일반논문-08-13-5-13

영역 대응을 이용한 다시점 영상 집합의 통합 영역화

이 수 찬[a], 권 동 진[a], 윤 일 동[b]‡, 이 상 욱[a]

Joint Segmentation of Multi-View Images by Region Correspondence

Soochahn Lee[a], Dongjin Kwon[a], Il Dong Yun[b]‡, Sang-Uk Lee[a]

요 약

본 논문은 다시점에서 물체를 촬영한 영상들의 집합, 즉, 다시점 영상 집합(multi-view image set)이 주어진 경우, 적은 사용자 입력을 통해 효율적으로 영상 집합 내 관심 물체의 영역을 추출하는 기법을 제안한다. 제안하는 기법은 사용자가 직접 입력을 통해 영역화한 하나의 영상을 바탕으로, 그 영상의 배경 및 전경과 인접 영상 간의 변형을 각각 근사하여 전경 및 배경에 대응되는 인접 영상의 영역을 파악하고, 이 영역들을 통해 인접 영상을 영역화한 후, 영역화된 영상을 바탕으로 다음 인접 영상을 영역화하는 과정을 순차적으로 반복하여 영상 집합 전체를 영역화한다. 이때 전경 및 배경의 변형은 각각 특징점 기반 레지스트레이션(registration) 기법과 선형성·거리비율 보존(affine) 변형을 가정한 대응점 기반 변형행렬(homography)을 통해 근사되며, 각 대응 영역을 기반으로 하는 화소 색 분포 및 형상 정보(shape prior)를 마르코프 랜덤 장(Markov random field)에서의 에너지 최소화에 기반을 둔 영역화 기법에 적용하여 영역화를 수행한다. 제시하는 실험 결과는 제안하는 기법이 적은 사용자 입력으로 다시점 영상 집합 전체를 효과적으로 영역화한다는 것을 뒷받침한다.

Abstract

This paper presents a method to segment the object of interest from a set of multi-view images with minimal user interaction. Specifically, after the user segments an initial image, we first estimate the transformations between foreground and background of the segmented image and the neighboring image, respectively. From these transformations, we obtain regions in the neighboring image that respectively correspond to the foreground and the background of the segmented image. We are then able to segment the neighboring image based on these regions, and iterate this process to segment the whole image set. Transformation of foregrounds are estimated by feature-based registration with free-form deformation, while transformation of backgrounds are estimated by homography constrained to affine transformation. Here, both are based on correspondence point pairs. Segmentation is done by estimating pixel color distributions and defining a shape prior based on the obtained foreground and background regions and applying them to a Markov random field (MRF) energy minimization framework for image segmentation. Experimental results demonstrate the effectiveness of the proposed method.

Keywords: Multi-view images, Image Segmentation, Region Correspondence, MRF Energy minimization, Shape prior

a) 서울대학교 전기·컴퓨터공학부
School of Electrical Engineering and Computer Science, Seoul National Univ.
b) 한국외국어대학교 용인캠퍼스 디지털정보공학과
School of Digital Information Engineering, Hankuk University of Foreign Studies
‡ 교신저자 : 윤일동(yun@hufs.ac.kr)
※ 이 논문은 2008년도 정부(과학기술부)의 재원으로 한국과학재단의 지원을 받아 수행된 연구임 (No. R01-2007-000-11425-0)

Ⅰ. 서 론

관련 있는 영역들로 영상을 분할하는 영역화(image segmentation)는 컴퓨터 비젼과 그래픽스 분야에서 오랫동안 연구되어온 주제로 폭넓은 응용 분야를 가진다. 인식과 3차

원 구조 복원 등에서는 전처리 과정으로 이용되며, 배경 대체를 비롯한 영상 합성과 가상현실 등 많은 문제에 필수적인 중간 과정의 역할을 하기도 한다.

단일 영상의 영역화 기법은 컴퓨터 비젼(Computer Vision)의 태동기부터 연구되어온 고전적인 문제로, 지난 십년간 특히 많은 수의 기법들이 제안되었다. 각 기법들은 위치와 색만을 고려하여 유사한 특성을 가진 화소들로 이루어진 영역들을 묶고자 하는지, 또는 특정 물체로 이루어진 영역을 구분하고자 하는지에 따라 두 가지로 분류될 수 있다. 화소 특성에 따른 영역화 기법들[4, 15, 19]은 일반적으로 사람들의 직관과 다른 결과를 주지만, 관계가 긴밀한 화소들을 하나로 묶어 추가 조작을 용이하게 한다. 반면 영상 내의 물체별로 영역화하는 기법들[1, 8, 13]은 대부분 데이터베이스에 기반을 둔 학습 기법이나 사용자의 직접 입력 등을 통한 고차원(high level) 정보를 필요로 한다. 특히 마르코프 랜덤 장(Markov Random Field, 이하 MRF)에서의 에너지 모델과 그래프 컷을 통한 에너지 최소화 기법을 이용한 Boykov와 Jolly의 기법[1]은 간단한 사용자 입력을 기반으로 뛰어난 결과를 도출하여 다른 기법들[8, 13]을 낳는 시발점이 되었다. Rother 등의 기법[13]은 흑백 이미지에 적용되던 이 기법[1]을 컬러로 확장하였으며, 초기 사용자의 입력을 기반으로 입력 영역 지정, 전경과 배경의 컬러 분포 생성, 그리고 영역화를 자동적으로 반복 수행하여 더욱 개선된 결과를 제시하였다.

다양한 단일 영상 영역화 기법들이 등장하면서 이를 응용한 다양한 비디오 영상 영역화 기법들도 제안되었다[20, 10, 18, 5, 21]. Boykov와 Jolly의 기법[1]은 2차원 단일 영상 뿐 아니라 비디오 데이터를 3차원으로 취급하여 뛰어난 영역화 결과를 제시하였다. Wang 등의 기법[20]과 Li 등의 기법[10] 등은 이를 발전시켜, 화소 특성에 따른 영역화 기법[4, 19]을 전처리 과정으로, 전체 비디오를 조작할 수 있는 세련된 사용자 입력체계(interface)를 제공하였다. 또한 웹캠 영상에 실시간으로 동작하는 기법들[18, 5, 21]도 제안되었는데, 이들은 카메라가 거의 움직이지 않는다는 가정 하에 영상의 특성을 학습하여 영역화를 수행하였다. 이 기법들은 영상 프레임간의 변형을 단순화한다는 공통점을 지니는데, 이 때문에 시점이 급격히 변하는 경우에 적용이 어렵다.

본 논문에서는 하나의 물체를 다양한 시점에서 촬영한 다시점 영상 집합(multi-view image set)을 전경 물체와 배경으로 영역화하는 문제를 다룬다. 다시점 영상 집합은 단일 영상과 비디오 영상의 중간적인 성격을 띠며, 일반 카메라 사용자가 같은 장소에서 여러 장의 영상을 촬영한 경우나 다시점 스테레오(multi-view stereo) 기법을 통한 물체의 3차원 구조 복원의 입력 데이터를 위해 다시점에서 그 물체를 촬영한 경우 생성된다. 이때, 일반 사용자에 의해 촬영된 영상들을 일괄적으로 편집하고자 하는 경우나 다시점 스테레오 기법의 정밀도를 높이기 위한 경우 영상 집합을 편리하게 영역화하는 기법을 필요로 하게 된다.

단일 영상이나 비디오 영역화 기법들에 비해 다시점 영상 집합의 영역화 기법에 대해서는 많은 연구가 이루어지지 않았다. Sormann 등의 기법[16]은 영상들 간의 대응점을 이용하여 영상들의 시점 정보를 획득하고, 이를 바탕으로 사용자가 직접 영역화한 영상의 경계를 인접 영상으로 전파한다는 점에서 본 논문에서 제안하는 기법과 유사하다. 하지만, 시점 정보를 포함한 카메라 정보(calibration)를 근사한 후 경계를 인접 영상의 동일면선상(epipolar line) 위의 화소 중 변형(gradient)가 가장 큰 지점으로 가정함으로써, 전경의 변형이 적고 배경이 복잡(cluttered)하지 않은 경우에 한해 작동한다는 단점이 있다. Sormann 등의 두 번째 기법[17]의 기법은 거리 변환(distance transform)을 통한 경계 전파 정보를 추가하였을 뿐, Wang 등의 비디오 영역화 기법[20]의 사용자 입력체계를 적용하여 영상 집합을 비디오 영상과 동일하게 취급하기 때문에 시점 변형이 다양한 영상 집합에 적용하기 어렵다. Campbell 등의 기법[3]은 카메라에 대한 정보가 주어졌다고 가정하고, 영역화와 전경 물체의 3차원 구조 복원의 과정을 반복적으로 순환하여 최종적인 영역화 결과를 구한다. 이 기법은 카메라 정보가 없는 경우에 적용할 수 없으며, 3차원 구조 복원 과정을 거치므로 계산이 복잡하다는 단점이 있다.

본 논문에서 제안하는 기법은 Sormann 등의 기법[16]과 유사하게 사용자가 입력을 제공하여 영역화한 영상의 정보를 전체 영상 집합으로 전파하여 전체 영상 집합을 영역화한다. 그러나 기존 기법과는 달리 특징점 기반 레지스트레이션(feature-based registration) 기법[9]과 다시점 기하학

(multiple view geometry)[7]을 이용하여 전경 및 배경과 인접 영상간의 변형을 분리하여 근사하고, 이를 통해 인접 영상에서 전경 및 배경에 대응되는 영역을 찾아서 영역화를 효과적으로 수행한다. 2장에서는 제안하는 기법을 개관하며, 3장에서는 각 과정을 자세히 설명한다. 4장에서는 다양한 실험 결과를 제시하며 5장의 결론을 통해 마무리 짓는다.

II. 제안하는 기법에 대한 개관

그랩컷(GrabCut) 기법[13]을 비롯한 그래프 컷 기반 영역화 기법들[1, 8, 13]은 각 화소마다 배경 화소 또는 전경 화소로 지정(label)될 확률과 마르코프 랜덤 장(MRF)을 바탕으로 이미지 전체에 대해 결합 확률(joint probability)을 극대화하는, 또는 에너지를 최소화하는(energy minimization), 화소들의 상태를 찾음으로써 영역화를 하게 되는데, 이때 각 화소가 어떤 영역으로 지정될지에 대한 확률은 전경 및 배경의 화소 색 분포와 인접 화소와의 색 밝기 차에 의해 결정된다. 이 색 분포에 대한 정보를 제공하기 위해 사용자는 전경 및 배경의 일부 영역을 지정하고, 이를 바탕으로 전경 및 배경의 화소 색 분포가 근사되며, 지정된 화소들과 색 차이가 적은 주변 화소들이 같은 영역으로 지정되는 등 사용자의 입력을 통해 영역화의 기준이 마련되는 것이다. 이와 같이 기존의 그랩컷(GrabCut) 기법[13]은 사용자 입력 정보를 바탕으로 하는데, 다시점 영상 집합은 영상들이 유사한 특성을 가지므로, 제안하는 기법은 사용자 입력을 통해 하나의 영상에 대한 정보를 인접 영상들로 전파하고 이를 각 영상마다 그랩컷 기법에 대한 입력으로 하여 다중 시점 영상 집합 전체를 영역화하는 것이다. 여기에서 인접 영상들은 인접한 시점에서 촬영된 영상들을 의미하며, 이 때문에 영상 집합은 시점의 흐름에 따라 촬영되고 촬영 순서대로 배열되어 있어야 한다는 가정을 한다.

전경 및 배경 영역이 각각 어떤 색의 화소로 이루어졌는지를 나타내는 색 분포와 어떤 화소가 실제 전경 및 배경인지에 대한 정보가 정확하고 풍부할수록 그랩컷 기법을 통한 영역화의 결과가 정확해진다. 이를 위해 제안하는 기법은 영역화된 초기 영상과 인접 영상간의 변형을 근사하여 인접 영상의 어떤 부분이 초기 영상의 전경 및 배경에 대응되는지를 근사한 후, 각 영역을 기반으로 인접 영상의 화소 색 분포를 계산하고 전경 및 배경을 영역화한다. 세부적으로, 특징점 정합(feature point matching)을 기반으로 하여 종래의 다중 시점 기하학[7]과 영상 레지스트레이션 기법[9]을 적용하여 두 영상 사이의 변형을 근사하는데, 이때 배경은 독립적인 움직임이 없다는 가정 하에 변형 행렬(homography)로 변형을 근사하고, 전경은 시점의 변형과 독립적인 물체 자체의 움직임 등 다양한 변형을 효과적으로 표현할 수 있는 레지스트레이션 기법을 적용하여 변형을 근사한다. 이런 과정을 통해 예상되는 인접 영상의 전경 및 배경

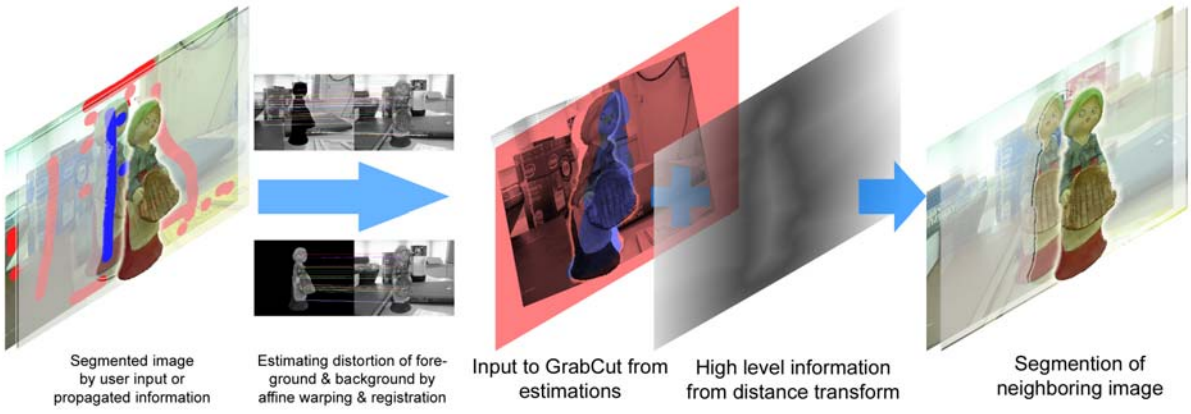


그림 1. 제안하는 기법에 대한 개관 및 흐름도
Fig. 1. Overview of the proposed method

영역을 기반으로 화소 색 분포를 근사하고, 여기에 예상되는 전경 경계의 거리 변환[6]을 형상 정보(shape prior)[8]로 이용하여 인접 영상의 영역화를 수행하게 된다. 영역화된 인접 영상을 바탕으로 다시 그 다음 인접 영상을 영역화하는 등 이 과정을 반복 수행하여 전체 영상의 영역화하게 되는 것이다. 그림 1은 이와 같은 전체 과정을 한 눈으로 정리한 그림이다.

III. 각 과정에 대한 세부 설명

1. 그랩컷(GrabCut)을 이용한 영역화 및 MRF(Markov Random Field) 에너지 모델

그랩컷 기법을 비롯하여 그래프 컷을 바탕으로 하는 영역화 기법들은 모두 하나의 화소 p를 노드(node)로, 인접 화소와의 관계를 에지(edge)로 보는 화소 쌍(pairwise) MRF를 생성하고, 이를 바탕으로 노드 집합 P의 각 화소가 해당하는 영역을 나타내는 레이블 $\mathbb{L}=(l_1,\cdots,l_p,\cdots,l_{|P|})$을 각 노드에 할당하는 문제로 영역화 문제를 치환하여 접근한다. 이때 레이블 할당은 할당된 레이블에 따른 에너지를 최소화시키는 형태로 이루어지며, 에너지는 다음과 같다:

$$E(\mathbb{L}) = R(\mathbb{L}) + \lambda B(\mathbb{L}) \tag{1}$$

여기에서 $R(\mathbb{L})$와 $B(\mathbb{L})$는 각각 에너지의 영역 항(region term)과 경계 항(boundary term)을 나타내는데, 영역 항은 각 화소가 전경 또는 배경에 속할 때의 에너지를 의미하며, 경계 항은 특정 지점에 경계가 위치하는 에너지를 의미한다. 각 항에 대한 구체적인 식은 다음과 같다:

$$\begin{aligned} R(\mathbb{L}) &= \sum_{p \in P} R_p(l_p), \\ B(\mathbb{L}) &= \sum_{\{p,q\} \in N} (B_{g,\{p,q\}} + \lambda_B B_{d,\{p,q\}}) \cdot \delta(l_p, l_q). \end{aligned} \tag{2}$$

이때, N은 MRF 상에서의 모든 인접 화소 쌍으로 이루어진 집합을 의미하며, $\delta(l_p,l_q)=\begin{cases}1, & \text{if } l_p \neq l_q \\ 0, & otherwise\end{cases}$ 이다. $R_p(l_p)$는 화소 p에 레이블 l_p를 할당할 때의 에너지로, p의 색 또는 밝기 값 I_p에 따라 근사된 전경 O 및 배경B의 화소 색 분포를 이용하여 계산되며, 다음의 수식으로 표현된다:

$$\begin{aligned} R_p(l_p = O) &= -\ln(Pr(I_p) \mid l_p = O)), \\ R_p(l_p = B) &= -\ln(Pr(I_p) \mid l_p = B)). \end{aligned} \tag{3}$$

경계 항은 인접 화소들의 색 또는 밝기 값의 차이에 의해 결정되는 $B_{g,\{p,q\}}$와 전파된 예상 경계의 거리 변환에 의해 결정되는 $B_{d,\{p,q\}}$로 이루어진다.

$$\begin{aligned} B_{g,\{p,q\}} &= dis(p,q) \cdot (exp(-\beta \mid I_p - I_q \mid), \\ \beta &= (2\langle \| I_p - I_q \| \rangle)^{-1} \end{aligned} \tag{4}$$

로 $dis(p,q)$는 화소 p와 q의 거리를, $|x|$는 벡터의 L_2 놈(norm)을, 그리고 $\langle \cdot \rangle$는 전체 영상 내의 평균을 의미하며, 영역 간의 경계는 화소 값의 차이가 큰 곳에 위치한다는 가정 하에 두 화소 값의 차이가 클수록 작은 값을 가지게 된다. B_d는 기존의 그랩컷 기법에서 적용되지 않았으나 형상 정보[8]를 사용하기 위해 포함되며, 자세한 정의는 Ⅲ.3절에 주어진다.

각 노드의 레이블을 도출하는 데에는 그랩컷 기법[13]을 이용한다. 그랩컷 기법은 그래프 컷을 이용한 영역화[1]를 통해 계산된 에너지를 최소화하는 레이블을 찾는 과정과, 찾은 레이블에 따른 결과 영역을 이용한 전경 및 배경의 화소 색 분포 근사를 반복 수행하여, 초기 입력 영역에 오차가 있는 경우에도 강인하게 작동한다. 이때 각 세부 단계에서 계산된 에너지를 최소화하는 레이블을 찾는데에는 그래프 컷을 통한 최소 분할/최대 흐름(min-cut/max-flow)을 찾는 알고리즘[2]을 통하여 계산된다.

2. 인접 이미지에서의 배경 및 전경의 변형 근사

제안하는 기법은 영역화된 현 영상의 전경 및 배경에 대응되는 인접 영상의 영역을 찾고 이를 인접 영상의 영역화에 필요한 사용자의 입력 대신에 활용하고자 하는 것이 핵심으

로, 이에 따라 현 영상에서 인접 영상으로 전경과 배경이 각각 어떻게 변형했는지 근사하게 된다. 변형의 근사는 각 영상에서 추출한 특징점을 기반으로 이루어지는데, 대응점의 오정합(outlier)을 줄이고, 전경 물체의 자체 움직임이 있는 경우를 반영하기 위해 전경과 배경을 분리하여 변형을 근사한다. 이때 대응점을 찾기 위한 특징점의 추출, 서술자의 생성, 그리고 정합은 크기 불변 특징 변환(Scale Invariant Feature Transform, 이하 SIFT) 기법[11]을 이용한다.

배경의 변형은 선형성·거리비율 보존(affine) 변형을 가정하여 변형 행렬을 이용한 선형 변형으로 근사된다. 세부적으로, 먼저 현 영상의 배경 영역과 정보를 전파받을 인접 영상의 특징점을 추출하여 서로 가장 가까운 네이버 정합(nearest neighbor matching)과 랜덤 샘플 합의 기법(RANdom SAmple Consensus, 이하 RANSAC)을 이용하여 정합한 후, 대응되는 특징점을 이용하여 변형 행렬을 계산한다[7].

전경의 변형은 특징점을 이용한 레지스트레이션 기법을 이용하여 근사되는데, 이 기법은 다음과 같이 요약된다. 우선 영상을 변형하여 목표 영상으로 레지스트레이션을 하기 위한 기본 변형 모델로 큐빅 B-스플라인(cubic B-spline)을 기반으로 한 일정한 그물망(regular mesh)의 자유 형태 변형(free-form deformation) 모델을 이용한다. 구체적으로, 영상의 크기가 높이 h, 너비 w인 경우, $\Omega = \{(x,y) \mid 0 < x < w, 0 < y < h\}$ 의 영역을 각 축에 따라 N개의 구간으로 나누어 $\delta_x = w/N$, $\delta_y = h/N$의 간격으로 벌어진 일정한 점으로 이루어진 일정한 그물망 Φ의 변화에 대한 에너지는

$$E(S, C, r) = \lambda_R E_D(S) + E_C(S, C, r) \tag{5}$$

으로 정의된다. 여기에서 S는 Φ의 변화 상태를 나타내는 벡터, C는 SIFT를 통해 추출된 대응점의 집합, $E_D(S)$는 Φ의 점들의 변형 에너지, 그리고 $E_C(S, C, r)$는 대응관계로 나타나는 에너지를 의미한다. 각 항은

$$E_D(S) = \sum_{(i,j,k) \in L} (-x_i + 2x_j - x_k)^2 + (-y_i + 2y_j - y_k)^2 \tag{6}$$

으로 i, j, k는 Φ의 점들에 대한 색인(index)을, L은 연속되는 세 점들의 색인의 집합을 나타내며,

$$E_C(S, C, r) = -\sum_{c \in C} w_c \rho(d, r) \tag{7}$$

으로

$$\rho(d,r) = \begin{cases} \dfrac{3(r^2 - d^2)}{4r^3} & d < r \\ 0 & otherwise \end{cases}, \tag{8}$$

$d = \| c_r - T_S(c_i) \|$, ρ은 강인한 측정기(robust estimator), 그리고 r은 강인한 측정기의 입력으로 대입되는 신뢰반경(confidence radius)[12]을 의미한다. 이를 자세히 설명하면, $E_D(S)$는 그물망 Φ의 점들의 위치가 변화한 형태 S에 따른 에너지로, 직선상에 위치하는 세 점이 변형된 후 직선과 어긋날수록 값이 증가하여 기존의 일정한 그물망과 다르게 일정하지 않을수록 큰 값을 가지며, $E_C(S, C, r)$는 영역화된 전경에서 추출된 특징점 c_i가 그물망 Φ와 변화 상태 S에 따른 변환 T_s에 의해 변환된 점 $T_s(c_i)$와 인접 영상에서 추출된 c_i의 대응점 c_r의 차이를 강인한 측정기에 대입한 값에 의해 결정된다. 결국 그물망의 변환 S으로 근사된 변환에 의해 초기 영상을 변환시킨 후 변환된 영상과 인접 영상의 각 대응점이 서로 같은 점이 되도록 하면서도, 그물망의 각 점들이 일정한 그물망으로부터 최소로 변화된 변환 S가 에너지를 최소화하는 변환이 되는 것이다. 이때 $T_s(c_i)$는 B-스플라인을 이용한 보간(interpolation)을 통해 계산되며 구체적인 식은 다음과 같다:

$$T_S(x) = \sum_{l=0}^{3} \sum_{m=0}^{3} B_l(u) B_m(v) \phi^S_{i+l, j+m}, \tag{9}$$

여기에서 $\lfloor \cdot \rfloor$는 버림 함수(floor function)를, $i = \lfloor x/\delta_x \rfloor + 1$, $j = \lfloor y/\delta_y \rfloor + 1$, $u = x/\delta_x - \lfloor x/\delta_x \rfloor$, $v = y/\delta_y - \lfloor y/\delta_y \rfloor$, 그리고 B_l과 B_m은 각각 l과 m차 큐빅 B-스플라인의 기저(basis) 함수이다. 위의 수식들을 기반으로 에너지를 최소화하는 상

태 변환 S를 구하는 데에는 비선형 켤레 그레디언트(non-linear conjugate gradient) 기법을 이용한다[14].

3. 전파된 정보의 적용

제안하는 기법에서는 영역화된 영상의 전경 및 배경에 대응되는 인접 영상의 영역을 영상 간 변형을 통해 근사한 후, 두 가지 방식으로 이용한다. 전경 및 배경의 화소 색 분포를 근사하기 위한 표본으로 이용하며, 근사된 전경의 경계와 영역화를 통해 구한 경계가 가깝도록 영향을 끼치는 형상 정보로 이용한다.

변형 근사를 통해 획득한 인접 영상의 예상 전경 및 배경 영역의 화소를 표본으로 화소 색 분포를 근사할 때에는, 화소들 중 전경이자 배경으로 중복되어 지정된 화소들과, 전경 화소로 근사되었으나 이전 영상의 화소 색 분포에 의해 배경에 속할 가능성이 높은 화소들 및 그 반대 경우의 화소들을 제외하고 근사한다. 이때, 근사된 전경 및 배경의 영역에 오차가 발생할 수 있으며, 인접한 영상은 유사한 색 분포를 가진다고 가정할 수 있으므로, 새로 계산된 인접 영상 색 분포와 영역화된 이전 영상 색 분포의 가중 평균을 이용한다. 세부적으로, 영역화된 배경 영역의 화소 수를 $n(seg_B)$, 재근사에 이용되는 화소의 수를 $n(new_B)$라 하면, 재근사된 전경과 영역화된 이전 영상의 전경의 가중치로 각각 $c_F=\begin{cases} 0.7 & \text{if } \frac{n(new_B)}{n(seg_B)} > 0.7 \\ \frac{n(new_B)}{n(seg_B)} & otherwise \\ 0.3 & \text{if } \frac{n(new_B)}{n(seg_B)} < 0.3 \end{cases}$, $1-c_F$을 이용한다. 배경에 대해서도 이와 같은 방법으로 가중치를 계산한다.

전경 영역의 경우 적용하는 레지스트레이션 기법을 통해 상당히 정확하게 근사된다. 이에 따라 근사된 전경의 경계를 실제 전경의 경계로 예상할 수 있게 된다. 이와 같은 형상 정보를 바탕으로 Ⅲ.1절에 언급한 경계 항의 식을

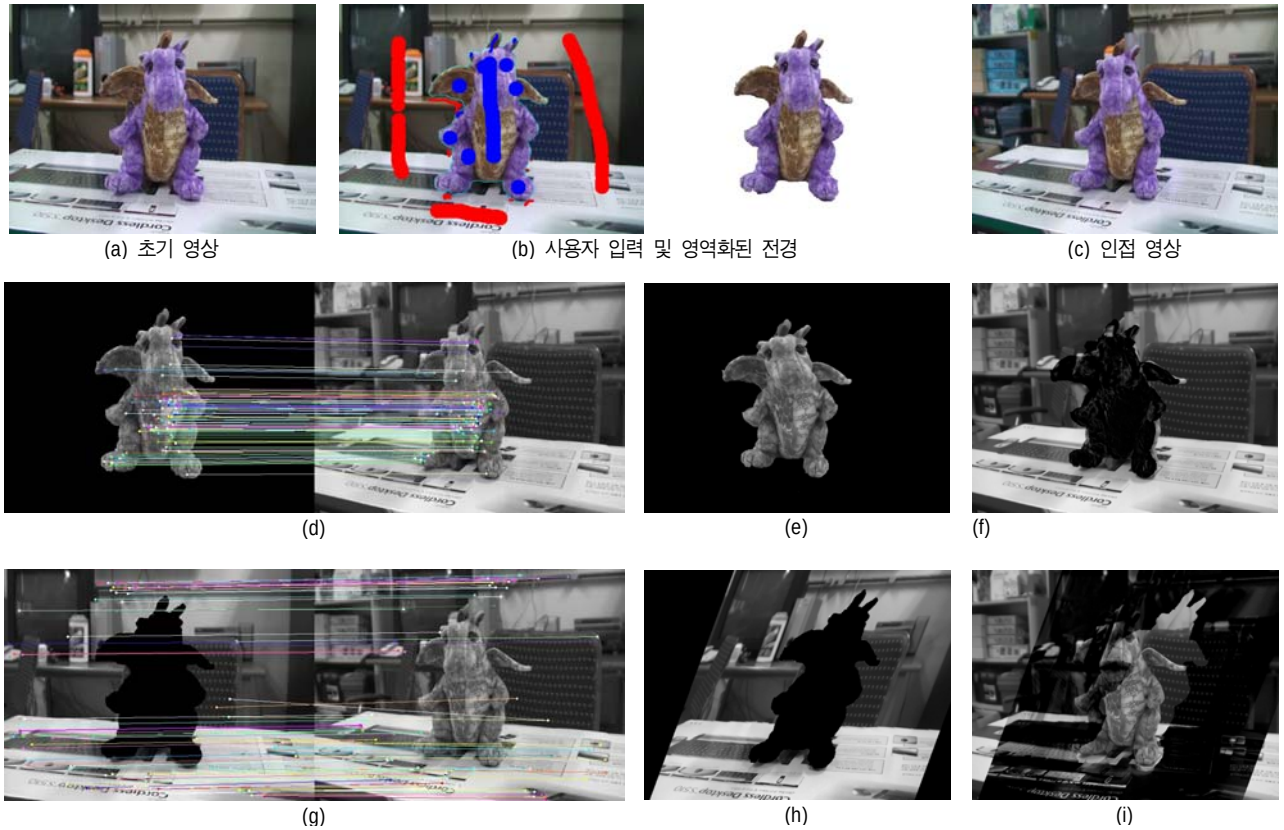
$$B_{d,\{p,q\}} = -\ln\left(\frac{1}{1+\lambda_{d1}\cdot exp(\lambda_{d2}\cdot\frac{dt(p)+dt(q)}{2})}\right) \quad (10)$$

으로 정의할 수 있다. 이때, $dt(p)$는 가장 가까운 예상 경계 지점으로부터 p까지의 거리를 의미하여 예상 경계 지점과 멀리 떨어진 지점에 영역화 경계가 있으면 에너지가 커지게 되는 것이다.

Ⅳ. 실험 결과

본 장에서는 제안하는 기법의 실험 결과를 제시한다. 우선 그림 2는 영역화된 영상의 전경 및 배경과 인접 영상과의 대응점을 찾고, 대응점들을 바탕으로 변형을 근사하는 과정을 나타낸다. 세부적으로, 그림 2(a), (c)와 같은 초기 영상과 인접 영상이 주어지고 사용자가 초기 영상을 그림 2(b)와 같이 영역화한 경우, 그림 2(d), (g)와 같이 영역화된 초기 영상의 전경 및 배경과 인접 영상의 대응점을 구하고, 구한 대응점을 바탕으로 그림 2(e), (h)와 같이 레지스트레이션 및 변형행렬을 통하여 전경 및 배경의 변형이 각각 근사되는 것이다. 그림 2(f), (i)는 변형된 배경 및 전경과 인접 영상의 차를 각각 나타내는데, 각 영역 간 근사된 변형의 오차를 보여준다. 그림 3은 대응점 정보의 부족(그림 3(a))으로 인하여 전경의 변형이 부정확하게 근사된 경우(그림 3(b))를 나타내는데, 제안하는 기법은 잘못 근사된 영역을 그대로 이용하지 않고 색 분포 근사 및 형상 정보 등 간접적으로 이용하기 때문에 이와 같이 영상 간 변형이 잘못 예측된 경우에도 강인한 영역화 결과를 도출한다(그림 5(b)). 이는 배경의 변형이 잘못 근사된 경우에도 유사하였다.

그림 4는 예상 경계와의 거리를 이용한 형상 정보를 적용하지 않은 결과와 적용한 결과를 비교하여 나타내며, 이는 거리 정보를 사용하는 경우에 더욱 정밀한 결과를 얻을 수 있다는 것을 나타낸다. 그림 5, 6, 7은 제시하는 기법을 통해 각각 공룡, 키든아낙, 오줌싸개소년 인형을 촬영한 영상 집합을 영역화한 결과를 제시한다. 모든 집합에서는 단 한 장의 영상만 사용자에 의해 영역화 되었으며, 각 집합은 각각 24장, 26장, 그리고 23장으로 이루어져 있다. 대부분의 물체 영역이 정확하게 검출된다는 것을 알 수 있으나, 일부 영역에서 물체의 영역이 잘못 지정된 경우를 발견할 수 있



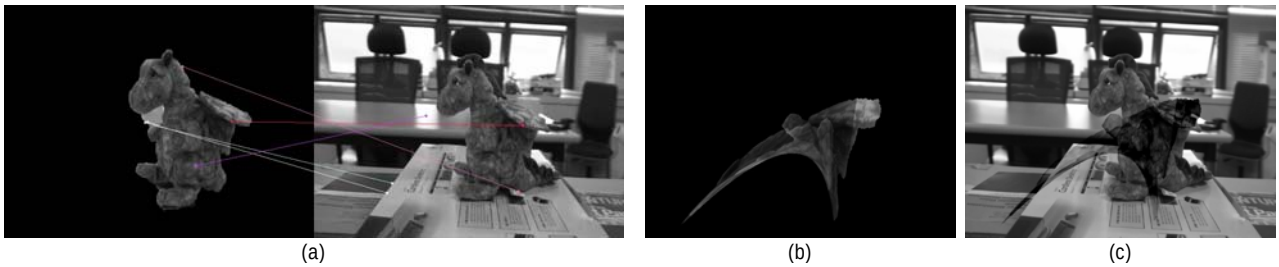
(a) 초기 영상 (b) 사용자 입력 및 영역화된 전경 (c) 인접 영상

(d) (e) (f)

(g) (h) (i)

그림 2. 근사된 영상간 변형의 예. (a) 초기 영상, (b) 사용자 입력 및 영역화된 전경, (c) 인접 영상, (d), (g) 전경 및 배경의 대응점, (e), (h) 근사된 변형, (f), (i) 변형된 전경 및 배경과 인접영상의 차이

Fig. 2. An example of estimated distortion. (a) Initial image, (b) user input and segmented foreground, (c) neighbor image, (d), (g) correspondence points of foreground, background, (e), (h) estimated distortion, (f), (i) Difference between distorted foreground, background and neighbor image

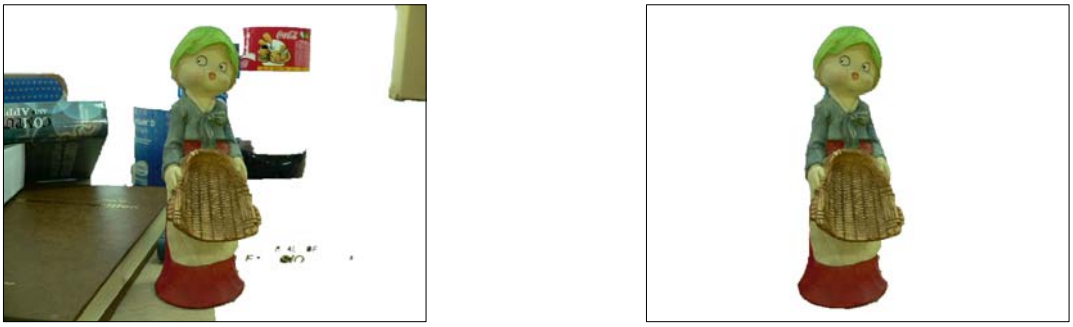


(a) (b) (c)

그림 3. 영상간 변형 근사가 실패한 예(공룡 집합 7번 프레임). (a) 영역화된 전경과 인접 영상의 대응점, (b) 근사된 변형, (c) 근사된 변형과 인접영상의 차이

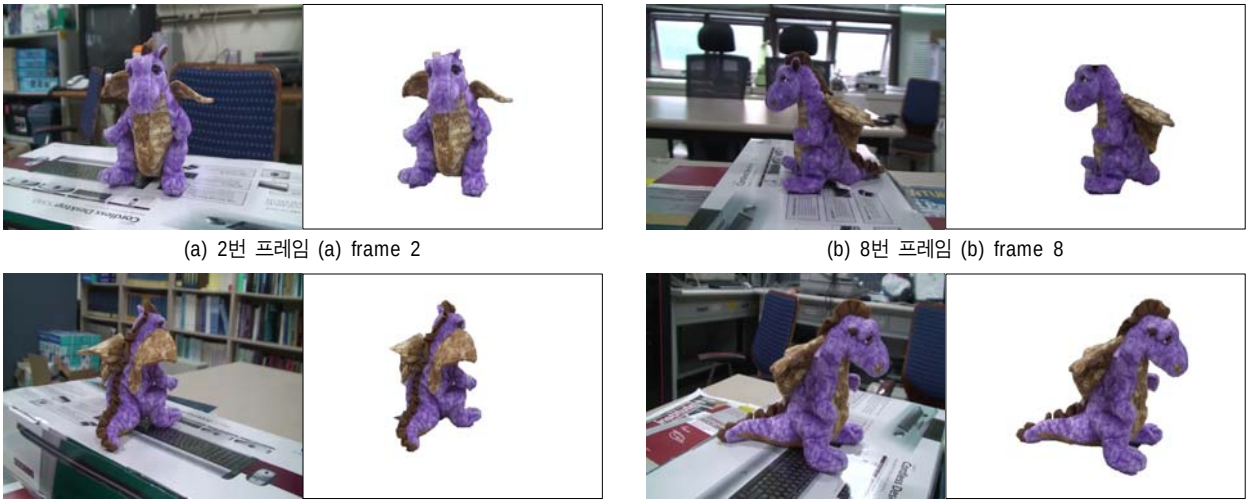
Fig. 3. An example where distortion estimation fails (frame 7 of DINO set). (a) Correspondence points of segemented foreground and the neighboring image, (b) Estimated distortion, (c) Difference between distorted foreground the neighboring image

다. 우선, 그림 5(b)의 경우 이전 영상에서 꼬리가 가리워져 있던 관계로 새롭게 시야로 들어온 공룡의 꼬리 부분이 유실되었고, 모양이 뾰족하고 배경과 색 차이가 적은 귀와 갈기 부분 또한 유실되었다. 또한, 그림 6(a)에서 오줌싸



(a) (b)

그림 4. (a) 형상 정보를 이용하지 않은 경우와, (b) 형상 정보를 이용한 경우의 영역화 결과
Fig. 4. Segmentation results (a) without shape prior, and (b) with shape prior



(a) 2번 프레임 (a) frame 2 (b) 8번 프레임 (b) frame 8

(c) 16번 프레임 (c) frame 16 (d) 21번 프레임 (d) frame 21

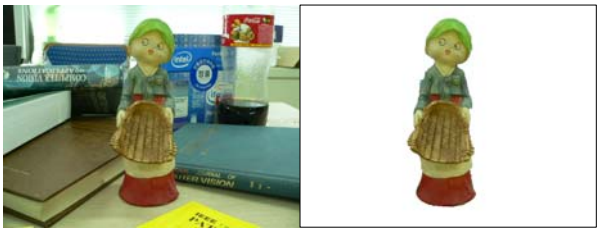
그림 5. 공룡 영상 집합의 영역화 결과
Fig. 5. Segmentation results of the DINO image set

개소년의 머리 부분에서 배경과 전경의 경계가 약해지면서 머리의 일부가 유실된 것을 알 수 있다. 앞으로 이와 같은 문제들을 해결하는데 추가적인 연구가 요구될 것으로 보인다.

제시된 결과에서 화소 확률 분포는 RGB 공간에서의 3차원 히스토그램으로 근사되었으며, 모든 실험 결과에서는 $\lambda_{d1} = 0.3$, $\lambda_{d2} = 0.1$ 이었으며, 공룡 영상 집합은 $\lambda = 3$, $\lambda_B = 0.167$, 키든아낙 집합은 $\lambda = 5$, $\lambda_B = 0.1$, 오줌싸개소년 집합은 $\lambda = 3$, $\lambda_B = 0.5$ 를 이용하였다.

V. 결 론

본 논문은 적은 양의 사용자 입력을 기반으로 물체의 다시점 영상 집합(multi-view image set)을 효율적으로 영역화하는 기법을 제안한다. 구체적으로, 사용자의 입력을 통해 영역화된 영상의 전경과 배경을 분리하여 인접 영상과의 대응점을 찾고, 이를 바탕으로 영역화된 영상의 전경 및 배경이 인접 영상에서 어떻게 변형되었는지를 근사하여 인접 영상에서 전경 및 배경 영역으로 예상되는 영역을 찾는다. 이를 통해 각 영역의 색 분포와 경계의 영역을 근사하고, 이를



(a) 3번 프레임 (a) frame 3

(b) 9번 프레임 (b) frame 9

(c) 15번 프레임 (c) frame 15

(d) 21번 프레임 (d) frame 21

그림 6. 키든아낙 영상 집합의 영역화 결과
Fig. 6. Segmentation results of the LADY image set

(a) 3번 프레임 (a) frame3

(b) 9번 프레임 (b) frame 9

(c) 15번 프레임 (c) frame 15

(d) 21번 프레임 (d) frame 21

그림 7. 오줌싸개소년 영상 집합의 영역화 결과
Fig. 7. Segmentation results of the BOY image set

MRF의 에너지 최소화 기법의 하나인 그랩컷(GrabCut) 기법의 입력으로 하여 인접 영상을 영역화하는 과정을 순차적으로 반복하여 영상 집합 전체를 영역화하게 된다.

제안하는 기법은 대응점을 기반으로 하므로 영상 간에

변형의 정도가 큰 경우에도 변형의 근사가 가능하며, 전경과 배경을 분리하여 대응되는 영역을 찾아 대응점의 오정합의 가능성을 줄이고 영상 간의 변형을 더욱 정밀하게 근사한다. 이때, 전경 영역의 경우 특징점 기반 레지스트레이션을 적용하여 전경 물체가 일정 수준 이상 움직임인 경우에도 변형을 근사할 수 있다. 제시된 실험 결과는 사용자가 단 한 장의 영상만을 영역화한 경우에 20장 넘는 영상이 정밀하게 영역화되는 등 적은 사용자 입력으로 물체의 다시점 영상 집합을 효과적으로 영역화할 수 있음을 뒷받침한다. 또한, 영역화된 영상과 인접 영상 간의 전경 및 배경의 변형이 잘못 근사된 경우에도 정보를 색 분포 근사 및 형상 정보 등 간접적으로 이용하기 때문에 실제 물체의 영역에 근접한 영역화가 이루어진다. 제안하는 기법은 다시점 스테레오(multi-view stereo) 기법에 대한 입력 및 같은 장소에서 촬영된 영상 집합을 영역화하여 일괄적으로 편집하는 데에 요긴하게 쓰일 것으로 예상된다.

참 고 문 헌

[1] Y. Boykov, M. P. Jolly, "Interactive Graph Cuts for Optimal Boundary & Region Segmentation of Objects in N-D images," In Proceedings of International Conference on Computer Vision (ICCV), vol. I, pp. 105-112, July 2001.

[2] Y. Boykov, V. Kolmogorov, "An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization in Vision," In IEEE transactions on Pattern Analysis and Machine Intelligence (PAMI), vol. 26, no. 9, pp. 1124-1137, Sept 2004.

[3] N. Campbell, G. Vogiatzis, C. Hernandez, R. Cipolla, "Automatic 3D Object Segmentation in Multiple Views Using Volumetric Graph-Cuts," In Proceedings of British Machine Vision Conference, vol. 1, pp. 530-539, September 2007

[4] D. Comaniciu, P, Meer, "Mean Shift: A Robust Approach Toward Feature Space Analysis," In IEEE Transactions on Pattern Analysis and Machine Intelligence, (PAMI) vol. 24, no. 5, pp. 603-619, May 2002

[5] A. Criminisi, G. Cross, A. Blake, V. Kolmogorov, "Bilayer Segmentation of Live Video," In Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), vol 1. pp 53-60, June 2006.

[6] P. F. Felzenswab, D. P. Huttenlocker, Distance "Transforms of Sampled Functions," Cornell University Technical Report, 2004

[7] R. Hartley, A. Zisserman, Multiple View Geometry in Computer Vision, Cambridge University Press; Second Ed., 2004

[8] M. P. Kumar, P. H. S Torr, "Obj Cut," In Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), vol 1. pp 18-25, June 2005.

[9] D. J. Kwon, I. D. Yun, K. H. Lee, S. U. Lee, "An Efficient Feature-Based Nonrigid Registration Using Free-Form Deformations: Application to Multiphase Liver CT Images," Submitted To Pattern Recognition

[10] Y. Li, J. Sun, H.-Y. Shum, "Video object cut and paste," In ACM Transactions on Graphics (SIGGRAPH), vol. 24, no. 3, pp. 595-600, August 2005.

[11] D. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," In International Journal of Computer Vision, vol. 60, no. 2, pp. 91-110, November 2004

[12] J. Pilet, V. Lepetit, P. Fua, "Real-Time Non-Rigid Surface Detection," In Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), vol. 1, pp. 822-828, June 2005

[13] C. Rother, V. Kolmogorov, A. Blake, "GrabCut - Interactive Foreground Extraction using Iterated Graph Cuts," In ACM Transactions on Graphics (SIGGRAPH), vol. 23, no. 3, pp. 309-314, August 2004.

[14] J. R. Shewchuk, An Introduction to the Conjugate Gradient Method Without the Agonizing Pain, Carnegie Mellon University, 1994

[15] J. Shi, J. Malik, "Normalized Cuts and Image Segmentation," In IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), vol. 22, no. 8, pp 885-905, August 2000.

[16] M. Sormann, C. Zach, J. Bauer, K. F. Karner, H. Bischof, "Automatic foreground propagation in image sequences for 3d re-construction," In Pattern Recognition, 27th DAGM Symposium, pp. 93-100, August 2005

[17] M. Sormann, C. Zach, K. F. Karner, "Graph Cut Based Multiple View Segmentation for 3D Reconstruction," In Proceedings of 3rd International Symposium on 3D Data Processing, Visualization and Transmission (3DPVT 2006), pp. 1085-1092, June 2006

[18] J. Sun, W. Zhang, X. Tang, H. Y. Shum, "Background Cut," In Proceedings of European Conference on Computer Vision, vol. 2, pp. 628-641, May 2006.

[19] L. Vincent, P. Soille, "Watersheds in Digital Spaces: an Efficient Algorithm Based On Immersion Simulations," IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), vol. 13, no. 6, pp. 583-598, 1991.

[20] J. Wang, P. Bhat, A. Colburn, M. Agrawala, M. F. Cohen, "Interactive video cutout," In ACM Transactions on Graphics (SIGGRAPH), vol. 24, no. 3, pp. 585-594, August 2005.

[21] P. Yin, A. Criminisi, J. Winn, I. A. Essa, "Tree-Based Classifiers for Bilayer Video Segmentation," In Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1-8, June 2007.

[22] 윤일동, 이경준, 이상욱, "카메라 보정 기법의 성능향상에 관한 연구," 제 20회 영상처리 및 이해에 관한 워크샵 논문집, 2008 2.

저 자 소 개

이 수 찬

- 2004년 : 서울대학교 전기공학부 학사
- 2004년 ~ 현재 : 서울대학교 전기공학부 박사과정 재학중
- 주관심분야 : 컴퓨터 비젼, 3차원 모델링

권 동 진

- 2001년 2월 : 서울대학교 자연과학대학 물리학,전산과학 학사
- 2005년 ~ 현재 : 서울대학교 전기공학부 박사과정 재학중
- 주관심분야 : 의료영상처리, 지문인식, 물체인식

윤 일 동

- 1989년 2월 : 서울대학교 제어계측공학과 공학사
- 1991년 2월 : 서울대학교 제어계측공학과 공학석사
- 1996년 8월 : 서울대학교 제어계측공학과 공학박사
- 1997년 3월 ~ 현재 : 한국외국어대학교 디지털정보공학과 교수
- 주관심분야 : 영상처리, 컴퓨터비젼

이 상 욱

- 1973년 : 서울대학교 전기공학부 학사
- 1976년 : 미국 Iwoa 주립대 전기공학과 석사
- 1980년 : 미국 Univ. of Southern California, 전기공학과 박사
- 1981년 : General Electric Co. 연구원
- 1983년 : M/A-COM Research Center 선임연구원
- 1983년 ~ 현재 : 서울대학교 전기공학부 정교수
- 주관심분야 : 영상처리, 컴퓨터 비젼