

# Assessing Local Descriptors for Feature-Based Registration of Whole-Slide Images

Arthur Elskens

*Laboratory of Image Synthesis and Analysis (LISA)*

*Université Libre de Bruxelles*

Brussels, Belgium

Arthur.Elskens@ulb.be

Adrien Foucart

*LISA*

*Université Libre de Bruxelles*

Egor Zindy

*DIAPath, Center for Microscopy and Molecular Imaging (CMMI)*

*Université Libre de Bruxelles*

Olivier Debeir

*LISA & CMMI*

*Université Libre de Bruxelles Université Libre de Bruxelles*

Christine Decaestecker

*LISA & CMMI*

**Abstract**—Feature-based registration has become increasingly popular in digital pathology for achieving initial global alignment between image pairs. However, the selection of algorithms used in this approach is often not well-justified. Specifically, the choice of local feature descriptor is rarely, if ever, discussed in the context of digital pathology. The majority of feature-based whole-slide image registration methods rely on the SIFT descriptor. In this study, we demonstrate that the choice of descriptor significantly influences the quality of registration results and that the BRIEF descriptor captures more optimal information for histological image registration.

**Index Terms**—Digital pathology, Whole-slide image, Feature-based registration

## I. INTRODUCTION

Whole-Slide Images (WSI) registration is a widely recognized and challenging computer vision task that lacks a straightforward out-of-the-box solution, especially when it comes to multi-stained WSI registration. In recent years, this particular task has been the focus of several challenges, namely ANHIR [1] and ACROBAT [2]. These challenges aimed to showcase the current state-of-the-art algorithms and their robustness when applied to real-world datasets.

Upon analyzing the submitted registration pipelines from each challenge, a consistent pattern emerged: achieving relevant results typically requires an initial global registration at low resolution before attempting non-rigid alignment of the image pair. Various approaches have been proposed to estimate the initial alignment parameters. These include: a comprehensive search of the rotation space after centroid alignment, template matching, and feature-based algorithm. The latter approach is the most commonly used, with six out

eight teams competing in the ACROBAT challenge employing it [2].

Feature matching approaches generally involve two main steps: (i) local feature extraction, and (ii) robust matching. More specifically, local feature extraction comprises detection and description of points of interest (PoI). This description part plays a crucial role as it significantly influences the robust matching process. Several local descriptors such as SIFT [3] and BRISK [4] exist. Studies comparing the impact of descriptor choices on the matching results have been conducted on LIDAR and building image datasets [5], [6]. However, to the best of our knowledge, no comparative analysis has yet been conducted on histological WSI.

This paper aims to assess the performance and influence of various local feature descriptors on the initial alignment of histological WSI. Our findings demonstrate that the selection of a descriptor significantly affects the feature-based registration process, highlighting its crucial role within the feature matching pipeline. To facilitate result replication, our code is readily available online <sup>1</sup>.

## II. RELATED WORKS

Feature matching is a well-explored topic in the field of computer vision, finding applications in various research areas such as pose estimation and 3D reconstruction. However, when it comes to digital pathology, the literature on feature matching is relatively sparse. The research conducted in this field has been limited, and this section aims to provide an overview of the algorithms that have been employed thus far.

### A. Pre-processing

Current feature matching algorithms have been developed primarily for single-channel images. To adapt these algorithms for multi-channel images, a common approach is to convert the images to grayscale as a pre-processing step. This technique

This research was supported by the Walloon Region (Belgium) in the framework of the PIT program “PROTHER-WAL” under grant agreement No. 7289. CD is a senior research associate with the National (Belgian) Fund for Scientific Research (F.R.S.-FNRS) and is an active member of the TRAIL Institute (Trusted AI Labs, <https://trail.ac/>, Fédération Wallonie-Bruxelles, Belgium). CMMI is supported by the European Regional Development Fund and the Walloon Region (Walloniabiomed, #411132-957270, project “CMMI-ULB”).

<sup>1</sup>[https://gitlab.com/prother-wal\\_ulb\\_lis\\_mnu/feature-based\\_registration/-/tree/sipaim23](https://gitlab.com/prother-wal_ulb_lis_mnu/feature-based_registration/-/tree/sipaim23)

has been widely employed and described in both the ANHIR and ACROBAT challenges [1], [2].

Furthermore, as a part of the pre-processing for the initial alignment step, it is common to downscale the WSI to a lower resolution image. This approach is dictated by the fact the object of interest is biological tissue with mechanical properties hence a certain local regularity/continuity but also serves two purposes: firstly, it helps speed up the registration process by reducing computational complexity, and secondly, it mitigates the risk of false matching. While the overall shape of the images may appear similar, employing higher-resolution images could potentially introduce erroneous correspondences at a finer level due to subtle differences in appearance.

### B. Feature extraction

Since the introduction of SIFT in [3], local features have become widely adopted tools in computer vision. Over the years, various traditional methods have been developed in addition to SIFT, including SURF [7], ORB [8], FREAK [9], BRIEF [10] and BRISK [4]. In the field of digital pathology, SIFT remains the most commonly used local feature detector and descriptor [11]–[13]. However, in some cases, alternative methods such as SURF and ORB are employed alongside SIFT to ensure robustness in the event of failure [12]. In addition, BRISK was chosen as the PoI detector in the approach ranked second in the ACROBAT challenge [14].

With the emergence of deep learning, there is a surge in the development of learning-based methods for local feature extraction. For example, in [14], BRISK-detected PoIs are described using a VGG network [15]. In another study reported in [16], SuperPoint [17], a Convolutional Neural Network (CNN), was employed for WSI registration and reached third place in the ANHIR challenge leaderboard (at the time of writing).

### C. Robust matching

The second main step in the feature matching pipeline is a robust estimation process.

This step can be further divided into: (i) matching, which establishes correspondences between the detected PoIs in the moving image and their counterparts in the fixed image, and (ii) robust estimation of the transformation between these sets of correspondences. RANSAC [18] is an extensively utilized robust algorithm in various disciplines, including digital pathology [13], [19], for identifying reliable correspondences among matched pairs.

## III. DATASET

The performance evaluation of the descriptors is conducted on a dataset merging three multi-stained WSI datasets: the ANHIR dataset and two others, the MSSC dataset and the ISC dataset, both introduced in [11]. These datasets were selected for their accessibility and the inclusion of landmarks that serve as reliable ground truth for evaluation purposes. Moreover, these data contain a wide spectrum of tissue types, with varied morphological characteristics, as well as regions

varying from homogeneous and smooth to highly textured. By incorporating such diversity, the evaluation accounts for the descriptors’ robustness to tissue and textural variations.

## IV. METHODS

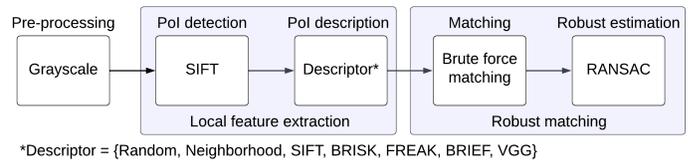


Fig. 1. Illustration of the feature-based matching pipeline, explicitly highlighting the algorithmic choices.

In this study, the algorithms employed in each step of the feature matching pipeline (cf. Figure 1) were fixed, with the exception of the description part. As recommended in the literature, the pre-processing step involves downscaling the original full-scale input images to x1 equivalent size ( $\sim 10 \mu m/px$ ) and converting them to grayscale. The OpenCV implementation of SIFT is used as a detection step, as it is widely recognized and commonly used in digital pathology. The matching process employed a brute force matching method, also implemented in OpenCV, with the particularity that only pairs of PoIs that are mutually the closest to each other are retained. This approach is similar to the “both” strategy proposed in [20]. The choice of distance metric depends on the specific descriptor utilized, which is listed in Table I. For robust estimation, RANSAC (OpenCV’s implementation) was selected as the inlier filter, as it is widely adopted in various fields, including digital pathology.

TABLE I  
EVALUATED DESCRIPTORS AND THEIR MATCHING DISTANCE.

Descriptor	Distance <sup>a</sup>
Random	Euclidean distance
Neighborhood	Euclidean distance
SIFT	Euclidean distance
BRISK	Hamming distance
FREAK	Hamming distance
BRIEF	Hamming distance
VGG	Euclidean distance

<sup>a</sup>As proposed in [6].

For the description step, a total of seven descriptors were evaluated, as outlined in Table I. The first two descriptors, namely *Random* and *Neighborhood*, were specifically included to assess the relevance of the proposed evaluation and are expected to perform poorly in comparison to the others. The *Random* descriptor generates a description vector filled with random numbers sampled from a uniform distribution ranging from 0 to 1. On the other hand, the *Neighborhood* descriptor employs a simple neighborhood-based approach by constructing the description vector from flattened intensity neighborhood values. The remaining descriptors are implementations from OpenCV of commonly used traditional methods. Additionally, the pre-trained OpenCV’s VGG, inspired by

the approach proposed in [14], was included in the evaluation. All descriptors, including the learning-based method, were assessed using their respective default parameters.

An affine transformation is estimated in this feature-based registration. This choice of transformation is commonly employed as a preliminary alignment step before proceeding to align the image pair in a non-linear, non-rigid manner.

## V. EVALUATION

The evaluation of the descriptor’s impact on the feature matching pipeline (cf. Figure 1) is conducted in two distinct parts. Firstly, the descriptive capabilities of each descriptor are assessed. This involves analyzing and comparing the ability of each descriptor to accurately represent and describe local image features. Secondly, the performance of the feature-based registration is evaluated for each descriptor. This stage focuses on evaluating how well the descriptors contribute to the overall registration process and their effectiveness in achieving accurate alignments between images.

### A. Descriptive capabilities

The evaluation of the descriptor’s descriptive capabilities relies on the available landmarks within each image. Specifically, for each descriptor, the landmark matching ranks between fixed and moving images are computed and normalized. To achieve this, pairwise distances are calculated in the latent space between each landmark in the moving image and those in the fixed image. The rank (starting at 0 for minimum distance) of the corresponding correct fixed-image landmark is retrieved and normalized by the number of available landmarks for that specific image pair. This normalization accounts for variations in the number of available landmarks among different images pairs, ensuring a fair comparison across the dataset. Normalized rank distributions are then compared between descriptors.

### B. Registration performances

The performance evaluation of the feature-based registration follows a conventional approach [21], utilizing the Root Mean Square Error (RMSE) as a metric. The RMSE is calculated by comparing the positions of the fixed image landmarks with their corresponding warped moving image landmarks. This error measurement quantitatively assesses the overall distortion between the landmarks in the fixed and moving images.

## VI. RESULTS

First, the descriptive capabilities of each descriptor are evaluated by comparing their distributions of normalized matching ranks, as described in Section V. Medians and interquartile values for these distributions are provided in Table II, where a rank of 0 represents the best possible performance, while a rank of 1 indicates the worst.

The results reveal that the *Random* descriptor performs the poorest, exhibiting a median normalized rank of 0.5 and a distribution that resembles a normal curve across the entire range. In comparison, the naive *Neighborhood* descriptor

TABLE II  
MEDIAN AND INTERQUARTILE VALUES FOR BOTH THE NORMALIZED RANKS AND THE RMSE DISTRIBUTIONS PER DESCRIPTOR. THE INITIAL RMSE DISTRIBUTION (BEFORE ANY ALIGNMENT) IS ALSO PROVIDED. FOR THE NORMALIZED RANKS DISTRIBUTION, THE CLOSER A NORMALIZE RANK TO ZERO THE BETTER. THE ROW IN BOLD IS CONSIDERED, STATISTICALLY, AS THE BEST RESULT ACCORDING TO A NEMENYI POST-HOC TEST.

Descriptor	Norma- lized ranks median	Norma- lized ranks IQ	RMSE median [ $\mu m$ ]	RMSE IQ [ $\mu m$ ]
Initial	n/a	n/a	725	[301-1615]
Random	0.5	[0.24-0.75]	5913	[3574-8296]
Neighborhood	0.31	[0.11-0.58]	6110	[3400-8754]
SIFT	0.09	[0.01-0.30]	216	[94-899]
BRISK	0.24	[0.04-0.59]	473	[162-2504]
FREAK	0.16	[0.02-0.49]	254	[115-1510]
<b>BRIEF</b>	<b>0.01</b>	<b>[0.0-0.11]</b>	<b>145</b>	<b>[68-295]</b>
VGG	0.15	[0.04-0.38]	254	[107-1870]

shows some improvement but falls short of being optimal. In contrast, the *BRIEF* descriptor appears to provide the most effective description of the WSI landmarks, followed by *SIFT*.

These observations are supported by statistical analysis, as follows. The Friedman test provides a p-value below  $10^{-15}$ , indicating highly significant differences among the descriptor distributions. Additionally, the post-hoc Nemenyi test reveals that all distributions are significantly different from each other with p-values below the 0.001 threshold, except the distributions of *FREAK* and *VGG* where no statistically significant difference is found, with a p-value of 0.9. Results on a representative image are shown in Figure 2.

In the second part of the evaluation, the performance of feature-based registration is assessed by examining the distribution of RMSE values after alignment. The medians and interquartile values of these distributions are provided in Table II, along with the initial RMSE values (before any alignment).

As expected and in contrast to the other descriptors, *Random* and *Neighborhood* yield completely unsatisfactory results, with final alignments being worse than the initial ones.

A Friedman test with the Nemenyi post-hoc confirms that the results are significantly different between the descriptors (global p-value below  $10^{-15}$ ) and that the results of *BRIEF* are significantly different from all the others (all p-values below 0.0035).

Therefore, in this particular configuration for registering WSI, both evaluation point to *BRIEF* as the best descriptor and most effective choice.

## VII. DISCUSSION

Our research demonstrates that the choice of descriptor has a significant impact on the feature matching process and, consequently, the overall feature-based registration in digital pathology. Although the widely used *SIFT* descriptor generally produces satisfactory results, our findings indicate that it is not the optimal choice in the tested pipeline configuration, despite its prevalent use in the field of digital pathology.

A major limitation of our study is that we only vary the descriptor selection while keeping the methods in the other

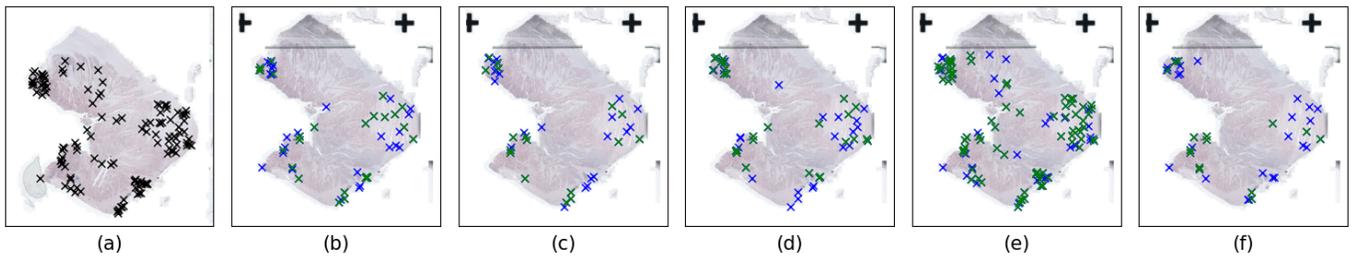


Fig. 2. Results of landmark matching on a representative image. (a) The fixed image showing its landmarks in black. (b) The moving image displaying landmarks correctly matched in the *SIFT* descriptor space in green, and landmarks with normalized ranks less than or equal to 0.05 in blue. Same is shown for the (c) *BRISK*, (d) *FREAK*, (e) *BRIEF* and (f) *VGG* descriptors.

steps of the feature matching pipeline fixed. The potential interactions between these different steps, which could have a significant impact on the performance of WSI registration, were not taken into account. The choice of the robust estimation process will affect the registration outcomes. Fine-tuning the parameters of the descriptors, which were assessed using their default values from the literature, could also improve their performance. Future work will also study whether pre-processing techniques such as color deconvolution may improve the robust matching stage by enhancing image similarity (particularly in the case of multi-stained images pairs).

Although it will be necessary to conduct a more exhaustive comparative analysis that combines various algorithms for each step of the pipeline, our results show that *BRIEF* may be a much more reliable choice of descriptor for WSI registration, leading to more accurate matching and, ultimately, a better estimation of the transform.

#### ACKNOWLEDGMENT

We would like to thank Christian Marzahl for sharing his WSI datasets.

#### REFERENCES

- [1] J. Borovec, J. Kybic, I. Arganda-Carreras, D. Sorokin, G. Bueno, A. Khvostikov, S. Bakas, E. Chang, S. Heldmann, K. Kartasalo, L. Latonen, J. Lotz, M. Noga, S. Pati, K. Punithakumar, P. Ruusuvoori, A. Skalski, N. Tahmasebi, M. Valkonen, and A. Muñoz-Barrutia, "AN-HIR: Automatic Non-rigid Histological Image Registration Challenge," *IEEE Transactions on Medical Imaging*, vol. PP, pp. 1–1, Apr. 2020.
- [2] P. Weitz, M. Valkonen, L. Solorzano, C. Carr, K. Kartasalo, C. Boissin, S. Koivukoski, A. Kuusela, D. Rasic, Y. Feng, S. S. Pouplier, A. Sharma, K. L. Eriksson, S. Robertson, C. Marzahl, C. D. Gatzenbee, A. R. A. Anderson, M. Wodzinski, A. Jurgas, N. Marini, M. Atzori, H. Müller, D. Budelmann, N. Weiss, S. Heldmann, J. Lotz, J. M. Wolterink, B. De Santi, A. Patil, A. Sethi, S. Kondo, S. Kasai, K. Hirasawa, M. Farrokh, N. Kumar, R. Greiner, L. Latonen, A.-V. Laenkholm, J. Hartman, P. Ruusuvoori, and M. Rantalainen, "The ACROBAT 2022 Challenge: Automatic Registration Of Breast Cancer Tissue," May 2023.
- [3] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer Vision*, vol. 60, pp. 91–110, Nov. 2004.
- [4] S. Leutenegger, M. Chli, and R. Y. Siegwart, "BRISK: Binary Robust invariant scalable keypoints," in *2011 International Conference on Computer Vision*, pp. 2548–2555, Nov. 2011. ISSN: 2380-7504.
- [5] J. L. Schönberger, H. Hardmeier, T. Sattler, and M. Pollefeys, "Comparative Evaluation of Hand-Crafted and Learned Local Features," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6959–6968, July 2017. ISSN: 1063-6919.
- [6] T. Mouats, N. Aouf, D. Nam, and S. Vidas, "Performance Evaluation of Feature Detectors and Descriptors Beyond the Visible," *Journal of Intelligent & Robotic Systems*, vol. 92, pp. 1–31, Sept. 2018.
- [7] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded Up Robust Features," in *Computer Vision – ECCV 2006* (A. Leonardis, H. Bischof, and A. Pinz, eds.), Lecture Notes in Computer Science, (Berlin, Heidelberg), pp. 404–417, Springer, 2006.
- [8] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *2011 International Conference on Computer Vision*, pp. 2564–2571, Nov. 2011. ISSN: 2380-7504.
- [9] A. Alahi, R. Ortiz, and P. Vandergheynst, "FREAK: Fast Retina Keypoint," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 510–517, June 2012. ISSN: 1063-6919.
- [10] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "BRIEF: Binary Robust Independent Elementary Features," vol. 6314, pp. 778–792, Sept. 2010.
- [11] C. Marzahl, F. Wilm, F. Dressler, L. Tharun, S. Perner, C. Bertram, C. Kröger, J. Voigt, R. Klopffleisch, A. Maier, M. Aubreville, and K. Breininger, "Robust Quad-Tree based Registration on Whole Slide Images," in *Proceedings of Machine Learning Research*, vol. 156, pp. 181–190, 2021.
- [12] M. Wodzinski and A. Skalski, "Multistep, automatic and nonrigid image registration method for histology samples acquired using multiple stains," *Physics in Medicine and Biology*, vol. 66, p. 025006, Jan. 2021.
- [13] C.-W. Wang, Y.-C. Lee, M.-A. Khalil, K.-Y. Lin, C.-P. Yu, and H.-C. Lien, "Fast cross-staining alignment of gigapixel whole slide images with application to prostate cancer and breast cancer analysis," *Scientific Reports*, vol. 12, p. 11623, July 2022. Number: 1 Publisher: Nature Publishing Group.
- [14] C. D. Gatzenbee, A.-M. Baker, S. Prabhakaran, R. J. C. Slebos, G. Mandal, E. Mulholland, S. Leedham, J. R. Conejo-Garcia, C. H. Chung, M. Robertson-Tessi, T. A. Graham, and A. R. A. Anderson, "VALIS: Virtual Alignment of pathoLogY Image Series," Nov. 2021. Pages: 2021.11.09.467917 Section: New Results.
- [15] K. Simonyan, A. Vedaldi, and A. Zisserman, "Learning Local Feature Descriptors Using Convex Optimisation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, pp. 1573–1585, Aug. 2014. Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence.
- [16] C. Zhang, Y. Jiang, N. Li, Z. Zhang, M. T. Islam, J. Dai, L. Liu, W. He, W. Qin, J. Xiong, Y. Xie, and X. Liang, "A Hybrid Deep Feature-Based Deformable Image Registration Method for Pathology Images," Apr. 2023.
- [17] D. DeTone, T. Malisiewicz, and A. Rabinovich, "SuperPoint: Self-Supervised Interest Point Detection and Description," Apr. 2018.
- [18] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.
- [19] V. A. Pyatov and D. V. Sorokin, "Affine Registration of Histological Images Using Transformer-Based Feature Matching," *Pattern Recognition and Image Analysis*, vol. 32, pp. 626–630, Sept. 2022.
- [20] Y. Jin, D. Mishkin, A. Mishchuk, J. Matas, P. Fua, K. M. Yi, and E. Trulls, "Image Matching across Wide Baselines: From Paper to Practice," *International Journal of Computer Vision*, vol. 129, pp. 517–547, Feb. 2021.
- [21] X. Moles Lopez, P. Barbot, Y.-R. Van Eycke, L. Verset, A.-L. Trépan, L. Larbanoix, I. Salmon, and C. Decaestecker, "Registration of whole immunohistochemical slide images: an efficient way to characterize biomarker colocalization," *Journal of the American Medical Informatics Association : JAMIA*, vol. 22, pp. 86–99, Jan. 2015.