Tobias Hanning

High Precision Camera Calibration

VIEWEG+TEUBNER RESEARCH

Tobias Hanning

High Precision Camera Calibration

VIEWEG+TEUBNER RESEARCH

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.d-nb.de.

Habilitation thesis University of Passau, 2009

1st Edition 2011

Editorial Office: Ute Wrasmann | Sabine Schöller

Vieweg+Teubner Verlag is a brand of Springer Fachmedien.
Springer Fachmedien is part of Springer Science+Business Media.
www.viewegteubner.de

Cover design: KünkelLopka Medienentwicklung, Heidelberg
Printing company: STRAUSS GMBH, Mörlenbach
Printed on acid-free paper
Printed in Germany

ISBN 978-3-8348-1413-5

Contents

Abstract

The main purpose of this work is to determine the camera mapping for optical measurement objectives. The standard approach models the camera mapping as a pinhole camera with distortion. We formulate different error functions for the pinhole camera model. Minimizing all error functions introduces a non-linear optimization. Therefore, we present initial values for the intrinsic and extrinsic camera parameters including distortion. In particular, the distortion can be determined by a linear least squares problem. This yields a semi-linear approach to camera calibration.

Stereo camera calibration introduces an additional constraint, which is used as epipolar line constraint in the literature. We extend this constraint to epipolar curves and present some calibration approaches for a stereo camera setup. These include the epipolar curve constraint.

When modelling the camera as a pinhole with distortion, we observe a residual error. We show that this error depends on the depth of the observed object. Thus, we present two approaches to introduce a depth-dependent distortion model: First, we propose a spline correction of the residual error, second, we suggest a two-plane distortion model. Several experimental results support both approaches.

Symbols

$A_{i,j}$	the element of the i-th row and j-th column of a matrix $A \in \mathbb{R}^{n \times m}$
$\operatorname{argmin}_{x \in X} g(x)$	$:= \{x_0 \in X : g(x_0) = \min \{g(x) \mid x \in X\}\}$
$\mathcal{C}_\infty(E, F)$	the set of functions from E to F which are differentiable for all degrees of differentiation
CCS	camera coordinate system
$\operatorname{dist}_d(A, B)$	$:= \inf \{\|a - b\| \mid a \in A, b \in B\}$ for $A, B \subset \mathbb{R}^d$
ICS	image coordinate system
$\mathcal{K}$	the set of all pinhole cameras
$\mathcal{K}_\Delta$	the set of all pinhole cameras with distortion model Δ
$\mathcal{L}^2(E, F)$	the set of all integrable functions $f : E \to F$ with $\int \lvert f\rvert^2 \mathrm{d}\lambda < \infty$
$\mathcal{L}(\mathbb{R}^3)$	$:= \left\{l \subset \mathbb{R}^3 \,\middle\vert\, l \text{ affin in } \mathbb{R}^3, \dim(l) = 1\right\}$ the set of all lines in $\mathbb{R}^3$
$\mathcal{L}(o)$	$:= \left\{l \in \mathcal{L}(\mathbb{R}^3) \,\middle\vert\, o \in l\right\}$ the set of all lines in $\mathbb{R}^3$ containing the point $o \in \mathbb{R}^3$
$l(p, q)$	$:= \{x \in \mathbb{R}^n \mid \exists \lambda \in \mathbb{R} : x = q + \lambda(q - p)\}$ the line defined by two point $p, q \in \mathbb{R}^n$
$\tilde{p} := (p_x, p_y, 1)^t \in \mathbb{R}^3$	for a point $p = (p_x, p_y) \in \mathbb{R}^2$
$\tilde{\bar{p}}^1 := (p_x, p_y, -1)^t \in \mathbb{R}^3$	for $p = (p_x, p_y) \in \mathbb{R}^2$ (see page 146)
O_3	orthogonal group in $\mathbb{R}^{3 \times 3}$
$\mathbb{P}^n$	the (real) projective space $\mathbb{R}^{n+1} \setminus \{O\} / \sim$ with $p \sim q \Leftrightarrow \exists \lambda \in \mathbb{R} \setminus \{0\} : \lambda p = q$
$\hat{p} = \binom{u}{v}$	with $(u, v, 1)^t = \Pi_z(p)$ for a point $p = (x, y, z) \in \mathbb{R}^2 \times \mathbb{R} \setminus \{0\}$

$\mathcal{P}(\mathbb{R}^2, \mathbb{R}^2)$	the vector space of all polynomials from $\mathbb{R}^2$ to $\mathbb{R}^2$
$\mathbf{P}$	a finite set of points in $\mathbb{R}^3$ determining a calibration pattern ("model")
$\text{Proj}_L p$	the orthogonal projection of $p \in \mathbb{R}^3$ on the line L
$Q_t := \begin{pmatrix} 0 & -t_3 & t_2 \\ t_3 & 0 & -t_1 \\ -t_2 & t_1 & 0 \end{pmatrix}$	matrix with $Q_t x = t \times x$ for all $x \in \mathbb{R}^3$ with $t = (t_1, t_2, t_3)^t \in \mathbb{R}^3$
$\mathbb{R}_+ := [0, \infty[$	non-negative real numbers
$\mathbb{R}_+^* :=]0, \infty[$	positive real numbers
RCS	reference coordinate system
$\begin{pmatrix} R & t \end{pmatrix} \in R^{3\times 4}$	matrix where the first three columns equal the columns of $R \in \mathbb{R}^{3\times 3}$ and the last column equals $t \in \mathbb{R}^3$
SO_3	special orthogonal group in $\mathbb{R}^3$ (rotation group)
$\mathcal{T}(\mathbb{R}^3, \mathbb{R}^3)$	isometric coordinate transformations in $\mathbb{R}^3$
$u \times v$	the cross product of two vectors $u, v \in \mathbb{R}^3$ for $u = (u_x, u_y, u_v)^t = (v_x, v_y, v_z)^t$ it is $u \times v = \begin{pmatrix} u_y v_z - u_z v_y \\ -(u_x v_z - u_z v_x) \\ u_x v_y - u_y v_x \end{pmatrix}$
$\text{span}(v_1, \ldots, v_n)$	$\{\lambda_1 v_1 + \cdots + \lambda_n v_n \mid \lambda_1, \ldots, \lambda_n \in \mathbb{R}\}$ the linear span of $v_1, \ldots, v_n$
w.r.t.	with respect to
$\{z = 0\}$	$:= \left\{(x, y, z) \in \mathbb{R}^3 \,\middle|\, z = 0\right\}$
$\{z = 1\}$	$:= \left\{(x, y, z) \in \mathbb{R}^3 \,\middle|\, z = 1\right\}$

Chapter 1

Introduction

1.1 Motivation

As indicated by its name most tasks in computer vision deal with an imaging device. If this imaging device is a camera, it performs a mapping from a 3D world to a 2D image. Determining the parameters of this mapping is called camera calibration. This problem also includes the modeling and parametrization of the observed imaging process.

Long before the computer vision community addressed the modeling and to determination the camera mapping, it was investigated by photogrammetry. Photogrammetry began nearly in parallel to the rise of the photography in the middle of the 19th century in France and Prussia. The name "photogrammetry" was established by Albrecht Meydenbauer (*1832, †1921), who published a procedure to measure buildings by photographies. He also founded the Königliche Preußische Messbild-Anstalt, the first administration for photogrammetric research. The main objective of photogrammetry is accuracy. Therefore, long and tedious calibration routines are taken into account to obtain all parameters of an observed camera mapping. Classic photogrammetry works on photographs. Nowadays, digital image sensors replace the classic cameras and introduce some other problems like the modeling of the digitization, which are yet not fully covered by the photogrammetric community.

On the other hand, the digitization of continuous signals is a well known object of research in computer vision. Many applications of computer vision do not need a high accuracy in reconstruction as photogrammetry provides. In robotics or driving assistance it is sometimes necessary to obtain information from an un-

T. Hanning, *High Precision Camera Calibration,* DOI 10.1007/978-3-8348-9830-2_1,

known environment. It is often more important to estimate a self-orientation of the sensor by coarse data than to measure exact distances (*e.g*. to obtain the egomotion of an autonomous system). Moreover, self-calibration defines an important task in computer vision. Self-calibration of a camera means to gather as much information from images of a more or less unknown nature as one can. If one assumes that a camera behaves like a pinhole camera, one can formulate the camera mapping as projective mapping. This opens the whole world of projective geometry for camera calibration and allows a boost in results since the 1990ies (summarized *e.g*. in [HZ00]).

To adjust the pinhole camera model for quality assurance in dimensional accuracy the model is augmented with a distortion mapping which is defined in the image plane. However, like every model, the model of a pinhole camera with distortion is only an approximation to the real behavior of a camera. In particular, most monochromatic aberrations of a lens system (*i.e*. the first four of the five Seidel aberrations: spherical aberration, coma, astigmatism and curvature of field) can not be modeled in a plane (see [Hec87]). Modelling the distortion as an in-plane mapping assumes that the observed distortion of a point does not depend on its distance to the lens plane. In photogrammetry it is well known that for the pinhole camera model the observed distortion varies not only with focusing the lens but also within the depth of field for a fixed focus (see *e.g*. [Atk96]).

In this work we address camera calibration particularly with regard to metric reconstruction. Since projective geometry is non-metrical, we only apply some results of the projective geometry to obtain initial solutions for problems w.r.t. a metric. We share the emphasis on the metric reconstruction with photogrammetry, therefore, we also use some results of the photogrammetric community. In particular, we use some results to motivate a dependence on depth in reconstruction.

1.2 Outline

In the following chapter we derive the camera mapping from geometric optics based on Snell's law. If we simplify Snell's law to first order optics we obtain the classic formulation of the camera mapping by the pinhole equation. The first order optics are only a rough approximation of the observed camera mapping. Any observed deviation to the pinhole camera model must be modeled additionally. The standard way to deal with aberrations is to introduce a distortion mapping in the image plane.

Camera calibration means to determine the camera model parameters which fit

best to the observed behavior of the actual camera. Therefore, we have to measure the distance of an observation to a given camera model. In the third chapter we introduce four approaches to define such a distance. The determination of the optimal camera mapping w.r.t. each of these distance functions defines a non-linear optimization problem.

Since the result of every non-linear optimization algorithm depends on the initial value, we address the problem to obtain such an initial value in the fourth chapter. We present some additional constraints for the starting solution according to Zhang ([Zha98]). These additional constraints allow a valid solution for the initial value problem.

The calibration of a stereo camera system is more than calibrating two camera separately: An additional constraint for the stereo camera system can be applied, since a calibration target is observed by both cameras. In the fifth chapter we extend the classic constraint (the so-called epipolar constraint) and present some results for this extension.

The limitations of the pinhole camera model with distortion become visible when it has to deal with blurring. The pinhole camera model depends on the lens maker's equation, which states that there is a determined object plane where the observation of an object appears sharp in the image. All objects outside this object plane appear blurred in the image. In the sixth chapter we analyze the blur induced by first order optics. As a main result will show that even in first order optics the blur depends not only on the depth of an object but also on its position in a fixed depth. Furthermore, the blur is not rotationally symmetric. Therefore, every point extraction method which assumes a rotationally invariant blur which is identical for each pixel, must be erroneous. Also, as experiments show, this error depends on the depth of the observed points. Thus, for a camera calibration which should provide a high precision, we need a component of the camera mapping which depends on the depth of the observed object. We present two approaches for a depth-dependent camera model in the last section. Several experimental results support the proposed non-standard camera models.

1.3 Contribution

This work contributes some new results in the following areas of camera calibration:

- Initial values for non-linear optimization

  Camera calibration is a non-linear problem. However, every non-linear optimization algorithm needs adequate initial values for an optimal performance. Thus, we present several methods to obtain initial values. In particular, we revisit Zhang's method to obtain initial values for the pinhole camera parameters and propose several ways to apply additional constraints which improve the result. Furthermore, we show that the determination of the distortion is a linear least squares problem provided that all other camera parameters are known.

- Semi-linearity in camera calibration

  Since the determination of the distortion forms a linear least squares problem, a part of the camera calibration error function can be minimized by linear methods in closed form. We call such a non-linear problem, which includes a linear part, a semi-linear problem. In the case of camera calibration we can decouple the calculation of the distortion parameters from the non-linear optimization. Thus, the number of parameters in the non-linear minimization will be reduced.

- Stereo camera calibration

  A stereo camera setup introduces additional constraints for the calibration. The well known epipolar constraint is extended to a generalized epipolar constraint. We present several approaches to calibrate a stereo camera w.r.t. this constraint instead of calibrating two single cameras.

- Non-standard camera models

  For the pinhole camera with distortion we observe a residual distortion which depends on the depth of the observed object. The standard camera misses a depth-dependent distortion model. Thus, we present two approaches to include the depth into the distortion model: First a correction by splines which depends on the depth of the observed object, second a novel two-plane distortion model which dissolves the pinhole assumption.

Chapter 2

Modelling the camera mapping

2.1 Geometric optics for computer vision

For our considerations we analyze the optical system by geometric optics. Since the effects are negligible for our purposes, we do not apply wave optical phenomena like diffraction of light. Furthermore, we assume that the lens is rotationally symmetric about a straight line which is called *the optical axis of the lens*.

2.1.1 The "thin lens" assumption and first order optics

Geometric optics are based on the refraction law (also known as Snell's law). Given two media with refraction indices n_1 and n_2 and a light ray, which passes from media one to media two, the refraction law states that the angles θ_1, θ_2 of the light beam to the normal of the interface of the medias obeys

$$n_1 \sin(\theta_1) = n_2 \sin(\theta_2) \tag{2.1}$$

(see Figure 2.1).

In computer vision it is widely accepted that the refraction index of air is so close to 1 that it can be treated as 1 [1]. A (spherical) lens has two media interfaces (air to lens, lens to air) which are described by two spheres with the same radius r (see Figure 2.2). Following Snell's law one can reconstruct the refraction of each light ray emitted from an object through the lens.

[1] In fact the refraction index of air is 1.0002926, whereas vacuum has index 1 for light with a wavelength of 589.3nm

T. Hanning, *High Precision Camera Calibration,* DOI 10.1007/978-3-8348-9830-2_2,

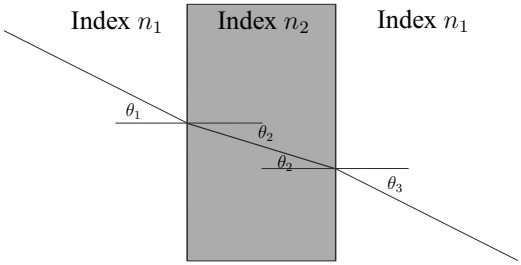

Figure 2.1: Illustration of Snell's law for parallel surfaces.

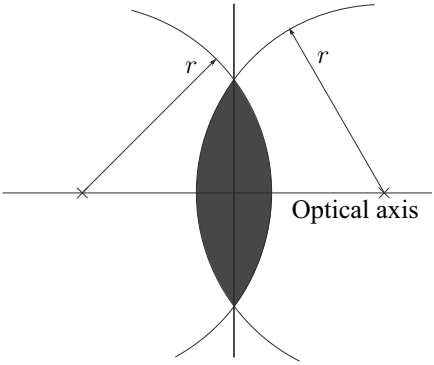

Figure 2.2: Sketch of two spheres with radius r defining a spherical lens (blue).

In first order optics it is assumed that all rays, which are considered in the camera model, are nearly parallel to the optical axis. Such rays are also called *paraxial*. Thus, we assume that there are only small angles to the normals of the lens surfaces and therefore the sine can be approximated by $\sin(x) \approx x$. Since this is the first term of the Taylor approximation of the sine at the origin, the derivations following this assumption are also called *first order optics*.

The second assumption in computer vision is called the *thin lens assumption*: the lens is assumed to be infinitesimally thin. From the refraction point of view the thin lens behaves like a spherical lens. Yet, the distance which a light ray covers inside the lens is infinitesimally small. Thus, from a localization point of view the lens is a plane. This plane is called *principal plane*. A light ray passing the thin lens is refracted at the surface air/lens and immediately after that at the surface lens/air. An immediate consequence of the thin lens assumption is that any light ray passing the thin lens at the optical axis will not be refracted, since both surfaces of the thin lens are parallel in this point (see Figure 2.1). Another consequence of the thin lens model combined with the paraxial optics simplification is that all (paraxial) light rays emitted at a point p at one side of the lens, which pass the lens, meet in one point i_p at the other side of the lens such that

$$\frac{1}{d_p} + \frac{1}{d_i} = \frac{1}{f} \tag{2.2}$$

holds, with $f = \frac{r}{2(n_l - 1)}$ and d_p resp. d_i being the distance of the object p resp. the image point i_p to the principal plane and n_l the refraction index of the lens (see [Hec87] for more details). This equation is often called *lens maker's equation*. This means that there is a relation between the object and the image and that this relation depends only on the distance of the object to the principal plane, but not on its distance to the optical axis. There is a determined distance behind the principal plane, denoted as d_i in Figure 2.3, at which an observation of a point from the object side will become a sharp image. On the other hand, any object plane determines a so-called *focal plane* behind the lens where the points in the object plane appear sharp. In Figure 2.3 three principal ways for (paraxial) rays from an object to its image passing through a thin lens model are displayed. Namely, these are

i. A ray that comes in parallel to the optical axis on one side proceeds towards a certain particular point F at a distance f to the principal plane on the other side. F is called *focal point*. The distance f to the principal plane is called *back focal length*.

ii. A ray that passes through the center of the lens will not change its direction. This ray is called *center ray*. The center of the lens is called *optical center*.

iii. A ray that arrives at the lens from the focus F (also at distance f to the principal plane) on one side goes out parallel to the axis on the other side. The distance of F' to the principal plane is called *front focal length*.

In this work we assume a spherical lens there the back focal length equals the front focal length. For the first order optics with the thin lens assumption, the focal length is independent from the distance of the ray to the optical axis.

2.1.2 The circle of confusion

In general, the image acquisition is performed by a planar imaging device (see section 2.1.3.1). In the following we call the plane where the image is acquired *image plane*. It is obvious that in general the image plane does not coincide with the focal plane. Therefore, not all light rays emitted from an observed object meet in a point in the image plane. Let us now consider a point light source in the object plane. All light rays, which are emitted from this light source and pass the lens, form a cone on the image side of the lens (in first order optics). The intersection of this cone with the image plane is called *circle of confusion*.

The three principal ways of the light ray through the lens can be used to determine the circle of confusion (see Figure 2.3). If the circle of confusion is smaller than the size of one element of the imaging device objects in the computer image appear sharp. The area where this is true is called *depth of field*.

The dependency of sharpness and depth can be used to estimate the depth of an object, *i.e.* the distance of the object to the principal plane, (known as "depth from focus", (see *e.g.* [AFM98], [SG87] or [Gro87], or "depth from de-focus", see [CR99] or [Asl03]).

The blurring effect of the first order optics should also be handled by the camera mapping. The common way to handle this is to convolute the "ideal image" with a kernel (see [HS92] or [FP02]).

The "ideal image" g_{ideal} is the image of the objects obtained by the center ray only. Of course, this image can not be observed anywhere. The input function g for the imaging device, which is sometimes called *sensor input function*, becomes

$$g = g_{\text{ideal}} * k \tag{2.3}$$

for a mollifier $k \in \mathcal{L}^2(\mathbb{R}^2, \mathbb{R}) \cap \mathcal{C}_\infty(\mathbb{R}^2, \mathbb{R})$. The center ray function g_{ideal} may be not continuous, but piecewise continuous. For mathematical reasons we assume

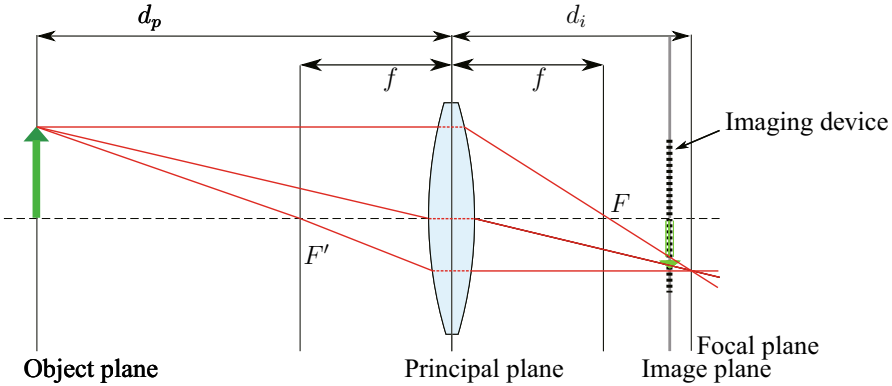

Figure 2.3: Sketch of the three main rays of the thin lens model (for the thin lens assumption the dashed red lines are assumed to have zero length).

that g_{ideal} is twice integrable. Note that the input function for the sensor array g is continuous as a convolution of a $\mathcal{L}^2$-function with a smooth function.

In computer vision the mollifier k is often called *point spread function* (PSF). The PSF can also been seen as the impulse response of the optical system (see *e.g.* [Goo96]). There are numerous ways to estimate the PSF assuming a symmetrical and identical kernel (at least for pixels in a local area) (see *e.g.* [BW93]) or almost without these assumptions (see [Pis06], see also section 6.2.3.1).

2.1.3 Image acquisition

2.1.3.1 The sensor array

Typical cameras use image sensors as imaging device in the image plane. An *image sensor* is a device that samples the sensor input function and converts it to an electric signal. The image sensor itself is a rectangular grid of photo-sensitive sensors. It can be an array of charge-coupled devices (CCD) or complementary metal–oxide–semiconductors (CMOS) sensors. The arrangement of the sensor array introduces a canonic coordinate system for the image.

Each sensor in the sensor array determines the value of the corresponding picture element (short: pixel). Therefore, a pixel represents in fact a rectangular area. Let $d_{p_u} \times d_{p_v}$ be the size of the photosensitive area of one sensor and d_{i_u} resp.

d_{i_v} be the distance between two sensors in the horizontal resp. vertical direction (see Figure 2.4). Then the distance between two pixel centers is $d_u := d_{i_u} + d_{s_u}$ in horizontal and $d_v := d_{i_v} + d_{s_v}$ in vertical direction.

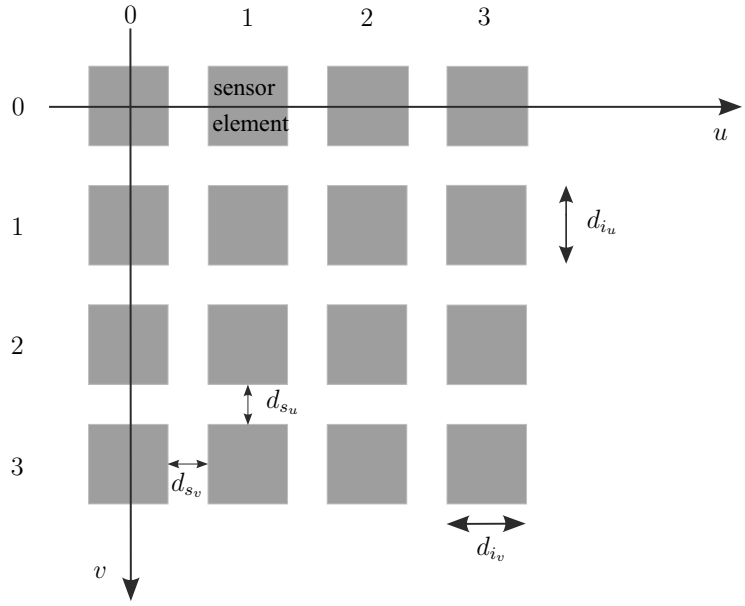

Figure 2.4: Sketch of the elements of an array of photosensitive sensors.

The value of a pixel is obtained by integrating the light intensity function over the area of the corresponding sensor of the sensor array: Let $g \in \mathcal{L}^2(\mathbb{R}^2 \times \mathbb{R}, \mathbb{R})$ be the light intensity function at the sensor array. Then the value $g_{i,j}$ of the sensor (i, j), which is also the value of the pixel (i, j), is

$$g_{i,j} = Q\left(\int_{\Delta t}\int_{A_{i,j}} g(x,t)s_{i,j}(x,t)\,\mathrm{d}x\,\mathrm{d}t\right) \tag{2.4}$$

Whereas

- $Q : \mathbb{R} \to G$ is a quantization function (which also corrects the characteristic line of the sensor element) for a finite set of gray-values G, and

- Δt is the exposure time of the whole optical device, and
- $A_{i,j}$ is called the *aperture* of the sensor (i,j). If $p_{i,j} = [ip_u - \frac{d_{p_u}}{2}, ip_u + \frac{d_{p_u}}{2}] \times [jp_v - \frac{d_{p_v}}{2}, jp_v + \frac{d_{p_v}}{2}]$ is the area of the element (i,j) w.r.t. the coordinate system of the sensor array, then obviously $p_{i,j} \subset A_{i,j}$ should hold. However, since the sensor elements in a typical CCD sensor neighbored sensor elements may interact with each other, we also may choose an aperture $A_{i,j}$ which is superset of $p_{i,j}$. Finally
- $s_{i,j} \in \mathcal{L}^2(\mathbb{R}^2 \times \mathbb{R}, \mathbb{R})$ is the sensor specific density function. It is a characteristic of the sensor element at the position (i,j) in the sensor array. The characteristic has a time component to model each influence on the sensor element which varies with time. This can be any physical influence of the whole device (like dark current, effects of the A/D converting, etc.).

2.1.3.2 A simplified sensor model

Since it is almost impossible to determine the time varying component of $s_{i,j}$, it is often replaced by a random variable ν which implements an additive noise (see [Pis06] or [För00]). The additive noise contains electronic noise, being independent on the intensity. Some authors (see [DS74]) claim that the noise characteristics is dominated by the Poisson distribution of the photon flux.

The most commonly accepted simplifications to the sensor model are that the light intensity function g does not change during Δt, and that the sensor characteristic $s_{i,j}$ is the same for every sensor of the array. Moreover, it is common that $s_{i,j}$ is constant on $p_{i,j}$ and zero on $A_{i,j} \setminus p_{i,j}$. With all these simplifications it is no longer necessary to model the sensor characteristic as a density function. It can be transformed to the discretization function:

$$g_{i,j} = \text{round}(r(\int_{p_{i,j}} g(x)\ \mathrm{d}x + \nu(i,j))) \tag{2.5}$$

The bijective function $r : \mathbb{R} \to \mathbb{R}$ is often called radiance function and is generally not linear. It represents a characteristic of the whole sensor array. Typical radiance functions are logarithmic, piecewise linear or have a linear and a logarithmic part. The most significant non-linearity on the response curve is at its saturation point, where any pixel with a radiance above a certain level is mapped to the same maximum image gray-value. However, for practical applications we