



Whole slide image registration via multi-stained feature matching

Md. Ziaul Hoque ^{a,b,*}, Anja Keskinarkaus ^{a,b}, Pia Nyberg ^{c,e}, Taneli Mattila ^{d,e},
Tapio Seppänen ^{a,b}

^a Physiological Signal Analysis Group, Center for Machine Vision and Signal Analysis, University of Oulu, Finland

^b Faculty of Information Technology and Electrical Engineering, University of Oulu, Finland

^c Biobank Borealis of Northern Finland, Oulu University Hospital, Finland

^d Department of Pathology, Oulu University Hospital, Finland

^e Translational & Cancer Research Unit, Medical Research Center Oulu, Faculty of Medicine, University of Oulu, Finland



ARTICLE INFO

Keywords:

Computer-aided diagnosis
Digital histopathology
Histological image registration
Hematoxylin and eosin staining
Scale invariant feature transform
Feature slope computing

ABSTRACT

In the recent decade, medical image registration and fusion process has emerged as an effective application to follow up diseases and decide the necessary therapies based on the conditions of patient. For many of the considerable diagnostic analyses, it is common practice to assess two or more different histological slides or images from one tissue sample. A specific area analysis of two image modalities requires an overlay of the images to distinguish positions in the sample that are organized at a similar coordinate in both images. In particular cases, there are two common challenges in digital pathology: first, dissimilar appearances of images resulting due to staining variances and artifacts; second, large image size. In this paper, we develop algorithm to overcome the fact that scanners from different manufacturers have variations in the images. We propose whole slide image registration algorithm where adaptive smoothing is employed to smooth the stained image. A modified scale-invariant feature transform is applied to extract common information and a joint distance helps to match keypoints correctly by eliminating position transformation error. Finally, the registered image is obtained by utilizing correct correspondences and the interpolation of color intensities. We validate our proposal using different images acquired from surgical resection samples of lung cancer (adenocarcinoma). Extensive feature matching with apparently increasing correct correspondences and registration performance on several images demonstrate the superiority of our method over state-of-the-art methods. Our method potentially improves the matching accuracy that might be beneficial for computer-aided diagnosis in biobank applications.

1. Introduction

Nowadays medical image registration is the most challenging and emerging optimization process of applying a variety of geometric transformations over one or more images; in order to match their spatial viewpoints with one of the target images to establish correspondences among them [1]. It is an indispensable part of many detection tasks such as image combination, viewpoints detection, fusion, area detection, similarity monitoring, feature detection, and feature matching. The target images must contain some common structures that are expected to be placed at comparable areas, similar locations, and orientations during the registration process [2]. Image registration is used in several analytical studies including research to understand population propensities of phenotypes, estimate longitudinal changes, monitor the extent of tumor tissues, find out cells or tissues that are affected by the

disease, relate individual anatomy with a standard system and execute guided medical processes among different applications [2,3].

In histopathology, tissue sample slides are prepared using special stains to investigate the samples under a microscope and find out tissues that are affected by the disease [4–7]. Staining makes tissue samples transparent and colored for examination by the pathologists or doctors. Different histological, special, or immunohistochemical stains highlight different cells, cell structures, and tissue structures, resulting in dissimilar appearances on nearby tissue areas. Hematoxylin and eosin dyes are commonly used where hematoxylin stain attaches with acid structures and eosin stain attaches with eosinophilic structures. Moreover, hematoxylin stain appears as a blue or purple color (e.g. cells, nuclei), and eosin stain appears as a pink color (e.g. cytoplasm, collagen, muscle fiber) [4,5,7]. In digital pathology sections of tissue mounted on glass slides are digitalized with specialized scanners producing virtual slides,

* Corresponding author. Physiological Signal Analysis Group, Center for Machine Vision and Signal Analysis, University of Oulu, Finland.

E-mail address: mdziaul.hoque@oulu.fi (Md.Z. Hoque).

also called whole slide images. Accordingly, the whole process; specimen preparations, staining solution, stain coloring, chemistry oxidation, staining time, pH level of solutions, manufacturer reactivity as well as the use of different scanners, equipments, etc. varies the color appearance in whole slide images [5,7].

Digital histopathology images need to be registered to create three dimensional reconstruction from scanned thin slices [8–10]. A pathologist usually observes numerous slides and after that combines the visual information to form a diagnosis. When applying special or immunohistological staining, sequential slides give integral information where the image registration step is required to ensure an efficient investigation. The staining variations caused by different processes involved in sample preparation and differences in scanner equipments can affect the deployment of computational pathology, especially when digital slides are shared in large-scale cohorts and biobanks. Accordingly, there is a need to understand the variation. Successful whole slide image registration is a necessary part for the deep analysis of slide differences, distortion compensation [11], fusing information [12–14], creating high resolution mosaic from two dimensional small tiles [15], and segmenting unknown stains using known stains [16]. However, there are numerous difficulties confronted while registering histopathological images.

Whole slide image (WSI) registration is challenging due to pyramidal structures with various levels of resolutions, artifacts, missing sections on account of sample preparation, differences in local structure between slices [17], and color variations created by sectioning, staining, mounting, and scanning processes. It is also difficult to handle the pyramidal structure due to the lack of computer memory and display characteristics [6,7]. The size of the image is high, and special consideration must be placed to implement while maintaining high quality for analysis. Usually, the images are scanned with high resolution (20X or 40X magnification), resulting in several gigabyte sizes of images [2–7, 17]. Currently, few WSI registration methods have been proposed.

Lotz et al. [18] proposed a WSI registration to fuse the information from multiple neighboring immune-histochemically stained whole slides. This method registers the images with low resolution using nonlinear deformation model to account for low-frequency global deformation of tissue and refines those results on patches. The deformations are computed on each patch to combine the interpolation of globally smooth nonlinear deformation. The authors claimed that the corrected high-frequency local parts of the deformation are invisible at low resolution representations. But in practice, it creates difficulties selecting the right patches to compute with significantly longer runtime and causes registration errors. Lopez et al. [19] introduced an area-based b-spline method for biomarker co-localization from adjacent sections of immune-histochemically stained whole slides. Linear transformations are used to increase the speed of the registration process. This method can only measure the affine relation between the pixel intensity values of two images and utilize the similarity on randomly sampled pixels. Therefore, this method cannot alleviate the registration error caused by artifacts, tissue splitting, and folding problems in whole slide images. Roberts et al. [20], proposed a method that aligns consecutive slides to create 3D histological reconstruction via non-rigid transformation. A patch-based registration is applied to produce high-resolution 3D reconstructions of histopathological images. Finally, global b-spline nonrigid transformation is evaluated for the slice-to-slice registration. This method is unable to estimate the common information in the presence of local tissue and imaging irregularities. Besides, this method is effective for the alignment not for the 3D histological reconstruction of histopathological images with high resolution.

Rossetti et al. [21] proposed a dynamic 3D registration method that produces registered sub volumes from the gigapixel serial whole slide image. In their study, serial hematoxylin and eosin-stained tissue sections are used. Then the mapping and propagation method connects the reference coordinate at low resolution level to the registered subvolume at high resolution level. But the registration of the serial tissue sections is

difficult using global rigid transform. In this case, local deformable transforms are needed. Therefore, the authors planned to extend their work with iterative registration corrections through pyramidal resolution levels to improve their method. However, the contrast enhancement with SIFT and B-spline techniques is robust for histopathology with the same staining differences where color normalization [3,7,22] is required. But this technique has poor performance for the staining difference from one image to another due to the lack of proper color standardization.

Senthilnath et al. [23] introduced a registration method based on intensity variations and spatial relationships between two images for two-dimensional and three-dimensional satellite image registration. This method is also robust for the stained whole slide images, but the original intensities may introduce errors in different stains and influence the registration process. Several techniques have been proposed for pathology bearing image registration, mostly in radiology. These techniques can be divided into two categories including feature-based methods and intensity-based methods [1,2]. Intensity-based methods are useful to identify the similarity between pixel intensities for the determination of alignment between two images. But intensity-based techniques experienced poor effects on optimal alignment [2]. Feature-based methods extract salient features using a correlation between features and influence optimal alignment. Normally, these features include edge, point, and centroid of a particular region [3,24]. Moreover, due to the invariance to image scaling and rotation, the Scale Invariant Feature Transform (SIFT) [25] has been applied in various application areas. Also, SIFT is invariant to change in illumination factors, which can be considered useful to the successful registration of whole slide images.

SIFT [25], SURF [26], ORB [27], KAZE [28] etc. are the classical techniques that match features efficiently. These classical image registration techniques are robust and used in various fields. Nonetheless, when SIFT is specifically applied to pathological images, the number of correct correspondences is not sufficient to confirm coordinate matching accuracy due to variation in intensity mapping [10] where the mapping of the intensity can be linear or nonlinear. To overcome this problem, Li et al. [29] proposed robust Fast SIFT (R-SIFT), where the gradient orientation of each pixel is refined, and gradually main orientations are transferred to each of the keypoints. To improve the matching performance, the scale orientation with joint distance measurement can be considered.

Kupfer et al. [30] proposed a quick mode seeking SIFT (MS-SIFT) algorithm where SIFT features are associated with position, scale, and orientation to perform into the transformation space. This eliminates outlying corresponding keypoints through outlier removal of uncertain SIFT keypoint correspondences and improves the overall matching. Sedaghat et al. [31] proposed the uniform SIFT algorithm which successfully creates enough distribution of aligned keypoints. The extracted features are initiated to a cross-matching process by checking consistency through a projective transformation model. Additionally, Dellingler et al. [32] proposed SAR-SIFT image registration algorithm. This algorithm introduces a new gradient computation yielding orientation and magnitude to minimize the speckle noise for the registration of SAR images with different incidence angles. Gong et al. [33] introduced a coarse to fine scheme for image registration based on SIFT and mutual information. The coarse outcomes provide a near-optimal introductory result for the enhancer in the calibrating process. We are motivated by the fact that remote sensing images have been commonly used in most of the current state-of-the-art developments. Inspired by the performance of different SIFT-based methods, we introduce a modified SIFT method to apply to whole slide images.

In this paper, we propose a gradient scale-invariant feature transform based registration method under feature slope grouping to align diagnostic lung adenocarcinoma, a lung cancer subtype, hematoxylin and eosin stained slides. However, whole slide image is a representation of colors in the original slide, but not always an exact reproduction of

them. The effect of scanner illumination, scanner, and slide optics, onboard image processing may cause color variation, artifacts, high-frequency noise, and variation on pixel intensities in the final virtual slide image [34]. Based on our observation, the size, lack of spatial information and interclass similarity, color variations, and artifacts introduced by sectioning, mounting, and imaging have a great impact on the performance of whole slide image registration. Therefore, we propose a method to overcome difficulties in registering whole slide images and enable further studies of these variations. Computational pathology is an area of growing interest, and our goal is to develop methods, which can provide more understanding of the aspects that might affect the accuracy and reproducibility of automatic image analysis methods in the area of digital pathology.

Finally, we compare the performance of our proposed method with SIFT [25], SURF [26], ORB [27], KAZE [28], and MS-SIFT [30] methods and discuss the results. Generally, corners are the interest points in the image that are invariant to illumination, translation, and rotation [35]. In particular, the hematoxylin and eosin stained whole slide images have fewer edge and corner areas. To utilize and valid this statement, we also test the corner detection methods in our experimental comparison. However, these are popular registration methods, with specific application areas, but not previously applied to the whole slide images.

The rest of the paper is organized as follows; Section 2 represents the proposed method. Section 3 shows the simulation results and the performance comparison of the proposed method with selected state of the art methods. A discussion on the methods including limitations and future perspectives of hybrid deep learning models is provided in Section 4. Finally, the main outcomes with conclusive statements and perspective works are concluded in Section 5.

2. Proposed method

The scale-invariant feature transform (SIFT) method might be the best solution for the large-scale whole slide image registration. However, the traditional SIFT method has still challenges in matching the correct number of accurate keypoint matches at the same time while eliminating keypoint mismatches. Therefore, we propose a modified SIFT algorithm under feature slope grouping to find out negative, positive, and approximately zero slopes and Euclidean transformation model to estimate rotation angle, vertical and horizontal shifts by averaging differences of keypoint locations. In addition, fast sample consensus [36] is applied for refining matches to calculate the correct correspondences from keypoint pairs and get registered image. To address dissimilar appearances due to staining variance difficulties, our proposed method extracts the common information and performs registration on scaled variants of images with thresholding [30]. After that, the successive transformation is applied to the full resolution images. It is not always expected to have a minimum Euclidean distance where the correct correspondence depends on descriptors formed by gradient magnitudes and orientations around the area of keypoints. Therefore, the robust descriptors are extracted by calculating the gradient orientations and magnitudes for each pixel in Gaussian scale space. In order to improve accuracy, we use the original input image to build the lowest level of the Gaussian pyramid.

We process gradient magnitude of the Gaussian scale-space image with adaptive smoothing [37]. The gradient magnitude, $D_{\delta}^{(t)}(x,y)$ and gradient orientation, $R_{\delta}^{(t)}(x,y)$ of t iteration are computed as follows,

$$D_{\delta}^{(t)}(x,y) = \sqrt{\left\{D_{x,\delta}^{(t)}(x,y)\right\}^2 + \left\{D_{y,\delta}^{(t)}(x,y)\right\}^2} \quad (1)$$

$$R_{\delta}^{(t)}(x,y) = \arctan \left[\begin{array}{c} D_{y,\delta}^{(t)}(x,y) \\ D_{x,\delta}^{(t)}(x,y) \end{array} \right] \quad (2)$$

where, $D_{x,\delta}^{(t)}(x,y) = \frac{1}{2}[I_{in}^{(t)}(x+1,y) - I_{in}^{(t)}(x-1,y)]$ and $D_{y,\delta}^{(t)}(x,y) = \frac{1}{2}[I_{in}^{(t)}(x,y+1) - I_{in}^{(t)}(x,y-1)]$ are the horizontal and vertical derivatives in scale of δ , Gaussian scale space respectively. The weight coefficients are computed as follows,

$$W_{\delta}^{(t)}(x,y) = \exp \left(-\frac{\left[D_{x,\delta}^{(t)}(x,y)\right]^2 + \left[D_{y,\delta}^{(t)}(x,y)\right]^2}{2b^2}\right) \quad (3)$$

The smoothed image $I_{in}^{(t+1)}(x,y)$ is defined as follows,

$$I_{in}^{(t+1)}(x,y) = \frac{\sum_{i=-1}^1 \sum_{j=-1}^1 I_{in}^{(t)}(x+i, y+j) W_{\delta}^{(t)}(i,j)}{\sum_{i=-1}^1 \sum_{j=-1}^1 W_{\delta}^{(t)}(x+i, y+j)} \quad (4)$$

The weight coefficient $W_{\delta}^{(t)}(x,y)$ is achieved by taking parameter b as 1.5 to 2 times standard deviation and the image gradient. The computed gradients using Equation (1) and Equation (2) are used to process descriptor extraction and assign orientation. For the evaluation of our method, the gradient orientations of the large image pairs are quantized in eight bins [38]. Fig. 1 shows the overall flow chart including different components of our proposed WSI registration method.

Our SIFT-based algorithm has five fundamental steps. First, a scale space extrema estimation utilizing the Difference of Gaussian. Second, keypoint candidate localization and elimination of the low contrast keypoints. Third, keypoint candidates' orientation assignment with local image gradient. Fourth, feature slope computing and grouping. Fifth, feature matching, outlier removing, and descriptor generator (based on image gradient magnitude and orientation) that help to compute the local image descriptor for each keypoint candidate [36]. The similarity transformation is essential to utilize the enhanced feature matching that consists of three parameters including translation, scale, and rotation.

Our algorithm estimates the rotation angle, vertical and horizontal shifts by averaging differences of keypoint locations. It works with three main modules such as keypoint detection, descriptor extraction, and keypoint matching [33]. The difference of Gaussian scale space is estimated as Laplacian of Gaussian approximation. Then local extrema are selected as the candidate keypoints. These keypoints are assigned at least one fundamental orientation depending on a local histogram of gradient orientation. In the similarity transformation model, the correct matching will have the same rotation angle in space, same scale ratio, same horizontal and vertical shifts in most cases. Consequently, we utilize the inherent information (i.e., position, scale, and fundamental orientation) of each keypoint to build the right correspondences.

2.1. Initialization of keypoints matching

The initial matching of our proposed algorithm is the keypoint matching proportion to Euclidean distance of nearest neighbor and second nearest neighbor of corresponding descriptors. After selecting the whole area of two images the threshold is applied proportionally. Pair set is achieved using keypoint detection represented by the histograms of scale ratio (s^*), main orientation difference ($\Delta\theta^*$), horizontal (Δx^*), and vertical (Δy^*) shifts. The initial transformation parameter (μ_i) from the image pair is calculated by the improved random sample consensus called fast sample consensus [36].

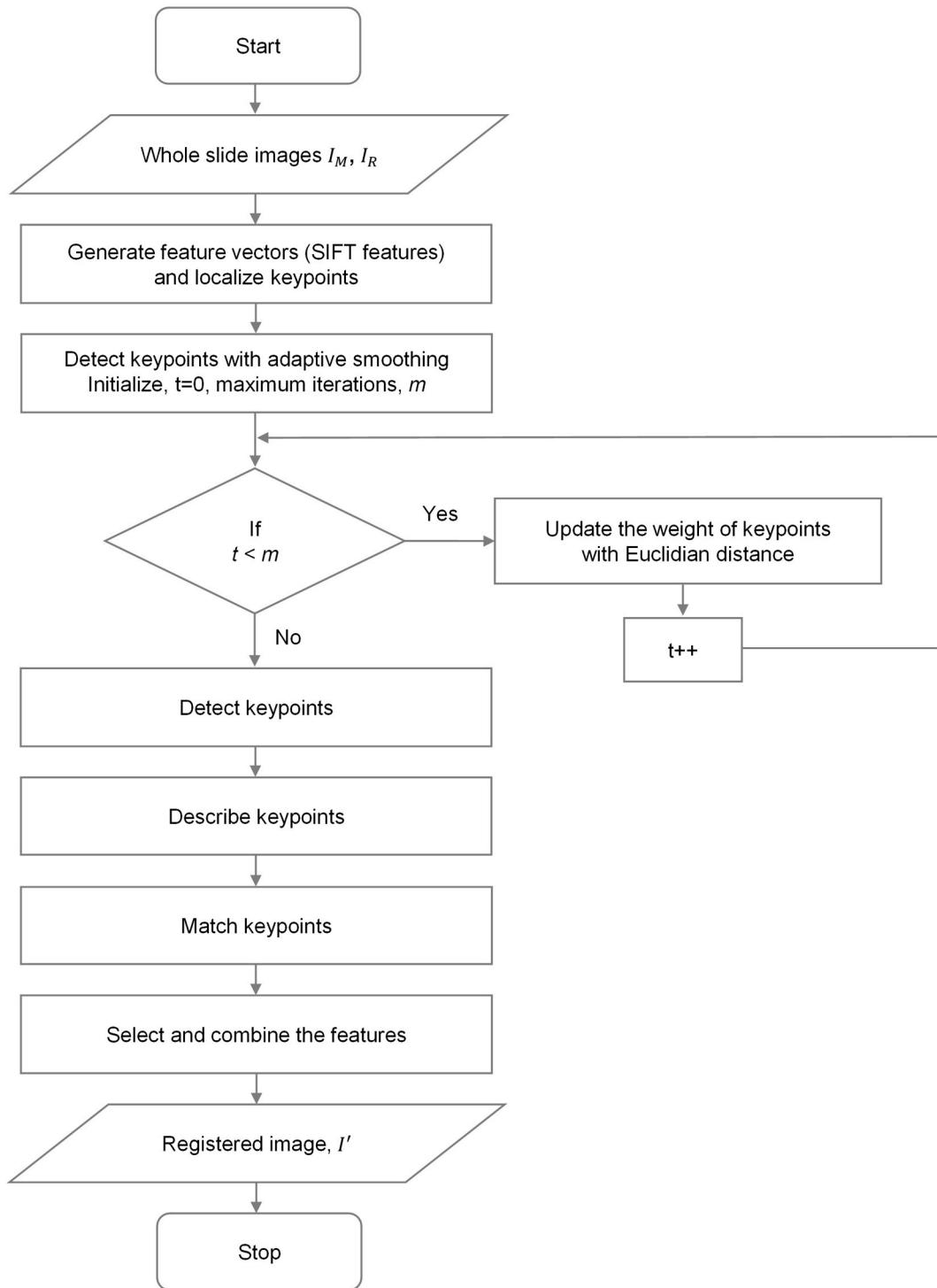


Fig. 1. Flow chart of the proposed multi-stained feature matching based whole slide image registration technique.

2.2. Rematching to obtain keypoint pairs

The primary orientation variation histogram has two modes where there are only two different combinations of s^* , $\Delta\theta^*$, Δx^* , and Δy^* . A joint distance is calculated to measure the distance for each mode combination as follows:

$$PD(k_i, k'_j) = (1 + \lambda_p)(1 + \lambda_s)(1 + \lambda_o)ED(k_i, k'_j) \quad (5)$$

where, $ED(k_i, k'_j)$ is the Euclidean distance of the descriptors corresponding to keypoints, $(k_i = k_1, k_2, \dots, k_N)$ and $(k'_j = k'_1, k'_2, \dots, k'_N)$ respectively. Moreover, Equation (5) depends on the following factors,

$$\lambda_p = l_p(k_i, k'_j) = \left| \left| (x_i, y_i) - F_t \left((x'_j, y'_j), \mu_i \right) \right| \right| \quad (6)$$

$$\lambda_s = l_s(k_i, k'_j) = \left| 1 - s^* \frac{P'_j}{P_i} \right| \quad (7)$$

$$\lambda_o = l_o(k_i, k'_j) = \text{abs}(\Delta\theta_{i,j} - \Delta\theta^*); \Delta\theta_{i,j} = \theta_i - \theta'_j \quad (8)$$

where, $(x_i, y_j), P_i, \theta_i$ and $(x'_j, y'_j), P'_j, \theta'_j$ are the positions of the keypoints k_i and k'_j respectively. $F_t((x_j, y_j), \mu_t)$ denotes similarity transformation function. s^* and θ^* are the scale ratio and orientation difference. The $PD(k_i, k'_j)$ will be minimized if the keypoint pairs will be effectively matched. Here, the proportion between the distance of nearest neighbor and second nearest neighbor helps to match keypoints correctly. Since coordinating is performed twice, keypoints in the reference image will have one or more considerable matching in another image. Thus, we regard the point pair with minimum $PD(k_i, k'_j)$ as the candidate coordinating pair. Accordingly, we obtain a set of keypoint pairs.

2.3. Feature slope computing and point grouping

In our approach, the feature slope computing [39] is used for conducting matched pairs as follows,

$$G_i = \frac{y_i^{Rf} - y_i^{Rg}}{(x_i^{Rf} + dx) - x_i^{Rg}} = \frac{\Delta v_i}{\Delta u_i + dx} \quad (9)$$

where, G_i is the computing slope for i^{th} matched pair at (x_i^{Rf}, y_i^{Rf}) in reference image and (x_i^{Rg}, y_i^{Rg}) in subject image, $\Delta u_i = x_i^{Rf} - x_i^{Rg}$ and $\Delta v_i = y_i^{Rf} - y_i^{Rg}$ are the horizontal and vertical distances, and dx is the displacement distance (distance between keypoints in reference image and corresponding keypoints in subject image). Here k-means clustering [40] is used to partition feature pairs into three groups e.g. positive, negative, and nearly zero slopes. A line length-based inequality helps to eliminate the potential mismatched features as follows,

$$\left[\sqrt{(\Delta u_i + dx)^2 + (\Delta v_i)^2} - \mu^n \right] > \epsilon \mu^n \quad (10)$$

where, first part is the length of i^{th} feature matching pair in n group, $\epsilon(0 < \epsilon < 1)$ is the tiny number to control the number of feature pairs to be removed and μ^n is the mean line length in n group that is defined as,

$$\mu^n = \frac{1}{M^n} \sum_{i=1}^{M^n} \sqrt{(\Delta u_i + dx)^2 + (\Delta v_i)^2} \quad (11)$$

where, M^n is the total feature pairs in the n group. When Equation (10) will be satisfied, the feature pair will be processed and considered as mismatched to remove. The covariance matrix is computed as follows,

$$C^n = \begin{bmatrix} \sigma_{xx}^n & \sigma_{xy}^n \\ \sigma_{yx}^n & \sigma_{yy}^n \end{bmatrix} \quad (12)$$

where, σ_{xx}^n and σ_{yy}^n denote the variances, and σ_{xy}^n and σ_{yx}^n denote the covariances in terms of remaining feature locations in group n . Finally, the maximum and minimum slopes in n are computed and grouped.

2.4. Outlier removal

Subsequently, our proposed method filters out the majority of outliers to eliminate false correspondences. The Fast Sample Consensus algorithm is applied in the outlier removal step. It helps to get more matching performance than random sample consensus [41] with few iterations. The horizontal and vertical displacements of corresponding keypoints are defined as follows,

$$\left. \begin{array}{l} \Delta x_1 = x_1 - s^*(x'_1 \cos(\Delta\theta^*) - y'_1 \sin(\Delta\theta^*)) \\ \Delta y_1 = y_1 - s^*(x'_1 \sin(\Delta\theta^*) + y'_1 \cos(\Delta\theta^*)) \end{array} \right\} \quad (13)$$

where, (x_1, y_1) and (x'_1, y'_1) are the coordinates of corresponding keypoints. The outliers are removed using a filter [30] as follows,

$$|\Delta x_1 - \Delta x^*| \geq \Delta x_{th}, |\Delta y_1 - \Delta y^*| \geq \Delta y_{th} \quad (14)$$

where, Δx_{th} and Δy_{th} are the horizontal and vertical displacement thresholds. The objective of image registration is to determine a transformation, which maximizes the similarity between two image pairs. To deal with large whole slide images, full size is down sampled to 10x apparent magnification scale for performing registration on downscaled images. Finally, our proposed model is applied to full-sized images by upscaling the extracted deformation details of correct matching interpolations and different color intensity levels. Using input images (source and target), the common information (correct correspondences and interpolation of color intensities) is extracted in the following computational transformations of the model. The overall procedure of our model is illustrated in Fig. 2.

2.5. Evaluation metrics

The quantitative comparisons of different reference methods are evaluated by measuring F-score, Precision, Recall, and average processing time concerning cumulative distributions. The F-score, Precision, and Recall are defined as follows,

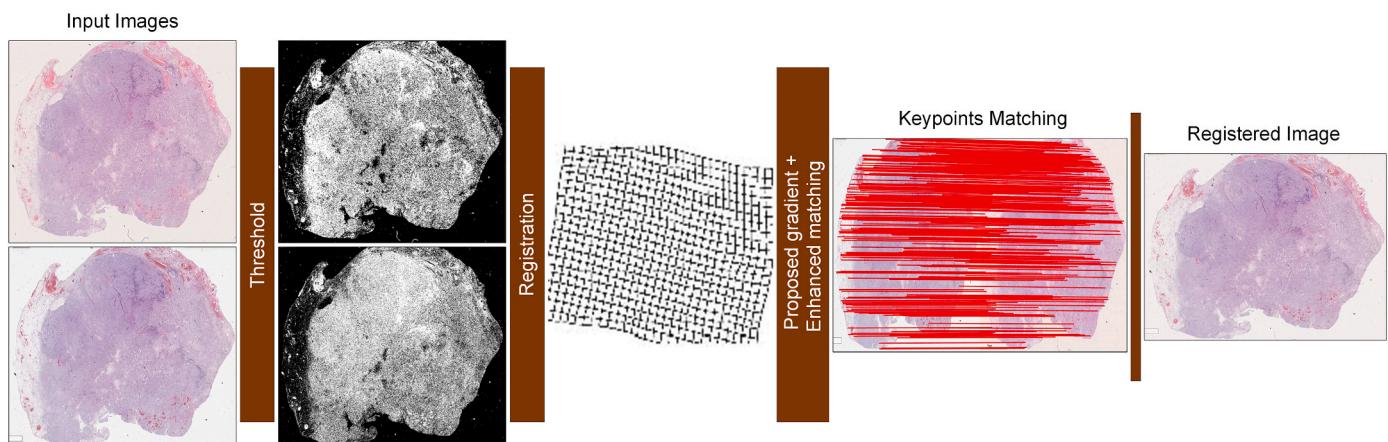


Fig. 2. General framework that applied to the source images to obtain the aligned image.

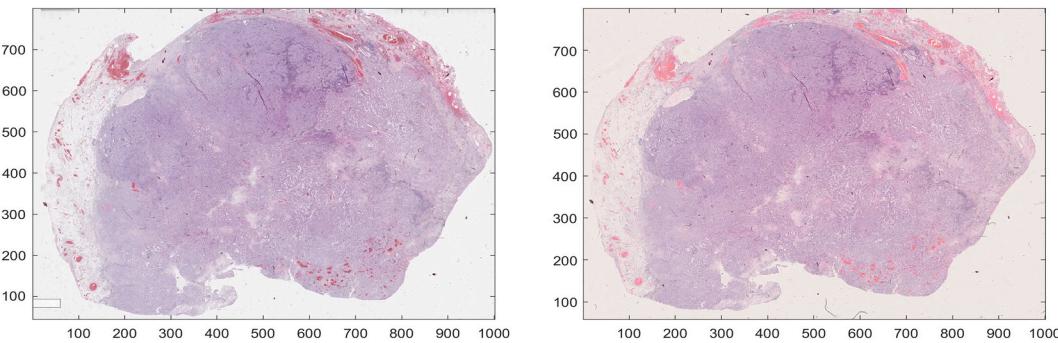


Fig. 3. An example of lung sample image pair; reference image (oys_aperio.svs on Left) and subject image (oys_hama.ndpi on Right).

$$F-score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (15)$$

$$Precision = \frac{\text{Number of identified correct match}}{\text{Number of preserved match}} \quad (16)$$

$$Recall = \frac{\text{Number of identified correct match}}{\text{Number of correct match contained in putative set}} \quad (17)$$

Moreover, the performance evaluation is measured by calculating RMSE (root mean square error) [30,33], MAE (maximum error), MEE (median error), and STD (standard deviation), where the number of N corresponding point pairs, $\{(x_i, y_i), (x'_i, y'_i)\}$ are picked from reference and subject images. The RMSE, MAE, MEE, and STD are formulated as follows,

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - x'_i)^2 + (y_i - y'_i)^2} \quad (18)$$

$$MAE = \max \left\{ \sqrt{(x_i - x'_i)^2 + (y_i - y'_i)^2} \right\}_{i=1}^N \quad (19)$$

$$MEE = median \left\{ \sqrt{(x_i - x'_i)^2 + (y_i - y'_i)^2} \right\}_{i=1}^N \quad (20)$$

$$STD = \sqrt{\frac{1}{N} \sum_{i=1}^N [d(x'_i, y'_i) - RMSE]^2} \quad (21)$$

where, (x'_i, y'_i) represents the transformed coordinates of (x_i, y_i) and $d(\cdot, \cdot)$ represents the distance. The algorithm is performed multiple times and the average of those is processed as final result for each test image pair. Finally, the keypoint quantity of right correspondences is utilized as the key element to evaluate the robustness of the proposed method. The performance results of matching also contain incorrect matching or mismatching points. Therefore, we calculate the rate of correct correspondences (ROCC) based on the number of false and correct correspondences (NOFC and NOCC) [42] as follows,

$$ROCC = \frac{NOCC}{NOCC + NOFC} \quad (22)$$

The speed of the registration process is another important index to evaluate the robustness of registration methods. In this context, we also measure the run time (RT).

3. Simulation results and performance study

In this section, the experimental results of our proposed method are demonstrated with illustrations. Then the effectiveness and robustness of our proposed method over the state-of-the-art methods are compared. Generally, there is not any specific quantified evaluation metric for the

Table 1

Details information of oys_aperio.svs (Leica Aperio) and oys_hama.ndpi (Hamamatsu) format images.

Image information	oys_aperio.svs	oys_hama.ndpi
Description	Aperio image library v12.0.15 11988x89547 [0, 100 117528x89447] (240x240) RGB Q = 60	NanoZoomer Digital Pathology Image
Magnification	40X	40X
Image Type	SVS	NDPI 13
Image Width	117 528 pixels	182 784 pixels
Image Height	89 447 pixels	107 520 pixels
Image Depth	1 pixel	1 pixel
Image Channels	3	3
Image Bit Depth	8 bits	8 bits
File Size	3653670000 bytes	4402072000 bytes
Compression Type	JPEG using libjpeg	JPEG
Compression Quality	60	N/A
Compression Ratio	8.63	13.39
Tile Width	240 pixels	256 pixels
Tile Height	240 pixels	256 pixels

stained whole slide image registration. The registration accuracy evaluation is a difficult task, and it becomes more difficult without objective ground truth and the gold standard. In case of stained whole slide image registration, we preferred to go for manually segmented larger structures that are identifiable in both slide images. Thus, the registration accuracies were evaluated and validated by computing the manually segmented structures (assigned by two experienced collaborating pathologists using Aperio ImageScope and ImageJ [43,44]) and automatically segmented corresponding structures. A similar evaluation and validation process has been done earlier [18,45].

Finally, the proposed method is validated on several high resolution lung cancer (adenocarcinoma) sample images (35 whole slide image pairs). The methods are implemented under MATLAB R2019b with a machine running Linux, with an Intel Core i7-8700MQ Quad-Core processor at 3.20 GHz. The two step registration methods based on deep and local features with area matching are implemented under python 3.6.

3.1. Data description

Diagnostic lung adenocarcinoma HE-slides were selected from surgical resection samples and were digitalized by Leica Aperio AT2 and Hamamatsu scanners with 40x magnification. 35 image pairs were scanned and used for the analyses. The use of diagnostic images was approved by the ethical committee of Borealis Biobank of Northern Finland (BB-2020-4005). An example of lung adenocarcinoma image pair, Leica Aperio (.svs) and Hamamatsu (.ndpi) images are shown in Fig. 3. The detailed information about particular example pair of the

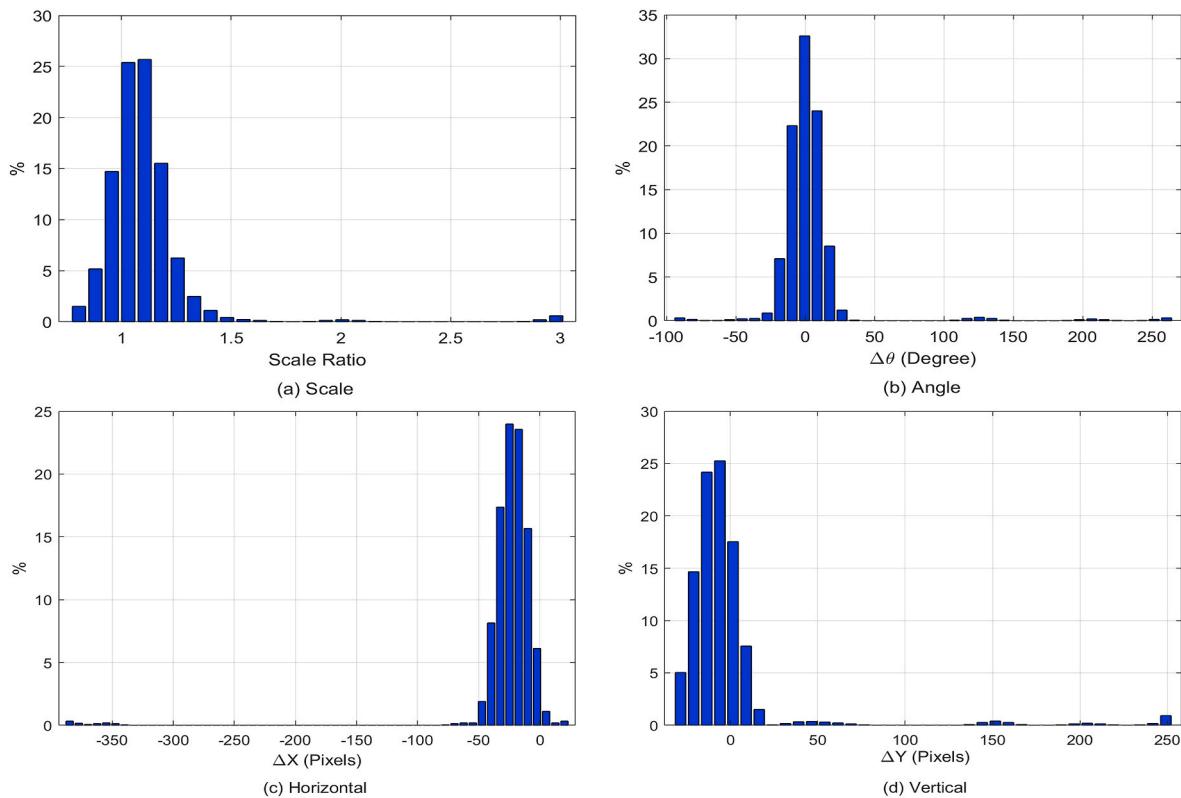


Fig. 4. Histograms of (a) scale ratio of the matched keypoints, (b) angle difference from Equation (14), (b) horizontal displacement difference from Equation (1), and (d) vertical displacement difference from Equation (1).

oys_hama.ndpi and oys_aperio.svs images is given in Table 1.

From Fig. 3 and Table 1, we can see that the overall geometrical structures of two images are same but the color constancy, background, artifacts, high-frequency noise availability, pixel intensities, size, etc. are not same.

3.2. Experimental results

We thoroughly validate the robustness and effectiveness of our method by comparing the output registered image with the input reference and subject images separately. The histograms of scale ratio, orientation difference, horizontal and vertical sifts, and angle difference of matched keypoints eventually exhibit a single mode apart from the orientation difference histograms (as shown in Fig. 4). The reason why orientation difference histograms work with two primary modes is that the circumferential angles are not constant in -180° and 180° . It is important to observe that, there is just a single fundamental mode in rotation angle.

Similar to the method [46] if we know one of the areas of two fundamental models, the area of another mode can be determined by the

following equation,

$$\Delta\theta' = \begin{cases} \Delta\theta + 360, & \Delta\theta \in [-360, 0) \\ \Delta\theta - 360, & \Delta\theta \in [0, 360) \end{cases} \quad (23)$$

where, $\Delta\theta$ and $\Delta\theta'$ are the locations of orientation difference histograms.

Fig. 4 represents the orientation difference histograms for the scale ratio due to variation of intensity mapping, circumferential angle difference, horizontal and vertical shifts. We can see a significant intensity variance resulted in different orientation differences of correspondences for the correct match.

Fig. 5 illustrates the three dimensional overview of various regions and the distribution of keypoint correspondences where yellow and brown color highlight smooth and more appearance variation areas respectively, and green and blue color highlight the area of high appearance variations.

It is important to determine the common information (e.g. tissue images contain comparative shape) before aligning any image for successful alignment. Usually, the general structure of the sample will be considered to perform registration. Once the information is segmented,

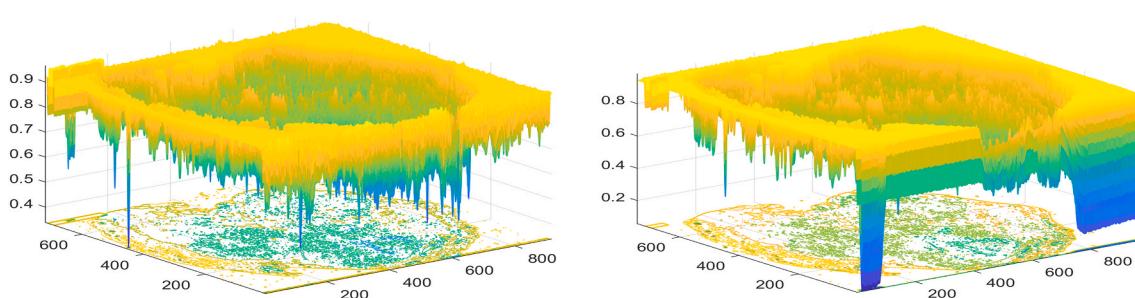


Fig. 5. Area visualization where yellow and brown color highlight smooth and more appearance variation areas respectively, and green and blue color highlight the area of high appearance variations; reference image (Left) and subject image (Right).

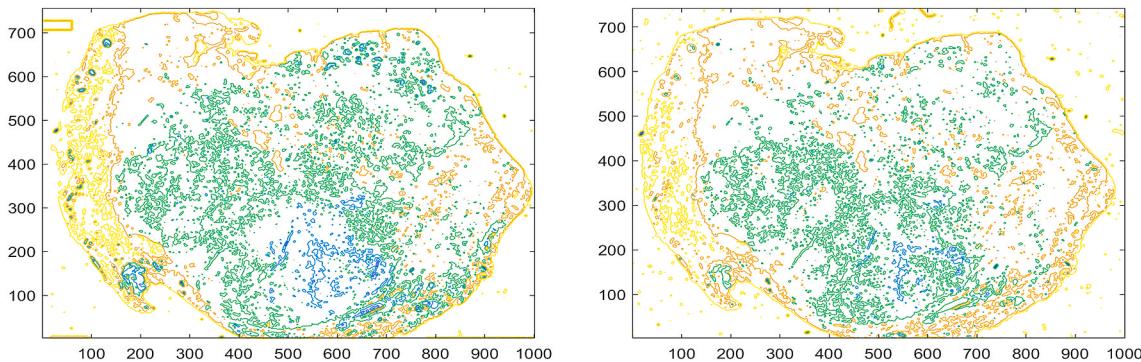


Fig. 6. Region classification where yellow and brown color highlight keypoints of smooth and more appearance variation areas respectively, and green and blue color highlight keypoints of the high appearance variation areas; reference image (Left), subject image (Right).

the registration can be processed and both images are aligned to locate correct correspondences and visualize the common information using region classification. Fig. 6 illustrates the region classification to visualize keypoint correspondences for the correct matching where yellow and brown color highlight keypoints of smooth and more appearance variation areas respectively, and green and blue color highlight keypoints of the high appearance variation areas.

We measure the intensity differences between reference image and subject image using difference of gaussian pyramid, orientation difference, and Gradient orientation as shown in Fig. 7. It is clear from Fig. 7 that the gradient definition is best to overcome the difference in image intensity. It is also better for an enhanced feature matching by the combination of position, scale, and orientation of keypoints to increase the correct correspondences. To measure the robustness and effectiveness of our method we calculated structural similarity index, average difference, mean cross-correlation, and absolute error by comparing the registered image processed by the proposed method with reference and subject images. The performance results of an example image pair (lung adenocarcinoma WSI) comparing the registered image processed by the proposed method with reference and subject images are given in Table 2.

The gradient orientations of the example image pair are quantized in

eight bins. It is seen from Fig. 8 that 877 and 1379 points are the initial point distributions from subject and reference images respectively and 292 points are the right number of points from subject and reference images. It is clear that our algorithm can precisely coordinate the test pair with the successful tuning of its parameters to acquire superior registration results.

To verify the robustness of the enhanced feature matching performance we compared the proposed gradient + fast sample consensus (improving random sample consensus to get more correct

Table 2

Performance comparisons by comparing the output registered image with the input reference and subject images.

Evaluation criterion	Comparison between registered and reference images	Comparison between registered and subject images
Structural similarity index	96.5825	94.0169
Average difference	3.4175	5.9831
Mean cross-correlation	0.9786	0.9697
Absolute Error	0.0679	0.0589

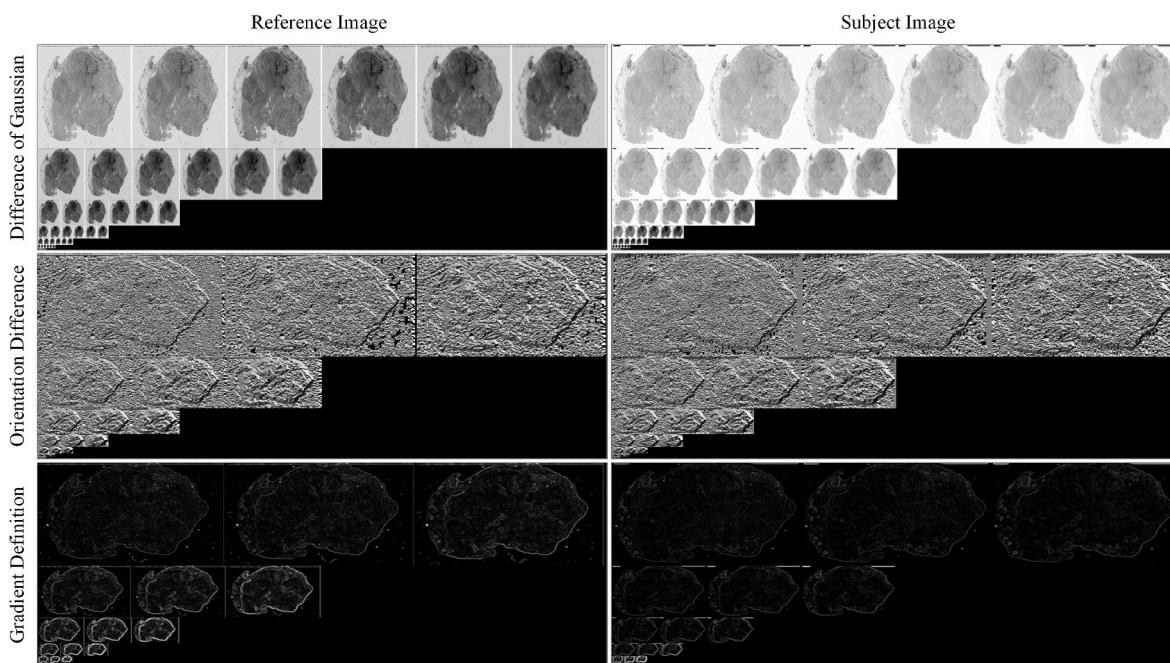


Fig. 7. Performance results of intensity mapping using difference of gaussian pyramid, orientation difference and gradient definition (including orientations and magnitudes).

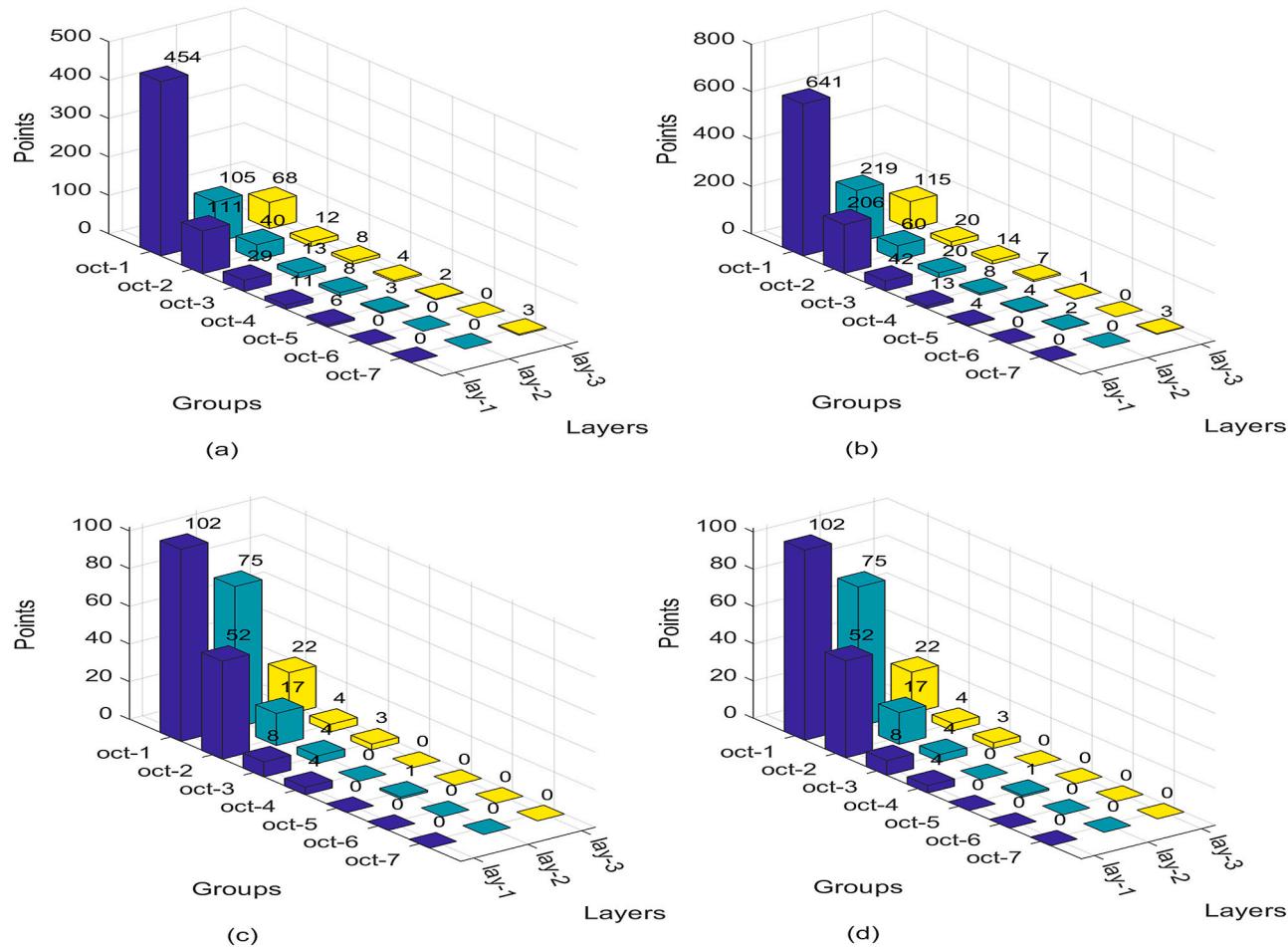


Fig. 8. Feature point distribution in eight bins of reference and subject images. (a) Initial point distribution 877 points from subject image, (b) Initial point distribution 1379 points from reference image, (c) Right number of points 292 from subject image, (d) Right number of points 292 from reference image.

correspondences) and proposed gradient + enhanced matching (using inherent information's e.g scale, position, and orientations for each keypoint pair to increase correct correspondences). From the illustration of results in Fig. 9, it can be observed that the enhanced matching using inherent information's e.g scale, position and orientations accurately increases the number of correctly matched keypoints. The proposed enhanced matching approach contrasted with fast sample consensus increases the right matches. It can also be observed that the edge and area of two image pairs are precisely overlapped, which demonstrates the superiority of our proposed algorithm. For the above consideration, we can say, our new gradient computational method is more robust to the complex nonlinear intensity transform of whole slide images.

3.3. Performance comparison with different methods

We compare the effectiveness and robustness of our proposed method with the state-of-the-art methods. The detailed performance results of coordinating precision, matching accuracy, effectively coordinated keypoint number, mismatching rates, and processing time are studied. The F-score, precision, recall in terms of cumulative distributions are measured to compare the efficiency and robustness of different methods. The image registration accuracies are also measured in terms of RMSE, MAE, MEE, and STD. In addition, different feature points are calculated by utilizing intensity differences and color variations. Fig. 10 presents different keypoint detection results, where green and red dots are the detected keypoints of reference image and subject image respectively and the yellow lines are the matched keypoints.

It is seen from Fig. 10 that our registration algorithm performs well in

dealing with the problem of intensity differences and detects the highest number of feature points compared to reference methods. Considering the outcomes, SIFT, MS-SIFT, KAZE, and SURF performed well. On the other hand, ORB has the poor experimental performance results. Table 3 summarizes the comparison of our method with state of the art methods.

It is clear from Table 3 that the experimental matching rate of SIFT, MS-SIFT, KAZE, and SURF are good, but they are less robust, and fewer points are detected in hematoxylin and eosin stained whole slide images. ORB has a poor experimental matching rate. This method is mostly suitable for object recognition, 3D reconstruction, image retrieval, and camera localization. Based on the experimental results we noted that these methods cannot achieve the best results on whole slide images. The detected points and matching rate of our method and MS-SIFT are comparatively better than other reference methods. The processing time of our method and MS-SIFT are 19.11s and 21.09s respectively, which are comparatively lower than other reference methods.

For the quantitative performance comparisons of different reference methods, we calculate the F-score, precision, recall in terms of cumulative distributions where comparatively high value indicates better registration performance. We also calculate the RMSE, MAE, MEE, and STD where a low value indicates the highly significant registration method. The high F-score, precision and recall values, and low RMSE, MAE, MEE, and STD values (Table 4 and Table 5) prove that our method and MS-SIFT are more effective and highly robust compared to other reference methods.

Fig. 11 shows precision, recall curves, and the graphical representations of the average matching and mismatching rate of our proposed method where the average matching rate is 95%, and the mismatching

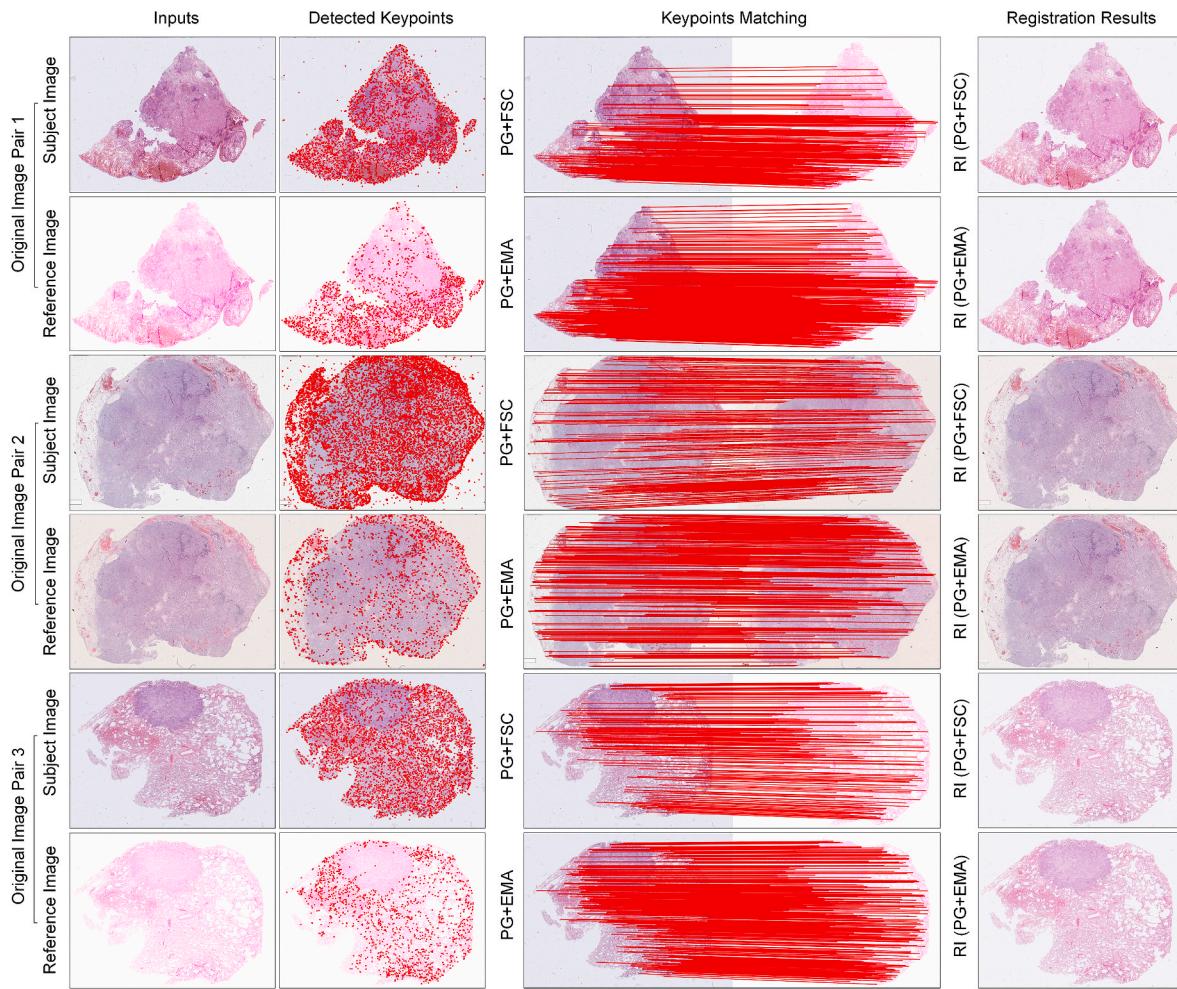


Fig. 9. Keypoints detection and multi-stained feature matching using our proposed gradient + fast sample consensus (PM + FSC) and proposed gradient + enhanced matching algorithm (PM + EMA) including final registration results. Note: our proposed gradient is used in both cases therefore the output registered images (RI) look quite similar.

rate is 5%.

The matching rate of detected points, matching and mismatching points, and processing time over the 35 whole slide image pairs are provided in Table A1 of the Appendix. Furthermore, (Mean \pm Standard Deviation) driven from RGB channels is calculated separately to describe and compare the main characteristics of different whole slide image pairs. It is clear that, our algorithm can accurately match the whole slide image pairs and is effective to the complex nonlinear intensity transform with gradient calculation to achieve high accuracy. The performance evaluation results show that our method is faster with better performance compared to other state-of-the-art methods in terms of aligning accuracy and matching number of keypoints.

However, our proposed method might be beneficial for the other fields (e.g. SAR image registration) apart from digital pathology and integrated to design hybrid approaches. To validate this statement we tested our proposed method in two-step registration combining deep and local features. The robustness, accuracy and advantages of our proposed method are verified by utilizing a two-step registration. We fine-tune the VGG-16 model and initialize the weight as similar to the previously introduced two-step registration method [42]. In first step, the approximate spatial relationship obtained by convolution neural network (CNN) is calculated to match and generate stable results of deep features. In second step, the matching strategy evaluating spatial relationship is applied to local feature based algorithms to get more accurate features in location to adjust the generated results of the first step. After calculating accurate transformation matrix according to the correct correspondences and

spatial relationship, registration results are obtained. In this case, random sample consensus is applied due to utilize the effectiveness of different state-of-the-art methods on homography estimation. SAR image pairs are used to evaluate all methods as similar to the two-step registration method based on deep and local features [42].

Table A2 shows the area matching results of different methods adopting CNN features (help to get the nearest neighbor). Most of the cases, two kinds of features are redundant which increases the complexity with comparatively high processing time requirements. It is also difficult to generate transformation matrix due to significant nonlinear differences between two image pairs. Our proposed method with deep features reduces the complexity by deep feature matching based on spatial relationships. From Table A2 it is seen that the performance results are comparatively better with taking full advantage of deep and local features. It is clear that the combination of deep and local features is a good way to SAR image registration, but it increases the complexities with registration error, high computational costs, and processing time. Therefore, it must take into consideration to explore the reduction of different complexities of this type of hybrid approach in the future.

4. Discussion

Our proposed algorithm effectively matches the whole slide image pairs and achieves significant matching accuracy over state-of-the-art methods. Perhaps surprisingly, most of the selected state-of-the-art methods used in our comparison are classical techniques. We spent a

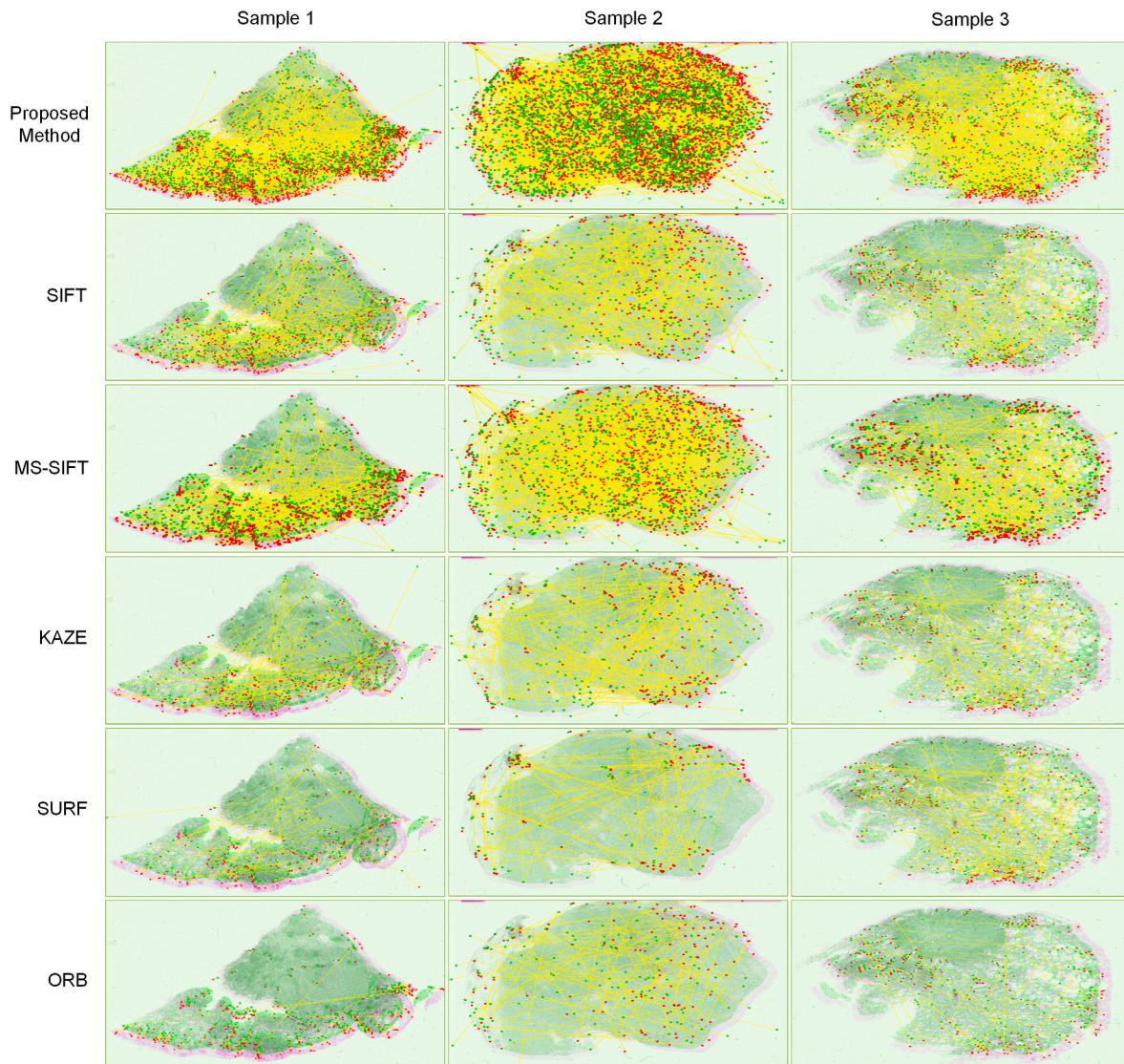


Fig. 10. Registration estimation performance comparison in terms of detected features where green and red dots represent the detected keypoints of reference image and subject image respectively and the yellow lines are the matched keypoints.

Table 3
Statistical matching rate and processing time comparisons of different reference methods.

Method	Detected Points		Matching Points	Mismatching Points	Matching Rate	Run Time (Second)
	Reference Image	Subject Image				
SIFT [25]	450	375	357	18	83%	25.31
SURF [26]	400	341	229	50	67%	28.23
ORB [27]	360	297	268	29	71%	24.16
KAZE [28]	903	612	597	15	84%	22.55
MS-SIFT [30]	1004	719	706	13	90%	21.09
Proposed Method	1379	877	873	4	95%	19.11

Table 4
Quantitative comparisons of different reference methods measuring F-score, Precision, and Recall with respect to cumulative distributions.

Method	F-score (%)	Precision (%)	Recall (%)
SIFT [25]	82.78	83.12	82.44
SURF [26]	69.70	70.19	69.21
ORB [27]	72.19	73.33	71.10
KAZE [28]	84.85	85.07	84.63
MS-SIFT [30]	88.60	88.05	89.15
Proposed Method	96.16	96.55	95.78

Table 5
Registration results comparison of the reference methods in terms of RMSE, MAE, MEE, and STD. The units are in pixel and the best results are highlighted with bold fonts.

Method	RMSE	MAE	MEE	STD
SIFT [25]	0.6103	26.14	3.61	4.935
SURF [26]	1.5190	56.23	16.59	8.237
ORB [27]	0.9441	28.51	6.94	5.844
KAZE [28]	0.5815	24.22	4.03	4.361
MS-SIFT [30]	0.5432	23.08	3.10	4.190
Proposed Method	0.4631	20.09	2.05	3.211

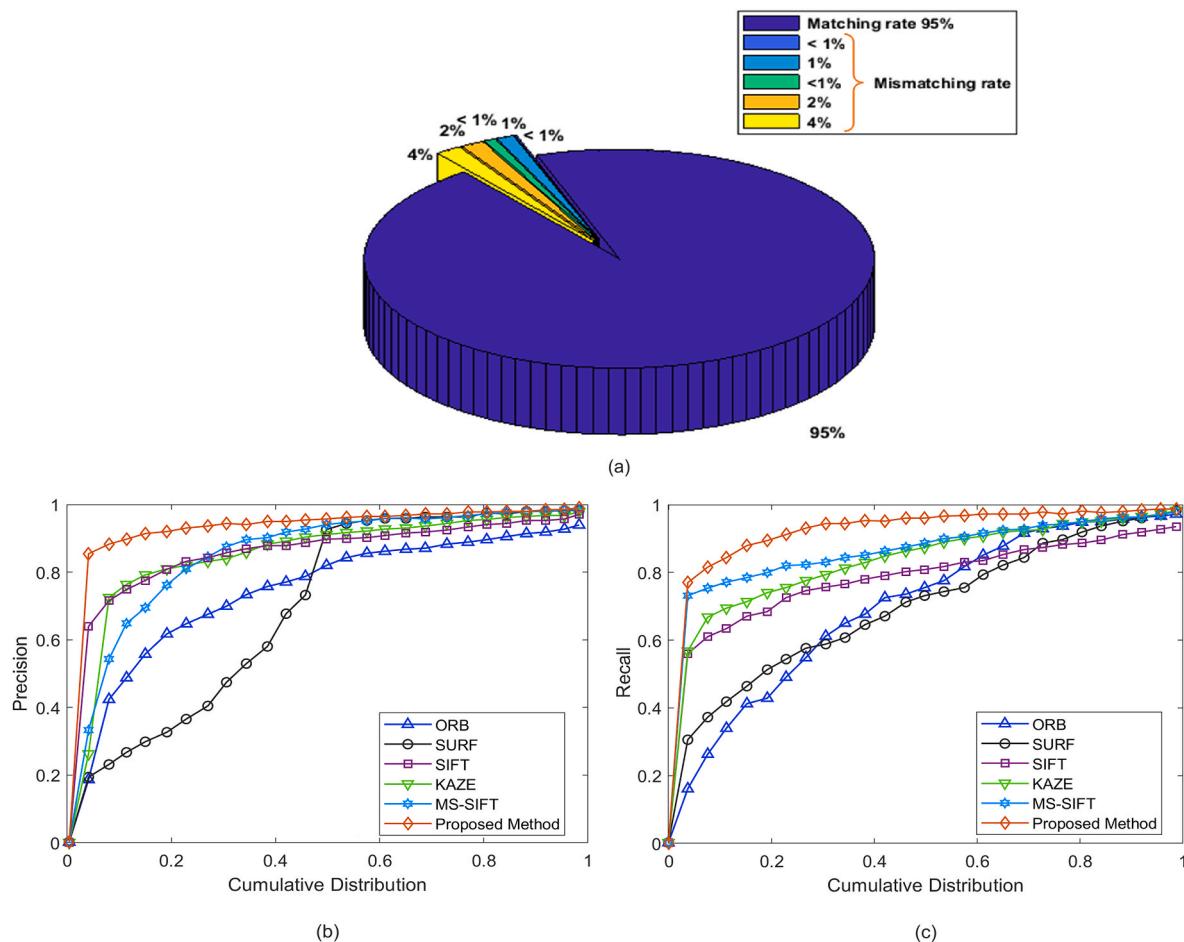


Fig. 11. (a) Graphical representation (matching and mismatching rates of our proposed method), (b) Precision in terms of cumulative distributions, (c) Recall in terms of cumulative distributions.

significant effort to find more robust solution for the whole slide image registration challenges. There is not any method that developed explicitly for the histological image registration challenges taking the advantage of high resolution with full color information. To the best of our knowledge, this is the first computational pathology research to automatically extract the multi-stained features working on large scale whole slide image avoiding patch-based registration.

Recently different deep learning based image registration techniques have been developed mostly suitable for the other domain apart from our research [42,47–49]. They come with high computational costs which limits the applicability of these techniques in particular image analysis systems. Some deep learning architectures are proposed based on different classical methods. Our proposed method might be integrated to design hybrid approaches in account of taking full advantage of deep and local features which we have experimentally proved in our experiment. As for the two steps method, two kinds of features are redundant which increases the complexity with comparatively high processing time requirements. Though our proposed method with deep features comparatively reduces these complexities we still need to work on the advanced algorithm to explore complexity reduction.

However, the most time consuming part of our experiment was to ensure objectivity by checking each putative correspondence to be correct or false as ground truth manually. Before conducting any of our experiments we checked each putative match to be true or false in each image pair. The annotated images with landmarks were determined by comparing the details at the manually segmented structures that was assigned by the pathologists using Aperio ImageScope and ImageJ. The quantitative experiment is conducted on the manually selected landmarks. These annotated landmarks were sufficient for our experiment,

but it is still difficult to identify the landmarks in large areas without salient corresponding structures. On the other hand, the preparation of this type of dataset that we used in our experiment is very hard and time consuming with a lot of labor [50]. Hence, another limitation of our research is that, we have only tested our proposed method on 35 stained whole slide image pairs of lung cancer (adenocarcinoma) samples. In particular, deep learning networks are complex with a huge number of training parameters having interaction difficulties with a single layer and lack of interpretability. Deep learning models are often seen as a black box that does not explain the exact and trustful predictions in a way that human can understand [50]. In digital pathology one of the biggest challenges of using deep learning is the lack of labeled information at pixel-level for segmentation tasks or patch-level for classification tasks. Because it is hard to label a huge number of whole slide images and a lot of labor with expert pathologists is required [5,51]. Moreover, it is still challenging to apply deep neural networks in large scaled high-resolution images (e.g. 40X magnification) for the limited size of the network and large computational memory with high processing time requirements.

Our observation is that none of these methods including our proposal would be probably accurate and robust enough for the routine automatic clinical trial, but this might be a good initialization to find the best solution. We would recommend addressing future research to create more robust algorithm that would be capable of handling full resolution by avoiding different whole slide image registration challenges. Converting multi-modal to a single modal image and then recombining back to multi-modal image might be an extension of our work to solve the complexity of extracting multi-stained feature points for better registration performance.

5. Conclusion

In this study, we address the advantages of digital pathology that offer the possibility of automated image analysis to assist on tasks that are manually done by a pathologist. Moreover, there is a growing interest in applying artificial intelligence to support different types of cancer diagnostics and research. In contrast to traditional pathology, digitalized images can easily establish image databases allowing collaborations and innovations. Also, minimizing the differences in inter-laboratory image properties will be required for efficient cooperation between diagnostic institutions. However, there is a variety of whole slide imaging scanners available on the market. Although the digital pathology systems and devices are calibrated in sites, potential inter-laboratory differences, variety of proprietary file formats and scanner differences. More in deep, analysis should be undertaken into the cause of color variation, artifacts, high-frequency noise, and variation on pixel intensities to take full advantage of the possibilities of computational pathology. One of the challenges with analyzing and comparing pathological images automatically from different scanners is the whole slide image registration. The challenges are related to large file sizes, the nature of the histological section, and the effects of the tissue preparation.

In this paper, we propose a new gradient computational registration method including a robust keypoint matching algorithm that combines scale, keypoint orientation, and position to increase the correct matching number of keypoints. The method is robust in the transformation of complex nonlinear intensity, which would allow further studies of the inherent laboratory to laboratory variations. This paper also demonstrates the capacities of registration utilizing a large tissue image with different staining variations. The affectability of SIFT, MS-SIFT, KAZE, and SURF methods are investigated. ORB method is also compared, and it is seen from the experiment that, this method has less robustness compared to our method in hematoxylin-eosin stained whole slide image analysis. To evaluate the performance, experiments are conducted on a dataset including 35 lung adenocarcinoma tissue sample whole slide image pairs. The experimental evaluation shows that our proposed method has high F-score, precision, recall values, and comparatively low root mean square error, maximum error, median error, and standard deviation values compared to the reference methods. It means our proposed method is consistently better with superior performance in registration of the challenging stained whole slide images and provides better accuracy in the common information

extraction by choosing the correct thresholding parameters.

Particularly, our algorithm performs well on stained histological images with overcoming the dissimilarities misaligned risk, which will be beneficial in further tissue whole slide image analysis in digital pathology. Most significantly, our proposed method might potentially be applied to image registration tasks in other fields apart from digital pathology. However, our future research is to work on deep convolutional pipeline and address the problem of multi-stained feature matching and registration with full resolution of histopathology images.

Ethical approval

The ethical committee of Northern Finland Biobank Borealis approved this research including all tissue samples.

Declaration of competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Md. Ziaul Hoque: Conceptualization, Methodology, Software, Validation, Visualization, Writing – original draft. **Anja Keskinarkaus:** Supervision, Data Curation, Investigation, Writing – review & editing. **Pia Nyberg:** Resources, Writing – review & editing. **Taneli Mattila:** Image Selection, Writing – review & editing. **Tapio Seppänen:** Principal Supervision, Project Administration, Resources, Funding Acquisition, Writing – review & editing.

Acknowledgment

The research work of this paper was conducted with the Physiological Signal Analysis group at Center for Machine Vision and Signal Analysis (CMVS) in the Faculty of Information Technology and Electrical Engineering (ITEE) at the University of Oulu. This research has been financially supported by the Academy of Finland 6 Genesis Flagship (Grant 318927) and Academy of Finland Identifying trajectories of healthy aging via integration of birth cohorts and biobank data (Grant 309112). We are thankful to MD, Ph.D., board-certified pathologist Johanna Mäkinen for selecting the lung HE-slides.

Appendix A

Table A1 represents the detailed statistics (Mean \pm Standard Deviation) of 35 whole slide image pairs driven from RGB channels; detected, matching, and mismatching points with processing time. **Table A2** shows the area matching results of two step registration combining deep and local features.

Table A1

Detailed statistics (Mean \pm Standard Deviation) derived from the separated RGB channels and matching rate of our proposed method over 35 different tissue sample whole slide image pairs.

Tissue Sample Image Pairs	Compute (Mean \pm Standard Deviation) from separated RGB channels						Detected Points		Matching Points	Mismatching Points	Run Time (Second)			
	Reference Image (oys_aperio.svs)			Subject Image (oys_hama.ndpi)			Reference Image	Subject Image						
	Channel 1	Channel 2	Channel 3	Channel 1	Channel 2	Channel 3								
Image Pair 1	0.94 \pm 0.05	0.57 \pm 0.05	0.65 \pm 0.03	0.95 \pm 0.03	0.54 \pm 0.04	0.60 \pm 0.04	1379	877	873	4	19.11			
Image Pair 2	0.95 \pm 0.05	0.51 \pm 0.05	0.68 \pm 0.03	0.96 \pm 0.04	0.58 \pm 0.03	0.62 \pm 0.03	1345	894	889	5	18.35			
Image Pair 3	0.94 \pm 0.06	0.71 \pm 0.05	0.72 \pm 0.03	0.96 \pm 0.05	0.57 \pm 0.06	0.62 \pm 0.04	1311	893	889	4	19.01			
Image Pair 4	0.95 \pm 0.05	0.54 \pm 0.04	0.59 \pm 0.04	0.96 \pm 0.04	0.53 \pm 0.03	0.59 \pm 0.04	1357	889	886	3	17.05			
Image Pair 5	0.94 \pm 0.06	0.58 \pm 0.04	0.71 \pm 0.03	0.96 \pm 0.03	0.54 \pm 0.03	0.65 \pm 0.03	1327	811	807	4	16.90			
Image Pair 6	0.95 \pm 0.05	0.59 \pm 0.05	0.70 \pm 0.02	0.97 \pm 0.04	0.38 \pm 0.03	0.70 \pm 0.02	1278	822	818	4	17.23			
Image Pair 7	0.95 \pm 0.05	0.64 \pm 0.05	0.65 \pm 0.05	0.97 \pm 0.04	0.49 \pm 0.03	0.57 \pm 0.03	1329	807	803	3	19.13			
Image Pair 8	0.95 \pm 0.05	0.58 \pm 0.04	0.49 \pm 0.04	0.95 \pm 0.06	0.53 \pm 0.05	0.51 \pm 0.04	1407	901	896	5	19.08			
Image Pair 9	0.95 \pm 0.11	0.58 \pm 0.04	0.59 \pm 0.04	0.95 \pm 0.05	0.44 \pm 0.04	0.63 \pm 0.03	1361	834	828	6	18.45			

(continued on next page)

Table A1 (continued)

Tissue Sample Image Pairs	Compute (Mean \pm Standard Deviation) from separated RGB channels						Detected Points		Matching Points	Mismatching Points	Run Time (Second)			
	Reference Image (oys_aperio.svs)			Subject Image (oys_hama.ndpi)			Reference Image	Subject Image						
	Channel 1	Channel 2	Channel 3	Channel 1	Channel 2	Channel 3								
Image Pair 10	0.96 \pm 0.09	0.56 \pm 0.05	0.69 \pm 0.03	0.96 \pm 0.05	0.53 \pm 0.03	0.64 \pm 0.02	1394	861	857	4	18.04			
Image Pair 11	0.97 \pm 0.04	0.40 \pm 0.04	0.69 \pm 0.04	0.97 \pm 0.03	0.54 \pm 0.04	0.61 \pm 0.03	1335	856	851	5	18.01			
Image Pair 12	0.96 \pm 0.04	0.55 \pm 0.03	0.71 \pm 0.02	0.97 \pm 0.04	0.43 \pm 0.03	0.63 \pm 0.03	1419	914	899	5	18.62			
Image Pair 13	0.95 \pm 0.04	0.50 \pm 0.06	0.70 \pm 0.03	0.96 \pm 0.03	0.44 \pm 0.04	0.56 \pm 0.03	1459	955	948	7	19.55			
Image Pair 14	0.97 \pm 0.04	0.42 \pm 0.03	0.65 \pm 0.02	0.97 \pm 0.04	0.56 \pm 0.03	0.73 \pm 0.04	1421	897	893	4	20.07			
Image Pair 15	0.96 \pm 0.04	0.54 \pm 0.05	0.60 \pm 0.04	0.97 \pm 0.03	0.51 \pm 0.03	0.65 \pm 0.03	1392	859	855	4	17.52			
Image Pair 16	0.97 \pm 0.04	0.50 \pm 0.03	0.61 \pm 0.02	0.97 \pm 0.04	0.57 \pm 0.04	0.63 \pm 0.03	1424	844	841	3	17.38			
Image Pair 17	0.96 \pm 0.05	0.48 \pm 0.04	0.64 \pm 0.03	0.97 \pm 0.03	0.54 \pm 0.03	0.50 \pm 0.04	1355	848	843	5	17.91			
Image Pair 18	0.96 \pm 0.05	0.41 \pm 0.03	0.62 \pm 0.02	0.97 \pm 0.03	0.57 \pm 0.05	0.78 \pm 0.03	1406	872	869	3	17.02			
Image Pair 19	0.97 \pm 0.04	0.40 \pm 0.02	0.59 \pm 0.02	0.97 \pm 0.04	0.62 \pm 0.06	0.60 \pm 0.03	1417	913	909	4	18.52			
Image Pair 20	0.96 \pm 0.04	0.47 \pm 0.06	0.58 \pm 0.04	0.97 \pm 0.03	0.63 \pm 0.04	0.81 \pm 0.02	1341	823	819	4	19.11			
Image Pair 21	0.96 \pm 0.04	0.51 \pm 0.04	0.64 \pm 0.03	0.96 \pm 0.05	0.58 \pm 0.04	0.49 \pm 0.04	1460	894	888	6	19.00			
Image Pair 22	0.96 \pm 0.04	0.52 \pm 0.07	0.65 \pm 0.03	0.97 \pm 0.04	0.55 \pm 0.03	0.70 \pm 0.02	1278	756	751	5	18.25			
Image Pair 23	0.95 \pm 0.05	0.61 \pm 0.06	0.56 \pm 0.04	0.97 \pm 0.04	0.57 \pm 0.03	0.59 \pm 0.03	1398	820	813	7	18.47			
Image Pair 24	0.96 \pm 0.04	0.51 \pm 0.06	0.61 \pm 0.03	0.96 \pm 0.04	0.53 \pm 0.04	0.54 \pm 0.04	1508	905	901	4	18.24			
Image Pair 25	0.97 \pm 0.04	0.49 \pm 0.05	0.78 \pm 0.02	0.97 \pm 0.04	0.44 \pm 0.03	0.63 \pm 0.03	1302	808	805	3	19.15			
Image Pair 26	0.97 \pm 0.04	0.54 \pm 0.05	0.79 \pm 0.02	0.97 \pm 0.03	0.58 \pm 0.02	0.64 \pm 0.02	1522	966	961	5	19.38			
Image Pair 27	0.97 \pm 0.04	0.53 \pm 0.05	0.69 \pm 0.02	0.97 \pm 0.04	0.56 \pm 0.03	0.60 \pm 0.03	1318	827	822	5	18.09			
Image Pair 28	0.96 \pm 0.03	0.59 \pm 0.04	0.49 \pm 0.04	0.97 \pm 0.02	0.48 \pm 0.03	0.63 \pm 0.03	1364	871	865	6	18.14			
Image Pair 29	0.97 \pm 0.03	0.49 \pm 0.05	0.77 \pm 0.02	0.97 \pm 0.03	0.58 \pm 0.03	0.69 \pm 0.03	1501	928	923	5	19.05			
Image Pair 30	0.97 \pm 0.03	0.58 \pm 0.04	0.70 \pm 0.02	0.97 \pm 0.03	0.56 \pm 0.02	0.73 \pm 0.02	1282	843	839	4	19.61			
Image Pair 31	0.96 \pm 0.04	0.71 \pm 0.04	0.70 \pm 0.03	0.97 \pm 0.02	0.57 \pm 0.03	0.63 \pm 0.03	1320	858	854	4	19.43			
Image Pair 32	0.96 \pm 0.05	0.50 \pm 0.06	0.62 \pm 0.03	0.96 \pm 0.03	0.62 \pm 0.04	0.67 \pm 0.04	1462	917	912	5	19.10			
Image Pair 33	0.96 \pm 0.05	0.57 \pm 0.07	0.60 \pm 0.04	0.97 \pm 0.04	0.65 \pm 0.03	0.71 \pm 0.03	1378	906	901	5	19.26			
Image Pair 34	0.93 \pm 0.07	0.50 \pm 0.06	0.53 \pm 0.05	0.95 \pm 0.05	0.55 \pm 0.04	0.59 \pm 0.04	1400	912	908	4	18.66			
Image Pair 35	0.92 \pm 0.08	0.39 \pm 0.06	0.45 \pm 0.06	0.94 \pm 0.05	0.53 \pm 0.04	0.48 \pm 0.04	1390	894	890	4	17.69			

Table A2

Experimental results (two step registration combining deep and local features) comparison of different methods in terms of RMSE, NOCC, ROCC, and RT.

Method	Performance results for image pair 1				Performance results for image pair 2				Performance results for image pair 3			
	RMSE	NOCC	ROCC	RT (S)	RMSE	NOCC	ROCC	RT (S)	RMSE	NOCC	ROCC	RT (S)
Manual	2.382	30	*	*	1.784	30	*	*	2.031	30	*	*
SIFT [25]	0.651	10	0.030	1.170	0.679	13	0.051	1.560	0.648	18	0.055	1.241
SURF [26]	0.680	18	0.019	14.97	0.644	6	0.012	2.130	0.637	6	0.008	2.320
MS-SIFT [30]	0.632	62	0.241	18.41	0.621	56	0.204	18.11	0.630	44	0.226	19.50
SAR-SIFT [32]	0.672	59	0.217	105.2	0.715	29	0.209	71.66	0.662	40	0.206	118.1
Proposed Method (MF-SIFT)	0.581	68	0.273	16.11	0.603	61	0.221	15.36	0.562	55	0.290	16.02
Area matching (CNN + SIFT)	0.588	50	0.276	18.23	0.637	39	0.255	16.92	0.594	61	0.399	16.88
Area matching (CNN + MS-SIFT)	0.613	68	0.257	33.50	0.571	45	0.264	32.10	0.531	64	0.306	30.10
Area matching (CNN + SURF)	0.601	55	0.125	23.68	0.611	43	0.132	19.13	0.641	58	0.141	19.83
Area matching (CNN + SAR-SIFT)	0.630	72	0.260	114.5	0.629	40	0.226	83.55	0.602	72	0.314	129.5
Area matching (CNN + MF-SIFT)	0.510	76	0.285	25.31	0.531	52	0.290	24.01	0.501	75	0.320	28.10

References

- [1] Y. Ou, H. Akbari, M. Bilello, X. Da, C. Davatzikos, Comparative evaluation of registration algorithms in different brain databases with varying difficulty: results and insights, *IEEE Trans. Med. Imag.* 33 (10) (2014) 2039–2065.
- [2] D.L. Hill, P.G. Batchelor, M. Holden, D.J. Hawkes, Medical image registration, *Phys. Med. Biol.* 46 (3) (2001) R1.
- [3] B. Glocker, A. Sotiras, N. Komodakis, N. Paragios, Deformable medical image registration: setting the state of the art with discrete methods, *Annu. Rev. Biomed. Eng.* 13 (2011) 219–244.
- [4] N. Alsubaie, N. Trahearn, S.E.A. Raza, D. Snead, N.M. Rajpoot, Stain deconvolution using statistical analysis of multi-resolution stain colour representation, *PLoS One* 12 (1) (2017) e0169875.
- [5] T.A.A. Tosta, P.R. da Faria, L.A. Neves, M.Z. do Nascimento, Computational normalization of H&E stained histological images: progress, challenges and future potential, *Artif. Intell. Med.* 95 (2019) 118–132.
- [6] M.N. Gurcan, L.E. Boucheron, A. Can, A. Madabhushi, N.M. Rajpoot, B. Yener, Histopathological image analysis: a review, *IEEE Rev. Biomed. Eng.* 2 (2009) 147–171.
- [7] M.Z. Hoque, A. Keskinarkaus, P. Nyberg, T. Seppänen, Retinex model based stain normalization technique for whole slide image analysis, *Comput. Med. Imag. Graph.* 90 (2021) 101901.
- [8] J. Pichat, J.E. Iglesias, T. Yousry, S. Ourselin, M. Modat, A survey of methods for 3D histology reconstruction, *Med. Image Anal.* 46 (2018) 73–105.
- [9] J. Pichat, M. Modat, T. Yousry, S. Ourselin, A multi-path approach to histology volume reconstruction, in: Proc. IEEE 12th Int. Symp. Biomed. Imag. (ISBI), 2015, pp. 1280–1283.
- [10] Y. Song, D. Treanor, A.J. Bulpitt, D.R. Magee, 3D reconstruction of multiple stained histology images, *J. Pathol. Inf.* 4 (Suppl) (2013).
- [11] A. du Bois d'Aische, M.D. Craene, X. Geets, V. Gregoire, B. Macq, S.K. Warfield, Efficient multi-modal dense field non-rigid registration: alignment of histological and section images, *Med. Image Anal.* 9 (6) (2005) 538–546.
- [12] D.F.G. Obando, A. Frafjord, I. Oynebraten, A. Corthay, J.-C. Olivo-Marin, V. Meas-Yedid, Multi-staining registration of large histology images, in: Proc. IEEE 14th Int. Symp. Biomed. Imag. (ISBI), 2017, pp. 345–348.
- [13] O. Déniz, D. Toomey, C. Conway, G. Bueno, Multi-stained whole slide image alignment in digital pathology, *Proc. SPIE* 9420 (2015), 94200Z.
- [14] N. Trahearn, D. Epstein, I. Cree, D. Snead, N. Rajpoot, Hyperstain inspector: a framework for robust registration and localised co-expression analysis of multiple whole-slide images of serial histology sections, *Sci. Rep.* 7 (1) (2017) 1–3.
- [15] S. Saalfeld, A. Cardona, V. Hartenstein, P. Tomancak, As-rigid-as-possible mosaicking and serial section registration of large ssTEM datasets, *Bioinformatics* 26 (12) (2010) i57–i63.
- [16] L. Gupta, B.M. Klinkhamer, P. Boor, D. Merhof, M. Gadermayr, Stain independent segmentation of whole slide images: a case study in renal histology, in: Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI), 2018, pp. 1360–1364.

- [17] J. Borovec, J. Kybic, I. Arganda-Carreras, D.V. Sorokin, G. Bueno, A.V. Khvostikov, S. Bakas, I. Eric, C. Chang, S. Heldmann, et al., ANHIR: automatic non-rigid histological image registration challenge, *IEEE Trans. Med. Imag.* 39 (10) (2020) 3042–3052.
- [18] J. Lotz, J. Olesch, B. Müller, T. Polzin, P. Galuschka, J. Lotz, S. Heldmann, H. Laue, M. González-Vallinas, A. Warth, et al., Patch-based nonlinear image registration for gigapixel whole slide images, *IEEE (Inst. Electr. Electron. Eng.) Trans. Biomed. Eng.* 63 (9) (2015) 1812–1819.
- [19] X. Moles Lopez, P. Barbot, Y.-R. Van Eycke, L. Verset, A.-L. Tré-pant, L. Larbanoix, I. Salmon, C. Decaestecker, Registration of whole immunohistochemical slide images: an efficient way to characterize biomarker colocalization, *J. Am. Med. Inf. Assoc.* 22 (1) (2015) 86–99.
- [20] N. Roberts, D. Magee, Y. Song, K. Brabazon, M. Shires, D. Crellin, N.M. Orsi, R. Quirke, P. Quirke, D. Treanor, Toward routine use of 3D histopathology as a research tool, *Am. J. Pathol.* 180 (5) (2012) 1835–1842.
- [21] B.J. Rossetti, F. Wang, P. Zhang, G. Teodoro, D.J. Brat, J. Kong, Dynamic registration for gigapixel serial whole slide images, in: *IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, 2017, IEEE, 2017, pp. 424–428.
- [22] Y. Zheng, Z. Jiang, H. Zhang, F. Xie, J. Shi, C. Xue, Adaptive color deconvolution for histological WSI normalization, *Comput. Methods Progr. Biomed.* 170 (2019) 107–120.
- [23] J. Senthilnath, S. Omkar, V. Mani, T. Karthikeyan, Multi objective discrete particle swarm optimization for multi sensor image alignment, *Geosci. Rem. Sens. Lett.* IEEE 10 (5) (2013) 1095–1099.
- [24] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, P. Suetens, Multimodality image registration by maximization of mutual information, *IEEE Trans. Med. Imag.* 16 (2) (1997) 187–198.
- [25] D.G. Lowe, Distinctive image features from scale-invariant keypoints, *Int. J. Comput. Vis.* 60 (2) (2004) 91–110.
- [26] H. Bay, A. Ess, T. Tuytelaars, L. Van Gool, Speeded-up robust features, *Comput. Vis. Image Understand.* 110 (3) (2008) 346–359.
- [27] E. Rublee, V. Rabaud, K. Konolige, G. Bradski, ORB: an efficient alternative to SIFT or SURF, in: *International Conference on Computer Vision*, 2011, IEEE, 2011, pp. 2564–2571.
- [28] P.F. Alcantarilla, A. Bartoli, A.J. Davison, KAZE features, in: *European Conference on Computer Vision*, Springer, 2012, pp. 214–227.
- [29] Y. Li, L. Liu, L. Wang, D. Li, M. Zhang, Fast SIFT algorithm based on Sobel edge detector, in: *2012 2nd International Conference on Consumer Electronics, Communications and Networks (CECNet)*, IEEE, 2012, pp. 1820–1823.
- [30] B. Kupfer, N.S. Netanyahu, I. Shimshoni, An efficient SIFT-based mode-seeking algorithm for sub-pixel registration of remotely sensed images, *Geosci. Rem. Sens. Lett.* IEEE 12 (2) (2014) 379–383.
- [31] A. Sedaghat, M. Mokhtarzade, H. Ebadi, Uniform robust scale-invariant feature matching for optical remote sensing images, *IEEE Trans. Geosci. Rem. Sens.* 49 (11) (2011) 4516–4527.
- [32] F. Dellinger, J. Delon, Y. Gousseau, J. Michel, F. Tupin, SAR-SIFT: ASIFT-like algorithm for SAR images, *IEEE Trans. Geosci. Rem. Sens.* 53 (1) (2014) 453–466.
- [33] M. Gong, S. Zhao, L. Jiao, D. Tian, S. Wang, A novel coarse-to-fine scheme. for automatic image registration based on SIFT and mutual information, *IEEE Trans. Geosci. Rem. Sens.* 52 (7) (2013) 4328–4338.
- [34] A. Gray, A. Wright, P. Jackson, M. Hale, D. Treanor, Quantification of histochemical stains using whole slide imaging: development of a method and demonstration of its usefulness in laboratory quality control, *J. Clin. Pathol.* 68 (3) (2015) 192–199.
- [35] J. Sánchez, N. Monzón, A. Salgado De La Nuez, An Analysis and Implementation of the Harris Corner Detector, *Image Processing Online*, 2018.
- [36] Y. Wu, W. Ma, M. Gong, L. Su, L. Jiao, A novel point-matching algorithm based on fast sample consensus for image registration, *Geosci. Rem. Sens. Lett.* IEEE 12 (1) (2014) 43–47.
- [37] P. Saint-Marc, J.-S. Chen, G. Medioni, Adaptive smoothing: a general tool for early vision, *IEEE Trans. Pattern Anal. Mach. Intell.* 13 (1991) 514–529, 06.
- [38] K. Mikolajczyk, C. Schmid, A performance evaluation of local descriptors, *IEEE Trans. Pattern Anal. Mach. Intell.* 27 (10) (2005) 1615–1630.
- [39] H.H. Chang, G.L. Wu, M.H. Chiang, Remote sensing image registration based on modified SIFT and feature slope grouping, *Geosci. Rem. Sens. Lett.* IEEE 16 (9) (2019) 1363–1367.
- [40] A. Fahim, A. Salem, F.A. Torkey, M. Ramadan, An efficient enhanced k-means clustering algorithm, *J. Zhejiang Univ. - Sci.* 7 (10) (2006) 1626–1633.
- [41] M.A. Fischler, R.C. Bolles, Random Sample Consensus: a paradigm for model fitting with applications to image analysis and automated cartography, *Commun. ACM* 24 (6) (1981) 381–395.
- [42] W. Ma, J. Zhang, Y. Wu, L. Jiao, H. Zhu, W. Zhao, A novel two-step registration method for remote sensing images based on deep and local features, *IEEE Trans. Geosci. Rem. Sens.* 57 (7) (2019) 4834–4843.
- [43] C.A. Schneider, W.S. Rasband, K.W. Eliceiri, NIH Image to ImageJ: 25 years of image analysis, *Nat. Methods* 9 (7) (2012) 671–675.
- [44] V. Della Mea, G.L. Baroni, D. Pilutti, C. Di Loreto, SlideJ, An ImageJ plugin for automated processing of whole slide images, *PLoS One* 12 (7) (2017) e0180540.
- [45] Y. Song, D. Treanor, A.J. Bulpitt, N. Wijayathunga, N. Roberts, R. Wilcox, D. R. Magee, Unsupervised content classification based nonrigid registration of differently stained histology images, *IEEE (Inst. Electr. Electron. Eng.) Trans. Biomed. Eng.* 61 (1) (2013) 96–108.
- [46] W. Ma, Z. Wen, Y. Wu, L. Jiao, M. Gong, Y. Zheng, L. Liu, Remote sensing image registration with modified SIFT and enhanced feature matching, *Geosci. Rem. Sens. Lett.* IEEE 14 (1) (2016) 3–7.
- [47] S. Miao, Z.J. Wang, R. Liao, A CNN regression approach for real-time 2d/3d registration, *IEEE Trans. Med. Imag.* 35 (5) (2016) 1352–1363.
- [48] D. DeTone, T. Malisiewicz, A. Rabinovich, Superpoint: self-supervised interest point detection and description, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 224–236.
- [49] A. Barroso-Laguna, E. Riba, D. Ponsa, K. Mikolajczyk, Key. net: keypoint detection by handcrafted and learned CNN filters, in: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 5836–5844.
- [50] M. Salvi, U.R. Acharya, F. Molinari, K.M. Meiburger, The impact of pre-and post-image processing techniques on deep learning frameworks: a comprehensive review for digital pathology image analysis, *Comput. Biol. Med.* 128 (2021) 104129.
- [51] D. Komura, S. Ishikawa, Machine learning methods for histopathological image analysis, *Comput. Struct. Biotechnol. J.* 16 (2018) 34–42.