

# 64-544 Grundlagen der Signalverarbeitung und Robotik

[http://tams.informatik.uni-hamburg.de/  
lectures/2012ss/vorlesung/GdSR](http://tams.informatik.uni-hamburg.de/lectures/2012ss/vorlesung/GdSR)

Jianwei Zhang

**T | A** Universität Hamburg  
**M | S** Fakultät für Mathematik, Informatik und Naturwissenschaften  
Fachbereich Informatik  
Technische Aspekte Multimodaler Systeme

Sommersemester 2012

## Agenda

### 5. Rekursive Zustandsschätzung

Grundlagen der Wahrscheinlichkeitsrechnung  
Stochastische Fortpflanzungsgesetze und Belief  
Bayes-Filter  
Selbst-Lokalisierung mobiler Roboter  
Literatur

## Gliederung

1. Einführung
2. Grundlagen der Robotik
3. Elementares der Sensorik
4. Verarbeitung von Scandaten
5. Rekursive Zustandsschätzung
6. Fuzzy-Logik
7. Steuerungsarchitekturen

## Rekursive Zustandsschätzung

- ▶ **Idee:** Schätzen eines Systemzustands, der nicht direkt gemessen werden kann, aber aus den Messungen ableitbar ist
  - ▶ Beobachtungen bestimmter Aspekte des Systems
  - ▶ Steueraktionen, die das System beeinflussen
  - ▶ Unsicherheit in Beobachtungen und Aktionen.
- ▶ Probabilistische Algorithmen zur Zustandsschätzungen berechnen eine **belief distribution** über die möglichen Zustände
- ▶ Der **belief** beschreibt das Wissen eines Systems über seinen Zustand

## Grundlagen der Wahrscheinlichkeitsrechnung

- ▶ Sensorwerte, Stellgrößen, der Zustand eines Systems und von dessen Umgebung können als **Zufallsvariable** modelliert werden
- ▶  $X$  sei eine Zufallsvariable und  $x$  ein Wert, den diese annehmen kann
- ▶ Ist der Wertebereich  $W$  (Ereignisbereich) von  $X$  diskret mit  $W = \{x_1, x_2, \dots, x_m\}$ , so schreibt man:

$$p(X = x) \quad \text{oder kurz} \quad p(x)$$

um die Wahrscheinlichkeit anzugeben, dass  $X$  den Wert  $x$  hat.

- ▶ es gilt:
  1.  $p(x_i) \geq 0$
  2.  $p(W) = 1$
  3.  $p(x_1 \cup x_2 \cup \dots \cup x_m) = p(x_1) + p(x_2) + \dots + p(x_m)$   
wenn die Paare  $x_i, x_j$  mit  $i \neq j$  disjunkt sind

## Grundlagen der Wahrscheinlichkeitsrechnung (cont.)

- ▶  $p(\emptyset) = 0$
- ▶  $p(x) = 1 - p(\bar{x})$ , mit  $\bar{x}$  Komplementärereignis zu  $x$
- ▶  $p(a - b) = p(a) - p(a \cap b)$
- ▶  $b \subseteq a \Rightarrow p(b) \leq p(a)$
- ▶  $p(a \cup b) = p(a) + p(b) - p(a \cap b)$

## Grundlagen der Wahrscheinlichkeitsrechnung (cont.)

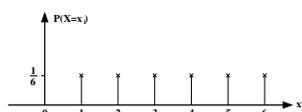
### Verteilungsfunktion

- ▶ Eine Funktion  $F$ , die durch die diskrete Zufallsvariable  $X$  bestimmt ist, heißt **Verteilungsfunktion** von  $X$ .
- ▶ Berechnung der Funktionswerte  $F(x)$ :  
Wert der Zufallsvariablen  $\leq$  als vorgegebener Wert:

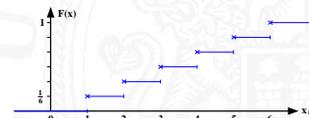
$$F(x) = P(X \leq x), \quad x \in \mathbb{R}$$

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} P(X = x_i)$$

- ▶ Beispiel Würfel:



Wahrscheinlichkeiten



Verteilungsfunktion

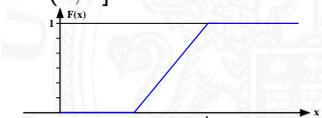
## Grundlagen der Wahrscheinlichkeitsrechnung (cont.)

### Dichtefunktion

- ▶ Eine Funktion  $f : \mathbb{R} \rightarrow \mathbb{R}$ , die über eine reellwertige Zufallsvariable  $X$  definiert ist, heißt **Wahrscheinlichkeitsdichte** oder **Dichtefunktion**.
- ▶ Die Funktion  $F(x) = p(X \leq x)$  mit  $(x \in \mathbb{R})$  heißt **Verteilungsfunktion** der stetigen Zufallsvariablen  $X$ . Mit ihr läßt sich berechnen mit welcher Wahrscheinlichkeit  $X$  einen Wert im Intervall  $(a, b]$  annimmt:  $p(a < X \leq b) = F(b) - F(a)$
- ▶ Beispiel Gleichverteilung im Intervall  $(a, b]$ :



Dichtefunktion



Verteilungsfunktion

## Grundlagen der Wahrscheinlichkeitsrechnung (cont.)

### Dichtefunktion

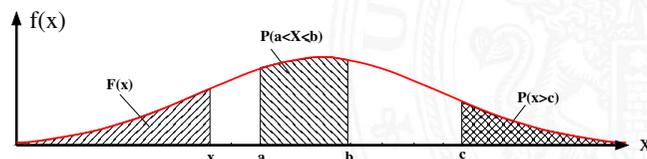
- Berechnung der stetigen Verteilung durch ihre Dichte  $f(x)$ :

$$F(x) = \int_{-\infty}^x f(x) dx$$

Wert der stetigen Zufallsvariablen im Intervall  $(a, b]$ :

$$P(a < X \leq b) = \int_a^b f(x) dx$$

- Beispiel: Dichte einer stetigen Zufallsvariablen  $X$



## Grundlagen der Wahrscheinlichkeitsrechnung (cont.)

### Erwartungswert

- Der **Erwartungswert** einer Zufallsvariablen ist gegeben als:

$$E[X] = \mu = \sum_x x p(x) \quad (\text{diskret})$$

$$E[X] = \mu = \int x p(x) dx \quad (\text{kontinuierlich})$$

- Der Erwartungswert ist eine lineare Funktion einer Zufallsvariablen, d.h. es gilt:

$$E[aX + b] = aE[X] + b$$

## Grundlagen der Wahrscheinlichkeitsrechnung (cont.)

### Varianz

- Die **Varianz**  $\sigma^2$  von  $X$  berechnet sich wie folgt:

$$\sigma^2 = E[X - \mu]^2 = E[X^2] - \mu^2$$

- Die Varianz misst das erwartete Quadrat der Abweichung vom Erwartungswert
- Für die Standardabweichung oder Streuung  $\sigma$  von  $X$  gilt:

$$\sigma = \sqrt{\sigma^2} = \sqrt{E[X - \mu]^2}$$

## Grundlagen der Wahrscheinlichkeitsrechnung (cont.)

### Stochastische Unabhängigkeit

- Für die Wahrscheinlichkeit, dass  $X$  den Wert  $x$  und  $Y$  den Wert  $y$  hat, schreibt man

$$p(x, y) = p(X = x \text{ und } Y = y)$$

- Sind die beiden Zufallsvariablen  $X$  und  $Y$  von einander unabhängig gilt:

$$p(x, y) = p(x)p(y)$$

## Grundlagen der Wahrscheinlichkeitsrechnung (cont.)

### Kovarianz

- Für die **Kovarianz** (Cov) zweier reeller Zufallsvariablen  $X, Y$  gilt:

$$\text{Cov}(X, Y) = E((X - E(X)) \cdot (Y - E(Y)))$$

unter der Voraussetzung, dass  $E(X), E(Y)$  und  $E(XY)$  existiert.

- $\text{Cov}(X, Y) > 0$ , monoton wachsender linearer Zusammenhang
- $\text{Cov}(X, Y) = 0$ , falls kein linearer Zusammenhang zwischen den Variablen; möglicherweise aber ein nichtlinearer Zusammenhang (Variablen sind nicht notwendigerweise stochastisch unabhängig)
- $\text{Cov}(X, Y) < 0$ , monoton fallender linearer Zusammenhang
- keine Aussage über die Stärke des Zusammenhanges
- $\text{Cov}(X, Y) = \text{Cov}(Y, X)$
- $\text{Cov}(X, X) = \text{Var}(X) = \sigma^2$
- falls  $X, Y$  stochastisch unabhängig, folgt  $\text{Cov}(X, Y) = 0$

## Grundlagen der Wahrscheinlichkeitsrechnung (cont.)

### Entropie

- **Entropie** einer Wahrscheinlichkeitsverteilung

Sei  $P = \{p_1, \dots, p_n\}$  eine Wahrscheinlichkeitsverteilung, dann gilt für die Entropie von  $P$ :

$$H(P) = - \sum_{i=1}^n p_i \log_2 p_i$$

- Entropie ist ein Maß für die Unordnung (in Anlehnung an den Unordnungsbegriff Entropie in der Thermodynamik); je größer die Wahrscheinlichkeit für das Auftreten eines Zeichens, umso kleiner die Entropie, die Unordnung im System
- entspricht dem Erwartungswert des Informationsgehaltes ( $E[I(x)]$  mit  $I(x) = \log_2 \frac{1}{p(x)} = -\log_2 p(x)$  [bit])

## Grundlagen der Wahrscheinlichkeitsrechnung (cont.)

### Exkurs: Informationsgehalt nach Shannon

- **Informationsgehalt**: mathematische Beschreibung des statistischen Informationsgehalts  $I(x)$  eines Zeichens  $x$
- Forderungen an Berechnungsfunktion:
  - je seltener ein bestimmtes Zeichen  $x$  auftritt, desto größer soll dessen Informationsgehalt sein
  - Gesamtinformation einer Zeichenkette,  $x_1 x_2 \dots x_n$  ergibt sich aus der Summe der Einzelinformationen, also:

$$I(x_1 x_2 \dots x_n) = I(x_1) + I(x_2) + \dots + I(x_n)$$

- falls Zeichen sicher auftritt ( $p(x) = 1$ ) soll gelten  $I(x) = 0$
- Logarithmusfunktion erfüllt gestellte Forderungen

## Grundlagen der Wahrscheinlichkeitsrechnung (cont.)

### Exkurs: Informationsgehalt nach Shannon (cont.)

- Berechnung des Informationsgehaltes:

$$I(x) = \log_b \frac{1}{p(x)} \quad [\text{shannon}(sh)]$$

- Basis  $b$  quasi Maßstab oder auch Anzahl der möglichen Zustände einer Signalquelle
- bei zwei Zuständen  $\{0,1\} \rightarrow b = 2$  und Einheit [bit]
- sendet Quelle immer identisches Zeichen ( $p(x) = 1$ ) ist  $I(x) = 0$
- ist im binären Fall  $p(0) = p(1) = 0.5$  ist  $I(0) = I(1) = 1$ ; dann entspricht Informationsgehalt Länge der Zeichenkette

## Grundlagen der Wahrscheinlichkeitsrechnung (cont.)

### Exkurs: Informationsgehalt nach Shannon (cont.)

- ▶ gibt es im Fall  $b = 2$   $2^k$  Symbole gleicher Wahrscheinlichkeit, so ist  $I(s) = k$  für jedes Symbol  $s$
- ▶ im diskreten Fall ist  $-\log_2 p(x)$  die Anzahl der Bits, die notwendig sind, um  $x$  optimal zu kodieren, wenn  $p(x)$  die Wahrscheinlichkeit für die Beobachtung von  $x$  ist

Beispiel: der Buchstabe  $s$  erscheint in deutschen Texten mit einer Wahrscheinlichkeit  $p(s) = 0,073$

$$I(s) = -\log_2(0.073) = 3.8\text{bit}$$

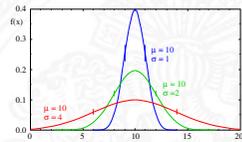
Hinweis:  $\log_b(x) = \frac{\log_a(x)}{\log_a(b)}$  hier:  $\log_2(x) = \frac{\log_{10}(x)}{\log_{10}(2)}$

## Grundlagen der Wahrscheinlichkeitsrechnung (cont.)

### Dichtefunktionen

- ▶ nimmt die Zufallsvariable kontinuierliche Werte an, wird sie durch eine **Dichtefunktion** beschrieben (engl. *probability density function, PDF*)
- ▶ eine typische Dichtefunktion ist die **Normalverteilung** mit Mittelwert  $\mu$  und Varianz  $\sigma^2$ :

$$p(x) = (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}\right\}$$



- ▶ für eine PDF gilt:  $\int p(x)dx = 1$

- ▶ Funktionswert einer PDF ist nicht wie bei diskreten Wahrscheinlichkeiten nach oben durch 1 begrenzt

## Grundlagen der Wahrscheinlichkeitsrechnung (cont.)

### Entropie (cont.)

- ▶ Für die Entropie  $P = \{p_1, \dots, p_n\}$  gelten folgende Schranken:

$$0 \leq H(p_1, \dots, p_n) \leq \log_2 n$$

- ▶ untere Schranke von  $H(p_1, \dots, p_n) = 0$  genau dann, wenn die  $W$ -Verteilung entartet ist; also wenn ein  $i$  existiert mit  $p_i = 1$
- ▶ obere Schranke von  $H(p_1, \dots, p_n) = \log_2 n$  genau dann, wenn  $p_i = \frac{1}{n}, \forall i, 1 \leq i \leq n$  (Gleichverteilung)

- ▶ Beispiel: Alphabet mit vier Zeichen  $\Sigma = \{r, s, t, u\}$

$$H(\Sigma) = 1/2 \cdot 1 + 1/4 \cdot 2 + 1/8 \cdot 3 + 1/8 \cdot 3 = 1.75$$

Folgerung: Um dieses Alphabet binär minimal zu kodieren, bräuchte es 2 bit.

s	p(s)	I(s)
r	1/2	1
s	1/4	2
t	1/8	3
u	1/8	3

## Grundlagen der Wahrscheinlichkeitsrechnung (cont.)

### Dichtefunktionen, mehrdimensional

- ▶ Im mehrdimensionalen Fall mit Mittelwertvektor  $\mu$  und Kovarianzmatrix  $\Sigma$ :

$$p(x) = \det(2\pi\Sigma)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(x-\mu)\Sigma^{-1}(x-\mu)^T\right\}$$

- ▶ Die Kovarianzmatrix  $\Sigma$  ist die Matrix der paarweisen Kovarianzen eines Zufallsvektors  $X = (X_1, \dots, X_n)$

$$\Sigma = \begin{bmatrix} \text{Cov}(X_1, X_1) & \dots & \text{Cov}(X_1, X_n) \\ \vdots & \ddots & \vdots \\ \text{Cov}(X_n, X_1) & \dots & \text{Cov}(X_n, X_n) \end{bmatrix}$$

- ▶ Die **Kovarianzmatrix**  $\Sigma$  ist *positiv semidefinit* und *symmetrisch* ( $xAx^T \geq 0$ , diagonalisierbar, Eigenwerte  $\geq 0$ )

## Grundlagen der Wahrscheinlichkeitsrechnung (cont.)

### bedingte Wahrscheinlichkeiten

- Wahrscheinlichkeit für  $p(X = x)$  unter der Voraussetzung  $p(Y = y)$  (bedingte Wahrscheinlichkeit):

$$p(x|y) = p(X = x | Y = y)$$

- wenn  $p(y) > 0$ , so gilt:  $p(x|y) = \frac{p(x, y)}{p(y)}$

- sind  $X$  und  $Y$  voneinander unabhängig, gilt:

$$p(x|y) = \frac{p(x)p(y)}{p(y)} = p(x)$$

→ wenn  $X$  und  $Y$  unabhängig voneinander sind, sagt  $Y$  nichts über  $X$

## Grundlagen der Wahrscheinlichkeitsrechnung (cont.)

### Satz von Bayes

- Satz von Bayes: (mit  $p(y) > 0$ )

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)} = \frac{p(y|x)p(x)}{\sum_{x'} p(y|x')p(x')} \quad (\text{diskret})$$

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)} = \frac{p(y|x)p(x)}{\int p(y|x')p(x')dx} \quad (\text{kontinuierlich})$$

Satz von Bayes beschreibt Umkehrung von Schlussfolgerungen

- die Berechnung von  $p(\text{Ereignis} | \text{Ursache})$  ist oft einfach
- aber gesucht ist  $p(\text{Ursache} | \text{Ereignis})$

## Grundlagen der Wahrscheinlichkeitsrechnung (cont.)

### totale Wahrscheinlichkeit

- Gesetz der totalen Wahrscheinlichkeit:

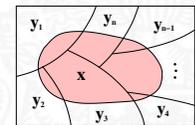
$$p(x) = \sum_y p(x|y)p(y) \quad (\text{diskret})$$

$$p(x) = \int p(x|y)p(y)dy \quad (\text{kontinuierlich})$$

Seien  $y_1, y_2, \dots, y_n$  wechselweise disjunkte Ereignisse, die den Ereignisraum  $S$  bilden, also  $S = (y_1 \cup y_2 \cup y_3 \cup \dots \cup y_n)$ ; ferner liege das Ereignis  $x$  ebenfalls in  $S$

$$x = S \cap x = (y_1 \cup y_2 \cup y_3 \cup \dots \cup y_n) \cap x \\ = (y_1 \cap x) \cup (y_2 \cap x) \cup \dots \cup (y_n \cap x)$$

$$p(x) = p(y_1, x) + p(y_2, x) + \dots + p(y_n, x) \\ = p(x|y_1)p(y_1) + p(x|y_2)p(y_2) + \dots + p(x|y_n)p(y_n) \\ = \sum_{i=1}^n p(x|y_i)p(y_i)$$



## Grundlagen der Wahrscheinlichkeitsrechnung (cont.)

### Satz von Bayes

Beweisidee:

Nach Definition der bedingten Wahrscheinlichkeit gilt:

$$p(x|y) = \frac{p(x, y)}{p(y)} = \frac{p(y, x)}{p(y)}$$

nach Ausnutzen des Multiplikationssatzes der bedingten Wahrscheinlichkeiten auf den Zähler und Einsetzen der totalen Wahrscheinlichkeit in den Nenner folgt:

$$p(x_k|y) = \frac{p(y|x_k)p(x_k)}{\sum_{i=1}^n p(y|x_i)p(x_i)}$$

## Grundlagen der Wahrscheinlichkeitsrechnung (cont.)

### Satz von Bayes: Beispiel

Beispiel:

In einer Zahnradfabrik werden Zahnräder identischer Abmessungen auf zwei verschiedenen Maschinen  $m_1$  und  $m_2$  gefräst. Die Maschine  $m_1$  stellt 40% der Zahnräder her, die Maschine  $m_2$  60%. Dabei sind 4% der Zahnräder der Maschine  $m_1$  fehlerhaft und 8% der Maschine  $m_2$ .

Aus der Gesamtproduktion der Zahnräder wird eines zufällig entnommen. Dieses ist fehlerhaft. Wie hoch ist die Wahrscheinlichkeit, dass es von Maschine  $m_1$  gefräst wurde?

$$p(m_1) = 0.4; \quad p(m_2) = 0.6; \quad p(f|m_1) = 0.04; \quad p(f|m_2) = 0.08$$

$$p(m_1|f) = \frac{p(f|m_1)p(m_1)}{\sum_{i=1}^2 p(f|m_i)p(m_i)}$$

$$= \frac{0.04 \cdot 0.4}{(0.04 \cdot 0.4) + (0.08 \cdot 0.6)} = \frac{0.016}{0.064} = \frac{1}{4}$$

## Grundlagen der Wahrscheinlichkeitsrechnung (cont.)

### Satz von Bayes

- ▶ Der Satz von Bayes ist vielfältig anwendbar.
- ▶ Sei  $x$  die Größe, die wir aus  $y$  ableiten wollen, dann bezeichnet man die Grundwahrscheinlichkeit  $p(x)$  als **A-Priori-Wahrscheinlichkeit** und  $y$  als **Daten** (z.B. Sensormessungen).
- ▶ Die Verteilung der  $p(x)$  beschreibt das Wissen über  $X$ , bevor die Messung  $y$  berücksichtigt wird.
- ▶ Die Verteilung der  $p(x|y)$  wird als **A-Posteriori-Wahrscheinlichkeit** von  $X$  bezeichnet.

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}$$

## Grundlagen der Wahrscheinlichkeitsrechnung (cont.)

### Satz von Bayes

- ▶ Der Satz von Bayes erlaubt die bedingte Wahrscheinlichkeit  $p(x|y)$  zu berechnen, wenn die Wahrscheinlichkeit für eine Messung  $y$  bei einem Zustand  $x$  ( $p(y|x)$ ) und die Grundwahrscheinlichkeiten  $p(x)$  und  $p(y)$  bekannt sind.

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}$$

## Grundlagen der Wahrscheinlichkeitsrechnung (cont.)

### Satz von Bayes

- ▶ Im Satz von Bayes hängt die Grundwahrscheinlichkeit  $p(y)$  nicht vom betrachteten  $x$  ab.
- ▶ Daher ist der Faktor  $p(y)^{-1}$  für jeden Wert  $x$  in  $p(x|y)$  gleich.
- ▶ Dieser Faktor wird meist als Normalisierungsfaktor im Satz von Bayes notiert:

$$p(x|y) = \eta p(y|x)p(x)$$

- ▶ Der Faktor  $\eta$  wird derart gewählt, dass die Normierung der Verteilung auf 1 erhalten bleibt.
- ▶  $\eta$  bezeichnet im Folgenden immer solche Normalisierungsfaktoren, deren eigentliche Werte sich aber unterscheiden können.

## Grundlagen der Wahrscheinlichkeitsrechnung

- ▶ Alle bisherigen Regeln können durch eine weitere Zufallsvariable, z. B.  $Z$ , bedingt sein.
- ▶ Dies bedeutet für den Satz von Bayes mit  $Z = z$ :

$$p(x|y, z) = \frac{p(y|x, z)p(x|z)}{p(y|z)}$$

solange  $p(y|z) > 0$

- ▶ Der Multiplikationssatz unabhängiger Zufallsvariablen ( $p(a, b) = p(a)p(b)$ ) gilt analog und definiert die bedingte Unabhängigkeit:

$$p(x, y|z) = p(x|z)p(y|z)$$

## Grundlagen der Wahrscheinlichkeitsrechnung bedingte Unabhängigkeit

- ▶  $p(x, y|z) = p(x|z)p(y|z)$  besagt, dass  $y$  keine Information über  $x$  trägt, wenn  $z$  bekannt ist

- ▶ Formel beschreibt eine **bedingte Unabhängigkeit** und ist äquivalent zu

$$\begin{aligned} p(x|z) &= p(x|z, y) \\ p(y|z) &= p(y|z, x) \end{aligned}$$

- ▶ impliziert **nicht**, dass  $X$  unabhängig von  $Y$  ist:

$$p(x, y|z) = p(x|z)p(y|z) \not\Rightarrow p(x, y) = p(x)p(y)$$

die Umkehrung gilt generell auch nicht:

$$p(x, y) = p(x)p(y) \not\Rightarrow p(x, y|z) = p(x|z)p(y|z)$$

## Stochastische Fortpflanzungsgesetze

- ▶ Der Zustand eines Systems lässt sich durch eine Wahrscheinlichkeitsverteilung

$$p(x_t | x_{0:t-1}, z_{1:t-1}, u_{1:t})$$

beschreiben, die abhängig ist von

- ▶ den bisherigen Zuständen  $x_{0:t-1}$
- ▶ allen bisherigen Messungen  $z_{1:t-1}$  und
- ▶ allen bisherigen Stellgrößen (Steuerkommandos)  $u_{1:t}$

## Stochastische Fortpflanzungsgesetze (cont.)

- ▶ Ist der Zustand  $x_t$  **komplett**, dann fasst die Stellgröße  $u_t$  und der Zustand  $x_{t-1}$  alle bisherigen Ereignisse zusammen:

$$p(x_t | x_{0:t-1}, z_{1:t-1}, u_{1:t}) = p(x_t | x_{t-1}, u_t)$$

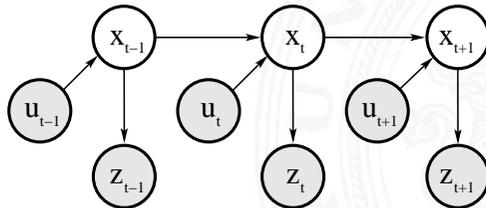
- ▶ Für die Messungen gilt dann entsprechend:

$$p(z_t | x_{0:t}, z_{1:t-1}, u_{1:t}) = p(z_t | x_t)$$

- ▶ Der Zustand  $x_t$  ist ausreichend um die Messung  $z_t$  vorherzusagen
- ▶ Die bedingte Wahrscheinlichkeit  $p(x_t | x_{t-1}, u_t)$  wird als **Zustandsübergangswahrscheinlichkeit** bezeichnet
  - ▶ Sie beschreibt wie sich der Zustand der Umgebung abhängig von den Stellgrößen ändert
- ▶ Die Wahrscheinlichkeit  $p(z_t | x_t)$  wird **Wahrscheinlichkeit der Messung** genannt

## Stochastische Fortpflanzungsgesetze (cont.)

- ▶ Zustandsübergangswahrscheinlichkeit und Wahrscheinlichkeit der Messung beschreiben zusammen ein **dynamisches stochastisches System**
- ▶ Eine solche Systembeschreibung ist auch bekannt als **hidden Markov model (HMM)** oder **dynamic Bayes network (DBN)**



## Zustandsschätzung: Belief

- ▶ Als **belief** bezeichnet man das Wissen (Schätzung) eines Systems über seinen Zustand.
- ▶ Der *wahre Zustand* eines Systems ist nicht dasselbe!
- ▶ Der **belief** ist die A-Posteriori-Wahrscheinlichkeit der Zustandsvariablen anhand der bisherigen Daten:

$$bel(x_t) = p(x_t | z_{1:t}, u_{1:t})$$

- ▶ Diese Definition berechnet den **belief** (als Wahrscheinlichkeit) nach der aktuellen Messung  $z_t$ .

## Zustandsschätzung: Belief (cont.)

- ▶ Der **belief** vor der Messung wird als

$$\overline{bel}(x_t) = p(x_t | z_{1:t-1}, u_{1:t})$$

definiert und meist als **Vorhersage** (engl. *prediction*) bezeichnet

- ▶ Der Schritt zur Berechnung von  $bel(x_t)$  aus der Vorhersage  $\overline{bel}(x_t)$  wird als **Korrektur** (engl. *correction* oder *measurement update*) bezeichnet

## Bayes-Filter

- ▶ Ein allgemeiner Algorithmus, um **beliefs** zu berechnen, ist der **Bayes-Filter-Algorithmus**
- ▶ Der Algorithmus ist rekursiv und berechnet  $bel(x_t)$  zum Zeitpunkt  $t$  aus folgenden Größen
  - ▶  $bel(x_{t-1})$  zum Zeitpunkt  $t - 1$  (alter belief)
  - ▶ der Stellgröße  $u_t$  (aktuelle Stellgröße)
  - ▶ der Messung  $z_t$  (aktuelle Messung)



## Bayes-Filter-Algorithmus

Bayes-Filter ( $bel(x_{t-1}), u_t, z_t$ ):

1. **for all**  $x_t$  **do**
2.      $\overline{bel}(x_t) = \int p(x_t | u_t, x_{t-1}) bel(x_{t-1}) dx_{t-1}$
3.      $bel(x_t) = \eta p(z_t | x_t) \overline{bel}(x_t)$
4. **endfor**
5. **return**  $bel(x_t)$



## Bayes-Filter-Algorithmus (cont.)

- ▶ Der Algorithmus muss mit einem *belief* zum Zeitpunkt  $t = 0$  initialisiert werden.
- ▶ In den meisten Fällen ist der Startzustand  $x_0$  bekannt und die Wahrscheinlichkeit kann für alle anderen Zustände gleich 0 gesetzt werden,
- ▶ oder der Zustand ist unbekannt und alle Zustände werden mit der gleichen Wahrscheinlichkeit initialisiert.



## Bayes-Filter-Algorithmus (cont.)

- ▶ In Zeile 2 wird die Stellgröße  $u_t$  verarbeitet
  - ▶  $\overline{bel}(x_t)$  ist das Integral des Produktes zweier Wahrscheinlichkeitsverteilungen:
    - ▶ die Wahrscheinlichkeit für Zustand  $x_{t-1}$  und
    - ▶ die Wahrscheinlichkeit das mit  $u_t$  in den Zustand  $x_t$  gewechselt wird
  - ▶ Das ist die **Vorhersage**
- ▶ In Zeile 3 findet die **Korrektur** statt
  - ▶ Vorhersage ( $\overline{bel}(x_t)$ ) wird mit der Wahrscheinlichkeit multipliziert, dass die Messung  $z_t$  in diesem Zustand wahrgenommen wird (Satz von Bayes)

$$2 \quad \overline{bel}(x_t) = \int p(x_t | u_t, x_{t-1}) bel(x_{t-1}) dx_{t-1}$$

$$3 \quad bel(x_t) = \eta p(z_t | x_t) \overline{bel}(x_t)$$



## Bayes-Filter-Algorithmus (cont.)

- ▶ Der Algorithmus kann in dieser Form nur für sehr einfache Probleme implementiert werden.
- ▶ Entweder muss die Integration in Zeile 2 und die Multiplikation in Zeile 3 in geschlossener Form möglich sein,
- ▶ oder ein endlicher Zustandsraum muss gegeben sein, so dass das Integral in Zeile 2 zu einer Summe wird.

$$2 \quad \overline{bel}(x_t) = \int p(x_t | u_t, x_{t-1}) bel(x_{t-1}) dx_{t-1}$$

$$3 \quad bel(x_t) = \eta p(z_t | x_t) \overline{bel}(x_t)$$

## Mathematische Herleitung des Bayes-Filter

- ▶ **Beweis durch Induktion:**  
Zeige, dass  $\overline{bel}(x_t) = p(x_t|z_{1:t-1}, u_{1:t})$  aus  $p(x_{t-1}|z_{1:t-1}, u_{1:t-1})$  hergeleitet werden kann
- ▶ **Beweis-Basis:** Korrekte Initialisierung von  $bel(x_0)$  zum Zeitpunkt  $t = 0$
- ▶ **Bedingungen:** Zustand  $x_t$  ist komplett und die Stellgrößen werden zufällig gewählt
- ▶ Nach dem Satz von Bayes gilt:

$$\begin{aligned} bel(x_t) = p(x_t|z_{1:t}, u_{1:t}) &= \frac{p(z_t|x_t, z_{1:t-1}, u_{1:t})p(x_t|z_{1:t-1}, u_{1:t})}{p(z_t|z_{1:t-1}, u_{1:t})} \\ &= \eta p(z_t|x_t, z_{1:t-1}, u_{1:t})p(x_t|z_{1:t-1}, u_{1:t}) \end{aligned}$$

## Mathematische Herleitung des Bayes-Filter (cont.)

- $bel(x_t) = p(x_t|z_{1:t}, u_{1:t}) = \eta p(z_t|x_t, z_{1:t-1}, u_{1:t})p(x_t|z_{1:t-1}, u_{1:t})$
- ▶ da der Zustand komplett ist, gilt (bedingte Unabhängigkeit)

$$p(z_t|x_t, z_{1:t-1}, u_{1:t}) = p(z_t|x_t)$$

- ▶ Daraus folgt:

$$bel(x_t) = p(x_t|z_{1:t}, u_{1:t}) = \eta p(z_t|x_t)p(x_t|z_{1:t-1}, u_{1:t})$$

- ▶ und damit

$$bel(x_t) = \eta p(z_t|x_t)\overline{bel}(x_t),$$

was Zeile 3 des Bayes-Filter-Algorithmus entspricht

## Mathematische Herleitung des Bayes-Filter (cont.)

- ▶ Nach dem Gesetz der totalen Wahrscheinlichkeit ( $p(a) = \int p(a|c)p(c)dc$ ) gilt:

$$\begin{aligned} \overline{bel}(x_t) &= p(x_t|z_{1:t-1}, u_{1:t}) \\ &= \int p(x_t|x_{t-1}, z_{1:t-1}, u_{1:t}) p(x_{t-1}|z_{1:t-1}, u_{1:t}) dx_{t-1} \end{aligned}$$

- ▶ wenn wir  $x_{t-1}$  kennen, liefern  $z_{1:t-1}$  und  $u_{1:t-1}$  keine zusätzliche Information (Annahme: Zustand komplett)
- ▶ daher gilt die Vereinfachung

$$p(x_t|x_{t-1}, z_{1:t-1}, u_{1:t}) = p(x_t|x_{t-1}, u_t)$$

$u_t$  bleibt hier, da diese Stellgröße erst zum Zustandsübergang führt

$$\overline{bel}(x_t) = \int p(x_t|x_{t-1}, u_t) p(x_{t-1}|z_{1:t-1}, u_{1:t}) dx_{t-1}$$

## Mathematische Herleitung des Bayes-Filter (cont.)

- $\overline{bel}(x_t) = \int p(x_t|x_{t-1}, u_t) p(x_{t-1}|z_{1:t-1}, u_{1:t}) dx_{t-1}$
- ▶ für den Übergang in den Zustand  $x_{t-1}$  ist die Steuergröße  $u_t$  irrelevant, kann als zufällig gewählte aufgefasst und daher in den Bedingungen für  $p(x_{t-1}|z_{1:t-1}, u_{1:t})$  weggelassen werden
- ▶ Es ergibt sich mit

$$\begin{aligned} \overline{bel}(x_t) &= \int p(x_t|x_{t-1}, u_t) p(x_{t-1}|z_{1:t-1}, u_{1:t-1}) dx_{t-1} \\ &= \int p(x_t|x_{t-1}, u_t) bel(x_{t-1}) dx_{t-1} \end{aligned}$$

die 2. Zeile des Bayes-Filter-Algorithmus

## Mathematische Herleitung des Bayes-Filter (cont.)

- ▶ Für konkrete Implementierungen des Bayes-Filter-Algorithmus werden drei Wahrscheinlichkeitsverteilungen benötigt:
  - ▶ der initiale *belief*  $p(x_0)$
  - ▶ die Wahrscheinlichkeit der Messungen  $p(z_t|x_t)$
  - ▶ die Zustandsübergangswahrscheinlichkeiten  $p(x_t|u_t, x_{t-1})$

## Markov-Annahme

- ▶ Die Annahme, dass ein Zustand komplett ist, wird **Markov-Annahme** (engl. *Markov assumption*) genannt.
- ▶ Die Annahme fordert, dass vergangene und zukünftige Daten unabhängig sind, wenn man den aktuellen Zustand  $x_t$  kennt.

## Markov-Annahme (cont.)

- ▶ Das folgende Beispiel soll verdeutlichen, wie hart diese Annahme ist:
  - ▶ zur Lokalisierung mobiler Roboter werden Bayes-Filter eingesetzt
  - ▶ dabei ist  $x_t$  die Position im Raum (Pose) des Roboters mit Bezug auf eine feste Karte
  - ▶ es gibt Effekte die die Sensormessungen systematisch verfälschen und damit die Markov-Annahme zunichte machen:
    - ▶ Einfluss sich bewegender Personen auf die Sensormessungen
    - ▶ Ungenauigkeiten in den probabilistischen Modellen  $p(z_t|x_t)$  und  $p(x_t|u_t, x_{t-1})$
    - ▶ Rundungsfehler, wenn Näherungen für die Repräsentation des *beliefs* verwendet werden
    - ▶ Variablen innerhalb der Software, die mehrere Stellgrößen beeinflussen

## Markov-Annahme (cont.)

- ▶ Einige dieser Variablen könnten im Zustand berücksichtigt werden, werden aber oft fallen gelassen, um den rechnerischen Aufwand zu verringern
- ▶ In der Praxis sind Bayes-Filter erstaunlich robust gegen die Verletzung der Markov-Annahme

## Darstellung und Berechnung

- ▶ Bayes-Filter können auf unterschiedliche Weise implementiert werden
- ▶ Die Techniken basieren auf unterschiedlichen Annahmen über die Wahrscheinlichkeiten für die Messungen, die Zustandsübergänge und den *belief*
- ▶ Die Algorithmen besitzen dadurch unterschiedliche rechnerische Eigenschaften
- ▶ In den meisten Fällen müssen die *beliefs* angenähert werden
- ▶ Dies hat Auswirkungen auf die Komplexität der Algorithmen
- ▶ Keine der verschiedenen Techniken ist generell zu bevorzugen

## Effizienz

- ▶ Einige Näherungen bedingen polynomiale Laufzeiten in Abhängigkeit von der Dimension des Zustands (z.B. Kalman-Filter)
- ▶ Einige haben exponentielle Laufzeiten (z. B. gridbasierte Techniken)
- ▶ Partikel-basierte Verfahren haben eine Laufzeit, die von der gewünschten Genauigkeit abhängt

## Beispiele

- ▶ Kalman-Filter
  - ▶ Fehlerverteilung der Messfehler durch Gaussverteilung modelliert
  - ▶ Veränderung des Zustandes linear und a-priori bekannt
- ▶ Partikelfilter
  - ▶ von gitterbasierter Lokalisierung abgeleitet
  - ▶ keine Annahme der Normalverteilung des geschätzten Wertes
  - ▶ Schätzung wird als Wahrscheinlichkeitsverteilung modelliert
  - ▶ Verteilung wird an ausgewählten Stellen, den sog. Partikeln diskretisiert und zwischen diesen wird interpoliert.

## Genauigkeit

- ▶ Einige Näherungen eignen sich besser, um eine Reihe von Wahrscheinlichkeitsverteilungen anzunähern
- ▶ Beispielsweise eignen sich Normalverteilungen für uni-modale Wahrscheinlichkeitsverteilungen
- ▶ Histogramme können multimodale Verteilungen besser annähern, aber auf Kosten der Genauigkeit
- ▶ Partikel-Techniken können eine Vielzahl von Verteilungen annähern, wodurch allerdings die Anzahl der Partikel sehr groß werden kann

## Einfachheit der Implementierung

- ▶ Abhängig von der Form der Verteilungen kann die Implementierung der Verfahren aufwändig sein
- ▶ Weil Partikel-Techniken meist sehr einfach für nicht-lineare Systeme implementiert werden können, sind sie zurzeit sehr populär

## Zusammenfassung (cont.)

- ▶ Der *belief* ist die A-Posteriori-Wahrscheinlichkeit des Zustands; gegeben: Alle bisherigen Messungen und Stellgrößen. Der *Bayes-Filter* ist ein Algorithmus um den *belief* zu berechnen. Er ist rekursiv, d.h. der *belief* zum Zeitpunkt  $t$  wird aus dem *belief* zum Zeitpunkt  $t-1$  berechnet.
- ▶ Der Bayes-Filter macht eine *Markov-Annahme*, d.h. der Zustand ist eine komplette Zusammenfassung der Vergangenheit. Diese Annahme trifft in der Praxis meist nicht zu.
- ▶ Bayes-Filter sollten vor ihrem Einsatz in speziellen Anwendungen anhand verschiedener Kriterien wie Genauigkeit, Effizienz und Einfachheit evaluiert werden .

## Zusammenfassung

- ▶ Die Interaktion zwischen einem Roboter und dessen Umgebung wird modelliert als gekoppeltes dynamisches System. Der Roboter setzt dazu Stellgrößen um die Umgebung zu manipulieren und nimmt die Umgebung über Sensormessungen wahr.
- ▶ Die Dynamiken werden charakterisiert durch zwei wahrscheinlichkeitstheoretische Gesetze
  - ▶ die Wahrscheinlichkeitsverteilungen für den Zustandsübergang
  - ▶ die  $p(z_t|x_t)$  die Messungen
 Erstere beschreibt wie der Zustand sich mit der Zeit ändert, die Zweite, wie Messungen wahrgenommen werden.

## Markov-Lokalisierung

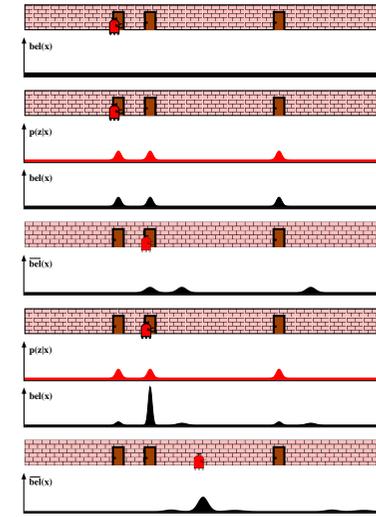
- ▶ Wahrscheinlichkeitstheoretische Lokalisierungsverfahren sind Varianten des Bayes-Filters
- ▶ direkte Anwendung des Bayes-Filters wird **Markov-Lokalisierung** genannt
- ▶ Markov-Lokalisierungen benötigen zusätzlich eine Karte  $m$  der Umgebung
- ▶ Karte  $m$  für das Messmodell erforderlich (teilweise auch im Bewegungsmodell berücksichtigt)
- ▶ Markov-Lokalisierung wird verwendet für
  - ▶ globale Lokalisierung
  - ▶ Positions-Tracking
  - ▶ das Kidnapped-Robot-Problem

## Markov-Lokalisierung (cont.)

Markov\_Lokalisierung( $bel(x_{t-1}), u_t, z_t, m$ ):

1. **for all**  $x_t$  **do**
2.      $\overline{bel}(x_t) = \int p(x_t | u_t, x_{t-1}, m) bel(x_{t-1}) dx_{t-1}$
3.      $bel(x_t) = \eta p(z_t | x_t, m) \overline{bel}(x_t)$
4. **endfor**
5. **return**  $bel(x_t)$

## Markov-Lokalisierung (cont.)



- ▶ falls Anfangsposition und -pose unbekannt, wird für  $bel(x_0)$  eine Gleichverteilung angenommen und

$$bel(x_0) = \frac{1}{|X|}$$

gesetzt, mit  $|X|$  entspricht der Anzahl der gültigen Posen in der Karte

- ▶ falls Anfangspose zwar bekannt, aber nur als Approximation, wird  $bel(x_0)$  mit einer Gaussverteilung initialisiert:

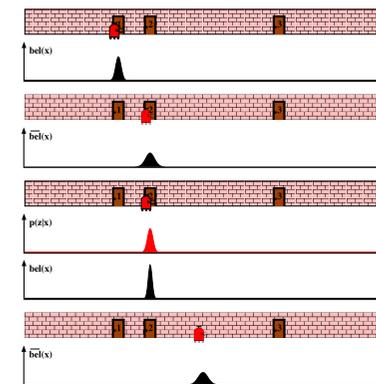
$$bel(x_0) = \det(2\pi\Sigma)^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (x - \bar{x}_0) \Sigma^{-1} (x - \bar{x}_0)^T \right\}$$

- ▶ grundsätzlich ist die Markov-Lokalisierung unempfindlich gegenüber der Repräsentation der zugrunde liegenden Verteilungen

## Extended Kalman Filter Lokalisierung

- ▶ **extended Kalman filter localisation** (oder EKF-Lokalisierung) ist eine spezielle Ausprägung der der Markov-Lokalisierung
- ▶ die EKF-Lokalisierung modelliert den *belief*  $bel(x_t)$  durch unimodale Verteilung mit Mittelwert  $\mu_t$  und Kovarianzmatrix  $\Sigma_t$
- ▶ die Karte ist dabei eine Sammlung von Merkmalen
- ▶ der Roboter misst die Entfernung und Richtung der Merkmale (Feature) in seiner Nähe
  - ▶ erstellt Merkmalsvektor  $z_t = (z_t^1, z_t^2, \dots, z_t^m)$ ; Merkmal  $z_t^i$  bedeutet: Merkmal  $i$  zum Zeitpunkt  $t$
  - ▶ berechnet Übereinstimmungsvektor  $c_t = (c_t^1, c_t^2, \dots, c_t^m)$ ;  $c_t^i = a$ , mit  $a \in \{1, \dots, n\}$  und  $n$  Anzahl der Merkmale der Karte, bedeutet: Merkmal  $i$  des Merkmalsvektors  $z$  zum Zeitpunkt  $t$  stimmt mit Merkmal  $a$  der Karte überein

## Extended Kalman Filter Lokalisierung (cont.)



Beispiel EKF:

- ▶ Annahmen:
  - ▶ Startpose hinreichend genau bekannt (durch Gaussverteilung abgebildet)
  - ▶ Korrespondenzen (jeweilige Übereinstimmungsvektoren  $c_t$ ) zum Zeitpunkt  $t$  bekannt
- ▶ Wahrscheinlichkeit der Messung:  $p(z_t | x_t, m, c_t)$
- ▶ Zustandsübergangswahrscheinlichkeit:  $p(x_t | x_{t-1}, m, c_t)$

## Grid-Lokalisierung

**Grid-Lokalisierung** approximiert den *belief* durch einen **Histogram-Filter** über die Grid-Dekomposition des Zustandsraumes

- ▶ Diskretisierung des Zustandsraumes durch Grid-Zellen  $x$
- ▶ erlaubt multimodale Verteilungen
- ▶ dieser diskrete Bayes-Filter verwaltet eine Menge diskreter Wahrscheinlichkeitswerte:

$$bel(x_t) = \{p_{k,t}\}$$

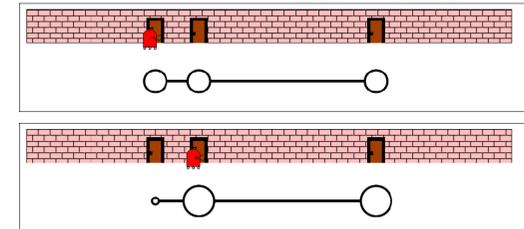
wobei jedes  $p_{k,t}$  zu einer Grid-Zelle  $x_k$  gehört

- ▶ die Vereinigung der Grid-Zellen zum Zeitpunkt  $t$  bildet den Zustandsraum  $X_t$ :

$$domain(X_t) = \{x_{1,t} \cup x_{2,t} \cup \dots \cup x_{K,t}\}$$

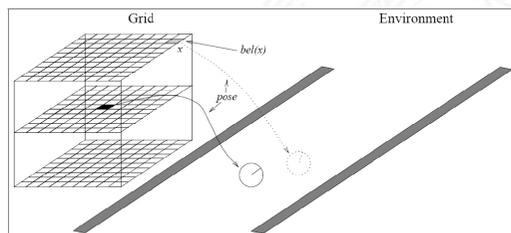
## Grid-Lokalisierung (cont.)

- ▶ zwei Varianten der Grid-Zerlegung üblich:
  - ▶ topologische Grid-Zerlegung
    - ▶ Zelle entspricht signifikanten Ort/Merkmal in der Karte (Flurbeispiel: Tür, Enmündung ...)
    - ▶ üblicherweise sehr grobes Grid
    - ▶ Grid abhängig von Gegebenheiten



## Grid-Lokalisierung (cont.)

- ▶ zwei Varianten der Grid-Zerlegung üblich:
  - ▶ metrische Grid-Zerlegung
    - ▶ gleiche Zellengröße
    - ▶ typische Werte der Zellengröße in der Praxis sind bis zu 15cm
    - ▶ Tiefenauflösung bei bis zu 5° Winkelauflösung
    - ▶ höhere Auflösung gegenüber topologischem Grid zulasten des Aufwandes



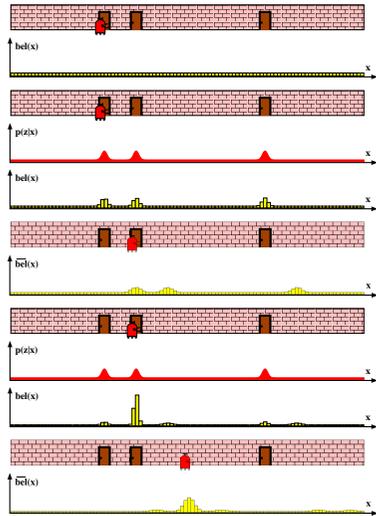
## Grid-Lokalisierung (cont.)

**Grid\_Lokalisierung**( $\{p_{k,t-1}, u_t, z_t, m\}$ ):

1. **for all**  $k$  **do**
2. 
$$\bar{p}_{k,t} = \sum_i [p_{i,t-1} \cdot \text{Bewegungsmodell}(\text{mean}(x_k), u_t, \text{mean}(x_i))]$$
3. 
$$p_{k,t} = \eta \cdot \text{Messmodell}(z_t, \text{mean}(x_k), m) \cdot \bar{p}_{k,t}$$
4. **endfor**
5. **return**  $p_{k,t}$

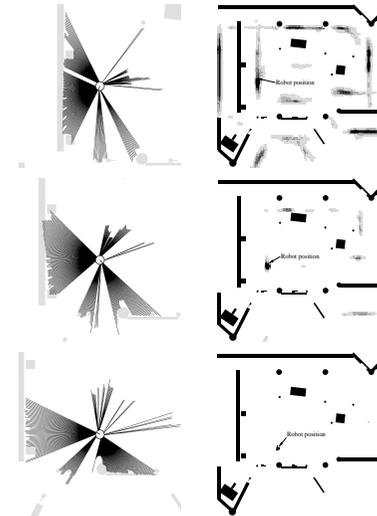
die Funktion *mean* berechnet den Schwerpunkt einer Zelle  $x_i$

## Grid-Lokalisierung (cont.)



- ▶ feingliedrige metrische Zerlegung
- ▶ Annahme der Gleichverteilung für den Startzustand  $x_0$
- ▶ belief  $bel(x_t)$  dargestellt als Histogramm der Zelle
- ▶ zugrundeliegende Verteilungen können mehrdimensional sein

## Grid-Lokalisierung (cont.)



Beispiel: globale Lokalisierung

- ▶ Winkelauflösung  $5^\circ$ , Tiefenauflösung 15 cm
- ▶ Annahme der Gleichverteilung für den Startzustand  $x_0$
- ▶ Messwerte außerhalb der Reichweite wurden fortgelassen
- ▶ graue Bereiche sind Orte erhöhter Aufenthaltswahrscheinlichkeit
- ▶ belief dreidimensional  $(x, y, \theta)$ ; zur Vereinfachung jedoch nur in den  $(x, y)$ -Raum projiziert

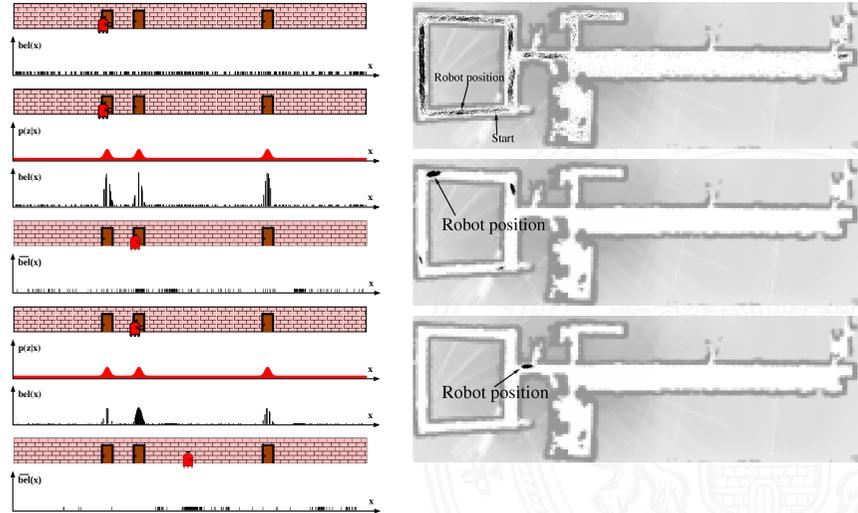
## Monte-Carlo-Lokalisierung

- ▶ **Monte-Carlo-Lokalisierung (MCL)** approximiert den *belief*  $bel(x_t)$  durch eine Menge  $S$  von gewichteten Partikeln
- ▶ Partikelmenge (Samples):  $S = \{s_i | i = 1 \dots N\}$ 
  - ▶ diskrete Näherung einer Wahrscheinlichkeits-Verteilung
  - ▶ Annäherung nahezu beliebiger Verteilungsfunktionen möglich
  - ▶ es können multimodale und unimodale Verteilungen approximiert und überganglos zwischen ihnen gewechselt werden
- ▶ Beschreibung eines Samples mit Wichtung  $m^i = p(z, s^i) \geq 0$ :  $((x, y, \theta), m)$  mit  $\sum_{n=1}^N m^n = 1$
- ▶ eine höhere Anzahl von Partikeln erhöht die Genauigkeit
- ▶ Partikelmengengröße kann adaptiv verändert werden
  - ▶ z. B. hohe Übereinstimmung mit Sensorwerten, d. h. Wichtungsfaktoren hoch, kann Samplmenge verkleinert werden

## Monte-Carlo-Lokalisierung (cont.)

- ▶ Berechnungsverfahren entsprechend Markov-Lokalisierung
  - ▶ Vorhersage: Anwendung des Bewegungsmodells auf die  $N$  Sample der Samplmenge  $S_{k-1}$ ; somit Näherung der Position nach der Bewegung; neue Samplmenge  $\bar{S}_k$  nähert die vorhergesagte Verteilung  $p(x_k | Z^{k-1})$  an
  - ▶ Update: Verarbeitung der aktuellen Messwerte  $z_k$  und Wichtung der Samples mit  $m'_k = p(z_k, s'_k)$ ; Samplmenge  $S_k$  nähert die vorhergesagte Verteilung  $p(x_k | Z^k)$  an

## Monte-Carlo-Lokalisierung (cont.)



[TBF05] Sebastian Thrun, Wolfram Burgard, Dieter Fox,  
Ronald C. Arkin (Hrsg.):  
*Probabilistic Robotics*.  
MIT Press; Cambridge, Massachusetts, 2005  
(Intelligent Robotics and Autonomous Agents). –  
Kapitel 2