The Journal of the Convergence on Culture Technology (JCCT)
Vol. 8, No. 1, pp.201-207, January 31, 2022. pISSN 2384-0358, eISSN 2384-0366

http://dx.doi.org/10.17703/JCCT.2022.8.1.201

JCCT 2022-1-22

선형회귀를 이용한 사과의 색상과 당도 분석

Analysis of Apple Colors and Sugar Contents Using Linear Regression

김선종*

Seon-Jong Kim*

요 약 본 논문에서는 같은 지역에서 수확된 사과 영상에 대해 선형회귀를 이용하여 RGB, HSV, La*b* 색상과 당도와의 연관 관계를 분석하였다. 먼저, 각 색상 레벨에 따른 당도와의 상관계수를 조사한 결과, 색상 레벨에 따라 양의 계수를 갖는 (+) 영역과 음의 계수를 갖는 (-) 영역으로 구분됨을 알 수 있었다. 또한 평균값으로 대표되는 색상과 당도와의 상관계수는 La*b* 색상 공간에서 0.342로 RGB, hsv 공간에서의 계수보다 높게 나타났다. 즉, 이는 당도는 La*b* 공간에서의 색상과 관계가 있다는 것을 의미하고 있다. 또한 당도와 관련이 높은 영역으로 구성된 복합 색상에서는 R^2=0.3627로 나타났으며, 이는 당도와 관련이 있음을 보여주고 있다. 9개 모든 색상 공간에서는 R^2=0.3668로 나타났다. 이 경우 v와 b*의 계수가 당도에 영향이 있음을 알 수 있었다. 이로 보아 노란색을 대변하는 b*가 높을수록 당도도 높게 나타난다는 경험적인 예측의 타당성을 확인할 수 있었다.

주요어 : 선형회귀, 색상-당도, 색상 공간, 사과영상

Abstract In this paper, the relationship between RGB, HSV, La*b* colors and sugar content was analyzed using linear regression on apples harvested in the same region. First, as a result of examining the correlation coefficient with sugar content according to each color level, it was found that the (+) region having a positive coefficient and a (-) region having a negative coefficient were separated according to the color level. Also, the correlation coefficient between color and sugar content, represented by the average value, was 0.342 in the La*b* color space, which was higher than the coefficient in the RGB and hsv space. That is, this means that the sugar content is related to the color in the La*b* space. Also, in the complex color composed of regions with high sugar content, it was found to be R^2=0.3627, indicating that it is related to sugar content. In all nine color spaces, it was found to be R^2=0.3668. In this case, it was found that the coefficients of v and b* had an effect on the sugar content. Due to this, it was possible to confirm the validity of the empirical prediction that the higher the b* representing yellow, the higher the sugar content.

Key words : Linear Regression, Color-Sugar Contents, Color Spaces, Apple Images

Ⅰ. 서 론

과일의 껍질은 각자의 고유한 색상을 가지고 있으며, 내부 특성과도 관련이 있으며, 과일 내부는 주로 신맛 또는 달콤한 맛으로 구성된 요소들로 구성되어 져 있다. 따라서 과일의 껍질 색상은 품질이나 특성을 표현하는 중요한 요인 중 하나이다 [1]. 특히, 사과는 우리나라 모든 지역에서 재배되고 있으며, 각 지역의 날씨와

*정회원, 부산대학교 IT응용공학과 교수 (제1저자)
접수일: 2021년 12월 27일, 수정완료일: 2022년 1월 1일
게재확정일: 2022년 1월 8일

Received: December 27, 2021 / Revised: January 1, 2022
Accepted: January 8, 2022
*Corresponding Author: ksj329@pusan.ac.kr
Dept. of Applied IT Eng., Pusan National Univ, Korea

환경에 따라 다양한 형태와 색상을 가지고 있다. 사과의 맛은 당도가 아주 큰 비중을 차지하고 있다. 당도는 주로 적외선의 회절 성질을 이용하여 측정하고 있으며, 경험적으로 껍질의 색상 분포와 관련이 있는 것으로 알려져 있다 [2, 3].

최근 카메라의 성능이 우수해지고, 머신 비전 기술이 발전됨에 따라, 컴퓨터 비전 기반의 기술들이 농업이나 식품 산업 등에 많이 응용되고 있다. 또한 이 기술들은 농업 산업의 고품질을 위한 생산이나 분류과정에 중요한 역할을 하고 있다 [4, 5]. 영상 처리 기술은 수확물의 품질을 분석하는 등 다양한 분야에 도입되고 있으며, 컴퓨터 비전 기술은 과일의 품질 검사, 문자 해석을 통해 인식, 품질 등급 및 분류에 대한 정보를 제공하고 있다 [6-8].

색상 센서를 이용하여 수확시기를 예측하거나 품질이나 등급을 측정하는 기술들은 사과의 생산과정 중에 적용될 수 있다. 오 등은 스마트팜에서 사용 가능한 비파괴 당도 측정, 색상 및 크기 센서로부터 다량의 데이터들을 수집 및 분석하여 인자들 간의 관계로부터 수확시기를 예측하는 시스템을 제안하였다 [9]. Lu 등은 근적외선을 이용하여 체리의 단단한 정도와 당도를 예측하는 등 색상을 이용하여 다양한 특성에 대해 조사 분석하고 있다 [10].

본 논문에서는 사과의 RGB, HSV, La*b* 색상 공간과 당도와의 연관 관계를 선형회귀 방법에 의해 분석하고자 한다. 각 색상 레벨과 당도와의 관계, 각 색상 공간에 따른 선형회귀 분석을 통해 당도와 색상간의 관계를 조사한다. 특히, 사과의 색상이 노란색을 띨 때에 당도가 높다는 경험적인 판단에 대한 정량적인 분석을 한다.

II. 선형회귀 분석과 상관계수

1. 데이터 수집

사과는 경험적으로 껍질에 나타나는 색상과 내부의 당도가 연관이 있는 것으로 알려져 있다. 특히, 사과의 배 영역의 색상이 관련이 있음이 확인되었다 [11]. 수집된 영상은 한 지역에서 출하된 사과를 카메라를 통해 얻었다. 그리고 수집된 사과는 믹서기로 갈아서 즙으로 만들고, 즙을 통해 당도를 측정하였다. 본 논문에서는 카메라 영상과 측정된 당도와의 관계를 분석하고자 한다. 그림 1은 사용된 사과의 배 부분의 영상이다. 색상 정보만을 얻기 위하여, 영상 (a)는 외부와 내부를 없애는 영상 편집 과정을 수행하여 (b)를 얻을 수 있다. 이는 배꼽 중앙의 색상이 당도와 무관하다고 판단되기 때문이다. 사과 영상은 120개를 사용하였다.

(a) (b)

그림 1. 사용된 사과의 (a) 배 영상과 (b) 데이터
Figure 1. The used partial image of an apple (a) original and (b) target image

그림 2는 데이터 셋의 당도 분포를 보여주고 있다. 당도는 12-19 Brix 사이에 분포되었다. 당도의 분포가 15 근처인 영상이 30개로 가장 많았으며, 최저 12부터 최고 19인 당도로 분포되어 있다. 색상은 24bit RGB로 표현되며, 각 색상은 8bit로 256 레벨을 가지도록 하였다.

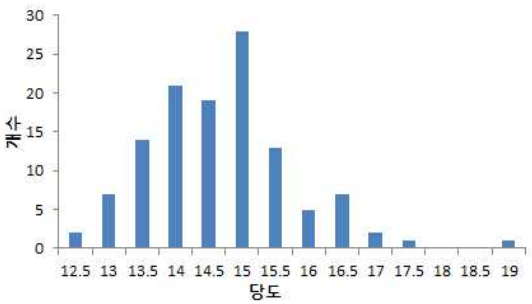
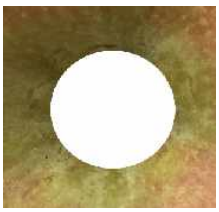


그림 2. 당도 분포
Figure 2. The distribution of sugar contents.

2. 색상 정보 추출

영상의 색상 공간은 일반적으로 RGB 뿐만 아니라 hsv, La*b* 등이 색상 공간도 사용된다. 각 색상 공간들은 사람의 눈의 특성을 이용한 색상 공간이며, 서로 다른 색상 공간과의 변환을 쉽게 할 수 있다. 그림 3은 당도 14.6을 가지는 사과 영상에 대한 R, G, B 공간에서의 히스토그램을 각각 보여주고 있다. 모든 색상 공간은 256레벨을 갖는다. 그림에서 알 수 있듯이 사과 영상

은 B 영역보다는 R이나 G의 영역이 넓은 분포를 하며, B의 레벨이 R보다 낮아 붉은색이나 녹색을 띠게 된다.



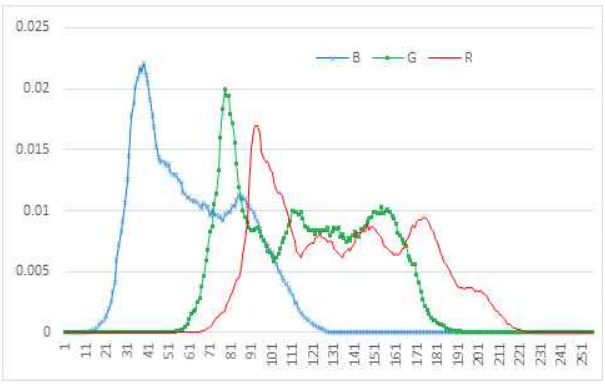


그림 3. RGB 색상 분포
Figure 3. The histograms for RGB colors

그림 4는 데이터셋 영상에 대한 R, G, B 히스토그램의 평균적인 특성을 보여주고 있다. R 값은 70 ~ 240레벨 사이로 분포되었으며, G는 30 ~ 220레벨, 그리고 B는 20 ~ 160레벨로 분포되는 것을 확인할 수 있다. B의 영역보다는 G와 R의 영역이 넓게 분포하는 것을 알 수 있다. 특히, R 색상의 분포가 높게 나타나서 수확된 사과는 주로 붉은 색상으로 나타난다는 것을 알 수 있다. 데이터 셋의 사과 색상은 전체적으로 비슷한 색상을 가지는 것처럼 보이지만 분포하는 색상 히스토그램은 조금씩 다르다.

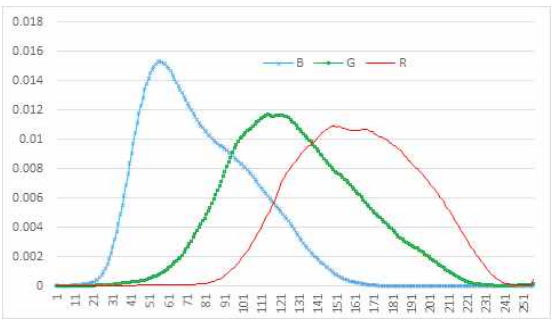


그림 4. 모든 영상에 대한 색상 분포
Figure 4. The histograms of colors for all used images

주어진 영상의 히스토그램은 256레벨로 나타나고, 모든 레벨에 대한 색상 분석은 복잡할 수 있다. 따라서 본 논문에서는 각 색상에 대한 평균값을 가지고 분석하고자 한다. 이때, 평균값은 각 색상에 대한 자세한 정보가 사라질 수 있다. 따라서 이 방법은 정확한 분석보다는 전체적인 경향을 조사하는 데 유용하게 사용될 수 있다.

그리고 색상 공간은 RGB, hsv, La^*b^* 색상 공간을 사용한다. 이 색상 공간은 인간이 인지하는 색상 공간과 근접시키고자 하는 특성이 있으며, 이를 통해 인간의 경험적인 색상과 당도와의 관계를 유추할 수 있기 때문이다. 본 논문에서는 각 색상 공간에 대해 당도와의 연관 관계뿐만 아니라 최적의 색상 공간 조합의 경우에도 분석한다.

3. 선형 회귀분석

회귀분석은 독립변수와 종속변수의 인과적 관계를 파악할 때 사용되는 통계적 분석 방법의 하나이다. 만약 독립변수와 종속변수가 인과관계가 있다면 선형회귀 결과는 어떤 형태의 인과관계인지 보여준다.

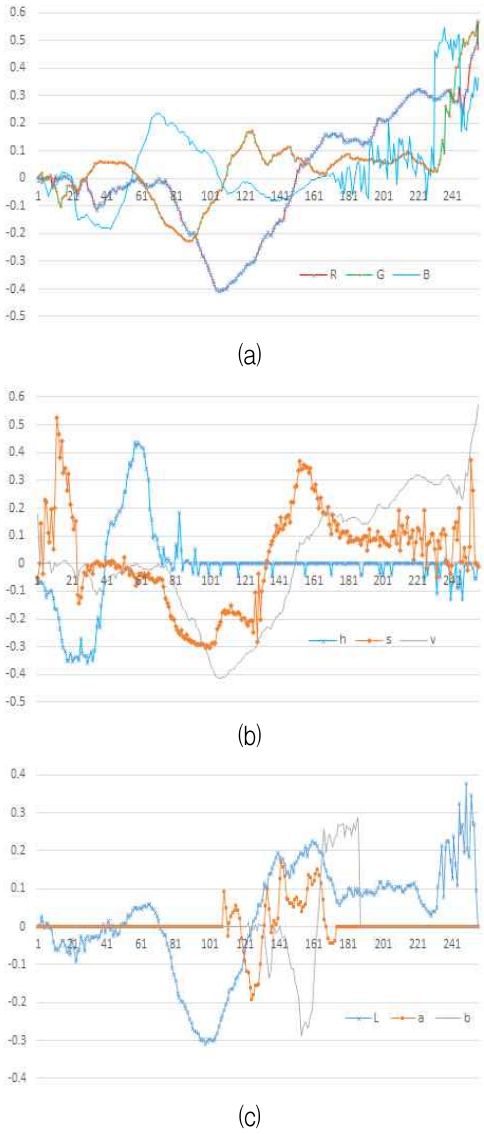
또한 회귀분석은 독립변수, 종속변수, 회귀계수로 단순화시킬 수 있다. 회귀계수는 독립변수와 종속변수의 관계를 표준화시켜서 보여주는 수치이다. 즉, 선형 함수에서는 기울기를 의미하는 수치이다. 회귀계수는 다양한 방식에 의해 얻을 수 있지만, 주로 최소 제곱법에 따라 산출되고 있다. 최소 제곱법은 어떤 방정식의 해를 근사적으로 구하는 방법으로, 근사해와 실제 해의 차의 제곱의 합을 최소로 만드는 방법이다.

4. 상관계수, R^2

회귀계수로 주로 사용되는 상관계수는 두 변수 간의 상관관계가 얼마인지를 나타내는 지수이다. 본 논문에서는 일반적으로 사용되는 피어슨 상관계수를 이용한다. 피어슨 상관계수는 공분산과 두 변수 간의 표준편차를 이용하며, 두 변수의 선형적 관계를 표현할 때 유용하다. 계수 값은 -1 에서 1 사이의 값을 가지며, +1은 양의 비례한 상관관계를, -1은 음의 상관관계를 가진다. 양의 값이면 두 변수 간의 분포가 우상향으로 나타나며, 음의 값이면 두 변수 간의 분포가 우하향하는 특성으로 나타난다.

III. 실험 및 결과 고찰

데이터 셋으로부터 주어진 색상 공간을 이용해 평균값을 얻고, 이를 통해 색상과 당도와의 관계를 조사하려고 한다. 먼저 RGB, hsv 및 lab 공간에서 각 색상 레벨과 당도와의 어떤 관계가 있는지 이에 대한 상관계수를 조사하였다. 그 결과를 그림 5에 도시하였다. (a)는 RGB 공간에서의 색상 레벨과 당도와의 상관계수를 나타낸 것이다. R인 경우, 레벨 58에서 108 사이에는 상관계수가 (-) 영역으로 나타났다. 즉, 빨간색의 레벨에 따라 당도와 (+) 영역 또는 (-) 영역으로 나타났다. (+) 영역은 이 색상과 당도는 양의 계수로 선형관계가 있음을, (-) 영역은 음의 관계가 있음을 각각 알려주고 있다. 물론 G와 B 색상 레벨도 비슷한 특성이 있다. 상관계수가 (+) 영역과 (-) 영역의 분포가 레벨에 따라 다른 것을 알 수 있다. (b)는 hsv 공간에 대한 색상 레벨과 당도와의 상관계수를 보여주고 있다. 그림에서 보면 h의 레벨이 가장 확실하게 나타나는 것을 알 수 있다. 즉, 레벨 41을 기준으로 음의 영역과 양의 영역이 구분된다는 것을 알 수 있다. 또한 (c)는 La*b* 공간에 대한 각 색상 레벨과 당도의 상관계수를 보여주고 있다. b* 의 레벨이 160을 기준으로 두 영역이 차이가 난다는 것을 알 수 있다. 이상과 같이 각 색상 레벨에 따라 당도와 상관계수는 다르게 나타나는 것을 보여주고 있다.



(a) (b) (c)

그림 5. (a) RGB, (b) hsv 및 (c) La*b* 공간에서 색상 레벨에 따른 당도의 상관 계수

Figure 5. The correlation coefficient with sugars and values in the (a) RGB, (b) hsv, and (c) La*b* space

표 1은 그림 5로부터 각 색상 공간의 색상 레벨과 당도와의 상관계수를 이용해 특정한 영역을 얻은 결과를 보여주고 있다. 상관계수가 0.1보다 큰 영역 또는 -0.1보다 작은 영역에 대해 모든 색상에 대한 당도와의 상관계수를 조사한 결과이다. 즉, RGB 공간에서 R > 0.1인 영역에 대한 상관계수는 0.356이다. 또한 v < -0.1인 영역과의 상관계수는 -0.371로, 이 영역에 대한 색상이 클수록 당도는 떨어진다는 것을 알 수 있다.

두 번째 실험에서는 각 색상 공간의 평균값과 당도와의 관계를 조사하였다. 그림 6은 각 색상 공간에서의 평균값과 당도와의 분포를 도시한 것이다. 그림 6(a)는 RGB 공간에 대한 분포로서, R 색상의 평균값은 당도와 R^2=0.1259의 관계가 있음을 알 수 있다.

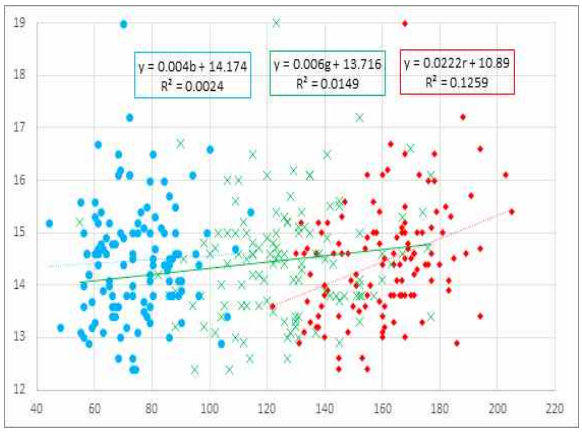
그림 6(b)는 hsv 공간에서 색상 평균값과 당도를 표시한 것이다. 그림에서 v는 당도와 R^2=0.1286 연관이 있음을 알 수 있다. 그리고 (c)는 La*b* 공간에 대한 분포

표 1. 상관계수가 높은 영역에 의한 상관계수 R^2

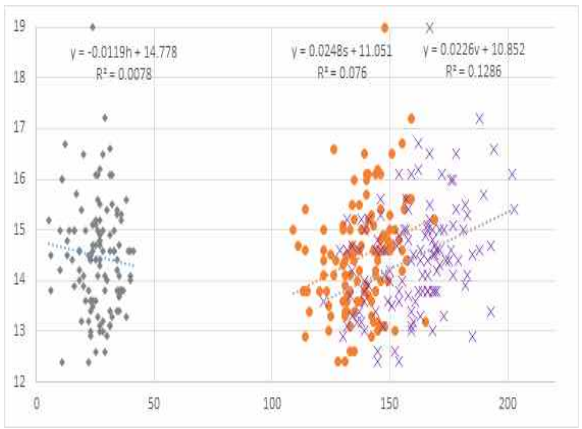
Table 1. R^2 using by the highly related region of the correlation coefficient

space	Region	R^2
RGB	R > 0.1	0.356
	R < -0.1	-0.370
	B > 0.1	0.281
hsv	S > 0.1	0.314
	V > 0.1	0.356
	V < -0.1	-0.371
La*b*	L < -0.1	-0.278
	b > 0.1	0.323
	b < -0.1	-0.321

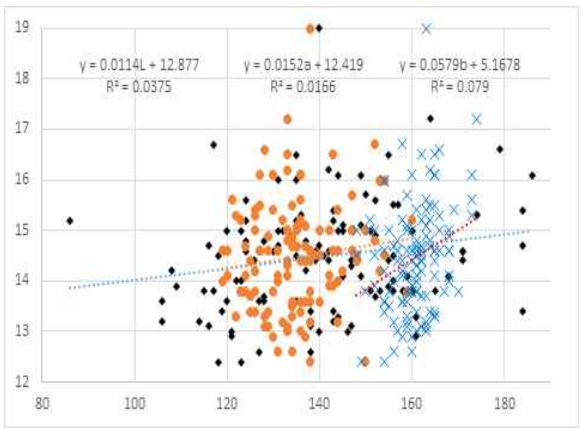
로서, 노란색을 의미하는 b*는 R^2=0.079로, 큰 의미가 있지는 않다는 것을 알 수 있다. 이처럼 각 색상과 당도와의 상관계수는 R과 v에서 약간 의미 있는 상관계수가 나왔다.



(a)



(b)



(c)

그림 6. (a) RGB, (b) hsv 및 (c) La*b* 공간에서 평균값과 당도의 분포
Figure 6. The distribution of the sugar contents in the (a) RGB, (b) hsv, and (c) La*b* space.

표 2. 각 색상 공간에 따른 선형회귀(LR) 계수
Table 2. LR coefficients according to the color space

Space	Sugar contents	R^2
RGB	0.0654R+0.0016G+0.0060B	0.332
hsv	0.0012h+0.0435s+0.0915v	0.327
La*b*	-0.0679L+0.0341a*+0.1631b*	0.342

세 번째 실험에서는 색상 공간과 당도와의 관계를 선형 회귀 분석에 따라 조사하였다. 표 2는 선형 회귀 분석에 따른 평균 색상 값과 당도의 관계를 조사한 결과이다. RGB 공간에서 0.332의 상관계수를 가졌으며, hsv와 La*b* 공간에서는 각각 0.327, 0.342의 계수로 각각 나타났다. 이로 보아 La*b* 색상 공간에서 가장 높은 상관계수로 나타나는 보아 당도를 예측할 때는 이 색상 공간을 사용하는 것이 유리함을 알 수 있다.

다음은 여러 색상 공간의 조합을 이용하여 상관계수를 조사하였다. 그림 7은 전술한 각 색상 레벨에서 상관계수가 높은 영역의 조합을 이용하여 당도와의 상관계수를 보여주고 있다. 이때, 사용한 색상 영역은 각각 S > 0.1, V > 0.1, R > 0.1 그리고 b > 0.1인 4개의 영역으로 한정시켰다. 그 결과 R^2=0.3627의 결과를 얻을 수 있었다.

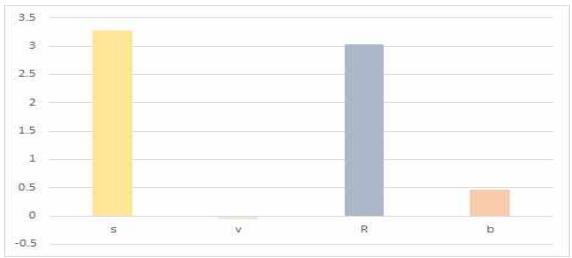


그림 7. 선택된 영역에 대한 LR 계수
Figure 7. LR coefficients when the selected colors used

마지막으로 색상 공간 RGBhsvLa*b*의 9개 색상과 당도와의 상관계수를 알아보았다. 각 색상에 대한 계수는 그림 8에 도시하였다. 그림에서 보면, 당도에 미치는 색상은 v와 b로 +영향으로 나타났다. 즉, v는 흑백 레벨을 의미하며, 흰색(255)에 가까울수록 당도도 높이 나타날 가능성이 있다는 것을 알려주고 있다. 그리고 얼마나 붉은색에 가까운가를 나타내는 a*는 당도와 관계가 약하게 나타났지만, 얼마나 노란색에 가까운가를 나타내는 b*는 당도에 큰 영향으로 나타난다는 것을 알 수 있다. 대표적으로 R=-0.1390, v=0.1652, b*=0.1648로 나타났다.

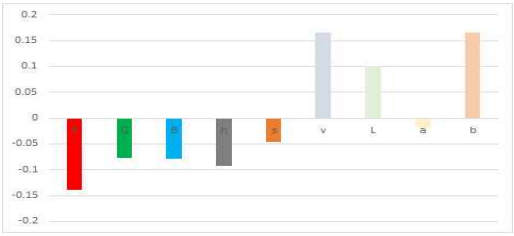


그림 8. 색상 9개를 사용한 경우의 LR 계수
Figure 8. LR coefficients when 9 parameters used

그림 9는 선형회귀 분석에 의한 예측된 당도와 실제 당도를 비교하여 도시한 결과이다. 이때, R^2=0.3668의 관계로 나타났다.

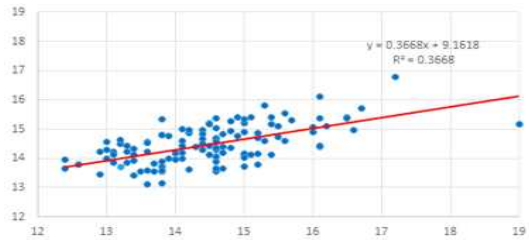


그림 9. 당도 예측선
Figure 9. The predicted line of the sugar contents

Ⅳ. 결 론

본 논문에서는 사과 영상에 대해 RGB, HSV, La*b* 공간에 대해 색상과 당도와의 관계를 분석하였다. 먼저, 각 색상과 당도는 레벨에 따라 양의 계수를 가지는 (+) 영역과 음의 계수를 가지는 (-) 영역이 분포하였다. 그리고 평균값을 이용한 회귀분석에서는 La*b* 공간과의 상관계수 R^2=0.342로 RGB, hsv 공간에서의 계수보다 높게 나타났다. 특히, 관련이 높은 레벨 색상으로 구성된 조합된 영역에서는 R^2=0.3627을 얻을 수 있었고, 9개 모든 색상을 사용할 때는 R^2=0.3668을 얻을 수 있었다. 이로 보아 사과의 색상은 당도와 어느정도 관련이 있다는 것을 알 수 있었다. 특히, 노란색의 특성을 나타내는 b*의 값이 클수록 당도도 높게 나타날 것이라는 경험적인 예측에 대한 타당성을 알 수 있었다. 이처럼 하나의 대표 색상과 당도와의 관계를 분석하였지만, 앞으로 확장된 색상을 이용하여 당도와의 관련성을 정밀하게 조사할 필요가 있다. 또한 사과의 당도는 색상뿐만 아니라 질감과도 연관이 있을 수 있어 이에 관한 연구가 계속되어야 한다.

References

[1] Bhargava A. and Bansal A. “Fruits and vegetables quality evaluation using computer vision: A review,” Journal of King Saud University – Computer and Information Sciences (2018). https://doi.org/10.1016/j.jksuci.2018.06.002.

[2] Hayrettin Toylan1 and Hilmi Kuscu, “A Real-Time Apple Grading System Using Multicolor Space,” Hindawi Publishing Corporation The Scientific World Journal Volume 2014. https://doi.org/10.1155/2014/292681

[3] G. C. Bora, R. Pathak, M. Ahmadi, and P. Mistry, “Image processing analysis to track colour changes on apple and correlate to moisture content in drying stages,” Food Quality and Safety, 2018, 2, 105 - 110. https://doi.org/10.1093/fqsafe/fyy003

[4] Z. A. Kus, B. Demir, I. Eski, F. Gurbuz, and S. Ercisli, “Estimation of the Colour Properties of Apples Varieties Using Neural Network,” Erwerbs-Obstbau (2017) 59:291 - 299. DOI:10.1007/s10341-017-0324-z

[5] S. N. Subhashree, S. Sunoj, J. Xue, and G. C. Bora, “Quantification of browning in apples using colour and textural features by image analysis,” Food Quality and Safety, 2017, 00, pp. 1 - 6. https://doi.org/10.1093/fqsafe/fyx021

[6] X. Cheng, T. Pang, H. Tao, M. Lin, and H. Yang “Prediction of apple internal qualities using hyperspectral imaging techniques,” Int. Conf. Adv. Mechatronics Syst., 450 - 455, 2017, DOI:10.1109/ICAMechS.2017.8316516

[7] J.-H. Lee, J.-H. Park, and S.-J. Kim, "A New Similarity Measurer for Color-Texture and Its Clustering for Apple," International Journal of Signal Processing Systems, Vol. 8, No. 3, pp. 59-63, September 2020. DOI:10.18178/ijsps.8.3.59-63.

[8] Soo-Mok Jung, “Color Image Watermarking Technique using Adjacent Pixels and Spatial Encryption Technique,” The Journal of the Convergence on Culture Technology (JCCT) Vol. 7, No. 3, pp.863-867, Aug. 2021. https://doi.org/10.7236/IJIBC.2021.13.1.187

[9] J.-W. Oh, H.-K. Kim, and I.-T. Kim, “Design and Implementation of Fruit harvest time Predicting System based on Machine Learning,” Smart Media Journal, Vol. 8, No. 1, pp. 74-81, 2019. https://doi.org/10.30693/SMJ.2019.8.1.74

[10]R. Lu, "Predicting Firmness and Sugar Content of Sweet cheeies Using Near-infrared Diffuse Reflectance Spectroscopy" Transactions of the ASAE, Vol. 44(5): 1265-1271, 2001. DOI: 10.13031/2013.6421
[11]S.-J. Kim, "Association analysis of sugar contents based on color information of apple fruit images," Journal of KIIT, Vol. 18, No. 4, pp. 41-49, 2020. DOI:10.14801/jkiit.2020.18.4.41

※ 이 과제는 부산대학교 기본연구지원사업(2년)
에 의하여 연구되었음.