https://doi.org/10.2352/ISSN.2470-1173.2020.9.IQSP-316

Effective ISP Tuning Framework Based on User Preference Feedback

Cheoljong Yang, Jinhyun Kim, Jungmin Lee, Younghoon Kim, Sung-Su Kim, TaeHyung Kim and JoonSeo Yim
Samsung Electronics Co., Ltd; Hwaseong-si / Gyeonggi-do / Republic of Korea

Abstract

This paper presents an effective tuning framework between CMOS Image Sensor (CIS) and Image Signal Processor (ISP) based on user preference feedback. One of key issue in ISP tuning is how to apply individual's subjectivity of Image Quality (IQ) in systematic way. In order to mitigate this issue, we propose a framework that efficiently surveys user preference of IQ and select ISP parameter based on those preferences. The overall processes are done on large-scale image database generated by an ISP simulator. In preference survey part, we make clusters that consist of perceptually similar images and gather user's feedback on representative images of each cluster. Next, for training user preference, we train a DNN model according to general preference, and fine-tune model to optimize individuals preference based on user feedback. The model provides ISP candidate most similar to the preferences. In order to assess performance, the proposed framework was evaluated with a state-of-art CIS and ISP system. The experimental results indicate that the proposed framework converges the IQ score according to user feedback and find the ISP parameters that have higher quality IQ as compared with hand-tuned results.

Introduction

As camera performance has become a key feature of mobile phone, manufacturers are putting more resource for getting better image quality. The RGB image of the mobile camera is the processed image of the raw data obtained from the sensor through the ISP as shown in **Figure 1**. Therefore, for high quality image acquisition, it is very important to use not only a good sensor but also to set up an optimal ISP tuning parameters. Since most of commercial ISPs does not public their internal algorithms in the form of a black-boxes, however, the tuning is an iterative process of a manual tuning and an IQ verification until the output IQ meets a level of satisfaction. Therefore, the tuning process is very difficult and inefficient, and is dependent on the experience of IQ experts.

To make an efficient optimization of the IQ tuning, recently limited efforts have been made to develop an effective ISP tuning method based on machine learning. Nishimura *et al*. proposed automatic tuning framework based on non-linear optimization algorithm [1]. For optimization, he generates reference images for de-noising and sharpening ISP blocks, and finds the best ISP parameters that minimize loss function of pixel-by-pixel Mean Squared Error (MSE) between the reference image and simulated images. Tseng *et al*. proposed an algorithm of differentiable mapping between a parameter space and evaluation metrics, parameterized by Convolution Neural Network (CNN) [2]. Both of their detailed methods are different, but they commonly tune the ISP parameters until the IQ reaches an objective value of reference

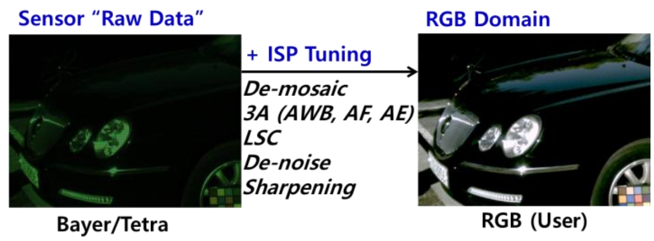

Figure 1.The CMOS and ISP flow.

images. These methods have shown that a machine learning methods can make ISP tuning process more efficient in an environment where objective target (e.g. reference image, IQ metric) can be set.

Besides the efficiency of the tuning process, a tuning method that reflects the subjectivity of IQ is another issue. Most of the tuning process is done by ISP tuning experts, so it is difficult to reflect the subjectivity of IQ. It has been shown that the subjectivity of human image quality varies, and development of ISP tuning reflecting individual subjectivity is required [3]. Juan *et al.* show that personal IQ preference can be grouped according to such other's similar preferences and proposed image enhancement method based on the grouped preference [4]. The subjectivity of IQ can be divided into several clusters, and ISP tuning method is applied efficiently by applying the ISP tuning method to share the same parameters among users belonging to the same group. Both of prior works have proved effective in reflecting subjective in the IQ tuning. However, the techniques were used only for color contrast correction, and it is necessity to extend the tuning range in order to apply it to the mobile camera ISP environment.

In this study, we present an effective ISP tuning system based on user preference feedback with perceptual modeling. Inspired by previous studies, we propose machine learning based efficient tuning method using clustering of IQ subjectivity. More specifically, the proposed system composes a DNN model of IQ metric and fine-tunes the model via user preference feedback. In main proposal, we focus on a machine learning based effective method of clustering user feedback and generation of preference model. Finally, the framework provides the tuned ISP parameters according to user's preference. Prior works about ISP tuning methods mainly focus on an optimization process not a method of incorporation of IQ subjectivity. Unlike prior works, we focus not only effective methods for tuning but how to reflect preference of IQ.

The paper is organized as follows. Section *User preference based ISP tuning framework* presents the methodology of the main proposal and its application to effective ISP parameter tuning. The proposed framework was experimentally validated on a state-of-the-art CMOS and ISP in Section *Experiment*. Future work and concluding remarks are also shown.

User preference based ISP tuning framework

The overall process of the proposed framework is illustrated in **figure 2**. The framework is divided into three steps: *Modeling for pre-filtering, User preference feedback* and *Model fine-tuning*. In the first step, we build a DNN model for filtering on the large-scale image generated by ISP simulator. This process dramatically reduces the number of entire images. Next, in *User preference feedback* part, we present an IQ preference survey method that considers perceptual metric. In the last step, we fine-tune the DNN model in the first step according to the surveyed result in second step.

In this paper, we are mainly focusing on methods of modeling the user's IQ preference and selecting good representative images for feedback. For efficient explanation, *User preference feedback* is described in the first session with a GUI of surveying. After that the first and last steps of framework are explained together in the *preference modeling* session.

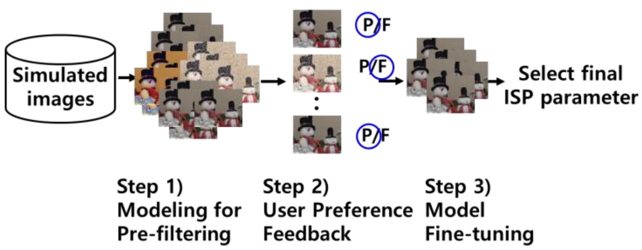

Figure 2. Flowchart of the ISP tuning framework based on user's preference feedback

User preference feedback

Since it is impossible to make a subjectivity model without any information about the user's preference, the framework will require to user about *N*-th pairwise image comparisons. In the subjective quality evaluation, *the pair-wise comparison* may show higher reliability than *the categorical rating method* [5]. *The pair-wise comparison* method simply selects the better one out of the two images, which is more intuitive than the other approach using the quantifying the difference in image quality. In particular, images in ISP tuning process have a very small difference, and it is difficult to quantify the difference. So the proposed framework uses *the pair-wise comparison method* for surveying user preference as in **figure 3**.

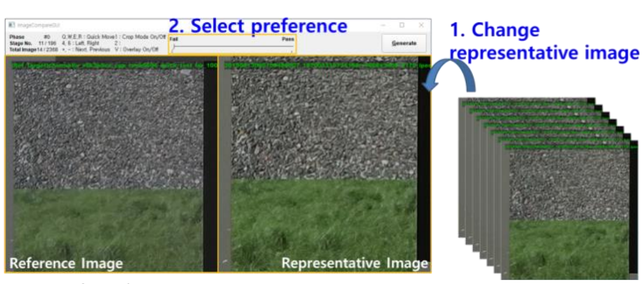

Figure 3.GUI of the pair-wise comparison based subjective image quality assessment. A left image in main GUI is reference and a right image is representative test image.

The GUI displays a fixed reference image with high IQ in left panel and a multiple set of representative images selected from filtered test images. In survey, user is always forced to choose whether the representative image has preference IQ or not. In this process, the user's feedback is specified as the image label for the preference (e.g. preference or non-preferred). The method is straightforward and thus expected to be accurate for non-expert on IQ verification or ISP tuning.

Selection of representative images

In whole process of the proposed framework, we use a large-scale simulation images by applying randomly selected ISP parameters of about 20 in the noise reduction and sharpening blocks. In the feedback process, due to the large number of images, we cannot feedback on all the images. Therefore, selecting a good set of representative images from the whole images is critical. The framework needs to identify a subset of different images that represent the original set.

In order to select reasonable representative image set, we clustering the large-scale images with considering both of objective and subjective IQ metric as in **Table 1**. The objective IQ metric is calculated statistically from one or more ROI in printed chart under controlled environment (illuminance, color temperature and distance). Therefore, various IQs can be expressed according to the ROI type (e.g. Noise, Texture, Resolution, Color and etc.). Subjective metric, on the other hand, can be defined as one score or distance by comprehensively determining the IQ level. Therefore, we make cluster based on the objective metric, which is high-dimensional data, and then performed post processing using the subjective metric.

First, the framework extracts objective image feature vector $\boldsymbol{x}$ of k-dimension that concatenate of pre-defined IQ metrics such as *visual noise min/max, acutance mean/max, texture, over/under sharpening* [6]. Using objective IQ metric, we applied the k-means algorithm to group the whole images to about a hundred clusters [7].

Table 1. Process of selection set of representative images

k-means with perceptual threshold

INPUT: $image\ I \in \mathbb{R}^n,\ IQM\ x \in \mathbb{R}^{n\times k}$
$x_{outlier} \leftarrow \emptyset$
$C \leftarrow kmeans(x)$
for j =1,…, T, i=1, …, N_j do
    $d_i \leftarrow dist(I_c, I_i)$
    if $d_i > \tau$
        $x_{outlier} \leftarrow x_{outlier} \cup x_i$
end for
$C \leftarrow C \cup kmeans(x_{outlier})$

Second, because k-means is vulnerable to outlier problems, we handled outliers using subjective IQ metric based on similarity distance of CNN features of Alexnet [8]. Compared with the traditional similarity metric such as SSIM, the CNN based metric shows a higher correlation with human perception [9]. We calculate the similarity distance between the center image and the rest of the cluster image, and determine outlier image if the distance is above the threshold. The outlier images, next, perform k-means clustering once more and concatenates with existing cluster. Finally, the center images of each cluster are used as the set of representative images to survey user feedback.

Preference modeling

The purpose of preference model is to determine the user's preference for a given image. In feedback part, user preferences were labeled as two categories: *preference* or *non-preferred*. Therefore, we used a fully connected DNN based binary classifier that consisting of 3 hidden layers as in **figure 4**. All hidden weight layers use the rectification (ReLU) activation function and the dropout technique is adapted in training process [10]. The output of the last fully connected layer is fed to the softmax of binary class, which produce a score for the preference as follows:

$$s_{\boldsymbol{w}}(\boldsymbol{x}) = \frac{\exp\left(z_{preference}(\boldsymbol{x},\boldsymbol{w})\right)}{\sum_L \exp\left(z_L(\boldsymbol{x},\boldsymbol{w})\right)} \quad (1)$$

where, $\boldsymbol{x}$ is input of IQM, $\boldsymbol{w}$ is weight parameters, $z(\cdot)$ is output of last hidden layer and L is output label, $\mathrm{L} \in \{\mathrm{preference}, \mathrm{non-preferred}\}$. The preference model accepts IQM $\boldsymbol{x}$ as input and returns preference c as output.

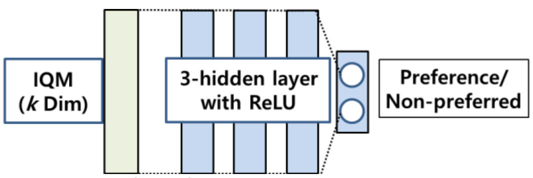

Figure 4.The DNN for preference modeling.

The preference model is utilized twice in the framework: pre-filtering and fine-tuning step. The first step roughly filters acceptable quality images from the entire image data. In the later fine-tuning process, based on those users' comparison result, we fine-tune the DNN model and measure a similarity score of users' preference images. Finally, the framework provides top *k* candidates according to user preferences and applies the selected final image' parameters to the ISP system.

Experiment

The framework was evaluated in a state-of-art CMOS and ISP system. In aspect of ISP optimization, we are focusing on noise reduction and sharpness blocks which have IQ trade-off relationship. Our experiment performed on day-light condition using studio chart that has various sub-chart such as *dead leaves*, *3 women faces* and etc. [6].

We evaluate the proposed framework in two aspects: 1) Fine-tuning performance and 2) ISP tuning quality. To evaluate the fine-tuning performance, we measure the convergence of IQ metrics at each progress as **table 2**. The *simulation* step means original dataset generated from ISP simulator. The *pre-filter* and *fine-tuned* are the filtering results using the DNN model before and after reflecting user preference. And the *user feedback* step is a mean and standard deviation of preferred images. As the process approaches to the final step (*fine-tuned* and *top-5* step), we confirm that the convergence of mean value to the scores of user feedback and the decrease of standard deviation. This means that the DNN model of binary classifiers can reflect the preferences well. Note that, we do not aim to find the highest IQ metric score, but to find images that are similar to preferred images based on user's IQ feedback.

Table 2. The mean and standard deviation of IQM

Step	# of image	Noise	Texture	Resolution
Simu-lation	10K	85.97(8.63)	67.00(6.09)	91.05(2.39)
Pre-Filtered	2.8K	86.10(5.43)	68.92(3.77)	91.40(1.48)
Fine-tuned	630	90.39(2.56)	69.56(3.10)	91.70(0.68)
Top-5	-	90.63(1.93)	69.60(1.60)	91.77(0.41)
User feedback	10	90.42(2.24)	69.08(3.07)	91.68(0.70)

Next, we compare of IQ between a base-tuned and the proposed method as shown in **figure 5**. The base-tuned result shows less noise than the proposed method in women face and it is hard to notice. In contrast, the texture of the proposed method is clearly better in *dead leaves* and *grass* region. The differences in IQM are described in **table 3**. Noise score decreased but it is hard to notice in real image. On the other hand, as texture and resolution are visually improved, we can see that the score has increased. These results show that the objective IQ metric cannot fully reflect the perceptual IQ of people, and it can be seen that ISP tuning can produce better images when the IQ tuning is performed through user feedback.

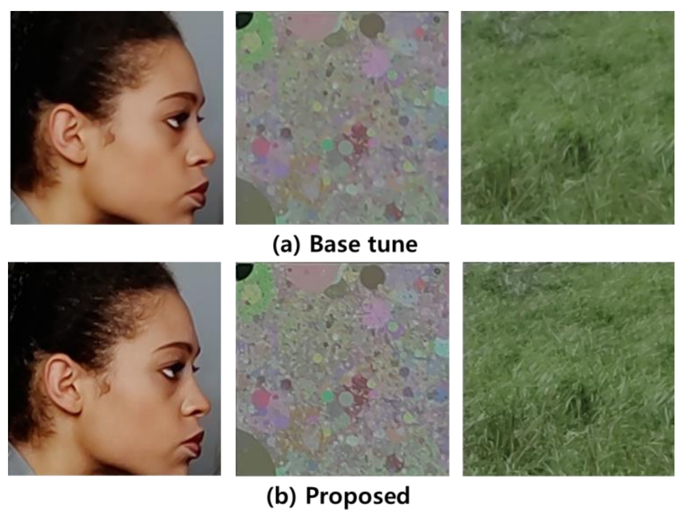

Figure 5. Comparison of tuning images: (a) Base-tune and (b) the proposed method.

Table 3. The final IQM score

	Noise	Texture	Resolution
Base tune	**95.52**	66.80	87.16
Proposed	91.85	**70.04**	**92.61**

Conclusion

This framework focused on how to reflect the subjective of IQ for finding optimum ISP tuning process. The framework has two key features: (1) image filtering based on personal preference and (2) selecting set of representative image with perceptual metric. Based on experimental evidence, the framework can find user optimized ISP parameters with straight forward feedbacks. In the future, the reliability of the proposed may be extended into the following approaches: (1) Verification for advanced ISP blocks such as multi-frame based ISP and (2) Expansion of perceptual metric to feature extraction part.

References

[1] Nishimura, Jun, et al. "Automatic ISP Image Quality Tuning Using Nonlinear Optimization." 2018 25th IEEE International Conference on Image Processing (ICIP). IEEE, 2018.

[2] Tseng, Ethan, et al. "Hyperparameter optimization in black-box image processing using differentiable proxies." ACM Transactions on Graphics (TOG) 38.4 (2019): 27.

[3] Kang, Sing Bing, Ashish Kapoor, and Dani Lischinski. "Personalization of image enhancement." CVPR 2010. IEEE, 2010.

[4] Caicedo, Juan C., Ashish Kapoor, and Sing Bing Kang. "Collaborative personalization of image enhancement." CVPR 2011. IEEE, 2011

[5] Mantiuk, Rafał K., Anna Tomaszewska, and Radosław Mantiuk. "Comparison of four subjective methods for image quality assessment." Computer graphics forum. Vol. 31. No. 8. Oxford, UK: Blackwell Publishing Ltd, 2012.

[6] Baxter, Donald, et al. "Development of the I3A CPIQ spatial metrics." Image Quality and System Performance IX. Vol. 8293. International Society for Optics and Photonics, 2012.

[7] Wagstaff, Kiri, et al. "Constrained k-means clustering with background knowledge." ICML. Vol. 1. 2001.

[8] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems. 2012.

[9] Zhang, Richard, et al. "The unreasonable effectiveness of deep features as a perceptual metric." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018.

[10] Srivastava, Nitish, et al. "Dropout: a simple way to prevent neural networks from overfitting." The journal of machine learning research 15.1 (2014): 1929-1958.

Author Biography

Cheoljong Yang *received the B.S degree in electrical engineering (2010) and his PhD in vision information processing (2017) from Korea University, Seoul, Korea. Since then he has worked in the system LSI division at Samsung Electronics. His current research interests are in machine learning based image quality assessment for CMOS.*

Jinhyun Kim *is currently working as an engineer at Samsung Electronics. He received his B.S. in computer science (2012) from Hanyang University in Korea. His research interests include image quality metric, machine learning.*

Jungmin Lee *is currently working as an engineer at Samsung Electronics. He received his B.S. in electronic engineering (2010) and M.S. in image processing and computer vision (2012) from Sogang University in Korea. His research interests include image processing, computer vision and objective image quality assessment.*

Younghoon Kim *is the engineer of Samsung Electronics Co. Ltd. He received B.S. (2011) in Computer Science and Engineering from Konkuk University, Korea. His research areas are Machine Learning, Big data. Recently, his current research interests include Deep Learning based Image Quality Metrics and calibration.*

Sung-Su Kim *received his B.S. in electronic engineering and his M.S. in human vision system from KyungPook Natiional University. Since 2004, he has worked in Samsung Advanced Institute of Technology (SAIT) and Samsung Electronics Co. Ltd., Korea. His research interests include pattern recognition, image understanding, image quality metric, and machine learning.*

Taehyung Kim *received his B.S degree in Control & Instrumentation Engineering from Seoul City University (1996). Since then he has worked in Samsung Electrics Co. Ltd, Hwasung, Korea. His work has focused on CMOS.*

JoonSeo Yim *received his B.S. and Ph.D. degree from Seoul National University (1991) and KAIST (1998) respectively, majored in Electrical and Electronics Engineering. He has been worked in Samsung Electronics. His research interests include camera sensor innovation, evolutionary computation and design optimization methodologies.*