

비선형 다중스케일 필터링을 사용한 비디오 객체 분할에 관한 연구

준회원 이 응 희*, 김 태 희**, 이 규 동*, 정회원 정 동 석*

A Study on Video Object Segmentation using Nonlinear Multiscale Filtering

Woong-Hee Lee*, Tae-Hee Kim**, Gyu-Dong Lee* Associate Members, Dong-Seok Jeong* Regular Member

요 약

MPEG-4와 같은 객체 기반 부호화는 멀티미디어 응용을 위한 다양한 내용 기반 기능들을 제공한다. 압축 효율의 향상과 더불어 이러한 기능들이 지원되도록 하기 위해서는 비디오 데이터의 각 프레임은 비디오 객체로 분할되어야 한다. 본 논문에서는 비선형 다중스케일 필터링과 시공간 정보를 사용한 효과적인 비디오 객체 분할 기법을 제안한다. 제안된 방법은 안정화된 역 확산 방정식(Stabilized Inverse Diffusion Equation : SIDE)에 기반한 비선형 다중스케일 필터링을 사용하여 공간적 분할을 수행한다. 또한 구해진 초기 분할된 영역들은 인접 영역 그래프 (Region Adjacency Graph : RAG)를 사용하여 병합된다. 본 논문에서는 통계적 유의성 검사(Statistical significance test)와 시변 메모리(Time-variant memory)를 시간적 분할 방법으로 사용하며 구해진 공간적 분할과 시간적 분할을 결합하여 최종 객체 영역을 효과적으로 분할한다. 본 논문에서 제안된 공간적 분할 방법은 기존의 형태학적 Watershed 알고리즘에 비해 잡음에 강인한 분할 특성을 나타내었으며 기존의 A. Neri의 방법과 비교하였을 때, 최종 분할된 객체 영역의 정확도 비율이 Akiyo는 43%, Claire는 29% 정도 향상됨을 확인할 수 있었다.

key Words : video object segmentation; nonlinear multiscale filtering; SIDE.

ABSTRACT

Object-based coding, such as MPEG-4, enables various content-based functionalities for multimedia applications. In order to support such functionalities, as well as to improve coding efficiency, each frame of video sequences should be segmented into video objects. In this paper, we propose an effective video object segmentation method using nonlinear multiscale filtering and spatio-temporal information. Proposed method performs a spatial segmentation using a nonlinear multiscale filtering based on the stabilized inverse diffusion equation(SIDE). And, the segmented regions are merged using region adjacency graph(RAG). In this paper, we use a statistical significance test and a time-variant memory as temporal segmentation methods. By combining of extracted spatial and temporal segmentations, we can segment the video objects effectively. Proposed method is more robust to noise than the existing watershed algorithm. Experimental result shows that the proposed method improves a boundary accuracy ratio by 43% on "Akiyo" and by 29% on "Claire" than A. Neri's Method does.

I. 서론

MPEG-4 비디오 표준에서 채택된 객체 기반 부호화는 기존 MPEG-1이나 MPEG-2와 같은 비디오

* 인하대학교 전자공학과 (g1991205@inhavision.inha.ac.kr), ** 한국전자통신연구원 전파방송연구실
 논문번호 : 030272-0625, 접수일자 : 2003년 6월 26일

부호화 표준과 구분되는 중요한 특징으로 나타난다. 이러한 객체 기반 부호화는 압축 효율의 향상과 더불어 개별 객체 단위의 전송, 조작, 객체간의 다양한 상호작용, 내용 정보의 재사용과 같은 멀티미디어 응용에 효과적으로 적용할 수 있는 다양한 기능들을 제공한다^[1]. 그러나 위와 같은 객체 기반 기능들이 구현되기 위해서는 비디오 데이터로부터 고정된 배경과 객체를 분할하는 비디오 객체 분할이 효과적으로 구현되어야 한다. 일반적인 비디오 객체 분할 기법은 분할을 결정하는 기준에 따라 크게 두 가지로 나눌 수 있다^[2]. 첫 번째로 공간적 균질성 (spatial homogeneity)을 분할기준으로 사용하는 방법이다. 이 방법은 다음과 같은 단계로 구성된다. 우선 형태학적 필터를 통해 영상을 단순화시키고 Watershed 알고리즘을 이용하여 영역을 추출한다. 이후 각 영역들의 움직임 벡터들을 움직임 추정에 의해 구한 뒤, 유사한 움직임을 갖는 영역들을 병합하여 최종 객체 영역을 결정한다^[3]. 하지만 Watershed 알고리즘을 사용한 이와 같은 방법은 계산이 복잡하고 영상의 기울기 잡음 (gradient noise)에 의해 과분할 되는 단점이 있다고 알려져 있다^[4]. 두 번째는 변화 검출 (change detection)을 분할 기준으로 사용하는 방법이다^{[5][6][7]}. 이 방법은 우선 연속되는 두 프레임간의 차분을 통해 움직임 객체의 영역을 검출한다. 그리고 영역 경계가 잘 나타나도록 시공간 정보에 기반한 후처리를 한다. 그러나 이 방법은 움직이는 객체의 내부가 평탄한 경우, 객체의 움직임에 의해 가려지는 배경 (covered background)과 나타나는 배경 (uncovered background)이 객체 영역으로 검출되어 실제 객체 경계를 정확하게 검출하지 못하는 단점이 있다. 최근에는 위와 같은 두 분할 기준을 결합한 방법으로 시공간 (spatio-temporal) 정보를 사용한 방법이 연구되고 있다. 이 방법은 공간적 분할 (spatial segmentation)을 통해 객체 영역의 경계를 구하고 시간적 분할 (temporal segmentation)을 통해 배경과 객체를 구분하여 객체를 분할한다^[8].

본 논문에서는 비선형 다중스케일 필터링과 시공간 정보를 사용한 효과적인 비디오 객체 분할 기법을 제안한다. 제안된 방법은 먼저 잡음에 취약한 Watershed 알고리즘의 단점을 해결하기 위해 안정화된 역 확산 방정식 (Stabilized Inverse Diffusion Equation : SIDE)에 기반한 비선형 다중스케일 필터링을 사용하여 잡음에 강인한 초기 분할을 한다. 이후 초기 분할된 각 영역들은 인접 영역 그래프

(Region Adjacency Graph : RAG)를 사용하여 병합된다. 제안된 공간적 분할 방법은 잡음에 강인한 분할 특성과 객체 경계를 정확하게 분리해 내는 장점이 있다. 또한 본 논문에서는 Y. Tsaig^[11]와 F. Long^[8]이 사용한 통계적 유의성 검사 (statistical significance test)^[5]와 시변 메모리 (time-variant memory)를 시간적 분할 방법으로 사용하며, 구해진 공간적 분할과 시간적 분할을 결합하여 최종 객체 영역을 분할한다.

본 논문의 구성은 다음과 같다. 2장에서는 공간적 분할 방법으로 SIDE에 기반한 비선형 다중스케일 필터링 분할 방법과 RAG를 통한 영역 병합 방법을 설명한다. 또한 3장에서는 본 논문에서 사용한 시간적 분할 방법과 시공간 정보의 결합을 통한 최종 객체 분할에 관하여 설명한다. 4장에서는 제안한 방법에 의해 구한 실험 결과를 제시하고 5장에서 결론 및 향후 과제에 관하여 논한다.

II. 공간적 분할 방법

1. 비선형 다중스케일 필터링

일반적으로 영상 신호를 분할하기 위해 영상의 값이 영역 내부에서는 유사하고 영역 경계에서는 변화가 크게 나타나는 성질을 이용한다. 이때 영역 경계를 유지하면서 영역 내부를 단순화시켜 유사성을 높이는 방법을 사용하면 영상 분할의 효율을 증대시킬 수 있다. 다중스케일 필터의 기본적인 개념은 그림 1과 같이 입력된 신호를 반복적인 필터링을 통해 간단하게 만드는 것이다. 이러한 다중스케일 필터링이 반복적으로 수행되면 영상 신호의 경우 잡음이 제거되고 영상이 단순화되어 영상 분할의 효율을 높일 수 있다. 이때 반복되는 각 단계의 수를 스케일 (scale)이라고 하며 각 출력 신호의 집합을 다중스케일 (multiscale)이라고 한다^[10].

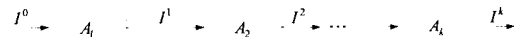


그림 1. 반복 필터링을 통한 multiscale 생성

두 개의 공간 변수 x 와 y 를 가지는 입력 영상 신호를 $I^0(x, y)$ 라고 할 때, 다중스케일 필터링은 편미분 방정식 (Partial Differential Equation : PDE)으로 식 (1)과 같이 표현된다.

$$I_n = A(I, I_x, I_y, I_{xx}, I_{yy}, I_{xy}) \quad (1)$$

$$K(0, x, y) = I^0(x, y),$$

이때 변수 n 은 스케일이라고 하며 $0 \leq n < \infty$ 의 범위를 가진다. 위의 식 (1)은 스케일 $0 \leq n < \infty$ 에 대해 다중스케일 영상 $I(n, x, y)$ 를 생성한다. 이때 사용되는 편미분 방정식은 영상의 값을 주변으로 확산시켜 평탄하게 하는 확산 방정식(diffusion equation)과 반대 경우의 역 확산 방정식(inverse diffusion equation)으로 구분된다. 또한 편미분 방정식이 선형인 경우를 선형 다중스케일 필터링이라고 하고 비선형인 경우를 비선형 다중스케일 필터링이라고 한다. 선형 다중스케일 필터링의 대표적인 예는 식 (2)와 같은 Tikhonov의 선형 열 전달 방정식(linear heat conduction equation)이 있다. 이때 식 (2)의 $div()$ 는 발산(divergence)이고 ∇I 는 영상 기울기(image gradient)이다.

$$I_n = I_{xx} + I_{yy} = div(\nabla I) \quad (2)$$

$$K(0, x, y) = I^0(x, y),$$

선형 열 전달 방정식을 사용한 다중스케일 필터링의 결과는 그림 2 (b)와 같이 영상의 기울기에 관계없이 확산만 이루어져 영역 경계를 포함한 영상 전체가 블러링 된다. 영상 분할을 위한 영상의 단순화는 영역의 경계가 유지되어야 하는데, 이러한 선형 열 전달 방정식을 사용한 경우에는 영역의 경계 정보가 블러링 되어 분할에 사용하기에는 어려움이 있다.

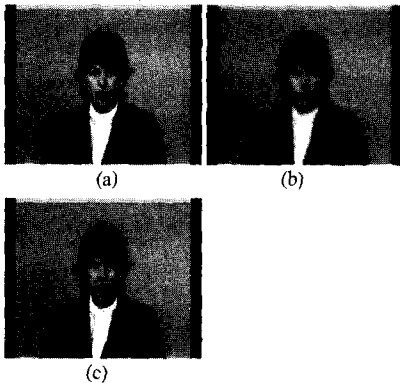


그림 2. 선형 및 비선형 다중스케일 필터링의 예 (a) 입력 영상, (b)Tikhonov 선형 열 전달 방정식, (c) Perona-Malik 비선형 확산 방정식

전체적으로 영상이 확산만 되는 선형 열 전달 방

정식의 단점을 해결하기 위해 영상의 기울기 값에 따라 확산 정도가 다르게 적용되는 비선형 확산 방정식을 사용한 방법이 Perona와 Malik에 의해 제안되었다^[9]. 이 방법은 영상의 기울기 값에 따라 평탄한 영역은 확산이 되고 영상의 경계 부분은 역 확산이 되어 영상을 단순화시키면서 영역의 경계를 그대로 유지하는 장점이 있다. 이러한 Perona-Malik 비선형 확산 방정식은 식 (3)과 같이 정의된다.

$$I_n = div(G(\|\nabla I\|)\nabla I) \quad (3)$$

$$K(0, x, y) = I^0(x, y),$$

식 (3)에서 나타나는 $G(\|\nabla I\|)$ 는 비선형 확산 계수로서 $G(0) = 1$ 이다. 이 비선형 확산 계수가 최대 값인 1이 되었을 때에는 식 (3)이 선형 열 전달 방정식과 같은 식이 되는 것을 알 수 있다.

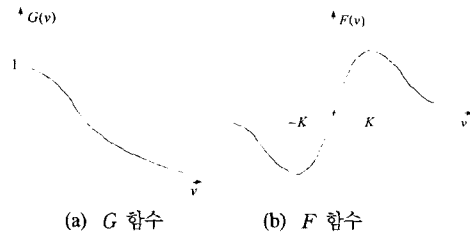


그림 3. Perona-Malik 비선형 확산 방정식에 사용되는 함수 특성

비선형 확산 계수 $G(\|\nabla I\|)$ 는 그림 3 (a)와 같이 영상의 기울기 값이 증가하면 확산이 감소된다. 이를 통하여 영상의 기울기 값이 작은 평탄한 영역 내부는 확산이 되고 기울기 값이 큰 영역 경계 부분은 역 확산이 되어 그림 2 (c)와 같이 영상이 단순화 되면서도 영역의 경계가 유지되는 결과를 얻을 수 있다. 이러한 결과는 앞서 설명한 바와 같이 영역 경계를 유지하면서 영역 내부의 유사성을 높이므로 영상 분할의 효율을 높일 수 있다.

Perona-Malik 비선형 확산 방정식에서 $F(\nabla I) = G(\|\nabla I\|)\nabla I$ 라고 정의하면 그림 3 (b)와 같은 함수로 나타나게 된다. 이때 상수 K 는 확산과 역 확산간의 기준이 된다. 즉 K 보다 작은 기울기 값을 갖는 부분은 평탄한 영역 내부로 판단되어 확산을 통한 단순화가 이루어지며 K 보다 큰 기울기 값을 갖는 부분은 영역의 경계로 판단되어 역 확산이 이루어진다. 그러나 이러한 Perona-Malik 비선형 확산 방정식에서는 어느 정도의 기울기를 영역 경계로 판단하여 K 를 설정해야 하는가가 문제이다.

만일 K 값이 작게 설정될 경우에는 평탄한 영역까지 지역 확산을 통해 강조가 되어 오히려 영역 내부의 잡음을 강조하는 결과가 나타난다. 또한 K 값이 크게 설정되면 선형 다중스케일 필터링과 같이 영역의 경계가 블러링 되는 결과가 나타난다. 이처럼 Perona-Malik 비선형 확산 방정식의 경우에는 확산과 비확산을 결정하는 K 값에 의해 안정화되지 않은 결과가 나타나게 된다^[12].

위와 같은 Perona-Malik 비선형 다중스케일 필터링을 거친 단순화된 영상을 분할하면 분할의 효율을 높일 수 있다. 그러나 앞서 설명한 바와 같이 안정화되지 않은 방정식의 사용으로 오히려 분할 결과가 나쁘게 나타날 수도 있다. 또한 이제까지 설명한 방법들은 영상 복원과 관련하여 영상의 잡음을 제거하기 위해 주로 사용되는 것으로 영상 분할에 직접 사용하기에는 한계가 있다. 본 논문에서는 Perona-Malik 방정식이 가진 불안정한 전개에 의한 단점을 없애기 위해 F 함수에서 K 값을 0으로 설정하여 구해지는 안정화된 역 확산 방정식(SIDE)을 사용하여 다중스케일 필터링을 정의한다. 또한 앞서의 방법들은 영상의 화소를 기준으로 화소값을 변화시키는 처리를 하지만 제안된 방법은 영상 분할을 위해 영역을 기준으로 처리를 한다.

2. SIDE에 기반한 비선형 다중스케일 필터링 분할 방법

본 논문에서는 I. Pollak에 의해 제안된 안정화된 역 확산 방정식(SIDE)에 기반한 다중스케일 필터링을 이용하여 영상의 공간적 분할을 한다. SIDE를 이용한 분할 방법은 잡음에 강인하며 안정화된 분할 결과를 나타낸다^{[10][11][12]}. SIDE는 Perona-Malik 방정식에서 사용되는 F 함수에서 K 값을 0으로 설정한 것과 같이 정의된다. 이때 K 값이 0으로 설정됨으로서 SIDE는 확산 부분이 없는 역 확산으로만 구성된 방정식이 되고 K 값을 설정할 필요가 없어 지므로 안정화된 방정식이 된다. 이러한 SIDE에서 사용되는 F 함수는 다음 식 (4)와 같이 정의되며 이를 그림 4 (a)에 나타내었다.

$$\begin{aligned} F'(v) &\leq 0 \text{ for } v \neq 0, \\ F(0^+) &> 0, \\ F(v_1) &= F(v_2) \leftrightarrow v_1 = v_2. \end{aligned} \quad (4)$$

본 논문에서는 식 (4)를 만족하는 SIDE로서 다음의 식 (5)와 그림 4 (b)와 같은 F 함수를 사용한다.

$$F(v) = \begin{cases} \operatorname{sgn}(v) - \frac{v}{L} & \text{if } -L \leq v \leq L, \\ 0 & \text{else} \end{cases} \quad (5)$$

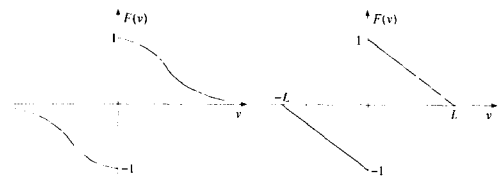


그림 4. SIDE F 함수 특성 및 본 논문에서 사용한 F 함수

본 논문에서 사용하는 SIDE에 기반한 다중스케일 필터링 방법은 영상 분할을 목적으로 하여 화소 단위가 아닌 영역 단위로 다중스케일 필터링이 수행된다. 이를 위해 위의 식 (5)의 F 함수를 사용하여 다음 식 (6)과 같이 영역 값의 변화를 구하는 방정식이 사용된다.

$$\dot{I}_{Ri} = \frac{1}{m_{Ri}} \sum_{Rj \in NRi} F(I_{Ri} - I_{Rj}) p_{ij} \quad (6)$$

여기서 I_{Ri} 는 영역 Ri 의 값이고 필터링이 진행될 때, 스케일에 따른 변화율은 \dot{I}_{Ri} 이다. 또한 NRi 는 영역 Ri 에 인접한 영역들의 집합이고 p_{ij} 는 영역 Ri 와 Rj 간의 인접되어 있는 픽셀의 개수이다. 그리고 m_{Ri} 는 영역 Ri 의 면적이 된다.

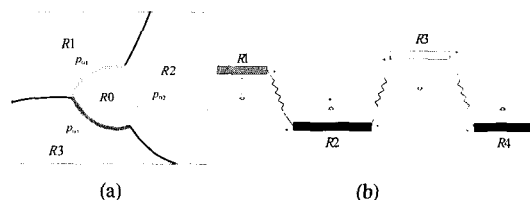


그림 5. 제안된 SIDE 기반 다중스케일 필터링 예제 및 분할 개념 (a) 새로운 영역 값 계산의 예, (b) SIDE 기반 영역 분할의 개념

예를 들어 영역이 그림 5 (a)와 같을 때, 영역 $R0$ 에는 영역 $R1$, $R2$, $R3$ 가 접해 있으며 이들 영역과 접해 있는 픽셀의 개수는 각각 p_{01} , p_{02} , p_{03} 가 된다. 이때 새로운 영역 값을 계산하기 위해 위의 식 (6)에 의하여 영역 값의 변화율 \dot{I}_{R0} 를 식 (7)과 같이 계산한다. 이와 같이 구해진 \dot{I}_{R0} 를 현재의 I_{R0} 값에 더하여 영역 $R0$ 의 새로운 값을 구한다.

$$I_{R0} = \frac{1}{m_{R0}} \{ F(I_{R0} - I_{R1})p_{01} + F(I_{R0} - I_{R2})p_{02} + F(I_{R0} - I_{R3})p_{03} \} \quad (7)$$

그림 5 (b)에는 인접한 영역 값의 차이에 의해 새로운 영역 값이 설정되는 개념을 1차원적으로 나타낸다. 그림 5 (b)와 같이 각 이웃하는 영역들은 서로 스프링으로 연결된 것처럼 생각할 수 있다. 이때 스프링에 가해지는 힘은 식 (6)에 의해 구해진다. 이렇게 구해진 힘에 의해 각 영역의 값은 서서히 변화된다. 영역의 값이 변화하다가 인접한 영역이 같은 값으로 위치하면 이 영역들은 하나의 영역으로 병합된다. 이때 식 (6)에서 합해진 힘이 영역의 면적으로 나누어지는 데, 이를 통해 면적이 큰 영역은 상대적으로 작은 변위가 생기고 면적이 작은 영역들은 큰 변위가 생기게 된다. 따라서 면적이 작은 영역이 우선적으로 면적이 큰 영역에 병합되는 효과가 나타난다.

본 논문에서는 영상의 공간적 분할을 위해 다음과 같은 단계를 거쳐 초기 분할을 수행한다.

- 1 단계 : 레이블링을 통하여 초기 영역을 설정한다. 이때 영역의 면적 m_{Rk} , $k=\{0,1,\dots,l-1\}$ 와 영역 값 I_{Rk} , $k=\{0,1,\dots,l-1\}$ 를 계산한다. 이때 전체 레이블 수는 l 이다.
- 2 단계 : 식 (6)에 의해 계산된 값으로 각 영역 값을 갱신해 준다.
- 3 단계 : 영역 값이 같은 인접한 영역은 병합한다. 이때 m_{Rk} 와 I_{Rk} 그리고 l 을 갱신해 준다.
- 4 단계 : 2 단계로 간다. 이때 l 값이 미리 정해 놓은 영역의 수 T , 보다 작으면 종료한다.

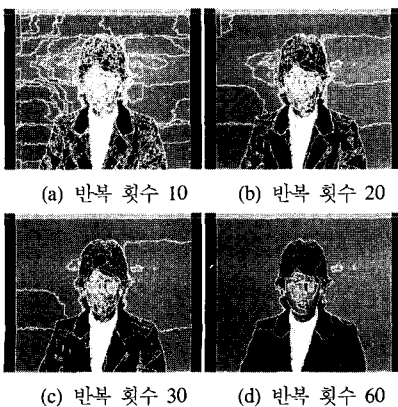


그림 6. 스케일에 따른 SIDE 기반 다중스케일 분할 결과

그림 6은 위와 같은 방법에 따라 분할된 결과이다. 필터링이 반복적으로 수행됨에 따라 영역들이 경계를 효과적으로 보존하면서 병합되는 것을 알 수 있다.

3. RAG를 통한 영역 병합 방법

본 논문에서는 앞 절에서 설명한 SIDE에 기반한 비선형 다중스케일 필터링을 통해 영상을 초기 분할한 후, 인접 영역 그래프(RAG)를 이용하여 영역을 병합한다. 인접 영역 그래프는 그림 7 (b)와 같이 정의된다. 이때 각 영역은 노드가 되고 노드사이의 거리는 인접한 두 영역간의 거리로 정의된다.

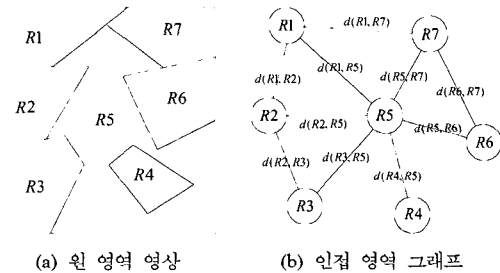


그림 7. 인접 영역 그래프

본 논문에서는 인접한 두 영역의 거리를 다음과 같은 식 (8)과 (9)를 이용하여 구한다. 즉 두 영역의 평균 값 차이를 영역간의 거리로 정의한다.

$$\mu_{Ri} = \frac{1}{m_{Ri}} \sum_{x,y \in Ri} I(x,y) \quad (8)$$

$$d(Ri, Rj) = | \mu_{Ri} - \mu_{Rj} | \quad (9)$$

본 논문에서는 RAG를 구한 뒤, 다음과 같은 두 가지 단계를 거쳐 영역 병합을 수행한다.

- 1 단계 : 우선 영역의 면적이 T_m 보다 작은 영역을 대상으로 거리가 최소인 인접한 영역과 병합을 수행한다. 영역이 병합되면 RAG를 갱신한다. 모든 영역의 면적이 T_m 보다 크게 되면 다음 단계로 간다.
- 2 단계 : 이 단계에서는 모든 영역에 대해 거리가 T_d 이하인 영역을 병합한다. 앞 단계와 마찬가지로 영역이 병합되면 RAG를 갱신한다. 영역 사이의 거리가 모두 T_d 보다 클 경우 영역 병합을 종료한다.

III. 시간적 분할 방법

1. 통계적 유의성 검사에 의한 변화 검출 마스크 추출

본 논문에서는 변화 검출 마스크(Change Detection Mask : CDM)를 구하기 위해 Y. Tsaig^[1]와 F. Long^[8]이 사용한 차분 제곱의 정규화된 합(normalized sum)을 검정통계량으로 사용한다. 이 검정통계량은 통계적으로 χ^2 분포(chi-square distribution)를 가진다고 알려져 있다^{[1][7][8]}. 만약 시간이 t 일 때의 검정통계량을 $\nabla_t(x, y)$ 라 하면 이때 구해지는 변화 검출 마스크 $CDM_{t-1,t}$ 은 식 (10)과 같이 문턱값 T_a 를 사용하여 구해진다. 이때 사용되는 문턱값 T_a 는 유의성 수준(significance level) α 에 의한 유의성 검사(significance test)를 통해 결정한다^{[1][8]}.

$$CDM_{t-1,t}(x, y) = \begin{cases} 1 & \text{if } \nabla_t(x, y) > T_a \\ 0 & \text{else} \end{cases} \quad (10)$$

본 논문에서는 3×3 크기 블록에서 $\nabla_t(x, y)$ 를 구한다. 또한 $\nabla_t(x, y)$ 를 구하기 위해 필요한 배경 영역의 분산은 다음과 같이 추정된 배경 영역에서 구한다. 우선 CDM을 구하려는 프레임이 장면 분할(scene-cut)이후 첫 번째 프레임과 같이 배경 영역이 결정되지 않은 경우에는 차분의 제곱이 큰 값에서부터 10%에 해당하는 부분을 제외한 배경 영역에서 분산을 구한다. 그리고 그 밖의 경우에는 시공간적으로 최종 분할된 결과에 의해 배경 영역의 분산을 구한다.

2. 시변 메모리 및 변화 검출 마스크 갱신 방법

본 논문에서는 변화 검출 마스크(CDM)와 객체 마스크(Object Mask : OM)를 갱신하기 위해서 두 개의 시변 메모리를 사용한다. 우선 변화 검출 마스크를 갱신하기 위해, 시간이 t 일 때 위치 (x, y) 의 변화 상태를 식 (11)과 같이 K_{CDM} 의 깊이(depth)를 가지는 시변 메모리 M_t^{CDM} 에 저장한다.

$$M_t^{CDM}(x, y) = \begin{cases} K_{CDM} & \text{if } (x, y) \in CDM_{t-1,t} \\ \max(0, M_{t-1}^{CDM}(x, y) - 1) & \text{if } (x, y) \notin CDM_{t-1,t} \end{cases} \quad (11)$$

이렇게 구해진 $t-1$ 번째 시변 메모리 M_{t-1}^{CDM} 와 $t-1$ 번째 최종 객체 마스크 OM_{t-1} 에 의해 t 번째 프레임에 해당하는 $CDM_{t-1,t}$ 를 식 (12)와 같이 갱신한다. 이를 통하여 일관적인 변화 검출이 가능하게 된다.

$$CDM_{t-1,t}(x, y) = CDM_{t-1,t} \vee \begin{cases} OM_{t-1}(x, y) & \text{if } M_{t-1}^{CDM}(x, y) > 0 \\ 0 & \text{if } M_{t-1}^{CDM}(x, y) = 0 \end{cases} \quad (12)$$

또한 본 논문에서는 이전 K_{OM} 개의 객체 마스크에서 검출된 영역을 참조하여 현재 프레임에 해당하는 최종 객체 마스크를 판단한다. 이를 위해 시간이 t 일 때의 최종 검출 마스크 OM_t 에서의 변화는 식 (13)과 같은 시변 메모리 M_t^{OM} 에 저장된다.

$$M_t^{OM}(x, y) = \begin{cases} \min(K_{OM}, M_{t-1}^{OM}(x, y) + 1) & \text{if } (x, y) \in OM_t \\ \max(0, M_{t-1}^{OM}(x, y) - 1) & \text{if } (x, y) \notin OM_t \end{cases} \quad (13)$$

3. 시공간 정보의 결합

본 논문에서는 앞서 설명한 공간적 분할 정보와 시간적 분할 정보를 결합하여 객체 마스크를 구한다. 우선 t 번째 프레임을 SIDE에 기반한 비선형 다중스케일 필터링으로 분할한 레이블링 영역을 S_t^{sp} 로 정의한다. 그리고 t 번째 프레임과 $t-1$ 번째 프레임간의 CDM인 $CDM_{t-1,t}$ 에 나타난 변화 영역을 S_t^{mp} 로 정의한다. 이 두 영역을 사용하여 공간적으로 레이블링된 영역 내부에 존재하는 변화 영역의 면적 비율을 식 (14)와 같이 구한다.

$$\eta = \frac{Area(S_t^{sp} \cap S_t^{mp})}{Area(S_t^{sp})} \quad (14)$$

또한 본 논문에서는 식 (13)를 이용하여 구해진 이전 $t-1$ 번째 프레임에 대한 시변 메모리 M_{t-1}^{OM} 의 값이 0 이 아닌 영역을 S_{t-1}^{OM} 으로 정의한다. 이렇게 정의된 영역 S_{t-1}^{OM} 과 앞서 정의한 레이블링 영역 S_t^{sp} 을 이용하여 다음 식 (15)과 같이 공간적으로 레이블링된 영역 내부에 존재하는 이전 객체 마스크 영역의 면적 비율을 구한다.

$$\gamma = \frac{Area(S_i^{spl} \cap S_{i-1}^{OM})}{Area(S_i^{spl})} \quad (15)$$

본 논문에서는 η 가 문턱치 T_η 보다 크거나 γ 가 문턱치 T_γ 보다 큰 경우를 객체 영역으로 판단하고, 만일 η 와 γ 가 모두 각각의 문턱치 T_η 와 T_γ 보다 작을 경우에는 배경 영역으로 판단하여 최종 객체 마스크 OM_i 를 구한다.

IV. 실험 결과

1. 제안된 객체 분할 방법 실험

본 논문에서는 제안된 방법의 객체 분할 성능을 평가하기 위해 352×288 크기를 가지는 CIF 포맷의 Akiyo와 Claire 시퀀스 데이터를 사용하였다. 제안된 방법에 사용되는 실험 파라미터 값은 표 1과 같이 각 데이터에 따라 적용하였으며 프레임 차분을 위한 프레임 간격은 2프레임으로 설정하였다.

표 1. 실험에 사용한 파라미터 값

Data	L	T _i	T _m	T _d	α	T _η	T _γ	K _{CDM}	K _{OM}
Akiyo	25	2000	25	0.1	10 ⁻³	0.5	0.5	4	2
Claire	20	2000	25	25.1	10 ⁻³	0.15	0.5	4	4

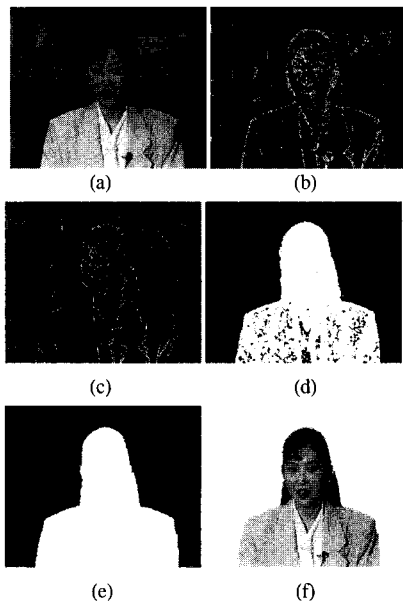


그림 8. Akiyo 영상에서의 객체 분할 과정 (18번째 프레임 영상) (a) 원영상, (b) 다중스케일 필터링 초기 분할, (c) RAG 병합 결과, (d) 변화 검출 마스크, (e) 최종 객체 영역 마스크, (f) 추출된 객체 영상

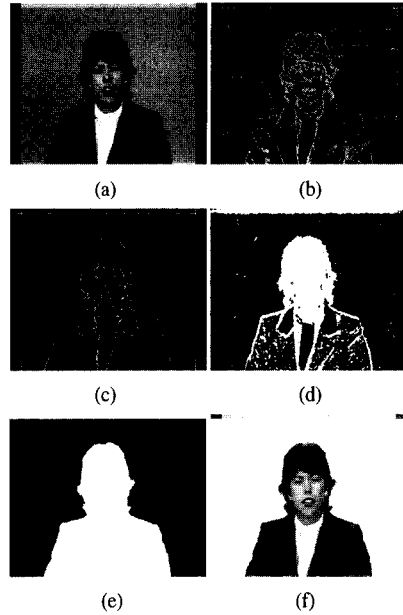


그림 9. Claire 영상에서의 객체 분할 과정 (17번째 프레임 영상) (a) 원영상, (b) 다중스케일 필터링 초기 분할, (c) RAG 병합 결과, (d) 변화 검출 마스크, (e) 최종 객체 영역 마스크, (f) 추출된 객체 영상

본 논문에서 제안한 객체 분할 과정은 그림 8과 9에 나타내었다. 그림 8과 9의 (b)에 나타난 제안된 다중스케일 필터링 분할을 사용한 초기 분할 영역은 Akiyo는 평균 1952개 그리고 Claire는 평균 1949개의 영역으로 분할되었다. RAG를 이용한 병합을 거친 후에는 그림 8과 9의 (c)에 나타난 바와 같이 Akiyo는 평균 168개의 영역으로 그리고 Claire는 평균 39개의 영역으로 효과적으로 병합된다. 시간적 분할을 통한 CDM은 그림 8과 9의 (d)에 그리고 공간 분할된 영역과 결합된 최종 객체 영역 마스크와 추출된 객체 영상은 그림 8과 9의 (e)와 (f)에 각각 나타내었다.

제안한 방법을 사용하여 각 프레임에서 추출된 객체 영상들을 그림 10과 11에 나타내었다. 이와 같이 제안된 비선형 다중스케일 필터링을 사용한 시공간적 분할 방법은 움직임 객체 영역을 효과적으로 분할할 수 있었다.



그림 10. Akiyo 각 프레임 영상에서 추출된 객체 영상



그림 11. Claire 각 프레임 영상에서 추출된 객체 영상

2. 기존 방법과 경계선 정확도 비교

분할된 객체 영역의 경계가 얼마나 정확하게 분할되었는지를 알아보기 위해, 제안된 방법과 기존 방법을 경계선 정확도 비율을 통하여 비교하였다. 본 논문에서는 기존 방법으로 A. Neri가 제안한 HOS(high order statistics)와 형태학적 필터링을 이용한 객체 분할 방법⁶⁾을 구현하여 비교하였다. 그림 12에 Akiyo영상에 대해 A. Neri의 방법으로 분할한 객체 영상과 제안된 방법으로 분할한 객체 영상을 나타내었다.

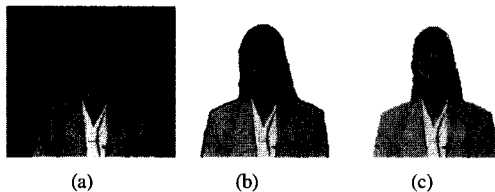
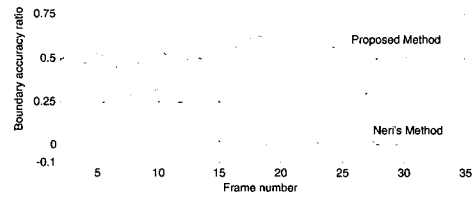
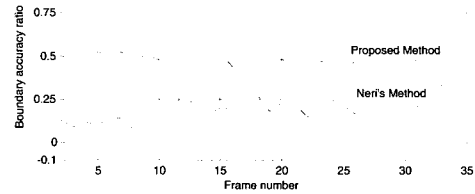


그림 12. Akiyo 영상에서의 A. Neri의 방법과 제안된 방법에 의한 분할된 객체 영역 (20번째 프레임 영상) (a) 원영상, (b) A. Neri의 방법, (c) 제안된 방법

본 논문에서는 각 분할 방법을 통한 객체 영역 경계의 정확도를 측정하기 위해 실제 객체 영역의 경계선 픽셀과 각 분할 방법을 통해 구해진 객체 영역의 경계선 픽셀과 일치하는 픽셀의 수를 계수하고 이를 실제 객체 영역의 경계선 픽셀수로 나누어 경계선 정확도 비율(boundary accuracy ratio)을 구하였다. 그림 13에 Akiyo와 Claire의 33개의 프레임을 대상으로 제안된 방법과 A. Neri의 방법을 이용하여 구해진 경계선 정확도 비율을 그래프로 나타내었다. 또한 표 2에 구해진 경계선 정확도 비율을 평균하여 각 데이터 별로 나타내었다. 실험 결과에 따르면 제안된 방법이 A. Neri의 방법에 비해 경계선 정확도 비율이 평균적으로 Akiyo는 43% 그리고 Claire는 29% 정도 높게 나오는 것을 알 수 있다. 즉 제안된 방법이 A. Neri의 방법보다 객체 영역의 경계를 정확히 분할하는 것을 알 수 있다.



(a) Akiyo에서의 경계선 정확도 비교



(b) Claire에서의 경계선 정확도 비교

그림 13. 제안된 방법과 A. Neri 방법의 경계선 정확도 비율 비교

표 2. 제안된 방법과 A. Neri 방법의 평균 경계선 정확도 비율

Data	제안한 방법	A. Neri의 방법
Akiyo	0.524194	0.095563
Claire	0.464356	0.176033

3. 잡음 환경에서의 제안된 비선형 다중스케일 분할 결과

본 논문에서 제안된 다중스케일 필터링 분할 방법이 잡음 영상을 효과적으로 분할할 수 있는지를 실험하기 위해 합성 영상에 가우시안 잡음을 포함시키고 공간 분할을 수행하였다. 그림 14는 이러한 잡음이 포함된 영상에 대한 Watershed 알고리즘의 분할 결과와 제안된 다중스케일 필터링 분할 방법의 분할 결과를 보여준다. 이때 사용한 Watershed 알고리즘은 L. Vincent 와 P. Soille가 제안한 Immersion 기반 Watershed 알고리즘을 사용하였다¹³⁾. 그림 14의 (b)와 (e)에 나타난 바와 같이 Watershed 알고리즘은 잡음에 민감하여 영역이 과분할되는 결과가 나타난다. 반면에 그림 14의 (c)와 (f)에 나타난 바와 같이 제안된 다중스케일 필터링 분할 방법은 Watershed 알고리즘에 비해 잡음에 강인하고 효과적인 분할이 이루어지는 것을 알 수 있다. 또한 제안된 방법을 통한 분할 결과에서는 잡음에 의해 나타나는 영역들이 객체 영역에 비해 상대적으로 크기가 작게 나타난다. 따라서 제안된 다중스케일 필터링 분할 방법을 거친 후에 영역 면적을

이러한 RAG 영역 병합이 수행되면 이러한 영역들은 효과적으로 제거될 수 있음을 알 수 있다.

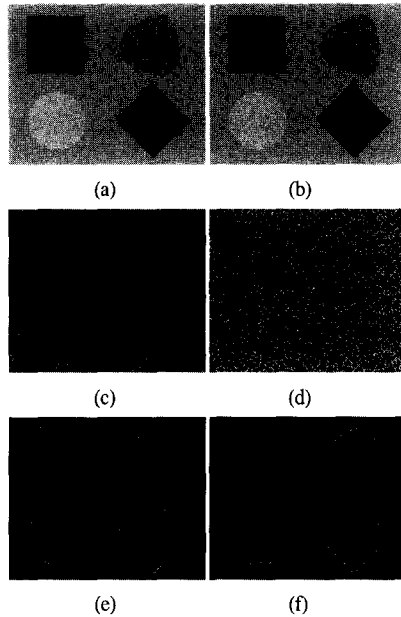


그림 14. 잡음 환경 하에서의 Watershed와 제안된 분할 방법의 분할 결과 비교 (a) $\sigma=5$ 인 가우시안 잡음 영상, (b) $\sigma=10$ 인 가우시안 잡음 영상, (c) (a)의 Watershed 분할 결과, (d) (b)의 Watershed 분할 결과, (e) (a)의 제안된 분할 결과, (f) (b)의 제안된 분할 결과

V. 결론

본 논문에서는 SIDE에 기반한 비선형 다중스케일 필터링을 통한 공간적 분할과 통계적 유의성 검사 방법을 통한 시간적 분할을 결합한 비디오 객체 분할 방법을 제안하였다. 제안된 공간적 분할 방법은 기존의 Watershed와 비교하여 잡음 환경에 강한 특징을 나타냈고 RAG와 결합하여 효과적인 공간 분할을 할 수 있었다. 또한 제안된 시간적 분할 방법은 시변 메모리를 사용하여 이전 변화 검출 마스크와 이전 프레임에서 검출한 객체 마스크를 유지하여 시간적으로 일관적인 객체 분할을 할 수 있었다. 제안된 방법은 기존의 A. Neri의 분할 방법에 비해 분할된 객체 경계의 정확도가 평균적으로 Akiyo에서는 43% 그리고 Claire에서는 29%정도 향상되는 것을 알 수 있었다. 향후 과제로는 제안된 방법에서 사용되는 파라미터들이 영상 데이터의 특성을 고려하여 설정되도록 각 파라미터에 대한 분석이 필요할 것으로 생각되며 배경과 비슷한 밝기 값을 가진 객체 영역이 배경에 병합되는 문제점을 해결하는 방법

이 추후 연구되어야 할 것으로 생각된다.

참고 문헌

- [1] Y. Tsaig and A. Averbuch, "Automatic segmentation of moving objects in video sequences: A region labeling approach," *IEEE Trans. on Circuits and Syst. Video Technol.*, vol. 12, no. 7, pp. 597-612, July 2002.
- [2] S. Y. Chien, S. Y. Ma, and L. G. Chen, "Efficient moving object segmentation algorithm using background registration technique," *IEEE Trans. on Circuit and Syst. Video Technol.*, vol. 12, no. 7, pp. 577-586, July 2002.
- [3] D. Wang, "Unsupervised video segmentation based on watersheds and temporal tracking," *IEEE Trans. on Circuit and Syst. Video Technol.*, vol. 8, no. 5, pp. 539-546, Sept. 1998.
- [4] M. Kim, J. G. Choi, D. Kim, H. Lee, M. H. Lee, C. Ahn, and Y. Ho, "A VOP generation tool : Automatic segmentation of moving objects in image sequences based on spatio-temporal information," *IEEE Trans. on Circuit and Syst. Video Technol.*, vol. 9, no. 8, pp. 1216-1226, Dec. 1999.
- [5] T. Aach, A. Kaup, and R. Mester, "Statistical model-based change detection in moving video," *Signal Processing*, vol. 31, pp.165-180, Mar. 1993.
- [6] A. Neri, S. Colonnese, G. Russo, and P. Talone, "Automatic moving object and background separation," *Signal Processing*, vol. 66, pp. 219-232, Apr. 1998.
- [7] R. Mech and M. Wollborn, "A noise robust method for 2D shape estimation of moving objects in video sequences considering a moving camera," *Signal Processing*, vol. 66, pp. 203-217, Apr. 1998.
- [8] F. Long, D. Feng, H. Peng, and W. Siu, "Extracting semantic video objects," *IEEE Computer Graphics and Applications*, vol. 21, no. 1, pp. 48-55, January/February 2003.
- [9] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. on Pattern Anal. Machine Intell.*, vol. 12,

