STATISTICAL ANALYSIS OF SECOND-ORDER RELATIONS OF 3D STRUCTURES

Preparation of Camera-Ready Contributions to INSTICC Proceedings

Sinan Kalkan, Florentin Wörgötter
Bernstein Centre for Computational Neuroscience, Unv. of Göttingen, Germany
{*sinan,worgott*}*@bccn-goettingen.de*

Norbert Krüger
Cognitive Vision Group, Unv. of Southern Denmark, Denmark
norbert@mip.sdu.dk

Keywords: Range Data Statistics, Indirect Depth Estimation

Abstract: Algorithmic 3D reconstruction methods like stereopsis or structure from motion fail to extract depth at homogeneous image structures where the human visual system succeeds and is able to estimate depth. In this paper, using chromatic 3D range data, we analyze in which way depth in homogeneous structures is related to the depth at the bounding 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 can statistically show that the likelihood to find a certain depth at a homogeneous image patch depends on the distance between the image patch and its edges. Furthermore, we find that this prediction is higher when there is a second edge which is proximate to and coplanar with the first edge. These results allow deriving statistically based prediction models for depth extrapolation into homogeneous image structures. We present initial results of a model that predicts depth based on these statistics.

1 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, 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, utilization of texture gradients or linear perspective use statistical and geometrical relations in 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 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.

In figure 1, the relation between the depth of homogeneous image structures and edges is shown. In figure 1(a), we see that the depth of homogeneous image structures is directly related to the depth of the bounding edges; however, this relation does not always exist as shown in figure 1(b,c) where the depth is cued in shading.

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

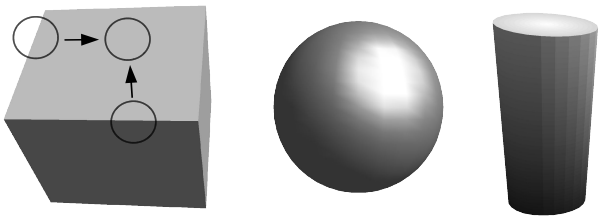
In this paper, by making use of chromatic range data (see figure 3 for examples), we investigate

whether the depth at homogeneous image structures are related to or predictable by the depth of the edges that bound them. This investigation is important because (1) it contributes to a better understanding of the intrinsic parameters of the 3D world, and (2) it suggests an indirect method to estimate the depth for homogeneous image structures; that is, using the depth estimations about the edges to predict the depth of homogeneous image structures instead of using the 2D image information itself as shown in figure 1(a).

There have been only a few studies that have investigated the 3D world from range data (Howe and Purves, 2004; Huang et al., 2000; Kalkan et al., 2006; Potetz and Lee, 2003; Yang and Purves, 2003). 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 (Huang 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 in shape from shading algorithms for estimating depth. In (Howe and Purves, 2002; Howe and Purves, 2004), range image statistics were analyzed for explanation of several visual illusions.

In (Kalkan et al., 2006), a higher-order representation of the 2D local image patches and the 3D local patches were considered; they represented 2D images in terms of homogeneous, edge-like and corner-like structures whereas 3D range data in terms of continuities, gap discontinuities and orientation discontinuities (see section 2). With these representations, they could compute the probability $P(\text{3D Structure} \mid \text{2D Structure})$ which among other things justifies and quantifies the assumption that if two image points do not have contrast difference in-between, then they are likely to be coplanar. This assumption is called 'no news is good news' and widely used in 3D reconstruction studies (see, *e.g.*, (Grimson, 1983)).

All the studies discussed above are first-order, analyzing the relation between the image data and the range data. In this work, however, we are interested in higher order relations between local 3D features. In this sense, our work is a natural extension of (Kalkan et al., 2006).



(a) (b) (c)

Figure 1: Illustration of the relation between the depth of homogeneous image structures and the bounding edges. (a) In the case of cube, the relation is eminent. However, in the case of round surfaces, (b) the depth of homogeneous image 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).

The outline of the paper is as follows: In section 2, different types of local 3D structures are introduced. In section 3, the methodology underlying our statistical analysis is presented. The results are presented and discussed in section 4. Finally, in section 5, the paper is concluded.

2 LOCAL 3D STRUCTURE TYPES

For our work, we have made use of the classification introduced in (Kalkan et al., 2006) where it is intuitively argued that 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.
- Regular Gap discontinuity: The underlying 3D structure can be described by a small set of surfaces with a significant depth difference. An example of gap discontinuity is shown in figure 2(d).
- Irregular Gap discontinuity: The underlying 3D structure shows high depth variation and cannot be described by two or three surfaces. An example of an irregular gap discontinuity is shown in figure 2(e).
- Orientation Discontinuity: The underlying 3D structure can be described by two surfaces with significantly different 3D orientations that meet at the point whose 3D structure is being questioned. In this type of discontinuity, no gap but a change

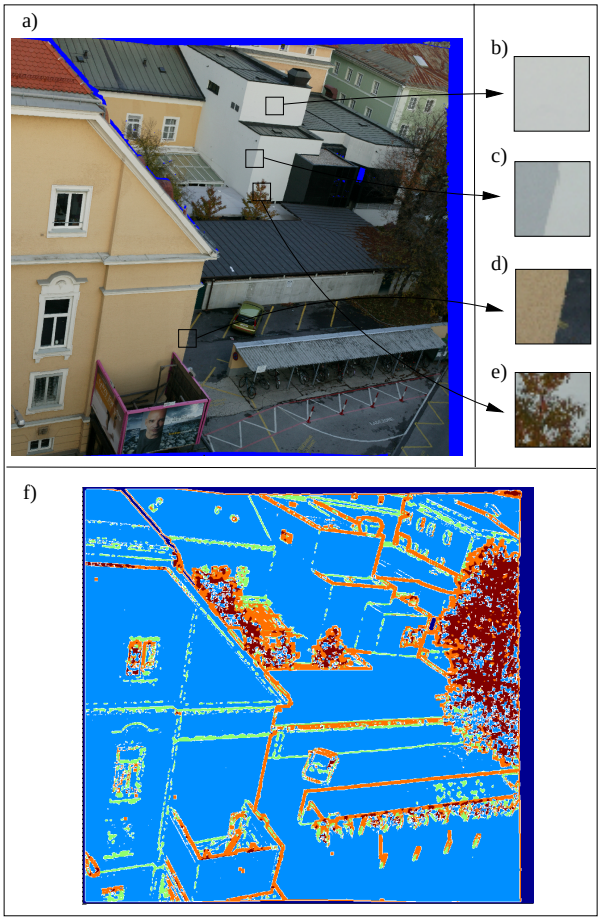


Figure 2: Illustration of the types of 3D discontinuities. (a) 2D image. (b) Continuity. (c) Orientation discontinuity. (d) Gap discontinuity. (e) Irregular gap discontinuity. (f) 3D discontinuity of each pixel is shown in different colors. Blue: continuous surfaces, light blue: orientation discontinuities, orange: gap discontinuities and brown: irregular gap discontinuities. Dark blue indicates points without range data.

in 3D orientation between the meeting surfaces occurs. An example for this type of discontinuity is shown in figure 2(c).

3D discontinuities are detected in studies which involve range data processing, using different methods and using 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 (Bolle and Vemuri, 1991; Hoover et al., 1996; Shirai, 1987).

For our analysis, we have adopted the measures defined in (Kalkan et al., 2006). In this work, a gap discontinuity is measured by simple edge detection in XYZ coordinate values. An orientation discontinuity is measured by exploiting the fact that two meeting surfaces with different orientations produce two clusters in the histogram distribution of the 3D orientation of the points. An irregular discontinuity is measured by exploiting the fact that the histogram distribution of the 3D orientation of the points should be flat.

Discontinuity types of each pixel for a scene is shown in figure 2(f) where the local 3D structure type of each point is shown in different colors.

Figure 3: A subset of the 20 3D data sets used in the analysis. The points without corresponding range data are marked in blue. The gray image shows the range data of the top-left scene. The resolution range of the whole data set is [512-2048]x[390-2290] with an average resolution of 1140x1001.

3 METHODS

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; due to space limitations, only two of them are shown in figure 3. The range of an object which does not reflect the laser beam back to the scanner or which is out of the range of the scanner cannot be measured. These points are marked with blue in figure 3 and are not processed in our analysis. The resolution range of the data set is [512-2048]x[390-2290] with an average resolution of 1140x1001.

3.1 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. For homogeneous and edge-like structures, different representations are needed due to different underlying structures (in the rest of the paper, a homogeneous image structure that corresponds to a 3D continuity will be called a *mono*.). 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:

$$\pi^M = (\mathbf{X}_{3D}, \mathbf{X}_{2D}, \mathbf{c}, \mathbf{p}), \quad (1)$$

$$\pi^E = (\mathbf{X}_{3D}, \mathbf{X}_{2D}, \phi_{2D}, \mathbf{c}_1, \mathbf{c}_2, \mathbf{p}_1, \mathbf{p}_2), \quad (2)$$

where $\mathbf{X}_{3D}$ and $\mathbf{X}_{2D}$ denote 3D and 2D positions of the 3D entity; ϕ_{2D} is 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 4). Note that π^M does not have any 2D orientation information (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 4.

In our analysis, the entities are regularly sampled from the 2D information. The sampling size is 10 pixels. See (Krüger et al., 2003; 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 4). 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 straightforward. For these structures, our approach was to fit unit-planes[1] 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 two-fold planar representation of the 3D entity.

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

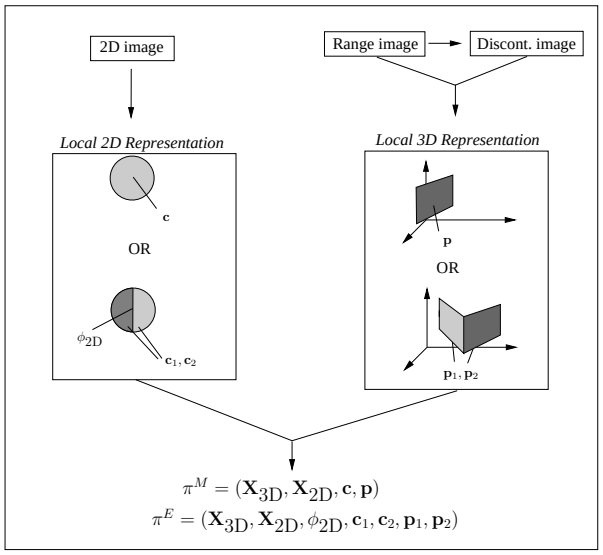


Figure 4: Illustration of the representation of a 3D entity. From the 2D and 3D information, local 2D and 3D representation is extracted.

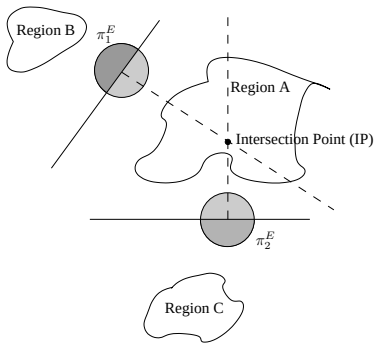


Figure 5: The parameters involved in second order 3D statistics.

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 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.

3.2 Collecting the Data Set

In our analysis, we form pairs out of π^Es that are close enough, and for each pair, we check whether monos in the scene are coplanar to the elements of the pair

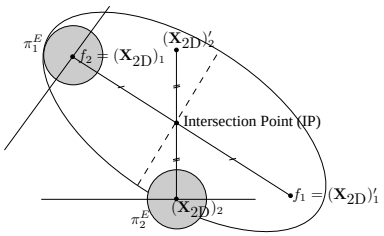


Figure 6: The ellipse in second order 3D statistics.

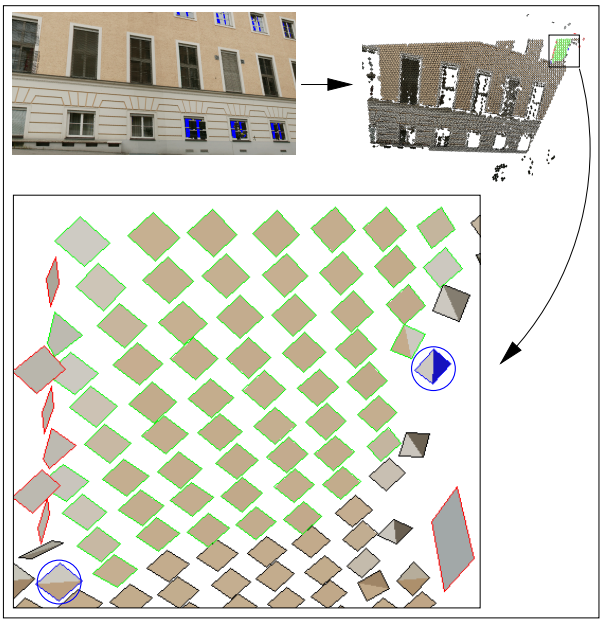


Figure 7: Illustration of a pair of π^E and the set of monos associated to them. Top-left shows the 2D image. Top-right shows the 3D representation of the scene in our 3D visualization software. At the bottom, a part of the 3D representation is displayed in detail where the edges are shown in blue; the monos coplanar with the edges are shown in green, and non-coplanar monos are shown in red. The entities are drawn in rectangles because of the high computational complexity of drawing circles.

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 5. 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 (*e.g.*, see figure 1(a)).

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

$$\mathcal{P} = \left\{(\pi_1^E, \pi_2^E) \mid \forall \pi_1^E, \pi_2^E,\ \pi_1^E \in \Omega(\pi_2^E),\ I(\perp(\pi_1^E), \perp(\pi_2^E))\right\}, \tag{3}$$

where $\Omega(\pi^E)$ is the N-pixel-2D-neighborhood of π^E; $\perp(\pi^E)$ 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.

Next, we have to determine which monos in region A should be analyzed for the relation; that is, what is the shape of region A? Empirically, it turns out that an ellipse (1) is the computationally cheapest shape and (2) fits to different configurations of π_1 and π_2 under different orientations and distances. Neither a rectangle nor a circle satisfy these two properties. Figure 6 demonstrates the ellipse for the example pair of edges in figure 5. The center of the ellipse is at the intersection of the normals of the edges which we call as the intersection point (IP) in the rest of the paper.

For each pair of edges in $\mathcal{P}$, we decide on which region to analyze the relation of depth by intersecting the normals of the edges. Then, we associate the monos inside the ellipse that are defined by the pair of edges.

Note that a π^E has two planes that represent the underlying 3D structure. When π^Es become associated to monos, only one plane that faces the ellipse becomes relevant. Let π^{sE} denote the semi-representation of π^E which can be defined as:

$$\pi^{sE} = (\mathbf{X}_{3D}, \mathbf{X}_{2D}, \mathbf{c}, \mathbf{p}). \tag{4}$$

Note that π^{sE} is equivalent to the definition of π^M in equation 2.

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

$$\mathcal{T} = \{(\pi_1^{sE}, \pi_2^{sE}, \pi^M) \mid (\pi_1^E, \pi_2^E) \in \mathcal{P}, \pi^M \in \mathcal{S}^M, \pi^M \in E(\pi_1^E, \pi_2^E)\}, \tag{5}$$

where $\mathcal{S}^M$ is the set of all π^M, and $E(\pi_1^E, \pi_2^E)$ represents the ellipse associated to π_1^E and π_2^E[2].

A pair of π^Es and the set of monos associated to them are illustrated in figure 7. The edges are shown in blue, and the coplanar and non-coplanar monos are shown in green and red, respectively.

[2]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 minor axis.

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:

$$\begin{aligned} l_1 &= ((\mathbf{X}_{2D})_1 - IP), \\ l_2 &= ((\mathbf{X}_{2D})_2 - IP). \end{aligned} \tag{6}$$

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

$$E(\pi_1^E, \pi_2^E) = \begin{cases} f_1 = (\mathbf{X}_{2D})_1, f_2 = (\mathbf{X}_{2D})'_1, b = |l_2| & \text{if } |l_1| > |l_2|, \\ f_1 = (\mathbf{X}_{2D})_2, f_2 = (\mathbf{X}_{2D})'_2, b = |l_1| & \text{otherwise.} \end{cases} \tag{7}$$

where $(\mathbf{X}_{2D})'$ is the symmetry of $\mathbf{X}_{2D}$ around the intersection point and on the line defined by $\mathbf{X}_{2D}$ and IP (as shown in figure 6).

3.3 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 applied:

$$cop(\pi_1^s,\pi_2^s) = \alpha(\mathbf{p}^{\pi_1^s},\mathbf{p}^{\pi_2^s}) < T_p \text{ AND } d(\mathbf{p}^{\pi_1^s},\pi_2^s)/d(\pi_1^s,\pi_2^s) < T_d, \quad (8)$$

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 and 0.5, respectively.

4 RESULTS AND DISCUSSIONS

The data set 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 8 shows the results of our analysis. In figure 8(a), 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 $P(cop(\pi^M,\pi_1^E \wedge \pi_2^E) \mid d_N(\pi^M,\pi^E))$ where $cop(\pi^M,\pi_1^E \wedge \pi_2^E)$ is defined as $cop(\pi_1^E,\pi_2^E) \wedge cop(\pi^M,\pi^E)$, and π^E is either π_1^E or π_2^E. The normalized distance measure[3] $d_N(\pi^M,\pi^E)$ is defined as:

$$d_N(\pi^M,\pi^E) = \frac{d(\pi^M,\pi^E)}{2\sqrt{d(\pi_1^E,IP)^2 + d(\pi_2^E,IP)^2}}, \quad (9)$$

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 8(a) that the likelihood decreases when a mono is more distant from an edge. However, when the distance measure gets closer to 1, the likelihood increases. This is because when the mono gets away from either π_1^E or π_2^E, it becomes closer to the other π^E.

In figure 8(b), we see the unconstrained case of figure 8(a); *i.e.,* the case where there is no information about the coplanarity of π_1^E and π_2^E, namely, $P(cop(\pi^M,\pi^E) \mid d_N(\pi^M,\pi^E))$ where π^E is either π_1^E or π_2^E. We see that the likelihood distribution is weaker than the case where π_1^E and π_2^E are coplanar. The comparison with figure 8(a) shows that the existence of another edge in the neighborhood increases the likelihood of finding coplanar structures.

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

In figure 8(c), the likelihood of the coplanarity of a mono against the distance to IP (*i.e.,* $P(cop(\pi^M,\pi_1^E \wedge \pi_2^E) \mid d_{NU}(\pi^M,IP), d_{NV}(\pi^M,IP))$) is shown. We see in the figure that the likelihood shows a flat distribution against the distance to IP.

In figure 8(d), the likelihood of the coplanarity of a mono against the distance to π_1^E *and* π_2^E (*i.e.,* $P(cop(\pi^M,\pi_1^E \wedge \pi_2^E) \mid d_N(\pi^M,\pi_1^E), d_N(\pi^M,\pi_2^E))$) 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 when π^M moves away from an equidistant point, it becomes closer to the other edge and in that case, as shown in figure 8(a), the likelihood increases.

The results, especially figure 8(a) and (b) confirm the importance of the relation illustrated in figure 1(a).

In figure 9, first results of an unpublished ongoing work on a depth prediction model based on the presented statistical framework are presented. 9(c) shows the results of feature-based stereo while in 9(d), depth predictions are shown in our 3D display software

5 CONCLUSION

In this paper, using 3D range data with real-world color information, we have analyzed whether the depth of a mono is predictable from the depth of the edges that bound the homogeneous image patch. We have analyzed the predictability of the depth of a mono given a single edge and a pair of coplanar edges.

We have shown that a mono is more likely to be coplanar with an edge when it is closer to the edge and when there is another coplanar edge in the neighborhood. We have shown that the existence of a coplanar edge in the neighborhood is a strong event and to our knowledge, is not recognized by the literature.

The results suggest that the depth estimation at homogeneous image structures can be achieved indirectly from the available information at the edges. We believe that this is a new approach to 3D reconstruction.

In this paper, we are only interested in second-order long-range relations between local features. For round objects like shown in figure 1(b,c), the depth information is given by the shading whose statistical properties can only be captured by different relations.

In our current work, we are developing a model that exploits the statistics presented in this paper to predict the depth of homogeneous image patches from the depth of edges. First results of this ongoing work are also presented in the paper.

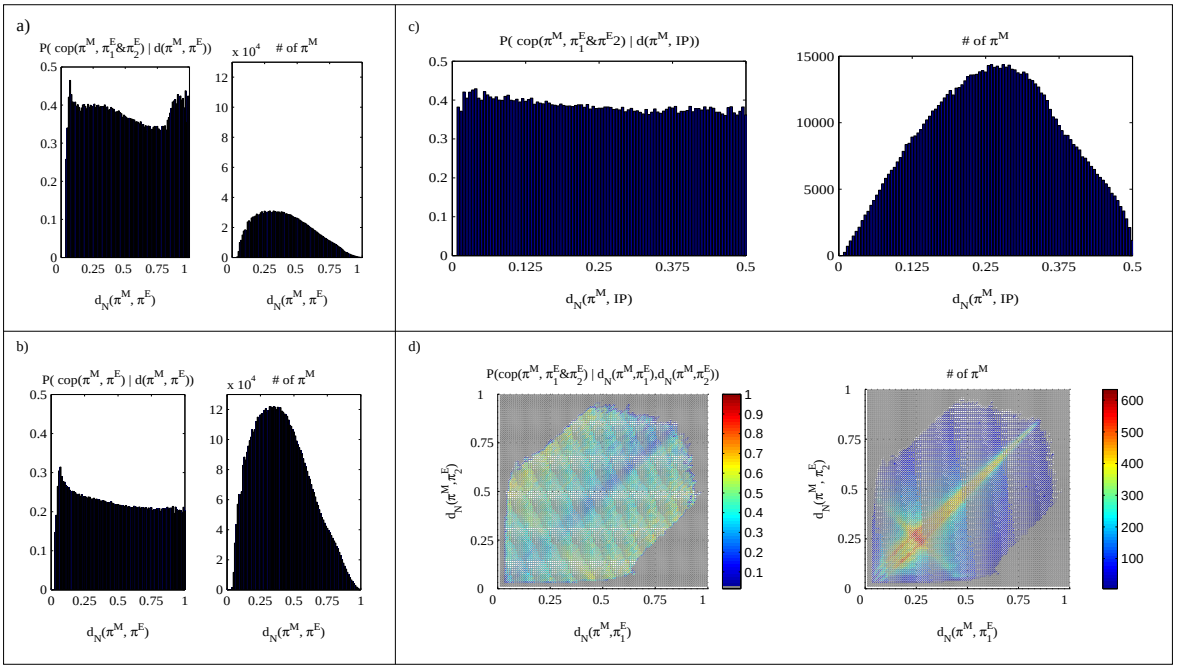


Figure 8: Likelihood distribution of coplanarity of monos. In each sub-figure, left-plot shows the likelihood distribution whereas right-plot shows the frequency distribution. (a) The likelihood of the coplanarity of a mono with π_1^E *and* π_2^E against the distance to π_1^E *or* π_2^E.(b) The likelihood 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.(c) The likelihood of the coplanarity of a mono against the distance to *IP*. (d) The likelihood of the coplanarity of a mono against the distance to π_1^E *and* π_2^E.

6 ACKNOWLEDGEMENTS

We would like to thank RIEGL UK Ltd. for providing us with 3D range data. This work is supported by the ECOVISION project.

REFERENCES

Bolle, R. M. and Vemuri, B. C. (1991). On three-dimensional surface reconstruction methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(1):1–13.

Bruce, V., Green, P. R., and Georgeson, M. A. (2003). *Visual Perception: Physiology, Psychology and Ecology*. Psychology Press, 4th edition.

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

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

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

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

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

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

Howe, C. Q. and 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, A. B., and Mumford, D. (2000). Statistics of range images. *CVPR*, 1(1):1324–1331.

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

Knill, D. C. and Richards, W., editors (1996). *Perception as bayesian inference*. Cambridge: Cambridge University Press.

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

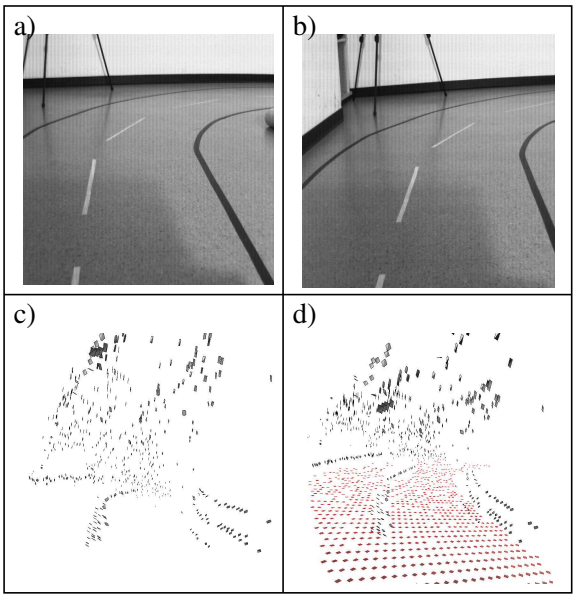


Figure 9: Initial depth prediction results on a toy example. (a) Left image. (b) Right image. (c) Feature-based stereo result giving depth only at edges. (d) Depth-prediction based on (c).

Krüger, N., Lappe, M., and 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*, pages 53–59.

Krüger, N. and Wörgötter, F. (2004). Statistical and deterministic regularities: Utilisation of motion and grouping in biological and artificial visual systems. *Advances in Imaging and Electron Physics*, 131:82–147.

Krüger, N. and 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*, pages 157–166.

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

Potetz, B. and Lee, T. S. (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., and Wörgötter, F. (2004). A non-local stereo similarity based on collinear groups. *Proceedings of the Fourth International ICSC Symposium on Engineering of Intelligent Systems*.

Purves, D. and Lotto, B., editors (2002). *Why we see what we do: an empirical theory of vision*. Sunderland, MA: Sinauer Associates.

Rao, R. P. N., Olshausen, B. A., and Lewicki, M. S., editors (2002). *Probabilistic models of the brain*. MA: MIT Press.

Shirai, Y. (1987). *Three-dimensional computer vision*. Springer-Verlag New York, Inc.

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

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