MULTI-ILLUMINATION FUSION WITH CRACK ENHANCEMENT USING CYCLE-CONSISTENT LOSSES

Milind G. Padalkar⋆   *Carlos Beltrán-González*⋆   *Alessio Del Bue*†

⋆ Pattern Analysis and Computer Vision (PAVIS), Istituto Italiano di Tecnologia, Genova, Italy
†Visual Geometry and Modelling (VGM), Istituto Italiano di Tecnologia, Genova, Italy

ABSTRACT

This paper addresses for the first time, the problem of multi-illumination fusion with crack enhancement. Our models are trained using cycle-consistent losses to combine crack details from several mutually registered multi-illumination images of ceramic tiles, into a single representative image. Using real-world industrial data, we show that the crack locations are enhanced in the fused images, making them easily noticeable for remote inspection, and demonstrate the effectiveness of our method compared to a multi-exposure fusion technique.

Index Terms— Multi-illumination fusion, crack detection, visual inspection, cycle-consistent loss

1. INTRODUCTION

Inspection and maintenance activities are periodically carried out for several industrial components to ensure both quality and safety. With the advancement in technology, there is an increasing demand for automating the inspection process. In this direction, there has been a steady progress in the use of image processing and computer vision based methods for automatic visual inspection. For example, use of photometric stereo to identify defects in casted steel [1], convolutional neural networks for inspection of metallic components in nuclear power plants [2], examining flaws in concrete structures [3], inspection with multiple lights [4, 5], have been used to automate visual inspection of cracks from the digital images and videos of the areas to be inspected. Similarly, defect detection in aeronautic components is discussed in [6, 7, 8].

In spite of this progress, an expert's opinion remains necessary whenever the cost of failure to detect the defects is high or even catastrophic. Such a visual inspection process requires the expert to carefully perform large number of assessments based on prior training and experience. Thus, at the moment, the role of a human expert is irreplaceable in many industrial inspection scenarios. However, experts may not always be available in person to perform inspections, especially in situations like the recent pandemic that restrict movement of people. So inspection has an increasing trend on being executed remotely. Further, to perform a more reliable assessment, the experts may rely on several images which may be acquired from different viewpoints and/or lighting conditions. With many objects to be inspected, the process can become exhausting, especially when the number of images is large and defects may be visible in only few of them.

In this paper, we address one such problem involving large number of images for every object to be inspected. Our proposed method fuses several mutually registered images of ceramic tiles, into a single representative image, highlighting cracks to help the visual inspection process. The images are acquired with different illumination conditions using a customized illumination setup, to improve the visibility of cracks for remote inspection. Our fusion method is based on training image generators with cycle-consistent losses motivated by [9], that allow transformation from one domain (acquired images) to another (fused) and back. Cracks are enhanced by constraining the image generators with loss networks that produce binary crack representations of the inputs. An example of fusion with crack enhancement using the proposed method is shown in Fig. 1.

Our main contribution is a method to combine and enhance crack details into a single representative image from an image sequence acquired using different illuminations. Although there exist techniques for image fusion, no previous attempts have been made for simultaneous crack enhancement using multi-illumination image sequence, to the best of our knowledge.

2. RELATED WORK

The problem we address is that of generating one representative image by fusing crack details present in several images acquired with different illuminations. Multi-Exposure Fusion (MEF) is a similar problem where the objective is to fuse a sequence of images acquired with different exposure times to get a single well exposed image [10, 11, 12, 13, 14, 15]. Nevertheless, cracks can have better visibility in a multi-illumination sequence than in a multi-exposure sequence. This is because varying illumination directions can easily create noticeable shadows on cracks, but changing the exposure may not have the same effect. Like in our problem, MEF also assumes the input images to be mutually registered. However, the change across pixels in the different images is

(a) Image sequence acquired with different illuminations   (b) Initial estimate   (c) Fusion with crack enhancement
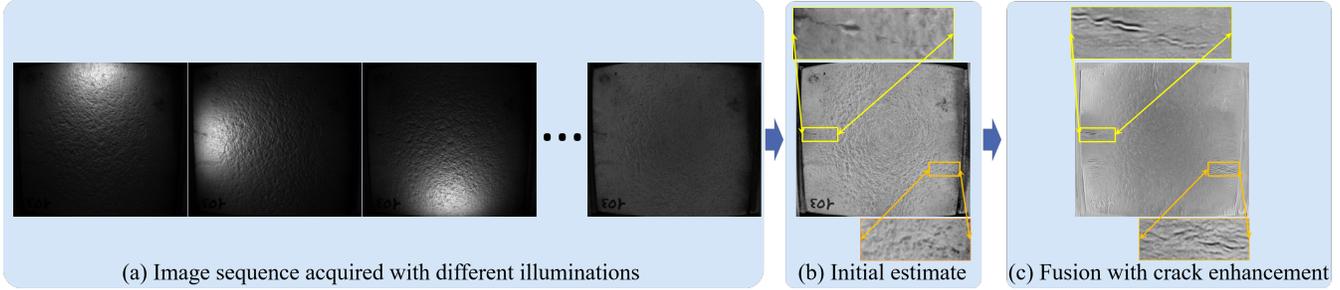
Fig. 1. A sequence of images is acquired with different illuminations since cracks can have better visibility only under certain illumination directions depending on their location and orientation. Manual inspection of details in several images is exhausting and one can easily miss seeing a crack that is visible in one or more of those images. Our method generates one representative image (c) having better visibility of cracks, by fusing and highlighting the crack details present in the image sequence (a).

consistent with MEF, while in our case it is not. Also, due to exposure bracketing in MEF, the pixels across different images are well distributed over the exposure range. In our case, the pixels are well exposed only in few images while being underexposed in most of them. These differences are mainly because of varying illumination directions in our problem.

Nevertheless, since the goal of MEF is similar to ours, we discuss the generic MEF methods [10, 11, 12]. In general, for patches (or pixels) x_n, in a sequence of N exposure bracketed images, the fused patch (or pixel) y is obtained as: $y = \sum_{n=1}^{N} w_n x_n$, where different MEF algorithms use different methods to calculate the weight w_n.

The method in [10] calculates the weight w_n based on contrast, saturation and well-exposedness of each pixel across the sequence. This method is straightforward and can work with an arbitrary number of input images. However, the resulting image appears dark with a large input sequence if pixels are not well exposed in most of them. This can be improved by using morphological operations for preprocessing and additional criteria for weight calculation. In our case, we use such a modified version of exposure fusion to generate an initial estimate of the fused patches required for training.

The first deep neural network architecture used to model the transformation from x_n to y for MEF was proposed in [11]. It optimizes an encoder-decoder model to match the structural similarity of the fused image patches with that of the input image patches, using the structural similarity index measure (SSIM) inspired from [16]. Here, fusion is performed by adding the features extracted from two exposure bracketed images using a shared encoder. The fused features are then decoded by the decoder. This method is interesting and can work with an arbitrary number of images. However, the shared encoder generates high dimensional features for each input, which becomes computationally expensive for backpropagation with large number of inputs. Moreover, SSIM does not address brightness enhancement when most of pixels in the sequence are underexposed.

The method proposed in [12] uses an iterative algorithm to optimize the SSIM between the fused and the input images.

For such an optimization, the number of parameters is equal to the number of pixels in the fused image. It, therefore, becomes computationally expensive with the increase in image size, to a point where it is impossible to use for large images.

All these methods perform fusion, which is also the goal of our work. In addition, we also perform a task specific enhancement in the same fusion framework, which has not been addressed previously to the best of our knowledge. Our proposed method uses fully convolutional architectures and patch based training similar to [11], but with a dedicated encoder that can work simultaneously with all patches in the sequence, which is computationally less expensive in comparison to a shared encoder. We address the issue of brightness enhancement when most of pixels in the sequence are underexposed, by using examples of initially fused patches that are well exposed. Additionally, we use cycle-consistent losses for both visual as well as corresponding crack representations, which helps in highlighting the crack details in the fused image.

3. PROPOSED APPROACH

The pipeline of our proposed approach is shown in Fig. 2. First, we acquire mutually registered tile images in several different illumination conditions using a customized setup. We then obtain the binary crack annotations with the help of experts who carefully inspect the tile as well as the acquired image sequence. From the acquired sequence, we generate an initial estimate of the fused image with a modified version of exposure fusion [10] using morphological preprocessing and additional weight-criteria to improve the contrast and brightness. Co-located patches extracted from the image sequence, initial fusion estimate and binary crack annotations are then used for training our models.

3.1. Models

Considering the sequence of N images as domain A, fused image as domain B and crack annotations as domain C, we train 4 models, viz., G_{A2B} (N to 1 channel image generator),
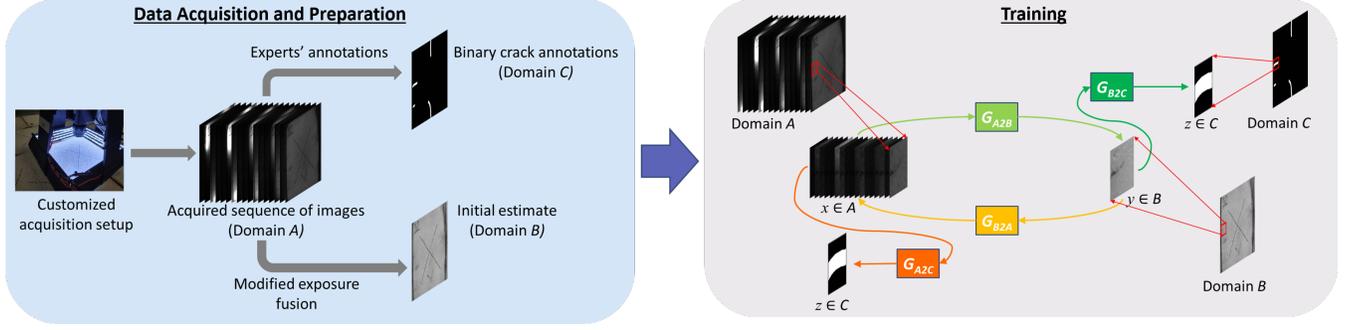
Fig. 2. Proposed approach pipeline: An acquired sequence of images is annotated by experts to generate ground truth binary crack annotations. An initial estimate for fusion is obtained from the acquired sequence using a modified version of exposure fusion [10] to improve the contrast and brightness. Co-located patches from these images are then used for training image generators G_{A2B}, G_{B2A} in a cycle consistent manner along with the crack generators G_{A2C}, G_{B2C}.

G_{B2A} (1 to N channel image generator), G_{B2C} (1 to 1 channel crack generator) and G_{A2C} (N to 1 channel crack generator). This framework is inspired from [9], except that we use crack generators in place of discriminators and train with paired patches. The architecture of all our models is based on U-Net [17] having skip connections. The image generators try to reconstruct the visual information while the crack generators try to extract the crack details. Using the crack generators as loss networks for the image generators helps in crack enhancement. The cycle-consistent losses further assist the image generators in preserving both visual information as well as crack details. This is done by enforcing crack enhancement not only for transformation across the two domains A and B, but also across cyclic transformation to the same domains.

3.2. Losses

For training the crack generators G_{A2C} and G_{B2C}, we use the balanced binary crossentropy (BBCE) loss, which has been effective for generating binary edge-like representations [18]. Here, the balancing is done using the proportion of pixels annotated as cracks in the ground truth patch. Considering patches $x \in A$, $y \in B$ and $z \in C$, the losses for training G_{A2C} and G_{B2C} are given by:

$$loss_{A2C}(x, z) = BBCE(x, z), \tag{1}$$

$$loss_{B2C}(y, z) = BBCE(y, z). \tag{2}$$

For the image generators, we use a combination of losses, viz., mean absolute error (MAE) or L1 loss and BBCE in a cycle-consistent fashion as follows:

$$\begin{aligned}loss_{A2B}(x, y, z) =& BBCE(G_{B2C}(G_{A2B}(x)), z) + \\ & BBCE(G_{A2C}(G_{B2A}(G_{A2B}(x))), z) + \\ & BBCE(G_{B2C}(G_{A2B}(G_{B2A}(y))), z) + \\ & MAE(G_{B2A}(G_{A2B}(x)), x) + \\ & MAE(G_{A2B}(G_{B2A}(y)), y), \end{aligned}$$
$$\tag{3}$$

$$\begin{aligned}loss_{B2A}(x, y, z) =& BBCE(G_{A2C}(G_{B2A}(y)), z) + \\ & BBCE(G_{B2C}(G_{A2B}(G_{B2A}(y))), z) + \\ & BBCE(G_{A2C}(G_{B2A}(G_{A2B}(x))), z) + \\ & MAE(G_{A2B}(G_{B2A}(y)), y) + \\ & MAE(G_{B2A}(G_{A2B}(x)), x). \end{aligned}$$
$$\tag{4}$$

3.3. Training

For each batch of patches $x \in A$, $y \in B$ and $z \in C$, training is performed in the following order:

1. Train G_{B2A} : $\underset{G_{B2A}}{\arg\min} \, loss_{B2A}(x, y, z)$.

2. Train G_{A2C} : $\underset{G_{A2C}}{\arg\min} \, loss_{A2C}(x, z)$.

3. Train G_{A2B} : $\underset{G_{A2B}}{\arg\min} \, loss_{A2B}(x, y, z)$.

4. Train G_{B2C} : $\underset{G_{B2C}}{\arg\min} \, loss_{B2C}(y, z)$.

While the crack generators G_{A2C} and G_{B2C} can be trained independently, training in the above order helps the image generators G_{A2B} and G_{B2A} gradually adapt to crack enhancement. This provides more stability to the models.

4. RESULTS

Our experiments are conducted on a real-world industrial data of 88 ceramic tiles, which consists of image sequences acquired using our customized multi-illumination setup [5]. Each sequence has $N = 65$ different illumination images of size 1944×2592. Patches of size 128×128 from 79 tiles are used for training, while 9 tiles are used for testing. Models are trained from scratch for 2 epochs on a single NVIDIA RTX-2080 GPU, with batches size of 8. Adam optimizer is used with learning rates 0.0001 and 0.00001 for training the image generators and crack generators, respectively.
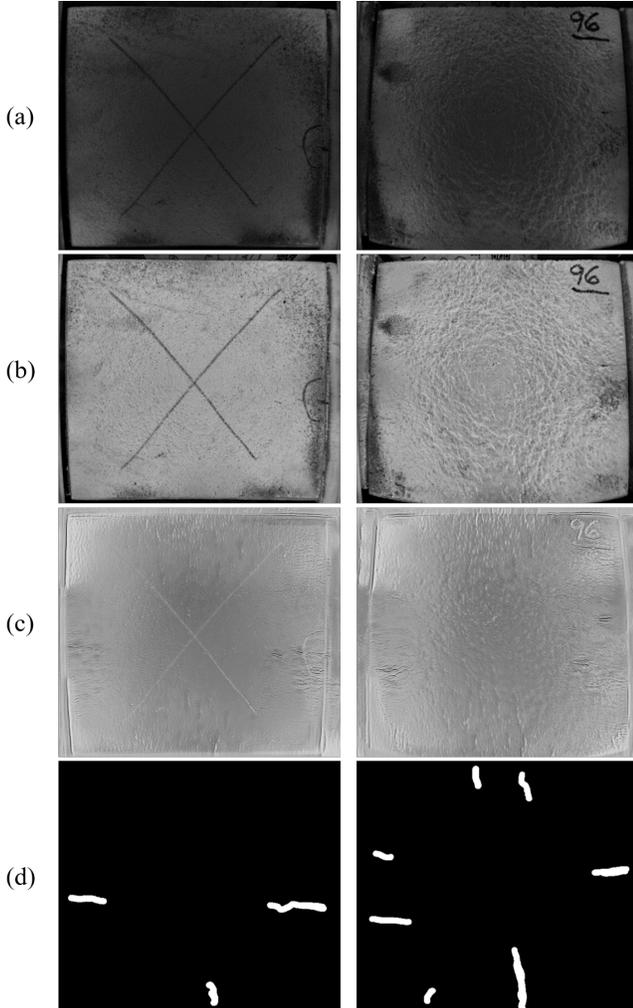
Fig. 3. Results: (a) Fusion with a MEF method [10], (b) initial estimate, (c) fusion with crack enhancement with our method (G_{A2B}) and (d) crack annotations provided by experts. Notice the high contrast in (c) for regions corresponding to the crack annotations in (d). (Zoom-in for better visualization)

Fusion results for two test tiles are shown in Fig. 3. The effect of MEF techniques for multi-illumination images is shown using exposure fusion [10] in Fig. 3(a). Our initial estimate and output of G_{A2B} are shown in Fig. 3(b)-(c), while the ground truth binary crack annotations provided by experts are shown in Fig. 3(d). We evaluate the performance using a measure that captures saliency in terms of edge strength. For an image I with salient regions $\Omega \in I$ having stronger edges, the edge strength in Ω should be higher than the global edge strength to make Ω easily noticeable. Using Laplacian of Gaussian (LoG), the edge strength ES is calculated as:

$$ES = \frac{mean(|L_p|)}{mean(|L_q|)}, \quad p \in \Omega, \quad q \in I, \quad L = LoG(I), \quad (5)$$

where L_p is value of L at pixel $p \in \Omega$ and L_q is value of L

Table 1. Performance comparison for test images with edge strength measure defined in Eq. (5) (The higher the better).

Image#	MEF [10]	Initial estimate	Proposed (G_{A2B})
1	1.2369	1.2354	**2.6695**
2	1.0719	1.1380	**3.1600**
3	1.0360	1.1272	**1.9077**
4	1.0825	1.1279	**1.7663**
5	1.0844	1.1723	**2.4617**
6	1.1637	1.0021	**2.3807**
7	0.9425	0.9081	**1.9220**
8	1.0956	0.9017	**2.4140**
9	1.0581	1.2385	**2.6135**

at pixel $q \in I$. Table 1 presents the edge strengths for test images considering the crack annotations (white regions) in Fig. 3(d) as Ω.

The MEF method [10] generates darker results (Fig. 3(a)) with little crack information. This is because pixels are underexposed in most of the images of the input sequence, an issue which is not addressed by the MEF methods. Initial estimates in Fig. 3(b) have improved contrast and brightness. However, the presence of dust particles on tile's surface (e.g., right side of the tile in first column) reduces visibility of the cracks. Our results (Fig. 3(c)) have distinct visibility of cracks even in the presence of dust particles. Moreover, tile details are better visible in our results and the cracks have higher contrast, making them easily noticeable. This is also confirmed by the comparison presented in Table 1 that indicates better performance of our method for all the test images.

5. CONCLUSIONS

In this paper we have proposed a method to combine and enhance crack details into a single representative image from several images acquired using different illuminations. The generated fused image helps in the visual inspection process, without which one can easily miss seeing a crack that is visible in one or more of acquired images. The visual information is fused using our image generators trained with cycle-consistent losses. Use of crack generators as loss networks for the image generators helps to enhance the crack details. We address the enhancement of pixels that are underexposed in most of the images of the acquired sequence, which is not addressed by MEF. Results show that our proposed method is better suited than using a MEF approach for fusion of multi-illumination images.

6. ACKNOWLEDGEMENT

7. REFERENCES

[1] A. Landstrom, M. J. Thurley, and H. Jonsson, "Sub-millimeter crack detection in casted steel using color photometric stereo," in *2013 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, Nov 2013, pp. 1–7. 1

[2] F. Chen and M. R. Jahanshahi, "Nb-cnn: Deep learning-based crack detection using convolutional neural network and naïve bayes data fusion," *IEEE Trans. on Industrial Electronics*, vol. 65, no. 5, pp. 4392–4400, May 2018. 1

[3] L. Yang, B. Li, G. Yang, Y. Chang, Z. Liu, B. Jiang, and J. Xiaol, "Deep neural network based visual inspection with 3d metric measurement of concrete defects using wall-climbing robot," in *2019 IEEE IROS*, Nov 2019, pp. 2849–2854. 1

[4] M. Aghaei, M. Bustreo, P. Morerio, N. Carissimi, A. Del Bue, and V. Murino, "Complex-object visual inspection: Empirical studies on a multiple lighting solution," in *25th International Conference on Pattern Recogntion (ICPR)*, 2020. 1

[5] M. G. Padalkar, C. Beltrán-González, M. Bustreo, A. Del Bue, and V. Murino, "A versatile crack inspection portable system based on classifier ensemble and controlled illumination," in *25th International Conference on Pattern Recogntion (ICPR)*, 2020. 1, 3

[6] S. Martelli, L. Mazzei, C. Canali, P. Guardiani, S. Giunta, A. Ghiazza, I. Mondino, F. Cannella, V. Murino, and A. Del Bue, "Deep endoscope: Intelligent duct inspection for the avionic industry," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 4, pp. 1701–1711, 2018. 1

[7] M. S. Biagio, C. Beltrán-González, S. Giunta, A. Del Bue, and V. Murino, "Automatic inspection of aeronautic components," *Machine Vision and Applications*, vol. 28, no. 5, pp. 591–605, Aug 2017. 1

[8] C. Beltrán-González, M. Bustreo, and A. Del Bue, "External and internal quality inspection of aerospace components," in *2020 IEEE 7th International Workshop on Metrology for AeroSpace (MetroAeroSpace)*, 2020, pp. 351–355. 1

[9] J. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Computer Vision (ICCV), 2017 IEEE International Conference on*, 2017. 1, 3

[10] T. Mertens, J. Kautz, and F. Van Reeth, "Exposure fusion," in *15th Pacific Conference on Computer Graphics and Applications (PG'07)*, 2007, pp. 382–390. 1, 2, 3, 4

[11] K. R. Prabhakar, V. S. Srikar, and R. V. Babu, "Deepfuse: A deep unsupervised approach for exposure fusion with extreme exposure image pairs," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 4724–4732. 1, 2

[12] K. Ma, Z. Duanmu, H. Yeganeh, and Z. Wang, "Multi-exposure image fusion by optimizing a structural similarity index," *IEEE Transactions on Computational Imaging*, vol. 4, no. 1, pp. 60–72, 2018. 1, 2

[13] F. Kou, Z. Li, C. Wen, and W. Chen, "Multi-scale exposure fusion via gradient domain guided image filtering," in *2017 IEEE International Conference on Multimedia and Expo (ICME)*, 2017, pp. 1105–1110. 1

[14] Jianrui Cai, Shuhang Gu, and Lei Zhang, "Learning a deep single image contrast enhancer from multi-exposure images," *IEEE Transactions on Image Processing*, vol. 27, no. 4, pp. 2049–2062, 2018. 1

[15] Q. Wang, W. Chen, X. Wu, and Z. Li, "Detail-enhanced multi-scale exposure fusion in yuv color space," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 8, pp. 2418–2429, 2020. 1

[16] K. Ma, K. Zeng, and Z. Wang, "Perceptual quality assessment for multi-exposure image fusion," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3345–3356, 2015. 2

[17] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, 2015, pp. 234–241. 3

[18] S. Xie and Z. Tu, "Holistically-nested edge detection," in *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 1395–1403. 3