Fourier Formulation of Illumination Optics
and
Computer-Automated Reflector Design

by
Robert Demetrius Stock

advised by
M. W. Siegel

Pittsburgh, Pennsylvania
August, 1995

A PHD THESIS

Keywords: near zone, illumination modelling, luminaire characterization, lighting invariants, extended light sources, analytical representation, ray tracing, reflector modelling, computer-aided design

Submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
at Carnegie Mellon University

ACKNOWLEDGMENTS

I thank my advisor, Melvin W. Siegel, for providing me not only with an interesting research topic, but also the resources and guidance I needed to bring it to a satisfactory conclusion in a timely manner. Mel has helped make the last three years both pleasant and rewarding for me. I have learned many things, including the proper use of the word "determine" and the difference between latitude and longitude. Many interesting conversations with Victor S. Grinberg also helped to sharpen our thinking about this topic. I would also like to thank Alan Guisewite for his invaluable assistance with CMU’s computing environment. His patience regarding my constant questions and problems has been commendable.

I am grateful to my committee, consisting of Physics Department faculty members Steve Garoff, Brad Keister, and Jim Russ, and Computer Science faculty member Paul Heckbert, for their comments and support.

I am also indebted to the Carnegie Mellon University Physics Department for providing me with a teaching assistantship and summer job for five years, without which this thesis would not be possible. The early stages of this work, before I became involved, were supported by General Motors Inland Fisher Guide Division, Computer Graphics Group, Anderson IN. Experimentation with the applicability of supercomputing to this topic was supported by grant #ASC930003P from the Pittsburgh Supercomputing Center.

Finally, I thank my parents, Robert Bryant Stock and Alice Ann Demetrius Stock, whose constant and selfless dedication to my educational development has provided me with the background and encouragement I needed to reach this point. I could not have done it without you.

ABSTRACT

This thesis solves three major problems related to illumination optics: far-zone, orientation invariant light source characterization for quality control and analysis; near-zone light source characterization for accurate image rendering and raytracing from a general extended source; and computer-automated reflector design, using the near-zone model. The main approach is to combine measurements with Fourier analysis, using judiciously chosen coordinate systems and orthogonal fitting functions.

This approach has several advantages over standard raytracing models: it provides for natural data compression and interpolation, it solves the problem of determining the radiance distribution of a real source by using actual pinhole CCD camera measurements, and it eliminates the computationally intensive ray-filament intersection problem by transforming the source into an equivalent nonuniform spherical radiator. A method for treating the occlusion of rays by the extended filament, with only spherical intersection calculations, is also discussed.

The problem of designing a reflector to cast a desired intensity distribution on a distant screen for a given source is attacked by combining the realistic image rendering properties of the general light source model with a smooth-surface specular reflector model. The problem then becomes a straighforward numerical optimization of perturbations to the base parabolic reflector shape. The minimization method chosen is conjugate variables, which is essentially a multidimensional binary search that is modified to adaptively choose efficient search axes in parameter space. An algorithm to provide a first guess for the perturbations based on the shape of the image perimeters is also described.

Fundamentally, the goal of this work is to develop a general light source model that is physically accurate, intuitively descriptive, computationally convenient, and applicable to real sources.

1. INTRODUCTION

1.1 History and Motivation

Photometry, founded by Lambert in 1760 [LAM 60], is a branch of geometric optics concerned with the calculation and measurement of light intensity. Photometry's first real stimulus emerged roughly sixty years ago with the advent of widespread electrical lighting. Complaining that "theoretical photometry constitutes a case of 'arrested development', and has remained basically unchanged since 1760 while the rest of physics has swept triumphantly ahead" [GER 39] and that theoretical "methods are fragmentary" [MOO 43], illumination engineers developed a theory for treating the light field as a vector quantity [GER 39, MOO 38]. After an initial period of excitement, however, interest in the light field waned because, in general, no scalar potential existed, and no new results were found. The fundamental problem is that the radiance (power per projected area per unit solid angle) produced by a light source depends not only on the position in space of the observation point, *but also on the direction of view*. It is therefore not a vector quantity, but a scalar function of the five variables[1] (x, y, z, θ, ϕ).

After languishing in obscurity for three more decades, photometry received its second stimulus from the availability of computer resources. Furthermore, the need to perform illumination calculations no longer resided solely in the domain of illumination engineering, but also in the realm of computer graphics and computer vision. The two primary image synthesis methods used today are called "radiosity" and "raytracing."

The radiosity method, used to calculate the feedback effects of interreflecting diffusing surfaces, was adapted by computer scientists from the heat transfer literature of the 1950s [SPA 63, HEC 91]. The solution to the problem was formulated in the optical literature as early as 1926 [MOO 40, YAM 26]. Before the advent of computers, however, only certain special cases could be calculated or even approximated.

Raytracing [WHI 80, GLA 89, HAL 89, ROG 91], the method of light transport favored for

[1] with one constraint discussed in section 1.3.1

handling specular surfaces, suffers several serious shortcomings when applied to real-world problems. One such problem concerns the difficulty of ray and surface intersection calculations in ray tracing for design, rendering, etc; another is the difficulty of characterizing any real luminaire that one might want to include as an object in a ray traced scene. Previous models were therefore forced to make simplifying assumptions regarding light source shape and composition, often employed empirical or idealized relations, and provided no standard procedure for characterizing the light distribution of a given real source. Thus a physically correct and computationally convenient treatment of a real light source in the near-zone,[2] where the extended nature of the light source matters, represented a nearly intractable problem for standard methods.

Over the last few years the notion that the radiance distribution of an extended source could be projected onto a sphere surrounding the body [LEV 71, VER 84, ASH 93] (rendering the intersection problem trivial), and stored as a set of coefficients from an expansion in orthogonal polynomials [NGA 87, NGA 92] (solving the storage problem), began to appear. Levin's presentation of this idea was in the form of brief observation; the others that reported actually working on the idea tended to use inconvenient interpolation or data compression schemes. None of this early work addressed the problem of occlusion by the source of secondary reflected rays or of rays from other sources. Although the need for accurate near-zone lighting calculations has been demonstrated [MIS 90], the field of near-zone source characterization is in its infancy.

Accurate near-zone rendering with real sources could be applied not only to computer graphics, but also to the tedious problem of luminaire reflector design. As a concrete example, automotive headlamps must have illumination patterns that satisfy both aesthetic and regulatory (SAE and federal) requirements. Styling, space, and cost considerations add further constraints. Customer preference is also leading designers away from reliance on lenses; more and more the job must be done with the reflector alone. The design of satisfactory reflectors, however, is an

[2]As a rule of thumb, the near-zone is defined by illumination engineers as the volume within five characteristic lengths of the source; see section 1.3.3.

expensive and time consuming process, consisting of iteratively raytracing, prototyping (using libraries of standard lens and/or reflector components in various combinations), and measuring; i.e. trial and error. This technique leads to the astonishing fact that headlights require the longest development time of any automobile component: typically four years!

Currently, computers have been of little help in this field, because most models for real light sources lack prescriptions for physically characterizing the true radiance distribution of the source, and rely on empirical or *ad hoc* relations. In addition, expensive filament and/or reflector intersection problems often lead to inefficient ray tracing. Near-zone source characteristics are usually ignored. For instance, a technique developed by researchers at Mitsubishi Motors, for example, treats the source as a set of radiating pieces, and finds a 2-dimensional reflector by "growing" reflector segments with an orientation that produces the best match to a desired illumination pattern [MYO 82]. This is essentially the process currently carried out by human designers [ELM 74].

Our goal then is to save time and money by developing a combined near-zone light source and reflector model which can provide accurate computer-simulated rendering[3] to minimize the need for actual prototypes, giving the designer a valuable new tool. It may also be part of a computer-driven optimization system,[4] the result of which could be a starting point for fine adjustment by a human expert, perhaps reducing days or weeks of analysis and design to a few hours.

1.2 Overview of Approach

Real light sources, i.e. extended bodies with nonuniform radiance distributions, therefore, still lack a physics-based near-zone model that is descriptive, practical, and general. We develop such a model by providing a prescription for combining actual measurements with expansions in orthogonal functions. This model is then applied to the problem of designing a pseudo-parabolic reflector for a real light source to create a desired illumination pattern on a screen.

Although standard illumination design practice evaluates a design by using a discrete number

[3] computer-*aided* design

[4] computer-*automated* design

of "test points" at which the lighting level must be in some range, we choose to attempt to solve the more highly constrained problem of trying to match an entire image for two reasons. First, designers worry about the aesthetics of illumination between test points as well, and this more strenuous problem includes the "test point" method as a subset. Second, we do not want this research to become too heavily associated with one specific industrial application of reflector design (i.e. headlamps), but to remain more general.

Operationally the model depends on two types of measurements taken at points on a sphere surrounding the light source. The first type of measurement simply uses a photocell to obtain the intensity integrated over viewing angles. The second type uses a pinhole camera with a charge-coupled device (CCD) on the image plane to obtain a series of 2-dimensional images for the extraction of the directional radiance information necessary for near-zone ray tracing. The measurements in both cases are encoded as a set of Fourier coefficients of appropriate orthogonal functions.

The model is intuitively descriptive because the photocell measurements are fit to spherical harmonics, giving the coefficients of the fit a well-understood physical meaning in terms of a multipole expansion. Furthermore, I find invariant quantities among the coefficients that capture the essence of the angular distribution of light intensity, independent of rotation artifacts. Finally, the coefficients can be combined to yield a measure of the non-uniformity or standard deviation of the illumination distribution. These invariant quantities can be used for far-zone source characterization and quality control, for example.

The model is practical for several reasons. One is the ease with which raytracing calculations can be performed, because the light source is transformed by our procedure into a spherical source of secondary wavelets and because the fitting process allows interpolation between measurement points; the intersection problem, which can be expensive in standard ray tracing, becomes trivial. For example, the standard alternative would be to model the filament as a polygonal mesh, which requires measuring the filament’s exact location in space. Another reason is in terms of computer memory requirements: intensity measurements (at the density with which we record the data) can be encoded by coefficients at a ratio of roughly a thousand to one, unlike typical ray tracing models which would demand a large number of points or polygons to be used.

The model is general because it is based on measurement. It is therefore applicable to any type of light source, unlike standard physics-based or computer graphics-based models which are forced to make simplifying assumptions regarding light source shape and composition, often employ *ad hoc* or empirical relations, and provide no standard procedure for characterizing the radiance distribution of a given real source. Our model provides a common language for conveniently describing, quantifying, and comparing the illumination patterns and radiance distributions of real light sources, regardless of their nature.

Because of prevailing design and manufacturing techniques, most reflector models consist of a set of facets, which means a very large number of parameters is required to specify the surface, and the extra problem of enforcing smoothness arises. Because there are too many parameters in such a model for standard optimization algorithms, Ashdown outlines how genetic algorithms may be used for computer-automated reflector design [ASH 94]. The difficulties of standard reflector models are solved in our approach by eliminating the use of facets in favor of a polynomial representation, reducing the number of parameters to a dozen or so and insuring smoothness. Though local control of the surface is lost, the simplicity gained allows the use of standard numerical optimization techniques. Furthermore, the reflector surface coefficients have understandable physical meanings, such as astigmatism and focal length. Thus, adjustments can be made by a human designer with a firmer intellectual basis than the simple trial and error practice of manipulating a vast array of facets.

In summary, we provide a compact method for real source characterization in both the far- and near-zones, and combine the computationally practical near-zone model with a smooth reflector representation to facilitate both computer-aided and computer-automated reflector design. The source model can also be applied to accurate rendering in computer graphics.

1.3 Basic Optics

Radiometry and photometry are both concerned with the measurement of light intensities, usually in architectural or other human vision design contexts. Radiometric quantities are expressed in terms of the energy received by light-sensitive sensors. Photometric quantities are essentially radiometric measurements that are weighted by the frequency response function of the human eye, and thus contain a physiological, perception-based component. For this thesis, I will rely solely on radiometric units.

Physical optics treats light as an oscillating electromagnetic wave [BOR 70]. If the light sources are radiating incoherently (so that interference and polarization effects can be ignored), and the wavelengths of the emitted light are small compared to any apertures (so that diffraction effects can be ignored), then physical optics reduces to geometric optics. Because the environments of typical macroscopic illumination calculations satisfy these requirements, geometric optics is the right model to use in the fields of radiometry and photometry, and in fact geometric optics is almost always used in computer graphics [SHI 91].

In geometric optics, light rays are independent and their only property is their radiance. Radiance (also called brightness, luminosity or, rarely, helios [MOO 81]) is a powerful and useful quantity for radiometric calculations. Its units are *watts meter* $^{-2}$ *steradian* $^{-1}$. In photometric terminology it is known as luminance, and it is expressed in units of *candelas meter* $^{-2}$ (*nits*), or *foot-lamberts* [HALS 93]. The radiance emitted by a surface may (and usually does) vary with direction, but *does not vary with distance from the emitting surface*[5] (for a constant index of refraction). Thus the radiance distribution for all directions over a closed surface (say, a sphere) surrounding a light source is a full description of the illumination properties of that source.

The theses of Heckbert [HEC 91] and Shirley [SHI 91] contain good descriptions of radiometry and geometric optics from a computer science point of view. For a rigorous physics-based approach, see of course Born and Wolf [BOR 70]. Tables of terms and units can be found

[5]The r^{-2} dependence of the emissive power is balanced by the r^{-2} dependence of the steradians; see section 1.3.1 for a detailed explanation.

in Heckbert [HEC 91] and Nicodemus [NIC 63]. A new set of terms is suggested by Moon [MOO 81].

1.3.1 Constancy of Radiance

The constancy of radiance along lines can be seen by a simple conservation of energy proof [NIC 63]. The radiance seen by a surface A_p due to light emitted from a surface A_q is the power received divided by the *etendue* [WEL 78] (also called "throughput" [NIC 63]). The *etendue* is a geometrical factor equal to the projected area of the emitting surface times the solid angle subtended by the receiving surface. Because this factor is symmetrical between the surfaces, and because all of the power emitted into a solid angle subtended by A_p must be intercepted by A_p (barring losses), the radiance seen at A_p coming from A_q must be the same as the radiance emitted from A_q in the direction of A_p. Since the distance between the surfaces is arbitrary, the radiance must be a constant along lines [NIC 63]. This proof depends on the assumption that the surfaces are small enough compared with the distances between them that the radiance in a given direction is constant over the surface.

Other forms of the proof exist as well. The above argument can be cast in terms of two surfaces of equal temperature in radiation equilibrium; a difference between the radiance received and the radiance emitted would result in a violation of the Second Law of Thermodynamics [RIE 82]. A simple proof that depends on the validity of the sine condition of geometric optics is given by Born and Wolf [BOR 70]. Another proof uses Liouville’s theorem, which states that the phase space volume of a system remains constant: since the volume in phase space of a beam of photons is proportional to the radiance [JOY 74], the radiance cannot change. According to Landau and Lifshitz, Liouville’s theorem holds exactly for photons, not just statistically, because they are noninteracting [LAN 69].

1.3.2 Illumination Calculations

As shown in figure 1-1, the illumination of surface A_p can be determined from the radiance B of A_q and geometrical factors. Assuming $r^2 \gg A_p$ and A_q, the power P striking A_p is

$$P = B A_q \cos(\beta) A_p \cos(\alpha) / r^2$$

where α and β are the angles from each surface’s normal to the line joining the centers of the

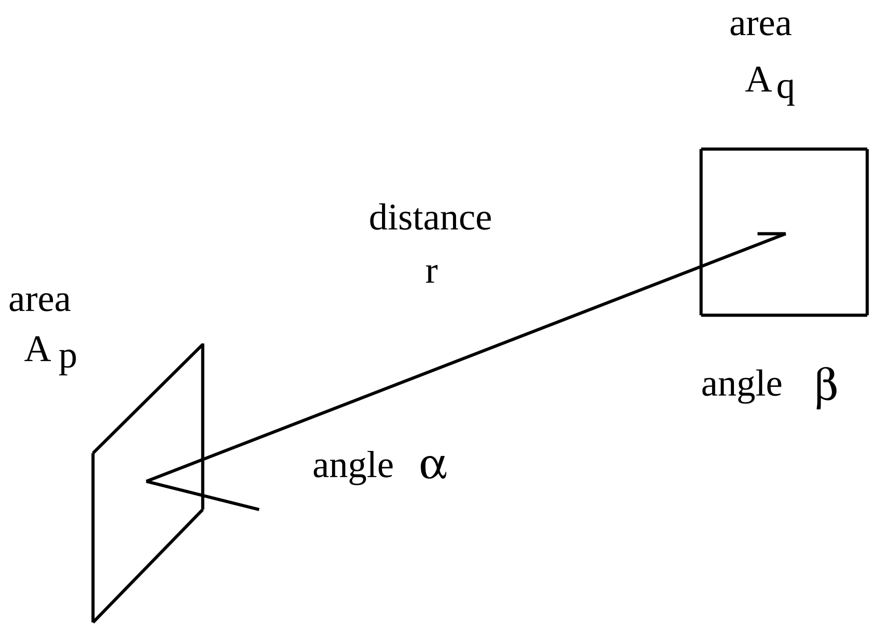

Figure 1-1: illumination of a surface

surfaces.

Good references for radiometric calculations are Walsh [WAL 65], Moon [MOO 38], and Boyd [BOY 83].

1.3.3 The Near-Zone

In the near-zone, the extended nature of light sources becomes important, and they cannot be treated as point sources with radial (though perhaps nonuniform) radiance distributions. As a rule of thumb, illumination engineers assume the near-zone ends at a distance of five times the largest dimension of the source [IES 85]. I believe this rule can be understood as follows. Figure 1-2 shows a detector receiving light from a long, narrow source of constant radiance B.

The source will appear point-like as $r \to \infty$ to the degree that the light energy received by the detector from a small area dA near the end of the source is the same as the light energy received by the detector from the center of the source. Using the methods from section 1.3.2, the power

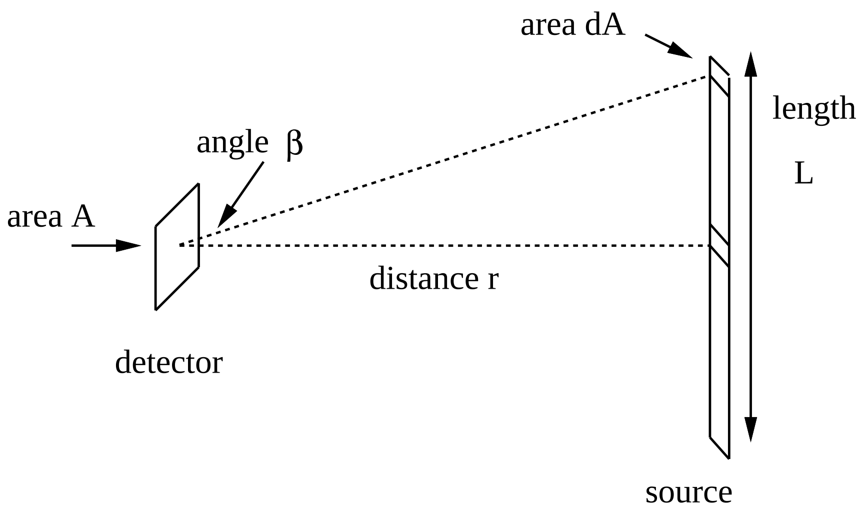

Figure 1-2: light from an extended source

from the center of the source is

$$P_c = BA\,dA/r^2$$

while the power striking the detector from the end is

$$P_e = BA\cos\beta\,dA\cos\beta/(r/\cos\beta)^2 = P_c\cos^4\beta$$

In terms of the length L of the source, however, $\cos^4\beta$ can be written as

$$\cos^4\beta = r^4/(r^2 + L^2/4)^2 = f(r)$$

so that

$$P_e = P_c f(r)$$

The source will therefore appear point-like to the degree that the function $f(r)$ approaches unity. A graph of $f(r)$ with r measured in units of L is shown in figure 1-3. When $r = 5L$, $P_e = 0.98P_c$ and the curve is becoming level, indicating that the extended nature of the source is becoming less important and the near-zone is giving way to the far-zone.

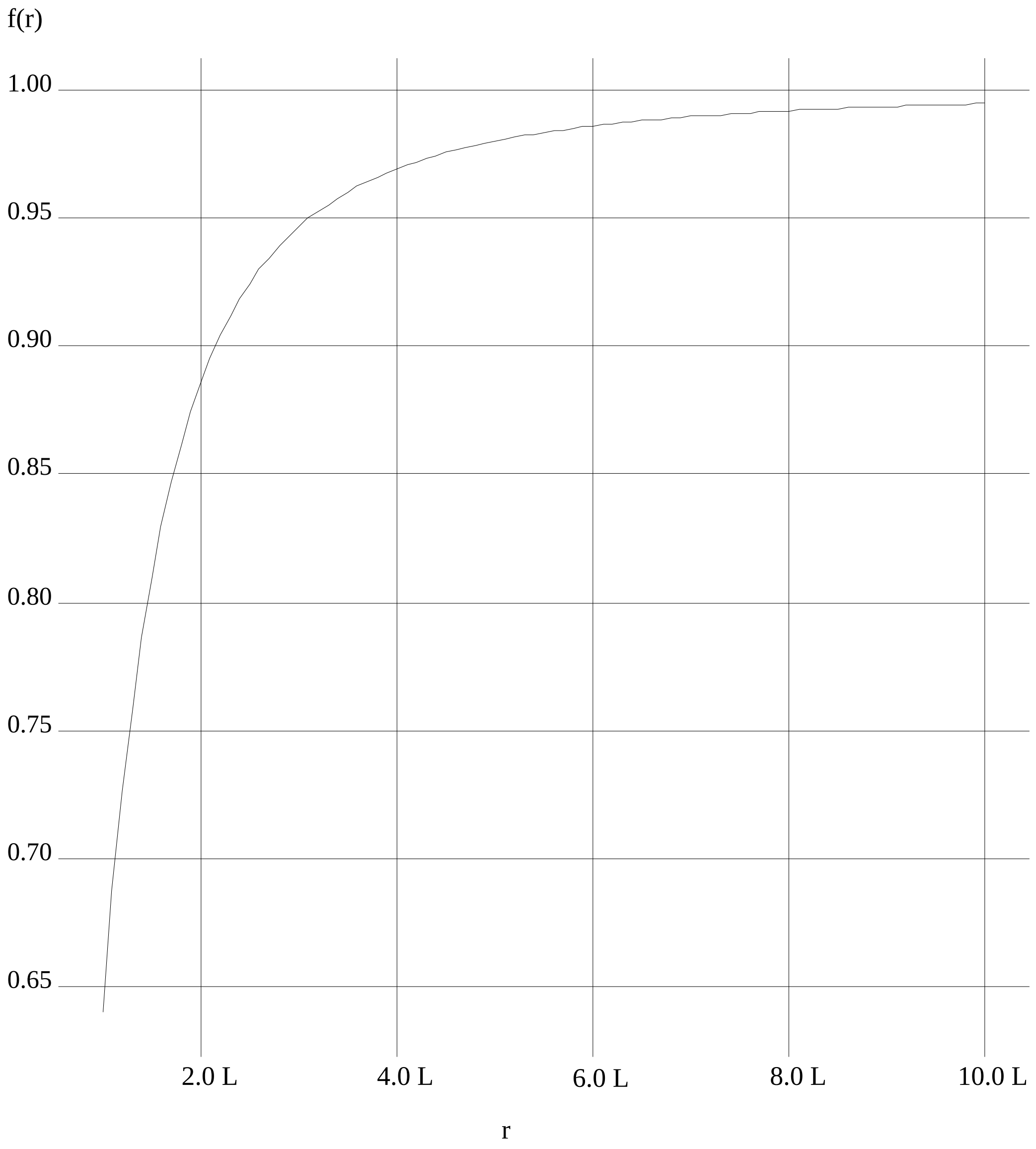

Figure 1-3: function f(r)

2. ORIENTATION INVARIANT PARAMETERS

2.1 Introduction

At present there is no common language for comparing two different light sources. Trying to quantify (or even to explain qualitatively) the difference in performance terms, for example, of an axial and a transverse filament, or a gas discharge tube and a light emitting diode, or a gas lamp mantle and a laser, is like comparing apples and oranges. We have developed a concept for a common language or representation for these diverse kinds of light sources, without reference to any coordinate system or light generation process. This concept, based on an expansion in spherical harmonics of an intensity dataset over a sphere, is therefore able to characterize the angular light distribution of a source without introducing orientation-dependent artifacts.

Section 2.2 presents an overview of this idea. Data from computer-generated light sources are shown in section 2.3, and data from real sources are in section 2.4. Details of the measurement and data fitting processes are described in sections 2.5 and 2.6. Appendices I and II provide formal proofs of the rotational invariance of these parameters, including a geometrical interpretation which leads to an amusing result. Appendix IX describes the experimental setup of the apparatus and includes pictures of the different bulb types. All of the datasets (for this chapter and the next) were taken by Mel and Alan; the apparatus no longer existed by the time I joined the project.

2.2 Orientation Independent Source Characterization

This luminaire model utilizes a goniometer to acquire data equally spaced in co-latitude θ and longitude ϕ over a sphere of radius r_0 surrounding the light source. Light emanating from the source in a particular (radial) direction is collected by a lens, focused on a stop, and then falls on the surface of a photodiode. The purpose of the lens and stop is to collimate the light, eliminating extraneous signals and illuminating the photodiode over its entire surface from a mainly forward direction.[6] The photodiode integrates the incoming light, resulting in a single

[6]No attempt is made to account for the angular response of the photodiode, but the optics keep the angle of incidence within a cone about the photodiode's surface normal.

intensity value for each point on the sphere. Because of its simplicity, this dataset is useful for calculating rotationally invariant quantities that characterize or describe the light source. Such measurements have been used for raytracing purposes [VER 84], but are only appropriate for that task in the far-zone.

The measured illumination distribution function[7] $I[\theta,\phi]$ is fit to a series of spherical harmonics times a set of complex coefficients $\{\alpha_{lm}\}$:

$$I[\theta,\phi] = \sum_{lm} \alpha_{lm} Y_{lm}[\theta,\phi]$$

Numerical details are in section 2.6. A comprehensive discussion of spherical harmonics is found in Varshalovich [VAR 88].

The coefficients α_{lm} themselves, however, cannot be used directly to quantitatively compare the performance qualities of different light sources; the same light source will produce a radically different set of coefficients if it is rotated with respect to the measuring apparatus. The problem then becomes one of finding rotational invariants among the coefficients to calculate numbers that capture the essence of the angular distribution of illumination produced by the source in question.

Because the Y_{lm} form an orthogonal basis in θ and ϕ, the α_{lm} can be thought of as the components of a vector whose length remains constant under rotations. Therefore,

$$\sum_{lm} |\alpha_{lm}|^2 = \text{constant} = N$$

That just says that the total mean-square intensity does not depend on the orientation of the light source about the center of the measurement sphere. But it is possible to go even further.

Consider now an arbitrary rotation of coordinates. The Y_{lm} transform as [ROS 57]

$$Y_{lm}(\theta',\phi') = \sum_{n} D^{l}_{nm} Y_{ln}(\theta,\phi)$$

where the **D** are the Wigner **D**-matrices used in the quantum mechanics of angular momentum [ROS 57], and depend only on the Euler angles of the rotation. They possess two important properties. The first is that there is a different **D** for each *l*-value. Thus a rotation can

---

[7]Throughout this thesis, experimental parameters will be in square brackets and free variables will be in parentheses.

transform Y_{20} into some combination of Y_{21} and Y_{22}, but it can never be transformed into Y_{30}; rotations do not mix l-values.[8] The second important fact is that the **D**-matrices all have determinant +1 and are therefore length preserving in each l subspace. It follows then that for **each** l, the length of the projection of the "coefficient vector" $\vec{\alpha}$ into the subspace l is rotationally invariant, or

$$\{\sum_m |\alpha_{lm}|^2/N\}^{1/2} = \text{constant} = d_l$$

A more detailed proof is given in appendix I.

The factor of N is included to normalize the d-numbers so that the squares of the d_l sum to unity, allowing them to be compared on an equal footing. The d_l, then, are rotationally independent quantities that can be used to characterize and compare the angular distributions of different light sources: d_0 represents the "monopole" strength (I_{avg}/I_{rms}), d_1 the "dipole" strength, and so on. Appendix II shows that physically the d_l represent the ratio of *rms* power distributed in the angular mode l to the total *rms* power. In the far-zone the normalized d_l become constant with radius as well as orientation, because the power (and hence all of the α_{lm}) is falling as $1/r^2$.

N and α_{00} have the following physical meanings. The mean-square intensity is defined as

$$\begin{aligned}
I_{ms} = I_{rms}{}^2 &= \int_{4\pi} d\Omega\, I(\Omega)^2/4\pi \\
&= \int_{4\pi} d\Omega \sum_{lm} \alpha_{lm} Y_{lm}(\Omega) \sum_{l'm'} \alpha_{l'm'} Y_{l'm'}(\Omega)/4\pi \\
&= \sum_{ll'mm'} \alpha_{lm}\alpha_{l'm'} \int_{4\pi} d\Omega\, Y_{lm}(\Omega)\, Y_{l'm'}(\Omega)/4\pi \\
&= \sum_{ll'mm'} \alpha_{lm}\alpha_{l'm'}\delta_{ll'}\delta_{mm'}/4\pi \\
&= \sum_{lm} |\alpha_{lm}|^2/4\pi = N/4\pi
\end{aligned}$$

Thus

$$N = 4\pi I_{ms}$$

[8]This is related to the fact that the Y_{lm} are the eigenfunctions of the angular momentum operator, where l depends on the total angular momentum and m is a component; rotations can change angular momentum components (m values), but cannot change the total angular momentum (l values).

Furthermore, the average intensity is

$$I_{avg} = \int_{4\pi} d\Omega\ I(\Omega)/4\pi$$

$$= \int_{4\pi} d\Omega\ I(\Omega)\, Y_{00}{}^2 = \alpha_{00}\, Y_{00}$$

so that

$$\alpha_{00} = \sqrt{4\pi}\, I_{avg}$$

Another characteristic parameter can be calculated from the coefficients to represent the anisotropy, or standard deviation σ of the illumination distribution. Because $\alpha_{00}(4\pi)^{-1/2}$ is the mean intensity and $N(4\pi)^{-1}$ is the mean-square intensity, $\sigma=(N-\alpha_{00}{}^2)^{1/2}/\alpha_{00}$ is the normalized "scatter" of the luminaire, in which there has been some recent interest in the illuminating engineering literature [ARM 90, MAT 89]. Therefore σ essentially measures the dominance (or lack thereof) of the first term of the expansion with respect to all of the others.

2.3 Simulated Demonstration

A simulated (simple, noise-free) source was used to demonstrate the rotational invariance of the d_l in practice. The illumination pattern of a glowing hollow cylinder of constant radiance was determined by raytracing for first a horizontal orientation and then a vertical one. Because of the lack of noise, any discrepancies in this example will quantify the distorting effect of finite, discrete sampling: the precise locations of sampled radiator patches will not exactly correspond between the two datasets, and thus the two datasets are not actually perfectly congruent under any rotation.

Although the two sets of $\{\alpha_{lm}\}$ coefficients are very different, when the d_l are calculated as shown in table 2-1,[9] the fact that the glowing cylinders are actually alike (to two decimal places) becomes obvious; the effect of sampling on this scale is seen in the third decimal.

[9]For this case, all of the odd l d_l are zero by cylindrical symmetry.

raw coefficients and computed d-numbers					
coef	vertical	horizontal	d	vertical	horizontal
$\alpha_{0,0}$	+2.490	+2.494	d_0	0.983	0.983
$\alpha_{2,0}$	-0.443	+0.223	d_2	0.175	0.176
$\alpha_{2,2}$	+0.000	+0.547	d_4	0.043	0.041
$\alpha_{4,0}$	-0.109	-0.039	d_6	0.022	0.023
$\alpha_{4,2}$	+0.000	-0.083	d_8	0.022	0.021
$\alpha_{4,4}$	+0.000	-0.109	d_{10}	0.014	0.014
$\alpha_{16,0}$	-0.017	-0.001	d_{12}	0.012	0.010

Figure 2-1: the coefficients are different but the d-numbers are similar

2.4 Application to Real Sources

Four different coiled-filament headlamps[10] were investigated. They are a horizontal filament "spot quality" lamp surrounded by a smooth glass bulb; a vertical axial filament "high beam" lamp with a cylindrical bulb that becomes conical on top; a vertical filament axial "low beam" lamp with a cylindrical bulb and a light-blocking cap on top; and a "dual beam" lamp with two horizontal filaments for high and low beams, with a cap on top. The spot quality bulb is used in the maufacturer's reflector design process to evaluate reflector prototypes, to hand finish dies, etc. The other three bulbs are standard headlamp bulbs, with part numbers 9005, 9006, and 9004 respectively (see appendix IX for photographs).

The intensity dataset for the spot quality bulb is shown pictorially in figure 2-2, where the horizontal direction represents azimuth ϕ and the vertical direction represents elevation θ. For this dataset, the stepsizes were 3 grads[11] in θ and 2 grads in ϕ. The dataset is shown as a sine-mapped spherical projection so that the pixel areas are proportional to the areas actually sampled over the sphere. The top of the dataset corresponds with the top of the bulb. The lower portion (roughly $1/8\,th$) of the dataset is unavailable; it is essentially a "hole" of (almost entirely) zero intensity caused by the physical inability of the goniometer to sample points near the shaft

[10] provided by a major automotive lighting supplier

[11] 100 grads = 90 degrees = $\pi/2$ radians. The stepper motors on our goniometer are built with 400 steps per revolution, so our measurements are naturally reported in integral numbers of grads.

holding the source (see appendix IX).

After fitting, when regenerated and interpolated at 1x1 grad steps, the processed data look like figure 2-3. Some features of this picture are worth mentioning. The two large dark spots are the ends of the horizontal filament, and the vertical posts holding the ends of the filament are also just visible. The "rings" are not artifacts; they appear clearly on real data taken with a finer grid, using steps of 1 grad each in θ and ϕ, as shown in figure 2-4. These dark rings are apparently due to "see-through" between the filament coils for viewing angles in ϕ that are nearly perpendicular to the filament.

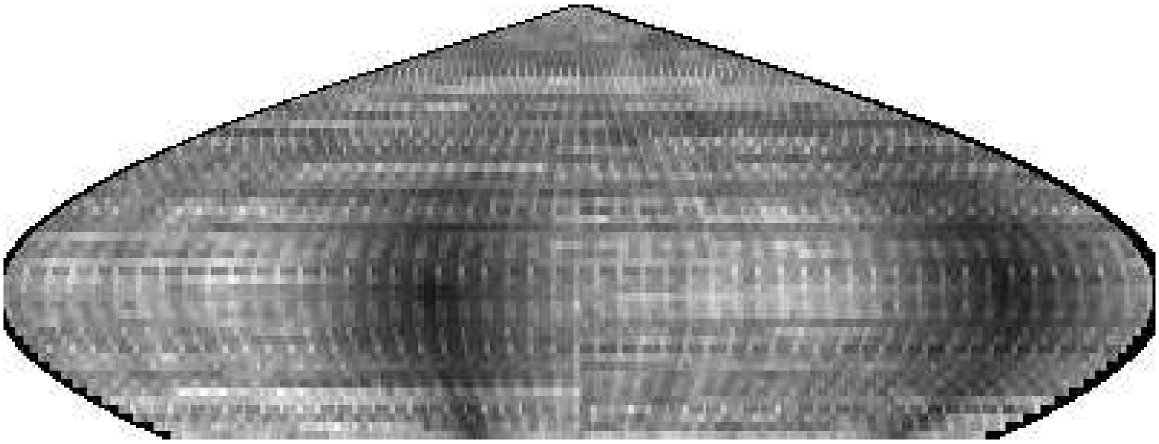

Figure 2-2: spot quality horizontal filament bulb, intensity data, 2x3 grad steps

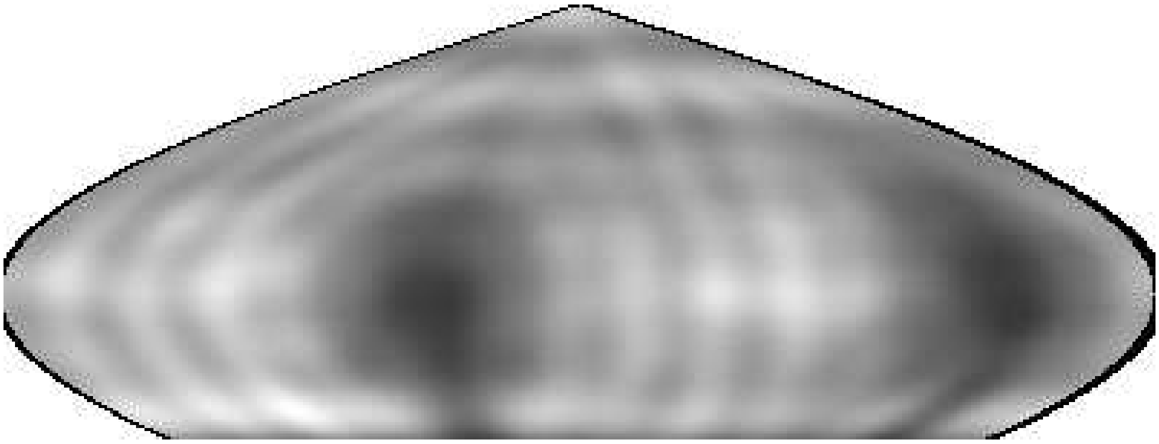

Figure 2-3: spot quality horizontal filament bulb, regenerated data, 1x1 grad steps

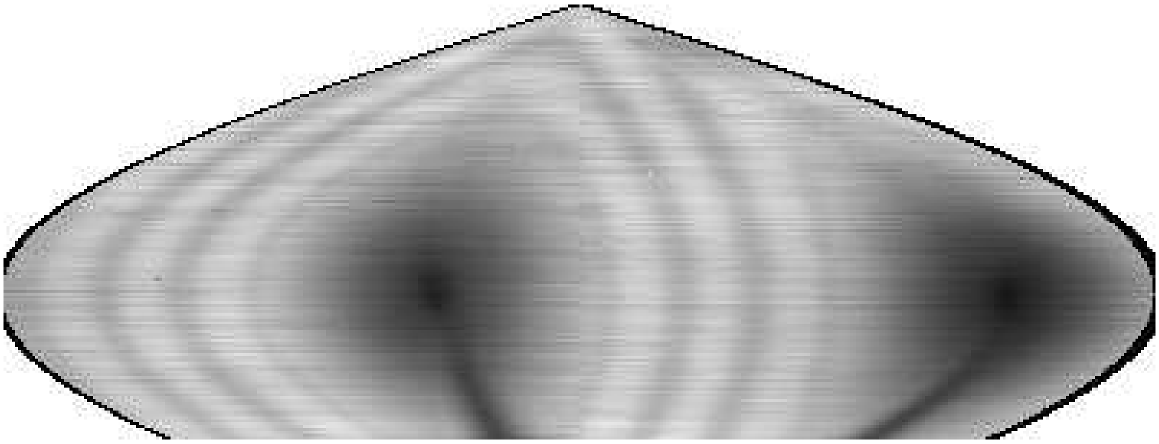

Figure 2-4: spot quality horizontal filament bulb, intensity data, 1x1 grad steps

A high beam lamp dataset is shown in figure 2-5, and its regeneration at six times the sampling density is shown in figure 2-6. Note the horizontal "see-through" rings near the center of the vertical filament. The dark stripe from pole to pole (of a few degrees azimuth) is the vertical filament mounting post; the other dark stripe 180° from the mounting post may be due to the post reflecting less light than the quartz bulb envelope.

Regenerated datasets for dual and low beam lamps are also shown in figures 2-7, 2-8, and 2-9.

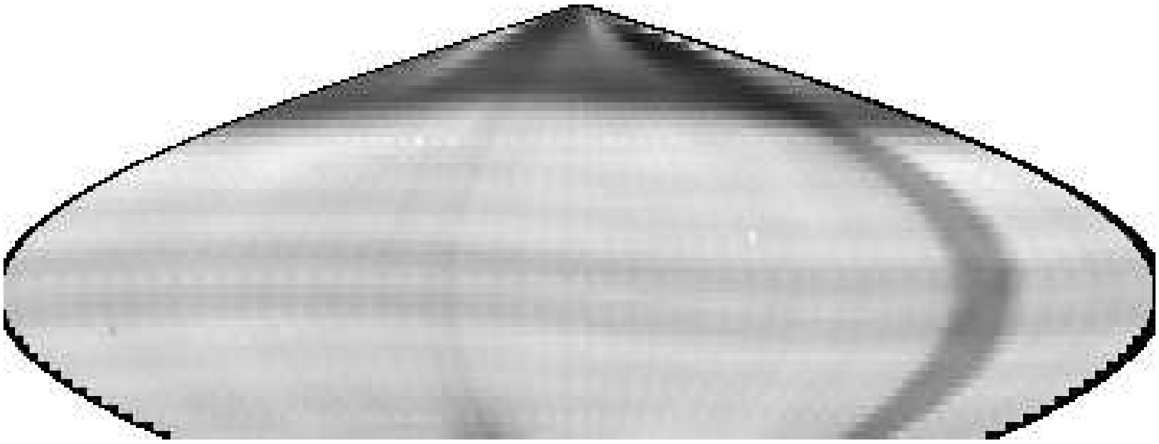

Figure 2-5: high beam vertical filament bulb, intensity data, 2x3 grad steps

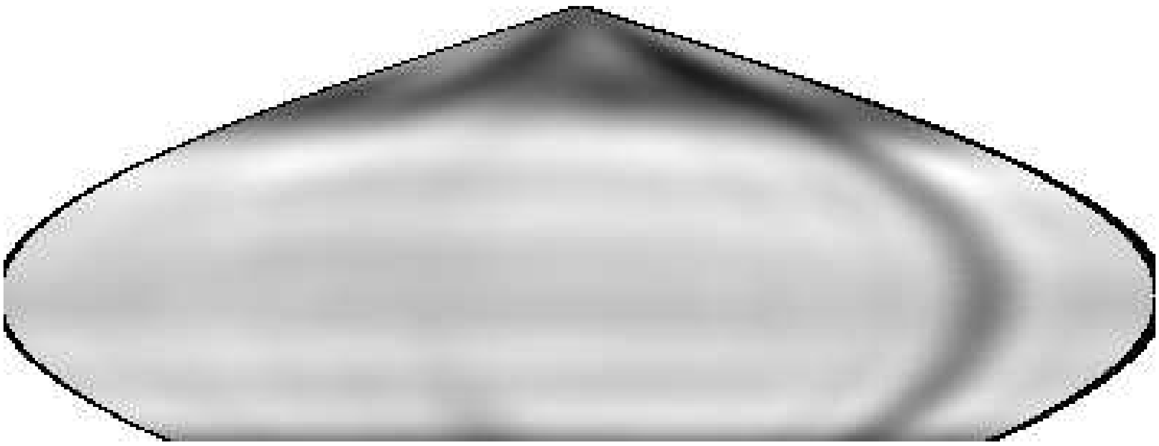

Figure 2-6: high beam vertical filament bulb, regenerated data, 1x1 grad steps

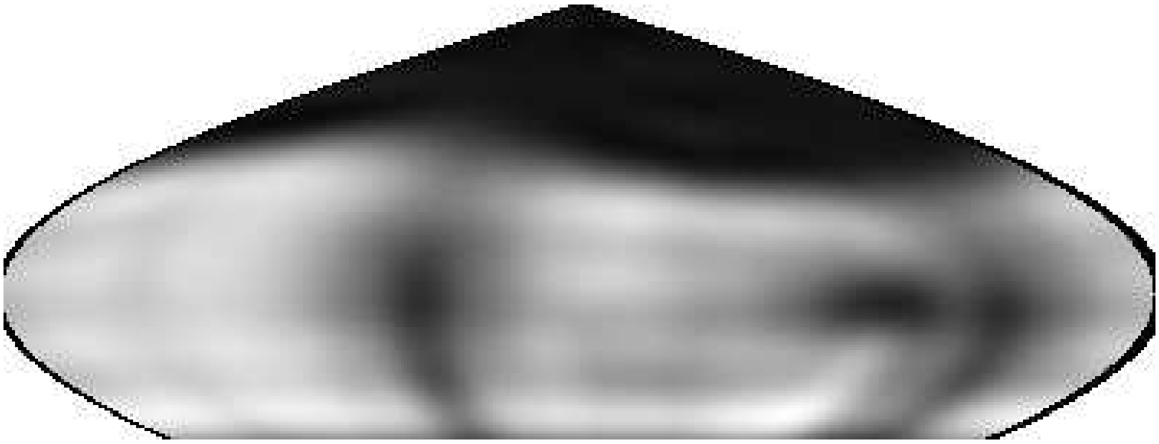

Figure 2-7: dual horizontal filament bulb, low beam, regenerated data, 1x1 grad steps

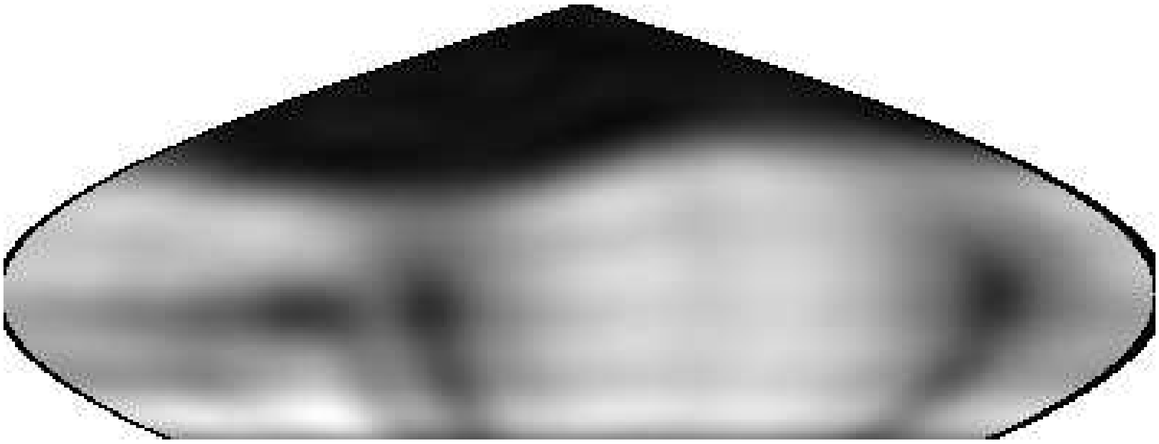

Figure 2-8: dual horizontal filament bulb, high beam, regenerated data, 1x1 grad steps

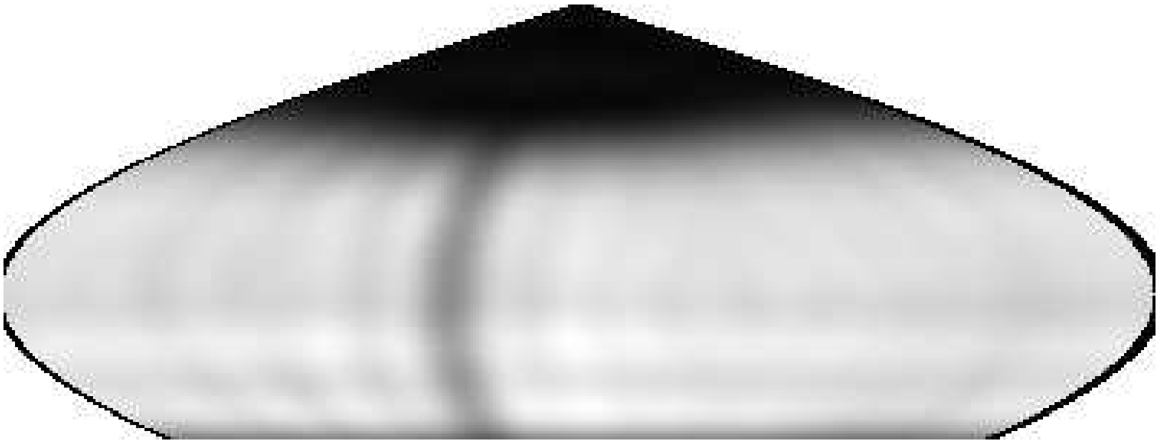

Figure 2-9: low beam vertical filament bulb, regenerated data, 1x1 grad steps

2.5 Measurement and Uncertainties

The goniometer operates by making a full sweep in ϕ from -200 to +200 grads before incrementing θ, which ranges from -150 to +150 grads (due to physical limitations, the goniometer could not sample the lower portion of the sphere). Thus each point is sampled twice (once with positive θ and once with negative θ), except for points on the "dateline" great circle which are sampled three times, and the pole which is sampled as many times as there are ϕ values. Multiple readings are averaged.[12]

The many readings at the pole are used to obtain an estimate of the uncertainty or standard deviation of the measurements, which is then used to estimate the uncertainties of the α_{lm} and, by extension, the uncertainties of the d_l by the usual "propagation of error" techniques [BEV 69]. Typical values for the reduced σ of the intensity readings at the pole are in the range of ten to twenty percent of the pole's mean, showing the datasets were plagued by noise (mostly of mechanical origin). Knowing the uncertainties would be important, for example, for creating a fuzzy-logic quality control system to compare the *d*-number profile of a light source with a standard library of profiles. The uncertainty in the data, coupled with the effect of finite sampling, implies the d_l are reliable to one or one and a half decimal places.

A method for automated quality control in lamp production has been outlined in a paper by Lewin [LEW 90]. The typical quantities to be measured, however, are somewhat primitive, consisting of isocandela diagrams and average intensities; the *d*-numbers present a much more sophisticated standard for comparison.

As an example, consider the bulbs of section 2.4. As expected, the intensity datasets for these filaments vary greatly, but the *d*-number profiles reveal certain similarities (to at least the first decimal) among the angular distributions of the various bulb types. The *d*-number profiles are shown in figure 2-10 for the first few values. The high beam dual filament, for instance, produces an intensity dataset that is rotated by 180° about a vertical axis and has more power than the low beam dual filament, but their respective *d*-number profiles are essentially identical.

[12] A baseline current was probably "bucked out" of the data when originally taken, but this value has unfortunately been lost.

The high beam and spot quality bulb are also similar due to the fact that both are identical filaments, but with differing orientations. Differences become apparent in higher orders, however, because the "holes" in the datasets, where data was unavailable, break the symmetries of the bulbs in different places. The low-beam bulb, with an axial filament and a cap to balance the lower octant hole, is nearly symmetric under parity transformations and thus has the smallest odd-l d_l of all the bulbs.

2.6 Data Fitting

The equation representing the integrated intensity dataset is

$$I[\Omega_i] = \sum_{lm} \alpha_{lm} Y_{lm}[\Omega_i]$$

where $[\Omega_i]$ is the ith measurement of $[\{\theta,\phi\}]$. This suggests the matrix equation

$$\boldsymbol{I} = \boldsymbol{a}\vec{\alpha}$$

Here

$$I_i = I[\Omega_i]$$

$$a_{ij} = Y_{lm}[\Omega_i]$$

and

$$\alpha_j = \alpha_{lm}$$

where the j in the previous two equations stands for the jth value of lm.

Because there are more measurements than coefficients, the system of equations can be solved by a least-squares ("LSQ") fit:

$$\boldsymbol{a}^T\boldsymbol{I} = \boldsymbol{a}^T\boldsymbol{a}\vec{\alpha} = \boldsymbol{b}\vec{\alpha}$$

Thus the coefficients are found by [BEV 69]

$$\vec{\alpha} = \boldsymbol{b}^{-1}\boldsymbol{I}$$

where the matrix $\boldsymbol{b}$ is invertible by Gauss-Jordan elimination.

It should be noted, however, that although the LSQ method produces a low-residual fit, the coefficients it finds may be wildly "unphysical" due to an instability in the solution for the coefficients. This instability results from the fact that our measuring apparatus cannot sample the lower octant of the sphere surrounding the light source. The incompleteness of the interval in θ causes a loss of orthogonality (implying a loss of linear independence) among the spherical harmonic basis functions, and consequently introduces an ambiguity into the contribution of

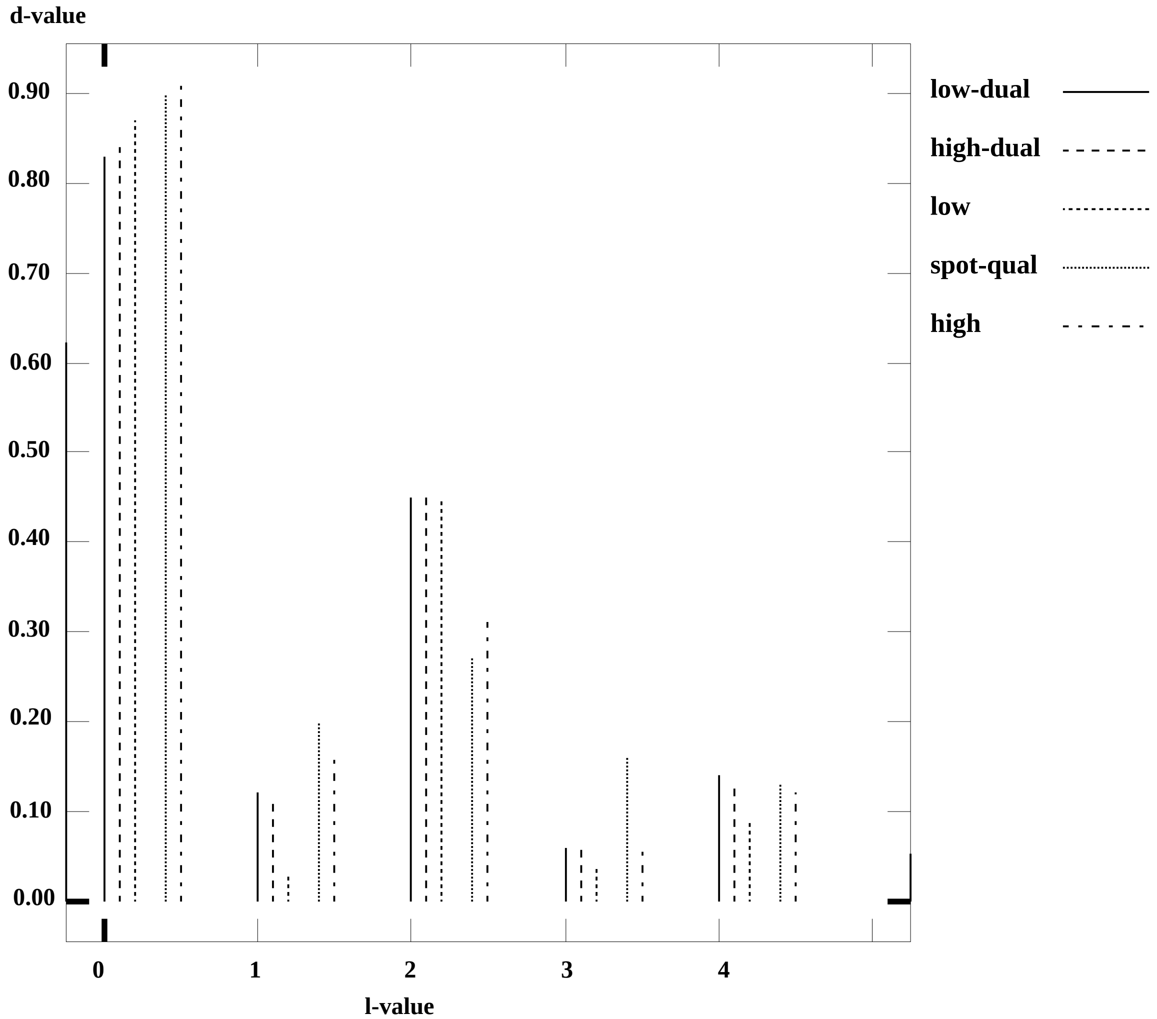

Figure 2-10: *d*-numbers for various bulb types

some of the terms. Thus the values of the coefficients change as more terms are added [HAM 73], and it becomes impossible to ascribe physical meaning to the *d*-numbers.

The point I wish to stress is that I am not simply interested in a good fit to the data, for which LSQ would be sufficient. Instead, I would like to extract information about the actual intensity distribution from the data. Some physical meaning must therefore apply to the coefficients, and a necessary requirement is that the coefficients must be calculable incrementally, without new terms affecting the values of previous terms. Thus I found the LSQ method to be unsuitable for this calculation with this data; had the data interval been complete, the LSQ method would work (although a direct discrete Fourier dot product would be faster).

In theory, linear independence is a well-defined notion; two vectors are either linearly independent or they aren't. There is no room for dispute. In practice, however, especially in the realm of finite-precision numerical calculation, vectors can be partially dependent. In fact, it is possible to quantify mathematically the degree to which two vectors are linearly dependent [FOR 77, PRE 88].

The method of singular value decomposition ("SVD") is specifically designed to address this problem [FOR 77, GOL 70, HAM 73]. Although the matrix $\boldsymbol{a}$ is not square, a pseudoinverse exists. Any arbitrary matrix can be written as the product of three trivially invertible submatrices:

$$\boldsymbol{a} = \boldsymbol{uwv}$$

If $\boldsymbol{a}$ is m by n then so is $\boldsymbol{u}$, and $\boldsymbol{w}$ and $\boldsymbol{v}$ are both n by n. Both $\boldsymbol{u}$ and $\boldsymbol{v}$ are orthogonal, meaning the pseudoinverse equals the transpose. $\boldsymbol{w}$ is diagonal, meaning its inverse is found simply by taking the reciprocal of its diagonal terms. Thus once this decomposition has been performed, the pseudoinverse of $\boldsymbol{a}$ is found immediately to be

$$\boldsymbol{a}^{-1} = \boldsymbol{v}^T \boldsymbol{w}^{-1} \boldsymbol{u}^T$$

So far, nothing new has been gained. The diagonal terms of the matrix $\boldsymbol{w}$, called the "singular values" of the matrix $\boldsymbol{a}$, however, allow the identification of exactly which basis functions are becoming the most linearly dependent: setting the most ill-behaved w_{ii} to zero prevents those terms whose contribution to the Fourier sum is not uniquely determined from influencing the coefficients, whereas the LSQ method assigns large, delicately cancelling values to these coefficients [PRE 88].

The relative stability of the two methods is shown in figure 2-11, which gives the coefficient α_{00} as a function of l_{max} for each method. The solid line shows the LSQ method, and the dashed line shows the SVD method. The two methods essentially agree up to $l_{max} = 4$, after which point SVD starts to zero out singular values. Thus a degree of stability can be achieved with SVD, but the "black art" involved in deciding exactly how large a singular value must become in order to be considered "ill-behaved" makes this method still less than satisfactory.

One solution to this problem would be to use a new set of fitting functions that would be designed to be orthogonal over our specific data interval. The introduction of nonstandard functions, however, would mean a corresponding loss of both acceptance and physical intuition. Another approach to this problem would be to use an interpolating *sinc* or spline function. Because strict interpolations go through every data point, this method is not desirable for data that are noisy or that contain occasional large outliers. The idea is to smooth the noise over the entire dataset, rather than fit to it at each datapoint. For a given order, increasing the density of sampled points also reduces the ambiguity of the solution, but does not competely eliminate it.

The approach adopted, then, is to "fill in" the missing octant with data of intensity zero, with the justification that this region corresponds roughly with the shadow of the base of the bulb. Because the spherical harmonic fitting functions are now orthogonal over our dataset, the least-squares fit becomes equivalent to a discrete Fourier "dot product" integral. Thus the coefficients are found directly by

$$\alpha_{lm} = \sum_i Y^*_{lm}[\theta,\phi]_i \, I[\theta,\phi]_i \sin\theta_i \, \Delta\theta \, \Delta\phi$$

where Y^*_{lm} is the complex conjugate of Y_{lm}, and the effort expended for an SVD matrix operation is avoided. The coefficients are also guaranteed to remain constant as more terms are added. Furthermore, if an LSQ matrix inversion is used with the complete filled-in dataset, statistical tools are built in [BEV 69] that allow for the automatic computation of the uncertainties of the coefficients α_{lm}, although the computation is a little bit slower than the direct evaluation of the dot product.

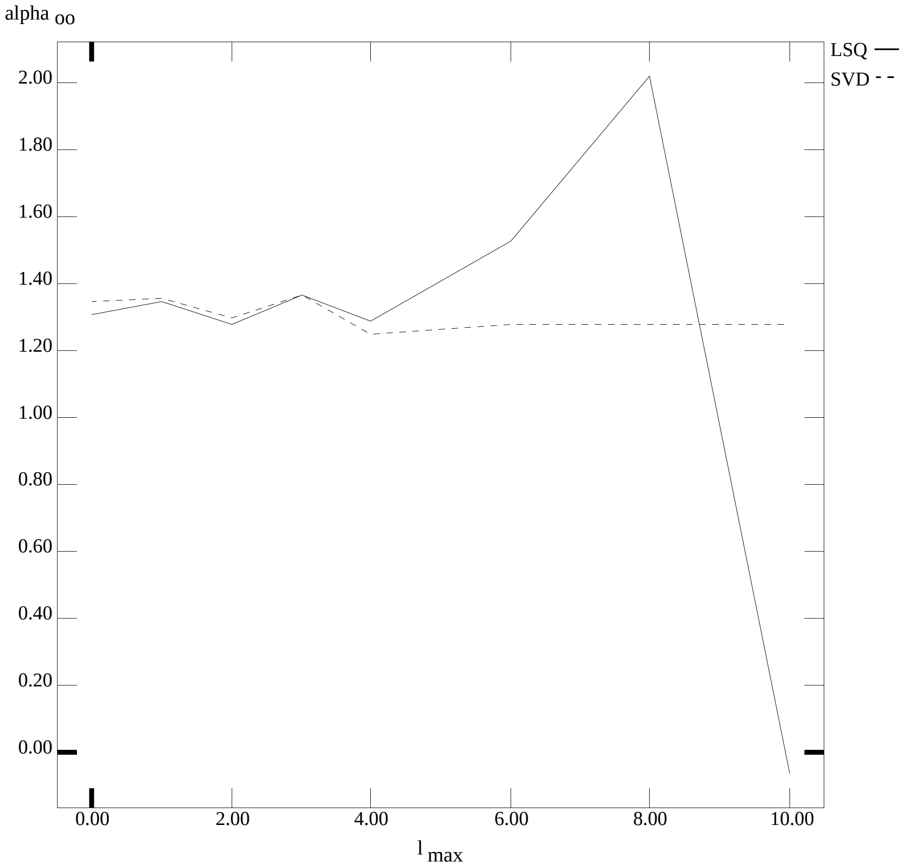

Figure 2-11: stability comparison of SVD and LSQ solution methods

Because the source is extended, the "shadow" region between light and dark should extend over a few values of θ.[13] I have observed that smoothing the cutoff from light to dark linearly (for example), as opposed to creating a sharp drop, does indeed reduce the residual of the fit. Because the extent of the shadow region is not known, however, the results of this chapter do not include smoothing of the cutoff (but see section 5.2).

An important remaining question concerns the maximum value of l used for the fit. The dataset pictures appear to regenerate well (by eye) typically around $l_{max} = 16$; adding more terms does not seem to improve the fit significantly enough to justify the additional computation. A graph of the reduced χ^2 of the fit for a high beam bulb shows that $\chi^2 \rightarrow 1.2$ at $l_{max} = 16$, beyond which it levels out (figure 2-12).

[13]Section 5.2 contains an estimate of this region's location and extent.

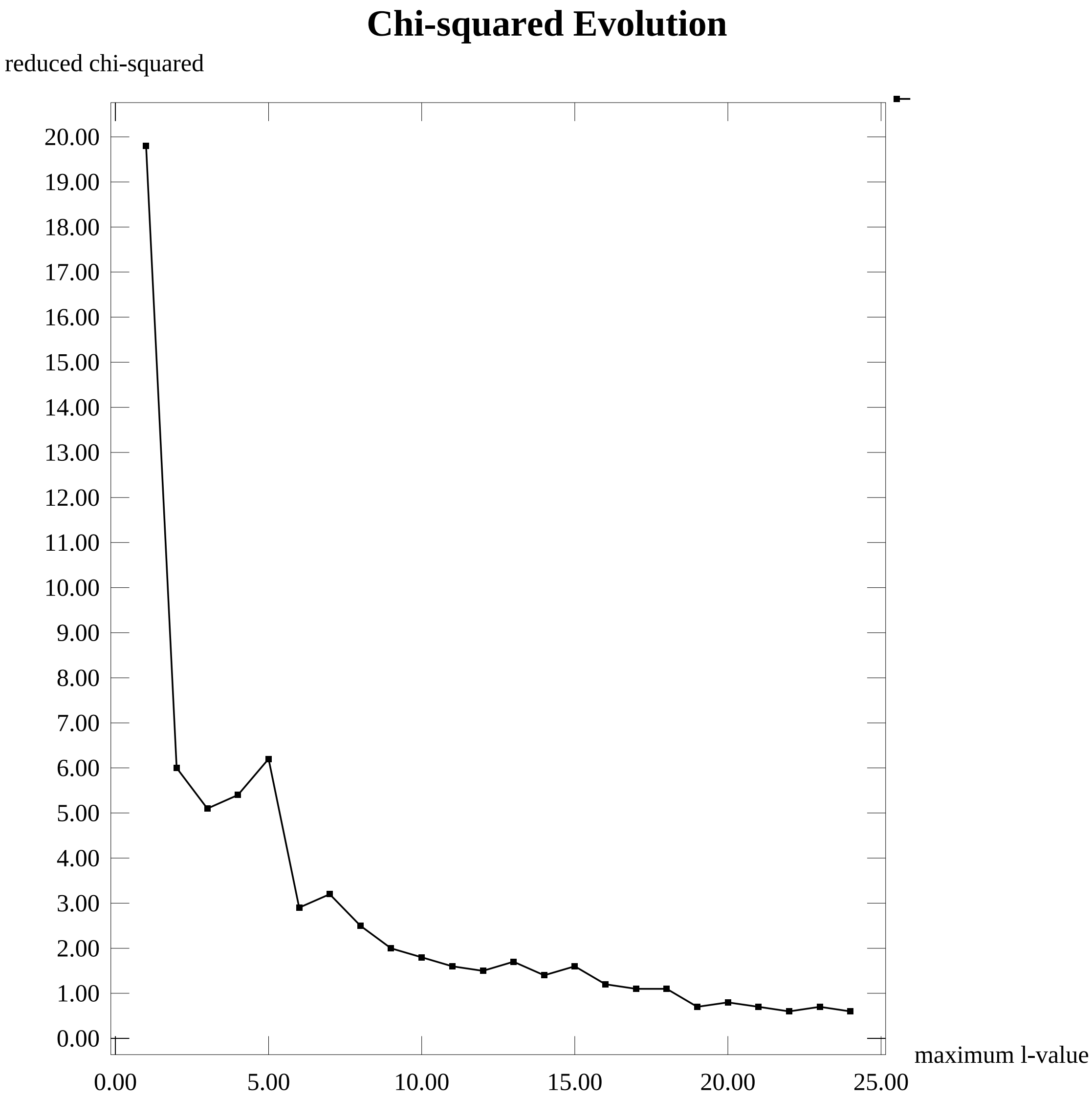

Figure 2-12: graph of reduced chi-squared versus l_{max}

3. A GENERAL NEAR-ZONE LIGHT SOURCE MODEL

3.1 Introduction

Real light sources (extended bodies with nonuniform radiance distributions) lack a physics-based near-zone model that is both generally applicable regardless of source type or geometry, and computationally efficient for ray tracing, computer graphics, luminaire design, etc. We develop such a model by providing

1. a prescription for making a set of directional radiance measurements that fully characterize the source,
2. a robust and highly compressed representation of the measurements in terms of the coefficients of Fourier expansions in suitable orthogonal functions, and
3. ray tracing to this spherical source representation, which reduces the normally expensive intersection problem to the algebraic solution of a quadratic equation.

The required physical measurements are a set of two-dimensional CCD pinhole camera images of the source taken, using a goniometer mechanism, at several-degree intervals of latitude and longitude on a sphere that surrounds the source. We show that this data *on a sphere* describes a virtual source that (outside the sphere) is exactly equivalent to the real source. However, in ray tracing, intersecting the sphere is algebraic, whereas intersecting the real source can be a difficult numerical search problem.

We represent the virtual source via Fourier expansion coefficients that describe the directional radiance distribution measurements. This gives our datasets approximately three orders of magnitude of compression (section 3.3.3), a natural interpolation method (section 3.3.4), and negligible sensitivity to outlier noise in the initial data (section 3.3.1). As a bonus, this approach gives us a simple (no search for intersections) way to check for the occlusion of any ray in the scene by radiant parts of the source (section 3.5).

Section 3.2 discusses how a CCD pinhole camera picture can be used to find the directional radiance distribution of a source. Section 3.3 describes our multidimensional Fourier decomposition of a complete dataset of pinhole camera pictures taken over a sphere surrounding the source, in order to replace the actual source with an equivalent spherical source, thereby rendering the intersection problem of classical raytracing trivial. Mathematical details of the fitting process are given in appendices III and IV. Section 3.4 shows how an actual light transport calculation would work, and section 3.5 explains how it may be possible to check for the self-occlusion of rays by light-emitting portions of the source with only sphere-intersection checks; the directional nature of the radiance distribution over the measurement sphere actually preserves some information about the shape and location of the actual source within the sphere. Section 3.6 shows why this model is generally applicable to all sources, regardless of the details of their geometry or light generation process. Appendix VI provides some mathematical details concerning ray tracing.

3.2 Radiance From Pinhole Camera Measurements

The source model begins with data acquired by using a goniometer to scan the measurement sensor over the surface of a sphere that encompasses the radiant source.

Goniometer-based measurements have previously been used in computer graphics for simulating extended light sources [VER 84]. These measurements have been of radiant intensity integrated over incoming directions. They are thus useless for raytracing *in the near-zone*:[14] the light flow is not entirely radial, so the radiance distribution is a function of four variables (3 for position, 2 for direction of view, and -1 for the constancy constraint along lines) and the scaling of intensity is therefore **not** $1/r^2$. In fact, in the near-zone of a real light source there is really no unique "center" (or line, plane, etc) from which to measure a distance with which radiant intensity might scale!

To correctly model extended sources in the near-zone we need measurements with additional detail: we need to measure the differential intensity per unit directed solid angle, not just the integral intensity over all directions. We obtain the required differential data from the goniometer by mounting a CCD pinhole camera on it, placing the source near the center of the goniometer, and collecting images. The pinhole defines latitudes and longitudes on the sphere, each pixel-pinhole pair defines a direction, and the gray level reported by each pixel is proportional to the radiant intensity transmitted by the pinhole into the solid angle subtended by

[14]They are useful in the far zone; see, for example, chapter 2.

that pixel as seen from the pinhole. Each pixel-aperture pair thus acts as a miniature radiance meter, and the many pixels effectively allow many measurements to be taken at once.

An Electrim EDC-1000 digitally controlled monochrome CCD camera was used to take the filament images. This camera has pixels of 13.75 μm by 16 μm in an array of 192 by 165, giving a sensing area of 2.64 mm by 2.64 mm. Five pinholes were obtained from Edmund Scientific, with diameters in the ranges of 5-12 μm, 20-35 μm, 40-70 μm, 85-120 μm, and 200-300 μm. The 40-70 μm pinhole produced the least image crosshatching,[15] and was used for subsequent image gathering. To reduce blooming, an infrared blocking filter and an 0.8 neutral density filter were also placed in front of the pinhole. Measurements were typically taken with a radius between 2.22 and 5.22 inches.

The pinhole camera image of a spot quality light bulb filament (see section 2.4 for a description of bulb types), shown as a negative for clarity, is typically a "blob" like the one shown in figure 3-1.

Figure 3-1: a "blob" (a pinhole CCD camera negative)

Because an aperture uniformly illuminated in a given direction by a radiance B from one side acts just like an area source of the same radiance with respect to objects on the other side, a single pinhole camera picture can measure the radiance distribution through a small patch of area corresponding in location, size, and orientation with the camera's aperture. Many such pictures

[15]This determination was performed by eye by Mr. Alan Guisewite.

can be used to characterize the full radiance distribution of the source over some closed (conveniently spherical) surface. The entire dataset[16] of all the pinhole camera pictures taken over the measurement sphere is shown in figure 3-2. A full dataset[17] of the low beam in a dual filament bulb is shown in figure 3-3 for comparison. Obviously, it is important to adjust the geometry to ensure that each blob fits entirely within the CCD array.

If, as in figure 3-4, the aperture is placed on a sphere at coordinates (r_0,θ_0,ϕ_0), the radiance B passing through this point in the direction of pixel (x_0,y_0) is

$$B(r_0,\theta_0,\phi_0;x_0,y_0) = I(r_0,\theta_0,\phi_0;x_0,y_0)\,f^2/(A_p A_q \cos^4\beta)$$

Here I is the power measured at the pixel, f is the distance to the image plane of the camera, β is the angle of incidence, and A_p and A_q are the pixel and aperture areas, respectively. Two cosine factors appear for the projected areas, and two are due to the squared pinhole-to-aperture distance.

Ashdown has outlined a similar data gathering procedure [ASH 93]. He too records the angular distribution of light at multiple points on the sphere. However his methods of reducing, representing, and using the data in downstream computations appear to be less developed than ours.

3.3 Data Analysis and Representation

3.3.1 The Harmonic Oscillator Wavefunction Expansion

First we decompose each 2-dimensional image into a multi-parameter representation in which the parameters are intimately connected with our intuitive perception about how best to convey the description of a "blob":

- *Parameters x_{00} and y_{00}:* the x and y coordinates of the center-of-intensity;
- *Parameters σ_x and σ_y:* the standard deviations of the intensity distributions in a frame centered on the center-of-intensity and oriented along and perpendicular to the tilt-axis;
- *Parameter a_{00}:* the tilt-angle, with respect to the x-axis, of the elliptical cross-

[16]This dataset is known as "v102591a.dat."

[17]This dataset is known is "v100891a.dat."

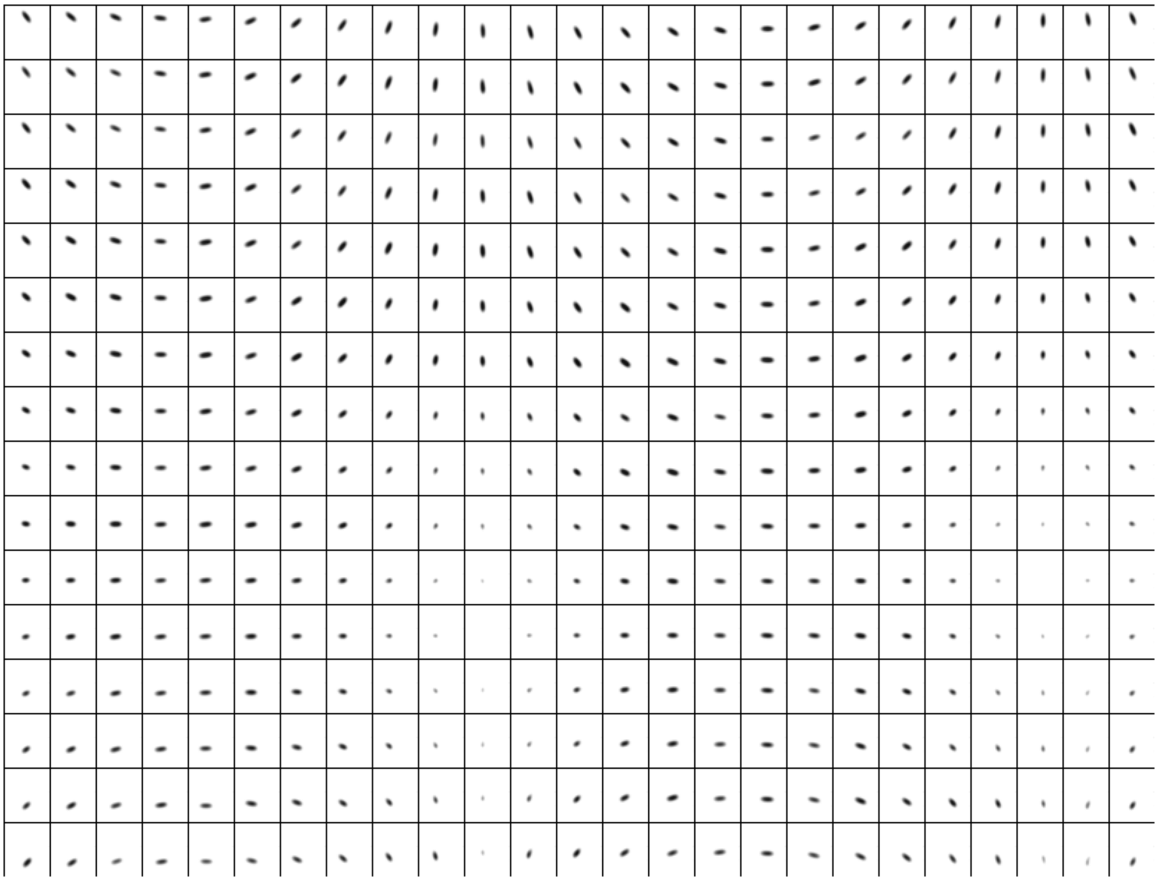

Figure 3-2: a set of spot quality "blobs" (vertical θ, horizontal ϕ)

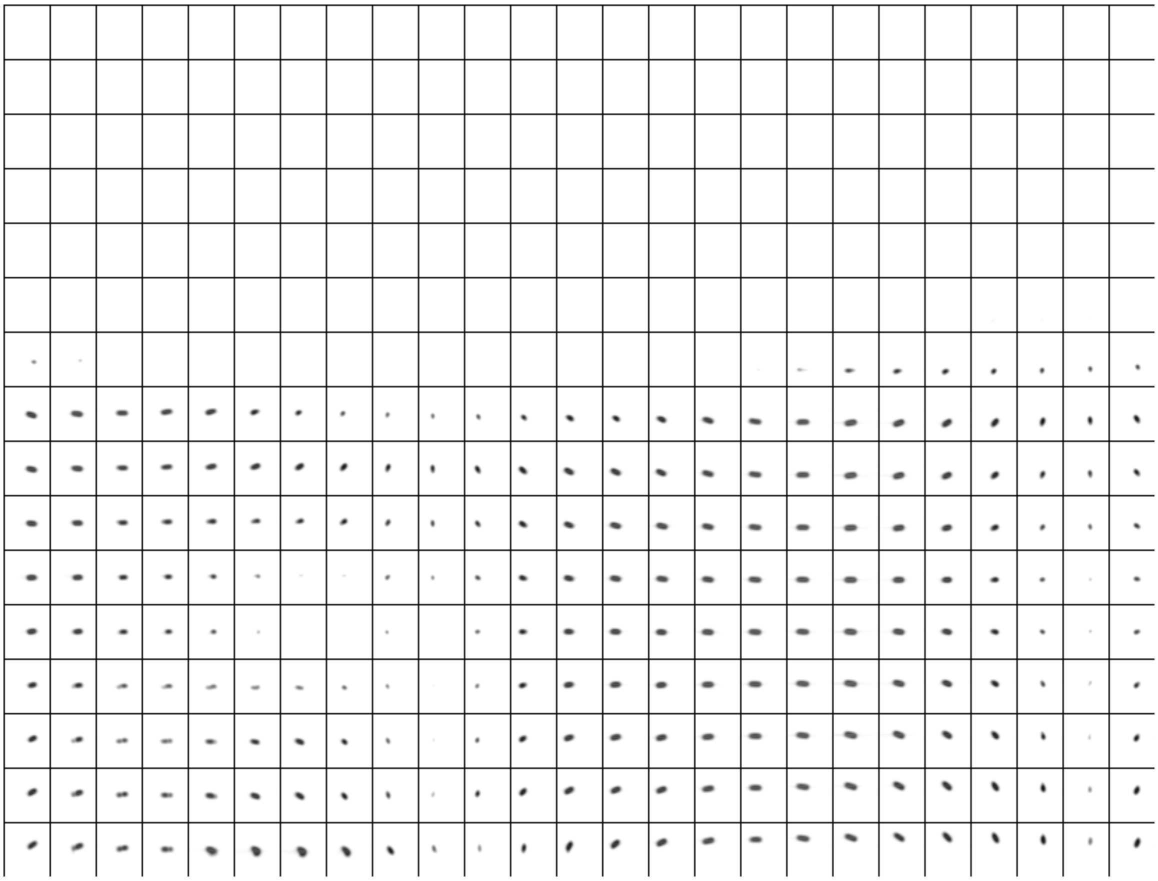

Figure 3-3: a set of dual low "blobs" (vertical θ, horizontal ϕ)

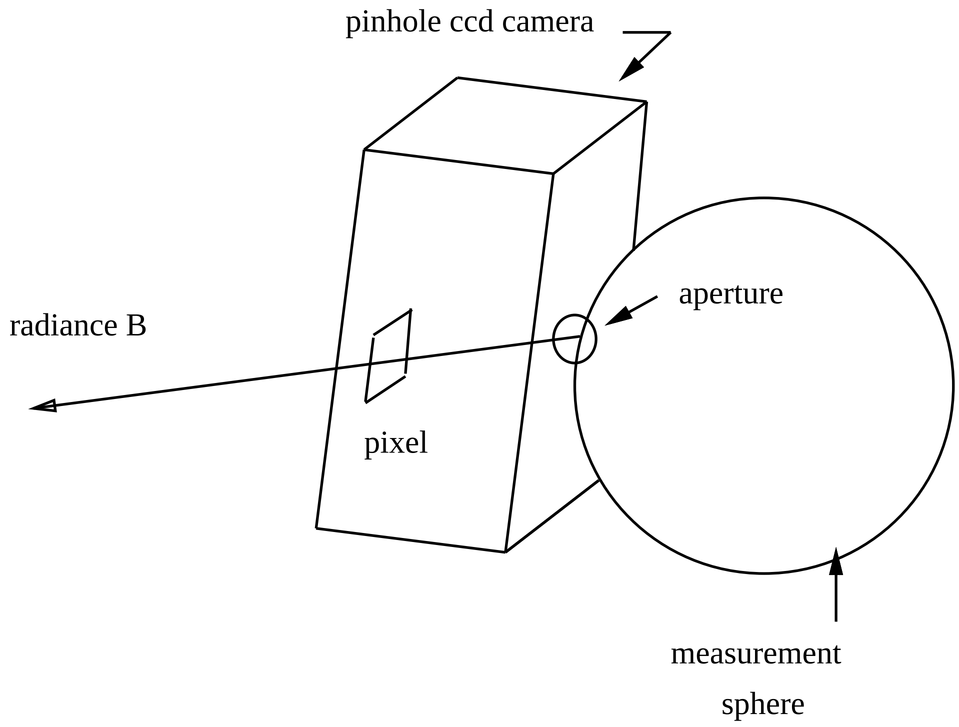

Figure 3-4: illumination of an aperture

section of the bi-Gaussian that "best-fits" the blob, derived analytically from a function of σ_x, σ_y, and σ_{xy} (see appendix III);

- *Higher Parameters* m_{ij}: coefficients of the Fourier expansion of the intensity distributions along and perpendicular to the tilt axis, using basis functions well known in quantum mechanics as "the wavefunctions of the harmonic oscillator" (they are not known by, e.g., a person's name), scaled by parameters σ_x and σ_y respectively.

The harmonic oscillator wavefunctions are an orthonormal set on the interval $-\infty$ to $+\infty$. They are products of a Gaussian with a Hermite polynomial scaled to the standard deviations of the bi-Gaussian that best-fits the blob. Appendix III reviews the mathematical details.

The image taken at camera pinhole position co-latitude θ and longitude ϕ thus has the intensity distribution

$$I[\theta,\phi;x,y] = \sum_{ij} m_{ij}[\theta,\phi]\,\Phi_i[x]\,\Phi_j[y]$$

where $[x,y]$ is a position, in units of σ_x and σ_y, with respect to the center-of-intensity $[x_{00},y_{00}]$, along axes tilted by a_{00} from horizontal.

We typically include terms in *i* and *j* from zero to six; higher order terms tend to "fit the noise" for our data rather than lend improved understanding of the underlying light distribution. For very regular spot quality filament images, a maximum index of four is often acceptable if regeneration speed is an issue.

The effectiveness of this parameterization is illustrated by figure 3-5, which is the image shown in figure 3-1 "regenerated" from its 2D parametric representation in the image plane. Note how the noise (consisting of the dots 1.3" to the right and ±0.25" up and down from the center of blob 3-1) has been "smoothed" away. The *rms* fitting error (over pixels) for the regenerated image (figure 3-5) is 1.6 grayscales over a dynamic range of 245; this is well within the typical CCD pixel non-uniformity of 10% (according to the manufacturer's spec sheet).

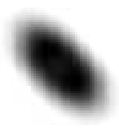

Figure 3-5: figure 3-1 regenerated from its 2D m_{ij} coefficients

3.3.2 Expanding the Expansion

By the preceeding method we obtain a set of coefficients for each picture taken over the sphere:

$$s[\theta,\phi] = \{m_{ij}, x_{00}, y_{00}, a_{00}, \sigma_x, \sigma_y\}[\theta,\phi] \quad \text{for } i,j = 0\ ...\ mmax$$

So far we have concentrated on each image individually: θ and φ are just labels that identify a particular image. Now it is useful to think about labelling the parts of the data set in another way: recognizing that each element s_k of set s is a smoothly-varying function of position (θ,φ) on the sphere, we can represent each of these elements by another set of coefficients $\{\alpha^k_{lm}\}$ times some convenient set of orthogonal basis functions $\Psi_{lm}[\theta,\phi]$:

$$s_k[\theta,\phi] = \sum_{lm} \alpha^k_{lm} \Psi_{lm}[\theta,\phi]$$

We obtain excellent and efficiently computable fits when we choose the 2-dimensional basis functions $\Psi_{lm}[\theta,\phi]$ to be products of 1-dimensional trigonometric Fourier series [TOL 62]:

$$\alpha^k_{lm}\Psi_{lm}[\theta,\phi] = a^k_{lm}\sin[l\theta]\sin[m\phi] + b^k_{lm}\sin[l\theta]\cos[m\phi]$$
$$+ c^k_{lm}\cos[l\theta]\sin[m\phi] + d^k_{lm}\cos[l\theta]\cos[m\phi]$$

The set of coefficients $\{\alpha^k_{lm}\}$ is shorthand for the larger set $\{a^k_{lm}, b^k_{lm}, c^k_{lm}, d^k_{lm}\}$. Spherical Harmonics would be inappropriate here because they lose their φ dependence at the poles, where the pictures still vary with φ. The coefficients are found by the standard Fourier integral on a discrete space.

The effectiveness of this parameterization is illustrated by figure 3-6, which is the image shown in figure 3-1 "regenerated" from its 4D parametric representation over the sphere. The *rms* fitting error (over pixels) for the regenerated image (figure 3-6) is 9.2 grayscales over a dynamic range of 245, again within the typical CCD pixel non-uniformity of 10%.

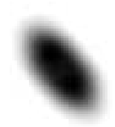

Figure 3-6: figure 3-1 regenerated from its 4D α_{lm}^{k} coefficients

Thus the dataset over the sphere is treated as if it were on a rectangular grid, and the incomplete data interval in θ from 0° to 135° is mapped to the range 0° to $360° - \Delta\theta\, 360/135$.[18] The number of θ coefficients $(2m+1)$ and ϕ coefficients $(2l+1)$ can be tuned to approximate the ratio of the number of θ data points to the number of ϕ data points.

For efficiency in regeneration, following an observation by Acton [ACT 90], the trigonometric functions of multiple angles should not be calculated as function calls, but from the three-term multiple angle recurrence relation. Thus, for example, after merely setting $\cos(0)$ to one and making a single function call to evaluate $\cos(\theta)$, all higher orders can be calculated recursively with a single multiplication and subtraction from the formula

$$\cos(n\alpha) = 2\cos(\alpha)\cos((n-1)\alpha) - \cos((n-2)\alpha)$$

[18]It is not mapped to 360° because the data may be different at $\theta' = 360°$ compared with $\theta' = 0°$, but the fitting functions at those two points are identical.

3.3.3 Compression Factor for Blob Images

Five blob-locating parameters and forty-nine m_{ij} coefficients were needed to represent the 196x165 pixel image of figure 3-1. Thus, the representation of a single image was successfully changed from 32,340 one-byte integers to 54 four-byte floats, for a compression ratio of about 150 to 1. For many (typically 425) images over a sphere, each of the 54 blob parameters can be stored by the $\{\alpha_{lm}^{k}\}$ with reasonable accuracy at compression ratios of five or ten to one. Therefore a compression ratio approaching three orders of magnitude can be obtained for the entire set of images over the measurement sphere.

3.3.4 Complete Representation and Regeneration

Assuming non-pathological behavior (which experimentally is the case), the intensity at any hypothetical pixel (x,y) in a pinhole camera located at any hypothetical point (θ,ϕ) on the measurement sphere can then be represented (and interpolated) via the two-step process:

$$s_k(\theta,\phi) = \sum_{lm} \alpha_{lm}^{k} \Psi_{lm}(\theta,\phi)$$

$$I(\theta,\phi;x,y) = \sum_{ij} m_{ij}(\theta,\phi)\,\Phi_i(x)\,\Phi_j(y)$$

where again (x,y) refers to a coordinate system described by the regenerated s_k parameters described in section 3.3.1.

With this represention entire pinhole camera pictures can be reconstructed at *and interpolated between* latitudes and longitudes at which pictures were actually taken. Figures 3-7 and 3-9 show the pictures (real data from figure 3-2) of the spot quality bulb filament taken at co-latitude $\theta = 60$ grads at two values of longitude ϕ separated by 16 grads. Figure 3-8 shows the picture computed (interpolated) midway between them via the coefficients $\{\alpha_{lm}^{k}\}$.

Figure 3-7: actual image at (60,-8) grads

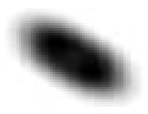

Figure 3-8: interpolated image at (60,0) grads

Figure 3-9: actual image at (60,8) grads

3.4 Light Transport

To transport light, the power incident on any distant surface patch can be found by sampling small solid angles in many directions from the surface, checking for intersection with the measurement sphere, regenerating the radiance of the source, and re-converting the radiance to incident power. This calculation is possible because the radiance is constant along lines, so the regenerated radiance emanating from the source is the same as that incident upon the receiving surface from any particular direction.

In detail, the procedure for finding the power incident on a small surface patch of area A_s at position $\vec{r}$ once the coefficients of the radiance distribution are known is to erect an imaginary pinhole camera C' whose aperture is located at $\vec{r}$ and whose normal is aligned parallel with the normal to A_s (figure 3-10). A ray R is traced from each imaginary pixel $[x',y']$ on the image plane of C' through the pinhole aperture and is tested for intersection with the measurement sphere of radius r_0 that now represents the light source, which involves nothing more complicated than finding the roots of a quadratic equation. If an intersection is made at the point (r_0,θ_0,ϕ_0) on the surface of the sphere, a second imaginary camera C_0 is constructed at this point with its aperture perpendicular to a radius and pointing towards the center of the sphere. Then the point (x_0,y_0) at which the ray R intersects the image plane of camera C_0 is easily found, and the radiance B along ray R in camera C_0 is regenerated as in sections 3.2 and 3.3.4.

The power incident upon $[x',y']$ is found by reconverting this radiance, because it is constant along lines (section 1.3):

$$I'[\vec{r};x',y'] = BA'_p A_s \cos^4\beta' / f'^2$$

where β' is the angle of incidence of the ray R with respect to A_s, f' is the image distance of camera C', and A'_p is the size of the sampling pixels. The total power entering C' is simply the sum of the power incident upon all of the pixels, or

$$I_T[\vec{r}] = \sum_{x',y'} I(x',y';\vec{r})$$

Finally, the total power per unit area striking the oriented surface A_s at point $\vec{r}$ is

$$I[A_s,\vec{r}] = I_T[\vec{r}] / A_s$$

The imaginary camera C' is simply a device, similar to the hemicube, that is used to perform

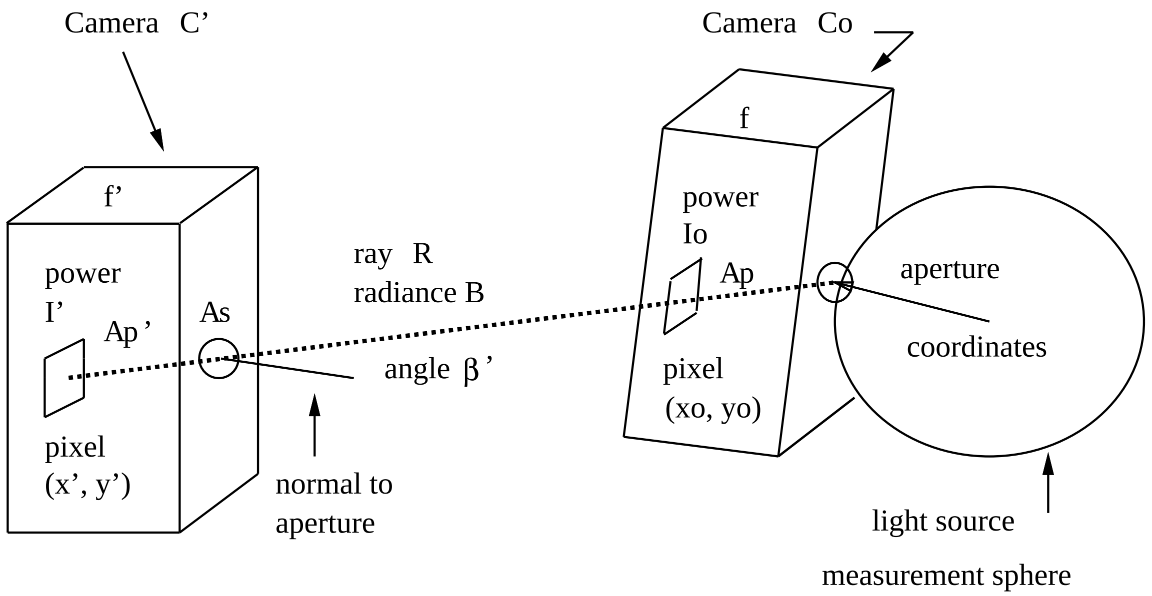

Figure 3-10: transporting light energy

numerical surface integration.

3.4.1 Efficient Sampling Limits

Knowledge of the size and location of the measurement sphere (or spheres), however, allows the sampling to be efficiently restricted to the geometrical projection of the measurement sphere onto the image plane of the sampling hemicube (figure 3-11).

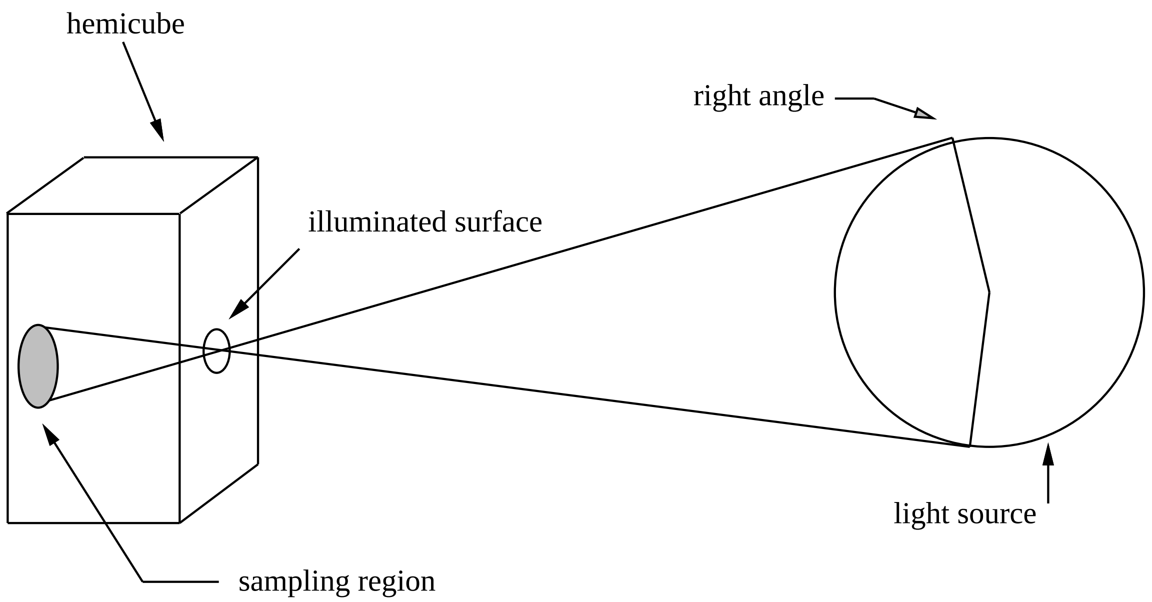

Figure 3-11: efficient sampling limits

3.4.2 Anti-Aliasing

The parameters A_p and f in the derivation of section 3.4 should be the same as those of the real camera C_0 which was used to collect the data. A'_p and f', which refer to the dimensions of the imaginary camera C' appear to be totally arbitrary, but they must actually be chosen carefully to avoid aliasing. The real camera measurements sampled patches of area of the source A_{src} on the order of

$$A_{src} = A_p d^2 / f^2$$

where d is the distance from the aperture on the surface of the measurement sphere to the actual patch of imaged source (by similar triangles taken through the pinhole camera's aperture). An estimate of the minimum d can be obtained from the experimental setup. The raytracing will be vulnerable to aliasing unless the imaginary camera C' samples the source on approximately the same scale; thus A'_p and f' are constrained by

$$A'_p / f'^2 \le A_p d^2 / (f^2 r^2)$$

where r is the distance from A_s (the aperture of C') to the source, and can be usually approximated as the distance from A_s to the center of the measurement sphere. Because of the inequality, it is safest to underestimate d and overestimate r.

3.5 Self-Occlusion

The occlusion of rays by light-emitting portions of the source can easily be checked within the framework of this model. If at some point on a surface (over which the radiance distribution is known) and in some direction the radiance is non-zero, then the line through that point, in that direction, *must intersect* a light-emitting part of the source with that radiance. Conversely, if at some point, in some direction, the radiance is zero, then the line through that point, in that direction, *does not intersect* a light-emitting part of the source. We assume the source is a simple emitter, and does not have a frosted or refractive glass envelope, for example.

As shown in figure 3-12, the exit position and direction coordinates of a ray B that intersects the measurement sphere surrounding the source can easily be calculated by solving a quadratic, and the source's radiance B' at these coordinates can be computed from the coefficients $\{\alpha^k_{lm}\}$. If B' is nonzero, then there must exist a light-emitting surface within the measurement sphere that lies in the path of the ray, thereby occluding that ray. If the radiance is zero, then the ray

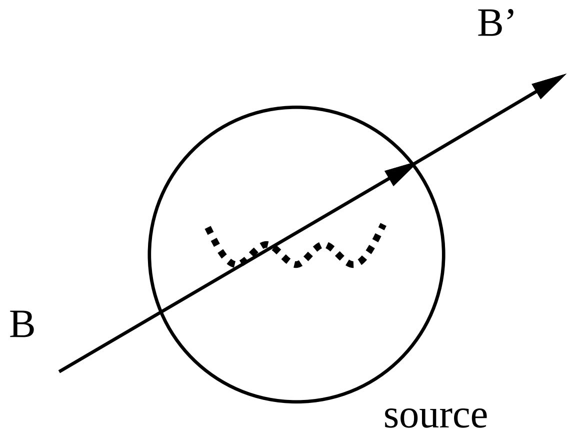

Figure 3-12: occlusion

will not be occluded by any light-emitting portion of the source.[19] The problem of intersecting, say, a complicated coiled filament to test for occlusion, is thus rendered trivial by the measurement sphere concept.

A series of pinhole camera measurements, with the filament off but backlit, may provide a way for describing the ray-blocking and transmitting portions of the source.

3.6 Generality of Light Source Model

The fact that this model is based on measurement gives it two great strengths. The first is its applicability to real light sources consisting of extended bodies with nonuniform radiance distributions; no assumptions of point-like or uniform behavior are necessary. The second is its complete generality; no restrictions on the nature of the light generation process (filaments vs discharges vs LEDs, etc.) are required. A comparison with other models will illustrate these points.

The Donahue and Joseph model [DON 75], for instance, considers a helical filament as a set of radiating points on the surface of a cylinder. The model is specificallly tailored to the problem

[19]"Dark" portions of the source (posts, baffles, etc) may still block such a ray, so in this model there is no way around explicitly testing for intersections with these structures. Often they are small, however, and can reasonably be handled via highly simplified geometrical models that are easy to intersect.

of designing a luminaire reflector for, e.g., an automobile headlamp. Filament points are assumed to be Lambertian radiators, with some reasonable distribution in brightness from point to point. No allowance is made for the exact geometrical details of the filament, or scattering from posts, the glass envelope, or other structures. Filament "see-in" and "see-through" effects are treated by heuristic approximations. Since the source is modelled as a set of points, only forward raytracing, in which rays are sent from these points, can be used. This results in constraints on the reflector model, forcing the reflector to consist of facets for easy intersection, and also leads to inefficient numerical surface integration (ray tracing). This light source and reflector model, though computationally simple, lacks generality because it is based on assumptions specific to filament sources of small-radius helical geometry. Within its domain of applicability, the Donahue and Joseph model has nevertheless had enormous practical importance and economic value for many years, becoming the industry standard.

Another model by Page [PAG 89] strives for generality, at least among incandescent filaments, by using principles of physics to balance power input and output. Unfortunately, the processes of heat radiation, conduction, and convection by real materials are often understood only in terms of empirical relationships. Furthermore, the equations of power balance turn out to be nonlinear, and can only be solved by successive approximations. Again, the assumptions made to simplify the problem depend on specific filament geometries and materials, resulting in a loss of general applicability. Furthermore, this model is not suitable for ray tracing because the filament geometry is too complicated to allow efficient or precise ray intersection calculations: where exactly is the filament?

Standard ray tracing methods used in computer graphics produce remarkable pictures, but have shortcomings when applied to real illumination system design, theoretical model checking, etc. The classic illumination model [WHI 80] takes liberties with real-world physics to simplify the computations, to achieve certain aesthetic effects, and to compensate for some artifacts of display technology. For instance, light sources are treated as lines for intensity calculations, the $1/r$ intensity dependence being chosen as a compromise between point-like and extended body behavior. The characteristic soft shadows cast by extended sources and other optical effects are typically added by hand during postprocessing (rather than calculated directly) or not simulated at all. To treat an extended light source with a nonuniform brightness distribution is possible within the standard computer graphics framework by dividing its surface into a large set of radiating patches. Such a light source, however, would be a purely fictional device for the sole purpose of rendering an imaginary scene; no prescription for decomposing an actual light source into such a model exists. The questions of what measurements would determine the subdivision and the radiating power of each patch remain unanswered. Even then, the resulting complex shape of the light source would be uneconomical both in terms of storage requirements and intersection calculations.

4. COMPUTER AUTOMATED REFLECTOR DESIGN

4.1 Introduction

Reflector design, for automotive headlamps for instance, is a time-consuming process that could benefit from computer assistance and automation. To this end, we describe a reflector model that can be used within a ray tracing framework, in conjunction with our previously discussed general near-zone light source model, to efficiently and accurately render an illumination pattern on a distant screen. The reflector surface is represented in an appropriate coordinate system as a paraboloid to which perturbations (with well-understood physical meanings) are added. We present an algorithm that swiftly calculates a first guess for the perturbations, and then we apply a search algorithm that performs an automated optimization to match a desired illumination pattern on a screen. The realistic rendering capabilities of the combined reflector and light source model also provide a convenient design tool for testing various reflector shapes.

The task of designing a reflector for a given light source to create a desired illumination pattern on a screen has attracted interest among luminaire designers in the automotive and architectural fields [DON 72, DON 75, MYO 82, ASH 94]. The application of a computer to this problem requires (1) a near-zone light source model, (2) a light transport method, (3) a reflector model, and (4) a figure of merit. With these four elements, the reflector design problem in principle becomes one of optimization, which can be solved by a straightforward search of the reflector model parameter space by, for example, steepest descent, conjugate variables, conjugate gradient, simulated annealing, or genetic algorithms.

Standard specular light transport relies on raytracing, which requires computationally intensive ray-surface intersection calculations for each of the potentially large number of optimization steps. Therefore previous formulations of the problem have compromised accuracy by simplifying the geometry of the light source, the reflector, or both. Furthermore, the use of light source and light tranport models that are applicable in the near-zone have only recently begun to appear [MIS 90, NGA 87, NGA 92, ASH 93]. Using the generally applicable measurement-based near-zone light source model of chapter 3, we formulate the problem in a way that is simultaneously computationally efficient and physically correct. The reflectors considered here are smooth paraboloids of revolution to which a few physically-based higher order perturbations are added; thus the number of terms to be optimized is greatly reduced compared with facetted mirror models. No restrictions are placed on the light source; real, extended sources can be correctly treated even in the near-zone.

We usually assume f (the focal length of the base-shape paraboloid reflector) is given, and do not allow it to vary (although it is possible to let the search algorithm perturb it if so desired). There are two main reasons for this approach, both related to the observation that f sets the scale of the problem. The first reason is that it is natural to measure all distances in units with $f = 1$; a change of f would suddenly change the meaning of all the other perturbations, resulting in a large jump in perturbation space and a less than systematic search. The second reason is that the size of the reflector is often constrained, and f is the single most important parameter in determining the overall size. It should be reasonable, therefore, to think of f as a given constant.

We describe an algorithm that finds a reasonable first guess to the solution, and then we use a modified version of conjugate variables, which has the best combination of speed and accuracy we have yet found for this problem, to carry out the optimization. Finally we present examples of automated reflector design using these tools. The light pattern of the computer-designed reflector can then be fine-tuned by a human designer by, for example, adding lenses or light baffles.

Section 4.2 describes the reflector model, and sections 4.3 and 4.4 show how the general near-zone light source model and light transport method of chapter 3 can be combined with this reflector model to quickly and accurately render the illumination pattern of a luminaire on a distant screen. Section 4.4.1 explains how the 2-dimensional Fourier components of an illumination pattern can be used to quantitatively characterize the similarity between two images. Section 4.5 builds on these results to outline a method for computer-automated optimization of the reflector surface to create a desired illumination pattern. Section 4.6 presents the results of computer-automated design attempts for a simulated source, while section 4.7 does the same using real source data. Appendix VI gives the formulas for ray-surface intersection calculations, and appendix VII provides the details of the algorithm that makes a "first guess" at the reflector surface shape.

4.2 Reflector Model

For illustration we present a relatively simple but nevertheless realistic parameter model of a reflector. The aim is to find values of the reflector parameters that, given the bulb coefficients, produce (within tolerance) a desired illumination pattern on, for example, a distant wall. The reflector is assumed to be perfectly specular (this assumption is used for simplicity but is not required). The base shape, to which higher order perturbations are sought, is a parabola of revolution aligned along the z axis with its focal point at the origin and its apex at $z=-1$. The reflector is therefore represented by a polynomial expansion of the form:

$$z=\frac{x^2+y^2}{4}-1+v+r(x^2+y^2)^{1/2}+l_1x+l_2y$$
$$+a_1(x^2-y^2)+a_2xy+c_1x^3+c_2y^3+c_3x^2y+c_4xy^2$$

Because x, y, and z are measured in units of f (the fixed focal length of the base shape parabolic reflector), the perturbation coefficients are dimensionless.

Some rule, based on design constraints, must be used to truncate the reflector. For this thesis we arbitrarily chose to use a rule that truncates the reflector above $z=0$. Other rules based on overall width, depth, or both may also of course be used, depending on the problem at hand.

These perturbations represent

$v \to$ vertical offset
$r \to$ conical term
$l_1, l_2 \to$ linear perturbations
$a_1, a_2 \to$ astigmatism terms
$c_1, c_2, c_3, c_4 \to$ cubic perturbations

These coefficients have understandable physical meanings. The offset, for example, is like the "object distance" in the Newtonian lens equation, and thus controls the convergence or divergence of the rays. The linear terms shift the paraboloid both horizontally and vertically, without actually changing its shape, causing a shift in the transverse position of the image.

[20] The three terms v, l_1, and l_2 thus control the horizontal and vertical position of the paraboloid. The terms r, c_1 and c_2 have effects qualitatively similar to the offset and linear terms respectively, but with attendant changes in reflector shape. Each mixed cubic term also acts like linear term, shifting the image, but along a pair of axes crossing at approximately 53°. The astigmatism term a_1 stretches or compresses the image along the x axis, and does the reverse to the image along the y axis. The other astigmatism term does the same, but along axes 45° to the coordinate axes of the image. Higher-order terms may be added as necessary, but the optimization will become slower and more local minima may appear.

The representation of the reflector surface as a polynomial expansion has several additional advantages, including enforcing smoothness and giving the surface normal for any (x,y) by an analytic derivative. The automated machining tools that would fabricate the mirror could therefore easily interpret this model (but terms like z^n would generally be ruled out). Furthermore, the smooth model, as opposed to the standard facetted approach, greatly reduces the number of terms to be optimized. It does, however, require foregoing purely local control over the light. The computer-generated mirror, however, can be used as a good starting point for an expert's "tweaking".

Section 5.2 mentions the possibility of using different base shapes and perturbations for creating illumination patterns that are very different from spotlight-type patterns.

4.3 Summary of Light Source Model and Transport Method

Our light source model and light transport method are described in detail in chapter 3. In summary, to fully characterize a source and properly transport light energy, the radiance distribution (power emitted per unit area, per outgoing directed solid angle) must be known over all space. Real light sources in the near-zone must be treated as extended bodies, so the light is not radial and thus, even when there is only a single source, the radiance at a fixed point in space will vary with the direction of view. The radiance distribution is therefore a scalar function of five variables: three for position in space and two for direction of view. Because radiance is a

[20] The use of linear perturbations to a paraboloid to aim the light has appeared in the literature [WAT 91].

constant along light rays [NIC 63, BOR 70, RIE 82], however, it is sufficient to know the radiance emanating in all directions from every point on a closed surface surrounding the source to fully characterize the remaining four degrees of freedom of the distribution (five variables and one constraint).

As previously explained in chapter 3, this distribution is learned by taking a series of pinhole CCD camera measurements in which the pinhole aperture positions lie on a sphere surrounding the source; the aperture sequentially samples points on the sphere and the CCD pixels sample (simultaneously and in parallel) numerous outgoing directions. The radiance is calculated from the optical power detected at a pixel and geometrical factors that depend on the position of the pixel with respect to the aperture.

A multidimensional Fourier analysis fits the measured radiance distribution to a set of coefficients $\{\alpha_{lm}^{k}\}$ times products of appropriate orthogonal functions. Assuming smooth variations in radiance, this approach allows natural prediction between measurement points by simply regenerating the radiance from the coefficients with the orthogonal functions evaluated at any coordinates of interest. Unlike standard interpolation techniques (*sinc* functions), the fit is not forced to pass through all of the measurement points. Thus localized noise spikes in the original data disappear in the reconstruction, becoming spread out over the entire dataset and appearing only as slightly increased experimental uncertainties in the values of the coefficients.

Light transport is easily accomplished using the constancy of radiance along light rays by backtracing rays toward the source, checking for intersection with the sphere on which the radiance measurements were taken, regenerating the outgoing radiance (if any), and converting radiance to power with simple geometrical factors (the possibility of self-occlusion is currently ignored). The measurement procedure transforms the complexly shaped, extended light source into an equivalent (nonuniform and nonradial) spherical radiator, reducing the intersection search to the algebraic solution of a quadratic. For points outside the measurement sphere, the calculation is physically correct even in the near-zone. The coefficient representation also provides a convenient data compression method of at least three orders of magnitude.

4.4 Image Rendering

The ray tracing space for this problem consists of three objects: the equivalent spherical source, the pseudo parabolic reflector, and the image plane. Because both the image plane and the source sphere are easy to intersect (algebraic calculation versus search), it is computationally feasible to represent the reflector surface $z(x,y)$ as a polynomial expansion as explained in section 4.2. Normally such an expression would be avoided within a raytracing framework because the intersection problem would require solving for the zeros of a high order polynomial; the standard solution is to treat the reflecting surface as a set of easily intersectible facets. With our spherical source model, however, we are free to parametrically step over the surface (by sampling points in x and y), and to cast specularly-related pairs of rays toward the easily-intersectible image plane and source sphere. Thus we use a combination of forward and backward raytracing, in which the entire reflector surface and the entire image plane are sampled.

The procedure for building up an image for a given light source is performed by parametrically stepping over the reflector surface by taking equal steps in x and y, and using the current coefficients to find the height $z(x,y)$ and normal $\boldsymbol{n}$ of each corresponding planar reflector element. From this reflector point, a ray $\boldsymbol{r}_1$ is cast to each pixel (x',y') on the image plane $z=D$. For each such ray, the corresponding specularly-reflected ray $\boldsymbol{r}_2$ is checked for intersection with the spherical light source (figure 4-1). Mathematical details are supplied in appendix VI.

For a set of reflector coefficients, the pseudocode for the rendering algorithm is:

- step over reflector plane
    - calculate reflector element position, orientation, and area
    - step over image plane
        - send ray from image to reflector
        - find specular ray S
        - if S intersects measurement sphere:
            - regenerate radiance along S
            - calculate image pixel power

To decrease the computation time, the rendering can be performed on a regular grid for only a fraction of the total image pixels (typically 1/4 or 1/9), with the remaining intensity values filled

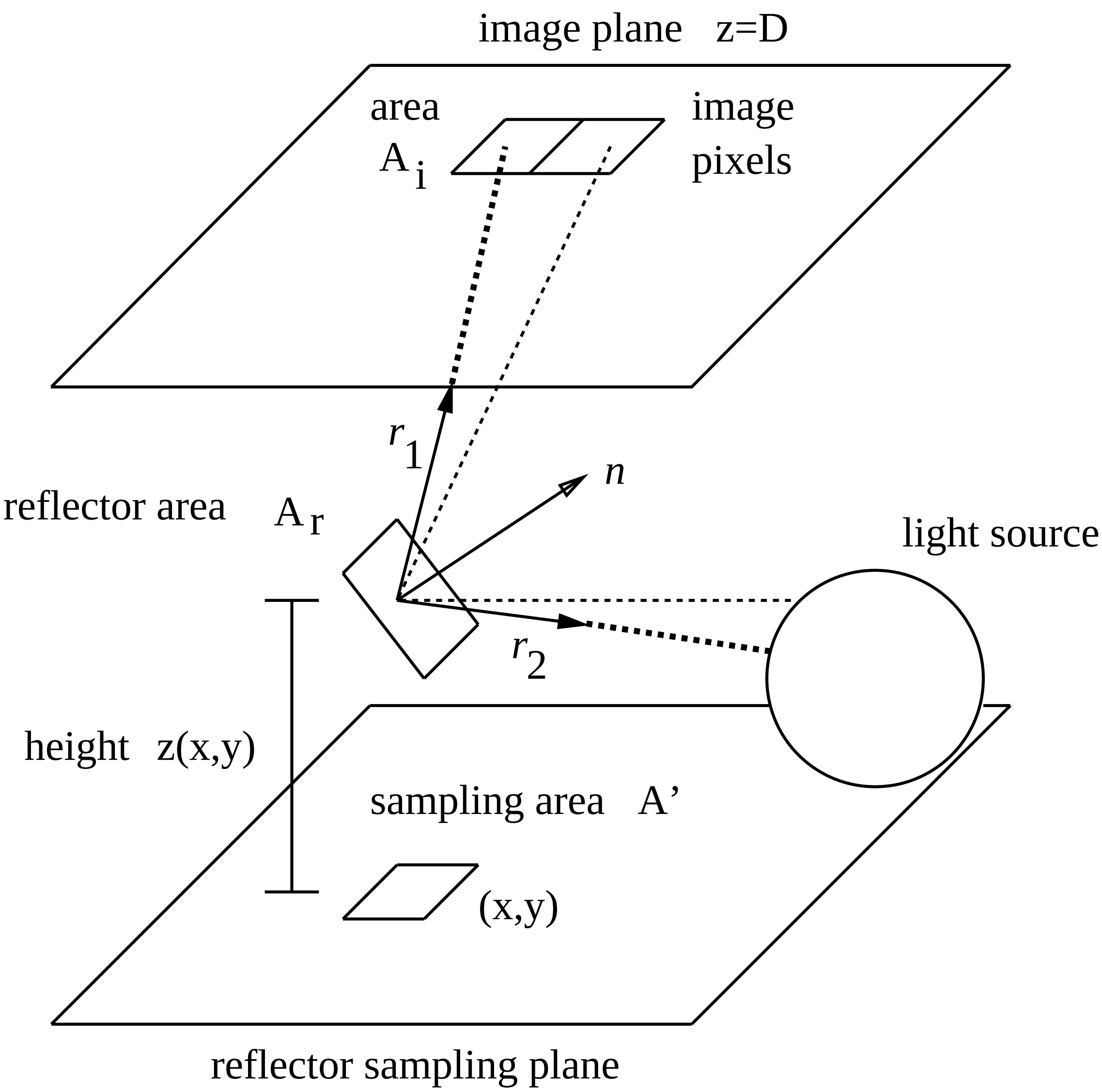

Figure 4-1: sampling the reflector surface

in by linear interpolation. Because the time required for the linear interpolation is negligible compared with actual rendering, the speed-up is often well worth the slight decrease in accuracy. As the final solution for the mirror coefficients is approached, the rendering density can be increased to fine-tune the coefficients, but this has not been found to appreciably change the solution.

4.4.1 Comparing Two Images

The resulting image can be built up pixel-by-pixel this way and stored as an array of intensity values. For compactness and generality, the image can alternatively be represented by its Fourier coefficients, for example, with harmonic oscillator wave functions used as the basis functions in an appropriate coordinate system, as described in section 3.3 and appendix III. This representation also provides a robust method of quantifying the "closeness" of two images.

The desired target image is supplied as a set of (typically 49) coefficients $\{m_{ij}\}$, plus five additional parameters to specify the coordinate system: two to locate the origin at the "center of intensity" of the image (x_{00},y_{00}), two to set the units of length in the x and y directions equal to the standard deviations of the image (σ_x,σ_y), and one to tilt the x-axis to correspond with the major axis of the elliptical cross-section of the bi-Gaussian which best fits the image (a_{00}).

For a given reflector, the Fourier coefficients $\{m'_{ij}\}$ of the rendered image can be calculated incrementally without the need for ever storing a large array of image intensity values. The coefficients are built up by

$$m'_{ij} = \sum_{xy} \sum_{x'y'} \Delta I'(x',y';x,y)\, \Phi_i(x'/\sigma_x)\, \Phi_j(y'/\sigma_y)$$

where Φ is a harmonic oscillator wave function, and the same coordinate system as the desired image is used. Here x and y refer to the reflector sampling plane, and x' and y' refer to the shifted and rotated coordinates of the image plane.

All optimization procedures essentially iteratively vary the reflector coefficients, calculate test images, and replace the old reflector coefficients with those that minimize some measure of "closeness" between the test image and the desired image. For two image arrays I_{xy} and I'_{xy}, the squared residual is defined as (dropping the primes on the image plane coordinates x and y):[21]

$$\Delta^2 = \int\int_{-\infty}^{+\infty} dxdy\,(I_{xy} - I'_{xy})^2$$

Thus it appears as if the desired image must be "unpacked" from its $\{m_{ij}\}$ representation for comparison with the current image, which is a memory intensive and time-consuming process.

An alternative definition of Δ^2 is possible, however. If the image I' is also characterized by coefficients $\{m'_{ij}\}$, using the same coordinate system parameters as the image I, then orthogonality among the basis functions results in

$$\begin{aligned}\Delta^2 &= \int\int_{-\infty}^{+\infty} dxdy\,\{\sum_{ij} m_{ij}\,\Phi_i(x/\sigma_x)\,\Phi_j(y/\sigma_y) - \sum_{ij} m'_{ij}\,\Phi_i(x/\sigma_x)\,\Phi_j(y/\sigma_y)\}^2 \\ &= \sigma_x\sigma_y \int\int_{-\infty}^{+\infty} dxdy\,\{\sum_{ij}(m_{ij} - m'_{ij})\,\Phi_i(x)\,\Phi_j(y)\}^2 \\ &= \sigma_x\sigma_y \sum_{ijkl}(m_{ij} - m'_{ij})(m_{kl} - m'_{kl}) \int_{-\infty}^{+\infty} dx\,\Phi_i(x)\,\Phi_k(x) \int_{-\infty}^{+\infty} dy\,\Phi_j(y)\,\Phi_l(y) \\ &= \sigma_x\sigma_y \sum_{ijkl}(m_{ij} - m'_{ij})(m_{kl} - m'_{kl})\,\delta_{ik}\,\delta_{jl}\end{aligned}$$

leaving

$$\Delta^2 = \sigma_x\sigma_y \sum_{ij}(m_{ij} - m'_{ij})^2$$

Dividing by $\sigma_x\sigma_y$, a measure of the characteristic size of the image, and taking the square root, results in the normalized "root mean square" quantity

$$R = (\sum_{ij}(m_{ij} - m'_{ij})^2)^{1/2}$$

Because the size has been divided out, R can be used as a standard, size-independent characterization of the closeness between two images. We have found this coefficient-based method to be more numerically robust than the pixel-based method.

Once the Fourier coefficients are found, the work required to compare two compressed images is greatly reduced; for a 200x200 pixel image represented by 49 coefficients, for example, the number of squaring operations required to calculate R is decreased by a factor of 800. The

[21]Because the filament image is completely contained within the camera's field of view, the intensity outside of the camera's boundaries can be set to zero, allowing the integral to proceed to infinity.

Fourier representation also has the desireable property of spreading out any noise (or atypical intensity values) localized in individual pixels over the entire image.

4.5 Automated Reflector Design

Working within this new model, we develop an algorithm for automated reflector design. The user supplies the position of the measurement sphere for a light source characterized by the α_{lm}^{k} coefficients, a desired illumination pattern on a screen a distance D from the light source, and a base focal length f for the initially parabolic mirror. The origin of the reflector's xyz coordinate system is chosen to coincide with the center of the light source's measurement sphere.

A judicious choice of coordinate axes with which to represent the reflector can minimize the final values of the perturbation coefficients and hence reduce the solution time. For instance, the linear perturbation terms to the reflector base shape which are responsible for pointing the image can be minimized by requiring the z axis to tilt from the vertical in such a way as to pass through the "center of intensity" of the desired image. Furthermore, the astigmatism terms can be "diagonalized" if the x and y axes are rotated such that their projections onto the image plane in the direction of the z axis are parallel to the major and minor axes of the ellipse that best fits the desired image; in this way the "cross-astigmatism" term a_2 will become small.

The x and y reflector axes are first rotated about the vertical by a_{00}, and then appropriate rotations of the coordinate system about these two axes are made to align the z axis of the paraboloid with the point (x_{00},y_{00}) on the image plane, which describes the location of the center-of-intensity of the desired image (see section 3.3 and appendix III).

For any given set of reflector coefficients, an image can be rendered. The "closeness" of this image to the desired image can then be measured. This measure of closeness R (section 4.4.1) is then a numerical function of the reflector coefficients. The problem has now been cast as a straightforward exercise in numerical optimization [POW 70].

Any reasonable method that samples the reflector parameter space (such as steepest descent, simulated annealing, multidimensional binary search, or genetic algorithms) will suffice. All optimization/minimization methods, however, must overcome two major obstacles: the presence of narrow, twisting ravines (or ridges), and local minima, especially for residual functions such as this one which consist of several parameters and contain no *a priori* derivative information. The lack of analytic knowledge of the slope of the residual function unfortunately excludes many of the most efficient minimization algorithms.

Simulated annealing robustly deals with these problems but can be time consuming. The gradient-based steepest descent method often overshoots the minimum or misses twists in ravines, and convergence is slow near a minimum; furthermore, derivatives are required. The multidimensional binary search [ACT 90], or "star search", is single-minded but inefficient. Ashdown has described the application of genetic algorithms to this problem, using Bezier and B-splines surfaces to model the reflector surface [ASH 94]. We choose to use Brent's modification [BRE 71, BRE 73] to Powell's improvement [POW 64] of the "conjugate variables" approach.

The conjugate variables method is essentially a multidimensional binary search without derivatives that is made more efficient by adaptively changing the search axes in order to be sure of finding the minimum of a quadratic form in a finite number of steps. Since all functions approach a quadratic form near a minimum, convergence should be rapid if a good initial guess is available. The basic procedure for iteration k is:

1. Let n search directions $\boldsymbol{v}_i^k$ and a starting point $\boldsymbol{x}_0^k$ in reflector coefficient space be given.
2. For $i = 1, \ldots, n$, find λ_i to minimize the residual function $R(\boldsymbol{x}_{i-1}^k + \lambda_i \boldsymbol{v}_i^k)$ described in section 4.4.1, and define $\boldsymbol{x}_i^k = \boldsymbol{x}_{i-1}^k + \lambda_i \boldsymbol{v}_i^k$.
3. Minimize from point $\boldsymbol{x}_n^k$ in the direction of $\boldsymbol{v}_{n+1}^k = \boldsymbol{x}_n^k - \boldsymbol{x}_0^k$, and call the resulting point $\boldsymbol{x}_0^{k+1}$.
4. Define a new set of search directions so $\boldsymbol{v}_i^{k+1} = \boldsymbol{v}_{i+1}^k$. The last k directions of the set can be shown to be mutually conjugate.
5. Increment k and go to 1.

Search directions $\boldsymbol{p}$ and $\boldsymbol{q}$ are said to be conjugate with respect to the quadratic surface

$$f(\boldsymbol{x}) = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b} \cdot \boldsymbol{x} + c$$

if

$$\boldsymbol{p}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{q} = 0$$

For n variables a complete set of conjugate search directions can theoretically be built up after n full iterations. It can be proved [POW 64] that if the current best solution for the minimum is close enough to the true local minimum such that the quadratic approximation to the residual surface is valid, then the local minimum will be found by simply minimizing exactly once along each of the conjugate search directions in any order. Convergence (within tolerance) for a quadratic form (or near a minimum) is therefore linear in the number of search parameters.

Powell's improvement for efficiently choosing the search directions, based on numerical realities, is that the actual search direction to be replaced, if any, is the one that will maximize the determinant of the matrix $\boldsymbol{V}$ of search directions $\boldsymbol{v}_i$. This prevents the search directions from becoming too linearly dependent, reducing the efficiency of the search or restricting it to a subspace. Some [FLE 65, ZAN 67] have suggested that Powell's method is too stringent, because the replacement of a conjugate direction will slow the convergence. Instead, they propose that the determinant of search directions merely be kept above some minimum value. Section 5.2 contains a suggestion of mine concerning which search direction to replace.

Brent discusses a variety of further modifications. For instance, the search direction to be replaced is restricted to the set of non-conjugate directions. Thus, there is some choice to allow the determinant of $\boldsymbol{V}$ to remain high, but conjugate directions are not thrown away. Next, once a complete set has been found, to avoid having the directions eventually become dependent, $\boldsymbol{V}$ is periodically reset to some orthogonal matrix. The standard choice is to simply start again with the identity matrix, but in order to preserve information about the surface, however, Brent resets $\boldsymbol{V}$ to the matrix whose columns are the principal axes of the quadratic form that best fits the residual surface at the current point. In this way the search directions remain conjugate. Scaling of the search directions is also periodically performed to reduce the condition number of $\boldsymbol{V}$ to continue to fight linear independence. A random step is also incorporated if a valley is suspected; the minimization from a nearby random point should find another point in the valley and then a search in the direction from one of these minima to the other should allow quick traversal.

Brent also describes a halting criterion. The search stops whenever

$$|\boldsymbol{x}_0^{k+1} - \boldsymbol{x}_0^k| \leq \varepsilon^{1/2} |\boldsymbol{x}_0^{k+1}| + T$$

holds for some number of iterations (typically one or two) where ε is the machine epsilon and T a positive tolerance.

Although conjugate variables provides a robust multidimensional minimization method when no derivatives are available, local rather than global minima can still be found. We deal with this remaining problem in several ways. The first is the realization that a true global minimum is not actually required; all we need is a minimum that is good enough to satisfy the design tolerance. Another strategy is to use more than one starting point in parameter space. Choosing an initial set of reflector coefficients that is close to the global minimum (or a "good enough" minimum) can also help; in appendix VII we therefore present a "first guess" algorithm that chooses a reasonable set of starting parameters based on the shape of the perimeter of the desired pattern. The problems of efficiently searching a tortuous, numerically determined surface have thus been addressed.

The situation of requiring a "good" minimum but not the exact mirror solution is analogous to the problem of finding $\boldsymbol{x}$ to solve the operator (or matrix) equation $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, where $\boldsymbol{x}$ represents the reflector, $\boldsymbol{b}$ the desired illumination pattern, and $\boldsymbol{A}$ the light transport operator. Let $\boldsymbol{x}_0$ be the exact solution, and $\boldsymbol{x}_1$ the solution found by the search. We require $\boldsymbol{A}\boldsymbol{x}_1$ to be close to $\boldsymbol{b}$, but we do not require $\boldsymbol{x}_1$ to be close to $\boldsymbol{x}_0$. In the realm of linear algebra those two conditions are only equivalent if the matrix representation of $\boldsymbol{A}$ is well-conditioned; Gaussian elimination, for example, is guaranteed to produce a low residual fit but not necessarily the "right" solution vector.[22] In fact, we should not even expect that an exact reflector solution exists at all in our model for an arbitrary desired illumination pattern.

---

[22]Note that our goal here is different from that of chapter 2, in which it was essential to not only produce a low residual fit, but also to keep $\boldsymbol{x}_1$ "close" to $\boldsymbol{x}_0$ in order to preserve the physical meaning of the d_l.

4.6 Results - Simulated Source

Reflectors were designed by this process using a 15x15 mirror plane grid with a step size of 0.4 *f*, and a 51x51 image plane grid with a step size of 1.0 *f*. These numbers were chosen to avoid visually obvious sampling errors in the images. The distance to the image plane was arbitrarily taken as 320 *f*, and the light source radius as 0.5 *f*. In order to avoid introducing artifacts due to light source anisotropy, and to more easily understand the results, we used a simplified synthetic light source with the property that any radial pinhole camera picture on the measurement sphere would result in a bi-Gaussian intensity distribution (i.e. all of the m_{ij} were set to zero except for m_{00}). Coincidentally, to the degree that a bi-Gaussian approximates cosine^4, the light source approximates a Lambertian radiating sphere.

Images (to become the target "desired" images) were created using this synthetic light source model so that the desired image size and intensity would be reasonable in terms of energy conservation for this source. The mirror used for creating the image was a vertical paraboloid with randomly chosen perturbation terms.

When numerically searching for a good minimum in the presence of several local minima, various starting points should be used for the search. Thus, each example will show the results for two different starting points: an unperturbed paraboloid that is automatically oriented to point at the center of intensity of the image as described in section 4.5, and a paraboloid oriented the same way but with the results of the "first guess" algorithm of appendix VII used as initial perturbations.

As a general rule of thumb, I classify solutions with final residuals less than 10.0 as absolute successes, between 10.0 and 20.0 as marginally satsifactory, and greater than 20.0 as clear failures.

For the ten standard perturbation coefficients of section 4.2, each iteration (minimizing in each of the coordinate directions) required approximately three minutes on a Silicon Graphics Indigo workstation running a 150 MHZ IP20 processor with a MIPS R4400 CPU, and included 20-25 renderings. The rendering was done on a regular grid containing 1/9 of the pixels, with the rest filled in by linear interpolation.

4.6.1 A Simple Example

A desired image was created using only the two astigmatism perturbations $a_1 = -2.0 \times 10^{-4}$ and $a_2 = -2.0 \times 10^{-4}$. The automated alignment procedure tries to find a solution in a coordinate system with its *x*-axis aligned with the major axis of the ellipse that best fits the image, so we should expect that no exact solution will exist within the reflector model in this rotated coordinate system. We should expect, however, that the solution will have a small value for the cross-astigmatism a_2. This is indeed the case, with the most successful solution returning values of $a_1 = -2.32 \times 10^{-4}$, $a_2 = 4.1 \times 10^{-5}$, and zero for all of the other terms.

A summary of the residual for each of the two starting points is shown in figure 4-2. Without the guess the residual began at 26.9 and ended at a respectable 6.0, but with the guess the initial residual was 12.1 and decreased to 2.4 before halting.

Figure 4-3 depicts the desired image, figure 4-4 displays the initial image of the first starting point (tilted paraboloid only), figure 4-5 shows the initial image of the second starting point (tilted paraboloid with first guess algorithm), and figure 4-6 shows the best solution found. Figure 4-7 shows a plot of the absolute difference between the desired and best-fit images. Figures 4-8 and 4-9 show the desired and derived images side by side for comparison. Each contour line in this section represents a change in intensity of five greyscales (out of 256), except for figure 4-7 where it represents one greyscale.

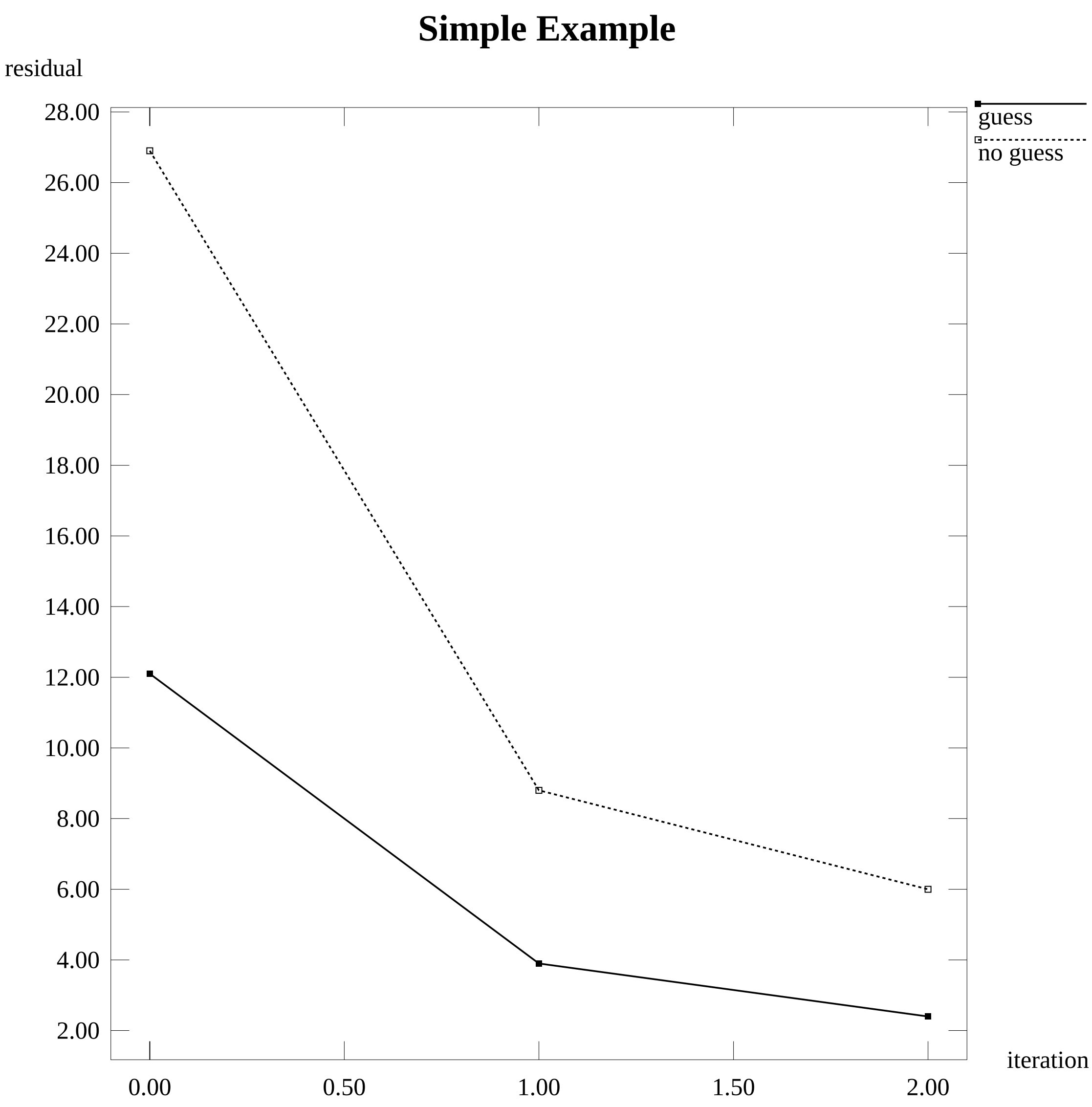

Figure 4-2: residual summary

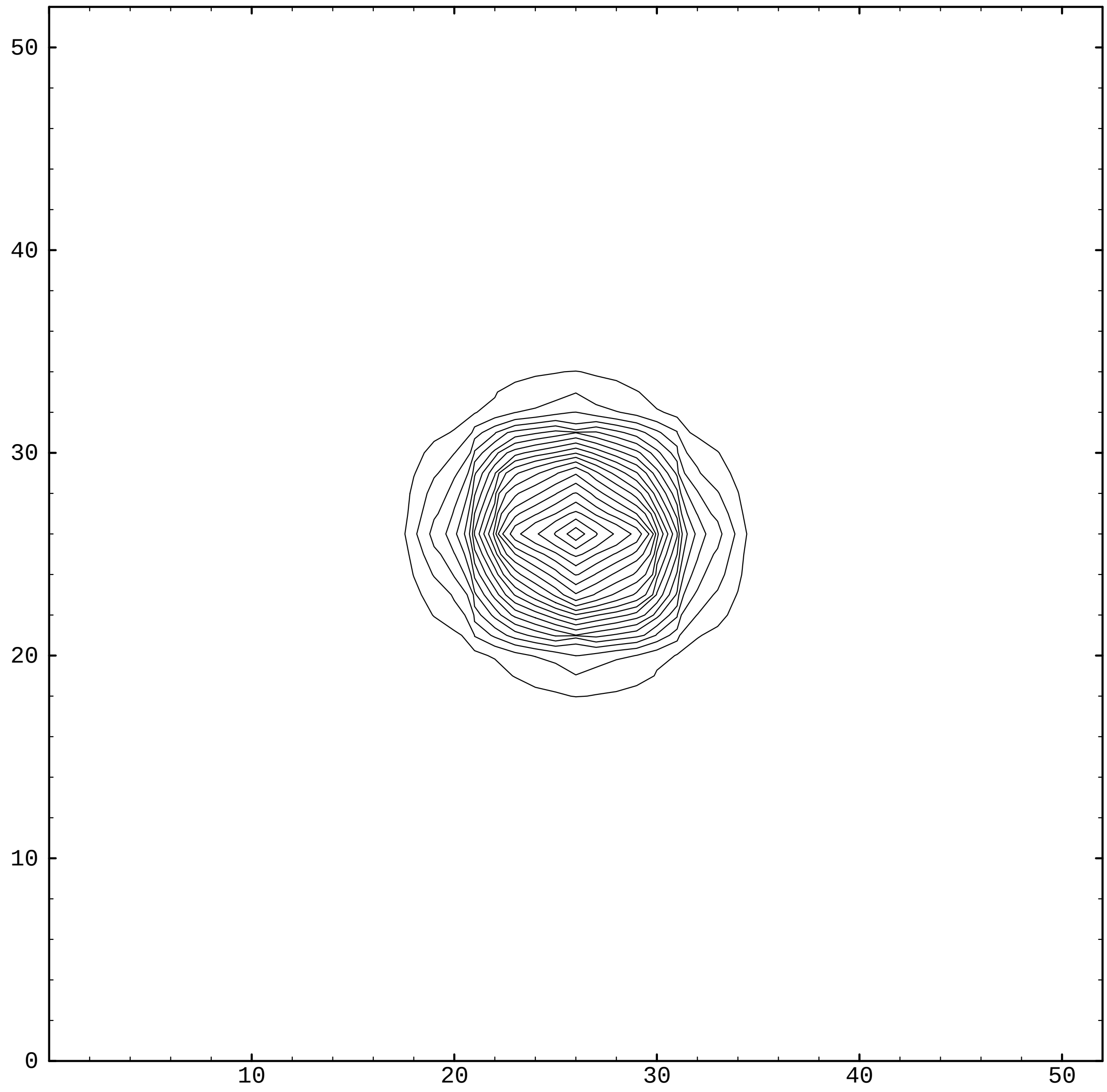

Figure 4-3: desired image (simple example)

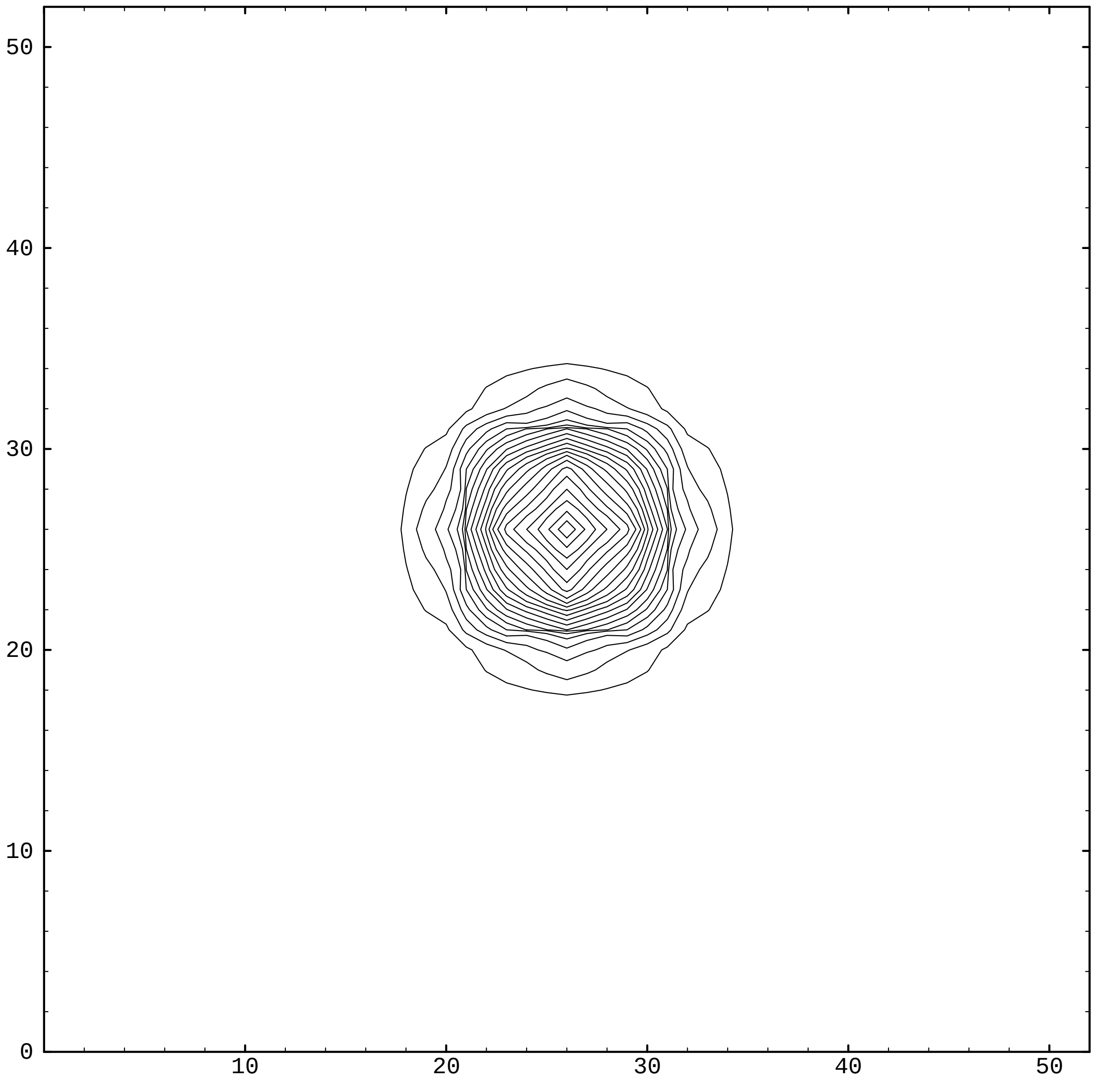

Figure 4-4: initial image (simple example)

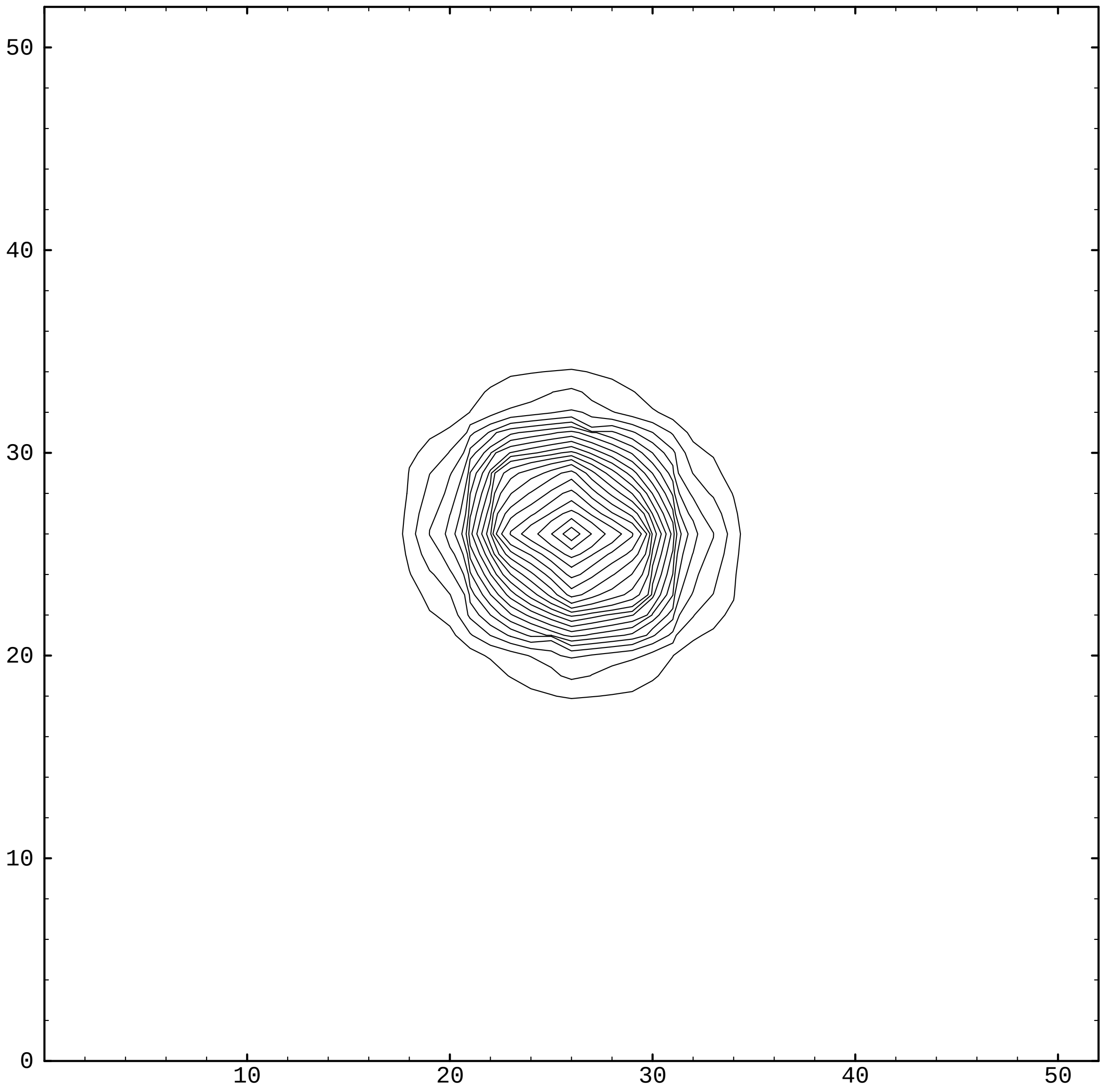

Figure 4-5: initial image, guess algorithm (simple example)

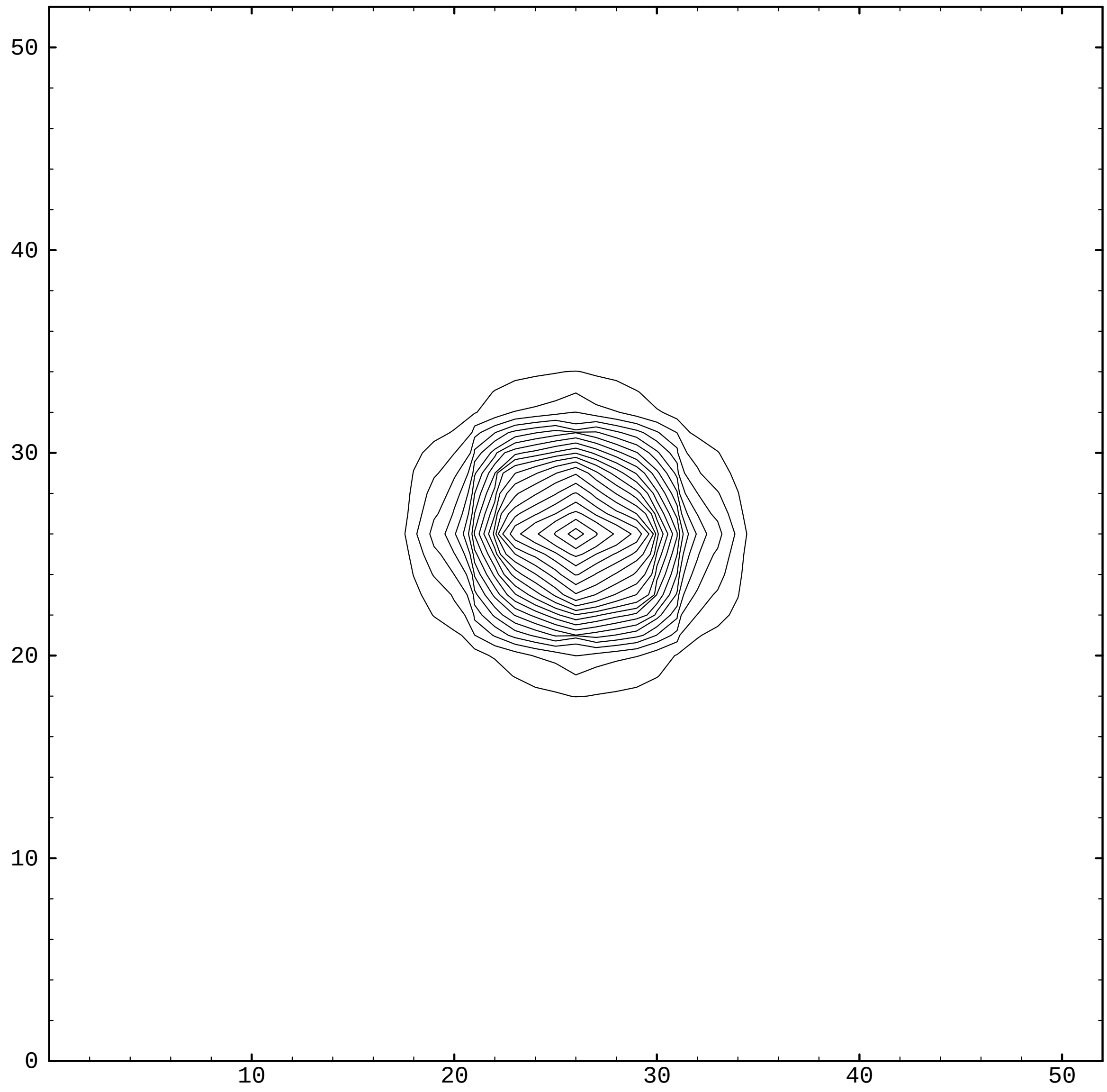

Figure 4-6: best solution, guess algorithm (simple example)

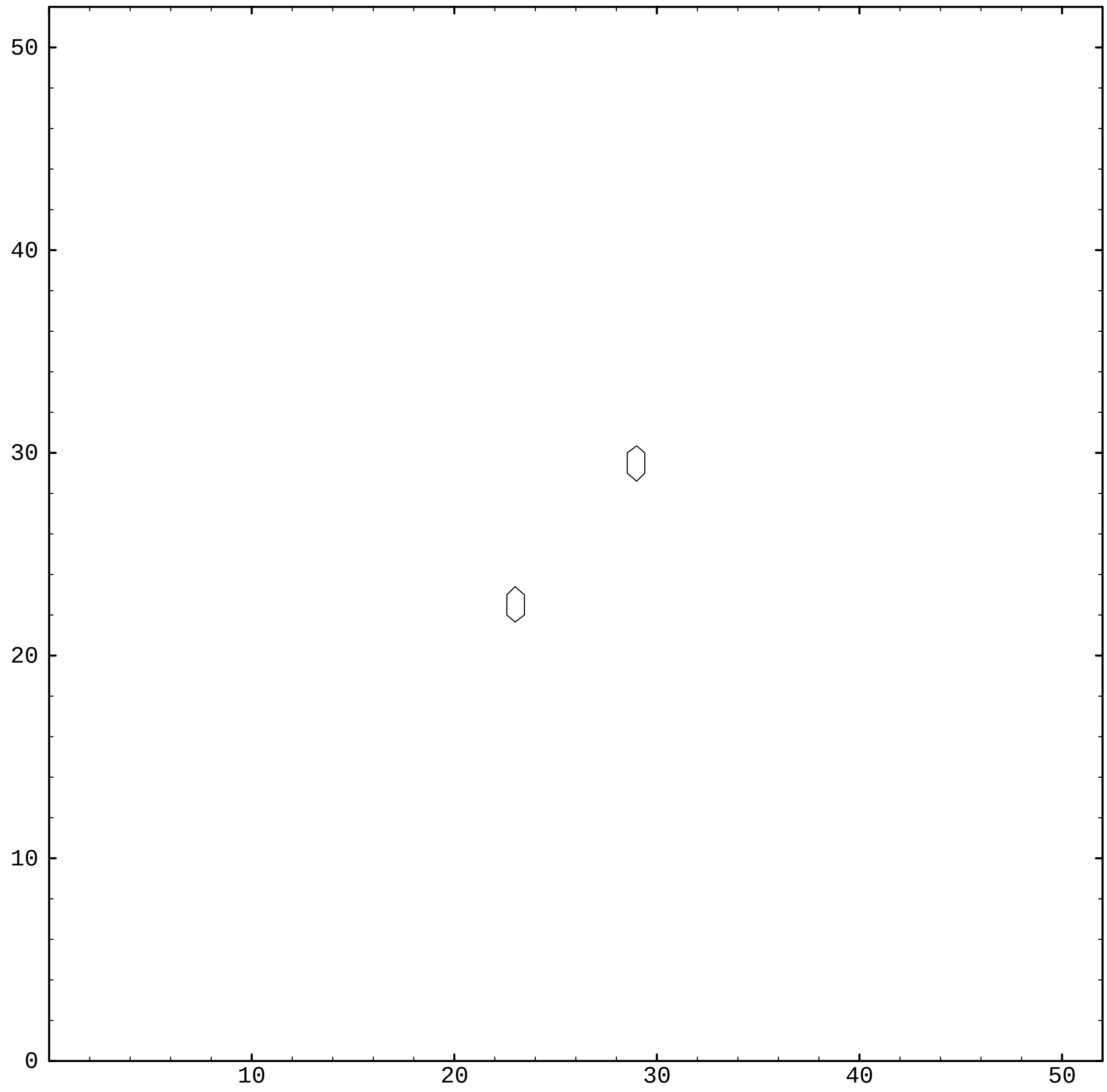

Figure 4-7: absolute difference plot (simple example)

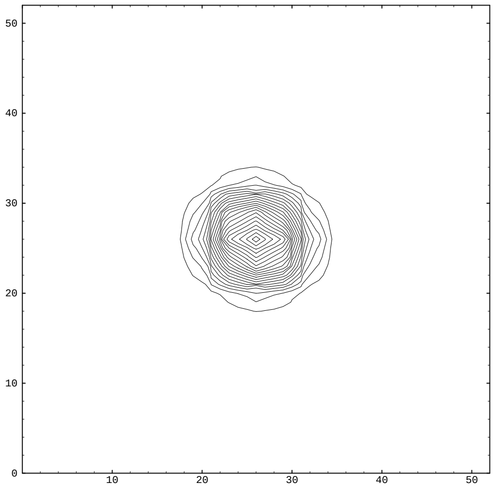

Figure 4-8: desired image (simple example)

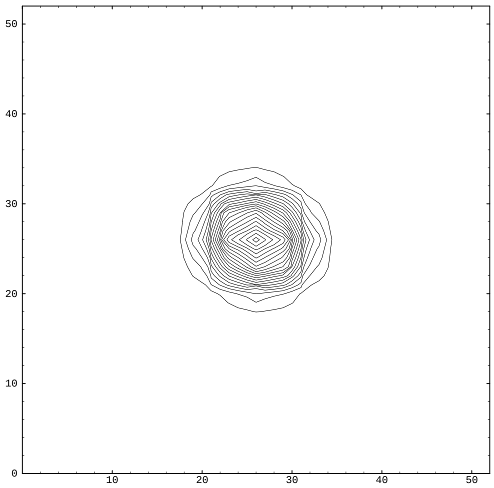

Figure 4-9: best solution, guess algorithm (simple example)

4.6.2 A Twin-Peaked Example

A more interesting image, consisting of two high-intensity regions, was created using ten perturbations. The resulting image was then rotated by 80 grads to produce the desired image. I consider this to be a difficult example because a double intensity peak is hard to construct within the parameters of this spotlight-type reflector model, especially with small perturbations, and furthermore the first guess algorithm merely attempts to match the outside border of the image and makes no allowance for strange internal intensity variations. We might expect, therefore, that the first guess, operating outside of its assumptions, will not perform satisfactorily.

A summary of the residual for each of the two starting points is shown in figure 4-10. Without the guess the residual began at 161.2 and ended at a successful 7.3, and with the guess the initial residual was 110.1 but only decreased to 10.2.

Figure 4-11 depicts the desired image, figure 4-12 displays the initial image of the first starting point (tilted paraboloid only), figure 4-13 shows the initial image of the second starting point (tilted paraboloid with first guess algorithm), and figure 4-14 shows the best solution found. Figure 4-15 shows a plot of the absolute difference between the desired and best-fit images. Figures 4-16 and 4-17 show the desired and derived images side by side for comparison. Each contour line in this section represents a change in intensity of 7.5 greyscales, except for figure 4-15 where it represents 3.5 greyscales.

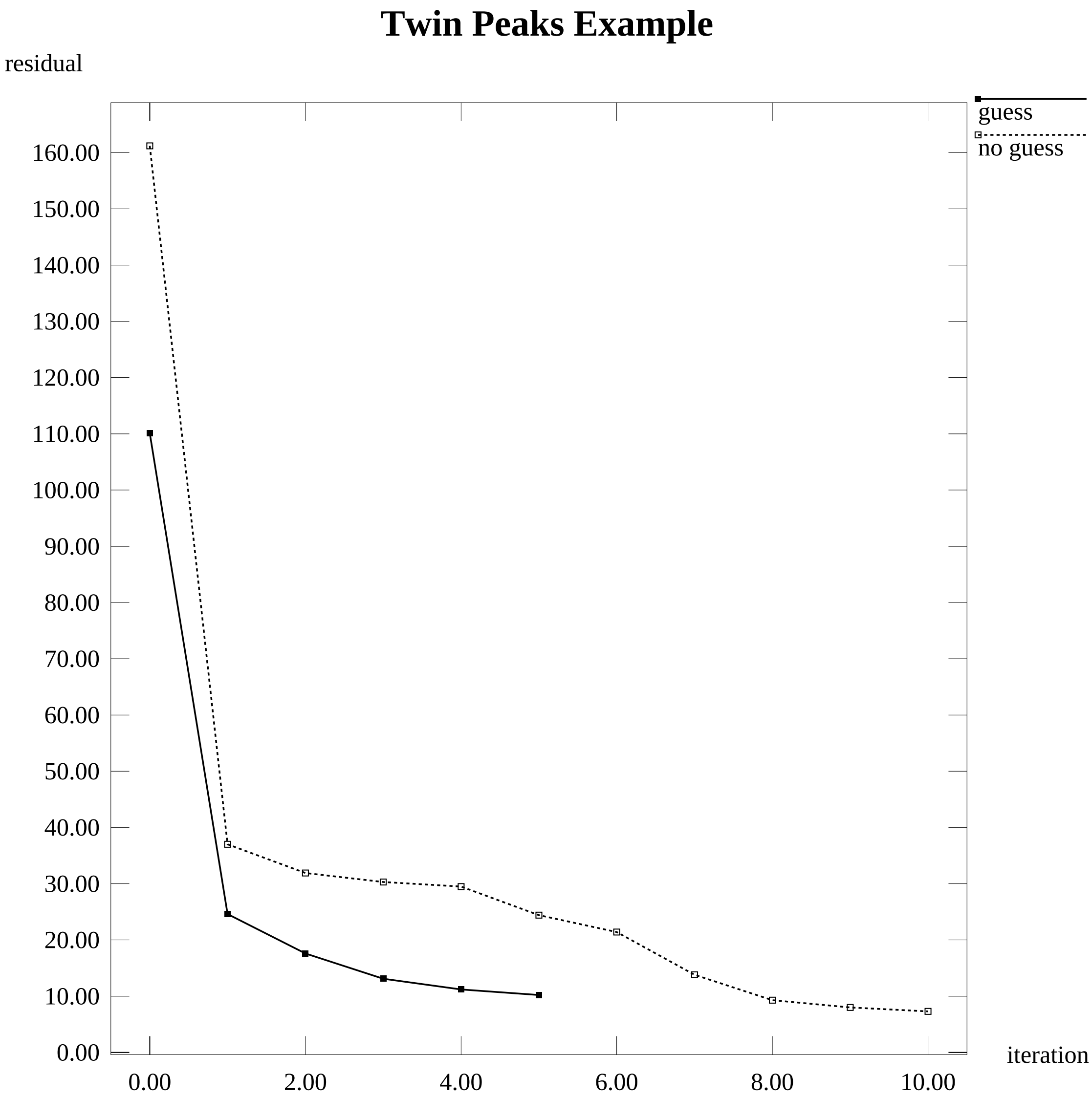

Figure 4-10: residual summary

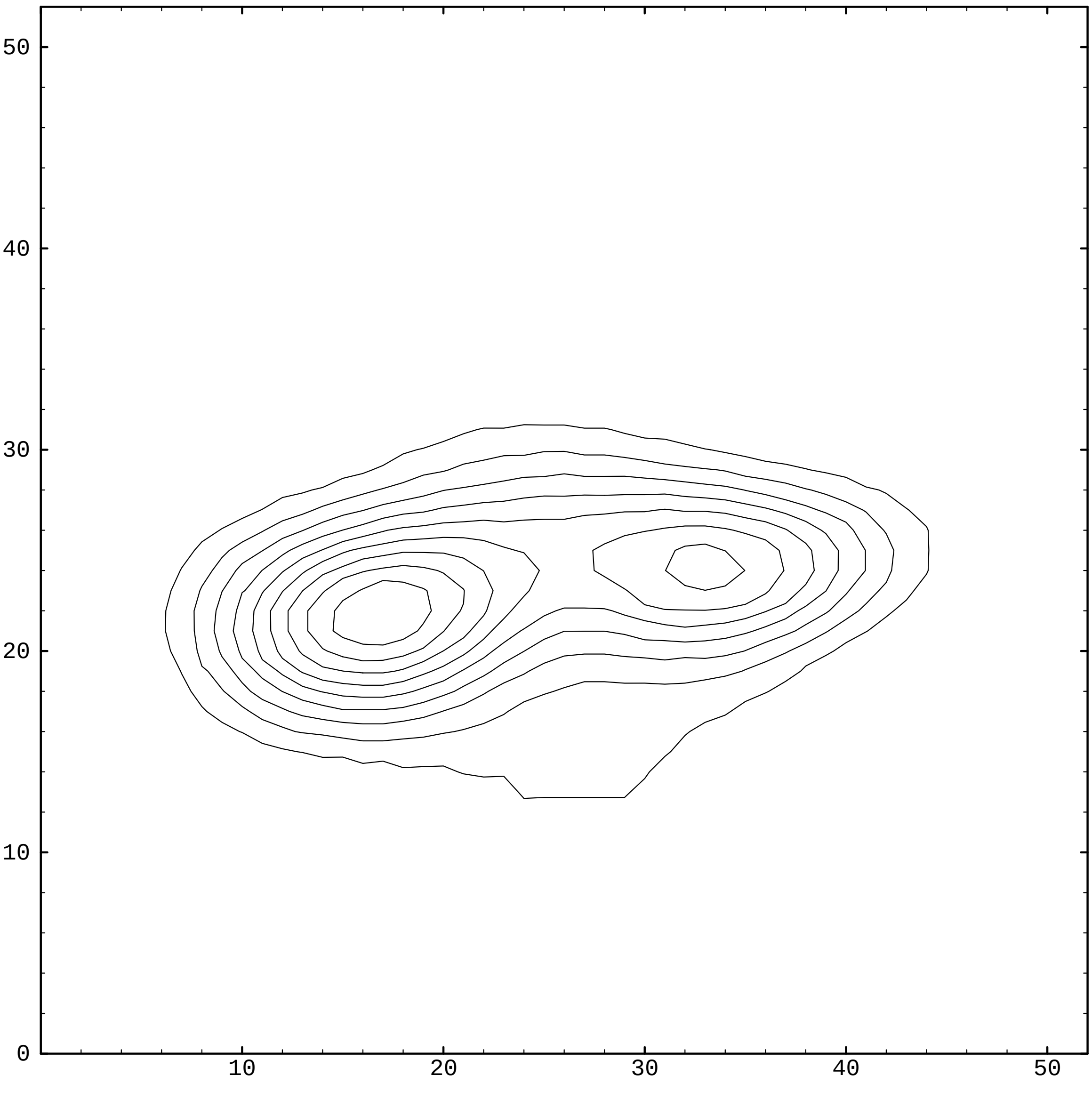

Figure 4-11: desired image (twin-peaked example)

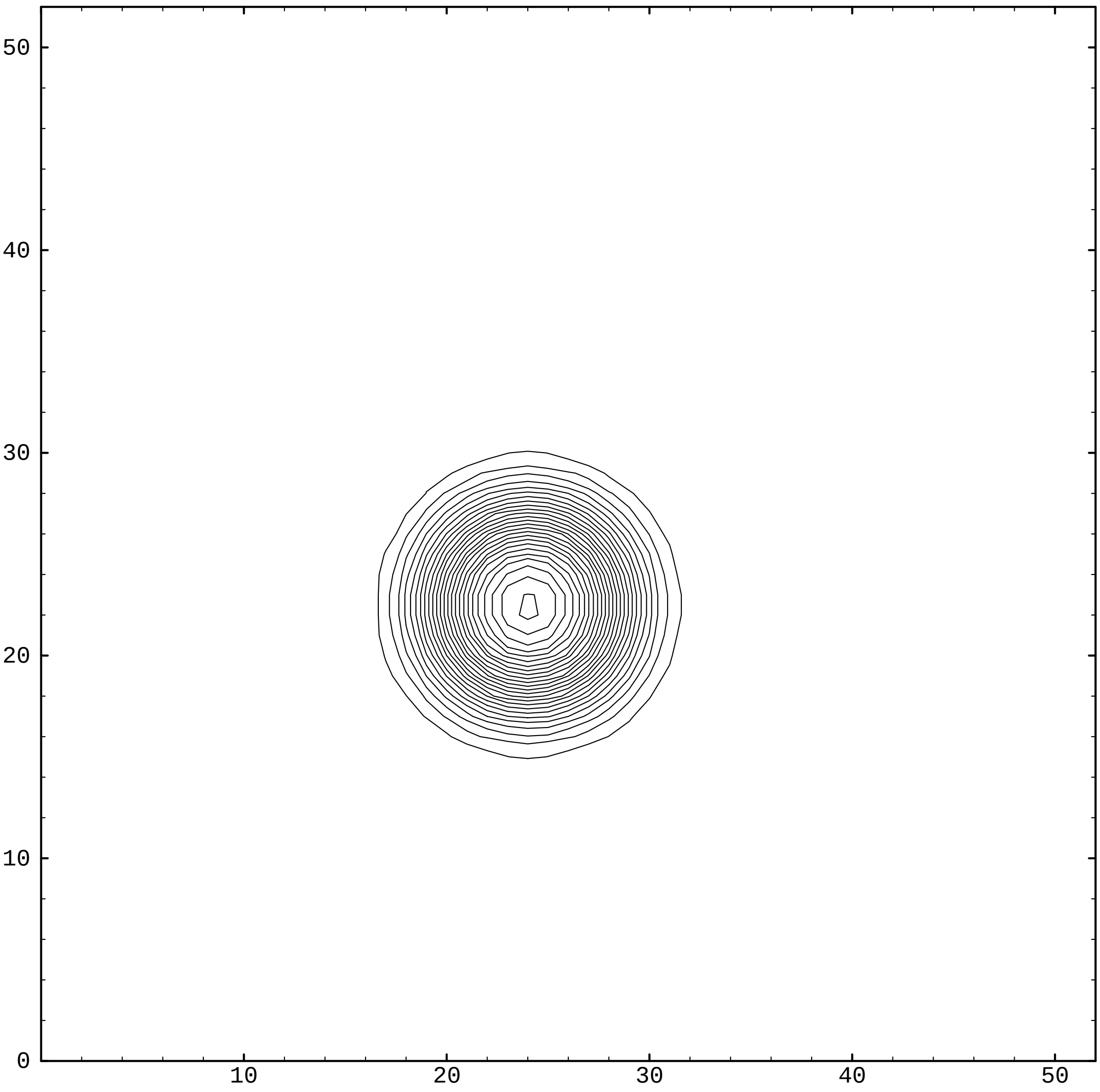

Figure 4-12: initial image (twin-peaked example)

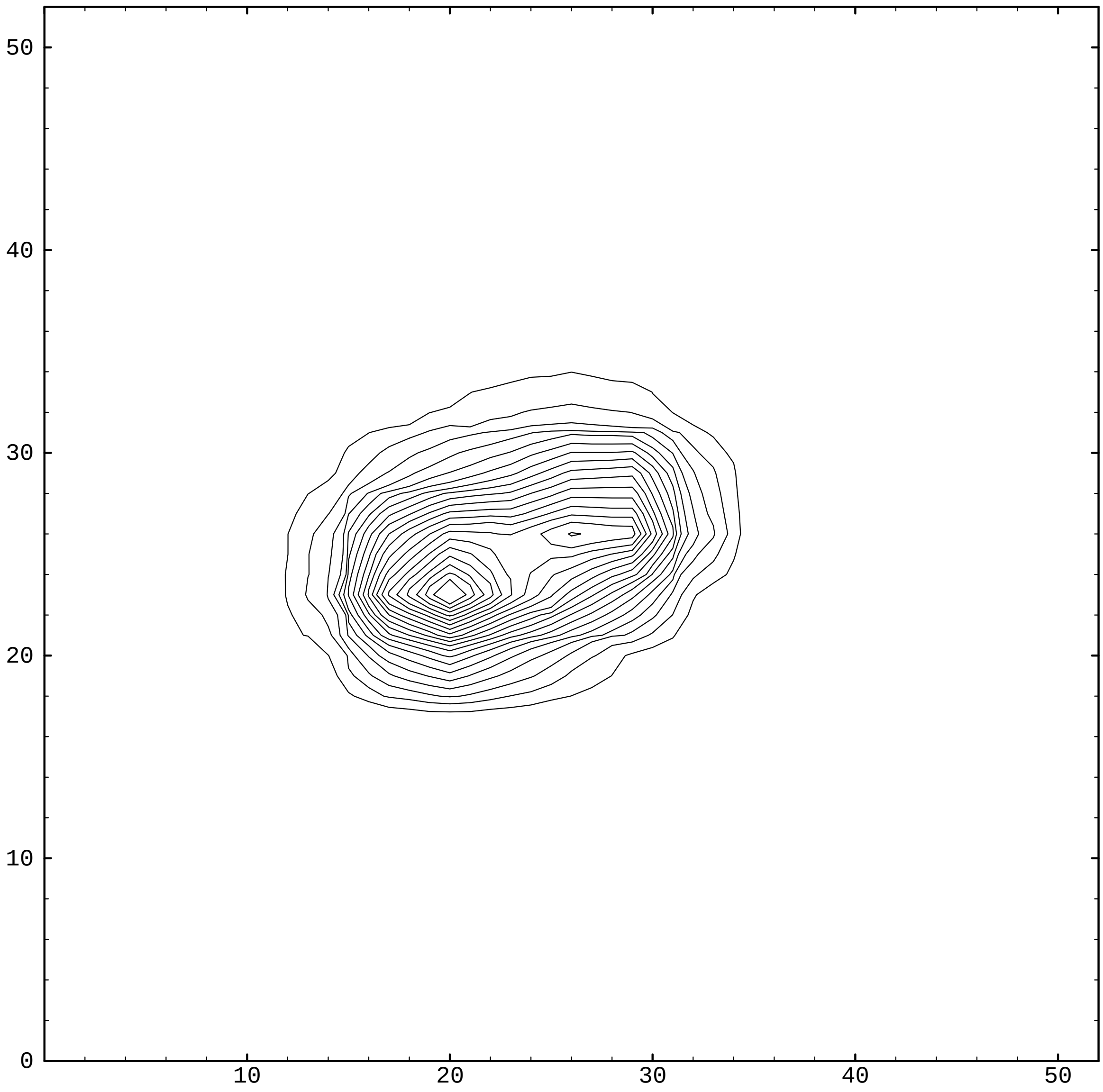

Figure 4-13: initial image, guess algorithm (twin-peaked example)

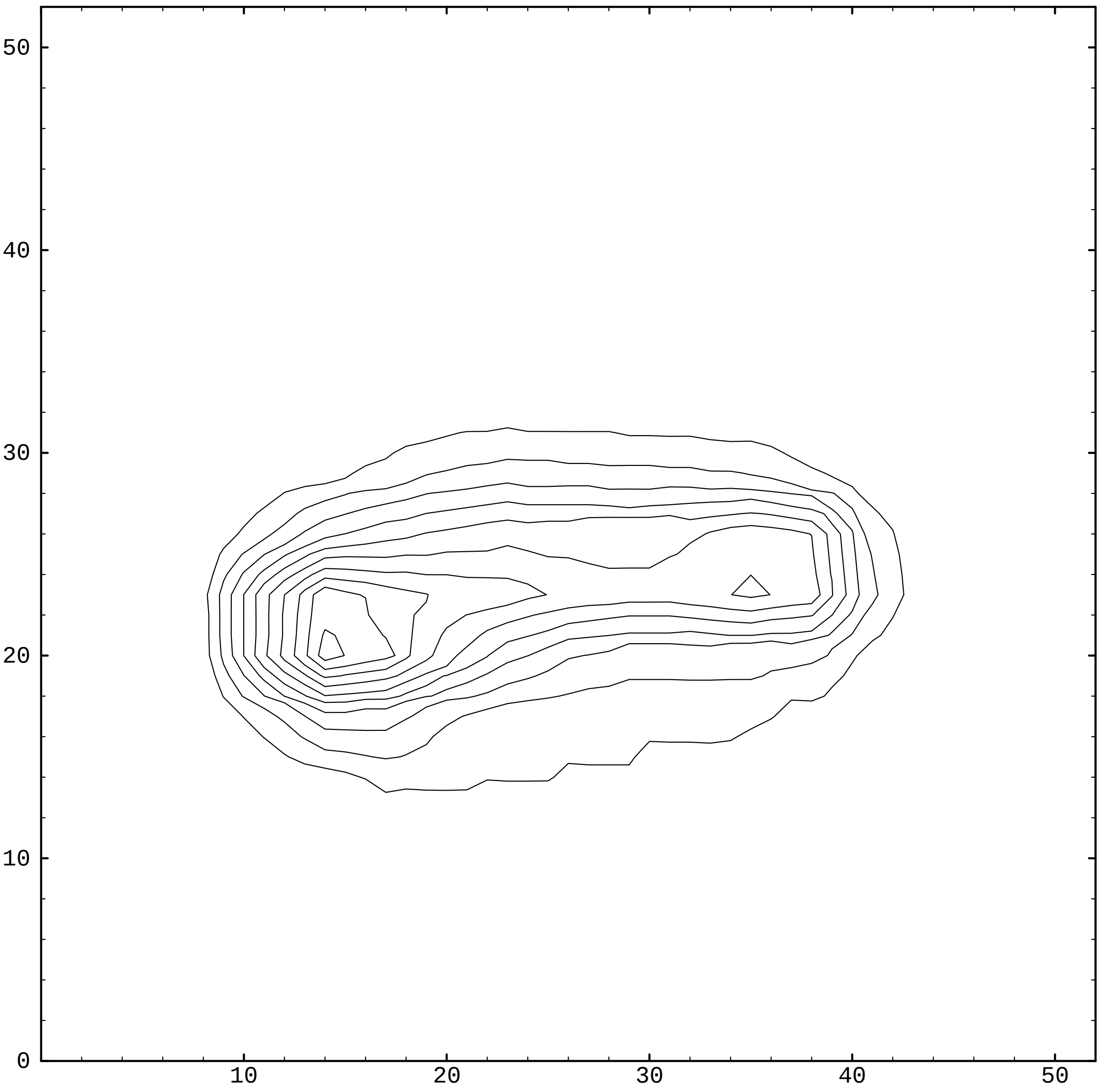

Figure 4-14: best solution, no guess (twin-peaked example)

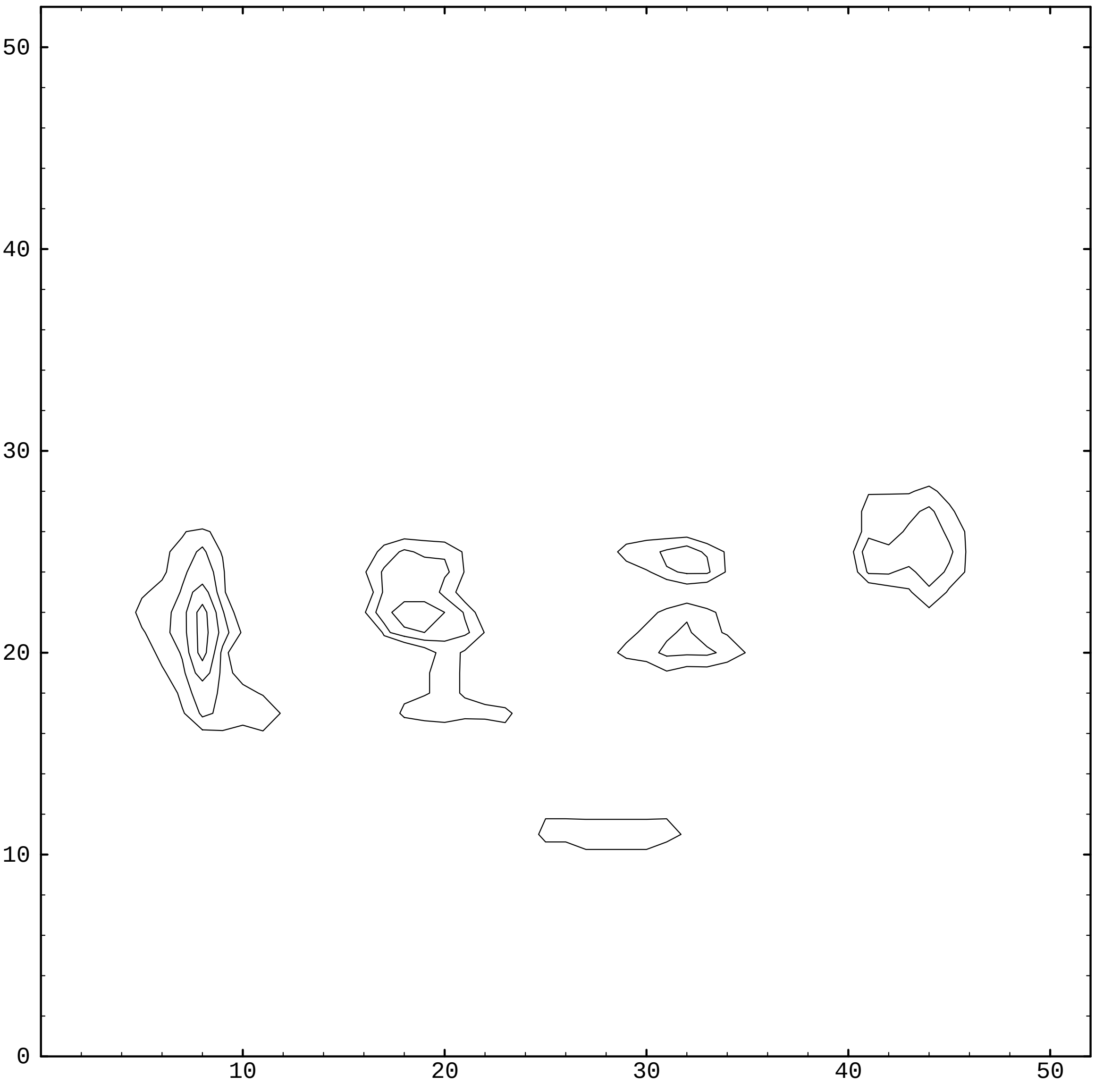

Figure 4-15: absolute difference plot (twin-peaked example)

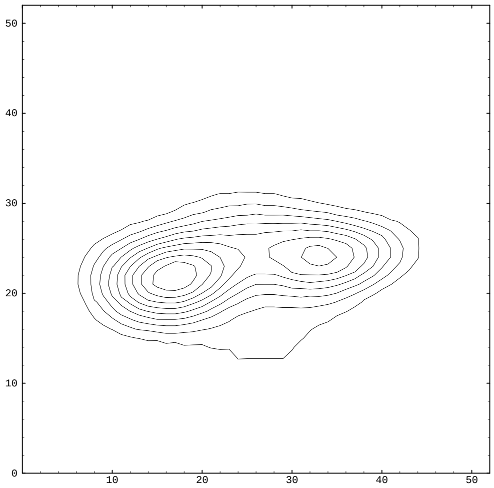

Figure 4-16: desired image (twin-peaked example)

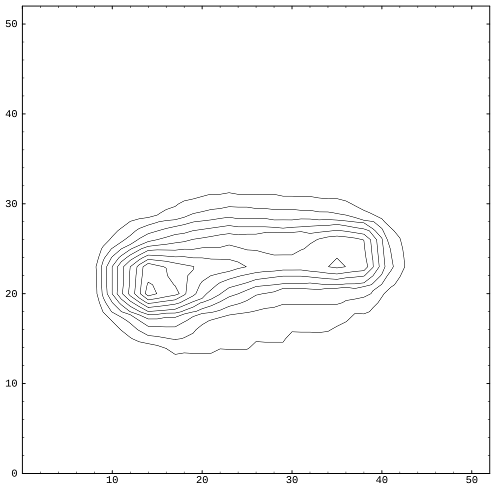

Figure 4-17: best solution, no guess (twin-peaked example)

4.6.3 A Tri-Lobed Example

Another image, with three lobes, was created using six perturbations: a converging offset, a linear term, an astigmatism, two quartics, and a diverging conical term. This example should also be difficult because of possible confusion between the proper contributions of the offset and radial terms (both having similar light-steering properties) in the solution. Furthermore, the first guess algorithm is not equipped to guess the offset term, which is important in this example. Most importantly, however, the quartic terms used to make the image are not present in the "standard" reflector model of section 4.2.

A summary of the residual for each of the two starting points is shown in figure 4-18. Without the guess the residual began at 119.5 and ended at 22.6, and with the guess the initial residual was 103.7 but only decreased to a disappointing 15.5. Because of the poor solution, this example is revisited in section 4.6.4 with an expanded parametric model for the reflector. The asymmetry of the solution in this case is due to the fact that the search directions are sampled sequentially and also change with each iteration.

Figure 4-19 depicts the desired image, figure 4-20 displays the initial image of the first starting point (tilted paraboloid only), figure 4-21 shows the initial image of the second starting point (tilted paraboloid with first guess algorithm), and figure 4-22 shows the best solution found. Figure 4-23 shows a plot of the absolute difference between the desired and best-fit images. Figures 4-24 and 4-25 show the desired and derived images side by side for comparison. Each contour line in this section represents a change in intensity of five greyscales, except for figure 4-23 where it represents 3.5 greyscales.

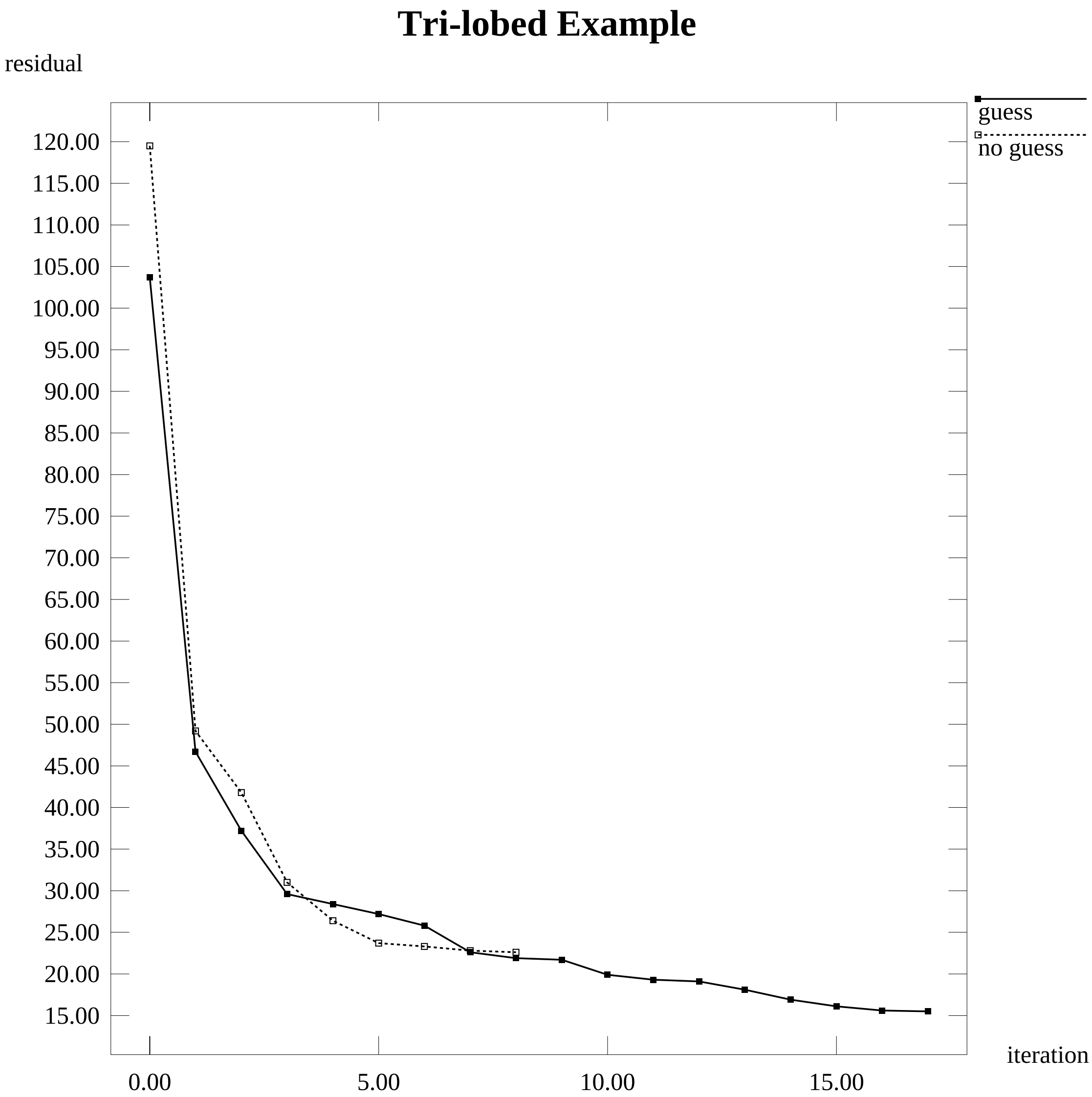

Figure 4-18: residual summary

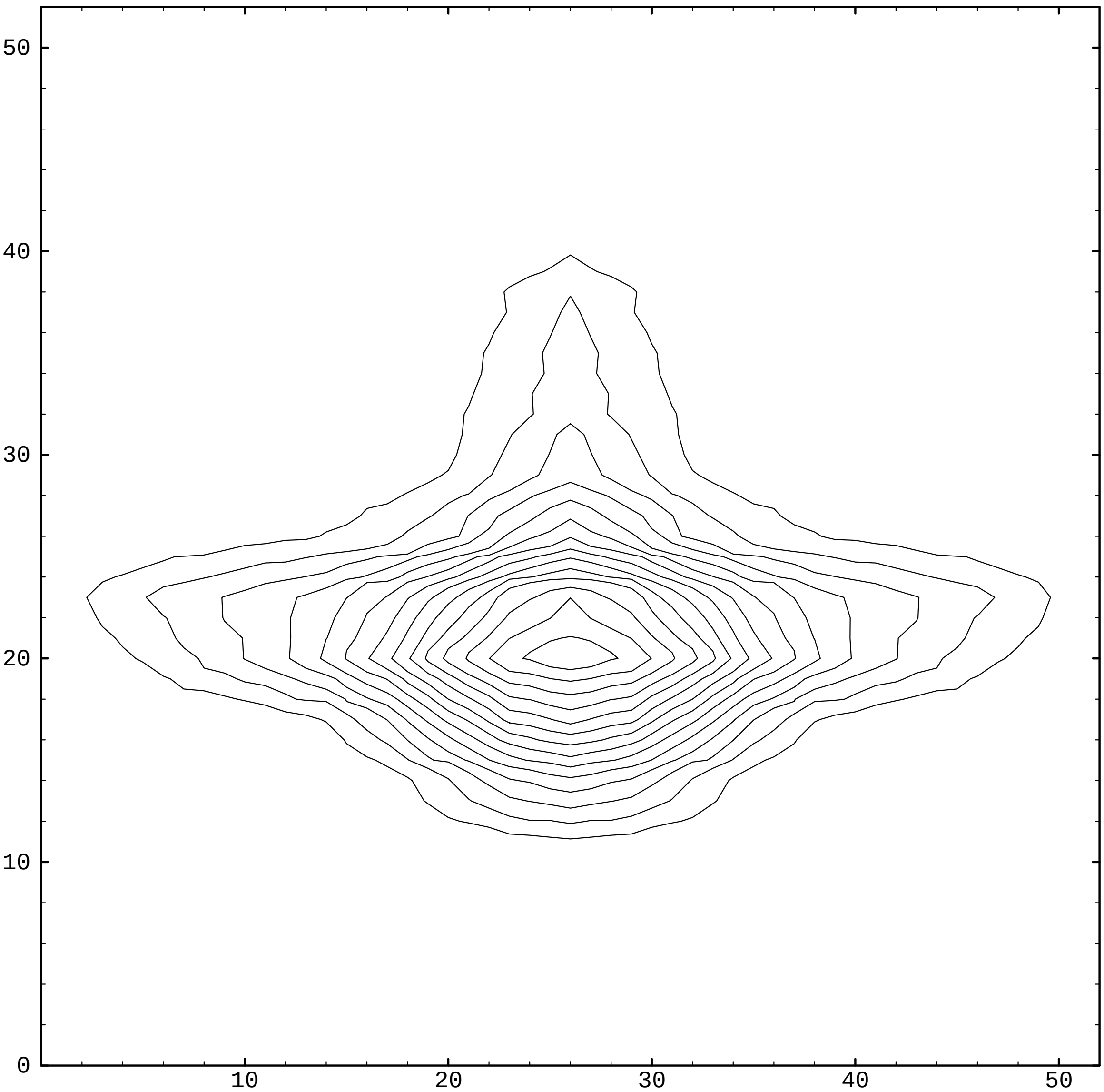

Figure 4-19: desired image (tri-lobed example)

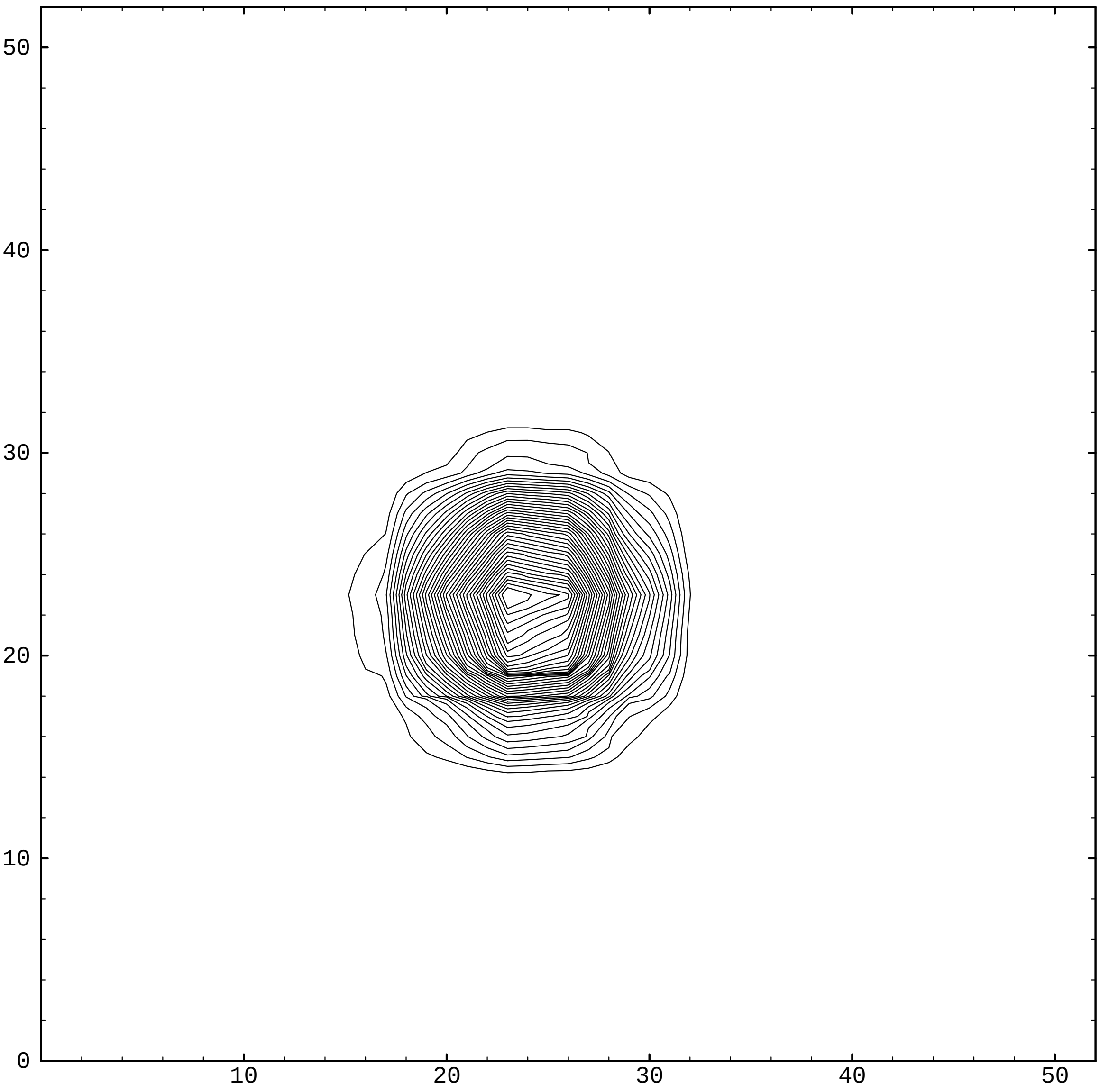

Figure 4-20: initial image (tri-lobed example)

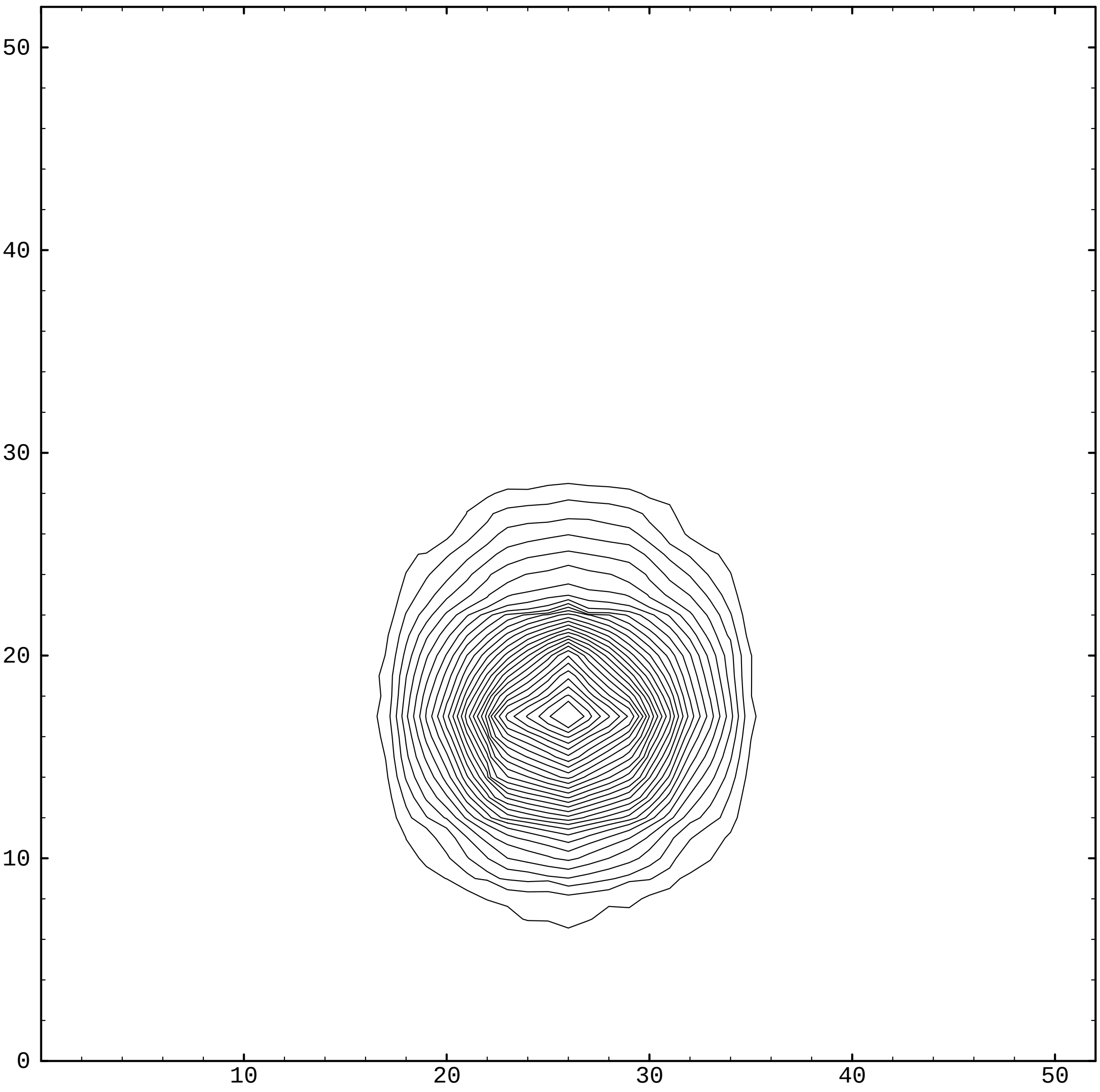

Figure 4-21: initial image, guess algorithm (tri-lobed example)

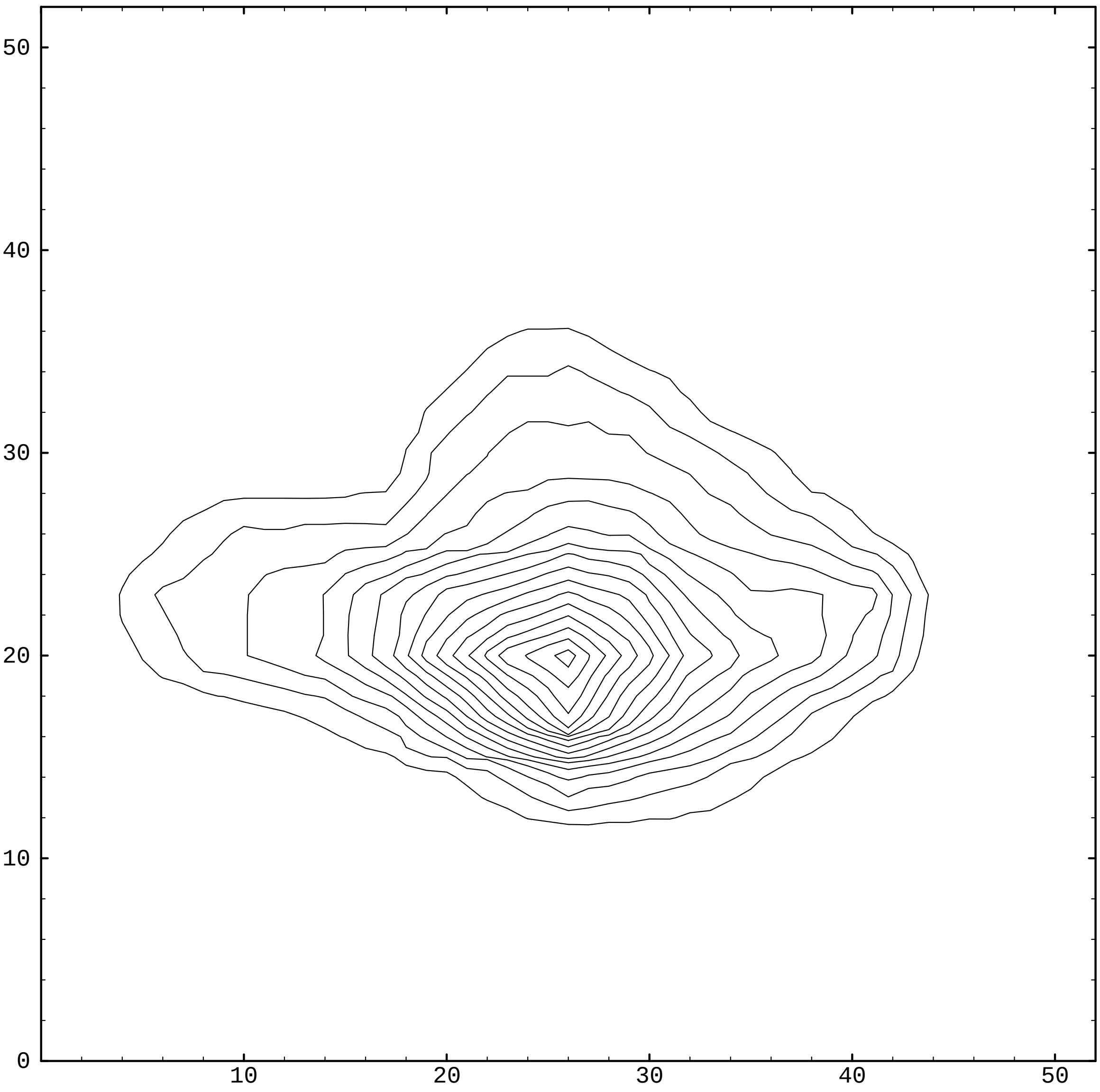

Figure 4-22: best solution, guess algorithm (trilobed example)

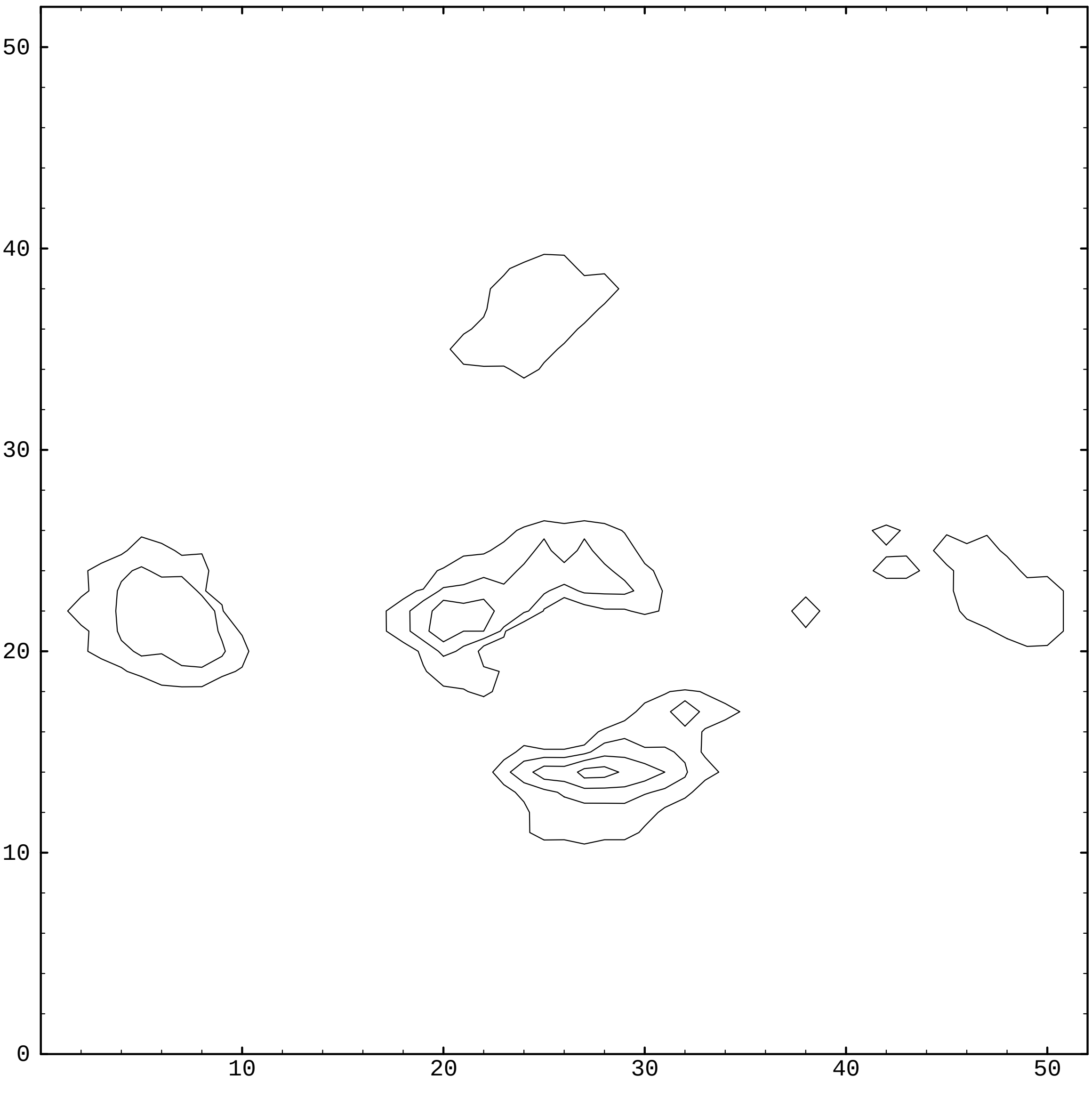

Figure 4-23: absolute difference plot (trilobed example)

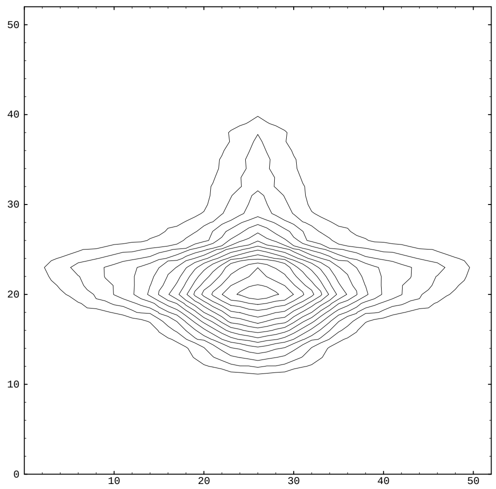

Figure 4-24: desired image (tri-lobed example)

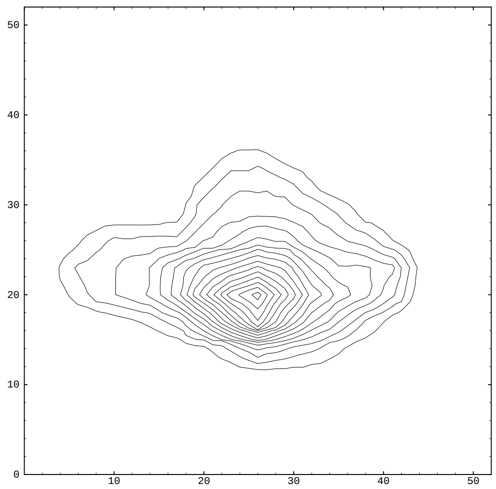

Figure 4-25: best solution, guess algorithm (tri-lobed example)

4.6.4 A Tri-Lobed Example Revisited

The example of section 4.6.3 was attempted again with an expanded reflector model. The model was extended to include the five fourth order terms x^4, y^4, x^3y, x^2y^2, and xy^3.

A summary of the residual for each of the two starting points is shown in figure 4-26. Without the guess the residual began at 119.5 and ended at 14, but with the guess the initial residual was 103.3 and ended at an astonishing 4.9, showing the importance of the quartic terms in this case.

Figure 4-27 shows the initial image of the second starting point (tilted paraboloid with first guess algorithm) and figure 4-28 shows the best solution found. Figure 4-29 shows a plot of the absolute difference between the desired and best-fit images. Figures 4-30 and 4-31 show the desired and derived images side by side for comparison. Each contour line in this section represents a change in intensity of five greyscales, except for figure 4-29 where it represents 3.5 greyscales.

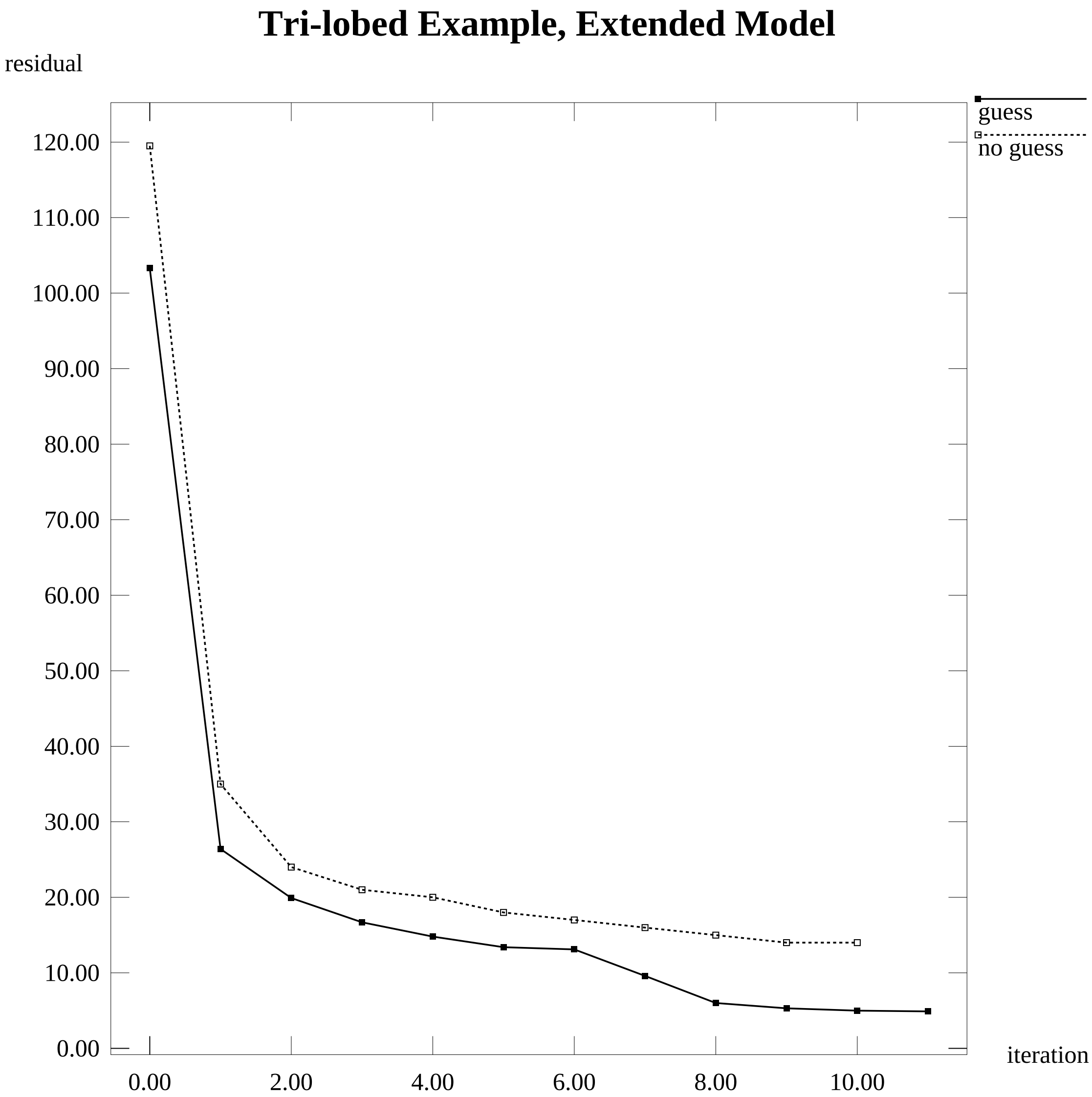

Figure 4-26: residual summary

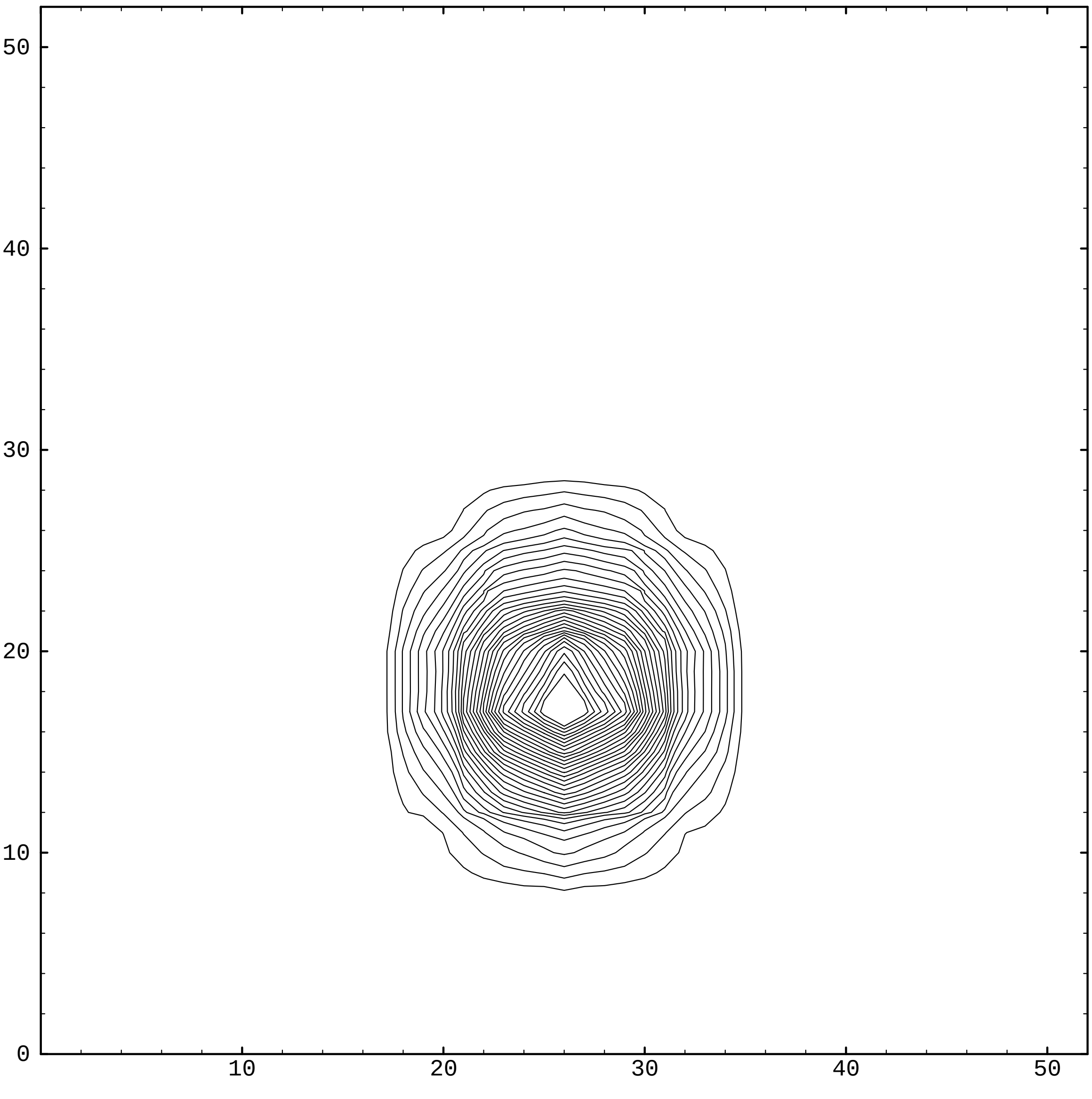

Figure 4-27: initial image, guess algorithm (extended tri-lobed example)

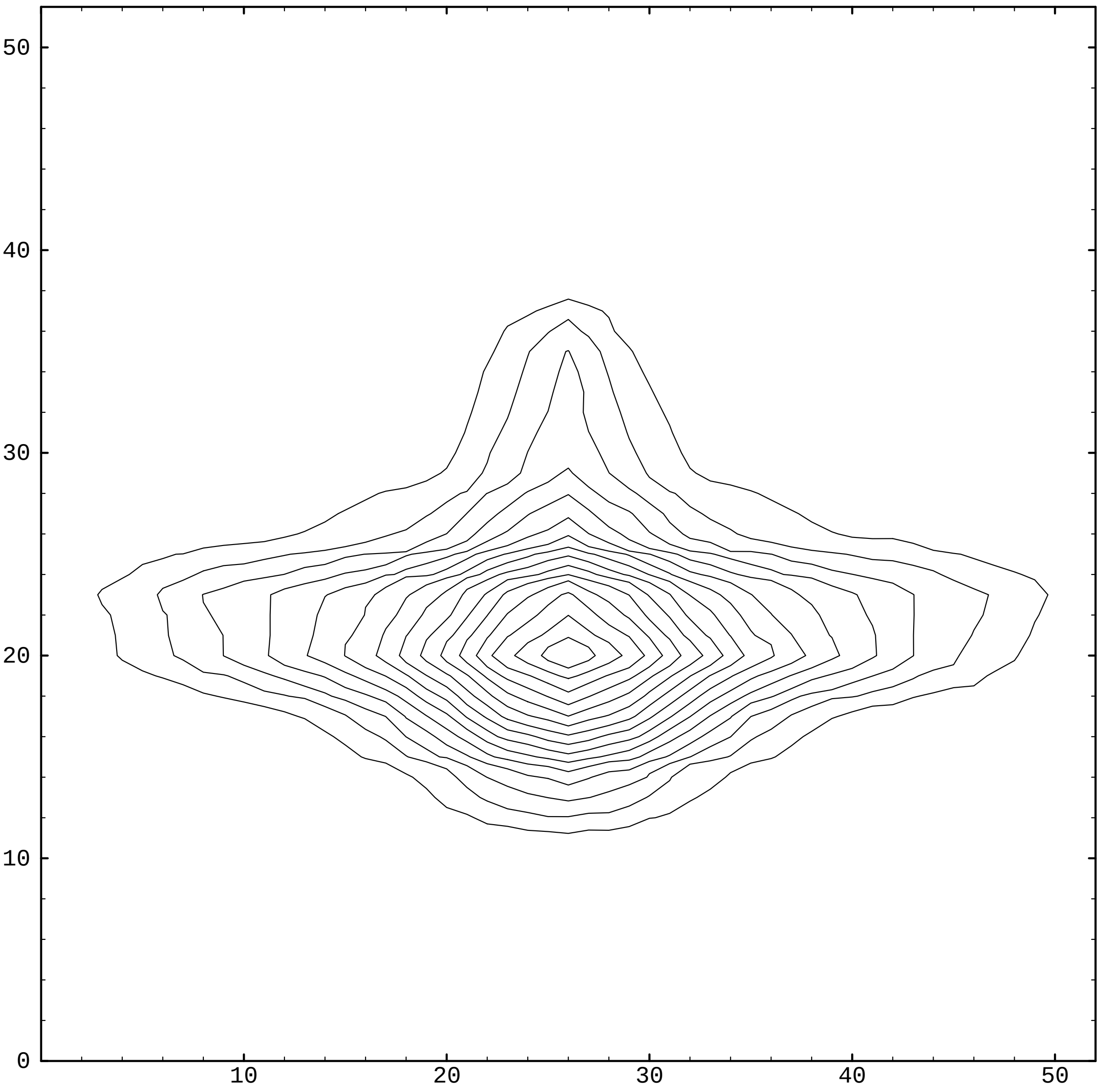

Figure 4-28: best solution, guess algorithm (extended trilobed example)

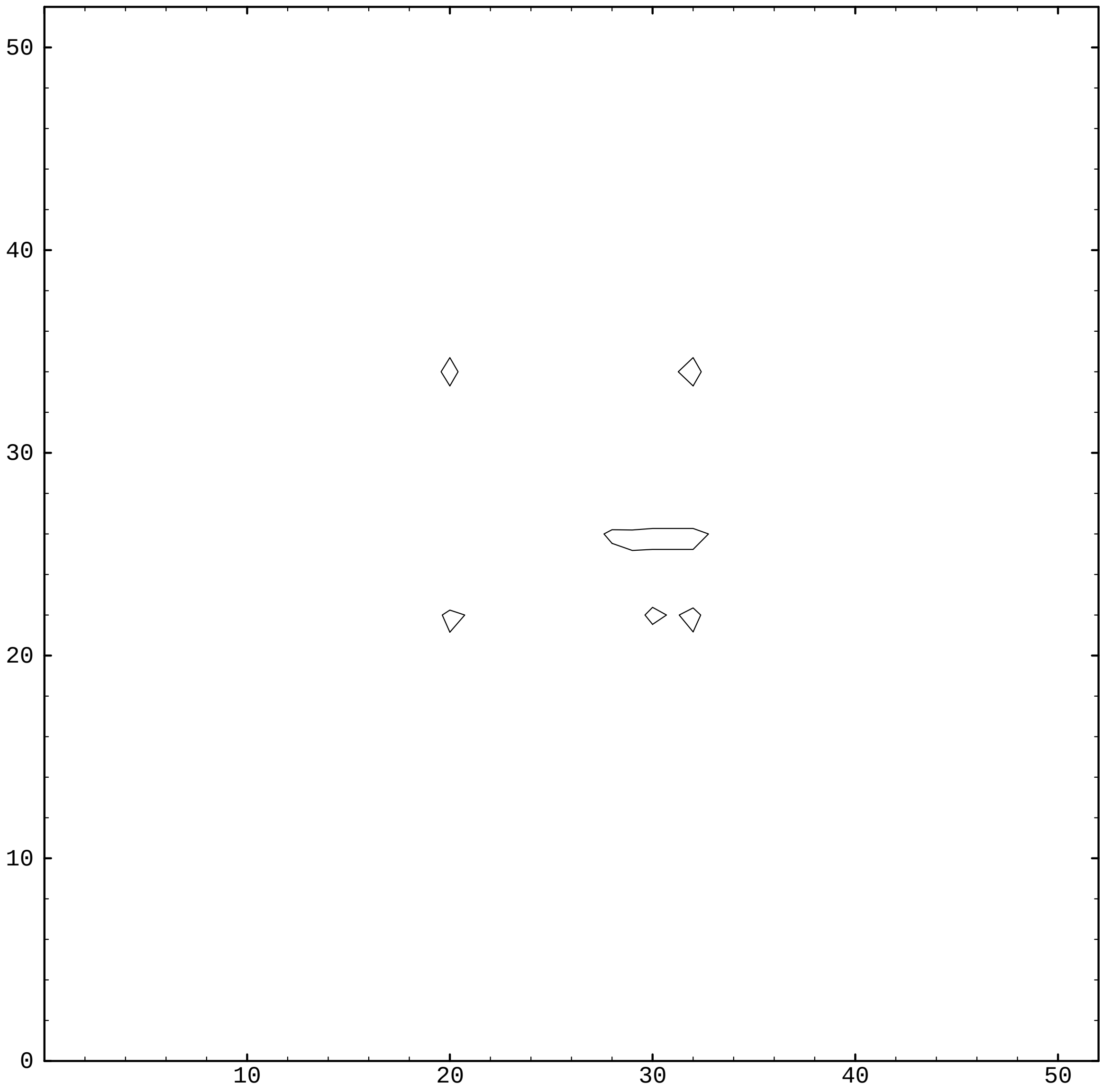

Figure 4-29: absolute difference plot (extended trilobed example)

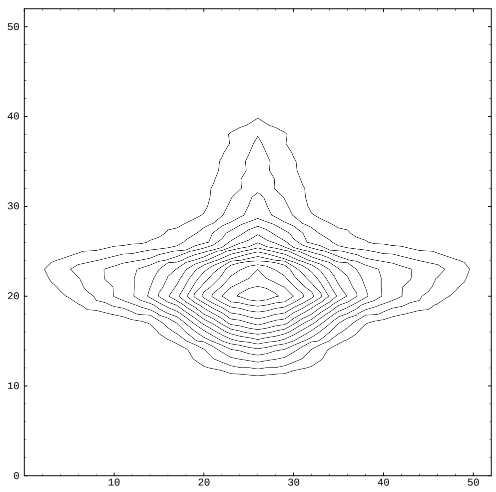

Figure 4-30: desired image (tri-lobed example)

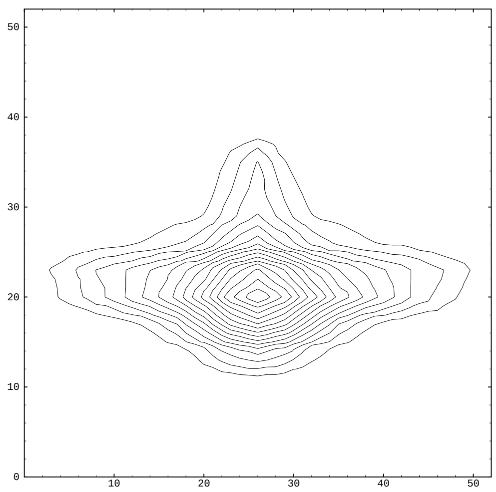

Figure 4-31: best solution, guess algorithm (extended tri-lobed example)

4.6.5 Discussion

For choosing good starting points, the first guess algorithm used in conjunction with pointing the light initially outperformed mere pointing in all cases. The first guess also led to better final solutions in all cases except one (section 4.6.2), in which the guessing algorithm's implicit assumption of a single intensity peak was violated. The conclusion to be drawn is that when no *a priori* information is known about the residual surface, it is best to sample as many different regions as possible by using multiple initial approximations.

4.7 Results - Real Source

As a proof of concept, we present a mirror design example that uses actual spot quality bulb data to obtain a lighting pattern from a real source. The bulb dataset used is shown in figure 3-2. Because the CCD camera measurements were not calibrated, a true automotive design is not attempted. Instead, the point of this section is to show that the data compression and handling concepts of chapter 3 can indeed be put into practice.

The image was created on a 48x48 grid, with a pixel width of 0.5 f, at a distance of 60 f from the source. Every fourth pixel on every fourth row was rendered, with the rest filled in by linear interpolation. The mirror surface was sampled by squares 0.2 f on a side on a 27x27 grid.

Each of the 400 pinhole camera pictures representing the source was fit to 25 m_{ij} terms instead of the usual 49 to increase the rendering speed; although the regeneration of individual pinhole camera images suffered, the effect on the overall projected image was small (an *rms* difference of less than 2.0 greyscales over a dynamic range of 140).

4.7.1 A Real Source Example

A desired image was created using six perturbations, including a strong linear term. A summary of the residual for each of the two starting points (with and without the first guess) is shown in figure 4-32. Without the guess the residual began at 92.0 and ended at 20.2, and with the guess the initial residual was 77.2 but only decreased to a tantalizing 11.7. Because of the barely acceptable solution, this example is revisited in section 4.7.2 with an expanded parametric model for the reflector.

Figure 4-33 depicts the desired image, figure 4-34 displays the initial image of the first starting point (tilted paraboloid only), figure 4-35 shows the initial image of the second starting point (tilted paraboloid with first guess algorithm), and figure 4-36 shows the best solution found. Figure 4-37 shows a plot of the absolute difference between the desired and best-fit images. Figures 4-38 and 4-39 show the desired and derived images side by side for comparison. Each contour line in this section represents a change in intensity of seven greyscales, except for figure 4-37 where it represents 3.5 greyscales.

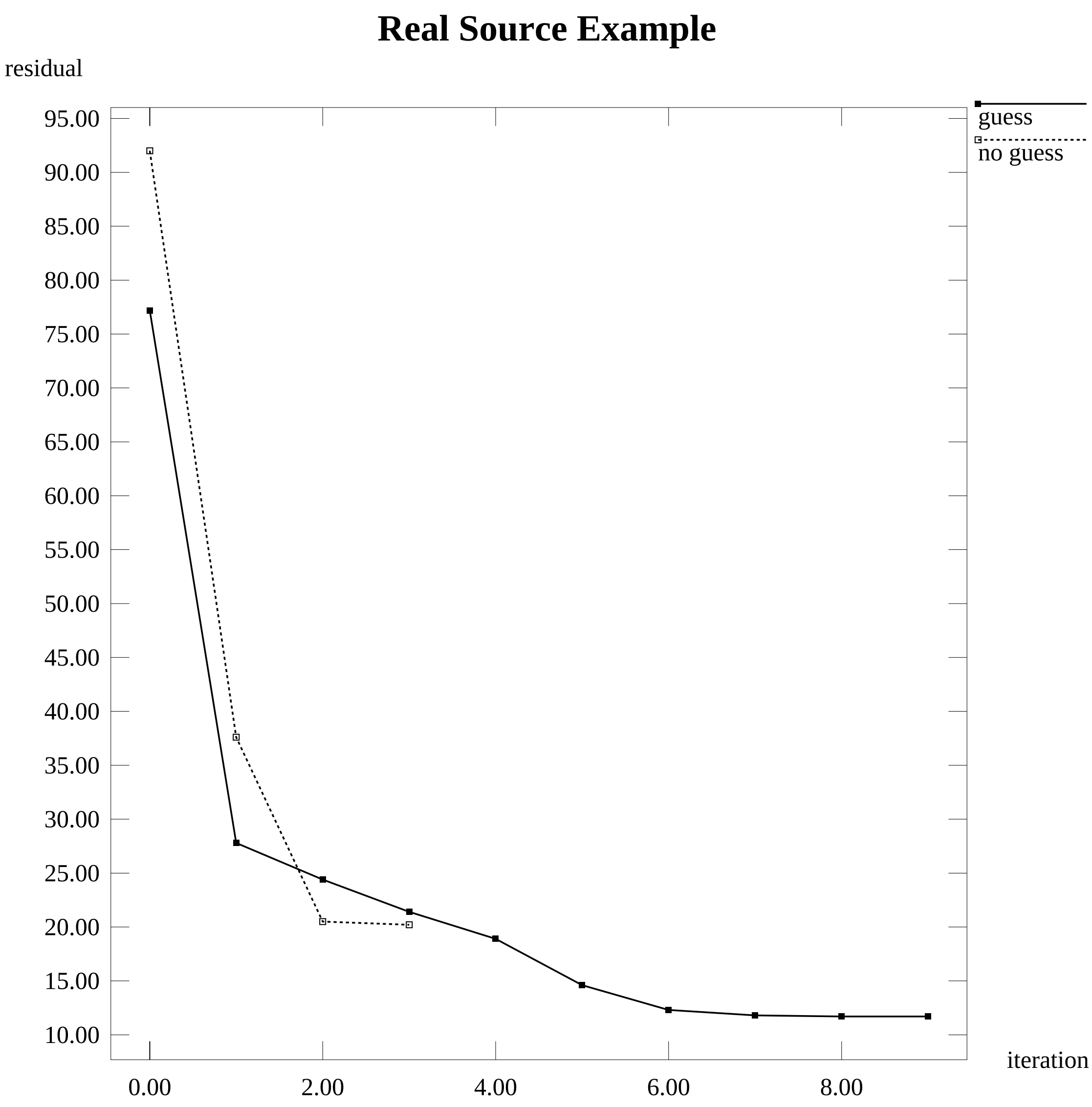

Figure 4-32: residual summary

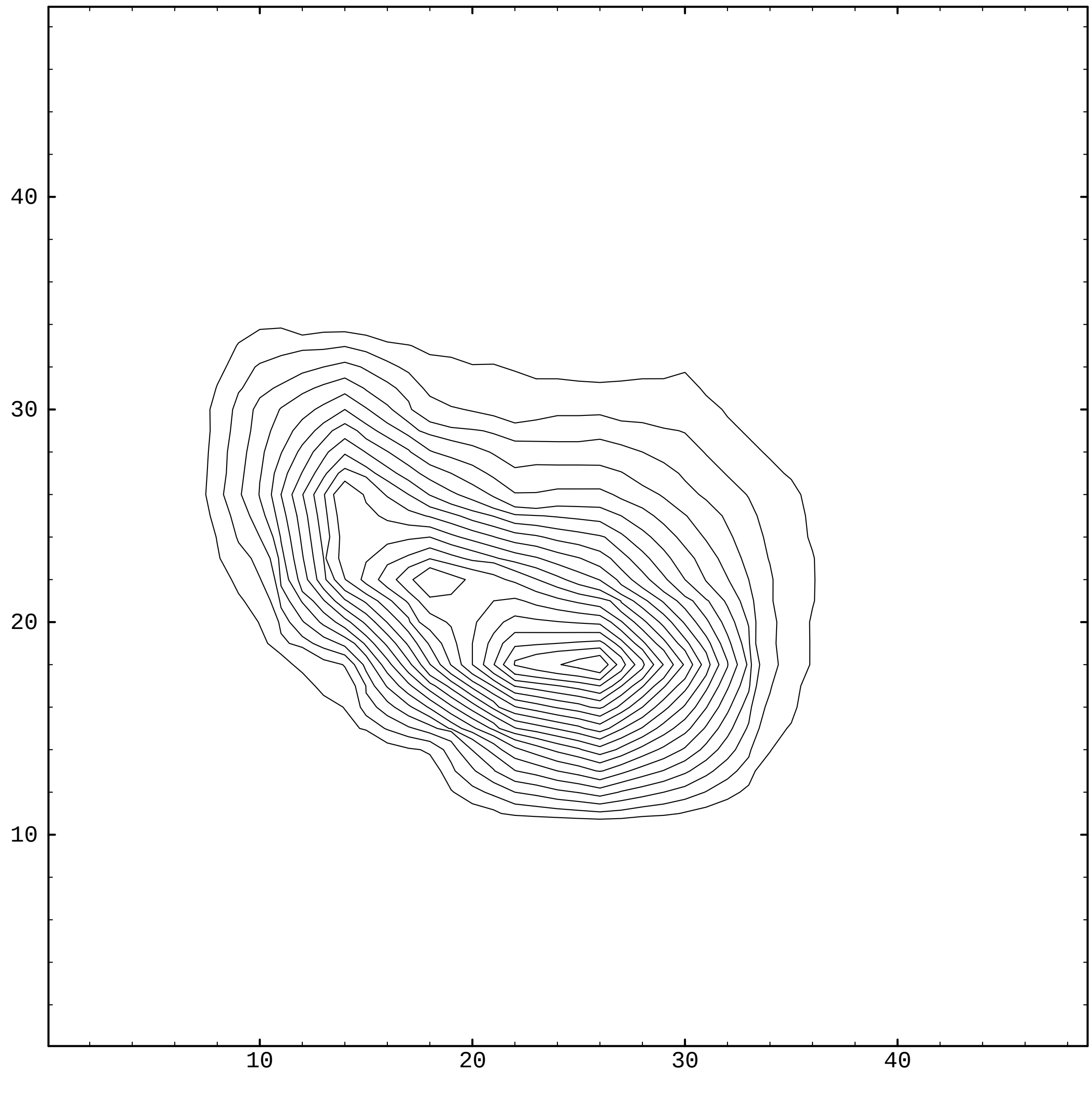

Figure 4-33: desired image (real source example)

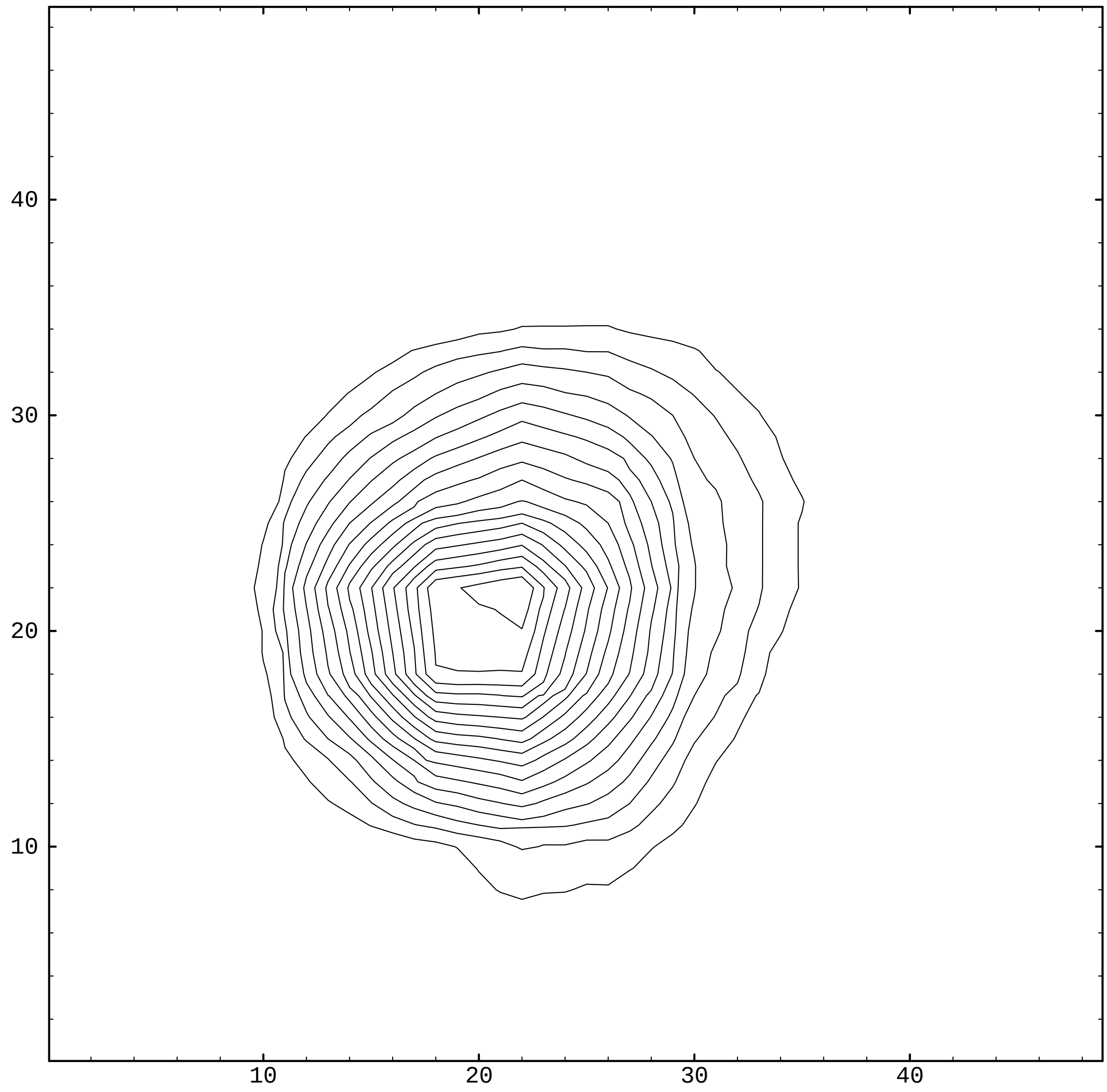

Figure 4-34: initial image (real source example)

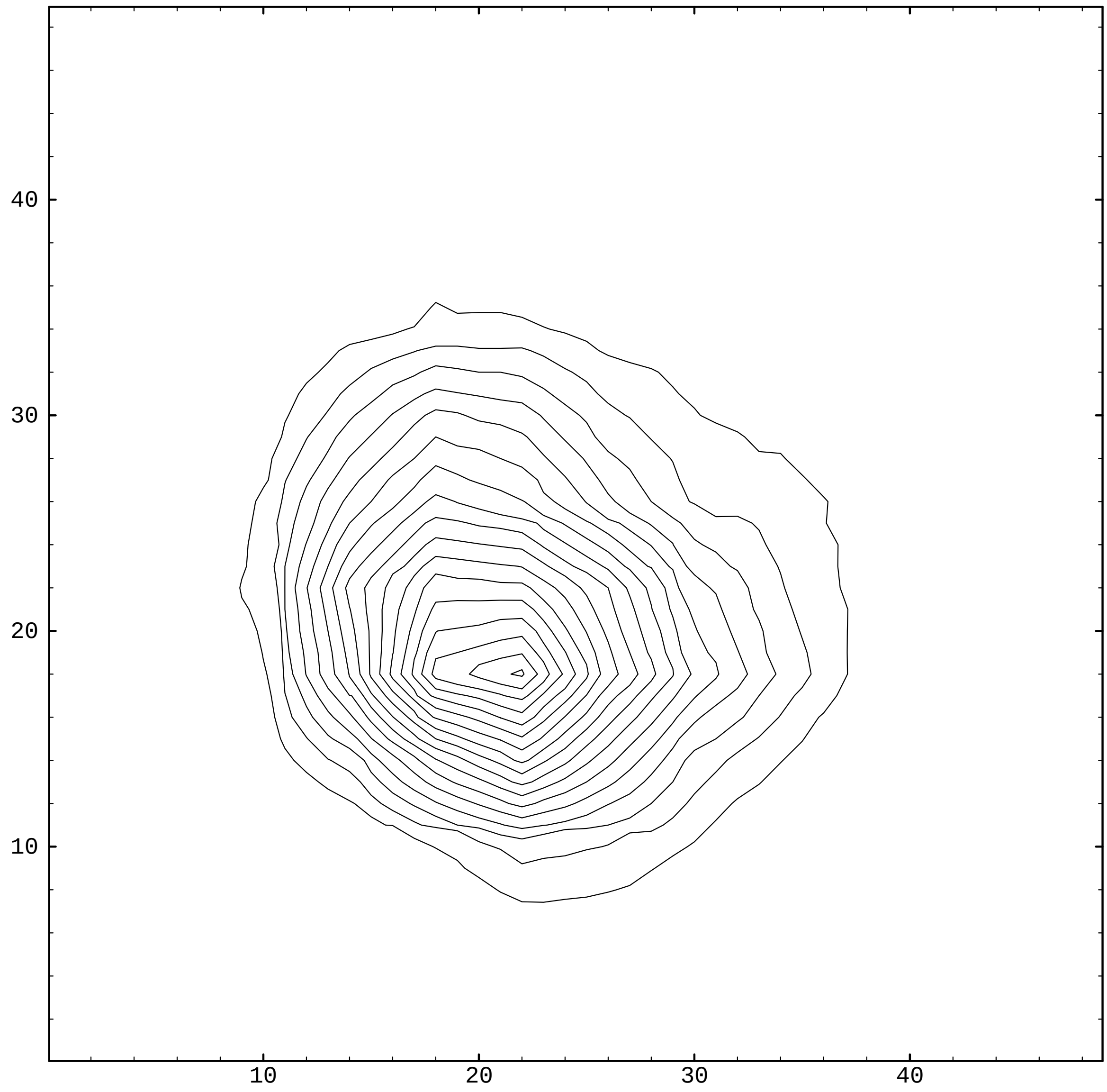

Figure 4-35: initial image, guess algorithm (real source example)

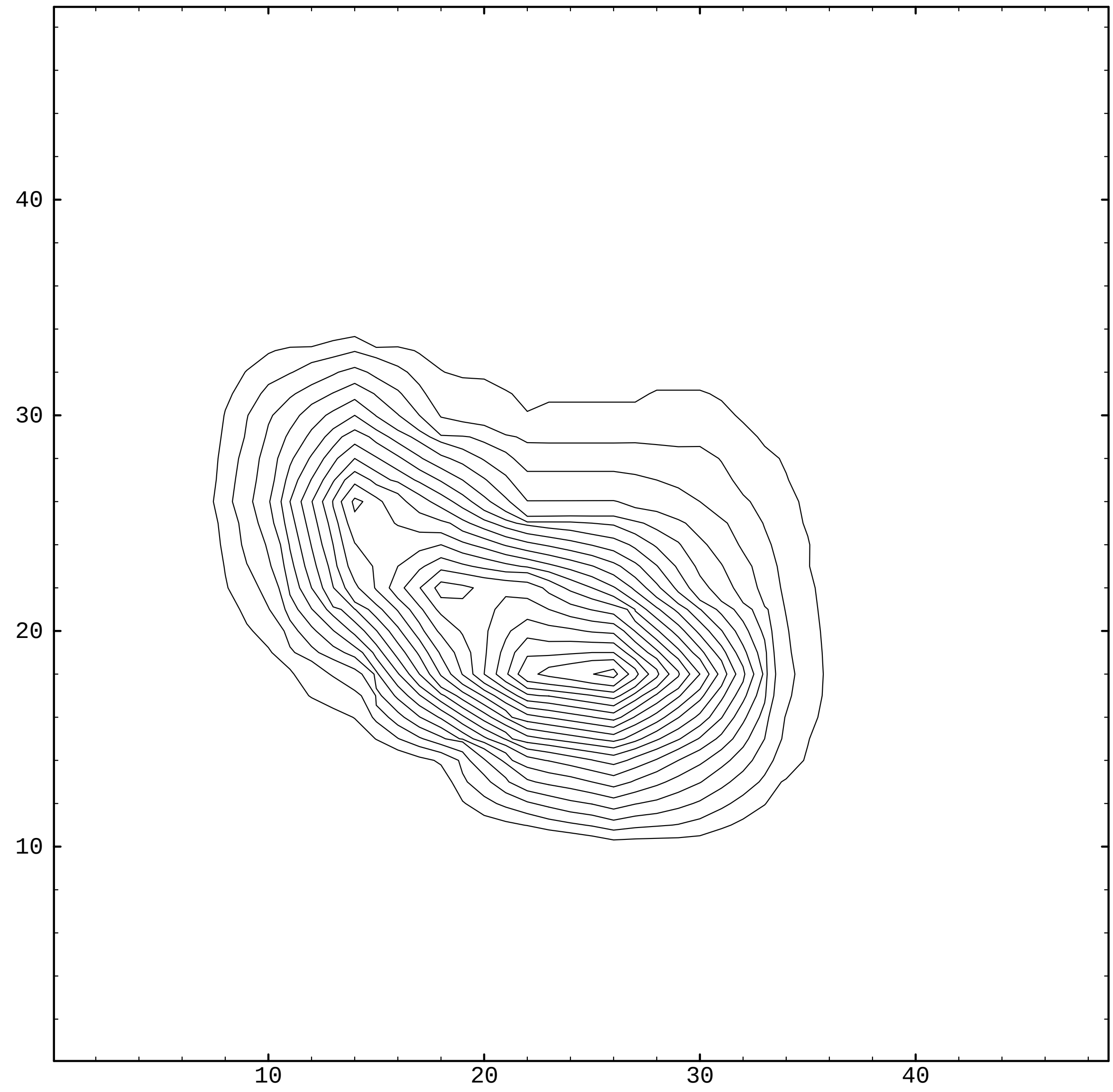

Figure 4-36: best solution, guess algorithm (real source example)

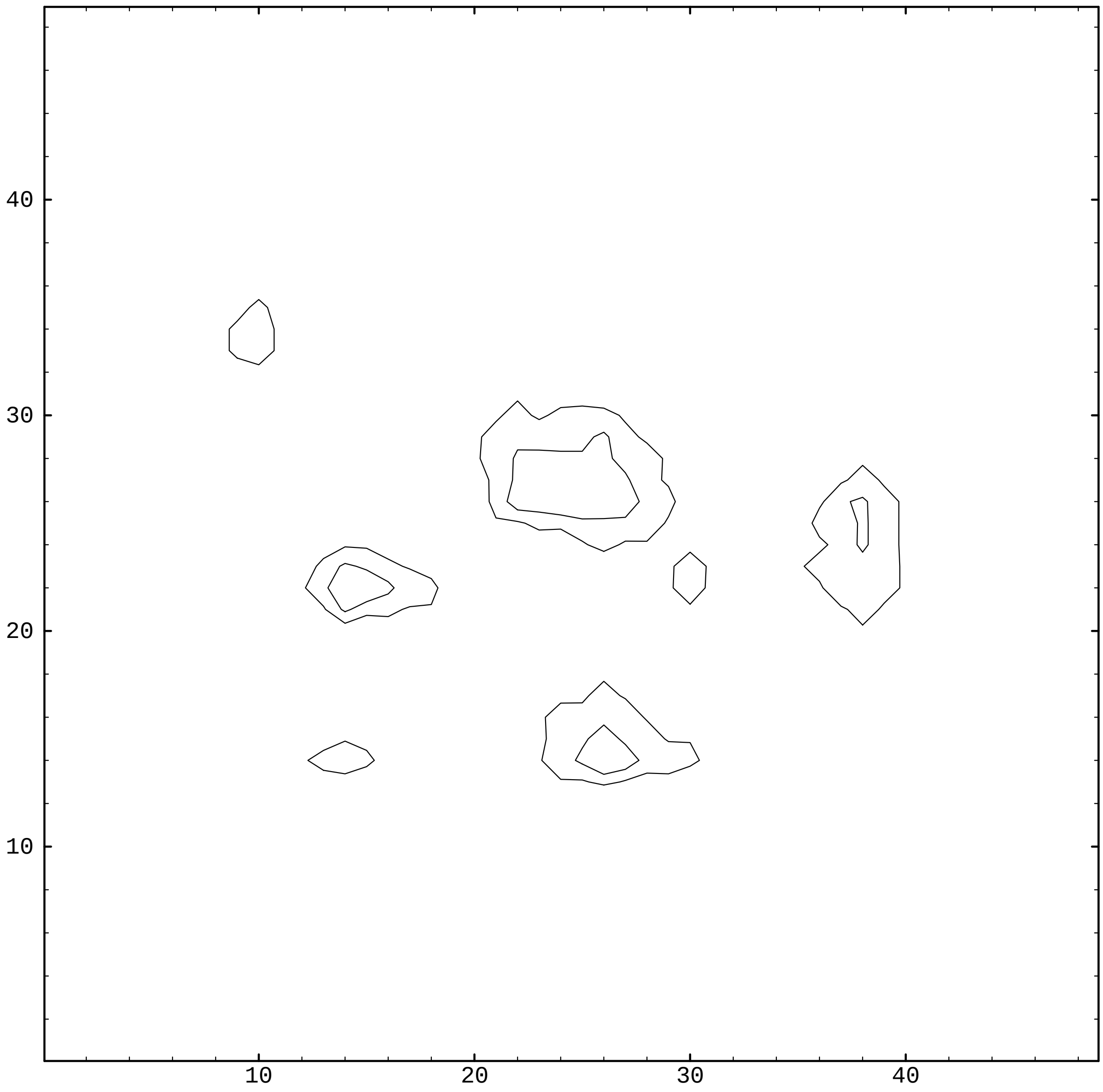

Figure 4-37: absolute difference plot (real source example)

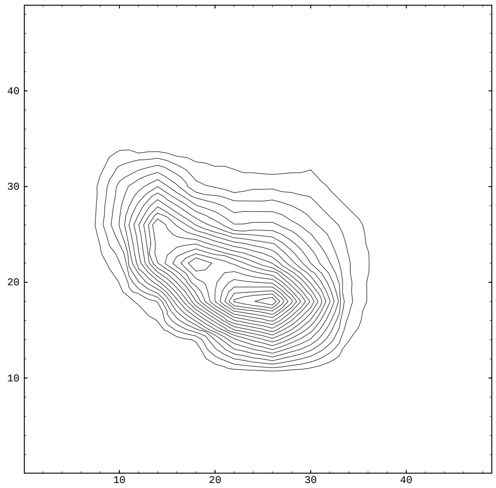

Figure 4-38: desired image (real source example)

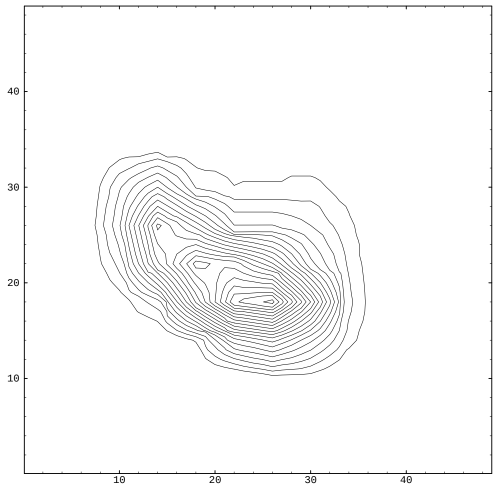

Figure 4-39: best solution, guess algorithm (real source example)

4.7.2 A Real Source Example Revisited

The parametric reflector model was again extended to include the five extra fourth order terms, as in section 4.6.4. Although it was not possible to obtain a better initial guess (because the guess only attempts to adjust the border of the image and not the interior), the searching algorithm found a better solution in fewer iterations with the extended model. As shown in figure 4-40, the residual was reduced to 10.7 after six tries. Figure 4-42 shows a plot of the absolute difference between the desired and best-fit images. Figure 4-41 shows the final derived image, and figures 4-43 and 4-44 show the desired and final images side by side for comparison. Each contour line in this section represents a change in intensity of seven greyscales, except for figure 4-42 where it represents 3.5 greyscales.

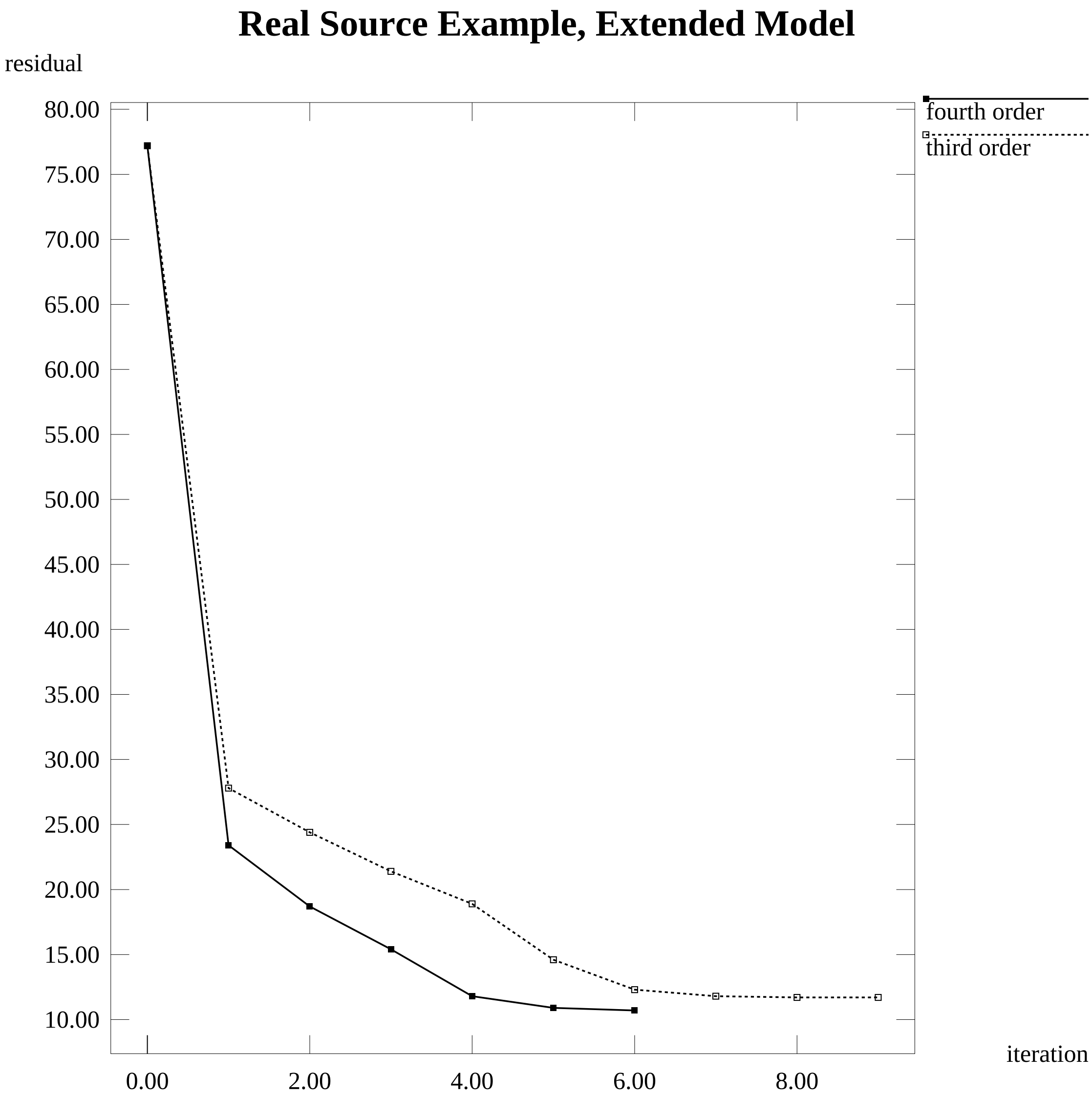

Figure 4-40: residual summary

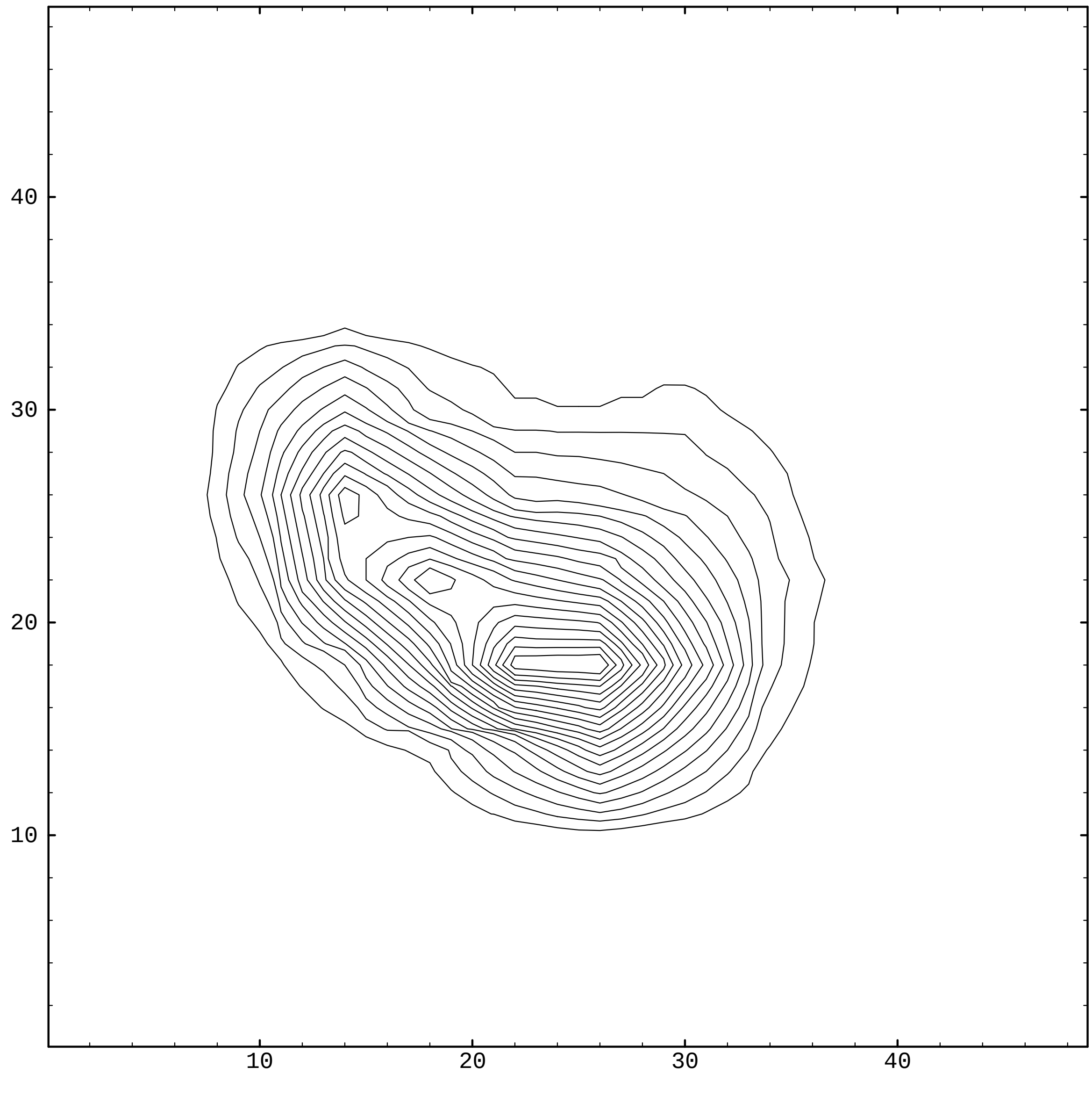

Figure 4-41: best solution, guess algorithm (real source example, extended model)

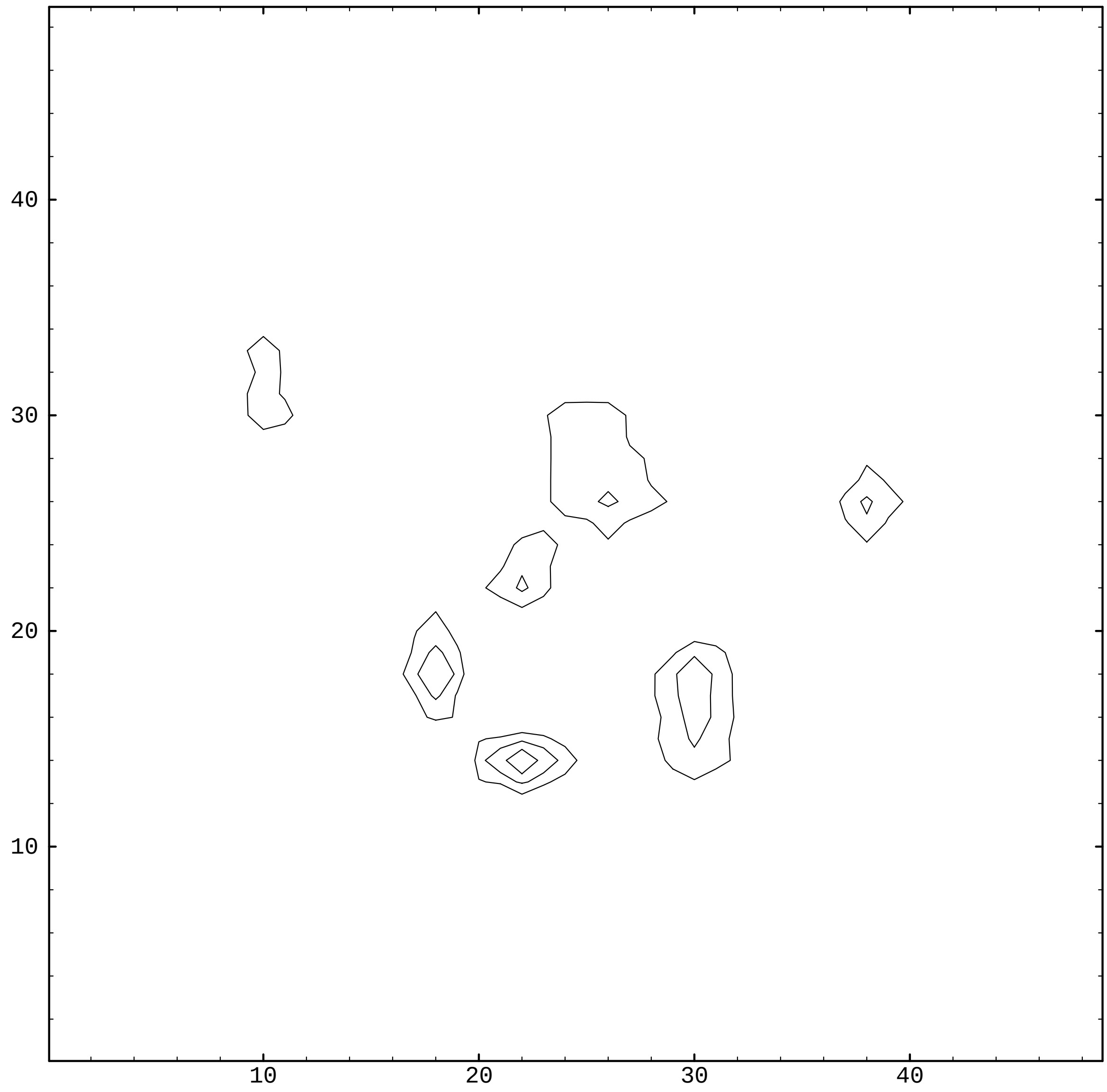

Figure 4-42: absolute difference plot (real source example, extended model)

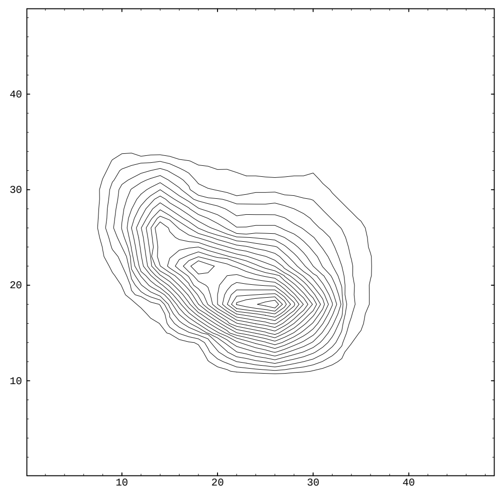

Figure 4-43: desired image (real source example)

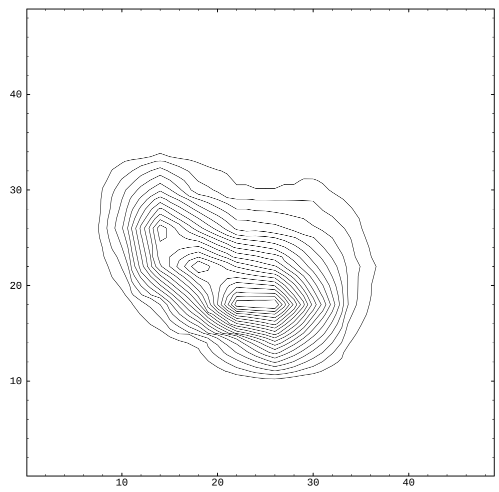

Figure 4-44: best solution, guess algorithm (real source example, extended model)

5. CONCLUSION

5.1 Significance

Fundamentally, the goal of this research is to develop a model of radiometry based on a combination of judicious measurement, efficient data representation, and physics based computation.

The impact of the discovery of the far-zone d_l parameters (chapter 2) is that light sources of inherently different natures can nevertheless be compared on an equal footing, i.e., the moment-like numbers characteristic of their spatial distributions offer a standard of comparison whatever the nature of the light source. Furthermore, the rotationally invariant nature of the d_l allows the sources to be compared without introducing rotational artifacts due to the arbitrary orientation of a local coordinate system.

The general near-zone light source model (chapter 3) will find applications in the fields of realistic scene rendering, luminaire modeling, and luminaire design. Currently, the Fourier expansion seems to be the most compact and convenient description of the radiance distribution of an emitting surface.

The near-zone model is efficient because the radiance distribution of an extended source is encoded as a set of coefficients of orthogonal functions over a sphere, which solves the data storage problem, the interpolation problem, and the brute-force ray tracing intersection problem. The model is physics-based because it obeys the laws of geometric optics without resorting to the simplifying assumptions concerning the source's shape or radiance distribution often found in computer graphics-based methods. The model applies to real light sources with complete generality because it relies on measurement, thus treating correctly extended bodies with nonuniform radiance distributions.

The general light source model, using a nonuniform, extended filament, is successfully applied to the problem of designing a reflector to produce a desired illumination pattern on a screen (chapter 4). To do this, a smooth reflector model that consists of a set of physically understood perturbations to a parabolic base shape was developed. Because it is easy to calculate the intersection of a ray with our physically correct near-zone light source model, it is possible to parametrically sample the reflector surface and render an image on a distant screen with a combination of forward and backward ray tracing. This system can be used as a design tool with a human in the loop to test various reflector shapes and source types, or it can be used in conjunction with a computer-driven optimization algorithm. A judicious choice of coordinate system and a good first guess at the perturbations based on the shape of the perimeter of the image can often result in a better reflector, or decrease the time required for the search of perturbation space. A numerical optimization method was chosen that is appropriate to this type of numerical problem, in which derivative information is lacking.

5.2 Directions for Further Research

The extent of the shadow region, for linearly smoothing the intensity data cutoff, should be investigated for the smoothing to be properly implemented. The size of the region depends on the goniometer apparatus in conjunction with the configuration of the source and the geometry of the bulb base. Since the apparatus is no longer available to me, a good estimate cannot be made. A rough estimate, however, based on the filament size alone of a dual-filament bulb suggests that the shadow region should begin at $\theta = 150$ grads (just where the data ended) and extend for perhaps 22 to 28 grads (about half of the unmeasured region). Vertical filament bulbs would of course have larger shadow regions. An experiment that linearized the cutoff over the entire unmeasured region for a vertical filament bulb, fitting to an l_{max} of 8, reduced the squared residual from 24 to 18. This issue must be fully explored for actual bulb characterization.

The method of calculating the occlusion of rays by the filament remains theoretical and has not been tested or implemented. The available CCD chip size, pinhole size, and goniometer (which had considerable wobble) caused the images to be small and fuzzy; thus the current real data on hand show fuzzy blobs instead of sharp filament images, and so will not allow "see-through" light to pass between the coils. For the idealized solid radiating sphere used as the source in most of the simulations, the occlusion method should perform accurately.

Currently, the focal length f must be supplied by the user; I do not know of a good way to automatically choose f based solely on the desired illumination pattern. Especially in an automotive headlamp design context, f will probably be constrained by space and styling considerations. Nor does the optimization algorithm search for f (although it could, though I have explained why I think it would be a bad idea in section 4.1). One way to approach this issue[23] is to consider the focal length to be a "magnification factor": given the length of the horizontal filament, the characteristic size of the image, and the distance to the image plane, it is possible to find f such that a ray from the end of the filament strikes the edge of the image.

Concerning which search direction to discard in the minimization algorithm, I would like to note that since the "condition number" of a matrix is a better measure of the linear dependence of its columns than the determinant [KAH 88], that the best direction to be replaced is the one that keeps $cond(\boldsymbol{V})$ small rather than the one that keeps $det(\boldsymbol{V})$ large. The condition number can be estimated as the ratio of the largest and smallest singular values of the matrix, found by a Singular Value Decomposition (section 2.6). The time involved for repeated decompositions may be prohibitive, however.

It should not be difficult to identify critical reflector segments (section 4.2) by finding the change in their individual contributions to the image residual as their heights are varied slightly, in order to choose proper manufacturing tolerances.

The optimization could be automated further, to include issues such as resolution, subsampling, and the number of terms used in the reflector model.

The significance of design tolerances with respect to the robustness of the solution should be investigated, though it should be a simple matter to determine the change of the image for a change in reflector perturbation. The effects of the manufacturing tolerances in the light sources themselves should also be tested; they will exhibit themselves as uncertainties in the $\{\alpha_{lm}^{k}\}$. Current headlight design practice calls for a half a degree range around the testpoints at 60 feet to account for variations in bulb manufacture and placement (bulb placement should be a very sensitive factor in our method).

---

[23]due to M. W. Siegel

The way to avoid fitting a_{00} over the measurement sphere[24] described in section IV should be inplemented.

If a non-spotlight type of image is desired, the general approach outlined here still holds, but a different reflector representation, especially for the base shape, should be used. A different set of functions for the image representation should also be found. In other words, it is a matter of choosing the proper basis functions for the problem at hand.

[24] suggested by P. Heckbert

I. PROOF OF ROTATIONAL INVARIANCE

Consider an arbitrary rotation of coordinates in which the point (θ,ϕ) is renamed (θ',ϕ'), resulting in a new set of coefficients $\{\alpha'_{lm}\}$. Requiring

$$I(\theta,\phi) = I(\theta',\phi')$$

results in

$$\sum_{lm} Y_{lm}(\theta,\phi)\,\alpha_{lm} = \sum_{lm} Y_{lm}(\theta',\phi')\,\alpha'_{lm}$$

Under a rotation of axes the Y_{lm} transform as [ROS 57]

$$Y_{lm}(\theta',\phi') = \sum_{n} D^{l}_{nm}\,Y_{ln}(\theta,\phi)$$

where the $\boldsymbol{D}$ are the Wigner $\boldsymbol{D}$-matrices from Quantum Mechanics, and depend on the Euler angles of rotation of the axes. Substituting, we obtain

$$\sum_{lm} Y_{lm}(\theta,\phi)\,\alpha_{lm} = \sum_{lmn} D^{l}_{nm}\,Y_{ln}(\theta,\phi)\,\alpha'_{lm}$$

or

$$\sum_{lm} Y_{lm}(\theta,\phi)\,\alpha_{lm} = \sum_{ln} Y_{ln}(\theta,\phi)\cdot\sum_{m} D^{l}_{nm}\,\alpha'_{lm}$$

Because the spherical harmonics are orthonormal, i.e.

$$\int d\theta \int d\phi\, Y^{*}_{l'm'}(\theta,\phi)\,Y_{lm}(\theta,\phi) = \delta_{l'l}\delta_{m'm}$$

multiplying by $Y^{*}_{l'm'}(\theta,\phi)$ and integrating over angles produces

$$\alpha_{lm} = \sum_{n} D^{l}_{mn}\,\alpha'_{lm}$$

Therefore, for fixed l we have

$$\sum_{m} |\alpha_{lm}|^2 = \sum_{m} \Big(\sum_{p} D^{*l}_{pm}\,\alpha'^{*}_{lp}\Big)\cdot\Big(\sum_{q} D^{l}_{mq}\,\alpha'_{lq}\Big)$$

or

$$\sum_{m} |\alpha_{lm}|^2 = \sum_{pq} \alpha'^{*}_{lp}\,\alpha'_{lq}\Big(\sum_{m} D^{*l}_{pm} D^{l}_{mq}\Big)$$

The $\boldsymbol{D}$-matrices are unitary, meaning

$$\sum_{m} D^{*l}_{pm} D^{l}_{mq} = \delta_{pq}$$

so that

$$\sum_{m} |\alpha_{lm}|^2 = \sum_{m} |\alpha'_{lm}|^2 = \text{constant under rotations.}$$

$$N \equiv \sum_{lm} |\alpha_{lm}|^2$$

and

$$d_l \equiv \left(\sum_m |\alpha_{lm}|^2 / N\right)^{1/2}$$

are also invariant under rotations. A set d_l depends on the shape of the angular distribution but not on the absolute intensity, which is related to α_{00}.

II. AN ABSTRACT SPACE

In an abstract, infinite-dimensional vector space spanned by an orthogonal set of unit vectors $\mathbf{y}_{lm}$, define the vectors

$$\mathbf{Y}(\theta,\phi) = \sum_{lm} Y_{lm}(\theta,\phi)\,\mathbf{y}_{lm}$$

and

$$\mathbf{A} = \sum_{lm} \alpha_{lm}\,\mathbf{y}_{lm}$$

Then the light source is represented in this space as the fixed vector **A**, and the intensity at a point on the sphere surrounding the source is expressed as the dot product

$$I(\theta,\phi) = \mathbf{A}\cdot\mathbf{Y}(\theta,\phi)$$

Now define the subvectors

$$\mathbf{Y}_l(\theta,\phi) = \sum_{m} Y_{lm}(\theta,\phi)\,\mathbf{y}_{lm}$$

and

$$\mathbf{A}_l = \sum_{m} \alpha_{lm}\,\mathbf{y}_{lm}$$

In this representation, the d_l can be interpreted as the normalized length of the projection of the vector **A** into the subspace l, or alternatively, as the normalized length of $\mathbf{A}_l$:

$$d_l = |\mathbf{A}_l| / |\mathbf{A}|$$

Note that

$$|\mathbf{A}| = \{\sum_{lm} |\alpha_{lm}|^2\}^{1/2} = \sqrt{N} = \sqrt{4\pi}\, I_{rms}$$

Physically, then, d_l^2 represents the rotationally invariant fraction of the total mean-square power distributed in the angular mode l.

Figure II-1 illustrates these relationships. Only three basis vectors ($\mathbf{y}_{00}$, $\mathbf{y}_{10}$ and $\mathbf{y}_{11}$) are used for clarity. Regardless of the orientation of the measuring apparatus, the tip of the vector **A** must lie on a circle of radius d_1 and height d_0; the closer **A** is to vertical, the more uniform is the light source. Furthermore, valid vectors $\mathbf{Y}(\theta,\phi)$ representing points on the measuring sphere must also lie on a horizontal circle of fixed dimensions because the length

$$|\mathbf{Y}_l(\theta,\phi)| = (2l+1)/(4\pi)$$

is a constant independent of θ and ϕ.

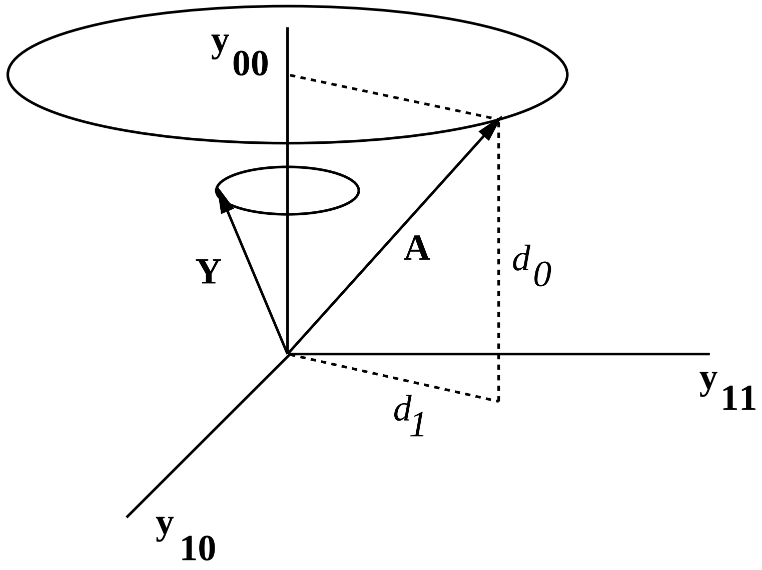

Figure II-1: vectors in an abstract space

In addition to providing an alternative description of the d_l, this representation also provides an amusing shortcut to the calculation of the coefficients α_{lm} in certain special cases. The dot product for $I(\theta,\phi)$ can be written as

$$I(\theta,\phi) = |\mathbf{A}|\,|\mathbf{Y}(\theta,\phi)|\cos(\beta)$$

Using

$$|\mathbf{A}| = \sqrt{4\pi}\,I_{rms}$$

and [JAC 75]

$$|\mathbf{Y}(\theta,\phi)| = \{\sum_{l=0}^{lmax}\sum_{m=-l}^{m=+l} |Y_{lm}(\theta,\phi)|^2\}^{1/2}$$

$$= \{\sum_{l=0}^{lmax}(2l+1)/(4\pi)\}^{1/2}$$

$$= (l_{max}+1)/\sqrt{4\pi}$$

results in

$$I(\theta,\phi) = I_{rms}(l_{max}+1)\cos(\beta)$$

Therefore, in the (admittedly unlikely) event that the dataset happens to have the easily-tested property that

$$I_{max}(\theta_0,\phi_0) = I_{rms}(l_{max} + 1)$$

then $\cos(\beta) = 1$, which means **A** is aligned with $\mathbf{Y}(\theta_0,\phi_0)$. Thus their coefficients must be directly proportional, i.e.

$$\alpha_{lm} = 4\pi Y_{lm}(\theta_0,\phi_0) I_{rms} / (l_{max} + 1)$$

or

$$\alpha_{lm} = 4\pi Y_{lm}(\theta_0,\phi_0) I_{rms}{}^2 / I_{max}$$

So for special distributions, the abstract space model gives the coefficients immediately, without the need for fitting the dataset to spherical harmonic functions.

III. BLOB FITTING DETAILS

III.1 Blobs and Moments

A "blob" is a roughly elliptical, roughly Gaussian 2-dimensional distribution. In a coordinate system in which a perfectly bi-Gaussian blob is centered on the origin with its major and minor axes coincident with the $\bar{x}$- and $\bar{y}$-axes, its intensity distribution can thus be written in the simple form:

$$G(\bar{x},\bar{y}) = A \exp\{-\bar{x}^2/(2\bar{\sigma}_x{}^2)\} \exp\{-\bar{y}^2/(2\bar{\sigma}_y{}^2)\}$$

(In this appendix, coordinates with overlines refer to the rotated and scaled frame, whereas those without are absolute pixel coordinates in the camera frame). The zeroth moment m_{00}, the volume under the bi-gaussian surface, is related to A by a constant factor:

$$m_{00} = A\sqrt{\pi}$$

All of the other m_{ij} are zero for this bi-Gaussian blob.

For a blob that is not exactly bi-Gaussian, but roughly so, m_{00} would still contain most of the shape information, and the other moments would be relatively small.

To separate details of camera orientation and blob location in the camera's field of view from representation of the inherent structure of the blob, we initially define and compute five parameters: the x- and y-coordinates of the center-of-illumination, the tilt-angle of the blob with respect to the x-axis, and the standard deviations of the blob along and perpendicular to the tilt-axis. Figure III-1 shows the coordinate system used for this analysis.

We name these offset, rotation, and distance scale parameters as follows:

The total intensity is:

$$S_{00} = \sum_{xy} S(x,y)$$

where $S(x,y)$ is the intensity measured at pixel (x,y).

The offsets are:

$$x_{00} = \langle x \rangle = \sum_{xy} x\, S(x,y) / S_{00}$$

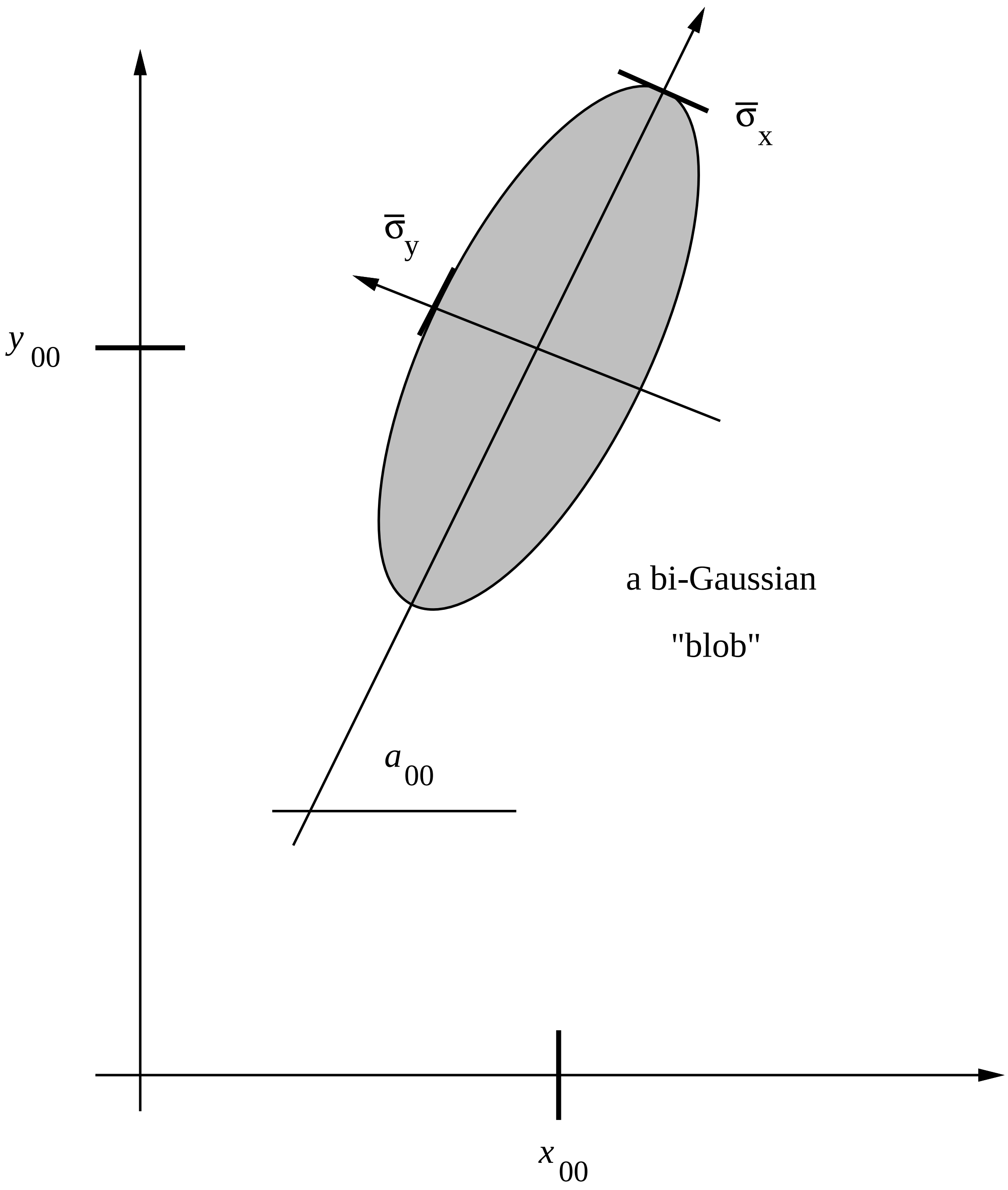

Figure III-1: coordinate system for decomposition of a blob and

$$y_{00} = \langle y \rangle = \sum_{xy} y\, S(x,y) / S_{00}$$

The squared standard deviations before rotating by the tilt angle are:

$$\sigma_{xx} = \langle x^2 \rangle - \langle x \rangle^2 = (\int\int_{-\infty}^{+\infty} dx\, dy\, S(x,y)\, x^2 / S_{00}) - x_{00}{}^2$$

$$\sigma_{yy} = \langle y^2 \rangle - \langle y \rangle^2 = (\int\int_{-\infty}^{+\infty} dx\, dy\, S(x,y)\, y^2 / S_{00}) - y_{00}{}^2$$

and

$$\sigma_{xy} = \langle xy \rangle - \langle x \rangle \langle y \rangle = (\int\int_{-\infty}^{+\infty} dx\, dy\, S(x,y)\, x\, y / S_{00}) - x_{00} y_{00}$$

The formula for the tilt angle is derived by noting that for a bi-Gaussian distribution, the expectation value $\langle \bar{x}\, \bar{y} \rangle$ equals zero in a coordinate system whose origin is located at the central maximum and whose $\bar{x}$-axis is aligned with the major axis of the distribution. Writing this expression in the original, unshifted coordinate system results in

$$\langle \{(x - x_{00}) \cos(a_{00}) + (y - y_{00}) \sin(a_{00})\} \{(x - x_{00}) \sin(a_{00}) - (y - y_{00}) \cos(a_{00})\} \rangle = 0$$

Expanding and using the above definitions gives:

$$(\sigma_{xx} - \sigma_{yy}) \cos(a_{00}) \sin(a_{00}) = \sigma_{xy} (\cos^2(a_{00}) - \sin^2(a_{00})$$

Rearranging and using a standard identity produces the tilt angle:

$$a_{00} = 0.5 \arctan(2\sigma_{xy} / \{\sigma_{xx} - \sigma_{yy}\})$$

After translating, scaling, and rotating the axes, the standard deviations along the major and minor axes of the blob are:

$$\bar{\sigma}_x = (\int\int_{-\infty}^{+\infty} dx\, dy\, S(x,y)\, \bar{x}(x,y)^2 / S_{00})^{1/2}$$

and

$$\bar{\sigma}_y = (\int\int_{-\infty}^{+\infty} dx\, dy\, S(x,y)\, \bar{y}(x,y)^2 / S_{00})^{1/2}$$

where

$$\bar{x} = (x - x_{00}) \cos(a_{00}) + (y - y_{00}) \sin(a_{00})$$

and

$$\bar{y} = (y - y_{00})\cos(a_{00}) - (x - x_{00})\sin(a_{00})$$

Or, in terms of a_{00},

$$\bar{\sigma}_x{}^2 = \sigma_{xx}\cos^2(a_{00}) + 2\sigma_{xy}\sin(a_{00})\cos(a_{00}) + \sigma_{yy}\sin^2(a_{00})$$

and

$$\bar{\sigma}_y{}^2 = \sigma_{xx}\sin^2(a_{00}) - 2\sigma_{xy}\sin(a_{00})\cos(a_{00}) + \sigma_{yy}\cos^2(a_{00})$$

III.2 Harmonic Oscillator Wave Functions

Because the interval over which a blob extends is arbitrary, generating a unique expansion independent of the number of terms [BEV 69, HAM 73] requires the basis functions to be orthogonal over the interval $(-\infty, +\infty)$.

The Gaussian and either its derivatives or its products with simple polynomials have previously been used in the computer vision literature as basis functions for fitting blobs, blurred edges, etc. However it appears that no one in computer vision has yet recognized that the harmonic oscillator wavefunctions are the right way to implement the intent, i.e., to use a set of basis functions that have the requisite mathematical properties *and* whose shapes are well matched to the shape of the data, thereby achieving good fits with a small number of terms.

The harmonic oscillator wavefunctions are:

$$\Phi_n(z) = (1/\sqrt{\pi^{1/2}\, 2^n\, n!})\ \exp(-z^2/2)\, h_n(z)$$

where the $h_n(z)$ are the Hermite polynomials:

$$h_0(z) = 1$$
$$h_1(z) = 2z$$
$$h_2(z) = 4z^2 - 2$$
$$h_3(z) = 8z^3 - 12z$$
$$h_4(z) = 16z^4 - 48z^2 + 12$$
$$\cdots$$
$$h_n(z) = 2z\, h_{n-1}(z) - 2(n-1)\, h_{n-2}(z)$$

They are orthonormal:

$$\int_{-\infty}^{+\infty} dz\, \Phi_i(z)\, \Phi_j(z) = \delta_{ij}$$

and complete:

$$\sum_i \Phi_i(z)\, \Phi_i(z') = \delta(z-z')$$

It is therefore convenient, and it has proven effective, to represent appropriately centered and oriented "blobs" as the coefficients of a Fourier expansion in products of harmonic oscillator functions for each coordinate direction:

$$S(x,y) = \sum_{ij} m_{ij}\, \Phi_i(\overline{x}(x,y)/\overline{\sigma}_x)\, \Phi_j(\overline{y}(x,y)/\overline{\sigma}_y)$$

where $\overline{\sigma}_x$ and $\overline{\sigma}_y$ are the standard deviations in $\overline{x}$ and $\overline{y}$. Using the standard deviation in each direction as the scale factor for distance in that direction ensures that the spatial frequencies contained in $\Phi_i(z)$ are neither so small (the wavelengths so long) that there is danger of undersampling and thus aliasing, nor so large (the wavelengths so short) that there is oversampling and thus loss of computational efficiency.

III.3 Method of Calculating Coefficients

The coefficients m_{ij} are calculated from representative data by

$$m_{ij} = \int\int_{-\infty}^{+\infty} dx\, dy\, S(x,y)\, \Phi_i(\overline{x}(x,y)/\overline{\sigma}_x)\, \Phi_j(\overline{y}(x,y)/\overline{\sigma}_y) \,/\, (\overline{\sigma}_x \overline{\sigma}_y)$$

The values of i and j range from zero to some predetermined limit *mmax*, typically equal to six.

This discrete Fourier integral "dot product" method for calculating the coefficients works well when the data are (in an appropriate sense) evenly distributed over the space in which the basis functions are orthogonal, and when the sum and the integral are really equivalent (in an appropriate sense). The dot product calculations are much faster than least-squares fitting, and (unlike least-squares fitting with non-orthogonal functions) adding higher order coefficients does not change the values of the previously calculated lower order coefficients. However the statistics that emerge as a side effect of the least-squares calculation are of course not obtained "for free" with the dot product.

IV. SPHERE FITTING DETAILS

This appendix describes how difficulties relating to aliasing and nonperiodicity were handled when the blob parameters were fit over the measurement sphere with the α_{lm}^{k}.

IV.1 Noncontinuity

Because the tilt angle a_{00} is calculated by an arctangent (appendix III), it ranges from –90° to +90°, resulting in large discontinuous jumps from one endpoint to the other. Figure IV-1 shows a set of raw tilt angle data in two-dimensional form; the right-left axis with 25 data points is ϕ_n, the front-back axis with 16 data points is θ_n, and the vertical axis is the tilt angle in radians. The spherical fitting routine then spends most of its energy fitting to these nonphysical jumps, sometimes resulting in a poor overall fit. In addition, a problem similar to aliasing can result when regenerating in a jump region. For example, assume two neighboring tilt angle data points are +89° and –89°. A proper fit between these two points should range from +89° to +90°, and then from –90° to –89°. The actual fit, however, will range from +89° down through 0° and then on to –89°, going through precisely the wrong points. Because the accuracy of the fit to the eye depends critically on the tilt of the image, this situation is serious.

The aliasing problem was remedied by adding offsets of ±180° to the tilt angle data whenever necessary to smooth it out. The result of this smoothing procedure is shown in figure IV-2. Reconstructed tilt angles that lie outside of the legal range are then "de-shifted" back into range. One question that remained was how to deal with slight fitting errors near ±90°. Is a reconstructed angle that is just over +90° supposed to be shifted by –180°, or is it merely a slight fitting error for an angle that is actually supposed to be just under +90°? The consequence of such a mistake in the fit angle results in a rotation of the image by 180°. Thus "Shannon sampling" interpolation theory [MAR 91] (appendix V) is used as a consistency check for angles which reconstruct "close" to the legal endpoints to determine whether or not they should be shifted by ±180°. The only drawback is that the interpolation theory requires the retention of the complete original dataset of tilt angles. The *rms* residual of one dataset of tilt angles was reduced by a factor of four by this method.

IV.2 Nonperiodicity

Not all datasets saw an improvement in the residual, however (though the aliasing problem was solved). For example, for a spot quality bulb dataset the *rms* residual increased from 0.10 to 1.12 after adding the offsets. This is because spot quality bulb filament images rotate fairly regularly and the tilt angles thus always go through +90° in the same direction, causing the offsets to accumulate instead of averaging out to zero. This results in the adjusted tilt angle of the last data point differing vastly from the first data point. The 2-D Fourier fit requires continuity of the data in both ϕ and θ or else convergence will be slow; not only will the offsets therefore prevent the data from "wrapping" in ϕ, but the accumulation of this effect will cause a severe discontinuity in θ. The discontinuity in θ is quite apparent in figure IV-2.

In the one-dimensional case, the standard way [HAM 73] to solve the discontinuity problem when using Fourier series is to subtract a line that goes through the first and last data points from the data, and then fitting to the (now continuous) result. Upon reconstruction, the line is added back, requiring the retention of only two extra parameters to specify the line. I extended this trick to the two-dimensional case by subtracting a plane from the data that goes through the first and last data points, and is independent of ϕ. After subtracting this plane, the processed tilt angle data are shown in figure IV-3. Notice the creation of continuity in θ; this decreased the *rms* residual by a factor of six to 0.19. The resulting regenerated fit, after postprocessing out the plane and offsets, is shown in figure IV-4.

Because the tilt angle data still did not wrap in ϕ, more coefficients were used to fit this parameter than any of the other image parameters (345 coefficients versus 49 coefficients for 425 data points). Other parameters such as m_{00} that do not suffer from ϕ-nonperiodicity are easier to fit; the data and fit for a typical m_{00} dataset are shown in figures IV-5 and IV-6. The subtraction of the plane did not tend to change the *rms* residual of the other parameters very much, partly because on physical grounds they were automatically continuous in ϕ. Although not necessarily continuous in θ, the discontinuities were still not as severe as the case of the tilt angle with its accumulating offsets.

A better solution to the problem of fitting a_{00} could be to dispense with fitting it altogether

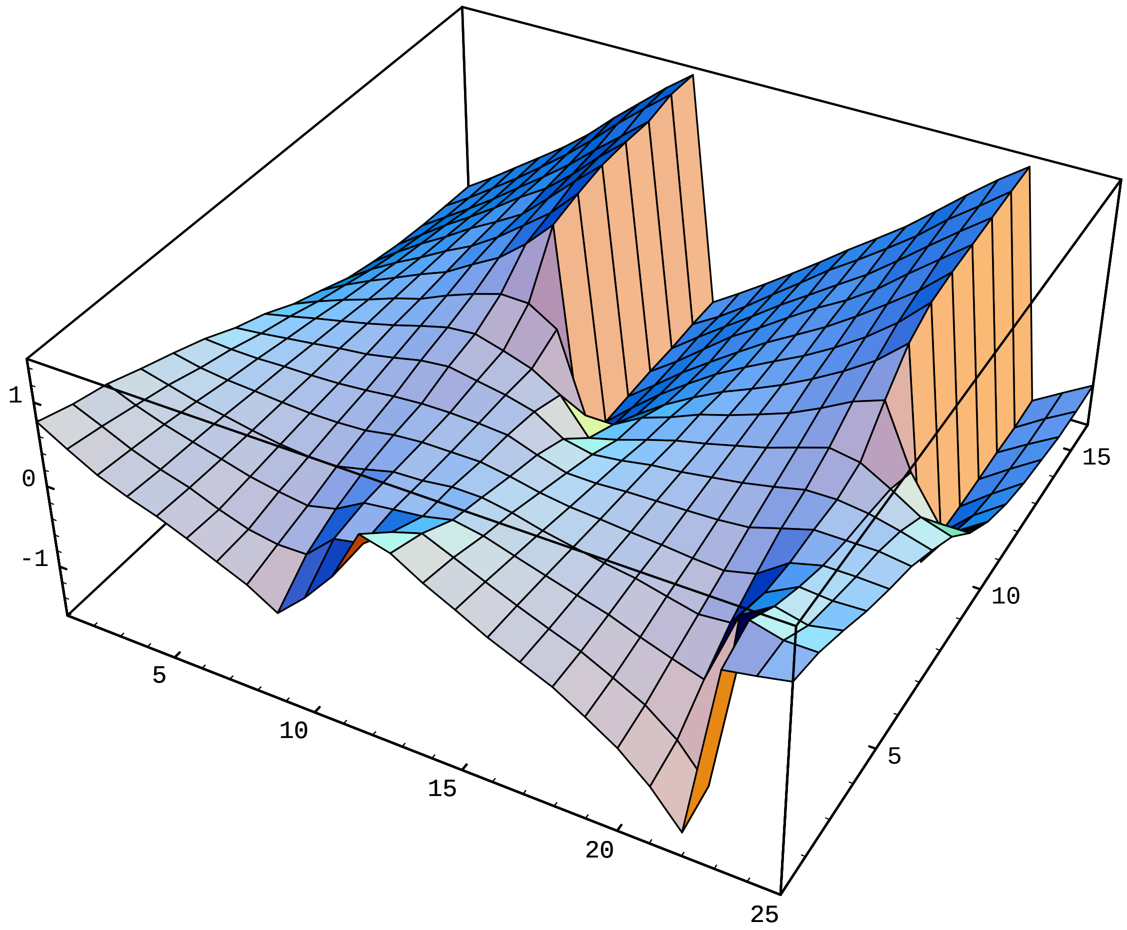

Figure IV-1: raw tilt angle data

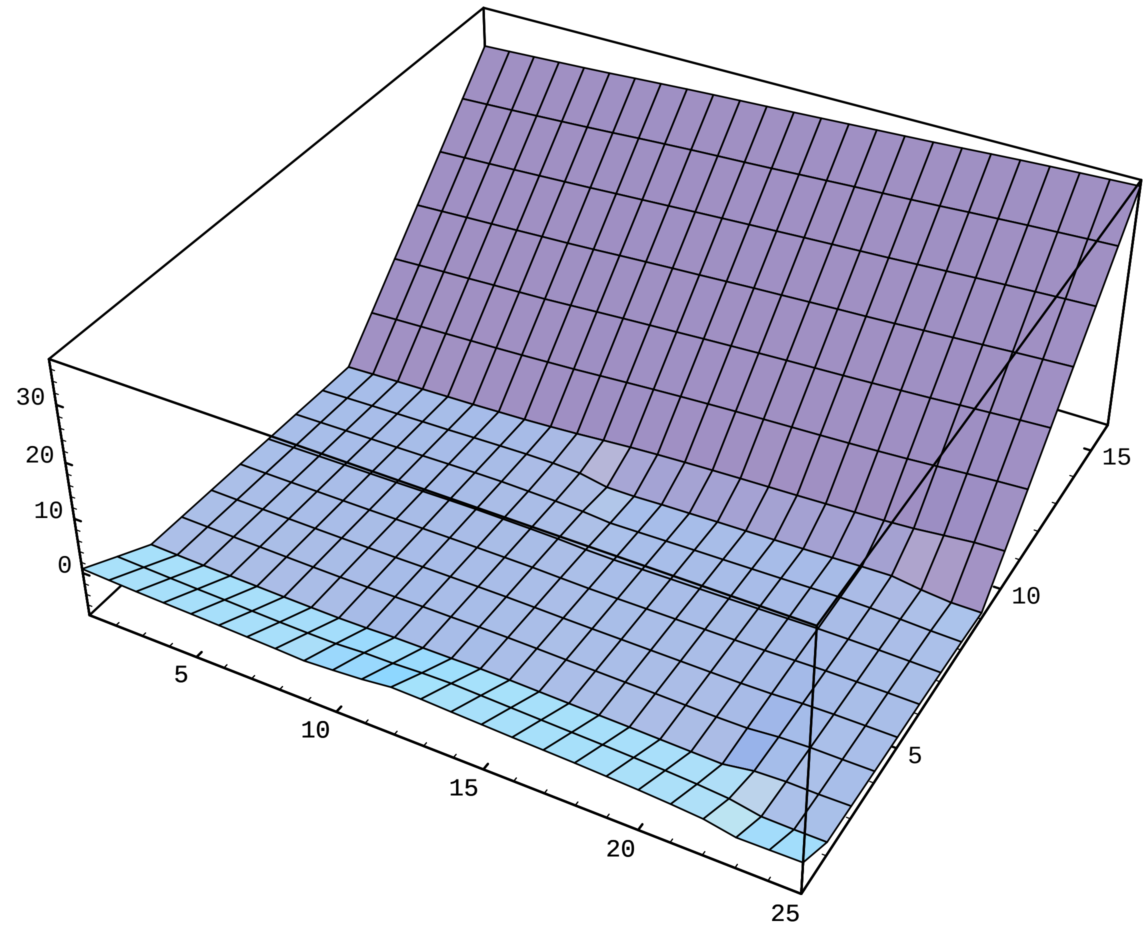

Figure IV-2: tilt angle data smoothed with offsets

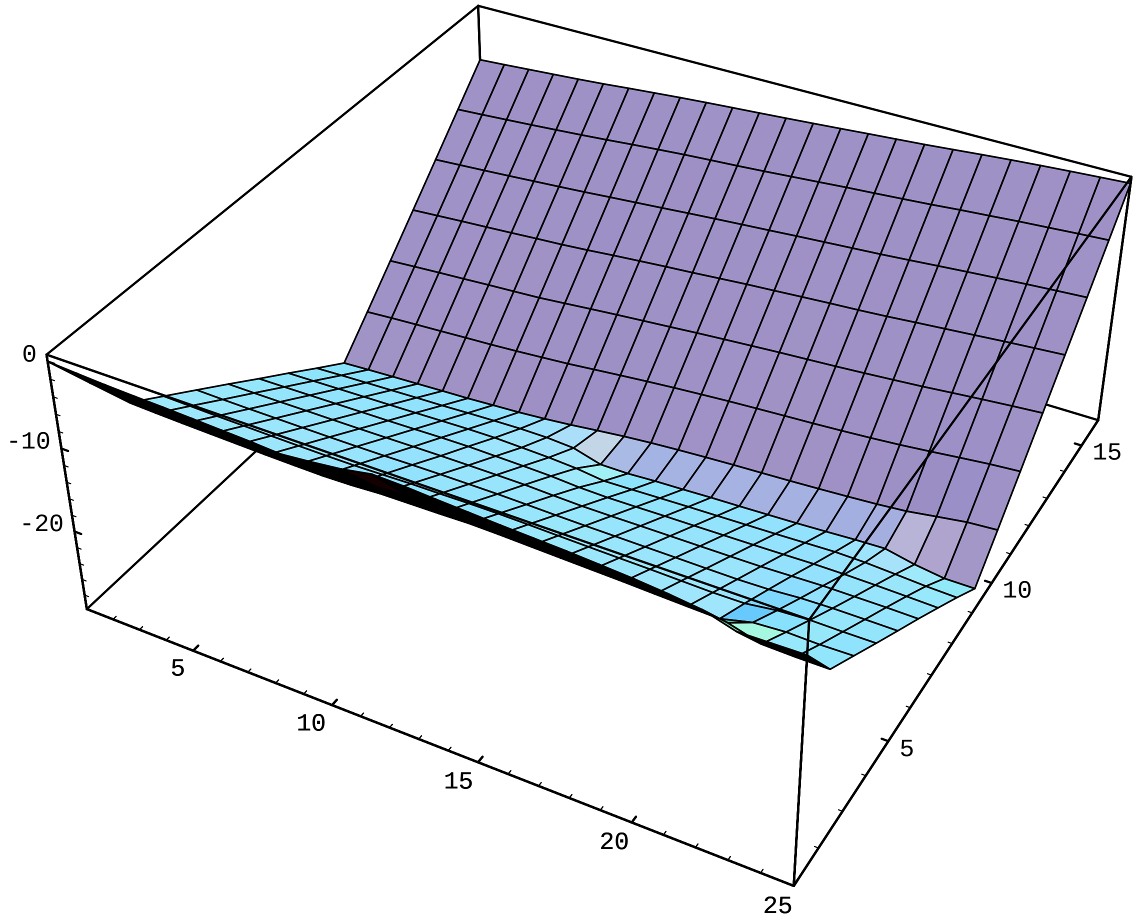

Figure IV-3: tilt angle data with a plane subtracted

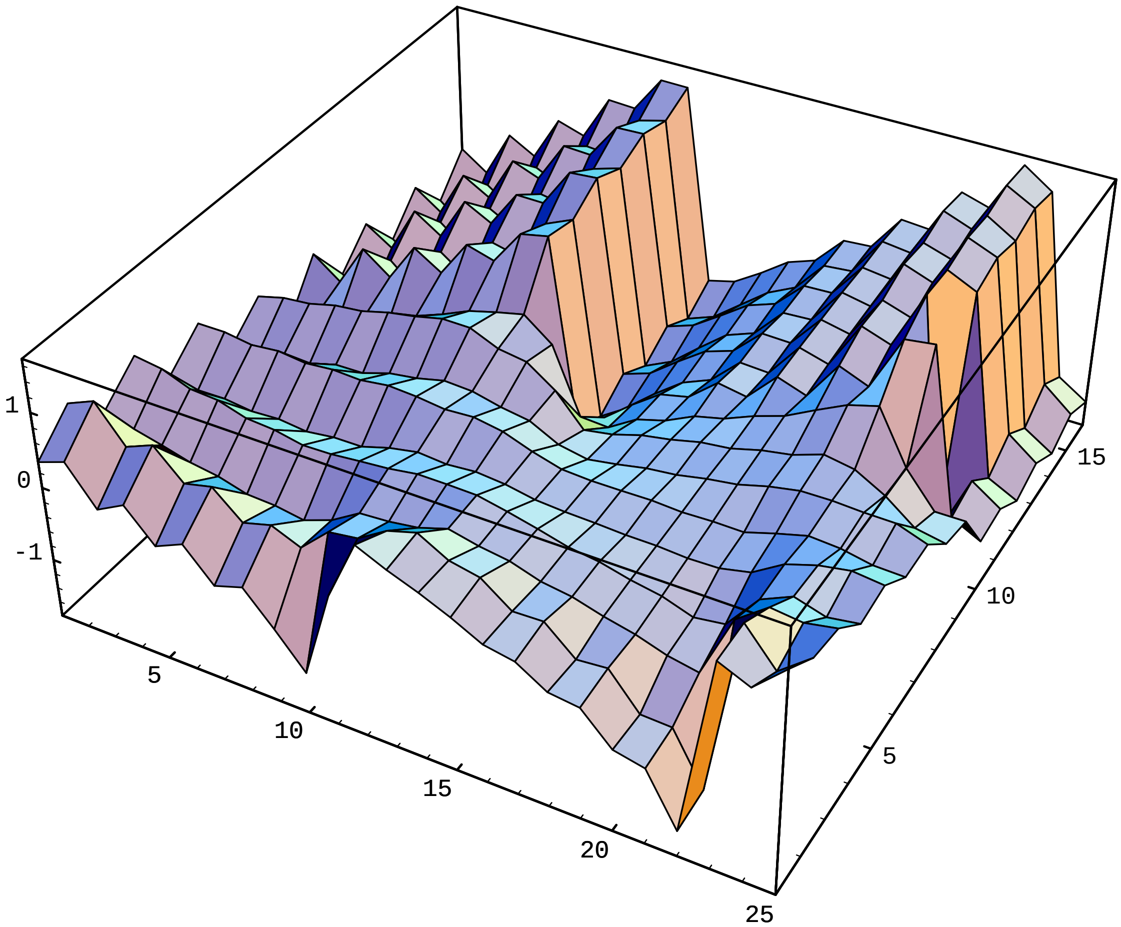

Figure IV-4: regenerated tilt angle data after postprocessing

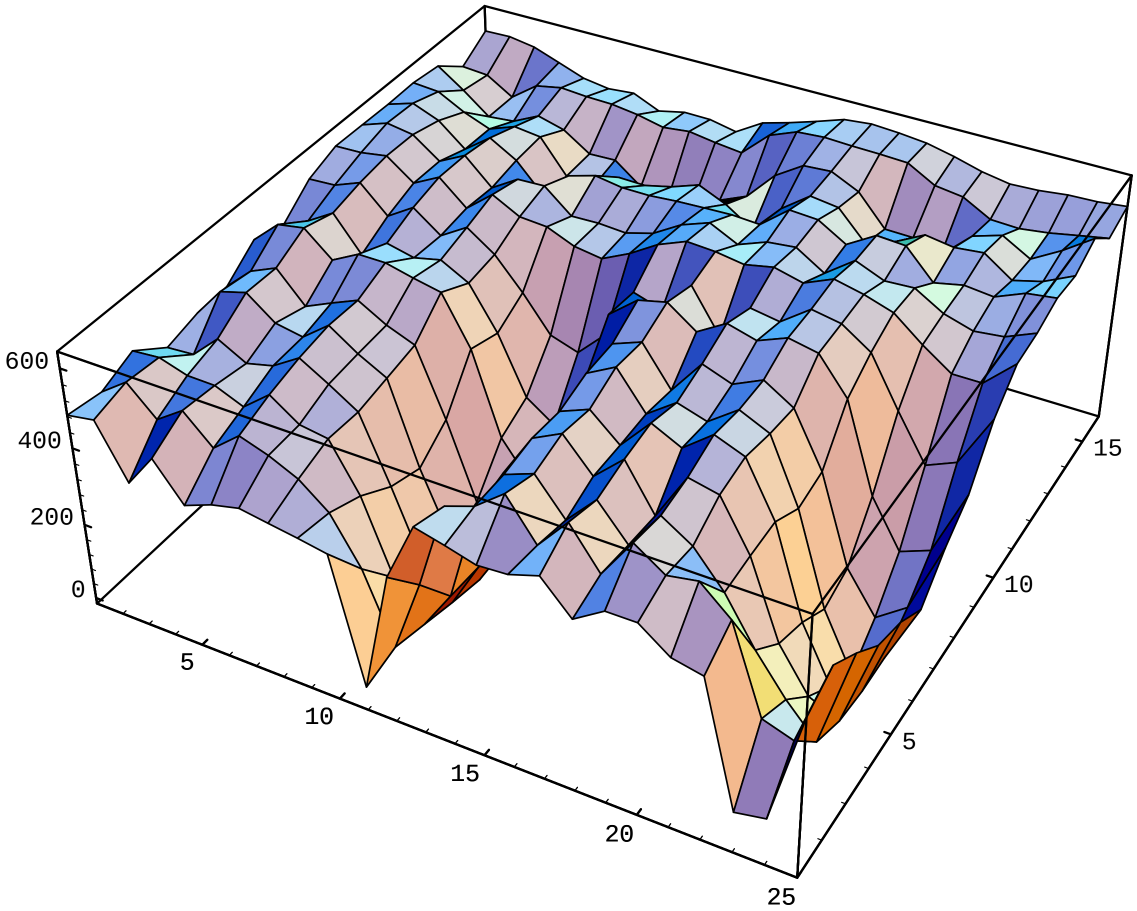

Figure IV-5: typical m_{00} dataset

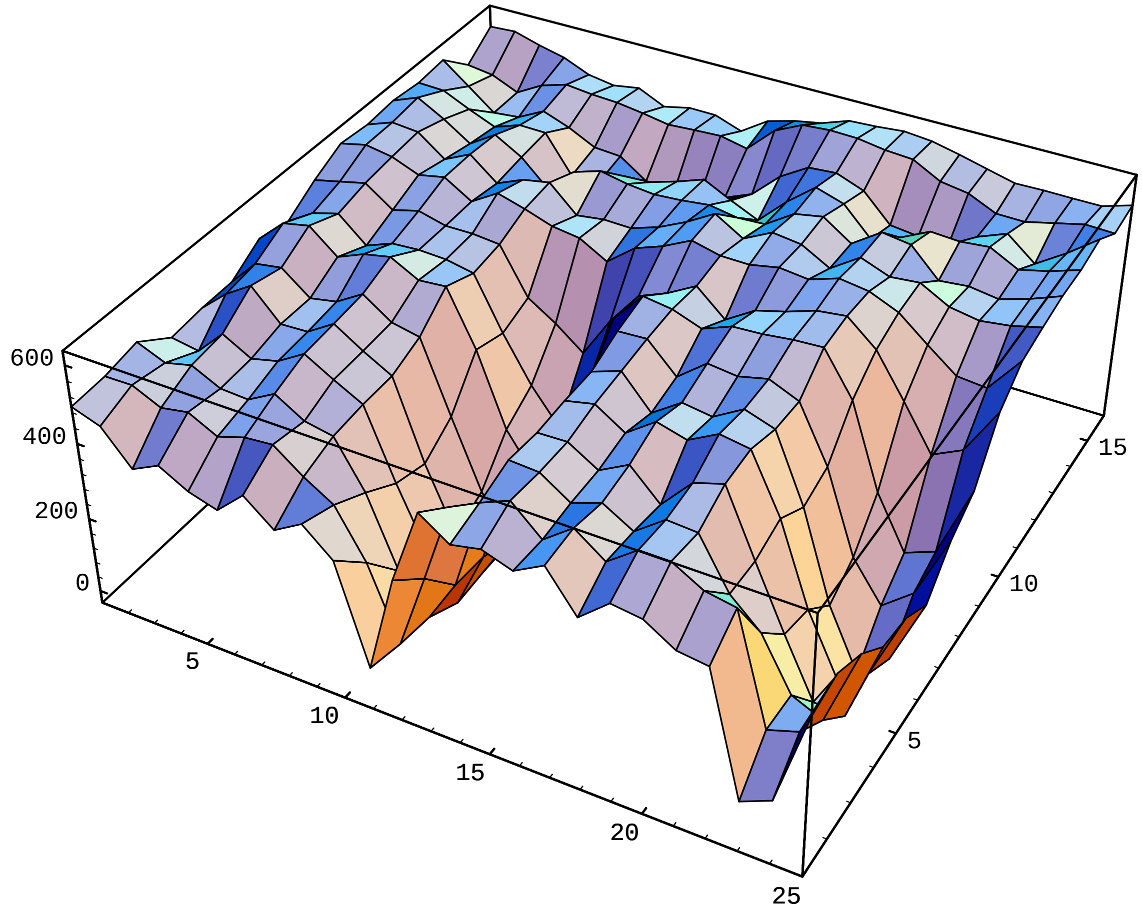

Figure IV-6: regenerated m_{00} dataset since it can be derived from other parameters. For example, it should be possible to fit the set $\{\sigma_{xx},\sigma_{yy},\sigma_{xy}\}$ over the sphere instead of the currently used set $\{a_{00},\bar{\sigma}_x,\bar{\sigma}_y\}$, and then calculate a_{00}, $\bar{\sigma}_x$, and $\bar{\sigma}_y$ from them using the formulas of appendix III.

V. SHANNON SAMPLING: A TUTORIAL

Shannon Sampling [MAR 91] is an interpolation method based on Fourier transforms. Let a signal $x(t)$ have Fourier transform $X(u)$. If $X(u)$ is bandlimited with bandwidth W, that is, $X(u)$ is zero for all $|u| > W$, then the continuous signal $x(t)$ is fully characterized by a finite set of samples:

$$x(t) = \sum_{n=-\infty}^{\infty} x(n/2W)\,\mathrm{sinc}\,(2Wt-n)$$

The samples $x(n/2W)$ of $x(t)$, where n is an integer, are determined by measurement, and the minimum sampling rate to avoid aliasing, $2W$, is called the *Nyquist rate*.

Define $s(t)$ as a function[25] of $x(t)$ such that

$$s(t) = x(t) \sum_{n=-\infty}^{\infty} \delta(t-n/2W)$$

This function is zero except for those t which fall on sampling points $n/2W$, and thus can be written as a function of the samples $x(n/2W)$:

$$s(t) = \sum_{n=-\infty}^{\infty} x(n/2W)\,\delta(t-n/2W)$$

If $x(t)$ can be recovered from $s(t)$, then $x(t)$ can be recovered from its samples.

Because the first form of $s(t)$ is a product of two functions, the Fourier transform $S(u)$ of $s(t)$ is a convolution of the Fourier transforms of the same two functions. Since the transform of a delta functions is another delta function,

$$S(u) = X(u) * \sum_{n=-\infty}^{\infty} 2W\delta(u-2nW)$$

or, using properties of the Dirac delta function,

$$S(u) = 2W \sum_{n=-\infty}^{\infty} X(u-2nW)$$

The effect is to create an infinity of replicas of $X(u)$, each centered at $2nW$ and spaced $2W$ apart. Because $X(u)$ is assumed to be bandlimited with width $2W$, the replicas do not overlap. A

[25]This function $s(t)$ is sometimes referred to as the "sampling" function, but this is misleading. Although it is a function of the samples, it does not represent the physical process of sampling because of the unphysical delta functions.

single replica may be picked out by multiplying $S(u)$ with a square aperture function Π of width $2W$ and centered on the origin:

$$X(u) = S(u)\,\Pi(u/2W)\,/\,2W$$

The Fourier transform of the aperture $\Pi(u/2W)/2W$ is the function $\mathrm{sinc}\,(2Wt)$. Therefore, conversion of $X(u)$ back to the time domain yields the continuous signal $x(t)$ as the convolution of the transforms of $S(u)$ and $\Pi(u/2W)$:

$$x(t) = s(t) * \mathrm{sinc}\,(2Wt)$$

Substituting the second form of $s(t)$ in terms of the samples $x(n/2W)$ and performing the convolution integral gives the desired result

$$x(t) = \sum_{n=-\infty}^{\infty} x(n/2W)\,\mathrm{sinc}\,(2Wt-n)$$

This proof depends on two major points. The first point is the trick of defining the function $s(t)$ so that it can be expressed both in terms of the continuous signal $x(t)$ and of the discrete samples $x(n/2W)$. The second point concerns the lack of aliasing. Aliasing occurs when the replicas of $X(u)$ overlap, which means the sampling rate was too low for the actual bandwidth of the function. Thus a pure $X(u)$ cannot be picked out by the aperture function, and frequencies from other replicas "leak" into the solution.

VI. RENDERING DETAILS

The normal $\boldsymbol{n}$ to each reflector surface patch is found by [HOR 90]

$$\boldsymbol{n} = \{-p, -q, 1\}$$

where

$$p = \frac{dz}{dx}$$

and

$$q = \frac{dz}{dy}$$

Then, using a combination of forwards and backwards specular raytracing, if $\boldsymbol{r}_1$ is the ray from a reflector patch to an image pixel, the corresponding reflected ray is found by

$$\boldsymbol{r}_2 = -\boldsymbol{r}_1 - 2(\boldsymbol{n}\cdot\boldsymbol{r}_1)\,\boldsymbol{n}$$

where all vectors have unit length.

The intersection of a ray from point $\boldsymbol{p}$ in the direction $\boldsymbol{v}$ with a sphere of radius r is found by solving the equation

$$|\boldsymbol{p} + n\boldsymbol{v}| = r$$

for n. This results in the quadratic

$$\boldsymbol{v}\cdot\boldsymbol{v}\,n^2 + 2\boldsymbol{p}\cdot\boldsymbol{v}\,n + \boldsymbol{p}\cdot\boldsymbol{p} - r^2 = 0$$

The reflector is sampled by taking equal areas $A' = \Delta x \Delta y$ in the xy plane. Because this area is the projection of the area A_r of the reflector patch onto this plane, assuming the patch is small enough to be considered locally planar,

$$A_r = A' / \cos\beta$$

where β is the angle between the reflector normal $\boldsymbol{n}$ and vertical. So $\cos\beta = 1/(p^2 + q^2 + 1)^{1/2}$ and

$$A_r = A'(p^2 + q^2 + 1)^{1/2}$$

Thus the intensity striking image pixel (x',y') due to the reflector element at (x,y), $\Delta I'(x',y';x,y)$, can be found by regenerating the brightness B of the source along this ray from the source intersection coordinates and the Fourier coefficients $\{\alpha^k_{lm}\}$, and then using

$$\Delta I'(x',y';x,y) = B A_i(x',y') A_r(x,y) / d(x',y';x,y)^2$$

where $A_i(x',y')$ is the area of the image plane pixel at (x',y',D), $A_r(x,y)$ is the area of the reflector element at $(x,y,z(x,y))$, and $d(x',y';x,y)$ is the distance between them. Finally, the total intensity

$I'(x',y')$ received by image pixel (x',y') due to all of the mirror elements is

$$I'(x',y') = \sum_{xy} \Delta I'(x',y';x,y)$$

This assumption of local planar behavior for small enough patches was tested by comparing the total sampled area with an analytical result for the area of a pure paraboloid of revolution truncated above $z = 0$. The correct total area is found by analytic surface integration to be $8\pi(\sqrt{8} - 1)/3 f^2 = 15.32 f^2$. The percent errors for various step sizes $\Delta x = \Delta y$ measured in units of f are shown in Table VI-1.

percent error				
step	0.5	0.25	0.125	0.05
%	2.93	2.38	1.06	0.22

Figure VI-1: percent error of reflector surface area versus sampling size

VII. FIRST-GUESS ALGORITHM

In outline, the method works by comparing the shape of the perimeter of the image cast by an unperturbed parabolic reflector with that of the desired image. The change in radius between the two images at a given angle around the perimeter is attributed entirely to the change in slope of the corresponding reflector element at the parabolic reflector's edge. The procedure is to start from the center of each image, move out along corresponding radial lines, and look for the first edge (or second edge in the case of the twin-peaked example of section 4.6.2). The required slope can be written in terms of the coordinates of the appropriate reflector edge element, and hence in terms of the perturbation coefficients. Considering several image radii and reflector edge elements, for small perturbations a set of overdetermined linear equations results, which can be solved for the most consistent set of perturbation coefficients by the method of least-squares.

The perimeters are found with the Roberts edge detector formula [ROB 65, PRA 78]. An edge occurs at $(x+1/2, y+1/2)$ when the function of image intensity

$$g(x,y) = |I(x,y) - I(x+1,y+1)| + |I(x+1,y) - I(x,y+1)|$$

is a maximum. Subpixel interpolation can be used to refine the edge calculation. Operating on the image of figure VII-1, for example, the Roberts edge detector found the edge points shown in figure VII-2.

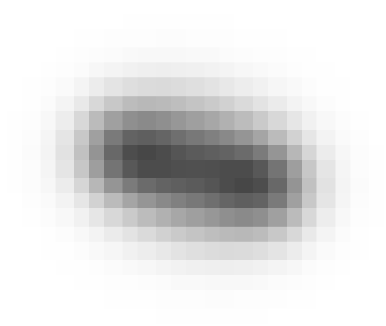

Figure VII-1: a sample image

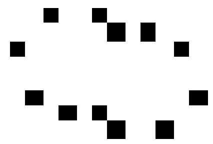

Figure VII-2: twelve edge points

The full superimposed edges of the desired and initial images might look like the perimeters shown in figure VII-3. Two pairs of corresponding radii are shown.

Figure VII-4 shows how the change in perimeter radius along a given line is attributed to the

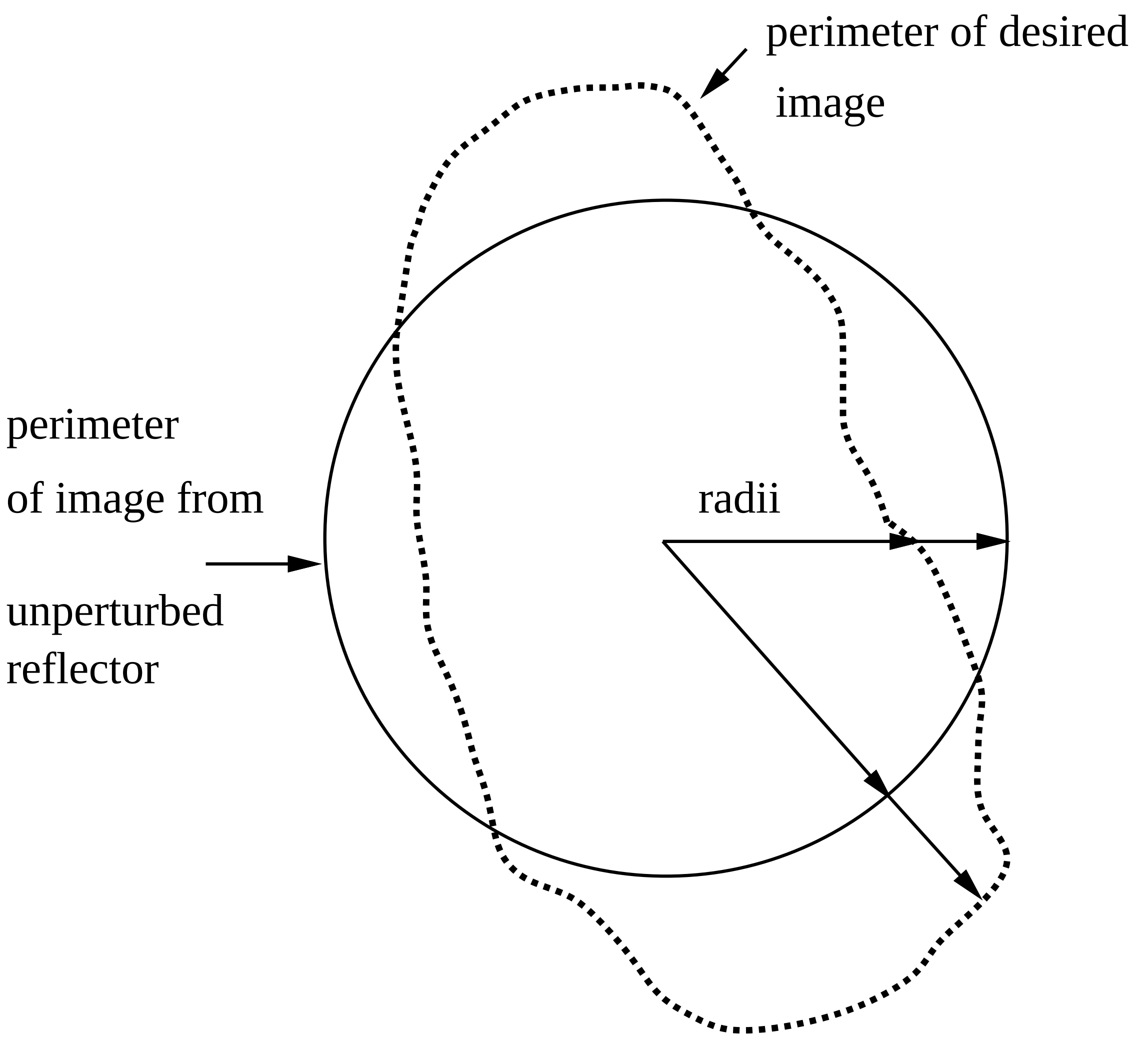

Figure VII-3: comparing radii of image perimeter edges

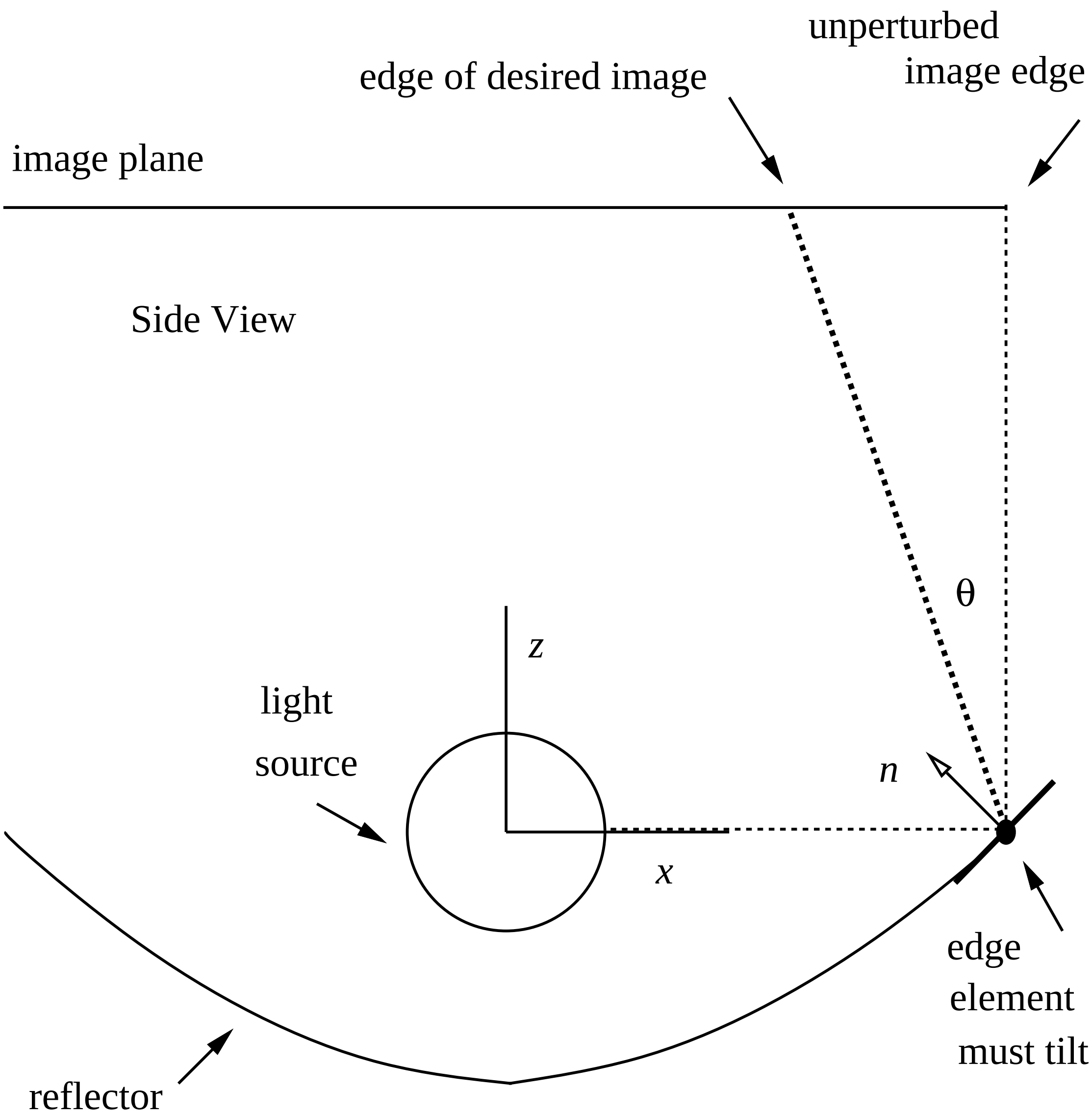

Figure VII-4: image edge shape roughly corresponds with reflector edge tilt

tilt of an edge element of the reflector. The z axis corresponds with the axis of the parabolic reflector, as explained in section 4.5 (in the figure the z axis happens to be perpendicular to the image plane, but this is not necessary). The new angle θ from the unit vector $\boldsymbol{z}$ to the required normal vector $\boldsymbol{n'}$ of the edge of the reflector at this point is easily calculated once the coordinates of the image edge points are known, and thus

$$\boldsymbol{n'} \cdot \boldsymbol{z} = 1 = (p^2 + q^2 + 1)^{1/2} \cos\theta$$

using the definitions of p and q from appendix VI. Expanding p^2 and q^2 and assuming the perturbations are small, this can be written as

$$2/\cos^2\theta - 4 = l_1 x + l_2 y + 2a_1(x^2 - y^2) + 2a_2 xy + \\ 3c_1 x^3 + 3c_2 y^3 + 3c_3 x^2 y + 3c_4 xy^2 + 2r(x^2 + y^2)^{1/2}$$

using the fact that $(x^2 + y^2)/4 = 1$ around the rim of the parabolic reflector and eliminating all terms that are second order or higher in the perturbation coefficients. The extension to the use of additional terms in the reflector model is straightforward.

For several radius pairs, each yielding an x, y, and θ, a system of equations linear in the perturbation coefficients results, which can be quickly solved by a least squares fit. If the perturbations are indeed small, the guess often performs remarkably well. For example, for a desired image with an exact solution of $l_1 = -5.0\text{x}10^{-4}$ and $a_1 = -3.0\text{x}10^{-5}$, the guess returns $l_1 = -5.3\text{x}10^{-4}$ and $a_1 = -2.6\text{x}10^{-5}$, reducing the residual from 36.7 (for a vertically oriented pure paraboloid) to 2.0. If the axis of the pure parabolic mirror is oriented to point at this image's center of intensity, as outlined in section 4.5, the residual immediately drops to 7.9. Applying the guess to this tilted reflector results in $l_1 = -1.7\text{x}10^{-5}$ and $a_1 = -3.0\text{x}10^{-5}$, reducing the residual to 5.1. This is not as good as the original guess in the vertical frame because an exact solution does not exist in the tilted reference frame. Note how the guess for the linear coefficient l_1 in the tilted frame decreases by an order of magnitude; this is because the linear term mainly represents a shift of the image position, which is largely taken care of simply by pointing the reflector the right way. If the perturbations are not in fact small, the guess may not be very useful however.

Trying to guess too many terms (typically eight) by the least-squares method often results in a small pivot and a spurious "inverse" matrix. This is because the dataset cannot distinguish between the contributions of perturbations with similar symmetry. It is therefore necessary to use the covariance matrix to calculate the relative uncertainty of each term [BEV 69], and then guess only the most significant terms, varying the number of data points used until the best overall fit can be found. In practice, two to three times as many data points as coefficients resulted in the best guesses.

VIII. RELATED FIELDS

VIII.1 Potential Scattering

The inverse scattering problem, in which scattering data are used to reconstruct an object, has received much attention in the field of Quantum Mechanics. Attempts have been made to extend this method to the realms of acoustic and electromagnetic scattering. Could the problem of designing a mirror be cast in the same terms, in which the scattering source is the light source, the unknown scattering potential is the reflector, and the scattering data is the desired intensity pattern?

Several papers have discussed the reconstruction from electromagnetic scattering data of the shape of a smooth reflecting object [POR 82], or of the index of refraction as a function of position [EFT 82]. The wave equation describing the incident field can be written as a Schrodinger-like equation with a frequency-dependent potential, to which the developed machinery of quantum mechanical scattering can be applied. Typically, however, the solutions depend on simplifying assumptions. Some success has been achieved experimentally, for example, in reconstructing a composite rectangular transmitting object in the Born (weak scattered field) approximation, using a laser for illumination [CAR 82].

Applying these methods to reflector design, however, seems to encounter insurmountable problems. For instance, without simplifying geometry, the scattered field must be known for several illumination directions separately; as posed, the current problem concerns the total scattered light for all illumination directions mixed together, and there is no way to separate out the sub-illumination patterns due to each illumination direction from the source. Furthermore, the reflector design problem is stated in terms of geometric optics (intensity), but the scattering theory uses concepts from physical optics (phase and amplitude); there simply is no phase information in the desired illumination pattern.

VIII.2 Adaptive Optics

Adaptive optics is the field concerned with modifying the optics of a system in real time to make images sharper [TYS 91]. Its main application is for the correction of atmospheric turbulence in telescopic images. Often the adjustable optic element is a mirror. For instance, a typical system in a large telescope employs an interferometric sensor to measure the slope of the income wavefront. The shape of the (constant phase) wavefront is reconstructed, and a mirror is deformed such that upon reflection, the undistorted plane-wave wavefront is restored, enhancing the resolution of the image.

Can these principles be applied in reverse, to find the shape of a mirror such that the outgoing light from a nearby source takes on a desired wavefront shape far from the mirror? That is, can we substitute "desired illumination pattern" for "incoming wavefront shape"? In general the answer would seem to be no, based on the following problems. Adaptive optics seems to assume, for instance, that the incoming light in the tube of the telescope is travelling parallel to the telescope's axis, and that therefore a one-to-one correspondence exists between wavefront patches and reflector segments. In the reflector design problem, however, the light at any point on a screen has contributions from many reflector elements, because the source is not a point. Making a correspondence between a nonuniform source and the telescopic image is even less clear, unless the light source could be approximated as a point like the ideal image of a star.

VIII.3 Analytical Results

Analytic solutions for the shape of a specular reflector have been obtained under certain simplifying conditions [WIN 91, GOR 92, RAB 94]. For instance, the source must be of uniform radiance, so the intensity at any illumination patch due to some mirror element depends only on the solid angle subtended by the element as seen from the patch; the image plane must be far enough away from the reflector so that this solid angle depends only on the orientation of the element (distance is essentially constant); and there exists a smooth one-to-one light-transfer correspondence between reflector elements and image patches. With these assumptions, a parametric equation can be written for the curve of the mirror in order for the reflector elements to have the right slope to send the required amount of light to their corresponding image patches.

These assumptions are not valid for the general problem of reflector design with extended, coiled-filament sources, with non-uniform radiance distributions.

VIII.4 Shape From Shading

"Shape from shading" is a method by which the shape (surface heights or gradients) of an illuminated object can be recovered from a single image [HOR 90, HOR 77, HOR 81]. As in section VIII.3, the fundamental assumption is that the radiance from a surface element depends only on its orientation relative to the viewer. The problem of achieving a one-to-one mapping from surface elements to reflector patches can be solved by using an optical system such as a pinhole camera to produce the image. The surfaces are usually assumed to have Lambertian reflectance properties, but this is not an absolute restriction. Even so, the fact that the reflector design problem simply uses the aggregate image cast by a luminaire on a screen precludes the possibility of assuming the necessary one-to-one correspondance between reflector elements and image patches. Also, the closeness and nonuniformity of the light source fail to satisfy the position independent radiance assumption of the reflector elements. Essentially, shape from shading in general scenes is more difficult than in "diffuse-only" scenes. Our method can perhaps be regarded as an example of shape from shading where the light source and receiver are known, but the specular reflector is unknown.

IX. APPARATUS

Figure IX-1 displays a schematic of the goniometer apparatus and details of the two "detector" configurations. The source is attached to the end of a rotating horizontal shaft, which holds the source above the center of a rotating horizontal platter. The shaft is also connected to the platter. The detector is placed on an independent fixed support above the platter at a controllable radius ranging between 2.22 and 5.22 inches. The goniometer operates by turning the source on the shaft in preset steps through ϕ, and collecting data at each step. Once the shaft makes a complete revolution, the platter rotates through a step in θ, moving the source with respect to the detector. The "detector" can be either the lens-stop-photodiode combination, or the pinhole CCD camera. The different headlight bulb types are shown in figure IX-2. In order, they are the spot quality, high beam, low beam, and dual beam bulbs.

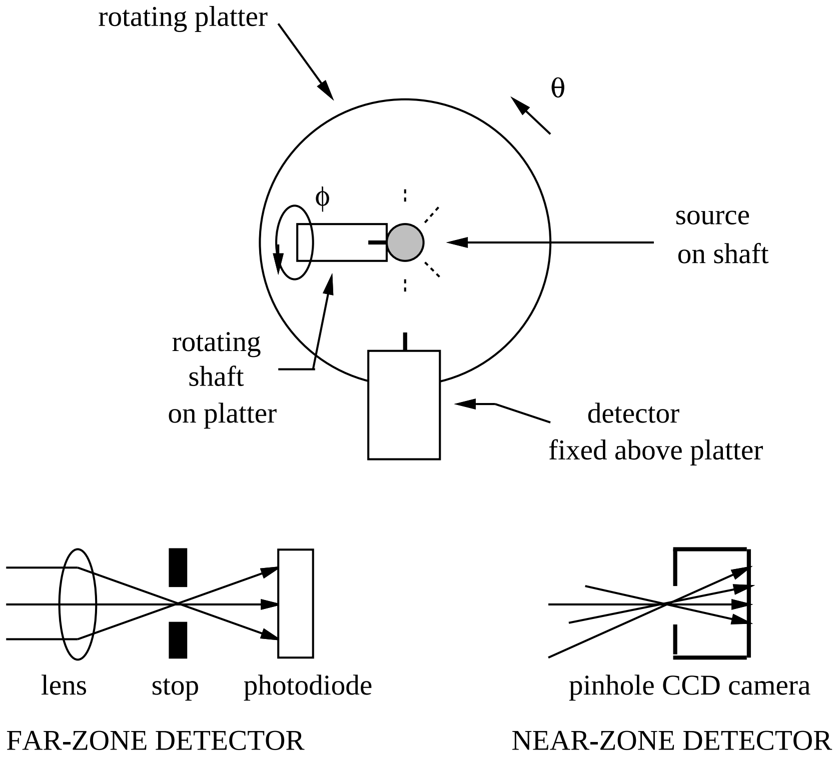

Figure IX-1: a sketch of the goniometer

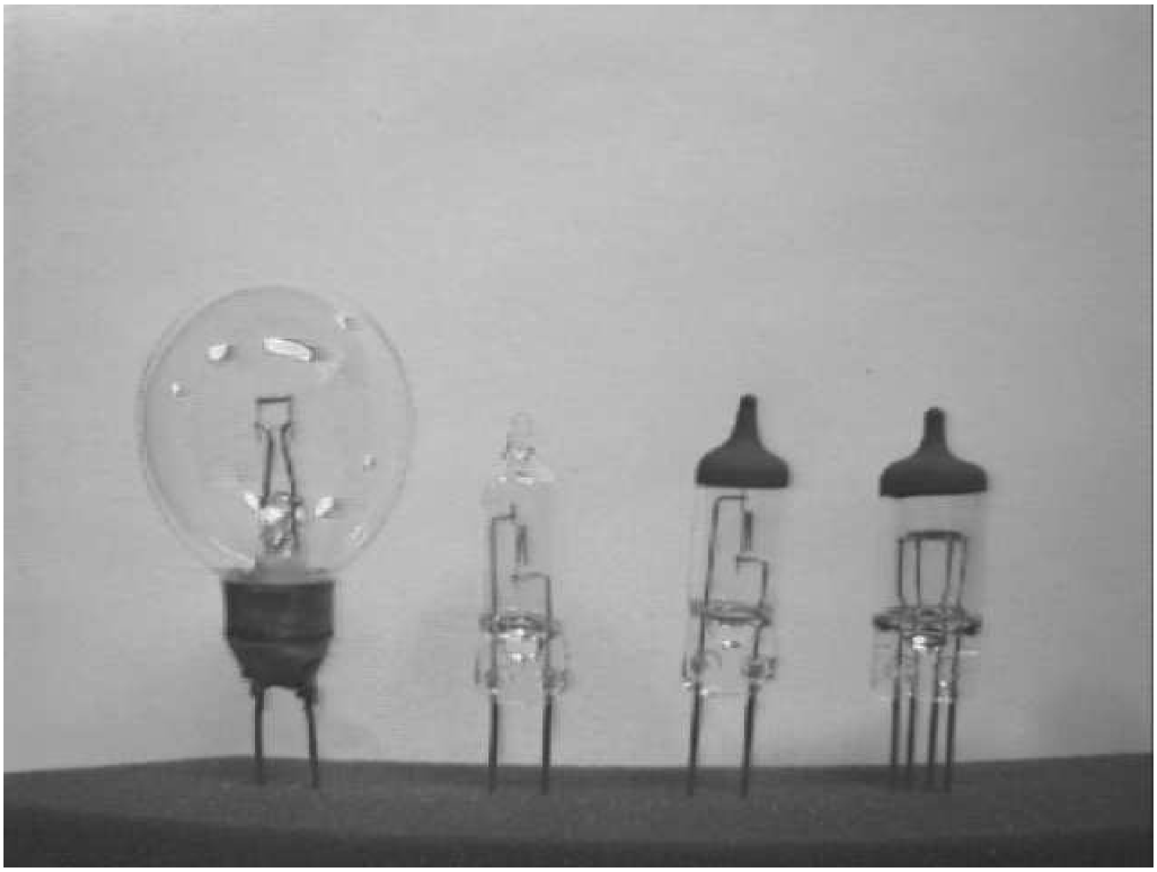

Figure IX-2: spot quality, high, low, and dual beam bulbs

References

[ACT 90] F. S. Acton.
Numerical Methods That Work.
Mathematical Association of America, Washington, D.C., 1990.

[ARM 90] J. D. Armstrong.
A New Measure of Uniformity for Lighting Installations.
Journal of the Illuminating Engineering Society 19(2):84-89, 1990.

[ASH 93] I. Ashdown.
Near-Field Photometry: A New Approach.
Journal of the Illuminating Engineering Society 22(1):163-180, 1993.

[ASH 94] I. Ashdown.
Non-Imaging Optics Design Using Genetic Algorithms.
Journal of the Illuminating Engineering Society 23(1):12-21, 1994.

[BEV 69] P. R. Bevington.
Data Reduction and Error Analysis for the Physical Sciences.
McGraw-Hill, New York, 1969.

[BOR 70] M. Born and E. Wolf.
Principles of Optics.
Pergamon, New York, 1970.

[BOY 83] R. W. Boyd.
Radiometry and the Detection of Optical Radiation.
Wiley, New York, 1983.

[BRE 71] R. P. Brent.
Algorithms for Finding Zeros and Extrema of Functions Without Calculating Derivatives.
Technical Report, Stanford University, 1971.

[BRE 73] R. P. Brent.
Algorithms for Minimization Without Derivatives.
Prentice Hall, Inc., Englewood Cliffs, New Jersey, 1973.

[CAR 82] W. H. Carter.
Inverse Scattering in the First Born Approximation.
In W. H. Carter (editor), *Applications of Mathematics in Modern Optics*, pages 122. SPIE, 1982.

[DON 72] R. J. Donohue and B. W. Joseph.
Computer Design of Automotive Lamps With Faceted Reflectors.
Journal of the Illuminating Engineering Society 2:36-42, October, 1972.

[DON 75] R. J. Donohue and B. W. Joseph.
Computer Synthesized Filament Images From Reflectors and Through Lens Elements for Lamp Design and Evaluation.
Applied Optics 14(10):2384-2390, 1975.

[EFT 82] C. Eftimiu and P. L. Huddleston.
Reconstruction of the Index of Refraction of a Sphere From Scattering Phase Shifts at One Frequency.
In W. H. Carter (editor), *Applications of Mathematics in Modern Optics*, pages 156. SPIE, 1982.

[ELM 74] W. B. Elmer.
Optical Design of Reflectors.
published by author, Andover, MA, 1974.

[FLE 65] R. Fletcher.
Function Minimization Without Evaluating Derivatives - A Review.
Computer Journal 8:33-41, 1965.

[FOR 77] G. E. Forsythe.
Computer Methods for Mathematical Computations.
Prentice-Hall, Englewood Cliffs, NJ, 1977.

[GER 39] J. Gershun.
The Light Field.
Journal of Mathematics and Physics 18:51, 1939.

[GLA 89] A. S. Glassner (editor).
An Introduction to Ray Tracing.
Academic, London, 1989.

[GOL 70] G. H. Golub and C. Reinsch.
Singular Value Decompsition and Least Squares Solutions.
Numerische Mathematik 14:403-420, 1970.

[GOR 92] J. M. Gordon, P. Kashin, and A. Rabl.
Nonimaging Reflectors for Efficient Uniform Illumination.
Applied Optics 31:6027, 1992.

[HAL 89] R. Hall.
Illumination and Color in Computer Generated Imagery.
Springer-Verlag, New York, 1989.

[HALS 93] C. P. Halsted.
Brightness, Luminance, and Confusion.
Information Display :21, March, 1993.

[HAM 73] R. W. Hamming.
Numerical Methods for Scientists and Engineers.
McGraw-Hill, New York, 1973.

[HEC 91] P. S. Heckbert.
Simulating Global Illumination Using Adaptive Meshing.
Technical Report UCB/CSD 91/636, University of California, Berkeley, 1991.

[HOR 77] B. K. P. Horn.
Understanding Image Intensities.
Artificial Intelligence (8), 1977.

[HOR 81] B. K. P. Horn.
Hill Shading and the Reflectance Map.
Proceedings of the IEEE 69(1), January, 1981.

[HOR 90] B. K. P. Horn.
Height and Gradient from Shading.
International Journal of computer Vision 5(1), 1990.

[IES 85] IES.
Photometric Testing of Indoor Fluorescent Luminaires.
Technical Report IES LM-41-1985, Illuminating Engineering Society of North America, New York, 1985.

[JAC 75] J. D. Jackson.
Classical Electrodynamics.
John Wiley & Sons, New York, 1975.

[JOY 74] W. B. Joyce.
Classical-Particle Description of Photons and Phonons.
Physical Review D 9(12):3234, 1974.

[KAH 88] D. Kahaner.
Numerical Methods and Software.
Prentice Hall, Englewood Cliffs, NJ, 1988.

[LAM 60] Lambert.
Photometria siva de mensura et gradibus luminis, colorum et umbrae.
1760.

[LAN 69] L. D. Landau and E. M. Lifshitz.
Statistical Physics.
Addison-Wesley, Reading, MA, 1969.

[LEV 71] R. E. Levin.
Photometric Characteristics of Light-Controlling Apparatus.
Illuminating Engineering 66(4):205-215, 1971.

[LEW 90] I. Lewin, R. Laird, and B. Carruthers.
Development of New Photometer Concepts for Quality Control Applications.
Journal of the Illuminating Engineering Society 19(2):90-97, 1990.

[MAR 91] R. J. Marks.
Introduction to Shannon Sampling and Interpolation Theory.
Springer-Verlag, New York, 1991.

[MAT 89] J.-P. Mathieu.
Statistical Uniformity, A New Method Of Evaluation.
Journal of the Illuminating Engineering Society 18(2):76-79, 1989.

[MIS 90] R. G. Mistrick and C. R. English.
Study of Near-Field Indirect Lighting Calculations.
Journal of the Illuminating Engineering Society 19(2):103-112, 1990.

[MOO 38] P. Moon.
Basic Principles of Illumination Calculations.
Journal of the Optical Society of America 29:108, 1938.

[MOO 40] P. Moon.
On Interreflections.
Journal of the Optical Society of America 30:195, 1940.

[MOO 43] P. Moon.
New Methods of Calculating Illumination.
Journal of the Optical Society of America 33:115, 1943.

[MOO 81] P. Moon and E. Spencer.
The Photic Field.
MIT Press, Cambridge, Mass., 1981.

[MYO 82] O. Myodo and M. Karino.
A New Method for Computer Aided Design of Luminaire Reflectors.
Journal of the Illuminating Engineering Society 11(1):98-105, 1982.

[NGA 87] P. Y. Ngai.
On Near-Field Photometry.
Journal of the Illuminating Engineering Society 16(2):129-136, 1987.

[NGA 92] P. Y. Ngai et al.
Near-Field Photometry: Measurement and Application for Fluorescent Luminaires.
Journal of the Illuminating Engineering Society 21(2):68-83, 1992.

[NIC 63] F. E. Nicodemus.
Radiance.
American Journal of Physics 31:368, 1963.

[PAG 89] T. D. Page, R. S. Bergman, and M. R. Vukcevich.
A Computational First Principles Approach to Incandescent Lamp Design.
Journal of the Illuminating Engineering Society 18(2):109, 1989.

[POR 82] R. P. Porter.
Inverse Scattering for Smooth Reflecting Objects.
In W. H. Carter (editor), *Applications of Mathematics in Modern Optics*, pages 134. SPIE, 1982.

[POW 64] M. J. D. Powell.
An Efficient Method for Finding the Minimum of a Function of Several Variables Without Calculating Derivatives.
Computer Journal 7:155-162, 1964.

[POW 70] M. J. D. Powell.
A Survey of Numerical Methods for Unconstrained Optimization.
SIAM Review 12(1):79-97, January, 1970.

[PRA 78] W. K. Pratt.
Digital Image Processing.
John Wiley & Sons, New York, 1978.

[PRE 88] W. H. Press et al.
Numerical Recipes in C.
Cambridge University Press, Cambridge, 1988.

[RAB 94] A. Rabl and J. M. Gordon.
Reflector Design for Illumination with Extended Sources: the Basic Solutions.
Applied Optics 33:6012, 1994.

[RIE 82] H. Ries.
Thermodynamic Limitations of the Concentration of Electromagnetic Radiation.
Journal of the Optical Society of America 72(3):380, March, 1982.

[ROB 65] L. G. Roberts.
Machine Perception of 3-Dimensional Solids.
In J. T. Tippet et al. (editors), *Optical and Electro-optical Information Processing*, pages 159-197. MIT Press, Cambridge, MA, 1965.

[ROG 91] D. Rogers.
State of the Art in Computer Graphics.
Springer-Verlag, New York, 1991.

[ROS 57] M. E. Rose.
Elementary Theory of Angular Momentum.
Wiley, New York, 1957.

[SHI 91] P. S. Shirley.
Physically Based Lighting Calculations for Computer Graphics.
PhD thesis, University of Illinois at Urbana-Champaign, 1991.

[SPA 63] E. M. Sparrow.
On the Calculation of Radiant Interchange Between Surfaces.
In W. Ibele (editor), *Modern Developments in heat Transfer*, pages 375-380. Academic Press, Inc., New York, 1963.

[TOL 62] G. P. Tolstov.
Fourier Series.
Prentice Hall, Inc., Englewood Cliffs, NJ, 1962.

[TYS 91] R. K. Tyson.
Principles of Adaptive Optics.
Academic Press, Inc., San Diego, 1991.

[VAR 88] D. A. Varshalovich et al.
Quantum Theory of Angular Momentum.
World Scientific Publishing Co Pte Ltd, Singapore, 1988.

[VER 84] C. Verbeck and D. Greenberg.
A Comprehensive Light-Source Description for Computer Graphics.
IEEE Computer Graphics and Applications 4:66-75, July, 1984.

[WAL 65] J. W. T. Walsh.
Photometry.
Dover Publications, New York, 1965.

[WAT 91] T. Watanabe and H. Kawashima.
Development of Smoothly Curved Reflector for Headlamps.
In D. Hoffmeister (editor), *Vehicle Lighting Design for Optimal Visibility and Performance (SP-857),* pages 69-75. Society of Automotive Engineers, 1991.

[WEL 78] W. T. Welford and R. Winston.
The Optics of Nonimaging Concentrators.
Academic Press, Inc., New York, 1978.

[WHI 80] T. Whitted.
An Improved Illumination Model for Shaded Display.
Communications of the ACM 23(6):343-349, June, 1980.

[WIN 91] R. Winston.
Nonimaging Optics: Optical Design at the Thermodynamic Limit.
In R. Winston, R. L. Holman (editors), *Nonimaging Optics: Maximum Efficiency Light Transfer*, pages 2. SPIE, 1991.

[YAM 26] Z. Yamauti.
The Light Flux Distribution of a System of Interreflecting Surfaces.
Journal of the Optical Society of America 13:561, 1926.

[ZAN 67] W. I. Zangwill.
Minimizing a Function Without Calculating Derivatives.
Computer Journal 10:293-296, 1967.

Table of Contents

List of Figures