Applications of Hidden Markov Chains in Image Analysis

Kjersti Aas   Line Eikvil   Ragnar Bang Huseby

Department of Pattern Recognition and Image Processing,
Norwegian Computing Center, P.O. Box 114 Blindern, N-0314 Oslo, Norway

e-mail: Kjersti.Aas@nr.no   Line.Eikvil@nr.no   Ragnar.Huseby@nr.no
Tel: (+47) 22 85 25 00   Fax: (+47) 22 69 76 60

Kjersti Aas received the M.Sc. degree in industrial mathematics from the Department of Industrial Mathematics, the Norwegian Institute of Technology, Norway 1990. She is currently working at the Norwegian Computing Center, Oslo, Norway. Her research interests include statistical pattern recognition, image analysis, industrial inspection, and stochastic modelling.

Line Eikvil received her M.Sc. degree from the Department of Informatics at the University of Oslo in 1988. She is currently working as a researcher at the Norwegian Computing Center and has been involved in several projects concerning document image analysis and machine vision. Her research interests lie mainly within the various applications of statistical pattern recognition.

Ragnar Bang Huseby received the M.Sc. degree in Mathematics from the Department of Mathematics at the University of Oslo in 1990. He is currently working at the Norwegian Computing Center, Oslo, Norway. His research interests include image analysis, pattern recognition, and mathematical modelling.

Applications of Hidden Markov Chains in Image Analysis

Summary

In this paper, we demonstrate that hidden Markov chains have a potential for use in different image analysis problems. Four examples are presented where the data source varies from a 1-band image to a live video sequence, and the applications range from machine vision to video surveillance.

The first example is fetched from text recognition. Here we use hidden Markov chain models for each character to obtain simultaneous segmentation and recognition of printed words from a grey level image.

The second example is a machine vision problem, where the issue is to recognize return bottles in crates imaged from above. A laser range scanner is used to aquire the images and the hidden Markov chain models are used to describe scanlines through the bottles.

In the remaining two examples, we have a sequence of images. In the first of these examples, the problem is to classify tumors as malignant or benign based on a dynamic Magnetic Resonance Image (MRI) sequence. The hidden Markov chain model is here used to describe the evolution of the image intensity with time after injection of contrast agent in the case of a malignant tumor.

The last example is a live video sequence, where the video images are used for traffic surveillance. The problem is to detect (and count) vehicles passing the surveillance site. Two detectors are used for this task, and the course of events reported when a vehicle is passing gives a sequence of two-dimensional feature vectors that is modelled as a hidden Markov chain.

Even though images are 2-dimensional, we have shown that the problem which is to be solved often is of a 1-dimensional character. Hence, hidden Markov chains may be applicable and their several strong points also can be exploited in image analysis.

Applications of Hidden Markov Chains in Image Analysis

Abstract

In image analysis, 2-dimensional Markov models, i.e. Markov field models, have been applied for segmentation purposes, but except for the area of text recognition, the application of hidden Markov chains has been rare. Through four very different examples, this paper demonstrates the applicability also for hidden Markov chains in image analysis, and shows that the problems of image analysis often may have 1-dimensional characteristics even though the images are 2-dimensional.

Keywords: hidden Markov chain models, image analysis, applications.

1 Introduction

The Markov models that have been used in image analysis, have traditionally been 2-dimensional, i.e. the so-called Markov random field (MRF) models. The Markov random fields model the local properties of image regions. Hence they are appropriate for image segmentation, for which context is important, and have typically been used for different segmentation (also called restoration) problems in remote sensing, machine vision, medical imaging, and document image analysis. An obvious drawback with methods based on Markov random field models is that they can only be used for small neighbourhoods in the image, due to the computational complexity and the modelling problems posed by large neighbourhoods.

The use of hidden Markov chain (HMC) models for recognition of structures in time series has been popular for a long time. The basic theory was published in a series of classic papers by Baum et al. in the late 1960s and early 1970s and it was implemented for speech processing applications by Baker at CMU and Jelinek et al. at IBM in the 1970s. The reader is referred to Rabiner [1] for an excellent tutorial on the subject.

Speech and text recognition are related research areas, and recently we have seen that hidden

Markov chain models have been introduced to the area of character recognition. In this paper we will show that such models are applicable for several different image analysis problems. The advantage of HMC models compared to the MRF models is that one by considering 1-dimensional neighbourhoods is able to combine the simplicity of local modelling with the strength of global dependence. Although the data source is a 2-dimensional image, the problem to be solved may often be of a 1-dimensional character and the features may be ordered in a time series.

This paper will give examples of applications of hidden Markov chain models within four areas of image analysis; document image analysis, machine vision, medical imaging, and video analysis. The problem to be solved is a different one for each example, and the data source varies from a 1-band image to a live video sequence. In all the examples the original image data are first reduced to a (possible multivariate) time series.

The paper is organized as follows. In Section 2 we will give the necessary theoretical background for hidden Markov chain models. In Section 3 the different examples of applications of HMC models for image analysis are presented, and the classification problems and the choice of models involved in the examples are described. The algorithms used for solving the problems of Section 3 will be discussed in Section 4, and in Section 5 we will present procedures for estimation of model parameters.

2 Hidden Markov Chain Model

A hidden Markov chain is a doubly stochastic process with an underlying stochastic process that is not observable (hidden), but can be observed through another set of stochastic processes that produce a sequence of observations [1]. We assume that observations are measured at regularly spaced discrete points in time; $t = 1, ..., n$. We will denote the sequence of observations by $Y = Y_1, ..., Y_n$ (Y_t may be multivariate) and the sequence of states in the hidden process by $C = C_1, ..., C_n$. (In the following, capital letters will be used for stochastic variables while small letters will be used for actual values.)

Although the sequence of states can not be observed directly, we have some prior information on how it is likely to behave, and this information is used in the design of the models. The model may be divided into two parts:

- A model for the state process.
- A model for the observations conditioned on the state process.

In Sections 2.1 and 2.2 we will discuss different choices for the two models.

2.1 Models for the state process

A full probabilistic description of the state process, in general, requires specification of the current state (at time t), as well as all the predecessor states, i.e.,

$$\pi(C) = \prod_{t=1}^{n} P(C_t = j | C_1 = c_1, ..., C_{t-1} = i) \tag{1}$$

Ordinary Markov model. For a first order hidden Markov chain, the probabilistic description in Equation 1 is truncated to just the current and the previous state, i.e.,

$$\pi(C) = \prod_{t=1}^{n} P(C_t = j | C_{t-1} = i) \tag{2}$$

Furthermore, if one only those processes for which the right-hand side of Equation 2 is independent of time are considered, we get a set of state transition probabilities a_{ij} of the form

$$a_{ij} = P(C_t = j | C_{t-1} = i) \tag{3}$$

with the state transition probabilities having the properties

$$a_{ij} \geq 0 \tag{4}$$

$$\sum_{j=1}^{K} a_{ij} = 1 \tag{5}$$

since they obey stochastic constraints. Here, K is the number of states of the model.

The implicit state duration of an ordinary Markov model will be geometric as the probability of duration d in state j is proportional to $(a_{jj})^d$. This means that the probability of remaining in a state will decrease exponentially with duration d (the time spent in the state).

For many applications it is not very likely that the probability of remaining in a state will decrease exponentially, meaning that the implicit geometric state duration will be inappropriate. However, short state durations may be compensated for by defining a larger number of states. An extension of the model to more general definitions of state duration is also possible, leading to so-called semi Markov models.

Semi Markov model. As stated in the previous paragraph, the perhaps major weakness of conventional hidden Markov models is the modelling of state duration. In many cases one would prefer to explicitly model duration density in some analytic form, so that a transition from one state to another is made only after an appropriate number of observations have occurred in the state (as specified by the duration density). For a semi Markov model the probabilistic description in Equation 1 is of the form

$$\pi(C) = \prod_{t=1}^{n} P(C_t = j | C_{t-1} = i, W_{t-1} = w) \tag{6}$$

where W_{t-1} is the time spent in the current state, i.e.,

$$W_{t-1} = \min\{w \geq 1 | C_{t-1-w} \neq C_{t-1}\}. \tag{7}$$

For the special case when $C_1 = C_2 = ... = C_{t-1}$, the waiting time W_{t-1} equals $t-1$. The probability $P(C_t = j|C_{t-1} = i, W_{t-1} = w)$ is on the form:

$$P(C_t = j|C_{t-1} = i, W_{t-1} = w) = \begin{cases} d_i(w) & \text{if} \quad j = i \\ (1 - d_i(w))a_{ij} & \text{if} \quad j \neq i \end{cases} \tag{8}$$

Here, $d_i(w)$ is the probability of staying in state i for (at least) one time unit longer, given that the start of the segment took place w units earlier, while a_{ij} is the probability of a transition from state i to state j, given that a transition has occurred.

Note, that by considering (C_t, W_t) as a multivariate process, this process will satisfy the Markov property. Hence, it can be interpreted as an ordinary Markov chain [2].

2.2 Models for the observations

When a new state is entered, an observation output symbol is produced according to a probability distribution, and the second part of the Markov model is a description of this distribution which is denoted $P(Y|C)$. In its most general form $P(Y|C)$ is given by

$$P(Y|C) = \prod_{t=1}^{n} P(Y_t|Y_1, ..., Y_{t-1}, C) \tag{9}$$

We consider two types of observation probability distributions, namely, discrete and continuous.

2.2.1 Discrete models.

In a discrete model, the observations are characterized as discrete symbols from a finite alphabet. Hence, a discrete observation symbol probability distribution is used within each state of the model. As opposed to the distributions described in the following section which requires parameters of the distribution to be specified, the discrete model used in this paper is characterized by the probabilities of symbols given a specific state.

2.2.2 Continuous models

For a continous hidden Markov model we generally assume that

$$Y_t = g(Y_1, ..., Y_{t-1}, C_t). \tag{10}$$

The expression at the right-hand side of Equation 10 may take many different forms dependent on the way the relationship between the state process and the observations is being modelled. In this paper we consider three different choices for this relationship; a constant model with Gaussian uncorrelated noise, a constant model with Gaussian correlated noise, and a linear model with Gaussian independent noise.

Constant model with uncorrelated Gaussian noise. In this model, the signal is assumed to be constant inside each segment, i.e.

$$Y_t = \mu + \varepsilon_t \tag{11}$$

where ε_t is assumed to be uncorrelated Gaussian noise with zero mean and variance σ_ε^2. Then, Y_t will be Gaussian distributed with mean:

$$E(Y_t) = \mu$$

and variance:

$$Var(Y_t) = \sigma_\varepsilon^2$$

Constant model with correlated Gaussian noise. Also in this model, the signal is assumed to be constant inside each segment, i.e.

$$Y_t = \mu + \varepsilon_t \tag{12}$$

but the noise ε_t is assumed to follow a random walk model, i.e.

$$\varepsilon_t = \varepsilon_{t-1} + \eta_t \tag{13}$$

where $\eta_t; t = 1, ..., n$ is a sequence of independent Gaussian variables with zero means and equal variances σ_η^2. Then, the probability distribution for Y_t is Gaussian with mean

$$E(Y_t) = y_{t-1} \tag{14}$$

and variance

$$Var(Y_t) = \sigma_\eta^2. \tag{15}$$

Linear model with uncorrelated Gaussian noise. This model is more complicated than the others in that the distribution of Y_t depends on W_t:

$$Y_t = Y_{t-W_t} + \beta W_t + \varepsilon_t.$$

In the special case of the linear model used in this paper, we assume that the slope β is Gaussian distributed with mean μ_β and variance σ_β^2. Moreover, we assume that $\varepsilon_t; t = 1, ..., n$ is a sequence of independent Gaussian variables with zero means and equal variances σ_ε^2. Then, the observation probability distribution for Y_t is Gaussian with mean

$$E(Y_t) = y_{t-w} + \mu_\beta w \tag{16}$$

and variance

$$Var(Y_t) = w^2 \sigma_\beta^2 + \sigma_\varepsilon^2. \tag{17}$$

3 Applications of hidden Markov chain models

In this section, we will describe four examples of image analysis problems where hidden Markov chain models have been applied successfully. The examples are fetched from four different application areas of image analysis; document image analysis, industrial inspection, medical imaging, and video analysis. The data source varies from a single 1-band image to a live video sequence. However, a common factor for the problems that these examples represent, is that the essential image data may be reduced to a time series.

The examples represent different phenomenons, where the data are acquired from different types of sources. For each of the four examples, a different problem should be solved. The problems can be summarized as finding

- the best matching state sequence,
- the best matching model,
- the probability of the presence of a specific state,
- the probability of the presence of specific state transitions,

given the observation sequence.

We will in the following describe the image type, the HMC model, and the problem to be solved for each application. A short summary of the problem characteristics is given in Table 1, while the model choices are summarized in Table 2.

	Text recognition	**Bottle recognition**	**Tumor classification**	**Vehicle detection**
Image origin	Scanned text page	Range scan	Dynamic MR	Video
Image type	1-band image	2-band image	finite sequence	live sequence
Size of feature vector	18	2	1	2
Number of classes	54	7	2	2

Table 1: *Problem characteristics.*

3.1 HMC models for document image analysis: Text recognition

The first applications of hidden Markov chain models for text recognition appeared around the beginning of this decade. Since then these methods have become quite popular within the field, and especially for problems like handwritten character recognition where segmentation of single characters is difficult, and it is better to consider a word as a whole. The example of HMC models for text recognition that we will present here, is fetched from a study presented in [7], where text recognition is performed on words from a grey level image.

The aim of our study was to investigate whether HMC models could be used for text recognition in such a way that the difficult problem of character segmentation was avoided. In OCR systems, the main problem is errors caused by poor segmentation resulting in connected or split characters. This will nearly always be a problem when segmentation and recognition are performed as two separate operations. Through the use of HMC models it is possible to jointly optimize recognition and segmentation in a single search procedure. An additional advantage is that such models give a nice framework for incorporation of contextual information like transition probabilities between characters and dictionary information. Hidden Markov chain models are therefore well suited for text recognition.

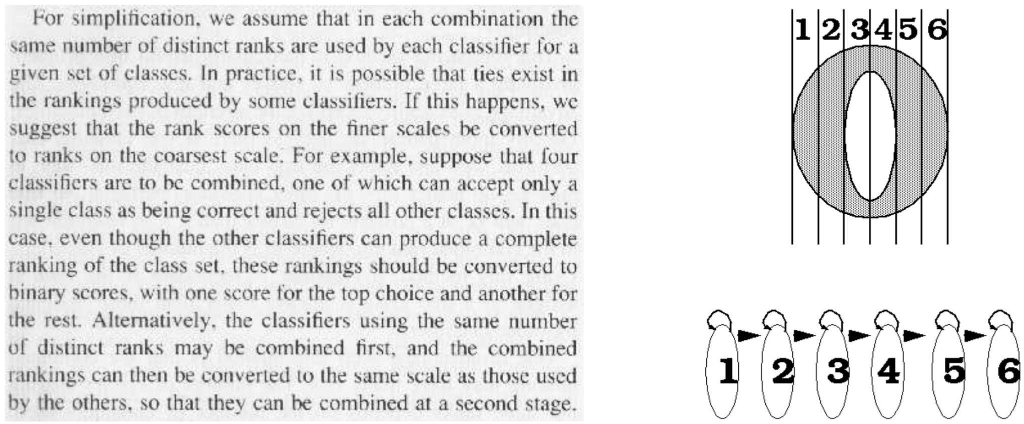
For simplification, we assume that in each combination the same number of distinct ranks are used by each classifier for a given set of classes. In practice, it is possible that ties exist in the rankings produced by some classifiers. If this happens, we suggest that the rank scores on the finer scales be converted to ranks on the coarsest scale. For example, suppose that four classifiers are to be combined, one of which can accept only a single class as being correct and rejects all other classes. In this case, even though the other classifiers can produce a complete ranking of the class set, these rankings should be converted to binary scores, with one score for the top choice and another for the rest. Alternatively, the classifiers using the same number of distinct ranks may be combined first, and the combined rankings can then be converted to the same scale as those used by the others, so that they can be combined at a second stage.

Figure 1: *Left: A scanned page of printed text. Upper right: The 6 states for the character "o". Lower right: The HMC model for a character.*

Features were extracted from the image as grey level columns of sub images corresponding to the circumscribing box of each word of the text. Each feature vector (column) had dimension 18, giving a multivariate time series of features from each word.

The characters were modelled as HMC models, with the states being segments of the character (See Figure 1). 54 different character classes were modelled (A-Z, a-z, . ,), and for all but the shortest characters, 6 states were used. This relatively large number of states was used to compensate for the implicit geometric waiting time of the Markov model. For each word, the sequence of feature vectors was matched against all possible combinations of states to find the sequence of states giving the best fit. Hence, we were looking at the problem of finding the optimal state sequence, given the observation sequence.

3.2 HMC models for industrial inspection: Bottle recognition

As previously mentioned, Markov random field models have been used for segmentation problems in industrial inspection and machine vision, but applications using Markov chains are very sparse within this area. We have, however, been working with one problem of this type, and we will here present this study which is fetched from [8, 9]. The task in this study was to segment and discriminate similar-looking bottles in a crate imaged from above.

The study was performed in cooperation with a company which produces different types of automatons for recirculation of bottles. Reliable recognition of bottles is important, since the amount reimbursed for a bottle depends on its type. Recognition of bottles in crates represents special problems, as the bottles can only be imaged from above. To get more information about the shape of the bottles, a laser range scanner is used for the image acquisition. This range scanner delivers both the distance to points in the scene and the strength of the reflected signal at the same point. The result is an image with two components as shown in Figure 2.

Each scan-line (in x-direction) of the 2-band image was treated as a bivariate time series. The x-direction was chosen for two reasons. First, the correlation between pixels is highest in this direction (due to the image acquisition process). Second, each scan-line is the result of one physical measurement, meaning that information about the scanner's behaviour could be utilized in the model. Each bivariate time series was modelled as a HMC. As for the text recognition problem, HMC models may be used for jointly optimizing recognition and segmentation in

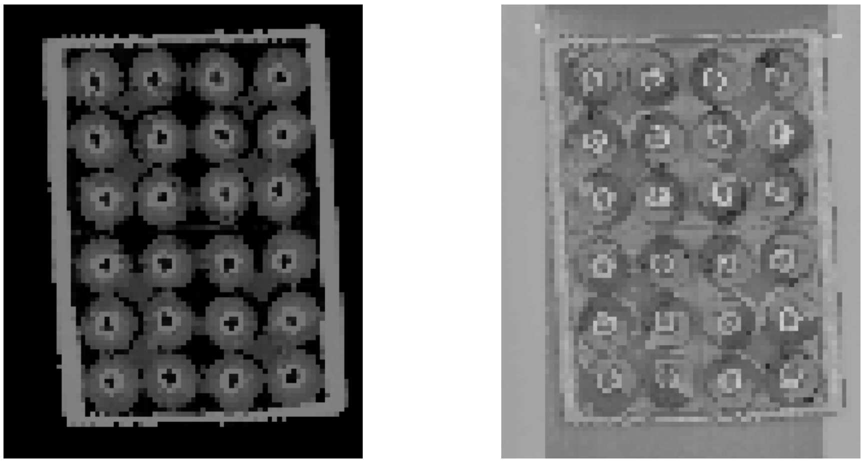

Figure 2: *Range (left) and intensity (right) images of a crate containing bottles.*

a single search procedure. However, for this application, the segmentation and recognition of bottles were treated as two separate problems. More information about the work on the segmentation problem can be found in [8], here we focus on the problem of classifying the bottles assuming that their location is known.

In [9] scan-lines going through the center of the bottle (like the ones in Figure 3, left) were extracted and each bottle type was modelled as a HMC model with 5 states, corresponding to five different parts of the bottle (Figure 3, right). The number of classes was 7, corresponding to the number of different bottle types in the test material, and the problem to be solved was to determine the bottle type that best fitted the observed sequence of feature vectors for the bottles in the image.

3.3 HMC models for medical imaging: Tumor classification

As for industrial inspection, Markov random fields have been used for segmentation of medical images, while applications using Markov chains are rare. Still, Markov chains can be useful also in this field, and we will in this section look at an example from a study presented in [11]. The aim of this study was to find automatic methods for characterizing tumors correctly as malignant or benign, based on a time series derived from a dynamic Magnetic Resonance Image (MRI) sequence.

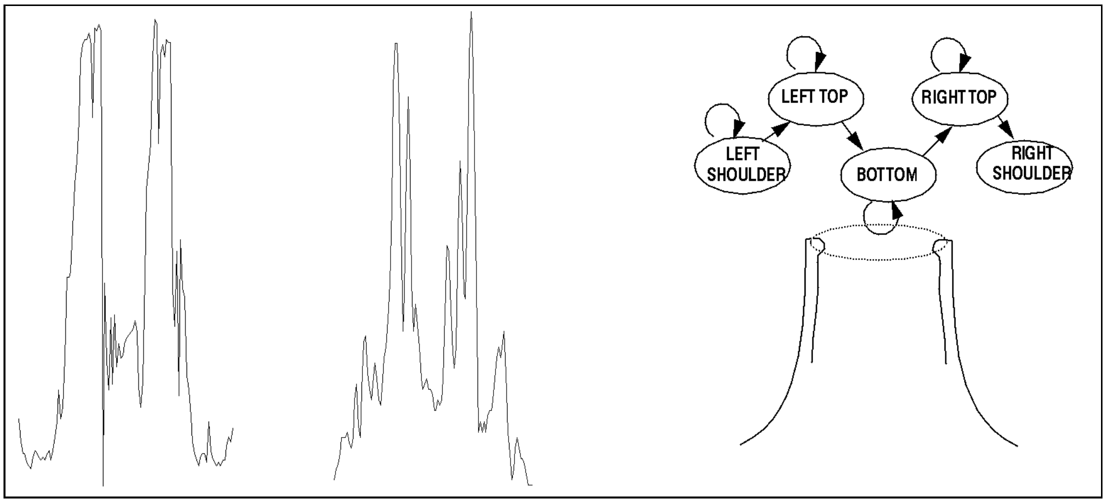

Figure 3: *Left: Scanlines (range and intensity) through the middle of the bottle. Right: The HMC model for a bottle.*

Figure 4: *MR image sequence from a breast. Left: Pre-contrast phase. Middle: Fast contrast enhancing phase. Right: Post-contrast enhanced plateau phase.*

Developments in MRI have made time-dynamic imaging possible, giving data consisting of a series of images acquired sequentially, where each image describes the same section of the body and has the same MR parameters. Figure 4 shows some samples from an MR image sequence acquired from a breast. By simply observing the images in the sequence separately, valuable information may get lost. Instead, the evolution of the image intensity with time, after injection of contrast agent, should be considered.

From the original MR image sequence, the time series is obtained by extracting a representative value from the same region of each image in the sequence. Inspection of the 1-dimensional signals from malignant and from benign lesions (Figure 5), indicates that the shapes of the two time series are very different. The malignant curves are characterized by a steep slope, while

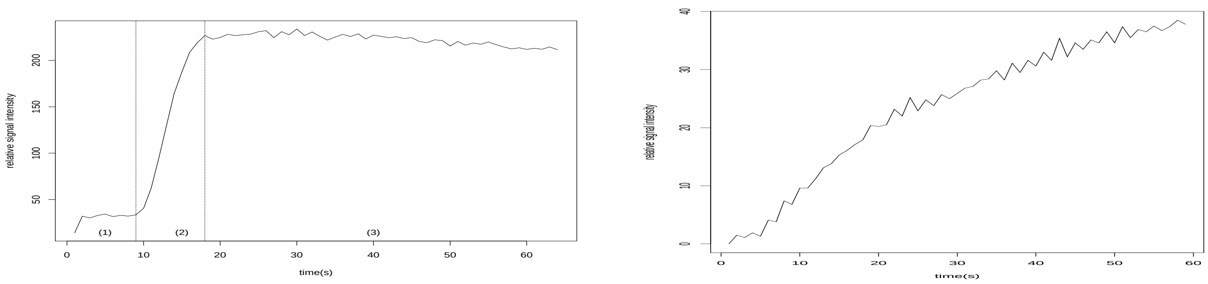

Figure 5: *Left: Time-intensity curve from a malignant tumor. States of HMC model shown by superimposed dashed lines. Right: Time-intensity curve from a benign tumor.*

the benign curves slant slightly.

The shape of a typical malignant curve was modeled as a Markov model with three states as illustrated in Figure 5, left. The first state describes the pre-contrast phase of the signal, the second state describes the fast contrast enhancing phase, while the third and last state describes the post-contrast enhanced plateau phase. It seemed natural to use a constant model for state 1 and 3, and a linear model for state 2.

For this problem we wanted to avoid the implicit geometric waiting time of the ordinary Markov model. Instead, uniform waiting time distributions were defined for all the states, meaning that the probability of leaving a state is independent of the time spent there. This gives us a semi Markov model.

Since it is the steep slope that characterises a malignant signal, the problem here was to determine, at each point in time, the probability of the observations matching a steep slope. This resulted in a probability curve for a malignant lesion for each image sequence. If the maximum value of this curve exceeded a threshold value, the curve was classified as malignant. Hence, the main problem was here to find the probability of the presence of a certain state given the observations.

3.4 HMC models for video analysis: Vehicle detection

Analysis of live video is a younger field in image analysis than the previously mentioned application areas. Markov models will probably have a large potential in analysis and recognition of movement, and a few papers on recognition of human movement have appeared [12, 13, 14]. In an ongoing project we are currently working with another kind of movement, namely the movement of vehicles in a traffic scene.

Improved systems for traffic surveillance can increase safety and efficiency on the roads, and current technology involving loop detectors buried in the road provides little flexibility and is expensive to maintain. Video-based systems give more flexibility, and faster hardware has made this a viable alternative. The main task of such a system is to count passing vehicles to estimate the traffic density. Other parameters like speed, may also be estimated.

In our study, we have used two consecutive lines within a lane to detect passing vehicles. Within these areas the image is analysed, and for each frame of the video, the detectors return TRUE (T) or FALSE (F) depending on whether a vehicle is present or not (Figure 6). The course of events reported by the detectors when a vehicle is passing gives a sequence of two-dimensional feature vectors, of which each component is either F or T. This means that there are four possible outcomes for the feature vectors; (F,F), (F,T), (T,F) and (T,T). This sequence is modelled as an HMM with 5 states as illustrated in Figure 6, right.

A vehicle is assumed to have passed the detection area if there is a large probability of transition from one of the states corresponding to a vehicle leaving the region (4 or 5) to one of the states corresponding to a new car entering the region (1 or 2). Hence, this problem can be defined as that of determining the probability of specific state transitions.

4 Solutions to hidden Markov chain problems

This section presents the solutions to each of the four classification problems stated in Section 3.

Figure 6: *Left: Frame from a video sequence showing the line detectors. Right: The five different states of the HMM for car detection.*

Text recognition. The problem is here to find the sequence of states that best matches the sequence of observations. That is, we want to recover the hidden part of the model, and we do this by finding the optimal state sequence, given the observations. Given a model λ we can find the single best state sequence by maximizing $P(C|Y,\lambda)$ with respect to C, which is equivalent to maximizing $P(C,Y|\lambda)$. A formal technique for finding the best state sequence exists and is called the Viterbi algorithm [3].

Bottle classification. The problem is here to find which of the models (bottle types) that best matches the sequence of observations. For each model, λ, we then have to compute the probability of the observation sequence, given the model: $P(Y_1 ... Y_n|\lambda)$ and classify an unknown bottle to the bottle type that corresponds to the model λ that maximizes this probability. We have that

$$P(Y_1 ... Y_n|\lambda) = \prod_{t=1}^{n} P(Y_t|Y_1, ... Y_{t-1}, \lambda). \tag{18}$$

From the model assumptions we have that the quantity $P(Y_t|Y_1, ... Y_{t-1}, \lambda)$ may be written as:

$$P(Y_t|Y_1, ... Y_{t-1}, \lambda) = \sum_{j=1}^{K} P(Y_t|C_t = j, \lambda)\, P(C_t = j|Y_1, ... Y_{t-1}, \lambda) \tag{19}$$

(K is the number of states of the model). This is an expression where the computation of

	Text recognition	**Bottle recognition**	**Tumor classification**	**Vehicle detection**
Number of states	6	5	3	5
Model for state process	Ordinary Markov	Ordinary Markov	Semi Markov	Ordinary Markov
Model for the observations	<u>All states:</u> Constant with uncorrelated Gaussian noise	<u>All states:</u> Constant with uncorrelated Gaussian noise	<u>States 1 and 3:</u> Constant with correlated Gaussian noise <u>State 2:</u> Linear with uncorrelated Gaussian noise	<u>All states:</u> Discrete

Table 2: *Choice of models.*

probabilities of the form $P(C_t = j|Y_1, ..., Y_{t-1})$ is an essential part, and such probabilities may be computed either by the approach described by Kitagawa in [4] or by Baum's forward recursion [5]. In this case the Kitagawa algorithm was used.

Tumor classification. The problem is here to determine, at each point in time, the probability of the observations belonging to a certain state j. Hence, for each time point t, we need to find $P(C_t = j|Y)$, i.e. the probability of the presence of state j, given all the observations. This probability can be written as:

$$P(C_t = j|Y_1, ..., Y_n) = \sum_w P(C_t = j, W_t = w|Y_1, ..., Y_n). \tag{20}$$

As it can be seen, this computation depends on the calculation of probabilities of the form $P(C_t = j, W_t = w|Y_1, ..., Y_n)$ where C_t is now a semi Markov chain. However, as stated in Section 2.1, it can be verified that the pair (C_t, W_t) may be interpreted as an ordinary Markov chain for which Baum's backward recursion [5] is valid.

Vehicle detection. The problem is here to find the probability of certain state transitions, corresponding to the passing of a vehicle at the time t. Hence, we compute the probability of transition between two states i and j at time t:

$$P(C_{t-1} = i, C_t = j | Y_1, ..., Y_n) = \frac{P(C_t = j | Y_1, ..., Y_t)}{P(C_t = j | Y_1, ..., Y_{t-1})} P(C_{t-1} = i | Y_1, ..., Y_{t-1}) \, a_{ij} \tag{21}$$

As for the bottle recognition, this problem involves computation of probabilities of the form $P(C_t = j | Y_1, ..., Y_t)$, and again the Kitagawa algorithm can be used.

	Text recognition	**Bottle recognition**	**Tumor classification**	**Vehicle detection**
Algorithm used for solving the problem	Viterbi	Kitagawa	Baum	Kitagawa
Parameter estimation	Segmental K-means	EM - algorithm	Manual	Manual

Table 3: *Choice of methods.*

5 Parameter estimation

The most difficult problem of HMMs is to determine the model parameters. The parameters may be divided into two categories; parameters involved in the distribution for the state process and parameters involved in the conditional distribution for the observations. The parameters involved in the distribution for the state process, are:

- The number of states N
- The prior probabilities, π_i
- The transition probabilities, a_{ij}
- The parameters of the waiting time distribution, $d_i(w)$

These parameters are often determined by an expert who has some prior knowledge about the state process. This was also the case for all the applications described in this paper.

The parameters involved in the conditional distribution for the observations depend on the choice of distribution. In what follows, the estimation of the observation distribution parameters is discussed for each of the four applications.

Text recognition. As stated in Section 3.1, 6 states were used for all but the shortest characters. For each state, a constant model with uncorrelated Gaussian noise was assumed. The components of each feature vector were assumed to be independently distributed. The reasons for assuming independence were mainly computational. The means and variances in the Gaussian distribution were estimated using the segmental K-means method [6].

Bottle classification. As for the text recognition problem, a constant model with uncorrelated noise was assumed. The dimension of the feature vectors was in this case only 2, meaning that assuming independence among the components of the feature vectors was not necessary. To estimate the expectation vectors and covariance matrices, we used the expectation-maximization (EM) method [15].

Tumor classification. For this application, a constant model with correlated Gaussian noise was assumed for the first and third states, while a linear model was assumed for the second state. The parameters μ_β and σ_β^2, as well as the noise variance σ_ε^2, were manually determined. The reason why manual parameter estimation was used for this application, was that we had relatively few samples and consequently would expect non-robust results using one of the automatic methods.

Vehicle detection. For this application, a discrete model with 5 states was assumed. The probabilities of each outcome given a specific state were manually determined. For each state, the probability of a detector being TRUE (FALSE) given that it should be TRUE (FALSE) was set to 0.95, and the probability of it being FALSE (TRUE) given that it should be TRUE (FALSE) to 0.05. The observations at the two detectors were assumed to be independent, i.e.

the probability of a feature vector was computed as the product of the probabilities of its two components.

6 Conclusions

The use of HMC models in image analysis have up to now been limited to the area of text recognition. Probably due to the 2-dimensional nature of images, focus has instead been on the 2-dimensional Markov random field models. However, MRF models can often be very complex and computationally expensive, which is an obvious drawback for practical applications in image analysis.

Even though images are 2-dimensional, this does not mean that the problem which is to be solved can not be of a 1-dimensional character, or at least be reduced to or expressed as a 1-dimensional problem. In this paper, we have demonstrated how this can be done for four different problems, which are initially very different in character. The four examples show that whether the original data source is a single image, a multiband image or a sequence of images, the essential data may be reduced to series of observations which can be handled as traditional time series.

Although Markov fields may lead to computationally expensive methods, this is not necessarily the case for Markov chains. For Markov chains the computational speed is mainly dependent on the number and dimensionality of the observations. For many problems the amount of observations does not need to be very large, and our example of vehicle detection in a video based traffic surveillance system demonstrates that Markov models can be used even for applications which require processing at video rate.

For applications in image analysis, the Markov models have several strong points:

- They enable integration of segmentation and classification for jointly optimization.
- They give a nice framework for easy integration of contextual and a priori information.
- They enable recognition of objects, events and phenomenons through their evolution over time.

In our examples we have exploited all these features of the Markov models. Integration of segmentation and classification is demonstrated through the text recognition example, integration of contextual information is utilized especially for text and bottle recognition, and integration of prior knowledge is obtained through the design of the models for all four examples. For the tumor classification and the vehicle detection where we have sequences of images, we have exploited the potential of recognition of events and phenomenons through their behaviour over time.

Generally, Markov chain models are well suited for problems that have a sequentially changing behaviour. Many real world problems have these characteristics, and when focusing at the characteristics of the problem to be solved, rather than at the characteristics of an image, we can see that this is also the case for many image analysis problems. The information in the image which is essential in solving the analysis problem, can, as we have seen, in many cases be viewed as a sequence of observations. Hence, we believe that Markov chain models have a potential in image analysis for a variety of applications, and with video entering the scene of image analysis, we believe that the potential of these methods will be even larger.

References

[1] L. R. Rabiner, A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition, *Proceeding of the IEEE* **77**, 257–285 (1989).

[2] S. E. Levinson, Continuously variable duration hidden Markov models for automatic speech recognition, *Computer Speech and Language* **1**, 29–45 (1986).

[3] G. Forney, The Viterbi Algorithm, *Proc. IEEE* **61**, 268–278 (1983).

[4] G. Kitagawa, Non-Gaussian State-Space Modeling of Non-stationary Time Series, *JASA* **82**, 1032–1041 (1987).

[5] L. E. Baum, T. Petrie, G. Soules and N. Weiss, A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Function of Markov Chains, *Annals of Mathematical Statistics* **41**, 164–171 (1970).

[6] B-H. Juang and L. R. Rabiner, The Segmental K-means Algorithm for Estimating Parameters of Hidden Markov Models, *IEEE Trans. on Acoust., Speech, Signal Processing* **38**, 1639–1641 (1990).

[7] K. Aas and L. Eikvil, Text page recognition using grey level features and hidden Markov models, *Pattern Recognition* **29**, 977–985 (1996).

[8] R. B. Huseby, G.T. Høgåsen, G. Storvik and K. Aas, Combining range and intensity data with a hidden Markov model, *Proc. 11th IAPR Conference* (1992).

[9] R.B. Huseby, Recognition of 3-D objects in coregistered range- and intensity images using a Hidden Markov Chain model, *Proc. 8th Scandinavian Conference on Image Analysis (SCIA)* (1993).

[10] J. Haslett, Maximum likelihood discriminant analysis on the plane using a Markovian model of spatial context, *Pattern Recognition* **18**, 287–296 (1985).

[11] K. Aas and M. Holden, Characterization of lesions in dynamical MRI using contrast agent and statistical analysis, *Proc. NOBIM-konferansen* (1994).

[12] J. Yamato, J. Ohya and K. Ishii. Recognizing human action in time-sequential images using hidden Markov models, *Proc. 1992 ICCV* 379–385 (1992).

[13] T. Starner and A. Pentland, Visual Recognition of American Sign Language Using Hidden Markov Models, *Proc. Int'l Workshop on Auto. Face and Gesture Recognition* 189–194 (1995).

[14] M. Brand, N. Oliver and S. Pentland, Coupled hidden Markov models for complex action recognition, *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR97)* (1997).

[15] A. P. Dempster, N. M. Laird and D. B. Rubin, Maximum Likelihood from incomplete data via the EM-algorithm, *J. Roy. Soc.* **39**, 1–38 (1977).