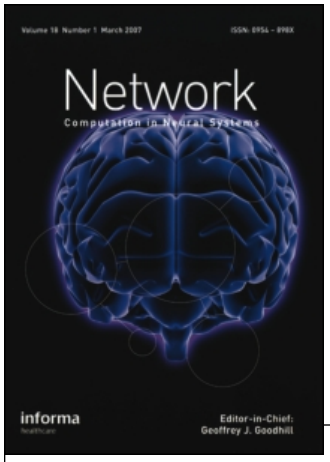

Network: Computation in Neural Systems
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713663148

First-order and second-order statistical analysis of 3d and 2d image structure

S. Kalkan [a]; F. Wörgötter [a]; N. Krüger [b]

[a] Bernstein Centre for Computational Neuroscience, University of GÖttingen, Germany
[b] Cognitive Vision Group, University of Southern Denmark, Denmark

Online Publication Date: 01 January 2007

To cite this Article: Kalkan, S., Wörgötter, F. and Krüger, N. (2007) 'First-order and second-order statistical analysis of 3d and 2d image structure', Network: Computation in Neural Systems, 18:2, 129 — 160

To link to this article: DOI: 10.1080/09548980701580444
URL: http://dx.doi.org/10.1080/09548980701580444

Network: Computation in Neural Systems
June 2007; 18(2): 129–160

informa
healthcare

First-order and second-order statistical analysis of 3d and 2d image structure

S. KALKAN[1], F. WÖRGÖTTER[1], & N. KRÜGER[2]

[1]*Bernstein Centre for Computational Neuroscience, University of Göttingen, Germany and* [2]*Cognitive Vision Group, University of Southern Denmark, Denmark*

(Received 5 March 2007; accepted 18 July 2007)

Abstract
In the first part of this article, we analyze the relation between local image structures (i.e., homogeneous, edge-like, corner-like or texture-like structures) and the underlying local 3D structure (represented in terms of continuous surfaces and different kinds of 3D discontinuities) using range data with real-world color images. We find that homogeneous image structures correspond to continuous surfaces, and discontinuities are mainly formed by edge-like or corner-like structures, which we discuss regarding potential computer vision applications and existing assumptions about the 3D world. In the second part, we utilize the measurements developed in the first part to investigate how the depth at homogeneous image structures is related to the depth of neighbor edges. For this, we first extract the local 3D structure of regularly sampled points, and then, analyze the coplanarity relation between these local 3D structures. We show that the likelihood to find a certain depth at a homogeneous image patch depends on the distance between the image patch and a neighbor edge. We find that this dependence is higher when there is a second neighbor edge which is coplanar with the first neighbor edge. These results allow deriving statistically based prediction models for depth interpolation on homogeneous image structures.

Keywords: *Natural scenes, visual system*

Introduction

Depth estimation relies on the extraction of 3D structure from 2D images which is realized by a set of inverse problems including structure from motion,

Correspondence: Sian Kalkan, Bernstein Centre for Computational Neuroscience, University of Göttingen, Germany.
E-mail: sinan@bccn-goettingen.de

ISSN 0954-898X print/ISSN 1361-6536 online
DOI: 10.1080/09548980701580444

stereo vision, shape from shading, linear perspective, texture gradients, and occlusion (Bruce et al. 2003). In methods which make use of multiple views (i.e., stereo and structure from motion), correspondences between different 2D views of the scene are required. In contrast, monocular or pictorial cues such as shape from shading, texture gradients or linear perspective use statistical and geometrical relations within one image to make statements about the underlying 3D structure.

Many surfaces have only weak texture or no texture at all, and as a consequence, the correspondence problem is very hard or not at all resolvable for these surfaces. Nevertheless, humans are able to reconstruct the 3D information for these surfaces, too. This gives rise to the assumption that in the human visual system, an interpolation process is realized that, *starting with the local analysis of edges, corners and textures* computes depth also in areas where correspondences cannot easily be found.

Processing of depth in the human visual system starts with the processing of local image structures (such as edge-like structures, corner-like structures, and textures) in V1 (Hubel and Wiesel 1969; Gallent et al. 1994; Lee et al. 1998). These structures (called 2D structures in the rest of the article) are utilized in stereo vision, depth from motion, depth from texture gradients and other depth cues, which are localized in different parts of the brain, starting from V1 and involving V2, V3, V4 and MT (Sereno et al. 2002).

There exists good evidence that depth cues which are not directly based on correspondences evolve rather late in the development of the human visual system. For example, pictorial depth cues are made use of only after approximately 6 months (Kellman and Arterberry 1998). This indicates that experience may play an important role in the development of these cues, i.e., we have to understand depth perception as a statistical learning problem (Knill and Richards 1996; Purves and Lotto 2002; Rao et al. 2002). A step towards such an understanding is the investigation and use of the statistical relations between the local 2D structures and the underlying 3D structure for each of these depth cues (Knill and Richards 1996; Purves and Lotto 2002; Rao et al. 2002).

With the notion that the human visual system is adapted to the statistics of the environment (Brunswik and Kamiya 1953; Knill and Richards 1996; Olshausen and Field 1996; Krueger 1998; Purves and Lotto 2002; Rao et al. 2002; Simoncelli 2003) and its successful applications to grouping, object recognition and stereo (Zhu, 1999; Elder and Goldberg 2002; Elder et al. 2003; Pugeault et al. 2004), the analysis and the usage of natural image statistics have become an important focus of vision research. Moreover, with the advances in technology, it has been also possible to analyze the 3D world using 3D range scanners Huang et al. 2000; Potetz and Lee 2003; Yang and Purves 2003; Howe and Purves 2004.

In this article, we analyze first-order and second-order relations[1] between 2D and 3D structures extracted from chromatic 3D range data.[2] For the first-order analysis, we investigate the relation between local 2D structures (i.e., homogeneous, edge-like, corner-like or texture-like structures) and the underlying local 3D structure. As for the second-order analysis, we investigate the relation between the depth at homogeneous 2D structures and the depth at the bounding edges.

There have been only a few studies that have analyzed the 3D world from range data (Huang et al. 2000; Potetz and Lee 2003; Yang and Purves 2003; Howe and Purves 2004), and these works have only been of first order. In Yang and Purves (2003), the distribution of roughness, size, distance, 3D orientation, curvature, and independent components of surfaces was analyzed. Their major conclusions were: (1) local 3D patches tend to be saddle-like, and (2) natural scene geometry is quite regular and less complex than luminance images. In Hunag et al. (2000), the distribution of 3D points was analyzed using co-occurrence statistics and 2D and 3D joint distributions of Haar filter reactions. They showed that range images are much simpler to analyze than optical images and that a 3D scene is composed of piecewise smooth regions. In Potetz and Lee (2003), the correlation between light intensities of the image data and the corresponding range data as well as surface convexity were investigated. They could justify the event that brighter objects are closer to the viewer, which is used by shape from shading algorithms in estimating depth. In Howe and Purves (2002, 2004), range image statistics were analyzed for explanation of several visual illusions.

Our first-order analysis differs from these works. For 2D local image patches, existing studies have only considered light intensity. As for 3D local patches, the most complex considered representation has been the curvature of the local 3D patch. In this work, however, we create a higher order representation of the 2D local image patches and the 3D local patches; we represent 2D local image patches using homogeneous, edge-like, corner-like or texture-like structures, and 3D local patches using continuous surfaces and different kinds of 3D discontinuities. By this, we relate established local 2D structures to their underlying 3D structures.

For the first-order analysis, we compute the conditional likelihood P (3D structure | 2D structure), by creating 2D and 3D representations of the local structure. Using this likelihood, we quantify some assumptions made by the studies that reconstruct the 3D world from dense range data. For example, we will show that the depth distribution varies significantly for different visual features, and we will quantify already established inter-dependencies such as 'no news is good news' (Grimson 1983). This work also supports the understanding of how intrinsic properties of 2D–3D relations can be used for the reconstruction of depth, for example, by using statistical priors in the formalization of depth cues.

For the second-order analysis, given two proximate co-planar edges, we compute the 'likelihood field' of finding co-planar surface patches which project as homogeneous 2D structures in the 2D image. This likelihood field is similar to the 'association field' (Field et al. 1993) which is a likelihood field also based on natural image statistics. The 'likelihood field' which we compute provides important information about (1) the predictability of depth at homogeneous 2D structures using the depth available at the bounding edges and (2) the relative complexity of 3D geometric structure compared to the complexity of local 2D structures.

The article is organized as follows: in the second and third sections, we define the types of local 2D structures and local 3D structures and how we extract them for our analysis. In the fourth section, we analyze the relation between the local 2D and 3D structures, and discuss the results. In the fifth section, we present our methods for analyzing the second-order relation between the homogeneous 2D structures and bounding edge structures, and discuss the results. Finally, we conclude the article in the last section with a discussion.

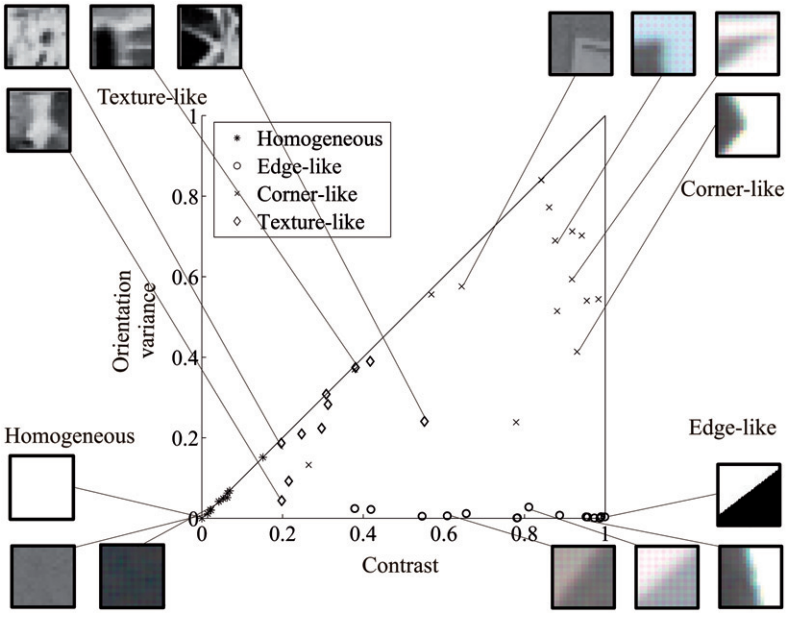

Figure 1. How a set of 54 patches map to the different areas of the intrinsic dimensionality triangle. Some examples from these patches are also shown. The horizontal and vertical axes of the triangle denote the contrast and the orientation variances of the image patches, respectively.

Local 2D structures

We distinguish between the following local 2D structures (examples of each structure are given in Figure 1):

- Homogeneous 2D structures: Homogeneous 2D structures are signals of uniform intensities, and they are not much made use of in the human visual system because retinal ganglion cells give only weak sustained responses and adapt quickly at homogeneous intensities (Bruce et al. 2003).
- Edge-like 2D structures: Edges are low-level structures which constitute the boundaries between homogeneous or texture-like signals. Detection of edge-like structures in the human visual system starts with orientation sensitive cells in V1 (Hubel and Wiesel 1969), and biological and machine vision systems depend on their reliable extraction and utilization (Koenderink and Dorn 1982; Marr 1982).
- Corner-like 2D structures: Corners[3] are image patches where two or more edge-like structures with significantly different orientations intersect (see, e.g., Guzman (1968) and Rubin (2001) for their importance in vision). It has been suggested that the human visual system makes use of them for different tasks like recovery of surface occlusion (Guzman 1968; Rubin 2001) and shape interpretation (Malik 1987).
- Texture-like 2D structures: Although there is not a widely agreed definition, textures are often defined as signals which consist of repetitive, random or directional structures (for their analysis, extraction and importance in vision,

see e.g., Tuceryan and Jain (1998). Our world consists of textures on many surfaces, and the fact that we can reliably reconstruct the 3D structure from any textured environment indicates that human visual system makes use of and is very good at the analysis and the utilization of textures.

In this article, we define texture as 2D structures which have low spectral energy and a lot of orientation variance (see Figure 1 and the following subsection).

It is locally hard to distinguish between these 'ideal' cases, and there are 2D structures that carry mixed properties of these 'ideal' cases. The classification of the features outlined above is a discrete one. However, a discrete classification may cause problems as the inherent properties of the 'mixed' structures are lost in the discretization process. Instead, in this article, we make use of a continuous scheme which is based on the concept of intrinsic dimensionality (see the following subsection for more details).

Detection of local 2D structures

In image processing, intrinsic dimensionality (iD) was introduced by Zetzsche and Barth (1990) and was used to formalize a *discrete distinction* between edge-like and junction-like structures. This corresponds to a classical interpretation of local 2D structures in computer vision.

Homogeneous, edge-like and junction-like structures are respectively classified by iD as *intrinsically zero-dimensional (i0D)*, *intrinsically one-dimensional (i1D) and intrinsically two-dimensional (i2D)*.

When looking at the spectral representation of a local image patch (Figure 2(a) and (b)), we see that the energy of an i0D signal is concentrated in the origin (Figure 2(b)-top), the energy of an i1D signal is concentrated along a line (Figure 2(b)-middle) while the energy of an i2D signal varies in more than one dimension (Figure 2(b)-bottom).

It has been shown in Felsberg and Krüger (2003) and Krüger and Felsberg (2003) that the structure of the iD can be understood as a triangle that is spanned by two measures: origin variance (i.e., contrast) and line variance. Origin variance describes the deviation of the energy from a concentration at the origin while line variance describes the deviation from a line structure (Figure 2(b) and (c)); in other words, origin variance measures non homogeneity of the signal whereas the line variance measures the junctionness. The corners of the triangle then correspond to the 'ideal' cases of iD. The surface of the triangle corresponds to signals that carry aspects of the three 'ideal' cases, and the distance from the corners of the triangle indicates the similarity (or dissimilarity) to *ideal* i0D, i1D, and i2D signals.

The triangular structure of the intrinsic dimension is counter-intuitive, in the first place, since it realizes a two-dimensional topology in contrast to a linear one-dimensional structure that is expressed in the discrete counting 0, 1, and 2. As shown in Krüger and Felsberg (2003) and Felsberg and Krüger (2003), this triangular interpretation allows for a *continuous formulation* of iD in terms of 3 confidences assigned to each discrete case. This is achieved by first computing two measurements of origin and line variance which define a point in the triangle

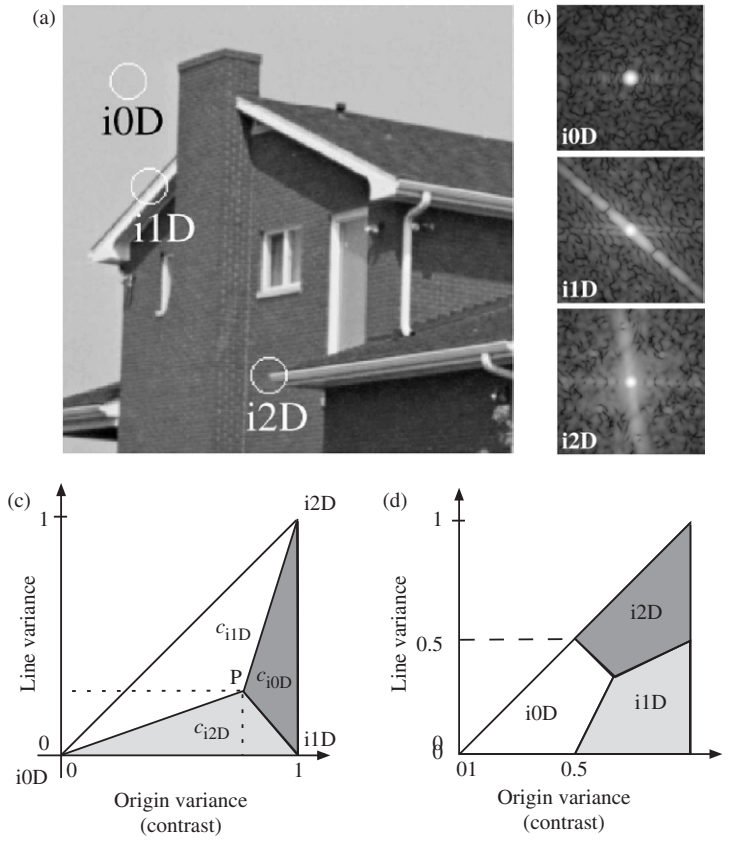

Figure 2. Illustration of iD (sub-figures (a) and (b) taken from Felsberg and Krüger (2003)). (a) Three image patches for three different intrinsic dimensions. (b) The 2D spatial frequency spectra of the local patches in (a), from top to bottom: i0D, i1D, i2D. (c) The topology of iD. Origin variance is variance from a point, i.e., the origin. Line variance is variance from a line, measuring the junctionness of the signal. c_{iND} for $N = 0, 1, 2$ stands for confidence for being i0D, i1D and i2D, respectively. Confidences for an arbitrary point P is shown in the Figure which reflect the areas of the sub-triangles defined by P and the corners of the triangle. (d) The decision areas for local 2D structures.

(Figure 2(c)). The bary centric coordinates (Coxeter 1969) of this point in the triangle directly lead to a definition of three confidences that add up to one:

$$c_{\mathrm{i0D}} = 1 - x, c_{\mathrm{i1D}} = x - y, c_{\mathrm{i2D}} = y. \tag{1}$$

These three confidences reflect the volume of the areas of the three sub-triangles which are defined by the point in the triangle and the corners of the triangle (Figure 2(c)). For example, for an arbitrary point P in the triangle, the area of the sub-triangle i0D-P-i1D denotes the confidence for i2D as shown in Figure 2(c). That leads to the decision areas for i0D, i1D and i2D as seen in Figure 2(d). See Appendix (Felsberg and Krüger 2003; Krüger and Felsberg 2003) for more details.

For the example image in Figure 2, computed iD is given in Figure 3.

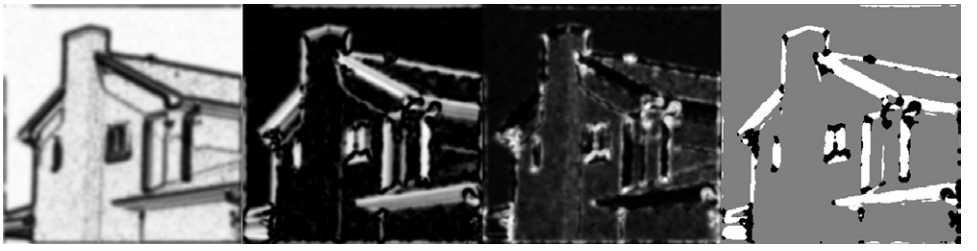

Figure 3. Computed iD for the image in Figure 2, black means zero and white means one. From left to right: c_{i0D}, c_{i1D}, c_{i2d} and highest confidence marked in gray, white and black for i0D, i1D and i2D, respectively.

Figure 1 shows how a set of example local 2D structures map on to it. In Figure 1, we see that different visual structures map to different areas in the triangle. A detailed analysis of how 2D structures are distributed over the intrinsic dimensionality triangle and how some visual information depends on this distribution can be found in Kalkan et al. (2005).

Local 3D structures

To our knowledge, there does not exist a systematic and agreed classification of local 3D structures like there is for 2D local structures (i.e., homogeneous structures, edges, corners and textures). Intuitively, the 3D world consists of continuous surface patches and different kinds of 3D discontinuities. During the imaging process (through the lenses of the camera or the eye), 2D local structures are generated by these 3D structures together with the illumination and the reflectivity of the environment.

With this intuition, any 3D scene can be decomposed geometrically into surfaces and 3D discontinuities. In this context, the local 3D structure of a point can be:

- Surface continuity: The underlying 3D structure can be described by one surface whose normal does not change or changes smoothly (Figure 4(a)).
- Regular gap discontinuity: Regular gap discontinuities are occlusion boundaries, whose underlying 3D structure can be described by a small set of surfaces with a significant depth difference. The 2D and 3D views of an example gap discontinuity are shown in Figure 4(d).
- Irregular gap discontinuity: The underlying 3D structure shows high depth-variation that cannot be described by two or three surfaces. An example of an irregular gap discontinuity is shown in Figure 4(e).
- Orientation discontinuity: The underlying 3D structure can be described by two surfaces with significantly different 3D orientations that meet at the center of the patch. This type of discontinuity is produced by a change in 3D orientation rather than a gap between surfaces. An example for this type of discontinuity is shown in Figure 4(c).

One interesting example is 3D corners of, for example, a cube. 3D corners would be classified as regular gap discontinuities or orientation discontinuities,

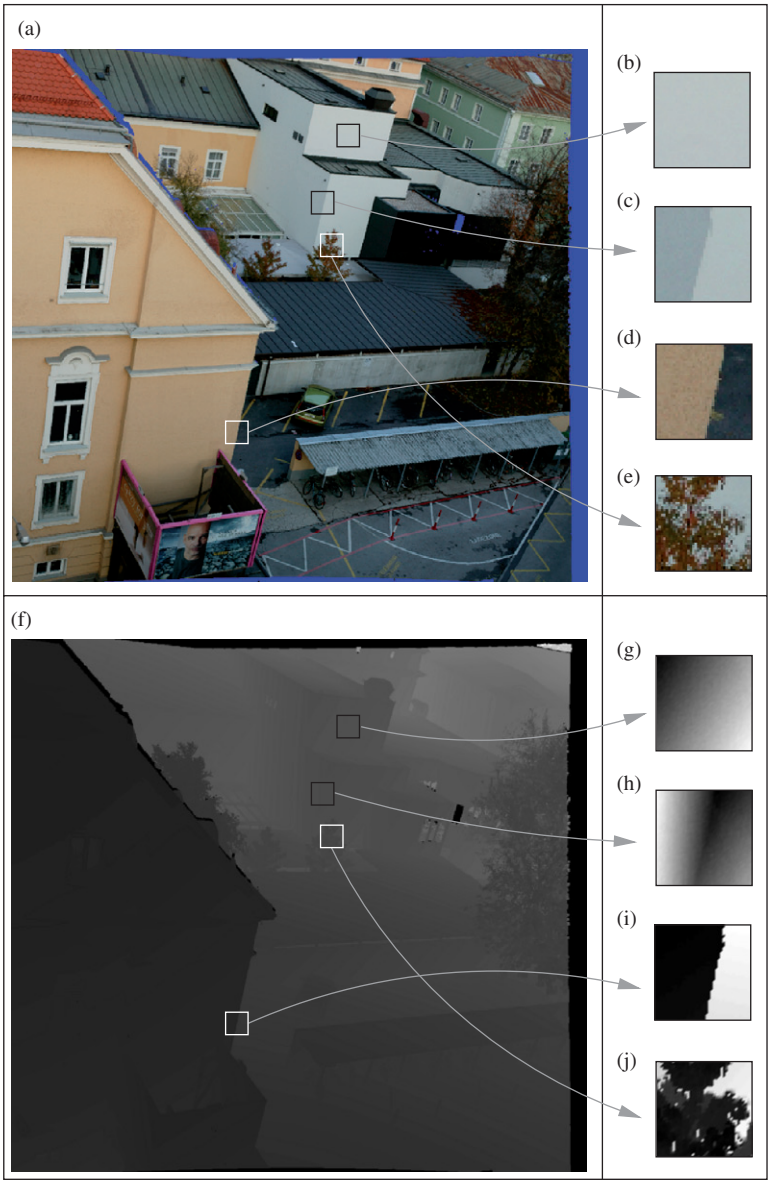

Figure 4. Illustration of the types of 3D discontinuities. (a) 2D image. (b) Continuity. (c) Orientation discontinuity. (d) Gap discontinuity. (e) Irregular gap discontinuity. (f)–(j) The range images corresponding to (a)–(e). Note that the range images are scaled independently for better visibility.

depending on the view. If the image patch includes parts of the background objects, then there is a gap discontinuity, and the 3D corner would be classified as a gap discontinuity. If, however, the camera centers the corner so that all the adjacent edges of the cube are visible and no parts of other objects are visible, then the 3D corner would be an orientation discontinuity.

Figure 5. 10 of the 20 3D data sets used in the analysis. The points without range information are marked in blue. The gray image shows the range data of the top-left scene.

Detection of local 3D structures

In this subsection, we define our measures for the three kinds of discontinuities that we described above; namely, gap discontinuity, irregular gap discontinuity and orientation discontinuity. The measures for gap discontinuity, irregular gap discontinuity and orientation discontinuity of a patch P will be denoted by $\mu_{\mathrm{GD}}(\mathrm{P})$, $\mu_{\mathrm{IGD}}(\mathrm{P})$ and $\mu_{\mathrm{OD}}(\mathrm{P})$, respectively. The reader who is not interested in the technical details can jump directly to the fourth section.

Three-dimensional discontinuities are detected in studies which involve range data processing, using different methods and under different names like two-dimensional discontinuous edge, jump edge or depth discontinuity for gap discontinuity; and, two-dimensional corner edge, crease edge or surface discontinuity for orientation discontinuity (Shirai 1987; Bolle and Vemuri 1991; Hoover et al. 1996).

In our analysis, we used chromatic range data of outdoor scenes which were obtained from RIEGL UK Ltd (http://www.riegl.co.uk/). There were 20 scenes in total, 10 of which are shown in Figure 5. The range of an object which does not reflect the laser beam back to the scanner or is out of the range of the scanner cannot be measured. These points are marked with blue in Figure 5 and are not processed in our analysis. The horizontal and the vertical resolutions of the scenes respectively have the following ranges: [512–2048] and [390–2290]. The average resolution of the scenes is 1140×1001.

Measure for gap discontinuity: μ_{GD}. Gap discontinuities can be measured or detected in a similar way as edges in 2D images; edge detection processes

RGB-coded 2D images, while for a gap discontinuity one needs to process XYZ-coded 2D images.[4] In other words, gap discontinuities can be measured or detected by taking the second-order derivative of XYZ values (Shirai, 1987).

Measurement of a gap discontinuity is expected to operate on both the horizontal and the vertical axes of the 2D image; that is, it should be a two-dimensional function. The alternative is to discard the topology and do an 'edge-detection' in sorted XYZ values, i.e., to operate as a one-dimensional function. Although we are not aware of a systematic comparison of the alternatives, for our analysis and for our data, the topology-discarding gap discontinuity measurement captured the underlying 3D structure better (of course, qualitatively, i.e., by visual inspection). Therefore, we have adopted the topology-discarding gap discontinuity measurement in the rest of the article.

For an image patch P of size $N \times N$, let,

$$\begin{aligned} \mathcal{X} &= \text{ascending\_sort}(\{X_i \mid i \in P\}), \\ \mathcal{Y} &= \text{ascending\_sort}(\{Y_i \mid i \in P\}), \\ \mathcal{Z} &= \text{ascending\_sort}(\{Z_i \mid i \in P\}), \end{aligned} \tag{2}$$

and also, for $i = 1, \ldots, (N \times N - 2)$,

$$\begin{aligned} \mathcal{X}^{\Delta} &= \{ \mid (\mathcal{X}_{i+2} - \mathcal{X}_{i+1}) - (\mathcal{X}_{i+1} - \mathcal{X}_i) \mid \}, \\ \mathcal{Y}^{\Delta} &= \{ \mid (\mathcal{Y}_{i+2} - \mathcal{Y}_{i+1}) - (\mathcal{Y}_{i+1} - \mathcal{Y}_i) \mid \}, \\ \mathcal{Z}^{\Delta} &= \{ \mid (\mathcal{Z}_{i+2} - \mathcal{Z}_{i+1}) - (\mathcal{Z}_{i+1} - \mathcal{Z}_i) \mid \}, \end{aligned} \tag{3}$$

where $\mathcal{X}_i, \mathcal{Y}_i, \mathcal{Z}_i$ represents 3D coordinates of pixel i. Equation (3) takes the absolute value of the $[+1, -2, +1]$ operator.

The sets $\mathcal{X}^{\Delta}, \mathcal{Y}^{\Delta}$ and $\mathcal{Z}^{\Delta}$ are the measurements of the jumps (i.e., second-order differentials) in the sets $\mathcal{X}, \mathcal{Y}$ and $\mathcal{Z}$, respectively. A gap discontinuity can be defined simply as a measure of these jumps in these sets. In other words:

$$\mu_{\text{GD}}(P) = \frac{h(\mathcal{X}^{\Delta}) + h(\mathcal{Y}^{\Delta}) + h(\mathcal{Z}^{\Delta})}{3}, \tag{4}$$

where the function $h : \mathcal{S} \rightarrow [0, 1]$ over the set $\mathcal{S}$ measures the homogeneity of its argument set (in terms of its 'peakiness') and is defined as follows:

$$h(\mathcal{S}) = \frac{1}{\#(\mathcal{S})} \times \sum_{i \in \mathcal{S}} \frac{s_i}{\max(\mathcal{S})}, \tag{5}$$

where $\#(\mathcal{S})$ is the number of the elements of $\mathcal{S}$, and s_i is the ith element of the set $\mathcal{S}$. Note that as a homogeneous set (i.e., a non-gap discontinuity) $\mathcal{S}$ produces a high $h(\mathcal{S})$ value, a gap discontinuity causes a low μ_{GD} value. Figure 8(c) shows the performance of μ_{GD} on one of our scenes shown in Figure 5.

It is known that derivatives, like in Equations (2) and (3), are sensitive to noise. Gaussian-based functions could be employed instead. In this article, we chose simple derivatives for their faster computation times, and instead employed a more robust processing stage (i.e., analyzing the uniformity of the distribution of derivatives) to make the measurement more robust to noise. As shown in Figure 8(c), this method can capture the underlying 3D structure well.

Measure for orientation discontinuity: μ_{OD}. The orientation discontinuity of a patch P can be detected or measured by taking the 3D orientation difference between the surfaces that meet in P. If the size of the patch P is small enough, the surfaces can be, in practice, approximated by 2-pixel wide unit planes.[5]

The histogram of the 3D orientation differences between every pair of unit planes forms one cluster for continuous surfaces and two clusters for orientation discontinuities.

For an image patch P of size $N \times N$ pixels, the orientation discontinuity measure is defined as:

$$\mu_{\mathrm{OD}}(P) = \psi\Big(H^n\Big(\Big\{\alpha(i,j) \;\mid\; i,j \in \mathrm{planes}(P), i \neq j\Big\}\Big)\Big), \tag{6}$$

where $H^n(S)$ is a function which computes the n-bin histogram of its argument set $\mathcal{S}$; $\psi(\mathcal{S})$ is a function which finds the number of clusters in $\mathcal{S}$; planes (P) is a function which fits 2-pixel-wide unit planes to 1-pixel-apart points in P using singular value decomposition;[6] and, $\alpha(i,j)$ is the angle between planes i and j.

For a histogram H of size N_H, the number of clusters is given by:

$$\psi(S) = \frac{\sum_{i=1}^{N_H+1} neq([H_i > \max(H)/10], [H_{i-1} > \max(H)/10])}{2}, \tag{7}$$

where the function *neq* returns 1 if its parameters are not equal and returns 0, otherwise; H_i represents the ith element of the histogram H; H_0 and H_{N_H+1} are defined as zero; and, $\max(H)/10$ is an empirically set threshold. Figure 6 shows two example clusters for a continuous surface and an orientation discontinuity.

Figure 8(d) shows the performance of μ_{OD} on one of our scenes shown in Figure 5.

Measure for irregular gap discontinuity: μ_{IGD}. Irregular gap discontinuity of a patch P can be measured using the observation that an irregular-gap discontinuous patch in a real image usually consists of small surface fragments with different 3D orientations. Therefore, the spread of the 3D orientation histogram of a patch P can measure the irregular gap discontinuity of P.

Similar to the measure for orientation discontinuity defined in the previous two subsections, the histogram of the differences between the 3D orientations of the unit planes (which are of 2 pixels wide) is analyzed. For an image patch P of size $N \times N$ pixels, the irregular gap discontinuity measure is defined as:

$$\mu_{\mathrm{IGD}}(P) = h\big(H^n\big(\big\{\alpha(i,j) \;\mid\; i,j \in \mathrm{planes}(P), i \neq j\big\}\big)\big), \tag{8}$$

where planes (P), $\alpha(i,j)$, $H^n(\mathcal{S})$ and $h(\mathcal{S})$ are as defined in the previous subsection. Figure 8(e) shows the performance of μ_{IGD} on one of our scenes shown in Figure 5.

Combining the measures. The relation between the measurements and the types of the 3D discontinuities are outlined in Table 1 which entails that an image patch P is:

- gap discontinuous if $\mu_{\mathrm{GD}}(P)T_g$ and $\mu_{\mathrm{IGD}}(P)T_{ig}$,
- irregular-gap discontinuous if $\mu_{\mathrm{GD}}(P) < T_g$ and $\mu_{\mathrm{IGD}}(P)T_{ig}$,
- orientation discontinuous if $\mu_{\mathrm{GD}}(P) \geq T_g$ and $\mu_{\mathrm{OD}} > 1$,
- continuous if $\mu_{\mathrm{GD}}(P) \geq T_g$ and $\mu_{\mathrm{OD}}(P) \leq 1$.

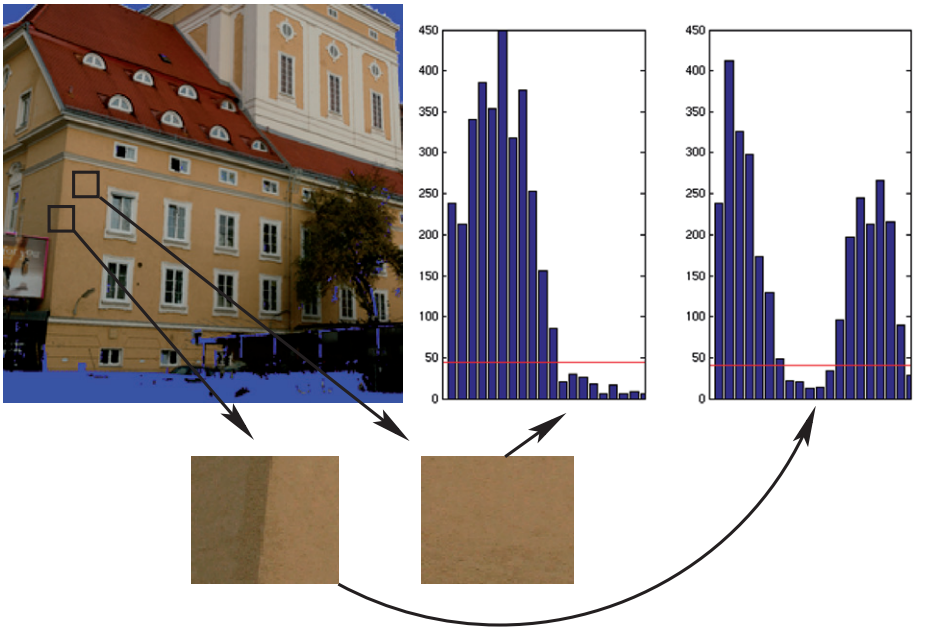

Figure 6. Example histograms and the number of clusters that the function $\psi(S)$ computes. $\psi(S)$ finds one cluster in the left histogram and two clusters in the right histogram. Red line marks the threshold value of the function. x axis denotes the values for 3D orientation differences.

Table I. The relation between the measurements and the types of the 3D discontinuities.

Discontinuity type	μ_{GD}	μ_{IGD}	μ_{OD}
Continuity	High value	Don't care	1
Gap discontinuity	Low value	Low value	Don't care
Irregular gap discontinuity	Low value	High value	Don't care
Orientation gap discontinuity	High value	Don't care	>1

For our analysis, N, where $N \times N$ is the size of the patches, is set to 10 pixels. Bigger values for N means larger support region for the measures, in which case different kinds of 3D discontinuities might interfere in the patch. On the other hand, using smaller values would make the measures very sensitive to noise. Other thresholds T_g and T_{ig} are respectively set to 0.4 and 0.6. These values are empirically determined by testing the measures over a large set of samples. Different values for these thresholds may result in wrong classifications of local 3D structures and may lead to different results than those presented in this article. Similarly, the number of bins, n, in H^n is empirically determined as 20.

Figure 7 shows the performance of the measures on two artificial scenes, one for gap discontinuity and one for orientation discontinuity for a set of depth and angle differences between planes. In the figure, the detected discontinuity type is shown for each pixel. We see that gap discontinuity can be detected reliably even if the gap difference is low. The sensitivity of the orientation discontinuity measure is around

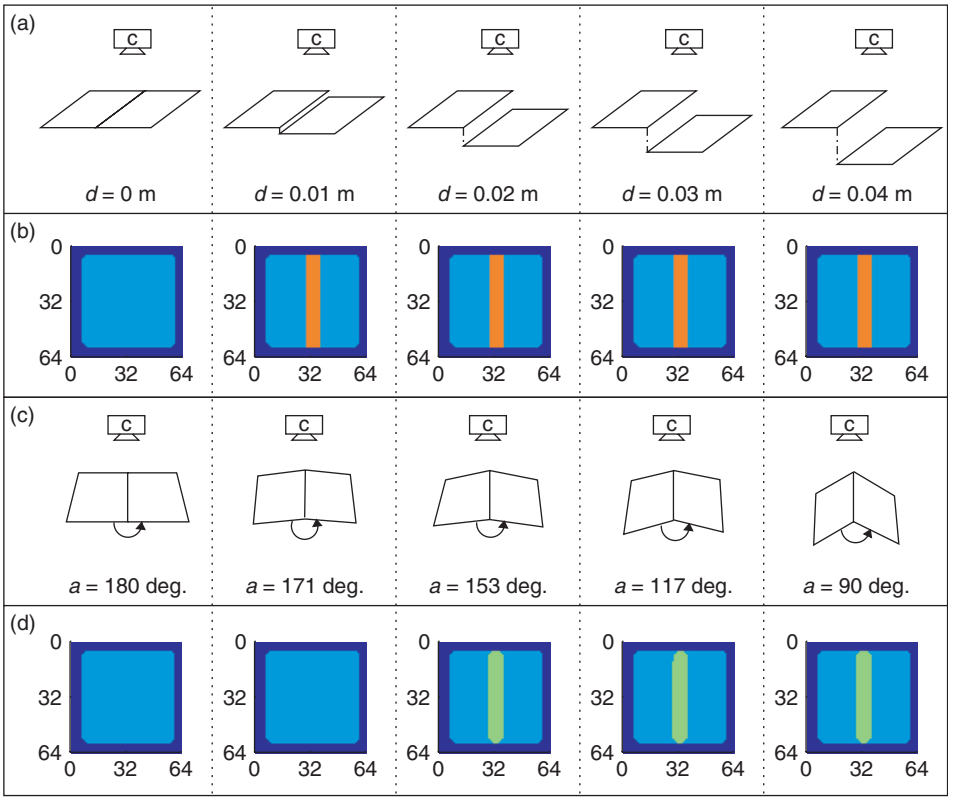

Figure 7. Results of the combined measures on artificial data. The camera and the range scanner are denoted by c. (a) Gap discontinuity tests. There are two planes which are separated by a distance d where $d = 0, 0.01, 0.02, 0.03, 0.04$ meters. (b) The detected discontinuities. Dark blue marks the boundary points where the measures are not applicable. Blue and orange respectively correspond to detected continuities and gap discontinuities. (c) Orientation discontinuity tests. There are two planes which are connected but separated with an angle a where $a = 180, 171, 153, 117, 90$ degrees. (d) The detected discontinuities. Dark blue marks the boundary points where the measures are not applicable. Blue and green respectively correspond to detected continuities and orientation discontinuities.

160 degrees. However, the sensitivity of the measures would be different in real scenes due to the noise in the range data.

For a real example scene from Figure 5, the detected discontinuities are shown in Figure 8(a). We see that the underlying 3D structure of the scene is reflected in Figure 8(a).

Note that this categorical combination of the measures appears to be against the motivation that has been provided for the classification of local 2D structures where we had advocated a continuous approach. There are two reasons: (1) with continuous 3D measures, the dimensionality of the results would be four (origin variance, line variance, a 3D measure and the normalized frequency of the signals), which is difficult to visualize and analyze. In fact, the number of triangles that had to be shown in Figure 9 would be 12, and it would be very difficult to interpret all the triangles together. (2) It has been argued by several studies (Huang et al. 2000; Yang and Purves 2003) that range images are much simpler and

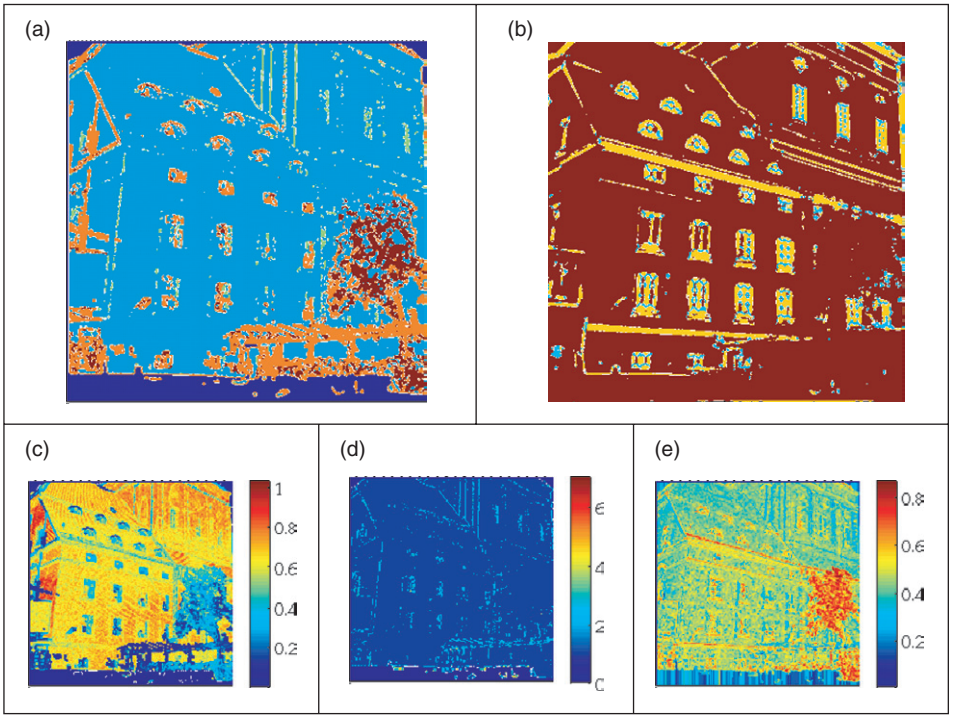

Figure 8. The 3D and 2D information for one of the scenes shown in Figure 5. Dark blue marks the points without range data. (a) 3D discontinuity. Blue: continuous surfaces, light blue: orientation discontinuities, orange: gap discontinuities and brown: irregular gap discontinuities. (b) Intrinsic dimensionality. Homogeneous patches, edge-like and corner-like structures are encoded in colors brown, yellow and light blue, respectively. (c) Gap discontinuity measure μ_{GD}. (d) Orientation discontinuity measure μ_{OD}. (e) Irregular gap discontinuity measure μ_{IGD}.

less complex to analyze than 2D images. This suggests that it might be safer to have a categorical classification for range images.

First-order statistics: Analysis of the relation between local 3d and 2d structure

In this section, we analyze the relation between local 2D structures and local 3D structure; namely, the likelihood of observing a 3D structure given the corresponding 2D structure (i.e., P(3D structure | 2D structure).

Results and discussion

For each pixel of the scene (except where range data is not available), we computed the 3D discontinuity type and the intrinsic dimensionality. Figure 8(a) and (b) show the images where the 3D discontinuity and the intrinsic dimensionality of each pixel are marked with different colors.

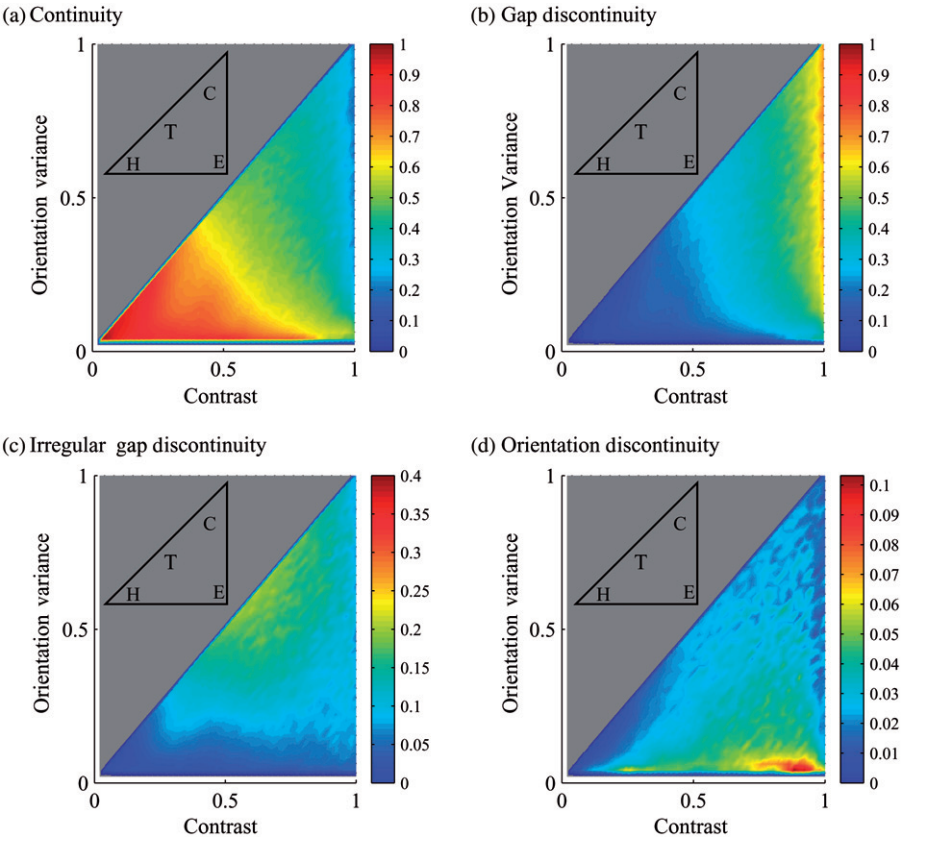

Figure 9. *P*(3D discontinuity | 2D structure). The schematic insets indicate the locations of the different types of 2D structures inside the triangle for easy reference (the letters C, E, H, T represent corner-like, edge-like, homogeneous and texture-like structures). (a) *P*(Continuity | 2D structure). (b) *P*(Gap discontinuity | 2D structure). (c) *P* (Irregular gap discontinuity | 2D structure). (d) *P*(Orientation discontinuity | 2D structure).

Having the 3D discontinuity type and the information about the local 2D structure of each point, we wanted to analyze what the likely underlying 3D structure is for a given local 2D structure; that is, the conditional likelihood (i.e., *P*(3D discontinuity | 2D structure). Using the available 3D discontinuity type and the information about the local 2D structure, other measurements or correlations between the range data and the image data could also be computed in a further study.

P(3D discontinuity | 2D structure) is shown in Figure 9. Note that the four triangles in Figure 9(a–d) add up to one for all points of the triangle.

In Figure 10, maximum likelihood estimates (MLE) of local 3D structures, given local 2D structures, are provided. Figure 10(a) shows the MLE from the distributions in Figure 9. Due to high likelihoods, gap discontinuities and continuities are the most likely estimates, given local 2D structures. Figure 10(b) shows the MLE from the *normalized* distributions: i.e., each triangle in Figure 9 is

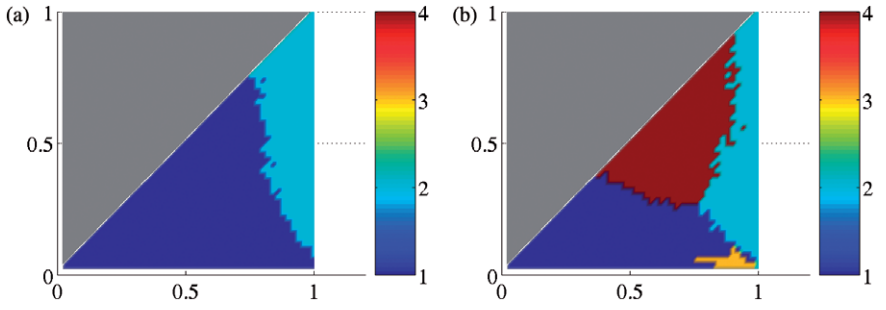

Figure 10. Maximum likelihood estimates of local 3D structures, given local 2D structures. Numbers 1, 2, 3 and 4 represent continuity, gap discontinuity, orientation discontinuity and irregular gap discontinuity, respectively. (a) Raw maximum likelihood estimates. Note that the estimates are dominated by continuities and gap discontinuities. (b) Maximum likelihood estimates from normalized likelihood distributions: the triangles provided in Figure 9 are normalized within themselves so that the maximum likelihood of $P(X \mid \text{2D structure})$ is 1 for X being continuity, gap discontinuity, irregular gap discontinuity and orientation discontinuity.

normalized within itself so that its maximum likelihood is 1. This way we can see the mostly likely *local 2D structures* for different local 3D structures.

- Figure 9(a) shows that homogeneous 2D structures are very likely to be formed by 3D continuities as the likelihood $P(\text{Continuity} \mid \text{2D structure})$ is very high (bigger than 0.85) for the area where homogeneous 2D structures exist (marked with H in Figure 9(a)). This observation is confirmed in the MLE estimates of Figure 10. Many surface reconstruction studies make use of a basic assumption that there is a smooth surface between any two points in the 3D world, if there is no contrast difference between these points in the image. This assumption has been first called as 'no news is good news' in Grimson (1983). Figure 9(a) quantifies 'no news is good news' and shows for which structures and to what extent it holds: in addition to the fact that no news is in fact good news, Figure 9(a) shows that news, especially texture-like structures and edge-like structures, can also be good news (see subsequently). Homogeneous 2D structures cannot be used for depth extraction by correspondence-based methods, and only weak or no information from these structures is processed by the cortex. Unfortunately, the vast majority of local image structure is of this type (Kalkan et al. 2005). On the other hand, homogeneous structures indicate 'no change' in depth which is the underlying assumption of interpolation algorithms.
- Edges are considered as important sources of information for object recognition and reliable correspondence finding. Approximately 10% of local 2D structures are of that type (Kalkan et al. 2005). Figure 9(a–d) together with the MLE estimates in Figure 10 show that most of the edges are very likely to be formed by continuous surfaces or gap discontinuities. Looking at the decision areas for different local 2D structures shown in Figure 2(d), we see that the edges formed by continuous surfaces are mostly low-contrast edges (Figure 9(a)); i.e., the origin variance is close to 0.5. Little percentage of the edges are formed by orientation discontinuities (Figure 9(d)).

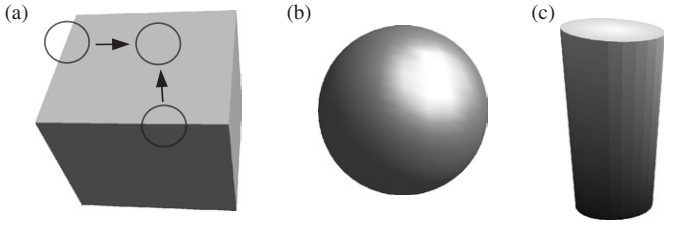

Figure 11. Illustration of the relation between the depth of homogeneous 2D structures and the bounding edges. (a) In the case of the cube, the depth of homogeneous image area and the bounding edges are related. However, in the case of round surfaces, (b) the depth of homogeneous 2D structures may not be related to the depth of the bounding edges. (c) In the case of a cylinder, we see both cases of the relation as illustrated in (a) and (b).

- Figures 9(a) and (b) show that well-defined corner-like structures are formed by either gap discontinuities or continuities.
- Figures 9(d) and 10 show that textures also are very likely to be formed by surface continuities and irregular gap discontinuities. Finding correspondences becomes more difficult with the lack or repetitiveness of the local structure. The estimates of the correspondences at texture-like structures are naturally less reliable. In this sense, the likelihood that certain textures are formed by continuous surfaces (shown in Figure 9(a)) can be used to model stereo-matching functions that include interpolation as well as information about possible correspondences based on the local image information.

It is remarkable that local 2D structures mapping to different sub-regions in the triangle are formed by rather different 3D structures. This clearly indicates that these different 2D structures should be used in different ways for surface reconstruction.

Second-order statistics: Analysis of co-planarity between 3D edges and continuous patches

As already mentioned in the first section, it is not possible to extract depth at homogeneous 2D structures (in the rest of the article, a homogeneous 2D structure that corresponds to a 3D continuity will be called a *mono*) using methods that make use of multiple views for 3D reconstruction. In this section, by making use of the ground truth range data, we investigate co-planarity relations between the depth at homogeneous 2D structures and the edges that bound them. This relation is illustrated for a few examples in Figure 11.

For the analysis, we used the chromatic range data set that we also used for the first-order analysis in the fourth section. Samples from the dataset are displayed in Figure 5.

In the following subsection, we explain how we analyze the relation. The results are presented and discussed in the subsection 'Results and discussions' further.

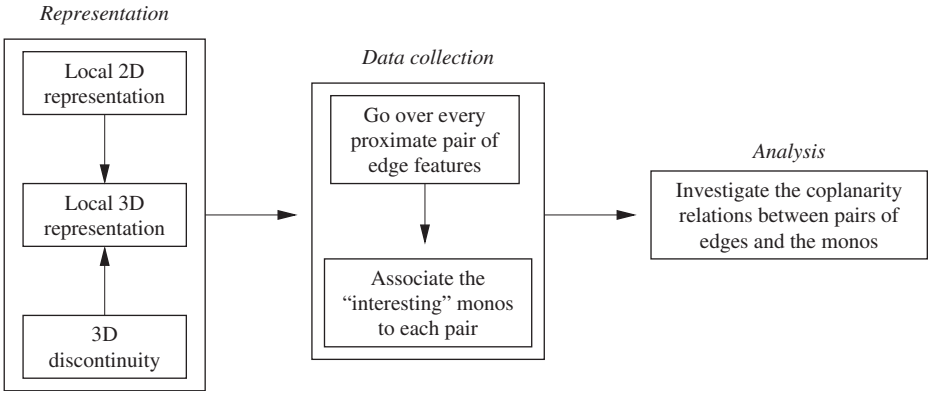

Figure 12. Overview of the analysis process.

Methods

This subsection provides the procedural details of how the analysis is performed.

The analysis is performed in three stages: First, local 2D and 3D representations of the scene are extracted from the chromatic range data. Second, a data set is constructed out of each pair of edge features, associating the monos that are likely to be coplanar to those edges to them (see subsection 'Collecting the data set' for what we mean by relevance). Third, the coplanarity between the monos and the edge features that they are associated to are investigated. An overview of the analysis process is sketched in Figure 12, which roughly lists the steps involved.

Representation. Using the 2D image and the associated 3D range data, a representation of the scene is created in terms of local compository 2D and 3D features denoted by π. In this process, first, 2D features are extracted from the image information, and at the locations of these 2D features, 3D features are computed. The complementary information from the 2D and 3D features are then merged at each valid position, where validity is only defined by having enough range data to extract a 3D representation.

For homogeneous and edge-like structures, different representations are needed due to different underlying structures. For this reason, we have two different definitions of π denoted respectively by π^{e} (for edge-like structures) and π^{m} (for monos) and formulated as:

$$\pi^{\mathrm{m}} = (\mathbf{X}_{3\mathrm{D}}, \mathbf{X}_{2\mathrm{D}}, \mathbf{c}, \mathbf{p}), \tag{9}$$

$$\pi^{\mathrm{e}} = (\mathbf{X}_{3\mathrm{D}}, \mathbf{X}_{2\mathrm{D}}, \phi_{2\mathrm{D}}, \mathbf{c}_1, \mathbf{c}_2, \mathbf{p}_1, \mathbf{p}_2), \tag{10}$$

where $\mathbf{X}_{3\mathrm{D}}$ and $\mathbf{X}_{2\mathrm{D}}$ denote 3D and 2D positions of the 3D entity; $\phi_{2\mathrm{D}}$ is the 2D orientation of the 3D entity; $\mathbf{c}_1$ and $\mathbf{c}_2$ are the 2D color representation of the surfaces of the 3D entity; $\mathbf{c}$ represents the color of π^{m}; $\mathbf{p}_1$ and $\mathbf{p}_2$ are the planes that represent the surfaces that meet at the 3D entity; and $\mathbf{p}$ represents the plane of π^{m} (see Figure 13). Note that π^{m} does not have any 2D orientation information

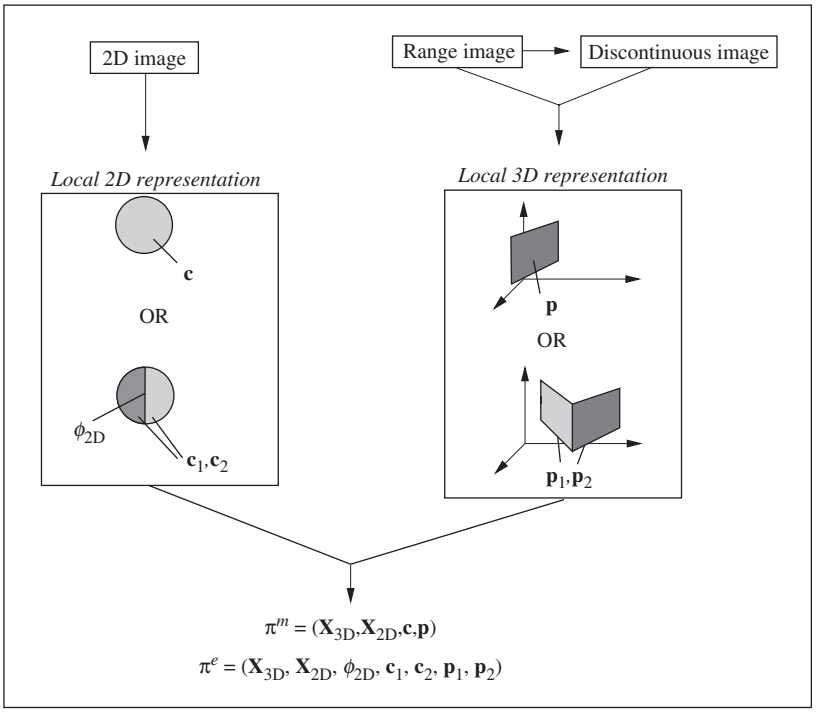

Figure 13. Illustration of the representation of a 3D entity. From the 2D and 3D information, local 2D and 3D representation is extracted.

(because it is undefined for homogeneous structures), and π^e has two color and plane representations to the 'left' and 'right' of the edge.

The process of creating the representation of a scene is illustrated in Figure 13.

In our analysis, the entities are regularly sampled from the 2D information. The sampling size is 10 pixels. See Krügar et al. (2003) and Krüger and Wörgötter (2005) for details.

Extraction of the planar representation requires knowledge about the type of local 3D structure of the 3D entity (see Figure 13). Namely, if the 3D entity is a continuous surface, then only one plane needs to be extracted; if the 3D entity is an orientation discontinuity, then there will be two planes for extraction; if the 3D entity is a gap discontinuity, then there will also be two planes for extraction.

In the case of a continuous surface, a single plane is fitted to the set of 3D points in the 3D entity in question. For orientation discontinuous 3D structures, extraction of the planar representation is not straight-forward. For these structures, our approach was to fit unit-planes[7] to the 3D points of the 3D entity and find the two clusters in these planes using *k*-means clustering of the 3D orientations of the small planes. Then, one plane is fitted for each of the two clusters, producing the bi-fold planar representation of the 3D entity.

Color representation is extracted in a similar way. If the image patch is a homogeneous structure, then the average color of the pixels in the patch is taken to be the color representation. If the image patch is edge-like, then it has two colors separated by the line which goes through the center of the image patch and which

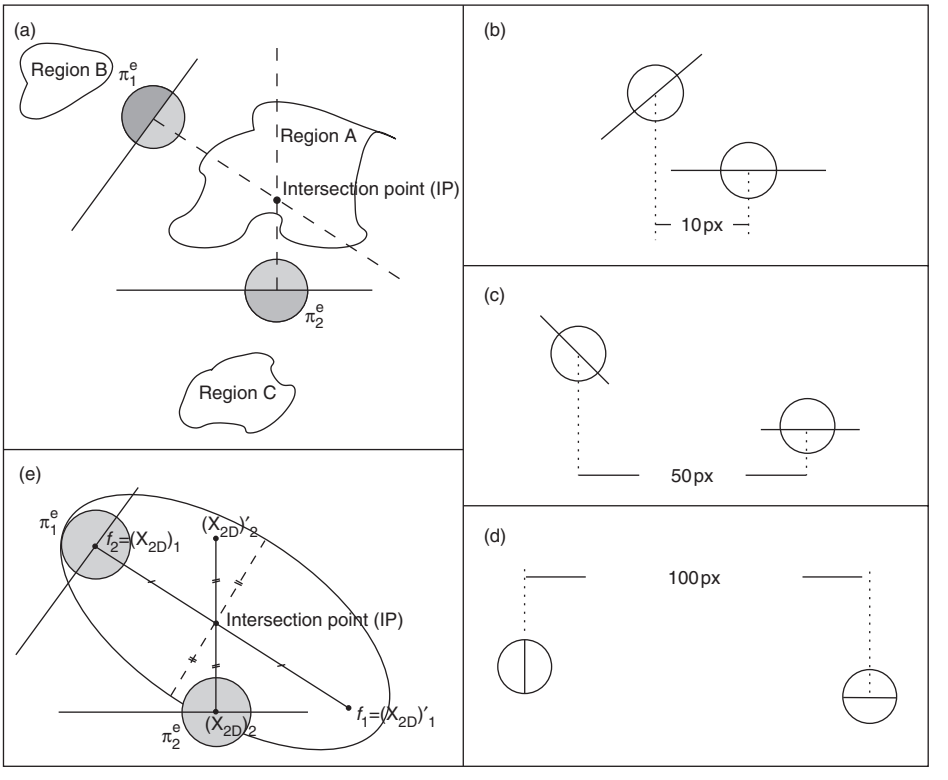

Figure 14. (a) Given a pair of edge features, coplanarity relation can be investigated for homogeneous image patches inside regions A, B and C. However, due to computational intractability reasons, this article is concerned in making the analysis only in region A (see the text for more details). (b)–(d) A few different configurations of edge features that might be encountered in the analysis. The difficult part of the investigation is to make these different configurations comparable, which can be achieved by fitting a shape (like square, rectangle, circle, parallelogram, ellipse) to these configurations. (e) The ellipse, among the alternative shapes (i.e., square, rectangle, circle, parallelogram) turns out to describe the different configurations shown in (b)–(d) better. For this reason, ellipse is for analyzing coplanarity relations in the rest of the article. See the text for details on how the parameters of the ellipse are set.

has the 2D orientation of the image patch. In this case, the averages of the colors of the different sides of the edge define the color representation in terms of c_1 and c_2. If the image patch is corner-like, the color representation becomes undefined.

Collecting the data set. In our analysis, we form pairs out of π^e's that are close enough (see subsequently), and for each pair, we check whether monos in the scene are coplanar to the elements of the pair or not. As there are plenty of monos in the scene, we only consider a subset of monos for each pair of π^e that we suspect to be relevant to the analysis because otherwise, the analysis becomes computationally intractable. The situation is illustrated in Figure 14(a). In this Figure, two π^e and three regions are shown; however, only one of these regions (i.e., region A) is likely to have coplanar monos (Figure 11(a)). This *assumption* is based on the observation

of how objects are formed in the real world: objects have boundaries which consists of edge-like structures who bound surfaces, or image areas, of the object. The image area that is bounded by a pair of edge-like structures is likely to be the area that has the normals of both structures. For convex surfaces of the objects, the area that is bounded belongs to the object; however, in the case of concave surfaces, the area covered may also be from other objects, and the extent of the effect of this is part of the analysis.

Let $\mathcal{P}$ denote the set of pairs of proximate π^{e}'s whose normals intersect. $\mathcal{P}$ can be defined as:

$$\mathcal{P} = \left\{(\pi_1^{e}, \pi_2^{e}) \mid \forall \pi_1^{e}, \pi_2^{e},\ \pi_1^{e} \in \Omega(\pi_2^{e}),\ I(\perp(\pi_1^{e}), \perp(\pi_2^{e}))\right\}, \tag{11}$$

where $\Omega(\pi^{e})$ is the N-pixel-2D-neighborhood of π^{e}[8]; $\perp(\pi^{e})$ is the 2D line orthogonal to the 2D orientation of π^{e}, i.e., the normal of π^{e}; and, $I(l_1, l_2)$ is true if the lines l_1 and l_2 intersect. We have taken N to be 100.

It turns out that there are a lot of different configurations possible for a pair of edge features based on relative position and orientation, which are illustrated for a few cases in Figure 14(b)–(d). The difficult part of the investigation is to be able to compare these different configurations. One way to achieve this is to fit a shape to region A which can *normalize* the coplanarity relations by its size in order to make them comparable (see subsection 'Results and discussions' further for more information).

The possible shapes would be square, rectangle, parallelogram, circle and ellipse. Among the alternatives, it turns out that an ellipse (1) is computationally cheap and (2) fits to different configurations of π_1 and π_2 under different orientations and distances *without* leaving region A much. Figure 14(e) demonstrates the ellipse generated by an example pair of edges in Figure 14(a). The center of the ellipse is at the intersection of the normals of the edges, which we call *the intersection point* (IP) in the rest of the article.

The parameters of an ellipse are composed of two focus points f_1, f_2 and the minor axis b. In our analysis, the more distant 3D edge determines the foci of the ellipse (and, hence, the major axis), and the other 3D edge determines the length of the minor axis. Alternatively, the ellipse can be constructed by minimizing an energy functional which optimizes the area of the ellipse inside region A and going through the features π_1 and π_2. However, for the sake of speed issues, the ellipse is constructed without optimization.

See 'Parameters of an ellipse' in Appendix for details on how we determine the parameters of the ellipse.

For each pair of edges in $\mathcal{P}$, the region to analyze coplanarity is determined by intersecting the normals of the edges. Then, the monos inside the ellipse are associated to the pair of edges.

Note that a π^{e} has two planes that represent the underlying 3D structure. When π^{e}'s become associated to monos, only one plane, the one that points into the ellipse, remains relevant. Let π^{se} denote the semi-representation of π^{e}which can be defined as:

$$\pi^{se} = (\mathbf{X}_{3D}, \mathbf{X}_{2D}, \mathbf{c}, \mathbf{p}). \tag{12}$$

Note that π^{se} is equivalent to the definition of π^{m} in equation (10).

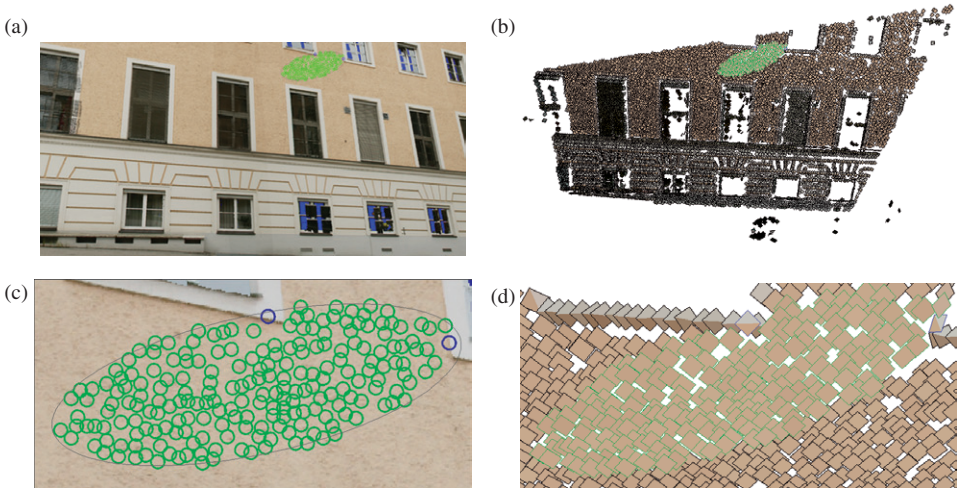

Figure 15. Illustration of a pair of π^{e} and the set of monos associated to them. (a) The input scene. A pair of edges (marked in blue) and the associated monos (marked in green) with an ellipse (drawn in black) around them shown on the input image. See (c) for a zoomed version. (b) The 3D representation of the scene in our 3D visualization software. This representation is created from the range data corresponding to (a) and is explained in the text. (c) The part of the input image from (a) where the edges, the monos and the ellipse are better visible. (d) A part of the 3D representation (from (b)) corresponding to the pair of edges and the monos in (c) is displayed in detail where the edges are shown with blue margins; the monos with the edges are shown in green (all monos are coplanar with the edges). The 3D entities are drawn in rectangles because of the high computational complexity for drawing circles.

Let $\mathcal{T}$ denote the data set which stores $\mathcal{P}$ and the associated monos which can be formulated as:

$$\mathcal{T} = \left\{(\pi_1^{\mathrm{se}}, \pi_2^{\mathrm{se}}, \pi^{\mathrm{m}}) \mid (\pi_1^{\mathrm{e}}, \pi_2^{\mathrm{e}}) \in \mathcal{P}, \pi^{\mathrm{m}} \in \mathcal{S}^{\mathrm{m}}, \pi^{\mathrm{m}} \in E(\pi_1^{\mathrm{e}}, \pi_2^{\mathrm{e}})\right\}, \tag{13}$$

where $\mathcal{S}^{\mathrm{m}}$ is the set of all π^{m}.

A pair of π^{e}'s and the set of monos associated to them are illustrated in Figure 15. The figure shows the edges and the monos (together with ellipse) in 2D and 3D.

Definition of coplanarity. Two entities are coplanar if they are on the same plane. Coplanarity of edge features and monos is equivalent to coplanarity of two planar patches: two planar patches A and B are coplanar if (1) they are parallel and (2) the planar distance between them is zero.

See 'Definition of coplanarity' in Appendix for more information.

Results and discussions

The data set $\mathcal{T}$ defined in Equation (13) consists of pairs of π_1^{e}, π_2^{e} and the associated monos. Using this set, we compute the likelihood that a mono is coplanar with π_1^{e} and/or π_2^{e} against a distance measure.

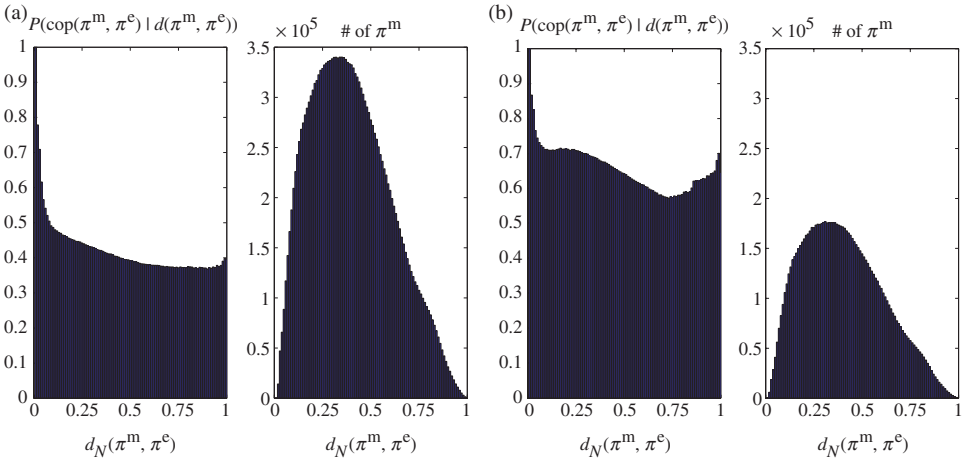

Figure 16. Likelihood distribution of coplanarity of monos. In each sub-figure, left-plot shows the likelihood distribution whereas right-plot shows the frequency distribution. (a) The likelihood of the coplanarity of a mono with π_1^e or π_2^e against the distance to π_1^e *or* π_2^e . This is the unconstrained case; i.e., the case where there is no information about the coplanarity of π_1^e and π_2^e . (b) The likelihood of the coplanarity of a mono with π_1^e and π_2^e against the distance to π_1^e *or* π_2^e .

The results of our analysis are shown in Figures 16, 18 and 19.

In Figure 16(b), the likelihood of the coplanarity of a mono against the distance to π_1^e or π_2^e is shown. This likelihood can be denoted formally as $P(\mathrm{cop}(\pi^m, \pi_1^e \text{ and } \pi_2^e) \mid d_N(\pi^m, \pi^e))$ where $\mathrm{cop}(\pi^m, \pi_1^e \text{ and } \pi_2^e)$ is defined as $\mathrm{cop}(\pi_1^e, \pi_2^e) \wedge cop(\pi^m, \pi^e)$, and π^e is either π_1^e or π_2^e. The normalized distance measure[9] $d_N(\pi^m, \pi^e)$ is defined as:

$$d_N(\pi^m, \pi^e) = \frac{d(\pi^m, \pi^e)}{2\sqrt{d(\pi_1^e, \mathrm{IP})^2 + d(\pi_2^e, \mathrm{IP})^2}}, \tag{14}$$

where π^e is either π_1^e or π_2^e, and IP is the intersection point of π_1^e and π_2^e. We see in Figure 16(b) that the likelihood decreases when a mono is more distant from an edge. However, when the distance measure gets closer to one, the likelihood increases again. This is because, when a mono gets away from either π_1^e or π_2^e, it gets closer to the other π^e.

In Figure 16(a), we see the unconstrained case of Figure 16(b); i.e., the case where there is no information about the coplanarity of π_1^e and π_2^e; namely, the likelihood $P(\mathrm{cop}(\pi^m, \pi^e) \mid d_N(\pi^m, \pi^e))$ where π^e is either π_1^e or π_2^e. The comparison with Figure 16(b) shows that the existence of another edge in the neighborhood increases the likelihood of finding coplanar structures. As there is no other coplanar edge in the neighborhood, the likelihood does not increase when the distance is close to one (compare with Figure 16(b)).

It is intuitive to expect symmetries in Figure 16. However, as (1) the roles of π_1^e and π_2^e in the ellipse are fixed, and (2) one π^e is guaranteed to be on the major axis, and the other π^e may or may not be on the minor axis, the symmetry is not observable in Figure 16.

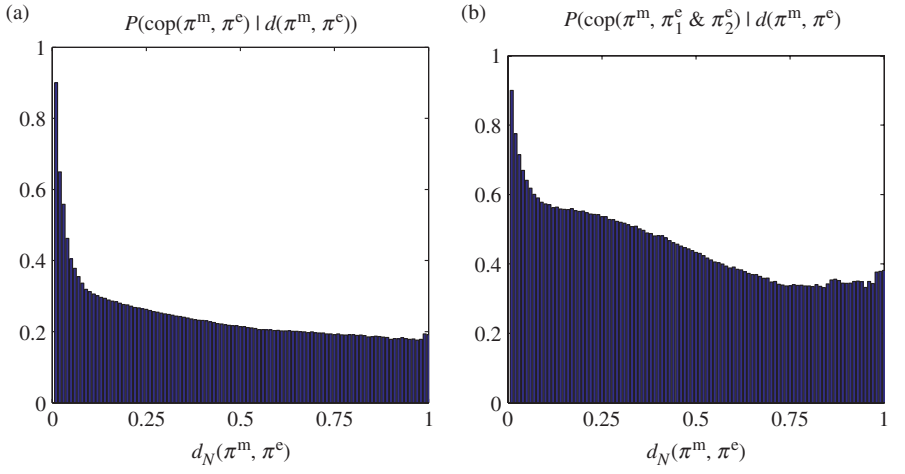

Figure 17. Likelihoods from Figure 16(a) and (b) with a more *strict* coplanarity relation (namely, we set the thresholds T_p and T_d to 10 degrees and 0.2, respectively. See Appendix for more information about these thresholds). (a) Figure 16(a) with more strict coplanarity relation. (b) Figure 16(b) with more strict coplanarity relation.

To see the effect of the coplanarity relation on the results, we reproduced Figure 16(a) and (b) with a more *strict* coplanarity relation (namely, we set the thresholds T_p and T_d to 10 degrees and 0.2, respectively. See Appendix for more information about these thresholds). The results with more constrained coplanarity relation are shown in Figure 17. Although the likelihood changes quantitatively, the figure shows the qualitative behaviours that have been observed with the standard thresholds. Moreover, we cross-checked the results for subsets of the original dataset (results not provided here) and confirmed the same qualitative results.

In Figure 18, the likelihood of the coplanarity of a mono against the distance to IP (i.e., $P(\text{cop}(\pi^m, \pi_1^e \text{ and } \pi_2^e) \mid d_N(\pi^m, \text{IP}))$) is shown. We see in the figure that the likelihood shows a flat distribution against the distance to IP.

In Figure 19, the likelihood of the coplanarity of a mono against the distance to π_1^e and π_2^e (i.e., $P(\text{cop}(\pi^m, \pi_1^e \text{ and } \pi_2^e) \mid d_N(\pi^m, \pi_1^e), d_N(\pi^m, \pi_2^e))$) is shown. We see that when π^m is close to π_1^e or π_2^e, it is more likely to be coplanar with π_1^e and π_2^e than when it is equidistant to both edges. The reason is that, when π^m moves away from an equidistant point, it becomes closer to the other edge, in which case the likelihood increases as shown in Figure 16(a).

The results, especially Figure 16(a) and (b) confirms the importance of the relation illustrated in Figure 11(a).

Discussion

Summary of the findings

The 'Results and discussion' in the fourth section analyzed the likelihood P(3D structure | 2D structure). In this section, we confirm and quantify the

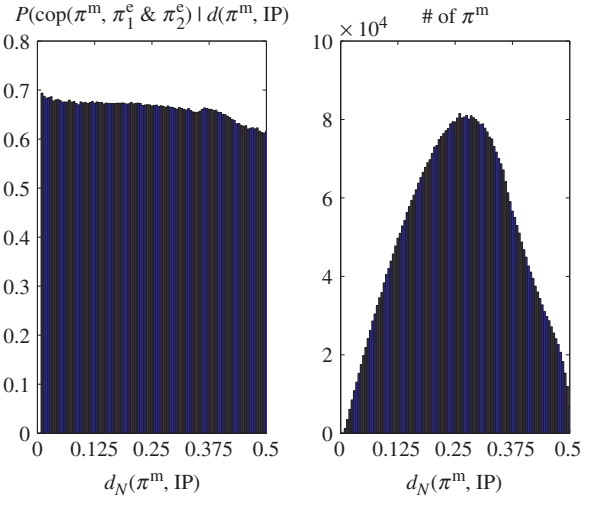

Figure 18. The likelihood of the coplanarity of a mono against the distance to IP. Left-plot shows the likelihood distribution whereas right-plot shows the frequency distribution.

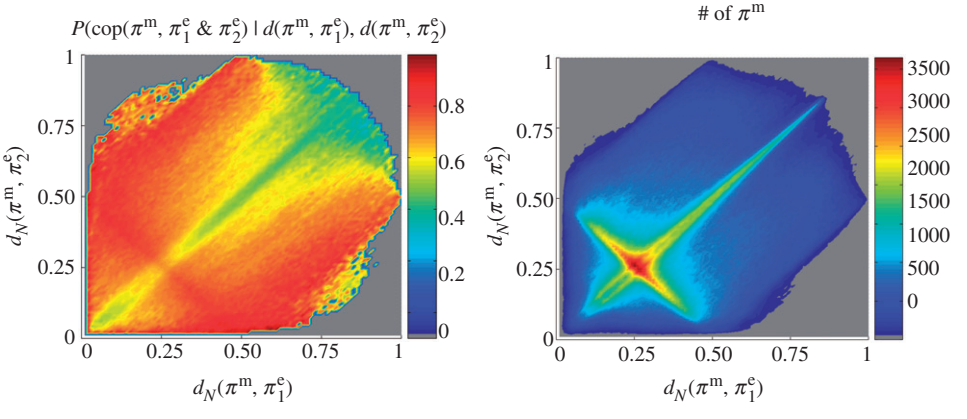

Figure 19. The likelihood of the coplanarity of a mono against the distance to π_1^e and π_2^e. Left-plot shows the likelihood distribution whereas right-plot shows the frequency distribution.

assumptions used in several surface interpolation studies. Our main findings from this section are as follows:

- As expected, homogeneous 2D structures are formed by continuous surfaces.
- Surprisingly, considerable amount of edges and texture-like structures are likely to be formed by continuous surfaces too. However, we confirm the expectation that gap discontinuities and orientation discontinuities are likely to be the underlying 3D structure for edge-like structures. As for texture-like structures, they may also be formed by irregular gap discontinuities.
- Corner-like structures, on the other hand, are mainly formed by gap discontinuities.

In 'Results and discussion' of the fifth section, we investigated the predictability of depth at homogeneous 2D structures. We confirm the basic assumption that closer entities are very likely to be coplanar. Moreover, we provide results showing that this likelihood increases if there are more edge features in the neighborhood.

Interpretation of the findings

Existing psychophysical experiments (Collett 1985; Anderson et al. 2002), computational theories (Barrow and Tenenbaum 1981; Grimson 1982; Terzopoulos 1988) and the observation that humans can perceive depth at weakly textured areas suggest that in the human visual system, *an interpolation process* is realized that, starting with the local analysis of edges, corners and textures, computes depth also in areas where correspondences cannot easily be found.

This article was concerned with the analysis of the statistics that might be involved in such an interpolation process, by making use of chromatic range data.

In the first part (fourth section), we analyzed which local 2D structures suggest a depth interpolation process. Using natural images, we showed that homogeneous 2D structures correspond to continuous surfaces, as suggested and utilized by some computational theories of surface interpolation (Grimson 1983). On the other hand, a considerable proportion of edge-like structures lie on continuous surfaces (Figure 9(a)); i.e., a contrast difference does not necessarily mean a depth discontinuity. This suggests that interpreting edges in combination with neighboring corners or edges is important for understanding the underlying 3D structure (Barrow and Tenenbaum 1981).

The results from the fourth section are useful in several contexts:

- Depth interpolation studies assume that homogeneous image regions are part of the same surface. Such studies can be extended with the statistics provided here as priors in a Bayesian framework. This extension would allow making use of the continuous surfaces that a contrast difference (caused by textures or edge-like structures) might correspond to. Acquiring range data from a scene is a time-consuming task compared to image acquisition, which lasts on the order of seconds even for high resolutions. In Torres-Mendez and Dudek (2006), for mobile robot environment modeling, instead of making a full-scan of the whole scene, only partial range scan is performed due to time constraints. This partial range data is completed by using a Markov Random Field which is trained from a pair of complete range and the corresponding image data. In Torres-Mendez and Dudek (2006), the partial range data is produced in a regular way; i.e., every n-th scan-column is neglected. This assumption, however, may introduce aliasing in the 3D data acquired from natural images using depth cues, and therefore, their method may not be applicable. Nevertheless, it could possibly be improved by utilizing the priors introduced in this article.
- Automated registration of range and color images of a scene is crucial for several purposes like extracting 3D models of real objects. Methods that align edges extracted from the intensity image with the range data already exist (Laycock and Day 2006). These methods can be extended with the results presented in this article in a way that not only edges but also other 2D structures are used for alignment. Such an extension also allows a probabilistic framework by utilizing

the likelihood P(3D structure | 2D structure). Moreover, making use of local 3D structure types that are introduced in this article can be more robust than just a gap discontinuity detection. Such an extension is possible by maximizing the following energy function:

$$E(R, T) = \int_{u,v} P(\text{3D structure at } (u, v) \mid \text{2D structure at } (u, v))\mathrm{d}u\mathrm{d}v, \qquad (15)$$

where R and T are translation and rotation of the range data in 3D space respectively.

In the second part (fifth section), we analyzed whether depth at homogeneous 2D structures is related to the depth of edge-like structures in the neighborhood. Such an analysis is important for understanding the possible mechanisms that could underlie depth interpolation processes. Our findings show that an edge feature provides significant evidence for making depth prediction at a homogeneous image patch that is in the neighborhood. Moreover, the existence of a second edge feature in its neighborhood which is not collinear with the first edge feature increases the likelihood of the prediction.

Using second-order relations and higher order features for representing the 2D image and 3D range data, we produce confirming results that the range images are simpler to analyze compared to 2D images (Huang et al. 2000; Yang and Purves 2003).

By extracting a more complex representation than existing range-data analysis studies, we could point to the intrinsic properties of the 3D world and its relation to the image data. This analysis is important because (1) it may be that the human visual system is adapted to the statistics of the environment (Brunswik and Kamiya 1953; Knill and Richards 1996; Olshausen and Field 1996; Krueger 1998; Purves and Lotto 2002; Rao et al. 2002), and (2) it may be used in several computer vision applications (for example, depth estimation) in a similar way as in (Zhu 1999; Elder and Goldberg 2002; Elder et al. 2003; Pugeault et al. 2004).

In our current work, the likelihood distributions are being used for estimating the 3D depth at homogeneous 2D structures from the depth of bounding edge-like structures.

Limitations of the current work

The first limitation is due to the type of scenes that have been used; i.e., scenes of man-made environments which also included trees. Alternative scenes could include pure forest scenes or scenes taken from an environment with totally round objects. However, we believe that our data set captures the general properties of the scenes that a human being encounters in daily life.

Different scenes might produce quantitatively different but qualitatively similar results. For example, forest scenes would produce much more irregular gap discontinuities than the current scenes; however, our conclusions regarding the link between textures and irregular gap discontinuities would still hold. Moreover, coplanarity relations would be harder to predict for such scenes since (depending on the scale) surface continuities are harder to find; however, on a bigger scale, some forest scenes are likely to produce the same qualitative results presented in

this article because of piecewise planar leaves which are separated by gap discontinuities.

It should be noted that acquisition of range data with color images is very hard for forest scenes since the color image of the scene is taken after the scene is scanned with the scanner. During this period, the leaves and the trees may move (due to wind etc.), making the range and the color data inconsistent. In office environments, a similar problem arises: due to lateral separation between the digital camera and range scanner, there is the parallax problem, which again produces inconsistent range–color association. For an office environment, a small-scale range scanner needs to be used.

The statistics presented in this article can be extended by analyzing forest scenes, office scenes etc. independently. The comparison of such independent analyses should provide more insights into the relations that this article have investigated but we believe that the qualitative conclusions of this article would still hold.

It would be interesting to see the results presented in the article by changing the measure for surface continuity so that it can separate planar and curved surfaces. We believe that such a change would effect only the second part of the article.

Acknowledgements

We would like to thank RIEGL UK Ltd for providing us with 3D range data, and Nicolas Pugeault for his comments on the text. This work is supported by the European-funded DRIVSCO project, and an extension of two conference publications of the authors: Kalkan et al. (2006); Kalkan et al. (2007).

Notes

[1] In this article, a relation is first-order if it involves two entities and an event between them. Analogously, a is second-order if there are three entities and (at least) two events between them.

[2] In this article, chromatic 3D range data means range data which has associated real-world color information. The color information is acquired using a digital camera which is calibrated with the range scanner.

[3] In this article, for the sake of simplicity, junctions are called corners, too.

[4] Note that XYZ and RGB coordinate systems are not the same. However, detection of gap discontinuity in XYZ coordinates can be assumed to be a special case of edge detection in RGB coordinates.

[5] Note that using bigger planes have the disadvantage of losing accuracy in positioning which is very crucial for the current analysis.

[6] Singular value decomposition is a standard technique for fitting planes to a set of points. It finds the perfectly fitting plane if it exists; otherwise, it returns the least-squares solution.

[7] By unit-planes, we mean planes that are fitted to the 3D points that are 1-pixel apart in the 2D image.

[8] In other words, the Euclidean image distance between the structures should be less than N.

[9] In the following plots, the distance means the Euclidean distance in the image domain.

References

Anderson BL, Singh M, Fleming RW. 2002. The interpolation of object and surface structure. Cognitive Psychology 44(43):148–190.

Barrow HG, Tenenbaum JM. 1981. Interpreting line drawings as three-dimensional surfaces. Artificial Intelligence 17:75–116.

Bolle RM, Vemuri BC. 1991. On three-dimensional surface reconstruction methods. IEEE Transactions on Pattern Analysis and Machine Intelligence 13(1):1–13.

Bruce V, Green PR, Georgeson MA. 2003. Visual perception: Physiology, psychology and ecology. 4th ed. East Sussex, UK: Psychology Press.

Brunswik E, Kamiya J. 1953. Ecological cue–validity of proximity and of other Gestalt factors. American Journal of Psychologie, LXVI:20–32.

Collett TS. 1985. Extrapolating and Interpolating Surfaces in Depth. Royal Society of London Proceedings Series B 224:43–56.

Coxeter H. 1969. Introduction to geometry. 2nd ed. New York: Wiley & Sons.

Elder H, Goldberg R. 2002. Ecological statistics of gestalt laws for the perceptual organization of contours. Journal of Vision 2(4):324–353.

Elder JH, Krupnik A, Johnston LA. 2003. Contour grouping with prior models. IEEE Transactions on Pattern Analysis and Machine Intelligence 25(25):1–14.

Felsberg M, Krüger N. 2003. A probablistic definition of intrinsic dimensionality for images. Pattern Recognition, 24th DAGM Symposium. Springer Berlin/Heidelberg.

Field DJ, Hayes A, Hess RF. 1993. Contour integration by the human visual system: Evidence for a local association field. Vision Research 33(2):173–193.

Gallant JL, Essen DCV, Nothdurft HC. 1994. Early vision and beyond, chapter: Two-dimensional and three-dimensional texture processing in visual cortex of the macaque monkey. MA: MIT Press. pp 89–98.

Grimson WEL. 1982. A computational theory of visual surface interpolation. Royal Society of London Philosophical Transactions Series B 298:395–427.

Grimson WEL. 1983. Surface consistency constraints in vision. Computer Vision, Graphics and Image Processing 24(1):28–51.

Guzman A. 1968. Decomposition of a visual scene into three-dimensional bodies. AFIPS Fall Joint Conference Proceedings 33:291–304.

Hoover A, Jean-Baptiste G, Jiang X, Flynn PJ, Bunke H, Goldgof DB, Bowyer K, Eggert DW, Fitzgibbon A, Fisher RB. 1996. An experimental comparison of range image segmentation algorithms. IEEE Transactions on Pattern Analysis and Machine Intelligence 18(7):673–689.

Howe CQ, Purves D. 2002. Range image statistics can explain the anomalous perception of length. PNAS 99(20):13184–13188.

Howe CQ, Purves D. 2004. Size contrast and assimilation explained by the statistics of natural scene geometry. Journal of Cognitive Neuroscience 16(1):90–102.

Huang J, Lee AB, Mumford D. 2000. Statistics of range images. CVPR 1(1):1324–1331.

Hubel D, Wiesel T. 1969. Anatomical demonstration of columns in the monkey striate cortex. Nature 221:747–750.

Kalkan S, Calow D, Wörgötter F, Lappe M, Krüger N. 2005. Local image structures and optic flow estimation. Network: Computation in Neural Systems 16(4):341–356.

Kalkan S, Wörgötter F, Krüger N. 2006. Statistical analysis of local 3d structure in 2d images. CVPR 1:1114–1121.

Kalkan S, Wörgötter F, Krüger N. 2007. Statistical analysis of second-order relations of 3d structures. Int. Conference on Computer Vision Theory and Applications (VISAPP), Barcelona, Spain.

Kellman P, Arterberry M, editors. 1998. The cradle of knowledge. Cambridge, Mass: MIT-Press.

Knill DC, Richards W, editors. 1996. Perception as bayesian inference. Cambridge: Cambridge University Press.

Koenderink J, Dorn A. 1982. The shape of smooth objects and the way contours end. Perception 11:129–173.

Krueger N. 1998. Collinearity and parallelism are statistically significant second order relations of complex cell responses. Neural Processing Letters 8(2):117–129.

Krüger N, Felsberg M. 2003. A continuous formulation of intrinsic dimension. Proceedings of the British Machine Vision Conference, Norwich, UK.

Krüger N, Lappe M, Wörgötter F. 2003. Biologically motivated multi-modal processing of visual primitives. Proc. the AISB 2003 Symposium on Biologically inspired Machine Vision, Theory and Application, Wales, pp 53–59.

Krüger N, Wörgötter F. 2005. Multi-modal primitives as functional models of hyper-columns and their use for contextual integration. Proc. 1st Int. Symposium on Brain, Vision and Artificial Intelligence, Naples, Italy, Lecture Notes in Computer Science, Springer, LNCS 3704, pp 157–166.

Laycock RG, Day AM. 2006. Image registration in a coarse three dimensional virtual environment. Computer Graphics Forum 25(1):69–82.

Lee TS, Mumford D, Romero R, Lamme VAF. 1998. The role of the primary visual cortex in higher level vision. Vision Research 38:2429–2454.

Malik J. 1987. Interpreting line drawings of curved objects. International Journal of Computer Vision 1:73–103.

Marr D. 1982. Vision: A computational investigation into the human representation and processing of visual information. San Fransisco, CA: Freeman.

Olshausen B, Field D. 1996. Natural image statistics and efficient coding. Network 7:333–339.

Potetz B, Lee TS. 2003. Statistical correlations between two-dimensional images and three-dimensional structures in natural scenes. Journal of the Optical Society of America 20(7):1292–1303.

Pugeault N, Krüger N, Wörgötter F. 2004. A non-local stereo similarity based on collinear groups. Proceedings of the Fourth International ICSC Symposium on Engineering of Intelligent Systems, Island of Madeira, Portugal.

Purves D, Lotto B, editors. 2002. Why we see what we do: An empirical theory of vision. Sunderland, MA: Sinauer Associates.

Rao RPN, Olshausen BA, Lewicki MS, editors. 2002. Probabilistic models of the brain. MA: MIT Press.

Rubin N. 2001. The role of junctions in surface completion and contour matching. Perception 30:339–366.

Sereno ME, Trinath T, Augath M, Logothetis NK. 2002. Three-dimensional shape representation in monkey cortex. Neuron 33(4):635–652.

Shirai Y. 1987. Three-dimensional computer vision. New York: Springer-Verlag, Inc.

Simoncelli EP. 2003. Vision and the statistics of the visual environment. Current Opinion in Neurobiology 13(2):144–149.

Terzopoulos D. 1988. The computation of visible-surface representations. IEEE Trans. Pattern Anal. Mach. Intell. 10(4):417–438.

Torres-Mendez LA, Dudek G. 2006. Statistics of visual and partial depth data for mobile robot environment modeling. Mexican International Conference on Artificial Intelligence (MICAI), Apizaco, Mexico.

Tuceryan M, Jain NK. 1998. Texture analysis. The handbook of pattern recognition and computer vision. 2nd ed. pp 207–248.

Yang Z, Purves D. 2003. Image/source statistics of surfaces in natural scenes. Network: Computation in Neural Systems 14:371–390.

Zetzsche C, Barth E. 1990. Fundamental limits of linear filters in the visual processing of two dimensional signals. Vision Research 30(7):1111–1117.

Zhu SC. 1999. Embedding gestalt laws in markov random fields. IEEE Transactions on Pattern Analysis and Machine Intelligence 21(11):1170–1187.

Appendix

Parameters of an ellipse

Let us denote the position of two 3D edges π_1^e, π_2^e by $(\mathbf{X}_{2D})_1$ and $(\mathbf{X}_{2D})_2$ respectively. The vectors between the 3D edges and IP (let us call l_1 and l_2) can be defined as:

$$\begin{aligned} l_1 &= ((\mathbf{X}_{2D})_1 - IP), \\ l_2 &= ((\mathbf{X}_{2D})_2 - IP). \end{aligned} \tag{A1}$$

Having defined l_1 and l_2, the ellipse $E(\pi_1^e, \pi_2^e)$ is as follows:

$$E(\pi_1^e, \pi_2^e) = \begin{cases} f_1 = (\mathbf{X}_{2D})_1, f_2 = (\mathbf{X}_{2D})'_1, b = |l_2| & \text{if } |l_1| > |l_2|, \\ f_1 = (\mathbf{X}_{2D})_2, f_2 = (\mathbf{X}_{2D})'_2, b = |l_1| & \text{otherwise.} \end{cases} \tag{A2}$$

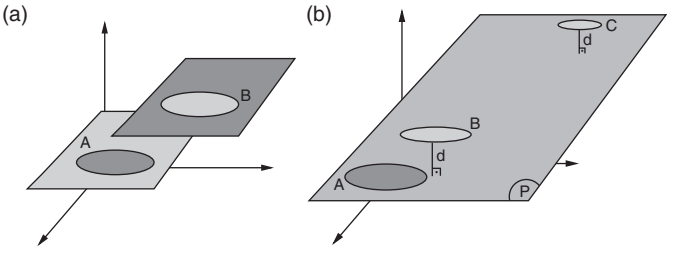

Figure A1. Criteria for coplanarity of two planes. (a) According to the angular-difference criterion of coplanarity, entities A and B will be measured as coplanar although they are on different planes. In (b), P is the plane defined by entity A. According to the distance-based coplanarity definition, entities B and C have the same measure of coplanarity. However, entity C which is more distant to entity A should have a higher measure of coplanarity than entity B although they have the same distance to plane P (see the text).

where $(\mathbf{X}_{2D})'$ is symmetrical with $\mathbf{X}_{2D}$ around the intersection point and on the line defined by $\mathbf{X}_{2D}$ and IP (as shown in Figure 14(e)).

Definition of coplanarity

Let π^s denote either a semi-edge π^{se} or a mono π^m. Two π^s are coplanar iff they are on the same plane. When it comes to measuring coplanarity, two criteria need to be tested:

(i) Angular criterion: For two π^s to be coplanar, the angular difference between the orientation of the planes that represent them should be less than a threshold. A situation is illustrated in Figure A1(a) where angular criterion holds but the planes are not coplanar.

(ii) Distance-based criterion: For two π^s to be coplanar, the distance between the center of the first π^s and the plane defined by the other π^s should be less than a threshold. In Figure A1(b), B and C are at the same distance to the plane P which is the plane defined by the planar patch A. However, C is more distant to the center of A than B, and in this article, we treat that C is more coplanar to A than B is to A. The reason for this can be clarified with an example: assume that A, B and C are all parallel, and that the *planar* and the Euclidean distances between A and B are both D units, and between A and C are respectively D and $n \times D$. It is straightforward to see that although B and C have the same planar distances to A, for $n \gg 1$, C should have a higher coplanarity measure.

It is sufficient to combine these two criteria as follows:

$$\begin{aligned} \mathrm{cop}(\pi_1^s, \pi_2^s) = \alpha(\mathbf{p}^{\pi_1^s},\ \mathbf{p}^{\pi_2^s}) < T_p \text{ and} \\ d(\mathbf{p}^{\pi_1^s},\ \pi_2^s)/d(\pi_1^s,\ \pi_2^s) < T_d, \end{aligned} \tag{A3}$$

where $\mathbf{p}^{\pi^s}$ is the plane associated to π^s; $\alpha(\mathbf{p}_1, \mathbf{p}_2)$ is the angle between the orientations of $\mathbf{p}_1$ and $\mathbf{p}_2$; and, $d(.,.)$ is the Euclidean distance between two entities.

In our analysis, we have empirically chosen T_p and T_d as 20 degrees and 0.5, respectively. Again, like the parameters set in 'Combining the measures' in third section, these values are determined by testing the coplanarity measure over

different samples. T_p is the limit for angular separation between two planar patches. Bigger values would relax the coplanarity measure, and vice versa. T_d restricts the distances between the patches; in analogy to T_p, T_d can be used to relax the coplanarity measure. As shown in Figure 17 for a stricter coplanarity definition (with T_p and T_d set to 10 degrees and 0.2), different values for these thresholds would quantitatively but not qualitatively change the results presented in the fifth section.