Depth map calculation for a variable number of moving objects using Markov sequential object processes

M.N.M. van Lieshout

CWI/Eurandom

PO Box 4079, 1009 AB Amsterdam, The Netherlands

Abstract: We advocate the use of Markov sequential object processes for tracking a variable number of moving objects through a video frame with a view towards depth map calculation. A regression model based on a sequential object process is related to the Hough transform; regularisation terms are incorporated to control within and between frame object interactions. We construct a Markov chain Monte Carlo method for finding the optimal tracks and associated depth maps, and illustrate the approach by a sport sequence.

1 Introduction

1.1 Motivation

Tracking moving objects is the dual task of detecting the objects in a video sequence and following their movements throughout the sequence. Thus, a tracking algorithm must decide whether there are any objects of a specified kind in each of the image frames; if so, it determines the number of objects, their locations, shapes, sizes and spatial relationships within each frame, as well as their movements across frames. In other words, for each object a record has to be kept of the frame in which it is first seen in the observation window, its trajectory over time, and possibly the frame in which the object is last observed.

Object tracking is important, as motion is a prime source of semantic information. Applications include monitoring and surveillance, robotics, and biomedical image analysis. Our motivation comes from the need to transform plentiful 2D content to a format suitable for display on a 3D-TV in light of the dearth of 'true' 3D content. To do so, our goal is to infer a depth map. Indeed, when two objects pass each other, their partial order relation becomes apparent, and can be propagated over frames. It is interesting to note that the occlusion that is often claimed to hinder higher order vision tasks is actually a great help in this context. We use the set of partial depth order relations between object pairs to determine this depth map. This depth map together with the original (2D) video may be viewed on a Philips 3D display. Such a display takes video and the associated depth map and creates in real-time 9 views enabling glasses free 3D-TV [www.philips.com/3DSolutions].

Motion tracking is a complex task, and the classical approach is to tackle easier subproblems [5, 24, 25]. In an initialisation phase, the number of objects to be followed is decided upon, and their positions and velocities are measured. To deal with measurement noise, a set of equations is derived for the movement of an object from one frame to the next. These in turn form the basis for a Kalman or data-association filtering phase that outputs cleaned, more robust object coordinates and relates these to the measurements [3, 11]. Note though

that the Kalman filter can be proved to be optimal only for the prediction of the unobserved state of a linear system under Gaussian noise, a condition that rarely holds for features extracted from video data. For this reason, particle filters [6] were proposed which use a Monte Carlo approximation to the posterior probability distribution. However, the approach suffers from initialisation problems when dealing with a variable number of objects [9, 25], and is not able to capture interactions between the objects [12].

An alternative to Gaussian modelling is to use the Hough transform [8], which translates complex feature recognition problems into easier to handle local peak detection problems. In tracking, the equations for the movement of an object from one frame to the next can be expressed in terms of translation and rotation parameters, evidence for which in turn is accumulated in Hough space [10]. The Hough transform is robust against noise and occlusion; its main disadvantage is the need for storage, although this may be alleviated somewhat by restriction of the parameter space.

1.2 Background and related work

The goal of this paper is to present a coherent theoretical framework for deriving partial depth order relations between a variable number of moving objects and using these relations to calculate a depth map. In particular, we advocate the use of (sequential) spatial object processes, building on successful work on the application of stochastic geometric ideas to computer vision problems.

The idea to use Markov object processes as priors in vision can be traced back to the early 1990s [1, 2, 19, 20]. By their very nature, such models – in combination with a term for assessing the fit of the hypothesized scene with the data image(s) – allow for an unknown, variable number of objects, and may exhibit complicated interactions between objects. Moreover, there is no need for linearity or Gaussianity assumptions, and the posterior distribution quantifies the uncertainty about the validity of the hypothesis. The ideas were taken up and applied to a variety of pattern recognition problems in fields such as confocal microscopy [22] or remote sensing [13, 23], to name but a few. Recently, [12] proposed an object process prior with a view towards the movements of a colony of ants. A sequential, data driven Metropolis–Hastings algorithm was designed, and shown to be effective in dealing with interactions between the ants, but less so in case of occlusion.

Being mostly concerned with objects that do not overlap or have a similar appearance, the above-mentioned papers do not take into account the relative depths of objects in the scene. The work [17] concerned with recognition of piles of mushrooms in a single image frame did, but at high computational cost. Recently, [14, 15] introduced so-called Markov sequential object processes, that seem to be well suited to deal with depth ordering and occlusion because – in contrast to classical Markov object processes – they explicitly model the permutation order of the objects involved. This paper is a first study into the use of such models for depth map estimation by tracking moving objects. Its plan is as follows. First, in Section 2, we fix notation. In Section 3 we propose a regression model based on a sequential object process to assess the likelihood of hypothesised scenes and relate it to the Hough transform. Section 4 introduces further regularisation terms to control within and between frame object interactions. Section 5 is devoted to the construction of a Markov chain

Monte Carlo method for finding the optimal tracks. Section 6 gives an example concerning a sports sequence, and the paper closes with a summary and discussion of future work.

2 Preliminaries and notation

2.1 Setup

The experimental data consist of a sequence of image frames $\mathbf{y} = (\mathbf{y}^i; i = 1, \ldots, I)$, where $\mathbf{y}^i = (y_t^i \ ; \ t \in T)$, $i = 1, \ldots, I$. The *'image space'* T is an arbitrary finite set of pixels, and $I \in \mathbb{N}$ is the number of frames. The observed value y_t^i at pixel $t \in T$ in frame i ranges over a set V that is arbitrary, typically $\{0, 1, \ldots, 255\}^d$ with $d = 1$ for grey level images and $d = 3$ for coloured ones.

Every image frame depicts a scene that contains objects of a certain type that we are interested in. Here, we suppose the *'object space'* $D \times L$ of possible objects to be a Cartesian product of location and object parameters. The set D is a compact subset of $\mathbb{R}^2$ and is used to specify the location of the object; L is an arbitrary complete separable metric space. Typically, the specification of an object includes shape and size parameters, as well as parameters for colour or texture. In the example to be presented later on, the objects will be coloured ellipses, so that $L = [a_{\min}, a_{\max}] \times [b_{\min}, b_{\max}] \times [0, \pi] \times \{0, \ldots, 255\}^3$. The first two components correspond to the ellipse minor and major axis respectively, the third one to its orientation; the three discrete components specify the ellipse's RGB colour. Note though, that far more general object classes are possible, for example that of all polygons in $\mathbb{R}^2$. We assume that each object $x \in D \times L$ determines a region $R(x) \subseteq T$ in image space that is 'occupied' by the object, and refer to it as the *'template'* of x in T.

An object configuration is simply a finite vector of objects $\vec{\mathbf{x}} = (x_1, \cdots, x_n)$ where $x_j \in D \times L$, $j = 1, \cdots, n$, $n \geq 0$. The objects may be in any spatial relation to each other; the number of objects is variable and may be zero. A configuration $\vec{\mathbf{x}}$ is mapped to a *'signal'* image $\theta_t(\vec{\mathbf{x}})$, $t \in T$. In the absence of blur and shadows,

$$\theta_t(\vec{\mathbf{x}}) := \begin{cases} \theta(x_j) & \text{if } t \in R(x_j) \setminus \cup_{k<j} R(x_k) \\ \theta_0 & \text{if } t \in T \setminus \cup R(x_j) \end{cases} \tag{1}$$

where $\theta(x) \in \Theta$ for some parameter space Θ is the object's assumed ideal (noise-free) image and θ_0 is the assumed 'background' signal. The example given in this work assumes a single colour for $\theta(x)$, but more complex images with texture variation may be possible. Note that among the objects occupying a given pixel, the one with the smallest index determines the signal. Thus, the model explicitly accounts for occlusion, in contrast to unordered object processes [2, 12] and in a simpler way than in [17].

2.2 Gibbsian modification of Poisson sequential object processes

The basic reference model for an object configuration is the Poisson object process. Under this model, the sequence length is Poisson distributed, and objects are independent and identically distributed. More formally, given a finite diffuse Borel measure μ on $(D, \mathcal{B}_D)$,

usually Lebesgue measure, with $\mu(D) > 0$, and a Borel probability measure μ_L on the space of object parameters $(L, \mathcal{B}_L)$, write

$$\nu(F) := \sum_{n=0}^{\infty} \frac{e^{-\mu(D)}}{n!} \int_{(D\times L)^n} \mathbf{1}\{(x_1, \ldots, x_n) \in F\}\, d(\mu \times \mu_L)(x_1) \ldots d(\mu \times \mu_L)(x_n)$$

defined for F in the σ-algebra on finite marked sequences generated by the product Borel σ-fields on $(D \times L)^n$ where the term for $n = 0$ should be read as $\exp[-\mu(D)]\,\mathbf{1}\{\emptyset \in F\}$.

Other distributions may be specified by a density with respect to ν. We use the Gibbs formulation and write $g(\vec{\mathbf{x}}) \propto \exp[-E(\vec{\mathbf{x}})]$. We refer to the exponent $E(\cdot)$ as a *'Hamiltonian'* or *'energy function'* in keeping with standard statistical physics terminology. Since infinite valued energies are allowed, no restriction other than absolute continuity is imposed.

Our data consist of video sequences, hence we consider sequences $\mathbf{x} = (\vec{\mathbf{x}}^1, \ldots, \vec{\mathbf{x}}^I)$ distributed according to

$$f(\mathbf{x}) \propto \exp[-U(\mathbf{x})] \tag{2}$$

with respect to the I-fold product measure ν^I, analogously to the single image frame case above, with energy function $U(\cdot)$.

3 The regression model

Suppose Θ and V are compatible in the sense that an L_p distance can be defined between the data and signal images. Then, upon observation of the video sequence $\mathbf{y}$, write

$$U(\mathbf{x}) = \sum_{i=1}^{I} \lambda_i L_p(\mathbf{y}^i, \theta(\vec{\mathbf{x}}^i))^p \tag{3}$$

for $p = 1, 2$ and $\lambda_i > 0$. This energy function describes the 'forward problem' of image formation and measures the goodness of fit between the hypothesised object configurations and the actual data sequence. In probabilistic terms, (3) amounts to assuming independent Gaussian noise at each pixel for $p = 2$, and independent Laplacian (double exponential) noise for $p = 1$ [2]. Clearly, where appropriate, other types of noise could be used instead.

Given observation of $\mathbf{y}$, we seek 'the' minimiser of the energy function. Since it is a sum of individual pixel error terms, optimisation of (3) over object configurations is equivalent to least squares respectively least absolute deviation regression. Note however that such a minimum may be non-existent or non-unique. Indeed, adding extra objects 'behind' the signal of those closer to the camera (having a lower index) does not affect the energy function.

3.1 Markov property

Note that the potential energy required for adding object ξ to the $\vec{\mathbf{x}}^i$ in frame i to obtain the vector $(\vec{\mathbf{x}}^i, \xi)$ for signal (1) and energy (3) is given by

$$\lambda_i \sum_{t \in R(\xi) \setminus \cup_k R(x_k^i)} \left[|y_t^i - \theta(\xi)|^p - |y_t^i - \theta_0|^p \right]. \tag{4}$$

Equation (4) clearly depends only on $R(\xi)$ and those objects $R(x_k^i)$ that overlap $R(\xi)$. Note that if ξ were added at position k, (4) would be replaced by

$$\sum_{t \in R(\xi) \setminus \cup_{l<k} R(x_l^i)} \left[|y_t^i - \theta(\xi)|^p - |y_t^i - \theta(\vec{\mathbf{x}}^i)|^p \right].$$

As $R(\xi) \setminus \cup_{l<k} R(x_l^i) = R(\xi) \setminus \bigcup_{l<k: R(x_l^i) \cap R(\xi) \neq \emptyset} R(x_l^i)$, the difference in energy does not depend on those $R(x_l^i)$ with $l < k$ that do not overlap $R(\xi)$. The second term involves only x_l^i with $l \geq k$, and then only those that overlap $R(\xi)$ (otherwise, the signal at t would be θ_0). Note that the role of the objects with $l < k$ and $l \geq k$ is different: the first determine the set of pixels over which to take the sum, the second ones' signal value is used.

Since the energy is finite, the above remarks amount to saying that the single frame energy function E defines a *'Markov sequential object process'* [15] with respect to the *overlapping objects relation*

$$u \sim v \Leftrightarrow R(u) \cap R(v) \neq \emptyset. \tag{5}$$

If $u \sim v$, the object v is said to be a (directed) *neighbour* of u.

3.2 Relation to Hough transform

Consider the track of a newly arrived object against an empty background. Write b for its birth frame, d its death frame, u_i for the object in frame $i = b, \ldots, d$, and $(v_i)_{i=b}^{d-1}$ for the translation vectors between frames. The track thus parameterised will be denoted by $\tilde{u}$. Furthermore, suppose that the signal is translation invariant, so that $\theta_t(u + \Delta) = \theta_{t-\Delta}(u)$ for all pixels t, Δ with $t - \Delta \in T$. Then, the decrease in energy is

$$\begin{aligned} U(\emptyset) - U(\tilde{u}) &= \lambda_b \sum_{t \in R(u_b)} \left[|y_t^b - \theta_0|^p - |y_t^b - \theta(u_b)|^p \right] \\ &+ \sum_{i=b}^{d-1} \lambda_{i+1} \sum_{t \in R(u_i)} \left[|y_{t+v_i}^{i+1} - \theta_0|^p - |y_{t+v_i}^{i+1} - \theta(u_i)|^p \right]. \end{aligned}$$

The first term corresponds to the Hough transform for detecting the initial presence of an object by letting each pixel vote for the objects that contain it with strength $|y_t^b - \theta_0|^p - |y_t^b - \theta(u_b)|^p$; the second term is a recursive Hough transform voting for the movement from u_i by v_i with strength

$$|y_{t+v_i}^{i+1} - \theta_0|^p - |y_{t+v_i}^{i+1} - \theta(u_i)|^p$$

for each pair of pixels $(t, t + v_i)$ with $t \in R(u_i)$.

4 Regularisation

Although model (3) tells us how to deal with Hough transforms in cases of occlusion, the optima thus obtained have undesirable features. Since the energy function depends on the signal only, one may add any number of objects hidden behind those in an optimum solution

without affecting the optimality. Thus, we impose a repulsive Markov overlapping object term to favour solutions with minimum numbers of objects. Also the correlation between objects in subsequent frames is not taken into account. Hence, we add link terms favouring objects close to similar ones in neighbouring video frames.

More formally, we shall specify the probability density of an I-variate sequential spatial process. An alternative equivalent definition is to specify a density of a sequential marked spatial process with marks in $\{1,\ldots,I\}$ denoting the frame in which the object lives. Write $\vec{\mathbf{s}}$ for the vector of link functions, and define the energy function

$$U(\mathbf{x};\vec{\mathbf{s}};\mathbf{y}) := \lambda_1 \sum_{i=1}^{I} V_1(\mathbf{y}^i|\vec{\mathbf{x}}^i) + \lambda_2 \sum_{i=1}^{I} V_2(\vec{\mathbf{x}}^i) + \lambda_3 \sum_{i=1}^{I-1} V_3(\vec{\mathbf{x}}^i, \vec{\mathbf{x}}^{i+1}, s^{i,i+1}) \tag{6}$$

and weights $\lambda_i > 0$. The reference distribution is the product measure of unit rate Poisson sequences for the ordered objects in each frame, and counting measure on the countable space of link functions for the matching between frames. The function V_1 corresponds to (3). In the next subsections, we shall discuss respectively V_2 and V_3 in more detail.

4.1 Within frame interaction

As in object recognition, in order to avoid over fitting, a natural condition for V_2 is to impose Markovianity with respect to the overlapping object relation. In this paper, we use the Strauss prior given by

$$V_2(\vec{\mathbf{x}}) = \beta n(\vec{\mathbf{x}}) - \gamma n_o(\vec{\mathbf{x}})$$

where $n(\vec{\mathbf{x}})$ is the length of the point sequence $\vec{\mathbf{x}}$, and $n_o(\vec{\mathbf{x}})$ is the number of pairs $\{u,v\}$ in $\vec{\mathbf{x}}$ for which $R(u) \cap R(v) \neq \emptyset$. The parameter β is a real, γ is negative. Clearly, if we add a new object u to the sequence,

$$V_2((\vec{\mathbf{x}},u)) - V_2(\vec{\mathbf{x}}) = \beta - \gamma \#\{x_i \in \vec{\mathbf{x}} : R(u) \cap R(x_i) \neq \emptyset\}$$

depends only on those existing objects that overlap $R(u)$. Therefore, such updates are 'local' operations.

More generally, one might use a pairwise interaction model

$$V_2(\vec{\mathbf{x}}) = \sum_k \beta(x_k) + \sum_{k<l} \varphi(x_k, x_l)$$

for some **symmetric function** $\varphi(\cdot,\cdot) \geq 0$, and intensity function $\beta(\cdot)$. The first order mark term could penalise colours close to the background, or favour large objects over small ones.

4.2 Propagation over frames

In the previous section, we defined inhibition between the objects in a single frame. Between frames, we would like to have attraction, that is, temporal cohesion. Let

$$S_{m,n} = \{(M,N,\pi) : M \subseteq \{1,\ldots,m\}; N \subseteq \{1,\ldots,n\}; |M| = |N|\}$$

with $m, n \in \mathbb{N}_0$, and $\pi : M \to N$ a bijection. Given an $s \in S_{m,n}$, the coordinates are written by $M(s)$, $N(s)$, and $\pi(s)$ respectively [16]. Then, for two configurations $\vec{\mathbf{x}}^i$, $\vec{\mathbf{x}}^{i+1}$ in consecutive frames, and $s^{i,i+1} \in S_{n(\vec{\mathbf{x}}^i),n(\vec{\mathbf{x}}^{i+1})}$, take V_3 of the form

$$V_3(\vec{\mathbf{x}}^i, \vec{\mathbf{x}}^{i+1}, s^{i,i+1}) = \sum_{l \in M(s^{i,i+1})} \tau(\vec{\mathbf{x}}^i_l, \vec{\mathbf{x}}^{i+1}_{(\pi(s^{i,i+1}))(l)}) + \sum_{l \notin M(s^{i,i+1})} \lambda(\vec{\mathbf{x}}^i_l) + \sum_{l \notin N(s^{i,i+1})} \lambda(\vec{\mathbf{x}}^{i+1}_l) \quad (7)$$

$$+ \sum_{x_l \sim x_k \in \vec{\mathbf{x}}^i; l<k \in M(s^{i,i+1})} \rho \mathbf{1}\left\{\pi(s^{i,i+1})(l) > \pi(s^{i,i+1})(k)\right\}$$

$$+ \sum_{x_l \sim x_k \in \vec{\mathbf{x}}^{i+1}; l<k \in N(s^{i,i+1})} \rho \mathbf{1}\left\{\pi^{-1}(s^{i,i+1})(l) > \pi^{-1}(s^{i,i+1})(k)\right\}.$$

The positive valued function $\lambda(\cdot)$ penalises unmatched objects, whereas the positive valued function $\tau(\cdot,\cdot)$ is a dissimilarity measure for its arguments (symmetric of course). The parameter ρ forces a similar ranking in index between objects that overlap in one frame in all other frames.

The function (7) can be thought of as a discrete Markov transition probability kernel in frame-time for the step from $\vec{\mathbf{x}}^i$ to $\vec{\mathbf{x}}^{i+1}$.

5 Metropolis–Hastings sampler

Our goal is to find the optimal configuration sequence and links, in the sense of minimising the energy function. We shall do this by simulated annealing within the Metropolis–Hastings framework.

More specifically, for any scene sequence $\mathbf{x} = (\vec{\mathbf{x}}^1, \ldots, \vec{\mathbf{x}}^I)$, matchings $\mathbf{s} = (s^{i,i+1})_{1=1,\ldots,I-1}$, and data sequence $\mathbf{y}$, our transition kernel has the following form:

$$P(\mathbf{x}; \mathbf{s}; \mathbf{y}, F) = \sum_{i=1}^{m} p_i P_i(\mathbf{x}; \mathbf{s}; \mathbf{y}, F) \quad (8)$$

where $p_i \in (0,1)$ is the probability of selecting move type i from among the m possibilities, and $P_i(\mathbf{x}, F)$ is the probability using move type i of updating state $\mathbf{x}$ into a new state belonging to the set F. Clearly, $\sum_i p_i = 1$. The probability kernels P_i follow the proposal/acceptance strategy common to all Metropolis–Hastings algorithms[7, 4]. For example,

$$\begin{aligned} P_1(\mathbf{x}; \mathbf{s}; \mathbf{y}, F) &= \sum_{i=1}^{I} \frac{1}{I\mu(D)} \frac{1}{n(\vec{\mathbf{x}}^i)+1} \sum_{j=1}^{n(\vec{\mathbf{x}}^i)+1} \int_{D \times L} 1\{s_{ij}(\mathbf{x}, \xi) \in F\} \alpha(\mathbf{x}, s_{ij}(\mathbf{x}, \xi)) d(\mu \times \mu_L)(\xi) \\ &+ 1\{\mathbf{x} \in F\} \sum_{i=1}^{I} \frac{1}{I\mu(D)} \frac{1}{n(\vec{\mathbf{x}}^i)+1} \sum_{j=1}^{n(\vec{\mathbf{x}}^i)} \int_{D \times L} (1 - \alpha(\mathbf{x}, s_{ij}(\mathbf{x}, \xi))) d(\mu \times \mu_L)(\xi) \end{aligned}$$

is the probability kernel for insertion of a randomly chosen new object without links at a random position in the sequence describing a randomly chosen frame. The notation $\alpha(\cdot,\cdot)$ is used for the acceptance probabilities.

Inspired by [4, 16], we propose the following move types: birth and death of matched or unmatched objects, reorderings and rematchings, and modification of object location and characteristics. Below, we shall derive the Hastings ratios for these move types.

Birth and death moves are implemented in the following way. The probability of such a move is p_{bd}. Then, with probability $1/2$, a birth is selected, with probability $1/2$, a death. Then, it is decided randomly whether the object to be born or die is linked to objects in either of the neighbouring frames or not. In the first case, a further random choice is made about the type of the connection.

We shall assume that the displacement kernel k is symmetric in the sense that $k(u|v) = k(v|u)$. Reversibility of change moves is guaranteed by the assumption that $K(u|v) > 0$ if and only if $K(v|u) > 0$. Furthermore, we shall assume that for every pair of objects $u \neq v$, there is a path $u = w_1, \ldots, w_n = v$ of distinct objects such that $K(w_2|w_1) \cdots K(w_n|w_{n-1}) > 0$.

Birth of singly matched object A frame is selected uniformly from the set $\{1, \ldots, I\}$. With probability $1/2$, the object to be born is a modified version of some object in the previous frame, with probability $1/2$ of one in the next image frame.

Thus, suppose frame i is chosen with current configuration $\vec{\mathbf{x}}^i$ and a new object would be matched to one in frame $i-1$. If $i = 1$, no transition occurs. Otherwise, try and choose one of the points in frame $i-1$ that is not yet matched to an object in the current frame uniformly at random, say x_k^{i-1}, again with the proviso that if such an object cannot be found, no transition occurs. Generate a new object according to the probability kernel $k(\xi|x_k^{i-1})d(\mu \times \mu_L)(\xi)$ and insert it into the object sequence $\vec{\mathbf{x}}^i$ at a uniformly chosen position j to obtain $c_j(\vec{\mathbf{x}}^i, \xi)$. Add the match between x_k^{i-1} and ξ to $s^{i-1,i}$ and adjust the indices to obtain $c_j(s^{i-1,i}, \xi, x_k^{i-1})$ and $c_j(s^{i,i+1}, \xi)$, the latter provided $i < I$. Accept the move according to the Hastings ratio

$$\frac{n(\vec{\mathbf{x}}^{i-1}) - |M(s^{i-1,i})|}{|N(s^{i-1,i}) \setminus M(s^{i,i+1})| + 1} \frac{1}{k(\xi|x_k^{i-1})} \exp\left[-U(\mathbf{x}'; \mathbf{s}'; \mathbf{y}) + U(\mathbf{x}; \mathbf{s}; \mathbf{y})\right]$$

where $\mathbf{x}'$ differs from $\mathbf{x}$ only in frame i with $\vec{\mathbf{x}}^{i\prime} = c_j(\vec{\mathbf{x}}^i, \xi)$, and $\mathbf{s}'$ is identical to $\mathbf{s}$ except for the matchings involving frame i which are replaced by $c_j(s^{i-1,i}, \xi, x_k^{i-1})$ and $c_j(s^{i,i+1}, \xi)$ whenever appropriate.

The move that adds an object matched to a similar one in the next image frame is completely analogous.

Birth of doubly matched object A frame is selected uniformly from the set $\{1, \ldots, I\}$. If frame 1 or I is selected, the current state remains unchanged.

Thus, suppose frame $i \in \{2, \ldots, I-1\}$ is chosen with current configuration $\vec{\mathbf{x}}^i$. Try and choose one of the points in frame $i-1$ that is not yet matched to an object in the current frame uniformly at random, say x_k^{i-1}, and a randomly chosen object in frame $i+1$ that is not yet matched to frame i either, say x_m^{i+1}. If either of such objects cannot be found, no transition occurs. Generate a new object according to the probability kernel $k(\xi|x_k^{i-1})d(\mu \times \mu_L)(\xi)$ and insert it into the object sequence $\vec{\mathbf{x}}^i$ at a uniformly chosen position j to obtain $c_j(\vec{\mathbf{x}}^i, \xi)$. Add the match between x_k^{i-1} and ξ to $s^{i-1,i}$ and adjust the indices to obtain $c_j(s^{i-1,i}, \xi, x_k^{i-1})$. Also add the match between ξ and x_m^{i+1} to $s^{i,i+1}$ taking into account the updated indices in

the object sequence in frame i to obtain $c_j(s^{i,i+1}, \xi, x_m^{i+1})$. Accept the move according to the Hastings ratio

$$\frac{\left(n(\vec{\mathbf{x}}^{i-1}) - |M(s^{i-1,i})|\right)\ \left(n(\vec{\mathbf{x}}^{i+1}) - |N(s^{i,i+1})|\right)}{(|M(s^{i,i+1}) \cap N(s^{i-1,i})| + 1)\ k(\xi|x_k^{i-1})}\ \exp\left[-U(\mathbf{x}';\mathbf{s}';\mathbf{y}) + U(\mathbf{x};\mathbf{s};\mathbf{y})\right]$$

where $\mathbf{x}'$ differs from $\mathbf{x}$ only in frame i with $\vec{\mathbf{x}}^{i\prime} = c_j(\vec{\mathbf{x}}^i, \xi)$, and $\mathbf{s}'$ is identical to $\mathbf{s}$ except for the matchings involving frame i which are replaced by $c_j(s^{i-1,i}, \xi, x_k^{i-1})$ and $c_j(s^{i,i+1}, \xi, x_m^{i+1})$.

Birth of unmatched object A frame is selected uniformly from the set $\{1, \ldots, I\}$.

Thus, suppose frame $i \in \{1, \ldots, I\}$ is chosen with current configuration $\vec{\mathbf{x}}^i$. Generate a new object according to $(\mu \times \mu_L)(D \times L)^{-1} d(\mu \times \mu_L)(\xi)$, and insert it into the object sequence $\vec{\mathbf{x}}^i$ at a uniformly chosen position j to obtain $c_j(\vec{\mathbf{x}}^i, \xi)$, and adjust the indices for the matchings involving frame i to obtain $c_j(s^{i-1,i}, \xi)$ and $c_j(s^{i,i+1}, \xi)$, if they exist. Accept the move according to the Hastings ratio

$$\frac{(\mu \times \mu_L)(D \times L)}{n(\vec{\mathbf{x}}^i) + 1 - |N(s^{i-1,i}) \cup M(s^{i,i+1})|}\ \exp\left[-U(\mathbf{x}';\mathbf{s}';\mathbf{y}) + U(\mathbf{x};\mathbf{s};\mathbf{y})\right]$$

where $\mathbf{x}'$ differs from $\mathbf{x}$ only in frame i with $\vec{\mathbf{x}}^{i\prime} = c_j(\vec{\mathbf{x}}^i, \xi)$, and $\mathbf{s}'$ is identical to $\mathbf{s}$ except for the matchings involving frame i which are replaced by $c_j(s^{i-1,i}, \xi, x_k^{i-1})$ and $c_j(s^{i,i+1}, \xi)$ whenever appropriate.

Death of singly matched object A frame is selected uniformly from the set $\{1, \ldots, I\}$. With probability $1/2$, the object to be deleted is a modified version of some object in the previous frame, with probability $1/2$ of one in the next image frame.

Thus, suppose frame i is chosen with current configuration $\vec{\mathbf{x}}^i$ and the object to be deleted should be matched to one in frame $i-1$. If $i = 1$ or there are no suitably matched objects in frame i, no transition occurs. Otherwise, choose one of the points in the current frame that is matched to an object in frame $i-1$ uniformly at random, say x_k^{i-1} and x_l^i, and delete x_l^i from $\vec{\mathbf{x}}^i$ to obtain $\vec{\mathbf{x}}^i_{(-l)}$. Delete the match between x_k^{i-1} and x_l^i from $s^{i-1,i}$ and adjust the indices to obtain $s^{i-1,i}_{(-l^N)}$ and $s^{i,i+1}_{(-l^M)}$, the latter provided $i < I$. Accept the move according to the Hastings ratio

$$\frac{|N(s^{i-1,i}) \setminus M(s^{i,i+1})|\ k(x_l^i|x_k^{i-1})}{n(\vec{\mathbf{x}}^{i-1}) - |M(s^{i-1,i})| + 1}\ \exp\left[-U(\mathbf{x}';\mathbf{s}';\mathbf{y}) + U(\mathbf{x};\mathbf{s};\mathbf{y})\right]$$

where $\mathbf{x}'$ differs from $\mathbf{x}$ only in frame i with $\vec{\mathbf{x}}^{i\prime} = \vec{\mathbf{x}}^i_{(-l)}$, and $\mathbf{s}'$ is identical to $\mathbf{s}$ except for the matchings involving frame i, which are replaced by $s^{i-1,i}_{(-l^N)}$ and $s^{i,i+1}_{(-l^M)}$ whenever appropriate.

The move that deletes an object matched to a similar one in the next image frame is completely analogous.

Death of doubly matched object A frame is selected uniformly from the set $\{1,\ldots,I\}$.

Thus, suppose frame i is chosen with current configuration $\vec{\mathbf{x}}^i$ and the object to be deleted should be matched to modified ones in frames $i-1$ and $i+1$. If $i=1$ or I, or there are no suitably matched objects in frame i, no transition occurs. Otherwise, choose one of the doubly matched objects in frame i uniformly at random, say x_i^l with its matches x_k^{i-1} and x_m^{i+1}, and delete x_l^i to obtain $\vec{\mathbf{x}}^i_{(-l)}$. Delete the match between x_k^{i-1} and x_l^i from $s^{i-1,i}$ and adjust the indices to obtain $s^{i-1,i}_{(-l^N)}$. Also delete the match between x_l^i and x_m^{i+1} and adjust the indices to obtain $s^{i,i+1}_{(-l^M)}$. Accept the move according to the Hastings ratio

$$\frac{|M(s^{i,i+1})\cap N(s^{i-1,i})|\; k(x_l^i|x_k^{i-1})}{(n(\vec{\mathbf{x}}^{i-1})-|M(s^{i-1,i})|+1)\,(n(\vec{\mathbf{x}}^{i+1})-|N(s^{i,i+1})|+1)}\;\exp\left[-U(\mathbf{x}';\mathbf{s}';\mathbf{y})+U(\mathbf{x};\mathbf{s};\mathbf{y})\right]$$

where $\mathbf{x}'$ differs from $\mathbf{x}$ only in frame i with $\vec{\mathbf{x}}^{i\prime}=\vec{\mathbf{x}}^i_{(-l)}$, and $\mathbf{s}'$ is identical to $\mathbf{s}$ except for the matchings involving frame i, which are replaced by $s^{i-1,i}_{(-l^N)}$ and $s^{i,i+1}_{(-l^M)}$.

Death of unmatched object A frame is selected uniformly from the set $\{1,\ldots,I\}$.

Thus, suppose frame i is chosen with current configuration $\vec{\mathbf{x}}^i$ and the object to be deleted should not be matched to any object in a consecutive frame. If there are no such objects in frame i, no transition occurs. Otherwise, choose one of the unmatched points in the current frame uniformly at random, say x_l^i, and delete x_l^i from $\vec{\mathbf{x}}^i$ to obtain $\vec{\mathbf{x}}^i_{(-l)}$. Adjust the indices to obtain $s^{i-1,i}_{(-l^N)}$ and $s^{i,i+1}_{(-l^M)}$, if they exist. Accept the move according to the Hastings ratio

$$\frac{n(\vec{\mathbf{x}}^i)-|N(s^{i-1,i})\cup M(s^{i,i+1})|}{(\mu\times\mu_L)(D\times L)}\;\exp\left[-U(\mathbf{x}';\mathbf{s}';\mathbf{y})+U(\mathbf{x};\mathbf{s};\mathbf{y})\right]$$

where $\mathbf{x}'$ differs from $\mathbf{x}$ only in frame i with $\vec{\mathbf{x}}^{i\prime}=\vec{\mathbf{x}}^i_{(-l)}$, and $\mathbf{s}'$ is identical to $\mathbf{s}$ except for the matchings involving frame i which are replaced by $s^{i-1,i}_{(-l^N)}$ and $s^{i,i+1}_{(-l^M)}$ whenever appropriate.

Modification of permutation order A frame is selected uniformly from the set $\{1,\ldots,I\}$. If the selected frame contains less than two objects, the current state remains unchanged.

Otherwise, select a current position k, and another position $l\neq k$ uniformly over the options. Then propose to interchange the objects x_l^i and x_k^i in the current configuration $\vec{\mathbf{x}}^i$ to obtain $c_{kl}(\vec{\mathbf{x}}^i)$. Adjust the indices for the matchings involving frame i to obtain $c_{kl}(s^{i-1,i})$ and $c_{kl}(s^{i,i+1})$, if they exist. Accept the move according to the Hastings ratio

$$\exp\left[-U(\mathbf{x}';\mathbf{s}';\mathbf{y})+U(\mathbf{x};\mathbf{s};\mathbf{y})\right]$$

where $\mathbf{x}'$ differs from $\mathbf{x}$ only in frame i with $\vec{\mathbf{x}}^{i\prime}=c_{kl}(\vec{\mathbf{x}}^i)$, and $\mathbf{s}'$ is identical to $\mathbf{s}$ except for the matchings involving frame i which are replaced by $c_{kl}(s^{i-1,i})$ and $c_{kl}(s^{i,i+1})$ whenever appropriate.

As for births and deaths of objects, assume the probability of proposing to add a match equals that of proposing to delete a match.

Birth of match Select a pair $(i-1,i)$, $i \in \{2,\ldots,I\}$ of consecutive frames uniformly at random.

Try and choose one of the points in frame $i-1$ that is not yet matched to an object in frame i uniformly at random, say x_k^{i-1}, and a randomly chosen object in frame i that is not yet matched to frame $i-1$, say x_l^i. If either of such objects cannot be found, no transition occurs. Otherwise, add the match (x_k^{i-1}, x_l^i) to $s^{i-1,i}$ to obtain $c_+(s^{i-1,i},k,l)$. Accept the move according to the Hastings ratio

$$\frac{\big(n(\vec{\mathbf{x}}^{i-1}) - |M(s^{i-1,i})|\big)\ \big(n(\vec{\mathbf{x}}^{i}) - |N(s^{i-1,i})|\big)}{|N(s^{i-1,i})|+1}\ \exp\big[-U(\mathbf{x};\mathbf{s}';\mathbf{y}) + U(\mathbf{x};\mathbf{s};\mathbf{y})\big]$$

where $\mathbf{s}'$ is identical to $\mathbf{s}$ except for $s^{i-1,i}$ which is replaced by $c_+(s^{i-1,i},k,l)$.

Death of match Select a pair $(i-1,i)$, $i \in \{2,\ldots,I\}$ of consecutive frames uniformly at random.

Try and choose one of the points in frame $i-1$ that is matched to an object in frame i uniformly at random, say x_k^{i-1} matched to x_l^i. If no such object can be found, no transition occurs. Otherwise, delete the match between x_k^{i-1} and x_l^i from $s^{i-1,i}$ to obtain $c_-(s^{i-1,i},k,l)$. Accept the move according to the Hastings ratio

$$\frac{|N(s^{i-1,i})|}{(n(\vec{\mathbf{x}}^{i-1}) - |M(s^{i-1,i})|+1)\ (n(\vec{\mathbf{x}}^{i}) - |N(s^{i-1,i})|+1)}\ \exp\big[-U(\mathbf{x};\mathbf{s}';\mathbf{y}) + U(\mathbf{x};\mathbf{s};\mathbf{y})\big]$$

where $\mathbf{s}'$ is identical to $\mathbf{s}$ except for $s^{i-1,i}$ which is replaced by $c_-(s^{i-1,i},k,l)$.

Modification of object A frame is selected uniformly from the set $\{1,\ldots,I\}$.

Thus, suppose frame i is chosen with current configuration $\vec{\mathbf{x}}^i$. Provided the length of $\vec{\mathbf{x}}^i$ is non-zero, choose one of the objects in the current frame uniformly at random, say x_j^i, and replace it by a new object generated according to the probability kernel $K(\xi|x_j^i)d(\mu\times\mu_l)(\xi)$ to obtain $c_j(\vec{\mathbf{x}}^i,\xi)$. Do not alter the match functions, in other words, give ξ the matches of x_j^i. Accept the move according to the Hastings ratio

$$\frac{K(x_j^i|\xi)}{K(\xi|x_j^i)}\ \exp\big[-U(\mathbf{x}';\mathbf{s};\mathbf{y}) + U(\mathbf{x};\mathbf{s};\mathbf{y})\big]$$

where $\mathbf{x}'$ differs from $\mathbf{x}$ only in frame i with $\vec{\mathbf{x}}^{i\prime} = c_j(\vec{\mathbf{x}}^i,\xi)$.

5.1 Proof of convergence

In addition to the conditions on k and K, assume all p_i in (8) are non-zero, and the (sequential) object process defined by the Hamiltonians $V_1,\ldots,V_3$ is locally stable and hereditary (so that (6) can be properly normalised into a probability distribution). Restrict the Metropolis–Hastings chain to the set of $(\mathbf{x},\mathbf{s})$ which have positive probability density of occurring. We claim that this chain Z_n, $n \in \mathbb{N}_0$ with the transitions outlined in the previous subsection converges in total variation to the distribution (6) from almost all initial states.

To prove the claim, we must establish aperiodicity and positive recurrence. To do so, consider the Dirac measure δ_0 on $z_a := \{(\emptyset, \emptyset)$. Then, Z_n is irreducible with respect to this measure, as the empty state z_a can be reached with positive probability by visiting each frame in turn, deleting the matches first and then deleting the by now unmatched objects, if necessary applying a series of change moves first. By assumption, the proposal probabilities are positive, and under the assumptions on the kernels k and K and that the target distribution is hereditary, so are the acceptance probabilities. We conclude that Z_n is irreducible.

By construction, (6) is an invariant measure, so that Z_n, being irreducible, is positive recurrent, and (6) is the unique invariant probability measure [21, p. 155].

The state z_a is an accessible atom for Z_n. Since the 1-step probability of staying in z_a exceeds for instance the death proposal probability, hence is strictly positive, Z_n is strongly aperiodic [21, p. 150].

It remains to establish that the atom z_a is Harris recurrent for the Markov chain Z_n. To do so, employ renewal theory. Note that the stopping times τ_j, $j = 1, \ldots, \infty$ defined by $\tau_0 := 0$ and recursively by

$$\tau_j := \inf\{n > \tau_{j-1} : Z_n = z_a\}$$

satisfy $E[\tau_1 | Z_0 = z_a] < \infty$, and a fortiori $P(\tau_1 < \infty | Z_0 = z_a) = 1$ by Thm 4.5.3 in [21]. (In equilibrium, Z_n spends a non-null fraction of time in z_a). Hence, when started in the empty state, the Markov chain returns to z_a infinitely often. In other words $\{z_a\}$ is a Harris recurrent set.

From a computational point of view, most of the Hastings ratios involve only the frame being updated, and only the links with direct neighbours. Therefore, the Markov properties discussed before imply a local computational effort.

6 Example

We study an example in sports tracking: the ball and bat in a table tennis sequence (see Figure 1). Both objects of interest can be conveniently described mathematically by an ellipse with three shape parameters: the orientation and half lengths of both axes. We took 5 pixels for the minimum half axis length, 40 for the maximum one. The reference measures for position and shape were Lebesgue. The colour of an ellipse was described by a discrete RGB space, equipped with the equal weight mixture of data frame histograms.

Regarding the parameters in model (6), the background colour was found by consideration of the RGB colour histograms. In V_1, an L_2 criterion was used with $\sigma = 128$ in Gaussian noise terms. The function V_2 adds a penalty 50 for each object, 5 for each pair of overlapping ellipses. In V_3, we took $\lambda(\cdot) \equiv 5 = \rho$ constant; the dissimilarity term $\tau(x_1, x_2)$ is the sum of $d^2(x_1, x_2)/800$ with $d(\cdot, \cdot)$ the Euclidean distances between centres, of the absolute differences in axes lengths and orientation (modulo π), and normalised absolute differences in RGB space, with normalisation $1/255$.

With respect to the Metropolis–Hastings algorithm, we used a Gaussian kernel with variance 800 for $k(\cdot|\cdot)$ just for the position; shape and colour attributes remain unchanged. The

modification kernel $K(\cdot|\cdot)$ is a mixture of kernels for modification of location, size, orientation, and colour with equal weights. The real valued characteristics are updated distributed uniformly in a neighbourhood corresponding to 5 percent of the allowed range, the discrete colour is chosen from the reference probability measure. The probability of selecting a matched birth/death, an unmatched birth/death, a change in index or match were all $1/10$. The probability of a change move was $2/5$. Within these classes, uniformly distributed choices were made.

A convenient way to represent the relative position of objects is to plot the average *depth map*

$$\sum_{j=1}^{k} d_t(\vec{X}_j); \quad t \in T, \tag{9}$$

over k Metropolis–Hastings steps after burn-in for each image frame, where the depth of object vector $\vec{\mathbf{x}} = (x_1, \ldots, x_n)$ at pixel $t \in T$ is defined as

$$d_t(\vec{\mathbf{x}}) := \begin{cases} \frac{n-j+1}{n} 255 & \text{if } t \in R(x_j) \setminus \cup_{k<j} R(x_k) \\ 0 & \text{if } t \in T \setminus \cup R(x_j) \end{cases}$$

for $n \geq 1$, and $d_t(\emptyset) \equiv 0$.

For the tennis sequence, after a burn-in of 30,000 steps, annealing was performed for temperatures $T_n = 1.0/(1+0.005*n)$ with 50 steps for each $n = 0, \ldots, 1000$. The near-optimal configuration is as depicted in Figure 1. The annealed depth map is the correct one with the bat under the ball. If we would not have included a temporal cohesion term, the depth values of both bat and ball in the first and third frame would have been $(3 \times 255)/4$. Note that there are slight deviations from an elliptic shape in the bat because of the perspective.

The depth map (9) at fixed temperature 5 with $k = 300,000$ Metropolis–Hastings steps after a burn-in of 50,000 steps is given in Figure 2. It can be seen that the correct relative depths are found, but the depth value of, say, the ball is less than 255 with smaller values near the edge.

7 Summary and future work

In this paper, we presented an application of Markov sequential object processes to the calculation of depth maps for scenes involving a variable number of interacting objects that may change over time with a view to 3D-TV. The model proposed here is able to cope with the occlusion caused by having objects at different depths, maintains the identity of objects as well as their relative depth over consecutive video frames, and ensures fit to the data. The computational complexity of the model can be handled by a suitably designed Metropolis–Hastings algorithm. In contrast to commonly used filtering methods, the sampler goes back and forth between frames, gathering depth information when objects overlap and transferring this information on to other frames that do not provide depth cues. As most interest focuses on the optimal depth map, a simulated annealing scheme may be used. The approach was illustrated by a table tennis video sequence.

This work concentrated on objects that are described by a few shape parameters. However, the theoretical framework presented here is not limited to such cases, indeed includes e.g. polygons of arbitrary shape, or even completely general closed sets [18]. In the future, we intend to formalise such a segmentation based approach and evaluate its effectiveness for scenes that are not composed of simple objects against a homogeneous background.

Acknowledgements

This research is supported by the Technology Foundation STW, the applied science division of NWO, and the technology programme of the Ministry of Economic Affairs (project CWI.6156 'Markov sequential point processes for image analysis and statistical physics'). The author would like to thank Dr. C. Varekamp (Philips Research) for data and interesting discussions.

References

[1] Baddeley, A.J. and Lieshout, M.N.M. van. Object recognition using Markov spatial processes. *Proceedings, 11th IAPR International Conference on Pattern Recognition*, pages B 136–139. IEEE Computer Society Press, Los Alamitos, California, 1992.

[2] Baddeley, A.J. and Lieshout, M.N.M. van. Stochastic geometry models in high-level vision. In *Statistics and Images, Volume 1*, K.V. Mardia and G.K. Kanji (Eds.) *Advances in Applied Statistics*, a supplement to *Journal of Applied Statistics*, 20:231–256. Carfax, Abingdon, 1993.

[3] Eubank, R.L. *A Kalman filter primer.* Boca Raton: Chapman & Hall/CRC, 2006.

[4] Geyer, C.J. and Møller , J. Simulation procedures and likelihood inference for spatial point processes. *Scandinavian Journal of Statistics*, 21:359–373, 1994.

[5] Goodman, I.R., Mahler, R.P.S. and Nguyen, H.T. Mathematics of data fusion. Volume 39 of Series B: Mathematical and Statistical Methods. Dordrecht: Kluwer, 1997.

[6] Gordon, N., Salmond, D., and Smith, A. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings*, 140:107–113, 1993.

[7] Green, P.J. Reversible jump MCMC computation and Bayesian model determination, *Biometrika*, 82:711–732, 1995.

[8] Hough, P.V.C. Method and means for recognizing complex patterns. US Patent 3069654, 1962.

[9] Hue, C., Le Cadre, J.-P., and Pérez, P. Sequential Monte Carlo methods for multiple target tracking and data fusion. *IEEE Transactions on Signal Processing*, 50:309–325, 2002.

[10] Illingworth, J. and Kittler, J. A survey of the Hough transform. *Computer Vision, Graphics and Image Processing*, 44:87–116, 1988.

[11] Kalman, R. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82:35–45, 1960.

[12] Khan, Z., Balch, T., and Dellaert, F. MCMC-based particle filtering for tracking a variable number of interacting targets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27:1805–1819, 2005.

[13] Lacoste, C., Descombes, X., and Zerubia, J. Point processes for unsupervised line network extraction in remote sensing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27:1568-1579, 2005.

[14] M.N.M. van Lieshout. Markovianity in space and time. In *Dynamics & stochastics: Festschrift in honour of M.S. Keane*, D. Denteneer, F. den Hollander, and E. Verbitskiy (Eds.), Lecture Notes - Monograph Series, Volume 48, pp. 154–168. Beachwood: Institute for Mathematical Statistics, 2006.

[15] M.N.M. van Lieshout. Campbell and moment measures for finite sequential spatial processes. In *Proceedings Prague Stochastics 2006*, M. Hušková and M. Janžura (Eds.), pp. 215–224. Prague: Matfyzpress, 2006.

[16] Lund, J., Penttinen, A., and Rudemo, M. Bayesian analysis of spatial point patterns from noisy observations. Research Report, Department of Mathematics and Physics, The Royal Veterinary and Agricultural University, Copenhagen, 1999.

[17] Mardia, K.V., Qian, W., Shah, D., and Desouza, K.M.A. Deformable template recognition of multiple occluded objects. *IEEE Proceedings on Pattern Analysis and Machine Intelligence*, 19:1035–1042, 1997.

[18] Matheron, G. *Random sets and integral geometry.* New York: John Wiley and Sons, 1975.

[19] Molina,R. and Ripley, B.D. Using spatial models as priors in astronomical image analysis. *Journal of Applied Statistics*, 16:193–206, 1989.

[20] Ripley, B.D. and Sutherland, A.I. Finding spiral structures in images of galaxies. *Philosophical Transactions of the Royal Society of London, Series A*, 332:477–485, 1990.

[21] Robert, C.P. and Casella, G. *Monte Carlo statistical methods.* New York: Springer, 1999.

[22] Rue, H. and Hurn, M.A. Bayesian object identification. *Biometrika*, 86:649–660, 1999.

[23] Stoica, R., Descombes, X., and Zerubia, J. A Gibbs point process for road extraction in remotely sensed images. *International Journal of Computer Vision*, 57:121-136, 2004.

[24] Stone, L.D., Barlow, C.A., and Corwin, T.L. *Bayesian multiple target tracking.* Norwood: Artech House, 1999.

[25] Vihola, M. *Random sets for multitarget tracking and data fusion.* Licentiate Thesis, Tampere University of Technology, 2004.

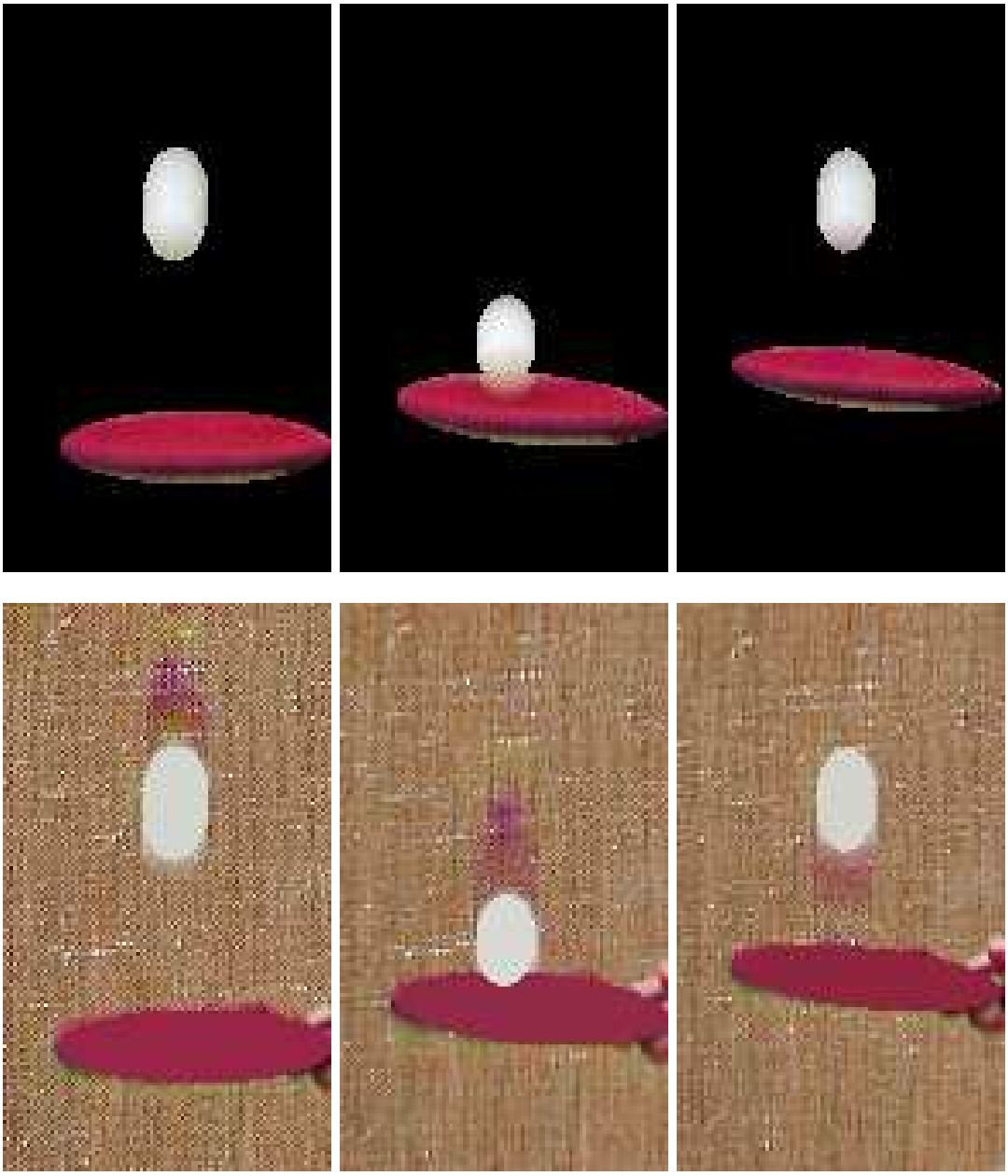

Figure 1: Data masked by annealed object sequence (top) and annealed object sequence overlaid upon the data (bottom).

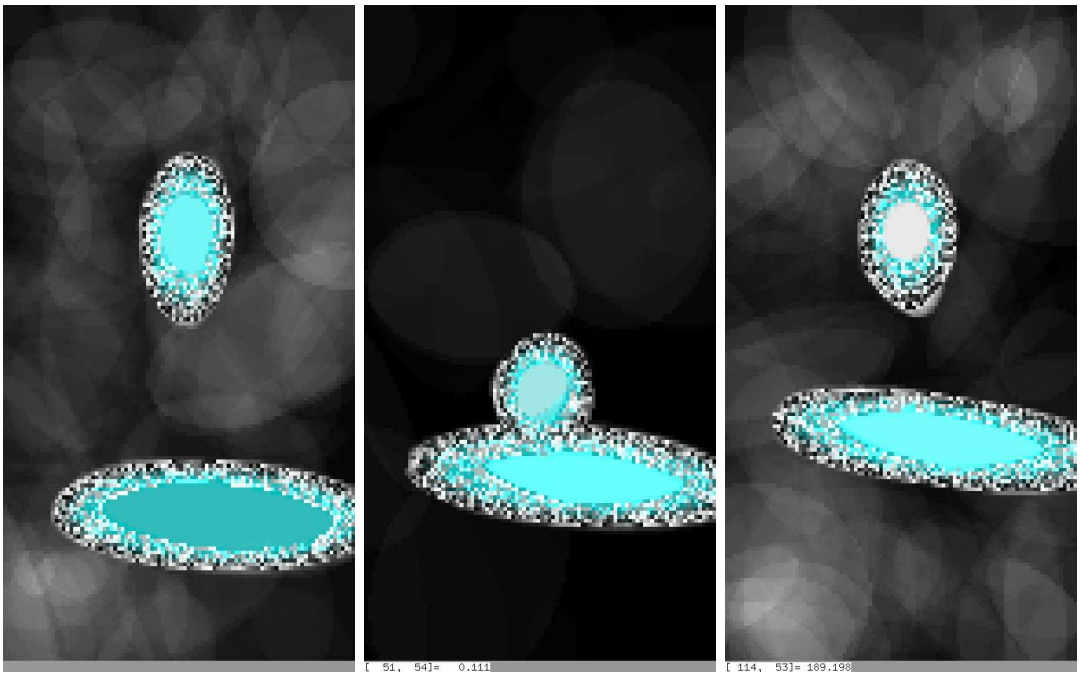

Figure 2: Depth map averaged over 300,000 Metropolis–Hastings steps after a burn-in of $50,000$ steps at temperature 5.