

Artistic Stylization of Images and Video

Part II – Vision for Stylisation

Eurographics 2011

John Collomosse

Centre for Vision Speech and Signal Processing (CVSSP),
University of Surrey, United Kingdom



- **Visual Interest and NPR: an Evaluation and Manifesto**
A. Santella and D. DeCarlo, NPAR 2004
- **Stylization and Abstraction of Photographs**
D. Decarlo, A. Santella, SIGGRAPH 2002
- **Segmentation-based 3D Artistic Rendering**
A. Kolliopoulos, J. Wang, A. Hertzmann, EGSR 2006.
- **Synergism in Low Level Vision (EDISON)**
C. Christoudias, B. Georgescu, P. Meer, ICPR 2002.
- **SIFT flow: dense correspondence across difference scenes**
C. Liu, J. Yuen, A. Torralba, J. Sivic, W. Freeman, ECCV 2008.
- **High Accuracy Optical Flow Estimation Based on a Theory for Warping**
T. Brox, A. Bruhn, N Papenberg, J. Weickert, ECCV 2004.
- **What dreams may come (movie)**
Dir. V. Ward. Universal. 1998.
- **Non-photorealistic Rendering SIGGRAPH Course notes**
D. Green, SIGGRAPH 1999
- **Processing Images and Video for Impressionist Effect**
P. Litwinowicz, SIGGRAPH 1997
- **Video Tooning**
J. Wang, Y. Xu, H. Shum, M. Cohen, SIGGRAPH 2004



- **Painterly Rendering for Video and Interaction**
A. Hertzmann, K. Perlin. NPAR 2000.
- **Painterly Rendering for Animation**
B. Meier. SIGGRAPH 1996
- **Image Analogies**
A. Hertzmann, C. Jacobs, N. Oliver, B. Curless, D. Salesin. SIGGRAPH 2001
- **Directional Texture Transfer**
H. Lee, S. Seo, S. Ryoo, K. Yoon. NPAR 2010.
- **Empathic Painting: Interactive stylization using observed emotional state**
M. Shugrina, M. Betke, J. Collomosse. NPAR 2006.
- **Genetic Paint: A Search for Salient Paintings**
J. Collomosse, P. Hall. EvoMUSART 2005 (J. IJAIT 2006).
- **The Art of Scale Space**
J. A. Bangham, S. Gibson, R. Harvey. BMVC 2003.
- **Visual interest and NPR: An evaluation and manifesto**
A. Santella, D. DeCarlo. NPAR 2004.
- **Segmentation-based 3D Artistic Rendering**
A. Kolliopoulos, J. Wang, A. Hertzmann. EGSR 2006.
- **Stylized Video Cubes**
A. Klein, P. Sloan, A. Colburn, A. Finkelstein, M. Cohen. EG SCA 2002.
- **Image and Video based Painterly Animation**
J. Hayes and I. Essa, NPAR 2004.



- **Stroke Surfaces: Temporally Coherent Artistic Animations from Video**
J. Collomosse, D. Rowntree, P. Hall. IEEE TVCG 2005.
- **Video Watercolorization using Bidirectional Texture Advection**
A. Bousseau, D. Neyret, J. Thollot, D. Salesin
- **Video Analysis for Cartoon-like Special Effects**
J. Collomosse, D. Rowntree, P. Hall. BMVC 2003.
- **Video Analysis for Dynamic cues and Futurist Art**
J. Collomosse, P. Hall. Graphical Models. 2006.
- **Motion Magnification**
C. Liu, A. Torralba, W. Freeman, F. Durand, E. Adelson. SIGGRAPH 2005
- **Video SnapCut: Robust Video Object Cutout Using Localized Classifiers**
X. Bai, J. Wang, D. Simons, G. Sapiro. SIGGRAPH 2009
- **Stylized Displays of Home Image and Video Collections**
T. Wang, R. Hu, J. Collomosse, D. Slatter, P. Cheatle, D. Greig. NPAR 2010 (CAG 2011)
- **Painterly animation using video semantics and feature correspondence**
L. Liang, K. Zeng, H. Lv, Y. Wang, Q. Xu, S. Zhu. NPAR 2010
- **From Image Parsing to Painterly Rendering**
K. Zeng, M. Zhao, C. Xiong, S. Zhu. ACM ToG 2010.



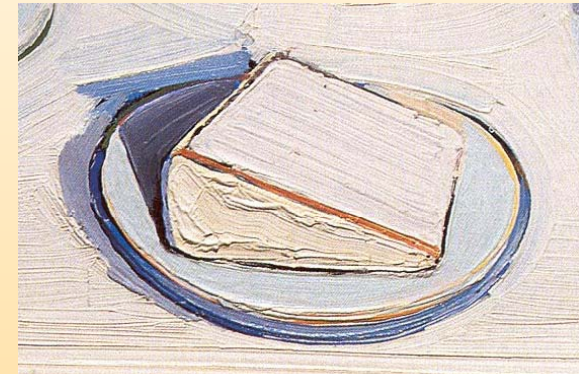
- **Artistic Stylization pre-2000**
 - Dependent on low-level image processing (e.g. Sobel) to drive preservation of local edge and high frequency content.
- An Artist does not paint a stroke by looking only at the image content under that stroke
- A higher level of visual analysis is needed:

Consider more than local edge information

Global analysis vs. greedy placement

- Computer Vision and Optimisation are solutions

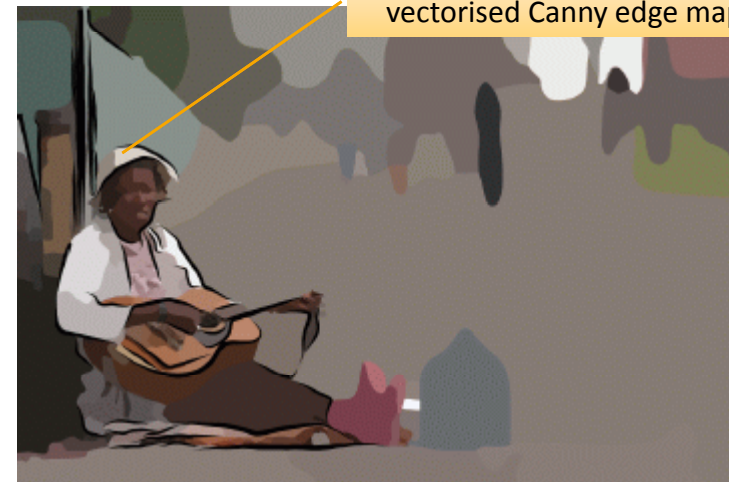
Region-based discrimination



Around the Cake (Thiebaud '62). Markup (Kolliopoulos '06)



- **Segmentation (EDISON / Mean-Shift)** [Christoudias et al, ICPR 2002]
 - Create a spatial hierarchy of regions
 - Strokes painted in a region have same prominence
 - Or render regions flat with black edges to create 'toon effect
 - Determine prominence of regions interactively
 - ...using an eye tracker





Stylization and Abstraction of Photographs Implementation Steps

- Segment levels of low-pass (Gaussian) pyramid
 - DeCarlo uses factor of $\sqrt{2}\sigma$ between layers
 - Discard regions < 500 pixels (on 640x480 image)
- Segments grouped into hierarchy from fine to coarse based on overlap and common colour

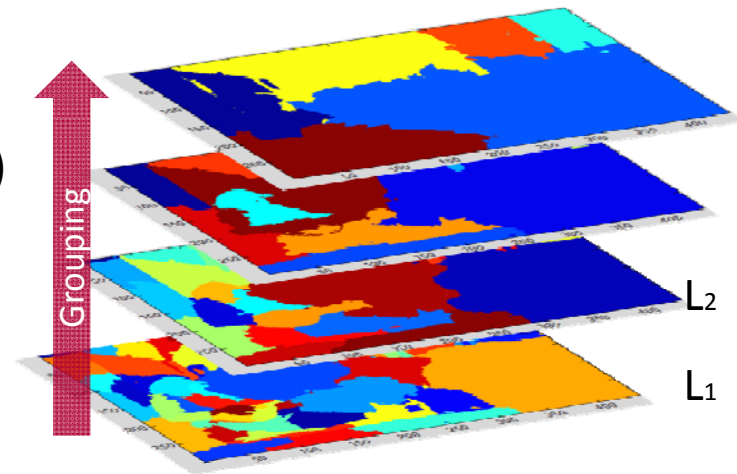
1. For each* region A at the current level e.g. L_1
2. Find the region B_i in level above e.g. L_2 maximising:

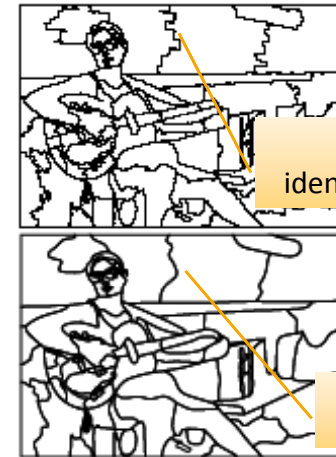
$$\text{overlap}(A, B_i) = \frac{\text{area}(A \cap B_i)}{\|\text{color}(A) - \text{color}(B_i)\| + 1}$$

3. Assign A's parent to B_i , providing $A \cap B_i$ is contiguous+

*At step 1, iterate through regions in order of increasing area.

+ After all levels are processed, any orphan regions become children of root node.





Region detail
identified via fixation

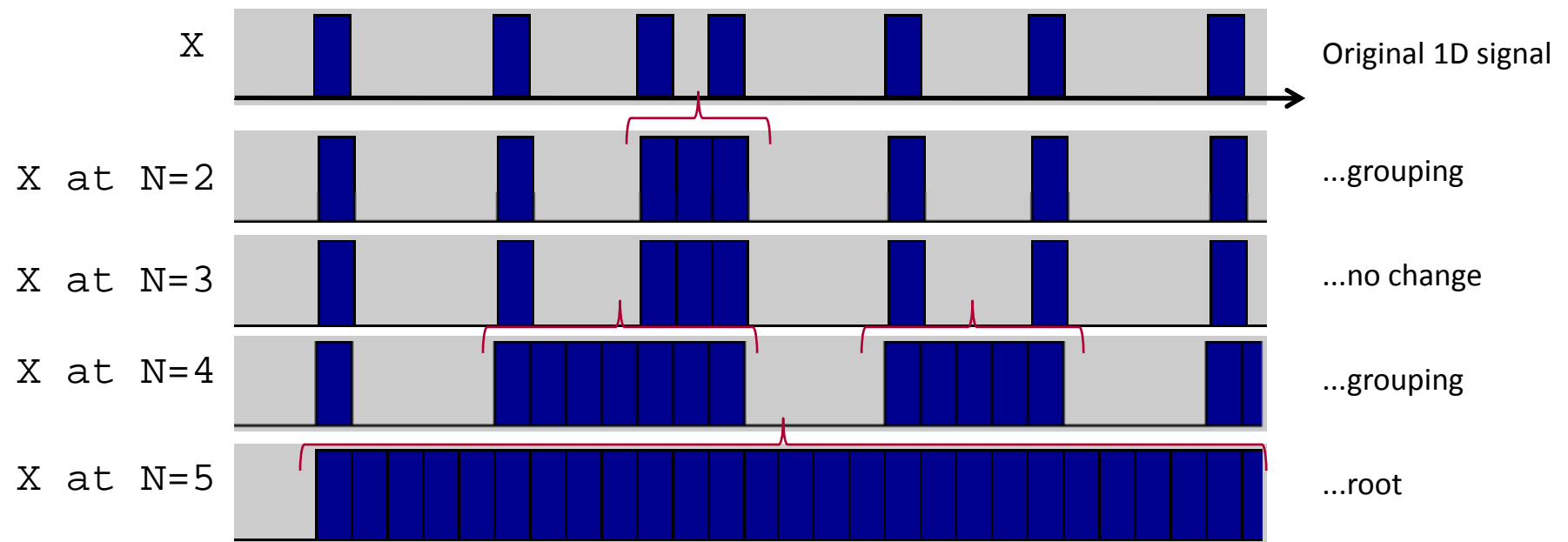
Post-smoothing

- Painting starts at the coarsest level of region detail
- A region is split if more than half its children are fixated upon
- The resulting region map is noisy, but aesthetics improve after smoothing and vectorisation



- **Alternative scale-space hierarchy using sieves**
 - **Morphological operations (closure followed by opening)**

```
X=imerode(imdilate(X,ones(1,N)),ones(1,N));
```





- **Sieves better preserve edges/corners vs. Gaussian**
 - Extended to 2D in [Bangham '99], NPR application [Bangham '03]. Colour sieves (Harvey '04)



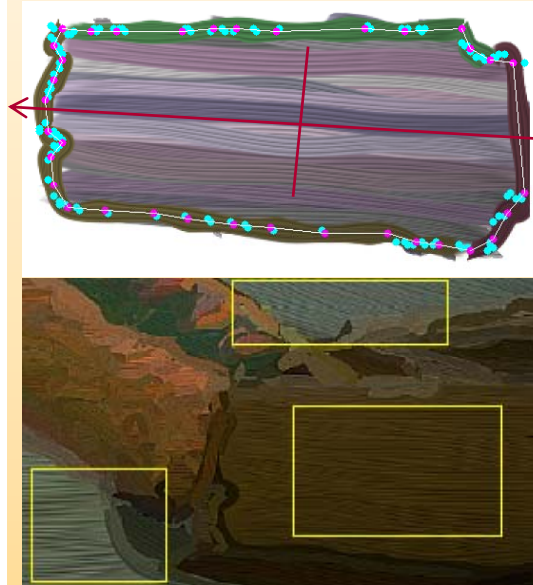
- Similar level of detail strategy to Decarlo/Santella can be applied to scale-space tree



■ Painting the regions



Paint via 3rd party algorithm e.g. Hertzmann with constant stroke size [Santella /DeCarlo NPAR'02]



Fill region with strokes in direction of principal axis [Shugrina et al, NPAR '06]

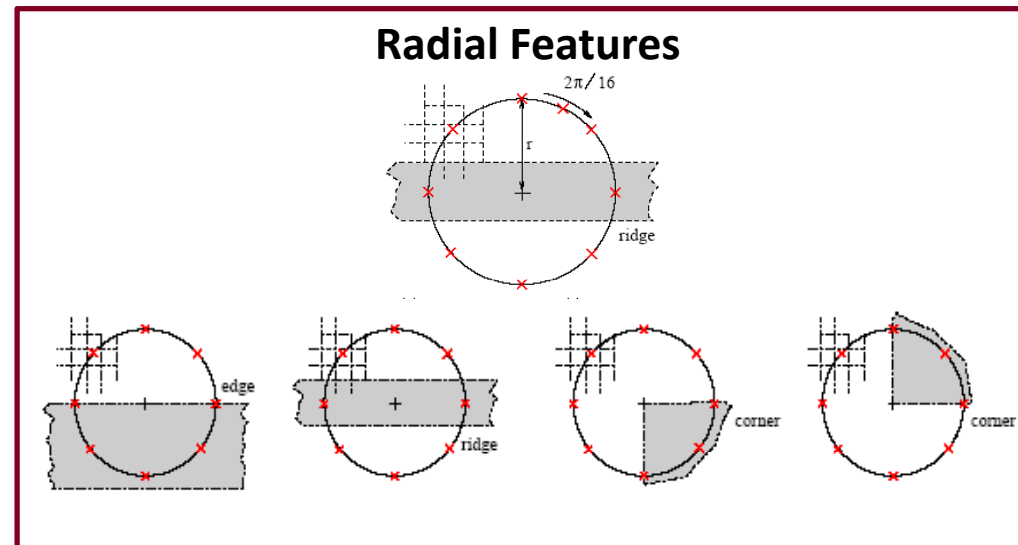
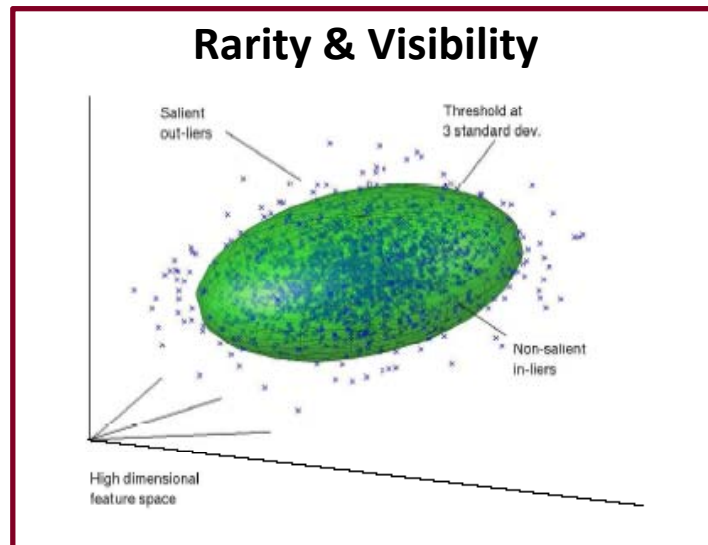


Fill with strokes in directions derived from region exterior contour [Wang et al, NPAR '10]

c.f. video painting...

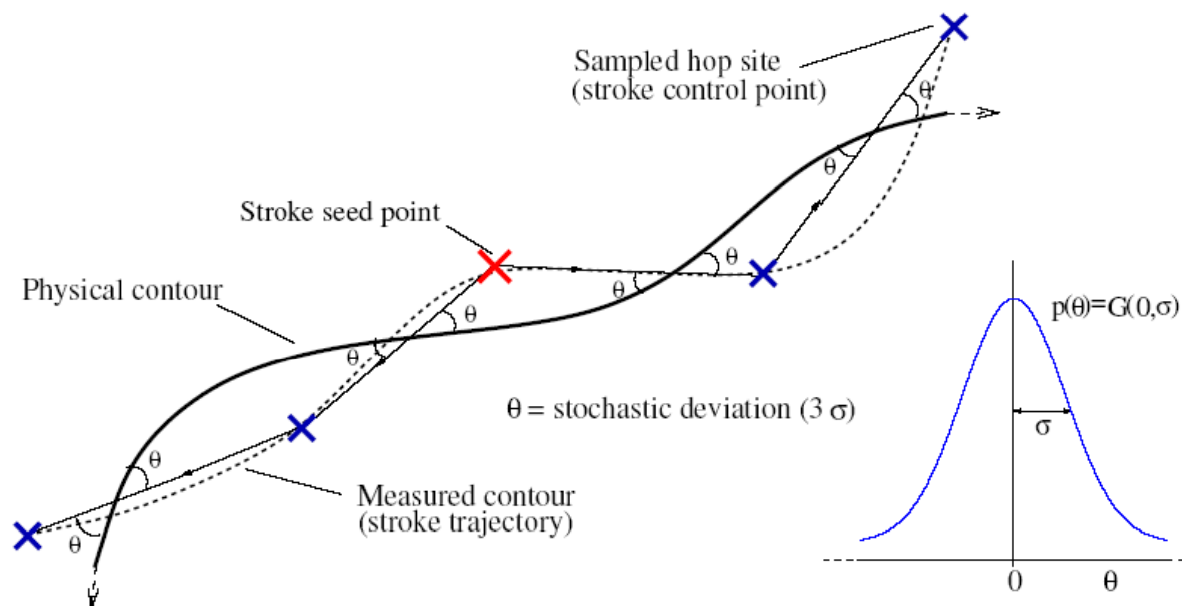


- **Automated Differential Emphasis in Painting**
 - **Prescriptive salience measures [Itti & Koch]**
 - Not closely correlated to human behaviour [Santella/DeCarlo NPAR'04]
 - Salience is subjective and task dependent
 - **Trainable measure of salience (GMM of radial features)**



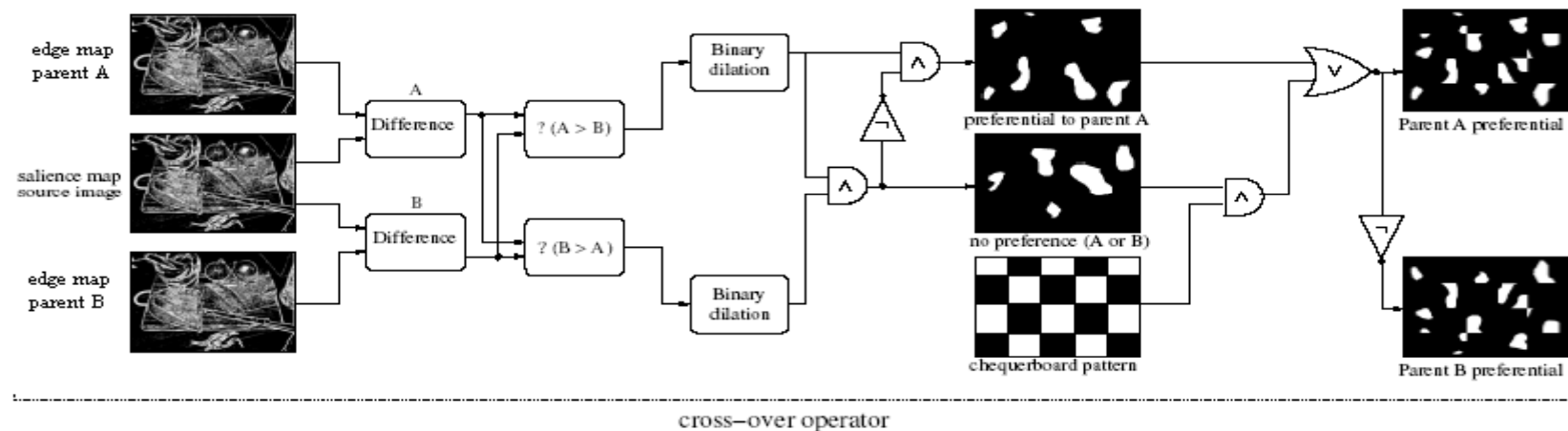


- Genetic Optimizaton to find “best” painting
 - The optimal painting preserves detail in salient areas, and removes non-salient detail
 - MSE between salience map and Sobel edge detail in the painting (c.f. Hertzmann ‘01)



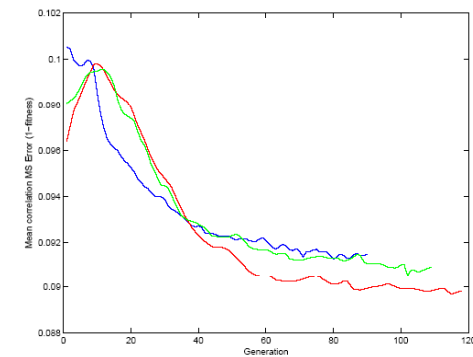


- Paintings are bred by cloning strokes from two individuals
- (Two parent cross-over)
 - fitness proportionate selection with replacement
- Promotion of rapid convergence
 - Top 10% carried over to next gen. automatically
 - Bottom 10% culled





- Iterative optimization improves detail in salient regions
 - Population of ~50 paintings
 - Convergence in ~200 iterations
 - Stochastic variation in stroke attributes creates diversity
 - GA combines favourable regions of parent paintings

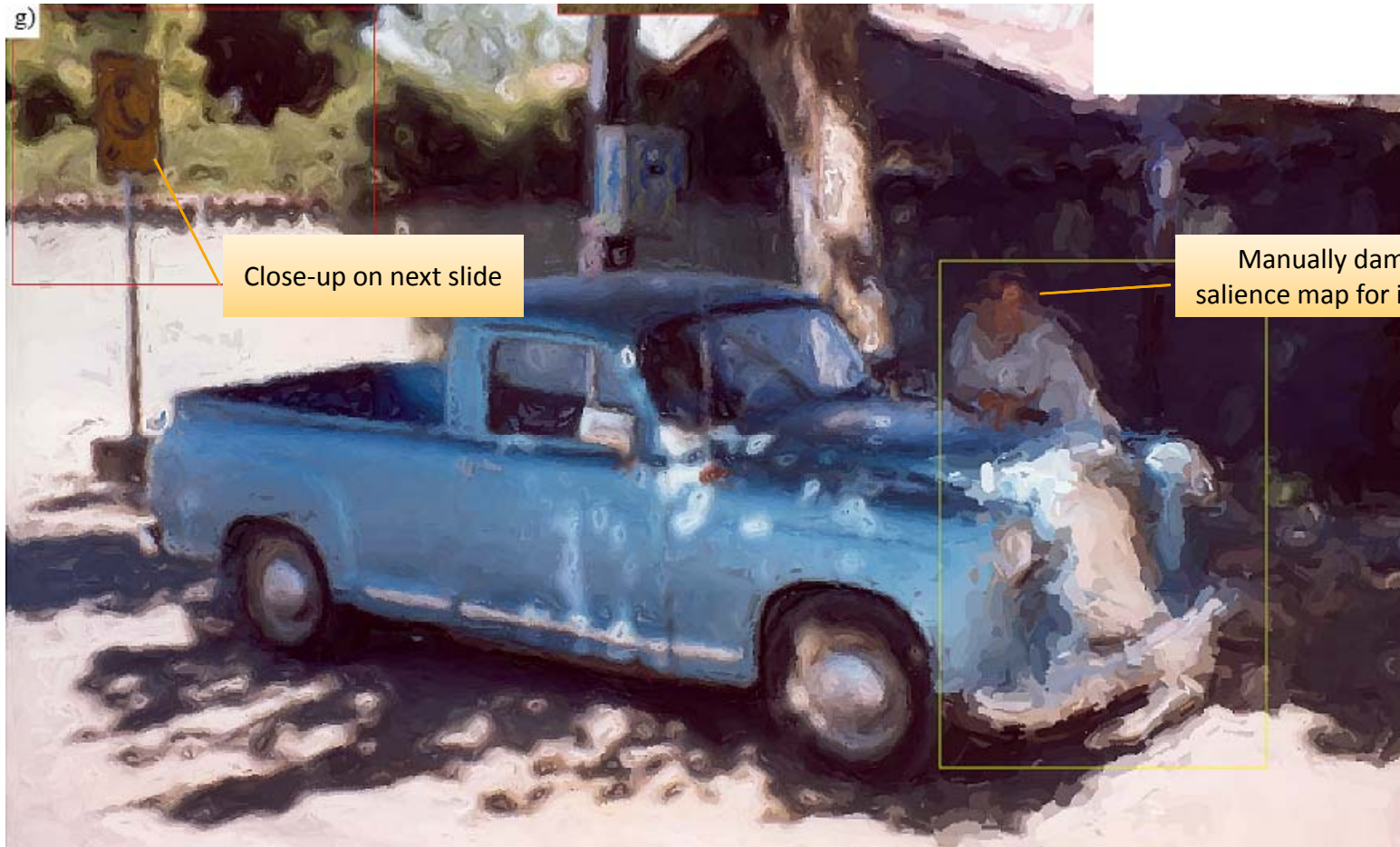




Eurographics 2011
LLANDUDNO UK
11-15 April 2011

Genetic Paint: Search for Salient Paintings

Collomosse et al. 2005.





- **Comparison of Sobel-driven and Saliency-driven painting**
 - **Detail on the sign is preferentially retained (wrt. Leaves of the tree)**
 - **Not all edges / high frequency texture are salient**



Original



Litwinowicz '97



Saliency driven



- **Painting research code available**
 - <http://www.collomosse.com/EG2011tut/summerschool.zip>
- **MATLAB based (experiment with different salience maps)**
- **Code adapted from Collomosse et al. 2005 – single iteration, spline strokes.**
- **Previously released as lab exercise at EPSRC VVG Summer School (2007)**



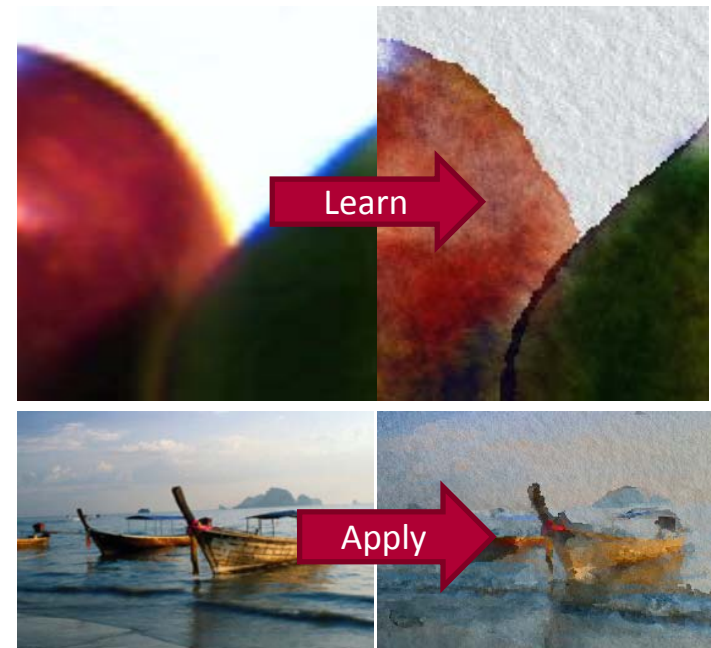
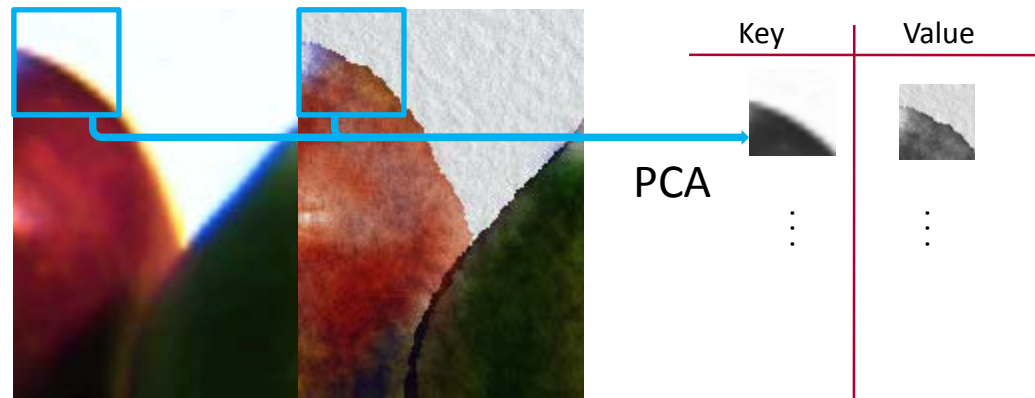


■ Style Transfer

- Learning vs Heuristic approach to stylise photos
- Patch based lookup (luminance only)

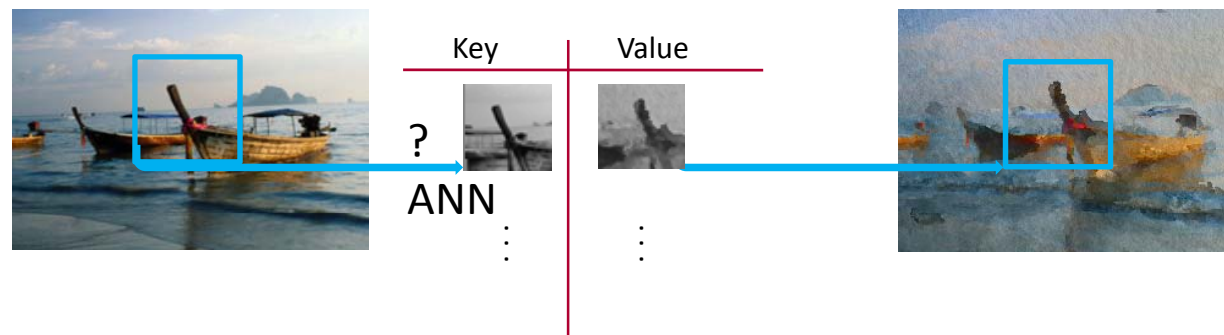
$$Y(p) \leftarrow \frac{\sigma_B}{\sigma_A}(Y(p) - \mu_A) + \mu_B$$

- Similar to Freeman texture synthesis but using external collection of patches
- Learned as lookup table

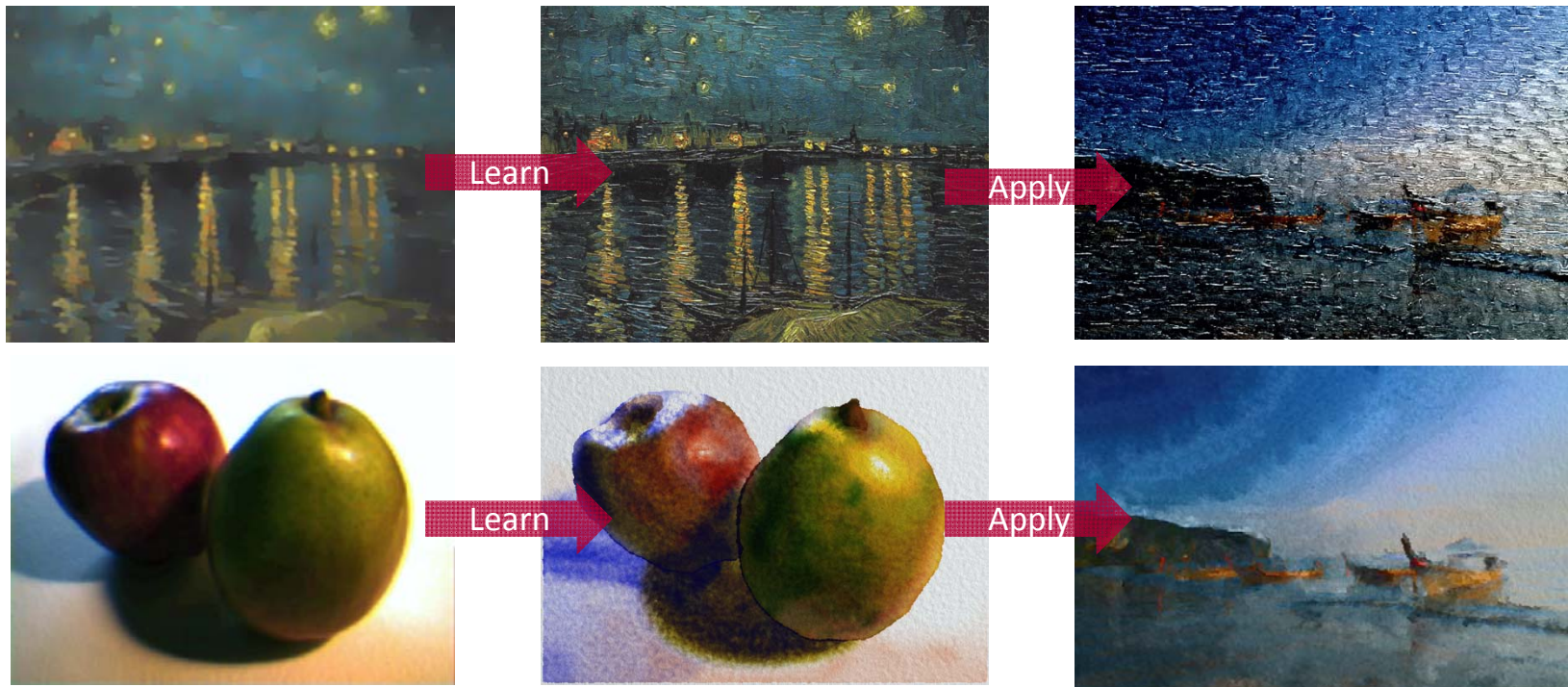


■ Style Transfer

- Synthesis has 'data' and 'smoothness' terms
 - Data (patch lookup)
 - Pixel-wise luminance comparison (after PCA)
 - Smoothness (derived from Ashikhmin)
 - Minimise MSE between proposed patch and existing neighbours
 - Gaussian weighted distance function (avoids discontinuity)



- **Style Transfer Examples**



- **Other extensions to video [Hors & Essa '02] and to take orientation into account [Lee et al. '10]**



- **Video Stylisation**
 - Techniques to create painterly animations or cartoons from video
 - Enabled by automated techniques for image stylization



Stylised Appearance



Stylised Motion



- Goal of video stylization
 - Create the desired aesthetic exhibiting good temporal coherence
- Temporal coherence is here defined as:
 1. Absence of distracting flicker
 2. Motion of brush strokes (or other component marks) is in agreement with the motion of content
- Naïve approaches
 - Repaint every frame independently
= *Flicker (violates 1.)*
 - Fix strokes in place and change attributes e.g. colour according to video content
= *Motion unmatched (violates 2.)*
“the shower door effect” – Barb Meier





- **Painterly animation using Optical Flow**
 - Brush strokes are pushed from frame to frame using flow estimate
 - Oscar winning visual effects in movie “What Dreams May Come” (1998)
 - Manual correction of flow estimate (~1000 person-hours [Green’99])

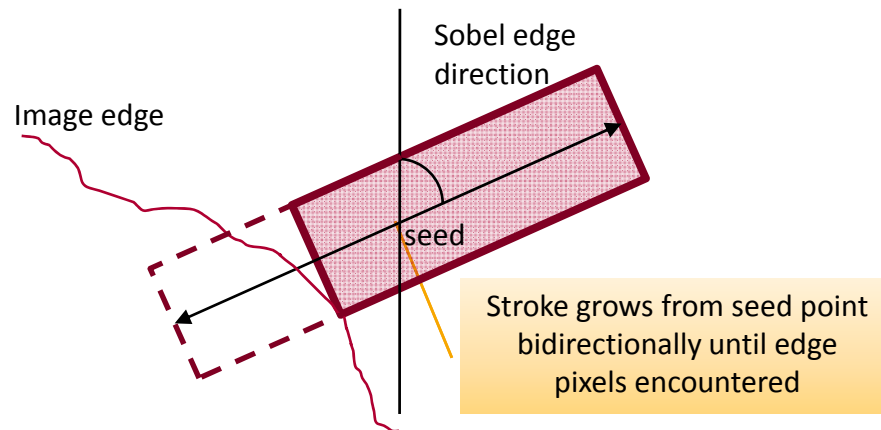
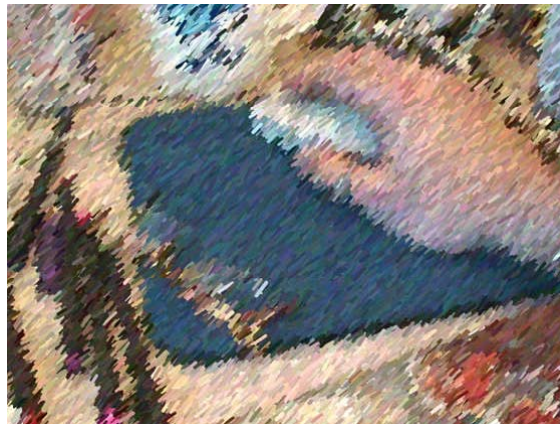




Eurographics 2011
LLANDUDNO UK
11-15 April 2011

Processing Images & Video for Impressionist Effect

Litwinowicz (1997)



No clipping



Clipping



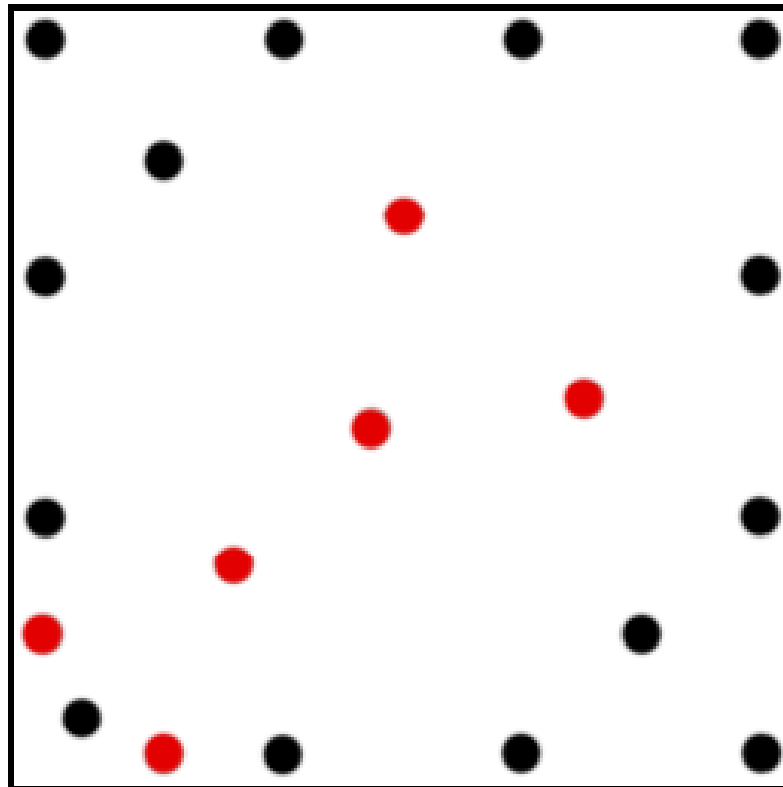


- **Initialisation as per single image (regular seeding)**
 - Randomise rendering order of strokes
- **Strokes translated to next frame via flow field**
- **Greedy approximation to avoid irregular coverage**
 - Delaunay triangulation of seeds (and image corners)
 - **Death.** Seeds too close together are deleted
 - Tested in random order
 - **Birth.** Triangles with area $>$ threshold are subdivided
 - New seeds are randomly place rendering order





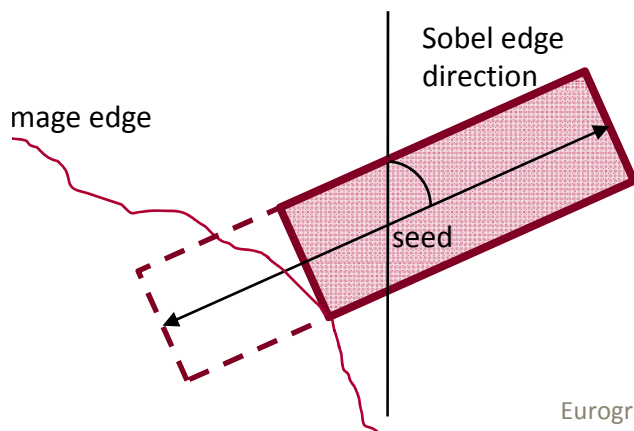
▪ Stroke Birth





■ Tips on reducing flicker

- Detect your own scene cuts and reinitialise
- Use a robust Optical Flow algorithm (!)
 - e.g. SIFTFlow or Brox
- Pre-filter heavily (Gaussian). Care with interlaced content.
- Interpolate orientations from strong edges only
 - Smooths out codec noise
 - Litwinowicz uses thin-plate spline (expensive) but can use Poisson filling (fast on GPU) to good effect



Sobel field



Without interpolation



With interpolation



- **Main sources of temporal incoherence**
- **Motion matching**
 - **Optical Flow = visual correspondence problem**
 - **Inevitable inaccuracies in estimate are cumulative**
 - **Content appears to slip below strokes = shower door effect**
 - **Manual correction of OF mitigates this but is expensive**
- **Flicker**
 - **Random order of new strokes disguises regularity**
 - **...but the noise generates flicker**
 - **Sudden disappearance of strokes exposes others = popping**
 - **Sobel edges are noisy at moderate scales**
 - **Strokes are clipped against flicking edge map**



- **Main sources of temporal incoherence**

- **Motion matching**

- **Optical Flow = visual correspondence problem**
- **Inevitable inaccuracies in estimate are cumulative**
- **Content appears to slip below strokes = shower door effect**
 - **Manual correction of OF mitigates this but is expensive**

Addressed by

Hertzmann and Perlin
NPAR 2000

- **Flicker**

- **Random order of new strokes disguises regularity**
 - **...but the noise generates flicker**
- **Sudden disappearance of strokes exposes others = popping**
- **Sobel edges are noisy at moderate scales**
 - **Strokes are clipped against flicking edge map**

Hays and Essa
NPAR 2004



- Repaint only the areas that change significantly
 - Fast – enables realtime interaction
 - Limits shower-door by repainting limited regions of canvas (“paint-over”)
- RGB Difference to detect regions
 - Optical flow *optionally* used to translate strokes

$$\frac{1}{|M|} \sum_{(i,j) \in M} ||I_{t+1}(i,j) - I_t(i,j)|| > T_V$$

- Control points shifted under flow





Eurographics 2011
LLANDUDNO UK
11-15 April 2011

Paint-Over





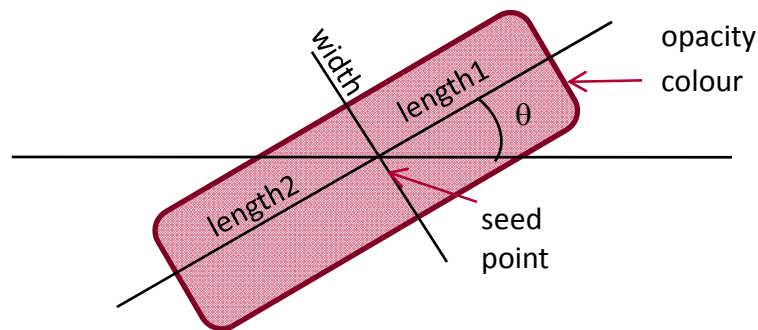
Eurographics 2011
LLANDUDNO UK
11-15 April 2011

Paint-Over and Optical Flow





- **Key Innovation**
 - Temporal smoothing of stroke attributes
- **Stroke Opacity for birth/death**
- **Orientation**
 - RBF interpolated field (similar to Litwinowicz)
 - But interpolated from strokes marked “strong”
not from per-frame orientation field
 - Strokes born on strong edges
- **Length and orientation are also smoothed**





Eurographics 2011
LLANDUDNO UK
11-15 April 2011

Image and Video-based Painterly Animation

Hayes & Essa (2004)





Eurographics 2011
LLANDUDNO UK
11-15 April 2011

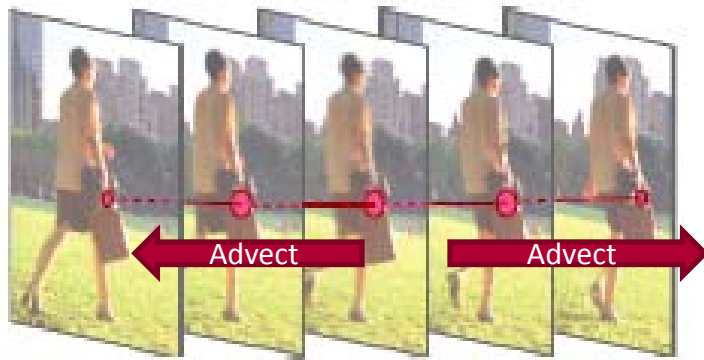
Image and Video-based Painterly Animation

Hayes & Essa (2004)





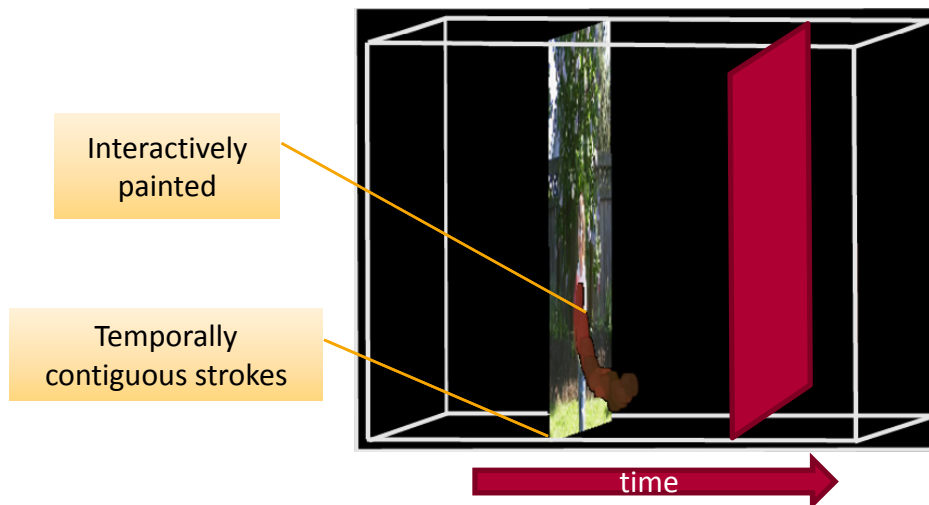
- **Bi-Directional Flow (of textures vs. strokes)**
 - Adaptation of “Texture advection” from flow visualisation
 - Frequent occlusions in video motivated bi-directional flow
 - Two textures seeded – one flows forward, one back



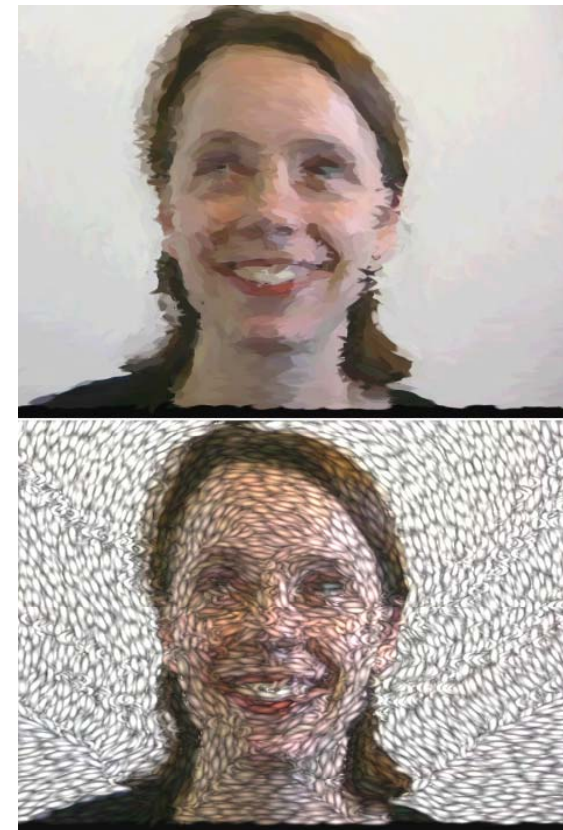
- **Trend towards more global temporal analysis**



- **Temporally local (inter-frame) approaches**
 - No long-view of video structure
 - necessitates averaging of past information
 - Averaging mitigates flicker but exaggerates the shower door effect
- **Spatio-temporal primitives**



timeslice



- **Automated Space-time Analysis**
 - Goal is coherent segmentation of video into semantic regions
 - Coherent space-time regions are smoothed then do not flicker



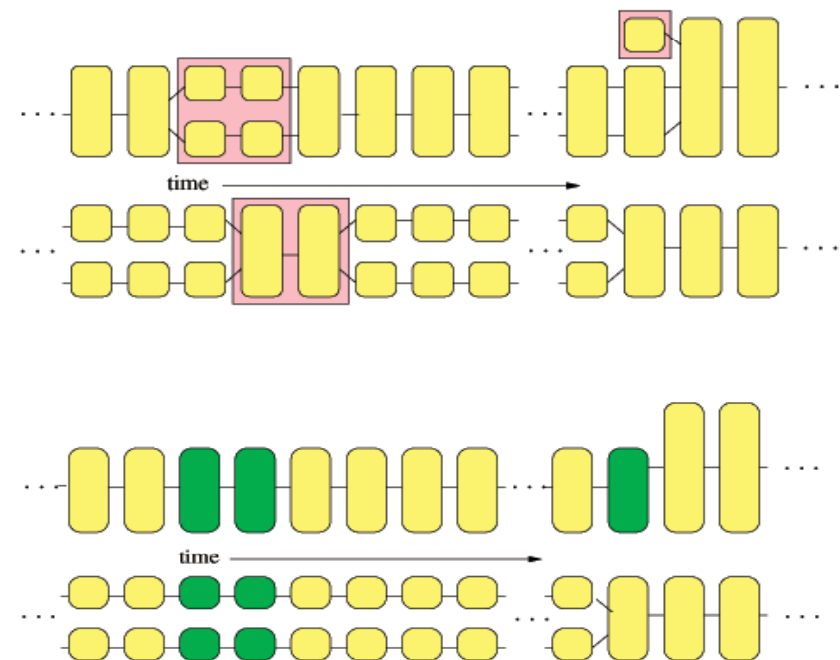
Extending Mean-shift to Space-time (3D)



EDISON segment frames and associate (2D+t)

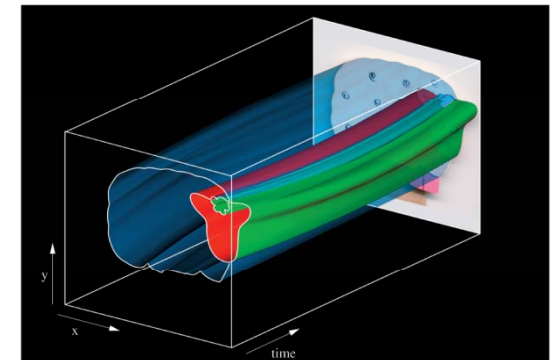


- **Region association (2D+t)**
 - Based on a weighted blend of heuristics
 - Shape (Fourier Descriptors)
 - Colour
 - Overlap (as DeCarlo/Santella)
 - Associations are filtered by locating
 - Short-time branches
 - Short-time cycles
 - Surface voxels between volumes are identified
 - Surfaces fragmented into “stroke surfaces” that abut only two volumes





- “Stroke Surfaces” separate volumes
 - Winged edge structure
 - Smoothing the surface smooths the volumes
 - Generalisation of snakes to 2D surfaces
 - Separate terms for spatial and temporal constraint



$$E = \int_0^1 \int_0^1 (E_{int}[\underline{Q}(s, t)] + E_{ext}[\underline{Q}(s, t)]) ds dt.$$

$$E_{int} = \alpha \left| \frac{\partial \underline{Q}(s, t)}{\partial s} \right|^2 + \beta \left| \frac{\partial \underline{Q}(s, t)}{\partial t} \right|^2 + \gamma \left| \frac{\partial^2 \underline{Q}(s, t)}{\partial s^2} \right|^2 + \delta \left| \frac{\partial^2 \underline{Q}(s, t)}{\partial t^2} \right|^2$$

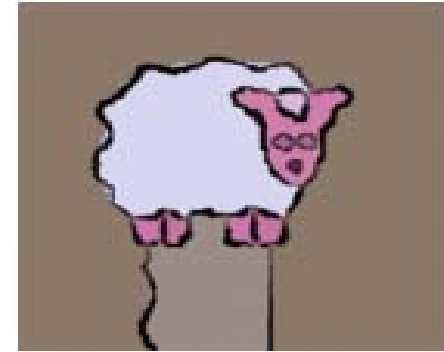
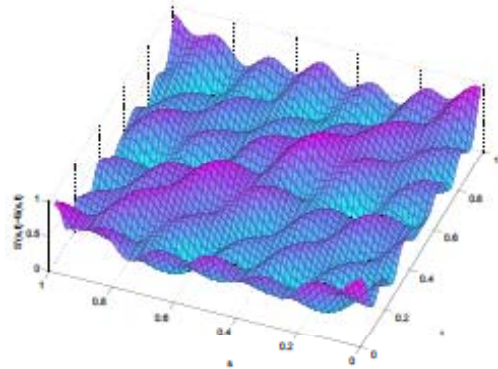
$$E_{ext} = \eta f(\underline{Q}(s, t)).$$



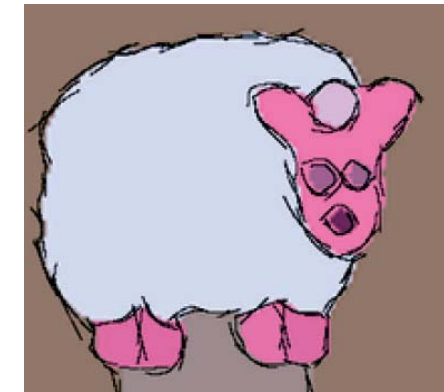
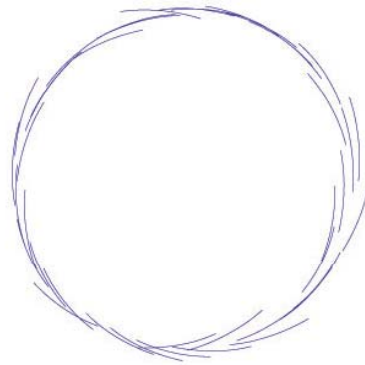
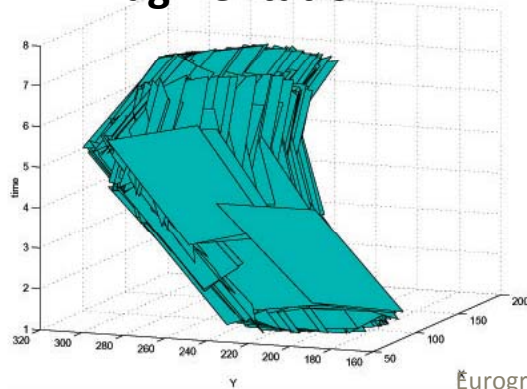


■ Surface Manipulation

■ Undulation



■ Fragmentation

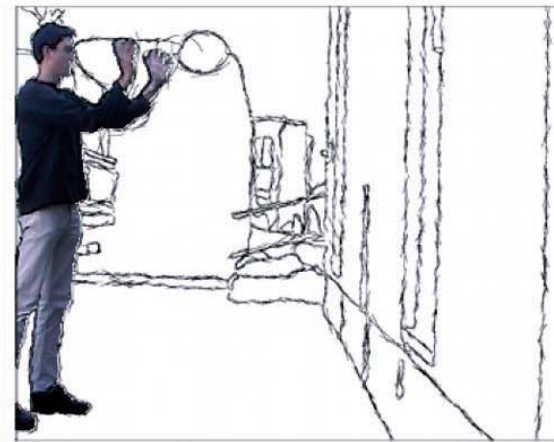
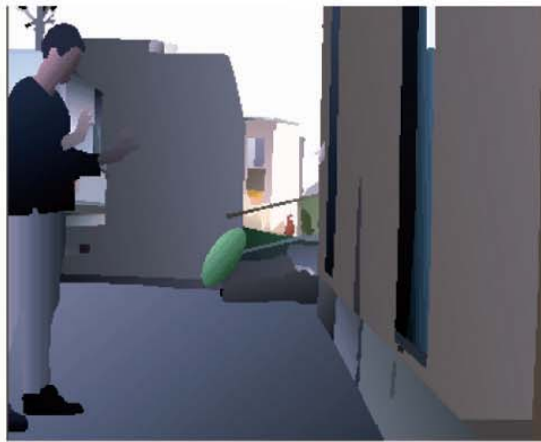




Eurographics 2011
LLANDUDNO UK
11-15 April 2011

Stroke Surfaces: Coherent Artistic Animations from Video

Collomosse et al. (2005)





Eurographics 2011
LLANDUDNO UK
11-15 April 2011

Stroke Surfaces: Coherent Artistic Animations from Video

Collomosse et al. (2005)

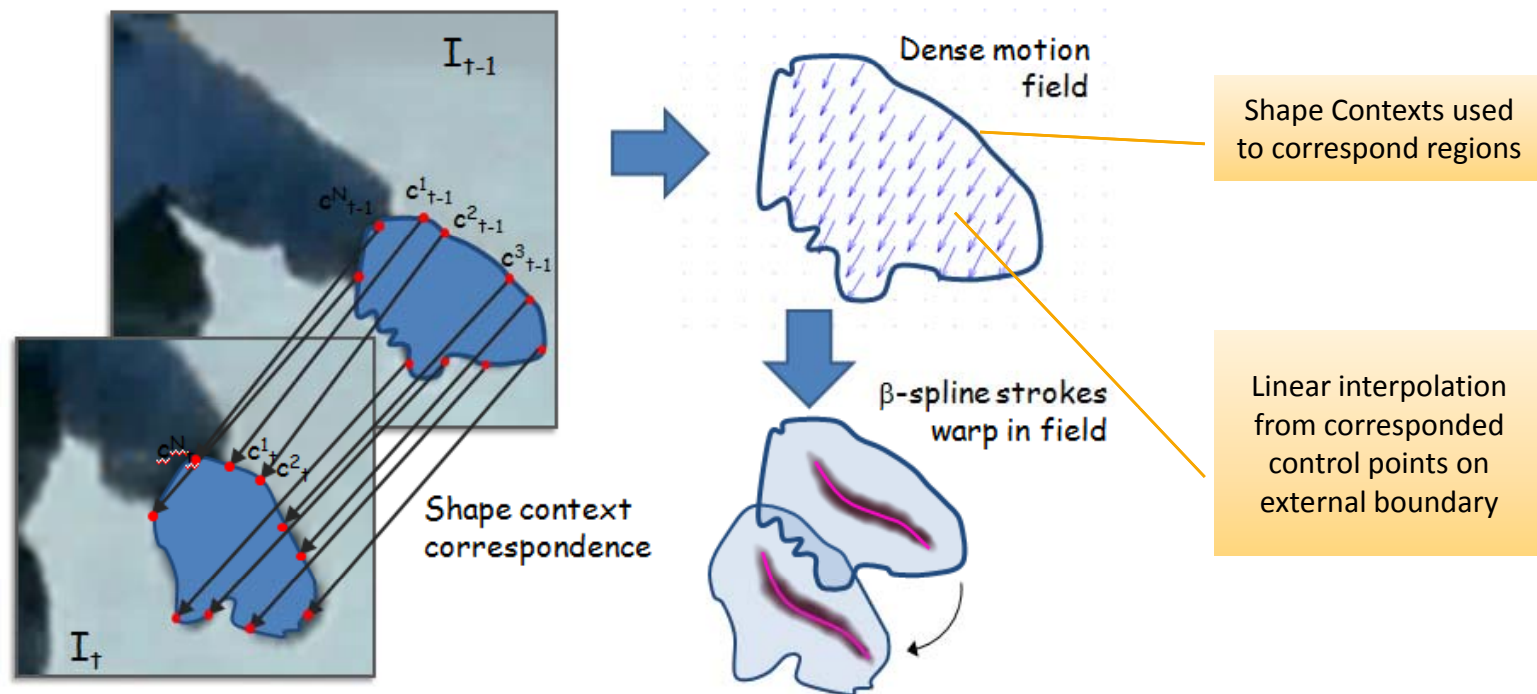
■ Coherent Segmentation



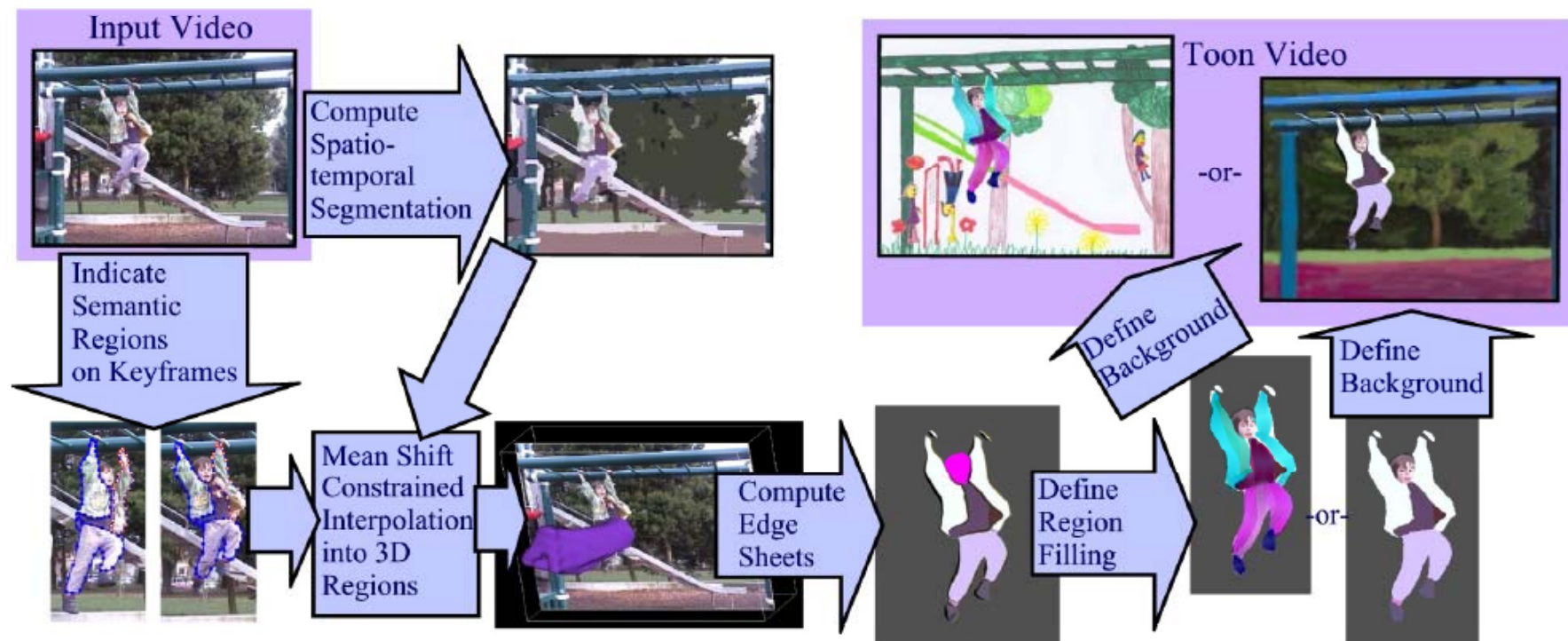


■ Rotoscoping

- Coherent motion of (groups of) regions can be exploited to paint coherently
- Interpolate internal points (e.g. stroke seeds) from region boundary



■ Rotoscoping





■ Rotoscoping

- Coherent motion of (groups of) regions can be exploited to paint coherently





■ Motion Emphasis

- Augmentation cue (Speed-lines, ghosting)
- Deformation (Squash and stretch, general deformation)
- Time warping (Anticipation/snap)





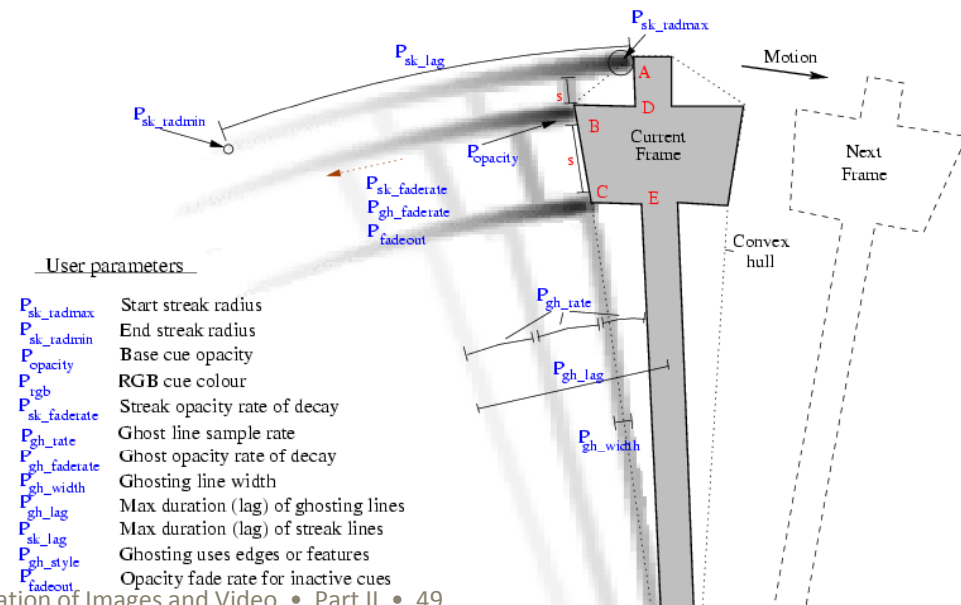
■ Augmentation Cues

- Segment trails of corresponded control points into smooth sections.
- Iteratively select smooth sections to maximising:

$$H(0) = 0$$

$$H(i+1) = H(i) + (\alpha v(x) + \beta L(x) - \gamma D(x) - \delta \omega(x, \sigma; w) + \zeta \rho(x))$$

$\alpha, \beta, \gamma, \delta, \zeta$: user weights
 σ selected sections
 $v(x)$ mean velocity over x
 $\rho(x)$ mean curvature at x
 $L(x)$ temporal extent of x
 $D(x)$ dist from convex hull at x
 $\omega(x, \sigma; w)$ overlap of x and σ

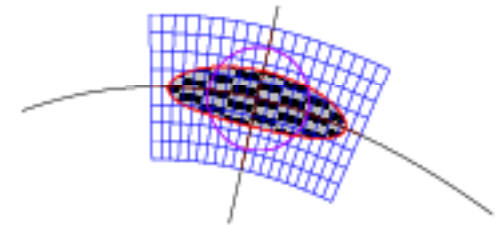




■ Deformation Cues

- A motion dependent curvilinear basis is formed using the trajectory of the region centroid, and its normal.

$$\underline{x} = \underline{G}_c(s) + r\underline{n}(s)$$



$U(\cdot)$ as the transformation from curvilinear space, keep the inverse as a lookup table.

$$\underline{x} = U(\underline{r}) \quad \underline{r} = (s, r)^T$$

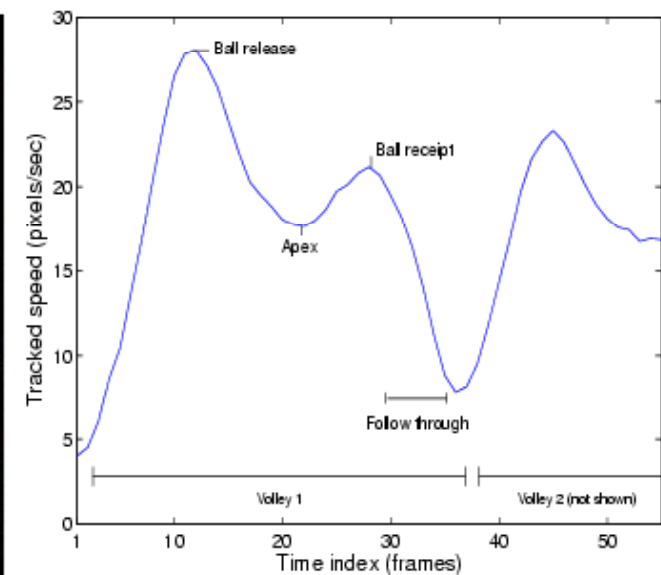
$$\underline{x} \leftarrow U\left(\begin{bmatrix} k & \mathbf{0} \\ \mathbf{0} & \frac{1}{k} \end{bmatrix} U^{-1}(\underline{x})\right)$$

$$k = 1 + \frac{K}{2} \left(1 - \cos\left(\pi \frac{v^2 + 1}{2}\right)\right)$$

$$v = \begin{cases} 0 & \text{if } |\dot{\underline{u}}| < V_{min} \\ 1 & \text{if } |\dot{\underline{u}}| \geq V_{max} \\ (|\dot{\underline{u}}| - V_{min}) / (V_{max} - V_{min}) & \text{otherwise} \end{cases}$$

Squash and stretch
(after Cheney et al '02)

- Deformation Cues
 - Squash and stretch in a camera motion compensated frame





■ Deformation Cues

- More general motion deformations can be created by specifying a transfer function dependent on a point's local acceleration and position, as well as its speed.

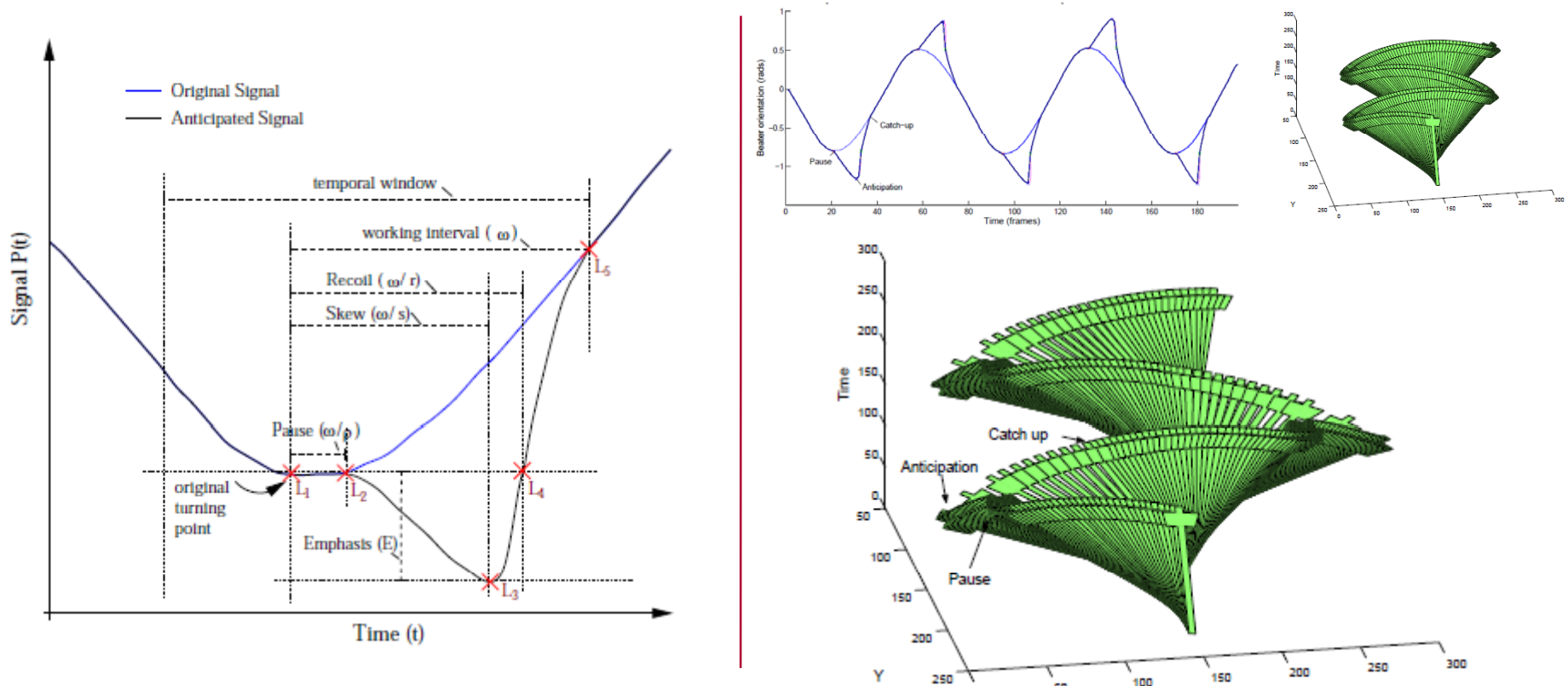
$$x' = U(T(U^{-1}(x), \dot{x}, \ddot{x}))$$

A function can operate on each component of $\underline{r} = (r_1, r_2)$ independently, to create effects suggesting drag we use...

$$r_1 = r_1 - F\left(\frac{2}{\pi} \text{atan}(|\dot{x}_i|)\right)^P \text{sign}(\dot{x}_i)$$



- Anticipation (Snap)**
 - Alter motion timing to introduce a lag then “catch up” prior to changes of motion





Eurographics 2011
LLANDUDNO UK
11-15 April 2011

Video Paintbox: The Fine Art of Video Painting

Collomosse and Hall (2006)

■ A Complete Video Paintbox



Segmentation + augmentation + deformation

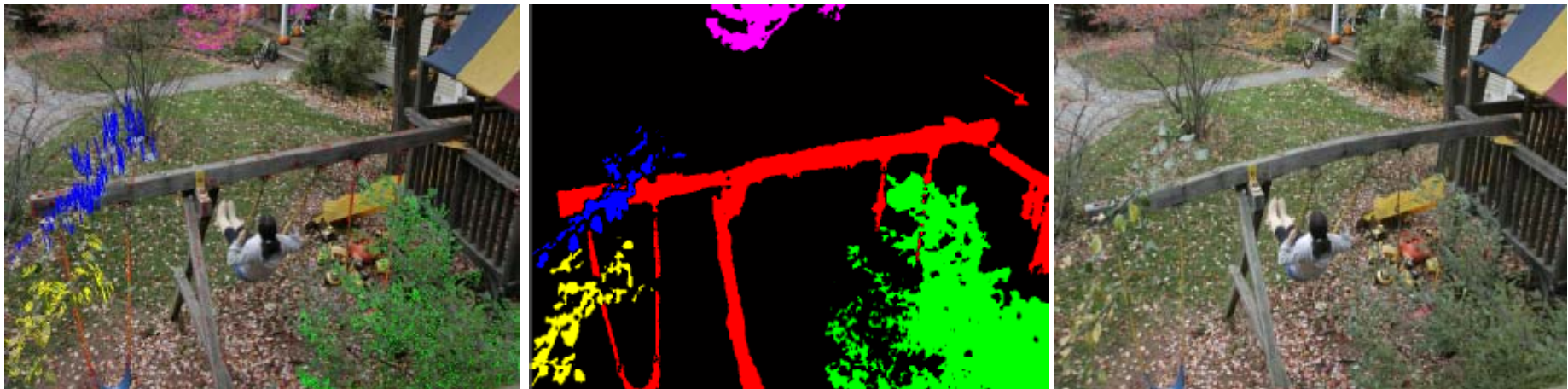


Anticipation + deformation



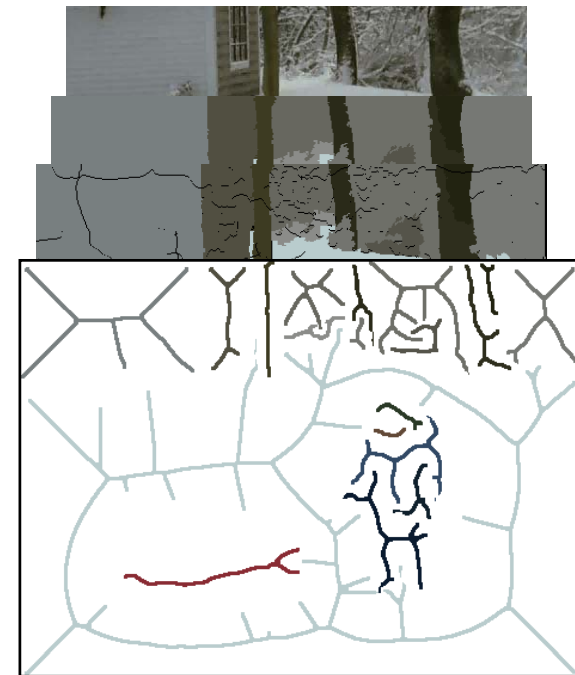
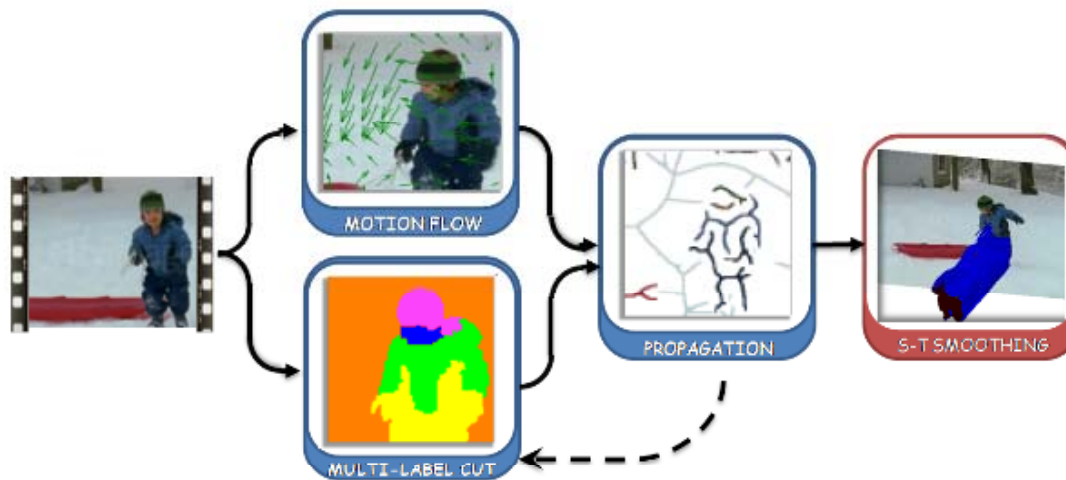
■ Deformation Cues

- General deformation technique using motion vector clustering to layer video
 - User intervention needed to fix noisy segmentation maps
- Per-pixel flow vector pushes pixels to exaggerate existing motion
- Texture filling compensates for holes





- Combining Segmentation and Optical Flow
 - Multi-label graph-cut segmentation with label prior propagated forward from previous frames
 - Region colour models are learned incrementally





Combining Segmentation and Optical Flow

- For each pixel $p \in \mathcal{P}$ within frame $I_t(p)$
 - Find best mapping $l : \mathcal{P} \rightarrow \mathcal{L}$
 $\mathcal{L} = (l(1), \dots, l(p), \dots, l(|\mathcal{P}|))$
 - Subset of \mathcal{L} are carried from t-1 by flow

$$E(\mathcal{L}, \Theta, \mathcal{P}) = U(\mathcal{L}, \Theta, \mathcal{P}) + V(\mathcal{L}, \mathcal{P}).$$

$$U(\mathcal{L}, \Theta, \mathcal{P}) = \sum_{p \in \mathcal{P}} -\log P_g(I_t(p) | l(p); \Theta).$$

$$P_g(I(p) | l(p) = l_i; \Theta) = \sum_{k=1}^{n_l} w_{ik} \mathcal{N}(I(p); \mu_{ik}, \Sigma_{ik})$$

$$V(\mathcal{L}, \mathcal{P}) = \gamma \sum_{(m,n) \in N} [l(m) \neq l(n)] e^{-\beta \|I(m) - I(n)\|^2}.$$

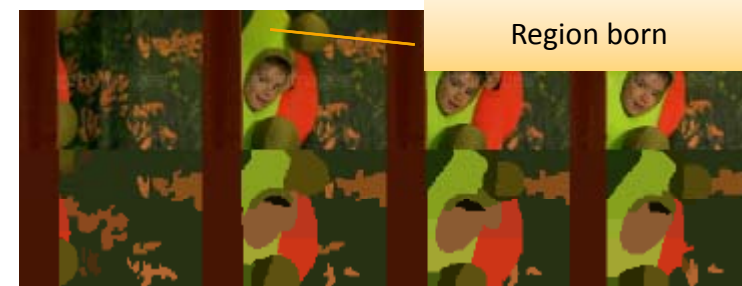
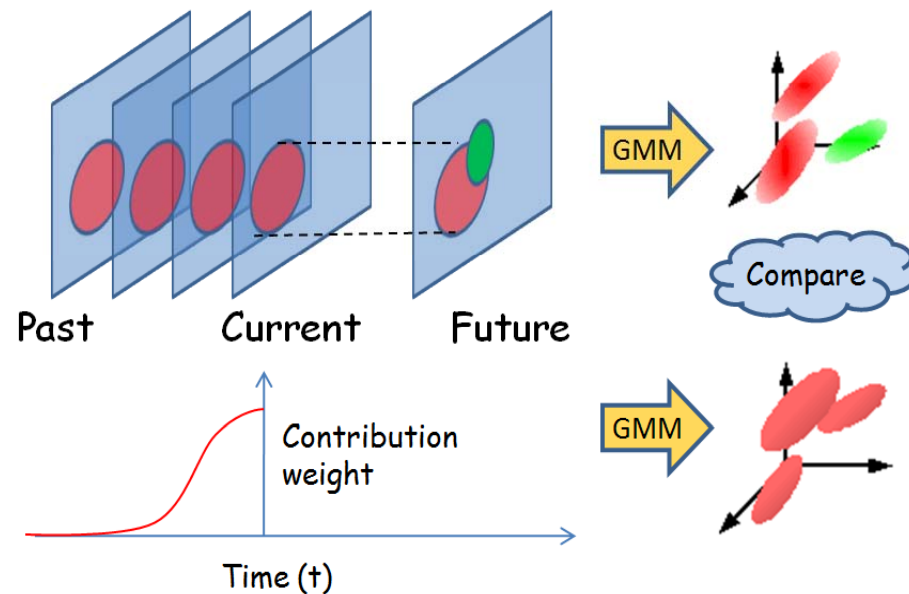
- Colour models are learned over time incrementally via Gaussian Mixtures

Learned colour model

Contrast adaptive

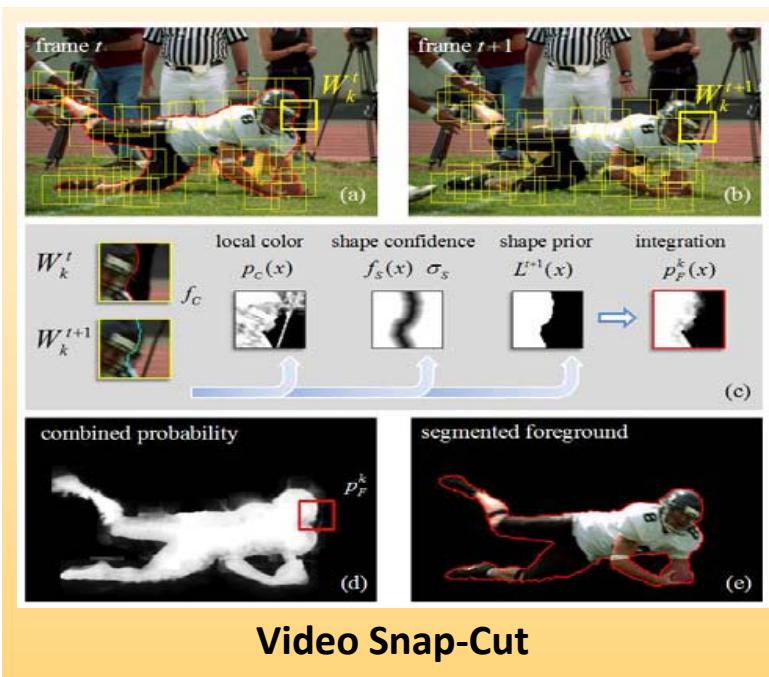


- Region colour distribution (GMM) updated with temporal weight
 - Comparison with historic model (χ^2) can detect region birth

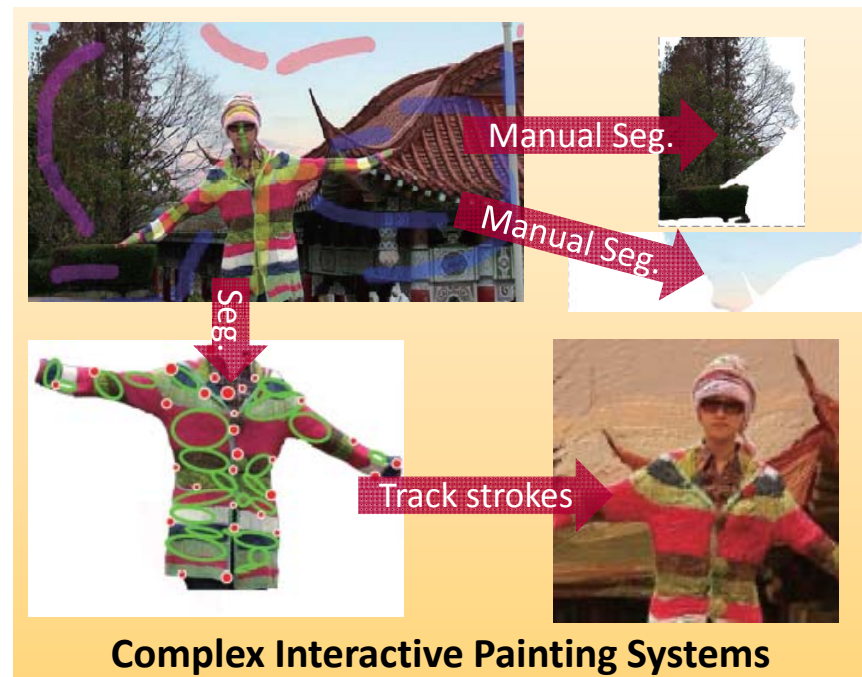


Wang et al. '10

- Trends from automatic 'Tooning to interactive tools
 - The necessity of interaction to solve the general segmentation problem



Bai et al. '09



Liang et al. '10

Yet many applications demand automation or real-time. Part III discusses solutions.



Coffee Q & A

- **After the break!**
 - **Part III - Anisotropy and Diffusion**
 - **Part IV - Future Challenges in NPR**

