

영상신호 특성 및 Human 특징을 이용한 실시간 영상 분류

*김민준 **김원하

경희대학교

*i8pq8i@naver.com

Video Segmentation Using Image signal and Human characteristic

*Kim, Min-Joon **Kim, Won-Ha

Kyung Hee University

요약

영상에서 배경으로부터 객체를 분류하는 영상 분류 알고리즘은 물체 인식 및 추적 등 다양한 응용분야에서 중요하다. 본 논문에서는 고정된 카메라에서 다수의 초기 프레임에 참조하여 실시간 영상 분류 방법을 제안한다. 먼저 전경과 배경을 구분하는 확률모델을 제안하였으며 초기 프레임 동안에 카메라의 특성을 추출하여 카메라에 적응적으로 영상을 분류한다. 또한 분류된 영상에서 human의 특징을 이용하여 분류된 결과를 보정하는 방법을 제안한다. 마지막으로 제안한 알고리즘의 실시간 분류 처리를 위하여 복잡도를 최소화 하였다.

1. 서론

영상을 객체 별로 분류하거나 관심 있는 영역으로 나누는 작업을 영상 분류라고 한다. 일반적으로 영상에서 물체를 추출하거나 영상의 전경/배경을 나누는 영상 처리 기법은 여러 응용분야에서 사용되고 있다. 대표적으로 intensity 정보만을 사용하는 영상 분류 방법이 있다. [1],[2]에서는 프레임간 비교를 통해 화소 단위로 Mixture of Gaussian 모델을 생성해서 배경을 분리하며 [8],[9]에서는 각각 Bayesian decision 모델, Kernel density 모델을 이용하여 영상을 분류하는 방법을 제시하였다. 하지만 이러한 방법들은 intensity 정보만을 사용하기 때문에 충분한 성능을 제공하지 못하는 단점이 있다. 이에 알고리즘의 성능을 개선시키기 위해 [3],[4],[5],[6],[7]에서는 texture 정보를 추가적인 feature 정보로 사용하였다. 이를 통해 알고리즘의 성능은 향상되었으나 feature들을 사용함에 따라 알고리즘 복잡도가 높아졌다. [10]에서는 복잡도가 높은 알고리즘에서 하드웨어적인 향상을 구현하였지만 추가적인 장치가 요구된다. 위와 같이 단일 영상에서 소프트웨어적인 연산만으로 물체를 분류하는 방식과 다르게 영상의 깊이 정보를 이용하는 연구도 이루어지고 있다. [11],[12],[13]에서는 깊이 센서나 stereo 영상으로 깊이 정보를 얻음으로써 간단하게 전경과 배경을 분리하는 방법을 제시한다. 그러나 깊이 센서라는 추가적인 하드웨어가 요구되기 때문에 장치 가격은 증가하게 된다. 또한 깊이 센서의 기술적인 한계로 저해상도의 깊이 영상을 제공함으로써 실제 색 영상과의 합성이 불가피하게 되었다.

앞서 소개한 논문과 비교하여 본 논문에서는 추가적인 장치 없이 단일 카메라로 실시간 영상 분류 알고리즘을 제안한다. 제안하는 알고리즘은 다수의 초기 프레임으로 카메라 특성을 분석하여 참조 프레임과 현재 프레임의 intensity, 색 그리고 texture정보 비교를 통해 영상 분류를 수행한다. 또한, 기존의 이진분류 방식과 다르게 전경과 배경을

확률적으로 분류함으로써 이진분류에서 나타나는 under/over 분류 문제를 감소시킨다. 본 논문에서는 실시간 처리를 위해 block 단위로 영상 분류를 수행하며 낮은 복잡도로 색, texture를 분석하는 방법을 제안한다. 마지막으로 human 인식 알고리즘을 통해 영상 분류의 성능을 향상시켰다. 본 논문은 다음과 같이 구성되어 있다. 두 번째 장은 영상 및 카메라 특성을 활용한 영상 분류 방법을 제안한다. 세 번째 장은 human 인식 알고리즘을 활용하여 분류된 결과를 보정하는 방법을 소개한다. 네 번째 장은 예시 영상을 통한 시뮬레이션 결과 비교 및 실제 알고리즘의 복잡도를 제시하고, 다섯 번째 장에서 결과 분석을 통해 결론을 내린다. 그림 1은 전체 흐름도를 나타낸다.

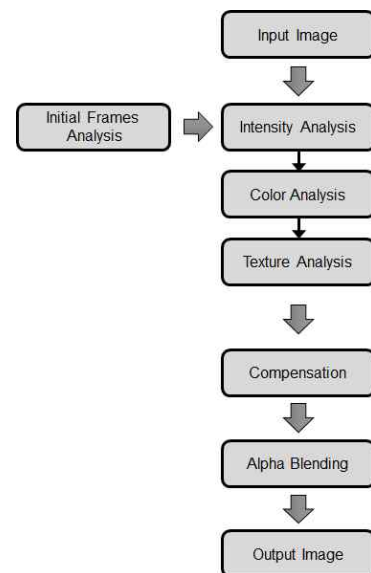


그림 1. 전체 흐름도
Fig 1. Overall Flowchart

2. RGB 분석 기반 Block DCT 이진 분류방법

DCT 변환에 앞서 RGB 크기 차이를 분석한다. 색 영상은 RGB 3개의 채널을 가지기 때문에 3채널 DCT를 수행하여 비교를 하는 것은 비효율적이다. 왜냐하면 3개의 채널 DCT를 한 후 비교를 하는 것은 계산 소요가 많기 때문이다. 기본적으로 프레임 간 RGB요소 중 가장 큰 변화만 고려하고 1개 채널에 대해서 DCT를 수행 하여도 같은 효과를 볼 수 있다. 따라서 식(1)과 (2)와 같이 RGB 가운데 가장 큰 차이를 갖는 채널을 구한 후 해당 채널만 Block DCT를 수행한다. R_r 은 참조 프레임에서 R 채널의 8x8 block 평균 화소값(Average pixel value)이고 R_c 는 현재 프레임에서 R 채널의 8x8 block 평균 화소값이다. G, B의 경우도 같은 방법으로 계산된다.

$$Diff_{max} = \max[(R_r - R_c), (G_r - G_c), (B_r - B_c)] \quad (1)$$

$$DCT_{current} \begin{cases} DCT_{R_c} & \text{if } Diff_{max} = (R_r - R_c) \\ DCT_{G_c} & \text{if } Diff_{max} = (G_r - G_c) \\ DCT_{B_c} & \text{if } Diff_{max} = (B_r - B_c) \end{cases} \quad (2)$$

이렇게 함으로써 DCT를 각 block 당 한 개 채널만 수행하면서, DC 요소의 값이 결국 두 프레임에서 RGB 차이의 최댓값으로 나타낼 수 있는 장점이 있다. 또한 식 (3)과 같이 $Diff_{max}$ 의 값이 경험적으로 설정된 ρ 값 이상일 경우는 확실한 전경으로 분류하고 나머지에 대해서만 DCT를 수행한다.

$$Current\ Block \begin{cases} Foreground & \text{if } Diff_{max} > \rho \\ DCT & \text{otherwise} \end{cases} \quad (3)$$

3. Guided 필터를 이용한 불충분 분류 개선방법

그림 5은 성능이 낮아지는 영역을 확대한 것이다. 이진 분류에서는 명확히 전경으로 결정된 영역들이 Guided 필터에 의해서 값이 낮아지고 unknown 값으로 필터링 되는 것을 관찰할 수 있다.

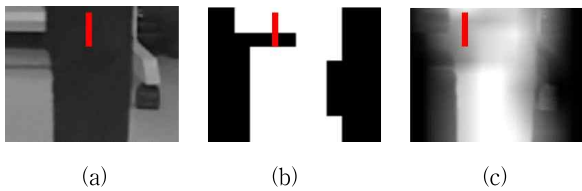


그림 5. Guided 필터의 한계 예시 (a) Guidance 영상 (b) 입력 영상 (c) 출력 영상

예시 영상에서 전경과 배경의 경계부분(빨간 표시 영역)을 1차원 신호로 분석하면 그림 6과 같이 나타낼 수 있다.

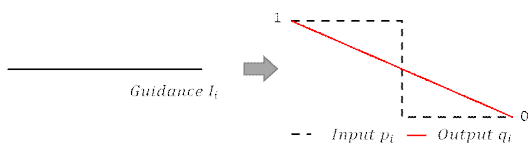


그림 6. Guided 필터 1차원 신호 분석

Guidance영상이 거의 평평하기 때문에 구하려는 분산과 공분산의 값이 0에 가까운 값을 갖는다. 이에 따라 식(5), (6)에서 계수를 구하면 식(7)과 같이 출력영상의 값이 결국 평균 입력영상이 된다.

$$\begin{aligned} a &\approx 0 \\ b &= 0 \\ q_i &= b \end{aligned} \quad (7)$$

where flat, $Var_I \approx 0$, $Cov_{Ip} \approx 0$

Guided 필터는 guidance 영상에 따라 필터 출력이 영향을 받기 때문에 flat한 영역에서는 결과 성능이 저하된다. 또한 이 영역들의 공통적인 특징은 실제로는 전경이지만 이진 분류과정에서 확실하게 분류되지 않은 상태에서 guidance 영상도 평평한 어려운 경우이다. 이 문제를 해결하기 위해서 이진 분류로 얻은 입력영상을 개선한 후 matting을 적용하는 방법을 제안한다. 이를 위해 본 논문에서는 Guided 필터를 두 번 적용함으로써 성능감소를 최소화한다. 먼저 이진 분류로 얻어진 입력영상을 1st Guided 필터를 적용하여 입력 영상을 개선한다. Unknown 값으로 된 영역을 찾기 위해 식(8)를 적용한다.

$$q_i^* = q_i / \mu_p \quad (8)$$

q_i^* 는 결정 파라미터로써 Guided 필터 출력 (q_i) 과 입력영상의 평균치(μ_p)의 비율 값이다. 이 값은 guidance가 평평한 경우와 edge인 경우로 나누어서 고려한다. 첫 번째 경우를 그림 7에서 살펴보면 guidance가 평평한 영역은 Guided 필터 출력이 평균 입력 결과와 같아지기 때문에 비율이 1이 되는 것을 볼 수 있다.

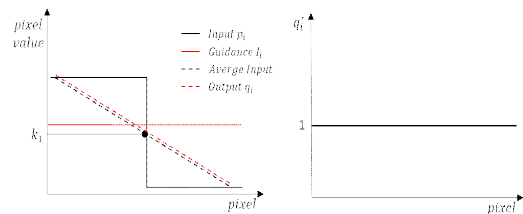


그림 7. Flat 영역에서의 Decision Parameter

이 경우에는 결정 파라미터가 1인 영역을 명확히 전경이라고 판단함으로써 안정적인 새로운 입력영상을 만들 수 있다. 그림 7에서 점 k_1 의 오른쪽부분은 실제 전경 영역이지만 이진 분류에서 전경으로 분류되지 못했던 영역으로 새로운 입력 생성과정에서 올바른 영역으로 재분류된다.

두 번째 경우를 그림 8에서 살펴보면 출력이 guidance에 영향을 받아 edge가 보존되어 결정 파라미터의 값이 진동하는 것을 관찰할 수 있다.

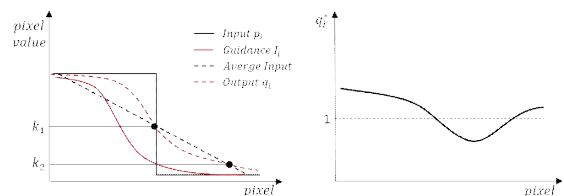


그림 8. Edge 영역에서의 Decision Parameter

이 경우에는서 결정 파라미터가 1일 때만 고려한다면 판별오류가 발생될 수 있으므로 결정 파라미터가 1보다 큰 경우를 진정으로 판단해야 한다. 그러면 점 k_1 까지 원래의 입력영역을 보존할 수 있게 된다. 그러나 그림에서 보이듯이 점 k_2 부터 출력과 평균 입력의 비율이 역전되는 것을 관찰할 수 있다. 왜냐하면 guidance의 영향을 받아 edge 보존력에 의해 이 영역에서 입력의 smoothing 정도가 평균 입력보다 커지기 때문이다. 이를 방지하기 위한 추가적인 파라미터를 고려한다. 그림 9은 두 번째 경우의 결과인데 β 를 적용한다면 역전현상이 일어난 영역이 새로운 입력영역으로 잘못 판단되는 것을 방지할 수 있다. 여기서 β 는 평균 입력에 대한 파라미터이며 실제 알고리즘 적용 시 평균 입력에 대해 적용 범위를 조절한다.

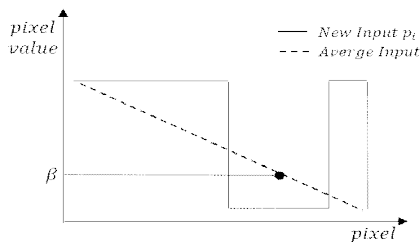


그림 9. Edge 영역에서의 새로운 입력 결과

이에 따라 제안하는 방법을 식(9)과 같이 나타낼 수 있다.

$$i \rightarrow \text{New input Region} \quad \text{if } q_i^* \geq 1 \ \& \ \mu_p > \beta \quad (9)$$

얻어진 결과영상과 기존의 이진 분류 영상을 OR Operator를 취해 새로운 입력영상을 생성한다.

그림 10는 제안하는 방법으로 개선한 결과를 보여준다.



그림 10. 입력영상 비교(a) 이진 입력 영상 (b) 새로운 입력 영상

위의 과정을 통해 성능이 향상된 새로운 입력 영상은 처음 입력과 마찬가지로 under/over 이진 분류 영상이므로 이를 해결함과 동시에 Alpha 영상으로 만들기 위해 2nd Guided 필터처리를 활용한 matting 처리를 한다.

4. 실험

그림 11는 기존 방법 중에서 texture정보를 분석하는 feature로써 복잡도가 낮고 주로 사용하는 LBP(Local Binary Pattern)의 Histogram 방법[4]과 DCT방법을 적용하여 이진 분류한 결과이다. LBP 검사 화소는 1 화소이며 비트 수는 8비트로 LBP를 설계하였다.

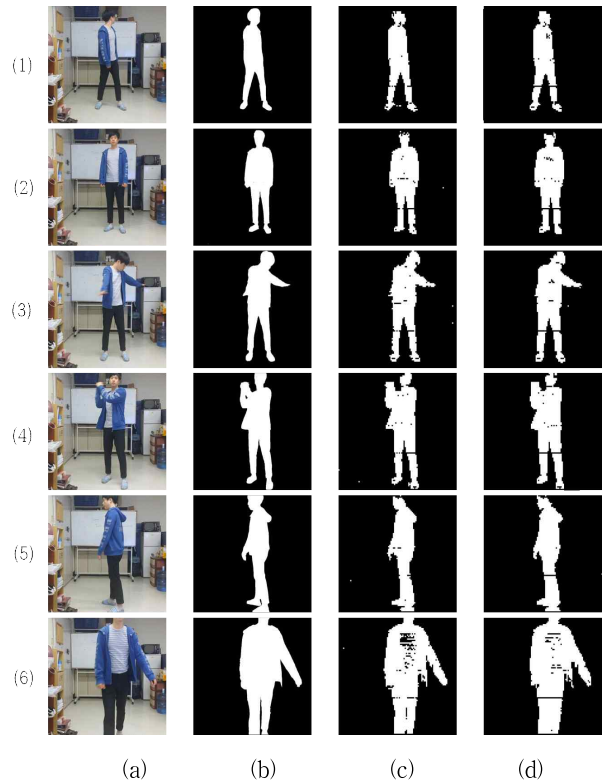


그림 11. 이진 분류 성능 비교 영상 (a)원본 영상 (b)실측 영상 (c) LBP 방법(d) DCT 방법

표 1은 DCT의 성능을 검증하기 위해서 각 영상의 (Ground Truth)영상과 이진분류영역의 화소 개수 차이를 측정된 것이다. (영상 크기 1280x720, 총 화소개수 921600)

Method \ Image No.	G.T-DCT	G.T.-LBP
Image (1)	1545	2569
Image (2)	352	3040
Image (3)	662	3434
Image (4)	1270	4214
Image (5)	1311	5607
Image (6)	6268	14716

표 1. 이진 분류 화소 개수 비교

표와 같이 DCT방법에서 실측 자료와 화소 개수가 상대적으로 유사한 것을 관찰할 수 있다.

그림 13은 동일한 입력으로 matting을 하였을 때 기존의 Guided 필터방법과 제안하는 Guided 필터방법을 적용한 결과이다.

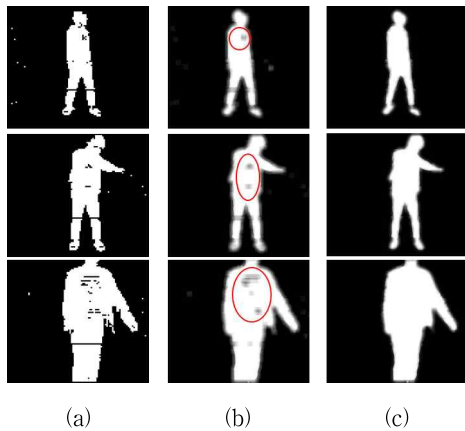


그림 13. Guided 필터 Matting 출력 비교 (a) 입력 영상 (b) 기존 결과 (c) 제안하는 결과

그림과 같이 기존의 Guided 필터에서 보였던 문제 영역(빨간 표시 영역)이 개선되는 것과 동시에 다른 영역은 그대로 보존됨을 관찰할 수 있다.

5. 결론

제안하는 알고리즘은 intensity와 texture의 변화를 측정하여 분류를 하는 방법으로 DCT 분석을 기반을 두었고 RGB 색정보를 추가적으로 포함하여 알고리즘의 성능과 속도를 높였다. Matting의 유용한 Guided 필터를 사용하는 것과 동시에 필터의 한계를 극복하여 성능을 향상시켰다.

제안하는 알고리즘은 결과 값이 영상 합성을 위한 Alpha값이기 때문에 다양한 응용분야에서 쉽게 사용될 수 있다는 장점이 있다. 이는 Alpha 값이 영상 전체에서 전경만을 색인화하기 때문에 영상 전송/압축하는 응용분야에서 활용도가 높아질 것이다. 본 논문에서는 움직이는 물체를 안정적이며 실시간으로 추출할 수 있는 새로운 matting 방법을 제시하고 있지만 몇 가지 제한사항이 존재한다. 참조 프레임과 현재 프레임을 block 단위로 비교하기 때문에 전경 영역일지라도 만약 block의 색 및 texture 특성이 유사하다면 전경으로 판단되기 어려울 것이다. 위의 한계점들은 후속연구에서 다루어 질 것이다.

감사의 글

이 논문은 2015년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(2015R1D1A1A01059722)

참고문헌

[1] J.J. Verbeek, N. Vlassis and B. Kröse, "Efficient Greedy Learning of Gaussian Mixture Models," Published in Neural Computation 15(2), pages 469-485, 2003.
 [2] T. Yu, C. Zhang, M. Cohen, Y. Rui and Y. Wu, "Monocular Video Foreground/Background Segmentation by Tracking Spatial-Color Gaussian Mixture Models," . In IEEE Workshop on Motion and Video Computing, pages 5 - 5, 2007.
 [3] B. Vishnyakov, V. Gorbachevich, S. Sidiyakin, Y. Vizilter, I. Malin and A. Egorov, "Fast Moving Objects detection Using iLBP

Background Model," The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XL-3, 2014.

[4] X. Sun and F. Chang, "Background Model Combining Gauss Model with Local Binary Pattern Feature," Journal of Convergence Information Technology(JCIT) Volume 7, Number 17, Sep 2012.

[5] A. Shimada and R.I Taniguchi, "Hybrid Background Model using Spatial-Temporal LBP," DOI: 10.1109/AVSS.2009.12 Conference: Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance, AVSS 2009, 2-4 September 2009.

[6] J. Yao and J.M. Odobez, "Multi-Layer Background Subtraction Based on Color and Texture," IN IEEE The CVPR Visual Surveillance Workshop (CVPR-VS), MINNEAPOLIS, June 2007.

[7] M.Heikkila and M.Pietikainen, "A Texture-Based Method for Modeling the Background and Detecting Moving Objects," IEEE Transactions on Pattern Analysis and Machine Intelligence, VOL. 28, NO. 4, April 2006.

[8] L. Li, W. Huang, I. Y.H. Gu, and Q. Tian, "Foreground Object Detection from Videos Containing Complex Background," In Proceedings of the Eleventh ACM International Conference on Multimedia, Berkeley, CA, USA, 2 - 8 November 2003.

[9] L. Jeisung and P. Mignon, "An Adaptive Background Subtraction Method Based on Kernel Density Estimation," Sensors journal, 2012.

[10] R. Rodriguez-Gomez, E. J. Fernandez-Sanchez, J. Diaz and E. Ros, "FPGA Implementation for Real-Time Background Subtraction Based on Horprasert Model," Sensors journal, 2012.

[11] V.Ganapathi, C. Plagemann, D. Koller and S. Thrun, "Real Time Motion Capture Using a Single Time-Of-Flight Camera," In Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13 - 18 pp. 755 - 762. June 2010.

[12] V. Kolmogorov, A. Criminisi, A. Blake, G. Cross and C. Rother, "Bi-layer segmentation of binocular stereo video," In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, CA, USA, 20 - 25 pp. 407 - 414. June 2005.

[13] E. J. Fernandez-Sanchez and J.Diaz and E.Ros, "Background Subtraction Based on Color and Depth Using Active Sensors," Sensors journal, 2013.

[14] R. Lienhart, A. Kuranov, and V. Pisarevsky, "Empirical Analysis of Detection Cascades of Boosted Classifiers for Rapid Object Detection," Proc. DAGM 25th Pattern Recognition Symp., pp. 297-304, 2003.

[15] R. C. Gonzalez, R. E. Woods and S. L. Eddins, Digital Image Processing Using MATLAB, Gatesmark Publishing, 2009.

[16] T. porter and T.Duff. "Compositing Digital Images," Computer Graphics Volume 18, Number 3 July 1984.