

논문 2006-43SP-4-9

SOM 기반의 계층적 군집 방법을 이용한 계산 효율적 비디오 객체 분할

(Computationally Efficient Video Object Segmentation using
SOM-Based Hierarchical Clustering)

정 찬 호*, 김 경 환**

(Chanho Jung and Gyeonghwan Kim)

요 약

본 논문에서는 계산 효율적이고 노이즈에 강건한 비디오 객체 분할 알고리즘을 제안한다. 움직임 분할과 색 분할을 효율적으로 결합한 시공간 분할 방법의 구현을 위해 SOM 기반의 계층적 군집 방법을 도입하여 특징 벡터들의 군집 관점에서 분할 과정을 해석함으로써 기존의 객체 분할 방법에서 정확한 분할 결과를 얻기 위해서 요구되어지는 많은 연산량과 노이즈에 의한 시스템의 성능 저하 문제를 최소화한다. 움직임 분할 과정에서는 움직임 추정 에러에 의한 영향을 최소화하기 위해서 MRF 기반의 MAP 추정 방법을 이용하여 계산한 움직임 벡터의 신뢰도를 이용한다. 또한 움직임 분할의 성능 향상을 위해서 움직임 신뢰도 히스토그램을 이용한 노이즈 제거 과정을 거칠 뿐만 아니라 자동으로 장면 내에 존재하는 객체의 수를 구하기 위해서 군집 유효성 지표를 이용한다. 객체 추적의 성능 향상을 위해 교차 투영 기법을 이용하며, 분할 결과의 시간적 일관성 유지를 위해 동적 메모리를 이용한다. 다양한 특성을 가지는 비디오 시퀀스들을 이용한 실험을 통해 제안하는 방법이 계산 효율적이고 노이즈에 강건하게 비디오 객체 분할을 수행함은 물론 기존의 구현 방법에 비해 정확한 분할 결과를 얻을 수 있음을 확인하였다.

Abstract

This paper proposes a robust and computationally efficient algorithm for automatic video object segmentation. For implementing the spatio-temporal segmentation, which aims for efficient combination of the motion segmentation and the color segmentation, an SOM-based hierarchical clustering method in which the segmentation process is regarded as clustering of feature vectors is employed. As results, problems of high computational complexity which required for obtaining exact segmentation results in conventional video object segmentation methods, and the performance degradation due to noise are significantly reduced. A measure of motion vector reliability which employs MRF-based MAP estimation scheme has been introduced to minimize the influence from the motion estimation error. In addition, a noise elimination scheme based on the motion reliability histogram and a clustering validity index for automatically identifying the number of objects in the scene have been applied. A cross projection method for effective object tracking and a dynamic memory to maintain temporal coherency have been introduced as well. A set of experiments has been conducted over several video sequences to evaluate the proposed algorithm, and the efficiency in terms of computational complexity, robustness from noise, and higher segmentation accuracy of the proposed algorithm have been proved.

Keywords : video object segmentation, self-organizing map(SOM), hierarchical clustering, motion vector reliability, rule-based region processing

* 학생회원, LG전자 디지털TV 연구소
(Digital TV Research Lab. LG Electronics)

** 정회원, 서강대학교 전자공학과
(Department of Electronic Engineering, Sogang University)

※ 이 연구는 2005년도 서강대학교 교내 연구비 지원에 의하여 이루어졌음.

접수일자: 2006년12월5일, 수정완료일: 2006년6월7일

I. 서 론

최근 내용 기반 비디오 응용분야에 대한 관심의 증가에 힘입어 비디오 시퀀스의 내용 기반 표현에 대한 연구에 상당한 진전이 있어왔다. 객체 기반 비디오 부호화, 내용 기반 비디오

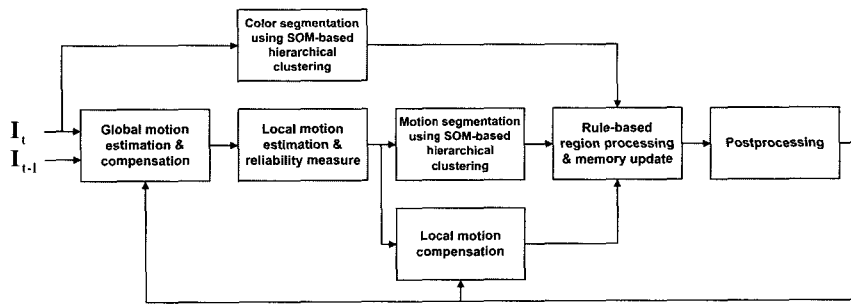


그림 1. 제안하는 시스템의 블록 다이어그램
Fig. 1. Block diagram of the proposed system.

색인 및 검색, 비디오 감시 시스템, 비디오 편집과 같은 내용 기반 기능성(content-based functionality)을 제공하는 응용분야에서는 비디오 시퀀스의 내용 기반 표현이 중요한 역할을 하기 때문이다^[1,2]. 비디오 시퀀스의 내용 기반 표현을 위해서 입력된 비디오 시퀀스는 임의의 형태를 가진 물리적인 객체들로 구분될 필요가 있다. 비디오 부호화 표준 MPEG-4에서는 비디오 시퀀스를 구성하는 의미 있는 객체들을 VOP(video object plane)로 정의한다^[3-5]. 각 객체의 형태, 움직임, 색과 같은 특징들은 비디오 시퀀스의 전송 및 저장 과정에서 부호화 되고, 효율적인 비디오 시퀀스의 색인 및 검색 과정에서 유용한 정보를 제공하기도 한다. 따라서 내용 기반 비디오 응용분야의 성능은 비디오 시퀀스의 정확하고 효율적인 분할 결과에 크게 영향을 받는다^[6].

비디오 시퀀스에서 움직이는 객체들을 분할해내기 위한 다양한 방법들이 연구되어 왔다. 시공간 분할 방법은 기존의 방법들 중에서 비디오 객체 분할을 위한 가장 정확하고 효율적인 방법 중 하나이며, 움직임 분할 과정과 색 분할 과정의 효율적인 결합을 필요로 한다. 시공간 분할을 위해서 K-평균 군집 방법(K-means clustering)^[7], 베이시안 추정(Bayesian estimation)^[8], RSST(recursive shortest spanning tree)^[9,10], 모폴로지를 이용한 watershed^[11]가 이용되기도 한다. 이러한 방법들은 비교적 정확하게 객체들의 경계선을 결정하지만 계산 복잡도가 높다는 단점이 있다^[1]. 이 중에서 RSST를 이용한 방법의 경우 다른 방법들에 비해서 여러 가지 장점을 갖는데, 파라미터의 초기화가 필요하지 않고 비용 함수를 구성하는 서로 다른 부분들의 가중치를 설정하지 않아도 되는 점뿐만 아니라 개략적인 분할 결과로부터 세밀한 분할 결과까지 제공해 주기 때문에 분할 결과에 대한 효율적인 제어가 가능하다^[10]는 점 등을 들 수 있다. 비슷한 성능을 보이면서 상대적으로 낮은 계산 복잡도를 갖는 M-RSST(multiscale recursive shortest spanning tree)^[12]가 이용되기도 하는데 영상의 크기가 클 경우에는 이 방법도 높은 계산 복잡도를 갖는다. 또한 기존의 방

법들에서는 움직임 추정 에러를 포함한 노이즈를 고려한 처리 과정이 충분하지 않으며, 장면 내에 존재하는 객체의 수를 결정해주는 과정에서 사용자 관여가 필요하거나 임계값의 설정이 필요한 경우가 있다^[3,5,10,13].

본 논문에서는 위에서 언급한 문제점들을 해결하기 위해 연속적인 프레임(그림 1에서 I_t 과 I_{t-1} 로 표시)을 입력받아 비디오 객체 분할을 수행하기 위한 기능단위들로 구성된 시스템을 그림 1과 같이 제안한다. 제안하는 시스템은 European COST 211 group이 제안한 AM(analysis model)^[9]과 비슷한 구조를 갖지만 각 기능단위들은 계산 효율적이고 노이즈에 강건하도록 설계되었다.

움직임 분할과 색 분할을 위해서는 SOM 기반의 계층적 군집 방법을 이용한 분할 알고리즘을 제안한다. 분할의 첫 번째 단계에서는 특징 벡터들을 이용하여 SOM을 학습시키며, 두 번째 단계에서는 학습된 SOM에 기반한 응집 계층적(agglomerative hierarchical) 군집 방법을 이용하여 최종 분할 결과를 얻는다. 이러한 두 단계 접근 방법은 계산 복잡도에 대한 장점뿐만 아니라 노이즈 감소의 효과를 보이는데^[14,15], 이는 SOM에 의해서 학습된 신경소자들이 특징 벡터들의 국부 평균 값을 나타내기 때문에 K-평균 군집 방법, RSST를 이용한 방법에서와 같이 특징 벡터들을 직접적으로 군집하는 경우보다 특징 벡터들의 불규칙적인 변화에 덜 민감하기 때문이다^[14].

움직임 벡터의 신뢰도 계산에는 MRF(Markov random fields) 기반의 MAP(maximum a priori) 추정 방법의 이용을 제안한다. 국부 움직임 추정 과정에서 aperture 문제, 가림(occlusion) 문제와 같은 근본적인 오류 발생 원인에 의해서 정확하지 않은 움직임 벡터가 추정된다. 이러한 움직임 벡터는 움직임 분할 과정에서 정확한 객체의 경계선을 구하는 데 방해 요소가 될 수 있다. 따라서 본 논문에서는 움직임 분할의 정확성을

향상시키기 위해서 움직임 벡터의 신뢰도를 구하여 움직임 분할 과정에서 이용한다. 이 과정에서 추가적인 노이즈를 발생시킬 수 있는 밝기 또는 색 정보의 기울기를 이용하는 것^[13,16]이 아니라 블록 정합 움직임 추정을 확률적으로 모델링함으로써 움직임 벡터의 신뢰도를 계산한다.

또한 제안하는 움직임 분할 과정에서는 군집 과정의 성능 향상을 위해서 움직임 신뢰도 히스토그램(motion reliability histogram)을 이용하여 노이즈와 이상치(outlier)를 제거하기 위한 전처리 과정을 수행한다. 움직임 분할 과정에서 발생하는 영역의 수를 자동적으로 결정해주는 것은 쉬운 일이 아니며, 비디오 객체 분할 시스템의 성능은 움직임 분할 과정에서 결정된 영역의 수에 크게 영향을 받는다^[9,10]. 그러므로 제안하는 시스템에서는 자동으로 객체의 수를 결정해 주기 위해서 군집 과정 중 군집 유효성 지표^[17]를 이용하여 장면 내에 존재하는 의미 있는 객체의 수를 결정하거나 근사할 수 있도록 한다.

객체 추적을 위해 규칙기반 영역처리 방법은 기존의 AM^[9]의 규칙기반 영역처리 방법에 비하여, 일시적으로 정지한 객체의 추적이 가능하게 하고 분할 결과의 시간적 일관성을 유지할 수 있도록 하는 등 보다 안정적인 객체 추적이 가능한 영역 처리 방법을 제안한다.

II. 전역 움직임 추정

제안하는 시스템의 첫 번째 단계는 카메라의 움직임 파라미터를 추정하고 카메라의 움직임을 보상해주는 것이다. 이러한 전역 움직임 추정 과정은 카메라의 움직임이 존재하는 비디오 시퀀스가 입력되더라도 국부 움직임 추정을 가능하게 해준다. MPEG-4 VM (verification model)을 포함한 많은 적용 분야에서는 화소 기반의 전역 움직임 추정 방법을 사용한다^[18]. 화소 기반의 추정 방법은 계산 복잡도가 매우 높아서 실제 응용분야에 적용하기에는 많은 어려움이 있다. 따라서 제안하는 시스템에서는 전체 시스템의 효율성을 향상시키기 위해서 블록 기반의 전역 움직임 추정 방법을 사용한다^[19].

III. 움직임 벡터의 신뢰도

인접한 두 프레임 간의 국부 움직임 추정을 위해 제안하는 시스템에서는 객체들의 움직임을 piecewise-constant

하게 근사할 수 있다고 가정하여 translational 모델^[20]을 적용하고, 블록 정합 알고리즘에 비해서 계산 효율적이고 국부 최소값 문제에 의해서 영향을 덜 받는 계층적 블록 정합 알고리즘을 이용한다^[3,9]. 따라서 piecewise-constant 움직임을 가정하게 되면 piecewise-smooth 움직임을 가정할 때에 비해서 계산 복잡도가 훨씬 낮아 매우 효율적인 구현이 가능하게 된다. 일반적으로 블록 기반의 움직임 추정은 aperture 문제, 가림(occlusion) 문제 외에도 움직이는 물체의 경계선 영역에서의 움직임 불연속성 문제와 같은 오류 발생 원인에 의해서 영향을 받는다^[21].

본 논문에서는 이와 같은 문제점들을 해결하기 위해서 MRF 기반의 MAP 추정 방법을 이용한 움직임 벡터의 신뢰도 계산 방법을 제안한다. 일반적으로 MRF는 공간적인 상관관계를 가지는 확률 변수들을 분석하는데 효과적이기 때문에 움직임 벡터의 신뢰도를 구하는 과정에서 MRF를 이용하면 인접한 블록들의 움직임 벡터들과의 상관관계에 대한 정보를 이용할 수 있게 된다^[22]. MRF의 결합 분포는 Gibbs 분포로 표현되며, 움직임 벡터의 신뢰도를 구하는 과정에서 움직임 벡터들의 공간적인 변화를 반영하기 위해서 적절한 에너지 함수를 정의해야 한다. 움직임 벡터의 신뢰도는 계층적 블록 정합 과정을 확률적으로 모델링하고 추정된 움직임 벡터의 사후 확률을 구함으로써 계산된다^[21]. 계층적 블록 정합 과정의 목적함수는 SAD(sum of absolute difference)로 정의되며, 이로부터 움직임 보상 밝기 차이(motion compensated intensity difference)의 조건부 확률 분포는 Laplacian 분포로 모델링 된다. 목적함수를 최소화 하는 움직임 벡터 \hat{v} 을 구하는 과정은 최대 우도 추정(maximum likelihood estimation) 과정이 된다. 움직임 벡터의 사후 확률을 구하는 과정에서 움직임 벡터의 사전 확률은 식 (1)과 같이 정의된다.

$$P(v_{i,j} = v; \sigma) = \frac{1}{Z(\sigma)} \exp(-\sum \epsilon(v; \sigma)) \quad (1)$$

식 (1)에서 v 는 움직임 벡터를 나타내며 (i, j) 는 블록의 위치를 나타낸다. Z 는 정규화 상수를 나타내며 σ 는 매개 변수의 집합을, 그리고 ϵ 은 에너지 함수를 나타낸다. 에너지 함수는 수평 방향 움직임 벡터에 대한 clique들과 수직 방향 움직임 벡터에 대한 clique들로 구성되며, 수평 방향 움직임 벡터에 대한 clique들에 관 계된 에너지 함수는 식 (2)와 같이 나타난다.

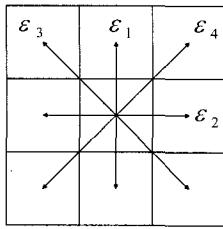
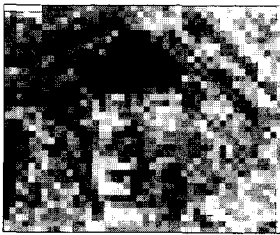


그림 2. 인접한 움직임 벡터들의 공간적인 상관관계를 반영하는 clique들의 방향

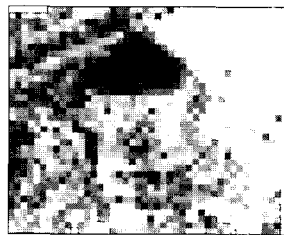
Fig. 2. Directions of cliques reflecting spatial correlation of neighboring motion vectors.



(a)



(b)



(c)

그림 3. Foreman 시퀀스에 대한 움직임 벡터의 신뢰도 계산: (a) 움직임 벡터, (b) MAP 추정 방법 (c) MRF 기반의 MAP 추정 방법

Fig. 3. Motion vector reliability for foreman sequence: (a) motion vector, (b) using MAP estimation, (c) MRF-based MAP estimation.

$$\begin{aligned}
 \epsilon_{1(h)}(\mathbf{v}) &= \left| \hat{\mathbf{v}}_{i-1,j(h)} - 2\mathbf{v}_{i,j(h)} + \hat{\mathbf{v}}_{i+1,j(h)} \right| \\
 \epsilon_{2(h)}(\mathbf{v}) &= \left| \hat{\mathbf{v}}_{i,j-1(h)} - 2\mathbf{v}_{i,j(h)} + \hat{\mathbf{v}}_{i,j+1(h)} \right| \\
 \epsilon_{3(h)}(\mathbf{v}) &= \left| \hat{\mathbf{v}}_{i-1,j-1(h)} - 2\mathbf{v}_{i,j(h)} + \hat{\mathbf{v}}_{i+1,j+1(h)} \right| \\
 \epsilon_{4(h)}(\mathbf{v}) &= \left| \hat{\mathbf{v}}_{i-1,j+1(h)} - 2\mathbf{v}_{i,j(h)} + \hat{\mathbf{v}}_{i+1,j-1(h)} \right|
 \end{aligned} \quad (2)$$

식 (2)에서 $\mathbf{v}_{(h)}$ 는 \mathbf{v} 의 수평 방향 성분을 나타내며, 수직 방향 움직임 벡터에 대한 clique들에 관계된 에너지 함수도 식 (2)와 같은 형태로 나타난다. 그림 2는 식 (2)의 clique들의 방향성을 보여준다. 그림 2와 식 (2)에서 보는 바와 같이 인접한 블록들의 상호 작용을 반영한 구조를 정의하는 과정에서 수평 방향, 수직 방향의 블록들

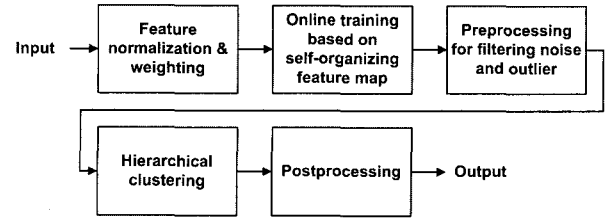


그림 4. 제안하는 움직임 분할 과정

Fig. 4. The proposed motion segmentation flow.

에 대한 움직임 벡터뿐만 아니라 대각 방향의 블록들에 대한 움직임 벡터도 고려한다. 따라서 추정된 움직임 벡터의 사후 확률을 구하는 과정에서 그림 2와 같이 인접한 블록들의 움직임 벡터들과의 관계에 대한 정보를 이용하게 된다.

그림 3의 (b)와 (c)는 (a)의 foreman CIF 시퀀스에 대한 움직임 벡터에 대하여 기존의 비교대상 방법^[21]과 제안하는 방법에 의한 움직임 벡터의 신뢰도를 각각 보여준다. 제안하는 방법에 의한 신뢰도가 기존의 방법에 의한 신뢰도에 비해 인접한 블록들의 움직임 벡터와 유사할수록 높은 값을 보여준다. 즉, 움직임 벡터에 의한 국부 최소값의 두드러짐이 큰 경우라도 인접한 블록들의 움직임 벡터와 큰 차이를 보이는 경우에는 제안하는 방법이 기존의 방법보다 작은 신뢰도 값을 가지며, 움직임 벡터에 대한 국부 최소값의 두드러짐이 작은 경우라도 인접한 블록들의 움직임 벡터와 유사할 경우에는 더 큰 신뢰도를 가지게 된다.

IV. SOM 기반의 계층적 군집 방법을 이용한 분할

분할 과정의 첫 번째 단계에서 이용되는 SOM의 학습은 소스 공간상의 특징 벡터들을 타겟 공간상의 신경소자들로 2차원 매핑시켜준다. 매핑 과정 중 소스 공간에서 근접한 위치에 있는 특징 벡터들은 타겟 공간에서도 근접한 신경소자들로 매핑된다. 계층적 군집 방법은 K-평균 군집 방법과 같은 평면적인 군집 방법과는 달리 초기화, 국부 최소값에 의해 영향을 받지 않는 장점을 가진다^[15]. 계층적 군집 과정에서는 순차적으로 정사각형 격자 구조상에서 가장 유사한 특성을 가지는 두 군집을 병합한다. 따라서 계층적 군집의 초기화 과정에서 SOM을 구성하는 각 신경소자들은 하나의 군집을 이룬다. 분할을 위한 두 단계 접근 방법의 특성상 SOM을 구성하는 신경소자들의 개수가 전체 분할 과정에서 발생할 것으로 예상되는 군집의 개수보다 커야한다.

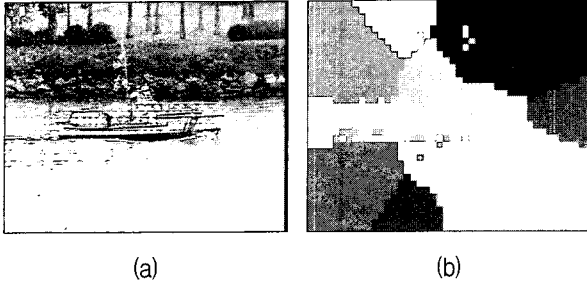


그림 5. *Coastguard* 시퀀스에 대한 SOM의 학습 결과 ($M=16$): (a) 움직임 벡터, (b) 학습 결과

Fig. 5. Training result of SOM for *Coastguard* sequence ($M=16$): (a) motion vector, (b) training result.

1. 움직임 분할

제안하는 움직임 분할 과정의 흐름도는 그림 4와 같다. 움직임 분할의 첫 번째 과정은 특징 벡터의 크기를 정규화하는 과정이다. 특징 벡터 \mathbf{x} 는 국부 움직임 추정 과정에서 이용된 블록에 대한 움직임 벡터의 수평, 수직 방향 성분과 위치의 수평, 수직 방향 성분으로 구성되며, 특징 벡터의 정규화는 식 (3)에 의해 이루어진다.

$$\tilde{x}_i = \frac{x_i - m_i}{\sigma_i} \quad \text{for } i = 1, \dots, 4 \quad (3)$$

식 (3)에서 x_i 는 특징 벡터 \mathbf{x} 를 구성하는 i 번째 특징 성분이며, m_i 와 σ_i 는 각각 i 번째 특징의 평균과 표준편차를 나타낸다. SOM을 이용한 특징 벡터들의 학습 과정에서는 RSST를 이용한 방법^[23]에서처럼 특징 벡터 간의 연결정보를 이용할 수 없기 때문에 움직임 정보뿐만 아니라 위치정보를 특징으로 활용하여 RSST 방법에서 군집 간의 연결정보를 이용한 것과 비슷한 효과를 가지도록 한다. 정규화 과정을 거친 특징 벡터에 대한 가중치 부여 과정은 SOM을 이용한 특징 벡터의 학습 과정에서 움직임 정보의 신뢰도에 적응적인 학습을 가능하게 해주며, 식 (4)에 의해 수행된다.

$$\begin{aligned} \hat{\mathbf{x}} &= \mathbf{w} \tilde{\mathbf{x}} \\ \mathbf{w} &= \text{diag}(w_1, w_2, w_3, w_4) \\ w_1 &= w_2 = m + nc, \quad w_3 = w_4 = 1 - (m + nc) \\ 0 &\leq c \leq 1, \quad 0 < m, n < 1, \quad 0 < m + n < 1 \end{aligned} \quad (4)$$

식 (4)에서 \mathbf{w} 는 가중치 행렬을 나타내며, w_1 와 w_2 는 움직임 정보에 대한 가중치를 나타낸다. c 는 해당 특징 벡터에 대한 움직임 정보의 신뢰도를 나타내며,

움직임 정보에 대한 가중치는 m 과 $m+n$ 사이의 값을 갖게 된다. 따라서 움직임 정보에 대한 가중치와 위치 정보에 대한 가중치의 합은 항상 1이 된다. 움직임 정보의 신뢰도가 0과 1사이의 값을 가지기 때문에 특징 벡터에 대한 가중치 부여 과정에서 색 또는 밝기 정보의 기울기를 이용한 방법^[13]과 달리 임계값이 필요 없고 향상된 학습 결과를 얻도록 해준다.

반복적으로 이루어지는 SOM의 학습은 추출된 임의의 특징 벡터에 대해서 SOM을 구성하는 모든 신경소자에 대한 특징 벡터 \mathbf{a}_m 과의 거리를 계산한다. 이 과정에서 BMU(best matching unit) b 는 식 (5)에 의해서 구해진다.

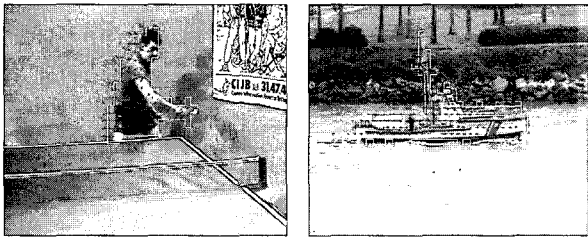
$$b = \arg \min_m \|\hat{\mathbf{x}} - \mathbf{a}_m\| \quad m \in \{1, \dots, M\} \quad (5)$$

식 (5)에서 M 은 신경소자의 개수를 나타낸다. BMU가 정해지면 BMU 및 BMU의 이웃 신경소자들에 대한 업데이트가 식 (6)과 같이 진행된다.

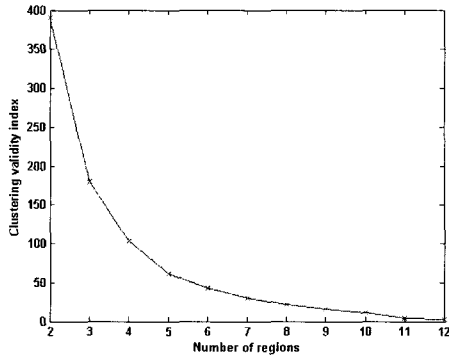
$$\begin{aligned} \mathbf{a}_m(t+1) &= \mathbf{a}_m(t) + \eta(t)h_{bm}(t)[\hat{\mathbf{x}} - \mathbf{a}_m(t)] \\ h_{bm}(t) &= \exp\left(-\frac{\|\mathbf{r}_b - \mathbf{r}_m\|^2}{2\phi^2(t)}\right) \end{aligned} \quad (6)$$

식 (6)에서 t 는 시간을 나타내며, $\eta(t)$ 와 $h_{bm}(t)$ 는 각각 learning rate와 neighborhood kernel을 나타낸다. \mathbf{r}_b 와 \mathbf{r}_m 는 각각 신경소자 b 와 m 의 위치를 나타내며, $\phi(t)$ 는 neighborhood width를 나타낸다. 안정적인 SOM 학습을 위해서 $\eta(t)$ 와 $\phi(t)$ 는 t 가 증가할수록 감소시킨다. 식 (6)에서 보는 바와 같이 각 반복 단계에서 BMU와 BMU의 이웃 신경소자들이 추출된 특징 벡터 $\hat{\mathbf{x}}$ 에 가까워지기 때문에 학습이 끝난 SOM의 각 신경소자들은 이웃 신경소자들과 비슷한 특성을 갖게 된다.

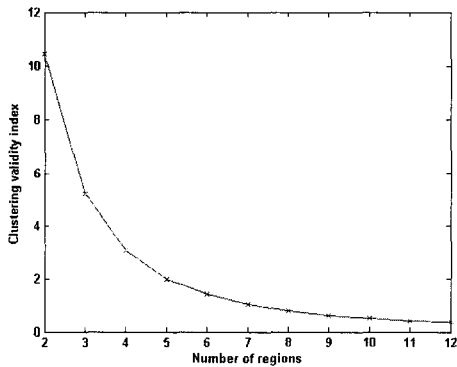
장면 내에 존재하는 각 물리적인 객체들은 하나 또는 여러 개의 신경소자들로 표현되며, 물리적인 객체들과 신경소자들과의 관계는 SOM을 구성하는 신경소자의 개수와 객체의 움직임 특성에 따라서 달라진다. *Coastguard* CIF 시퀀스에 대한 SOM의 학습 결과를 보여주는 그림 5의 (b)에서 동일한 신경소자는 동일한 색상으로 표현된다. 그림 5에서 보는 바와 같이 배경을 포함한 물리적인 객체들은 하나 또는 여러 개의 신경소자들로 표현되기 때문에 SOM 학습 과정은 특징 벡터의 불규칙적인 변화에 덜 민감할 뿐만 아니라 객체들에



(a) (b)



(c)



(d)

그림 6. 움직임 분할 결과 및 군집 유효성 지표: (a), (c) Table tennis 시퀀스, (b), (d) Coastguard 시퀀스
 Fig. 6. The results of motion segmentation and clustering validity index: (a), (c) Table tennis sequence, (b), (d) Coastguard sequence.

대한 특징 공간상에서의 표현도 향상시킬 수 있다.

학습을 거친 SOM에 대한 계층적 군집 과정에서 기존의 방법들^[9,10,13]처럼 입력되는 모든 정보를 이용할 경우 노이즈 및 이상치에 의한 군집 성능의 저하를 최소화하기 위해 전처리 과정이 필요하다. 전처리 과정은 각 신경소자에 속하는 특징 벡터들의 개수에 대한 통계적인 정보와 각 신경소자에 대한 특징 벡터와 신경소자에 속하는 특징 벡터들의 평균 벡터 간의 편이에 대한 통계적인 정보를 이용한 단계^[115]와 신경소자를 구성하

는 특징 벡터들에 대한 움직임 정보의 신뢰도를 이용한 단계로 나누어 다음과 같이 이루어진다.

1. 각 신경소자에 속하는 특징 벡터의 개수 $count_i$ 를 계산하고 모든 신경소자에 대한 평균 m_{count} 과 표준편차 σ_{count} 를 구한다. $count_i < m_{count} + 2\sigma_{count}$ 를 만족하는 신경소자 i 를 제거한다.
2. 각 신경소자에 대한 특징 벡터 a_i 와 평균 벡터 m_i 간의 편이 $deviation_i$ 를 구하고 모든 신경소자에 대한 편이의 평균 $m_{deviation}$ 과 표준편차 $\sigma_{deviation}$ 을 구한다. $deviation_i > m_{deviation} + 2\sigma_{deviation}$ 을 만족하는 신경소자 i 를 제거한다.
3. 통계적인 정보를 이용한 전처리 단계에서 제거된 신경소자와 보간 신경소자를 제외한 모든 신경소자들에 대해서 각 신경소자에 속하는 특징 벡터들에 대한 움직임 정보의 신뢰도를 이용하여 움직임 신뢰도 히스토그램을 구한다. 움직임 신뢰도 히스토그램을 구성하는 특징 벡터들 중 낮은 신뢰도를 가지는 $T\%$ 를 제외한 나머지 특징 벡터들을 이용하여 신경소자에 대한 특징 벡터를 다시 계산한다.

제안하는 움직임 분할 과정에서 이용하는 2차원 정사각형 격자 구조를 가지는 SOM에 대한 군집 과정에서는 군집 간의 연결정보를 이용할 수 있기 때문에 각 신경소자에 대한 특징 벡터를 구성하는 성분들 중 움직임 정보에 대한 성분만을 이용한다. 계층적 군집 과정에서의 병합은 격자 구조상에서 이웃 군집간에 일어나며, 제안하는 움직임 분할 과정에서는 최소 분산 기준을 이용한다^[10,24]. 따라서 계층적 군집의 반복적인 병합 단계에서 해당 군집 개수에 대해서 전역 최소 분산 조건을 만족하는 두 이웃 군집들에 대해 병합이 일어난다. 병합 과정에서는 전처리 과정에서 제거된 신경소자와 보간 신경소자는 고려하지 않는다. 병합 과정은 남아 있는 군집의 개수가 전처리 과정을 거친 SOM이 가질 수 있는 최소 군집의 개수와 같아질 때까지 계속된다.

제안하는 움직임 분할 과정에서는 사용자 관여 없이 계층적 군집 과정에서 발생하는 영역의 수를 결정해 주기 위해서 군집 유효성 지표^[17]를 이용한다. 식 (7)의 해 정의되는 군집 유효성 지표 $I(K)$ 는 영역의 수 K 에 대한 함수이며, SOM이 가질 수 있는 최대 군집의 개수부터 최소 군집의 개수에 대해서 구해진다.

$$I(K) = \left(\frac{1}{K} \times \frac{E_1}{E_k} \times D_k \right)^p$$

$$\text{where } E_K = \sum_{i=1}^K \sum_{j=1}^n u_{ij} \| \mathbf{f}_j - \mathbf{c}_i \|$$

$$D_K = \max \| \mathbf{c}_i - \mathbf{c}_j \| \quad (i \neq j)$$

여기서 u_{ij} 는 특징 벡터들의 분할 상태를 나타내주는 분할 행렬의 구성 요소이고, \mathbf{c}_i 와 \mathbf{f}_j 는 각각 전처리 과정에서 다시 계산된 신경소자의 특징 벡터와 신경소자에 속한 특징 벡터를 나타내며, 움직임 정보만을 이용하기 때문에 2차원 벡터로 표현된다. p 는 군집 간의 대비를 제어하기 위한 매개 변수이다. 즉 E_K 는 군집 내의 거리를 나타내며, D_K 는 군집 간의 거리를 나타낸다.

따라서 장면을 구성하는 물리적인 객체의 수는 식 (7)에 의해서 I 를 최대로 하는 영역의 수로 근사되는데 물리적인 객체의 수와 최적 영역의 수의 관계는 객체들의 움직임 특성에 따라 달라지며, 객체의 움직임 특성이 복잡한 경우에는 최적 영역의 수가 실제 장면 내에 존재하는 객체의 수보다 큰 값을 가지게 된다. 그러므로 계층적 군집 과정이 끝난 후에는 비슷한 움직임 특성을 가지는 이웃 영역 간의 병합 과정이 필요하게 된다.

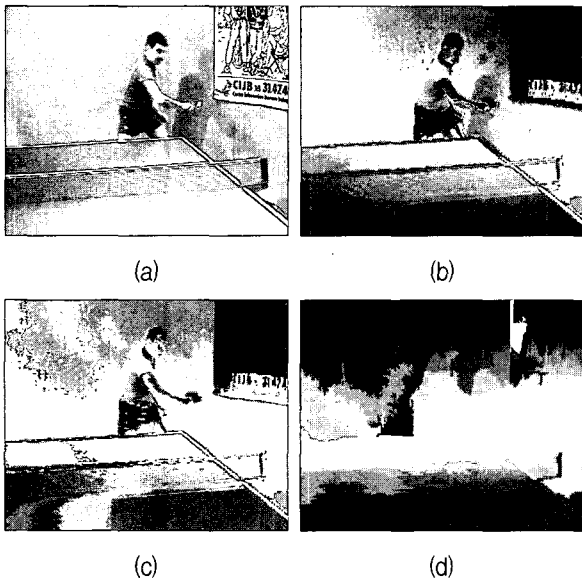


그림 7. Table tennis 시퀀스에 대한 색 분할 결과: (a) Table tennis 시퀀스, (b) SOM 학습, (c) 계층적 군집, (d) 후처리

Fig. 7. Color segmentation results for table tennis sequence: (a) Table tennis sequence, (b) SOM training, (c) hierarchical clustering, (d) postprocessing.

움직임 분할의 후처리 과정에서는 SOM 학습, 계층적 군집을 거쳐서 생성된 분할 결과에 대한 연결 제약 조건을 고려하고 움직임 추정 에러에 의해서 잘못 나누어진 영역들이 병합된다. SOM 학습 결과에서 동일한 신경소자에 속하는 블록들이 항상 연결되어 있는 것은 아니기 때문에 영역을 구성하는 블록들에 대한 연결 제약 조건이 필요하다. 잘못 나누어진 영역들의 병합 과정에서는 첫째로 사용자가 미리 정해놓은 영역의 크기보다 작은 영역들의 이웃 영역으로의 병합이 필요하고, 둘째로 비슷한 움직임 특성을 가지는 이웃 영역들에 대한 병합이 필요하다. 특히, 하나의 물리적인 객체가 비슷한 움직임 특성을 가지는 여러 영역들로 나누어져 있을 경우 병합하게 되면 향상된 분할 결과를 얻을 수 있게 된다. 그림 6은 table tennis (352 × 240)와 coastguard CIF 시퀀스에 대한 움직임 분할 결과 및 움직임 벡터의 신뢰도를 보여준다. 그림 6의 (a)의 경우 움직임 벡터의 신뢰도 결과에서 보는 바와 같이 aperture 문제와 객체의 그림자에 의해 움직임 추정 에러가 발생했음에도 불구하고 분할이 잘 이루어진 것을 알 수 있다. 그림 6의 (c)와 (d)는 움직임 분할 결과에 대한 군집 유효성 지표를 보여준다. 식 (7)의 군집 유효성 지표에 의해서 장면을 구성하는 물리적인 객체의 수가 최적 영역의 수로 근사되었음을 볼 수 있다.

2. 색 분할

계산 효율적이고 정확한 색 분할을 위해 움직임 분할에서와 마찬가지로 두 단계 접근 방법을 이용한다. 그러나 색 분할 과정에서는 움직임 분할 과정에서와 같이 특징 벡터에 대한 신뢰도를 이용하여 가중치를 부여하거나 노이즈 및 이상치 제거를 위한 전처리 과정을 거치지 않는다. 색 분할 과정은 실제 객체들의 경계선이 색 분할 결과를 구성하는 영역들의 경계선과 일치하도록 해준다. 따라서 분할 과정에서 발생한 각 영역들은 반드시 하나의 물리적인 객체에 포함되어야 한다^[9]. 색 분할을 위한 특징 추출 과정에서는 perceptually uniform component를 제공하는 CIELAB 색 공간이 사용된다^[6]. 또한 특정 영역이 다른 영역들에 비해서 상대적으로 커지는 것을 방지하기 위해서 SSE(sum of squared error) 목적 함수에 기반한 step-wise optimal 계층적 군집 방법을 이용한다. 색 분할의 후처리 과정에서는 영역들의 연결 제약 조건이 고려되며, 작은 영역들에 대한 병합이 진행된다. 작은 영역들에 대한 병합 과정에서는 사용자가

미리 정해놓은 영역의 크기보다 작은 영역들을 영역의 크기순으로 정렬한다. 가장 작은 크기를 가지는 영역에 대한 병합이 반복적으로 일어나는데, 해당 영역보다 크고 가장 비슷한 색 특징을 가지는 이웃 영역으로 병합된다. 그림 7의 (d)에서 보는 바와 같이 색 분할 결과의 영역들의 경계선이 물리적인 객체들의 경계선과 거의 일치하는 것을 볼 수 있다.

3. 분할 알고리즘의 계산 복잡도

제안하는 시스템에서 움직임 분할과 색 분할을 위해 이용된 두 단계 접근 방법은 기존의 RSST를 이용한 방법과 M-RSST를 이용한 방법에 비해서 낮은 계산 복잡도를 갖는다. 두 단계 접근 방법의 경우 전체 과정에 대한 계산 복잡도 중 SOM 학습에 대한 계산 복잡도가 차지하는 비율이 매우 크다^[14]. 그림 8은 그림 4의 움직임 분할을 구성하는 각 과정의 처리 시간 비율을 보여주며, 색 분할에 대한 처리 시간 비율도 비슷한 양상을 나타낸다. 그림 8에서 보는 바와 같이 두 단계 접근 방법의 계산 복잡도는 SOM 학습에 대한 계산 복잡도에 의해서 결정되며, SOM이 M 개의 신경소자로 이루어질 경우 두 단계 접근 방법에 대한 계산 복잡도의 하한은 $O(M)$ 으로 근사된다^[25]. 반면에 RSST를 이용한 방법의 경우 특징 벡터들의 개수가 N 일 때 $O(N^2)$ 의 계산 복잡도를 가진다^[26]. 또한 M-RSST를 이용한 방법에서는 영상 피라미드에 대한 scale 매개 변수가 P 일 경우 가장 낮은 해상도를 갖는 영상에 대해서만 $O\left(\frac{N}{4^P}\right)^2$ 의 계산 복잡도를 가지게 된다.

두 단계 접근 방법에서 M 은 N 에 비해서 작은 값을 가지기 때문에 SOM 기반의 계층적 군집 방법을 이용할 경우 RSST를 이용하거나 M-RSST를 이용할 때에 비해서 연산량을 크게 감소시킬 수 있다. 표 1은 Pentium IV 1.8 GHz의 Window XP 환경에서 Visual C++를 이용하여 시스템을 구현한 경우 움직임 분할을 위해 RSST를 적용한 경우와 두 단계 접근 방법을 적용한 경우의 상대적인 처리 시간 비율을 보여주며,

표 1. 상대적인 처리 시간 비율
Table 1. Relative processing time ratio.

영상 크기	RSST	두 단계 접근 방법
176×144	1	21.8(×10 ⁻³)
352×288	1	8.3(×10 ⁻³)
720×576	1	0.78(×10 ⁻³)

□ SOM training ■ hierarchical clustering
□ remaining processes

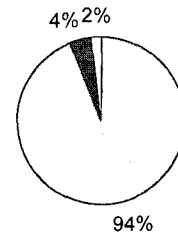


그림 8. 움직임 분할의 처리 시간 비율
Fig. 8. Ratio of processing time of the motion segmentation.

RSST를 적용한 경우를 기준으로 영상의 크기에 따라 나타내었다. 움직임 분할 과정에서 4×4 또는 5×5 구조를 가지는 SOM을 이용하였으며, 10번의 epoch에 걸쳐서 학습을 수행하였다. 두 단계 접근 방법을 적용한 경우의 실제 처리 시간은 영상의 크기순으로 23, 47, 134 msec이고 표 1과 [12]에 제시된 처리 시간 비교에서 나타난 바와 같이 제안하는 방법은 기존의 방법들에 비해서 연산량이 적으며, 특히 영상의 크기가 커질수록 연산량의 감소 정도가 증가한다.

V. 규칙 기반 영역 처리

본 논문에서 제안하는 규칙 기반 영역 처리 과정의 블록 흐름도는 그림 9와 같다. 그림 9에서 R_t^I , R_t^M , R_t^{MC} 는 각각 현재 프레임에 대한 색 분할 결과, 움직임 분할 결과, 움직임 보상 결과를 나타내며, R_t^O 는 최종 분할 결과를 나타낸다. 움직임 보상 과정에서는 전역 움직임 추정 및 국부 움직임 추정 과정에서 구해진 움직임 정보와 이전 프레임의 분할 결과를 이용하여 현재 프레임에 대한 분할 결과의 예측을 가능하도록 한다. 규칙 기반 영역 처리의 첫 번째 과정은 투영 연산자, 군집 연산자^[9]를 이용하여 R_t^M , R_t^{MC} 를 구성하는

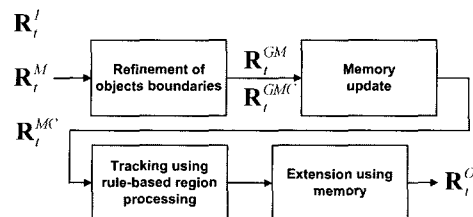


그림 9. 제안하는 규칙 기반 영역 처리 과정
Fig. 9. Flowchart for the proposed rule-based region processing.

영역들의 경계선을 물리적인 객체들의 경계선에 가깝도록 만들어주는 것이며, 첫 번째 과정에서 발생하는 결과는 R_t^{GM} , R_t^{GMC} 로 표현된다.

제안하는 규칙 기반 영역 처리 방법에서는 일시적으로 정지한 객체에 대한 추적을 가능하게 할 뿐만 아니라 분할 결과의 시간적 일관성 유지를 위해 동적 메모리^[20]를 이용하여, 일시적으로 정지한 객체에 대한 추적이 불가능한 European COST 211 group이 제안한 규칙 기반 영역 처리 방식의 한계를 극복한다. European COST 211 group이 제안한 AM에서는 객체들의 추적을 위해서 auxiliary set B 를 이용한다. AM에서 제안된 규칙들은 B 의 공집합 여부에 영향을 받으며, B 가 공집합일 경우 해당 B 에 대한 이전 프레임의 객체와 현재 프레임의 객체가 정합되었다고 가정한다. 그러나 다음 두 가지 경우에 의해 객체가 정합되어야 하는 경우에도 B 는 공집합이 아닐 수 있게 된다. 첫째는 이전 프레임과 현재 프레임에 존재하는 동일한 객체임에도 불구하고 해당 객체의 움직임이 영역이 같지 않은 경우이고, 둘째는 움직임 추정 에러에 의해서 움직임 보상이 정확하게 이루어지지 않은 경우이다. B 의 공집합 여부를 결정하는 과정이 정확하게 이루어지지 않을 경우 규칙 기반 영역 처리 과정에서 정확한 객체추적이 불가능하다. 따라서 제안하는 영역 처리 방법에서는 위의 두 가지 경우에도 객체추적이 가능하도록 하는 교차 투영법을 소개한다. 교차투영은 식 (8)과 같이 $R_t^{GM} = \bigcup_{m=0}^{m_{\max}-1} GM_{t,m}$, $R_t^{GMC} = \bigcup_{c=0}^{c_{\max}-1} GMC_{t,c}$ 간에 일어나며, 객체 추적을 위한 규칙들은 교차 투영 결과에 의해서 결정된다.

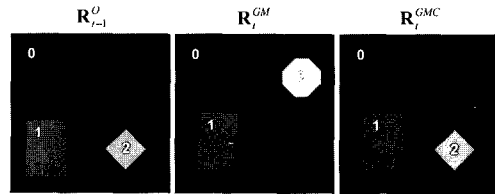
$$\begin{aligned} P_{MC}(m) &= \arg \max_c \{a(GMC_{t,c} \cap GM_{t,m})\} \\ P_M(c) &= \arg \max_m \{a(GM_{t,m} \cap GMC_{t,c})\} \\ 0 \leq m \leq (m_{\max} - 1), 0 \leq c \leq (c_{\max} - 1) \end{aligned} \quad (8)$$

식 (8)에서 a 는 AM에서와 같이 area operator^[9]를 나타낸다. $GMC_{t,0}$, $GM_{t,0}$ 은 각각 R_t^{GMC} , R_t^{GM} 을 구성하는 영역들 중 배경 영역을 나타낸다. 객체 추적을 위한 규칙들을 정의하기 위해 식 (9), (10)과 같이 교차 관계 함수 $N_M(m)$, $N_{MC}(c)$ 를 정의한다. 식 (8-10)에 의해 $R_t^O = \bigcup_{g=0}^{g_{\max}-1} O_{t,g}$ 를 구하기 위한 규칙들이 표 2와 같이 다섯 가지 경우로 나누어진다. 표 2에서 $MEM_{t,c}$ 는 $GMC_{t,c}$ 영역에서의 동적 메모리를 나타낸다. 표 2에서

표 2. 규칙 기반 영역 처리

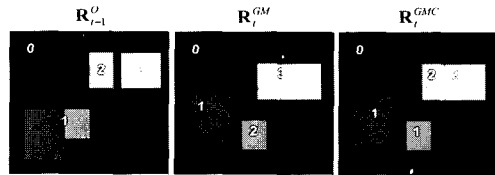
Table 2. Rule-based region processing.

	Condition	Processing
Rule 1 (Matched)	$N_M(m)=1$ and $N_{MC}(c)=1$ ($P_M(c)=m$)	$O_{t,c} = GM_{t,m} \vee (GMC_{t,c}, MEM_{t,c})$
Rule 2 (Newly appearing single moving object)	$N_M(m)=0$ and $P_{MC}(m)=0$	$O_{t,new} = GM_{t,m}$
Rule 3 (Stopping single moving object)	$P_M(c)=0$	$O_{t,c} = GMC_{t,c} \wedge MEM_{t,c}$
Rule 4 (Multiple moving objects are merged to single moving object)	$N_M(m) \geq 2$ ($P_{MC}(m)=c$)	$O_{t,c} = GM_{t,m}$
Rule 5 (Single moving object is divided to multiple moving objects)	$N_M(m)=1$ and $N_{MC}(c) \geq 2$ ($P_M(c)=m$)	$O_{t,c} = GM_{t,m}$
	$N_M(m)=1$ and $P_{MC}(m) \neq 0$ or $N_M(m)=1$ and $N_{MC}(c)=0$ ($P_M(c)=m$)	$O_{t,new} = GM_{t,m}$



m	$P_{MC}(m)$	$N_M(m)$	c	$P_{MC}(c)$	$N_{MC}(c)$
1	1	1	1	1	1
2	0	0	2	0	0

(a)



m	$P_{MC}(m)$	$N_M(m)$	c	$P_{MC}(c)$	$N_{MC}(c)$
1	1	1	1	1	2
2	1	0	2	3	0
2	3	2	3	3	1

(b)

그림 10. 규칙 기반 영역 처리에 대한 도식적인 예: (a) 규칙 1, 규칙 3, (b) 규칙 2, 규칙 4, 규칙 5

Fig. 10. Schematic example for rule-based region processing: (a) rule 1 and rule 3, (b) rule 2, rule 4, and rule 5.

보는 바와 같이 $O_{t,g}$ 를 구하는 과정에서 동적 메모리는 이전 프레임과 현재 프레임의 객체가 정합된 경우(rule 1)와 일시적으로 객체가 정지한 경우(rule 3)에만 이용된다. 이전 프레임의 분할 결과에 대한 움직임 보상은 좀 더 정확한 분할 결과를 얻을 수 있게 해줄 뿐만 아니라 객체들의 추적을 가능하게 한다. 그림 10은 표 2의 규칙들에 대한 도식적인 예를 보여준다. 후처리 과정에서는 움직임 분할의 후처리 과정에서와 마찬가지로 영역들의 연결 제약 조건을 고려하고 사용자가 미리 정해

높은 영역의 크기보다 작은 영역들에 대한 병합 과정이 필요하다.

$$N_M(m) = \sum_{c=1}^{c_{\max}-1} f(m,c) \tag{9}$$

$$f(m,c) = \begin{cases} 1 & \text{if } P_M(c) = m \\ 0 & \text{otherwise} \end{cases}$$

$$N_{MC}(c) = \sum_{m=1}^{m_{\max}-1} g(c,m) \tag{10}$$

$$g(c,m) = \begin{cases} 1 & \text{if } P_{MC}(m) = c \\ 0 & \text{otherwise} \end{cases}$$

VI. 실험 결과 및 토의

제안하는 시스템의 효율성을 알아보기 위해서 *table tennis* 시퀀스(352×240), *coastguard* CIF 시퀀스, *bream* CIF 시퀀스를 실험에 이용한다.

그림 11의 *table tennis* 시퀀스에서 *player* 객체는 비강체 움직임의 특성을 가지며, *player* 객체를 구성하는 영역은 강체 움직임의 특성을 가지는 여러 영역들로 구성된다. 그림 11에서 보는 바와 같이 제안하는 시스템이 비교대상 방법에 비해서 정확하게 객체들의 경계선을 보여준다.

Coastguard 시퀀스에서는 수평 방향의 카메라 움직임이 존재한다. 그림 12에서 보는 바와 같이 제안하는 시스템이 비교대상 방법보다 *boat* 객체에 대해 더욱 정확한 분할 결과를 제공해준다. 특히 비교대상 방법에 의

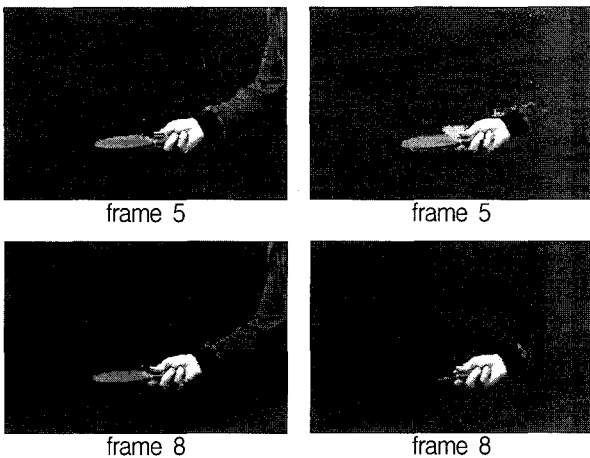


그림 11. *Table tennis* 시퀀스에 대한 분할 결과: 제안하는 방법(좌측) 및 COST 211 AM(우측)

Fig. 11. Segmentation results for *table tennis* sequence: by the proposed method(left column) and COST 211 Analysis Model(right column).

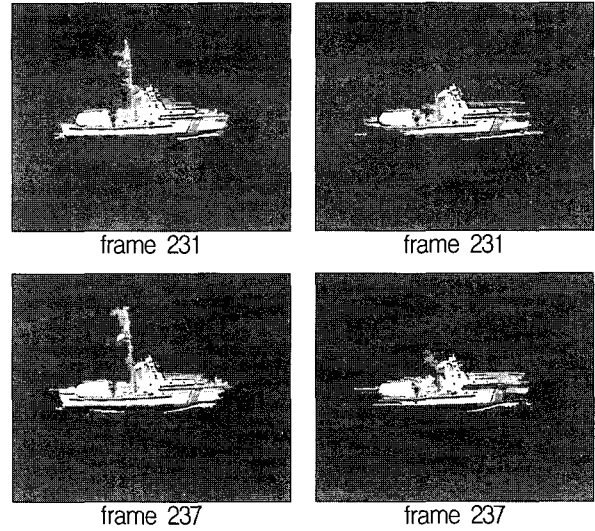


그림 12. *Coastguard* 시퀀스에 대한 분할 결과: 제안하는 방법(좌측) 및 COST 211 AM(우측)

Fig. 12. Segmentation results for *Coastguard* sequence: (by the proposed method(left column) and COST 211 Analysis Model(right column)).

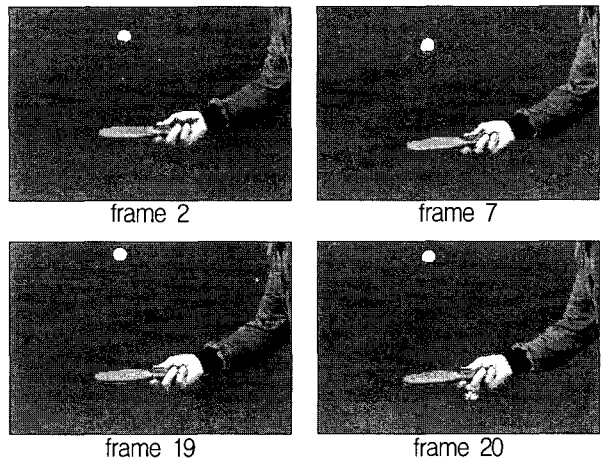


그림 13. *Table tennis* 시퀀스에 대한 분할 결과

Fig. 13. Segmentation results for *Table tennis* sequence.

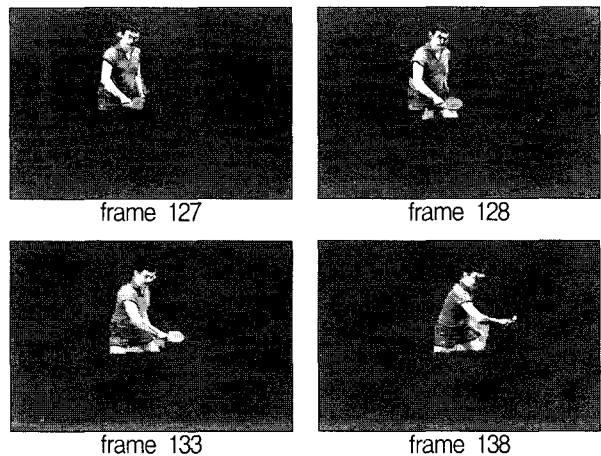


그림 14. *Table tennis* 시퀀스에 대한 분할 결과

Fig. 14. Segmentation results for *Table tennis* sequence.

해서 추출되지 못한 *flag*를 포함하는 *boat* 객체의 윗부분도 제안하는 시스템에 의해서 성공적으로 추출되었음을 볼 수 있다.

Table tennis 시퀀스에 대해서는 일부 프레임에서 *ball* 객체 추출에 실패한 기존의 방법^[6]과는 달리 제안하는 시스템은 그림 13에서 보는 바와 같이 모든 프레임에 걸쳐서 성공적으로 *ball* 객체를 추출한다.

그림 14와 그림 15는 그림 11-13에서 보여준 프레임들과는 다른 성격을 가지는 프레임들에 대해 제안하는 방법을 적용한 분할 결과를 보여준다. 비강체 움직임의 특성을 갖는 그림 14의 *player* 객체의 그림자 영역이 객체의 정확한 경계선을 구하는 데 방해 요소가 될 수 있다. 그림 15의 *boat* 객체들의 경우 서로 다른 움직임 방향과 속도를 가진다. 그림에도 불구하고 그림 14와 그림

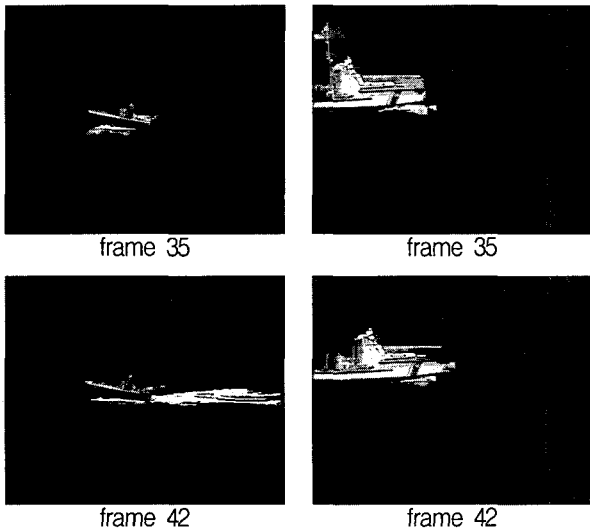


그림 15. *Coastguard* 시퀀스에 대한 분할 결과
Fig. 15. Segmentation results for *Coastguard* sequence.

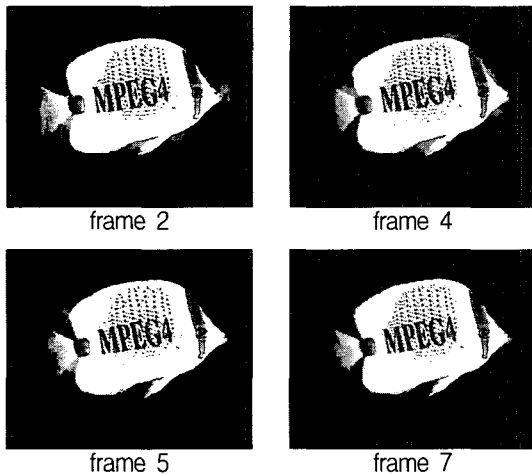


그림 16. *Bream* 시퀀스에 대한 분할 결과
Fig. 16. Segmentation results for *Bream* sequence.

15에서 보는 바와 같이 제안하는 시스템에 의해서 실제 객체들의 경계선에 가까운 분할 결과를 얻을 수 있음을 볼 수 있다.

그림 16은 *bream* 시퀀스에 대한 제안하는 시스템에

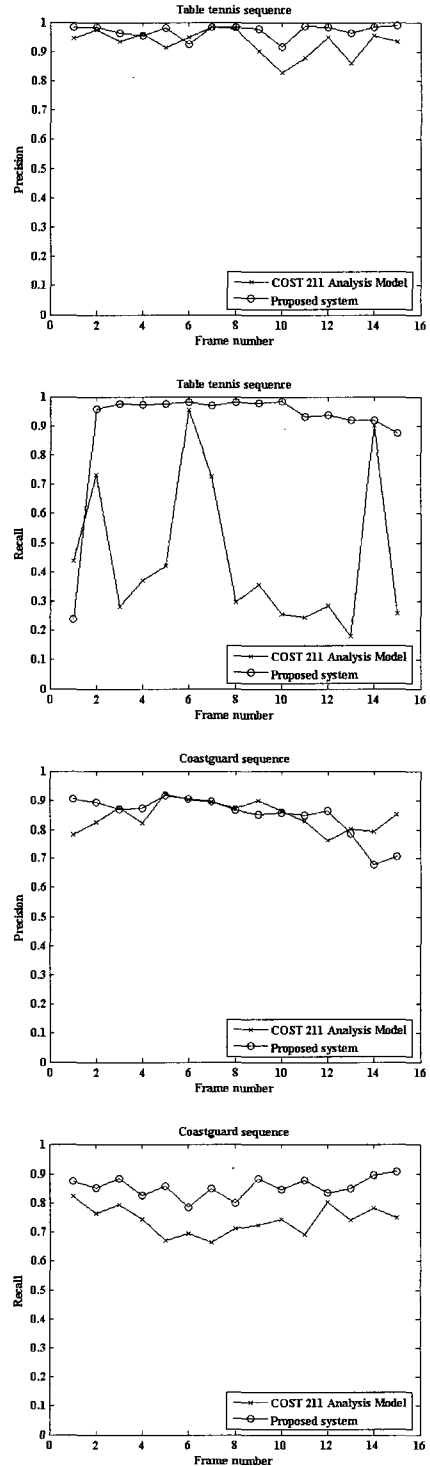


그림 17. *Table tennis* 시퀀스와 *coastguard* 시퀀스에 대한 precision과 recall
Fig. 17. Precision and recall for the *table tennis* sequence and *coastguard* sequence.

의한 분할 결과를 보여준다. 그림 16의 *fish* 객체가 복잡한 움직임 특성을 가짐에도 불구하고 제안하는 시스템에 의해서 효과적인 분할이 이루어졌음을 볼 수 있다.

본 논문에서는 제안하는 시스템의 효율성을 평가하기 위해서 식 (11)과 식 (12)의 객관적 평가 기준을 이용한다^[2]. 객관적 평가는 ground truth와 비디오 객체 분할 알고리즘에 의해서 구해진 분할 결과를 비교함으로써 이루어진다. *Table tennis* 시퀀스와 *coastguard* 시퀀스에 대한 객관적 평가 결과를 보여주는 그림 17에서 제안하는 시스템에 의해서 높은 precision과 recall을 얻을 수 있음을 알 수 있다.

$$\text{precision} = \frac{\text{Actual object pixels extracted}}{\text{Total number of pixels extracted}} \quad (11)$$

$$\text{recall} = \frac{\text{Actual object pixels extracted}}{\text{Total number of actual object pixels}} \quad (12)$$

VII. 결 론

본 논문에서는 두 단계 접근 방법을 이용한 계산 효율적이고 노이즈에 강건한 자동 비디오 객체 분할 시스템을 제안하였다. 기존의 비디오 객체 분할 시스템에서 특징 벡터들을 직접적으로 군집한 것과는 달리 제안하는 시스템에서는 SOM 기반의 계층적 군집 방법을 이용하여 특징 벡터들을 군집함으로써 계산 효율적인 동시에 노이즈에 강건한 비디오 객체 분할을 가능하게 해주었다. 또한 움직임 분할의 성능 향상을 위해 움직임 벡터들의 공간적인 변화를 반영하는 새로운 움직임 벡터의 신뢰도 계산 방법을 제안하였으며, 이 과정에서 MRF 기반의 MAP 추정 방법을 이용하였다. 노이즈에 강건한 움직임 분할을 위해서 움직임 신뢰도 히스토그램을 이용한 전처리 과정을 제안하였으며, 사용자 상호작용의 개입 없이 움직임 분할이 가능하도록 군집 유효성 지표를 이용하였다. 안정적인 객체 추적과 분할 결과의 시간적 일관성 유지가 가능하도록 하기 위해서 교차 투영과 동적 메모리를 이용한 규칙 기반 영역 처리 방법을 제안하였다. 다양한 특성을 가지는 비디오 시퀀스들에 대해서 실험한 결과 제안하는 시스템이 기존의 방법에 비해서 연산량을 감소시킬 뿐만 아니라 정확한 분할 결과를 제공할 수 있었다. 따라서 높은 효율성과 정확성을 제공하는 제안하는 시스템은 내용 기반 비디오 애플리케이션에서 유용하게 사용될 수 있을 것이다.

참 고 문 헌

- [1] Y. Liu and Y. F. Zheng, "Video object segmentation and tracking using ψ -learning classification," *IEEE Trans. on Cir. and Sys. for Video Tech.*, vol. 15, pp. 885-899, 2005.
- [2] H. Xu, A. Younis, and M. Kabuka, "Automatic moving object extraction for content-based applications," *IEEE Trans. on Cir. and Sys. for Video Tech.*, vol. 14, pp. 796-812, 2004.
- [3] T. Meier and K. N. Ngan, "Video segmentation for content-based coding," *IEEE Trans. on Cir. and Sys. for Video Tech.*, vol. 9, pp. 1190-1203, 1999.
- [4] 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. on Cir. and Sys. for Video Tech.*, vol. 9, pp. 1216-1226, 1999.
- [5] A. Doulamis, N. Doulamis, K. Ntalianis, and S. Kollias, "An efficient fully unsupervised video object segmentation scheme using an adaptive neural-network classifier architecture," *IEEE Trans. on Neu. Net.*, vol. 14, pp. 616-630, 2003.
- [6] V. Mezaris, I. Kompatsiaris, and M. G. Strintzis, "Video object segmentation using Bayes-based temporal tracking and trajectory-based region merging," *IEEE Trans. on Cir. and Sys. for Video Tech.*, vol. 14, pp. 782-795, 2004.
- [7] J. Y. A. Wang and E. H. Adelson, "Representing moving images with layers," *IEEE Trans. on Imag. Proc.*, vol. 3, pp. 625-638, 1994.
- [8] D. W. Murray and B. F. Buxton, "Scene segmentation from visual motion using global optimization," *IEEE Trans. on Pat. Anal. and Mach. Intel.*, vol. 9, pp. 220-228, 1987.
- [9] A. A. Alatan, L. Onural, M. Wollborn, R. Mech, E. Tuncel, and T. Sikora, "Image sequence analysis for emerging interactive multimedia services-the European COST 211 framework," *IEEE Trans. on Cir. and Sys. for Video Tech.*, vol. 8, pp. 802-813, 1998.
- [10] E. Tuncel and L. Onural, "Utilization of the recursive shortest spanning tree algorithm for video-object segmentation by 2-D affine motion modeling," *IEEE Trans. on Cir. and Sys. for Video Tech.*, vol. 10, pp. 776-781, 2000.
- [11] D. Wang, "Unsupervised video segmentation based on watersheds and temporal tracking," *IEEE Trans. on Cir. and Sys. for Video*

- Tech.*, vol. 8, pp. 539-546, 1998.
- [12] A. D. Doulamis, N. Doulamis, and S. Kollas, "Non-sequential video content representation using temporal variation of feature vectors," *IEEE Trans. on Cons. Elec.*, vol. 46, pp. 758-768, 2000.
- [13] J. Kim and T. Chen, "Multiple feature clustering for image sequence segmentation," *Pat. Rec. Let.*, vol. 22, pp. 1207-1217, 2001.
- [14] J. Vesanto and E. Alhoniemi, "Clustering of the self-organizing map," *IEEE Trans. on Neu. Net.*, vol. 11, pp. 586-600, 2000.
- [15] S. Wu and T. W. S. Chow, "Clustering of the self-organizing map using a clustering validity index based on inter-cluster and intra-cluster density," *Pat. Rec.*, vol. 37, pp. 175-188, 2004.
- [16] R. Cucchiara, A. Prati, and R. Vezzani, "Real-time motion segmentation from moving cameras," *Real-Time Imag.*, vol. 10, pp. 127-143, 2004.
- [17] U. Maulik and S. Bandyopadhyay, "Performance evaluation of some clustering algorithms and validity indices," *IEEE Trans. on Pat. Anal. and Mach. Intel.*, vol. 24, pp. 1650-1654, 2002.
- [18] F. Dufaux and J. Konrad, "Efficient, robust, and fast global motion estimation for video coding," *IEEE Trans. on Imag. Proc.*, vol. 9, pp. 497-501, 2000.
- [19] Y. Su, M.-T. Sun, and V. Hsu, "Global motion estimation from coarsely sampled motion vector field and the applications," *IEEE Trans. on Cir. and Sys. for Video Tech.*, vol. 15, pp. 232-242, 2005.
- [20] Y. Tsaig and A. Averbuch, "Automatic segmentation of moving objects in video sequences: a region labeling approach," *IEEE Trans. on Cir. and Sys. for Video Tech.*, vol. 12, pp. 597-612, 2002.
- [21] I. Patras, E. A. Hendriks, and R. L. Lagendijk, "Confidence measures for block matching motion estimation," *IEEE Int. Conf. on Imag. Proc.*, vol. 2, pp. 277-280, 2002.
- [22] R. Neher and A. Srivastava, "A Bayesian MRF framework for labeling terrain using hyperspectral imaging," *IEEE Trans. on Geo. and Rem. Sen.*, vol. 43, pp. 1363-1374, 2005.
- [23] O. J. Morris, J. M. Lee, A. G. Constantinides, "A unified method for segmentation and edge detection using graph theory," *IEEE Int. Conf. on Acou, Spe., and Sig. Proc.*, vol. 11, pp. 2051-2055, 1986.
- [24] M. Y. Kiang, "Extending the Kohonen self-organizing map networks for clustering analysis," *Comp. Stat. & Data Anal.*, vol. 38, pp. 161-180, 2001.
- [25] P. Xu, C.-H. Chang, and A. Paplinski, "Self-organizing topological tree for online vector quantization and data clustering," *IEEE Trans. on Sys., Man and Cyb.*, vol. 35, pp. 515-526, 2005.
- [26] S. H. Kwok, A. G. Constantinides, and W.-C. Siu, "An efficient recursive shortest spanning tree algorithm using linking properties," *IEEE Trans. on Cir. and Sys. for Video Tech.*, vol. 14, pp. 852-863, 2004.

 저 자 소 개



정 찬 호(학생회원)
 2004년 2월 서강대학교
 전자공학과 학사
 2006년 2월 서강대학교
 전자공학과 석사
 2006년 3월~현재 LG전자 연구원
 <주관심분야 : 영상신호처리, 컴
 퓨터비전, DTV>

김 경 환(정회원)
 대한전자공학회논문지 제41권 SP편 제5호 참조