Contents

8.6 Course of dimensionality

Parameter estimation requires the knowledge of the likelihood p.d.f. and/or the prior p.d.f.. There are very few situations where these p.d.f.s can be obtained from mathematical deduction or physical reasoning. In almost any of these situations these p.d.f.s have to be "learned" using training data sets, or directly from the observations. The question that arises is: "Which is the size of a training data set necessary to derive the statistics of the data?"

We consider first the simple example of the one-dimensional p.d.f. $p(x)$, the image histogram. For images represented with 256 gray levels (8 bits), we consider the most defavorable situation, the uniform p.d.f. and assume that we are satisfied with an estimator of $p(x)\,\mathrm{d}x$ obtained using 16 realizations. It results that the minimum size of an 8-bit image required to estimate its histogram is 64×64 pixels. We repeat the same reasoning for the 2-dimensional p.d.f. $p(x_1, x_2)$. The required minimum size of the image to estimate $p(x_1, x_2)$ will be 1024×1024 pixels. This is the *course of dimensionality.*

Another simple example for the course of dimensionality is finding the "center" of a "unit interval" in an n-dimensional space (see Fig. 8.1).

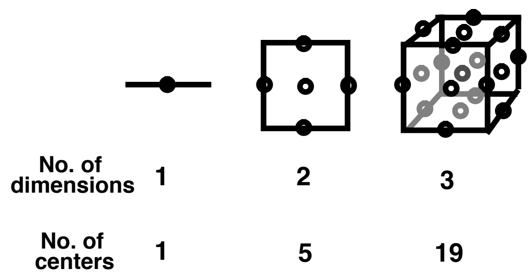

Figure 8.1: Exemplification of the circular assumption, the image is considered cylindrically periodic in both directions.

The large amount of required training data prohibits the learning of higher order p.d.f.s. The size of the required images is very high and strongly contradictory to our desire to have high resolution detail observables. There are two possible solutions: utilization of analytical models (e.g. Gaussian), or models described by only *local* statistics (Markov models, to be discussed in the next chapter).

8.7 Computational aspects of linear estimators

Until now we made only a short remark concerning the algebraic calculations involved when images are viewed as two-dimensional signals. Instead we considered the image as a one-dimensional vector where the pixels are arranged in TV-scan order. In this chapter we give more details. The model used for the estimation process is

$$\boldsymbol{y} = H\boldsymbol{x} + \boldsymbol{n} \tag{8.1}$$

where $\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{n}$ are vectors and H is a linear operator in matrix form. We further assume that this operator is a circular convolution.

$$y_k = \sum_{m=0}^{M-1} x_m h_{k-m}, \qquad k = 0, \ldots, M-1 \tag{8.2}$$

with

$$\boldsymbol{x} = \begin{bmatrix} x_0 \\ \vdots \\ x_{M-1} \end{bmatrix} \quad \boldsymbol{y} = \begin{bmatrix} y_0 \\ \vdots \\ y_{M-1} \end{bmatrix} \quad H = \begin{bmatrix} h_0 & h_{-1} & h_{-2} & \cdots & h_{-M+1} \\ h_1 & h_0 & h_{-1} & \cdots & h_{-M+2} \\ h_2 & h_1 & h_0 & \cdots & h_{-M+3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ h_{M-1} & h_{M-2} & h_{M-3} & \cdots & h_0 \end{bmatrix} \tag{8.3}$$

Due to the circular assumption we have

$$h_m = h_{M+m} \tag{8.4}$$

Thus

$$H = \begin{bmatrix} h_0 & h_{M-1} & h_{M-2} & \cdots & h_1 \\ h_1 & h_0 & h_{M-1} & \cdots & h_2 \\ h_2 & h_1 & h_0 & \cdots & h_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ h_{M-1} & h_{M-2} & h_{M-3} & \cdots & h_0 \end{bmatrix} \tag{8.5}$$

H is a circulant matrix: the rows (columns) are related by a circular shift. If the extent of the convolution kernel $h(\cdot)$ is limited to the range $r < M$ where $h_k = 0$ for $r < k < M$ then H is a band matrix.

We generalize the circular convolution to two dimensions (Fig. 8.2).

$$y_{k,l} = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} x_{m,n} h_{k-m,l-n}, \qquad k = 0, \ldots, M-1; \qquad l = 0, \ldots, N-1; \tag{8.6}$$

The indices are computed modulo M and modulo N, respectively (Fig. 8.3). Let $\boldsymbol{y}$ and $\boldsymbol{x}$ represent MN-dimensional column vectors formed by stacking the rows of the images of size $M \times N$ (Fig. 8.4). Using this convention the convolution can be written as matrix-vector product.

$$\boldsymbol{y} = H\boldsymbol{x} \tag{8.7}$$

where $\boldsymbol{x}$ and $\boldsymbol{y}$ are of dimension $MN \times 1$ and H is of dimension $MN \times MN$. The matrix consists of M^2 partitions each being of size $N \times N$ and ordered according to

$$H = \begin{bmatrix} H_0 & H_{M-1} & H_{M-2} & \cdots & H_1 \\ H_1 & H_0 & H_{M-1} & \cdots & H_2 \\ H_2 & H_1 & H_0 & \cdots & H_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ H_{M-1} & H_{M-2} & H_{M-3} & \cdots & H_0 \end{bmatrix} \tag{8.8}$$

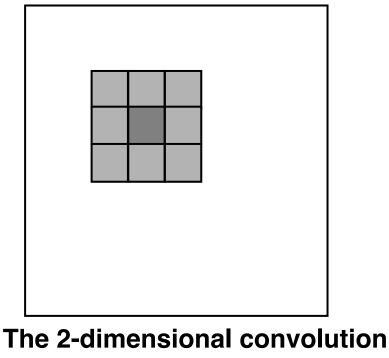

Figure 8.2: The 2-dimensional convolution, an example using a 3×3 blurring kernel. The convolution kernel runs over the image and all pixels are replaced by the weighted average of the neighbors.

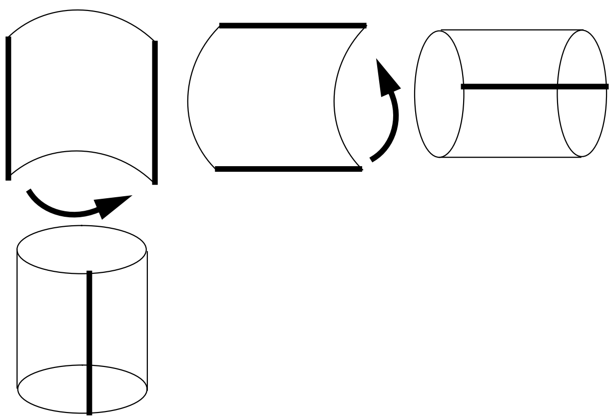

Figure 8.3: Exemplification of the circular assumption, the image is considered cylindrically periodic in both directions.

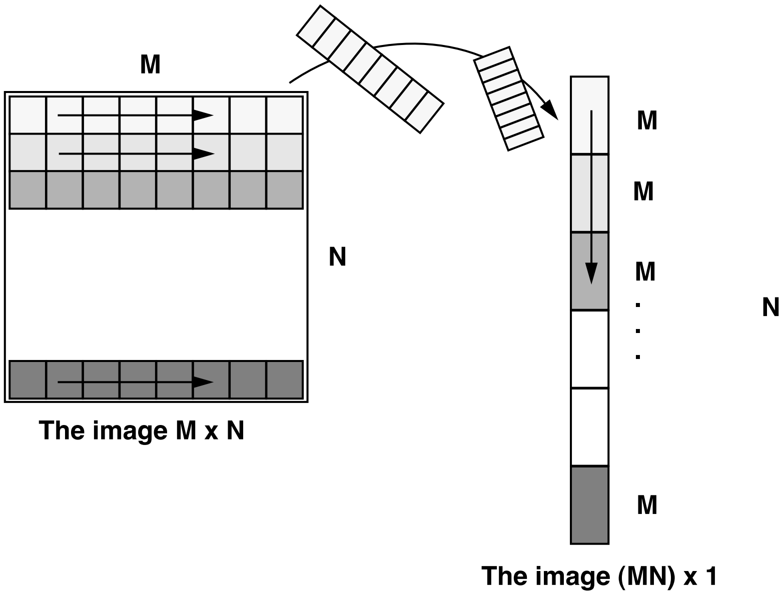

Figure 8.4: Diagram showing the re-arrangement of an image in vector by staking its rows.

Each partition is constructed from the j-th row of the extended convolution kernel $h_{m,n}$

$$H_j = \begin{bmatrix} h_{j,0} & h_{j,M-1} & h_{j,M-2} & \cdots & h_{j,1} \\ h_{j,1} & h_{j,0} & h_{j,M-1} & \cdots & h_{j,2} \\ h_{j,2} & h_{j,1} & h_{j,0} & \cdots & h_{j,3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ h_{j,M-1} & h_{j,M-2} & h_{j,M-3} & \cdots & h_{j,0} \end{bmatrix} \tag{8.9}$$

H_j is a circulant matrix and the blocks of H are also subscripted in a circular way. H is a *block-circulant* matrix. All the matrix computation involved in the linear estimation proceeds on that basis. The diagram in Fig. 8.5 shows qualitatively the arrangement of the image.

Observation For a *small* image $N = M = 512$ the size of the matrix H is 262144×262144 ! The computation task is immense.

8.7.1 Diagonalization of circulant and block-circulant matrices

Given a circulant matrix H we consider a scalar function $\lambda(k)$

$$\begin{aligned} \lambda(k) \;&=\; h_0 + h_{M-1}\,\exp\left[j\frac{2\pi}{M}k\right] + h_{M-2}\,\exp\left[j\frac{2\pi}{M}2k\right] + \\ &\cdots + h_1\,\exp\left[j\frac{2\pi}{M}(M-1)k\right] \end{aligned} \tag{8.10}$$

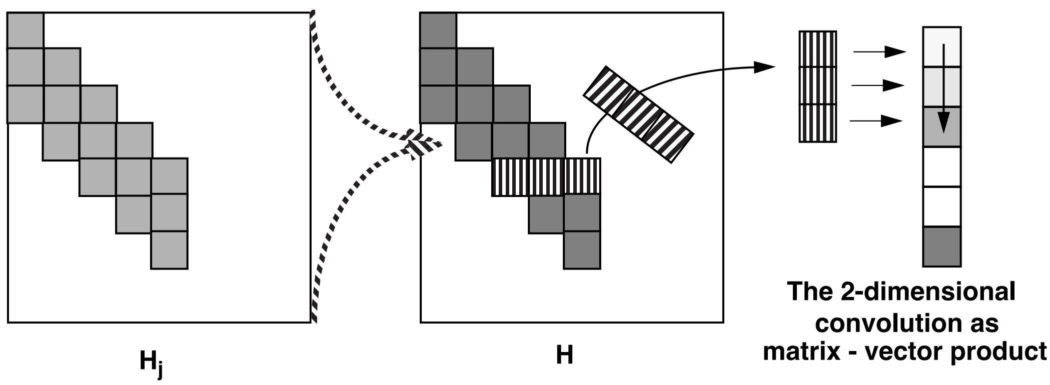

Figure 8.5: Diagram showing the computation of 2-dimensional convolution as matrix-vector product. The exemplification is for a spatially limited convolution kernel, thus the matrices H_j and H are band diagonal. The 2-dimensional spatial vicinity computation is obtained by the "filtering" effect of blocks of matrix H over the stacked rows of the image.

$$\boldsymbol{\omega}(k) = \begin{bmatrix} 1 \\ \exp\left[j\frac{2\pi}{M}k\right] \\ \vdots \\ \exp\left[j\frac{2\pi}{M}(M-1)k\right] \end{bmatrix}, \quad k = 0,\ldots,M-1 \tag{8.11}$$

It can be shown that

$$H\,\boldsymbol{\omega}(k) = \lambda(k)\boldsymbol{\omega}(k). \tag{8.12}$$

$\omega(k)$ are *eigenvectors* of the matrix H and $\lambda(k)$ are the *eigenvalues*. We build now a matrix Ω

$$\Omega = [\boldsymbol{\omega}(0), \boldsymbol{\omega}(1), \ldots, \boldsymbol{\omega}(M-1)] \tag{8.13}$$

having as columns the eigenvectors of H. The elements of Ω are

$$\Omega_{k,l} = \frac{1}{M}\exp\left[-j\frac{2\pi}{M}k\,l\right] \tag{8.14}$$

The matrix Ω is orthogonal

$$\Omega\Omega^{-1} = \Omega^{-1}\Omega = I \tag{8.15}$$

I is the $M \times N$ identity matrix. It follows that H may be expressed as

$$H = \Omega\Lambda\Omega^{-1} \tag{8.16}$$

thus

$$\Lambda = \Omega^{-1} H \Omega \tag{8.17}$$

where Λ is the diagonal matrix

$$\Lambda_{k,k} = \lambda(k) \tag{8.18}$$

formed by the eigenvalues of H. A similar result is obtained by generalization for the block-circulant matrices.

Observation The observation model, the convolution can be written

$$\boldsymbol{y} = \Omega \Lambda \Omega^{-1} \boldsymbol{x} + \boldsymbol{n} \tag{8.19}$$

and

$$\Omega^{-1} \boldsymbol{y} = \Lambda \Omega^{-1} \boldsymbol{x} + \Omega^{-1} \boldsymbol{n} \tag{8.20}$$

where the vector $\Omega^{-1} \boldsymbol{y}$ corresponds to the stacked rows of the Fourier transform matrix of the image. Similar reasoning applies to the vectors $\Omega^{-1} \boldsymbol{x}$ and $\Omega^{-1} \boldsymbol{n}$.

Remark The solution of the linear estimation problems reduces to evaluation of several discrete Fourier transforms. The FFT algorithm makes the computation efficient.

8.8 Measures of information

In this section we will shortly describe the concept of information and give some properties that are essential for its use in estimation theory.

The best known information measures are *Shannon's entropy, Kullback-Leibler divergence and Fisher's information matrix.* Some of the concepts which are involved in the notion of information are *statistical entropy, uncertainty, accuracy, coding, questionaires, statistical independence, probabilistic distance* and *discrimination ability.*

We will consider mainly the concept of accuracy related to the performance of parameter estimators and address two problems

1. How much information is contained in the observations of $\boldsymbol{y}$ about a parameter $\boldsymbol{x}$?

2. What is a "good" estimator?

8.8.1 Shannon measure of information

Shannon called the average information contained in a random variable X with p.d.f. $p(\boldsymbol{x})$, *entropy* $S(X)$

$$S(X) = -\int \mathrm{d}\boldsymbol{x}\, p(\boldsymbol{x}) \log p(\boldsymbol{x}) \tag{8.21}$$

If we consider the observation model

$$\boldsymbol{y} = H\boldsymbol{x} + \boldsymbol{n} \tag{8.22}$$

the quantity

$$S(X,Y) = -\int\int \mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{y}\,p(\boldsymbol{x},\boldsymbol{y})\log p(\boldsymbol{x},\boldsymbol{y}) \tag{8.23}$$

is a measure of the average information in the joint space (X,Y). In a similar way the conditional entropies are defined

$$S(X|Y) = -\int\int \mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{y}\,p(\boldsymbol{x},\boldsymbol{y})\log p(\boldsymbol{x}|\boldsymbol{y}) \tag{8.24}$$

and

$$S(Y|X) = -\int\int \mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{y}\,p(\boldsymbol{y},\boldsymbol{x})\log p(\boldsymbol{y}|\boldsymbol{x}). \tag{8.25}$$

In the frame of our discussion it is of interest to calculate the mutual information $I(X,Y)$

$$\begin{aligned} I(X,Y) &= S(X) + S(Y) - S(X,Y) && (8.26)\\ &= S(X) - S(X|Y) && (8.27)\\ &= S(Y) - S(Y|X) && (8.28)\end{aligned}$$

which is a measure of the information in the random variable Y about the random variable X. For example, if the observation of X is done under very noisy conditions, X and Y will be statistically independent. No "information" about X will be recognized in Y and

$$I(X,Y) = 0 \tag{8.29}$$

Until now we considered the images to be modeled only by the causality

$$X \to Y$$

described in the likelihood p.d.f. $p(\boldsymbol{y}|\boldsymbol{x})$. We consider a more general causality

$$U \to X \to Y \to V$$

described by

$$p(\boldsymbol{u},\boldsymbol{x},\boldsymbol{y},\boldsymbol{v}) = p(\boldsymbol{v}|\boldsymbol{y})\,p(\boldsymbol{y}|\boldsymbol{x})\,p(\boldsymbol{x}|\boldsymbol{u})\,p(\boldsymbol{u}). \tag{8.30}$$

A more detailed discussion of this model follows in the next chapter, but a simple example is the causality of the observation of some physical parameter of a scene, like the mass of wood in a forest (U), which is mapped in a noiseless image (X), observed with a noisy camera (Y) and finally received in an the measurement (V).

The following inequality is verified (data processing theorem)

$$I(U,V) \leq I(X,Y) \tag{8.31}$$

Loosely speaking, "processing" never increases information.

An estimator $\hat{\boldsymbol{x}}$ was obtained applying some linear or non-linear transform to the observations. Thus

$$I(X,\hat{X}) \leq I(X,Y) \tag{8.32}$$

The maximum information contained in an estimator $\hat{\boldsymbol{x}}$ (which is a random variable) can not be higher than the information in our process.

8.8.2 Kullback-Leibler divergence

Kullback-Leibler divergence is a generalization of Shannon's measure of information. The divergence is a function of two p.d.f.s potentially characterizing a random variable X.

$$L(p,q) = \int \mathrm{d}\boldsymbol{x}\, p(\boldsymbol{x}) \, \log \frac{p(\boldsymbol{x})}{q(\boldsymbol{x})} \tag{8.33}$$

We will interpret the divergence as a measure that describes the amount of information that a measurement gives about the truth of a given model as compared to a second alternative model.

Observation If $q(\boldsymbol{x})$ is a uniform p.d.f. the divergence is nothing else but the Shannon entropy for the r.v. X. Thus the Shannon entropy can be interpreted as the amount of information in a model $p(\boldsymbol{x})$ of X compared to the maximum incertitude model – the uniform distribution. Note that the uniform distribution is the one with maximum entropy.

8.8.3 Quality of an estimator

As we have seen, an estimator $\hat{x}$ is itself a random variable because it is a function of the random variable Y, the observations. Estimation theory studies various criteria for making good estimates.

Intuitively one hopes to choose the estimator $\hat{x}(\boldsymbol{y})$ so that

1. Its conditional mean satisfies

$$\mathcal{E}[\hat{x}(\boldsymbol{y})] = x. \tag{8.34}$$

   This means that the estimator is unbiased.

2. Its conditional variance $\sigma^2_{\hat{x}}$ is as small as possible.

Information theory provides us with the lower bound on the conditional variance of any estimator. We approach the problem of estimation quality by comparing the "models" used to describe two unbiased scalar estimators $\hat{x}_0$ and $\hat{x}_1$. Considering that $\hat{x}_0$ and $\hat{x}_1$ are close to each other and using a Taylor expansion for $\xi \approx 1$.

$$\xi \log \xi = (\xi - 1) + \frac{1}{2}(\xi - 1)^2 + \dots \tag{8.35}$$

therefore

$$\frac{p(\boldsymbol{y}|x_0)}{p(\boldsymbol{y}|x_1)} \log \frac{p(\boldsymbol{y}|x_0)}{p(\boldsymbol{y}|x_1)} \approx \left(\frac{p(\boldsymbol{y}|x_0)}{p(\boldsymbol{y}|x_1)} - 1\right) + \frac{1}{2}\left(\frac{p(\boldsymbol{y}|x_0)}{p(\boldsymbol{y}|x_1)} - 1\right)^2. \tag{8.36}$$

And assuming also that

$$\lim_{x_1 \to x_0} \max_{\boldsymbol{y}} \left| \frac{p(\boldsymbol{y}|x_0) - p(\boldsymbol{y}|x_1)}{p(\boldsymbol{y}|x_0)} \right| = 0 \tag{8.37}$$

we can write

$$L\left(p(\boldsymbol{y}|x_0), p(\boldsymbol{y}|x_1)\right) \approx \frac{1}{2}\int \mathrm{d}\boldsymbol{y}\, p(\boldsymbol{y}|x_0)\left(\frac{p(\boldsymbol{y}|x_1)}{p(\boldsymbol{y}|x_0)} - 1\right)^2. \tag{8.38}$$

Based on this assumptions we demonstrate the following inequality

$$\sigma^2_{\hat{x}_1} \cdot \left[\int \mathrm{d}\boldsymbol{y}\, p(\boldsymbol{y}|x_0)\left(\frac{p(\boldsymbol{y}|x_0)}{p(\boldsymbol{y}|x_1)} - 1\right)\right] \geq (x_0 - x_1)^2 \tag{8.39}$$

Expectations are evaluated with respect to $p(\boldsymbol{y}|x_0)$. We use the Schwarz inequality

$$\mathcal{E}\left[(\hat{x}_1(\boldsymbol{y}) - \mathcal{E}[\hat{x}_1])^2\right] \mathcal{E}\left[\left(1 - \frac{p(\boldsymbol{y}|x_0)}{p(\boldsymbol{y}|x_1)}\right)^2\right] \tag{8.40}$$

$$\geq \left(\mathcal{E}\left[(\hat{x}_1(\boldsymbol{y}) - \mathcal{E}[\hat{x}_1])\left(1 - \frac{p(\boldsymbol{y}|x_0)}{p(\boldsymbol{y}|x_1)}\right)\right]\right)^2 \tag{8.41}$$

$$= \left(\int \mathrm{d}\boldsymbol{y}\, p(\boldsymbol{y}|x_1)\,(\hat{x}_1(\boldsymbol{y}) - x_1)\left(1 - \frac{p(\boldsymbol{y}|x_0)}{p(\boldsymbol{y}|x_1)}\right)\right)^2 \tag{8.42}$$

$$= (x_1 - x_0)^2 \tag{8.43}$$

And now for $x_0 = x_1 + \triangle x_1$ and taking $\triangle x_1 \to 0$ we find

$$\sigma^2_{\hat{x}} \geq \frac{1}{\mathcal{E}\left[\frac{\partial}{\partial x}\log p(\boldsymbol{y}|x)\right]^2} \tag{8.44}$$

the *Cramer-Rao* bound for the variance of an unbiased estimator. The Cramer-Rao bound can also be written as a function of the divergence

$$\sigma^2_{\hat{x}} \geq \left[\lim_{\hat{x}_1 \to \hat{x}_0} \frac{2\, L(p(\boldsymbol{y}|\hat{x}_0), p(\boldsymbol{y}|\hat{x}_1))}{(\hat{x}_0 - \hat{x}_1)^2}\right]^{-1} \tag{8.45}$$

For vector parameter the inferior bound of the accuracy is given by the Fisher information matrix. $I(X)$ with the elements

$$I_{i,j}(x) = \int \mathrm{d}\boldsymbol{y}\, p(\boldsymbol{y}|x)\left[\frac{\partial}{\partial x_i}\log p(\boldsymbol{y}|x)\right]\left[\frac{\partial}{\partial x_j}\log p(\boldsymbol{y}|x)\right], \qquad i,j = 1,\dots N \tag{8.46}$$

The covariance of the estimator

$$C_{\hat{x}} \geq I^{-1}(x) \tag{8.47}$$

is bounded by the Fisher information matrix.

Observation The Cramer-Rao bound is obtained as a particular case of the Fisher information matrix for $N = 1$. It can be expressed in an equivalent formula

$$\sigma^2_{\hat{x}} \geq \frac{-1}{\mathcal{E}\left[\frac{\partial^2}{\partial x^2}\log p(\boldsymbol{y}|x)\right]} \tag{8.48}$$

if the derivative exists. The Fisher information matrix can also by expressed as a function of the divergence.

Example We consider N independent and identically distributed Gaussian observations $y_1, \ldots y_N$. The mean x is unknown and the variance is $\sigma^2 = 1$.

$$p(\boldsymbol{y}|x) = \frac{1}{(\sqrt{2\,\pi})^M} \exp\left(-\frac{1}{2}\sum_{k=1}^{N}(y_k - x)^2\right). \tag{8.49}$$

We evaluate the Cramer-Rao bound for the estimator $\hat{x}$.

$$\frac{\partial}{\partial x}\log p(\boldsymbol{y}|x) = \sum_{k=1}^{N}(y_k - x) \tag{8.50}$$

$$\frac{\partial^2}{\partial x^2}\log p(\boldsymbol{y}|x) = -N \tag{8.51}$$

The estimator is

$$\hat{x} = \frac{1}{N}\sum_{k=1}^{N} y_k \tag{8.52}$$

and its variance is

$$\sigma^2_{\hat{x}} = \frac{1}{N}. \tag{8.53}$$

Remark An estimator which reaches its inferior Cramer-Rao bound is called *efficient*.

8.9 Image formation and image models

A simplified image formation process is presented in Fig. 8.6. Images are recorded in order to obtain useful information about a scene. The image formation process assumes illumination of the scene with certain radiation: light, laser, microwaves, X-rays, acoustic waves etc. The reflected (or transmitted) field contains information of the scene geometry, like shapes and shadows, and also scattering (or transmittance) information of the physical quantities (e.g. density, humidity, roughness etc.). The reflected (transmitted) radiation called cross-section propagates to a sensor. The propagation can be affected by disturbances: refraction, absorbation, etc. The sensor records the incident field and transforms it in an electrical signal, the image. Thus the model for image formation can be summarized as a sequential process:

scene $\rightarrow$ cross-section $\rightarrow$ image.

We identify two types of problems:

1. Scene understanding: Given an observed image, reconstruct the scene identity, i.e. surface or object geometry reconstruction.

2. Image understanding: Given an observed image, reconstruct and recognize the structures in the "undisturbed" image, the cross-section, i.e. edge and texture identification.

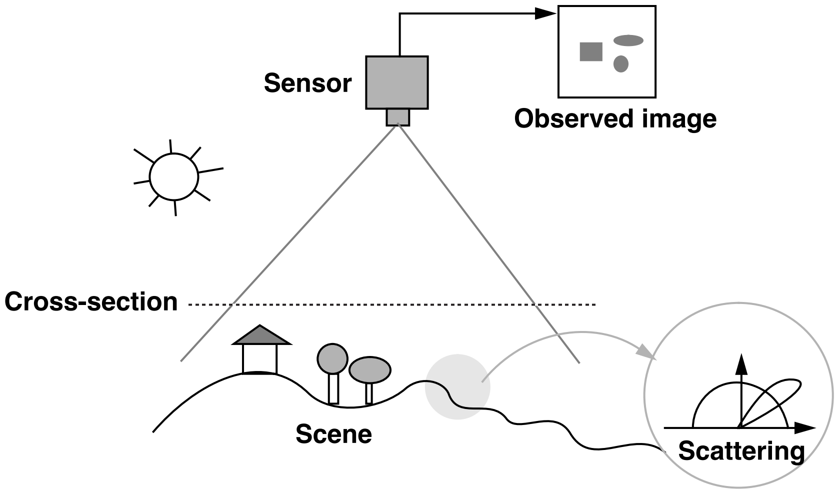

Figure 8.6: A simple image formation example.

Both problems are *inverse problems* and require *information extraction.* In these lectures we limit ourselves to the second case, and approached information extraction as a parameter estimation task. In the previous chapters the example used was the estimation of the image intensity. We extend now the parameter estimation problem to the more general case of image understanding.

The hierarchy proposed below can be described by certain stochastic models.

- $p(\boldsymbol{\xi})$ the scene model, $\boldsymbol{\xi}$ is the parameter vector characterizing the geometry and radiometry of th scene.

- $p(\boldsymbol{x}, \boldsymbol{\theta})$ the cross-section model, where $\boldsymbol{x}$ represents the intensity process of the "ideal" image, and $\boldsymbol{\theta}$ is the vector parameter characterizing the structures in the cross-section image.

Both $\boldsymbol{x}$, and $\boldsymbol{\theta}$ depend on $\boldsymbol{\xi}$. This causality is modeled by the conditional process

$$p(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{\xi}).$$

The propagation and the sensor are modeled by an operator H and the stochastic noise. Thus the observed image will be

$$\boldsymbol{y} = H\boldsymbol{x}_\theta + \boldsymbol{n} \tag{8.54}$$

The notation $\boldsymbol{x}_\theta$ represents the dependence of the cross-section from the structural parameter. $\boldsymbol{\theta}$. Thus the image understanding problem can be understood as inverting $\boldsymbol{x}$ and $\boldsymbol{\theta}$ from the observation $\boldsymbol{y}$. This task can be accomplished using the Bayes formula

$$p(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta})\, p(\boldsymbol{x}, \boldsymbol{\theta})}{p(\boldsymbol{y})} \tag{8.55}$$

and the solution will be a MAP estimator. The same reasoning applies for the inversion of the scene parameter $\boldsymbol{\xi}$. The diagram of Fig. 8.7 summarizes the image formation process.

Table 8.1 presents a classification of the tasks of image understanding. The estimation of $\boldsymbol{x}$ was already discussed, it is the estimation of the image intensity. Estimation of parameter $\boldsymbol{\theta}$ means recognition of structures: lines, borders, textures. It will be discussed in the next chapter.

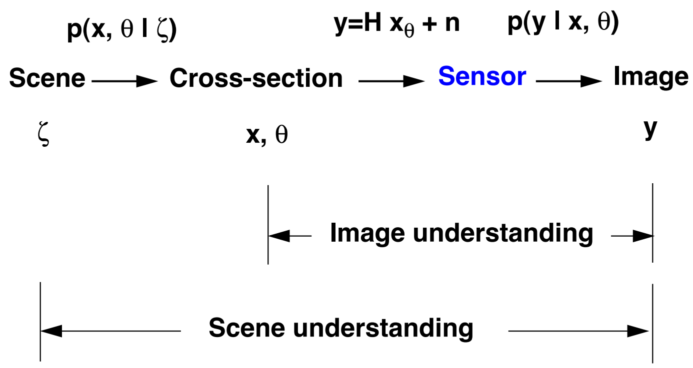

Figure 8.7: Hierarchical modeling of the image formation process. The causalities scene, cross-section, image are stochastical models described by probability density functions, the sensor is characterized by a deterministic model, the operator H, and the noise as another stochastic process. The task of image understanding is the inversion of the cross-section (the ideal image) and its parameters from the observations, the task of scene understanding is the inversion of the scene parameters from the observed image.

	Problem	Description	Models
1.	SMOOTHING	Given noisy image data, filter it to smooth out the noise variations.	Noise & image Power Spectra
2.	ENHANCEMENT	Bring out or enhance certain features of the image e.g., edge enhancement, contrast stretching, etc.	Features
3.	RESTORATION & FILTERING	Restore an image with known (or unknown) degradation as close to its a original form as possible, e.g., image De-blurring, image reconstruction, image registration, geometric correction etc.	Degradations, Criterion of "closeness"
4.	DATA COMPRESSION	Minimize the Number of Bits required to store/transmit an image for a give level of Distortion.	Distortion Criterion, Image as an Information source
5.	FEATURE EXTRACTION	Extract certain features from an image, e.g., edges.	Features, detection criterion
6.	DETECTION AND IDENTIFICATION	Detect and identify the presence of an object from a scene e.g., matched filter, pattern recognition and image segmentation, texture analysis etc.	Detection criterion, object and scene.
7.	INTERPOLATION AND EXTRAPOLATION	Given image data at certain points in a region, estimate the image values of all other points in inside this region (interpolation) and also at points outside this region (extrapolation).	Estimation Criterion, and Degree of smoothness of the data
8.	SPECTRAL ESTIMATION	Given image data in a region estimate its power spectrum.	Criterion of Estimation, A priori model for data
9.	SPECTRAL FACTORIZATION	Given the magnitude of the frequency response of a filter, design a realizable filter e.g., a stable "causal" filter.	Criterion of realizability
10.	SYNTHESIS	Given a description or some features of an image, design a system which reproduces a replica of that image; e.g., texture synthesis.	Features, Criterion of reproduction

Table 8.1: Typical Problems in Image Processing

8.10 Outlook

In this lecture we first identified two computational aspects:

1. The learning of a stochastic model from training data, which is strongly limited by the course of dimensionality thus being practically impossible to derive more than the 2-dimensional joint p.d.f. of a process. Possible solutions are the derivation of the model based on mathematical or physical reasoning, the utilization of ad hoc parametric models, and the application of models with local properties (Markov).

2. The representation of images, which are 2-dimensional signals, in a format suitable for matrix computation involved in the linear estimation. The solution presented used the circular assumption for the signals and blurring operator, and allows efficient computation using FFT.

The next topic approached was the quality of the estimators. The Shannon entropy, the Kullback-Leibler divergence and Fisher information matrix have been discussed. A demonstration for Cramer-Rao bound is given.

The last part of the lecture systematized the image formation process and introduced the hierarchical modeling concept. Based on this concepts the basic tasks of image analysis have been summarized.

Bibliography

[1] R. Gonzales, P. Wintz, *Digital Image Processing*, Addison-Wesley, 1987.

[2] A. Jain, *Advances in Mathematical Models for Image Processing*, Proceedings of IEEE, Vol.69, No.5, May 1981, pp. 502-528.

[3] R. Blahut, *Principles and practice of information theory*, Addison-Wesley, 1990.

[4] Al. Spataru, *Fondéments de la théorie de la transmission de l'information,* Presses Polytechniques Romandes , 1987