# LEARNING-BASED TONE MAPPING OPERATOR FOR IMAGE MATCHING

*Aakanksha Rana*

LTCI, Télécom ParisTech
Université Paris Saclay

*Giuseppe Valenzise, Frederic Dufaux*

Laboratoire des Signaux et Systmes (L2S, UMR 8506)
CNRS - CentraleSupelec - Université Paris-Sud

**ABSTRACT**

In this paper, we propose a new framework to optimally tone-map a high dynamic range (HDR) content for image matching under drastic illumination variations. This task is of fundamental importance for many computer vision applications. To design such a framework, we build a luminance invariant *guidance model* using a Support Vector Regressor (SVR) and learn it to facilitate the extraction of invariant descriptors from scenes subject to wide variety of appearance changes such as day/night transition. To this end, we initially generate appropriate training samples using a simple similarity-maximization mechanism. We then employ the learned model to predict optimal modulation maps that help to locally alter the intrinsic characteristics (such as shape, size) of the tone mapping function. We evaluate the proposed model performance in terms of matching score and mean average precision rate using state-of-the-art descriptor extraction schemes. We demonstrate that our tone mapping framework significantly outperforms the existing perceptually-driven state-of-the-art TMOs on the benchmark datasets.

***Index Terms*—** Descriptor, Image matching, High dynamic range, Tone mapping operator.

## 1. INTRODUCTION

High Dynamic Range (HDR) [1,2] imaging captures high contrast information from the very dark and bright regions of a scene. As a result, it has brought potential interest in solving illumination-related challenges in computer vision problems such as image matching [3, 4] where performance of the algorithms degrades substantially with drastic lighting variations.

Image matching algorithms [3] look for distinctive feature descriptors that are capable of describing the detected regions and remain invariant under different transformations such as geometrical or lighting variations. Traditionally, such algorithms have been designed and optimized for low dynamic range (LDR) imagery which is represented using gamma-corrected 8-bit integer representation and is approximately linear to human perception. On the other hand, HDR images consist of real-valued pixels which are proportional to the physical luminance of the scene and are expressed in $cd/m^2$. As a consequence, HDR linear values are inappropriate when used with LDR-optimized descriptor extraction designs. In such a scenario, a simple solution opted by recent studies [5–9] is to convert HDR into an LDR representation using a Tone Mapping Operator (TMO) [1].

Tracing roots from computer graphics, TMOs have been designed to map HDR content in a suitable 8-bit LDR representation for display purposes [10,11]. For instance, TMOs such as [12,13] rely on compressing the estimated luminance (using Gaussian, bilateral filter) in HDR images to obtain a visually pleasant tone-mapped output. Nevertheless, their design objectives are to preserve human-vision attributes such as brightness and perceptual contrast. Differently from visual perception, the image matching pipelines are designed for machines where features are extracted from pixel-level information. These algorithms extract unique signatures, such as *histogram of gradient* orientations from image locations which can be matched when the same scene is captured under different transformations. As a result, existing TMOs might be sub-optimal for different computer vision applications ranging from scene recognition [14] to image retrieval [15] where algorithms rely on such descriptors extraction algorithms.

Though some recent studies emphasize the necessity and explain the requisites for designing TMOs optimal for keypoint detection [5–7], the descriptor-optimal TMO designs have not been studied so far. Hence, the problem of designing such optimal TMO remains open.

In this paper, we address this problem and propose a novel framework for designing a descriptor-optimal tone mapping operator (DoTMO). Our proposed method aims at facilitating the matching of descriptors that are extracted from high-contrast areas of the scenes under complex real-world illumination transitions, such as day/night change. To this end, we initially introduce a tone mapping function which can be locally modulated by spatially varying its parameters. We then propose to predict these modulation maps by means of a learned illumination-invariant guidance model which relies on gradient orientation-based features that are extracted from densely sampled patches from the HDR content.

Our idea is motivated by the conclusions of our previous work [8] where significant gains in Repeatability Rate [3] were observed when optimal TMO parameters (controlling TMO's shape and size) were learned pixel-wise. However, in that work, we mainly focused on designing a tone mapping model for corner-like keypoint detection task, while in this paper we consider a different problem, i.e., an optimal TMO for the extraction of discriminative descriptors.

Unlike corner detection, descriptor extraction depends on the large set of neighborhood pixel-set (or patch) which are processed altogether to formulate the discriminative unique signature. Hence, in this work, we propose to *learn* the TMO parameters locally but based on patch-level information from the scenes. Specifically, since each descriptor is restricted to a patch size such as $16 \times 16$ in SIFT and SURF, we learn the TMO parameters on patches of the same size.

The design of our guidance model is inspired by the regression-based "task-optimization" models [16]. In this paper, we formulate the problem of predicting optimal modulation maps as a regression problem and solve it by using a Support Vector Regressor (SVR) to cope with large variability in the input training samples.

Since there is no standard dataset to train or test any model in the context of our DoTMO, we additionally propose a simple descriptor

The work presented in this document was supported by BPIFrance and Région Ile de France, in the framework of the FUI 18 Plein Phare project

similarity-maximization approach to generate appropriate training samples. To this end, we define an objective function aiming to maximize the similarities of descriptors if they are extracted from images with lighting variations but from the same location. We carry out the optimization using stochastic gradient descent (SGD) [17] by deriving the required partial derivative architecture. We finally present the comparison of our approach with state-of-the-art TMOs using different descriptor extraction schemes. Our results show consistent gains in term of overall matching scores [18] and mean Average Precision (mAP) [3] rate across different illumination conditions with respect to popular tone mapping approaches proposed in the literature.

The paper is organized as follows. In Section 2, we provide the details of our proposed approach. We present the experimental results and analysis in Section 3. Finally, the conclusions are drawn in Section 4, along with future research directions.

## 2. PROPOSED TONE MAPPING MODEL

### 2.1. Model Overview

Fig. 1 outlines the framework of our proposed algorithm. It primarily consists of a tone mapping function $\varphi$ which maps the linear-valued HDR content of an image $I$ to an output LDR $I'$. More specifically, it is expressed as

$$I'(x) = \varphi(I(x), \boldsymbol{\theta}), \tag{1}$$

where $I \in \Re^{m\times n}$, $I'$ is of size $m\times n$ with pixel values in the $[0, 255]$ range, and $\boldsymbol{\theta}$ represents a vector of modulation maps, $\boldsymbol{\theta} = \{\theta_1, \theta_2\}$, where $\theta_k$ is of size $m \times n$. Secondly, the framework consists of a guidance model where an SVR predicts the optimal values of these modulation maps $\boldsymbol{\theta}$ by using the densely extracted local features from the HDR content. To this end, initially, the HDR image is densely sampled into patches of size $s \times s$ and from each such patch a SIFT feature $f$ is extracted. Then, these features are fed to the regressor which in turn predicts parameter values for modulation map $\theta_1, \theta_2$. Note that the regressor output for each feature is applied over the size $s\times s$ in these modulation maps, corresponding to exact location of the sampled patch from which the feature is extracted. Such patch level tuned vector parameters $\theta_1, \theta_2$ are later used by $\varphi$ to obtain the tone mapped image $I'$.

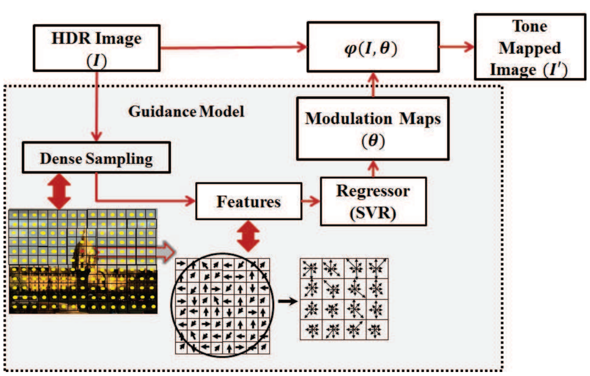


**Fig. 1**: *DoTMO*. The architecture of our proposed TMO.

### 2.2. Tone Mapping Function

Inspired by illumination normalization TMOs [12,13,19], our tone mapping function $\varphi$ in Eq. (1) is expressed as: $\varphi = I \cdot L^{-1}$, where the illumination component $L$ is estimated by a variant of bilateral filtering [20] and is given as:

$$L(x, \boldsymbol{\theta}) = \frac{1}{W} \cdot \sum_{y\in\Omega} \mathcal{G}_{\theta_1(x)}(\|x-y\|) \cdot \mathcal{G}_{\theta_2(x)}(\|I(x) - I(y)\|)I(y), \tag{2}$$

where $\mathcal{G}$ is a Gaussian kernel and for each pixel location $x$, the pixel $y$ is in the neighborhood set $\Omega$. The normalization factor $W$ is equal to $\sum_{y\in\Omega} \mathcal{G}_{\theta_1(x)}(\|x-y\|) \cdot \mathcal{G}_{\theta_2(x)}(\|I(x) - I(y)\|)$. Here, the modulation vector $\boldsymbol{\theta}$ has two components: $\theta_1$ and $\theta_2$. They are often globally referred to as *spatial* and *range* variance respectively and control the behavior of function $\varphi$. For example, if $\theta_2$ is predicted higher at a patch location, its corresponding Gaussian kernel widens and flattens behaving like a Gaussian blur [20], and finally, a blurred luminance $L$ is estimated. In such condition, the final tone mapped pixels, which are obtained by normalizing the estimated $L$ for the corresponding patch, will preserve the structures such as gradients.

Notice that we opted for bilateral filtering because its proposed formulation facilitates the integration of the core concept of local modulation. However, any other tone mapping function with parametric formulations such as [10,12] could be used.

### 2.3. Guidance Model based on SVR

Suppose we are given a training set $\{(f_1, o_1), ..(f_n, o_n)\}$, where $f_i$ is the feature sample and $o_i$ represents its corresponding observation (scalar or vector), $i = 1...n$. A classical linear regressor would solve the problem of fitting a prediction function as: $r(f_i) = (\omega^T f_i + b)$, where $\omega, b$ are estimated by minimizing the mean square error. However, such function is often incapable of separating the non-linearly sampled data, like our case where $f_i$ is the SIFT feature with size 128, and $o(i) = \theta_{k(i)}$, where $k = 1, 2$. Therefore, with such given inputs, we use the non-linear SVR [21] which maps the input vector $f_i$ into high dimensional space using the kernel $\psi$ where data becomes linearly separable and is given as $r(f_i) = (\omega^T\psi(f_i) + b)$. To fit the desired non-linear SVR prediction function, the following optimization problem is solved:

$$\min_{\omega,b,\xi,\xi^*} \frac{1}{2}\|\omega^2\| + C\sum_{i=1}^{n}(\xi_i + \xi_i^*)$$

subject to:

$$\theta_{k(i)} - (\omega^T\psi(f_i) + b) \leq \chi + \xi_i,$$
$$(\omega^T\psi(f_i) + b) - \theta_{k(i)} \leq \chi + \xi_i^*,$$
$$\xi_i, \xi_i^* \geq 0, i = 1..n$$

where $\xi, \xi^*$ are the slack variables, $C$ represents the cost which is imposed for samples that exceed the error $\chi$. For further understanding of the non-linear SVR optimization problem, we refer the reader to [21].

### 2.4. Generation of Samples

To train the SVR, we need to find appropriate training features and their corresponding *supervised* observations $\theta_1, \theta_2$ as shown in Fig. 2. To this end, we propose a two step solution. First, we identify *key* locations in a scene, where we can extract meaningful descriptor features. Second, we build a model to find the optimal $\theta_1$

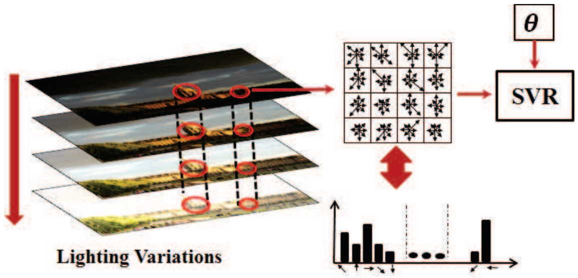


**Fig. 2**: *Training Pipeline*


**Fig. 3**: Scenes from *HDR luminance dataset*. The dataset is composed of 8 scene from different indoor/outdoor locations.

and $\theta_2$ that maximize the similarity between those descriptors which are captured from the same key locations of the scene.

To identify key locations, we first detect keypoints independently in each log-scaled HDR image of the scene using the DoG [14] detector. We then iteratively check, for each detected keypoint, whether it is found at about the same location in other images of the same scene, taken under different illumination conditions. If it is detected in the majority of these images, we call it a *key* location. As we just want to collect 'meaningful' *key* locations with majority occurrence under lighting variations, any other format could also be used instead of log-HDR.

From each key location, we use SIFT [14] as training feature, extracted from linear HDR content. More specifically, it is given as concatenation of 16 unnormalized cells *i.e.*, $[\mathbf{x}_1, ...\mathbf{x}_{16}]$ where each cell can be compactly defined as [22,23]:

$$h(\Theta|p)[\mathbf{x}] = \int \mathcal{G}_\delta(\Theta - \angle\nabla p(y))\mathcal{G}_{\hat{\sigma}}(y-x)\|\nabla p(y)\|d(y) \quad (3)$$

where $\mathbf{x}$ is the center location of the cell in the restricted square patch $p$ of size $16 \times 16$. The independent variable $\Theta$ represents the gradient orientation ranging from $[0, 2\pi]$. Moreover, $\mathcal{G}$ represents the Gaussian kernel with standard deviation $\hat{\sigma}$ and an angular dispersion parameter $\delta$.

**Similarity model**: We assume a scene $S$ consisting of $n$ HDR images with lighting variations as shown in Fig. 2. We consider $P = \{(1,2),(2,3)..,\}$ to be the set of $K = \binom{N}{2}$ pair combinations of $N$ descriptors extracted from a key location. Our aim is to minimize the following objective function:

$$\mathcal{F}(\boldsymbol{\theta}) = \frac{1}{K}\sum_{\{i,j\}\in P}\Phi(\boldsymbol{h}_i(\boldsymbol{\theta}), \boldsymbol{h}_j(\boldsymbol{\theta})). \quad (4)$$

We define function $\Phi$ using the logistic penalty (similar to max-margin formulations in [15]), $\Phi(\boldsymbol{h}_i, \boldsymbol{h}_j) = \log(1+\exp(\epsilon - \boldsymbol{h}_i^T\boldsymbol{h}_j))$. We optimize the objective function in Eq. (4) using a robust optimization technique, Stochastic Gradient Descent (SGD) [17]. SGD update rule to estimate $\theta$ maps at each iteration $t$ is given as: $\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t - \gamma_t \cdot \nabla\Phi_{\{i,j\}t}(\boldsymbol{\theta}_t)$, where $\gamma_t$ is a learning rate which is decayed with $t$ as $\gamma_t = \gamma_0/(t+1)$ and the gradient for the objective in Eq. (4) is replaced (as detailed in [17]) with the gradient of a randomly chosen sample pair $\{i,j\}$ at time $t$, *i.e.*, $\nabla\Phi_{\{i,j\}}(\boldsymbol{\theta}_t) \triangleq \frac{\partial\Phi(\boldsymbol{h}_i,\boldsymbol{h}_j)}{\partial\boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}_t}$.

## 3. RESULTS AND DISCUSSION

### 3.1. Experimental Setup

We build the test setup for image matching using the HDR luminance dataset shown in Fig. 3 which consists of 4 indoor and 4 outdoor scenes as detailed in [8]. We compare our proposed DoTMO with the classical perception based TMOs: BTMO [19], ChiuTMO [12], DragoTMO [24], ReinhardTMO [10] and MantiukTMO [11].

The BTMO in [19] and ChiuTMO [12] are also based on normalizing the estimated luminance $L$ but use global parametric settings. DragoTMO [24] maps the HDR content based on adaptive logarithmic scaling. ReinhardTMO [10] and MantiukTMO [11] are well known tone mapping techniques for high visual quality outputs with appealing brightness and contrast. We considered these TMOs as they have been previously applied for HDR evaluation studies [7,19] for the related task of feature detection.

To effectively evaluate the impact of descriptor extraction scheme, we selected the strongest 500 keypoints using the DoG detector [14] for each tone mapped image. Then, we use four popular and widely used descriptor schemes SURF [25] and SIFT [14], FREAK[26] and BRISK [27] (binary descriptors).

**Metrics**: We evaluated the descriptor performance using the standard measures of Matching Score and mAP rates as detailed in [3,18]. Matching Score is defined as the fraction of correct matches in the minimum of total number of correspondences in the image pair. mAP is calculated as the mean of the area under the precision-recall (PR) curves where recall is defined as the fraction of true positives over total correspondences and precision is given as the ratio of true positives to the total number of matches.

To define a match, we use the standard nearest neighbor distance ratio (NNDR) matching strategy. According to NNDR, a descriptor finds a good match if the ratio between its distance from first closest match and its distance from second closest match is less than a given threshold $th$. Hamming and Euclidean distances are used for binary (BRISK) and non-binary (SIFT, SURF) descriptors, respectively. Two descriptors yield a true positive match if they correspond to two keypoints/regions which are indeed repeated [3] in the reference and query images. Similarly, a match is labeled as a false positive if the corresponding keypoints are not repeated. A PR curve is generated by varying the NNDR threshold.

#### 3.1.1. *Training and Implementation details*

For each test scene, we build the training set with 5000 training samples and use it to train and validate the SVR model. Given a test scene from our dataset (Fig. 3), the training set is drawn from the other 7 scenes. For each training sample, we compute the SIFT feature on a patch size of $16 \times 16$.

**Implementation**. We use the SVR implementation of LibSVM [28]

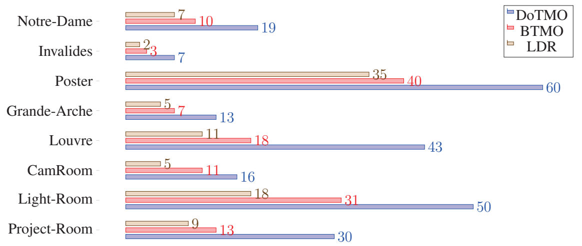


**Fig. 4**: **Matching Score** computed using DoTMO, BTMO and LDR for each test scene using SURF descriptor.

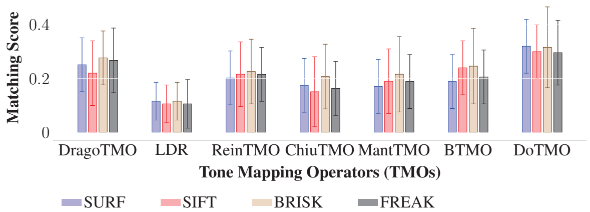


**Fig. 5**: **Average Matching Scores** computed on different TMOs using SURF, SIFT, FREAK, BRISK descriptor extraction schemes. The average is calculated over all test scenes.

**Fig. 7**: Day/Night matching using SURF. Row I: 2 HDR images from *Invalides* scene are displayed after log scaling[24]. Correct and incorrect matches are shown with yellow and red lines respectively. Green lines represent the special case of mismatch due to repetitive structure. Row II: the feature matching using our proposed DoTMO (11 correct and 3 incorrect matches). Row III: using Reinhard TMO (3 correct and 11 incorrect matches). Row IV: using MantiukTMO (4 incorrect and 3 correct matches).

using the Radial Basis Function (RBF) kernel. To obtain the optimal values of SVR parameters, the regularization cost and epsilon-SVR are tuned by 10-fold cross validation from the range of $[2^{-5}, 2^{15}]$ and $[2^{-10}, 2^{5}]$, respectively. We use the HDR Toolbox [29] for the implementation of the considered TMOs, Matlab's Computer Vision toolbox for SURF, FREAK, BRISK and Vlfeat [23] for SIFT.

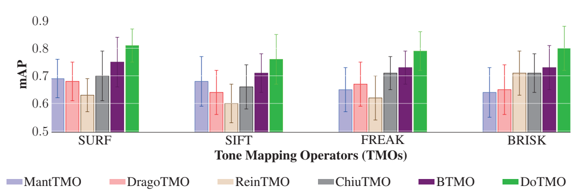


**Fig. 6**: **Mean Average Precision** (mAP) rates computed on different TMOs using SURF, SIFT, FREAK, BRISK descriptor extraction schemes. The average is calculated over all test scenes.

### 3.2. Evaluation Results

We perform a thorough evaluation of our proposed DoTMO quantitatively using the matching score and mAP. We initially show in Fig. 4 the performance of our method over all test scenes using the SURF descriptor, where we compare our algorithm with BTMO [19] and the best exposure LDR. Our results clearly show that predicted local modulation of the bilateral filtering helps in preserving the invariance of the local gradient and hence, boosts the average number of correct matches in both the indoor and outdoor scenes. However, we observe small gains in outdoor scenes such as Invalides. This can be explained by strong lighting transitions and is partially due to increased false matches due to repetitive structures in the images as shown in Fig. 7. Note that, we use threshold $th = 0.2$ to avoid ambiguous matches and to improve the readability of descriptor matching in Fig. 7.

**Comparison with popular TMOs**. We evaluate the performance of our method across different descriptor extraction schemes including both gradient based and binary descriptors. In terms of average matching score, we observe that by using every extraction scheme our DoTMO overall yields a higher number of correct matches, as shown in Fig. 5. Furthermore, in Fig. 6, we compute the mAP rates by averaging the area-under-the-curve of PR curves of the complete dataset. We observe that for every descriptor extraction scheme our proposed model outperforms all the other TMOs. Additionally, we compare our proposed TMO with popular and visually pleasing Reinhard TMO [10] and MantiukTMO [11] in Fig. 7, where we show that our method produces a higher number of correct matches in difficult day/night matching.

## 4. CONCLUSIONS

We propose a novel TMO approach to improve the descriptor discriminability under drastic changes of lighting conditions. To this end, we train a SVR using SIFT features to learn a model which spatially modulates the pixel-wise adaptive TMO. Further, we introduce a simple and effective method for generating the training set to learn the SVR for the given problem. We evaluate our model on our proposed HDR benchmark dataset of indoor/outdoor scenes. Our model achieves significantly better matching score and mean average precision than state-of-the-art TMOs on the HDR dataset and across different descriptor extraction algorithms. In the future, we plan to extend our model to combine detector and descriptor, and explore its usability for real-time problems such as object matching.

# References


[1] F. Dufaux, P. Le Callet, R. Mantiuk, and M. Mrak, *High Dynamic Range Video: From Acquisition, to Display and Applications*, Academic Press, 2016.

[2] A. Chalmers and K. Debattista, "HDR Video Past, Present and Future: A Perspective ," *Signal Processing: Image Communication*, vol. 54, pp. 49 – 55, 2017.

[3] K. Mikolajczyk and C. Schmid, "A Performance Evaluation of Local Descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 10, pp. 1615–1630, Oct. 2005.

[4] H. Zhou, T. Sattler, and D. W. Jacobs, "Evaluating Local Features for Day-Night Matching," in *Computer Vision - ECCV 2016 Workshops - Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part III*, 2016, pp. 724–736.

[5] P. Bronislav, A. Chalmers, P. Zemcík, L. Hooberman, and M. Cadík, "Evaluation of Feature Point Detection in High Dynamic Range Imagery," *Journal of Visual Communication and Image Representation*, vol. 38, pp. 141 – 160, 2016.

[6] A. Rana, G. Valenzise, and F. Dufaux, "Evaluation of Feature Detection in HDR Based Imaging Under Changes in Illumination Conditions," in *IEEE International Symposium on Multimedia, ISM 2015, Miami,USA, December, 2015*, 2015, pp. 289–294.

[7] R. Suma, G. Stavropoulou, E. Stathopoulou, L. V. Gool, A. Georgopoulos, and A. Chalmers, "Evaluation of the Effectiveness of *HDR* Tone Mapping Operators for Photogrammetric Applications," *Virtual Archaeology Review*, vol. 7, no. 15, 2016.

[8] A. Rana, G. Valenzise, and F. Dufaux, "Learning-based Adaptive Tone Mapping for Keypoint Detection," in *IEEE International Conference on Multimedia & Expo (ICME'2017)*, Hong Kong, China, July 2017.

[9] A. Rana, G. Valenzise, and F. Dufaux, "An Evaluation of HDR Image Matching under Extreme Illumination Changes," in *The International Conference on Visual Communications and Image Processing (VCIP)*, Chengdu, China, Nov. 2016.

[10] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, "Photographic Tone Reproduction for Digital Images," *ACM Trans. Graph.*, pp. 267–276, July 2002.

[11] R. Mantiuk, K. Myszkowski, and H. P. Seidel, "A Perceptual Framework for Contrast Processing of High Dynamic Range Images," *ACM Trans. Appl. Percept.*, vol. 3, no. 3, pp. 286–308, July 2006.

[12] K. Chiu, M. Herf, P. Shirley, S. Swamy, C. Wang, and K. Zimmerman, "Spatially Nonuniform Scaling Functions for High Contrast Images," in *Proceedings of Graphics Interface '93*, Toronto, Ontario, Canada, 1993, GI '93, pp. 245–253.

[13] F. Durand and J. Dorsey, "Fast Bilateral Filtering for the Display of High-dynamic-range Images," in *Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques*, 2002, SIGGRAPH '02, pp. 257–266.

[14] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *Int. J. Comput. Vision*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[15] A. Rana, J. Zepeda, and P. Pérez, "Feature Learning for the Image Retrieval Task," in *Computer Vision - FSLCV, ACCV 2014 - Singapore, November 1-2, 2014*, 2014, pp. 152–165.

[16] K. S. Ni and T. Q. Nguyen, "Image Superresolution Using Support Vector Regression," *Trans. Img. Proc.*, vol. 16, no. 6, June 2007.

[17] L. Bottou, *Stochastic Gradient Descent Tricks*, pp. 421–436, Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.

[18] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool, "A Comparison of Affine Region Detectors," *International Journal of Computer Vision*, 2005.

[19] A. Rana, G. Valenzise, and F. Dufaux, "Optimizing Tone Mapping Operators for Keypoint Detection under Illumination Changes," in *2016 IEEE Workshop on Multimedia Signal Processing (MMSP 2016)*, Montréal, Canada, Sept. 2016.

[20] C. Tomasi and R. Manduchi, "Bilateral Filtering for Gray and Color Images," in *Computer Vision, 1998. Sixth International Conference on*. IEEE, 1998, pp. 839–846.

[21] A. J. Smola and B. Schölkopf, "A Tutorial on Support Vector Regression," *Statistics and Computing*, vol. 14, no. 3, Aug. 2004.

[22] J. Dong, N. Karianakis, D. Davis, J. Hernandez, J. Balzer, and S. Soatto, "Multi-view Feature Engineering and Learning," in *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, June 7-12, 2015*, 2015.

[23] A. Vedaldi and B. Fulkerson, "VLFeat: An Open and Portable Library of Computer Vision Algorithms," 2008.

[24] F. Drago, K. Myszkowski, T. Annen, and N. Chiba, "Adaptive Logarithmic Mapping For Displaying High Contrast Scenes," *Computer Graphics Forum*, pp. 419–426, 2003.

[25] H. Bay, T. Tuytelaars, and L. V. Gool, "Surf: Speeded up Robust Features," in *In ECCV*, pp. 404–417. 2006.

[26] R. Ortiz, "FREAK: Fast Retina Keypoint," in *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Washington, DC, USA, 2012, CVPR '12, pp. 510–517.

[27] S. Leutenegger, M. Chli, and R. Y. Siegwart, "BRISK: Binary Robust Invariant Scalable Keypoints," in *Proceedings of the 2011 International Conference on Computer Vision*, Washington, DC, USA, 2011, ICCV '11, pp. 2548–2555.

[28] C.-C. Chang and C.-J. Lin, "LIBSVM: A Library for Support Vector Machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, May 2011.

[29] F. Banterle, A. Artusi, K. Debattista, and A. Chalmers, *Advanced High Dynamic Range Imaging: Theory and Practice*, AK Peters (CRC Press), Natick, MA, USA, 2011.