Motion Texture
A Two-level Statistical Model for Motion Synthesis

Yan Li      Tianshu Wang      Heung-Yeung Shum
Microsoft Research, Asia

Motion capture data re-use
- Motion editing
  - Interactive motion editing
    - [Bruderlin’95, Witkin’95, Lee’99, Pullen’02]
  - Adaptation to new characters and environments
    - [Hodgins’95, Gleicher’98, Popovic’99]
- Motion synthesis
  - Retain the realism of original captured data
  - Allow the user to control and direct the character

Motion synthesis
- Reordering the original data
  - Chop into motion clips & model their transitions
    - [Kovar’02, Lee’02, Arikan’02] Video Texture
  - Problems
    - No generalization ability
    - Difficult to edit at the frame level
- Generative models
  - Hidden Markov Model (HMM)
    - [Brand’00, Tanco’00]
  - Auto-regressive model for simple movements
    - [Pavlovic’00] BLDS, [Soatto’01] Dynamic Texture

Motion texture
Representation
- A two-level statistical model
- Motion texton — Linear dynamic system (LDS)
- Texton distribution — Markov process

Analogous to 2D texture image
- Textons (Julesz’81)
- 2D spatial distribution of textons

Motion texture analysis/synthesis
Challenges
- How to learn the motion texture
- How to synthesize from motion texture
- How to deal with high dimensional motion data
Motion data representation
- Configuration of articulated characters
  - Linear components: \( p(t) \in \mathbb{R}^l \)
  - Angular components: \( q^i(t) \in S^i \)
- \( m(t) = \begin{pmatrix} p(t) \\ q^1(t) \\ \vdots \\ q^n(t) \end{pmatrix} \)
  - Position of the root joint
  - Orientation of the root joint
  - Orientations of the body joint

Dealing with motion data

Motion texture
- A graphical model representation
- Markov model: \( P(x_{ij} | x_i) \)
- Dynamics model: \( P(x_{ij+1} | x_i, x_j) \)
- Observation model: \( P(y_j | x_j) \)
- An auto-regressive moving average process (ARMA)

Linear dynamic system (LDS)
- Previous work
  - Tracking and gait recognition [Bregler’97, North’00, Pavlovic’00, Bissacco’01]
  - Video synthesis [Soatto’01, Fitzgibbon’01]
- Challenge for character motion synthesis
  - Captured motion is non-stationary
- Our solution
  - Segment-based stationary processes (LDS’)

Motion textons
- A set of basic motion elements
- Representation: LDS
- Dynamics model: \( X_{i+1} = A X_i + V_i \)
- Observation model: \( Y_i = C X_i + W_i \)
- \( \theta = (A, C, V, W) \)

Learning motion texture
- Estimation by EM algorithm (segmentation!)
  - E-step: how many segments and where they are
  - M-step: fitting LDS parameters \( \theta \) for the segments labeled by the same texton
- Transition graph (M): counting segment labels
- A maximum likelihood solution
  \[ \{\hat{\theta}, \hat{M}\} = \arg\max_{\theta, M} P(Y_{1:M} | \theta, M) \]
  \( \theta \): LDS parameters, \( M \): transition graph
Unsupervised learning

With the texton distribution
- Random walk
  - ABDCBDEC……
- Multiple paths between any pairs
  - \{ABDC\}, \{AC\}
- Add variations to the synthesized path
  - \{ABC\}

A two-step synthesis algorithm
- Texton path planning using DP
  - Finding the lowest cost path
  - Specifying the path length
- Synthesizing a single texton
  - Smooth transition between textons

Texton synthesis by constrained LDS
Sampling noise and incorporating both boundary constraints

Texton synthesis by sampling noise
Without end constraints

Video: synthesis and editing results
Synthesis with Motion Texture
- Single Texton
- Two Adjacent Textons
- Cyclic Motion
- Synthesizing Fine Details
Video: synthesized dance motion

Conclusion
- **Motion texture**
  - Two levels: textons and distribution
  - An unsupervised learning algorithm
- **Real-time synthesis**
  - Texton path planning
  - Constrained synthesis
  - Generate fine details
  - Interactive editing

Acknowledgements
- Dr. Ying-Qing Xu of MSRA
- Dr. Michael Cohen of MSR

Thanks!

hshum@microsoft.com

An example: dance motion

Problem
- We assume that motion textons are realizations of second-order stationary stochastic processes
- But in synthesis, the dynamics may deviate from the original one as time progresses
- Error accumulation in state space results in motion artifacts

Original | Synthesized
An illustration in the state space

\[ X_{i+1} = A_i X_i + V_i \]

Data transformation

Transformation between angular and linear signals

Learning motion texture

- A maximum likelihood solution:
  \[ (\theta, M) = \arg \max_{(\theta, M)} P(Y_{i+1} | 0, M) \]
  \[ (\theta : \text{LDS} \text{ parameters, } M : \text{transition matrix}) \]

- Introduce two hidden variables
  - \( L \) — segment labels
  - \( H \) — segment points

- Estimation by EM algorithm
  - E-step: infer \( L \) and \( H \), and estimate optimal state sequence
  - M-step: update \( \theta \) by fitting LDS', \( M \) by counting segment labels

Q/A: why not using derivative of rotation

- In synthesis, we would like to constrain the end pose of the motion instead of the end angular speed of it.
- We are assuming that the identity in the exp map is the key pose.
- This assumption may result in singularity.
- Fortunately, the rotation change in each texton is small (< π).

Q/A: Related work

- Motion texture
  [Bregler’00]
- Texton
  [Julesz’81, Malik’99, Zhu’02, Guo’01, Liang’01]
- Linear dynamic system
  [Bregler’97, Soatto’01, Fitzgibbon’01]
- Modeling nonlinear dynamics
  [Blake’02, Brand’00, Pavlovic’00]
Texton synthesis with constrained LDS

Learning Motion Texture

Motion Textons

Switched LDS

Motion Textures
Texton synthesis by sampling noise

Dynamics model:
\[ X_{t+1} = A_t X_t + V_t \]
Observation model:
\[ Y_t = C_t X_t + W_t \]

Sample the initial pose
Synthesize a frame by the observation model
Draw samples from the noise term \( V_t \)
Compute \( X_t \) by the dynamic model

An analogy to 2D texture images

- An image texture: two-dimensional spatial distribution
- In psychology, basic texture elements are called "texton" or "texel" [Julesz'81]
- In early vision, natural visual patterns consist of multiple layers of stochastic processes [Marr'82]

Patch-based sampling (Liang et.al. '01)

Patches + MRF

Integrating descriptive and generative models -- (Guo et.al. '01)

But what are image textons?

Goal: model mo-cap data

- Generalize the stochastic and dynamic nature
- Synthesize realistic character motion
- Edit at both the high and low levels