Segmentation of hand radiographs by using multi-level connected active appearance models

Joost A. Kauffman^a, Cornelis H. Slump^a and Hein J. Bernelot Moens^b

 ^aSignals and Systems group, Faculty of Electrical Engineering, Mathematics and Computer Science, University of Twente, Enschede, The Netherlands;
 ^bDepartment of Rheumatology, Hospital Group Twente, Hengelo, The Netherlands

ABSTRACT

Robust and accurate segmentation methods are important for the computerized evaluation of medical images. For treatment of rheumatoid arthritis, joint damage assessment in radiographs of hands is frequently used for monitoring disease progression. Current clinical scoring methods are based on visual measurements that are time-consuming and subject to intra and inter-reader variance. A solution may be found in the development of partially automated assessment procedures. This requires reliable segmentation algorithms.

Our work demonstrates a segmentation method based on multiple connected active appearance models (AAM) with multiple search steps using different quality levels. The quality level can be regulated by setting the image resolution and the number of landmarks in the AAMs.

We performed experiments using two models of different quality levels for shape and texture information. Both models included AAMs for the carpal region, the metacarpals, and all phalanges. By starting an iterative search with the faster, low-quality model, we were able to determine the initial parameters of the second, highquality model. After the second search, the results showed successful segmentation for 22 of 30 test images. For these images, 70% of the landmarks were found within 1.3 mm difference from manual placement by an expert.

The multi-level search approach resulted in a reduction of 50% in calculation time compared to a search using a single model. Results are expected to improve when the model is refined by increasing the number of training examples and the resolution of the models.

1. INTRODUCTION

Hand radiographs are often used to assess joint damage in patients diagnosed with rheumatoid arthritis. Rheumatoid arthritis (RA) is a chronic inflammatory disease that causes pain, swelling, stiffness, and loss of function in the joints. The disease usually affects the joints, particularly in the wrist and the fingers. Also other parts of the body besides the joints can be affected. Since there is no proven cure for RA available yet, current treatments mainly focus on pain relief, inflammation reduction, and slowing down or stopping joint damage.

To prevent irreversible joint damage, it is essential to detect RA at an early stage. For an effective medical treatment it is important that the disease can be monitored closely. Joint damage assessment in hand radiographs is a frequently used method for monitoring the progression of RA. Several proposed scoring methods¹ are based on measurements of carpal/metacarpal ratio,² joint space widths (JSW), erosion volumes, and the use of damage classification schemes.^{3–5} In general, these methods are time-consuming and depend on subjective measurements. Especially in an early stage of RA, it is important that the applied scoring method is sensitive to change, so the effects of treatments can be monitored closely and the treatments altered if necessary. Several studies have been conducted into this subject.^{6–8}

Our aim is to develop an automated assessment procedure with a higher sensitivity to change compared to manual measurements. This paper shows a first step towards this development by presenting a segmentation algorithm which can identify the bones of the hand skeleton and outline their contours. In future research we will use this algorithm as a starting point for developing automated scoring methods.

Further author information: (Send correspondence to J.A.K.)

J.A.K.: E-mail: j.a.kauffman@utwente.nl, Telephone: $+31\ (0)\ 53\ 489\ 2673$

Our proposed segmentation algorithm is based on an iterative search method with active appearance models (AAM).⁹ AAMs are deformable models that include information about both shape and texture. These models are trained statistically and are suitable for detecting objects with large variability. This property may be useful in cases where joint damage is present (Figure 1). Due to severe erosions, the JSW can disappear, which makes it difficult to find a joint with local image processing techniques. Since AAM search algorithms use shape and



Figure 1. Result of joint damage: healthy joint (left) with JSW clearly visible and severely damaged joint (right).

texture information about the total image in combination with prior trained knowledge, these locations may be estimated from the fitted model. Other research with AAMs on parts of hand radiographs has been described $in^{10,11}$.

In this work, we use multiple connected AAMs^{12, 13} to model each bone (or combination of bones for the carpals), and connect these in a single model of the entire hand. The reason for this is that we want to model the bone shape variabilities independently from hand positioning. An important part of the variability in hand radiographs is caused by differences in hand and finger positioning. To reduce the calculation effort in determining these general hand features, we start out with a simplified model. After convergence of the search error by using this faster, low-quality model, we switch to a slower, higher quality model to find the bone edges. The following section will describe our segmentation algorithm.

2. SEGMENTATION

For segmenting hand radiographs, we are interested in finding the locations and the contours of the bones. For this purpose, we create a set of AAMs from a training set of 50 hand radiographs. This results in a statistical model that can be fitted to a series of images in a test set. Since we know the properties of our model, we can find the locations of the bones in which we are interested and estimate their contours.

2.1. Modelling the hand

The images that we used for this research are digital scans from film radiographs. The scanning resolution is 600 dpi (in both directions) with 16-bit gray values (effectively 12-bit), and images were stored without using lossy image compression. We created the hand model from a set of 50 example radiographs (anteroposterior view) of single hands. To be able to use both left-hand and right-hand images in a single model, we mirrored all left-hand images.

2.1.1. Landmarking

We have created two models from the training set; a low-quality (LQ) model and a high-quality (HQ) model. For the HQ model, we use manually outlined contours in each training image to set the shape parameters of the sub-AAMs. With the help of a 'fish bone' shaped grid (Figure 2), we determine 64 landmarks on each of the contours of the metacarpals, the proximal phalanges and the middle phalanges. The distal phalanges are



Figure 2. A 'fish bone' shaped grid and the manually outlined contour define the landmark locations. The grid is placed manually by selecting the proximal and distal ends of a bone. The origins of the radiating grid lines are located at 15% of the bone length from both ends.



Figure 3. The left image shows the landmark locations for the LQ model and the right for the HQ model. The dashed lines show the sub-AAM boundaries.

considered to be less important for joint damage assessment and are therefore marked by only four landmarks. We mark the carpal region by ten landmarks (Figure 3). To each bone shape we add four extra landmarks. These points define the corners of the bone's bounding box and are also used to constrain the texture patch of the model. By doing this we include more information about the texture around the edges of the contour in the model. For the LQ model, we only use the corners of the bounding boxes to define the shapes.

Since a hand radiograph is a two-dimensional projection of a three-dimensional collection of bones, parts of bones may overlap one another. To model this effect, it is necessary to separate the model into submodels. There-fore we model each bone, except for the carpal bones, separately in a sub-AAM. In the current implementation we use 20 submodels; the carpal bones are collected in a single sub-AAM.

2.1.2. Overall alignment

Before determining the AAM submodels, we align the combined shapes to one another through Procrustes analysis¹⁴ to make the model invariant to scaling, rotation and translation. For this overall alignment we only make use of the landmarks of metacarpals 2-5, since the shapes of these bones are roughly independent of finger positioning.

2.1.3. Submodel alignment

Next, we split models into submodels and make our calculations separately for each bone. For the anteroposterior projection of the hand we consider seven joints that can introduce lateral positioning variability. Figure 4 shows these joints as dots in a wireframe model. Differences in finger positioning cause concurrent rotational and translational changes in the relationships between submodels. Using a single (x,y)-coordinate frame for the hand model, these positioning variabilities introduce nonlinearities. This makes it difficult to create an accurate model, especially with few training examples. For the LQ model this is not important, since we only use this to roughly determine the locations of the joints. For the HQ model we try to compensate for this.

To make the HQ model invariant to positioning variabilities, we disconnect the bone shapes that are connected to these joints and give them new coordinate frames. This results in a parent-child structure with seven dependent sub-coordinate frames, as shown in Figure 4. The proximal ends of the bones connected to these joints are used as the origin of the sub-coordinate frames. Next, we align all shapes by translating and rotating them to their mean origin and angle with respect to their parent's coordinate frame. Figure 5 shows how these alignment operations reduce the shape variabilities in our training set.



Figure 4. Wireframe model of the hand with connected submodels. The dots mark the joints that enable lateral displacement of the fingers. Seven dependent sub-coordinate frames (1.1-5, 1.1, 1.1.1) are used to make the model invariant to changes in finger positioning.

2.1.4. Shape and texture modelling

After aligning the shapes, we create appearance models for the remaining shape and texture variabilities in our training set.

The x- and y-coordinates of the landmarks of each submodel are stored in a 2L element vector \mathbf{x} (L is the number of landmarks).

$$\mathbf{x} = (x_1, x_2, \dots, x_L, y_1, y_2, \dots, y_L)^{\mathrm{T}}$$
(1)

The mean shape $\bar{\mathbf{x}}$ is calculated for each submodel and is used for creating a shape-free texture patch to which we can map the grey intensities of the images.

$$\bar{\mathbf{x}} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{x}_{i} \tag{2}$$

For the HQ model, we set the texture resolution of the model to 15 pixels per cm and for the LQ model to 8 pixels per cm. These resolution reductions will limit the models' accuracies, but are necessary to reduce the calculation effort needed. On the other hand, it also makes the model more robust for small texture variations, since noise and small defects are also reduced. Thin-plate spline warps¹⁵ are used to warp each training image to the mean shape of the sub-AAM. To compensate for differences in image contrast, we normalize the intensities and remove any offset. For every image, the pixel intensity values of the shape-free patch are stored in a texture vector \mathbf{g} of length M. We also calculate the mean texture $\bar{\mathbf{g}}$.

$$\mathbf{g} = (g_1, g_2, \dots, g_M)^{\mathrm{T}} \tag{3}$$

$$\bar{\mathbf{g}} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{g}_i \tag{4}$$

Next we join the submodels together again by concatenating the shape and texture vectors in \mathbf{x}_{tot} and \mathbf{g}_{tot} , and their means in $\mathbf{\bar{x}}_{tot}$ and $\mathbf{\bar{g}}_{tot}$:

$$\mathbf{x}_{\text{tot}} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \dots \\ \mathbf{x}_{20} \end{pmatrix}, \ \, \bar{\mathbf{x}}_{\text{tot}} = \begin{pmatrix} \bar{\mathbf{x}}_1 \\ \bar{\mathbf{x}}_2 \\ \dots \\ \bar{\mathbf{x}}_{20} \end{pmatrix}$$
(5)



Figure 5. Wireframes of training set before (left) and after (right) aligning the submodels to their sub-coordinate frame. This shows how the the variability caused by differences in finger positioning is extracted from the HQ model.

$$\mathbf{g}_{\text{tot}} = \begin{pmatrix} \mathbf{g}_1 \\ \mathbf{g}_2 \\ \dots \\ \mathbf{g}_{20} \end{pmatrix}, \ \mathbf{\bar{g}}_{\text{tot}} = \begin{pmatrix} \mathbf{\bar{g}}_1 \\ \mathbf{\bar{g}}_2 \\ \dots \\ \mathbf{\bar{g}}_{20} \end{pmatrix}$$
(6)

We express the differences between the training examples and the mean in \mathbf{x}_{tot} and \mathbf{g}_{tot} :

$$\tilde{\mathbf{x}}_{tot} = \mathbf{x}_{tot} - \bar{\mathbf{x}}_{tot} \tag{7}$$

$$\tilde{\mathbf{g}}_{tot} = \mathbf{g}_{tot} - \bar{\mathbf{g}}_{tot} \tag{8}$$

To combine shape and texture information in a single model, we concatenate these vectors in vector \mathbf{b} . The ratio w of the mean total intensity variance to the mean total shape variance is used to weight the shape information:

$$\mathbf{b} = \begin{pmatrix} w \tilde{\mathbf{x}}_{\text{tot}} \\ \tilde{\mathbf{g}}_{\text{tot}} \end{pmatrix} \tag{9}$$

$$w = \frac{\operatorname{var}(\tilde{\mathbf{g}}_{\text{tot}})}{\operatorname{var}(\tilde{\mathbf{x}}_{\text{tot}})}$$
(10)

Principal component analysis $(PCA)^{16}$ is used on the vectors **b** from the training set to reduce the number of variations in the models. For our experiments, we reduced the LQ model to 30 parameter variations covering 89% of the variance displayed in the training set. The HQ model, without the positioning variabilities, was reduced to 25 parameters covering 85%.

$$\mathbf{b} \approx \mathbf{P}_c \mathbf{c} \tag{11}$$

Now \mathbf{P}_c is a matrix of eigenvectors and \mathbf{c} can be used to control the shape (\mathbf{x}_{synt}) and texture (\mathbf{g}_{synt}) parameters:

$$\begin{pmatrix} \mathbf{x}_{\text{synt}} \\ \mathbf{g}_{\text{synt}} \end{pmatrix} = \begin{pmatrix} \bar{\mathbf{x}}_{\text{tot}} \\ \bar{\mathbf{g}}_{\text{tot}} \end{pmatrix} + \mathbf{P}_c \mathbf{c}$$
(12)

The resulting models are used for creating synthetic hand images. Since the models consist of submodels that may have a certain overlap, the overlapping regions of texture are averaged. By tuning the models' parameters c the modes of variation can be made visible (Figure 6).

In case of the HQ model we still need to add several pose parameters (rotation and translation) for the submodels that are in a sub-coordinate frame as mentioned in section 2.1.3. Since there are seven sub-coordinate frames for this model, we have to add 21 pose parameters (t').

$$\mathbf{c}_{\mathrm{HQ}} = \begin{pmatrix} \mathbf{c} \\ \mathbf{t}' \end{pmatrix} \tag{13}$$

Finally, since the current model has been made invariant to overall changes in pose (translation, rotation and scaling) and intensity, we complete the model by adding parameters for pose (t) and intensity (u) transformations:

$$\mathbf{p}_{\mathrm{LQ}} = \begin{pmatrix} \mathbf{c} \\ \mathbf{t} \\ \mathbf{u} \end{pmatrix}, \tag{14}$$

$$\mathbf{p}_{\mathrm{HQ}} = \begin{pmatrix} \mathbf{c}_{\mathrm{HQ}} \\ \mathbf{t} \\ \mathbf{u} \end{pmatrix}, \tag{15}$$

where parameter vector \mathbf{p}_{LQ} controls the LQ model and \mathbf{p}_{HQ} the HQ model.

2.2. AAM search

The goal of the AAM search is to minimize the difference between a model and the actual image. This is done by estimating new model parameters from the grey value differences $\delta \mathbf{I}$ between both images:

$$\delta \mathbf{I} = \mathbf{I}_{\text{image}} - \mathbf{I}_{\text{model}} \tag{16}$$

To simplify the calculations, we assume there is a linear relationship between $\delta \mathbf{I}$ and the parameter change $\delta \mathbf{p}$:

$$\delta \mathbf{p} = \mathbf{R} \delta \mathbf{I} \tag{17}$$

We computed **R** by performing a series of experiments as described in.⁹ The linearity assumption works best for small parameter changes. For larger changes, the estimate of $\delta \mathbf{p}$ will be poor. By iterating it is possible to converge to a (local) optimal solution.

To speed up the search, we used a decreasing coefficient (starting with 1.0) for adapting the model's parameters. For each iteration, the overall fitting error $E_{\rm fit}$ is calculated by determining the mean square difference between the grey value intensities of the test image $\mathbf{g}_{\rm test}$ and the created synthetic image $\mathbf{g}_{\rm synt}$:

$$E_{\rm fit} = \frac{1}{G} (\mathbf{g}_{\rm test} - \mathbf{g}_{\rm synt})^{\rm T} (\mathbf{g}_{\rm test} - \mathbf{g}_{\rm synt}), \tag{18}$$

where G is the number of pixels that are compared. The coefficient is decreased by 30 percent if the error increases between two following iterations. The algorithm stops when there is no improvement for three following iterations at decreasing step sizes.

We initialize the first AAM search with the LQ model by setting the shape and texture parameters to their mean (zero). The model's pose parameters for rotation and scaling are also set to their mean, which are calculated



Figure 6. Synthetic images created with the LQ model for the three largest modes of variation. For each mode a single parameter was changed with two standard deviation difference from the mean in both directions.

from the training set. The translation parameters are calculated by aligning the centers of gravity of the model and the image. In this way the bottom 3 cm of the test image is omitted to compensate for possible artifacts, such as name tags in the image.

2.3. Switching between models

The resulting model parameters from the first AAM search are used to initialize the second, more detailed search. To switch from one model to another, it is necessary to determine the relation between the two models. A linear relationship can be deduced from the optimal parameters of both models belonging to the training set.

Considering the matrices A and B of parameter vectors from the training set of N examples for both models, we can estimate the cross covariance matrix Σ_{ba} (mean parameter values are zero):

$$\Sigma_{ba} = \frac{BA^{\mathrm{T}}}{N-1}.$$
(19)

The more training examples available, the better the estimation of the cross covariance matrix will be.

Next, the relation between A and B becomes:

$$B = \Sigma_{ba} A. \tag{20}$$

Since both models are different linear approximations, we can expect a certain error between the models' appearances. Also the parameter reductions by PCA for both models may result in errors caused by uncorrelated features.

2.4. Implementation

For recording landmarks in the image data set, we have written and implemented a C++ application for the Windows operating system.

We have implemented the AAM search algorithm in Matlab with the help of some functions from the image processing toolbox. The AAM search experiments were executed on a PC with a 2 GHz Intel Pentium IV processor and 533 MHz front side bus.

3. RESULTS

For the experimental evaluation, we gathered a data set of 80 hand radiographs. To test the algorithm's performance, the image collection was split into a training set of 50 images and a test set of 30 images.

The first search step with the LQ model found all bones correctly in 21 images of the test set. On average, convergence was reached within 8 iterations, which took about 0.3 seconds per iteration in our development environment. Seven images showed small failures in detecting the phalanges; the detected locations were off by less than 0.5 cm. Two large search failures were caused by a bad initialization; they had started too far off from the center of the hand and therefore did not find the fingers correctly.

After convergence was reached, the search algorithm switched to the HQ model. As expected (subsection 2.3), occasionally small segmentation differences were visible just after the model transition. Usually this was corrected after one or two further iterations.

The second search step with the HQ model converged within 7 iterations on average, at 2.5 seconds for each iteration. The 21 images that were correctly segmented in the first search also succeeded in the second. Also, one of the images that initially showed small failures was segmented successfully in the second search. For these 22 images, 70% of the landmarks were found within 1.3 mm distance from manual segmentation. Figure 8 shows the results after several iterations during the two search levels for one of our test images. The histogram in Figure 7A shows the total mean error for the images in the test set.

We also used two alternative AAM search configurations to show the effects of the multi-level search and the pose invariant submodels. In the first alternative, we used the HQ model without initialization by the LQ model. The segmentation results were not as good as with the two search levels, as is shown in the first histogram of



Figure 7. Histograms of the total mean error for the images in the training set for three different search configurations.

Figure 7. Since we did not change the HQ model itself, this shows that a better initialization of our model improves the results. Also, the execution time had increased: on average it took 16 iterations before convergence was reached at an average 2.5 seconds for each iteration. Overall, this shows that the single-level search method is about two times slower than the multi-level method.

In the second alternative, we modified the HQ model by leaving out the alignment step for the submodels. This makes the model variant for pose variabilities. This configuration also performed worse than the initial method. Figure 7 shows the results of these tests. There was no noticeable change in execution time for this method.

4. DISCUSSION AND CONCLUSIONS

Using our segmentation method we detected the hand skeleton's bone contours correctly in 22 out of 30 test images. 70% of the landmarks were found within 1.3 mm distance (at 0.66 mm pixel size) compared to manual segmentation by an expert.

The maximum texture resolution of 15 pixels per cm in the second search step sets a limit to the accuracy of the results. Increasing the resolution will improve accuracy, but will also add extra processing time. The current accuracy of the algorithm is suitable for finding regions of interest. Though, if we want to measure JSWs, we will need to increase the model's texture resolution.



Figure 8. Example of a successful AAM search with the LQ model (upper series) and the HQ model (lower series) after t iterations. The landmarks of the submodels have been connected and are visible as contours. The models' textures are not visible in these images.

Using two search levels of increasing model quality seems to be effective in reducing the calculation time. The search with the combination of the two models performed about twice as fast as the one with the single HQ model. Also, the total mean error of the results decreased, which is caused by a better initialization of the HQ model. Since both models are different linear approximations, switching between the two models sometimes results in minor deviations between their appearances. Usually this deviation is corrected within one or two iterations.

To compensate for large variabilities caused by differences in finger positioning, we made several submodels of the fingers invariant to rotation and translation. Our results show that this improves detection without changing the search's execution time.

Some small segmentation failures may be caused by a poorly trained model and a limited number of permitted variations. This can be improved by increasing the size of the training set. Poor initialization can also cause larger failures in the convergence of the search. Therefore we will need to find a more robust initialization method that can roughly determine the position of the hand.

REFERENCES

- J. T. Sharp, "An overview of radiographic analysis of joint damage in rheumatoid arthritis and its use in meta-analysis," *The Journal of Rheumatology* 27(1), pp. 254–260, 2000.
- 2. D. E. Trentham and A. T. Masi, "Carpo:metacarpal ratio. A new quantitative measure of radiologic progression of wrist involvement in rheumatioid arthritis," *Arthritis and Rheumatism* 19(5), pp. 939–944, 1976.
- J. T. Sharp, M. D. Lidsky, L. C. Collins, and J. Moreland, "Methods of scoring the progression of radiologic changes in rheumatoid arthritis. Correlation of radiologic, clinical and laboratory abnormalities," *Arthritis* and Rheumatism 14(6), pp. 706–720, 1971.
- A. Larsen, "How to apply larsen score in evaluating radiographs of rheumatoid arthritis in long-term studies," *The Journal of Rheumatology* 22(10), pp. 1974–1975, 1995.
- 5. D. van der Heijde, "How to read radiographs according to the sharp/van der heijde method," *The Journal of Rheumatology* **27**(1), pp. 261–263, 2000.
- D. van der Heijde, T. Dankert, F. Nieman, R. Rau, and M. Boers, "Reliability and sensitivity to change of a simplification of the Sharp/van der Heijde radiological assessment in rheumatoid arthritis," *Rheumatology* 38(10), pp. 941–947, 1999.
- M. N. Lassere, D. van der Heijde, K. R. Johnson, M. Boers, and J. Edmonds, "Reliability of measures of disease activity and disease damage in rheumatoid arthritis: implications for smallest detectable difference, minimal clinically important difference, and analysis of treatment effects in randomized controlled trials," *The Journal of Rheumatology* 28(4), pp. 892–903, 2001.
- K. Bruynesteyn, D. van der Heijde, M. Boers, M. Lassere, A. Boonen, J. Edmonds, H. Houben, H. Paulus, P. Peloso, Saudan, and oters, "Minimal clinically important difference in radiological progression of joint damage over 1 year in rheumatoid arthritis: preliminary results of a validation study with clinical experts," *The Journal of Rheumatology* 28(4), pp. 904–910, 2001.
- 9. T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23(6), pp. 681–685, 2001.
- M. B. Stegmann, R. Fisker, B. K. Ersboll, H. H. Thodberg, and L. Hyldstrup, "Active appearance models: Theory and cases," 9th Danish Conference on Pattern Recognition and Image Analysis 1, pp. 49–57, 2000.
- 11. H. H. Thodberg, "Hands-on experience with active appearance models," in *Medical Imaging 2002: Image Proceedings*, Sonka and Fitzpatrick, eds., **4684**, pp. 495–506, SPIE, 2002.
- M. G. Roberts, T. F. Cootes, and J. E. Adams, "Linking sequences of active appearance sub-models via constraints: an application in automated vertebral morphology," in *Proceedings of British Machine Vision Conference 2003*, 1, pp. 349–358, 2003.
- 13. M. Bukovec, R. Truyen, B. Likar, R. Bernard, and F. Pernus, "Image segmentation by connected parametrical models," in *Medical Imaging 2004: Image Processing*, **5370**, pp. 398–409, SPIE, 2004.
- C. Goodall, "Procrustes methods in the statistical-analysis of shape," Journal of the Royal Statistical Society Series B-Methodological 53(2), pp. 285–339, 1991.
- F. L. Bookstein, "Principal warps thin-plate splines and the decomposition of deformations," *IEEE Trans*actions on Pattern Analysis and Machine Intelligence 11(6), pp. 567–585, 1989.
- J. J. Gerbrands, "On the relationships between SVD, KLT and PCA," *Pattern Recognition* 14(1-6), pp. 375–381, 1981.