

No-Reference Color Dot Artifact Assessment for Remosaiced Images

Subin Han, Seungwan Jeon, Sara Lee, Yu Gyeong Lee, Hee-shin Kim, Kichul Park, Sung-Su Kim, Yitae Kim; S.LSI Division, Samsung Electronics; Hwaseong-si, Gyeonggi-do, Republic of Korea

Abstract

In complementary metal oxide semiconductor image sensor (CIS) industry, advances of techniques have been introduced and it led to unexpected artifacts in the photograph. The color dots, known as false color, also appear in images from CIS employing the modified color filter arrays and the remosaicing image signal processors (ISPs). Therefore, the objective metric for image quality assessments (IQAs) have become important to minimize artifacts for CIS manufacturers. In our study, we suggest a novel no-reference IQA metric to quantify the false color occurring in practical IQA scenarios. We propose a pseudo-reference to overcome the absence of reference image, inferring an ideal sensor output. As we detected the distorted pixels by specifying outlier colors with a statistical method, the pseudo-reference was generated while correcting outlier pixels with the appropriate colors according to an unsupervised clustering model. With the derived pseudo-reference, our method suggests a metric score based on the color difference from an input, as it reflects the results of our subjective false color visibility analysis.

Introduction

With the rapid spread of mobile devices, users have demanded improved experiences on the mobile camera with performance enhancement and compactness. Following this trend, the complementary metal oxide semiconductor (CMOS) image sensor (CIS) industry has adopted several technical advances in both hardware and software to overcome physical constraints.

As part of the ongoing technical trials, color filter array (CFA) structures have been modified. A CMOS image sensor captures color information using a mosaic pattern of color-filtered pixels, known as CFA pattern (Fig. 1). Nowadays, the modified patterns of CFAs have been adopted to the novel CIS to improve sensitivity in low light conditions. Unlike the conventional Bayer pattern which consists of 2 green pixels, 1 red pixel and 1 blue pixel (Fig. 1a), non-Bayer pattern CFAs (Fig. 1b) have employed the pixel patterns split in more identical pixels to act as a single large pixel through binning in low light conditions.

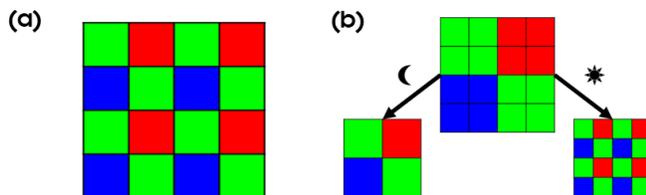


Figure 1. Color filter array patterns: (a) Bayer pattern, and (b) non-Bayer pattern.

The remosaicing ISP is a specific image signal processor (ISP) algorithm adopted together with the non-Bayer CFA to provide high resolution especially for high luminance environments. It interpolates the color data to rearrange it into the Bayer pattern in high luminance conditions. However, due to the intact limitation of 3-Ch outputs, the acquired raw information is insufficient in spatial domain to reproduce actual colors, and it led to unexpected artifacts (e.g., false color, zipper artifact, and etc.) which the CIS manufacturers are always concerned with. Thus, it has been a challenge to minimize the artifacts when optimizing remosaicing ISP parameters. Moreover, several objective image quality assessments (IQAs) for evaluating the outputs has been conducted actively.

False color is one of the color artifacts generated in remosaicing ISP outputs. It appears as an unexpected color dot when remosaicing ISPs failed to recover original colors from CFA-patterned signals. These dots are typically observed in high-frequency, and high-contrast areas such as edges of lush leaves, and sparkling water surface (Fig. 2). In order to measure them, various objective metrics have been used: full-reference methods suggested for general artifacts, comparing sensor outputs against an undistorted reference image [1, 2], and a no-reference method applicable only for a black-and-white chart [3]. In fact, those methods do not adequately address most practical situations of false color artifact on which the subjective assessments have been only conducted.

In our study, we propose a false color metric based on no-reference method available for natural scene images. During the method, we generated a pseudo-reference which would be acted as an actual reference when comparing it with an input image. The key idea for generating a pseudo-reference is to correct outlier colors of the input into proper colors. To achieve this, we designed the outlier color detection method composed of the luminance clustering by the k -means clustering algorithm [4] and a color distribution estimating method using the Gaussian mixture model [5] for conducting outlier extraction. The detected outlier pixels were, and then, repainted into the appropriate ideal colors adopting color quantization method with the k -means clustering algorithm according to [6] in a pseudo-reference generation. Finally, we calculated the false color score using the color difference from CIEDE2000 formula [7]. Additionally, we investigated false color visibility through an analysis of mean opinion score (MOS) and verified that our proposed score correlates with the qualitative assessment.



Figure 2. False color artifacts occurred in high-frequency, high-contrast areas of natural images.

No-reference False Color Metric

Our metric consists of three steps: false color detection; pseudo-reference image generation; and score calculation (Fig. 3).

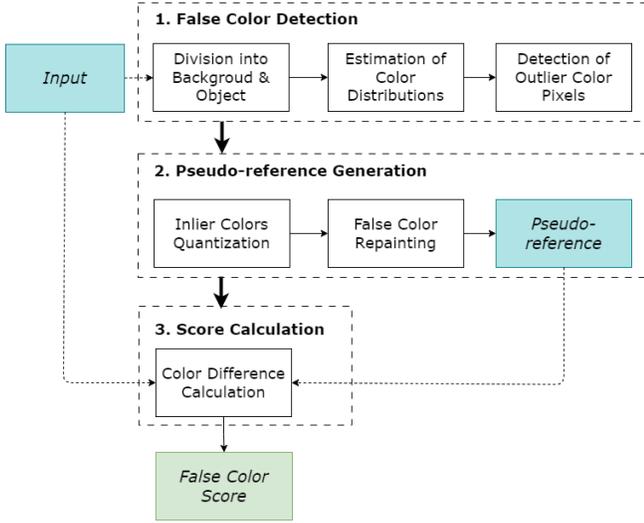


Figure 3. The flowchart of no-reference false color metric.

1. False Color Detection

To identify the false color artifacts, we detected outlier colors in an input and labeled outlier pixels as false color pixels (Fig. 4). Before detecting outlier colors, we divided pixels into two groups by luminance due to the properties of an input scenarios. Assuming that the input (Fig. 4a) consists of a background area illuminated brightly and a N-color object (N commonly 1 or 2), two areas exhibit significantly different Chroma_{HSV}(R, G, B) = $V(R, G, B) \times S(R, G, B)$ densities in the Hue-Chroma-Value (HCV) space, because the maximum Chroma depends on the lightness [8]. Thus, we categorized pixels into the backgrounds and the object by luminance beforehand to effectively analyze colors. However, as a luminance range is not a constant for all inputs, we used k -means clustering algorithm, one of unsupervised clustering methods [4]. After converting the RGB image into grayscale using an equation $Y' = 0.2125R + 0.7154G + 0.0721B$ recommended by Poynton [9], the k -means clustering model grouped pixels into the 2 luminance groups (Fig. 4bc).

Next, we estimated color distributions while clustering pixels according to the colors. To characterize the color distributions, we trained a Gaussian mixture model (GMM) [5] from 2-D color vectors, \mathbf{X} , expressed in Hue-Chroma polar coordinates domain. Because of the linear relationship between the Chroma and the RGB combination, as described by Romani, *et al.* [8], we used the

Chroma rather than Saturation of the HSV color space. As we mentioned, we used the GMM which is a probabilistic model represented by a finite number of mixed Gaussian distributions according to Deng *et al.* [5]. When a GMM is fitted as the function f_{GMM} , the model clusters data \mathbf{X} into a finite number N groups and estimates N Gaussian parameters, consisting of mean, covariance, and posterior ($\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\pi}$), expressed as follows:

$$f_{\text{GMM}}(\mathbf{X}, N) \rightarrow (\mathbf{X}_1, \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \boldsymbol{\pi}_1), \dots, (\mathbf{X}_N, \boldsymbol{\mu}_N, \boldsymbol{\Sigma}_N, \boldsymbol{\pi}_N). \quad (1)$$

In our method, we estimated the distributions separately through the model.

For \mathbf{X}_{bg} :

$$f_{\text{GMM}}(\mathbf{X}_{bg}, 1) = \{(\mathbf{X}_{bg}, \boldsymbol{\mu}_{bg}, \boldsymbol{\Sigma}_{bg})\}, \quad (2)$$

where $\boldsymbol{\mu}_{bg}$ denotes a background mean color, $\boldsymbol{\Sigma}_{bg}$ is a covariance.

For \mathbf{X}_{obj} :

$$f_{\text{GMM}}(\mathbf{X}_{obj}, N) = \{(\mathbf{X}_1, \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \boldsymbol{\pi}_1), \dots, (\mathbf{X}_N, \boldsymbol{\mu}_N, \boldsymbol{\Sigma}_N, \boldsymbol{\pi}_N)\}, \quad (3)$$

where $\boldsymbol{\mu}_k$ denotes a mean color of object color \mathbf{X}_k and $\boldsymbol{\Sigma}_k$ denote covariance and $\boldsymbol{\pi}_k$ is a posterior for $k = 1, \dots, N$.

To detect outliers of each distributions, we employed the outlier detection method that utilizes the Mahalanobis distance as the distance formula for two points in the multivariate distributions, as proposed by Gallego *et al.* [10]. The Mahalanobis distance between each color point \mathbf{x}_i and $\boldsymbol{\mu}$ is written as follows:

$$d(\mathbf{x}_i, \boldsymbol{\mu}_k) = \sqrt{(\mathbf{x}_i - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_k)}. \quad (4)$$

Furthermore, according to Gallego *et al.*, when defining the probability, P , that a color is present in a region within a specific Mahalanobis distance, d , the d corresponding to P can be derived as follows [10]:

$$d(P) = \sqrt{-2 \log(1 - P)}. \quad (5)$$

With the equation (5), we defined cutoff distances for the inliers with P set to 0.5 and 0.95, for the background and the object respectively as follows:

$$d_{cutoff,bg} = d_{bg}(0.5), \quad (6)$$

$$d_{cutoff,k} = d_k(0.95) \text{ for } k = 1, \dots, N, \quad (7)$$

where $d_{cutoff,bg}$ is the background distance cutoff and $d_{cutoff,k}$ is the object distance cutoff for all $k = 1, \dots, N$. Note that, we set the background color cutoff more strictly than object colors to generate the pseudo-reference stably in the next step. After that, we detected inlier pixels, G_{bg} , whose colors are within the $d_{cutoff,bg}$ among \mathbf{X}_{bg} and inlier pixels, G_k , whose colors are within the $d_{cutoff,bg}$ among \mathbf{X}_k for all $k = 1, \dots, N$. Then, we designated the false color pixels as the complement of the union of all inliers as follows:

$$G^c = (G_{bg} \cup G_1 \cup \dots \cup G_N)^c. \quad (8)$$

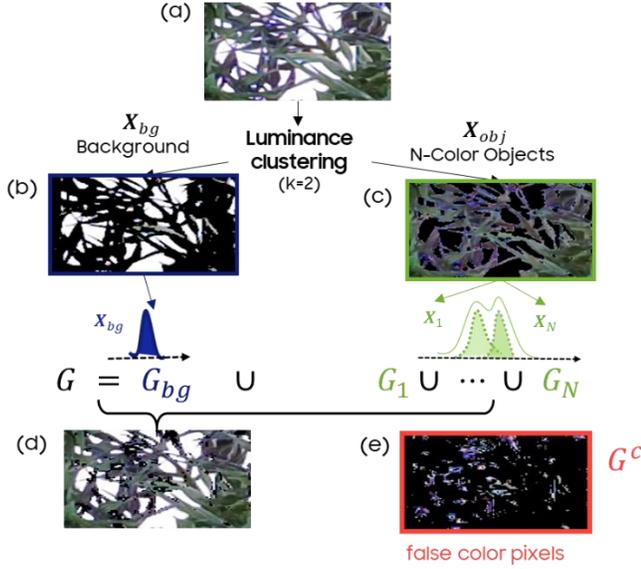


Figure 4. In false color detection, input (a) is divided into background regions (b) and N-Color object regions (c) by luminance, and inliers of each group are detected as (d). The false color pixels (e) are designated as the complement of (d).

2. Pseudo-reference Generation

In order to measure artifacts without an actual reference, we introduced a pseudo-reference (Fig. 5b) generated from the input image (Fig. 5a). In this method, we employed the color quantization method with k -means clustering algorithm suggested by Celebi [6]. In Celebi's method, colors in the image are quantized through the process of palette design and pixel mapping. Modifying the inputs of each process, we quantized the inlier colors in G and corrected G^c to the quantized inlier colors to generate the pseudo-reference.

First, we fitted a k -means clustering model ($k=32$) with inlier colors in G to quantize them into 32 representative colors. When fitting the model, we calculated the distance with the Euclidean distance in the CIELAB [7] space and grouped colors into 32 clusters empirically. After this process, we obtained the mean colors k_1, k_2, \dots, k_{32} from each cluster and these mean colors formed the inlier color palette. This palette was then mapped to each pixel in G^c through model prediction. With the inlier-trained model, we predicted the nearest k of each false color to assign one of inlier colors to each false color pixel. Therefore, we generated the pseudo-reference by coloring G^c with the RGB values of the assigned colors. As f_k represents the mapping function in the prediction, a pseudo-reference image R is defined as:

$$R = \{i_{(h,w)} \in \mathbb{R}^3 | (h,w) \in G\} \cup \{f_k(i_{(h,w)}) \in \mathbb{R}^3 | (h,w) \in G^c\},$$

$$f_k: i \rightarrow k \in \{k_1, k_2, \dots, k_{32}\} \quad (9)$$

where (h,w) represents the coordinated of the image, and G is the set of inlier pixels. In other words, we inferred a pseudo-reference (Fig. 5b) by repainting false color pixels with adaptively predicted inlier colors.



Figure 5. An example result: (a) input, and (b) pseudo-reference.

3. Score Calculation

After detecting false color pixels and generating the pseudo-reference, we measured the color difference between the input and pseudo-reference. The difference between two images, calculated with Delta E formula of CIEDE2000 [7], indicates how much colors are distorted in false color pixels. When aggregating the color differences into the score, we considered several features to reflect the characteristics of false color artifact. First, we added up the color differences using power mean formula [11] with exponent p to increase sensitivity to intense artifacts than minor frequent artifacts. To assign higher weight to significant color distortions, we suggested that the exponent p should be greater than 1, possibly 2 or higher. Furthermore, as most false color artifacts occur dominantly on the edge, we used the edge image of grayscale image (Y') [9] as a weighted map. This map was obtained with Sobel filter [12], and we combined two directional edge maps with l-2 norm. Finally, we formulated the false color score equation as:

$$Score = \left[\frac{\sum \Delta E(x,y)^p \text{edge}(x,y)}{\sum \text{edge}(x,y)} \right]^{\frac{1}{p}}, \text{ where } p > 1 \quad (10)$$

where $\Delta E(x,y)$ is the color difference of each pixel between the original image and the pseudo-reference image R .

Experiment Designs for Analysis

False color, unlike general image quality components (e.g., sharpness, resolution, etc.), could not be evaluated enough by general image quality metrics. Therefore, we verified our proposed score by figuring out the correlation with qualitative scores. In order to gather the qualitative scores, we conducted an investigation for human visibility of false color artifact, categorizing the visibility factors of false color into two main aspects: strength and quantity. Considering these factors, we constructed a subjective IQA survey that enabled us to analyze the impact of each factor on human perception.

Mean Opinion Score (MOS) Investigation

To gather the qualitative scores, we conducted a subjective IQA survey and obtained the mean opinion scores (MOS). For the survey, we generated a dataset using false color simulation, in which color dot noises were added to undistorted images along the edges. With the two main aspects of false color previously defined, we independently varied these factors in the simulation across 4 different levels. Each of the 16 combinations of factors was then applied in 5 sample references, resulting in 80 distorted images were created (Fig. 6a).

To assess how sensitive human sight is to strength and quantity factors separately, we reorganized 80 images in respect of each artifact factors and repeated for twice to constitute two types of test subsets (Fig. 6bc). We designed each subset consisting of 20 conditions of tests with 4 quality levels of images per test. For

instance, in the strength test subset (Fig. 6b), 4 different strength levels of images were assigned in a test and there were totally 20 different tests. In the MOS score survey, we presented those 4 images in a sequence to 36 IQA experts, requesting them to rate each image in the range from 1 to 3. A rating of 1 indicates good image quality with no visible artifacts, 2 represents a marginal level, and 3 signifies that the image quality is definitively poor. This survey structure allowed us to collect MOS values under specific false color factors.

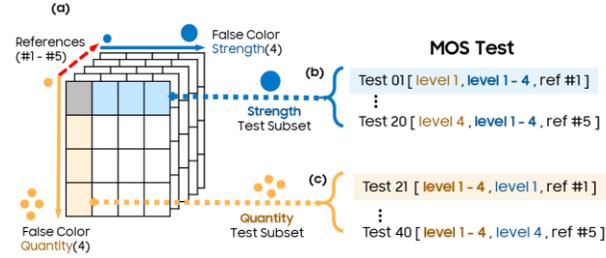


Figure 6. Structure of mean opinion score (MOS) survey dataset: (a) false color dataset, (b) strength test subset, and (c) quantity test subset.

T test for Liner Correlation

In our experiments, Pearson correlation coefficients (PCC) were calculated to measure the linear relationship between MOS and our suggested metric score. We calculated a single PCC from 20 images for each individual artifact levels within a single test subset, resulting in 8 PCC values, (Fig. 7). Furthermore, to confirm that a certain PCC proved the high correlation, we employed the t -test for the correlation coefficients [13]. In [13], the t -statistic for PCC evaluation is denoted as follows:

$$t = \rho \sqrt{\frac{N-2}{1-\rho^2}}, \text{ degrees of freedom (d.f.)} = N - 2 \quad (12)$$

where ρ is a Pearson correlation coefficient and N is a sample size, which was 20 in our experiments. If the absolute value of t derived from a certain PCC was greater than the critical, we could conclude that the PCC was sufficiently high to prove the linear relationship of our scores and MOS values. In other words, each $|t|$ indicates a confidence level of the linear relationship under the specific false color quality condition.

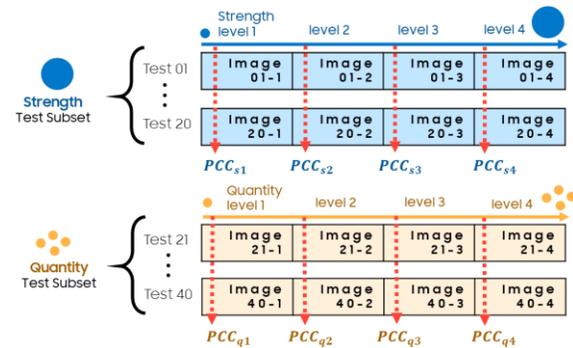


Figure 7. Pearson correlation coefficients (PCC) are gathered from each level and each subset of 20 images. Total 8 PCC results are obtained from the dataset.

Experimental Results and Discussion

Table 1. and Table 2. show the t -test results for linear correlation between MOS values and our proposed false color scores as described in the further section. This results demonstrate the linear relationship between MOS values and false color scores with various exponent p values in the equation (10). Through these results, we selected the optimal value of p .

In Table 2., all t results for the quantity test subset exceed 3.92, which is the critical value for $\alpha=0.0005$ (Table 3.), and it means our metric has a linear relationship with MOS in a 99.9% confidence level. According to these results, our scores are well matched regardless of p in terms of false color quantity.

Table 1. T-test results for strength test subset.

Strength level	t					
	$p=4$	$p=3$	$p=2$	$p=1$	$p=1/2$	$p=1/3$
1	3.25	2.72	2.27	1.81	1.38	1.18
2	4.40	4.37	4.01	2.98	2.10	1.75
3	6.35	7.51	8.10	5.71	3.66	2.98
4	6.09	7.13	7.44	5.03	3.02	2.36

Table 2. T-test results for quantity test subset.

Quantity level	t					
	$p=4$	$p=3$	$p=2$	$p=1$	$p=1/2$	$p=1/3$
1	17.0	16.9	16.5	13.2	8.36	6.41
2	14.7	15.0	15.3	13.0	8.41	6.45
3	16.7	16.8	16.3	12.2	7.48	5.72
4	16.1	15.6	14.2	10.0	6.14	4.68

Table 3. T- critical value table for 2-sided distribution

d.f.=18	t (2-sided)	
	$\alpha = 0.05$ (90.0%)	$\alpha = 0.0005$ (99.9%)
	1.73	3.92

In contrast, the results of the strength test subset in Table 1. show that it is difficult to confirm the linear relationship regardless of the score exponent parameter. Comparing the scatter plot of the strength test subset and another plot of the quantity test subset (Fig. 8), it is evident that MOS values in strength subset varied more depending on strength levels than quantity levels. It demonstrates that the human sights are more influenced by the intensity of false color signals than the amount of them.

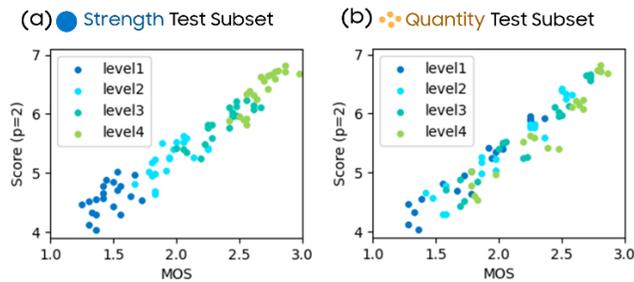


Figure 8. Scatter plots of mean opinion score (MOS) and false color score ($p=2$) results from (a) the strength test subset, and (b) the quantity test subset, with dots colored according to artifact levels.

Therefore, we set p to above 1 in the equation (10) to prioritize the strong artifact signal than other minor signals in the metric score. In Table 1., we can observe all t ($p > 1$) results exceed 1.73 and it can confirm the linear correlation with 90.0% confidence level (Table 3.). That is the reason why we adopted the power mean formula and set the p to the number above 1 in equation (10). In addition, we recommended setting p to 2 because t -test results, each indicating a level of reliability, are generally high across most strength levels. As a result, we could confirm that our method scores are aligned well with the MOS in 90.0% confidence level as we verified our method with a statistical hypothesis test.

Fig. 9 shows a part of the results in our experiments, while the object color N set to 1 and p set to 2. It is evident that our method can quantify artifacts in context of human visibility. In Fig. 9e-h, the pseudo-references show that false color artifacts are well corrected through our method and result scores show an increase as the image quality degrades due to artifacts. It means that the scores exhibit a similar trend to human perception. In essence, the results demonstrate that our method successfully reflected subjective image quality opinions and quantified them to certain objective values effectively.

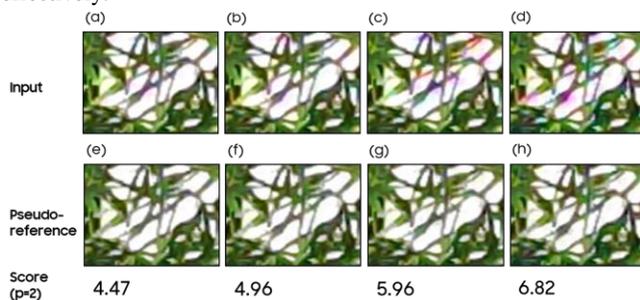


Figure 9. Sample results of experiments: (a-d) input images, (e-h) pseudo-references of (a-d), and false color scores, where a high score indicates poor image quality.

Conclusion

In summary, we proposed a no-reference IQA metric for false color to overcome the constraints of an input in false color IQA scenarios as being restricted to black-and-white charts or requiring real inferences. To achieve this, we inferred a pseudo-reference to substitute for the ideal sensor output. The pseudo-reference was generated using unsupervised clustering methods which detected outlier colors and replaced them with inlier colors. With our method,

we could measure artifacts by comparison of an input and pseudo-reference with regard to human visibility. Because the score is highly correlated with qualitative scores, our method would serve as an objective indicator for false color IQA in the CIS industry. Furthermore, when employed in various scenarios such as natural scenes, it is expected to contribute to ISP parameter optimization through the evaluations of ISPs

References

- [1] A. Eskicioglu and P. Fisher, "Image quality measures and their performance," *IEEE Transactions on Communications*, vol. 43, no. 12, pp. 2959-2965, 1995.
- [2] Z. Wang, A. Bovik, H. Sheikh and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600-612, 2004.
- [3] J. Lee, S. June, J. Bang, S.-S. Kim and J. Yim, "Color image distortion assessment based on synthetic ground truth recovery," *Electronic Imaging*, vol. 34, no. 9, pp. 342-1--342-5, 2022.
- [4] C. K. Reddy and B. Vinzamuri, "Data clustering," in *A survey of partitioned and hierarchical clustering algorithms*, Chapman and Hall/CRC, 2018, pp. 87-110.
- [5] H. Deng and J. Han, "Probabilistic models for clustering," in *Data Clustering*, Chapman and Hall/CRC, 2018, pp. 61-86.
- [6] M. E. Celebi, "Improving the performance of k-means for color quantization," *Image and Vision Computing*, vol. 29, no. 4, pp. 260-271, 2011.
- [7] M. Luo, G. Cui and B. Rigg, "The development of the CIE 2000 colour-difference formula: CIEDE2000," *Color Research & Application*, vol. 26, no. 5, pp. 340-350, 2001.
- [8] S. Romani, P. Sobrevilla and E. Montseny, "Variability estimation of hue and saturation components in the HSV space," *Color Research & Application*, vol. 4, no. 261-271, p. 37, 2012.
- [9] C. Poynton, "Color FAQ-Frequently Asked Questions Color," 28 11 2006. [Online]. Available: <http://www.poynton.com/notes/colour-and-gamma/ColorFAQ.html>. [Accessed 11 2 2024].
- [10] G. Gallego, C. Cuevas, R. Mohedano and N. Garcia, "On the Mahalanobis distance classification criterion for multidimensional normal distributions," *IEEE Transactions on Signal Processing*, vol. 61, no. 17, pp. 4387-4396, 2013.
- [11] P. Bullen, "The Power Means," in *Handbook of Means and Their Inequalities*, vol. 560, Dordrecht, Springer Netherlands, 2003, pp. 175-265.
- [12] N. Kanopoulos, N. Vasanthavada and R. L. Baker, "Design of an image edge detection filter using the Sobel operator," *IEEE Journal of solid-state circuits*, vol. 23, no. 2, pp. 358-367, 1988.

- [13] M.-T. Puth, M. Neuhäuser and G. D. Ruxton, "Effective use of Pearson's product-moment correlation coefficient," *Animal behaviour*, vol. 93, pp. 183-189, 2014.

Author Biography

Subin Han received her B.S. degree in electrical engineering from Korea University in 2020. Since 2021, she has worked in Samsung Electronics, Republic of Korea, as an engineer. Her work has focused on computer vision, image/video signal processing, and image quality metric.