

논문-98-3-1-05

## 확률적 클러스터링에 의한 움직임 파라미터 추정과 세그멘테이션

정 차 근\*

### Motion Parameter Estimation and Segmentation with Probabilistic Clustering

Cha Keon Cheong\*

#### 요 약

본 논문에서는 콤팩트한 동영상 표현과 객체기반의 generic한 동영상압축을 위한 파라미터릭 움직임 모델의 파라미터 추정과 세그멘테이션 기법에 관해서 기술한다. 동영상의 optical flow와 같은 국소적 움직임 정보와 파라미터릭 움직임 모델의 특징을 이용해서 영상의 콤팩트한 구조적 표현을 추출하기 위해, 본 논문에서는 2 스텝의 과정 즉, 초기 영역을 추출하는 과정과, 파라미터릭 움직임 파라미터의 추정과 세그멘테이션을 동시에 수행하는 과정으로 구성된 새로운 알고리즘을 제안한다. 혼합 모델의 ML 추정에 의거한 확률적 클러스터링에 의해 움직임 물체의 움직임과 형상을 반영한 초기영역을 추출하고, 파라미터릭 움직임 모델을 사용해서 각각의 초기 영역마다 움직임 파라미터를 추정하고 세그멘테이션을 수행한다. 또한, CIF 표준 동영상을 사용한 모의 실험을 통해 본 제안 알고리즘의 유효성을 평가한다.

#### Abstract

This paper addresses a problem of extraction of parameteric motion estimation and structural motion segmentation for compact image sequence representation and object-based generic video coding. In order to extract meaningful motion structure from image sequences, a direct parameteric motion estimation based on a pre-segmentation is proposed. The pre-segmentation which considers the motion of the moving objects is carried out based on probabilistic clustering with mixture models using optical flow and image intensities. Parametric motion segmentation can be obtained by iterated estimation of motion model parameters and region reassignment according to a criterion using Gauss-Newton iterative optimization algorithm. The efficiency of the proposed method is verified with computer simulation using CIF real image sequences.

#### I. 서 론

정보화 사회의 진전에 따라, 다양한 영상 품질의 전송을 포함해서 제작, 검색, 편집, 원격제어등과 같은 여러 가지의 많은 응용에 대응할수 있는 영상 처리 기법의 개발이 요구되고 있다. DCT를 중심으로 하는 파형부호화 기

법은 국제표준화가 진행되어 실용화되고 있지만, 영상의 작은 참조 범위내에서의 상관에 의거해서 처리되기 때문에 초저 비트 영상 부호화나 검색등의 응용에는 적합치 않다.<sup>[1][2][17]</sup> 따라서, 입력 동영상으로부터 내용기반이나 객체기반의 구조정보를 추출, 동영상을 보다 콤팩트한 표현으로 기술함으로써 동영상 해석이나 새로운 부호화 기법의 개발에 적합한 알고리즘의 개발이 기대되고 있다.<sup>[12][16][18][19][21]</sup>

동영상에는 움직임의 유무에 따라 움직임 영역과 정지 영역, 움직임에 의해서 물체가 은폐되거나 새롭게 나타나

\* 호서대학교 제어계측공학과

Dept. of Control &amp; Instrumentation, college of Eng. Hoseo University

는 영역이 발생한다. 또한, 움직임 영역은 카메라등의 움직임에 의해 global motion을 갖는 영역과, 물체의 국소적 움직임에 의한 local motion을 갖는 영역으로 나눌수 있다. 따라서, 다양한 응용에 적합한 동영상의 해석을 위해서는, (i)물체의 움직임에 따른 각 영역의 추출, (ii)각 영역마다 콤팩트한 움직임 정보의 복원, (iii)각 영역의 기술에 필요한 최소한의 휘도나 색 정보의 추출, (iv)각 영역간 공간 정보의 검출 등과 같은 과정이 필요하다.

움직임이나 각 프레임의 휘도값에 의거해서 동영상의 각 영역을 분할하기 위한 많은 연구가 이루어져 왔다. 초기의 연구<sup>1)3)20)</sup>에서는 주로 프레임간의 국소적 움직임 정보나 텍스처 정보를 이용해서 영역을 분할하고자 시도했지만, 분할된 영역에는 움직임 물체의 형상이나 움직임에 관한 정보가 충분히 반영되어 있지 않다. 이러한 문제점을 개선해서 동영상을 보다 콤팩트한 표현으로 기술하기 위해, 단순한 국소적 움직임이나 텍스처 정보에 의한 영역분할이 아니라, global motion, 움직임 물체의 영역 형상, 각 영역의 깊이 정보등을 고려한 영역 분할과 동영상 기술에 관한 연구가 최근에 많은 주목을 받고 있다.<sup>11)4)6)18)</sup>

이들의 기법에서는 움직임 물체를 파라미터릭 움직임 모델에 근사시켜 움직임 파라미터의 추정과 동시에 세그멘테이션을 실행하고 있다. 이때, 각 움직임 물체의 움직임 추정과 세그멘테이션을 수행하는 방법에 따라 optical flow와 같은 국소적 움직임 정보에 기반을 둔 방법<sup>2)21)</sup>과 파라미터릭 움직임 모델에 기반을 두고 동영상으로부터 움직임 파라미터의 추정과 동시에 세그멘테이션을 하나의 과정으로 실행하는 직접법<sup>4)8)14)18)</sup>으로 나눌수 있다.

국소적 움직임 정보에 기반을 둔 방법에서는 국소적 움직임 정보로부터 선형회귀 과정에 의거해서 global 움직임 파라미터를 추정하고, 파라미터 공간내에서 K-means 법<sup>12)1)</sup>이나 Hough 변환<sup>2)</sup>을 이용해서 세그멘테이션을 수행하고 있다. 이들의 기법은 알고리즘이 비교적 간단하고 계산량이 적은 반면에, 국소적 정보만을 사용해서 클러스터링을 실행함으로써 움직임 물체의 경계부근이나 은폐되는 영역에서의 움직임 추정과 세그멘테이션이 정확하지 않게 된다. 또한, 움직임 파라미터 공간내에서의 클러스터링은 주어진 클러스터수에 좌우될 가능성도 크다.

파라미터릭 움직임 모델에 기반을 둔 직접법은 2D/3D 파라미터릭 움직임 모델에 의거해서 입력 동영상으로부터 움직임 파라미터의 추정과 세그멘테이션을 동시에 실행하는 기법이다. 이 기법에서는 파라미터를 추정할 때, M추정과 같은 robust 추정법<sup>4)6)18)</sup>을 사용해서 무용한 데이터의 영향을 억제하거나, 고차의 움직임 모델을 적용함으로써 물체의 움직임을 보다 정확하게 추정할수 있다. 그러나, 이들 기법은 계산량이 방대해지고, 움직임 파라미터의 추정과 세그멘테이션 결과의 유효성, 알고리즘의 수렴속도 등이 설정한 초기정보와 추정 알고리즘에 크게 좌우된다.

본 논문에서는 움직임 정보의 추정과 세그멘테이션을

위한 새로운 알고리즘을 제안한다. 제안 기법은 움직임 물체의 형상이나 움직임을 반영한 구조정보를 추출하기 위해, 초기 영역 정보를 추출, 이 초기 영역에 의거해서 파라미터릭 움직임 모델의 파라미터 추정과 동시에 영역을 수정해서 구조 정보를 추출하는 것이다. 초기 영역 정보는 국소적 움직임 정보나 각 프레임의 휘도값에 의거해서 확률적 클러스터링을 통해 추출된다. 추출된 초기 영역정보에 기반을 두고, Gauss-Newton의 최적화 기법을 사용해서 파라미터릭 움직임 모델의 파라미터 추정과 세그멘테이션을 최적화 하기 때문에 수렴속도가 빠르고 보다 정확한 구조정보의 추출이 가능하다. 또한, 초기 영역 정보로 물체의 국소적 움직임을 반영하기 때문에 계산량의 감소와 더불어 보다 정확한 움직임 정보의 추정 및 세그멘테이션이 가능하게 된다.

이하, 2절에서 혼합 모델의 ML (Maximum Likelihood) 추정법에 관해서 간단히 기술하고, ML 추정에 의거 국소적 움직임 정보와 휘도값의 확률적 클러스터링에 의한 세그멘테이션의 초기 정보 추출에 관해서 고찰한다. 3절에서는 추출된 초기 영역에 기반을 두고 파라미터릭 움직임 모델의 파라미터 추정과 세그멘테이션을 동시에 수행하는 알고리즘에 관해서 기술한다. 제안 기법의 유효성을 평가하기 위해, CIF 표준 동영상을 사용해 모의 실험한 결과에 관해 4절에서 고찰하고, 결론과 향후 과제에 관해 5절에서 간단히 기술한다.

## II. 혼합모델의 ML추정과 초기영역의 분할

통계학 분야에서 많이 연구되어 온 혼합모델에 대한 파라미터의 반복ML (Maximum Likelihood) 추정법<sup>9)15)</sup>에 관해서 간단히 설명하고, 물체의 움직임을 반영한 초기영역을 추출하기 위해서 혼합모델에 의거 동영상의 optical flow나 텍스처와 같은 국소적인 정보에 의한 확률적 클러스터링에 관해서 기술한다.

### 1. 혼합모델의 ML추정

관측된 데이터 Y로 부터 통계적 모델의 파라미터를 추정하는 문제는 불완전한 데이터를 사용하는 추정의 문제이고, 그 추정 방법으로 반복 ML추정 알고리즘, 즉 EM (Expectation and Maximization) 알고리즘이 대표적으로 알려져 있다.<sup>15)</sup>

일반적으로 관측 데이터 Y는 복수의 모델 또는 과정으로부터 관측된 것으로 하고, 이것을 불완전한 데이터 (incomplete data)로 본다. 한편 관측전의 원래의 모델 데이터 X를 완전한 데이터 (complete data)로 하고, 관측된 데이터와의 사이에

$$Y = H(X) \tag{1}$$

의 관계를 갖는 것으로 한다. 여기서  $H(\cdot)$  는 불가역 변환작용소 즉 1대1 대응이 아닌 변환을 나타낸다. 이때,  $Y$ 의 모델 파라미터  $\phi$ 의 ML추정은 데이터  $Y$ 에 의거한 확률밀도함수 (probability density function: pdf)  $f(Y|\phi)$ 를 다음과 같이 최대가 되도록 추정하는 알고리즘이다.

$$\hat{\phi} = \arg \max_{\phi} \log f(Y|\phi) \quad (2)$$

관측 데이터로 부터 파라미터를 단일 과정에 의해서 추정하는 것은 어렵기 때문에, 반복 알고리즘이 주로 사용된다. 이때, EM 알고리즘에 의한 파라미터의 ML추정에서는 미리 주어진 초기치  $\hat{\phi}^{(0)}$ 에 대해서 다음의 두 과정을 반복 적용해서 파라미터  $\phi$ 를 추정한다.

E-Step :

$$Q(\phi | \hat{\phi}^{(k)}) = E[\log f(X|\phi) | Y, \hat{\phi}^{(k)}] \quad (3)$$

M-Step :

$$\hat{\phi}^{(k+1)} = \arg \max_{\phi} Q(\phi | \hat{\phi}^{(k)}) \quad (4)$$

여기서  $k$ 는 반복회수를 나타내는 계수이다.

파라미터 추정을 위한 EM 알고리즘은 함수  $Q(\cdot)$ 가 연속인 경우 발산하지 않고 지수함수적으로 수렴하는 것이 알려져 있다.<sup>[9][15]</sup> 그러나, 주어진 초기치  $\hat{\phi}^{(0)}$ 에 따라서 국소적인 값에 수렴하거나, 수렴속도가 늦어지기도 한다. 이상의 파라미터 추정을 위한 EM 알고리즘은 다양한 분야에 걸쳐서 응용되고 있다. 특히, 영상처리 분야에서의 응용으로써 optical flow의 추정<sup>[14]</sup> MRF (Markov Random Field) 모델에 의거한 정지화상의 영역분할<sup>[23]</sup>, 동영상의 움직임 파라미터 추정과 세그멘테이션<sup>[4][8][22]</sup> 등이 있다.

## 2. 확률적 클러스터링에 의한 세그멘테이션

Optical flow와 같은 국소적인 움직임 정보는 공간적 smoothness와 같은 제약조건하에서 움직임 파라미터를 추정하기 때문에 움직임 물체의 경계부근이나 은폐되는 영역에서의 움직임 추정에 오차가 크고, 물체의 움직임을 정확히 반영한 움직임 정보의 추정은 어렵게 된다. 또한, 움직임을 추정할 때 제한된 범위내의 영상만을 사용하기 때문에 참조범위를 넘는 움직임이나 global한 움직임의 경우, 정확한 움직임 파라미터의 추정은 용이하지 않다. 따라서, 국소적인 정보에 기반을 두고 추정된 움직임 정보로부터 파라미터적 움직임 모델을 근사시켜 동영상의 구조 정보를 추출하는 기법에는 많은 제한이 따른다.

또다른 방법으로 파라미터적 움직임 모델에 의거해서 움직임 파라미터의 추정과 세그멘테이션을 동시에 수행하는 것이다. 그러나, 일반적으로 움직임 파라미터의 추정과 세그멘테이션은 밀접한 관계를 갖고 있다. 즉, 움직임 파라미터의 추정을 정확하게 하기 위해서는 움직임 물체에 관한 세그멘테이션 정보가 필요하고, 움직임 물체를 세그멘테이션 하기 위해서는 보다 정확한 움직임 정보가 요구된다. 이 때문에 종래의 많은 기법에서는 움직임 파라미터의 추정과 세그멘테이션의 과정을 서로 반복하거나, 불력으로 주어진 처리대상 영역이 서로 다른 움직임 모델을 갖는 복수 모델로 구성되어 있다는 가정하에서, 움직임 파라미터의 추정과 세그멘테이션을 단일 과정으로 처리하고자 하는 기법이 사용되어 왔다. 이들 종래의 기법에서는, 반복과정의 초기치를 위한 영역으로서 임의 크기의 불력이 주로 사용되고 있다. 따라서, 움직임 파라미터의 추정과 세그멘테이션을 위한 알고리즘의 수렴속도가 늦고, 방대한 계산량을 필요로 하는 경우가 많다.

본 논문에서는 이상의 문제점을 개선하고 물체의 움직임과 형상을 반영한 움직임 파라미터의 추정과 세그멘테이션을 수행하기 위해서, 관측 데이터  $Y$ 로 입력 동영상의 optical flow 정보와 휘도값을 사용하고, 확률적 클러스터링을 수행해서 초기 영역정보를 추출한다. K-means법과 같은 기법에서는, 단순히 클러스터링과 각 클러스터의 모델 파라미터 추정의 과정을 반복해서 최종적인 클러스터링을 얻기 때문에 관측 데이터의 클러스터링에 데이터의 사후확률이 직접 반영되지 않는다. 따라서, 평가함수와 초기 클러스터수에 의해서 부적당한 결과가 얻어질 가능성이 높게되고, 분리 및 통합 (split and merge)을 수행할 때, 최적의 문턱값 (threshold value)를 자동적으로 결정하기 어렵다. 이에 반해, 확률적 클러스터링 기법에서는 관측 데이터의 사후확률에 의거해서 클러스터링을 수행하기 때문에, 분리와 통합의 과정이 필요치 않고, 자동적으로 클러스터링을 수행한다. 따라서 관측 데이터의 통계적 성질을 보다 정확히 반영한 클러스터링의 결과를 얻을 수 있다.

입력 동영상으로 부터의 관측 데이터  $Y$ 를, 수식의 간략화를 위해 1차원의 데이터로 하고, 1 프레임의 화소수를  $N$ , 모델수 즉 클러스터수를  $K$ 로 한다. 이때, 클러스터링은 다음과 같이 관측 데이터  $Y$

$$Y = \{ Y_1, Y_2, \dots, Y_N \} \quad (5)$$

에 대한 클러스터  $R_i$ 의 사후확률  $f(R_i; Y)$ 을 최대가 되도록 분할하는 것이다.

$$\max_i f(R_i; Y) \quad (i=1, 2, \dots, K) \quad (6)$$

여기서,  $f(R_i; Y) = \sum_{j=1}^N \Pr.(Y_j \in R_i; Y)$ 를 의미한다.

관측 데이터 Y를 K 클러스터로 부터의 데이터로 하면, 임의의 계수  $\gamma_i$  를 이용해서 관측 데이터 Y의 pdf를

$$f(Y; \phi) = \sum_{i=1}^K \gamma_i f(Y; \theta_i) \quad (7)$$

$$\text{단, } \sum_{i=1}^K \gamma_i = 1, \gamma_i \geq 0 \quad (8)$$

와 같이 각 클러스터에 대한 pdf의 가중치합 (weighted sum)으로 나타낸다. 여기서  $f(Y; \theta_i)$ 는 파라미터  $\theta_i$  에 의한 데이터 Y의 pdf이고,  $\theta$  는 각 클러스터의 모델에 대한 미지의 통계 파라미터 벡터이다. 또한  $\phi$  는  $\phi = [\gamma^T, \theta^T]^T$  를 의미한다.

$Y_j$  를 각 클러스터에 할당되는 데이터로서 본 논문에서는 다음과 같은 2진 변수  $z_j = [z_{1j}, \dots, z_{Kj}]^T$  를 이용한다.

$$z_{ij} = \begin{cases} 1, & \text{if } Y_j \in R_i \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

이 식으로부터 알 수 있는 바와 같이  $z_j$  는 데이터  $Y_j$  가 클러스터  $R_i$  에 속한 경우,  $z_{ij}$  이외에는 전부 0, 즉  $z_j = e_i$  가 되는 2진 변수이다. 이 경우 데이터 X는 (Y, z)로 정의된다.

데이터  $Y_j$  ( $j=1, \dots, N$ )는 pdf  $f(Y; \theta)$ 가 독립이고 동일분포(independently and identically distributed: i.i.d)의 함수인 것으로 하고, 변수 z가 조건부로 주어진 경우에도 이러한 성질이 유지된다고 가정한다. 이 경우, 주어진 파라미터  $\phi$  에 대한 X의 log-likelihood pdf는

$$\begin{aligned} \log f(X; \phi) &= \sum_{j=1}^N \log f(Y_j, z_j; \phi) \\ &= \sum_{j=1}^N \log f(Y_j; z_j, \phi) + \sum_{j=1}^N \log f(z_j; \phi) \\ &= \sum_{j=1}^N z_j \log f(Y_j; \phi) + \sum_{j=1}^N z_j \log f(z_j; \phi) \end{aligned} \quad (10)$$

가 된다. 또한, 식 7을 식 10에 대입하면,  $\log f(X; \phi)$ 는

$$\begin{aligned} \log f(X; \phi) &= \sum_{i=1}^K \sum_{j=1}^N z_{ij} [\log \gamma_i + \log f(Y_j; \theta_i)] \\ &+ \sum_{i=1}^K \sum_{j=1}^N z_{ij} \log f(e_j; \phi) \end{aligned} \quad (11)$$

가 됨을 알수 있다. 따라서 파라미터의 추정치  $\hat{\phi}$  에 대한 EM 알고리즘의 E-step  $Q(\phi | \hat{\phi})$ 는

$$\begin{aligned} Q(\phi | \hat{\phi}) &= E[\log f(X; \phi) | Y, \hat{\phi}] \\ &= \sum_{i=1}^K \sum_{j=1}^N E[z_{ij} | Y, \hat{\phi}] [\log \gamma_i + \log f(Y_j; \theta_i)] \\ &+ \sum_{i=1}^K \sum_{j=1}^N E[z_{ij} | Y, \hat{\phi}] \log f(e_j; \phi) \end{aligned} \quad (12)$$

가 된다.

식 12에서,  $E[z_{ij} | Y, \hat{\phi}]$ 는 식 6의 데이터  $Y_j$ 가 클러스터  $R_i$ 에 속할 사후확률에 해당된다. 이것을  $\tau_i(Y_j; \hat{\phi})$ 로 하면,

$$\begin{aligned} \tau_i(Y_j; \hat{\phi}) &= E[z_{ij} | Y, \hat{\phi}] \\ &= \Pr(Y_j \in R_i | Y_j; \hat{\phi}) \\ &= \frac{\gamma_i f(Y_j; \hat{\theta}_i)}{\sum_{k=1}^K \gamma_k f(Y_j; \hat{\theta}_k)} \end{aligned} \quad (13)$$

와 같이 된다.  $Y_1, \dots, Y_N$ 의 확률적 클러스터링을 수행하기 위해서는 각  $Y_j$ 를 사후확률이 가장 높은 클러스터에 할당하면 된다. 즉,

$$\tau_k(Y_j; \hat{\phi}) > \tau_i(Y_j; \hat{\phi}) \quad (i=1, 2, \dots, K; i \neq k) \quad (14)$$

인 경우,  $Y_j$ 는 클러스터  $R_k$ 에 속하는 것으로 한다.

식 12의  $\log f(e_j; \phi)$ 는 파라미터  $\phi$ 의 추정과 직접적인 관계가 없기 때문에 식 8의 조건을 식 12에 대입해서 식 4의 M-step을 수행하면 파라미터  $\phi$ 의 ML추정은  $\hat{\phi} = [\hat{\gamma}^T, \hat{\theta}^T]^T$

$$\hat{\gamma}_i = \frac{1}{N} \sum_{j=1}^N \hat{\tau}_{ij} \quad (i=1, 2, \dots, K) \quad (15)$$

$$\sum_{i=1}^K \sum_{j=1}^N \hat{\tau}_{ij} \frac{\partial \log f(Y_j; \hat{\theta}_i)}{\partial \theta_i} = 0 \quad (16)$$

와 같이 유도할수 있다.<sup>[15]</sup>

여기서  $\tau_{ij}$ 는  $\tau_{ij} = \tau_i(Y_j; \hat{\phi})$ 이다. 또한,  $Y_j$ 의 pdf를 Gaussian분포로 하면, 파라미터 벡터  $\theta$ 는 평균치 벡터  $\overline{Y}_1, \dots, \overline{Y}_K$ 와 분산치 벡터  $\Sigma_1, \dots, \Sigma_K$ 가 된다. Gaussian 정규분포의 조건에서,  $\gamma_i, \overline{Y}_i$ 와  $\Sigma_i$ 의 ML추정은 다음과 같이 유도된다.

$$\hat{\gamma}_i = \frac{1}{N} \sum_{j=1}^N \hat{\tau}_{ij} \quad (17)$$

$$\hat{\overline{Y}}_i = \frac{1}{N \hat{\gamma}_i} \sum_{j=1}^N \hat{\tau}_{ij} Y_j \quad (18)$$

$$\widehat{\Sigma}_i = \frac{1}{N\widehat{\gamma}_i} \sum_{j=1}^N \widehat{\tau}_{ij} [Y_j - \overline{Y}_i] [Y_j - \overline{Y}_i]^T \quad (19)$$

또한,  $Y_j$ 가 클러스터  $R_i$ 에 속하는 사후확률  $\tau_{ij}$ 는 식 13으로 부터

$$\widehat{\tau}_{ij} = \frac{\gamma_i |\Sigma_i|^{-\frac{1}{2}} \exp(\Omega_{ij})}{\sum_{k=1}^K \gamma_k |\Sigma_k|^{-\frac{1}{2}} \exp(\Omega_{jk})} \quad (20)$$

$$\Omega_{ij} = -\frac{1}{2} [Y_j - \overline{Y}_i]^T \Sigma_i^{-1} [Y_j - \overline{Y}_i]$$

이 된다.

확률적 클러스터링을 수행하기 위해, 본 논문에서 사용한 관측 데이터  $Y$ 는 정규화한 optical flow의 크기성분 (u,v)과 방향성분  $\alpha = \tan^{-1}(\frac{v}{u})$ , 각 프레임의 휘도값 (I)의 4차원으로 하고, 이상의 과정을 반복 수행함으로써 초기 영역을 추출한다.

### III. 움직임 파라미터의 추정과 움직임 물체의 추출

2절에서의 알고리즘에 의해 추출된 동영역내의 초기영역에 대해서 움직임 모델을 정해, 그 파라미터를 각 영역마다 직접 추정해서 예측오차가 가장 적게되도록 각 영역을 수정한다. 계산량의 관점에서 다음과 같은 Gauss-Newton 반복알고리즘<sup>[3]</sup>을 이용해서 각 영역의 움직임 파라미터의 추정과 세그멘테이션을 수행한다. 그림 1은 이를 위한 알고리즘의 블럭도를 나타낸 것이다.

연속 프레임간의 휘도치의 변화가 크지 않다고 가정하면, 시간  $t$ 와  $t-1$ 사이의 움직임에 의한 화소 (x,y)에서의 오차  $e$ 는

$$e = I(x,y,t) - I(x-u,y-v,t-1) \quad (21)$$

와 같이 주어진다. 여기서,  $I(x,y,t)$ 는 시간  $t$ , 화소 (x,y)에서의 휘도값이고, (u,v)는 다음과 같은 affine 파라미터의 움직임 벡터이다.

$$\begin{aligned} u &= a_1 + a_2x + a_3y \\ v &= a_4 + a_5x + a_6y \end{aligned} \quad (22)$$

여기서  $(a_1, \dots, a_6)$ 는 움직임 파라미터이다. 식 21을 Taylor 급수로 전개하고 고차항을 제거해서 정리하면

$$e = I_x u + I_y v + I_t \quad (23)$$

가 된다. 단,  $I_x, I_y, I_t$ 는 각변수에 대해서 편미분한 것이다. 임의의 영역  $R$ 에서 오차의 2승  $E^2$ 는

$$E^2 = \sum_R e^2 = \sum_R (I_x u + I_y v + I_t)^2 \quad (24)$$

로 주어진다. 이 식에 식 22와 식 23을 대입해서 정리하면,

$$E^2 = \sum_R \nabla I X P + \Delta I]^2 \quad (25)$$

이 된다. 식 25에서는 표기의 간단화를 위해 다음의 행렬을 사용했다.

$$\begin{aligned} \nabla I &= [I_x, I_y], \quad \Delta I = I_t, \\ X &= \begin{bmatrix} 1 & x & y & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & x & y \end{bmatrix}, \\ P &= [a_1, a_2, a_3, a_4, a_5, a_6]^T \end{aligned}$$

식 25를 파라미터  $P$ 에 대해 미분하고, 그 값이 영이 되도록 하면, 파라미터  $P^{(n+1)}$ 은

$$P^{(n+1)} = P^{(n)} + M^{-1} B^{(n)} = P^{(n)} + \Delta P^{(n)} \quad (26)$$

$$(\text{단 } M = \sum_R X^T (\nabla I)^T (\nabla I) X, \quad B^{(n)} = - \sum_R X^T (\nabla I)^T \Delta I^{(n)})$$

와 같이 추정된다. 여기서,  $P^{(n)}$ 은  $n$ 회 반복에 의해서 추정된 파라미터이고,  $\Delta P^{(n)}$ 은  $n$ 번째의 반복에서 경신되는 값이다. 이상의 알고리즘을 이용해서 분할된 각영역마다 움직임 파라미터를 추정한다. 또한 bilinear interpolation을 이용해서 각영역의 움직임 파라미터에 의한 움직임 보상 예측화상을 구해서 예측오차를 계산한다.

국소적인 움직임이나 텍스쳐정보의 클러스터링만에 의한 세그멘테이션에는 움직임 물체의 형상이나 global motion을 반영시키기 어렵다. 따라서 이들의 영역을 추정된 움직임 파라미터를 사용해서 수정하는 것이 필요하다. 화소 (x,y)에서의 영역  $R_k$ 의 움직임 파라미터  $\widehat{P}_k$ 에 의거해서 구한 움직임 보상을 화소를  $\widehat{T}_k(x,y,t)$ 로 하고, 원영상과의 오차를  $e_k(x,y)$ 로 한다.

$$\begin{aligned} \widehat{T}_k(x,y,t) &= I(x-u_k, y-v_k, t-1) \\ e_k(x,y) &= I(x,y,t) - \widehat{T}_k(x,y,t) \end{aligned} \quad (26)$$

오차  $e_k(x,y)$ 를 다음과 같이 적당한 크기의 창 (window)내에서 계산하고 이를  $d_k(x,y)$ 로 한다.

$$d_k(x,y) = \sum_{i,j} w(i,j) e_k(x+i, y+j) \quad (27)$$

여기서  $W(i, j)$ 는 가중치 파라미터이고,  $3 \times 3$ 의 창합수를 사용했다. 각영역에서 계산한 예측오차  $d_k(x, y)$ 는 평균이 0, 분산이  $\sigma_k$ 인 Gaussian분포로 모델링하고 다음의 평가함수  $\beta_k(x, y)$ 를 도입해서 사용한다.

$$\beta_k(x, y) = \frac{1}{\sigma_k} \exp\left(-\frac{d_k^2(x, y)}{2\sigma_k^2}\right) \quad (28)$$

화소  $(x, y)$ 에서

$$\beta_i(x, y) < \beta_k(x, y) \quad (k=1, \dots, K, k \neq i) \quad (29)$$

의 조건을 만족하는 경우, 화소  $(x, y)$ 는 영역  $R_k$ 에 속한 것으로 한다. 이와 함께  $3 \times 3$ 의 메디안 필터를 사용해서 수정된 각영역의 경계선을 부드럽게 함과 동시에 작은 고립 영역을 제거한다.

#### IV. 실험 결과

본 논문에서 제안한 알고리즘은, 2절에서 기술한 확률적 클러스터링에 의한 초기 세그멘테이션을 수행하는 과정과, 3절에서 기술한 파라미터릭 움직임 모델에 의거한 움직임 파라미터의 추정과 세그멘테이션을 동시에 수행하는 과정으로 구성된다. 또한, 각 과정을 처리할 때, 전 프

레이의 세그멘테이션 결과를 반영해서 이용한다. 그림 2는 제안 알고리즘의 블럭도를 나타낸 것이다. 알고리즘의 유효성을 평가하기 위해, 실제의 동영상을 사용해서 컴퓨터에 의한 모의 실험을 수행했다. 실험에서 사용한 동영상은 CIF표준 영상의 "flower garden (352 x 240)"과 "foreman (352 x 288)"이다.

초기 세그멘테이션의 과정에서는 optical flow를 추정해서 각 프레임의 휘도값과 함께 확률적 클러스터링을 수행하는데 이용했다. optical flow의 추정은 움직임이 큰 동영상에도 대응할수 있도록 하기 위해 계층적 구배법<sup>15)</sup>을 사용했다. 이때, 계층영상은 Gaussian 필터를 사용해서 구성했다. 또한, 시공간 3차원으로 필터링을 해서 미분할때의 잡음의 영향을 최소화 하도록 했다.

동영상 "foreman"은, 초기 수십 프레임에서 global motion을 갖는 배경과 local motion을 갖는 움직임 물체(인물)로 이루어진 영상이다. 따라서, 정지영역은 없으며 하나의 프레임전체가 움직임 영역으로 검출된다. 그림 3(a)와 3(b)는 연속 원영상 (frame no. 31과 32)의 하나를 나타낸 것이다. 그리고 클러스터수  $K$ 를  $K=2$ 로 해서 확률적 클러스터링을 수행해서 얻은 세그멘테이션의 초기 정보의 추출 결과를 그림 3(c)에 보이고 있다. 이 정보에 근거해서 움직임 파라미터의 추정과 동시에 세그멘테이션을 행한 결과를 그림 3(d)에 나타낸다. 여기서 헬멧 내부는, 초기 영역의 추출에서는 배경으로 분류되지만 인물의 얼굴 영역과 동일한 움직임을 갖는다. 따라서 파라미터릭 움직임 모델에 의거, 움직임 파라미터의 추정과 세그멘테이션을 반복함에 따라 인물의 영역에 포함되어 움직임 물체

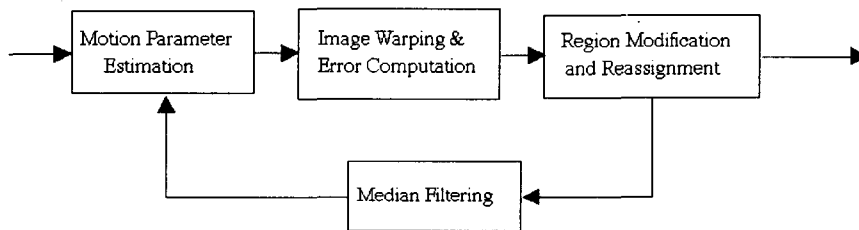


그림 1. 움직임 파라미터의 추정과 세그멘테이션을 위한 알고리즘의 블럭도  
 Fig. 1. The block diagram for the motion parameter estimation and segmentation

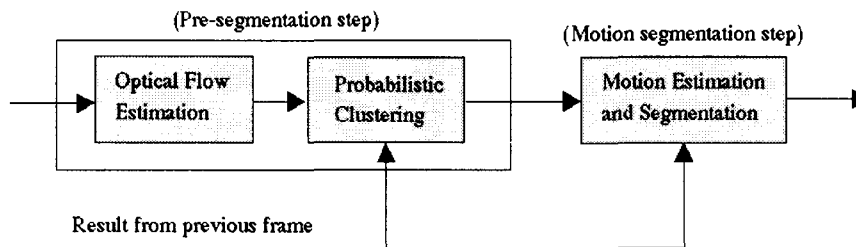


그림 2. 동영상의 구조적 움직임 세그멘테이션을 위한 제안 알고리즘의 블럭도  
 Fig. 2. The block diagram for the proposed structural motion segmentation

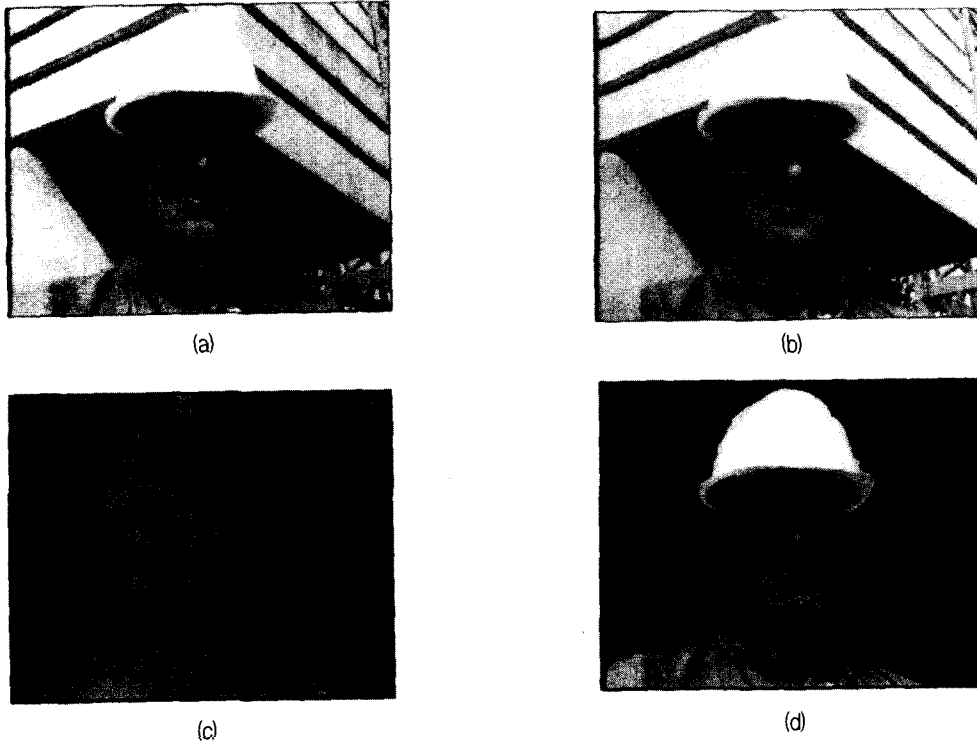


그림 3. 움직임 세그멘테이션의 결과 ("foreman"): 원영상 ((a) 프레임 31, (b) 프레임 32), (c) 초기 영역의 분할, (d) 움직임 세그멘테이션.  
 Fig. 3. The motion segmentation result for the image sequence "foreman": (a)The original image at frame no. 31 and (b) frame no. 32, (c) Presegmentation, (d) Motion segmentation

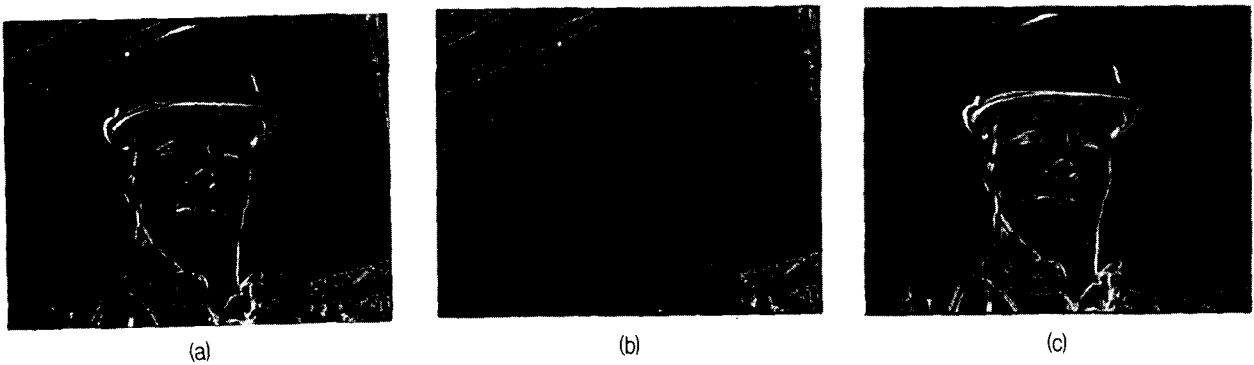


그림 4. 각 물체의 움직임에 대한 움직임 보상 예측 오차: (a) 프레임간 차, (b) 세그먼트 "인물", (c) 세그먼트 "배경"  
 Fig. 4. The warped residual motion compensated error with respect to the estimated motion parameters for the image sequence "foreman" for each motion segmentation: (a) Interframe difference, (b) Man, (c) Overall background

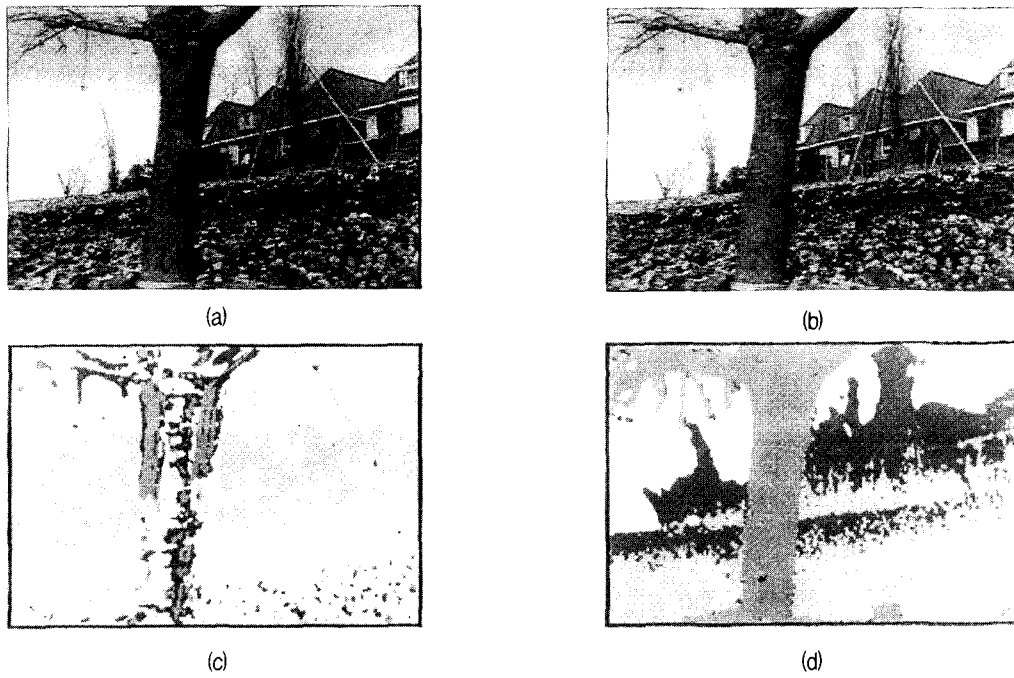


그림 5. 움직임 세그멘테이션의 결과 ("flower garden"): 원영상 ((a)프레임 16 (b)프레임 17), (c) 초기영역의 분할, (d) 움직임 세그멘테이션.  
 Fig. 5. The motion segmentation result for the image sequence "flower garden": (a) The original image at frame no. 16, (b) The original image at frame no. 17, (c) Presegmentation, (d) motion segmentation

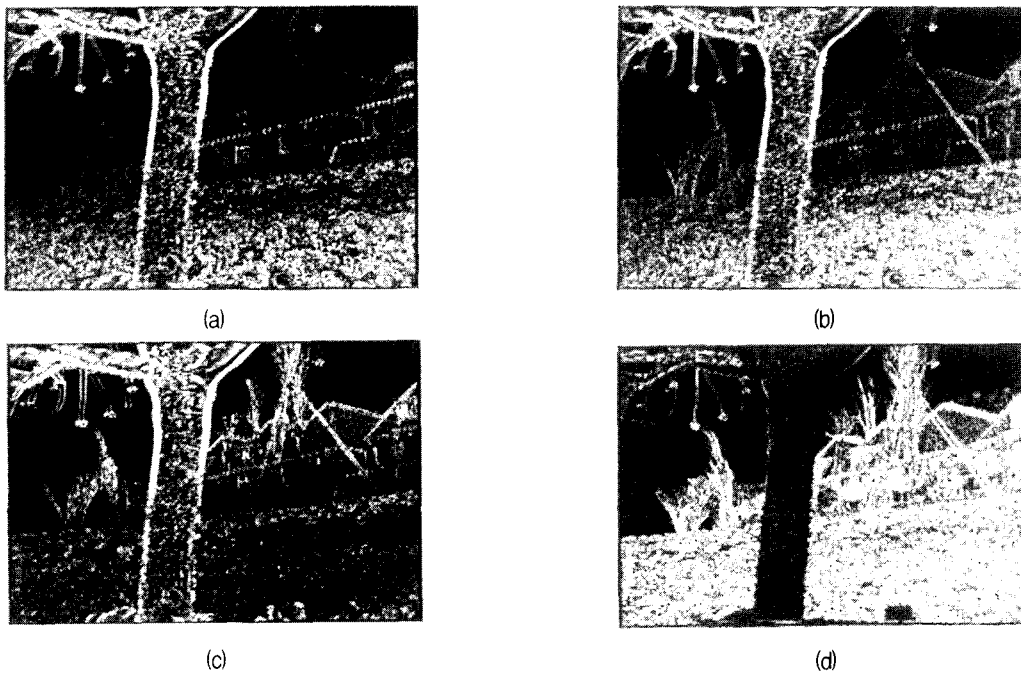


그림 6. 각 물체의 움직임에 대한 움직임 보상 예측 오차: (a) 세그먼트 "하늘", (b) 세그먼트 "집", (c) 세그먼트 "화단", (d) 세그먼트 "나무".  
 Fig. 6. The warped residual motion compensation error to the estimated motion parameters for the image sequence "flower garden" for each motion segmentation: (a) Heaven area, (b) House, (c) Flower garden, (d) Frontal tree.



로서 인물과 배경이 정확하게 분리됨을 확인할 수 있다. 또한, 각 움직임 물체에 대한 파라미터의 추정 정도를 확인하기 위해, 각 움직임 물체의 움직임 파라미터에 의한 움직임 보상 예측 오차를 그림 4에 나타낸다. 그림 4(a)는 움직임 보상을 하지 않은 프레임간 차의 절대값이고, 그림 4(b)와 4(c)는 각 움직임 물체의 움직임 파라미터에 의한 예측오차를 나타낸 것이다. 이때, 추정된 움직임 파라미터에 의한 움직임 보상 전,후의 예측오차로부터 계산한 PSNR은 각각 75.96dB와 85.5dB이다. 이 결과로부터 각각의 움직임 물체는 그들의 움직임이 반영되어 예측오차가 크게 줄어든 것을 알 수 있다.

동일한 실험을 동영상 "flower garden"에 적용한 결과를 그림 5에 나타낸다. 이 경우 클러스터수  $K$ 를  $K=4$ 로 해서 클러스터링하고 초기 세그멘테이션을 추출했다. 또한, 예측오차에 의한 각 움직임 물체의 움직임 파라미터의 추정 정도를 그림 6에 나타냈다. 여기서, 그림 6(a) ~ 6(d)는 각각 하늘, 집, 화단, 큰 나무의 부분에 해당되는 영역의 움직임 보상 예측의 결과에 대응되는 것이다. 이때, 하늘의 영역에 대한 부분은 움직임이 제로에 가깝고 따라서 예측오차는 프레임간 차의 절대값과 거의 동일하다. 그림 4와 마찬가지로, 움직임 파라미터에 의한 움직임 보상 전, 후의 예측오차로부터 계산한 PSNR은 각각 65.52dB와 78.74dB가 얻어졌다.

본 논문에서는 계산량을 줄이기 위해 클러스터수  $K$ 값을 사전에 가정하고 실험을 수행했으나, 제안 알고리즘은 미지의  $K$ 값에 대해서도 정상적으로 동작한다. 그러나 그 계산량은 확률적 클러스터링을 수행할때의 클러스터수  $K$ 에 크게 좌우된다. 특히, 사후확률  $\tau_{ij}$  (식 20)의 계산은 각 클러스터마다 프레임 전체에 걸쳐 계산해야 되고 그 계산량은 클러스터수에 비례해서 증가한다. 따라서, 클러스터수가 많은 경우 전체 계산량은 많아지게 되고 실시간 처리와는 거리가 있으나, 많은 계산량이 요구되는 EM 알고리즘은 초기영역을 검출하는데에만 사용되므로, EM 알고리즘의 반복회수를 제한함으로써 계산량의 제어가 가능하다. 또한, 추출된 초기영역을 중심으로 파라미터 움직임 모델의 움직임 파라미터의 추정과 세그멘테이션을 수행하기 때문에, 화소단위를 초기정보로 활용하는 종래의 파라미터 움직임 모델만에 의한 직접법보다는 수렴속도가 빠르게 되고 그 계산량을 줄일 수 있다.

일반적인 동영상에서 물체의 움직임과 영역형상을 정확히 반영한 움직임 파라미터의 추정과 세그멘테이션은 성질이 나쁜 문제 (ill-posed problem)의 하나로 그 정확한 해를 구하는 것이 어렵다. 따라서, 클러스터수  $K$ 값을 자동으로 결정하고, 계산량을 대폭적으로 줄임과 동시에 물체의 움직임과 영역형상을 반영한 파라미터의 추정과 세그멘테이션을 수행하는 것은 어려운 과제이며 향후 많은 연구가 필요하다. 본 제안 알고리즘의 실효성은 임의의  $K$ 값에 대해, 동영상의 움직임과 영역형상을 반영한 정확한 파라미터의 추정과 세그멘테이션이 가능하다는 데 있다.

또한, 본 논문의 가장 큰 목적은, 보다 적은 움직임 모델에 의해 정확한 움직임 파라미터를 추정하고 동영상을 보다 콤팩트한 표현으로 기술함으로써, 동영상의 검색이나 편집, 제작, 실시간성이 요구되지 않는 영상통신등을 위한 새로운 처리 기법의 개발에 활용하는 것이다. 이러한 관점에서 실험 결과를 고찰하면, 움직임 물체의 움직임과 영역형상에 관한 정보가 충분히 반영되어 있음을 알 수 있다. 즉, 그림 3의 "foreman"에서는 인물과 배경이 분리되어 2개의 파라미터 움직임 모델에 의해 동영상을 기술할 수 있다. 국소적인 움직임 정보만에 기초한 기법에서는 헬멧의 부분이나 몸의 일부만이 배경의 영역에 포함되거나, 배경의 일부만이 인물의 영역에 포함되어 그림 3(d)와 같은 결과를 얻을 수 없는 가능성이 높다. 또한, 그림 5의 "flower garden"에서는 global motion으로부터 움직임의 크기에 따라 4개의 영역이 잘 분리되어 있음을 알 수 있다. 화단과 집의 영역이 뚜렷하게 분리되어 있지 않으나, 종래의 기법<sup>[18][21]</sup>에 비해, 분리 추출된 각 영역에는 움직임만이 아니라 형상에 관한 정보가 잘 반영되어 있음을 알 수 있다.

## V. 결 론

동영상으로부터 콤팩트한 구조정보를 추출하기 위한 방법으로 혼합 모델의 확률적 클러스터링 기법을 이용해서 국소적 움직임 정보와 휘도값으로 초기영역을 분할하고, 각 초기 영역마다 파라미터 움직임 모델의 움직임 파라미터의 추정과 세그멘테이션을 수행하는 알고리즘에 관해서 기술했다. 모의 실험을 통해, 움직임 물체의 형상이나 움직임을 반영한 세그멘테이션과 콤팩트한 움직임 정보가 추출되는 것을 확인하고 그 유효성을 입증했다.

그러나, 해결해야 할 과제 또한 많이 남아 있다. 계산량을 줄이기 위해 각 동영상마다 클러스터수를 미리 제한해서 사용했다. 따라서 적절한 클러스터수의 자동 결정을 포함한 알고리즘의 일반화와 계산량을 줄이기 위한 알고리즘의 개선이 요구되고 있다. 향후, MPEG4에서 논의된 객체 기반의 generic한 동영상압축이나, 영상의 검색, 편집 등의 응용을 고려한 보다 많은 연구의 진행이 필요하다.

## 참 고 문 헌

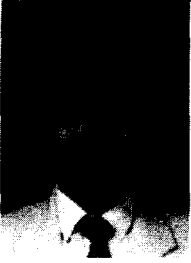
- [1] E. H. Adelson, "Layered representation for vision and video," *IEEE Workshop on Representation of Visual Scenes, MIT*, Jun. 1995.
- [2] G. Adiv, "Determining three dimensional motion and structure from optical flow generated by several moving objects," *IEEE Trans. Pattern Anal. and Machine Intell.*, vol. 17, no. 4, pp.

- 384-401, Jul. 1985.
- [3] P. Anandan, J. R. Bergen, K. J. Hanna, and R. Hingorani, "Hierarchical model-based motion estimation," *Motion Analysis and Image Sequence Processing*, edited by M. I. Sezan and R. L. Lagendijk, KAP, pp. 1-22, 1993.
- [4] S. Ayer and H. S. Sawhney, "Layered representation of motion video using robust maximum likelihood estimation of mixture models and MDL encoding," *Int. Conf. on Computer Vision'95*, pp. 777-784, Jun. 1995.
- [5] J. L. Barron, D. J. Fleet, and S. S. Beauchemin, "Performance of optical flow techniques," *Proc. IEEE Computer Vision and Pattern Recon.*, pp. 236-242, 1992.
- [6] M. J. Black and A. Jepson, "Estimating multiple independent motions in segmented images using parametric models with local deformations," *IEEE Workshop on Motion of Nonrigid and Artificial Objects*, pp. 220-227, Nov. 1994.
- [7] C. K. Cheong and K. Aizawa, "Structural motion segmentation based on probabilistic clustering," *Proc. IEEE Int. Conf. Image Processing*, vol. I.505-I.508, Sep. 1996.
- [8] C. K. Cheong, K. Aizawa, T. Saito, M. Kaneko, and H. Harashima, "Structural motion segmentation for compact image sequence representation," *Proc. SPIE's Symp. on Visual Commun. and Image Proc.'96*, vol. 2727, pp. 1152-1163, Mar. 1996.
- [9] A. P. Dempster, N. M. Laird, and D. J. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Royal Soc. Statist. Series B*, no. 1, pp. 1-38, 1977.
- [10] N. Diehl, "Object-oriented motion estimation and segmentation in image sequences," *Signal Processing: Image Commun.*, vol. 3 no. 1 pp. 23-56, 1991.
- [11] L. Haibo and R. Forchheimer, "Image sequence coding at very low bitrates: A review," *IEEE Trans. Image Processing*, vol. 3, no. 3, pp. 589-609, Sep. 1994.
- [12] M. Irani, S. Hsu, and P. Anandan, "Mosaic based video compression," *Proc. of SPIE Visual Commun. and Image Processing*, vol. 2419, pp. 242-252, Feb. 1995.
- [13] R. Jain, W. N. Martin, and J. K. Aggarwal, "Segmentation through the detection of changes due to motion," *Computer Graphics and Image processing*, vol. 11, pp. 13-34, 1979.
- [14] A. Jepson and M. J. Black, "Mixture models for optical flow computation," *Technical report RBCVTR-93-44, Univ. of Toronto*, Apr. 1993.
- [15] G. J. McLachlan and K. E. Basford, "Mixture models: inference and applications to clustering," *Mercel Dekker Inc.* 1988.
- [16] T. Saito and K. Aizawa, "Development of video compression techniques," *J. Institute Television Eng. of Japan*, vol. 49, no. 4, pp. 523-534, Apr. 1995.
- [17] P. Salembier etc, "Segmentation-based video coding system allowing the manipulation of objects," *IEEE Trans. Circuit and Systems for video Tech.*, vol. 7, no. 1, pp. 60-74, Feb. 1997.
- [18] H. S. Sawhney, S. Ayer and M. Gorkani, "Model-based 2D & 3D dominant motion estimation for mosaicing and video representation," *Int. Conf. on Computer Vision'95*, pp. 583-590, Jun. 1995.
- [19] T. Sikora, "The MPEG-4 video standard verification model," *IEEE Trans. Circuit and Systems for video Tech.*, vol. 7, no. 1, pp. 19-31, Feb. 1997.
- [20] W. B. Thompson, "Combining motion and contrast for segmentation," *IEEE Trans. Pattern Anal. and Machine Intell.*, vol. 2, no. 6, pp. 543-549, Nov. 1980.
- [21] J. Y. A. Wang and E. H. Adelson, "Representing moving images with layers," *IEEE Trans. Image Processing*, vol. 13, no. 5, pp. 625-638, Sep. 1994.
- [22] Y. Weiss and E. H. Adelson, "Perceptually organized EM: a framework for motion segmentation that combines information about form and motion," *MIT Technical Report*, vol. 315, 1995.
- [23] J. Zhang, J. W. Modestino, and D. A. Langan, "Maximum likelihood parameter estimation for unsupervised stochastic model-based image segmentation," *IEEE Trans. Image Processing*, vol. 3, no. 4, pp. 404-420, Jul. 1994.

---

 저 자 소 개
 

---



## 정 차 근

1978. 3. ~ 1982. 2. 경북대학교 전자공학과

1982. 3. ~ 1984. 2. 서울대학교 대학원 전기공학과 석사

1990. 4. ~ 1994. 3. 일본 동경대학 전기공학과 공학박사

1984. 1. ~ 1990. 2. LG전자 영상미디어연구소 선임연구원

1994. 2. ~ 1997. 8. LG종합기술원 책임연구원

1997. 9. ~ 현재 호서대학교 제어계측공학과 교수

주관심분야: 디지털 신호처리, 영상처리 및 부호화, 디지털 영상 통신, Image Sensor