A Sub-pixel Accurate Quantification of Joint Space Narrowing Progression in Rheumatoid Arthritis

Yafei Ou, Prasoon Ambalathankandy, *Member, IEEE*, Ryunosuke Furuya, Seiya Kawada, Tianyu Zeng, Yujie An, Tamotsu Kamishima, Kenichi Tamura, and Masayuki Ikebe, *Member, IEEE*

***Abstract*— Rheumatoid arthritis (RA) is a chronic autoimmune disease that primarily affects peripheral synovial joints, like fingers, wrists and feet. Radiology plays a critical role in the diagnosis and monitoring of RA. Limited by the current spatial resolution of radiographic imaging, joint space narrowing (JSN) progression of RA for the same reason above can be less than one pixel per year with universal spatial resolution. Insensitive monitoring of JSN can hinder the radiologist/rheumatologist from making a proper and timely clinical judgment. In this paper, we propose a novel and sensitive method that we call partial image phase-only correlation which aims to automatically quantify JSN progression in the early RA. The majority of the current literature utilizes the mean error, root-mean-square deviation and standard deviation to report the accuracy at pixel level. Our work measures JSN progression between a baseline and its follow-up finger joint images by using the phase spectrum in the frequency domain. Using this study, the mean error can be reduced to 0.0130 mm when applied to phantom radiographs with ground truth, and 0.0519 mm standard deviation for clinical radiography. With the sub-pixel accuracy far beyond usual manual measurements, we are optimistic that the proposed work is a promising scheme for automatically quantifying JSN progression.**

***Index Terms*— Rheumatoid Arthritis, Frequency Domain Analysis, Joint Space Narrowing, Phantom Imaging, Radiology, Computer-aided Diagnosis.**

I. INTRODUCTION

Rheumatoid arthritis (RA) is a progressive, chronic autoimmune disease characterized by synovitis that can ultimately cause deformities and ankylosis in peripheral synovial joints and impair the movement and flexibility of digits, and as well as the patient's whole hand. The major radiographic changes on hand, wrist and feet joints are cartilage damage and bone destruction (like bone erosion and joint space narrowing (JSN)). Those damage and destruction typically lead to painful joints, progressive joint destruction, deformity, followed by functional limitation and severe disability [1], [2]. There are substantial evidence that RA can be managed in a low level of disease activity and clinical remission with disease-modifying antirheumatic drugs [3], [4]. Early diagnosis by precise quantification of subtle radiographic changes is essential for successful treatment, as it can improve outcomes and effectively manage the progression of RA [3], [4].

Radiology plays a crucial role in diagnosis and monitoring of RA. Clinical radiologist/rheumatologist can assess the radiographic progression of RA by using the Sharp/van der Heijde scoring method (SvdH). This method relies on scoring of the radiographies by subjectively assessing JSN and bone erosion of 38 hand or foot joints [5]. As one of the most important indicators for the diagnosis and monitoring of RA, the joint space has always attracted extensive attention of researchers. In recent years, researchers have invested great efforts to study automatic quantification of joint space in RA [6]–[13]. The JSN progression quantification pipline in radiographs is performed in two steps; joint position detection and joint space quantification.

1) Related works about finger joint position detection: The earliest studies about finger joint location detection were based on using pixel information. Those algorithms extracted the finger midlines based on ridge detection, thus, finger joint location can be detected according to the gradient or intensity information of finger midline [8]. However, these method may break finger midline at the metacarpophalangeal (MCP) joint because of decrease in bone density. This may lead to mismatch of joint position for the following reasons: (i) Bone overlap caused by finger bending in the vertical plane. (ii) Marginal density decrease caused by ankylosis or complete luxation [8].

In recent years, machine learning (ML) based methods have become a very important tool to solve complex medical image processing tasks [14], [15]. For finger joint detection, there are some ML-based studies utilizing key point detection

Manuscript received 13 February 2022; revised 9 September 2022; accepted 24 October 2022. This work was supported in part by the Japan Society for the Promotion of Science (JSPS) Grants-in-Aid for Scientific Research (KAKENHI) under Grants 18H05288 and 21K07611, and in part by JST SPRING under Grant JPMJSP2119. *(Corresponding author: Masayuki Ikebe.)*

Yafei Ou, Prasoon Ambalathankandy, Ryunosuke Furuya, Seiya Kawada and Masayuki Ikebe are with the Research Center For Integrated Quantum Electronics, Hokkaido University, Sapporo 060-0813, Japan (e-mail: yafei.ou.x5@elms.hokudai.ac.jp; prasoon.ak@ist.hokudai.ac.jp; huruya@rciqe.hokudai.ac.jp; kawada@rciqe.hokudai.ac.jp; ikebe@ist.hokudai.ac.jp).

Tianyu Zeng, Yujie An and Tamotsu Kamishima are with the Graduate School of Health Sciences, Hokkaido University, Sapporo 060-0812, Japan (e-mail: zengtianyu1223@yahoo.co.jp; anyujie19960922@kamiken.net; e-mail: ktamotamo2@hs.hokudai.ac.jp).

Kenichi Tamura is with Department of Mechanical Engineering, College of Engineering, Nihon University, Koriyama 963-8642, Japan (e-mail: tamura.kennichi@nihon-u.ac.jp).

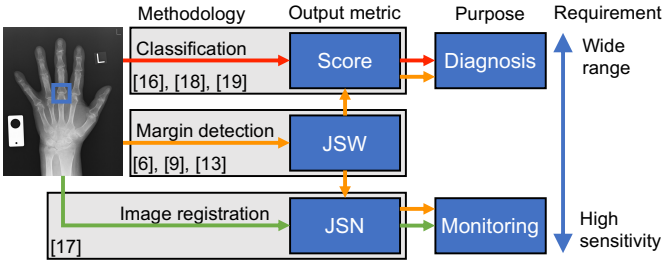


Fig. 1. A schematic diagram of the classification of joint space quantification works in the computer-aided diagnosis and monitoring of RA. According to the methodology and its output metric, joint space quantification works can be broadly grouped into three groups, ML-based classification based SvdH scoring method (red lines), margin detection based JSW quantification method (orange lines) and image registration based JSN progression quantification method (green lines, the framework we proposed). JSW can be used to score the SvdH score to qualitative diagnosis, or calculate the JSN to quantitative monitoring. Diagnosis methodologies should be used in all stages of RA, which have higher requirement on the detectable range. Meanwhile, monitoring methodologies require higher timeliness and sensitivity.

for convolutional neural network (CNN) [7], [9], support vector machine (SVM) [12] and Haar-like adaptive boosting (AdaBoost) [10].

2) Literature survey of joint space quantification in RA: As shown in Fig. 1, according to the nature of purpose, previous works on joint space quantification in RA can be grouped into two groups, qualitative diagnosis and quantitative monitoring. Diagnosis is a qualitative judgment of the RA stage, usually based on the absolute width of the joint space. Monitoring is the basis for RA patient drug management, which requires high sensitivity quantitative quantification, usually based on the relative narrowing of the joint space.

In the literature, the earliest joint space quantification method detects the upper and lower bone margins to measure the joint space width (JSW) (orange lines in Fig. 1). According to the margin detection method, they can be broadly grouped into two groups; supervised ML-based [6], [7] and image features based [8], such as intensity, gradient, derivative or differential.

As shown in Fig. 1, margin detection based JSW quantification methods can combine both qualitative diagnosis and quantitative monitoring. According to the quantified JSW combined with the SvdH scoring standard, a qualitative diagnosis of the RA stage can be obtained. And the JSN can be obtained by calculating the difference of the JSW during the two imaging to realize the quantitative monitoring of RA. However, margin detection based JSW quantification works have three main limitations: (i) Margin detection based studies [6]–[8] can best achieve only pixel-level accuracy (please see Section IV-B for more details). Furthermore, limited by the current spatial resolution of radiographic imaging, JSN progression over a period of one year can be less than one pixel. This means that the pixel-level accuracy algorithm requires more time to detect for any changes in the joint space. Nevertheless, this can lead to insensitive monitoring of JSN progression, and this may hinder the radiologist/rheumatologist from making a proper monitoring in the ”window of opportunity” [16]. (ii) Considering that margin detection require clear margin information, it cannot be used in the advanced RA when the bone margin is destroyed [8].

At present, ML-based algorithms have increasingly become the mainstream of computer-aided diagnosis in RA [17]. The popular research direction in joint space quantification is the SvdH scoring method [5] based on ML image classification (red lines in Fig. 1).

ML-based classification can quickly determine the RA condition in the early RA and the advanced RA. Considering that ML-based classification is proven now, these works can achieve very low false negative and false positive rates. Nevertheless, SvdH scoring standard with only five levels limits the sensitivity and timeliness of the tool. If it is not feasible to increase the number of levels in the scoring standard, that would make it difficult for radiologist/rheumatologist to make accurate scores for training data. But blindly increasing the number of levels will increase the difficulty and error rate for radiologist/rheumatologist to manually label data. It is precisely because of this conflict that the upper limit of the sensitivity of these algorithms is severely limited. Compared to margin detection based JSW quantification works, these works greatly improves the detectable range at the expense of low sensitivity. And abandon the application of the monitoring and strengthen the qualitative diagnosis in the early RA and the advanced RA.

In our conference paper [10], we proposed a JSN progression quantification method by calculating the relative widths of the joint space based on an image registration method (green lines in Fig. 1).

The experiments in our conference paper [10] show that image registration based JSN progression quantification framework has the potential for higher sensitivity and lower mean error when compare to margin detection based JSW quantification framework and image classification based scoring framework. These advantages indicate that our proposed framework can fill the gap in monitoring the JSN progression with high sensitivity. Thereby it has a broad application prospect in the monitoring of RA. Nevertheless, the image registration based JSN progression quantification framework has two limitations: (i) This framework can only calculate the difference of JSW between the baseline and its follow-up joint image, which is a relative width of joint space. Although it can be used for RA monitoring with high sensitivity, it cannot provide absolute width information of joint space. It needs to cooperate with other algorithms for qualitative diagnosis in RA. (ii) Take the MCP joint as an example, this framework should segment the proximal phalanx bone and metacarpal bone. Segmentation require a certain joint space, and this framework cannot be used in the advanced RA when the joint space completely disappears.

3) Our contributions: Original contribution of this work can be summarized as follows:

i) Describe a detection method for finger midline and joint position.

ii) Propose an image segmentation algorithm to segment joint images.

iii) Present an improved phase-only correlation method named partial image phase-only correlation (PIPOC) to mea-

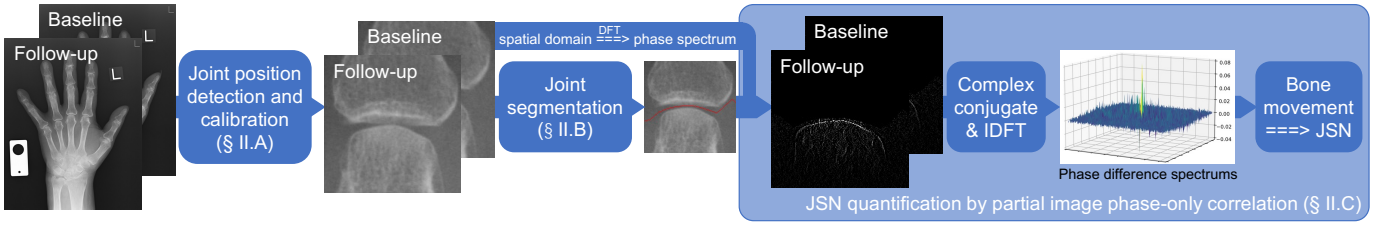


Fig. 2. The overview of our proposed image registration based JSN progression quantification framework. Our framework can be understood in three steps.§ II-A Use image processing algorithms to detect and calibrate joint positions. § II-B Take a MCP joint as an example, segment the proximal phalanx bone and metacarpal bone based on gradient information. § II-C Combining the discrete Fourier transform and the joint segmentation curve, segment the proximal phalanx bone and metacarpal bone in the phase spectrum. Then, Measure the movement of the proximal phalanx bone and metacarpal bone between baseline and follow-up radiographs respectively by calculating the phase difference, thus resulting in JSN progression quantification.

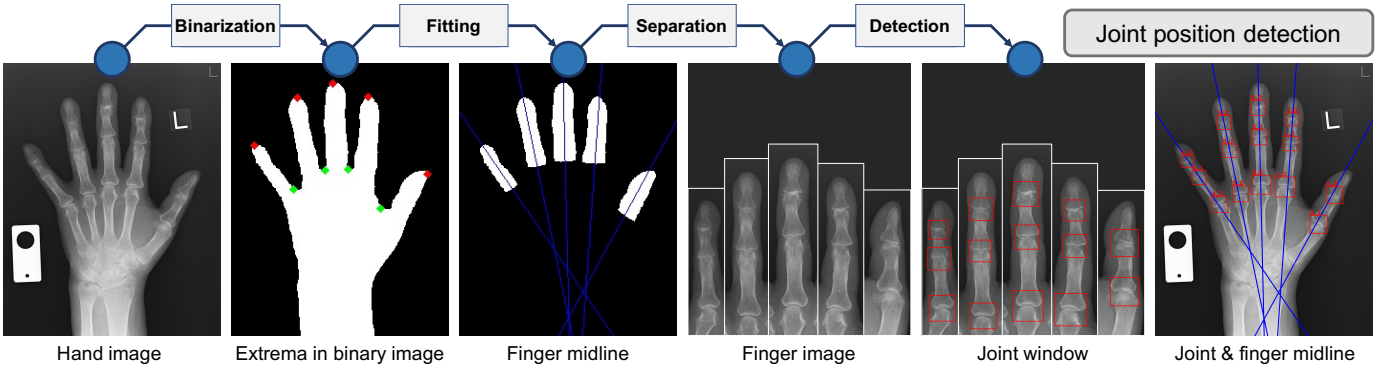


Fig. 3. Schematic overview of joint position detection. The approximate regions of fingers are obtained according to each pair of local maxima (red points) and local minima (green points) in binary image. Then, the finger midlines (blue lines) are calculated by fitting to each region. Finally, an AdaBoost based joint classifier is used to detect the joint positions (red windows).

sure JSN progression in the early RA.

iv) Automate the features listed in (i-iii). Using our method the JSN progression can be measured from a group of input sequential radiographs.

v) The proposed work can achieve sub-pixel accuracy on JSN progression measurement.

This rest of the paper is organized as follows: Section II reports a fully automatic method for the localization of the joint position, and we propose an image registration algorithm (partial image phase-only correlation) for the JSN progression quantification in RA. In Section III, we provide information on the datasets used in this work; including phantom study, and clinical dataset. Section IV, presents the joint position detection results using clinical data and the JSN progression quantification evaluation results for both phantom study and clinical study. Section V presents a detailed discussion with concluding remarks. And an account of the future research directions for computer-aided monitoring in RA, especially the JSN progression quantification based on ML-based image registration.

II. METHODOLOGY

The main objective of the proposed JSN quantification algorithm is to improve sensitivity, accuracy and robustness so that radiologist/rheumatologist can closely monitor the JSN progression in early RA. The schematic overview of this work is shown in Fig. 2. The methodology in this work can be divided into (i) joint position detection and calibration (ii) joint segmentation (iii) joint space narrowing quantification.

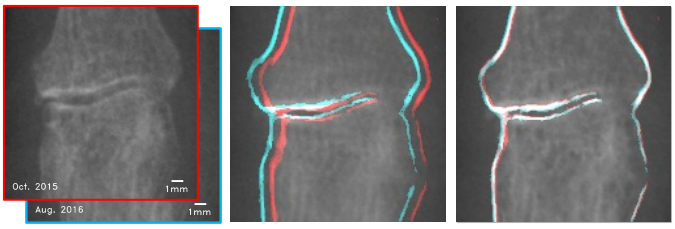


Fig. 4. Results of joint position calibration: (a) A proximal interphalangeal joint (PIP) of little finger in baseline radiograph (red border) and follow-up radiograph (cyan border). (b) The margin of PIP radiograph in (a) before position calibration (red: baseline radiograph, cyan: follow-up radiograph, white: overlap). (c) The margin information after position calibration.

A. Joint position detection and calibration

As shown in Fig. 3, the pipeline of joint position detection and calibration can be briefly explained as follows: (i) Obtain the approximate estimates of the finger midlines in binary image. (ii) Detect joint positions by using a ML-based joint classifier. (iii) Calibrate the relative position deviations in joint windows.

1) Finger midline detection: Finger position estimation can significantly reduce the potential region, thus reduces the calculation of joint detection. The scheme of finger midline detection is shown in Fig. 3. The approximate area and angle of fingers are estimated using the binary images obtained from respective hand radiographs.

Given a radiograph, we binarize the X-ray using Otsu's

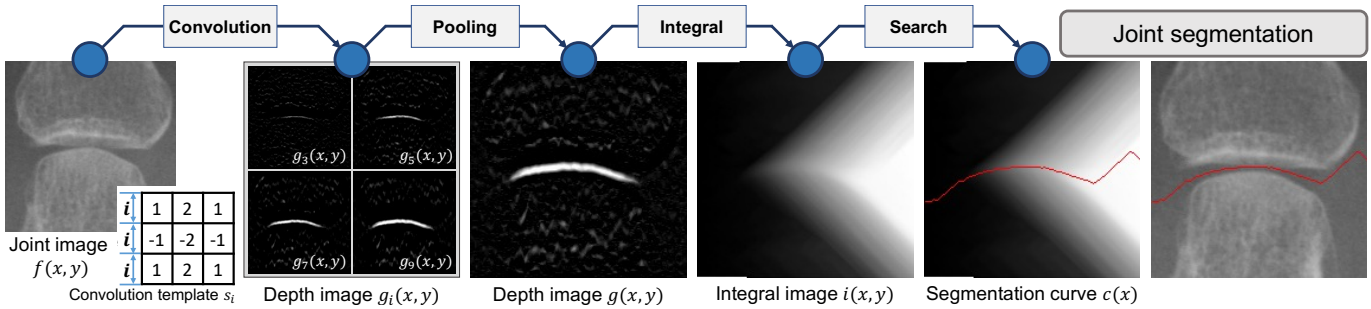


Fig. 5. Overview of joint segmentation based on gradient information. Gully depth map $\boldsymbol{g(x,y)}$ is calculated to define depth feather. Independent margin filtering on the upper and lower side determines the pixel depth. Height-adjustable convolution template $\boldsymbol{S_i}$ ensures that a given range of gully can be detected. The integral image $\boldsymbol{i(x,y)}$ is calculated to search the segmentation curve $\boldsymbol{c(x)}$ with the maximum depth-sum.

method [18], and smooth its margin by using morphological opening and closing. We obtain the local maxima (red point) and local minima (green point) of hand margins as shown in Fig. 3. From our experiments we found that using polygonal approximation can significantly improve robustness when searching for extrema, that one could obtain using pure margin [19]. Next, the approximate region of fingers are obtained according to each pair of local maxima and minima, which can calculate the midline of each finger by fitting to each finger region based on least squares method.

From our experiments and analysis we found that reducing the width and height of the binary hand images to one-fifth does not significantly effect the accuracy of the finger midline detection, and this results in accelerating the detection process (17.7 $\times$ faster).

2) Joint position detection: As shown in Fig. 3, the position, angle, and size of each finger image is obtained according to the finger midline and its region in the binary image. Then, the joint windows are detected in finger images with a joint classifier which is trained by using haar-like feature based AdaBoost [20].

3) Joint position calibration: We propose a low computational solution based on FIPOC to calibrate the joint position, a detailed discussion of FIPOC implementation is presented in Section II-C.As show in Fig. 4 the joint position calibration which relies on FIPOC cannot reduce the deviation with ground truth. It can limit relative position deviation between baseline and follow-up joint windows within one pixel.

B. Joint segmentation

Take a MCP joint as an example, the proximal phalanx bone and metacarpal bone is segmented from the joint image, based on gradient information, so that the displacements of the proximal phalanx bone and metacarpal bone can be measured separately.

1) Depth image: The depth map is used to gauge the depth of each pixel within a given range of width. Only the vertical depth is detected in this work, because all joint images are arranged vertically. Nevertheless, the depth of any direction can be detected with a customized convolution template S_i. The detailed explanation is shown in Fig. 5.

In order to detect depth within a range of width, a height-adjustable convolution template S_i is used (i is odd) to calculate the depth of i pixels gully height. Consider a joint image $f(x,y)$ with M pixel width and N pixel height. The convolution of $f(x,y)$ can be formulated as shown in Eq. 1.

$$\begin{aligned} g_{ia}(x,y) &= \sum_{k=-1}^{1} \sum_{l=-(i-1)/2}^{(3i-1)/2} s_i(k,l) f(x+k, y+l) \\ g_{ib}(x,y) &= \sum_{k=-1}^{1} \sum_{l=-(3i-1)/2}^{(i-1)/2} s_i(k,l) f(x+k, y+l) \end{aligned} \tag{1}$$

$g_{ia}(x,y)$ represents the gradient above, and $g_{ib}(x,y)$ represents the gradient below. The smaller gradient is defined as the depth, $g_i(x,y) = \min(g_{ia}(x,y), g_{ib}(x,y))$.

The depth images $g_i(x,y)$ of the joint image $f(x,y)$ are shown in Fig. 5 when i is 3, 5, 7 or 9. The i represents the height of the height-adjustable convolution template S_i. In depth map $g_i(x,y)$, narrow gullies can have high intensity when i is small, conversely, wide gullies have high intensity when i is large.

A max-pooling is performed on depth map $g_i(x,y)$ to downsample, $g(x,y) = \max(g_{i_{min}}(x,y), \cdots, g_{i_{max}}(x,y))$. The $[i_{min}, i_{max}]$ represents the range of the height of convolution template S_i. This range is positively correlated with the spatial resolution. In our experience this range is defined as $[i_{min} = 1, i_{max} = 9]$ when the spatial resolution is 0.175 mm/pixel.

2) Integral image: The integral image $i(x,y)$ is an intermediate matrix, which is used to find the segmentation curve with the maximum depth-sum. It can be expressed as the local maximum in the left column plus depth map $g(x,y)$, as shown in Eq. 2.

$$i(x,y) = \begin{cases} g(x,y) & x=0 \\ \max(i(x-1,y-1), i(x-1,y), i(x-1,y+1)) + g(x,y) & x>0 \end{cases} \tag{2}$$

3) Segmentation curve: The segmentation curve with the maximum depth-sum can be determined from integral image $i(x,y)$ as follows. First, determine the maximum value of the rightmost column in $i(x,y)$ as the end point of the segmentation curve. Then, select the maximum of the three adjacent pixels in the left column in $i(x,y)$ as the next point of the segmentation curve until arriving at the leftmost column. The segmentation curve $c(x)$ is defined as Eq. 3. The $\arg\max_y i(x,y)$ indicates the index of the maximum value on

the y axis for a given x value in a given y range.

$$c(x)=\arg\max_{y} i(x,y)\begin{cases} y\in[0,N-1] & x=M-1 \\ y\in[c(x+1)-1,c(x+1)+1] & x<M-1 \end{cases} \quad (3)$$

The binary matrix of the proximal phalanx bone $s_0(x,y)$ and the metacarpal bone $s_1(x,y)$ can be expressed as Eq. 4, according to the segmentation curve $c(x)$.

$$s_0(x,y)=\begin{cases}1 & y<c(x)\\ 0 & \text{otherwise}\end{cases} \quad s_1(x,y)=\begin{cases}1 & y>c(x)\\ 0 & \text{otherwise}\end{cases} \quad (4)$$

An example of finger joint segmentation is shown in Fig. 5.

C. JSN progression quantification by partial image phase-only correlation

The basic concept of JSN progression quantification by FIPOC or PIPOC can be described using the flowchart in Fig 6. Consider two images, $f(x,y)$ and $g(x,y)$, which are divided into k regions. Consider a region i, let α_i and β_i represent sub-pixel displacement from $f(x,y)$ to $g(x,y)$ in x and y directions respectively, and a binary matrix $s_i(x,y)$ that includes segmentation information. So, $g(x,y)$ can be represented as $g(x,y)=\sum_{i=0}^{k} f(x-\alpha_i,y-\beta_i)*s_i(x,y)$.

A 2D Hanning window function is applied to input images $f(x,y)$ and $g(x,y)$ to reduce the effect of discontinuity at image border [21]. The Hanning window $w(x,y)$ can be defined as $w(x,y)=\frac{1+cos(\frac{\pi x}{M})}{2}\frac{1+cos(\frac{\pi y}{N})}{2}$.

Let $F(u,v)$ and $G(u,v)$ denote the 2D Discrete Fourier Transforms (DFT) of the two images. Considering the properties of DFT $\mathcal{F}$, $F(u,v)$ and $G(u,v)$ can be expressed as Eq. 5.

$$F(u,v)=\mathcal{F}(f(x,y)w(x,y)) \quad G(u,v)=\mathcal{F}(g(x,y)w(x,y)) \quad (5)$$

Next, extract the phase component of $F(u,v)$ and $G(u,v)$, the functions are divided by the amplitude, as follows:

$$\hat{F}(u,v)=\frac{F(u,v)}{|F(u,v)|} \quad \hat{G}(u,v)=\frac{G(u,v)}{|G(u,v)|} \quad (6)$$

FIPOC will calculate the phase difference spectrum $\hat{r}(u,v)$ between $\hat{F}(u,v)$ and $\hat{G}(u,v)$ immediately (the dotted line in Fig. 6). But when the displacement of each region is different, there will be several dirac delta functions in phase difference spectrum, as show in Eq. 7.

$$\hat{r}(u,v)=\sum_{i=0}^{k} p_i\delta(\alpha_i,\beta_i) \quad (7)$$

Different from FIPOC, PIPOC segments the phase spectrum in spatial domain. Next, the phase spectrum $\hat{f}(x,y)$ of image $f(x,y)$ and the phase spectrum $\hat{g}(x,y)$ of image $g(x,y)$ in spatial domain are obtained by Inverse Discrete Fourier Transform (IDFT) $\mathcal{F}^{-1}$.

$$\hat{f}(x,y)=\mathcal{F}^{-1}(\hat{F}(u,v)) \quad \hat{g}(x,y)=\mathcal{F}^{-1}(\hat{G}(u,v)) \quad (8)$$

Segmenting region i by using segmentation matrix $s_i(x,y)$.

$$\hat{f}_i(x,y)=\hat{f}(x,y)*s_i(x,y) \quad \hat{g}_i(x,y)=\hat{g}(x,y)*s_i(x,y) \quad (9)$$

Subsequently combining DFT $\mathcal{F}$ and Eq. 9 to develop the phase spectrum of region i in frequency domain.

$$\hat{F}_i(u,v)=\mathcal{F}(\hat{f}_i(x,y)) \quad \hat{G}_i(u,v)=\mathcal{F}(\hat{g}_i(x,y)) \quad (10)$$

The normalized cross phase spectrum $\hat{R}_i(u,v)$ of region i between $F(u,v)$ and $G(u,v)$ can be obtained respectively as given in Eq. 11. Here, $\overline{G_i(u,v)}$ in Eq. 11 denotes the complex conjugate of $G_i(u,v)$.

$$\hat{R}_i(u,v)=\frac{\hat{F}_i(u,v)\overline{\hat{G}_i(u,v)}}{|\hat{F}_i(u,v)\overline{\hat{G}_i(u,v)}|} \quad (11)$$

Next, the phase difference spectrums $\hat{r}_i(x,y)$ of region i between the two images are obtained by IDFT $\mathcal{F}^{-1}$. The location of the dirac delta function δ represents the displacement between two images.

$$\hat{r}_i(x,y)=\mathcal{F}^{-1}(\hat{R}_i(u,v))=\delta(\alpha_i,\beta_i) \quad (12)$$

In case of Fourier Transform, the location of the peak of dirac delta function δ in the phase difference spectrum $\hat{r}_i(x,y)$ can be determined according to the maximum peak.

$$(\alpha_i{}',\beta_i{}')=\arg\max_{(x,y)}\hat{r}_i(x,y) \quad (13)$$

Consider the DFT, the least-square fitting method employed to estimate displacement (α_i,β_i) around the maximum peak $(\alpha_i{}',\beta_i{}')$. Since the δ function has a very sharp peak, limited number of data points 5×5 are used to fit δ function [21] in this work. Thus, the JSN_{fg} between image $f(x,y)$ and image $g(x,y)$ can be quantified according to the displacement difference between the proximal phalanx bone $s_0(x,y)$ and the metacarpal bone $s_1(x,y)$, $\text{JSN}_{fg}=\beta_0-\beta_1$.

D. FIPOC and PIPOC

Previously, we had proposed a JSN progression quantification method [10], [13], which is based on FIPOC [21]. FIPOC is a well-known method for image registration, it can estimate the relative displacement between two images and it is based on the frequency domain analysis. Considering that there is a single displacement α in the spatial domain between the two signals $f(x)$ and $g(x)$, this displacement appears in the phase difference spectrum as a dirac delta function with coordinate α. FIPOC can measure displacement by quantifying the coordinate of the dirac delta function to within 0.01 pixels of error [21].

Nevertheless, when the displacement between two signals is different in different regions, the phase dispersion will appear in the phase difference spectrum. Take the baseline MCP joint image $f(x,y)$ and its follow-up image $g(x,y)$ as an example (like Fig. 7), let α_0 and α_1 represent the vertical displacement of the proximal phalanx bone and metacarpal bone respectively. When α_0 and α_1 is different, the displacement of each region will show a dirac delta function in the phase difference spectrum, its center coordinate represents the displacement, and its intensity is positively related to the region. Thus, the JSN progression can be quantified by calculating the distance difference between these two dirac delta function. This basic idea is widely be used in image processing. In [22] Shimada

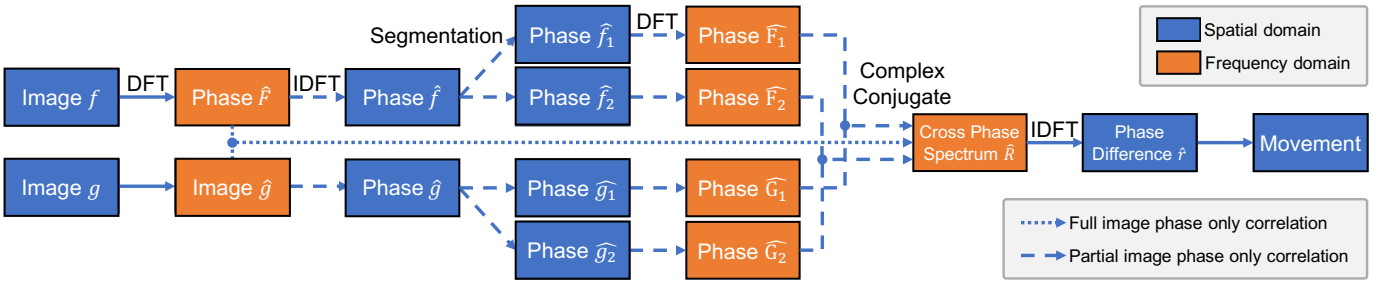


Fig. 6. A flowchart describing the sequence of operations for implementing FIPOC and PIPOC algorithms.

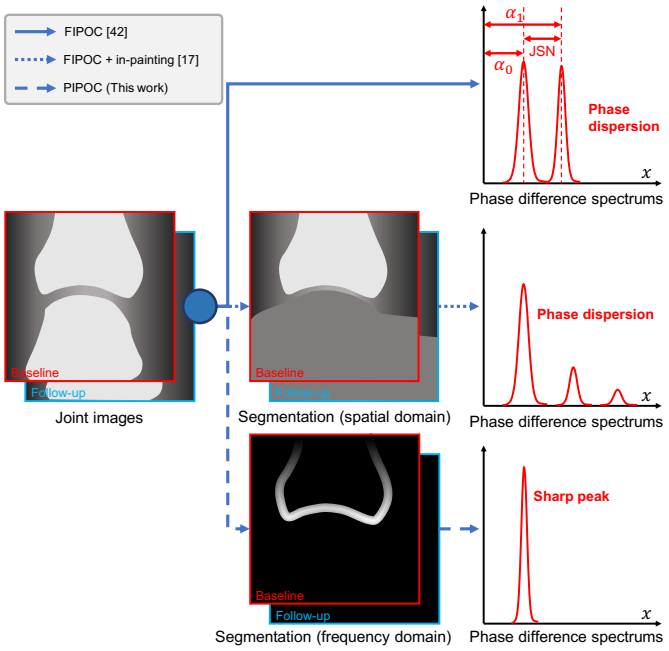


Fig. 7. Take two MCP joint image as an example, a schematic diagram of the comparison of the dirac delta function in the phase difference spectrum when using FIPOC (without segmentation), FIPOC (segment in spatial domain and combine with in-painting algorithm) or PIPOC (segment in frequency domain) when there are multiple regions with different displacement.

et al. proposed a sparse disparity estimation algorithm based on FIPOC for depth prediction in binocular stereo vision. [22] estimate sparse disparity in different depths of field by separately calculating the positions of multiple dirac delta functions in phase difference spectrum. However, considering that the dirac delta functions in close proximity can affect each other and even overlap, and could reduce the accuracy of their coordinate measurement. The precise position of each dirac delta function can be obtained if and only if the displacement differences between multiple regions are large enough (about 3 pixels [22]).

When the displacement difference is small and cannot be simply quantified by using the above method, then alternately in previous works they would segment images in spatial domain [10]. Compared to the FIPOC method to directly quantify the displacements of different regions in the phase difference spectrum, the proposed method combines FIPOC, spatial domain segmentation and image in-painting algorithm to quantify smaller displacement differences, additionally, it can have wider application prospects in issues where the displacement differences are small (like JSN progression quantization). Indeed, this approach can quantify the small displacements of multiple regions from independent images respectively. However, if we don't fill vacant space by using image in-painting algorithm, we cannot apply fast Fourier transform for FIPOC processing. The in-painting algorithm can generate some non-existent phase features, and thus lead to phase dispersion in phase difference spectrum. This phase dispersion can increase quantification error and even cause mismatches (the second row in Fig. 7). Therefore, we have found a method to exploit the effectiveness of FIPOC, no matter what shape the segmentation is in.

PIPOC, the algorithm proposed in Section II-C, can determine the location of each dirac delta functions and avoids the impact of the various object segmentation including any form background on accuracy in the phase spectrum. In our JSN progression quantification experiments, PIPOC has a lower phase dispersion compared to the method combining FIPOC, spatial domain segmentation and in-painting algorithm. PIPOC can avoid the effects of in-painting algorithms, and obtain a sharper delta function in phase spectrum (the third row in Fig. 7). The proposed PIPOC can further improve the accuracy and robustness of the JSN progression quantification.

III. Materials

Imaging phantom based experiments were studied to evaluate our algorithm's performance. From our experiments we observed that the manually labeled JSW has pixel level mean error, and it is discussed in Section IV-B.1. To evaluate the performance of our algorithm, we prepared phantom and clinical images. Phantom images with ground truth were used to measure accuracy in terms of absolute error.

A. Phantom study

Imaging phantoms are specially designed physical test objects for simulating anatomical structures of the body, which makes them easier and more convenient to use. The two-layer phantom can simulate the X-ray absorption coefficient (CT value) of bone cortex and cancellous bone. This is an advantage in X-ray imaging due to dose issues, phantom devices can be thus tested without using human volunteers.

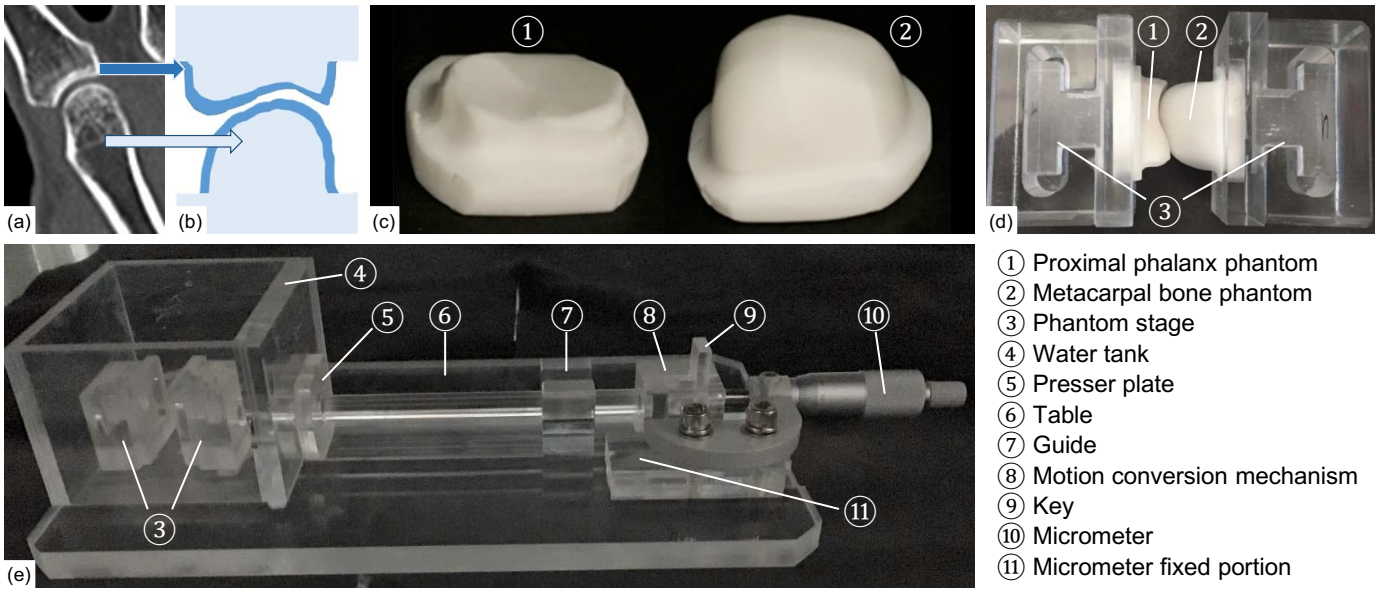


Fig. 8. A MCP joint-shaped two-layer phantom design and phantom imaging environment. (a) A MCP joint in radiographic clinical imaging. (b) A diagram of two-layer structure bone (dark blue: Bone cortex, light blue: Cancellous bone). (c) A set of MCP joint-shaped two-layer phantom. (d) The phantom joint connect with the attaching portion. (e) The phantom imaging environment.

TABLE I
PHANTOM DESIGN PREPARATION

	Bone cortex	Cancellous bone
TMA : adhesive	1:1.2	1:5
Particle Size (μm)	107 $\sim$ 250	107 $\sim$ 250
Temperature (K)	1370	1370

1) Phantom design: A phalanx-shaped phantom was produced using vacuum-sintered bodies of a novel apatite called Titanium medical apatite (TMA) [23]. The chemical formula of TMA is $Ca_{10}(PO_4)_6 \cdot TiO_2 \cdot (OH)_2 \cdot nH_2O$. TMA powder was kneaded with distilled water, and solid cylinders of compacted TMA were formed by compression molding at $10MPa$. TMA was vacuum sintered using a resistance furnace at about $10^{-3}Pa$.

Using TMA to design imaging phantom has the following advantages: (1) The CT value of phantom in radiographs can be easily modified by changing the ratio of TMA and adhesive. (2) TMA bodies are easy to process and model with a 3D-modeling machine or a lathe. (3) TMA vacuum-sintered bodies has a density of approximately $2300kg/m^3$ (corresponding to that of a compact bone or a tooth).

The phantom used in our experiment is a two-layer TMA vacuum-sintered body to simulate the X-ray absorption coefficient (CT value) of bone cortex and cancellous bone. The diagram of the two-layered bone is shown in Fig. 8 (a) and (b). The phantom mimics MCP joint, proximal phalanx, and the metacarpal bone. The assembled phantoms are shown in Fig. 8 (c). The important properties of our phantom bones are given in Table I [23].

2) Imaging environment: The phantom joint was mounted on to the stage as shown in Fig. 8 (d). The phantom stage was connected to a micrometer, and thereby the JSW of phantom could be easily adjusted using the micrometer controls. The JSW range is up to 13 mm, and has a minimum scale of 0.01 mm. There is substantial evidence that JSW has a close relationship with age and sex in healthy populations [16], [24]. In addition, RA is more frequent in females who are between 30 and 50 years of age, and their JSW is around 1.70 mm [16], [24]. In our work, the JSW standard of phantom was set as 1.70 mm. Following two sets of phantom images with different specifications were provided. (i) JSW range: 1.20 mm - 2.20 mm, and increment step size: 0.10 mm (ii) JSW range: 1.65 mm - 1.75 mm, and increment step size: 0.01 mm.

Figure 9 shows some MCP phantom images that were used in our experiments. Clinical researches show that tomosynthesis is superior to radiography for the manual evaluation of JSN and the manual detection of bone erosion in patients with RA [25], [26]. However, considering that radiography is the most widely used imaging technique in the diagnosis of RA, we utilized both imaging techniques (radiography and tomosynthesis) in our phantom experiments.

The radiographic imaging device used in our phantom experiment is *FUJIFILM DR CALNEO Smart C47* from *Fujifilm* Corporation, Tokyo, Japan, with a 1.5 mm X-ray aluminum filter thickness. The tomosynthesis imaging device is *Side Station i3* from *Shimadzu* Corporation, Kyoto, Japan. The reconstruction function used in tomosynthesis is IR Hand F, and the effective slice thickness is 2 mm. Digital imaging and communications in medicine (DICOM) standard was used in managing our datasets in phantom study. The spatial resolution used in our radiographic phantom study is 0.15 mm/pixel at 12 bit depth. And 0.26 mm/pixel at 12 bit depth in our tomosynthesis dataset. For detailed imaging parameter descriptions, please refer to Table II.

In phantom-based experiments, filling air is easier to implement, for comparison with existing and possible future air-filled phantom studies, we imaged phantom in the air. Nevertheless, in clinical studies, the X-ray beam can be

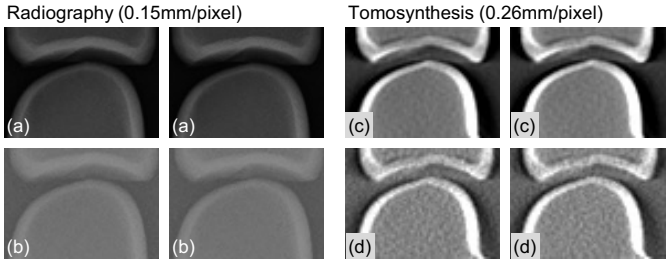


Fig. 9. MCP phantom images in different imaging conditions (tank filled with air or water). Radigraphic imaging: (a) air (b) water. Tomosynthesis: (c) air (d) water. In each group we show two images with different joint space width, left: 1.20mm, right : 1.30mm.

TABLE II
RADIOGRAPHIC IMAGING CONFIGURATION PARAMETERS

	Radiography		Tomosynthesis
	Phantom	Clinical	Phantom
Tube voltage (kV)	50	42	47
Tube current (mA)	100	100	250
Exposure time (mSec)	20	20	105
Source to image (cm)	100	100	N/A

attenuated by the tissue, these attenuations are displayed as noise in the radiography. In related phantom studies, water is usually used to simulate the noise generated by the beam attenuation in the tissue [27], [28]. In our phantom study, the phalanx-shaped phantom was mounted on the stage as shown in Fig. 8 (d), and placed in a tank. We can image the phantom with low noise when the tank is filled with air, or filled with distilled water which has an X-ray absorbing properties similar to normal tissue. Our experimental phantom imaging setup is shown in Fig. 8 (e).

B. Clinical dataset

1) Study population: For clinical assessment, we prepared dataset from *Sagawa Akira Rheumatology Clinic* (Sapporo, Japan), *Sapporo City General Hospital* (Sapporo, Japan) and *Hokkaido Medical Center for Rheumatic Diseases* (Sapporo, Japan). This dataset contains 1120 hand posteroanterior projection (PA) radiographs from patients diagnosed with early-stage RA. All images were used in the joint position detection experiments. Considering that several images were required to evaluate our work when calculating standard deviation. Thus, images of patients who were radiographed at least three times were retained, which contains 549 hand PA radiographs of 77 RA patients out of which 88.0% are female. Detailed patients information are summarized in Table III (please note, the gender and age information of a small number of patients were not included upon patient request).

This study was conducted in accordance with the guidelines of the Declaration of Helsinki and approved by the Ethics Committee of the Faculty of Health Sciences, Hokkaido University (approval number: 19 - 46).

2) Imaging environment: The radiographic imaging device used in our clinical study is *DR-155HS2-5* from *Hitachi* Corporation, Tokyo, Japan, with 1.5 mm X-ray aluminum filter thickness. The centering point of the X-ray beam was the MCP joint of the middle finger. DICOM standard was used in managing our cilical dataset, and the image resolution is 2010 pixels $\times$ 1490 pixels, and a 0.175 mm $\times$ 0.175 mm pixel size at 12 bit depth. For detailed imaging parameter descriptions, please refer to Table II.

TABLE III
PATIENT INFORMATION IN THE CLINICAL DATASET

	Mean $\pm$ SD	Range
Age at enrollment (year)	55.83 $\pm$ 13.86	20.68 $\sim$ 88.00
Number of Photography*	4.30 $\pm$ 2.54	3 $\sim$ 17
Treatment Duration (year)	4.01 $\pm$ 3.43	0.88 $\sim$ 12.10

* Patients did two-handed or one-handed radiographic imaging.

TABLE IV
FALSE NEGATIVE AND FALSE POSITIVE COUNTS AND RATIOS OF JOINT LOCATION DETECTION

	False Negative				False positive	
	IP	DIP	PIP	MCP	CMC	Others
Thumb	24	N/A	N/A	5	63	2
Index	N/A	0	2	0	0	7
Middle	N/A	1	1	0	0	2
Ring	N/A	2	0	2	0	4
Small	N/A	11	1	0	0	1
Overall	**24(2.14%)**	**14(0.31%)**	**4(0.09%)**	**7(0.16%)**	**63**	**16**

* **IP**: Interphalangeal joint. **DIP**: Distal interphalangeal joint. **PIP**: Proximal interphalangeal joint. **CMC**: Carpometacarpal joint.

IV. EXPERIMENTS AND DISCUSSION

A. Joint position detection

We selected a 22-layer AdaBoost based classifier for joint position detection, which has a false negative ratio and a false positive ratio of 0.31% and 0.50% respectively. The performance on each joint is shown in the Table IV. From this table, we can observe that false positives occurred mainly in the carpometacarpal (CMC) joint of the thumb, which is the joint that most closely resembles the target joints in a hand radiograph. And false negative appeared mainly in the thumb, especially the interphalangeal (IP) joint. In our opinion, the main reason for this situation is that the radiographic angle of the thumb is different from other fingers, resulting a difference in radiography. Differentiation of the joint position detection on the thumb may be effective in improving detection accuracy.

B. JSN progression quantification

1) Phantom study: Phantom images with ground truth were used in this experiment to calculate the absolute error of PIPOC, and to compare it with manual measurements. Our manual measurement experiments were performed by one radiologist and one radiological technologist after substantial training. They did not know the ground truth of the out-of-order phantom images. They were asked to determine the center of the proximal phalanx bone phantom by drawing straight lines horizontally connecting both ends of the phantom base, then a straight line was drawn from the center vertically, and the JSW overlapping the straight line was measured.

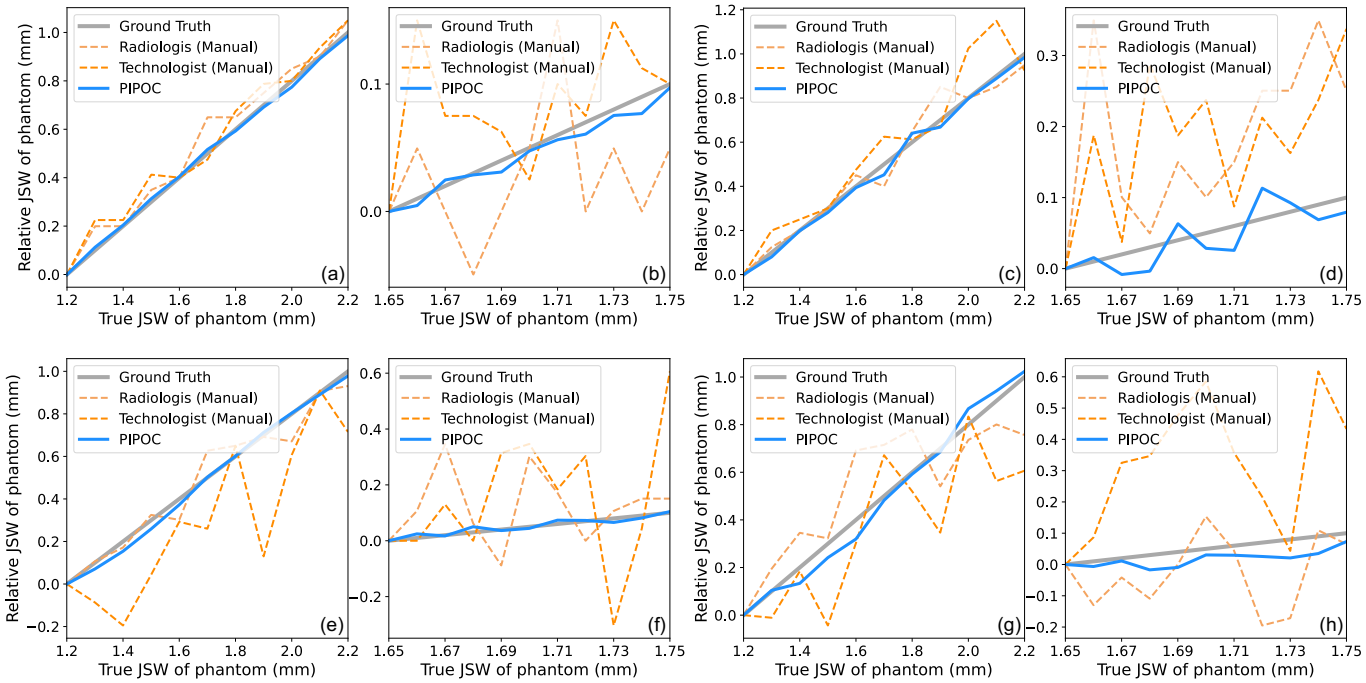


Fig. 10. The measurement result of PIPOC and manual in phantom study. Blue lines are the relative JSW of each image to the first image obtained by PIPOC. Orange dot lines are the difference of manually measured JSW between every image and the first image. We preformed two imaging techniques in our phantom study, radiography: (a) ∼ (d), Tomosynthesis: (e) ∼ (h). The phantom of sub-figure (a), (b), (e) and (f) is placed in air. The phantom of sub-figure (c), (d), (g) and (h) is placed in distilled water. The true JSW of phantom is from 1.20mm to 2.20mm at increments of 0.10mm in sub-figure (a), (c), (e) and (g). And it is from 1.65mm to 1.75mm at increments of 0.01mm in sub-figure (b), (d), (f) and (h).

TABLE V
THE PERFORMANCE ANALYSIS IN MILLIMETER FOR PIPOC AND MANUAL MEASUREMENT WHEN USING PHANTOM IMAGES

	Mean Error						Root-Mean-Square Deviation					
	Air			Water			Air			Water		
Radiography	Fig.10(a)	Fig.10(b)	Average	Fig.10(c)	Fig.10(d)	Average	Fig.10(a)	Fig.10(b)	Average	Fig.10(c)	Fig.10(d)	Average
Radiologist (Manual)	0.0509	0.0620	0.0565	0.0727	0.1196	0.0961	0.0665	0.0758	0.0711	0.0923	0.1450	0.1186
Technologist (Manual)*	0.0595	0.0497	0.0546	0.1186	0.1034	0.1110	0.0709	0.0632	0.0671	0.1440	0.1237	0.1339
Mean of Manual	0.0552	0.0559	0.0555	0.0957	0.1115	0.1036	0.0687	0.0695	0.0691	0.1182	0.1343	0.1263
PIPOC (Ours)	**0.0193**	**0.0066**	**0.0130**	**0.0251**	**0.0200**	**0.0226**	**0.0220**	**0.0081**	**0.0150**	**0.0303**	**0.0245**	**0.0274**
FIPOC [10]	0.0400	-	0.0400	-	-	-	-	-	-	-	-	-
Tomosynthesis	Fig.10(e)	Fig.10(f)	Average	Fig.10(g)	Fig.10(h)	Average	Fig.10(e)	Fig.10(f)	Average	Fig.10(g)	Fig.10(h)	Average
Radiologist (Manual)	0.0815	0.1477	0.1146	0.2009	0.1316	0.1662	0.1000	0.1835	0.1418	0.2365	0.1603	0.1984
Technologist (Manual)*	0.2155	0.2671	0.2413	0.2210	0.2342	0.2276	0.2616	0.3307	0.2962	0.2678	0.2800	0.2739
Mean of Manual	0.1485	0.2074	0.1780	0.2110	0.1829	0.1969	0.1808	0.2571	0.2190	0.2521	0.2201	0.2361
PIPOC (Ours)	**0.0180**	**0.0124**	**0.0152**	**0.0329**	**0.0243**	**0.0286**	**0.0220**	**0.0154**	**0.0187**	**0.0407**	**0.0292**	**0.0349**

* Measured manually by a radiological technologist.

Figure 10 and Table V presents the measurement result of phantom study. The manual measurement result of the radiologist and the radiological technologist showed high similarity in terms of mean error and RMSD in multiple phantom data sets. The mean error of manual measurements are about 0.0555 mm (0.37 pixel) in low noise environment (air sets), and 0.1036 mm (0.69 pixel) in high noise environment (water sets) in radiographic phantom studies. For tomosynthesis datasets, the mean error is 0.1780 mm (0.68 pixel) in low noise environment, and 0.1969 mm (0.76 pixel) in high noise environment. This shows that visual measurement also can be greatly affected by the noise. On the other hand, this also indicates the manually annotated data have sub-pixel level mean error. Hence, the manually annotated ground truth may result in sub-pixel level deviation in algorithm evaluation of other works.

In paper [10], only one phantom dataset (environment: air, JSW range: 1.20 mm - 2.20 mm, increment step size: 0.10 mm) is used in experiment. The mean error of FIPOC is slightly lower than manual measurement. When compared to FIPOC, PIPOC can further improve the accuracy and robustness in JSN progression quantification, by eliminating the impact of image in-painting algorithm. As show in Table V, our work only has a 11.9% to 35.0% mean error, and a 11.7% to 32.0% RMSD when compared to manual measurement in radiographic phantom study. For the tomosynthesis datasets, our work only has 6.0% to 15.6% mean error, and a 6.0% to 16.1% RMSD in comparison to manual measurement. This illustrates the improved performance of JSN progression quantification when using phantom datasets. Considering the spatial

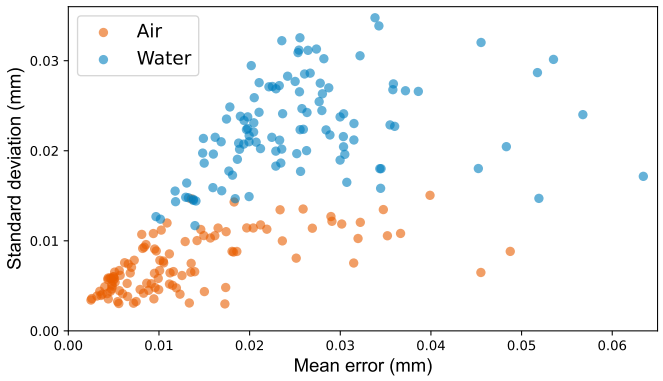
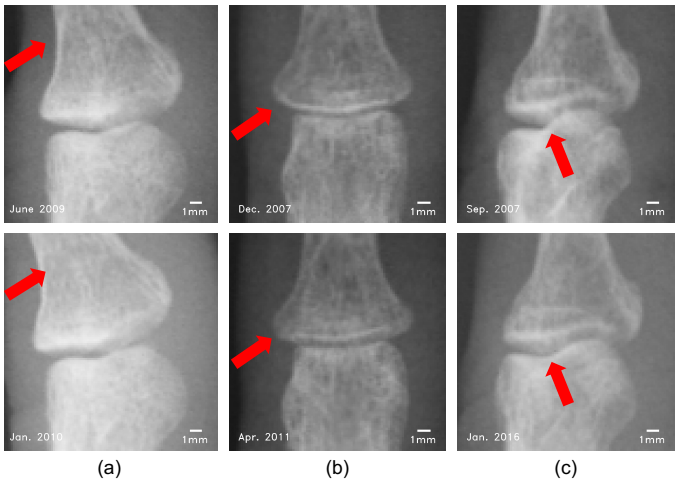


Fig. 11. The relationship between the mean error $\boldsymbol{E}$ and the standard deviation $\boldsymbol{\sigma}$ of all sets of JSN in radiographic phantom data.



(a) (b) (c)

Fig. 12. Joints with mismatched registration. (a) Inconsistent joint angle. (b) Bended finger. (c) Inconsistent projection angle.

resolution of radiography (0.15 mm/pixel) and tomosynthesis (0.26mm/pixel), we can notice that PIPOC performs better on tomosynthesis images when compared in pixels. We regard this difference is due to the sharper edge information of the tomosynthesis. Thus, we recommend using PIPOC in tomosynthesis preferentially with same spatial resolution.

In related works, the ground truth of joint space is usually measured by the radiologist or the rheumatologist manually. But as discussed above, manual measurement also have sub-pixel level mean error. Thus, manually measured ground truth may result in sub-pixel deviation in algorithm evaluation. This deviation is negligible when evaluating the algorithm on a pixel scale. But it can be inaccurate on a sub-pixel scale. To the best of our knowledge there are no published algorithms/methods which can compute ground truth RA joint space with sub-pixel accuracy. We propose to use the standard deviation σ of multiple measurements to demonstrate the reliability of PIPOC without ground truth. The definition of standard deviation can be described as follows.

In case of three images f, g and k, the JSN_{fg-k} between image f and image g can be indirectly calculated by introducing intermediate image k, $\text{JSN}_{fg-k} = \text{JSN}_{fk} + \text{JSN}_{kg}$. Considering a set of images, the $\overline{\text{JSN}_{fg}}$ can be obtained by taking the average of multiple measurements, $\overline{\text{JSN}_{fg}} = \frac{1}{n}\sum_{k=1}^{n}\text{JSN}_{fg-k}$. So, the standard deviation σ of JSN_{fg} is defined as Eq. 14.

$$\sigma_{fg} = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(\text{JSN}_{fg-k} - \overline{\text{JSN}_{fg}})^2} \tag{14}$$

The standard deviation σ_{fg} represents a dispersion of a set of JSN_{fg-k} ($k \in [1, n]$). According to our experiments when using phantom datasets, the standard deviation σ and the mean error E has a high positive correlation, as show in Fig. 11. The Pearson correlation coefficient between σ and E is 0.641 (count: 220, p-value: $< .001$). For the above reason, and the most important advantage that the standard deviation σ not relying on the ground truth, we used it to measure the performance of our work in clinical databases. In addition, we also found that noise in radiography due to beam attenuation in tissue can greatly affects the accuracy of measurements especially in terms of standard deviation.

TABLE VI
THE MEAN STANDARD DEVIATION IN MILLIMETER AND THE MISMATCHING RATIOS FOR PIPOC

	Clinical Data				Phantom Data	
	IP	DIP	PIP	MCP	Air	Water
Thumb	0.093(7.2%)	N/A	N/A	0.078(4.5%)	-	-
Index	N/A	0.047(4.0%)	0.065(5.2%)	0.051(1.6%)	-	-
Middle	N/A	0.055(5.8%)	0.061(3.4%)	0.057(4.3%)	-	-
Ring	N/A	0.029(4.1%)	0.033(1.8%)	0.023(1.8%)	-	-
Small	N/A	0.044(5.9%)	0.053(3.7%)	0.038(2.0%)	-	-
Overall	**0.093(7.2%)**	**0.044(5.0%)**	**0.053(3.5%)**	**0.050(2.8%)**	**0.007**	**0.025**

2) Clinical data: 549 hand PA radiographs have been analyzed in this subsection. Compared to phantom data, clinical data lay out additional challenges. The major challenge in this work is the uncertainty of hand posture, different hand postures can present differentiated bone contours.

According to our experiments, changes in bone contours can affects the accuracy of JSN progression quantification. Here, we showcase (see Fig. 12) majority of mismatch bone contour cases. The most frequent reason is the inconsistent angle between the upper and lower bones of joint, as show in Fig. 12 (a). This mainly occurs on IP and MCP joints. PIPOC has high accuracy for translation detection, but weak resistance to rotation. Another important reason of mismatched registration is the bending of the fingers, which appears on DIP and PIP joints, for an example see Fig. 12 (b). Finger bending can result in the changes of the far margin appearance of upper bone. Besides, inconsistent projection angle also can be the reason, see Fig. 12 (c). Most of the time it happens only on the IP joint, which is caused by inconsistent joint position or thumb roll. The individuated finger movements differ greatly as studied in [29]. Movements of the thumb, index finger, and little finger typically were more highly individuated than were movements of the middle or ring fingers. The angular motion tended to be greatest at the PIP joint of each digit [29]. It is worth noting that, the flexibility of joint and standard deviation express high positive correlation (refer Table VI).

In summary, the hand posture should be consistent and avoid bending of the fingers, especially the thumb when using our

TABLE VII

COMPARISON WITH RELATED WORKS. MEAN ERROR AND STANDARD DEVIATION IN MILLIMETER. NUMBERS IN BRACES INDICATE THE CORRESPONDING PERCENTAGE OF THE GROUND TRUTH VALUE FOR THE RESPECTIVE JOINT.

		Dataset	Resolution	Mean Error				Standard Deviation			
		(radiographs)	(mm/pixel)	DIP	PIP	MCP	Overall	DIP	PIP	MCP	Overall
Neural Network [6]	'00	54	0.1	0.118	0.071	0.091	0.093	-	-	-	-
Active Shape Models [7]	TMI'08	160 MCP*	0.0846	-	-	0.283(16.1%)	0.283(16.1%)	-	-	0.080(4.5%)	0.080(4.5%)
Edge Detection [8]	TBME'15	104	0.1	(5.8%)	(7.2%)	(7.1%)	(6.8%)	(4.8%)	(5.3%)	(4.4%)	(4.8%)
PIPOC (Ours)	**-**	**549**	**0.175**	**-**	**-**	**-**	**-**	**0.044**	**0.053**	**0.050**	**0.049**
FIPOC [10]	ISBI'19	Phantom	0.15	-	-	0.040	0.040	-	-	-	-
Manual Measurement	-	Phantom	0.15	-	-	0.056	0.056	-	-	-	-
PIPOC (Ours)	**-**	**Phantom**	**0.15**	**-**	**-**	**0.013**	**0.013**	**-**	**-**	**0.007**	**0.007**

* The dataset in [7] contains 160 MCP joint radiographs. Considering that each hand radiograph contains 5 MCP joints, this dataset can be equivalent to 32 hand radiographs.

work for JSN progression quantification. Thus, we strongly recommend that using guide lines lines to standardize hand posture in taking radiography, this simple step can greatly improve the accuracy of PIPOC.

3) Comparison with related works: Table VII compares our work with previous JSW/JSN quantification works. In paper [6], they only used RMSD instead of mean error to evaluate the accuracy of their work, so we standardized the error metric accordingly. Considering that the error should conform to a Gaussian distribution, the mean error and RMSD can be transformed by Eq. 15.

$$\begin{aligned} E &= \int_{-\infty}^{+\infty} |x| \cdot \frac{1}{\sqrt{2\pi} \cdot \mathrm{RMSD}} e^{-\frac{x^2}{2 \cdot \mathrm{RMSD}^2}} dx \\ &= \sqrt{\frac{2}{\pi}} \cdot \mathrm{RMSD} \end{aligned} \tag{15}$$

In paper [8], authors only give the corresponding percentage of the error to the ground truth. Considering the JSW of MCP is around 1.70 mm [16], [24], the mean error of MCP in millimeter is around 0.121. It is noteworthy that, papers [6]–[8] used manual measurement results as ground truth. As discussed above and in Table VII, manual measurement has an error about 0.056 mm (low noise) / 0.104 mm (high noise) when using phantom data (spatial resolution: 0.15 mm/pixel). Although this value can decrease with higher spatial resolution, it is undeniable that in these works which employ manual measurement as the ground truth, the mean error may have a deviation.

The calculation procedure of standard deviation in paper [7] is different from ours. They measured JSW of each joint 10 times with varying clipping of the entire radiograph. The standard deviation quantified the uncertainty of measuring a radiograph. In our work, an intermediate radiograph is introduced in standard deviation calculation. The JSN progression between the two radiographs and the intermediate image is calculated respectively, thus, the standard deviation can be obtained by changing the intermediate image. When using the standard deviation calculation method given in paper [7], we measured a lower standard deviation (DIP: 0.0099 mm, PIP: 0.0095 mm, MCP: 0.0061 mm. These standard deviations do not include mismatched data, the mismatching ratios are shown in Table VI).

Compared to the method which combines FIPOC, spatial domain segmentation and image in-painting algorithm [10], this work has a lower mean error in phantom study, and faster processing speed (on our clinical dataset, PIPOC (Ours): 0.0121 sec/time, FIPOC [10]: 0.0358 sec/time). These improvements are due to the removal of the in-painting algorithm.

We can observe from Table VII that even though the spatial resolution of our work is poorer than those in the related works, our mean error and the standard deviation are significantly lower.

V. CONCLUSION AND FUTURE WORKS

This work aims for computer-aided monitoring and diagnosis of rheumatoid arthritis (RA). We proposed a joint space narrowing (JSN) progression quantization framework based on image registration. Our extensive experiments demonstrate that our proposed framework promises high precision monitoring when compared to two mainstream related works; the scoring framework based on machine learning classification, and the joint space width (JSW) quantization framework based on edge detection. Our proposed framework can fill the gap in monitoring the JSN progression with high sensitivity, and offers a broad application prospect.

In this work, we have improved the full image phase only correlation (FIPOC) algorithm by adding a phase spectrum segmentation step, so that it can measure the displacements of multiple regions at the same time. We named the improved FIPOC algorithm as partial image phase only correlation (PIPOC). As an image registration algorithm, PIPOC is used in JSN progression quantification. Compared with FIPOC, PIPOC can effectively avoid the impact of the segmentation and in-painting process and reduce the phase dispersion in the phase difference spectrum. In our phantom study, PIPOC has a much lower mean error than FIPOC.

The proposed framework implements fully automatic quantization of JSN progression in RA. The algorithm pipeline consists of joint position detection, joint segmentation and JSN progression quantification. In this work, the joint positions are detected with a joint classifier which is trained by haar-like feature based AdaBoost, and a ridge detection-like algorithm is proposed to calculate the segmentation curve. The JSN progression is quantified by PIPOC.

In our phantom studies, and experiments using clinical dataset our algorithm can measure the displacements of upper and lower bones with sub-pixel accuracy. The measured mean

error of our algorithm is in range of 11.9% - 35.0% in comparison to manual measurements using radiographic phantom datasets, in range of 6.0% - 16.1% when using tomosynthesis phantom datasets, and with a standard deviation of 0.0519 mm when using radiographic clinical dataset. Our work greatly improves the accuracy and sensitivity of JSN progression quantification, which might help radiologists/rheumatologists to make more timely judgments on diagnosis and prognosis in rheumatoid arthritis patients.

Currently, machine learning (ML) is applied to difficult tasks in medical image processing [14], [15]. We anticipate future studies in this direction. Our experiments in this study has shown the superiority of image registration based JSN progression quantification framework in RA monitoring compared to the current most popular classification based SvdH scoring framework, and the margin detection based JSW quantification framework. Those ML-based image registration algorithms can be used for JSN quantization in RA. To address the posture (finger movement) related constraints and inconsistent joint angle which is likely to result in mismatched registration. Given the advantages of ML, it may be possible to achieve higher robustness (lower mismatching ratios) at the expense of a small amount of accuracy [15]. Furthermore, we can quantify JSN by ML using the image features extracted by our work, this can improve the overall performance of the algorithm.

VI. Acknowledgments

The authors would like to sincerely thank Akira Sagawa, MD, Sagawa Akira Rheumatology Clinic (Sapporo, Japan), Masaya Mukai, MD, Sapporo City General Hospital (Sapporo, Japan) and kazuhide Tanimura, MD, Hokkaido Medical Center for Rheumatic Diseases (Sapporo, Japan) for image data preparation.

References

[1] J. J. Cush, A. Kavanaugh, and M. E. Weinblatt, *Rheumatoid arthritis: early diagnosis and treatment*. Professional Communications, 2010.

[2] A. Young, J. Dixey, N. Cox, P. Davies, J. Devlin, P. Emery, S. Gallivan, A. Gough, D. James, P. Prouse *et al.*, "How does functional disability in early rheumatoid arthritis (ra) affect patients and their lives? results of 5 years of follow-up in 732 patients from the early ra study (eras)," *Rheumatology*, vol. 39, no. 6, pp. 603–611, 2000.

[3] K. G. Saag, G. G. Teng, N. M. Patkar, J. Anuntiyo, C. Finney, J. R. Curtis, H. E. Paulus, A. Mudano, M. Pisu, M. Elkins-Melton *et al.*, "American college of rheumatology 2008 recommendations for the use of nonbiologic and biologic disease-modifying antirheumatic drugs in rheumatoid arthritis," *Arthritis Care & Research: Official Journal of the American College of Rheumatology*, vol. 59, no. 6, pp. 762–784, 2008.

[4] V. Majithia and S. A. Geraci, "Rheumatoid arthritis: diagnosis and management," *The American journal of medicine*, vol. 120, no. 11, pp. 936–939, 2007.

[5] D. Van der Heijde, "How to read radiographs according to the sharp/van der heijde method." *The Journal of rheumatology*, vol. 27, no. 1, p. 261, 2000.

[6] J. Duryea, Y. Jiang, M. Zakharevich, and H. Genant, "Neural network based algorithm to quantify joint space width in joints of the hand for arthritis assessment," *Medical physics*, vol. 27, no. 5, pp. 1185–1194, 2000.

[7] G. Langs, P. Peloschek, H. Bischof, and F. Kainberger, "Automatic quantification of joint space narrowing and erosions in rheumatoid arthritis," *IEEE transactions on medical imaging*, vol. 28, no. 1, pp. 151–164, 2008.

[8] Y. Huo, K. L. Vincken, D. van der Heijde, M. J. De Hair, F. P. Lafeber, and M. A. Viergever, "Automatic quantification of radiographic finger joint space width of patients with early rheumatoid arthritis," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 10, pp. 2177–2186, 2015.

[9] T. Hirano, M. Nishide, N. Nonaka, J. Seita, K. Ebina, K. Sakurada, and A. Kumanogoh, "Development and validation of a deep-learning model for scoring of radiographic finger joint destruction in rheumatoid arthritis," *Rheumatology advances in practice*, vol. 3, no. 2, p. rkz047, 2019.

[10] Y. Ou, P. Ambalathankandy, T. Shimada, T. Kamishima, and M. Ikebe, "Automatic radiographic quantification of joint space narrowing progression in rheumatoid arthritis using poc," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. IEEE, 2019, pp. 1183–1187.

[11] K. Üreten, H. Erbay, and H. H. Maraş, "Detection of rheumatoid arthritis from hand radiographs using a convolutional neural network," *Clinical rheumatology*, vol. 39, no. 4, pp. 969–974, 2020.

[12] K. Nakatsu, K. Morita, N. Yagi, and S. Kobashi, "Finger joint detection method in hand x-ray radiograph images using statistical shape model and support vector machine," in *2020 International Symposium on Community-centric Systems (CcS)*. IEEE, 2020, pp. 1–5.

[13] A. Taguchi, S. Shishido, Y. Ou, M. Ikebe, T. Zeng, W. Fang, K. Murakami, T. Ueda, N. Yasojima, K. Sato *et al.*, "Quantification of joint space width difference on radiography via phase-only correlation (poc) analysis: a phantom study comparing with various tomographical modalities using conventional margin-contouring," *Journal of Digital Imaging*, vol. 34, no. 1, pp. 96–104, 2021.

[14] N. Asiri, M. Hussain, F. Al Adel, and N. Alzaidi, "Deep learning based computer-aided diagnosis systems for diabetic retinopathy: A survey," *Artificial intelligence in medicine*, vol. 99, p. 101701, 2019.

[15] G. Haskins, U. Kruger, and P. Yan, "Deep learning in medical image registration: a survey," *Machine Vision and Applications*, vol. 31, no. 1, pp. 1–18, 2020.

[16] T. K. Kvien, T. Uhlig, S. ØDEGÅRD, and M. S. Heiberg, "Epidemiological aspects of rheumatoid arthritis: the sex ratio," *Annals of the New York Academy of Sciences*, vol. 1069, no. 1, pp. 212–222, 2006.

[17] M. Jiang, Y. Li, C. Jiang, L. Zhao, X. Zhang, and P. E. Lipsky, "Machine learning in rheumatic diseases," *Clinical Reviews in Allergy & Immunology*, vol. 60, no. 1, pp. 96–110, 2021.

[18] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE transactions on systems, man, and cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[19] J. Sklansky and V. Gonzalez, "Fast polygonal approximation of digitized curves," *Pattern Recognition*, vol. 12, no. 5, pp. 327–331, 1980.

[20] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001*, vol. 1. IEEE, 2001, pp. I–I.

[21] K. Takita, T. Aoki, Y. Sasaki, T. Higuchi, and K. Kobayashi, "High-accuracy subpixel image registration based on phase-only correlation," *IEICE transactions on fundamentals of electronics, communications and computer sciences*, vol. 86, no. 8, pp. 1925–1934, 2003.

[22] T. Shimada, M. Ikebe, P. Ambalathankandy, M. Motomura, T. Asai *et al.*, "Sparse disparity estimation using global phase only correlation for stereo matching acceleration," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 1842–1846.

[23] K. Tamura, T. Fujita, and Y. Morisaki, "Vacuum-sintered body of a novel apatite for artificial bone," *Open Engineering*, vol. 3, no. 4, pp. 700–706, 2013.

[24] A. Pfeil, J. Böttcher, B. E. Seidl, J.-P. Heyne, A. Petrovitch, T. Eidner, H.-J. Mentzel, G. Wolf, G. Hein, and W. A. Kaiser, "Computer-aided joint space analysis of the metacarpal-phalangeal and proximal-interphalangeal finger joint: normative age-related and gender-specific data," *Skeletal radiology*, vol. 36, no. 9, pp. 853–864, 2007.

[25] Y. Ono, R. Kashihara, N. Yasojima, H. Kasahara, Y. Shimizu, K. Tamura, K. Tsutsumi, K. Sutherland, T. Koike, and T. Kamishima, "Tomosynthesis can facilitate accurate measurement of joint space width under the condition of the oblique incidence of x-rays in patients with rheumatoid arthritis," *The British Journal of Radiology*, vol. 89, no. 1062, p. 20150967, 2016.

[26] T. Aoki, M. Fujii, Y. Yamashita, H. Takahashi, H. Oki, Y. Hayashida, K. Saito, Y. Tanaka, and Y. Korogi, "Tomosynthesis of the wrist and hand in patients with rheumatoid arthritis: comparison with radiography and mri," *American Journal of Roentgenology*, vol. 202, no. 2, pp. 386–390, 2014.

[27] R. A. Brooks and G. Di Chiro, "Statistical limitations in x-ray reconstructive tomography," *Medical physics*, vol. 3, no. 4, pp. 237–240, 1976.
[28] D. A. Chesler, S. J. Riederer, and N. J. Pelc, "Noise due to photon counting statistics in computed x-ray tomography." *Journal of computer assisted tomography*, vol. 1, no. 1, pp. 64–74, 1977.
[29] C. Häger-Ross and M. H. Schieber, "Quantifying the independence of human finger movements: comparisons of digits, hands, and movement frequencies," *Journal of Neuroscience*, vol. 20, no. 22, pp. 8542–8550, 2000.