

Active Shape Model Based Segmentation of Bone Structures in Hip Radiographs

Nabil Boukala, Eric Favier, Bernard Laget
 Laboratoire DIPI
 ENI de St-Étienne
 58, rue Jean Parot
 42023 Saint-Étienne - FRANCE
 boukala|favier|laget@enise.fr

Petia Radeva
 Computer Vision Center - CVC
 Campus UAB
 08193 - Bellaterra - Cerdanyola
 Barcelona- SPAIN
 petia.radeva@cvc.uab.es

Abstract— This paper presents a novel method for segmentation of bone structures in anterior-posterior (AP) radiographs based on active shape models. A priori global knowledge of the geometric structure of each hip is captured by a statistical deformable template integrating a set of admissible deformations. Then, to represent the image structure expected at each shape point, two statistical models are built: the first model contains knowledge about the edges extracted from the radiograph. This information is summarized in shape context histograms. The second model describes the local image structure around each model point. After gathering these two types of data over the training set, an Independent Component Analysis allows us to derive two data representations for each contour part of the shape. The search is performed over a median pyramid using the shape context model. The obtained segmentation is refined at the original radiograph resolution using the more local model. A leave-one-out test was used to evaluate the performance of the proposed method and to compare it with other conventional methods. The results demonstrate that the method is very robust and precise, and that it can be useful in the context of preoperative planning of hip surgery.

Key Words: Hip, Radiograph, Active Shape Model, Shape Context, Independent Component Analysis.

1 Introduction

Prior to hip replacement or osteotomies surgery, preoperative planning is required to decide on the type and size of the prosthesis and the reconstruction of the joint geometry. Despite the advantages that 3D image analysis methods offer, to date they are applied only in a few hospitals with a low percentage of the operated patients[1, 2, 3] due to technical, time and cost-related factors. To date the developments in 3D image visualization and simulation, and robot-assisted surgery cannot replace the conventional daily used methods of 2D radiodiagnosics. Delineation of bone structures in a

standard AP-radiograph is a prerequisite in preoperative planning. Since annotating manually these structures is very tedious and time-consuming, fast and accurate computer-aided segmentation methods are required. The difficulty to achieve a robust and fast segmentation of AP-radiographs is caused by several highly variable factors which substantially affects their aspect. These variations (Fig. 1) are induced by patient pose during acquisition, morphological differences, pathological deformations, X-ray device and acquisition protocol, and radiographic film, etc...

Among the large family of methods called deformable models, the learning-based Active Shape (ASM) and Active Appearance (AAM)[17, 18] Models have proven very successful [7, 8] to segment non-rigid objects with great variability regarding shape and appearance in medical imaging. In ASM, the shape variability is learned through observation, the objects being represented as sets of labelled points. The obtained shape representations are called Point Distribution Models (PDM). In practice, this is accomplished by a training set of annotated examples followed by a Procrustes Analysis[20] in order to align shapes with regards to position, scale and orientation. In this normalized frame, the statistics of the points coordinates can be obtained by principal component analysis (PCA). The characteristic pattern of a shape class is described by the average shape vector and a linear combination of eigenvectors of the variations around the average shape. During image search, new target points are searched in a region of the image around each model point and the model is updated to best fit these new target points. The new target points can simply be considered as points of highest gradient magnitude along a normal through the current point. Better results can be obtained by using a training set to build a statistical model to represent the image structure expected at each point.

A direct extension of the ASM approach has lead to the Active Appearance Models. Besides shape information, the textual information, i.e. the pixel intensities across the object in question, is included into the model. To obtain texture information from the training set, each shape is warped to a reference shape (mean shape) and sampled. Hereafter a photometric normalization

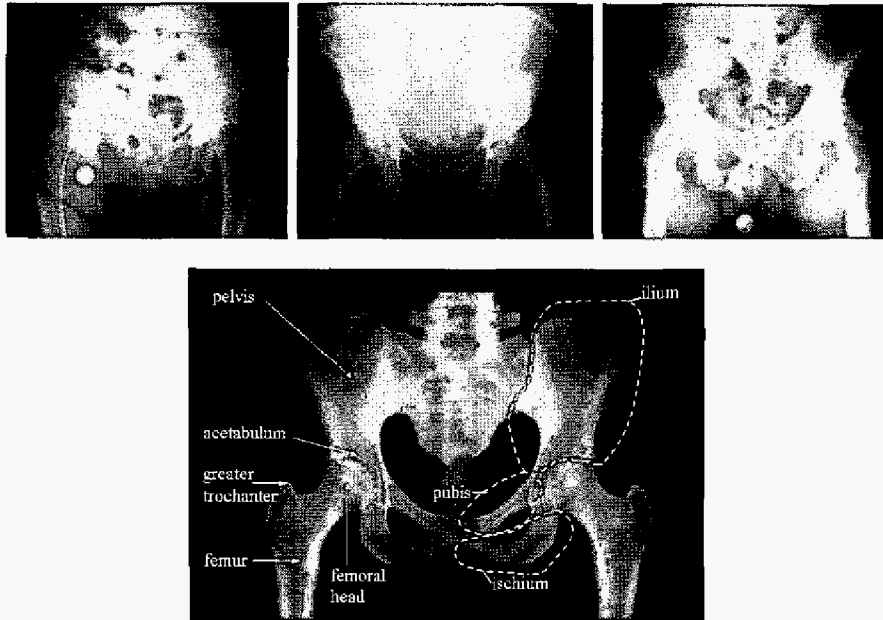


Figure 1: Variations in AP radiographs. Pelvis anatomy, anterior view.

of the obtained textures is done to remove influence from global linear changes in pixel intensities. Hence, a compact PCA representation is derived to deform the texture in a manner similar to what is observed in the training set. The AAM search attempts to minimize the difference between an actual image and the synthesized object obtained from the current shape, appearance and pose parameters. But this optimisation problem is not solved using optimisation techniques such as gradient based methods. Instead AAMs circumvent these potential problems by learning a linear relationship linking pixel differences (between the actual image and the synthesised object) to pose/model parameter displacements thanks to a multivariate linear regression. Though the AAM search provides a fast way of optimising the AAM using prior knowledge this might not lead to the optimal solution, primarily due to weakness in the assumption that the optimisation problems in an AAM search is strictly similar to those observed in a training set. It is suggested to fine-tune the model fit by using a general purpose optimisation method or to complete the AAM search with a traditional ASM search. AAMs and ASMs are also inherently dependent on a good initialisation.

Recently, Blind Source Separation (BSS) by Independent Component Analysis (ICA) has shown promise in signal processing application including speech enhancement systems, telecommunications and medical signal processing. ICA[9, 10, 11] is a linear transformation whose objective is to provide statistically independent projections. In other words ICA gives a representation in which the product of the marginal probabilities of the projected features best approximates the probability of the original features. In the context of statistical

pattern recognition of high-dimensional data and used together with the naive Bayes classifier, ICA transforms a D-dimensional density estimation into the estimation of D one-dimensional densities.

The proposed method uses a shape context descriptor[5] to perform a first optimization of the pose/shape parameters, which are then refined by a search using the learned image structure around each model point. The shape context data and local image structures are both represented using ICA. This allows us to easily estimate their probability density function (pdf). During search, a median pyramid [16] of the actual image is built; the model is initialised at the coarsest resolution, and updated at each higher resolution. Finally, at the highest resolution we refine the pose and shape parameters by a traditional ASM search using the learned ICA representation for each model point.

This paper is organized as follows. Section 2 describes the global methodology of the work. The results are presented and discussed in section 3 and our conclusions follow in section 4.

2 Segmentation Method

2.1 Dataset and Statistical Deformable Model

Our data set is composed of 24 X-ray images, of 13 female and 11 male patients coming from four different hospitals. Each patient presents a pathological deformation which can affect the visibility around the acetabulum and femoral head regions (the pelvis anatomy is depicted in Fig. 1). We generate a training set by segmenting manually the bone boundaries and hand annotating them with 20 easily located biological landmarks

(Fig. 2). We augment the set of correspondence points by subdividing regularly along the boundaries between these landmarks resulting in a model composed of 294 points. Following the method described by Cootes in [17, 18], we derive from this training set a mean shape and its main modes of variations (16 modes of variations capturing 98% of the shape variations).

Articulated shapes with pivotal rotations around one or more points are inherently non-linear. Consequently, it should be better to consider the three anatomies, i.e., pelvis, right and left hip, separately as in [8] or to consider other models[14, 15] which are not based on PCA. We define the position of the femur relatively to the pelvis by the angle between the two lines shown in Fig.2. The first line connects the acetabulum extremities while the other goes through the middle of the femoral neck and also passes through the centre of femoral head which is approximated by a circle. This angle was computed for each shape of the training set giving us the average angle value θ . The non-linearity, which is mainly caused by the variations in the patient posture is removed by rotating the femur around the centre of the femoral head so that each shape example presents the same angle value θ . The PDM is computed with this set of corrected shapes. During search, we update the pose/shape parameters using the following procedure: each anatomy (pelvis and femurs) is brought independantly in the PDM normalized frame by three procrustes analysis, the shape parameters are computed, the whole shape is brought back in the image space using the transformation matrix of the pelvis, and finally, each femur is rotated around the centre of the femoral head and scaled to best fit the new points. Removing the non-linearity in such a way improves the quality of the PDM. A leave-one-out test was used to evaluate the performance of the proposed PDM on our radiographs dataset, this gives us the lower bound in terms of precision that can be reach for each radiographs with such a shape model. The average error (average distance between homologous points)we obtained is 4.1 mm whereas we obtained 5.2 mm with the initial model without considering the non linearity.

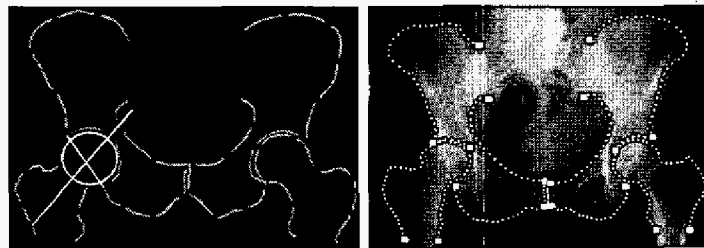


Figure 2: The femur is rotated around the femoral head center to adjust the angle between the drawn lines. The PDM with the 20 expert-labelled landmarks.

2.2 Shape Context Descriptor

We use the shape context descriptor [5] as a global attribute to register robustly and precisely our shape model with a new image. This first registration must be precise enough so that a further ASM search based on local image structures could find the optimal solution and avoid local minima. This descriptor was specially designed to find pointwise correspondences between an image shape and a stored prototype shape. By definition, the shape context is a shape descriptor which captures the coarse arrangement of a set of points sampled from the contours of an object with respect to another point of the contour. For a point p_i on the shape, the shape context at this point is the histogram h_i of the relative coordinates of the remaining $n - 1$ points,

$$h_i(k) = \#\{q \neq p_i : (q - p_i) \in \text{bin}(k)\}$$

For matching two sampled shapes, these histograms are computed at each sample point and the correspondences between the two point sets are found by solving a bipartite weighted graph matching problem, the weights being the similarity distances between histograms. We use the shape context descriptor to describe the edge information obtained after applying the canny edge detector to the radiographs.

2.3 Contour Classes

The shape model is divided in contour patches of equal size (Fig. 3). We derive a general representation of each contour class by first picking a set of samples points inside the contour part(Fig. 3). Then, for each radiograph, each sample point is equipped with two attributes vectors: the first one is a shape context histogram aimed to initialise the shape model in a new image and register it with relative precision, the other one is the image patch of predefined size and centred on the pixel in question, used for the final segmentation. If our dataset is composed of N images and that in each contour patch we select M samples, the training data used to learn representations of a contour class is composed of $N \times M$ histograms and $N \times M$ image patches. These two groups of data are considered separately and for each an ICA representation is derived.

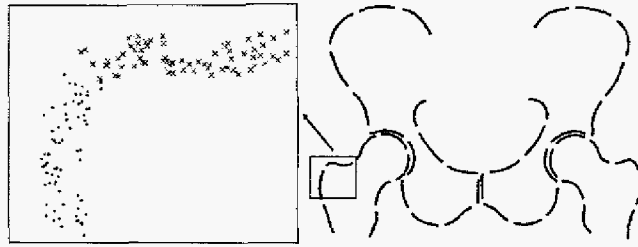


Figure 3: From right to left: the mean shape divided in contour patches and the samples points randomly picked inside two of these contour patches.

2.4 Independent Component Analysis

Given a set of observations represented by the M dimensional random vector x (zero-centered), assume the following ICA model:

$$x = As \text{ and } Wx = s,$$

where s is the vector of independent components s_m , A is called the basis matrix and W the filter or projection matrix which is the inverse of A (complete case: A is a square matrix). Based on the independence assumption on the sources s_m and the change of variables theorem, we have:

$$p(x) = |\det W| p(s) \approx |\det W| \prod_{m=1}^M p_m(s_m),$$

where p_m is the unidimensional marginal distribution of the m -th independent component. Instead of one representation, we have to learn a representation for each shape histogram class and each contour class. So for a certain class C^k , the filter and basis matrices are class-dependent $W = W^k$ and $A = A^k$. If $s^k = W^k(x - \bar{x}^k)$ where \bar{x}^k is the class mean estimated from the training data, we have the following class-conditional probability:

$$p(x|C^k) = |\det W^k| p(s^k|C^k)$$

Most ICA methods require data whitening as preprocessing. Since some simple denoising is also recommended, dimensionality reduction and whitening through PCA is very common practice as a preprocessing stage for ICA. In this case, W^k can be decomposed as

$$W^k = B^k(D^k)^{-1/2}V^k,$$

where V^k and D^k are the matrices composed by the eigenvectors and eigenvalues of the class covariance matrix, and B^k the ICA filter matrix. For this case, B results in an orthogonal matrix. The class-conditional probability becomes:

$$p(x|C^k) = |\det W^k| \prod_{m=1}^M p^k(s_m)$$

Using the fact that the absolute value of the determinant of an orthogonal matrix is 1 and that the determinant of a diagonal matrix is the product of the terms in the diagonal, we have:

$$|\det W^k| = |\det B^k| |(det D^k)^{-1/2}| |\det V^k| = \prod_{m=1}^M \frac{1}{\sqrt{\lambda_m^k}},$$

where λ_m^k are the eigenvalues of the covariance matrix of class C^k . Taking the logarithm, we obtain:

$$\log p(x|C^k) = \sum_{m=1}^M \log p^k(s_m) + \log |\det W^k|$$

For local image data, we applied the following preprocessing steps: normalisation, zero-centering saving the class mean, reduction of dimensionality and whitening. Then we applied the FastICA algorithm[12] on this preprocessed data. For the shape histogram data, we normalized the histograms and we applied Non-Negative Sparse Coding[13]. For both data sets, the unidimensional pdf were approximated by a mixture of three gaussians whose parameters are obtained by applying the expectation-maximization (EM) algorithm[4].

2.5 Training Procedure

The presented ASM requires the following training steps:

- Each image of the training set is annotated manually in order to build the PDM,
- The shape model is divided in contours patches,
- For each contour class, a local representation is learned by applying ICA and by estimating the densities in the resulting ICA space,
- For each contour class and at each level of a median pyramid, an ICA representation is learned based on the shape context histograms, followed by a density estimation.

2.6 Search Procedure

Given a new image, the segmentation consists in the following steps:

- For each contour class, compute the shape context likelihood map over the lowest resolution image of the pyramid,
- Segment the image in regions of dominant class: each pixel is associated to the contour class of highest likelihood,
- Extract the maximum of each region, and derive an affine transformation mapping each contour patch on its corresponding pixel by Procrustes Analysis,
- Traditional ASM search over the likelihood maps: each model point is moved along its normal to the pixel of highest likelihood, then the shape model is adjusted to best fit these new positions.
- For each pyramid level of higher resolution:
 1. Infer the pose and shape parameter from preceding pyramid level,
 2. Traditional ASM search based on shape context information: for each model point, compute the shape context histogram on the neighbour pixels along the normal and estimate their likelihood, update the position of the model point. Then adjust the shape/pose parameters.
- On the original image, use the local image structure representation to perform an ASM search.

3 Results and Discussion

We compared the performance of our method with a combination of the ASM and AAM algorithms. All the experiments were performed using a leave-one-out procedure. The AAM was built to represent 30000 pixels (we did not observe any improvements for higher resolutions). Multiresolution search was used with 3 resolution levels (25% 50% and 100% of the original image). We first compare the results of the AAM method with the multiresolution search based on shape context histograms. We used 100 bins histograms. During training, non-negative sparse coding was applied to the set of normalized histograms in order to obtain 60 components. Fig. 4 (left picture) shows the RMS point-to-point error for each image tested, the number of iterations was 10 per pyramid level with both methods (30 iterations). Over all the training set AAM gave a precision of $11.66 \pm 4.93\text{mm}$, while our approach reached a precision of $11.10 \pm 2.57\text{mm}$. It is interesting to see that grey-level textures and binary features summarized in shape contexts can lead to similar results in terms of robustness and precision. For 5 images, AAM gave a RMS

error below 7 mm while the minimum reached with our method is 8mm. It seems that the shape context descriptor cannot permit to reach a better precision, but it shows in this experiment a better stability than the AAM. In terms of speed, eventhough we precomputed the pdfs and stored them in look-up tables, AAM is more rapid.

In the second experiment, we used ASM to refine the search of the AAM search. The local grey-level models of the ASM were profiles 21 pixels long (10 either side of the point) and the search was performed on 5 pixels either side until convergence. We compared four different image features for these local models (intensity, normalized intensity, gradient, normalized gradient) leading us to the same conclusion as G. Behiels and al. in [7] that normalized intensities are the better image features in combination with the Mahalanobis distance. The ASM search was initiated with the shape parameters obtained from the multiresolution AAM search. In 75% of the images tested, the ASM improved the results of the multiresolution AAM search. We performed the same experiment using our ICA model: they were built using 16×16 pixels patches, reduced to 60 independent components after PCA-whitening. Here, we used the Fast-ICA algorithm. Fig. 4 (right picture) shows the RMS point-to-point errors given by an ASM search using traditional profile models and an ASM search using our ICA model. We cannot observe any significative difference due to the small size of our dataset, but we strongly think that bringing together our two ICA models (shape context and local structure) in an unique formulation could lead to a very efficient method.

4 Conclusion

We have presented an original statistical method of bone segmentation in AP radiographs based on a priori knowledge of the geometric structure of each hip and using both the contours extracted from radiographic images and local pixel intensities structures. The method performs the segmentation task with reasonable computational complexity and makes an original use of the shape context descriptor. The obtained robustness and precision proves that it can be useful to improve performance and fully automate the segmentation of bone structures in AP radiographs. The method can also be applied to compare preoperative to postoperative radiographs and time series in general. It still has to be validated on a larger dataset: a sufficient number of shape examples is necessary to integrate other types of PDMs more adapted to non-linearities. The proposed method remains sufficiently general to be applied to other medical registration/segmentation problems. We now intend to regroup the two contour class representations in a single representation and test other shape descriptors which are more computationally effective than the shape context.

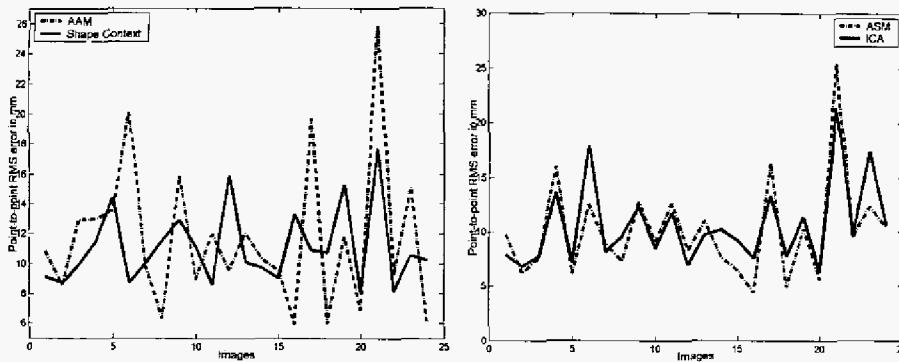


Figure 4: Correspondence RMS error for each image tested. Left picture: AAM results in dotted line, solid line for ICA model using shape context. Right picture: dotted line for ASM search using traditional profile models, solid line for ASM search using the local ICA model.

References

- [1] Babisch, J., Layher, F., Ritter, B., Venbrocks, R.-A.: Computer-assisted biomechanically based 2D planning of hip Surgery. *Orthopdische Praxis* **37** (2001) 29–38
- [2] Igljic, A., Kralj-Igljic, V.: Computer system for determination of hip joint contact stress distribution from standard AP pelvic radiograph. *Rad. and Onc.* **33** (1999) 263–266
- [3] Jaklic, A., Pernus, F.: Morphometric analysis of AP pelvic and hip radiographs. *Proc. of the 3rd Slov. Elect. and Comp. Sc. Conf.* (1994) 352–355
- [4] Dempster, A.P., Laird, N., Rubin, D.: Maximum Likelihood for Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society*, **39** (1977) 1–38
- [5] Belongie, S., Malik, J., Puzicha, J.: Shape Context: A new descriptor for shape matching and object recognition. *NIPS* (2000) 831–837
- [6] Belongie, S., Malik, J.: Matching with shape context. *IEEE Workshop on Content-based Access of Image and Video Libraries* (2000) 831–837
- [7] Behiels, G., Vandermeulen, D., Maes, F., Suetens, P., Dewaele, P.: Active shape model-based segmentation of digital x-ray images. *MICCAI* **1679** (1999) 128–137
- [8] Bernard, R., Pernus, F.: Statistical approach to anatomical landmark extraction in AP radiographs. *SPIE* **4322** (2001) 537–544
- [9] Jutten, C., Herault, J.: Blind Separation of Sources, Part I: An Adaptive Algorithm-Based on Neuromimetic Architecture. *Signal Processing* **24** (1991) 1–10
- [10] Bell, A.J., Sejnowski, T.J.: An Information-Maximization Approach to Blind Separation and Blind Deconvolution. *Neural Computation* **7** (1995) 1129–1159
- [11] Cardoso, J.F., Laheld, B.: Independent Component Analysis—A New Concept?. *Signal Processing* **36** (1994) 287–314
- [12] Hyvaerinen, A., Oja, E.: A Fast Fixed-Point Algorithm for Independent Component Analysis. *Neural Computation* **9** (1997) 1483–1492
- [13] Hoyer, P.: Non-Negative Sparse Coding. *IEEE Workshop on Neural Networks for Signal Processing* (2002) 557–565
- [14] Cootes, T.F., Taylor, C.J.: A Mixture Model for Representing Shape Variation. *Image and Vision Computing* **17** (1999) 567–574
- [15] Bressan, M., Vitria, J.: Independent Modes of Variation in Point Distribution Models. *Visual Form 2001, 4th International Workshop on Visual Form, LNCS 2059, Springer Verlag* (2001) 567–574
- [16] Starck, J.L., Murtagh, F., Louys, M.: Astronomical Image Compression Using the Pyramidal Median Transform. *Astron. Data Anal. Software Syst IV* (1995)
- [17] Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Active shape models: Their training and application. *CVIU* **61** (1995) 38–59
- [18] Cootes, T.F., Edwards, G. J., Taylor, C.J.: Active appearance models. *European conf. on computer vision* **2** (1998) 484–498
- [19] Bookstein, F.L.: Principal Warps: Thin Plate Splines and the decomposition of deformations. *IEEE Trans. on PAMI* **11**(1989) 567–585
- [20] Gower, J.C.: Generalized Procrustes Analysis. *Psychometrika* **40** (1975) 33–51