HAL
open science

An Occupancy-Depth Generative Model of Multi-view Images

Pau Gargallo, Peter Sturm, Sergi Pujades

To cite this version:

Pau Gargallo, Peter Sturm, Sergi Pujades. An Occupancy-Depth Generative Model of Multi-view Images. ACCV 2007 - 8th Asian Conference on Computer Vision, Nov 2007, Tokyo, Japan. pp.373-383, 10.1007/978-3-540-76390-1_37 . inria-00384276

HAL Id: inria-00384276

https://inria.hal.science/inria-00384276v1

Submitted on 14 May 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

An Occupancy–Depth Generative Model of Multi-view Images

Pau Gargallo, Peter Sturm, and Sergi Pujades

INRIA Rhône-Alpes and Laboratoire Jean Kuntzmann, France
name.surname@inrialpes.fr

Abstract. This paper presents an occupancy based generative model of stereo and multi-view stereo images. In this model, the space is divided into empty and occupied regions. The depth of a pixel is naturally determined from the occupancy as the depth of the first occupied point in its viewing ray. The color of a pixel corresponds to the color of this 3D point.

This model has two theoretical advantages. First, unlike other occupancy based models, it explicitly models the deterministic relationship between occupancy and depth and, thus, it correctly handles occlusions. Second, unlike depth based approaches, determining depth from the occupancy automatically ensures the coherence of the resulting depth maps.

Experimental results computing the MAP of the model using message passing techniques are presented to show the applicability of the model.

1 Introduction

Extracting 3D information from multiple images is one of the central problems of computer vision. It has applications to photorealistic 3D reconstruction, image based rendering, tracking and robotics among others. Many successful methods exist for each application, but a satisfactory general formalism is still to be defined. In this paper we present a simple probabilistic generative model of multi-view images that accurately defines the natural relationship between the shape and color of the 3D world and the observed images.

The model is constructed with the philosophy that if one is able to describe the image formation process with a model, then Bayesian inference can be used to invert the process and recover information about the model from the images. The approach yields to a generic widely applicable formalism. The price to pay for such a generality is that the results for each specific application will be poor compared to those of more specialized techniques.

There are mainly two approaches to the stereo and multi-view stereo problems. In small-baseline situations, the 3D world is represented by a depth map on a reference image and computing depth is regarded as a matching problem [1]. In wide-baseline situations, it is often more convenient to represent the shape of the objects by a surface or an occupancy function [2]. Depth and occupancy are

Y. Yagi et al. (Eds.): ACCV 2007, Part II, LNCS 4844, pp. 373–383, 2007.
© Springer-Verlag Berlin Heidelberg 2007

obviously highly related, but most of the algorithms concentrate on finding one of the two.

The main problem with either approach are occlusions. The fact that a 3D point is not always visible from all the cameras makes the extraction of 3D information from images hard. The two main approaches to solve this issue are to treat occluded points as outliers [3] or to explicitly model the geometrical reason for the occlusion.

Making a accurate generative model for multi-view images, as we wish to do, necessarily involves modeling occlusions geometrically, because geometric occlusions really exist in the true image formation process. In fact, geometric occlusions are so important that there are reconstruction techniques, like those based on the visual hull, that use only the occlusion information [4,5].

Geometric occlusions can be modeled effectively in depth based approaches by computing a depth map for every input image [6,7,8]. This requires to add constraints, so that the multiple depth maps are coherent and form a single surface. These constrains are not necessary in shape based approaches that implicitly incorporate them because they compute a single model for all the images.

Shape based approaches usually deal with geometric occlusions in an alternating way [9,10,11]. They first compute the visibility given the current estimate of the shape; and then modify the shape according to some criteria. This procedure disregards the fact that the visibility will change while modifying the shape. A voxel carving technique carving inside the object, or a shrinking surface evolution are consequences of this oversight. Our recent work [12] avoids these problems by explicitly taking into account the visibility changes during the surface evolution. This is nothing but the continuous version of the model presented in this paper.

The model presented in this paper explicitly characterizes the relationship between depth and shape and profits of the benefits of both worlds. The shape occupancy automatically gives coherence to the depth maps. Properly deriving the depth maps from the occupancy implicitly encodes the geometric occlusions.

Our model is very related both in objective and approach to the recent work of Hernández *et al.* [13]. In that work depth cues, are probabilistically integrated for inferring occupancy. The principal difference between their model and the one presented here is that they make some independence assumptions that we do not. In particular, they assume the depth of different pixels to be independent. This greatly simplifies the inference and good results are achieved. In our model, as in the real world, depth is determined by the occupancy and therefore the depth of different pixels are not independent. This create a huge very loopy factor graph representation of the join probability of occupancy and depth. Inference in such a graph is hard as we will see in the experiments.

2 The Model

This section presents the occupancy-depth model. We first introduce the random variables involved in the generative process. Then we decompose their joint probability distribution into simpler terms and give a form to each of them.

2.1 Occupancy, Depth and Color Variables

Consider a discretization of the 3D space in a finite set of sites $\mathcal{S} \subset \mathbb{R}^3$. A given site $\mathrm{x} \in \mathcal{S}$, can be in the free space or inside an object. This defines the occupancy of the site that will be represented by a binary random variable u_{x} (1 meaning occupied and 0 meaning free). The occupancy of the whole space will be represented by the random process $u : \mathcal{S} \to \{0, 1\}$, which defines the shape of the objects in the space.

The shape of the objects is not enough to generate images. Their appearance is also needed. In the simplest case, under the constant brightness assumption, the appearance can be represented by a color at each point on the surface of the objects. As we do not know the shape of the objects right now, we will need to define the color of all sites in $\mathcal{S}$, even if only the color of the sites lying on the surface is relevant. The color will be represented by a random process $\mathcal{C} : \mathcal{S} \to \mathbb{R}^3$.

The depth of a pixel is defined as the distance (measured along the camera's z-axis) between the optical center of the camera and the 3D point observed at the pixel. Given the occupancy of the space and the position and calibration of a camera i, the depth $\mathcal{D}^i_{\mathrm{p}}$ of a pixel p is determined as the depth of the first occupied point in its viewing ray.

The *observed* color at that pixel will be denoted by $\mathcal{I}^i_{\mathrm{p}}$. This color should ideally correspond to the color of the site observed at that pixel. i.e. the point of the viewing ray of p which is at depth $\mathcal{D}^i_{\mathrm{p}}$.

2.2 Decomposition

Having defined all the variables depths, colors, occupancies and observed colors – we will now define their joint probability distribution. To do so, we first decompose the distribution terms representing the natural dependence between the variables. One can think of this step as defining the way the data (the images) were generated.

The proposed decomposition is

$$p(u, \mathcal{C}, \mathcal{D}, \mathcal{I}) = p(u)p(\mathcal{C}|u) \prod_{i,\mathrm{p}} p(\mathcal{D}^i_{\mathrm{p}}|u) \prod_{i,\mathrm{p}} p(\mathcal{I}^i_{\mathrm{p}}|\mathcal{D}^i_{\mathrm{p}}, \mathcal{C}) \ . \tag{1}$$

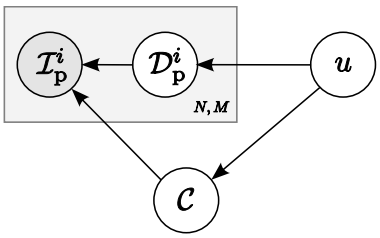

Fig. 1. Bayesian network representation of the joint probability decomposition

It is represented graphically in Figure 1. Each term of the decomposition corresponds to a variable and, thus, to a node of the network. The arrows represent the statistical dependencies between the variables. In other words, the order that one has to follow to generate random samples of the model from scratch.

Therefore, the data generation process is as follows. First one builds the objects of the world by generating an occupancy function. Then one paints them by choosing the space colors. Finally, one takes pictures of the generated world: one first determines which points are visible from the camera by computing the depth of the pixels, and then sets the color of the pixels to be the color of the observed 3D points.

In the following sections, we define each of the terms of the decomposition (1).

2.3 World Priors

Not all the possible occupancy functions are equally likely a priori. One expects the occupied points to be gathered together forming objects, rather than randomly scattered over the 3D space. To represent such a belief, we choose the occupancy u to follow a Markov Random Field distribution. This gives the following prior,

$$p(u) \propto \exp\{-\sum_{\mathrm{x,y}} \psi(u_{\mathrm{x}}, u_{\mathrm{y}})\} \tag{2}$$

where the sum extends to all the neighboring points (x, y) in a grid discretization $\mathcal{S}$ of the 3D space. The energy potentials are of the form $\psi(u_{\mathrm{x}}, u_{\mathrm{y}}) = \alpha|u_{\mathrm{x}} - u_{\mathrm{y}}|$, so that they penalize neighboring sites of different occupancies by a cost α.

This prior is isotropic in the sense that two neighboring points are equally likely to have the same occupancy regardless of their position and color. From experience, we know that the discontinuities or edges in images often correspond to discontinuities in the occupancy (the occluding contours). Therefore, one could be tempted to use the input images to derive a smoothing prior for the occupancy that is weaker at the points projecting to image discontinuities. While effective, this would not be correct from a Bayesian point of view, as one would be using the data to derive a prior for the model. We will now see how to obtain this anisotropic smoothing effect in a more theoretically well funded way.

In the proposed decomposition, the prior on the color of the space depends on the occupancy. This makes it possible to express the following idea. Two neighboring points that are both either occupied or free, are likely to have similar colors. The colors of two points with different occupancies are not necessarily related. This can be expressed by the MRF distribution

$$p(\mathcal{C}|u) \propto \exp\{-\sum_{\mathrm{x,y}} \phi(\mathcal{C}_{\mathrm{x}}, \mathcal{C}_{\mathrm{y}}, u_{\mathrm{x}}, u_{\mathrm{y}})\} \tag{3}$$

with

$$\phi(\mathcal{C}_{\mathrm{x}}, \mathcal{C}_{\mathrm{y}}, u_{\mathrm{x}}, u_{\mathrm{y}}) = \begin{cases} \varrho(\mathcal{C}_{\mathrm{x}} - \mathcal{C}_{\mathrm{y}}) & \text{if } \ u_{\mathrm{x}} = u_{\mathrm{y}} \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

where ϱ is some robust penalty function, that penalize the difference of colors of neighboring points with the same occupancy.

Now, combining the prior on the occupancy with the prior on the color we have

$$p(u, \mathcal{C}) \propto \exp\{-\sum_{\mathrm{x,y}} \bar{\psi}(\mathcal{C}_{\mathrm{x}}, \mathcal{C}_{\mathrm{y}}, u_{\mathrm{x}}, u_{\mathrm{y}})\} \tag{5}$$

with

$$\bar{\psi}(\mathcal{C}_{\mathrm{x}}, \mathcal{C}_{\mathrm{y}}, u_{\mathrm{x}}, u_{\mathrm{y}}) = \begin{cases} \varrho(\mathcal{C}_{\mathrm{x}} - \mathcal{C}_{\mathrm{y}}) & \text{if } \ u_{\mathrm{x}} = u_{\mathrm{y}} \\ \alpha & \text{otherwise} \end{cases} \tag{6}$$

If we are given the color of the space, then $p(u|C) \propto p(u, C)$ is a color driven smoothing prior on the occupancy. Neighboring points with the same color are more likely to have the same occupancy than neighboring points with different colors. As the color will be estimated from the images, the color discontinuities will coincide with the edges in the images. Thus, this term will represent our experience based knowledge that object borders coincide with image edges.

2.4 Pixel Likelihood

The color $\mathcal{I}^i_{\mathrm{p}}$ observed at a pixel should be equal to the color of the 3D point visible at that pixel, up to the sensor noise and other unmodeled effects, e.g. specularities. If we denote the color of the observed 3D point as $\mathcal{C}(\mathcal{D}^i_p)$, we have

$$p(\mathcal{I}^i_{\mathrm{p}}|\mathcal{D}^i_{\mathrm{p}}, \mathcal{C}) \propto \exp\{\ -\rho\big(\mathcal{I}^i_{\mathrm{p}} - \mathcal{C}(\mathcal{D}^i_p)\big)\ \} \tag{7}$$

where ρ is some robust penalty function.

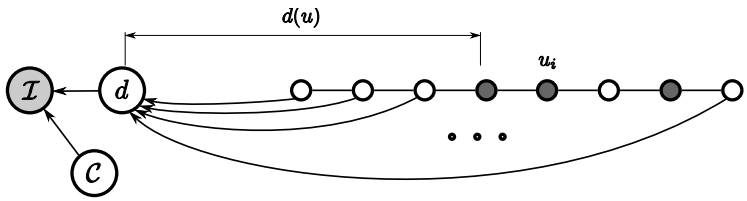

Fig. 2. Bayesian network of the model associated to a single viewing ray. The occupancy variables u_i along the viewing ray determine de depth d. The color of the image $\mathcal{I}$ corresponds to the color $\mathcal{C}$ at depth d.

Note that unlike traditional stereo algorithms, here there are no occlusions to be taken into account by the function ρ. We are matching the color of a pixel with the color of the observed scene point, not with the color of pixels in the other images. The observed scene point is, by definition, non-occluded, so no occlusion problem appears here.

Depth Marginalization. The likelihood (7) of a pixel depends only on its depth and not on the whole occupancy function. However, the relationship between occupancy and depths is simple and deterministic. Therefore, it is easy to marginalize out the depth and express the likelihood directly in terms of the occupancy of the points on the viewing ray of the pixel.

To simplify the notation we will do the computations for a single pixel. Figure2 shows the Bayesian network associated to a single pixel. The points of its viewing ray will be denoted by the natural numbers $\{0, 1, 2, \cdots\}$, ordered by increasing distance to the camera. Their occupancy is a vector u such that u_i is the occupancy of the i-th point in the viewing ray. The depth will be denoted by d.

With this language, the probability of a depth given the occupancy is

$$p(d|u) = \prod_{i<d} (1 - u_i) u_d \tag{8}$$

This equation encodes the fact that if d is the depth of a pixel, then occupancy of the points with lower depths must be zero and the occupancy of the point at depth d must be one. It is easy to check that $p(d|u)$ is 1 only when d is the actual depth determined by u and 0 otherwise.

Now, if we note the likelihood of the pixel having a depth d by $L(d) = p(\mathcal{I}^i_{\mathrm{p}}|\mathcal{D}^i_{\mathrm{p}} = d, \mathcal{C})$ and the likelihood of a pixel given the occupancy u by $L(u)$, then

$$L(u) = p(\mathcal{I}|u, \mathcal{C}) = \sum_d p(\mathcal{I}|d, \mathcal{C}) p(d|u) = \sum_d L(d) \prod_{i<d} (1 - u_i) u_d \ . \tag{9}$$

Note that the summand is null for all depths d except for the one that corresponds to the occupancy u.

3 Inference

The last section presented the generative occupancy–depth model by defining the joint probability of occupancy, color, depth and image pixels. In this section we will present an algorithm for inverting the process and recover occupancy from multiple images.

Given a set of observed images $\{\mathcal{I}^i\}$ the goal is to find the posterior probability of occupancy and color, $p(u, \mathcal{C}|\mathcal{I})$. In a tracking application, for example, one may be interested in computing the occupancy *marginals* at each point in the 3D space. This can be used as input for a volumetric 3D tracker. Alternatively, in a novel view synthesis application one may be more interested in finding the *most probable* world in order to render it from other points of view.

As we will see these are difficult tasks, challenging the most recent inference algorithms. The main problem is the interdependence of the occupancies of the points in a viewing ray, which creates high order cliques, in addition to the extrem loopiness of the network. We present here a first try of solving the inference problem by using EM and message passing.

The optimization is done by alternating between the optimization of the occupancy and the color. In the E-step, the probabilities of the occupancies are computed using message passing. Depending on the goal, the sum-product or the max-product algorithm is used. In the M-step, the color estimation is improved by maximizing its expected log-posterior.

It is interesting to note that simpler optimization techniques like iterative conditional modes will lead to algorithms strongly similar to voxel carving or surface evolution. In this case, one will start with an initial guess of the occupancy and will try to change the occupancy of the voxels at the border of the objects in order to improve the posterior. The message passing algorithm presented below is smarter in the sense that it will make decisions about the occupancy of a whole viewing ray at a time.

3.1 Factor Graph Representation

In the E-step, for a fixed coloring of the space, $\mathcal{C}$, the posterior probability $p(u|\mathcal{I},\mathcal{C})$ must be computed (or maximized). This distribution can be represented as a factor graph. Again, for notational simplicity, we detail here the factor graph corresponding to a single pixel,

$$p(u|\mathcal{I},\mathcal{C}) \propto L(u) \prod_{ij} \exp(-\bar{\psi}(u_i, u_j)) \ . \tag{10}$$

Figure 3 shows the factor graph for a single pixel and sketches the one corresponding to a pair of pixels in different images.

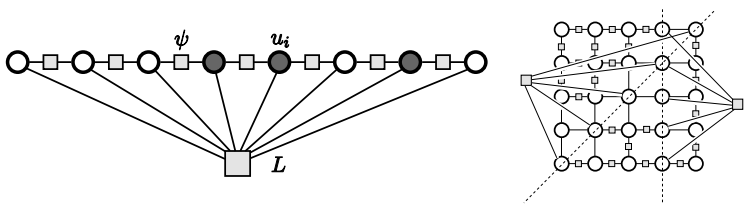

Fig. 3. Factor graph representation of the distribution for a single pixel (left) and for two pixel with intersecting viewing rays (right)

The graph contains two types of factors. The likelihoods of the pixels L are huge factors connecting all the points in the viewing ray of each pixel. The smoothing potentials $\exp(-\bar{\psi})$ are standard pairwise factors. Notice the extreme loopiness of the complete graph where the viewing rays of different pixels intersect to each other.

3.2 Message Passing

Inference will be done by message passing in the factor graph. Messages will be sent from the likelihood and smoothing factors to the occupancy variables and

vice-versa. One can visualize this process as a dialog between the images and the occupancy. Each image tells the occupancy what to be. The occupancy gathers all the messages and replies to each image with a summary of the messages sent by the other images. The process continues until a global agreement is found.

We derive here the reweighted message passing equations [14,15] for the occupancy factor graph. The equations are presented for the sum-product algorithm, but the max-product equivalents can be obtained easily by replacing all the sums in the equations by maximums.

Using the notation of Minka [14], the posterior probability of the occupancy $p(u|\mathcal{I},\mathcal{C})$ will be approximated by a fully factorized variational distribution $q(u) = \prod_i q_i(u_i)$. The belief q_i of an occupancy variable u_i is the product of the messages that it receives. That is

$$q_i(u_i) = \prod_{a\in\mathcal{N}(i)} m_{a\to i}(u_i)\,, \tag{11}$$

where $\mathcal{N}(i)$ are the indices of the factors connected to u_i. The message that a factor f_a sends to a variable is given by

$$m_{a\to i}(u_i)^{\alpha_a} = \sum_{u_a\setminus u_i} f_a(u_a)^{\alpha_a} \prod_{j\in\mathcal{N}(a)\setminus i} n_{j\to a}(u_j)\ , \tag{12}$$

where u_a is the set of variables connected to the factor f_a and $\mathcal{N}(a)$ their indices. α_a is the weight of the factor and it is a parameter of the message passing algorithm. Finally the replying message from a variable to a factor is

$$n_{i\to a}(u_i) = q_i(u_i) m_{a\to i}(u_i)^{-\alpha_a}\ . \tag{13}$$

The pair-wise smoothing potentials are simple and a direct implementation of these formulas is straightforward. The likelihood factors, however, need more attention. These factors link a lot of variables. A direct implementation of the equations above will involve computing sums over all the possible configurations of occupancy along each viewing ray. The number of configurations grows exponentially with the number of sites and becomes quickly intractable.

Luckily, the likelihood factor L is simple enough that the sum can be simplified analytically. The message that a pixel likelihood factor sends to the occupancy of the grid point i is

$$m_{L\to u_i}(x)^{\alpha_L} = \sum_{u\setminus u_i} L(u)^{\alpha_L} \prod_{j\neq i} n_{j\to L}(u_j)\ . \tag{14}$$

Substituting $L(u)$ from equation (9) we get

$$\sum_d \sum_{u\setminus u_i} L(d)^{\alpha_L} \prod_{j<d}(1-u_j) u_d \prod_{j\neq i} n_{j\to L}(u_j)\ . \tag{15}$$

And now we can split and simplify the sum over d into 3 sums with d smaller, equal and bigger than i respectivelly,

$$\begin{aligned} & \sum_{d<i} L(d)^{\alpha_L} \prod_{j<d} n_{j\to L}(0)\, n_{d\to L}(1) \prod_{j>d \wedge j\neq i} \Big(\sum_{u_j} n_{j\to L}(u_j)\Big) \\ + \quad & u_i\, L(i)^{\alpha_L} \prod_{j<i} n_{j\to L}(0) \prod_{j>d} \Big(\sum_{u_j} n_{j\to L}(u_j)\Big) \\ + \quad & (1-u_i) \sum_{d>i} L(d)^{\alpha_L} \prod_{j<d \wedge j\neq i} n_{j\to L}(0)\, n_{d\to L}(1) \prod_{j>d} \Big(\sum_{u_j} n_{j\to L}(u_j)\Big). \end{aligned} \tag{16}$$

Finally, defining

$$\tau(d) = L(d)^{\alpha_L} \prod_{j<d} n_{u_j\to L}(0) n_{u_d\to L}(1) \prod_{j>d} \Big(\sum_{u_j} n_{u_j\to L}(u_j)\Big), \tag{17}$$

we have

$$m_{L\to u_i}(u_i)^{\alpha_L} = \frac{1}{\sum_{u_i} n_{u_i\to L}(u_i)} \sum_{d<i} \tau(d) + \frac{u_i}{n_{u_i\to L}(1)} \tau(i) + \frac{1-u_i}{n_{u_i\to L}(0)} \sum_{d>i} \tau(d), \tag{18}$$

which can be computed in linear time.

3.3 Color Estimation

In the M-step the color of the space is computed by maximizing the expected log-posterior,

$$\langle \ln p(u, \mathcal{C}|\mathcal{I}) \rangle_q = \langle \ln p(\mathcal{I}|u, \mathcal{C}) \rangle_q + \langle \ln p(u, \mathcal{C}) \rangle_q + const. \tag{19}$$

where $\langle \cdot \rangle_q$ denotes the expectation with respect to u assuming that it follows the variational distribution q. Again, the expectation is a sum over all possible occupancy configurations. Simplifications similar to the ones done above yield to

$$\langle \ln p(\mathcal{I}|u, \mathcal{C}) \rangle_q = -\sum_d \prod_{i<d} q_i(0)\, q_d(1)\, \rho\big(\mathcal{I} - \mathcal{C}(d)\big) \tag{20}$$

and

$$\begin{aligned} \langle \ln p(u, \mathcal{C}) \rangle_q = -\sum_{ij} \Big[& \big(q_i(0)\, q_j(0) + q_i(1)\, q_j(1)\big)\, \varrho(\mathcal{C}_i - \mathcal{C}_j) \\ & + \big(q_i(0)\, q_j(1) + q_i(1)\, q_j(0)\big)\, \alpha \Big]\,. \end{aligned} \tag{21}$$

The optimization is done by a gradient descent procedure.

The initialization of the whole EM procedure is as follows. The messages are all initialized to uniform distributions. The color is initialized at each site by computing the mean of the color of its projections to the input images.

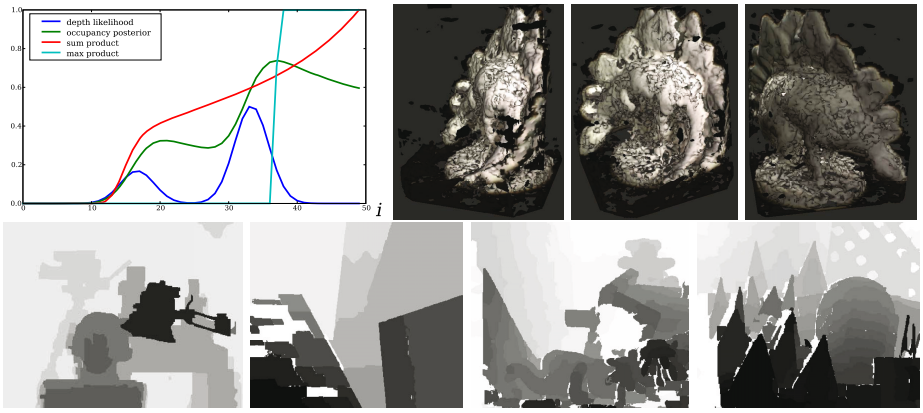

Fig. 4. Results for the single pixel (top-left), the multi-view (top-right) and stereo pair (bottom) experiments

4 Experimental Validation

In order to validate the generative model and to test the possibility of inverting the process using the inference method described above, we performed three types of experiments.

The first experiment was done to test the performance of the reweighted message passing algorithm. A single pixel camera model was used and different likelihood function $L(d)$ were manually chosen. The ground truth posterior of the occupancy was computed by sampling. First image of Figure 4 shows an example result. In general, both the sum-product and the max-product performed poorly, only roughtly approximating the accual marginals and MAPs.

Although this poor performance, we tested the algorithm with real images with the hope that the interconnections between different viewing rays will constrain the problem and possibly get better results. Last row of Figure 4 shows the results obtained for the four Middlebury stereo pairs [1]. The displayed depthmaps were computed from the inferred occupancy of frontoparallel layers placed in front of the cameras.

The results are generally correct, but large errors are present, specially in the textureless regions. Holes appear in these regions. A possible explanation for these holes is that, as we are smoothing the occupancy of the layers (which are mostly empty) the likelihood term in the textureless regions is too weak to prevent the smoothing from erasing any occupancy on the layer.

The last test was done with wide angle multi-view stereo data, to test the exact same algorithm in a very different setup. Figure 4 shows preliminary results of the inference for the dataset *dinoSparseRing* [2]. The shape of the dinosaur is globally recovered, but here also there are numerous holes in the object. We did not perform a numerical evaluation of the results because the scores will be dominated by the errors described above and, therefore, uninteresting.

5 Conclusion

We presented a generative model of multi-view images, developped an inference algoritm to compute occupancy from images and test it experimentally. The results of the experiments are not comparable to the ones obtained by the state-of-the-art techniques, but served well to validate the model. Future work is needed for finding better inference methods and to determine to which degree the errors are due to the inference method or to the generative model itself.

References

1. Scharstein, D., Szeliski, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. Int. J. Comput. Vision 47(1-3), 7–42 (2002)
2. Seitz, S.M., Curless, B., Diebel, J., Scharstein, D., Szeliski, R.: A comparison and evaluation of multi-view stereo reconstruction algorithms. In: Computer Vision and Pattern Recognition 2006, Washington, DC, USA, pp. 519–528 (2006)
3. Strecha, C., Fransens, R., Gool, L.V.: Combined depth and outlier estimation in multi-view stereo. In: Computer Vision and Pattern Recognition 2006, Washington, pp. 2394–2401. IEEE Computer Society Press, Los Alamitos (2006)
4. Laurentini, A.: The visual hull concept for silhouette-based image understanding. IEEE Trans. Pattern Anal. Mach. Intell. 16(2), 150–162 (1994)
5. Solem, J.E., Kahl, F., Heyden, A.: Visibility constrained surface evolution. In: Computer Vision and Pattern Recognition 2005, San Diego, USA, pp. 892–899 (2005)
6. Kang, S.B., Szeliski, R., Chai, J.: Handling occlusions in dense multi-view stereo. In: Computer Vision and Pattern Recognition 2001, Kauai, Hawaii, pp. 103–110 (2001)
7. Kolmogorov, V., Zabih, R.: Multi-camera scene reconstruction via graph cuts. In: European Conference on Computer Vision 2002, London, UK, pp. 82–96 (2002)
8. Gargallo, P., Sturm, P.: Bayesian 3d modeling from images using multiple depth maps. In: Computer Vision and Pattern Recognition 2005, San Diego, vol. 2, pp. 885–891 (2005)
9. Faugeras, O.D., Keriven, R.: Complete dense stereovision using level set methods. In: Burkhardt, H., Neumann, B. (eds.) ECCV 1998. LNCS, vol. 1407, pp. 379–393. Springer, Heidelberg (1998)
10. Kutulakos, K.N., Seitz, S.M.: A theory of shape by space carving. Int. J. Comput. Vision 38(3), 199–218 (2000)
11. Paris, S., Sillion, F.X., Quan, L.: A surface reconstruction method using global graph cut optimization. Int. J. Comput. Vision 66(2), 141–161 (2006)
12. Gargallo, P., Prados, E., Sturm, P.: Minimizing the reprojection error in surface reconstruction from images. In: Proceedings of the International Conference on Computer Vision, Rio de Janeiro, Brazil, IEEE Computer Society Press, Los Alamitos (2007)
13. Hernández, C., Vogiatzis, G., Cipolla, R.: Probabilistic visibility for multi-view stereo. In: Computer Vision and Pattern Recognition 2007, Minneapolis (2007)
14. Minka, T.: Divergence measures and message passing. Technical report, Microsoft Research (2005)
15. Wainwright, M., Jaakkola, T., Willsky, A.: Map estimation via agreement on trees: message-passing and linear programming. Information Theory, IEEE Transactions on 51, 3697–3717 (2005)