WSEAS TRANSACTIONS on COMPUTERS
DOI: 10.37394/23205.2025.24.30

Fast and Robust Intensity-Based Image Matching Using FFT

LEILA ESSANNOUNI[1], MANAL TAOUFIKI[2], FEDWA ESSANNOUNI[3]
[1] LASTIMI Laboratory, Higher School of Technology of Salé,
Mohammed V University, Rabat,
MOROCCO

[2] Maisonneuve College, Montreal,
CANADA

[3] LRIT Laboratory, Faculty of Sciences,
Mohammed V University, Rabat,
MOROCCO

Abstract: - This paper presents a fast and robust image matching approach based on the cosine M-estimator kernel and the Fast Fourier Transform (FFT). We show that the robust cosine M-estimator can be effectively used to compare image intensities through correlation of transformed images. The speed of the method derives from the use of FFT to compute correlation. Its computational complexity is $O(N \log N)$, compared to $O(N^2)$ for direct matching. Experimental results demonstrate that the proposed method maintains high matching accuracy even in the presence of up to 60% outliers and 60% occlusion. It outperforms traditional approaches such as Sum of Squared Differences (SSD) and Normalized Cross-Correlation (NCC). Moreover, the proposed approach achieves superior performance compared to recent deep learning-based methods such as LoFTR. By combining the cosine M-estimator with FFT, the method is suitable for real-time applications that require robust matching under challenging conditions.

Key-words: Image Matching, Robust Correlation, Cosine M-estimator, Robust Statistics, FFT, Fast Algorithms.

Received: June 13, 2025. Revised: July 15, 2025. Accepted: August 17, 2025. Published: December 31, 2025.

1 Introduction

Image matching is a fundamental task in computer vision. It refers to the process of aligning two or more images by detecting and matching corresponding features. This process has numerous applications in fields such as medical imaging, remote sensing, robotics, and visual surveillance [1], [2], [3], [4], [5]. In real-world scenarios, the performance of image matching is often degraded by noise, outliers, and occlusion. These challenges can lead to inaccurate correspondences. Robust statistical approaches and mutual information techniques [6], [7], [8], [9] have gained significant attention due to their improved performance under varying conditions. However, these methods are often time-consuming.

Recent advances in deep learning have achieved significant improvements in image matching performance. Methods such as LoFTR [3] and SuperGlue [6] have demonstrated impressive results on challenging datasets. Nevertheless, these approaches typically require large amounts of training data and significant computational resources, and they may not generalize well to specific domains. In contrast, traditional approaches based on robust correlation can be implemented efficiently without the need for training data.

In this paper, we propose a robust approach for image intensity matching that combines the statistical robustness of the cosine M-estimator with the computational efficiency of the Fast Fourier Transform (FFT). We show that the proposed method effectively applies the cosine M-estimator [10] to image intensity matching. By the use of the FFT to compute correlation, the proposed approach has a computational complexity of $O(N \log N)$, compared to $O(N^2)$ for direct spatial-domain matching. Thus, it is suitable for real-time applications while being robust to outliers and occlusion.

The remainder of the paper is organized as follows. Section 2 describes the proposed method. Section 3 demonstrates its robustness. Section 4 discusses the computational cost. Section 5 presents experimental results. Finally, conclusions are drawn Section 6.

2 Fast and robust image intensity matching using the Cosine M-estimator

Robust statistics are defined by small variations to the data, or large variations to a small number of data (outliers) [9], [8], [10]. In this paper, we propose to use the Cosine M-estimator kernel as similarity measure. M-estimators have been successfully applied to various robust image matching tasks [10], and we extend this concept by implementing it efficiently in the frequency domain using FFT.

Let f be the reference image with dimensions $N = N1 \times N2$, and g be the second image with dimensions $n = n1 \times n2$. The images are grayscale and normalized to the range [0,1]. The images are compared using the cosine similarity function. For a location *(x,y)* in *f*, we calculate the similarity *E(x,y)* between *f* and the template *g* as follows:

$$E(x,y) = (\mathrm{n1xn2}) - \sum_{i=1}^{n1} \sum_{j=1}^{n2} \cos\big(f(x+i, y+j) - g(i,j)\big) \quad (1)$$

The cosine function provides a robust measure of similarity that is less sensitive to outliers compared to traditional squared difference measures.

The influence of the error measure on the matching can be understood using the influence function. The influence function of the standard SSD error kernel is:

$$\varphi(e) = e \quad (2)$$

For the proposed error measure, the influence function is:

$$\varphi(e) = \sin(e) \quad -\pi \le e \le \pi \quad (3)$$

Figure 1 plots the influence function of the SSD kernel and the cosine kernel. As seen in the figure the cosine influence function is less increasing than the SSD influence function in the interval [0 1].

The matching using the cosine function can be computed using correlation. From the images *I1* and *I2*, we reconstruct the images *f* and *g* as follows:

$$\mathrm{f(x,y)} = e^{\mathrm{i(I1(x,y))}} \quad (4)$$

Here, *i* is the complex imaginary unit. The image *g* is constructed in the same way as *f*.

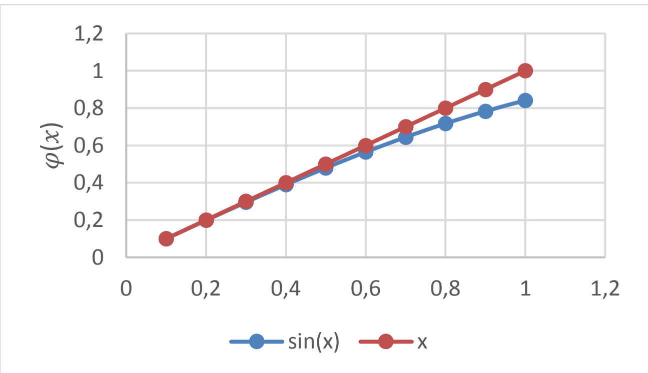


Fig. 1: Influence function of the cosine and the SSD kernels.
Source: created by the authors

The transformed images *I1* and *I2* are matched using correlation. The correlation is calculated rapidly using the FFT. Given *F(j,k)*, the FFT of *f(x,y)*, and *G(j,k)*, the FFT of *g(x,y)*, and IFFT() the Inverse Fast Fourier Transform function, the correlation matching surface is :

$$\Re\mathrm{eal}\{IFFT(F(j,k)G^*(j,k))\} (5)$$

The best match is given by the position of the maximum in 5, we assume that *I1* and *I2* have the same size. Images with different sizes are matched using zero padding.

3 Robustness

In this section, we demonstrate the theoretical robustness of the proposed correlation method. To show how the proposed correlation employs the Cosine M-estimator kernel, we analyze the mathematical properties of the transformed images and their correlation in the frequency domain.

We consider two pixels *p1* and *p2* from the transformed image *f*, and *g* respectively, given by the equation (4). Taking the complex conjugate of *p2*, their expressions are:

$$p1 = e^{i\theta 1} \qquad p2^* = e^{-i\theta 2}, \text{ Where}$$
$$\theta 1 = \mathrm{I1(x,y)} \ and \ \theta 2 = \mathrm{I2(x,y)} \quad (6)$$

Correlation shifts one image relative to the other and measures the sum of the products of their corresponding pixels.

The product of pixel *p1* and the complex conjugate of pixel *p2* is:

$$p1p2^* = e^{i(\theta 1 - \theta 2)} \quad (7)$$

In Cartesian form, the real part is expressed as:

$$\Re\mathrm{eal}\{p1p2^*\} = \cos(\theta 1 - \theta 2) \quad (8)$$

From this last equation, we see that the proposed technique applies the cosine kernel to the difference of the normalized image intensities.

The real part of the correlation of the two transformed images (Eq. 5) is equal to the sum of the cosine of the normalized image intensities difference.

4 Computational complexity analysis

The efficiency of the proposed method is largely driven by the use of the FFT. Its overall complexity is O(N log N), where N denotes the largest dimension of the two images. This represents a substantial improvement over the O(N^2) cost of direct matching. In practice, this advantage becomes especially significant for large image sizes, making the approach well suited for real-time applications that require both speed and robustness.

In terms of memory usage, the proposed method is highly efficient, as it requires storing only the two input images along with their corresponding frequency-domain representations. As the image size increases, memory consumption grows linearly, enabling the method to be used even with limited computational resources.

5 Experimental results

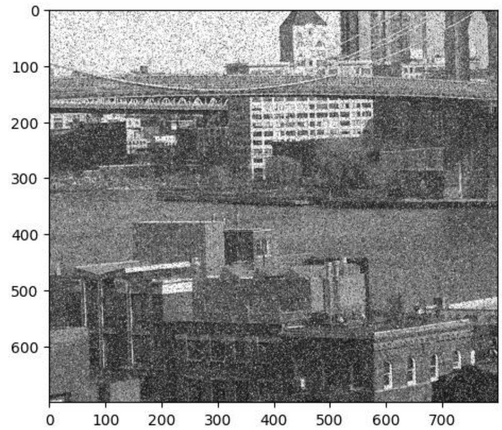
To evaluate the efficacy of the proposed method in the presence of outliers and occlusion, extensive matching experiments were performed on the Google Colab platform. The grayscale images are normalized to the range [0, 1]. Images with different sizes are matched using zero padding. The outliers are simulated using Salt and pepper noise.

Our approach was compared against standard Sum of Squared Differences (SSD), Normalized Cross-Correlation (NCC), and the recent deep learning-based LoFTR method [3]. The matching rate corresponds to the detection percentage of the second image at the correct position.

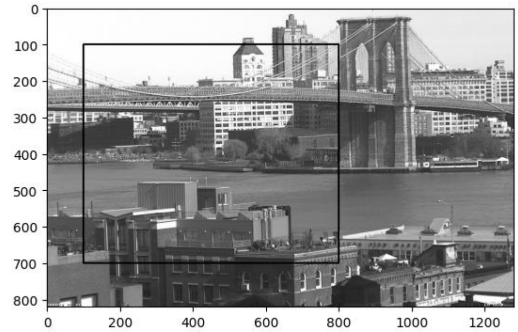
The evaluation utilized a diverse dataset comprising 200 images selected from the HPatches database [11]. For each test, the second image was extracted from the original image, and controlled amounts of

Table 1. Matching rates (%) at various outlier levels
Source: created by the authors

Percentage of outliers	SSD	NCC	LoFTR	**Proposed method**
10%	100%	100%	100%	**100%**
20%	100%	100%	100%	**100%**
30%	95%	100%	30%	**100%**
40%	90%	100%	0%	**100%**
50%	65%	80%	0%	**100%**
60%	40%	60%	0%	**100%**



(a) Second image.



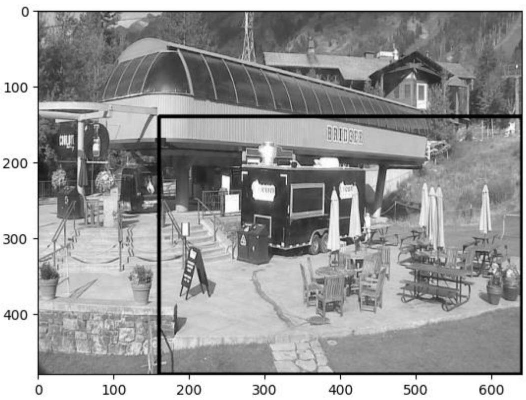
(b) Reference image.

Fig. 2: Example of matching in the case of Salt and pepper noise.
Source: created by the authors

Table 2. Matching rates (%) under different occlusion levels
Source: created by the authors

Percentage of occlusion	SSD	NCC	LoFTR	**Proposed method**
10%	100%	100%	100%	**100%**
20%	100%	100%	100%	**100%**
30%	95%	100%	100%	**100%**
40%	80%	90%	80%	**100%**
50%	70%	80%	70%	**100%**
60%	40%	60%	50%	**98%**

(a) Second image.



(b) Reference image.

Fig. 3: Example of image matching under occlusion.
Source: created by the authors

Salt and pepper noise and occlusion were introduced. Table 1 presents the matching rates for the proposed method, SSD, NCC, and LoFTR method at various levels of Salt and Pepper noise. As illustrated in the table, the proposed approach gives better results compared to the Sum of Squared Differences, Normalized Cross-Correlation, and LoFTR method. LoFTR is based on keypoint descriptors, which are sensitive to pixel-level corruption such as salt and pepper noise, whereas our method operates directly on image intensities, making it more robust in such cases. The proposed method successfully matches the images even in the presence of up to 60% noise. Figure 2 shows an example of the matching results in the presence of Salt and pepper noise.

In the second experiment, we simulated occlusion by overlaying a white square (10–60 % area) on the second image. As shown in Table 2, the proposed approach gives the best performance. The proposed method finds the best match with up to 60% occlusion. Figure 3 shows an example of matching in the presence of occlusion.

6 Conclusion and future work

In this paper, we introduced a novel robust image intensity matching method using the Cosine M-estimator and FFT-based correlation. The proposed method demonstrates superior performance compared to traditional approaches such as SSD, NCC, and the deep learning method LoFTR. It achieves high matching accuracy even in the presence of significant outliers and occlusion (up to 60%).

Future work will focus on extending the method to handle more complex geometric transformations, such as rotation and scaling.

References

[1] Verykokou, S., and Ioannidis, C., "Image Matching: A Comprehensive Overview of Conventional and Learning-Based Methods", Encyclopedia, Vol. 5, No. 1, Article 4, 2025. https://doi.org/10.3390/encyclopedia5010004

[2] Cui, Z., Qi, W., and Liu, Y., "A Fast Image Template Matching Algorithm Based on Normalized Cross Correlation", Journal of Physics: Conference Series, Vol. 1693, Article 012163, Bristol, UK, 2020. https://doi.org/10.1088/1742-6596/1693/1/012163

[3] Sun, J., Shen, Z., Wang, Y., Bao, H., and Zhou, X., "LoFTR: Detector-Free Local Feature Matching with Transformers ", in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2021), IEEE, 2021, pp. 8918–8927. https://doi.org/10.1109/CVPR46437.2021.00881

[4] Luo, C., Yang, W., Huang, P., and Zhou, J., "Overview of Image Matching Based on ORB Algorithm", Journal of Physics: Conference Series, Vol. 1237, Article 032020, Bristol, UK, 2019. https://doi.org/10.1088/1742-6596/1237/3/032020

[5] Kaso, A., "Computation of the Normalized Cross-Correlation by Fast Fourier Transform", PLoS ONE, Vol. 13, No. 8, e0203434, 2018. https://doi.org/10.1371/journal.pone.0203434

[6] Sarlin, P.-E., DeTone, D., Malisiewicz, T., and Rabinovich, A., "SuperGlue: Learning Feature Matching with Graph Neural

Networks", in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition CVPR, Seattle, WA, USA, pp. 4938–4947, 2020. https://doi.org/10.1109/CVPR42600.2020.00499

[7] Ayubi, G. A., Kowalski, B., and Dubra, A., "Normalized weighted cross correlation for multi-channel image registration", Optics Continuum, Vol. 3, No. 5, pp. 649–665, 2024. https://doi.org/10.1364/OPTCON.525065

[8] Öfverstedt, J., Lindblad, J., and Sladoje, N., "Fast computation of mutual information in the frequency domain with applications to global multimodal image alignment", Pattern Recognition Letters, Vol. 159,pp. 1–8, 2022. https://doi.org/10.1016/j.patrec.2022.05.022

[9] Tombari, F., Di Stefano, L., Mattoccia, S., and Galanti, A., "Performance evaluation of robust matching measures", in Proceedings of the Third International Conference on Computer Vision Theory and Applications VISAPP, Funchal, Madeira, Portugal, Vol. 1, pp. 473–478, 2008. https://doi.org/10.5220/0001087304730478

[10] Fitch, A., Kadyrov, A., Christmas, W., and Kittler, J., "Fast Exhaustive Robust Matching", in Proceedings of the 16th International Conference on Pattern Recognition (ICPR 2002), IEEE, 2002, pp. 903–906. https://doi.org/10.1109/ICPR.2002.1048178

[11] Balntas, V., Lenc, K., Vedaldi, A., and Mikolajczyk, K., "HPatches: A Benchmark and Evaluation of Handcrafted and Learned Local Descriptors", in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition CVPR, Honolulu, HI, USA, 2017.

Contribution of individual authors to the creation of a scientific article (ghostwriting policy)

All authors contributed equally to this work, from the formulation of the problem to the analysis of the experimental results.

Sources of Funding for the Research Presented in the Scientific Article or for the Scientific Article Itself

No funding was received for conducting this study.

Conflict of Interest

The authors declare that they have no conflicts of interest relevant to the content of this article.