https://doi.org/10.2352/ISSN.2470-1173.2021.9.IQSP-196

Understanding the importance of Artistic Aspect of Camera Image Quality Tuning and Quantifying the Artistic Attributes by using Scientific Principles

Vickrant Zunjarrao, Ranga Burada, Sharvin Shobhane, Lekha Lokhande, Anusha Devi T. T.
Microsoft Surface Team, One Microsoft Way, Redmond, WA

Abstract

The camera tuning process contains multiple stages under the image signal processing (ISP) pipeline through which the RAW image gets processed and displayed. The objective image quality is well defined and important when tuning the color imaging pipeline in the Camera. The ISP pipeline is tuned by optimizing objective criteria, resulting image and video may not be aesthetically appealing to end-users and may not be sufficient to provide the best visual experience. There are certain key artistic factors affecting the overall user experience in the process of ISP pipeline tuning. However, currently there is no industry standard for artistic image quality (IQ) quantification and is mostly based on expert based subjective evaluation. In this paper, we emphasize the importance of artistic attributes and to quantify them with two studies. First study focusses on importance of artistic attribute by two alternate forced choice pairwise comparison user study on artistic vs objective tuning. Based on this study, artistic tuning is statistically significantly better than objective tuning with 95% confidence interval based on fisher least significant difference test. Second study develop a method to quantify holistic IQ. The novel mathematical equation has been formulated to calculate the weight of different IQ attributes which are impacting on overall IQ as a first step. Rank based user study was performed on expert and non-expert users to find their preference on different attributes with the use of novel mathematical equation. In this equation, user preferences were converted into weights of artistic attributes in the holistic IQ. The study shows that color saturation and memory color attributes have higher impact for both expert and non-expert user with weightage of 14.04% and 13.62% in holistic IQ, respectively. This study validates the importance of the artistic approach and our first step to quantify these artistic attributes in a scientific way. It also exhibits the need of a novel image quality assessment criteria to tune and validate the final visual experience of the consumer camera. Especially, in the era of Artificial Intelligence (AI), quantifying artistic attributes is far more important than ever.

Introduction

In this revolutionary camera world, mobile cameras have gained significant attention in recent times as the usage has increased dramatically. The camera IQ became essential to standout in this competitive market to satisfy consumer's needs. Image quality is dependent on camera system which consists of hardware components and software stacks. ISP pipeline and its optimization play a significant role to improve overall IQ. ISP pipeline is one of the main components of the camera system which commonly includes shading correction, black level subtraction, denoising, demosaicing, color transformations, tone mapping, edge enhancement, and image encoding for final display. Image quality tuning can be objective or combination of objective and artistic. The objective IQ tuning is based on the guidance of the standards like IEEE CPIQ [1], Skype certification [2], ISO 12233 [3], ISO 15739 [4], etc. In various organization, IQ engineers customize these standards based on product requirements. This objective criterion takes an image to a certain accepted level, but this is not enough to sustain in this competitive consumer market. Effort to enhance aesthetic appearance of the images to make them more appealing to the users is the need of the hour. The aesthetic appearance is mostly analyzed by the human cognitive skills based on non-linear preference of human experience like the interaction between emotional-valuation, sensory motor, and meaning-knowledge neural systems [5] which is affected by different IQ attributes.

The ideal image quality assessment (IQA) consists of objective subjective testing and user studies. The objective IQA is generally divided into three groups: full-reference image quality assessment (FR-IQA) algorithms, e.g.,[6], [7], reduced-reference image quality assessment (RR-IQA) algorithms, and no-reference/blind image quality assessment (NR-IQA) algorithms, e.g.,[8], [9]. Subjective image quality assessment is based on a visual assessment by the group of IQ experts. IQ experts use reference and non-reference methods to provide feedback in a non-numeric way. User study is based on psychophysical experiments and incorporating its results to enhance images. With the help of different methods like forced choice pairwise comparison, pairwise similarity judgements and difference mean opinion score, etc. [10] we can conduct preferential study based on psychophysical principles.

There are different machine learning and deep learning methods [11] to evaluate quality or artistic quality of images. Conventionally, the overall image quality contributed by different IQ attributes are usually summed by metrics like Minkowski [12]. Especially in the era of deep learning, it is possible to automatically predict the perceptual image quality[13] and enhance it [14]. The quantification of holistic IQ based on its attributes have been persistent problem in modern day IQ and none of the method can be directly used to optimize visual experience of the end use of mobile consumer cameras. Hence, the new method to quantify different image quality attributes is needed.

In this paper, we are going to address the following research questions:

1) What is the user's preference on artistic vs objective tuning?
2) What are the most impactful artistic IQ attributes and how to quantify different artistic IQ attributes based on user's preference?

To answer these questions, we have conducted two different subjective user studies. The first part of the study was performed to analyze the user preference on artistic approach vs objective tuning criteria images under different condition based on two alternate forced choice (AFC) pairwise comparison [15]. The second part of the user study provide an innovative method to identify user's preference and to quantify the impact of artistic attribute on holistic IQ (overall visual experience) based on rank-order method [16]. The study will demonstrate a new approach to quantify the IQ and need of novel image quality criteria which will help in assessment of final visual experience of the consumer camera.

The rest of the paper is organized as follows: in methodology, a brief introduction on experimental setups for two different user studies are described. It is followed by results and discussions where inferences from the two different studies were examined. Finally, the first study is concluded with importance of different artistic attributes and the second study results in the new novel equation to quantify the artistic attributes.

Methodology

Part one of Study: Comparative Study Between Objective Vs Artistic Image Quality

The goal of part one study is to demonstrate the importance of artistic image quality. A user study was conducted to evaluate if participants prefer artistically tuned images over the objectively tuned images.

Separate tuning process has been followed to capture objectively and artistically tuned images. Tuning process contains multiple stages under the ISP pipeline through which the whole image gets processed and displayed.

The objective tuning is generally based on different numeric standards outlined by ISO, IEEE, and CIE. With the help of these standards, IQ Engineers working at various organizations create their own criteria for objective IQ evaluation per the product needs.

Artistic tuning is done considering the educated guesses of IQ experts where it is a combination of numeric and visual feedback based off reference and / or non-reference methods. This is a closed loop feedback system. It becomes very important to involve highly trained and skilled engineers for the feedbacking session to achieve an optimal level of image quality. In exceptional cases or unique scenarios artistic tuning might not satisfy the objective criteria but in this case; visuals supersede numeric.

Experiment

We set up the studies to learn about overall population's preference in the various use case scenarios for two set of images:
Set 1: Objectively Tuned Images
Set 2: Artistically Tuned Images

The alternate forced choice comparison method was used for this study. Images were randomly placed during the studies to avoid special pattern in participant's mind while choosing images. The online portal was used to conduct this study as shown in Figure 1.

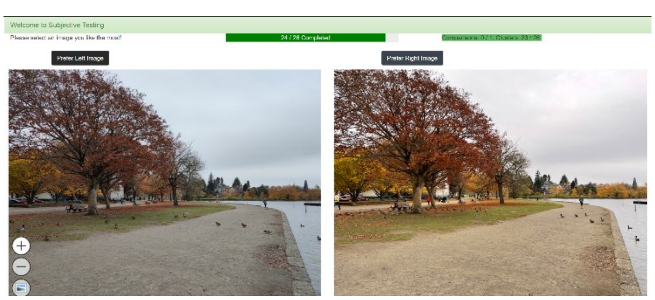
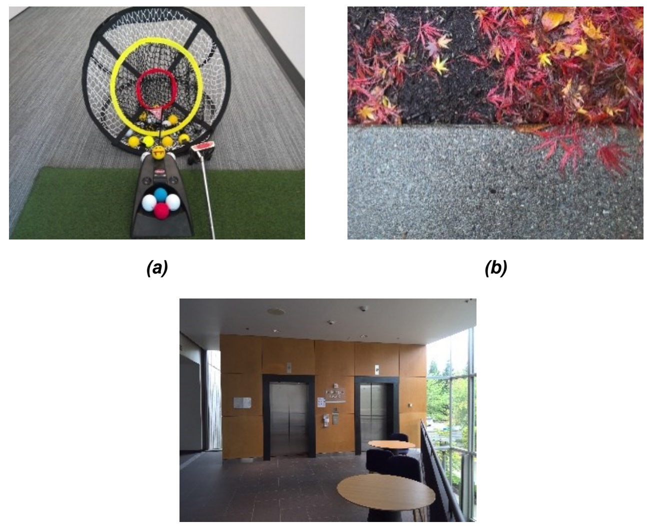


Figure 1. *Online portal interface*

Scene collection was one of the most important steps while setting up the study. We came up with the scenes and frames which had various consumer camera use case scenarios per these standard categories:

- Indoor, outdoor, Mixed light
- High lux (Above 800 lux), mid lux (100 to 800 lux), low lux (below 100 lux) level
- Different standard illuminants: D65, D50, TL84, U30, A

We carefully selected 20 scenes to satisfy above categories. The examples of different scene categories are shown in Figure 2.



(a) *(b)*

(c)

Figure 2. *Example of different types of scene (a) Indoor scene (b) Outdoor scene (c) Mixed light scene*

Indoor Scene

An example of indoor scene representing highly saturated colors in the scene is shown in Figure 2 (a). This scene has U30 fluorescent lighting condition with mid lux. It also has variety of colors to evaluate hue and saturation aspects.

Outdoor Scene

The typical high lux outdoor cloudy D65 scene as shown in Figure 2 (b). It is a good representation of texture, colors, and contrast.

Mixed Light Scene

The scene shown in Figure 2 (c) has a glass window to welcome outside light and indoor environment. This is a good scene to represent the mixed light scenario as the scene lighting conditions change with changes in the weather outside. This scene has U30 from indoor fluorescent lights and D65 lighting from the window with high lux. It also has variety of colors to evaluate hue and saturation aspects.

Participants were asked to register and complete their profile; they were asked about their age, gender, and geographic up-bringing. During this test, side by side comparisons of image pairs were displayed. Users were asked to choose based on two alternate forced choice pairwise comparison method. Users were requested to use professionally color calibrated surface screen to conduct the study which reduces discrepancies in viewing conditions.

The data was collected from the population within the organization irrespective of their age and gender. Participants here refer to the target audience - a team of experts, trained image quality professionals, scientists, non-IQ engineers, artists, and others.

Part two of Study: Quantification of different attributes

The objective of part two of our study was to identify the most impactful artistic IQ parameters and quantify these parameters based on user preference.

From an extensive literature survey[17] and based on camera tuning team's experience on dealing with the end user's preferences, we found out various image quality attributes which a regular user perceives. We also observed the attributes most frequently pointed out by users. Based on these results we were able to narrow down to eight most perceived image quality attributes as shown in Figure 3. We grouped multiple similar attributes under a single attribute as shown in Figure 3.

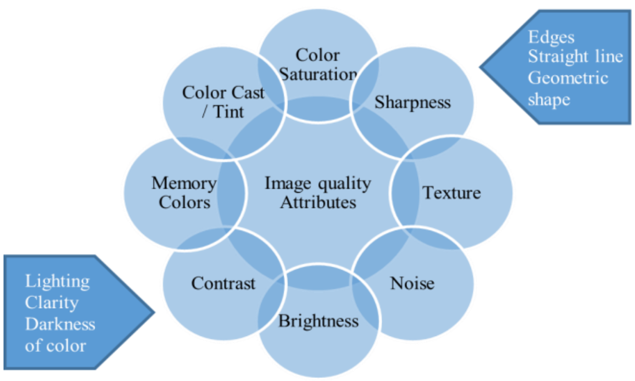


Figure 3. *Most commonly perceived image quality attributes*

Our aim was to find out the weightage of these individual image quality attributes observed by users. These weights may be different for each individual observer. The experiment was set up in such a way that each user was shown 12 different slides. An example of the slide is shown in Figure 4. We randomly selected the images and made sure that a single IQ attribute was not dominating a specific slide based on our personal opinion. These IQ attributes were presented to the participant in random order.

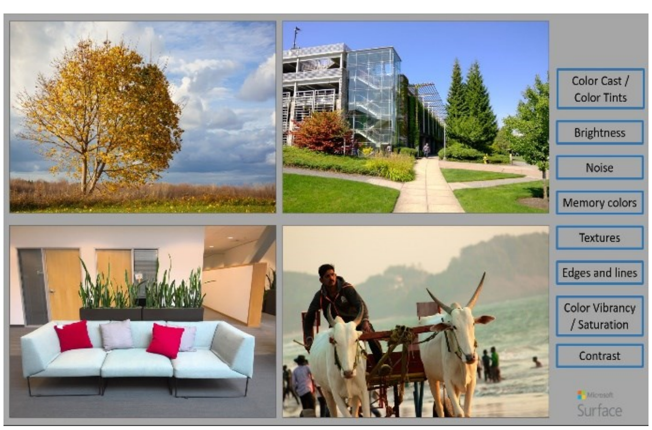


Figure 4. *Example of the slide shown to participants*

The participant was educated about these eight IQ attributes with simple definitions and example images to increase the amount of concordance and reliability in the data. This explanation was especially useful for non-expert participants.

The definition of the parameters was as follows:

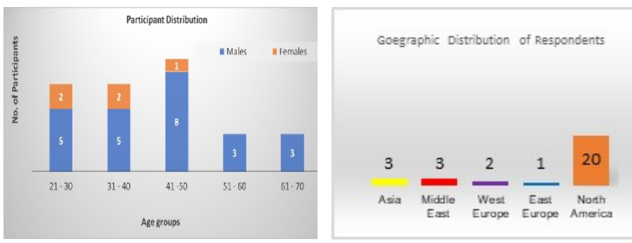
- Color Vibrancy / Saturation: Color Vibrancy or Saturation is defined as the intensity of color in an image. Colors with higher saturation appear to be purer whereas colors with lower saturation appear to be tending towards grey. In color theory, color saturation is one of the parameters of Hue, Saturation, Value (HSV) color space [18].
- Sharpness: Edges and straight lines help to provide structure and clarity in an image. Clear, defined lines can help to create a more aesthetically pleasing frame for a scene. Sharpness helps to provide separation of subjects from the background. In an image containing buildings and windows, the definition of edges of the subject is perceived as sharpness. Metric used to measure sharpness is Modular Transfer Function (MTF) [19].
- Texture: Texture gives an idea of the feel of an object in the scene in the viewer's mind. Seeing the details that provide texture help to evoke the feel of a scene such as the smoothness of glass or the coarseness of sand. The metric used to gauge texture is Acutance[20].
- Noise: Noise is the appearance of small particles in an image that is visible as either different colors or different patches of brightness. The patch size and color intensity can vary depending on the type of noise. Noise, when constrained, can help give the feel of classic film photography grain in an image, providing a "vintage" effect (desired noise). Under certain limits of noise, the feeling of texture can also be created, however, excessive noise can ruin the clarity of an image (undesired noise).
- Brightness: Brightness is defined as an attribute which makes the image look more or less intense. It can be perceived as the luminance of the image. Bright image makes it look appealing whereas images with low brightness look dull. Brightness is a co-ordinate on the Hue Saturation Lightness (HSL) scale where Lightness corresponds to brightness[21].
- Contrast: Contrast is the separation of light and dark areas in a scene. The most common contrast is between white and black to give separation of two areas of a scene. In some scenes, contrast can help to provide a feeling of depth, especially when using shallow depth of field, which provides the focus onto the main subject of an image. Higher contrast is perceived to produce images which have higher clarity, and this can be modified by changing the Tone Curve.
- Memory Color: Memory colors are colors that are immediately recognizable due to their association. Examples of these colors are red apples, blue skies, green grass, yellow bananas, and skin tones. Although there may be subtle shifts in the color, whenever an object is mentioned, such as an apple, it is instantly associated with the color red.
- Color Cast / Tint: Color casts in an image are subtle shifts of a single color across an entire image that help evoke the feel and emotion of a scene. One of the most common shifts is between orange and blue, which is present with sunlight and shifts depending on the time of day. This attribute has a strong association with the overall white balancing of an image. The final white point of an image has to be modified for an image to have a specific color tint.

Each participant was shown the 12 slides one by one and was asked to see all four images together as one. These slides had a list of all the IQ attributes as a sidebar for quick reference. Then, the user was asked to note down up to eight most prominent image quality attributes that they spot. The order in which the participant mentions the attribute is important as it will help us understand that which attribute did the participant perceive to have a strong impact on that particular slide.

Results and Discussion

Part one of Study:

Total 29 participants responded to the user study. The participant distribution was 83% male and 17% female as shown below. The geographic distribution of the participants is also mentioned as shown in Figure 5.



(a) (b)

Figure 5. *Participant distribution for part one study (a) per age group (b) per geographic location*

In summary, artistic image quality has been preferred over objective tuning across various age groups and gender based on 20 AFC pairwise comparisons using Fisher Test. In eighteen scenes, artistic tuning performed significantly better than objective tuning. Only in two scenes artistic tuning did not perform better than objective tuning. Those two scenes consisted of indoor less chromatic scene and person with outdoor scene.

Further, Kendall's concordance test was conducted to see the agreement between observers. In this study, Kendall's coefficient of concordance W, was found to be 0.77 where W value ranges from 0 (no agreement) to 1 (perfect agreement). Higher value of W indicates that there is stronger agreement between the participants among themselves for all scenes and strengthens our conclusion.

As shown in Figure 6; approximately 72 % participants had higher preference towards Artistic tuning and 17% of overall judges had higher preference for Objective tuning. In this study, there were 10% of subjects who were not conforming towards any of the tuning categories.

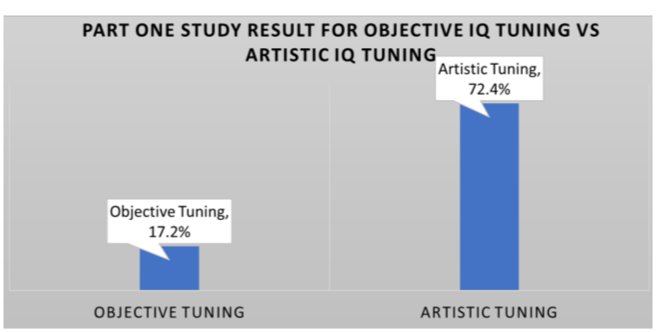


Figure 6. *Participant Preference on Objective Vs Artistic Tuning*

Based on statistically significant difference, Fisher test is used to determine whether there is a significant perceived difference between objective tuning and artistic tuning [22]. It was confirmed that Artistic image quality was better than objective image quality with 95 % confidence interval.

Part two of Study:

The objective of this study was to quantify the overall image quality based on certain IQ attributes and find the contribution of each attribute in the overall image quality.

For this study, 30 non expert and 12 experts participated as shown in Figure 7. We grouped the participants in these two groups as we wanted to understand about the similarities and differences in their perception of IQ and impact on the calculation of weights. Experts were defined as participants who work in or have experience in camera image quality domain. Whereas non-experts were defined as mobile or professional camera users who may or may not pursue photography as a hobby and do not expertise in IQ domain.

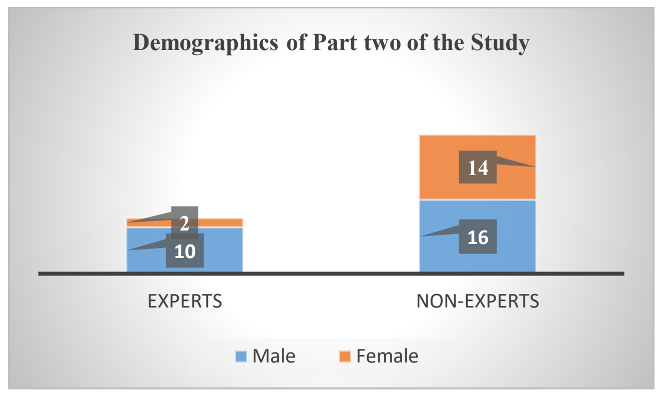


Figure 7. *Population distribution in part two of the study*

Based on the study of 42 participants, we can plot the frequency of mentions for all the attributes as shown in Figure 8. We can observe that color saturation and memory color have been mentioned the most followed by nearly equal mentions of all other attributes except noise.

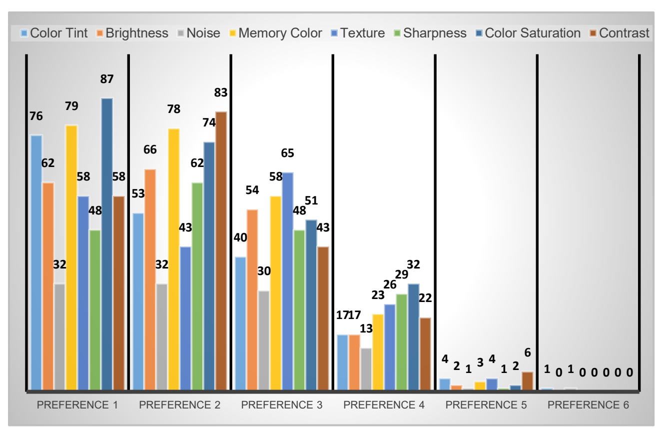


Figure 8. *Number of mentions of each attribute vs preference level*

The initial step is to measure the frequency of mentions of all the attributes at each preference level. The maximum number of IQ attributes mentioned may be different for different participants. Thus, we needed to come up with a weight distribution method which will try to give unbiased weightage to each attribute irrespective of the number of attributes a user mentioned for each slide. This was achieved by using the formula shown in equation (1).

$$W(i,n) = \left[\frac{a}{N} + \sum_{j=i}^{n} \frac{(1-a)}{j \times 2^{(n-1-\max(0,j-2))}}\right] X\,100 \qquad (1)$$

N = Total number of attributes
a = Maximum percentage to distribute initial offset
n = maximum number of preferences mentioned by a user
i = preference level currently iterated

To calculate the weights of each IQ attribute, we began with the frequency data for each individual user. This data was then normalized and an offset of one was added to the frequency data. The total number of attributes (N) for this study is 8. Every image generally has all these attributes present, and we thought of giving them an equal initial weight. We experimented with the value of a to be 50%, 33%, and 20% and concluded that 33% distributed in 8 attributes gives justified initial weight of 4.17% to every attribute. Suppose a participant never mentioned more than 4 attributes, then all the IQ attributes in preference levels 5 to 8 will have equal weightage of 4.17% due to this offset, rather than completely disregarding those preference levels as shown in Figure 9. This frequency data is then multiplied with the weights calculated by our novel Rank order-based weight formula.

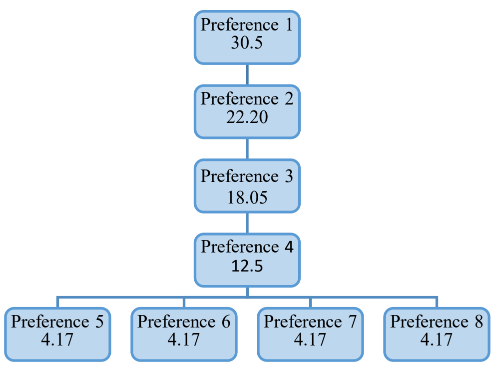


Figure 9. Weight distribution based on a participant mentioning maximum 4 preferences.

Output of this calculation is the weightage of each IQ attribute for that particular user per slide. Then, we average these weights in all the slides together to obtain the weights for all the IQ attributes for that particular user. We found the average weights of all the IQ attributes for overall participants, experts, and non-experts separately and tabulated the results in Table 1.

Table 1: Average calculated weights of each IQ attribute in overall Image Quality

Rank	Image Quality Attribute	Overall Average Weights for 42 users (%)	Weights for 30 non-experts (%)	Weights for 12 experts (%)
1	Color Saturation	14.04	13.69	14.91
2	Memory Color	13.62	13.32	14.38
3	Contrast	12.86	13.12	12.21
4	Brightness	12.55	13.09	11.20
5	Texture	12.40	11.77	13.96
6	Color Tint	12.30	12.38	12.08
7	Sharpness	12.22	12.11	12.49
8	Noise	10.01	10.51	8.76

Color saturation, memory color followed by contrast have the highest weightage based on the average of 42 participants studied. We can observe that experts and non-experts have preferred color saturation and memory color as their top two choices and thus can infer that these parameters have a higher impact on overall artistic image quality.

Conclusion

This study showcases the importance of the artistic attributes and our scientific approach to quantify the effect of artistic attributes on the final IQ. The first part of this study demonstrated the importance of the artistic approach and proved that artistic tuning is statistically significantly better than objective tuning with 95% confidence interval based on fisher LSD test. In the next part of this study, the novel mathematical equation has been proposed with the rank-order based user studies to calculate the weight of different artistic IQ attributes which are impacting on overall IQ. We also quantified weights for experts and non-experts separately to understand if there is any difference between these two populations. Color saturation and memory color were found to have higher significance on visual IQ for both experts and non-experts from this study. This quantification method of artistic attributes can be used to optimize camera imaging pipeline during tuning process to enhance visual experience of an end-user. Future work will focus on finding out statistical model which will help to quantify each artistic attribute eventually to quantify overall image quality and automate IQ tuning for consumer mobile camera using artificial intelligence.

References

[1] "IEEE 1858-2016 - Camera Phone Image Quality (CPIQ) - Corrigendum 1," 2017.
[2] K. T. Chen, C. Y. Huang, H. Polly, and C. L. Lei, "Quantifying Skype user satisfaction," in *Computer Communication Review*, Oct. 2006, vol. 36, no. 4, pp. 399–410.
[3] "ISO 12233: Resolution, Photography - Electronic still-picture," 2005.
[4] "ISO 15739: Imaging Noise, Photography - Electronic still-picture," 2003.

[5] A. Chatterjee and O. Vartanian, "Neuroscience of aesthetics," *Ann. N.Y. Acad. Sci*.
[6] A. Alaei, · R Raveaux, and · D Conte, "Image quality assessment based on regions of interest," *Signal, Image and Video Processing*, vol. 11, pp. 673–680, 2017.
[7] S. Bianco, G. Ciocca, and F Marini, "Image quality assessment by preprocessing and full reference model combination," in *Proceedings of SPIE*, Jan. 2009, vol. 7242, p. 72420O.
[8] A. Mittal and R Soundararajan, "Making a completely blind image quality analyzer," 2012.
[9] S. Mahmoudpour and M. Kim, "No-reference image quality assessment in complex-shearlet domain," in *Signal, Image and Video Processing*, Nov. 2016, vol. 10, no. 8, pp. 1465–1472.
[10] P. Mohammadi, A. Ebrahimi-Moghadam, and S. Shirani, "Subjective and Objective Quality Assessment of Image: A Survey," 2014.
[11] Y. Deng, C. Loy, and X. T. Magazine, "Image aesthetic assessment: An experimental survey," *IEEE Signal Processing*, pp. 80–106, 2017.
[12] M. P. Eckert and A. P. Bradley, "Perceptual quality metrics applied to still image compression," *Signal Processing*, vol. 70, no. 3, pp. 177–200, Nov. 1998.
[13] S. Bianco, L. Celona, P. Napoletano, and R. Schettini, "On the use of deep learning for blind image quality assessment," *Signal, Image and Video Processing*, vol. 12, no. 2, pp. 355–362, Feb. 2018.
[14] A. C. Bovik, "Automatic prediction of perceptual image and video quality," *Proceedings of IEEE*.
[15] R. K. Mantiuk, A. Tomaszewska, and R. Mantiuk, "Comparison of Four Subjective Methods for Image Quality Assessment," in *Computer Graphics Forum*, Dec. 2012, vol. 31, no. 8, pp. 2478–2491.
[16] M. Danielson and L. Ekenberg, "Rank ordering methods for multi-criteria decisions," in *Lecture Notes in Business Information Processing*, 2014, vol. 180 LNBIP, pp. 128–135.
[17] G. Tuomas Leisti, J. Radun, T. Virtanen, R. Halonen, and G. Nyman, "Subjective experience of image quality: attributes, definitions, and decision making of subjective image quality," *Proceedings of SPIE*, 2009.
[18] G. H. Joblove and D. Greenberg, "Color spaces for Computer Graphics," in *Comput Graph (ACM)*, Aug. 1978, vol. 12, no. 3, pp. 20–25.
[19] G. Boreman, *Modulation transfer function in optical and electro-optical systems*. 2001.
[20] D. Präkel, *The visual dictionary of photography*. Ava Publishing, 2010.
[21] S. Bezryadin, P. Bourov, and D. Ilinih, "Brightness Calculation in Digital Image Processing," *International Symposium on Technologies for Digital Photo Fulfillment*, vol. 2007, no. 1, pp. 10–15, Sep. 2014.
[22] Freeman, Jenny V., and Steven A. Julious, "The analysis of categorical data," 2007.