미생물 자동 분류 및 계수를 위한 영상 분석 시스템

권영희[0] 김진형 윤수영* 윤세왕*
한국과학기술원 전자전산학과, 대한제당(주) 중앙연구소*
(kyhee[0], jkim)@ai.kaist.ac.kr, (yun54*, swyoon*)@ts.co.kr

Image Analysis for Classifying and Counting Microbes

Younghee Kwon[0], Jin-Hyung Kim, Su-Young Yun*, Se-Wang Yoon*
Dept. of Electrical Engineering & Computer Science, KAIST
R&D Center, TS Corporation*

요 약

본 연구에서는 식품 안정성을 검사하기 위해 식품 내 미생물 농도를 계산하는 방법을 제안한다. 기존의 생물학적 방법을 보완하여 영상 분석에 기반한 자동화 방법으로 빠르고 보다 정확한 방법을 제안한다. 현미경으로 확대한 시료의 영상을 자동으로 분석하여 영상 내의 미생물의 수를 측정하면 여러 영상에 대한 결과를 통계적으로 분석하여 식품 내 미생물 농도를 보다 정확히 측정할 수 있다. 영상 처리 과정을 통해 영상으로부터 미생물군을 찾아내고, 해당 미생물군을 모양에 따라 세 가지 종류 – 간균, 구균, 효모로 분류하고 계수한다. 제안하는 시스템의 성능을 실험으로 보였다. 영상 내에서 측정된 미생물의 수와 실제 미생물의 수의 비를 검토해 본 바, 본 시스템이 실용적임을 보였다.

1. 서 론

매년 많은 식중독 사건이 발생하고 있으므로 식품의 안정성을 점검하는 것은 매우 중요한데, 식품 부패의 주요 원인은 미생물이다 [1]. 따라서 식품의 안정성은 식품 내의 미생물 농도를 측정함으로써 판단할 수 있다.

미생물의 농도 측정법에는 직접 배양법, ATP 측정법, 면역 분석법 등이 있으나[2] 각각이 내포한 한계점으로 인해 널리 사용되지 못하고 있다. 직접 배양법은 식품 내 미생물을 평판 배지에 수일간 배양한 후 그 수를 육안으로 측정하는 방법으로서, 소요 시간이 길고 정확성이 떨어지는 단점이 있다. ATP 측정법은 세포 구성 물질인 ATP를 측정하는 방식으로, 미생물 뿐 아니라 유기 물질까지 계산하는 단점이 있다. 면역 분석법은 항체를 이용하는 방법으로 특정 미생물을 검사할 때에 적용되는 방법이다.

이에 식품 안정성을 짧은 시간 내에 검사할 수 있도록 새로운 자동화 방법론을 제안하고자 한다. 새로운 방법은 식품으로부터 미생물을 분리하여 시료화한 후 현미경을 사용하여 시료 내 미생물을 직접 계수한 결과를 이용해 식품의 안전성을 판단하는 방법이다. 본 논문에서는 이 방법론의 일부로서, 현미경 상의 영상으로부터 미생물을 계수하는 영상 처리 시스템을 제안하고자 한다.

본 시스템은 검사 대상인 식품 시료를 현미경으로 확대하여 촬영한 디지털 영상을 입력으로 하고, 영상 분석 과정을 통해 해당 영상 내의 미생물을 분류, 계수한 결과를 출력으로 하는 시스템이다. 2장에서는 시스템 개요에 대해, 3장에서는 미생물을 분류하고 계수하는 방법에 대해, 4장에서는 실험 및 결과에 대해 설명하겠다.

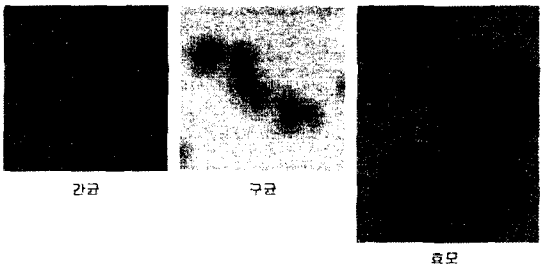

그림 1. 미생물의 종류

2. 시스템 개요

본 시스템의 대략적인 흐름은 다음과 같다.

영상을 입력 받아, 명도 영상으로 바꾸고 영상 내의 잡영을 제거하기 위해 Gaussian Smoothing을 해 주는 등의 전처리를 거친 후 관심 영역을 추출한다. 해당 관심 영역들에 대해 미생물을 모양에 따라 세 종류 – 간균, 구균, 효모로 분류하고 계수하여 결과를 출력한다. (그림2)

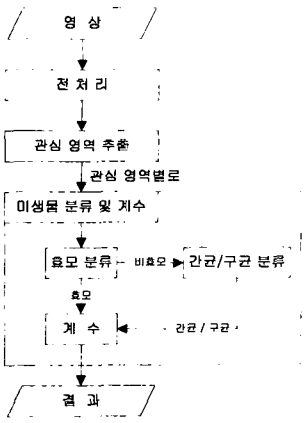

그림 2. 시스템의 흐름도

3. 분류 및 계수

미생물의 분류는 관심 영역을 추출하여 각 관심 영역별로 수행하고, 먼저 효모를 분류해 낸 후 간균과 구균을 분류한다. 그 후 분류된 관심 영역에 대해 계수를 한다.

관심 영역의 추출

미생물 영상에는 빈 공간이 많아 전체를 처리 대상으로 하기보다 미생물이 존재하는 부분만 선택하여 처리하는 것이 효율적이므로 관심 영역을 추출한다. 미생물이 배경에 비해 명도가 낮은 특성을 이용하여 이진화를 수행하는데, Otsu의 지역적 이진화 방법[3]을 일정 단위 크기마다 적용하고 픽셀별 명도 기준을 내삽법으로 계산하여 이진화하였다. 이진화하여 미생물 영역으로 구분된 픽셀들의 연결 화소(connected component)를 구하면 각각의 연결 영역이 관심 영역이다.

군집을 이루는 미생물은 모두 한 종류임을 가정하여 관심 영역 단위로 미생물을 분류하고 계수한다.

효모와 비효모의 분류

효모는 간균이나 구균에 비해 개체 크기가 큰 특징을 가지고 있어 크기를 이용해 구분하기가 쉬우므로, 미생물을 먼저 효모와 효모가 아닌 것으로 분류한 다음 효모가 아닌 것들을 간균과 구균으로 분류한다.

미생물을 분류하는 데 있어 가장 어려운 것은 미생물이 군체를 이룰 때 모이는 형태를 예측할 수 없다는 점이다. 즉, 일정한 구조를 이루기보다는 불규칙한 형태로 붙어 있는 경우가 많으므로 전체적인 모양을 살피는 것은 미생물을 분류하는 데 그리 도움을 주지 못하므로 지역적인 특성을 이용하여 미생물을 분류하는 것이 좋다.

효모는 한 개체의 크기가 크기 때문에 윤곽선이 상대적으로 완만한 반면, 간균이나 구균의 경우 개체 당 크기가 작기 때문에 모여 있어도 윤곽선에 급한 부분이 자주 나타나는 것을 볼 수 있다. 따라서 이러한 윤곽선의 완만도 분포를 이용하여 효모와 타종을 분류할 수 있다.

윤곽선이 완만하다는 것은 윤곽선을 따라가며 각 지점에서 접선을 그을 때 접선의 기울기 변화량이 작다는 것을 의미한다. 윤곽선의 일정 길이마다의 접선의 기울기 변화량 a는 윤곽선을 해당 길이로 분할한 각 지점을 연결한 선의 기울기를 차례대로 p, q라 하였을 때, $p-q$와 같게 된다. 즉 $p-q$가 크면 완만도가 큰 것이고 작으면 완만도가 작다. $p-q<0$인 경우는 해당 지점에서 윤곽선이 오목한 경우이다. (그림3)

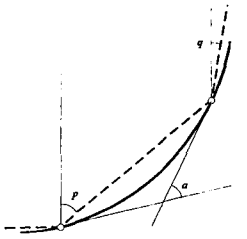

그림 3. 윤곽선의 접선 간 각도

윤곽선의 완만도를 나타내기 위해 일정 길이 분할선 간 각도를 순서대로 t벌씩 묶은 각도벌($a_i, a_{i+1}, \cdots, a_{i+t-1}$)을 특징으로 사용한다. 그리고 이 각도벌들을 윤곽선 전체에 대해 누적 분포를 구해 각도벌 분포를 계산한다. 이 각도벌 분포를 간균, 구균, 효모에 대해 각각 평균하여 종류별로 모델을 만든다.

주어지는 관심 영역을 분류하기 위해서는 관심 영역의 윤곽선을 구해 각도벌 분포를 구해 각 모델의 각도벌 분포와 원형 정합을 통해 가장 유사한 모델을 찾아 관심 영역을 분류한다.

간균과 구균의 분류

개체 하나하나를 보았을 때, 간균은 모양의 변이가 심하지만 구균은 모양의 변이가 적어 작은 구형의 모양을 가지므로 한 마리만 있을 때에는 간균과 구균을 분류하는 것이 비교적 수월한 편이다. 그러나 [그림1]에서 볼 수 있듯이, 간균과 구균 모두 군집을 이룰 때에는 전체적인 형태가 불규칙하게 된다. 따라서 간균과 구균을 효과적으로 분류하기 위해서는 군집 자체에 대해 분석하는 것보다 군집에서 각 개체를 분리해 낸 후에 각 개체에 대해 분류하는 것이 낫다.

군집에서 개체 영상을 분리해 내는 방법에는 능동 윤곽선을 이용하는 방법[4]과 이진화를 하는 방법, 형태를 미리 알고 있을 경우 템플릿을 이용해 위치를 찾아내는 방법[5] 등이 있을 수 있다.

영상에서 미생물 개체를 분리해 내기 위해서 본 연구에서는 지형학적 특징(Topographic Feature)[6]을 이용하였다. 이 방법은 명도 영상을 일종의 지형도로 보고

픽셀의 명도를 지면의 높이로 보아 각 픽셀을 주위의 명도 분포에 따라 지형학적으로 분류한 것이다. 이 방법을 사용하면 미생물이 만나는 지점에서 영상을 분할하는 것이 가능하다.

Trier가 제안한 방법에 따라 분할 가능한 지점을 분할한 후 연결 화소를 구하면 이렇게 구해진 연결 화소들은 미생물의 개체별 영역이 된다. 각 연결 화소에 대해 중간값 필터 등의 후처리를 통해 미생물 영역을 부드럽게 한다. 각 연결 화소에 해당하는 픽셀을 취하여 미생물 개체의 영상을 얻을 수 있다.

각 미생물 개체 영상이 간균인지 구균인지 분류하기 위해 간균과 구균을 구분하는 특징을 추출하여 인식기를 훈련시킨다. 간균과 구균을 효과적으로 구분할 수 있는 이심률, 면적, 압축률(compactness; 둘레^2/면적), 연장률(elongatedness; 면적/폭^2)을 특징으로 이용하였고, 은닉층이 1개인 다층 신경망을 이용하여 훈련하였다.

구균의 군체 중 가깝게 붙은 것은 개체 분할 단계에서 분리가 되지 않아 간균으로 오인식 되는 경우가 있다. 이러한 오인식을 바로잡기 위해 구균의 군체와 간균을 구분하는 인식기를 추가로 훈련하여 간균인지 구균인지 검증한다. 이 검증기는 압축률, 연장률, 그리고 불변 모멘트[7]를 특징으로 사용하여 훈련, 검증하였다.

각 개체별로 인식 결과를 얻은 후에, 이를 종합하여 관심 영역의 분류를 행하게 된다. 미생물의 분류는 관심 영역 단위로 수행하므로, 개체별로 분류 결과가 다를 경우 이를 종합하는 과정이 필요하다. 이를 위해서 관심 영역 내의 개체별 인식 결과를 0과 1사이의 실수로 정규화 하여 평균한 결과값을 이용하여 관심 영역을 분류한다.

미생물의 계수 역시 관심영역 단위로 하는데, 개체 분할 단계에서 분할된 개체 수를 기준으로 측정을 한다. 이때, 효모는 개체 분할 단계를 거치지 않으므로 효모의 경우에는 관심 영역의 면적을 계산하고 이를 미리 계산된 평균 면적으로 나누어 개체 수를 계산한다. 그리고 구균의 경우, 미생물이 개체별로 분리되지 않는 경우에 관심 영역의 면적을 평균 면적으로 나누어 개체 수를 계산한다.

4. 실 험

실험에 사용한 시스템은 펜티엄 III 600Mhz, 256M RAM의 IBM 호환 PC에 Microsoft Windows 2000 운영체제, Visual C++ 6.0 환경이었다. 실험에 사용한 데이터는 미생물 영상 63 장을 둘로 나누어 42 장은 인식기 훈련에 사용하였고, 21 장은 성능 측정에 사용하였다. 미생물 영상은 장당 30 개체 내외의 미생물을 포함한다.

실험 결과는 21장의 영상에 대해 각각 계측 결과를 측정한 후에 평균값을 보였다. 성능 확인을 위해 계측 결과를 실제 미생물 수로 나눈 비를 함께 보였다 (표 1).

표 1. 종류별 미생물 계수 결과

계측결과 평균	간균	구균	효모
분석 결과(R)	19.8	15.5	9.9
실제 미생물 수(N)	11.9	19.2	8.4
R / N	1.66	0.80	1.18
log (R/N)	0.22	-0.09	0.07

계측 결과(R)를 실제 미생물 수(N)로 나눈 비(R/N) 값을 살펴 보면 간균의 경우 1.66, 구균의 경우 0.80, 효모의 경우 1.18을 기록하고 있다. 일반적으로 미생물 농도 측정법에 요구되는 정확도는 log scale로 1의 범위 내의 정확도이다. 그에 따라 log(R/N)이 -0.5~0.5 사이에 있을 것이 요구되는데, 본 연구에서 제안하는 방법은 해당 정확도를 만족하므로 실용성이 있음을 알 수 있다.

5. 결 론

영상 처리 방법에 기반한 미생물 계수 시스템을 제안하였다. 기존의 방법에 비해 속도가 빠르고, 검사 반복을 통해 정확도를 높일 수 있다.

또한 제안하는 시스템의 성능을 실험으로 보였다. 영상 내에서 측정된 미생물의 수와 실제 미생물의 수의 비의 log 값이 -0.09~0.22 사이에서 나타났는데, 실용적인 방법이 될 수 있는 기준이 log 값으로 -0.5~0.5 사이에 있는 것이므로 본 시스템은 실용적이라고 볼 수 있다.

참고 문헌

[1] 김창남 외, 우리나라의 가공식품에 대한 미생물학적 위험도 평가, 한국식품위생학회지, 12(4), p.340, 1997

[2] Food and Drug Administration : Bacteriological Analytical Manual, 8th ed, AOAC International, 1995

[3] N. Otsu, "A threshold selection method from gray-level histograms," IEEE Trans. Systems, Man, and Cybernetics, 9(1), p.62, 1979

[4] R. Goldenberg, et al., "Fast geodesic active contour," IEEE Trans. Image Processing, 10(10), p.1467, 2001

[5] A. Garrido, et al., "Applying deformable templates for cell image segmentation," Pattern Recognition, 33, p.821, 2000

[6] O. D. Trier, et al., "Recognition of Digits in Hydrographic Maps: Binary vs. Topographic Analysis," IEEE Trans. PAMI, 19(4), p.399, 1997

[7] M. K. Hu, "Pattern recognition by moment invariants," Proc. IEEE, 49(9), p.1428, 1961