# Automatic Finger Joint Detection for Volumetric Hand Imaging

Johannes Bopp[1], Mathias Unberath[1,2], Stefan Steidl[1], Rebecca Fahrig[3], Isabelle Oliveira[4], Arnd Kleyer[4], Andreas Maier[1,2]

[1]Pattern Recognition Lab, FAU Erlangen-Nürnberg, Germany
[2]Graduate School in Advanced Optical Technologies, Erlangen, Germany
[3]Radiological Sciences Lab, Stanford University, Stanford, USA
[4]University Hospital, FAU Erlangen-Nürnberg, Germany
Johannes.Bopp@fau.de

**Abstract.** We propose a fully automatic method for robust finger joint detection in T1 weighted magnetic resonance imaging (MRI) sequences for initialization of statistical shape model (SSM) based segmentation. We propose a robust method that only relies on few training samples. Therefore, a parallel-beam forward projection is calculated on the MRI volume. A trained Bagging classifier will detect the joints in 2D which are then splatted into the 3D volume. For evaluation, leave-one-out cross validation was performed. The detection of the joints in 2D yielded a Dice score of $0.67 \pm 0.056$ with respect to a manually obtained ground truth. For the initialization of SSM-based segmentation algorithms, the results are very promising.

## 1 Introduction

Magnetic resonance imaging (MRI) is the best imaging modality to monitor rheumatoid arthritis and is especially used for difficult diagnosis. A typical acquisition consists of T1 and T2 weighted sequences that allow assessment of the three main pathologies associated with the disease: synovitis, edema, and erosions. For the assessment the EULAR-OMERACT system was introduced. In this atlas-based assessment the physician estimates the ratio between pathological bone and healthy bone in 10% steps [1]. Because of the large slice thickness of 2.75 mm and the subjectivity of the evaluation, volumetric assessment is challenging. This may give rise to a large inter-observer variability, indicating that EULAR-OMERACT scores are hardly ever comparable. Therefore, automatic approaches to arthritis detection may be advantageous. In order to promote such methods, automatic hand bone segmentations are needed. As arthritis may heavily affect bone appearance, statistical shape models (SSMs) are a promising way to estimate the outline of the non-diseased bone as they learn valid behavior from a correctly annotated training set [2]. SSM based segmentation algorithms have successfully been applied in bone age assessment [3]. However, these algorithms require reliable initialization. For the wrist bones Koch *et al.* proposed a marginal space learning based algorithm trained with 100 manually segmented

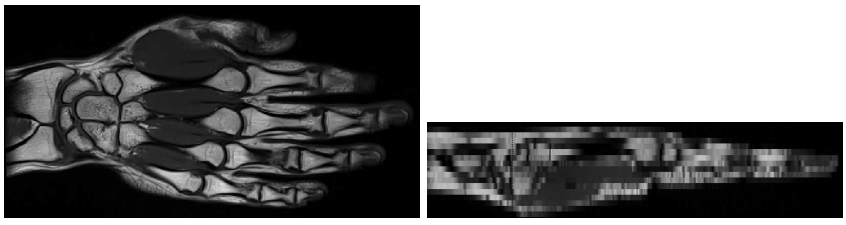
training samples [4]. The large amount of training samples needed for marginal space learning based approaches makes them impractical, as the manual annotation is cumbersome and time consuming. Instead of initialization via marginal space learning, we propose to automatically detect the finger joints. Therefore, a parallel beam forward projection is performed on the MRI volume. Then a trained classifier detects the joints in the 2D projections. From the 2D locations the corresponding 3D joints will be searched. The pipeline is implemented in CONRAD, a Java based software framework original developed for cone-beam imaging in radiology [5].

## 2 Materials and Methods

### 2.1 Data

The data set consists of 10 volumetric images with a spacing of 0.5 mm ×0.5 mm ×2.75 mm showing the right hand in coronal view (Figure 1). All scans were acquired in a T1 weighted turbo-spin echo sequence. Therefore, bone-marrow and trabecular bone appear bright while cortical bone is not visible at all. To train the classifier for the joint detection a ground truth has to be created manually in the 2D forward projections. These labels will also be used later as ground truth for evaluation. For the evaluation in 3D the bones are segmented manually.

**Fig. 1.** (a) A representative coronal slice of the volumetric MRI images is shown. The large slice thicknessof 2.75 mm in (b) complicates volumetric segmentation.



(a) (b)

### 2.2 Parallel-Beam Forward Projection

To automatically detect the finger joints, first a parallel-beam forward projection is performed on the MRI volume. Therefore, a summation of voxel intensities $\boldsymbol{u}(x, y, z)$ along the ray (in z direction) is performed:

$$\boldsymbol{v}(x, y) = \sum_{z} \boldsymbol{u}(x, y, z). \tag{1}$$

After the projection we apply our prior knowledge, that the finger joints appear in the upper 2/3 of the image, and crop the image accordingly. Finally the

projection $\boldsymbol{v}(x, y)$ is normalized. An example forward projection can be seen in Figure 2(a). In these projection images the finger joints are labeled manually.

### 2.3 Classification and Feature Selection

To train a classifier for the detection of finger joints in the projected 2D image the WEKA workbench is used with 5 different features [6]. All features are calculated on a Gaussian Pyramid allowing for a multi-scale representation. The following features are used and can be categorized as following:

- Gaussian blur (voxel intensity-averaging),
- Sobel filter (edge detection),
- Hessian matrix (orientation),
- difference of gaussians (blob detection),
- membrane projection (line detection) [7].

For the Gaussian blur a convolution with an isotropic Gaussian kernel

$$P(\boldsymbol{r}) = \frac{1}{\sigma\sqrt{2\pi}} e^{-||\boldsymbol{r}-\boldsymbol{\mu}||_2^2/2\sigma^2} \tag{2}$$

is performed at all positions $\boldsymbol{r}$ using different standard deviations $\sigma$. The larger the standard deviation the larger the affected area of the kernel which results in smoother pixel values.

For edge detection features the Sobel filter is used. It is a discrete differentiation operator which calculates an approximation of the gradient of the image intensity function. The Sobel filter is based on two convolutions while $*$ is the convolution operator. The resulting approximation of the gradient in $x$ direction ($\boldsymbol{G}_x$) is calculated with the mask $\boldsymbol{S}_x$ in $x$ direction and the gradient $\boldsymbol{G}_y$ in $y$ direction is calculated with the transposed of the mask in $x$ direction. Both are displayed in the following equations

$$\boldsymbol{S}_x = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ +1 & +2 & +1 \end{bmatrix}, \qquad \boldsymbol{G}_x = \boldsymbol{S}_x * \boldsymbol{v}, \qquad \boldsymbol{G}_y = \boldsymbol{S}_x^T * \boldsymbol{v}. \tag{3}$$

the gradient magnitude can be calculated by

$$\boldsymbol{G}(x, y) = \sqrt{\boldsymbol{G}_x(x, y)^2 + \boldsymbol{G}_y(x, y)^2}, \tag{4}$$

and the gradient direction

$$\boldsymbol{\Theta}(x, y) = \arctan\left(\frac{\boldsymbol{G}_y(x, y)}{\boldsymbol{G}_x(x, y)}\right) \tag{5}$$

can be calculated. The third feature is derived from the Hessian matrix

$$H(\boldsymbol{r}) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1}(\boldsymbol{r}) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(\boldsymbol{r}) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(\boldsymbol{r}) & \frac{\partial^2 f}{\partial x_2 \partial x_2}(\boldsymbol{r}) \end{pmatrix}. \tag{6}$$

The features are based on the second derivative of the image intensity function. As main feature the first and the second eigenvalue as well as the orientation are calculated.

The difference of Gaussians is similar to the Hessian matrix, both are based on the second derivative since the Laplace operator can be approximated by the difference of Gaussians. It is commonly used to detect "blob"-like structures.

The fifth feature is the membrane projection which was originally developed for detecting cell membranes. The image is convolved with 30 different kernels. Each kernel shows a straight line at a distinct slope. The resulting 30 images are z-projected into one single image via 5 different methods using the

- sum of pixels in each image,
- standard deviation of the pixels in each image,
- median of the pixels in each image,
- maximum of the pixels in each image, and
- minimum of the pixels in each image.

As finger joints appear as dark lines in between bright bones, line detection filters such as membrane projections may yield high responses.

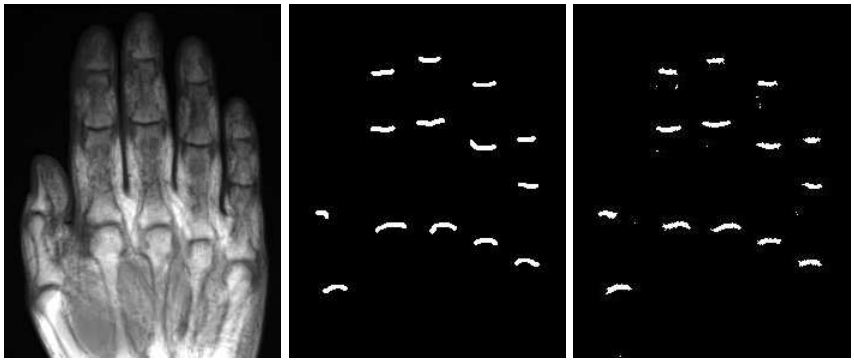
### 2.4 Classifier

With the features a combined classifier by bootstrap aggregation (Bagging) is trained. The idea is to produce and combine a variety of different weak classifiers to a strong one. Therefore, the base-learners are trained over varying training sets to make them different. By bootstrapping, many samples are generated out of a single sample by randomly picking a subset of samples from the original ensemble with replacement. That means some samples may be repeated while others are never drawn. After generating different training sets, the base learners are trained on these sets with an unstable training procedure. The outcomes of the base learners are averaged during testing to generate a single prediction. In bagging every model has the same weight, so it is not important if the model was more or less successful [8]. For this experiment REPTrees were used as base-learners [9].

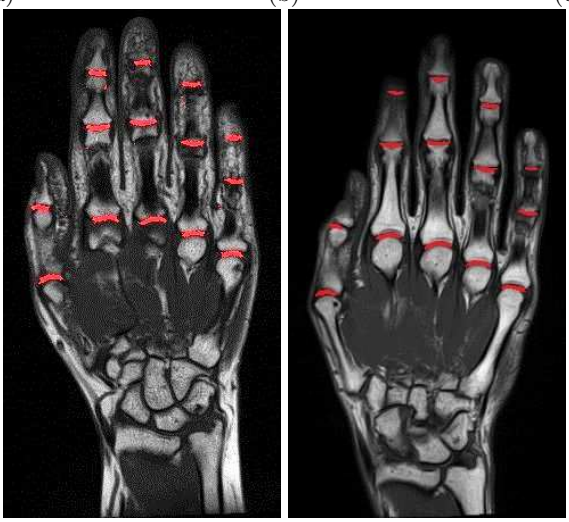
## 3 Experiment

For the experiment 10 bagging classifiers for the leave-one-out cross validation are trained on the projected images with the WEKA workbench using the features in Section 2. Because of the imbalance of the two classes, joint and no joint, the feature stack was balanced. The manually segmented joints serve as a ground truth for training and evaluation. As ground truth for the 3D detection, manually labeled hand bones are used. For the evaluation of the detection the Dice score is calculated both in 2D and in 3D giving a score for the overlap. For the 2D case a high overlap between the ground truth and the predicted joints is best which results in a higher Dice score. In the 3D case, however, since the joints are in between the bones, we want a low overlap with the segmented bones in the 3D volume, indicating that a low Dice score is best.

**Fig. 2.** An example of the parallel-beam forward projection is shown in (a). In (b) the ground truth is displayed, while in (c) the result of the classification is shown. In (d) and (e) the detected joints are reconstructed into the 3D volume. Figure 2(d) shows the volume corresponding to the prediction shown in (c), while Figure 2(e) shows the 3D result for a patient not previously shown.



(a) (b) (c)



(d) (e)

## 4 Results

An example of the result of the experiment for the 2D case is shown in Figure 2. A classifier that did not have this example in the training set received the parallel-beam projection (Figure 2(a)) of the original volume. The resulting predictions are shown in Figure 2(c). For comparison the ground truth is displayed in Figure 2(b). The average Dice score over all 10 trained classifiers was 0.67 with a standard deviation of 0.056.

For the 3D case the average dice score was 0.025 with a standard deviation of 0.0077.

## 5 Discussion

We found a robust method to detect joints in T1 weighted MRI volumes of the hands of patients suffering from rheumatoid arthritis. Even if the Dice score of 0.67 leaves room for improvements, the false positive rate is low as can be understood from Figures 2(b) and 2(c). This is crucial for robust initialization as we are mostly interested in the location of the joint and not so much in its overall size. To increase the Dice score and to get a better volumetric segmentation of the joints more features could be used for the training set to get better results. In the next step the detected joints will be skeletonized before the reconstruction into the 3D volume. They will be used for the initialization of the SSMs for volumetric bone segmentation.

## References

1. Østergaard M, Edmonds J, McQueen F, Peterfy C, Lassere M, Ejbjerg B, et al. An introduction to the EULAR–OMERACT rheumatoid arthritis MRI reference image atlas. Annals of the Rheumatic Diseases. 2005;64(suppl 1):i3–i7.
2. Unberath M, Maier A, Fleischmann D, Hornegger J, Fahrig R. Open-Source 4D Statistical Shape Model of the Heart for X-ray Projection Imaging. Biomedical Imaging, 2015 IEEE 12th International Symposium on. 2015; p. 739–742.
3. Thodberg H, Kreiborg S, Juul A, Pedersen K. The BoneXpert method for automated determination of skeletal maturity. Medical Imaging, IEEE Transactions on. 2009;28(1):52–66.
4. Koch M, Schwing A, Comaniciu D, Pollefeys M. Fully automatic segmentation of wrist bones for arthritis patients. In: Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on. IEEE; 2011. p. 636–640.
5. Maier A, Hofmann H, Berger M, Fischer P, Schwemmer C, Wu H, et al. CONRAD – A software framework for cone-beam imaging in radiology. Medical physics. 2013;40(11):111914.
6. Holmes G, Donkin A, Witten I. Weka: A machine learning workbench. In: Intelligent Information Systems, 1994. Proceedings of the 1994 Second Australian and New Zealand Conference on. IEEE; 1994. p. 357–361.
7. Staniewicz L, Midgley P. Machine learning as a tool for classifying electron tomographic reconstructions. Advanced Structural and Chemical Imaging. 2015;1(1):1–15.
8. Breiman L. Bagging predictors. Machine learning. 1996;24(2):123–140.
9. Witten IH, Frank E. Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann; 2005.