

# Soft Histograms for Belief Propagation

Erik Jonsson and Michael Felsberg

Computer Vision Laboratory  
Dept. of Electrical Engineering, Linköping University  
erijs@isy.liu.se, mfe@isy.liu.se

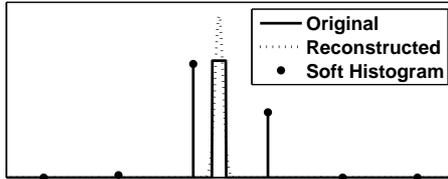
**Abstract.** Belief propagation methods are powerful tools for various problems in computer vision. While most methods assume a discrete set of labels for each node in the graphical model, there has recently been an increased interest in using real-valued labels and continuous probability density functions for such problems. We propose using channel representations (soft histograms) as a new way of moving from discrete to real-valued labels. The soft histograms are related to continuous density functions through the maximum entropy principle, and a complete soft histogram-based belief propagation method is evaluated and compared to hard discretization methods on simulated and real data.

## 1 Introduction

Bayesian networks are a powerful modeling tool for many problems in artificial vision and other areas. Two important special cases of Bayesian networks are Hidden Markov Models (HMM) and Markov Random Fields (MRF). HMMs are the underlying model for most tracking algorithms, and the Kalman filter is derived from the special case of HMMs where the probabilities are all Gaussian. Markov Random Fields have been successfully used for image restoration [1], segmentation [11] and interpretation [2].

In the Kalman filtering case, there is an analytical solution available, which makes it possible to use continuous state variables in a simple way, but for more complicated graphical structures or non-Gaussian models, there is no such simple solution. Most work on general Bayesian networks is performed in a discrete setting, where each node in the probabilistic structure can attain one out of several discrete labels. However, in many application areas including image processing, the underlying signals are continuous in nature, and the amount of discretization introduced is rather arbitrary. Recently, there has been an increased interest in representing real-valued labels and continuous density functions in belief propagation methods using e.g. ideas from particle filtering [8] or mixtures of Gaussians [16].

This work examines another way of dealing with real-valued labels. The basic idea is to replace hard discretizations with *soft histograms*, which use smooth, overlapping basis functions instead of rectangular bins. When this type of representation is used to represent single values, it is referred to as a *channel representation*, and there is some biological motivation for such representations [15]



**Fig. 1.** A continuous PDF measured using soft bins and reconstructed.

related to the concept of *ensemble coding* and *population vectors* [5], [4]. A similar type of representation is found in the visual cortex, where different neurons are selective for e.g. different local orientations of visual stimuli, but where the response patterns are overlapping, such that no single neuron responds uniquely for one orientation. In [6], channel representations are used for machine learning, and in [3] for denoising of image features. In [10], it was shown how to reconstruct continuous probability density functions (PDFs) from channel vectors using the maximum entropy method, which strengthens the theoretical basis for using channel vectors as a representation of PDFs.

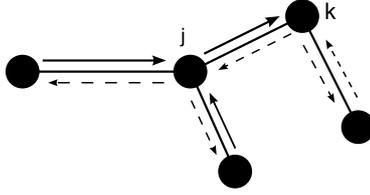
Using soft histograms, peaks in the PDF can be represented with an accuracy higher than the bin spacing (known as *hyperacuity* in the perceptual sciences). This is illustrated in Fig. 1, where a PDF is encoded into a soft histogram and then reconstructed. Since the complexity of discrete message passing is quadratic in the number of labels (or bins in the histogram), it would be of great benefit if the state space resolution could be reduced without impairing the accuracy. As an example, if the resolution can be reduced with a factor 4 in each dimension of a 3D state space, the total number of bins required is reduced by a factor  $4^3 = 64$  and the complexity by a factor  $64^2 = 4096$ . The main purpose of this paper is not to present a ready-to-use algorithm, but rather to introduce important principles and show the potential of using soft histograms for belief propagation.

The paper is organized as follows: First, a brief review of the belief propagation algorithm is given, mainly to introduce our notation. In Sect. 3 the *maximum entropy* principle is introduced and related to soft histograms. In Sect. 4, it is shown how to perform all operations on the soft histograms required for the belief propagation algorithm. Finally, in Sect. 5 the method is evaluated experimentally.

## 2 Belief Propagation

The most general formulation of belief propagation is excellently described in [12]. Here, we briefly present a less general version of the algorithm, suitable for the scope of this paper.

A Bayesian Decision Network (BDN) is a graph  $G$  where each node is a random variable that is dependent of other nodes in the graph only through its neighbors (the Markov property). Usually, the values of some nodes are known,



**Fig. 2.** A Bayesian network. The solid arrows illustrates one step in the belief propagation algorithm: two incoming messages are combined to give an outgoing message. After the execution of the entire algorithm, messages between all connected pairs of nodes have been evaluated, in both directions.

and the rest are to be found. Let  $\mathcal{N}$  be the set of node connections, such that  $(i, j) \in \mathcal{N}$  if there is an edge between node  $i$  and  $j$ . Let  $\mathcal{N}(i)$  denote the set of neighbors to node  $i$ . If the maximal *clique* size is 2, e.g. if  $G$  is a tree, the probability of a certain node labeling  $\mathbf{x}$  can be factorized to

$$p(\mathbf{x}) = \prod_{(i,j) \in \mathcal{N}} \psi_{i,j}(x_i, x_j) , \quad (1)$$

where  $\psi_{i,j}$  is a pairwise *compatibility function* of neighboring nodes  $i, j$ . These compatibility functions can be defined either from some real probabilistic knowledge of the problem or in a heuristic manner.

The *marginal posterior probability*  $p(x_i)$  of a single label  $x_i$  is obtained by inserting the known labels into (1) and integrating out all other labels  $x_j$  from  $p(\mathbf{x})$ . These marginals can be calculated efficiently when  $G$  is a tree through the well-known belief propagation algorithm. Here, a message  $m_{i \rightarrow j}(x_j)$  represents the belief from node  $i$  about the label of a neighboring node  $j$ , with consideration also of all nodes behind node  $i$ . Each message is a probability distribution on the set of labels, in our case a continuous probability density. Messages are recursively evaluated as

$$m_{j \rightarrow k}(x_k) = \int \psi_{j,k}(x_j, x_k) \tilde{p}(x_j) dx_j , \quad (2)$$

where

$$\tilde{p}(x_j) = \prod_{i \in \mathcal{N}(j) \setminus \{k\}} m_{i \rightarrow j}(x_j) \quad (3)$$

is the aggregated incoming message to node  $j$  (see Fig. 2). Messages are propagated in all directions in the tree, starting at the leaves, and the marginal posterior of a given node  $j$  is the product of all incoming messages from all directions:

$$p(x_j) = \prod_{i \in \mathcal{N}(j)} m_{i \rightarrow j}(x_j) . \quad (4)$$

### 3 PDF Representations and Maximum Entropy

During the execution of the belief propagation algorithm, messages need to be multiplied in (3) and transformed through  $\psi$  in (2). In the classical, discrete case, messages are vectors which can be multiplied element-wise, and (2) gets replaced by a matrix product. In the continuous case, things get more tricky. Since we cannot represent arbitrary continuous functions, we must restrict ourselves to some subset of continuous PDFs which is representable by a finite number of parameters. In this section, the maximum entropy principle is used as a way of relating a finite set of measurements to a continuous PDF. This knowledge will then be used to understand what (2) and (3) become when using the soft histogram representation.

#### 3.1 Maximum Entropy Principle

Let  $\{f_n\}$  be a set of arbitrary real-valued functions, and assume that we know a PDF only through a finite number of measurements

$$c_n = \frac{1}{T} \sum_{t=1}^T f_n(x_t) \approx \int f_n(x)p(x)dx = \langle f_n, p \rangle . \quad (5)$$

We now want to reconstruct a continuous density function from these  $c_n$ 's. According to the *maximum entropy method* (MEM), we should choose the *most uniform* distribution, i.e. the one with the highest entropy. The resulting continuous reconstruction from (5) is given by

$$\max H(p) \quad \text{subject to } \langle f_n, p \rangle = c_n, \forall n , \quad (6)$$

where

$$H(p) = - \int p(x) \ln p(x) dx \quad (7)$$

is the (differential) entropy of  $p(x)$ . A well-known result [7] gives that the optimal  $p(x)$  belongs to the *exponential family*

$$p(x) \propto \exp \left( \sum_n \lambda_n f_n(x) \right) , \quad (8)$$

where  $\lambda_n$  are some parameters to be determined. The vector  $\boldsymbol{\lambda} = [\lambda_1, \dots]$  is sometimes referred to as the *natural* or *exponential parameters*, the vector  $\mathbf{c} = [c_1, \dots]$  as a *mean parameter vector*, and the functions  $f_n(x)$  as *natural statistics* [18]. There is a one-to-one mapping between the vectors  $\mathbf{c}$  and  $\boldsymbol{\lambda}$ ; these vectors can be viewed as dual representations of the PDF. Note that two PDFs in the same exponential family can be multiplied simply by summing their exponential parameters, and that the product remains in the same family.

### 3.2 Gaussians

Assume that all we know about  $p(x)$  is the first two moments  $c_1 = \int xp(x)dx$  and  $c_2 = \int x^2p(x)dx$ . Using the maximum entropy principle, we get  $p(x) \propto \exp(\lambda_1x + \lambda_2x^2)$ , i.e. a Gaussian distribution. This famous result gives a theoretical motivation for approximating a PDF with a Gaussian when only the mean and variance is known. In this case, it is possible to express the relationship between  $\mathbf{c}$  and  $\boldsymbol{\lambda}$  analytically, and it is easy to verify that  $\lambda_1 = c_1/(c_2 - c_1^2)$  and  $\lambda_2 = -1/(2(c_2 - c_1^2))$ . Two Gaussians represented with a mean and a variance can now be multiplied by computing the  $\lambda$ -coefficients, adding them and switching back to  $c_1$  and  $c_2$ .

### 3.3 Hard Histograms

Assume that we have measured a hard histogram of  $p(x)$ , i.e. our  $\mathbf{c}$  vector consists of<sup>1</sup>

$$c_n = \int 1(|x - \xi_n| < w/2)p(x)dx = \int_{\xi_n-w/2}^{\xi_n+w/2} p(x)dx, \quad \forall n, \quad (9)$$

where  $\xi_n$  are the bin centers. The MEM choice of  $p(x)$  is now of the form  $p(x) = \exp(\sum_i \lambda_i 1(|x - \xi_i| < w/2))$ , i.e. we have a piecewise constant expression in the exponent, which makes the entire  $p(x)$  piecewise constant. But then

$$c_n = \int_{\xi_n-w/2}^{\xi_n+w/2} \exp(\lambda_n)dx, \quad (10)$$

so we must have  $\lambda_n = \ln c_n/w$ . In this case,  $p(x)$  can truly be treated as a discrete distribution, and two PDFs can be added and multiplied just by adding and multiplying the  $c_n$  coefficients.

### 3.4 Soft Histograms

Now consider what happens if we create a histogram where the bins are no longer rectangular. Let

$$c_n = \sum_{t=1}^T B_n(x_t) \approx \int B_n(x)p(x)dx, \quad (11)$$

where  $B_n(x) = B_0(x - \xi_n)$  is some smooth, local function centered at  $\xi_n$ . This produces a soft histogram (or a *channel vector*, where each  $c_n$  is a *channel coefficient*). In this work, the 2<sup>nd</sup> order B-spline kernel [17] will be used. This is a continuous piecewise quadratic function with compact support, and the centers  $\xi_n$  are located such that exactly three  $B_n$ -functions are non-zero for each  $x$ . From a channel vector  $\mathbf{c}$ , the MEM choice of  $p(x)$  is [10]

$$p(x) \propto \exp\left(\sum_n \lambda_n B(x - \xi_n)\right). \quad (12)$$

<sup>1</sup> If  $P(x)$  is a logical proposition depending on  $x$ ,  $1(P(x))$  is the function that is 1 when  $P(x)$  is true and 0 otherwise

Unfortunately, there is no simple closed-form solution for finding the vector  $\lambda$  from a channel vector  $\mathbf{c}$ . In [10], this transformation was done iteratively using a Newton method, which converges in around 10 – 20 iterations. Furthermore, in the hard histogram case, we could do additions and multiplications directly on the histogram  $\mathbf{c}$ , but for soft histograms, we need to explicitly compute and sum the exponential parameters in order to multiply two PDFs.

Reconsider (12). Since  $B(x)$  is a piecewise quadratic function, so is the entire exponent of (12). This means that  $p(x)$  is actually a *piecewise Gaussian* function. This is an interesting and perhaps surprising observation. But since we know that global quadratic measurements result in a Gaussian PDF, and that the B-splines make local quadratic measurements of the PDF, it is natural that we end up with a locally Gaussian-looking function. In contrast, the popular Gaussian Mixture Models (GMMs) are not piecewise Gaussian, since the sum of two Gaussians is not again a Gaussian.

The soft histogram technique described here is related to kernel density estimation (KDE), but in classical KDE there is never an attempt to reconstruct a density which is consistent to some measurements. Instead, an over-smoothed PDF is accepted as the final estimate.

## 4 The Integral Operator

As a step in the belief propagation algorithm, we need to propagate the aggregated messages at a current node through  $\psi$  to obtain the new outgoing message:

$$m(x_{\text{out}}) = \int \psi(x_{\text{in}}, x_{\text{out}}) \tilde{p}_{\text{in}}(x_{\text{in}}) dx_{\text{in}} . \quad (13)$$

This is essentially (2), but with some indices dropped to simplify the notation. In general the resulting  $m(x_{\text{out}})$  is not in exponential family of  $\tilde{p}_{\text{in}}(x_{\text{in}})$ . In order to stay within this family, we let the output message be  $\tilde{m}(x_{\text{out}})$  such that  $\langle m, B_i \rangle = \langle \tilde{m}, B_i \rangle, \forall i$ , i.e.  $m$  and  $\tilde{m}$  have the same soft histogram. This choice of  $\tilde{m}$  minimizes the KL-divergence  $D(m||\tilde{m})$  subject to  $\tilde{m}$  being in the desired exponential family [14].

In order to avoid double subscripts, we let vector indexing be denoted by  $[\cdot]$ . Each coefficient  $c_{\text{out}}[n]$  of the soft histogram  $\mathbf{c}_{\text{out}}$  representing the outgoing message  $\tilde{m}(x_{\text{out}})$  is then

$$c_{\text{out}}[n] = \langle B_n, m \rangle = \int B_n(x_{\text{out}}) \left( \int \psi(x_{\text{in}}, x_{\text{out}}) \tilde{p}_{\text{in}}(x_{\text{in}}) dx_{\text{in}} \right) dx_{\text{out}} = \quad (14)$$

$$= \int \left( \int B_n(x_{\text{out}}) \psi(x_{\text{in}}, x_{\text{out}}) dx_{\text{out}} \right) \tilde{p}_{\text{in}}(x_{\text{in}}) dx_{\text{in}} . \quad (15)$$

By defining

$$q_n(x_{\text{in}}) = \int B_n(x_{\text{out}}) \psi(x_{\text{in}}, x_{\text{out}}) dx_{\text{out}} , \quad (16)$$

we get the simple relationship

$$c_{\text{out}}[n] = \langle q_n, \tilde{p}_{\text{in}} \rangle . \quad (17)$$

We see that it is enough to consider a finite number of functions  $q_n$  that measure the contribution to each output channel from the input density. In order to represent  $\psi$  efficiently, we can restrict ourselves to functions  $q_n(x_{\text{in}})$  that can be expressed using a finite number of parameters. The goal is to find the equivalent of the matrix representing  $\psi(x_{\text{in}}, x_{\text{out}})$  in the discrete case. The following two choices fit well with the soft histogram representation:

#### 4.1 Linear $q$ representation

By letting

$$q_n(x_{\text{in}}) = \sum_{\nu} a_{n,\nu} B_{\nu}(x_{\text{in}}) , \quad (18)$$

(17) becomes

$$c_{\text{out}}[n] = \sum_{\nu} a_{n,\nu} \langle B_{\nu}, \tilde{p}_{\text{in}} \rangle = \sum_{\nu} a_{n,\nu} c_{\text{in}}[\nu] , \quad (19)$$

and the entire operation can be expressed as a linear operation directly on the channel vectors, i.e. we have

$$\mathbf{c}_{\text{out}} = \mathbf{A} \mathbf{c}_{\text{in}} \quad (20)$$

This is analogous to the hard histogram case, and is intuitively appealing since the operation remains a linear mapping. An optimal  $\mathbf{A}$  could be found by creating training samples of input vectors  $\mathbf{c}_{\text{in}}$  and  $\mathbf{c}_{\text{out}}$  and using the least squares technique described in [9].

#### 4.2 Exponential $q$ representation

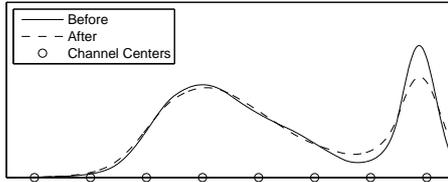
The second approach is to consider functions  $q_n(x_{\text{out}})$  that can be expressed in the same exponential form as  $\tilde{p}_{\text{in}}$ , i.e.:

$$q_n(x_{\text{in}}) = \exp \left( \sum_{\nu} a_{n,\nu} B_{\nu}(x_{\text{in}}) \right) . \quad (21)$$

Now, the scalar product (17) can be computed by adding the exponential parameters of  $\tilde{p}_{\text{in}}$  and  $q_n$  and integrating the corresponding continuous PDF, e.g. using some erf lookup-table operation. Each such integral operation gives one element of the output channel vector, and by repeating the process for each  $q_n$ , we get the entire soft histogram of  $m(x_{\text{out}})$ . The coefficients  $a_{n,\nu}$  can be organized in a matrix  $\mathbf{A}$  which summarizes the entire  $\psi$ .

In a preprocessing step, the functions  $q_n$  can be computed from  $\psi$  according to (16). The exponential parameter of each  $q_n$  can then be computed by finding the soft histogram of each  $q_n$  and do the “ $\mathbf{c}$  to  $\boldsymbol{\lambda}$  conversion”, i.e. projecting  $q_n$  into our exponential family.

In Fig. 3 an example is shown where a PDF is transformed through a smoothing kernel using this method. Note the accuracy compared to the channel centers.



**Fig. 3.** An example of a PDF represented as a soft histogram transformed through a smoothing integration kernel

### 4.3 Closedness of the integral operator

For a general linear mapping  $\mathbf{A}$  operating on a channel vector  $\mathbf{c}$ , there is no guarantee that the output  $\mathbf{c}_{\text{out}} = \mathbf{A}\mathbf{c}$  is a valid channel vector, i.e. a vector containing the mean parameters of some PDF. One example of an invalid vector is  $[0, 0, 1, 0, 0]$ . Since the channels are overlapping, even the most peaked PDF would produce non-zero entries in several bins. For the exponential representation above, there is also no theoretical guarantee of closedness, even though this has not been a problem in practice so far. Finding compact, expressive representations of  $\psi(x_{\text{in}}, x_{\text{out}})$  free of this problem definitely requires more attention.

## 5 Experiments

### 5.1 Hidden Markov Model

As a first experiment, a simple Hidden Markov Model was considered. A sequence of hidden states  $\{x_t\}$  following a Brownian motion was constructed, from which a sequence of observations  $\{\mathbf{y}_t\}$  was obtained. Formally, we have

$$x_t = x_{t-1} + v_t \quad (22)$$

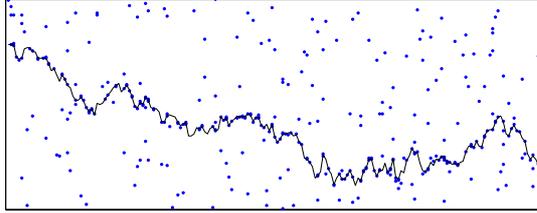
$$\mathbf{y}_t = g(x_t) \quad (23)$$

where  $v_t$  is an independent Gaussian noise term. The observation function  $g$  gives a random-length vector of measurements, where each element  $y_t[i]$  is distributed according to

$$p(y_t[i]) = k_1 + k_2 \exp[-(y_t[i] - x_t)^2 / \sigma^2] \quad , \quad (24)$$

i.e. each measurement is either a normally distributed inlier or an outlier. The goal is to find the MSE estimate of the hidden state sequence by first finding the marginal distribution of the hidden state at each time step. The MSE estimate is the expectation of each  $x_t$  from these marginals. The marginals are found using belief propagation in a forward and a backward step.

To have an “ideal” method to compare with, the state space was quantized to a fine grid of 200 levels, such that the standard discrete belief propagation algorithm on this fine grid gives the optimal solution. This is referred to as the



**Fig. 4.** Example of a true process together with the observations

“high resolution” method in the experiments. In addition to this, the state space was further quantized down to just 10 bins, using both hard and soft histograms. The true state sequence and observations of a typical run is shown in Fig. 4. The RMS errors for different number of bins for the hard and soft histogram are shown in Fig. 5 using both the linear and exponential representation of  $\psi$ . The result was averaged over 10 runs for each bin configuration. Since the exponential representation produced the best results, this method was selected for a more detailed examination.

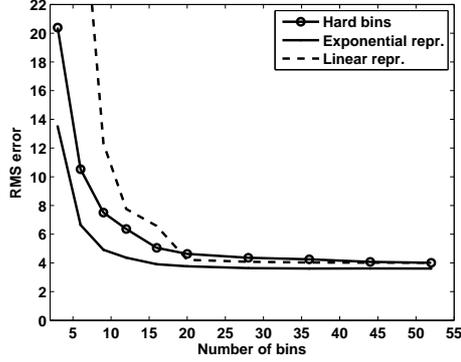
Figure 6 shows the posterior densities and the resulting MSE state sequence estimates of the same example. We clearly see the quantization caused by the hard bins. On the contrary, the soft histogram method is faithful to the measurements and gives nearly the same result as the high resolution method.

Figure 7 shows the posterior at two distinct time steps as 1D plots, to further visualize the representative power of soft histograms. In the left plot, the high-resolution marginal is a peak much more narrow than the bin spacing, which is represented well in the soft histogram. This plot is taken from a time  $t$  when the trajectory passes exactly through a measurement. In the right plot, there is more uncertainty about the exact position of the trajectory, visible as a small tail extending at the right side of the peak of the high-resolution posterior. The soft histogram is not able to represent this, and gives a more blurred PDF. However, the peak is still quite accurate compared to the hard histogram.

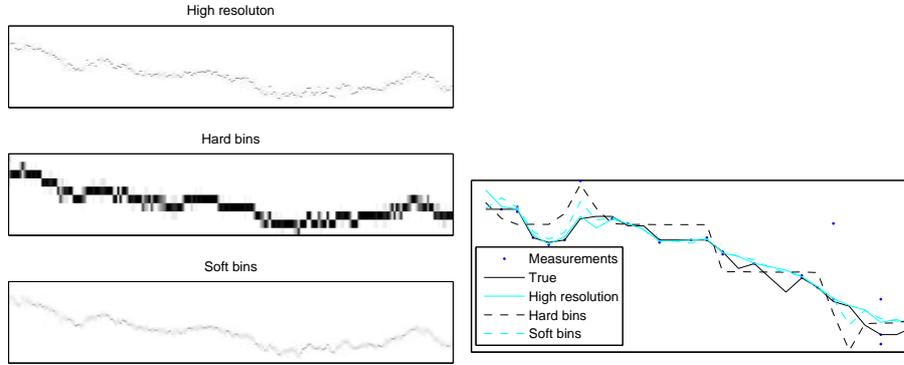
## 5.2 Boundary Detection

As a more realistic demonstration, the method was applied to boundary detection in ultrasonic imagery. The image is of a blood vessel, and the objective is to locate the boundary between two different types of tissue. In [13], this problem was solved using dynamic programming, which can be viewed as a special case of the Viterbi algorithm for MAP estimation in a hidden markov model. Each column of the image is viewed as an observation, and the position of the boundary is the sought hidden variable.

In absence of a true statistical model, and in order to keep things simple, a heuristic model was constructed from the sobel filter response, such that the probability of boundary position  $y_x$  at column  $x$  given the observed image column



**Fig. 5.** The RMS error for different number of bins, averaged over 10 runs



**Fig. 6.** Left: The posterior marginal densities produced from the three methods. Right: Estimated state trajectories for the three compared methods, zoomed for better visibility.

$\text{obs}_x$  is

$$p(y_x | \text{obs}_x) = k_1 \exp[-k_2 I_y(x, y_x)] \quad , \quad (25)$$

where  $I_y$  is the vertical component of the image gradient. Two adjacent boundary positions are related by

$$\psi(y_x, y_{x+1}) = k_3 \exp[-k_4 (y_x - y_{x+1})^2] \quad (26)$$

The constants  $k_i$  were selected manually for good results.

The space of boundary positions was quantized to 16 levels, and the marginals were calculated using both hard discretization and using soft histograms with the exponential representation of  $\psi$ . Each marginal was then maximized to produce a boundary position for each image column. The qualitative results are shown in Fig. 8. The bin centers are displayed to the left. The superior resolution produced by the soft histograms can clearly be seen, even though the result is not perfect.

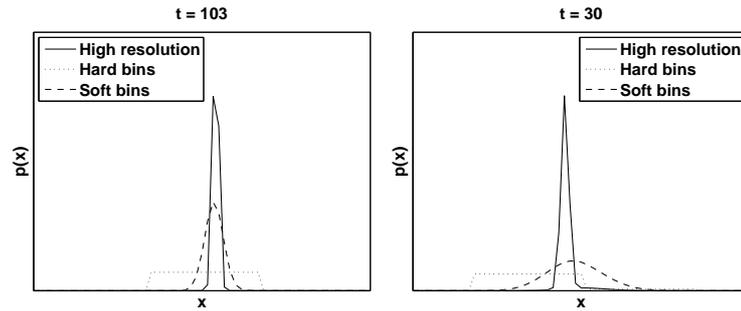


Fig. 7. Zoomed 1D plot of the marginals at two distinct time steps

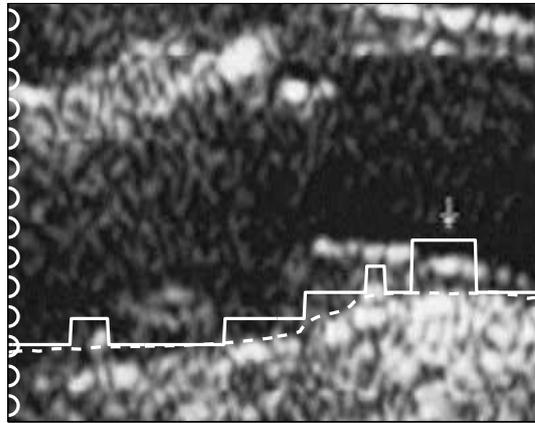


Fig. 8. Detection results in an ultrasonic image using both hard quantization and soft histograms

## 6 Conclusions

Using soft histograms instead of hard discretizations is intuitively appealing. The maximum entropy principle provides a natural framework in which to link the histograms to continuous density functions. Within this framework, hard discretizations, Gaussian representations and soft histograms are just three different instances of the same underlying principle. The equivalents of multiplying two PDFs and transforming a PDF through an integration kernel have been examined, which are the two key operations of belief propagation methods. Furthermore, the neurological basis of the channel representation makes this combination a plausible candidate for how information is represented, transported and integrated in the human visual system.

There is however still some open questions regarding the representation of the integral operator, as discussed in Sect. 4, and the computational complexity

of the conversion between  $\mathbf{c}$  and  $\boldsymbol{\lambda}$  is prohibitive. In the current implementation, the conversion is performed using an iterative method, where each step involves calculating integrals over Gaussians. For low state-space dimensionalities, this can be done using lookup-table techniques, but for higher dimensionalities, there is currently no satisfying solution. For some applications, like grayscale image reconstruction and stereo, one-dimensional state spaces are sufficient. However, it is the belief of the authors that it is possible to find some approximation that enables this technique also for higher dimensionalities. Finding good approximations is a subject for future research.

## Acknowledgements

This work has been supported by EC Grant IST-2003-004176 COSPAL. This paper does not represent the opinion of the European Community, and the European Community is not responsible for any use which may be made of its contents. The ultrasonic image was kindly provided by Prof. Tomas Gustavsson, S2, Chalmers Univ. of Technology.

## References

1. Julian Besag. On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society*, 48(3):259–302, 1986.
2. A. El-Yacoubi, M. Gilloux, R. Sabourin, and C.y. Suen. An hmm-based approach for off-line unconstrained handwritten word modeling and recognition. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 21(8):752–760, August 1999.
3. M. Felsberg, P.-E. Forssén, and H. Scharr. Channel smoothing: Efficient robust smoothing of low-level signal features. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(2):209–222, February 2006.
4. Michael S. Gazzaniga, R. B. Ivry, and G. R. Mangun. *Cognitive Neuroscience*. W. W. Norton & Company, second edition, 2002.
5. A. P. Georgopoulos, A. B. Schwartz, and R. E. Kettner. Neuronal population coding of movement direction. *Science*, 233:14161419, 1986.
6. G. H. Granlund. An associative perception-action structure using a localized space variant information representation. In *Proceedings of Algebraic Frames for the Perception-Action Cycle (AFPAC)*, Kiel, Germany, September 2000.
7. Simon Haykin. *Neural Networks, A Comprehensive Foundation*. Prentice Hall, second edition, 1999.
8. M. Isard. PAMPAS: Real-valued graphical models for computer vision. In *Proc. Conf. Computer Vision and Pattern Recognition*, 2003.
9. Björn Johansson. *Low Level Operations and Learning in Computer Vision*. PhD thesis, Linköping University, Sweden, SE-581 83 Linköping, Sweden, December 2004. Dissertation No. 912, ISBN 91-85295-93-0.
10. E. Jonsson and M. Felsberg. Reconstruction of probability density functions from channel representations. In *Proc. 14th Scandinavian Conference on Image Analysis*, 2005.
11. Santhana Krishnamachari and Rama Chellappa. Multiresolution gaussmarkov random field models for texture segmentation. *IEEE Trans. Image Processing*, 6(2), February 1997.

12. Frank R. Kschischang, Brendan J. Frey, and Hans-Andrea Loeliger. Factor graphs and the sum-product algorithm. *IEEE Transactions on Information Theory*, 47(2), February 2001.
13. Q. Liang, I. Wendelhag, J. Wikstrand, and T. Gustavsson. A multiscale dynamic programming procedure for boundary detection in ultrasonic artery images. *IEEE Transactions on medical imaging*, 19(2):127–142, February 2000.
14. Thomas Minka. Expectation propagation for approximate bayesian inference. In *Proc. 17th Conf. in Uncertainty in Artificial Intelligence*, pages 362–369, 2001.
15. Herman P. Snippe and Jan J. Koenderink. Discrimination thresholds for channel-coded systems. *Biological Cybernetics*, 66:543–551, 1992.
16. E.B. Sudderth, A.T. Ihler, W.T. Freeman, and A.S. Willsky. Nonparametric belief propagation. In *Proc. Conf. Computer Vision and Pattern Recognition*, 2003.
17. Michael Unser. Splines: A perfect fit for signal and image processing. *IEEE Signal Processing Magazine*, 16(6):22–38, November 1999.
18. John Winn and Christopher M. Bishop. Variational message passing. *Journal of Machine Learning Research*, 6:661–694, 2005.