

RGB영상의 독립성분분석을 이용한 건조추영상 분류

권기현*, 임정대**

Dried pepper sorting using independent component analysis on RGB images

Kihyeon Kwon*, Jung-Dae Lim**

요약

고추는 건조과정에서 부패되거나 색이 변하는 경우가 발생하므로 건조추 품질을 높이기 위해서는 건조추를 선별할 수 있는 기법이 필요하다. 독립성분분석은 블라인드소스분리에서 가장 널리 사용되는 방법으로 이 기법을 사용하여 건조시킨 고추 영상에서 가장 중요한 성분에 대한 농축영상을 얻는다. 취득한 농축영상은 일반 이진(BW) 영상과 달리 주요 성분만 반영한 것으로 영상의 주요 성분 분포 상태를 알 수 있으며 품질을 판단하여 선별하는 것이 가능하다. 또한, 추출된 농축영상의 크기는 고추의 매운 맛을 내는 주요 성분인 캡사이신류의 양과 관련성이 있음을 알 수 있다. ICA 독립성분을 기반으로 한 농축영상 추출을 통해 고추 건조과정에서 부패되어 색상이 좋지 않거나 캡사이신류와 같은 주요 성분이 없게 된 고추를 선별해하는 방법을 제안한다.

▶ Keyword : 독립성분분석, 건조추, 농축 영상

Abstract

Hot pepper can be easily faded or discolored in drying process, so we need to use the sorting technique to improve the quality for dried hot pepper. Independent Component Analysis (ICA) is one of the most widely used methods for blind source separation. In this paper we use this technique to get a concentration image of the most important component which plays a role in the dried pepper. This concentration image is different from the binary image and it reflects the characteristics of major components, so that we know the distribution and quality of the component and how to sort the dried pepper. Also, the size of the concentration image can tell the relation with capsaicinoids which make hot taste. We propose a sorting method of the dried hot pepper that is faded or discolored and lacked a major component likes capsaicin in drying process using ICA concentration image.

▶ Keyword : Independent Component Analysis, dried pepper, concentration image

• 제1저자 : 권기현 • 교신저자 : 임정대

• 투고일 : 2011. 09. 27, 심사일 : 2011. 10. 22, 게재확정일 : 2011. 11. 26.

* 강원대학교 전자정보통신공학부(Dept. of Information & Communication Engineering, Kangwon National University)

** 강원대학교 생약자원개발학과(Dept. of Herbal Medicine Resource, Kangwon National University)

I. 서론

고추의 외관적인 색깔과 신미성분은 고추품질 평가기준으로서 중요하다. 고추 특유의 매운 맛을 내는 신미성분은 캡사이노이드(capsaicinoid)이며 이것은 캡사이신(capsaicin)과 디히드로캡사이신(dihydrocapsaicin) 등의 여러 가지 포화 및 불포화 아마이드 화합물의 혼합체인데, 이러한 성분들이 복합적으로 함유되어 매운 맛을 결정하는 것으로 알려져 있다[1][2].

캡사이노이드 함유량은 고추가 숙성됨에 따라 상당히 변하며 붉은 색이 되게 하는데 중요한 역할을 한다[1]. 고추가 익는 것은 엽록소가 분해되어 캡사이노이드를 형성하는 과정에서 색이 변하게 되는 것이다. Polder et al. [3]에서는 스펙트럼 영상과 HPLC(High-Pressure Liquid Chromatography)에 의해 측정된 측정치에 높은 상관관계가 있다는 것을 발견하였다. 그러나 스펙트럼 영상을 사용하기 위해서는 스펙트럼 장비의 칼라브레이션, 영상 스펙트럼 추출 등에 따른 시간 및 비용이 많이 발생한다.

Polder G.[4]의 연구와 여러 문헌의 대부분의 연구들은 감독에 의한 방법(supervised techniques)에 기반하고, 컬러브레이션 단계와 소팅 단계가 있으며 영상 획득과 영상을 농축 영상으로 변환하는 것이 필요하다. 이때 영상 획득 및 소팅매트릭스와 곱하는 작업 모두 병목이 될 수 있다. 비감독(unsupervised) 분류가 가능한 기법을 BSS(Blind Source Separation)이라고 한다. 가장 많이 사용되는 BSS 방법이 ICA(Independent Component Analysis) [5]이다.

본 논문에서는 임의 고추를 선택하여 HPLC 분석을 통해 색상과 손상 상태에 따른 주요 화합물 상태를 구하고 동일 컬러 영상에 대해 독립성분분석을 통해 농축영상(concentration image)을 구해 건조 상태가 용이한 고추 영상 선별이 가능함을 보인다.

본 논문의 구성은 2장에서 건조 고추의 HPLC 분석결과를 살펴보고, 3장에서 ICA에 대해 설명하며, 4장에서는 영상에 대해 농축 영상을 구하는 방법 및 결과를 설명하고 5장에서 결론을 맺는다.

II. 고추 샘플의 HPLC 분석

선택된 건조 고추 영상(그림 1)의 캡사이신류의 분포를 구하기 위해 다음과 같이 HPLC 분석을 수행하였다.

캡사이노이드 성분은 Ahn[6] 방법에 준하여 에탄올로 추출한 후 캡사이신과 디히드로캡사이신 함량을 LC-10AT HPLC 시스템(Shimadzu Instrument Co. LTD, Japan)을 사용하여 정량하였다. 즉 추출방법은 건조된 고추분말 1g를 달아서 에탄올 8ml를 넣은 다음 교반기를 이용하여 2시간 동안 추출하였으며 추출액은 원심분리 (5,000rpm, 10min)하여 상등액을 얻어 0.50 μ m syringe filter로 여과하여 시료를 준비하였다.

컬럼은 Waters사의 Atlantis[®] dC18 컬럼(5 μ m, 4.6 \times 250mm)을 사용하였고 이동상(mobile phase)은 메탄올과 물로 사용하였고 선형 그라디언트 시스템 방법을 적용하여 분석하였다. 시료의 주입부피(injection volume)는 20 μ l로 하였고 유량(flow rate)은 1.5 ml/min로 조정하였으며 검출기(detector)는 SPD-10A UV 검출기(detector)를 이용, 파장(wavelength)을 280 nm에서 고정하여 사용하였다.

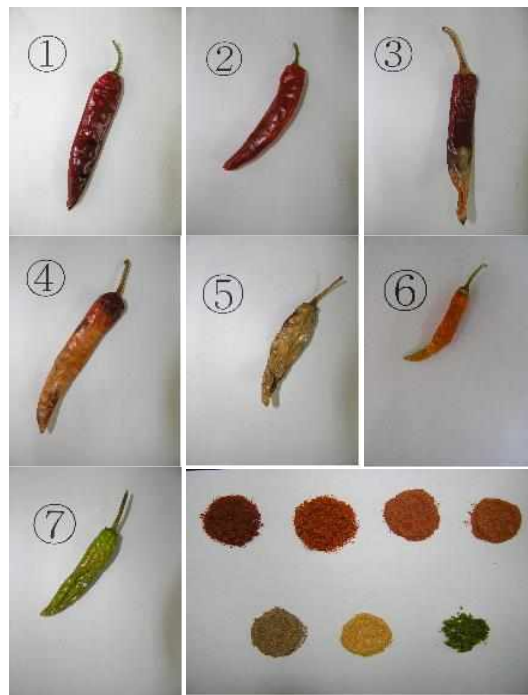


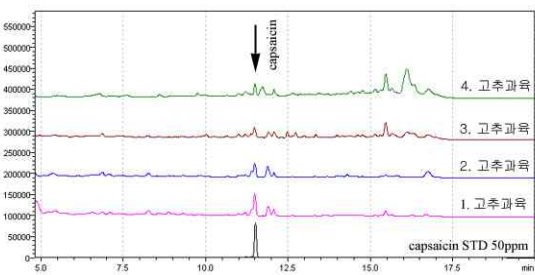
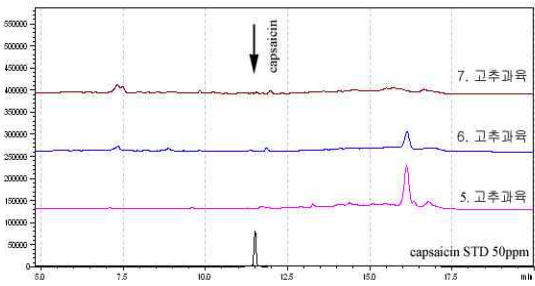
그림 1. 선택된 건조고추 및 시료샘플
Fig. 1. Selected Dried Hot Pepper and Test Sample

표 1. Capsaicin, dihydrocapsaicin 내용량
Table 1. Capsaicin & Dihydrocapsaicin of Dried Hot Pepper

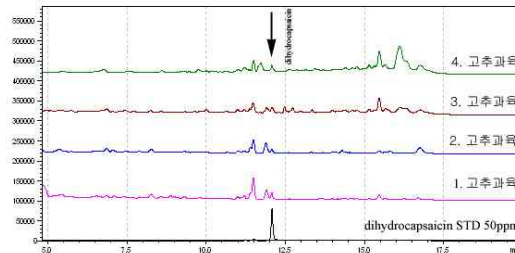
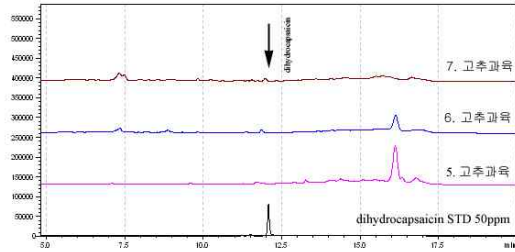
	capsaicin $\mu\text{g/g}$ FW	dihydrocapsaicin $\mu\text{g/g}$ FW
1. 고추과육	322.77 \pm 5.62	150.07 \pm 1.45
2. 고추과육	153.49 \pm 2.13	113.55 \pm 1.33
3. 고추과육	182.84 \pm 8.95	171.36 \pm 3.64
4. 고추과육	151.81 \pm 3.14	216.78 \pm 3.25
5. 고추과육	ND	ND
6. 고추과육	ND	Tr
7. 고추과육	34.14 \pm 1.23	ND

ND: not detected; Tr : trace amount

시료의 정량은 표준 화합물인 캡사이신 (8-methyl-N-vanillyl-6-nonenamide, Sigma)과 디히드로캡사이신 (8-methyl-N-vanillyl-nonenamide, Sigma)을 10, 50, 100, 500, 1000ppm으로 조제하여 검량선(calibration curve)을 구한 후 검량선을 작성하였으며 표준물질의 회귀곡선을 이용하여 면적을 y에 대입하여 농도(x)를 구한 다음 단위를 $\mu\text{g/g}$ 로 하여 함량을 정량하였다. [$R^2=0.9997503$ (캡사이신), $R^2=0.9996671$ (디히드로캡사이신)]



(a) capsaicin



(b) dihydrocapsaicin

그림 2. capsaicin과 dihydrocapsaicin의 HPLC 크로마토그램
Fig. 2. HPLC chromatogram of capsaicin & dihydrocapsaicin

선택된 건조 고추 영상(그림 1)의 캡사이신과 디히드로캡사이신 량은 표 1 및 그림 2와 같다. 캡사이신, 디히드로캡사이신 모두 고추과육 1~4에서 많이 나타나고 있으며, 고추과육 5~7에서는 거의 나타나고 있지 않다. 대체로, 건조 상태가 좋거나 양호한 경우 캡사이신과 디히드로캡사이신 량이 많은 것을 알 수 있으며, 건조 상태가 좋지 않은 고추과육 5~6은 거의 나타나지 않거나 매우 소량인 것을 알 수 있고, 익지 않은 고추과육 7에서는 캡사이신 만 소량 있다.

III. ICA 농축영상 계산

이 절에서는 ICA[5]의 개념에서 ICA 농축영상을 구하는 방법을 설명한다.

3.1. ICA 및 농축영상

노이즈프리 ICA모델은 Hyvärinen & Oja [5]와 Comon [7]의 논문에 기반하고 있다.

통계적으로 독립적인 소스($s_1, s_2, \dots, s_m, \dots, s_M$)가 선형시스템을 통해 혼합되어 관찰된 것이 N 개($x_1, x_2, \dots, x_n, \dots, x_N$) 있다고 가정하면 식(1)과 같이 표현할 수 있다.

$$x_n = a_{n1}s_1 + \dots + a_{nm}s_m + \dots + a_{nM}s_M \quad (1)$$

$$n = 1, 2, \dots, N$$

혼합치 x_n 과 소스 s_m 은 랜덤변수라는 가정 하에 혼합된 관찰치는 계수 A_{nm} 에 의해 특성이 부여된다. 행렬을 사용하면 편한데, $X = [x_1, x_2, \dots, x_N]^T$, $s = [s_1, s_2, \dots, s_M]^T$ 로 표현할 수 있으므로, 식(1)의 선형시스템은 식(2)이 된다.

$$X = AS \quad (2)$$

A 는 믹싱체계 A_{nm} 을 가지는 $N \times M$ 스칼라 행렬이다. 이 통계적인 모델을 노이즈프리 ICA모델이라고 한다. A 의 역행렬이 s 에 대한 추정치가 된다. 언믹싱(Unmixing) 행렬 W 는 식(3)과 같다.

$$u = Wx = WAs \approx s \quad (3)$$

u 는 S 의 추정치이고 W 는 A^{-1} 의 추정치이다. 이 행렬들을 사용하여 원본 소스의 값을 구하게 된다.

PCA(Principal Component Analysis)에서 주성분은 각 주성분의 분산에 따라 정렬되고, 해당 고유값을 통해 계산된다. 이것은 독립성분의 경우 가능하지 않은데, 독립성분의 분산 모두 알 수 없기 때문이다. 실제로, PCA는 ICA의 전처리 단계에서 자주 사용되어 성분의 정렬, 성분의수를 추정하는데 사용된다.

A 매트릭스의 크기는 X 매트릭스 픽셀 수에 비례한다. 혼란 영상의 모든 픽셀을 사용할 경우에, A 는 너무 커져서 실제로 응용이 불가능하다. 그러므로 임의로 선택된 픽셀의 서브셋(X'_{im})이 사용되어진다. A 와 S 를 측정된 다음 모든 농축 영상(A'_{im})을 식(4)에 의해 구할 수 있다.

$$A'_{im} = SX'_{im} \quad (4)$$

3.2. 전처리: 백색화 및 차원 감축

데이터에 ICA알고리즘을 적용하기 전에, 백색화(whitening) 처리를 하는 것이 필요하다. 이를 위해 X 의 평균값을 0평균으로 만든다($M = E\{\widehat{X}\}$). 이 전처리를 통해 ICA 알고리즘이 더 간단하게 된다. 데이터에서 혼합행렬 A 를 추정하는 다음에는, 우리는 S 의 평균을 S 의 중심추정치에 더하여 추정을 완성하게 된다. S 의 평균행렬은 $A^{-1}M$ 으로 주어진다.

백색화는 매트릭스 X 을 변환하여 잡음을 줄이는 기술로

PCA 처리로 백색화가 가능하며 데이터의 차원을 줄이는데도 사용된다.

3.3. FastICA

ICA 처리 방법은 Infomax, JADE, FastICA 등으로 다양하나 본 논문에서는 FastICA[8]을 사용하여 ICA 처리를 한다. 이 알고리즘은 ICA 처리에 계산 효율적인 알고리즘으로 전통적인 ICA 방법보다 10-100배 더 빠르다. 중요한 특징은 수렴성이 우수하고 독립성분이 순차로 구해지며, 계산이 간단하고 메모리소모도 적다.

IV. 구현 및 성능평가

4.1 구현 조건

건조 상태에 따라 덜 익은 것부터 익은 것까지 7개의 고추(그림 1)를 사용하였다. 카메라로 촬영 후 HPLC 분석을 우선 수행하였다. 실험에 사용된 영상은 Nikon D70 RGB카메라를 사용하였으며 1571×317 해상도의 영상을 취득 후 배경을 제거하고 축소하여 150×30 해상도의 영상을 구하였다.

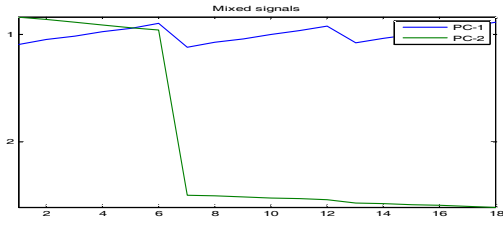
각 영상에서 임의로 선택한 100개의 픽셀이 관측변수 X 의 집합이 된다. 고추 모양에 따른 차이는 정규화를 통해 해결하였다[9]. 백색화작업을 위해 PCA가 사용되었으며 분산은 1로 하였다. FastICA 알고리즘을 사용하여, 혼합 매트릭스 A 와 독립성분이 계산되어진다. 이 알고리즘은 항상 가장 좋은 결과가 얻어지는 것을 보장하지는 않기 때문에 테스트를 위해 ICA 알고리즘을 동일한 X 매트릭스에 대해 100회 반복하였다.

X 의 빠른 처리를 위해 A 매트릭스는 가능한 작은 것이 좋다. ICA알고리즘의 처리속도는 데이터 크기에 비례한다. 나은 측정을 위해 얼마나 많은 샘플이 필요한지 알기 위해서, 하나의 고추 당 선택한 픽셀의 수를 변경해가면서 ICA 알고리즘을 반복하여 100개의 픽셀을 선택했을 때 처리 속도와 정확도가 만족함을 보였다.

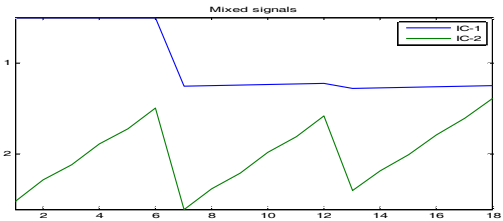
ICA가 독립성분을 뽑아내는데 사용되므로, 측정된 성분은 실제 고추의 성분과 관련된 것임을 기대할 수 있다.

4.2 주성분 및 독립성분 비교

선택된 건조 고추 영상(그림 1)에서 PCA를 통해 평균 99.24%의 고유치를 가지는 첫 두 주성분을 구하였다. 이것은 2개의 독립성분이 발견 될 수 있다는 것을 의미하는 것이다.



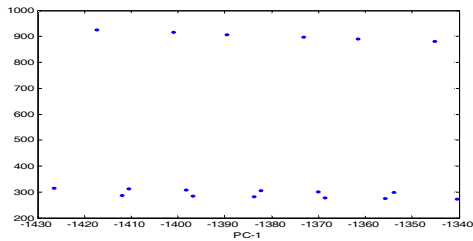
(a) PCA Plot



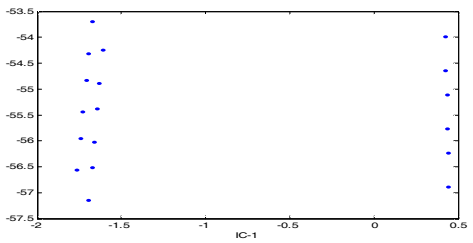
(b) ICA Plot

그림 3. PCA & ICA Plot
Fig. 3. PCA & ICA Plot

ICA 독립성분을 구하기 위해 FastICA 알고리즘을 사용하여, 150*30 픽셀 영상에서 두개의 독립성분(IC-1, IC-2)을 구하였다. 그림 3은 PCA와 ICA에서 구한 두 개의 주성분과 독립성분이다. PCA 주성분과 ICA 독립성분의 패턴이 유사하나, ICA 독립성분의 분리가 나은 것으로 알 수 있다(그림 3).



(b) PCA Scatter Plot



(b) ICA Scatter Plot

그림 4. PCA & ICA Scatter Plot
Fig. 4. PCA & ICA Scatter Plot

주성분과 독립성분을 평면에 투영하여 살펴보면 ICA 독립성분이 성분의 방향과 보다 높은 관련성을 보이고 있음을 알 수 있다(그림 4).



그림 5. 고추과육의 원본영상, 이진영상, 독립성분영상(BW), 독립성분영상(RGB)

Fig. 5. Original Image, Binary Image(BW), IC Image(BW) and IC Image(RGB) of Dried Hot Pepper

더 많은 독립성분을 구하려 했으나, ICA 알고리즘은 더 이상 커버하지 못했다. HPLC 분석에서도 캡사이신과 디히드로캡사이신이 건조추에서 보이는 가장 높은 화합물이라는 것을 알 수 있다(그림 6).

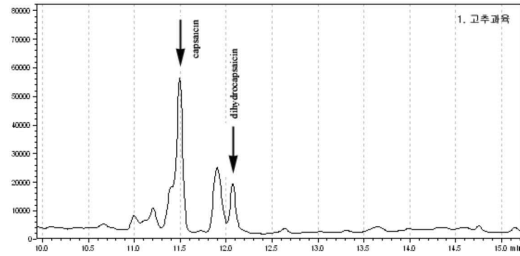


그림 6. 고추과육1의 capsaicin과 dihydrocapsaicin 내용량
Fig. 6. Capsaicin & Dihydrocapsaicin of Dried Hot Pepper1

4.3 독립성분 농축영상

선택된 건조 고추 영상(그림 1)에 식 4에 의거 독립성분 S를 원본 영상에 곱하여 독립성분 농축영상을 얻을 수 있다.

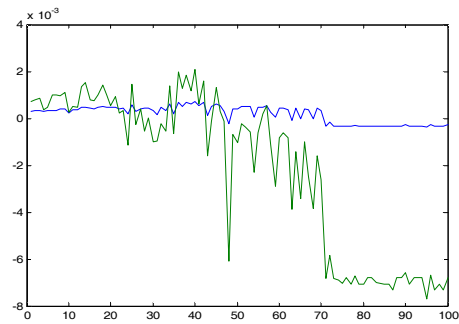


그림 7. 독립성분 가중치
Fig. 7. The Weight of IC

이 영상은 각 픽셀이 가지는 RGB 값의 한계로 독립성분

1에 대해서만, 흑백 영상으로 나타나며, 독립성분 2는 값이 미미하여 영상으로 표현하는 것이 어렵다. 그림 7은 독립성분 1, 2에 대한 가중치로 독립성분 2의 가중치의 경우 거의 0에 가까운 것으로 알 수 있다.

표 2 독립성분 농축영상 면적
Table 2. The Area of IC Concentration Image

	면적(%)
1. 고추과육	84.1602
2. 고추과육	76.7470
3. 고추과육	73.7149
4. 고추과육	74.9234
5. 고추과육	5.5974
6. 고추과육	2.4024
7. 고추과육	1.0000

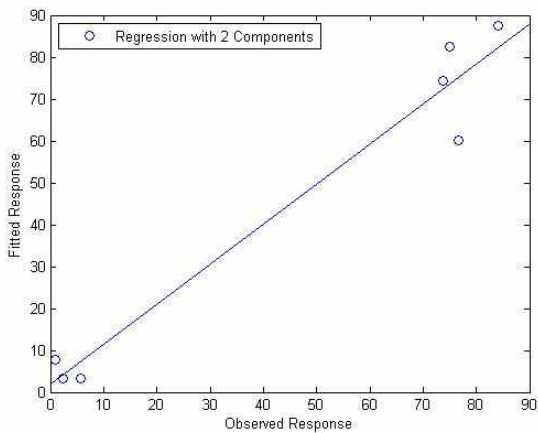


그림 8. 독립성분 농축영상, HPLC 화합물 회귀분석($R^2=0.9585$)
Fig. 8. Regression Analysis of IC Concentration Image and HPLC Compound($R^2=0.9585$)

그림 5는 선택된 건조 고추 영상(a)에 대해 우선 이진영상(b)을 구하고, 독립성분 1에 대한 농축영상을 만든 다음(c), 독립성분 농축영상을 RGB 영상(d)으로 변환시킨 것이다. 그림 5에서 ①, ②는 정상적인 상태로 독립성분 농축영상이 잘 표현되어 있다. ③, ④는 부분적인 손상이 있는 경우이며 ⑤, ⑥, ⑦은 부패했거나 덜 익은 경우로 독립성분 농축영상을 구하기 어려운 경우이다.

독립성분 농축영상의 면적(표 2)과 HPLC로 구한 화합물(표 1)간의 관계를 회귀분석을 통해 살펴보면 그림 8과 같이 나타나며 $R^2=0.9585$ 임을 알 수 있다. 즉, 독립성분 농축영상의 면적이 넓은 고추 과육과 캡사이신류의 양이 서로 비례한

다는 것을 알 수 있다. 이를 통해 2차원 독립성분 농축영상에 관한 실시간 영상 분석은 건조 고추의 상태를 판별하는데 적용될 수 있다.

V. 결론

ICA는 비감독 시스템에 적합한 기법으로 RGB 영상을 사용한 건조추의 실시간 선별에 사용될 수 있다. 구해진 독립성분과 실제 고추 화합물의 값을 연계시키기 위해서는 초기 캘리브레이션이 여전히 필요하며 광원, 센서 드리프트, 원산지에 따른 다양성 등을 고려하여야 한다. ICA 특성상 잘못 측정된 독립성분이 결정될 수 있기 때문에, 실제 응용에서는 분석 단계에서 이들을 배제하는 것이 필요하다.

감독(supervised) 기반 시스템에 비해 이 시스템의 큰 장점은 캘리브레이션에 참조 데이터를 적게 사용하여 시스템을 쉽고, 빠르고, 싸게 만들 수 있다는 점이다. 건조추 영상당 100개의 작은 수의 픽셀로도 잘 동작하는 것을 보였고, 이는 FastICA 알고리즘의 속도를 증가시킬 수 있었다. 픽셀을 직접 추출할 수 있는 CMOS 카메라를 사용하면 데이터 취득 속도는 실질적으로 증가시킬 수 있을 것이다. 제안된 방법은 실질적인 품질 선별 시스템에서 구현되어 건조추 선별을 위해 사용될 수 있다.

참고문헌

- [1] K. Iwai, T. Suzuki, H. Fujiwake and S. Oka, "Simultaneous microdetermination of capsaicin and its four analogues by using high-performance liquid chromatography and gas chromatography-mass spectrometry," Journal of chromatography, v.172 no.1, pp.303-311, April 1979.
- [2] A. Muller-Stock, RK. Joshi and J. Buchi, "Study of the components of capsaicin. Quantitative gas chromatographic determination of individual homologs and analogs of capsaicin in mixtures from a natural source and of vanillyl pelargonic amide as adulteration," Journal of chromatography v.63, pp.281-287, 1971.
- [3] G. Polder, "Tomato sorting using independent

component analysis on spectral images”, Real-time imaging v.9 no.4 , pp.253-259, August 2003.

[4] G. Polder, van der Heijden GWAM, van der Voet H, Young IT. "Measuring surface distribution of compounds in tomatoes using imaging spectrometry", Postharvest biology and technology v.34 no.2, pp.117-129 , Nov. 2004.

[5] A. Hyvarinen, E. Oja, "Independent component analysis: algorithms and applications", Neural Networks v.13, pp.411 -30, 2000.

[6] HW. Ahn, "Studies on component analysis related to quality of Korean red pepper", MS Thesis. Kyung-Hee University, 1997

[7] P. Comon, "Independent component analysis, a new concept", Signal Processing v.36, pp.287 - 314, April 1994.

[8] A. Hyvarinen, "Fast and robust fixed-point algorithms for independent component analysis", IEEE transactions on neural networks v.10 no.3, pp.626-634 , 1999.

[9] G. Polder, GWAM Heijden, IT. Young, "Spectral image analysis for measuring ripeness of tomatoes", Transactions of the ASAE v.45 no.4, pp.1155-1162, 2002.

[10] GY. Heo, JS. Seo, IG. Lee, "A Non-linear Variant of Improved Robust Fuzzy PCA". Journal of the Korea society of computer and information, v.16 no.4, pp.15-22, 2011.

저자 소개



권 기 현
 1993 : 강원대학교 전자계산학과 이학사.
 1995 : 강원대학교 전자계산학과 이학석사.
 2000 : 강원대학교 컴퓨터과학과 이학박사
 현 재 : 강원대학교 전자정보통신공학부 교수
 관심분야 : 임베디드시스템, 패턴인식
 Email : kweon@kangwon.ac.kr



임 정 대
 1998 : 강원대학교 자원식물학과 농학사.
 2000 : 강원대학교 자원식물학과 농학석사.
 2004 : 강원대학교 생물공학과 농학박사
 현 재 : 강원대학교 생약자원개발학과 교수
 관심분야 : 약용식물대사공학, 생리활성평가
 Email : ijdae@kangwon.ac.kr

