

텐서보팅과 마르코프 랜덤 필드를 이용한 자연 영상의 텍스트 이진화

Natural Scene Text Binarization using Tensor Voting and Markov Random Field

최현수*, 이귀상**

(Hyun Su Choi, Guee Sang Lee)

요약

본 논문에서는 텐서보팅을 이용하여 기존 마르코프 랜덤 필드 메소드 안의 가우시안 혼합 모델 함수의 성능을 향상시킬 수 있는 적합한 클러스터 개수 검출 방법을 제시한다. 제안하는 방법의 핵심 포인트는 텐서보팅의 인풋 데이터 토큰의 연속성인 saliency map을 통한 중심점 개수의 추출이다. 우리는 가장 먼저 주어진 자연 영상에서 전경 및 배경 후보 영역을 분리한다. 다음으로, 분리된 각 후보 영역에 대하여 텐서보팅을 적용하여 적절한 클러스터 개수를 추출한다. 우리는 검출된 클러스터 개수를 이용하여 정확한 가우시안 혼합 모델 모델링을 수행할 수 있다. 그리고 이를 적용한 마르코프 랜덤 필드의 unary term과 pairwise term을 계산하여 자연 영상의 텍스트 이진화 결과를 반환한다. 실험 결과, 제안된 방법이 최적의 클러스터 개수를 반환하고, 향상된 텍스트 이진화 결과를 반환함을 확인하였다.

■ 중심어 : 이진화 ; 마르코프 랜덤 필드 ; 가우시안 혼합 모델 ; 텐서 보팅

Abstract

In this paper, we propose a method for detecting the number of clusters. This method can improve the performance of a gaussian mixture model function in conventional markov random field method by using the tensor voting. The key point of the proposed method is that extracts the number of the center through the continuity of saliency map of the input data of the tensor voting token. At first, we separate the foreground and background region candidate in a given natural images. After that, we extract the appropriate cluster number for each separate candidate regions by applying the tensor voting. We can make accurate modeling a gaussian mixture model by using a detected number of cluster. We can return the result of natural binary text image by calculating the unary term and the pairwise term of markov random field. After the experiment, we can confirm that the proposed method returns the optimal cluster number and text binarization results are improved.

■ keywords : Binarization ; Markov random field ; Gaussian mixture model ; Tensor voting

I. 서론

클러스터의 개수 K 를 결정하는 것은 가우시안 혼합 모델을 다루기 이전에 선행되어야 하고, 또 매우 중요한 작업이다. 클러스터 개수를 다르게 함에 따라서 가우시

안 혼합 모델 함수의 분류 결과는 매우 달라질 수 있다. 우리는 일반적으로 가장 데이터 간의 밀집도를 높게 가지면서도, 개수는 많지 않은 적절한 개수 K 개를 선택하여 사용한다. 데이터의 밀집이 적절할 경우 정확도를 얻을 수 있고, 클러스터의 개수가 적절할 경우 시간을 얻을 수 있다. 이 둘은 서로 반비례 관계이다. 많은 경

* 학생회원, 전남대학교 전자컴퓨터공학과

** 정회원, 전남대학교 전자컴퓨터공학과

이 논문은 2015년도 정부(교육부)의 재원으로 한국연구재단(NRF-2013R1A1A4A03006535 and NRF-2015R1D1A1A01060172)의 지원을 받아 수행되었음.

접수일자 : 2015년 08월 28일

수정일자 : 2015년 12월 28일

게재확정일 : 2015년 12월 30일

교신저자 : 이귀상 e-mail : gslee@jnu.ac.kr

우에 있어서 클러스터링 대상영역의 색깔에 대한 집합과 비슷하게 클러스터 개수 K 가 선택되었을 때, 클러스터링 작업을 수행한 후 최적의 분류 결과가 획득되는 것을 확인할 수 있다.

이처럼 K 의 개수가 결과에 중요한 영향을 미친다는 것을 알고 있기 때문에, 우리는 K 의 개수를 선택하는 여러 가지 방법을 사용한다. 가장 좋은 결과를 낼 수 있는 방법은 유저가 하나의 파라미터로서 K 를 일일이 설정해 주는 것이다. 이는 유저에게 매우 의존적이고, 수동적인 방법이다. 가장 많이 사용하는 방법으로는 Mishra의 논문에서와 같이[1], 평균적으로 결과가 좋은 클러스터 개수를 사용하는 것이다. 하지만 이는 샘플에 따라 결과의 편차가 매우 클 수 있고, 좋은 결과를 반환한다는 보장이 없다. 그림 1은 클러스터 개수를 고정함에 따라 달라지는 가우시안 혼합 모델 함수의 결과를 포함한 텍스트 이진화 결과를 나타낸 것이다. 자연 영상의 텍스트 이진화를 구현한 Mishra의 논문에서는 클러스터의 개수를 각 5개로 고정하는 경우를 추천했지만, 실제로 예외 사항이 존재하는 것을 확인할 수 있다.

우리는 또한 클러스터 개수 K 를 결정하기 위해서, 클러스터링 전에 K 를 미리 추정하는 방법을 사용할 수 있다. 보편적으로 알려진 방법은 validity 계산에 기반한 방법이다[2-5].

이 방법은 클러스터링 영역의 자료를 이용하여 validity를 계산한 후, 이를 기반으로 클러스터의 개수 K 를 결정한다. Validity의 계산은 얼마나 클러스터가 잘 퍼져 있으면서도 잘 분리되어 있는지를 계산하여 이루어진다. 또 다른 클러스터 개수 K 를 미리 추정하는 방법으로, Toan의 논문에서는 텐서보팅을 이용하여 클러스터링 방법 K-means에서 초기 파라미터 K 값을 결정하는 방법이 제안되었다[6]. 우리는 이 논문에서, Toan의 방법을 기존의 마르코프 랜덤 필드 안에서 likelihood를 계산하는 또 다른 클러스터링 방법인 가우시안 혼합 모델과 결합시킬 것이다.

기존 마르코프 랜덤 필드를 적용한 텍스트 이진화 결과는 클러스터의 개수에 매우 의존적이고, 개수를 고정할 경우 결과가 불안정하다. 이 논문에서는 마르코프 랜덤 필드의 후보 영역 생성 후, 입력 데이터인 토큰의 연속성으로 나타내어지는 saliency map을 통한 중심점 개수의 추출을 통해 클러스터 개수 K 를 결정한다. 이는 텐서보팅[11]을 이용하여 이루어지고, 적응적인 클러스터 개수의 생성을 통해서 더욱 향상된 텍스트 이진화 결과[12]를 반환한다.



그림 1. 클러스터 개수 선택에 따른 텍스트 이진화 결과 (a,d) 원본 영상, (b,e) 배경/전경 각 2개, (c,f) 배경/전경 각 5개

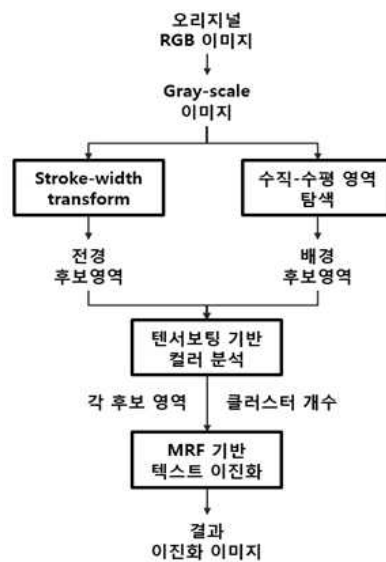


그림 2. 제안 알고리즘에 대한 블록도

본 논문은 다음과 같은 순서로 구성되어 있다. 먼저 2장에서는 본 논문에서 제안한 알고리즘 및 그 배경 지식에 대하여 설명한다. 다음으로 3장에서는 실험 결과를 보인 후, 그 결과에 대하여 설명한다. 마지막으로 4장에서는 결론 및 향후 연구에 대하여 기술한다.

II. 본 론

1. 제안 알고리즘

전체 시스템 알고리즘은 다음 3단계로 나뉜다 : 전경

-배경 후보영역 분리, 텐서보팅 기반의 클러스터 개수 검출, 검출된 클러스터를 이용한 가우시안 혼합 함수 클러 모델링 및 마르코프 랜덤 필드를 이용한 텍스트 이진화. 그림 2는 위와 같은 전체 알고리즘을 도면으로 나타낸 것이다. 알고리즘의 입력값은 오리지널 RGB 이미지이고, 결과값은 픽셀 값이 0과 1로 이루어진 이진화 이미지이다. 결과값인 이진화 이미지에서의 0은 배경을 뜻하고, 글자 외의 배경 영역을 나타낸다. 반면 1은 전경을 뜻하고, 텍스트 영역을 나타낸다.

가. 후보영역 추출

제한한 알고리즘의 첫 번째 작업은 텐서보팅 및 가우시안 혼합 모델을 적용할 배경-전경 후보 영역을 추출해 내는 것이다. 이진화 결과는 첫 단계에서 얼마나 정확한 후보 영역을 추출해 내는 것에 의존한다. 우리는 텍스트를 이진화 하기 위해 가장 적합한 추출 방법을 선택하였다.

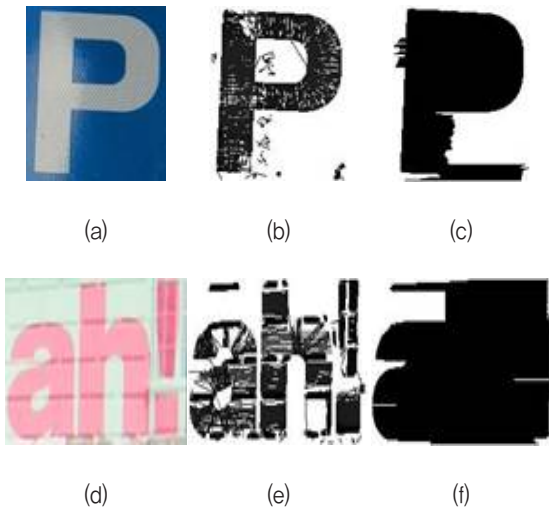


그림 3. 입력 영상의 전경 및 배경 후보영역 추출 결과 (a,d) 원본 영상, (b,e) 전경 후보영역, (c,f) 배경 후보영역

먼저, 전경 후보영역은 stroke width transform을 이용하여 추출된다[7]. 이는 입력 이미지 boundary의 gradient 방향을 이용하여 계산된다. 전경 후보 영역 검출 결과, 글자의 획이 포함된 영역이 반환된다. 배경 후보 영역은 비교적 간단하다. 먼저 영상의 boundary를 구한 후, 수직-수평 탐색을 시행한다. Boundary를 구하는 방법은 캐니 엣지 탐색 알고리즘을 사용한다[8]. 우리는 캐니 엣지 알고리즘의 파라미터를 조정하여 강력한 픽셀값의 변화만 엣지로 받아들일 수 있다. 그 후 예지에

대한 수직-수평 탐색을 수행하여, 예지를 만나지 않은 확실한 배경 후보 영역만 반환해 낸다. 그림 3은 입력 이미지에 대한 전경 및 배경 후보영역 추출 결과를 나타낸 것이다.

나. 클러스터 개수 검출

적절한 후보영역을 추출해 낸 후, 우리는 이 seed 영역을 기반으로 하여 클러스터 개수를 추정한다. 우리는 적절한 클러스터 개수를 추정하기 위해, Toan이 제안한 텐서보팅을 이용한 파라미터 추정을 각각의 후보 영역에 적용한다[6]. 이는 본래 K-means 군집화의 초기 파라미터 결정에 이용되었다. 텐서는 데이터 값과 이웃한 데이터와의 관계를 반영한 하나의 특징벡터라고 볼 수 있다. 불이나 스틱 등의 설정 영역 안의 이웃하는 토큰을 통해 생성 및 재생성된 텐서는 두 개의 고유값 λ 및 고유벡터 e 를 가지고 있다. 이를 포함한 텐서 정보는 다음과 같이 공분산 행렬로 나타내어질 수 있다.

$$T = \lambda_1 \hat{e}_1 \hat{e}_1^T + \lambda_2 \hat{e}_2 \hat{e}_2^T = (\lambda_1 - \lambda_2) \hat{e}_1 \hat{e}_1^T + \lambda_2 (\hat{e}_1 \hat{e}_1^T + \hat{e}_2 \hat{e}_2^T) \quad (1)$$

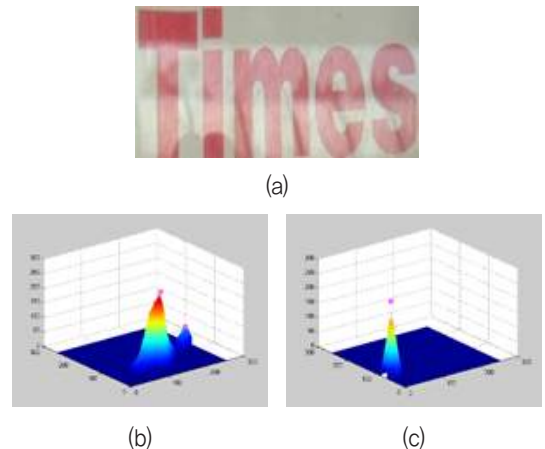


그림 4. 입력 이미지에 대한 surface saliency map (a) 원본 영상, (b) 전경 후보영역, (c) 배경 후보영역

우리는 여기서 2D stick 텐서의 saliency를 나타내는 $(\lambda_1 - \lambda_2)$ 정보를 사용한다. 해당 정보를 시각화하여 확인하면, 우리가 목표로 한 것처럼 각 색 집합의 분포에 맞는 꼭지점을 각각 확인할 수 있다. 우리는 여기서 꼭지점의 최소 높이 및 꼭지점간의 거리 설정을 통해 적합한 꼭지점의 개수, 즉 적합한 클러스터의 개수를 반환한다. 텐서보팅의 더욱 자세한 개념은 G. S. Lee의 논문에서 확인할 수 있다[9]. 그림 4는 입력 이미지에 대한 surf

ace saliency map을 나타낸 것이다. 우리는 이 그림을 통해서 각 saliency 에너지의 집합이 우리가 원하는 클러스터링 대상영역의 색깔에 대한 집합과 유사함을 확인할 수 있다.

다. MRF 기반 텍스트 이진화

적절한 후보영역 및 클러스터 개수를 얻은 후, 우리는 마르코프 랜덤 필드를 이용하여 텍스트 이진화 작업을 수행한다. 이진화 문제는 또 다른 측면에서 픽셀 라벨링 작업으로 정의할 수 있다. 우리는 Mishra의 논문에서 제안한 마르코프 랜덤 필드를 사용하여 라벨링 작업을 수행한다[1]. 라벨 0은 전경, 라벨 1을 배경을 나타내고 이를 수식화하면 라벨 $x_i \in \{0, 1\}$ 이다. 라벨링 문제를 해결하기 위한 에너지 식은 다음과 같다.

$$E(x, \theta, z) = E_i(x, \theta, z) + E_{ij}(x, z) \quad (2)$$

여기서 z 는 각 픽셀의 색 값이고 θ 는 전경/배경 후보 영역에 의해 생성된 가우시안 혼합 모델 파라미터 셋이다. 이웃한 픽셀 i, j 에 대해서, 우리는 i 에 대한 unary term과 i 와 j 의 관계인 pairwise term을 계산한다. 먼저, unary term에서는 가우시안 혼합 모델을 이용해 likelihood를 구하는 작업을 수행한다. Likelihood계산에 대한 unary term 식은 다음과 같다.

$$E_i(x, \theta, z) = -\sum_i \log P(z_i | x_i, \theta) \quad (3)$$

이는 앞에서 구한 후보 영역에 의해 생성된 가우시안 혼합 모델을 통해 구해질 수 있다. 사용된 likelihood에 대한 식은 다음과 같다.

$$P(z_i | x_i, \theta) = N(z, \theta; \mu(x_i | k_{x_i}), (\sum_i (x_i | k_{x_i}))) \quad (4)$$

여기서 k_{x_i} 는 각 레이블 x_i 에 대한 클러스터 개수이고 이를 수식화 하면 $k_{x_i} \in \{0, \dots, N_{x_i}\}$ 이다. 다음으로 pairwise term에서는 이웃한 픽셀과의 값의 차이와 gradient 성분의 차이를 이용하여 에너지를 계산한다. 이는 Mishra가 제안한 방법으로[1], 입력 데이터에서 전경과 배경의 각 픽셀이 이웃한 같은 레이블의 픽셀과 비슷한 값을 가질 것이라는 전제로 만들어진 식이다. 이를 구현한 pairwise term에 대한 식은 다음과 같다.

$$E_{ij}(x, z) = [x_i \neq x_j] ((\lambda_1 \exp \sum_{i, j \in N} (-\beta (z_i - z_j))) + (\lambda_2 \exp \sum_{i, j \in N} (-\beta (w_i - w_j)))) \quad (5)$$

여기서 w 는 각 gradient 성분 값이다. 우리는 파라미터 λ 와 β 의 조정을 통해서 unary와 pairwise의 가중치 및 값 차이에 대한 가중치를 조정할 수 있다. Pairwise 에너지는 식 (2)와 같이 unary 에너지와 함께 에너지를 생성한다. 여기서 에너지는 일종의 cost 값으로 우리는 낮은 에너지를 반환하는 결과를 선택한다. 이는 graph cut을 이용하여 이루어진다.

2. 실험 결과

제안된 알고리즘의 성능 실험을 위해서 ICDAR 2003 테스트 이미지를 사용하였다. 이중에 입력 이미지는 하나의 텍스트 이미지이다. 결과로 나온 이진화 이미지는 이전의 각각 5개로 고정된 파라미터를 사용한 이진화 결과와 비교된다. 구현 코드는 Mishra의 논문을 구현한 코드를 사용하였다[1].

그림 5는 입력 이미지에 대해서 적절한 클러스터 개수 사용과 고정된 클러스터 개수 사용에 따른 텍스트 이진화 결과를 서로 비교한 것이다. 실험 결과, 우리는 적절한 클러스터 추정 시 더 개선된 이진화 이미지를 반환하는 것을 눈으로 확인할 수 있다.



그림 5. 기존 방법과 제안된 방법의 텍스트 이진화 결과 (a,d) 원본 영상. (b,e) 고정된 클러스터 개수 사용. (c,f) 제안된 방법

성능 평가는 자체 생성한 ground truth를 적용한 F-s

core를 이용하여 수행되었고, F-score를 위한 정확도 precision과 recall의 계산 방법은 다음과 같다.

$$Precision = \frac{T}{T+F} \times 100\% \tag{6}$$

$$Recall = \frac{T}{S} \times 100\% \tag{7}$$

우리는 이 계산된 precision과 recall을 이용하여 F-score를 계산할 수 있다.

$$Fscore = \frac{2 \times (Precision + Recall)}{Precision + Recall} \times 100\% \tag{8}$$

아래 표 1은 위 식을 적용한 15개의 이미지에 대한 성능 평가 결과이다. 우리는 표 안에서 제안하는 알고리즘이 기존의 제안 방법인 클러스터의 개수를 선행 시험 결과 평균적으로 좋은 결과를 반환하는 어떠한 수로 고정(배경 5/전경5) 하는 방법에 비하여 더욱 향상된 결과를 반환함을 확인할 수 있다.

표 1. 각 방법에 대한 이진화 결과 비교 (백분율, %)

클러스터 개수 선택 방법	Prec.	Recall	F-score
기존 제안 방법 (전경/배경 5개 고정)	92.32	95.31	93.79
제안된 알고리즘 (텐서보팅 결과값 사용)	96.87	97.51	97.19

III. 결론

이 논문에서는 가우시안 혼합 모델의 성능을 향상시킬 수 있는 적합한 클러스터 개수 추출 방법이 제안되었다. 일반적으로 유저들이 일일이 선택하거나, 미리 실험한 결과를 이용하여 가장 확률이 높은 몇 가지의 숫자가 고정적으로 사용되어 왔다. 우리의 알고리즘으로 인해, 이러한 선행 방법들의 불안정성이 해결되었다. 실험 결과는 제시한 알고리즘을 적용한 텍스트 이진화 결과가 더욱 적응적이면서도 향상된 결과임을 보여 주고 있다. 향후 연구로는 더욱 정확한 중심점 검출을 위해 잡음 제거 등의 효과적인 처리와의 결합 등의 방법이 필요하다고 여겨진다.

참고 문헌

- [1] Mishra A., Alahari K., Jawahar C.V., "An MRF Model for Binarization of Natural Scene Text," *Proc. 11th International Conference on Document Analysis and Recognition*, pp. 11-16, 2011.
- [2] U. Maulik and S., Bandyopadhyay, "Performance Evaluation of Some Clustering Algorithms and Validity Indices," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, pp. 1650-1654, 2002.
- [3] J.C. Bezdek and N.R. Pal, "Some New Indices of Cluster Validity," *IEEE Trans. System, Man and Cybernetics*, vol. 28, no. 3, pp. 301-315, 1998.
- [4] D.L. Davies and D.W. Bouldin, "A cluster separation measure," *IEEE Trans. Pattern Anal. Machine Intell*, vol. 1, pp. 224-227, 1979.
- [5] S Ray and R H Turi, "ADetermination of number of clusters in K-means clustering and application in colour image segmentation," *Proceedings of the 4th International Conference on Advances in Pattern Recognition and Digital Techniques*, pp. 137-143, 1999.
- [6] Toan Nguyen, Park, J., Lee, G., "Determination of Initial Parameters for K-means Clustering by Tensor Voting," *Korea Computer Congress*, vol. 36, 2009.
- [7] Epshtein B., Ofek E., Wexler Y, "Detecting text in natural scenes with stroke width transform," *Proc. Computer Vision and Pattern Recognition*, pp. 2963-2970, 2010.
- [8] Canny J. " Computational Approach to Edge Detection," *IEEE Trans. on PAMI* 8, pp. 679-698, 1986.
- [9] Lee, G., Toan Nguyen, Park, J., "Extraction of Text Alignment by Tensor Voting and its Application to Text Detection," *Korea Institute of Information Scientists and Engineers(KIISE)*, vol. 36, 2009.
- [10] Y. Boykov and M.-P. Jolly, "Interactive graph cuts for optimal boundary and region segmentation of objects in n-d images," *In ICCV*, pp. 105-112, 2009
- [11] 박종현, 박순영, 이귀상, "텐서보팅에 기반한 영상처리 및 응용," *스마트미디어저널*, vol. 1, no. 2,

pp.23-33, 2012.

- [12] Le Thi Khue Van and Guesang Lee, "Text Extraction in HIS Color Space by Weighting Scheme," Smart Media Journal, vol. 2, no. 1, pp. 31-36, 2013.

저 자 소 개



최현수 (학생회원)

2014년 전남대학교 전자컴퓨터공학
부(컴퓨터공학전공) 학사 졸업.
2015년 전남대학교 전자컴퓨터공학
과 석사 재학.

〈주관심분야 : 멀티미디어 영상 처리, 패턴 인식〉



이귀상 (정회원)

1980년 서울대학교 전자공학과 학사
졸업.
1982년 서울대학교 컴퓨터공학과 석
사 졸업.

1991년 펜실베이니아주립대학교 컴퓨터과학과 박사 졸업.

〈주관심분야 : 영상 처리, 컴퓨터 비전, 비디오 과학 기술〉