

# 5

## Kamera-Modelle

Um in der Bildbearbeitung reale Kameras nachbilden zu können, wurden Kamera-Modelle entwickelt, welche deren physikalische Eigenschaften mathematisch möglichst exakt zu beschreiben versuchen. Dieses Kapitel beschreibt nach einer Einführung in die Grundlagen der geometrischen Transformationen sowohl die intrinsische als auch die extrinsische Orientierung eines Kamera-Modells.

### 5.1 Geometrische Transformationen

In der Computergraphik werden hauptsächlich zwei Arten von mathematischen Gebilden verwendet, um geometrische Transformationen im Raum auszudrücken. Im folgenden Abschnitt wird die Anwendung von Transformationsmatrizen erklärt, im Anschluss daran folgt eine kurze Einführung in die Quaternionen-Rechnung.

#### 5.1.1 Transformationsmatrizen

Im Gegensatz zu den in der linearen Algebra für die Darstellung von dreidimensionalen Daten verwendeten Vektoren und Transformationsmatrizen in drei Dimensionen werden in der Computergraphik meist Vektoren und Matrizen für homogene Koordinaten verwendet [15, 14]. Sie können als Erweiterung der dreidimensionalen Vektoren und Matrizen um eine zusätzliche, sogenannte homogene Dimension angesehen werden, was bedeutet, dass der homogene Raum vierdimensional ist. Damit kann jeder Punkt im dreidimensionalen Raum als Gerade im homogenen Gegenstück interpretiert werden.

Der Vorteil von homogenen Koordinaten gegenüber den dreidimensionalen ist [14], dass die Translation eines Punktes nicht mehr durch eine Vektoraddition ausgedrückt werden muss, sondern einheitlich wie die Skalierung und die Rotation durch eine Multiplikation des zu transformierenden Vektor mit einer Transformationsmatrix. Ausserdem können auch perspektivische Abbildungen nach dem Lochkameramodell im homogenen Raum durch eine Matrixmultiplikation beschrieben werden. Da die massgeblichen Komponenten für Translation und Rotation zusätzlich an unterschiedlichen Positionen in den homogenen Matrizen zu stehen kommen, vereinfacht sich die Darstellung von solchen mathematischen Gebilden.

Die drei wichtigsten Transformationsmatrizen beschreiben Translationen, Skalierungen und Rotationen. Soll ein homogener Vektor  $\mathbf{v} = (x, y, z, 1)^T$  um  $(t_x, t_y, t_z)$  in den drei Koordinatenrichtungen verschoben werden, so kann man ihn mit der Transformationsmatrix

$$T(t_x, t_y, t_z) = \begin{bmatrix} 1 & 0 & 0 & t_x \\ 0 & 1 & 0 & t_y \\ 0 & 0 & 1 & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (5.1)$$

multiplizieren. Der resultierende Vektor erhält die Form  $\mathbf{v}' = (x + t_x, y + t_y, z + t_z, 1)^T$ , was genau dem gewünschten Resultat entspricht. Ähnlich sieht die Skalierungsmatrix aus, welche die Komponenten des homogenen Vektors  $\mathbf{v}$  um die reellen Faktoren  $(s_x, s_y, s_z)^T$  skaliert:

$$S(s_x, s_y, s_z) = \begin{bmatrix} s_x & 0 & 0 & 0 \\ 0 & s_y & 0 & 0 \\ 0 & 0 & s_z & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (5.2)$$

Für die Darstellung von Rotationen um einen Winkel  $\theta$  existiert für jede Koordinatenachse eine gesonderte Variante einer homogenen Rotationsmatrix, welche jeweils einen Vektor um den gegebenen Winkel um die gefragte Achse rotiert.

$$R_x(\theta) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta & 0 \\ 0 & \sin\theta & \cos\theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{für die x-Achse} \quad (5.3)$$

$$R_y(\theta) = \begin{bmatrix} \cos\theta & 0 & \sin\theta & 0 \\ 0 & 1 & 0 & 0 \\ -\sin\theta & 0 & \cos\theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{für die y-Achse} \quad (5.4)$$

$$R_z(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta & 0 & 0 \\ \sin\theta & \cos\theta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{für die z-Achse} \quad (5.5)$$

In einem Rechtskoordinatensystem mit nach oben zeigender y-Achse und in negative z-Richtung schauender Kamera wird eine Rotation um die y-Achse als *Gieren* (*yaw*) bezeichnet, die Rotation um die z-Achse als *Rollen* (*roll*) und diejenige um die x-Achse als *Nicken* (*pitch*). Die Bezeichnungen wurden aus der Aviatik übernommen.

### 5.1.2 Quaternionen

Eine weitere Möglichkeit, Rotationen auszudrücken, sind nach [14] sogenannte Quaternionen. Ihre wesentlichen Vorteile sind die im Vergleich zu den homogenen Transformationsmatrizen kompakte Darstellungsweise sowie eine höhere numerische Stabilität. Ein Quaternion ist ein mathematisches Element der Form

$$\dot{\mathbf{q}} = c + xi + yj + zk \quad (5.6)$$

wobei  $c, x, y, z$  reelle und  $i, j, k$  imaginäre Zahlen darstellen. Oft wird auch die Darstellung  $\dot{\mathbf{q}} = c + u$  verwendet, wobei  $c$  der Realteil und  $u$  das reine Quaternion ist. Ein Quaternion kann somit durch ihre sinngemässe Erweiterung einer komplexen Zahl in die vierte Dimension als hyperkomplexe Zahl aufgefasst werden. Für die imaginären Zahlen  $i, j, k$  gelten folgende Regeln:

- $i^2 = j^2 = k^2 = -1$
- $ij = k, ji = -k, jk = i, kj = -i, ki = j, ik = -j$

Ein Rotations-Einheitsquaternion für eine Einheits-Rotationsachse  $\mathbf{r} = (x_r, y_r, z_r), |\mathbf{r}| = 1$  und einen Rotationswinkel  $\vartheta$  wird auf folgende Art gebildet:

$$\dot{\mathbf{q}} = \cos\left(\frac{1}{2} \cdot \vartheta\right) + \sin\left(\frac{1}{2} \cdot \vartheta\right) \cdot n \quad \text{mit } n = \begin{bmatrix} x_r \\ y_r \\ z_r \end{bmatrix} \cdot \begin{bmatrix} i \\ j \\ k \end{bmatrix} \quad (5.7)$$

Die Rotation eines Punktes  $P(x_p, y_p, z_p)$  respektive seines zugehörigen Quaternion  $\dot{\mathbf{p}} = x_p \cdot i + y_p \cdot j + z_p \cdot k$  mit dem Rotations-Quaternion aus Gleichung 5.7 geschieht durch zwei Multiplikationen mit  $\dot{\mathbf{q}} = c + u$  und dem konjugierten Element  $\dot{\mathbf{q}}^* = c - u$ :

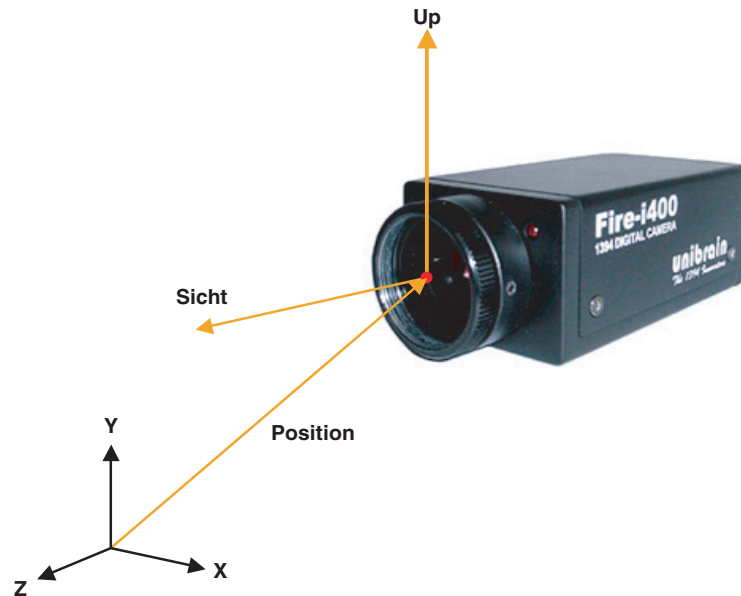
$$\dot{\mathbf{p}}' = \dot{\mathbf{q}} \dot{\mathbf{p}} \dot{\mathbf{q}}^* \quad (5.8)$$

## 5.2 Extrinsische Orientierung

Unter extrinsischer oder äusserer Orientierung eines Kamerasystems versteht man die Positionierung und Ausrichtung der Kamera im Objekt- oder Weltkoordinatensystem. Das entspricht dem Aufstellen, Montieren und Ausrichten einer realen Kamera in einer geeigneten Entfernung zum zu photographierenden Objekt. Mathematisch beschrieben wird die extrinsische Orientierung durch eine Transformationsmatrix. Diese platziert den Koordinatenursprung in der Eintrittspupille der Kamera und lässt sie nach Konvention (Anhang B) in die negative z-Richtung schauen, die y-Achse zeigt nach oben.

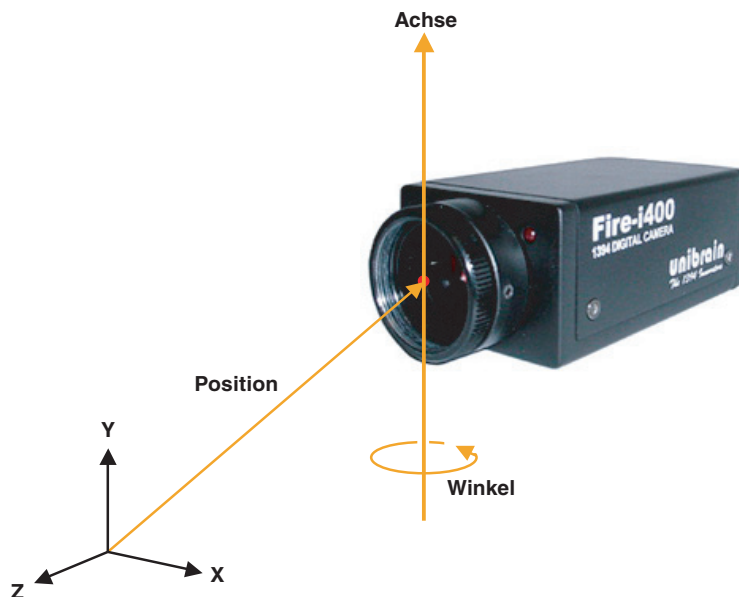
Für eine bereits im Raum positionierte Kamera gibt es zwei unterschiedliche Ansätze, wie sie auf ein Objekt ausgerichtet werden kann, wie also die restlichen Freiheitsgrade bestimmt werden können:

- Durch Angabe eines Sichtvektors, der vom Objektiv zum anvisierten Objekt zeigt, sowie die Angabe eines Up-Vektors, der definiert, in welche Richtung die vertikale Achse des Bildes zeigt (Abbildung 5.1).



**Abbildung 5.1:** Extrinsische Orientierung, definiert durch Positionsvektor, Sichtvektor und Up-Vektor

- Durch Angabe einer Rotationsachse sowie eines Rotationswinkels um diese Achse (Abbildung 5.2).



**Abbildung 5.2:** Extrinsische Orientierung, definiert durch Positionsvektor, Rotationsachse und Rotationswinkel

Da die beiden Orientierungsarten ineinander überführt werden können, wird im Folgenden nur die Transformationsmatrix für die Beschreibung mittels der Rotationsachse und dem Rotationswinkel hergeleitet.

In einem ersten Schritt werden die Rotationsachse  $\mathbf{r} = (x, y, z)^T$  und der Winkel  $\zeta$  in ein Einheitsquaternion der Form  $\mathbf{q} = c + li + mj + nk$  umgewandelt:

$$\dot{\mathbf{q}} = \cos\left(\frac{1}{2}\zeta\right) + \frac{\sin\left(\frac{1}{2}\zeta\right)}{|\mathbf{r}|}xi + \frac{\sin\left(\frac{1}{2}\zeta\right)}{|\mathbf{r}|}yj + \frac{\sin\left(\frac{1}{2}\zeta\right)}{|\mathbf{r}|}zk \quad (5.9)$$

Für Quaternionen wiederum existiert nach [11] eine Umwandlung in eine Rotationsmatrix  $R$ , welche wir als Rotationsanteil für die Beschreibung der extrinsischen Transformationsmatrix benötigen:

$$R = \begin{bmatrix} w^2 + l^2 - m^2 - n^2 & 2lm - 2wn & 2ln + 2wm & 0 \\ 2lm + 2wn & w^2 - l^2 + m^2 - n^2 & 2mn - 2wl & 0 \\ 2ln - 2wm & 2mn + 2wl & w^2 - l^2 - m^2 + n^2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (5.10)$$

Neben der Rotationsmatrix in Gleichung 5.10 wird für die Vervollständigung der extrinsischen Transformationsmatrix noch eine Translation vom Weltkoordinatensystem-Ursprung zur Kameraposition  $\mathbf{p} = (X_0, Y_0, Z_0)^T$  gebraucht, was, wie in Kapitel 5.1 beschrieben, durch eine Matrizen-Multiplikation mit einer homogenen Translationsmatrix erreicht wird:

$$T_{extr} = \begin{bmatrix} 1 & 0 & 0 & X_0 \\ 0 & 1 & 0 & Y_0 \\ 0 & 0 & 1 & Z_0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot R \quad (5.11)$$

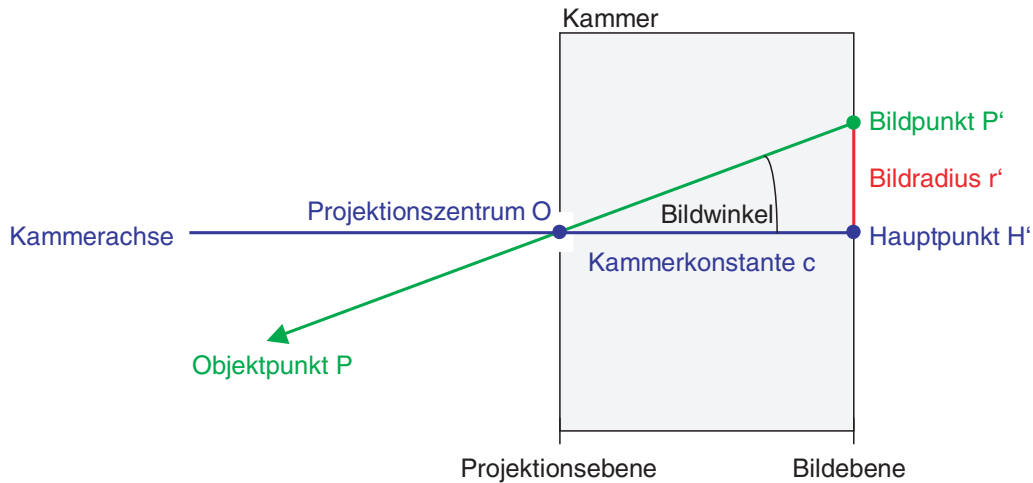
### 5.3 Intrinsische Orientierung

Nach der Positionierung und Ausrichtung der Kamera im Raum muss in einem nächsten Schritt die intrinsische oder innere Orientierung definiert werden. Die daraus entstehende intrinsische Transformationsmatrix bildet den Objektraum auf die Bildebene ab.

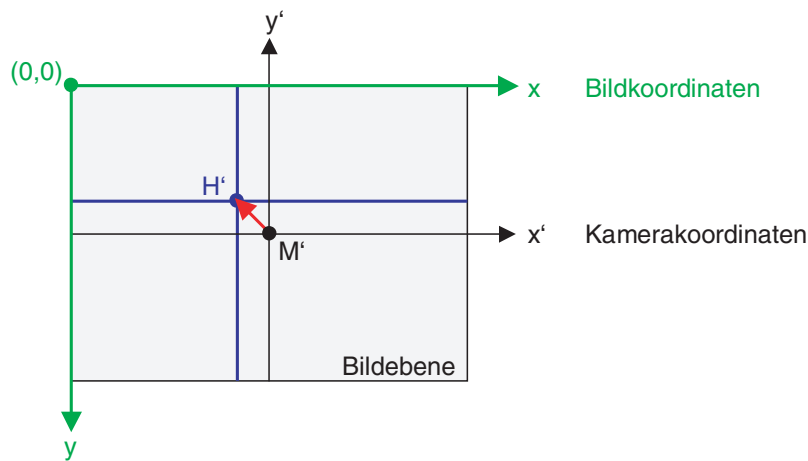
#### 5.3.1 Lochkameramodell

Das Lochkameramodell beschreibt die einfachste mögliche Kamera. Sie enthält keine optischen Linsen und lässt somit Lichtstrahlen als eine einzige Gerade vom Objekt bis auf die Bildebene beschreiben, wodurch das Lochkameramodell eine geometrische Vereinfachung komplizierter optischer Systeme ist.

Die innere Orientierung einer Lochkamera wird durch die Kammerkonstante sowie die Position des Hauptpunkts in der Bildebene beschrieben. Die Kammerkonstante wird oft mit der Brennweite einer Kamera verwechselt, was genau genommen zwei unterschiedliche Parameter sind. Die Kammerkonstante ist eine rein rechnerische Grösse [16], welche die Distanz vom Projektionszentrum zum Hauptpunkt beschreibt. Eine Lochkamera besitzt keine Brennweite, denn diese Eigenschaft wird nur in Linsensystemen beschrieben und hängt, wie im Kapitel 4 physikalisch begründet, von variablen Parametern wie der Wellenlänge ab. Um im Zusammenhang mit weiteren zu besprechenden Kameramodellen, welche eine Brennweite besitzen, nicht für Verwirrung in der Begriffsbildung zu sorgen, ist es trotzdem vertretbar, im Falle der Lochkamera die Kammerkonstante als Brennweite zu bezeichnen, was im Folgenden getan werden soll.



**Abbildung 5.3:** Parameter der intrinsischen Orientierung



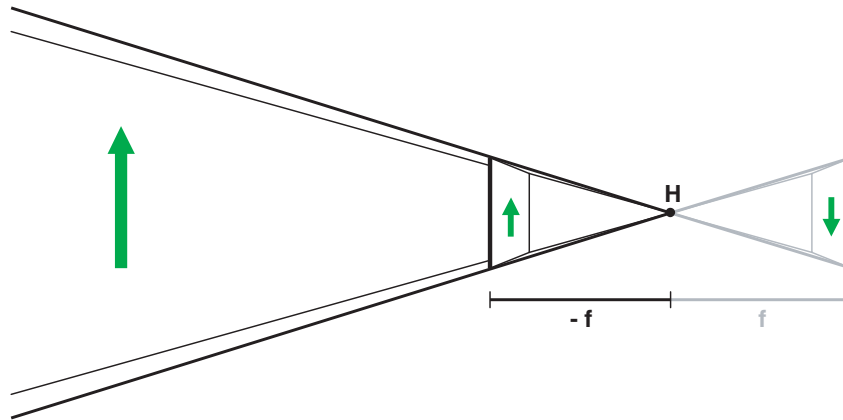
**Abbildung 5.4:** Verschiebung des Hauptpunktes vom Ursprung des Kamerakoordinatensystems

Die Bildung der intrinsischen Transformationsmatrix sei hier in mehreren Schritten hergeleitet.

In einem ersten Schritt werden die Weltkoordinaten in das Kamerakoordinatensystem skaliert. Da in diesem System alle Punkte in einer Ebene, die sich in einem Abstand der Größe der Brennweite vom Projektionszentrum befindet und senkrecht zum Hauptsehstrahl steht, abgebildet werden, muss es als zweidimensional bezeichnet werden. Die entsprechende Skalierungsmatrix sieht folgendermassen aus:

$$F = \begin{bmatrix} -f & 0 & 0 & 0 \\ 0 & -f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad (5.12)$$

Die Begründung für die negativen Brennweiten  $f$  wird in Abbildung 5.5 gegeben. Es wird dadurch möglich, die Bildebene nicht wie in realen Kameras angetroffen hinter das Projektionszentrum zu platzieren, sondern innerhalb des Frustrums. Diese Darstellungsweise vereinfacht oft das Verständnis von projektiven Betrachtungen.



**Abbildung 5.5:** Verwendung der negativen Brennweite für einfachere Darstellbarkeit

Mit einer ersten Translationsmatrix können wir den Hauptpunkt vom Ursprung des Kamerakoordinatensystems um  $(u_0, v_0)$  wegbewegen. Der Ursprung des Koordinatensystems ist dort platziert, wo in einer Lochkamera die Kameraachse die Bildebene durchsticht. Durch ungenaue Platzierung des CCD-Chips oder einer Filmplatte kann es nun geschehen, dass der Bildmittelpunkt nicht genau dem Ursprung des Koordinatensystems entspricht. Da mehrere Artefakte (Vignetting, Verzerrung) radial um den Hauptpunkt angesiedelt sind, ist eine genaue Positionierung desselben aber äusserst wichtig. Durch die Matrix  $H$  können wir diese Korrektur erreichen:

$$H = \begin{bmatrix} 1 & 0 & u_0 & 0 \\ 0 & 1 & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad (5.13)$$

Kameras mit Linsensystemen können auch durch Produktionsfehler der optischen Linsen oder eine schlechte gegenseitige Ausrichtung derselben eine Verschiebung des Hauptpunktes erfahren, sodass die Kameraachse nicht mehr auf den Kamerakoordinatensystem-Ursprung abgebildet wird. Diese Ursache betrifft zwar mangels optischer Linsen nicht die Lochkamera, soll aber hier der Vollständigkeit halber erwähnt sein.

Mit der Skalierungsmatrix  $S$  wird die Einheitenumrechnung vom metrischen Kamerakoordinatensystem ins pixelorientierte Bildkoordinatensystem erzielt. Bedingt durch den strukturellen Aufbau von CCD-Chips (Kapitel 3.2.2) unterscheidet sich das horizontale Auflösungsvermögen oft leicht vom vertikalen. Mit den Skalierungsfaktoren  $k_u$  und  $k_v$  in  $S$  kann dieser Unterschied im Seitenverhältnis korrigiert werden:

$$S = \begin{bmatrix} k_u & 0 & 0 & 0 \\ 0 & k_v & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad (5.14)$$

Die letzte Matrix  $I$  bestimmt die Ausrichtung des Bildkoordinatensystems. Die Skalierung der  $y$ -Achse mit  $-1$  lässt dieselbe nach unten schauen, sodass die Bildkoordinaten wie von Bilddateien gewohnt links oben entspringen und nach rechts und nach unten wachsen.

$$I = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad (5.15)$$

Zusammenfassend resultiert aus den besprochenen Matrizen die folgende intrinsische Transformationsmatrix, welche die innere Orientierung nach dem Lochkameramodell vollständig beschreibt:

$$T_{intr} = \begin{bmatrix} -f \cdot k_u & 0 & u_0 & 0 \\ 0 & f \cdot k_v & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad (5.16)$$

### 5.3.2 Kameramodell nach Tsai / Zhang

Zwei Erweiterungen des Lochkameramodells sind die Modelle nach Tsai [39] und Zhang [40]. Im Gegensatz zum oben besprochenen Modell berücksichtigen diese nun folgenden Kameramodelle einfachste Linsensysteme mit den dadurch entstehenden Problematiken.

#### Tiefenschärfe

Die erste Besonderheit von Linsensystemen gegenüber der Lochkamera ist die Tiefenschärfe, was bedeutet, dass nur ein von der Fokussierung abhängiger Tiefenbereich des Objekt-raums scharf auf die Bildebene abgebildet wird. Objekte, die sich weiter entfernt oder näher bei der Kamera befinden, werden erkennbar unscharf dargestellt. Typischerweise wird die Tiefenschärfe in der Portraitphotographie eingesetzt, um den Hintergrund durch seine Unschärfe vom Vordergrund zu trennen.

Drei intrinsische Faktoren bestimmen die Tiefenschärfe: die Brennweite, die Aperturgrösse respektive Blendenöffnung sowie die Korngrösse des verwendeten Filmmaterials respektive die Grösse eines CCD-Elements. Zusätzlich spielt die Entfernung des fotografierten Objekts von der Kamera eine wichtige Rolle. Die Auswirkungen der einzelnen Faktoren lassen sich in einfachen Regeln zusammenfassen:

- Je kleiner die Brennweite (Weitwinkel), desto grösser die Tiefenschärfe.
- Je kleiner die Apertur, desto grösser die Tiefenschärfe.
- Je grösser die Distanz zum Objekt, desto grösser die Tiefenschärfe.
- Je grösser die Korngrösse, desto grösser der maximal tolerierbare Zerstreungskreis und damit die Tiefenschärfe.

Die Aussage mit dem Zerstreungskreis muss relativ gesehen werden, denn der maximale tolerierbare Zerstreungskreis ist eine subjektive Grösse. Die traditionelle Photographie beschreibt diese Kreisfläche als die auf einer Photographie von Auge identifizierbare Punktgrösse. Beim Kleinbildfilm ist der allgemein akzeptierte Durchmesser eines solchen Kreises  $1/30$ mm. In der aktuellen digitalen Photographie werden Bilder hingegen oft am Bildschirm vergrössert oder es werden Ausschnitte extrahiert, wodurch meist die Auflösungsgrenze erreicht wird. Damit muss die Fläche eines tolerierbaren Zerstreungskreises derjenigen eines CCD-Elementes entsprechen, denn das ist die betrachtete Auflösungsschranke.

Nach [32] ist die optimale Tiefenschärfe  $d$  durch folgende Formel gegeben:

$$d = \frac{\lambda \sqrt{n^2 - A^2}}{A^2} \quad (5.17)$$

wobei  $\lambda$  die Wellenlänge des einfallenden Lichts ist,  $n$  der Brechungsindex der Optik und  $A$  die Aperturgrösse. Die Erweiterung dieser Formel um den maximal tolerierbaren Zerstreungskreis  $c$  nach [33] ergibt die beiden Grenztiefen  $d_n$  und  $d_f$  für die nahe und die entfernte Schärfenbegrenzungsdistanz von der Kamera:

$$\begin{aligned} d_n &= d - \frac{sc(s-f)}{fA + c(s-f)} \\ d_f &= d + \frac{sc(s-f)}{fA - c(s-f)} \end{aligned} \quad (5.18)$$

mit der Brennweite  $f$  und der Objektentfernung  $s$ .

### Radiale Verzerrung

Ebenfalls bedingt durch Linsen entsteht die im Kapitel 4.1.1 beschriebene radiale Verzeichnung. In der Optik wird dieser Effekt durch ein Polynom folgender Form modelliert [23, 9]:

$$\tilde{r} = r + k_1 r^3 + k_2 r^5 + k_3 r^7 + \dots \quad (5.19)$$

Roger Y. Tsai [39] und Zhengyou Zhang [40] haben die Auswirkungen von radialer Verzerrung empirisch untersucht und sind der Meinung, dass es für die realistische Nachbildung eines Kamerasystems und insbesondere eine genügend genaue Kamerakalibrierung reicht, sich auf die ersten Terme des Polynoms 5.19 zu beschränken. Nach Tsai genügt eine einzige Verzeichnungskonstante  $k_1$ , Zhang verwendet zusätzlich die fünfte Potenz des Bildradius und benötigt dafür noch die Konstante  $k_2$ .

### Tangentiale Verzerrung

Ein zweiter auftretender Verzerrungstyp ist die tangentielle Verzeichnung. Sie kann nicht mehr alleine als Funktion des Bildradius modelliert werden, sondern verknüpft zusätzlich die Verzeichnungskonstanten  $p_1$  und  $p_2$  direkt mit den Koordinaten eines betrachteten Bildpunktes:

$$\begin{aligned} \tilde{x} &= x + x(2p_1 xy + p_2(r^2 + 2x^2)) \\ \tilde{y} &= y + y(2p_2 xy + p_1(r^2 + 2y^2)) \end{aligned} \quad (5.20)$$

Sowohl Tsai wie auch Zhang sind der Meinung, die entsprechenden Effekte könnten ausser Betracht gelassen werden.

