Advanced Robot Perception

Fortgeschrittene Konzepte der Wahrnehmung für Robotersysteme

Georg von Wichert, Siemens Corporate Technology

Heutiges Thema: Aktive Wahrnehmung

- Wahrnehmungsprozess immer fehlerbehaftet
  - „Signal-Rausch-Verhältnis“ in der Robotik immer sehr niedrig, d.h. massive Störungen
  - Expliziter Umgang mit den Fehlern!
  - Bayes-Filter als Modell für Wahrnehmungsvorgang unter Unsicherheit

$$\boxed{Bel(x_t) = \underbrace{\eta\, P(z_t \mid x_t)}_{\text{Korrektur/ Mess-Update}} \underbrace{\int P(x_t \mid u_t, x_{t-1})\, Bel(x_{t-1})\, dx_{t-1}}_{\text{Prädiktion / Zeit-Update}}}$$

Heutiges Thema: Aktive Wahrnehmung

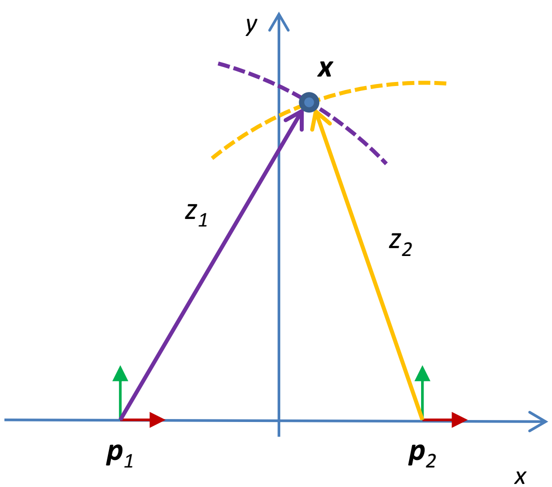
- Was kann ich als Maschine tun, um meine Wahrnehmungsergebnisse zu verbessern?

- Speziell in der Robotik: Roboter sind beweglich
  - Aktive Steuerung der Wahrnehmung möglich
  - Beispiel: Einnehmen eines „guten“ Standpunktes
  - Beispiel: „Gute“ Konfiguration der Sensoren

- Was heißt hier „gut“?

Beispiel: Datenfusion mit dem Extended Kalman Filter

- Beobachtung eines beweglichen Objekts in der Ebene mit zwei ungerichteten Abstandsensoren



$$\mathbf{p}_i = \begin{bmatrix} p_{i1} \\ p_{i2} \end{bmatrix} \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \quad \mathbf{z} = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}$$

$$z_i = \|\mathbf{x} - \mathbf{p}_i\| = \sqrt{(x_1 - p_{i1})^2 + (x_2 - p_{i2})^2}$$

Sensormodell:

$$\mathbf{z}_t = \mathbf{h}(\mathbf{x}_t) + \nu_t = \begin{bmatrix} z_1(\mathbf{x}_t) \\ z_2(\mathbf{x}_t) \end{bmatrix} + \nu_t$$

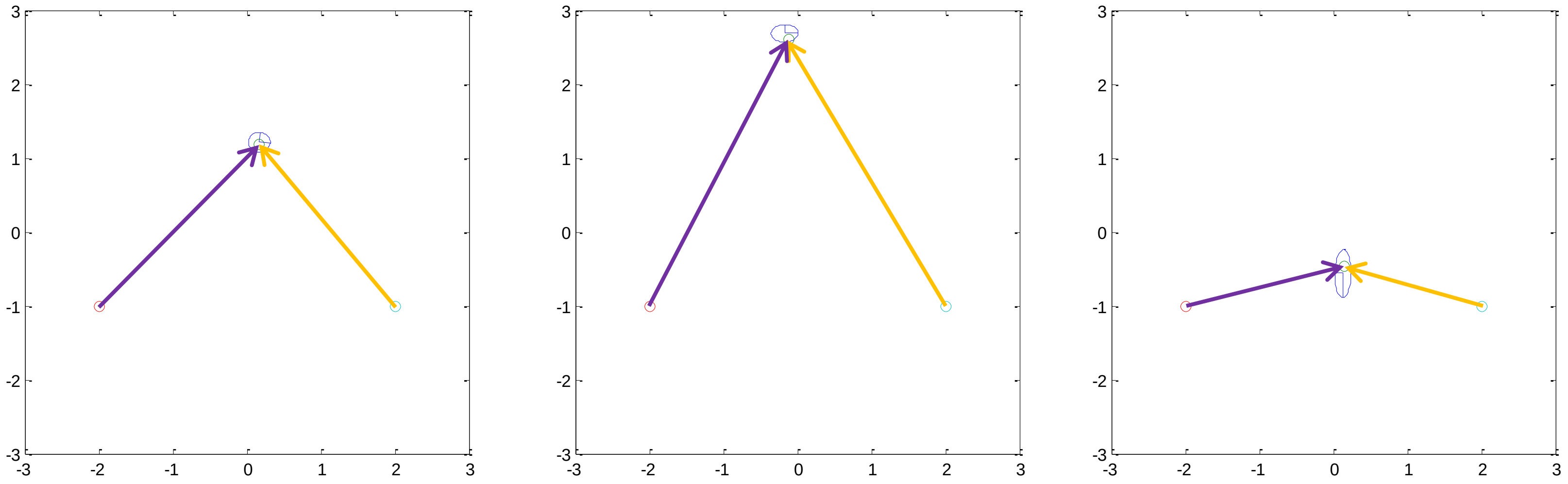
Messrauschen: $\nu_t \sim \mathcal{N}(0, \mathrm{Q})$

Prozessmodell:

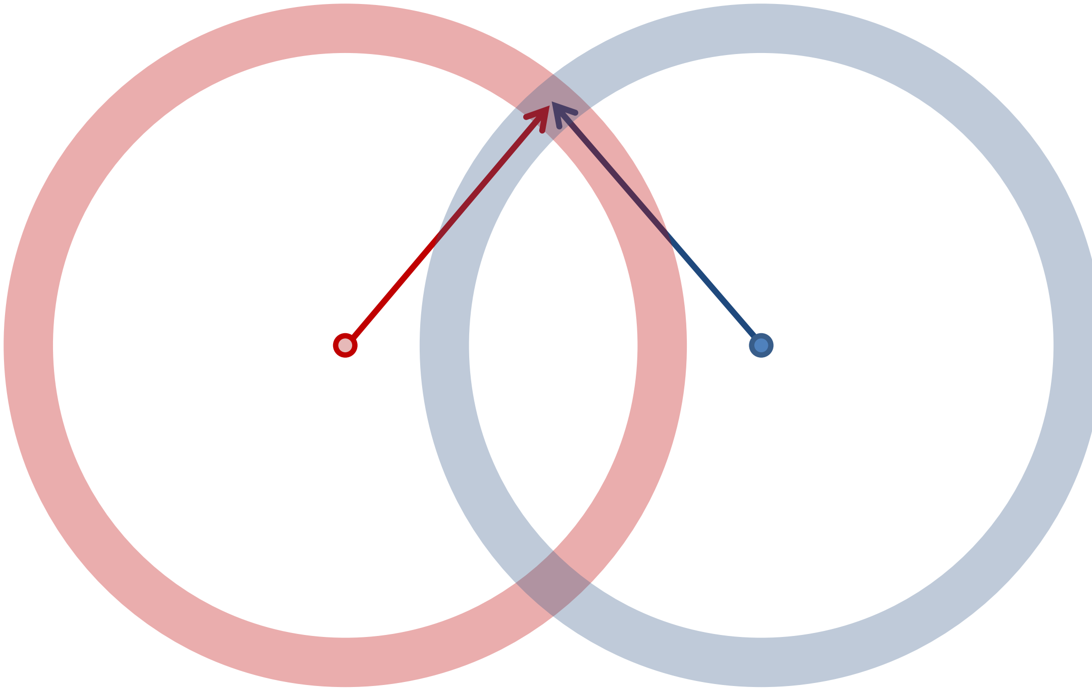
$$\mathbf{x}_t = \mathbf{g}(\mathbf{x}_{t-1}, \mathbf{u}_t) + \xi_t = \begin{bmatrix} 1 & 0 & T & 0 \\ 0 & 1 & 0 & T \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \mathbf{x}_{t-1} + \xi_t$$

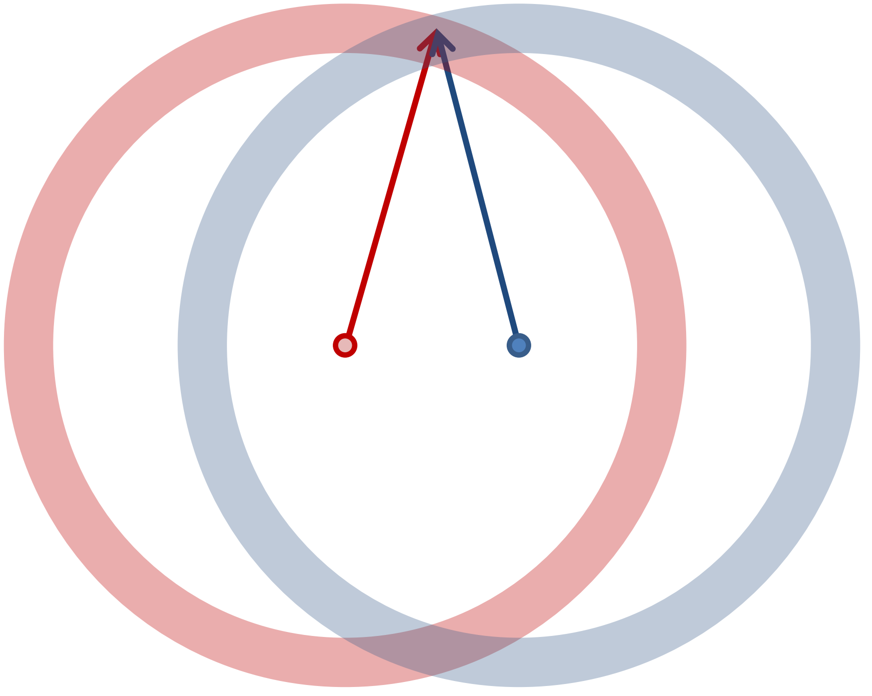
Prozessrauschen: $\xi_t \sim \mathcal{N}(0, \mathrm{R})$

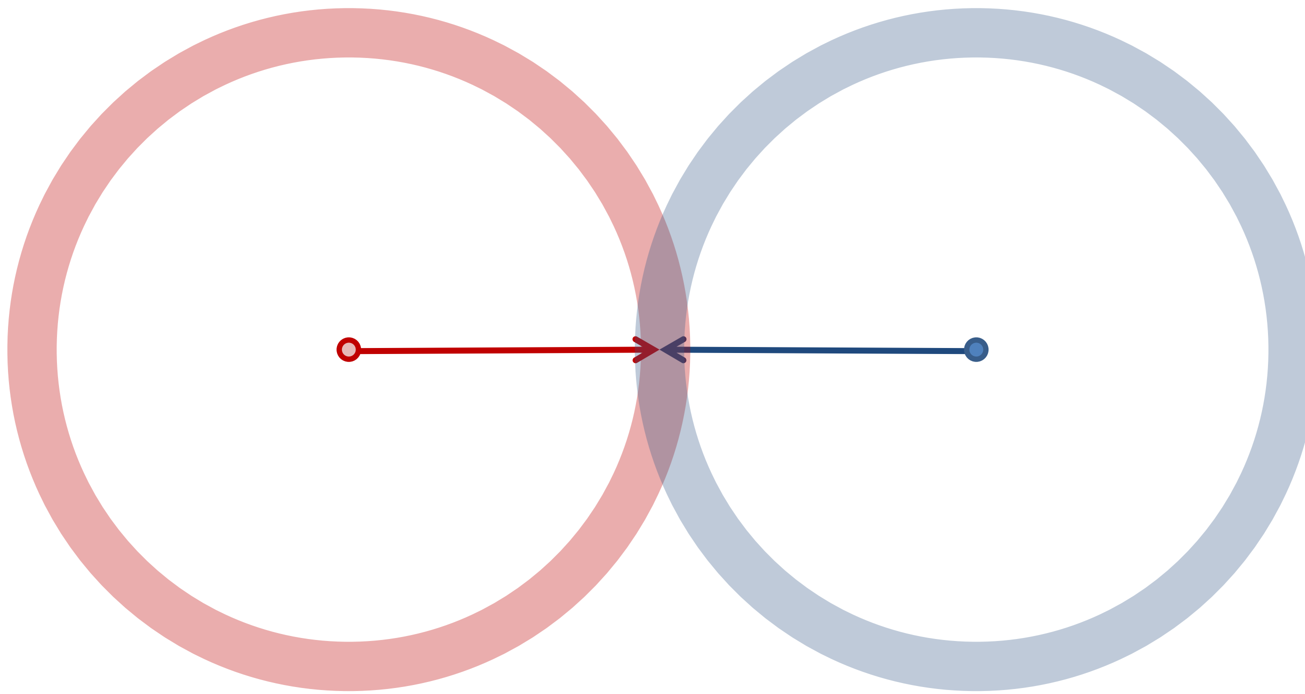
Beispiel: Datenfusion mit dem Extended Kalman Filter



- Beobachtung eines beweglichen Objekts in der Ebene mit zwei ungerichteten Abstandsensoren
- Wie sieht die ideale Konfiguration aus?
  - Rechtwinkliges Dreieck

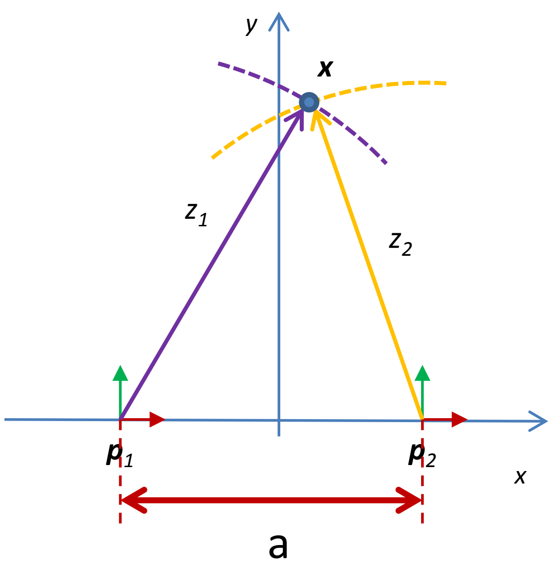






Beispiel: Datenfusion mit dem Extended Kalman Filter

- Beobachtung eines beweglichen Objekts in der Ebene mit zwei ungerichteten Abstandsensoren



$$\mathbf{p}_i = \begin{bmatrix} p_{i1} \\ p_{i2} \end{bmatrix} \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \quad \mathbf{z} = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}$$

$$z_i = \|\mathbf{x} - \mathbf{p}_i\| = \sqrt{(x_1 - p_{i1})^2 + (x_2 - p_{i2})^2}$$

Sensormodell:

$$\mathbf{z}_t = \mathbf{h}(\mathbf{x}_t) + \nu_t = \begin{bmatrix} z_1(\mathbf{x}_t) \\ z_2(\mathbf{x}_t) \end{bmatrix} + \nu_t$$

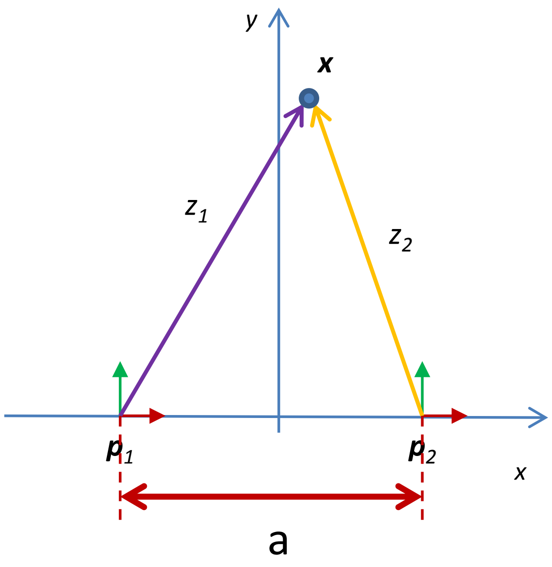
Messrauschen: $\nu_t \sim \mathcal{N}(0, \mathrm{Q})$

Prozessmodell:

$$\mathbf{x}_t = \mathbf{g}(\mathbf{x}_{t-1}, \mathbf{u}_t) + \xi_t = \begin{bmatrix} 1 & 0 & T & 0 \\ 0 & 1 & 0 & T \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \mathbf{x}_{t-1} + \xi_t$$

Prozessrauschen: $\xi_t \sim \mathcal{N}(0, \mathrm{R})$

Aktive Wahrnehmung



- Steuerung von a zur Laufzeit
- Denkbare Stellstrategie:
  - Mögl. rechtwinkliges Dreieck herstellen!
  - Natürlich nur bei diesem Aufbau sinnvoll!
- Wir suchen ein grundlegendes Prinzip für die Steuerung der Wahrnehmungsparameter bzw. Wahrnehmungsaktionen!
- Frage: Können wir quantifizieren, wie viel wir die **Unsicherheit** über den unsicheren System/Umweltzustand durch die jeweilige Messung verringern?
  - Maximieren!

Informationstheorie

- Entwickelt für Kommunikationstechnik
  - "A Mathematical Theory of Communication“, 1948
    - Zentrales Thema: Übertragung von Information über einen gestörten Kanal
    - Applikationen in Kodierung und Kryptographie
- Wir werden das hier ohne Herleitung einführen
  - Details: Spezialvorlesung, **Lehrstuhl für Nachrichtentechnik,** Prof. Kramer
  - Wir wollen hier nur sehen, dass/wie wir die Informationstheorie für unsere Zwecke nutzen können!

Claude Elwood Shannon
30.4.1916 – 24.2.2001

Was ist Information?

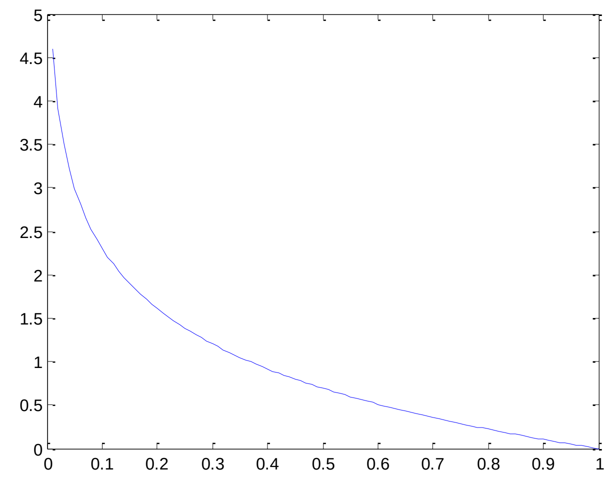
- Für eine diskrete „Quelle“ mit einem endlichen „Alphabet“ $\mathcal{X} = \{x_o, x_1, \ldots, x_{M-1}\}$, wobei die Wahrscheinlichkeit für die jeweiligen Ereignisse durch $P(X = x_k) = p_k$ gegeben ist, ist der Informationsgehalt

$$I(x_k) = \log \frac{1}{p_k} = -\log(p_k)$$

- Logarithmus zur Basis 2: Information in bit

Was ist Information?

$$I(x_k) = \log \frac{1}{p_k} = -\log(p_k)$$



- Der Informationsgehalt entspricht dem „Grad der Überraschung“ bei der Beobachtung eines Ereignisses x_k
- Je seltener das Ereignis, umso größer ist der Informationsgehalt
  - Wenn ein häufiges Ereignis auftritt, erhalten wir wenig Information

Entropie

- Entropie ist der **mittlere Informationsgehalt** eines Ereignisses

$$H(X) = E_x[I(x)] = \sum_k P(X = x_k) I(x_k) = -\sum_k p_k \log p_k$$

- Das Äquivalent für kontinuierliche Zufallsvariablen ist die differentielle Entropie

$$H(X) = -\int p(x) \log(p(x)) \; dx$$

- Die Entropie ist ein Maß für die Unsicherheit der Verteilung p(x).
- Sie ist maximal, wenn $p(x)$ eine Gleichverteilung ist.

Bedingte (differentielle) Entropie

- Nehmen wir an, wir haben zwei Zufallsvariablen *x* und *z* mit einer Verbundverteilung *p*(*x*, *z*)
- Wie groß ist die Entropie von *x*, wenn wir *z* kennen?

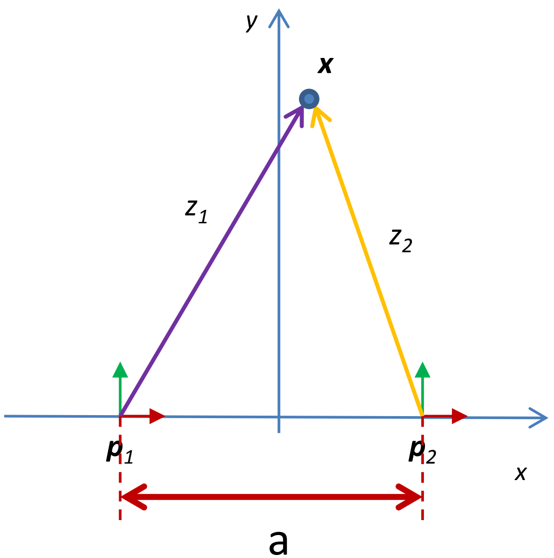
$$H(X|Z) = -\iint p(x,z)\log(p(x|z))\;dx\;dz$$

$$= -\iint p(x|z)p(z)\log(p(x|z))\;dx\;dz$$

$$= -\int p(z)\underbrace{\int p(x|z)\log(p(x|z))\;dx}_{H(X|z)}\;dz = \int p(z)H(X|z)\;dz = E_z(H(X|z))$$

z.B. Entropie von *x*, nachdem wir eine bestimmte Messung *z* gemacht haben

Aktive Wahrnehmung



- Steuerung von a zur Laufzeit

- Denkbare Stellstrategie:
  - Mögl. rechtwinkliges Dreieck herstellen!
  - Natürlich nur bei diesem Aufbau sinnvoll!

- Wir suchen ein grundlegendes Prinzip für die Steuerung der Wahrnehmungsparameter bzw. Wahrnehmungsaktionen!

- Minimieren der Entropie nach der Messung

$$H_a(X_t \mid Z_t) \quad \text{wobei} \quad bel(x_t) = p(x_t \mid z_t(a))$$

durch **geeignete Wahl** der Aktion ***a***

Entropie & Kalman-Filter (EKF)

Prädiktion:

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

$$\bar{\mu}_t = g(u_t, \mu_{t-1})$$

$$\bar{\Sigma}_t = G_t \Sigma_{t-1} G_t^T + R_t$$

$$G_t = \left. \frac{\partial g(u_t, x_{t-1})}{\partial x_{t-1}} \right|_{\mu_{t-1}}$$

Korrektur:

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

$$bel(x_t) = p(x_t \mid z_t) \qquad \begin{cases} K_t = \bar{\Sigma}_t H_t^T (H_t \bar{\Sigma}_t H_t^T + Q_t)^{-1} \\ \mu_t = \bar{\mu}_t + K_t (z_t - h(\bar{\mu}_t)) \\ \Sigma_t = (I - K_t H_t) \bar{\Sigma}_t \end{cases}$$

$$\sim N(x_t; \mu_t, \Sigma_t)$$

$$H_t = \left. \frac{\partial h(x_t)}{\partial x_t} \right|_{\bar{\mu}_t}$$

Aktive Wahrnehmung mit EKF

Ansatz: Minimieren der Entropie nach der Messung *z*(*a*)

$$H_a(X_t|Z_t) = -\int p(z_t(a)) \int p(x_t|z_t(a)) \log(p(x_t|z_t(a)))\ dx_t\ dz_t(a)$$

$$= \int p(z_t(a)) H(X_t|z_t(a))\ dz_t(a)$$

Einsetzen!

Spezialisierung auf EKF:

$$p(x\,|\,z_t(a)) \sim N(x_t; \mu_t, \Sigma_t)$$

Entropie der n-dimensionalen Normalverteilung:

$$H(X) = \frac{n}{2} + \frac{1}{2}\log((2\pi)^n|\Sigma|) \qquad x \in \mathbb{R}^n$$

Aktive Wahrnehmung mit EKF

Ansatz: Minimieren der Entropie nach der Messung *z*(*a*)

$$H_a(X_t|Z_t) = -\int p(z_t(a)) \int p(x_t|z_t(a)) \log(p(x_t|z_t(a)))\ dx_t\ dz_t(a)$$

$$= \int p(z_t(a)) H(X_t|z_t(a))\ dz_t(a)$$

$$= \int p(z_t(a)) (\frac{n}{2} + \frac{1}{2} \log((2\pi)^n |\Sigma_t|))\ dz_t(a)$$

$$= \int (\frac{n}{2}\ p(z_t(a)) + \frac{1}{2}\ p(z_t(a)) \log((2\pi)^n |\Sigma_t|))\ dz_t(a)$$

$$= \underbrace{\frac{n}{2} \int p(z_t(a)) dz_t(a)}_{c} + \frac{1}{2} \int p(z_t(a)) \log((2\pi)^n |\Sigma_t|)\ dz_t(a)$$

Aktive Wahrnehmung mit EKF

Achtung!

$H_a(X_t|Z_t) \neq H_{t,a}$

Ansatz: Minimieren der Entropie nach der Messung *z*(*a*)

$$H_a(X_t|Z_t) = c + \frac{1}{2} \int p(z_t(a)) \log((2\pi)^n |\Sigma_t|) \; dz_t(a)$$

Kalmanfilter: Σ_t ist unabhängig von z !!!

$$= c + \log((2\pi)^n \boxed{|\Sigma_{t,a}|}) \underbrace{\int \frac{p(z(a))}{2} \; dz(a)}_{d}$$

$$K_{t,a} = \bar{\Sigma}_t H_{t,a}^T (H_{t,a} \bar{\Sigma}_t H_{t,a}^T + Q_t)^{-1}$$

$$\mu_t = \bar{\mu}_t + K_{t,a}(z_t - h_a(\bar{\mu}_t))$$

$$\Sigma_{t,a} = (I - K_{t,a} H_{t,a}) \bar{\Sigma}_t$$

$$H_{t,a} = \left. \frac{\partial h_a(x_t)}{\partial x_t} \right|_{\bar{\mu}_t}$$

Um $H_a(X_t/Z_t)$ zu minimieren, müssen wir durch passende Wahl von a_t die Determinante von $\Sigma_{t,a}$ minimieren!!!

$$a_t^* = \arg\min_{a_t} |\Sigma_{t,a}|$$

Weil $\Sigma_{t,a}$ nicht von z_t abhängt, können wir das bereits vor der tatsächlichen Messung ausrechnen!

Aktive Wahrnehmung

Ansatz: Minimieren der Entropie nach der Messung *z*(*a*)

$$a_t^* = \arg\min_{a_t} H_a(X_t | Z_t) \xrightarrow{\text{Kalman-Filter}} a_t^* = \arg\min_{a_t} |\Sigma_{t,a}|$$

- Warum ist das toll?
  - Grundlegendes Prinzip! Das wollten wir doch….
  - Unabhängig von dem spezifischen Problem!
    - Das genaue Prozess und Beobachtungsmodell haben wir für die prinzipielle Betrachtung gar nicht erst benötigt
    - Das Modell können wir jetzt einfach in die Formeln einsetzen

$$H_{t,a} = \left. \frac{\partial h_a(x_t)}{\partial x_t} \right|_{\bar{\mu}_t}$$

Environment Perception:
MPG Framework for 6D Pose Estimation

SIEMENS

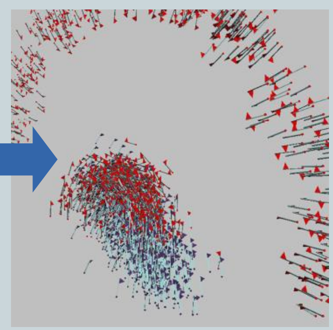
Objectives

- Full exploitation of all available information related to pose
- Consistent handling of uncertainties
- Appropriate handling of orientation
- Efficiency

Approach

- 6D pose representation:
  Dual quaternions
- Strictly probabilistic (poses, observation models)
- Representation of pdfs:
  “Mixtures of *projected* Gaussians” (MPGs)

Inferred pdfs for 6D poses



After fusion

Using more features

- Unified framework for incremental fusion of diverse – possibly “weak” - observations / information sources (“Forward reasoning”)
- Efficiency due to parametric representation and controlled approximations

W. Feiten, P. Atwal, R. Eidenberger, T. Grundmann, *6D Pose Uncertainty in Robotic Perception*, In: Advances in Robotics Research , pp 89-98, Springer, 2009

SIEMENS

Active Perception for Robots: Example

Initial scenario

Observations