Selectively De-Animating Video

Jiamin Bai, Aseem Agarwala, Maneesh Agrawala, Ravi Ramamoorthi

SIGGRAPH 2012

Inspiration

http://cinemagraphs.com/

Cinemagraphs

Input Video

Final Result

De-Animating Video

Input Video

Final Result

Example Walkthrough

Input Video

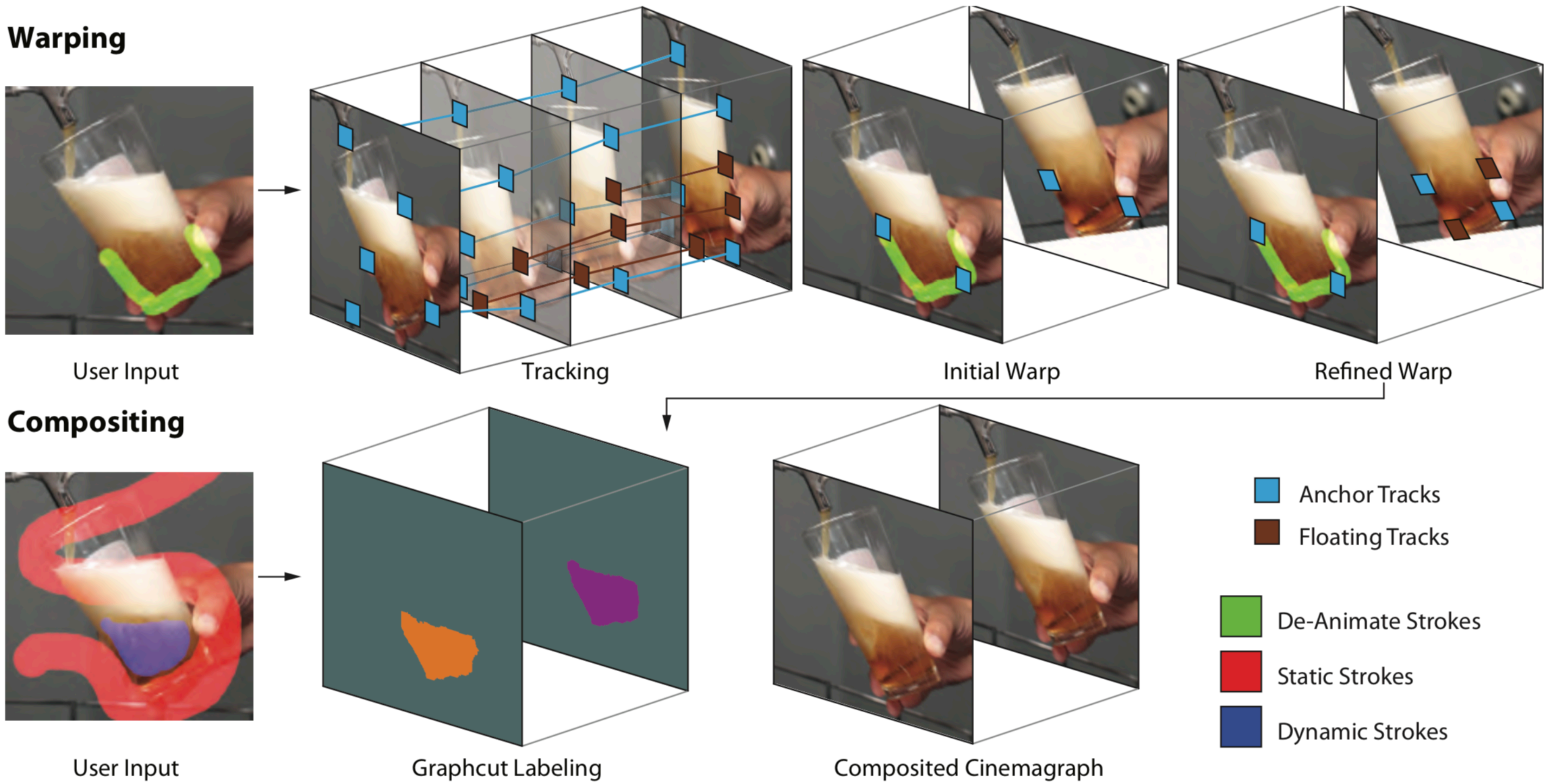
Example Walkthrough

Cinemagraphs

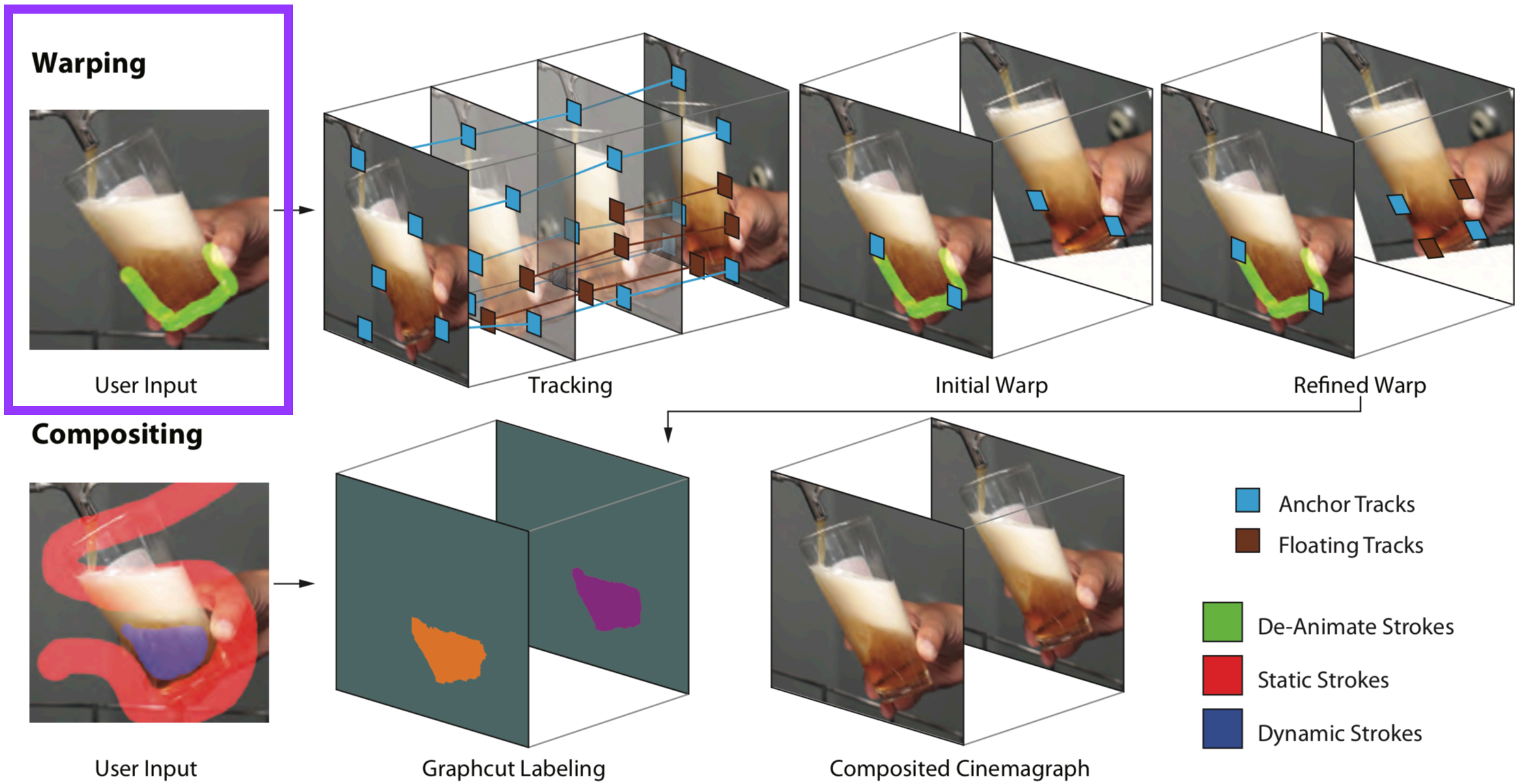
Input Video

Final Result

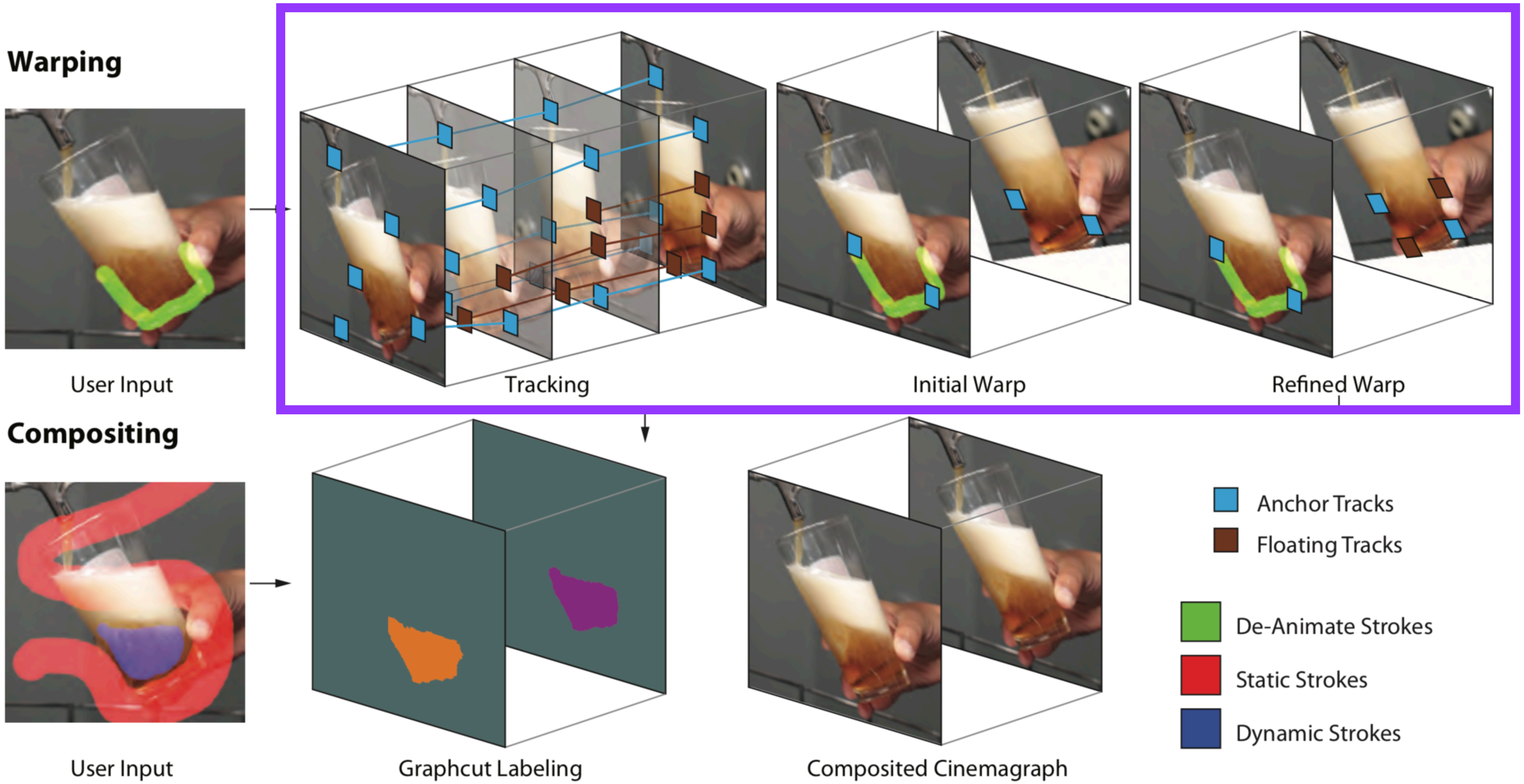
System Diagram



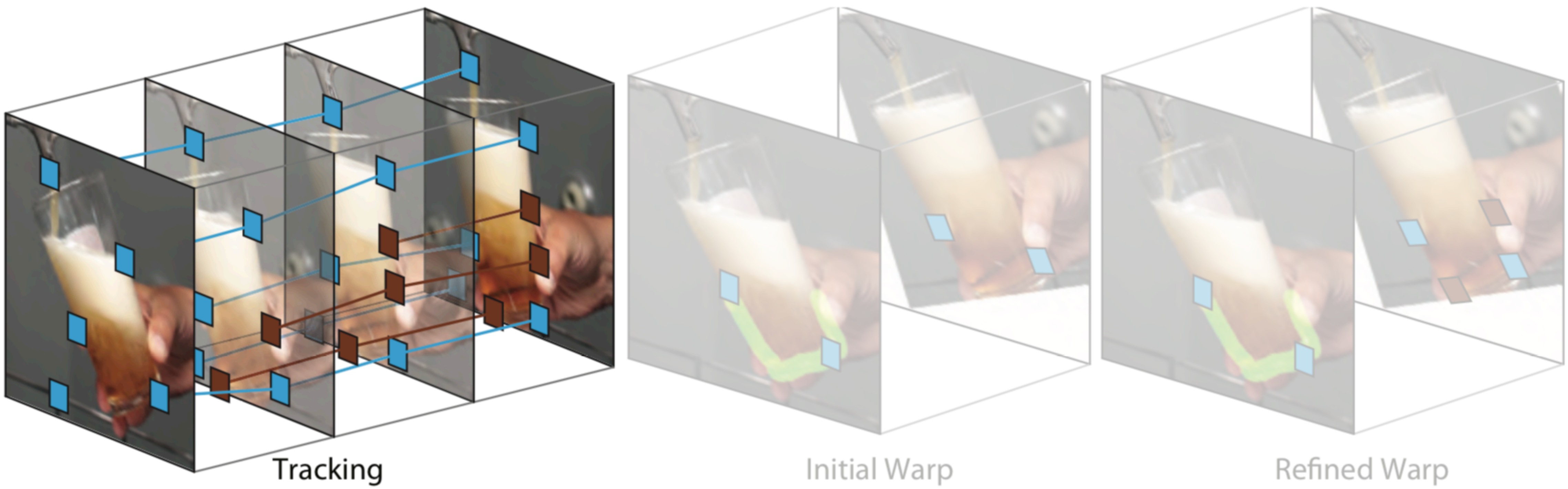
System Diagram



System Diagram



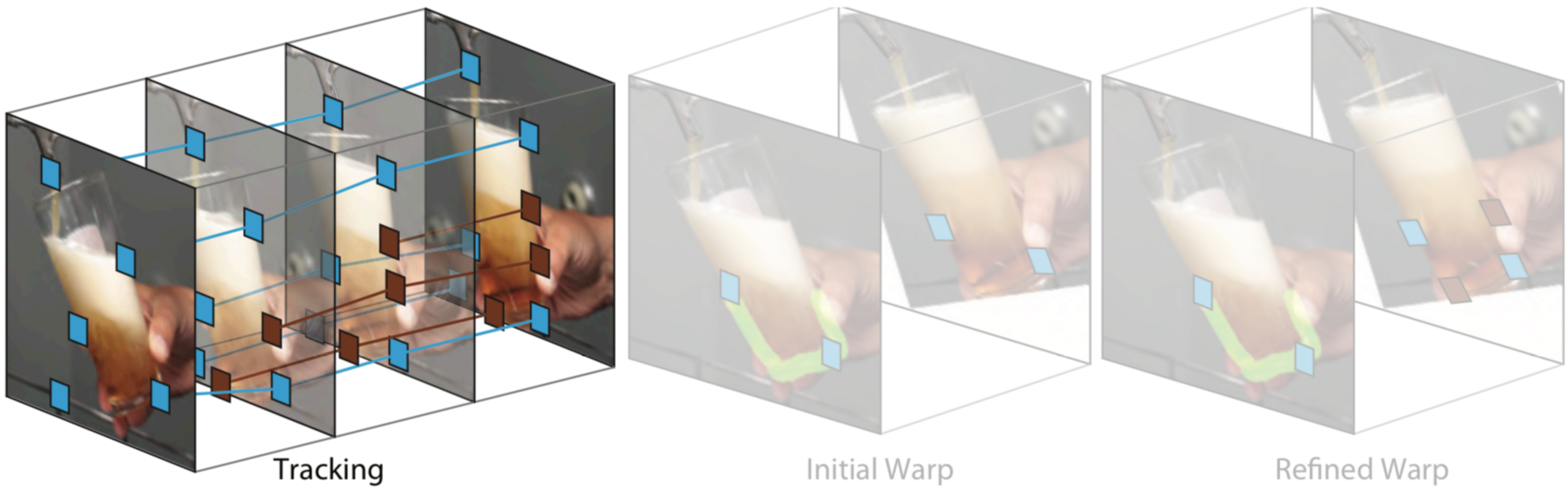
Warping: Tracking



K (s, t) = set of tracks as a table of 2D coordinates

s = track index
t = time (frame number)

Warping: Tracking



$K(s, t)$ = set of tracks as a table of 2D coordinates

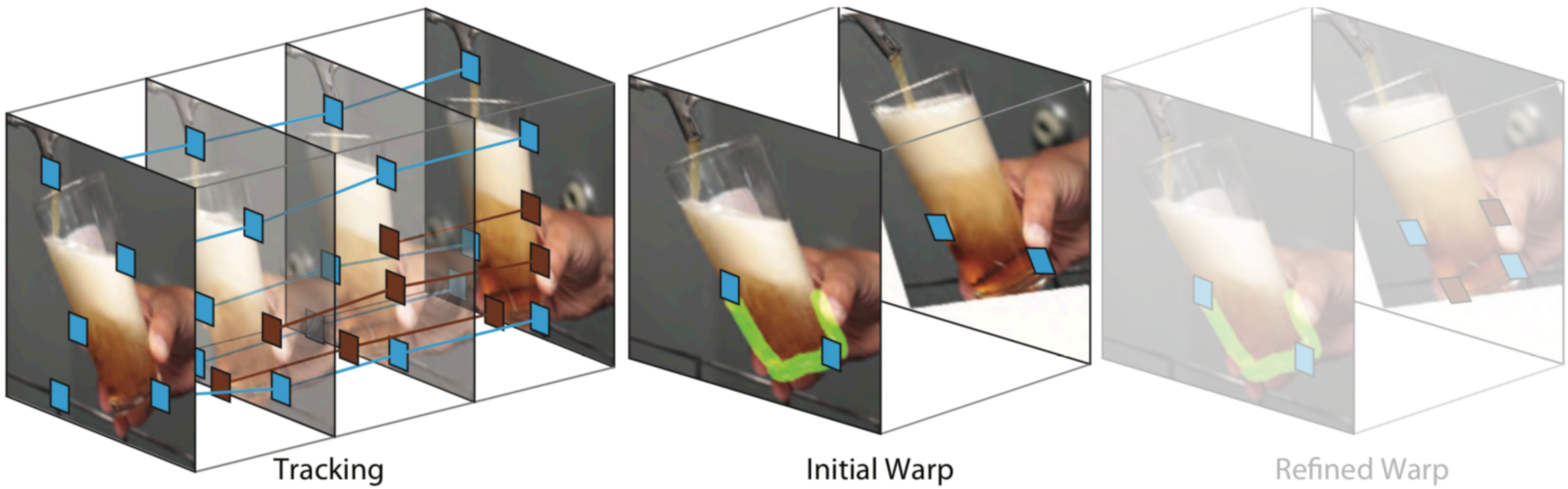
$K_G(s, t)$ = subset of tracks that lie on the user indicated region

$K'_G(s, t)$ = locations of tracks after warping

$K'_G(s, t) = K_G(s, t_a)$

s = track index
t = time (frame number)
t_a = reference frame

Warping: Tracking



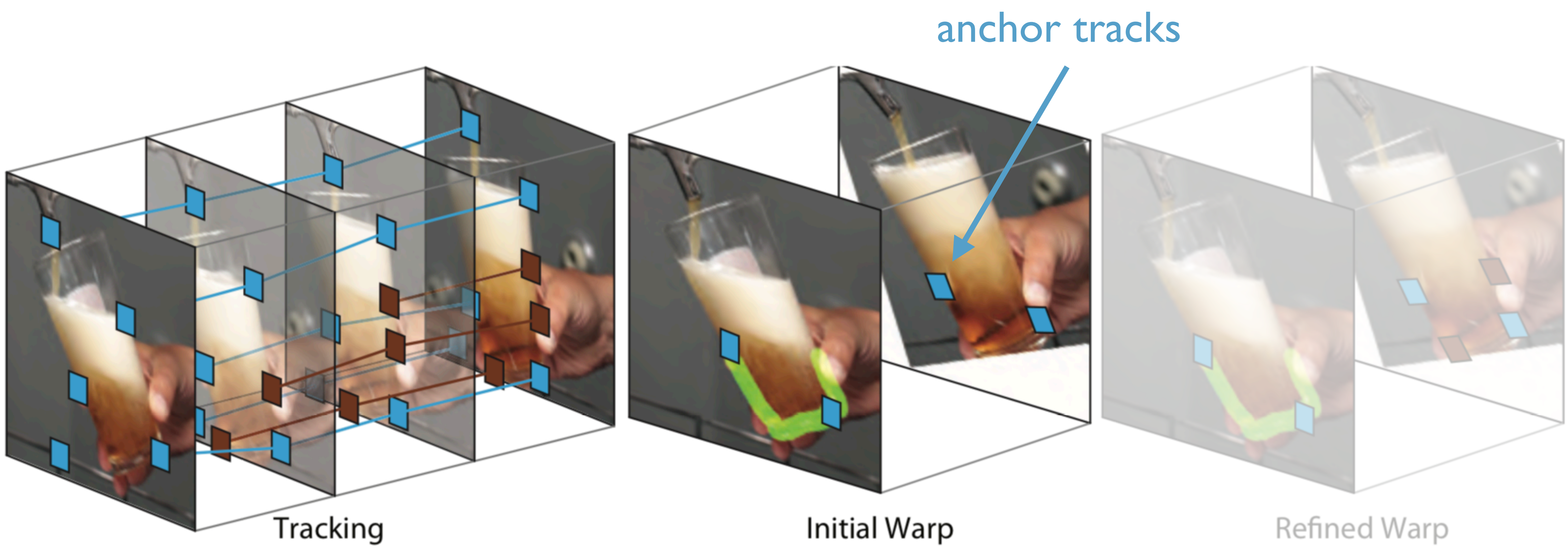
$K(s, t)$ = set of tracks as a table of 2D coordinates

$K_G(s, t)$ = subset of tracks that lie on the user indicated region
$K'_G(s, t)$ = locations of tracks after warping
$K'_G(s, t) = K_G(s, t_a)$

$K_G = K_A$

s = track index
t = time (frame number)
t_a = reference frame

Warping: Tracking



$K(s, t)$ = set of tracks as a table of 2D coordinates

$K_G(s, t)$ = subset of tracks that lie on the user indicated region

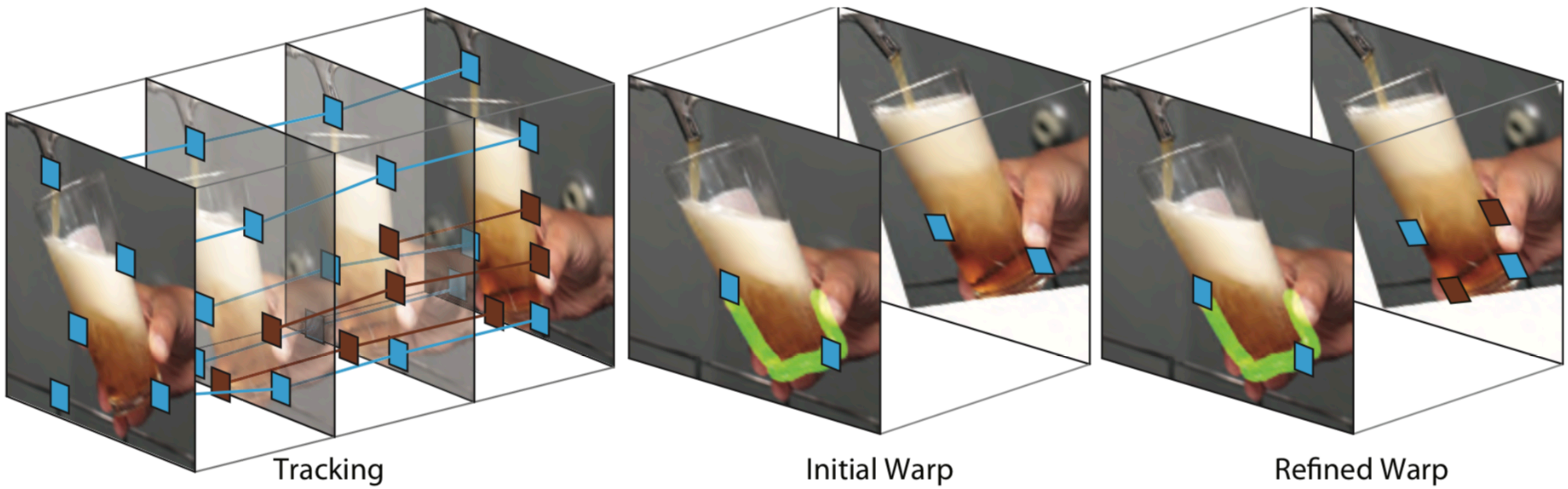
$K'_G(s, t)$ = locations of tracks after warping

$K'_G(s, t) = K_G(s, t_a)$

$K_G = K_A$

s = track index
t = time (frame number)
t_a = reference frame

Warping: Tracking



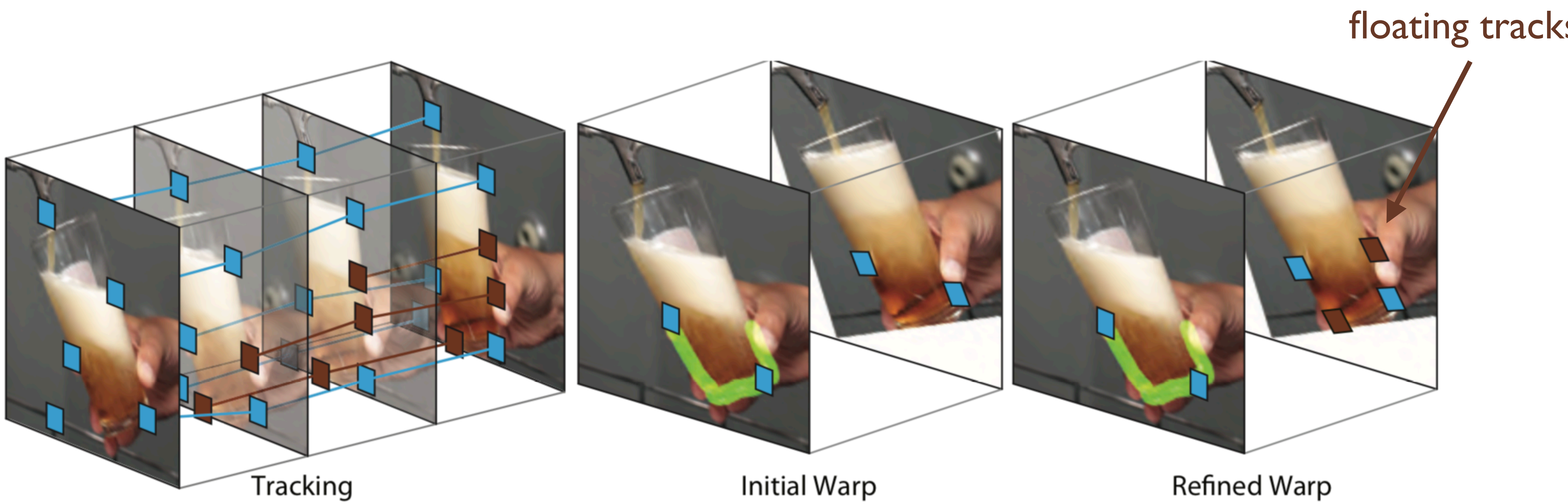
$K(s, t)$ = set of tracks as a table of 2D coordinates

$K_G(s, t)$ = subset of tracks that lie on the user indicated region
$K'_G(s, t)$ = locations of tracks after warping
$K'_G(s, t) = K_G(s, t_a)$

$K_G = K_A \cup K_F$

s = track index
t = time (frame number)
t_a = reference frame

Warping: Tracking



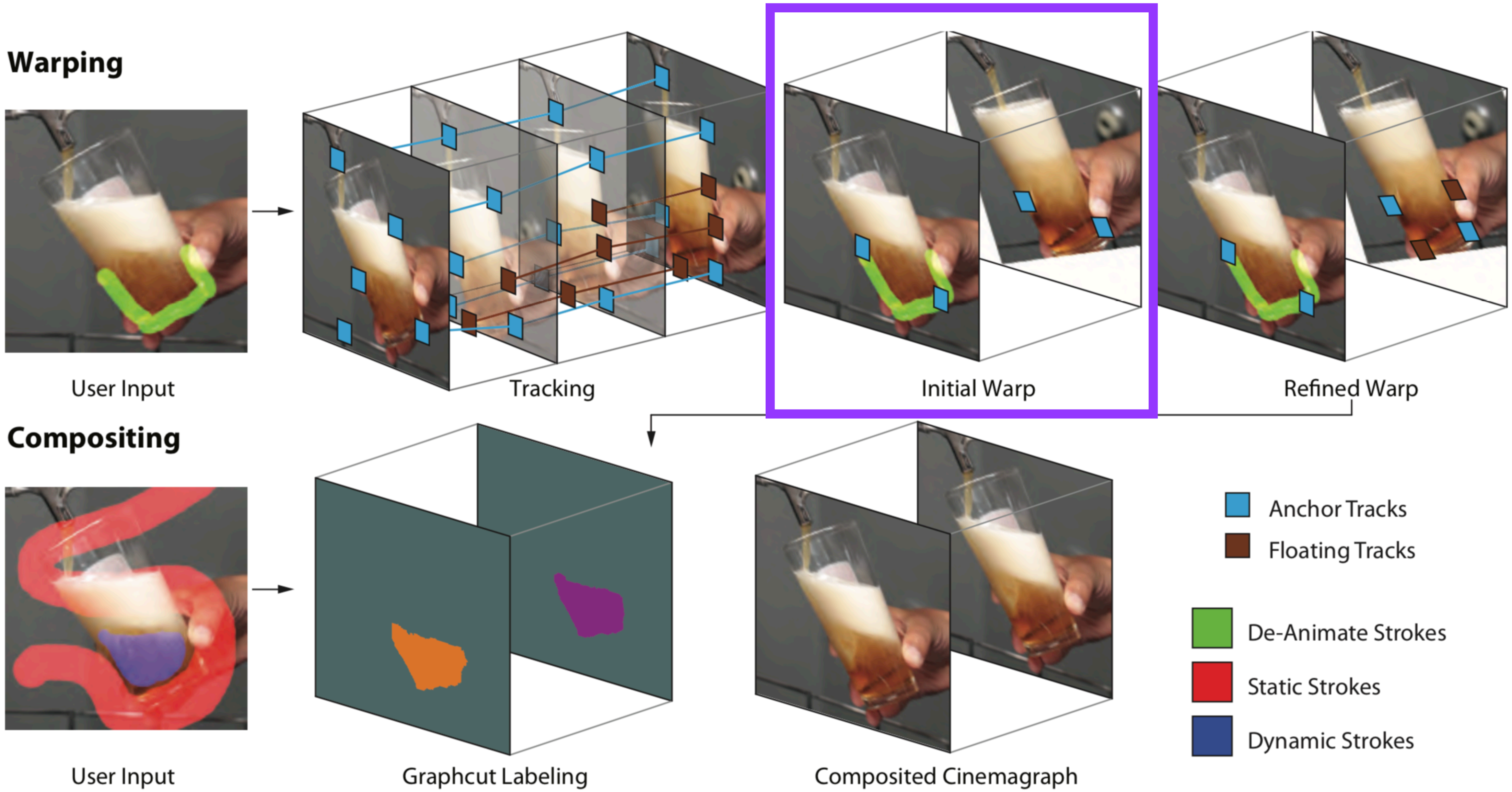
$K(s, t)$ = set of tracks as a table of 2D coordinates

$K_G(s, t)$ = subset of tracks that lie on the user indicated region
$K'_G(s, t)$ = locations of tracks after warping
$K'_G(s, t) = K_G(s, t_a)$

$K_G = K_A \cup K_F$

s = track index
t = time (frame number)
t_a = reference frame

System Diagram



Warping: Initial Warp

$$E = E_a + \omega E_s$$

Warping: Initial Warp

$$E = E_a + \omega E_s$$

main constraint

Warping: Initial Warp

$$E = E_a + \omega E_s$$

shape-preserving

Warping: Initial Warp

$$E = E_a + \omega E_s$$

$$\boxed{E(\hat{V}) = E_d(\hat{V}) + \alpha E_s(\hat{V}),}$$

Warping: Initial Warp

$$E = E_a + \omega E_s$$

$$E_a = \sum_{s \in K_A, t} l(s,t) \left| K_A(s, t_a) - \mathbf{w}(s,t) \cdot \mathbf{V}'(s,t) \right|^2$$

Warping: Initial Warp

$$E = E_a + \omega E_s$$

$$E_a = \sum_{s \in K_A, t} l(s,t) | K_A(s, t_a) - \underbrace{\boxed{\mathbf{w}(s,t) \cdot \mathbf{V}'(s,t)}}_{K'_A(s,t)} |^2$$

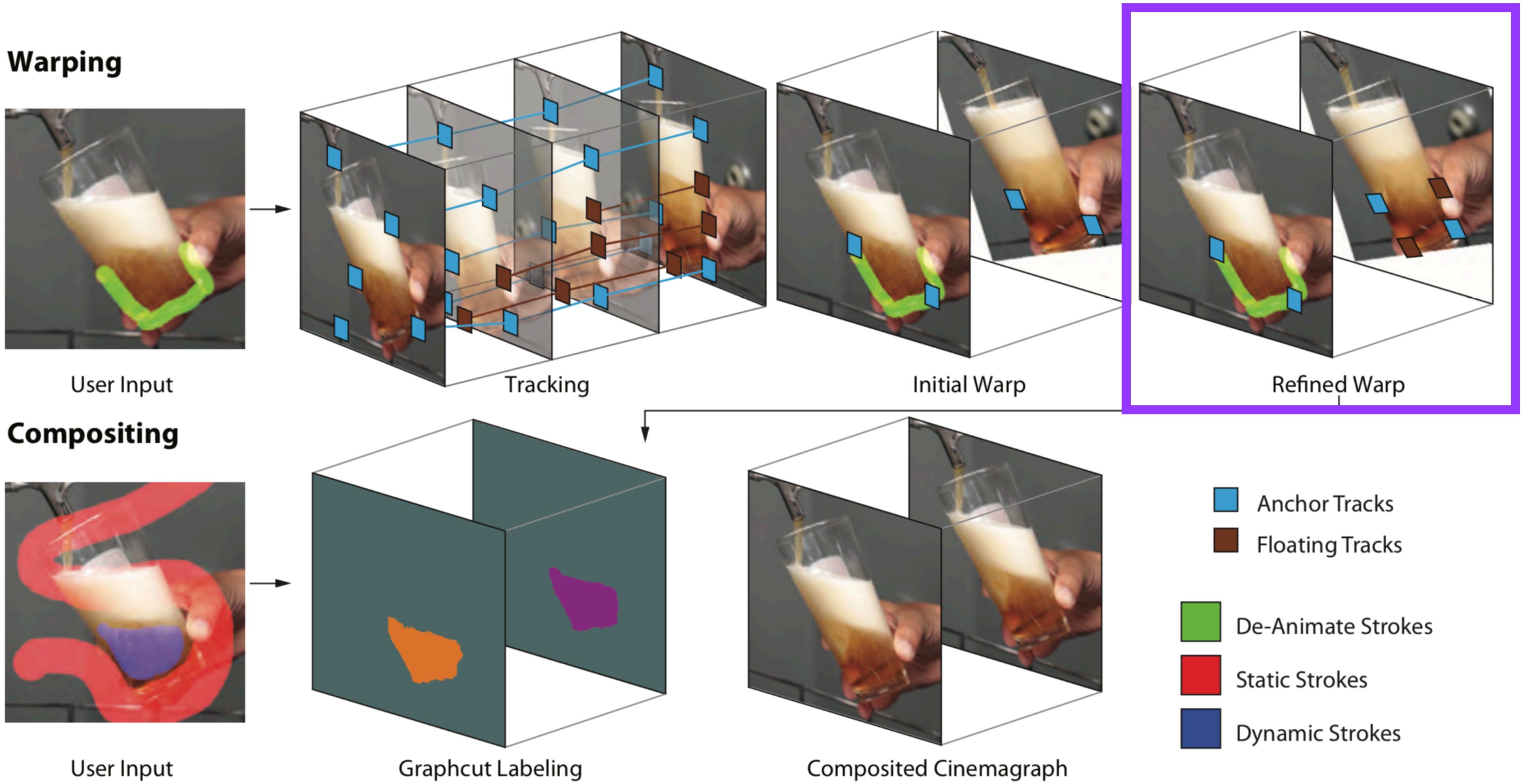
Warping: Initial Warp

$$E = E_a + \omega E_s$$

$$E_a = \sum_{s \in K_A, t} \boxed{l(s,t)} \left| K_A(s, t_a) - \mathbf{w}(s,t) \cdot \mathbf{V}'(s,t) \right|^2$$

weighting function

System Diagram



Warping: Refined Warp

$$E = E_a + E_f + \omega E_s$$

Warping: Refined Warp

$$E = E_a + E_f + \omega E_s$$

$$E_a = \sum_{s \in K_A, t} l(s,t) \left| \boxed{K_A(s, t_a)} - \mathbf{w}(s,t) \cdot \mathbf{V}'(s,t) \right|^2$$

$K'_F\ (s, t_a)$???

Warping: Refined Warp

$$E = E_a + E_f + \omega E_s$$

$$E_f = \sum_{s \in K_F, t} l(s,t) \left| \mathbf{w}(s,t) \cdot \mathbf{V}'(s,t) - \mathbf{w}(s,t{+}1) \cdot \mathbf{V}'(s,t{+}1) \right|^2$$

Warping: Refined Warp

E = E$_a$ + E$_f$ + **ω**E$_s$

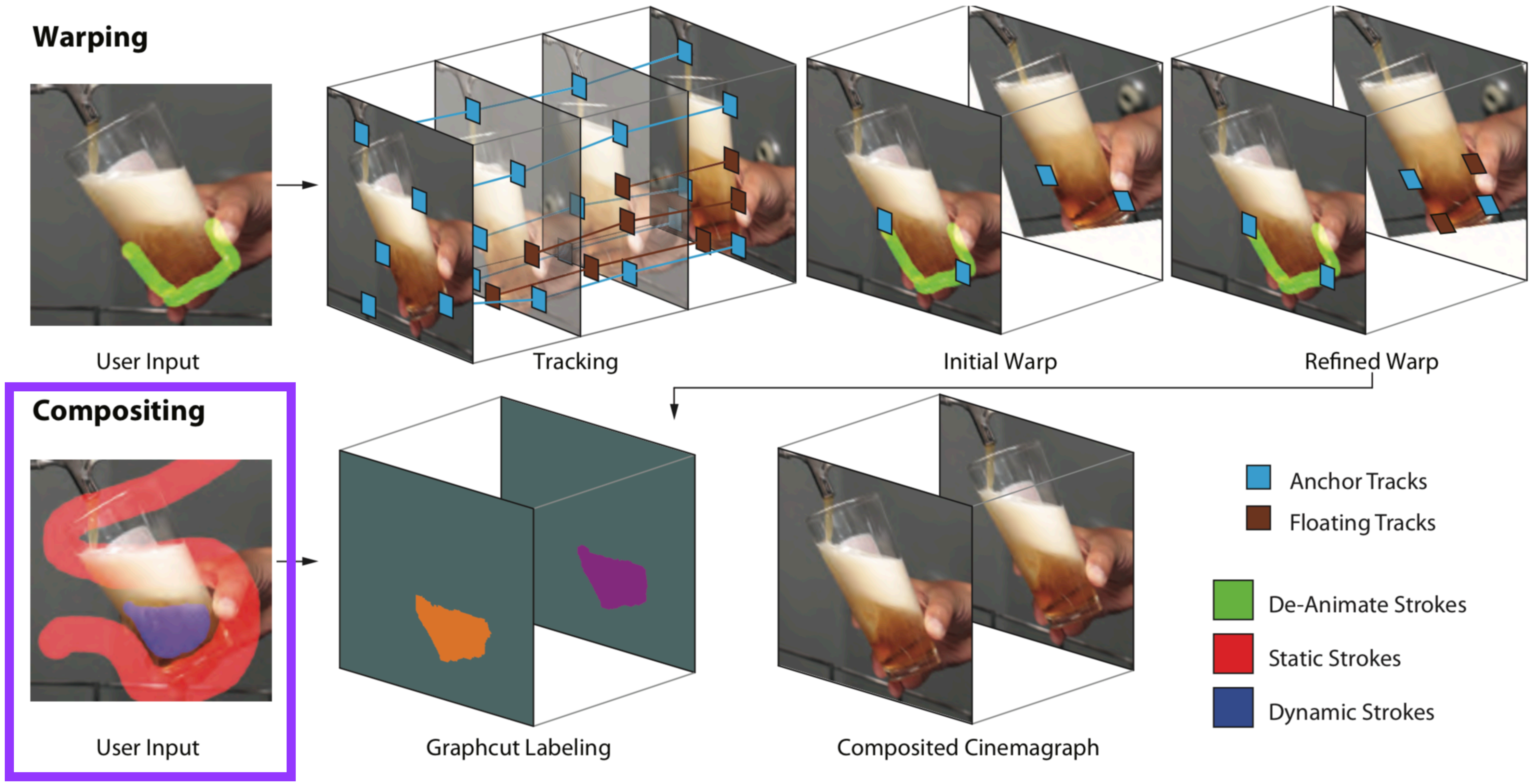
$$E_f = \sum_{s \in K_F, t} l(s,t) \Big| \underbrace{\boxed{\mathbf{w}(s,t) \cdot \mathbf{V}'(s,t)}}_{K'_F(s,\,t)} - \underbrace{\boxed{\mathbf{w}(s,t{+}1) \cdot \mathbf{V}'(s,t{+}1)}}_{K'_F(s,\,t+1)} \Big|^2$$

Warping: Result

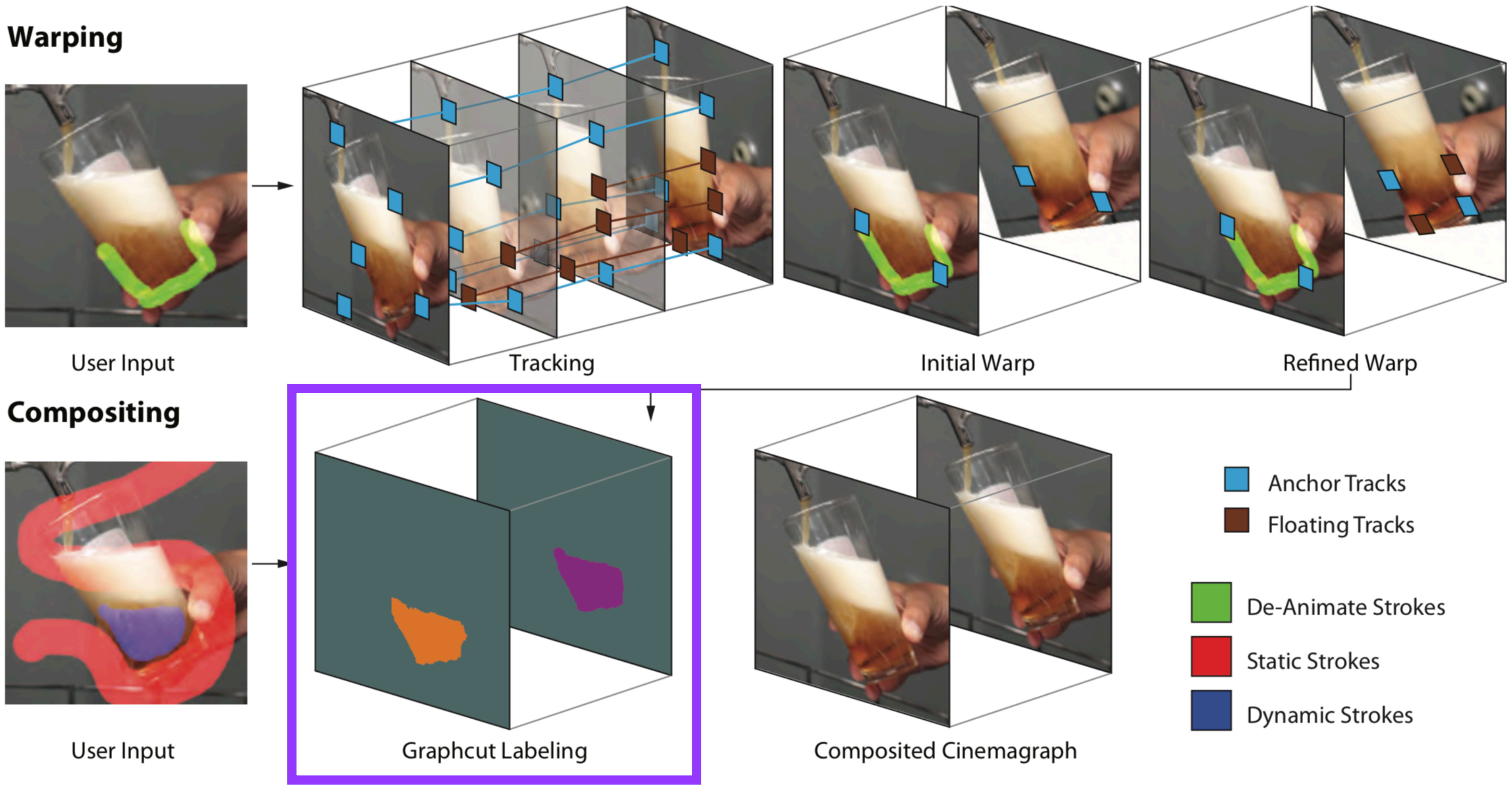
Input Sequence

Our Warped Video

System Diagram



System Diagram



Candidate Video Volumes

Labels L = W U S

Candidate Video Volumes

Labels $L = W \cup S$

dynamic: copies of warped video $W(x, y, t)$

$W = \{W\}$ or $\{W_i, W_j\}$ (if loop seamlessly)

Candidate Video Volumes

Labels $L = W \cup S$

dynamic: copies of warped video $W(x, y, t)$

$W = \{W\}$ or $\{W_i, W_j\}$ (if loop seamlessly)

static: still-frames from input video repeated to fill duration of output

$S = \{I_b, I_{2b}, \ldots\ I_{5b}\}$

b = time interval that evenly samples the input five times
I_b = video where both frame of input video is repeated for duration of output

Candidate Video Volumes

Labels $L = W \cup S$

dynamic: copies of warped video $W(x, y, t)$

$W = \{W\}$ or $\{W_i, W_j\}$ (if loop seamlessly)

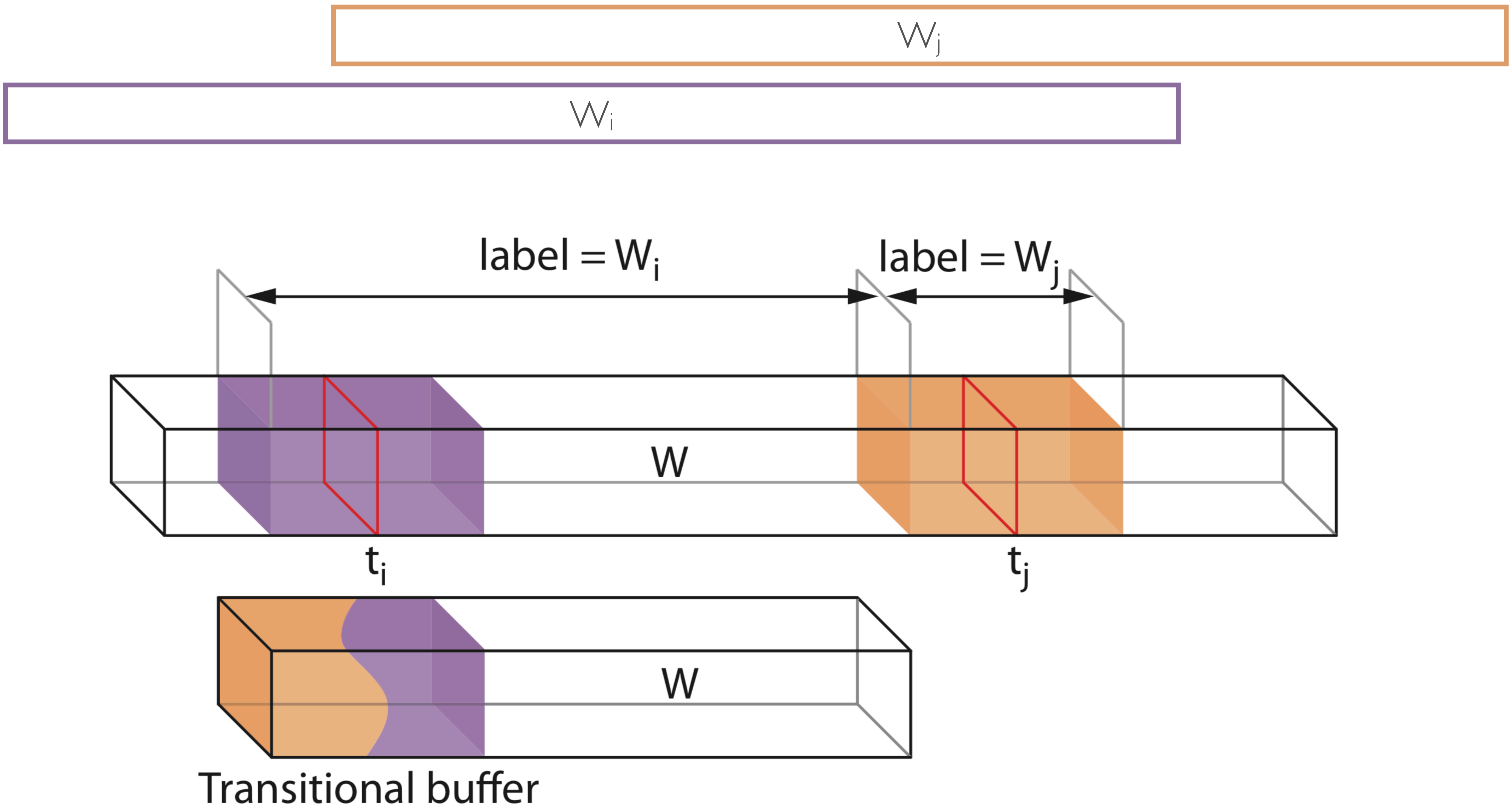
static: still-frames from input video repeated to fill duration of output

$S = \{I_b, I_{2b}, \ldots\ I_{5b}\}$

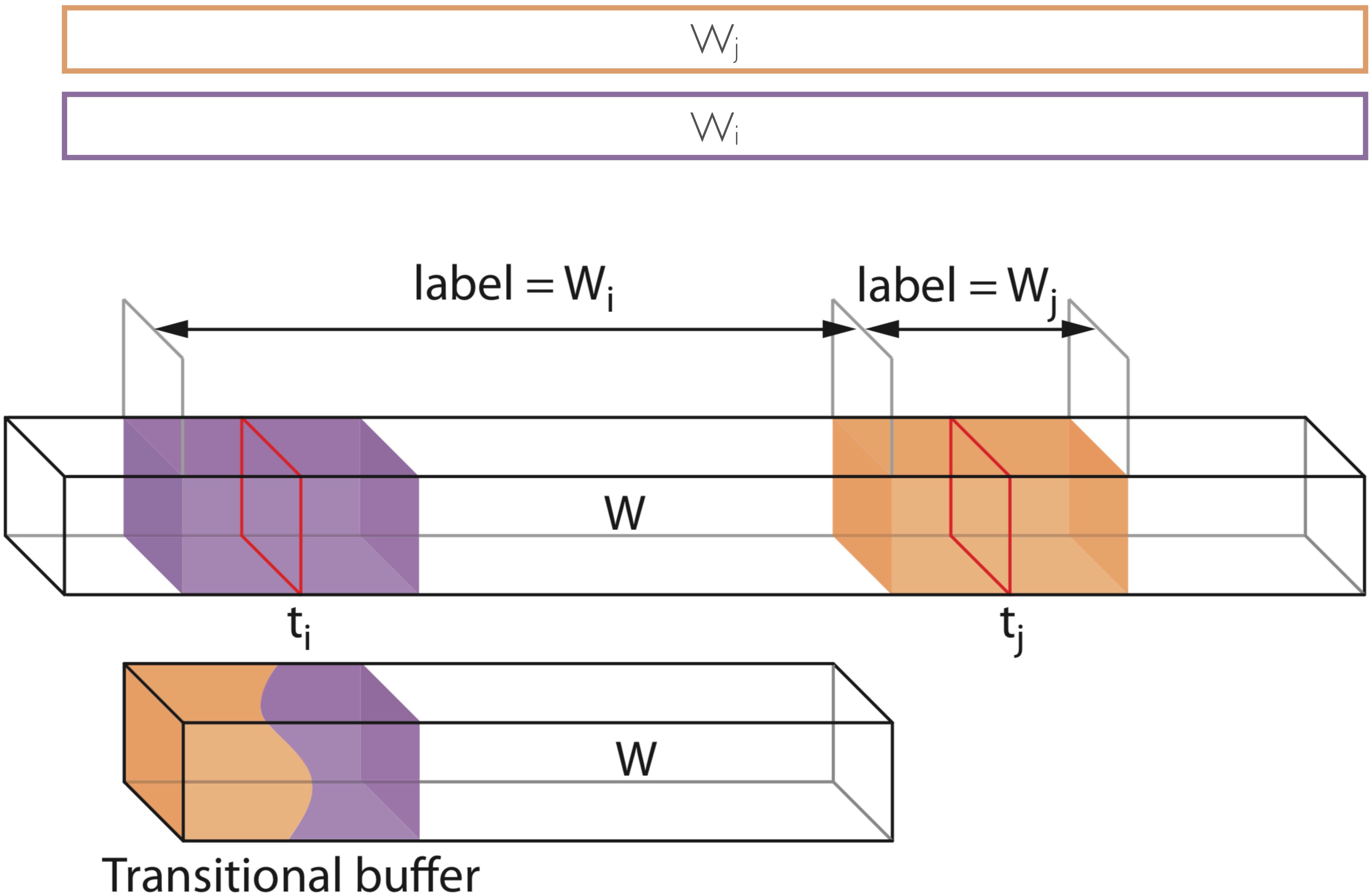
b = time interval that evenly samples the input five times
I_b = video where both frame of input video is repeated for duration of output

or "clean plate"

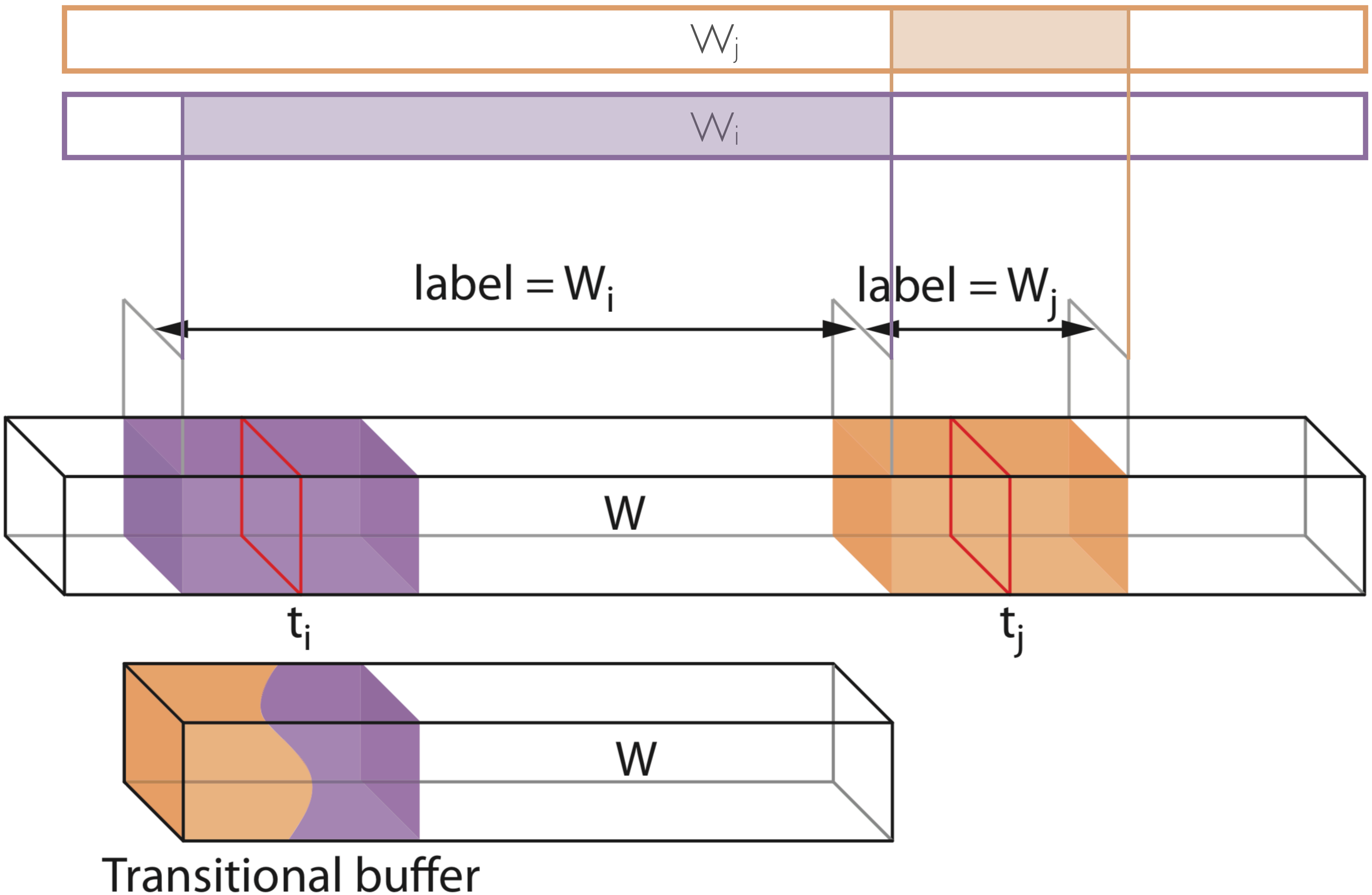
Compositing: Graph-cut



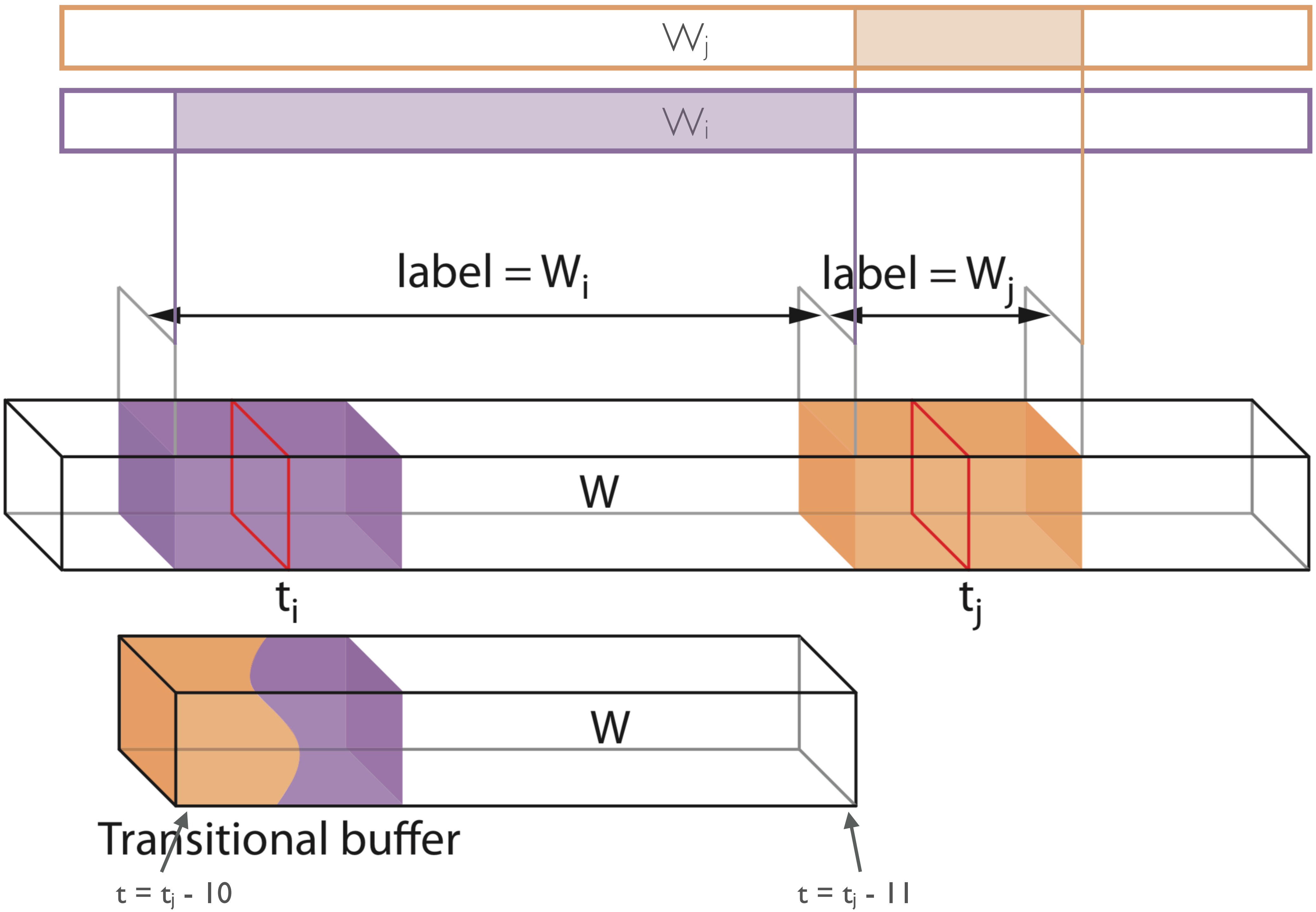
Compositing: Graph-cut



Compositing: Graph-cut



Compositing: Graph-cut



Compositing: Labeling Constraints

Compositing: Labeling Constraints

From user-drawn compositing strokes:

- If $v(x, y) =$ **blue**, $\boldsymbol{\lambda}(x, y, t) \in W$
- If $v(x, y) =$ **red**, $\boldsymbol{\lambda}(x, y, t) \in S$

$v(x, y) =$ strokes $\{$red, blue, NULL$\}$

Compositing: Labeling Constraints

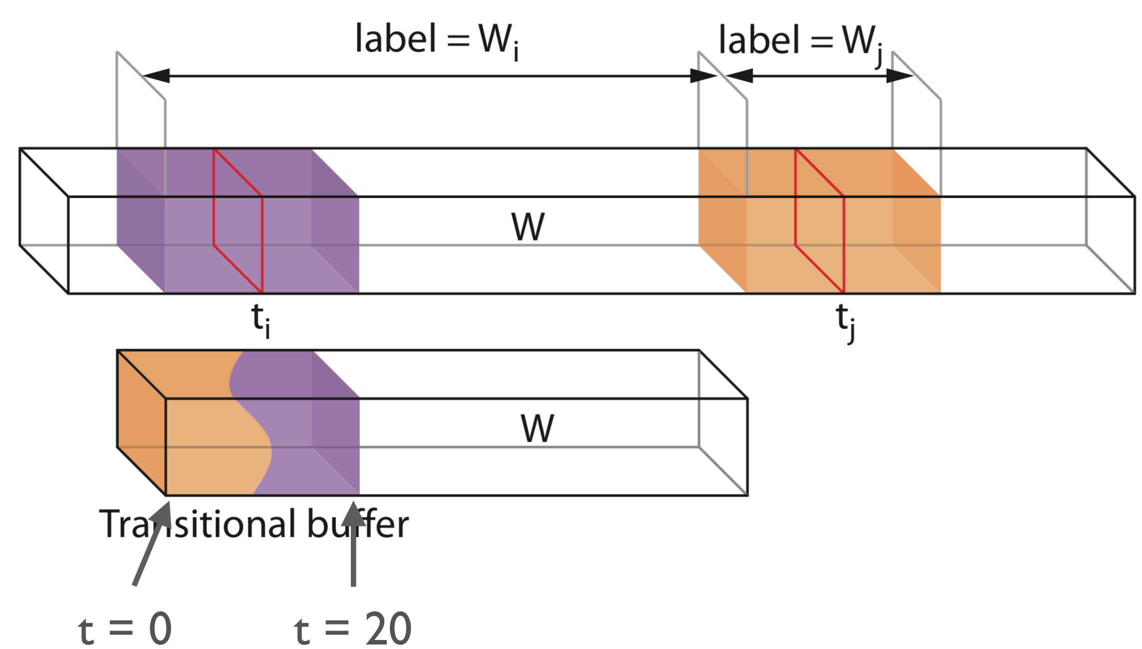
From user-drawn compositing strokes:

- If $v(x, y) =$ **blue**, $\lambda(x, y, t) \in W$
- If $v(x, y) =$ **red**, $\lambda(x, y, t) \in S$

$v(x, y) =$ strokes {red, blue, NULL}

For seamless looping:

- $\lambda(x, y, 0) \neq W_i$
- $\lambda(x, y, 20) \neq W_j$



Compositing: Energy Function

$$\Phi(p_1, p_2, \lambda_1, \lambda_2) = \frac{\gamma(p_1, p_2, \lambda_1, \lambda_2)}{Z(p_1, p_2, \lambda_1, \lambda_2)}$$

Compositing: Energy Function

$$\Phi(p_1, p_2, \lambda_1, \lambda_2) = \frac{\boxed{\gamma(p_1, p_2, \lambda_1, \lambda_2)}}{Z(p_1, p_2, \lambda_1, \lambda_2)}$$

RGB differences

Compositing: Energy Function

$$\Phi(p_1, p_2, \lambda_1, \lambda_2) = \frac{\gamma(p_1, p_2, \lambda_1, \lambda_2)}{\boxed{Z(p_1, p_2, \lambda_1, \lambda_2)}}$$

edge strengths

Compositing: Energy Function

$$\Phi(p_1, p_2, \lambda_1, \lambda_2) = \frac{\boxed{\gamma(p_1, p_2, \lambda_1, \lambda_2)}}{Z(p_1, p_2, \lambda_1, \lambda_2)}$$

RGB differences

$$\begin{aligned}\gamma(p_1, p_2, \lambda_1, \lambda_2) &= |C(p_1, \lambda_1) - C(p_1, \lambda_2)|^2 \\ &+ |C(p_2, \lambda_1) - C(p_2, \lambda_2)|^2\end{aligned}$$

Compositing: Energy Function

$$\Phi(p_1, p_2, \lambda_1, \lambda_2) = \frac{\gamma(p_1, p_2, \lambda_1, \lambda_2)}{Z(p_1, p_2, \lambda_1, \lambda_2)}$$

$$\begin{aligned}\gamma(p_1, p_2, \lambda_1, \lambda_2) &= |C(p_1, \lambda_1) - C(p_1, \lambda_2)|^2 \\ &+ |\boxed{C(p_2, \lambda_1)} - C(p_2, \lambda_2)|^2\end{aligned}$$

color of pixel p_2 in candidate video volume $\lambda(p_1)$

Compositing: Energy Function

$$\Phi(p_1, p_2, \lambda_1, \lambda_2) = \frac{\gamma(p_1, p_2, \lambda_1, \lambda_2)}{\boxed{Z(p_1, p_2, \lambda_1, \lambda_2)}}$$

edge strengths

$$\begin{aligned} \gamma(p_1, p_2, \lambda_1, \lambda_2) &= |C(p_1, \lambda_1) - C(p_1, \lambda_2)|^2 \\ &+ |C(p_2, \lambda_1) - C(p_2, \lambda_2)|^2 \end{aligned}$$

$$Z(p_1, p_2, \lambda_1, \lambda_2) = \begin{cases} \sigma(p_1, p_2, \lambda_1) & \lambda_1 \in \mathbf{W} \wedge \lambda_2 \in \mathbf{S} \\ \sigma(p_1, p_2, \lambda_2) & \lambda_1 \in \mathbf{S} \wedge \lambda_2 \in \mathbf{W} \\ \frac{1}{2}[\sigma(p_1, p_2, \lambda_1) + \sigma(p_1, p_2, \lambda_2)] & \text{Otherwise} \end{cases}$$

Compositing: Energy Function

$$\Phi(p_1, p_2, \lambda_1, \lambda_2) = \frac{\gamma(p_1, p_2, \lambda_1, \lambda_2)}{\boxed{Z(p_1, p_2, \lambda_1, \lambda_2)}}$$

edge strengths

$$\begin{aligned}\gamma(p_1, p_2, \lambda_1, \lambda_2) &= |C(p_1, \lambda_1) - C(p_1, \lambda_2)|^2 \\ &+ |C(p_2, \lambda_1) - C(p_2, \lambda_2)|^2\end{aligned}$$

$$Z(p_1, p_2, \lambda_1, \lambda_2) =$$

only consider dynamic candidates for seams between dynamic and static

$$\begin{cases} \boxed{\sigma(p_1, p_2, \lambda_1)} & \lambda_1 \in \mathbf{W} \wedge \lambda_2 \in \mathbf{S} \\ \boxed{\sigma(p_1, p_2, \lambda_2)} & \lambda_1 \in \mathbf{S} \wedge \lambda_2 \in \mathbf{W} \\ \frac{1}{2}[\sigma(p_1, p_2, \lambda_1) + \sigma(p_1, p_2, \lambda_2)] & \text{Otherwise} \end{cases}$$

Compositing: Energy Function

$$\Phi(p_1, p_2, \lambda_1, \lambda_2) = \frac{\gamma(p_1, p_2, \lambda_1, \lambda_2)}{\boxed{Z(p_1, p_2, \lambda_1, \lambda_2)}}$$

edge strengths

$$\begin{aligned}\gamma(p_1, p_2, \lambda_1, \lambda_2) &= |C(p_1, \lambda_1) - C(p_1, \lambda_2)|^2 \\ &+ |C(p_2, \lambda_1) - C(p_2, \lambda_2)|^2\end{aligned}$$

$$Z(p_1, p_2, \lambda_1, \lambda_2) = \begin{cases} \sigma(p_1, p_2, \lambda_1) & \lambda_1 \in \mathbf{W} \wedge \lambda_2 \in \mathbf{S} \\ \sigma(p_1, p_2, \lambda_2) & \lambda_1 \in \mathbf{S} \wedge \lambda_2 \in \mathbf{W} \\ \boxed{\frac{1}{2}[\sigma(p_1, p_2, \lambda_1) + \sigma(p_1, p_2, \lambda_2)]} & \text{Otherwise} \end{cases}$$

Compositing: Energy Function

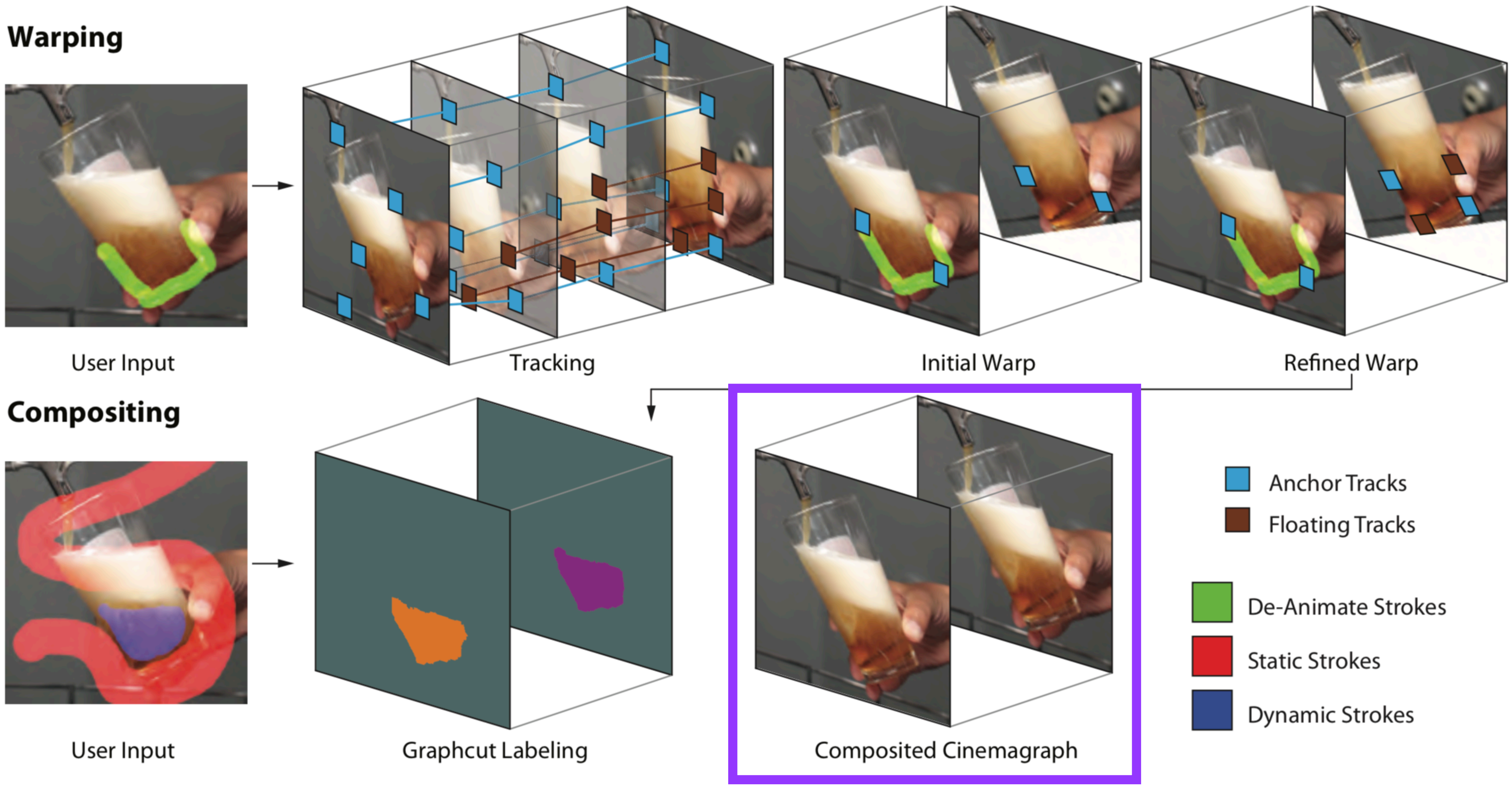
$$\Phi(p_1, p_2, \lambda_1, \lambda_2) = \frac{\gamma(p_1, p_2, \lambda_1, \lambda_2)}{Z(p_1, p_2, \lambda_1, \lambda_2)}$$

$$\begin{aligned}\gamma(p_1, p_2, \lambda_1, \lambda_2) &= |C(p_1, \lambda_1) - C(p_1, \lambda_2)|^2 \\ &+ |C(p_2, \lambda_1) - C(p_2, \lambda_2)|^2\end{aligned}$$

$$Z(p_1, p_2, \lambda_1, \lambda_2) =$$
$$\begin{cases} \sigma(p_1, p_2, \lambda_1) & \lambda_1 \in \mathbf{W} \wedge \lambda_2 \in \mathbf{S} \\ \sigma(p_1, p_2, \lambda_2) & \lambda_1 \in \mathbf{S} \wedge \lambda_2 \in \mathbf{W} \\ \frac{1}{2}[\sigma(p_1, p_2, \lambda_1) + \sigma(p_1, p_2, \lambda_2)] & \text{Otherwise} \end{cases}$$

minimize $\displaystyle\sum_{p1,p2} \Phi(p_1, p_2, \lambda_1, \lambda_2)$

System Diagram



Results: Beer

Results: Model K

Input Video

Final Result

Results: Glass


Input Video


Final Result

Results: Glass

Input Sequence

De-Animate Strokes

Results: Glass

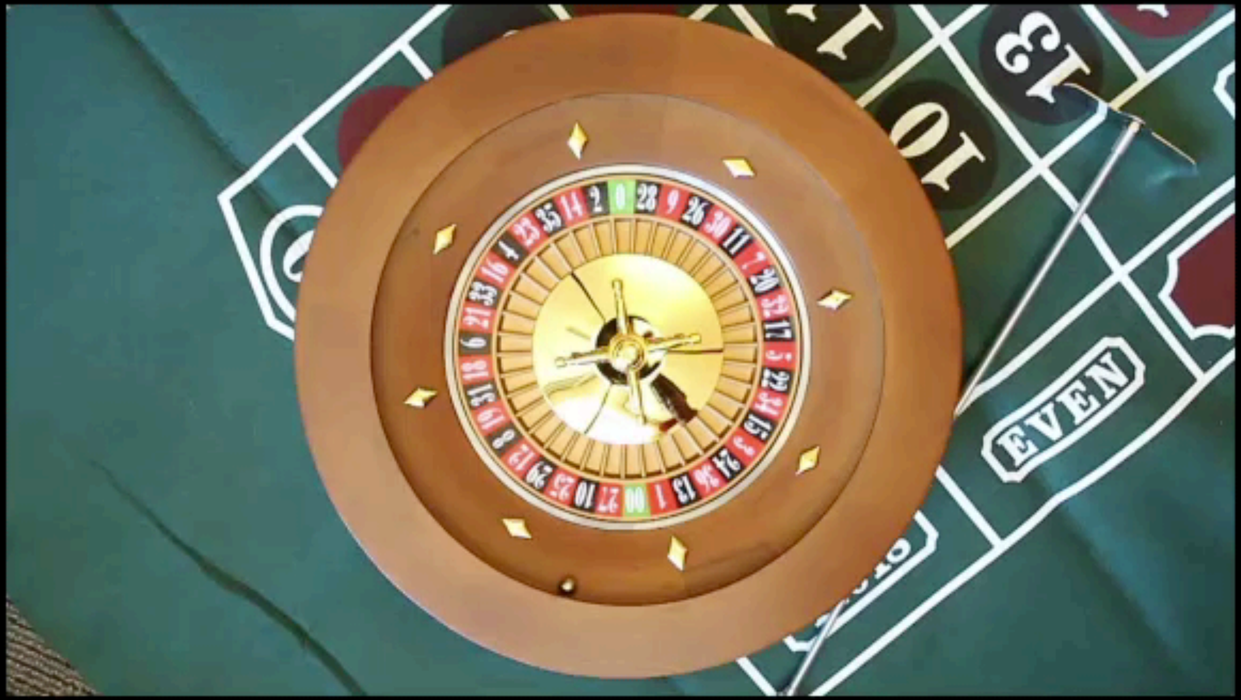
Compositing Strokes

Our Warped Video

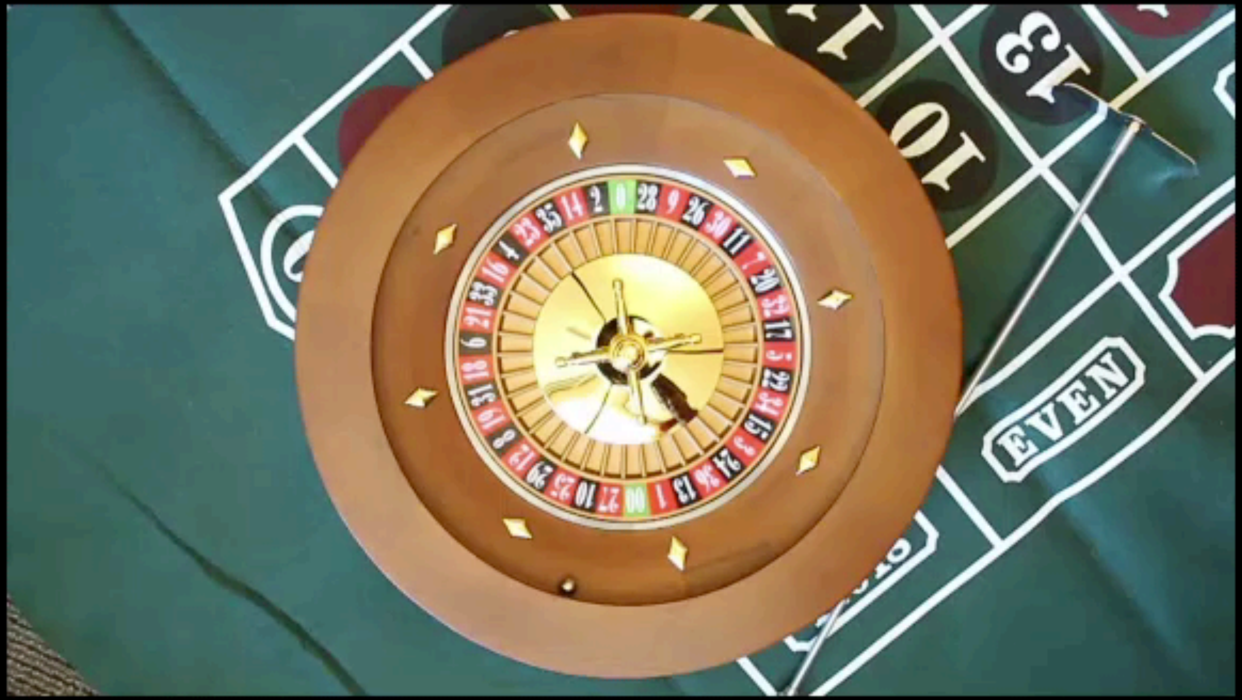
Final Result

Results: Video Editing

Results: Roulette

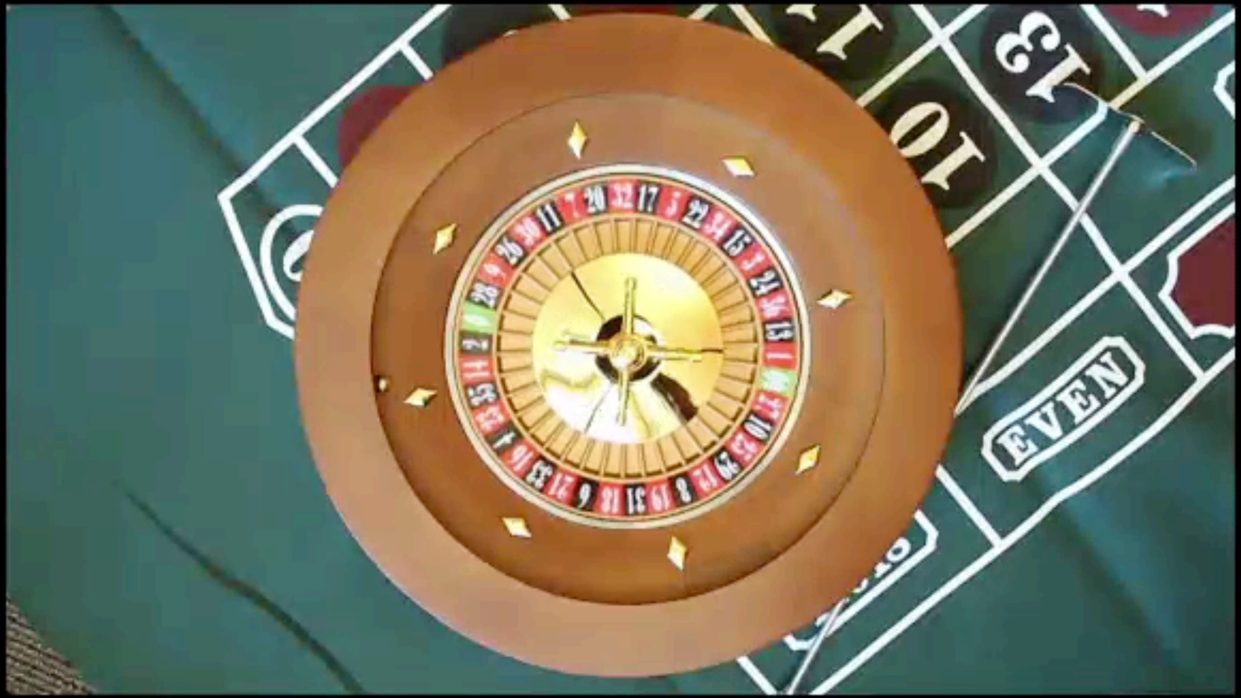


Input Video

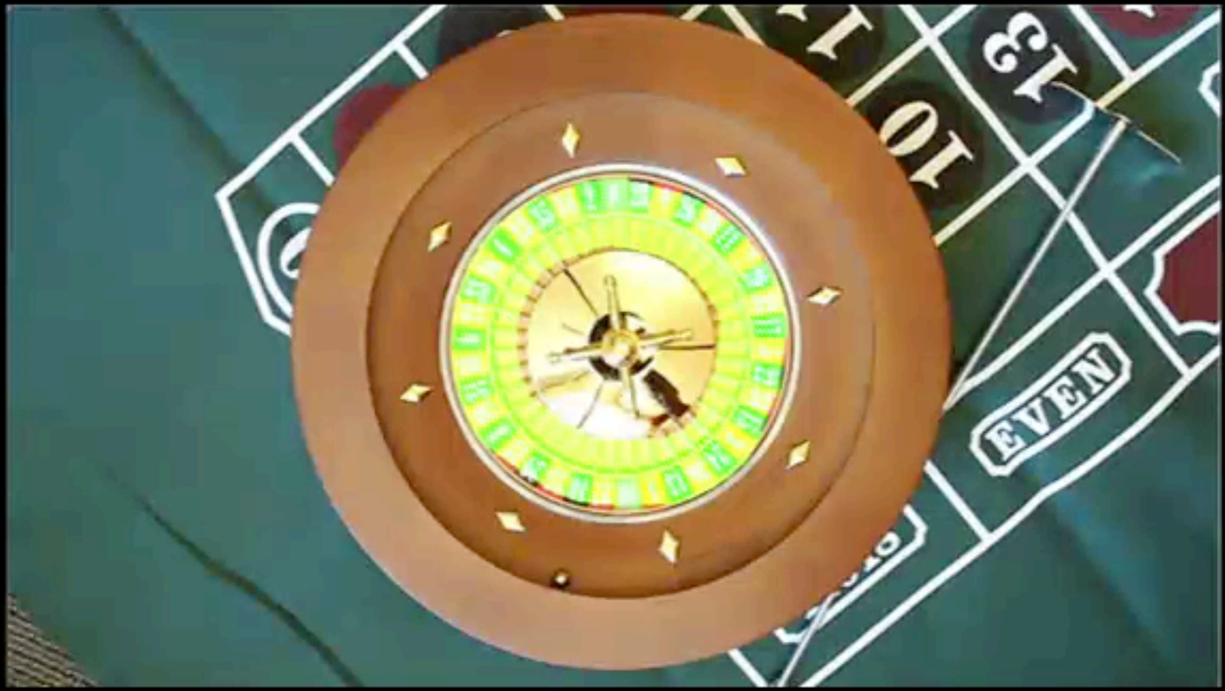


Final Result

Results: Roulette



Input Sequence



De-Animate Strokes

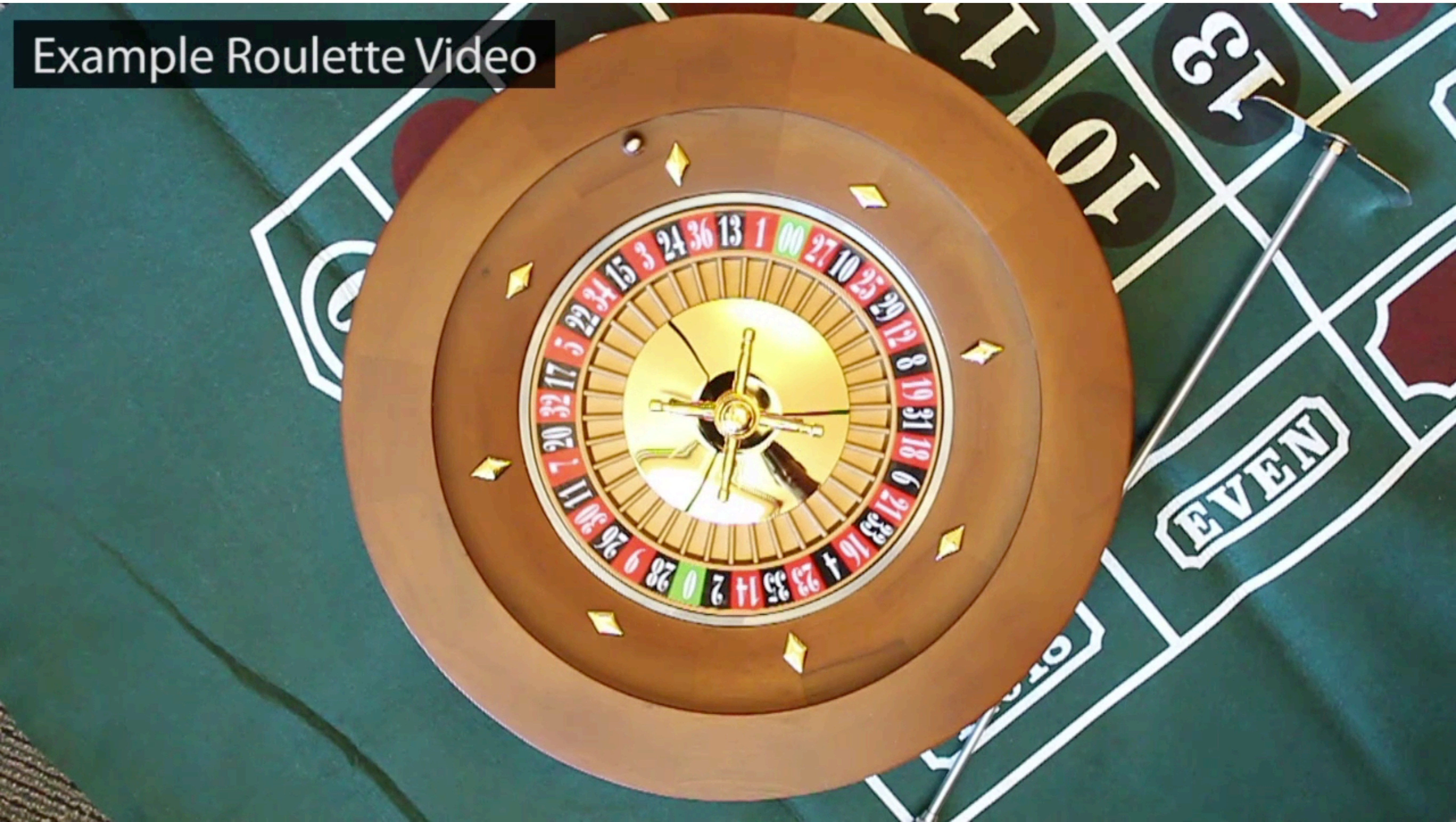
Results: Roulette

Compositing Strokes

Our Warped Video

Final Result

Results: Video Editing



Assumptions

Assumptions

- Input captured with a tripod (or previously stabilized)

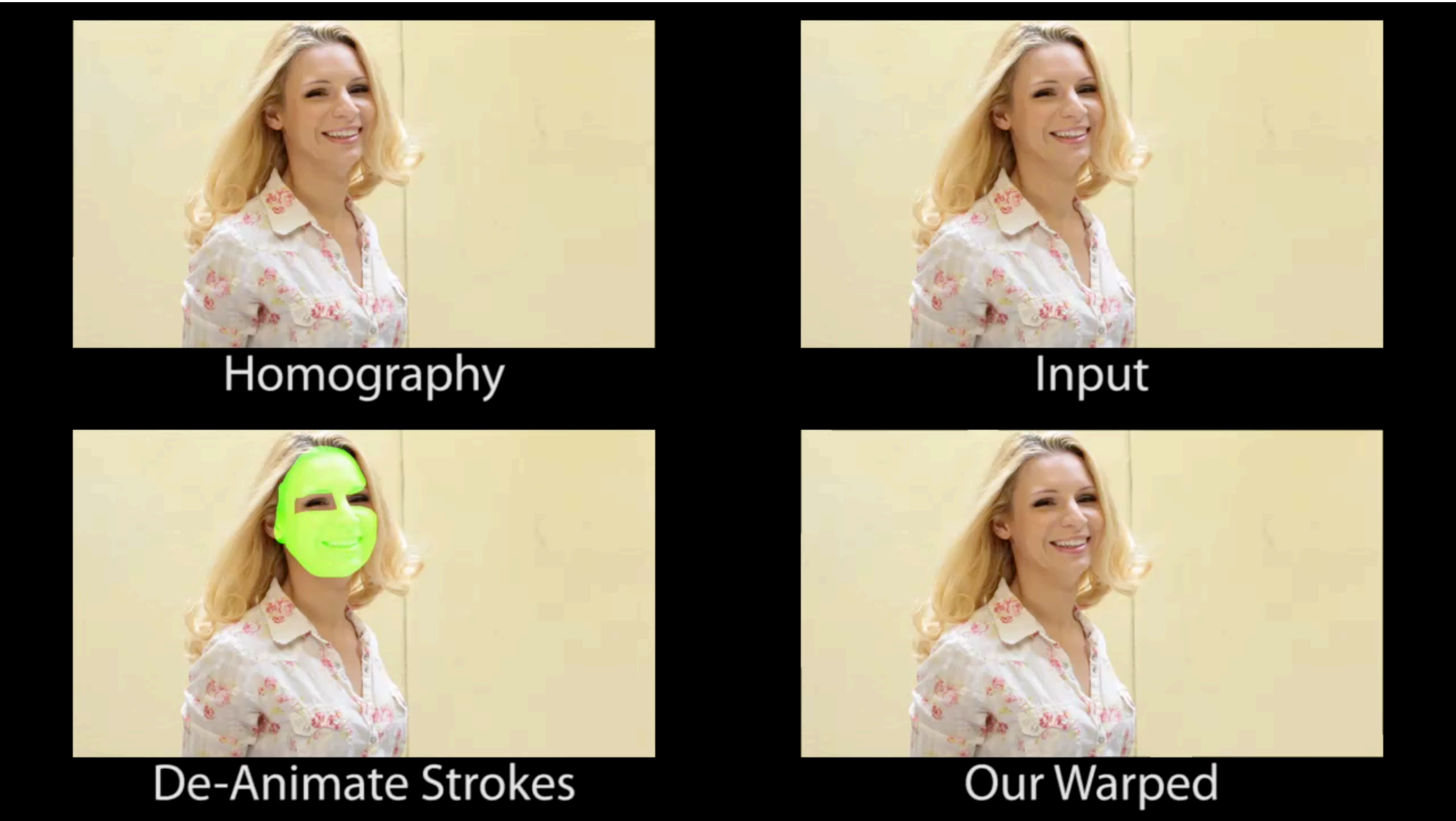
Assumptions

- Input captured with a tripod (or previously stabilized)
- Assume large-scale motions can be be de-animated with 2D warps

Assumptions

- Input captured with a tripod (or previously stabilized)
- Assume large-scale motions can be be de-animated with 2D warps
- Objects to de-animate shot in front of a defocused, uniform, or uniformly-textured background

Limitations: 3D Motion



Limitations: Background

Hard Constraints

GMM Constraints

Limitations

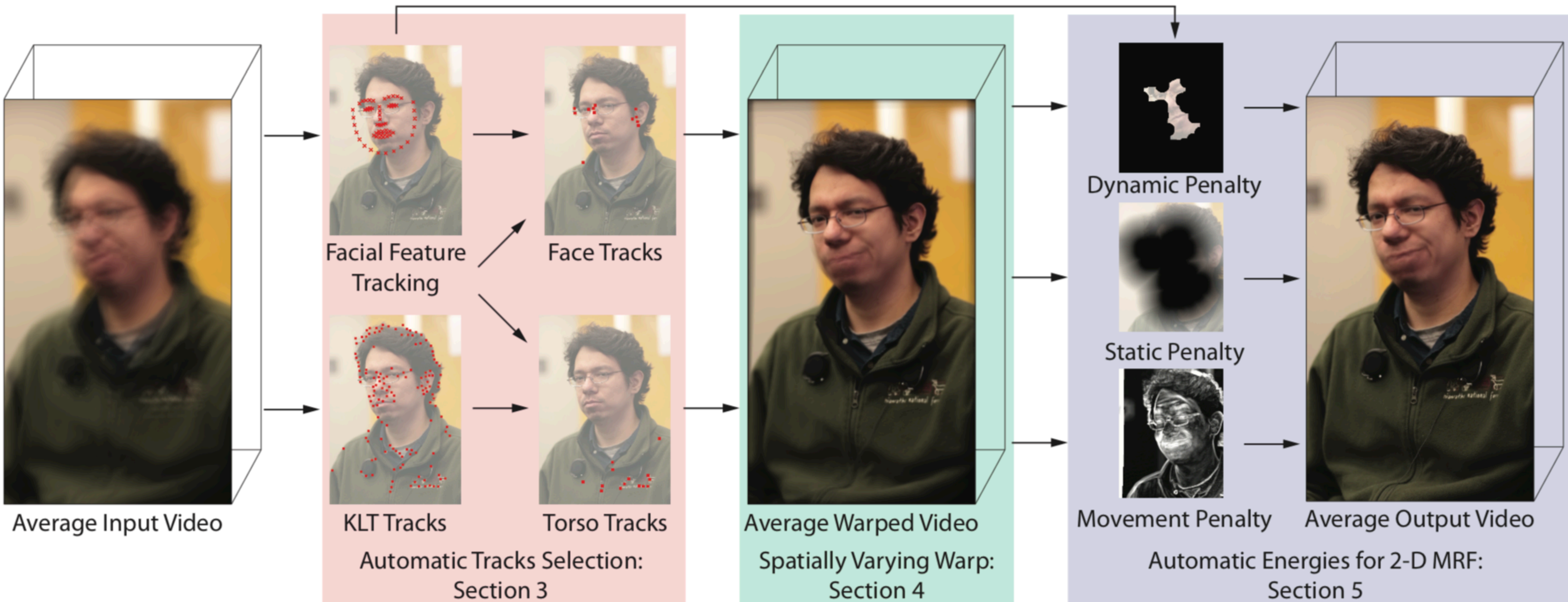
- What happens if the input video is not stabilized?

Follow-up

- This system includes some manual annotation, how would you automate the user input?

- Specifically, what would you do for faces?

Follow-up: Cinemagraph Portraits



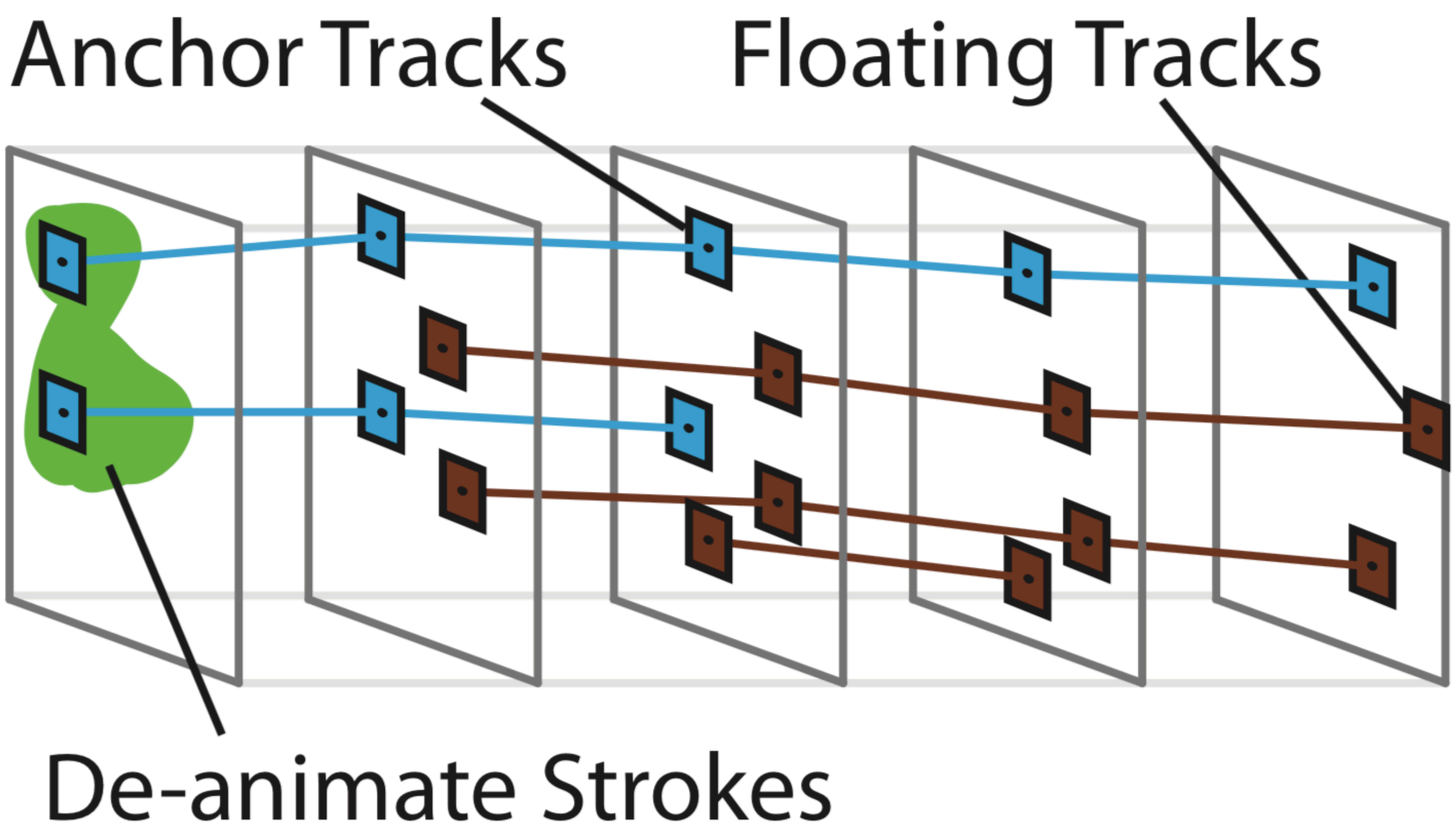
"Automatic Cinemagraph Portraits" Bai et al. EGSR 2013

Selectively De-Animating Video

Jiamin Bai, Aseem Agarwala, Maneesh Agrawala, Ravi Ramamoorthi

SIGGRAPH 2012

Warping: Tracking



Warping: Initial vs Refined

Results: Existing Techniques

Warp Stabilizer

Our Warped Video

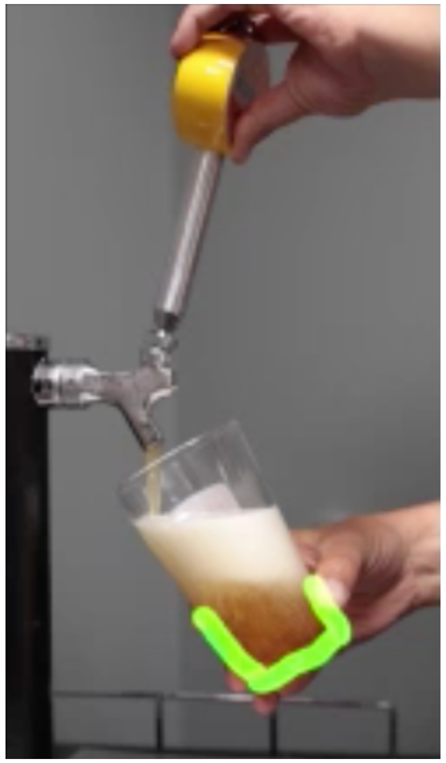
Adapted Cost Function

$$M'(s,t,\mathbf{A},\mathbf{B}) = \frac{M(s,t,\mathbf{A},\mathbf{B})}{\|\mathbf{G}^d_{\mathbf{A}}(s)\| + \|\mathbf{G}^d_{\mathbf{A}}(t)\| + \|\mathbf{G}^d_{\mathbf{B}}(s)\| + \|\mathbf{G}^d_{\mathbf{B}}(t)\|} \quad (5)$$

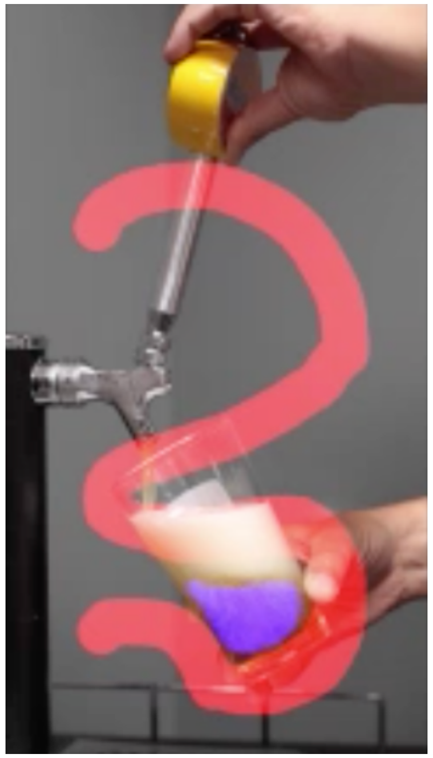
Graph-cut

$$Z(p_1, p_2, \lambda_1, \lambda_2) =$$
$$\begin{cases} \sigma(p_1, p_2, \lambda_1) & \lambda_1 \in \mathbf{W} \wedge \lambda_2 \in \mathbf{S} \\ \sigma(p_1, p_2, \lambda_2) & \lambda_1 \in \mathbf{S} \;\; \wedge \lambda_2 \in \mathbf{W} \\ \frac{1}{2}[\sigma(p_1, p_2, \lambda_1) + \sigma(p_1, p_2, \lambda_2)] & \text{Otherwise} \end{cases}$$

User Input: De-animated Static



de-animate strokes



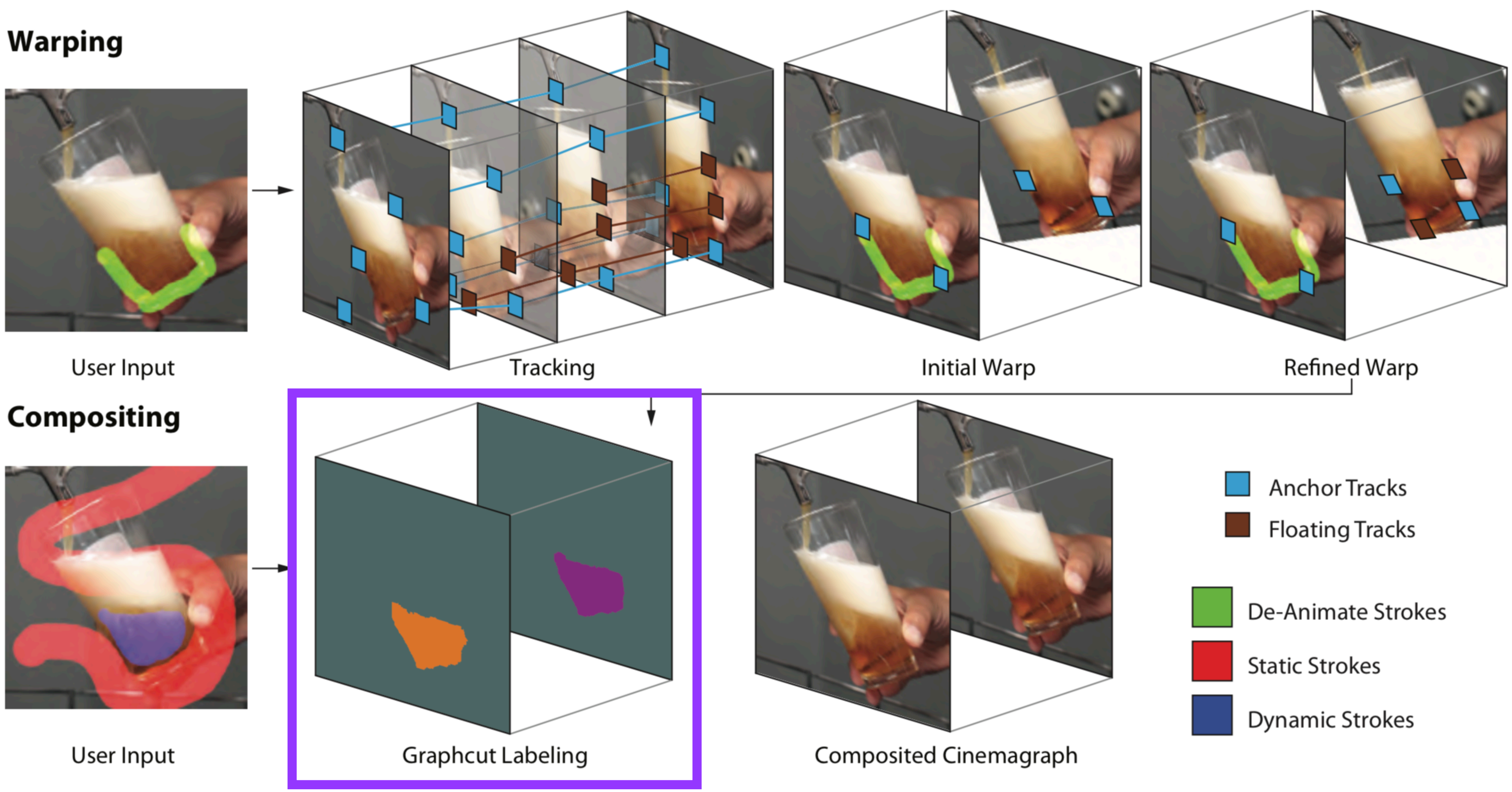
compositing strokes

User Input: De-animated Dynamic

de-animate strokes

compositing strokes

System Diagram



System Diagram