객체 특징점 모델링을 이용한 시멘틱 단서 기반 영상 분류

(Semantic Cue based Image Classification using Object Salient Point Modeling)

박 상 혁 † 변 혜 란 ††
(Sanghyuk Park) (Hyeran Byun)

요 약 대부분의 영상들은 여러 객체 영역들의 시각적인 특징과 각각의 의미들의 조합으로 구성되어 있다. 그러나 일반적으로 영상 처리를 위한 컴퓨터 시스템들은 영상을 특정 객체 영역의 의미 정보 단위로 해석하지 못하기 때문에 사람이 영상을 인지하는 것과 의미적인 차이(semantic gap)가 발생한다. 본 논문에서는 이러한 문제점을 극복하기 위하여 각 객체 영역 단위에서 추출한 고유한 특징점들을 고차원의 의미 정보로 모델링하여 영상을 분류하는 방법을 제안한다. 제안하는 방법은 객체 단위로 추출된 고유한 특징점들의 의미 정보를 특정 객체 영역을 인식하기 위한 의미 단서로 이용한다. 이를 통하여 기존의 영상 분류 방법들에 비하여 인간의 인지 능력과 유사하고 보다 효율적으로 영상을 분류할 수 있는 장점이 있다. 실험 결과는 다양한 카테고리 종류의 영상에 대하여 제안하는 방법의 효과적인 분류 성능을 보여준다.

키워드 : 영상 분류, 의미 특징점 모델링, 객체 기반 영상 분류

Abstract Most images are composed as union of the various objects which can describe meaning respectively. Unlike human perception, The general computer systems used for image processing analyze images based on low level features like color, texture and shape. The semantic gap between low level image features and the richness of user semantic knowledges can bring about unsatisfactory classification results from user expectation. In order to deal with this problem, we propose a semantic cue based image classification method using salient points from object of interest. Salient points are used to extract low level features from images and to link high level semantic concepts, and they represent distinct semantic information. The proposed algorithm can reduce semantic gap using salient points modeling which are used for image classification like human perception. and also it can improve classification accuracy of natural images according to their semantic concept relative to certain object information by using salient points. The experimental result shows both a high efficiency of the proposed methods and a good performance.

Key words : Image Classification, Salient Point Modeling, Object-based Image Classification

1. 서 론

오늘날 정보통신기술의 발전과 다양한 멀티미디어 장치의 보급과 더불어 폭발적으로 증가하는 영상 데이터들을 보다 효율적으로 관리하기 위한 연구들이 지속적으로 이루어지고 있다. 특히 영상데이터를 효율적으로 분류하기 위한 연구들은 과거와 같은 사용자의 수작업에 의한 영상 분류 과정 등에서 발생하는 시간과 비용의 비효율성 문제를 극복하기 위한 방법으로 주목받고 있다. 이러한 연구들은 유사한 속성을 가지는 영상들을 일정한 카테고리 집합으로 자동 분류하는데 목적이 있다. 일반적으로 사용자들은 영상 데이터를 구분하기 위하여 영상의 전체적인 시각적 속성(색상, 형태, 질감 등)보다는 영상에서 보다 큰 의미를 차지하는 특정한 객체나 영역에 관심을 가진다. 그러나 컴퓨터 시스템은 영상을 화소 단위의 특징으로 분석하기 때문에 인간이 오랜 기간 동안 다양한 경험과 지식에 의하여 객체를 인지하고 영상을 해석하는 방법과 의미적인 차이가 발생하게 된다.

최근의 영상 분류를 위한 연구들은 영상의 전역적인 시각적 속성에 기반을 둔 저차원의 특징 정보만을 사용하던 기존의 방법들의 단점을 극복하기 위해 영상에서 보다 중요한 특징을 가지는 객체 영역을 고려하여 분류하는 연구가 많이 진행되고 있다. 영상을 일정한 영역 단위로 분할하고 특정 영역에 문맥 정보를 부여하여 이들의 부분 영역 정보를 이용한 분류 방법[1-3]등과 영상이 다수의 객체 범주로 해석될 수 있는 복합 구조임을 고려하여 다양한 정보들을 이용하여 조건 확률과 같

· 이 논문은 2009년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No. 2009-0067625)
· 이 논문은 2009 한국컴퓨터종합학술대회에서 '객체 특징점 모델링을 이용한 시멘틱 단서 기반 영상 분류 알고리즘'의 제목으로 발표된 논문을 확장한 것임

† 학생회원 : 연세대학교 컴퓨터과학과
korealovewe@naver.com
†† 종신회원 : 연세대학교 컴퓨터과학과 교수
hrbyun@yonsei.ac.kr

논문접수 : 2009년 8월 13일
심사완료 : 2009년 10월 18일

정보과학회논문지: 컴퓨팅의 실제 및 레터 제16권 제1호(2010.1)

은 식별 모델을 생성하여 영상을 구성하는 각 부분 객체 영역에 대한 자동 분류가 가능한 방법[4,5]등이 있다. 그러나 이러한 방법들은 분할된 영역 또는 객체들에 특정한 의미 정보를 부여하기 위해 복잡한 사전 학습 알고리즘이 필요한 단점이 있고, 다수의 객체 영역이 서로 겹쳐서 존재하는 경우 복잡한 배경 등으로 인하여 중요 객체 영역의 판단에 어려움이 발생한다.

본 논문에서는 영상을 구성하는 여러 객체들에서 고유한 지역 특징점들을 추출하고 각각의 특징점마다 부여된 의미적인 개념을 이용하여 해당 객체를 해석함으로써 다양한 종류의 영상 데이터들을 인간의 인지 능력과 유사하게 해석하여 분류하는 방법을 제안하고자 한다. 제안하는 방법은 우선 영상으로부터 특정한 의미 정보를 가지는 객체 영역들을 추출하기 위하여 영상 분할 알고리즘을 이용하여 영상을 다수의 객체 영역 단위로 분할한다. 그리고 가우시안 근사화 필터를 이용하여 각 객체 영역마다 고유한 지역 특징점들을 빠르게 추출한다. 이때 추출된 특징점 주변의 색상과 질감 정보를 이용하여 각각의 특징점마다 해당 객체가 가지는 의미적 개념에 해당하는 특정 키워드를 부여하여 모델링한다. 그리고 영상을 구성하는 각 객체 영역들은 영상 분할을 통한 객체 영역의 크기 가중치 값과 각 객체 내의 특징점들로부터 K-Nearest Neighbors 분류 과정을 통하여 가장 높은 분포를 가지는 의미 정보와의 조합을 통하여 분류된다.

2. 객체 특징점 모델링

2.1 영상 분할

대부분의 영상은 다양한 의미로 해석이 가능한 여러 객체 영역들의 조합으로 구성되어 있다. 이때 객체 영역이란 하늘, 구름, 건물 등과 같은 의미 정보에 해당하는 특정한 키워드 등으로 구체화 할 수 있고, 영상의 내용을 대표하여 표현할 수 있는 핵심적인 영역이다. 따라서 영상을 특정한 객체 영역 단위에 기반하여 해석하기 위해서는 우선 영상을 구성하는 각 객체 영역을 효율적으로 분할하는 알고리즘이 요구된다. 본 논문에서는 효율적인 영상 분할을 위하여 Deng과 Manjunath가 제안한 JSEG[6]알고리즘을 이용하였다. JSEG알고리즘은 특정한 모델을 추정하는 기존의 분할 방법들과 달리 영상의 공간적 컬러 분포에 기반한 질감과 동질성을 검사하여 영상을 효과적으로 자동 분할하는 방법이다. JSEG알고리즘의 전체적인 흐름도는 그림 1과 같다.

2.2 가우시안 근사화 필터를 이용한 특징점 추출

영상을 구성하는 각각의 특정 객체 영역들에서 고유한 특징점들을 추출하기 위하여 Lowe에 의하여 제안된 SIFT알고리즘[7]의 이론적인 개념을 이용하였다. SIFT

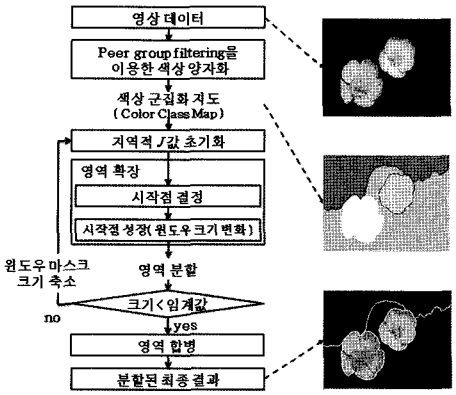

그림 1 JSEG 알고리즘의 흐름도

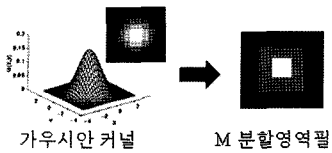

그림 2 제안하는 근사화 가우시안 필터

알고리즘의 키포인트 탐색기는 스케일 공간상에서 극점을 검출하는데 이러한 극점은 영상의 크기 변환이나 영상 변환에 강인하며, 노이즈 및 조명의 변화에도 강인한 특성이 있다. 이러한 특성에 의하여 추출된 극점들을 영상을 구성하는 특정 객체들마다의 고유한 특징점으로 추출하여 이용하였다. 본 논문에서는 SIFT 알고리즘의 가우시안 커널의 속도문제를 개선하기 위하여 가우시안 근사화 필터를 이용한 특징점 추출 방법을 제안한다. 이는 그림 2와 같이 오리지널 가우시널 필터 대신에 M분할영역으로 구성된 가우시안 근사화 필터를 생성함으로써 기존의 가우시안 커널보다 빠른 속도로 특징점 추출이 가능하게 한다.

또한 스케일 변화에 강인한 특징 점들을 추출하기 위하여 Fast approximated SIFT[8]은 Viola *et al.*[9]에서 소개된 적분 영상의 조합을 이용하여 극대와 극소에 해당하는 특징점을 간편하고 빠르게 추출할 수 있다. 가우시안 근사화 필터와 적분 영상 방법을 조합한 결과는 영상의 부분 영역 $I'(S_i)$에 대하여 식 (1)과 같이 표현할 수 있고, 이때 $I'(S_i)$의 가우시안 근사화 필터의 가중치는 식 (2)와 같다. 결국 가우시안 근사화 필터 방법으로 생성한 스케일 공간상에서의 영상은 식 (3)으로 표현할 수 있다.

$$I'(S_i) = \frac{1}{N_i \times N_i} \sum_{(x,y)\in S_i} I(x,y) \quad (1)$$

$$w_i(S_i) = \frac{1}{N_i \times N_i} \sum_{(x,y)\in S_i} G(x,y,\sigma) \quad (2)$$

$$L'(x,y,\sigma)=\sum_{i=1}^{M} w_i(\sigma)I'(s_i) \quad (3)$$

제안하는 가우시안 근사화 필터 $G'(x,y,\sigma)$와 오리지널 가우시안 필터 방법의 에러율을 계산하기 위한 목적함수 $\varepsilon(x,y,\sigma)$는 식 (4)를 이용하였고, 분할 영역 M의 크기와 스케일 펙터 σ에 따른 에러량 변화는 그림 3과 같다.

$$\varepsilon(x,y,\sigma)=\|G(x,y,\sigma)*I(x,y)-G'(x,y,\sigma)*I(x,y)\|^2 \quad (4)$$

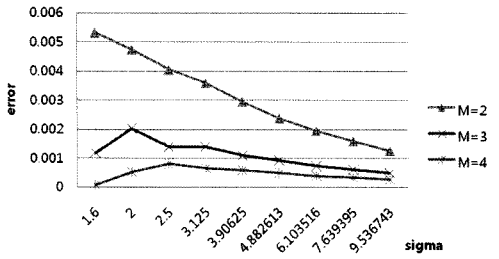

그림 3 M 크기와 스케일 펙터σ에 의한 에러량 변화

그림 3의 결과로부터 스케일 펙터 σ가 증가하거나, 분할 영역 M의 개수가 증가할수록 근사화 필터에 따른 에러량이 감소함을 알 수 있다. 본 논문에서는 실험에 의하여 M=4, 스케일펙터 σ(1.6,2,2.5)를 이용하였다.

2.3 객체 특징점 모델링

• 특징 정보 추출

기존의 SIFT와 같은 지역 특징점 기술자에서는 특징점을 중심으로 방향 성분에 해당하는 128차원의 특징을 생성하였으나, 본 논문에서는 각각의 특징점을 중심으로 주변의 N×N 윈도우 영역에 해당하는 질감 정보와 색상 정보를 추출하여 사전에 정의된 객체의 카테고리 정보로 인덱싱한 후 모델을 학습하였다.

영상에서 질감 정보를 추출하기 위해 사용한 소벨 필터는 윤곽선을 검출하기 위한 필터로 편미분 연산자 대신에 식 (5)와 같은 3×3 필터 마스크를 이용함으로써 빠른 윤곽선 추출이 가능하다.

$$S_x=\begin{bmatrix}-1 & 0 & 1\\ -2 & 0 & 2\\ -1 & 0 & 1\end{bmatrix}\ S_y=\begin{bmatrix}-1 & -2 & -1\\ 0 & 0 & 0\\ 1 & 2 & 1\end{bmatrix} \quad (5)$$

여기서 S_x는 X축 방향에 해당하는 필터이고, S_y는 Y축 방향에 해당하는 필터로써 기준이 되는 픽셀 $I(x,y)$와 식 (6)을 통해 기울기 값을 계산하고 그 크기 정보 M_s를 질감 정보 $H_t\in R^{N\times N}$로 추출하였다.

$$\nabla_x(x,y)=[I(x+1,y+1)+2I(x+1,y)+I(x+1,y-1)]-[I(x-1,y+1)+2I(x-1,y)+I(x-1,y-1)]$$

$$\nabla_y(x,y)=[I(x+1,y+1)+2I(x,y+1)+I(x-1,y+1)]-[I(x+1,y-1)+2I(x,y-1)+I(x-1,y-1)]$$

$$M_s=\sqrt{\nabla_x^2+\nabla_y^2} \quad (6)$$

그리고 색상 정보를 추출하기 위하여 입력 영상을 HSV 색상 공간상의 36개의 값으로 양자화 시킨 후, NxN 윈도우 영역 내의 픽셀들의 색상 평균값이 아닌 대표 색상을 추출하였다. 일반적으로 색상 정보를 추출함에 있어서 영역내의 픽셀들의 평균 색상을 추출하는 경우에는 그림 4와 같이 원영상이 표현하는 직관적인 색상과는 다르게 추출되는 경우가 발생한다. 따라서 양자화된 HSV 공간에서 윈도우 영역내의 픽셀들은 각 색상 히스토그램의 빈으로 누적되고 윈도우 영역 내에서 가장 많은 분포를 가지는 대표 색상을 특징점의 색상 특징 $H_c\in R^3$로 이용하였다.

1(a) 1(b) 1(c) 2(a) 2(b) 2(c)

그림 4 평균 색상과 대표 색상의 차이
(a) 원영역 (b) 평균색상 (c) 대표색상

• 특징점 의미 정보 모델링

영상 데이터베이스로부터 추출된 특징점들은 그림 5와 같이 각각의 클래스에 해당하는 카테고리별로 군집화 시켰다.

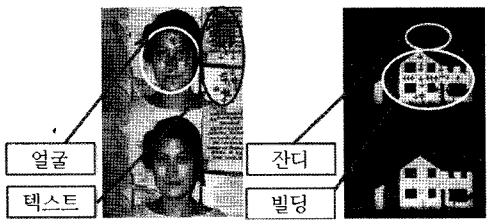

그림 5 특징점 모델링을 위한 카테고리별 군집화

그리고 위의 방법을 통하여 인덱싱된 각 특징 점들을 K-평균 알고리즘을 이용하여 모델링하였다. 영상에서 추출된 특징점 주변의 NxN 영역의 특징 점들을 기반으로 데이터의 군집수 K를 정하고 정의된 K는 Seed Point로 사용된다. 일반적으로 K-평균 알고리즘은 복잡도가 낮고, 좋은 결과를 보여주는 반면에 K값과 초기 평균값에 따라 큰 결과 차이를 보여준다. 본 논문에서는 실험에 의하여 K=40으로 선택하였다. 일반적으로 영상의 객체로부터 추출한 특징점들이 보다 높은 식별 성분을 가지기 위해서는 고차원의 특징 벡터가 요구된다. 즉, 윈도우 크기를 증가시켜서 특징 정보를 많이 추출하는 것은 분명 영상의 객체마다 추출되는 고유의 특징점들의 식별성을 높여주는 효과가 있다. 그러나 이러한 특

징 벡터의 차원이 증가하면 영상 분류를 위한 특징점의 색인 기법의 효율성이 떨어지는 문제가 발생한다. 따라서 본 논문에서는 특징 벡터의 양을 효율적으로 증가시키면서 이러한 문제를 해결하기 위하여 실험을 통하여 윈도우 크기를 21로 조절하고, 각각의 특징점들로부터 추출된 특징 벡터들을 주성분 분석법을 이용하여 20차원으로 축소시킴으로서 데이터의 양을 줄이고 계산량을 감소시켰다.

데이터의 특징 벡터 축소를 위하여 먼저 영상의 특징 데이터에 대하여 자기상관행렬 R_x와 그의 정규화된 고유치를 다음의 식 (7), (8)로부터 계산한다.

$$x(t)=[x_1,x_2,x_3,...,x_{444}]^T \quad (7)$$

$$R_x=E[xx^T] \quad (8)$$

여기서, $x(t)$는 영상의 색상,질감 특징 데이터이고, R_x는 특징데이터의 자기상관행렬이다. 이때 자기상관행렬 R_x의 정규화된 고유치(e_i) ($\|e_i\|=1,\ i=1,2,...,m$) 에 의하여 특징데이터 $x(t)$는 1차원으로 투영될 수 있다. 따라서 특징데이터 $x(t)$의 주성분은 식 (9)와 같이 R_x의 정규화된 고유치 중에 가장 큰 고유치인 (e_i)의 고유 요소로 표현할 수 있다.

$$a(t)=e_1^T x(t) \quad (9)$$

그 다음 추출한 R_x의 정규화된 고유치 $\{e_i\}$를 크기 역순으로 배열하고 그 중 20개를 이용하여 특징 데이터를 20차원으로 축소하였다.

3. 실험 결과 및 분석

본 논문에서는 다양한 객체로부터 특징점들을 추출하고 영상 분류를 위한 객체의 의미 정보로 사용하기 위한 카테고리 정보를 세분화하기 위하여 Caltech DB[10]와 MSRC DBv1[11] 그리고 Corel 1000DB로부터 8861장의 영상데이터를 선택하여 총 407,091개의 특징점을 추출하였다. 그리고 영상 데이터베이스에 존재하는 객체의 모든 내용을 표현할 수 있는 핵심 키워드 28개를 선정하고 이들을 객체의 의미 정보로 각 특징 점들을 인덱싱하였다. 영상 데이터베이스로부터 추출된 각각의 특징점들을 각각의 의미 정보에 해당하는 카테고리별로 군집화시킨 후 축소된 특징 벡터의 차원에서 K-평균 알고리즘을 적용하여 각 카테고리별로 K 평균에 해당하는 40개의 특징점들을 추출하여 모델링하였다. 28개의 카테고리로 세분화되어 추출된 특징점들에서 각각의 60%에 해당하는 데이터는 특징점들을 각각의 객체의 의미 정보를 인덱싱하기 위한 모델링에 사용하였고 나머지 40%의 객체 특징점들은 K-NN 방법을 통하여 가장 근접하게 속하는 한 개의 객체 카테고리로 분류하여

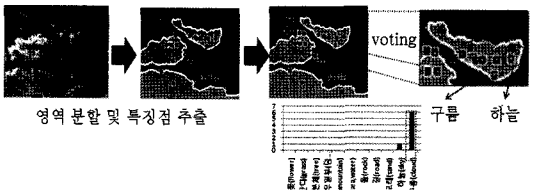

그림 6 K-NN 분류를 통한 객체 영역의 의미 정보 추출

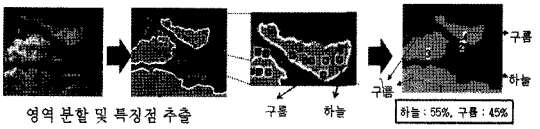

그림 7 특징점과 객체 영역 가중치를 이용한 영상 분류

성능을 평가하기 위해 사용되었다. 그림 6에서와 같이 객체 영역은 K-NN 방법에 의하여 구름이라는 의미 정보와 정합된다.

위의 과정을 이용하여 각각의 객체 단위에서 추출한 특징점들의 의미 정보와 각 객체 영역의 크기 가중치 값을 이용하여 영상은 최종적으로 그림 7과 같이 각 객체 영역들의 의미 정보의 조합으로 해석이 가능하며, 이 중에서 가장 큰 비중의 의미 정보를 가지는 카테고리 영상으로 분류 가능하다.

전체 영상 데이터베이스 내의 28개 객체 카테고리 정보에 의하여 추출된 특징점들의 개수와 이들을 이용한 객체의 분류 성능 결과는 표 1과 같다. 실험 결과를 통하여 본 논문에서 제안하는 방법이 28개의 객체 카테고리에 영상에 대해서 평균 81.2%의 높은 분류 성능을 보여줌을 알 수 있다. 또한 J.Shotto *et al.*[5]에서 사용된 방법과 제안하는 방법과의 성능 비교를 위하여 16개의 객체 카테고리에 따른 분류(인식) 성능을 비교 평가한 결과 그림 8에서와 같이 본 논문에서 제안하는 방법이 평균 82.8%로 [5]의 65.74%보다 17%가량 높았으며, 전체 16개의 카테고리에 대해서도 전반적으로 고른 분류 성능을 보임을 확인할 수 있다.

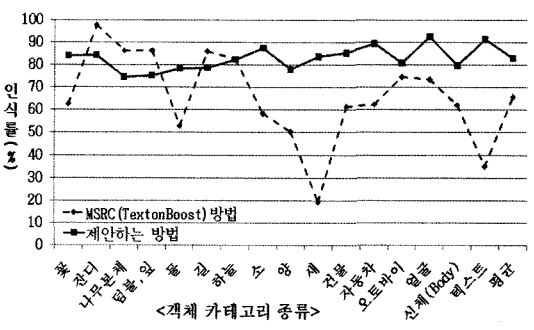

그림 8 16개 카테고리에 대한 [5]의 방법과의 성능 비교

표 1 분류 성능 결과

객체의미정보	인덱스	추출된 전체 특징점 개수	분류성능(%)
꽃	0	25300	84.1
잔디	1	16244	84.3
나무	2	6127	74.5
나뭇잎,덤불	3	97865	75.3
산	4	15502	83.6
물, 바다	5	6370	78.4
돌,바위	6	3290	80.2
길,도로	7	253	78.7
모래	8	3719	77.7
하늘	9	347	82.4
구름	10	2776	73.1
석양	11	602	72.6
소	12	3374	87.4
말	13	2476	70.2
양	14	5986	77.9
코끼리	15	5399	74.7
표범	16	9402	86.8
새	17	858	83.6
건물	18	18806	85.3
굴뚝	19	469	70.4
창문,현관문	20	2580	76.2
자동차	21	50403	89.3
버스	22	30633	88.2
비행기	23	44528	95.0
오토바이	24	19344	80.8
얼굴	25	8061	92.4
사람신체	26	255	79.6
텍스트,간판	27	26122	91.4
전체 분류 성능			81.2

4. 결 론

본 논문에서는 영상에서 의미적인 개념을 가지는 객체에 대해서 각각의 객체 단위로 추출이 가능한 고유한 특징점들을 이용하여, 특징점 주변의 내용 기반의 특징 정보와 이를 고차원의 객체 단위 의미 정보로 결합하여 영상을 효율적으로 분류하는 방법을 제안하였다. 기존의 객체 인식 및 영상 분류 방법들은 영상을 특정 객체 단위로 효과적으로 분할하기 어렵고 사용자가 가지는 주요 관심의 객체 영역에 대한 정의도 애매하기 때문에 객관적인 영상 분류가 어려웠다. 본 논문에서 제안하는 방법은 영상을 구성하는 다수의 객체 영역들에 대해서 각각의 의미 정보들을 추출하여 영상을 해석함으로써, 인간의 시각적 인지 능력과 유사하게 영상 내 존재하는 각 객체 영역들의 의미 정보에 의하여 영상을 분류할 수 있는 장점이 있다. 그리고 실험 결과를 통하여 제안하는 방법이 다수의 객체들에 대해서도 높은 분류 성능을 보임을 확인할 수 있었다. 본 논문에서 제안하는 방법은 영상 내 존재하는 특정 객체 인식을 통한 영상의 장면 이해와 다수 객체 영역들의 의미 정보를 이용한 자동 색인, 다중 의미 기반의 영상 검색 및 분류 등과 같은 다양한 분야에서 유용하게 활용 될 수 있을 것이다.

참 고 문 헌

[1] Z.Agabari, T.Ohasih and A.makinouchi, "Semantic Approach to Image Database Classification and Retrieval," *IEICE*, vol.103(192), pp.109-114, 2003.

[2] J.Vogel and B.Schiele, "Semantic Modeling of Natural Scenes for Content- Based Image Retrieval," *IJCV*, vol.72(2), pp.133-157, 2007.

[3] K.Grauman and T.Darrell, "The pyramid match kernel: Discriminative classification with sets of image features," *ICCV*, vol.2, pp.1458-1465, 2005.

[4] A.Bosch, A.Zisserman and X.Munoz, "Image Classification using Random Forests and Ferns," In *Proc. ICCV*, pp.1-8, 2007.

[5] J.Shotton, J.Winn, C.Rother and A.Criminisi, "TextonBoost : Joint Appearance, Shape and Context Modeling for Multi-Class Object Recognition and Segmentation," *ECCV*, vol.1, pp.1-15, 2006.

[6] Y.Deng and B.S.Manjunath, "Unsupervised segmentation of color-texture regions in images and video," *PAMI*, vol.8, pp.800-810, 2001.

[7] David G.Lowe, "Distinctive Image Features from Scale Invariant Keypoints," *IJCV*, vol.60(2), pp.91-110, 2004.

[8] M.Grabner, H.Grabner and H. Bischof, "Fast Approximated SIFT," *ACCV*, LNCS 3851, pp.918-927, 2006.

[9] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," *In Proc. CVPR*, vol.1, pp.511-518, 2001.

[10] Caltech categories dataset, http://www.vision.caltech.edu/html-files/archive.html

[11] MSRC_ObjectRecognition database(v1.0), http://research. microsoft.com/vision/cambridge/recognition