

# *k*-Means 클러스터링을 활용한 색각 검사 방안

이혜진, 박영호\*

숙명여자대학교 IT 공학과

e-mail : {adorablehye96, yhpark}@sookmyung.ac.kr

\*교신저자

## Color vision test using *k*-Means clustering

Hye-Jin Lee, Young-Ho Park

Dept. of IT Engineering, Sookmyung Women's University

### 요 약

본 논문에서는 *k*-Means 클러스터링을 활용한 컬러 기반 이미지 추출을 통한 색각 검사 방안 연구를 진행한다. 이를 위해, RGB 컬러스페이스 기반의 이미지를 특별한 컬러스페이스 이미지로 변환 후 컬러 패턴 분포에 따라 *k*-Means 클러스터링을 적용하여 다양한 형태의 이미지를 추출하는 실험을 수행한다. 위의 실험을 통해 하나의 이미지를 컬러 분포 패턴을 통해 클러스터링하여 이미지를 추출을 통하여 정상인과 색각 이상자를 판별할 수 있었다. 실험 결과, 다양한 형태와 색을 가진 이미지를 추출하여 정상인이 보는 이미지와 색각 이상자가 보는 이미지가 다른 것을 확인하였다.

### 1. 서론

최근 다양한 컬러 정보 디스플레이 기기들이 발전함으로써 예전보다 더 다양한 색상들을 접할 수 있게 되었다. 하지만 대부분의 사람들이 다양한 컬러를 접하고 있음에도 불구하고 전 세계 인구의 약 10%를 차지하는 색각 이상자들은 여전히 컬러 정보들에 있어서 제한적이다. 색각 이상이란 망막 원뿔 세포의 선천적 기능 이상 또는 후천적인 망막 원뿔 세포의 손상이나 시각 경로의 이상으로 색깔을 정상적으로 구분하지 못하는 현상이다. 이처럼 색각 이상자는 정상인과 다른 색깔을 가진 사람으로 제 1·제 2·제 3 이상의 3종과 정도에서도 3색형·2색형·1색형으로 나뉜다. 제 1 색각 이상자는 적에서 녹색에 걸친 색, 제 2 색각 이상자는 적부터 황록에 걸친 색, 제 3 색각 이상자는 청영역의 색의 판별을 잘 못한다. 이러한 색각 이상자를 판단하는 데에는 물체 색과 숫자를 이용하여 검사하는 이시하라 색각 검사표를 사용한다. 그러나, 이 검사는 숫자를 인지하기 어려운 아동이나 장애우들이 색각 검사를 하기에는 어려움이 있다.

따라서, 본 논문에서는 기존의 숫자 그림을 보고 검사하는 방법이 아닌 숫자를 인지하지 못하여도 검사를 할 수 있도록 일반 이미지를 사용하여 검사하는 방안을 제안하고자 한다. 이 방법은 RGB 컬러스페이스 기반의 이미지를  $L^*a^*b^*$  컬러스페이스로 이미지를 변환하여 처리하는 방법이다.  $L^*a^*b^*$  컬러스페이스 방법은 3.1 절에서 설명한다.

본 논문의 구성은 다음과 같다. 2 장에서는 관련 연구와 *k*-Means 클러스터링을 활용한 색각 이상 검사에

사용된 알고리즘을 설명한다. 3 장에서는 제안하는 방법을 설명한다. 4 장에서는 실험을 진행한다. 마지막으로 5 장에서는 결론 및 향후 연구에 대해 소개한다.

### 2. 관련 연구

본 장에서는 *k*-Means 클러스터링 알고리즘과 색각 이상을 판단하는 기존 검사 방법에 대해 설명한다. 먼저, 2.1 절에서는 *k*-Means 클러스터링의 수식과 관련하여 설명하고, 2.2 절에서는 색각 이상을 판단하는 기존 검사 방법에 대해 설명한다.

#### 2.1 *k*-Means 클러스터링 알고리즘

*k*-Means 클러스터링 알고리즘은 비지도 학습방법 알고리즘 중 하나이며 주어진 데이터를 *k*개의 클러스터 수로 분류하는 알고리즘이다[1,2]. 입력 받은 *n*개의 데이터를 *n*보다 작거나 같은 *k*개의 그룹으로 나눈다. 이 과정에서 거리 기반의 그룹 간 비유사도와 같은 비용 함수를 최소화하는 방식으로 이루어지며 같은 그룹 내의 유사도는 증가하고 다른 그룹 간의 유사도는 감소한다[1]. 여기서 *n*개의 데이터는 이미지 내 컬러의 개수를 나타내며, *k*개의 클러스터 수는 이미지의 색 분포 패턴에 따라 분류된 색 그룹의 수이다. 식 (1)[2]은 *k*-Means 클러스터링 알고리즘을 수식화한 것이다.

$$V = \sum_{i=1}^k \sum_{i \in S_i} |x_j - u_i|^2 \quad (1)$$

이 알고리즘의 장점으로는 간단한 알고리즘으로 계산 시간이 짧으며 주어진 자료에 대한 사전 정보없이

의미 있는 자료 구조를 찾아 낼 수 있다는 것이다. 단점으로는 분석 결과가 관찰치 사이의 거리 또는 유사성을 어떻게 정의하느냐에 따라 크게 좌우된다는 점이다.

**2.2 색각 이상을 판단하는 기존 검사 방법**

색각 이상을 검사할 수 있는 방법은 많이 개발되어 현재까지 이용되고 있다. 색상 배열법, 색광을 사용하는 방법, 물체 색을 이용하는 방법 크게 3 가지 방법이 있다. 색상 배열법은 FM 100-hue 검사와 이중 15 색상 배열검사기 등이 있고 색광을 이용하는 방법으로는 Nagel 의 색각경이 있으며 물체 색을 이용하는 방법에는 색각 검사표를 이용하여 검사하는 이시하라 색각 검사표, 한식색각표검사, HRR 등이 있다[3]. 색상 배열법이나 색광을 이용해서 검사하는 방법은 검사 시간을 단축하거나 결과의 정확도를 높이기 위한 연구는 해왔지만 물체 색을 이용하는 방법 중에 있어서는 여전히 숫자 그림인 이시하라 검사표를 제일 많이 이용되고 있다.

본 논문에서는 숫자를 인지하기 힘든 아동이나 장애우를 위해 흔히 사용되는 정해진 숫자 그림을 이용한 검사가 아닌 원하는 임의의 이미지를 이용하여 자유롭게 색각 이상을 검사하는 방법을 살펴보고자 한다.

**3. 제안하는 방법**

본 장에서는 *k*-Means 클러스터링을 활용한 색각 검사 방안 에 관하여 제안하는 방법을 설명한다. 먼저, 3.1 절에서는 RGB 컬러스페이스를 *L\*a\*b\** 컬러스페이스로 변환하는 이유와 정의를 설명하고, 3.2 절에서는 *L\*a\*b\** 컬러스페이스 기반의 *k*-Means 클러스터링을 이용한 색각 검사 방안 에 대하여 설명한다.

**3.1 *L\*a\*b\** 컬러스페이스**

*L\*a\*b\** 컬러 스페이스는 인간의 시각에 대한 연구를 바탕으로 정의되었으며, '*L\**'층은 각 색의 밝기 값을 포함하고 모든 색의 정보는 '*a\**'층과 '*b\**'층이 갖고 있다[4]. 그러므로 색 구별을 잘 못하는 색각 이상자들을 판별하는 색각 검사 방안 에 더 적합함으로 RGB 컬러스페이스를 *L\*a\*b\** 컬러스페이스로 변환하여 사용한다. *L\*a\*b\** 컬러스페이스(CIELAB 또는 CIEL\*a\*b\*)는 CIE XYZ 3 자극값에서 파생되며 광도층 '*L\**'과 빨간색-녹색 축에 놓인 색의 위치를 나타내는 색도층 '*a\**'와 파란색-노란색 축에 놓인 색의 위치를 나타내는 색도층 '*b\**'로 구성된다. 식(2), (3)[6]은 *L\*a\*b\** 컬러스페이스로 변환하는 수식을 나타낸 것이다.

$$\begin{aligned} L^* &= 116f(Y/Y_n) - 16 \\ a^* &= 500[F(X/X_n) - f(Y/Y_n)] \\ b^* &= 200[F(Y/Y_n) - f(Z/Z_n)] \end{aligned} \quad (2)$$

$f(t) =$

$$\begin{cases} t^{1/3} , & \text{if } t > \left(\frac{6}{29}\right)^3 \\ \frac{1}{3}\left(\frac{29}{6}\right)^2 t + \frac{4}{29} , & \text{otherwise} \end{cases} \quad (3)$$

$X_n, Y_n$  및  $Z_n$ 은 CIE XYZ 를 표준 흰색에 대해 정규화한 값이다[5].

**3.2 *L\*a\*b\** 컬러스페이스 기반의 *k*-Means 클러스터링을 이용한 색각 검사 방안**

본 절에서는 아동이나 장애우를 위한 색각 방안을 제안한다. 제안하는 방법은 MATLAB 기반에서 *k*-Means 클러스터링 알고리즘과 RGB 컬러스페이스 기반의 이미지를 사용하여 *L\*a\*b\** 컬러스페이스 기반의 이미지를 추출하는 단계에 대하여 설명한다. 단계는 총 세 단계로 구성된다.

- a) *Step 1 (컬러스페이스 변환 단계)*: 컬러스페이스 변환 단계에서는 RGB 컬러스페이스 기반인 이미지를 각 색의 밝기 값을 포함하고 있는 '*L\**' 층과 모든 컬러 정보를 갖고 있는 '*a\**' 층과 '*b\**' 층을 가진 *L\*a\*b\** 컬러스페이스 기반으로 변환한다.
- b) *Step 2 (k-Means 클러스터링 단계)*: *k*-Means 클러스터링 단계에서는 *k*-Means 클러스터링을 사용하기 위해 분할할 군집 수와 두 객체가 서로 얼마나 가까운지를 수량화하는 거리 측정법을 지정하고 클러스터링을 하기 위해 데이터의 데이터 형을 single 형으로 변환한다. 클러스터링 결과는 변환된 *k* 개의 군집 수로 알 수 있다.
- c) *Step 3 (컬러 기반 이미지 추출 단계)*: 컬러 기반 이미지 추출 단계에서는 모든 객체에 대해 군집에 해당하는 인덱스 또는 레이블을 반환한 후, 해당 군집의 픽셀을 남긴 후 나머지 군집은 모두 제거한다.

색각 이상자들은 판별을 잘 하지 못하는 색상에 따라 다양한 형태의 이미지를 볼 수 있다. 결과 이미지는 4.1 절 실험 결과에서 확인할 수 있다.

**4. 실험**

본 장에서는 개발 환경과 실험 결과에 대해 설명한다. 먼저, 4.1 절에서는 개발 환경을 설명하고, 4.2 절에서는 추출한 이미지와 실험 결과를 설명한다.

**4.1 실험 환경**

본 연구에서는 하나의 이미지를 색 분포 패턴에 따라 클러스터링하여 다섯 가지의 결과를 수집하였다. 해당 이미지는 헤마톡실린과 에오신(H&E)으로 염색한 조직을 촬영한 이미지를 사용하였으며 개발 툴은 MATLAB 을 사용하였다. 개발을 진행한 컴퓨터 사양은 Intel i7-7500U CPU 를 사용하였다.

#### 4.2 실험 결과

Fig.1 은 크게 흰색, 보라색, 핑크색 세 가지 색을 갖고 있는 원본 이미지이며 Fig.2, Fig.3(a)~(d)는 색깔 이상자들이 바라본 이미지를  $L*a*b*$ 컬러스페이스를 기반으로  $k$ -Means 클러스터링을 하여 나타낸 것이다. 만약 Fig.1 이 Fig.2 와 Fig.3 (a)~(d)의 이미지들 중 하나로 보인다면 그 사람은 색깔 이상자임을 알 수 있으며 [4]에서 참조하였다. 더 나아가 Fig.2 로 보인다면 컬러 구분을 아예 하지 못하는 단색형 색깔 이상자임을 알 수 있고 Fig.3 (a)~(d)의 이미지 중 하나로 보인다면 제 1,2,3 색깔 이상자 중 하나임을 알 수 있다.

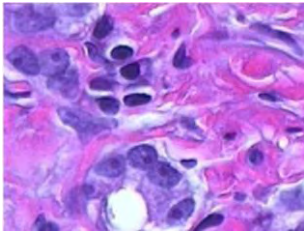


Fig.1 원본 이미지

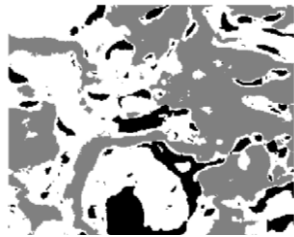
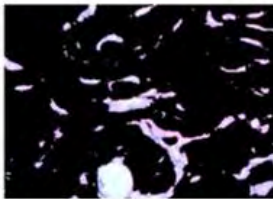
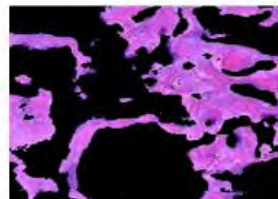


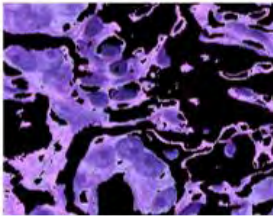
Fig.2 단색형 색깔 이상자가 본 이미지



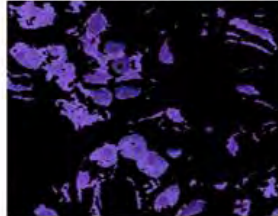
(a) Image1



(b) Image2



(c) Image3



(d) Image4

Fig.3 색깔 이상자들이 본 이미지

#### 5. 결론 및 향후 연구

본 논문에서는 물체 색을 이용하여 색깔 검사를 하는 방법 중 숫자 이미지를 사용하는 이시하라 색깔 검사표가 제일 많이 사용된다. 하지만 숫자를 인지하기 힘든 아동이나 장애우가 색깔 검사를 하기에는 어려움이 있다고 생각되어 어떤 이미지로도 색깔 검사를 할 수 있는 방안을 제안하였다. 실험 결과, 원본 이미지를 기준으로  $k$ -Means 클러스터링을 이용하여 컬러 기반 이미지를 추출하여 색을 인지하지 못하는 단색형 색깔 이상자와 제 1, 2, 3 색깔 이상자들이 본 이미지를 추출하였다. 이를 통해 기존의 정해진 숫자 이미지가 아닌 다양한 이미지를 이용하여 컬러를 기반으로 클러스터링하고 이미지를 추출하여 사람들에게

게 보이는 형태에 따라 색깔 검사를 할 수 있음을 확인하였다. 하지만 이미지에 따라 갖고 있는 컬러의 수가 다르기 때문에 클러스터의 개수  $k$  를 매번 다르게 지정해주어야 한다는 문제점에 대하여 보완 방법이 필요하다. 그러므로  $k$ -Means 클러스터링 알고리즘에서 색의 수를 결정하는  $k$ 의 예측 방법은 추후 연구로 남겨둔다.

#### 사사문구

이 논문은 2019 년도 정부(미래창조과학부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임. (No.2016-0-00406. (기반 SW-창조씨앗 2 단계)SIAT형 CCTV 클라우드 플랫폼 기술 개발)

#### 참고문헌

- [1] B. Subbiah and S. Christopher, "Image Classification through Integrated K-means Algorithm," *International Journal of Computer Science Issues*, Vol.9, No.2, pp.518-524, 2012.
- [2] H.-H. Bock, "Clustering Methods: A History of k-Means Algorithms", *Selected contributions in data analysis and classification*. pp.161-172, 2007.
- [3] 김민섭, 노문남, 이건, 위원량, 이진학. "서한전산화 색깔검사 (1) : 전산화된 색깔검사의 개발 및 시험 연구", *대한안과학회지*, Vol.41, No.1, pp.205-214, 2000
- [4] 컬러 기반 K-means 클러스터링, <https://kr.mathworks.com/>
- [5] Lab 컬러스페이스 설명, <https://ko.wikipedia.org>
- [6] C.Connolly and T. Fleiss. "A study of efficiency and accuracy in the transformation from RGB to CIELAB color space." *IEEE Transactions on Image Processing*, Vol.6, No.7, pp.1046-1048, 1997.