Knowledge based automated feature extraction to categorize secondary digitized radiographs

Michael Kohnen*[a], Frank Vogelsang[a], Berthold B. Wein[a], Markus Kilbinger[a], Rolf W. Günther[a], Frank Weiler[b], Jörg Bredno[c], and Jörg Dahmen[d]

[a] Dept. of Diagnostic Radiology, Aachen University of Technology, D–52057 Aachen, Germany

[b] Parsytec AG, Auf der Hüls 183, D-52068 Aachen, Germany

[c] Institute of Medical Informatics, Aachen University of Technology, D–52057 Aachen, Germany

[d] Dept. for Computer Sciences VI, Aachen University of Technology, D–52057 Aachen, Germany

ABSTRACT

An essential part of the IRMA–project (Image Retrieval in Medical Applications) is the categorization of digitized images into predefined classes using a combination of different independent features. To obtain an automated and content–based categorization, the following features are extracted from the image data: Fourier coefficients of normalized projections are computed to supply a scale– and translation–invariant description. Furthermore, histogram information and Co–occurrence matrices are calculated to supply information about the grey value distribution and textural information. But the key part of the feature extraction is the shape information of the objects represented by an Active Shape Model. The Active Shape Model supports various form variations given by a representative training set; we use one particular Active Shape Model for each image class. These different Active Shape Models are matched on preprocessed image data with a simulated annealing optimization. The different extracted features were chosen with regard to the different characteristics of the image content. They give a comprehensive description of image content using only few different features. Using this combination of different features for categorization results in a robust classification of image data, which is a basic step towards medical archives that allow retrieval results for queries of diagnostic relevance.

Keywords: image retrieval, categorization, projections, histogram, Co–occurrence matrix, Active Shape Model

1. INTRODUCTION

In medical informatics, the importance of digital image archives in daily clinical routine is growing fast. Among other aspects, the IRMA–Project (Image Retrieval in Medical Applications) has the goal to provide methods which enable an automatic categorization of secondary digitized radiographs retrieved from a conventional archive.[6] The description and categorization of the experimental database, containing almost 2500 radiographs up to now, was developed by radiologic experts to guarantee a representative distribution of radiographs. In this approach, a combination of different features is used to guarantee a robust categorization of the images into predefined classes.

The features are either non–knowledge based like projections, histograms and textural informations or incorporate a–priori knowledge using Active Shape Models first described by Cootes et. al.[2,3] to include shape information into the classification process. The Active Shape Models are trained on representative training sets of contours to built up a proper shape model for each class. These manually generated shape models contain a large amount of a–priori knowledge about all possible shape variations of objects in different categories.

This knowledge based automatic categorization is the first step towards archive queries of diagnostic relevance. This means for later steps of the IRMA–project, that far more differentiated queries (e.g. "Give me more chest radiographs with rib fractures", etc.) can be answered successfully without requiring user input about the image contents at archiving time.

* Correspondence: Email: kohnen@rad.rwth–aachen.de; Telephone: ++49 241 80 89627; Fax: ++49 241 8888 480

2. METHODS

This section first describes the different methods used to extract the non–knowledge based features of the image material. Following, the computation and optimization of the Active Shape Models is explained. Furthermore, we describe which describes the computation of an energy image which is needed for the optimization of the Active Shape Models. Finally the extracted features are combined to a feature vector, which is used for the classification.

2.1. Computation of non–knowledge based features

These features are more or less standard features, which were used in digital image processing in many different contexts. But here the difference especially for the projections is that we use a scale– and translation–invariant representation of the features. Thus, we are able to compute features which are independent of the size and orientation of images.

But these standard features are too unspecific to allow a robust and satisfying classification. In many cases, it is impossible to distinguish between different categories of radiographs without using the knowledge based shape information. Nevertheless, these features are an important part of the whole classification process and are able to improve the results of a pure shape based classification.

2.1.1. Scale– and translation–invariant representation of projections

Usually the contents of the radiographs have a defined orientation, which derives from the imaging process. In most cases these features differ only in size, if the images were digitized in different resolutions, and translation along the projection axes. Therefore, it is necessary to compute a scale– and translation invariant representation of the projections.

At first we define a projection as the normalized summation of the pixel grey–values of an image $f(f : M \times N \rightarrow G)$ horizontally and vertically. The normalized projections are represented by the vectors P_{hor} and P_{ver}, which include the normalized row– and column–sums of the grey–values respectively:

$$\begin{aligned} P_{hor}(n) &= \frac{1}{\sum\limits_{i,j} f(i,j)} \sum_{j=0}^{M-1} f(n,j) \\ P_{ver}(m) &= \frac{1}{\sum\limits_{i,j} f(i,j)} \sum_{i=0}^{N-1} f(i,m) \end{aligned} \tag{1}$$

To achieve a scale invariant representation a linear interpolation method is used to resize the projections down to a predefined length of 32 values. This results in a feature vector of reduced dimension that is faster to compute without losing too much relevant information.

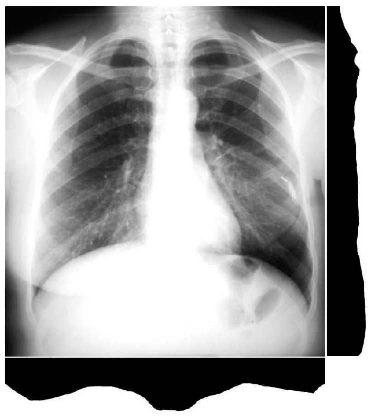

Figure 1. Horizontal and vertical projections of a thorax radiography. The vertical projection shows characteristic features resulting from the mediastinum.

This representation is not translation invariant yet. Therefore, we use the discrete Fourier transform (DFT) to obtain a frequence analysis of the projection vector. The Fourier domain of the transformed values can be divided into a phase– and an amplitude spectrum. The phase spectrum contains the location information whereas the amplitude spectrum holds the translation independent frequence description of the projection data. This quality of the DFT is used here to generate a translation invariant representation of the projections. The amplitude spectrum of the Fourier transformed projections P^*_{hor} and P^*_{ver} is computed using the following formulas:

$$P^*_{hor}(n) = |\sum_k P_{hor}(k)\omega^{kn}|$$

$$P^*_{ver}(n) = |\sum_k P_{ver}(k)\omega^{kn}| \tag{2}$$

$$\text{with} \quad \omega = e^{\frac{2\pi i}{32}}.$$

According to equation 2 this the transformed values are real values as they can be interpreted as the absolute value of the complex values. The loss of information due to this representation is restricted to the location based information of the Fourier transformed projection.

2.1.2. Histogram and textural information

The histogram information is also used to distinguish particular categories in the archive from each other. Histograms obtained from soft tissue images for example are quite different from histograms computed from skeletal radiographs. As visualized in figure 2, the histograms of a mammography and a femoral have very different gray–value distributions, which is taken as a characteristic feature to distinguish these categories from each other.

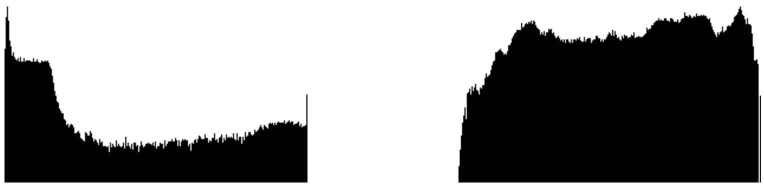

Figure 2. Left: histogram generated from a mammography (soft tissue). Right: histogram computed from a skeletal radiograph.

The images stored in the database are 8–bit gray–value images. Therefore, the histograms $h(g)$ consist of 256 entries, which is too large to be computed as a feature by the classifier. That's why we also scale the histogram by linear interpolation down to a length of 32 values. Another advantage of the normalization is the possibility to handle images of any gray–value range (e.g. 12–bit images resulting from CT scans).

Further textural features are an important measure to reliably distinguish medical images from each other. The success of a wide variety of radiologic diagnoses depends on the proper inclusion of textural information into the diagnose process. Co–occurrence matrices first described by Haralick[5] are a proper representation of textural information occurring in medical images.

Co–occurrence matrices are $G \times G$ matrices, where the elements $P_{\vec{d}}(i,j)$ are the number of the occurrences of the grey–values i and j with a defined displacement $\vec{d}$ in the image. Additionally the Co–occurrence matrix we use is normalized and rotation invariant.

But this matrix is far to big (256×256 values leads to a feature vector length of 65536) to be taken into consideration as feature for the classifier. Therefore we compute the six linear independent Haralick texture measures $h_1, \ldots, h_6$ from the matrix. These texture measures can be combined to a single feature vector

$$c = (h_1, \ldots, h_6), \tag{3}$$

which is taken as an input for the classifier.

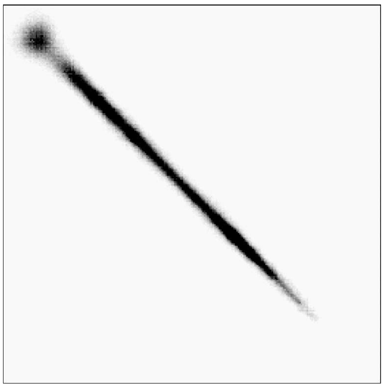

Figure 3. A visualization of a rotation invariant Co–occurrence matrix from a hand radiography. The matrix shows the characteristic accumulation of values greater zero along the diagonal axis.

2.2. Shape models as a knowledge–based feature

The Active Shape Model used for generating knowledge based features for the classification process is the key part of our approach. The Shape Model used here is a contour or polygon v which is deformed to represent biological shape variations by a reduced set of parameters to achieve a best possible correlation with the edges occurring in the image. The optimization process is necessary to find the optimal parameter settings using an energy image D.

Assuming that the images can be divided into a predefined number of categories $K_i, i \in \{1, \ldots, N\}$, the shape information of the different objects occurring in the classes can be trained a–priorly. The Active Shape Model is able to model possible shape variations of an object, preferred by a set of training shapes, with a comparatively low number of parameters. Here the large variety of biological objects must be reflected by the training set.

2.2.1. Computation of the shape model

To compute an Active Shape Model, an affine normalized set of training contours, containing m contours with n sorted vertices each, is needed. The vertices of the contours are sorted so that the starting point is always fixed (e.g. middle of left side of object edge), the vertices are ordered clockwise around the object, and the vertices of the contour have all the same distance from each other.

The affine normalization can be computed minimizing the quadratic error of every normalized contour $\tilde{v}_i$ to the mean contour $\overline{v}$ of the training set. An affine transformation of a vertex $v_{i_j} = \binom{x_{i_j}}{y_{i_j}}$ of the contour v_i is defined as follows:

$$\begin{pmatrix} \tilde{x}_{i_j} \\ \tilde{y}_{i_j} \end{pmatrix} = \begin{pmatrix} a & -b \\ b & a \end{pmatrix} \cdot \begin{pmatrix} x_{i_j} \\ y_{i_j} \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \end{pmatrix} \quad \text{with} \quad a = s \cdot \cos(\alpha) \quad \text{and} \quad b = s \cdot \sin(\alpha). \tag{4}$$

Therefore, the affine transformed coordinates of the contour which is normalized are computed using

$$\begin{aligned} \tilde{x}_{i_j} &= a \cdot x_{i_j} - b \cdot y_{i_j} + t_x \\ \tilde{y}_{i_j} &= b \cdot x_{i_j} + a \cdot y_{i_j} + t_y. \end{aligned} \tag{5}$$

The resulting contour must minimize the quadratic error

$$\lambda_i = \sum_{j=1}^{n} \left((\overline{x}_j - \tilde{x}_{i_j})^2 + (\overline{y}_j - \tilde{y}_{i_j})^2 \right), \tag{6}$$

between the normalized contour $\tilde{v}_i$ and the mean contour $\overline{v}$ of the training set. To solve this equation we substitute the unknown variables $\tilde{x}_{i_j}$ and $\tilde{y}_{i_j}$ by using equation 5 and obtain

$$\lambda_i = \sum_{i=1}^{n} \left((\overline{x}_j - a \cdot x_{i_j} + b \cdot y_{i_j} - t_x)^2 + (\overline{y}_j - b \cdot x_{i_j} - a \cdot y_{i_j} - t_y)^2 \right), \tag{7}$$

which can be easily solved by partially deriving it by its affine parameters and solving the resulting equation system.

The normalization leads to a set of contours, which is independent from affine shape variations of the objects. The advantage is, that we achieve a separation from the affine deformation parameters and the real shape deformation parameters of the Active Shape Model.

Subsequently a contour $\tilde{v}_i$ of the normalized training set, can be defined as a vector $\vec{v}_i$ with length $2n$, which contains the coordinates of its sorted vertices:

$$\vec{v}_i = (x_{i_1}, y_{i_1}, x_{i_2}, y_{i_2}, \ldots, x_{i_n}, y_{i_n})^T \tag{8}$$

Therefore, the contours of the training set can be interpreted as elements of a $2n$–dimensional vector–space. The motivation is, that the shape characteristics are lying in a subspace of this $2n$–dimensional vector–space. To describe the shape characteristics with a lower number of basis vectors, the Karhunen Loéve transform is used. The $2n \times 2n$–covariance–matrix C is computed, which entries c_{ij} describe the dependencies between the different vertices of the contours:

$$c_{ij} = \frac{1}{2n} \sum_{l=1}^{2n} (v_{l_i} - \overline{v}_i)(v_{l_j} - \overline{v}_j), \tag{9}$$

where $\overline{v}_i$ is the mean contour of the normalized training set.

The Karhunen Loéve transform of the covariance matrix C generates an orthonormal basis of eigenvectors with their corresponding eigenvalues. The different sizes of the eigenvalues describe the variance of the training set along the accompanying eigenvector. The eigenvectors are sorted by the size of their eigenvalues. Usually the size of the eigenvalues is shrinking rapidly and so only a few eigenvectors are needed to describe the relevant shape variations.

A shape variation v^* of the Active Shape Model is computed by an reduced orthonormal basis P^*, which consists of the k eigenvectors with the greatest eigenvalues, and the corresponding varied eigenvalues ξ_j:

$$v^* = \overline{v} + P\xi = \overline{v} + \left[\begin{pmatrix} p_{1,1} & \cdots & p_{1,k} \\ \vdots & & \vdots \\ p_{n,1} & \cdots & p_{n,k} \end{pmatrix} \begin{pmatrix} \xi_1 \\ \vdots \\ \xi_k \end{pmatrix} \right] \tag{10}$$

Translation, scaling, and rotation is applied by an affine transformation on the shape variation v^* as described previously in equation 4. This computed shape model is affine invariant and allows proper shape variations given by the training set.

A comparatively low number of eigenvectors is sufficient to model possible shape variations given by the training set. Here the used shape models are computed with satisfying accuracy using only ten different parameters. These parameters divide into 5 affine parameters (translation in x– and y–direction, scaling in x– and y–direction, rotation by α), and the variations of the five largest eigenvalues. The allowed amount of possible variations is limited by a discrete interval with fixed borders. The borders of the intervals $I_j = [a_j, b_j]$ for every varied eigenvalue ξ_j can be computed from the normalized training set. For this we compute the different variations $\hat{\xi}_{i_j}$ of the eigenvalues of the Active Shape Model which describe exactly the particular normalized training shapes $\tilde{v}_i$:

$$\begin{pmatrix} \hat{\xi}_{i_1} \\ \vdots \\ \hat{\xi}_{i_n} \end{pmatrix} = \begin{pmatrix} p_{1,1} & \cdots & p_{1,n} \\ \vdots & & \vdots \\ p_{n,1} & \cdots & p_{n,n} \end{pmatrix}^{-1} \cdot (\tilde{v}_i - \overline{v}_i) \tag{11}$$

Further, the borders of the intervals can be calculated taking the maximum and minimum values from the computed variations of the eigenvalues $\hat{\xi}_{i_j}$:

$$I_j = [\min_i \xi_{i_j}, \max_i \xi_{i_j}] \tag{12}$$

The interval borders of the affine parameters t_x, t_y, and s are set in respect to the image size and the magnitude of the mean contour $\overline{v}$. Thus, the Active Shape Model is capable of being optimized successfully on images of previously unknown scale. The interval borders used for the rotation are manually set as they are independent of the image– and shape model size.

Figure 4 shows the behavior of the Active Shape Model of the hand varying the largest eigenvalue with respect to the interval borders.

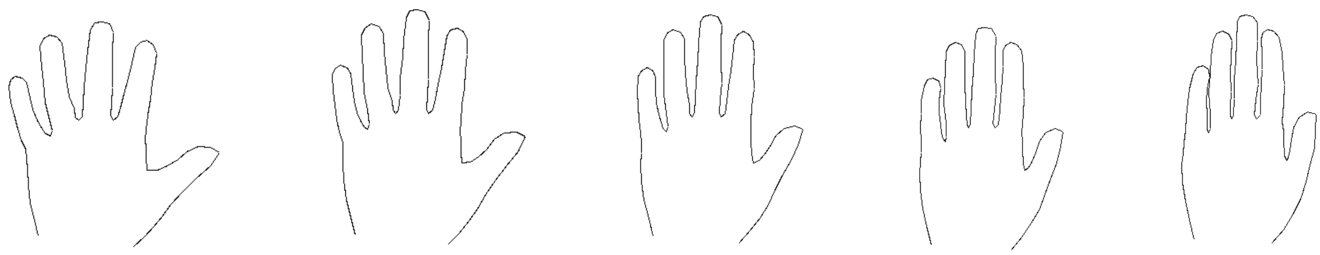

Figure 4. Variation of the Active Shape Model in the interval $I_1 = [a_1, b_1]$ along the eigenvector with the largest eigenvalue.

2.2.2. Preprocessing of the image data

A preprocessing of the image data is necessary for the optimization of the Active Shape Model. Goal of this preprocessing step is an image, which contains a reduced number of local minima to improve the optimization process. Local minima are caused by edges in the images, which do not belong to relevant objects.

The contour information of the image is extracted by the Canny–Edge–Detector.[4] Because this edge filter classifies the pixels using the grey–value gradient, a large amount of irrelevant contours in the image is detected, which are caused by image disturbances or other artifacts (e.g. medical markers, metallic implants, etc.). Therefore, contours are sorted out, which do not fulfill criteria like minimum length and inhomogeneity of the contour run. The minimum length of the contours is computed using a tracer algorithm, which counts the number of pixels the contour consists of. If the number is below a threshold t_1, the contour is removed from the image. The inhomogeneity of the contour is computed counting the number of direction changes of the contour pixels. Alike before, if this value is greater than a threshold t_2, the contour is also removed from the image. Therefore the resulting binary image contains only contours which have an increased probability to belong to the relevant objects. Different variations of the parameters t_1 and t_2 cause resulting images, which contain more or less contours respectively.

Nevertheless, this method runs the danger of removing relevant contours as well. We thought this aspect to be negligible compared to the improved optimization based on the sufficient amount of relevant contours remaining in the image.

Finally a distance transformed version of the computed contour image is created for the optimization process. The grey–values of the distance–image are a measure for the city–block– (or chessboard–) distance to the next contour in the image. The main advantage of the distance image is that a computation of the distance of every pixel to the nearest edge (local minimum) can be computed in a complexity of $O(1)$. Figure 5 shows a binary contour image of the hand and the distance transformed version.

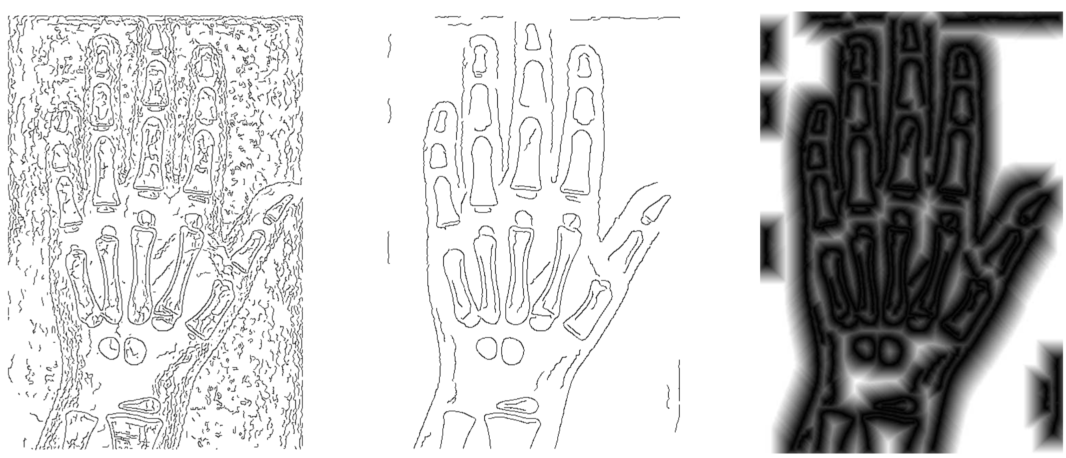

Figure 5. Left: binary contour image of a hand radiography. Middle: binary contour image after removing irrelevant contours. Right: distance transformed contour image (contrasted due to visualization aspects).

2.2.3. Optimization of the shape models

Goal of the optimization is to determine the parameterized variation of the Active Shape Model with the minimal energy E_{v^*} under consideration of the distance image D. The energy is computed by a weighted summation of the overprinted grey–values of the distance image by the vertices belonging to the shape model:

$$E_{v^*} = \sum_{i=1}^{n} w_i v_i^* \cdot D_{xy}, \quad \text{with} \quad \sum_{i=1}^{n} w_i = 1 \tag{13}$$

The number of possible shape variations of the Active Shape Models with k parameters, where every parameter can accept l different values is l^k. Because of the exponential magnitude the exact solution of the search problem is NP–complete.

Therefore, the simulated annealing method is used, which provided satisfactory solutions of large search problems containing several local optima in the past.[3] A necessity for the application of this stochastic optimization method is a topologically sorted search space. This ordering of the search space is modeled by a neighbor relationship. In the case of the shape models used here every shape variation has twenty neighbor configurations, which can be computed by increasing or decreasing one of the ten parameters by a discrete step. The simulated annealing method evolves in the search space by computing new neighbors randomly and determines whether to accept these or not. The probability of accepting a new neighbor depends on a Boltzmann distribution.

Thus, the shape model is capable of leaving local minima and should approximate under successive computation of new neighbors towards the global minimum. The simulated annealing method is therefore a very suitable method for finding the global optimum in a search space containing many local optima. Figure 6 shows the evolution of the optimization on the example of a hand radiograph.

$E^*_{v^*} = 0.772$ $E^*_{v^*} = 0.363$ $E^*_{v^*} = 0.215$

Figure 6. Variations of the Active Shape Model of the hand after 50, 200 and 400 iterations of the simulated annealing optimization.

2.2.4. Computing features from shape models

Each a–priorly known category c_i possesses its own Active Shape Model, which describes the possible shape variations of the occurring objects. We created the models using the computation method described above.

After optimization of the models for every category c_i, we obtain an minimum energy $E_{v_i^*}$ for each model, which is a good measure for the correlation between the shape and the displayed objects in the image. In other words, the shape model with the lowest energy fits best to the contours of the image. To grant the comparability of the different energy measures of the models, the energies must be normalized in respect to their contour length. Thus we obtain the following normalized energy measures:

$$E^*_{v_i^*} = \frac{1}{n_i} \cdot E_{v_i^*} \tag{14}$$

The normalized energy measures can be interpreted as a measure for the probability of the image belonging to the particular category the Active Shape Model was generated from. Intention of this approach is that the category has the highest probability where the optimized the shape model had the lowest energy. The energy measures of all the Active Shape Models are combined into a vector s^*, which can be taken as a feature vector for the classification process:

$$s = (E^*_{v^*_1}, \ldots, E^*_{v^*_n}) \tag{15}$$

2.3. Categorization using a combined feature vector

In this section, the categorization of the images using knowledge and non–knowledge based extracted features is described. Summarizing, we generated a variety of different features from each image presented in 1. Therefore, the resulting feature vector V, which is taken as input vector for a classifier, has a length of $102 + N$, where N is the number of different categories:

Table 1. Features extracted from the images, their length and functional symbol.

Description of the feature	length of feature	functional symbol
Fourier transformed scaled horizontal projection	32	$P^*_{hor}(n)$
Fourier transformed scaled vertical projection	32	$P^*_{ver}(n)$
Scaled histogram	32	$h^*(g)$
Texture measures by Haralick computed from the Co–occurrence matrix	6	$c(h_i)$
Knowledge based feature computed by the Active Shape Model	# categories	$s(E^*_{v^*_i})$
$\sum$	102 + # categories	

$$V = (P^*_{hor}, P^*_{ver}, h^*, c, s) \tag{16}$$

The size of V is not too large for the most common classifiers (e.g. nearest neighbor classifier) as we consider the differentiation of 7 categories. Here we use the so called C5.0 classifier, which generates a decision tree from a training set of feature vectors to minimize the entropy of the training set.[7] The advantage of this classifier is the transparency of the generated decision tree to the user in a way that later modifications on the tree can be easily performed and classification errors can be determined depending on the rules stored in the tree.

3. RESULTS

The different extracted features like projections, histograms, texture and shape measures were chosen regarding to the different aspects of the image contents. They give a good measure of what is visible in the image described by a comparatively low number of different features. Seven different categories were defined from the image archive for the test of the algorithms. The different categories are hand, foot, skull, spine, chest, pelvis and extremities. At this state of the project, the non–knowledge based features are tested separately from the knowledge based ones to determine the quality separately of the two different kinds of features.

First tests based on the non–knowledge features produced an error rate of almost 30%. This result is not surprising considering the image contents. In fact, images containing skeletal objects often cannot be separated satisfactory by histogram an textural information. Mainly the projections are a good feature to separate the contents of skeletal objects, but in several cases the projections are disturbed by image artifacts like medical markers or metallic implants, which make it impossible for the classifier to distinguish the classes satisfactorily.

Therefore we incorporate knowledge based shape information about the objects occurring in the images. By now, two different Active Shape Models were generated, which model the shape characteristics of spine and hands. The variabilities were analyzed by a training set of 12 manually generated shape models each. At first we tested the behavior of the models separately on new test images. Optimizing the Active Shape Model of the hand on an amount of 30 unknown hand images, a mean energy measure of $\mu_h = 0.280$ with a standard deviation of $\sigma_h = 0.169$ was calculated. In contrast, the optimization of the spine model on the same set of images produced a mean lasting energy of $\mu_w = 0.556$ with a standard deviation of $\sigma_w = 0.258$. This constellation shows the capability of the knowledge based features to distinguish the different categories from each other. The images used for testing the shape features partly contained artifacts, which had no effect on the behavior of the shape model because their particular shape could not be represented by the shape models and therefore was not detected. But this method also ends up at its limits if too large parts of the relevant objects could not be detected by the Canny filter or the shape of the objects runs too atypical.

4. DISCUSSION

Our approach is trying to categorize images in a medical archive into predefined classes using a–priori knowledge based shape information and non–knowledge based arbitrary features. Beforehand a categorization using a balloon model to detect arbitrarily shaped biological objects[1] did not produce satisfying features for categorization.

But it must be considered that this approach maybe not suitable for image material which contains less or too noisy contour information (e.g. ultrasound images). Satisfying results are expected on skeletal radiographic images, because on these images the relevant objects (bones) have usually clear contours. Also the non–knowledge based features lead to better classification results on skeletal radiographs because of the better contrasted grey–value distribution in these images. There are also problems especially with the shape based features and projections if the variability of the contours is too high to be modeled by an Active Shape Model. These aspects must be taken into consideration for example for the classification of images of the abdomen with contrast media. Therefore, the IRMA–project will improve the classification process with a combination of many different features to grant a robust and satisfying categorization for medical archives.

Nevertheless,our approach shows an interesting opportunity to integrate a–priori known shape information into a classification process. Particularly the robustness of the algorithm in regard to image artifacts distinguishes it from classic approaches. Considering the results it is obvious that the non–knowledge based features can only improve the classification results of a knowledge based categorization. In addition this approach is able to provide a pre–segmentation of the visible objects, which can be very helpful for subsequent segmentation algorithms.

REFERENCES

1. J. Bredno, T. Lehmann, K. Spitzer, *A General Finite Element Model for Segmentation in 2, 3, and 4 Dimensions*, Procs. Medical Imaging SPIE, San Diego, 2000, (this issue)
2. T. Cootes, C. Taylor, D. Cooper, J. Graham, *A Trainable Method of Parametric Shape Description*, Image Vision Comput. Vol. 10, pp. 289–294, 1992
3. T. Cootes, C. Taylor, D. Cooper, J. Graham, *Active Shape Models – Their Training and Application*, Computer Vision and Image Understanding, Vol. 61, No. 1, Jan, pp. 38–59, 1995
4. J. Canny, *A computational Approach to Edge Detection*, IEEE Transactions PAMI, Vol 8(6), pp. 679–697, 1986
5. R.M. Haralick, K. Shanmugam, I. Dinstein, *Textural Features for Image Classification. IEEE*, SMC 3, pp. 610–621, 1973
6. T. Lehmann, B. Wein, J. Dahmen, J. Bredno, F. Vogelsang, M. Kohnen *Content–Based Image Retrieval in Medical Applications: A Novel Multi–Step Approach*, Proc. SPIE 3972, 2000 (in press.)
7. J.R. Quinlan, *Induction of Decision Trees*, Machine Learning, Vol. 1, pp. 81–106, 1986