

## 가우시안 혼합모델을 이용한 솔라셀 색상분류

고진석 · 임재열<sup>†</sup>

<sup>†</sup>한국기술교육대학교 대학원 전기전자통신공학과

### Solar Cell Classification using Gaussian Mixture Models

Jin Seok Ko and Jae Yeol Rheem<sup>†</sup>

<sup>†</sup>Dept. of Electrical, Electronics & Communication Engr., Korea University of Technology and Education

#### ABSTRACT

In recent years, worldwide production of solar wafers increased rapidly. Therefore, the solar wafer technology in the developed countries already has become an industry, and related industries such as solar wafer manufacturing equipment have developed rapidly. In this paper we propose the color classification method of the polycrystalline solar wafer that needed in manufacturing equipment. The solar wafer produced in the manufacturing process does not have a uniform color. Therefore, the solar wafer panels made with insensitive color uniformity will fall off the aesthetics. Gaussian mixture models (GMM) are among the most statistically mature methods for clustering and we use the Gaussian mixture models for the classification of the polycrystalline solar wafers. In addition, we compare the performance of the color feature vector from various color space for color classification. Experimental results show that the feature vector from YCbCr color space has the most efficient performance and the correct classification rate is 97.4%.

**Key Words** : Color classification, GMM, Solar cell classification

## 1. 서 론

최근 수년간 전세계 태양전지 생산량은 급속도로 증대되고 있다. 이에 따라 선진국에서는 이미 태양광 기술이 하나의 산업으로 자리잡고 있으며, 이와 관련하여 태양전지 제조장비와 같은 관련산업이 급속도로 발전하고 있다[1].

태양전지 제조 공정에서 생산된 다결정 태양전지 웨이퍼는 균일한 색상을 가지지 못한다. 따라서 색상구분 없이 태양전지 패널을 제작하면, 패널 표면 색상의 균일성이 떨어져서 미관이 떨어지게 된다. 따라서 태양전지 제조장비 업체들은 태양전지의 색상을 분류해야 한다.

Fig. 1은 동일한 공정에서 생산된 다결정 솔라셀 제품이다. 동일한 공정임에도 불구하고 색상의 밝기편차가 있는 것을 확인할 수 있다. 또한 동일한 밝기수준의 솔라셀도 내부의 결정 때문에 표면의 밝기가 불균일하

게 보여지는 특징이 있다. 이는 솔라셀 구성이 다결정으로 이뤄졌기 때문이다.

본 논문에서는 태양전지의 색상을 분류하는 방법에 대하여 제안하였다. 색상분류를 위한 방법으로 가우시안 혼합모델을 이용하였다. 또한 색상분류를 위한 특징 벡터는 컬러 히스토그램의 채널 별 peak값을 이용하였다. 다양한 컬러모델에서 솔라셀 색상분류의 성능을 비교하기 위하여 HSV, YCbCr, Lab, RGB 컬러모델에서 특징을 추출하여 비교하였다[2-3].

## 2. 솔라셀 색상분류를 위한 특징벡터

색상은 영상이 가지는 하나의 속성으로 영상의 전반에 걸친 색상 정보는 영상분류 및 내용기반 검색분야 등에서 활용된다. 영상 내부 각 영역의 색상인식 또는 영상에 포함된 물체의 색상인식은 영상분할, 물체인식, 물체 추적 등 여러 분야에서 광범위하게 사용된다.

본 논문에서는 색상분류를 위한 특징으로 다양한 컬러

<sup>†</sup>E-mail : rheem@kut.ac.kr

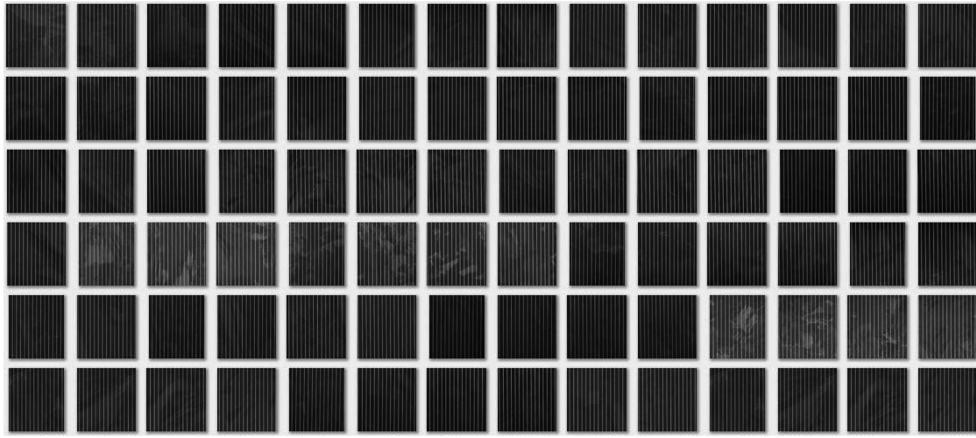


Fig. 1. Color difference of polycrystalline solar wafers.

러공간에서의 히스토그램 기반의 특징벡터를 이용하였다. 컬러모델은 어떤 특정한 상황에서 컬러의 특징을 설명하기 위한 방법이다. 그러나 하나의 컬러모델을 이용하여 컬러의 모든 성질을 표현하는 것은 불가능하기 때문에 여러 컬러모델을 정의하여 사용한다. 이러한 컬러모델들은 보통 세가지 요소를 사용하여 색을 표현하기 때문에, 각각의 요소를 하나의 축으로 하는 3차원 좌표 시스템에 대응할 수 있다.

본 논문에서는 색상분류를 위하여 가장 많이 사용되는 컬러모델인 RGB, HSV, YCbCr 그리고 Lab 컬러모델에 대해서 히스토그램 기반의 특징벡터를 추출하여 성능을 비교하였다[2].

히스토그램을 분류기의 특징벡터로 직접 사용할 경우, 특징벡터의 차수가 높아서 분류기의 학습과정에서 많은 시간을 수요한다. 따라서 적당한 차수의 특징벡터로 차원을 축소할 필요가 있다. 본 논문에서는 히스토그램의 채널 별 peak점을 특징벡터로 이용하였다. 정확한 peak점을 구하기 위하여 히스토그램에 smoothing 필터를 취한 후, peak점을 구하였다.

Fig. 2는 컬러 히스토그램에서 peak 점을 구하는 모습을 보여주고 있다. 1개의 영상에 대해서 채널 별 히스토그램 peak점을 구하여, 3 by 1 크기의 특징벡터를 생성한다.

### 3. Gaussian mixture model을 이용한 솔라셀 컬러분류

GMM은 출력 확률밀도 함수가 가우시안 밀도 혼합인 1개의 상태만으로 구성된 CHMM의 한 형태이다. 영상으로부터 추출된 특징벡터들의 분포가 가우시안

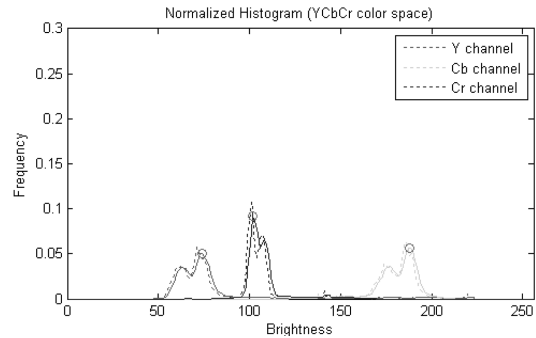


Fig. 2. Feature extraction based on color histogram.

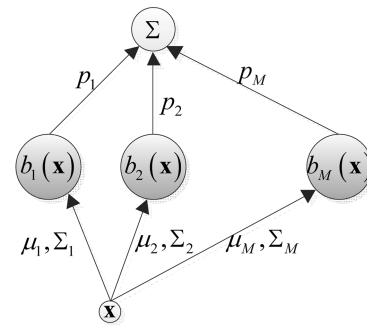


Fig. 3. M-component Gaussian mixture densities.

혼합밀도로 모델링 되므로  $M$ 차원 특징벡터  $\mathbf{x}$ 를 위한 혼합밀도는 Fig. 3과 같다[3].

Fig. 3의  $M$ 성분 가우시안 혼합밀도함수의 분포는 식 (1)과 같이 나타낼 수 있다.

$$p(\mathbf{x}|\lambda) = \sum_{i=1}^M p_i b_i(\mathbf{x}) \quad (1)$$

여기서  $M$ 은 혼합성분 개수이고,  $p_i$ 는  $i$ 번째 혼합성분의 가중치이며,  $\sum_{i=1}^M p_i = 1$  를 만족해야 한다.

각 성분밀도  $b_i(\mathbf{x}), i=1, \dots, M$ 는 정규확률분포를 사용하며 식 (2)로 정의된다.

$$b_i(\mathbf{x}) = \frac{1}{(2\pi)^{D/2} |\Sigma_i|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mu_i)' \Sigma_i^{-1}(\mathbf{x}-\mu_i)\right\} \quad (2)$$

여기서  $\mu_i$ 는  $i$ 번째 성분의 평균벡터이고,  $\Sigma_i$ 는 공분산 행렬이다. 모델 파라미터  $\lambda$ 는 식 (3)과 같이 가우시안 혼합성분 밀도의 가중치, 평균벡터, 공분산 행렬로 구성된다.

$$\lambda = \{p_i, \mu_i, \Sigma_i\} \quad i=1, \dots, M. \quad (3)$$

GMM의 우도를 최대화하기 위한 파라미터  $\lambda$ 를 추정하기 위하여 최대우도 추정이 사용된다. 최대우도를 위한 파라미터는 EM 알고리즘을 반복적으로 계산하여 구할 수 있다[4]. 다음의 재추정식은 GMM을 이용한 웨이퍼 컬러 모델의 유사도를 단조증가 시킨다.

Mixture weight:

$$\bar{p}_i = \frac{1}{T} \sum_{t=1}^T p(i | \mathbf{x}_t, \lambda) \quad (4)$$

Mean vector:

$$\bar{\mu}_i = \frac{\sum_{t=1}^T p(i | \mathbf{x}_t, \lambda) \mathbf{x}_t}{\sum_{t=1}^T p(i | \mathbf{x}_t, \lambda)} \quad (5)$$

Variance vector:

$$\bar{\sigma}_i^2 = \frac{\sum_{t=1}^T p(i | \mathbf{x}_t, \lambda) \mathbf{x}_t^2}{\sum_{t=1}^T p(i | \mathbf{x}_t, \lambda)} - \bar{\mu}_i^2 \quad (6)$$

컬러분류는 Bayes' rule에 따라  $C$ 개의 컬러 모델들 중 사후확률  $P(\lambda_c | X), 1 \leq c \leq C$ 를 최대화 하는 컬러모델  $\hat{c}$ 을 찾는 것이다. 여기에서  $C$ 개의 컬러 클래스는 GMM 모델  $\lambda_1, \dots, \lambda_C$ 로 각각 나타낸다. 이때 컬러 분류 시스템은 다음과 같은 식(7)로 표현할 수 있다.

$$\hat{c} = \arg \max_{1 \leq c \leq C} \log p(\mathbf{x} | \lambda_c) \quad (7)$$

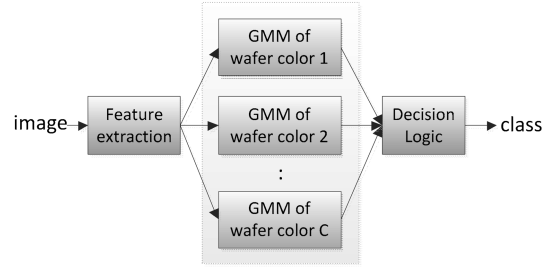


Fig. 4. Color classifier based on Gaussian mixture models.

Fig. 4는 가우시안 혼합모델을 이용한 색상 분류 시스템을 나타내고 있다.

## 4. 실험결과

### 4.1. 실험데이터

실험에는 총 936개의 솔라셀 이미지가 사용되었으며, 색상 분류 클래스는 3가지로 설정하였다. 색상분류 클래스의 개수를 3개로 한 이유는 솔라셀 자동 검사장비의 분류 기준을 참고하였다. 분류 클래스가 증가할수록 검사장비의 sorter 부분이 복잡해지고, 전체 tack time도 증가하게 된다.

클래스 분류기준은 솔라셀 제조업체의 생산 오퍼레이터의 분류 기준이 필요하다. 그러나 현실적으로 제조업체의 분류 기준을 얻을 수 없기 때문에 벡터 양자화 과정을 통하여 3개의 클래스로 분류하였다. 사용된 알고리즘은 k-means 알고리즘을 사용하였으며, HSV, YCbCr, Lab, RGB 컬러모델에서 추출한 특징벡터를 이용하여 벡터 양자화를 수행하였다[5].

벡터양자화를 수행한 결과를 육안으로 확인한 결과, YCbCr 컬러모델과 Lab 컬러모델에서 벡터양자화를 수행한 결과가 다른 컬러모델에 비해서 상대적으로 클래스 분류가 잘 된 것을 확인할 수 있었다.

Lab 컬러모델과 YCbCr 컬러모델에서 추출한 특징벡터를 이용한 결과를 실험데이터로 활용하였다. 각각의 분류 클래스 별 실험 데이터에서 학습을 위하여 30%, 분류 테스트를 위하여 70%의 데이터를 이용하였다. Table 1과 Table 2는 실험에서 사용한 데이터의 수량을 보여주고 있다.

### 4.2. 컬러모델에 따른 색상 분류 실험

Table 1에서 나타내는 Lab 컬러모델을 이용한 벡터양자화 결과를 실험데이터 1, Table 2에서 나타내는 YCbCr 컬러모델을 이용한 벡터양자화 결과를 실

**Table 1.** Experimental data sets in Lab color model.

Classes	# of total images	# of training images	# of test images
Class 1	506	138	324
Class 2	228	59	140
Class 3	202	82	193
Total	936	279	657

**Table 2.** Experimental data sets in YCbCr color model.

Classes	# of total images	# of training images	# of test images
Class 1	506	146	342
Class 2	228	63	147
Class 3	202	71	167
Total	936	280	656

험데이터 2로 표현할 때, 실험데이터 1을 이용하여 각각의 특징벡터에 따른 분류율을 비교한 결과를 Table 3에 나타내었으며, 실험데이터 2를 이용하여 각각의 특징벡터에 따른 분류율을 비교한 결과를 Table 4에 나타내었다.

실험결과 Lab 컬러모델을 이용하여 벡터양자화를 수행한 경우, Lab 컬러 특징벡터를 이용하는 경우에 성능이 가장 뛰어났다. 마찬가지로 YCbCr 컬러모델을 이용하여 벡터양자화를 수행한 경우, YCbCr 컬러 특징벡터를 이용하는 경우에 성능이 가장 뛰어났다. 두 실험 모두 HSV 컬러 특징벡터에서 성능이 낮게 나왔다.

#### 4.3. 혼합성분 개수에 따른 분류 성능 비교

가우시안 혼합모델은 혼합성분의 개수가 성능에 중요한 영향을 미친다. 본 실험에서는 혼합성분 개수에 따른 분류율을 비교하였으며, Table 5에 나타내고 있다.

실험결과 혼합성분의 개수가 4-6개일 경우에 성능이 가장 뛰어났으며, 혼합성분의 개수가 6개를 넘어서면 성능이 감소하는 것을 확인할 수 있었다. 전체적인 성능은 YCbCr 컬러 특징벡터를 사용한 경우가 Lab 컬러 특징벡터를 사용하는 것에 비해서 우수하였으며, 최대 약 2.3% 정도 우수한 성능을 보여줬다.

**Table 5.** Comparison of classification rate on the number of mixture components (%).

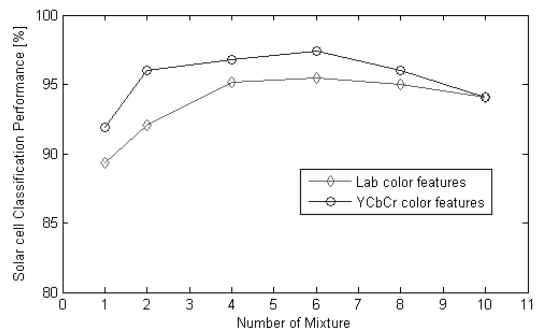
# of mixtures		1	2	4	6	8	10
Correct classification rate (%)	Lab color model	89.3455	92.0852	95.1294	95.4338	94.9772	94.0639
	YCbCr color model	91.9207	96.0366	96.7988	97.4085	96.0366	94.0549

**Table 3.** Comparison of classification rate on the type of feature vectors: Experimental data set 1 (result of VQ using Lab color model), # of mixture is 4.

Type of feature vectors	Correct classification rate (%)	Error rate (%)
RGB color feature	93.9117	6.0883
HSV color feature	55.5556	44.4444
YCbCr color feature	92.3896	7.6104
Lab color feature	95.1294	4.8706

**Table 4.** Comparison of classification rate on the type of feature vectors: Experimental data set 2 (result of VQ using YCbCr color model), # of mixture is 4.

Type of feature vectors	Correct classification rate (%)	Error rate (%)
RGB color feature	92.3780	7.6220
HSV color feature	54.7256	45.2744
YCbCr color feature	96.7988	3.2012
Lab color feature	88.8720	11.1280

**Fig. 5.** Comparison of classification rate on the number of mixture components.

## 4. 결 론

본 논문에서는 효율적인 솔라셀 색상분류를 위하여 가우시안 혼합모델을 이용한 색상 분류기를 제안하였다. 제안된 방법은 솔라셀 영상의 컬러 히스토그램 채

널 별 peak값을 특징벡터로 이용하였다. 다양한 컬러 모델에서 솔라셀 색상분류의 성능을 비교하였으며, YCbCr 컬러모델에서 가장 뛰어난 성능을 보여주는 것을 확인하였다. 또한 4-6개의 혼합성분이 존재할 때, 최적의 분류율을 보이는 것을 확인함으로써 실시간 처리에도 적합함을 확인할 수 있었다. 제안된 분류기의 유용성을 검증하기 위하여 실시간 처리가 가능한 시스템을 구현하고, 실제 생산현장에 적용하여 그 유용성을 입증할 계획이다.

### 참고문헌

1. D.M. Tsai, C.C. Chang, and S.M. Chao, "Micro-crack inspection in heterogeneously textured solar wafers using anisotropic diffusion," *Image and Vision Computing*, vol. 28, 2010, pp. 491-501.
2. M.W. Schwarz, W.B. Cowan, and J.C. Beatty, "An experimental comparison of RGB, YIQ, LAB, HSV, and opponent color models," *ACM Transactions on Graphics (TOG)*, vol. 6, 1987, pp. 123-158.
3. D.A. Reynolds and R.C. Rose, "Robust text-independent speaker identification using Gaussian mixture speaker models," *Speech and Audio Processing, IEEE Transactions on*, vol. 3, 2002, pp. 72-83.
4. A.P. Dempster, N.M. Laird, D.B. Rubin, and others, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 39, 1977, pp. 1-38.
5. J. A. Hartigan and M. A. Wong, "A k-means clustering algorithm," *JR Stat. Soc., Ser. C*, vol. 28, pp. 100-108, 1979.

---

접수일: 2011년 3월 7일, 심사일: 2011년 4월 5일,  
 게재확정일: 2011년 5월 31일