

연속영상에서 non-rigid object의 자동 분할

정회원 정철곤*, 김중규*, 안치득**

Automatic segmentation of non-rigid object in image sequences

Cheol-kon Jung*, Joong-kyu Kim*, Chiteuk Ahn** *Regular Members*

요 약

본 논문은 연속영상에서 non-rigid object를 자동으로 분할하는 알고리즘을 제안하였다. Non-rigid object는 형태의 변화가 일정하지 않은 object로서 기존의 분할 알고리즘과는 다른 새로운 분할 알고리즘을 필요로 한다. 본 논문에서는 특히 구름이나 연기와 같이 형태의 변화가 큰 non-rigid object를 자동으로 분할하는 알고리즘을 제안하였다. 제안된 알고리즘은 공간분할, 시간분할, 그리고 공간분할과 시간분할의 결합의 세가지 단계로 구성되어 있다. 공간분할은 영상에서 픽셀의 intensity를 마코프 랜덤 필드로 가정하고 에너지 최소화를 통해 영상을 분할한다. 시간분할은 속도벡터를 기반으로 하여 움직임이 있는 영역만을 분할한다. 마지막으로 공간분할과 시간분할을 결합하여 non-rigid object의 최종적인 분할을 수행한다. 실험결과, 제안된 알고리즘은 연속영상에서 non-rigid object를 자동으로 분할함을 확인할 수 있었다.

ABSTRACT

This paper presents automatic segmentation algorithm of non-rigid object in image sequences. Non-rigid object is deformable object with fuzzy, blurred and indefinite boundary. To segment this object, we have to use the new segmentation algorithm different from conventional segmentation algorithms. In this paper, we present automatic segmentation algorithm to apply to non-rigid object like smoke. The procedure toward complete segmentation consists of three steps: spatial segmentation, temporal segmentation, and fusion of spatial and temporal segmentation. Spatial segmentation segments current frame using markov random field modeling. Temporal segmentation segments moving region using velocity vector between sequential frames. Then, non-rigid object is segmented by fusion of spatial and temporal segmentation. Experimental results demonstrate the efficiency of the presented method.

I. Introduction

멀티미디어 서비스를 위한 부호화 표준으로 제정된 MPEG-4는 대화형 멀티미디어 방송, 이동 멀티미디어 통신, 인터넷 등 다양한 형태의 차세대 멀티미디어 서비스를 지원하는 데에 그 목적을 두고 있다. 특히 MPEG-4는 비디오 객체 평면(video object plane : VOP)의 개념을 도입하여 의미와 내용이 있는 객체만을 부호화하는 방식으로 화질은 개선하고 압축률은 높이는 객체 기반 부호화 방식을 채택하

고 있다. 따라서 비디오 객체 평면을 생성하기 위한 움직이는 객체의 분할은 매우 중요한 기술이다. 효과적이고 정확한 객체 분할 알고리즘은 MPEG-4의 성능을 좌우한다^{[1][3]}.

지금까지 이 분야에 대한 연구는 주로 rigid object나 형태의 변화가 크지 않은 non-rigid object를 대상으로 하여 진행되어 왔다. 대표적인 예로 유사변환 방법이나 워터셰드 변환과 같은 모팔로자를 이용한 방법이 있다. 최근에는 위성영상이나 적외선 영상 등을 이용하여 구름이나 연기와 같이 시간이

* 성균관대학교 전기전자 및 컴퓨터공학부 디지털신호처리연구실(jkkim@yurim.skku.ac.kr), ** 한국전자통신연구원
 논문번호 : K01083-0219, 접수일자 : 2001년 2월 19일

흐름에 따라 형태의 변화가 큰 non-rigid object를 대상으로 하는 연구가 점차로 진행되고 있다^{[4][5]}.

본 논문에서는 연속영상에서 non-rigid object를 자동으로 분할하는 알고리즘을 제안하였다. 제안된 알고리즘은 형태의 변화가 큰 non-rigid object에 대해서도 우수한 성능을 나타낸다. 제안된 알고리즘은 공간 분할과 시간 분할, 공간 분할과 시간 분할의 결합으로 구성되어 있다. 공간 분할에서는 영상의 intensity를 마코프 랜덤 필드로 가정하고 모델링하여 객체를 분할한다. 또한 시간 분할에서는 연속 영상에 대하여 움직임이 있는 영역만을 분할한다. 마지막으로 공간 분할과 시간 분할을 결합하여 비디오 객체를 최종적으로 분할한다.

본 논문의 구성은 다음과 같다. 2장에서는 본 논문에서 제안된 알고리즘의 전체 흐름도에 대하여 살펴보고, 3장에서는 공간 분할에 대하여, 4장에서는 시간 분할에 대하여 설명하였다. 또한 5장에서는 공간 분할과 시간 분할을 결합하여 non-rigid object를 최종적으로 분할하는 방법을 설명하였다. 6장에서는 실험결과를, 7장에서는 결론을 기술하였다.

II. Overview of proposed algorithm

본 논문에서 제안된 알고리즘은 공간분할과 시간 분할, 그리고 공간분할과 시간분할의 결합의 세 단계로 구성되어 있다. 그림 1은 제안된 알고리즘의 흐름도이다. 공간분할에서는 마코프 랜덤 필드를 이용해 영상을 공간적으로 분할한다. 공간 분할을 통해 얻어진 분할영상은 객체의 boundary 정보만을 나타낸다. 시간 분할은 연속 영상에서 움직임을 분할하여 의미있는 정보를 분할한다. 마지막으로 시간 분할에 의해 생긴 분할 영상을 이용해 공간분할에 의해 생겨난 각각의 영역이 움직임이 존재하는 영역인지 아닌지를 결정하여 비디오 객체를 분할한다.

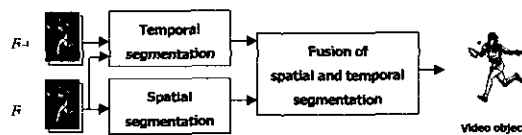


그림 1. Flow chart of presented algorithm

III. Spatial segmentation

공간분할은 전체 영상을 intensity에 따라 동질성

이 있는 영역으로 나누기 위해 수행이 되어진다. 본 논문에서는 각 픽셀의 intensity를 마코프 랜덤 필드로 모델링하여 공간적인 분할을 수행하였다. 각각의 영역들은 공간분할을 통해 얻어지는 boundary에 의해 구분되어진다.

3.1 Optimal determination of cluster's number

영상을 마코프 랜덤 필드로 모델링하여 분할을 수행하기 위해서는 최적의 클러스터의 개수를 입력하는 과정이 반드시 필요하다. 3.1절에서는 클러스터의 최적의 개수를 결정하는 과정을 설명하였다. 영상에서의 한 픽셀 (x, y) 에 대한 intensity를 불규칙변수 $z(x, y)$ 라고 가정한다. 여기서 불규칙변수 $z(x, y) = z$ 라고 하면, z 는 그림 2와 같은 분포를 나타낸다^[6].

그림 2와 같이 영상에서 intensity z 의 분포 $h(z)$ 는 K 개의 확률밀도 함수의 합으로 (1)식과 같이 표현할 수 있다. 확률밀도함수가 가우시안 분포를 나타낸다고 가정하면 $h(z)$ 는 다음과 같다.

$$h(z) = \sum_{i=0}^{K-1} \frac{P_i}{\sigma_i \sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{z - m_i}{\sigma_i} \right)^2 \right] \quad (1)$$

여기서 P_i 은 각 확률밀도 함수들이 가지는 사전 확률을 의미하며, m_i 과 σ_i 은 각 확률밀도 함수의 평균과 표준편차이다. 또한 확률밀도함수의 개수 K 는 곧 z 의 cluster의 개수와 동일하므로, 본 논문에서는 Rose에 의해 제안된 cluster validity measure를 개선하여 사용하였다. 이 방법은 최적의 K 값이 cluster의 평균값과 소속된 점들 사이의 거리의 합을 최소화시키고, 각 평균사이의 거리를 최대화시킨다는 점을 이용하였다^[7].

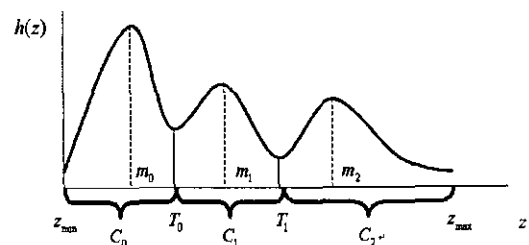


그림 2. z 의 분포

개선된 cluster validity measure *validity* 는 (2) 식과 같다.

$$validity = w \cdot \frac{intra}{inter} \quad (2)$$

여기서 w 는 가중치 상수이다. 또한 N 을 전체 픽셀의 개수라 하고, l 번째 cluster를 C_l

$$(l=0,1,\Lambda, K-1) \text{라고 하면, } intra = \frac{1}{N} \sum_{l=0}^{K-1} \sum_{z \in C_l} |z - m_l|$$

$$\text{이고, } inter = \frac{1}{K-1} \sum_{l=1}^{K-1} |m_l - m_{l-1}| \text{이다.}$$

따라서 최적의 K 는 *validity* 를 최소화시키는 값이다

3.2 Pixel labeling by optimal threshold value

각 픽셀이 소속된 cluster를 결정하기 위해 최적의 문턱값 $T_n (n=0,1,\Lambda, K-2)$ 을 결정해야 한다. 각 분포의 표준편차 $\sigma_0 = \Lambda = \sigma_{K-1} = \sigma$ 라고 가정하면, 최적의 문턱값 T_n 은 다음 식에 의해 결정된다^[8].

$$T_n = \frac{m_n + m_{n+1}}{2} + \frac{\sigma^2}{m_n - m_{n+1}} \ln \frac{P_{n+1}}{P_n} \quad (3)$$

(3)식에서 $P_n = P_{n+1}$ 이라고 가정하면, 문턱값 T_n 은 $\frac{m_n + m_{n+1}}{2}$ 이 된다. 본 논문에서는 $P_n = P_{n+1}$ 라고 가정하고 영상을 분할하였다.

(3)식에 의해 결정된 T_n 을 이용하여 픽셀의 레이블 l 을 결정할 수 있다. 영상에서 한 픽셀 (x, y) 의 레이블 필드 $L(x, y)$ 라고 하면 (4)식과 같이 각 픽셀의 레이블 l 은 소속된 cluster의 index와 같다.

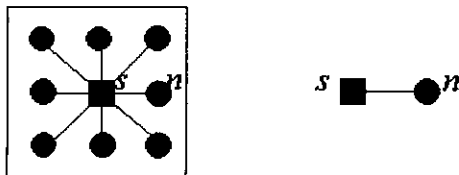


그림 3. Neighborhood system and binary clique
(s : center pixel n : neighboring pixel)

$$L(x, y) = l, \quad z(x, y) \in C_l \quad (4)$$

3.3 Spatial segmentation by MRF modeling

그림 2.의 문턱값 T_n 을 이용해 thresholding을 하게 되면, *intensity*의 변화에 의한 조잡한 *boundary* 를 많이 발생시켜 영상을 불필요하게 많이 작은 영역들로 나누게 한다. 이러한 조잡한 *boundary* 를 제거하고 동질성을 가지는 영역을 합병하기 위해 마코프 랜덤 필드 모델링을 하였다. 마코프 랜덤 필드 모델은 공간적인 동질성을 이용한 모델로서 *intensity*의 변화에 의한 조잡한 *boundary* 를 제거하고, 동질성을 가지는 작은 영역들을 합병할 수가 있다.

영상의 한 픽셀 (x, y) 의 *intensity* z 를 마코프 랜덤 필드라 가정하면, neighborhood system과 이진 클릭은 그림 3.와 같이 표현되어진다. 여기서 z 가 가지는 확률 $P(z)$ 는 Hammersley-Clifford 등가성에 의해 (5)식과 같이 표현되어진다^{[9][10]}.

$$P(z) = Q^{-1} \times e^{-U(z)} \quad (5)$$

여기서 Q 는 분할 함수이고, $U(z)$ 는 에너지 함수이다.

각 픽셀의 z 에 대한 최종 레이블 l^* 은 (6)식과 같이 MAP(maximum a posteriori criterion) 이론과 베이스 이론에 의해 결정된다.

$$l^* = \arg \max_l (P(l|z)) = \arg \max_l (P(z|l) \cdot P(l)) \quad (6)$$

또한, (5)식에 의해 (6)식을 다시 쓰면 (7)식과 같이 된다.

$$\max_l (P(l|z)) = \min_l (U(l|z)) \quad (7)$$

따라서 최종 레이블 l^* 은 에너지의 최소값을 갖는 레이블과 같다. 그리고, (7)식의 에너지 함수는 (8)식과 같다^[9].

$$U(l|z) = U_a(z|l) + U_m(l) \quad (8)$$

여기서 $U_a(z|l)$ 은 레이블과 데이터를 연결하는 연결 에너지로서 (9)식과 같이 표현된다.

$$U_a(z|I) = \frac{1}{2\sigma_0^2} \sum_i (z - m_i)^2 \quad (9)$$

또한 $U_m(I)$ 은 모델 에너지로서 잡음에 의해 생긴 불필요한 영역을 제거하고, 움직이는 객체에 의한 영역에 공간적인 유사성을 부여한다. $U_m(I)$ 은 (10)식처럼 그림 3.의 이진 클러스터들의 포텐셜 함수 $V_c(l_s, l_n)$ 의 합이다. (l_s : s 의 레이블, l_n : n 의 레이블)

$$U_m(I) = \sum_c V_c(l_s, l_n) \quad (10)$$

여기서 β 를 양의 상수라고 하면, 포텐셜 함수 $V_c(l_s, l_n)$ 는 (11)식과 같다.

$$V_c(l_s, l_n) = \begin{cases} -\beta, & \text{if } l_s = l_n \\ +\beta, & \text{if } l_s \neq l_n \end{cases} \quad (11)$$

본 논문에서는 에너지 함수 $U(I|z)$ 의 최소화를 위해 수행속도가 빠른 ICM(iterated conditional mode)을 사용하였다. ICM은 영상의 각 픽셀에 대한 두 개의 레이블 값의 에너지를 비교하여 최소의 에너지를 갖는 레이블을 각 픽셀에 부여한다. 이 과정을 수렴할 때까지 반복하여 최종적인 레이블 값을 영상에 부여한다. 본 논문에서 수렴조건으로 다음 식을 사용하였다^{[11][12]}.

$$\sum_{(x,y)} |I^k - I^{k-1}| < \epsilon \quad (12)$$

여기서 k 는 반복횟수를 의미하고, ϵ 은 임의로 지정한 작은 수이다.

IV. Temporal segmentation

시간분할은 공간분할을 통해 얻어진 분할 영상에 대하여 움직임이 존재하는 영역을 결정하기 위해 수행된다. 만약 시간의 변화에 따라 조명의 변화가 거의 없고, 카메라가 정지되어 있다면, 비디오 객체의 motion은 intensity의 변화와 밀접한 관련이 있다. 즉 motion이 존재하는 영역에서는 intensity의 변화에 의한 속도벡터가 발생하게 된다. 따라서 시간분할은 연속 영상에서 객체의 움직임에 의해 발

생하는 속도벡터를 이용하여 수행한다. 시간분할은 'motion decision'과 'false correction'의 두 가지 과정을 거쳐 움직임이 존재하는 영역을 결정한다.

4.1 Velocity vector by optical flow analysis

시간 t 에 대한 영상에서의 한 점 (x, y) 의 intensity를 $I(x, y, t)$ 라고 한다. 연속 영상에서 객체의 motion이 적다고 가정하면, optical flow constraint 방정식은 (13)식과 같이 된다^{[6][13]}.

$$I_x u + I_y v + I_t = 0 \quad (13)$$

여기서 $u = \frac{dx}{dt}, v = \frac{dy}{dt}, I_x = \frac{\partial I}{\partial x}, I_y = \frac{\partial I}{\partial y}, I_t = \frac{\partial I}{\partial t}$ 이고, u 와 v 는 속도벡터를 이루는 구성 성분들이다.

본 논문에서는 (13)식의 속도벡터를 계산하기 위해 계산속도가 가장 빠른 Lucas와 kanade가 제안한 방법을 사용하였다. Lucas와 Kanade는 작은 영역 B 에 소속된 픽셀에 대한 motion 벡터는 동일하다는 가정하에 속도벡터를 계산하였다. 계산된 속도벡터는 (14)식과 같다.

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \sum_{(x,y) \in B} I_x^2 & \sum_{(x,y) \in B} I_x I_y \\ \sum_{(x,y) \in B} I_x I_y & \sum_{(x,y) \in B} I_y^2 \end{bmatrix}^{-1} \begin{bmatrix} - \sum_{(x,y) \in B} I_x I_t \\ - \sum_{(x,y) \in B} I_y I_t \end{bmatrix} \quad (14)$$

4.2 Motion decision by signal detection theory

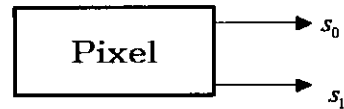


그림 4. 두 가지 경우

(14)식에 의해 얻어진 속도벡터를 이용해 영상에서 motion의 존재 유무를 결정하게 되는데, 그 과정은 아래와 같다. 신호 탐지 이론에 따라 motion의 존재 유무를 신호의 존재 유무로 가정한다. 그림 4.와 같이 영상의 각 픽셀은 두 가지 경우로 나뉘어 지는데, s_0 는 motion이 존재하지 않는다는 것을 나타내고, s_1 은 motion이 존재함을 나타낸다^[14].

그리고, 어떤 픽셀 (x, y) 에 대한 속도벡터의 크

기를 불규칙변수 $r(x, y)$ 라고 하면, 불규칙변수 $r(x, y)$ 는 (15)식과 같이 나타내어진다.

$$r(x, y) = \sqrt{u(x, y)^2 + v(x, y)^2} \quad (15)$$

불규칙변수 $r(x, y) = r$ 라고 하면, r 는 그림 5와 같이 s_0 와 s_1 에 대한 평균이 a_0, a_1 인 두 개의 조건부 확률밀도함수 $p(r|s_0), p(r|s_1)$ 을 가지며, 이를 s_0 와 s_1 의 *likelihood*라고 부른다^[15].

각각의 *likelihood*는 (16),(17)식과 같이 나타내어진다.

$$p(r|s_0) = \frac{1}{\sigma_0 \sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{r - a_0}{\sigma_0} \right)^2 \right] \quad (16)$$

$$p(r|s_1) = \frac{1}{\sigma_0 \sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{r - a_1}{\sigma_0} \right)^2 \right] \quad (17)$$

여기서 σ_0^2 은 잡음의 분산이다. 또한 $p(r|s_0)$ 은 s_0 인 경우 r 의 확률밀도를 나타내고, $p(r|s_1)$ 은 s_1 인 경우 r 의 확률밀도를 나타낸다.

각 픽셀에서 motion의 존재 유무를 결정하기 위해 그림 5와 같이 문턱값 γ_0 를 사용한다. 문턱값 γ_0 는 두 개의 *likelihood*가 서로 대칭이기 때문에 $(a_0 + a_1)/2$ 이다. 또한 γ_0 을 이용해 각 픽셀에 대하여 레이블링을 할 수 있다. Motion이 존재하는 픽셀에 대한 레이블 값을 1이라 하고, motion이 존재하지 않는 픽셀에 대한 레이블 값을 0이라고 하면, 각 픽셀의 레이블 l 을 (18)식에 의해 결정할 수 있다.

$$l = \begin{cases} 1, & r \geq \gamma_0 \\ 0, & r < \gamma_0 \end{cases} \quad (18)$$

여기서 평균값 a_0, a_1 은 평균값의 개수를 2로 두고, K-means 알고리즘을 수행하여 얻었다^[5].

4.3 False correction using MRF model

그림 5에서 문턱값 γ_0 로 thresholding을 하게 되면, 두 가지 문제가 발생하게 된다. 하나는 motion이 없는 경우에 motion이 있다고 결정하는 *false*

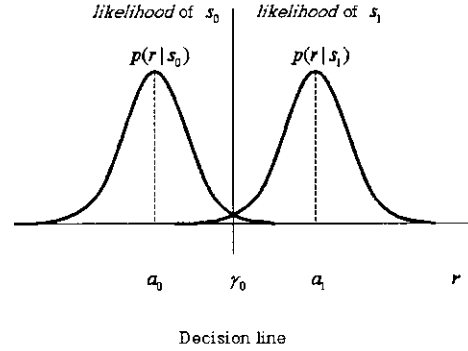


그림 5. 조건부 확률밀도함수 : $p(r|s_0), p(r|s_1)$

*alarm*과 또 하나는 motion이 있는 경우에 motion이 없다고 결정하는 *miss*이다. 이러한 두 가지 경우를 보상하기 위해 3.3절에서 언급한 베이스 규칙에 기반한 마코프 랜덤 필드 모델을 사용하여 오류를 수정하였다^{[9][10]}.

V. Fusion of spatial and temporal segmentation

공간분할을 통해 얻어진 분할 영상은 boundary 정보만을 가지고 있기 때문에 의미 있는 객체를 분할하는 마스크로 사용되기가 어렵다. 따라서 시간분할을 통해 얻어진 분할 영상을 이용하여 의미 있는 객체 분할을 위한 최종 분할 마스크를 생성한다.

먼저 공간분할을 통해 얻은 분할 영상에 대하여 connected component labeling을 수행하여 각각의 영역에 대한 레이블링을 수행한다^[16]. 그리고 시간분할을 통해 얻은 분할영상으로 그 영역이 움직임이 존재하는 영역인지 아닌지를 결정한다. 결정하는 방법은 [2]에서 제안된 방법을 수정하여 사용하였다.

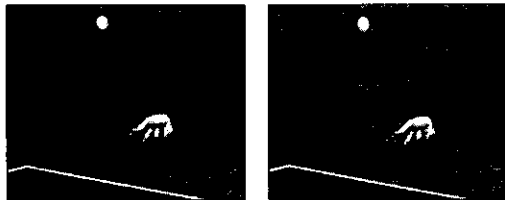
연속영상에서 현재 프레임을 F_i , 이전 프레임을 F_{i-1} 라고 할 때, 시간분할에 의해 얻어진 분할 마스크를 TM_i (temporal mask)라고 한다. 또한 공간분할에 의해 분할된 i 번째 영역을 $R_{i,j}$ 라고 하고, TM_i 에 의해 영역 $R_{i,j}$ 에 투영된 영역을 $R_{i,j}^{proj}$ 라고 한다면, 움직이는 객체의 영역임을 결정하는 식은 아래와 같다^[2].

$$P = \frac{N(R_{i,j}^{proj} \cap R_{i,j})}{N(R_{i,j})} \begin{cases} \geq \tau: & \text{moving region} \\ < \tau: & \text{background region} \end{cases} \quad (19)$$

여기서 $N(R_{i,t})$ 는 $R_{i,t}$ 에 포함되어 있는 픽셀의 개수를 의미한다. 즉 (19)식에서 P 가 임계값 τ 보다 크거나 같으면 움직이는 영역이고, 반대인 경우에는 배경 영역이다. 본 논문에서는 임계값 τ 를 0.5로 설정하였는데, 이는 50% 이상이 되어야 영역 $R_{i,t}$ 를 움직이는 영역으로 결정한다는 것을 의미한다.

VI. Experimental results

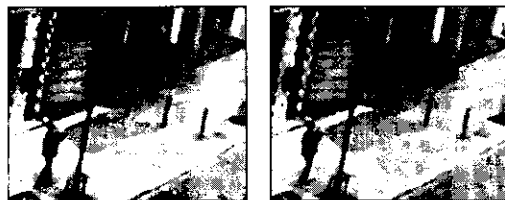
6장에서는 실제 연속영상에 대하여 제안된 알고리즘을 적용한 결과를 기술하였다. 실험환경은 P-III 733MHz CPU, 256MB의 메모리를 가지는 IBM호환 PC상에서 수행하였으며, OS는 Windows 98, 개발틀은 Visual C++ 6.0을 사용하였다. 실험에 사용된 영상은 176×144의 해상도를 갖는 QCIF format의 'Table tennis', 'Foreman', 'Street'와 'Smoke' sequences 중에서 연속되는 두 프레임이다. 그림 6.은 실험에 사용된 원영상이다. 'Table tennis', 'Foreman' sequences는 형태의 변화가 크지 않은 non-rigid object이며, 'Street'와 'Smoke' sequences는 형태의 변화가 큰 non-rigid object이다. 실험결과, 제안된 알고리즘은 연속 영상에서 non-rigid object를 분할하는데 우수한 성능을 나타내었다.



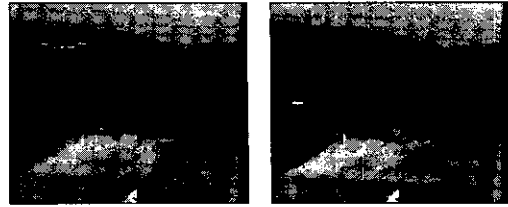
(a) 'Table tennis' sequence with QCIF



(b) 'Foreman' sequence with QCIF



(c) 'Street' sequence with QCIF



(d) 'Smoke' sequence with QCIF

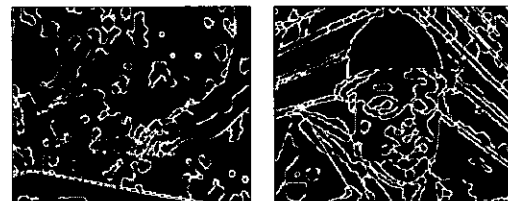
그림 6. Original images

6.1 Spatial segmentation

공간분할은 영상의 intensity를 마코프 랜덤 필드로 모델링하여 에너지 최소화 과정을 통해 영상을 분할한다. 에너지 최소화를 수행하기 전에 먼저 클러스터의 개수를 입력하는 과정이 필요하다. 표 1.은 validity를 클러스터의 개수 K 에 대하여 구한 결과이다. K 는 2에서 10까지의 값을 사용하였고, w 는 1로 설정하였다. 최적의 K 값은 validity를 최소화시키는 값으로 'Table tennis' sequence에서는 9, 'Foreman' sequence에서는 10, 'Street' sequence에서는 6, 'Smoke' sequence에서는 7이다.

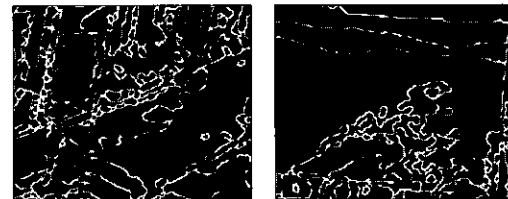
표 1. validity

K	Table tennis	Foreman	Street	Smoke
2	0.04232	0.06757	0.07306	0.02579
3	0.01685	0.03920	0.04285	0.01956
4	0.01880	0.02463	0.04593	0.01784
5	0.01393	0.01694	0.02706	0.02286
6	0.00854	0.02188	0.01908	0.02060
7	0.00777	0.02499	0.02095	0.00583
8	0.00650	0.02541	0.02345	0.00591
9	0.00535	0.01831	0.02160	0.00773
10	0.21581	0.01618	0.02150	0.00890



(a)

(b)



(c)

(d)

그림 7. Spatial segmentation results

이렇게 얻은 최적의 K 를 입력 값으로 하여 각 sequence의 m_i 과 T_n 을 구하여 각 픽셀이 소속된 클러스터의 레이블을 각 픽셀에 할당한다. 그리고 마코프 랜덤 필드 모델링을 하여 영상을 분할한다. 그림 7.은 이러한 공간분할을 통해 얻은 결과이다.

6.2 Temporal segmentation

공간분할이 수행되는 동안, 연속되는 두 프레임에 대하여 시간분할이 수행된다. 먼저 optical flow를 구하여 각 픽셀에 대한 속도벡터를 얻고, 이 속도벡터에 대하여 motion의 존재 여부를 결정한다. 작은 영역 B 의 크기는 3×3 이고, ϵ 은 1이며, σ_0 의 값은 100이다. 그리고 반복 횟수 k 는 약 5~15이다.

그림 8.은 시간분할을 수행함에 의해 얻어진 분할 영상이다. 그림에서 motion이 존재하는 영역의 레이블 값은 1이며, motion이 존재하지 않는 영역의 레이블 값은 0이다. 형태의 변화가 큰 non-rigid object의 motion을 포함하고 있는 'Street'와 'Smoke' sequences에서는 움직이는 객체와는 상관 없는 영역들도 분할되어 있는 것을 알 수 있다.

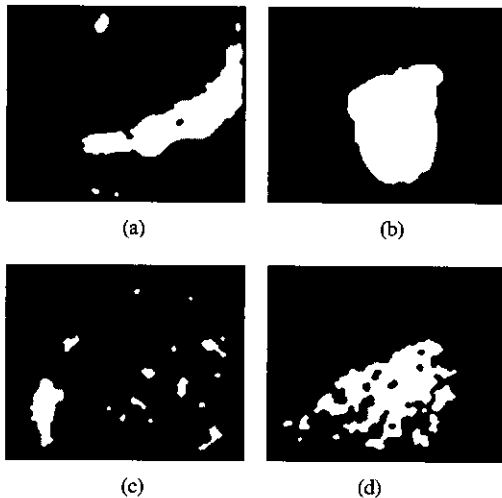


그림 7. Temporal segmentation results

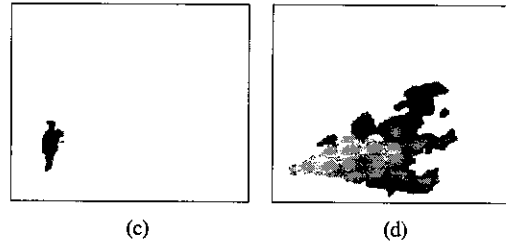
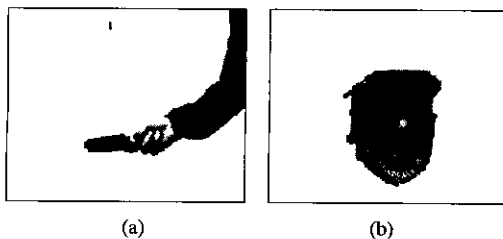


그림 9. Final segmentation results

6.3 Fusion of spatial and temporal segmentation

공간분할과 시간분할을 수행함으로 얻어진 두 분할 결과를 결합하여 최종 분할 영상을 획득한다. 이를 위해 먼저 공간 분할된 영상에 대하여 connected component labeling을 수행하여 각 영역에 대하여 레이블 값을 부여한다. 그리고 시간 분할된 영상을 공간 분할된 영상에 투영하여 (19)식에 의해 실제 움직임이 존재하는 영역을 결정한다. 그림 9.는 공간분할과 시간분할의 결합에 의해 최종적으로 분할된 영상을 나타낸 그림이다. 최종분할 결과에서 보는 바와 같이 제안된 알고리즘은 non-rigid object를 효율적으로 잘 분할함을 확인할 수 있다.

VII. Conclusion

본 논문에서는 연속영상에서 non-rigid object를 자동으로 분할하는 알고리즘을 제안하였다. 제안된 알고리즘은 마코프 랜덤 필드 모델을 이용한 공간 분할과 속도벡터에 의한 시간분할, 그리고 공간분할과 시간분할의 결합에 의한 최종 분할의 세 가지 과정으로 구성되어 있다. 공간분할을 하여 영상을 객체에 따라 나누고, 시간분할을 하여 분할된 영역이 motion이 존재하는 영역인가 아닌가를 결정한다. 제안된 알고리즘은 형태의 변화가 일정하지 않은 non-rigid object에 대해서 분할 성능이 뛰어난 실험결과와 확인할 수 있었다.

현재, 영상 분할 기술은 멀티미디어 서비스를 위한 부호화 표준안으로 제정된 MPEG-4, MPEG-7에 응용되고 있다. MPEG-4와 MPEG-7은 객체 기반 부호화 방식을 채택하고 있으며, 비디오 객체 평면 (VOP : Video Object Plane)의 개념을 도입하여 부호화 효율을 높이고 있다. 본 논문에서 제안된 분할 알고리즘은 특히 구름이나 연기, 또는 적외선 영상과 같은 형태의 변화가 큰 non-rigid object의 분할에서 좋은 결과를 나타내고 있다. 따라서 제안된

알고리즘은 non-rigid object의 비디오 객체 평면 생성을 위한 기반 기술로 응용이 가능하리라 생각된다.

참고 문헌

[1] T. Sikora, "The MPEG-4 Video Standard Verification Model", *IEEE Trans. Circuits and Systems for Video Technology*, 7, pp. 19-31, 1997.

[2] M. Kim, J. G. Choi, D. Kim, H. Lee, M. H. Lee, C. Ahn, and Y. S. Ho, "A VOP Generation Tool: Automatic Segmentation of Moving Objects in Image Sequences Based on Spatio-Temporal Information", *IEEE Trans. Circuits and Systems for Video Technology*, 9, pp. 1216-1999, 1999.

[3] M. Kim, J. G. Jeon, J. S. Kwak, M. H. Lee, and C. Ahn, "Moving Object Segmentation in Video Sequences by User Interaction and Automatic Object Tracking", *Image and Vision Computing*, 19, pp. 245-260, 2001.

[4] J. Fan, J. Yu, G. Fujita, T. Onoye, L. Wu, and I. Shirakawa, "Spatiotemporal Segmentation for Compact Video Representation", *Signal Processing: Image Communication*, 16, pp. 553-566, 2001.

[5] C. K. Jung, K. S. Kim, and J. K. Kim, "Automatic Moving Object Segmentation using Automatic Region Growing Algorithm", *Korean Institute of Communication Sciences*, 26, pp. 187-193, 2001.

[6] A. M. Tekalp, *Digital video processing*, Prentice Hall, pp. 72-94 508-517, 1995.

[7] S. Ray and R. H. Turi, "Determination of Number Clusters in K-means Clustering and Application in Colour Image Segmentation", *ICAPRDT' 99*, pp. 137-143, 1999.

[8] R. C. Gonzalez and R. E. Woods, *Digital image processing*, Addison Wesley, pp. 443-458, 1992.

[9] F. Luthon, A. Caplier, and M. Lievin, "Spatiotemporal MRF approach to video segmentation: Application to motion detection and lip segmentation", *Signal Processing*, 76, pp 61-80, 1999.

[10] M. Zimanyi, "Reconstruction of Tomographic Data by Markov Random Fields", *Central European Seminar on Computer Graphics*, 1998.

[11] R. C. Dubes, A. K. Jain, S. G. Nadabar, and C. C. Chen, "MRF Model-based Algorithms for Image Segmentation", *Pattern Recognition, Proceedings, 10 th International Conference I*, pp. 808-814, 1990.

[12] J. Wei and Z. Li, "An efficient two-pass MAP-MRF algorithm for motion estimation based on mean field theory", *IEEE trans. on circuits and systems for video technology*, 9, pp. 960-972, 1999.

[13] B. K. P. Horn and B. G. Schunck, "Determining Optical Flow", *Artificial Intelligence*, 17, pp. 185-203, 1981.

[14] M. Barkat, *Signal detection & estimation*, Artech House, pp. 115-174, 1991.

[15] B. Sklar, *Digital communication*, Prentice Hall, pp. 132-138, 1988.

[16] R. Jain, R. Kasturi, and B. G. Schunck, *Machine Vision*, McGraw-Hill, pp. 25-72, 1995.

정 철 곤(Cheol-kon Jung)

정회원



1995년 2월 : 성균관대학교
전자공학과 학사
1997년 2월 : 성균관대학교
전자공학과 석사
1997년~현재 : 성균관대학교
전기전자 및 컴퓨터
공학부 박사과정

<주관심 분야> 레이더신호처리, 영상처리, 컴퓨터비전, 컴퓨터그래픽스, 디지털비디오처리

김 중 규(Joong-kyu Kim)

정회원



1980년 : 서울대학교 전자공학과
졸업
1982년 : 서울대학교 전자공학과
석사
1989년 : The University of
Michigan, Ann Arbor,
Department of Electri

cal and Computer Engineering Ph.D.

1980년~1981년 : 한국전자통신연구소 위촉연구원

1989년~1990년 : University of Michigan, Post

Doctoral

1990년~1991년 : 삼성전자 선임연구원

1992년~현재 : 성균관대학교 전기전자 및 컴퓨터공학부 교수

<주관심 분야> 적응신호처리, 레이더신호처리, 의학 영상신호처리, 음향신호처리, 디지털비디오 처리

안 치 득(Chieteuk Ahn)

정회원

1980년 : 서울대학교 전자공학과 졸업

1982년 : 서울대학교 전자공학과 석사

1991년 : The University of Florida, Gainesville

1982년~한국전자통신연구원 방송시스템연구부장

<주관심 분야> 비디오 신호처리, 비디오 통신