

Mean Shift 분석을 이용한 그래프 컷 기반의 자동 칼라 영상 분할

(Graph Cut-based Automatic Color Image Segmentation using Mean Shift Analysis)

박 안 진 [†] 김 정 환 [†] 정 기 철 ^{**}
(Anjin Park) (Jungwhan Kim) (Keechul Jung)

요 약 그래프 컷(graph cuts) 방법은 주어진 사전정보와 각 픽셀간의 유사도를 나타내는 데이터 항(data term)과 이웃하는 픽셀간의 유사도를 나타내는 스무드 항(smoothness term)으로 구성된 에너지 함수를 전역적으로 최소화하는 방법으로, 최근 영상 분할에 많이 이용되고 있다. 기존 그래프 컷 기반의 영상 분할 방법에서 데이터 항을 설정하기 위해 GMM(Gaussian mixture model)을 주로 이용하였으며, 평균과 공분산을 각 클래스를 위한 사전정보로 이용하였다. 이 때문에 클래스의 모양이 초구(hyper-sphere) 또는 초타원(hyper-ellipsoid)일 때만 좋은 성능을 보이는 단점이 있다. 다양한 클래스의 모양에서 좋은 성능을 보이기 위해, 본 논문에서는 mean shift 분석 방법을 이용한 그래프 컷 기반의 자동 영상분할 방법을 제안한다. 데이터 항을 설정하기 위해 $L^*u^*v^*$ 색상공간에서 임의로 선택된 초기 mean으로부터 밀도가 높은 지역인 모드(mode)로 이동하는 mean의 집합들을 사전정보로 이용한다. Mean shift 분석 방법은 군 집화에서 좋은 성능을 보이지만, 오랜 수행시간이 소요되는 단점이 있다. 이를 해결하기 위해 특징공간을 3차원 격자로 변형하였으며, mean의 이동은 격자에서 모든 픽셀이 아닌 3차원 윈도우내의 1차원 모멘트(moment)를 이용한다. 실험에서 GMM을 이용한 그래프 컷 기반의 영상분할 방법과 최근 많이 이용되고 있는 mean shift와 normalized cut기반의 영상분할 방법을 제안된 방법과 비교하였으며, Berkeley dataset 을 기반으로 앞의 세 가지 방법보다 좋은 성능을 보였다.

키워드: 그래프 컷, Mean Shift 분석, 영상분할

Abstract A graph cuts method has recently attracted a lot of attentions for image segmentation, as it can globally minimize energy functions composed of data term that reflects how each pixel fits into prior information for each class and smoothness term that penalizes discontinuities between neighboring pixels. In previous approaches to graph cuts-based automatic image segmentation, GMM(Gaussian mixture models) is generally used, and means and covariance matrixes calculated by EM algorithm were used as prior information for each cluster. However, it is practicable only for clusters with a hyper-spherical or hyper-ellipsoidal shape, as the cluster was represented based on the covariance matrix centered on the mean. For arbitrary-shaped clusters, this paper proposes graph cuts-based image segmentation using mean shift analysis. As a prior information to estimate the data term, we use the set of mean trajectories toward each mode from initial means randomly selected in $L^*u^*v^*$ color space. Since the mean shift procedure requires many computational times, we transform

· 이 논문은 2006년도 정부재원(교육인적자원부 학술연구조성사업비)으로 한국 학술진흥재단의 지원(KRF-2006-331-D00523)과 지식경제부 및 정보통신연구진흥원의 대학 IT연구센터 지원사업(NIPA-2009-(C1090-0902-0007))의 연구 결과로 수행되었음

· 이 논문은 2008 한국컴퓨터종합학술대회에서 'Mean Shift 분석을 이용한 그래프 컷 기반의 자동 칼라 영상 분할'의 제목으로 발표된 논문을 확장한 것임

[†] 학생회원 : 숭실대학교 미디어학과
anjin@ssu.ac.kr

kim598@ssu.ac.kr

^{**} 종신회원 : 숭실대학교 미디어학부 교수
kcjung@ssu.ac.kr

논문접수 : 2008년 8월 25일

심사완료 : 2009년 8월 28일

Copyright©2009 한국정보과학회: 개인 목적이나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.

정보과학회논문지: 소프트웨어 및 응용 제36권 제11호(2009.11)

features in continuous feature space into 3D discrete grid, and use 3D kernel based on the first moment in the grid, which are needed to move the means to modes. In the experiments, we investigate the problems of mean shift-based and normalized cuts-based image segmentation methods that are recently popular methods, and the proposed method showed better performance than previous two methods and graph cuts-based automatic image segmentation using GMM on Berkeley segmentation dataset.

Key words : Graph Cuts, Mean Shift Analysis, Image Segmentation

1. 서론

영상분할의 목적은 입력영상을 동질의 속성을 가진 영역으로 나누는 것으로, 분할된 영역은 겹치지 않아야 하고 분할된 영역의 합집합은 입력영상이 되어야 한다. 이런 영상분할 기법은 컴퓨터 비전이나 영상처리 분야에서 물체 추적 및 인식(object tracking and recognition), 영상 검색(image retrieval), 얼굴 추출 및 인식(face detection and recognition) 등 다양한 응용분야를 위한 전처리 단계로써 매우 중요하다[1].

기존의 영상분할 방법은 크게 특징공간(feature space) 기반의 방법과 영상공간(image domain) 기반의 방법 두 가지로 분류된다[2]. 특징공간 기반의 방법은 특징공간에 표현된 색상이나 텍스처와 같은 특징값을 이용하는 방법으로, 일반적으로 군집화 방법을 이용하여 특징값을 분류하고 분류된 특징값의 레이블이 영상 분할의 결과로 이용된다[2]. 특징공간 기반의 방법은 확실하게 군집화될 수 있는 특징값을 가진 영상에서 좋은 성능을 보이지만, 영상공간 상에서 픽셀의 위치관계를 고려하지 못하기 때문에 잡음이 많거나 복잡한 특징값을 가진 영상(색상이 복잡한 영상)에서 좋은 성능을 보이지 못하는 단점이 있다. 이를 해결하기 위해 Watershed 알고리즘[3]과 영상공간에서의 mean shift 알고리즘[4] 등과 같은 영상공간 기반의 방법이 많이 이용되고 있다. 하지만 이 방법들은 에지 정보를 고려하기 때문에 잡음이 많은 영상에서 작은 영역으로 많이 분할하는 문제점(over-segmentation)을 가지며, 이 때문에 작은 영역을 병합하는 추가적인 알고리즘이 필요하다[5].

최근 위 두 방법의 문제점들을 해결하기 위해 특징공간과 영상공간을 동시에 고려하는 그래프 컷(graph cut)[6] 기반의 방법들이 영상분할을 위해 다양하게 시도되고 있다[7-11]. 그래프 컷은 데이터 항(data term)과 스무드 항(smoothness term)으로 구성된 에너지 함수를 전역적으로 최소화하는 방법으로, 데이터 항은 특징공간에서 주어진 레이블을 위한 사전정보와 각 픽셀(pixel)간의 유사도를 나타내며, 스무드 항은 영상공간에서 이웃하는 픽셀간의 유사도를 나타낸다. 이로 인해 그래프 컷은 두 공간을 동시에 고려할 수 있는 장점이 있다. Li 등[7], Boykov와 Funka-Lea[8], 그리고 Rother

등[9]은 그래프 컷을 이용한 대화식(interactive) 영상분할 툴을 제안하였다. Li 등[7]과 Boykov와 Funka-Lea [8]가 제안한 툴에서는 마우스 버튼을 누른 상태에서 드래그하여 전경과 배경에 대한 사전정보를 얻었으며, 마우스 왼쪽 버튼은 전경에 대한 정보를 오른쪽 버튼은 배경에 대한 정보를 획득하기 위해 이용되었다. Rother 등[9]이 제안한 툴에서는 전경 물체의 주위를 사각형으로 표시하여 기존방법[7,8] 보다 쉽게 전경과 배경에 대한 사전정보를 얻었다. 위에서 언급한 방법들[7-9]은 그래프 컷을 이용하여 좋은 성능을 보였지만, 데이터 항을 설정할 때 전경, 배경을 위한 사전정보를 사용자가 직접 입력해야 하는 단점을 가지고 있다.

이와 같은 대화식 영상분할 툴은 수동적인 방법으로 데이터 항을 자동으로 설정하기 위해 GMM(Gaussian mixture models)이 주로 이용되었으며, EM 알고리즘에 의해 계산된 평균과 공분산을 기반으로 데이터 항을 설정하였다[10,11]. 이 방법은 특징공간 상에 표현된 특징값의 밀도함수를 추정함으로써 쉽게 데이터 항을 설정할 수 있지만, 평균과 공분산으로 클래스를 표현하기 때문에 클래스의 모양이 초구(hyper-sphere) 또는 초타원(hyper-ellipsoid)일 때만 좋은 성능을 보이며, 클래스의 수 또한 미리 알아야 한다. 결과적으로 GMM을 이용한 방법들은 특징공간 상에서 실제 데이터를 정확하게 분석, 추정하기에 적당하지 않다[4].

다양한 클래스의 모양에서 좋은 성능을 유지하기 위해, 본 논문에서는 mean shift 분석 방법을 이용한 그래프 컷 기반의 자동 영상 분할 방법을 제안한다. mean shift 기반의 방법은 특징공간에서 다양한 클래스의 모양을 보다 정확하게 분석할 수 있는 장점 때문에 GMM 기반 방법의 첫 번째 문제점을 해결할 수 있으며, 클래스의 개수를 자동으로 찾을 수 있기 때문에 GMM 기반 방법의 두 번째 문제를 해결할 수 있다[4]. 본 논문에서는 특징공간에서 임의로 선택된 초기 mean으로부터 밀도가 높은 지역인 모드로 이동하는 mean들의 집합을 각 클래스를 위한 사전정보로 이용하며, 이 사전정보를 기반으로 데이터 항을 자동으로 설정한다. mean shift 분석 방법은 특징값을 군집화할 때 다양한 클래스의 모양에서 좋은 성능을 보이지만[12,13], 가장 큰 문제점은

많은 수행시간이 소요되는 것이며, 시간 복잡도(complexity)는 $O(nm)$ 이다. 여기서 n 은 입력 영상의 픽셀 수이며, m 은 임의로 선택된 초기 mean의 수이다[14]. 우리는 수행시간을 줄이기 위해 먼저 본 논문에서 사용한 $L^*u^*v^*$ 칼라공간을 3차원 격자(grid)로 변형하며, 이렇게 변형된 격자에서 $h \times h \times h$ 3차원 윈도우내에서 중심값을 계산하는 1차원 모멘트(moment)를 이용하여 mean을 이동한다. 본 논문에서 $L^*u^*v^*$ 칼라공간을 사용한 이유는 $L^*u^*v^*$ 칼라공간은 인간이 색상을 인식하는 방식과 유사하며, $L^*u^*v^*$ 각 축들은 독립적으로 균일한 색상 공간에 가깝게 설계되어 있어 기본적인 칼라공간 보다 유용하기 때문[15]으로 실제로 실험한 다른 몇몇 색상공간 RGB, HSV, YUV 보다 평균적으로 좋은 결과를 보여주었다.

물체분할을 위해 이용된 최초의 그래프 컷 방법은 단지 두 개의 터미널만을 가지기 때문에 이진 레이블링 문제에만 적용이 가능하다. 본 논문에서 다중(multi) 레이블링을 위해 α -expansion 방법을 이용하며, 스무드 항을 설정하기 위해 α -expansion 방법에 최적화된 Potts 모델을 이용한다[6]. 실험에서 최근에 많이 이용되고 있는 mean shift와 normalized cut기반의 영상분할 방법에 대한 문제점을 분석하였고, 제안된 방법이 기존의 문제점을 해결함으로써 Berkeley Segmentation Dataset을 기반으로 이전의 두 가지 방법과 GMM을 이용한 그래프 컷 기반의 방법보다 더 좋은 성능을 보였다.

본 논문의 구성은 다음과 같다. 제 2장에서 영상 분할을 위한 에너지 함수를 설명하며, 제 3장에서 자동 레이블링 방법과 다중 레이블링 방법에 대해 기술한다. 제 4장에서 실험 및 결과에 대해 언급하고, 제 5장에서 결론 및 향후 연구 방향을 기술한다.

2. 영상 분할을 위한 에너지 함수

본 논문에서는 영상분할 문제를 레이블링(labeling) 문제로 고려한다. 레이블링 문제는 주어진 영상 P 의 각 픽셀 p 에 레이블 집합 중 하나를 할당하는 것으로 $F = \{f_1, f_2, \dots, f_n\}$ 로 표시하며, F 의 각 요소는 레이블의 집합 중 하나가 할당되고 n 은 픽셀의 수를 나타낸다. 입력영상의 각 픽셀에 적합한 레이블을 할당하기 위해, 영상분할 문제를 먼저 에너지 함수로 표현하며 식 (1)과 같다.

$$E(F) = \sum_{p \in P} D_p(f_p) + \lambda \sum_{(p,q) \in N} V_{p,q}(f_p, f_q) \quad (1)$$

여기서 $D_p(f_p)$ 는 픽셀 p 가 f_p 에 얼마나 적합한지를 나타내는 데이터 항으로, 픽셀 p 가 레이블 f_p 에 대한 사전 정보와 유사하면 $D_p(f_p)$ 에 높은 비용(cost)을 할당하며, 유사하지 않으면 낮은 비용을 할당한다. $V_{p,q}(f_p, f_q)$ 는

스무드 항으로 영상공간에서 이웃하는 픽셀과의 연속성(continuity)을 나타내며, 이웃하는 두 픽셀이 유사하면 높은 비용(cost)을 할당한다. 우리는 스무드항을 Potts 모델로 표현하기 위해 $V_{p,q}(f_p, f_q)$ 을 $V_{p,q} \cdot \delta(f_p, f_q)$ 로 대체한다. 여기서 $\delta(f_p, f_q)$ 는 델타 함수로 f_p 와 f_q 가 같은 레이블을 가지면 0, 아니면 1의 값을 가지며, 두 픽셀이 다른 레이블을 가질 때의 비용만을 고려하기 위해 사용한다. λ 는 두 항의 상대적인 중요성을 나타내며, N 은 입력영상에서 이웃하는 픽셀 쌍의 집합으로 n -links(neighborhood links)라 불린다.

에너지 함수를 최소화하기 위해 우리는 그래프 컷을 사용하며 이를 위해 먼저 픽셀과 동일한 정점(vertex)을 가진 그래프 $G = \langle V, E \rangle$ 를 생성한다. 그래프에 레이블 정보를 반영하기 위해 두 개의 추가적인 정점 source (S)와 sink(T)가 필요하며 각 정점에 t -links(terminal links)라 불리는 에지 $\{p, S\}$ 와 $\{p, T\}$ 를 추가하게 된다. 결론적으로 그래프의 v 와 ϵ 는 식 (2)와 같이 나타낼 수 있다.

$$v = P \cup \{S, T\}, \quad \epsilon = N \cup_{p \in P} \{\{p, S\}, \{p, T\}\} \quad (2)$$

그래프의 가중치는 t -links와 n -links에 할당되며, 터미널과 각 정점을 연결하는 t -links는 픽셀과 레이블 사이의 유사도를 나타내는 데이터 항이, 근접한 이웃 정점을 연결하는 n -links에는 두 근접한 픽셀 사이의 연속성을 나타내는 스무드 항이 그래프의 가중치로 각각 할당된다.

에너지 함수에서 데이터항의 비용을 할당하기 위해 각 레이블을 위한 사전정보가 필요하다. 본 논문에서 우리는 특징공간에서 임의로 선택된 초기 mean으로부터 각 모드로 이동하는 mean의 집합을 각 레이블을 위한 사전정보로 이용하며, 자세한 내용은 3장에서 설명할 것이다. 이렇게 획득한 각 레이블의 집합은 수행시간을 줄이기 위해 히스토그램으로 표현하며, 데이터 항은 식 (3)과 같이 정의한다.

$$\begin{cases} D_p(f_p = 0) = \infty, & D_p(f_p = 1) = 0, & \forall p \in L_0, \\ D_p(f_p = 0) = 0, & D_p(f_p = 1) = \infty, & \forall p \in L_1, \\ D_p(f_p = 0) = \ln P(p|H_0), & D_p(f_p = 1) = \ln P(p|H_1), & \forall p \in U. \end{cases} \quad (3)$$

여기서 레이블 0과 1을 위한 사전정보의 집합을 L_0 과 L_1 로 표현하며, U 는 L_0 과 L_1 에 속하지 않는 픽셀들을 나타낸다. $\ln P(p|H_i)$ 는 집합 L_i 의 히스토그램 H_i 가 주어졌을 때 p 의 확률값을 나타낸다.

스무드 항의 비용은 각 픽셀 사이의 불연속 보전(discontinuity-preserving)을 위해 할당되고 이를 위해 우리는 Potts 모델을 사용하였으며 $V_{p,q}$ 는 식 (4)와 같이 정의된다.

$$V_{p,q} = \text{dis}(p, q)^{-1} \exp(-\beta \|p - q\|^2) \quad (4)$$

여기서 $dis(\cdot)$ 는 영상공간에서 픽셀 사이의 거리를 나타낸다. 상수 β 가 0일 때 스무드 함은 기존에 많이 이용되었던 Ising 모델로 영상공간에서 평탄화(smoothness)를 위해 이전에 많이 이용되었지만, 에지와 같은 대비가 높은 영역 역시 평탄하게 만드는 경향이 있다. 이 때문에 최근에는 β 에 0보다 높은 값을 할당하여 이런 문제를 해결하였으며, 상수 β 를 식 (5)와 같이 설정하면 더 좋은 성능을 보였다[8].

$$\beta = (\|p - q\|^2)^{-1} \quad (5)$$

여기서 $\langle \cdot \rangle$ 는 모든 p 와 q 의 비유사도에 대한 평균을 의미한다.

그래프 컷은 에너지 함수를 최소화시키기 위해 제안된 여러 방법들 중 빠르고 좋은 결과를 보여주는 방법 [16,17]이며 그래프 G 가 설정된 후에 가중치의 합이 최소가 되는 컷(cut)을 찾는 방법으로 컷된 두 터미널 S 와 T 에 각각 연결된 각 픽셀에 레이블을 할당한다[18]. 최소 컷을 찾기 위한 max flow를 구하기 위한 방법으로 Dinic[19]의 이론, Ford와 Fulkerson[20]의 이론, 그리고 HPRF[21] 방법과 QPRF[21]의 방법이 제안되었고 그 중 가장 빠른 방법은 Ford와 Fulkerson[20]의 이론이다[17].

따라서 본 논문은 Ford와 Fulkerson[20]의 이론을 기반으로 S 에서 T 로 흐르는 max flow를 찾는 것으로 최소 컷을 구한다. max flow는 정점을 두 개의 중복되지 않는 영역으로 나누는 그래프의 에지들을 산출하며, 그 에지의 가중치 합이 최소 컷의 비용과 같다. 따라서 그래프에서 max flow가 t -link와 n -link가 할당된 가중치의 합이 최소가 되는 에지를 산출하기 때문에, 본 논문에서 max flow를 에너지 함수를 전역으로 최소화하기 위해 이용한다[18].

3. 자동 다중 레이블링 방법

2장에서 설명한 에너지 함수는 각 클래스를 위한 사전정보가 필요한 문제점이 있으며, 에너지 함수를 최소화하기 위해 이용된 그래프 컷 방법은 두 개의 터미널만을 가지기 때문에 이진 레이블링 문제에만 적용이 가능한 문제점이 있다. 3.1에서 영상분할을 자동으로 레이블링하는 방법에 대해 설명하며, 3.2에서 이진 레이블링 문제를 다중 레이블링 문제로 확장하는 방법에 대해 설명한다.

3.1 영상분할을 위한 자동 레이블링 방법

우리는 데이터 항을 설정하기 위해 특징공간에서 임의로 선택된 초기 mean으로부터 클래스의 중심으로 이동하는 모드로 이동하는 mean의 경로를 각 클래스의 사전정보로 이용한다. 이는 하나의 모드를 향하는 mean

의 경로들이 명시적으로 하나의 클래스에 속하며, mean shift 분석 방법이 다양한 모양의 클래스를 가진 특징공간을 정교하게 분석할 수 있는 장점 때문이다[4]. 그래서 우리는 먼저 mean shift 분석 방법에 대해 간단하게 기술한다.

최근 커널 밀도추정(kernel density estimation) 함수가 다양한 모양의 클래스를 가진 특징공간에서 밀도를 추정하기 위해 많이 이용되고 있으며, 추정된 밀도의 모드는 특징공간에 특징값이 밀집되어 있는 위치와 같기 때문에 모드는 클래스의 중심으로 이용될 수 있다. mean shift는 커널 밀도추정 함수에서 기울기 정보를 기반으로 모드를 찾아가는 방법이다.

커널 밀도추정 함수는 식 (6)과 같이 정의된다.

$$\hat{f}_k(p) = \sum_{j=1}^n k(\|p - p_j\|^2) w(p_j) \quad (6)$$

여기서 $w(p_j)$ 는 가중치 함수이며, $k(\cdot)$ 는 커널 함수이다. 모드는 추정 함수의 기울기가 0인 위치이며, 만약 c 가 상수인 $k'(r) = c \cdot sk(r)$ 를 만족하는 커널 $SK(p) = sk(\|p - p_j\|^2)$ 가 존재한다면 기울기는

$$\begin{aligned} \nabla \hat{f}_k(p) &= \sum_{j=1}^n (p - p_j) k'(\|p - p_j\|^2) w(p_j) \\ &= \left[\sum_{j=1}^n sk(\|p - p_j\|^2) w(p_j) \right] \times \left[\frac{\sum_{j=1}^n p_j sk(\|p - p_j\|^2) w(p_j)}{\sum_{j=1}^n sk(\|p - p_j\|^2) w(p_j)} - p \right] \\ &= \hat{f}_{sk}(p) [mean_{sk}(p) - p]. \end{aligned} \quad (7)$$

와 같이 기술할 수 있다. 식 (7)에서 $mean_{sk}(p)$ 와 p 가 같으면 기울기가 0이 되며, 이때 p 가 모드이다.

mean shift는 초기단계에서 랜덤하게 선택된 데이터 집합을 mean의 초기값으로 설정하며, 각각의 mean ms_j 는 $mean_{sk}(ms_j)$ 를 이용하여 갱신한다(즉 $ms_{j+1} \leftarrow mean_{sk}(ms_j)$). 그리고 ms_j 는 모드에 도착할 때까지 반복해서 갱신되며, k 번째 모드로 향하는 연속적인 mean의 집합을 $ms^k = \{ms_j^k\}$, $j=1, 2, \dots$,와 같이 표현한다. 여기서 ms_j^k 은 mean의 초기 위치이며, ms^k 은 클래스 k 를 위한 사전정보로 2장의 L_k 로 이용된다.

$$\text{식 (7)에서 } mean_{sk}(ms) = \frac{\sum_{j=1}^n p_j sk(\|ms - p_j\|^2) w(p_j)}{\sum_{j=1}^n sk(\|ms - p_j\|^2) w(p_j)}$$

으로 모든 픽셀들을 고려하기 때문에 많은 수행시간이 소요되며, 시간 복잡도는 $O(mn)$ 이다[14]. 여기서 n 은 입력 영상의 픽셀 수이며, m 은 임의로 선택된 초기 mean의 수이다. 우리는 수행시간을 줄이기 위해 먼저 본 논문에서 사용한 $L^*u^*v^*$ 칼라공간을 3차원 격자(TDG : Three Dimensional Grid)로 변형하였으며, TDG는 정수화된 특징값의 좌표를 기반으로 0과 1값으로 설정된다. 이렇게 변형된 격자 TDG에서 mean을 이

동하기 위해 전체 픽셀이 아닌 $h \times h \times h$ 윈도우 내에서 1차원 모멘트를 이용하여 중심값을 구하면, 식 (8)과 같다.

$$x_{C_{j+1}} = \frac{M_{100}}{M_{000}}, y_{C_{j+1}} = \frac{M_{010}}{M_{000}}, z_{C_{j+1}} = \frac{M_{001}}{M_{000}}, \quad (8)$$

여기서

$$M_{abc} = \sum_{xt=xc_j+h/2}^{xt=xc_j-h/2} \sum_{yt=yc_j+h/2}^{yt=yc_j-h/2} \sum_{zt=zc_j+h/2}^{zt=zc_j-h/2} xt^a yt^b zt^c TDG(xt, yt, zt)$$

이며, mean ms_j 는 $(x_{C_j}, y_{C_j}, z_{C_j})$ 로 표현된다. 1차원 모멘트를 이용할 경우 모든 픽셀이 아닌 $2xt \times 2yt \times 2zt$ 격자 내의 픽셀만을 고려하면 되기 때문에 모든 픽셀을 고려하는 것 보다 빠른 수행시간을 보인다. 식 (7)에서 $means_k(ms_j)$ 는 커널 SK 에 의한 ms_j 의 가중치 평균을 의미하며, 만약 커널 함수를 평면(flat) 함수(식 (9))로 가정하면 식 (8)과 같은 결과를 보인다. 따라서 본 논문에서 사용한 mean shift 함수(식 (8))는 식 (7)을 이용한 방법과 유사한 결과를 보이지만 보다 빠르게 모드를 찾아가는 장점이 있다.

$$SK(ms) = \begin{cases} 1, & \|ms - p_j\|^2 \leq h, \\ 0, & \|ms - p_j\|^2 > h. \end{cases}$$

3.2 다중 레이블링을 위한 자동 군집화

2장에서 영상분할을 두 개의 클래스로 레이블링하는 방법을 설명하였다. 이 방법은 배경에서 물체를 추출하는 것과 같이 이진화 문제에서 좋은 성능을 보이지만, 대부분의 레이블링 문제는 세 개 이상의 클래스를 포함한다. 결론적으로 이진 레이블링 문제를 다중 레이블링 문제로 확장하는 다중 그래프 컷 방법이 요구된다. 우리는 다중 레이블링을 위해 스무드 항이 본 논문에서 이용한 Potts 모델일 때 에너지 함수를 전역적으로 최적할 수 있는 장점을 가지고 있는 α -expansion 알고리즘을 이용한다[6].

- | |
|---|
| <ol style="list-style-type: none"> 1. Start with an arbitrary labeling F 2. Set success := 0 3. For each label $\alpha \in L$ <ol style="list-style-type: none"> 3.1. Find $F^* = \text{argmin} E(F')$ within one α-expansion of F 3.2. If $E(F^*) < E(F)$, set $F := F^*$ and success := 1 4. If success = 1 goto 2 5. Return F |
|---|

그림 1 α -expansion 알고리즘의 의사 코드

초기에는 각 픽셀에 레이블을 임의로 할당하며, 그 중 하나의 레이블을 α 라고 할 때, 만약 $l \neq \alpha$ 일 때 $P_\alpha \subset P'_\alpha$, $P'_l \subset P_l$ 를 만족하게 레이블링 F에서 새로운 레이블링 F'로 변환하면 이 움직임을 α -expansion이라고 하며, 여

기서 $P_l = \{p \in P \mid f_p = l\}$ 이며 레이블 l 이 할당된 픽셀의 집합을 나타낸다. 다시 말해서 α -expansion은 α 가 레이블된 픽셀(P_α)의 수를 확장하는 방법이다. 레이블링 F가 주어졌을 때 expansion의 움직임은 굉장히 많은 수가 존재한다. 여기서 그래프 컷 방법이 최적의 expansion 움직임을 찾기 위해 이용되며, 그림 1은 α -expansion 알고리즘은 보여준다. 그림 1에서 3.1, 3.2를 '반복'(iteration), 2, 3, 4를 '순환'(cycle)이라고 부른다. 알고리즘의 각 '순환'에서 모든 레이블을 위해 '반복'을 수행하며, 레이블의 순서는 임의로 설정된다.

최적의 expansion 움직임을 찾는 것은 그래프 $G_\alpha = \langle v_\alpha, \epsilon_\alpha \rangle$ 에서 최소 컷과 일치하는 레이블링을 계산하는 것으로, 이 그래프의 구조는 현재의 레이블링 F와 레이블 α 에 의해 결정되고, 그래프는 각 '반복' 후에 동적으로 변한다.

그래프의 구조는 다음과 같다. 정점의 집합은 두 개의 터미널 α, α' 와 영상 픽셀 $p \in P$ 를 포함하며, 추가적으로 현재 레이블링 F에서 다른 레이블을 가진 이웃하는 픽셀들 $\{p, q\} \in N$ 의 각 쌍들($f_p \neq f_q$)을 위해 보조 정점 $a_{(p,q)}$ 을 포함한다. 따라서 정점의 집합은

$$v_\alpha = \left\{ \alpha, \alpha', P, \bigcup_{\substack{\{p,q\} \in N, \\ f_p \neq f_q}} a_{(p,q)} \right\}$$

로 정의된다.

그래프의 에지는 각 픽셀 $p \in P$ 와 터미널 사이를 연결하는 t -links $t_p^\alpha, t_p^{\alpha'}$ 와 이웃하는 픽셀을 연결하는 n -links로 구성되며, 여기서 n -links는 보조 정점을 고려해야 한다. f_p 와 f_q 가 다르면 세 쌍의 에지 $e_{(p,q)} = \{e_{(p,\alpha)}, e_{(\alpha,q)}, t_p^{\alpha'}\}$ 를 할당하며, 여기서 $e_{(p,q)}$ 는 이웃하는 픽셀 p 와 q 를 연결하는 에지(n -link)이며, $\alpha = a_{(p,q)}$ 는 예비 정점과 대응한다. $e_{(p,\alpha)}$ 와 $e_{(\alpha,q)}$ 에지는 픽셀 p, q 와 $a_{(p,q)}$ 를 연결하는 에지이며, t -link $t_p^{\alpha'}$ 는 예비 노드 $a_{(p,q)}$ 와 α' 을 연결하는 에지이다. 그래서 모든 에지는 다음과 같이 정의할 수 있다.

$$\epsilon_\alpha = \left\{ \bigcup_{p \in P} \{t_p^\alpha, t_p^{\alpha'}\}, \bigcup_{\substack{\{p,q\} \in N, \\ f_p \neq f_q}} \epsilon_{(p,q)}, \bigcup_{\substack{\{p,q\} \in N, \\ f_p = f_q}} e_{(p,q)} \right\}$$

에지에 할당되는 가중치는 표 1과 같이 정의된다.

4. 실험 및 결과

제안된 방법의 성능을 평가하기 위해 Berkeley Segmentation Dataset[22]을 이용하였으며, 테스트 영상의 크기는 320×480 또는 480×320 이다. 모든 실험은 팬티엄

표 1 다중 레이블을 위한 가중치

에지	t_p^α	t_p^α	$t_p^{\alpha'}$	$e_{\{p,a\}}$	$e_{\{a,p\}}$	$t_p^{\alpha'}$	$e_{\{q,p\}}$
가중치	∞	$D_p(\alpha)$	$D_p(f_p)$	$V(f_p, \alpha)$	$V(\alpha, f_q)$	$V(f_p, f_q)$	$V(f_p, f_q)$
조건	$p \in P_\alpha$	$p \notin P_\alpha$	$p \in \mathbf{P}$	$\{p, q\} \in \mathbf{N}, f_p \neq f_q$			$\{p, q\} \in \mathbf{N}, f_p = f_q$

4 2.66GHz 컴퓨터에서 수행하였으며, 공개 소스인 그래프 컷과 에너지 최소화 라이브러리[6], Mean Shift 기반의 영상분할 코드[4], Normalized cut을 위한 Matlab 코드[23]를 이용하여 구현 및 테스트하였다.

그림 2는 제안된 방법의 결과 영상이다. 본 실험에서 사용자에게 의해 입력받는 매개 변수는 mean을 mode로 이동하기 위해 이용된 윈도우의 대역너비(bandwidth) h 와 두 에너지 항의 상대적인 중요도 λ 이다. 본 실험에서 h 의 값은 각 영상에서 수동으로 설정하였으며, λ 는 15로 고정하였다.

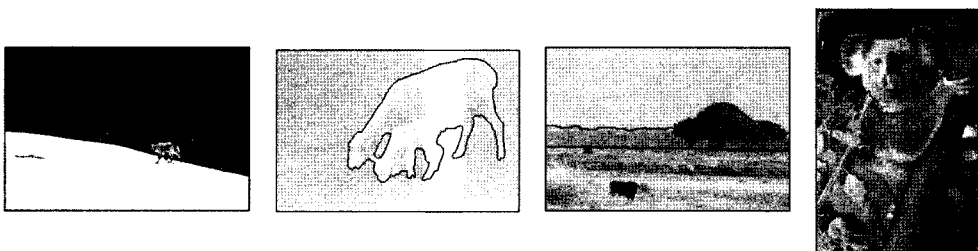
제안된 방법을 평가하기 위해 Normalized cut(NCut) 기반의 분할방법[23], Mean Shift(MS) 기반의 분할방법[4], GMM을 이용한 그래프 컷 기반의 분할방법[11]과 비교하였으며, 테스트 영상으로 Berkeley Segmentation Dataset으로부터 선택된 60개 컬러 영상을 이용하였다. NCut 알고리즘과 GMM기반의 알고리즘에서 각 영상의 클래스 개수는 테스트 영상[24]에서 제공하는 사람이 나눈 영역의 평균 개수와 같게 설정하였으며 그림 3은 각 방법과 제안된 방법의 결과를 비교한 영상이다.

Berkeley Segmentation Dataset은 각 영상마다 5~7개의 사람이 분할한 결과를 가지며, 그림 3(b)는 사람이 분할한 결과를 누적한 영상으로 누적 빈도가 높을수록 검은색에 가깝다. 우리는 이 결과를 ground truth로 이용하였으며, 영상분할의 수량적 평가는 precision, recall 그리고 F-measure를 이용하였다[22]. 우리는 5~7개의 ground truth와 비교한 결과의 평균으로 하나의 영상에 대한 precision(P)과 recall(R)을 계산하였으며, F-measure(F)는 precision, recall로부터 $F=PR/(\zeta R+(1-\zeta)P)$ 과 같이 계산된다. 여기서 ζ 는 precision과 recall 사이의

중요도를 나타내며 본 논문에서는 0.5로 설정하였다. 표 2는 기존의 3가지 방법과 제안된 방법의 precision, recall, F-measure를 보여주며, 기존의 3가지 방법보다 F-measure에서 좋은 성능을 보여주었고 제안된 방법은 기존의 방법보다 사람이 분할한 성능에 보다 가까웠다.

기존 방법들의 문제점은 다음과 같다. NCut기반의 결과는 그림 3에서 보는 바와 같이 거의 동일한 크기로 영상을 분할하며, 이는 NCut 알고리즘이 기본적으로 disassociation 함수를 이용하여 가중치 그래프를 거의 균형적으로 분할하기 위해 제안된 방법이기 때문이다[25]. MS기반의 결과는 다수의 영역으로 나누어지는 단점이 있다. 이는 영상공간에서 MS를 수행할 때 모든 에지 정보를 고려하기 때문이며, 영상공간만을 고려할 때 발생하는 근본적인 문제점이다[25].

마지막으로 GMM기반의 방법은 앞의 두 방법과 비교해서 좋은 성능을 보일지라도, 그림 4(a,b,e,f)에서 보는 바와 같이, 대부분의 경우 평균을 중심으로 한 공분산을 기반으로 클래스를 설정하기 때문에 특징공간에서 특징값을 정확하게 분석하지 못하는 단점이 있다. 뿐만 아니라 EM 알고리즘에 의해 계산된 평균과 공분산은 초기 값에 영향을 받기 때문에 항상 같은 결과를 보이지는 않는다. 그림 5는 제안된 방법의 결과 영상과 추정된 mean의 궤적을 보여준다. 그림 5에서 mean의 궤적을 3차원에서 그리기 어려워 두 개의 2차원 공간 L^*u^* (그림 5(a,d))와 L^*u^* (그림 5(b,e))로 나누어 보여주며, 사각형은 모드를 의미한다. 제안된 방법의 장점은 그림 5에서 보는 바와 같이 특징공간에서 특징값을 GMM보다 정확하게 분석하며, 초기값에 관계없이 항상 같은 결과를 보인다.



(a) (b) (c) (d)

그림 2 제안된 방법의 결과: (a) 대역너비 $h=9$, (b) $h=13$, (c) $h=13$, (d) $h=15$

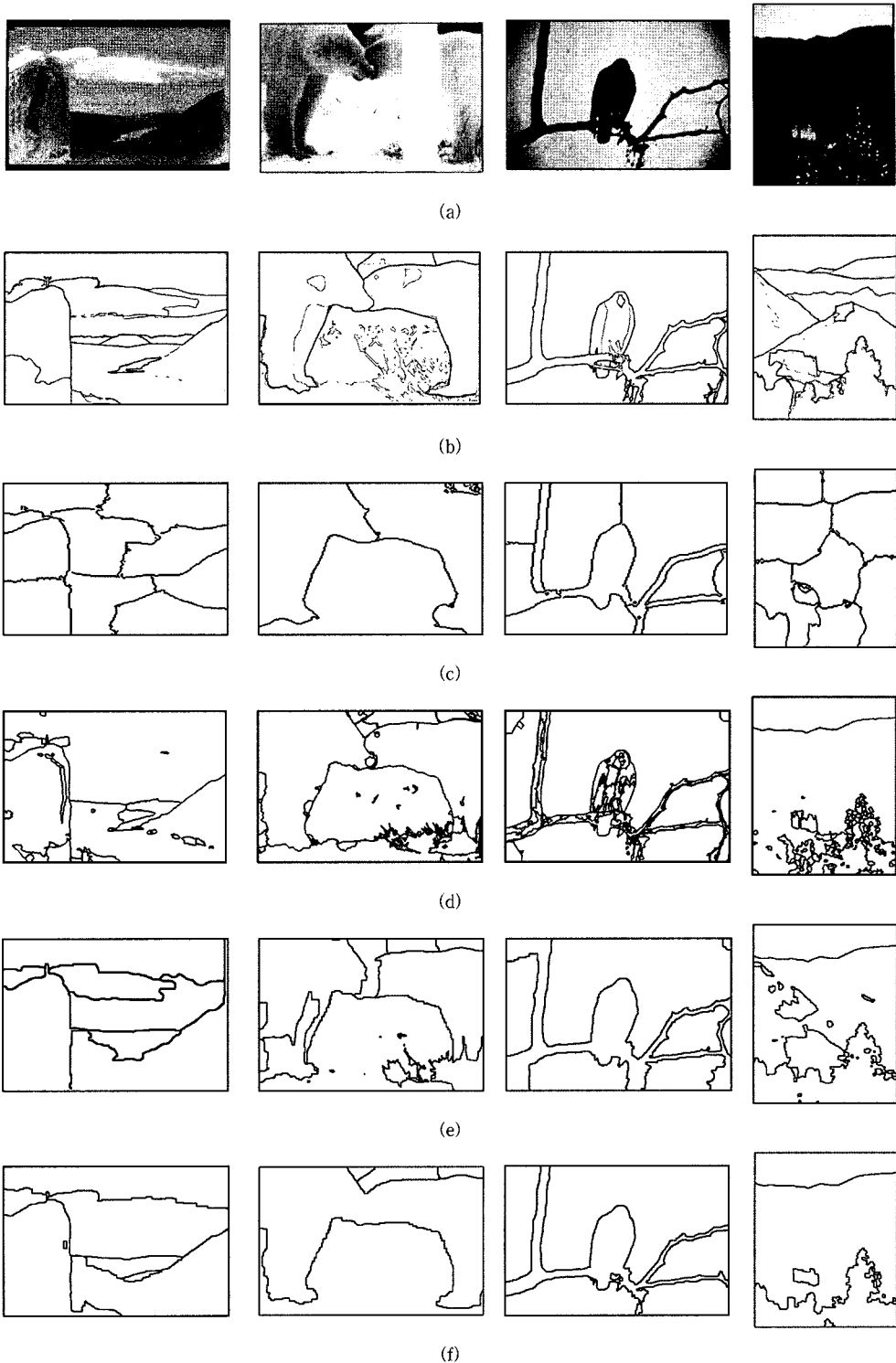


그림 3 외곽선 비교: (a) 입력영상, (b) 사람에 의해 분할된 결과[24], (c) NCut기반의 결과[23], (d) MS기반의 결과 [4], (e) GMM기반의 결과[11], (f) 제안된 방법의 결과

표 2 기존 방법과 제안된 방법의 수량적 평가

	Human[24]	NCut [23]	MS [4]	GMM[11]	제안된 방법
Precision	0.80	0.42	0.46	0.63	0.79
Recall	0.77	0.52	0.68	0.60	0.54
F-Measure	0.79	0.47	0.56	0.61	0.64

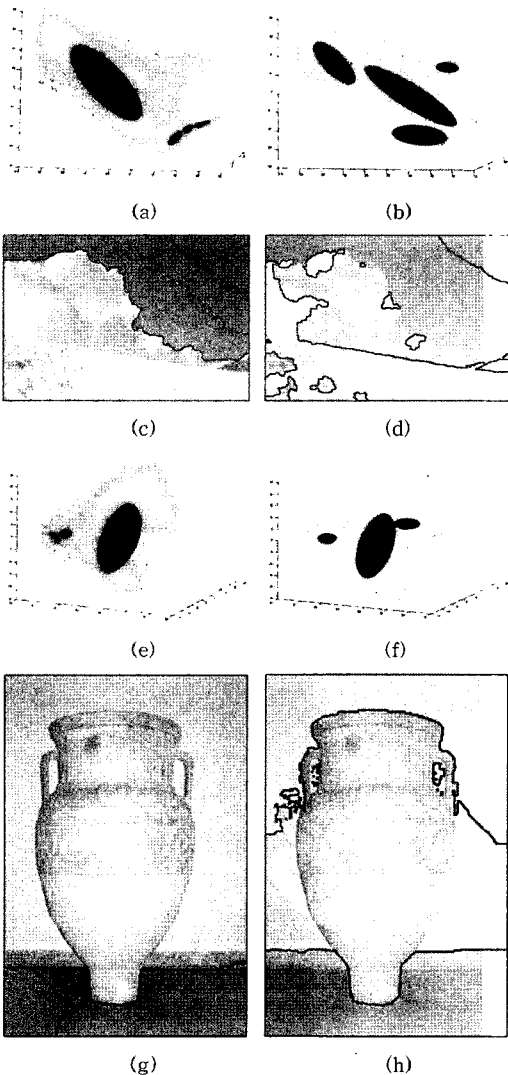


그림 4 GMM기반의 결과: (a,b,e,f) EM에 의해 추정된 매개변수, 평균과 공분산¹⁾, (c,d,g,h) 추정된 매개변수에 의한 결과영상

표 3은 제안된 방법의 각 단계별 수행시간을 나타낸다. 1차원 모멘트를 이용하여 모드를 찾아가는 mean shift단계는 약 1초의 수행시간을 소요되며, 이는 영상분

1) 3차원 각 축은 칼라공간 Luv 를 의미한다.

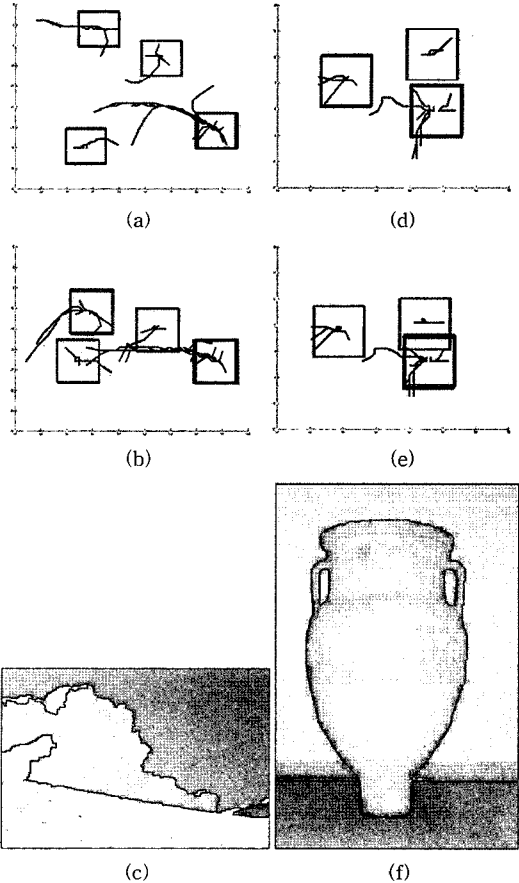


그림 5 그림 4에 대한 제안된 방법의 결과: (a,b,d,e) 특징공간에서 mean의 궤적, (c,f) mean의 궤적을 기반으로 한 결과영상

할을 실시간으로 처리하기에 어려움이 있다. 그러나 기존의 mean shift 방법을 이용할 경우 약 60초 정도의 수행시간을 필요하며, 이와 비교할 때 본 논문에서 이용한 방법이 빠른 수행시간을 보였다. 제안된 방법의 수행시간 중 가장 많은 시간이 소요되는 부분은 다중 레이블링을 위한 α -expansion이며, 이는 그림 1에서 '순환'이 조건을 만족할 때까지 계속 반복되기 때문이다. 표 4는 기존의 3가지 방법에 대한 수행시간을 보여주며, 제안된 방법이 실시간으로 수행하기 어렵지만, 기존의 방법과 비교해서 빠른 수행시간을 보인다. 그림 6은 제안된 방법의 추가적인 결과를 보여준다.

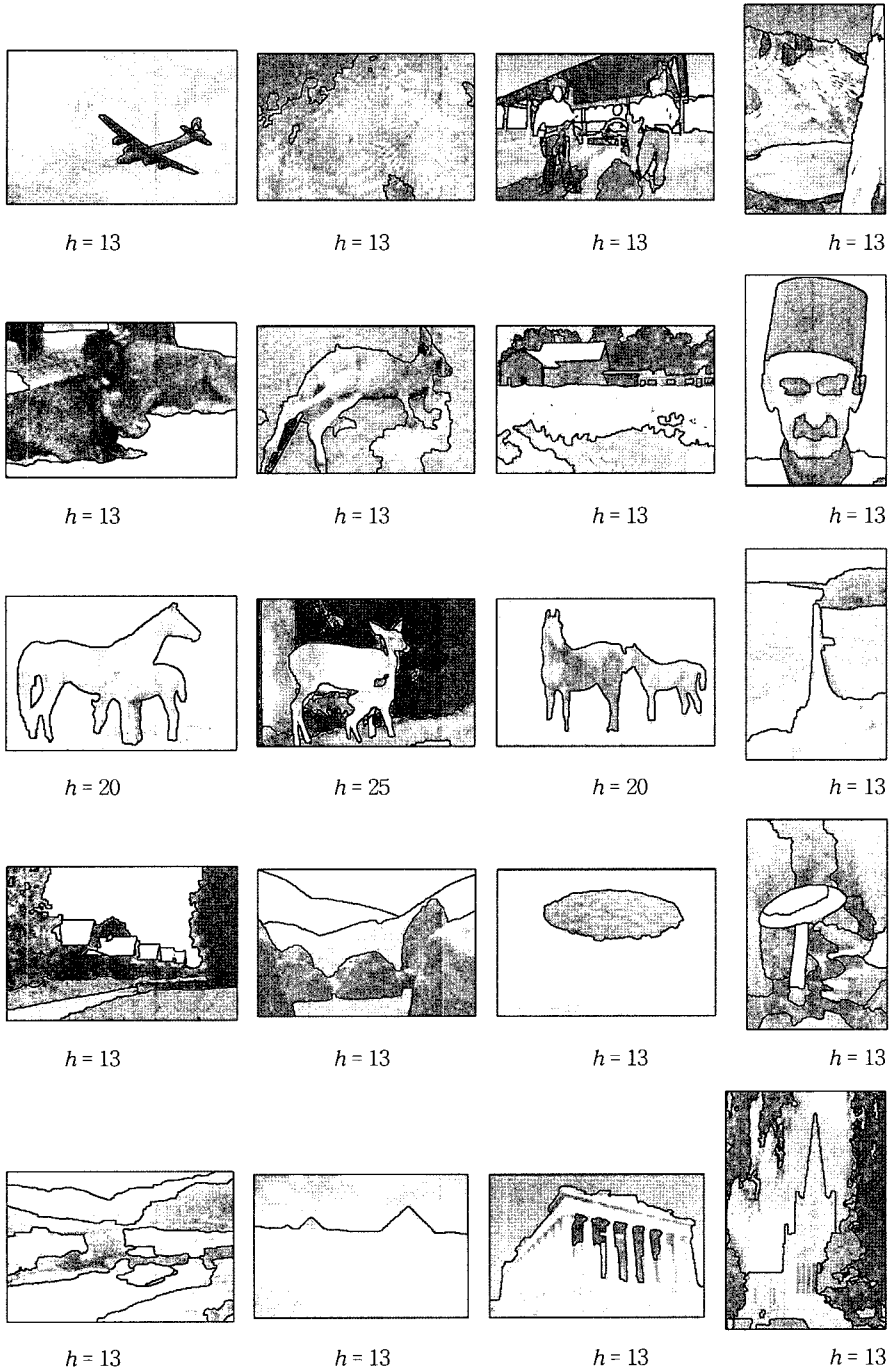


그림 6 제안된 방법의 추가적인 결과

표 3 각 단계별 수행시간

수행과정	Mean Shift	그래프 구축	α -expansion	합
수행시간(sec.)	1.0	0.5	6.0	7.5

표 4 기존의 방법에 대한 수행시간

	<i>NCut</i>	<i>MS</i>	<i>GMM</i>
수행시간(sec.)	55.0	9.5	80.0

5. 결론

그래프 컷 방법은 데이터 항과 스무드 항으로 구성된 에너지 함수를 전역적으로 최소화하는 방법으로 컴퓨터 비전과 이미지 처리에 최근까지 좋은 결과를 보였지만 자동적으로 데이터 항을 위한 사전정보를 획득하기 어렵다. 앞서 GMM을 이용한 그래프 컷 기반의 영상 분할 방법이 일반적으로 사용되고 있지만 특징 공간상에서 클래스의 모양이 초구나 초타원일 때만 좋은 성능을 보이는 단점이 있다. 본 논문에서 다양한 클래스의 모양에서 좋은 성능을 보이기 위해, Mean shift 기반으로 $L^*u^*v^*$ 색상 공간에서 모드를 찾아가는 mean들의 경로를 각 클래스의 사전정보로 사용하는 그래프 컷 기반의 영상 분할방법을 제안했고 제안된 방법은 앞선 세 가지 방법들의 문제점을 극복하였으며 Berkeley Segmentation Dataset을 이용하여 앞선 세 가지 방법보다 좋은 결과를 보여주었다. 그러나 데이터 항을 설정할 때 mean을 이동하기 위한 윈도우 대역너비 h 를 수동으로 설정하는 단점이 있으며, 우리는 h 를 자동으로 설정하기 위한 방법에 대해 연구할 것이다. 뿐만 아니라 그래프 컷을 이용하여 다중 레이블링을 수행할 때의 수행시간을 최소화하기 위한 방법을 연구할 것이다.

참고 문헌

- [1] L.G. Shapiro and G.C. Stockman, "Computer Vision," New Jersey, Prentice-Hall, 2001.
- [2] N. Pal and S. Pal, "A review on image segmentation techniques," *Pattern Recognition*, vol.26, no.9, pp.1277-1294, 1993.
- [3] L. Vincent and P. Soille, "Watersheds in Digital Spaces: An Efficient Algorithm based on Immersion Simulation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.13, no.6, pp.583-597, 1991.
- [4] D. Comaniciu and P. Meer, "Mean Shift: A Robust Approach toward Feature Space Analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.23, no.5, pp.603-619, 2002.
- [5] S. Makrogiannis, G. Economou, and S. Fotopoulos, "A Region Dissimilarity Relation that Combines Feature-Space and Spatial Information for Color Image Segmentation," *IEEE Transactions on System, Man, and Cybernetics*, Part B, vol.35, no.1, pp.44-53, 2005.
- [6] Y. Boykov, O. Veksler, and R. Zabih, "Fast Approximate Energy Minimization via Graph Cuts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.23, no.11, pp.1222-1239, 2001.
- [7] Y. Li, J. Sun, C-K. Tang, and H-Y. Shum, "Lazy Snapping," *ACM Transactions on Graphics*, vol.23, Issue 3, pp.303-308, 2004.
- [8] Y. Boykov and G. Funka-Lea, "Graph Cuts and Efficient N-D Image Segmentation," *International Journal of Computer Vision*, vol.70, no.2, pp.109-131, 2006.
- [9] C. Rother, V. Kolmogorov, and A. Blake, "Grab-Cut: Interactive Foreground Extraction using Iterated Graph Cuts," *ACM Transactions on Graphics*, vol. 23, Issue 3, pp.309-314, 2004.
- [10] A. P. Eriksson, C. Olsson, and F. Kahl, "Image Segmentation with Context," *Proceedings of Scandinavian Conference on Image Analysis, Lecture Notes on Computer Science*, vol.4522, pp. 283-292, 2007.
- [11] J-S. Kim and K-S. Hong, "A New Graph Cut-based Multiple Active Contour Algorithm without Initial Contours and Seed Points," *Machine Vision and Applications*, vol.19, no.3, pp. 181-193, 2008.
- [12] K. Fukunaga and L.D. Hostetler, "The Estimation of the Gradient of a Density Function, with Applications in Pattern Recognition," *IEEE Transactions on Information Theory*, vol.21, pp.32-40, 1975.
- [13] Y. Cheng, "Mean Shift, Mode Seeking, and Clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.17, no.8, pp.790-799, 1995.
- [14] D. Comaniciu, "Nonparametric Robust Methods for Computer Vision," Ph.D. Thesis, ECE Department, Rutgers University, 2000.
- [15] G. Wyszecki and W. S. Stiles, *Color Science : Concepts and Methods, Quantitative Data and Formulae*, 2nd ed. New York : Wiley, 2000.
- [16] Y. Boykov and V. Kolmogorov, "An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization in Vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.26, no.9, pp.1124-1137, 2004.
- [17] R. Szeliski, R. Zabih and C. Rother, "A Comparative Study of Energy Minimization Methods for Markov Random Fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.30, no.6, pp.1068-1080, 2008.
- [18] D. H. Greig, B. T. Porteous, and A. H. Seheult, "Exact Maximum A Posterior Estimation for Binary Images," *Journal of Royal Statistical Society, Series B (Methodological)*, vol.51, no.2, pp.271-278, 1989.
- [19] E. A. Dinic. Algorithm for solution of a problem of maximum flow in networks with power estimation. Soviet math. Dokl., 11:1277-1280, 1970.
- [20] L. Ford and D. Fulkerson, "Flows in Networks," Princeton University Press, 1962.
- [21] Andrew V. Goldberg and Robert E. Tarjan. A New approach to the maximum-flow problem. Journal of the Association for Computing Machi-

- nery, 35(4):921-940, October 1988.
- [22] D. Martin, C. Fowlkes, and J. Malik, "Learning to Detect Natural Image Boundaries using Local Brightness, Color, and Texture Cues," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.26, no.5, pp.530-549, 2004.
- [23] J. Shi and J. Malik, "Normalized Cuts and Image Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.22, no.8, pp. 888-905, 2000.
- [24] D. Martin, C. Fowlkers, D. Tal, and J. Malik, "A Database of Human Segmented Natural Images and Its Application to Evaluating Segmentation Algorithms and Measuring Ecological Statics," *Proceedings of IEEE International Conference on Computer Vision*, pp.416-423, 2001.
- [25] W. Tao, H. Jin, Y. Zhang, "Color Image Segmentation based on Mean Shift and Normalized Cuts," *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, vol.37, no.5, pp.1382-1390, 2007.

박 안 진

정보과학회논문지 : 소프트웨어 및 응용
제 36 권 제 4 호 참조



김 정 환

2008년~현재 숭실대학교 IT대학 미디어
학과 석사과정. 관심분야는 영상처리, 컴
퓨터비전, 패턴인식

정 기 철

정보과학회논문지 : 소프트웨어 및 응용
제 36 권 제 4 호 참조