

Model-Based Lower Limb Segmentation using Weighted Multiple Candidates

André Gooßen¹, Eugen Hermann¹, Thorsten Gernoth¹
Thomas Pralow², Rolf-Rainer Grigat¹

¹Vision Systems, Hamburg University of Technology, 21079 Hamburg, Germany

²General X-Ray, Philips Healthcare, 22335 Hamburg, Germany

andre.goossen@tu-harburg.de

Abstract. In this paper we propose an extension of active shape model based bone segmentation. We examine the benefit of using multiple candidates for new landmark positions during segmentation. To incorporate this information we compare three strategies of adapting the fitting algorithm. For evaluation we segmented the hip, knee and ankle joints in more than 100 digital radiographs of the lower limbs. We achieve superior accuracy compared to the classic algorithm and prove that segmentation results benefit from multiple candidates.

1 Introduction

Segmentation of bone structures in digital radiographs is a prerequisite for quantitative orthopaedic examinations, such as length and angle [1], bone density or bone age measurements and preoperative planning [2]. However, an accurate delineation of these structures by experts is time-consuming [3] and often subject to intra- and inter-observer variability thus requiring a robust computer-aided segmentation capable of processing the large variety of images in clinical practice.

A well-known approach to segmenting digital radiographs are active shape models (ASM) [4]. However when it comes to modelling the appearance of medical image data it is often impossible to determine an accurate and robust feature. Spurious edges and noise produce additional candidates and might cause the segmentation to fail. Taking multiple candidates of a single or different image features [5] increases the probability of including the true position. We propose and evaluate three approaches for the task of segmenting the lower limbs in long-leg radiographs and compare the results in terms of accuracy and robustness.

2 Materials and Methods

In the original Active Shape Model (ASM) algorithm [4] the shape \mathbf{x} is modelled by adding a linear combination of eigenvectors $\mathbf{P} = (\mathbf{p}_1|\mathbf{p}_2|\dots|\mathbf{p}_t)$ describing possible shape variation to the mean shape $\bar{\mathbf{x}}$. The appearance is modeled using sampling vectors around each landmark $(x_i, y_i)^\top$, $i = 1, 2, \dots, n$. In an iterative manner new landmark positions (\hat{x}_i, \hat{y}_i) are estimated from the local appearance.

We call these positions candidates. A new shape \mathbf{x} is generated from these candidates by least-squares fitting, i.e. minimizing the error

$$\Delta = (\hat{\mathbf{x}} - (\bar{\mathbf{x}} + \mathbf{P}\mathbf{b}))^\top (\hat{\mathbf{x}} - (\bar{\mathbf{x}} + \mathbf{P}\mathbf{b})) \quad (1)$$

with \mathbf{b} denoting the parameters of the shape model.

However, this procedure results in outlying points disproportionately affecting the fit, which may or may not be desirable depending on the problem. The least-squares minimization yields optimal results for Gaussian distributed errors. Due to the non-uniform appearance and detailed structure in medical images the assumption of a Gaussian model for the residual distribution is seldom accurate. For that reason we add a quality measure for each candidate to control the candidate's influence on the fit. The quality of each candidate is represented in form of weights w_i , $0 \leq w_i \leq 1$ combined to a weight matrix

$$\mathbf{W} = \begin{bmatrix} w_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & w_{2n} \end{bmatrix} \quad (2)$$

with $w_{n+i} = w_i \forall 1 \leq i \leq n$. Using this extension, we derive the new error term

$$\Delta = (\hat{\mathbf{x}} - (\bar{\mathbf{x}} + \mathbf{P}\mathbf{b}))^\top \mathbf{W} (\hat{\mathbf{x}} - (\bar{\mathbf{x}} + \mathbf{P}\mathbf{b})) \quad (3)$$

This technique only considers single candidates per landmark. However, there might be additional information such as multiple local optima or candidates of different image features [5]. This follows the idea of boosting, i.e. taking many weak candidates and combining them into a robust and accurate result.

To incorporate this information we extend the fitting algorithm to multiple candidates per landmark and examine three possible strategies. Let κ_i denote the number of candidates for the i -th landmark and let $w_{i,k}$, $k = 1, 2, \dots, \kappa_i$ denote the corresponding weights:

- *Best/Fittest Candidate*: For each landmark we simply keep the candidate with the highest weight $w_i = \max(w_{i,k})$, regardless of the overall-rating of this weight compared to neighbouring landmarks. For an appropriate generation of the weights, this strategy always selects the fittest candidate. However, if all the weights of a certain landmark are small, this conflicts with the idea of boosting.
- *Multiple Candidates*: Instead of neglecting candidates with small weights, we integrate all candidates for a specific landmark into the error term. The new equations are given with respect to the candidate's coordinates. With κ_i denoting the number of estimations for the i -th landmark and the total number of feature points $\kappa = \sum_{i=1}^n \kappa_i$, the extended target vector $\hat{\mathbf{x}} \in \mathbb{R}^{2\kappa}$ is given by

$$\hat{\mathbf{x}} = (x_{1,1}, \dots, x_{1,\kappa_1}, \dots, x_{n,1}, \dots, x_{n,\kappa_n}, y_{1,1}, \dots, y_{1,\kappa_1}, \dots, y_{n,1}, \dots, y_{n,\kappa_n})^\top \quad (4)$$

The mean vector $\bar{\mathbf{x}} \in \mathbb{R}^{2\kappa}$ becomes

$$\bar{\mathbf{x}} = \left(\underbrace{x_1, \dots, x_1}_{\kappa_1}, \dots, \underbrace{x_n, \dots, x_n}_{\kappa_n}, \underbrace{y_1, \dots, y_1}_{\kappa_1}, \dots, \underbrace{y_n, \dots, y_n}_{\kappa_n} \right)^\top \quad (5)$$

In the same way each eigenvector \mathbf{p}_k , with $k = 1, \dots, t$ has to be extended to

$$\mathbf{p}_k = \left(\underbrace{p_1, \dots, p_1}_{\kappa_1}, \dots, \underbrace{p_n, \dots, p_n}_{\kappa_n}, \underbrace{p_{n+1}, \dots, p_{n+1}}_{\kappa_1}, \dots, \underbrace{p_{2n}, \dots, p_{2n}}_{\kappa_n} \right)^\top \quad (6)$$

- *RANSAC/MSAC*: This statistical approach [6] has the advantage of implicitly defining neither Gaussian nor any other error distribution and thus should work for any kind of candidate distribution, e.g. bimodal for spurious edges or equally distributed for image parts dominated by noise.

We derive a model by randomly sampling the minimum number of candidates t required to generate a shape. Subsequently we identify the number of candidates within a corridor τ around this shape. However we do not count, following the RANSAC algorithm, but rather sum up the weights of all candidates within this consensus set

$$C = \sum_i w'_{i,\kappa_i}, \quad w'_{i,\kappa_i} = \begin{cases} 0 & \delta((\hat{x}_{i,\kappa_i}, \hat{y}_{i,\kappa_i})^\top) > \tau \\ w_{i,\kappa_i} & \delta((\hat{x}_{i,\kappa_i}, \hat{y}_{i,\kappa_i})^\top) \leq \tau \end{cases} \quad (7)$$

Thus, the shape with the largest sum wins. This technique can be further extended; by selecting the candidates with a probability proportional to their weight, instead of equal probabilities, we enhance the chance to select a set of candidates on the true shape.

We applied the three strategies to the task of segmenting the joints in 109 clinical long-leg radiographs [7]. We measured the mean and maximum curve-to-curve (Fréchet) distance to manual expert delineations and compared the errors. To evaluate the lower bound for the accuracy we also quantified the inter-observer error and performed a study with five observers. The reference implementation is the classic ASM algorithm with an unweighted single candidate per landmark.

3 Results

The average processing time for a full segmentation (six joints) is 7 s for best candidate strategy, 11 s for multiple candidates and 30 s for RANSAC on a standard 4×3 GHz machine. Refer to Fig. 2 for a depiction of the overall quantitative results. The segmentation accuracy is increased for all three methods in terms of a lower mean as well as maximum error. For each of the partial segmentations of the whole leg, using multiple candidates results in smallest errors. The mean error is 0.59 mm for the hip, 0.47 mm for the knee, and 0.37 mm for the ankle joint compared to an inter-observer variability of 0.49 mm, 0.43 mm, and 0.35 mm respectively. Fig. 1 shows typical results of the three methods.

4 Discussion

We introduced and evaluated three approaches to incorporate multiple candidates for new landmark positions into classic ASM segmentation. Experiments have shown that weighted fitting of multiple candidates is superior to using single candidates or unweighted fitting.

It is evident from the mean error that the use of weighted multiple candidates considerably improves the segmentation accuracy for delineating bone structures in long-leg radiographs. The significant drop in the maximum error in all three methods indicates that, in addition to the higher accuracy, also the robustness is increased compared to classic ASM segmentation.

The remaining error is close to the inter-observer variability which forms the lower bound when measuring the accuracy. Moreover, further improvement of the segmentation is limited by the shape model's reconstruction error. It is thus

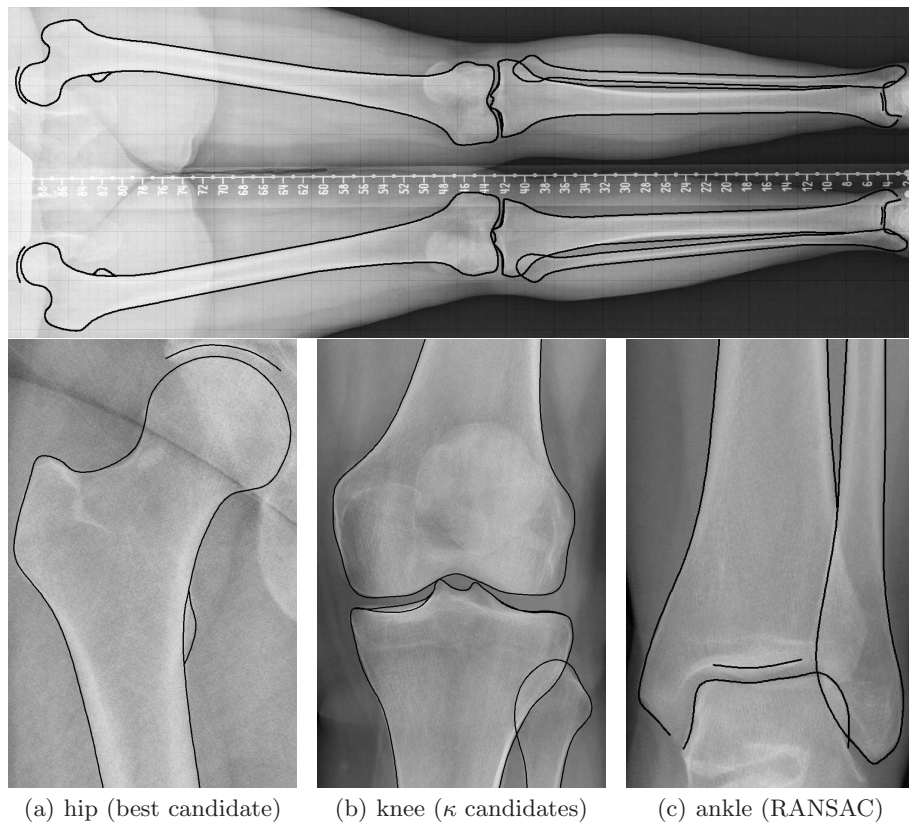
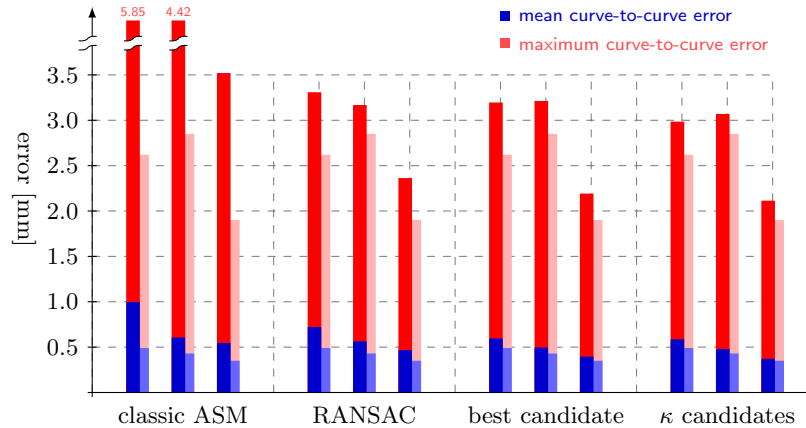


Fig. 1. Full leg segmentation in an oversized radiograph and typical delineation results of the segmentation for (a) hip joint (mean error 0.59 mm), (b) knee joint (mean error 0.47 mm) and (c) ankle joint (mean error 0.37 mm) using the three proposed methods.

Fig. 2. Mean and maximum curve-to-curve error for the classic ASM algorithm [4] and RANSAC strategy, weighted least squares with best candidate and multiple candidate strategy. The bars depict results for the hip (left), knee (mid) and ankle joint (right). The shaded bars in the background result from the inter-observer study and thus depict a measure for the accuracy of the manual delineation.



necessary to apply a separate post-processing to the ASM result, which will be addressed in future work.

References

1. Boewer M, Arndt H, Ostermann PW, et al. Length and angle measurements of the lower extremity in digital composite overview images. *Eur Radiol.* 2005;15(1):158–64.
2. Pafilas D, Nayagam S. The pelvic support osteotomy: indications and preoperative planning. *Strategies Trauma Limb Reconstr.* 2008;3(2):83–2.
3. Hochhausen C. Vergleich der software-unterstützten Vermessung von Röntgenaufnahmen der unteren Extremität im Rahmen der präoperativen Planung für Korrekturingriffe mit der manuellen Planung [PhD Thesis]. Medizinische Hochschule Hannover; 2004.
4. Cootes TF, Taylor CJ, Cooper DH, et al. Active shape models - their training and application. *Comput Vis Image Underst.* 1995;61(1):38–59.
5. Gooßen A, Peters D, Gernoth T, et al. Intelligent feature selection for model-based bone segmentation in digital radiographs. In: *Proc Int Technol Appl in Biomed.* IEEE; 2009. p. 1–4.
6. Rogers M, Graham J. Robust active shape model search. *Lect Notes Computer Sci.* 2002;2352:517–30.
7. Gooßen A, Schlüter M, Pralow T, et al. A stitching algorithm for automatic registration of digital radiographs. *Lect Notes Computer Sci.* 2008;5112:854–62.