Stereovision: Grundlagen

Matthias Behnisch *mbehnisc@techfak.uni-bielefeld.de*

Februar 2005

Inhaltsverzeichnis

1 Einleitung **2**

2 Räumliches Sehen **2**

3 Binokulare Perspektive **4**
3.1 Hering Koordinaten 5

4 Disparitäten **6**
4.1 Horizontale Disparität 6
4.2 Vertikale Disparität 7

5 Berechnung der Tiefe **7**

6 Fazit **8**

7 Quellen **8**

1 Einleitung

Bei der Entwicklung humanoider Roboter orientiert man sich stark am Vorbild des Menschen. Nicht nur der Körperbau ist ähnlich, zusätzlich versucht man seine kognitiven Fähigkeiten zu simulieren. Der Mensch nimmt seine Umgebung mit Hilfe seiner Sinne war, sie liefern ihm ein Bild seiner Umwelt und ohne sie hätte er keine Möglichkeit mit ihr in Interaktion zu treten. Ein wichtiger, vielleicht der wichtigste Sinn, ist die Fähigkeit zu sehen. Auf keine andere Weise kann ein so präzises und umfassendes Bild der Umgebung aufgenommen werden. Jedoch ist ein inhärentes Problem des Sehens mit Hilfe von Augen bzw. Kameras das die eigentlich 3-dimensionale Umgebung auf eine nur 2-dimensionale Ebene abgebildet wird, d.h die unmittelbare Information über die Position von Gegenständen im Raum geht verloren. Trotzdem hat der Mensch die Fähigkeit räumlich zu sehen, also lässt sich die 3-dimensionale Struktur der Umwelt aus 2-dimensionalen Bildern rekonstruieren. In dieser Ausarbeitung werde ich die Grundlagen des räumlichen Sehens *(Stereovision)* beschreiben.

2 Räumliches Sehen

Es gibt verschiedene Möglichkeiten aus Bildern auf räumliche Tiefe zu schließen, nämlich aus einzelnen Bildern, aus durch Bewegung entstehenden Bildfolgen und aus Bildpaaren.
einzelne Bilder: Aus einzelnen Bildern lässt sich aufgrund der *Schattierung* Tiefeninformation gewinnen, wie Abbildung 1 zeigt. Es ist deutlich zu erkennen, dass hier eine Kugel abgebildet ist. Dies geht aus dem Grauwertverlauf auf der Oberfläche hervor, der die für kugelförmige Objekte spezifische Form zeigt. Des weiteren erhält man einen zusätzlichen Hinweis durch die *Verdeckung*. Da der Hintergrund nur teilweise zu sehen ist und von der Kugel verdeckt wird, muss dieser hinter der Kugel sein. Diese relative Position von Objekten zueinander lässt sich auch durch die *Größe bekannter Objekte* erkennen. Beispielweise ist

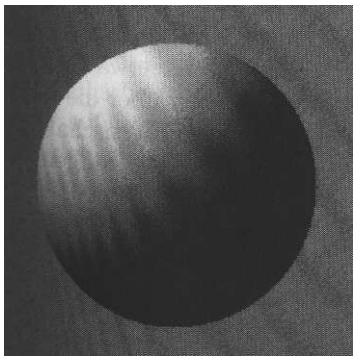

Abbildung 1: Räumliche Wahrnehmung durch Schattierungen

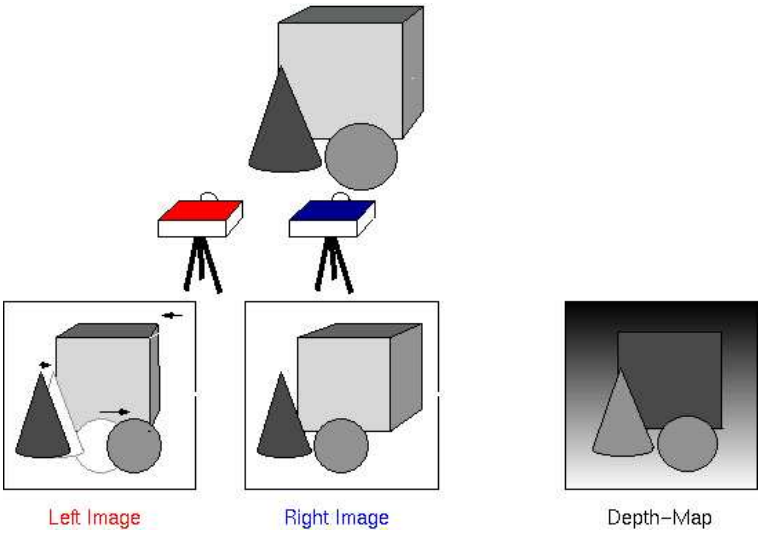

Abbildung 2: Räumliches Sehen durch zwei Kameras

die ungefähre Größe eines Autos bekannt und ein Auto, das wesentlich kleiner erscheint als ein anderes, muss sich folglich in einer größeren Entfernung befinden.
Bildfolgen: Bei Bildfolgen kann die Position verschiedener Objekte durch die *Bewegungsparallaxe* bestimmt werden. Bewegt sich ein Beobachter mit Blick auf mehrere Gegenstände mit unterschiedlicher Entfernung zu ihm, so scheint es, als ob sich diese mit einer jeweils unterschiedlichen Geschwindigkeit bewegen. Nahgelegene Objekte bewegen sich wesentlich schneller als weiter entfernte. Zusätzlich kann es durch die Bewegung zu wechselnden Verdeckungen zwischen Objekten kommen, aus denen sich auf die relative Positionierung schliessen lässt.
Bildpaare: Durch das Sehen mit zwei Augen werden zwei unterschiedliche Perspektiven derselben Szene aufgenommen. Wie in Abbildung 2 zu sehen resultieren daraus zwei leicht unterschiedliche Bilder, aus denen dann die Tiefeninformation erzeugt werden kann.

Obwohl der Mensch alle diese Tiefenhinweise simultan benutzt, um sich ein Bild von der räumlichen Struktur seiner Umwelt zu machen, liefert die *binokulare* Wahrnehmung, d.h. das Sehen mit zwei Augen, die genauste Information. Während die anderen Techniken zumeist nur relative Positionen bestimmen oder nur sehr ungenau sind, lässt sich die Tiefe durch das binokulare Sehen sehr präzise und in absoluten Werten berechnen. Daher beschränken sich die folgenden Kapitel auf das räumliche Sehen mit Bildpaaren.

3 Binokulare Perspektive

Die Ursache für die Unterschiede in den beiden Bildern beim binokularen Sehen liegt in der Positionsdifferenz der Augen. Bei der Untersuchung der Abbildungsgeometrie unterscheidet man zwei Fälle. Zum einem den Fall paralleler Blickachsen, in dem beide Augen parallel ausgerichtet sind und sich die Achsen nicht schneiden. Dieser ist einfacher zu handhaben, jedoch im Bezug auf das Vorbild Mensch nicht sehr wirklichkeitsnah, da sich bei ihm die Augen unabhängig voneinander bewegen können. Beim zweiten Fall schneiden sich die Blickachsen in einem Punkt, sind also konvergent. Den Fall der konvergierenden Achsen werde ich nun genauer erläutern.

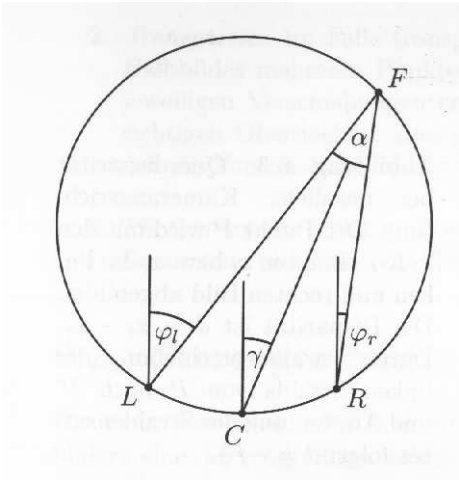

Abbildung 3: Binokulare Perspektive: Vergenz α, Version γ

Abbildung 3 zeigt die Geometrie der binokularen Perspektive. Die Punkte L und R bezeichnen die Position des linken bzw. rechten Auges, deren Blickrichtungen in Bezug auf die Ruheposition sind die Winkel φ_l und φ_r. Diese Achsen schneiden sich im Fixierpunkt F unter dem *Vergenzwinkel* α, der für alle Fixierpunkte auf dem abgebildeten Kreis durch L, R und F gleich ist. Den Kreis konstanter Vergenz nennt man auch den *Vieth-Müller-Kreis*. Für die Vergenz gilt:

$$\alpha = \varphi_l - \varphi_r \tag{1}$$

Den Mittelwert der beiden Blickrichtungen bezeichnet man als *Version* γ:

$$\gamma = \frac{1}{2}(\varphi_l + \varphi_r) \tag{2}$$

Anschaulich kann man die Version als die Blickrichtung der zyklopischen Sicht interpretieren. Beide Halbbilder der binokularen Perspektive werden zu einem Bild verschmolzen, welches der Sicht eines imaginären zyklopischen Auges C

zwischen dem linken und rechten Auge entspricht.
Als Vergenz und Version bezeichnet man auch die Augenbewegungen, die zur Veränderung dieser beiden Winkel führen. Gegensinnige Bewegungen beider Augen zur Nase heißen Konvergenz und die umgekehrte Richtung zu den Schläfen Divergenz. Gleichsinnige Bewegungen nach links oder rechts nennt man dann Versionen.

3.1 Hering Koordinaten

Jeden Punkt im Gesichtsfeld des Beobachters kann man nun mit Hilfe der Vergenz und der Version lokalisieren, diese beiden Winkel bilden die sogenannten *Hering-Koordianten*. In Abbildung 4 sind die Hering Koordianten der horizontalen Ebene dargestellt, also die Ebene, die sich bei nicht geneigten Kopf des Betrachters horizontal von den Augen zum Horizont erstreckt. Punkte mit konstanter Vergenz liegen auf den *Vieth-Müller-Kreisen* durch die Knotenpunkte und den Fixierpunkt. Die Kurven konstanter Version werden *Hillebrand-Hyperbeln* genannt und verlaufen duch jeweils einen der beiden Knotenpunkte. Wenn die Vergenz sowie die Version eines Punktes auf dieser Ebene bekannt sind, kann nun auch die absolute Position zum Betrachter berechnet werden. Die Verhältnisse für die binokulare Perspektive des Menschen erhält man mit der Längeneinheit 3,25 cm, dies entspricht der Hälfte des Augenabstandes.

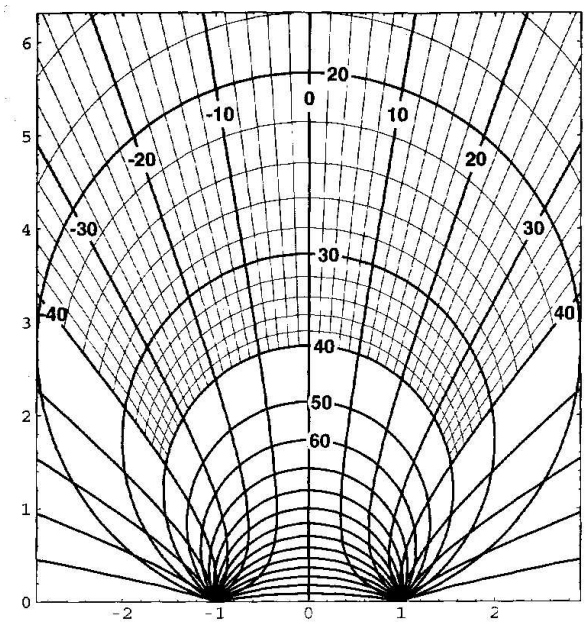

Abbildung 4: Hering-Koordinaten der horizontalen Ebene

4 Disparitäten

Wie bereits erwähnt beruht die binokulare Stereovision auf den Unterschieden zwischen den beiden Bildern des linken und des rechten Auges (Abbildung 2). Diese Bildunterschiede werden auch *Disparitäten* genannt. Es gibt verschiedene Typen:

- **Horizontale Disparität (Querdisparität):**
  Ein einzelner Punkt wird in beiden Bildern an unterschiedliche Bildkoordinaten abgebildet. Die horizontale Differenz zwischen diesen nennt man horizontale Disparität.

- **Vertikale Disparität:**
  Außerhalb der horizontalen Ebene treten auch vertikale Differenzen zwischen korrespondierenden Bildpunkten auf.

- **Orientierung von Linien:**
  Geraden im Raum werden in beiden Augen mit unterschiedlicher Steigung abgebildet.

- **Form- und Größenunterschiede:**
  Ausgedehnte Körper werden durch die zwei verschiedenen Perspektiven unterschiedlich verzerrt, z.B. kann das Seitenverhältnis von Rechtecken in beiden Bildern unterschiedlich abgebildet werden.

- **Schattierungsdisparitäten:**
  Die Reflexion von Licht an glatten Oberflächen ist von der Position des Betrachers abhängig, daher wird das Licht in unterschiedlicher Stärke zu den Positionen beider Augen reflektiert. Daraus resultieren für einen identischen Punkt auf einer Oberfläche unterschiedliche Schattierungen in beiden Bildern.

- **Monokulare Verdeckung:**
  An den Rändern von Objekten sind oft Teile der weiter hinten liegenden Flächen nur von einem Auge aus sichtbar. So existieren Bildelemente, die keine Entsprechung im Bild des anderen Auges haben.

4.1 Horizontale Disparität

Abbildung 5 zeigt, wie man die horizontale Disparität berechnen kann. Die Augen (L und R) fixieren den Fixierpunkt F. Der Punkt P erscheint unter den Winkeln ϱ_l und ϱ_r im Bild des linken bzw. rechten Auges. Wir definieren nun die Querdisparität dieses Punktes als die Differenz dieser Winkel:

$$\delta := \varrho_r - \varrho_l \tag{3}$$

Mit Hilfe der Winkelsummen der Dreiecke LFS und RPS kann man die Disparität auch als die Differenz der Vergenzwinkel α_F und α_P ausdrücken:

$$\alpha_F + \varrho_l = \alpha_P + \varrho_r$$

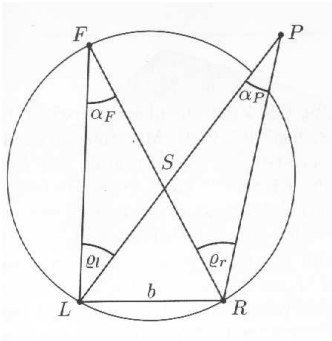

Abbildung 5: Berechnung der horizontalen Disparität

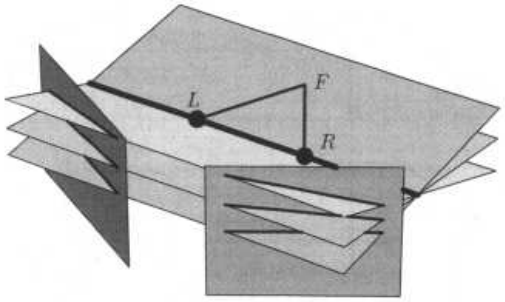

Abbildung 6: Epipolare Ebenen

$$\delta = \varrho_r - \varrho_l = \alpha_F - \alpha_P \tag{4}$$

Alle Punkte auf dem Vieth-Müller-Kreis haben keine Querdisparität, während Punkte innerhalb des Kreises negative und außerhalb positive Querdisparität haben.

4.2 Vertikale Disparität

Bei Punkten, die außerhalb der horizontalen Sichtebene des Betrachters liegen, treten neben den horizontalen auch vertikale Disparitäten auf.
Abbildung 6 zeigt die epipolare Ebene, die durch den Fixierpunkt sowie durch die beiden Knotenpunkte verläuft, sowie die Bildebenen der Kameras, die senkrecht zu den gedachten Geraden durch den linken bzw. rechten Knotenpunkt und den Fixierpunkt stehen. Befindet sich der Fixierpunkt nicht auf der horizontalen Ebene, so ist die entsprechende epipolare Ebene um die Achse LR geneigt. Schneidet man die epipolaren Ebenen mit den Bildebenen, erhält man die epipolaren Geraden. Je höher oder tiefer der Fixierpunkt zur horizontalen Ebene steht, desto stärker sind diese Geraden geneigt. Sie werden in beiden Bildebenen mit entgegengesetzter Neigung abgebildet. Legt man die beiden korrespondierdenden epipolaren Geraden übereinander, so scheiden sich diese in einem Punkt, an dem keine vertikale Disparität auftritt. Links und rechts dieses Schnittpunktes wird durch die Geraden ein festes Verhältnis zwischen vertikaler und horizontaler Disparität vorgegeben. Dieses Verhätnis ist fest an den Vergenzwinkel gekoppelt, d.h durch Berechnung der beiden Disparitäten hat man die Möglichkeit den Vergenzwinkel zu bestimmen.

5 Berechnung der Tiefe

Um aus den zwei Bildern der binokularen Perspektive die Tiefeninformation zu bestimmen, müssen die Disparitätswerte für jeweils zwei korrespondierende

Bildpunkte berechnet werden. Dies geschieht mit Hilfe von Stereo-Algorithmen, die ich hier jedoch nicht weiter erläutern werde. Hat man diese Disparität gegeben und kennt man den Basisabstand der Kameras sowie deren Blickrichtung, so kann man nun die absolute Tiefe wie folgt berechnen:

1. Aus den Blickrichtungen wird die Vergenz und Version des Fixierpunktes berechnet:
$$\alpha_F = \varphi_l - \varphi_r$$
$$\gamma_F = \frac{1}{2}(\varphi_l + \varphi_r)$$

2. Wir kennen die Disparität eines Punktes, daraus folgt seine Vergenz (Gleichung 4):
$$\alpha_P = \alpha_F - \delta$$
Mit Hilfe der Hering-Koordinaten (Abbildung 4) wissen wir nun, auf welchen Vieth-Müller-Kreis sich der Punkt befindet.

3. Zusätzlich wird die Version des Punktes benötigt, um die genaue Position zu bestimmen. Diese lässt sich mit Kenntnis der Kameraparameter aus den Bildkoordianten berechnen.

4. Nun kennen wir die Position in Hering-Koordinaten, die einfach in die Raumposition und somit auch in Tiefenwerte umgerechnet werden können.

Wendet man dieses Verfahren auf jeden Bildpunkt an, so erhält man im Idealfall die präzise Tiefeninformation der gesamten Szene.

6 Fazit

Es gibt verschiedenste Möglichkeiten aus Bildern auf die räumliche Struktur der abgebildeten Szene zu schließen. Bei der binokularen Perspektive lässt sich die Lösung des Problems auf das Berechnen von Disparitäten zwischen den beiden Teilbildern zurückführen. In der technischen Anwendung ist dies oft kompliziert und fehleranfällig, da einige Schwierigkeiten existieren. So ist es im Allgemeinen nicht möglich, Disparitäten kontinuierlich über das gesamte Bild zu bestimmen. Durch Verdeckungen, Reflektionen, Transparenz und Rauschen kann nicht jedem Punkt in einem Bild ein koorespondierender Punkt im anderen zugeordnet werden. Jedoch existieren Algorithmen, um eine recht genaue, approximative Tiefenstruktur zu bestimmen.

7 Quellen

- Mallot, Hanspeter A.: Sehen und die Verarbeitung visueller Information, Vieweg 1998
- http://axon.physik.uni-bremen.de/research/stereo/index.html