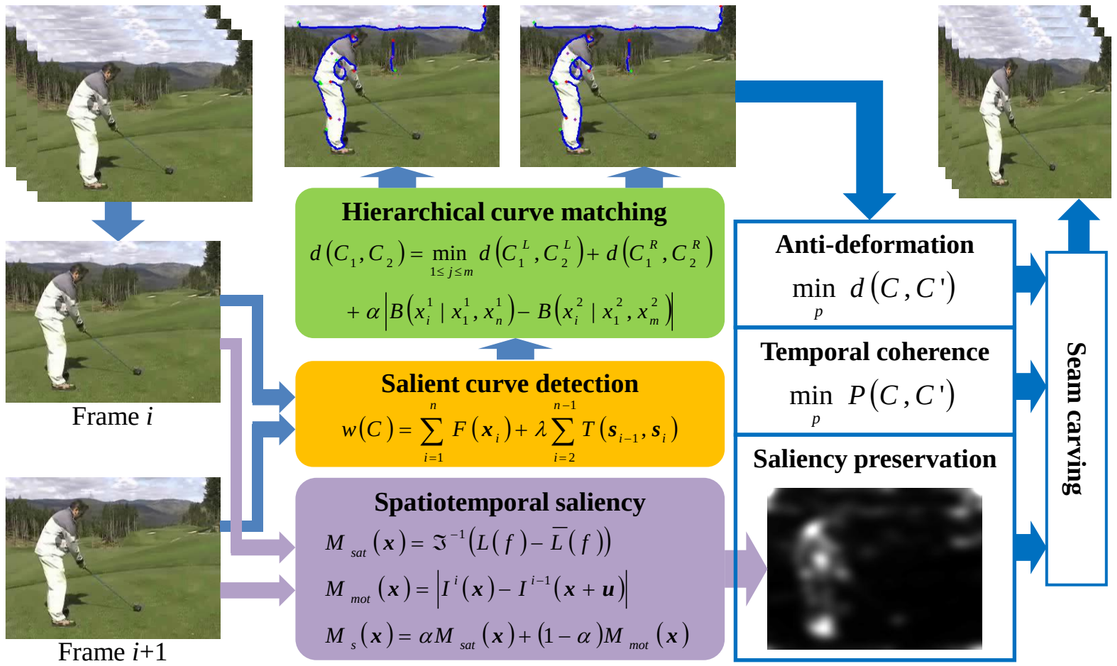
Visual Preserving Video Retargeting with Deformable Shape Consistency

Zhiquan Ren, Botao Wang, Yuchen Zhang, Hongkai Xiong
Image, Video, and Multimedia Communication Lab (IVM Lab)
Department of Electronic Engineering, Shanghai Jiao Tong University

1. Introduction

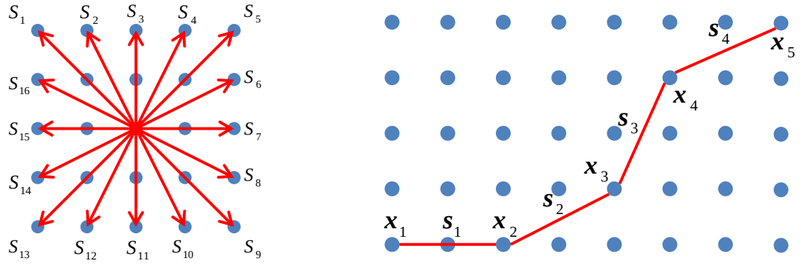
Video retargeting is a technique that resizes the aspect ratio and resolution of a video in a content aware manner. In this paper, we propose a novel shape preserving video retargeting method to reduce deformations and maintain temporal consistency via matching salient curves in the video frames. Our method transforms the anti-deformation and temporal consistency problems in video retargeting into a curve matching cost minimization problem. By incorporating a deformation cost and temporal inconsistency cost into the seam carving framework, the quality of the retargeted videos can be significantly improved.



System framework

2. Salient curve detection

A curve is represented as a sequence of oriented segments $c=(x_1,x_2,\cdots,x_n)$, where $s_i=x_{i+1}-x_i\in S$



Oriented segments Curve representation

The weight of a curve is defined as

$$W(c)=\sum_{i=1}^{n}F(x_i)+\lambda\sum_{i=2}^{n-1}T(s_{i-1},s_i)$$

where $F(\boldsymbol{x})$ is the saliency map, and $T(\boldsymbol{u},\boldsymbol{v})$ is a smoothness term that encourages consistent orientations of consecutive fragments $\boldsymbol{u}$ and $\boldsymbol{v}$. Salient curves can be extracted by detecting curves with large weights and sufficient lengths.

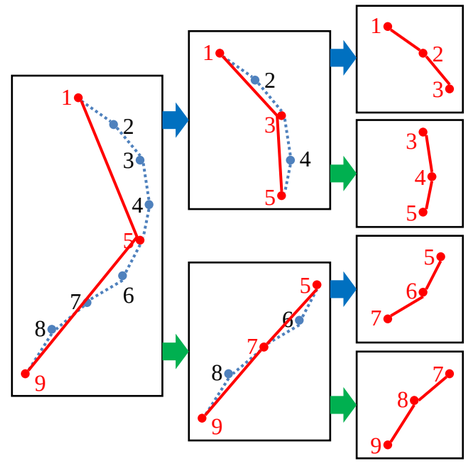
3. Hierarchical shape matching

A curve is represented by a binary shape tree. Each node in the shape tree stores the relative location of the midpoint with respect to the two endpoints of a sub-curve. The left child of the node stores the first half sub-curve and the right child stores the second half sub-curve in the same manner.

Given two endpoints $x_1=(x_1,y_1)$ and $x_n=(x_n,y_n)$ of a curve, the relative location of the midpoint $x_i=(x_i,y_i)$ is represented in the Bookstein coordinate

$$x_i'=\frac{(x_n-x_1)(x_i-x_1)+(y_n-y_1)(y_i-y_1)}{(x_n-x_1)^2+(y_n-y_1)^2}-\frac{1}{2},\ y_i'=\frac{(x_n-x_1)(y_i-y_1)+(y_n-y_1)(x_i-x_1)}{(x_n-x_1)^2+(y_n-y_1)^2}$$

Matching two curves in the Bookstein coordinate is invariant to translation, rotation and scaling.



Shape tree

To match curve $c_1=(x_k^1)_{k=1}^n$ and $c_2=(x_k^2)_{k=1}^m$, we fix the shape tree of $c1$, and look for a mapping from the points in $c2$ to the points in $c1$, which minimizes the deformation cost

$$d(c_1,c_2)=\min_{1<j<m}\left\{d(c_1^L,c_2^L)+d(c_1^R,c_2^R)+\alpha\left|B(x_i^1\mid x_1^1,x_n^1)-B(x_j^2\mid x_1^2,x_m^2)\right|\right\}$$

where x_i^1 is the midpoint of $c1$, x_j^2 is the midpoint of $c2$, $d(c_1^L,c_2^L)$ is the matching cost of the left subtrees of $c1$ and $c2$, and $d(c_1^R,c_2^R)$ is the matching cost of the right subtrees of $c1$ and $c2$.

3. Shape preserving seam carving

The energy of a seam is

$$E(\boldsymbol{p})=E_s(\boldsymbol{p})+\beta E_d(\boldsymbol{p})+\gamma E_t(\boldsymbol{p})$$

where *Es* is the visual saliency term, *Ed* is the deformation term and *Et* is the temporal term.

Visual saliency term *Es* is the sum of the saliency intensities of seam $\boldsymbol{p}$ in the saliency map *Ms*:

$$E_s(\boldsymbol{p})=\sum_{j=1}^{H}M_s(p_j,j)$$

The deformation term *Ed* is the sum of deformation costs of corresponding curves after removing seam $\boldsymbol{p}$:

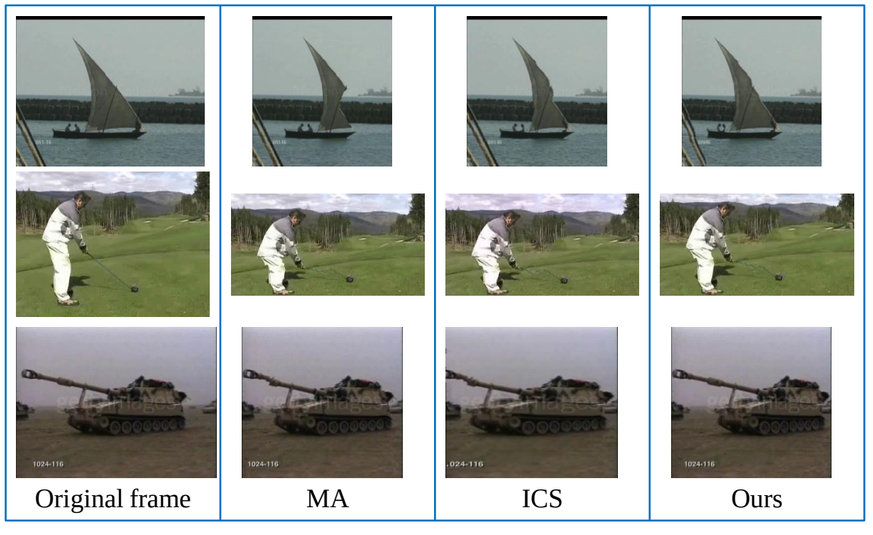
$$E_d(\boldsymbol{p})=\sum_{i=1}^{N}d(c_i,c_i')$$

where c_i' is the deformed curve of c_i. The temporal term *Et* penalizes the variation of relative positions of curves.

The optimal seam is obtained via dynamic programming.

$$\boldsymbol{p}^*=\arg\min_p\ E_s(\boldsymbol{p})+\beta E_d(\boldsymbol{p})+\gamma E_t(\boldsymbol{p})$$

4. Experimental results



Original frame MA ICS Ours

Psycho-visual evaluation result

	Ours	ICS	MA
Ours	-	17	15
ICS	59	-	31
MA	61	45	-
Total	120	62	46
Prefer	78.9%	40.8%	30.3%

ICS: M. Rubinstein et al., "Improved seam carving for video retargeting," SIGGRAPH, 2008.
MA: Y. Bo et al., "Matching area based seam carving for video retargeting," *TCSVT*, 2008.