**Lecture notes**

# Mathematical methods of signal and image processing


Benjamin Berkels


WS 2021/22


RWTH Aachen



last change: October 11, 2021

typeset using KOMA-Script and LaTeX

# 1. Digital images and point operators

## References

[1] K. Bredies and D. Lorenz. *Mathematical Image Processing.* 1st ed. Vieweg+Teubner Verlag, 2018.

[2] R. C. Gonzalez and R. E. Woods. *Digital Image Processing.* 3rd Edition. Prentice Hall, 2007.

[3] S. B. Damelin and W. Miller Jr. *The Mathematics of Signal Processing.* Cambridge University Press, 2012.

[4] S. Mallat. *A Wavelet Tour of Signal Processing.* 3rd ed. Elsevier, 2009.

[5] G. Aubert and P. Kornprobst. *Mathematical Problems in Image Processing.* 2nd ed. Springer New York, 2006.

## 1.1. Digital images

Mathematically, an image is a mapping from a domain $\Omega \subset \mathbb{R}^d$ to a value range $V$. For instance, the mapping

$$f : (0,1)^2 \to [0,1]$$

is a typical (continuous) grayscale image. As domain, we consider a $d$-dimensional cuboid, i. e. $\Omega = (a_1, b_1) \times \cdots \times (a_d, b_d) \subset \mathbb{R}^d$. Depending on the acquisition technique, there are quite different value ranges:

- Grayscale images: $V = \mathbb{R}$ or $V = [0,1]$.
- Color images: $V = \mathbb{R}^3$ or $V = [0,1]^3$ (different color models, e. g. RGB, HSV).
- Diffusion Tensor Imaging (DTI): $V = \big\{M \in \mathbb{R}^{3\times 3} \;:\; M \text{ symmetric, positive definite}\big\}$
- Spectral Imaging: $V = \mathbb{R}^n$ or $V = [0,1]^n$ with $n \gg 1$.
- ...

**Definition 1.1** (Digital image)**.**

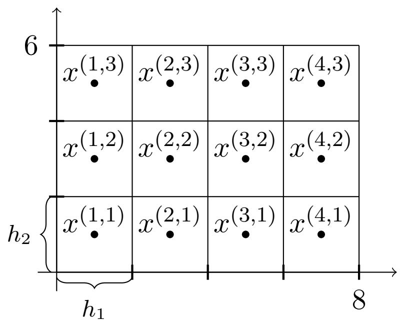
(i) $d \in \mathbb{N}$ is called *spatial dimension.*

(ii) Let $\underline{m} = (m_1, \ldots, m_d) \in \mathbb{N}^d$. The tuple $\underline{m}$ is called *image size.*

(iii) Combined, $\underline{m}$ and the cuboid $\Omega$ imply a grid: An index $\underline{j} \in \mathcal{I} := \{1, \ldots, m_1\} \times \cdots \times \{1, \ldots, m_d\}$ corresponds to the grid node $x^{\underline{j}} = \Big(x_1^{\underline{j}}, \ldots, x_d^{\underline{j}}\Big) \in \Omega$, where

$$x_i^{\underline{j}} = a_i + \big(j_i - \tfrac{1}{2}\big)\, h_i,\; i = 1, \ldots, d.$$

Here, the *grid size* $\underline{h} \in \mathbb{R}^d$ is defined as $h_i = \frac{b_i - a_i}{m_i}$, $i = 1, \ldots, d$.




Grid for $d = 2$, $\Omega = (0, 8) \times (0, 6)$, $\underline{m} = (4, 3)$

(iv) A *grid* $\underline{X}$ is the $d$-dimensional array (tensor of order $d$, short: $d$-array) of all grid nodes, i. e.

$$\underline{X} = (x^{\underline{j}})_{\underline{j} \in \mathcal{I}}.$$

Such a grid is called *regular grid.* The special case $h_1 = \ldots = h_d$ is called *cartesian grid.*

(v) A grid implies a partition of $\Omega$ into cells

$$c_{\underline{j}} = \left\{ x \in \Omega \,:\, \left| x_i - x_i^{\underline{j}} \right| < \tfrac{h_i}{2} \text{ for all } i \in \{1, \ldots, d\} \right\}.$$

We have,

(1) $\overline{\Omega} = \bigcup\limits_{\underline{j} \in \mathcal{I}} \overline{c_{\underline{j}}}$

(2) $c_{\underline{j}} \cap c_{\underline{k}} = \emptyset$ for $\underline{j}, \underline{k} \in \mathcal{I}$ with $\underline{j} \neq \underline{k}$

(3) $c_{\underline{j}} \neq \emptyset$ for $\underline{j} \in \mathcal{I}$

This kind of grid is called *cell-centered* and is usually used with finite differences. For multilinear finite elements, *mesh-centered* grids are more common.

(vi) If a value (e. g. a gray value) $f_{\underline{j}}$ is assigned to each cell, the values can be arranged in a $d$-array $F = (f_{\underline{j}})_{\underline{j} \in \mathcal{I}}$. This kind of $F$ is called *digital image*, *discrete image* or *pixel image.*

For actual image data one has to take into account that the value range is quantized and only contains finitely many elements. If a grayscale image is stored with a precision of $n$ bit, there are $2^n$ possible values. For grayscale images, the precision is typically 8 bit, i. e. $V = \{0, \ldots, 255\}$. Color images are usually stored with a precision of 8 bit per channel. Modern video formats like HDR10 or Dolby Vision use 10 resp. 12 bit.

(vii) The tuple $\left( c_{\underline{j}}, f_{\underline{j}} \right)$ is called *pixel* (for $d = 2$) or *voxel* (for $d = 3$). Pixel is short for “picture element” and voxel the analogue for volume elements.

(viii) A function $f : \Omega \to V$ that is constant on each cell, i. e. $f|_{c_{\underline{j}}}$ is constant for all $\underline{j} \in \mathcal{I}$, is called *pixelated.* Since

$$L := \Omega \setminus \bigcup_{\underline{j} \in \mathcal{I}} c_{\underline{j}}$$

is a Lebesgue null set, the values can be freely chosen in the sense of the Lebesgue measure. If $x \in L$, then $x$ is between (at least) two cells, i. e. there are $\underline{j} \in \mathcal{I}$ and $i \in \{1, \dots, d\}$, such that $x \in \overline{c_{\underline{j}}}$ and $x_i = x_i^{\underline{j}} - \frac{h_i}{2}$. If multiple such indices $\underline{j}$ exist, we choose the index such that $\#\left\{i \in \{1, \dots, d\} \,:\, x_i = x_i^{\underline{j}} - \frac{h_i}{2}\right\}$ is maximal. Here, $\#A$ denotes the number of elements in the set $A$. For a pixelated function, we then assume that $f(x) = f(c_{\underline{j}})$.

*Example* 1.2. For $d = 2$, $\Omega = (0, 8) \times (0, 6)$, $\underline{m} = (4, 3)$ we have

$$\underline{X} = \begin{pmatrix} (1,5) & (3,5) & (5,5) & (7,5) \\ (1,3) & (3,3) & (5,3) & (7,3) \\ (1,1) & (3,1) & (5,1) & (7,1) \end{pmatrix}.$$

An image on this grid is, for instance,

$$F = \begin{pmatrix} 0 & 1 & 2 & 3 \\ 0 & 1 & 2 & 3 \\ 0 & 1 & 2 & 3 \end{pmatrix}.$$

*Remark* 1.3.

- In case $d = 1$, one more commonly talks about *signals* rather than images.
- In case $d = 2$, the cuboid $\Omega$ is usually chosen such that $h_1 = h_2 =: h$. Moreover, the cuboid is usually such that either $h = 1$ (thus $b_i - a_i = m_i$) or $h = \min \frac{1}{m_i}$ (thus $b_i - a_i = hm_i$).
- Volumetric data (e. g. MRT or CT) is mostly acquired slice-wise. Here, the grid size in a slice is usually constant ($h_1 = h_2$) but not orthogonal to the slices ($h_1 \neq h_3$). In this case, the voxels are no cubes.

*Remark* 1.4 (Scanning). Digital images can be acquired by scanning: Let $f : \Omega \to \mathbb{R}$ be a function (signal/image). If $f \in L^2(\Omega)$, a measurement of $f$ is generally described by

$$(f, \psi)_{L^2} = \int_\Omega f(y)\psi(y)\, \mathrm{d}y\,.$$

Here, $\psi \in L^2(\mathbb{R}^d)$ is a test function with $\psi \geq 0$ and $\int_{\mathbb{R}^d} \psi(x)\, \mathrm{d}x = 1$.

A simple approximation of a CCD sensor is, for example, given by the test function

$$r_h(x) = \begin{cases} \frac{1}{h^d} & \|x\|_\infty < \frac{h}{2} \\ 0 & \text{else} \end{cases} \quad \text{and } (f, r_h(x - \cdot))_{L^2} = \int_\Omega f(y) r_h(x - y)\, \mathrm{d}y =: r_x^h[f].$$

For a given grid $\underline{X}$ one gets the discrete image $(r^h_{x^{\underline{j}}}[f])_{\underline{j} \in \mathcal{I}}$.

This approach can also be used to represent a high resolution image with a smaller resolution, which is required, for instance, in so-called multilevel methods.

*Remark* 1.5 (Coordinate systems). We consider the unit square $(0, 1)^2$ as image domain. Mathematically, the origin, i. e. the point $(0, 0)$, is in the "lower left" of the domain. For digital images, when they are displayed on the screen, this is different. The origin pixel, i. e. the pixel corresponding to the index (1,1) (or (0,0) in C indexing), is "top left". Moreover, the first coordinate index corresponds to the vertical axis, the second coordinate index to the horizontal axis. This corresponds to a rotation of the image domain by 90 degrees. This is important, for instance, when expressing angles.

## 1.2. Intensity transformations

In image processing, one distinguishes the following classes of methods, which create a new image from a given image (e. g. a denoised image from a noisy image):

- Point operators / intensity transformations
- Local operators
- Global operators

First, we address point operators, which can be used, for instance, to correct (global) illumination problems.

**Definition 1.6.** Let $f : \Omega \to \mathbb{R}$ be an image and $T : \mathbb{R} \to \mathbb{R}$. $T$ is called *intensity transformation* and $T \circ f$ is called *intensity transformed image.*

(i) For $a < b$, the mapping

$$T^{\text{clip}}_{[a,b]} : \mathbb{R} \to [a, b], s \mapsto \begin{cases} a & s \leq a \\ b & s \geq b \\ s & \text{else} \end{cases}$$

is called *intensity clipping.*

(ii) For $a < b$, the mapping

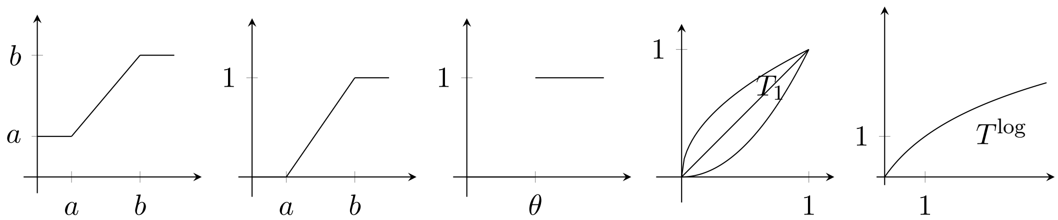
$$T^{\text{norm}}_{[a,b]} : \mathbb{R} \to [0, 1], s \mapsto T^{\text{clip}}_{[0,1]} \left( \frac{s - a}{b - a} \right)$$

is called *intensity normalization.* If $f$ is not constant and $f(\Omega)$ closed and bounded, $\hat{f} = T^{\text{norm}}[f] := T^{\text{norm}}_{[\min f, \max f]} \circ f$ has value range $[0, 1]$ and it holds that $\min \hat{f} = 0$ and $\max \hat{f} = 1$.

(iii) The mapping

$$T^{\text{threshold}}_{\theta} : \mathbb{R} \to \{0, 1\}, s \mapsto \begin{cases} 0 & s \leq \theta \\ 1 & s > \theta \end{cases}$$

is called *thresholding.* $H := T^{\text{threshold}}_0$ is known as *Heaviside function* and also plays a role in areas other than image processing.



(iv) For $\gamma > 0$, the mapping

$$T_\gamma : [0,1] \to [0,1], s \mapsto s^\gamma$$

is called *gamma correction.* For $\gamma < 1$, the mapping is strictly concave, for $\gamma > 1$ it is strictly convex.

(v) The mapping

$$T^{\log} : [0,\infty) \to [0,\infty), s \mapsto \log_2(1+s)$$

is called *log transformation* and is often used to visualize the *power spectrum* (absolute value of the Fourier transform). The base of the logarithm is 2, so that $[0,1]$ is mapped to $[0,1]$ bijectively.

*Remark* 1.7 (Discrete histogram)*.* If $F$ is a discrete image with value range $V$,

$$H_F : V \to \mathbb{N}_0, s \mapsto \#\left\{\underline{j} \in \mathcal{I} \,:\, f_{\underline{j}} = s\right\}$$

is called *(discrete) histogram* of $F$. It states which values occur how often in the image. Using the Kronecker delta

$$\delta_{i,j} := \begin{cases} 1 & i = j \\ 0 & \text{else} \end{cases},$$

the histogram can be expressed as a sum:

$$H_F(s) = \sum_{\underline{j} \in \mathcal{I}} \delta_{s, f_{\underline{j}}}$$

If the value range is $V := \{0,\dots,n\}$, the *(discrete) cumulative distribution function* of $F$ is defined as

$$G_F : \{0,\dots,n\} \to \mathbb{N}_0, s \mapsto \sum_{r=0}^{s} H_F(r).$$

The histogram can be easily extended to pixelated images. To this end, we first introduce some additional notation.

**Definition 1.8.**

(i) For a set $A \subset \mathbb{R}^d$,

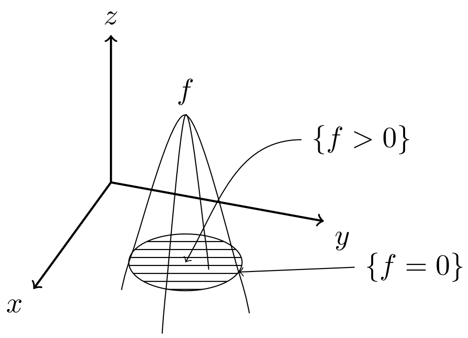
$$\chi_A : \mathbb{R}^d \to \{0,1\}, x \mapsto \begin{cases} 1 & x \in A \\ 0 & \text{else} \end{cases}$$

is called *characteristic function* of $A$.

(ii) Let $\Omega \subset \mathbb{R}^d$ be a domain and $f : \Omega \to \mathbb{R}$. For $s \in \mathbb{R}$,

$$\{f = s\} := \{x \in \Omega \,:\, f(x) = s\}$$

is called *s-level set* of $f$. Analogously, one defines $\{f \le s\}$, the *s-sublevel set* of $f$ and $\{f \ge s\}$, the *s-superlevel set* of $f$. We have $f^{-1}(s) = \{f = s\}$, $f^{-1}((-\infty, s]) = \{f \le s\}$ and $f^{-1}([s,\infty)) = \{f \ge s\}$. $\{f = s\}$ is also called *level line/isoline* ($d = 2$) or *level surface/isosurface* ($d = 3$).




(iii) For a bounded set $A \subset \mathbb{R}^d$,

$$|A| := \mathrm{Vol}(A) := \int_A \mathrm{d}x = \int_{\mathbb{R}^d} \chi_A(x)\,\mathrm{d}x$$

denotes the *volume* of $A$. In particular, $\mathrm{Vol}(A) = \lambda(A)$, where $\lambda$ is the Lebesgue measure.

*Remark* 1.9 (Histogram)*.* Let $f$ be a pixelated image and $F$ the corresponding discrete image. Then,

$$H_F(s) = \sum_{\underline{j} \in \mathcal{I}} \frac{1}{\mathrm{Vol}(c_{\underline{j}})} \int_{c_{\underline{j}}} \chi_{\{f=s\}}(x)\,\mathrm{d}x = \frac{1}{\prod h_i} \int_\Omega \chi_{\{f=s\}}(x)\,\mathrm{d}x = \frac{1}{\prod h_i} \mathrm{Vol}(\{f = s\}).$$

Except for the scaling with the cell volume, the discrete histogram is equal to the volume of the level sets. The latter is a property that can also be considered on images that are not pixelated.

This motivates to consider the following mapping as basis for a continuous histogram:

$$H_f : \mathbb{R} \to [0, \infty), s \mapsto \mathrm{Vol}(\{f = s\})$$

Since all level sets of a continuous image could be Lebesgue null sets (e. g. $f(x) = x$ on $[0, 1]$), we extend $H_f$ from $\mathbb{R}$ as domain to subsets of $\mathbb{R}$ as domain, specifically to $\mathcal{B}(\mathbb{R})$, the set of Borel measurable subsets of $\mathbb{R}$:

$$H_f : \mathcal{B}(\mathbb{R}) \to [0, \infty), A \mapsto \mathrm{Vol}(\{f \in A\}).$$

Thus, $H_f$ is a (positive) measure. In particular,

$$H_f(A) = \mathrm{Vol}(\{f \in A\}) = \mathrm{Vol}(f^{-1}(A)) = \lambda(f^{-1}(A)).$$

Therefore, $H_f$ is the push-forward measure (or image measure) of the Lebesgue measure under the mapping $f$, cf. Proposition A.8. For $g : \mathbb{R} \to \mathbb{R}$, we have

$$\int_{\mathbb{R}} g(s)\mathrm{d}H_f(s) = \int_\Omega (g \circ f)(x)\,\mathrm{d}x,$$

provided $f : \Omega \to \mathbb{R}$ is $\mathcal{B}(\Omega)$-$\mathcal{B}(\mathbb{R})$ measurable and $g$ is $f(\lambda)$-integrable, cf. Proposition A.9.

The *cumulative distribution function* of $f$ is defined as

$$G_f(s) = \mathrm{Vol}(\{f \leq s\}).$$

The change of this function at a position $s$ is also a measure for the histogram of $f$ at $s$. However, $G_f$ is not necessarily differentiable in the classical sense. For example, for pixelated images,

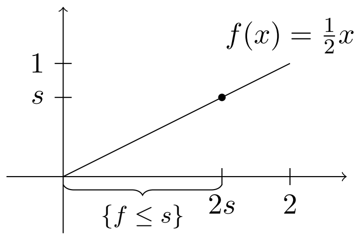
$G_f$ is a step function and thus not differentiable. One can show that $H_f$ is the distributional derivative of $G_f$ (Exercise).

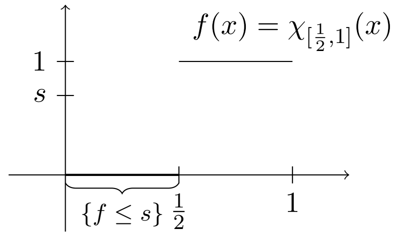
Examples:

- For $f : (0,2) \to [0,1], x \mapsto \frac{1}{2}x$, we have $G_f : [0,1] \to [0,2], s \mapsto 2s$ and $G_f'(s) = 2$.
- For $f : (0,1) \to [0,1], x \mapsto \chi_{[\frac{1}{2},1]}(x)$, we have

$$G_f(s) = \begin{cases} 0 & s < 0 \\ \frac{1}{2} & 0 \le s < 1 \\ 1 & s \ge 1 \end{cases} \text{ and } H_f = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1.$$

Here, $\delta_x$ denotes the Dirac delta function centered at $x$.




If $f$ is pixelated with value range $V \coloneqq \{0, \dots, n\}$ and $F$ the corresponding discrete image, then

$$G_f(s) = \mathrm{Vol}\left(\bigcup_{r=0}^{s} \{f = r\}\right) = \sum_{r=0}^{s} \mathrm{Vol}\left(\{f = r\}\right) = \left(\prod h_i\right) \sum_{r=0}^{s} H_F(r) = \left(\prod h_i\right) G_F(s).$$

Except for the scaling with the cell volume, the discrete cumulative distribution function is equal to the volume of the sublevel sets.

*Remark* 1.10 (Binning). The histogram of a continuous function can be approximated discretely in a straightforward manner. In the same way, a discrete histogram can be approximated. To this end, the value range $V$ is partitioned into intervals (so-called *bins*).

Choose $m \in \mathbb{N}$ and $r_0 < r_1 < \ldots < r_m \in \mathbb{R}$. Then, the bins are defined as

$$B_k = \begin{cases} [r_{k-1}, r_k) & k \in \{1, \dots, m-1\} \\ [r_{k-1}, r_k] & k = m \end{cases}$$

and the binned discrete histogram of $f : \Omega \to \mathbb{R}$ is

$$H_k[f] = \mathrm{Vol}(\{f \in B_k\}) \text{ for all } k \in \{1, \dots, m\}.$$

If the value range of $f$ is bounded, a typical choice of the interval boundaries is

$$r_k = \begin{cases} \inf f + \frac{k}{m-1}(\sup f - \inf f) & k \in \{0, \dots, m-1\} \\ \sup f + 1 & k = m \end{cases}.$$

As example, we consider the image

$$F = \begin{pmatrix} 1 & 2 & 1 & 2 \\ 4 & 3 & 4 & 2 \\ 7 & 1 & 4 & 1 \end{pmatrix}$$

on a cartesian grid with grid size 1 and the following bins

$$B_1 = [1, 3),\ B_2 = [3, 5),\ B_3 = [5, 7),\ B_4 = [7, 8].$$

Thus, we get

$$H_1 = 7,\ H_2 = 4,\ H_3 = 0,\ H_4 = 1.$$

*Example* 1.11 (Histogram equalization). If the contrast of an image $f$ is still not sufficient after a normalization with $T^{\text{norm}}$, the contrast can potentially be further improved using the histogram:

Here, the idea is to distribute the gray values as uniformly as possible in the value range $[0, 1]$. Let now $f : \Omega \to [0, 1]$. We look for a bijective, increasing intensity transformation $T : [0, 1] \to [0, 1]$ such that

$$G_{T\circ f}(s) = s \operatorname{Vol}(\Omega) \text{ for all } s \in [0, 1]. \tag{$*$}$$

Then, $G'_{T\circ f} \equiv \operatorname{Vol}(\Omega)$, i.e. the gray values are distributed uniformly. If $T$ fulfills $(*)$, we get

$$T(s) \operatorname{Vol}(\Omega) \overset{(*)}{=} G_{T\circ f}(T(s)) = \operatorname{Vol}(\{T \circ f \leq T(s)\}) \overset{T \text{ bijective, increasing}}{=} \operatorname{Vol}(\{f \leq s\}) = G_f(s).$$

If $f$ is such that $G_f \in C([0, 1], [0, \operatorname{Vol}(\Omega)])$, $G_f$ is strictly increasing and $G_f(0) = 0$, then

$$T : [0, 1] \to [0, 1], s \mapsto \frac{G_f(s)}{\operatorname{Vol}(\Omega)}$$

has the desired properties: The continuity and strict monotonicity of $G_f$ combined with $G_f(0) = 0$ imply the invertibility of $G_f$ on $G_f([0, 1]) = [0, \operatorname{Vol}(\Omega)]$. Thus, $T^{-1}(s) = G_f^{-1}(s \operatorname{Vol}(\Omega))$ and we get

$$G_{T\circ f}(s) = \operatorname{Vol}(\{T \circ f \leq s\}) = \operatorname{Vol}(\{f \leq T^{-1}(s)\}) = G_f(T^{-1}(s)) = s \operatorname{Vol}(\Omega).$$

If the value range is discrete, we cannot create a uniformly distributed histogram (since a gray value can only be mapped to exactly one gray value), but the idea can still be applied in this case. Let $f : \Omega \to V$ be a pixelated image with $V := \{0, \dots, n\}$ and $F$ the corresponding discrete image. Moreover, let $\operatorname{rd} : \mathbb{R} \to \mathbb{Z}$ denote the rounding mapping, i.e. a real number is mapped to its nearest integer (rounding up if there are two nearest integers). More precisely,

$$\operatorname{rd}(x) = \begin{cases} \lfloor x \rfloor & x - \lfloor x \rfloor < \frac{1}{2} \\ \lfloor x \rfloor + 1 & \text{else} \end{cases},$$

where

$$\lfloor x \rfloor := \max \{k \in \mathbb{Z} : k \leq x\}$$

is the floor function. Then, one gets a discrete histogram equalization using the intensity transformation

$$T(s) = \operatorname{rd}\left(\frac{n}{\operatorname{Vol}(\Omega)} G_f(s)\right) = \operatorname{rd}\left(\frac{n}{\#\mathcal{I}} G_F(s)\right).$$

Using the first equality, the idea can also be applied to non-pixelated images with discrete value range.

*Remark* 1.12 (Histogram matching). If images are taken from the same scene but with different sensors, the histograms can be used to compensate sensor-dependent intensity differences.

Let $f, g : \Omega \to [0,1]$ be such that $G_f$ and $G_g$ are invertible. Moreover, let

$$T(s) = \frac{G_f(s)}{\mathrm{Vol}\,(\Omega)} \text{ and } S(s) = \frac{G_g(s)}{\mathrm{Vol}\,(\Omega)}$$

be the intensity transformations for histogram equalization. With those, the histogram of one of the images can be transferred to the other image: We have

$$\begin{aligned} G_{S^{-1}\circ T\circ f}(s) &= \mathrm{Vol}(\{S^{-1} \circ T \circ f \le s\}) = \mathrm{Vol}(\{T \circ f \le S(s)\}) = G_{T\circ f}(S(s)) \\ &= S(s)\,\mathrm{Vol}(\Omega) = G_g(s). \end{aligned}$$

Thus, the histograms of $S^{-1} \circ T \circ f$ and $g$ coincide.

Note that above we did not use that $G_g$ is the cumulative distribution function of an image $g$ but just the bijectivity of $G_g$ as mapping from $[0,1]$ to $[0, \mathrm{Vol}(\Omega)]$. Thus, this idea allows transferring an arbitrary, bijective, increasing distribution function to $f$.

*Example* 1.13 (Segmentation by thresholding). We consider an image $f : \Omega \to [0,1]$ of a handwritten text (e. g. scan of a note or a photo of a blackboard). For such an image, it is useful to reduce the number of gray values to two, i. e. to segment the image. This compensates differences in illumination and reduces noise. Moreover, the data size is reduced significantly. The segmentation can be achieved using $T_\theta^{\text{threshold}}$, but requires a threshold $\theta$ to be chosen.

$T_\theta^{\text{threshold}}$ partitions the image into two components, $\Sigma = \{f \le \theta\}$ and $\Omega \setminus \Sigma = \{f > \theta\}$. Now, the idea is to chose the threshold $\theta$ in such a way that it corresponds to the average of the average gray values in the two components, i. e.

$$\theta = \frac{1}{2}\left(\fint_{\{f\le\theta\}} f(x)\,\mathrm{d}x + \fint_{\{f>\theta\}} f(x)\,\mathrm{d}x\right) =: \varphi(\theta),$$

where

$$\fint_A f(x)\,\mathrm{d}x = \frac{1}{\mathrm{Vol}(A)}\int_A f(x)\,\mathrm{d}x\,.$$

In case $\mathrm{Vol}(\{f \le \theta\}) = 0$ or $\mathrm{Vol}(\{f > \theta\}) = 0$, we define

$$\fint_{\{f\le\theta\}} f(x)\,\mathrm{d}x = \theta \text{ and } \fint_{\{f>\theta\}} f(x)\,\mathrm{d}x = \theta, \text{ respectively.}$$

This equation can be solved with the fixed point iteration

$$\theta^{n+1} = \varphi(\theta^n) \text{ for } n \in \mathbb{N},$$

given an initial value $\theta^1 \in (0,1)$. A typical initial value is $\theta^1 = \frac{1}{2}(\inf f + \sup f)$. This method to determine the threshold is called *isodata algorithm*.

Since $f(\Omega) \subset [0,1]$, we have $\varphi([0,1]) \subset [0,1]$. Moreover, for $\theta_1 \leq \theta_2$,

$$
\begin{aligned}
\varphi_a(\theta_2) &:= \fint_{\{f\leq\theta_2\}} f(x)\,\mathrm{d}x = \frac{1}{|\{f\leq\theta_2\}|}\left(\int_{\{f\leq\theta_1\}} f(x)\,\mathrm{d}x + \int_{\{\theta_1<f\leq\theta_2\}} f(x)\,\mathrm{d}x\right) \\
&\geq \frac{1}{|\{f\leq\theta_2\}|}\left(\int_{\{f\leq\theta_1\}} f(x)\,\mathrm{d}x + \theta_1\,|\{\theta_1<f\leq\theta_2\}|\right) \\
&\geq \frac{1}{|\{f\leq\theta_2\}|}\left(\int_{\{f\leq\theta_1\}} f(x)\,\mathrm{d}x + \fint_{\{f\leq\theta_1\}} f(x)\,\mathrm{d}x\,|\{\theta_1<f\leq\theta_2\}|\right) \\
&= \frac{1}{|\{f\leq\theta_2\}|}\left(|\{f\leq\theta_1\}| + |\{\theta_1<f\leq\theta_2\}|\right)\fint_{\{f\leq\theta_1\}} f(x)\,\mathrm{d}x \\
&= \fint_{\{f\leq\theta_1\}} f(x)\,\mathrm{d}x = \varphi_a(\theta_1).
\end{aligned}
$$

This means that $\varphi_a$ is increasing. Analogously, one shows that $\varphi_b(\theta) := \fint_{\{f>\theta\}} f(x)\,\mathrm{d}x$ is increasing. Thus, as sum of two increasing functions, $\varphi$ is also increasing. Let $\theta^1 \in [0,1]$ be arbitrary but fixed. If $\theta^2 = \varphi(\theta^1) \geq \theta^1$, we get $\theta^{n+1} \geq \theta^n$, i. e. the sequence is increasing. Otherwise, we have $\theta^2 < \theta^1$ and we get $\theta^{n+1} \leq \theta^n$, i. e. the sequence is decreasing. In both cases, the sequence is monotonic and bounded, thus convergent.

Moreover, one can show: Every monotonic function $\varphi : [0,1] \to [0,1]$ has a fixed point (exercise). In case $\varphi$ is not continuous, the limit of the sequence $(\theta^n)_{n\in\mathbb{N}}$ is not necessarily a fixed point of $\varphi$ though.

The average gray values can be computed with the histogram. We have (exercise)

$$
\fint_{\{f\leq\theta\}} f(x)\,\mathrm{d}x = \int_{\mathbb{R}} s\chi_{(-\infty,\theta]}(s)\mathrm{d}H_f(s) \Big/ \int_{\mathbb{R}} \chi_{(-\infty,\theta]}(s)\mathrm{d}H_f(s),
$$

$$
\fint_{\{f>\theta\}} f(x)\,\mathrm{d}x = \int_{\mathbb{R}} s\chi_{(\theta,\infty)}(s)\mathrm{d}H_f(s) \Big/ \int_{\mathbb{R}} \chi_{(\theta,\infty)}(s)\mathrm{d}H_f(s).
$$

If $F$ is a discrete image with value range $V := \{0,\dots,n\}$, we have

$$
\varphi(\theta) = \frac{1}{2}\left(\sum_{s=0}^{\lfloor\theta\rfloor} sH_F(s) \Big/ \sum_{s=0}^{\lfloor\theta\rfloor} H_F(s) + \sum_{s=\lfloor\theta\rfloor+1}^{n} sH_F(s) \Big/ \sum_{s=\lfloor\theta\rfloor+1}^{n} H_F(s)\right)
$$

and the fixed point iteration can be implemented using this expression.