Kapitel 2

Grundlagen der Mustererkennung

Was ist Mustererkennung?

Mustererkennung ist die Theorie der bestmöglichen Zuordnung eines unbekannten Musters oder Beobachtung $\mathbf{z}_i$ zu einer ***Bedeutungs-*** oder ***Äquivalenzklasse*** $\mathcal{E}_j$ (Klassifikation).

Eine Äquivalenzklasse $\mathcal{E}$ besteht aus einer Menge von Mustern $\{\mathbf{x}_i\}$ und einer zweistelligen Verknüpfung (Äquivalenzrelation) mit den folgenden drei Eigenschaften:

a) $\mathbf{x}_i \sim \mathbf{x}_i$ reflexiv (jedes Element ist zu sich selbst äquivalent)

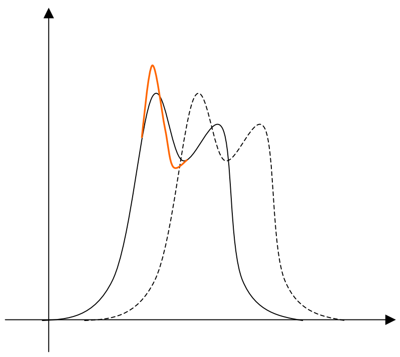
b) $\mathbf{x}_i \sim \mathbf{x}_j \Rightarrow \mathbf{x}_j \sim \mathbf{x}_i$ symmetrisch

c) $(x \sim y) \& (y \sim z) \Rightarrow x \sim z$ transitiv

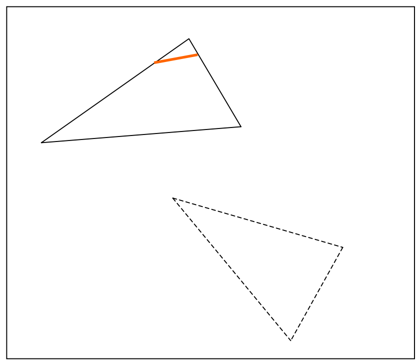
$\mathbf{x}_i \sim \mathbf{x}_j$: d.h. $\mathbf{x}_i$ ist äquivalent zu $\mathbf{x}_j$ in Bezug auf die Relation ~

Relevante and irrelevante Änderungen in Signalen und Bildern

1D



2D



Bedeutungs- oder Äquivalenzklassen

Äquivalenzklassen $\mathcal{E}$ können durch zwei Arten definiert werden, nämlich

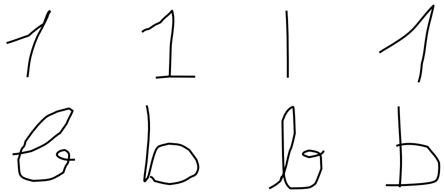
1) Angabe aller Repräsentanten von $\mathcal{E}$, da die Veränderungen nicht systematisch formuliert werden können, oder:

2a) Durch ein erzeugendes Element $\mathbf{x}_0$ und eine mathematische Gruppe $\mathcal{G}$ (Abgeschlossenheit der Daten)

2b) Abgeschlossene Abbildung mit anschließender Abbildung auf einen Unterraum (Projektion, Okklusion), z.B. Bewegung eines 3D-Objektes im Raum mit anschließender Projektion auf die Kameraebene

Beispiel für 1):

Die Menge aller handgeschriebenen Buchstaben oder Ziffern:



Hier ist eine parametrische Beschreibung der Äquivalenzklasse praktisch unmöglich.

Äquivalenzklasse für den Druckbuchstaben A

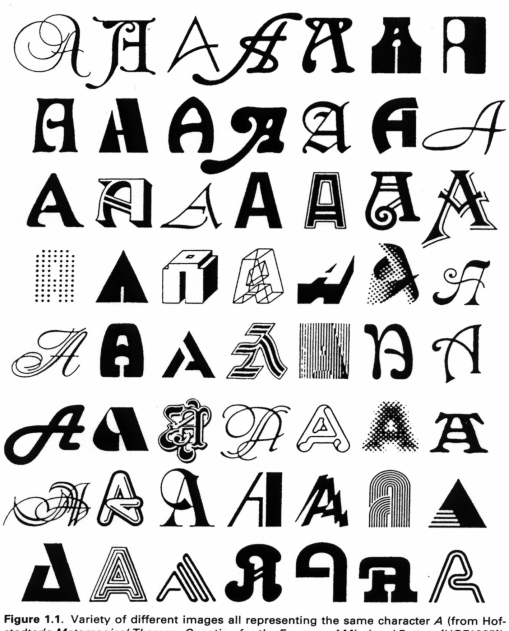
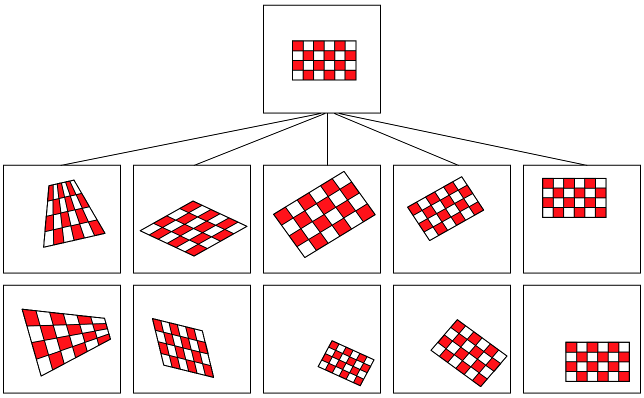


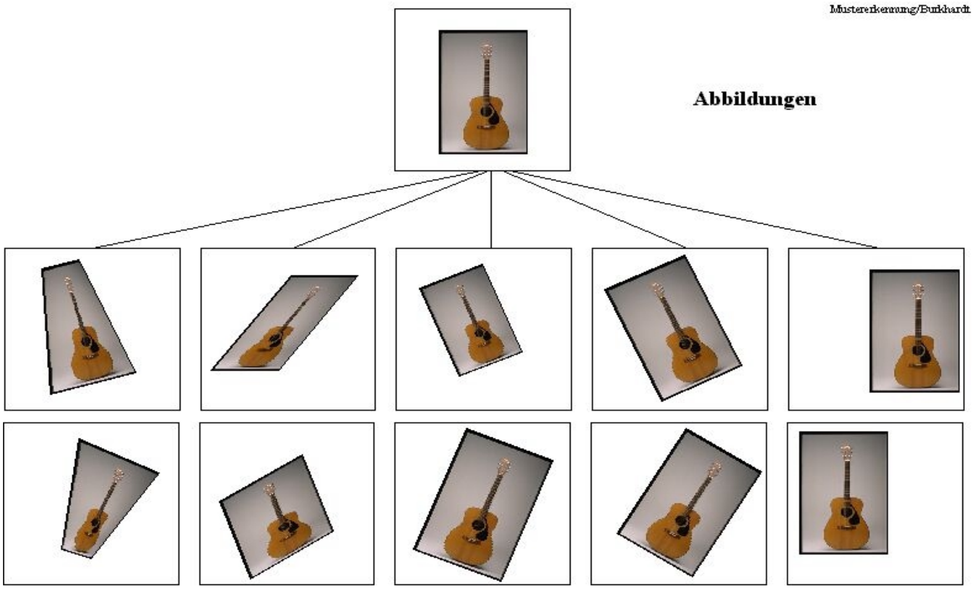
Figure 1.1. Variety of different images all representing the same character *A* (from Hofstadter's *Metamagical Themes: Questing for the Essence of Mind and Pattern* [HOF1985]).

Beispiel für 2a: Geometrische Transformationen



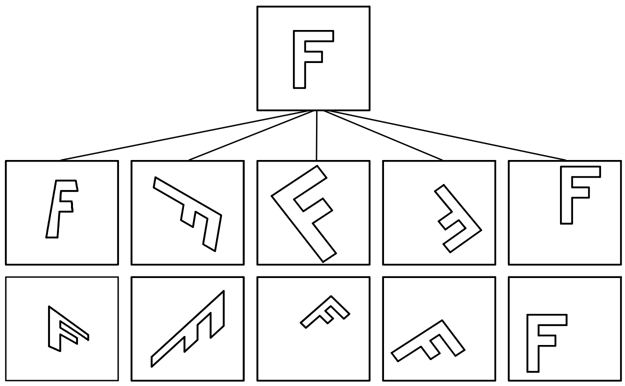
Zentral-projektionen | **Affine Abbildungen** (paralleltreu) | **Ähnlichkeiten** (winkeltreu) | **Kongruenzen** | **Translationen**

Abbildungen



Zentralprojektionen | **Affine Abbildungen** (paralleltreu) | **Ähnlichkeiten** (winkeltreu) | **Kongruenzen** | **Translationen**

Geometrische Transformationen



Zentralprojektionen | **Affine Abbildungen** (erhält Parallelitäten) | **Ähnlichkeiten** (erhält Winkel) | **Kongruenzen** | **Translationen**

Beispiel für 2b: Bewegung im Raum (Translation und Rotation) mit anschliessender Abbildung auf einen Unterraum; unvollständige Beobachtungen

Gruppe

Def.: Eine algebraische Struktur $\mathcal{G}$ mit einer zweistelligen inneren Verknüpfung • heißt Gruppe, wenn für beliebige Elemente $a,b,c \in \mathcal{G}$ folgende Gesetze gelten:

1) $a \bullet (b \bullet c) = (a \bullet b) \bullet c$ assoziativ
2) Es existiert ein ***Einselement*** $e \in \mathcal{G}$, mit
   $a \bullet e = e \bullet a = a \quad \forall a \in \mathcal{G}$
3) Zu jedem a gibt es ein ***inverses Element*** $a^{-1} \in \mathcal{G}$, mit:
   $a \bullet a^{-1} = a^{-1} \bullet a = e$

Aus 1)-3) folgt, daß es genau ein Einselement und zu jedem $a \in \mathcal{G}$ genau ein inverses Element $a^{-1} \in \mathcal{G}$ gibt.

Für eine abelsche Gruppe gilt die Kommutativeigenschaft:

$$a \bullet b = b \bullet a \qquad \forall a,b \in \mathcal{G}$$

Beispiel für 2a):

Die Gruppe $\mathcal{G}(\mathbf{p})$ der geometrische Abbildungen, welche die Bewegung von Objekten, einschließlich projektiver Abbildungen charakterisieren. Man erhält eine Äquivalenzklasse durch Angabe eines erzeugenden Elementes $\mathbf{x}_0$ sowie einer Bewegungsgruppe, welche auch parametrisiert mit dem Vektor **p** beschrieben werden kann. Dies kann z. B. die Gruppe der ebenen (Euklidschen) Bewegungen (Translation und Rotation):

$$\mathcal{E}_{\mathcal{G}}(\mathbf{x}_0) := \{g_i(\mathbf{x}_0) \mid \forall g_i \in \mathcal{G}\}$$

Für zwei Objekte **x**,**y** der selben Äquivalenzklasse gilt demnach:

$$x \overset{\mathcal{G}}{\sim} y \Leftrightarrow \exists g_i \in \mathcal{G}: x = g_i(y)$$

Affine Transformation bzgl. der zweidimensionalen Koordinaten $\mathbf{t}=(t_0,t_1)$:

$$\boxed{\mathbf{t'} = \mathbf{At} + \mathbf{a}}$$

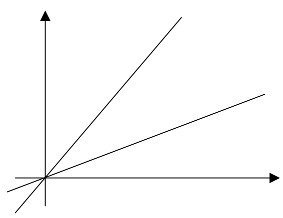
mit:

$\mathbf{A}=\mathbf{I}$	die Gruppe der Translationen	$\mathcal{T}$	(dim($\mathbf{p}$)=2)
$\mathbf{A}^T\mathbf{A}=\mathbf{I}$	die Gruppe der Kongruenzen	$\mathcal{C}$	(dim($\mathbf{p}$)=3)
$\mathbf{A}^T\mathbf{A}=k\mathbf{I}$	die Gruppe der Ähnlichkeiten	$\mathcal{S}$	(dim($\mathbf{p}$)=4)
$\det(\mathbf{A})\neq 0$	die Gruppe der affinen Abbildungen	$\mathcal{A}$	(dim($\mathbf{p}$)=6)

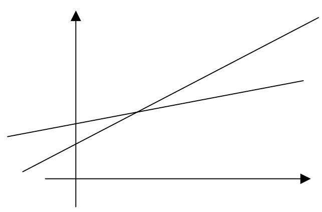
Die affine Abbildung

$$\boxed{\mathbf{t'} = \mathbf{At} + \mathbf{a}}$$

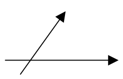
6 Freiheitsgrade



lineare (homogene) Abbildungen



affine Abbildungen

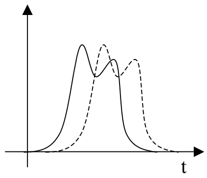


Affine (schiefwinklige) Koordinaten, als Verallgemeinerung der kartesischen

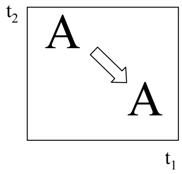
Die Gruppe der Translationen $\mathcal{T}$ für kontinuierlich definierte Signale und Bilder

Die Menge der Translationen $\{\tau\}$ bilden hinsichtlich der Verknüpfung durch Zusammensetzung $\tau_1(\tau_2(...))=(\tau_1 \bullet \tau_2)(...)$ eine abelsche Gruppe.

1D



2D



$$x(t') = x(t-a)$$

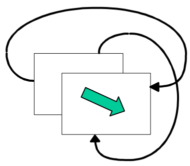
$$X(t_1', t_2') = X(\mathbf{t})$$
$$= X(t_1' - a_1, t_2' - a_2) = X(\mathbf{t} - \mathbf{a})$$

Die Gruppe der Translationen $\mathcal{T}$ als abgeschlossene Operation

Auf ***unendlich*** ausgedehnten Koordinaten lässt sich die Translation natürlich als abgeschlossene Operation definieren. Die Abgeschlossenheit ist erforderlich, damit nicht Elemente bei der Operation verschwinden und andere hinzukommen. Man verlangt:

$$\mathbf{x} \in \mathcal{X} \Rightarrow \tau(\mathbf{x}) \in \mathcal{X} \qquad \forall \tau \in \mathcal{X}$$

Die Abgeschlossenheit der Daten bei einer Translation, angewendet auf Signale oder Bilder, welche nur auf einem ***endlichen*** Bereich definiert sind (Signal- oder Bildfenster), erreicht man durch ***zyklisches Verschieben***. Übertragen auf unendlich ausgedehnte Koordinaten ließe sich dies auch durch eine periodische Fortsetzung eines endlichen Definitionsbereichs oder Fensters erreichen.
Eine mathematische Gruppe garantiert die Abgeschlossenheit, da ein inverses Element exisitiert!



Die Gruppe der Translationen $\mathcal{T}$ für abgetastete endliche Signale und Bilder

Abgetastete endliche Muster:

$$\begin{aligned} \mathbf{x} &:= \{x_i\} \qquad & i &= 0, \cdots, N-1 \\ \mathbf{X} &:= \{X_{i,j}\} \qquad & i, j &= 0, \cdots, N-1 \end{aligned}$$

Translation als zyklische Permutation:

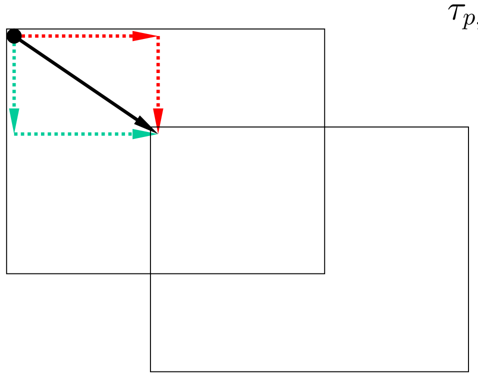
$$\begin{aligned} \tau_k(\mathbf{x}) &:= \{x_{(i+k)\mathrm{mod}N}\} = \{x_{<i+k>_N}\} \\ \tau_{k,l}(\mathbf{X}) &:= \{X_{<i+k>_M, <j+l>_N}\} \end{aligned}$$

Beispiel:

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \\ 13 & 14 & 15 & 16 \end{bmatrix} \Rightarrow \tau_{2,1}(\mathbf{A}) = \begin{bmatrix} 10 & 11 & 12 & 9 \\ 14 & 15 & 16 & 13 \\ 2 & 3 & 4 & 1 \\ 6 & 7 & 8 & 5 \end{bmatrix}$$

Die 2D-Translation kann in zwei 1D-Translationen faktorisiert werden (Zeilen/Spaltenpermutation):

$$\tau_{p,r} = \tau_p \bullet \tau_r = \tau_r \bullet \tau_p$$



Die zyklische Permutation von Zeilen und Spalten einer Bildmatrix mit Hilfe von Permutationsmatrizen:

$$\tau_{1,1}(\mathbf{A}) = \overbrace{\mathbf{P}^T\mathbf{A}}^{\text{Spaltenpermutation}}\!\!\!\!\underbrace{\mathbf{P}}_{\text{Zeilenpermutation}} \quad \text{bzw:} \quad \tau_{p,r}(\mathbf{A}) = (\mathbf{P}^p)^T\mathbf{A}\mathbf{P}^r$$

Die Permutationsmatrix ist orthogonal und somit gilt: $\mathbf{P}^{-1} = \mathbf{P}^T$

So zum Beispiel für N=4:

$$\mathbf{P} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix} \quad \text{mit:} \quad \mathbf{P}^0 = \mathbf{P}^4 = \mathbf{I}$$

Die Gruppe der Kongruenzen $\mathcal{C}$ für kontinuierlich definierte Bilder

Die Gruppe der Kongruenzen entstehen durch Translation und Rotation, was man auch mit Euklidscher Bewegung bezeichnet.

$$\mathbf{t'} = \mathbf{At} + \mathbf{a}$$

$$\mathbf{A} = \begin{pmatrix} \cos(\varphi) & \sin(\varphi) \\ -\sin(\varphi) & \cos(\varphi) \end{pmatrix}; \quad \mathbf{a} \neq \mathbf{0}$$

Die Drehmatrix **A** ist orthogonal und es gilt:

$$\mathbf{A}^{-1} = \mathbf{A}^T \text{ sowie: } det(\mathbf{A}) = \mathbf{1}$$

Die Gruppe der Ähnlichkeiten $\mathcal{S}$ für kontinuierlich definierte Bilder

Die Gruppe der Ähnlichkeiten entsteht durch Verschieben, Drehen und Stauchen:

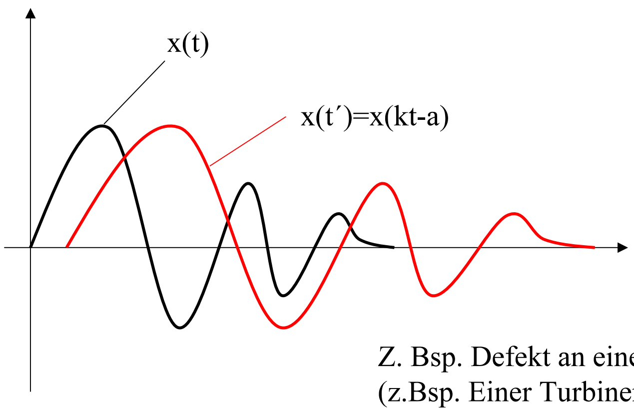
$$\mathbf{t'} = \mathbf{At} + \mathbf{a}$$

$$\mathbf{A} = k\begin{pmatrix} \cos(\varphi) & \sin(\varphi) \\ -\sin(\varphi) & \cos(\varphi) \end{pmatrix}; \quad \mathbf{a} \neq \mathbf{0}$$

Daraus folgt: $\mathbf{A}\mathbf{A}^T = k^2\mathbf{I}$ sowie: $det(\mathbf{A}) = k^2$

Sowie: $\mathbf{A}^{-1} = \frac{1}{k^2}\mathbf{A}^T$

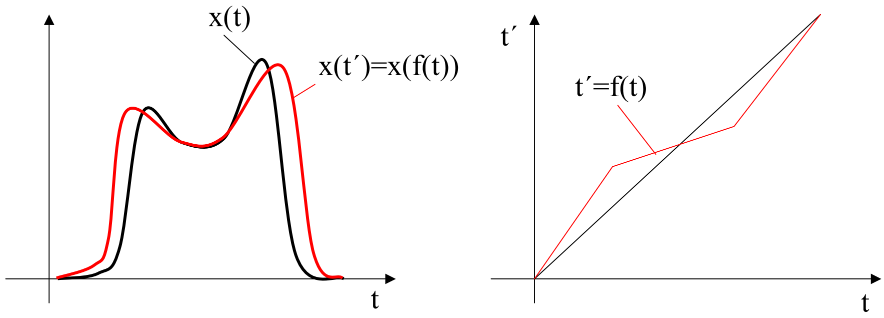
Z. Bsp.: stauchen und verschieben eines eindimensionalen Signals



Z. Bsp. Defekt an einem rotierenden Teil (z.Bsp. Einer Turbinenschaufel) bei unterschiedlicher Drehzahl (oder sogar zeitvarianter Drehzahl (siehe unten).

Allgemeinere Äquivalenzklassen: Beliebige Zeitmodulationen

$x(t') = x(f(t))$ $f(t)$ monoton (Kausalität, Zeit läuft nicht zurück), aber sonst beliebig



Ähnlichkeiten im Sinne einer verallgemeinerten Metrik!

Allgemeinere Äquivalenzklassen: Beliebige Zeitmodulationen

- z. Bsp.: unterschiedliche Sprechergeschwindigkeit bei der Spracherkennung oder variierende Tempi bei der Musikerkennung
- Dieser Fall ist schwieriger als das Beispiel mit der Turbinenschaufel, da man ja dort einen Drehzahlmesser anbringen könnte und die Zeitschwankungen korrigieren könnte