Binocular Projection of a Random Scene

Miles Hansard
www.eecs.qmul.ac.uk/~milesh/

School of Elec. Eng. & Comp. Sci.
Queen Mary, University of London,
Mile End Road, London E1 4NS.

Abstract

Current approaches to large-scale visual reconstruction would benefit from a statistical model of multiple-view projection. In particular, global constraints based on scene-clutter and occlusion are required. This work presents a new statistical model, starting from the simplest example of a random scene, viewed by two cameras. It has previously been shown that, if the scene is modelled as a Poisson process of identical objects, then the distance to a visible object follows an exponential distribution. But this is not qualitatively realistic, because the mode of the exponential is at zero, implying that the optical centre is fully amid the clutter. Real range-data, in contrast, follows a two-tailed distribution along each ray. A more realistic visibility density is proposed, in which the optical centre is displaced from the Poisson scene by a Gaussian shift. This means that the distance to an un-occluded object is now distributed according to the convolution of the exponential and Gaussian distributions, which has a variable mode. This distribution is re-parameterized along the corresponding epipolar line in another view, via the appropriate Jacobian. The resulting correspondence density gives the prior probability of a binocular match. The parameters of the density are estimated from a data-set of outdoor laser range-scans. This makes it possible to generate full 3D Poisson models that are statistically consistent with the real data. These synthetic scenes are used to further investigate the correspondence density, by Monte Carlo simulation.

1 Introduction

Multiple-view correspondence problems are hard because of inconsistencies and ambiguities between the images. Some parts of the scene may appear very different from one view to another, and yet many scenes may be compatible with the given images. These observations motivate the Bayesian approach to such problems [3, 9, 11]. In particular, if the scene is assumed to be piecewise-smooth, then dense correspondence can often be estimated by global optimization routines [1, 5]. These methods work well, given calibrated cameras and a limited class of scenes (e.g. 'carpentered' interiors).

Strong local priors, such as piecewise surface-smoothness, are less appropriate for intricate [16] or large-scale [22] visual reconstruction tasks. Typical scenes are *cluttered* at the scale of buildings and trees, which suggests a role for weak, but *global* priors. In particular, it would be useful to have probabilistic models of visibility and occlusion. Similar questions have previously been addressed in relation to volumetric scene representations [2, 13, 18, 19]. These models emphasize the weak but global constraints that are placed on the scene by *photometric* relations, across large numbers of images. It is argued here that there is a complementary role for statistical models of multiple-view *geometry*. In particular,
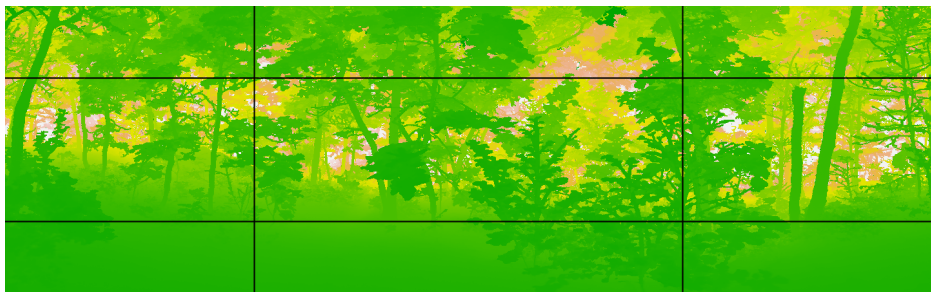
Figure 1: **Range data**. An example image from the Brown data-set [12]. Lighter surfaces are further away. Data for the experiments was taken from the the central $120° \times 40°$ rectangle of these images.

a **correspondence density** is introduced, which assigns a prior probability to the projection of a point in one image, given its location in another. The approach is based on the statistics of natural scenes [4, 20], in the same way that 2-D restoration methods can be based on the statistics of natural images [9, 11, 21].

The statistical properties of 3-D scenes are difficult to study, for two reasons. Firstly, there is a huge variety of visual environments, and so it seems that any practical model must be quite generic. Secondly, global models are difficult to test, because most of the available data is extremely biased, although this may not be immediately apparent. For example, reconstructions that have been acquired from a sparse set of viewpoints tend to be unions of façades, orthogonal to each optical axis, rather than fair representations of the internally-structured scene volume. This applies to laser and time-of-flight data, as well as to conventional multi-view reconstructions. The most obvious manifestation of the bias is that new viewpoints will usually reveal gaps, or gross interpolations, in any optical reconstruction.

The first problem is addressed here by basing the scene model on a homogeneous Poisson process [8], which is as generic as possible. The second problem is avoided, by testing on simulated scenes that are *statistically consistent* with a real data set, given the Poisson model. Specifically, the parameters of a **visibility density** are estimated from a collection of laser range-scans. No viewpoint-bias is introduced, because no actual 3-D structure is retained. The estimated parameters, however, are sufficient to generate any number of synthetic scenes, according to the Poisson model. These synthetic scenes have the property that, if the visibility parameters were re-estimated from them, the same values would be obtained as from the laser scans. Millions of these random scenes have been generated, and used to evaluate the correspondence density.

The present work is closely related to that of Langer [14], in which the exponential occlusion model (cf. [12, 17]) was developed. This is the basis of the new visibility density that is introduced in section 2. The empirical evaluation of the density is based on laser range-scan data that was recorded and analyzed (differently) by Huang, Lee & Mumford [12, 15]. The new binocular correspondence density is interestingly related to the monocular SLAM model of Civera, Davison & Montiel [7]. There the authors propose an inverse-depth parameterization, based on a Gaussian model of parallax in the images (as opposed to distance in the scene). The relevance of this will be considered in section 4.3.

1.1 Overview and contributions

Section 2 develops the monocular case, based on the existing exponential model (2.1). The placement of the cameras with respect to the scene is then considered (2.2), leading to the new **visibility density** (2.3). Section 3 develops the binocular case, including the geometry (3.1) and parameterization (3.2), leading to the new **correspondence density** (3.3). Section 4 reports the experiments, including fits to laser range-scan data (4.1), as well as simulations (4.2). The new model is then evaluated (4.3), followed by a general discussion in section 5. The chief contributions of the paper are the ex-Gaussian scene model (2.3), the binocular Jacobian (3.2) and the scene analysis and synthesis procedures in (4.1) and (4.2) respectively.

1.2 Preliminaries

Some recurring assumptions and notation will be established here. Firstly, it will be assumed that the scene $\mathcal{S}$ consists of many identical objects. Suppose that the *numbers* of objects found in disjoint regions of the scene are statistically independent; it follows that the objects are distributed in space according to a *Poisson process* [8]. This work is concerned with the binocular case and so it will be convenient to work in an arbitrary *epipolar plane*, $\mathcal{E}$, that passes through both optical centres. This captures all of the essential geometry, and the resulting model can easily be extended to the full 3-D space. The scene-elements, in this plane, are *discs*.

The cameras, which are assumed to be calibrated, will be modelled simply as bundles of rays through their optical centres. In other words, wide-angle spherical projection is assumed, which is also appropriate for laser ranging devices. The relative orientation of the cameras will not be modelled, in order to simplify the presentation. However, in the spherical-imaging model, the mapping from the bundle of rays to the oriented image-surface is just an overall angular shift (vs. a homography for planar imaging). It will be assumed that a point in the left (L) view is observed, and that the location of the corresponding point in the right (R) image is sought. The model is of course symmetric with respect to these labels, and indeed a mixture of $L \to R$ and $R \to L$ processes may be considered simultaneously.

The exponential and Gaussian probability distributions will be used repeatedly, so let them be expressed as $F(x,\lambda) = \lambda \exp(-\lambda x)$ and $G(x,\mu,\sigma) = (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)$ respectively. Note that the usual condition $F(x,\lambda) = 0$ for negative x has *not* been imposed here, as this will allow a cleaner development below (though in practice, the argument will always be a length or volume, and therefore non-negative).

2 Monocular model

This section develops a new model of the **visibility density** along a line of sight, which defines the probability that an observed object lies at a given distance from the optical centre. Two underlying processes will be identified in sections 2.1 and 2.2, and combined in section 2.3.

2.1 Occlusion process

It is assumed, as described in sec. 1.2, that objects in the epipolar plane $\mathcal{E}$ are distributed according to a homogeneous Poisson process [8]. This means that the probability $\text{pr}(k|\mathcal{R})$
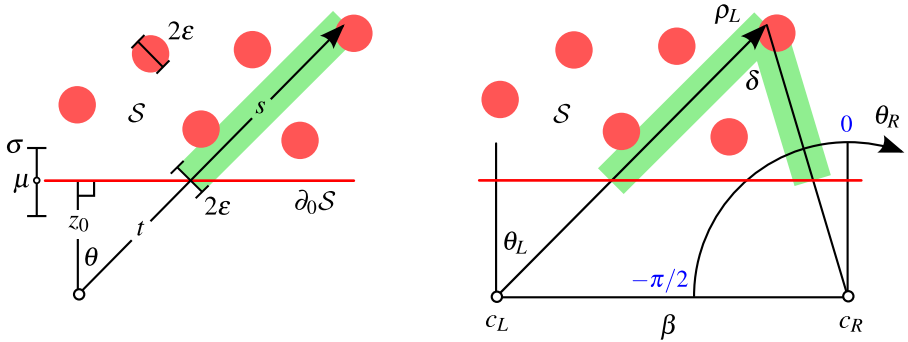
Figure 2: **Viewing geometry**. *Monocular case, left:* The scene $\mathcal{S}$ comprises discs of radius ε, with near-boundary $\partial_0 \mathcal{S}$. The depth z_0 of the latter is normally distributed around μ. A ray in direction θ extends to distance t through empty space, followed by distance s through the scene, before striking an object. Equivalently, there are no disc-centres in the $2\varepsilon \times s$ green rectangle, and so s is exponentially distributed. *Binocular case, right:* both rectangles must be empty for the disc to be binocularly visible. Note that equal increments of θ_R would span increasingly long segments of $\rho_L = s + t$ as the difference-angle δ decreases.

of k objects being located in a region $\mathcal{R} \subset \mathcal{E}$ of area $|\mathcal{R}|$ is given by the Poisson density

$$\mathrm{pr}(k|\mathcal{R}) = \exp(-\lambda|\mathcal{R}|)\frac{(\lambda|\mathcal{R}|)^k}{k!} \tag{1}$$

where λ is the *intensity* of the process (number of objects per unit area). Now, following Langer [14] and Huang *et al.* [12], suppose that the objects are discs of radius ε. In order for a disc at distance s, in direction θ to be visible, no other disc-centre can be within a perpendicular distance ε of the corresponding ray-segment. Geometrically, the $2\varepsilon \times s$ rectangle $\mathcal{R}$ must be empty, as shown in fig. 2 (left). Algebraically, $k = 0$, and it follows from (1) that

$$\mathrm{pr}(s|\lambda, \varepsilon) = F(s, 2\varepsilon\lambda) \tag{2}$$

which is an exponential density (as defined in sec. 1.2), after incorporating the normalization with respect to distance, $\int_0^\infty \mathrm{pr}(s|\lambda, \varepsilon)\,\mathrm{d}s = 1$.

2.2 Offset process

The chief limitation of (2) as a visibility model is that the mode of the distribution is always at the optical centre. This is not characteristic of empirical scene distributions, as seen in fig. 3, because the camera (or head) is not placed randomly amid the scene clutter. Rather, the optical centre tends to be placed in relatively empty space, and pointed *at* the cluttered region. Suppose that the near-boundary $\partial_0 \mathcal{S}$ of the clutter is a fronto-parallel plane, lying at a random distance z_0 in the straight-ahead direction $\theta = 0$. A ray of azimuth θ intersects the near-plane at a distance $t = z_0/\cos\theta$. Now suppose that z_0 follows a Gaussian distribution $G(z_0, \mu, \sigma)$, where μ is the mean 'viewing distance'. It follows, by exact uncertainty propagation, that

$$\mathrm{pr}(t|\theta, \mu, \sigma) = G\left(t, \frac{\mu}{\cos\theta}, \frac{\sigma}{\cos\theta}\right) \tag{3}$$

is the distribution of the distance t to $\partial_0 \mathcal{S}$, along a ray of azimuth θ. The parameter σ accounts for variations in the overall viewing-distance, irregularity of the scene-boundary, and for additive Gaussian measurement errors.

2.3   Visibility density

Recall from sec. 2.1 that a ray in direction θ extends to distance s through the scene $\mathcal{S}$. However, given the offset model (3), the scene begins at distance t along the ray. The total range is the sum of the two distances

$$\rho = s + t \tag{4}$$

which must be distributed according to the *convolution* of the individual densities. It is convenient to first define $H = F \star G$ from the original exponential F and Gaussian G densities:

$$H(\rho, \lambda, \mu, \sigma) = \int_{-\infty}^{\rho} F(\rho - \tau, \lambda) \, G(\tau, \mu, \sigma) \, d\tau \tag{5}$$

$$= F\left(\rho - \mu - \tfrac{1}{2}\lambda \sigma^2, \lambda\right) \int_{-\infty}^{\rho} G(\tau - \mu - \lambda \sigma^2, 0, \sigma) \, d\tau \tag{6}$$

where, in practice, the value of the remaining integral is easily obtained from the standard normal CDF, evaluated at $(\rho - \mu)/\sigma - \lambda \sigma$. The density $H(\rho, \lambda, \mu, \sigma)$ is the *ex-Gaussian*, which is commonly used to model human reaction-times [6]. It can be seen from (6) that H is the product of a falling exponential and a rising sigmoid, so it has a left-skewed shape. The mean and variance of H are $\mu + 1/\lambda$ and $\sigma^2 + 1/\lambda^2$, respectively. Another appealing property of this density is that additive Gaussian measurement error $\Delta\rho$ is absorbed so that the density of $\rho + \Delta\rho$ remains ex-Gaussian. It is straightforward to define the full range-density, by combining the parameterizations of (2) and (3), to give

$$\text{pr}(\rho | \theta, \lambda, \mu, \sigma) = H\left(\rho, \, 2\varepsilon\lambda, \, \frac{\mu}{\cos\theta}, \, \frac{\sigma}{\cos\theta}\right). \tag{7}$$

This then, is the probability that a line-of-sight extends to distance ρ, in direction θ. The near-boundary of the scene is, on average, at distance μ in direction $\theta = 0$. Each unit-region of the scene contains, on average, λ discs of radius ε.

3   Binocular model

In order for a point $\boldsymbol{p}$ to be *binocularly* visible, the $\rho_L \times \varepsilon$ and $\rho_R \times \varepsilon$ rectangles around the rays in directions θ_L and θ_R must *both* be empty [14], as in fig. 2. It follows that the binocular visibility density is a product of ex-Gaussians (7),

$$\text{pr}(\rho_L, \rho_R) = \text{pr}(\rho_L | \theta_L) \times \text{pr}(\rho_R | \theta_R) \tag{8}$$

with everything implicitly conditioned on the scene-parameters ε, λ, μ and σ. Note that this joint probability is defined over the epipolar-plane $\mathcal{E}$, where (ρ_L, ρ_R) is just another pair of coordinates for $\boldsymbol{p}$. It would, of course, be much more interesting to consider the distributions that arise in the *images*, rather than the scene. In particular, the *conditional* distributions $\text{pr}(\theta_R | \theta_L)$ and $\text{pr}(\theta_L | \theta_R)$ are most important. For example, the distribution $\text{pr}(\theta_R | \theta_L)$ gives the probability of seeing a point in direction θ_R from $\boldsymbol{c}_R$, given that it was seen in direction θ_L in $\boldsymbol{c}_L$. These densities will be defined in sec. 3.3, but it is first necessary to develop the underlying geometry and resulting change of variables.
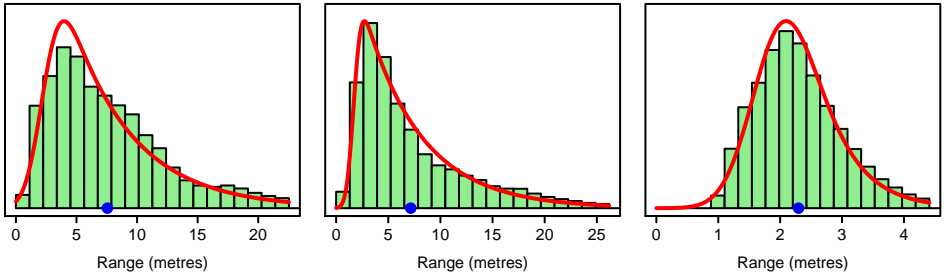
Figure 3: **Range densities**. *Left:* maximum likelihood ex-Gaussian fit to the upper 'canopy' regions of 14 range-scans. *Middle:* fit to the central $120° \times 40°$ regions (cf. fig. 1), which is most representative of heterogeneous clutter. *Right:* fit to the lower 'ground' regions. The blue dot is the mean of each data-set.

3.1 Viewing geometry

The left and right optical centres, c_L and c_R, are separated by an interocular distance β. The baseline of the system is perpendicular to the two zero-azimuth planes, as in fig. 2. As described in sec. 1.2, the binocular geometry will be analyzed in the epipolar plane $\mathcal{E}$ that is spanned by c_L, c_R and a given scene-point p. These three points form a triangle, with angle $\delta = \theta_R - \theta_L$ at p. The viewing distances ρ_L and ρ_R can be obtained from the angles θ_L and θ_R via the sine rule,

$$\rho_L = \frac{\beta \cos \theta_R}{\sin \delta} \quad \text{and} \quad \rho_R = \frac{\beta \cos \theta_L}{\sin \delta} \quad \text{where} \quad \delta = \theta_R - \theta_L. \tag{9}$$

Note that the angle $0 \leq \delta \leq \pi$ is non-negative for any point in front of the baseline, and non-zero for any finite point. Furthermore, it follows from the triangle inequality that $\rho_L + \rho_R \geq \beta$, and that for finite points $|\rho_R - \rho_L| \leq \beta$.

3.2 Change of variables

Suppose that $\mathrm{pr}_X(x)$ and $\mathrm{pr}_Y(y)$ are the densities of a random variable x, before and after a transformation $y = f(x)$. Now recall that if f is smooth and invertible, then the two densities are related by the conservation of probability $\mathrm{pr}_Y(y)\,|\mathrm{d}y| = \mathrm{pr}_X(x)\,|\mathrm{d}x|$. It follows that the transformed density is properly defined as $\mathrm{pr}_Y(y) = \mathrm{pr}_X\big(f^{-1}(y)\big)\,|\mathrm{d}f^{-1}/\mathrm{d}y|$. The transformed variable y and inverse-transformation f^{-1} are, in the present context, θ_R and the vector-valued function $\big(\rho_L(\theta_R), \rho_R(\theta_R)\big)$ respectively. The latter will be abbreviated to (ρ_L, ρ_R), with the two components defined by (9). The conserved probability is then

$$\mathrm{pr}(\theta_R \,|\, \theta_L) \times |\mathrm{d}\theta_R| = \mathrm{pr}(\rho_L, \rho_R) \times |\mathrm{d}(\rho_L, \rho_R)|. \tag{10}$$

This is more complicated than the scalar case, because the inverse function maps θ_R to a *pair* of numbers (ρ_L, ρ_R), such that a curve is traced out in the 2-D parameter-space. The quantity $|\mathrm{d}(\rho_L, \rho_R)/\mathrm{d}\theta_R|$ is the *length* of the tangent vector to the curve, at the point defined by θ_R. This can also be interpreted as the '1-D volume' of the 2×1 Jacobian matrix $\mathrm{d}(\rho_L, \rho_R)/\mathrm{d}\theta_R$, if the volume of a rank-r matrix is defined as the product $\Pi_i^r \sigma_i$ of nonzero singular-values (as $r = 1$ and $\sigma_1 = |\mathrm{d}(\rho_L, \rho_R)/\mathrm{d}\theta_R|$ in the present case). The actual Jacobian term can be

obtained by differentiating (9) with respect to θ_R, and evaluating the norm:

$$J_R(\theta) = \sqrt{(d\rho_L/d\theta_R)^2 + (d\rho_R/d\theta_R)^2}$$
$$= \rho_R\sqrt{\csc^2\delta + \cot^2\delta}, \quad \text{where} \quad \delta = \theta - \theta_L. \tag{11}$$

The explicit form of (11) is obtained by noting that the derivatives of (9) can be expressed as $d\rho_L/d\theta_R = -\rho_R/\sin\delta$ and $d\rho_R/d\theta_R = -\rho_R/\tan\delta$. It is worthwhile, remembering that the direction θ_L is fixed, to consider the form of the curve (ρ_L, ρ_R) in the parameter space. In particular, it is asymptotic to a $45°$ line, which can be seen with reference to the derivatives, by evaluating the local slope $\frac{d\rho_R}{d\theta_R}/\frac{d\rho_L}{d\theta_R} = \cos\delta$, as θ_R becomes equal to θ_L and so $\delta \to 0$. Conversely, at the point $d\rho_R/d\theta_R \propto \cot\delta$, the directions θ_L and θ_R are perpendicular, $\delta = \pi/2$, and ρ_R is minimized. Equation (11), as noted above, gives the length of the tangent-vector to this curve. The factor $\sqrt{\csc^2\delta + \cot^2\delta}$ is U-shaped over the range $0 \le \delta \le \pi$, equal to one at $\delta = \pi/2$, and accelerates very rapidly at either end. This has significant consequences for the shape of the final density, as described below.

3.3 Correspondence density

The complete conditional density can now be constructed from (10) and (11). The probability of seeing a point in direction θ_R from c_R, given that it was observed in direction θ_L from c_L is

$$\text{pr}(\theta_R \,|\, \theta_L) = \text{pr}(\rho_L, \rho_R) \times J_R(\theta_R)/S_R(\theta_L) \tag{12}$$

where $J_R(\theta_R)$ is the Jacobian function (11), and $S_R(\theta_L)$ is the normalizing constant. The latter is defined by integrating from the epipole to the vanishing point,

$$S_R(\theta_L) = \int_{-\frac{\pi}{2}}^{\theta_L} \text{pr}(\rho_L, \rho_R)\, J_R(\theta_R)\, d\theta_R \tag{13}$$

so that $\int_{-\pi/2}^{\theta_L} \text{pr}(\theta_R \,|\, \theta_L)\, d\theta_R = 1$. It may seem that the integral (13) could be zero (and 12 undefined) if a point in direction θ_L is occluded from c_R in a particular scene. However, over the *ensemble* of scenes the integral must be positive, as certainly each point is *sometimes* binocularly visible. The constant $S_L(\theta_R)$ for $\text{pr}(\theta_L \,|\, \theta_R)$ is similarly defined, but with the integration running from θ_R to $\pi/2$.

It is important to note that the density (12) is well-defined at infinity, which corresponds to the vanishing-point $\text{pr}(\theta_R \to \theta_L \,|\, \theta_L)$. This limit is always equal to zero, even though $J_R(\theta_R) \to \infty$, because of the *exponential* decay $\text{pr}(\rho_L, \rho_R) \to 0$ in the tail of the ex-Gaussian product. This ensures that the normalizing constant (13) can easily be obtained by numerical integration.

4 Experiments

Three stages of experiment are reported. In sec. 4.1 the parameters of the monocular model from sec. 2.3 are estimated for real data. In sec 4.2 these parameters are used to synthesize a full 3-D scene that is consistent with the data. Finally, in 4.3, the binocular model from sec. 3.3 is tested on the synthetic scenes. The data consists of 14 wide-angle range images, showing woodland scenes, chosen from the Brown set [12]. These were acquired by a Riegl LMS-Z210 laser scanner, with angular resolution of $0.18°$, and a range resolution of 0.8cm.

4.1   Scene analysis

Maximum-likelihood estimates of the parameters λ, μ and σ, in the model $\text{pr}(\rho \,|\, \theta, \lambda, \mu, \sigma)$ from (7) will now be computed. Firstly, the two additional parameters θ and ε, that appear on the right-hand side of (7), must be considered. The azimuth θ_i is known for each pixel, and so the range-data ρ_i is first transformed to depth-data $\rho_i \cos \theta_i$, thereby eliminating azimuth from the model. In effect, all data is projected onto the straight-ahead direction. The disc-radius ε appears as $2\varepsilon\lambda$ in (7), and is therefore not observable; instead, the product $\Lambda = 2\varepsilon\lambda$ is estimated. This results in a one-parameter family of fitted models, varying from coarse/sparse to fine/dense.

Thenegative log-likelihood of the model was numerically minimized over the parameters (Λ, μ, σ), using an initizalization based on the empirical mean (which would be the ML estimate of $1/\Lambda$ if μ and σ were zero). The resulting fit is shown on the left of fig. 3. It can be seen that the shape of the empirical distribution is well-represented by the ex-Gaussian model. In particular, the local asymmetry at the mode is present the fit.

4.2   Scene synthesis

The level curves of $\text{pr}(\rho_L, \rho_R)$ can be approximated by confocal ellipses, around the two optical centres, because the density depends essentially on the sum $\rho_L + \rho_R$. It follows that the natural far-boundary $\partial_1 \mathcal{S}$ of the scene is a semi-ellipse in the epipolar plane $\mathcal{E}$. But typical scenes are large in relation to the baseline separation β and so, for practical purposes, the foci coincide and the far-boundary is a semi-circle. The near-boundary $\partial_0 \mathcal{S}$ is a plane, as detailed in 2.2, which intersects $\mathcal{E}$ in a line. The area of the remaining D-shaped region is $|\mathcal{S}| = \frac{z_1}{2}(\phi - \sin\phi)$ where ϕ is the field-of-view, and $\cos(\phi/2) = \mu/z_1$. Hence $|\mathcal{S}|$ is determined by μ, which has been estimated, and z_1 which can either be imposed, or set to the maximum range of the laser-scanner. In fact it makes little difference for dense scenes, in which the boundary $\partial_1 \mathcal{S}$ is rarely seen.

If N discs are generated, then the intensity of the simulated data will be $\lambda' = N/|\mathcal{S}|$. This should match λ in the estimate $\Lambda = 2\varepsilon\lambda$, and so $N = \frac{\Lambda}{2\varepsilon}|\mathcal{S}|$ after rounding the right-hand side to the nearest integer. Note that ε remains a free parameter, as previously discussed. The points themselves are generated by rejection-sampling over the bounding rectangle of the scene. Let x be uniformly distributed on $[-z_1, z_1]$, and let z be uniformly distributed on $[z_0, z_1]$ where z_0 is itself normally distributed with the previously estimated parameters μ and σ. Each point (x, z) is accepted if $x^2 + z^2 < z_1^2$, until N have been added to the scene. This scheme is quite efficient, because the D-shaped region fills most of its bounding rectangle.

4.3   Evaluation

It is now possible to evaluate the binocular model from sec. 3, using the 3-D scenes that were synthesized in sec. 4.2. This is not a trivial exercise, because the model involves several simplifications, and it is not immediately clear that these are reasonable. In particular, according to (8), the probability of binocular visibility is a function of two disjoint rectangles, $\rho_L \times 2\varepsilon\lambda$ and $\rho_R \times 2\varepsilon\lambda$. This is a slight underestimate of the probability, because the two rectangles actually *overlap* as they meet, by an amount that depends on δ. Likewise, it is not immediately clear that the scene-boundary has been suitably defined in 4.2.

The parameters Λ, μ and σ, estimated from the range data, were used to simulate random scenes as described in sec. 4.2. The object radius ε was set to 10cm, which is plausible for
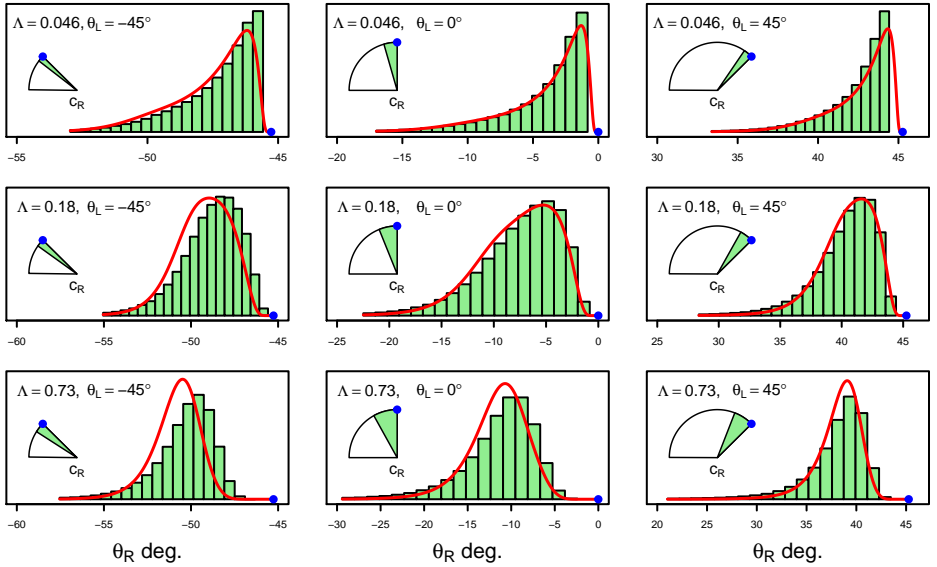
Figure 4: **Correspondence densities**. Histograms show the angular distributions $\mathrm{pr}(\theta_R|\theta_L)$ obtained by Monte Carlo simulation. The red lines are the *predicted* (not fitted) distributions defined by (12). The angular sectors extend from the epipole $(-90°)$ to θ_L, which is the vanishing-point (blue dot) of the ray through c_L, as seen from c_R. The extent of each distribution is coloured green. *Top row:* at low densities, the Jacobian dominates because even very distant points $(\delta \approx 0)$ may be seen. *Middle row:* these distributions are matched to the statistics of the laser-range data in the central panel of fig. 3. *Bottom row:* at high densities, the occlusion process dominates, and the mode is shifted back from the vanishing-point.

foliage mixed with larger surfaces. The baseline separation β was set to 50cm, which would be realistic for an outdoor stereo system. It is interesting, given these settings, to study the effect of varying θ_L around zero, and of varying λ around the estimated density $\frac{\Lambda}{2\varepsilon}$. The results of these Monte Carlo simulation are shown in fig. 4, where each graph represents 10^6 binocular projections. The red lines are the distributions that are directly predicted (not fitted) by (12).

Firstly, it is clear that the broad location of the distribution is correct in all cases (note that the horizontal axis varies according to the plot). Secondly, the shapes of the distributions are complicated, and vary according to the density Λ. At low densities more distant parts of the scene are binocularly visible, and the mass moves towards each vanishing point. At high densities occlusions occur just inside $\partial_0 \mathcal{S}$, and more symmetric distributions are obtained. The shapes of the distributions also vary according to the azimuth θ_L. All of these variations are well captured by the model. The least accurate predictions are for the left-hand column of fig. 4. This is attributable to the overlap effect discussed above; when $\theta_L \ll 0$, the left and right rays are near-parallel over the whole range of θ_R, and so there is substantial overlap between the two surrounding rectangles. It is interesting to note that, except at very low densities, the correspondence distributions (fig. 4) are 'more Gaussian' than the visibility densities (fig. 3). This supports the assumptions made by Civera et al. in their inverse-depth parameterization [7].

5 Discussion

An ex-Gaussian model of the monocular visibility density has been introduced. It has been shown that this is a good fit for outdoor data from a laser range-scanner. The Jacobian of the projection from the supporting ray to the corresponding epipolar line has been derived. The binocular correspondence density, which assigns a probability to the projection of a point in one image, given its location in the other, has been constructed from these components. Random scenes, that are statistically consistent with the range-data, were synthesized and used to validate the model. Future work will consider more sophisticated scene models, including spatially inhomogeneous processes. The correspondence density, as discussed in the introduction, can be interpreted as a global prior for binocular matching. The practicality of this application will be investigated.

References

[1] P. Belhumeur. A Bayesian approach to binocular stereopsis. *International Journal of Computer Vision*, 19(3):237–260, 1996.

[2] R. Bhotika, D.J. Fleet, and K.N. Kutulakos. A Probabilistic Theory of Occupancy and Emptiness. In *Proc. ECCV*, pages 112–132, 2002.

[3] A. Blake and A. Zisserman. *Visual Reconstruction*. MIT Press, 1987.

[4] A. Blake, B. Bascle, M. Isard, and J. MacCormick. Statistical models of visual shape and motion. *Proceedings of the Royal Society of London A*, 356:1283–1302, 1998.

[5] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11):1222–1239, 2001.

[6] S.L. Burbeck and R.D. Luce. Evidence from auditory simple reaction-times for both change and level detectors. *Perception and Psychophysics*, 32:117–133, 1982.

[7] J. Civera, A.J. Davison, and J.M.M. Montiel. Inverse depth parametrization for monocular slam. *IEEE Transactions on Robotics*, 24(5):932–945, 2008.

[8] D.R. Cox and V. Isham. *Point Processes*. Chapman & Hall, 1980.

[9] S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721–741, 1984.

[10] U. Grenander and A. Srivastava. Probability models for clutter in natural images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(4):424–429, 2001.

[11] G.E. Hinton and Z. Ghahramani. Generative models for discovering sparse distributed representations. *Philosophical Transactions of the Royal Society of London B*, 352:1177–1190, 1997.

[12] J. Huang, A.B. Lee, and D. Mumford. Statistics of range images. In *Proc. CVPR*, pages 324–331, 2000.

[13] K.N. Kutulakos and S.M. Seitz. A theory of shape by space carving. *International Journal of Computer Vision*, 38(3):199–218, 2000.

[14] M.S. Langer. Surface visibility probabilities in 3D cluttered scenes. In *Proc. ECCV*, pages 401–412, 2008.

[15] A.B. Lee, D. Mumford, and J. Huang. Occlusion models for natural images: A statistical study of a scale-invariant dead leaves model. *International Journal of Computer Vision*, 41(1):35–59, 2001.

[16] G. Li and S.W. Zucker. Contextual inference in contour-based stereo correspondence. *International Journal of Computer Vision*, 69(1):59–75, 2006.

[17] S.G. Narasimhan and S.K. Nayar. Vision and the atmosphere. *International Journal of Computer Vision*, 48(3):233–254, 2002.

[18] T. Pollard and J.L. Mundy. Change Detection in a 3D World. In *Proc. CVPR*, pages 1–6, 2007.

[19] M.I. Restrepo, B.A. Mayer, and J.L. Mundy. Object recognition in probabilistic 3D volumetric scenes. In *Proc. ICPRAM*, pages 180–190, 2012.

[20] D.L. Ruderman and W. Bialek. Statistics of natural images: Scaling in the woods. *Physical Review Letters*, 73(6):814–817, 1994.

[21] E.P. Simoncelli and B.A. Olshausen. Natural image statistics and neural representation. *Annual Reviews of Neuroscience*, 24:1193–1215, 2001.

[22] N. Snavely, S.M. Seitz, and R. Szeliski. Modeling the world from internet photo collections. *International Journal of Computer Vision*, 80(2):189–210, 2008.