

논문 2011-48SP-1-18

클러스터링과 마르코프 랜덤 필드를 이용한 배경 모델링 기법 제안

(Improving Clustering-Based Background Modeling Techniques Using
Markov Random Fields)

한 희 일*, 박 수 빈**

(Hee-il Hahn and Soobin Park)

요 약

본 논문에서는 마르코프 랜덤 필드(Markov random fields: MRF) 기반으로 배경을 모델링하는 방식과 함께 관련 파라미터들을 추정하는 알고리즘을 제안한다. 화소 기반의 배경 모델링 기법은 인근 화소 간의 연관성을 고려하지 않고 화소 단위의 시간적 변화에 대한 통계적 특성에 주로 의존하므로 판정 오류를 줄이는데 한계가 있다. 제안 알고리즘은 화소 기반으로 배경 모델을 일차적으로 수행한 다음 MRF를 이용하여 시공간적으로 인근한 화소 간의 상호 의존성을 활용하여 배경모델의 정확도를 향상시키는데 그 목적을 두고 있다. MRF는 기본적으로 파라미터의 크기에 매우 민감하므로 기존의 MRF 기반 알고리즘은 이미지에 따라 적절한 값을 사전에 구하여 적용하고 있다. 제안한 방식은 초기에 임의의 파라미터로 배경/전경 상태변수를 구한 후에 이의 통계적 특성을 이용하여 파라미터들을 추정하고 추정된 파라미터를 적용하여 상태변수를 재차 구하는 과정을 반복함으로써 최적의 파라미터에 적응적으로 수렴하도록 조정한다. 실내외의 다양한 환경에서 촬영한 비디오를 이용하여 제안한 방식 성능을 확인한다.

Abstract

It is challenging to detect foreground objects when background includes an illumination variation, shadow or structural variation due to its motion. Basically pixel-based background models including codebook-based modeling suffer from statistical randomness of each pixel. This paper proposes an algorithm that incorporates Markov random field model into pixel-based background modeling to achieve more accurate foreground detection. Under the assumptions the distance between the pixel on the input image and the corresponding background model and the difference between the scene estimates of the spatio-temporally neighboring pixels are exponentially distributed, a recursive approach for estimating the MRF regularizing parameters is proposed. The proposed method alternates between estimating the parameters with the intermediate foreground detection and estimating the foreground detection with the estimated parameters, after computing it with random initial parameters. Extensive experiment is conducted with several videos recorded both indoors and outdoors to compare the proposed method with the standard codebook-based algorithm.

Keywords : Visual surveillance, Background subtraction, Pixel-based background modeling, Markov random fields

I. 서 론

아직까지는 특정 위치를 감시하기 위하여 DVR 등으로 촬영하여 보관하거나 감시자가 모니터로 현장을 확인하는 방식을 주로 채택하고 있다. 이러한 방식은 방대한 양의 비디오 저장이 필요하고 감시자의 노동력이 요구된다. 하지만, 현장에서 발생하는 이벤트를 자동으

* 평생회원, ** 학생회원, 한국외국어대학교 정보통신공학과

(Department of Information and Communications Engineering, Hankuk University of Foreign Studies)

※ 이 논문은 2010년도 한국외국어대학교 교내 학술연구 지원에 의하여 연구되었음

접수일자: 2010년11월16일, 수정완료일: 2011년1월7일

로 인식하여 특히 행위를 검출할 때에만 촬영한 비디오 신호를 전송하거나 저장하는 동시에 감시자에게 알람을 제공한다면 이와 같은 비용과 불편을 크게 줄일 수 있을 것이다. 이를 위해서는 기본적으로 비디오 신호에서 움직임 물체를 검출하고 이를 추적하는 기능이 구현되어야 한다. 움직임 물체를 검출하고 어떤 형태인지를 알아내려면 배경 이미지를 정확히 모델링하는 것이 선행되어야 한다. 배경 정보를 정확히 알 수 있으면 비디오의 매 프레임과 기준 배경 모델을 비교하여 각 화소가 배경 모델과 일치하면 제거하고 그렇지 않으면 움직임 물체로 판정함으로써 전경을 검출할 수 있다.

카메라를 고정시켜 특정 장면을 촬영하더라도 카메라 잠움뿐만 아니라 강이나 호수의 물결, 바람에 흔들리는 나뭇가지나 조명의 변화 등은 배경 정보의 정확한 모델링을 어렵게 한다. 배경 모델링은 기본적으로 화소 단위로 수행하는데, 가장 간단한 배경 모델링은 배경만 포함된 프레임들의 각 화소 별 평균, 표준편차 등의 통계적 특성을 이용하여 각 화소에 대응되는 배경정보를 규정하는 방식이다. 이 방식은 각 입력 프레임과 배경 모델의 차를 구함으로써 쉽게 움직임 물체를 구할 수는 있으나 정확한 검출은 기대하기 어렵다. 위에서 언급한 바와 같이, 예를 들어 흔들리는 나무 가지가 있는 장면에서는 해당 화소 위치에서 나뭇가지와 그 뒤의 배경정보가 시간에 따라 번갈아 나타날 수 있어서 각 화소 값의 통계적 특성은 한 개의 확률분포 함수로는 모델링하는데 큰 한계가 있다. 이러한 문제를 해결하기 위하여 Stauffer *et al.*^[1]은 여러 개의 가우시안 확률분포 함수(mixture of Gaussian)로 각 화소의 통계적 특성을 구하고자 하였다. 하지만, 이 방식은 화소 별로 몇 개의 함수를 이용해야 하는지 정해야 하고, 각 함수 별로 평균과 표준편차 등의 파라미터들을 구하는데 있어서 어려움이 내재되어 있다. 게다가, 배경이 빨리 변하면 이에 제대로 적용하지 못하는 문제를 보이고 있다. Elgammal *et al.*^[2]은 파라미터들을 별도로 구하지 않고도 각 화소 별 확률분포 함수를 구하는 방식을 제안하였다. 이 방식은 배경정보만을 포함하는 여러 장의 프레임을 미리 저장해 놓아야 하며 이로부터 가우시안 커널을 이용하여 각 화소에 해당하는 확률분포 함수를 구하고 있다. 하지만, 확률분포함수를 모델링하는데 이용되는 프레임에는 배경정보만 포함되어 있는, 주로 초기의 여러 프레임을 이용하게 되며 시간에 따라 모델을 적용시키기 위해서는 중간에 움직임 물체가 없는 프레

임 밖에 활용할 수 없으므로 적용속도가 매우 느린 문제가 있다. 게다가, 확률분포 모델링을 위하여 여러 프레임을 항상 저장하고 있어야 하고 그 계산량이 엄청나게 요구되므로 실시간 구현이 사실상 매우 어려운 문제를 안고 있다. Buttler *et al.*^[3]는 각 화소를 다수의 클러스터로 모델링하는데, 여기서 각 클러스터는 중심점(centroid)과 비중계수로 표현된다. 입력 화소가 주어지면 각 클러스터의 중심점과 비교하여 가장 일치하는 클러스터를 찾고 그 클러스터의 비중계수보다 큰 계수들을 합한 다음 임계값으로 이진화하여 판정한다. 여기서, 일치하는 클러스터가 없으면 입력화소를 중심점으로 하는 새 클러스터를 추가한다. 하지만, 움직임 물체가 어떠한 형태로 얼마나 자주 출현할지에 대한 사전정보를 구하기 어려우므로 임계값에 따라 성능이 크게 좌우된다는 한계가 있다. Kim *et al.*^[4]은 벡터 양자화 기법을 원용하여 각 화소의 통계적 특성을 여러 개의 코드워드로 표현하는 방식을 제안하였다. 이 방식은 각 화소를 다수의 클러스터로 모델링한다는 점에서 [3]의 방식과 개념적으로 유사한데, 움직임 패턴이 있는 배경이나 조명의 변화에 매우 강한 특성을 보이고 있다. [3]과 [4]의 방식은 모두 배경모델을 유지하기 위한 여러 프레임의 이미지를 저장할 필요가 없어서 적은 저장 공간과 계산량으로 구현할 수 있는 장점이 있다. 게다가 움직임 물체가 존재하는 프레임도 배경 모델을 적용시키는데 이용할 수 있어서 적응력이 매우 우수한 것으로 알려져 있다. Migdal *et al.*^[5]는 Gibbs 분포를 이용하여 에너지를 정의하고 이로부터 MRF를 모델링한 다음 [1]의 mixture of Gaussian와 결합하여 전경을 검출하는 알고리즘을 제안하였다. Wu *et al.*^[6]는 [4]의 코드북 기반 배경모델 알고리즘과 MRF를 결합하고 Metropolis dynamics^[7]를 이용하여 구현하였다. McHugh *et al.*^[12]는 [2]의 방식을 채택하여 배경과 전경에서의 확률분포를 구하고 이들의 비를 쓰레숄드함으로써 배경/전경 분류를 수행하였다. 이 방식은 [2]와는 달리 전경 확률분포를 균일 분포로 가정하지 않고 별도로 구하여, 결과적으로 확률분포의 비에 대한 문턱값을 적용적으로 조정함으로써 성능향상의 효과를 얻고 있다. 이에 더하여 Gibbs 분포를 이용한 MRF 모델을 적용하여 배경/전경 판정결과를 반복적으로 개선하는 방식을 제안하였다. 하지만 [5]와 [6]을 포함한 MRF 기반 방식의 문제점은 내부에 적용된 여러 파라미터들을 실험적으로 구한 상수로 고정하여 이용하기 때문에 입력 이미지의 통계적

특성에 따라 적절한 값을 미리 구하여야 한다는 것이다.

이러한 한계를 극복하기 위하여 본 논문에서는 MRF의 내부 파라미터를 입력 이미지와 출력 이미지의 시공간적인 확률특성에 따라 적응적으로 파라미터를 유도하는 알고리즘을 제안한다. 본 논문은 기본적으로 [4]의 방식을 기반으로 배경정보를 일차적으로 모델링한 다음 MRF를 이용하여 시공간적으로 인근한 화소 간의 상호 의존성을 활용하여 배경 모델의 정확도를 향상시키는 알고리즘을 제안하는데 그 목적이 있다. MRF는 기본적으로 파라미터의 크기에 매우 민감하므로 기존의 MRF 기반 알고리즘은 이미지에 따라 적절한 값을 사전에 구하여 적용하고 있다. 제안한 방식은 초기에 임의의 파라미터로 배경/전경 상태변수를 구한 후에 이의 통계적 특성을 이용하여 파라미터들을 추정하고 추정된 파라미터를 적용하여 상태변수를 재차 구하는 과정을 반복함으로써 최적의 파라미터에 적응적으로 수렴하도록 조정한다. 따라서, 이미지에 따라 파라미터들을 사전에 구하여야 하는 문제를 해결하는 특징이 있다.

본 논문의 구성은 다음과 같다. 2장에서는 코드북 기반의 배경 모델링 결과를 바탕으로 MRF를 이용하여 전경을 검출하는 알고리즘과 함께 내부의 여러 MRF 파라미터를 구하는 방식을 설명하고 본 논문에서 제안하는 알고리즘에 대해 기술한다. 제안 알고리즘을 적용하였을 때의 실험결과와 이에 대한 성능분석을 3장에서 기술한다. 마지막으로 4장에서는 결론을 맺고 향후 연구 진행방향에 대하여 논의한다.

II. MRF를 이용한 전경 검출

시간 t 에서 이미지 y_i^t 가 주어졌을 때 그에 해당하는 장면에서의 상태변수 x_i^t 를 추론하는 문제를 상정한다. 여기서, 인덱스 i 는 화소의 위치를 나타낸다. 랜덤 변수 (random variable) X 를 정의하고 시간 t 에서 i 번째 원소 $X_i^t \in X$ 는 상태정보 $x_i^t \in A$ 를 갖는다고 가정한다. 여기서 $A = \{0, 1\}$ 는 배경($x_i^t = 0$) 또는 전경($x_i^t = 1$)을 의미한다. 각 화소에 대하여 x_i^t 와 y_i^t 사이에는 통계적인 연관성이 있는데, 이를 $\Phi_i(x_i^t, y_i^t)$ 라고 표기한다. $\Phi_i(x_i^t, y_i^t)$ 는 화소 y_i^t 가 배경모델에 속하면서 $x_i^t = 0$ 이면 큰 값을 갖지만, $x_i^t = 1$ 이면 매우 작아야 한다. 반

대로 배경모델에 속하지 않으면 $x_i^t = 0$ 일 때 작은 값을 갖는다. 랜덤 변수 ξ_i 가 각 화소 y_i^t 에 할당되는데, 이는 배경 모델에 속하는지를 나타낸다. 즉 y_i^t 가 배경 모델에 속하면 ξ_i 는 0를, 그렇지 않으면 1을 각각 갖는다. 본 논문에서 $\Phi_i(x_i^t, y_i^t)$ 는 다음과 같이 배경에서는 지수분포를, 전경에서는 균일분포를 각각 갖는다고 가정한다.

$$\Phi_i(x_i^t, y_i^t) = \begin{cases} \zeta e^{-\mu d(y_i^t)} & \xi_i = 0 \\ \frac{1}{M} & \xi_i = 1 \end{cases} \quad (1)$$

여기서 μ 는 지수 분포의 감소율을 나타내고 ζ 는 정규화 상수이다. $d(y_i^t)$ 는 화소 y_i^t 가 배경 모델에 속할 때 그에 해당하는 클러스터 중심과의 최소거리를 나타낸다. 그리고 전경에서 $d(y_i^t)$ 는 일반적으로 큰 값을 갖고 있으나 화소의 위치에 따라 변화 폭이 매우 크므로 균일분포를 갖는다고 가정한다. y_i^t 가 배경 모델에 속할 확률 $p(\xi_i = 0)$ 를 λ 라고 정의하면 식 (1)은 다음과 같이 표현될 수 있다.

$$\Phi_i(x_i^t, y_i^t) = \lambda \zeta e^{-\mu d(y_i^t)} + \frac{1-\lambda}{M} \quad (2)$$

위 식에서 정규화 상수 ζ , λ 와 뒤에서 설명할 파라미터 μ 가 결정되면 M 은 다음과 같이 구할 수 있다.

$$M = \frac{\sum_{y_i^t \in f} 1}{1 - \lambda \zeta \sum_{y_i^t \in b} e^{-\mu d(y_i^t)}} \quad (3)$$

여기서 b 와 f 는 각각 배경과 전경 집합을 나타낸다.

또한, 각 화소 위치 i 에서 추론하고자 하는 값인 x_i^t 는 그 인근 화소의 x_j^t 와 상호연관성이 매우 높는데, 이를 $\Psi_{ij}(x_i^t, x_j^t)$ 라고 정의한다. $\Psi_{ij}(x_i^t, x_j^t)$ 는 x_i^t 와 x_j^t 사이의 연관성을 나타내는 확률에 비례하므로 이들이 동일하면 큰 값을 갖는다. 따라서 $\Psi_{ij}(x_i^t, x_j^t)$ 는 다음과 같이 나타낼 수 있다.

$$\Psi_{ij}(x_i^t, x_j^t) = \eta e^{-\nu |x_i^t - x_j^t|} \quad (4)$$

여기서 ν 와 η 는 각각 지수 분포의 감소율, 정규화 상수를 나타낸다. 위 식에서 $|x_i^t - x_j^t|$ 는 0 또는 1의 경우

밖에 없으므로 위와 같이 지수 분포를 가정해도 무방하다. 이와 함께 x_i^t 와 바로 전 시간에서 동일한 화소위치의 x_j^{t-1} 의 연관성 $\Theta_{ij}(x_i^t, x_j^{t-1})$ 를 다음과 같이 정의한다.

$$\Theta_{ij}(x_i^t, x_j^{t-1}) = \delta e^{-\sigma|x_i^t - x_j^{t-1}|} \quad (5)$$

위 식 (2), (4)와 마찬가지로 δ 와 σ 는 각각 지수 분포의 감소율, 정규화 상수를 나타낸다. 전경의 움직임이 급격하지 않다고 가정하면 $\Theta_{ij}(x_i^t, x_j^{t-1})$ 도 x_i^t 와 x_j^{t-1} 가 동일한 값을 가질 때 그 값이 커지는 특성이 있다. 따라서 장면 x_i^t, x_j^{t-1} 와 이미지 y_i^t 들이 인근 화소에 대해서만 서로 관련이 있다고 가정하면 이들의 확률분포 함수(joint probability density function)는 다음과 같이 정리할 수 있다.

$$P(\{x^{t-1}\}, \{x^t\}, \{y^t\}) \quad (6)$$

$$= \frac{1}{Z} \prod_{(i,j)} \Psi_{ij}(x_i^t, x_j^t) \prod_i \Phi_i(x_i^t, y_i^t) \prod_{(i,j)} \Theta_{ij}(x_i^t, x_j^{t-1})$$

위 식에서 Z 는 정규화 상수이고, (i, j) 는 이미지 격자에서 4개의 인근 화소를 나타낸다. 이 랜덤 변수들의 상호 관련성을 인근 화소로 제한을 두는 가정 즉, 마르코프 과정(Markov process)이라고 가정함으로써 장면 변수 x_i^t 에 대한 추정을 쉽게 유도할 수 있다.

식 (2), (4), (5)를 식 (6)에 대입하면 다음과 같이 정리할 수 있다.

$$P(\{x^{t-1}\}, \{x^t\}, \{y^t\}) \quad (7)$$

$$= \frac{1}{Z} \prod_i \left\{ \lambda \zeta e^{-\mu d(y_i^t)} + \frac{1-\lambda}{M} \right\} \prod_{(i,j)} \eta e^{-\nu|x_i^t - x_j^t|} \prod_{(i,j)} \delta e^{-\sigma|x_i^t - x_j^{t-1}|}$$

계산을 용이하게 하기 위하여 양변에 $-\log$ 를 취하여 다음과 같이 정리한 다음

$$E = -\ln P(\{x^{t-1}\}, \{x^t\}, \{y^t\}) \quad (8)$$

$$= \ln Z - \sum_i \ln \left\{ \lambda \zeta e^{-\mu d(y_i^t)} + \frac{1-\lambda}{M} \right\}$$

$$+ \sum_{(i,j)} (\nu|x_i^t - x_j^t| - \ln \eta) + \sum_{(i,j)} (\sigma|x_i^t - x_j^{t-1}| - \ln \delta)$$

이 식에서 상수항을 제거하면 다음과 같이 줄일 수 있다.

$$E \propto - \sum_i \ln \left\{ \lambda \zeta e^{-\mu d(y_i^t)} + \frac{1-\lambda}{M} \right\} \quad (9)$$

$$+ \sum_{(i,j)} (\nu|x_i^t - x_j^t| + \sigma|x_i^t - x_j^{t-1}|)$$

$$\text{그런데 } \alpha = -\ln(a+c), \quad \beta = \frac{ab}{a+c},$$

$$\gamma = \ln \left(\frac{a+c}{c} \right) \quad \text{이면} \quad \min(\beta|x|, \gamma) + \alpha = -\ln(ae^{-\beta|x|} + c) \text{의 upper bound이다. 따라서 식 (9)를 최소화하는 것은 다음 식으로 나타낼 수 있다.}$$

$$\min E \propto \sum_i \min(\kappa|d(y_i^t)|, \theta) \quad (10)$$

$$+ \min \left(\nu \sum_{(i,j)} |x_i^t - x_j^t| + \sigma \sum_{(i,j)} |x_i^t - x_j^{t-1}| \right)$$

$$\text{여기서 } \kappa = \frac{\lambda \zeta \mu}{\lambda \zeta + \frac{1-\lambda}{M}} \text{ 이고 } \theta = \ln \left(1 + \frac{\lambda \zeta M}{1-\lambda} \right) \text{이다.}$$

결국 위식에 나타낸 바와 같이 E 를 최소화하도록 κ, ν, σ 등의 파라미터를 업데이트하고 이에 따라 배경/전경 분류를 수행함으로써 성능을 향상시킬 수 있다.

다음에서는 이 파라미터들을 추정하는 방식을 설명하고 이에 기반한 배경/전경 분류 방법을 기술한다.

1. 파라미터 ν 와 σ 의 추정

우선 코드북 기반 배경 모델링 알고리즘을 구동하여 구한 배경/전경에 대한 중간 분류 결과를 이용함으로써 파라미터 ν 와 σ 를 다음과 같이 초기화한다. 즉, 시간 t 에서 x_i^t 과 x_j^t 가 동일할 확률과 서로 다를 확률의 비를 다음과 같이 계산함으로써 식 (4)에 주어진 파라미터 ν 를 구할 수 있다.

$$\nu = -\log \left\{ \frac{\sum_{(i,j)} \text{Prob}(x_i^t \neq x_j^t)}{\sum_{(i,j)} \text{Prob}(x_i^t = x_j^t)} \right\} \quad (11)$$

동일한 방법으로 파라미터 σ 는 다음과 같이 구할 수 있다.

$$\sigma = -\log \left\{ \frac{\sum_{(i,j)} \text{Prob}(x_i^t \neq x_j^{t-1})}{\sum_{(i,j)} \text{Prob}(x_i^t = x_j^{t-1})} \right\} \quad (12)$$

2. 파라미터 λ 와 μ 의 추정

우선 배경 모델에 속한 화소의 거리 정보 $d(y_i^t)$ 의 가능한 개수 L 을 다음과 같이 구한다.

$$L = \max\{|d(y_i^t)|\} + 1 \tag{13}$$

식 (1)에 나타낸 랜덤 변수 ξ_i 의 조건부 확률은 다음과 같이 구할 수 있다.

$$\begin{aligned} \rho_i &\equiv P(\xi_i = 0/d(y_i^t), \lambda, \mu) \\ &= \frac{\lambda \zeta e^{-\mu|d(y_i^t)|}}{\lambda \zeta e^{-\mu|d(y_i^t)|} + \frac{1-\lambda}{M}} \end{aligned} \tag{14}$$

본 논문에서는 EM(expectation maximization) 알고리즘을 적용하여 기대값 $E_{\xi_i}[\log P(d(y_i^t), \xi_i/\lambda, \mu)]$ 를 최대화시키는 파라미터들을 구함으로써 λ 와 μ 를 추정한다^[11]. 여기서 $P(d(y_i^t), \xi_i/\lambda, \mu)$ 는 다음과 같이 주어진다.

$$\begin{aligned} P(d(y_i^t), \xi_i = 0/\lambda, \mu) &= \lambda \zeta e^{-\mu|d(y_i^t)|} \\ P(d(y_i^t), \xi_i = 1/\lambda, \mu) &= \frac{(1-\lambda)}{M} \end{aligned} \tag{15}$$

위 식을 이용하면 $E_{\xi_i}[\log P(d(y_i^t), \xi_i/\lambda, \mu)]$ 는 다음과 같이 정리될 수 있다.

$$\begin{aligned} E_{\xi_i}[\log P(d(y_i^t), \xi_i/\lambda, \mu)] & \\ &= \sum_{y_i^t \in b} \rho_i \log P(d(y_i^t), \xi_i = 0/\lambda, \mu) \\ &\quad + \sum_{y_i^t \in f} (1-\rho_i) \log P(d(y_i^t), \xi_i = 1/\lambda, \mu) \\ &= \sum_{y_i^t \in b} \rho_i (\log(\lambda \zeta) - \mu|d(y_i^t)|) + \sum_{y_i^t \in f} (1-\rho_i) \log \frac{1-\lambda}{M} \end{aligned} \tag{16}$$

위 식을 λ 와 μ 에 대하여 편미분한 다음 이를 0으로 놓고 풀면 λ 는 다음과 같이 주어지는데,

$$\lambda = \frac{\sum_{y_i^t \in b} \rho_i}{\sum_{y_i^t \in b} \rho_i + \sum_{y_i^t \in f} (1-\rho_i)} \tag{17}$$

구현 과정에서는 코드북 기반 배경 모델링 알고리즘 결과를 이용하여 전체 화소에서 배경으로 판정된 화소의 비를 λ 의 초기값으로 정하고 식 (14)를 이용하여 ρ_i

를 구한 다음 식 (17)을 통하여 λ 를 추정한다. 그리고 μ 는 다음 식의 근이다.

$$\frac{1}{e^\mu - 1} - \frac{L}{e^{\mu L} - 1} = \frac{\sum_{y_i^t \in b} \rho_i |d(y_i^t)|}{\sum_{y_i^t \in b} \rho_i} \tag{18}$$

그런데 실험에 의하면 L 은 30 이상의 값을 가지므로 위 식에서 좌변의 두 번째 항인 $\frac{L}{e^{\mu L} - 1}$ 는 무시할 수 있다. 따라서 μ 는 다음과 같이 구할 수 있다.

$$\mu = \log\left(1 + \frac{1}{\chi}\right) \tag{19}$$

여기서 χ 는 식 (18)의 우변을 나타낸다.

이와 같이 추정된 파라미터 ν , σ , λ 와 μ 등을 이용하여 MRF를 구동함으로써 배경/전경 분류를 수행한다. 새롭게 구한 배경/전경 분류결과를 이용하여 위와 같이 파라미터들을 재차 추정하는 과정을 반복함으로써 최적의 파라미터에 적응적으로 수렴하도록 조정한다.

3. Belief propagation을 이용한 전경 검출

식 (10)에 주어진 MRF를 MAP(maximum a posteriori)으로 해결하기 위하여 belief propagation BP^[9-10]을 채택하였다. BP는 전 노드에서 구한 메시지를 병렬적으로 전송하여 정보공유를 반복적으로 수행한다. 시간 t 에서 노드 i 에서 노드 j 로 전송하는 메시지를 m_{ij}^t 라고 나타내면 이는 다음과 같이 업데이트된다.

$$\begin{aligned} m_{ij}^t(x_j^t) &= \min_{x_i^t} \sum_i (\min(\kappa|d(y_i^t)|, \theta) \\ &\quad + \nu|x_j^t - x_i^t| + \sigma|x_j^t - x_i^{t-1}| \\ &\quad + \sum_{k \in N(i)/j} m_{ki}(x_i^t)) \end{aligned} \tag{20}$$

여기서 $N(i)$ 는 노드 i 의 인근 노드(4-neighbors)를, $N(i)/j$ 는 노드 j 를 제외한 $N(i)$ 를 각각 나타낸다.

업데이트된 메시지를 이용하여 노드 i 에서의 신뢰도 (belief) $b_i(x_i^t)$ 는 다음과 같이 구할 수 있다.

$$b_i(x_i^t) = \min(\kappa|d(y_i^t)|, \theta) + \sum_{k \in N(i)} m_{ki}(x_i^t) \tag{21}$$

신뢰도 $b_i(x_i^t)$ 는 $-\log(P(x_i^t))$ 에 해당되므로 상태변수 x_i^t 는 다음과 같이 구한다.

$$x_i^t = \begin{cases} 1 & b_i(0) > b_i(1) \\ 0 & b_i(0) \leq b_i(1) \end{cases} \quad (22)$$

4. 제안 알고리즘

식 (1)의 $\Phi_i(x_i^t, y_i^t)$ 를 구하려면 우선 화소 y_i^t 가 배경 모델에 속하는 정도 또는 확률을 추정하여야 하는데, 이를 위해서는 배경모델을 설정하여야 한다. 카메라를 고정하여 배경을 촬영하더라도 바람에 움직이는 나무, 물결, 흔들리는 커튼 등, 다양한 원인으로 인하여 배경 이미지도 시간에 따라 동적으로 변할 수 있다. 이런 경우에 각 화소마다 그 값이 그룹 지어져 덩어리(blob) 형태로 나타난다. 위에서도 언급한 바와 같이 많은 연구자들이 이런 덩어리를 다양한 방법으로 모델링하는 방식을 제안하고 있다. 예를 들어, 덩어리를 평균과 공분산을 파라미터로 갖는 가우시안 분포로 모델링할 수 있다^[2]. 또는 덩어리가 여러 개 나타날 수 있음을 고려하여 그 개수에 맞게 다수의 가우시안 분포(mixed of Gaussian distribution)로 모델링하든지^[1] 커널 함수를 이용하여 파라미터 없이 직접 확률분포를 모델링하기도 한다^[2]. 본 논문에서는 적용력이 우수하면서도 적은 계산량으로 구현이 가능한 코드북 기반 배경모델링 기법을 채택하였다. 이 방식은 벡터 양자화 기법에서와 같이 각 덩어리를 코드워드로 표현한다. 즉, 각 화소마다 코드북을 할당하고 코드북 내에 덩어리 수에 해당하는 코드워드를 생성한다^[4]. 따라서 코드북 내의 코드워드 수는 화소마다 다를 수 있다.

충분한 길이의 학습 이미지 프레임으로부터 각 화소에 대한 코드북을 구성한 다음, 각 입력화소 y_i^t 가 이에 해당하는 코드북 C_i 의 코드워드 $\{c_1, c_2, \dots, c_N\}$ 내에 속하는지를 판단한다. 여기서 화소의 거리정보 $d(y_i^t)$ 는 화소 y_i^t 와 가장 근거리에서 있는 코드워드의 중심 화소 간의 유클리디안 거리를 계산함으로써 얻는다.

이와 같이 코드북 기반 알고리즘을 바탕으로 화소 단위의 배경모델을 구한 다음 BP를 이용하여 시공간 상의 인근 화소에서의 정보 교류를 활성화함으로써 배경 모델의 정확도를 향상시킨다.

III. 실험 결과

제안 알고리즘의 동작을 확인하고 성능을 분석, 평가

하기 위하여 본 논문에서는 실내와 실외에서 촬영한 비디오를 이용하였다. 실내 비디오는 배경 이미지가 비교적 정적인 특성을 갖지만 조명의 변화와 함께 백색 균형이 깨지는 현상을 보여 주고 있다. 이에 반하여, 실외 비디오는 바람에 흔들리는 나뭇가지를 배경으로 갖고 있어서 제안 알고리즘이 다양한 동적 배경 특성 하에서 움직임 물체를 제대로 검출할 수 있는지를 파악할 수 있도록 하였다. 실험에 이용한 비디오는 모두 320×240 의 해상도를 갖고 있다. 본 실험은 기존의 화소기반 전경 검출 알고리즘에 MRF를 결합하였을 때의 성능향상을 측정하는데 초점을 두고 있어서 검출 결과에 대한 후처리 작업은 수행하지 않았다. 본 논문에서는 화소기반 전경검출 알고리즘으로 [4]의 코드워드 기반 알고리즘을 채택하였다.

1. 정성적 평가

그림 1은 실내에서 촬영한 비디오를 이용하여 구한 실험결과를 보여 주고 있는데 카메라 내부 잡음과 함께 전경의 움직임에 따른 그림자의 변화에 대응하는 제안 알고리즘의 검출능력을 나타내고 있다. 이 비디오 프레임은 카메라와 전경의 거리에 따라 조명이 변하는 특성과 함께 백색 균형(white balance)이 깨지는 현상을 보이고 있다. 즉, 전경이 카메라에 근접하면 좌중간 벽의 색상이 좀 더 푸른 빛을 띠는 현상을 보이고 있다. 그림 1의 첫째 열은 서로 다른 프레임의 입력 이미지를, 둘

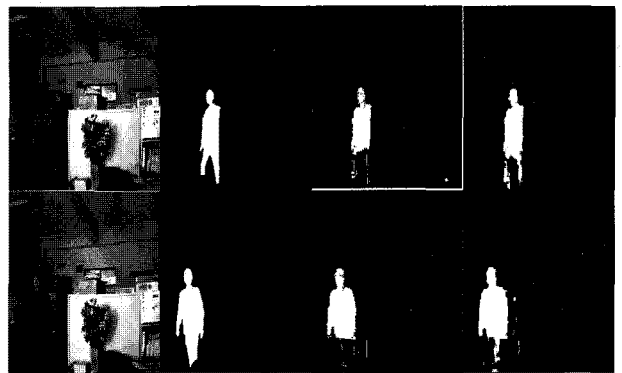


그림 1. 실내 비디오에 대한 실험결과. 1열: 입력 이미지 2열: ground_truth 영상 3열: 코드북 기반 알고리즘의 검출 결과 4열: 제안 알고리즘의 검출 결과

Fig. 1. The comparative experimental results on the video recorded indoors. Column 1: original images. Column 2: ground-truths. Column 3: detection results of codebook-based algorithm. Column 4: detection results of our method.

제 열은 수작업으로 전경물체를 판정한 ground_truth 이미지를 각각 나타내고 있다. 이와 함께 셋째 열과 넷째 열은 코드워드 기반 알고리즘과 제안 알고리즘의 전경 검출 결과를 각각 보여 주고 있다. 이 그림에서 알 수 있듯이 전경의 상체 부분은 배경과 색상 차가 매우 커 쉽게 검출할 수 있으나 하체는 검은 색을 띠고 있는데 그 주변의 배경도 매우 어두워 육안으로도 분별하기 어려울 정도로 전경과 배경 간에 밝기와 색상차가 매우 작다. 이러한 환경에서 코드북 기반 알고리즘은 하체를 제대로 분별하지 못하지만 제안 알고리즘은 비교적 선명하게 검출하는 특성을 보여주고 있다.

그림 2는 실외에서 촬영한 비디오에 대한 코드북 기반 알고리즘과 제안 알고리즘의 전경 검출 성능을 보여주고 있다. 입력 이미지의 중간 부분을 보면 낮은 언덕의 잔디 색과 전경의 상의 색상이 거의 유사함을 알 수 있다. 기본적으로 화소 기반 알고리즘은 배경과 전경의 화소가 유사한 값을 가지면 이를 구별하기가 매우 어렵다.

그림 2의 3열에서 볼 수 있듯이 코드북 기반 알고리즘은 전경이 이와 유사한 색상을 갖는 배경 앞을 지날 때 전경을 배경으로 판정하는 오류(false negatives)를 범하고 있다. 그 결과로 검출된 전경의 가슴 부위에 띠 형태의 오류를 확인할 수 있다. 즉 전경의 상의와 배경의 누런 잔디를 구별하지 못하여 그 부분에 오류가 발생하고 있다. 반면에 제안 알고리즘은 판정대상 화소와

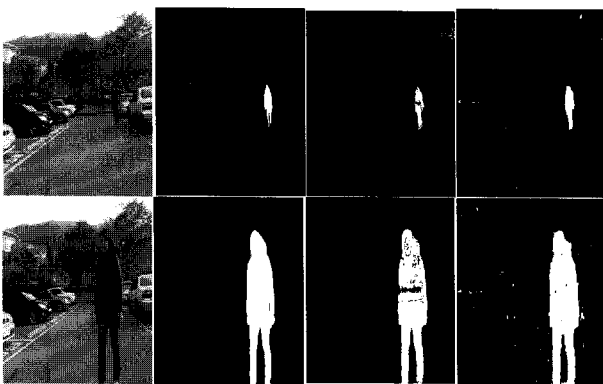


그림 2. 실외 비디오에 대한 실험결과. 1열: 입력 이미지 2열: ground_truth 영상 3열: 코드워드 기반 알고리즘의 검출 결과 4열: 제안 알고리즘의 검출 결과

Fig. 2. The comparative experimental results on the video recorded outdoors. Column 1: original images. Column 2: ground-truths. Column 3: detection results of codebook-based algorithm. Column 4: detection results of our method.

그 주변 화소 간의 상호연관성을 파악하여 정보를 교류하기 때문에 일부분에서 전경과 배경이 유사한 값을 갖더라도 그 주변 정보를 종합하여 이를 보정하는 능력이 있음을 그림 2의 4열을 통해 확인할 수 있다. 다만 배경으로 판정된 일부 영역에서 그 주변과의 정보교류를 통하여 오히려 전경으로 검출함으로써 블롭(blob) 형태의 오류를 보이는 문제점을 노출하고 있다.

2. 정량적 평가

위의 두 비디오를 이용하여 제안 알고리즘의 성능을 정량적으로 평가하였다. 실내에서 촬영한 비디오에서는 앞 부분의 320개 프레임을 이용하여 코드워드를 훈련시키고 그 뒤로 4 프레임마다 수작업으로 ground_truth를 생성한 다음 이를 이용하여 [8]이 제안한 유사도(similarity)를 측정하였다. 검출 결과 L_t 와 이에 해당되는 ground_truth G_t 의 유사도는 다음과 같이 정의된다.

$$S(L_t, G_t) = \frac{|L_t \cap G_t|}{|L_t \cup G_t|} \tag{23}$$

검출 결과 L_t 가 G_t 와 일치할수록 유사도 $S(L_t, G_t)$ 는 1에 수렴하고 반대로 관련이 없으면 0에 수렴한다.

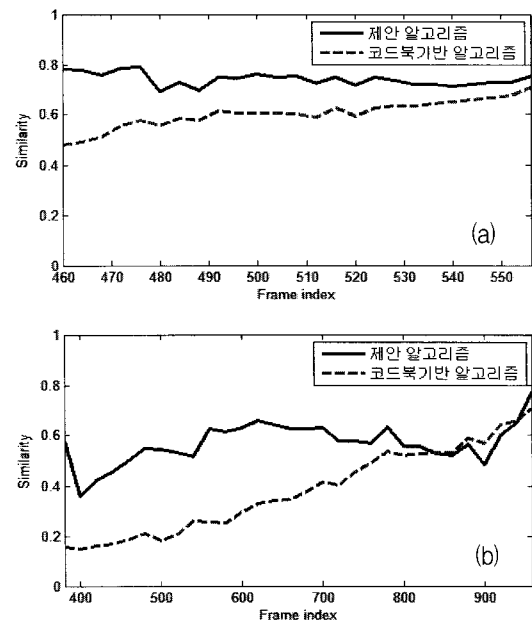


그림 3. 코드북 기반 알고리즘과 제안 알고리즘의 유사도 측정결과 (a) 실내 비디오 (b) 실외 비디오

Fig. 3. The similarity curves for codebook-based algorithm and our method on the video recorded (a) indoors and (b) outdoors.

그림 3은 실내 비디오와 실외 비디오에 대한 코드북 기반 알고리즘과 제안 알고리즘의 유사도 측정결과를 보여 주고 있다. 그림 3에서는 모든 프레임에서 제안 알고리즘이 코드북 기반 방식보다 우수한 성능을 보여 주고 있다. 하지만 실외비디오에서는 제안 알고리즘이 그림자의 영향을 더 많이 받음으로 인하여 전경이 카메라에 근접하는 후반부에서 성능이 다소 저하되는 현상을 보이고 있다. 이 그림을 통해서 알 수 있듯이 유사도 측정 결과는 그림 1과 그림 2의 정성적 결과와 매우 일치한다. 즉 제안 알고리즘의 유사도는 거의 모든 프레임에서 코드북 기반 알고리즘보다 높은 값을 유지하고 있음을 알 수 있다.

IV. 결 론

본 논문에서는 화소 기반 배경 모델링 기법과 MRF를 결합하여 전경검출의 성능을 향상시키는 방안을 제안하였다. 지금까지 MRF는 그 내부의 파라미터들을 실험적으로 구한 값으로 정하였기 때문에 이미지에 따라 별도로 미리 구하여야 하는 한계를 지니고 있다. 하지만 본 논문에서는 입력 이미지와 전경 검출 결과의 시공간적 특성에 따라 이들에 대한 확률모델을 제시하고 이에 따라 MRF의 내부 파라미터들을 구하는 기법을 제안하였다. 제안 알고리즘의 성능을 확인하기 위하여 본 논문에서는 여러 입력 비디오 프레임의 검출결과를 비교해 보고 수작업으로 ground truth를 생성한 다음 ground truth와 검출결과에 대한 유사도를 측정하여 그 결과를 비교하였다. 제안 알고리즘은 배경과 전경의 화소 값이 유사하여 전경을 배경으로 인식하는 오류는 크게 줄일 수 있음을 확인하였으나 카메라 잡음이나 그림자 주변에서 화소 간의 정보교류를 통해 오히려 배경을 전경으로 오인식하는 문제점도 노출하였다. 이러한 문제점을 해결하기 위한 연구가 앞으로 수행해야 할 시급한 과제이다.

참 고 문 헌

- [1] C. Stauffer, W.E.L. Grimson, "Adaptive background mixture models for real-time tracking," international Journal of Computer Vision, IEEE International Conference on Computer Vision and Pattern Recognition, Vol. 2, pp.246-252, 1999.
- [2] A. Elgammal, R. Duraiswami, D. Harwood, L.S. Davis, "Background and foreground modeling using nonparametric kernel density estimation for visual surveillance," pp. 1151-1163, Proc. IEEE, vol. 90, no. 7, Jul, 2002.
- [3] D. Buttler, S. Sridharan, V.M. Bove Jr., "Real-time adaptive background segmentation," ICASSP'03, pp. 349-352, 2003.
- [4] K. Kim, T.H. Chalidabhongse, D. Harwood, "Real-time foreground-background segmentation using codebook model," Elsevier Real-Time Imaging, vol. 11, pp. 172-185, 2005.
- [5] J. Migdal and W. E. Grimson, "Background subtraction using Markov thresholds," Proceedings of the IEEE Workshop on Motion and Video Computing(WACV/MOTION'05), 2005.
- [6] M. Wu and X. Peng, "Spatio-temporal context for codebook-based dynamic background subtraction," International Journal of Electronics and Communications, pp. 739-747, 2010.
- [7] M. Berthod, Z. Kato, S. Yu, J. Zerubia, "Bayesian image classification using Markov random fields," Image Vision Comput 1996; 14:285-95.
- [8] Y. Chen, C. Chen, C. Huang, Y. Hung, "Efficient hierarchical method for background subtraction," Pattern Recognition, pp. 2706-2715, 2007.
- [9] J. S. Yedidia, W. T. Freeman and Y. Weiss, "Understanding belief propagation and its generalizations," TR-2001-22, Jan. 2002.
- [10] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient belief propagation for early vision," Proc. Conf. Computer Vision and Pattern Recognition, pp.261-268, 2004.
- [11] L. Zhang and S. M. Seitz, "Estimating optimal parameters for MRF stereo from a single image pair," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.29, no.2, pp.331-342, Feb. 2007.
- [12] J. M. McHugh, J. Konrad, V. Saligrama, and P. Jodoin, "Foreground-adaptive background subtraction," IEEE Signal Processing Letters, vol.16, issue 5, pp.390-393, 2009.

— 저 자 소 개 —



한 희 일(평생회원)
 1984년 서울대학교 제어계측
 공학과 학사 졸업.
 1988년 서울대학교 제어계측
 공학과 석사 졸업.
 1995년 University of Arizona
 전기및컴퓨터공학과
 박사 졸업.

2011년 현재 한국외국어대학교 정보통신공학과
 교수.

<주관심분야 : 영상처리, 컴퓨터비전, 패턴인식>



박 수 빈(학생회원)
 2010년 한국외국어대학교 정보
 통신공학과 학사 졸업.
 2010년 한국외국어대학교 정보
 통신공학과 석사 과정.
 <주관심분야 : 신호처리, 영상처
 리>