Modeling Textured Motion : Particle, Wave and Sketch

Yizhou Wang

Computer Science Department
University of California, Los Angeles
Los Angeles, CA 90095

Song Chun Zhu

Department of Statistics

Abstract

In this paper, we present a generative model for textured motion phenomena, such as falling snow, wavy river and dancing grass, etc. Firstly, we represent an image as a linear superposition of image bases selected from a generic and over-complete dictionary. The dictionary contains Gabor bases for point/particle elements and Fourier bases for wave-elements. These bases compete to explain the input images. The transform from a raw image to a base or a token representation leads to large dimension reduction. Secondly, we introduce a unified motion equation to characterize the motion of these bases and the interactions between waves and particles, e.g. a ball floating on water. We use statistical learning algorithm to identify the structure of moving objects and their trajectories. Then novel sequences can be synthesized easily from the motion and image models. Thirdly, we replace the dictionary of Gabor and Fourier bases with symbolic sketches (also bases). With the same image and motion model, we can render realistic and stylish cartoon animation. In our view, cartoon and sketch are symbolic visualization of the inner representation for visual perception. The success of the cartoon animation, in turn, suggests that our image and motion models capture the essence of visual perception of textured motion.

1 Introduction

Natural scenes contain rich stochastic motion patterns which are characterized by the movement of a large number of distinguishable or indistinguishable elements, such as falling snow, flock of birds, river waves, etc. These motion patterns, called textured motion [16], temporal texture [14] and dynamic textures[13] in the literature, cannot be analyzed by conventional optical flow fields [6] and have stimulated growing interests in both graphics and vision.

In graphics, the objective is to render photorealistic video sequences or non-photorealistic but stylish cartoon animations. Both physics-based[10] methods, such as partial differential equations, and image-based, such as video texture[12] and volume texture [17], are studied to simulate fire, fluid and gaseous phenomena.

In vision, Szummer and Picard studied a spatial-temporal auto-regression (STAR) model [14], which is a causal Gaussian Markov random field model (GMRF) in space and time. Bar-Joseph *et. al.* extended 2D texture synthesis work to generate volume texture using a tree structured representation [1]. Soatto *et. al.* [13] represented an image by linear combination of principal components which are computed from the training sequence. Then an auto-regression (AR) model was applied to the coefficients of those principal components for motion dynamics. Fitzgibbon [3] used a similar image representation with a global camera motion component to analyze textured motions for registration purpose.

In the above works, the basic moving elements are either represented by pixels or by the entire image or its principle components. Such representations are either too local or too global to capture the semantic structures of the objects in the video sequences, despite their success in synthesis. Recently, Wang and Zhu [16] proposed a generative model to represent an image with a number of Gabor bases. They computed the moving objects, like snow flakes, by grouping spatially close and temporally consistent trajectories of the bases to form what they called the "movetons". This model, however, is inefficient to represent motion with indistinguishable elements, such as water waves. It is generally considered challenging, in both vision and graphics, to model the interactions between particle objects and waves, for example, a ball or a boat on a river.

Motivated by these observations, this paper presents a general representation for textured motion in four aspects.

1. Photometric model. An image is represented as a superposition of bases from an over-complete dictionary, including Fourier bases and Gabor sin/cos bases at different scales, orientations. As Fig.1 shows, the Gabor and Fourier bases are selected through a competition and explain-away mechanism and are effective for particle and wave patterns, respectively. It is shown in Table 1 that large dimension reduction is achieved after transforming a raw image to a

Model	Parameters to Remember	Compression Ratio
Training Sequence	$150 \times 200(I) \times 100(nfrm) = 3 \times 10^6$	NA
Video Textures	$150 \times 200(I) \times 100(nfrm) = 3 \times 10^6$	$1:1$
Dynamic Textures	$150 \times 200(I) + 150 \times 200 \times 20(\text{PCA}) + 20 \times 20(A) + 20(\sigma) \approx 6.3 \times 10^5$	$1:5$
Textured Motion	$10^3(\text{magn}) + 10^3(\text{phase}) \times 8(\text{Cov}) + 10^3(\sigma) + [20(p) + 20(\gamma) + 1(\sigma_\gamma)] \times 8 \approx 10^4$	$1:300$

Table 1: Comparison of the compress ratios among 3 typical models for wavy river sequence.

token representation.

2. Geometric model. Each object in the scene is represented by a number of bases with deformable structures. For instance, a ball is represented by a few Gabor bases moving together and a river is represented by a number of Fourier bases with displacements in phases.

3. Dynamic model. We adopt a general motion equation which includes an AR component for each base, external forces and the interactions with other bases. For example, a ball (Gabor bases) on a river is driven by wind and water waves (Fourier bases).

4. Sketch model. We replace the dictionary of Gabor and Fourier bases with sketches (symbolic tokens), thus change the photometric model to a sketch model. Together with the same motion model, we can render non-photorealistic and stylish cartoon animation. In our view, cartoon and sketch are symbolic visualization of the inner representation for visual perception. The success of the cartoon animation, in turn, suggests that our representation captures the essence of visual perception of textured motion.

In summary, our representation is much more parsimonious compared with other models. Table. 1 lists the compression rates of each model for a wavy river sequence. The training sequence is 100-frame long and each frame has 150×200-pixels. The video texture method [12] stores the entire sequence, and synthesizes a new sequence by reordering the training frames to achieve smooth transition. Dynamic textures [13] characterizes the stochastic process by remembering a number of parameters, including 1 mean image, 20 principle components of the frames, a dynamics matrix A and 20 noise terms. Therefore, the model achieves better compact rate of about $1:5$. Our model uses about 1000 Fourier bases with 1000 magnitudes and 1000 phases to represent the image without noticeable loss, and the dynamics are fitted by a 20th (p) order AR model on the coefficients with some noise terms. The compression rate is about $1:300$, due to the use of a generic dictionary.

In the following of the paper, we present four models sequentially – photometric, geometric, dynamic, and sketch. A number of synthesized movies and cartoon animation are shown as results, which are better evaluated from the supplementary file.

2 Textured motion representation

Let $\mathbf{I}[0,\tau]$ denote an image sequence on a 2D lattice Λ in a discretized and time interval $[0,\tau] = \{0,1,...,\tau\}$. For $t \in [0,\tau]$, $\mathbf{I}(u,v)$ or $\mathbf{I}(u,v,t) \in \mathbf{I}[0,\tau]$ denotes a pixel on a frame.

2.1 Photometric model– particles vs waves

There are two general image coding paradigms [2] in the literature: compact coding and sparse coding. Compact coding uses a complete basis/dictionary or tight frame, such as Fourier transform and wavelets. In this scheme, an input signal is represented by a combination of all bases in the dictionary. Sparse coding uses an over-complete dictionary, e.g. Gabor bases, LoG (Laplacian of Gaussian) bases. As the basis is over complete, an input signal is represented by a very small population of the bases which are experts for the input signal. This leads to an effective representation with large dimension reduction.

We employ the over-complete dictionary Δ with both Gabor and LoG bases Δ_{pcl} to represent point like objects – particles and Fourier bases Δ_{wav} for wave patterns.

$$\Delta = \Delta_{\text{pcl}} \cup \Delta_{\text{wav}}.$$

It is known what the Gabor bases are specified by variables $\beta = (x, y, \sigma, \theta)$ for position, scale and orientation, the LoG bases are isotropic with variables $\beta = (x, y, \sigma)$ and Fourier bases are defined by spatial frequency and phase $\beta = (\xi, \eta, \phi)$. So we obtain

$$\begin{aligned} \Delta_{\text{pcl}} &= \{\text{Gcos}(u,v;\beta),\ \text{Gsin}(u,v;\beta),\ \text{LoG}(u,v;\beta) :\ \forall\beta\}, \\ \Delta_{\text{wav}} &= \{\text{FB}(u,v;\beta) :\ \forall\beta\}. \end{aligned}$$

An image $\mathbf{I}$ is a superposition of N image bases.

$$\mathbf{I} = \sum_{j=1}^{N} \alpha_j \psi_j + \mathbf{n}, \quad \psi_j \in \Delta, \tag{1}$$

where α_j is a coefficient of base ψ_j, $\mathbf{n}$ is a noise process for the residues. In general, we have

$$|\Delta| = O(100\,|\Lambda|), \text{ and } N = O(|\Lambda|/100).$$

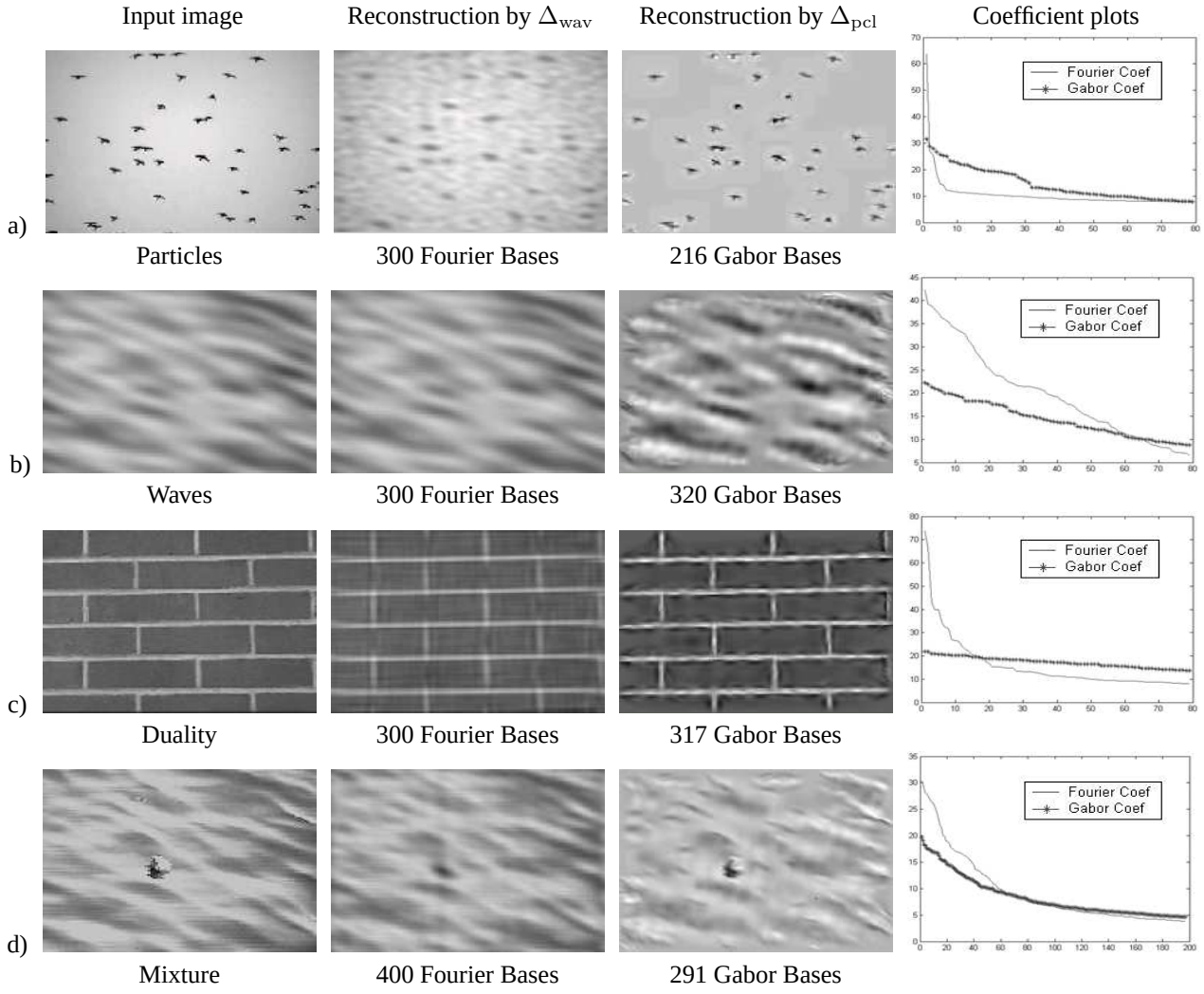


Figure 1: Image reconstructions by Fourier bases Δ_{wav} and Gabor/LoG bases Δ_{pcl}. The curves plot the coefficients in the match pursuit process. The thick curve is for Δ_{pcl}.

Thus a raw image is transformed into a parsimonious token representation, called *base map*.

$$\mathbf{B} = \{\mathbf{b}_j = (\alpha_j, \beta_j),\ j = 1, 2, ..., N\}. \tag{2}$$

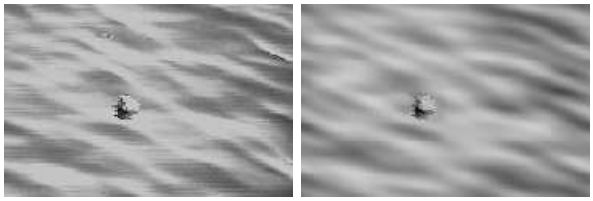
Furthermore, we divide the base map into a particle map $\mathbf{B}_{\text{pcl}}$ and a wave map $\mathbf{B}_{\text{wav}}$.

$$\mathbf{B} = \mathbf{B}_{\text{pcl}} \cup \mathbf{B}_{\text{wav}}.$$

Similarly, we transform an image sequence $\mathbf{I}[0, \tau]$ into a token representation $\mathbf{B}[0, \tau]$ where each base $\mathbf{b}_j(t)$ is tracked frame by frame. It is worth mentioning that each Fourier base *FB* can move only in one dimension [4], thus the 2D spatial velocity is transformed to a 1D phase speed

$$\frac{\phi_j(t)}{dt} = \xi_j \frac{dx}{dt} + \eta_j \frac{dy}{dt}. \tag{3}$$

Figures 1 and 2 compare the particle bases Δ_{pcl} and the wave bases Δ_{wav} by representing different textured motion patterns. We select four typical images for illustration. From each image, we obtain two reconstructions: one by wave (Fourier) bases Δ_{wav} and the other by particle (Gabor and LoG) bases Δ_{pcl}. We select the bases from each dictionary using a match pursuit procedure[7]. This is a greedy algorithm that picks a base which has the highest response on the current residue image. So at each step, it reduces the reconstruction error in a steep descent way. We plot the coefficients $\alpha_j, (j = 1, 2, ..., N)$ of the selected bases from Δ_{wav} and Δ_{pcl}. The slopes of the curves reflect the coding efficiencies of the dictionary. The steeply decreasing curve implies that the bases are very effective in reconstructing the image and thus capture the essential objects in the im-



Input image 80 Fourier + 21 Gabor

Figure 2: Reconstruct floating-ball image by mixture of Fourier and Gabor.

age, whereas a flat curve means the opposite.

Fig. 1. a) is a flock of birds. In contrast with the roughly reconstructed image by $N = 300$ Fourier bases, the image with $N = 216$ particle bases is reconstructed very well. The curve plot shows that the first few Fourier bases have large responses capturing the global lighting effects in the sky. Therefore, the best representation for this image is a few Fourier bases for lighting plus the particle bases for individual birds. Fig. 1.b) is a water wave image. The Fourier bases are obviously better than the particle bases. Both the reconstructed images and the two curves serve as the evidence. Fig.1.c) is a brick image. It has both periodic global structures and local high contrast features. Given a small size of the image, the two dictionaries explain it about equally well. Finally, Fig.1.d) shows a ball floating on river. We can see that neither type of bases alone is able to effectively represent this image. However, Fig.2 exhibits a better reconstruction using a combination of 80 Fourier bases and 21 Gabor bases.

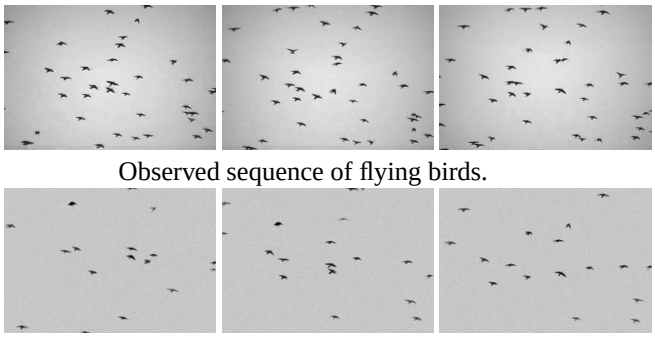
These examples demonstrate that textured motion sequences have both wave and particle patterns, and the combined dictionary representation is both concise and meaningful.

2.2 Geometric model – the moving elements

In the photometric model, as the dictionary Δ is generic, a particle object, like a bird or ball, is represented by a few bases moving together with closely tangled trajectories. It is also the case for the Fourier bases for waves as they also travel in groups [15]. The water waves that we observe are summation of travelling sinusoid waves caused by different sources of motion, such as wind, boat, earthquake, etc. Therefore, the moving elements are represented by a number of bases (n_b) with some deformable configuration in space and phase domain.

$$\pi = (n_b, \{\mathbf{b}_j = (\alpha_j, \Theta_j), j = 1, 2, \ldots, n_b\})$$

Furthermore, as each moving element has a lifespan $[t^b, t^e] \subset [0, \tau]$ [16], e.g., a bird can fly into our view at time t^b and out of our view at time t^e, we represent the moving



Observed sequence of flying birds.

Synthesized sequence with fewer flying birds to show video editing effect.

Figure 3: Example of modeling and synthesizing a flying-bird sequence.

object in space, frequency and time by a representation

$$\mathcal{C}[t^b, t^e] = (\pi(t^b), \pi(t^{b+1}), ..., \pi(t^e)).$$

Thus we obtain a meaningful semantic representation of the textured motion in terms of the moving elements, their deformation, and trajectories.

$$\mathbf{W} = (K, \{\mathcal{C}_i[t_i^b, t_i^e], i = 1, 2, ..., K\}),$$

where K is the number of objects in the sequence.

In summary, we have the following generative model for an image $\mathbf{I}$,

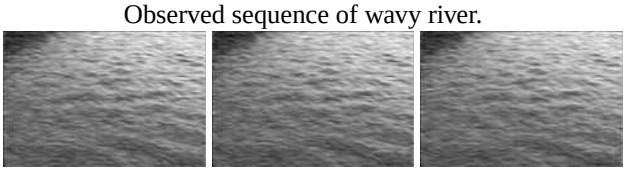
$$\mathbf{W} \xrightarrow{\Phi} \{\mathbf{B}_{\text{wav}}, \mathbf{B}_{\text{pcl}}\} \xrightarrow{\triangle} \mathbf{I}.$$

This representation is not only low-dimensional and generic, but also captures the essence of visual perception of textured motion. In Section 4, we use this generative model to synthesize cartoon animation by only replacing the bases $\mathbf{B}$ and moving elements π with a symbolic representation.

2.3 Dynamic model – the interactions

In this section, we present a dynamic model for the motion of moving elements and especially their interactions. We are especially interested in two types of interactions. (1). The influence of waves on particles. Particle elements in textured motion are often driven by waves, for example, a ball floating on a wavy river. This kind of effect is previously hard to simulate [12, 13]. (2). The interactions among wave components. The relative motion of different Fourier bases must be constrained to keep certain phase alignments. Other interactions, such as particle collision, particle-wave collision (splash) are not considered in this paper.

Observed sequence of wavy river.



Synthesized sequence of wavy river.

Figure 4: Results of river sequence.

Observed sequence of grassland.

Synthesized sequence of grassland.

Figure 5: Results of grassland sequence.

Let π be the motion status (e.g. $\pi = x$ is the position for particles, and $\pi = \phi$ is the phase for Fourier bases). The motion equation is a p-th order AR model with coefficients α, driven by a simple Brownian motion n and a force $U(\mathbf{B}_{\text{wav}}(t), \mathbf{B}_{\text{pcl}}(t))$ from other bases.

$$\pi(t) = \sum_{j=1}^{p} \alpha_j \pi(t-j) + U(\mathbf{B}_{\text{wav}}(t), \mathbf{B}_{\text{pcl}}(t)) + c(t) + n \quad (4)$$

This general motion equation is reduced to four categories of special cases in the rest of this section.

Case 1: Dynamic model for free moving particles - 0D case, e.g. falling snow, flying birds.

In this case, the location $x(t)$ of a particle is the status of the object $\pi(t)$ in Eq.4. By assuming there is minimum interaction among particles, we obtain its motion equation as a degenerated case from Eq.4.

$$x(t) = \sum_{j=1}^{p} \alpha_j x(t-j) + c + n,$$

where, the external force field $c(t)$ is assumed to be spatially and temporally constant.

Case 2: Dynamic model for waves - 2D case, e.g. wavy river.

In this case, Fourier bases are selected to represent the image sequence. Their motion is characterized by phase change $\phi(t)$ in frequency domain [4]. Eq.3 shows that if a wave travels at a constant velocity, ϕ_j is a line in phase-temporal domain. Once wrapped into $[0, 2\pi)$, it appears to be periodic.

We model the motion of the Fourier bases by a Gaussian Markov Random Field (GMRF). AR model is applied to the coefficients of eigen-phase vectors which are extracted from the covariance matrix of phases to estimate wave motion. Thus the $\pi(t)$ in Eq.4 is the coefficient of the principle components. As another special case of the general motion equation, the dynamics of which is shown as follows.

$$\begin{cases} \phi(t) = \sum_{j=1}^{m} \gamma_j(t) \varphi_j + n_\phi \\ \gamma(t) = \sum_{j=1}^{q} \lambda_j \gamma(t-j) + n_\gamma \end{cases}$$

where φ_j is the jth eigen-phase vector with coefficient $\gamma_j(t)$ at frame t, m is the number of eigen-phase vectors, q is the order of AR model, and λ is the AR coefficient for the eigen-phase coefficients , n_ϕ and n_γ are the noise of ϕ and γ respectively, which are both assumed to be Gaussian.

Case 3: Dynamic model for particles & waves - 0D-2D case, e.g. floating ball or foams.

In this case, not only should particles motion in space and waves motion in phase domain be modelled, but also their interactions. Here we assume waves have more influence on particles instead of vice versa.

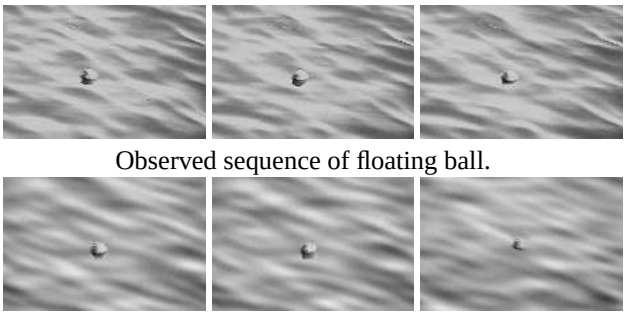
$$\begin{cases} x(t) = \sum_{j=1}^{p} \alpha_j x(t-j) + \beta \Delta\phi(t) + c + n \\ \phi(t) = \sum_{j=1}^{m} \gamma_j(t) \varphi_j + n_\phi \\ \gamma(t) = \sum_{j=1}^{q} \lambda_j \gamma(t-j) + n_\gamma \end{cases}$$

Similar to the previous 2 cases, the first equation models the motion of particles, and the rest equations are about waves. A regression model $\beta\Delta\phi(t)$ is introduced to describe the influence of the waves on particles. $\Delta\phi(t)$ is the phase shift between frame t and $t-1$, β is the regression coefficient. After we combine the two motions, the model becomes a coupled Gaussian Markov Random Field (GMRF).

Case 4: Dynamic model for splines & waves - 1D-2D case, e.g. dancing grass.

In this case, particles are elongated and elastic like springs and are densely clustering together. Since the motion of each grass is heavily constrained by a large number of its neighbors, the group action becomes obvious. Globally, you will see the wave pattern of a grassland. The motion equations are derived as follows.

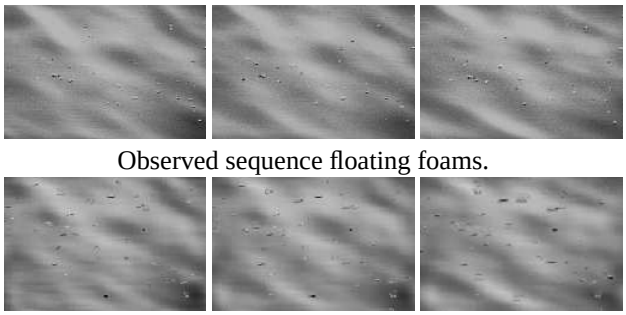
$$\begin{cases} x(t) = & \sum_{j=1}^{p} \alpha_j x(t-j) + \kappa(x - x_0) \\ & + \beta \sin[< \omega_x, x(t) > + \phi(t)] + n \\ \phi(t) = & \sum_{j=1}^{m} \gamma_j(t) \varphi_j + n_\phi \\ \gamma(t) = & \sum_{j=1}^{q} \lambda_j \gamma(t-j) + n_\gamma \end{cases}$$



Observed sequence of floating ball.

Synthesized sequence of floating ball.

Figure 6: Learned results from floating-ball sequence.



Observed sequence floating foams.

Synthesized sequence floating foams.

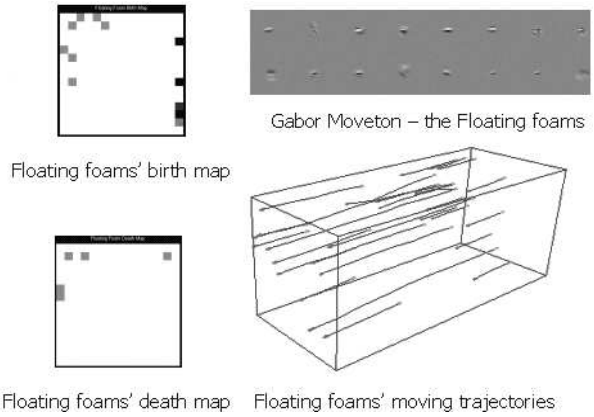


Figure 7: Learned results from floating-foam sequence.

where κ describes the elasticity of a grass, of which the free position of its tip is x_0. To synthesize the photo-realistic image, we use the method as described in Case 3. To generate cartoon animation, we manually track several grass tips at different location as particles. A regression model is applied to describe the influence of a group of grasses on a single grass, which is a function of both time and spatial location. Thus we introduce $\sin[< \omega_x, x(t) > +\phi(t)]$ to govern the variation in time and space, where ω_x is the spatial frequency of the grassland. Then hundreds of new grasses are produced around those tracked grasses. They will follow the motion of their nearest tracked neighbor. Finally, spline curves are connected between the tips and the corresponding roots of grasses as sketches of the grassland.

3 Learning and inference

Given an input sequence $\mathbf{I}^{\text{obs}}_{[0,\tau]}$, our objective is to learn both geometric and dynamic models. The geometric model parameterized by Φ identifies the moving elements in the sequence. The dynamic model specified by parameters Γ characterizes the motion of the moving elements. To learn these models, we should also compute the hidden variables $\mathbf{W}$ and $\mathbf{B}$ from $\mathbf{I}^{\text{obs}}_{[0,\tau]}$ – the inner representation.

Based on the models and analysis described above, we divide our learning scheme into three parts. 1)"Particle learning" in the spatial-temporal domain. It includes computing the forces created by the waves, which is resolved by estimating the regression coefficient β in Case 3 & 4. 2)"Wave learning" in frequency-temporal domain. 3) Fuse waves and particles in the image domain. As the particles displacement will be influenced by the phases of the waves while the phases are also affected by the identified particles, the "particle learning" and "wave learning" processes are inseparable.

The problem is posed as statistical learning by maximum likelihood estimate (MLE). The log-likelihood function for an observed training sequence $\mathbf{I}^{obs}_{[0,\tau]}$ is

$$\mathcal{L}(\Theta) = \log p(\mathbf{I}^{obs}_{[0,\tau]}; \Theta) = \log \int p(I|W;\Phi)p(W;\Gamma)dW$$

where $\Theta = (\Phi, \Gamma)$ is the parameter governing the textured motion patterns, and W denotes the hidden variables related to the specific sequence $\mathbf{I}^{\text{obs}}_{[0,\tau]}$. The goal of learning is to estimate Θ with maximum likelihood,

$$\Theta^* = \arg\max \mathcal{L}(\Theta).$$

To solve the MLE in the above equation, we set $\frac{\partial \mathcal{L}(\Theta)}{\partial \Theta} = 0$.

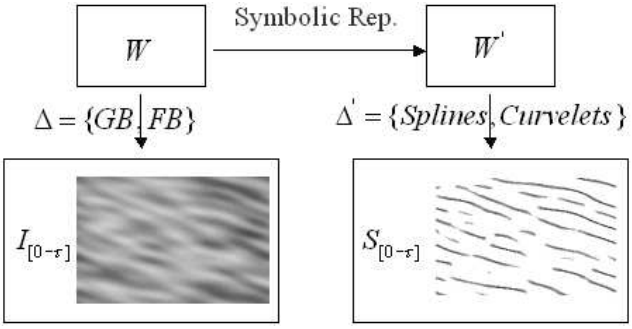


Figure 8: From photorealistic to semantic representation.

It follows that

$$\int [\frac{\partial \log p(\mathbf{I}^{obs}|W;\Phi)}{\partial \Phi} + \frac{\partial \log p(W;\Gamma)}{\partial \Gamma}] p(W|\mathbf{I}^{obs};\Phi,\Gamma) dW = 0.$$

We adopt the stochastic gradient algorithm used in [5]. The learning iterates in three steps.

1. Sampling $W^{syn}_{[0,\tau]} \sim p(W|\mathbf{I}^{obs};\Phi,\Gamma)$, which includes two parts. Part 1, Computing particle bases forming $\mathbf{B}_{\rm pcl}$, grouping bases into movetons and tracking the movetons. The computation is realized by Markov Chain Monte Carlo (MCMC) techniques. Part 2, Computing wave bases forming $\mathbf{B}_{\rm wav}$. This is done by Fourier transformation on the remaining images after sift out the particles.

2. Update the motion dynamics parameter Γ for both particles and waves at each step s.

$$\Gamma(s+1) = (1-\rho)\Gamma(s) + \rho \frac{\partial \log p(W^{syn}_{[0,\tau]};\Gamma)}{\partial \Gamma},$$

3. Update the moveton parameter Φ by clustering and grouping. In this step, only the particles moveton is concerned, because we assume each Fourier base forms a moveton by itself.

$$\Phi(s+1) = (1-\rho)\Phi(s) + \rho \frac{\partial \log p(\mathbf{I}^{obs}|W^{syn};\Phi)}{\partial \Phi}.$$

Some learned results are shown in Fig.4, 5, 6, 7. And more results can be seen from the attached video clips.

4 Sketch model

In this section, we present a sketch model and a novel way to render cartoon animation from the input and synthesized sequences.

In our view, cartoon is a simplified and symbolic visualization of our inner representation W. Suppose we have computed $W_{[0,\tau]}$ from the input sequence $\mathbf{I}^{\rm obs}_{[0,\tau]}$ or $W_{[0,\tau]}$ has been synthesized from the learned model Θ. In photorealistic rendering, we use the dictionaries $\Delta_{\rm pcl}$ and $\Delta_{\rm wav}$ to render a motion sequence

$$W_{[0,\tau]} \overset{\Delta_{\rm pcl},\, \Delta_{\rm wav}}{\longrightarrow} \mathbf{I}_{[0,\tau]}.$$

To render cartoon animation, it takes two steps. Firstly, we extract a subset of hidden variables $W'_{[0,\tau]}$, which is supposed to capture the essential semantics, from $W_{[0,\tau]}$ to simplify the description. For example, let $\pi \in W$ be a snow flake, we represent its geometric shape in the cartoon but ignore its photometric properties. Secondly, we replace the two dictionaries $\Delta_{\rm pcl}$ and $\Delta_{\rm wav}$ by symbolic bases $\Delta'_{\rm pcl}$ and $\Delta'_{\rm wav}$, respectively to render cartoon animation $S_{[0,\tau]}$ using the generative model:

$$W'_{[0,\tau]} \overset{\Delta'_{\rm pcl},\, \Delta'_{\rm wav}}{\longrightarrow} S_{[0,\tau]}.$$

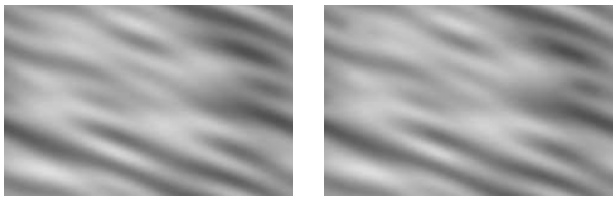
where $S_{[0,\tau]}$ is a sketch for an observed or newly synthesized sequence. This procedure is illustrated in Fig.8 and 10. The animated cartoon is attached to the supplementary file.

Now we briefly explain how we choose a symbolic representation for particles and waves. It is obvious that $\Delta'_{\rm pcl}$ and $\Delta'_{\rm wav}$ represent the style of the cartoon. Therefore, (1). We render a particle element π by a contour outline for birds and by spline curves for grass. (2). As we cannot sketch each individual Fourier base for waves, we combine all the Fourier bases to generate a wave function instead:

$$\mathbf{J} = \sum_{j=1}^{n} \alpha_j \psi_j, \quad \psi_j \in \Delta_{\rm wav}.$$

We sketch $\mathbf{J}$ by spline curves at the ridges and valleys $\nabla^2 \mathbf{J} = 0$. Fig.8 shows an example.

We claim that wave sketch is a almost sufficient semantic image representation as we can recover the original image from those extracted sketches. Fig.9 shows an example for the river image. For each point (x, y) on the symbolic sketch, i.e. $\nabla^2 \mathbf{J}(x, y) = 0$, we remember the pixel intensity $\mathbf{J}(x, y)$ and the slope $\nabla J(x, y)$. Then we interpolate



a)Observed image b)Interpolated image

Figure 9: Image interpolation from extracted sketches.

the rest of the image by spline or simple heat diffusion using the sketch as boundary condition. This is related to Marr's conjecture of the primal sketch. Marr conjectured that the zero-crossing and their slope are sufficient for recovering band-pass filtered image.

Fig. 10 shows a combined cartoon animation. We choose three natural sequences: flying birds, floating ball on a river and wavy grassland, and learn the geometric and dynamic models for each of them. Then we render synthesized sequences and generate their cartoons using the sketch model (The floating ball is replaced by a boat). A static background – mountain, sun, and river bank is drawn manually (Fig. 10.a). Finally, we fill the three cartoons into the blank areas of the background image.

5 Summary and future work

In this paper, we presented a generative model for textured motion and introduced an image representation scheme using over-complete generic basis to model natural images containing local particles and global wave patterns. A general motion equation is derived to characterize the interaction of waves and particles. A sketch model for rendering non-photorealistic sequences from the learned geometric and dynamic models is also presented.

In the future, we would extend this work by (1) modelling the interaction among particles, e.g. collision, (2) studying the influence of particles on waves, e.g. splash effect of a stone dropped into water.

References

[1] Z. Bar-Joseph, R. El-Yaniv, D. Lischinski, and M. Werman. "Texture mixing and texture movie synthesis using statistical learning", *IEEE Trans. on Vis. & Comp. Graph., 7, 2001.*

[2] D.J. Field, "What is the goal of sensory coding?", *Neural Computation, 6:559-601, 1994*

[3] A. W. Fitzgibbon, "Stochastic ridigity: image registration for nowhere-static senes", *Proc. ICCV, pp 662-669, July 2001.*

[4] D.J. Fleet, A.D. Jepson, "Stability of phase information", *IEEE Trans. on PAMI, 15(12):1253-1268, 1993*

[5] M. Gu, "A stochastic approximation algorithm with MCMC method for incomplete data estimation problems", *Preprint, Dept. of Stat. McGill Univ. 1998*

[6] R. Mann and M. S. Langer, "Optical snow and the aperture problem", *ICPR, Vol. IV, pp.264-7, Aug. 2002.*

[7] S. Mallat and Z. Zhang, "Matching Pursuit in a Time-Frequency Dictionary", *IEEE Trans. on Sigmal Processing*, Vol. 41, 3397-3415, 1993.

[8] D. Marr, "Vision", *IW.H. Freeman, 1983*

[9] B.A. Olshausen, D.J. Field, "Sparse coding with an over-complete basis set: A strategy employed by V1?", *Vision Research, 37:3311-3325, 1997.*

[10] C. H. Perry and R. W. Picard, "Synthesizing Flames and Their Spreading," *Proc. of the 5th Eurographics Workshop on Animation and Simulation, Oslo, Norway, Sept. 1994.*

[11] P. Saisan, G. Doretto, Y. Wu, and S. Soatto, "Dynamic Texture Recognition," *CVPR, 2001.*

[12] A. Schodl, R. Szeliski, D. Salesin and I. Essa, "Video Textures", *SIGGRAPH, 2000.*

[13] S. Soatto, G. Doretto, and Y. N. Wu, "Dynamic Texture", *ICCV, 2001*

[14] M. Szummer and R. W. Picard, "Temporal texture modeling" , *Int'l. Conf. on Image Proc., Vol. 3, pp. 823-826, Sept. 1996.*

[15] R.A.R. Tricker, *Bores, Breakers, Waves and Wakes*, American Elsevier, New York, 1965.

[16] Y. Wang, S. Zhu, "A Generative Method for Textured Motion: Analysis and Synthesis", *ECCV, 2002*.

[17] L.Y. Wei and M. Levoy, "Fast Texture Synthesis using Tree-structured Vector Quantization", *SIGGRAPH, 2000.*

[18] S. Zhu, C. Guo, Y. Wu, and Y. Wang, "What are Textons?", *ECCV, 2002*.

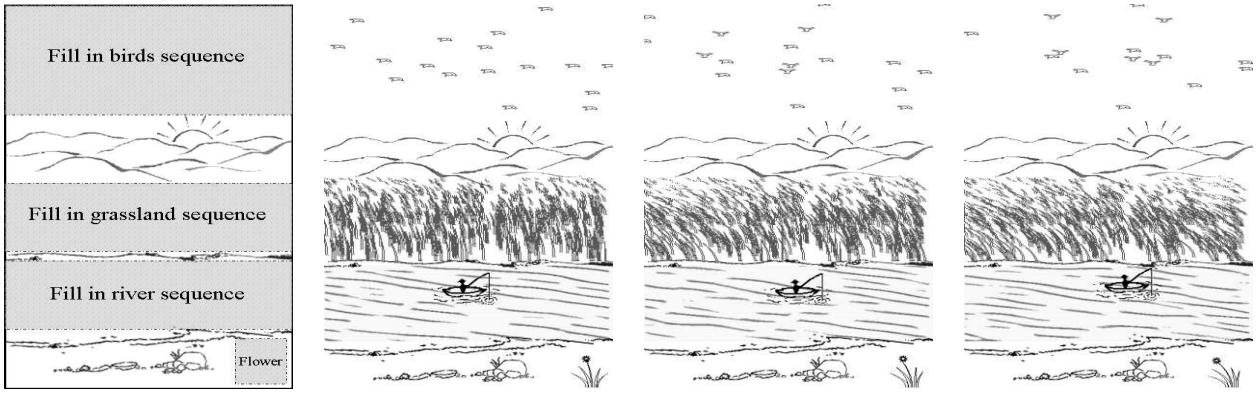


Figure 10: Synthesized cartoon sequence based on learned textured motions.