필터링 시스템을 위한 카테고리 식별 방법을 제안한다. 실험 결과는 제안된 다중 특징값의 조합 방법이 카테고리 식별에 매우 효과적임을 확인하였다. 특징값 추출 방법으로 MPEG-7 기술자를 사용하였기 때문에, 다른 데이터베이스와의 호환성을 높일 수 있다. 앞으로는 카테고리를 보다 늘리고, 그들의 구조를 보다 체계화할 필요가 있으며, 단일 특징값 간의 카테고리 식별 패턴의 관계를 통계적으로 증명하는 실험이 필요하다. 그리고, 이미지 뿐만 아니라, 비디오 콘텐츠에도 적용할 수 있도록 알고리즘을 보완해야 할 것이다.

참고문헌

- [1] Y. J. Jung, J. H. Yoon, S. J. Yang, H. K. Lee, Y. M. Ro, "Content-based Image Filtering Technique for Information Filtering Agent", *Real-Time Imaging VI on Proc. SPIE*, vol. 4666, pp.10-17, Jan. 2002.
- [2] S. Newsam, B. Sumengen, B. S. Manjunath, "Category-based Image Retrieval", *IEEE International Conference on Image Processing*, vol.2, pp.596-599, 2001.
- [3] MPEG-7 Visual part of eXperimentation Model Version 10.1, ISO/MPEG, m7357, Sydney, July 2001
- [4] F. Idris, S. Panchanathan, "Review of Image and Video indexing Techniques", *Journal of Visual Communication and Image Representation*, Vol. 8, No. 2, pp. 146-166, June 1997
- [5] H. K. Kang, Y. J. Jung, H. K. Lee, Y. M. Ro, "Multimedia database system embedding MPEG-7 meta data", *SPIE Electronics Imaging*, Jan. 2000.
- [6] Image database: <http://www3.corel.com>