# Bimodal Histogram Transformation Based on Maximum Likelihood Parameter Estimates in Univariate Gaussian Mixtures

Nette Schultz and Jens Michael Carstensen

Department of Mathematical Modelling, Building 321
Technical University of Denmark, DK-2800 Lyngby, Denmark. *

**Abstract.** This paper presents a bimodal histogram transformation procedure where conjugate gradient optimization is used for estimating maximum likelihood parameters of univariate Gaussian mixtures. The paper only deals with bimodal distributions but extension to multimodal distributions is fairly straightforward. The transformation is suggested as a preprocessing step that provides a standardized input to e.g. a classifier. This approach is used for pixelwise classification in RGB-images of meat.

## 1 Introduction

A transformation of image data distributions can be used in image enhancement to enhance specific sub levels in the intensity range. It can also be used as preprocessing before classification or segmentation. Common transformations transform the image data to more or less predetermined distributions with fixed parameters. By using a histogram match with predetermined distributions the histogram information is totally neglected. The bimodal histogram transformation described here does not remove all the histogram information, and is therefore interesting as standardization algorithm for data that is bimodal by nature. We will describe the bimodal histogram transformation, which estimates the parameters for a bimodal data set using maximum likelihood estimation, and only changes some of the parameters to predetermined fixed values. It could as well be a histogram transformation with more than two Gaussian distributions.

The method is used as a preprocessing step in a classification case study. The case study is to classify meat images pixel-wise into lean or fat, where the meat images are cross-sectional cuts from pork carcasses. The images are 8-bit rgb color. The light exposures vary from image to im-

* We would like to thank Hans Henrik Thodberg from the Danish Meat Research Institute and professor Knut Conradsen, associated professor Bjarne Ersbøll and the rest of the Image Group at Department of Mathematical Modelling, DTU.

age. Thus without some kind of standardization of the data the pixel-wise classification of lean and fat will be deteriorated, because the light intensity ranges in the images used as training set for the classifications model estimation are changing in relation to the rest of the data set. We therefore apply the bimodal histogram transformation to each rgb band.
When the intensity-function of the change in light exposure is unknown, and too much change in the distribution of an image is undesirable, the bimodal histogram transformation can be one approach to overcome the difficulties of light exposure changes.

## 2 Dealing with Varying Light Exposure

Our approach is to assume that the image is a mixture of Gaussians. We estimate the parameters - averages, variances and weights - and only change a minimum of parameters necessary to get the desired standardization or enhancement.
In this actual case study we will only be dealing with bimodality, but it can easily be extended to a mixture of a larger number of Gaussians.
When a data set is classified by means of supervised models, a training set is necessary. The training set is used for estimating parameters in the model. A common problem in many applications is, that the data set distributions can vary for different incidents. This means that the training set used for model estimation will be different from the new data set, and can cause a poor classification.
For pixel classification in an image, the light intensities can change over the image or from one image to another over time, dependent of the environment and image acquisition applications. In the meat classification case study, the pixels' light intensities are varying from dark and under exposed to light and over exposed images. Fig. 1 shows two different examples of meat images. The image intensities are changing in the range from dark to over exposed. These unstable light conditions makes a classification on the raw images troublesome.
One way to overcome this problem is by using features which are invariant in relation to the intensity changes. Features which are invariant to linear transformations are e.g. skewness, kurtosis and orthogonal transformations like canonical variables and principal components based on the correlation matrix.
The transformation is, however, often nonlinear. Another approach is to transform the raw data to a more or less fixed histogram distribution. A gentle transformation where only some of the distribution parameters are fixed can therefore be suitable. This transformation can be an univariate color scale calibration for images with bimodal histograms, where only the averages have been changed.

# 3 Maximum Likelihood Estimation of Parameters for Bimodal Data Sets

Assuming the data set is a mixture of Gaussians, see e.g. [3] the parameters can be estimated and used in the histogram transformation.
In the univariate case the likelihood and its derivatives can easily be derived, and conjugate gradient can be used for estimating the maximum likelihood parameters. The multivariate case induces difficulties with the derivatives. Approximations to maximum likelihood estimation like the EM algorithm in [2] could be used in the multivariate case.

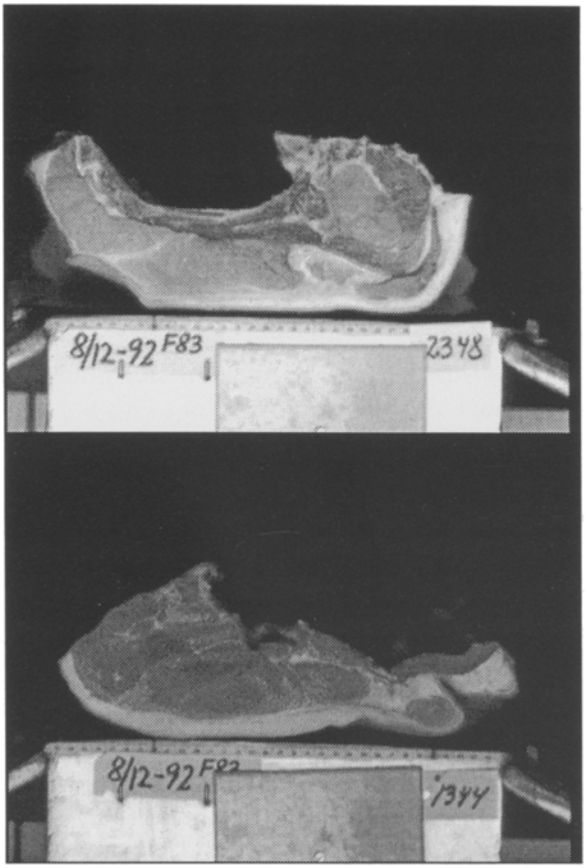


**Fig. 1.** From top to bottom an over exposed and under exposed image of meat is shown.

In the meat image pixel classification application, the histogram is assumed to be a mixture of two Gaussian distributions. Therefore we will only be dealing with bimodality, but the results can easily be extended to mixtures of more than two Gaussian distributions.
The likelihood is

$$L(\mu_1,\mu_2,\sigma_1,\sigma_2,\alpha) = \prod_{i=1}^{N}(\alpha\frac{1}{\sqrt{2\pi}}\frac{1}{\sigma_1}\exp(-\frac{1}{2}(\frac{t_i-\mu_1}{\sigma_1})^2)$$
$$+(1-\alpha)\frac{1}{\sqrt{2\pi}}\frac{1}{\sigma_2}\exp(-\frac{1}{2}(\frac{t_i-\mu_2}{\sigma_2})^2)), \quad (1)$$

where $N$ is the number of observations, and $t_i$ is the value of observation number $i$.
The log likelihood is

$$\ln L(\mu_1,\mu_2,\sigma_1,\sigma_2,\alpha) = \sum_{i=1}^{N}\ln(\alpha\frac{1}{\sqrt{2\pi}}\frac{1}{\sigma_1}\exp(-\frac{1}{2}(\frac{t_i-\mu_1}{\sigma_1})^2)$$
$$+(1-\alpha)\frac{1}{\sqrt{2\pi}}\frac{1}{\sigma_2}\exp(-\frac{1}{2}(\frac{t_i-\mu_2}{\sigma_2})^2)) \quad (2)$$

Let the data set be the pixel values in an image. The number of observations is the number of pixels in the image, and the pixel values are integers in the range $[0; Maxval]$. Then the log likelihood can be written as

$$\ln L(\mu_1,\mu_2,\sigma_1,\sigma_2,\alpha) = \sum_{i=0}^{Maxval} n_i f^*(i,\mu_1,\mu_2,\sigma_1,\sigma_2,\alpha) \quad (3)$$

where $n_i$ is the number of pixels with pixel value $i$, and

$$f^*(i,\mu_1,\mu_2,\sigma_1,\sigma_2,\alpha) = \ln(\alpha\frac{1}{\sqrt{2\pi}}\frac{1}{\sigma_1}\exp(-\frac{1}{2}(\frac{i-\mu_1}{\sigma_1})^2)$$
$$+(1-\alpha)\frac{1}{\sqrt{2\pi}}\frac{1}{\sigma_2}\exp(-\frac{1}{2}(\frac{i-\mu_2}{\sigma_2})^2)) \quad (4)$$

We want to estimate the parameters $(\mu_1,\mu_2,\sigma_1,\sigma_2,\alpha)$ as

$$\arg\max \ln L(\mu_1,\mu_2,\sigma_1,\sigma_2,\alpha). \quad (5)$$

As we are dealing with parameters belonging to Gaussian distributions, we have the following constraints

$$\sigma_1 > 0, \tag{6}$$
$$\sigma_2 > 0, \tag{7}$$

and

$$0 \leq \alpha \leq 1. \tag{8}$$

Unfortunately the log likelihood has maxima outside the range of the constraints. A way to deal with the constraints is to let

$$\sigma_1 = v_1^2, \tag{9}$$
$$\sigma_2 = v_2^2, \tag{10}$$
$$\alpha = \exp^{(-a^2)}, \tag{11}$$

and

$$\begin{aligned} &f(i, \mu_1, \mu_2, v_1, v_2, a) = \\ &\ln(\exp^{(-a^2)} \frac{1}{\sqrt{2\pi}} \frac{1}{v_1^2} \exp(-\frac{1}{2}(\frac{i-\mu_1}{v_1^2})^2) \\ &+(1-\exp^{(-a^2)}) \frac{1}{\sqrt{2\pi}} \frac{1}{v_2^2} \exp(-\frac{1}{2}(\frac{i-\mu_2}{v_2^2})^2))). \end{aligned} \tag{12}$$

Then

$$\ln L(\mu_1, \mu_2, v_1, v_2, a) = \sum_{i=0}^{Maxval} n_i f(i, \mu_1, \mu_2, v_1, v_2, a). \tag{13}$$

Now estimating the parameters $(\mu_1, \mu_2, v_1, v_2, a)$ for the maximum likelihood

$$\arg\max \ln L(\mu_1, \mu_2, v_1, v_2, a) \tag{14}$$

will satisfy the constraints in 6-8.
When we use conjugate gradient as optimization algorithm the first partial derivatives have to be calculated. We have

$$\frac{d(\ln(L(\mu_1, \mu_2, v_1, v_2, a)))}{d\mu_1} = \sum_{i=0}^{Maxval} n_i \frac{df}{d\mu_1} \tag{15}$$

$$\frac{d(\ln(L(\mu_1, \mu_2, v_1, v_2, a)))}{d\mu_2} = \sum_{i=0}^{Maxval} n_i \frac{df}{d\mu_2} \tag{16}$$

$$\frac{d(\ln(L(\mu_1, \mu_2, v_1, v_2, a)))}{dv_1} = \sum_{i=0}^{Maxval} n_i \frac{df}{dv_1} \tag{17}$$

$$\frac{d(\ln(L(\mu_1, \mu_2, v_1, v_2, a)))}{dv_2} = \sum_{i=0}^{Maxval} n_i \frac{df}{dv_2} \tag{18}$$

$$\frac{d(\ln(L(\mu_1, \mu_2, v_1, v_2, a)))}{da} = \sum_{i=0}^{Maxval} n_i \frac{df}{da}. \tag{19}$$

Where

$$\frac{df}{d\mu_1} = \tag{20}$$

$$\frac{\exp(-a^2)\frac{1}{v_1^6}(i-\mu_1)\exp(-\frac{1}{2}(\frac{i-\mu_1}{v_1^2})^2)}{\exp^{(-a^2)}\frac{1}{v_1^2}\exp(-\frac{1}{2}(\frac{i-\mu_1}{v_1^2})^2) + (1-\exp^{(-a^2)})\frac{1}{v_2^2}\exp(-\frac{1}{2}(\frac{i-\mu_2}{v_2^2})^2)}$$

$$\frac{df}{d\mu_2} = \tag{21}$$

$$\frac{(1-\exp(-a^2))\frac{1}{v_2^6}(i-\mu_2)\exp(-\frac{1}{2}(\frac{i-\mu_2}{v_2^2})^2)}{\exp^{(-a^2)}\frac{1}{v_1^2}\exp(-\frac{1}{2}(\frac{i-\mu_1}{v_1^2})^2) + (1-\exp^{(-a^2)})\frac{1}{v_2^2}\exp(-\frac{1}{2}(\frac{i-\mu_2}{v_2^2})^2)}$$

$$\frac{df}{dv_1} = \tag{22}$$

$$\frac{2\exp(-a^2)\frac{1}{v_1^3}\exp(-\frac{1}{2}(\frac{i-\mu_1}{v_1^2})^2)((\frac{i-\mu_1}{v_1^2})^2-1)}{\exp^{(-a^2)}\frac{1}{v_1^2}\exp(-\frac{1}{2}(\frac{i-\mu_1}{v_1^2})^2) + (1-\exp^{(-a^2)})\frac{1}{v_2^2}\exp(-\frac{1}{2}(\frac{i-\mu_2}{v_2^2})^2)}$$

$$\frac{df}{dv_2} = \tag{23}$$

$$\frac{2(1-\exp(-a^2))\frac{1}{v_2^3}\exp(-\frac{1}{2}(\frac{i-\mu_2}{v_2^2})^2)((\frac{i-\mu_2}{v_2^2})^2-1)}{\exp^{(-a^2)}\frac{1}{v_1^2}\exp(-\frac{1}{2}(\frac{i-\mu_1}{v_1^2})^2) + (1-\exp^{(-a^2)})\frac{1}{v_2^2}\exp(-\frac{1}{2}(\frac{i-\mu_2}{v_2^2})^2)}$$

$$\frac{df}{da} = \tag{24}$$

$$\frac{2a(-\frac{1}{v_1^2}\exp(-\frac{1}{2}(\frac{i-\mu_1}{v_1^2})^2) + \frac{1}{v_2^2}\exp(-\frac{1}{2}(\frac{i-\mu_2}{v_2^2})^2))\exp(-a^2)}{\exp^{(-a^2)}\frac{1}{v_1^2}\exp(-\frac{1}{2}(\frac{i-\mu_1}{v_1^2})^2) + (1-\exp^{(-a^2)})\frac{1}{v_2^2}\exp(-\frac{1}{2}(\frac{i-\mu_2}{v_2^2})^2)}.$$

A good starting value of the parameters can speed up the conjugate gradient algorithm, and make it less likely that the maximum likelihood estimation jumps into local maxima caused by noisy data. We obtain a starting guess by estimating the parameters on the one-dimensional mean filtered histogram. The averages are located as local maxima, the alfa by the local minimum in between the averages and the variances by the 5% and 95% fractiles.

Conjugate gradient is used because it is relatively fast and the derivatives can be calculated.

The bimodal transformation procedure is as follows

1 Estimate a starting guess of the parameter values.
2 Calculate the derivatives of the parameters.
3 Estimate new parameter values with conjugate gradient.
4 If the parameter-changes are below a pre-set limit, we may proceed, otherwise go to step 2.
5 Transform the image histogram to a histogram with the predetermined parameter values and the other estimated parameter values.

# 4 Case Study: Pixel-wise Classification of Meat Images

The Danish slaughter-houses estimate the carcass meat percent by using measurements as total weight, different anatomical lengths and local meat and fat measurements from optical insertion probes.
By use of vision, we can get alternative features, for estimation of the meat percentage, see e.g. [1].
The areas of lean and fat in a slice of meat are intuitively correlated with the meat percent. Therefore, after the carcass has been divided into front part, middle part and ham as it is usual in a slaughter house, one image in color rgb, has been taken of the cut between front and middle part, and a second image has been taken of the cut between middle part and ham. Fig. 1 shows an image of the front (top) and an image of the ham (bottom).
Assuming that the background has been removed from the image (by use of deformable templates as described in [1]). Our aim on this step is to classify the meat into lean and fat pixel by pixel.

## 4.1 The Histogram Transformation

The data set consists of 572 images. 283 images of the front and 289 images of the ham. In [4] several different classification methods have been used and compared on a subset of the data set. In the comparison a neural network without hidden layers worked fairly well. The input features to the network was the raw rgb bands, the local median filtered bands and local standard deviation filtered bands. Therefore we will use this classification method. A representative training set has been selected from 12 of the images for model estimation.

Because of the changes in light exposure over images we apply a bimodal histogram transformation to the raw rgb bands before any feature extraction. Here the averages are changed to predetermined values.

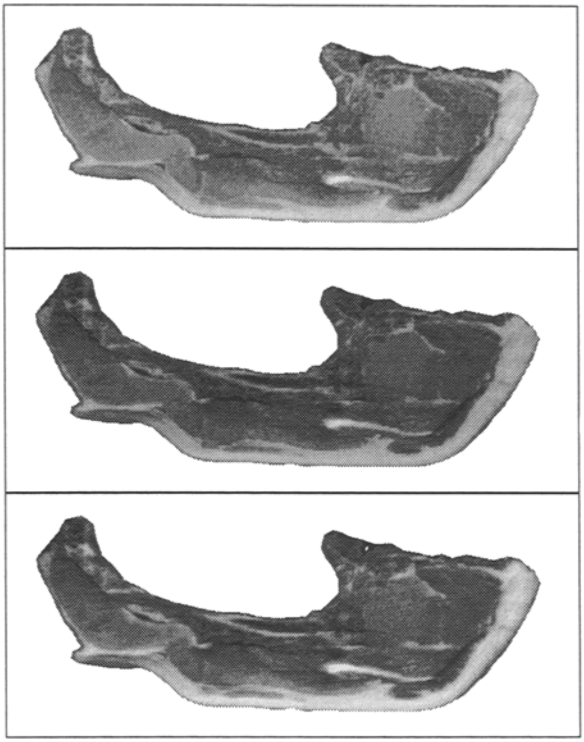

**Fig. 2.** Each image consist of three bands of format byte.From top to bottom the red, green and blue band is shown.

In Fig. 2 the red, green and blue band of an image from the front is shown. The corresponding histograms are shown in Fig. 3. The pixel values are in the range 0-255. The histograms do not look very nice.

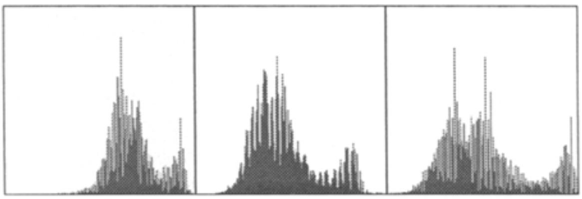

**Fig. 3.** Histograms in range 0-255 of the red, green and blue band from left to right.

They are very spiky, probably due to a preference with regard to the least significant bit. Furthermore there is a clear difference between the green compared to the red and blue band recording method. Therefore a change in the sample space can be preferable.

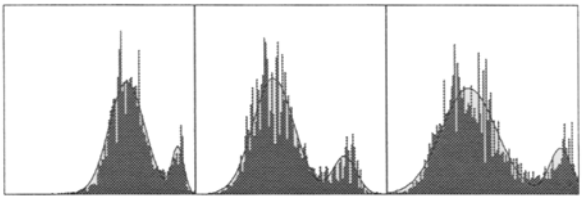

**Fig. 4.** Histograms in range 0-128 of the red, green and blue band from left to right.

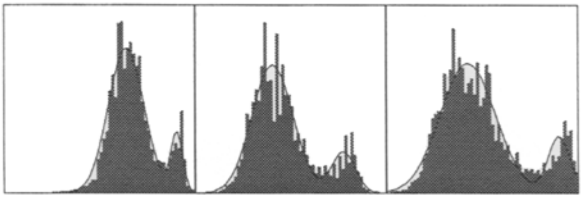

**Fig. 5.** Histograms in range 0-64 of the red, green and blue band from left to right.

In Fig. 4 and 5 the histograms with sample ranges 0-128 and 0-64 are shown. The corresponding estimated density, where the maximum likelihood estimation of parameters is used, is also shown as an overlaid graph. The densities seem to fit the data set fairly well, though the data set is noisy and actually could look like a mixture of three Gaussians. Because the lean can vary in darkness coursed by anatomical differences in the muscles, the lean class distribution could actually be a mixture of two normal distributions. But these differences are only visible in some of the images.

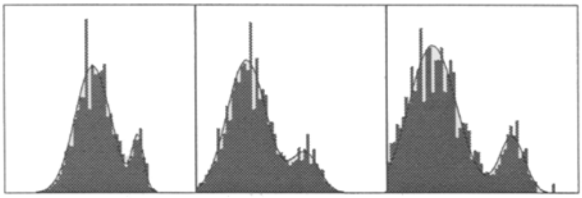

**Fig. 6.** Histograms in range 0-64 of the bimodal histogram transformed red, green and blue band from left to right.

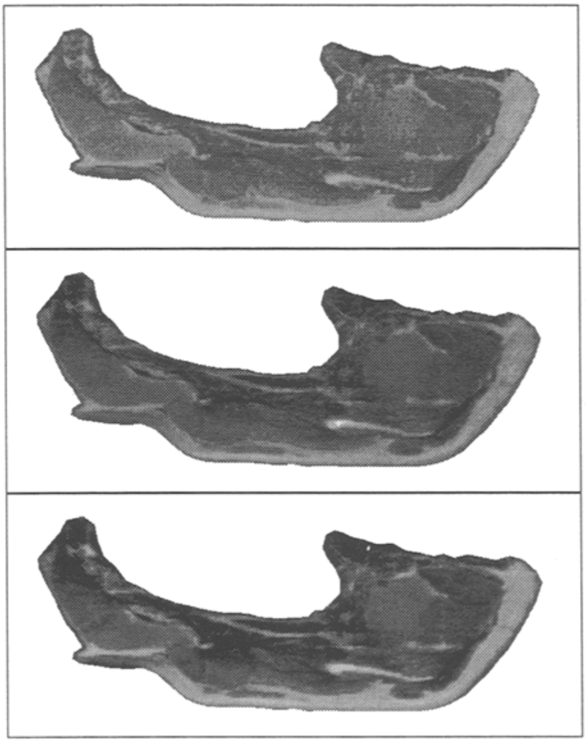

**Fig. 7.** The red, green and blue band - top to bottom - after bimodal histogram transformation.

The result of a bimodal histogram transformation is shown in Fig. 6 as histograms and 7 as images. We have used the histograms in the range 0-64 in Fig. 5 and, as described in Section 3, transformed them by changing the mean values and keeping the variances and alpha. In this case the resulting image has become darker, and the classification between lean and fat will be improved.

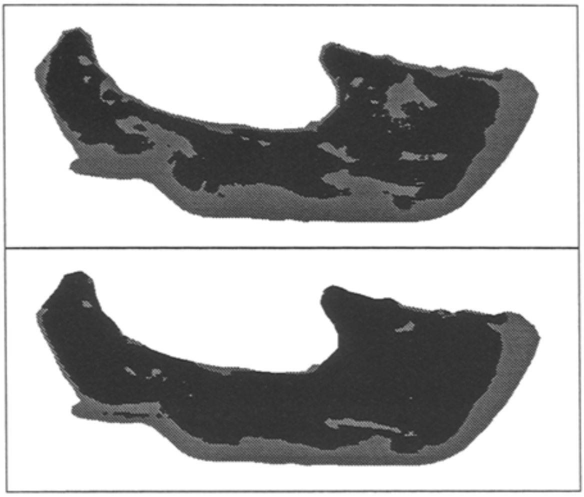

**Fig. 8.** The meat image classified into lean - black - and fat - grey - without and with bimodal histogram transformation - top to bottom.

In Fig. 8 a pixel-wise classification of the meat from Fig. 2 without - the top - and with - the bottom - bimodal histogram transformation is shown. It is seen that without the histogram transformation preprocessing, many lean pixels are misclassified into fat. Especially on the upper boundary of the meat, and inside in the lean area.

# 5 Conclusion

We have described a bimodal histogram transformation procedure where conjugate gradient is used for estimating the maximum likelihood parameters.
The bimodal histogram transformation have been used as a preprocessing step in a classification case study, where rgb color images of meat were segmented into lean and fat. The histogram transformation was done in order to gain some kind of standardization, which can be needed when supervised models are used for classification.
The bimodal histogram transformation seems to improve the classification, when the image distributions can vary dependent of the environment and image acquisitions. When the function of image intensity changes is unknown, the bimodal histogram transformation is a good alternative, because the histogram's parameters are not completely changed.

# References

1. N. Schultz. *Segmentation and classification of biological objects.* PhD thesis, Institute of Mathematical Modelling , Technical University of Denmark, Lyngby, 1995. 194 pp.
2. T. Taxt, N.L. Hjort, and L. Eikvil. Statistical classification using a linear mixture of two multinormal probability densities. *Pattern Recognition Letters*, 12(12):731–737, 1991.
3. D.M. Titterington, A.F.M. Smith, and U.E. Makov. *Statistical analysis of finite mixture distributions.* J. Wiley & Sons, Great Britain, 1985. 243 pp.
4. C. Vind. A comparison of neural networks and statistical methods for classification. Master's thesis, Institute of Computer Science, Copenhagen University and the Danish Meat Research Institute, University of Copenhagen, Denmark, 1994.