A New Framework for Efficient Low-light Image Enhancement using Approximated Gaussian Process

Yuen Peng Loh
Faculty of Computing and Informatics, Multimedia University, Cyberjaya, Selangor, Malaysia

ABSTRACT
Gaussian Process (GP) is a robust distribution modeling technique that is very promising for computer vision systems. In particular, its multivariate distribution modeling is especially effective for low-light image enhancement where localized enhancement is required to address the over- and under-enhancement problem, and also retrieval of features which has been lost due to low illumination. However, GP lacks practicality due to its computation complexity that increases cubically following data increment. This paper proposes a sparse GP regression based solution whereby clustering is exploited to reduce the training cost of a GP model. Instead of utilizing all values from an image, clustering groups similar training pixels or image patches pairs into clusters and the cluster centers are used to train an approximate GP enhancement model. Experiments conducted showed the proposed framework can achieve training time reduction of as much as 75% from the baseline. In line with this, the proposed approach also improved the enhancement performance in both PSNR and features retrieval metric, and is competitive with the current state-of-the-art.

Keywords : *Gaussian Process, image enhancement, low-light, sparse approximation.*

I. INTRODUCTION

Gaussian process ($\mathcal{GP}$) [1] is a sophisticated distribution modeling technique for regression problems. Unlike common model based regression technique that models the relationship of data and variables using a distribution function, the $\mathcal{GP}$ governs them as a distribution of functions instead. This characteristic is promising for image enhancement problems such as super-resolution [2] and low-light image contrast enhancement [3] as each local region and pixel of an image calls for specific functions for optimal enhancement.

However, the fundamental drawback that restricts the $\mathcal{GP}$ to be practically implemented into a deployable system is its computation complexity. For a typical $\mathcal{GP}$, the computational cost is $O(n^3)$ for n training data, whereas image data are of varying resolution which can range from 200×200 to 2000×2000 and beyond. As the resolution increases, the cost of enhancement by $\mathcal{GP}$ increases cubically as well, thus render the processing time impractical.

In this work, a simplistic approach is introduced to accelerate the training of a Gaussian Process by decreasing

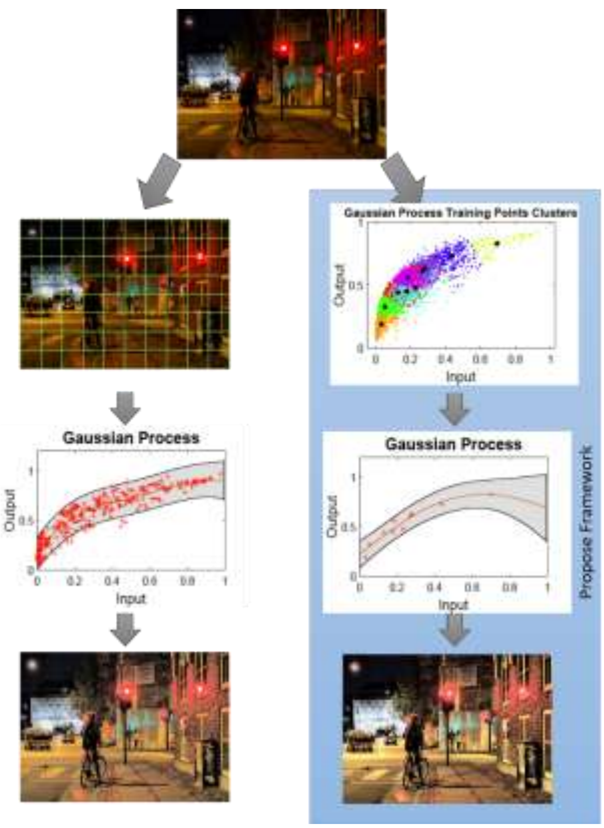

Figure 1: Illustration comparison of proposed framework with prototype by [3]

the training data required to build a $\mathcal{GP}$ model. Specifically, this paper proposes a framework to speed up the runtime training of local $\mathcal{GP}$ for the practical application of low-light image enhancement. This framework exploits image region similarity to form clusters and utilizes the cluster centers as representative training data to approximate a $\mathcal{GP}$ enhancement model. The resultant $\mathcal{GP}$ closely approximates a $\mathcal{GP}$ trained using all the data, but obtained using a significantly less time due to the reduction in data points, i.e. number of pixels, n to a predetermined number of data points, i.e. number of clusters, k where $k \ll n$.

II. RELATED WORKS

Various sophisticated approaches has been proposed to reduce the computation complexity of $\mathcal{GP}$ [4]–[6] mainly by mathematical approximations to scale down the cost $O(n^3)$ [4] is an early study using a two-step framework to provide a unifying view on sparse approximations for $\mathcal{GP}$ regression. Covering approximation paradigms such as likelihood approximations, projection methods, matrix approximations etc., the study provides a theoretical ranking of the approaches based on how well they approximate a full $\mathcal{GP}$. However, the study has noted that a theoretical ranking such as the given, may not be transferable to practical situations.

[5] is one such method that that reduces the complexity to $O(m^2n)$ where m represent pseudo data points and $m \ll n$. The points are learned by gradient descent optimization to parameterize the covariance of the $\mathcal{GP}$ model. The highly sparse solution is able to achieve comparable performance to a full $\mathcal{GP}$ when applied to larger datasets at the time, e.g. 2000 training data points reduced to 25 pseudo data points.

On the other hand, [6] approach the problem from an algorithmic point of view instead of data approximation as done [5]. They proposed to hierarchically factor the covariance matrix that enables the inversion operation to be performed at $O(n \log^2 n)$ and the matrix determinant evaluation at $O(n \log n)$ where the combined costs is a reduction from the original cost of $O(n^3)$.

The proposed framework is a form of the sparse $\mathcal{GP}$ regression model [5] which has the complexity cost $O(m^2n)$. However, the core difference is that instead of using a pseudo dataset to parameterize the $\mathcal{GP}$, data points are clustered based on their similarity and the cluster centers, k are used as representative data points to approximate the distribution. Thus, the training complexity cost remains at $O(k^3)$ but with $k \ll n$.

III. TRAINING DATA SELECTION FOR EFFICIENT ENHANCEMENT

A. $\mathcal{GP}$ in Low-light Image Enhancement

Proposed by Loh and Chan [3], low-light contrast enhancement using $\mathcal{GP}$ shows promising results for the non-traditional purpose of features retrieval instead of image quality. Due to the non-uniformity of illumination in low-light environments, enhancement of low-light pixels require highly localized functions. This is to counteract different luminance levels and provide optimal enhancement so as to avoid the over- and under-enhancement problem that is commonly faced by low-light enhancement algorithms.

Unlike more popular regression methods that models data distribution based on a function, $\mathcal{GP}$ models the data as a distribution of functions, which can be intuitively understood as modeling each data point with its own function. Hence, successfully addressing the over- and under-enhancement issue. Additionally, this prototype model also encompass a Convolutional Neural Network (CNN) that is trained to capture object features and maps them into pixel space. The mapped pixel values serves as an intermediate reference that instill feature enhancement functions into the construction of the $\mathcal{GP}$, consequently achieving features retrieval.

Specifically, given the input data $x = \{x_1, x_2, \dots, x_n\}$, which in this domain, are the pixel values of the low-light image to be enhanced, the $\mathcal{GP}$ defines a distribution of functions $f(x) \sim \mathcal{GP}(\mu, K)$ where μ is the mean of the distribution that is set as 0 for simplified implementation, and K is the covariance function that defines the relationship between the training and testing inputs. The output, which are the values of enhanced pixels, y can therefore be estimated from the marginal distribution of functions $P\big(f(x_1), f(x_2), \dots, f(x_n)\big)$, i.e. distribution of localized enhancement functions.

As noted that the complexity of training a $\mathcal{GP}$ is in the third order and the model is trained at runtime for every low-light image to achieve localized enhancement. Hence, the computational time is not feasible for practical implementation on a task such as image enhancement that could contain tens of thousands of pixels as data points for each image. Thus, the prototype model of [3] implemented a simple patch similarity approach to select unique training values based on average intensity of each training patch pairs. While this method managed to reduce the computation time, there are still notable drawbacks.

First, the reduction is arbitrary whereby if the local variation of the low-light image illumination intensity are high, significant reduction is not achievable due to large unique pair samples. Thus, it only enables the prototype to be viable on single channel images, such as the luminance channel, Y of from the $YCbCr$ color space. Subsequently, the second drawback is the poor color quality of the resultant enhanced image. This is contributed by the single channel processing that heavily relies on the channels decomposition provided by color space conversion. Therefore, this paper seek to reduce the training computation while minimizing the compromise on quality.

B. Training Data Reduction by k-means clustering

To this end, data clustering is engaged as the approach for training data reduction by exploiting image patch similarity, i.e. using single image patch to represent similar patches to construct the enhancement $\mathcal{GP}$. Explicitly, the computationally efficient centroid-based technique, k-means [7] with k-means++ initialization [8] is used to obtain centers of image patch clusters as representative data points for training.

Consider a low-light image, I_L and its corresponding reference image generated by the CNN model, I_R, each containing n pixels. They are both partitioned into patches of preset size $p \times q$ pixels to produce N patches each where $N < n$. Each of the image patches from I_L and its spatially corresponding patch from I_R, are flattened into row vectors and concatenated to form data points, $X^d = \{X_1, X_2, \dots, X_N\}^d$ with dimension, $d = p \times q \times 2$ for each point.

Following the generation of X^d that represents image patch pairs, the number of clusters to be formed, k is first selected, which is equivalent to the amount of training data to approximate the $\mathcal{GP}$. The k-means++ algorithm will first greedily pick the initial k cluster centers, $c = \{c_1, c_2, \dots, c_k | c \in X\}$ that are maximally different.

Table I: Results of model trained using k training points applied only on the illumination channel, Y.

Training points, k	Quality PSNR	Feature retrieval (SIFT) Precision	Recall	F_1	Comp. Time (s)
Dark	10.44	0.4514	0.1711	0.2090	-
GPE [3]	16.25	0.4745	**0.6563**	0.5292	1.11
200	**16.36**	**0.5180**	0.5943	**0.5308**	0.91
100	16.25	0.5036	0.5928	0.5222	0.46
50	16.20	0.4872	0.5929	0.5119	0.37
40	16.18	0.4820	0.5926	0.5092	0.36
30	16.09	0.4745	0.5914	0.5040	0.32
20	15.99	0.4600	0.5862	0.4928	0.29
10	15.63	0.4229	0.5754	0.4644	**0.27**

Using the initialized c, the k-means algorithm then distributes the remaining X data points into clusters C_i corresponding to the center point c_i, where $i \in \{1,2,\dots,k\}$ based on the nearest distance determined by the measure $\phi(c_i, X_j)$ where $j \in \{1,2,\dots,N\}$. In this case, the distance measure will reflect the similarity of a pair of training patch with another pair.

In the next step, the mean of each cluster C_i is calculated using all the data points within the cluster, which are then assigned as new cluster centers. Subsequently, the data points X are redistributed following the new clusters centers, forming new clusters. These steps are then repeated to improve the intra-cluster variation iteratively. The iterations continue until a stable set of c are obtained. Therefore, the converged cluster centers c, illustrated as black points in the scatter plot in Fig. 1, are used as the training data points for the $\mathcal{GP}$. The resultant distribution can be viewed as a approximate of the enhancement $\mathcal{GP}$ model, obtained with significantly less training data, i.e. k training points where $k \ll n$, but still achieves similar enhancement results, as shown in Fig. 1.

IV. EXPERIMENTS

The proposed framework is tested following the schemes provided by [3], i.e. the same 300 testing images from the Microsoft COCO dataset [9] used in the paper are used for the quantitative evaluations in this section whereby each of the images are synthetically darkened to produce 26 different images using gamma correction, including the original bright image. First, is an ablation study on the effect brought about by varying amounts of clusters centers as well as when the model is applied on different color spaces. The metrics used for the evaluation include PSNR for quality, precision, recall, and F_1-score for SIFT features retrieval, as well as computational time to gauge the efficiency.

Then a comparative study is conducted against other low-light image enhancement algorithms [10],[11], both quantitatively using the same data in the ablation experiments, and qualitatively using the real low-light images from the Exclusively Dark (ExDark) image dataset [13] to look into the feasibility of such a framework in solving this real world problem. Failure cases are discussed in the end of the section as well.

A. Implementation Details

In the implemented framework, Euclidean norm is used as the distance measure for the k-means clustering with varying

Table II: Results of model applied on $\boldsymbol{RGB}$ images with varying $\boldsymbol{k}$ value.

Training points, k	Quality	Feature retrieval (SIFT)			Comp. Time (s)
	PSNR	Precision	Recall	F_1	
Dark	10.44	0.4514	0.1711	0.2090	-
200	**16.70**	**0.5717**	0.5385	**0.5245**	1.74
100	16.67	0.5600	0.5476	0.5244	0.75
50	16.66	0.5478	0.5549	0.5231	0.56
40	16.66	0.5416	0.5579	0.5216	0.54
30	16.61	0.5315	**0.5602**	0.5177	0.49
20	16.46	0.5105	0.5590	0.5063	0.45
10	16.27	0.4713	0.5582	0.4846	**0.41**

values of k, where $k \ll n$. The squared exponential function is selected as the covariance function for the $\mathcal{GP}$, and two CNN models were implemented, one for generating single channel luminance, Y reference image, whereas another for generating three channel RGB reference image.

For the computational time evaluation, all the experiments were done using Intel i7 processor @ 3.60 GHz without any GPU acceleration. Real images from the ExDark dataset were used and resized to 256×256 pixels for fair comparison.

B. $\boldsymbol{k}$ Values Evaluation

Two experiments were conducted to investigate the feasibility of the proposed framework. The first is to incorporate the clustering into the existing model provided by [3], replacing their unique pair selection algorithm. In this original model, the CNN and $\mathcal{GP}$ model are only applied on the illumination component, Y of the $YCbCr$ color space image due to computation restrictions. By inserting the k-means into the model, the training data required for the regression is significantly reduced. This model is referred to as the Luminance Model (LM). The evaluation results are shown in Table I including the performance of the prototype obtained from the paper except the computational which is implemented with the aforementioned hardware. Dark refers to the evaluation applied on the synthetically darkened data that is not enhanced. The training points by [3] is not stated as the approach uses arbitrary amount of points based on the unique pairs found.

It can be observed in Table I that the training point reduction effectively shortens the processing time. By using as low as $k = 10$ data points for training, the computation time can be reduced by as much as 75.8% with only a compromise of 3.8% in PSNR, 10.9% in Precision, 12.3% in Recall, and 12.2% in F_1-score for SIFT features retrieval. Interestingly, by increasing the k value, the performance not only improves the performance at the trade-off of increasing computation time, the models manage to outperform the baseline set by [3]. Notably, using $40 \leq k \leq 200$ training data points, the improvement of the computation time and Precision are at the range of (67.7%,17.5%) and (1.5%,9.2%) respectively, with relatively stable Recall that is an average of 9.6% lower than the baseline. This counter intuitive improvement can be attributed to the patch based similarity clustering and cluster centers generation used in the k-means algorithm. The baseline algorithm represents patches by a single average value to select unique training values which

Table III: Results comparison of proposed framework with state-of-the-art low-light image enhancement methods.

Training points, k	Quality	Feature retrieval (SIFT)			Comp. Time (s)
	PSNR	Precision	Recall	F_1	
Dark	10.44	0.4514	0.1711	0.2090	-
FBE [13]	14.68	**0.5959**	0.4659	0.4831	0.12
WVM [14]	12.88	0.5794	0.3458	0.3722	3.29
LIME [15]	15.08	0.3205	0.6463	0.4076	**0.04**

SRRM [16]	14.10	0.5779	0.3395	0.3714	5.85
GBLE [17]	13.25	0.2284	0.3303	0.2513	0.35
GPE [3]	16.25	0.4745	**0.6563**	0.5292	1.11
LM(k=200)	16.36	0.5180	0.5943	**0.5308**	0.91
CM(k=30)	**16.61**	0.5315	0.5602	0.5177	0.49

results in a significant loss of information. Hence, it is more crucial to have good quality representative data to achieve reliable approximation as compared to increasing the training data quantity.

The second experiment is conducted by applying the enhancement model on RGB images, using all three channels. In order to implement this configuration, the CNN model that provides the feature reference image needs to be modified to work on three channel images instead of one. Therefore, the

CNN architecture is altered to take in three channel RGB images as input and output the same color space image. The model is then trained from scratch using images from synthesized from the Microsoft COCO training set following the scheme detailed in [3], i.e. total images used for training is 2,152,358. This model will be referred to as the Color Model (CM) Table II shows the comparison results of using varying k values. Results using the prototype implementation without data reduction is not included due to the unreasonable time required to process a single image.

Based on Table II, it is noted that the computation time required by CMs are higher than LMs, where for $k = 10$, the reduction compared to the baseline is at 63.0%. Nonetheless, this is an expected increase due to the increase in color channels processed especially the reference image generation by CNN. The more noteworthy observation is the PSNR results where all values of k outperform the baseline. Furthermore, the Precision for SIFT feature retrieval exceeds the baseline and LMs by using only $k = 30$, i.e. 12.0% and 2.6% improvement from the baseline and LM($k = 200$) respectively. These significant improvements can be attributed to the usage of all channels representing and image as compared to only a single channel enhancement model by LM. However, it is noted that the CMs record greater deterioration in the Recall, and there does not seem to be an observable correlation between the increase and decrease of k with the Recall performance.

C. Comparison with State-of-the-Arts

In this section, look into the quantitative results of a few state-of-the-art algorithms in low-light image enhancement as compared to the selected models of the proposed framework. As the experiments are conducted using the exact same data and settings detailed by [3], the state-of-the-art results shown in Table III are imported from the paper, except for the computational time which is obtained by re-implementing all the methods using the same hardware.

The compared methods are fusion-based enhancement (FBE) [13], weighted variational model (WVM) [14], low-light image enhancement via illumination map estimation (LIME) [15], structure-revealing low-light image enhancement (SRRM) [16], gradient-based low-light enhancement (GBLE) [17], and the baseline Gaussian Process enhancement (GPE) [3]. Two models using the proposed method, LM($k = 200$) and CM($k = 30$) are included in Table III for the comparison.

As seen in Table III, the performance of the proposed models are comparable on all evaluations, especially the PSNR metric. Nevertheless, the trade-off between performance and speed is evident in the proposed models.

D. Qualitative Comparison

The qualitative examples corresponding to the methods and results discussed in Section IV C are shown in Fig. 3. It can be visually observed that each method has its strengths and drawbacks. FBE is poor at handling noise while WVM suffers from under-enhancement. LIME shows the best color however it suffers from over-enhancement when the illumination difference within an image is too large, such as the presence of a light source. SRRM shows a satisfactory enhancement, though on closer inspection, the saturation is relatively high, e.g. faces of people has a slight orange hue or the greenish tint in the enhanced image. GBLE performs the poorest as its results look artificial and GPE shows an obvious lack of color. The proposed model shows a balanced enhancement where there are not apparent presence of the aforementioned issues and it shows clear improvement in terms of color when compared to the baseline, GPE. However, in selected cases, its enhancement is not as bright as LIME, as seen in Fig. 3.

E. Failure cases

While the proposed framework performs relatively well, there are some noticeable failures cases, which are shown in Fig. 2. An interesting case found in the results of the proposed framework when applied to real low-light RGB images is that the enhancement skews the color spectrum of a small subset of images, as shown in Fig. 2 (a). This is likely due to the design of using one $\mathcal{GP}$ distribution to perform regression for three color channels. Given a low-light pixel value from a channel, the enhanced value is obtained from the same distribution regardless of the channel it is from. It is hypothesized that this can be rectified by introducing additional constrain when training the $\mathcal{GP}$ model and performing the prediction.

The other failure case is encountered when the low-light image is severely deteriorated by image noise, as seen in Fig. 2 (b). Capturing images in low-light conditions are prone to noise signals such as poison noise. This is because in an environment of insufficient light particles, photographers will prolong the exposure in order to capture enough light and details. At the same time, noise signals are capture as well, hence producing such problems in the images. The current model does not explicitly handle the issue of noise, however, as shown in Fig. 3, it shows robustness towards noise as compared to some of the state-of-the-art methods.

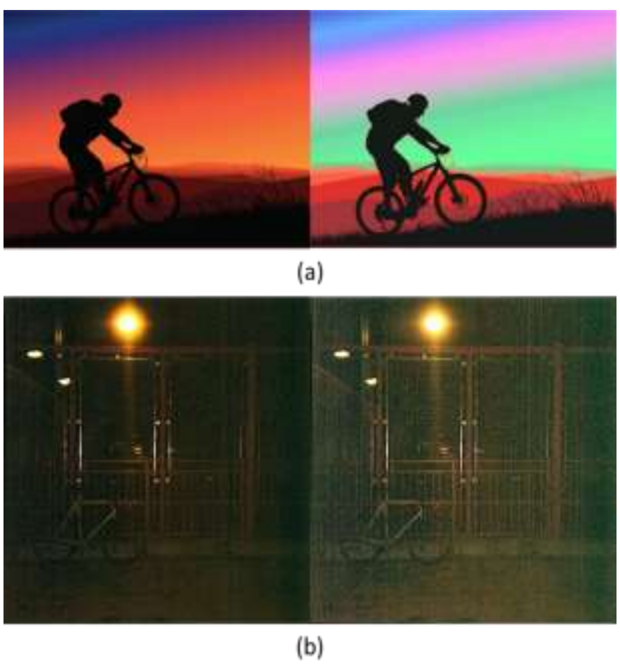

Figure 2: Examples of failure cases. (a) Skewed color spectrum; (b) Severe noise. Left: Original low-light image, Right: Image enhance by proposed CM ($\boldsymbol{k = 30}$).

V. CONCLUSION

A simple framework to reduce the computational cost of $\mathcal{GP}$ training is proposed for the application of low-light image enhancement. Specifically, k-means clustering is used to group similar training data points, i.e. pixels or patches pairs of a low-light image its corresponding reference image. From the clustered data, the derived cluster centers are used to approximate the $\mathcal{GP}$ enhancement model, effectively reducing the amount of training data, from the total of n pixels or patches, to k cluster centers, where $k \ll n$.

Experiments on various values of k shows that an approximate $\mathcal{GP}$ model can be trained in less than 25% of the time required by the baseline and with a compromise of less than 15% on the PSNR and features retrieval evaluation. Moreover, this major reduction in the computational has allow the model to be applicable to a full color space image instead of the luminance channel only. As a result, the performance of the model is further improved, in particular the PSNR and Precision of features retrieval, and it is competitive with the current state-of-the-art low-light image enhancement methods.

An analysis on the failure cases brought up two issues of the proposal. The first is the skewing of the color spectrum in some enhancements, and the model is unable to satisfactorily enhance images with severe noise. The future work of this research aims to address these issues in efforts to produce a more robust modeling of low-light image enhancement.

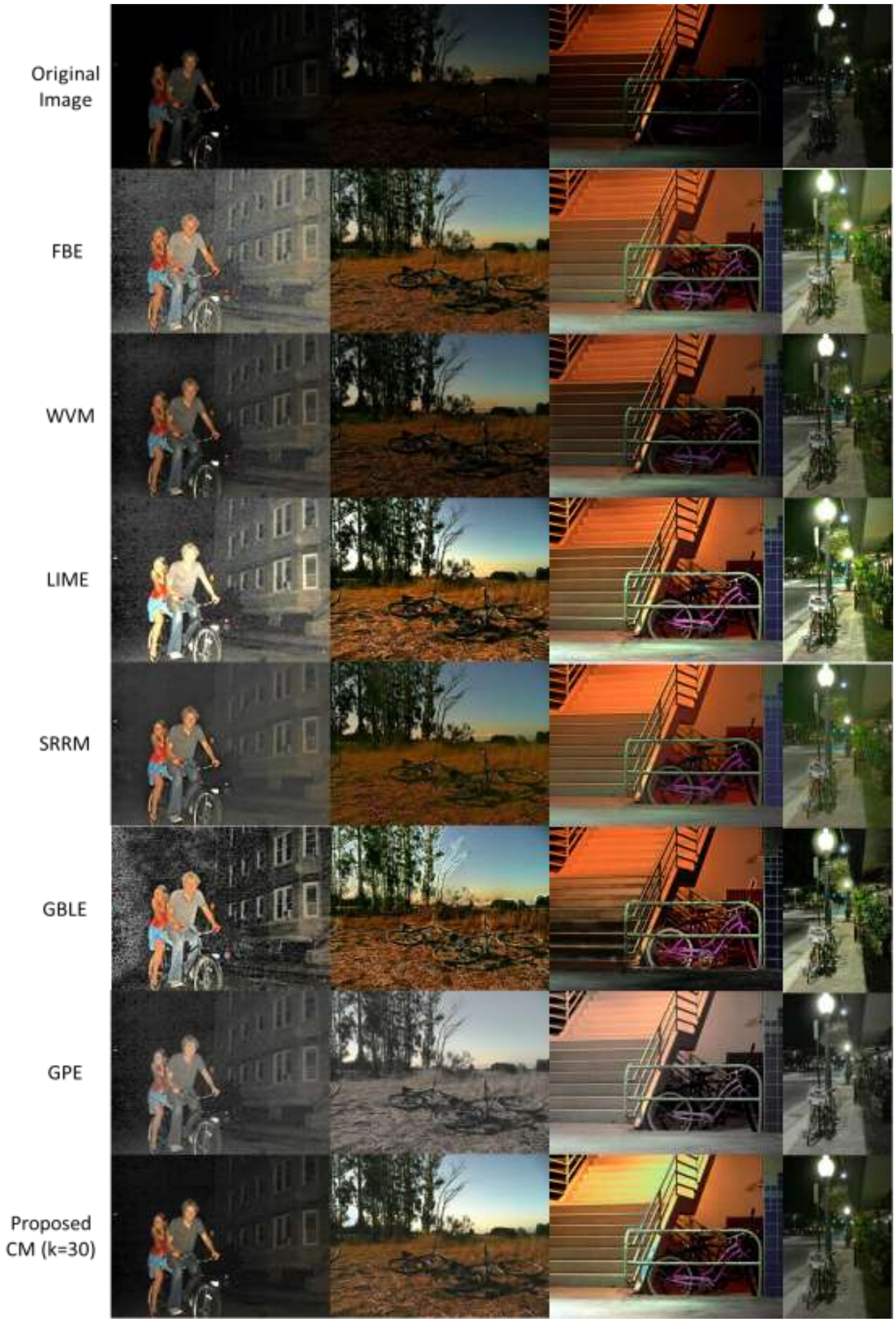

Figure 3: Comparison of low-light image enhancement methods applied to real low-light images from the ExDark dataset.

REFERENCES

[1] Williams, C. K., & Rasmussen, C. E. (2006). "*Gaussian processes for machine learning*" (Vol. 2, No. 3, p. 4). Cambridge, MA: MIT press.

[2] He, H., & Siu, W. C. (2011, June). "*Single image super-resolution using Gaussian process regression*". In 2011 IEEE Conference on Computer Vision and Pattern Recognition (pp. 449-456).

[3] Loh, Y. P., Liang, X., & Chan, C. S. (2019). "*Low-light image enhancement using Gaussian Process for features retrieval*". Signal Processing: Image Communication, 74, 175-190.

[4] Quiñonero-Candela, J., & Rasmussen, C. E. (2005). "*A unifying view of sparse approximate Gaussian process regression*". Journal of Machine Learning Research, 6(Dec), 1939-1959.

[5] Snelson, E., & Ghahramani, Z. (2006). "*Sparse Gaussian processes using pseudo-inputs*". In Advances in neural information processing systems (pp. 1257-1264).

[6] Ambikasaran, S., Foreman-Mackey, D., Greengard, L., Hogg, D. W., & O'Neil, M. (2015). ***"Fast direct methods for Gaussian processes".*** IEEE transactions on pattern analysis and machine intelligence, 38(2), 252-265.

[7] Lloyd, S. (1982). "*Least squares quantization in PCM. IEEE transactions on information theory*", 28(2), 129-137.

[8] Arthur, D., & Vassilvitskii, S. (2007, January). *"k-means++: The advantages of careful seeding. In Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms" (*pp. 1027-1035). Society for Industrial and Applied Mathematics.

[9] Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P & Zitnick, C. L. (2014, September). "*Microsoft coco: Common objects in context. In European conference on computer vision*" (pp. 740-755). Springer, Cham.

[10] Li, L., Wang, R., Wang, W., & Gao, W. (2015, September). "*A low-light image enhancement method for both denoising and contrast enlarging*". In 2015 IEEE International Conference on Image Processing (pp. 3730-3734). IEEE.

[11] Fu, X., Zeng, D., Huang, Y., Zhang, X. P., & Ding, X. (2016). "*A weighted variational model for simultaneous reflectance and illumination estimation*". In 2016 IEEE Conference on Computer Vision and Pattern Recognition (pp. 2782-2790).

[12] Loh, Y. P., & Chan, C. S. (2019). "*Getting to know low-light images with the exclusively dark dataset. Computer Vision and Image Understanding*", 178, 30-42.

[13] Fu, X., Zeng, D., Huang, Y., Liao, Y., Ding, X., & Paisley, J. (2016). "*A fusion-based enhancing method for weakly illuminated images. Signal Processing*", 129, 82-96.

[14] Fu, X., Zeng, D., Huang, Y., Zhang, X. P., & Ding, X. (2016). "*A weighted variational model for simultaneous reflectance and illumination estimation*". In 2016 IEEE Conference on Computer Vision and Pattern Recognition (pp. 2782-2790).

[15] Guo, X., Li, Y., & Ling, H. (2016). "*LIME: Low-light image enhancement via illumination map estimation*". IEEE Transactions on image processing, 26(2), 982-993.

[16] Li, M., Liu, J., Yang, W., Sun, X., & Guo, Z. (2018). "*Structure-revealing low-light image enhancement via robust Retinex model*". IEEE Transactions on Image Processing, 27(6), 2828-2841.

[17] Tanaka, M., Shibata, T., & Okutomi, M. (2019, January). "*Gradient-based low-light image enhancement*". In 2019 IEEE International Conference on Consumer Electronics (pp. 1-2). IEEE.