ity of Groningen Faculty of Mathematics and Physical Sciences
Department of
Computing Science

The Use of Fuzzy Spatial and Geometrical Relations in a Visual Recognition System Based on Deformable Templates

Inge Gerrits

supervisors:
dr.ir. J.A.G. Nijhuis
prof.dr.ir. L. Spaanenburg
drs. H. Stevens

August 1997

University of Groningen **Faculty of Mathematics and Physical Sciences**

Department of
Computing Science

The Use of Fuzzy Spatial and Geometrical Relations in a Visual Recognition System Based on Deformable Templates

Inge Gerrits

supervisors:
dr.ir. J.A.G. Nijhuis
prof.dr.ir. L. Spaanenburg
drs. H. Stevens

August 1997

University of Groningen Faculty of Mathematics and Physical Science
Department of Computing Science

THE USE OF FUZZY SPATIAL AND GEOMETRICAL RELATIONS IN A VISUAL RECOGNITION SYSTEM BASED ON DEFORMABLE TEMPLATES

I. H. Gerrits
student Technical Computer Science
University of Groningen
August 1997

supervisors:
dr.ir. J.A.G. Nijhuis
prof.dr.ir. L. Spaanenburg
drs. H. Stevens

TABLE OF CONTENTS

ABSTRACT

This is a research for the use of fuzzy spatial and geometrical relations in a visual recognition system. Without fuzzy spatial and geometrical relation it is only possible to make a visual recognition system starting from a good image quality, which is practically impossible with currently recording techniques.

The application of the visual recognition system is to locate the vertebrae in X-ray images of scoliosis patients. Scoliosis is a lateral curvature of the spine, which is normally straight. In these X-ray images the vertebrae are not clearly visualised, i.e. edges are not sharp defined, some vertebrae are even not visible and vertebrae who are visible do not have the same light intensity. In spite of this bad quality it is not difficult for a physician to locate the vertebrae. A physician is able, owing to his prior knowledge to make a good estimate of the position of these vertebrae. The aim of our research is to automate this process.

A human approach is used for the system, which means that not only the question what prior knowledge does a physician use is asked, but also the question what do we see in these images. The prior knowledge can be formulated by fuzzy sets. These fuzzy sets defines the relations between vertebrae of the spine, such as likely deviation in direction and position. By combining fuzzy set theory with a visual recognition system one can try to analyse X-ray images like physicians can do. Below a simple model of object recognition in the human visual system is scheduled.

human approach

- **chaos**
- **edge detection**
- **label edges**
- **build objects**
- **correct results**

This model exist of five successive phases. Initial when the image is presented there is chaos, then edges of object are located and labelled by importancy. Objects are build from these important edges by using our prior knowledge. In this process we can make mistakes, but due to the prior knowledge we are able to correct these mistakes if the constructed objects does not fit to the expected one. The artificial visual recognition system is build similar to these successive phases.

The visual recognition system is tested on several images. Results indicate that in all these images the system is able to locate the vertebrae.

CHAPTER 1: INTRODUCTION

The aim of this research is to make a visual recognition system, that locate vertebrae in X-ray images. In these X-ray images the lateral spine is shown of **scoliosis** patients. Normally the lateral spine is straight, but the lateral spine of scoliosis patients is characterised by a curvature. These patients can be helped by a brace, that exercises pressure on certain vertebrae to push the spine in the normal state. Making a brace one needs a good three-dimensional model of the spine, modelled by a computer. This present research is limited to two-dimensional, because of the limited time to spend. But if one has a two-dimensional model, it is easier to make a three-dimensional model. A future three-dimensional recognition system naturally builds on the integration of information from two-dimensional images from different perspectives.

Vertebrae in X-ray images are not clear visualised, i.e. some parts of the vertebrae are not visible or vertebrae are not visible at all, the lighting of the image is not equally distributed and some part of the image are over-exposed. These factors makes it difficult for a visual recognition system to locate the vertebrae. In spite of the missing information a physician is able to make a good estimate of the position of the vertebrae owing to his **prior knowledge**. The missing information in the X-ray image is supplemented by the prior knowledge. This prior knowledge can be formulated by fuzzy spatial and geometrical relations of the vertebrae. Examples of **spatial relations** are 'above', 'below', 'near' and 'far', in which prior knowledge can be defined about the relation of the vertebrae in the spine. Examples of **geometrical relations** are 'height', 'width' and 'squares', that can be used to define prior knowledge about the shape of a vertebra. The strength of the relation is measured with a membership function expressing various "degrees" of strength of the relation on the unit interval [0,1]. By combining **fuzzy set theory** with a **visual recognition system** one can try to analyse an X-ray image like physicians can do.

With the use of **deformable templates** one can locate the contour of an object in an image. There exist a large variety of deformable templates, some will be discussed in chapter 3. The template is the prototype of the object and this prototype can be deformable to a certain degree, if it fits the object. The degree of deformation is determined by **local information** of the image. Due to additional noise, missing information and uncertainty in the image, the use of local information can be misleading. By using fuzzy geometrical relations of the object one can correct the results of the deformable template with the expected one.

In the following pages we first discuss theory (chapter 2,3 and 4), explaining the terms which can be used in the rest of the paper. Chapter 2 and 3 briefly discuss fuzzy set theory and deformable templates as they are used in literature. In chapter 4 the seriousness of scoliosis will be indicated. The research is defined in chapter 5, and chapter 6 explains the approach of the visual recognition system and shows the results. In chapter 7 this approach will be discussed.

CHAPTER 2: FUZZY SET THEORY

2.1 Fuzzy Set Theoretic Approach to Computer Vision

computer vision

Computer vision is the study of theories and algorithms for **automating** the process of visual perception [14], i.e. automatically extracting useful information by carrying out computations on images . This can be divided into three classes:

- low-level vision
- medium-level vision
- high-level vision

Uncertainty can arise in every phase of computer vision. In figure below the connection of these phases are shown.

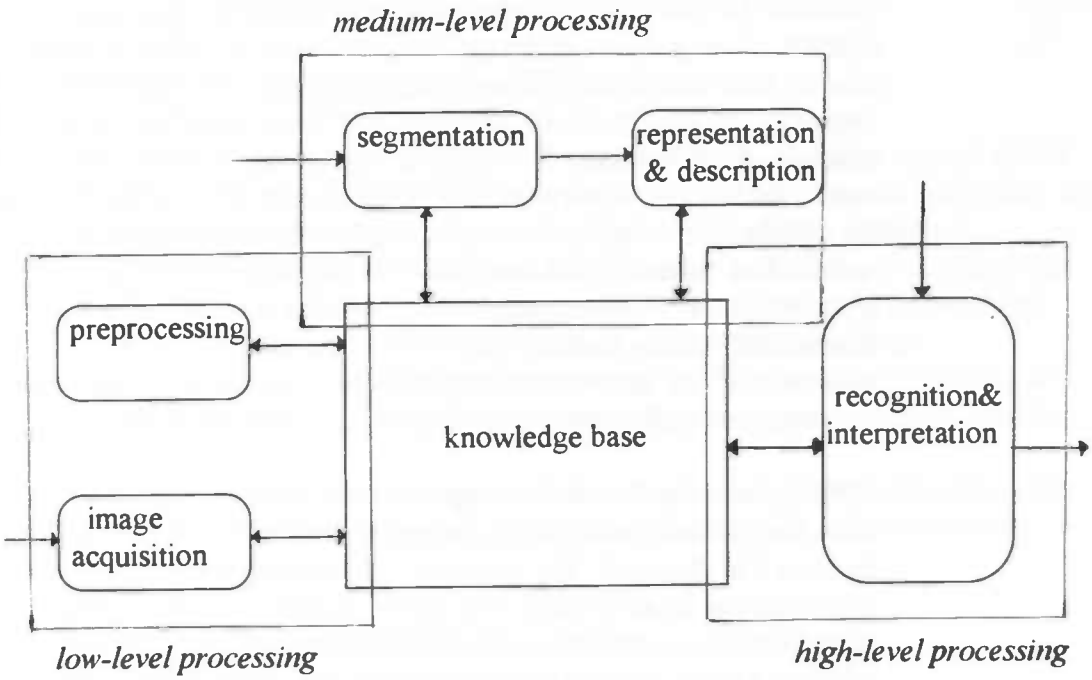


Figure: Levels in image analysis.

low-level

On the low-level, image processing is to obtain **elementary features**. Examples are noise removal, smoothing and sharpening of contrast. The image may contain additive and non-additive noise of various sorts and distributions. Removing this noise while keeping useful information undisturbed is very difficult. This causes the uncertainty in low-level vision.

medium-level

Grouping the elementary features of low-level vision belongs to medium-level vision. Examples are segmentation, region growing and matching. Imprecision in computations and vagueness in class definitions are examples of uncertainty.

high-level

Interpretation of the scene is high-level vision. **Interpretation** of the scene is to extract information of interest from a background of irrelevant details and to make inferences from incomplete information [24]. Ambiguities in interpretations and ill-posed questions are examples of uncertainty on high level vision.

uncertainty

To deal with **uncertainty** problems in computer vision different methods are developed.

- Bayesian belief networks
- Fuzzy set theory

Bayesian belief

This traditional method is based on the use of probabilities for expression the amount of belief in a property. The Bayesian paradigm consists of four successive stages [4]:

- Construction of a **prior probability distribution** $p(x)$ where x is to be reconstructed (contour of the object).
- Combining the observed image z with the underlying contour x through a **conditional probability density** $p(z|x)$.
- Constructing the **posteriori probability density** $p(x|z)$ from $p(x)$ and $p(z|x)$ by Bayes Theorem giving $p(x|z) = p(x)\,p(z|x)$.
- Base any inferences about x on the posterior distribution $p(x|z)$. One choice of inference can be to find the **Maximum a Posteriori** (MAP) estimate.

The Bayesian objective function $p(x|z)$ consists of two terms. The first term is a Bayesian conditional probability $p(z|x)$, which is a potential energy linking the edge positions and the gradient directions in the input image to the specified object boundary. The second term is a Bayesian prior probability $p(x)$, which penalises the various deformations of the specified object boundary (large deviations from the prototype result in a large penalty).

Fuzzy set theory

The idea proposed by Lotfi Zadeh suggested that set membership is the key to decision making when faced with **uncertainty** [25]. Classical sets contain objects that satisfy precise properties of membership; fuzzy sets contain objects that satisfy imprecise properties of membership, i.e. membership of an object in a fuzzy set can be approximate. The degree of membership is on the real continuous interval [0,1], where the endpoints of 0 and 1 conform to no

membership and full membership, respectively. The sets on the domain Ω that can accommodate "**degrees of membership**" were termed by Zadeh as "fuzzy sets". A grey level image is a function f: $R^2 \rightarrow R$ where f(x,y) is called the grey level of pixel (x,y). In order to apply the assortment of fuzzy set theoretic operators to an image, the grey levels must be converted to membership values. Let Ω be the domain over which the image function is defined. Then a fuzzy subset of Ω is a membership function $\mu_f : \Omega \rightarrow [0,1]$, where the value of μ_f (x,y) depends on the original grey level f(x,y). Many of the basic geometric properties of image regions and relationships among regions can be generalised to fuzzy subsets.

Properties of objects and **spatial relations** between objects play an important role in scene understanding [14]. Humans are able to quickly ascertain the relationship between objects, for example "B is to the right of A", but this has turned out to be a somewhat illusive task for automation. The determination of spatial relationships is critical for high-level vision processes involved in, for example, autonomous navigation, medical diagnosis, or more generally, scene interpretation.

2.2 Fuzzy Sets Used in Literature

Huntsberger, Rangarajan and Jayaramamurthy

Huntsberger et al.

Huntsberger, Rangarajan and Jayaramamurthy [13] developed a system called FLASH (Fuzzy Logic Analysis System in Hardware) which treats uncertainty in the input cue representation in the context of Zadeh's fuzzy set theory. Individual pixels in the input image are represented by their fuzzy membership to clusters, returned from an iterative segmentation technique. The low level portion of the FLASH system uses an iterative algorithm for image segmentation, based on clustering in an image colour space. The clustering in colour space is done with the fuzzy c-means algorithm generalised by Bezdek.

objective function

Optimisation of an objective function which encodes similarities between pixels gives the desired description of image colour characteristics. The objective function is defined as

$$J_m(U,v) = \Sigma^n_{k=1} \Sigma^c_{i=1} (\mu_{ik})^m (d_{ik})^2$$

where μ_{ik} is the fuzzy membership value of pixel k in cluster centre i, and d_{ik} is any inner product induced norm metric. The set of cluster centres v would be vectors in a colour space and as such represent the global colour characteristics of an image. The exponent m can be used to vary the nature of the clustering, ranging from absolute "hard" clustering at m = 1 to increasingly fuzzier clustering as m increases.

algorithm

The fuzzy c-means algorithm relies on the appropriate choices of U and v to minimise the objective function given above. This can be accomplished using the algorithm given below:

1) Fix the number of clusters c, $2 \leq c < n$ where n = number of data items. Fix m,
$1 \leq m < \infty$. Choose any inner product induced norm metric $\| * \|$.
2) Initialise the fuzzy c partition, $U^{(o)}$.
3) At step b, b = 0,1,2,···.
4) Calculate the c cluster centres $\{v_i^{(b)}\}$ with $U^{(b)}$ and the formula: cluster centre for cluster i equals

$$v_i = \Sigma^n{}_{k=1} (\mu_{ik})^m x_k / \Sigma^n{}_{k=1} (\mu_{ik})^m$$

5) Update $U^{(b)}$: calculate the memberships in $U^{(b+1)}$ as follows:

a) Calculate I_k and T_k :
$I_k = \{i \mid 1 \leq i \leq c: d_{ik} = \| x_k - v_i \| = 0\}$
$T_k = 1,2, ,c-I_k$.

b) For data item k, compute new membership values :

1) if I_k empty,
$\mu_{ik} = 1 / [\Sigma^c{}_{j=1} (d_{ik} / d_{jk})^{2(m-1)}]$

2) else, $\mu_{ik} = 0 \ \forall i$ member T_k,
and $\Sigma_{i \in I_k} \mu_{ik} = 1$.

6) Compare $U^{(b)}$ and $U^{(b+1)}$ in a convenient matrix norm: if $\|U^{(b)} - U^{(b+1)}\| \leq \varepsilon_L$, stop; otherwise, set b = b + 1 , and return to step 4.

initialisation

The initial partition $U^{(0)}$ can be done randomly with relative independence from the membership values returned after convergence. Data values x_k used in the induced norm metric are the input colour vectors. The fuzzy membership values μ_{ik} are used in the FLASH system to build a low-level representation of the pixels in the image which contains global information, while at the same time maintaining local information at the pixel level.

texture

The information returned from the segmentation phase of the FLASH system is by itself insufficient for general computer vision tasks. Since RGB colour is the feature space that the segmentation was performed in, regions that are found are fuzzy sets in that space. Previous studies using colour histogramming techniques have resulted in only 54 percent accuracy as far as meaningful region identification. This is due in part to the lack of shape and texture information in the segmentation phase. Texture has been shown to be a very important cue for region identification in humans.

distance metric

Since the FLASH system uses the fuzzy c-means algorithm for segmentation purpose, it relies on a distance metric. Selection of a distance metric in a texture feature space has many difficulties associated with it. Even using normalised texture measures is not sufficient, since the distance between two texture measures may not even be in the same space.

homogeneity

The behaviour of the fuzzy membership values in the transition between colour regions will be an indication of the strength of the edge between regions. In order to define an edge detector, information about the relative homogeneity of colours within regions and mixing of colours across the discrete digitised transitions between regions must be included in the definition. The relative homogeneity of a colour region given in terms of membership values to fuzzy sets in the segmentation can be written as

$$HOMOG_k(\mu_i, \mu_j) = \mu_i - \mu_j$$

where μ_i and μ_j are the fuzzy membership values associated with pixel k to sets i and j in the lowest level representation. This operation is applied to each pixel in an image after a descending sort is performed on the membership values. Values of homogeneity close to 1.0 would indicate that class i is the dominant colour characteristic for pixel k.

colour edge

The location of a colour edge can now be defined as the spatial location of the zero crossings of

$$EDGELOC_{kl} = HOMOG_k - HOMOG_l$$

where k and l are two adjacent pixels in the horizontal or vertical direction.

The strength of a colour edge in the FLASH system is defined as

$$\mu_E = |EDGELOC_{kl}|/2$$

This operation gives the membership value of the detected colour edge to the class of colour step edges. Values close to 0.25 and slightly above indicate diffuse edges (shadows), while sharp edges are characterised by values to 1.0.

The representation in the FLASH system for regions or blobs is n-sided polygons. These polygons are defined in terms of linking the edge elements given by the zero crossings of the operator μ_E.

Pathak and Pal

pathak et al.

Pathak and Pal [15] use a syntactic method, named fuzzy grammar, of structural pattern recognition. They are concerned with the syntactic classification into one of the possible stages of skeletal maturity. A three-stage hierarchical syntactic approach is presented for automatic recognition of the ages of

different bone. Two algorithms based on six-tuple fuzzy grammars and seven-tuple fractional fuzzy grammars have been used separately for classification at each stage. The primitives considered are a line segment of unit length, clockwise and counter-clockwise curves and a "dot". For any such curve they have defined its membership values corresponding to fuzzy sets of "sharp", "fair" and "gentle" curves.

stages

Two algorithms are illustrated with the help of an X-ray image of the radius of a 10-12 year old boy. They distinguish nine different stages of skeletal maturity of radius of hand and wrist.

stage A: Epiphysis[1] is totally absent.

stage B: Epiphysis gradually appears above the metaphysis[2] as a single deposit of calcium with irregular outline.

stage C: Epiphysis gradually assumes a well-defined oval shape, its maximum diameter being less than half the width of the metaphysis.

stage D: The Epiphysis continues to grow in size but becomes slightly tapering at its medial end, being more rounded at the lateral end.

stage E: Its maximum diameter now exceeds half the width of the metaphysis. The shape is more or less the same though it becomes larger and a thickened white line representing the edge of the palmar surface[3] appears within it at the distal border.

stage F: The palmar surface of the proximal border also develops and becomes visible as a thickened white line at the proximal edge of the epiphysis.

stage G: The palmar surface of the medial border also becomes apparent as a white line so that the three visible palmar surfaces combine to appear as a single continuous, thickened $\subset$-shaped contour.

stage H: The epiphysis caps the metaphysis almost entirely.

stage I: Fusion of the epiphysis and the metaphysis begins.

---

[1] An epiphysis, in some bones, is a separate terminal ossification which only becomes united with the main bone at the attainment of maturity.

[2] A metaphysis of a long bone is the end of the shaft where it joins the epiphysis.

[3] The palmar surface of any bone in the hand and wrist is that surface which is towards the palm of the hand.

definition 1a A **fuzzy grammar** (FG) is a six-tuple

$FG = (V_N, V_T, P, S, J, f)$

where
V_N : set of nonterminals, i.e. labels of certain fuzzy sets on V_T^* called fuzzy syntactic categories.
V_T : set of terminals such that $V_N^* \cap V_T = \varnothing$.
V_T^* : set of finite strings constructed by concatenation of elements of V_T.
P: set of production rules.
S: starting symbol ($\in V_N$).
J: $\{r_i \mid i = 1,2,...,n,\ n = \#(P)\}$, a set of distinct labels for all productions in P, where #(P) is the number of elements in the set P.
f: mapping f: $J \rightarrow [0,1]$, $f(r_i)$ denoting the fuzzy membership in P of the rule labelled r_i.

definition 1b For any string x having m (≥ 1) derivation(s) in the language L (FG), its **membership in L(FG)** is given by

$$\mu_{L(FG)}(x) = \max_{1 \leq k \leq m} [\min_{1 \leq i \leq l_k} f(r_i^k)],$$

where
k : index of a derivation chain leading to x.
l_k : length of the k th derivation chain.
r_i^k : label of the i th production rule in the k th derivation.

definition 2a A **fractional fuzzy grammar** (FFG) is a seven-tuple

$FFG = (V_N, V_T, P, S, J, g, h)$

where V_N, V_T, P and S are as above, and g and h are mappings from J into the set of nonnegative integers such that
$g(r_k) \leq h(r_k)$, $\forall r_k \in J$.

definition 2b The **membership** of any string x having m (≥ 1) derivation(s) **in the language L(FFG)** generated by FFG is

$$\mu_{L(FFG)}(x) = \sup_{1 \leq k \leq m} \Sigma_{j=1}^{l_k} g(r_j^k) / \Sigma_{j=1}^{l_k} h(r_j^k)$$

where 0/0 is taken to be zero by convention.

sharp, fair gentle

For any curve b, the degree of arcness $\mu_{arc}(b)$ has been defined in the primitive extraction algorithm as

$$\mu_{arc}(b) = (1 - l / p)^{Fc} \ , \ \mu_{arc} \in [0,1)$$

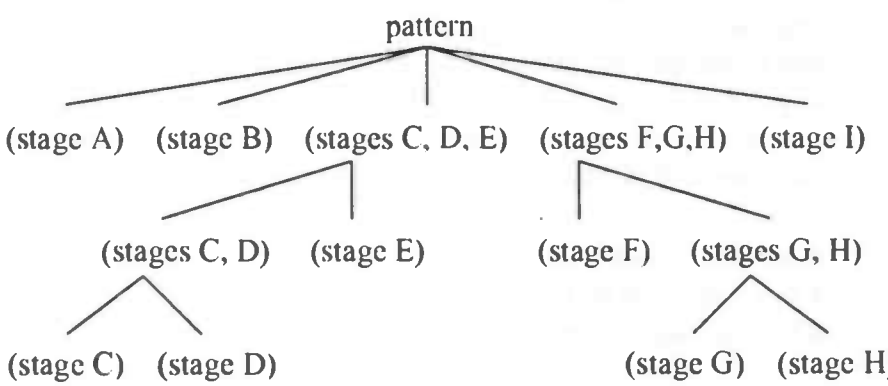
where l is the length of the line segment joining the two extreme points of the arc b, p is the length of the arc b, and F_c is a suitably chosen exponential fuzzifier with $F_c > 0$. The degrees of membership $\mu_S(b)$, $\mu_F(b)$, and $\mu_G(b)$ to the fuzzy sets of sharp, fair and gentle curves may be defined as

$$\mu_S(b) = f_S(\mu_{arc}(b))$$
$$\mu_F(b) = f_F(|\mu_{arc}(b) - 1/2 |)$$
$$\mu_G(b) = f_G(\mu_{arc}(b))$$

where
$f_S(\cdot)$ is a monotonically increasing function over [0,1].
$f_G(\cdot)$ and $f_F(\cdot)$ are monotonically decreasing functions over [0,1] and [0,1/2], respectively.
$\mu_S(b)$, $\mu_F(b)$, and $\mu_G(b)$ all take values in [0,1] only.

algorithm

The algorithm is like a tree with depth four. The first step classify a pattern into stage A, stage B and stage I and a group stages C, D, E and a group stages F, G, H. The second step classify the first group into stage E and a group stages C, D and classify the second group into stage F and a group stages G, H. The third step classify these last groups into stages C, D, G and H.



A context-free grammar is used with $V_T = \{a, b, b^*, c\}$. The a, b, b*, c denote a line segment of unit length, a clockwise curve, an anticlockwise curve, and a dot, respectively. Let x denote the string representing the contour of the epiphysis and y the string representing the interior of the epiphysis, i.e. the boundaries of the image of the palmar surface of the epiphysis. If x is found to be the empty string, they infer the class of stage A. If not, and if x is parsed by the first stage grammar, then

$$x \in_C L(G_k), \text{ if } \mu_{L(Gk)}(x) = \max_{2 \le i \le 5} \mu_{L(Gi)}(x) \quad k= 2,3,4,5.$$

Each of the stages A, B, and I is unique in itself. If x in stage A or stage B or stage I then stop; otherwise go to second step. If x is in the first group (stage C,D,E) then parse y by means of the second stage grammar and if y $\in_C L(G_E)$ then decide on stage E. If x is in the second group (stage F,G,H) then parse y by means of the second grammar and if y $\in_C L(G_F)$ decide on stage F. In the third step determine for the first group (stage C, D) D_E (the maximum diameter of the epiphysis) and W_M (width of the metaphysis). If $r = D_E/W_M \leq 0.5$, decide on event C; otherwise, decide on D. For the second group (stage G, H) determine S_E (the slope of the proximal edge of the epiphysis at the medial end) and S_W (the slope of the distal edge of metaphysis at the medial end). If $S = S_E - S_W$ is less than some predetermined α, suitably small, then decide on event H; otherwise, decide on event G.

Keller, Krishnapuram and Ma

Keller et al.

Keller, Krishnapuram and Ma [18] propose direct methods to analyse properties of fuzzy image regions and spatial relations between fuzzy image regions quantitatively. The methods use membership functions generated by a fuzzy segmentation algorithm such as the fuzzy C-means algorithm or by an unconstrained possibilistic C-means.

fuzzy subsets

The membership function μ_F for an object is defined by $\mu_F : \Omega \rightarrow [0,1]$. Each point x in Ω is assigned a membership grade, $\mu_F(x)$. It is further convenient to represent this object or region in terms of α-cut level sets, F^α, such that : $F^\alpha = \{\alpha \mid \mu_F(x) \geq \alpha\}$, where $\alpha \in [0,1]$. One popular method for assigning multiclass membership values to pixels, for either segmentation or other types of processing, is the fuzzy C-means algorithm. Let R be the set of real numbers and R^n be the n-dimensional vector space over the reals. Let X be a finite subset of R^n, $X = \{x_1, x_2, \ldots x_n\}$. In our case, each x_i is a feature vector for a pixel in the image. For an integer C, $2 \leq C \leq N$, a C x N matrix $U = \{u_{ik}\}$ is called a fuzzy C partition of X whenever the entries of U satisfy three constraints:

- $u_{ij} \in [0,1]$ for all i and j
- $\Sigma_{i=1}^{C} u_{ij} = 1$ for all j
- $0 < \Sigma_{j=1}^{N} u_{ij} < N$ for all i

However, being unsupervised, it is not possible to predict ahead of time what type of clusters will emerge from the fuzzy C-means from a perceptual standpoint. The partition generated by it may also be sensitive to noisy features and outliers. Also, the number of classes must be specified for the algorithm to run.

spatial relation

Properties and attributes of fuzzy image regions may be both geometric and nongeometric. Geometric properties that are frequently encountered are area, height, diameter, roundness and elongatedness. Nongeometric properties are brightness, colour and texture. The primitive spatial relations between two

objects are 1) LEFT OF, 2) RIGHT OF, 3) ABOVE, 4) BELOW, 5) BEHIND, 6) IN FRONT OF, 7) NEAR, 8) FAR, 9) INSIDE, 10) OUTSIDE, and 11) SURROUND. They define the relations as fuzzy sets over the universe of discourse of the α-cut values $\{\alpha_1,...,\alpha_n\}$. The general approach they use is as follows. Let A and B be two fuzzy sets defined on Ω. At each α-cut value α_i, they compute the membership value for "$A^{\alpha i}$ RELATION $B^{\alpha j}$" based on certain measurements γ on the relative positions of the pairs of elements (a,b), $a \in A^{\alpha i}$ and $b \in B^{\alpha i}$. These measurements are aggregated for all pairs of elements to give an aggregated measurement γ_i. The membership value for "$A^{\alpha i}$ RELATION $B^{\alpha j}$" denoted by $\mu_{A\_REL\_B}(\alpha_i)$ is then computed by evaluating a membership function μ_{REL} at γ_i. The overall membership value for "A RELATION B " is then computed using

$$\mu_{A\_REL\_B} = \Sigma_{i=1}^{n} (\alpha_i - \alpha_{i+1}) \mu_{A\_REL\_B}(\alpha_i).$$

example

Take for example the LEFT OF relation. Suppose we have two points A and B. Denote AB as the line connecting A and B. Let θ be the angle between AB and the horizontal line. The membership function for "A is to the LEFT of B" may be defined as a function of θ as

$$\mu_{LEFT}(\theta) = \begin{cases} 1, & |\theta| < a\pi/2 \\ \pi/2 - |\theta| / \pi/2(1-a), & a\pi/2 \le |\theta| \le \pi/2 \\ 0, & |\theta| > \pi/2 \end{cases}$$

A large value for a tends to give an optimistic result, and a small value would give a pessimistic result.

CHAPTER 3: DEFORMABLE TEMPLATES

3.1 Deformable Template Matching

matching

Template matching [2] is a filtering method of **detecting a particular feature** in an image. Providing that the appearance of this feature in the image is known accurately, one can try to detect it with an operator called a template. This template is, in effect, a subimage that looks just like the image of the object. A similarity measure is computed which reflects how well the image data match the template for each possible template location. The point of maximal match can be selected as the location of the feature. The similarity measure can be the cross-correlation or the sum of the squared or absolute image intensity differences of corresponding pixels [1].

deformable template

Deformed templates are obtained by applying parametric transforms to the prototype template, and the variability in the template shape is archived by imposing a probability distribution on the admissible mappings. Deformed templates are less sensitive to the signal-to-noise content of the images than the prototype template matching, where matching completely depends on the quality of the images. Deformable models that have been proposed in literature can be partitioned into two classes [3]:

- free-form
- parametric

free-form

Free-form deformable models **do not have a global structure** of the template, the template is constrained by only local continuity and smoothness constraints. Since there is no global structure for the template, it can represent an arbitrary shape as long as the continuity and the smoothness constraints are satisfied. An example of free-form deformable model is the energy-minimising spline, called snakes.

parametric

A parametric deformable template is used when some **prior information of the geometrical shape** is available, which can be encoded using a small number of parameters. The prior shape information of the object of interest is specified as a template [3]. This prototype template is not parameterised, but it contains the edge/boundary information in the form of a bitmap.

3.2 Deformable Templates Used in Literature

Jolly, Lakshmanan and Jain

Jolly et al.

Jolly, Lakshmanan and Jain [11] propose a segmentation algorithm using deformable template models to segment a vehicle of interest both from the stationary complex background and from other moving vehicles in an image sequence. This segmentation is difficult due to the complex nature of the background. By using a deformable template based Bayesian scheme they are able to overcome this inherent difficulty. There is a considerable amount of **domain-specific knowledge** in their segmentation problem:

1) the object of interest is a vehicle,
2) it is located in the lane closest to the camera,
3) it is moving,
4) its edges in the image are well defined.

They incorporate this prior knowledge by using deformable template models, and pose the vehicle segmentation problem as a Bayesian energy minimisation problem.

template

They define the prototype template of a generic vehicle as a polygon characterised by N vertices $\Theta = (\theta_1,\theta_2,...,\theta_N)$ in the 2D plane as shown in the figure, where $\theta_i = (X_i,Y_i)$ and X_i,Y_i are the Cartesian co-ordinates of θ_i. Each pair of successive points (θ_i, θ_{i+1}), i = 1,..., N defines a boundary segment φ_i of the template. They assume that the boundary is closed, so for notation purposes, $\theta_{N+1} = \theta_1$.

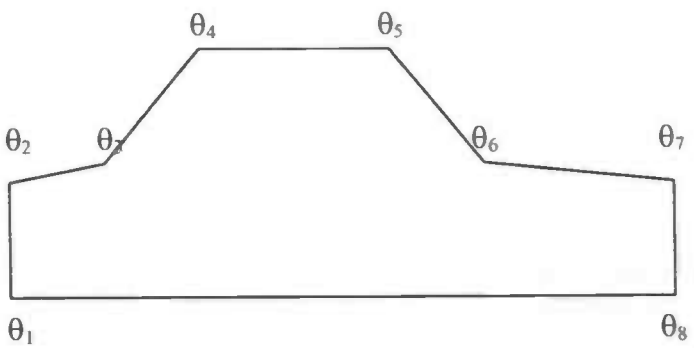


Figure: Prototype template

In order for the deformed template to resemble a vehicle, some constraints on the relationship between the different vertices $(\theta_1, \theta_2, \ldots, \theta_N)$ must be satisfied. For example, φ_4 should always be above φ_8, φ_1 should always be located to the left of φ_7, etc. A set of rules on the template parameters Θ constrains the shape of the vehicle template.

algorithm

They use the Bayesian method as following:

1) $T(\Theta)$ defines the template.
2) The prior probability density function $p(\Theta)$ is defined as
   $p(\Theta) = k_l \exp \{- \Sigma_{l=1}^{M} g_l(\Theta)\}$,
   where g_l is a function for a rule and k_l denotes the normalising constant.
3) They model the likelihood probability density function $p(Z|\Theta)$ as a Gibbs distribution whose exponent comprises of two terms. The first term is a function which is derived from the fact that the vehicle of interest is moving. It attains its maximum value when the deformed template encompasses only those pixels that are moving. The second term is a directional edge-based function. It attains its maximum value when the contours of the deformed template coincide with underlying image edges that have strong gradient magnitude and whose gradient orientation is perpendicular to the contour.
4) The **Metropolis algorithm** is used for finding the maximum a posteriori (MAP) probability $p(\Theta|Z)$. The Metropolis algorithm is a simulated annealing procedure, it minimises the energy function $E(\Theta, Z)$ by constructing a sequence of template deformations $\Theta^{(1)}, \Theta^{(2)}, \ldots$, starting from a prototype template $\Theta^{(0)}$, such that $\lim_{k\to\infty}\Theta^{(k)} = \Theta^{*}$. At each iteration k, the algorithm determines a new value $\Theta^{(k+1)}$ of the deformation parameters, based on their current value $\Theta^{(k)}$. First a parameter Θ^{*} is selected at random from the neighbours of $\Theta^{(k)}$, and then $\Theta^{(k+1)}$ is determined as

$$\Theta^{(k+1)} = \begin{cases} \Theta^{*} & \text{with probability } p_k \\ \Theta^{(k)} & \text{otherwise} \end{cases}$$

where

$$p_k = \min (\exp\{[E(\Theta^{*},Z) - E(\Theta^{(k)},Z)]/ T_k\},1)$$

and T_k is a monotonically decreasing sequence such that
a) $T_k \geq T/(1+ \log k)$ for sufficiently large T,
b) $\lim_{k\to\infty} T_k = 0$.

Temperature Schedule

The temperature schedule T_k is a critical component of the Metropolis algorithm. The temperature schedule $T_k = T/(1+\log k)$ requires a sufficiently large value of T. This choice is difficult, because if T is chosen too large, then the algorithm requires too many iterations for convergence, whereas if T is chosen too small, then the algorithm converges to a local minimum relatively close to the starting position $\Theta^{(0)}$. An alternative is the **geometric temperature schedule**:

$$T_k = T_0(T_f/ T_0)^{k/Kmax}$$

where T_0 is the starting temperature, T_f is the final temperature, and K_{max} is the number of iterations. In the geometric temperature schedule, compared to the

logarithmic schedule, the temperature does not initially drop too rapidly, but it approaches the zero value much faster.

results

They do, however, experience some difficulties even with the geometric temperature schedule, particularly in choosing values for T_0 and K_{max} . It can be seen that the segmentation and classification results are strongly affected by the choice of T_0 and K_{max}. In 12 of the 18 cases, the vehicle was correctly classified as a van, but in two of those cases, the shape of the template is not very accurate. In the remaining six cases, the vehicle was mistaken for a sedan or a pickup truck. Unfortunately, the best set of values for T_0 and K_{max} varies with the input image.

Rueckert and Burger

Rueckert et al.

Rueckert and Burger [12] propose a hybrid **Multi-Temperature Simulated Annealing** optimisation to minimise the energy of a deformable model. The performance and robustness of the algorithm have been tested on spin-echo MR images of the cardiovascular system. The elasticity of an artery like the aorta plays as important role in cardiovascular haemodynamics and its measurement may be a method of detecting and monitoring cardiovascular diseases. The elasticity of the aorta can be measured as aortic compliance, which is the change in volume per unit change in blood pressure. MR imaging methods of the cardiovascular system provide a direct and non-invasive method of studying the elasticity or compliance of the aorta. The aorta compliance can be calculated from two separate spin-echo images taken at the end of systole and diastole during the cardiac cycle. The task was to segment the ascending and descending aorta in 15 datasets of different individuals in order to measure regional aortic compliance. The quality of the images can vary considerably and is often poor due to the low signal-to-noise ratio of the images which is caused by short acquisition times. Moreover, the boundary of the aorta can be obscured by motion and blood flow artefacts.

template

Geometrically Deformable Models (GDM) are defined as a set of vertices which are connected by edges. A more efficient and stable representation is obtained by using the vertices as control points $P = \{P_1, ..., P_n\}$ which are connected by a set of locally controlled, C^2 continuous spline curves $Q = \{Q_1, ..., Q_n\}$. This representation has several advantages: It yields an analytic and differentiable curve representation. Moreover, it provides a compact, smooth and continuous object representation. An example of such spline-based GDM representation is given in the figure below.

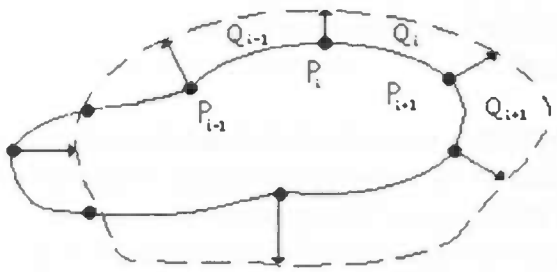


Figure: A spline-based GDM representation is defined by control points P_i which are connected by spline curves Q_i.

deformation

During the deformation process, the initial model is deformed by moving the vertices along the direction of their normal vectors n_i at each vertex. The deformation is controlled by the minimisation of its associated energy function E.

segmentation algorithm

The segmentation algorithm starts by constructing a linear scale-space of an image through convolution of the original image with a Gaussian kernel at different levels of scale, where the scale corresponds to the standard deviation of the Gaussian kernel. Large scales suppress may smaller scale features as well as noise. This often leads to a fast convergence of the model because only strong edges are retained, but the model might miss contours with weak edges. Small scales identify small and large scale features with accurate location but at the same time the model becomes more sensitive to noise. In order to maximise the accuracy of the segmentation, the contour of the object of interest is tracked in scale-space from coarse to fine scales. A multiscale representation $L(x,y,\sigma)$ of a 2-D image $I(x,y)$ can be obtained by a convolution with a Gaussian kernel,

$$L(x,y,\sigma) = I(x,y) \otimes 1/(2\pi\sigma^2) \exp (- (x^2 + y^2) / 2\sigma^2)$$

resampling

The resampling process controls the resolution of the deformable model. The resolution of the deformable model depends on the number of vertices and their spacing and should be closely related to the scale at which the image is treated. The resampling algorithm automatically adapts the resolution to the current scale. This is done by fixing a distance d_l which is proportional to the actual scale σ_l:

$$d_l = d_0 \, \sigma_l / \sigma_0$$

where d_0 is the resolution of the deformable models at lowest-scale. For each curve segment Q_i which is longer than d_l a new control point is inserted between P_i and P_{i+1}. After resampling the deformable model, the deformation process is repeated at the next lower level of scale l - 1.

energy function

The total energy can be written as the sum of the energies of all curve segments:

$$E_{total}(Q,\sigma) = \Sigma_{i=1}^{n} E_{intern}(Q_i,\sigma) + E_{extern}(Q_i,\sigma)$$

The **internal forces** of a curve segment Q_i enforce a smooth contour and can be expressed by the first and second derivative of the contour:

$$E_{intern}(Qi, \sigma) = \int_0^1 \alpha_1 |Q_i`(\tau)|^2 + \alpha_2 |Q_i``(\tau)|^2 \, \delta(\tau)$$

The first derivative term controls the elasticity of the contour while the second derivative term controls the bending. Setting α_1 to zero would allow the contour to develop gaps; setting α_2 to zero would allow the contour to develop corners.

The **external energy** introduce a coupling of the deformable model to image features. Image features can be expressed in terms of differential invariants which are invariant to the choice of the underlying co-ordinate system. The magnitude of the gradient operator ∇ I of the image intensity function measures the strength of an edge and can be used to attract the contour towards prominent edges. The zero-crossings of the Laplace operator ΔI, or the second derivative of the image intensity function, indicate a local minimum or maximum of the first derivative which is equivalent to the centre of an edge. The Euclidean distance D(ΔI) of the zero-crossings can be used to attract the contour towards the centre of edges.

$$E_{extern}(Q_i, \sigma) = \Sigma_{(x,y) \in Q_i} - \alpha_3 | \nabla L(x,y,\sigma) |^2 + \alpha_4 | D(\Delta L(x,y,\sigma)) |^2$$

The relative importance of the edge strength is controlled by the constant α_3 while the constant α_4 controls the importance of the edge localisation accuracy. The weighting factors α_i for the internal and external energy terms has to be fixed by the user.

multiscale algorithm

The multiscale algorithm requires an energy minimisation technique which is able to change its behaviour with decreasing scale. At **coarse scales** the minimisation technique should be independent if its initialisation and insensitive to local minima. At **fine scales** the minimisation technique should be conditionally dependent of its initialisation and locally controlled. As a consequence they have implemented a hybrid minimisation technique, Multi-Temperature Simulated Annealing (MTSA). The basic idea of MTSA is the following: At the highest level of scale (lowest resolution) the Simulated Annealing process is started at very high temperatures which enable the algorithm to escape local minima in the energy space. The algorithm stops if the system is frozen which corresponds to a global minimum of the energy function. This final result is then resampled and subdivided and used as an initialisation for the next lower level of scale.

initial temperature

A geometric schedule for selecting the initial temperature at different levels of scale is given by

$$T^{l+1}_{initial} = b \cdot T^{l}_{initial} \quad \text{with } 0 < b < 1$$

where $T^{l}_{initial}$ denotes the temperature at the first iteration at scale l. The annealing process at each level of scale is then carried out according to the **cooling schedule**

$$T(k) \geq c / \log (1+k)$$

where c is a constant depending on the amount of energy which is necessary to escape local minima.

results

They have randomly selected a subset of 15 subjects and compared the multiscale and the classic contour fitting algorithm with the clinically accepted reference method which is the manual segmentation of the aorta in both images by an experienced radiologist. The manual segmentation is repeated four times by the same observer and the average of these four segmentations is considered to be the true segmentation result. The results of the computer-based algorithm have been compared with intra-observer variability of the radiologist which is 3.76 % in these images.

The first set of tests have been made using the classic contour fitting algorithm at fixed scales. The energy function is minimised with Simulated Annealing minimisation. The segmentation error at a coarse scale is 18.72 % and continuously decreases until at a finer scale a minimal segmentation error with 3.89 % is reached.

The second set of tests have been made using the proposed multi contour fitting algorithm and MTSA energy minimisation. At high levels of scale they found very similar results to those of the classic contour fitting algorithm. However, the multiscale algorithm does not increase the segmentation error for finer

scales. Instead the segmentation error decreases constantly. The segmentation error at the lowest level of scale is 2.87 %

Jain, Zhong and Lakshmanan

Jain et al.

Jain, Zhong and Lakshmanan [3] propose a general object localisation and retrieval scheme based on object shape using deformable templates. A Bayesian scheme, which is based on prior knowledge and the edge information in the input image, is employed to find a match between the deformed template and the objects in the image.

template

The prototype template T_0 consists of a set of points on the object contour, which is not necessary closed, and can consist of several connected components. They represent such a template as a bitmap, with bright pixels lying on the contour and dark pixels elsewhere. Such a scheme captures the global structure of a shape without specifying a parametric form for each class of shapes. This model is appropriate for general shape matching since the same approach can be applied to objects of different shapes by drawing different prototype templates.

algorithm

A Bayesian inference scheme is employed to integrate the prior shape knowledge of the template and the observed object in the input image.

1) T_0 defines the prototype template.
2) $T_{s,\Theta,\xi,d}$ defines a deformation of the prototype. This deformation is realised by scaling the local deformation by a factor s, rotating the prototype template by an angle Θ, locally deforming the rotation by a set of parameters ξ, and translating the scaled version along the x and y direction by an amount d. Assuming that s, Θ, ξ, and d are all independent of each other, they get the following prior probability density function:

$$p(s, \Theta, \xi, d) = p(s) \times p(\Theta) \times p(\xi) \times p(d)$$

3) The likelihood they propose only uses the edge information in the input image. The deformable template is attracted and aligned to the salient edges in the input image via a directional edge potential field. This field is determined by the positions and directions of the edges in the input image. For a pixel (x, y) its edge potential is defined as:

$$\Phi(x, y) = - \exp \{- \rho (\delta^2_x + \delta^2_y)^{1/2}\},$$

where (δ_x, δ_y) is the displacement to the nearest edge point in the image, and ρ is a smoothing factor which controls the degree of smoothness of the potential field. They modify this edge potential by adding to it a directional component.

This new edge potential induces an energy function that relates a deformed template $T_{s,\Theta,\xi,d}$ to the edges in the input image I:

$$E(T_{s,\Theta,\xi,d}, I) = 1/ n_T \Sigma (1 + \Phi(x,y) | \cos(\beta(x,y)) |),$$

where the summation is over all the pixels on the deformed template, n_T is the number of pixels on the template, $\beta(x,y)$ is the angle between the tangent of the nearest edge and the tangent direction of the template at (x,y), and the constant 1 is added so that the potentials are positive and take values between 0 and 1. This definition requires that the template boundary agrees with the image edges not only in position, but also in the tangent direction. The lower this energy function the better the deformed template matches the edges in the input image. Using this energy function they define the likelihood probability density as follow:

$$p(I | s,\Theta,\xi,d) = \alpha \exp\{ - E(T_{s,\Theta,\xi,d}, I) \}$$

where α is a normalising constant to ensure that the above function integrates to 1.

4) The a posteriori probability density of the deformed template is given as:

$$p(s,\Theta,\xi,d | I) = p(I | s,\Theta,\xi,d) p(s,\Theta,\xi,d) / p(I)$$

They have employed a **multiresolution approach**. At the coarsest stage, a smooth potential field is used with a large value of ρ (smoothing factor in likelihood). This stage attempts to roughly locate the global optima efficiently without regard to accuracy. At finer stages, finer energy potential fields are used because more accurate localisation is desired. The deformed templates obtained from the previous stage with low energy (below a threshold) are used as initial templates for the matching. A larger number of deformation parameters and finer step sizes are used, only if the energy is below a certain threshold, to obtain better matches.

results

Given an input image, they start out by sketching a prototype template which resembles an object of interest in the image. Their experimental results have been divided into six sets. The first set demonstrates how the given prototype templates deform themselves locally to match the object contours in the input images when placed in a neighbourhood of the desired objects. The second set illustrates that their matching scheme can localise objects independent of their location, and orientation in the image. The third set demonstrates that the matching scheme is able to retrieve all the objects in an input image that resemble the prototype template. In the fourth set they illustrate the fact that the matching scheme can handle prototype templates that are not closed. The fifth set illustrates the scale invariance aspect of the matching scheme. And the sixth set illustrates that the objective function also can reject the hypothesis that a certain object is present in an image.

Storvik

Storvik

Storvik [4] discuss a method for curve detection based on a fully Bayesian approach. A model for image contours which allows the number of nodes on the contours to vary is introduced. Iterative algorithms based on **stochastic sampling** is constructed, which make it possible to simulate samples from the posterior distribution, making estimates and uncertainty measures of specific quantities available. In practice, computational aspects must taken into consideration when choosing the models. The approach is applied to ultrasound images of the left ventricle and to Magnetic Resonance images of the human brain.

template

Under the assumption that the image consists of only one object having a simply connected domain, the image x of interest is completely defined by the contour of the object. A polygon representation of an object is a representation where the contour is defined by a set of nodes giving co-ordinates of points on the contour in a circular (clockwise) manner. Between each node, the contour is defined by a straight line. The figure below illustrates the representation.

Figure: polygon-representation of the contour for a simply connected object.

algorithm

They follow the Bayesian approach for developing models for simply connected objects.

1) The template is defined by $x = (X_1, \ldots, X_N)$ where each X_i gives the co-ordinates of a point on the contour. N is the number of nodes, and may be stochastic.
2) Energy-functions are usually easier to formalise than probabilities. They therefore assume that p(x) is of the form

$$p(x) = 1/Z \exp\{ -U(x)/T \}$$

where U(x) is called the energy function. The vector $U(x) = \{U1(x), \ldots, Up(x)\}$ contains components measuring various characteristics of the contour. Z is a normalising constant ensuring the probability function to be a proper distribution function, usually unknown because of huge number of possible configurations x. The constant T is usually referred to as the "temperature". The a priori distribution should capture the knowledge available about x.

A common assumption is that the energy-function is built up by potentials measuring local characteristics. In that case

$$U(x) = \Sigma_i \, V_i(x)$$

where the sum is over all nodes on the contour and $V_i(x)$ is some measure depending only on the nodes in a small neighbourhood of node i. In the case of random number of nodes, an alternative could be to use the average of the potential measures,

$$U(x) = 1/|x| \; \Sigma_i \; V_i(x)$$

where $|x|$ is the number of nodes on the contour x.

For the active contour approaches, the derivatives and second derivatives of the contour have been used as smoothing measures. The derivatives are approximated by

$$V_i'(x) = \| x_i - x_{i-1} \|^2,$$
$$V_i''(x) = \| x_{i+1} - 2\, x_i + x_{i-1} \|^2,$$

which are used as potentials in the energy function.

3) The probability density is related to the specification of the observed image data.

$$f(z \mid x) = 1/Z \, \Pi_i \exp\{-h(x_i; z)\}$$

where $h(x_i; z)$ is some local measure from the observed image z at location x_i, and Z is a normalisation constant.

4) To obtain the MAP estimate they use the Metropolis algorithm. Construction of a **Metropolis algorithm** mainly involves the definition of a transition-matrix defining the possible transitions at each iteration step. The transition matrix has to be constructed such that the resulting Markov chain {X(s)} will be irreducible and aperiodic. Randomness in the number of nodes complicates the matter. In particular, making changes on x by moving only a node from one location to another will not suffice. A technique for simplifying this problem is however to extend the configuration space Ω by introducing a stochastic variable indicating the position of an object moving around the curve. That is, they are defining the extended configuration space

$$\Omega^* = \{ x^* = (x, p^x); x \in \Omega, p^x \in x\}$$

where p^x is the position node and $p^x \in x$ means that p^x is a node on the contour. Conditioned on a configuration, the position node is assumed to have a uniform distribution on the set of nodes.

Start with an arbitrary $X^*(0)$. For each iteration s carry out the following steps.

a) Assume that $X^*(s) = x^* = (x,p^x)$.

b) Select a state $y^* = (y,p^x)$ by the distribution given by the x*th row of the transition matrix Q.

c) Change the contour to z = x with probability $1 - \alpha_{x^* y^*}(s)$, where

$$\alpha_{x^* y^*}(s) = \text{MIN}\{1, p(y\,;\,s)/\,p(x\,;\,s)\}$$

and

$$p(x\,;\,s) = 1/Z_s \exp\{-U(x) / T(s)\}$$

where T(s) is a decreasing sequence, $T(s) = c/\log(s+1)$.

d) Draw a new position node p^z using the (z, p^x)th row of transition matrix R and put $X^*(s+1) = (z, p^z)$.

Implementation Issues

When implementing the approach discussed above, care has to be taken in order to keep the computations cost reasonable. For a given model, the computational considerations are influenced by the choice of both contour representation and algorithms. To make the computation of models easy, they assume that each pixel in the observed image z will either be completely inside or completely outside the contour. Furthermore, the length between each node on the contour is assumed to be fixed and equal to the side-length of the pixels (assuming pixels to have quadratic shapes). This restriction forces the contour to follow the pixel-sides. Furthermore, the set of possible locations of nodes is restricted to the corners of the pixels (see figure below).

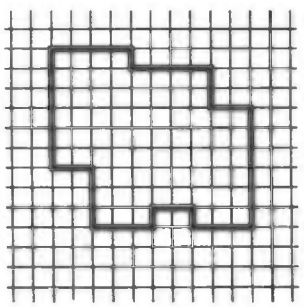


Figure: Example of a configuration when the contour follows the pixel-sides.

Changes at each iteration is performed by changing one or a few pixels through adding or subtracting these from the object region. Care has to be taken in order to make the new configuration legal. Up to three pixels were allowed to be changed at each iteration.

results

They consider two examples, one from ultrasound images of the left ventricle, and one from Magnetic Resonance images of the human brain. The first example shows after 100 000 000 iterations (taking 137 minutes of computer time on a DECstation 5000/25) with c = 100, a solution that provides a very

algorithm

Their general procedure is to consider the transformations $(s_0, s_1, ..., s_{n-1})$, without the regularity conditions, to be given by an S-valued **Markov process** on the edge graph associated with the connector graph (σ):

$$1/Z \ \Pi_{(i1,i2)} A(s_{i1}, s_{i2}) \ \Pi_i Q(s_i)$$

where the first product is over all neighbouring edges (i_1,i_2) and the second is over all edges. The acceptor function A expresses the stochastic dependence between adjacent group elements s_{i1} and s_{i2}, while Q is a weight function which expresses the varying preferences for different s values. The constant Z is a normalising constant. Then they obtain the prior measure on $E(c^0)$ by conditioning on the regularity constraints. The actual form of A and Q is dictated by the application.

example

In the case of modelling two-dimensional shapes one can model them discretely via polygons, then the components would be:

1) generators (G): polygonal edges,
2) connector graph (σ): cyclic graph,
3) regularity (R): polygons being simple and closed,
4) transformation group (S): GL(2) Euclidean of scale.

One method for constructing the probability measures is to have the components of matrices $(s_0, s_1, ..., s_{n-1})$ be independent, first-order Markov processes:

$$S_i = \begin{bmatrix} 1+u_i & v_i \\ wi & 1+z_i \end{bmatrix}, \quad S_i = \begin{bmatrix} 1+u_i & v_i \\ -v_i & 1+u_i \end{bmatrix} \text{ or } \quad S_i = \begin{bmatrix} 1+u_i & 0 \\ 0 & 1+u_i \end{bmatrix}$$

In the first case, S is GL(2) and $\{u_i\}$, $\{v_i\}$, $\{w_i\}$, $\{z_i\}$ are zero-mean, first-order Markov processes on the cyclic graph. In the second case, S is Euclidean group × scale and there are just two processes. In the third case, S is the scale group and there is only one process, $\{u_i\}$. As an example, for S equal to Euclidean × scale, before requiring regularity, the density is:

$$1/Z \ \Pi_{i=0}^{n-1} \exp\{-1/2\sigma_i^2 (u_i - a_i u_{i+1})^2 - 1/2\tau_i^2 (v_i - b_i v_{i+1})^2\}$$

where $\{a_i, \sigma_i, b_i, \tau_i\}_{i=0}^{n-1}$ are the parameters of the two Markov processes and Z is the normalising constant. Their prior probability measure is then obtained by requiring regularity; that is, conditioning on $(s_0, s_1, ..., s_{n-1})$ satisfying the closure constraints:

$$\Sigma_{j=0}^{n-1} s_j g_j^0 = 0$$

where $\{g_j^0\}$ are the edges of the polygonal template c^0.

results

In one experiment adult-male right hands are modelled. The observed images are 128 × 120, digitised visible light pictures taken under varying degree of optical noise. The template c^0 was data based. It was estimated from hand boundaries obtained from pictures with high signal-to-noise ratio. The approach taken in this application was actually multi-stage in nature. In the first stage, the group S is the affine group and the transformation is applied to the entire template, resulting in a new template. In the successive stages the transformation group elements are allowed to differ on different edges and the group S used in these later stages was Euclidean × scale. The hands experiment included pattern synthesis and restoration, along with the development of a general framework for shape modelling. Many analytical and computational questions need to be answered before this approach is fully implementable.

CHAPTER 4: SCOLIOSIS

4.1 Introduction

Scoliosis is defined as a **lateral curvature** of the spine [26]. The presence of a lateral curvature is abnormal, the lateral spine of a person is normally straight. Scoliosis as a physical deformity is accompanied by functional changes in the thoracic and abdominal organs, and psychic and emotional disturbances. The extent of the functional changes in the heart, lungs, and other viscera is in direct proportion to the degree of physical deformity [27]. For severe deformities there may be many and marked changes in the visceral functions, and life itself may be threatened.

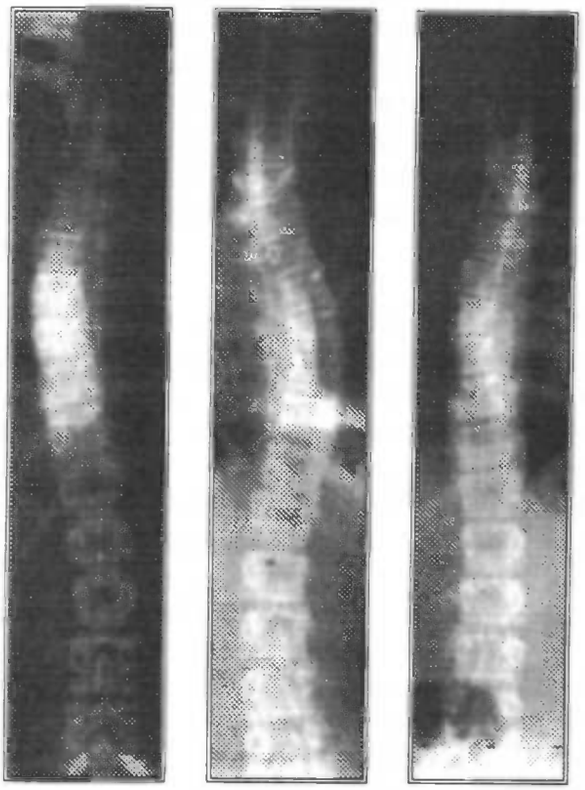


Figure 4.1: The lateral spine of scoliosis patients.

4.2 Basic Causes of Scoliosis

Traditionally patients with scoliosis can be categorised in one of three groups of structural scoliosis [26], namely:

1. Congenital.
2. Paralytic
3. Idiopathic.

congenital

The first term congenital is used to designate abnormalities, that are present at **birth**. In the extreme there are multiple malformed vertebrae, many ribs are fused or absent, the curve is severe and long, there is no compensation and the prognosis is poor. The causes of congenital deformities can be inherited, due to mutations, or due to toxic or mechanical influences during early development in utero.

paralytic

The second group contains those with **paralysis** of one or more of trunk muscles. Asymmetrical muscle action may lead to deformity. It is noteworthy that the problem of deformity arises only in a growing child, where this contracture is solely due to a lack of stretching of the muscles, because the antagonist is paralysed or to an actual shortening. It is the failure of a muscle to keep up with skeletal growth rather than actual shortening of muscle and tendon length.

idiopathic

Idiopathic is a group in which the exact cause of deformity is **unknown**. They are the most common curves and with the virtual disappearance of paralytic scoliosis they take on an even greater importance than in the past. Scoliosis of idiopathic origin occurs in the thoracic and lumbar vertebrae and its onset may be seen at all ages, from the new-born infant to the almost fully grown boy or girl, who may in the last years of growth suddenly develop a curve.

other diseases

There are also many rare but important diseases that produce scoliosis. The diseases are individually uncommon and therefore scoliosis arising from these diseases is also rare, but scoliosis occurs with great frequency amongst those afflicted. The conditions, which may cause scoliosis, occurs total over 50 different diseases and syndromes; some are exceedingly rare, others are frequently seen [26]. In nearly all these cases the causative disease is more significant than the scoliosis, which is often only one of several problems associated with the primary disease.

4.3 Treatment

The treatment of scoliosis was originally empirical and consisted of measures that seemed to reduce the deformity. These measures, crude and violent though they were, nevertheless depended on therapeutic principles, which still constitute essential features of modern treatment, namely, the elimination of the force of gravity, the use of traction as the basic corrective force, and the application of pressure over the convexity of the curve. In treatment, there was no great change until the eighteenth century, when fixation of the back in an improved position was combined with corrective forces (a brace). Gymnastic exercises came into vogue and have remained an important aid in therapy ever since.

CHAPTER 5: RESEARCH SITUATION AND RESEARCH DEFINITION

5.1 Research Situation

The use of images for recognition implies intensive calculations starting from a good image quality, which is practically impossible with current recording techniques. Owing to this there will be vagueness and inaccuracy in the images. This is the reason why there are few industrial applications for visual recognition systems.

Scoliosis, as can be concluded from the previous chapter, is a very serious disease. To have a good treatment for these patients, it is important to have an accurate picture of the spine. X-ray images are used, but to study the development of the treatment a lot of X-ray images are needed. Because of the **limited amount of X-rays** one may receive in one year, it is not possible to produce a lot of X-ray images. It is preferable to make a model of an X-ray image by a computer program and update this model by less harmless scanning-methods.

To produce a model of the spine one needs to locate the vertebrae in the X-ray image. These X-ray images are of **bad quality**, i.e. edges are not sharp defined, some vertebrae are even not visible, some pieces of a vertebra are not visible and vertebrae which are visible do not have the same light intensity. In spite of this bad quality it is not difficult for a physician to locate the vertebrae. A physician is able owing to his **prior knowledge** to make a good estimate of the position of these vertebrae.

5.2 Research Definition

The vertebrae in an X-ray image are determined manually by a physician. The aim of our research is to automate this process. Because of missing information in the X-ray image, it is impossible to make a visual recognition system based only on computer vision. This missing information need to be supplemented like a physician does. The prior knowledge can be formulated by fuzzy sets. By combining **fuzzy set theory** with a **visual recognition system** one can try to analyse an X-ray image like physicians can do. The precise research definition is as follows:

The use of fuzzy spatial and geometrical relations in a visual recognition system based on deformable templates.

5.3 Prior Knowledge

vertebra

The human spine consists of 30 vertebrae, from head downwards 8 cervical vertebrae, 12 thoracic vertebrae, 5 lumbar vertebrae and 5 sacral vertebrae [30]. Each vertebra has his own name, counting from head downwards they are C1, .. , C8, T1, .. , T12, L1, .. , L5, S1, .. , S5. The X-ray images used in this research are dorsal, which means seen from the back. Not all 30 vertebrae are registered in these images. The lowest vertebra that is visible is lumbar 5, denoted by L5, and is the first one above the hip. The highest vertebra that is visible is one of the thoracic vertebrae. The vertebrae increase in size from cervical to sacral. A dorsal vertebra can be visualised as in figure 5.1. The two elliptical areas in the vertebra are the pedicles.

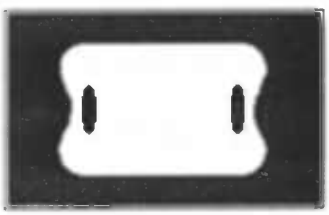
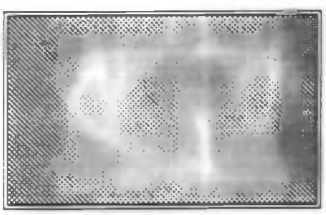
a)  b) 

Figure 5.1: a) A model of a vertebra. b) X-ray of a vertebra.

relations

The spatial and geometrical relations of vertebrae can be estimated, because the spine has to be one smoothing line. The position of a vertebra is dependent on the vertical neighbouring vertebrae and the width of a vertebra is always larger than its height. By using such information a physician is able to make a very precise estimation of the position of a missing vertebra.

CHAPTER 6: APPROACH AND RESULTS

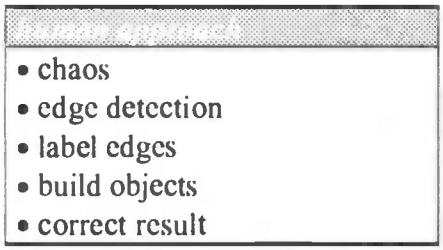
6.1 Introduction

literature

As can be read in literature, see chapter 2 and 3, a lot of difficulties arise while implementing a visual recognition system. Sometimes restrictions are made, such as: the edges of the objects are well-defined [11], the format of image has to be of a particular size [5] or only one object is in the image [4]. Another problem arises when the results are affected by the choice of the parameters, which of course is not suitable for an automatic recognition system. Some programs need 100 000 000 iterations and take 137 minutes computing time [4], it depends on the application whether this is acceptable. The literature did not inspire me to follow their ideas because of the appearance of certain difficulties, such as mentioned before, and in my opinion the morphologically detailed approach mostly used does not work on undetailed raw X-ray images.

human approach

Further we have chosen for a human approach, which means that not only the question what prior knowledge does a physician have is asked, but also the question what do we see in these images. How does our visual system work when we look at these X-ray images. The first feature we notice in an image are edges, not objects, but particular straight edges. After finding them, we label edges by importancy and build the objects from these important edges by using our prior knowledge. In this process we can make mistakes, but due to the prior knowledge we are able to correct these mistakes if the constructed object does not fit to the expected one. Physiologically our visual system is complexly organised, therefore it is not the intention to rebuild this system, but to use the global idea. This global idea can be scheduled as follow:



template

As can be seen in the previous chapter, the shape of a vertebra is rectangular, with some inlets on the sides. We have chosen to neglect the inlets, and therefore the template has a rectangular shape. This simplifies the prototype construction with no direct influence on the resulting quality. Because of the unknown size of the vertebrae, we have chosen to use an active contour fitting algorithm, which shrinks shapes until it fits around the object edges.

6.2 Image preprocessing

chaos

When an image is first captured by the retinal photoreceptors in the eye, it consists of some 120 million individual pointwise measurements of light intensity [29]. A first problem that arises is how to transform this large and unstructured set of measurements into a more useful representation, that is, a representation that is more concise and more convenient for subsequent processing stages.

edges

In the fields of computer vision and image processing, the dominant suggestion has been to begin the analysis of the incoming image by producing an edge representation using a process called edge detection. Edges in the image are locations where the light intensity changes significantly. The main rationale for using an edge representation as the primary representation is that intensity edges are usually physically significant: they correspond to object boundaries and to discontinuities in surface properties, such as spatial orientation or reflectance. The important role of edges is also supported by the fact that for humans a line sketch of an image often conveys most of the essential information, although judging from the underlying intensity distributions, the line sketch and the image are radically different.

derivative

To produce an edge representation of the image, one can identify edges as locations in the image where the (directional) derivative of intensity reaches a local maximum in its absolute value. Locating derivative maxima therefore appears to be a reasonable step toward identifying sharp edges.

noise

The differentiation of a signal, however, raises certain technical difficulties because this operation enhances considerably the high-frequency components of the signal. Because images often contain high-frequency noise, the noise is enhanced by differentiation. To overcome this problem, the differentiation of a signal is usually preceded by a smoothing operation that removes the noisy high-frequency components.

contrast stretching

Images with low light intensity differences, mostly due to bad light circumstances during registration of the images, can be improved by a contrast stretching filter. This filter is represented by:

$$f(t) = 255 \cdot (t - P_{min}) / (P_{max} - P_{min})$$

Where P_{min} is the minimum pixel value and P_{max} is the maximum pixel value of the input image.

Figure 6.1 shows the contrast stretching on an one-dimensional intensity profile, which is a vertical slide of the X-ray image. One should note the vertical axis of these images.

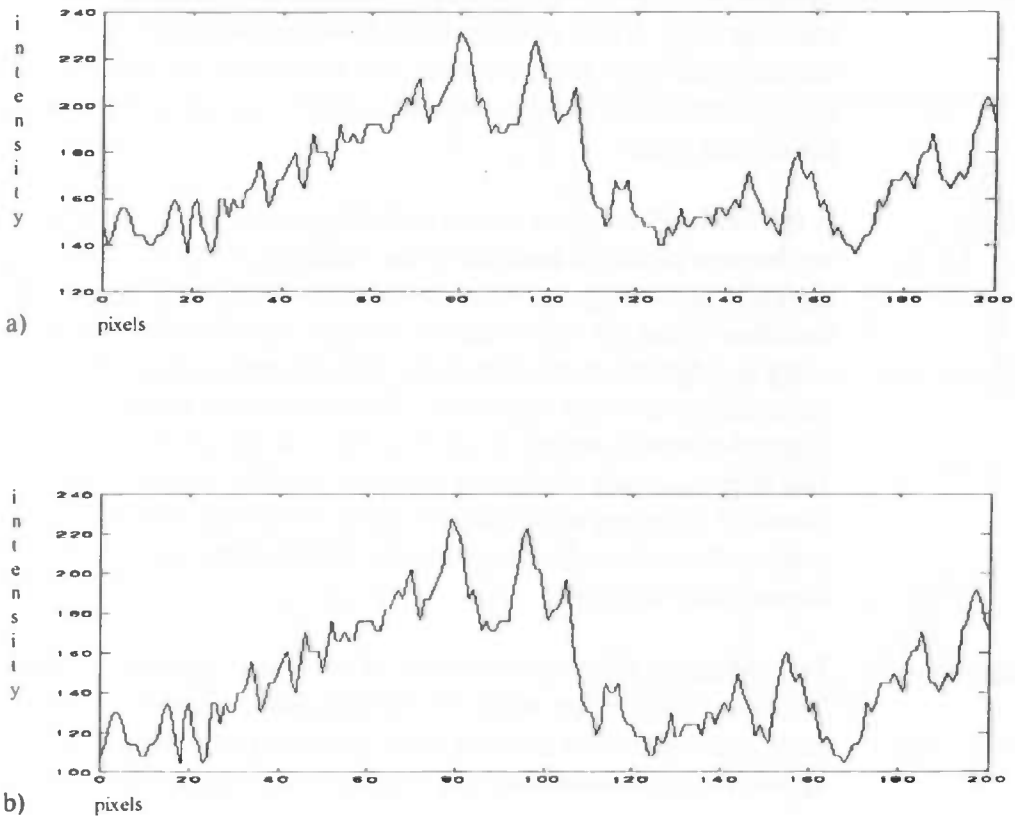


Figure 6.1: a) Original image. b) After contrast stretching.

smoothing

A Hamming window is taken as smoothing operator. This is defined as [32]

$$H(t_1,t_2) = 0.54 + 0.46 \cos 2\pi\sqrt{(t_1^2 + t_2^2)}/T_0 , \qquad \sqrt{(t_1^2 + t_2^2)} \le T_0/2$$
$$= 0, \qquad \sqrt{(t_1^2 + t_2^2)} > T_0/2$$

The convolution of the original image $x_0(t_1,t_2)$ with the Hamming window gives

$$x(t_1,t_2) = x_0(t_1,t_2) * H(t_1,t_2) = \Sigma_{t1',t2'}\, x_0(t_1',t_2')\, H_{T0}(t_1-t_1', t_2-t_2')$$

In figure 6.2 we see the results of the smoothing operator on a one-dimensional intensity profile taken from an X-ray image, where the constant T_0 has value six.

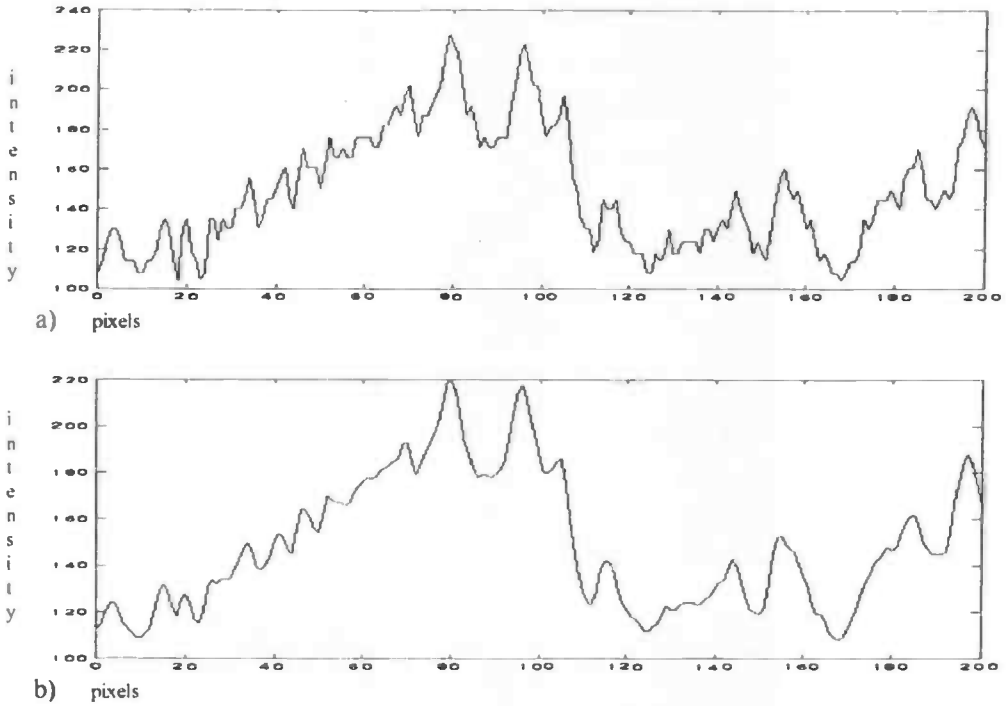


Figure 6.2: a) Input image b) After Hamming window

lateral inhibition

A process which takes place in the natural visual system is lateral inhibition. Lateral inhibition ordinarily serves to heighten the contrast at borders; it can also produce some other curious effects. For example, examine figure 6.3 [30]. Do you see the dark diamonds at the crossroads among the black squares?

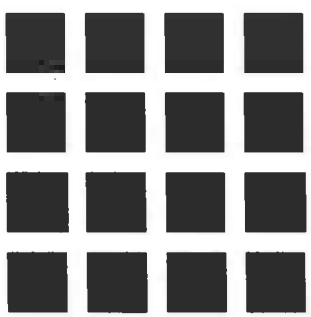


Figure 6.3: An illustration of lateral inhibition.

Lateral inhibition surpresses the surroundings of an edge. In the figure above the white adjacent to the edge of a black square looks "brighter" than it really is. The dark diamonds are the "normal" white light. Due to this we see dark diamonds at the crossroads among the black squares.

Lateral inhibition can be visualised in computer vision by the following filter:

$$[\,-1 \;\; 0 \;\; \overline{|3|} \;\; 0 \;\; -1\,]$$

Figure 6.4 shows the effects of the lateral inhibition filter on an one-dimensional intensity profile, note the vertical axis.

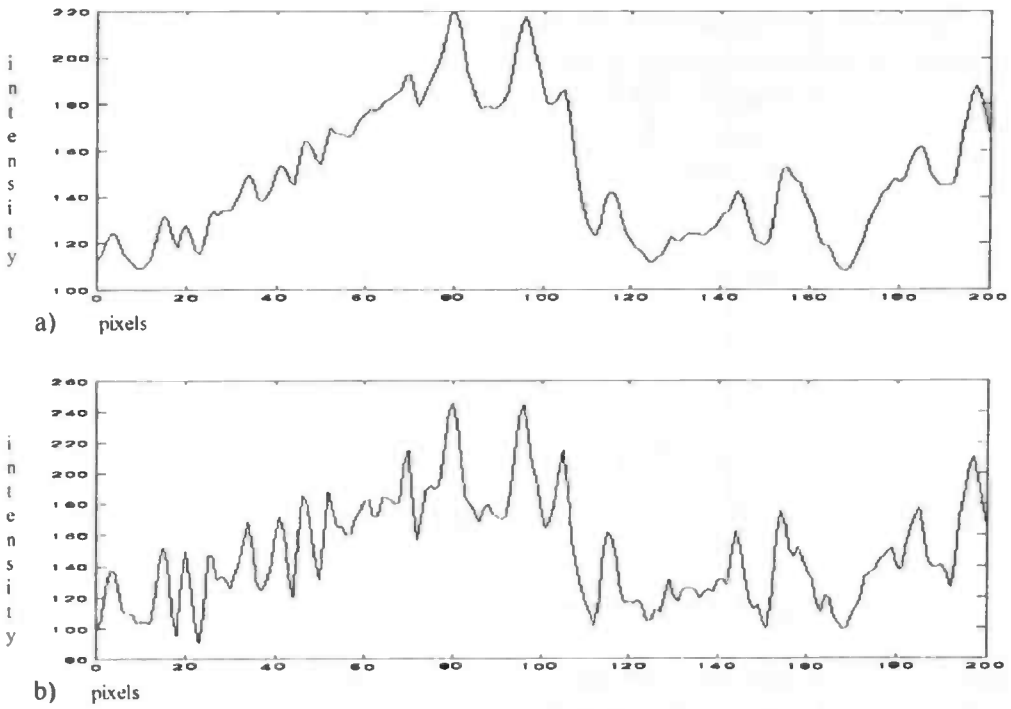


Figure 6.4: a) Input image. b) After lateral inhibition.

summary

To summarise the preprocessing, the original X-ray image is filtered by the following processes.

- Contrast stretching
- Hamming window
- Lateral inhibition

The aim of such processes is to improve the visualisation of the edges of objects in the images, see figure 6.5.

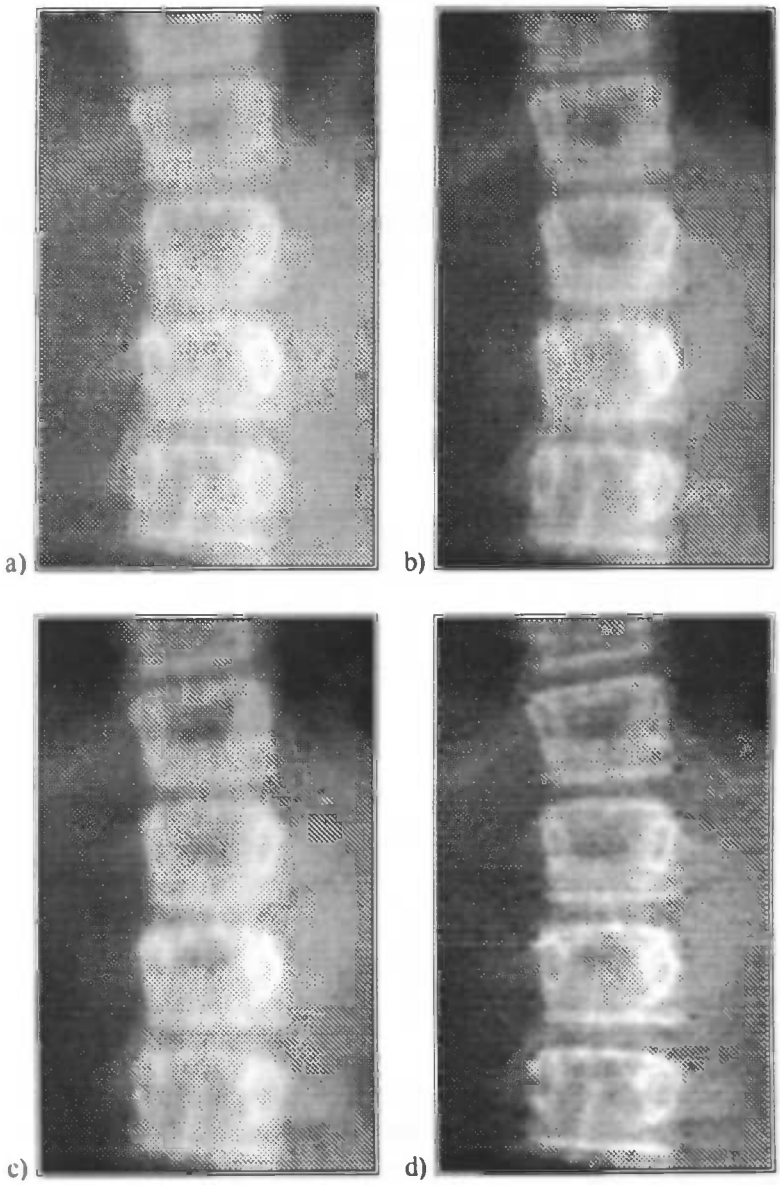


Figure 6.5: a) Original b) Contrast stretching c) Hamming window d) Lateral inhibition

6.3 Edge Detection

sampled

As shortly discussed in the previous section one can identify edges as locations in the image where the (directional) derivative of intensity reaches a local maximum in its absolute value. By locating derivative maxima one can identify edges of objects in an image. Most images are not continuous, but sampled, like X-ray images are. Therefore differentiation in a sampled domain is an approximation of differentiation in continuous domain. For this reason differentiation in a sampled domain is called pseudo-differentiation.

gradient

A gradient is a vector with a size and a direction, which indicates the light intensity change and its corresponding direction. This change can be measured by the first derivative. The gradient of position t in an image x is defined as $\nabla x[t]$, where ∇ is called the nabla operator and can be rewritten as $\nabla x[t] = \delta/\delta x[t]$ [31]

A number of filters can be used for pseudo-differentiation. In the following three filters, which determine the derivatives of an image in vertical direction, will be discussed. Two of them, Prewitt and Sobel, are frequently used.

difference

This filter can be written as:

$$1/2(\Delta t)^3 \begin{bmatrix} -1 \\ [0] \\ 1 \end{bmatrix}$$

Prewitt

The Prewitt filter can be written as:

$$1/6(\Delta t)^3 \begin{bmatrix} -1 & -1 & -1 \\ 0 & [0] & 0 \\ 1 & 1 & 1 \end{bmatrix} = 1/2(\Delta t)^3 \begin{bmatrix} -1 \\ [0] \\ 1 \end{bmatrix} * 1/3(\Delta t)^3 \; [1 \; [1] \; 1]$$

Sobel

The Sobel filter can be written as:

$$1/8(\Delta t)^3 \begin{bmatrix} -1 & -2 & -1 \\ 0 & [0] & 0 \\ 1 & 2 & 1 \end{bmatrix} = 1/2(\Delta t)^3 \begin{bmatrix} -1 \\ [0] \\ 1 \end{bmatrix} * 1/4(\Delta t)^3 \; [1 \; [2] \; 1]$$

comparison

As showed above the Sobel and Prewitt filters are extensions of the difference filter. The differences between the Sobel and Prewitt filter is the degree in which they take the neighbourhood pixels into account. The number between brackets in the filter is the position of the pixel in the image for which the derivative is calculated. The advantage of taking neighbourhood pixels into account is that the filter becomes less sensitive to noise, i.e. an edge of an object does not exist solely of one derivative, but needs also adjacent derivatives. The disadvantage of these filters, Sobel and Prewitt, is that they prefer straight lines, horizontal or vertical; the Prewitt filter even more so. Horizontal or vertical edges get a higher derivative value than oblique edges, even when they have exactly the same sharp edges.

In figure 6.6 the effect of the three filters (difference, Sobel and Prewitt) on an one-dimensional slide of the X-ray image is shown. We can see that the differences between these filters are small

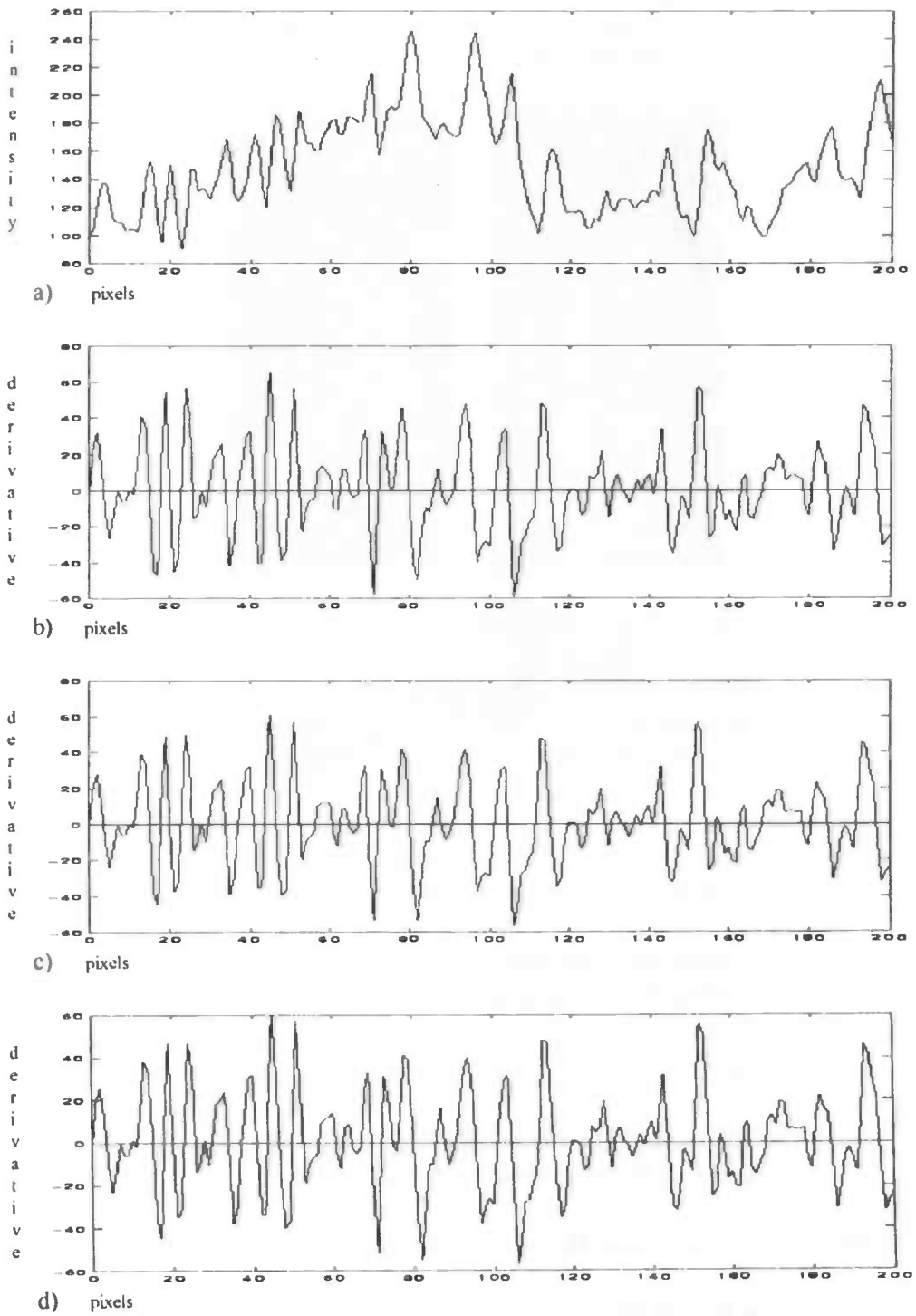


Figure 6.6: a) original image b) Difference filter. c) Sobel filter. d) Prewitt filter.

We have selected the Prewitt filter, because vertebrae in X-ray images are not clear visualised, so it is less dependent of noise. The input X-ray images consist

of pixel values between 0 and 255. After filtering, the X-ray images consist of pixel values between -128 and 128, negative derivatives and positive derivatives. Because a grey pixel value must have a value between 0 and 255, the negative values are mostly changed through an absolute function. We have chosen to keep these negative values, because positive and negative derivatives give information about the type of edges (upper or under) of the object.

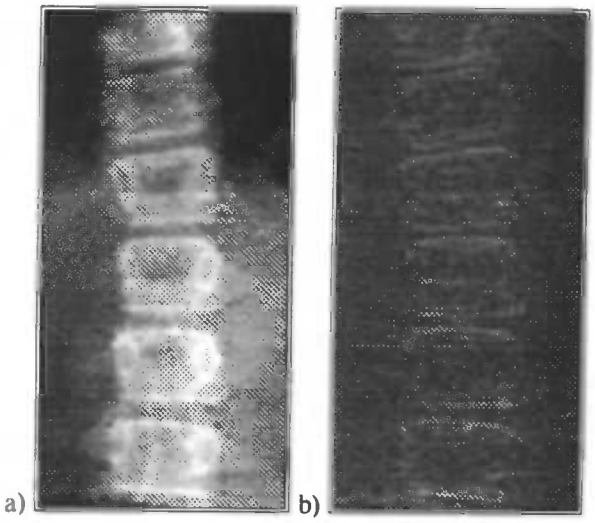


Figure 6.7 : a) original image b) after Prewitt filter

scaling

Because the different X-ray images have different light intensity, it is important to scale the images after an edge filter to get a general value distribution. This scale function first looks for the maximum and minimum value in the input image; if the maximum is greater than the absolute value of the minimum, the maximum value is used for scaling, else the absolute value of the minimum is used for scaling. This is because negative and positive derivatives are considered as dependent parts of an image, one may not split the image into two parts, positive and negative derivatives. After determining this scaling factor the image values are scaled by multiplying with 255 and dividing through the scaling factor. The scaling function S can be formalised as follow:

$$S(x[t_1,t_2]) = (x[t_1,t_2] \cdot 255) / \mathrm{Max}(|P_{min}|, P_{max})$$

where $x[t_1,t_2]$ is the image, Max is a maximum operator, P_{min} the minimum pixel and P_{max} the maximum pixel in the image.

summary

To summarise this section the edge detection is made through

- Prewitt filter
- Scaling function

6.4 Landmarks

uncertainty

After filtering the X-ray images with a Prewitt filter, we get a landscape with derivatives, see figure 6.8. High positive derivatives are part of upper edges of an object and low negative derivatives are part of under edges of an object. Because the **lighting** of the X-ray image is **not equally distributed** in the whole image, moderate derivatives can be part of an edge or can be noise. At this phase it is not possible to distinguish between them. High positive and low negative derivatives can also be part of **over-exposure**, so it is not always part of an edge. These uncertainties make it difficult to identify objects in the X-ray images.

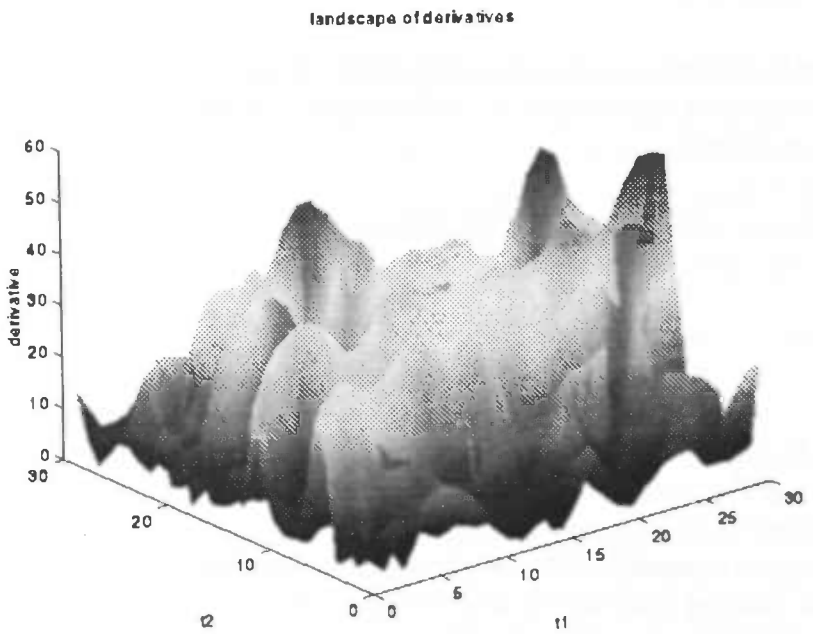


Figure 6.8: A landscape of positive derivatives

landmarks

Because upper edges of vertebrae only consist of positive derivatives, and under edges of vertebrae only consist of negative derivatives and these edges can be seen as straight lines, one can scan the image for edge lines. These edge lines are called landmarks, because they mark the derivative landscape with possibly vertebra edges.

scanning

The X-ray images are scanned for landmarks, these are lines containing information about its

- left co-ordinate,
- right co-ordinate,
- Y_L value pertaining by x = 0,
- α_L, which indicates the direction,
- sign, increasing or decreasing gradient relating to previous line,
- activity, left and right co-ordinates both must be inside the image,
- width, the number of pixels lying on the line,
- gradient ∇L.

gradient

The gradient of a line is the mean value of the derivatives on this line from the left to the right co-ordinate. This can be formulated as:

$$\nabla L = \Sigma_{t1} \nabla x[t_1, Rnd(Y_L - \tan(\alpha_L \cdot \pi/180) \cdot t_1)]$$

The gradient is an indication of the strength in which the landmark matches an edge. It labels the landmark by importancy. Landmarks with high positive gradient are likely to be an upper edge of a vertebrae and landmarks with low negative gradient are likely to be an under edge of a vertebrae. The strength of a fuzzy region F on the unit interval [0,1] can be expressed as

$$S(F) = |\nabla L| / 255$$

where F is defined over the referential set $\Omega = [-255, 255]$. To distinguish between possible upper edges and under edges, one can formulate $S_{up}(F)$ and $S_{be}(F)$ as follows

$$S_{up}(F) = \begin{cases} \nabla L / 255 & \text{if } \nabla L > 0 \\ 0 & \text{if } \nabla L \leq 0 \end{cases}$$

$$S_{be}(F) = \begin{cases} |\nabla L| / 255 & \text{if } \nabla L < 0 \\ 0 & \text{if } \nabla L \geq 0 \end{cases}$$

where $S_{up}(F)$ and $S_{be}(F)$ denotes strength of the possible upper edges and the strength of the possible under edges respectively.

algorithm

The image is scanned in different directions for local gradient maxima or minima. Starting from the top of an image, a line with a fixed direction moves downward, if the sign of this line in a certain position is positive, which means the gradient is increasing relating to previous line, and the sign of the next line is negative, which means the gradient is decreasing relating to its previous line, and its own gradient is positive, than this line is marked as a positive landmark. Inversely, if the sign of the line is negative and the sign of the next line is positive and its gradient is negative, then the line is marked as a negative landmark.

This process is illustrated by figure 6.9. The successive points are the gradients of the successive lines with a fixed direction; the square points are marked as landmarks.

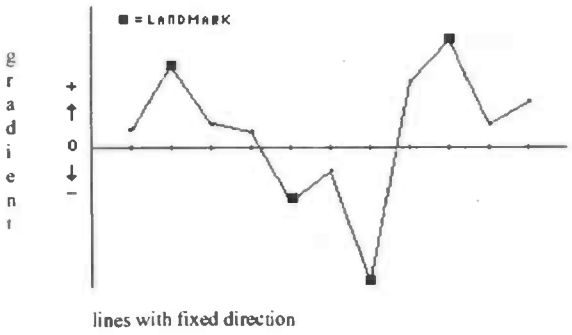


Figure 6.9: Illustration of the landmark process.

Because of many local gradient maxima, there appear a lot of landmarks with different directions. There are certain criteria according to which landmarks can be deleted, because landmarks who do not satisfy on these criteria are very unlikely to be an edge of a vertebra. One can reduce the number of landmarks by:

- looking at their distribution,
- deleting landmarks with minimal gradient,
- looking at neighbour landmarks.

Care should be taken, because landmarks, the possible vertebra edges, can have small gradients, due to unequal distributed lightning.

distribution

The distribution of the derivatives on a landmark, which matches a vertebra edge is shown in figure 6.10. The horizontal line in the figure is the gradient ∇L of the landmark.

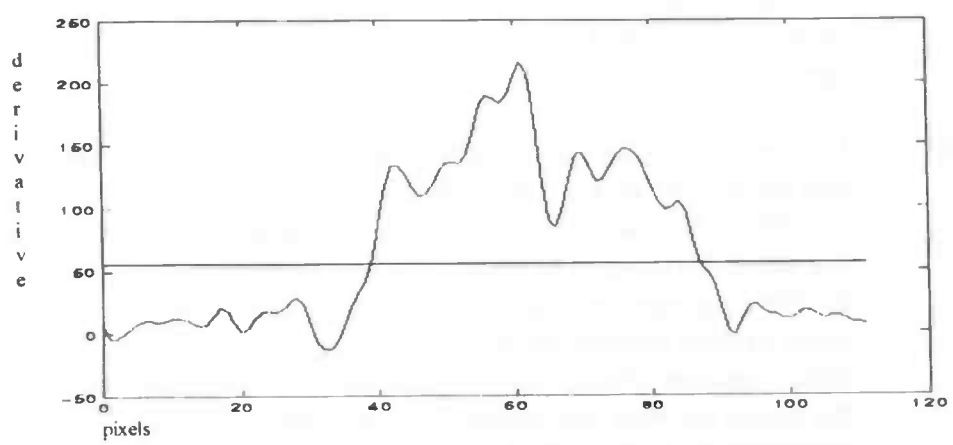
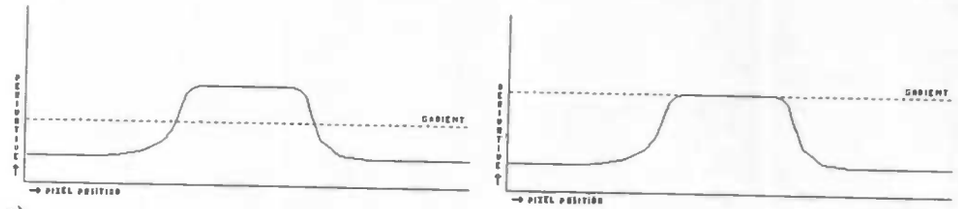


Figure 6.10: The distribution of a landmark, that matches a vertebra edge.

The maximum derivatives on a landmark should lie close together in the midst of the landmark, which indicates the edge of the vertebra, and the minimum derivatives on a landmark should lie on the endings of the landmark, which indicates the background of the vertebra.

The gradient ∇L of a landmark, which denotes the average derivative value, can be maximised by shortening the length of the landmark. Under ideal circumstances the maximum gradient belong to the landmark with the length which is about the same as the width of the vertebra, see figure 6.11.



a) b)

Figure 6.11: a) The ideal distribution b) The ideal maximum gradient.

Because we do not know the width of an edge found by a landmark, it is not possible to verify the results. Perhaps due to lack of information in the X-ray image the width of the edge is half the width as it should be. One thing we know for sure is that the width of an edge is never as large as the width of the X-ray image. Because otherwise the next or previous vertebrae will go out of the image, for scoliosis is a curvature of the spine. By maximising the gradient of the landmark, i.e. shrinking the width of the edges, it is possible to delete some irrelevant landmarks. Note, the gradient of a line is the mean value of the derivatives on this line from the left to the right co-ordinate. Deleting of irrelevant landmarks is done in the following process, first maximise temporarily the gradients of all landmarks and calculate its width, secondly calculate the average width of these landmarks and the standard deviation and last delete all landmarks which have a width larger than the average width plus two times the standard deviation. Because of this process, it is no problem to have images with a latticework, that normally should disturb the results with their high gradients.

minimal energy

Landmarks with very small local gradient maxima can also be deleted. Of course one needs to be very careful with this, because landmarks with a certain low gradient in the bottom of the image are irrelevant, while they are very useful in the top of the image (see figure 4.1). But a certain threshold α can be assumed to delete the very small gradients. The threshold is calculated by a scalar, mostly three, divided through the length of the landmark. In other words, if less than a certain amount of pixels on the landmark have the maximum derivative and all the others are zero, the landmark can be deleted. The calculation of the threshold can be formalised as

$$\alpha = a / L, \qquad \alpha \in [0,1]$$

where a is the scalar and L denotes the length of the landmark. Now the strength of a landmark is represented in terms of a α-cut level set, F^{α}, such that: $F^{\alpha} = \{x | \mu_F(x) \geq \alpha\}$. The strength of a fuzzy region F is as follows

$$S(F) = \begin{cases} S(F^{\alpha}) \\ 0, \quad \text{otherwise} \end{cases}$$

neighbours To reduce the number of landmarks one can also look at the neighbours of a landmark, these neighbours are the landmarks which vary little in position and in direction in comparison to the original landmark. If two landmarks only vary in a certain degree of direction $\delta\theta$ or in a certain position δy, the landmark with the minimum gradient can be deleted. This can be formalised as follows, where θ denotes the direction and y the position of the original landmark.

$$\text{Max}(L_{\theta-\delta\theta,y-\delta y}, ..,L_{\theta,y}, .., L_{\theta+\delta\theta,y+\delta y}), \qquad \text{for a certain } \theta \text{ and } y$$

results After deleting irrelevant landmarks by its distribution, minimal gradient and neighbours, landmarks are left which are possible vertebra edges. In figure 6.12 two original images and their landmarks are shown. To illustrate the landmarks, the minimal_energy process is used with a higher scalar than usual to emphasise the position of the higher landmark gradients .

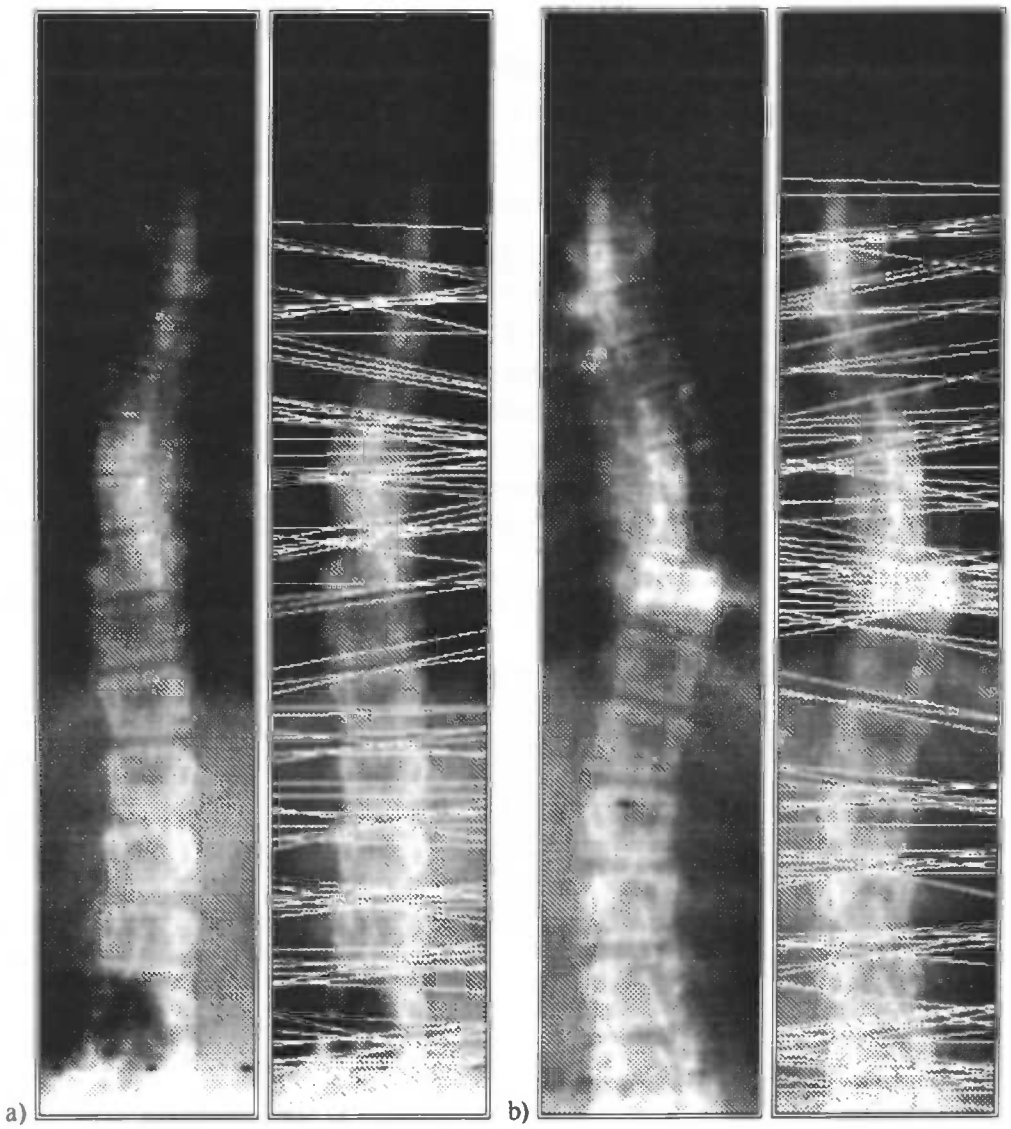


Figure 6.12: Two examples of an original image and its landmarks.

summary To summarise this section the following steps are taken.

- Landmarks
- Distribution
- Minimal energy
- Delete neighbours

6.5 Pattern finding

In the previous section we have determined the landmarks that are the possible vertebra edges. Now it is time to start reasoning with these landmarks by using fuzzy spatial relations. First we are looking for a global pattern in these landmarks. If we have found this pattern or part of this pattern, it will be corrected by using fuzzy relations and properties such as angle, distance and strength.

global pattern

To find a pattern in the landmarks, that are the possible upper and under edges of the vertebrae, we make use of the spatial relations between these vertebrae. Because the X-ray image visualises the lower lumbar vertebrae better than the higher thoracic vertebrae, which indicates more certainty, we are starting the process of finding a pattern at the bottom of the X-ray image. This pattern is built based on the spatial relation of the previous already found vertebrae. The intention is not to find the location of the vertebrae, but to define some information of the global pattern. One needs information to reason. After finding a global pattern in the landmarks one can calculate the height of a vertebra and the height between vertebrae, the interspace. This height need not to be the exact height, but some indication of the vertebrae positions.

algorithm

The prior knowledge used in the algorithm is based on the height of the vertebrae. From sacral to cervical the vertebrae decrease in height [33]. It does not make sense to use exact data of these heights, because of the lack of information in the X-ray images. When the height of the first bottom vertebra in the X-ray image is denoted as h_1, the relation of the vertebrae can be formulated as

$$h_{i-1} \leq h_i \leq h_{i+1}$$

where i indicates the i-th vertebra, the number of the vertebrae is unknown. The following reasoning is used in the algorithm, if h_i denotes the height of a vertebra, the vertebra above should be smaller in height and the under edge of this vertebra should be found in the next area with height $h_i/2$. When the under edge of this vertebra is found, the upper edge should be found in the next area with height h_i, to prevent pedicles who are in the midst of a vertebra the upper edge have to be found in the upper part of this area. The process will search for a landmark with the strongest gradient in these areas, i.e. landmarks who are most likely to be an edge.

❶. If less than three vertebra found, define the height of the first vertebra, denoted as h_1, else **stop**.

❷. Define y_{bc} , initial equal to the height of the image.

❸. Calculate $y_{up} = y_{bc} - h_i$, if y_{up} outside image or h_i > Max **stop**.

❹. Search for the **maximum $S_{bc}(F)$** in an area between y_{bc} and $\frac{1}{2}(y_{bc} - h_i)$, except for the first negative landmark this area is between y_{bc} and $(y_{bc} - 2h_i)$. If not found, $y_{bc} = y_{up} + 5$. If not found and previous also not found, go to ①. If found, $y_{bc} = y_L$ the y value of the landmark beneath and $y_{up} = (y_{bc} - h_i)$. If found and if found previous landmark calculate and store the space between the two landmarks.

❺. Search for the **maximum $S_{up}(F)$** in an area between $\frac{5}{8}(y_{bc}-h_i)$ and y_{up}. If not found, $y_{bc} = y_{up}$. If not found and previous also not found, go to ①. If found calculate and store real h_i, $i = i+1$, $y_{bc} = y_L$ the y value of the upper landmark. Go to ③.

The height of the first vertebra is unknown, so it need to be estimated by a value h_1. This value can vary in a certain range, from small to large. If this value is too small or too large less vertical edges will be found in the specified areas. When successively two landmarks are not found, the process is stopped or the height of the first vertebra will be enlarged, depending on the already found pattern. The values y_{bc} and y_{up} defines the area in which a vertebra is situated, h_i is the height of the previous vertebra, except in the case of the first vertebra. The scalar 5/8 helps to avoid pedicles.

results Figure 6.13 shows the results of the global_pattern process.

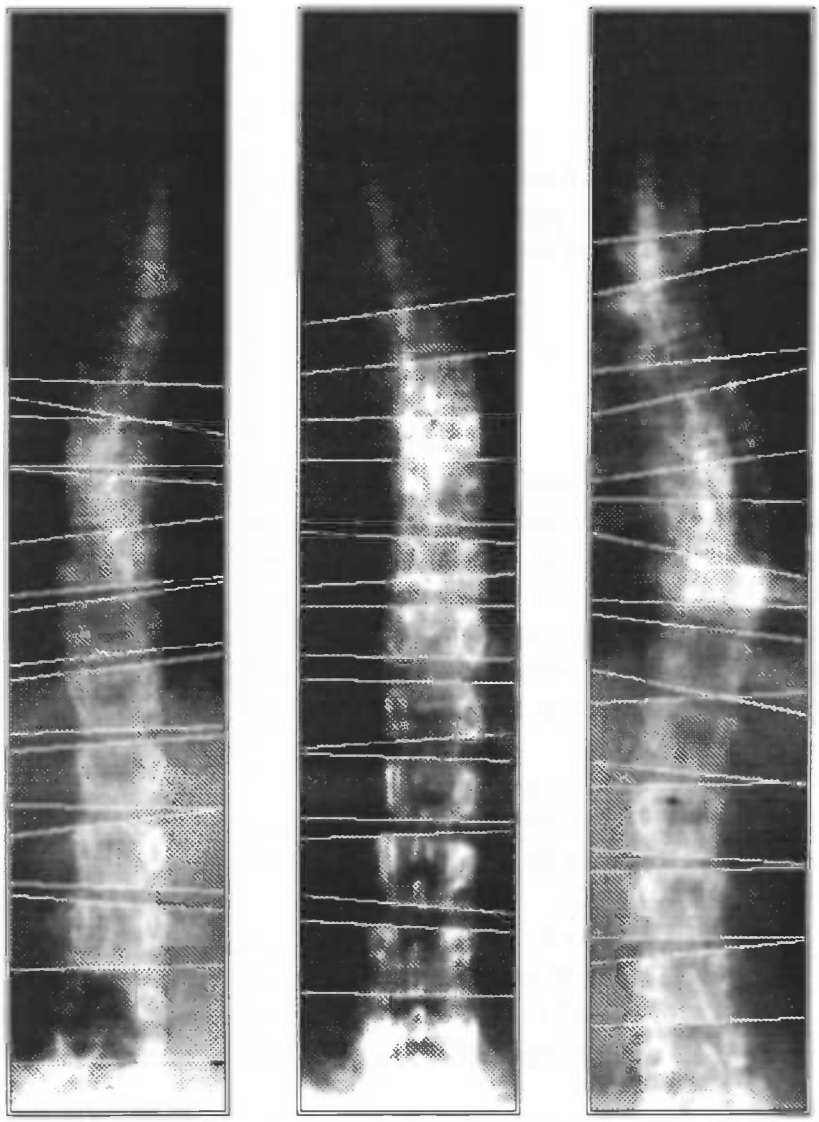


Figure 6.13: A pattern found in the landmarks by using global_pattern.

correction Now we have some useful information about the vertebrae and therefore we can start correcting by using fuzzy sets. The membership function for "angle between two edges" may be defined as a function of θ as

$$\mu_{ANGLE}(\theta) = \begin{cases} 1 & \theta = 0 \\ 2 - e^{(\theta/8)} & 0 < \theta \leq 5 \\ 0 & \theta > 5 \end{cases}$$

This membership express the relation of a given under edge with an already known angle and its most likely neighbour edge. If the under edge has a weak strength S_{be} compared to the upper edge, more deviation in angle should be permitted, because this indicates more uncertainty. Therefore another

membership function is used, which expresses the relation of strength of the upper and under edges. The membership value for "S_{be} related to S_{up}" is computed using

$$\mu_{Sbe\_REL\_Sup} = \begin{cases} 1, & S_{be} \geq S_{up} \\ S_{be}/S_{up}, & S_{be} < S_{up} \end{cases}$$

Another membership function is used which express the relation of the expected position, based on the pattern, and the position of a vertebra. The membership value for "deviation of expected position" may be computed using

$$\mu_{POSITION}(t) = \begin{cases} 1, & t = 0 \\ 2\text{-}e^{(t/10)} & 0 < t \leq 8 \\ 0, & t > 8 \end{cases}$$

algorithm

Again the process is started at the bottom of the X-ray image. Two arrays acquired from the previous process are used, namely h_i and I_i, that denotes respectively the height of vertebra i and the inter space of vertebra i and vertebra (i +1). Initial the first landmark is used calculated from the previous process. On purpose no account is taken for deviation in direction between the edges of adjacent vertebrae, because this would cause a preference to straight spines, which is not most likely for scoliosis patients.

❶. Store $S_{be}(F^{\alpha})$ and α_b and Y_b.
❷. Calculate h. If first vertebra, $h = \frac{1}{2}(h_1 + h_2)$, else if ($h_i = 0$) or ($h_i > (h_{i-1} + h_{i-2})$) or ($h_i < \frac{1}{2}(h_{i-1} + h_{i-2})$), $h = \frac{1}{2}(h_{i-1} + h_{i-2})$, else $h = \frac{1}{2}(h_i + h_{i-1})$.
❸. Calculate $Y_u = Y_b - h$. If Y_u outside image, **stop**.
❹. Search for the upper edge with the strongest $S_{up}(F)$ in an area ($Yu \pm C_u$). If ($\mu_{POSITION}(t)$ is high) or ($\mu_{Sbe\_REL\_Sup}$ is low and $\mu_{ANGLE}(\theta)$ is average) or ($\mu_{ANGLE}(\theta)$ is low), repeat searching for less strongest $S_{up}(F)$, else edge found.
❺. If no edge found, make one with $\alpha_u = \alpha_b$, $Y_u = (Y_b - h)$. Calculate new h_i and store α_u and Y_u.
❻. Calculate I inter space. If first inter space, $I = \frac{1}{2}(I_i + I_{i+1})$, else if ($I_i = 0$) or ($I_i > (I_{i-1} + I_{i-2})$) or ($I_i < \frac{1}{2}(I_{i-1} + I_{i-2})$), $I = \frac{1}{2}(I_{i-1} + I_{i-2})$, else $I = \frac{1}{2}(I_i + I_{i-1})$.
❼. Calculate $Y_b = Y_u - I$. If Y_b outside image, **stop**
❽. Search for the under edge with the strongest $S_{be}(F)$ in an area ($Y_b \pm C_b$). If ($\mu_{POSITION}(t)$ is high), repeat searching for less strongest $S_{be}(F)$, else edge found.
❾. If no landmark found. make one with $\alpha_b = \alpha_u$, $Y_b = (Y_u - I)$. Calculate new I_i and go to ①.

results

In figure 6.14 the results are shown of the correction process.

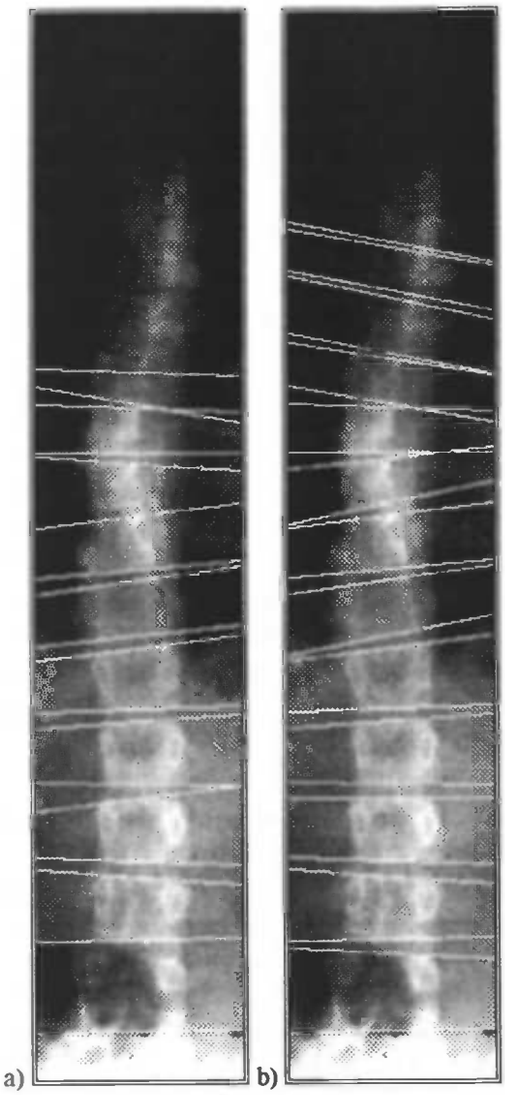


Figure 6.14: Finding a pattern by using fuzzy sets. a) before. b) after.

Now it is possible to isolate a vertebra and to use contour shrinking to get its contour, which will be explained in next section.

summary

In this section the following processes are used.

- Global_pattern
- Correction process

6.6 Active contour model

template

The template has a rectangular shape. This simplifies the prototype construction with no direct influence on the resulting quality. Because physicians extract information such as the corner points of the vertebrae, this prototype makes it easy to determine such points. In figure 6.15 an example is shown of the template.

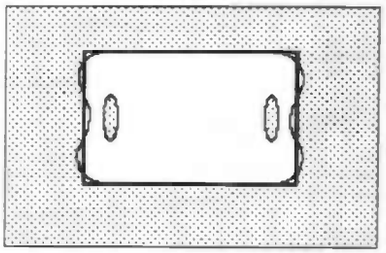


Figure 6.15: Example of the template.

contour

After the preceding sequentially used processes, like Landmarks, Pattern finding, vertebrae are isolated. This is very important, because active contour algorithms need to be initialised with a contour, very close to the edges of the objects [24]. In a visual recognition system this initialisation needs to be automated and not made by hand.

Because the template has a rectangular shape, it can easily defined by only four corner points. The points on the snake need not to be evenly distributed nor may the points on the snake develop sharp corners, like most snake algorithm assume [4][12][24]. Besides due to th it is less sensitive to noise, because moving does not depend on the local information of one pixel.

algorithm

The energy function of the algorithm is simple. It calculates the mean intensity of the area inside the template. This energy need to be maximised. Because we have already found the horizontally edges of a vertebra, only the vertical edges have to be located. Initially these vertical edges are at the borders of the image. Stepwise shrinking of the contour takes place until the energy is maximised. First shrinking of the contour is performed with vertical edges. When the energy does not increase anymore, the edges are rotated to find a better fit. One can require that the rotation of the edges should be the same. Because of the horizontal distributions of a vertebra and being less dependent on local information, this process will pretty well find the contour of the vertebra (if visible).

results

In figure 6.16 the results are shown of the active contour model. In this process prior knowledge is used with regards to the shape of the vertebra. Due to lack of information the width of a vertebra can be too small or too large, see first vertebra of the first spine in figure 6.16. We know that the width of a vertebra

should always be larger than its height and that the spine should be a smoothing line. This information is used in the next section elastic_constrains.

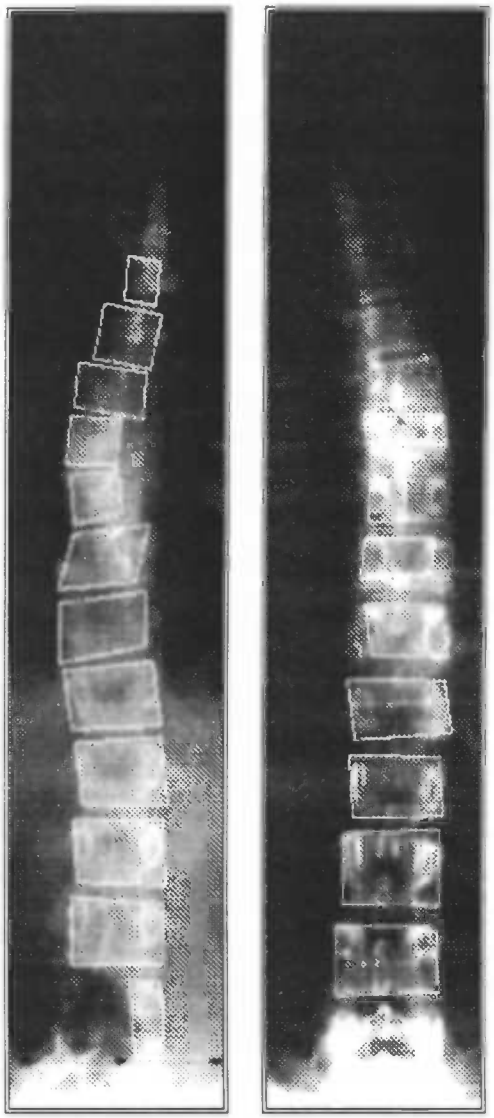


Figure 6.16 : Results of the active contour model.

6.7 Elastic Constraints

In this process no information is used from the images, the only aim is to correct the geometrical relations and smoothing the curvature. The global idea of this process is to put elastics between the corners of a vertebra and the corners of its next vertebra. By shortening these elastics the spine will be smoothed. The geometrical relation of the width/height of a vertebra can be formalised as

$$G(F) = \begin{cases} 0, & \text{if } w < h \\ 1, & \text{if } w \geq h \end{cases}$$

algorithm

❶ Calculate the **height** and the **width** of the vertebra.

❷ Calculate closest point lying on the edge of the vertebra related to the **left** corner of the edge **above** and calculate the distance d1 between the original point and the closest point. Calculate closest point lying on the edge of the vertebra related to the **left** corner of the edge **beneath** and calculate the distance d2 between the original point and the closest point.

❸ If G(F) = 0, and if vertebra is an upper edge, move distance d1 on the edge, else move d2 on the edge. Else d = (d1 + d2)/2, if (d > d2) and upper edge, move distance d2 on the edge, else if (d >d1) and under edge, move distance d1 on the edge, else move distance d.

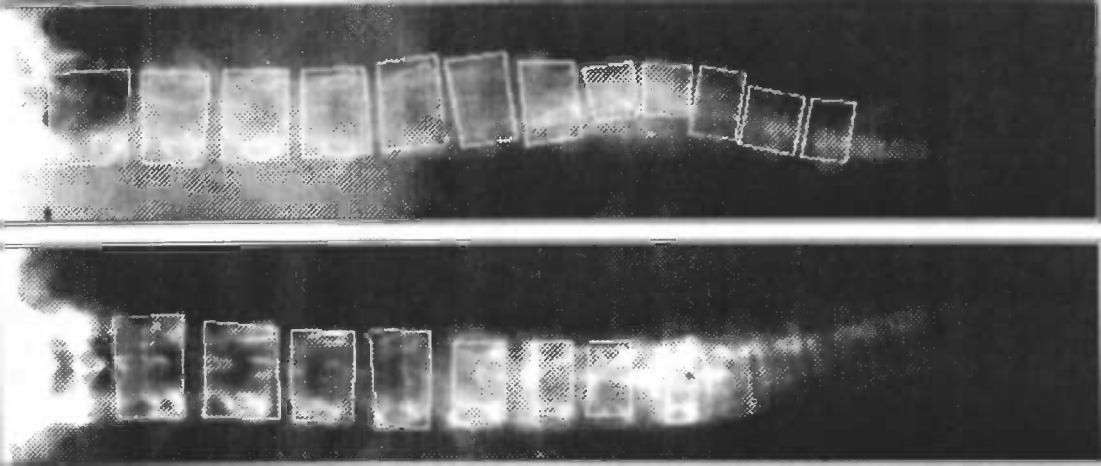
❹ Calculate closest point lying on the edge of the vertebra related to the **right** corner of the edge **above** and calculate the distance d1 between the original point and the closest point. Calculate closest point lying on the edge of the vertebra related to the **right** corner of the edge **beneath** and calculate the distance d2 between the original point and the closest point.

❺ Calculate width.

❻ If G(F) = 0, and if vertebra is an upper edge, move distance d1 on the edge, else move d2 on the edge. Else d = (d1 + d2)/2, if (d > d2) and upper edge, move distance d2 on the edge, else if (d >d1) and under edge, move distance d1 on the edge, else move distance d.

❼ Repeat until last vertebra.

The under edge of the first vertebra will not be corrected, because there is no vertebra beneath it, it will only be corrected if the width of this vertebra does not correspond to the geometrical relation of a vertebra. The same situation appears for the last vertebra, where the upper edge will not be corrected, because there is no vertebra above, unless the width of the vertebra does not correspond to the geometrical relation of a vertebra.



CHAPTER 7: DISCUSSION

7.1 Discussion

human approach

We have chosen for a human approach, which means that not only the question what prior knowledge does a physician have, is asked, but also the question what do we see in these images. We schedule the global idea of our visual system. Now we can compare the approach of the visual recognition system to the human approach, see table 7.1. Just like the first phase of the human approach model, the initial state of the visual recognition system is chaos, i.e. thousand unstructured pixels are offered to the system. The second phase, edge detection, can be compared to the Prewitt and Lateral inhibition filtering in the visual recognition system. The third phase, label edges by importancy, is the landmark and pattern finding process in the visual recognition system. Objects were built in the active contour model and the results were corrected by the elastic constraints process.

human approach	visual recognition system
• **chaos**	• **chaos**
• **edge detection**	• **Prewitt et al.**
• **label edges**	• **landmarks, pattern finding**
• **build objects**	• **active contour model**
• **correct results**	• **elastic constraints**

Table 7.1

The used human approach is a simple model of our visual system. The fysiology of the human visual system is in reality much more complexly organised. One can improve the used model to a more realistic human process, by after correcting the results, again start labeling the edges by the new importancy, building objects of them and correcting the result, repeating this process until the results fits to the expected results, specified by prior knowledge.

prior knowledge

On the question what prior knowledge does a physician have, we formulate fuzzy spatial and geometrical relations to define this knowledge. Fuzzy knowledge was added to the visual recognition system. In table 7.2 an overview is given where and what kind of information is added. In the preprocessing and in the Prewitt filtering et al. only information of the image is used. In the landmark process mainly information of the image itself is used. But also we used knowledge concerning the edges of vertebrae. Pattern finding used landmarks, and extra information of the spatial relation of the spine. Prior knowledge of the shape of the vertebra is used in the active contour model. In elastic constraints, information of the geometrical relations of a vertebra is added, and prior knowledge of curvature of the spine is used. No information of the image is used in this latter process.

Prewitt et al.	+	-
landmarks	+	shape edges
pattern finding	+	spatial relations spine
active contour model	+	shape vertebra
elastic constraints	--	geometrical relations, shape spine

Table 7.2

results

The results of the visual recognition system are shown in the previous chapter. Because of the limited time to spend, our aim was not to make a robust visual recognition system, but to work out the global idea and approach. The system is tested on some X-ray images, with the result that in all these image vertebrae are located. The number of vertebrae that are located in an image, is dependent on the lighting in the upper part of the X-ray. The system does not work on scoliosis of a very serious degree, because the pattern finding process expects that the vertebrae are situated above each other, and not next to each other. Fortunately, these serious cases do not appear frequently anymore. If a serious case appears it is usually associated with another more significant disease.

system

The visual recognition can be improved on certain points.The active contour model is formalised with a very simple energy function, this can be refined with more image parameters to get a more precise matching of the template.

The strong part of the system is the isolation of the vertebrae with the landmark process. A critical point in active contour models is the initialisation. In most cases the restriction of one object in the image is made. Because the spine exist of many vertebrae, i.e. objects, the isolation of the different vertebrae makes it possible to use an active contour model.

The preprocessing of the visual recognition system is dependent on the image, due to lightning of the X-ray image that is not equally distributed in the whole image and due to over-exposure. Contrast stretching improves the contrast in the X-ray image, but not the distribution of the intensity and over-exposure prevents stretching of the higher intensity in the X-ray image, while there is a real gap between the intensity of the over-exposure and the intensity of the vertebrae visualised. With the new X-ray scanningsmethods the quality of X-ray images will be improved, i.e. no over-exposure and a better visualisation of the vertebrae. Of course, there always will be a connection between the quality of the images used in a visual recognition system and the results of it.

7.2 Further investigation

Because the aim of this research is to work out the global idea and approach of adding prior knowledge to a visual recognition system, there are a lot issues in the system that can be improved to be a more robust system. Even it should be possible to make the preprocessing independent of the images, by first filtering the over-expose out of the image and secondly by looking at the distribution of the intensity in the images, i.e. calculation mean value, calculation the standard deviation and using this information to stretch and scale the image to a standard distribution. As mentioned before, one should refine the active contour model by using more information of the image, by not only using the intensity inside the contour but also the derivatives of the edges. One can improve the system also by repeating the last three step iteratively of the human approach. In the last step of the approach, the width of the vertebrae is determined with the use of the information of the heights. Going back to the phase of labelling edges to importancy, one can also use the information of the width of the vertebrae to determine the height of the vertebrae. One does always need one of both first to reason about the shape of a vertebra. It is also possible to extract more information of the images, which can be used to make a more robust system. The possible improvements of the system is summarised in the following points

- Make use of the pedicles in the vertebrae in the process pattern_finding.
- Take a closer look at the distribution of the landmarks.
- Repeat iteratively the last three steps of the approach until the results fits to the expected results.
- Eliminate over-exposure
- Scale and stretch the distribution of the intensity to a standard distribution
- Refine the active contour model.

By improving the visual recognition system is it possible to make a real robust system. One can extract and use more information in the visual recognition system than it does use now. The weakest issues in the system, as mentioned above, can all be eliminated.

LITERATURE

[1] **R. M. Haralick and L. G. Shapiro,**" *Computer and robot vision*", Addison-Wesley Publishing Company, Volume 2, 1993.

[2] **D. H. Ballard and C.M. Brown,**" *Computer vision* ", Prentice-Hall, 1982.

[3] **K. Jain, Y. Zhong and S. Lakshmanan,** " *Object matching using deformable templates*", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 18, no. 3, March 1996, pp 267-277.

[4] **G. Storvik,** "*A Bayesian approach to dynamic contours through stochastic sampling and simulated annealing*", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 16, no. 10, October 1994, pp 976-986.

[5] **U. Grenander and D. Keenan,** "*Toward automated image understanding*", Journal of Applied Statistic, 16, pp 207-221.

[6] **S. Geman and D. Geman,** "*Stochastic relaxation. Gibbs distributions. and the Bayesian restoration of images*", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. Pami-6, no. 6, November 1984, pp 721-741.

[7] **D.J. Burr,** " *Elastic matching of line drawings*", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. Pami-3, no. 6, November 1981, pp 708-713.

[8] **Y. Amit, U. Grenander and M. Piccioni,** " *Structural image restoration through deformable templates*", Journal of the American Statistical Association, Vol. 86, No. 414, June 1991, pp 376-387.

[9] **L.D. Cohen and I. Cohen,** " *Finite-element methods for active contour models and balloons for 2-D and 3-D images*", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 15, no. 11, November 1993, pp 1131-1147.

[10] **S. Lakshmanan and D. Grimmer,** "*A deformable template approach to detecting straight edges in radar images*", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 18, no. 4, April 1996, pp 439-443.

[11] **M.D. Jolly, S. Lakshmanan and A.K. Jain,** " *Vehicle segmentation and classification using deformable templates*", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 18, no. 3, March 1996, pp 293-308.

[12] **D. Rueckert and P. Burger,** " *A multiscale approach to contour fitting for MR images*", 1995/1996.

[13] **T.L. Huntsberger, C. Rangarajan and S.N. Jayaramamurthy,** " *Representation of uncertainty in computer vision using fuzzy sets*", IEEE Trans. Comput., Vol. C-35, no. 2, February 1986, pp 145-156.

[14] **R. Krishnapuram and J.M. Keller,** " *Fuzzy set theoretic approach to computer vision: an overview*", 1992, pp 135-142.

[15] **A. Pathak and S.K. Pal,** " *Fuzzy grammars in syntactic recognition of skeletal maturity from X-rays*", IEEE Trans. Syst., Vol. SMC-16, no. 5, September/October 1986, pp 657-667.

[16] **S.K. Pal,** " *Fuzzy sets in image processing and recognition*", 1992, pp 119-126.

[17] **A. Rosenfeld**, " *The fuzzy geometry of image subsets*", Pattern Recognition Letters, Vol. 2, September 1984, pp 311-317.
[18] **R. Krishnapuram, J.M. Keller and Y. Ma**, " *Quantitative analysis of properties and spatial relations of fuzzy image regions*", IEEE Transactions on Fuzzy Systems, Vol. 1, no. 3, August 1993, pp 222-233.
[19] **R. Dousat, W. Konen, M. Lades, C. v.d. Malsburg, J. Vorbruegggen, L. Wiskott and R. Wuertz**, " *Neural mechanisms of elastic pattern matching*", Proceedings of a BMFT---Workshop held at Maurach, Germany, 20/21 October, 1992.
[20] **A. Rosenfeld**, "*Geometric properties of sets of lines*", July 1994.
[21] **A. Rosenfeld**, " *Fuzzy plane geometry: triangles*", November 1993.
[22] **P. Radeva and J. Serrat**, " *Rubber snake: implementation on signed distance potential*", Vision Conference SWISS'93, September 1993, pp 187-194.
[23] **L. Wiskott, J. Fellous, N. Krueger and C. von der Malsburg**, " *Face recognition by elastic bunch graph matching*", Internal Report, Bochum, 1996.
[24] **J.B.T.M. Roerdink**, "*Computer Vision*", RUG, September 1996.
[25] **T.J. Ross**, "*Fuzzy Logic with Engineering Applications*", McGraw-Hill Inc., 1995.
[26] **J.I.P. James**, "*Scoliosis*", Butler & Tanner Ltd. , New York, 1976.
[27] **S. Kleinberg**, "*Scoliosis; pathology. etiology, and treatment*", Waverly Press Inc., Baltimore, 1951.
[28] **R. Roaf**, "*Scoliosis*", Neill & Co. Ltd, Edinburgh, 1966.
[29] **M.I. Posner**, "*Foundations of cognitive science*", The MIT Press, Cambridge, 1993.
[30] **J.W. Kalat**, "*Biological psychology*", Brooks/Cole Publishing Company, 1995.
[31] **R. Moddemeijer**, "*Beeldbewerking*", RUG, September 1996.
[32] **R.L. Fante**, "*Signal analysis and estimation*", Massachusetts, September 1987.
[33] **A.E. Linton, M.E. Levy, B.F. DiGiovanni, B.M. Latimer and P.V. Scoles**,"*Vertebral morphology in the human spine*", Departments of Physical Anthropology and Orthopaedics, Hamann-Todd Osteological Collection.