Sensor Signal Processing

Prof. Dr.-Ing. Andreas König
Lehrstuhl Integrierte Sensorsysteme

FB Elektrotechnik und Informationstechnik
Technische Universität Kaiserslautern

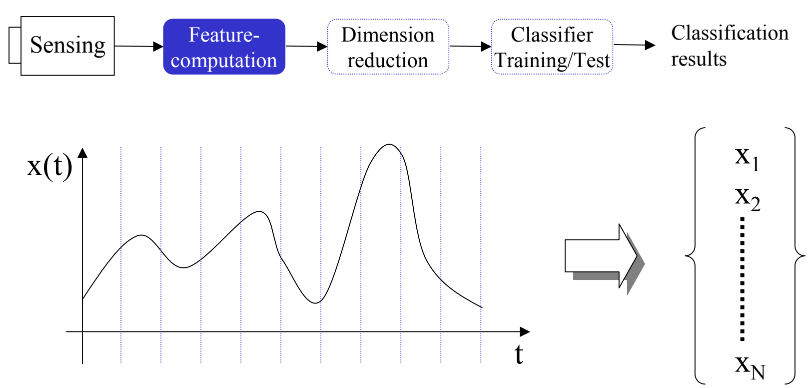
Fall Semester 2005

Course Contents

Chapter Contents

Motivation

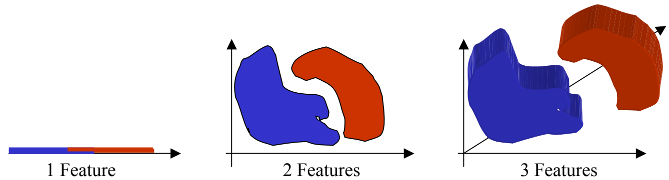
- Signal processing helps to condition sensor signals by signal to signal transforms
- Feature computation extracts & condenses information by heuristic techniques using iconic to symbolic transform



- Large number of potential approaches !

Motivation

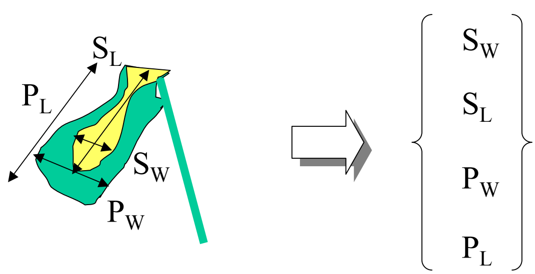
- Computed features shall provide a compact & invariant description
- Sufficient features must be computed from one or combination of several methods with optimum parameter settings



- This combination of several heuristic methods can generate a sufficient, but necessarily minimum number of features
- Redundance & irrelevance in the feature set must be eliminated (*Curse of dimensionality, computational complexity*)
- Dimensionality reduction: (un)supervised, (non)linear methods

Geometrical Features

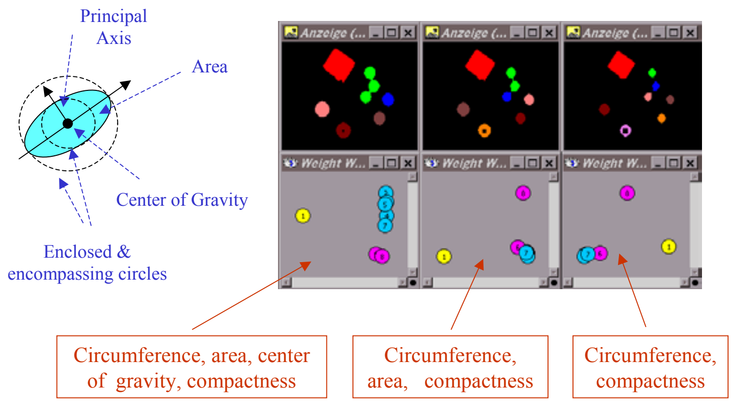
- In practical cases, measurement values can immediately serve as significant variables or features for problem description
- Example: Medical database with diagnostic and laboratory information
- Very often, geometrical information can serve as immediate feature for classification
- Example: Iris data



- Setal and petal widths and lengths contribute 4 features

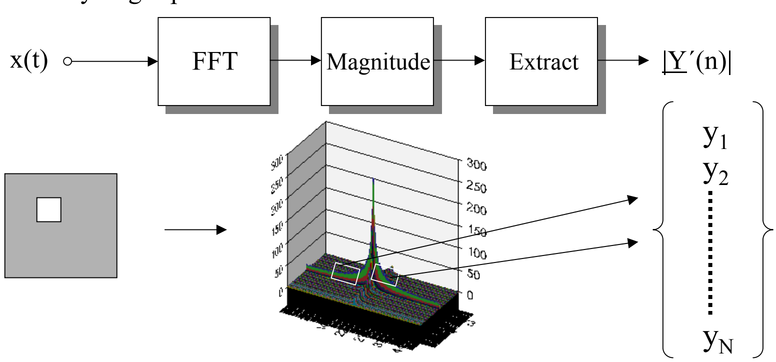
Geometrical Features

- Example of geometric feature computation & feature elimination issue approach



Spectral Features

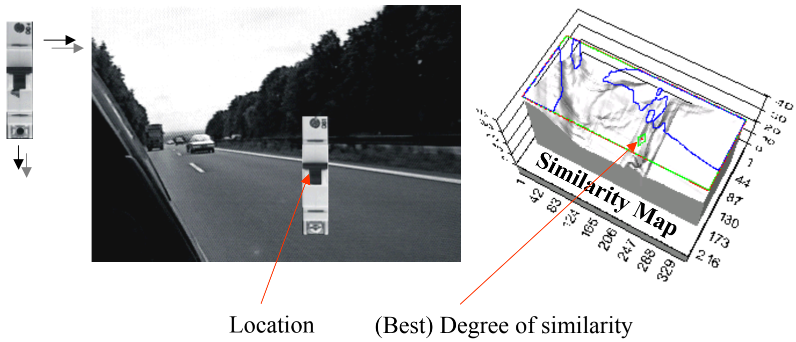
- FFT computation can be exploited to extract features enjoying inherent invariance properties
- Potentially large spectra can be reduced to several coefficients of interest



- The choice of single, isolated coefficients or frequency usually is extremely sensitive to shifts in the spectral composition
- Context knowledge, e.g., revolutions/s, and coefficient windows are employed instead trading compactness for robustness

Correlation Features

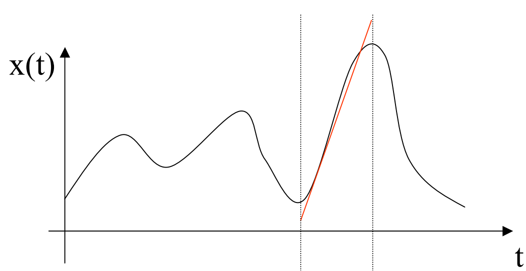
- Correlation serves multiple purpose in signal and image processing
- Single object is separated from background using, e.g., template correlation
- Correlation of a signal with one or several prototype signals or templates can also serve for feature computation



Model Matching

- The signal information can be condensed in the parameters of one (global) or several (local) models
- Simplest model: linear (regression) model

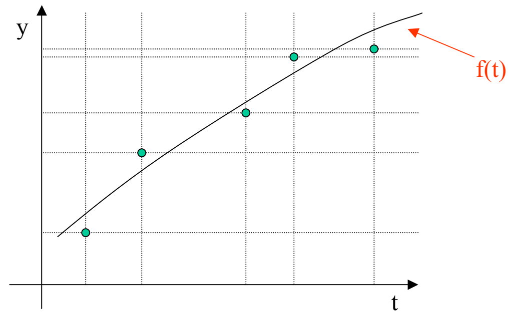
$$y(t) = m \cdot t + b \tag{3.1}$$



- Description of a signal (section) by the parameters m and b
- Similar approach to function approximation to obtain model for transfer function

Model Matching

- The general case of a limited number, not necessarily equidistantly sampled measurements is depicted by the following sketch:



- A model f(t) can be constructed to achieve a compact description by the model parameters
- Further, the system behavior between samples (y,t) is modeled, too
- The linear case is a special case !

Model Matching

- The determination of such a model function can take place by interpolation or approximation
- In the case of interpolation function f(x) assumes all values of the sample positions, i.e.

$$f(t_i) = y_i \tag{3.2}$$

- Due to common noise contributions in the signal data, interpolation will lead to bad generalization behavior for f(t)
- Better smoothness and generalisation properties desirable
- In the case of approximation according to an assessement function f(t) assumes values close to the measured values, but not necessarily coinciding with those
- Common assessment measures or fitting criterias of f(t) for given sample data are:

$$\sum_{i=1}^{N} \left| f(t_i) - y_i \right| = \min ! \tag{3.3}$$

Model Matching

- This criterion is denoted as Manhattan distance or L_1-Approximation Euclidean Distance or L_2-approximation:

$$\sum_{i=1}^{N}\left(f(t_i)-y_i\right)^2 = \min ! \qquad (3.4)$$

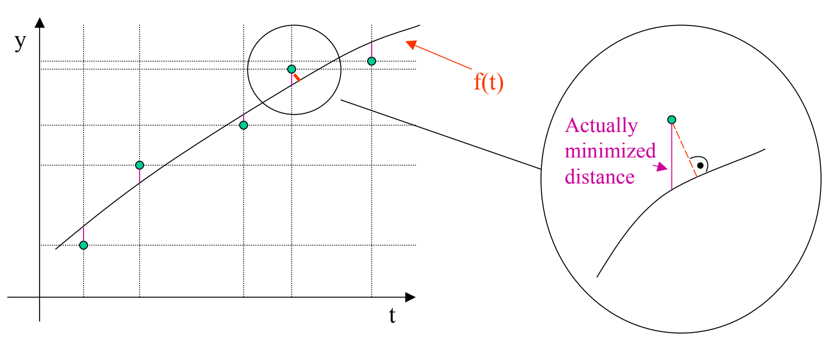
- Further, the Tschebycheff- oder L_∞-approximation:

$$\max\left|f(t_i)-y_i\right| \leq D ! \qquad (3.5)$$

- Commonly, L_2-Approximation is applied
- As illustrated in the following slides sketch, only the y-coordinates differences of actual measurement and model output are optimized

Model Matching

- Fitting process for f(x):



- Several options for choice of f(x)
- Dependent on available measurement data, e.g. obvious/assumed linear relationship or nonlinear relationsship (typical case for sensor data)

Model Matching

- One potential choice for f(t) or f(x) in the case of function approximation is the choice of polynomial function
- Based on the validity of certain restrictions of the signal or transfer characteristics shape the order of the applicable polynomial can be limited
- Assumptions of a monotonically decreasing or increasing functions free of extrema and oscillations allow, e.g., the limitation to polynomials of 3rd order:

$$P(x) = a + bx + cx^2 + dx^3 \tag{3.6}$$

- The polynomial function returns values for all samples with an overall minimized accumulated error for all samples:

$$S = \sum_{i=1}^{N} \underbrace{w(x_i)}_{} \big(\overbrace{P(x_i) - y_i}^{r_i}\big)^2 = \min ! \tag{3.7}$$

Optional weighting function
In the following $w(x_i)=1$

Model Matching

- One option to obtain an optimum solution for the least squares approach is the determination of the four coefficients of the 3rd order polynomial by computation of the partial derivatives of S with regard to these coefficients
- An extremum is given by:

$$\frac{\partial S}{\partial a} = 0; \quad \frac{\partial S}{\partial b} = 0; \quad \frac{\partial S}{\partial c} = 0; \quad \frac{\partial S}{\partial d} = 0 \tag{3.8}$$

- In the regarded case, the extremum will always be a minimum, as the matrix of 2nd order derivatives is always positively definite
- Thus the approach delivers four equations for four variables:

$$\frac{\partial S}{\partial a} = \sum_{i=1}^{N} 2\left(a + bx_i + cx_i^2 + dx_i^3 - y_i\right) \cdot 1 = 0$$

$$\frac{\partial S}{\partial b} = \sum_{i=1}^{N} 2\left(a + bx_i + cx_i^2 + dx_i^3 - y_i\right) \cdot x_i = 0 \tag{3.9}$$

Model Matching

$$\frac{\partial S}{\partial c} = \sum_{i=1}^{N} 2\left(a + bx_i + cx_i^2 + dx_i^3 - y_i\right) \cdot x_i^2 = 0$$

$$\frac{\partial S}{\partial d} = \sum_{i=1}^{N} 2\left(a + bx_i + cx_i^2 + dx_i^3 - y_i\right) \cdot x_i^3 = 0 \qquad (3.10)$$

- This can be reformulated to give the following normal equations:

$$aN + b\sum x_i + c\sum x_i^2 + d\sum x_i^3 = \sum y_i$$
$$a\sum x_i + b\sum x_i^2 + c\sum x_i^3 + d\sum x_i^4 = \sum x_i y_i$$
$$a\sum x_i^2 + b\sum x_i^3 + c\sum x_i^4 + d\sum x_i^5 = \sum x_i^2 y_i$$
$$a\sum x_i^3 + b\sum x_i^4 + c\sum x_i^5 + d\sum x_i^5 = \sum x_i^3 y_i \qquad (3.11)$$

- Solving of this system of equations returns parameters a,b,c,d
- Generally, gradient descent techniques could also be applied (Iteration/step number, stopping criterion, learn parameter required)

Model Matching

- As initially suggested in eq. (3.1), in particular cases, an underlying linear relationship can be assumed, i.e., a 1rst oder polynomial or a straight line can be fitted:

$$P(x) = f(x) = a + bx \qquad (3.12)$$

- Error function to be minimized:

$$S = \sum_{i=1}^{N} \left(P(x_i) - y_i\right)^2 = \sum_{i=1}^{N} \left(a + bx_i - y_i\right)^2 \min ! \qquad (3.13)$$

- Corresponding normal equations:

$$aN + b\sum x_i = \sum y_i \qquad (3.14a)$$
$$a\sum x_i + b\sum x_i^2 = \sum x_i y_i \qquad (3.14b)$$

Model Matching

- Resolving of eq. (3.14a) returns:

$$a = \frac{\sum y_i}{N} - \frac{b \sum x_i}{N} \tag{3.15}$$

- Inserting eq. (5.85) in eq. (5.84b) returns:

$$\frac{\sum x_i \sum y_i}{N} - \frac{b \sum x_i \sum x_i}{N} + b \sum x_i^2 = \sum x_i y_i \tag{3.16}$$

$$\Rightarrow b = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{N}}{\sum x_i^2 - \frac{\sum x_i \sum x_i}{N}} \tag{3.17}$$

- By inserting data values in eq. (3.16) and eq. (3.17) the corresponding model coefficients can be computed

Model Matching

- **Example 3.1:** For the following sample values a regression line shall be computed:

i	1	2	3	4	5
x_i	1	2	3	4	5
y_i	1	2	3	3	4

To compute the coefficients by eq. (3.16) and eq. (3.17), required subexpressions will be computed first from the data:

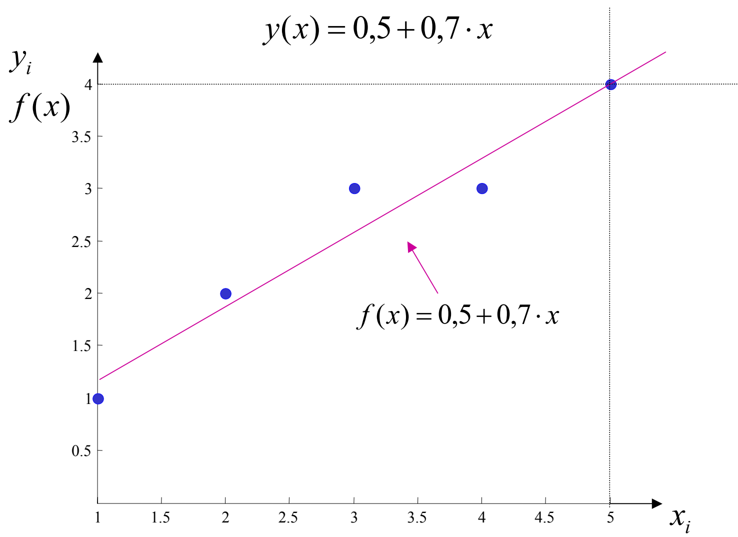
$$\sum_{i=1}^{5} x_i = 15; \quad \sum_{i=1}^{5} x_i^2 = 55; \quad \sum_{i=1}^{5} y_i = 13; \quad \sum_{i=1}^{5} x_i y_i = 46$$

$$\Rightarrow b = \frac{46 - \frac{15 \cdot 13}{5}}{55 - \frac{15 \cdot 15}{5}} = 0{,}7; \quad a = \frac{13}{5} - 0{,}7 \frac{15}{5} = 0{,}5$$

Model Matching

With the determined two coefficients the linear model is given by:

$$y(x) = 0{,}5 + 0{,}7 \cdot x$$



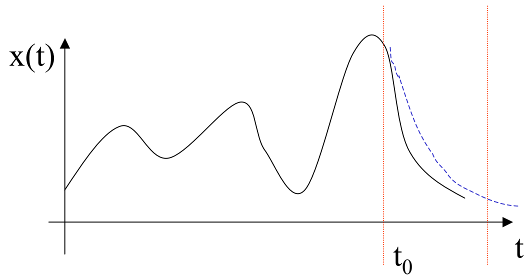
Model Matching

- The simple linear model, or a 2nd or 3rd order polynomial can be sufficient for many practical cases
- In the general case significantly higher orders can be required
- Polynomials can lead to substantial problems, e.g. ripples, in the approximation and problems handling noisy data
- Alternative approaches base on piece-wise approximation using spline functions or kernel functions
- Relationship to the general procedure designing classifiers, where probability densities are likewise modeled or approximated
- Other options of model building and fitting can arise, when a priori knowledge of the underlying relationships exist
- For instance, physical constraints and mechanisms can lead to a functional behavior similar to charging and discharging charateristics
- These might be described by exponential models with characteristic time constants

Model Matching

- Assuming a signal with an envelope, that shows an exponentially decaying amplitude, an exponential model could be fitted to the data

$$y = f(t) = x(t_0) \cdot e^{-\frac{t-t_0}{\tau}} \tag{3.18}$$



- An error function could be defined as:

$$S = \frac{1}{2}\sum_{i=0}^{N}\left(x(t_0)\cdot e^{-\frac{t_i-t_0}{\tau}} - x(t_i)\right)^2 = \min! \tag{3.19}$$

Model Matching

- The partial derivative required for gradient descent is given by

$$\frac{\partial S}{\partial \tau} = \sum_{i=0}^{N}\left(\left[x(t_0)\cdot e^{-\frac{t_i-t_0}{\tau}} - x(t_i)\right]\cdot x(t_0)\cdot e^{-\frac{t_i-t_0}{\tau}}\cdot\left(\frac{t_i-t_0}{\tau^2}\right)\right) \tag{3.20}$$

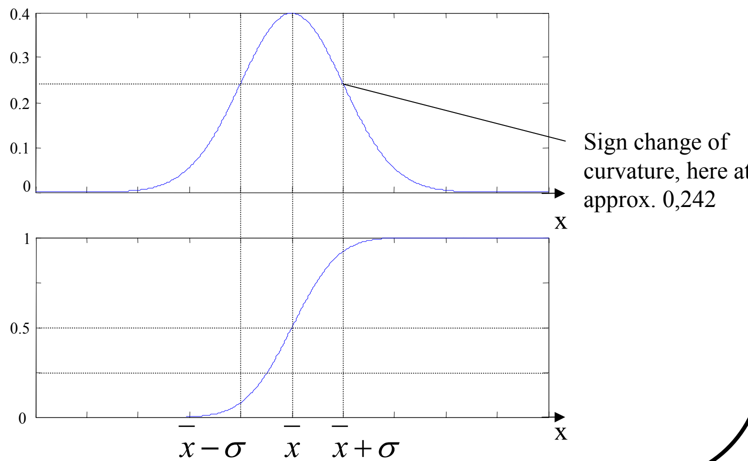
- The model parameter is adapted according the following adaptation rule

$$\tau^{new} = \tau^{old} - \eta\cdot\frac{\partial S}{\partial \tau} \tag{3.21}$$

- This approach returns after convergence a value for the parameter τ, that corresponds with best fit of the exponential (local) model to the data
- Appropriate initialization of τ required
- Error threshold or fixed number of steps required for adaptation control
- Extensions of the model (residual value, location, etc.) feasible

Statistical Moments

- Commonly, for a signal or image region, statistical properties or distribution assumptions can be assumed
- A common assumption is the Gaussian distribution $(\bar{x}; \sigma)$ – *normal distribution*



Statistical Moments

- The density is described by the following relationship:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-\bar{x})^2}{\sigma^2}} \qquad (3.22)$$

- Features can be computed based on this model assumption, determinating the model parameters mean and variance (standard deviation):

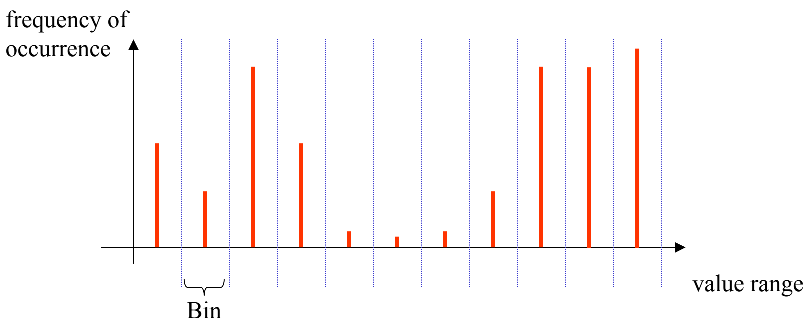
$$\hat{\bar{x}} = \frac{1}{N}\sum_{n=1}^{N} x_n \qquad (3.23)$$

$$s^2 = \frac{1}{N-1}\sum_{n=1}^{N}\left(x_n - \hat{\bar{x}}\right)^2 \qquad (3.24)$$

- These parameters serve as simple features !
- Higher order statistical moments can be computed

Histograms

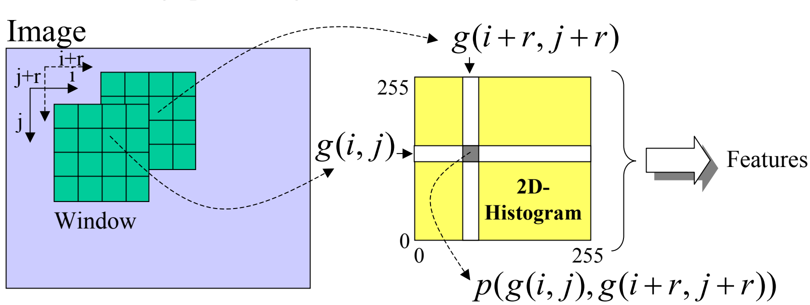
- In the case of unknown distribution, an estimation can be obtained by histogram techniques



- Bin size determines quantization
- Not necessarily equidistant binning or thresholding !
- Frequencies commonly are normalized
- Cumulative histogram can be computed
- Special case: Direction/Orientation histograms

Second order statistics

- Histograms assume first oder statistics, more information can be extracted by higher order statistics
- Common approach: second order statistics, denoted as co-occurrence-matrices in image processing



- Rotation invariance not inherent to approach, common remedy is computation of the mean of four matrices for 0°, 45°, 90°, and 135°
- Compact features by Haralicks moment invariants ($M_{1,2,4,5,9,12,13}$)
- Quantization (bin discretization) applicable !

Geometrical Moment Invariants

- Binary or gray-value image structures can be described by geometrical moment of increasing order (s. e.g., [7])
- The moment m_{pq} of the order (p+q) is given as

$$m_{pq} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^p y^q f(x,y)dxdy \tag{3.25}$$

- In the discrete case of signal and image processing eq. (3.25) changes to

$$m_{pq} = \sum_{-\infty}^{\infty}\sum_{-\infty}^{\infty} x^p y^q f(x,y) \tag{3.26}$$

- From these moments m_{pq} invariants with regard to translation, rotation, and scale can be computed
- Subtraction of the center of gravity leads to centralized moments and translation invariants

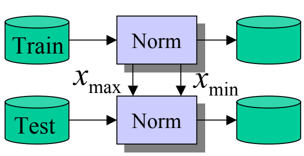
$$m_{pq} = \sum_{-\infty}^{\infty}\sum_{-\infty}^{\infty} \left(x-\bar{x}\right)^p \left(y-\bar{y}\right)^q f(x,y) \tag{3.27}$$

Geometrical Moment Invariants

- A couple of the m_{pq} allow immediate interpretation, e.g., m_{00} corresponds to the blob area
- The center of gravity of blobs can be expressed as

$$\bar{x} = \frac{m_{10}}{m_{00}} \quad ; \quad \bar{y} = \frac{m_{01}}{m_{00}} \tag{3.28}$$

- Moment invariants show excessive dynamic range, which requires normalization for balanced evaluation by distance metric



$$x_j^{norm} = \frac{x_j - x_{min}}{x_{max} - x_{min}} \tag{3.29}$$

Summary

- The chapter gave a survey of the basic idea and relevant selection of feature computation methods
- A large number of alternative methods are available, in particular domain specific techniques, applicable for, e.g., image processing
- Each presented method provides one or several features, which can be salient for ensuing classification
- Several methods can be combined to achieve a sufficient number of features for the regarded application and data samples !
- Not necessarily are all computed features required, i.e., systematic redundancy and irrelevancy reduction is required
- Optimal choice, parameterization, and weeding out of irrelevant/redundant features are key steps for efficient feature computation in system design
- In the literature, this step of operation is also denoted as feature extraction