Optimized Contrast Enhancements to Improve Robustness of Visual Tracking in a SLAM Relocalisation Context

Xi Wang, Marc Christie, Eric Marchand

***Abstract*— Robustness of indirect SLAM techniques to light changing conditions remains a central issue in the robotics community. With the change in the illumination of a scene, feature points are either not extracted properly due to low contrasts, or not matched due to large differences in descriptors. In this paper, we propose a multi-layered image representation (MLI) in which each layer holds a contrast enhanced version of the current image in the tracking process in order to improve detection and matching. We show how Mutual Information can be used to compute dynamic contrast enhancements on each layer. We demonstrate how this approach dramatically improves the robustness in dynamic light changing conditions on both synthetic and real environments compared to default ORB-SLAM. This work focalises on the specific case of SLAM *relocalisation* in which a first pass on a reference video constructs a map, and a second pass with a light changed condition relocalizes the camera in the map.**

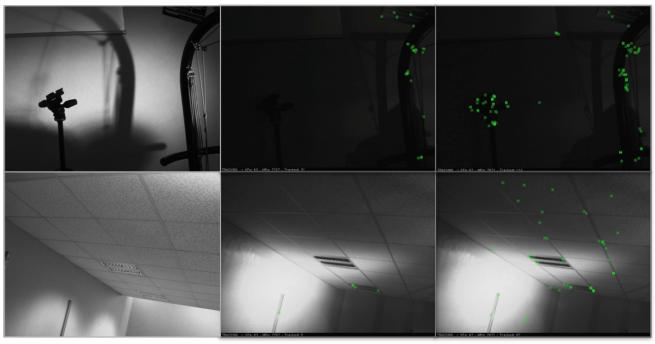
I. INTRODUCTION

Visual tracking systems such as SLAM and visual odometry are widely used in industrial and consumer devices. Except for *direct methods*, *eg.* [1] working on the analysis of changes in pixel gradients, most visual SLAMs rely on corner detection with *extractors* that extract keypoints (KP) and *descriptors* that identify and match the extracted KPs over different frames.

Unfortunately, the corner detection process and consequently the matching problem are strongly dependent on the illumination condition at the moment of capturing images. Although the matching process usually relies on gradient information that is more or less independent from intensity, SLAM methods still suffer from illumination changes at different degrees and may yield inaccurate maps and even tracking failures [2], [3].

In this paper, we propose to improve the robustness of indirect SLAM techniques to light changing conditions by using a multi-layered image representation (MLI). We target the specific case of relocalisation, in which a first pass in a given lighting condition is used to construct a map, and a second pass in a different and dynamic lighting condition is performed to relocalise the camera in a way similar to [4] or more recently [2]. The idea of MLI is to dynamically generate k contrast enhancements of an input image into k layers and apply KP detection on each image layer to improve detection and hence KP matching and tracking (see Fig. 1). The challenge therefore holds in how to compute the optimal parameters of each contrast enhancement to maximize keypoint detection and matching.

Author are with Univ Rennes, Inria, CNRS, IRISA, France.
Email:{xi.wang}@inria.fr
{marc.christie, eric.marchand}@irisa.fr



a. Reference Light Condition b. ORB-SLAM c. Our method

Fig. 1: Keypoint tracking results in different lighting conditions. Only a few keypoints are matched between reference condition (image a) and standard ORB-SLAM (image b), compared to our novel MLI method (image c).

In this paper, we rely on the information theory approach of Mutual Information to compute the optimal contrast enhancements on each frame of a video sequence.

The contributions of this paper are:

- a multi-layered image representation using different contrast enhancements to improve detection and matching of keypoints
- an efficient process to compute optimal parameters for these contrast enhancements using an information theoretic approach
- a dramatic improvement of ORB-SLAM tracking in light changing conditions.

Results displayed in Fig. 1 shows that our approach outperforms the default ORB-SLAM technique in terms of percentage of tracked frames in a sequence, hence demonstrating a stronger robustness to light changing conditions.

II. RELATED WORK

Robustness to light changing conditions is a central issue that has received increased attention. The issue has often been tackled at the *extractor* level by searching an optimal contrast threshold in the KP extractor with respect to the current lighting condition. For example, in SuperFast [5] the FAST contrast threshold – a threshold value that triggers a brighter, darker or similar decision on per-pixel comparison – is dynamically computed using a feedback-like optimization method that yields a new threshold value per region in the image. Lowering the threshold however tends to generate a large number of KPs that influence the computational capac-

ity of other processes, and the proposed technique requires specific adaptations to be applied to other KP detectors.

Another possibility consists in applying image transformations (*eg.* contrast enhancers) on captured images before applying KP detectors. Interestingly, it has been demonstrated that KP extractors gain significant performance by using HDR images as input, converted to SDR images through tone-mapping operators [6], [7]. Among these techniques, a learning-based optimal tone-mapping operator has been proposed for SIFT-like detectors [8]. But the high computational cost and specific HDR devices required, as well as HDR-customized extractors hamper the wider applicability of such approaches. In comparison, for SDR images, research has mainly focused on contrast enhancement operators for aesthetic and perceptional goals through changes in the exposure times [9] which remain limited in addressing robustness of KP tracking.

For direct and semi-direct SLAM methods, *i.e.* methods that rely on analysis of pixel intensities rather than extracting intermediate features, robustness to illumination changes has been addressed by optimizing an affine brightness transfer transformation between consecutive frames [10], [1] or matching a dimension reduced deep-learning feature [11]. Using mutual information instead of photometric error as the metric during the optimization process of pose estimation has also demonstrated its benefits [12], [2]. While exhibiting a good robustness to illumination changes, these methods remain computationally expensive.

III. MULTI-LAYERED IMAGES

Our approach consists in computing, for every frame, a number of contrast enhancements of the original camera image into different images (called layers) before applying keypoint detection on each layer. The idea is inspired by the significant gain obtained in keypoint tracking on tone mapped images of High Dynamic Range (HDR) images [6], [7], but replacing an HDR image by artificially-exposed multiple images in a Multi-Layered Image representation called MLI. These enhancements are computed in a way to improve the detection and matching of keypoints. The parameters of the first contrast enhancement are searched by maximising the shared information between a reference image (in a good lit condition) and the contrast-enhanced image. We rely on a Mutual Information metric, an information theory measure of dependence between two random variables (images), and demonstrate the relevance of this metric in keypoint tracking (see Section IV). The parameters of the second layer are then searched by maximising the mutual information with the reference image, without the information already provided by the first layer. Other layers are computed in a similar incremental way. Given that landscapes of the Mutual Information metrics are difficult to optimize, a specific smoothing process is proposed that enables the use of straightforward gradient descent optimization techniques.

The contrast enhancement technique relies on a saturated affine brightness transfer per-pixel function (SAT). We use a SAT form that defines a *contrast band* $\mathbf{u} = (a, b)^\top$ which

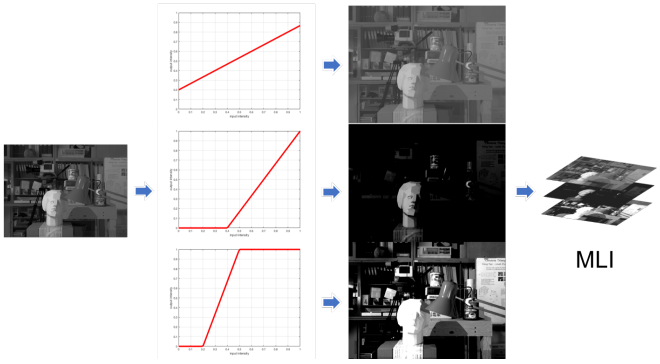


Fig. 2: Using SAT function with different contrast bands to generate a multi-layered image representation (MLI).

conveniently models the lower cut point (a) and higher cut point (b) of the saturation, with a linear interpolation between a and b on pixel intensity i (see Fig. 2). A given contrast $\mathbf{u} = (a, b)^\top$ is defined in a contrast space $\Gamma \subseteq \mathbb{R}^2$, where Γ is the space of all contrast bands where $b > a$.

$$f_{SAT}(i, (a, b)^\top) = \mathit{min}(\mathit{max}(0, i/(b - a)), 1) \qquad (1)$$

Parameters a and b naturally represent the *band* region where the contrast is enhanced, which motivated the choice of this operator. To ensure enhancement or compression of contrasts, we define the range of values for $\mathbf{u} = (a, b)^\top$ as $a \in [-\infty, 1]$ and $b \in [0, \infty]$. The computation of a layer k in our MLI representation is performed by applying the following operator MLI_k on all pixel intensities of the image using a contrast band $\mathbf{u}_k$. A MLI is therefore represented as a set of k image layers where $MLI_k(I) = f_{SAT}(I, \mathbf{u}_k)$ for an image I, where $f_{SAT}(I, \mathbf{u}_k)$ is the application of f_{SAT} on all pixels of I.

This MLI representation can be seamlessly integrated in a state of the art SLAM tracking technique such as ORB-SLAM, replacing the keypoint detection of the camera image, by (i) a set of contrast enhancements of the camera image, and (ii) keypoint detection applied on each enhanced image (see Section V).

IV. OPTIMAL IMAGE ENHANCEMENT

The challenge consists in computing the best parameters for each contrast enhancement on each camera image in a way to improve detection and matching of keypoints. Our hypothesis is that we can compute a close to optimal value of $\mathbf{u}_k$ by maximising Mutual Information between a well lit reference image I^* and a transformed test image $f_{SAT}(I, \mathbf{u}_k)$.

A. Mutual Information

Mutual information (MI) was initially introduced in information theory [13], and then widely applied in the field of computer vision for image alignment, model registration as well as visual tracking and SLAM [2], [14], [15]. The MI built from the image entropy of two different images provides a measure of their mutual dependence. In image alignment

tasks, the higher the mutual information, the better the alignment since mutual information considers the distribution of the intensities as well as the intensities themselves.

Entropy $h(I)$ is a variability measure of a random variable I. In image alignment or illumination evaluation scenarios, I is regarded as one image with r the possible values (gray-level intensities) of I. Equation $p_I(r)=P(I=r)$ therefore expresses the probability distribution function of r, in other words the normalized histogram of the image. The Shannon entropy $h(I)$ of an image I is expressed as:

$$h(I) = -\sum_{r} p_I(r) log(p_I(r)) \tag{2}$$

With the same principle, the joint entropy $h(I, I^*)$ of two images I and I^* can be defined in the following way:

$$h(I, I^*) = -\sum_{t,r} p_{II^*}(t, r) log(p_{II^*}(t, r)) \tag{3}$$

where t and r are the possible grey-level intensities of the I and I^*. The joint probability distribution function is defined as $p_{II^*}(t, r)=P(I=t \cap I^*=r)$, which can also be regarded as a normalized bi-dimensional histogram of images I and I^*.

With the above notations of entropy and joint entropy, the mutual information (MI) is expressed as the intersection of two random variables I and I^* (see Fig. 3):

$$MI(I, I^*) = h(I) + h(I^*) - h(I, I^*) \tag{4}$$

B. Optimal enhancement for the first layer

We need to search for the optimal parameter $\mathbf{u}^*$ that maximises the MI between a reference image I^* (*eg.* an image lit in normal lighting conditions) and a contrast enhanced version of camera image $f_{SAT}(I, \mathbf{u})$.

$$\mathbf{u}^* = \underset{\mathbf{u}}{\operatorname{argmax}}\, MI(f_{SAT}(I, \mathbf{u}), I^*) \tag{5}$$

We can empirically show that the MI has similar behavior to the ground truth wrt illumination changes. Given a reference image I^* under a given light condition, and a test image I in a different lighting condition, the computation of the ground truth (*i.e.* the absolute optimal enhancement) can be performed by an exhaustive sampling of the contrast band parameter $\mathbf{u}_k$, applying corresponding image transforms on I and evaluating the number of matched keypoints between I^* and $f_{SAT}(I, \mathbf{u}_k)$, as displayed in Fig. 7.

We illustrate this on an example from the NewTsukuba data set [16]. We compare the landscapes generated by sampling $\mathbf{u}_k$ on (1) the ground truth ORB detector and on (2) mutual information Eq. (4). Despite differences, we observe the optimums are positioned at similar contrast band values. In an obvious way, the more information is shared between a reference image and a contrast-enhanced image, the better are the detection and matching.

C. Optimal enhancements for other layers

The parameters of the second layer are searched by maximising the Mutual Information with the reference image, as well as the information already provided by the first layer (see Fig. 3).

This can be expressed as multivariate mutual information with the definition of higher dimensional joint probability distribution, to account for multiple image layers. From Eq. (2) and (3), a joint entropy of 3 random image variables is obtained with a definition of normalized tri-dimensional histogram $p_{II^*I^0}(t, r, w)=P(I=t \cap I^*=r \cap I^0=w)$ where t, r, w are possible gray-levels of each image respectively.

$$h(I, I^*, I^0) = -\sum_{t,r,w} p_{II^*I^0}(t, r, w) log(p_{II^*I^0}(t, r, w)) \tag{6}$$

Similarly, a multivariate mutual information between three images can be formulated:

$$\begin{aligned} MI(I, I^*, I^0) = h(I, I^*, I^0) + h(I) + h(I^*) + h(I^0) \\ -h(I, I^*) - h(I, I^0) - h(I^0, I^*) \end{aligned} \tag{7}$$

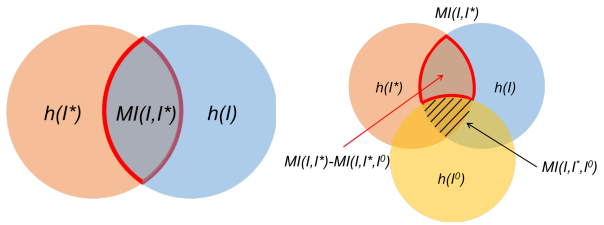


Fig. 3: Mutual information between two images (left), and three images (right) defined as the shared entropy between images.

Given I^* an image under reference light conditions, I the test image for which the contrast bands need to be computed, and $I^0 = f_{SAT}(I, \mathbf{u}_0)$ the first contrast band computed by Eq. (9), we can express the tri-variable mutual information to represent the low-correlated information generated in second layer (see Fig. 3 right red line).

$$\begin{aligned} MI(I, I^*) - MI(I, I^*, I^0) = MI(I^*, I|I^0) \\ = h(I, I^0) + h(I^*, I^0) - h(I, I^*, I^0) - h(I^0) \end{aligned} \tag{8}$$

Given the contrast band from first layer $\mathbf{u}_0$, the optimization of second layer is carried out as follows:

$$\mathbf{u}^* = \underset{\mathbf{u}}{\operatorname{argmax}}\, MI(I^*, f_{SAT}(I, \mathbf{u})|f_{SAT}(I, \mathbf{u}^0)) \tag{9}$$

The computation of further layers can be expressed in a similar way. The mutual information between four images $MI(I, I^*, I^0, I^1)$ or more is computationally expensive to achieve. However a reasonable approximation can be computed using mutual information of previous optimal image $MI(I, I^*, I^1)$ to replace $MI(I, I^*, I^0, I^1)$, which balances the computational cost and preciseness.

D. Smoothing Mutual Information

Derivative optimization approaches favor smoother objective function landscape to make sure an efficient descent to the optimum. Unlike image alignment where reducing the number of bins is important to smooth the cost function (image alignment indeed concentrates more on the geometric information instead of illumination information in one image [17], [14], [2]). For illumination estimation, lowering the histogram bins during the estimation does not gain any benefits but loses information. This is illustrated on another example from the NewTsukuba data set in Fig. 4 by presenting the landscape of the cost function $MI(I^*, f_{SAT}(I, \mathbf{u}))$ (see Fig. 5).

a. reference condition b. test condition

Fig. 4: Images I^* and I from NewTsukuba data set [16]: a synthetic data set with identical camera trajectories and variant illumination conditions

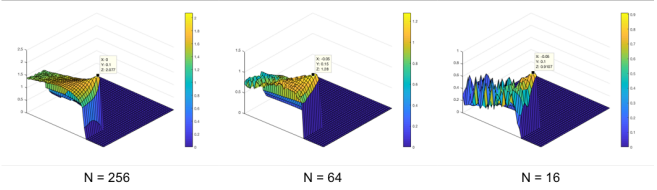


Fig. 5: Lowering the number of histogram bins (which is classical for image alignment tasks) leads to a difficult-to-optimize cost function landscape during illumination estimation process (axis represent parameters a and b of contrast band $\mathbf{u} = (a, b)^\top$).

In our work, we selected 256 bins for the purpose of illumination estimation combined with an *sigmoid-smoothed* SAT function Eq. (11). $f_{SgSAT}(I, \mathbf{u})$ based on Eq. (1) supporting a more curved transition at cutting points which leads to a smoother and less aliased cost function landscape (see Fig. 6). Here we show our sigmoid function $Sg(x)$ with $k = 1/(b-a)$:

$$Sg(x) = \frac{1}{1 + 8ke^{(-(x-(b+a)/2)))}} \quad \textit{with } \mathbf{u} = (a, b)^\top \tag{10}$$

With the definition of d as the length of contrast band, $d = b - a$, we have the *sigmoid-smoothed* SAT function $f_{SgSAT}(I, \mathbf{u})$ defined as:

$$f_{SgSAT}(I, \mathbf{u}) = max(1-d, 0) \times Sg(I) + min(d, 1) \times f_{SAT}(I, \mathbf{u}) \tag{11}$$

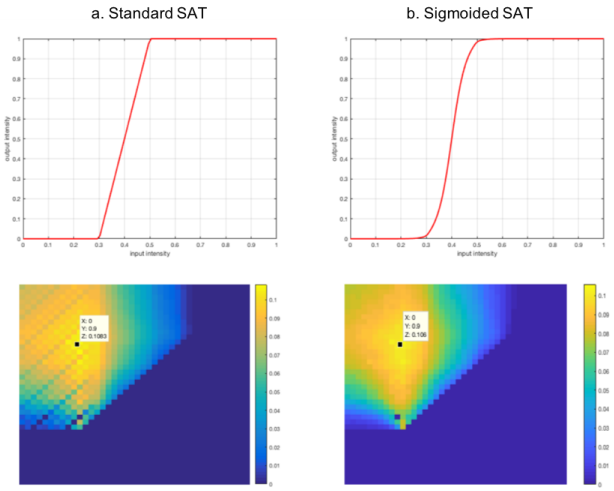


Fig. 6: Comparison between standard SAT function and sigmoid-smoothed SAT function wrt $\mathbf{u} = (0.3, 0.5)^\top$. Second row shows the conditional mutual information $MI(I^*, I|I^0)$ computed by Eq. (8), notations keep as aforecited for I^* and I with I^0 generated by the optimal contrast band from $MI(I^*, f_{SAT}(I, \mathbf{u}))$. We see clearly that standard SAT causes aliasing-like effect in the landscape, due to derivative discontinuity at cutting points a, b

E. Optimization Framework

Using the concepts introduced, we propose an optimization framework that computes a multi-layered image representation for each frame of a video sequence.

As illustrated in Algo. 1, the first step relies on the cost function of standard mutual information Eq. (5) to compute the optimal result of the first layer. The following layers are then computed by optimizing a cost function of conditional mutual information with the previous result aiming to find the best contrast band which benefits most low-correlated information knowing the existence of previous ones. I^* represents the image under reference light condition and I the current test image to optimize, $\mathbf{u}_i, i = 0..N$ is referring to the computed optimal contrast band of each layer with a layer number N.

Algorithm 1 Optimal MLI Generated by MI

1: $i \leftarrow 0$
2: $\mathbf{u}_0 \leftarrow \text{argmax}_{\mathbf{u}}(MI(I^*, f_{SgSAT}(I, \mathbf{u})))$
3: **while** $i < N$ **do**
4: $\quad \mathbf{u}_i \leftarrow \text{argmax}_{\mathbf{u}}(MI(I^*, f_{SgSAT}(I, \mathbf{u})|f_{SgSAT}(I, \mathbf{u}_{i-1})))$
5: $\quad i \leftarrow i + 1$
6: **end while**
7: **return** $\{\mathbf{u}_k\}_{k=1..N}$

A demonstration with the NewTsukuba data set (see Fig. 4) in Fig. 7 illustrates the idea of our multiple step optimization framework. First step is the computation of MI between two images, shown in Fig. 7 representing the

first layer (layer 1). The second and third layer rely on the computation of the first layer and instead of optimizing a standard MI cost function, a conditional mutual information cost function is optimized using Eq. (9). Comparing with the ground truth generated by ORB [18] detector, our proposed method presents a highly similar behavior as well as a characteristic of derivability for gradient optimization methods.

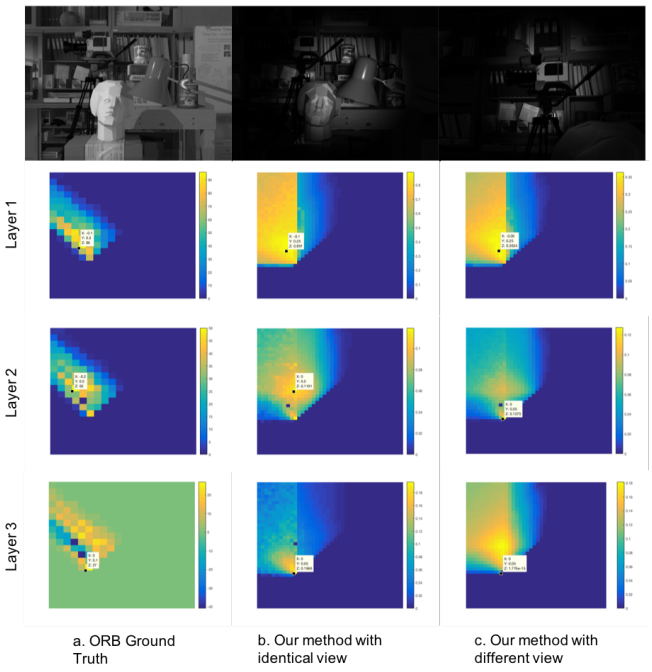


Fig. 7: In each layer, similar behaviors are shown with regard to optimums (see image a,b), compared with ORB detector. Ground truth for layers 2 and 3 are generated by a subtraction between the common keypoints detected from the previous optimal contrast band $\mathbf{u}^*$ and the keypoints from all others contrast bands *i.e.* the ground truth of layer 2 is computed by removing all keypoints common with keypoints detected in previous optimum $\mathbf{u}^* = (0, 0.15)^\top$ in layer 1. Empirical results also show that even with relatively different reference images (c), the landscapes are similar.

V. EVALUATION AND EXPERIMENTS

To evaluate the benefits of our approach in visual SLAM relocalisation tasks, we first select a synthetic scene benchmark under different static and dynamic lighting conditions [16]. This dataset encompasses four videos rendered with identical virtual camera trajectories in a synthetic scene with different illumination conditions (Daylight, Fluorescent, Lamps, Flashlight).

We then designed a real scene benchmark in different static and dynamically changing lighting conditions by executing a same camera trajectory using a robotic arm (see companion videos). In both benchmarks, using a reference video in a given lighting condition, we tested the robustness of our approach compared to default ORB-SLAM in localising the camera from the second video sequence against the keyframes generated from the first video sequence, in a way similar to [4] or more recently NID-SLAM [2]. In each benchmark, we report the success rate, *i.e.* the percentage of the frames from second video successfully relocated. Our implementation is integrated in ORB-SLAM [19].

Results of NewTsukuba data set are displayed in Table. I. The table reports an improved success rate against illumination changing environments in all but one condition, and provides a 96.5% success rate where both default ORB-SLAM and NID-SLAM fail (0%) in the Lamps to Flashlight condition, the reason about the ill-optimized performance needs to be investigated in the future work. Optimal contrast bands for the sequences are computed with a relative low sampling frequency wrt to image acquisition frequency (renew a contrast band every 5 to 10 frames).

V_2 \ V_1	Daylight			Fluo			Lamps			Flash		
	NID	ORB	MLI	NID	ORB	MLI	NID	ORB	MLI	NID	ORB	MLI
Daylight	99.3	100	**100**	96.7	96.2	**100**	73.9	97.6	**99.7**	74.6	79.8	**90.7**
Fluo	95.0	88.1	**99.8**	99.7	100	**100**	85.3	93.9	**100**	95.8	**100**	90.5
Lamps	88.3	55.7	**99.0**	93.6	79.8	**94.1**	93.1	100	**100**	84.3	37.9	**92.4**
Flash	23.8	30.7	**92.8**	92.2	90.6	**94.6**	0.00	0.00	**96.5**	92.0	**100**	99.3

TABLE I: SLAM keyframe retrieval success rate between our MLI implementation, default ORB-SLAM and NID-SLAM.

In the case of real scenes, we placed a monocular camera on a trajectory memory 7 DoF Franka robot arm to guarantee that the camera movement in each video is identical. In comparison with the synthetic scene, a strongly dynamic lighting condition is introduced, by having two operators randomly move spots lights in the experiment scene. Using the same *success rate* criterion, our MLI ORB implementation managed to track **100%** of the dynamically lit scene against a keyframe map generated under normal lighting condition, while normal ORB-SLAM only retrieved **52.12%** of the keyframes. Fig. 9 shows the inlier keypoints after graph optimization process (details see ORB-SLAM [19]), which can be regarded as trusty tracked points generated in the current frame. It demonstrates that MLI performs dramatically better than default ORB which frequently lost tracking during the video. Screenshots of the experiment video are displayed in Fig. 8. A better tracking quality especially around dark or over-exposed area of non-uniformly light images can be observed.

VI. CONCLUSIONS

We have introduced a novel multi-layered image representation based on mutual information optimization to tackle the illumination robustness problem in SLAM relocalization tasks. Each layer in MLI provides low-correlated information which helps to enhance the contrast and therefore increase the robustness during keypoints tracking process under varying illumination conditions. The optimal parameters are computed using a multiple steps mutual information optimization framework. The proposed method shows significant improvements on both synthetic and real videos.

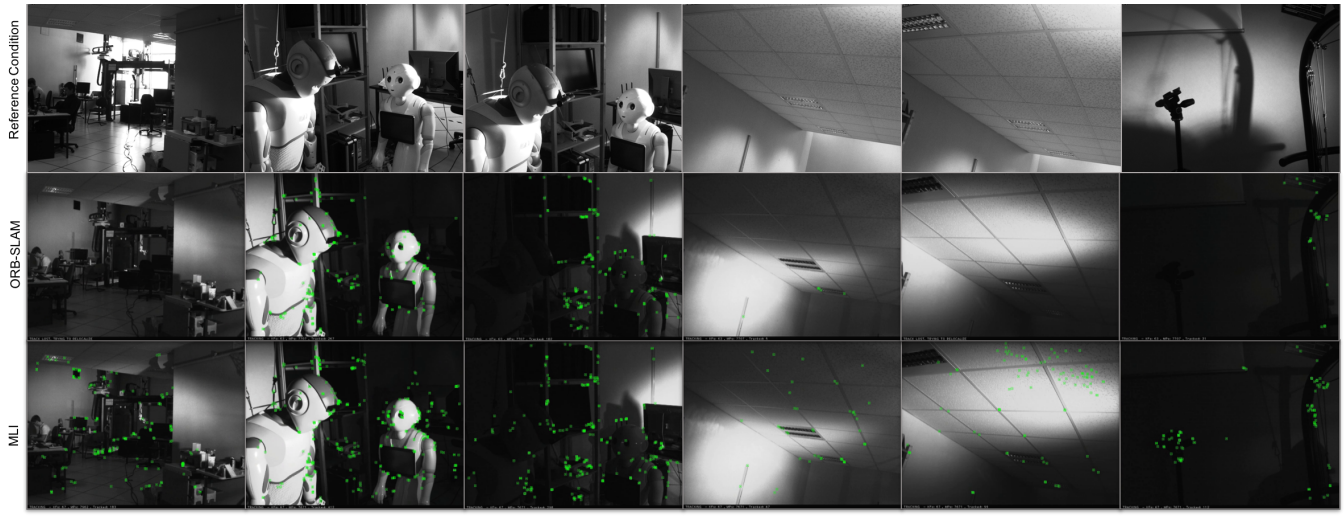


Fig. 8: Results of real scene against dynamic illumination variance. With a keyframe map generated under reference condition, MLI shows a better retrieve capacity especially when encountering non-uniform illumination variance. In contrast, standard ORB-SLAM only tracks the well lighted parts in the image.

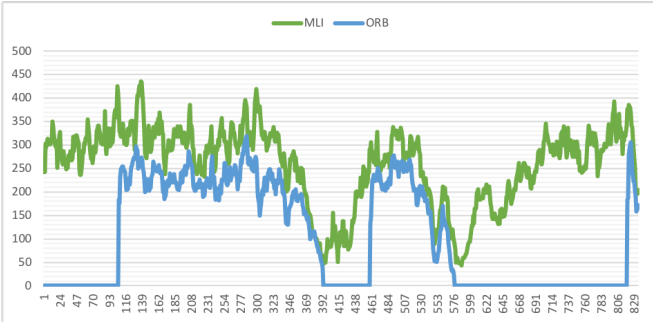


Fig. 9: The number of trusty inlier keypoints after graph optimization of ORB-SLAM which is a critical indicator displaying the tracking quality. MLI generates significantly better results than standard ORB under the dynamic light changing environment.

REFERENCES

[1] J. Engel, T. Schöps, and D. Cremers, "LSD-SLAM: Large-scale direct monocular SLAM," in *European Conference on Computer Vision (ECCV)*, September 2014.

[2] G. Pascoe, W. Madder, M. Tanner, P. Piniés, and P. Newman, "Nid-slam: Robust monocular slam using normalised information distance," in *Conf. on Computer Vision and Pattern Recognition*, 2017.

[3] P. Seonwook, S. Thomas, and P. Marc, "Illumination change robustness in direct visual slam," in *2017 IEEE Int. Conf. on Robotics and Automation (ICRA)*, 2017.

[4] E. Royer, M. Lhuillier, M. Dhome, and J.-M. Lavest, "Monocular vision for mobile robot localization and autonomous navigation," *International Journal of Computer Vision*, vol. 74, no. 3, pp. 237–260, Sep 2007. [Online]. Available: https://doi.org/10.1007/s11263-006-0023-y

[5] G. Florentz and E. Aldea, "Superfast: Model-based adaptive corner detection for scalable robotic vision," in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 2014, pp. 1003–1010.

[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, pp. 289–294.

[7] B. Přibyl, A. Chalmers, and P. Zemčík, "Feature point detection under extreme lighting conditions," in *Proceedings Spring Conf. on Computer Graphics*. ACM, 2013, pp. 143–150.

[8] A. Rana, G. Valenzise, and F. Dufaux, "Learning-Based Tone Mapping Operator for Image Matching," in *IEEE Int. Conf. on Image Processing (ICIP)*. Beijing, China: , Sep. 2017.

[9] J. Yuan, L. Sun, "Automatic exposure correction of consumer photographs," *European Conf. on Computer Vision (ECCV)*, pp. 771–785, 2012.

[10] J. Engel, V. Koltun, and D. Cremers, "Direct sparse odometry," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PP, no. 99, pp. 1–1, 2017.

[11] H. Zhang, X. Wang, X. Du, M. Liu, and Q. Chen, "Dynamic environments localization via dimensions reduction of deep learning features," in *Computer Vision Systems*, M. Liu, H. Chen, and M. Vincze, Eds. Cham: Springer International Publishing, 2017, pp. 239–253.

[12] G. Caron, A. Dame, and E. Marchand, "Direct model based visual tracking and pose estimation using mutual information," *Image and Vision Computing*, vol. 32, no. 1, pp. 54–63, 2014.

[13] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948. [Online]. Available: http://dx.doi.org/10.1002/j.1538-7305.1948.tb01338.x

[14] A. Dame and E. Marchand, "Second-order optimization of mutual information for real-time image registration," *IEEE Transactions on Image Processing*, vol. 21, no. 9, pp. 4190–4203, Sept 2012.

[15] P. Thevenaz and M. Unser, "Optimization of mutual information for multiresolution image registration," *IEEE Transactions on Image Processing*, vol. 9, no. 12, pp. 2083–2099, Dec 2000.

[16] M. Peris, S. Martull, A. Maki, Y. Ohkawa, and K. Fukui, "Towards a simulation driven stereo vision system," in *Proceedings of Int. Conf. on Pattern Recognition (ICPR)*, Nov 2012, pp. 1038–1042.

[17] M. A. V. Josien P.W. Pluim, J. B. Antoine Maintz, "Mutual information matching and interpolation artifacts," *Proc.SPIE*, vol. 3661, pp. 3661 – 3661 – 10, 1999. [Online]. Available: https://doi.org/10.1117/12.348605

[18] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "Orb: An efficient alternative to sift or surf," in *IEEE Int. Conf. on Computer Vision (ICCV)*, 2011, pp. 2564–2571.

[19] R. Mur-Artal, J. M. M. Montiel, and J. D. Tards, "Orb-slam: A versatile and accurate monocular slam system," *IEEE Trans. on Robotics*, vol. 31, no. 5, pp. 1147–1163, Oct 2015.