

SHAPE FROM TEXTURE USING LOCALLY SCALED POINT PROCESSES

EVA-MARIA DIDDEN^{✉,1}, THORDIS L. THORARINSDOTTIR², ALEX LENKOSKI² AND CHRISTOPH SCHNÖRR³

¹Institut für Sozial- und Präventivmedizin (ISPM), Universität Bern, Finkenhubelweg 11, CH-3012 Bern, Switzerland; ²Norwegian Computing Center, P.O. Box 113 Blindern, 0314 Oslo, Norway; ³Image & Pattern Analysis Group, Universität Heidelberg, Speyerer Straße 6, D-691115 Heidelberg, Germany
e-mail: evamaria.didden@ispm.unibe.ch; thordis@nr.no; alex@nr.no; schnoerr@math.uni-heidelberg.de
(Received October 31, 2013; revised March 11, 2015; accepted July 27, 2015)

ABSTRACT

Shape from texture refers to the extraction of 3D information from 2D images with irregular texture. This paper introduces a statistical framework to learn shape from texture where convex texture elements in a 2D image are represented through a point process. In a first step, the 2D image is preprocessed to generate a probability map corresponding to an estimate of the unnormalized intensity of the latent point process underlying the texture elements. The latent point process is subsequently inferred from the probability map in a non-parametric, model free manner. Finally, the 3D information is extracted from the point pattern by applying a locally scaled point process model where the local scaling function represents the deformation caused by the projection of a 3D surface onto a 2D image.

Keywords: 3D scenes, convex texture elements, locally scaled point processes, near regular texture, perspective scaling, shape analysis.

INTRODUCTION

Natural images contain a variety of perceptual information enabling the viewer to infer the three-dimensional shapes of objects and surfaces (Tuceryan and Jain, 1998). Stevens (1980) observed that surface geometry mainly has three effects on the appearance of texture in images: foreshortening and scaling of texture elements, and a change in their density. Gibson (1950) proposed the slant, the angle between a normal to the surface and a normal to the image plane, as a measure for surface orientation. Stevens amended this by introducing the tilt, the angle between the surface normal's projection onto the image plane and a fixed coordinate axis in the image plane. In this paper, we will directly infer the surface normal from a single image taken under standard perspective projection as shown in Fig. 1.

Statistical procedures for estimating surface orientation often make strong assumptions on the regularity of texture. Witkin (1981) assumes observed edge directions provide the necessary information, while Blostein and Ahuja (1989) consider circular texture elements with uniform intensity. Blake and Marions (1990) consider the bias of the orientation of line elements isotropically oriented on a plane in 3D space, induced by the plane's orientation under

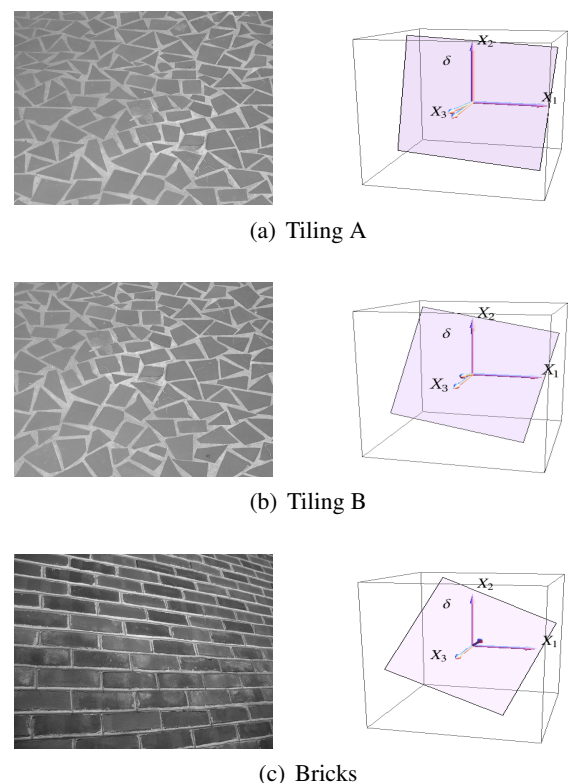


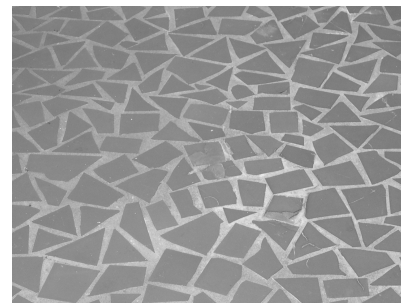
Fig. 1. Original natural scenes (left) and the estimated 3D orientation towards the camera (right). The surface normal is indicated by δ in the coordinate system defined by (X_1, X_2, X_3) . The field of view of the camera is assumed to be driven by a wide angle setting of $\phi_c = 54^\circ$.

orthographic projection, along with a computational approach related to Kanatani’s texture moments (Kanatani, 1989).

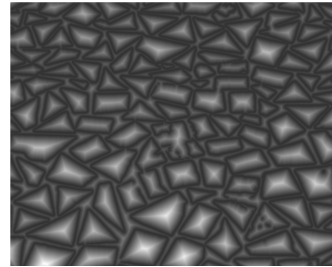
Malik and Rosenholtz (1997) locally estimate “texture distortion” in terms of an affine transformation of adjacent image patches. The strong homogeneity assumption underlying this approach has been relaxed by Clerc and Mallat (2002), to a condition that is difficult to verify in practice. Forsyth (2006) eliminates assumptions on the non-local structure of textures (like homogeneity) altogether and aims to estimate shape from the deformation of individual texture elements. Loh and Hartley (2005) criticize prior work due to the restrictive assumptions related to homogeneity, isotropy, stationarity or orthographic projection, and claim to devise a shape-from-texture approach in the most general form. Their work, however, also relies on estimating the deformation of single texture elements, similar to Forsyth (2006).

We propose a general framework for inferring shape from near regular textures, as defined by Liu *et al.* (2009), by applying the locally scaled point process model of Hahn *et al.* (2003). This framework enables the simultaneous representation of local variability and global regularity in the spatial arrangement of texture elements which are thought of as a marked point process (Fig. 2). We preprocess the image (Fig. 2.1) to obtain a probability map (Fig. 2.2) representing an unnormalized intensity estimate for the underlying point process, subsequently apply a non-parametric framework to infer the point locations (Fig. 2.3) and based on the resulting point pattern (Fig. 2.4), learn the parameters of a locally scaled point process model to obtain a compact description of 3D image attributes (Fig. 2.5).

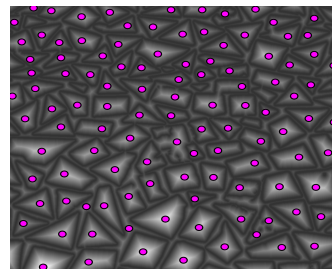
Point process models have previously been applied in image analysis applications where the goal is the detection of texture elements, see, *e.g.*, Lafarge *et al.* (2010) and references therein. These approaches usually apply a marked point process framework, with marks describing the texture elements. Such set-ups rely on a good geometric description of individual texture elements, limiting the class of feasible textures. As our goal is not the detection of individual texture elements but the extraction of 3D information, we omit the modeling of each texture element and infer the latent point locations in a model free manner. Thus, our sole assumption regarding texture element shape is approximate convexity which offers considerable flexibility.



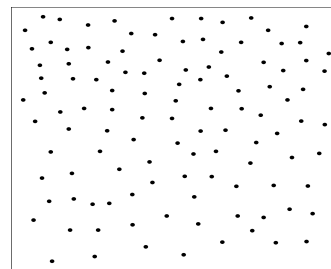
1. Original image



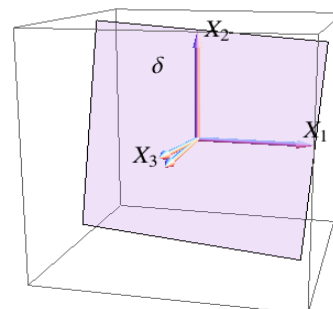
2. Probability map



3. Point assignment



4. Point process realization



5. 3D orientation

Fig. 2. Sketch of the step sequence proposed for inferring shape from texture.

The remainder of the paper is organized as follows. The next section contains preliminaries on image geometry followed by the method section describing the image preprocessing, the point pattern detection and the point process inference framework. We then present results for both simulated and real images with near regular textures. The paper closes with a short discussion section.

PRELIMINARIES

Let

$$P = \{X \in \mathbb{R}^3 : \langle \delta, X \rangle + h = 0\}, \quad (1)$$

with $\|\delta\| = 1$ and $\langle \delta, X \rangle < 0$ denote a plane in 3D with unknown unit normal δ and distance h from the origin. We assume δ to be oriented towards the camera, forming obtuse angles $\langle \delta, X \rangle < 0$ with projection rays X . The world coordinates $X = (X_1, X_2, X_3)^\top$ and image coordinates $x = (x_1, x_2)^\top$ are aligned as shown in Fig. 3. Here, we denote the image domain by D and assume the image to be scaled to have fixed area, $v^2(D) = a$, where v^2 is the two-dimensional Lebesgue measure.

We consider the basic pinhole camera (Hartley and Zisserman, 2000) and among the internal parameters, we only look at the focal length $f > 0$ which depends on the field of view (Fig. 3). As usual, we identify image points and rays of the projective plane through

$$X = (x_1, x_2, -f)^\top. \quad (2)$$

An image point X given by Eq. 2 meets P in λX with

$$\lambda = -\frac{h}{\langle \delta, X \rangle}, \quad \lambda > 0. \quad (3)$$

It follows that a point X_P in P is related to the image point X through

$$X_P = X_P(x_1, x_2) = -\frac{h}{\langle \delta, X \rangle} X. \quad (4)$$

A homogeneous texture covering P induces an inhomogeneous texture on the two-dimensional image plane with density given by the surface element

$$\begin{aligned} dX_P &= \|\partial_{x_1} X_P \times \partial_{x_2} X_P\| v^2(dx) \\ &= -\frac{h^2 f}{\langle \delta, X \rangle^3} v^2(dx). \end{aligned} \quad (5)$$

Taking, for instance, the fronto-parallel plane $\delta = (0, 0, 1)^\top$ results by Eq. 2 merely in the constant scale factor $(h/f)^2$, i.e., the homogeneous density $(h/f)^2 v^2(dx)$. However, for arbitrary orientation δ ,

this factor depends on X , as illustrated in Fig. 4. Eq. 5 then quantifies perspective foreshortening and inhomogeneity of the texture, respectively, as observed in the image, and mathematically represents the visually apparent texture gradient.

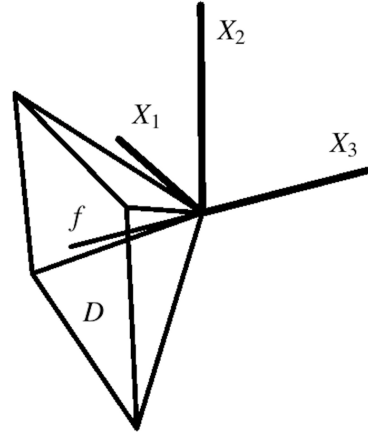


Fig. 3. The camera with focal length f is oriented towards the negative X_3 -halfspace. The scaled visible image domain is $D = [-a/2, a/2] \times [-1/2, 1/2]$. Given the field of view in terms of an angle ϕ_c , we have $f = (a/2) / \tan(\phi_c/2)$.

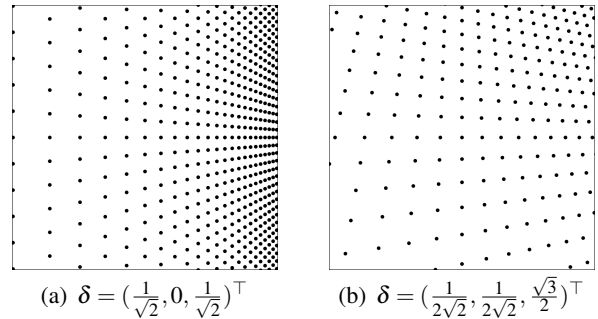


Fig. 4. Mappings of regular homogeneous point patterns in \mathbb{R}^3 onto a 2D-plane. The simulations are based on the parameters $D = [-1/2, 1/2] \times [-1/2, 1/2]$, $h = 20$ and $\phi_c = 27^\circ$ ($f = 0.98$).

METHODS

In a first step, we apply image preprocessing that generates a probability map $Y = \{Y(x) : x \in D_0, 0 \leq Y(x) \leq 1\}$ representing the spatial arrangement of texture elements in the original image plane D_0 . Here, we assume that D_0 is rectangular. To this end, two elementary techniques are locally applied: Boundary detection and the corresponding distance transform. The former step entails either gradient magnitude

computation using small-scale derivative-of-Gaussian filters (Canny, 1986) or, for texture elements with less regular appearance, the earth-mover's distance (Pele and Werman, 2009) between local histograms. Inspecting in turn the histogram of the resulting soft-indicator function for boundaries enables one to determine a threshold and apply the distance transform.

In our framework, the texture elements are regarded as a realization of a marked point process where the underlying point pattern is latent. The value of the probability map $Y(x)$ in $x \in D_0$ denotes the probability that one of the latent points is located in x . To recover the latent point pattern based on the information in Y , we first search for local maxima in Y . That is, let $W_x = [x_1 - k_1, x_1 + k_1] \times [x_2 - k_1, x_2 + k_1]$ and set

$$\Phi = \{x \in D_0 : W_x \subset D_0, Y(x) = \max_{z \in W_x} Y(z)\}, \quad (6)$$

for some fixed positive k_1 with an upper bound ensuring that $\bigcap_{x \in D_0} W_x = \emptyset$. We then define a neighbourhood relation on Φ by setting $x^1 \sim x^2$ if

$$\min_{z \in [x^1, x^2]} Y(z) \geq k_2 \max\{Y(x^1), Y(x^2)\}, \quad (7)$$

where $x^1, x^2 \in \Phi$, $[x^1, x^2]$ denotes the line from x^1 to x^2 and k_2 is a constant with $0 < k_2 < 1$. We may now write Φ as a union of disjoint neighbourhood components, $\Phi = \bigcup_{i=1, \dots, n_0} C_i$, where

$$C_i = \{\{x^{j_1}, \dots, x^{j_{m_i}}\} \subseteq \Phi : \forall k \in \{1, \dots, m_i\}, \exists l \in \{1, \dots, m_i\}, k \neq l : x^{j_k} \sim x^{j_l}\}. \quad (8)$$

That is, each $x \in C_i$ is neighbour with at least one point in $C_i \setminus x$. In the unlikely case that $\Phi = C_1$, k_1 and k_2 need to be re-specified, or the image is not appropriate or not appropriately preprocessed for our shape-from-texture approach. For fixed k_1 and k_2 , the decomposition of Φ is almost surely unique, provided that the pixel values are measurable on a continuous scale.

Under the assumption that the texture elements are close to convex sets, two points x^1 and x^2 in Φ are neighbours if and only if they likely fall within the same texture element. Hence, we first estimate the latent point process Ψ_0 in the observation window D_0 as

$$\Psi_0 = \{x^1, \dots, x^{n_0} : Y(x^i) = \max_{z \in C_i} Y(z)\}. \quad (9)$$

To avoid boundary effects, we afterwards eliminate all elements of Ψ_0 not located in $D := [\min_x(D_0) + k_1, \max_x(D_0) - k_1] \times [\min_x(D_0) + k_1, \max_x(D_0) - k_1]$ and

continue the analysis with the resulting point process realization in D ,

$$\Psi = \{x \in \Psi \cap D\}, \quad (10)$$

where $\min_x(D_0)$ and $\max_x(D_0)$ define the bounding box of the original image plane D_0 . The total number of points in Ψ is denoted by n .

Formally, a point process can be described as a random counting measure $N(\cdot)$, where $N(A)$ is the number of events in A for a Borel subset A of the relevant state space, in our context the image domain D . The intensity measure of the point process is given by $\Lambda(A) = \mathbb{E}N(A)$ and the associated intensity function is

$$\alpha(x) = \lim_{v^2(dx) \rightarrow 0} \frac{\mathbb{E}N(dx)}{v^2(dx)}. \quad (11)$$

For a homogeneous point process, it holds that $\alpha(x) = \beta$ for some $\beta > 0$, while for an inhomogeneous point process where the inhomogeneity stems from local scaling (Hahn *et al.*, 2003), we obtain

$$\alpha(x) = \beta c_\eta^{-2}(x), \quad (12)$$

for some scaling function $c_\eta : \mathbb{R}^2 \rightarrow \mathbb{R}_+$ which we assume depends on a set of parameters η . That is, η controls the amount of scaling. The point process is homogeneous if $\eta = 0$, whereas $\eta \neq 0$ indicates heterogeneity. Examples are given in Fig. 5 (a). The scaling function c_η acts as a local deformation in that it locally affects distances and areas. More precisely, $v_c^d(A) = \int_A c_\eta(x)^{-d} v^d(dx)$, where v^d denotes the d -dimensional Lebesgue measure and v_c^d its scaled version for $d = 1, 2$.

For identifiability reasons, Prokešová *et al.* (2006) propose normalizing c_η to conserve the total area of the state space. That is, they require that

$$v^2(D) = \int_D c_\eta^{-2}(x) v^2(dx) = \int_D \gamma(\eta) \tilde{c}_\eta^{-2}(x) v^2(dx), \quad (13)$$

where $\tilde{c}_\eta^{-2}(x)$ denotes the unnormalized scaling function and $\gamma(\eta)$ the normalizing constant. Hahn *et al.* (2003) and Prokešová *et al.* (2006) specifically consider the exponential scaling function with $c_\eta(x) \propto \exp(\eta^\top x)$. This scaling function is particularly attractive in that locally scaled distances can be calculated explicitly,

$$v_c^1(x^i, x^j) = v^1(x^i, x^j) \left| \frac{c_\eta^{-1}(x^i) - c_\eta^{-1}(x^j)}{\eta^\top (x^j - x^i)} \right|, \quad (14)$$

for any $x^i, x^j \in D$, where $v^1(\cdot)$ corresponds to the length of the line segment $[x^i, x^j]$ between x^i and x^j , and $v_c^1(\cdot)$ denotes its scaled version. Examples of exponentially scaled distances are given in Fig. 5b.

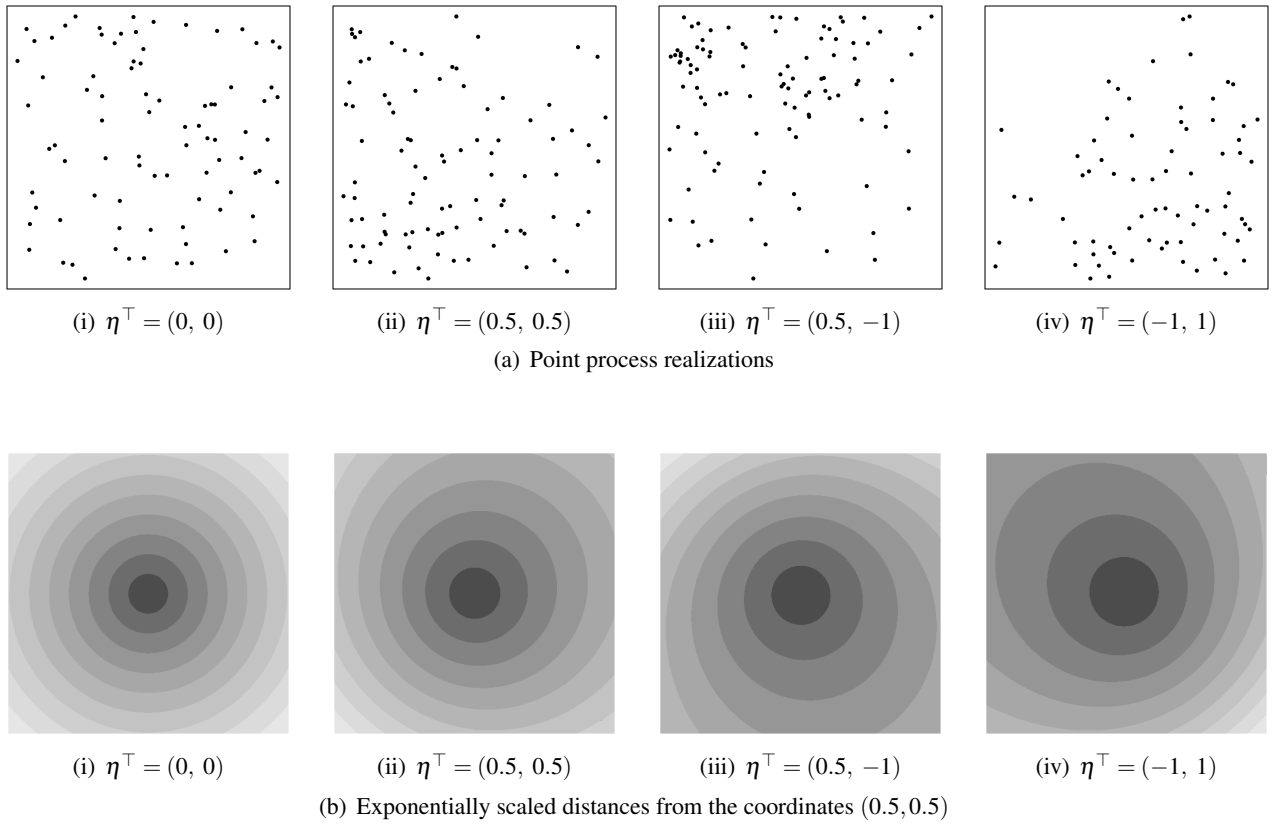


Fig. 5. Exponential scaling effects of varying strength in $D = [0, 1] \times [0, 1]$. (a) shows four point process realizations sampled from a locally scaled point process model due to Hahn et al. (2003). Apart from η , all model components are kept fixed. (b) exemplifies the exponentially scaled distances under the different specifications of η in (a). Increasing darkness indicates decreasing distance.

Here, we employ the density in Eq. 5 as scaling kernel, i. e. $\tilde{c}_{\delta(\eta)}^{-2}(x|f, h)v^2(dx) = dX_P$, where we choose spherical coordinates

$$\begin{aligned} \delta &= \delta(\eta_1, \eta_2) \\ &= (\sin \eta_1 \cos \eta_2, \sin \eta_1 \sin \eta_2, \cos \eta_1)^\top, \end{aligned} \quad (15)$$

with $\eta_1 \in [0, u]$ and $\eta_2 \in [0, 2\pi]$. The upper limit u restricting the range of the scaling parameter η_1 ensures that $\langle \delta, X \rangle < 0$ and therefore depends on the focal length f as well as on the size and location of the observation window D . As suggested by Prokešová et al. (2006), we normalize the scaling function such that Eq. 13 holds. That is, we solve

$$v^2(D) = a = \int_D \gamma(\delta, h, f) dX_P. \quad (16)$$

It follows that

$$\begin{aligned} \gamma(\delta, h, f) &= \frac{1}{16h^2 f^2 \delta_3} (a\delta_1 - 2f\delta_3 - \delta_2) \\ &\quad \times (a\delta_1 - 2f\delta_3 + \delta_2) \\ &\quad \times (a\delta_1 + 2f\delta_3 - \delta_2) \\ &\quad \times (a\delta_1 + 2f\delta_3 + \delta_2). \end{aligned}$$

A more general result for $D = [a_1, a_1] \times [b_1, b_2]$ is given in the Appendix.

Under the model in Eq. 5, the intensity function in Eq. 12 becomes

$$\alpha(x) = \beta \frac{\gamma(\delta(\eta_1, \eta_2), h, f) h^2 f}{|\langle \delta(\eta_1, \eta_2), X \rangle|^3}, \quad (17)$$

with $X = (x_1, x_2, -f)^\top$ as in Eq. 2. As a byproduct, the unknown plane parameter h cancels. It sets the absolute scale and cannot be inferred from a single image. Furthermore, the scaling function is computationally tractable and, as for the exponential scaling discussed above, the scaled distance function is available in closed form,

$$\begin{aligned} v_c^1([x^i, x^j]) &= v^1([x^i, x^j]) \times \gamma(\delta, h, f)^{\frac{1}{2}} \\ &\quad \times \left| \frac{2h\sqrt{f}}{\langle \delta, X^i - X^j \rangle} \left(\frac{1}{\langle \delta, -X^i \rangle^{\frac{1}{2}}} - \frac{1}{\langle \delta, -X^j \rangle^{\frac{1}{2}}} \right) \right|, \end{aligned} \quad (18)$$

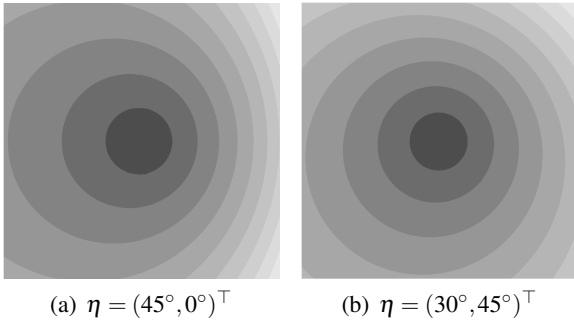


Fig. 6. *Examples of distances from the point $(0,0)$ within the observation window $D = [-1/2, 1/2] \times [-1/2, 1/2]$, under scaling assumptions due to Eq. 18. Darker shades of gray indicate smaller distances.*

provided that the basic requirement $\langle \delta, X^i \rangle < 0$ is fulfilled for all $i = 1, \dots, n$. Eq. 18 can easily be recalculated by applying the coarea formula,

$$v_c^1([x^i, x^j]) = v^1([x^i, x^j]) \int_0^1 c_{\delta}^{-1}(x^i + t(x^j - x^i)) dt, \quad (19)$$

see Krantz and Parks (2008, Ch. 5) for more details.

Examples of scaled distances are given in Fig. 6. When compared with Fig. 5, we see that the perspective scaling in Eq. 17 results in similar distance scaling as the exponential scaling while it also provides a coherent description of the perspective foreshortening.

For a given image, we assume that the focal length f is known. It remains to estimate the parameters (β, η_1, η_2) of the intensity function in Eq. 17 based on the estimated point pattern Ψ . The desired 3D image information, the slant and the tilt of the surface, may then be characterized by the scaling parameter estimates $\hat{\eta}_1$ and $\hat{\eta}_2$. The parameter estimation is performed by maximizing the composite likelihood, see, *e.g.*, Møller (2010), that takes the form

$$L(\Psi | \beta, \eta_1, \eta_2) \propto \exp(-\beta v^2(D)) \beta^n \prod_{i=1}^n c_{\eta}^{-2}(x^i). \quad (20)$$

The maximum composite likelihood estimate for β is $\hat{\beta} = \frac{n}{v^2(D)}$. For the remaining two parameters—the parameters of interest in our setting—we maximize the function

$$l(\Psi | \hat{\beta}, \eta_1, \eta_2) \quad (21)$$

$$= n \log \left(\frac{n}{v^2(D)} - 1 \right) + \sum_{i=1}^n \log(c_{\eta}^{-2}(x^i)).$$

Note that Eq. 21 is equal to the profile composite log-likelihood, *i. e.* the logarithm of the right-hand side in Eq. 20 with β replaced by its estimate. If Ψ is a Poisson process, the estimate of (β, η_1, η_2) is the maximum likelihood estimate.

RESULTS

SIMULATION STUDIES

We first present the results of a simulation study where we analyse sets of 3D point coordinates sampled from either a perfectly regular pattern or a homogeneous Poisson process and subsequently projected onto the 2D-plane $D = [-1/2, 1/2] \times [-1/2, 1/2]$ (Fig. 4 and Fig. 7).

We estimate the scaling parameters associated with the synthetic patterns via the composite likelihood in Eq. 21. The true parameter values and the corresponding estimates are given in Table 1. While the estimation procedure is able to reconstruct the true values with a reasonable accuracy, the results are slightly better for the regular patterns than for the random patterns. These results are representative for several further such examples (results not shown). We conclude that the composite likelihood is able to identify the scaling parameters of the perspective scaling function even though it is specified using the intensity of the point process only and thus ignores higher order characteristics. That is, for quantifying the scaling effects, a model not accounting for interaction is sufficient even if the process is repulsive as in Fig. 4.

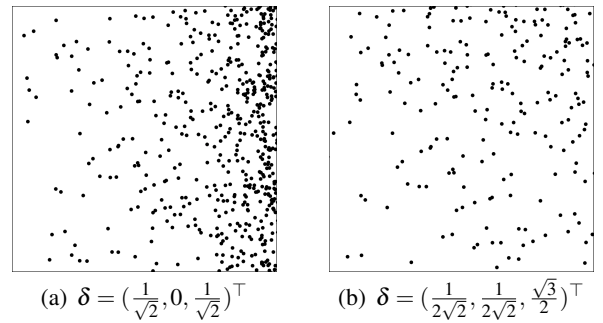


Fig. 7. *Simulated Poisson point patterns with 3D shape given by the outer normals in the subfigure captions. The internal parameters correspond to the settings in Fig. 4 and Fig. 6.*

Table 1. True angles and composite likelihood estimates for the surface normals of the simulated point patterns in Fig. 4 and 7. Regular pattern type refers to the images in Fig. 4 and Poisson type to the images in Fig. 7.

Pattern type	(η_1, η_2)	$(\hat{\eta}_1, \hat{\eta}_2)$
Regular	$(45^\circ, 0^\circ)$	$(45.5^\circ, 0.0^\circ)$
Poisson	$(45^\circ, 0^\circ)$	$(46.2^\circ, 0.7^\circ)$
Regular	$(30^\circ, 45^\circ)$	$(29.9^\circ, 45.7^\circ)$
Poisson	$(30^\circ, 45^\circ)$	$(26.2^\circ, 45.5^\circ)$

In the following, we evaluate and discuss the point detection procedure proposed in Eq. 6–Eq. 10 before estimating the scaling parameters. To investigate how violations of convexity in the shapes of the texture elements may affect the estimation of the scaling parameters, we generate images of size 1800×1800 pixels with varying proportions and arrangements of nonconvex shapes (Fig. 8). The true scaling effects are $\eta_1 = 20^\circ$ and $\eta_2 = 25^\circ$, and the overall packing density is the same in each image.

Fig. 8 shows the textured scenes and the estimated point process realizations for $k_1 = 65$. For the image in Fig. 8a containing convex shapes only, all texture elements are correctly detected if the neighbourhoods are determined by $k_1 = 65$ or $k_1 = 75$. For the scenes in Fig. 8b,c that are partly covered by concave shapes, $k_1 = 65$ seems more appropriate than $k_1 = 75$. To avoid boundary effects, the outermost 75 pixels are not considered as potential point locations. Note that the threshold k_2 is of no relevance here, since all images are binary.

The estimated scaling parameters for the point patterns in Fig. 8 are given in Table 2. While the estimates are quite accurate when all the texture elements are convex, the accuracy deteriorates somewhat with increased proportion of nonconvex shapes. In particular, we see that a grouping of the nonconvex shapes yields a lower accuracy compared to a random distribution of the nonconvex shapes.

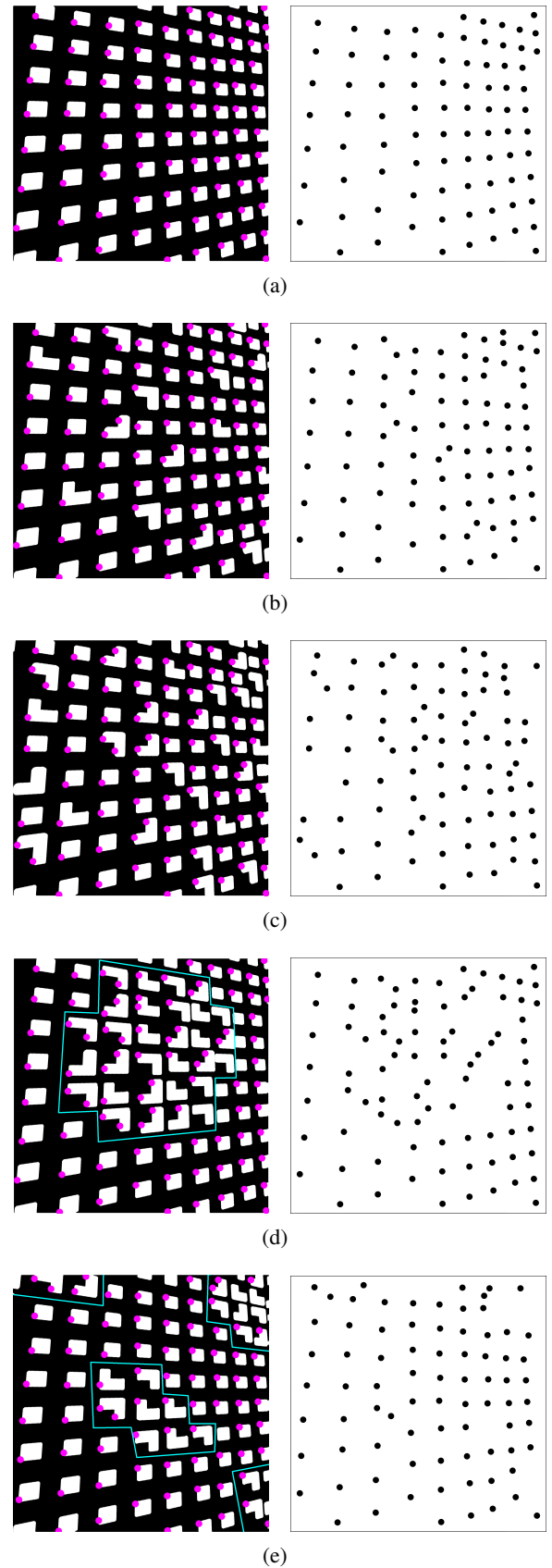


Fig. 8. Point process estimation in scenes with different amounts of nonconvex texture elements. The focal length corresponds to $f = 1.17$ ($\phi_c = 54^\circ$).

Table 2. Composite likelihood estimates for the surface normals of the simulated point patterns in Fig. 8. Random means that the concave elements are randomly distributed among the convex elements, whereas in a clustered arrangement, the concave shapes appear in groups.

	# points	$(\hat{\eta}_1, \hat{\eta}_2)$	concave shapes [%]	arrangement
(a)	84	$(20.32^\circ, 23.66^\circ)$	0	—
(b)	86	$(19.21^\circ, 31.00^\circ)$	20	random
(c)	85	$(13.66^\circ, 20.89^\circ)$	33	random
(d)	86	$(16.34^\circ, 39.20^\circ)$	33	one cluster
(e)	79	$(14.31^\circ, 38.36^\circ)$	33	four clusters

NATURAL SCENES

For the analysis of real natural scenes, we apply our methodology to the set of tiling and brick images shown in Fig. 1. The original images are of size 1280×960 pixels, and during the preprocessing, they are downsized to 1166×846 pixels in order to eliminate boundary effects in the point detection. More precisely, we cut off bounding boxes of width 115 pixels from the original scenes, since we consider estimation of the latent point process Ψ with $k_1 \in \{35, 45, \dots, 105, 115\}$. For each value of k_1 , we generate a point process realization and subsequently estimate its scaling parameters, i. e. the angles η_1 and η_2 which in turn determine the unit normal vector δ of the original plane in 3D.

The point detection is very robust in the selection of the threshold value k_2 , and threshold values from

0.15 to 0.5 have limited effects on the results. It is somewhat more sensitive to changes in the neighbourhood size k_1 as illustrated graphically in Fig. 9. For the tiling images neighbourhoods from 55×55 to 95×95 result in similar point patterns and hence in similar scaling parameter estimates. For the bricks scene, on the other hand, the estimates vary considerably and slightly smaller neighbourhoods seem to be needed.

In the following, we employ neighbourhoods of size 75×75 pixels for the tiling scenes and 55×55 pixels for the bricks scene, with a threshold of $k_2 = 0.25$ for the neighbourhood relation in all cases. The probability maps and the resulting point patterns are shown in Fig. 10.

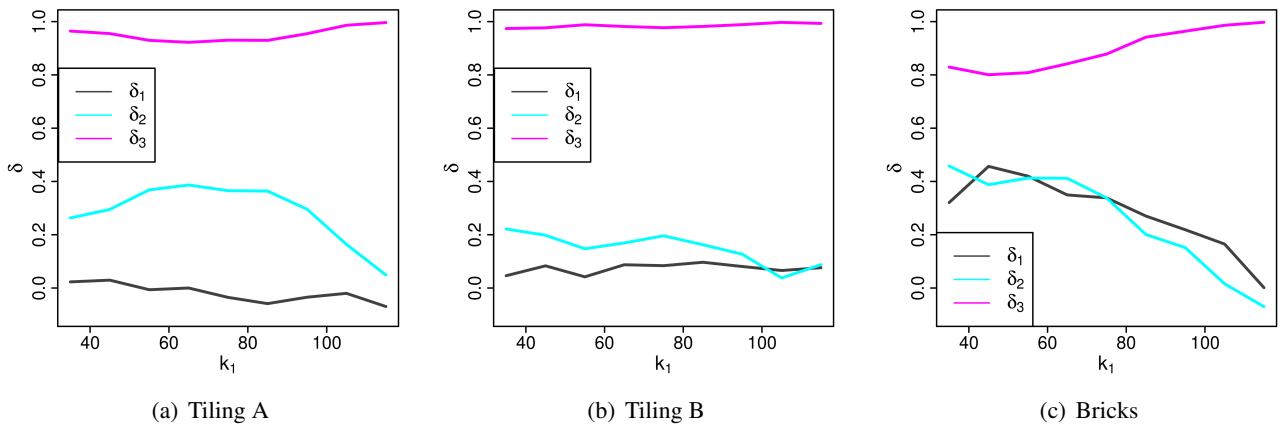
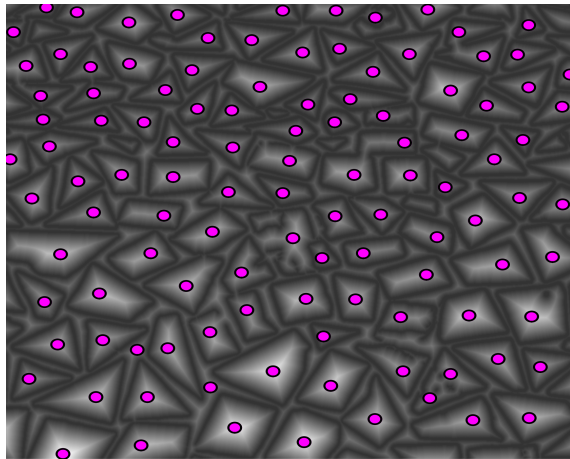
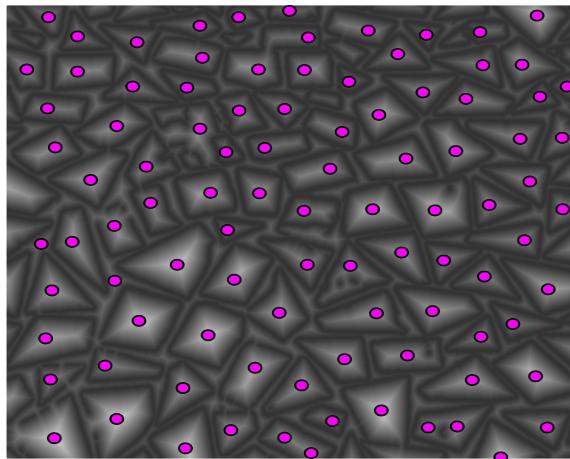


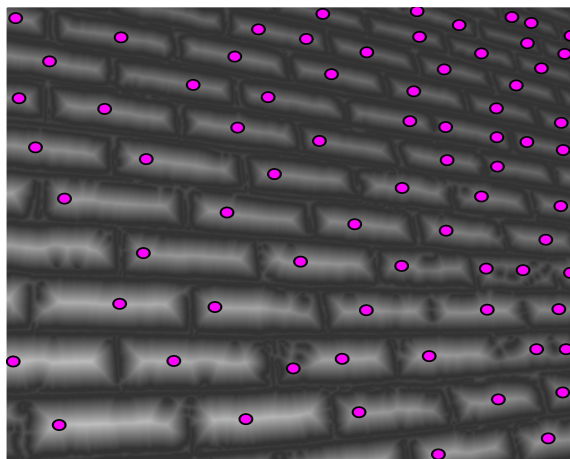
Fig. 9. Estimation of the unit normal δ subject to different specifications of the neighbourhood size k_1 .



(a) Tiling A



(b) Tiling B



(c) Bricks

Fig. 10. *Estimated probability maps and point configurations for the natural scenes in Fig. 1.*

For deriving the information on camera positioning and angle from the point configurations in Fig. 10, we project the point process realizations onto an observation window D of dimension $[-0.69, 0.69] \times$

$[-0.50, 0.50]$. We further assume that the field of view corresponds to a standard wide angle setting of $\phi_c = 54^\circ$ and hence take $f = 1.35$ as a basis, the same settings as we applied in the simulation examples above. The resulting scaling parameter estimates are listed in Table 3 and the 3D orientation of the camera toward the textures is illustrated in Fig. 1.

Table 3. *Perspective scaling parameter estimates for the natural scenes in Fig. 1.*

Texture type	$(\hat{\eta}_1, \hat{\eta}_2)$
(a) Tiling A	$(22.1^\circ, 94.7^\circ)$
(b) Tiling B	$(12.2^\circ, 66.7^\circ)$
(c) Bricks	$(36.0^\circ, 44.1^\circ)$

DISCUSSION

This paper introduces a framework for extracting 3D information from a textured 2D image building on the recently developed locally scaled point processes (Hahn *et al.*, 2003). The perspective scaling function quantifies perspective foreshortening and the resulting inhomogeneity of the texture. The framework is quite flexible regarding assumptions on the texture composition in that it only requires the texture elements to be close to convex in shape and it successfully extracts useful information related to camera orientation.

The separation of image preprocessing and point detection on one hand and the estimation procedure for the scaling parameters on the other hand offers great flexibility. We believe that the locally scaled point process framework can be applied in more general settings to analyze point patterns in images, for instance, as a new additional inference step in the texture detection algorithms discussed in Lafarge *et al.* (2010) and references therein. Due to the low computational budget of our framework, it also seems feasible to combine it with image segmentation where 3D information is needed for several segments within an image, each of which might be covered with a different type of texture elements.

There are further considerable avenues for development. One area for future development is to build a large hierarchical framework where the three inference steps, the image preprocessing, the point detection and the parameter estimation, are joined in an iterative fashion. A fully Bayesian inference framework along the lines of the work of Rajala and Penttinen (2012) could also be an alternative to

the composite likelihood estimation performed here. Future work will concentrate on embellishing our inference framework.

ACKNOWLEDGMENTS

We thank Ute Hahn for sharing her expertise. This work has been supported by the German Science Foundation (DFG), grant RTG 1653. The work of Thordis L. Thorarinsdottir and Alex Lenkoski was further supported by Statistics for Innovation, *sf²*, in Oslo.

REFERENCES

Blake A, Marinos C (1990). Shape from texture: Estimation, isotropy and moments. *Artif Intell* 45:323–80.

Blostein D, Ahuja N (1989). Shape from texture: Integrating texture-element extraction and surface estimation. *IEEE T Pattern Anal* 11:1233–51.

Canny J (1986). A computational approach to edge detection. *IEEE T Pattern Anal* 8:679–98.

Clerc M, Mallat S (2002). The texture gradient equation for recovering shape from texture. *IEEE T Pattern Anal* 24(4):536–49.

Forsyth D (2006). Shape from texture without boundaries. *Int J Comput Vision* 67(1):71–91.

Gibson J (1950). *The perception of the visual world*. Boston, MA: Houghton Mifflin.

Hahn U, Jensen EBV, Van Lieshout, MC, Nielsen, L (2003). Inhomogenous spatial point processes by location-dependent scaling. *Adv Appl Probab* 35:319–36.

Hartley R, Zisserman A (2000). *Multiple view geometry in computer vision*. Cambridge: Cambridge Univ Press.

Kanatani K (1989). Shape from texture: General principle. *Artif Intell* 38:1–48.

Krantz SG, Parks HR (2008). *Geometric integration theory*. Boston: Birkhäuser.

Lafarge F, Gimel’Farb G, Descombes X (2010). Geometric feature extraction by a multi-marked point process. *IEEE T Pattern Anal* 32(9):1597–609.

Liu Y, Hel-Or H, Kaplan C, Van Gool L (2009). Computational symmetry in computer vision and computer graphics. *Found Trends Comput Graph Vision* 5(1-2):1–195.

Loh A, Hartley R (2005). Shape from non-homogeneous, non-stationary, anisotropic, perspective texture. In: Clocksin W, Fitzgibbon A, Torr P, eds. *Proc 16th Brit Mach Vision Conf* 8.1–8.10.

Malik J, Rosenholtz R (1997). Computing local surface orientation and shape from texture for curved surfaces. *Int J Comput Vision* 23(2):149–68.

Møller J (2010). Spatial point patterns: Parametric methods. In: Gelfand AE, Diggle PJ, Fuentes M, Guttorp P, eds. *Handbook of spatial statistics*. Boca Raton, FL: CRC Press. 317-37

Møller J, Waagepetersen RP (2004). *Statistical inference and simulation for spatial point processes*. Boca Raton, FL: Chapman & Hall/CRC.

Pele O, Werman W (2009). Fast and robust Earth mover’s distances. In: *Proc 12th Int Conf Comput Vision (ICCV)*: 460-7.

Prokešová M, Hahn U, Jensen, EBV (2006). Statistics for locally scaled point processes. In: Baddeley A, Gregori P, Mateu J, Stoica R, Stoyan D, eds. *Case studies in spatial point process modelling*. New York: Springer. 99–123.

Rajala T, Penttinen A (2014). Bayesian analysis of a Gibbs hard-core point pattern model with varying repulsion range. *Comput Stat Data Anal* 71:530-41.

Stevens KA (1980). *Surface perception from local analysis of texture and contour*. MIT Tech Rep AI-TR 512. Cambridge, MA: MIT.

Tuceryan M, Jain AK (1998). Texture analysis. In: Chen CH, Pau LF, Wang PSP, eds. *Handbook of pattern recognition and computer vision*, 2nd ed. Singapore: World Scientific. 207–48.

Witkin AP (1981). Recovering surface shape and orientation from texture. *Artif Intell* 17:17–45.

APPENDIX

In our data analysis, we assume that the image domain is normalized such that $D = [-a/2, a/2] \times [-1/2, 1/2]$. More generally, the image domain could be of the form $D = [a_1, a_2] \times [b_1, b_2]$ for some $a_1, a_2, b_1, b_2 \in \mathbb{R}$ with $a_1 < a_2$ and $b_1 < b_2$. In this case, the condition of conservation of the total area in (13) becomes

$$v^2(D) = (a_2 - a_1)(b_2 - b_1) = \int_D \gamma(\delta, d, f) dX_P. \quad (22)$$

It follows that

$$\begin{aligned} \gamma(\delta, h, f) = & \frac{2}{h^2 f} (-(a_1 + a_2)\delta_1 - (b_1 + b_2)\delta_2 + f\delta_3)^{-1} \\ & \times (a_1\delta_1 + b_1\delta_2 - f\delta_3) \\ & \times (a_1\delta_1 + b_2\delta_2 - f\delta_3) \\ & \times (a_2\delta_1 + b_1\delta_2 - f\delta_3) \\ & \times (a_2\delta_1 + b_2\delta_2 - f\delta_3). \end{aligned} \quad (23)$$