Comments and questions can be sent to: michael@cs.uu.nl


ELSEVIER

Pattern Recognition Letters 20 (1999) 521–533

Pattern Recognition Letters

www.elsevier.nl/locate/patrec

Recognition of radiopaque markers in X-ray images using a neural network as nonlinear filter

M. Egmont-Petersen *,[1], T. Arts

Department of Biophysics, Maastricht University, The Netherlands

Received 13 March 1998; received in revised form 24 March 1999

Abstract

Neural networks are developed to recognise radiopaque markers in biplane cineangiographic video-images with a background composed of different objects. Our connectionist approach is compared theoretically as well as experimentally with linear template matching. Theoretically, neural networks are likely to give better recognition results as they can implement nonlinear discriminants. Experiments confirm that the networks result in better marker recognitions than template matching. © 1999 Elsevier Science B.V. All rights reserved.

Keywords: Object detection; Object recognition; Radiopaque markers; Model-based image processing; Neural networks; Nonlinear filtering; Cardiac deformation

1. Introduction

Recognition of small circular-symmetric objects in images is a well-known problem for which a number of different techniques has been developed. Some techniques recognise an object from its distinct contours, others by a convolution with a kernel resembling the object or by fitting the parameters of an appropriate geometric model using the Hough transform (Hough, 1962). Depending on the shape and size of the object, the image background and the signal-to-noise-ratio in the image material, one approach may be preferred above the others while accounting for differences in recognition performance and computational complexity. In this paper, it is argued that convolution-based (filtering) approaches will give the best detection results when the objects to recognise are rotation invariant and relatively small (diameter <10–15 pixels). Two different techniques based on convolution will be compared theoretically as well as experimentally: the linear filter and a (nonlinear) convolution neural network. The objects we want to recognise are radiopaque makers in radiographic images.

Our goal is to develop a technique to distinguish markers from the background in radiographic images which contain many objects such as catheters, electrodes, calibration devices and ribs, objects that impede marker recognition (Fig. 1). First, we describe the physiological background and explain why a method based on radiopaque markers is used to measure cardiac deformation. Secondly, we consider different ap-

* Corresponding author. Tel.: +071 526 2285; fax: +071 524 8256; e-mail: michael@lkeb.azl.nl

[1] Present address: Department of Radiology, Division of Image Processing, Leiden University Medical Centre, P.O. Box 9600, 2300 RC Leiden, The Netherlands.

PII: S0167-8655(99)00024-0

Fig. 1. Grabbed video-image with 13 markers, ring, catheters and pace electrode, 288 × 384 pixels.

proaches to object recognition in images. Linear template matching is analyzed in detail. It is suggested instead to use a convolution neural network for recognition of markers because its theoretical properties are superior. A training set is built by combing artificially generated images of markers with samples of background extracted from video-images. The neural network is compared with linear template matching in two experiments.

2. Background

The functioning of the cardiac muscle is studied in medical applications as well as in physiological experiments. One important feature often used to characterise cardiac functioning is the *regional deformation pattern* of the heart chambers, the ventricles. The human heart is composed of a left and a right ventricle and two atria. In the left ventricle a major force is exerted in the process of pumping the blood into the aorta, the main trunk by which oxygen-rich blood starts its course to the tissues. The (smaller) right ventricle pumps blood into the pulmonary artery, which transports blood to the lungs. For example myocardial ischemia (a lack of blood supply to the myocardium) manifests itself as a deviating deformation pattern of one or both ventricles in the sense that a part of the cardiac muscle hardly contracts.

Several techniques are being used for studying cardiac deformation, among others the MRI-centerline method (Geest et al., 1997), MRI-tagging (Aelen et al., 1997; Kraitchman et al., 1995), ul-

trasound (Ascah et al., 1990; Gerova et al., 1989), angiograms (Zwet et al., 1998) and implantation of radiopaque markers (Gerbrands et al., 1979). On the one hand, standard ultrasound and angiograms are due to their two-dimensional character less suited for a precise assessment of cardiac deformation. On the other hand, the MRI-techniques lend themselves for characterising the short term deformation (within one heart cycle) using e.g. deformable active shape models such as SNAKES (Kass et al., 1987), point distribution models (Cootes et al., 1995) or a kinematic model (Aelen et al., 1997). The assessment of long-term cardiac deformation over months as a result of changes in tissue volume and shape, however, requires matching of fixed landmarks. Anatomical landmarks lend themselves for this purpose, but to few well-suited landmarks can be defined to ensure a precise assessment of regional deformation. Furthermore, the localization of such landmarks is ill-defined. In our situation, a physiological experiment was conducted in which the long-term cardiac growth was measured in animals whose aorta had been narrowed. We chose a technique based on implantation of radiopaque markers. In total 12–18 gold spheres (diameter = 1.5 mm) were implanted near the inner and outer boundaries of the ventricle wall of dogs. The cardiac deformation is quantified by 3-dimensional localisation of the markers in biplane cineangiographic (stereo) video-images (Fig. 2) before and after the intervention (narrowing the aorta). This 3-dimensional localisation requires an exact detection of each marker in the biplane video-images followed by stereoscopic analysis, see (Muijtjens, 1995).

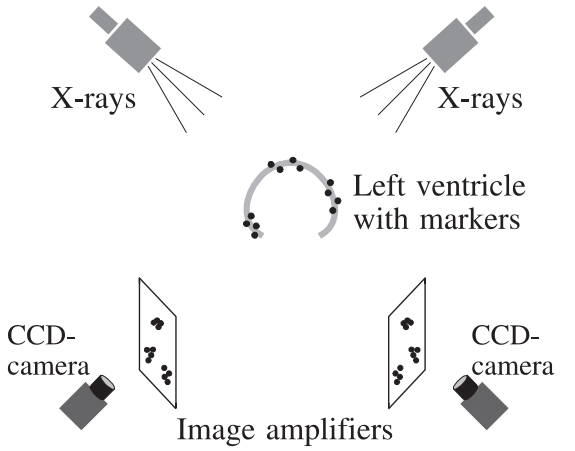

Fig. 2. Illustration of how the biplane cineangiographic video-images are recorded. The 3-dimensional marker localisations are obtained by matching the marker positions in the two frames.

3. Approaches for object recognition

In the field of digital image processing, recognition of objects is an essential and often difficult to tackle problem. Several techniques for automatic object recognition are available:

1. Contour-based recognition of each object using e.g. Fourier descriptors (Shridhar and Baldredin, 1984), or SNAKES (Kass et al., 1987; Lobregt and Viergever, 1995; Vieren et al., 1995).
2. Recognition based on the spatial intensity distribution of the object (Philips et al., 1983; Santamore et al., 1984) among others linear and inverse filtering.
3. Fitting a parametric model to the contours of an object using the Hough transform (Vrooman et al., 1998).

Contour-based methods are in general suited for recognition of large objects with a characteristic shape and are less appropriate for our problem. SNAKES, in particular, would be too tedious for marker recognition as the user would manually have to indicate the start position of each marker in an image sequence. Also the Hough transform has been applied for detection of radiographic markers (Vrooman et al., 1998). However, its major disadvantage is that the parameters are fitted on the contours of an object. These contours are obtained from a gradient image which is more sensitive to noise than the original (radiographic) image. Methods that utilise the complete intensity distribution, on the other hand, are most promising for automatic recognition of small circular-symmetric objects such as markers.

3.1. Filtering as detection method

Define an image as a two-dimensional signal $\boldsymbol{I}$ sampled on a grid. An object is a band-limited signal $\boldsymbol{S}$ and object recognition can be obtained by

filtering the image $\boldsymbol{I}$ using e.g. a linear filter (template matching, Gonzalez and Woods, 1992). When used for object recognition, linear filtering entails a convolution of the image with a kernel that resembles the spatial intensity distribution of the object, $\boldsymbol{D} = \boldsymbol{C} * \boldsymbol{I}$, i.e.,

$$\mathrm{d}(x,y) = \int_{\min(x')}^{\max(x')} \int_{\min(y')}^{\max(y')} c(x',y') \cdot i(x-x', y-y')\, \mathrm{d}x'\, \mathrm{d}y'. \tag{1}$$

Linear filtering discriminates two classes by a hyperplane of which the dimension equals the number of pixels in the slided window. The filter coefficients specified by the kernel determine the orientation of the hyperplane, the chosen threshold its location. The resulting output image contains bright spots at positions, where the correlation between window and kernel is high so thresholding this image results in marker recognitions.

3.2. Linear discriminant analysis

The pixels in the image $\boldsymbol{I}$ can also be seen as successive realisations of a random variable $i(x,y)$, the intensity of pixel (x,y). A convolution operation (Eq. (1)) entails a sample function which to each co-ordinate pair (x,y) results in a sample of pixel intensities, i.e., a window with the centre (x,y). Let us denote this sample function $w(x,y)$, a rectangular window with r rows and l columns. Object recognition can be seen as a classification task in which one wants to discriminate relevant objects from a, possibly heterogeneous, background. In this context, the linear filter Eq. (1) implements a linear discriminant function.

It is well-known that linear discriminants only result in optimal separation between the classes – markers and background – under strict assumptions as to how samples pertaining to the classes are distributed. Denote the class-conditional probability density function (PDF) of samples containing markers with $p(w(x,y)|\omega_m)$ and the PDF of samples containing (solely) background with $p(w(x,y)|\omega_b)$. Denote with $P(\omega_m)$ and $P(\omega_b)$ the prior probabilities of the two classes. The linear discriminant function can perform the optimal discrimination when $p(w(x,y)|\omega_m)$ and $p(w(x,y)|\omega_b)$ are both Gaussian and their covariance matrices are equal, $p(w(x,y)|\omega_m) \sim g(\boldsymbol{\mu}_m, \Sigma_m)$ and $p(w(x,y)|\omega_b) \sim g(\boldsymbol{\mu}_b, \Sigma_b)$, $\Sigma_m = \Sigma_b$ (see Duda and Hart, 1973). Under other specific circumstances, the linear discriminant may also result in the optimal discrimination. However, when the class-conditional distributions have different covariance matrices or are not Gaussian (e.g., multimodal), the linear discriminant gives inferior results compared to the minimal error-rate classifier, the optimal discriminant function (Duda and Hart, 1973).

The background in our image material contains catheters, a calibration ring, ribs and light shadows caused by the cardiac muscle. So it cannot be assumed that the intensity distribution of the background is (unimodal) Gaussian. The markers are occasionally blurred by motion so also their intensity distribution cannot be assumed to be Gaussian. Hence, the recognition problem cannot be performed optimally by a linear discriminant/filter.

3.3. Nonlinear discriminant analysis

Feed-forward neural networks have obtained a widespread use since the introduction of the Backpropagation algorithm (Rumelhart et al., 1986). This learning algorithm makes it possible to train nonlinear networks with so-called "hidden layers". A neural network consists of processing elements, so-called nodes. Each node has many inputs and generates one output value which is a nonlinear function of the weighted sum of the input signals the node receives.

A convolution neural network can be trained to recognise a marker from the intensity distribution of the sampling window that is slid across the image. Such a neural network featuring one hidden layer performs a nonlinear mapping from the $r \times l$-dimensional input space to the υ-dimensional class space

$$\begin{aligned} \boldsymbol{o} &= N(w(x,y)) \\ &= f(\boldsymbol{T}^2 f(\boldsymbol{T}^1 \mathrm{vec}(w(x,y)) - \boldsymbol{q}^1) - \boldsymbol{q}^2), \end{aligned} \tag{2}$$

with $w(x,y)$ being the sampling window, $\boldsymbol{T}^1$ is the weight matrix that connects the $r \times l$ input nodes with the h hidden nodes and $\boldsymbol{T}^2$ the weight matrix connecting the h hidden with the υ output nodes. $\boldsymbol{q}^1$ and $\boldsymbol{q}^2$ are the bias vectors of the hidden and output nodes, respectively. The function $f: \mathbb{R}^{\dim(\boldsymbol{a})} \rightarrow [\gamma,\eta]^{\dim(\boldsymbol{a})}$ is the nonlinear, bounded activation function applied to each element in the activation vector $\boldsymbol{a}$. Each element in the vector $\boldsymbol{o}$ represents the activation of a node in the output layer.

It has been proven that a feed-forward neural network is a nonlinear classifier capable of fitting any continuous (nonlinear) discriminant function to an arbitrary precision when provided with a sufficient number of nodes in the hidden layer (Funahashi, 1989; Hornik, 1989; Hornik et al., 1990; Villiers and Bernard, 1992). Additionally, it has been proven that the output vector of the network approximates the Bayesian posterior probabilities of the classes when the network is trained with a sufficient number of learning cases and equipped with enough hidden nodes (Richard and Lippmann, 1991). Therefore, in theory a neural network approximates a Bayes minimal error-rate classifier.

3.4. Nonlinear filtering with a neural network

In the previous section, it was shown that neural networks are in theory minimal error-rate classifiers which can give a better detection result than linear discriminants. However, object recognition can only be performed optimally when the sampling window $w(x,y)$ is large enough to capture the frequency components that are needed to discriminate markers from the background. Theoretically, the window should be so large that all potentially relevant frequency information is provided to the classifier. In practice, however, larger windows entail more complex classifiers which leads to an increased training time and a higher risk of overfitting the training set.

Neural networks have been trained to detect markers in our image material (Egmont-Petersen and Arts, 1996). The training set was composed of subimages of manually detected markers and of background. However, the scale of a marker occasionally varies depending on the relative distances between the X-ray source, the marker and the image amplifier. Secondly, in parts of the cardiac cycle the markers move so fast that they become blurred and take an elliptic shape. Therefore, we decided to build a model of the intensity distribution of a marker.

4. Intensity model

We will build an intensity model that makes it possible to generate images of (possibly blurred) markers with different contrast, position and size. The image background, however, is too complex to model and is extracted from existing image material. During training, the network will obtain as input the luminosity of the $r \times l$ pixels of the quadratic (convolution) window $w(x,y)$, $r = l$. After training, the convolution neural network produces an output image in which each pixel equals the difference $o_1 - o_2$ ($o_1 \sim p(w(x,y)|\omega_m)$ and $o_2 \sim p(w(x,y)|\omega_b)$).

4.1. The intensity distribution of a marker

The intensity distribution of a sphere in a radiograph depends on the absorption of the material and of the components and settings of the X-ray imaging system that is used. When the modulation transfer functions (MTFs) of the system components such as the focus, image intensifier and video camera are known, the absorption of an X-ray penetrating the sphere can be modelled by combining these MTFs with the general absorption equation (Lehmann et al., 1997; Ramze Rezaee et al., 1996). The parameters of the MTFs are difficult to estimate so we chose to define a parametric model that specifies the intensity distribution of markers observed in our image material.

We investigated different transcendental functions that lend themselves as approximation for the absorption function. We suggest to use the trigonometric function

$$z(e) = \alpha + \beta(1 + \cos(e/\lambda\pi - \pi)), \quad e \leqslant \lambda, \tag{3}$$

to approximate the absorption of a sphere as a function of the (Euclidean) length $e = (x^2 + y^2)^{1/2}$

of the projection vector in the image plane. The parameter α is the intensity of the background, β the amplitude of the marker signal and λ the radius of the marker.

Model validation was performed by estimating the parameters α, β and λ in Eq. (3) as follows. We extracted in total 45 subimages (7×7 pixels) of markers from 3 identically scaled video-images, each from a different sequence. Each subimage was centred on the (manually detected) marker. The intensities of the subimage were normalised to the interval (0,10), the image was inverted and the centre of gravity was computed using the intensities as weights (Gonzalez and Woods, 1992). In each subimage, the Euclidean distance e of each pixel to the centre of gravity was computed. The parameters of Eq. (3) were estimated by linear regression while varying the radius, $\lambda = 3.55, 3.60, \ldots, 4.15$. The best LMS-fit resulted in the parameters $\lambda = 3.9$, $\alpha = 0.495$ (± 1.224, standard error), $\beta = 4.034$ (± 0.058) and the correlation coefficient $R^2 = 0.786$ (Fig. 3). Third- and fourth-order polynomials were also fitted on the same data. The third-order polynomial resulted in a poor fit whereas the fourth-order polynomial gave a better fit, $R^2 = 0.792$. The tails of the two residual plots Fig. 4(a) and (b) indicate that the trigonometric function Eq. (3) is positioned centrally in the residual distribution whereas the fourth-order polynomial is not. Both plots depict heteroscodasity. These results indicate that Eq. (3) is a satisfactory approximation of the intensity distribution of a marker in our image material.

4.2. Adding background

Besides the intensity distribution of the marker itself, a marker image should also have a

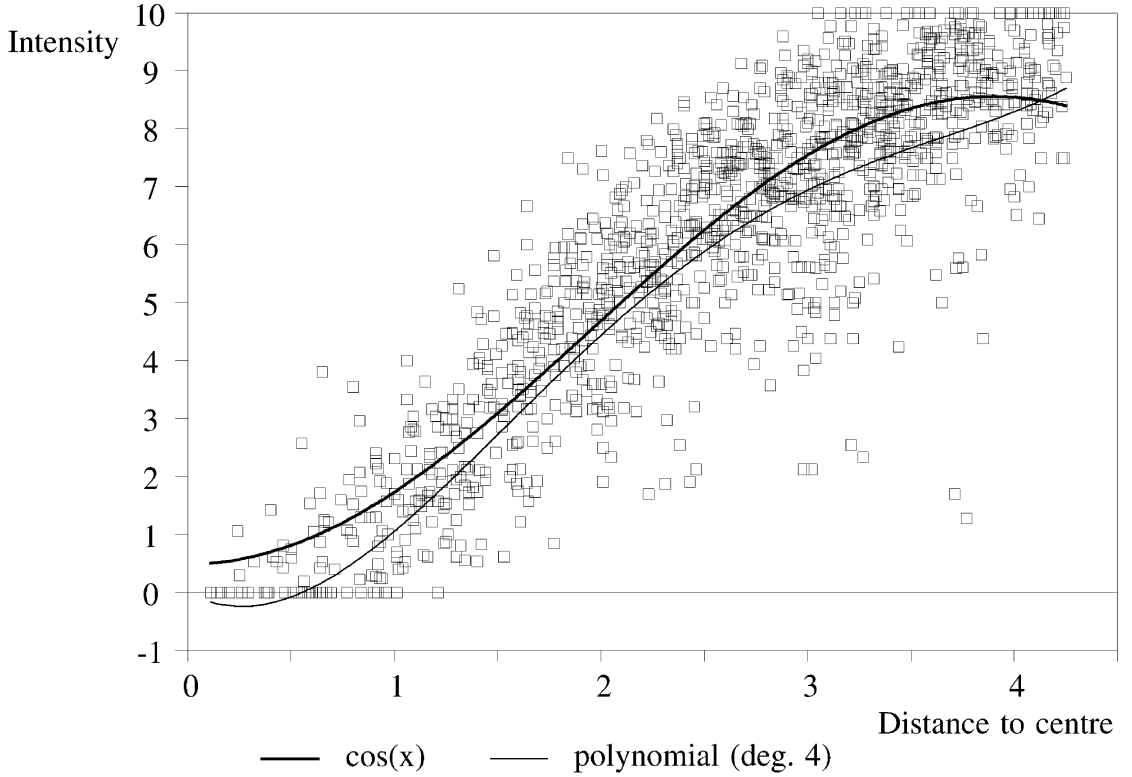

Fig. 3. Regression plot of pixel intensity as a function of Euclidean distance of the pixel to the centre of gravity (cos(x) and fourth-order polynomial).

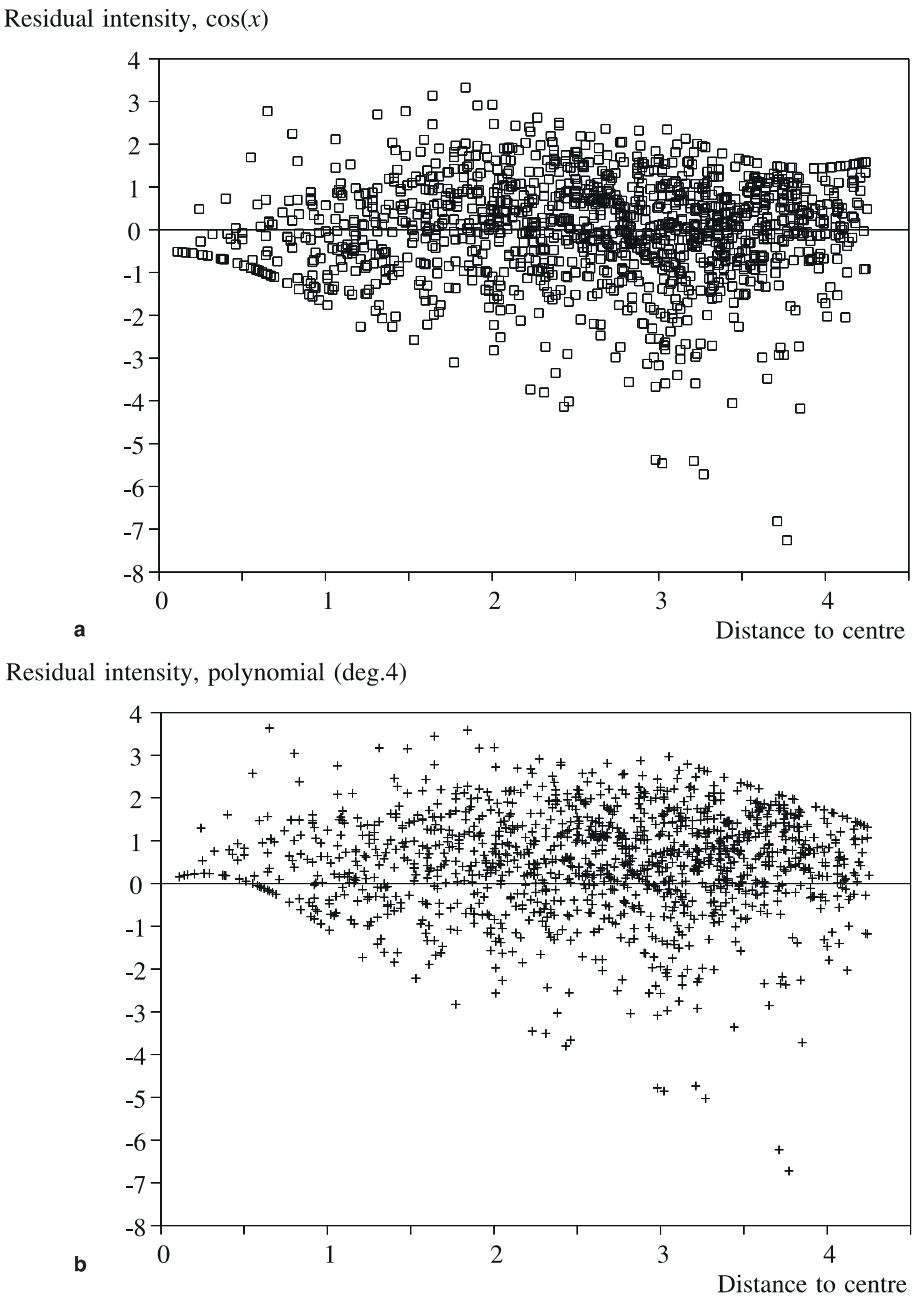

Fig. 4. Residual plots of regressions: (a) cos(x) and (b) a fourth-order polynomial. The tails indicate that the trigonometric function is positioned centrally in the residual distribution whereas the fourth-order polynomial is not.

background texture. In our case, the background is obtained by drawing at random a subimage $b(x, y)$ from one of the video-images at a position which does not contain a marker. The average intensity is subtracted from this background texture.

The intensity of each pixel in a marker subimage is given by the sum of the absorption of the sphere and the background structures along the X-rays. As the amplitude of the background texture is small compared to the signal obtained from the marker, we chose to add the background texture to the marker image with a weight that decreases linearly with the absorption of the marker,

$$s(x,y) = k\left(z\left(\sqrt{x^2+y^2}\right) - \mu_{\text{marker}}\right) + (b(x,y) - \mu_{\text{background}})z\left(\sqrt{x^2+y^2}\right). \quad (4)$$

The (integer) co-ordinates have the range $x,y \in -\frac{1}{2}r+\frac{1}{2},\ldots,\frac{1}{2}r-\frac{1}{2}$, with $r \times r$ the size of the marker image $s(x,y)$. The letter k denotes the contrast of the marker in relation to the background. The parameters of Eq. (3) were set to $\alpha = 0$ and $\beta = \frac{1}{2}$ such that $\max_e(z(e)) = 1$ and μ_{marker} the average absorption of the marker given by $z(x,y)$ and $\mu_{\text{background}}$ the average absorption of the background texture $b(x,y)$.

The centre of a marker can be chosen with subpixel precision by performing a phase transition in the frequency domain

$$\mathscr{F}_2\{s(\Delta x,\Delta y)\} = S(u,v)\exp(-\mathrm{j}2\pi(ux_0+vy_0)/r), \quad (5)$$

with $\Delta x = (x-x_0)$, $\Delta y = (y-y_0)$, $\mathscr{F}_2$ the two-dimensional Fourier transform, j the imaginary unit, (x,y) the co-ordinate pair in the spatial domain, (x_0,y_0) the new origin after translation and (v,u) the co-ordinate pair in the frequency domain.

5. Experiments

Two experiments with neural networks were conducted. The performance of the networks was compared with that of template matching (Gonzalez and Woods, 1992, p. 584). In our implementation of template matching, the average intensities of the window and the kernel were both set to zero but the signal amplitudes were not normalised. In both experiments, all neural networks were trained 3000 cycles with back-propagation (off-line learning). The learning rate was set to 0.0001, the momentum to 0.5 in all experiments.

The performance of the neural network and template matching was compared according to two quality metrics. The first is the correctness defined as

$$\rho = \frac{\iota}{n},$$

with ι being the number of subimages of markers plus the number of subimages of background that are correctly classified and n the total number of subimages in the training or test set (for confidence bounds on ρ see (Egmont-Petersen et al., 1994)). The second metric we use is the area under the ROC-curve (Receiver Operating Characteristic, see e.g. (Fogel et al., 1998)) which is computed by Riemann integration. Although the area under the ROC-curve is independent of a specific threshold value, this measure values the area obtained by integrating over the *low* false-positive rates equally with those obtained from integrating over the *high* false-positive rates. When ROC-curves cross the preferences of the user have to be taken into account in choosing the best recognition approach (Fogel et al., 1998).

A whole cardiac cycle typically lasts one second so 50 (field) video-images (50 Hz) are sufficient to distinguish the cardiac motion from the long term deformation. Regular stereo video-image series (2×64 images, 768×288 pixels) were grabbed with a frame grabber from Data Translation. The images were slightly blurred with an elliptic Gaussian kernel ($\sigma^2_{\text{horz}} = 1$, $\sigma^2_{\text{vert}} = \frac{1}{2}$, see (Lindeberg and ter Haar Romeny, 1994)) and subsampled with a factor two in the horizontal direction.

5.1. Experiment with nonmoving markers

In the first experiment, we investigated the effect varying the size of the window $r \times r$ provided as input to the neural network, the number of hidden nodes and the contrast k of each marker. We investigated three sizes of the quadratic window, $r \times r \in \{7\times 7, 9\times 9, 11\times 11\}$ pixels. The diameter of the marker $2\lambda = 7.8$ pixels so it fits completely within a window with the size $r \times r = 9 \times 9$ pixels. A smaller window size than 7×7 would discard too many pixels that contribute to the discrimination between marker and background.

Analyses of our image material indicated that the contrast of a marker varies between 10 and 80 intensities. In this experiment, the contrast was drawn at random from one of two uniform distributions, [2] $k_{\text{high}} \sim t(40,70)$ or $k_{\text{low}} \sim t(20,40)$. The subpixel translation was drawn at random from the uniform distribution $\Delta x, \Delta y \sim t(-\frac{1}{2},\frac{1}{2})$. The subimages representing background were drawn at random from one of three video-images and appended to the training set. The average intensity of each subimage was set to zero.

The markers with a high and a low contrast, respectively, were used to generate six learning sets, two sets for each of the three window sizes. In total 54 neural networks were trained while varying: the window size (3 settings), the marker contrast (2 settings) and the number of hidden nodes $(2,\ldots,4)$. Three networks, each with a different initial weight configuration, were trained for each combination of window size, contrast and number of hidden nodes.

The number of subimages with markers and solely background, respectively, were chosen in accordance with the "no free lunch" theorem (Wolpert and Macready, 1995) which states that in general statistical classifiers built from training sets with randomly chosen a priori distributions perform no better on a test set than random guessing. In our radiographic images, we used as ROI a circular mask that covers the image part inside the black frame (Fig. 1). Its area (about 40 000 pixels) corresponds with the area of about 850 markers. This implies a prior probability of 0.016 of observing a marker in a video-image containing 14 markers. All learning sets contained 7519 subimages of which 100 were markers and the remaining 7419 background, $P(\omega_m) = 0.013$ and $P(\omega_b) = 0.987$. In this experiment, 80% of the 7519 subimages were used for training the neural networks and the remaining randomly selected 20% of the subimages exclusively for testing purposes.

Three test video-images containing real markers, each from a different image sequence, were used to compare the performance of the neural networks and template matching (kernel computed from Eq. (3), 7×7 pixels). Among the 3 networks with the same number of hidden nodes, the ROC-curve of the network with the highest correctness on the test set was computed. The right boundary of the ROC-integral was chosen as the threshold value for which the poorest performing classifier could not recognise more true positive markers without recognising more than in total 30 (true and false positive) markers.

The networks trained with markers with a low contrast could not learn. We ascribe this to the high resemblance between a low contrast marker and the catheters which should be classified as background by the neural network. Solely the results of the networks trained with the learning sets in which the markers have high contrast are shown in Table 1. We computed the average correctness on the test set among each set of three neural networks with identical topology and window size. The networks based on a 7×7 window with 5 hidden nodes and on an 11×11 window with 3 hidden nodes seem to perform best.

When applied on real images, all networks with larger window sizes, $r \times r \in \{9 \times 9, 11 \times 11\}$, recognised parts of the ring and pace electrode as markers (see discussion). Therefore, we only computed the area under the ROC-curves for the networks with the smallest window size and for template matching (Table 2). The areas under the ROC-curves indicate that the neural networks with 4 and 5 hidden nodes perform better than template matching on all 3 test images.

5.2. Experiment with blurred markers

In parts of the systolic phase (contraction of the ventricles), rapid cardiac motion blurs the markers along the direction of movement. To obtain an optimal performance on these images, the neural networks need to be trained specifically to recognise blurred markers. For objects that move with a constant velocity while the camera shutter is open, the effect of motion blur is equivalent to convolving the image of the (stationary) object with a moving average filter (Yitzhaky and Kopeika, 1995). We built a training set in which the marker images were blurred artificially by a convolution

[2] The notation $x \sim t(a,b)$ indicates that x is drawn from a uniform distribution having the range (a,b).

Table 1
Average correctness (1-error rate) and standard deviation of the networks computed on a test set with 2000 cases, first experiment; three networks were trained for each combination of topology and window size

Window size	Contrast	Correctness	3 hidden	4 hidden	5 hidden
7×7	40–70	μ	0.9945	0.9978	0.9988
		σ	0.0016	0.0007	0.0008
9×9	40–70	μ	0.9983	0.9981	0.9982
		σ	0.0007	0.0008	0.0008
11×11	40–70	μ	0.9988	0.9987	0.9986
		σ	0.0010	0.0008	0.0010

Table 2
The areas under the ROC-curves for the neural networks with 3, 4 and 5 hidden nodes, which were trained with non-moving markers, and the linear filter, $r \times r = 7 \times 7$ pixels, first experiment; the network with 5 hidden nodes performs best

	Template matching	NN 3 hidden	NN 4 hidden	NN 5 hidden
Image 1	0.848	0.934	0.969	0.969
Image 2	0.904	0.958	0.987	0.985
Image 3	0.957	0.946	0.995	1.000

with one of the moving average filters $\boldsymbol{t}_2 = (0.5\ 0.5)$ or $\boldsymbol{t}_3 = (0.33\ 0.33\ 0.33)$. The probability that the marker image was not blurred, blurred with 2 ($\boldsymbol{t}_2$) or with 3 pixels ($\boldsymbol{t}_3$) equals 0.33. Each blurred marker image was rotated using bilinear interpolation with a randomly chosen angle $\phi \sim t(0, \pi)$. The other settings remained identical to the experiment with nonmoving markers.

For each window size $r \times r \in \{7 \times 7, 9 \times 9, 11 \times 11\}$ pixels, 9 neural networks, 3 with 3, with 4 and with 5 hidden nodes, were trained. The learning sets from the first experiment were used except from the marker images which we replaced by images of blurred markers. Among the 27 networks, the 3 with the best performance on the test set were evaluated on 2 video-images containing real, blurred markers. In the first image, almost all 14 markers were blurred whereas in the second image, mainly one group consisting of 5 markers was blurred.

Also in this second experiment, the networks with the larger window sizes ($r \times r \in \{9 \times 9, 11 \times 11\}$ pixels) gave poor recognition results so solely the ROC-curves of the networks with a window size $r \times r = 7 \times 7$ pixels were computed. The area under the ROC-curves for the neural networks and for template matching are shown in Table 3, which shows also the ROC-integral that results from applying the best neural network from the first experiment (5 hidden nodes, 7×7 window) on these 2 video-images. The neural network with 5 hidden nodes trained with images of *blurred* markers, performs best. However, the best network trained with *nonmoving* markers in the first experiment performs also well on the two test images. Fig. 5 shows the ROC-curves of 4 different neural networks and of template matching computed on the test image in which all markers were blurred. It is clear that the neural network trained with images of nonmoving markers performs best

Table 3
The areas under the ROC-curves for the neural networks with 3, 4 and 5 hidden nodes trained with blurred markers, template matching and the neural network with 5 hidden nodes, no blur, $r \times r = 7 \times 7$ pixels, second experiment

	Template matching	NN 3 hidden	NN 4 hidden	NN 5 hidden	NN 5 hidden no smear
Image 1	0.505	0.552	0.495	0.657	0.574
Image 2	0.621	0.815	0.812	0.862	0.831

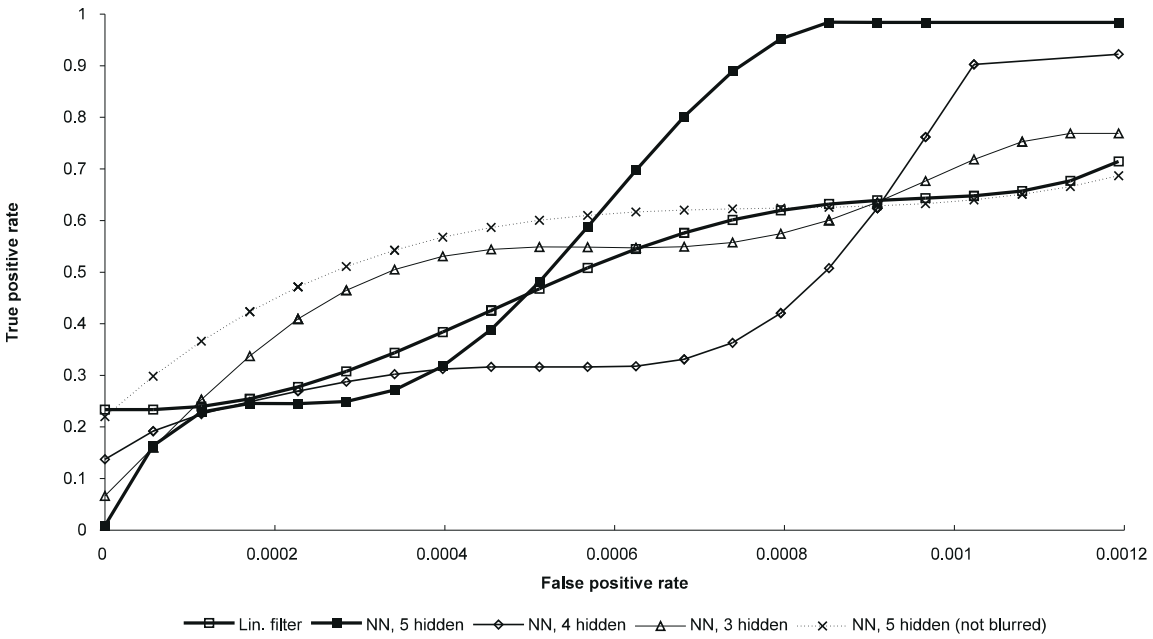

Fig. 5. ROC-curves of neural networks trained with blurred and non-moving markers tested on an end-systolic video-image. To avoid overlapping curves, each ROC-curve was fitted with a polynomial with 5–6 terms. All ROC-integrals were computed directly from the data.

when not all markers need to be recognised. On the other hand, the network with 5 hidden nodes trained with blurred markers could eventually recognise more true markers when the threshold was lowered enough.

6. Discussion

The results of the first experiment indicate very small differences in error-rates among all the neural networks (Table 1). However, the networks with larger window sizes perform poorly on real video-images because they overfit the learning set. In our experiments, the ratio ζ/n between the number of parameters of the neural network (denoted ζ) and the number of training images (denoted n), varies with the window size. The ratio ζ/n is directly related to the so-called Vapnik-Chervonenkis dimension (Baum and Haussler, 1989) which states that ζ/ξ training cases are needed to guarantee that a trained (feed-forward) neural network with ζ parameters has an error-rate less than ξ (the VC-dimension is an upper bound). However, we believe that the problem of overfitting is caused by the fact that each input (pixel) in the window $w(x,y)$ is processed independently of its neighbours by the neural network. For pixels outside a marker (border pixels in windows larger than 7×7), the expected mean difference in intensity between a window containing a marker and a window containing solely background is per definition zero. These border pixels contribute only indirectly to the discrimination. To better cope with this problem, one could develop an approach that takes the local frequency information into account.

In the training set, we chose the class distribution in accordance with the prior probabilities that the window $w(x,y)$ contains a marker or background, respectively. The ratio between the prior probabilities $P(\omega_m)/P(\omega_b)$ was set equal to the ratio between the area in pixels of the (circular) region of interest and the area occupied by 10–20

markers. It should be noticed that the ratio $P(\omega_m)/P(\omega_b)$ becomes smaller if during the convolution with the neural network a marker is only recognised when it appears centrally in the window. However, preliminary experiments with neural networks indicated that decreasing the fraction of marker images leaded to a poor performance.

7. Conclusion

In this article, we have developed an approach for recognition of markers in cineangiographic images based on neural networks. The learning set used to train the neural networks was generated with a stochastic model of the intensity distribution of a marker whereas the superimposed background was drawn at random from available video-images. The experiments indicate that the neural networks perform better than the standard technique – template matching – with respect to recognition of circular-symmetric markers.

The networks trained with blurred markers perform better than the linear filter on images containing blurred markers. In our application, only 10% of an image sequence contains any blurred markers so we apply the best network trained to recognise nonmoving markers on the whole video-sequence. The exact position of each marker is computed as the centre of gravity (pixels within a circle with a diameter of 7.8 pixels).

Convolving an image with a neural network takes about 10 s so off-line analysis of a sequence consisting of 64 stereo pairs of images takes 20 min which is sufficient in our application. Generally, each image contains 1–3 false positive markers whereas 1–2 markers are missed. The incomplete marker tracks are subsequently interpolated by applying an SVD-based technique on the coordinate pairs (Muijtjens, 1995).

An interesting finding is that well-performing neural networks could be trained with artificially generated images of markers. On the one hand, a pure model-based approach for generating the whole learning set is impeded by the complexity of the heterogeneous background in our image material. On the other hand, our combined model- and data-driven approach has the advantage over a pure data-driven one that images of (blurred) markers with different contrast, position and size can be generated. Such a set-up is especially suited for applications, where the object to recognize is simple to characterize but the characteristics of the ever present background are too complex to model.

References

Aelen, F.W.L., Arts, T., Sanders, D.G.M., Thelissen, G.R.P., Muijtjens, A.M.M., Prinzen, F.W., Reneman, R.S., 1997. Relation between torsion and cross-sectional area change in the human left ventricle. Journal of Biomechanics 30 (3), 207–212.

Ascah, K.J., King, M.E., Gillam, L.D., Weyman, A.E., 1990. The effects of right ventricular hemodynamics of left ventricular configuration. Canadian Journal of Cardiology 6 (3), 99–106.

Baum, E., Haussler, D., 1989. What size net gives a valid generalization?. Neural Computation 1 (1), 151–160.

Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J., 1995. Active shape models – their training and application. Computer Vision and Image Understanding 61 (1), 38–59.

Duda, R.O., Hart, P.E., 1973. Pattern Classification and Scene Analysis. Wiley, New York.

Egmont-Petersen, M., Arts, T., 1996. Detection of implanted markers in radiographic image sequences. In: Lehmann, T., Spitzer, K. (Eds.), Tagesband zur Aachener Workshop Bildverarbeitung in der Medizin. In: Proceedings of the Aachen Workshop on Medical Image Processing, Aachen, pp. 209–214.

Egmont-Petersen, M., Talmon, J.L., Brender, J., NcNair, P., 1994. On the quality of neural net classifiers. Artificial Intelligence in Medicine 6 (5), 359–381.

Fogel, D.B., Wasson, E.C., Boughton, E.M., Porto, V.W., Angeline, P.J., 1998. Linear and neural models for classifying breast masses. IEEE Transactions on Medical Imaging 17 (3), 485–488.

Funahashi, K.-I., 1989. On the approximate realization of continuous mappings by neural networks. Neural Networks 2, 183–192.

Gerbrands, J.J., Booman, F., Reiber, J.H.C., 1979. Computer analysis of moving radiopaque markers from X-ray cinefilms. Computer Graphics and Image Processing 11, 35–48.

Geest, R.J.v.d., de Roos, A., Wall, E.E.v.d., Reiber, J.H.C., 1997. Quantitative analysis of cardiovascular MR images. International Journal of Cardiac Imaging 13, 247–258.

Gerova, M., Barta, E., Stolarik, M., Gero, J., 1989. Geometry of the conduit coronary artery in diastole is determined by the volume of the left and right ventricles. Basic Research in Cardiology 84 (6), 583–590.

Gonzalez, R.C., Woods, R.E., 1992. Digital Image Processing. Addison-Wesley, Reading, MA.

Hornik, K., 1989. Multilayer feedforward networks are universal approximators. Neural Networks 2, 359–366.

Hornik, K., Stinchcombe, M., White, H., 1990. Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks. Neural Networks 3, 551–560.

Hough, P.V.C., 1962. A method and means for recognizing complex pattern. US Patent Application No. 3069654.

Kass, M., Witkin, A., Terzopoulos, D., 1987. Snakes: Active contour models. International Journal of Computer Vision 13, 321–331.

Kraitchman, D.L., Young, A.A., Chang, C.N., Axel, L., 1995. Semi automatic tracking of myocardial motion in MR tagged images. IEEE Transactions on Medical Imaging 14 (4), 422–433.

Lehmann, T., Oberschelp, W., Pelikan, E., Repges, R., 1997. Bildverarbeitung für die Medizin. Springer, Berlin.

Lindeberg, T., ter Haar Romeny, B.M., 1994. Linear scale-space theory, Chapter 2. In: ter Haar Romeny, B.M. (Ed.), Geometry-driven Diffusion in Computer Vision. Kluwer Academic Publishers, Dordrecht.

Lobregt, S., Viergever, M.A., 1995. A discrete dynamic contour model. IEEE Transactions on Medical Imaging 14 (1), 12–24.

Muijtjens, A.M.M., 1995. The reconstruction of 3-D maker trajectories in measuring heart motion. Doctoral Dissertation, Faculty of Medicine, Maastricht University, Maastricht.

Philips, C.M., Prenis, J., Santamore, W.P., Bove, A.A., 1983. Recognition and storage of metal heart marker position from biplane X-ray images at video rates. IEEE Transactions on Biomedical Engineering 30 (1), 10–17.

Ramze Rezaee, M., Zwet, P.M.J.v.d., Gerbrands, J.J., Reiber, J.H.C., 1996. A model for the modulation transfer function of cardiovascular X-ray systems. Investigative Radiology 31 (3), 161–172.

Richard, M.D., Lippmann, R.P., 1991. Neural network classifiers estimate bayesian a posteriori probabilities. Neural Computation 3, 461–483.

Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning internal representations by error propagation. In: Rumelhart, D.E., McClelland, J.L. (Eds.), Parallel Distributed Processing. Exploration into the Microstructure of Cognition, Vol. 1. MIT Press, Cambridge, MA, pp. 318–364.

Santamore, W.P., Whiteman, J., Negin, M., Heddy, R.F., 1984. Computer recognition of radiopaque myocardial markers. Computers and Biomedical Research 17 (6), 495–508.

Shridhar, M., Baldredin, A., 1984. High accuracy character recognition algorithm using Fourier and topological descriptors. Pattern Recognition 17 (5), 515–524.

Vieren, C., Cabestaing, F., Postaire, J.-G., 1995. Catching moving objects with snakes for motion tracking. Pattern Recognition Letters 16 (7), 679–685.

Villiers, J.D., Bernard, B., 1992. Backpropagation neural nets with one and two hidden layers. IEEE Transactions on Neural Networks 4, 136–141.

Vrooman, H.A., Valstar, E.R., Brand, G.-J., Admiraal, D.R., Rozing, P.M., Reiber, J.H.C., 1998. Fast and accurate measurements in digitized stereophotogrammetric radiographs. Journal of Biomechanics 31, 491–498.

Wolpert, D.H., Macready, W.G., 1995. No free lunch theorems for search. Technical Report SFI-95-02-010, The Santa Fe Institute.

Yitzhaky, Y., Kopeika, N.S., 1995. Identification of the blur extent from motion blurred images. In: Proceedings of the SPIE, Vol. 2470, pp. 2–11.

Zwet, P.M.J.v.d., Nettesheim, M., Gerbrands, J.J., Reiber, J.H.C., 1998. Derivation of optimal filters for the detection of coronary arteries. IEEE Transactions on Medical Image Processing 17 (1), 108–120.