



On Image Enhancement for Unsupervised Image Description and Matching

Michela Lecca¹ , Alessandro Torresani^{1,2} , and Fabio Remondino¹ 

¹ Fondazione Bruno Kessler - ICT, 38123 Trento, Italy
{lecca, atorresani, remondino}@fbk.eu

² Department of Computer Science, Università degli Studi di Trento,
38123 Trento, Italy

Abstract. An image enhancer improves the visibility and readability of the content of any input image by modifying one or more features related to vision perception. Its performance is usually assessed by quantifying and comparing the level of these features in the input and output images and/or with respect to a gold standard, often regardless of the application in which the enhancer is invoked. Here we provide an empirical evaluation of six image enhancers in the specific context of unsupervised image description and matching. To this purpose, we use each enhancer as pre-processing step of the well known algorithms SIFT and ORB, and we analyze on a public image dataset how the enhancement influence image retrieval. Our analysis shows that improving perceptual features like image brightness, contrast and regularity increases the accuracy of SIFT and ORB. More generally, our study provides a scheme to evaluate image enhancement from an application viewpoint, promoting an aware usage of the evaluated enhancers in a specific computer vision framework.

1 Introduction

An image enhancer is an algorithm that takes as input an image and processes one or more of its features in order to improve the visibility and readability of the visual content. These features usually reflect *perceptual* quality properties, i.e. visual characteristics that are highly significant for the human vision system, like image brightness and contrast, entropy of the color/intensity distribution, level of noise. The performance of an image enhancer is in general assessed by measuring the level of the modified perceptual properties and/or their variations between the input and output images or with respect to an *ideal* image, taken as gold-standard. Many measures have been designed so far to characterize perceptually image enhancement (see e.g. [9, 11, 19]), while to the best of our knowledge, few work has been done to investigate the impact of the perceptual changes in machine vision applications.

In this work we present an empirical evaluation of the image enhancement in the specific context of unsupervised image retrieval. In this framework, image enhancement is often needed to provide a rich and reliable description of the

visual content to be matched under many different circumstances, including difficult conditions due for instance to a wrong set-up of the camera parameters (e.g. low resolution or low exposure time) or to bad illumination (e.g. low-light or back-light) that may adversely affect the detail visibility. In our study, we consider six image enhancers and two image retrieval algorithms. We use each enhancer as pre-processing step of each retrieval routine and we study how the enhancement affects the retrieval performance on a set of images with and without enhancement. To this purpose, we analyze how improving a set of perceptual features (i.e. image brightness, contrast, regularity and color distribution entropy) may influence the retrieval performance that is here measured in terms of number of image descriptors, correct matches and their spatial distribution, and retrieval dissimilarity score.

The enhancers considered here have been chosen among many others available in the literature since they are representative of three different methodologies: statistical local or global analysis (histogram equalization (HE) and contrast-limited adaptive histogram equalization (CLAHE)), spatial color processing with random or deterministic feature sampling (the Milano Retinex algorithms Light-RSR [4] and STAR [13]), and reflectance/illumination image decomposition in constrained domains (LIME [8] and NPEA [20]).

The image description and matching algorithms used here are SIFT [14] and ORB [5], two well known and widely employed methods that, just because based on key-point extraction, require a good visibility of the image details.

We conducted our empirical analysis on the dataset MEXICO recently published on the net [3]. This dataset consists of 40 scenes of real-world indoor and outdoor environments characterized by issues challenging for both the enhancement and retrieval tasks, like the co-existence of dark and bright regions at different proportions, back-light, shadows, chromatic dominants of the illuminant, presence regions with different granularity, from uniform to highly textured.

Our study entails the following contributions: (1) it shows that image retrieval benefits from image enhancement, that enables a richer and more uniform description and matching; (2) it provides a general scheme to evaluate and characterize any image enhancer from an application viewpoint; (3) it promotes an aware use of enhancement techniques in the important field of image description and comparison; (4) finally, since carried out on a public image dataset, it enables further comparison with other methods.

2 Evaluated Algorithms for Image Enhancement

In this Section we briefly describe the six image enhancers considered in our empirical analysis. In the following, we grouped them in three classes upon the methodology and the assumptions they use.

Statistic-Based Image Enhancers - The histogram equalization (HE) and the contrast-limited adaptive histogram equalization (CLAHE) enhance any input image by stretching the probability density functions of one or more image

components in a given color space. In the RGB space, considered here, HE processes the R, G, B channels separately and adjusts the channel intensities to flatten as much as possible the intensity histogram. To this purpose, HE maps any intensity value k of the channel I to the value $T(k)$ given by:

$$T(k) = \text{floor}\left(255 \sum_{i=0}^k h(i)\right), \quad (1)$$

where function *floor* rounds its argument to the nearest greatest integer value and h is the histogram of I normalized to sum up to 1. CLAHE is similar to HE, but works on a set of image patches by redistributing their pixel intensities so that their histogram bins do not exceed a pre-defined threshold (called *clip limit*) that prevents the over-enhancement of uniform image areas.

Retinex Inspired Image Enhancers - Milano Retinexes [17] are spatial color algorithms derived by Retinex theory [12] and thus related to human color vision. They enhance any real-world image by processing spatial and visual features extracted independently from each color channel, according to this equation:

$$L(x) = \frac{I(x)}{w(x)}, \quad (2)$$

where L is the so-called *lightness*, i.e. the enhanced version of the channel I , x is an image pixel and $w(x) \in (0, +\infty)$ is an intensity level named *local reference white at x* . The value of $w(x)$ is computed by processing a set of intensities (in some implementations along with other features) sampled from a neighborhood $N(x)$ of x . Milano Retinexes provide different levels of image enhancement, since the value of $L(x)$ depends on the spatial sampling of $N(x)$, on the features selected from $N(x)$ and on the mathematical expression of $w(x)$. Here we consider Light-RSR [4] and STAR [13] for their computational efficiency.

For each pixel x , Light-RSR samples $N(x)$ by a random spray, i.e. by a set of m pixels randomly selected with radial density around x . The value $L(x)$ is obtained by dividing the intensity $I(x)$ by the maximum intensity over the spray and by smoothing and blurring the result in order to reduce the chromatic noise due to the random sampling.

STAR extracts the features contributing to $w(x)$ from M regions R_1, \dots, R_M obtained by segmenting I with [7]. Precisely, from each segment R_i , STAR selects the maximum intensity $I(R_i)$ and the set $S(R_i)$ of pixels which are most internal to R_i . For any $x \in R_i$, STAR computes the mean value $u(x)$ of the intensities $I(R_j) > I(x)$, each of them weighted by a function inversely proportional to the minimum Euclidean distance between $S(R_i)$ and $S(R_j)$. The value $w(x)$ is obtained by dividing $u(x)$ by the sum of the weights contributing to $w(x)$.

Image Enhancers Based on Illuminant Estimation - Both the algorithms NPEA [20] and LIME [8] rely on the image formation model that represents the color image I as the product of the reflectance \mathcal{R} of the materials depicted in the scene and the illumination \mathcal{I} . Precisely, for any pixel x of I ,

$$I(x) = \mathcal{R}(x)\mathcal{I}(x). \quad (3)$$

In this model, \mathcal{I} and \mathcal{R} express respectively the low- and the high- frequencies of the image. Discounting \mathcal{I} from I allows to retain significant image details while smoothing unessential details, therefore it is a way to enhance the image. NPEA and LIME are grounded on this principle. They estimate \mathcal{I} in a *constrained* domain, since in general the computation of \mathcal{I} and \mathcal{R} from I is an ill-posed problem. Both NPEA and LIME start from a coarse estimate of \mathcal{I} as the maximum intensity over the color channels, then they refine this estimation according to different assumptions. Precisely, NPEA hypothesizes that the reflectance is limited to a specific range and that the local relative order of the image intensities (i.e. the *image naturalness*) slightly changes over adjacent regions. LIME assumes the dark prior channel hypothesis [10] along with slight variations of the illuminant over the image. In addition, LIME imposes the fidelity between the coarse and the final estimation of \mathcal{I} . In NPEA, the enhanced image E is obtained as the product $E(x) = \mathcal{R}_e(x)\sigma(\mathcal{I}_e)$ where \mathcal{R}_e is an estimate of \mathcal{R} obtained by dividing I by the estimate \mathcal{I}_e of \mathcal{I} and σ is a smoothing function introduced to preserve image naturalness. In LIME, no reflectance is estimated, and E is computed from Eq. (3) by as the pixel-wise ratio between I by the estimated illumination \mathcal{I} . Of course, division by zero is always prevented.

3 Evaluated Methods for Unsupervised Image Description and Matching

This section describes the main principles and characteristics of SIFT [14] and ORB [5]. The goal of these algorithms is to match the content of a set of images to identify the common image regions. This is achieved in two phases: (a) *feature extraction*, i.e. identification of salient and locally distinguishable regions of the image, called key-points; (b) *feature description*, i.e. the computation and matching of the descriptors, which are discrete representations summarising the local structure around the detected key-points. The descriptors, in order to be effective, should be invariant to variations such as rotating, scaling and re-lighting.

Scale Invariant Feature Transform (SIFT) - Given an image I , SIFT builds up a pyramid structure whose base level contains the image I at full resolution, while the higher levels contain versions of I sequentially down-sampled. SIFT smooths each down-sampled version I_l of I by n Gaussian filters with increasing variance and computes the so-called *differences of Gaussians*, which encode the pixel-wise differences between the $n - 1$ pairs of subsequent Gaussian smoothed versions of I_l . The key-points are defined as the corners corresponding to local maxima of the differences of the Gaussians within the pyramid. Every key-point is then identified by the quadruple $\langle p, s, r, f \rangle$, where p is the key-point position in I , s is the scale (pyramid level), r is the orientation and f is the descriptor, which is a vector of 128 elements encoding the distribution of the orientation of the image gradients in the 16×16 window $W(x)$ centered at p . To make f invariant to rotations, the dominant orientation of the gradients in $W(x)$ is computed and used to rotate the image before computing f .

In SIFT, the dissimilarity measure between two key-points is defined as the L^2 -distance between their descriptors.

Oriented FAST and Rotated BRIEF (ORB) - ORB is a combination of the feature extractor FAST [18] and the feature descriptor BRIEF [6] with some modifications which enable multi-scale matching and guarantee rotation invariance. In FAST, a pixel x is a key-point if its intensity exceeds by a pre-defined threshold the intensities of a set of pixels y_1, \dots, y_n equi-spaced on a circumference $\Gamma(x)$ centered at x . BRIEF associates to each FAST key-point x the n - dimensional vector whose i -th entry is zero if $I(x) < I(y_i)$ and one otherwise. To achieve invariance against re-scaling, ORB detects the FAST key-points (that are corners) at multiple scales. Moreover, for each key-point x , ORB defines the orientation $\theta(x)$ of x as the angle between x and the intensity weighted centroid of a circular region $C(x)$ around x . Finally, to grant robustness to rotation and noise, ORB computes the BRIEF descriptors of x on the patch $C(x)$ steered by $\theta(x)$ and smoothed by a Gaussian filter.

In ORB, the dissimilarity measure between two key-points is defined as the Hamming distance between their binary descriptors.

4 Evaluation

We assess the performance of each image enhancer by accounting for the variations of both the perceptual features and the retrieval accuracy.

Evaluation in Terms of Perceptual Changes - We quantify numerically the perceptual changes by four features, which reflect perceptual properties usually modified by an image enhancer: mean brightness, multi-resolution contrast [16], histogram flatness and NIQE [15].

Given a color image J , the mean brightness B of J is the mean value of the intensities of the mono-chromatic image \mathcal{B} , obtained by averaging pixel by pixel the channel intensities of J . The multi-resolution contrast \mathcal{C} is the average of the mean contrasts of Z images $\mathcal{B}_1, \dots, \mathcal{B}_Z$ obtained by half-scaling \mathcal{B} sequentially. Here, the mean contrast of \mathcal{B}_s ($s \in 1, \dots, Z$) is the average value of the pixel contrasts $C(\mathcal{B}_s(x))$ with $x \in \mathcal{B}_s$, where $C(\mathcal{B}_s(x))$ is the mean value of the differences $|\mathcal{B}_s(x) - \mathcal{B}_s(y)|$ with y belonging to a 3×3 window centered at x . The histogram flatness F measures the entropy of the probability density function h of \mathcal{B} as the L^1 difference between h and an uniform probability density function. Finally, NIQE [15], here denoted by N , is a measure of *image naturalness*: it quantifies departures of J from image regularity, which is defined in terms of local second-order statistics.

Usually, an image enhancer increases the values of B and C , while decreases those of F and N , namely it makes the input image brighter and more contrasted, while it flattens its color distribution and smooths local irregularities. We observe that the exact amount of B , C , F and N and their variation after enhancement depend on the image at hand. In particular, for already clear images, the variation of B , C , F and N are negligible, while they are remarkable for unreadable images.

Evaluation in Terms of Image Description and Matching - We consider a dataset \mathcal{D} with n indoor and outdoor scenes, each of them represented by m images differing to each other only for the exposure time under which they have been captured. We define the reference of each scene as the image with the lowest value of F : this criterion guarantees that the reference has a good detail visibility, being its brightness distribution the most uniform among those of that scene. We describe the references and the queries by SIFT and ORB with and without enhancement, then we match each input (enhanced, resp.) query Q against the corresponding input (enhanced, resp.) reference R . We evaluate the description and matching performance of SIFT and ORB by the following measures:

- the percentage N_d of images of \mathcal{D} described by at least one key-point: if $N_d < 100\%$, then some images have no key-points;
- the numbers K_R and K_Q of key-points detected respectively on R and Q : in general, when $K_Q \ll K_R$, Q is poorly described with respect to R ; when $K_R \ll K_Q$, the query is over-described and this is often due to a high percentage of noisy pixels that are wrongly detected as key-points; when $K_Q \simeq K_R$, R and Q are likely described similarly, but of course this does not grant that the key-points of R and Q are effectively similar;
- the number M_g of key-points of Q matching key-points of R with the same position on the image (*correct matches*);
- the number M_b of key-points of Q matching key-points of R with different position on the image (*wrong matches*);
- the *mean dissimilarity ratio* σ of Q , computed as follows: we match each key-point x of Q to the key-points of R , we order the key-points of R by their dissimilarity with x (from low to high) and we compute the ratio $\sigma(x)$ between the first and second dissimilarity scores in the ranked list of key-points of R ; σ is the average of the ratios $\sigma(x)$ where x is a key-point of Q correctly matched; the lower σ , the higher the discrimination capability of the algorithm is;
- the flatness S of the spatial distribution of M_g over the image: to this purpose, we partition each query Q in four rectangular, non overlapping blocks Q_1, Q_2, Q_3, Q_4 whose top left corners are defined respectively by $(0, 0), (0, W/2), (H/2, 0), (H/2, W/2)$, where H and W denote the height and width of Q ; the flatter the distribution of the correct matches over these blocks, the more uniform the image description and matching and the higher the robustness of the retrieval algorithm to occlusions are.

The exact values of $N_d, K_R, K_Q, M_g, M_b, \sigma$ depend on the image at hand: for instance, almost uniform images have a low number of key-points that do not vary by enhancement. Nevertheless, we expect that the use of an enhancer as pre-processing step of description and matching procedures increases the values of N_d, K_R, K_Q, M_g and σ , while decreases the value of S . As a drawback, in some cases, the enhancement may increase M_b since it may highlight noisy pixels.

Finally, we also report the retrieval performance of SIFT and ORB obtained by comparing the input (enhanced, resp.) queries versus the input (enhanced,

resp.) references without the constraint on the spatial correspondence between query and reference key-points.

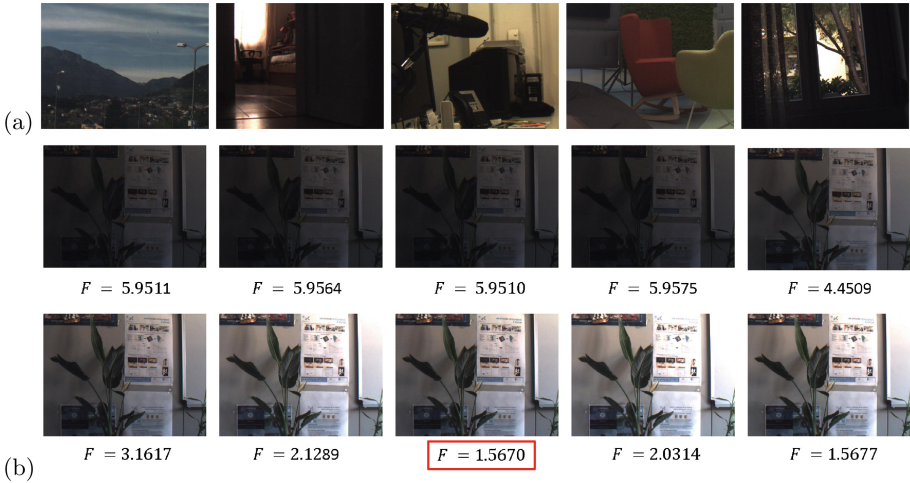


Fig. 1. (a) Some scenes from MEXICO. (b) A scene from MEXICO taken with increasing exposure times and the corresponding values of F . The lowest value of F (in the red box) identifies the reference image of this scene. (Color figure online)

5 Experiments, Results and Conclusions

In our test we employed the dataset MEXICO (Multi-Exposure Image Collection) [3], which consists of 40 scenes of indoor and outdoor environments captured by the FLIR camera [1] and each represented by 10 images acquired with increasing exposure time, ranging from 3 to 30 ms with regular steps of 3 ms (see Fig. 1). In all these images, the blocks Q_i 's are not uniform. As already mentioned in Sect. 1, these scenes present challenging issues for image enhancement, description and matching, like dark and bright regions at different proportions, surfaces differently textured, several light conditions, including shadows, color cast, back-light. The parameters of STAR, LIME and NPEA are set as in their original paper, the clip limit of CLAHE is 8, and the number of spray pixels of Light-RSR is 250. We exploit the ORB and SIFT C++ routines included in OpenCV library [2]. We notify that the implementation of ORB sets to 500 the maximum number of key-points to be extracted from any image.

Figure 2 reports the distributions of the perceptual features B , C , F , N for the MEXICO images. By analyzing their joint distributions, we observed that too low and too high values of B , reported on very dark and saturated image regions, correspond to low values of C (i.e. low visibility of the details) and high values of F and N (i.e. poorly readable image content and noise). Table 1(a)

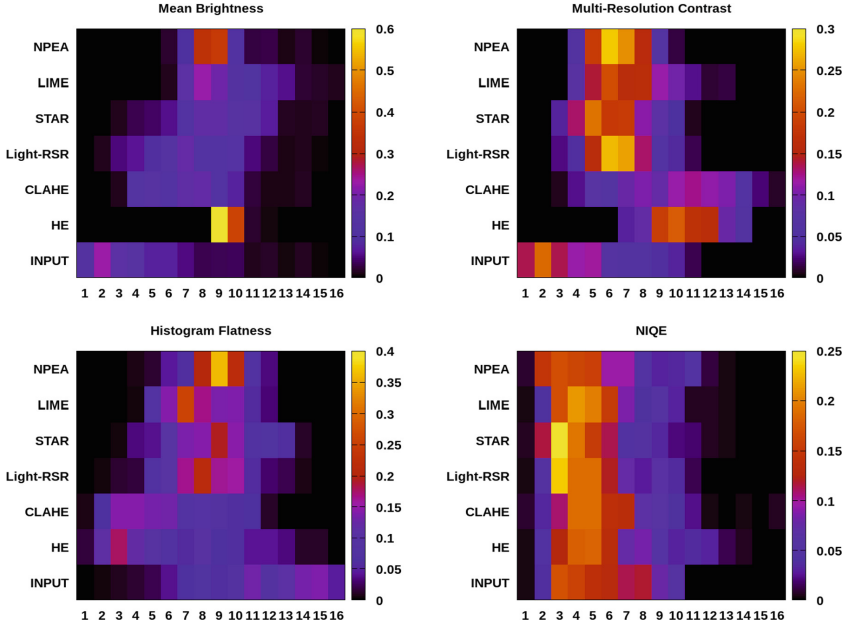


Fig. 2. Distributions (with 16 bins on the x -axis) of the perceptual features of MEXICO.

shows the mean values of B , C , F , N broken down by enhancers. On average, all the enhancers we considered increase the values of B and C , while decrease those of F . The mean value of N obtained on the MEXICO pictures without enhancement (case ‘INPUT’) is smaller than that output by all the enhancers, except for HE and CLAHE that generally tend to over-enhance the images and in this way introduce irregularities and emphasize noise.

For all the cases considered here, both SIFT and ORB described the references of MEXICO by at least one key-point, i.e. for the references $N_d = 100\%$. When no enhancement is used, both ORB and SIFT return a value of N_d smaller than 100%, meaning that no key-points have been detected on some queries (see Fig. 3, left for an example). Precisely, ORB and SIFT cannot describe the 13.89% and the 14.17% of the queries. On the contrary, all the enhanced queries are described by at least one key-point (i.e. $N_d = 100\%$), with K_Q ranging over [183, 500] for ORB and over [28, 7142] for SIFT.

Increasing the contrast is the key-point to improve the performance of ORB and SIFT, since these algorithms are based on the detection of key-points defined in terms of local intensity variations.

We observe that both too dark and saturated image areas present a low contrast value: while enhancers can improve the detail visibility in the dark regions, they cannot recover the visual signal in the saturated portions. Therefore, the values of K_R , K_Q , M_g and σ are higher on the enhanced versions of the queries that originally have been acquired with low exposure time or display

dark regions than of those that originally have an already good detail visibility or that contains saturated areas. The main drawback of image enhancement is due to the generation of many false positive key-points: the enhancement of dark regions where the visual signal is corrupted due to difficult light conditions, often magnify also noisy pixels that are erroneously detected as key-points and thus matched against the reference. As a consequence, the value of M_b increases proportionally to K_Q , determining mismatches that should be removed by post-processing. This phenomenon is particularly evident for HE and CLAHE, that, as already observed above, yield the highest value of N .

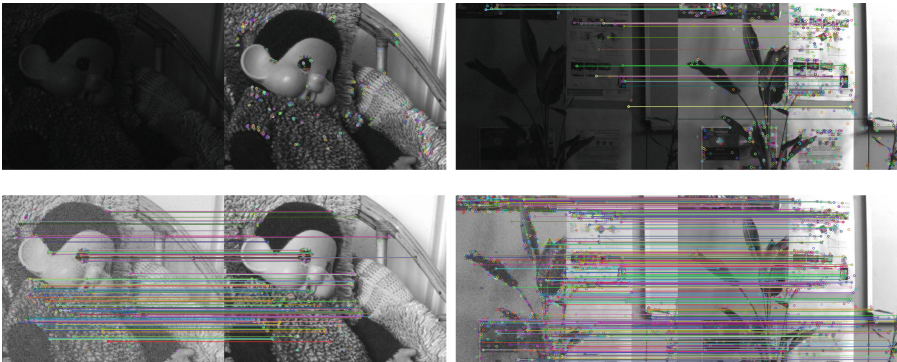


Fig. 3. On top: examples of key-point matching between a query and its reference by ORB (left) and SIFT (right) without enhancement. No key-points are detected on left, while the key-points detected on right are not uniformly distributed over the images. On bottom: key-point matching by ORB (left) and SIFT (right) on the images on top enhanced by STAR: the key-points have been uniformly detected over the images.

The spatial analysis of the distribution of M_g shows that the enhancement enables a more uniform image description, making the matching process more robust to occlusions with respect to the case ‘INPUT’ (see Fig. 3, right). In fact, as displayed in Tables 1(b) and (c), for all the enhancers, the values of S reported by ORB and SIFT are smaller than the case ‘INPUT’. Finally, Tables 1(d) and (e) show that the description of the image is remarkably more uniform when an enhancer is applied. The best results are in general obtained by SIFT.

Additional tests were performed to measure the accuracy of the key-point matching when the queries are matched against all the references. To this purpose, for each image group e (‘INPUT’, ‘HE’, . . . , ‘NPEA’) let Q_e and R_e be the sets of the queries and the references of e . We match each query $q_e \in Q_e$ against each reference $r_e \in R_e$, and we compute the dissimilarity between q_e and r_e as the mean value of the dissimilarities between the key-points of q_e matched with those of r_e , without the check on their spatial location. Table 1(f) shows the rate ρ of image retrieval on the different image groups, i.e. the number of queries assigned to the correct reference divided by the number of queries. The value of

ρ obtained on ‘INPUT’ is smaller than that achieved by enhancing the images, apart from ‘NPEA’ where noisy pixel adversely affect the SIFT performance. The bad results on ‘INPUT’ depend partly on the existence of dark images in which neither SIFT or ORB did not extract and match any feature.

We conclude that our experiments proved that modifying perceptual features like brightness, contrast, color distribution entropy and image regularity generally increases the description and matching performance since the enhancers allow to highlight the relevant details over the whole image. Future work will address the analysis of image enhancement in other machine vision applications.

Table 1. Evaluation summary

(a) Dataset Characterization

| Algorithm | \mathcal{B} | \mathcal{C} | $F[\times 10^{-3}]$ | \mathcal{N} |
|-----------|---------------|---------------|---------------------|---------------|
| INPUT | 51.97 | 9.08 | 5.36 | 4.20 |
| HE | 132.62 | 27.33 | 2.95 | 4.26 |
| CLAHE | 93.86 | 23.90 | 2.88 | 4.23 |
| Light-RSR | 102.65 | 15.92 | 3.80 | 3.94 |
| STAR | 121.56 | 14.97 | 4.11 | 3.81 |
| LIME | 128.07 | 17.62 | 3.68 | 4.02 |
| NPEA | 121.37 | 15.95 | 4.15 | 3.96 |

(b) Results of ORB

| Algorithm | N_d | K_R | K_Q | M_g | M_b | σ | S |
|-----------|-------|-------|-------|-------|-------|----------|------|
| INPUT | 86.11 | 469 | 272 | 133 | 37 | 0.51 | 0.78 |
| HE | 100 | 500 | 500 | 232 | 73 | 0.50 | 0.59 |
| CLAHE | 100 | 500 | 499 | 194 | 89 | 0.58 | 0.73 |
| Light-RSR | 100 | 456 | 487 | 237 | 79 | 0.52 | 0.71 |
| STAR | 100 | 499 | 481 | 233 | 75 | 0.52 | 0.65 |
| LIME | 100 | 499 | 493 | 204 | 93 | 0.58 | 0.66 |
| NPEA | 100 | 488 | 493 | 161 | 103 | 0.66 | 0.76 |

(c) Results of SIFT

| Algorithm | N_d | K_R | K_Q | M_g | M_b | σ | S |
|-----------|-------|-------|-------|-------|-------|----------|------|
| INPUT | 85.83 | 747 | 304 | 150 | 21 | 0.40 | 0.73 |
| HE | 100 | 2240 | 3741 | 545 | 333 | 0.54 | 0.45 |
| CLAHE | 100 | 3705 | 3259 | 657 | 402 | 0.59 | 0.48 |
| Light-RSR | 100 | 961 | 1169 | 349 | 82 | 0.46 | 0.54 |
| STAR | 100 | 1016 | 1123 | 365 | 88 | 0.47 | 0.51 |
| LIME | 100 | 1629 | 1415 | 483 | 145 | 0.54 | 0.49 |
| NPEA | 100 | 1625 | 2442 | 357 | 175 | 0.60 | 0.54 |

(d) Missed Block Matching [ORB]

| Algorithm | Q_1 | Q_2 | Q_3 | Q_4 |
|-----------|-------|-------|-------|-------|
| INPUT | 133 | 117 | 129 | 114 |
| HE | 10 | 23 | 6 | 31 |
| CLAHE | 28 | 34 | 26 | 56 |
| Light-RSR | 41 | 38 | 43 | 44 |
| STAR | 20 | 27 | 29 | 25 |
| LIME | 26 | 27 | 21 | 33 |
| NPEA | 42 | 36 | 47 | 70 |

(e) Missed Block Matching [SIFT]

| Algorithm | Q_1 | Q_2 | Q_3 | Q_4 |
|-----------|-------|-------|-------|-------|
| INPUT | 121 | 115 | 124 | 119 |
| HE | 1 | 9 | 0 | 3 |
| CLAHE | 1 | 13 | 4 | 7 |
| Light-RSR | 11 | 20 | 1 | 9 |
| STAR | 4 | 16 | 7 | 12 |
| LIME | 0 | 14 | 5 | 12 |
| NPEA | 2 | 12 | 9 | 11 |

(f) Retrieval Results

| Algorithm | ρ_{ORB} | ρ_{SIFT} |
|-----------|--------------|---------------|
| INPUT | 0.769 | 0.731 |
| HE | 0.992 | 0.767 |
| CLAHE | 0.986 | 0.983 |
| Light-RSR | 0.925 | 0.853 |
| STAR | 0.969 | 0.881 |
| LIME | 1.000 | 0.917 |
| NPEA | 0.856 | 0.639 |

Acknowledgements. The authors would like to thank Alessio Xompero (with Fondazione Bruno Kessler, IT) and (Queen Mary University, UK), for the fruitful discussions about this topic.

References

1. FLIR camera, datasheet. <https://www.ptgrey.com/firefly-mv-03mp-color-usb-20-micron-mt9v022>
2. OpenCV library. <https://opencv.org/>
3. MEXICO: Multi-exposure image collection (2019). <https://tev.fbk.eu/technologies/image-enhancement-datasets-and-software>
4. Banić, N., Lončarić, S.: Light random sprays Retinex: exploiting the noisy illumination estimation. *IEEE Signal Process. Lett.* **20**(12), 1240–1243 (2013)
5. Bradski, G., Konolige, K., Rabaud, V., Rublee, E.: ORB: an efficient alternative to SIFT or SURF. In: 2011 IEEE International Conference on Computer Vision (ICCV 2011) (ICCV), pp. 2564–2571, November 2011
6. Calonder, M., Lepetit, V., Ozuysal, M., Trzcinski, T., Strecha, C., Fua, P.: BRIEF: computing a local binary descriptor very fast. *IEEE Trans. Pattern Anal. Mach. Intell.* **34**(7), 1281–1298 (2012)
7. Felzenszwalb, P.F., Huttenlocher, D.P.: Efficient graph-based image segmentation. *Int. J. Comput. Vis.* **59**(2), 167–181 (2004)
8. Guo, X., Li, Y., Ling, H.: LIME: low-light image enhancement via illumination map estimation. *IEEE Trans. Image Process.* **26**(2), 982–993 (2017)
9. Lihuo, H., Fei, G., Weilong, H., Lei, H.: Objective image quality assessment: a survey. *Int. J. Comput. Math.* **91**(11), 2374–2388 (2014)
10. He, K., Sun, J., Tang, X.: Single image haze removal using dark channel prior. *IEEE Trans. Pattern Anal. Mach. Intell.* **33**(12), 2341–2353 (2011)
11. Kamble, V., Bhurchandi, K.M.: No-reference image quality assessment algorithms: a survey. *Optik* **126**(11), 1090–1097 (2015)
12. Land, E.H., McCann, J.J.: Lightness and Retinex theory. *J. Opt. Soc. Am.* **1**, 1–11 (1971)
13. Lecca, M.: STAR: a segmentation-based approximation of point-based sampling Milano Retinex for color image enhancement. *IEEE Trans. Image Process.* **27**(12), 5802–5812 (2018)
14. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **60**(2), 91–110 (2004)
15. Mittal, A., Soundararajan, R., Bovik, A.C.: Making a completely blind image quality analyzer. *IEEE Signal Process. Lett.* **20**(3), 209–212 (2013)
16. Rizzi, A., Algeri, T., Medeghini, G., Marini, D.: A proposal for contrast measure in digital images. In: 2nd European Conference on Color in Graphics, Imaging, and Vision and Sixth International Symposium on Multispectral Color Science, CGIV 2004, Aachen, pp. 187–192 (2004)
17. Rizzi, A., Bonanomi, C.: Milano Retinex family. *J. Electron. Imaging* **26**(3), 031207 (2017)
18. Rosten, E., Drummond, T.: Machine learning for high-speed corner detection. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) *ECCV 2006*. LNCS, vol. 3951, pp. 430–443. Springer, Heidelberg (2006). https://doi.org/10.1007/11744023_34
19. Phadikar, B.S., Maity, G.K., Phadikar, A.: Full reference image quality assessment: a survey. In: Bhattacharyya, S., Sen, S., Dutta, M., Biswas, P., Chattopadhyay, H. (eds.) *Industry Interactive Innovations in Science, Engineering and Technology*. LNNS, vol. 11, pp. 197–208. Springer, Singapore (2018). https://doi.org/10.1007/978-981-10-3953-9_19
20. Wang, S., Zheng, J., Hu, H.-M., Li, B.: Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE Trans. Image Process.* **22**(9), 3538–3548 (2013)