Automatic Joint Part Detection Method for Joint Space Measurement

Kosuke Goto and Tomio Goto
Dept. of Computer Science, Nagoya Institute of Technology, Nagoya, Japan
Email: airien0610@splab.nitech.ac.jp, t.goto@nitech.ac.jp

Koji Funahashi
Dept. of Orthopaedic Surgery, Kariya Toyota General Hospital, Kariya, Japan

***Abstract*—Early detection of rheumatoid arthritis is very important for its treatment. However, it can be difficult to detect changes in medical conditions by visually inspecting medical images. Computer-based applications that can support doctors' diagnoses can be helpful. In this study, we propose a diagnostic application based on computer-based recognition of features in medical images. Specifically, joint learning is performed by using Haar-like features those capture differences in brightness of joints. Furthermore, we aimed to improve detection accuracy by removing false positives based on pixel values and positional relationships of joint detection results. As a case study, we apply it to the detection of third finger joints in X-ray images of hands. The application is able to correctly identify these regions in most cases, thereby aiding doctors in the early detection of rheumatoid arthritis.**

***Index Terms*—rheumatoid arthritis, third finger joint detection, joint space**

I. Introduction

Medical and information processing technologies have both developed rapidly in recent decades. Accordingly, the use of medical imaging techniques such as endoscopy, MRI, and radiography has increased. Moreover, there is increasing demand for high-accuracy diagnosis based on such medical imaging. However, in practice, doctors often inspect medical images visually to determine medical conditions and make subjective evaluations before and after the detection of symptoms to judge progress. Therefore, there is the potential for errors in medical examination to occur, which is a serious problem to be solved. As mentioned earlier, there are many opportunities to use medical images. For example, multiple images are used to determine the progress of medical conditions. Therefore, judging multiple images for multiple patients visually imposes a great burden on doctors. In addition to such subjective visual evaluations, objective evaluations made using a computer are often required. Furthermore, in consideration of the effects of radiation on patients, X-ray and CT images are often taken rapidly, which limits their resolution [1]. If the resolution is too low, there is a possibility that false detections and oversights of medical conditions will increase. In addition, when low-resolution images are used in medical applications, errors are likely to occur. So, there are increasing demands for the development of medical applications that process medical images using computers. In our previous study, we developed a diagnostic application for early detection of rheumatoid arthritis by measuring distances between joints from X-ray images of hands [2]. However, there is a problem in this application, which has manual procedures by doctors, and it is a burden on doctors. Doctors have to cut out of joints from whole hands X-ray images, and edges (contours) of joints need to be selected by mouse clicks when measuring inter-articular cleft distances. Therefore, in this study, we focus on automatic cropping of joints from X-ray images and propose an automated application for detections of third finger joints.

II. Rheumatoid Arthritis

Rheumatoid disease is a general term for diseases accompanied by pain in the joints, muscles, bones, and ligaments that make up the motor organs. There are more than 100 types of rheumatic diseases, of which rheumatoid arthritis is a common one. It is an inflammatory disease that causes destructive deformation of joints and physical dysfunction by forming granulation tissue called *pannus* and destroying bones and cartilage via chronic and persistent inflammation of the synovium, which is called *synovitis*. The true cause of rheumatoid arthritis remains unknown but is related to genetic factors, environmental factors (such as infection and smoking), non-self antigens (foreign substances that trigger an immune response), and self-antigens. It is believed to accidentally act on other substances and tissues, leading to joint inflammation and tissue destruction [3], [4]. In the hand, the Joint Space Distance (JSD), which is the distance between joints, narrows and the joints tend to become distorted so that the fingers deform. Currently, the main diagnostic method is for doctors to directly observe X-ray images. However, visual diagnosis depends on subjective evaluation, which varies according to the doctor making the diagnosis. Therefore, it is necessary to develop an application that can

Manuscript received January 18, 2021; revised July 23, 2021.

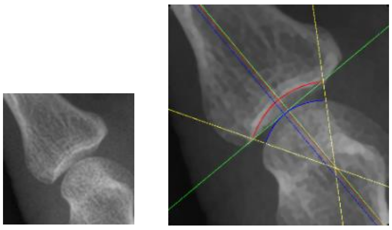
doi: 10.18178/joig.9.3.103-108

quantitatively diagnose symptom progression with high accuracy for objective evaluation.

III. Joint Space Distance Measurement Application

A joint space measurement application was developed in a previous work [2]. In this application, a joint image (e.g. Fig. 2 (a)) is selected from an X-ray image that includes both hands (e.g. Fig. 1). Then, the selected part is enlarged by the super-resolution method [5]-[9]. Next, we manually select the edges that form the joint on a computer and perform curve fitting, as shown in Fig. 2 (b). Finally, the JSD is measured by the fitting function with normal and integral calculations. Super-resolution enlargement enables high definition images, which enables very fine measurement of the joint cleft gap distance. Therefore, this application requires two manual operations: selecting the joints from the main image, then selecting their edges. As a result, there is some burden on doctors and potential for measurement error. Therefore, in this paper, we describe a method for automatically detecting the third joint part from an X-ray image of the whole hand using image recognition technology.

Figure 1. X-ray image of both hands.



(a) Input image (b) Output image

Figure 2. Measurement of joint space distance.

IV. Automation Overview

The procedure for creating joint images is first to locate the joints in an X-ray image of the entire hand, then specify areas to be cropped. In order to automate this operation, we learn features by machine learning to automatically recognize joint parts. However, this requires large numbers of correct and incorrect images for training. In a previous study, a single image was used for training by rotating it; however, the detection accuracy was poor. Therefore, in this paper, we propose a method that automatically creates a large number of images of the third finger joints as correct images and images of other bones as incorrect images. Furthermore, we improve detection accuracy by cropping images and applying a Gaussian filter.

A. Crop Detection of Target Images

Since the purpose of this study is automatic detection of third finger joints, as shown in Fig. 3, target images are clipped to exclude erroneous detections of parts away from the third joints. To select image regions for detection, we search the images with PSNR using the following procedure. 1) Enter a third joint image and a wrist part as a template. 2) Set a window size and vertical and horizontal shift widths. The window size is the same as the template image. 3) The window is applied from the upper-left corner of the X-ray image of the entire hand, and the PSNR between the image inside the window and the template image is calculated. This operation is repeated while moving the window in shift-width increments. After reaching the lower-right corner of the image, the image is cropped above the point where the PSNR is maximal. Fig. 4 shows a flowchart for this process. The learning-image creation algorithm described below uses the cropped image to create a learning image.

(a) Input image (b) Output image

Figure 3. Detection of target area for cropping.

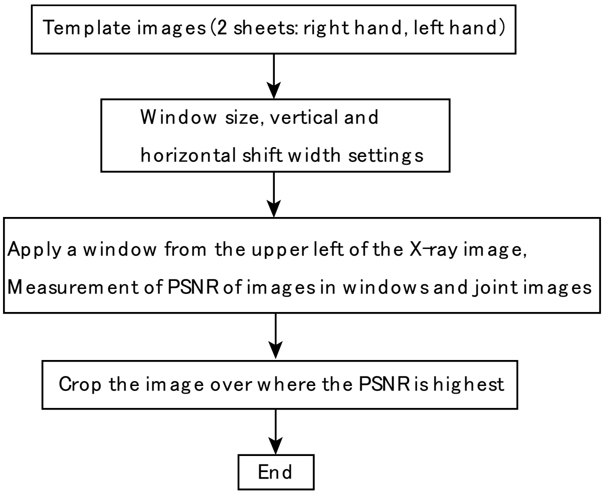


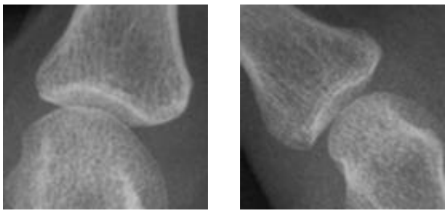
Figure 4. Flowchart for cropping target areas of images.

B. Learning Image Creation Algorithm

We describe a method for automatically creating learning images of third finger joints (correct images) and other bone parts (incorrect images).

1) Correct image creation algorithm

In the algorithm used to create correct images, the joint center coordinates of the third joints are first input. Subsequently, a region of $128\sqrt{2}$ pixels around the selected point is cut out and randomly rotated within the range of ± 20°. The center 128×128 pixel region of the rotated image is output as a correct image. As shown in Fig. 5, by using this process, we create an image that excludes black areas to eliminate adverse effects on learning.

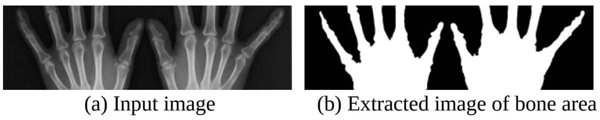


(a) Correct image 1 (b) Correct image 2

Figure 5. Examples of correct images.

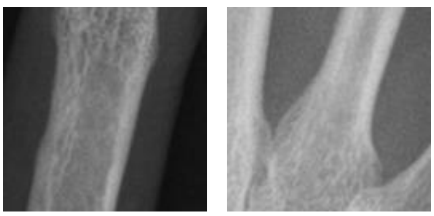
2) Incorrect image creation algorithm

In the algorithm used to create incorrect images, the joint center coordinates of the third joints are first input. Next, the bone region is extracted as shown in Fig. 6 because the incorrect images must not include areas outside the bone region. To extract bone areas, binarization processing is performed, then isolated points are removed by closing processing using morphological operations. Morphological operations can handle shapes of objects as those are. Therefore, it is applied in various forms to image analysis and is also used for feature extraction. The basic process of morphological operations includes *contraction processing* and *expansion processing*. Contraction processing is a process of referring to the eight pixels in the vicinity of a pixel of interest and replacing the lowest pixel value with that of the pixel of interest. Dilation processing is similar but replaces the maximum pixel value with the value of the pixel of interest. There are various types of ranges and shapes of neighboring pixels, which are called *structuring elements*. In this algorithm, a circular structuring element with a 5-pixel radius is used. Finally, in order to delete the unnecessary small area, the area below 50,000 pixels is deleted. The threshold for binarization is 60. Afterward, a point in the image is randomly selected; if the point is within the bone area and the distance to the third joint center point is $128\sqrt{2}$ or more (as shown in Fig. 7), then 128×128-pixel areas centered on those points are output as incorrect images. The centers of the images are included in bone areas, and parts away from the third joint center points can also be output as incorrect images. In other words, it is possible to output incorrect images that include various bone regions without including the third joints.



(a) Input image (b) Extracted image of bone area

Figure 6. Bone region extraction.



(a) Incorrect image 1 (b) Incorrect image 2

Figure 7. Examples of incorrect images.

C. Haar-like Feature

A requirement in automatic detection of joints is that it can be performed with any X-ray image. For that purpose, it is necessary to learn features with joints from many images to create a more adaptive detector. In X-ray images, the joints of the hands are characterized by the joint bones being bright and the gaps between joints being dark. Therefore, we thought it would be possible to capture the features of joints using Haar-like features based on differences in brightness [10]. In this paper, we used Haar-like features for training. These were obtained by applying patterns to learning images, as shown in Fig. 8 [11]. The patterns are applied to part of an image and the two parts are designated as A and B. Local edge and line components are detected from the brightness difference between areas A and B. Feature $H(r_1, r_2)$ can be obtained by calculating the average luminance difference between areas A and B using Equation (1).

$$H(r_1, r_2) = S(r_1) - S(r_2) \tag{1}$$

where $S(r_1)$ and $S(r_2)$ are the average luminance values of regions A and B, respectively.

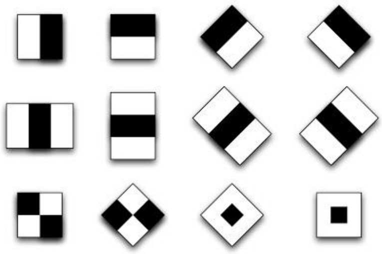


Figure 8. Haar-like patterns.

D. AdaBoost

A learning method called Adaboost was used for learning. In AdaBoost, different discriminators are created by sequentially updating the weights of learning samples, and the final discriminator is constructed by the weighted majority decision of these plural discriminators [12]. An individual classifier is called a Weak Classifier, and a combination of them is called a Strong Classifier. AdaBoost uses adaptive boosting, which improves the accuracy of the discriminator by updating the learning sample weights. In the recognition phase of image recognition, a classification is performed using a classifier created using the AdaBoost algorithm. Fig. 9 shows the classifier obtained by AdaBoost. A strong discriminator is constructed by combining T weak discriminators $h_t(x)$ that output the label y of the class to which the input pattern x belongs, weighted with reliability α_t. The final strong classifier $H(x)$ is given by the following equation.

$$H(x) = \sin\left[\sum_{t=1}^{T} \alpha_t h_t(x)\right] \tag{2}$$

where α_t is the reliability of the t^{th} weak classifier $h_t(x)$.

In addition, Viola and Jones used Haar-like features for weak classifiers and improved the accuracy and speed of recognition by cascading strong classifiers [13]. Normally, if an image can be scanned through the detection window, it is likely that something other than the object you want to detect will be found. Therefore, the first classifier performs classification using a very small number of weak classifiers for the detection window. When the first discriminator identifies the object to be detected, the second discriminator, which uses more weak discriminators than the first discriminator, is used. Similar processing is performed for the third and subsequent discriminators. The detection window that passed through the final classifier becomes the object to be detected. High-speed and high-accuracy detection is realized by first using a very small number of feature patterns for discrimination with loose criteria, then gradually tightening the criteria.

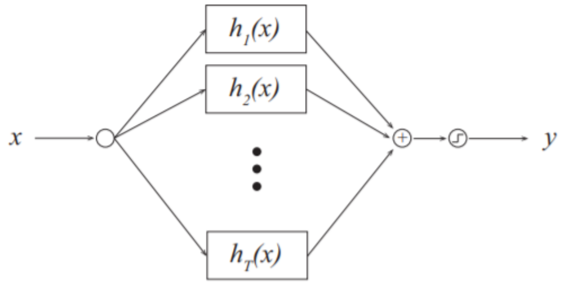


Figure 9. Structure of AdaBoost.

E. Applying a Gaussian Filter

In the field of character recognition, it has been reported that blurring an image can allow more features to be recognized than without blurring. Thus, it is expected that smoothing X-ray images and blurring them will allow more features to be captured, leading to improved accuracy. A Gaussian filter is used for smoothing. The moving average filter simply averages the brightness values around the pixel of interest. However, in most images, pixels close to the pixel of interest often have similar brightness, while more distant pixels have dissimilar brightness. Taking this into consideration, the Gaussian filter uses the Gaussian distribution function shown in Equation (3) so that the weight for calculation increases as it gets closer to the pixel of interest and decreases as it gets further away.

$$f(x,y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2+y^2}{2\sigma^2}\right) \qquad (3)$$

where σ represents the standard deviation of the Gaussian distribution. The larger the value, the greater the smoothing effect and the stronger blur that can be added. In this paper, it has been confirmed that accuracy is best when $\sigma = 2$. So, we used this value with the Gaussian filter, as shown in Fig. 10, which shows that an X-ray image can be blurred.

(a) Before application

(b) After application

Figure 10. Images before and after Gaussian filtering.

V. Experimental Results

A. Joint Detection Results

Joints were detected by learning from images created by a learning image creation algorithm and by creating a detector. As conditions, the learning was performed using 700 correct images and 300 non-correct images. It is normally assumed that several thousand learning images are needed, but since the detection accuracy did not improve with more than 700 images, we retained this number. Furthermore, since learning images are cropped to 128×128 pixels, the minimum detection size was set to 100×100 pixels, and the maximum detection size was set to 150×150 pixels to remove large and small false-positives with respect to correct detections. The experimental conditions are as shown in Table I. By using the learning images created by the learning image creation algorithm, detection accuracy was increased and all third joints could be detected, as shown in Fig. 11.

TABLE I. Numbers of Correct and Incorrect Images, and Minimum and Maximum Rectangle Size

Feature	Value
Number of correct images	700
Number of incorrect images	300
Minimum rectangle size	100×100 pixels
Maximum rectangle size	150×150 pixels

(a) Joint part detection 1

(b) Joint part detection 2

Figure 11. Examples of learning results.

However, erroneous detections are seen in the second joint of the little finger and the first joint of the thumb in Fig. 11. It is considered that these results were obtained because they had the same features as the third joints. Also, in Fig. 11(a), an erroneous detection occurs in the black portion in the center of the image, which is a problem. When the detection results of 28 X-ray images were examined, 10 false detections were found. Since all the third joints can be detected, it is considered that the degree of detection can be further increased by adding processing to eliminate such false detections. Therefore, we have since tried to deal with false-positives.

B. Elimination of False Detections due to Pixel Values

In the removal of erroneous detections, we first tried to remove the erroneous detections by using pixel values. In the X-ray images of the hands, the pixel values of the bone portions or white portions near the bones are around

90 to 120, and the dark portions (not bones) are around 50. Also, the pixel values of the other background portions are around 10–30. When the third joints are normally detected, most of the detection rectangles should contain bone and the bright parts around it. Therefore, a detection rectangle including many black portions can be erroneously detected. Using the pixel values of the X-ray images, we remove rectangles containing ≥ 10% pixels with values < 60. By adding this process, it is possible to eliminate erroneous detections, including dark portions as shown in Fig. 12. After applying this process to 28 X-ray images, there were seven false detections, of which three could be deleted. However, false-positives that did not include dark areas, such as the first joint of the thumb, could not be ruled out.

(a) Before processing

(b) After processing

Figure 12. Elimination of false detections based on pixel values.

C. *Elimination of False Detections due to Positional Relationships*

By eliminating false detections with dark areas based on pixel values, detection accuracy is improved. However, this process does not remove false detections without dark areas, such as the first joints of the thumb. Next, we remove erroneous detections based on positional relationships. False-positives are often found in regions containing the first joints of the thumb and the second joints of the little finger. In other words, there may be a false-positive detection above a correctly detected region in an X-ray image. We use this to eliminate false-positives in the following three steps.

1) Calculate the difference between the x-coordinates of two consecutive detection rectangles;
2) If the difference is < 70, compare the y-coordinates of those rectangles;
3) Remove the rectangle with the smaller y-coordinate.

Step 1 is applied to all detections. Since the rectangle size for joint detections is around 130 pixels, the difference between consecutive x-coordinates is usually at least 130 pixels. However, if there are false-positives, the difference will be less. In this paper, we determine false detections using a threshold difference between consecutive x-coordinates of ≤ 70 pixels. If the difference between consecutive x-coordinates is < 70 pixels, the false detection will be located above the correct detection in the image, as described above. Then, the y-coordinates are compared and the region with the lower value is removed. This process can eliminate false-positives above the correct answer rectangle as shown in Fig. 13.

By adding this process, all false detections were removed from the 28 X-ray images and all third joints were detected correctly.

(a) Before processing

(b) After processing

Figure 13. Elimination of a false detection based on positional relationship.

VI. Conclusion

Rheumatoid arthritis destroys bones and cartilage as the symptoms progress, resulting in a narrowing of joint fissure distances. Based on this feature, an application that measures joint gap distances was developed in a previous study. However, two manual operations were required, creating a burden for doctors. The first operation was to manually delineate the joint regions of an X-ray image of both hands based on the doctor's knowledge, and the second was to select bone edges to measure the joint gap distances. The purpose of this study was to automate the first operation.

To do this automatically, it is necessary for a computer to recognize joints using image recognition technology. We aimed to automate the detection of third joint parts by learning their features. By using learning images created by the learning image creation algorithm, a highly accurate detector was created. In addition, by adding an erroneous detection removal process based on pixel values and positional relationships, it was possible to detect all third joints in 28 X-ray images without any erroneous detections. Further research will confirm the detection accuracy of this process by applying it to more X-ray images. In addition, when measuring joint fissure distances, it is still necessary to select the joint edges manually. Automating this process would also be useful.

Conflicts of Interest

The authors declare no conflicts of interest associated with this manuscript.

Author Contributions

Mr. Goto designed and executed the experiments and wrote the manuscript, and Dr. Goto is a supervisor and edited the manuscript. Dr. Funahashi is also a supervisor for medical viewpoints. All authors reviewed and approved the final manuscript.

REFERENCES

[1] G. N. Hounsfield, "Computerized transverse axial scanning (Tomography) Part I. Description of system," *British Journal of Radiology*, vol. 46, pp. 1016-1022, 1973.

[2] M. Shimizu, H. Kariya, T. Goto, S. Hirano, and M. Sakurai, "Super-Resolution for X-ray images," in *Proc. IEEE 4th Global Conference on Consumer Electronics*, 2015, pp. 246-247.

[3] T. Takeuchi and H. Kameda, "The Japanese experiences with biologic therapies for rheumatoid arthritis," *Nature Reviews Rheumatology*, vol. 6, pp. 544-562, 2010.

[4] D. Aletaha, T. Neogi, A. J. Silman, J. Funovts, *et al.*, "2010 rheumatoid arthritis classification criteria: An American college of Rheumatology/European league against rheumatism collaborative initiative," *Arthritis & Rheumatism*, vol. 62, no. 9, pp. 2569-2581, 2010.

[5] L. Rudin, S. Ohser, and E. Fetami, "Nonlinear total variation based noise removal algorithm," *Physica D*, vol. 60, pp. 259-268, 1992.

[6] W. T. Freeman, T. R. Jones, and E. C. Paztor, "Example based super-resolution," *IEEE Computer Graphics and Applications*, vol. 22, no. 2, pp. 56-65, 2002.

[7] J. Sun, Z. Xu, and H. Y. Shum, "Image super-resolution using gradient profile prior," in *Proc. IEEE International Conference on Computer Vision and Pattern Recognition*, 2008, pp. 1-8.

[8] P. P. Gajjar and M. V. Joshi, "New learning based super-resolution: Use of DWT and IGMRF prior," *IEEE Transactions on Image Processing*, vol. 19, no. 5, pp. 1201-1213, 2010.

[9] J. Sun, Z. Xu and H. Y. Shum, "Gradient profile prior and its applications in image super-resolution and enhancement," *IEEE Transactions on Image Processing*, vol. 20, no. 6, pp. 1529-1542, 2011.

[10] C. P. Papageorgiou, M. Oren, and T. Poggio, "A general framework for object detection," in *Proc. IEEE International Conference on Computer Vision*, 1998, pp. 555-562.

[11] B. Wu, H. Ai, C. Huang, and S. Lao, "Fast rotation invariant multi-view face detection based on real AdaBoost," in *Proc. IEEE International Conference on Automatic Face and Gesture Recognition*, 2004, pp. 79-84.

[12] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of Computer and System Sciences*, vol. 55, pp. 119-139, 1997.

[13] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2001, pp. 511-518.

Kosuke Goto received a B.S. degree from the Department of Electrical and Electronic Engineering, Nagoya Institute of Technology, in 2019. He is currently pursuing an M.S. degree from the Department of Computer Science, Nagoya Institute of Technology, Japan. His research interests are signal and image processing and medical image processing.

Tomio Goto (M'01) received B.S., M.S., and Ph.D. degrees from the Nagoya Institute of Technology in 1997, 1999, and 2007, respectively. He is currently an Associate Professor at the Department of Computer Science, Nagoya Institute of Technology, Japan. His research interests include coding of images, non-linear filtering, noise reduction, and super-resolution, as well as most other applications of digital image processing. He is a senior member of the IEEE, IEICE, ITE, and IEEJ in Japan.

Koji Funahashi received his Dr. degree from the Nagoya University in 2004. He is currently a Cheif Director at the Department of Orthopaedic Surgery, Kariya Toyota General Hospital, Japan.